No. 621 **December 2019**

Matrix-free subcell residual distribution for Bernstein finite elements: Monolithic limiting

**H. Hajduk, D. Kuzmin, T. Kolev, V. Tomov,
I. Tomas, J. N. Shadid**

ISSN: 2190-1767

Matrix-free subcell residual distribution for Bernstein finite elements: Monolithic limiting

Hennes Hajduk[a], Dmitri Kuzmin[a], Tzanio Kolev[b], Vladimir Tomov[b],
Ignacio Tomas[c,d], John N. Shadid[c]

[a] *Institute of Applied Mathematics (LS III), TU Dortmund University,*
Vogelpothsweg 87, D-44227 Dortmund, Germany
[b] *Center for Applied Scientific Computing, Lawrence Livermore National Laboratory,*
P.O. Box 808, L-561, Livermore, CA 94551, USA
[c] *Computational Mathematics Department, Sandia National Laboratories*
P.O. Box 5800 MS 1321, Albuquerque, NM 87185-1321, USA
[d] *Department of Mathematics and Statistics, University of New Mexico*
MSC01 1115, Albuquerque, NM 87131, USA

Abstract

This paper is focused on the aspects of limiting in residual distribution (RD) schemes for high-order finite element approximations to advection problems. Both continuous and discontinuous Galerkin methods are considered in this work. Discrete maximum principles are enforced using algebraic manipulations of element contributions to the global nonlinear system. The required modifications can be carried out without calculating the element matrices and assembling their global counterparts. The components of element vectors associated with the standard Galerkin discretization are manipulated directly using localized subcell weights to achieve optimal accuracy. Low-order nonlinear RD schemes of this kind were originally developed to calculate local extremum diminishing predictors for flux-corrected transport (FCT) algorithms. In the present paper, we incorporate limiters directly into the residual distribution procedure, which makes it applicable to stationary problems and leads to well-posed nonlinear discrete problems. To circumvent the second-order accuracy barrier, the correction factors of monolithic limiting

Email addresses: hennes.hajduk@math.tu-dortmund.de (Hennes Hajduk), kuzmin@math.uni-dortmund.de (Dmitri Kuzmin), tzanio@llnl.gov (Tzanio Kolev), tomov2@llnl.gov (Vladimir Tomov), itomas@sandia.gov (Ignacio Tomas), jnshadi@sandia.gov (John N. Shadid)

approaches and FCT schemes are adjusted using smoothness sensors based on second derivatives. The convergence behavior of presented methods is illustrated by numerical studies for two-dimensional test problems.

Keywords: advection problems, high-order finite elements, Bernstein polynomials, matrix-free methods, discrete maximum principles, residual distribution, limiters

1. Introduction

Many applications of practical interest require numerical solution of advection problems on unstructured meshes. In the context of finite element discretizations, robust, accurate, and efficient schemes are far more difficult to construct than for elliptic problems with smooth exact solutions. The presence of discontinuities and steep gradients requires a modification of the standard Galerkin approximation to rule out formation of spurious oscillations and violations of maximum principles. The design of bound-preserving high-resolution finite element schemes typically involves construction and limiting of artificial diffusion operators. Fully algebraic correction techniques which guarantee the validity of discrete maximum principles (DMP) can be found, e.g., in [5, 6, 14, 19, 23]. Most of them are designed for low-order (linear or multilinear) Lagrange elements, and the (element or global) matrices of the target scheme must be provided as input for the computation of artificial diffusion coefficients. The first successful extensions of matrix-based *flux-corrected transport* (FCT) algorithms to high-order finite elements were recently developed in [4, 24] using the Bernstein basis representation. Matrix-free approaches belonging to the family of *residual distribution* (RD) schemes [1] were extended to high-order Bernstein finite elements in [2, 3, 16].

The RD framework proposed in [16] introduces a set of new tools for the design of nonlinear low-order schemes and FCT-like methods. In the process of residual distribution, a given element vector is modified in a manner which preserves the sums of positive and negative components, while providing the *local extremum diminishing* (LED) property as well as conservation of mass. If no data from neighbor elements and previous time steps or Runge-Kutta stages is used, the resulting compact-stencil RD schemes are at most first-order accurate. The optimal ones exhibit p-independent convergence behavior as the polynomial degree p is increased while keeping the total number of degrees of freedom fixed. The derivation of such low-order schemes involves localization of the RD procedure to subcells [16]. To recover the high accuracy of the underlying Galerkin approximation at least in smooth regions, the FCT postprocessing step performs a LED correction of the provisional low-order solution. Predictor-corrector algorithms of this kind are well suited for time-dependent advection problems and conservative projections of data [19, 23]. However, the use of large (pseudo)-time steps leads to increased levels of numerical diffusion, and convergence to steady state solutions is inhibited by the alternating use of diffusive and antidiffusive corrections.

In contrast to FCT, monolithic limiting techniques adjust the artificial diffusion terms **before** adding them to the residual or matrix of the high-order target scheme. In the context of low-order continuous finite element approximations, *algebraic flux correction* (AFC) schemes of this kind were developed and analyzed in [7, 8, 17, 18, 19, 23]. Many of them are formally applicable to high-order Bernstein finite elements but require the knowledge of matrix entries and/or cannot match the accuracy of a piecewise (multi-) linear Lagrange approximation with the same number of nodal points.

In the present paper, we make the first step towards the derivation of matrix-free monolithic correction procedures that have the potential of becoming **more** accurate as the degree p of Bernstein basis functions is increased while keeping the number of nodes approximately fixed. We begin with the presentation of limiting and mass correction operators that improve the process of residual distribution in an appropriate manner. We also generalize the FCT algorithms employed in [4, 16, 24] and compare the results obtained with different limiting approaches. Finally, we equip selected limiters with smoothness indicators which make it possible to avoid unnecessary modifications of the Galerkin discretization without losing any favorable properties (preservation of local bounds in the neighborhood of discontinuities and steep fronts, continuous dependence on data).

2. Galerkin finite element discretization

First of all, let us summarize the notation introduced in [16] for finite element discretizations of the linear advection equation

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

where $\mathbf{v} = \mathbf{v}(\mathbf{x})$ is a continuous velocity field and $\Omega \subset \mathbb{R}^d,\ d \in \{1, 2, 3\}$ is a bounded domain. Let $\mathcal{T}_h$ be a computational mesh composed of $E_h = |\mathcal{T}_h|$

elements. Inside each element $K^e \in \mathcal{T}_h$, the solution u is approximated by a polynomial $u_h^e = \sum_{j=1}^N u_j^e \varphi_j^e$ of degree $p \geq 1$. The Bernstein basis functions φ_j^e are nonnegative and form a partition of unity. It follows that [16]

$$u_{\min}^e := \min_{1\leq j\leq N} u_j^e \leq u_h^e(\mathbf{x}) \leq \max_{1\leq j\leq N} u_j^e =: u_{\max}^e \qquad \forall \mathbf{x} \in K^e. \tag{2}$$

The basis function $\varphi_j^e : K^e \to [0,1]$ attains its maximum at the j-th nodal point $\mathbf{x}_j^e$ of element K^e. The global number $i_j^e = \mathcal{I}^e(j)$ of the local degree of freedom u_j^e can be retrieved using the DOF mapping $\mathcal{I}^e : \{1,\dots,N\} \to \{i_1^e,\dots,i_N^e\} =: \mathcal{N}^e$ to perform the index conversion $j \mapsto i_j^e$. The set of elements containing a node with the global number $i \in \{1,\dots,N_h\}$ is denoted by $\mathcal{E}_i$. The global index notation is used, e.g., for the piecewise-polynomial global basis functions $\varphi_i,\ i = 1,\dots,N_h$, which satisfy $\varphi_i|_{K^e} = \varphi_j^e$ with $i = \mathcal{I}^e(j)$. If global node numbers are used in the same equation as φ_i^e or u_h^e, the required index conversion is performed automatically. For example, we use the shorthand notation φ_j for $\varphi_{\mathcal{I}^e(j)}$ in such formulas [16].

