Preprint
2016; **00**:1–20

Gradient-based nodal limiters for artificial diffusion operators in finite element schemes for transport equations

Dmitri Kuzmin[1,*] and John N. Shadid[2,3]

[1] *Institute of Applied Mathematics (LS III), TU Dortmund University, Vogelpothsweg 87, D-44227 Dortmund, Germany*
[2] *Computational Mathematics Dept., Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87185, USA*
[3] *Department of Mathematics and Statistics, University of New Mexico, MSC01 1115, Albuquerque, NM 87131, USA*

SUMMARY

This paper presents new linearity-preserving nodal limiters for enforcing discrete maximum principles in continuous (linear or bilinear) finite element approximations to transport problems with steep fronts. In the process of algebraic flux correction, the oscillatory antidiffusive part of a high-order base discretization is decomposed into a set of internodal fluxes and constrained to be local extremum diminishing. The proposed nodal limiter functions are designed to be continuous and satisfy the principle of linearity preservation which implies the preservation of second-order accuracy in smooth regions. The use of limited nodal gradients makes it possible to circumvent angle conditions and guarantee that the discrete maximum principle holds on arbitrary meshes. A numerical study is performed for linear convection and anisotropic diffusion problems on uniform and distorted meshes in two space dimensions.

Received . . .

KEY WORDS: convective transport, anisotropic diffusion, finite element schemes, discrete maximum principles, algebraic flux correction, limiters, gradient recovery

1. INTRODUCTION

Important structural properties of exact solutions to conservation law systems (nonnegativity, monotonicity, nonincreasing total variation) play an important role in the design of physics-compatible finite element methods. For example whereas it is easy to derive sufficient conditions under which a discrete maximum principle (DMP) holds for linear and bilinear elements, even the standard Galerkin discretization of the Laplace operator may violate these conditions on a nonuniform mesh. In the case of a hyperbolic transport equation, a linear DMP-conforming approximation of the convective term can be at most first-order accurate by the Godunov theorem [12]. 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. These methods often combine nonlinear residual-based shock-capturing terms and linear stabilization which localizes nonphysical undershoots and overshoots to a small neighborhood of steep gradients. However the techniques are generally insufficient to enforce the DMP constraints (for a review and comparative study of existing schemes see, e.g., [10, 17, 18]).

The design of nonlinear high-resolution schemes backed by the DMP theory typically involves

- construction of linear artificial diffusion operators that lead to algebraic systems satisfying the relevant DMP criteria for the coefficients of the (semi-)discrete problem;

*Correspondence to: Dmitri Kuzmin (kuzmin@math.uni-dortmund.de)

- limiting the corresponding edge or element contributions in an adaptive manner.

In this context the *algebraic flux correction* (AFC) [21] paradigm provides a set of general design principles for construction of artificial diffusion operators and limiter functions for continuous (linear or multilinear) finite elements. In particular, various generalizations of the *flux-corrected transport* (FCT) algorithm [5, 39] and *total variation diminishing* (TVD) schemes [15, 16] are available [19, 20, 25, 27]. A theoretical framework for analysis of AFC schemes was recently developed in [2, 3]. Alternative approaches to enforcing DMP can be found in [1, 6, 7, 14]. Some proofs of the DMP property impose restrictions on the angles or aspect ratios of mesh elements (triangulations of weakly acute type [4, 8], rectangular meshes of nonnarrow type [9, 11]). Other drawbacks of existing schemes include the presence of free parameters and strong numerical dissipation.

In the present paper, we focus on the design of limiter functions for artificial diffusion operators in algebraic flux correction schemes and related methods. The objective of this work is to develop new tools for enforcing DMP and linearity preservation on unstructured meshes. Instead of imposing upper and/or lower bounds on the sum of antidiffusive edge/element contributions to a given node, we define the nodal correction factors in terms of solution gradients. It turns out that the limiters proposed in [1] and [2] can be generalized and improved leading to an algorithm that combines the most attractive features of existing limiting techniques (simplicity, linearity preservation and DMP property on arbitrary meshes, continuous dependence on the data, low levels of numerical dissipation). In this paper, we use it to constrain antidiffusive fluxes in an edge-based algebraic flux correction scheme but the same limiter functions may be employed in the element-based version [20] and other nonlinear diffusion operators [2]. The convergence behavior of constrained P_1 finite element approximations is illustrated by a numerical study for linear convection and anisotropic diffusion problems in 2D. The results for distorted triangular meshes demonstrate the benefit of using limiters that do not impose any restrictions on the geometric properties of the mesh.

2. GALERKIN DISCRETIZATION

We begin with a brief introduction to the principles of algebraic flux correction for finite element approximations to the generic convection-diffusion equation

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

where u is a conserved scalar quantity, $\mathbf{v}$ is a given velocity field and $\mathcal{D}$ is a (possibly anisotropic) diffusion tensor. The model problems to be considered in this paper also include the limiting cases $\mathcal{D} = 0$ (pure convection), $\mathbf{v} = 0$ (pure diffusion), and $\frac{\partial u}{\partial t}$ (steady state).

Let $\Omega \subset \mathbb{R}^d,\ d \in \{1, 2, 3\}$ be a bounded domain with Lipschitz boundary $\Gamma = \partial\Omega$. The boundary conditions for our model problem are given by

$$\left\{ \begin{array}{rl} u = g_1 & \text{on } \Gamma_1, \\ -(\mathcal{D}\nabla u) \cdot \mathbf{n} = g_2 & \text{on } \Gamma_2, \\ (\mathbf{v}u - \mathcal{D}\nabla u) \cdot \mathbf{n} = g_3 & \text{on } \Gamma_3, \end{array} \right. \tag{2}$$

where $\mathbf{n}$ is the unit outward normal and $\Gamma_1, \Gamma_2, \Gamma_3$ are nonoverlapping subsets of Γ.

If the transient solution to (1) is of interest, we prescribe an initial condition of the form

$$u(\mathbf{x}, 0) = u_0(\mathbf{x}), \quad \mathbf{x} \in \Omega. \tag{3}$$

Steady state solutions can be obtained by solving the time-dependent problem until the time derivative vanishes. In this case, the initial condition may be chosen arbitrarily.

Substituting the boundary conditions (2) into the weak form of (1) with an admissible test function w vanishing on the Dirichlet boundary Γ_1, one obtains the variational formulation

$$\begin{aligned}\int_\Omega w\frac{\partial u}{\partial t}\,\mathrm{d}\mathbf{x} - \int_\Omega \nabla w\cdot(\mathbf{v}u - \mathcal{D}\nabla u)\,\mathrm{d}\mathbf{x} \quad &+ \quad \int_{\Gamma_2} wu\mathbf{v}\cdot\mathbf{n}\mathrm{d}s \\ &= \quad -\int_{\Gamma_2} wg_2\mathrm{d}s - \int_{\Gamma_3} wg_3\mathrm{d}s. \end{aligned} \tag{4}$$

In this paper, we discretize (4) in space using a finite element approximation defined by

$$u_h(\mathbf{x},t) = \sum_{j=1}^{N} u_j(t)\varphi_j(\mathbf{x}), \tag{5}$$

where $\{\varphi_1,\ldots,\varphi_N\}$ are the global basis functions for linear or multilinear Lagrange elements and $u_j(t)$ is the time-dependent nodal value associated with a vertex $\mathbf{x}_j$ of the computational mesh.

Approximating u by u_h and using admissible test functions $w\in\{\varphi_i : \varphi_i|_{\Gamma_1}=0\}$ in the variational formulation (4), we obtain a semi-discrete system of the form

$$M_C\frac{\mathrm{d}u}{\mathrm{d}t} = Ku + g, \tag{6}$$

where u is the vector of unknowns, $M_C=\{m_{ij}\}$ is the consistent mass matrix, $K=\{k_{ij}\}$ is the discrete transport operator and $g=\{g_i\}$ is a vector incorporating the boundary conditions.

Let $0=t^0<t^1<t^2<\cdots<t^M=T$ be a sequence of discrete time levels. Using the two-level θ-scheme for integration in time, we obtain the fully discrete problem

$$[M_C-\theta\Delta tK]u^{n+1} = [M_C+(1-\theta)\Delta tK]u^n + g^{n+\theta}, \tag{7}$$

where $\theta\in[0,1]$ is the degree of implicitness and $\Delta t = t^{n+1}-t^n$ is the time step. The forward Euler ($\theta=0$) method is unstable for convection-dominated transport problems and gives rise to severe time step restrictions in the case of dominating diffusion. For this reason, we restrict ourselves to the unconditionally stable Crank-Nicolson $\left(\theta=\frac{1}{2}\right)$ and backward Euler ($\theta=1$) time stepping.

3. ALGEBRAIC FLUX CORRECTION

To enforce sufficient conditions for the discrete maximum principle (if applicable) and/or positivity preservation, we will modify the standard Galerkin discretization (6) by adding some artificial diffusion. At the semi-discrete level, a numerical approximation of the form

$$M_L\frac{\mathrm{d}u}{\mathrm{d}t} = Lu + g \tag{8}$$

is positivity-preserving if $M_L=\mathrm{diag}\{m_i\}$ is a diagonal matrix with positive diagonal entries and all off-diagonal coefficients of $L=\{l_{ij}\}$ are nonnegative. If the diagonal coefficient is given by $l_{ii}=-\sum_{j\neq i}l_{ij}$, then the condition $l_{ij}\geq 0$ for $j\neq i$ is sufficient for the semi-discrete scheme to be local extremum diminishing (LED). After the discretization in time, the corresponding discrete maximum principle holds, perhaps under a CFL-like restriction for the time step [21, 25].

