Explicit and implicit FEM-FCT algorithms with flux linearization

Dmitri Kuzmin

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

Abstract

A new approach to the design of flux-corrected transport (FCT) algorithms for linear/bilinear finite element approximations of convection-dominated transport problems is pursued. The raw antidiffusive fluxes are linearized about an intermediate solution computed by a positivity-preserving low-order scheme. By virtue of this linearization, the costly evaluation of correction factors needs to be performed just once per time step, and no nonlinear algebraic systems need to be solved if the governing equation is linear. Furthermore, no questionable 'prelimiting' of antidiffusive fluxes is required, which eliminates the danger of artificial steepening. Three FEM-FCT algorithms based on the Runge-Kutta, Crank-Nicolson, and backward Euler time-stepping are proposed. Numerical results are presented for the linear convection equation as well as for the shock tube problem of gas dynamics.

Key words: Computational Fluid Dynamics, convection-dominated problems, high-resolution schemes, flux-corrected transport, finite element method
PACS: 02.60.Cb, 02.70.Dh, 47.11.Fg

1 Introduction

Modern high-resolution schemes for convection-dominated flows trace their origins to the *flux-corrected transport* (FCT) paradigm introduced in the early 1970s by Boris and Book [1]. The fully multidimensional generalization proposed by Zalesak [18] has formed a general framework for the design of FCT algorithms that represent a nonlinear blend of high- and low-order approximations. Unlike many other high-resolution schemes, flux correction of FCT type is feasible for finite element discretizations on unstructured meshes [14,15].

Email address: kuzmin@math.uni-dortmund.de (Dmitri Kuzmin).

Classical flux-corrected transport algorithms are based on an explicit correction of a provisional low-order solution, whose local maxima/minima provide the upper/lower bounds to be enforced. In the case of an implicit time discretization, the same strategy can be used to secure the positivity of the right-hand side, whereas the left-hand side is required to possess the M-matrix property [11]. The rationale for the development of (semi-)implicit FCT schemes stems from the fact that the CFL stability condition may become overly restrictive in the case of strongly nonuniform velocity fields and/or locally refined meshes. On the other hand, a properly configured implicit solver can be very efficient. If the time step is very small, a single Richardson iteration preconditioned by the lumped mass matrix may suffice. Therefore, the total CPU time can be comparable to that for a potentially unstable explicit algorithm.

In a series of recent publications [9–11,17], the FEM-FCT methodology was generalized to implicit time discretizations. However, the overhead cost of iterative flux/defect correction is rather high since the raw antidiffusive fluxes and the solution-dependent correction factors need to be updated in an iterative way. In addition, the presence of the consistent mass matrix has an adverse influence on the convergence rates [17]. Therefore, it is worthwhile to linearize raw antidiffusive fluxes so that the correction factors need to be computed just once per time step. The new linearization procedure proposed in this paper makes it possible to construct both explicit and implicit FEM-FCT algorithms which are slightly less accurate but much more efficient than those based on iterative flux correction. The fully implicit backward Euler FCT scheme is unconditionally positivity preserving. The ones based on the Runge-Kutta and Crank-Nicolson time-stepping must satisfy a CFL-like condition for the time step which can be readily inferred from the matrix properties.

Flux limiters of FCT type are to be recommended for strongly time-dependent problems since they are designed to accept more antidiffusion as the time step is refined. On the other hand, the use of large time steps results in a loss of accuracy. If the time derivative is expected to be small as compared to other terms in the governing equation, then mass lumping is appropriate and an upwind-biased flux limiter of TVD type [8,11] is to be preferred.

2 Design criteria

As a model problem, consider the time-dependent continuity equation

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

and a generic space discretization that can be represented as a system of ordinary differential equations for the vector of time-dependent nodal values

$$\frac{\mathrm{d}u}{\mathrm{d}t} = Cu. \tag{2}$$

The coefficients of the matrix $C = \{c_{ij}\}$ result from the discretization of the troublesome convective term. In general, they depend on the computational mesh, on the velocity field, and on the employed approximations.

The right-hand side of the $i-$th equation contains a (possibly nonlinear) combination of the solution values at node i and its nearest neighbors

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \sum_j c_{ij} u_j. \tag{3}$$

It is easy to prove that if all off-diagonal coefficients of C are nonnegative

$$c_{ij} \geq 0, \qquad \forall j \neq i \tag{4}$$

then the above semi-discrete scheme is *positivity-preserving*, i.e., [6]

$$u_i(t) = 0, \quad u_j(t) \geq 0, \quad \forall j \neq i \qquad \Rightarrow \qquad \frac{\mathrm{d}u_i}{\mathrm{d}t} \geq 0. \tag{5}$$

Incompressible velocity fields ($\nabla \cdot \mathbf{v} = 0$) typically give rise to coefficient matrices with zero row sums [11] so that $c_{ii} = -\sum_{j \neq i} c_{ij}$ and, therefore,

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \sum_{j \neq i} c_{ij}(u_j - u_i). \tag{6}$$

Such a semi-discrete scheme is not only positivity-preserving (PP) but also *local extremum diminishing* (LED) under the sufficient (but not necessary) condition (4). The LED criterion was introduced by Jameson [5] as a handy generalization of Harten's theorem that forms the basis for the design of one-dimensional *total variation diminishing* (TVD) schemes.

After the discretization in time by a two-level $\theta-$scheme, one obtains

$$[I - \theta \Delta t C]u^{n+1} = [I + (1-\theta)\Delta t C]u^n, \qquad 0 \leq \theta \leq 1. \tag{7}$$

This algebraic system can be written in the generic form

$$Au^{n+1} = Bu^n, \tag{8}$$

where $A = I - \theta\Delta t C$ and $B = I + (1-\theta)\Delta t C$. The properties of the resulting scheme depend on the choice of the implicitness parameter $\theta \in [0, 1]$.

