ruhr.paD

A New Approach to Enforcing Discrete Maximum Principles in Continuous Galerkin Methods for Convection-Dominated Transport Equations

D. Kuzmin and J. N. Shadid

A new approach to enforcing discrete maximum principles in continuous Galerkin methods for convection-dominated transport equations

Dmitri Kuzmin[a], John N. Shadid[b,c]

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

[b]*Computational Mathematics Department, Sandia National Laboratories P.O. Box 5800 MS 1321, Albuquerque, NM 87185-1321, USA*

[c]*Department of Mathematics and Statistics, University of New Mexico MSC01 1115, Albuquerque, NM 87131, USA*

Abstract

This paper presents a set of design principles and new algorithmic tools for enforcing maximum principles and/or positivity preservation in continuous finite element approximations to convection-dominated transport problems. Using a linear first-order advection equation as a model problem, we address the design of first-order artificial diffusion operators and their higher-order counterparts at the element matrix level. The proposed methodology leads to a nonlinear high-resolution scheme capable of resolving moving fronts and internal/boundary layers as sharp localized nonoscillatory features. The amount of numerical dissipation depends on the difference between the solution value at a given node and a local maximum or minimum. The shock-capturing numerical diffusion coefficient is designed to vanish as the nodal values approach a mass-weighted or linearity-preserving average. The universal applicability and simplicity of the element-based limiting procedure makes it an attractive alternative to edge-based algebraic flux correction.

Email addresses: `kuzmin@math.math-dortmund.de` (Dmitri Kuzmin), `jnshadi@sandia.gov` (John N. Shadid)

Additionally a Lipschitz continuous version of the limiter is presented that guarantees unique solvability of the nonlinear system associated with the steady state limit of the time-dependent problem. A grid convergence study is performed for two-dimensional test problems.

Keywords: convective transport, finite elements, discrete maximum principles, artificial diffusion, limiters, algebraic flux correction

1. Introduction

The Galerkin finite element discretization of convection-dominated transport equations is known to produce numerical approximations that may violate the discrete maximum principle and/or the criterion of positivity preservation on meshes that are too coarse to resolve certain fine-scale features (moving fronts, interior and boundary layers). The most common approach to avoiding nonphysical undershoots and overshoots in finite element methods is based on the use of nonlinear shock-capturing terms within the framework of variationally consistent Petrov-Galerkin methods (see, e.g., [10, 28, 29] for a review and comparative study of existing schemes). Additionally entropy viscosity approaches that attempt to control oscillations by introducing artificial dissipation that is based on an auxiliary entropy production residual have been proposed [22]. The main drawback of many existing approaches is the presence of problem-dependent free parameters along with the lack of provable nonlinear stability properties such as positivity and monotonicity preservation on general meshes. While small spurious oscillations can be tolerated in some applications, many models are very sensitive to nonphysical values of the transported variable. For this reason, the use of physics-compatible finite element approximations may be appropriate or even indispensable.

Nonlinear shock-capturing operators backed by the theory of discrete maximum principles (DMP) were recently developed and analyzed in [1, 2, 6, 7, 18]. In the case of [1] and [7], the proof of the DMP property imposes restrictions of sufficient mesh regularity. Moreover, mass lumping is required in applications to transient problems [1]. The explicit second-order method proposed in [18] satisfies a discrete maximum principle for arbitrary meshes and employs the consistent mass matrix. The way in which [18] enforces

2

DMP constraints is closely related to the concept of *algebraic flux correction* [32]. This approach provides a general framework for the design of artificial diffusion operators that render a finite element discretization *local extremum diminishing* (LED) or positivity-preserving. In nonlinear high-resolution schemes based on algebraic flux correction, the antidiffusive part of a high-order discretization is constrained using a conservative *limiter*.

The most prominent representative of algebraic flux correction schemes is the *flux-corrected transport* (FCT) algorithm introduced by Boris and Book [5] and Zalesak [47] in the context of explicit finite difference schemes. Unlike many other limiting techniques, FCT can be extended to finite element approximations using conservative decompositions of the antidiffusive term into internodal fluxes associated with edges of the sparsity graph [44, 11, 40, 35, 30] or element contributions associated with individual mesh cells [38, 39, 35]. The simplicity and efficiency of predictor-corrector FCT schemes make them very attractive in situations when the problem is time-dependent and the time steps are small [29]. Additionally, edge-based generalizations of *total variation diminishing* (TVD) schemes [20, 21] can be designed using reconstruction of 1D stencils [41, 42, 37] or algebraic flux correction schemes proposed in [31, 32]. In contrast to FCT, multidimensional extensions of iterative TVD limiters are directly applicable to stationary transport equations and produce steady-state solutions independent of the time step.

In this paper, new element-based limiting techniques for artificial diffusion operators constructed on the element matrix level are proposed. The main advantages of the presented methodology are computational simplicity, LED property on arbitrary meshes, low levels of numerical dissipation, and applicability to time-dependent problems without mass lumping.

After reviewing the design philosophy behind algebraic flux correction schemes, we consider local bilinear forms associated with low-order monotonicity-preserving artificial diffusion. Instead of constraining the sums of antidiffusive element contributions as in element-based FCT schemes, we define nodal correction factors so that the modified scheme is LED, and reduces to the underlying Galerkin discretization whenever the nodal value is close enough to a suitably defined average value to indicate the local function is nearly linear. To improve phase accuracy, an optional correction of the consistent mass matrix is performed using a local version of the element-based FCT al-

3

gorithm. A Lipschitz-continuous upwind limiter is developed for steady state computations. The numerical behavior of the proposed schemes is illustrated by a grid convergence study for two-dimensional transport problems.

2. Continuous problem

As a linear model problem, consider the time-dependent linear first-order advection equation

$$\frac{\partial u}{\partial t} + \nabla \cdot (\mathbf{v}u) = 0 \quad \text{in } \Omega, \tag{1}$$

where $u : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}$ is the conserved quantity, $\mathbf{v} : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}^d$ is a given velocity field, and Ω is a bounded domain in $\mathbb{R}^d$, $d \in \{1, 2, 3\}$.

The initial condition for the linear convection model is given by

$$u(\mathbf{x}, 0) = u_0(\mathbf{x}), \quad \mathbf{x} \in \Omega. \tag{2}$$

At the inlet $\Gamma_{\text{in}} := \{\mathbf{x} \in \Gamma \,|\, \mathbf{v} \cdot \mathbf{n} < 0\}$, where $\mathbf{n}$ is the unit outward normal to the boundary $\Gamma := \partial\Omega$, a Dirichlet boundary condition is prescribed

$$u = u_{\text{in}} \quad \text{on } \Gamma_{\text{in}}. \tag{3}$$

Let $\Sigma := \{(\mathbf{x}, t) \,|\, \mathbf{x} \in \Gamma_{\text{in}} \vee t = 0\}$ denote the set of points and time instants such that $u(\mathbf{x}, t)$ is known from the data prescribed in (2) or (3).

The solution to problem (1)–(3) is known to be positivity-preserving, i.e.,

$$u(\mathbf{x}, t) \geq 0 \quad \forall (\mathbf{x}, t) \in \Sigma \quad \Rightarrow \quad u(\mathbf{x}, t) \geq 0 \quad \forall (\mathbf{x}, t) \in \bar{\Omega} \times \mathbb{R}_+, \tag{4}$$

which can be easily shown using the method of characteristics.

Moreover, the following maximum principle holds in the case $\nabla \cdot \mathbf{v} = 0$:

$$\min_{\Sigma} u \leq u(\mathbf{x}, t) \leq \max_{\Sigma} u. \tag{5}$$

If $\nabla \cdot \mathbf{v} \neq 0$, then the solution to (1)–(2) will satisfy (4) but may violate (5).

4

3. Constrained approximations

Discretizing (1) in space, one obtains a sparse system of ordinary differential equations for the nodal values $u_i(t) \approx u(\mathbf{x}_i, t)$. For a number of discretizations the equation associated with an interior node $\mathbf{x}_i \in \Omega$ can be written as

$$m_i \frac{\mathrm{d}u_i}{\mathrm{d}t} = \sum_{j \in \mathcal{N}(i)} l_{ij} u_j + f_i(u), \tag{6}$$

where $\mathcal{N}(i) := \{i\} \cup \{j \,|\, l_{ij} \neq 0\}$ defines the local stencil of the numerical scheme and m_i is a positive entry of a diagonal mass matrix.

A well-designed discrete numerical approximation should preserve the important properties of the exact continuous solution. In algebraic flux correction schemes [32], the coefficients l_{ij} and $f_i(u)$ are defined to produce a solution that satisfies a (semi-)discrete maximum principle and/or the requirement of positivity preservation. Design criteria for these methods are often based on the following set of sufficient conditions.

3.1. Positivity preservation

Considering a system of ODEs in $\mathbb{R}^m$ resulting from spatial discretization for $t \geq 0$

$$\frac{\mathrm{d}v}{\mathrm{d}t} = F(v, t) \tag{7}$$

this system is termed positivity-preserving (or more accurately - nonnegativity preserving) if

$$v(0) \geq 0 \Rightarrow v(t) \geq 0 \text{ for all } t > 0, \tag{8}$$

where the inequalities are understood to hold for each component. The following result from [25] establishes necessary and sufficient conditions for positivity preservation (see Theorem 7.1 in [25] for details).

Suppose that F(v,t) is Lipschitz continuous with respect to v. Then the ODE system (7) is positivity-preserving iff for any vector $v \in \mathbb{R}^m$ and for all $i = 1, ..., m$, and $t \geq 0$,

$$v \geq 0, \quad v_i = 0 \Rightarrow F_i(v, t) \geq 0. \tag{9}$$

It is straightforward to verify that these conditions hold for the semi-discretized form (6) of the linear first-order advection equation (1) with the following assumptions

$$
\begin{aligned}
f_i &= 0 && \forall i, \\
l_{ij} &\geq 0 && \forall j \in \mathcal{N}(i)\backslash\{i\}, \\
m_i &> 0 && \forall i
\end{aligned}
$$

and by rewriting the semi-discretized system (6) as

$$
\frac{\mathrm{d}u_i}{\mathrm{d}t} = \frac{l_{ii}}{m_i}u_i + \sum_{j \in \mathcal{N}(i)\backslash\{i\}} \frac{l_{ij}}{m_i}u_j. \tag{10}
$$

From this result it follows that for $u \geq 0$ and $u_i = 0$ we have $F_i(u,t) \geq 0$ and that $F(u,t)$ is Lipschitz continuous with a Lipschitz constant $C = \|M^{-1}L\|$ where $M = \mathrm{diag}\{m_i\}$, and L is the discrete linear transport operator.

