Synchronized flux limiting for gas dynamics variables

Christoph Lohmann, Dmitri Kuzmin

Institute of Applied Mathematics (LS III), TU Dortmund University, Vogelpothsweg 87, D-44227 Dortmund, Germany

Abstract

This work addresses the design of failsafe flux limiters for systems of conserved quantities and derived variables in numerical schemes for the equations of gas dynamics. Building on Zalesak's multidimensional flux-corrected transport (FCT) algorithm, we construct a new positivity-preserving limiter for the density, total energy, and pressure. The bounds for the underlying inequality constraints are designed to enforce local maximum principles in regions of strong density variations and become less restrictive in smooth regions. The proposed approach leads to closed-form expressions for the synchronized correction factors without the need to solve inequality-constrained optimization problems. A numerical study is performed for the compressible Euler equations discretized using a finite element based FCT scheme.

Keywords: systems of conservation laws, local extremum diminishing limiters, positivity preservation, flux-corrected transport/remapping

1. Introduction

The ability to enforce local discrete maximum principles and/or positivity preservation for a set of coupled gas dynamics variables is a highly desired property of high-resolution schemes for the compressible Euler equations [10, 14, 18] and constrained interpolation (remapping) algorithms [2, 12, 13, 21]

Email addresses: christoph.lohmann@math.tu-dortmund.de (Christoph Lohmann), kuzmin@math.uni-dortmund.de (Dmitri Kuzmin)

for systems of conserved quantities. Many existing tools for constraining the quantities of interest are based on the use of limiting techniques for numerical fluxes associated with oscillatory antidiffusive components of a high-order approximation. The underlying design principles trace their origins to the classical *flux-corrected transport* (FCT) algorithm introduced by Boris and Book [3, 4, 5] and Zalesak [23] in the 1970s. Löhner et al. [15] extended the FCT methodology to unstructured grid finite element methods and systems of conservation laws. The first use of flux limiters in the context of remapping goes back to the work of Smolarkiewicz and Grell [19] who proposed a class of nonconservative monotone interpolation schemes. Conservative flux-corrected remap (FCR) methods were developed in [11, 14, 12, 21]. As shown by Bochev et al. [2], the FCR approach to calculating the correction factors is equivalent to solving an optimization problem with simple box constraints corresponding to a worst-case scenario. Advanced algorithms for constrained optimization-based data transfer were proposed in [1, 2, 13].

Flux limiting techniques for systems of coupled variables can be classified into sequential [12] and synchronized [11, 14, 13, 18] algorithms. A sequential limiter constrains each quantity of interest under worst-case assumptions regarding the fluxes that depend on other variables. In synchronized FCT algorithms [10, 11, 15, 14], the antidiffusive fluxes are multiplied by the minimum of the correction factors for selected control variables. Due to the involved linearizations, such algorithms may require additional a posteriori corrections to guarantee the nonnegativity of the pressure and internal energy [11, 24]. In optimization-based synchronized algorithms, different correction factors may be used in different conservation laws provided that the imposed constraints are satisfied for each quantity of interest [13]. However, the cost of coupled flux optimization is rather high, which has led Bochev et al. [1] to favor globally conservative formulations of the constrained remap problem.

In this paper, we improve the synchronized FCT algorithm presented in [10, 11] by introducing new limiters for the energy and pressure. In contrast to approaches that rely on linearized transformations of variables, the proposed limiting strategy does not involve any linearizations and guarantees positivity preservation without a posteriori fixes. Moreover, the bounds for FCT are designed to prevent unnecessary limiting in regions of constant pressure. The calculation of correction factors for the synchronized FCT limiter does not require solving inequality-constrained optimization problems, which makes it an inexpensive alternative to synchronized optimization-based limiters [2, 13]. The ability of the proposed algorithm to handle shocks and contact discontinuities is illustrated by a numerical study for the Euler equations.

2. Synchronized flux limiting

Consider a system of conservation laws for $U = [\rho, \rho\mathbf{v}, \rho E]^T$, where ρ is the density, $\mathbf{v}$ is the velocity and E is the total energy. In the case of an ideal polytropic gas, the pressure p is given by the equation of state

$$p = (\gamma - 1)\left(\rho E - \frac{|\rho\mathbf{v}|^2}{2\rho}\right), \tag{1}$$

where γ stands for the constant ratio of specific heats ($\gamma = 1.4$ for air).

Let U_i denote a numerical approximation to the vector U of gas dynamics variables at the ith nodal point or control volume. The simplest representatives of flux-corrected transport (FCT) and flux-corrected remapping (FCR) algorithms are based on the following predictor-corrector strategy:

1. Calculate a low-order approximation U_i^L using a numerical scheme which is guaranteed to satisfy all relevant maximum principles.
2. Decompose the difference between U_i^L and a high-order approximation U_i^H into a sum of antidiffusive fluxes $F_{ij} = [f_{ij}^\rho, \mathbf{f}_{ij}^{\rho v}, f_{ij}^{\rho E}]^T$ such that

$$m_i U_i^H = m_i U_i^L + \sum_{j \neq i} F_{ij}, \qquad F_{ji} = -F_{ij}, \tag{2}$$

   where m_i is a positive diagonal entry of the (lumped) mass matrix.
3. Multiply F_{ij} and its companion F_{ji} by a solution-dependent correction factor $\alpha_{ij} \in [0, 1]$ such that the flux-corrected approximation

$$m_i U_i = m_i U_i^L + \sum_{j \neq i} \alpha_{ij} F_{ij}, \qquad \alpha_{ji} = \alpha_{ij} \tag{3}$$

   satisfies inequality constraints of the form

$$u_i^{\min} \leq u_i \leq u_i^{\max} \tag{4}$$

   for each scalar quantity of interest u (density, energy, pressure etc.).

Following [11, 15], we will limit all components of F_{ij} using the same scalar correction factor α_{ij}. The choice $\alpha_{ij} \equiv 1$ corresponds to the high-order approximation U_i^H, whereas $\alpha_{ij} \equiv 0$ corresponds to the low-order approximation U_i^L. Since the latter is assumed to satisfy the maximum principles, the bounds for (4) are commonly defined in terms of U^L as follows:

$$u_i^{\max} = \max_{j \in \mathcal{N}(i)} u_j^L, \qquad u_i^{\min} = \min_{j \in \mathcal{N}(i)} u_j^L, \tag{5}$$

where u_i^L is the low-order approximation to the quantity of interest and $\mathcal{N}(i)$ is the set of nodes containing i and its nearest neighbors $j \neq i$. Throughout this paper, the shorthand notation "$j \neq i$" is used for $j \in \mathcal{N}(i) \backslash \{i\}$.