In general, a solution update of the form (8) is positivity-preserving if A is a so-called *M-matrix* and all coefficients B are nonnegative [11]. By definition, the inverse of an M-matrix has no negative entries either so that

$$u^n \geq 0 \quad \Rightarrow \quad b = Bu^n \geq 0 \quad \Rightarrow \quad u^{n+1} = A^{-1}b \geq 0. \tag{9}$$

A simple test of the M-matrix property is based on the following set of sufficient conditions which are widely used to prove the *discrete maximum principle* (DMP) for finite element discretizations of elliptic and parabolic problems

$$a_{ii} > 0, \quad a_{ij} \leq 0, \quad \sum_j a_{ij} > 0, \quad \forall i,\ \forall j \neq i. \tag{10}$$

By virtue of (4), the matrix $A = I - \theta\Delta t C$ satisfies (10) automatically and $B = I + (1-\theta)\Delta t C$ has no negative off-diagonal coefficients. The diagonal entries of B are also nonnegative if the time discretization is fully implicit ($\theta = 1$) or the time step Δt satisfies the additional CFL-like condition

$$1 + (1-\theta)\Delta t \min_i c_{ii} \geq 0. \tag{11}$$

Thus, the positivity constraint for (7) holds under conditions (4) and (11) which are easy to check and enforce for arbitrary space discretizations.

3 Algebraic flux correction

According to the well-known Godunov theorem, all linear positivity-preserving schemes are doomed to be (at most) first-order accurate. On the other hand, a high-order discretization that fails to satisfy the imposed algebraic constraints can be fixed *a posteriori* by adding a linear or nonlinear artificial diffusion operator designed so as to achieve the desired matrix properties. This approach to the derivation of high-resolution schemes can be classified as *algebraic flux correction* [11]. In this section, we explain how to enforce the positivity constraint and achieve high resolution in a mass-conserving fashion.

The standard Galerkin discretization of equation (1) can be written as

$$M_C \frac{\mathrm{d}u}{\mathrm{d}t} = Ku, \tag{12}$$

where the consistent mass matrix $M_C = \{m_{ij}\}$ and the discrete transport operator $K = \{k_{ij}\}$ fail to satisfy the imposed algebraic constraints.

In order to render (12) positivity-preserving, we replace the mass matrix M_C by its lumped counterpart $M_L = \mathrm{diag}\{m_i\}$ with diagonal entries

$$m_i = \sum_j m_{ij} \tag{13}$$

and add an artificial diffusion operator $D = \{d_{ij}\}$, which yields

$$M_L \frac{\mathrm{d}u}{\mathrm{d}t} = Lu, \qquad L = K + D. \tag{14}$$

This semi-discrete scheme is positivity-preserving if $l_{ij} = k_{ij} + d_{ij} \geq 0,\ \forall j \neq i$. Hence, it is natural to define the off-diagonal coefficients of D as follows

$$d_{ij} = \max\{-k_{ij}, 0, -k_{ji}\}, \quad \forall j \neq i. \tag{15}$$

The diagonal coefficients d_{ii} are defined so that the row and column sums of the symmetric matrix D are equal to zero

$$d_{ii} := -\sum_{j \neq i} d_{ij}. \tag{16}$$

In practice, there is no need to assemble the global matrix D. Instead, the discrete transport operator K should be modified directly [11]

$$\begin{aligned} &k_{ii} := k_{ii} - d_{ij}, \qquad k_{ij} := k_{ij} + d_{ij}, \\ &k_{ji} := k_{ji} + d_{ij}, \qquad k_{jj} := k_{jj} - d_{ij}. \end{aligned} \tag{17}$$

In the 1D case, the elimination of negative off-diagonal coefficients transforms a lumped-mass Galerkin scheme into the classical upwind difference discretization which represents the least diffusive linear LED scheme [9,11].

The low-order scheme (14) is nonoscillatory but its accuracy is very poor. In order to achieve high resolution, it is necessary to remove excessive artificial diffusion. Since both D and $M_C - M_L$ are symmetric matrices with zero row and column sums, the difference between the residuals of (12) and (14)

$$f = (M_L - M_C)\frac{\mathrm{d}u}{\mathrm{d}t} - Du \tag{18}$$

admits a conservative decomposition into a sum of *raw antidiffusive fluxes* that can be associated with edges of the sparsity graph of the global matrix

$$f_i = \sum_{j \neq i} f_{ij}, \qquad f_{ij} = \left[m_{ij} \frac{\mathrm{d}}{\mathrm{d}t} + d_{ij} \right] (u_i - u_j) = -f_{ji}. \tag{19}$$

In the course of algebraic flux correction, every antidiffusive flux f_{ij} is multiplied by a solution-dependent correction factor $\alpha_{ij} \in [0, 1]$ and inserted into the right-hand side of (14). The flux-corrected equation for node i reads

$$m_i \frac{\mathrm{d}u_i}{\mathrm{d}t} = \sum_j l_{ij} u_j + \sum_{j \neq i} \alpha_{ij} f_{ij}. \tag{20}$$

By construction, the high-order system (12) is recovered for $\alpha_{ij} \equiv 1$ and the low-order one (14) for $\alpha_{ij} \equiv 0$. For accuracy reasons, the correction factors α_{ij} should approach 1 in regions where the solution is sufficiently smooth. At the same time, some artificial diffusion must be retained in the vicinity of steep gradients, where spurious undershoots and overhoots are likely to occur. The computation of α_{ij} for a FEM-FCT scheme is based on Zalesak's multidimensional flux limiter [18] to be presented in the next section.

4 Zalesak's FCT limiter

After the discretization in time, the flux-corrected end-of-step solution u^{n+1} should satisfy an algebraic system of the form $Au^{n+1} = b$, where A is supposed to possess the M-matrix property. In order to enforce the positivity constraint for the right-hand side b, it is sufficient to guarantee that

$$u^n \geq 0 \;\Rightarrow\; \tilde{u} \geq 0 \;\Rightarrow\; b \geq 0, \tag{21}$$

where $\tilde{u}$ denotes a positivity-preserving auxiliary solution and

$$b_i = m_i \tilde{u}_i + \Delta t \sum_{j \neq i} \alpha_{ij} f_{ij}. \tag{22}$$

The flux correction process may start with an optional 'prelimiting' step as described in [3,11,15]. This adjustment is intended to prevent an antidiffusive flux from flattening the solution profile. If it turns out that

$$f_{ij}(\tilde{u}_j - \tilde{u}_i) > 0 \tag{23}$$

then f_{ij} is 'diffusive' and needs to be canceled. Prelimiting improves the resolution of steep gradients but it may also result in artificial steepening.