The continuous or discontinuous Galerkin discretization of (1) yields a linear system of semi-discrete equations which can be written as

$$m_i \frac{\mathrm{d}u_i}{\mathrm{d}t} = \dot{\rho}_i^H + \rho_i^H + \sigma_i^H, \qquad i = 1,\dots,N_h, \tag{3}$$

where $m_i = \int_\Omega \varphi_i \mathrm{d}\mathbf{x}$ is a positive diagonal entry of the lumped mass matrix. The contribution of element $K^e \in \mathcal{T}_h$ to the right-hand side is given by

$$\dot{\rho}^{e,H} = (M_L^e - M_C^e)\dot{u}^{e,H}, \quad \rho^{e,H} = C^e u^e, \quad \sigma^{e,H} = S^e(\hat{u}^e - u^e), \tag{4}$$

where M_C^e is the consistent element mass matrix and M_L^e is its lumped counterpart. The element matrices C^e and S^e are composed from volume and surface integrals associated with the discretization of the advective term. The entries of all element matrices are defined in [16]. The element vector $\hat{u}^e$ is used to express $\sigma^{e,H}$ as a matrix-vector product. If $\mathbf{x}_i^e \in K^e$ lies on the inflow boundary Γ_- of Ω, then $\hat{u}_i^e$ is a Bernstein coefficient of the Dirichlet boundary data. If $\mathbf{x}_i^e \notin \Gamma_-$ lies on the boundary of K^e, then $\hat{u}_i^e$ is the Bernstein coefficient of the upwind-sided trace. For all nodes $\mathbf{x}_i \in \Omega$ at which u_h is continuous, we have $\hat{u}_i^e = u_i^e$. The element vector $\dot{u}^{e,H}$ is a sub-vector of $\dot{u}^H = M_C^{-1}(Cu + S(\hat{u} - u))$, where matrices without superscripts are global matrices. In a practical implementation, the element contributions $\dot{\rho}^{e,H}$, $\rho^{e,H}$, and $\sigma^{e,H}$ can be calculated in a matrix-free manner, see [16] for details.

3. Residual distribution framework

Let us now present a general framework for converting a semi-discrete problem like (3) into a bound-preserving discretization of the form

$$m_i \frac{\mathrm{d}u_i}{\mathrm{d}t} = f_i(u), \qquad i = 1, \dots, N_h, \tag{5}$$

where the right-hand side $f_i = \sum_{e \in \mathcal{E}_i} f_i^e$ is assembled from element contributions $f^e = \dot{\rho}^e + \rho^e + \sigma^e$. If integration in time is performed using a strong stability preserving (SSP) Runge-Kutta method, the fully discrete scheme fulfills the LED property if the time step Δt satisfies

$$u_{i,\min} \leq u_i + \frac{\Delta t}{m_i} f_i(u) \leq u_{i,\max} \qquad \forall i = 1, \dots, N_h, \tag{6}$$

where $u_{i,\min}$ and $u_{i,\max}$ are defined in terms of $u^e_{\min}$ and $u^e_{\max}$ as follows:

$$u_{i,\min} = \min_{e \in \bar{\mathcal{E}}_i} u^e_{\min}, \qquad u_{i,\max} = \max_{e \in \bar{\mathcal{E}}_i} u^e_{\max}. \tag{7}$$

The set $\bar{\mathcal{E}}_i$ contains the numbers of all elements to which the point $\mathbf{x}_i$ belongs.

The low-order LED schemes presented in [16] use $\dot{\rho}^e = 0$ and $\sigma^e = \tilde{S}^e(\hat{u}^e - u^e)$, where $\tilde{S}^e$ is a lumped diagonal approximation to S^e. The element contribution $\rho^e = \mathcal{R}(C^e u^e, u)$ is constructed using a correction operator $\mathcal{R}$ which transforms $\rho^{e,H}$ into $\rho^{e,L} = \mathcal{R}(\rho^{e,H}, u)$ subject to the mass conservation constraint $\sum_{i=1}^N \rho_i^{e,L} = \sum_{i=1}^N \rho_i^{e,H}$ and the following LED criterion.

Theorem 1 (Localized LED criterion). *Suppose that the coefficients*

$$\kappa_i^e = \begin{cases} \frac{f_i^e}{u_{i,\max} - u_i^e} & \textit{if } f_i^e > 0, \\ \frac{f_i^e}{u_{i,\min} - u_i^e} & \textit{if } f_i^e < 0, \\ 0 & \textit{if } f_i^e = 0 \end{cases} \tag{8}$$

are bounded for all $e \in \mathcal{E}_i$. Then the LED conditions (6) hold if

$$\Delta t \sum_{e \in \mathcal{E}_i} \kappa_i^e \leq m_i \qquad \forall i = 1, \dots, N_h. \tag{9}$$

Proof. See Theorems 1 and 2 in [16].

Example 1. The simplest example of a residual distribution procedure that produces $\rho^{e,L} = \mathcal{R}(\rho^{e,H}, u)$ with the above properties is [16]

$$\rho_i^{e,L} = \frac{\rho_{*,+}^e(u_{\max}^e - u_i^e)}{\sum_{j=1}^N(u_{\max}^e - u_j^e) + \varepsilon} + \frac{\rho_{*,-}^e(u_{\min}^e - u_i^e)}{\sum_{j=1}^N(u_{\min}^e - u_j^e) - \varepsilon}, \tag{10}$$

where

$$\rho_{*,+}^e = \sum_{i=1}^N \max\{0, \rho_i^{e,H}\}, \qquad \rho_{*,-}^e = \sum_{i=1}^N \min\{0, \rho_i^{e,H}\}. \tag{11}$$

The infinitesimally small positive number ε is used in formulas like (10) to avoid indeterminacy in the case of a vanishing denominator. In a practical implementation, the use of if-statements is preferable since the mass conservation constraint is satisfied exactly only in the limit $\varepsilon \searrow 0$.

As shown in [16], the simple definition of weights in (10) leads to a low-order LED approximation which becomes increasingly diffusive as the degree p of Bernstein basis functions is increased while keeping the total number of degrees of freedom N_h fixed. To achieve p-independent convergence behavior, a localized subcell residual distribution operator $\mathcal{R}$ was defined in [16].

Remark 1. If the coefficients (8) are bounded and the corresponding LED scheme converges to a steady state solution, this solution can be shown to satisfy generalized discrete maximum principles (see Lemma 4.16 in [23]).

4. Monolithic limiting strategy

To achieve higher than first-order accuracy, a bound-violating element contribution $\eta_i^{e,H} \in \{\dot{\rho}_i^{e,H}, \rho_i^{e,H}, \sigma_i^{e,H} - \sigma_i^{e,L}\}$ should be corrected using the nodal bounds (7) which depend on the data in neighbor elements if $\mathbf{x}_i^e \in \partial K^e$. Using limiting functions based on the theory of algebraic flux correction schemes [7, 8, 23], we will define nodal correction factors $\alpha_i^e \in [0,1]$ such that the limited element contributions $\bar{\eta}_i^{e,H} = \alpha_i^e \eta_i^{e,H}$ satisfy

$$|\bar{\eta}_i^{e,H}| \leq q_i^e \min\{u_{i,\max} - u_i^e, u_i^e - u_{i,\min}\} \tag{12}$$

for some bounded coefficients $q_i^e \geq 0$ and local extrema defined by (7). If $\bar{\eta}_i^{e,H} \geq 0$, then (12) implies that $\bar{\eta}_i^{e,H} \leq q_i^e(u_{i,\max} - u_i^e)$. If $\bar{\eta}_i^{e,H} \leq 0$, then $-\bar{\eta}_i^{e,H} \leq q_i^e(u_i^e - u_{i,\min})$, whence $\bar{\eta}_i^{e,H} \geq q_i^e(u_{i,\min} - u_i^e)$. In either case, the contribution of $\bar{\eta}_i^{e,H}$ to κ_i^e defined by (8) cannot exceed q_i^e.

By Theorem 1, condition (12) implies the LED property of $\bar{\eta}_i^{e,H}$. However, the mass conservation constraint $\sum_{i=1}^N \eta_i^e = \sum_{i=1}^N \eta_i^{e,H}$ does not generally hold for $\eta^e = \bar{\eta}^{e,H}$. Hence, an additional RD step is required to adjust the LED part $\bar{\eta}^{e,H}$ or the bound-violating remainder $\eta^{e,A} = \eta^{e,H} - \bar{\eta}^{e,H}$.

The adjustment of $\eta^{e,A}$ can be performed, e.g., using one of the residual distribution procedures developed in [16] to convert $\rho^{e,H}$ into $\rho^{e,L}$. Replacing $\eta^{e,A}$ with $\bar{\eta}^{e,A} = \mathcal{R}(\eta^{e,A}, u)$, where $\mathcal{R}$ is a suitably chosen RD operator, we will construct an element vector $\eta^e = \bar{\eta}^{e,H} + \bar{\eta}^{e,A}$ satisfying the conditions of Theorem 1 as well as the mass conservation constraint.