Suppose that the Galerkin discretization (6) admits an equivalent representation

$$M_L\frac{\mathrm{d}u}{\mathrm{d}t} = Lu + f(u) + g \tag{9}$$

such that the matrices M_L and L satisfy the conditions of positivity preservation

$$m_i>0, \qquad l_{ij}\geq 0, \quad j\neq i. \tag{10}$$

Then the possibly oscillatory antidiffusive part of (6) is represented by the vector

$$f(u) = (M_L - M_C)\frac{\mathrm{d}u}{\mathrm{d}t} - (L - K)u. \tag{11}$$

To cast (6) into the form (9), we introduce the lumped mass matrix

$$M_L = \mathrm{diag}\{m_i\}, \qquad m_i = \sum_{j \neq i} m_{ij}, \qquad m_{ij} = \int_\Omega \varphi_i \varphi_j \, \mathrm{d}\mathbf{x}$$

and an artificial diffusion operator $D = \{d_{ij}\}$ defined by [21, 27]

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

Adding D to the Galerkin transport operator K, we construct the modified transport operator

$$L = K + D$$

satisfying the nonnegativity condition (10) for the off-diagonal coefficients. For linear finite elements in 1D, this modification yields a first-order accurate upwind approximation [21, 27]. Other approaches to constructing D with desired properties are discussed in [6, 14, 25].

As shown in [25], the semi-discrete scheme (9) is positivity-preserving (LED) if its low-order counterpart (8) is positivity-preserving (LED) and the antidiffusive term f is LED. Let

$$u_i^{\min} = \min_{j \in \mathcal{N}(i)} u_j, \qquad u_i^{\max} = \max_{j \in \mathcal{N}(i)} u_j \tag{12}$$

denote the local minimum and maximum over the stencil $\mathcal{N}(i)$ of node i. Then $f_i(u)$ is of LED type if $f_i(u) \leq 0$ for $u = u_i^{\max}$ and $f_i(u) \geq 0$ for $u = u_i^{\min}$.

An oscillatory high-order scheme can be repaired by limiting $f(u)$ in a conservative manner. In edge-based algebraic flux correction schemes [14, 21, 30, 33, 37], the antidiffusive term $f(u)$ is decomposed into internodal fluxes. The flux from node j into node i is defined by

$$f_{ij} = \left(m_{ij}\frac{\mathrm{d}}{\mathrm{d}t} + d_{ij}\right)(u_i - u_j), \qquad j \neq i. \tag{13}$$

The flux f_{ji} has the same magnitude and opposite sign. The limited LED counterpart of

$$f_i(u) = \sum_{j \neq i} f_{ij} \tag{14}$$

is given by

$$\bar{f}_i(u) = \sum_{j \neq i} \alpha_{ij} f_{ij}, \tag{15}$$

where $\alpha_{ij} \in [0, 1]$ are correction factors such that $\alpha_{ji} = \alpha_{ij}$ and [25]

$$c_i^{\min}(u_i^{\min} - u_i) \leq \bar{f}_i(u) \leq c_i^{\max}(u_i^{\max} - u_i) \tag{16}$$

for some bounded coefficients $c_i^{\min} > 0$ and $c_i^{\max} > 0$. To assure that the high-order approximation is recovered whenever the unconstrained antidiffusive correction is LED, the correction procedure should be designed to guarantee that $\alpha_{ij} \approx 1$ whenever $\bar{f}_i = f_i$ satisfies (16).

In general, an algebraic flux correction scheme can be designed using a decomposition of the global vector $f(u)$ into subvectors $f^e(u)$ associated with sets of neighboring nodes (e.g., vertices of the same mesh element) and having zero sums. We refer the reader to [24, 25, 31, 32] for a presentation of element-based limiting techniques in which $\bar{f}$ is defined in terms of element-level matrix-vector products and assembled using traditional finite element data structures. In this development we consider edge-based approaches formulated in terms of fluxes.

The edge-based correction factors α_{ij} for pairs of antidiffusive fluxes (f_{ij}, f_{ji}) and element-based correction factors α_e for antidiffusive element contributions f^e can be defined as the minimum of nodal correction factors $\Phi_i \in [0,1]$ designed to enforce inequality constraints (16) at a given node. In the edge-based version (15), we set $\alpha_{ij} = \min\{\Phi_i, \Phi_j\}$. In accordance with the design principles formulated in [25], the nodal limiter function Φ_i should possess the following properties:

- $\Phi_i \in [0,1]$ depends continuously on the nodal values u_j, $j \in \mathcal{N}(i)$;
- $\Phi_i = 0$ at a local maximum ($u_i = u_i^{\max}$) or minimum ($u_i = u_i^{\min}$);
- $\Phi_i = 1$ if u_h is linear on the patch of elements Ω_i containing node i.

The first property is needed to secure the well-posedness of the modified discrete problem. Existing theory [2, 3] guarantees convergence of fixed-point iterations for the nonlinear system under the assumption that the limited antidiffusive term $\bar{f}(u)$ is a Lipschitz-continuous function of u.

The second property guarantees that the sum of antidiffusive element contributions to node i is local extremum diminishing by (16) (see Section 8). The third property is commonly referred to as *linearity preservation* [2, 20, 21, 26] and is essential for maintaining consistency of the constrained Galerkin scheme in applications to anisotropic diffusion problems [26]. In the context of hyperbolic conservation laws, linearity preservation is not mandatory but highly desirable because it implies that the high order of the unconstrained approximation is recovered in smooth regions.

4. GRADIENT-BASED NODAL LIMITERS

In this paper, we consider limiting techniques based on generalizations of one-dimensional linearity-preserving limiters for uniform meshes. Given the one-sided gradient approximations

$$\partial_x^- u_i = \frac{u_i - u_{i-1}}{h}, \qquad \partial_x^+ u_i = \frac{u_{i+1} - u_i}{h}, \tag{17}$$

the nodal jump $[\![\cdot]\!]$ and average $\{\cdot\}$ are defined by

$$[\![\partial_x u_i]\!] = \partial_x^+ u_i - \partial_x^- u_i, \qquad \{\partial_x u_i\} = \frac{\partial_x^+ u_i + \partial_x^- u_i}{2}. \tag{18}$$

A gradient-based limiter function Φ_i satisfying the above design criteria is given by

$$\Phi_i = \begin{cases} 1 - \frac{|[\![\partial_x u_i]\!]|}{2\{|\partial_x u_i|\}} & \text{if } \{|\partial_x u_i|\} \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

This limiter is demonstrated to be both LED and linearity-preserving by observing

1. The case $\{|\partial_x u_i|\} = (|\partial_x^+ u_i| + |\partial_x^- u_i|)/2 = 0$ corresponds to $u_{i-1} = u_i = u_{i+1}$ and thus $u_i = u_i^{\max} = u_i^{\min}$. Since $\Phi_i = 0$ by (19) the appropriate behavior of the limiter is obtained for this case.
2. Considering the case $\{|\partial_x u_i|\} = (|\partial_x^+ u_i| + |\partial_x^- u_i|)/2 > 0$
   - If $u_i = u_i^{\max}$ or $u_i = u_i^{\min}$ then the one sided gradients have opposite signs and it follows that
   $$\Phi_i = 1 - h\frac{|\partial_x^+ u_i - [-sign(\partial_x^+ u_i)]\partial_x^- u_i|}{|u_{i+1} - u_i| + |u_i - u_{i-1}|} = 1 - \frac{|u_{i+1} - u_i| + |u_i - u_{i-1}|}{|u_{i+1} - u_i| + |u_i - u_{i-1}|} = 0.$$
   - If $u_i \neq u_i^{\max}$ and $u_i \neq u_i^{\min}$ then the one sided gradients have the same sign and
   $$0 \leq \Phi_i = 1 - \frac{|u_{i+1} - 2u_i + u_{i-1}|}{|u_{i+1} - u_i| + |u_i - u_{i-1}|} \leq 1.$$
   That is, the value of Φ_i depends on the ratio of the discretized first and second derivatives at node i in this case.

Finally linearity preserving is demonstrated by observing that

- If $\{|\partial_x u_i|\} = 0$ then $u_i = u_i^{\max} = u_i^{\min}$ and again $\Phi_i = 0$ as is required.
- If $\{|\partial_x u_i|\} > 0$, then the one-sided gradients coincide and, therefore by (19), the above formula yields $\Phi_i = 1$.

Hence, the limiter function defined by (19) is both LED and linearity-preserving.

4.1. Generalizations to Multidimensions

In edge-based finite element schemes, slope limiting is commonly performed using reconstruction of one-dimensional stencils [30, 33, 34, 35, 36, 37]. This approach leads to straightforward generalizations of 1D slope limiters like (19) which may or may not guarantee the LED property, depending on the employed reconstruction procedure [34]. Badia and Hierro [1] generalize (19) using the maxima of nodal jumps and averages over all directional derivatives. While this definition complies with the three fundamental design principles (continuity, LED criterion, linearity preservation), the programming effort and computational cost associated with its use for computation of Φ_i on unstructured meshes seem to be rather high.