In the worst case, all antidiffusive fluxes into node i have the same sign. Hence, it is worthwhile to treat the positive and negative ones separately using the fully multidimensional FCT algorithm proposed by Zalesak [18]

(1) Compute the sums of positive/negative antidiffusive fluxes into node i

$$P_i^+ = \sum_{j\neq i} \max\{0, f_{ij}\}, \qquad P_i^- = \sum_{j\neq i} \min\{0, f_{ij}\}. \tag{24}$$

(2) Compute the distance to a local extremum of the auxiliary solution

$$Q_i^+ = \max\{0, \max_{j\neq i}(\tilde{u}_j - \tilde{u}_i)\}, \qquad Q_i^- = \min\{0, \min_{j\neq i}(\tilde{u}_j - \tilde{u}_i)\}. \tag{25}$$

(3) Compute the nodal correction factors that enforce positivity for node i

$$R_i^+ = \min\left\{1, \frac{m_i Q_i^+}{\Delta t P_i^+}\right\}, \qquad R_i^- = \min\left\{1, \frac{m_i Q_i^-}{\Delta t P_i^-}\right\}. \tag{26}$$

(4) Limit the fluxes f_{ij} and f_{ji} using the minimum of $R_i^\pm$ and $R_j^\mp$

$$\alpha_{ij} = \begin{cases} \min\{R_i^+, R_j^-\}, & \text{if } f_{ij} > 0, \\ \min\{R_i^-, R_j^+\}, & \text{otherwise.} \end{cases} \tag{27}$$

This limiting strategy guarantees that the right-hand side (22) is bounded by

$$\tilde{u}_i^{\min} = \tilde{u}_i + Q_i^- \leq b_i/m_i \leq \tilde{u}_i + Q_i^+ = \tilde{u}_i^{\max} \tag{28}$$

and the scheme is positivity-preserving due to the M-matrix property of A.

Note that the nodal correction factors $R_i^\pm$ as defined in (26) are inversely proportional to Δt. Hence, a larger portion of the raw antidiffusive flux f_{ij} is retained as the time step is refined. This is why FCT works so well for strongly time-dependent problems. On the other hand, the use of large time steps results in a loss of accuracy and severe convergence problems are observed in the steady state limit. For a detailed positivity proof and further information regarding the properties of Zalesak's limiter we refer to [9–11,17].

5 Linearization of antidiffusive fluxes

One of the main difficulties concerning the design of high-resolution finite element schemes for time-dependent problems is the treatment of the consis-

tent mass matrix. After the discretization in time by the $\theta-$scheme, the time derivative that appears in (18) and (19) is replaced by the finite difference

$$(M_L - M_C)\frac{u^{n+1} - u^n}{\Delta t} \approx (M_L - M_C)\frac{\mathrm{d}u}{\mathrm{d}t}. \tag{29}$$

The corresponding raw antidiffusive fluxes f_{ij} and the correction factors α_{ij} depend on the unknown solution u^{n+1} and must be updated in an iterative way. This can be accomplished, e.g., by means of a simple defect correction scheme but the convergence rates typically leave a lot to be desired. A discrete Newton method [17] converges faster but the costly assembly of the (approximate) Jacobian operator is unlikely to pay off for truly transient problems which call for the use of small time steps and, in many cases, explicit algorithms.

In order to avoid a high overhead cost due to iterative flux correction, it is possible to evaluate the fluxes f_{ij} using the high-order solution $\bar{u}^{n+1}$

$$\begin{aligned} f_{ij}^{n+\theta} = &[m_{ij}/\Delta t + \theta d_{ij}]\ (\bar{u}_i^{n+1} - \bar{u}_j^{n+1}) \\ &- [m_{ij}/\Delta t - (1-\theta)d_{ij}]\ (u_i^n - u_j^n). \end{aligned} \tag{30}$$

This idea leads to a linearized FEM-FCT scheme of the form [9,17]

(1) Compute the high-order solution $\bar{u}^{n+1}$ from the algebraic system

$$[M_C - \theta\Delta t K]\bar{u}^{n+1} = [M_C + (1-\theta)\Delta t K]u^n. \tag{31}$$

(2) Compute the intermediate solution $u^{n+1-\theta}$ by the low-order scheme

$$u^{n+1-\theta} = u^n + (1-\theta)\Delta t M_L^{-1} L u^n. \tag{32}$$

(3) Perform flux correction using $\tilde{u} = u^{n+1-\theta}$ and $f_{ij}^{n+\theta}$ given by (30)

$$f_{ij}^* = \alpha_{ij}^{n+\theta} f_{ij}^{n+\theta}. \tag{33}$$

(4) Apply the limited antidiffusive fluxes f_{ij}^* to the intermediate solution

$$u_i^* = u_i^{n+1-\theta} + \frac{\Delta t}{m_i}\sum_{j\neq i} f_{ij}^*. \tag{34}$$

(5) Solve the linear system for the end-of-step solution

$$[M_L - \theta\Delta t L]\, u^{n+1} = M_L u^*. \tag{35}$$

This algorithm is positivity-preserving under the CFL-like condition (11) with $c_{ii} = l_{ii}/m_i$. If the governing equation is linear, then two algebraic systems per

time step need to be solved for $\theta > 0$. Since $A = M_L - \theta\Delta t L$ was designed to be an M-matrix, the final solution update (35) is positivity-preserving and the convergence of iterative solvers is very fast. However, the computation of the high-order predictor $\bar{u}^{n+1}$ from (31) is usually more expensive (or even impossible) due to unfavorable matrix properties. Moreover, the final solution may exhibit spurious ripples if the prelimiting step is omitted.