3.2. Local extremum diminishing schemes

In the case $f_i = 0$ and $\sum_{j \in \mathcal{N}(i)} l_{ij} = 0$, equation (6) reduces to

$$
m_i \frac{\mathrm{d}u_i}{\mathrm{d}t} = \sum_{j \in \mathcal{N}(i)\backslash\{i\}} l_{ij}(u_j - u_i). \tag{11}
$$

Under the assumption that $m_i > 0$ and $l_{ij} \geq 0$, we have

$$
u_i \geq u_j \quad \forall j \in \mathcal{N}(i)\backslash\{i\} \quad \Rightarrow \quad \frac{\mathrm{d}u_i}{\mathrm{d}t} \leq 0, \tag{12}
$$

$$
u_i \leq u_j \quad \forall j \in \mathcal{N}(i)\backslash\{i\} \quad \Rightarrow \quad \frac{\mathrm{d}u_i}{\mathrm{d}t} \geq 0. \tag{13}
$$

It follows that a local maximum cannot increase, and a local minimum cannot decrease. A space discretization satisfying this semi-discrete maximum principle is called *local extremum diminishing* (LED) [26, 27, 32]. It can be easily verified that LED implies positivity preservation, but the converse is not true.

3.3. LED criterion for antidiffusive correction terms

By the Godunov theorem [14], linear positivity-preserving and LED schemes can be at most first-order accurate. In higher-order schemes for transport equations, a nonvanishing term $f_i(u)$ incorporates an antidiffusive correction to a nonoscillatory coarse-scale approximation (see Section 3.4). This correction is LED if there exist some positive bounded coefficients $c_i^{\min}$ and $c_i^{\max}$ such that

$$c_i^{\min}(u_i^{\min} - u_i) \leq f_i(u) \leq c_i^{\max}(u_i^{\max} - u_i), \tag{14}$$

where

$$u_i^{\min} = \min_{j \in \mathcal{N}(i)} u_j, \tag{15}$$

$$u_i^{\max} = \max_{j \in \mathcal{N}(i)} u_j \tag{16}$$

are the local minimum and maximum over the stencil of node i.

It is easy to verify that properties (12),(13) hold if $f_i(u)$ satisfying (14) is added to the right-hand side of (11). Clearly if u_i is a local maximum, the right inequality in (14) implies $f_i(u) \leq 0$. In the case of a local minimum, the left inequality implies $f_i(u) \geq 0$. In general, a semi-discrete scheme of the form (6) will be positivity-preserving if $m_i > 0$, $l_{ij} \geq 0$ for $j \in \mathcal{N}(i)\backslash\{i\}$, and $f_i(u)$ satisfies the LED criterion (14) for some $c_i^{\min} > 0$ and $c_i^{\max} > 0$.

3.4. Constrained high-resolution schemes

In general a high-order scheme violating the LED constraints can be repaired by limiting the oscillatory part in a conservative manner. For example in edge-based algebraic flux correction schemes [44, 40, 37, 32, 18], the antidiffusive term is decomposed into skew-symmetric internodal fluxes f_{ij} such that $f_{ji} = -f_{ij}$ and

$$f_i(u) = \sum_{j \in \mathcal{N}(i)\backslash\{i\}} f_{ij}. \tag{17}$$

The limited LED counterpart of $f_i(u)$ is defined by

$$\bar{f}_i(u) = \sum_{j \in \mathcal{N}(i)\backslash\{i\}} \alpha_{ij} f_{ij}, \tag{18}$$

where $\alpha_{ij} \in [0, 1]$ are correction factors such that $\alpha_{ji} = \alpha_{ij}$ and

$$c_i^{\min}(u_i^{\min} - u_i) \leq \bar{f}_i(u) \leq c_i^{\max}(u_i^{\max} - u_i) \qquad (19)$$

for some bounded coefficients $c_i^{\min} > 0$ and $c_i^{\max} > 0$. To assure that the high-order approximation is recovered when the unconstrained antidiffusive correction is LED the correction procedure should be designed to guarantee that $\alpha_{ij} \approx 1$ whenever $f_i(u)$ satisfies (14).

In the context of traditional element-based data-structures a more flexible approach to constraining the antidiffusive part of a semi-discrete high-order method can be devised. This algorithm is based on a decomposition of the global vector $f = \{f_i\}$ into subvectors f^e associated with sets of neighboring nodes (e.g., vertices of the same mesh element). In Section 4.1, we will use this generalization to construct element-based limiting techniques for enforcing LED constraints in the context of continuous finite element discretizations.

3.5. Time step restrictions

For the fully discrete scheme to inherit the LED property of a given space discretization, the time-stepping method must be consistent with the discrete maximum principle, at least under certain time step restrictions. For example, consider (11) discretized in time using the two-level θ-scheme

$$m_i \frac{u_i^{n+1} - u_i^n}{\Delta t} = \theta \sum_{j \in \mathcal{N}(i) \backslash \{i\}} l_{ij}(u_j^{n+1} - u_i^{n+1}) + (1-\theta) \sum_{j \in \mathcal{N}(i) \backslash \{i\}} l_{ij}(u_j^n - u_i^n), \quad (20)$$

where $u_i^n \approx u(\mathbf{x}_i, t^n)$ denotes an approximate solution value at the time level $t^n = n\Delta t$ and $\theta \in [0, 1]$ is the implicitness parameter.

The solution of the fully discrete problem (20) satisfies a discrete maximum principle if u_i^{n+1} is bounded by the maximum and minimum of the other nodal values that appear in (20). For a LED space discretization, this will be the case if the time step Δt satisfies the CFL-like condition [35, 32]

$$\frac{1}{\Delta t} \geq (1 - \theta) \sum_{j \in \mathcal{N}(i) \backslash \{i\}} l_{ij} \qquad \forall i = 1, \dots, N. \qquad (21)$$

8

The fully implicit backward Euler method ($\theta = 1$) preserves the LED property for arbitrary time steps. The Crank-Nicolson ($\theta = \frac{1}{2}$) and forward Euler ($\theta = 0$) time discretizations are LED for time steps satisfying (21). In explicit *strong stability preserving* (SSP) Runge-Kutta methods [15, 16, 17], the end-of-step solution u^{n+1} represents a convex combination of forward Euler predictors and is LED under the corresponding time step restrictions. Implicit SSP Runge-Kutta methods are presented in the book by Gottlieb et al. [15].

4. Finite element methods

In this paper, we will use the LED criterion to analyze and constrain the antidiffusive part of a continuous (piecewise-linear or bilinear) finite element approximation. Multiplying (1) by a test function w, integrating by parts, and invoking the Dirichlet boundary condition (3), we obtain

$$\int_\Omega \left[w \frac{\partial u}{\partial t} - \nabla w \cdot (\mathbf{v}u) \right] \mathrm{d}\mathbf{x} + \int_{\Gamma_{\mathrm{out}}} wu(\mathbf{v} \cdot \mathbf{n}) \mathrm{d}s = - \int_{\Gamma_{\mathrm{in}}} wu_{\mathrm{in}}(\mathbf{v} \cdot \mathbf{n}) \mathrm{d}s, \quad (22)$$

where $\Gamma_{\mathrm{out}} := \{\mathbf{x} \in \Gamma \,|\, \mathbf{v} \cdot \mathbf{n} > 0\}$ is the outflow boundary. The so-defined variational formulation with weakly imposed boundary conditions is globally conservative since it reduces to the integral form of (1) in the case $w \equiv 1$.

Let $\{\varphi_1, \ldots, \varphi_j, \ldots, , \varphi_N\}$ be a set of global basis functions associated with vertices of the (possibly unstructured) mesh. The numerical solution is defined by

$$u_h(\mathbf{x}, t) = \sum_{j=1}^{N} u_j(t)\varphi_j(\mathbf{x}). \quad (23)$$

Substituting this approximation into the variational form (22) and using $w = \varphi_i$ as a test function, one obtains the Galerkin space discretization

$$\sum_{j \in \mathcal{N}(i)} m_{ij} \frac{\mathrm{d}u_j}{\mathrm{d}t} = \sum_{j \in \mathcal{N}(i)} k_{ij} u_j + g_i, \quad (24)$$

where

$$m_{ij} = \int_\Omega \varphi_i \varphi_j \, \mathrm{d}\mathbf{x},$$

9

$$k_{ij} = \int_{\Omega} (\mathbf{v} \cdot \nabla \varphi_i) \varphi_j \, \mathrm{d}\mathbf{x} - \int_{\Gamma_{\mathrm{out}}} \varphi_i \varphi_j \mathbf{v} \cdot \mathbf{n} \mathrm{d}s,$$

$$g_i = - \int_{\Gamma_{\mathrm{in}}} \varphi_i u_{\mathrm{in}} \mathbf{v} \cdot \mathbf{n} \mathrm{d}s.$$

The global matrix form of the semi-discrete finite element scheme is given by

$$M_C \frac{\mathrm{d}u}{\mathrm{d}t} = Ku + g, \tag{25}$$

where u is the vector of nodal values, M_C is the consistent mass matrix, K is the discrete convection operator and g is a vector of fluxes across Γ_{in}.

To cast equation (24) into the form (6), we introduce the lumped mass matrix

$$M_L = \mathrm{diag}\{m_i\}, \qquad m_i = \sum_{j \in \mathcal{N}(i)} m_{ij}$$

and consider an artificial diffusion operator $D = \{d_{ij}\}$ such that

$$d_{ii} = - \sum_{j \neq i} d_{ij}, \qquad d_{ij} = d_{ji}, \quad d_{ij} \geq \max\{0, -k_{ij}\} \quad \forall j \neq i.$$

Here and below we use "$j \neq i$" as a shorthand notation for "$j \in \mathcal{N}(i)\backslash\{i\}$".