As an alternative generalization of (19), we consider the gradient-based smoothness sensor

$$\Phi_i = \begin{cases} 1 - \dfrac{\left|\int_{\partial\Omega_i} \mathbf{n}\cdot\nabla u_h \mathrm{d}s\right|}{\int_{\partial\Omega_i} |\mathbf{n}\cdot\nabla u_h|\mathrm{d}s} & \text{if } \int_{\partial\Omega_i} |\mathbf{n}\cdot\nabla u_h|\mathrm{d}s \neq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{20}$$

where $\mathbf{n}$ is the unit outward normal to the outer boundary $\partial\Omega_i$ of the element patch Ω_i containing node i. On a uniform mesh of 1D linear finite elements, definition (20) is equivalent to (19) and to the limiters considered in [1, 6]. In the multidimensional case, the integration of normal derivatives can be performed in a loop over elements, which makes (20) a handy and efficient alternative to generalizations based on nodal jumps and averages of directional derivatives. However this limiter is not LED for arbitrary meshes.

Suppose that u_h is linear on the patch Ω_i. By the divergence theorem and linearity of u_h, we have

$$\int_{\partial\Omega_i} \mathbf{n}\cdot\nabla u_h \mathrm{d}s = \int_{\Omega_i} \Delta u_h \mathrm{d}s = 0$$

and, therefore, $\Phi_i = 1$. Thus, the limiter function defined by (20) is linearity preserving.

For linear finite elements on simplex meshes, the unit outward normal $\mathbf{n}$ is given by

$$\mathbf{n} = -\frac{\nabla\varphi_i}{\|\nabla\varphi_i\|},$$

whence

$$\int_{\partial\Omega_i} \mathbf{n}\cdot\nabla u_h \mathrm{d}s = -\sum_j u_j \int_{\partial\Omega_i} \frac{\nabla\varphi_i\cdot\nabla\varphi_j}{\|\nabla\varphi_i\|}\mathrm{d}s = \sum_{j\neq i}(u_i - u_j)\int_{\partial\Omega_i} \frac{\nabla\varphi_i\cdot\nabla\varphi_j}{\|\nabla\varphi_i\|}\mathrm{d}s.$$

Under the angle conditions known from proofs of the discrete maximum principle for the Laplace operator [4, 8], we have $\nabla\varphi_i\cdot\nabla\varphi_j < 0$ for $j \neq i$. Suppose that the angle conditions hold and u_i is a local maximum ($u_i \geq u_j$ for all $j \neq i$) or minimum ($u_i \leq u_j$ for all $j \neq i$). Then we have

$$\left|\int_{\partial\Omega_i} \mathbf{n}\cdot\nabla u_h \mathrm{d}s\right| = \int_{\partial\Omega_i} |\mathbf{n}\cdot\nabla u_h|\mathrm{d}s$$

and, therefore, $\Phi_i = 0$. Thus, the limiter function defined by (20) is local extremum diminishing on simplex meshes satisfying the angle conditions.

To circumvent the angle conditions, one may replace (20) by the formula

$$\Phi_i = \begin{cases} 1 - \dfrac{\left|\int_{\partial S_i} (\mathbf{x}-\mathbf{x}_i)\cdot\nabla u_h \mathrm{d}s\right|}{\int_{\partial S_i} |(\mathbf{x}-\mathbf{x}_i)\cdot\nabla u_h|\mathrm{d}s}, & \text{if } \int_{\partial S_i} |(\mathbf{x}-\mathbf{x}_i)\cdot\nabla u_h|\mathrm{d}s \neq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{21}$$

where S_i is a sphere of radius $r_i = \min_j \|\mathbf{x}_j - \mathbf{x}_i\|$ centered at the point $\mathbf{x}_i$. This approach would guarantee both linearity preservation and the LED property on arbitrary meshes. An obvious drawback is the overhead cost associated with numerical integration over ∂S_i.

If linearity preservation is not essential, the LED constraint can be easily enforced using

$$\Phi_i = \begin{cases} 1 - \frac{\left|\int_{\partial\Omega_i} (\mathbf{x}-\mathbf{x}_i)\cdot\nabla u_h \mathrm{d}s\right|}{\int_{\partial\Omega_i} |(\mathbf{x}-\mathbf{x}_i)\cdot\nabla u_h| \mathrm{d}s}, & \text{if } \int_{\partial\Omega_i} |(\mathbf{x}-\mathbf{x}_i)\cdot\nabla u_h| \mathrm{d}s \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

Due to the fact that $(\mathbf{x} - \mathbf{x}_i) \cdot \nabla u_h = u_h(\mathbf{x}) - u_i$ for linear finite elements, this definition guarantees that $\Phi_i = 0$ at a local extremum. If u_h is linear on Ω_i, then we have

$$\int_{\partial\Omega_i} (\mathbf{x} - \mathbf{x}_i) \cdot \nabla u_h \mathrm{d}s = \nabla u_h \cdot \int_{\partial\Omega_i} (\mathbf{x} - \mathbf{x}_i) \mathrm{d}s.$$

Hence, the limiter defined by (22) is linearity preserving if

$$\mathbf{x}_i = \frac{1}{|\partial\Omega_i|} \int_{\partial\Omega_i} \mathbf{x} \mathrm{d}s.$$

As another multidimensional generalization of (19), consider the formula [2]

$$\Phi_i = \begin{cases} 1 - \frac{\left|\sum_{j\neq i}(u_i - u_j)\right|}{\sum_{j\neq i}|u_i - u_j|} & \text{if } \sum_{j\neq i} |u_i - u_j| \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

It is easy to verify that this limiter is LED on any mesh but the proof of linearity preservation requires that the triangulation be symmetric with respect to its internal nodes [2].

The above discussion indicates that it is easy to construct nodal limiters which are

- LED on regular meshes and linearity preserving on any mesh **or**
- LED on any mesh and linearity preserving on uniform meshes.

The main result of this paper is a new gradient-based nodal limiter which is linearity preserving **and** LED on arbitrary meshes. The proposed limiter function is defined by

$$\Phi_i = \begin{cases} 1 - \frac{\left|\sum_{j\neq i} w_{ij}(u_i - u_j - \delta u_{ij})\right|}{\sum_{j\neq i} w_{ij}|u_i - u_j|} & \text{if } \sum_{j\neq i} w_{ij}|u_i - u_j| \neq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{24}$$

where w_{ij} are nonnegative weights with $w_{ij} > 0$ for $j \in \mathcal{N}(i)$ and δu_{ij} is an approximation to $u_i - u_j$ which is exact for linear functions and vanishes at a local extremum. We define

$$\delta u_{ij} := \bar{\mathbf{g}}_i \cdot (\mathbf{x}_i - \mathbf{x}_j) \tag{25}$$

in terms of a limited nodal gradient $\bar{\mathbf{g}}_i$ such that $\bar{\mathbf{g}}_i = \nabla u_h(\mathbf{x}_i)$ for $u_h \in P_1(\Omega_i)$ and $\bar{\mathbf{g}}_i = \mathbf{0}$ for $u_i = u_i^{\max}$ or $u_i = u_i^{\min}$. To ensure the nonnegativity of Φ_i, we also require that

$$|u_i - u_j - \bar{\mathbf{g}}_i \cdot (\mathbf{x}_i - \mathbf{x}_j)| \leq |u_i - u_j| \qquad \forall j \neq i. \tag{26}$$

An algorithm for calculating $\bar{\mathbf{g}}_i$ satisfying the above requirements is presented in the next section.

If the numerical solution u_h is linear on the patch Ω_i, then we have

$$\delta u_{ij} = \nabla u_h(\mathbf{x}_i) \cdot (\mathbf{x}_i - \mathbf{x}_j) = u_i - u_j,$$

whence $\Phi_i = 1$. At a local extremum, we have $\bar{\mathbf{g}}_i = \mathbf{0}$ and, therefore $\delta u_{ij} = 0$. It follows that

$$\Phi_i = 1 - \frac{\left|\sum_{j\neq i} w_{ij}(u_i - u_j)\right|}{\sum_{j\neq i} w_{ij}|u_i - u_j|} = 0$$

because all solution differences $u_i - u_j,\ j \neq i$ have the same sign at a local extremum. Hence, the limiter defined by (24) is linearity-preserving and LED under the above assumptions regarding $\bar{\mathbf{g}}_i$.

Note that formula (23) corresponds to (24) with $w_{ij} = 1$ and $\delta u_{ij} = 0$ for all $j \neq i$. However with this choice of δu_{ij} the proof of linearity preservation restricts the mesh geometry to symmetric triangulations [2]. Our specific choice above of δu_{ij} and the condition placed on $\mathbf{g}_{ij}$ remove this restriction. In the numerical study below, we use the coefficients $w_{ij} = m_{ij}$ of the consistent mass matrix M_C as weights and define the linearity-preserving slope correction δu_{ij} using (25).

