

No. 624

January 2020

**Locally bound-preserving enriched Galerkin
methods for the linear advection equation**

D. Kuzmin, H. Hajduk, A. Rupp

ISSN: 2190-1767

Locally bound-preserving enriched Galerkin methods for the linear advection equation

Dmitri Kuzmin^a, Hennes Hajduk^a, Andreas Rupp^b

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

^b*Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg,
Im Neuenheimer Feld 205, 69120 Heidelberg Germany*

Abstract

In this work, we introduce algebraic flux correction schemes for enriched ($\mathbb{P}_1 \oplus \mathbb{P}_0$ and $\mathbb{Q}_1 \oplus \mathbb{P}_0$) Galerkin discretizations of the linear advection equation. The piecewise-constant component stabilizes the continuous Galerkin approximation without introducing free parameters. However, violations of discrete maximum principles are possible in the neighborhood of discontinuities and steep fronts. To keep the cell averages and the degrees of freedom of the continuous $\mathbb{P}_1/\mathbb{Q}_1$ component in the admissible range, we limit the fluxes and element contributions, the complete removal of which would correspond to first-order upwinding. The first limiting procedure that we consider in this paper is based on the flux-corrected transport (FCT) paradigm. It belongs to the family of predictor-corrector algorithms and requires the use of small time steps. The second limiting strategy is monolithic and produces nonlinear problems with well-defined residuals. This kind of limiting is well suited for stationary and time-dependent problems alike. The need for inverting consistent mass matrices in explicit strong stability preserving Runge-Kutta time integrators is avoided by reconstructing nodal time derivatives from cell averages. Numerical studies are performed for standard 2D test problems.

Keywords: linear advection equation, enriched Galerkin method, discrete maximum principles, flux-corrected transport, convex limiting

Email addresses: kuzmin@math.uni-dortmund.de (Dmitri Kuzmin),
hennes.hajduk@math.tu-dortmund.de (Hennes Hajduk), andreas@rupp.ink (Andreas Rupp)

1. Introduction

The locally conservative enriched Galerkin (EG) method considered in this paper adds a piecewise-constant \mathbb{P}_0 component to a continuous piecewise-linear (\mathbb{P}_1) or multilinear (\mathbb{Q}_1) finite element discretization of the linear advection equation. The $\mathbb{P}_1 \oplus \mathbb{P}_0$ version of this method was introduced by Becker et al. [7] who proved that it is stable and converges at the same rate as the \mathbb{P}_1 discontinuous Galerkin (DG- \mathbb{P}_1) method. At the same time, the EG approach requires less degrees of freedom in the multidimensional case. The \mathbb{P}_0 enrichment of the \mathbb{P}_1 continuous Galerkin (CG- \mathbb{P}_1) approximation renders it intrinsically stable and provides the local conservation property. These important advantages of the EG approach were recognized by Sun and Liu [37] who extended it to mathematical models of coupled flow and transport processes in porous media. An efficient EG method for general elliptic and parabolic problems was designed and analyzed by Lee et al. [30]. The possibility of enriching the CG- \mathbb{P}_1 or CG- \mathbb{Q}_1 space with discontinuous quadratic edge bubbles was explored in [6, 8, 10]. The resulting approximations are continuous at the vertices of the mesh and discontinuous across the edges. They are as stable and accurate as DG- \mathbb{P}_2 but not locally conservative.

To enforce the validity of discrete maximum principles, numerical methods for conservation laws are commonly equipped with flux or slope limiters. Recent years have witnessed significant advances in the analysis and design of algebraic flux correction (AFC) tools for nonlinear high-resolution finite element schemes [3, 5, 12, 25, 32]. However, no bound-preserving limiters are currently available for locally conservative enriched Galerkin approximations, which restricts their practical applicability to problems with smooth solutions. In the present paper, we constrain the EG discretization of the linear advection equation in a way which guarantees preservation of local bounds for the cell averages and for the nodal values of the CG component. The evolution of cell averages is governed by a discrete conservation law which we equip with a flux limiter. The subproblem for the CG degrees of freedom is equipped with a limiter for antidiffusive element contributions.

This paper is organized as follows. In Section 2, we present the unconstrained EG scheme for our linear advection model. Semi-discrete subproblems are formulated for the cell averages and CG unknowns. The inversion of the consistent mass matrix is avoided using an inexpensive reconstruction of nodal time derivatives from cell averages. In Section 3, we perform first-order upwinding to identify the numerical fluxes and element contributions

that require limiting. In [Section 4](#), we present a predictor-corrector limiting strategy based on the flux-corrected transport (FCT) algorithms proposed in [\[9, 34\]](#). In [Section 5](#), we derive a monolithic version of the nonlinear EG approximation building on the convex limiting framework developed in [\[27\]](#). Limiters for L^2 projections into CG and DG spaces are presented in [Section 6](#). In [Section 7](#), we perform numerical studies for stationary and time-dependent advection problems. Preliminary conclusions are drawn in [Section 8](#).

2. Enriched Galerkin method

Let $u(\mathbf{x}, t)$ be a scalar conserved quantity advected by a given solenoidal velocity field $\mathbf{v}(\mathbf{x})$ in a bounded domain $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$ with Lipschitz boundary $\Gamma = \partial\Omega$. Consider the initial-boundary value problem

$$\frac{\partial u}{\partial t} + \nabla \cdot (\mathbf{v}u) = 0 \quad \text{in } \Omega \times \mathbb{R}_+, \quad (1a)$$

$$u(\cdot, 0) = u_0 \quad \text{in } \Omega, \quad (1b)$$

$$(u - u_{\text{in}})\mathbf{v} \cdot \mathbf{n} = 0 \quad \text{on } \Gamma_-, \quad (1c)$$

where u_0 is the initial data, u_{in} is the boundary data, \mathbf{n} is the unit outward normal, and $\Gamma_- = \{\mathbf{x} \in \Gamma : \mathbf{v}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) < 0\}$ is the inflow boundary.

We discretize the linear advection equation [\(1a\)](#) in space using a conforming mesh \mathcal{T}_h , the vertices and elements of which are denoted by $\mathbf{x}_1, \dots, \mathbf{x}_{N_h}$ and K^1, \dots, K^{E_h} , respectively. The volume of an element K^e is denoted by $|K^e|$. The numbers of vertices belonging to K^e are stored in the integer set \mathcal{N}^e . The numbers of elements containing a vertex \mathbf{x}_i are stored in the set \mathcal{E}_i . The numbers of vertices belonging to these elements are stored in $\mathcal{N}_i = \bigcup_{e \in \mathcal{E}_i} \mathcal{N}^e$. The boundary ∂K^e of each element is defined as a union of S^e components ∂K_s^e on which the outward unit normal \mathbf{n}^e is constant. The integer set containing the numbers of elements sharing such a boundary component (an edge in 2D or face in 3D) with K^e is referred to as \mathcal{E}^e . The notation $\partial K_-^e = \{\mathbf{x} \in \partial K^e : \mathbf{v}(\mathbf{x}) \cdot \mathbf{n}^e(\mathbf{x}) < 0\}$ and $\partial K_+^e = \partial K^e \setminus \partial K_-^e$ is used in surface integrals over the boundary $\partial K^e = \bigcup_{s=1}^{S^e} \partial K_s^e$.

The finite element space of the locally conservative enriched Galerkin method [\[7, 30, 37\]](#) is given by $V_h^c \oplus V_h^d$, where V_h^c is the CG- \mathbb{P}_1 or CG- \mathbb{Q}_1 space spanned by the Lagrange basis functions $\varphi_1, \dots, \varphi_{N_h}$ and V_h^d is the DG- \mathbb{P}_0 space spanned by the characteristic functions $\chi_1, \dots, \chi_{E_h}$ of mesh elements. That is, any $u_h \in V_h$ admits a decomposition into a CG component

$u_h^c = \sum_{j=1}^{N_h} u_j^c \varphi_j \in V_h^c$ and a DG component $u_h^d = \sum_{e=1}^{E_h} u_e^d \chi_e \in V_h^d$. Integration by parts in the weak form of (1a) yields the semi-discrete problem

$$\begin{aligned} \sum_{e=1}^{E_h} \int_{K^e} w_h \frac{\partial u_h}{\partial t} d\mathbf{x} &= \sum_{e=1}^{E_h} \int_{K^e} \nabla w_h \cdot (\mathbf{v} u_h) d\mathbf{x} \\ &\quad - \sum_{e=1}^{E_h} \int_{\partial K^e} w_h \hat{u}_h \mathbf{v} \cdot \mathbf{n}^e ds \quad \forall w_h \in V_h, \end{aligned} \quad (2)$$

where \hat{u}_h is the upwind-sided limit of $u_h = u_h^c + u_h^d$, i.e.,

$$\hat{u}_h(\mathbf{x}) = \begin{cases} u_h^c(\mathbf{x}) + u_e^d & \text{if } \mathbf{x} \in \partial K_+^e, \\ u_h^c(\mathbf{x}) + u_{e'}^d & \text{if } \mathbf{x} \in \partial K_-^e \cap \partial K^{e'}, \\ u_{\text{in}} & \text{if } \mathbf{x} \in \partial K_-^e \cap \Gamma. \end{cases} \quad (3)$$

Using the DG- \mathbb{P}_0 test functions $w_h^d \in \{\chi_1, \dots, \chi_{E_h}\}$, we find that the cell averages $\bar{u}_e = \frac{1}{|K^e|} \int_{K^e} u_h d\mathbf{x}$ of the EG approximation u_h must satisfy

$$|K^e| \frac{d\bar{u}_e}{dt} = \int_{K^e} \frac{\partial u_h}{\partial t} d\mathbf{x} = - \int_{\partial K^e} \hat{u}_h \mathbf{v} \cdot \mathbf{n}^e ds, \quad e = 1, \dots, E_h. \quad (4)$$

The set of equations associated with $w_h^c \in \{\varphi_1, \dots, \varphi_{N_h}\}$ is given by

$$\begin{aligned} \sum_{e \in \mathcal{E}_i} \int_{K^e} \varphi_i \frac{\partial u_h}{\partial t} d\mathbf{x} &= \sum_{e \in \mathcal{E}_i} \int_{K^e} \nabla \varphi_i \cdot (\mathbf{v} u_h) d\mathbf{x} \\ &\quad - \sum_{e \in \mathcal{E}_i} \int_{\partial K^e} \varphi_i \hat{u}_h \mathbf{v} \cdot \mathbf{n}^e ds, \quad i = 1, \dots, N_h. \end{aligned} \quad (5)$$

In view of the fact that $\sum_{i=1}^{N_h} \varphi_i \equiv 1$, the sum of (4) over e coincides with the sum of (5) over i . Indeed, the system of semi-discrete equations (4) and (5) is underdetermined because a globally constant function can be represented exactly in the spaces V_h^c and V_h^d alike. As noticed by Becker et al. [7], this non-uniqueness and the associated singularity of the global mass matrix can be avoided, e.g., by choosing the discontinuous field u_h^d to be *massless*.

To obtain a well-posed semi-discrete problem, we define the degree of freedom u_e^d as the difference between the cell averages of u_h and u_h^c , i.e.,

$$u_e^d = \frac{1}{|K^e|} \int_{K^e} (u_h - u_h^c) d\mathbf{x} = \bar{u}_e - \bar{u}_e^c \quad \forall e = 1, \dots, E_h. \quad (6)$$

The corresponding EG formulation can be interpreted as a variational multiscale (VMS) method [18, 19] in which \bar{u}_e is the coarse-scale component of $u_h|_{K^e}$, while $\bar{u}_e^c = \frac{1}{|K^e|} \int_{K^e} u_h^c d\mathbf{x}$ is the coarse-scale component of \bar{u}_e .

In this work, we evolve the cell averages \bar{u}_e while treating $u_e^d = \bar{u}_e - \bar{u}_e^c$ as derived quantities because the local conservation laws (4) and the discrete maximum principles to be enforced below are formulated in terms of \bar{u}_e . The semi-discrete problem consisting of systems (4) and (5) can be written as

$$\begin{bmatrix} \bar{M} & 0 \\ 0 & M_L \end{bmatrix} \frac{d}{dt} \begin{bmatrix} \bar{u} \\ u^c \end{bmatrix} = \begin{bmatrix} \bar{g}^H(\bar{u}, u^c) \\ g^H(\bar{u}, u^c) + \dot{g}^H\left(\frac{d\bar{u}}{dt}, \frac{du^c}{dt}\right) \end{bmatrix}, \quad (7)$$

where $\bar{u} = (\bar{u}_e)_{e=1}^{E_h}$ is the vector of cell averages, $u^c = (u_j^c)_{j=1}^{N_h}$ is the vector of CG degrees of freedom, $\bar{M} = \text{diag}(|K^e|)_{e=1}^{E_h}$ is the diagonal mass matrix of the space V_h^d , and $M_L = \text{diag}(m_i)_{i=1}^{N_h}$ is the lumped mass matrix of the space V_h^c . Let m_{ij}^e denote the contribution of element K^e to the entry m_{ij} of the consistent mass matrix $M_C = (m_{ij})_{i,j=1}^{N_h}$. The entries of M_C and M_L are assembled from such element contributions as follows:

$$m_{ij} = \sum_{e \in \mathcal{E}_i \cap \mathcal{E}_j} m_{ij}^e, \quad m_{ij}^e = \int_{K^e} \varphi_i \varphi_j d\mathbf{x}, \quad (8)$$

$$m_i = \sum_{e \in \mathcal{E}_i} m_i^e, \quad m_i^e = \sum_{j \in \mathcal{N}^e} m_{ij}^e = \int_{K^e} \varphi_i d\mathbf{x}. \quad (9)$$

If K^e is a simplex, parallelogram or parallelepiped, then $m_i^e = \frac{|K^e|}{|\mathcal{N}^e|}$, where $|\mathcal{N}^e|$ denotes the cardinality of \mathcal{N}^e , i.e., the number of vertices of K^e .

