On the design of general-purpose flux limiters for finite element schemes. I. Scalar convection

D. Kuzmin*

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

Abstract

The *algebraic flux correction* (AFC) paradigm is extended to finite element discretizations with a consistent mass matrix. It is shown how to render an implicit Galerkin scheme positivity-preserving and remove excessive artificial diffusion in regions where the solution is sufficiently smooth. To this end, the original discrete operators are modified in a mass-conserving fashion so as to enforce the algebraic constraints to be satisfied by the numerical solution. A node-oriented limiting strategy is employed to control the raw antidiffusive fluxes which consist of a convective part and a contribution of the consistent mass matrix. The former offsets the artificial diffusion due to 'upwinding' of the spatial differential operator and lends itself to an upwind-biased flux limiting. The latter eliminates the error induced by mass lumping and calls for the use of a symmetric flux limiter. The concept of a *target flux* and a new definition of upper/lower bounds make it possible to combine the advantages of algebraic FCT and TVD schemes introduced previously by the author and his coworkers. Unlike other high-resolution schemes for unstructured meshes, the new algorithm reduces to a consistent (high-order) Galerkin scheme in smooth regions and is designed to provide an optimal treatment of both stationary and time-dependent problems. Its performance is illustrated by application to the linear advection equation for a number of 1D and 2D configurations.

Key Words: convection-dominated problems; high-resolution schemes; flux correction; finite elements; consistent mass matrix

1 Introduction

For decades, the development of numerical methods for convection-dominated flows has been one of the primary research directions in Computational Fluid Dynamics. A variety of stabilization techniques and high-resolution schemes based on flux/slope limiting were proposed to combat the onset of nonphysical oscillations but no universally effective remedy has been found to date. A typical disadvantage of currently available discretization techniques is the lack of generality. The foundations of modern high-resolution schemes were developed in the finite difference framework using essentially one-dimensional concepts and, typically, geometric design criteria. As a result, many popular algorithms are limited to Cartesian meshes and/or explicit time-stepping schemes.

*Correspondence to: kuzmin@math.uni-dortmund.de

The design of genuinely multidimensional high-resolution schemes for finite element discretizations on unstructured meshes has proved to be a particularly challenging task. In the late 1980s and early 1990s, flux-corrected transport (FCT) and total variation diminishing (TVD) algorithms were carried over to explicit Galerkin schemes based on linear/bilinear finite elements [2],[27],[28],[29],[33],[34]. In spite of some inherent limitations to be mentioned below, these straightforward extensions produced very promising results but were met with little enthusiasm by FEM practitioners. The current trend in the unstructured grid community is to use finite volume upwinding [12],[36], residual distribution / fluctuation splitting [6],[9] or discontinuous Galerkin methods [7],[8]. Stabilization without shock capturing (streamline diffusion, edge stabilization / interior penalty) has also been widely used in the FEM context, especially for incompressible flows.

In a series of recent publications, the author and his collaborators introduced an algebraic approach to the design of high-resolution schemes which has made it possible to incorporate flux limiters of FCT and TVD type into implicit finite element schemes [20],[21],[22],[23]. The underlying *algebraic flux correction* paradigm can be summarized as follows: take the matrix resulting from an arbitrary discretization of the convective term and modify it so as to enforce the M-matrix property making sure that

- all modifications are conservative, i.e., there is no loss or gain of 'mass';
- the original high-order discretization is recovered in regions of smoothness.

To this end, a positivity-preserving low-order scheme is constructed by resorting to mass lumping followed by a conservative elimination of negative off-diagonal coefficients. Then the accuracy is enhanced by adding a limited amount of compensating antidiffusion, whereby the raw antidiffusive fluxes are limited node-by-node so as to satisfy the imposed algebraic constraints. Remarkably, all the necessary information is provided by the matrix coefficients, so that flux limiting can be performed in a "black-box" fashion.

Flux correction of FCT type is applicable to Galerkin schemes with a consistent mass matrix [21],[27] and yields highly accurate solutions to time-dependent problems. However, the amount of admissible antidiffusion is inversely proportional to the time step, which compromises the advantages of unconditionally stable implicit schemes. Moreover, severe convergence problems are observed in the steady-state limit. On the other hand, flux correction of TVD type is independent of the time step and optimal for the treatment of stationary problems. Standard TVD limiters can be integrated into unstructured grid codes and applied edge-by-edge [2],[28] or node-by-node [22],[23], so as to control the slope ratio for a local 1D stencil or the net antidiffusive flux, respectively. In either case, the resulting scheme proves local extremum diminishing (LED) but mass lumping is mandatory and there is an alarming ambiguity in the choice of the limiter function.

In fact, standard limiters like *minmod* or *superbee* are designed to constrain the antidiffusive flux for the 1D convection equation discretized by finite differences on a uniform mesh. They are defined as functions lying in the second-order TVD region, which corresponds to a nonlinear combination of the Lax-Wendroff and Beam-Warming methods. At the same time, the flux limiter for a finite element scheme should be designed to recover a consistent-mass (Taylor-)Galerkin discretization, as in the case of multidimensional FEM-FCT schemes. The use of one-dimensional TVD limiters is not to be recommended because a certain amount of artificial (anti-)diffusion is added even if there is no need for limiting. Hence, the resulting approximation does not reduce to the original Galerkin scheme and cannot be guaranteed to be second-order accurate for smooth data.

In the present paper, we recapitulate the principles of algebraic flux correction and focus on the choice of upper/lower bounds for the antidiffusive flux which corresponds to a conventional Galerkin discretization. Building on our experience with algebraic FCT and TVD schemes, we design a symmetric flux limiter for the contribution of the consistent mass matrix and blend it with an upwind-biased flux limiter for the discretized convective term. As a result, we obtain a new high-resolution finite element scheme which yields time-accurate solutions to transient problems and, moreover, does not suffer from a loss of accuracy if large time steps are employed when the solution approaches a highly convective steady state. Numerical examples are presented for 1D and 2D benchmark problems.

2 Conservative flux decomposition

Let us start with the definition of diffusive and antidiffusive fluxes for finite element discretizations. The reader who is already familiar with algebraic flux correction [23] may want to skip this section. Consider the time-dependent continuity equation

$$\frac{\partial u}{\partial t} + \nabla \cdot (\mathbf{v}u) = 0 \tag{1}$$

discretized in space by a high-order finite element (Galerkin or Taylor-Galerkin) method which yields a DAE system for the vector of time-dependent nodal values

$$M_C \frac{\mathrm{d}u}{\mathrm{d}t} = Ku, \tag{2}$$

where $M_C = \{m_{ij}\}$ denotes the consistent mass matrix and $K = \{k_{ij}\}$ is the discrete transport operator resulting from the discretization of the convective term.

It is well known that even stabilized high-order methods may produce nonphysical undershoots and overshoots in the vicinity of steep gradients. On the other hand, upwind-like approximations are nonoscillatory but overly diffusive. This is why modern high-resolution schemes use flux or slope limiters to switch between such linear approximations in an adaptive way. Roughly speaking, the high-order method is used in regions where the solution is sufficiently smooth and the low-order method elsewhere. In order to blend these methods automatically without resorting to artificial parameters typical of hybrid upwind discretizations [36], one needs to define certain mathematical criteria which guarantee that the numerical solution remains nonoscillatory. These criteria can be expressed as algebraic constraints to be imposed on a linear high-order discretization like (2).

A very handy criterion, which represents a generalization of Harten's TVD theorem, was introduced by Jameson [15],[16] who proved that a semi-discrete scheme of the form

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \sum_{j \neq i} c_{ij}(u_j - u_i), \qquad c_{ij} \geq 0, \quad \forall j \neq i \tag{3}$$

is *local extremum diminishing* (LED). After the discretization in time, such schemes remain positivity-preserving provided that each solution update $u^n \quad \to \quad u^{n+1}$ or the converged steady-state solution $u^{n+1} = u^n$ satisfies an equivalent algebraic system [23]

$$Au^{n+1} = Bu^n, \tag{4}$$

where $A = \{a_{ij}\}$ is an *M-matrix* and $B = \{b_{ij}\}$ has no negative entries so that

$$u^n \geq 0 \qquad \Rightarrow \qquad u^{n+1} = A^{-1}Bu^n \geq 0. \tag{5}$$

This extra requirement yields a readily computable upper bound for admissible time steps.

In the linear case, the above algebraic criteria can be readily enforced by means of 'discrete upwinding' as proposed in [19],[20]. For a finite element scheme of the form (2), the required matrix manipulations are as follows

- replace the consistent mass matrix M_C by its lumped counterpart $M_L = \mathrm{diag}\{m_i\}$,
- render the operator K local extremum diminishing by adding an artificial diffusion operator $D = \{d_{ij}\}$ so as to eliminate all negative off-diagonal coefficients.

At the end of the day, this gives a linear LED counterpart of (2) which reads

$$M_L \frac{\mathrm{d}u}{\mathrm{d}t} = Lu, \qquad \text{where} \quad L = K + D. \tag{6}$$

The artificial diffusion operator D is designed to be a symmetric matrix with zero row and column sums. Therefore, the term Du can be decomposed into a sum of skew-symmetric internodal fluxes associated with the edges of the sparsity graph [23]

$$(Du)_i := -\sum_{j \neq i} f_{ij}^d, \qquad f_{ij}^d = d_{ij}(u_i - u_j) = -f_{ji}^d. \tag{7}$$

A natural choice of the artificial diffusion coefficient d_{ij} for the edge $\overrightarrow{ij}$ is [20]

$$d_{ij} = \max\{-k_{ij}, 0, -k_{ji}\} = d_{ji}. \tag{8}$$

Alternatively, one can apply discrete upwinding to the skew-symmetric part $\frac{1}{2}(K - K^T)$ of the high-order transport operator K, which corresponds to

$$d_{ij} = \frac{|k_{ij} - k_{ji}|}{2} - \frac{k_{ij} + k_{ji}}{2} = d_{ji}. \tag{9}$$

In either case, the off-diagonal coefficients of the low-order operator $l_{ij} := k_{ij} + d_{ij} \geq 0$ are nonnegative so that the LED criterion is satisfied. Without loss of generality, the edge $\overrightarrow{ij}$ is oriented so that $l_{ij} \leq l_{ji}$, which implies that node i is located 'upwind' and corresponds to the row number of the eliminated negative coefficient [22],[23].