For a scalar conserved quantity u, nearly optimal correction factors α_{ij} can be calculated using Zalesak's multidimensional FCT limiter [23] which we use to constrain the density ($u = \rho$) in the next section. The design of FCT algorithms for systems is more involved because of the strong coupling between the quantities of interest [10, 15]. For example, any antidiffusive correction to ρ_i may produce an undershoot or overshoot in $\mathbf{v}_i := \frac{(\rho \mathbf{v})_i}{\rho_i}$ and/or $E := \frac{(\rho E)_i}{\rho_i}$ even if the values of $(\rho \mathbf{v})_i$ and $(\rho E)_i$ remain unchanged. Similarly, any adjustment of the conservative variables may result in a violation of local bounds for the pressure p defined by the equation of state (1). Hence, possible changes in the values of derived quantities must be taken into account when it comes to limiting the changes in the conservative variables.

In the next three sections, we present a new synchronized FCT algorithm for constraining the density, energy, and pressure. After formulating the inequality constraints for each variable, we derive upper bounds for the correction factors α_{ij} and design practical algorithms for enforcing these bounds.

3. The density limiter

The density ρ is easy to limit and represents a natural control variable because it is discontinuous at shocks and contact discontinuities alike (in contrast to the velocity $\mathbf{v}$ and pressure p which are continuous at a contact discontinuity). For this reason, the value of the synchronized correction factor α_{ij} should not exceed that of $\alpha_{ij}^{(\rho)}$ such that the flux-corrected nodal value ρ_i satisfies

$$\rho_i^{\min} \le \rho_i = \rho_i^L + \frac{1}{m_i}\sum_{j\neq i} \alpha_{ij} f_{ij}^{(\rho)} \le \rho_i^{\max} \qquad \forall \alpha_{ij} \le \alpha_{ij}^{(\rho)}. \tag{6}$$

The bounds $\rho_i^{\max}$ and $\rho_i^{\min}$ are defined by (5) with $u = \rho$. That is,

$$\rho_i^{\max} = \max_{j\in\mathcal{N}(i)} \rho_j^L, \qquad \rho_i^{\min} = \min_{j\in\mathcal{N}(i)} \rho_j^L. \tag{7}$$

Assuming the worst-case scenario (no cancellation of positive and negative fluxes), the provisional correction factor $\alpha_{ij}^{(\rho)}$ should be chosen so that

$$m_i Q_i^{-,\rho} \le R_i^{-,\rho} P_i^{-,\rho} \le \sum_{j\neq i} \alpha_{ij}^{(\rho)} f_{ij}^{(\rho)} \le R_i^{+,\rho} P_i^{+,\rho} \le m_i Q_i^{+,\rho}, \tag{8}$$

where

$$P_i^{+,\rho} = \sum_{j\neq i} \max\left\{0, f_{ij}^{(\rho)}\right\}, \qquad P_i^{-,\rho} = \sum_{j\neq i} \min\left\{0, f_{ij}^{(\rho)}\right\}, \tag{9}$$

$$Q_i^{+,\rho} = \rho_i^{\max} - \rho_i^L, \qquad Q_i^{-,\rho} = \rho_i^{\min} - \rho_i^L, \tag{10}$$

$$R_i^{+,\rho} = \min\left\{1, \frac{m_i Q_i^{+,\rho}}{P_i^{+,\rho}}\right\}, \qquad R_i^{-,\rho} = \min\left\{1, \frac{m_i Q_i^{-,\rho}}{P_i^{-,\rho}}\right\}. \tag{11}$$

Additionally, the symmetry condition $\alpha_{ji}^{(\rho)} = \alpha_{ij}^{(\rho)}$ must hold for the limited antidiffusive fluxes to remain skew-symmetric. Correction factors satisfying the above criteria can be determined using Zalesak's formula [23]

$$\alpha_{ij}^{(\rho)} = \left\{ \begin{array}{ll} \min\left\{R_i^{+,\rho}, R_j^{-,\rho}\right\} & \text{if } f_{ij}^{(\rho)} \ge 0, \\ \min\left\{R_i^{-,\rho}, R_j^{+,\rho}\right\} & \text{if } f_{ij}^{(\rho)} < 0. \end{array} \right. \tag{12}$$

This yields the first provisional bound $\alpha_{ij}^{(\rho)}$ for the synchronized correction factor α_{ij}. Using this result, we define the tight density bounds

$$\tilde{\rho}_i^{\max} = \rho_i^L + \frac{1}{m_i}\sum_{j\neq i} \max\left\{0, \alpha_{ij}^{(\rho)} f_{ij}^{(\rho)}\right\}, \tag{13}$$

$$\tilde{\rho}_i^{\min} = \rho_i^L + \frac{1}{m_i}\sum_{j\neq i} \min\left\{0, \alpha_{ij}^{(\rho)} f_{ij}^{(\rho)}\right\} \tag{14}$$

which we will need to construct the bounds for the energy and pressure below.

4. The energy limiter

The second natural control variable for a synchronized FCT algorithm is the total energy. The inequality constraints for $(\rho E)_i$ are given by

$$(\rho E)_i^{\min} \leq (\rho E)_i = (\rho E)_i^L + \frac{1}{m_i} \sum_{j \neq i} \alpha_{ij} f_{ij}^{(\rho E)} \leq (\rho E)_i^{\max}, \tag{15}$$

$$(\rho E)_i^{\max} = \max_{j \in \mathcal{N}(i)} (\rho E)_j^L, \qquad (\rho E)_i^{\min} = \min_{j \in \mathcal{N}(i)} (\rho E)_j^L \tag{16}$$

and the upper bound for α_{ij} can be determined using Zalesak's limiter.

The local maximum principle for E_i can be formulated as follows:

$$E_i^{\min} \leq E_i = \frac{(\rho E)_i}{\rho_i} = \frac{(\rho E)_i^L + \frac{1}{m_i} \sum_{j \neq i} \alpha_{ij} f_{ij}^{(\rho E)}}{\rho_i^L + \frac{1}{m_i} \sum_{j \neq i} \alpha_{ij} f_{ij}^{(\rho)}} \leq E_i^{\max}, \tag{17}$$

$$E_i^{\max} = \max_{j \in \mathcal{N}(i)} \frac{(\rho E)_j^L}{\rho_j^L}, \qquad E_i^{\min} = \min_{j \in \mathcal{N}(i)} \frac{(\rho E)_j^L}{\rho_j^L}. \tag{18}$$

According to (17), the limited antidiffusive fluxes must satisfy

$$\rho_i^L \left(E_i^{\min} - E_i^L \right) \leq \frac{1}{m_i} \sum_{j \neq i} \alpha_{ij} \left(f_{ij}^{(\rho E)} - E_i^{\min} f_{ij}^{(\rho)} \right), \tag{19}$$

$$\rho_i^L \left(E_i^{\max} - E_i^L \right) \geq \frac{1}{m_i} \sum_{j \neq i} \alpha_{ij} \left(f_{ij}^{(\rho E)} - E_i^{\max} f_{ij}^{(\rho)} \right). \tag{20}$$

An explicit formula for α_{ij} satisfying such energy constraints can be derived following the methodology developed in [12] for enforcing velocity constraints in sequential FCR schemes. In our experience, this way to constrain E_i produces poor results in the context of synchronized FCT because it imposes artificial constraints on the fluxes $f_{ij}^{(\rho)}$, especially in the limit $f_{ij}^{(\rho E)} \to 0$.

