Flux-Balanced Patankar-type Schemes for the Compressible Euler Equations

Thomas Izgin[1,2*], Andreas Meister[1], Chi-Wang Shu[2], Davide Torlo[3]

[1*]Institute of Mathematics, University of Kassel, Heinrich-Plett-Str. 40, Kassel, 34132, Hessen, Germany.
[2]Division of Applied Mathematics, Brown University, 182 George Street, Providence, 02906, Rhode Island, USA.
[3]Department of Mathematics, Università di Roma La Sapienza, Piazzale Aldo Moro 5, Rome, 00185 , Italy.

*Corresponding author(s). E-mail(s): izgin@mathematik.uni-kassel.de, thomas_izgin@brown.edu, ORCID: 0000-0003-3235-210X;
Contributing authors: meister@mathematik.uni-kassel.de;
chi-wang_shu@brown.edu, ORCID: 0000-0001-7720-9564;
davide.torlo@uniroma1.it, ORCID: 0000-0001-5106-1629;

Abstract

Positivity preservation of key physical quantities in the context of fluid flows, such as density and internal energy, is an essential property of a numerical scheme as otherwise the solution lacks physical relevance and has a not well-defined equation of state. One time integration technique that is capable of preserving the positivity of quantities for every time step size is the Patankar-trick and its variants. However, in the context of the Euler equations of gas dynamics, we wonder whether the Patankar-trick should be applied to the density and total energy equations or only to one of them. In this work, we discuss one drawback of the schemes when blindly applied to every positive conserved variable and additionally point out how to overcome the issue by balancing the involved numerical fluxes correctly. To illustrate our findings, we investigate modified Patankar–Runge–Kutta (MPRK) schemes in the context of the compressible Euler equations with and without stiff source terms. We discover that it is beneficial to only apply the Patankar-trick in the density equation and to balance the remaining numerical fluxes consistently rather than applying the trick also to the energy equation. This leads also to the preservation of contact discontinuities. We perform numerical experiments to demonstrate that the accuracy of the methods is maintained while the performance of our approach is superior to the traditional application of MPRK schemes.

1 Introduction

For many natural phenomena, including the flow of fluids such as gases and water, analytic solutions are rarely derivable, and hence, numerical methods are essential for approximating solutions. However, these methods must preserve critical physical properties – positivity (e. g., of density and pressure), conservation, and entropy stability – to ensure both physical relevance and numerical robustness. In particular, failing to maintain positivity (e. g., negative mass or pressure) yields nonphysical solutions, as well as meaningless equations of state, rendering the simulation invalid [1, 2]. Moreover, neglecting conservation properties also result in physically incorrect behavior of the numerical approximation [3]. Often, this requires to evolve conserved quantities, leaving other variables, like pressure in the compressible Euler equations, to be derived using an equation of state (EoS). This makes their control especially challenging during the numerical approximation. There are existing and ongoing works concerning numerical positivity, e. g., based on Rusanov-like fluxes for the Euler equations [4, 5], by using subcell limiters [6, 7], bound preserving filters [8, 9], geometric projection [10], or modified time stepping schemes [11, 12]. However, none of these approaches yields an unconditionally positive method, meaning that there are potentially severe restrictions on the time step size to ensure positivity. To guarantee positivity unconditionally while achieving high accuracy, numerical schemes must be nonlinear [13], i. e. the iterates are described by a nonlinear function even if the PDE is linear. Overall, several strategies have been developed to address the challenges related to unconditional positivity-preservation:

1. *Clipping techniques*, which forcibly set negative values to zero, either result in a mass-shifting optimization problem or otherwise compromise conservation, and, up-to-date, lack a proof of stability [14].
2. *Fully implicit, nonlinear methods* [15, 16] can enforce positivity but require costly iterative solvers, which may fail to converge (to a positive solution), and thus, still produce nonphysical results.
3. *Patankar-type methods* represent a family of explicit or linearly implicit yet nonlinear schemes, which are unconditionally positive [17–21].

The main idea behind Patankar-type schemes is to modify an existing time-stepping method by introducing nonlinear weights in such a way that the resulting numerical scheme becomes unconditionally positive. The primary challenge lies in designing these weights so that the modified scheme preserves the accuracy of the original (baseline) method. This nonlinear modification is achieved using the so-called *Patankar-trick* [17], which gives this family of methods its name. A notable example is the incorporation of modified Patankar (MP) weights into classical Runge–Kutta (RK) schemes, leading to

the development of modified Patankar–Runge–Kutta (MPRK) methods [18, 22], which in addition to being unconditionally positive, are also conservative. Motivated by their strong numerical performance, the Patankar-trick has since been successfully extended to a variety of time integration frameworks, including strong-stability-preserving Runge–Kutta (SSPRK) methods [2, 23], arbitrary high-order Deferred Correction (DeC) schemes [24], and linear multistep methods [25]. The resulting modified schemes all belong to the broader Patankar-type family, which themselves can be recast as non-standard additive Runge–Kutta (NSARK) methods, see [26, 27]. Despite their advantages, analytical results for Patankar-type schemes have been difficult to derive due to their nonlinear structure. It is only recently, in the PhD thesis [27], that rigorous results on Lyapunov stability [28] and high-order accuracy conditions [26] were developed. It is worth mentioning that this Lyapunov stability investigation is the first natural step when searching for reasonable well-balanced [29] schemes as it ensures that initial errors around a constant steady state do not accumulate during the course of the simulation. Indeed, it is the only nonlinear stability analysis currently available for these nonlinear schemes.

Despite progress, controlling the positivity of quantities defined via the EoS (like pressure or internal energy) remains a challenging task. Existing workarounds (e.g., taking absolute values or modifying parameters post hoc) lack reliability and physical grounding [2]. Using algebraic manipulations [30] of the fluxes is a viable solution, but it requires an intrusive modification of existing codes. Further approaches consider giving up strict conservation of certain quantities (e.g., momentum or total energy) in favor of evolving directly primitive variables such as pressure and may lead to easier control on such variables, making use of some correction terms where conservation is needed [31, 32].

In this work, we aim at making progress towards the design of unconditionally positive schemes for balance laws when conservative variables are evolved. In particular, we propose to apply the MP trick only to the density equation and to balance the remaining numerical fluxes consistently. This has the first effect of preserving contact discontinuities with constant pressure and velocity, which is a desirable property for the Euler equations. Moreover, we notice that during the simulations, it allows to maintain positivity of the pressure for larger values of the CFL with respect to classical MP methods, outperforming explicit methods and the traditional application of MPRK schemes.

The paper is organized as follows. We recall the compressible Euler equations with a single as well as with three reacting species in Section 2 and present the classical MPRK schemes in Section 3. Then, we discuss how to apply the MPRK schemes in the presence of stiff source terms in Section 4, proposing a novel flux-balanced version of the momentum and total energy fluxes to guarantee the preservation of contact discontinuities and to improve the performance of the method with respect to classical MPRK schemes. Finally, we perform numerical experiments in Section 5 and summarize our work in Section 6.

3

2 Governing Equations

In this work, we consider the compressible Euler equations of gas dynamics to describe the physical quantities of a single fluid, or of multi-species gas with chemical reactions. The Euler equations describe the evolution of density ρ (or densities ρ_i), the momentum ρu and the total energy ρE of a gas. It is of paramount importance to preserve some physical properties of the gas dynamics such as positivity of the density and internal energy, and preservation of equilibrium states with constant velocity and pressure.

2.1 Euler equations

The Euler equations and their reactive versions in one dimension can be written as hyperbolic balance laws of the type

$$\partial_t \mathbf{U}(x,t) + \partial_x \mathbf{F}(\mathbf{U}(x,t)) = \mathbf{S}(\mathbf{U}(x,t)), \qquad x \in \Omega \subset \mathbb{R}, \, t \in \mathbb{R}^+, \tag{1}$$

where $\mathbf{U} : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}^d$ is the vector of the evolved quantities (when $\mathbf{S} \equiv \mathbf{0}$ these are conserved quantities), $\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^d$ is the flux and $\mathbf{S} : \mathbb{R}^d \to \mathbb{R}^d$ the source term. The compressible Euler equations for a one species gas in a one dimensional domain is defined by

$$\mathbf{U} = (\rho, \rho u, \rho E)^T, \, \mathbf{F}(\mathbf{U}) = (\rho u, \rho u^2 + p, u(\rho E + p))^T \text{ and } \mathbf{S}(\mathbf{U}) = (0,0,0)^T \tag{2}$$

together with the equation of state (EoS) for ideal gas, i.e.

$$p = (\gamma - 1)(\rho E - \tfrac{1}{2}\rho v^2),$$

where we use $\gamma = 1.4$ for air.

2.2 Multi-Species Reactive Euler equations

The three species reactive Euler equations, where two monoatomic gasses react with each other and there is a third diatomic gas, are defined [2, 33] by (1) with

$$\mathbf{U} = (\rho_1, \rho_2, \rho_3, \rho u, \rho E)^T, \, \mathbf{F}(\mathbf{U}) = (\rho_1 u, \rho_2 u, \rho_3 u, \rho u^2 + p, (\rho E + p)u)^T, \tag{3}$$

$$\mathbf{S}(\mathbf{U}) = \delta(2M_1 \omega(\mathbf{U}), -M_2 \omega(\mathbf{U}), 0, 0, 0)^T, \text{ with } \rho = \sum_{i=1}^{3} \rho_i, \tag{4}$$

where

$$\omega(\mathbf{U}) = \left(k_f(T) \frac{\rho_2}{M_2} - k_b(T) \left(\frac{\rho_1}{M_1} \right)^2 \right) \sum_{s=1}^{3} \frac{\rho_s}{M_s},$$

$$T = \frac{p}{R \sum_{s=1}^{3} \frac{\rho_s}{M_s}}, \qquad\qquad k_f(T) = \frac{C}{T^2} \exp\left(-\frac{\hat{E}}{T} \right),$$

4

$$k_b(T) = \frac{k_f(T)}{\exp(b_1 + b_2 \ln(z) + b_3 z + b_4 z^2 + b_5 z^3)}, \qquad z = \frac{10^4}{T}.$$

Here, we have defined the inverse of the reaction characteristic time δ and we will make it vary during the simulations to test very severe stiffness in the source for large δ. The involved constants are

$$M_1 = 0.016, \qquad M_2 = 0.032, \qquad M_3 = 0.028 \qquad R = 287,$$
$$C = 2.9 \cdot 10^{17}, \qquad \hat{E} = 59750, \qquad b_1 = 2.85, \qquad b_2 = 0.988,$$
$$b_3 = -6.181, \qquad b_4 = -0.023, \qquad b_5 = -0.001, \qquad h_1^0 = 1.558 \cdot 10^7.$$

The equation of state (EoS) reads

$$p = \frac{\sum_{s=1}^{3} \frac{\rho_s}{M_s}}{\frac{3}{2}\frac{\rho_1}{M_1} + \frac{5}{2}\frac{\rho_2}{M_2} + \frac{3}{2}\frac{\rho_3}{M_3}} \left(\rho E - \rho_1 h_1^0 - \frac{m^2}{2\rho} \right) \tag{5}$$

and the speed of sound, c, is given by

$$c(\mathbf{U}) = \sqrt{\frac{\gamma p}{\rho}}, \quad \gamma = 1 + \frac{p}{T\sum_{s=1}^{3} \rho_s e_s'(T)}, \quad e_s(T) = \begin{cases} \frac{3}{2}\frac{RT\rho_s}{M_s}, & s = 1,2, \\ \frac{5}{2}\frac{RT\rho_s}{M_s}, & s = 3. \end{cases}$$

It is trivial to see that the total mass is conserved also by the source term as $\mathbf{S}_1(\mathbf{U}) = \delta 2 M_1 \omega(\mathbf{U}) = -\mathbf{S}_2(\mathbf{U}) = \delta M_2 \omega(\mathbf{U})$.