The *raw antidiffusive fluxes* which offset the error induced by mass lumping and discrete upwinding so that the original Galerkin scheme (2) is recovered are given by

$$f_{ij} = \left[m_{ij}\frac{\mathrm{d}}{\mathrm{d}t} + d_{ij}\right](u_i - u_j) = f_{ij}^m + f_{ij}^d, \qquad f_{ij}^m = m_{ij}(\dot{u}_i - \dot{u}_j). \tag{10}$$

Note that the above expression contains a time derivative which still needs to be discretized (cf. [20],[21]). In order to prevent the formation of nonphysical local extrema, the raw antidiffusive fluxes are multiplied by suitable correction factors (see below)

$$f_{ij}^* := \alpha_{ij} f_{ij}, \qquad \text{where} \qquad 0 \leq \alpha_{ij} \leq 1. \tag{11}$$

Inserting these fluxes into the right-hand side of (6) one obtains a nonlinear combination of the low-order scheme ($\alpha_{ij} \equiv 0$) and the original high-order one ($\alpha_{ij} \equiv 1$). The task of the flux limiter is to determine an optimal value of each correction factor α_{ij} so as to remove as much artificial diffusion as possible without generating spurious oscillations.

3 Flux correction in one dimension

In order to introduce some useful concepts in a rather simple setting, let us start with flux correction for the one-dimensional linear advection equation

$$\frac{\partial u}{\partial t} + v \frac{\partial u}{\partial x} = 0, \qquad v > 0 \tag{12}$$

discretized in space on a uniform mesh of linear finite elements. It is well known that the lumped-mass Galerkin scheme is equivalent to the central difference method. In this case, the elimination of negative off-diagonal coefficients leads to the classical upwind difference scheme [23]. The corresponding artificial diffusion coefficient (8) equals $d_{ij} = v/2$.

In the one-dimensional case, the overly diffusive upwind approximation can be transformed into a second-order scheme by adding antidiffusive fluxes of the form

$$f_{ij} = \phi_i d_{ij}(u_i - u_j), \tag{13}$$

where $j = i + 1$ and ϕ_i is a function of the slope ratio evaluated at node i, for instance

$$\phi_i = \xi + (1 - \xi) r_i, \qquad r_i = \frac{u_{i-1} - u_i}{u_i - u_{i+1}}. \tag{14}$$

For any value of $0 \leq \xi \leq 1$, the scaled antidiffusive correction (13) renders the upwind discretization second-order accurate. The central difference scheme is recovered for $\xi = 1$, whereas $\xi = 0$ corresponds to a backward difference approximation of second order. In general, the multiplier ϕ_i is supposed to adjust the antidiffusion coefficient d_{ij} so that a certain high-order discretization ('target scheme') is recovered if no flux limiting is performed. This interpretation of (13) leads to the following definition, cf. [39]

Definition. A *target flux* represents the amount of raw antidiffusion that converts a low-order approximation of the convective term into the desired high-order one.

In the course of flux correction, each target flux f_{ij} is replaced by its limited counterpart $f_{ij}^* = \alpha_{ij} f_{ij}$ to make sure that the resulting semi-discrete scheme

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} + \frac{vu_{i+1/2} - vu_{i-1/2}}{\Delta x} = 0, \qquad vu_{i+1/2} := vu_i + f_{ij}^* \tag{15}$$

remains local extremum / total variation diminishing. For our purposes, it is worthwhile to represent the limited antidiffusive flux for a classical TVD scheme as follows

$$f_{ij}^* := \max\{0, \min\{2, \phi_i, 2r_i\}\} d_{ij}(u_i - u_j). \tag{16}$$

Note that the effective correction factor $\alpha_{ij} = f_{ij}^*/f_{ij}$ is bounded by 0 and 1, whereas the limited coefficient ϕ_i may vary between 0 (backward difference) and 2 (forward difference). By construction, the limited antidiffusive flux admits the representation $f_{ij}^* = c_{ik}(u_k - u_i)$, where $k = i - 1$ and $c_{ik} \geq 0$. The LED criterion and Harten's TVD conditions for the downwind node j are also satisfied, since the antidiffusive flux f_{ji} is neutralized by the diffusive contribution $l_{ji}(u_i - u_j)$ of the low-order operator (see the next section).

The above interpretation of TVD schemes, which can be traced back to [39], reveals that the numerous 'limiter functions' proposed in the literature differ merely in the definition of the underlying target flux. The most popular representatives are

minmod	$\phi_i = \min\{1, r_i\}$,
Van Leer	$\phi_i = 2r_i/(1 + r_i)$,
MC	$\phi_i = (1 + r_i)/2$,
Koren	$\phi_i = (2 + r_i)/3$,
superbee	$\phi_i = \max\{1, r_i\}$.

The best accuracy attainable within Sweby's second-order TVD region is provided by Koren's limiter [18] which has been repeatedly reinvented under different names [1],[35]. Due to the fact that the leading terms in the modified equation cancel out, the resulting scheme is third-order accurate for sufficiently smooth data.

Flux limiters of TVD type based on the above definitions of ϕ_i can also be interpreted as limited average operators [15],[16] and used to enforce the LED property in the finite element framework [22],[28],[30]. However, the associated target fluxes (13) are certain to ensure second-order accuracy only for a constant velocity v on a uniform mesh, whereas the real target fluxes for a finite element scheme are uniquely defined by (10). Therefore, straightforward generalizations of TVD schemes to multidimensions (including those proposed by the author) are likely to pollute the solution in smooth regions, and second-order accuracy can no longer be guaranteed. For this reason, the use of standard TVD limiters is not to be recommended for finite element discretizations on unstructured meshes.

On the other hand, it is instructive to examine the mechanism which guarantees that the limited antidiffusive flux does not violate the LED criterion. As a matter of fact, the constraints imposed in (16) are not optimal, since the left boundary of the TVD region depends on the Courant number [14],[39]. However, ignoring this dependence in favor of the simple formula $\alpha_{ij}\phi_i = \max\{0, \min\{2, \phi_i, 2r_i\}\}$ makes such limiters remarkably efficient and directly applicable to stationary problems. That is, instead of computing a sharp bound for a given time step (which is particularly expensive in multidimensions) one can use some reasonable fixed bounds and adjust the time step if this is necessary to satisfy a CFL-like condition. In what follows, we will use a similar approach to design the upper and lower bounds for our algebraic flux correction schemes.

4 Flux correction in multidimensions

The discussion of one-dimensional TVD discretizations in the previous section gives an insight into the design philosophy of modern high-resolution schemes which carries over to multidimensions. Antidiffusive fluxes which violate the LED criterion (3) and, therefore, need to be limited are of the form $f_{ij} = p_{ij}(u_j - u_i)$, where $p_{ij} \leq 0$. On the other hand, edge contributions with nonnegative coefficients resemble diffusive fluxes and are harmless. Therefore, some antidiffusion is admissible as long as there exists a solution-dependent coefficient $q_{ik} \geq 0$ such that $f_{ij} = q_{ik}(u_k - u_i)$. In other words, the antidiffusive flux from node j into node i should be interpreted as a diffusive flux from another node. In order to enforce this sufficient condition, we resort to a node-based limiting strategy which was largely inspired by Zalesak's limiter [38] but is even more general. As we are about to see, it can be used to construct a variety of algorithms which differ in the definition of upper/lower bounds as well as in the type of flux limiting (upwind or symmetric).

In the multidimensional case, the net antidiffusive correction to each node may consist of both positive and negative edge contributions. Assuming the worst-case scenario, we shall limit them separately according to the following generic algorithm

1. Compute the sums of positive and negative antidiffusive fluxes represented as edge contributions $f_{ij} = p_{ij}(u_j - u_i)$ with negative coefficients $p_{ij} \leq 0$

$$P_i^+ = \sum_{j \neq i} p_{ij} \min\{0, u_j - u_i\}, \qquad P_i^- = \sum_{j \neq i} p_{ij} \max\{0, u_j - u_i\}. \tag{17}$$

2. Define the upper and lower bounds to be imposed in the course of flux correction as a sum of edge contributions with nonnegative coefficients $q_{ij} \geq 0$

$$Q_i^+ = \sum_{j \neq i} q_{ij} \max\{0, u_j - u_i\}, \qquad Q_i^- = \sum_{j \neq i} q_{ij} \min\{0, u_j - u_i\}. \tag{18}$$

3. Evaluate the *nodal correction factors* for positive/negative antidiffusive fluxes

$$R_i^+ = \min\{1, Q_i^+/P_i^+\}, \qquad R_i^- = \min\{1, Q_i^-/P_i^-\}. \tag{19}$$

4. Multiply the target flux f_{ij} by a combination of $R_i^{\pm}$ and $R_j^{\mp}$ such that

$$f_{ij}^* = \alpha_{ij} f_{ij}, \qquad \alpha_{ij} = \left\{ \begin{array}{ll} \alpha(R_i^+, R_j^-), & \text{if } f_{ij} > 0, \\ \alpha(R_i^-, R_j^+), & \text{otherwise.} \end{array} \right. \tag{20}$$

The last part calls for further explanation. Recall that the edges of the sparsity graph are oriented so that $0 \leq l_{ij} \leq l_{ji} = k_{ji} + d_{ij}$. Furthermore, the nodal correction factors (19) are designed so as to enforce the LED constraint $|R_i^{\pm} P_i^{\pm}| \leq |Q_i^{\pm}|$ for the upwind node i. After flux limiting, the contribution of the edge $\vec{ij}$ to the downwind node j is given by

$$l_{ji}(u_i - u_j) - f_{ij}^* = (l_{ji} + \alpha_{ij} p_{ij})(u_i - u_j), \tag{21}$$

and also proves local extremum diminishing provided that the (negative) antidiffusion coefficient p_{ij} and the correction factor α_{ij} satisfy the inequality $l_{ji} + \alpha_{ij} p_{ij} \geq 0$.

In light of the above, algebraic flux correction can be performed in two different ways:

- Upwind-biased flux correction: 'prelimit' the target flux $f_{ij} = p_{ij}(u_j - u_i)$ to satisfy the positivity constraint for node j **before** computing the sums $P_i^\pm$ in (17)
$$f'_{ij} = \min\{-p_{ij}, l_{ji}\}(u_i - u_j) \tag{22}$$
and use the correction factors $\alpha_{ij} = R_i^\pm$ to enforce the LED property for node i.