As a slightly less accurate but much more robust and efficient alternative, linearization can be performed about an intermediate solution $u^{n+1/2}$ computed explicitly or implicitly by the positivity-preserving low-order scheme. In what follows, we adopt the following definition of the raw antidiffusive flux

$$f_{ij}^{n+1/2} = m_{ij}(\dot{u}_i^{n+1/2} - \dot{u}_j^{n+1/2}) + d_{ij}(u_i^{n+1/2} - u_j^{n+1/2}), \tag{36}$$

where $\dot{u}^{n+1/2} \approx \frac{u^{n+1}-u^n}{\Delta t}$ can be defined as $\dot{u}^{n+1/2} = M_L^{-1} L u^{n+1/2}$ or simply

$$\dot{u}^{n+1/2} = 2\frac{u^{n+1/2} - u^n}{\Delta t}. \tag{37}$$

Note that the approximation of the time derivative is second-order accurate with respect to the time level $t^{n+1/2}$ at which the flux $f_{ij}^{n+1/2}$ is defined.

6 Runge-Kutta FCT scheme

Since the forward Euler method is known to be unstable, explicit FEM-FCT algorithms are usually designed on the basis of Lax-Wendroff/Taylor-Galerkin schemes [14]. This approach works well in practice but no proof of positivity seems to be available for the computationally efficient two-step version (Richtmyer scheme). Therefore, it is worthwhile to perform the discretization in time by a positivity-preserving TVD Runge-Kutta method as proposed in [4].

The following algorithm is based on the optimal second-order RK scheme

(1) Predict the provisional end-of-step solution

$$\bar{u}^{n+1} = u^n + \Delta t M_L^{-1} L u^n. \tag{38}$$

(2) Compute the intermediate solution of low order

$$u^{n+1/2} = \frac{\bar{u}^{n+1} + u^n}{2}. \tag{39}$$

(3) Correct the provisional end-of-step solution

$$\tilde{u}^{n+1} = u^{n+1/2} + \frac{\Delta t}{2} M_L^{-1} L \bar{u}^{n+1}. \tag{40}$$

(4) Perform flux correction using $\tilde{u} = \tilde{u}^{n+1}$ and $f_{ij}^{n+1/2}$ given by (36)–(37)

$$f_{ij}^{*} = \alpha_{ij}^{n+1/2} f_{ij}^{n+1/2}. \tag{41}$$

(5) Apply the limited antidiffusive fluxes f_{ij}^{*} to the end-of-step solution

$$u_i^{n+1} = \tilde{u}_i^{n+1} + \frac{\Delta t}{m_i} \sum_{j \neq i} f_{ij}^{*}. \tag{42}$$

It can readily be seen that this FEM-FCT algorithm satisfies the positivity constraint under the CFL-like condition (11) with $c_{ii} = l_{ii}/m_i$ and $\theta = 0$.

7 Crank-Nicolson FCT scheme

The semi-implicit Crank-Nicolson discretization yields an algebraic system

$$\left[M_L - \frac{\Delta t}{2} L \right] u^{n+1} = \left[M_L + \frac{\Delta t}{2} L \right] u^n + \Delta t f^{*}, \tag{43}$$

where f^{*} denotes the sum of limited antidiffusive fluxes. A fractional-step approach to solution of (43) leads to the following FEM-FCT algorithm:

(1) Compute the intermediate solution of low order

$$u^{n+1/2} = u^n + \frac{\Delta t}{2} M_L^{-1} L u^n. \tag{44}$$

(2) Perform flux correction using $\tilde{u} = u^{n+1/2}$ and $f_{ij}^{n+1/2}$ given by (36)–(37)

$$f_{ij}^{*} = \alpha_{ij}^{n+1/2} f_{ij}^{n+1/2}. \tag{45}$$

(3) Apply the limited antidiffusive fluxes f_{ij}^{*} to the intermediate solution

$$u_i^{*} = u_i^{n+1/2} + \frac{\Delta t}{m_i} \sum_{j \neq i} f_{ij}^{*}. \tag{46}$$

(4) Solve the linear system for the end-of-step solution

$$\left[M_L - \frac{\Delta t}{2} L \right] u^{n+1} = M_L u^{*}. \tag{47}$$

The largest admissible time step is given by (11) with $c_{ii} = l_{ii}/m_i$ and $\theta = \frac{1}{2}$.

8 Backward Euler FCT scheme

The last FEM-FCT algorithm to be presented is based on the backward Euler method, whereby the intermediate solution $u^{n+1/2}$ is computed implicitly:

(1) Solve the linear system for the intermediate solution

$$\left[M_L - \frac{\Delta t}{2} L\right] u^{n+1/2} = M_L u^n. \tag{48}$$

(2) Perform flux correction using $\tilde{u} = u^{n+1/2}$ and $f_{ij}^{n+1/2}$ given by (36)–(37)

$$f_{ij}^* = \alpha_{ij}^{n+1/2} f_{ij}^{n+1/2}. \tag{49}$$

(3) Apply the limited antidiffusive fluxes f_{ij}^* to the intermediate solution

$$u_i^* = u_i^{n+1/2} + \frac{\Delta t}{m_i} \sum_{j \neq i} f_{ij}^*. \tag{50}$$

(4) Solve the linear system for the end-of-step solution

$$\left[M_L - \frac{\Delta t}{2} L\right] u^{n+1} = M_L u^*. \tag{51}$$

On one hand, the fully implicit backward Euler method corresponds to first-order upwinding in time, which makes it overly diffusive at large time steps. On the other hand, it is unconditionally stable and every solution update is guaranteed to be positivity-preserving for arbitrary time steps.

9 Numerical examples

In this section, we apply our explicit and implicit FEM-FCT algorithms to solid body rotation in 2D as well as to the shock tube problem of gas dynamics. These well-documented test problems are chosen because they admit analytical solutions that can be used to assess the quality of numerical results.