2.3 Properties of Euler equations

In both systems, the preservation of the positivity of the densities is fundamental not only to guarantee the physical relevance of the solution but also to ensure the well-posedness of the problem. In particular, the positivity of the density is required to guarantee that the speed of sound is real-valued and that the system is hyperbolic. Moreover, it is also crucial to preserve the positivity of the pressure in order to compute the speed of sound and ensure that the EoS is well-defined.

Additionally, in absence of reactions, both systems admit a family of solutions with constant velocity and pressure, which are important to preserve in order to avoid spurious oscillations around these equilibria. Indeed, when $u(x,t) = \bar{u}$ and $p(x,t) = \bar{p}$ for some constants $\bar{u}, \bar{p} > 0$, the fluxes in (2) and (3) are constant for the variables ρu and ρE, while the densities are described by transport equations with constant velocity $\bar{u}$. This is the case with a contact discontinuity solution, described by

$$\rho(t,x) = \begin{cases} \rho_L, & x < \bar{u}t, \\ \rho_R, & x > \bar{u}t, \end{cases} \quad u(t,x) = \bar{u}, \quad p(t,x) = \bar{p}, \tag{6}$$

which naturally appears even in simple Riemann problems. Given that most of the numerical methods for the Euler equations are based on the solution of Riemann

problems, it is crucial to preserve such equilibria to avoid spurious oscillations around them.

In what follows, we will try to preserve such properties. The use of MPRK schemes will guarantee the positivity of the densities, while the preservation of the equilibria with constant velocity and pressure will be achieved by balancing the involved numerical fluxes correctly. The positivity of the pressure is less straightforward to guarantee, but we will show that the new flux-balanced version of MPRK schemes, introduced in Section 4, will require a less restrictive time step size to maintain the positivity of the pressure compared to the traditional application of MPRK schemes.

3 Modified Patankar–Runge–Kutta schemes

We will now recall the definition of MPRK schemes and their main properties. We will then discuss how to apply them to the compressible Euler equations in the presence of non-periodic boundary conditions and source terms in Section 4. Following [27, 29, 34, 35], MPRK schemes have been developed to solve the so-called production-destruction-rest systems (PDRS),

$$u_i'(t) = r_i(\mathbf{u}(t), t) + \sum_{j=1}^{N} (p_{ij}(\mathbf{u}(t), t) - d_{ij}(\mathbf{u}(t), t)), \quad \mathbf{u}(0) = \mathbf{u}^0 \in \mathbb{R}_{>0}^N, \qquad (7)$$

where $p_{ij}(\mathbf{u}(t), t), d_{ij}(\mathbf{u}(t), t) \geq 0$ for all $\mathbf{u}(t) > \mathbf{0}$, $t \geq 0$ for $i, j = 1, \ldots, N$. Here and in the following, vector inequalities should be understood in a pointwise way. The expressions p_{ij} and d_{ij} represent production and destruction terms, respectively, while r_i refer to rest terms. In this notation, we require that the PDS part satisfies $p_{ij} = d_{ji}$ as well as $p_{ii} = d_{ii} = 0$ for $i, j = 1, \ldots, N$. For later use and in contrast to [34], the rest terms are also split for $i = 1, \ldots, N$ according to

$$r_i(\mathbf{u}(t), t) = r_i^p(\mathbf{u}(t), t) - r_i^d(\mathbf{u}(t), t) \qquad (8)$$

with $r_i^p, r_i^d \geq 0$ for $t \geq 0$ and $\mathbf{u}(t) \geq \mathbf{0}$. Note that r_i^p and r_i^d can always be constructed, see [35] or (19) for an illustrative example.

The PDRS (7) is said to be *positive* if any solution $\mathbf{u} \colon I \to \mathbb{R}^N$ with $I \subseteq \mathbb{R}_{\geq 0}$ and $\mathbf{u}(0) > \mathbf{0}$ satisfies $\mathbf{u}(t) > \mathbf{0}$ for all $t \in I$, and *conservative* if $r_i = 0$ for all $i = 1, \ldots, N$. In this context, MPRK schemes are defined as follows.

Definition 1. *Given an explicit s-stage RK method described by a non-negative Butcher array, i. e. $\mathbf{A}, \mathbf{b}, \mathbf{c} \geq \mathbf{0}$, we define the corresponding MPRK scheme applied to*

the PDRS (7), (8) by

$$u_i^{(k)} = u_i^n + \Delta t \sum_{\nu=1}^{k-1} a_{k\nu} \left(r_i^p(\mathbf{u}^{(\nu)}, t_n + c_\nu \Delta t) + \sum_{j=1}^{N} p_{ij}(\mathbf{u}^{(\nu)}, t_n + c_\nu \Delta t) \frac{u_j^{(k)}}{\pi_j^{(k)}} \right.$$
$$\left. - \left(r_i^d(\mathbf{u}^{(\nu)}, t_n + c_\nu \Delta t) + \sum_{j=1}^{N} d_{ij}(\mathbf{u}^{(\nu)}, t_n + c_\nu \Delta t) \right) \frac{u_i^{(k)}}{\pi_i^{(k)}} \right), \quad k = 1, \dots, s,$$
$$u_i^{n+1} = u_i^n + \Delta t \sum_{k=1}^{s} b_k \left(r_i^p(\mathbf{u}^{(k)}, t_n + c_k \Delta t) + \sum_{j=1}^{N} p_{ij}(\mathbf{u}^{(k)}, t_n + c_k \Delta t) \frac{u_j^{n+1}}{\sigma_j} \right.$$
$$\left. - \left(r_i^d(\mathbf{u}^{(k)}, t_n + c_k \Delta t) + \sum_{j=1}^{N} d_{ij}(\mathbf{u}^{(k)}, t_n + c_k \Delta t) \right) \frac{u_i^{n+1}}{\sigma_i} \right)$$

$$(9)$$

for $i = 1, \dots, N$, where $\pi_i^{(k)}, \sigma_i$ are the so-called Patankar-weight denominators *(PWDs), which are required to be positive for any $\Delta t \geq 0$. Additionally, $\pi_i^{(k)} = \pi_i^{(k)}(\mathbf{u}^n, \mathbf{u}^{(1)}, \dots, \mathbf{u}^{(k-1)})$ is independent of $\mathbf{u}^{(k)}$, and $\sigma_i = \sigma_i(\mathbf{u}^n, \mathbf{u}^{(1)}, \dots, \mathbf{u}^{(s)})$ is independent of $\mathbf{u}^{n+1}$.*

As described in [20, 27], MPRK schemes can be written in matrix notation that, for the PDRS presented above, read as follows.

Remark 1. *In matrix notation, (9) can be rewritten as*

$$\mathbf{M}^{(k)}\mathbf{u}^{(k)} = \mathbf{u}^n + \Delta t \sum_{\nu=1}^{k-1} a_{k\nu}\mathbf{r}^p(\mathbf{u}^{(\nu)}, t_n + c_\nu \Delta t), \quad k = 1, \dots, s,$$
$$\mathbf{M}\mathbf{u}^{n+1} = \mathbf{u}^n + \Delta t \sum_{k=1}^{s} b_k \mathbf{r}^p(\mathbf{u}^{(k)}, t_n + c_k \Delta t),$$

$$(10)$$

where $\mathbf{r}^p = (r_1^p, \dots, r_N^p)^T$, $\mathbf{M}^{(k)} = (m_{ij}^{(k)})_{1\leq i,j\leq N}$, and $\mathbf{M} = (m_{ij})_{1\leq i,j\leq N}$ with

$$m_{ii}^{(k)} = 1 + \Delta t \sum_{\nu=1}^{k-1} a_{k\nu} \left(r_i^d(\mathbf{u}^{(\nu)}, t_n + c_\nu \Delta t) + \sum_{j=1}^{N} d_{ij}(\mathbf{u}^{(\nu)}, t_n + c_\nu \Delta t) \right) \frac{1}{\pi_i^{(k)}},$$
$$m_{ij}^{(k)} = -\Delta t \sum_{\nu=1}^{k-1} a_{k\nu} p_{ij}(\mathbf{u}^{(\nu)}, t_n + c_\nu \Delta t) \frac{1}{\pi_j^{(k)}}, \quad i \neq j$$

$$(11)$$

as well as

$$m_{ii} = 1 + \Delta t \sum_{k=1}^{s} b_k \left(r_i^d(\mathbf{u}^{(k)}, t_n + c_k \Delta t) + \sum_{j=1}^{N} d_{ij}(\mathbf{u}^{(k)}, t_n + c_k \Delta t) \right) \frac{1}{\sigma_i},$$
$$m_{ij} = -\Delta t \sum_{k=1}^{s} b_k p_{ij}(\mathbf{u}^{(k)}, t_n + c_k \Delta t) \frac{1}{\sigma_j}, \quad i \neq j.$$

$$(12)$$

The unconditional positivity of these schemes is then proved by showing that $\mathbf{M}^{(k)}$ and $\mathbf{M}$ in (11) and (12) are M-matrices, i.e. have non-negative inverses [20]. Since Definition 1 does not determines the PWDs, these can be chosen to ensure the order of accuracy of the method.

Below, we give the specific choices used in this work, however, we note that there is a comprehensive theory for the order conditions [20, 24, 26, 27, 36].

First-Order MPRK Scheme (MPE)

Note that $\mathbf{u}^{(1)} = \mathbf{u}^n$, so $\boldsymbol{\pi}^{(1)}$ does not have to be defined. The first-order scheme is based on forward Euler, which means that only $\boldsymbol{\sigma}$ needs to be defined. To obtain first order of accuracy, we use $\boldsymbol{\sigma} = \mathbf{u}^n$, resulting in the MPE scheme from [18].