To prevent unnecessary limiting of $f_{ij}^{(\rho)}$, the energy bounds can be extended to cover the full range of admissible density values. We have

$$\rho_i E_i^{\max} = \left(\rho_i^L + \frac{1}{m_i} \sum_{j \neq i} \alpha_{ij} f_{ij}^{(\rho)} \right) E_i^{\max} \leq \tilde{\rho}_i^{\max} E_i^{\max}, \tag{21}$$

$$\rho_i E_i^{\min} = \left(\rho_i^L + \frac{1}{m_i} \sum_{j \neq i} \alpha_{ij} f_{ij}^{(\rho)} \right) E_i^{\min} \geq \tilde{\rho}_i^{\min} E_i^{\min}, \tag{22}$$

where $\tilde{\rho}_i^{\max}$ and $\tilde{\rho}_i^{\min}$ are defined by (13) and (14), respectively. Hence,

$$\tilde{\rho}_i^{\min} E_i^{\min} - (\rho E)_i^L \leq \frac{1}{m_i} \sum_{j \neq i} \alpha_{ij} f_{ij}^{(\rho E)} \leq \tilde{\rho}_i^{\max} E_i^{\max} - (\rho E)_i^L \tag{23}$$

is a sufficient condition for the flux-corrected energy E_i to satisfy

$$\tilde{\rho}_i^{\min} E_i^{\min} \leq \rho_i E_i = (\rho E)_i \leq \tilde{\rho}_i^{\max} E_i^{\max} \tag{24}$$

for any ρ_i in the range $[\tilde{\rho}_i^{\min}, \tilde{\rho}_i^{\max}]$ of admissible values, as determined previously by the density limiter $\alpha_{ij}^{(\rho)}$. Combining this estimate and (15), we define the local bounds for the proposed energy limiter as follows:

$$\max\left\{(\rho E)_i^{\min}, \tilde{\rho}_i^{\min} E_i^{\min}\right\} \leq (\rho E)_i \leq \min\left\{(\rho E)_i^{\max}, \tilde{\rho}_i^{\max} E_i^{\max}\right\}. \tag{25}$$

Division by ρ_i yields the corresponding inequality constraints for E_i

$$\max\left\{\frac{(\rho E)_i^{\min}}{\rho_i}, \frac{\tilde{\rho}_i^{\min} E_i^{\min}}{\rho_i}\right\} \leq E_i \leq \min\left\{\frac{(\rho E)_i^{\max}}{\rho_i}, \frac{\tilde{\rho}_i^{\max} E_i^{\max}}{\rho_i}\right\}. \tag{26}$$

Since ρ_i is allowed to float in the range $[\tilde{\rho}_i^{\min}, \tilde{\rho}_i^{\max}]$, we have the estimate

$$\max\left\{\frac{(\rho E)_i^{\min}}{\tilde{\rho}_i^{\max}}, \frac{\tilde{\rho}_i^{\min} E_i^{\min}}{\tilde{\rho}_i^{\max}}\right\} \leq E_i \leq \min\left\{\frac{(\rho E)_i^{\max}}{\tilde{\rho}_i^{\min}}, \frac{\tilde{\rho}_i^{\max} E_i^{\max}}{\tilde{\rho}_i^{\min}}\right\}. \tag{27}$$

That is, the sharpness of the local bounds for E_i depends on the ratio of the tight density bounds $\frac{\tilde{\rho}_i^{\min}}{\rho_i^{\max}}$ which approaches 1 in the limit $\alpha_{ij}^{(\rho)} \to 0$.

To enforce the energy constraints, we use $\alpha_{ij} \leq \alpha_{ij}^{(\rho,E)} \leq \alpha_{ij}^{(\rho)}$ such that

$$m_i Q_i^{-,E} \leq R_i^{-,E} P_i^{-,E} \leq \sum_{j \neq i} \alpha_{ij}^{(\rho,E)} f_{ij}^{(\rho E)} \leq R_i^{+,E} P_i^{+,E} \leq m_i Q_i^{+,E}, \tag{28}$$

where

$$P_i^{+,E} = \sum_{j \neq i} \max\left\{0, \alpha_{ij}^{(\rho)} f_{ij}^{(\rho E)}\right\}, \quad P_i^{-,E} = \sum_{j \neq i} \min\left\{0, \alpha_{ij}^{(\rho)} f_{ij}^{(\rho E)}\right\}, \tag{29}$$

$$\begin{aligned} Q_i^{+,E} &= \min\left\{(\rho E)_i^{\max}, \tilde{\rho}_i^{\max} E_i^{\max}\right\} - (\rho E)_i^L, \\ Q_i^{-,E} &= \max\left\{(\rho E)_i^{\min}, \tilde{\rho}_i^{\min} E_i^{\min}\right\} - (\rho E)_i^L, \end{aligned} \tag{30}$$

$$R_i^{+,E} = \min\left\{1, \frac{m_i Q_i^{+,E}}{P_i^{+,E}}\right\}, \qquad R_i^{-,E} = \min\left\{1, \frac{m_i Q_i^{-,E}}{P_i^{-,E}}\right\}. \tag{31}$$

The formula for calculating the correction factors $\alpha_{ij}^{(\rho,E)}$ is given by

$$\alpha_{ij}^{(\rho,E)} = \begin{cases} \min\left\{R_i^{+,E}, R_j^{-,E}\right\} \alpha_{ij}^{(\rho)} & \text{if } f_{ij}^{(\rho E)} \geq 0, \\ \min\left\{R_i^{-,E}, R_j^{+,E}\right\} \alpha_{ij}^{(\rho)} & \text{if } f_{ij}^{(\rho E)} < 0. \end{cases} \tag{32}$$

This FCT algorithm uses the knowledge that $\alpha_{ij}^{(\rho,E)} \leq \alpha_{ij}^{(\rho)}$. In a practical implementation, we apply $\alpha_{ij}^{(\rho)}$ to all components of $F_{ij} = [f_{ij}^{\rho}, \mathbf{f}_{ij}^{\rho v}, f_{ij}^{\rho E}]^T$ and pass the density-limited fluxes $\alpha_{ij}^{(\rho)} F_{ij}$ to the energy limiter.

Remark. The algorithm presented in this section can also be used to constrain the components of $(\rho\mathbf{v})_i$ and $\mathbf{v}_i$. However, componentwise limiting of vector fields violates the principle of frame indifference. For this reason, the use of frame invariant velocity/momentum limiters is recommended [16, 25].

5. The pressure limiter

In many cases, the difference between the solutions produced by the energy limiter $\alpha_{ij} = \alpha_{ij}^{(\rho,E)}$ and the density limiter $\alpha_{ij} = \alpha_{ij}^{(\rho)}$ is marginal. However, it is essential to ensure that the pressure p does not become negative in the process of flux correction. If the limiting procedure does not guarantee this, it must be equipped with a 'failsafe' postprocessing technique for canceling the offending antidiffusive fluxes [11, 24]. Otherwise, the approximate Riemann solver may crash when it comes to calculating the speed of sound.