To reduce the amount of numerical dissipation, a linearity-preserving limiter may be configured to produce $\Phi_i = 1$ not only if u_h is linear on the patch Ω_i but also for approximations which are sufficiently close to a linear function on Ω_i. To this end, any nodal limiter of the form

$$\Phi_i = 1 - \frac{P_i}{Q_i} \tag{27}$$

with $0 \leq P_i \leq Q_i > 0$ can be modified as follows (cf. [25], Section 5.4):

$$\Phi_i = 1 - \frac{\max\{0, P_i - \beta Q_i\}}{(1-\beta)Q_i}, \qquad \beta \in [0, 1). \tag{28}$$

This modification preserves the property that $\Phi_i = 0$ for $P_i = Q_i$ and $\Phi_i = 1$ for $P_i = 0$.

Choosing a larger value of β makes the limiter less diffusive but increases the number of fixed-point iterations when it comes to solving the nonlinear system associated with the constrained Galerkin discretization. In the numerical study below, we use $\beta = \frac{3}{4}$. This setting yields a marked improvement of accuracy (compared to $\beta = 0$) without causing convergence problems.

5. CALCULATION OF LIMITED NODAL GRADIENTS

To define the limited nodal gradient $\bar{\mathbf{g}}_i$, we consider the continuous lumped-mass L^2 projection

$$\mathbf{g}_i := \frac{1}{m_i} \sum_j \mathbf{c}_{ij} u_j, \tag{29}$$

where m_i is a positive diagonal entry of the lumped mass matrix M_L, and $\mathbf{c}_{ij}$ is a vector-valued coefficient of the discrete gradient operator $\mathbf{C} = \{\mathbf{c}_{ij}\}$ defined by

$$\mathbf{c}_{ij} = \int_\Omega \varphi_i \nabla \varphi_j \, \mathrm{d}\mathbf{x}. \tag{30}$$

If u_h is linear on Ω, then the above reconstruction of $\mathbf{g}_i$ from ∇u_h is exact since

$$m_i = \sum_j m_{ij} = \int_\Omega \varphi_i \left(\sum_j \varphi_j \right) \mathrm{d}\mathbf{x} = \int_\Omega \varphi_i \, \mathrm{d}\mathbf{x}$$

and therefore

$$\mathbf{g}_i = \frac{1}{m_i} \int_\Omega \varphi_i \nabla u_h \, \mathrm{d}\mathbf{x} = \nabla u_h \left(\frac{1}{m_i} \int_\Omega \varphi_i \, \mathrm{d}\mathbf{x} \right) = \nabla u_h. \tag{31}$$

However, the recovered nodal gradient $\mathbf{g}_i$ may fail to satisfy the requirement that $\mathbf{g}_i = \mathbf{0}$ at a local extremum. Therefore, it may need to be limited in the same manner as reconstructed gradients in finite volume and discontinuous Galerkin methods [22, 23]. Introducing the correction factors

$$\psi_{ij} = \begin{cases} \min\left\{1, \frac{2(u_i - u_j)}{\mathbf{g}_i \cdot (\mathbf{x}_i - \mathbf{x}_j)}\right\} & \text{if } (u_i - u_j)\mathbf{g}_i \cdot (\mathbf{x}_i - \mathbf{x}_j) > 0, \\ 0 & \text{otherwise,} \end{cases}$$

we define

$$\Psi_i = \min_{j \in \mathcal{N}(i) \setminus \{i\}} \psi_{ij}, \qquad \bar{\mathbf{g}}_i = \Psi_i \mathbf{g}_i. \tag{32}$$

It is easy to verify that $\bar{\mathbf{g}}_i$ satisfies all design criteria formulated at the end of Section 4.

6. GRADIENT-BASED BACKGROUND DISSIPATION

To improve the phase accuracy of the constrained Galerkin approximation, some high-order dissipation may be added to the raw antidiffusive fluxes f_{ij} defined by (13). Following the approach used to construct background dissipation in [25], we generalize (13) as follows:

$$f_{ij} = \left(m_{ij} \frac{\mathrm{d}}{\mathrm{d}t} + d_{ij} \right) (u_i - u_j) + f_{ij}^{\mathrm{stab}}, \qquad j \neq i, \tag{33}$$

where

$$f_{ij}^{\mathrm{stab}} = \omega d_{ij} \left[\frac{\mathbf{g}_i + \mathbf{g}_j}{2} \cdot (\mathbf{x}_i - \mathbf{x}_j) - (u_i - u_j) \right]. \tag{34}$$

The amount of background dissipation is controlled using the blending factor $\omega \in [0, 1]$. In the case $\omega = 0$, the antidiffusive flux f_{ij} reduces to (13). Setting $\omega = 1$ corresponds to replacing $u_i - u_j$ by the smooth approximation in terms of the (unlimited) averaged gradients. If u_h is locally linear, then the nodal gradients are exact, whence $f_{ij}^{\mathrm{stab}} = 0$ for any value of ω. On a uniform mesh of 1D linear finite elements, the stabilizing flux f_{ij}^{stab} introduces fourth-order artificial dissipation [30].

7. LIMITING THE TIME DERIVATIVES

In applications to unsteady transport equations, we decompose the antidiffusive flux (33) into

$$f_{ij}^M = m_{ij} \frac{\mathrm{d}}{\mathrm{d}t} (u_i - u_j) \tag{35}$$

and

$$f_{ij}^K = d_{ij} (u_i - u_j) + f_{ij}^{\mathrm{stab}}. \tag{36}$$

Whereas the LED property can be enforced using the correction factor $\alpha_{ij} := \min\{\Phi_i, \Phi_j\}$ for both components, the flux f_{ij}^M may become dominant and produce significant phase errors. To balance f_{ij}^M and f_{ij}^K, we will handle f_{ij}^M using an edge-based version of the algorithm developed in [25].

Let $\dot{u}$ denote the vector of nodal time derivatives that corresponds to

$$\dot{u}^C = M_L^{-1} (Lu + g + \bar{f}), \tag{37}$$

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

$$\bar{f}_i = \sum_{j \neq i} (\min\{\alpha_{ij}, \beta_{ij}\} f_{ij}^M + \alpha_{ij} f_{ij}^K), \tag{38}$$

where $\beta_{ij} \in [0, 1]$ is a time derivative limiter to be defined below. Setting β_{ij} equal to zero, one obtains the lumped-mass approximation

$$\dot{u}^L = M_L^{-1} (Lu + g + \bar{f}^K), \qquad \bar{f}_i^K = \sum_{j \neq i} \alpha_{ij} f_{ij}^K \tag{39}$$

such that

$$\dot{u}^C = \dot{u}^L + M_L^{-1} \bar{f}^M, \qquad \bar{f}_i^M = \sum_{j \neq i} \min\{\alpha_{ij}, \beta_{ij}\} f_{ij}^M. \tag{40}$$

Hence, $\bar{f}^M$ can be interpreted as a high-order correction to $\dot{u}^L$. To constrain the changes of the time derivative due to this correction, we choose β_{ij} 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.,

$$\dot{u}_i^{\min} = \min_{j \in \mathcal{N}(i)} \dot{u}_j^L, \qquad \dot{u}_i^{\max} = \max_{j \in \mathcal{N}(i)} \dot{u}_j^L. \tag{41}$$

Substituting $\dot{u}^L$ for the time derivative in the raw antidiffusive flux

$$f_{ij}^M := m_{ij}(\dot{u}_i^L - \dot{u}_j^L),$$

we use the edge-based version [21, 30] of the multidimensional FCT algorithm

$$P_i^+ = \sum_{j\neq i} \max\{0, f_{ij}^M\}, \qquad P_i^- = \sum_{j\neq i} \min\{0, f_{ij}^M\}, \tag{42}$$

$$Q_i^+ = m_i(\dot{u}_i^{\max} - \dot{u}_i^L), \qquad Q_i^- = m_i(\dot{u}_i^{\min} - \dot{u}_i^L), \tag{43}$$

$$R_i^+ = \min\left\{1, \frac{Q_i^+}{P_i^+}\right\}, \qquad R_i^- = \min\left\{1, \frac{Q_i^-}{P_i^-}\right\} \tag{44}$$

to calculate the correction factors

$$\beta_{ij} = \begin{cases} \min\{R_i^+, R_j^-\} & \text{if } f_{ij}^M > 0, \\ \min\{R_i^-, R_j^+\} & \text{if } f_{ij}^M < 0. \end{cases} \tag{45}$$

8. PROOF OF THE LED PROPERTY

To verify the LED property, we need to show that estimates of the form (16) hold for the proposed choice of the correction factors $\alpha_{ij} = \min\{\Phi_i, \Phi_j\}$. At an interior node i, we have

$$\bar{f}_i = \sum_{j\neq i} (\min\{\alpha_{ij}, \beta_{ij}\} f_{ij}^M + \alpha_{ij} f_{ij}^K), \qquad 0 \le \alpha_{ij} \le \Phi_i,$$

where Φ_i is a nodal limiter such that $\Phi_i = 0$ at a local extremum. It follows that

$$\Phi_i f_i^- \le \bar{f}_i \le \Phi_i f_i^+, \tag{46}$$

where

$$f_i^+ = \sum_{j\neq i} \max\{0, f_{ij}^M\} + \sum_{j\neq i} \max\{0, f_{ij}^K\},$$

$$f_i^- = \sum_{j\neq i} \min\{0, f_{ij}^M\} + \sum_{j\neq i} \min\{0, f_{ij}^K\}.$$