9.1 Solid body rotation

A properly designed numerical algorithm should be capable of resolving both smooth and discontinuous profiles without excessive smoothing or steepening. To test the resolving power of our linearized FEM-FCT schemes, we consider a 2D benchmark problem proposed by LeVeque [13]. Figure 1 displays the

initial data that comprise a slotted cylinder, a sharp cone, and a smooth hump. The incompressible velocity field $\mathbf{v} = (0.5 - y, x - 0.5)$ corresponds to a counterclockwise rotation about the center of the computational domain $\Omega = (0,1) \times (0,1)$. After one full revolution ($t = 2\pi$) the exact solution of equation (1) coincides with the initial data. Thus, the main challenge of this test is to preserve the shape of the rotating bodies as far as possible.

The numerical solutions presented in Fig. 2–4 were computed on a uniform mesh of 128×128 bilinear finite elements using the second-order accurate Crank-Nicolson time-stepping with $\Delta t = 10^{-3}$. Algorithm (31)–(35) produces the most accurate results (see Fig. 2) but the iterative solver for the high-order system (31) converges very slowly even for such a small time step [17]. Moreover, raw antidiffusive fluxes of 'wrong' (according to (23)) sign must be canceled to prevent a strong distortion of the solution profiles. For an in-depth numerical study of this linearized FCT algorithm we refer to [17].

The use of an explicit low-order predictor $u^{n+1/2}$ eliminates the need for solving an ill-conditioned linear system, and the dubious prelimiting step can be safely omitted. The numerical solution computed by algorithm (44)–(47) is comparable to that produced by the upwind-biased MC limiter of TVD type, see Fig. 3–4. It is worth mentioning that implicit TVD schemes give rise to a nonlinear algebraic system which must be solved iteratively. Moreover, only the converged solution is guaranteed to be positivity-preserving. The new linearization strategy leads to a simple and efficient FCT algorithm, which justifies the noticeable loss of accuracy as compared to Fig. 2. In our experience, it is also possible to evaluate $f_{ij}^{n+1/2}$ and $\dot{u}^{n+1/2}$ using a higher-order approximation to $u^{n+1/2}$. However, this inevitably results in a higher computational cost and, typically, reintroduces the need for prelimiting.

To quantify the difference between the exact solution u and its numerical approximations u_h, we define the discrete error norms

$$E_1 = \sum_i m_i |u(x_i, y_i) - u_i| \approx \int_\Omega |u - u_h| \, \mathrm{d}x = ||u - u_h||_1, \tag{52}$$

$$E_2 = \sqrt{\sum_i m_i |u(x_i, y_i) - u_i|^2} \approx \sqrt{\int_\Omega |u - u_h|^2 \, \mathrm{d}x} = ||u - u_h||_2, \tag{53}$$

where m_i are the diagonal coefficients of the lumped mass matrix M_L.

The values of E_1 and E_2 for the FEM-FCT schemes under investigation are presented in Figs. 2–4 and Table 1. For time steps as small as $\Delta t = 10^{-3}$ all of the algorithms produce comparable results. As the time step is increased to $\Delta t = 10^{-2}$, the first order accuracy of the backward Euler (BE-FCT) scheme results in errors that are considerably larger than those for the second-order

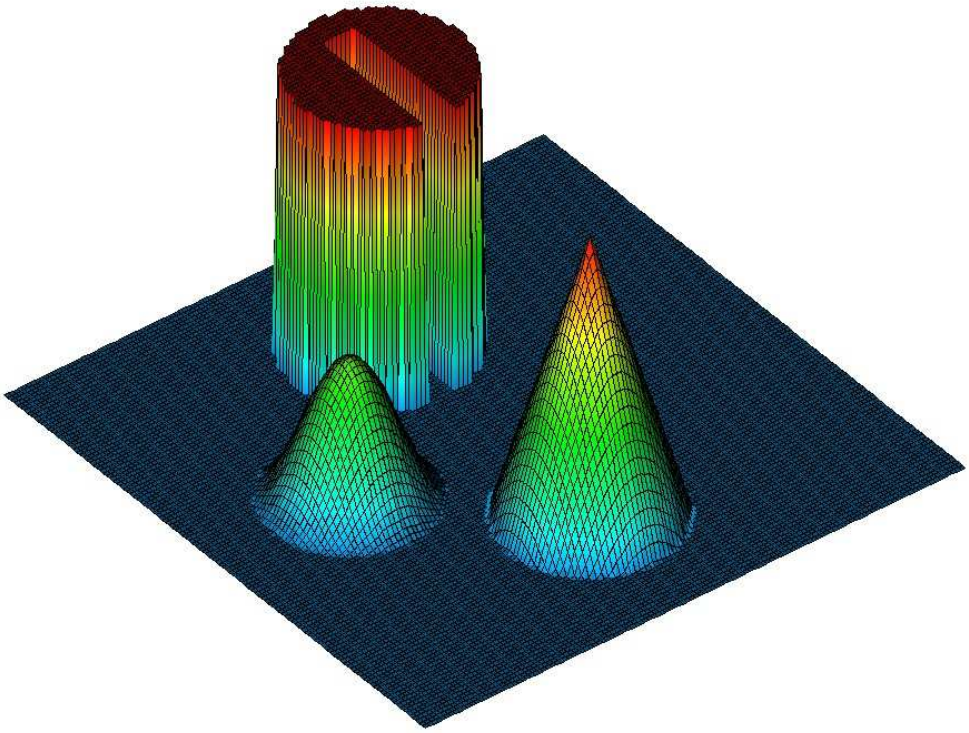

Fig. 1. Solid body rotation: initial data / exact solution.