In this section, we design a pressure limiter which guarantees positivity preservation *a priori*. Let the local bounds for p_i be defined by

$$\tilde{\rho}_i^{\min} p_i^{\min} \leq \rho_i p_i = \rho_i(\gamma - 1)\left[(\rho E)_i - \frac{|(\rho\mathbf{v})_i|^2}{2\rho_i}\right] \leq \tilde{\rho}_i^{\max} p_i^{\max}, \tag{33}$$

$$p_i^{\max} = \max_{j\in\mathcal{N}(i)} (\gamma - 1)\left[(\rho E)_j - \frac{|(\rho\mathbf{v})_j|^2}{2\rho_j}\right], \tag{34}$$

$$p_i^{\min} = \min_{j\in\mathcal{N}(i)} (\gamma - 1)\left[(\rho E)_j - \frac{|(\rho\mathbf{v})_j|^2}{2\rho_j}\right]. \tag{35}$$

The use of $\tilde{\rho}_i^{\max}$ and $\tilde{\rho}_i^{\min}$ in the pressure constraints (33) prevents the limiter from canceling all antidiffusive fluxes in regions of constant pressure. Note that the lower bound $\tilde{\rho}_i^{\min} p_i^{\min}$ is nonnegative and $\tilde{\rho}_i^{\substack{\max\\ \min}} \to \rho_i$ as $\alpha_{ij}^{(\rho)} \to 0$.

The constrained pressure p_i depends on the synchronized correction factors $\alpha_{ij} \le \alpha_{ij}^{(\rho,E)} \le \alpha_{ij}^{(\rho)}$ and conservative fluxes $(f_{ij}^{\rho}, \mathbf{f}_{ij}^{\rho v}, f_{ij}^{\rho E})$ as follows:

$$
\begin{aligned}
\frac{\rho_i p_i}{\gamma - 1} &= \left(\rho_i^L + \frac{1}{m_i}\sum_{j\neq i}\alpha_{ij} f_{ij}^{(\rho)}\right)\left((\rho E)_i^L + \frac{1}{m_i}\sum_{j\neq i}\alpha_{ij} f_{ij}^{(\rho E)}\right) \\
&- \frac{1}{2}\left|(\rho\mathbf{v})_i^L + \frac{1}{m_i}\sum_{j\neq i}\alpha_{ij}\mathbf{f}_{ij}^{(\rho\mathbf{v})}\right|^2 \\
&= \rho_i^L(\rho E)_i^L + \frac{1}{m_i}\sum_{j\neq i}\alpha_{ij}\left(\rho_i^L f_{ij}^{(\rho E)} + (\rho E)_i^L f_{ij}^{(\rho)}\right) \\
&+ \left(\frac{1}{m_i}\sum_{j\neq i}\alpha_{ij} f_{ij}^{(\rho)}\right)\left(\frac{1}{m_i}\sum_{k\neq i}\alpha_{ik} f_{ik}^{(\rho E)}\right) \\
&- \frac{1}{2}|(\rho\mathbf{v})_i^L|^2 - \frac{1}{m_i}\sum_{j\neq i}\alpha_{ij}(\rho\mathbf{v})_i^L\cdot\mathbf{f}_{ij}^{(\rho\mathbf{v})} \\
&- \frac{1}{2}\left|\frac{1}{m_i}\sum_{j\neq i}\alpha_{ij}\mathbf{f}_{ij}^{(\rho\mathbf{v})}\right|^2.
\end{aligned}
$$

The estimates corresponding to the worst-case scenario are given by

$$
\frac{m_i^2 Q_i^{-,p}}{\gamma - 1} \le R_i^{-,p}P_i^{-,p} \le P_i^{(p)} \le R_i^{+,p}P_i^{+,p} \le \frac{m_i^2 Q_i^{+,p}}{\gamma - 1}, \tag{36}
$$

where

$$
P_i^{(p)} = \frac{m_i^2(\rho_i p_i - \rho_i^L p_i^L)}{\gamma - 1} \tag{37}
$$

$$
\begin{aligned}
&= m_i \sum_{j\neq i}\alpha_{ij}\left(\rho_i^L f_{ij}^{(\rho E)} + (\rho E)_i^L f_{ij}^{(\rho)} - (\rho\mathbf{v})_i^L\cdot\mathbf{f}_{ij}^{(\rho\mathbf{v})}\right) \\
&+ \left(\sum_{j\neq i}\alpha_{ij} f_{ij}^{(\rho)}\right)\left(\sum_{k\neq i}\alpha_{ik} f_{ik}^{(\rho E)}\right) - \frac{1}{2}\left|\sum_{j\neq i}\alpha_{ij}\mathbf{f}_{ij}^{(\rho\mathbf{v})}\right|^2,
\end{aligned} \tag{38}
$$

$$
\begin{aligned}
P_i^{+,p} &= m_i \sum_{j\neq i} \max\left\{0, \alpha_{ij}^{(\rho,E)} \left(\rho_i^L f_{ij}^{(\rho E)} + (\rho E)_i^L f_{ij}^{(\rho)} - (\rho\mathbf{v})_i^L \cdot \mathbf{f}_{ij}^{(\rho\mathbf{v})}\right)\right\} \\
&\quad + \sum_{j\neq i}\sum_{k\neq i} \max\left\{0, \alpha_{ij}^{(\rho,E)} f_{ij}^{(\rho)} \alpha_{ik}^{(\rho,E)} f_{ik}^{(\rho E)}\right\}, \qquad (39)
\end{aligned}
$$

$$
\begin{aligned}
P_i^{-,p} &= m_i \sum_{j\neq i} \min\left\{0, \alpha_{ij}^{(\rho,E)} \left(\rho_i^L f_{ij}^{(\rho E)} + (\rho E)_i^L f_{ij}^{(\rho)} - (\rho\mathbf{v})_i^L \cdot \mathbf{f}_{ij}^{(\rho\mathbf{v})}\right)\right\} \\
&\quad + \sum_{j\neq i}\sum_{k\neq i} \min\left\{0, \alpha_{ij}^{(\rho,E)} f_{ij}^{(\rho)} \alpha_{ik}^{(\rho,E)} f_{ik}^{(\rho E)}\right\} - \frac{1}{2}\left(\sum_{j\neq i} \alpha_{ij}^{(\rho,E)} |\mathbf{f}_{ij}^{(\rho\mathbf{v})}|\right)^2, \qquad (40)
\end{aligned}
$$

$$
Q_i^{+,p} := \tilde{\rho}_i^{\max} p_i^{\max} - \rho_i^L p_i^L, \qquad Q_i^{-,p} = \tilde{\rho}_i^{\min} p_i^{\min} - \rho_i^L p_i^L, \qquad (41)
$$

$$
R_i^{+,p} = \min\left\{1, \frac{m_i^2 Q_i^{+,p}}{(\gamma-1)P_i^{+,p}}\right\}, \qquad R_i^{-,p} = \min\left\{1, \frac{m_i^2 Q_i^{-,p}}{(\gamma-1)P_i^{-,p}}\right\}. \qquad (42)
$$

The formula for $P_i^{-,p}$ was derived using the triangle inequality to estimate the second quadratic term in (38). The proof of the estimates

$$
m_i^2 Q_i^{-,p} \leq R_i^{-,p} P_i^{-,p}, \qquad R_i^{+,p} P_i^{+,p} \leq m_i^2 Q_i^{+,p}
$$

in the inequality chain (36) exploits the fact that $(\alpha_{ij})^2 \leq \alpha_{ij}$ for $\alpha_{ij} \in [0,1]$.