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 $\bar{f}_i > 0$. Thus

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

In the case $\bar{f}_i < 0$, a similar argument yields a lower bound for the limited antidiffusive term

$$\bar{f}_i \ge \Phi_i f_i^- = \frac{\Phi_i f_i^-}{u_i^{\min} - u_i}(u_i^{\min} - u_i).$$

This proves the existence of the two-sided LED estimate (16) with coefficients

$$c_i^{\min} = \frac{\Phi_i f_i^-}{u_i^{\min} - u_i}, \qquad c_i^{\max} = \frac{\Phi_i f_i^+}{u_i^{\max} - u_i}.$$

By continuity of Φ_i, the values of $c_i^{\min}$ and $c_i^{\max}$ remain bounded as u_i approaches $u_i^{\min} = u_k$ or $u_i^{\max} = u_k$ for $k \neq i$. For any $\epsilon > 0$, substitution of $u_i = u_i^{\max} - \epsilon$ into (24) and evaluation of $c_i^{\max}$

yields

$$c_i^{\max} = \frac{f_i^+}{\epsilon} \frac{\sum_{j\neq i} w_{ij}|u_i^{\max} - u_j - \epsilon| - \left|\sum_{j\neq i} w_{ij}(u_i^{\max} - u_j - \epsilon - \delta u_{ij})\right|}{\sum_{j\neq i} w_{ij}|u_i - u_j|}. \tag{47}$$

Let $k \neq i$ be the number of a neighbor node at which the local maximum $u_k = u_i^{\max}$ is attained. By construction, the linearity-preserving gradient correction δu_{ij} satisfies

$$|\delta u_{ij}| = |\Psi_i \mathbf{g}_i \cdot (\mathbf{x}_i - \mathbf{x}_j)| \leq 2\epsilon\gamma_{ij}, \qquad \epsilon + \delta u_{ij} \leq \epsilon(1 + 2\gamma_{ij}),$$

$$\gamma_{ij} = \begin{cases} \frac{|\mathbf{g}_i \cdot (\mathbf{x}_i - \mathbf{x}_j)|}{|\mathbf{g}_i \cdot (\mathbf{x}_i - \mathbf{x}_k)|} & \text{if } |\mathbf{g}_i \cdot (\mathbf{x}_i - \mathbf{x}_k)| \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that for sufficiently small $\epsilon > 0$ we have

$$\left|\sum_{j\neq i} w_{ij}(u_i^{\max} - u_j - \epsilon - \delta u_{ij})\right| = \left|\sum_{j\neq i} w_{ij}(u_i^{\max} - u_j) - \sum_{j\neq i} w_{ij}(\epsilon + \delta u_{ij})\right|$$

$$= \sum_{j\neq i} w_{ij}(u_i^{\max} - u_j) - \sum_{j\neq i} w_{ij}(\epsilon + \delta u_{ij}).$$

The first sum in the numerator of (47) can be estimated thus:

$$\sum_{j\neq i} w_{ij}|u_i^{\max} - u_j - \epsilon| \leq \sum_{j\neq i} w_{ij}(u_i^{\max} - u_j) + \epsilon \sum_{j\neq i} w_{ij},$$

whence

$$c_i^{\max} \leq 2f_i^+ \frac{\sum_{j\neq i} w_{ij}(1 + \gamma_{ij})}{\sum_{j\neq i} w_{ij}|u_i - u_j|}. \tag{48}$$

The proof of boundedness for the LED coefficient $c_i^{\min}$ is similar.

9. TIME DISCRETIZATION

The constrained semi-discrete finite element scheme is given by the nonlinear ODE system

$$M_L \frac{\mathrm{d}u}{\mathrm{d}t} = Lu + g + \bar{f}. \tag{49}$$

We discretize this system in time using the two-level θ method

$$M_L \frac{u^{n+1} - u^n}{\Delta t} = \theta(Lu^{n+1} + g^{n+1} + \bar{f}^{n+1}) + (1 - \theta)(Lu^n + g^n + \bar{f}^n). \tag{50}$$

In the case $\theta = 1$, the fully discrete scheme is unconditionally positivity-preserving by the M-matrix property of the low-order operator $M_L + \Delta t L$ and definition of $\bar{f}$. In the case of $\theta < 1$, the fully discrete scheme is positivity preserving if the coefficient associated with the nodal value u_i^n is nonnegative [21, 25]. This requirement leads to the CFL-like positivity condition

$$\frac{1}{\Delta t} \geq (1 - \theta) \left[\sum_{j\neq i} l_{ij} + c_i\right] \qquad \forall i = 1, \ldots, N, \tag{51}$$

where $c_i = \max\{c_i^{\min}, c_i^{\max}\}$ is defined by the above analysis of the LED property for $\bar{f}_i$.

We remark that the proposed limiting strategy can also be used in conjunction with other time discretizations including strong stability preserving (SSP) Runge-Kutta schemes [13].

Due to the dependence of the correction factors α_{ij} and β_{ij} on the unknown solution, the algebraic system (50) is nonlinear. It can be solved using the fixed-point iteration

$$u^{(m+1)} = u^{(m)} + \left[\frac{1}{\Delta t} M_L - \theta L\right]^{-1} r^{(m)}, \qquad m = 0, 1, 2, \dots \tag{52}$$

$$r^{(m)} = \theta(Lu^{(m)} + g^{n+1} + \bar{f}^{(m)}) + (1-\theta)(Lu^n + g^n + \bar{f}^n) - M_L \frac{u^{(m)} - u^n}{\Delta t}. \tag{53}$$

In applications to anisotropic diffusion problems, the rates of convergence to steady state solutions can be greatly improved using Anderson acceleration for fixed-point iterations [20, 38].

10. NUMERICAL EXAMPLES

In this section, we perform a numerical study of algebraic flux corrections schemes equipped with the gradient-based limiters (20) and (24) (to be referred to as GL1 and GL2, respectively) for pure convection and anisotropic diffusion problems discretized using linear finite elements on uniform and nonuniform triangular meshes. Given a uniform grid with spacing h, its distorted counterpart is generated by applying random perturbations to the Cartesian coordinates of internal nodes

$$x_i := x_i + \alpha h \xi_i \qquad y_i := y_i + \alpha h \eta_i, \tag{54}$$

where $\xi_i, \eta_i \in [-0.5, 0.5]$ are random numbers. The parameter $\alpha \in [0, 1]$ quantifies the degree of distortion. In this numerical study, we use $\alpha = 0.75$ to generate grid deformations strong enough to violate the angle conditions under which the GL1 limiter satisfies the LED principle.

Given a reference solution u, we measure the errors in numerical approximations u_h on successively refined meshes using the discrete L^1 norm defined by [19, 20, 21]

$$E_1(h) := \sum_i m_i |u(\mathbf{x}_i) - u_i| \approx \int_\Omega |u - u_h| \, \mathrm{d}\mathbf{x} =: \|u - u_h\|_1, \tag{55}$$

where $m_i = \int_\Omega \varphi_i \, \mathrm{d}\mathbf{x}$ is a diagonal coefficient of the lumped mass matrix M_L. The convergence behavior of GL1 and GL2 is illustrated by the experimental order of convergence [28]

$$p = \log_2 \left(\frac{E_1(2h)}{E_1(h)} \right). \tag{56}$$

Additionally, the largest and smallest nodal values are presented for each mesh to detect possible violations of the discrete maximum principle (for GL1) and quantify numerical dissipation.

10.1. Solid body rotation

The solid body rotation benchmark [28] is a standard test for numerical advection schemes. Its use in this numerical study enables direct comparison with other types of algebraic flux correction [19, 20, 21, 25] and residual-based shock capturing techniques [18]. The governing equation (1) with $\mathbf{v}(x, y) = (0.5 - y, x - 0.5)$ and $\mathcal{D} = 0$ is solved in $\Omega = (0, 1) \times (0, 1)$ subject to homogeneous Dirichlet boundary conditions on the inflow boundary $\Gamma_1 = \{(x, y) \in \Gamma \, : \, \mathbf{v} \cdot \mathbf{n} < 0\}$. The exact solution u corresponds to counterclockwise rotation of the initial profile u_0 shown in Fig. 1(a) about the point $(0.5, 0.5)$. For a detailed description of this benchmark, we refer to [18, 21, 28].

Figures 1(b)-(c) display numerical solutions at the final time $T = 2\pi$ calculated using the low-order method ($\alpha_{ij} = 0$ for all $j \neq i$) and GL2 on triangular meshes consisting of $32,768$ linear elements. The results for GL1 are similar (not shown here). However, the violation of angle conditions for GL1 gives rise to significant undershoots and overshoots at early stages of computation (see Fig. 2). The convergence history for GL1 and GL2 is presented in Tables I-IV. The GL2 solutions satisfy the discrete maximum principle on all meshes, and the values of E_1 tend

to be smaller that those for GL1. The nodal limiter based on (21) is LED on all meshes but is less accurate than GL1 and GL2 and does not guarantee linearity preservation on general meshes. For this reason, it represents a usable but inferior alternative to the methods considered in this study.

The results for Q_1 finite element approximations are similar. The solutions calculated using GL2 on uniform and perturbed meshes of 128×128 rectangular elements are shown in Fig. 3.