Second-Order MPRK Scheme (MPHeun)

Although there is a one-parameter family of second-order MPRK schemes [20], we follow the suggestions from [34] and we will use the scheme based on Heun's method, given by $\boldsymbol{\pi}^{(2)} = \mathbf{u}^n$ and $\boldsymbol{\sigma} = \mathbf{u}^{(2)}$.

4 Modified Patankar Finite Volume Schemes

We start explaining the application of MPRK methods in the simple case of the one species Euler equations (1)-(2) and then we will generalize for the multispecies Euler equations (1)-(3). Afterwards, we will highlight an issue of the method when applied to the one species Euler equations (1)-(2) with a contact discontinuity (37), and how to overcome this problem.

Given $\Omega = [a, b]$ and equispaced points $x_{j+\frac{1}{2}} = a + j\Delta x$ with $\Delta x = \frac{b-a}{N}$, $j = 0, \ldots, N$, we consider the cells $\Omega_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ with $i = 1, \ldots, N$. Along this work, we will use classical Finite Volume methods [37] describing the solution by its cell averages $\mathbf{U}_i \approx \frac{1}{\Delta x} \int_{\Omega_i} \mathbf{U}(x) \mathrm{d}x$. The evolution of these values will depend on interface fluxes that are defined starting from the neighboring values. In this work, we use the local Lax–Friedrichs (LLF) numerical flux, which, using the notation

$$\{\!\{\mathbf{g}(\mathbf{U})\}\!\}_{i+\frac{1}{2}} := \frac{\mathbf{g}(\mathbf{U}_i) + \mathbf{g}(\mathbf{U}_{i+1})}{2}, \quad [\![\mathbf{g}(\mathbf{U})]\!]_{i+\frac{1}{2}} := \mathbf{g}(\mathbf{U}_{i+1}) - \mathbf{g}(\mathbf{U}_i) \qquad (13)$$

can be written as

$$\begin{aligned}
\hat{\mathbf{F}}_{i+\frac{1}{2}} &= \{\!\{\mathbf{F}(\mathbf{U})\}\!\}_{i+\frac{1}{2}} - \frac{\alpha_{i+\frac{1}{2}}}{2}[\![\mathbf{U}]\!]_{i+\frac{1}{2}}, \\
\alpha_{i+\frac{1}{2}} &= \max\{|u_i| + c(\mathbf{U}_i), |u_{i+1}| + c(\mathbf{U}_{i+1}), |\{\!\{u\}\!\}_{i+\frac{1}{2}}| + c(\{\!\{\mathbf{U}\}\!\}_{i+\frac{1}{2}})\}.
\end{aligned} \qquad (14)$$

This allows to write the semi-discrete form of (1) as

$$\frac{\mathrm{d}\mathbf{U}_i}{\mathrm{d}t} = \frac{1}{\Delta x}\left(\hat{\mathbf{F}}_{i-\frac{1}{2}} - \hat{\mathbf{F}}_{i+\frac{1}{2}}\right) + \mathbf{S}(\mathbf{U}_i), \qquad (15)$$

where $\mathbf{U}_i = (U_{i,1}, U_{i,2}, U_{i,3})^T = ((\rho)_i, (\rho u)_i, (\rho E)_i)^T$ represent the cell averages of the physical quantities for each cell Ω_i, and we applied the midpoint rule to the source term. In the following, we describe how to apply the MPE scheme to the semi-discrete form of Euler equations in order to guarantee positivity of the density and total energy. The generalization to higher-order schemes is then straightforward.

To illustrate the main idea, and for the sake of simplicity, let us first focus on the density components of (15) for the Euler system (1)-(2), but the same technique will be applied to other positive variables. In the simulations, we will test also the version where both density and total energy are treated with the MP trick. Those version will be denoted with a -ρE, e.g. MPE-ρE. The density equation reads

$$\frac{\mathrm{d}\rho_i}{\mathrm{d}t} = \frac{1}{\Delta x}\left(\hat{F}_{i-\frac{1}{2},1} - \hat{F}_{i+\frac{1}{2},1}\right), \quad i = 1, \ldots, N, \tag{16}$$

as $S_1(\mathbf{U}_i) = 0$. First, note that due to the boundary conditions, the sum $\sum_{i=1}^{N}\left(\hat{F}_{i-\frac{1}{2},1} - \hat{F}_{i+\frac{1}{2},1}\right) = \hat{F}_{\frac{1}{2},1} - \hat{F}_{N+\frac{1}{2},1}$ does not vanish in general, so that (16) must be viewed as a PDRS in order to apply the MPE scheme. Hence, we rewrite (16) as a PDRS

$$\frac{\mathrm{d}\rho_i}{\mathrm{d}t} = \frac{1}{\Delta x}\left(r_{i,1}^p - r_{i,1}^d + \sum_{j=1}^{N}(p_{i,j,1} - d_{i,j,1})\right), \quad i = 1, \ldots, N, \tag{17}$$

where $r_{i,1}^p, r_{i,1}^d, p_{i,j,1}, d_{i,j,1} \geq 0$, and $p_{i,j,1} = d_{j,i,1}$. In particular, we have

$$p_{i+1,i,1} = d_{i,i+1,1} = \max\left\{0, \hat{F}_{i+\frac{1}{2},1}\right\}, \quad p_{i,i+1,1} = d_{i+1,i,1} = -\min\left\{0, \hat{F}_{i+\frac{1}{2},1}\right\} \tag{18}$$

for $i = 1, \ldots, N-1$, and $p_{i,j,1} = d_{j,i,1} = 0$, if $|i - j| \neq 1$. Furthermore, the terms $\mathbf{r}_1^p$ and $\mathbf{r}_1^d$ are given by

$$\begin{aligned}
\mathbf{r}_1^p &= \max\left\{0, \hat{F}_{\frac{1}{2},1}\right\}\mathbf{e}_1 + \max\left\{0, -\hat{F}_{N+\frac{1}{2},1}\right\}\mathbf{e}_N, \\
\mathbf{r}_1^d &= -\left(\min\left\{0, \hat{F}_{\frac{1}{2},1}\right\}\mathbf{e}_1 + \min\left\{0, -\hat{F}_{N+\frac{1}{2},1}\right\}\mathbf{e}_N\right)
\end{aligned} \tag{19}$$

with the unit vectors $\mathbf{e}_1, \mathbf{e}_N \in \mathbb{R}^N$. Now, the MPE scheme for the density reads

$$\rho_i^{n+1} = \rho_i^n + \frac{\Delta t}{\Delta x}r_{i,1}^{p,n} + \frac{\Delta t}{\Delta x}\left(\sum_{j=1}^{N}p_{i,j,1}^n\frac{\rho_j^{n+1}}{\rho_j^n} - \left(r_{i,1}^{d,n} + \sum_{j=1}^{N}d_{i,j,1}^n\right)\frac{\rho_i^{n+1}}{\rho_i^n}\right), \tag{20}$$

for $i = 1, \ldots, N$ and only differs from the forward Euler (FE) by the fractional weights. We note that the MPE scheme is also first-order accurate when applied to a PDRS, see [26, 35] for more details. Moreover, the solution $\mathbf{u}_1^{n+1} = (\rho_1^{n+1}, \ldots, \rho_N^{n+1})^T$ of (20) can be obtained by solving the linear system $\mathbf{M}_1\mathbf{u}_1^{n+1} = \mathbf{u}_1^n + \frac{\Delta t}{\Delta x}\mathbf{r}_1^{p,n}$, where the matrix $\mathbf{M}_1 \in \mathbb{R}^{N \times N}$ is an M-matrix and therefore possesses a non-negative inverse. Furthermore, due to $p_{i,j,1}^n = d_{j,i,1}^n = 0$ for $|i - j| \neq 1$, the matrix $\mathbf{M}_1$ is tridiagonal

and sparse for large N. In particular, writing $\mathbf{M}_1 = (m_{i,j,1})_{1 \le i,j \le N}$ as a special case of (12), we observe

$$m_{i,i,1} = 1 + \frac{\Delta t}{\Delta x} \frac{r_{i,1}^{d,n} + \sum_{j=1}^{N} d_{i,j,1}^n}{U_{i,1}^n} \quad \text{and} \quad m_{i,j,1} = -\frac{\Delta t}{\Delta x} \frac{p_{i,j,1}^n}{U_{j,1}^n}, \quad |j - i| = 1$$

and $m_{i,j,1} = 0$ otherwise.

Similarly, this approach could also be used to guarantee the positivity of the total energy by solving the corresponding linear system $\mathbf{M}_k \mathbf{u}_k^{n+1} = \mathbf{u}_k^n + \frac{\Delta t}{\Delta x} \mathbf{r}_k^{p,n}$ for $k = 3$ resulting in a total of two decoupled linear systems of size $N \times N$ that must be solved in each step, while the computation of the velocity remains unchanged, since there are no positivity constraints on the velocity.

MPE for Multi-species Euler

In the case of multi-species Euler (1)-(3), one has to perform the same MPE method on each density equation and on the total energy, solving $\mathbf{M}_k \mathbf{u}_k^{n+1} = \mathbf{u}_k^n + \frac{\Delta t}{\Delta x} \mathbf{r}_k^{p,n}$ for $k = 1, 2, 3, 5$, while using the forward Euler scheme for the velocity equation. One can define the vectors $\mathbf{u} = (\mathbf{u}_1^T, \mathbf{u}_2^T, \mathbf{u}_3^T, \mathbf{u}_5^T)^T$ and $\mathbf{r}^p = ((\mathbf{r}_1^p)^T, (\mathbf{r}_2^p)^T, (\mathbf{r}_3^p)^T, (\mathbf{r}_5^p)^T)^T$ as well as the block diagonal matrix $\mathbf{M} = \text{diag}(\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_5)$, so that $\mathbf{M}\mathbf{u}^{n+1} = \mathbf{u}^n + \frac{\Delta t}{\Delta x} \mathbf{r}^p$ represents the four decoupled linear systems that have to be solved. However, since source terms are present for the densities ρ_1 and ρ_2, we have to adjust our scheme as follows.