To enforce the pressure bounds, the correction factors α_{ij} are defined thus:

$$
\alpha_{ij} = \min\left\{R_i^{+,p}, R_i^{-,p}, R_j^{+,p}, R_j^{-,p}\right\} \alpha_{ij}^{(\rho,E)}. \qquad (43)
$$

Due to the presence of quadratic terms in (38), this formula for α_{ij} does not distinguish between positive and negative antidiffusive fluxes.

Remark. The pressure limiter can be configured to enforce the constraints

$$
p_i^{\min} \leq p_i = \frac{\rho_i^L p_i^L + \frac{\gamma-1}{m_i^2} P_i^{(p)}}{\rho_i^L + \frac{1}{m_i}\sum_{j\neq i} \alpha_{ij} f_{ij}^{(\rho)}} \leq p_i^{\max}
$$

using a modification of the above algorithm to find $\alpha_{ij} \leq \alpha_{ij}^{(\rho,E)}$ such that

$$
\rho_i^L (p_i^{\min} - p_i^L) \leq \frac{\gamma-1}{m_i^2} P_i^{(p)} - \frac{1}{m_i}\sum_{j\neq i} \alpha_{ij} f_{ij}^{(\rho)} \leq \rho_i^L (p_i^{\max} - p_i^L).
$$

However, the use of sharp pressure bounds is not recommended in the context of synchronized FCT limiting for reasons explained in Section 4.

6. FEM-FCT for the Euler equations

In what follows, we apply the synchronized FCT limiter to a continuous piecewise-(bi)linear finite element discretization of the Euler equations

$$\frac{\partial U}{\partial t} + \nabla \cdot \mathbf{F}(U) = 0 \qquad \text{in } \Omega \subset \mathbb{R}^d, \quad d \in \{1,2,3\} \tag{44}$$

which represent a system of conservation laws for the gas dynamics variables $U = [\rho, \rho\mathbf{v}, \rho E]^T$. The triple of flux functions $\mathbf{F}$ is defined by

$$\mathbf{F}(U) = \{\rho\mathbf{v}, \rho\mathbf{v} \otimes \mathbf{v} + p\mathcal{I}, (\rho E + p)\mathbf{v}\} = \mathbf{A}(U)U, \tag{45}$$

where $\mathbf{A}(U) = \frac{\partial \mathbf{F}}{\partial U}$ is the Jacobian matrix and $\mathcal{I}$ is the identity tensor.

Let $\{\varphi_1, \ldots, \varphi_N\}$ be a set of finite element basis functions associated with the vertices of the computational mesh. Substituting the approximations

$$U_h = \sum_j U_j \varphi_j, \qquad \mathbf{F}_h = \sum_j \mathbf{F}_j \varphi_j \tag{46}$$

into the Galerkin weak form of (44), one obtains the semi-discrete problem

$$\sum_j \left(m_{ij} \frac{\mathrm{d}U_j}{\mathrm{d}t} \right) = -\sum_j \mathbf{c}_{ij} \cdot \mathbf{F}_j = -\sum_j (\mathbf{c}_{ij} \cdot \mathbf{A}_j) U_j, \tag{47}$$

where $\mathbf{F}_j = \mathbf{A}_j U_j$ and the coefficients are given by [11, 10]

$$m_{ij} = \int_\Omega \varphi_i \varphi_j \,\mathrm{d}\mathbf{x}, \qquad \mathbf{c}_{ij} = \int_\Omega \varphi_i \nabla \varphi_j \,\mathrm{d}\mathbf{x}. \tag{48}$$

Replacing m_{ij} by $\delta_{ij} \sum_j m_{ij}$, we define the lumped mass matrix

$$M_L = \mathrm{diag}\{m_i\}, \qquad m_i = \int_\Omega \varphi_i \,\mathrm{d}\mathbf{x} = \sum_j m_{ij}. \tag{49}$$

To construct a low-order scheme for the FEM-FCT algorithm, we add scalar artificial diffusion to the lumped-mass version of (47). This yields [10, 11]

$$m_i \frac{\mathrm{d}U_i}{\mathrm{d}t} = -\sum_j (\mathbf{c}_{ij} \cdot \mathbf{A}_j) U_j + \sum_{j \neq i} d_{ij} (U_j - U_i). \tag{50}$$

Since approximate Riemann solvers based on Roe's linearization [17] may fail to satisfy local maximum principles for some quantities of interest, we employ Rusanov-type dissipation proportional to the fastest wave speed. The corresponding artificial diffusion coefficients d_{ij} are defined by [10, 11]

$$d_{ij} = \max\{|\mathbf{c}_{ij} \cdot \mathbf{v}_j| + |\mathbf{c}_{ij}|c_j, |\mathbf{c}_{ji} \cdot \mathbf{v}_i| + |\mathbf{c}_{ji}|c_i\}, \tag{51}$$

where $c_i = \sqrt{\gamma p_i/\rho_i}$ denotes the local speed of sound.

The low-order predictor U^L can be calculated using an explicit or implicit time discretization of (50). In the below numerical study, we use an explicit strong stability preserving (SSP) Runge-Kutta scheme of third order [6, 7] for the 1D test problems and Crank-Nicolson time-stepping in 2D.

The antidiffusive fluxes for the predictor-corrector FCT scheme [8, 11] based on the above high- and low-order approximations are given by

$$F_{ij} = \Delta t \left[m_{ij} \left(\frac{\mathrm{d}U_i^L}{\mathrm{d}t} - \frac{\mathrm{d}U_j^L}{\mathrm{d}t} \right) + d_{ij}(U_i^L - U_j^L) \right], \tag{52}$$

where Δt is the time step and the nodal time derivatives are defined by (50).

For a detailed description of the FEM-FCT algorithm, we refer to [10, 11].

7. Numerical examples

In this section, we solve standard test problems for the Euler equations using the FEM-FCT scheme equipped with the new synchronized flux limiter.