- Symmetric flux correction: limit f_{ij} using the minimum of nodal correction factors for both nodes, i.e., $\alpha_{ij} = \min\{R_i^\pm, R_j^\mp\}$ so that the following estimates hold
$$Q_i^- \leq R_i^- P_i^- \leq \sum\nolimits_{j\neq i} \alpha_{ij} f_{ij} \leq R_i^+ P_i^+ \leq Q_i^+. \tag{23}$$

In the FEM context, the optimal choice of the limiting strategy depends on the magnitude of the antidiffusion coefficient $p_{ij} = f_{ij}/(u_j - u_i)$ for the target flux f_{ij} as defined by (10).

The above methodology is not to be confused with Zalesak's multidimensional FCT algorithm [38],[40]. In fact, standard (two-step) FCT methods do not fit into this framework, since the computation of a provisional low-order solution makes them inherently explicit. However, we intentionally use the same notation to emphasize the common features such as the **node-oriented** approach to flux correction which makes it possible to control the interplay of multiple antidiffusive fluxes acting in concert. The main advantage of the algorithm (17)–(20) as compared to the classical Zalesak limiter is a remarkable flexibility in the choice of upper/lower bounds $Q_i^\pm$ which makes it possible to bridge the gap between algebraic flux correction schemes of FCT and TVD type.

4.1 Treatment of convective antidiffusion

For the time being, let us assume that the problem at hand is stationary and neglect the contribution of the consistent mass matrix which will be considered in the next subsection. The prelimited target flux (22) for a lumped-mass Galerkin discretization is given by
$$f'_{ij} = \min\{d_{ij}, l_{ji}\}(u_i - u_j), \tag{24}$$
where d_{ij} is the artificial diffusion coefficient for discrete upwinding. It is worth mentioning that there is actually no need for prelimiting as long as $l_{ji} - \alpha_{ij} d_{ij} = k_{ji} + (1-\alpha_{ij})d_{ij} \geq 0$. Therefore, the above target flux reduces to f_{ij}^d as defined in (7), unless both off-diagonal coefficients of the high-order operator K were negative (a rather unusual situation).

In this particular case, the upwind-biased limiting strategy is preferable. The total amount of raw antidiffusion received by node i from its downwind neighbors is given by
$$P_i^\pm = \sum_{j\in\mathcal{J}_i} \begin{matrix}\max\\ \min\end{matrix}\,\{0, f'_{ij}\}, \quad \text{where} \quad \mathcal{J}_i = \{j \neq i \,|\, 0 = l_{ij} < l_{ji}\}. \tag{25}$$

The nonnegative off-diagonal coefficients of the low-order operator L can be used to define the upper/lower bounds as in the case of algebraic TVD schemes [22],[23]
$$Q_i^\pm = \sum_{j\neq i} l_{ij} \begin{matrix}\max\\ \min\end{matrix}\,(u_j - u_i), \qquad l_{ij} \geq 0, \quad \forall j \neq i. \tag{26}$$

Flux limiting is performed using the nodal correction factor for the upwind node:

$$f_{ij}^* = \begin{cases} R_i^+ f'_{ij}, & \text{if } f'_{ij} > 0, \\ R_i^- f'_{ij}, & \text{otherwise,} \end{cases} \qquad f_{ji}^* := -f_{ij}^*. \tag{27}$$

Remarkably, all the necessary information is extracted from the original matrix K and there is no need to know the coordinates of nodes or any other geometric details.

In one dimension, the resulting algorithm reduces to the flux-limited central difference scheme which corresponds to $f_{ij}^* = \max\{0, \min\{1, 2r_i\}\} d_{ij}(u_i - u_j)$ in accordance with (16). A family of local extremum diminishing schemes based on standard TVD limiters can be derived using target fluxes of the form (13), where the artificial diffusion coefficient d_{ij} is given by (9) and ϕ_i is a function of the smoothness indicator r_i. The latter is redefined as the ratio of edge contributions with positive and negative coefficients [22],[23]

$$r_i^{\pm} = \frac{\sum_{j\neq i} \max\{0, k_{ij} - k_{ji}\} \begin{smallmatrix}\max\\ \min\end{smallmatrix}\{0, u_j - u_i\}}{\sum_{j\neq i} \min\{0, k_{ij} - k_{ji}\} \begin{smallmatrix}\min\\ \max\end{smallmatrix}\{0, u_j - u_i\}}$$

which reduces to the usual slope ratio in the 1D case. Even though this *ad hoc* approach to the design of target fluxes works well in practice, it is no longer possible to guarantee that a high-order Galerkin approximation is recovered in smooth regions. For the flux-limited scheme to be consistent with the original one (2), it is necessary to use the target flux f_{ij} given by (10), as in the case of multidimensional FEM-FCT algorithms [21],[27]. Hence, it is a waste of time to design an optimal formula for ϕ_i as a function of the smoothness sensor r_i. In the finite element context, the accuracy of algebraic flux correction schemes should depend solely on the *resolving power* [40] of the underlying high-order method and can be enhanced by a suitable choice of basis functions in the variational formulation.

4.2 Treatment of mass antidiffusion

For genuinely time-dependent problems, mass lumping degrades the phase accuracy of finite element schemes and deprives them of a significant advantage in comparison to finite difference and finite volume methods. Berzins [3],[4] recognized the need for including the consistent mass matrix in a positivity-preserving fashion and presented some ideas as to how this can be accomplished. As of this writing, no truly multidimensional extension of his methodology seems to be available, so we need to look for another way to embed the consistent mass matrix into algebraic flux correction schemes.

The contribution of the mass matrix to target fluxes of the form (10) may be large enough to render the upwind-biased limiting strategy impractical. Furthermore, the upper and lower bounds based on the coefficients of the low-order operator (26) are independent of the time step and may turn out to be too restrictive. In this subsection, we concentrate on the treatment of mass antidiffusion $(M_L - M_C)\dot{u}$ assuming that the convective part f_{ij}^d of the target flux vanishes. In this case, the flow direction (upwind/downwind) is unknown and the antidiffusive flux may violate the positivity condition for both nodes. Therefore, we adopt the symmetric limiting strategy and discuss the choice of constraints to be imposed on the fully discretized target flux f_{ij}^m which corresponds to

$$f_{ij} = \frac{m_{ij}}{\Delta t}(u_i^{n+1} - u_j^{n+1}) - \frac{m_{ij}}{\Delta t}(u_i^n - u_j^n). \tag{28}$$

Interestingly enough, this flux consists of a truly antidiffusive implicit part and a diffusive explicit part which has a strong damping effect. In fact, explicit mass diffusion of the form $(M_C - M_L)u^n$ has frequently been used to construct a 'monotone' low-order method in the framework of high-resolution finite element schemes [11],[33],[34].

If the standard FEM-FCT algorithm is employed, the corresponding upper and lower bounds $Q_i^\pm$ depend on the local extrema $\tilde{u}_i^\pm$ of the low-order solution $\tilde{u} = u^n + \Delta t M_L^{-1} L u^n$ which reduces to u^n in the case $L = 0$ (no convection). In order to avoid the computation of $\tilde{u}$ and accommodate the contribution of the convective term in what follows, we use a weaker constraint and redefine the auxiliary quantities $P_i^\pm$ and $Q_i^\pm$ as follows

$$P_i^\pm = \sum_{j \neq i} \begin{matrix}\max\\\min\end{matrix} \{0, f_{ij}\}, \qquad Q_i^\pm = \sum_{j \neq i} \frac{m_{ij}}{\Delta t} \begin{matrix}\max\\\min\end{matrix} \{0, u_j^n - u_i^n\}, \tag{29}$$

where the off-diagonal coefficients of the consistent mass matrix m_{ij} are tacitly assumed to be nonnegative. Note that the nodal correction factors $R_i^\pm = \min\{1, Q_i^\pm / P_i^\pm\}$ are independent of the time step, since both $P_i^\pm$ and $Q_i^\pm$ are inversely proportional to it.

If the coefficient $p_{ij}^n = f_{ij}/(u_j^n - u_i^n)$ is nonnegative, the target flux (28) turns out to be diffusive, which may or may not be desirable. Otherwise, it may violate the positivity constraint for both nodes and should be limited in a symmetric fashion

$$f_{ij}^* = \begin{cases} \min\{R_i^+, R_j^-\} f_{ij}, & \text{if } f_{ij} > 0, \\ \min\{R_i^-, R_j^+\} f_{ij}, & \text{otherwise,} \end{cases} \qquad f_{ji}^* = -f_{ij}^*. \tag{30}$$

Another way to define $Q_i^\pm$ is to replace u_j^n in (29) by the local extrema $u_i^+ = \max_j u_j^n$ and $u_i^- = \min_j u_j^n$ evaluated over j such that $m_{ij} \neq 0$. This yields $Q_i^\pm = \frac{m_i - m_{ii}}{\Delta t}(u_i^\pm - u_i^n)$, where $m_i - m_{ii} = \sum_{j \neq i} m_{ij}$ is the difference between the diagonal entries of M_L and M_C.

4.3 General-purpose flux limiter

Now that we have a stand-alone flux limiter for convective antidiffusion (see Section 4.1) and a stand-alone flux limiter for mass antidiffusion (see Section 4.2) at our disposal, we can proceed to the treatment of antidiffusive fluxes (10) which involve both contributions. The operator splitting approach, i.e., a segregated limiting of f_{ij}^d and f_{ij}^m is feasible but the results are rather disappointing, especially if the two components have different signs. In particular, the magnitude of the antidiffusive flux may increase, which is clearly unacceptable. Furthermore, our experience with flux correction of FCT type indicates that it is worthwhile to prelimit f_{ij} so as to prevent it from becoming diffusive and creating numerical artifacts [20],[23]. Therefore, let us adjust the target fluxes thus:

$$f_{ij} := \min\{0, p_{ij}\}(u_j - u_i), \qquad p_{ij} = (f_{ij}^d + f_{ij}^m)/(u_j - u_i). \tag{31}$$

It remains to specify the upper/lower bounds $Q_i^\pm$ and choose the flux limiting strategy. Both algorithms considered so far are directly applicable to target fluxes of the form (31) but their performance is highly problem-dependent. It is not unusual that $p_{ij} + l_{ji} < 0$ if mass antidiffusion is strong enough, which means that a significant portion of the target flux cannot be recovered by the upwind-biased flux limiter alone. In other cases, symmetric flux limiting may produce inferior results because taking the minimum of nodal correction factors turns out to be more restrictive than prelimiting based on (22).