The use of mass lumping on the left-hand side of (7) does not introduce any error if it is compensated by the term $\dot{g}^H\left(\frac{d\bar{u}}{dt}, \frac{du^c}{dt}\right)$ on the right-hand side. Since the matrices \bar{M} and M_L are diagonal, the lagged treatment of \dot{g}^H is particularly advantageous in the context of explicit schemes for numerical solution of (7). Invoking (4) and (5), we find that right-hand side vectors $\bar{g}^H \in \mathbb{R}^{E_h}$, $g^H \in \mathbb{R}^{N_h}$, and $\dot{g}^H \in \mathbb{R}^{N_h}$ of system (7) are defined by

$$\bar{g}_e^H = - \int_{\partial K^e} \hat{u}_h \mathbf{v} \cdot \mathbf{n}^e ds, \quad (10)$$

$$g_i^H = \sum_{e \in \mathcal{E}_i} \int_{K^e} \nabla \varphi_i \cdot (\mathbf{v} u_h) d\mathbf{x} - \sum_{e \in \mathcal{E}_i} \int_{\partial K^e} \varphi_i \hat{u}_h \mathbf{v} \cdot \mathbf{n}^e ds, \quad (11)$$

$$\dot{g}_i^H = \sum_{e \in \mathcal{E}_i} \int_{K^e} \varphi_i \left[\left(\frac{du_i^c}{dt} - \frac{du_h^c}{dt} \right) - \left(\frac{d\bar{u}_e}{dt} - \frac{d\bar{u}_e^c}{dt} \right) \right] d\mathbf{x}. \quad (12)$$

The superscript H refers to a ‘high-order’ target. The superscript L is reserved for the ‘low-order’ approximations to be presented below.

Since $\sum_{e=1}^{E_h} \bar{g}_e^H = \sum_{i=1}^{N_h} g_i^H$ and the components of \dot{g}^H sum to zero, the total masses of u_h and u_h^c change at the same rate. We have

$$\frac{d}{dt} \sum_{e=1}^{E_h} \int_{K^e} u_h d\mathbf{x} = \sum_{e=1}^{E_h} |K^e| \frac{d\bar{u}_e}{dt} = \sum_{i=1}^{N_h} m_i \frac{du_i^c}{dt} = \frac{d}{dt} \sum_{e=1}^{E_h} \int_{K^e} u_h^c d\mathbf{x}. \quad (13)$$

It follows that $\sum_{e=1}^{E_h} \int_{K^e} u_h^d(\mathbf{x}, t) d\mathbf{x} = 0$ for all $t > 0$ if $\sum_{e=1}^{E_h} \int_{K^e} u_h^d(\mathbf{x}, 0) d\mathbf{x} = 0$. That is, an initially massless discontinuous field u_h^d remains massless.

To transform (7) into a system of ordinary differential equations (ODEs), the right-hand side of which does not contain time derivatives, we replace \dot{g}^H with $(M_L - M_C)\dot{u}^H$, where \dot{u}^H is a vector of discrete time derivatives expressed in terms of \bar{u} and \bar{u}^c . This representation makes it possible to avoid the need for inverting consistent mass matrices and decompose

$$\tilde{g}^H(\bar{u}, u^c) = g^H(\bar{u}, u^c) + (M_L - M_C)\dot{u}^H \quad (14)$$

into element contributions suitable for algebraic flux correction (see below). The semi-discrete problem for the continuous component u_h^c becomes

$$M_L \frac{du^c}{dt} = \tilde{g}^H(\bar{u}, u^c). \quad (15)$$

It is easy to verify that (15) is equivalent to (5) if $\dot{u}_h^H \in V_h^c$ satisfies

$$\int_{K^e} \varphi_i (\dot{u}_i^H - \dot{u}_h^H) d\mathbf{x} = \dot{g}_i^H \quad \forall i = 1, \dots, N_h \quad \forall e \in \mathcal{E}_i. \quad (16)$$

The matrix form of this system is given by $(M_L - M_C)\dot{u}^H = \dot{g}_H$, where $M_L - M_C$ is a discrete diffusion (alias *graph Laplacian*) operator. The exact solution \dot{u}^H of this linear system exists and is defined up to a constant for any right-hand side \dot{g}^H satisfying the zero sum condition $\sum_{i=1}^{N_h} \dot{g}_i^H = 0$.

Clearly, it would be highly impractical to solve Poisson-like problems for \dot{u}^H in numerical methods for system (15). The involved computational effort would not pay off because $\dot{u}_h^H \approx \dot{u}_h \approx \dot{u}_h^c$ for the EG method in which \dot{u}_h^d represents a small correction of \dot{u}_h^c . In pseudo-time stepping methods for

steady state problems, we use the lumped-mass approximation $\dot{u}^H \approx 0$. As we show in [Section 7](#), it also produces remarkably accurate results for time-dependent advection problems discretized using the EG method. As another inexpensive approximation, we consider the lumped-mass L^2 projection

$$\dot{u}_i = \frac{1}{m_i} \sum_{e \in \mathcal{E}_i} m_i^e \frac{d\bar{u}_e}{dt} = \frac{1}{m_i} \sum_{e \in \mathcal{E}_i} \frac{m_i^e}{|K^e|} \bar{g}_e(\bar{u}, u^c) \quad (17)$$

of $\frac{d\bar{u}_e}{dt} = \frac{1}{|K^e|} \bar{g}_e(\bar{u}, u^c)$ into the space V_h^c . Taking advantage of the fact that the time derivatives of u_h^d are small compared to those of u_h^c and u_h , we reconstruct $\dot{u}_i \approx \frac{du_i^c}{dt}$ from $\frac{d\bar{u}_e}{dt}$ in this way. Note that $\dot{u}_i = 0$ if $\frac{d\bar{u}_e}{dt} = 0 \forall e \in \mathcal{E}_i$. If the cell average \bar{u}_e is evolved using (4), then $\bar{g}_e = \bar{g}_e^H$, as defined by (10), should be used to calculate the approximate time derivative $\dot{u}_i \approx \dot{u}_i^H$.

3. Algebraic splitting

A numerical scheme for a time-dependent conservation law is called *local extremum diminishing* (LED) [20, 25] if the solution values or cell averages stay bounded by the local maxima and minima at the previous time level or Runge-Kutta stage. The LED property of the space discretization implies positivity preservation and the validity of generalized discrete maximum principles at steady state [32]. For example, the first-order upwind DG- \mathbb{P}_0 discretization of the linear advection equation is provably LED. The \mathbb{P}_0 enrichment of a CG- \mathbb{P}_1 or CG- \mathbb{Q}_1 discretizations has a positive impact on the rates of convergence to smooth solutions but may be insufficient to prevent spurious undershoots and overshoots in the neighborhood of steep gradients. This unsatisfactory behavior of EG solutions can be cured using the *algebraic flux correction* (AFC) methodology [3, 5, 25, 32], a general framework for constraining a high-order discretization to be LED. As a first step toward that end, we need to write our EG scheme in the equivalent form

$$\bar{M} \frac{d\bar{u}}{dt} = \bar{g}^L(\bar{u}) + \bar{f}(u^c), \quad (18)$$

$$M_L \frac{du^c}{dt} = g^L(u^c) + f(\bar{u}, u^c), \quad (19)$$

where $\bar{g}^L(\bar{u})$ and $g^L(u^c)$ correspond to low-order LED approximations. We construct an algebraic splitting of this kind in this section. In [Sections 4](#)

and 5, we present algorithms for constraining $\bar{f}(u^c) = \bar{g}^H(\bar{u}, u^c) - \bar{g}^L(\bar{u})$ and $f(\bar{u}, u^c) = \tilde{g}^H(\bar{u}, u^c) - g^L(u^c)$ to preserve LED property.

A low-order LED counterpart of the integral conservation law (4) can be defined as the DG- \mathbb{P}_0 upwind approximation in which \hat{u}_h is replaced with the upwind-sided cell average. Substituting (3) into (4) and recalling that $u_e^d = \bar{u}_e - \bar{u}_e^c$, we obtain the corresponding algebraic splitting

$$|K^e| \frac{d\bar{u}_e}{dt} = - \sum_{s=1}^{S^e} \int_{\partial K_s^e} \hat{u}_e \mathbf{v} \cdot \mathbf{n}^e ds + \sum_{s=1}^{S^e} \int_{\partial K_s^e} (\hat{u}_e^c - \hat{u}_h^c) \mathbf{v} \cdot \mathbf{n}^e ds, \quad (20)$$

where

$$\hat{u}_e = \begin{cases} \bar{u}_e & \text{on } \partial K_+^e, \\ \bar{u}_{e'} & \text{on } \partial K_-^e \cap \partial K^{e'}, \\ u_{\text{in}} & \text{on } \partial K^e \cap \Gamma_-, \end{cases} \quad \hat{u}_e^c = \begin{cases} \bar{u}_e^c & \text{on } \partial K_+^e, \\ \bar{u}_{e'}^c & \text{on } \partial K_-^e \cap \partial K^{e'}, \\ u_{\text{in}} & \text{on } \partial K^e \cap \Gamma_-, \end{cases}$$

$$\hat{u}_h^c = \begin{cases} u_h^c & \text{on } \partial K^e \setminus \Gamma_-, \\ u_{\text{in}} & \text{on } \partial K^e \cap \Gamma_-. \end{cases}$$

The right-hand side of the e -th equation in system (18) is defined by

$$\bar{g}_e^L = \sum_{s=1}^{S^e} \bar{g}_e^s, \quad \bar{g}_e^s = - \int_{\partial K_s^e} \hat{u}_e \mathbf{v} \cdot \mathbf{n}^e ds, \quad (21)$$

$$\bar{f}_e = \sum_{s=1}^{S^e} \bar{f}_e^s, \quad \bar{f}_e^s = \int_{\partial K_s^e} (\hat{u}_e^c - \hat{u}_h^c) \mathbf{v} \cdot \mathbf{n}^e ds. \quad (22)$$

In this splitting, the flux \bar{g}_e^s is a first-order upwind approximation to the rate of inflow across the boundary component ∂K_s^e , while \bar{f}_e^s is an antidiffusive flux depending on the upwind-sided fluctuation $\hat{u}_e^c - \hat{u}_h^c$ of u_h^c .

If \bar{u}_e is advanced in time using an explicit *strong stability preserving* (SSP) Runge-Kutta method [11], then each stage has the structure of a forward Euler step. The LED property of the fully discrete first-order upwind scheme

$$\bar{u}_e^L = \bar{u}_e + \frac{\Delta t}{|K^e|} \bar{g}_e^L(\bar{u}), \quad e \in \{1, \dots, E_h\}, \quad (23)$$

where Δt is a positive time step, is guaranteed by the following Theorem.

Theorem 1 (LED property of the DG- \mathbb{P}_0 upwind scheme). *If the time step Δt satisfies the CFL-like condition*

$$\Delta t \int_{\partial K^e} |\mathbf{v} \cdot \mathbf{n}^e| ds \leq |K^e|, \quad (24)$$

then the local discrete maximum principle

$$\min_{\partial K^e} \hat{u}_e =: \bar{u}_e^{\min} \leq \bar{u}_e^L \leq \bar{u}_e^{\max} := \max_{\partial K^e} \hat{u}_e \quad (25)$$

holds for the first-order upwind approximation \bar{u}_e^L defined by (23).

PROOF. Following Guermond and Popov [15], we show that \bar{u}_e^L is a convex combination of \bar{u}_e and some intermediate states $\bar{u}_e^s \in [\bar{u}_e^{\min}, \bar{u}_e^{\max}]$. Let us first suppose that $\partial K^e \cap \Gamma = \emptyset$. Define the coefficients

$$k_e^s = -\frac{1}{2} \int_{\partial K_s^e} \mathbf{v} \cdot \mathbf{n}^e ds, \quad d_e^s = \frac{1}{2} \int_{\partial K_s^e} |\mathbf{v} \cdot \mathbf{n}^e| ds. \quad (26)$$

Since $\nabla \cdot \mathbf{v} = 0$, we have $\bar{u}_e \int_{\partial K^e} \mathbf{v} \cdot \mathbf{n}^e ds = 0$ and, therefore,

$$\begin{aligned} \bar{g}_e^L(\bar{u}) &= - \int_{\partial K^e} \hat{u}_e \mathbf{v} \cdot \mathbf{n}^e ds = \int_{\partial K_-^e} (\hat{u}_e - \bar{u}_e) |\mathbf{v} \cdot \mathbf{n}^e| ds \\ &= \sum_{s=1}^{S^e} (k_e^s + d_e^s) (\bar{u}_{e'} - \bar{u}_e) = \sum_{s=1}^{S^e} 2d_e^s (\bar{u}_e^s - \bar{u}_e), \end{aligned} \quad (27)$$

where $e' \in \mathcal{E}^e$ is the number of $K^{e'}$ such that $\partial K_s^e = \partial K^e \cap \partial K^{e'}$ and

$$\bar{u}_e^s = \begin{cases} \frac{\bar{u}_{e'} + \bar{u}_e}{2} + \frac{k_e^s}{d_e^s} \frac{\bar{u}_{e'} - \bar{u}_e}{2} & \text{if } d_e^s > 0, \\ \bar{u}_e & \text{if } d_e^s = 0. \end{cases} \quad (28)$$

Note that we have $\bar{u}_e^s \in [\bar{u}_e^{\min}, \bar{u}_e^{\max}]$ by definition of k_e^s and d_e^s . In the case $\partial K^e \cap \Gamma \neq \emptyset$, we represent the boundary integral

$$\begin{aligned} \bar{g}_e^L(\bar{u}) &= - \int_{\partial K^e} \hat{u}_e \mathbf{v} \cdot \mathbf{n}^e ds = \int_{\partial K^e \cap \Gamma_-} (u_{\text{in}} - \bar{u}_e) |\mathbf{v} \cdot \mathbf{n}^e| ds \\ &\quad + \int_{\partial K_-^e \setminus \Gamma} (\hat{u}_e - \bar{u}_e) |\mathbf{v} \cdot \mathbf{n}^e| ds = \sum_{s=1}^{S^e} 2d_e^s (\bar{u}_e^s - \bar{u}_e) \end{aligned} \quad (29)$$

in terms of the generalized intermediate states

$$\bar{u}_e^s = \begin{cases} \frac{\bar{u}_{e'} + \bar{u}_e}{2} + \frac{k_e^s}{d_e^s} \frac{\bar{u}_{e'} - \bar{u}_e}{2} & \text{if } \partial K_s^e = \partial K^e \cap \partial K^{e'} \wedge d_e^s > 0, \\ \bar{u}_e + \frac{1}{2d_e^s} \int_{\partial K_s^e \cap \Gamma_-} (u_{\text{in}} - \bar{u}_e) |\mathbf{v} \cdot \mathbf{n}^e| ds & \text{if } \partial K_s^e \subset \Gamma \wedge d_e^s > 0, \\ \bar{u}_e & \text{otherwise} \end{cases} \quad (30)$$

which are also in the range $[\bar{u}_e^{\min}, \bar{u}_e^{\max}]$ by definition of d_e^s . Substituting $\bar{g}_e^L(\bar{u}) = \sum_{s=1}^{S^e} 2d_e^s(\bar{u}_e^s - \bar{u}_e)$ into (23), we find that (cf. [12, 14, 15])

$$\begin{aligned} \bar{u}_e^L = \bar{u}_e + \frac{\Delta t}{|K^e|} \sum_{s=1}^{S^e} 2d_e^s(\bar{u}_e^s - \bar{u}_e) &= \left(1 - \frac{\Delta t}{|K^e|} \int_{\partial K^e} |\mathbf{v} \cdot \mathbf{n}^e| ds\right) \bar{u}_e \\ &+ \frac{\Delta t}{|K^e|} \left(\sum_{s=1}^{S^e} \bar{u}_e^s \int_{\partial K_s^e} |\mathbf{v} \cdot \mathbf{n}^e| ds \right). \end{aligned} \quad (31)$$