7.1. Shock tube problem

Sod's shock tube problem [20] is a well-known benchmark for the 1D Euler equations. It models the flow of an inviscid gas in a tube initially separated by a membrane into two sections. Reflective boundary conditions are prescribed at the endpoints of the domain $\Omega = (0, 1)$. The initial condition for the nonlinear Riemann problem is given in terms of the primitive variables

$$\begin{bmatrix} \rho_L \\ v_L \\ p_L \end{bmatrix} = \begin{bmatrix} 1.0 \\ 0.0 \\ 1.0 \end{bmatrix}, \qquad \begin{bmatrix} \rho_R \\ v_R \\ p_R \end{bmatrix} = \begin{bmatrix} 0.125 \\ 0.0 \\ 0.15 \end{bmatrix}, \tag{53}$$

where the subscripts refer to the subdomains $\Omega_L = (0, 0.5)$ and $\Omega_R = (0.5, 1)$.

The numerical solutions presented in Fig. 1 were calculated on a uniform mesh of 100 linear finite elements using the time step $\Delta t = 10^{-3}$. The snapshots correspond to the final time $T = 0.231$. Figure 1(a) shows the low-order approximation to the density, velocity, and pressure. The solution profiles are strongly smeared and the local bounds are preserved for all primitive variables. The FEM-FCT solution displayed in Fig. 1(b) demonstrates the ability of the proposed limiter to capture shocks in a crisp and nonoscillatory manner. The smearing of the density profile at the contact discontinuity is inevitable due to the lack of self-steepening. Note that the resolution is much better than in the case of the underlying low-order scheme. The nodal values of the density and pressure are in the range defined by the inequality constraints for the synchronized flux limiter. In particular, positivity preservation is guaranteed without any additional postprocessing. The velocity profile exhibits small overshoots behind the rarefaction wave. These overshoots are not suppressed by the synchronized FCT limiter because the velocity is not included in the set of control variables to be constrained.

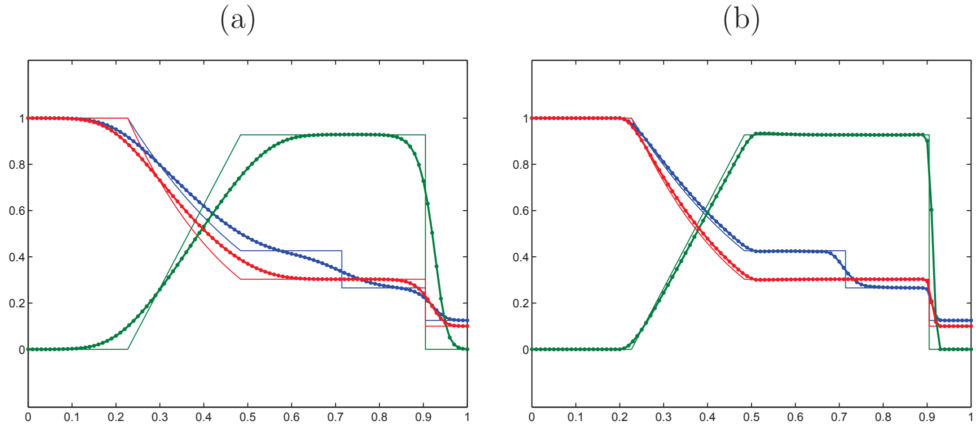


Figure 1: Sod's shock tube problem: (a) low-order scheme vs. (b) synchronized FCT algorithm, $h = 10^{-2}$, $\Delta t = 10^{-3}$, $T = 0.231$. Density: blue, velocity: green, pressure: red (for color graphics see the online version of this article).

7.2. *Blast wave problem*

The blast wave problem of Woodward and Colella [22] is a far more challenging test. The flow of a gamma-law gas, with $\gamma = 1.4$, takes place between reflecting walls, and the initial condition consists of the three constant states

$$\begin{bmatrix}\rho_L\\ v_L\\ p_L\end{bmatrix}=\begin{bmatrix}1.0\\ 0.0\\ 1000.0\end{bmatrix},\qquad \begin{bmatrix}\rho_M\\ v_M\\ p_M\end{bmatrix}=\begin{bmatrix}1.0\\ 0.0\\ 0.01\end{bmatrix},\qquad \begin{bmatrix}\rho_R\\ v_R\\ p_R\end{bmatrix}=\begin{bmatrix}1.0\\ 0.0\\ 100.0\end{bmatrix}\tag{54}$$

associated with $\Omega_L = (0, 0.1)$, $\Omega_M = (0.1, 0.9)$, and $\Omega_R = (0.9, 1)$.

The above initial conditions give rise to two strong blast waves which eventually collide. The flow evolution involves complex interactions of shocks, rarefactions, and contact discontinuities in a small region of space. These interactions are particularly difficult to capture using FCT algorithms which tend to clip peaks and distort steep fronts within the local bounds. The latter phenomenon is known as *terracing*. It can be alleviated by *prelimiting* the fluxes or improving the phase accuracy of the high-order scheme [24].

Figures 2 and 3 display the numerical approximations to the density and pressure at the final time $T = 0.038$. The mesh size and time step are given by $h = 10^{-3}$ and $\Delta t = 10^{-6}$, respectively. Again, the FEM-FCT solution is more accurate than its low-order counterpart and the bounds are preserved. The clearly visible terracing effect is caused by the poor phase accuracy of the standard Galerkin scheme and could be cured by adding high-order entropy viscosity (or another dissipative component) to the antidiffusive flux.

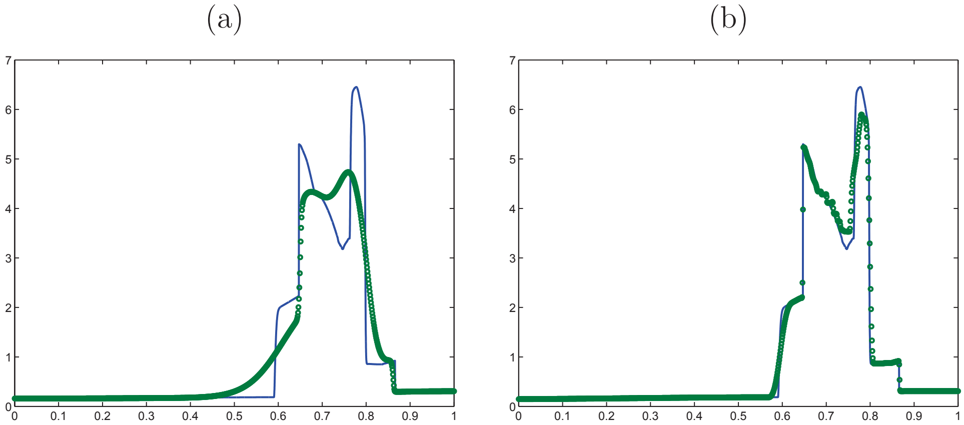


Figure 2: Blast wave problem: density distribution calculated using (a) low-order scheme and (b) synchronized FCT algorithm, $h = 10^{-3}$, $\Delta t = 10^{-6}$, $T = 0.038$.