A straightforward but inefficient way to combine the two flux limiting techniques is to apply them sequentially. For instance, one can use the upwind-biased algorithm (25)–(27) to predict f_{ij}^* and limit the rejected antidiffusion $\Delta f_{ij} = f_{ij} - f_{ij}^*$ according to (29)-(30) or vice versa. In any event, the effective upper and lower bounds for the sum of limited antidiffusive fluxes $f_{ij}^* + \Delta f_{ij}^*$ consist of the 'stationary' upwind part (26) and the 'time-dependent' symmetric part (29) which complement each other in the following way

- the former makes sure that a certain fraction of admissible antidiffusion is independent of the time step, which prevents a loss of accuracy in steady-state computations;
- the latter makes sure that solutions to truly time-dependent problems become more accurate as Δt is refined, since a larger portion of the target flux may be retained.

Both constituents of $Q_i^\pm$ were constructed using heuristic arguments rather than the intrinsic 'CFL' condition which requires that the diagonal coefficient in the right-hand side of (4) be nonnegative for a given Δt. Such estimates would be expensive to obtain and sometimes overly restrictive, e.g., for stationary problems solved by time marching. Therefore, we deliberately relax them to make the algorithm more efficient, improve the convergence rates, and satisfy the discrete maximum principle in the steady-state limit.

Instead of limiting the target fluxes by the algorithms (25)–(27) and/or (29)-(30) in a segregated way or sequentially, it is worthwhile to combine the corresponding quantities $P_i^\pm$ and $Q_i^\pm$, which leads to the following general-purpose (GP) limiting strategy

1. Use prelimiting (22) to split the target flux (31) into the 'upwind' part f_{ij}' and the remainder $\Delta f_{ij} := f_{ij} - f_{ij}'$ which violates the positivity constraint for node j.

2. Compute the total sums of raw antidiffusive fluxes which need to be constrained
$$P_i^\pm = \sum_{j \in \mathcal{J}_i} \begin{matrix}\max\\ \min\end{matrix} \{0, f_{ij}'\} + \sum_{j \neq i} \begin{matrix}\max\\ \min\end{matrix} \{0, \Delta f_{ij}\}. \tag{32}$$

3. Define the combined upper/lower bounds to be enforced on $P_i^\pm$ as follows
$$Q_i^\pm = \sum_{j \neq i} \left[\frac{m_{ij}}{\Delta t} + l_{ij}\right] \begin{matrix}\max\\ \min\end{matrix} (u_j - u_i). \tag{33}$$

4. Evaluate the nodal correction factors (19) for the flux limiting step
$$R_i^\pm = \min\{1, Q_i^\pm / P_i^\pm\}. \tag{34}$$

5. In a loop over edges, compute the antidiffusive correction $f_{ij}^* + \Delta f_{ij}^*$, where
$$f_{ij}^* = R_i^\pm f_{ij}', \qquad \Delta f_{ij}^* = \min\{R_i^\pm, R_j^\mp\} \Delta f_{ij}. \tag{35}$$

Note that the first sum in (32) is evaluated over the set of downwind nodes $\mathcal{J}_i$ (see (25)) while the second one contains antidiffusive edge contributions from all neighboring nodes.

The above algorithm reduces to its prototypes (25)–(27) and (29)-(30) in the special cases of a lumped mass matrix ($m_{ij} = 0$) or zero velocity ($d_{ij} = 0,\ l_{ij} = 0$), respectively. Of course, there are many other ways to select and enforce the upper/lower bounds. This flexibility may be used to incorporate additional (geometric) constraints such as *linearity preservation* [6] so as to provide optimal accuracy and/or consistency on irregular meshes. Ideally, the limiter should be designed so that the high-order scheme is recovered if the solution is smooth enough or $h \to 0$. On the other hand, the accuracy of the target flux rather than the choice of constraints and the type of flux limiting is decisive in many cases. Hence, the use of higher-order finite elements and/or time-stepping schemes appears to be a promising way to improve the performance of algebraic flux correction schemes.

5 Practical implementation

To make the presentation self-contained, we touch upon the iterative treatment of nonlinearities and discuss the practical implementation of the GP flux limiter (32)–(35) at the end of this section. In this paper, emphasis is laid on implicit time discretizations, since the fully explicit case is trivial from the viewpoint of linear algebra (no linear systems need to be solved). Moreover, if the use of small time steps is dictated by accuracy considerations, the explicit FEM-FCT algorithm of Löhner *et al.* [27] can be employed to constrain the target fluxes (31) in an efficient manner. Our goal is to develop a general methodology which is applicable to implicit finite element discretizations and provides a sufficiently accurate treatment of both stationary and time-dependent problems.

After the discretization in time by an implicit $\theta-$scheme such that $0 < \theta \leq 1$, the flux-limited Galerkin scheme can be represented in the form

$$[M_L - \theta\Delta t L]u^{n+1} = [M_L + (1-\theta)\Delta t L]u^n + \Delta t f^*, \tag{36}$$

where the last term is assembled from the limited antidiffusive fluxes given by (35)

$$f_i^* = \sum_{j\neq i}[f_{ij}^* + \Delta f_{ij}^*]. \tag{37}$$

This nonlinear algebraic system must be solved iteratively. Let us compute successive approximations to the solution u^{n+1} using the straightforward defect correction scheme

$$u^{(m+1)} = u^{(m)} + [A(u^{(m)})]^{-1}r^{(m)}, \qquad m = 0, 1, 2.\dots \tag{38}$$

where the residual $r^{(m)}$ consists of a low-order part plus limited antidiffusion

$$r^{(m)} = [M_L + (1-\theta)\Delta t L]u^n - [M_L - \theta\Delta t L]u^{(m)} + \Delta t f^* \tag{39}$$

and $A(u^{(m)})$ is a suitably chosen 'preconditioner'. Some typical choices are

$$A = M_L \tag{40}$$

(only suitable for very small Δt) and the low-order operator [21],[22]

$$A = M_L - \theta\Delta t L \tag{41}$$

which was designed to be an M-matrix. Alternatively, algebraic flux/defect correction schemes may be preconditioned by the nonlinear LED operator

$$A = M_L - \theta \Delta t L^*(u), \tag{42}$$

where $L^*(u)$ includes limited antidiffusion. The existence of this operator is guaranteed by the flux limiter [22],[23]. This kind of preconditioning renders all intermediate solutions $u^{(m)}$ positivity-preserving [17] but convergence is a prerequisite for mass conservation.

In practice, the 'inversion' of A is performed by solving the linear subproblem

$$A\Delta u^{(m+1)} = r^{(m)}, \qquad m = 0, 1, 2, \dots \tag{43}$$

After a certain number of inner iterations, the solution increment $\Delta u^{(m+1)}$ is applied to the last iterate, whereby u^n provides a reasonable initial guess

$$u^{(m+1)} = u^{(m)} + \Delta u^{(m+1)}, \qquad u^{(0)} = u^n. \tag{44}$$

The iteration process is terminated when a certain norm of the defect $r^{(m)}$ or that of the relative changes $\Delta u^{(m+1)}$ becomes small enough. Explicit and/or implicit underrelaxation techniques may be invoked to secure the convergence of outer iterations [13].

Let us summarize what we have said so far and piece together a practical algorithm for node-oriented flux correction based on the general-pupose flux limiter (32)–(35)

1. For each pair of neighboring nodes i and j, orient the edge $\vec{ij}$ so that $l_{ij} \leq l_{ji}$, prelimit the flux f_{ij} in accordance with (22) and compute $\Delta f_{ij} := f_{ij} - f'_{ij}$.

2. Add the corresponding edge contributions to the sums of positive/negative fluxes

$$P_i^\pm := P_i^\pm + \begin{matrix}\max\\\min\end{matrix}\{0, f_{ij}\}, \qquad P_j^\pm := P_j^\pm + \begin{matrix}\max\\\min\end{matrix}\{0, -\Delta f_{ij}\}. \tag{45}$$

3. Update the combined upper/lower bounds (33) for both nodes as follows

$$Q_i^\pm := Q_i^\pm + \left[\frac{m_{ij}}{\Delta t} + l_{ij}\right] \begin{matrix}\max\\\min\end{matrix}\{0, u_j - u_i\},$$
$$Q_j^\pm := Q_j^\pm + \left[\frac{m_{ji}}{\Delta t} + l_{ji}\right] \begin{matrix}\max\\\min\end{matrix}\{0, u_i - u_j\}. \tag{46}$$

4. In a loop over nodes, compute the nodal correction factors to be applied

$$R_i^\pm = \min\{1, Q_i^\pm / P_i^\pm\}. \tag{47}$$

5. Multiply the upwind part f'_{ij} by $R_i^\pm$ and add its contribution to the defect r

$$f_{ij}^* = \begin{cases} R_i^+ f'_{ij}, & \text{if } f'_{ij} > 0, \\ R_i^- f'_{ij}, & \text{otherwise,} \end{cases} \qquad \begin{aligned} r_i &:= r_i + \Delta t f_{ij}^*, \\ r_j &:= r_j - \Delta t f_{ij}^*. \end{aligned} \tag{48}$$

6. Limit the remainder Δf_{ij} in a symmetric fashion and insert it into the defect

$$\Delta f_{ij}^* = \begin{cases} \min\{R_i^+, R_j^-\}\Delta f_{ij}, & \text{if } \Delta f_{ij} > 0, \\ \min\{R_i^-, R_j^+\}\Delta f_{ij}, & \text{otherwise,} \end{cases} \qquad \begin{aligned} r_i &:= r_i + \Delta t \Delta f_{ij}^*, \\ r_j &:= r_j - \Delta t \Delta f_{ij}^*. \end{aligned} \tag{49}$$

This 'black-box' algorithm can be readily integrated into existing finite element codes based on both conventional (element-based) and edge-based data structures.

6 Numerical examples

In order to illustrate the ideas presented in this paper, we apply the new limiting strategy to the continuity equation (1) discretized in space by P_1/Q_1 finite elements. Throughout this section, the numerical error will be estimated by measuring the difference between the exact solution u and its finite element approximation u_h in the discrete L_1-norm

$$E_1 = \sum_i m_i |u(x_i, y_i) - u_i| \approx \int_\Omega |u - u_h| \,\mathrm{d}x = ||u - u_h||_1 \tag{50}$$

as well as in the discrete L_2-norm defined by the formula

$$E_2 = \sqrt{\sum_i m_i |u(x_i, y_i) - u_i|^2} \approx \sqrt{\int_\Omega |u - u_h|^2 \,\mathrm{d}x} = ||u - u_h||_2, \tag{51}$$

where $m_i = \int_\Omega \varphi_i \,\mathrm{d}x$ are the diagonal coefficients of the lumped mass matrix. The error norms will be presented in Fig. 1-8 along with the corresponding numerical solutions.