The assertion of the theorem follows from the fact that the result is a convex combination of \bar{u}_e and \bar{u}_e^s for time steps Δt satisfying (24). \square

Remark 1. If $\partial K_s^e = \partial K^e \cap \partial K^{e'} = \partial K_{s'}^{e'}$ is an internal edge/face, then

$$k_e^s = -k_{e'}^{s'}, \quad d_e^s = d_{e'}^{s'}, \quad \bar{u}_e^s = \bar{u}_{e'}^{s'}.$$

The first-order upwind fluxes $\bar{g}_{ee'} = \bar{g}_e^s$ and $\bar{g}_{e'e} = \bar{g}_{e'}^{s'}$ are given by

$$\bar{g}_{ee'} = (k_e^s + d_e^s)(\bar{u}_{e'} - \bar{u}_e), \quad \bar{g}_{e'e} = -\bar{g}_{ee'}.$$

Remark 2. In the one-dimensional case, the time step restriction imposed by condition (24) underestimates the optimal CFL bound by a factor of 2. However, the simple convexity-based approach to proving the LED property is well suited for the design of high-order extensions (see Section 5).

Let us now represent the CG subproblem in the form (19). Integrating $\int_{K^e} \nabla \varphi_i \cdot (\mathbf{v} u_h^c) d\mathbf{x}$ by parts and exploiting continuity of φ_i , which implies

$$\int_{\partial K^e \cap \partial K^{e'}} \varphi_i \hat{u}_h \mathbf{v} \cdot \mathbf{n}^e ds + \int_{\partial K^e \cap \partial K^{e'}} \varphi_i \hat{u}_h \mathbf{v} \cdot \mathbf{n}^{e'} ds = 0 \quad \forall e' \in \mathcal{E}^e$$

for the uniquely defined upwind limit $\hat{u}_h = u_h^c + \hat{u}_h^d = u_h^c + \hat{u}_e - \hat{u}_e^c$, the i -th component of $\tilde{g}^H(\bar{u}, u^c)$ that we defined in (15) can be written as

$$\begin{aligned} \tilde{g}_i^H(\bar{u}, u^c) &= \sum_{e \in \mathcal{E}_i} \int_{\partial K^e \cap \Gamma_-} \varphi_i(u_{\text{in}} - u_h^c) |\mathbf{v} \cdot \mathbf{n}^e| ds \\ &\quad - \sum_{e \in \mathcal{E}_i} \int_{K^e} \varphi_i \nabla \cdot (\mathbf{v} u_h^c) d\mathbf{x} + \sum_{e \in \mathcal{E}_i} \int_{K^e} \varphi_i (\dot{u}_i^H - \dot{u}_h^H) d\mathbf{x} \\ &\quad + \sum_{e \in \mathcal{E}_i} (\bar{u}_e - \bar{u}_e^c) \left(\int_{K^e} \mathbf{v} \cdot \nabla \varphi_i d\mathbf{x} - \int_{\partial K^e \cap \Gamma} \varphi_i \mathbf{v} \cdot \mathbf{n}^e ds \right), \end{aligned} \quad (32)$$

where $\dot{u}_h^H = \sum_{j=1}^{N_h} \dot{u}_j^H \varphi_j$ stands for the V_h^c interpolant of the nodal time derivatives \dot{u}_i^H reconstructed from $\bar{g}_e^H(\bar{u}, u^c)$, $e \in \mathcal{E}_i$.

The low-order LED approximation $g^L(u^c)$ to $\tilde{g}^H(\bar{u}, u^c)$ can be extracted using the element matrix version [26, 34] of the *discrete upwinding* method [29] which is commonly employed in algebraic flux correction schemes for finite elements [3, 16, 25, 32]. Introducing the coefficients

$$\begin{aligned} k_{ij}^e &= - \int_{K^e} \varphi_i \mathbf{v} \cdot \nabla \varphi_j d\mathbf{x} = \int_{K^e} \nabla \varphi_i \cdot \mathbf{v} \varphi_j d\mathbf{x} - \int_{\partial K^e} \varphi_i \varphi_j \mathbf{v} \cdot \mathbf{n}^e d\mathbf{x} \\ &= -k_{ji}^e - \int_{\partial K^e} \varphi_i \varphi_j \mathbf{v} \cdot \mathbf{n}^e d\mathbf{x} \end{aligned} \quad (33)$$

of the element matrix associated with the discretized advective term

$$\int_{K^e} \varphi_i \nabla \cdot (\mathbf{v} u_h^c) d\mathbf{x} = \int_{K^e} \varphi_i \mathbf{v} \cdot \nabla u_h^c d\mathbf{x} = - \sum_{j \in \mathcal{N}^e} k_{ij}^e u_j^c \quad (34)$$

and using them to calculate the artificial diffusion coefficients (cf. [16, 34])

$$d_{ij}^e = \begin{cases} \max\{-k_{ij}^e, 0, -k_{ji}^e\} & \text{if } j \in \mathcal{N}^e \setminus \{i\}, i \in \mathcal{N}^e, \\ -\sum_{k \in \mathcal{N}^e \setminus \{i\}} d_{ik}^e & \text{if } j = i \in \mathcal{N}^e, \\ 0 & \text{otherwise} \end{cases} \quad (35)$$

such that $k_{ij}^e + d_{ij}^e \geq 0$ for all $i \in \mathcal{N}^e$ and $j \in \mathcal{N}^e \setminus \{i\}$, we define

$$g_i^L(u^c) = \sum_{e \in \mathcal{E}_i} g_i^{e,L}, \quad f_i(\bar{u}, u^c) = \sum_{e \in \mathcal{E}_i} f_i^e \quad (36)$$

in terms of the element contributions

$$g_i^{e,L} = \int_{\partial K^e \cap \Gamma_-} \varphi_i(u_{\text{in}} - u_i^c) |\mathbf{v} \cdot \mathbf{n}^e| ds + \sum_{j \in \mathcal{N}^e \setminus \{i\}} (k_{ij}^e + d_{ij}^e)(u_j^c - u_i^c) \quad (37)$$

$$f_i^e = \int_{\partial K^e \cap \Gamma_-} \varphi_i(u_i^c - u_h^c) |\mathbf{v} \cdot \mathbf{n}^e| ds - \sum_{j \in \mathcal{N}^e \setminus \{i\}} [m_{ij}^e(\dot{u}_j^H - \dot{u}_i^H) + d_{ij}^e(u_j^c - u_i^c)] \\ + (\bar{u}_e - \bar{u}_e^c) \left(\int_{K^e} \mathbf{v} \cdot \nabla \varphi_i d\mathbf{x} - \int_{\partial K_+^e \cap \Gamma} \varphi_i \mathbf{v} \cdot \mathbf{n}^e ds \right). \quad (38)$$

In algebraic flux correction schemes for CG methods [26, 34], the antidiffusive element contributions f_i^e require additional high-order stabilization to avoid ripples within the range of values satisfying discrete maximum principles [26, 27, 34]. The last term on the right-hand side of (38) represents the contribution of the discontinuous EG component $u_e^d = \bar{u}_e - \bar{u}_e^c$. It provides the desired stabilization effect without introducing free parameters, which makes EG a better limiting target than artificially stabilized CG methods.

The forward Euler update corresponding to a single stage of an SSP Runge-Kutta method for the low-order CG subproblem is defined by

$$u_i^{c,L} = u_i^c + \frac{\Delta t}{m_i} g_i^L(u^c), \quad i \in \{1, \dots, N_h\}. \quad (39)$$

The so-defined approximation proves LED w.r.t. the nodal stencil \mathcal{N}_i .

Theorem 2 (LED property of the low-order CG scheme [12, 15]). *If*

$$\Delta t \sum_{e \in \mathcal{E}_i} \left(\int_{\partial K^e \cap \Gamma_-} \varphi_i |\mathbf{v} \cdot \mathbf{n}^e| ds + \sum_{j \in \mathcal{N}^e \setminus \{i\}} 2d_{ij}^e \right) \leq m_i, \quad (40)$$

then $u_i^{c,L}$ defined by (39) satisfies the local discrete maximum principle

$$u_i^{\min} \leq u_i^{c,L} \leq u_i^{\max}, \quad (41)$$

$$u_i^{\max} = \max \left\{ \max_{\partial \Omega_i \cap \Gamma_-} u_{\text{in}}, \max_{j \in \mathcal{N}_i} u_j^c \right\}, \quad (42)$$

$$u_i^{\min} = \min \left\{ \min_{\partial \Omega_i \cap \Gamma_-} u_{\text{in}}, \min_{j \in \mathcal{N}_i} u_j^c \right\}, \quad (43)$$

where $\partial \Omega_i$ is the boundary of the element patch $\Omega_i = \bigcup_{e \in \mathcal{E}_i} K^e$.

PROOF. The proof is similar to that of Theorem 1. We have

$$\begin{aligned} g_i^L(u^c) &= \sum_{e \in \mathcal{E}_i} \left[\int_{\partial K^e \cap \Gamma_-} \varphi_i(u_{\text{in}} - u_i^c) |\mathbf{v} \cdot \mathbf{n}^e| ds + \sum_{j \in \mathcal{N}^e \setminus \{i\}} 2d_{ij}^e (\bar{u}_{ij}^e - u_i^c) \right] \\ &= \sum_{e \in \mathcal{E}_i} \sum_{j \in \mathcal{N}^e} 2\bar{d}_{ij}^e (\bar{u}_{ij}^e - u_i^c), \end{aligned} \quad (44)$$

where $\bar{d}_{ij}^e = d_{ij}^e$ for $j \neq i$ and $\bar{d}_{ii}^e = \frac{1}{2} \int_{\partial K^e \cap \Gamma_-} \varphi_i |\mathbf{v} \cdot \mathbf{n}^e| ds$. The intermediate states \bar{u}_{ij}^e associated with these coefficients are defined by

$$\bar{u}_{ij}^e = \begin{cases} \frac{u_j^c + u_i^c}{2} + \frac{k_{ij}^e}{d_{ij}^e} \frac{u_j^c - u_i^c}{2} & \text{if } j \neq i \wedge d_{ij}^e > 0, \\ u_i^c + \frac{1}{2\bar{d}_{ii}^e} \int_{\partial K^e \cap \Gamma_-} \varphi_i (u_{\text{in}} - u_i^c) |\mathbf{v} \cdot \mathbf{n}^e| ds & \text{if } j = i \wedge \bar{d}_{ii}^e > 0, \\ u_i^c & \text{otherwise.} \end{cases} \quad (45)$$

The assertion of the theorem follows from the fact that $u_i^{c,L}$ is a convex combination of states belonging to the interval $[u_i^{\min}, u_i^{\max}]$. \square

Remark 3. In view of (33), the ‘edge bar states’ \bar{u}_{ij}^e and \bar{u}_{ji}^e corresponding to a pair of nodes $i \in \mathcal{N}^e$ and $j \in \mathcal{N}^e \setminus \{i\}$ with $d_{ij}^e > 0$ are related by

$$\begin{aligned} \bar{u}_{ij}^e &= \frac{u_j^c + u_i^c}{2} + \frac{k_{ji}^e}{d_{ij}^e} \frac{u_i^c - u_j^c}{2} + \frac{u_i^c - u_j^c}{2d_{ij}^e} \int_{\partial K^e} \varphi_i \varphi_j \mathbf{v} \cdot \mathbf{n}^e dx \\ &= \bar{u}_{ji}^e + \frac{u_i^c - u_j^c}{2d_{ij}^e} \int_{\partial K^e} \varphi_i \varphi_j \mathbf{v} \cdot \mathbf{n}^e dx \\ &= \bar{u}_{ji}^e - \frac{k_{ij}^e + k_{ji}^e}{2d_{ij}^e} (u_i^c - u_j^c). \end{aligned}$$

Remark 4. Preservation of local bounds for implicit time discretizations of semi-discrete LED schemes and the validity of generalized discrete maximum principles at steady state can be shown using more sophisticated proof techniques. For details, we refer the interested reader to [5, 32].

Remark 5. The provable order of accuracy of the DG- \mathbb{P}_0 upwind scheme and of its low-order CG counterpart is $\frac{1}{2}$ under suitable smoothness assumptions [17, 32]. For linear advection with constant velocity on a uniform

1D mesh, both schemes produce the first-order upwind finite difference approximation. That is why they are frequently called ‘first-order’ schemes in the literature. Stable ‘second-order’ \mathbb{P}_1 finite element discretizations of the pure advection equation converge to smooth exact solutions at the rate $\frac{3}{2}$ on general meshes [17, 39]. The EG- \mathbb{P}_1 method delivers this optimal order of accuracy [7] without the need for parameter-dependent stabilization.

Decompositions (22) and (36) make it possible to blend ‘first-order’ and ‘second-order’ approximations in a conservative manner. Importantly, the zero sum condition $f_{ee'} + f_{e'e} = 0$ holds for each pair of fluxes $f_{ee'} = \bar{f}_e^s$ and $f_{e'e} = f_{e'}^{s'}$, where $e' \in \mathcal{E}^e$ is the number of an element $K^{e'}$ such that $\partial K_s^e = \partial K^e \cap \partial K^{e'} = \partial K_{s'}^{e'}$. Similarly, our definition (38) of the element contributions f_i^e has the property that $\sum_{i \in \mathcal{N}^e} f_i^e = 0$ if $\partial K^e \cap \Gamma = \emptyset$. Multiplying flux pairs $\{f_e^s, f_{e'}^{s'}\}$ and components of $f^e = (f_i^e)_{i \in \mathcal{N}^e}$ by adaptively chosen correction factors $\bar{\alpha}^s = \bar{\alpha}^{s'}$ and α_i^e , respectively, their contribution to the residual of the semi-discrete scheme can be constrained to be LED. In the next two sections, we define $\bar{\alpha}^s$ and α_i^e for two kinds of bound-preserving EG schemes.