Rearranging (25), we obtain the following equivalent representation

$$M_L \frac{\mathrm{d}u}{\mathrm{d}t} = Lu + f(u), \tag{26}$$

where $L = K + D$ is a stabilized discrete transport operator and

$$f(u) = (M_L - M_C) \frac{\mathrm{d}u}{\mathrm{d}t} - Du + g \tag{27}$$

is the antidiffusive term which may fail to satisfy the LED criterion (14).

4.1. Element-by-element assembly

In a typical implementation of the finite element method, the global matrices that appear in equations (25)–(27) are assembled from element matrices and the global vector $f(u)$ is assembled from element vectors of the form

$$f^e = (M_L^e - M_C^e) \frac{\mathrm{d}u^e}{\mathrm{d}t} - D^e u^e + g^e. \tag{28}$$

Since the element matrices D^e and $M_C^e - M_L^e$ have zero column sums, the components of the corresponding matrix-vector products sum to zero. Hence, the multiplication of these matrix-vector products by an arbitrary correction factor α_e has the property of being a conservative correction to the numerical solution.

Replacing (28) by the corrected vector of element contributions

$$\bar{f}^e = \min\{\alpha_e^M, \alpha_e^K\}(M_L^e - M_C^e)\frac{\mathrm{d}u^e}{\mathrm{d}t} - \alpha_e^K D^e u^e + g^e, \tag{29}$$

we will choose the correction factors $\alpha_e \in [0,1]$ so as to enforce the LED constraint (19) for the sums of element contributions to interior nodes.

4.2. Low-order dissipation

Setting all correction factors equal to zero, one obtains the modified system

$$M_L\frac{\mathrm{d}u}{\mathrm{d}t} = Lu + g, \tag{30}$$

where g is the vector of weakly imposed inflow boundary conditions.

The global lumped mass matrix M_L is assembled from lumped element mass matrices $M_L^e = \mathrm{diag}\{m_I^e\}$ and the discrete operator L is assembled from element matrices of the form $L^e = K^e + D^e$, where D^e is a local artificial diffusion operator such that the sufficient condition of positivity preservation ($l_{IJ}^e \geq 0$ for $J \neq I$) is satisfied for all pairs of local degrees of freedom (indexed by I, J).

4.2.1. Upwinding by artificial diffusion associated with pairs of nodes

By analogy with global algebraic flux correction schemes [32], we perform "discrete upwinding" using the artificial diffusion operator defined by

$$d_{IJ}^e = \begin{cases} \max\{-k_{IJ}^e, 0, -k_{JI}^e\} & \text{if } J \neq I, \\ -\sum_{J \neq I} d_{IJ}^e & \text{if } J = I. \end{cases} \tag{31}$$

This definition of d_{IJ}^e guarantees discrete conservation and nonnegativity of the off-diagonal coefficients $l_{IJ}^e := k_{IJ}^e + d_{IJ}^e$ for all $J \neq I$.

4.2.2. Upwinding by artificial diffusion associated with a bilinear form

Alternatively, the element matrix D^e may be constructed, e.g., using the monotone local bilinear form proposed by Guermond and Nazarov [18]

$$
b_e(\varphi_I, \varphi_J) = \begin{cases} -\frac{1}{n_e - 1}|K_e| & \text{if } J \neq I, \\ |K_e| & \text{if } J = I, \end{cases}
\tag{32}
$$

where n_e is the number of local degrees of freedom and $|K_e|$ is the d-volume of element K_e. The entries of the artificial diffusion operator D^e become

$$
d^e_{IJ} = -\nu_e b_e(\varphi_I, \varphi_J), \qquad \nu_e = \max_{J \neq I} \frac{\max\{0, -k^e_{IJ}\}}{|K_e|}.
\tag{33}
$$

4.2.3. Upwinding by artificial diffusion associated with mass lumping

The difference between the consistent and lumped mass matrices M_C and M_L is also often used to enforce positivity constraints in low-order finite element schemes [44, 11, 38, 6]. At the element matrix level, this approach leads to

$$
d^e_{IJ} = \nu_e(m^e_{IJ} - \delta_{IJ} m^e_I), \qquad \nu_e = \max_{J \neq I} \frac{\max\{0, -k^e_{IJ}\}}{m^e_{IJ}}.
\tag{34}
$$

We remark that definitions (33) and (34) are more diffusive than (31) but admit interpretations (anisotropic diffusivity, inexact quadrature) which may be useful for theoretical analysis and error estimation.

4.3. High-order dissipation

The fully explicit ($\theta = 0$) Galerkin discretization of a time-dependent transport equation is unconditionally unstable and must be stabilized by incorporating high-order dissipation into the vector of antidiffusive element contributions f^e. Additionally while implicit Galerkin schemes are unconditionally stable for $\theta \geq \frac{1}{2}$, they tend to produce global oscillations which degrade the rate of convergence to smooth solutions. Even if a violation of the discrete maximum principle is ruled out by the limiter for f^e, significant phase errors may result in spurious distortions of smooth profiles. Hence, the use of high-order background dissipation may be appropriate in this context.

To improve the phase accuracy of the constrained Galerkin approximation, an additional contribution of high-order dissipation can be added to $\bar{f}^e$ defined by (29). We have

$$(D^e u^e)_I = \sum_{J \neq I} d^e_{IJ}(u^e_J - u^e_I).$$

A smoother approximation to the solution difference $u^e_J - u^e_I$ can be constructed using the reconstructed nodal gradient [31, 32]

$$(\nabla u)_i \approx \frac{1}{m_i} \sum_{j \neq i} \mathbf{c}_{ij}(u_j - u_i), \qquad \mathbf{c}_{ij} = \int_\Omega \varphi_i \nabla \varphi_j \, d\mathbf{x}.$$

This kind of gradient recovery corresponds to a lumped-mass L^2 projection. It satisfies the discrete maximum principle and is exact for linear functions [31, 32].

Approximating $u^e_J - u^e_I$ by the arithmetic mean of linear reconstructions

$$\delta u^e_{IJ} = \frac{(\nabla u)_I + (\nabla u)_J}{2} \cdot (\mathbf{x}_J - \mathbf{x}_I),$$

we replace (29) by

$$\bar{f}^e = \min\{\alpha^M_e, \alpha^K_e\}(M^e_C - M^e_L)\frac{du^e}{dt} + \alpha^K_e(f^e_{\mathrm{stab}} - D^e u^e) + g^e, \qquad (35)$$

where $f^e_{\mathrm{stab},I}$ is an element vector with components

$$f^e_{\mathrm{stab},I} = \omega \sum_{J \neq I} d^e_{IJ}\left(u^e_J - u^e_I - \delta u^e_{IJ}\right).$$

The amount of background dissipation is controlled using the blending factor $\omega \in [0, 1]$. In the case $\omega = 0$, the standard Galerkin scheme is recovered by setting $\alpha^M_e = \alpha^K_e = 1$ in the definition (35) of $\bar{f}^e$. Setting $\omega = 1$ corresponds to replacing $u^e_J - u^e_I$ by the smooth approximation δu^e_{IJ}.

If the numerical solution u_h is locally linear, then the nodal gradients are exact, whence $\delta u_{IJ} = u^e_J - u^e_I$ for any value of the parameter ω. Hence, the stabilization term $f^e_{\mathrm{stab},I}$ vanishes for linear functions. On a uniform mesh of 1D linear finite elements, it introduces fourth-order artificial dissipation [37].

13

5. Limiting strategy

In this section, we present algorithms for calculating the correction factors α_e^M and α_e^K for the antidiffusive components of the element vector $\bar{f}^e$. The proposed limiting strategy sets α_e^K equal to zero if a local extremum is attained at any vertex of the element. A continuous piecewise-linear function is used to provide a transition to $\alpha_e^K = 1$ in regions where the discrete nodal solution approaches a local linear function. The formula for α_e^M is designed to prevent the contribution of the consistent mass matrix from creating new local extrema in the discretized time derivatives.

5.1. Constrained transport operator

To constrain the high-order antidiffusive correction we seek a nodal correction factor $\Phi_i \in [0, 1]$ such that the LED constraint (19) holds for the sum of limited antidiffusive element contributions

$$\bar{f}_i = \sum_{e \in \mathcal{E}(i)} f_i^e,$$

where $\mathcal{E}(i)$ is the set of elements containing node i. To enforce the LED conditions for all nodes of a given element K_e, we define the element-based correction factor α_e^K as the minimum of the nodal correction factors:

$$\alpha_e^K = \min_{i \in \mathcal{V}(e)} \Phi_i, \tag{36}$$

where $\mathcal{V}(e)$ is the set of vertices of element K_e. In contrast to element-based FCT limiters [35, 38, 39], α_e^K does not depend on the signs of f_i^e and the same nodal correction factor Φ_i is used in all elements containing node i.

The value of Φ_i is determined using a limiter function $\Phi : [u_i^{\min}, u_i^{\max}] \mapsto [0, 1]$. Our approach to construction of Φ is based on the following design principles:

- $\Phi_i = \Phi(u_i)$ depends continuously on the data;

- $\Phi_i = 0$ at a local extremum ($u_i = u_i^{\min}$ or $u_i = u_i^{\max}$);

- $\Phi_i = 1$ if $u_i = \bar{u}_i$ for some $\bar{u}_i \in (u_i^{\min}, u_i^{\max})$.

The first property is needed to secure the well-posedness of the nonlinear discrete problem and convergence of fixed-point iterations [3]. The second property guarantees that the sum of antidiffusive element contributions to node i is local extremum diminishing. The third property is used to define a suitable limiter function that should allow the unconstrained antidiffusive flux to be employed in regions where the discrete nodal solution approaches an admissible average value $\bar{u}_i$.

5.1.1. Definition of useful average values $\bar{u}_i$

For example, consider the mass-weighted average $\bar{u}_i^M$ defined by the formula

$$\bar{u}_i^M = \frac{1}{m_i} \sum_j m_{ij} u_j. \tag{37}$$

If $u_i = \bar{u}_i^M$ then $u_i = \frac{\sum_{j\neq i} m_{ij} u_j}{\sum_{j\neq i} m_{ij}}$ is a convex combination of the solution values at neighbor nodes. Note that $u_i^{\max} = u_i$ or $u_i^{\min} = u_i$ implies the strong discrete maximum principle ($u_j = u_i$ for all $j \neq i$).