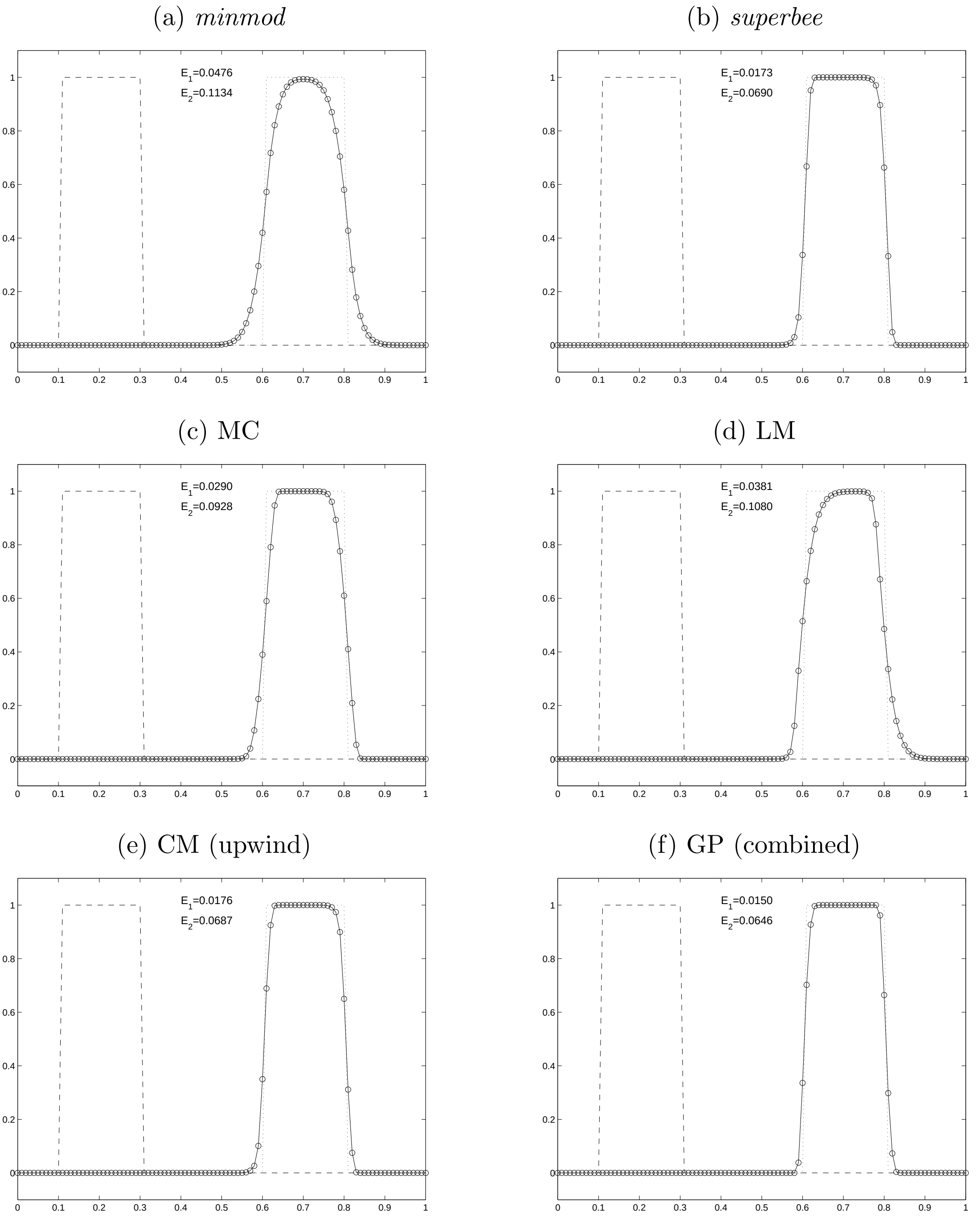
6.1 Convection of a square wave

Let us start with a classical test problem which consists of solving the one-dimensional convection equation (12) for the discontinuous initial data

$$u(x, 0) = \begin{cases} 1 & \text{if} \quad |x - 0.2| \leq 0.1 \\ 0 & \text{otherwise} \end{cases} \tag{52}$$

depicted as dashed lines in Fig. 1. The dotted lines show the exact solution for $v = 1$ and $t = 0.5$ which is obtained by translation of the initial profile along the $x-$axis. The domain $(0, 1)$ is discretized by linear finite elements of equal length $\Delta x = 10^{-2}$, so that the time step $\Delta t = 10^{-3}$ used to compute the solutions in Fig. 1 corresponds to the Courant number $\nu = 0.1$. The discretization in time is performed by the second-order accurate Lax-Wendroff method so that $d_{ij} = (1 - \nu)v/2$. The accuracy of numerical solutions is evaluated in terms of the discrete error norms (50) and (51) which are included in all diagrams. The behavior of standard TVD schemes for this simple test problem is well known. As usual, the most diffusive results are produced by the *minmod* limiter, while *superbee* performs best on such discontinuous solutions but tends to corrupt smooth profiles due to artificial steepening. Limiters like MC produce acceptable results in either case and are typically used by default. For the square wave problem, the MC limiter proves far superior to *minmod* but less accurate than *superbee*, see Fig. 1a-c.

The target flux for the Lax-Wendroff scheme corresponds to that for a lumped-mass (LM) Taylor-Galerkin method of second order. As shown in Fig. 1d, the resulting solution is asymmetric, whereby the right flank of the square wave is reproduced much better than the left one. The latter is smeared as much as that for *minmod*, which is due to mass lumping. Adding the contribution of the consistent mass (CM) matrix yields a target flux with improved phase characteristics [10]. Limiting it as before in accordance with (16) is equivalent to the use of algebraic flux correction based on (25)–(27). The numerical



(a) *minmod*

(b) *superbee*

(c) MC

(d) LM

(e) CM (upwind)

(f) GP (combined)

Figure 1. Convection of a square wave: numerical solutions at $t = 0.5$.

solution displayed in Fig. 1e resembles that produced by *superbee*. Note that the upper right corner of the square wave remains 'rounded' because the bounds (26) turn out to be too restrictive for transient problems. This can be rectified by invoking the general-purpose limiter (32)–(35) which is designed to become more accurate as the time step is refined. An equally crisp resolution of both flanks is obtained for $\Delta t = 10^{-4}$, see Fig. 1f. We conclude that the use of a consistent mass matrix is essential not only for the definition of the target flux but also for the estimation of upper and lower bounds.

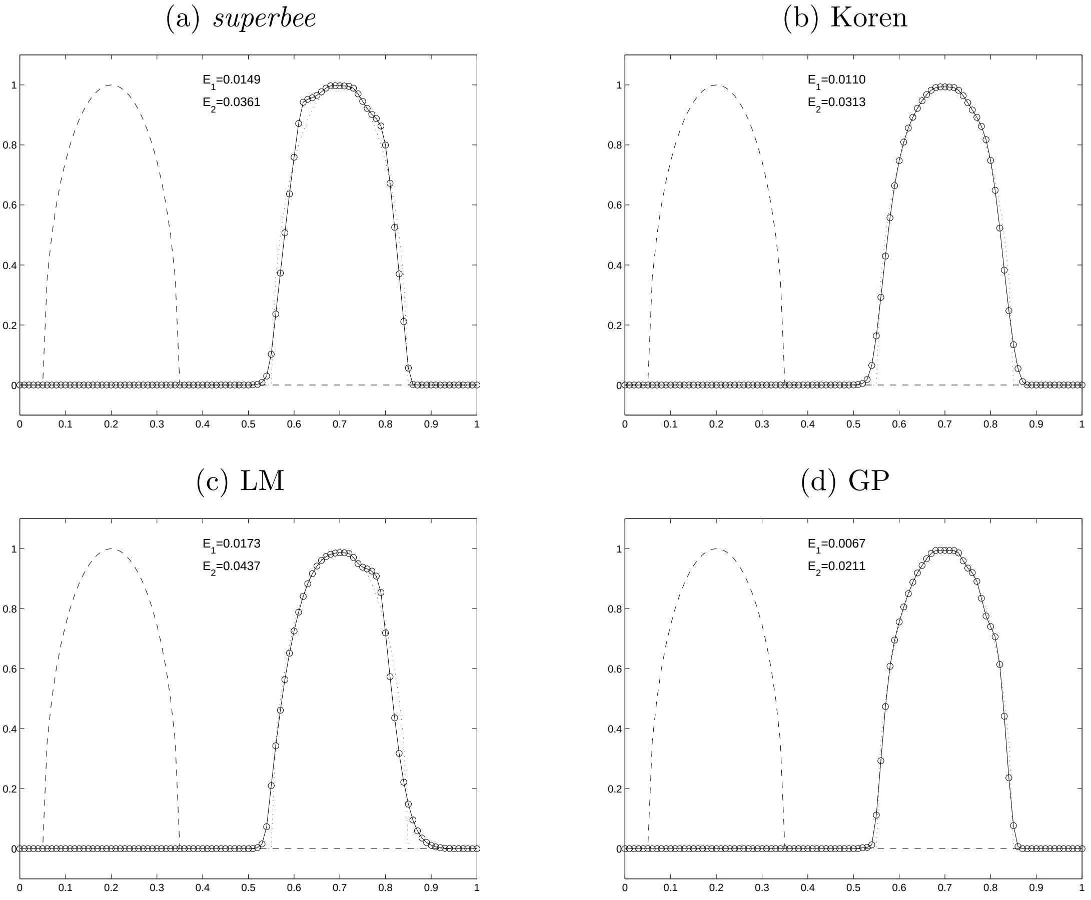
6.2 Convection of a semi-ellipse

Our second test problem is a slightly modified version of the one used in [32],[39],[40] to expose the 'terracing' phenomenon, an infamous byproduct of flux limiting. The linear convection equation is solved for continuous initial data given by the formula

$$u(x,0) = \sqrt{1 - \left(\frac{x - 0.2}{0.15}\right)^2} \qquad \text{if} \quad |x - 0.2| \leq 0.15 \tag{53}$$

and $u(x,0) = 0$ otherwise. All discretization parameters are the same as in the first example. The challenge of the second test consists in resolving the steep parts of the otherwise smooth profile without generating spurious kinks or plateaus. Such a nonphysical solution behavior, which is a common drawback of many modern high-resolution schemes, is referred to as terracing and can be interpreted as 'an integrated, nonlinear effect of residual phase errors' [32] or, loosely speaking, 'the ghosts of departed ripples' [5].