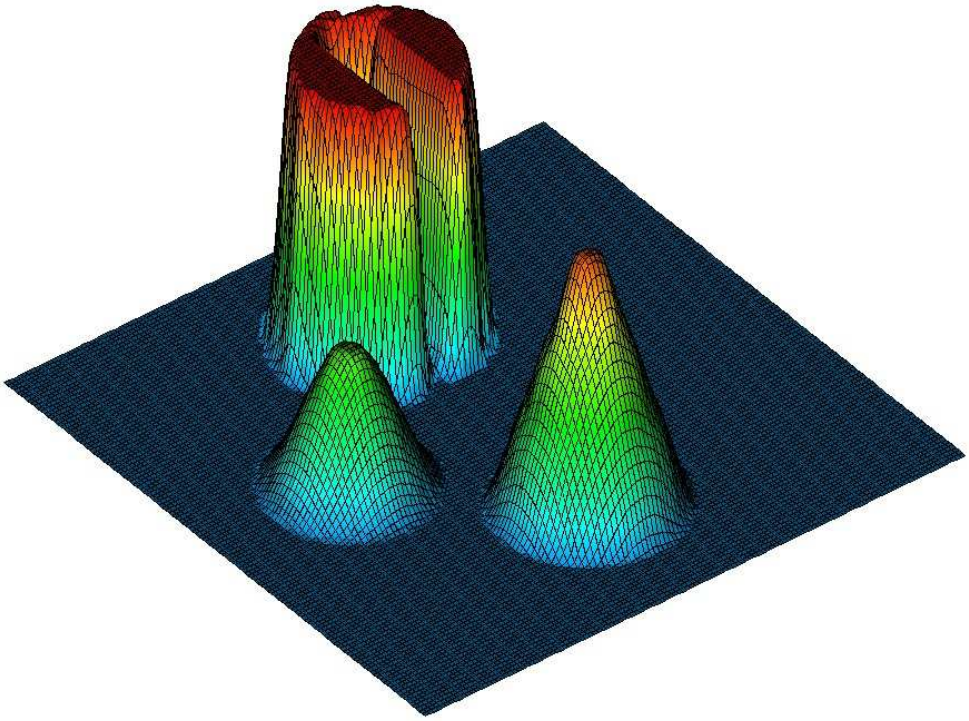

Fig. 2. CN-FCT linearized about $\bar{u}^{n+1}$, $E_1 = 0.0110,\ E_2 = 0.0574$.

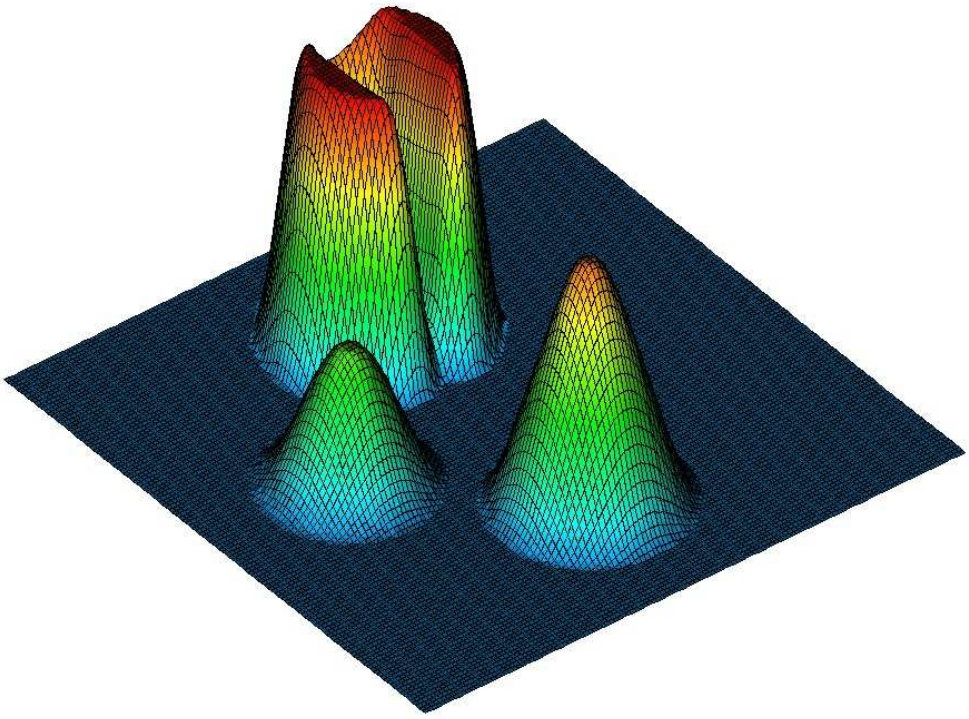

Fig. 3. CN-FCT linearized about $u^{n+1/2}$, $E_1 = 0.0227,\ E_2 = 0.0851$.

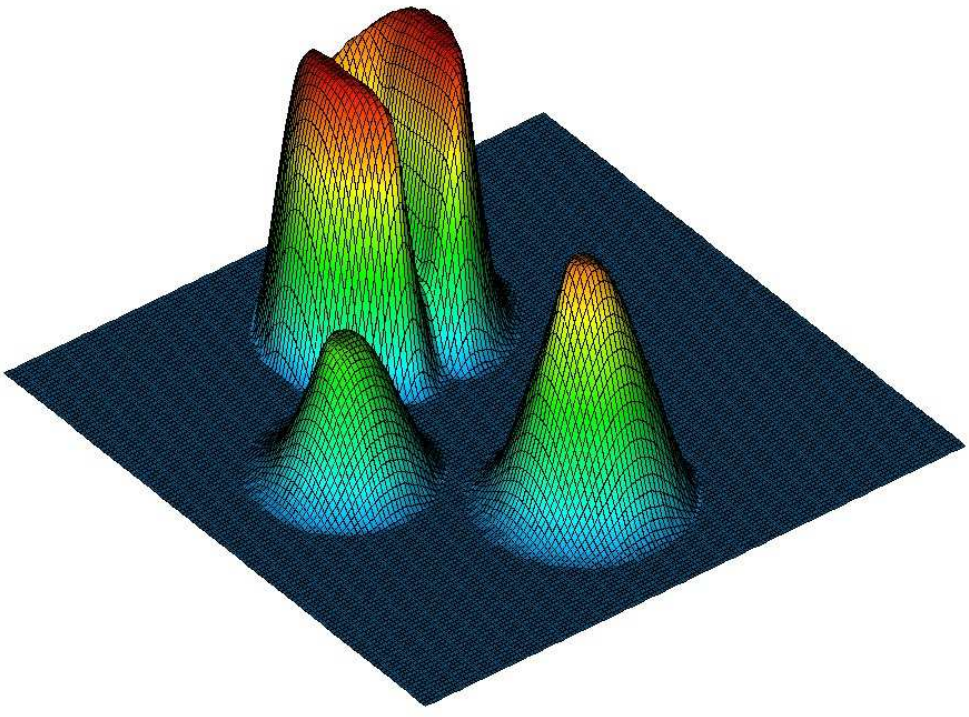

Fig. 4. CN-TVD with MC limiter, $E_1 = 0.0258,\ E_2 = 0.0894$.