If the components of $\eta^{e,H}$ sum to zero, then the zero sum property of $\eta^e = \mathcal{R}(\bar{\eta}^{e,H}, u)$ can be readily enforced using an RD operator $\mathcal{R}$ which performs additional limiting of positive or negative components. The LED part $\bar{\eta}^{e,H}$ of $\eta^{e,H}$ violates the zero sum condition if $\bar{\eta}^e_{*,+} \neq -\bar{\eta}^e_{*,-}$, where

$$\bar{\eta}^e_{*,+} = \sum_{j=1}^N \max\left\{0, \bar{\eta}_j^{e,H}\right\}, \qquad \bar{\eta}^e_{*,-} = \sum_{j=1}^N \min\left\{0, \bar{\eta}_j^{e,H}\right\}. \tag{13}$$

The two-step FCT algorithms employed in [16, 24] correct the mass conservation error (if any) using the following kind of residual distribution:

$$\eta_i^e = \frac{\bar{\eta}_*^e}{\sum_{j=1}^N \Phi^e_{j,+} + \varepsilon}\Phi^e_{i,+} - \frac{\bar{\eta}_*^e}{\sum_{j=1}^N \Phi^e_{j,-} - \varepsilon}\Phi^e_{i,-}, \tag{14}$$

where $\bar{\eta}_*^e = \min\{\bar{\eta}^e_{*,+}, -\bar{\eta}^e_{*,-}\}$ and the RD weights $\Phi^e_{i,\pm}$ are defined by

$$\Phi^e_{i,+} = \max\{0, \bar{\eta}_i^{e,H}\}, \qquad \Phi^e_{i,-} = \min\{0, \bar{\eta}_i^{e,H}\}. \tag{15}$$

Note that the positive components of $\bar{\eta}_i^e$ remain unchanged if there is a loss of mass ($\bar{\eta}_*^e = \bar{\eta}^e_{*,+} = \sum_{j=1}^N \Phi^e_{j,+}$). Their negative counterparts remain unchanged if there is a surplus of mass ($\bar{\eta}_*^e = -\bar{\eta}^e_{*,-} = \sum_{j=1}^N \Phi^e_{j,-}$).

A more sophisticated nonlinear correction procedure was proposed in [4] to penalize changes of $\bar{\eta}_i^{e,H}$ at nodes where $\bar{\eta}_i^{e,H} = \eta_i^{e,H}$.

In the remainder of this section, we discuss the definition of nodal correction factors and the choice of limiting strategy for different components of the element contribution $f^e = \dot{\rho}^e + \rho^e + \sigma^e$ to the right-hand side of (5).

4.1. Limiting for volume integrals

The advective Galerkin element contribution $\rho_i^{e,H}$ is defined by [16]

$$\rho_i^{e,H} = -\int_{K^e} \varphi_i^e \mathbf{v} \cdot \nabla u_h^e \mathrm{d}\mathbf{x}, \qquad i = 1, \dots, N. \tag{16}$$

Using the diagonal matrix A^e of nodal correction factors (to be defined shortly), we decompose $\rho^{e,H}$ into the nonconservative LED part $\bar{\rho}^{e,H} = A^e\rho^{e,H}$ and the remainder $\rho^{e,A} = \rho^{e,H} - \bar{\rho}^{e,H}$ which requires residual distribution. The resulting element vector can be formally written as

$$\rho^e = \bar{\rho}^{e,H} + \mathcal{R}(\rho^{e,A}, u), \tag{17}$$

where $\mathcal{R}$ is a residual distribution operator. In this work, we use the subcell RD procedure which was found to produce the best results for $\rho^{e,A} = \rho^{e,H}$ in [16]. As we will see later, the idea of leaving the LED-part $\bar{\rho}^{e,H}$ of the Galerkin residual unchanged and applying $\mathcal{R}$ to the (possibly vanishing) non-LED-part $\rho^{e,H} - \bar{\rho}^{e,H}$ results in further marked improvements.

When it comes to calculating the nodal correction factors $\alpha_i^e \in [0,1]$ for $\bar{\rho}_i^{e,H} = \alpha_i^e \rho_i^{e,H}$, we use the following definition (cf. [18, 23])

$$\alpha_i^e = \min\left\{1, \beta_i^e \frac{\min\{u_{i,\max} - u_i^e, u_i^e - u_{i,\min}\}}{\max\{u_{i,\max} - u_i^e, u_i^e - u_{i,\min}\} + \varepsilon}\right\}. \tag{18}$$

In accordance with the LED design criterion, this formula produces $\alpha_i^e = 0$ if $u_i^e = u_{i,\max}$ or $u_i^e = u_{i,\min}$. The free parameter $\beta_i^e > 0$ determines the amount of limiting away from local extrema. It can be chosen sufficiently large for (18) to produce $\alpha_i^e = 1$ at least for locally linear functions u_h. This property is known as *linearity preservation*. As shown in [18], it is guaranteed for

$$\beta_i^e \geq \frac{\max_{e\in\mathcal{E}_i} h^e}{\max_{e\in\mathcal{E}_i} r^e}, \tag{19}$$

where h^e is the diameter of K^e and r^e is the diameter of the largest inscribed ball. In general, larger values of β_i^e result in more accurate approximations but imply larger values of the LED coefficients q_i^e which make condition (9) more restrictive and may require the use of inordinately small time steps (or cause convergence problems in the steady-state limit). In all simulations employing the monolithic limiter we set $\beta_i^e = 10$.

Theorem 2 (LED property of $\boldsymbol{\rho_i^e}$). *Let α_i^e defined by (18) be the nodal correction factor for $\eta^{e,H} = \rho_i^{e,H}$. Then condition (12) holds for*

$$q_i^e = \beta_i^e \sum_{\substack{j=1\\ j\neq i}}^{N} \left| \int_{K^e} \varphi_i^e \mathbf{v} \cdot \nabla \varphi_j^e \mathrm{d}\mathbf{x} \right|. \tag{20}$$

Proof. By definition of the Galerkin element contribution $\rho_i^{e,H}$ to the discretized advective term, we have the estimate

$$|\rho_i^{e,H}| = \left|\sum_{\substack{j=1\\j\neq i}}^{N}(u_j^e - u_i^e)\int_{K^e}\varphi_i^e\mathbf{v}\cdot\nabla\varphi_j^e\mathrm{d}\mathbf{x}\right| \leq \sum_{\substack{j=1\\j\neq i}}^{N}|u_j^e - u_i^e|\left|\int_{K^e}\varphi_i^e\mathbf{v}\cdot\nabla\varphi_j^e\mathrm{d}\mathbf{x}\right|,$$

where $|u_j^e - u_i^e| \leq \max\{u_{i,\max} - u_i^e, u_i^e - u_{i,\min}\}$ because

$$\begin{aligned} u_j^e \geq u_i^e \qquad &\Rightarrow \qquad |u_j^e - u_i^e| = u_j^e - u_i^e \leq u_{\max}^e - u_i^e, \\ u_j^e \leq u_i^e \qquad &\Rightarrow \qquad |u_j^e - u_i^e| = u_i^e - u_j^e \leq u_i^e - u_{\min}^e. \end{aligned}$$

Invoking (18), we conclude that

$$\alpha_i^e|\rho_i^{e,H}| \leq \beta_i^e \min\{u_{i,\max} - u_i, u_i^e - u_{i,\min}\}\sum_{\substack{j=1\\j\neq i}}^{N}\left|\int_{K^e}\varphi_i^e\mathbf{v}\cdot\nabla\varphi_j^e\mathrm{d}\mathbf{x}\right|,$$

which proves the statement of the Theorem. □

Remark 2. Similarly to the FCT algorithms employed in [4, 16, 24], the pointwise limiting of $\rho^{e,H}$ is followed by a bound-preserving correction of the possibly nonconservative LED part $\bar{\rho}^{e,H} = A^e\rho^{e,H}$. The difference lies in the fact that the components of $\rho^{e,H}$ may not sum to zero and mass conservation is enforced using residual distribution rather than additional limiting.

4.2. Limiting for boundary terms

In the DG version, the Galerkin discretization of the advective term produces additional boundary terms which are stored in the element contribution $\sigma^{e,H} = S^e(\hat{u}^e - u^e)$. The symmetric element matrix S^e is defined by [16]

$$s_{ij}^e = -\int_{\partial K^e}\varphi_i^e\varphi_j^e\min\{0, \mathbf{v}\cdot\mathbf{n}^e\}\mathrm{ds}, \qquad i,j = 1,\ldots,N, \tag{21}$$

where $\mathbf{n}^e$ is the unit outward normal to the boundary of K^e. The diagonal element matrix $\tilde{S}^e$ of the low-order LED approximation $\sigma^{e,L} = \tilde{S}^e(\hat{u}^e - u^e)$ is constructed from S^e using row-sum lumping, i.e., [16]

$$\tilde{s}_{ij}^e = \delta_{ij}\sum_{k=1}^{N}s_{ik}^e, \qquad i,j = 1,\ldots,N. \tag{22}$$

The components of the element vector $\eta^{e,H}=\sigma^{e,H}-\sigma^{e,L}$ sum to zero. Therefore, we can constrain $\eta^{e,H}$ using the zero sum preserving limiting strategy described at the beginning of this section.