4. Flux-corrected transport

The presence of the additional terms $\bar{f}(u^c)$ and $f(\bar{u}, u^c)$ on the right-hand sides of the EG subproblems (18) and (19) may lead to a loss of the LED property that we proved in Theorems 1 and 2. To prevent violations of local discrete maximum principles in our EG schemes, we will adjust the magnitudes of the raw antidiffusive fluxes \bar{f}_e^s and element contributions f_i^e . An algorithm for calculating the corresponding correction factors $\bar{\alpha}^s \in [0, 1]$ and $\alpha_i^e \in [0, 1]$ is called *limiter*. It should preserve the discrete conservation properties, provably enforce preservation of local bounds, and leave the ‘second-order’ target discretization unchanged in smooth regions. The algebraic flux correction (AFC) methodology [3, 25, 32] provides a general framework for the design and analysis of such limiters.

The oldest representative of AFC schemes for finite elements is the flux-corrected transport (FCT) algorithm [29, 36, 38]. FCT adds limited antidiffusive corrections to the degrees of freedom calculated using a bound-preserving low-order scheme. Adopting this predictor-corrector strategy, the low-order

approximations (23) and (39) can be improved as follows:

$$\bar{u}_e^{\text{FCT}} = \bar{u}_e^L + \frac{\Delta t}{|K^e|} \sum_{s=1}^{S^e} \bar{\alpha}^s \bar{f}_e^s, \quad e \in \{1, \dots, E_h\}, \quad (46)$$

$$u_i^{c,\text{FCT}} = u_i^{c,L} + \frac{\Delta t}{m_i} \sum_{e \in \mathcal{E}_i} \alpha_i^e f_i^e, \quad i \in \{1, \dots, N_h\}. \quad (47)$$

Of course, the accuracy of the so-defined EG-FCT scheme depends on the choice of the limiting parameters $\bar{\alpha}^s$ and α_i^e . Varying them between 0 and 1, the antidiffusive correction terms can be adjusted to keep the values of \bar{u}_e^{FCT} and $u_i^{c,\text{FCT}}$ in the range determined by the extended local bounds

$$\bar{u}_e^{\text{max,FCT}} = \max_{i \in \mathcal{N}^e} \max \left\{ \max_{e' \in \mathcal{E}_i} \bar{u}_{e'}, u_i^c, \max_{e' \in \mathcal{E}_i} \bar{u}_{e'}^L \right\} \geq \bar{u}_e^L, \quad (48)$$

$$\bar{u}_e^{\text{min,FCT}} = \min_{i \in \mathcal{N}^e} \min \left\{ \min_{e' \in \mathcal{E}_i} \bar{u}_{e'}, u_i^c, \min_{e' \in \mathcal{E}_i} \bar{u}_{e'}^L \right\} \leq \bar{u}_e^L, \quad (49)$$

$$u_i^{\text{max,FCT}} = \max \left\{ \max_{j \in \mathcal{N}_i} u_j^c, \max_{e \in \mathcal{E}_i} \bar{u}_e, \max_{j \in \mathcal{N}_i} u_j^{c,L} \right\} \geq u_i^{c,L}, \quad (50)$$

$$u_i^{\text{min,FCT}} = \min \left\{ \min_{j \in \mathcal{N}_i} u_j^c, \min_{e \in \mathcal{E}_i} \bar{u}_e, \min_{j \in \mathcal{N}_i} u_j^{c,L} \right\} \leq u_i^{c,L}. \quad (51)$$

Since the values of \bar{u}_e^L and $u_i^{c,L}$ are in this range, the FCT constraints

$$\bar{u}_e^{\text{min,FCT}} \leq \bar{u}_e^{\text{FCT}} \leq \bar{u}_e^{\text{max,FCT}}, \quad (52)$$

$$u_i^{\text{min,FCT}} \leq u_i^{c,\text{FCT}} \leq u_i^{\text{max,FCT}} \quad (53)$$

can always be enforced by tuning $\bar{\alpha}^s$ and α_i^e in an appropriate manner.

Zalesak's multidimensional FCT algorithm [38] yields the following definition of the correction factors $\bar{\alpha}^s$ for step (46) of our AFC scheme:

$$\bar{\alpha}^s = \min_{1 \leq s \leq S^e} \begin{cases} \min\{R_e^+, R_{e'}^-\} & \text{if } \partial K_s^e = \partial K^e \cap \partial K^{e'} \wedge \bar{f}_e^s > 0, \\ \min\{R_e^-, R_{e'}^+\} & \text{if } \partial K_s^e = \partial K^e \cap \partial K^{e'} \wedge \bar{f}_e^s < 0, \\ R_e^+ & \text{if } \partial K_s^e \subset \Gamma \wedge \bar{f}_e^s > 0, \\ R_e^- & \text{if } \partial K_s^e \subset \Gamma \wedge \bar{f}_e^s < 0, \\ 1 & \text{otherwise,} \end{cases} \quad (54)$$

where the element-based correction factors R_e^\pm are defined by

$$R_e^+ = \min \left\{ 1, \frac{\frac{|K^e|}{\Delta t} (\bar{u}_e^{\max, \text{FCT}} - \bar{u}_e^L)}{\sum_{s=1}^{S^e} \max\{0, \bar{f}_e^s\}} \right\}, \quad (55)$$

$$R_e^- = \min \left\{ 1, \frac{\frac{|K^e|}{\Delta t} (\bar{u}_e^{\min, \text{FCT}} - \bar{u}_e^L)}{\sum_{s=1}^{S^e} \min\{0, \bar{f}_e^s\}} \right\}. \quad (56)$$

Importantly, the same correction factor $\bar{\alpha}^s = \bar{\alpha}^{s'}$ is used for fluxes $\bar{f}_{ee'} = \bar{f}_e^s$ and $\bar{f}_{e'e} = \bar{f}_{e'}^{s'} = -\bar{f}_{ee'}$ associated with $\partial K_s^e = \partial K^e \cap \partial K^{e'} = \partial K_{s'}^{e'}$. Hence, the FCT correction step (46) preserves the local conservation property.

In contrast to edge-based approaches (cf. [12, 25, 27, 32]), element-based FCT limiters [1, 9, 29, 34, 36] are readily applicable to the antidiffusive element contributions defined by (38). Following [1, 34], we will choose the correction factors α_i^e so as to satisfy the localized inequality constraints

$$f_i^{e, \min} \leq \alpha_i^e f_i^e \leq f_i^{e, \max} \quad \forall e \in \mathcal{E}_i, \quad (57)$$

where

$$f_i^{e, \max} = \frac{m_i^e}{\Delta t} (u_i^{\max, \text{FCT}} - u_i^{c, L}), \quad f_i^{e, \min} = \frac{m_i^e}{\Delta t} (u_i^{\min, \text{FCT}} - u_i^{c, L}). \quad (58)$$

As we show below, conditions (57) guarantee the validity of the FCT constraints (53) for the nodal value $u_i^{c, \text{FCT}}$ of the CG component.

The corresponding range of admissible correction factors α_i^e is given by

$$0 \leq \alpha_i^e \leq R_i^e = \begin{cases} \min \left\{ 1, \frac{f_i^{e, \max}}{f_i^e} \right\} & \text{if } f_i^e > 0, \\ \min \left\{ 1, \frac{f_i^{e, \min}}{f_i^e} \right\} & \text{if } f_i^e < 0, \\ 1 & \text{otherwise.} \end{cases} \quad (59)$$

Recall that the raw antidiffusive element contributions f_i^e satisfy the zero sum condition $\sum_{i \in \mathcal{N}^e} f_i^e = 0$ for $\partial K^e \cap \Gamma = \emptyset$. For the CG component to remain globally conservative, the choice of the nodal correction factors α_i^e must guarantee that $\sum_{i \in \mathcal{N}^e} \alpha_i^e f_i^e = 0$ for $\partial K^e \cap \Gamma = \emptyset$. This requirement is clearly satisfied if the same element-based correction factor $\alpha^e = \min_{j \in \mathcal{N}^e} R_j^e$ is used for all $i \in \mathcal{N}^e$. However, this straightforward definition does not guarantee continuous dependence of $\alpha^e f_i^e$ on the data because small variations of one component may cause large changes of other components.

The CG-FCT schemes proposed in [1, 34] adjust the values of $\tilde{f}_j^e = R_j^e f_j^e$ in a manner which guarantees the validity of the zero sum condition for $f_j^{e,*} = \alpha_j^e f_j^e$. Adopting this limiting strategy, we calculate the sums

$$\tilde{f}^{e,+} = \sum_{j \in \mathcal{N}^e} \max\{0, \tilde{f}_j^e\}, \quad \tilde{f}^{e,-} = \sum_{j \in \mathcal{N}^e} \min\{0, \tilde{f}_j^e\} \quad (60)$$

and balance them using the nodal correction factors defined by [1, 16, 34]

$$\alpha_i^e = \begin{cases} \min \left\{ 1, -\frac{\tilde{f}^{e,-}}{\tilde{f}^{e,+}} \right\} R_i^e & \text{if } \tilde{f}_i^e > 0 \wedge \partial K^e \cap \Gamma = \emptyset, \\ \min \left\{ 1, -\frac{\tilde{f}^{e,+}}{\tilde{f}^{e,-}} \right\} R_i^e & \text{if } \tilde{f}_i^e < 0 \wedge \partial K^e \cap \Gamma = \emptyset, \\ R_i^e & \text{otherwise.} \end{cases} \quad (61)$$

In practice, we bypass the computation of R_i^e . Given $f_i^{e,\max}$ and $f_i^{e,\min}$, we calculate the tentative bound-preserving element contributions

$$\tilde{f}_i^e = R_i^e f_i^e = \min\{f_i^{e,\max}, \max\{f_i^e, f_i^{e,\min}\}\} \quad (62)$$

and the sums $\tilde{f}^{e,\pm}$ defined by (60). The mass correction procedure corresponding to definition (61) of the nodal correction factors α_i^e yields

$$f_i^{e,*} = \alpha_i^e f_i^e = \begin{cases} \min \left\{ 1, -\frac{\tilde{f}^{e,-}}{\tilde{f}^{e,+}} \right\} \tilde{f}_i^e & \text{if } \tilde{f}_i^e > 0 \wedge \partial K^e \cap \Gamma = \emptyset, \\ \min \left\{ 1, -\frac{\tilde{f}^{e,+}}{\tilde{f}^{e,-}} \right\} \tilde{f}_i^e & \text{if } \tilde{f}_i^e < 0 \wedge \partial K^e \cap \Gamma = \emptyset, \\ \tilde{f}_i^e & \text{otherwise.} \end{cases} \quad (63)$$

The design philosophy behind this element-based limiting strategy can be summarized as ‘pointwise clipping + mass correction’. As we show in the next section, it is well suited for monolithic AFC schemes as well.

Theorem 3 (LED property of the FCT scheme). *Let the correction factors $\bar{\alpha}^s$ and α_i^e be defined by (54) and (61), respectively. Then approximations (46) and (47) satisfy the local maximum principles (52) and (53).*

PROOF. The FCT correction (46) of the low-order value \bar{u}_e^L is LED because

$$\begin{aligned} \bar{u}_e^{\min, \text{FCT}} - \bar{u}_e^L &\leq \frac{\Delta t}{|K^e|} \sum_{s=1}^{S^e} R_e^- \min\{0, \bar{f}_e^s\} \leq \frac{\Delta t}{|K^e|} \sum_{s=1}^{S^e} \bar{\alpha}^s \bar{f}_e^s \\ &\leq \frac{\Delta t}{|K^e|} \sum_{s=1}^{S^e} R_e^+ \max\{0, \bar{f}_e^s\} \leq \bar{u}_e^{\max, \text{FCT}} - \bar{u}_e^L. \end{aligned}$$

Following [9, 34], we prove that (47) is LED as well by noticing that

$$u_i^{\min, \text{FCT}} \leq \bar{u}_i^e := u_i^{c, L} + \frac{\Delta t}{m_i^e} f_i^{e, *}, \leq u_i^{\max, \text{FCT}}. \quad (64)$$

The LED property of (47) follows from the fact that $u_i^{c, \text{FCT}} = \frac{1}{m_i} \sum_{e \in \mathcal{E}_i} m_i^e \bar{u}_i^e$ is a convex combination of the bound-preserving states \bar{u}_i^e . \square

5. Monolithic convex limiting

Algebraic flux correction of FCT type is performed after the discretization in time and introduces a splitting error which is proportional to the (pseudo-) time step size. For that reason, FCT-like predictor-corrector approaches are not to be recommended for applications that require the use of large time steps or computation of steady state solutions. As an alternative to FCT, we propose a *monolithic convex limiting* (MCL) strategy based on the one developed in [27]. Instead of correcting low-order LED approximations to the EG degrees of freedom, let us add the sums of limited antidiffusive fluxes $\bar{f}_e^{s, *} = \bar{\alpha}^s \bar{f}_e^s$ and element contributions $f_i^{e, *} = \alpha_i^e f_i^e$ to the right-hand sides $\bar{g}_e^L(\bar{u})$ and $g_i^L(u^c)$ of the semi-discrete low-order schemes. This limiting approach approximates (18) and (19) by the nonlinear system

$$|K^e| \frac{d\bar{u}_e}{dt} = \bar{g}_e^L(\bar{u}) + \sum_{s=1}^{S^e} \bar{f}_e^{s, *}(u^c), \quad e = 1, \dots, E_h, \quad (65)$$

$$m_i \frac{du_i^c}{dt} = g_i^L(u^c) + \sum_{e \in \mathcal{E}_i} f_i^{e, *}(\bar{u}, u^c), \quad i = 1, \dots, N_h. \quad (66)$$

For the steady state of this problem to be well-defined, the nonlinear correction terms $\bar{f}_e^{s, *}(u^c)$ and $f_i^{e, *}(\bar{u}, u^c)$ should be Lipschitz continuous [3, 32]. Additionally, the element contributions f_i^e should be calculated using $\bar{g}_e = \bar{g}_e^L$ in formula (17) for \dot{u}_i . This reconstruction of the nodal time derivatives ensures that their contribution to f_i^e vanishes and full mass lumping is performed in (66) if the solution to (65) reaches a steady state.