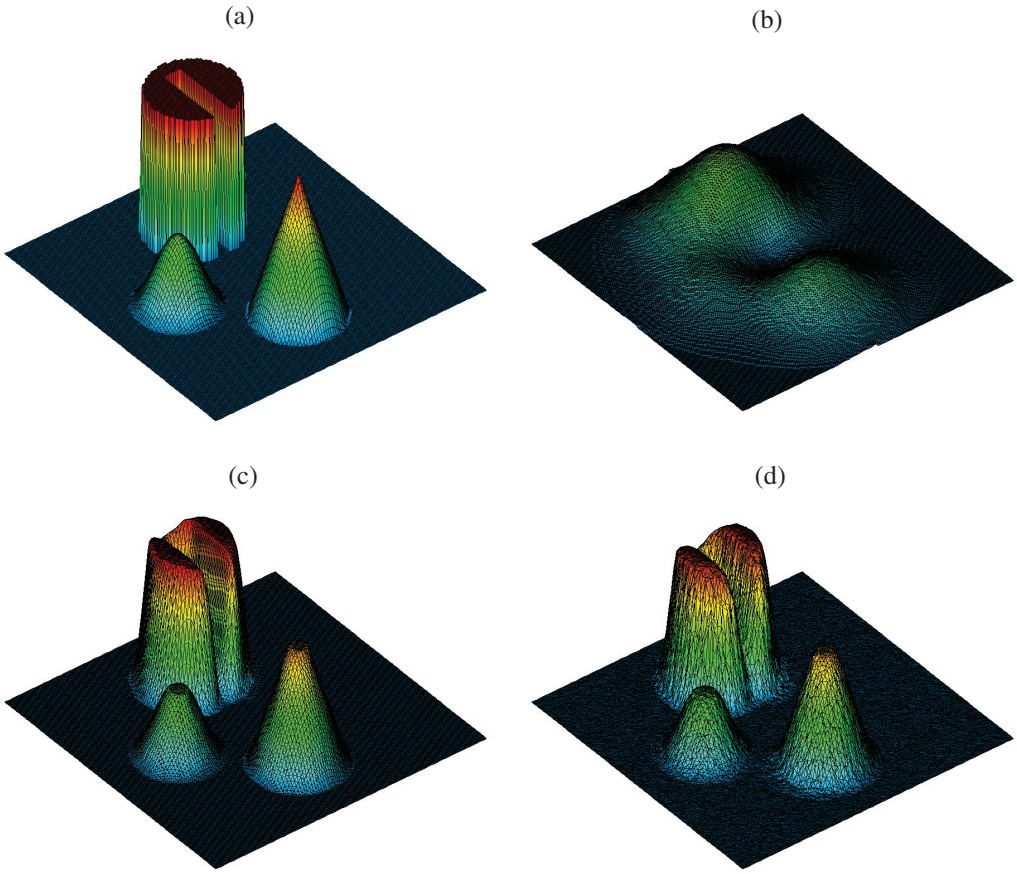

Figure 1. Solid body rotation: (a) initial data / exact solution, uniform mesh, (b) low-order solution, uniform mesh, (c) GL2 solution, uniform triangular mesh, (d) GL2 solution, perturbed triangular mesh, Discretization: $2 \times 128 \times 128$ $\mathcal{P}_1$ elements, Crank-Nicolson time-stepping, $\Delta t = 10^{-3}$, $T = 2\pi$.

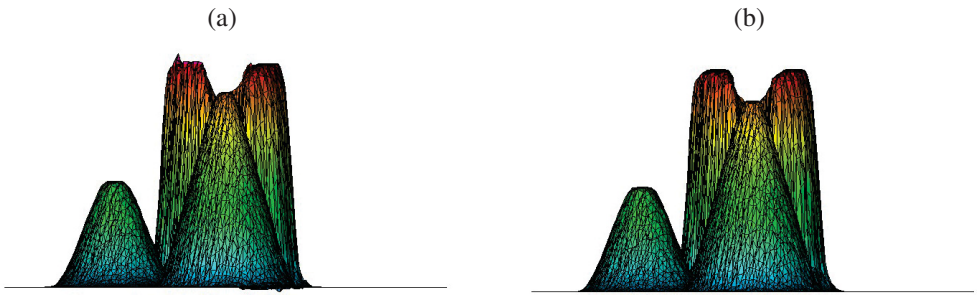

Figure 2. Solid body rotation: (a) GL1 vs. (b) GL2 on the perturbed triangular mesh at $T = \frac{\pi}{2}$.

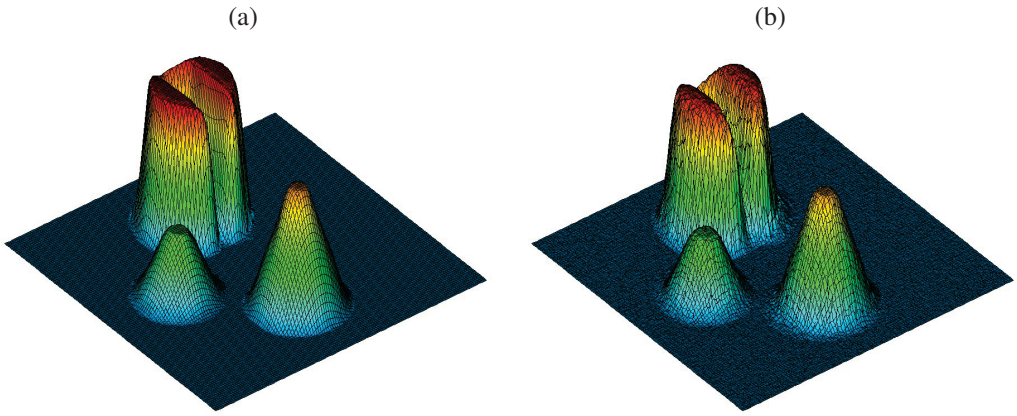

Figure 3. Solid body rotation: GL2 on (a) uniform and (b) perturbed quadrilateral meshes. Discretization: 128×128 $\mathcal{Q}_1$ elements, Crank-Nicolson time-stepping, $\Delta t = 10^{-3}$, $T = 2\pi$.

h	E_1	EOC	$\min u_h$	$\max u_h$
1/32	0.589e-01		0.0	0.748
1/64	0.374e-01	0.66	0.0	0.852
1/128	0.180e-01	1.06	0.0	0.996
1/256	0.964e-02	0.90	0.0	1.0

Table I. Solid body rotation: GL1, uniform mesh, $T = 2\pi$.

h	E_1	EOC	$\min u_h$	$\max u_h$
1/32	0.679e-01		-9.286e-3	0.699
1/64	0.429e-01	0.66	-4.247e-3	0.839
1/128	0.254e-01	0.76	-1.104e-3	0.978
1/256	0.147e-01	0.79	-1.170e-3	0.999

Table II. Solid body rotation: GL1, perturbed mesh, $T = 2\pi$.

h	E_1	EOC	$\min u_h$	$\max u_h$
1/32	0.578e-01		0.0	0.763
1/64	0.371e-01	0.64	0.0	0.859
1/128	0.181e-01	1.04	0.0	0.996
1/256	0.961e-02	0.91	0.0	1.0

Table III. Solid body rotation: GL2, uniform mesh, $T = 2\pi$.

h	E_1	EOC	$\min u_h$	$\max u_h$
1/32	0.675e-01		0.0	0.678
1/64	0.430e-01	0.65	0.0	0.799
1/128	0.247e-01	0.80	0.0	0.964
1/256	0.139e-01	0.83	0.0	0.999

Table IV. Solid body rotation: GL2, perturbed mesh, $T = 2\pi$.

10.2. Circular convection

In the second test, we solve the steady convection equation $\nabla \cdot (\mathbf{v}u) = 0$ with $\mathbf{v}(x, y) = (y, -x)$ in $\Omega = (0, 1) \times (0, 1)$. The exact solution is constant along the circular streamlines. The inflow boundary condition and the exact solution at any point in $\bar{\Omega}$ are defined by

$$u(x,y) = \begin{cases} 1, & \text{if } 0.15 \le r(x,y) \le 0.45, \\ \cos^2\left(10\pi \frac{r(x,y)-0.5}{3}\right), & \text{if } 0.55 \le r(x,y) \le 0.85, \\ 0, & \text{otherwise}, \end{cases}$$

where $r(x, y) = \sqrt{x^2 + y^2}$ denotes the distance to the corner point $(0, 0)$.

Stationary numerical solutions calculated using $32,768$ linear elements are presented in Fig. 4. The results produced by GL1 and GL2 look alike but the GL1 solutions exhibit undershoots and overshoots on perturbed meshes. The low-order solution is bounded by 0 and 1 but strongly smeared by numerical diffusion. The range of solution values, discrete L^1 errors, and convergence rates for the two gradient-based limiters on uniform and perturbed meshes are presented in Tables V–VIII.

h	E_1	EOC	$\min u_h$	$\max u_h$
1/32	0.271e-01		0.0	1.0
1/64	0.114e-01	1.25	0.0	1.0
1/128	0.515e-02	1.15	0.0	1.0
1/256	0.269e-02	0.94	0.0	1.0

Table V. Circular convection: GL1, uniform mesh.

h	E_1	EOC	$\min u_h$	$\max u_h$
1/32	0.285e-01		-3.579e-2	1.022
1/64	0.116e-01	1.30	-4.647e-2	1.004
1/128	0.632e-02	0.88	-1.842e-2	1.055
1/256	0.345e-02	0.87	-3.351e-2	1.045

Table VI. Circular convection: GL1, perturbed mesh.

h	E_1	EOC	$\min u_h$	$\max u_h$
1/32	0.248e-01		0.0	1.0
1/64	0.111e-01	1.16	0.0	1.0
1/128	0.507e-02	1.13	0.0	1.0
1/256	0.265e-02	0.94	0.0	1.0

Table VII. Circular convection: GL2, uniform mesh.

h	E_1	EOC	$\min u_h$	$\max u_h$
1/32	0.335e-01		0.0	1.0
1/64	0.148e-01	1.18	0.0	1.0
1/128	0.722e-02	1.04	0.0	1.0
1/256	0.394e-02	0.87	0.0	1.0

Table VIII. Circular convection: GL2, perturbed mesh.