Theorem 3 (LED property of $\boldsymbol{\sigma_i^e}$). *Let α_i^e defined by (18) be the nodal correction factor for $\eta^{e,H}=\sigma^{e,H}-\sigma^{e,L}$. Then condition (12) holds for*

$$q_i^e = 3\beta_i^e\sum_{\substack{j=1\\j\neq i}}^N s_{ij}^e. \tag{23}$$

Proof. Recall that $\sum_{j=1}^N s_{ij}^e=\tilde{s}_{ii}^e$ and $\tilde{s}_{ij}^e=0$ for $j\neq i$. We have

$$\begin{aligned}\left|\sum_{j=1}^N(s_{ij}^e-\tilde{s}_{ij}^e)(\hat{u}_j^e-u_j^e)\right| &= \left|\sum_{\substack{j=1\\j\neq i}}^N s_{ij}^e(\hat{u}_j^e-\hat{u}_i^e-u_j^e+u_i^e)\right|\\ &\le \sum_{\substack{j=1\\j\neq i}}^N s_{ij}^e(|\hat{u}_j^e-u_i^e|+|\hat{u}_i^e-u_i^e|+|u_j^e-u_i^e|)\\ &\le 3\sum_{\substack{j=1\\j\neq i}}^N s_{ij}^e\max\{u_{i,\max}-u_i^e,u_i^e-u_{i,\min}\},\end{aligned}$$

where we have used the triangle inequality and the sparsity pattern of S^e. Invoking definition (18), we obtain the desired result as in the proof of Theorem 2. □

Remark 3. In our experience, the effect of correcting $\sigma^{e,L}$ is marginal compared to improvements that are achieved by correcting $\rho^{e,L}$. Therefore, the correction of $\sigma^{e,L}$ may sometimes be skipped to speed up calculations.

4.3. Limiting for mass matrices

The Galerkin element contribution $\dot{\rho}^{e,H}$ to the right-hand side of the semi-discrete high-order system (3) is defined by $\dot{\rho}^{e,H}=(M_L^e-M_C^e)\dot{u}^{e,H}$. A matrix-free approach to calculating $\dot{u}^{e,H}$ and $\dot{\rho}^{e,H}$ is described in [16]. In the context of monolithic limiting, the vector of time derivatives $\dot{u}^e$ is defined by

$$m_i\dot{u}_i=\sum_{e\in\mathcal{E}_i}(\dot{\eta}_i^e+\rho_i^e+\sigma_i^e), \tag{24}$$

where ρ_i^e and σ_i^e are the LED element contributions of the RD scheme and $\dot\eta_i^e$ is a limited counterpart of the correction term

$$\dot\eta_i^{e,H} = \dot\rho_i^e + \theta_i^e(\rho_i^{e,A} + \sigma_i^{e,A}), \tag{25}$$

depending on a parameter $\theta_i^e \in [0,1]$ and the element contributions

$$\dot\rho^e = (M_L^e - M_C^e)\dot u^e, \qquad \rho^{e,A} = \rho^{e,H} - \rho^e, \qquad \sigma^{e,A} = \sigma^{e,H} - \sigma^e. \tag{26}$$

Note that the original Galerkin discretization (3) is recovered for $\dot\rho_i^e = \dot\rho_i^{e,H}$ and $\theta_i^e = 1$. The corresponding correction term $\dot\eta_i^{e,H} = \dot\rho_i^{e,H} + \rho_i^{e,A} + \sigma_i^{e,A}$ can be limited using the FCT algorithm presented in Section 5. However, it is not well suited for monolithic limiting because correction factors that provide the LED property while producing the same steady-state solutions as the lumped-mass RD scheme are difficult to define. The use of (25) with $\theta_i^e = 0$ makes it easy to find such correction factors but tends to produce large phase errors during the transient phase because the sign of $\dot\eta_i^{e,H} = \dot\rho_i^e$ may differ from that of the limiting target $\dot\eta_i^{e,T} = \dot\rho_i^e + \rho_i^{e,A} + \sigma_i^{e,A}$, the addition of which to $\rho_i^e + \sigma_i^e$ would produce $\dot\rho_i^e + \rho_i^{e,H} + \sigma_i^{e,H}$. To make sure that $\dot\eta_i^{e,H}$ defined by (25) has the same sign as $\dot\eta_i^{e,T}$ or equals zero, we use

$$\theta_i^e = \min\left\{1, \frac{|\dot\rho_i^e|}{|\rho_i^{e,A} + \sigma_i^{e,A}| + \varepsilon}\right\}. \tag{27}$$

Obviously, this definition produces $\dot\eta_i^{e,H} = \dot\rho_i^e = 0$ in the steady state limit. If the steady state solution is of primary interest or the problem at hand is weakly time dependent, the right-hand side of the RD scheme (24) can be assembled using $\dot\eta^e = 0$. This approximation corresponds to full mass lumping which may significantly degrade the overall accuracy of high-order finite element discretizations for strongly time dependent problems.

Since the components of $\dot\eta^{e,H}$ sum to zero, the limiting strategy based on (14) and (15) is applicable. In view of (27), the choice of the corresponding correction factors $\dot\alpha_i^e$ should guarantee the validity of (12) for $\bar\eta_i^{e,H} = \dot\alpha_i^e\dot\rho_i^e$. Time derivative limiters based on this design criterion were recently proposed in [18, 23]. Adapting them to our RD setting, we define $\dot\alpha_i^e$ as follows:

$$\dot\alpha_i^e = \min\left\{1, \dot\beta_i^e \frac{\min\{u_{i,\max} - u_i^e, u_i^e - u_{i,\min}\}}{\max\{\dot u_{\max}^e - \dot u_i^e, \dot u_i^e - \dot u_{\min}^e\} + \varepsilon}\right\}. \tag{28}$$

The parameter $\dot\beta_i^e$ should have units of the reciprocal second. Larger values of $\dot\beta_i^e$ reduce the amount of limiting but the LED property is guaranteed for smaller time steps. In the numerical examples below, we use (cf. [18, 23])

$$\dot\beta_i^e = \beta_i^e \frac{\max_{1\le i\le N} |\mathbf{v}(\mathbf{x}_i^e)|}{2h^e/p}, \tag{29}$$

where β_i^e is the parameter used in (18) and h^e is the diameter of K^e, and $p \in \mathbb{N}$ is the polynomial degree of Bernstein basis functions.

Theorem 4 (LED property of $\dot{\boldsymbol{\rho}}_{\boldsymbol{i}}^{\boldsymbol{e}}$). *Let $\dot\alpha_i^e$ defined by (28) be the nodal correction factor for $\eta^{e,H} = \dot\rho^e$. Then condition (12) holds for*

$$q_i^e = \dot\beta_i^e m_i^e. \tag{30}$$

Proof. We have $\dot\rho_i^e = \int_{K^e} \varphi_i(\dot u_i^e - \dot u_h^e)\mathrm{d}\mathbf{x}$. Since $\dot u_h^e$ is bounded by its Bernstein coefficients and $\int_{K^e} \varphi_i^e \mathrm{d}\mathbf{x} = m_i^e$, we have

$$\dot\alpha_i^e |\dot\rho_i^e| \le \dot\alpha_i^e m_i^e \max\{\dot u_{\max}^e - \dot u_i^e, \dot u_i^e - \dot u_{\min}^e\}.$$

The statement of the Theorem follows by definition of $\dot\alpha_i^e$. □

Due to the dependence of the element contributions $\dot\rho^e$ and correction factors $\dot\alpha_i^e$ on $\dot u^e$, system (24) is solved using the fixed-point iteration

$$m_i \dot u_i^{(m+1)} = \sum_{e\in\mathcal{E}_i} (\dot\eta_i^e(\dot u^{(m)}) + \rho_i^e + \sigma_i^e) \tag{31}$$

which can be initialized using $\dot u^{(0)} = 0$. In the DG version, $\mathcal{E}_i = \{e\}$ consists of a single element and (31) simplifies to $m_i^e \dot u_i^{(m+1)} = \dot\eta_i^e(\dot u^{(m)}) + \rho_i^e + \sigma_i^e$. Upon convergence, the corrected element contributions $f_i^e = \dot\eta_i^e(\dot u) + \rho_i^e + \sigma_i^e$ are inserted into the right-hand side of the monolithic LED scheme (5).