As before, we restrict explanations of the proposed methodology and its theoretical analysis to explicit SSP Runge-Kutta (pseudo-)time stepping. However, the limiters to be presented in this section are independent of the time discretization and directly applicable to stationary problems.

In the proofs of Theorems 1 and 2 we exploited the fact that

$$\bar{g}_e^L(\bar{u}) = \sum_{s=1}^{S^e} 2d_e^s(\bar{u}_e^s - \bar{u}_e), \quad (67)$$

$$g_i^L(u^c) = \sum_{e \in \mathcal{E}_i} \sum_{j \in \mathcal{N}_i} 2\bar{d}_{ij}^e(\bar{u}_{ij}^e - u_i^c), \quad (68)$$

where \bar{u}_e^s and \bar{u}_{ij}^e are bound-preserving intermediate states. Introducing

$$\bar{u}_i^e = \frac{1}{\bar{d}_i^e} \sum_{j \in \mathcal{N}^e} \bar{d}_{ij}^e \bar{u}_{ij}^e, \quad \bar{d}_i^e = \sum_{j \in \mathcal{N}^e} \bar{d}_{ij}^e, \quad (69)$$

we require that $\bar{f}_e^{s,*}$ and $f_i^{e,*}$ satisfy the FCT-like inequality constraints

$$\bar{u}_e^{\min, \text{MCL}} \leq \bar{u}_e^{s,*} := \bar{u}_e^s + \frac{\bar{f}_e^{s,*}}{2d_e^s} \leq \bar{u}_e^{\max, \text{MCL}}, \quad (70)$$

$$u_i^{\min, \text{MCL}} \leq u_i^{e,*} := \bar{u}_i^e + \frac{f_i^{e,*}}{2\bar{d}_i^e} \leq u_i^{\max, \text{MCL}}. \quad (71)$$

The basic local bounds for the monolithic version are defined by

$$\bar{u}_e^{\max, \text{MCL}} = \max \left\{ \max_{\partial K^e \cap \Gamma_-} u_{\text{in}}, \max_{i \in \mathcal{N}^e} \max \left\{ \max_{e' \in \mathcal{E}_i} \bar{u}_{e'}, u_i^c \right\} \right\}, \quad (72)$$

$$\bar{u}_e^{\min, \text{MCL}} = \min \left\{ \min_{\partial K^e \cap \Gamma_-} u_{\text{in}}, \min_{i \in \mathcal{N}^e} \min \left\{ \min_{e' \in \mathcal{E}_i} \bar{u}_{e'}, u_i^c \right\} \right\}, \quad (73)$$

$$u_i^{\max, \text{MCL}} = \max \left\{ \max_{\partial \Omega_i \cap \Gamma_-} u_{\text{in}}, \max_{j \in \mathcal{N}_i} u_j^c, \max_{e \in \mathcal{E}_i} \bar{u}_e \right\}, \quad (74)$$

$$u_i^{\min, \text{MCL}} = \min \left\{ \min_{\partial \Omega_i \cap \Gamma_-} u_{\text{in}}, \min_{j \in \mathcal{N}_i} u_j^c, \min_{e \in \mathcal{E}_i} \bar{u}_e \right\}. \quad (75)$$

A linearity-preserving (LP) version of these bounds can be constructed as proposed in Section 6.1 of [27]. The use of limiters that guarantee linearity preservation (i.e., produce $\bar{f}_e^{s,*} = \bar{f}_e^s$ and $f_i^{e,*} = f_i^e$ for locally linear input functions u_h) is essential for achieving optimal convergence to smooth solutions [4]. The MCL bounds (72)–(75) are LP on uniform meshes.

As we show below, an explicit SSP Runge-Kutta time discretization of (65) produces a convex combination of \bar{u}_e and $\bar{u}_e^{s,*}$, $s = 1, \dots, S^e$ for time steps satisfying (24), while the discrete counterpart of (66) produces a convex

combination of $u_i^{c,L}$ and $\bar{u}_i^{e,*}$, $e \in \mathcal{E}_i$ for time steps satisfying (40). Hence, the conditional LED property can be shown as in Theorems 1 and 2.

In contrast to FCT, the MCL constraints for $\bar{u}_e^{s,*}$ and $u_i^{e,*}$ are independent of the time step and no sums appear in the definition of $\bar{u}_e^{s,*}$. However, conditions (70) and (71) exhibit the same structure as (64). Therefore, they can be enforced in the much same way as in localized FCT algorithms.

Rearranging (70) and (71), we find that $\bar{f}_e^{s,*}$ and $f_i^{e,*}$ should satisfy

$$2d_e^s(\bar{u}_e^{\min,\text{MCL}} - \bar{u}_e^s) =: \bar{f}_e^{s,\min} \leq \bar{f}_e^{s,*} \leq \bar{f}_e^{s,\max} := 2d_e^s(\bar{u}_e^{\max,\text{MCL}} - \bar{u}_e^s), \quad (76)$$

$$2\bar{d}_i^e(u_i^{\min,\text{MCL}} - \bar{u}_i^e) := f_i^{e,\min} \leq f_i^{e,*} \leq f_i^{e,\max} := 2\bar{d}_i^e(u_i^{\max,\text{MCL}} - \bar{u}_i^e), \quad (77)$$

where the upper bounds are nonnegative and the lower bounds are nonpositive by definition of the low-order intermediate states (30) and (45).

Remark 6. In practice, we use the equivalent representations (cf. [27])

$$\bar{f}_e^{s,\max} = 2d_e^s \bar{u}_e^{\max,\text{MCL}} - \bar{w}_e^s, \quad f_i^{e,\max} = \sum_{j \in \mathcal{N}^e} (2\bar{d}_{ij}^e u_i^{\max,\text{MCL}} - \bar{w}_{ij}^e), \quad (78)$$

$$\bar{f}_e^{s,\min} = 2\bar{d}_e^s \bar{u}_e^{\min,\text{MCL}} - \bar{w}_e^s, \quad f_i^{e,\min} = \sum_{j \in \mathcal{N}^e} (2\bar{d}_{ij}^e u_i^{\min,\text{MCL}} - \bar{w}_{ij}^e) \quad (79)$$

in terms of the products $\bar{w}_e^s = 2d_e^s \bar{u}_e^s$ and $\bar{w}_{ij}^e = 2\bar{d}_{ij}^e \bar{u}_{ij}^e$, the direct calculation of which avoids rounding errors due to division by d_e^s and \bar{d}_i^e .

We begin with the presentation of the flux limiter for cell averages. Instead of finding a correction factor $\bar{\alpha}^s$ which guarantees the validity of (70) and (76) for $\bar{f}_e^{s,*} = \bar{\alpha}^s \bar{f}_e^s$, we define the limited flux $\bar{f}_e^{s,*}$ directly as in [27]. For any $e' \in \mathcal{E}^e$, the fluxes $\bar{f}_{ee'}^* = \bar{f}_e^{s,*}$ and $\bar{f}_{e'e}^* = \bar{f}_{e'}^{s',*}$ associated with the edge or face $\partial K_s^e = \partial K^e \cap \partial K^{e'} = \partial K_{s'}^{e'}$ must have the same magnitude and opposite signs for the local conservation property to be preserved. Hence, the formula for $\bar{f}_{ee'}^*$ must guarantee the validity of inequality constraints for $\bar{f}_{e'e}^*$ and vice versa. This requirement leads to the definition (cf. [27])

$$\bar{f}_e^{s,*} = \begin{cases} \min\{\bar{f}_e^s, \bar{f}_e^{s,\max}, -\bar{f}_{e'}^{s',\min}\} & \text{if } \bar{f}_e^s > 0, \partial K_s^e = \partial K_{s'}^{e'}, \\ \max\{\bar{f}_e^s, \bar{f}_e^{s,\min}, -\bar{f}_{e'}^{s',\max}\} & \text{if } \bar{f}_e^s < 0, \partial K_s^e = \partial K_{s'}^{e'}, \\ \min\{\bar{f}_e^s, \bar{f}_e^{s,\max}\} & \text{if } \bar{f}_e^s > 0, \partial K_s^e \subseteq \Gamma, \\ \max\{\bar{f}_e^s, \bar{f}_e^{s,\min}\} & \text{if } \bar{f}_e^s < 0, \partial K_s^e \subseteq \Gamma, \\ 1 & \text{otherwise.} \end{cases} \quad (80)$$

To satisfy conditions (77) for all nodes of element K^e , we define the limited antidiffusive element contributions $f_i^{e,*} = \alpha_i^e f_i^e$ using formula (63). That is, our monolithic convex limiting approach differs from the FCT version only in the definition of the bounds $f_i^{e,\max}$ and $f_i^{e,\min}$ for the computation of \tilde{f}_i^e .

Theorem 4 (LED property of the monolithic limiter). *For $\bar{f}_e^{s,*}$ and $f_i^{e,*}$ defined by (80) and (63), the fully discrete EG-MCL scheme*

$$\bar{u}_e^{\text{MCL}} = \bar{u}_e + \frac{\Delta t}{|K^e|} \left[\bar{g}_e^L(\bar{u}) + \sum_{s=1}^{S^e} \bar{f}_e^{s,*}(u^c) \right], \quad e = 1, \dots, E_h, \quad (81)$$

$$u_i^{c,\text{MCL}} = u_i^c + \frac{\Delta t}{m_i} \left[g_i^L(u^c) + \sum_{e \in \mathcal{E}_i} f_i^{e,*}(\bar{u}, u^c) \right], \quad i = 1, \dots, N_h \quad (82)$$

is LED w.r.t. the MCL bounds (72)–(75) under the same time step restrictions as the low-order schemes analyzed in Theorems 1 and 2.

PROOF. By virtue of (67)–(71), the update defined by (81) and (82) yields

$$\begin{aligned} \bar{u}_e^{\text{MCL}} &= \bar{u}_e + \frac{\Delta t}{|K^e|} \sum_{s=1}^{S^e} 2d_e^s (\bar{u}_e^{s,*} - \bar{u}_e), \\ &= \left(1 - \frac{\Delta t}{|K^e|} \sum_{s=1}^{S^e} 2d_e^s \right) \bar{u}_e + \left(\frac{\Delta t}{|K^e|} \sum_{s=1}^{S^e} 2d_e^s \right) \bar{u}_e^{s,*} \end{aligned}$$

and

$$\begin{aligned} u_i^{c,\text{MCL}} &= u_i^c + \frac{\Delta t}{m_i} \sum_{e \in \mathcal{E}_i} 2\bar{d}_i^e (\bar{u}_i^{e,*} - u_i^c) \\ &= \left(1 - \frac{\Delta t}{m_i} \sum_{e \in \mathcal{E}_i} \sum_{j \in \mathcal{N}^e} 2\bar{d}_{ij}^e \right) u_i^c + \left(\frac{\Delta t}{m_i} \sum_{e \in \mathcal{E}_i} \sum_{j \in \mathcal{N}^e} 2\bar{d}_{ij}^e \right) \bar{u}_i^{e,*}, \end{aligned}$$

where the states $\bar{u}_e^{s,*}$ and $\bar{u}_i^{e,*}$ are constrained to satisfy (70) and (71). \square

The steady-state limits of the nonlinear discrete problems (81) and (82) are well-defined and can be analyzed as in [3, 32]. The continuity requirements for the MCL version of (63) are satisfied because the limited element contributions can be written as $f_i^{e,*} = c_i^+ f_i^{e,\max} + c_i^- f_i^{e,\min}$ with bounded coefficients $c_i^\pm \geq 0$. We remark that boundedness of c_i^\pm cannot be guaranteed for

element-based limiters that use the same correction factor $\alpha^e = \min_{j \in \mathcal{N}^e} R_j^e$ for all components. As a consequence, steady-state residuals of resulting schemes may begin to stagnate after reaching a certain threshold.

An implicit treatment of our semi-discrete problem requires iterative solution of nonlinear systems [22, 25]. For example, let us discretize (65) and (66) in time using an implicit two-level θ scheme corresponding, e.g., to the Crank-Nicolson ($\theta = \frac{1}{2}$) or backward Euler ($\theta = 1$) method. A simple fixed-point iteration is defined by the deferred correction scheme

$$|K^e| \bar{u}_e^{m+1} - \theta \Delta t \bar{g}_e^L(\bar{u}^{m+1}) = |K^e| \bar{u}_e + (1 - \theta) \Delta t \bar{g}_e^L(\bar{u}) + \Delta t \sum_{s=1}^{S^e} [\theta \bar{f}_e^{s,*}(u^{c,m}) + (1 - \theta) \bar{f}_e^{s,*}(u^c)], \quad (83)$$

$$m_i u_i^{c,m+1} - \theta \Delta t g_i^L(u^{c,m+1}) = m_i u_i^c + (1 - \theta) \Delta t g_i^L(u^c) + \Delta t \sum_{e \in \mathcal{E}_i} [\theta f_i^{e,*}(\bar{u}^m, u^{c,m}) + (1 - \theta) f_i^{e,*}(\bar{u}, u^c)], \quad (84)$$

where $m \geq 0$ is the iteration index and \bar{u}, u^c without iteration index refer to values at the old time step. The initial guesses \bar{u}^0 and $u^{c,0}$ can be defined using the solution values from the previous time step or calculated by solving linear systems associated with the low-order LED approximation.

The convergence behavior of fixed point iterations for nonlinear AFC problems often leaves a lot to be desired, especially for time steps larger than the CFL bounds for explicit schemes. Significant speedups can be achieved, e.g., using Anderson acceleration [25] or quasi-Newton methods [2, 33].