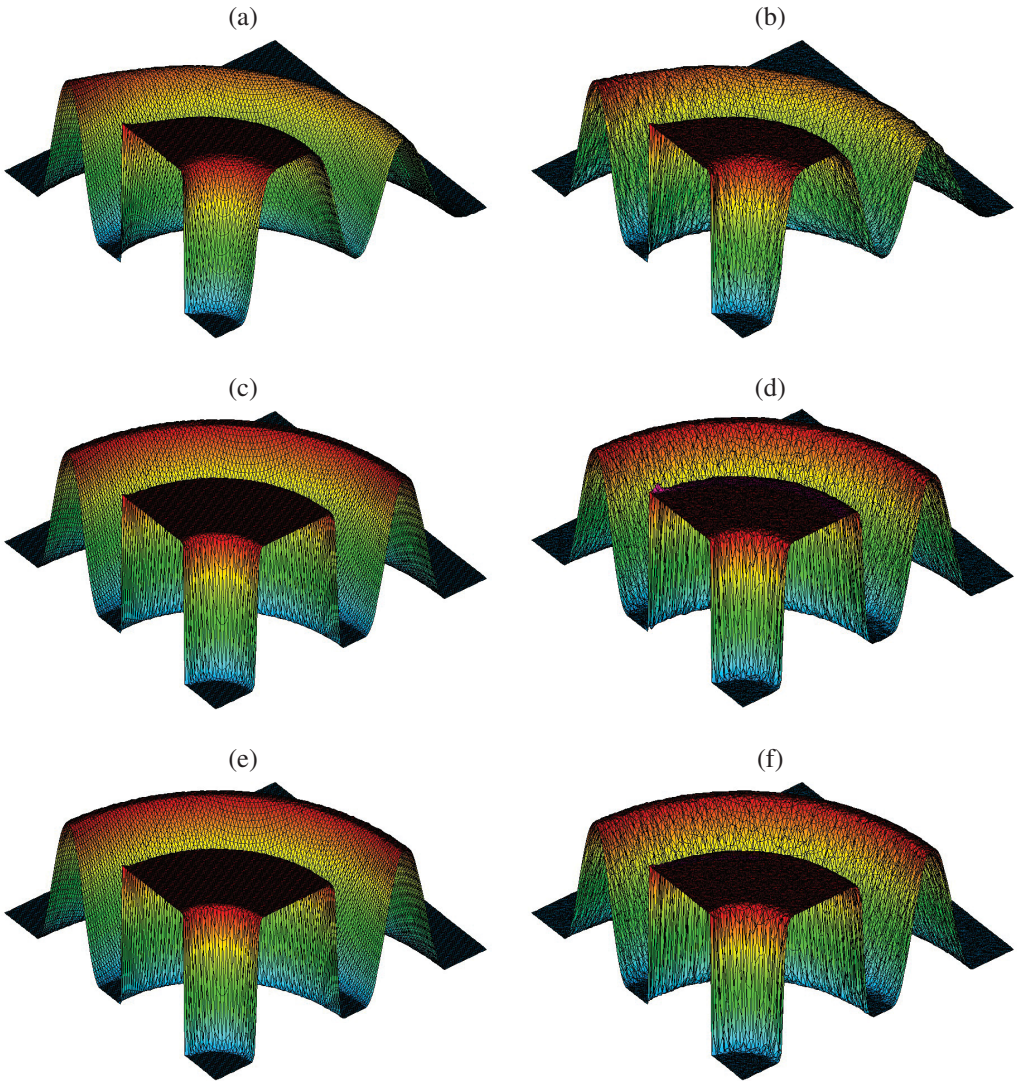

Figure 4. Circular convection: (a) low-order solution, uniform mesh, (b) low-order solution, perturbed mesh, (c) GP1 solution, uniform mesh, (d) GP1 solution, perturbed mesh, (e) GP2 solution, uniform mesh, (f) GP2 solution, perturbed mesh. Discretization: $2 \times 128 \times 128$ $\mathcal{P}_1$ elements.

11. ANISOTROPIC DIFFUSION

In the third example, we consider a steady diffusion equation of the form $-\nabla \cdot (\mathcal{D}\nabla u) = 0$. The domain $\Omega = (0,1)^2 \backslash [4/9, 5/9]^2$ is a square with a square hole in the middle [20, 26, 29]. The outer and inner boundary of Ω are denoted by Γ_0 and Γ_1, respectively (see Fig. 5(a)).

The following Dirichlet boundary conditions are prescribed in this test:

$$u(x,y) = \begin{cases} -1 & \text{if } (x,y) \in \Gamma_0, \\ 1 & \text{if } (x,y) \in \Gamma_1. \end{cases} \tag{57}$$

The diffusion tensor $\mathcal{D}$ is a symmetric positive definite matrix defined as

$$\mathcal{D} = \mathcal{R}(-\theta) \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix} \mathcal{R}(\theta), \tag{58}$$

where k_1 and k_2 are the positive eigenvalues and $\mathcal{R}(\theta)$ is a rotation matrix

$$\mathcal{R}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{59}$$

The eigenvalues of $\mathcal{D}$ represent the diffusion coefficients associated with the axes of the Cartesian coordinate system rotated by the angle θ. In this numerical example, we use

$$k_1 = 100, \qquad k_2 = 1, \qquad \theta = -\frac{\pi}{6}.$$

By the continuous maximum principle, the exact solution to the Dirichlet problem is bounded by the prescribed boundary data. However, the diffusion tensor (58) is highly anisotropic, which may result in a violation of the DMP even if a mesh satisfying the angle conditions is employed.

Since no exact solution is available, the reference solution depicted in Fig. 5(b) was calculated using the standard Galerkin method on a fine mesh ($h = 1/1152$). Even this solution has a small undershoot ($\min_\Omega u_h = -1.0015$). The solutions produced by the unconstrained Galerkin method and its GL2-constrained counterpart on uniform triangular meshes with spacing $h = 1/144$ are displayed in Fig. 5(c),(d). The Galerkin solution has an undershoot of 1.8%, whereas the GL2 solution is within the range $[-1, 1]$ of admissible values. The results of a grid convergence study for both methods are summarized in Tables IX and X. As the mesh is refined, the undershoots produced by the unconstrained Galerkin method become smaller. The GL2-constrained version exhibits similar rates of convergence and the discrete maximum principle holds on all meshes.

Interestingly enough, even the limiter based on (23) produces reasonable results on perturbed meshes on which it is not provably linearity preserving. The corresponding solutions (labeled GL3) are compared to GL2 solutions in Fig. 6. For a better comparison, the weights $w_{ij} = 1$ were used in the GL2 version based on (24). The GL2 results are less diffusive but the use of limited gradient reconstruction makes it more difficult to achieve convergence of the nonlinear solver.

h	E_1	EOC	$\min_\Omega u_h$	$\max_\Omega u_h$
1/36	0.513e-01		-1.055	1.0
1/72	0.299e-01	0.78	-1.039	1.0
1/144	0.157e-01	0.93	-1.018	1.0
1/288	0.712e-02	1.14	-1.001	1.0

Table IX. Anisotropic diffusion: Galerkin, uniform mesh.

h	E_1	EOC	$\min_\Omega u_h$	$\max_\Omega u_h$
1/36	0.430e-01		0.0	1.0
1/72	0.253e-01	0.77	0.0	1.0
1/144	0.143e-01	0.82	0.0	1.0
1/288	0.723e-02	0.98	0.0	1.0

Table X. Anisotropic diffusion: GL2, uniform mesh.