Terracing was first discovered in the FCT context but it is also typical of compressive TVD limiters like *superbee*, see Fig. 2a. At the same time, an excellent solution is produced by Koren's limiter (Fig. 2b) which is based on a third-order accurate target flux. In the finite element framework, mass lumping tends to aggravate phase errors, which manifests



(a) *superbee*

(b) Koren

(c) LM

(d) GP

Figure 2. Convection of a semi-ellipse: numerical solutions at $t = 0.5$.

itself in a pronounced terracing (Fig. 2c). The solution displayed in Fig. 2d illustrates the benefits of using the consistent mass matrix in conjunction with the general-purpose flux limiter. Remarkably, the corresponding error norms are even smaller that those for the optically perfect solution in Fig. 2b. The observed improvement in comparison to the lumped-mass version supports the conjecture that terracing can be cured to some extent by increasing the resolving power of the target flux so as to reduce the dispersive errors [39],[40]. Numerical experiments indicate that the small but still noticeable deviations from the exact shape at the right edge of the semi-ellipse in Fig. 2d are caused by the fluxes that prove insufficiently antidiffusive (more diffusive than *minmod* and, consequently, not linearity-preserving) in spite of the prelimiting performed in (31). Indeed, false diffusion cannot be detected by the flux limiter and should be filtered out beforehand.

6.3 Solid body rotation

Let us proceed to the two-dimensional benchmark problem proposed by LeVeque [26] which makes it possible to assess the ability of a high-resolution scheme to preserve both smooth and discontinuous profiles. To this end, a slotted cylinder, a sharp cone and a smooth hump are exposed to the nonuniform velocity field $\mathbf{v} = (0.5 - y, x - 0.5)$ and undergo a counterclockwise rotation about the center of the unit square $\Omega = (0,1) \times (0,1)$. Each solid body lies within a circle of radius $r_0 = 0.15$ centered at a point with Cartesian coordinates (x_0, y_0). In the rest of the domain, the solution is initialized by zero. The shapes of the three bodies as depicted in Fig. 3 can be expressed in terms of the normalized distance function for the respective reference point (x_0, y_0) thus:

$$r(x,y) = \frac{1}{r_0}\sqrt{(x-x_0)^2 + (y-y_0)^2}.$$

The center of the slotted cylinder is located at $(x_0, y_0) = (0.5, 0.75)$ and its geometry in the circular region $r(x,y) \leq 1$ is given by

$$u(x,y,0) = \begin{cases} 1 & \text{if } |x - x_0| \geq 0.025 \vee y \geq 0.85, \\ 0 & \text{otherwise.} \end{cases}$$

The corresponding analytical expression for the conical body reads

$$u(x,y,0) = 1 - r(x,y), \qquad (x_0, y_0) = (0.5, 0.25),$$

whereas the shape and location of the hump at $t = 0$ are as follows

$$u(x,y,0) = 0.25[1 + \cos(\pi \min\{r(x,y), 1\})], \quad (x_0, y_0) = (0.25, 0.5).$$

After one full revolution ($t = 2\pi$) the exact solution of the continuity equation (1) coincides with the initial data. The numerical solutions presented in Fig. 4-6 were computed on a uniform mesh of 128 × 128 bilinear finite elements using the second-order accurate Crank-Nicolson time-stepping ($\theta = 0.5$) with $\Delta t = 10^{-3}$. The general-purpose (GP) algorithm (32)-(35) produces the most accurate results shown in Fig. 4. The cone and hump are reproduced very well and even the narrow bridge of the slotted cylinder

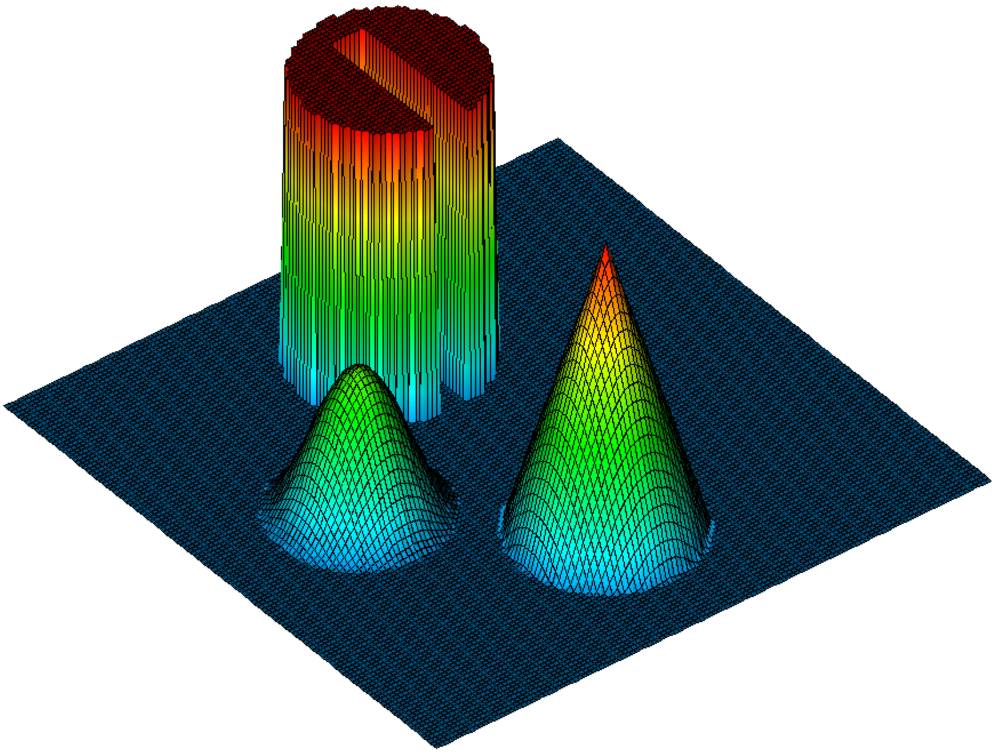


Figure 3. Solid body rotation: initial data / exact solution.

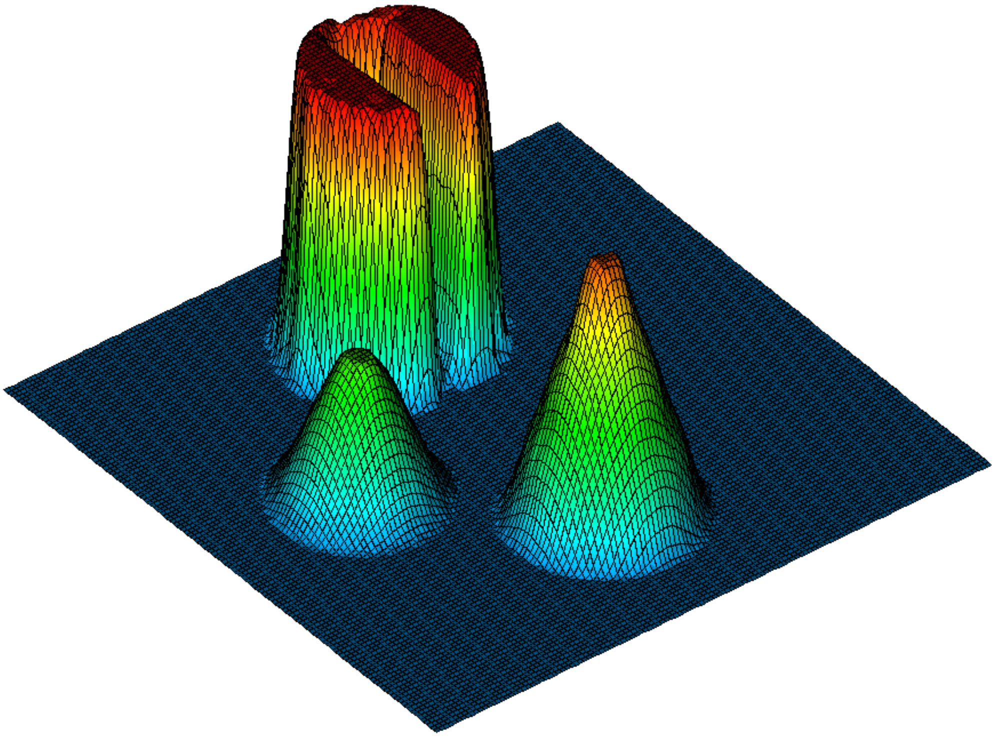


Figure 4. Solid body rotation (GP), $E_1 = 0.0111$, $E_2 = 0.0567$.

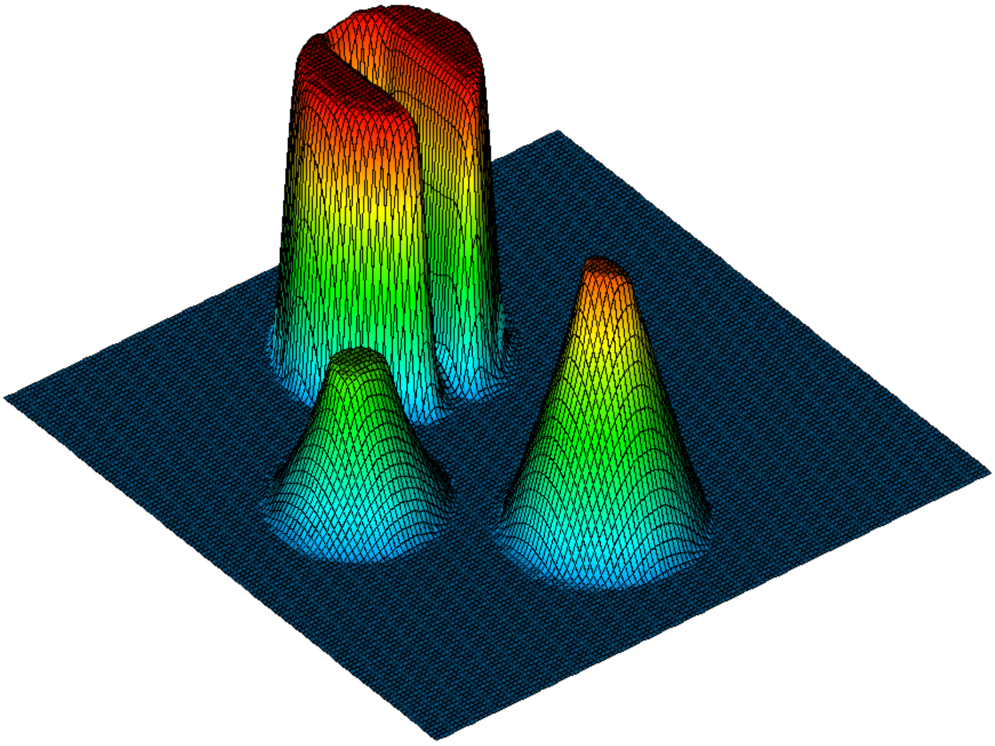


Figure 5. Solid body rotation (*superbee*), $E_1 = 0.0139$, $E_2 = 0.0610$.