6. Constrained projections

In Sections 4 and 5, we constrained the cell averages \bar{u}_e and CG components u_i^c to satisfy local discrete maximum principles. The restriction of the EG solution $u_h = u_h^c + u_h^d$ to a mesh cell K^e , $e = 1, \dots, E_h$ is given by

$$u_h(\mathbf{x}) = \bar{u}_e + [u_h^c(\mathbf{x}) - \bar{u}_e^c] \quad \forall \mathbf{x} \in K^e \quad (85)$$

and is generally not bound-preserving. It is neither easy nor necessary to constrain \bar{u}_e and u_h^c in a way which guarantees the LED property for u_h . Similarly to finite volume methods, the main physically meaningful unknowns of a locally conservative approximation are the cell averages. Additional unknowns of DG and EG finite element schemes are needed just to evolve the

cell averages in a more accurate way and to define more accurate reconstructions from cell averages. When it comes to visualization of EG solutions or their use in other equations of a coupled problem, the reconstruction defined by (85) can be postprocessed using an algebraic form of slope limiting to enforce preservation of local bounds for the output.

Let u_h^* denote a limited projection of the EG approximation (85) into the space V_h^c or its DG counterpart. The vertex-based limiting procedure presented in [24] can be used to enforce the inequality constraints

$$\bar{u}_i^{\min} \leq u_h^{e,*}(\mathbf{x}_i) \leq \bar{u}_i^{\max} \quad \forall i \in \mathcal{N}^e, \quad (86)$$

$$\bar{u}_i^{\max} = \max \left\{ \max_{e \in \mathcal{E}_i} \bar{u}_e, \max_{j \in \mathcal{N}_i} u_j^c \right\}, \quad \bar{u}_i^{\min} = \min \left\{ \min_{e \in \mathcal{E}_i} \bar{u}_e, \min_{j \in \mathcal{N}_i} u_j^c \right\} \quad (87)$$

using the element-based correction factor

$$\gamma^e = \min_{i \in \mathcal{N}^e} \begin{cases} \min \left\{ 1, \frac{\bar{u}_i^{\max} - \bar{u}_e}{u_h^c(\mathbf{x}_i) - \bar{u}_e^c} \right\} & \text{if } u_h^c(\mathbf{x}_i) > \bar{u}_e^c, \\ \min \left\{ 1, \frac{\bar{u}_i^{\min} - \bar{u}_e}{u_h^c(\mathbf{x}_i) - \bar{u}_e^c} \right\} & \text{if } u_h^c(\mathbf{x}_i) < \bar{u}_e^c, \\ 1 & \text{if } u_h^c(\mathbf{x}_i) = \bar{u}_e^c \end{cases} \quad (88)$$

to define the ‘slope-limited’ DG reconstruction

$$u_h^{e,*}(\mathbf{x}) = \bar{u}_e + \gamma^e [u_h^c(\mathbf{x}) - \bar{u}_e^c] \quad \forall \mathbf{x} \in K^e. \quad (89)$$

Note that the multiplication of the massless $\mathbb{P}_1/\mathbb{Q}_1$ component $u_h^c(\mathbf{x}) - \bar{u}_e^c$ by any $\gamma^e \in \mathbb{R}$ does not change the cell average.

A globally conservative constrained L^2 projection into the CG space V_h^c can be performed using the FCT methodology (cf. [28, 35]) to enforce (86). Let \tilde{u} be a given function that may represent, e.g., the initial data u_0 or the EG solution u_h at the final time. The standard L^2 projection defined by

$$\sum_{e \in \mathcal{E}_i} \sum_{j \in \mathcal{N}^e} m_{ij}^e u_j^H = \sum_{e \in \mathcal{E}_i} \int_{K^e} \varphi_i \tilde{u} d\mathbf{x}, \quad i = 1, \dots, N_h \quad (90)$$

provides the high-order target u_h^H for flux-corrected remapping (FCR). Using the cell averages of \tilde{u} , let us define the low-order reconstruction

$$u_i^L = \frac{1}{m_i} \sum_{e \in \mathcal{E}_i} m_i^e \tilde{u}_e, \quad \tilde{u}_e = \frac{1}{|K^e|} \int_{K^e} \tilde{u} d\mathbf{x} \quad (91)$$

as in formula (17). Since u_i^L is a convex combination of \tilde{u}_e , it is bounded by the local extrema $\bar{u}_i^{\max} = \max_{e \in \mathcal{E}_i} \tilde{u}_e$ and $\bar{u}_i^{\min} = \min_{e \in \mathcal{E}_i} \tilde{u}_e$.

To construct a better LED approximation to u_h^H , we notice that

$$u_i^H = u_i^L + \frac{1}{m_i} \sum_{e \in \mathcal{E}_i} \tilde{f}_i^e, \quad i = 1, \dots, N_h, \quad (92)$$

where

$$\tilde{f}_i^e = \int_{K^e} \varphi_i(u_i^H - u_h^H + \tilde{u} - \tilde{u}_e) \mathrm{d}\mathbf{x}, \quad \sum_{i \in \mathcal{N}^e} \tilde{f}_i^e = 0. \quad (93)$$

Using this relationship between u_h^H and u_h^L , we define

$$u_i^* = u_i^L + \frac{1}{m_i} \sum_{e \in \mathcal{E}_i} \gamma^e \tilde{f}_i^e, \quad i = 1, \dots, N_h \quad (94)$$

using the element-based correction factor

$$\gamma^e = \min_{i \in \mathcal{N}^e} \begin{cases} \min \left\{ 1, \frac{m_i^e (\bar{u}_i^{\max} - u_i^L)}{\tilde{f}_i^e} \right\} & \text{if } \tilde{f}_i^e > 0, \\ \min \left\{ 1, \frac{m_i^e (\bar{u}_i^{\min} - u_i^L)}{\tilde{f}_i^e} \right\} & \text{if } \tilde{f}_i^e < 0, \\ 1 & \text{otherwise.} \end{cases} \quad (95)$$

Note that this formula has the same structure as (88). The LED property of the element-based FCR algorithm can be shown as in Theorem 3.

Remark 7. For an EG approximation $\tilde{u} = u_h^c + u_h^d$, we have $\tilde{u}|_{K^e} - \tilde{u}_e = u_h^c|_{K^e} + (\bar{u}_e - \bar{u}_e^c) - \bar{u}_e = u_h^c|_{K^e} - \bar{u}_e^c$. Hence, the element contributions to be limited are given by $\tilde{f}_i^e = \int_{K^e} \varphi_i(u_i^H - u_h^H + u_h^c - \bar{u}_e^c) \mathrm{d}\mathbf{x}$. The cost of solving the linear system (90) for u^H can be reduced using a truncated series approximation to the inverse of the consistent mass matrix [13, 26].

7. Numerical examples

In this section, we study the numerical behavior of the presented limiting techniques for the EG method in 2D. Computations are performed on triangular and rectangular meshes. We illustrate the stabilizing effect of the DG- \mathbb{P}_0 enrichment and verify the ability of FCT and MCL limiters (as presented in Sections 4 and 5, respectively) to enforce discrete maximum principles in EG schemes. For visualization purposes and for quantitative

assessment of numerical errors, we project the results into the CG space V_h^c using the FCR procedure presented in [Section 6](#). The error of a constrained EG approximation is defined as $u - u_h^*$, where u is the exact (or reference) solution and $u_h^* \in V_h^c$ is the FCR projection of the numerical solution. We measure the accuracy of u_h^* using the discrete error norms [\[25\]](#)

$$E_1(h) := \sum_{i=1}^{N_h} m_i |u(\mathbf{x}_i) - u_i^*| \approx \int_{\Omega} |u - u_h^*| d\mathbf{x} = \|u - u_h^*\|_{L^1(\Omega)}, \quad (96)$$

$$E_2(h) := \sqrt{\sum_{i=1}^{N_h} m_i (u(\mathbf{x}_i) - u_i^*)^2} \approx \sqrt{\int_{\Omega} (u - u_h^*)^2 d\mathbf{x}} = \|u - u_h^*\|_{L^2(\Omega)}. \quad (97)$$

The experimental order of convergence is determined using the formula [\[31\]](#)

$$p = \log_2 \left(\frac{E_i(2h)}{E_i(h)} \right), \quad i = 1, 2. \quad (98)$$

The values of the global maxima $u^{\max} = \max_{1 \leq j \leq N_h} u_j^*$ and minima $u^{\min} = \min_{1 \leq j \leq N_h} u_j^*$ are reported as well to show the LED properties of the schemes under investigation and quantify the levels of numerical dissipation.

7.1. Solid body rotation

We begin with the solid body rotation benchmark [\[31\]](#). This difficult 2D test is widely used in numerical studies of algebraic flux correction schemes [\[16, 25, 26, 27, 32\]](#) and variational shock capturing techniques for stabilized finite element methods [\[23\]](#). Hence, the results to be presented can be compared to numerical solutions obtained with CG and DG approaches.

In the solid body rotation experiment, the divergence-free velocity field $\mathbf{v}(x, y) = (0.5 - y, x - 0.5)^\top$ is used to rotate a slotted cylinder, a sharp cone, and a smooth hump around the center $(0.5, 0.5)$ of the domain $\Omega = (0, 1)^2$. The initial condition, as defined by LeVeque [\[31\]](#), is given by

$$u_0(x, y) = \begin{cases} u_0^{\text{hump}}(x, y) & \text{if } \sqrt{(x - 0.25)^2 + (y - 0.5)^2} \leq 0.15, \\ u_0^{\text{cone}}(x, y) & \text{if } \sqrt{(x - 0.5)^2 + (y - 0.25)^2} \leq 0.15, \\ 1 & \text{if } \left\{ \left(\sqrt{(x - 0.5)^2 + (y - 0.75)^2} \leq 0.15 \right) \wedge \left(|x - 0.5| \geq 0.025 \vee y \geq 0.85 \right) \right\}, \\ 0 & \text{otherwise,} \end{cases}$$

where

$$u_0^{\text{hump}}(x, y) = \frac{1}{4} + \frac{1}{4} \cos \left(\frac{\pi \sqrt{(x - 0.25)^2 + (y - 0.5)^2}}{0.15} \right), \quad (99)$$

$$u_0^{\text{cone}}(x, y) = 1 - \frac{\sqrt{(x - 0.5)^2 + (y - 0.25)^2}}{0.15}. \quad (100)$$

Homogeneous Dirichlet boundary conditions are prescribed at the inlets.

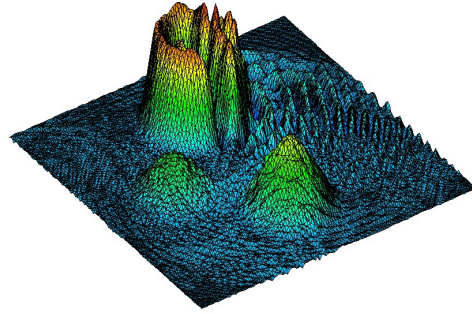
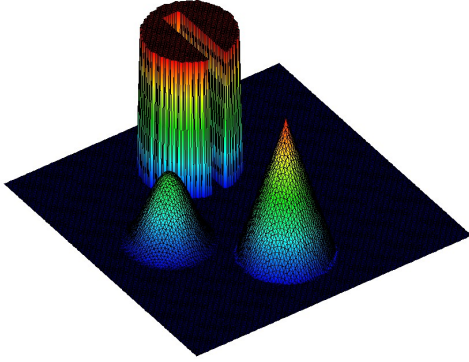
After each full rotation, the exact solution $u(\cdot, 2\pi k)$, $k \in \mathbb{N}$ coincides with the initial data u_0 . Time integration is performed using the optimal second-order explicit Runge-Kutta method and the time step $\Delta t = 10^{-3}$. The $\mathbb{P}_1 \oplus \mathbb{P}_0$ and $\mathbb{Q}_1 \oplus \mathbb{P}_0$ versions of the EG method are used for discretization in space on triangular and rectangular uniform meshes with spacing $h = \frac{1}{128}$.

In Figures 1 and 2, we present the continuous projection $u_h^0 = u_h(\cdot, 0)$ of the initial data u_0 into V_h^c and (FCR projections of) numerical solutions u_h after one full rotation ($t = 2\pi$). For a better comparison of the methods under investigation, the E_1 error w.r.t. the reference solution $u(\cdot, 2\pi) = u_h^0$ and the range of numerical solution values are shown above each plot. The CG- \mathbb{P}_1 and CG- \mathbb{Q}_1 results are displayed in Figs 1(b) and 2(b), respectively. They are highly oscillatory even in regions where the exact solution is smooth. The global spreading of numerical errors is responsible for the suboptimal convergence behavior of the continuous Galerkin method and strong ‘terracing’ effects (spurious distortions) in flux-limited CG approximations [34].

The diagrams shown in the bottom rows of Figs 1 and 2 demonstrate that the stabilizing effect of the \mathbb{P}_0 enrichment provides an effective cure for global oscillations. The stability of the $\mathbb{P}_1 \oplus \mathbb{P}_0$ version is backed by the theoretical analysis of Becker et al. [7]. The $\mathbb{Q}_1 \oplus \mathbb{P}_0$ approach performs much better than CG- \mathbb{Q}_1 . However, no rigorous proof of stability is available, and the solutions presented in Figs 2(c),(d) are not quite as accurate as their $\mathbb{P}_1 \oplus \mathbb{P}_0$ counterparts shown in Figs 1(c),(d). The smooth waves emanating from the slotted cylinder indicate that a piecewise-constant enrichment of the CG- \mathbb{Q}_1 approximation might be insufficient to ensure optimal convergence behavior. The use of $\dot{u}_i^H = 0$ in the definition of f_i^e introduces more numerical damping than formula (17). This property of the lumped-mass approximation is reflected in the smaller magnitude of spurious undershoots/overshoots and smaller values of the E_1 errors. The activation of limiters makes \dot{u}_i^H defined by (17) a better choice than $\dot{u}_i^H = 0$ because stronger damping tends to have an adverse effect on the overall accuracy of an LED approximation.