Consider the corresponding vectors $\mathbf{u}_k = (U_{1,k}, \dots, U_{N,k})^T = (\rho_{1,k}, \dots, \rho_{N,k})^T$ with $k \in \{1, 2\}$. The semi-discrete form (15) of the reactive Euler equations for these two quantities reads

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}\rho_{i,1} = \frac{\mathrm{d}}{\mathrm{d}t}U_{i,1} = r_{i,1}^p - r_{i,1}^d + \sum_{j=1}^{N}(p_{i,j,1} - d_{i,j,1}) + S_1(\mathbf{U}_i), & i = 1, \dots, N, \\ \frac{\mathrm{d}}{\mathrm{d}t}\rho_{i,2} = \frac{\mathrm{d}}{\mathrm{d}t}U_{i,2} = r_{i,2}^p - r_{i,2}^d + \sum_{j=1}^{N}(p_{i,j,2} - d_{i,j,2}) + S_2(\mathbf{U}_i), & i = 1, \dots, N. \end{cases}$$

Next, we split each source term $S_k(\mathbf{U}_i) = S_k^p(\mathbf{U}_i) - S_k^d(\mathbf{U}_i)$ into a production and a destruction part. In particular,

$$S_1^p(\mathbf{U}_i) = \delta 2 M_1 k_f(T) \frac{U_{i,2}}{M_2} \sum_{s=1}^{3} \frac{U_{i,s}}{M_s} \ge 0 \quad \text{for } U_{i,s} \ge 0$$

is the source production term and

$$S_1^d(\mathbf{U}_i) = \delta 2 M_1 k_b(T) \left(\frac{U_{i,1}}{M_1}\right)^2 \sum_{s=1}^{3} \frac{U_{i,s}}{M_s} \ge 0 \quad \text{for } U_{i,s} \ge 0$$

is the source destruction term of S_1. According to (4) and $2M_1 = M_2$, we observe $S_2^p(\mathbf{U}_i) = S_1^d(\mathbf{U}_i)$ and $S_2^d(\mathbf{U}_i) = S_1^p(\mathbf{U}_i)$. As a consequence, the MPE scheme for the

first two densities reads for $i = 1, \ldots, N$

$$
\begin{cases}
\rho_{i,1}^{n+1} = U_{i,1}^{n+1} = U_{i,1}^n + \dfrac{\Delta t}{\Delta x} \left(r_{i,1}^{p,n} + \sum_{j=1}^N p_{i,j,1}^n \dfrac{U_{j,1}^{n+1}}{U_{j,1}^n} - \left(r_{i,1}^{d,n} + \sum_{j=1}^N d_{i,j,1}^n \right) \dfrac{U_{i,1}^{n+1}}{U_{i,1}^n} \right) \\
\qquad\qquad + \Delta t \left(S_1^p(\mathbf{U}_i^n) \dfrac{U_{i,2}^{n+1}}{U_{i,2}^n} - S_1^d(\mathbf{U}_i^n) \dfrac{U_{i,1}^{n+1}}{U_{i,1}^n} \right) \\
\rho_{i,2}^{n+1} = U_{i,2}^{n+1} = U_{i,2}^n + \dfrac{\Delta t}{\Delta x} \left(r_{i,2}^{p,n} + \sum_{j=1}^N p_{i,j,2}^n \dfrac{U_{j,2}^{n+1}}{U_{j,2}^n} - \left(r_{i,2}^{d,n} + \sum_{j=1}^N d_{i,j,2}^n \right) \dfrac{U_{i,2}^{n+1}}{U_{i,2}^n} \right) \\
\qquad\qquad + \Delta t \left(S_2^p(\mathbf{U}_i^n) \dfrac{U_{i,1}^{n+1}}{U_{i,1}^n} - S_2^d(\mathbf{U}_i^n) \dfrac{U_{i,2}^{n+1}}{U_{i,2}^n} \right)
\end{cases}
$$

and hence, the full scheme for computing the densities and energy can be written as

$$
(\mathbf{M} + \hat{\mathbf{S}}) \mathbf{u}^{n+1} = \mathbf{u}^n + \frac{\Delta t}{\Delta x} \mathbf{r}^{p,n} \tag{21}
$$

with $\mathbf{M}$ as defined above and $\hat{\mathbf{S}} = \Delta t \begin{pmatrix} \mathbf{S}_{d,1} & \mathbf{S}_{p,2} & \mathbf{0} & \mathbf{0} \\ \mathbf{S}_{p,1} & \mathbf{S}_{d,2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$, where $\mathbf{S}_{d,k} = $ diag $\left(\frac{S_k^d(\mathbf{U}_1^n)}{U_{1,k}^n}, \ldots, \frac{S_k^d(\mathbf{U}_N^n)}{U_{N,k}^n} \right)$ and $\mathbf{S}_{p,k} = -\mathbf{S}_{d,k}$. Note that the linear system (21) is only coupled through the sparse matrix $\hat{\mathbf{S}}$.

Lemma 1. *The Matrix $\mathbf{B} := \mathbf{M} + \hat{\mathbf{S}}$ from* (21) *is an M-matrix.*

Proof. It is enough to restrict to the upper left $2N \times 2N$ block as the other blocks on the diagonal are M-matrices [27]. Hence, fix $k \in \{1, 2\}$ and observe that $\mathbf{1}^T \mathbf{S}_{d,k} = \|\mathbf{1}^T \mathbf{D}_{p,k}\|_1$. Hence, since the diagonal of $\mathbf{B}$ is positive and $\mathbf{M}_k^T$ is diagonally dominant, we deduce that $\mathbf{B}^T$ is also diagonally dominant along the same lines as in [20, Lemma 2.8]. Together with $b_{ij} \leq 0$ for $i \neq j$, we conclude that $\mathbf{B}^T$ is an M-matrix, which means that the same holds for $\mathbf{B}$. $\qquad\square$

4.1 Contact Discontinuities Preservation

In a contact discontinuity, i.e., when $u(x) = \bar{u}$ and $p(x) = \bar{p}$ are constant values, whatever form takes the density, the density equation becomes a transport equation with velocity $\bar{u}$, while the momentum and energy conservation laws guarantee that velocity and pressure remain constant.

Proposition 2 (Contact discontinuity preservation of explicit Euler LLF). *Consider the Euler equations* (1)-(2) *with homogeneous Neumann or periodic boundary conditions, and apply the LLF scheme (explicit Euler method in time) together with the EoS $p = (\gamma - 1)e$, where $e := \rho E - 0.5\rho u^2$ is the internal energy. If the initial conditions*

contain only contact discontinuities, i.e., $u = \bar{u}$ and $p = \bar{p}$, then velocity and pressure stay constant for all time steps of the method.

Proof. Since $u = \bar{u}$ and $p = \bar{p}$ are constant at the beginning, we observe that the LLF (14) for the momentum equation becomes

$$\hat{F}_{i+\frac{1}{2},2} = \bar{u}\hat{F}_{i+\frac{1}{2},1} + \bar{p}. \tag{22}$$

This leads to the following update equation for the momentum

$$\frac{(\rho u)_i^{n+1} - (\rho \bar{u})_i^n}{\Delta t} = -\frac{1}{\Delta x}\left(\hat{\mathbf{F}}_{i+\frac{1}{2},2}^n - \hat{\mathbf{F}}_{i-\frac{1}{2},2}^n\right) = -\bar{u}\frac{1}{\Delta x}\left(\hat{\mathbf{F}}_{i+\frac{1}{2},1}^n - \hat{\mathbf{F}}_{i-\frac{1}{2},1}^n\right), \tag{23}$$

which is clearly verified by $u_i^{n+1} = \bar{u}$ and by the LLF update of the density

$$\frac{\rho_i^{n+1} - \rho_i^n}{\Delta t} = -\frac{1}{\Delta x}\left(\hat{\mathbf{F}}_{i+\frac{1}{2},1}^n - \hat{\mathbf{F}}_{i-\frac{1}{2},1}^n\right). \tag{24}$$

Note that it follows from (22)-(24) that u will stay constant after one step of the LLF scheme. Furthermore, since $p = \bar{p}$ we deduce from the EoS for ideal gases that the internal energy e is also constant since $e = \rho E - 0.5\rho u^2 = \frac{p}{\gamma - 1}$. Hence, writing

$$F_3(\mathbf{U}) = u(\rho E + p) = u(e + 0.5\rho u^2 + p),$$

we obtain

$$\hat{F}_{i+\frac{1}{2},3} = 0.5\bar{u}^2\hat{F}_{i+\frac{1}{2},1} + \bar{u}(\bar{e} + \bar{p}). \tag{25}$$

From (25), we conclude that

$$(\rho E)_i^{n+1} = (\rho E)_i^n - \frac{\Delta t}{\Delta x}(\hat{F}_{i+\frac{1}{2},3} - \hat{F}_{i-\frac{1}{2},3}) = (\rho E)_i^n + 0.5\bar{u}^2(\rho_i^{n+1} - \rho_i^n).$$

Thus,

$$e_i^{n+1} = (\rho E)_i^{n+1} - 0.5\rho_i^{n+1}\bar{u}^2 = (\rho E)_i^n - 0.5\rho_i^n\bar{u}^2 = \bar{e},$$

which means that also pressure is constant in time. If the boundary conditions verify the same constraints $u = \bar{u}$ and $p = \bar{p}$, then the same arguments can be applied to boundary cells. $\square$

4.1.1 Modified Patankar Euler LLF for Euler Equations

However, the situation is different when MPRK schemes are applied, as the MP trick is never applied to the momentum equation. Hence, since ρ is computed with an MPRK scheme, the ρu must be computed in a consistent manner to keep u constant. Indeed, the main idea of the proof of Proposition 2 is that we can substitute the equation for the computation of ρ into the equations for momentum and total energy. Hence, if we only modify the density fluxes $\hat{F}_{i+\frac{1}{2},1}$ using Patankar weights, all we have to do is to appropriately weigh the other numerical fluxes with the same Patankar-weights as for

the density equation. To that end, we note that (22) and (25) can be split into a part that contains the numerical flux of the density, which should be weighted exactly as in the density equation, while the other terms should be kept as in the explicit update of the original fluxes.

This strategy is in conflict with the weighting of the energy equation with an MP trick, so we could not guarantee the positivity of the total energy. Nevertheless, the pressure is the quantity whose positivity has to be guaranteed, and from the numerical experiments it seemed that the novel weighting due to the density MP trick facilitates the positivity of the pressure much more than using the MP trick on the energy equation. Hence, we will stick to this strategy and we will consider in future works the combination of MP trick on the energy equation and the preservation of contact discontinuities. From here on, we focus on the ansatz of applying the MP trick to the density equation and weigh the other fluxes consistently.