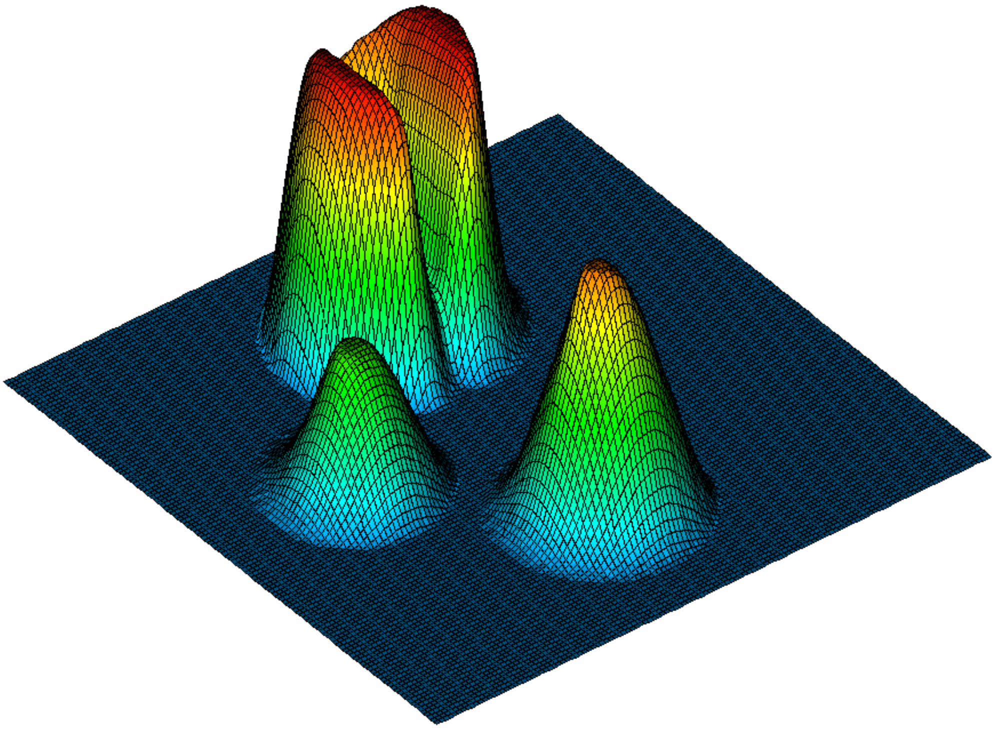


Figure 6. Solid body rotation (MC), $E_1 = 0.0255$, $E_2 = 0.0889$.

is largely preserved. Not surprisingly, this solution is very similar to that computed by an FCT algorithm based on the same target flux [21]. In either case, the prelimiting of antidiffusive fluxes in (31) is essential. If it is not performed, the ridges of the cylinder are subject to spurious erosion which can be interpreted as a sort of terracing.

By contrast, the performance of standard TVD limiters for this time-dependent test problem leaves a lot to be desired. The strong antidiffusion inherent to *superbee* alleviates the diffusive effect of mass lumping and yields a fairly good resolution of the slotted cylinder (Fig. 5) but entails a pronounced flattening of the smooth peaks. The numerical solution produced by the 'default' MC limiter (Fig. 6) exhibits both a strong smearing of the slotted cylinder and a noticeable distortion of the cone and hump.

6.4 Convection in space-time

If the problem at hand is stationary, the time derivative vanishes and so does the contribution of the consistent mass matrix. Therefore, mass lumping is appropriate, i.e., the raw antidiffusive flux is given by (24) and the upwind-biased algorithm (25)–(27) can be employed. Due to the fact that the underlying upper/lower bounds (26) are independent of the time step, it is possible to compute the steady-state solution directly or by means of pseudo-time-stepping based on the fully implicit backward Euler scheme ($\theta = 1$). In the latter case, the time step represents a variable underrelaxation parameter [13] which should be chosen as large as possible to reduce the computational cost. For an FCT-like limiter, whereby each solution update is required to be positivity-preserving, this would entail an irrecoverable loss of accuracy, since the nodal correction factors are inversely proportional to Δt. At the same time, our general-purpose algorithm is free of this drawback because it becomes equivalent to (25)–(27) for large time steps.

Let us return to the square wave test and reformulate the one-dimensional convection equation with $v = 0.5$ as a stationary problem of the form (1) with $\mathbf{v} = (0.5, 1)$. This corresponds to computing the solution for all time levels simultaneously instead of doing it step-by-step as usual [23]. The boundary conditions to be imposed at the 'inlet' of the space-time domain $\Omega = (0,1) \times (0,1)$ can be inferred from the exact solution given by

$$u(x,t) = \begin{cases} 1 & \text{if } |x - 0.5t - 0.2| \leq 0.1, \\ 0 & \text{otherwise.} \end{cases} \tag{54}$$

The initial data can be chosen arbitrarily since they do not affect the converged steady-state solution. For instance, the approximate solution can be initialized using (54). The numerical results obtained using algebraic flux correction (25)–(27) based on the lumped-mass (LM) Galerkin flux and the standard *minmod* limiter are presented in Fig. 7 and Fig. 8, respectively. Both solutions were marched to the steady state by the backward Euler method, whereby the time step $\Delta t = 1.0$ was intentionally chosen to be very large. The discontinuous initial profile is shown in the background, while the solution at time $t = 1$ appears in the front. This example demonstrates that the algorithm to which our GP limiter reduces in the stationary case performs much better than *minmod*, the only standard TVD limiter which is consistent with the underlying finite element scheme.

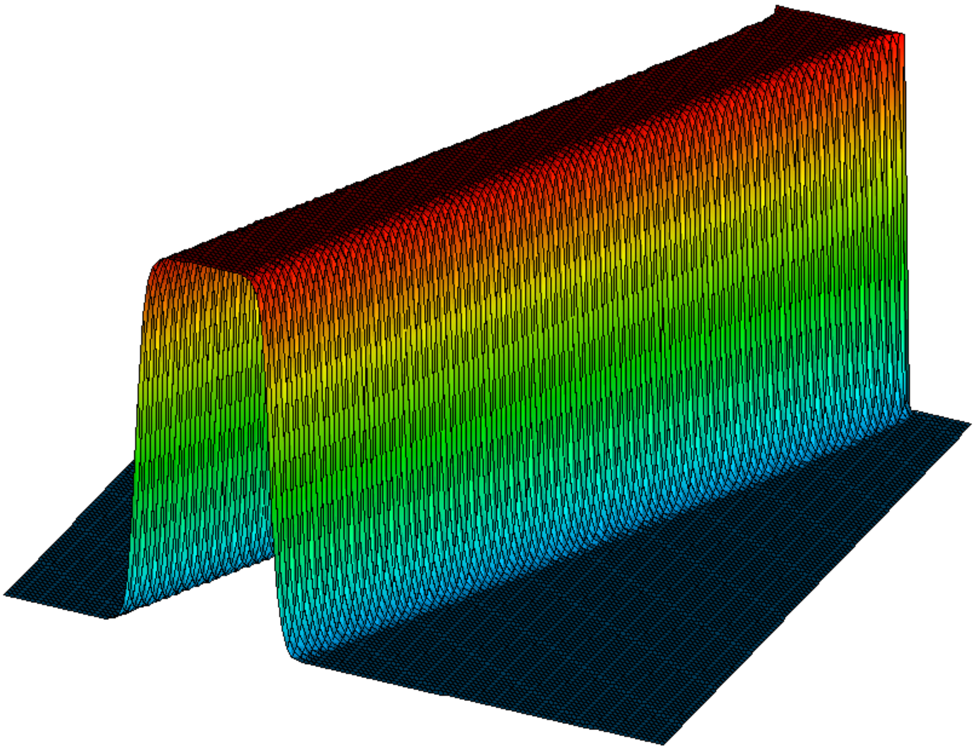


Figure 7. Convection in space-time (LM), $E_1 = 0.0179,\ E_2 = 0.0698$.

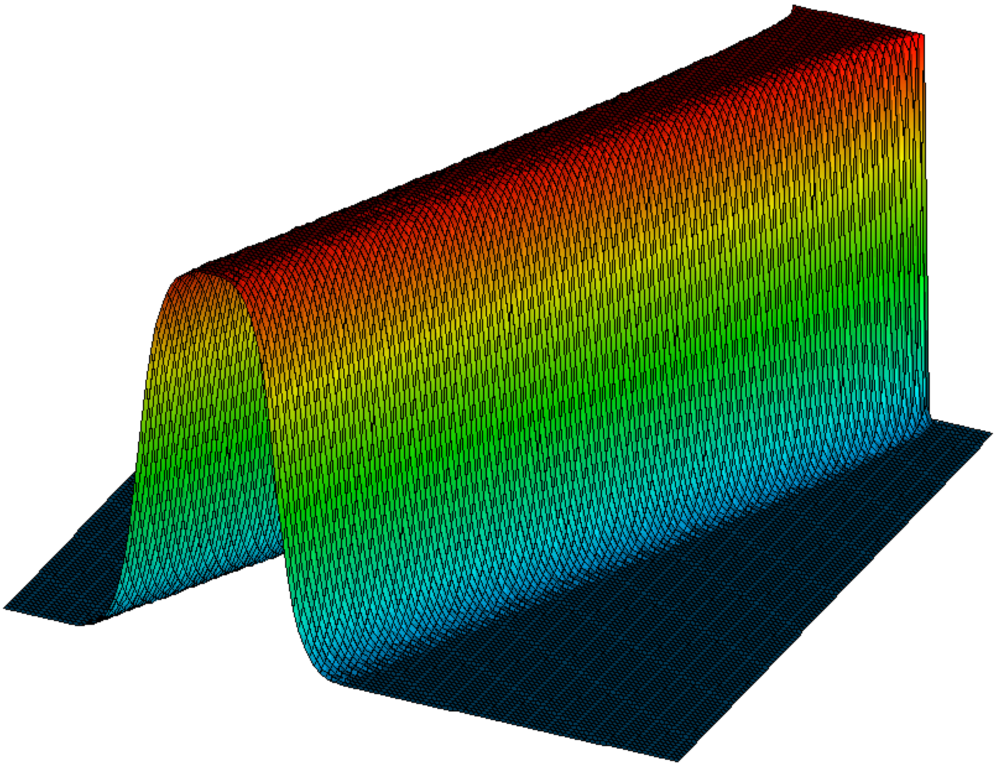


Figure 8. Convection in space-time (*minmod*), $E_1 = 0.0340,\ E_2 = 0.0971$.