(a) $E_1 = 0.00\text{e-}0$, $u_h \in [0.0, 1.0]$

(b) $E_1 = 4.43\text{e-}2$, $u_h \in [-0.354, 1.451]$



(c) $E_1 = 2.57\text{e-}2$, $u_h \in [-0.116, 1.169]$

(d) $E_1 = 3.13\text{e-}2$, $u_h \in [-0.207, 1.313]$

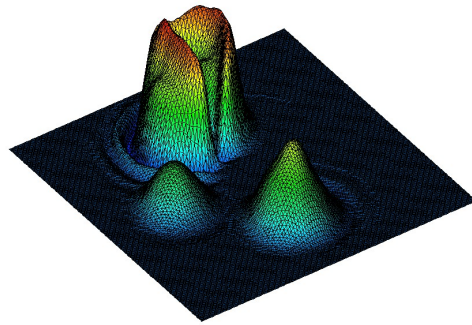
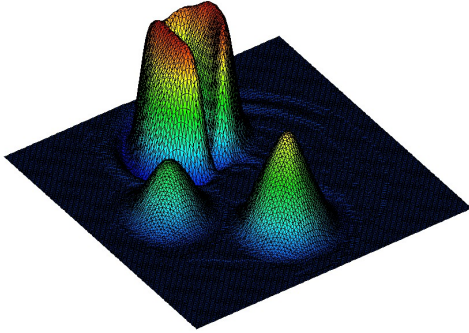
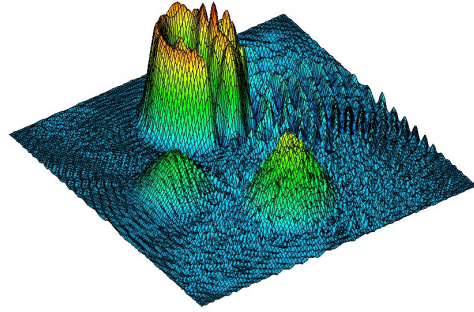
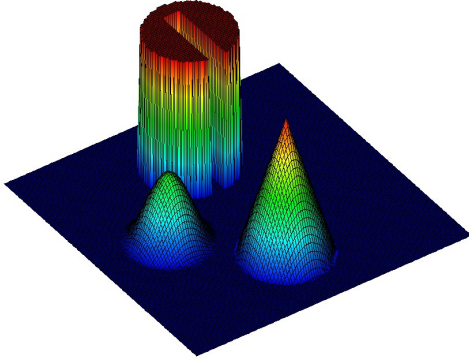


Fig. 1: Solid body rotation, triangular mesh, $h = \frac{1}{128}$, $\Delta t = 10^{-3}$. The plots show (a) CG- \mathbb{P}_1 interpolant of u_0 and numerical solutions at $t = 2\pi$ obtained with (b) CG- \mathbb{P}_1 , (c) EG- $\mathbb{P}_1 \oplus \mathbb{P}_0$ with $\dot{u}_i^H = 0$, (d) EG- $\mathbb{P}_1 \oplus \mathbb{P}_0$ with \dot{u}_i^H defined by (17).

(a) $E_1 = 0.00\text{e-}0$, $u_h \in [0.0, 1.0]$

(b) $E_1 = 4.43\text{e-}2$, $u_h \in [-0.369, 1.483]$



(c) $E_1 = 2.86\text{e-}2$, $u_h \in [-0.098, 1.150]$

(d) $E_1 = 3.13\text{e-}2$, $u_h \in [-0.207, 1.314]$

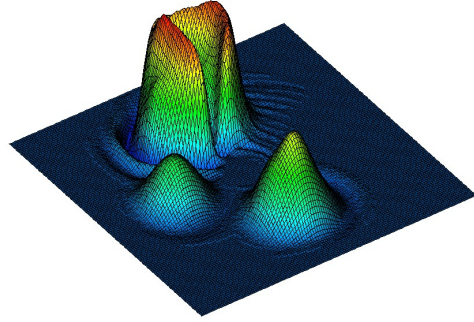
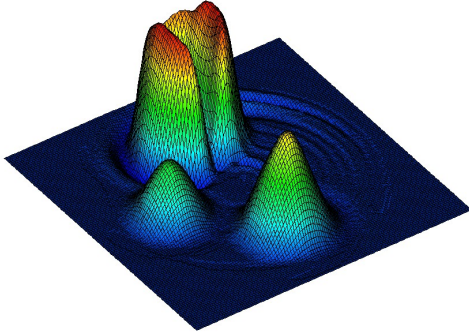
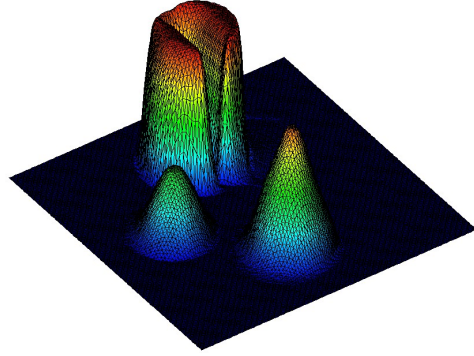
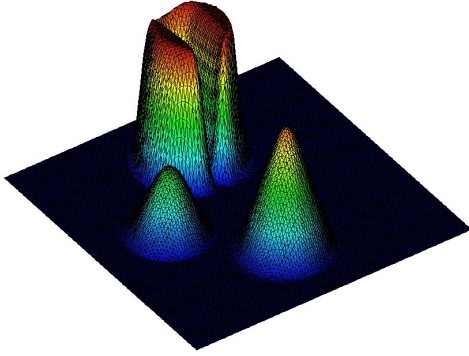


Fig. 2: Solid body rotation, rectangular mesh, $h = \frac{1}{128}$, $\Delta t = 10^{-3}$. The plots show (a) CG- Q_1 interpolant of u_0 and numerical solutions at $t = 2\pi$ obtained with (b) CG- Q_1 , (c) EG- $Q_1 \oplus P_0$ with $\dot{u}_i^H = 0$, (d) EG- $Q_1 \oplus P_0$ with \dot{u}_i^H defined by (17).

(a) $E_1 = 1.96\text{e-}2$, $u_h \in [0.0, 1.0]$

(b) $E_1 = 1.90\text{e-}2$, $u_h \in [0.0, 1.0]$



(c) $E_1 = 1.94\text{e-}2$, $u_h \in [0.0, 1.0]$

(d) $E_1 = 1.88\text{e-}2$, $u_h \in [0.0, 1.0]$

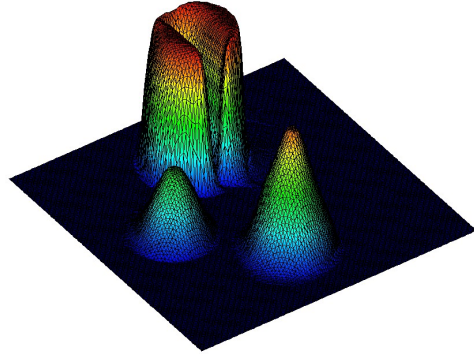
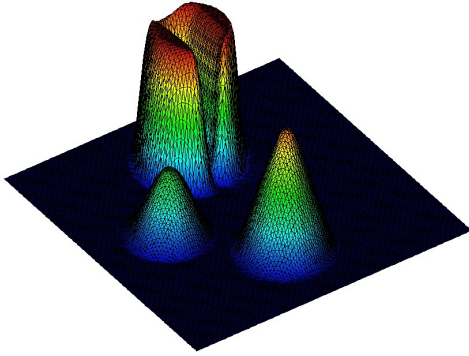
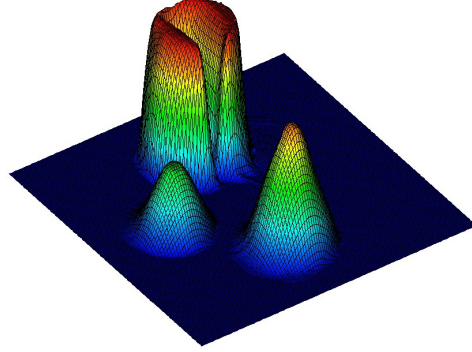
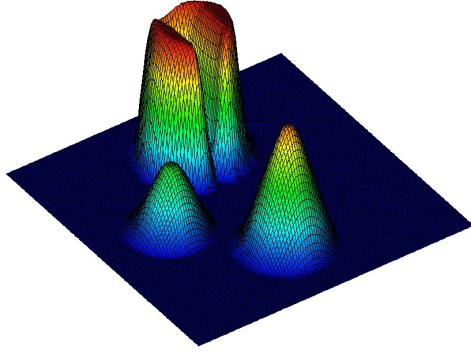


Fig. 3: Solid body rotation, triangular mesh, $h = \frac{1}{128}$, $\Delta t = 10^{-3}$. The plots show the $\text{EG-}\mathbb{P}_1 \oplus \mathbb{P}_0$ solutions at $t = 2\pi$ obtained using (a) FCT and $\dot{u}_i^H = 0$, (b) FCT and \dot{u}_i^H defined by (17), (c) MCL and $\dot{u}_i = 0$, (d) MCL and \dot{u}_i defined by (17).

(a) $E_1 = 2.19\text{e-}2$, $u_h \in [0.0, 0.997]$

(b) $E_1 = 2.12\text{e-}2$, $u_h \in [0.0, 998]$



(c) $E_1 = 2.23\text{e-}2$, $u_h \in [0.0, 0.995]$

(d) $E_1 = 2.11\text{e-}2$, $u_h \in [0.0, 0.999]$

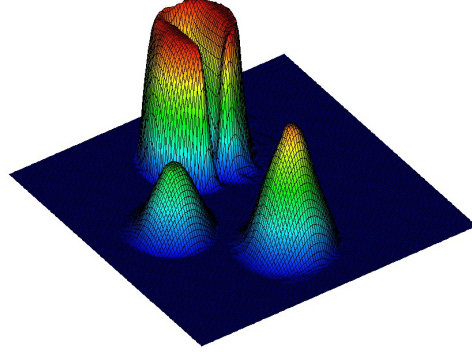
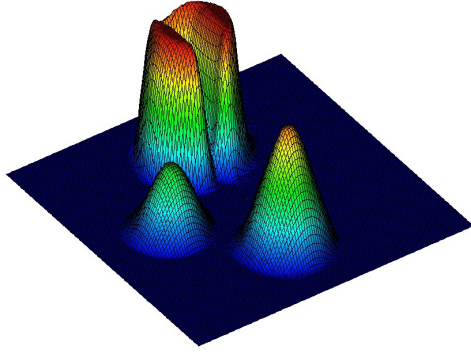


Fig. 4: Solid body rotation, rectangular mesh, $h = \frac{1}{128}$, $\Delta t = 10^{-3}$. The plots show the EG- $\mathbb{Q}_1 \oplus \mathbb{P}_0$ solutions at $t = 2\pi$ obtained using (a) FCT and $\dot{u}_i^H = 0$, (b) FCT and \dot{u}_i^H defined by (17), (c) MCL and $\dot{u}_i = 0$, (d) MCL and \dot{u}_i defined by (17).

The results presented in Figs 3 and 4 illustrate the ability of our limiting techniques to enforce discrete maximum principles without distorting smooth solution profiles. Remarkably, even the peaks of the rotating cone and hump are preserved very well, although our LED approximations set the antidiffusive terms to zero at any local extremum. The presented FCT and MCL results are virtually indistinguishable. Hence, the accuracy of the constrained EG approximations is largely determined by the choice of \dot{u}_i^H and by the definition of the local bounds. We conclude that both FCT and MCL are well suited for explicit EG discretizations of time-dependent advection problems. However, the FCT correction factors depend on the time step, which makes MCL a better limiter for implicit schemes and steady-state computations.

7.2. Steady circular advection

In the second numerical experiment, we solve the steady-state counterpart of (1a) in $\Omega = (0, 1)^2$ using the velocity field $\mathbf{v}(x, y) = (y, -x)$. The inflow boundary condition and the exact solution at any point in $\bar{\Omega}$ are given by

$$u(x, y) = \begin{cases} 1, & \text{if } 0.15 \leq r(x, y) \leq 0.45, \\ \cos^2 \left(10\pi \frac{r(x, y) - 0.7}{3} \right), & \text{if } 0.55 \leq r(x, y) \leq 0.85, \\ 0, & \text{otherwise,} \end{cases} \quad (101)$$

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

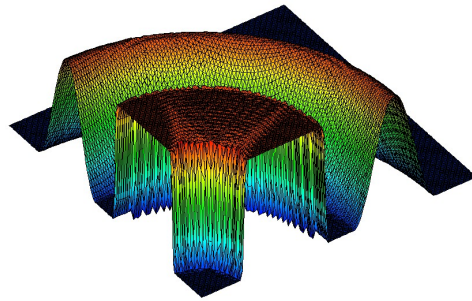
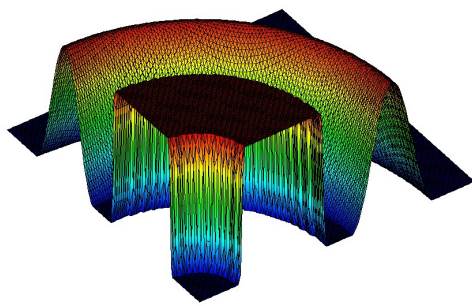
Numerical solutions are marched to the steady state using uniform triangular meshes and the same explicit SSP Runge-Kutta scheme as in the solid body rotation test. Computations are terminated when the L^1 norm of the pseudo-time derivative \dot{u}_h , as defined by (17), becomes smaller than the prescribed tolerance. The CG- \mathbb{P}_1 projection of the exact solution and the numerical approximations corresponding to $h = \frac{1}{128}$ are shown in Fig. 5. The enrichment of the oscillatory CG- \mathbb{P}_1 approximation with the \mathbb{P}_0 component of the EG method reduces the magnitude of spurious undershoots/overshoots and localizes them to thin layers around the discontinuities of the exact solution. The nonlinear MCL correction enforces strict preservation of local bounds, while leaving smooth portions of the EG solution unchanged.