A linearity-preserving limiter can be designed using an average $\bar{u}_i$ such that $u_i = \bar{u}_i$ for a locally linear function. Such an average can be defined, e.g., using the coefficients of a discrete Laplacian operator $S = \{s_{ij}\}$ to define

$$\bar{u}_i^S = (1 - \gamma)u_i - \frac{\gamma}{s_{ii}} \sum_{j\neq i} s_{ij} u_j. \tag{38}$$

By definition of a discrete Laplacian operator, we have $\sum_j s_{ij} u_j = 0$, i.e., $u_i = -\frac{1}{s_{ii}} \sum_{j\neq i} s_{ij} u_j = \bar{u}_i^S$ for linear functions. It follows that $\bar{u}_i^S$ is linearity-preserving for any $\gamma \in [0, 1]$. In the numerical study below we use $\gamma = 0.5$.

If the coefficients of S satisfy $s_{ij} \geq 0$ for $j \neq i$, then the LED criterion holds in the case $u_i = \bar{u}_i^S$. The Galerkin approximation of the Laplacian yields

$$s_{ij} = \int_\Omega \nabla \varphi_i \cdot \nabla \varphi_j \, d\mathbf{x}.$$

It is well known that $s_{ij} \geq 0$ for $j \neq i$ under certain restrictions on the angles or aspect ratios of mesh elements (triangulations of weakly acute type [4, 8], rectangular meshes of nonnarrow type [9, 12]). If the given mesh

15

violates these geometric conditions, nonnegative weights s_{ij} for $j \neq i$ can be constructed using various generalizations of barycentric coordinates to arbitrary polytopes [13, 23, 43, 46]. For example, the *maximum entropy coordinates* [23] may be used to construct a LED-type linearity-preserving average $\bar{u}_i^S$.

After choosing the average $\bar{u}_i$, a piecewise-linear continuous function Φ with desired properties can be constructed using the bounds

$$\bar{u}_i^{\max} = \frac{u_i^{\max} + \bar{u}_i}{2}, \qquad \bar{u}_i^{\min} = \frac{u_i^{\min} + \bar{u}_i}{2} \tag{39}$$

to define the range of safe values $u_i \in [\bar{u}_i^{\min}, \bar{u}_i^{\max}]$ for which the limiter function will return $\Phi_i = 1$. In accordance with the LED criterion, we must use $\Phi_i = 0$ for $u_i = u_i^{\max}$ and $u_i = u_i^{\min}$. Using linear interpolation on the intervals $(u_i^{\min}, \bar{u}_i^{\min})$ and $(\bar{u}_i^{\max}, u_i^{\max})$, we end up with the formula

$$\Phi_i = \begin{cases} \frac{u_i^{\max} - u_i}{u_i^{\max} - \bar{u}_i^{\max}} & \text{if} \quad u_i > \bar{u}_i^{\max}, \\ \frac{u_i^{\min} - u_i}{u_i^{\min} - \bar{u}_i^{\min}} & \text{if} \quad u_i < \bar{u}_i^{\min}, \\ 1 & \text{otherwise.} \end{cases} \tag{40}$$

The graph of the so-defined nodal limiter function Φ is sketched in Fig. 1.
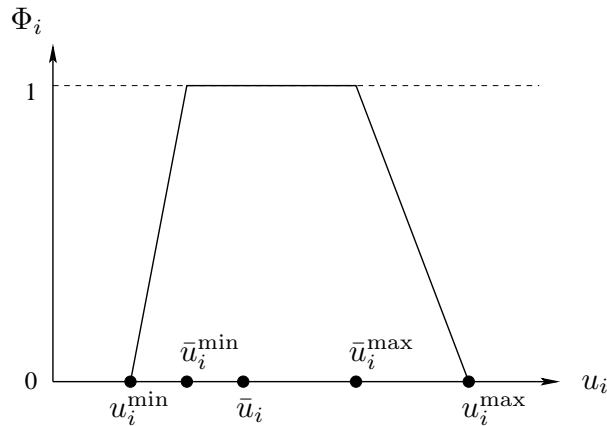

Figure 1: Nodal correction factor for the element-based limiter.

5.2. Constrained mass matrix

In applications to unsteady transport equations, we use the correction factor α_e^M in (35) to balance the contribution of the consistent mass matrix with that of the limited convective term and prevent it from producing corrections that have an opposite effect. Let $\dot{u}$ denote the vector of nodal time derivatives that corresponds to the constrained approximation

$$\dot{u}^C = M_L^{-1}(Lu + \bar{f}), \tag{41}$$

where $\bar{f}$ is assembled from limited antidiffusive element contributions

$$\bar{f}_i = \sum_{e \in \mathcal{E}(i)} \bar{f}_i^e, \qquad \bar{f}_i^e = \bar{f}_i^{e,M} + \bar{f}_i^{e,K}, \tag{42}$$

$$\bar{f}^{e,M} = \min\{\alpha_e^M, \alpha_e^K\} f^{e,M}, \qquad f^{e,M} = (M_C^e - M_L^e)\frac{\mathrm{d}u^e}{\mathrm{d}t},$$

$$\bar{f}^{e,K} = \alpha_e^K f^{e,K} + g^e, \qquad f^{e,K} = f_{\mathrm{stab}}^e - D^e u^e.$$

Setting α_e^M equal to zero, one obtains the lumped-mass approximation

$$\dot{u}^L = M_L^{-1}(Lu + \bar{f}^K), \qquad \bar{f}_i^K = \sum_{e \in \mathcal{E}(i)} \bar{f}_i^{e,K} \tag{43}$$

such that

$$\dot{u}^C = \dot{u}^L + M_L^{-1}\bar{f}^M, \qquad \bar{f}_i^M = \sum_{e \in \mathcal{E}(i)} \bar{f}_i^{e,M}. \tag{44}$$

Hence, the contribution of $\bar{f}^M$ can be interpreted as a high-order correction to $\dot{u}^L$. In order to constrain the changes of the time derivative due to this correction, we choose α_e^M so as to enforce the inequality constraints

$$\dot{u}_i^{\min} \leq \dot{u}^C \leq \dot{u}_i^{\max},$$

where $\dot{u}_i^{\min}$ and $\dot{u}_i^{\max}$ denote the local maxima and minima of $\dot{u}^L$, i.e.,

$$\dot{u}_i^{\min} = \min_{j \in \mathcal{N}(i)} \dot{u}_j^L, \tag{45}$$

$$\dot{u}_i^{\max} = \max_{j \in \mathcal{N}(i)} \dot{u}_j^L. \tag{46}$$

17

Substituting $\dot{u}^L$ for the time derivative in the constrained element vector

$$\bar{f}^{e,M} \approx \min\{\alpha_e^M, \alpha_e^K\}(M_C^e - M_L^e)\dot{u}^L,$$

we use a local version of the element-based FCT algorithm [35] to calculate

$$\alpha_e^M = \min_{i \in \mathcal{V}(e)} \Psi_i^e, \tag{47}$$

where Ψ_i^e are the nodal correction factors defined by

$$\Psi_i^e = \begin{cases} \min\left\{1, \frac{m_i^e(\dot{u}_i^{\max} - \dot{u}_i^L)}{f_i^{e,M}}\right\} & \text{if} \quad f_i^{e,M} > 0, \\ 1 & \text{if} \quad f_i^{e,M} = 0, \\ \min\left\{1, \frac{m_i^e(\dot{u}_i^{\min} - \dot{u}_i^L)}{f_i^{e,M}}\right\} & \text{if} \quad f_i^{e,M} < 0, \end{cases} \tag{48}$$

where m_i^e is the ith diagonal entry of the lumped element mass matrix M_L^e.

The above choice of α_e^M implies that the limited element contributions satisfy

$$m_i^e(\dot{u}_i^{\min} - \dot{u}_i^L) \leq \bar{f}_i^{e,M} \leq m_i^e(\dot{u}_i^{\max} - \dot{u}_i^L).$$

Summing over all elements containing node i, it is easy to verify that the values of $\dot{u}^C$ are bounded by the local extrema $\dot{u}_i^{\max}$ and $\dot{u}_i^{\min}$.

5.3. Proof of the LED property

To verify the LED property, we need to show that estimates of the form (19) hold for the proposed choice of the correction factors α_e^M and α_e^K.

Consider an interior node i such that the contribution of inflow boundary conditions $g_i = \sum_{e \in \mathcal{E}(i)} g_i^e$ is equal to zero. We have

$$\bar{f}_i = \sum_{e \in \mathcal{E}(i)} (\min\{\alpha_e^M, \alpha_e^K\} f_i^{e,M} + \alpha_e^K f_i^{e,K}), \qquad 0 \leq \alpha_e^K \leq \Phi_i,$$

where Φ_i is the nodal correction factor given by (40). It follows that

$$\Phi_i f_i^- \leq \bar{f}_i \leq \Phi_i f_i^+, \tag{49}$$

where

$$f_i^+ = \sum_{e \in \mathcal{E}(i)} \max\{0, f_i^{e,M}\} + \sum_{e \in \mathcal{E}(i)} \max\{0, f_i^{e,K}\},$$

18

$$f_i^- = \sum_{e \in \mathcal{E}(i)} \min\{0, f_i^{e,M}\} + \sum_{e \in \mathcal{E}(i)} \min\{0, f_i^{e,K}\}.$$

Suppose that $\bar{f}_i > 0$. Then we must have $u_i < u_i^{\max}$ because $u_i = u_i^{\max}$ implies $\Phi_i = 0$ and, therefore, $\bar{f}_i = 0$ in contradiction to the assumption that $\bar{f}_i > 0$. By definition (39), it follows that

$$u_i^{\min} < \bar{u}_i^{\min} < \bar{u}_i < \bar{u}_i^{\max} < u_i^{\max}.$$

Using (49) and (40), we will show that estimates of the form (19) hold on each interval on which the limiter function Φ_i is linear.