7 Conclusions

In this paper, we focused on the design of general-purpose flux limiters for implicit finite element discretizations including a consistent mass matrix. Algebraic constraints were imposed node-by-node so as to control the sum of edge contributions with negative coefficients. The choice of target fluxes was addressed and a fully multidimensional limiting strategy was presented. The upper/lower bounds for the sum of positive/negative antidiffusive fluxes were designed so as to enforce the LED property. A combination of flux limiters derived separately for two special cases (consistent-mass L_2-projection and lumped-mass Galerkin approximation) was found to strike the balance between accuracy and efficiency. The new algorithm, which combines the advantages of algebraic FCT and TVD schemes [21],[22], proves sufficiently accurate for stationary and time-dependent problems alike. An extension of the proposed methodology to the Euler and Navier-Stokes equations of fluid dynamics can be readily performed as explained in [24],[37] in the context of algebraic TVD schemes. The design of general-purpose flux limiters for hyperbolic systems will be addressed in a forthcoming publication [25].

Algebraic flux correction of the form (17)-(20) provides a very general framework for the derivation of new high-resolution schemes for finite element discretizations on unstructured meshes. The resolving power and phase characteristics of the high-order scheme can be improved by adding some background diffusion or using a time-accurate approximation from the family of Taylor-Galerkin methods [10]. High-order finite elements / bubble functions lend themselves to the design of target fluxes, whereas fluctuation splitting techniques [6] seem to be a useful tool for the definition of upper/lower bounds. Last but not least, the use of flux limiters as implicit subgrid scale models for Monotonically Integrated Large Eddy Simulation (MILES) and/or error indicators for adaptive mesh refinement [31] constitutes another promising direction for further research.

References

[1] M. Arora and P. L. Roe, A well-behaved TVD limiter for high-resolution calculations of unsteady flow. *J. Comput. Phys.* **132** (1997) 3-11.

[2] P. Arminjon and A. Dervieux, Construction of TVD-like artificial viscosities on 2-dimensional arbitrary FEM grids. *INRIA Research Report* **1111** (1989).

[3] M. Berzins, Modified mass matrices and positivity preservation for hyperbolic and parabolic PDEs. *Commun. Numer. Methods Engrg.* **17** (2001) no. 9, 659-666.

[4] M. Berzins, Variable-order finite elements and positivity preservation for hyperbolic PDEs. *Appl. Numer. Math.* **48** (2004) no. 3-4, 271-292.

[5] D. L. Book, The conception, gestation, birth, and infancy of FCT. In: D. Kuzmin, R. Löhner and S. Turek (eds.) *Flux-Corrected Transport: Principles, Algorithms, and Applications.* Springer, 2005, 5-28.

[6] J.-C. Carette, H. Deconinck, H. Paillère und P.L. Roe, Multidimensional upwinding: Its relation to finite elements. *Int. J. Numer. Methods Fluids* **20** (1995) no. 8-9, 935-955.

[7] B. Cockburn und C.-W. Shu, The Runge-Kutta discontinuous Galerkin method for conservation laws. V: Multidimensional systems. *J. Comput. Phys.* **141** (1998) 199–224.

[8] B. Cockburn, G. E. Karniadakis und C.-W. Shu, The development of discontinuous Galerkin methods. In: *Discontinuous Galerkin methods. Theory, computation and applications,* LNCSE **11**, Springer, 2000, 3-50.

[9] H. Deconinck, H. Paillère, R. Struijs und P.L. Roe, Multidimensional upwind schemes based on fluctuation-splitting for systems of conservation laws. *Comput. Mech.* **11** (1993) no. 5-6, 323-340.

[10] J. Donea, L. Quartapelle and V. Selmin, An analysis of time discretization in the finite element solution of hyperbolic problems. *J. Comput. Phys.* **70** (1987) 463–499.

[11] J. Donea, V. Selmin and L. Quartapelle, Recent developments of the Taylor-Galerkin method for the numerical solution of hyperbolic problems. *Numerical methods for fluid dynamics III*, Oxford, 171-185 (1988).

[12] M. Feistauer, J. Felcman und M. Lukacova-Medvid'ova, Combined finite element-finite volume solution of compressible flow. *J. Comput. Appl. Math.* **63** (1995) no. 1-3, 179-199.

[13] J. H. Ferziger and M. Peric, *Computational Methods for Fluid Dynamics.* Springer, 1996.

[14] K. Hain, The partial donor cell method. *J. Comput. Phys.* **73** (1987) 131-147.

[15] A. Jameson, Computational algorithms for aerodynamic analysis and design. *Appl. Numer. Math.* **13** (1993) 383-422.

[16] A. Jameson, Positive schemes and shock modelling for compressible flows. *Int. J. Numer. Meth. Fluids* **20** (1995) 743–776.

[17] T. Jongen and Y.P. Marx, Design of an unconditionally stable, positive scheme for the $K-\varepsilon$ and two-layer turbulence models. *Comput. Fluids* **26** (1997) no. 5, 469-487.

[18] B. Koren, A robust upwind discretization method for advection, diffusion and source terms. In: C. B. Vreugdenhil *et al.* (eds.), *Numerical methods for advection - diffusion problems.* Braunschweig: Vieweg. *Notes Numer. Fluid Mech.* **45** (1993) 117-138.

[19] D. Kuzmin, Positive finite element schemes based on the flux-corrected transport procedure. In: K. J. Bathe (ed.), *Computational Fluid and Solid Mechanics*, Elsevier, 887-888 (2001).

[20] D. Kuzmin and S. Turek, Flux correction tools for finite elements. *J. Comput. Phys.* **175** (2002) 525-558.

[21] D. Kuzmin, M. Möller and S. Turek, High-resolution FEM-FCT schemes for multidimensional conservation laws. *Computer Meth. Appl. Mech. Engrg.* **193** (2004) 4915-4946.

[22] D. Kuzmin and S. Turek, High-resolution FEM-TVD schemes based on a fully multidimensional flux limiter. *J. Comput. Phys.* **198** (2004) 131-158.

[23] D. Kuzmin and M. Möller, Algebraic flux correction I. Scalar conservation laws. In: D. Kuzmin, R. Löhner and S. Turek (eds.) *Flux-Corrected Transport: Principles, Algorithms, and Applications.* Springer, 2005, 155-206.

[24] D. Kuzmin and M. Möller, Algebraic flux correction II. Compressible Euler equations. In: D. Kuzmin, R. Löhner and S. Turek (eds.) *Flux-Corrected Transport: Principles, Algorithms, and Applications.* Springer, 2005, 207-250.

[25] D. Kuzmin and M. Möller, On the design of general-purpose flux limiters for finite element schemes. II. Euler equations. In preparation.

[26] R. J. LeVeque, High-resolution conservative algorithms for advection in incompressible flow. *Siam J. Numer. Anal.* **33** (1996) 627–665.

[27] R. Löhner, K. Morgan, J. Peraire and 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.

[28] P. R. M. Lyra, *Unstructured Grid Adaptive Algorithms for Fluid Dynamics and Heat Conduction.* PhD thesis, University of Wales, Swansea, 1994.

[29] P. R. M. Lyra, K. Morgan, J. Peraire and J. Peiro, TVD algorithms for the solution of the compressible Euler equations on unstructured meshes. *Int. J. Numer. Meth. Fluids* **19** (1994) 827–847.

[30] R.J. MacKinnon and G.F. Carey, Positivity-preserving, flux-limited finite difference and finite element methods for reactive transport. *Int. J. Numer. Methods Fluids* **41** (2003), no. 2, 151-183.

[31] M. Möller and D. Kuzmin, Adaptive mesh refinement for high-resolution finite element schemes. Technical report **297**, University of Dortmund, 2005.

[32] E. S. Oran and J. P. Boris, *Numerical Simulation of Reactive Flow.* 2nd edition, Cambridge University Press, 2001.

[33] V. Selmin, Finite element solution of hyperbolic equations. I. One-dimensional case. *INRIA Research Report* **655** (1987).

[34] V. Selmin, Finite element solution of hyperbolic equations. II. Two-dimensional case. *INRIA Research Report* **708** (1987).

[35] A. Sokolichin, *Mathematische Modellbildung und numerische Simulation von Gas-Flüssigkeits-Blasenströmungen*. Habilitation thesis, University of Stuttgart, 2004.

[36] S. Turek, *Efficient Solvers for Incompressible Flow Problems: An Algorithmic and Computational Approach*, LNCSE **6**, Springer, 1999.

[37] S. Turek and D. Kuzmin, Algebraic flux correction III. Incompressible flow problems. In: D. Kuzmin, R. Löhner and S. Turek (eds.) *Flux-Corrected Transport: Principles, Algorithms, and Applications*. Springer, 2005, 251-296.

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

[39] S. T. Zalesak, A preliminary comparison of modern shock-capturing schemes: linear advection. In: R. Vichnevetsky and R. Stepleman (eds.) *Advances in Computer Methods for PDEs*. Publ. IMACS, 1987, 15-22.

[40] S. T. Zalesak, The design of Flux-Corrected Transport (FCT) algorithms for structured grids. In: D. Kuzmin, R. Löhner and S. Turek (eds.) *Flux-Corrected Transport: Principles, Algorithms, and Applications*. Springer, 2005, 29-78.