To study the numerical behavior of EG- $\mathbb{P}_1 \oplus \mathbb{P}_0$ approximations to problems with smooth solutions, we perform grid convergence studies for [32]

$$u(x, y) = \exp \left(-100(r(x, y) - 0.7)^2 \right), \quad 0 \leq x, y \leq 1 \quad (102)$$

(a) $E_1 = 0.00e-0$, $u_h \in [0.0, 1.0]$

(b) $E_1 = 8.50e-3$, $u_h \in [-0.097, 1.106]$



(c) $E_1 = 3.91e-3$, $u_h \in [-0.049, 1.063]$

(d) $E_1 = 4.04e-3$, $u_h \in [0.0, 1.0]$

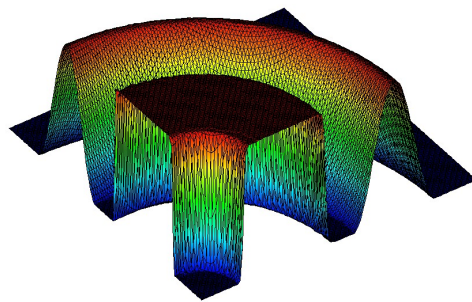
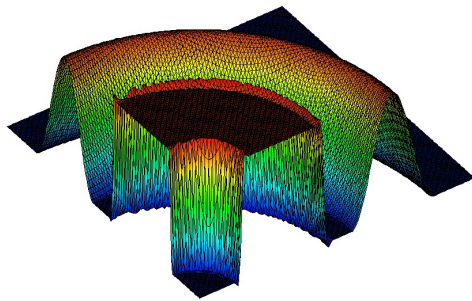


Fig. 5: Steady advection, triangular mesh, $h = \frac{1}{128}$. The plots show (a) CG- \mathbb{P}_1 interpolant of the exact solution and stationary numerical solutions obtained with (b) CG- \mathbb{P}_1 , (c) EG- $\mathbb{P}_1 \oplus \mathbb{P}_0$, (d) EG- $\mathbb{P}_1 \oplus \mathbb{P}_0 + \text{MCL}$.

using the same velocity field. The convergence history presented in Tables 1 and 2 demonstrates that second-order superconvergence can be achieved with EG- $\mathbb{P}_1 \oplus \mathbb{P}_0$ on uniform triangular meshes. The unlimited solutions have small undershoots and overshoots on all mesh levels. The EG- $\mathbb{P}_1 \oplus \mathbb{P}_0$ scheme equipped with the MCL limiter produces larger E_2 errors on coarse meshes but converges quadratically as well and guarantees the validity of the global discrete maximum principle w.r.t. the weakly imposed Dirichlet boundary condition on Γ_- . As the mesh is refined, the global maxima and minima of the numerical solutions approach the exact values $u^{\max} = 0$ and $u^{\min} = 1$. On the finest mesh, the MCL convergence rate drops to 1.70, which is still higher than the provable order of accuracy for EG discretizations of pure advection problems. The lack of superconvergence for $h = \frac{1}{256}$ can be attributed to the increased influence of rounding errors in the computation of limited fluxes and correction factors. Further research is required to verify this conjecture and design efficient implementations of EG-MCL for stationary problems.

h	E_2	p	u^{\min}	u^{\max}
$\frac{1}{32}$	5.33e-3		-6.210e-3	1.0073
$\frac{1}{64}$	1.26e-3	2.08	-2.726e-3	1.0028
$\frac{1}{128}$	2.91e-4	2.11	-1.037e-3	1.0010
$\frac{1}{256}$	7.08e-5	2.04	-3.638e-4	1.0002

Table 1: Steady advection. EG- $\mathbb{P}_1 \oplus \mathbb{P}_0$ convergence history.

h	E_2	p	u^{\min}	u^{\max}
$\frac{1}{32}$	9.00e-3		0.0	0.984
$\frac{1}{64}$	2.40e-3	1.91	0.0	0.995
$\frac{1}{128}$	6.19e-4	1.96	0.0	0.999
$\frac{1}{256}$	1.90e-4	1.70	0.0	1.000

Table 2: Steady advection. EG- $\mathbb{P}_1 \oplus \mathbb{P}_0 + \text{MCL}$ convergence history.

In Figure 6, we show the evolution history of the steady-state residuals $\int_{\Omega} |\dot{u}_h| dx$ for EG and EG-MCL on the finest mesh. It can be seen that our monolithic limiting strategy does not inhibit convergence to stationary solutions. The MCL-corrected residuals converge monotonically to machine zero and exhibit the same qualitative behavior as the EG residuals. The

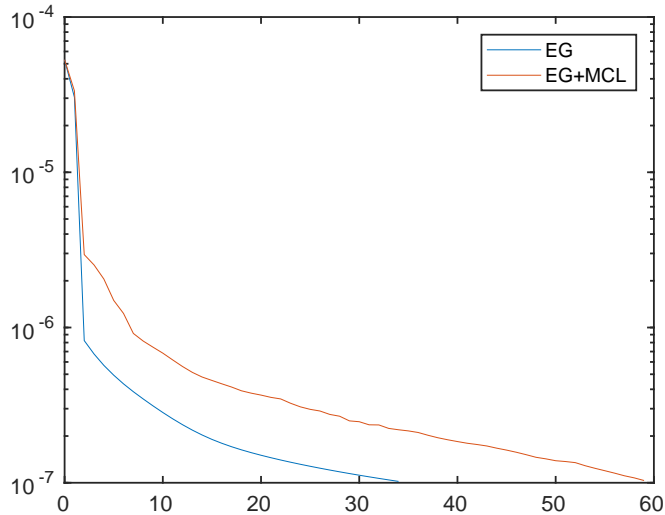


Fig. 6: Steady advection, $h = \frac{1}{256}$, steady state residuals $\int_{\Omega} |\dot{u}_h| dx$ vs. pseudo-time.

numbers of pseudo-time steps required to reach the tolerance 10^{-7} differ by a factor of 1.7. This is a remarkably small difference, given the fact that the steady-state problem is linear for EG and nonlinear for EG-MCL.

8. Conclusions

The methodology developed in this paper exploits the variational multi-scale nature of the locally conservative EG method leading to a well-organized interplay of its CG and DG components. The primary unknowns are the EG cell averages which we evolve directly. The CG component provides the antidiffusive fluxes for the evolution of cell averages, while the cell averages and their time derivatives are used to calculate the antidiffusive element contributions of the algebraic limiting procedure for the CG subproblem. The bound-preserving properties of the presented limiting techniques were shown in the context of explicit SSP Runge-Kutta schemes. The monolithic limiting approach is readily applicable to implicit time discretizations and stationary problems as well. The well-posedness of nonlinear discrete problems and validity of generalized discrete maximum principles can be verified using the proof techniques developed in [3, 5, 27, 32] for edge-based algebraic flux correction schemes. The design of limiters for EG discretizations of parabolic

and elliptic problems requires additional theoretical and numerical studies. It is hoped that the present paper provides a good stepping stone toward further development of provably bound-preserving EG methods.

Acknowledgments

The development of limiters for the enriched Galerkin method was initiated by Prof. Mary F. Wheeler during the first author's visit to the University of Texas at Austin. The authors thank Dr. Florian Frank (University Erlangen-Nuremberg) and Prof. Vadym Aizinger (University of Bayreuth) for helpful remarks regarding the properties of the EG approximation and early versions of the proposed limiting framework. This work was supported by the German Research Association (DFG) under grant KU 1530/23-1. A. Rupp acknowledges financial support by the DFG EXC 2181.

In memoriam

This paper is dedicated to the memory of Dr. Douglas Nelson Woods (*January 11th 1985 - †September 11th 2019), promising young scientist and post-doctoral research fellow at Los Alamos National Laboratory. Our thoughts and wishes go to his wife Jessica, to his parents Susan and Tom, to his sister Rebecca and to his brother Chris, whom he left behind.

References

- [1] R. Anderson, V. Dobrev, Tz. Kolev, D. Kuzmin, M. Quezada de Luna, R. Rieben, and V. Tomov, High-order local maximum principle preserving (MPP) discontinuous Galerkin finite element method for the transport equation. *J. Comput. Phys.* **334** (2017) 102–124.
- [2] S. Badia and J. Bonilla, Monotonicity-preserving finite element schemes based on differentiable nonlinear stabilization. *Computer Methods Appl. Mech. Engrg.* **313** (2017) 133-158.
- [3] G. Barrenechea, V. John, and P. Knobloch, Analysis of algebraic flux correction schemes. *SIAM J. Numer. Anal.* **54** (2016) 2427–2451.
- [4] G. Barrenechea, V. John, and P. Knobloch, A linearity preserving algebraic flux correction scheme satisfying the discrete maximum principle on general meshes. *Mathematical Models and Methods in Applied Sciences (M3AS)* **27** (2017) 525–548.

- [5] G. Barrenechea, V. John, P. Knobloch, and R. Rankin, A unified analysis of algebraic flux correction schemes for convection-diffusion equations. *SeMA* **75** (2018) 655–685.
- [6] R. Becker, M. Bittl, and D. Kuzmin, Analysis of a combined CG1–DG2 method for the transport equation. *SIAM J. Numer. Anal.* **53** (2015) 445–463.
- [7] R. Becker, E. Burman, P. Hansbo, and M.G. Larson, A reduced P^1 -discontinuous Galerkin method. Chalmers Finite Element Center Preprint 2003-13, Chalmers University of Technology, Göteborg, Sweden, 2003.
- [8] M. Bittl, D. Kuzmin, and R. Becker, The CG1–DG2 method for convection-diffusion equations in 2D. *J. Comput. Appl. Math.* **270** (2014) 21–31.
- [9] C.J. Cotter and D. Kuzmin, Embedded discontinuous Galerkin transport schemes with localised limiters. *J. Comput. Phys.* **311** (2016) 363–373.
- [10] S. Franz, Analysis of a family of continuous–discontinuous Galerkin FEM for convection–diffusion problems. *Applied Numerical Mathematics* **110** (2016) 93–109.
- [11] S. Gottlieb, C.-W. Shu, and E. Tadmor, Strong stability-preserving high-order time discretization methods. *SIAM Review* **43** (2001) 89–112.
- [12] J.-L. Guermond, M. Nazarov, B. Popov, and I. Tomas, Second-order invariant domain preserving approximation of the Euler equations using convex limiting. *SIAM J. Sci. Computing* **40** (2018) A3211–A3239.
- [13] J.-L. Guermond, M. Nazarov, B. Popov, and Y. Yang, A second-order maximum principle preserving Lagrange finite element technique for nonlinear scalar conservation equations. *SIAM Journal on Numerical Analysis* **52** (2014) 2163–2182.
- [14] J.-L. Guermond, M. Nazarov, and I. Tomas, Invariant domain preserving discretization-independent schemes and convex limiting for hyperbolic systems. *Computer Methods Appl. Mech. Engrg.* **347** (2019) 143–175.

- [15] J.-L. Guermond and B. Popov, Invariant domains and first-order continuous finite element approximation for hyperbolic systems. *SIAM J. Numer. Anal.* **54** (2016) 2466–2489.
- [16] H. Hajduk, D. Kuzmin, Tz. Kolev, and R. Abgrall, Matrix-free subcell residual distribution for Bernstein finite element discretizations of linear advection equations. *Computer Methods Appl. Mech. Engrg.* **359** (2020) 112658.
- [17] R. Hartmann and T. Leicht, Higher order and adaptive DG methods for compressible flows. Lecture notes, DLR (German Aerospace Center), 2013.
- [18] T.J.R. Hughes, Multiscale phenomena: Green’s functions, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods. *Computer Methods Appl. Mech. Engineering* **127** (1995) 387–401.
- [19] T.J.R. Hughes, G. R. Feijóo, L. Mazzei, and J.-B. Quincy, The variational multiscale method – a paradigm for computational mechanics. *Computer Methods Appl. Mech. Engineering* **166** (1998) 3–24.
- [20] A. Jameson, Computational algorithms for aerodynamic analysis and design. *Appl. Numer. Math.* **13** (1993) 383–422.
- [21] A. Jameson, Positive schemes and shock modelling for compressible flows. *Int. J. Numer. Methods Fluids* **20** (1995) 743–776.
- [22] A. Jha and V. John, A study of solvers for nonlinear AFC discretizations of convection-diffusion equations, *Computers & Mathematics with Applications* **79** (2019) 3117–3138.
- [23] 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.
- [24] D. Kuzmin, A vertex-based hierarchical slope limiter for p -adaptive discontinuous Galerkin methods. *J. Comput. Appl. Math.* **233** (2010) 3077–3085.

- [25] D. Kuzmin, Algebraic flux correction I. Scalar conservation laws. In: D. Kuzmin, R. Löhner and S. Turek (eds.) *Flux-Corrected Transport: Principles, Algorithms, and Applications*. Springer, 2nd edition: 145–192 (2012).
- [26] D. Kuzmin, Gradient-based limiting and stabilization of continuous Galerkin methods. Chapter 29 in: van Brummelen, E.H., Corsini, A., Perotto, S., Rozza, G. (eds.) *Numerical Methods for Flows: FEF 2017 Selected Contributions*, LNCSE, Springer, 2019.
- [27] D. Kuzmin, Monolithic convex limiting for continuous finite element discretizations of hyperbolic conservation laws. *Comput. Methods Appl. Mech. Engrg.* **361** (2020) 112804.
- [28] D. Kuzmin, M. Möller, J.N. Shadid, and M. Shashkov, Failsafe flux limiting and constrained data projections for equations of gas dynamics. *J. Comput. Phys.* **229** (2010) 8766–8779.
- [29] D. Kuzmin and S. Turek, Flux correction tools for finite elements. *J. Comput. Phys.* **175** (2002) 525–558.
- [30] S. Lee, Y.-J. Lee, and M.F. Wheeler, A locally conservative enriched Galerkin approximation and efficient solver for elliptic and parabolic problems. *SIAM J. Sci. Comput.* **38** (2016) A1404–A1429.
- [31] R.J. LeVeque, High-resolution conservative algorithms for advection in incompressible flow. *SIAM Journal on Numerical Analysis* **33**, (1996) 627–665.
- [32] C. Lohmann, *Physics-Compatible Finite Element Methods for Scalar and Tensorial Advection Problems*. Springer Spektrum, 2019.
- [33] C. Lohmann, On the solvability and iterative solution of algebraic flux correction problems for convection-reaction equations. Preprint *Ergebnisber. Angew. Mathematik* **612** (August 2019), TU Dortmund University.
- [34] C. Lohmann, D. Kuzmin, J.N. Shadid, and S. Mabuza, Flux-corrected transport algorithms for continuous Galerkin methods based on high order Bernstein finite elements. *J. Comput. Phys.* **344** (2017) 151-186.

- [35] R. Löhner, *Applied CFD Techniques: An Introduction Based on Finite Element Methods* (2nd edition). John Wiley & Sons, Chichester, 2008.
- [36] R. Löhner, K. Morgan, J. Peraire, and M. Vahdati, Finite element flux-corrected transport (FEM-FCT) for the Euler and Navier-Stokes equations. *Int. J. Numer. Meth. Fluids* **7** (1987) 1093–1109.
- [37] S. Sun and J. Liu, A locally conservative finite element method based on piecewise constant enrichment of the continuous Galerkin method. *SIAM J. Sci. Comput.* **31** (2009) 2528–2548.
- [38] S.T. Zalesak, Fully multidimensional flux-corrected transport algorithms for fluids. *J. Comput. Phys.* **31** (1979) 335–362.
- [39] G. Zhou, How accurate is the streamline diffusion finite element method? *Math. Comp.* **66** (1997) 31–44.