- In the case $u_i \in (\bar{u}_i^{\max}, u_i^{\max})$, we have

$$\Phi_i = \frac{u_i^{\max} - u_i}{u_i^{\max} - \bar{u}_i^{\max}},$$

$$\bar{f}_i \leq \Phi_i f_i^+ = \frac{f_i^+}{u_i^{\max} - \bar{u}_i^{\max}}(u_i^{\max} - u_i).$$

- The estimate for $u_i \in [\bar{u}_i^{\min}, \bar{u}_i^{\max}]$ is given by

$$\Phi_i = 1,$$

$$\bar{f}_i \leq \Phi_i f_i^+ = f_i^+ = \frac{f_i^+}{u_i^{\max} - u_i}(u_i^{\max} - u_i) \leq \frac{f_i^+}{u_i^{\max} - \bar{u}_i^{\max}}(u_i^{\max} - u_i).$$

- In the case $u_i \in (u_i^{\min}, \bar{u}_i^{\min})$, we obtain

$$\Phi_i = \frac{u_i^{\min} - u_i}{u_i^{\min} - \bar{u}_i^{\min}}$$

$$\bar{f}_i \leq \Phi_i f_i^+ < f_i^+ = \frac{f_i^+}{u_i^{\max} - u_i}(u_i^{\max} - u_i) < \frac{f_i^+}{u_i^{\max} - \bar{u}_i^{\max}}(u_i^{\max} - u_i).$$

It follows that

$$\bar{f}_i = c_i(u_i^{\max} - u_i), \qquad 0 < c_i \leq \frac{f_i^+}{u_i^{\max} - \bar{u}_i^{\max}} =: c_i^{\max} < \infty. \qquad (50)$$

19

This proves the existence of the LED representation (50) in the case $\bar{f}_i > 0$.

In the case $\bar{f}_i < 0$, the same arguments lead to

$$\bar{f}_i = c_i(u_i^{\min} - u_i), \qquad 0 < c_i \leq \frac{f_i^-}{u_i^{\min} - \bar{u}_i^{\min}} =: c_i^{\min} < \infty, \qquad (51)$$

which proves that the antidiffusive correction of the low-order scheme is LED.

5.4. Lipschitz continuity

A further desirable property is Lipschitz continuity of the limited antidiffusive term. In the context of constrained Galerkin schemes, the practical importance of this property lies in the fact that it guarantees unique solvability of the semi-discrete problem [25] and that of the nonlinear system associated with the steady state limit of the time-dependent problem [3].

The use of the element-based limiter based on (36) and (40) in steady state computations may cause convergence problems which can be attributed to the lack of Lipschitz continuity. In this section, we present a Lipschitz-continuous version supported by the theory developed in [2, 3].

The construction of Lipschitz-continuous limiters for algebraic flux correction schemes [3] begins with a conservative decomposition of the antidiffusive term into internodal fluxes. For example, an antidiffusive element contribution of the form $f^e(u) = -D^e u^e$ admits the following representation:

$$f_i^e = \sum_{\substack{j \in \mathcal{V}(e) \\ j \neq i}} f_{ij}^e, \qquad f_{ij}^e = d_{ij}^e(u_i - u_j). \qquad (52)$$

Note that $f_{ji}^e = -f_{ij}^e$ since the artificial diffusion coefficients satisfy $d_{ji}^e = d_{ij}^e$.

When it comes to limiting, the antidiffusive fluxes f_{ij}^e and f_{ji}^e are multiplied by a correction factor $\alpha_{ij}^e = \alpha_{ji}^e$. Setting $\alpha_{ij}^e := \alpha_e$ for all pairs of nodes corresponds to the isotropic limiting strategy ($\bar{f}^e = \alpha_e f^e$) employed so far.

According to Lemma 3.5 from [3], the nonlinear function $\alpha_{ij}^e(u)(u_i - u_j)$ is Lipschitz-continuous for correction factors of the form

$$\alpha_{ij}^e(u) = \frac{A_{ij}(u)}{|u_j - u_i| + B_{ij}(u)}, \qquad (53)$$

20

where $A_{ij}(u)$ and $B_{ij}(u)$ are nonnegative Lipschitz-continuous functions. For example, this criterion is satisfied by the following limiter function [2]:

$$\alpha_{ij}^e(u) = \begin{cases} \dfrac{|\sum_{j\neq i}(u_i - u_j)|}{\sum_{j\neq i}|u_i - u_j|} & \text{if } u_i^{\max} \neq u_i^{\min}, \\ 0 & \text{otherwise.} \end{cases} \tag{54}$$

However, this definition leads to a rather diffusive approximation. Therefore, we will generalize it using the ideas that have led us to formula (40).

Let $\bar{u}_i$ be a weighted average defined as in Section 5.1.1 in terms of some positive weights $w_{ij} > 0$ s.t. $\sum_j w_{ij} = 1$ (e.g., $w_{ij} = \frac{m_{ij}}{m_i}$ for $\bar{u}_i^M$).

Introducing a parameter $\beta \in [0, 1)$, we define the generalized limiter

$$\bar{u}_i^{\max} = \beta u_i^{\max} + (1 - \beta)\bar{u}_i, \quad \bar{u}_i^{\min} = \beta u_i^{\min} + (1 - \beta)\bar{u}_i, \tag{55}$$

$$\Phi_i = \begin{cases} 1 - \dfrac{\max\{0, u_i - \bar{u}_i^{\max}\} + \max\{0, \bar{u}_i^{\min} - u_i\}}{(1 - \beta)\sum_{j\neq i} w_{ij}|u_i - u_j|}, & \text{if } u_i^{\max} \neq u_i^{\min}, \\ 0 & \text{otherwise.} \end{cases} \tag{56}$$

The so-defined limiter function Φ_i satisfies the Lipschitz condition (53) and the LED criterion. As before, we have $\Phi_i(u) = 1$ for $u_i \in [\bar{u}_i^{\min}, \bar{u}_i^{\max}]$ and $\Phi_i(u) = 0$ for $u_i = u_i^{\min}$ or $u_i = u_i^{\max}$. Indeed, we have

$$u_i - \bar{u}_i^{\max} = \beta(u_i - u_i^{\max}) + (1 - \beta)(u_i - \bar{u}_i),$$

$$u_i - \bar{u}_i^{\min} = \beta(u_i - u_i^{\min}) + (1 - \beta)(u_i - \bar{u}_i),$$

where

$$u_i - \bar{u}_i = \sum_{j\neq i} w_{ij}(u_i - u_j).$$

In the case $u_i = u_i^{\max}$ we have $u_i - u_j = |u_i - u_j| \ \forall j \neq i$, whence

$$\Phi_i = 1 - \frac{u_i - \bar{u}_i^{\max}}{(1 - \beta)\sum_{j\neq i} w_{ij}|u_i - u_j|} = 0.$$

For $u_i = u_i^{\min}$, we have $u_i - u_j = -|u_i - u_j| \ \forall j \neq i$ and, therefore,

$$\Phi_i = 1 - \frac{u_i - \bar{u}_i^{\min}}{(1 - \beta)\sum_{j\neq i} w_{ij}|u_i - u_j|} = 0.$$

It is worth mentioning that $\alpha_{ij}^e := \Phi_i$ reduces to (54) if we set $\beta = 0$ and

$$w_{ij} = 1 \quad \forall j \neq i.$$

In the numerical study below, we used $\beta = 0.5$ and $w_{ij} = \frac{m_{ij}}{m_i}$ for $j \neq i$.

The proof of Lipschitz continuity may be invalidated when the minimum of nodal correction factors is taken in (36) or higher-order dissipation is added to the antidiffusive flux f_{ij}^e. To enforce the Lipschitz condition, we switch off the background dissipation by setting $\omega := 0$ and limit the local internodal fluxes f_{ij}^e using an upwind-biased limiting strategy which traces its origins to edge-based algebraic flux correction schemes [3, 32, 33, 34].

Let $\alpha_{ij}^e = \Phi_k = \alpha_{ji}^e$, where Φ_k is the nodal correction factor defined by (56) and k is the number of the upwind node defined by [32, 34]

$$k = \begin{cases} i & \text{if } k_{ij}^e \leq k_{ji}^e, \\ j & \text{otherwise.} \end{cases} \tag{57}$$

The proof of the LED property remains valid for Φ_k given by (56). The LED constraint for the downwind node is satisfied automatically because the corresponding antidiffusive flux is compensated by the contribution of the low-order operator. Hence, the choice of the correction factor α_{ij}^e is based solely on the LED criterion for the upwind node. We refer to [32, 34] for a presentation of the design philosophy behind upwind-biased LED limiters.

The above approach combines the high accuracy of formula (40) with superior robustness of Lipschitz-continuous upwind limiters analyzed in [2, 3]. This justifies the increased complexity of the anisotropic limiting strategy.

5.5. Time discretization

In this paper, we discretize the constrained semi-discrete finite element scheme

$$M_L \frac{\mathrm{d}u}{\mathrm{d}t} = Lu + \bar{f} \tag{58}$$

in time using the two-level θ method

$$M_L \frac{u^{n+1} - u^n}{\Delta t} = \theta(Lu^{n+1} + \bar{f}^{n+1}) + (1 - \theta)(Lu^n + \bar{f}^n). \tag{59}$$

22

The inclusion of the limited antidiffusive term leads to the CFL-like condition

$$\frac{1}{\Delta t} \geq (1 - \theta) \left[\sum_{j \in \mathcal{N}(i) \setminus \{i\}} l_{ij} + c_i \right] \qquad \forall i = 1, \ldots, N, \qquad (60)$$

where c_i is the bounded coefficient of the LED representation (50) or (51).

The convergence behavior of implicit schemes is also affected by the value of the coefficient c_i. Choosing $\bar{u}_i^{\max}$ too close to $u_i^{\max}$ or $\bar{u}_i^{\min}$ too close to $u_i^{\min}$ may produce large values of c_i and cause convergence problems or require the use of inordinately small time steps. The proposed definition of $\bar{u}_i^{\min}$ and $\bar{u}_i^{\max}$ offers a reasonable compromise between accuracy and robustness.

We remark that the semi-discrete nature of the proposed approach makes it possible to use a wide range of time discretizations including explicit and implicit strong stability preserving (SSP) Runge-Kutta schemes [15].

Due to the dependence of α_e^M and α_e^K on the unknown solution, the algebraic system (59) is nonlinear. It can be solved using the fixed-point iteration

$$u^{(m+1)} = u^{(m)} + \left[\frac{1}{\Delta t} M_L - \theta L \right]^{-1} r^{(m)}, \qquad m = 0, 1, 2, \ldots \qquad (61)$$

$$r^{(m)} = \theta (Lu^{(m)} + \bar{f}^{(m)}) + (1 - \theta)(Lu^n + \bar{f}^n) - M_L \frac{u^{(m)} - u^n}{\Delta t}. \qquad (62)$$

The rates of convergence to steady state solutions can be greatly improved using Anderson acceleration for fixed-point iterations [31, 45].