Overall, we will split the numerical fluxes into a part that should be weighted and another one that should be left untouched. The first will have a superscript index w and the latter will be indexed by u, i.e. we split $\hat{F}_{i+\frac{1}{2},k} = \hat{F}^w_{i+\frac{1}{2},k} + \hat{F}^u_{i+\frac{1}{2},k}$. According to (22), we define for the momentum equation

$$\hat{F}^w_{i+\frac{1}{2},2} = \{\{\rho u^2\}\}_{i+\frac{1}{2}} - \frac{\alpha_{i+\frac{1}{2}}}{2}[\![\rho u]\!]_{i+\frac{1}{2}}, \quad \hat{F}^u_{i+\frac{1}{2},2} = \{\{p\}\}_{i+\frac{1}{2}}$$

using the jump $[\![\cdot]\!]_{i+\frac{1}{2}}$ and mean $\{\{\cdot\}\}_{i+\frac{1}{2}}$ operators from (13). From (25), the energy fluxes can be written in the form

$$\hat{F}^w_{i+\frac{1}{2},3} = \{\{0.5\rho u^3\}\}_{i+\frac{1}{2}} - \frac{\alpha_{i+\frac{1}{2}}}{2}[\![0.5\rho u^2]\!]_{i+\frac{1}{2}}, \quad \hat{F}^u_{i+\frac{1}{2},3} = \{\{u(e+p)\}\}_{i+\frac{1}{2}} - \frac{\alpha_{i+\frac{1}{2}}}{2}[\![e]\!]_{i+\frac{1}{2}}.$$

In what follows, we deduce the variant of the MPE scheme. The generalization to higher order is straightforward. Now, recall that the production and destruction terms for the density are chosen analogously to (18). Hence, the weighting of $\hat{F}^w_{i+\frac{1}{2},k}$ with $k \in \{2,3\}$ for momentum and energy will depend on the sign of $\hat{F}_{i+\frac{1}{2},1}$. In particular, if $\hat{F}_{i+\frac{1}{2},1} < 0$ then this corresponds to a non-zero $p_{i,i+1,1}$ term with a weight of

$$w^{n+1}_{i+1} := \frac{\rho^{n+1}_{i+1}}{\rho^n_{i+1}}, \tag{26}$$

and, if $\hat{F}_{i+\frac{1}{2},1} > 0$, the term corresponds to a destruction term $d_{i,i+1,1}$, which is weighted with w^{n+1}_i. The rest terms can be treated analogously. We define for $k = 2,3$ and $i = 1, \ldots, N-1$ the auxiliary production and destruction terms by

$$\mathfrak{p}_{i,i+1,k} := \begin{cases} -\hat{F}^w_{i+\frac{1}{2},k}, & \hat{F}_{i+\frac{1}{2},1} < 0, \\ 0, & \text{otherwise} \end{cases}, \quad \mathfrak{d}_{i,i+1,k} := \begin{cases} \hat{F}^w_{i+\frac{1}{2},k}, & \hat{F}_{i+\frac{1}{2},1} \geq 0, \\ 0, & \text{otherwise} \end{cases}, \tag{27}$$

13

and set $\mathfrak{p}_{i,j,k} = \mathfrak{d}_{j,i,k}$ for all $i, j = 1, \ldots, N,$

$$\mathfrak{r}_{1,k}^d := \begin{cases} -\hat{F}_{\frac{1}{2},k}^w, & \hat{F}_{\frac{1}{2},1} < 0, \\ 0, & \text{otherwise}, \end{cases} \quad \mathfrak{r}_{2,k}^d = \ldots = \mathfrak{r}_{N-1,k}^d = 0, \, \mathfrak{r}_{N,k}^d := \begin{cases} \hat{F}_{N+\frac{1}{2},k}^w, & \hat{F}_{N+\frac{1}{2},1} \geq 0, \\ 0, & \text{otherwise}, \end{cases}$$

(28)

as well as $\mathfrak{r}_{i,k}^p := r_{i,k} + \mathfrak{r}_{i,k}^d + (\hat{F}_{i-\frac{1}{2},k}^u - \hat{F}_{i+\frac{1}{2},k}^u)$. Then, we can prove the following result.

Theorem 3. *The scheme*

$$U_{i,1}^{n+1} = U_{i,1}^n + \frac{\Delta t}{\Delta x} r_{i,1}^{p,n} + \frac{\Delta t}{\Delta x} \left(\sum_{j=1}^N p_{i,j,1}^n w_j^{n+1} - \left(r_{i,1}^{d,n} + \sum_{j=1}^N d_{i,j,1}^n \right) w_i^{n+1} \right)$$

$$U_{i,k}^{n+1} = U_{i,k}^n + \frac{\Delta t}{\Delta x} \mathfrak{r}_{i,k}^{p,n} + \frac{\Delta t}{\Delta x} \left(\sum_{j=1}^N \mathfrak{p}_{i,j,k}^n w_j^{n+1} - \left(\mathfrak{r}_{i,k}^{d,n} + \sum_{j=1}^N \mathfrak{d}_{i,j,k}^n \right) w_i^{n+1} \right), \, k = 2, 3,$$

(29)

with $w_i = U_{i,1}^{n+1}/U_{i,1}^n$ is first-order accurate and unconditionally preserves the positivity of the density. Additionally, if u and p are initially constant, they remain constant at all times.

Proof. The positivity of the density is already known. Additionally, we chose the auxiliary production, destruction and rest terms in such a way that we can mimic the proof of Proposition 2 since we have

$$p_{i,i\pm 1,1} = \mp \hat{F}_{i\pm\frac{1}{2},1} \implies \mathfrak{p}_{i,i\pm 1,k} = \mp \hat{F}_{i\pm\frac{1}{2},k}^w,$$

and similar relations hold for the $\mathfrak{d}$ and $\mathfrak{r}^d$ terms. Finally, the consistent weighting allows us to substitute the weighted numerical density-fluxes into the other equations as we did in Proposition 2. Finally, since the MPE scheme for the density equation is first-order accurate, we know from [26] that $w_i = 1 + \mathcal{O}(\Delta t)$. Substituting this relation into the momentum and total energy equation, we end up with the original LLF scheme plus an additional $\mathcal{O}(\Delta t^2)$ error. $\square$

We want to remark that the weights added to the momentum and total energy equations are explicitly computed from the density update, which is the only equation where the MP trick is applied. Hence, this flux-balancing does not need further linear solves in contrast to classical MPRK schemes.

In the numerical results, we will see that this splitting and weighting of the fluxes of momentum and energy will also help to keep the pressure positive. It will not lead to a provably positivity preserving scheme for the pressure, but the simulations run with this split outperform by huge factors the scheme where the MP is applied to the density and total energy and explicit Euler is used for the momentum equation.

4.1.2 Application to Multi-Species Euler

Now, we can extend this result to multi-species Euler equations (1), where the contact discontinuity preservation property is directly relevant only in the case without source

terms. Nevertheless, we hope that the increased stability of the scheme, in particular in the positivity of the pressure, will also be transferred to the reactive case.

The methodology is very similar to the previous section: the densities equations are each of them treated with MP Euler, then we use the sum of the weights to weigh the part of the numerical fluxes of momentum and energy that should resemble a density numerical flux. For $k = 1, \ldots, 3$, we have that the three densities evolve as

$$\frac{\rho_{i,k}^{n+1} - \rho_{i,k}^n}{\Delta t} = \frac{1}{\Delta x} \left(r_{i,k}^{p,n} + \sum_{j=1}^N (p_{i,j,k}^n w_{j,k}^{n+1} - (r_{i,k}^{d,n} + d_{i,j,k}^n) w_{i,k}^{n+1}) \right), \qquad (30)$$

with $w_{i,k}^{n+1} = \rho_{i,k}^{n+1}/\rho_{i,k}^n$. Then, we take the sum of these equations over k, to obtain

$$\frac{\rho_i^{n+1} - \rho_i^n}{\Delta t} = \frac{1}{\Delta x} \sum_{k=1}^3 \left(r_{i,k}^{p,n} + \sum_{j=1}^N (p_{i,j,k} w_{j,k}^{n+1} - (r_{i,k}^{d,n} + d_{i,j,k}) w_{i,k}^{n+1}) \right). \qquad (31)$$

The update of ρ is the quantity that should be factored out in the momentum and energy equations. So, let us define for the momentum for $k = 1, \ldots, 3$

$$\hat{F}_{i+\frac{1}{2},4}^{w,k} = \{\{\rho_k u^2\}\}_{i+\frac{1}{2}} - \frac{\alpha_{i+\frac{1}{2}}}{2} [\![\rho_k u]\!]_{i+\frac{1}{2}}, \quad \hat{F}_{i+\frac{1}{2},4}^u = \{\{p\}\}_{i+\frac{1}{2}},$$

and for the energy, we define for $k = 1, \ldots, 3$ fluxes

$$\hat{F}_{i+\frac{1}{2},5}^{w,k} = \{\{0.5\rho_k u^3\}\}_{i+\frac{1}{2}} - \frac{\alpha_{i+\frac{1}{2}}}{2} [\![0.5\rho_k u^2]\!]_{i+\frac{1}{2}}, \quad \hat{F}_{i+\frac{1}{2},5}^u = \{\{u(e+p)\}\}_{i+\frac{1}{2}} - \frac{\alpha_{i+\frac{1}{2}}}{2} [\![e]\!]_{i+\frac{1}{2}}.$$

We define for $\ell = 4, 5$, $k = 1, \ldots, 3$ and $i = 1, \ldots, N-1$ the auxiliary production and destruction terms by

$$\mathfrak{p}_{i,i+1,\ell}^k := \begin{cases} -\hat{F}_{i+\frac{1}{2},\ell}^{w,k}, & \hat{F}_{i+\frac{1}{2},k} < 0, \\ 0, & \text{otherwise} \end{cases}, \quad \mathfrak{d}_{i,i+1,\ell}^k := \begin{cases} \hat{F}_{i+\frac{1}{2},\ell}^{w,k}, & \hat{F}_{i+\frac{1}{2},k} \geq 0, \\ 0, & \text{otherwise} \end{cases}, \qquad (32)$$

and set $\mathfrak{p}_{i,j,\ell}^k = \mathfrak{d}_{j,i,\ell}^k$ for all $i, j = 1, \ldots, N$,

$$\mathfrak{r}_{1,\ell}^{d,k} := \begin{cases} -\hat{F}_{\frac{1}{2},\ell}^{w,k}, & \hat{F}_{\frac{1}{2},k} < 0, \\ 0, & \text{otherwise} \end{cases}, \mathfrak{r}_{2,\ell}^{d,k} = \ldots = \mathfrak{r}_{N-1,\ell}^{d,k} = 0, \mathfrak{r}_{N,\ell}^{d,k} := \begin{cases} \hat{F}_{N+\frac{1}{2},\ell}^{w,k}, & \hat{F}_{N+\frac{1}{2},k} \geq 0, \\ 0, & \text{otherwise} \end{cases},$$

$$(33)$$

as well as $\mathfrak{r}_{i,\ell}^p := r_{i,\ell} + \mathfrak{r}_{i,\ell}^d + (\hat{F}_{i-\frac{1}{2},\ell}^u - \hat{F}_{i+\frac{1}{2},\ell}^u)$. Then, we can prove the following result.