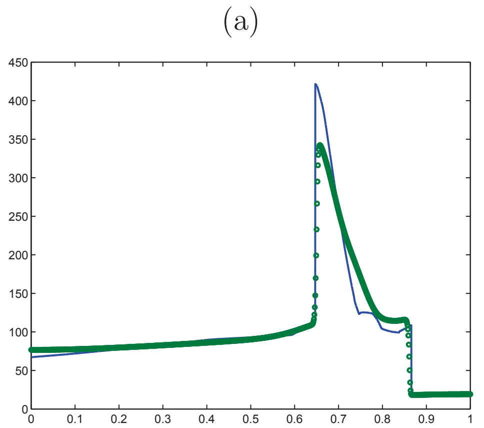


Figure 3: Blast wave problem: pressure distribution calculated using (a) low-order scheme and (b) synchronized FCT algorithm, $h = 10^{-3}$, $\Delta t = 10^{-6}$, $T = 0.038$.

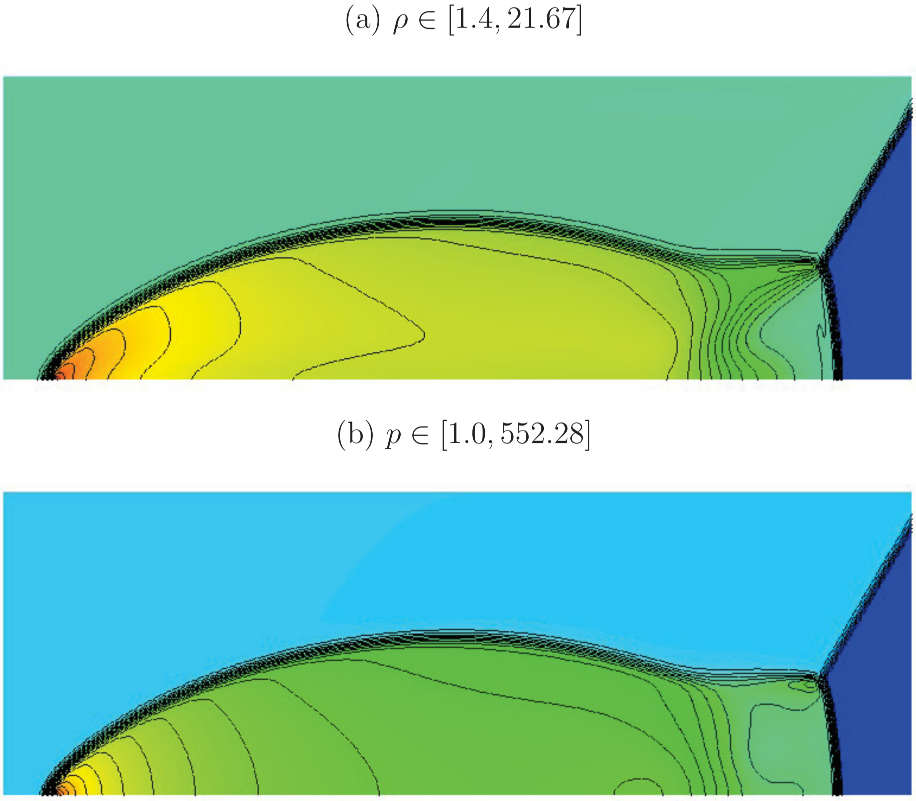
7.3. Double Mach reflection

In the last example, we consider the double Mach reflection benchmark [22] for the two-dimensional Euler equations. The computational domain for this test is the rectangle $\Omega = (0, 4) \times (0, 1)$. The flow pattern features a propagating Mach 10 shock in air ($\gamma = 1.4$) which initially makes a 60° angle with a reflecting wall. The following pre-shock and post-shock values of the flow variables are used to define the initial and boundary conditions

$$\begin{bmatrix} \rho_L \\ u_L \\ v_L \\ p_L \end{bmatrix} = \begin{bmatrix} 8.0 \\ 8.25\cos(30^\circ) \\ -8.25\sin(30^\circ) \\ 116.5 \end{bmatrix}, \qquad \begin{bmatrix} \rho_R \\ u_R \\ v_R \\ p_R \end{bmatrix} = \begin{bmatrix} 1.4 \\ 0.0 \\ 0.0 \\ 1.0 \end{bmatrix}. \tag{55}$$

Initially, the post-shock values (subscript L) are prescribed in the subdomain $\Omega_L = \{(x, y) \mid x < 1/6 + y/\sqrt{3}\}$ and the pre-shock values (subscript R) in $\Omega_R = \Omega\backslash\Omega_L$. The reflecting wall corresponds to $1/6 \le x \le 4$ and $y = 0$. No boundary conditions are required along the line $x = 4$. On the rest of the boundary, the post-shock conditions are assigned for $x < 1/6 + (1 + 20t)/\sqrt{3}$ and the pre-shock conditions elsewhere. The so-defined values along the top boundary describe the exact motion of the initial Mach 10 shock.

The density and pressure distributions produced by the low-order scheme and by the synchronized FCT algorithm at the final time $T = 0.2$ are displayed in Figs 4 and 5, respectively. Computations were performed on a uniform mesh of bilinear elements ($h = 1/128$). The low-order solution/predictor was advanced in time using the Crank-Nicolson time-stepping and the time step $\Delta t = 10^{-4}$. The results of this 2D simulation confirm the ability of the synchronized FCT limiter to remove significant amounts of numerical dissipation without generating negative pressures and/or nonphysical oscillations.



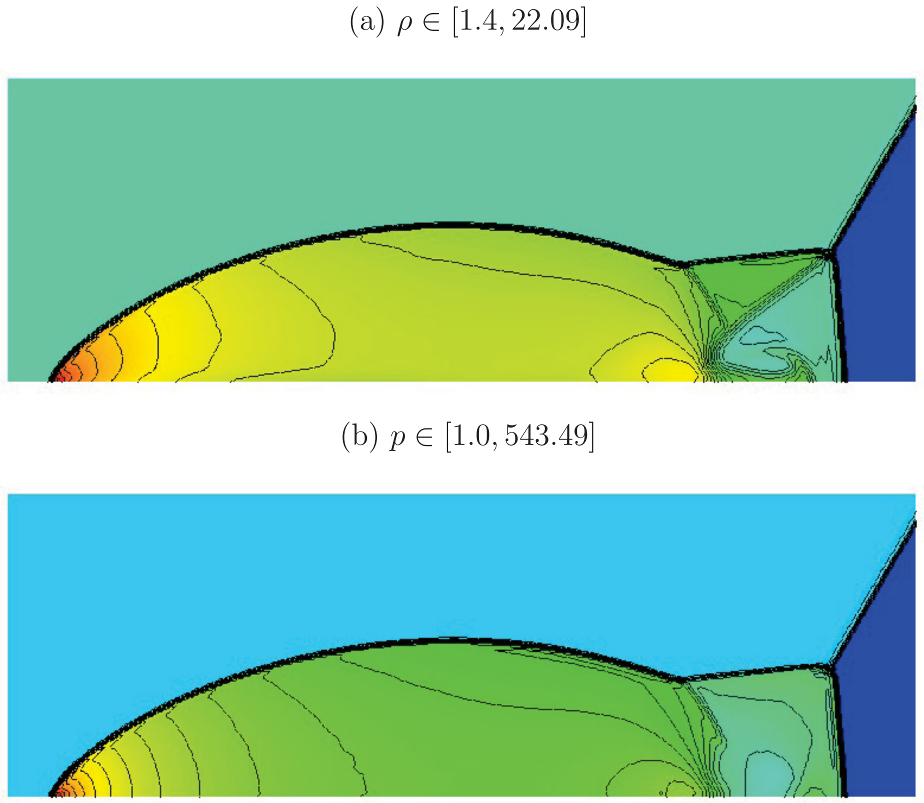
(a) $\rho \in [1.4, 21.67]$

(b) $p \in [1.0, 552.28]$

Figure 4: Double Mach reflection problem: (a) density and (b) pressure distribution at $T = 0.2$. Discretization: low-order scheme, 65,536 Q_1 elements, $\Delta t = 10^{-4}$.