5. Flux-corrected transport

In contrast to the above limiting procedures, which were largely inspired by recent advances in the development of monolithic algebraic flux correction schemes [7, 18, 23], classical FCT algorithms update the Bernstein coefficients of the finite element solution u_h using a low-order LED approximation

$$u_i^L = u_i + \frac{\Delta t}{m_i} \sum_{e\in\mathcal{E}_i} f_i^{e,L} \tag{32}$$

and perform LED antidiffusive corrections in the following manner:

$$\bar{u}_i = u_i^L + \frac{\Delta t}{m_i} \sum_{e \in \mathcal{E}_i} \mathcal{R}(f_i^{e,A}, u^L). \tag{33}$$

The low-order element contributions $f_i^{e,L}$ and the antidiffusive correction terms $f_i^{e,A}$ for the RD-FCT schemes to be considered in this work are defined in [16]. The components of $f^{e,A}$ sum to zero. The limiting operator $\mathcal{R}$ should preserve this property while enforcing the local maximum principle

$$u_{i,\min} \leq \bar{u}_i \leq u_{i,\max}. \tag{34}$$

This task can again be accomplished using the zero sum preserving limiter presented in Section 4. The FCT correction factor α_i^e for the antidiffusive element contribution $f_i^{e,A}$ is formally defined by [16, 24]

$$\alpha_i^e = \begin{cases} \min\left\{1, \frac{m_i^e(u_{i,\max} - u_i^{e,L})}{\Delta t f_i^{e,A}}\right\} & \text{if } f_i^{e,A} > 0, \\ \min\left\{1, \frac{m_i^e(u_{i,\min} - u_i^{e,L})}{\Delta t f_i^{e,A}}\right\} & \text{if } f_i^{e,A} < 0, \\ 1 & \text{otherwise,} \end{cases} \tag{35}$$

where m_i^e is a diagonal entry of the lumped element mass matrix M_L^e. As shown in [16, 24], this definition guarantees the validity of (34).

Remark 4. In practical implementations of the FCT algorithm based on componentwise limiting of $f^{e,A}$ and mass correction [4, 16], the nonconservative LED approximation $\bar{\eta}_i^{e,H} = \alpha_i^e f_i^{e,A}$ is calculated using the formula

$$\bar{\eta}_i^{e,H} = \frac{m_i^e}{\Delta t}(\bar{u}_i^{e,H} - u_i^{e,L}), \tag{36}$$

where $\bar{u}_i^{e,H} = \min\{u_{i,\max}, \max\{u_i^{e,H}, u_{i.\min}\}\}$ is a limited approximation to the high-order value $u_i^{e,H} = u_i^{e,L} + \frac{\Delta t}{m_i^e} f_i^{e,A}$ of the Bernstein coefficient.

6. Smooth extrema preserving limiting

The LED criterion imposes a second-order barrier on the accuracy of numerical approximations in the neighborhood of local extrema [29]. To avoid unnecessary limiting, smoothness indicators and troubled cell detectors are commonly employed in the literature [9, 10, 12, 15, 20, 21, 27]. A suitably defined smoothness indicator γ_i^e can be used to adjust the range $[u_{i,\min}, u_{i,\max}]$ of admissible values or the correction factor α_i^e which provides the LED property w.r.t. this range. Ideally, the definition of α_i^e should guarantee

- optimal convergence behavior in smooth regions;
- preservation of local bounds in nonsmooth regions;
- preservation of global bounds in the whole domain;
- continuous dependence of $\bar{\eta}_i^e = \alpha_i^e \eta_i^{e,H}$ on the data.

Let $\tilde{V}_h^{\rm CG} = \mathrm{span}\{\tilde{\varphi}_1^{\rm CG}, \ldots, \tilde{\varphi}_{N_h^{\rm CG}}^{\rm CG}\}$ and $\tilde{V}_h^{\rm DG} = \mathrm{span}\{\tilde{\varphi}_1^{\rm DG}, \ldots, \tilde{\varphi}_{N_h^{\rm DG}}^{\rm DG}\}$ denote the finite element spaces corresponding to the continuous and discontinuous piecewise-linear ($\mathbb{P}_1$) or multilinear ($\mathbb{Q}_1$) approximations on subcells of $\mathcal{T}_h$. Given the Bernstein coefficients $u_i,\ i = 1, \ldots, N_h$ of a high-order finite element approximation u_h, we will define the nodal smoothness indicators γ_i^e using reconstructed second derivatives of the piecewise-linear interpolant $\tilde{u}_h \in \tilde{V}_h$, where $\tilde{V}_h = \tilde{V}_h^{\rm CG}$ if $N_h = N_h^{\rm CG}$ and $\tilde{V}_h = \tilde{V}_h^{\rm DG}$ if $N_h = N_h^{\rm DG}$. The nodal values $\tilde{u}_h^e(\mathbf{x}_i^e) = \tilde{u}_i^e = u_h^e(\mathbf{x}_i^e)$ are determined using pointwise interpolation. Note that $\tilde{u}_i^e$ may not coincide with the value of the Bernstein coefficient u_i^e if $\mathbf{x}_i^e$ is not a vertex of the macroelement K^e.

Using a suitably defined discrete Laplacian operator $\Delta_h : \tilde{V}_h \to \tilde{V}_h^{CG}$ to calculate $g_i = (\Delta_h \tilde{u}_h)_i$, we define the nodal smoothness indicators

$$\gamma_i^e = \begin{cases} 1 - \left(\frac{|g_{i,\min} - g_{i,\max}|}{|g_{i,\min}| + |g_{i,\max}| + \varepsilon}\right)^q & \text{if } \mathbf{x}_i^e \in \Omega, \\ 1 & \text{if } \mathbf{x}_i^e \in \Gamma \end{cases} \tag{37}$$

and

$$\gamma_i^e = \begin{cases} \min\left\{1, \frac{C \max\{0, g_{i,\min} g_{i,\max}\}}{\max\{g_{i,\min}^2, g_{i,\max}^2\} + \varepsilon}\right\} & \text{if } \mathbf{x}_i^e \in \Omega, \\ 1 & \text{if } \mathbf{x}_i^e \in \Gamma, \end{cases} \tag{38}$$

where $q \geq 1$ and $C \geq 1$ are sensitivity parameters. In the numerical studies below, we use $q = 5$ and $C = 3$. The minimum and maximum

$$g_{i,\min} = \min_{j \in \tilde{\mathcal{N}}_i^{\rm CG}} g_j, \qquad g_{i,\max} = \max_{j \in \tilde{\mathcal{N}}_i^{\rm CG}} g_j \tag{39}$$

are taken over the set $\tilde{\mathcal{N}}_i^{\rm CG}$ of nodes that share a subcell with node i. This set contains the index i and the numbers of nearest neighbors.

For any $q \in \mathbb{N}$, formula (37) produces $\gamma_i^e = 0$ if the signs of g_i and g_j differ for any $j \in \tilde{\mathcal{N}}_i^{\mathrm{CG}}$. If the signs of g_i and g_j are the same for all $j \in \tilde{\mathcal{N}}_i^{\mathrm{CG}}$, then the value of the smoothness sensor $\gamma_i^e \in [0,1]$ depends on the difference between $g_{i,\min}$ and $g_{i,\max}$, i.e., on the local variations of the reconstructed Laplacians. The value $\gamma_i^e = 1$ is obtained only if $g_j = g_i$ for all $j \in \tilde{\mathcal{N}}_i^{\mathrm{CG}}$. Formula (38) produces $\gamma_i^e = 0$ as well if the signs of $g_{i,\min}$ and $g_{i,\max}$ differ. The maximal value $\gamma_i^e = 1$ is attained if the signs of the two extrema are the same and their magnitudes do not differ by more than a factor of C.

The discrete Laplacian operator $\Delta_h : \tilde{V}_h \to \tilde{V}_h^{CG}$ that we use to calculate the input data g_j for the smoothness indicator is based on the direct variational recovery of nodal Laplacians via (cf. [25], Section 5.3.1)

$$\begin{aligned}\sum_{j=1}^{N_h^{\mathrm{CG}}} g_j \int_\Omega \tilde{\varphi}_i^{\mathrm{CG}} \tilde{\varphi}_j^{\mathrm{CG}} \mathrm{d}\mathbf{x} &= \int_{\partial\Omega} \tilde{\varphi}_i^{\mathrm{CG}} \mathbf{n} \cdot \nabla \tilde{u}_h^{\mathrm{CG}} \mathrm{ds} \\ &\quad - \int_\Omega \nabla \tilde{\varphi}_i^{\mathrm{CG}} \cdot \nabla \tilde{u}_h^{\mathrm{CG}} \mathrm{d}\mathbf{x}, \qquad i = 1, \dots, N_h^{\mathrm{CG}}, \qquad (40)\end{aligned}$$

where $\tilde{u}_h^{CG} \in \tilde{V}_h^{CG}$ is the consistent-mass L^2 projection of $\tilde{u}_h \in \tilde{V}_h^{DG}$ in the DG version and $\tilde{u}_h^{CG} = \tilde{u}_h$ in the CG version.