Theorem 4. *The scheme*

$$U_{i,k}^{n+1} = U_{i,k}^n + \frac{\Delta t}{\Delta x}\left(r_{i,k}^{p,n} + \sum_{j=1}^N p_{i,j,k}^n w_{j,k}^{n+1} - \left(r_{i,1}^{d,n} + \sum_{j=1}^N d_{i,j,1}^n\right)w_{i,k}^{n+1}\right), \, k = 1, \ldots, 3,$$

$$U_{i,\ell}^{n+1} = U_{i,\ell}^n + \frac{\Delta t}{\Delta x}\left(\mathfrak{r}_{i,\ell}^{p,n} + \sum_{k=1}^3 \left(\sum_{j=1}^N \mathfrak{p}_{i,j,\ell}^{k,n} w_{j,k}^{n+1} - \left(\mathfrak{r}_{i,\ell}^{d,k,n} + \sum_{j=1}^N \mathfrak{d}_{i,j,\ell}^{k,n}\right)w_{i,k}^{n+1}\right)\right), \, \ell = 4, 5,$$

(34)

with $w_{i,k} = U_{i,k}^{n+1}/U_{i,k}^n$, discretizing the multi-species Euler without reaction source terms, unconditionally preserves the positivity of the densities. Additionally, if u and p are initially constant, they remain constant at all times. Moreover, the scheme is first-order accurate.

Proof. The proof is similar to the previous one, given the linearity of the weighted numerical fluxes on the densities. Concerning the accuracy, we know from [26] that $w_{j,k}^{n+1} = 1 + \mathcal{O}(\Delta t)$, and hence, the method has the same accuracy as the underlying explicit scheme. □

Remark 2. *In the second order case, we know that $w_{j,k}^{n+1} = 1 + \mathcal{O}(\Delta t^2)$. However, there will be also weights for the second stage of the RK scheme, i. e. $w_{j,k}^{(2)} = \frac{U_{j,k}^{(2)}}{\pi_{j,k}^{(2)}} = 1 + \mathcal{O}(\Delta t)$ for the densities [26, Theorem 18]. Furthermore, the same theorem in [26] then guarantees that the second-order in time will be maintained for the other quantities, and any high-order space discretization can be applied to obtain an overall second-order of accuracy. In what follows, we use linear reconstruction with a minmod limiter to obtain second order away from extrema and non-smooth regions.*

5 Numerical Results

In this section, we will validate the presented numerical method with balanced fluxes comparing it with the corresponding explicit schemes as well as with MP schemes applied to densities and with MP schemes applied to density and total energy equations. In particular, we will use the forward Euler (FE) method and Heun's method as explicit methods of first and second order, respectively. The corresponding MP schemes will be denoted by MPE and MPHeun, where we apply the MP trick to the densities equations only, MPE-ρE and MPHeun-ρE for the versions where both density and total energy are treated with the MP trick, and MPE-s and MPHeun-s for the versions where the MP trick is applied to the density and the fluxes of momentum and energy are weighted consistently with the density MP trick as described in the previous section.

5.1 Choice of the CFL

Most of the numerical experiments we will perform, undergo severe discontinuities and stiff source terms, which makes the choice of the time step crucial for the stability of the method. Except for the first smooth test case, we will set the CFL number on a reference solution using the largest timestep we can use without obtaining negative

densities or pressure. This will not guarantee that the simulation gives qualitatively good results, hence, we will introduce a *safety factor* that will premultiply the CFL, in order to obtain reliable simulations. The comparison between the methods will be then mainly on their computational costs at, more or less, the same level of accuracy and reliability.

5.2 Convergence Test

We start demonstrating that the claimed convergence can be observed in smooth regions. For this test case, we consider the Euler equations (1)-(2) on $[0, 1]$, with the initial conditions

$$(\rho, u, p)(0, x) = (1, 1, 1 + \cos(0.5\pi x)^4) \tag{35}$$

and final time $t_{\text{end}} = 0.03$. We compute the error on the density and the experimental order of convergence (EOC) for various number of cells for the first and second-order schemes using the L_1 and L_2 norms and a CFL number of 0.5 for all schemes. To that end, we define a reference solution using the same method with double the amount of cells, respectively. As can be seen in Table 1, the claimed accuracy is achieved. We note that the experimental order of convergence is reduced using the L_2 norm, see Table 2 as we use a minmod limiter that locally reduces the order to one at extrema. This is enough to impact the EOC when computed with the L_2 norm, but not the L_1 norm.

Table 1: EOC for Euler equations (1)-(2) with IC (35) using L_1 norm on ρ

N	First order				Second order			
	FE	MPE-ρE	MPE	MPE-s	Heun	MPHeun-ρE	MPHeun	MPHeun-s
160	0.996	0.984	0.983	0.982	1.908	1.905	1.907	1.901
320	0.996	0.998	0.998	0.997	1.948	1.950	1.950	1.944
640	1.000	1.000	1.000	0.999	1.957	1.957	1.960	1.955
1280	1.000	1.001	1.001	1.000	1.966	1.966	1.968	1.965
2560	1.000	1.000	1.000	1.000	1.971	1.972	1.971	1.969

Table 2: EOC for Euler equations (1)-(2) with IC (35) using L_2 norm on ρ

N	First order				Second order			
	FE	MPE-ρE	MPE	MPE-s	Heun	MPHeun-ρE	MPHeun	MPHeun-s
160	0.989	0.974	0.973	0.970	1.715	1.708	1.709	1.709
320	0.988	0.990	0.989	0.987	1.716	1.721	1.722	1.710
640	0.992	0.994	0.994	0.992	1.704	1.714	1.713	1.700
1280	0.995	0.996	0.996	0.995	1.690	1.697	1.694	1.686
2560	0.997	0.997	0.997	0.996	1.678	1.683	1.680	1.675

Table 3: Comparison of numerical schemes applied to multi-species Euler equations (1)-(3) with source (4) with $\delta = 10^4$ with $N = 4000$ cells: stable CFL and runtime in seconds

First order				Second order			
Method	CFL	safety factor	Runtime	Method	CFL	safety factor	Runtime
FE	0.008	1	1494.76	Heun	0.008	1	5185.29
MPE-ρE	0.21	0.4	325.18	MPHeun-ρE	0.11	0.4	1960.50
MPE	0.18	0.4	371.04	MPHeun	0.12	0.4	1671.91
MPE-s	0.19	0.4	470.39	MPHeun-s	0.86	1	121.29

5.3 Reactive Euler Equations

For the reactive Euler equations as a special case of the multi-species Euler equations, we consider the Riemann problem

$$U_i(0, x) = \begin{cases} U_{i,L}, & x < 0, \\ U_{i,R}, & x \geq 0, \end{cases} \tag{36}$$

defined by [2, 33]

$$\begin{pmatrix} \rho_{1,L} \\ \rho_{2,L} \\ \rho_{3,L} \\ v_L \\ p_L \end{pmatrix} = \begin{pmatrix} 5.251896311257205 \cdot 10^{-5} \\ 3.748071704863518 \cdot 10^{-5} \\ 2.962489471973072 \cdot 10^{-4} \\ 0 \\ 10^3 \end{pmatrix}, \begin{pmatrix} \rho_{1,R} \\ \rho_{2,R} \\ \rho_{3,R} \\ v_R \\ p_R \end{pmatrix} = \begin{pmatrix} 8.341661837019181 \cdot 10^{-8} \\ 9.45418692098664 \cdot 10^{-11} \\ 2.748909430004963 \cdot 10^{-7} \\ 0 \\ 1 \end{pmatrix}$$

together with homogeneous Neumann boundary conditions, and final time $t_{\text{end}} = 10^{-4}$. We will use $\delta = 10^4$ as the reaction rate, which makes the problem very stiff.

We use $N = 4000$ cells on the interval $[a, b] = [-1, 1]$, i.e. $\Delta x = \frac{2}{4000}$. The first-order schemes and second-order schemes produce similar approximations, respectively. However, due to the stiffness of the problem, the time step restrictions to have stability for the explicit schemes are very severe, see Table 3. On the other hand, the classical MPRK schemes have weaker constraints on the time-step (around CFL ~ 0.15), while the flux-balanced version of Heun's method can run with incredibly high CFL ~ 0.86, leading to by far the best performance among the schemes.

As shown in Table 3, the source terms are not directly included in the computation of the time step Δt. Instead, we select the maximal Δt based only on the fluxes and the CFL number. We choose a CFL that ensures the positivity of ρ_i, p, and ρE. The CFL values in Table 3 are "maximal" in the sense that increasing the last significant digit results in negative densities, pressure, or total energy. We then run the simulations with that CFL number multiplied with a *safety factor* specified in Table 3 to suppress spurious oscillations. We note that for the explicit schemes as well as for the second-order flux-balanced scheme we used a *safety factor* of 1, whereas the other schemes have a *safety factor* of 0.4 to mitigate the development of oscillations near the rarefaction wave.
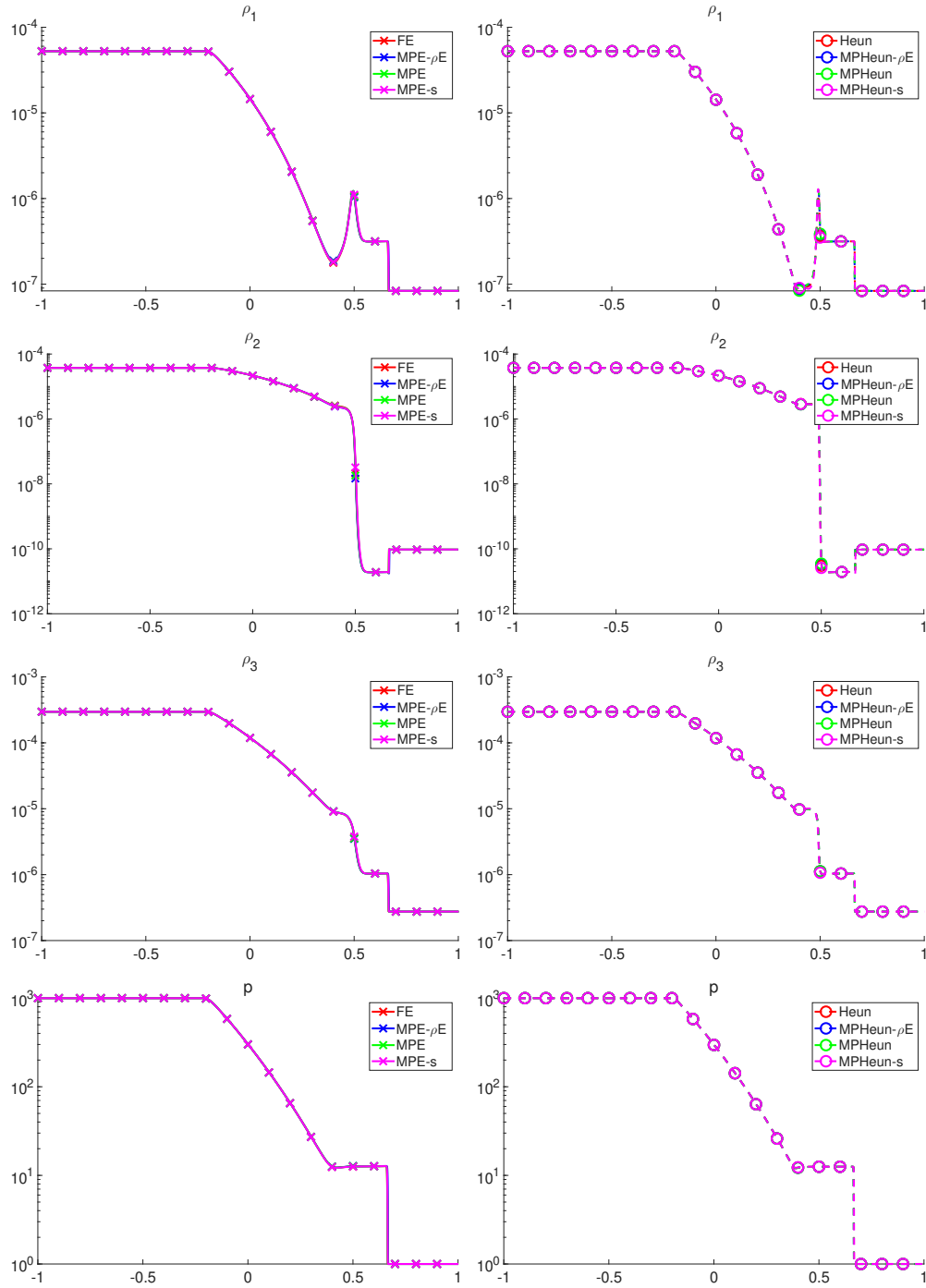
Fig. 1: Plots of the numerical solution of multi-species Euler equations (1)-(3) with source (4) using $\delta = 10^4$, $N = 4000$ cells as well as a CFL and safety factor as given in Table 3. First-order methods are at the left, second-order methods are at the right.