6. Numerical examples

In this section, we apply the proposed methodology to two-dimensional test problems that have been used to study edge-based algebraic flux correction schemes in [29, 30, 31, 32]. Given a reference solution u, we use the following norms to assess the accuracy of a finite element approximation u_h

$$E_1(h) = \sum_i m_i |u(\mathbf{x}_i) - u_i| \approx \|u - u_h\|_1, \qquad (63)$$

$$E_2(h) = \sqrt{\sum_i m_i |u(\mathbf{x}_i) - u_i|^2} \approx \|u - u_h\|_2, \tag{64}$$

where $m_i = \int_\Omega \varphi_i \, d\mathbf{x}$ is a diagonal coefficient of the lumped mass matrix M_L.

To study the dependence of E_1 and E_2 on the mesh size h, the numerical solutions computed on two different meshes are used to estimate the experimental order of convergence (EOC) using the formula [36]

$$p = \log_2 \left(\frac{E_1(2h)}{E_1(h)} \right). \tag{65}$$

In grid convergence studies for time-dependent problems, the ratio of the time step and mesh size is held constant in the process of refinement.

6.1. Solid body rotation

The solid body rotation test [36, 47] is often used to evaluate numerical advection schemes. The problem to be solved is the continuity equation

$$\frac{\partial u}{\partial t} + \nabla \cdot (\mathbf{v} u) = 0 \qquad \text{in } \Omega = (0, 1) \times (0, 1). \tag{66}$$

The velocity $\mathbf{v}$ describes a counterclockwise rotation about the center

$$\mathbf{v}(x, y) = (0.5 - y, x - 0.5). \tag{67}$$

After each full revolution, the exact solution u coincides with the given initial data u_0. Hence, the challenge of this test is to preserve the shape of u_0.

Following LeVeque [36], we simulate solid body rotation of a profile that consists of a slotted cylinder, a sharp cone, and a smooth hump (see Fig. 2a). The geometry of each body is described by a given function $G(x, y)$ defined on a circle of radius $r_0 = 0.15$ centered at some point (x_0, y_0). Let

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

be the normalized distance from (x_0, y_0). Then $r(x, y) \leq 1$ inside the circle.

The slotted cylinder is centered at the point $(x_0, y_0) = (0.5, 0.75)$ and

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

The cone is centered at $(x_0, y_0) = (0.5, 0.25)$, and its shape is given by

$$G(x, y) = 1 - r(x, y).$$

The hump is centered at $(x_0, y_0) = (0.25, 0.5)$, and the shape function is

$$G(x, y) = \frac{1 + \cos(\pi r(x, y))}{4}.$$

In the rest of the domain, the solution to (66) is initialized by zero, and homogeneous Dirichlet boundary conditions are prescribed at the inlets.

The snapshots presented in Figs 2 and 3 show the numerical solutions at the final time $T = 2\pi$ which corresponds to one full rotation. All computations are performed on a uniform mesh of 128×128 bilinear elements using the Crank-Nicolson time-stepping with the time step $\Delta t = 10^{-3}$. The low-order solution ($\alpha_e^K := 0$) obtained using the default definition (31) of the artificial diffusion operator is shown in Fig. 2b. The standard Galerkin scheme (Fig. 2c) produces global nonphysical oscillations, whereas the use of high-order background dissipation (Fig. 2d) localizes them to a small neighborhood of steep gradients. The results produced by four different versions of the constrained Galerkin scheme are presented in Fig. 3. The difference between the constrained solutions obtained using the mass-weighted ($\bar{u}_i = \bar{u}_i^M$) and Laplacian-weighted ($\bar{u}_i = \bar{u}_i^S$) averages in the formula for α_e^K is marginal. The use of background dissipation ($\omega = 0.1$) alleviates spurious distortions of the sharp cone, which are more pronounced in the case of the standard Galerkin approximation ($\omega = 0.0$). The convergence history and EOCs for all methods under investigation are listed in Tables 1-9. All constrained solutions are calculated using the artificial diffusion operator defined by (31) since definitions (33) and (34) are more diffusive (see Tables 1-3).

6.2. Circular convection

In the second test problem [24], we consider the steady convection equation

$$\nabla \cdot (\mathbf{v} u) = 0 \qquad \text{in } \Omega = (-1, 1) \times (0, 1), \tag{68}$$
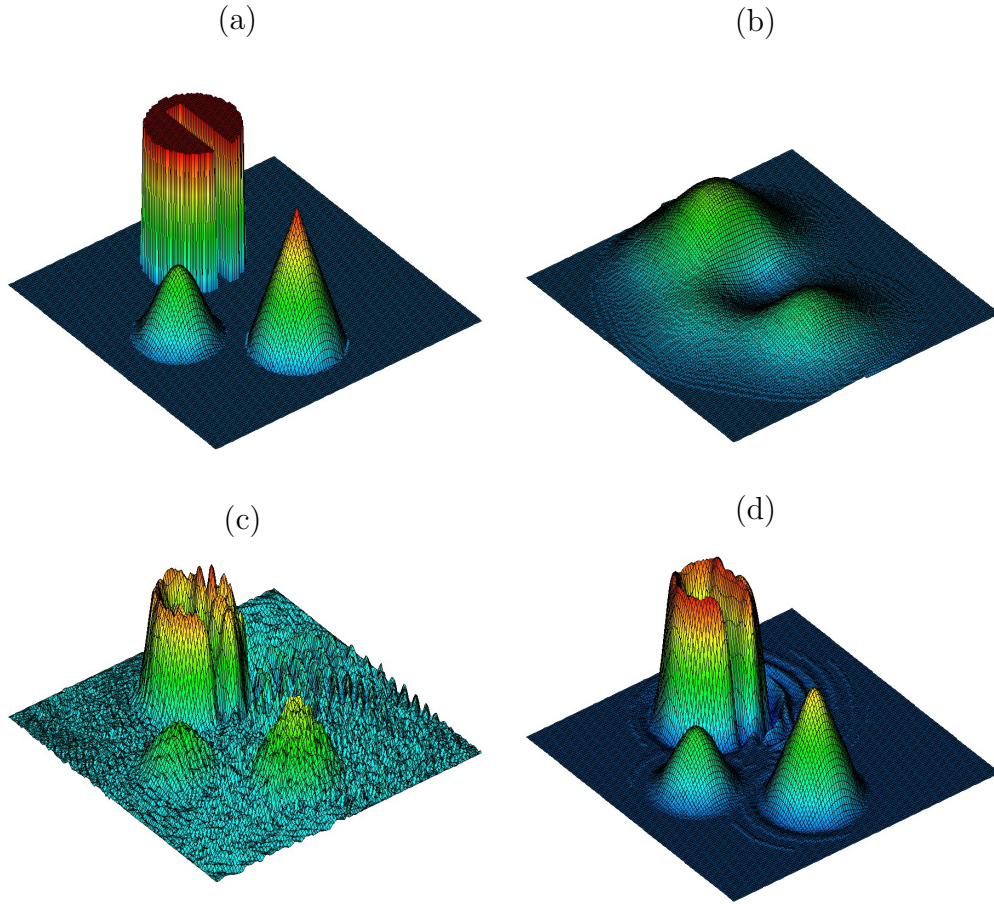
Figure 2: Solid body rotation: (a) initial data / exact solution, (b) $\alpha_e^K = 0$, (c) $\alpha_e^M = \alpha_e^K = 1$, $\omega = 0.0$, (d) $\alpha_e^M = \alpha_e^K = 1$, $\omega = 0.1$. Space discretization: $\mathcal{Q}_1$ elements ($h = 1/128$), Time discretization: Crank-Nicolson, $\Delta t = 10^{-3}$. Simulation time: $T = 2\pi$.

h	E_1	EOC	E_2	EOC
1/32	0.115e+00		0.230e+00	
1/64	0.111e+00	0.05	0.209e+00	0.14
1/128	0.968e-01	0.20	0.186e+00	0.17
1/256	0.795e-01	0.28	0.164e+00	0.18

Table 1: Solid body rotation: low-order scheme based on (31).
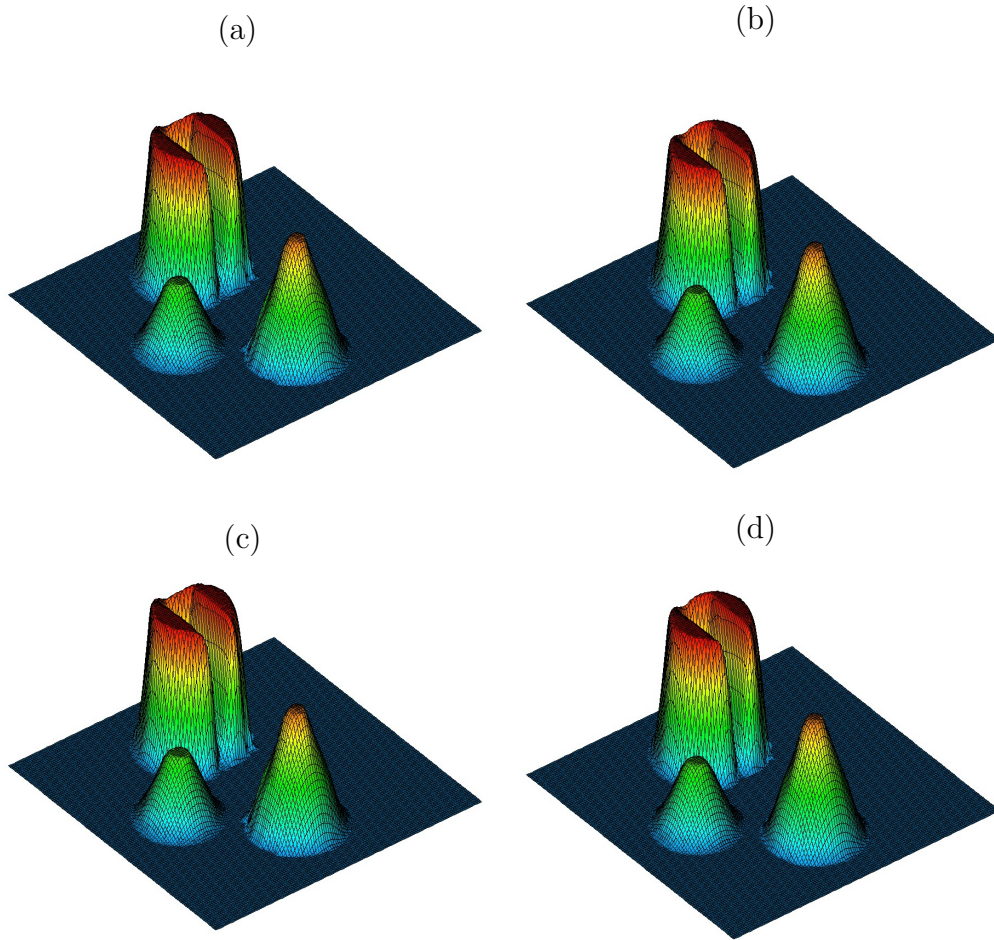
Figure 3: Solid body rotation: (a) $\bar{u}_i = \bar{u}_i^M$, $\omega = 0.0$, (b) $\bar{u}_i = \bar{u}_i^M$, $\omega = 0.1$, (c) $\bar{u}_i = \bar{u}_i^S$, $\omega = 0.0$, (d) $\bar{u}_i = \bar{u}_i^S$, $\omega = 0.1$, Space discretization: $\mathcal{Q}_1$ elements ($h = 1/128$), Time discretization: Crank-Nicolson, $\Delta t = 10^{-3}$. Simulation time: $T = 2\pi$.