12. CONCLUSIONS

The proposed generalizations of a one-dimensional slope limiter based on jumps and averages of one-sided nodal gradients lead to robust high-resolution finite element schemes for stationary and time-dependent transport problems. In particular, the use of limited averaged gradients makes it possible to enforce the discrete maximum principle and guarantee linearity preservation on arbitrary meshes. An element-based version of the presented algebraic flux correction scheme can

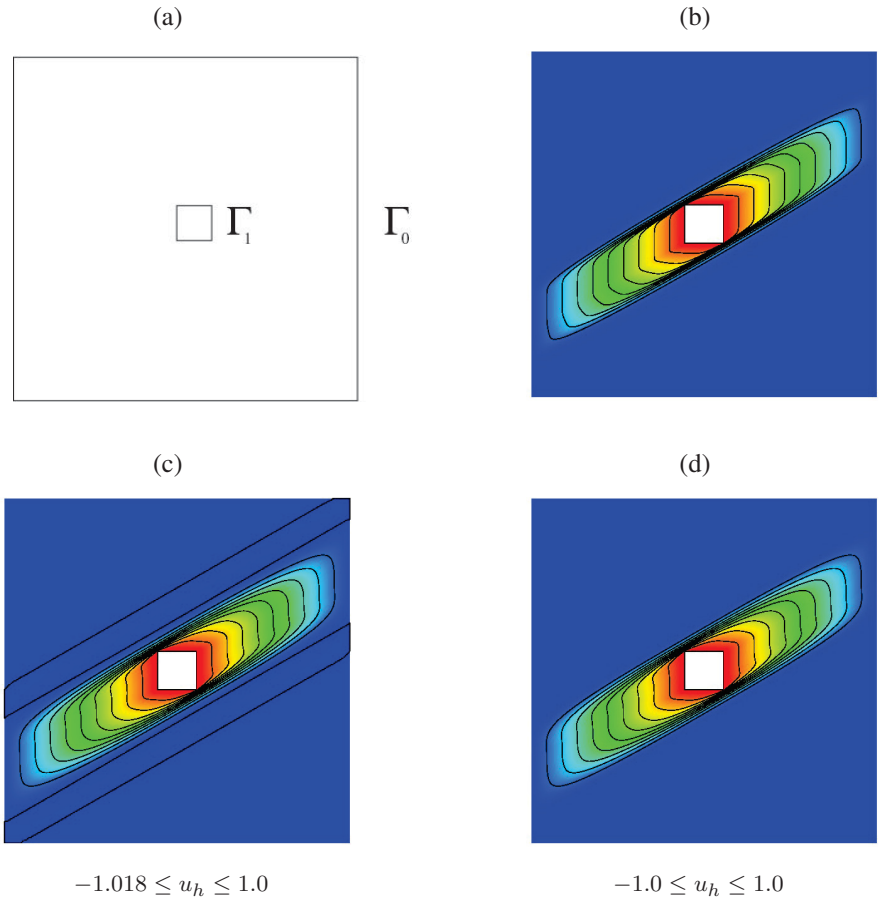

Figure 5. Anisotropic diffusion on a uniform triangular mesh: (a) computational domain, (b) Galerkin solution, $h = 1/1152$, (c) Galerkin solution, $h = 1/144$, (d) GP2 solution, $h = 1/144$.

be constructed using the new limiting strategy to constrain antidiffusive element contributions in the algorithm presented in [25]. The relationship to gradient-based limiters developed in [2, 6] may be exploited to provide further theoretical justification, e.g., by proving Lipschitz continuity or modifying the new limiters in a way which makes it possible to prove Lipschitz continuity.

ACKNOWLEDGEMENT

This research of D. Kuzmin was supported by the German Research Association (DFG) under grant KU 1530/15-1. The work of J. N. Shadid was partially supported by the DOE Office of Science Applied Mathematics Program at Sandia National Laboratories under contract DE-AC04-94AL85000.

REFERENCES

1. S. Badia and A. Hierro, On monotonicity-preserving stabilized finite element approximations of transport problems. *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.

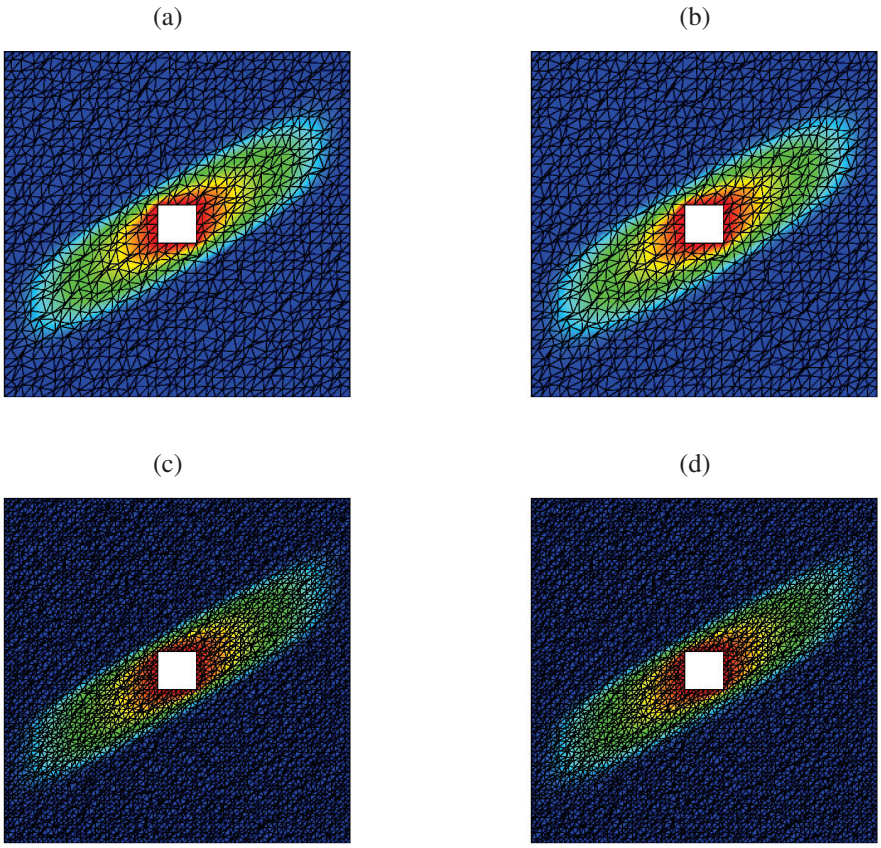

Figure 6. Anisotropic diffusion on perturbed triangular meshes: (a) GL2, 2,560 cells, (b) GL3, 2,560 cells, (c) GL2, 10,240 cells, (d) GL3, 10,240 cells.

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. I. Faragó, R. Horváth, S. Korotov, Discrete maximum principle for linear parabolic problems solved on hybrid meshes. *Appl. Numer. Math.* **53** (2005) 249–264.
12. 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.
13. S. Gottlieb, D. Ketcheson, C.-W. Shu, *Strong Stability Preserving Runge-Kutta and Multistep Time Discretizations*. World Scientific, 2011.
14. 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.
15. A. Harten, High resolution schemes for hyperbolic conservation laws. *J. Comput. Phys.* **49** (1983) 357–393.
16. A. Harten, On a class of high resolution total-variation-stable finite-difference-schemes. *SIAM J. Numer. Anal.* **21** (1984) 1-23.
17. 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.
18. 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.
19. D. Kuzmin, Explicit and implicit FEM-FCT algorithms with flux linearization. *J. Comput. Phys.* **228** (2009) 2517-2534.

20. D. Kuzmin, Linearity-preserving flux correction and convergence acceleration for constrained Galerkin schemes. *J. Comput. Appl. Math.* **236** (2012) 2317–2337.
21. 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.
22. D. Kuzmin, A vertex-based hierarchical slope limiter for p-adaptive discontinuous Galerkin methods. *J. Comput. Appl. Math.* **233** (2010) 3077–3085.
23. D. Kuzmin, Hierarchical slope limiting in explicit and implicit discontinuous Galerkin methods. *J. Comput. Phys.* **257** (2014) 1140–1162.
24. D. Kuzmin and J. Hämäläinen, *Finite Element Methods for Computational Fluid Dynamics: A Practical Guide*. SIAM, 2014, ISBN 978-1-611973-60-0.
25. D. Kuzmin and J.N. Shadid, A new approach to enforcing discrete maximum principles in continuous Galerkin methods for convection-dominated transport equations. Preprint: *Ergebnisberichte des Instituts für Angewandte Mathematik* **529** Department of Mathematics, TU Dortmund University, 2015. Submitted to *J. Comput. Phys.*
26. D. Kuzmin, M.J. Shashkov, and D. Svyatskiy, A constrained finite element method satisfying the discrete maximum principle for anisotropic diffusion problems. *J. Comput. Phys.* **228** (2009) 3448-3463.
27. D. Kuzmin and S. Turek, Flux correction tools for finite elements. *J. Comput. Phys.* **175** (2002) 525–558.
28. R.J. LeVeque: High-resolution conservative algorithms for advection in incompressible flow. *SIAM J. Numer. Anal.* **33** (1996) 627–665.
29. K. Lipnikov, M. Shashkov, D. Svyatskiy, Yu. Vassilevski: Monotone finite volume schemes for diffusion equations on unstructured triangular and shape-regular polygonal meshes. *J. Comput. Phys.* **227** (2007) 492–512.
30. R. Löhner, *Applied CFD Techniques: An Introduction Based on Finite Element Methods*. John Wiley & Sons, 2nd edition, 2008.
31. 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.
32. 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.
33. 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.
34. P.R.M. Lyra, *Unstructured Grid Adaptive Algorithms for Fluid Dynamics and Heat Conduction*. PhD thesis, University of Wales, Swansea, 1994.
35. 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.
36. 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.
37. V. Selmin, Finite element solution of hyperbolic equations. II. Two-dimensional case. *INRIA Research Report* **708**, 1987.
38. H.W. Walker, P. Ni, Anderson acceleration for fixed-point iterations. *SIAM J. Numer. Anal.* **49** (2011) 1715–1735.
39. S.T. Zalesak, Fully multidimensional flux-corrected transport algorithms for fluids. *J. Comput. Phys.* **31** (1979) 335–362.