Table 4: Comparison of numerical schemes applied to one species Euler equations
(1)-(2) with IC (37) (Contact Discontinuity) and $N = 1000$: stable CFL and runtime
in seconds

First order				Second order			
Method	CFL	safety factor	Runtime	Method	CFL	safety factor	Runtime
FE	1	0.7	5.12	Heun	1	0.7	8.88
MPE-ρE	0.28	0.7	15.79	MPHeun-ρE	0.39	0.7	29.82
MPE	0.38	0.7	10.70	MPHeun	0.42	0.7	10.70
MPE-s	1	0.7	6.95	MPHeun-s	1	0.7	17.60

Looking at the runtime in Table 3, even if the MPRK schemes have to solve many
linear systems, they are favorable over the explicit schemes as they are faster. However,
the second-order flux-balanced version with a runtime of around 2 minutes is the best
performing scheme while the other second-order schemes range from 25 minutes to
almost 90 minutes.

Despite these differences in computational speed, the numerical approximations
are nearly indistinguishable in the visualizations (see Figure 1) for the same order of
accuracy. We notice better approximations of the contact discontinuity for all second
order methods.

5.4 Contact Discontinuity

The contact discontinuity test case, we will consider, is a variation of the 1-2-3 Problem
[38, 39], where the density jump and the velocity are much larger. It is a Riemann
problem for the Euler equations with

$$\mathbf{U}_L = \begin{pmatrix} 1 \\ 20 \\ 3 \end{pmatrix}, \quad \mathbf{U}_R = \begin{pmatrix} 10^{-6} \\ 20 \\ 3 \end{pmatrix} \tag{37}$$

as initial condition for the primitive variables (ρ, u, p).

The advantage of the flux-balanced version over the classical MPRK schemes
becomes evident when comparing the schemes to the test case with the contact
discontinuity, i.e. (1)-(2), with (37). Here, the classical MPRK schemes failed at
maintaining constant velocity and pressure, see Figure 2. In contrast, the flux-balanced
version was able to preserve the constant states. Comparing the first-order schemes,
we note that the LLF scheme is positivity-preserving for CFL=0.5, which means that
the Patankar-trick is not needed for this test case.

5.5 Vacuum

Finally, we challenge the flux-balanced version with a test leading to a vacuum. It is
another Riemann problem with a double rarefaction wave forming from the center.
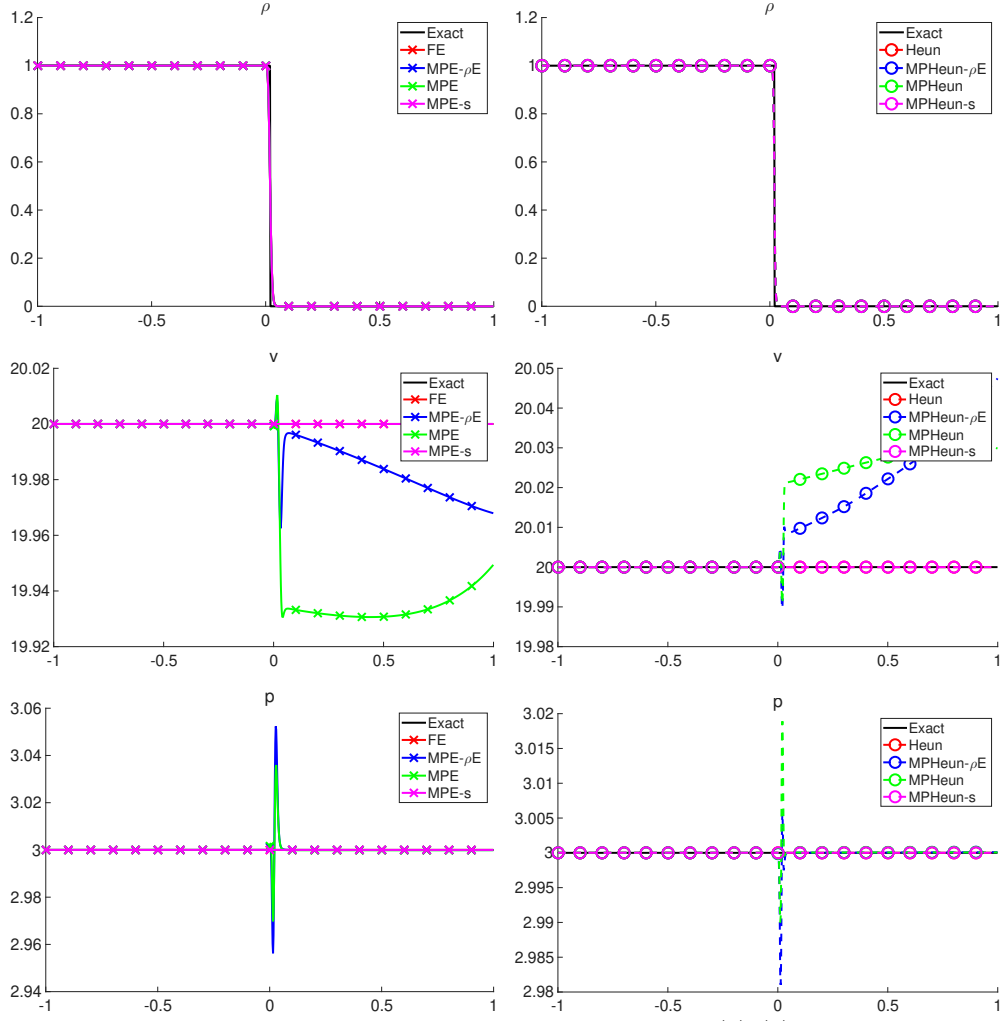
Fig. 2: Plots of the numerical solution of Euler equations (1)-(2) with contact discontinuity IC (37) with $N = 1000$ using a CFL and safety factor given in Table 4. First-order methods are left, second-order methods are on the right.

The Riemann problem is defined as

$$\mathbf{U}_L = \begin{pmatrix} 1 \\ -20 \\ 0.4 \end{pmatrix}, \quad \mathbf{U}_R = \begin{pmatrix} 1 \\ 20 \\ 0.4 \end{pmatrix} \tag{38}$$

in primitive variables. The final time is $t_{\text{end}} = 0.03$. To obtain the exact solution, we use the exact Riemann solver for the Euler equations [39] available at [40]. In Table 5, we see that the flux-balanced version is comparable to the explicit variants in terms

21

Table 5: Comparison of numerical schemes applied to Euler equations (1)-(2) with IC (38) (Vacuum) and $N = 1000$: stable CFL and runtime in seconds

First order				Second order			
Method	CFL	safety factor	Runtime	Method	CFL	safety factor	Runtime
FE	1	0.7	2.08	Heun	1	0.7	2.51
MPE-ρE	0.01	0.7	141.69	MPHeun-ρE	0.10	0.7	32.14
MPE	0.01	0.7	131.54	MPHeun	0.13	0.7	24.83
MPE-s	1	0.7	2.18	MPHeun-s	1	0.7	5.24
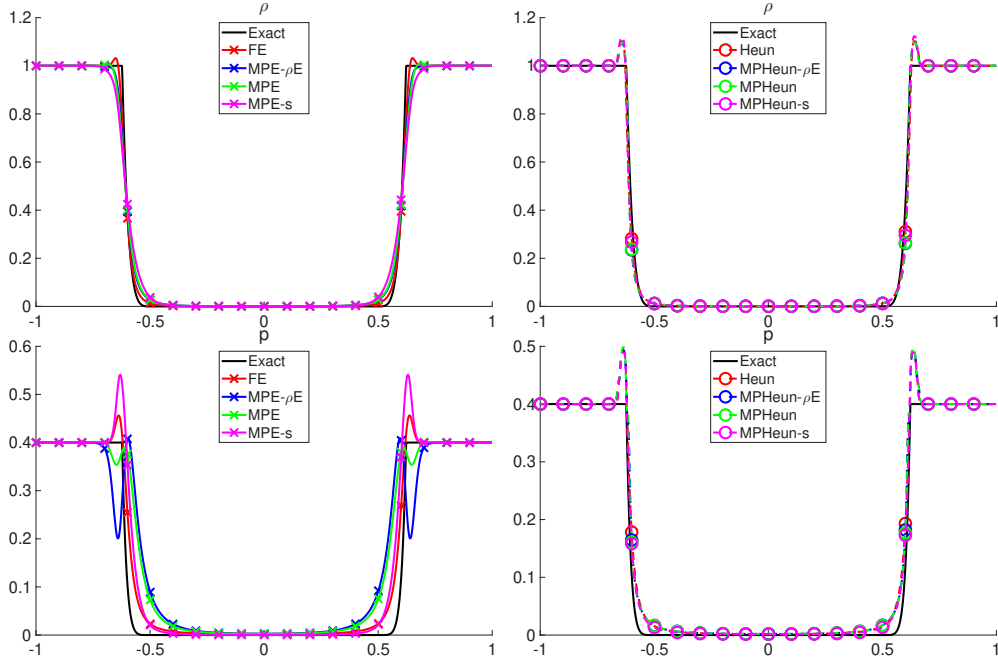
Fig. 3: Plots of the numerical solution of Euler equations (1)-(2) with vacuum IC (38) with $N = 1000$, and with CFL and safety factor given in Table 5. First-order methods are left, second-order methods are on the right.

of runtime, but it is much faster than the MPRK schemes. Additionally, they show similar numerical results, see Figure 3, while the MPRK schemes show nonphysical oscillations, which resulted in negative approximations for the pressure with a CFL number greater or equal to 0.02.