If row-sum mass lumping is performed, the formula for g_i simplifies to

$$g_i = \frac{\int_{\partial\Omega} \tilde{\varphi}_i^{\mathrm{CG}} \mathbf{n} \cdot \nabla \tilde{u}_h^{\mathrm{CG}} \mathrm{ds} - \int_\Omega \nabla \tilde{\varphi}_i^{\mathrm{CG}} \cdot \nabla \tilde{u}_h^{\mathrm{CG}} \mathrm{d}\mathbf{x}}{\int_\Omega \tilde{\varphi}_i^{\mathrm{CG}} \mathrm{d}\mathbf{x}}, \qquad i = 1, \dots, N_h^{\mathrm{CG}}. \qquad (41)$$

This definition produces a rather smooth approximation to the nodal Laplacians and, therefore, tends to overestimate the local regularity of $\tilde{u}_h$.

Remark 5. The reader may wonder why we reconstruct the second derivatives of $\tilde{u}_h$ instead of differentiating u_h to compute g_i. The reason for this lies in the fact that the second derivatives of u_h^e are continuous at all internal nodes of K^e. Therefore, formula (37) would produce $\gamma_i^e = 1$ at these nodes. Indeed, polynomials are smooth by definition, so smoothness indicators based on jumps of second (or higher-order) derivatives are useful only for nodes located at the corners of K^e. The idea of interpolating the corner values to define γ_i^e at the remaining nodes, as proposed in [20], produces reasonable results for $p = 2$ on relatively fine meshes. However, sharp peaks may remain undetected in the interior of large high-order elements.

In the context of RD-FCT schemes, we relax the local bounds using γ_i^e and the coefficient $u_i^{e,H}$ of the unlimited high-order solution as follows:

$$\begin{aligned} u_{i,\min}^{e,*} &= \gamma_i^e u_i^{e,H} + (1-\gamma_i^e) u_{i,\min}, \\ u_{i,\max}^{e,*} &= \gamma_i^e u_i^{e,H} + (1-\gamma_i^e) u_{i,\max}. \end{aligned} \tag{42}$$

For $\gamma_i^e = 1$, the so-defined relaxed bounds reduce to $u_{i,\min}^{e,*} = u_i^{e,H} = u_{i,\max}^{e,*}$ leading to $\bar{u}_i^{e,H} = u_i^{e,H}$, i.e. no limiting. This property makes it possible to avoid unnecessary limiting in smooth regions, while enforcing the LED bounds $u_{i,\min}^{e,*} = u_{i,\min}$ and $u_{i,\max}^{e,*} = u_{i,\max}$ at troubled nodes where $\gamma_i^e = 0$.

To rule out violations of a priori known global bounds $u_{\min}^{\text{glob}}$ and $u_{\max}^{\text{glob}}$, the relaxed bounds (42) are further confined to be in the interval $[u_{\min}^{\text{glob}}, u_{\max}^{\text{glob}}]$ via

$$\overline{u}_{i,\min}^e = \max\{u_{i,\min}^{e,*}, u_{\min}^{\text{glob}}\}, \qquad \overline{u}_{i,\max}^e = \min\{u_{i,\max}^{e,*}, u_{\max}^{\text{glob}}\}. \tag{43}$$

The bounds (43) are then utilized in the FCT step instead of $u_{i,\min}$, $u_{i,\max}$.

An extremum preserving version of the monolithic limiter based on (18) can be defined using the modified correction factors

$$\alpha_i^e = \begin{cases} \min\left\{1, \dfrac{\beta_i^e(u_i^e - u_{i,\min}^{e,*})}{u_{i,\max} - u_i^e + \varepsilon}\right\} & \text{if } u_{i,\max} - u_i^e > u_i^e - u_{i,\min} + \varepsilon, \\ \min\left\{1, \dfrac{\beta_i^e(u_{i,\max}^{e,*} - u_i^e)}{u_i^e - u_{i,\min} + \varepsilon}\right\} & \text{if } u_{i,\max} - u_i^e < u_i^e - u_{i,\min} - \varepsilon, \end{cases} \tag{44}$$

defined in terms of the relaxed local bounds

$$\begin{aligned} u_{i,\min}^{e,*} &= \gamma_i^e (2u_i^e - u_{i,\max}) + (1-\gamma_i^e) u_{i,\min}, \\ u_{i,\max}^{e,*} &= \gamma_i^e (2u_i - u_{i,\min}) + (1-\gamma_i^e) u_{i,\max}. \end{aligned} \tag{45}$$

Again, preservation of global bounds can be enforced by confining the relaxed bounds (45) within the global ones through (43). The time derivative limiters $\dot{\alpha}_i^e$ defined by (28) require similar modifications.

7. Numerical examples

In this section, we assess the accuracy of presented limiters and smoothness indicators in a series of numerical experiments for the DG version of the high-order Bernstein finite element approximation. As in the numerical study presented in [16], computations are performed with different values of

p for mesh sizes $h(p)$ corresponding to exactly the same number of unknowns N_h. The schemes under investigation are abbreviated by

RDS(p)	subcell RD scheme without limiting (as defined in [16]);
MON(p)	subcell RD with monolithic limiting (Section 4);
FCT(p)	subcell RD with FCT-type limiting (Section 5).

The extremum-preserving versions of MON and FCT will be identified using the suffix S_1 for γ_i^e defined by (37) and S_2 for γ_i^e defined by (38). The grid convergence studies for MON-S_i and FCT-S_i, $i \in \{1,2\}$ are restricted to stationary and time-dependent test problems, respectively.

The implementation of subcell RD procedures, limiting techniques and smoothness indicators is based on the open-source C++ finite element library MFEM [26]. Steady problems are solved using pseudo time stepping, whereas in the time dependent case we employ the third order SSP Runge Kutta method.

7.1. Unsteady advection in 1D

In the first series of numerical experiments, we perform grid convergence studies for the 1D smoothed step function

$$u_0(x) = \frac{1}{4}\left(1 + \tanh\left(\frac{x - 0.15}{0.03}\right)\right)\left(1 - \tanh\left(\frac{x - 0.35}{0.03}\right)\right)$$

which is advected in $\Omega = (0,1)$ using the constant velocity $v = 1$. The exact solution and boundary condition are given by $u(x,t) = u_0(x - vt)$ and $u_{\mathrm{in}}(t) = u(0,t)$, respectively. We solve this unsteady advection problem using the FCT version of the limited RD scheme. At the final time $T = 0.5$, we compute the errors $\|u_h(x,T) - u(x,T)\|_{L^1(\Omega)}$ and the corresponding experimental orders of convergence (EOC). The convergence behavior of the unconstrained DG scheme is summarized in Table 1. The L^1 error converges at the rate $p+1$ or faster for finite elements of degree p. The results presented in Table 2 illustrate the fact that even nonlinear LED schemes are at most second-order accurate. As stated earlier, the use of a smoothness indicator is a prerequisite for achieving higher than second order accuracy even for problems with smooth solutions. Tables 3 and 4 demonstrate the ability of FCT-S_1 and FCT-S_2 to deliver optimal convergence rates for advection problems with smooth exact solutions. These two extremum preservig FCT schemes produce virtually identical L^1 errors which differ from those presented in Table 1 only for polynomials of degree $p = 1,2$ on coarse meshes.

Hence, both versions of the Laplacian-based smoothness indicator identify the advected profile as a smooth function and prevent unnecessary limiting.

h	$p = 1$	EOC	$p = 2$	EOC	$p = 3$	EOC	$p = 4$	EOC
1/ 48	9.12E-03		7.46E-04		2.27E-05		1.45E-06	
1/ 64	5.06E-03	2.05	2.65E-04	3.60	4.66E-06	5.50	2.56E-07	6.02
1/ 96	1.97E-03	2.32	6.42E-05	3.50	6.83E-07	4.74	3.21E-08	5.12
1/128	9.56E-04	2.52	2.48E-05	3.30	2.08E-07	4.14	7.49E-09	5.05
1/192	3.34E-04	2.60	6.67E-06	3.24	4.06E-08	4.03		
1/256	1.61E-04	2.53	2.71E-06	3.13				
1/384	5.96E-05	2.45						

Table 1: $\|\cdot\|_{L^1(\Omega)}$ errors and corresponding EOC for standard DG using $p \in \{1, \ldots, 4\}$.