h	E_1	EOC	E_2	EOC
1/32	0.107e+00		0.252e+00	
1/64	0.112e+00	-0.07	0.234e+00	0.11
1/128	0.107e+00	0.07	0.210e+00	0.16
1/256	0.955e-01	0.16	0.186e+00	0.18

Table 2: Solid body rotation: low-order scheme based on (33).

h	E_1	EOC	E_2	EOC
1/32	0.108e+00		0.249e+00	
1/64	0.112e+00	-0.05	0.230e+00	0.11
1/128	0.105e+00	0.09	0.205e+00	0.16
1/256	0.927e-01	0.18	0.181e+00	0.18

Table 3: Solid body rotation: low-order scheme based on (34).

h	E_1	EOC	E_2	EOC
1/32	0.103e+00		0.175e+00	
1/64	0.693e-01	0.57	0.125e+00	0.49
1/128	0.472e-01	0.55	0.883e-01	0.50
1/256	0.355e-01	0.41	0.734e-01	0.27

Table 4: Solid body rotation: Galerkin scheme, $\omega = 0.0$.

h	E_1	EOC	E_2	EOC
1/32	0.621e-01		0.141e+00	
1/64	0.356e-01	0.80	0.101e+00	0.48
1/128	0.200e-01	0.83	0.711e-01	0.51
1/256	0.132e-01	0.60	0.573e-01	0.31

Table 5: Solid body rotation: Galerkin + background dissipation, $\omega = 0.1$.

h	E_1	EOC	E_2	EOC
1/32	0.559e-01		0.136e+00	
1/64	0.365e-01	0.61	0.114e+00	0.25
1/128	0.168e-01	1.12	0.708e-01	0.69
1/256	0.897e-02	0.91	0.522e-01	0.44

Table 6: Solid body rotation: constrained Galerkin, $\bar{u}_i = \bar{u}_i^M$, $\omega = 0.0$.

where

$$\mathbf{v}(x, y) = (y, -x).$$

h	E_1	EOC	E_2	EOC
1/32	0.564e-01		0.137e+00	
1/64	0.363e-01	0.64	0.115e+00	0.25
1/128	0.169e-01	1.10	0.720e-01	0.68
1/256	0.922e-02	0.87	0.536e-01	0.43

Table 7: Solid body rotation: constrained Galerkin, $\bar{u}_i = \bar{u}_i^M$, $\omega = 0.1$.

h	E_1	EOC	E_2	EOC
1/32	0.567e-01		0.136e+00	
1/64	0.371e-01	0.61	0.115e+00	0.24
1/128	0.170e-01	1.13	0.710e-01	0.70
1/256	0.902e-02	0.91	0.522e-01	0.44

Table 8: Solid body rotation: constrained Galerkin, $\bar{u}_i = \bar{u}_i^S$, $\omega = 0.0$.

h	E_1	EOC	E_2	EOC
1/32	0.571e-01		0.137e+00	
1/64	0.369e-01	0.63	0.116e+00	0.24
1/128	0.171e-01	1.11	0.723e-01	0.68
1/256	0.902e-02	0.92	0.522e-01	0.47

Table 9: Solid body rotation: constrained Galerkin, $\bar{u}_i = \bar{u}_i^S$, $\omega = 0.1$.

The exact solution and inflow boundary conditions are given by

$$u(x, y) = \begin{cases} G(r), & \text{if } 0.35 \leq r = \sqrt{x^2 + y^2} \leq 0.65, \\ 0, & \text{otherwise}, \end{cases}$$

where $G(r)$ is a given function that defines the shape of the solution profile.

To evaluate the ability of the proposed limiter to handle smooth data and discontinuous solutions, we consider the shape functions

$$G_1(r) = \cos^2\left(5\pi\frac{2r + 1}{3}\right), \qquad G_2(r) \equiv 1.$$

As before, computations are performed on a uniform mesh of bilinear finite elements which is successively refined to perform a grid convergence study. In

this numerical study, the solution of the time-dependent transport equation (1) is marched to the steady state using the backward Euler method.

The results produced by the constrained Galerkin scheme using $\bar{u}_i = \bar{u}_i^M$, $\omega = 0.1$, $h = 1/128$ are displayed in Fig. 4. The errors and EOCs for the inflow profiles G_1 and G_2 are listed in Tables 10-19. All constrained approximations employ local edge-based artificial diffusion of the form (31). The convergence rates for constrained Galerkin schemes based on low-order artificial diffusion of the form (33) and (34) are similar (not presented here).

As mentioned in Section 5.4, the impossibility of proving the Lipschitz condition for the isotropic element-based limiter (36) may give rise to convergence problems in steady state computations on fine meshes. These problems can be cured using the upwind-biased anisotropic limiter (56) to constrain anti-diffusive fluxes associated with pairs of nodes belonging to the same element. In our example, the Lipschitz continuity of this version makes it possible to obtain fully converged steady state solutions which are as accurate as those produced by the isotropic limiter (compare Tables 18-19 and 20-21).

h	E_1	EOC	E_2	EOC
1/32	0.164e+00		0.216e+00	
1/64	0.116e+00	0.50	0.164e+00	0.40
1/128	0.745e-01	0.64	0.113e+00	0.54
1/256	0.442e-01	0.75	0.713e-01	0.66

Table 10: Circular convection of G_1, low-order scheme based on (31).

h	E_1	EOC	E_2	EOC
1/32	0.228e+00		0.280e+00	
1/64	0.185e+00	0.30	0.237e+00	0.24
1/128	0.136e+00	0.44	0.187e+00	0.34
1/256	0.915e-01	0.57	0.135e+00	0.47

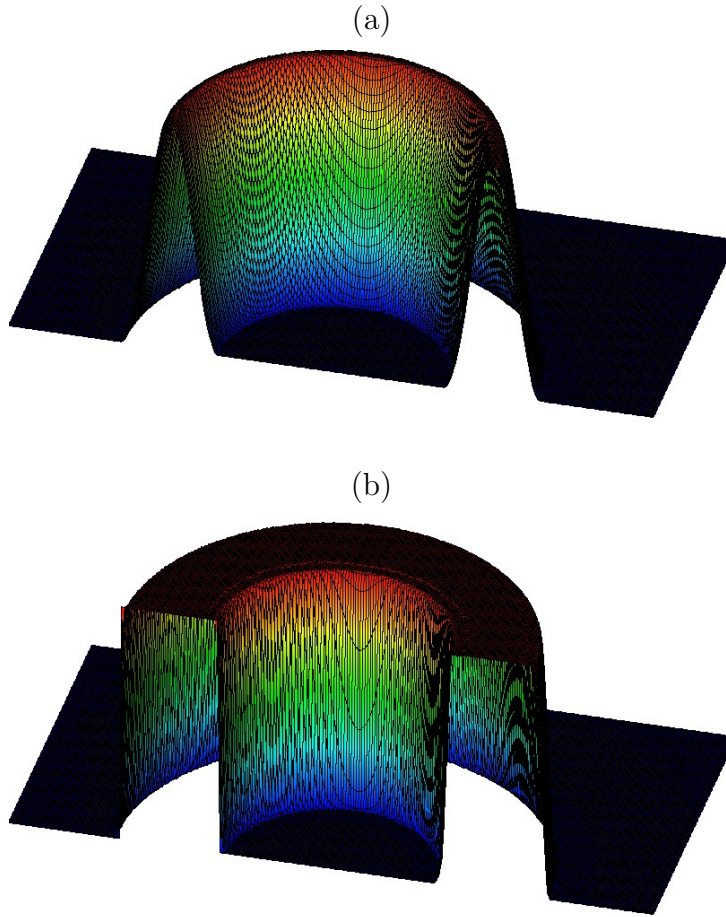Table 11: Circular convection of G_1, low-order scheme based on (33).