accurate Runge-Kutta (RK-FCT) and Crank-Nicolson (CN-FCT) algorithms. A further increase of the time step by a factor of 10 leads to a violation of the CFL stability condition for the fully explicit RK-FCT scheme, whereas CN-FCT and BE-FCT remain stable, although the accuracy of the resulting numerical solutions is very poor. In this example, the choice of the time step is dictated by accuracy considerations but if the CFL condition is violated only in small subdomains then a local loss of accuracy can be tolerated and the use of an unconditionally stable implicit time-stepping becomes attractive.

Table 1
Solid body rotation: discrete error norms for the linearized FEM-FCT schemes.

	$\Delta t = 10^{-3}$		$\Delta t = 10^{-2}$		$\Delta t = 10^{-1}$	
	E_1	E_2	E_1	E_2	E_1	E_2
RK-FCT	2.1646e-2	8.2602e-2	2.4055e-2	8.9201e-2	∞	∞
CN-FCT	2.1793e-2	8.2790e-2	2.2744e-2	8.5166e-2	8.2158e-2	1.7087e-1
BE-FCT	2.4689e-2	8.8203e-2	4.5507e-2	1.2202e-1	9.6400e-2	1.8907e-1

9.2 Shock tube problem

The presented FEM-FCT algorithms are applicable not only to linear convection problems but also to hyperbolic systems such as the Euler equations

$$\begin{aligned} &\tfrac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \\ &\tfrac{\partial (\rho \mathbf{v})}{\partial t} + \nabla \cdot (\rho \mathbf{v} \otimes \mathbf{v}) + \nabla p = 0, \\ &\tfrac{\partial (\rho E)}{\partial t} + \nabla \cdot (\rho E \mathbf{v}) + \nabla \cdot (p \mathbf{v}) = 0. \end{aligned} \tag{54}$$

The high-order discretization of (54) can be performed using the group finite element formulation which leads to a semi-discrete system of the form [12]

$$\left[M_C \frac{\mathrm{d}\mathrm{U}}{\mathrm{d}t} \right]_i = \sum_{j \neq i} \mathrm{A}_{ij} (\mathrm{U}_j - \mathrm{U}_i), \tag{55}$$

where $\mathrm{U}_i = [\rho_i, (\rho \mathbf{v})_i, \rho E \mathbf{v})_i]$ is the vector of nodal values. Due to the hyperbolicity of the Euler equations, the Roe matrices A_{ij} are diagonalizable and have real eigenvalues. Therefore, algebraic flux correction can be performed in terms of local characteristic variables, as explained in [8,12,19]. Flux limiting in terms of the conservative or primitive variables is also feasible but requires a proper synchronization of the correction factors [14,15].

Fig. 5 illustrates the performance of the characteristic CN-FCT scheme as applied to Sod's shock tube problem. The initial conditions for the one-dimensional Riemann problem to be solved are given by

$$\begin{bmatrix} \rho_L \\ v_L \\ p_L \end{bmatrix} = \begin{bmatrix} 1.0 \\ 0.0 \\ 1.0 \end{bmatrix}, \quad \forall x \in [0, 0.5], \qquad \begin{bmatrix} \rho_R \\ v_R \\ p_R \end{bmatrix} = \begin{bmatrix} 0.125 \\ 0.0 \\ 0.1 \end{bmatrix}, \quad \forall x \in (0.5, 1]. \tag{56}$$

The analytical solution at $t = 0.231$ and the low-order approximation are depicted by the dashed and dotted lines, respectively. The numerical solutions are computed using 100 linear finite elements and the time step $\Delta t = 10^{-3}$. Since the contribution of the consistent mass matrix is taken into account, the CN-FCT scheme proves more accurate than the lumped-mass version based on an upwind-biased TVD limiter (see Fig. 5). Due to the nonlinearity of the Euler equations, subproblem (47) for the final solution must be solved iteratively. However, a single defect correction step with a diagonal preconditioner turns out to be enough for $\Delta t = 10^{-3}$. In this particular case, the RK-FCT and CN-FCT schemes produce essentially the same results at the same cost.