Remark 3. *Using a CFL number of 0.5 and setting $\rho_R = 10^{-2}$, we observe that the flux-balanced MPE scheme yields a negative pressure after one time step. Hence, more research is needed in this field to provably control the pressure.*

6 Summary and Outlook

In this work, we showed how to embed an MPRK scheme into a Finite Volume framework for Euler equations of gas dynamics, also taking stiff source terms into consideration. Furthermore, we showed that the MP-trick should not be applied blindly to all components of the balance law that should stay positive, but that it is instead of interest to use MP only for the densities and to explicitly weigh in a consistent manner also the other numerical fluxes to be consistent with the contact discontinuities. We performed numerical experiments confirming the expected order of accuracy and the performance of the different numerical schemes when applied to various test cases. The results show that the flux-balanced version of the MPRK is much more stable than the classical MPRK schemes for large CFLs.

Future works include the investigation of a hybrid variant of the explicit and flux-balanced schemes including further Patankar-type schemes, and the application of the flux-balanced schemes to further hyperbolic systems such as the shallow water equations. We will also consider in future works the combination of the MP trick on the energy equation and the simultaneous preservation of contact discontinuities. Finally, our future work will also involve studying algorithms that are capable of preserving entropy properties and positivity simultaneously. We will investigate the algorithm's capability of controlling the positivity of internal energy or pressure via the EoS, and additionally, we will pursue the approach of solving the Euler system in primitive variables to evolve the internal energy or pressure directly while keeping conservation.

Statements and Declarations

Conflict of interest. On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] Sandu, A.: Positive numerical integration methods for chemical kinetic systems. J. Comput. Phys. **170**(2), 589–602 (2001) https://doi.org/10.1006/jcph.2001.6750

[2] Huang, J., Shu, C.-W.: Positivity-preserving time discretizations for production-destruction equations with applications to non-equilibrium flows. J. Sci. Comput. **78**(3), 1811–1839 (2019)

[3] Lax, P., Wendroff, B.: Systems of conservation laws. Comm. Pure Appl. Math. **13**(2), 217–237 (1960) https://doi.org/10.1002/cpa.3160130205

[4] Perthame, B., Shu, C.-W.: On positivity preserving finite volume schemes for Euler equations. Numer. Math. **73**(1), 119–130 (1996) https://doi.org/10.1007/s002110050187

[5] Clayton, B., Guermond, J.-L., Maier, M., Popov, B., Tovar, E.J.: Robust second-order approximation of the compressible Euler equations with an arbitrary equation of state. J. Comput. Phys. **478**, 111926–21 (2023) https://doi.org/10.1016/j.jcp.2023.111926

[6] Rueda-Ramírez, A.M., Pazner, W., Gassner, G.J.: Subcell limiting strategies for discontinuous Galerkin spectral element methods. Comput. & Fluids **247**, 105627–19 (2022) https://doi.org/10.1016/j.compfluid.2022.105627

[7] Gaburro, E., Dumbser, M.: A posteriori subcell finite volume limiter for general $P_N P_M$ schemes: applications from gasdynamics to relativistic magneto-hydrodynamics. J. Sci. Comput. **86**(3), 37–41 (2021) https://doi.org/10.1007/s10915-020-01405-8

[8] Dzanic, T.: Continuously bounds-preserving discontinuous Galerkin methods for hyperbolic conservation laws. J. Comput. Phys. **508**, 113010–25 (2024) https://doi.org/10.1016/j.jcp.2024.113010

[9] Dzanic, T.: A note on higher-order and nonlinear limiting approaches for continuously bounds-preserving discontinuous Galerkin methods. J. Comput. Phys. **516**, 113367–7 (2024) https://doi.org/10.1016/j.jcp.2024.113367

[10] Wu, K., Shu, C.-W.: Geometric quasilinearization framework for analysis and design of bound-preserving schemes. SIAM Rev. **65**(4), 1031–1073 (2023) https://doi.org/10.1137/21M1458247

[11] Ern, A., Guermond, J.-L.: Invariant-domain-preserving high-order time stepping: I. explicit Runge-Kutta schemes. SIAM J. Sci. Comput. **44**(5), 3366–3392 (2022) https://doi.org/10.1137/21M145793X

[12] Ern, A., Guermond, J.-L.: Invariant-domain preserving high-order time stepping: II. IMEX schemes. SIAM J. Sci. Comput. **45**(5), 2511–2538 (2023) https://doi.org/10.1137/22M1505025

[13] Sandu, A.: Positive numerical integration methods for chemical kinetic systems. J. Comput. Phys. **170**(2), 589–602 (2001) https://doi.org/10.1006/jcph.2001.6750

[14] Blanes, Sergio, Iserles, Arieh, Macnamara, Shev: Positivity-preserving methods for ordinary differential equations. ESAIM: M2AN **56**(6), 1843–1870 (2022) https://doi.org/10.1051/m2an/2022042

[15] Hubbard, M.E., Ricchiuto, M., Sármány, D.: Space-time residual distribution on moving meshes. Comput. Math. Appl. **79**(5), 1561–1589 (2020) https://doi.org/

10.1016/j.camwa.2019.09.019

[16] Ricchiuto, M.: Contributions to the development of residual discretizations for hyperbolic conservation laws with application to shallow water flows. Habilitation thesis, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland (December 2011). Habilitation thesis

[17] Patankar, S.V.: Numerical Heat Transfer and Fluid Flow. Series in computational methods in mechanics and thermal sciences. Hemisphere Pub. Corp. New York, Washington (1980). http://opac.inria.fr/record=b1085925

[18] Burchard, H., Deleersnijder, E., Meister, A.: A high-order conservative Patankar-type discretisation for stiff systems of production-destruction equations. Appl. Numer. Math. **47**(1), 1–30 (2003) https://doi.org/10.1016/S0168-9274(03)00101-6

[19] Martiradonna, A., Colonna, G., Diele, F.: GeCo: Geometric Conservative nonstandard schemes for biochemical systems. Appl. Numer. Math. **155**, 38–57 (2020) https://doi.org/10.1016/j.apnum.2019.12.004

[20] Kopecz, S., Meister, A.: On order conditions for modified Patankar-Runge-Kutta schemes. Appl. Numer. Math. **123**, 159–179 (2018)

[21] Ávila, A.I., Kopecz, S., Meister, A.: A comprehensive theory on generalized BBKS schemes. Appl. Numer. Math. **157**, 19–37 (2020) https://doi.org/10.1016/j.apnum.2020.05.027

[22] Kopecz, S., Meister, A.: Unconditionally positive and conservative third order modified Patankar-Runge-Kutta discretizations of production-destruction systems. BIT **58**(3), 691–728 (2018) https://doi.org/10.1007/s10543-018-0705-1

[23] Huang, J., Zhao, W., Shu, C.-W.: A third-order unconditionally positivity-preserving scheme for production-destruction equations with applications to non-equilibrium flows. J. Sci. Comput. **79**(2), 1015–1056 (2019)

[24] Öffner, P., Torlo, D.: Arbitrary high-order, conservative and positivity preserving Patankar-type deferred correction schemes. Appl. Numer. Math. **153**, 15–34 (2020)

[25] Izzo, G., Messina, E., Pezzella, M., Vecchio, A.: Modified patankar linear multistep methods for production-destruction systems. Journal of Scientific Computing **102**(3), 87 (2025)

[26] Izgin, T., Ketcheson, D.I., Meister, A.: Order conditions for Runge–Kutta-like methods with solution-dependent coefficients. Communications in Applied Mathematics and Computational Science **20-1**, 29–66 (2025) https://doi.org/10.2140/camcos.2025.20.29

[27] Izgin, T.: A unifying theory for Runge-Kutta-like time integrators: Convergence

and stability. PhD thesis, University of Kassel (2024). https://doi.org/10.17170/kobra-202402059522

[28] Izgin, T., Kopecz, S., Meister, A.: On the stability of unconditionally positive and linear invariants preserving time integration schemes. SIAM J. Numer. Anal. **60**(6), 3029–3051 (2022) https://doi.org/10.1137/22M1480318

[29] Ciallella, M., Micalizzi, L., Öffner, P., Torlo, D.: An arbitrary high order and positivity preserving method for the shallow water equations. Comput. & Fluids **247**, 105630 (2022) https://doi.org/10.1016/j.compfluid.2022.105630

[30] Kuzmin, D., Möller, M.: Algebraic flux correction ii. compressible euler equations. In: Kuzmin, D., Löhner, R., Turek, S. (eds.) Flux-Corrected Transport: Principles, Algorithms, and Applications, pp. 207–250. Springer, Berlin, Heidelberg (2005). https://doi.org/10.1007/3-540-27206-2_7 . https://doi.org/10.1007/3-540-27206-2_7

[31] Abgrall, R., Bacigaluppi, P., Tokareva, S.: A high-order nonconservative approach for hyperbolic equations in fluid dynamics. Computers & Fluids **169**, 10–22 (2018)

[32] Gaburro, E., Boscheri, W., Chiocchetti, S., Ricchiuto, M.: Discontinuous Galerkin schemes for hyperbolic systems in non-conservative variables: quasi-conservative formulation with subcell finite volume corrections. Computer Methods in Applied Mechanics and Engineering **431**, 117311 (2024)

[33] Wang, W., Shu, C.-W., Yee, H.C., Sjögreen, B.: High-order well-balanced schemes and applications to non-equilibrium flow. Journal of Computational Physics **228**(18), 6682–6702 (2009) https://doi.org/10.1016/j.jcp.2009.05.028

[34] Torlo, D., Öffner, P., Ranocha, H.: Issues with positivity-preserving Patankar-type schemes. Appl. Numer. Math. **182**, 117–147 (2022) https://doi.org/10.1016/j.apnum.2022.07.014

[35] Izgin, T., Ranocha, H.: Using Machine Learning to Design Time Step Size Controllers for Stable Time Integrators. https://arxiv.org/abs/2312.01796 (2025) arXiv:2312.01796 [math.NA]

[36] Kopecz, S., Meister, A.: Unconditionally positive and conservative third order modified Patankar-Runge-Kutta discretizations of production-destruction systems. BIT **58**(3), 691–728 (2018)

[37] LeVeque, R.J.: Finite Volume Methods for Hyperbolic Problems. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge (2002)

[38] Einfeldt, B., Munz, C.-D., Roe, P.L., Sjögreen, B.: On Godunov-type methods near low densities. Journal of computational physics **92**(2), 273–295 (1991)

[39] Toro, E.F.: Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction. Springer, Berlin Heidelberg (2009)

[40] Wu, Y.: exact Riemann solver for compressible Euler equations (code). https://github.com/YueWu2002/Euler_exact_Riemann (2024)