8. Summary

The presented approach to synchronized flux limiting for the density, energy and pressure guarantees positivity preservation at a fraction of the cost associated with optimization-based alternatives. At the same time, the revised



(a) $\rho \in [1.4, 22.09]$

(b) $p \in [1.0, 543.49]$

Figure 5: Double Mach reflection problem: (a) density and (b) pressure distribution at $T = 0.2$. Discretization: synchronized FCT algorithm, 65,536 Q_1 elements, $\Delta t = 10^{-4}$.

definition of local bounds for the energy and pressure makes it failsafe and less diffusive than the linearized FCT limiter presented in [11]. The proposed limiting procedure is readily applicable to flux-based remapping algorithms in the context of Arbitrary Lagrangian Eulerian (ALE) methods. We also envisage that it can be easily adapted to discontinuous Galerkin discretizations of the Euler equations equipped with vertex-based slope limiters [9].

Acknowledgments

This research was supported by the German Research Association (DFG) under grant KU 1530/13-1.

References

[1] P. Bochev, D. Ridzal, K. Peterson, Optimization-based remap and transport: A divide and conquer strategy for feature-preserving discretizations. *J. Comput. Phys.* **257** (2014) 1113–1139.

[2] P. Bochev, D. Ridzal, G. Scovazzi, M. Shashkov, Constrained-optimization based data transfer: A new perspective on flux correction. In: D. Kuzmin, R. Löhner, S. Turek (eds), *Flux-Corrected Transport: Principles, Algorithms, and Applications*. Springer, 2nd edition, 2012, pp. 345–398.

[3] J.P. Boris and D.L. Book, Flux-Corrected Transport: I. SHASTA, a fluid transport algorithm that works. *J. Comput. Phys.* **11** (1973) 38–69.

[4] D.L. Book, J.P. Boris, K. Hain, Flux-Corrected Transport: II. Generalizations of the Method. *J. Comput. Phys.* **18** (1975) 248–283.

[5] J.P. Boris and D.L. Book, Flux-Corrected Transport: III. Minimal-error FCT algorithms. *J. Comput. Phys.* **20** (1976) 397–431.

[6] S. Gottlieb and C. W. Shu, Total Variation Diminishing Runge-Kutta schemes. *Math. Comp.* **67** (1998) 73–85.

[7] S. Gottlieb, C.-W. Shu, E. Tadmor, Strong stability-preserving high-order time discretization methods. *SIAM Review* **43** (2001) 89–112.

[8] D. Kuzmin, Explicit and implicit FEM-FCT algorithms with flux linearization. *J. Comput. Phys.* **228** (2009) 2517-2534.

[9] D. Kuzmin, Hierarchical slope limiting in explicit and implicit discontinuous Galerkin methods. *J. Comput. Phys.* **257** (2014) 1140–1162.

[10] D. Kuzmin, M. Möller, M. Gurris, Algebraic flux correction II. Compressible flow problems. In: D. Kuzmin, R. Löhner, S. Turek (eds), *Flux-Corrected Transport: Principles, Algorithms, and Applications.* Springer, 2nd edition, 2012, pp. 193–238.

[11] D. Kuzmin, M. Möller, J.N. Shadid, M. Shashkov: Failsafe flux limiting and constrained data projections for equations of gas dynamics. *J. Comput. Phys.* **229** (2010) 8766–8779.

[12] R. Liska, M. Shashkov, P. Váchal, B. Wendroff, Optimization-based synchronized flux-corrected conservative interpolation (remapping) of mass and momentum for Arbitrary Lagrangian-Eulerian methods. *J. Comput. Phys.* **229** (2010) 1467–1497.

[13] R. Liska, M. Shashkov, P. Váchal, B. Wendroff, Synchronized flux corrected remapping for ALE methods. *Computers & Fluids* **46** (2011) 312317

[14] R. Löhner, *Applied CFD Techniques: An Introduction Based on Finite Element Methods*. Second Edition, John Wiley & Sons, 2008.

[15] R. Löhner, K. Morgan, J. Peraire, M. Vahdati, Finite element flux-corrected transport (FEM-FCT) for the Euler and Navier-Stokes equations. Int. J. Numer. Meth. Fluids, 7 (1987) 1093–1109.

[16] G. Luttwak and J. Falcovitz, Slope limiting for vectors: A novel vector limiting algorithm. *Int. J. Numer. Methods Fluids* **65** (2011) 1365–1375.

[17] P. L. Roe, Approximate Riemann solvers, parameter vectors and difference schemes. *J. Comput. Phys.* **43** (1981) 357–372.

[18] C. Schär and P. K. Smolarkiewicz, A synchronous and iterative flux-correction formalism for coupled transport equations. *J. Comput. Phys.* **128** (1996) 101–120.

[19] P. K. Smolarkiewicz and G. A. Grell, A class of monotone interpolation schemes. *J. Comput. Phys.* **101** (1992) 431–440.

[20] G. Sod, A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws. *J. Comput. Phys.* **27** (1978) 1–31.

[21] P. Váchal and R. Liska, Sequential flux-corrected remapping for ALE methods. In: A. Bermudez de Castro, D. Gomez, P. Quintela, and P. Salgado (eds.) *Numerical Mathematics and Advanced Applications* (ENUMATH 2005). Springer, 2006, pp. 671-679.

[22] P.R. Woodward and P. Colella, The numerical simulation of two-dimensional fluid flow with strong shocks. *J. Comput. Phys.* **54** (1984) 115–173.

[23] S.T. Zalesak, Fully multidimensional flux-corrected transport algorithms for fluids. *J. Comput. Phys.* **31** (1979) 335–362.

[24] S. T. Zalesak, The design of Flux-Corrected Transport (FCT) algorithms for structured grids. In: D. Kuzmin, R. Löhner, S. Turek (eds), *Flux-Corrected Transport: Principles, Algorithms, and Applications.* Springer, 2nd edition, 2012, pp. 23-66.

[25] X. Zeng and G. Scovazzi, A frame-invariant vector limiter for flux corrected nodal remap in arbitrary Lagrangian-Eulerian flow computations. *J. Comput. Phys.* **270** (2014) 753–783.