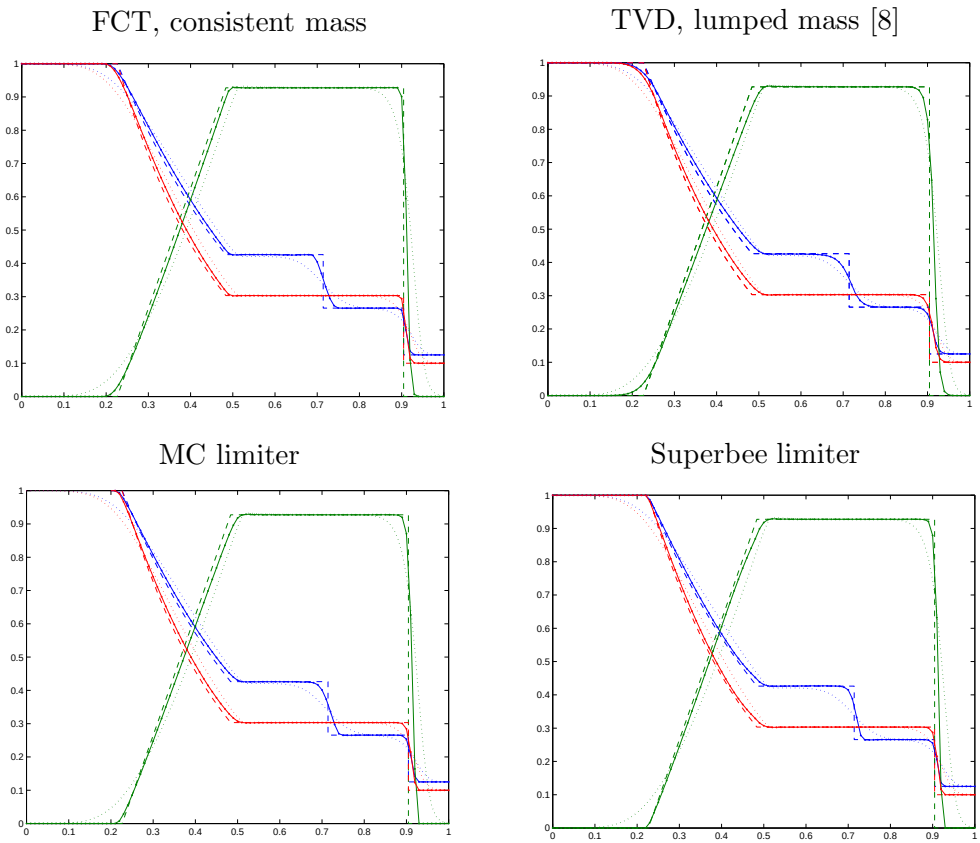

Fig. 5. Shock tube: 100 linear elements, $\Delta t = 10^{-3}$, solutions at $t = 0.231$.

10 Conclusions

Flux correction of FCT type is a useful tool for the design of high-resolution finite element schemes. The amount of artificial diffusion reduces as the time step is refined, and the consistent mass matrix can be included in a positivity-preserving fashion. A linearization of the raw antidiffusive flux about an intermediate solution of low order eliminates the need for iterative flux correction and *ad hoc* prelimiting. The resulting FEM-FCT schemes based on the Runge-Kutta, Crank-Nicolson, and backward Euler time stepping are more robust and efficient than their predecessors proposed in [9–11]. Since flux correction needs to be performed just once per time step, the use of an implicit time-stepping method is likely to pay off, for example, in the case of a strongly varying velocity field and/or a locally refined unstructured mesh. For sufficiently small time steps, the associated linear systems can be solved using a simple Richardson iteration preconditioned by the lumped mass matrix.

As always, the optimal choice of the time-stepping method and of the iterative solver is highly problem-dependent. The availability of both explicit and implicit FEM-FCT schemes that fit into a common framework makes it possible to adapt the solution strategy and the parameter settings in the course of simulation so as to capture the evolution details at the least possible cost. Mesh adaptation makes it possible to achieve a further gain of accuracy, whereby a usable error indicator can be constructed using the correction factors α_{ij} or slope-limited gradient recovery [16]. The combination of local mesh refinement with flux correction can be interpreted as a sort of $h-p$ adaptivity because the order of approximation is tailored to the local solution behavior.

References

[1] J. P. Boris, D. L. Book, Flux-corrected transport. I. SHASTA, A fluid transport algorithm that works, J. Comput. Phys. 11 (1973) 38–69.

[2] D. L. Book, The conception, gestation, birth, and infancy of FCT, in: D. Kuzmin, R. Löhner, S. Turek (Eds.), Flux-Corrected Transport: Principles, Algorithms, and Applications, Springer, Berlin, 2005, pp. 5–28.

[3] C. R. DeVore, An improved limiter for multidimensional flux-corrected transport, NASA Technical Report, AD-A360122 (1998).

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

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

[6] T. Jongen, Y. P. Marx, Design of an unconditionally stable, positive scheme for the $k - \varepsilon$ and two-layer turbulence models, Comput. Fluids 26 (5) (1997) 469–487.

[7] D. Kuzmin, On the design of general-purpose flux limiters for implicit FEM with a consistent mass matrix. I. Scalar convection, J. Comput. Phys. 219 (2006) 513–531.

[8] D. Kuzmin, Algebraic flux correction for finite element discretizations of coupled systems, in: E. Oñate, M. Papadrakakis, B. Schrefler (Eds.), Computational Methods for Coupled Problems in Science and Engineering II, CIMNE, Barcelona, 2007, pp. 653–656.

[9] D. Kuzmin, S. Turek, Flux correction tools for finite elements, J. Comput. Phys. 175 (2002) 525–558.

[10] D. Kuzmin, M. Möller, S. Turek, High-resolution FEM-FCT schemes for multidimensional conservation laws, Comp. Meth. Appl. Mech. Eng. 193 (2004) 4915–4946.

[11] D. Kuzmin, M. Möller, Algebraic flux correction I. Scalar conservation laws, in: D. Kuzmin, R. Löhner, S. Turek (Eds.), Flux-Corrected Transport: Principles, Algorithms, and Applications, Springer, Berlin, 2005, pp. 155–206.

[12] D. Kuzmin, M. Möller, Algebraic flux correction II. Compressible Euler equations, in: D. Kuzmin, R. Löhner, S. Turek (Eds.), Flux-Corrected Transport: Principles, Algorithms, and Applications, Springer, Berlin, 2005, pp. 207–250.

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

[14] 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.

[15] R. Löhner, J. D. Baum, 30 Years of FCT: Status and Directions, in: D. Kuzmin, R. Löhner, S. Turek (Eds.), Flux-Corrected Transport: Principles, Algorithms, and Applications, Springer, Berlin, 2005, pp. 131–154.

[16] M. Möller, D. Kuzmin, Adaptive mesh refinement for high-resolution finite element schemes, Int. J. Numer. Meth. Fluids, 52 (2006) 545–569.

[17] M. Möller, D. Kuzmin, D. Kourounis, Implicit FEM-FCT algorithms and discrete Newton methods for transient convection problems, Int. J. Numer. Meth. Fluids, in press.

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

[19] 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, Berlin, 2005, pp. 29–78.