Figure 4: Circular convection: (a) smooth and (b) discontinuous data. Space discretization: $\mathcal{Q}_1$ elements ($h = 1/128$), constrained Galerkin scheme ($\bar{u}_i = \bar{u}_i^M$, $\omega = 0.1$).

h	E_1	EOC	E_2	EOC
1/32	0.221e+00		0.272e+00	
1/64	0.176e+00	0.33	0.228e+00	0.25
1/128	0.127e+00	0.47	0.178e+00	0.36
1/256	0.840e-01	0.60	0.126e+00	0.50

Table 12: Circular convection of G_1, low-order scheme based on (34).

h	E_1	EOC	E_2	EOC
1/32	0.270e+00		0.301e+00	
1/64	0.196e+00	0.46	0.243e+00	0.31
1/128	0.139e+00	0.50	0.203e+00	0.26
1/256	0.985e-01	0.50	0.170e+00	0.26

Table 13: Circular convection of G_2, low-order scheme based on (31).

h	E_1	EOC	E_2	EOC
1/32	0.387e+00		0.394e+00	
1/64	0.309e+00	0.33	0.325e+00	0.28
1/128	0.228e+00	0.44	0.265e+00	0.29
1/256	0.162e+00	0.49	0.219e+00	0.28

Table 14: Circular convection of G_2, low-order scheme based on (33).

h	E_1	EOC	E_2	EOC
1/32	0.374e+00		0.382e+00	
1/64	0.292e+00	0.36	0.313e+00	0.29
1/128	0.214e+00	0.45	0.255e+00	0.30
1/256	0.152e+00	0.49	0.212e+00	0.27

Table 15: Circular convection of G_2, low-order scheme based on (34).

h	E_1	EOC	E_2	EOC
1/32	0.850e-02		0.130e-01	
1/64	0.151e-02	2.49	0.271e-02	2.26
1/128	0.271e-03	2.48	0.645e-03	2.07
1/256	0.522e-04	2.38	0.166e-03	1.96

Table 16: Circular convection of G_1, Galerkin + dissipation ($\omega = 0.1$).

7. Summary

In this paper, we explored an element-based approach to constraining the consistent mass matrix and the discrete transport operator in Galerkin finite

h	E_1	EOC	E_2	EOC
1/32	0.756e-01		0.145e+00	
1/64	0.433e-01	0.80	0.105e+00	0.47
1/128	0.265e-01	0.71	0.863e-01	0.28
1/256	0.153e-01	0.79	0.630e-01	0.45

Table 17: Circular convection of G_2, Galerkin + dissipation ($\omega = 0.1$).

h	E_1	EOC	E_2	EOC
1/32	0.230e-01		0.443e-01	
1/64	0.511e-02	2.17	0.125e-01	1.83
1/128	0.103e-02	2.31	0.327e-02	1.93
1/256	0.191e-03	2.43	0.815e-03	2.00

Table 18: Circular convection of G_1, constrained Galerkin ($\bar{u}_i = \bar{u}_i^M$, $\omega = 0.1$).

h	E_1	EOC	E_2	EOC
1/32	0.698e-01		0.148e+00	
1/64	0.384e-01	0.86	0.107e+00	0.47
1/128	0.229e-01	0.75	0.861e-01	0.31
1/256	0.128e-01	0.84	0.623e-01	0.47

Table 19: Circular convection of G_2, constrained Galerkin ($\bar{u}_i = \bar{u}_i^M$, $\omega = 0.1$).

h	E_1	EOC	E_2	EOC
1/32	0.256e-01		0.452e-01	
1/64	0.692e-02	1.89	0.138e-01	1.71
1/128	0.156e-02	2.15	0.371e-02	1.90
1/256	0.330e-03	2.24	0.954e-03	1.96

Table 20: Circular convection of G_1, anisotropic limiter ($\bar{u}_i = \bar{u}_i^M$, $\omega = 0.0$).

element schemes. The proposed methodology leads to simple local extremum diminishing corrections of standard bilinear forms at the semi-discrete level, which makes it universally applicable to stationary and time-dependent problems on arbitrary meshes. Moreover, we introduced a Lipschitz-continuous

33

h	E_1	EOC	E_2	EOC
1/32	0.621e-01		0.136e+00	
1/64	0.336e-01	0.89	0.938e-01	0.54
1/128	0.203e-01	0.73	0.779e-01	0.27
1/256	0.113e-01	0.85	0.539e-01	0.53

Table 21: Circular convection of G_2, anisotropic limiter ($\bar{u}_i = \bar{u}_i^M$, $\omega = 0.0$).

version of the nodal limiter function which improves the convergence behavior of fixed point iterations and may be used in edge-based nonlinear diffusion operators as a generalization of the limiter employed by Barrenechea et al. [3]. Finally the paper also describes approaches to enforce linearity preservation which prevents erroneous use of nonlinear stabilization in smooth regions.

Acknowledgments

References

[1] S. Badia and A. Hierro, On monotonicity-preserving stabilized finite element approximations of transport problem. *J. Comp. Physics* **36** (2014) A2673-A2697.

[2] G. Barrenechea, E. Burman, F. Karakatsani, Edge-based nonlinear diffusion for finite element approximations of convection-diffusion equations and its relation to algebraic flux-correction schemes. Preprint `arXiv:1509.08636v1 [math.NA]` 29 Sep 2015.

[3] G. Barrenechea, V. John, P. Knobloch, Analysis of algebraic flux correction schemes. WIAS Preprint No. **2107** (2015).

[4] T.J. Barth, Aspects of unstructured grids and finite volume solvers for the Euler and Navier-Stokes equations. In: Lecture Series 1994-05, von Karman Institute for Fluid Dynamics, Brussels, 1994.

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

[6] E. Burman, A monotonicity preserving, nonlinear, finite element upwind method for the transport equation. *Applied Mathematics Letters* **49** (2015) 141–146.

[7] E. Burman and A. Ern, Stabilized Galerkin approximation of convection-diffusion-reaction equations: discrete maximum principle and convergence. *Math. Comp.* **74** (2005) 1637–1652.

[8] P.G. Ciarlet and P.-A. Raviart, Maximum principle and convergence for the finite element method. *Comput. Methods Appl. Mech. Engrg.* **2** (1973) 17–31.

[9] I. Christie and C. Hall, The maximum principle for bilinear elements. *Int. J. Numer. Methods Engrg.* **20** (1984) 549–553.

[10] J. Donea and A. Huerta, *Finite Element Methods for Flow Problems.* John Wiley & Sons, Chichester, 2003.

[11] J. Donea, V. Selmin, L. Quartapelle, Recent developments of the Taylor-Galerkin method for the numerical solution of hyperbolic problems. *Numerical Methods for Fluid Dynamics III*, Oxford, 1988, 171–185.

[12] I. Faragó, R. Horváth, S. Korotov, Discrete maximum principle for linear parabolic problems solved on hybrid meshes. *Appl. Numer. Math.* **53** (2005) 249–264.

[13] M.S. Floater, Géza Kós, M. Reimers, Mean value coordinates in 3D. *Computer Aided Geometric Design* **22** (2005) 623–631.

[14] S.K. Godunov, A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Mat. Sb.* **47(89)**:3 (1959) 271–306.

[15] S. Gottlieb, D. Ketcheson, C.-W. Shu, *Strong Stability Preserving Runge-Kutta and Multistep Time Discretizations.* World Scientific, 2011.

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

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

[18] J.-L. Guermond, M. Nazarov, B. Popov, Y. Yang, A second-order maximum principle preserving Lagrange finite element technique for nonlinear scalar conservation equations. *SIAM J. Numer. Anal.* **52** (2014) 2163–2182.

[19] J.-L. Guermond and R. Pasquetti, A correction technique for the dispersive effects of mass lumping for transport problems. *Comput. Methods Appl. Mech. Engrg.* **253** (2013) 186–198.

[20] A. Harten, High resolution schemes for hyperbolic conservation laws. *J. Comput. Phys.* **49** (1983) 357–393.

[21] A. Harten, On a class of high resolution total-variation-stable finite-difference-schemes. *SIAM J. Numer. Anal.* **21** (1984) 1-23.

[22] J-l. Guermond, R. Pasquetti, B. Popov, Entropy viscosity method for nonlinear conservation laws. *SIAM J. Sci. Comput.* **230** (2011) 2484267.

[23] K. Hormann and N. Sukumar, Maximum entropy coordinates for arbitrary polytopes. *Computer Graphics Forum* **27** (2008) 1513–1520.

[24] M.E. Hubbard, Non-oscillatory third order fluctuation splitting schemes for steady scalar conservation laws. *J. Comput. Phys.* **222** (2007) 740–768.

[25] W. Hundsdorfer and J.G. Verwer, *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations.* Springer, 2003.

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

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

[28] V. John and P. Knobloch, On spurious oscillations at layers diminishing (SOLD) methods for convection-diffusion equations: Part I - A review. *Comput. Methods Appl. Mech. Engrg.* 196:17–20 (2007) 2197–2215.

[29] V. John and E. Schmeyer, On finite element methods for 3D time-dependent convection-diffusion-reaction equations with small diffusion. *Comput. Meth. Appl. Mech. Engrg.* **198** (2008) 475–494.

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

[31] D. Kuzmin, Linearity-preserving flux correction and convergence acceleration for constrained Galerkin schemes. *J. Comput. Appl. Math.* **236** (2012) 2317–2337.

[32] D. Kuzmin, Algebraic flux correction I. Scalar conservation laws. In: D. Kuzmin, R. Löhner, S. Turek (eds), *Flux-Corrected Transport: Principles, Algorithms, and Applications.* Springer, 2nd edition, 2012, pp. 145–192.

[33] 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, 653–656.

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

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

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

[37] R. Löhner, *Applied CFD Techniques: An Introduction Based on Finite Element Methods.* John Wiley & Sons, 2nd edition, 2008.

[38] 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.

[39] R. Löhner, K. Morgan, M. Vahdati, J.P. Boris, D.L. Book, FEM-FCT: combining unstructured grids with high resolution. *Commun. Appl. Numer. Methods* **4** (1988) 717–729.

[40] H. Luo, J.D. Baum, R. Löhner, J. Cabello, Adaptive edge-based finite element schemes for the Euler and Navier-Stokes equations; AIAA-93-0336, 1993.

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

[42] J. Peraire, M. Vahdati, J. Peiro, K. Morgan, The construction and behaviour of some unstructured grid algorithms for compressible flows. *Numerical Methods for Fluid Dynamics* **IV**, Oxford University Press, 1993, 221-239.

[43] R. DeRose, M. Meyer, Harmonic coordinates. Tech. rep. *Pixair Animation Studios* (2006).

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

[45] H.W. Walker, P. Ni, Anderson acceleration for fixed-point iterations. *SIAM J. Numer. Anal.* **49** (2011) 1715–1735.
WPI Math. Sci. Dept. Report MS-9-21-45, September 2009. Submitted to *SIAM J. Numer. Anal.*

[46] M. Wardetzky, S. Mathur, F. Kälberer, E. Grinspun, Discrete Laplace operators: No free lunch. In: A. Belyaev and M. Garland (Eds), *Eurographics Symposium on Geometry Processing* (2007), pp. 33-37.

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