h	$p = 1$	EOC	$p = 2$	EOC	$p = 3$	EOC	$p = 4$	EOC
1/ 48	7.02E-03		1.06E-03		1.02E-04		1.43E-04	
1/ 64	4.84E-03	1.29	4.47E-04	3.01	4.44E-05	2.90	6.32E-05	2.84
1/ 96	1.99E-03	2.19	1.22E-04	3.20	1.55E-05	2.59	2.12E-05	2.69
1/128	9.66E-04	2.51	5.36E-05	2.87	7.78E-06	2.40	1.01E-05	2.58
1/192	3.39E-04	2.59	1.73E-05	2.78	2.99E-06	2.36		
1/256	1.64E-04	2.52	8.02E-06	2.68				
1/384	6.08E-05	2.45						

Table 2: $\|\cdot\|_{L^1(\Omega)}$ errors and corresponding EOC for FCT and $p \in \{1, \ldots, 4\}$.

h	$p = 1$	EOC	$p = 2$	EOC	$p = 3$	EOC	$p = 4$	EOC
1/ 48	7.29E-03		1.08E-03		2.27E-05		1.45E-06	
1/ 64	4.97E-03	1.33	3.64E-04	3.78	4.66E-06	5.50	2.56E-07	6.02
1/ 96	1.97E-03	2.28	6.42E-05	4.28	6.83E-07	4.74	3.21E-08	5.12
1/128	9.56E-04	2.52	2.48E-05	3.30	2.08E-07	4.14	7.49E-09	5.05
1/192	3.34E-04	2.60	6.67E-06	3.24	4.06E-08	4.03		
1/256	1.61E-04	2.53	2.71E-06	3.13				
1/384	5.96E-05	2.45						

Table 3: $\|\cdot\|_{L^1(\Omega)}$ errors and corresponding EOC for FCT-S$_1$ and $p \in \{1, \ldots, 4\}$.

7.2. Unsteady advection in 2D

A standard 2D test problem involving the solution of the unsteady linear advection equation (1) is the solid body rotation benchmark [22]. In this

h	$p=1$	EOC	$p=2$	EOC	$p=3$	EOC	$p=4$	EOC
1/ 48	7.29E-03		1.03E-03		2.27E-05		1.45E-06	
1/ 64	4.96E-03	1.34	3.98E-04	3.33	4.66E-06	5.50	2.56E-07	6.02
1/ 96	1.97E-03	2.27	6.42E-05	4.50	6.83E-07	4.74	3.21E-08	5.12
1/128	9.56E-04	2.52	2.48E-05	3.30	2.08E-07	4.14	7.49E-09	5.05
1/192	3.34E-04	2.60	6.67E-06	3.24	4.06E-08	4.03		
1/256	1.61E-04	2.53	2.71E-06	3.13				
1/384	5.96E-05	2.45						

Table 4: $\|\cdot\|_{L^1(\Omega)}$ errors and corresponding EOC for FCT-S$_2$ and $p \in \{1,\dots,4\}$.

example, the velocity field $\mathbf{v}(x,y) = (0.5-y, x-0.5)^T$ is used to rotate a slotted cylinder, a sharp cone, and a smooth hump around the center of $\Omega = (0,1)^2$. The initial condition, as shown in Figs. 1a–1c, is given by

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

where

$$u_0^{\text{hump}}(x,y) = \frac{1}{4} + \frac{1}{4}\cos\left(\frac{\pi\sqrt{(x-0.25)^2+(y-0.5)^2}}{0.15}\right), \tag{46}$$

$$u_0^{\text{cone}}(x,y) = 1 - \frac{\sqrt{(x-0.5)^2+(y-0.25)^2}}{0.15}. \tag{47}$$

Homogeneous Dirichlet boundary conditions are prescribed at the inlets.

After each full rotation, the exact solution $u(x,y,2\pi k),\ k \in \mathbb{N}$ coincides with $u_0(x,y)$. The challenge of this test is to preserve the shape of the projected initial condition $u_h(\cdot,0)$ as accurately as possible.

We approximate the solution of this test problem using unstructured quadrilateral meshes with the total number of unknowns $N_h = 31968$ in each case. The results after one revolution ($T = 2\pi$) are shown in Figs. 1–3 for different schemes and discretizations. In contrast to the low-order version of the subcell RD scheme (cf. Figs. 1d–1f), its limited high-order extensions based on FCT do not exhibit p-independent convergence rates (cf. Figs. 2a–2c). As the polynomial degree p is increased while keeping N_h fixed, peak

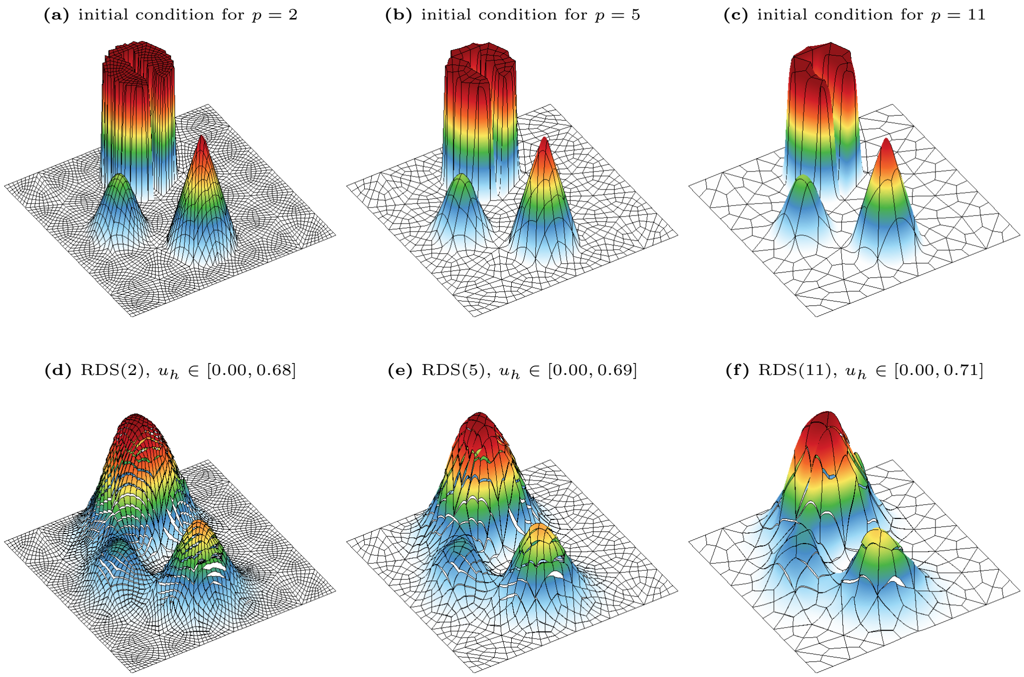
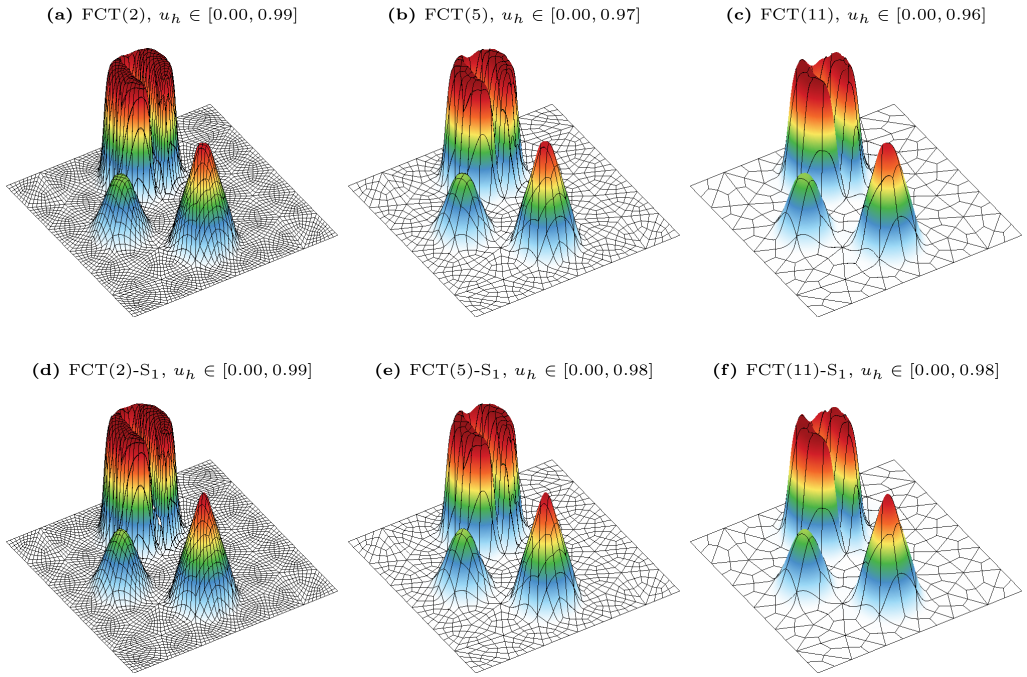

(a) initial condition for $p = 2$
(b) initial condition for $p = 5$
(c) initial condition for $p = 11$
(d) RDS(2), $u_h \in [0.00, 0.68]$
(e) RDS(5), $u_h \in [0.00, 0.69]$
(f) RDS(11), $u_h \in [0.00, 0.71]$

Fig. 1: Initial conditions and corresponding low-order solutions $u_h(\cdot, 2\pi)$ obtained with RDS for $p \in \{2, 5, 11\}$.

clipping effects become more pronounced. The results presented in Figs. 2d–2f confirm that significant improvements in preservation of the smooth and sharp peaks can be achieved by using a smoothness indicator to avoid unnecessary limiting in regions where the second derivatives vary smoothly and γ_i^e approaches 1.

The RD results obtained with the monolithic limiter (MON) illustrate the importance of including the element residuals $\dot{\eta}_i^e$ and the viability of the limiting strategy proposed in Section 4.3. The numerical solution shown in the left column of Fig. 3 was calculated using $\dot{\eta}_i^e = 0$. The failure to include the contribution of the consistent mass matrix gives rise to significant phase errors. The addition of $\dot{\eta}_i^e$ leads to a marked improvement but the MON solutions shown in the middle column of Fig. 3 are far less accurate than the corresponding FCT results in Fig. 2b. This state of affairs is due to the increased importance of mass matrix contributions for larger p and the fact that the FCT approach is designed to produce the least diffusive results for time-dependent problems and small time steps. The ability of MON to



(a) FCT(2), $u_h \in [0.00, 0.99]$

(b) FCT(5), $u_h \in [0.00, 0.97]$

(c) FCT(11), $u_h \in [0.00, 0.96]$

(d) FCT(2)-S$_1$, $u_h \in [0.00, 0.99]$

(e) FCT(5)-S$_1$, $u_h \in [0.00, 0.98]$

(f) FCT(11)-S$_1$, $u_h \in [0.00, 0.98]$

Fig. 2: Numerical solutions $u_h(\cdot, 2\pi)$ obtained using using FCT (top) and FCT-S$_1$ (bottom) for $p \in \{2, 5, 11\}$.

resolve steep gradients and smooth peaks can be greatly improved by using the proposed smoothness indicators. The accuracy of the MON-S$_1$ results presented in Fig. 3f is similar to that of the FCT-S solutions in Fig. 2e.

7.3. Steady advection in 2D

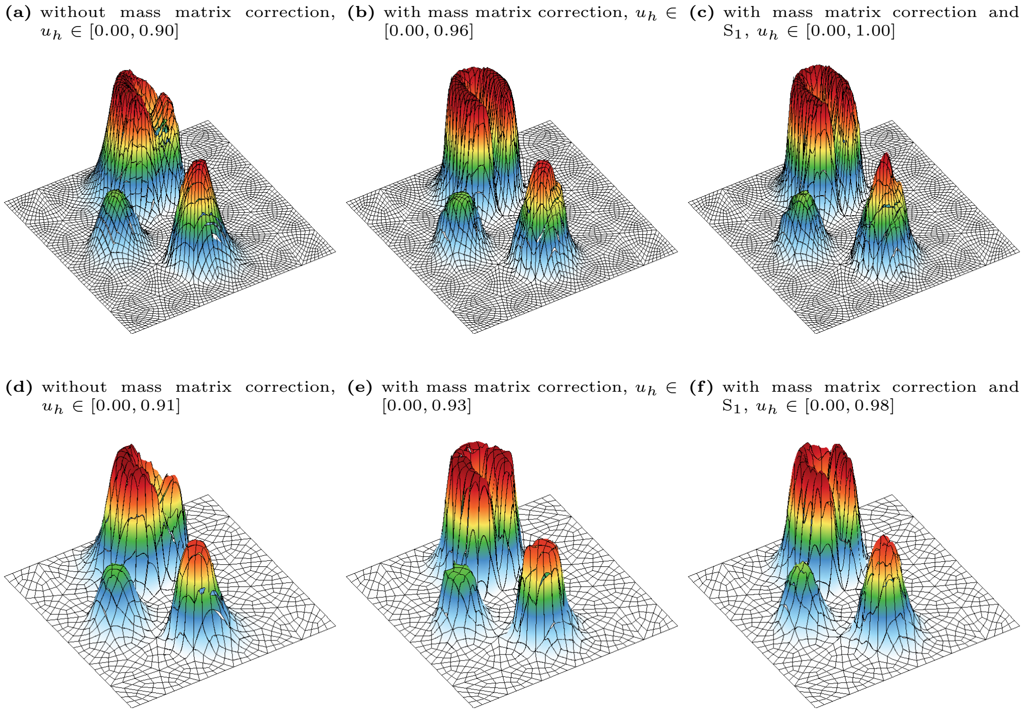
To show the merit of monolithic limiting, we solve the steady linear advection equation

$$\mathbf{v} \cdot \nabla u = 0 \quad \text{in } \Omega = (0,1)^2 \tag{48}$$

using the divergence-free velocity field $\mathbf{v}(x, y) = (y, -x)$ for two test configurations which differ in the choice of the inflow boundary data.

The Test 1 inflow boundary condition and the exact solution in $\bar{\Omega}$ are given by

$$u(x,y) = \frac{1}{4}\left(1 + \tanh\left(\frac{r(x,y) - 0.4}{0.03}\right)\right)\left(1 - \tanh\left(\frac{r(x,y) - 0.6}{0.03}\right)\right),$$



(a) without mass matrix correction, $u_h \in [0.00, 0.90]$ **(b)** with mass matrix correction, $u_h \in [0.00, 0.96]$ **(c)** with mass matrix correction and S_1, $u_h \in [0.00, 1.00]$

(d) without mass matrix correction, $u_h \in [0.00, 0.91]$ **(e)** with mass matrix correction, $u_h \in [0.00, 0.93]$ **(f)** with mass matrix correction and S_1, $u_h \in [0.00, 0.98]$

Fig. 3: Numerical solutions $u_h(\cdot, 2\pi)$ obtained using different versions of MON for $p = 2$ (top) and $p = 5$ (bottom).

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

$$u(x,y) = \begin{cases} 1, & \text{if } \ 0.15 \le r(x,y) \le 0.45, \\ \cos^2\left(10\pi \frac{r(x,y)-0.7}{3}\right), & \text{if } \ 0.55 \le r(x,y) \le 0.85, \\ 0, & \text{otherwise.} \end{cases} \tag{49}$$

Predictor-corrector algorithms of FCT type do not converge to stationary solutions and their accuracy depends on the (pseudo-)time step. For that reason, no grid convergence studies are performed for the FCT version. To study the convergence behavior of MON in the well-defined steady state limit, we calculate experimental orders of convergence for $L^1(\Omega)$ errors on successively refined uniform quadrilateral meshes. In Tables 5–6, we show the outcomes of the grid convergence study for Test 1 with smoothness indicators S_1 and S_2. The convergence rates are optimal for all considered polynomial degrees. Thus, both S_1 and S_2 perform well in the case of smooth

h	$p=1$	EOC	$p=2$	EOC	$p=3$	EOC	$p=4$	EOC
1/48	1.63E-03		8.94E-05		5.88E-06		6.82E-07	
1/64	9.66E-04	1.82	3.22E-05	3.55	1.77E-06	4.17	1.38E-07	5.56
1/96	3.59E-04	2.44	8.48E-06	3.29	3.44E-07	4.04	1.58E-08	5.35
1/128	1.76E-04	2.48	3.48E-06	3.10	1.09E-07	4.01	3.70E-09	5.04
1/192	6.63E-05	2.41	1.02E-06	3.02	2.15E-08	4.00		
1/256	3.43E-05	2.29	4.32E-07	3.00				
1/384	1.42E-05	2.17						

Table 5: $\|\cdot\|_{L^1(\Omega)}$ errors and corresponding EOC for MON-S$_1$ and $p \in \{1,\dots,4\}$.

h	$p=1$	EOC	$p=2$	EOC	$p=3$	EOC	$p=4$	EOC
1/48	1.62E-03		8.94E-05		5.88E-06		6.82E-07	
1/64	9.66E-04	1.80	3.22E-05	3.55	1.77E-06	4.17	1.38E-07	5.56
1/96	3.59E-04	2.44	8.48E-06	3.29	3.44E-07	4.04	1.58E-08	5.35
1/128	1.76E-04	2.48	3.48E-06	3.10	1.09E-07	4.01	3.70E-09	5.04
1/192	6.63E-05	2.41	1.02E-06	3.02	2.15E-08	4.00		
1/256	3.43E-05	2.29	4.32E-07	3.00				
1/384	1.42E-05	2.17						

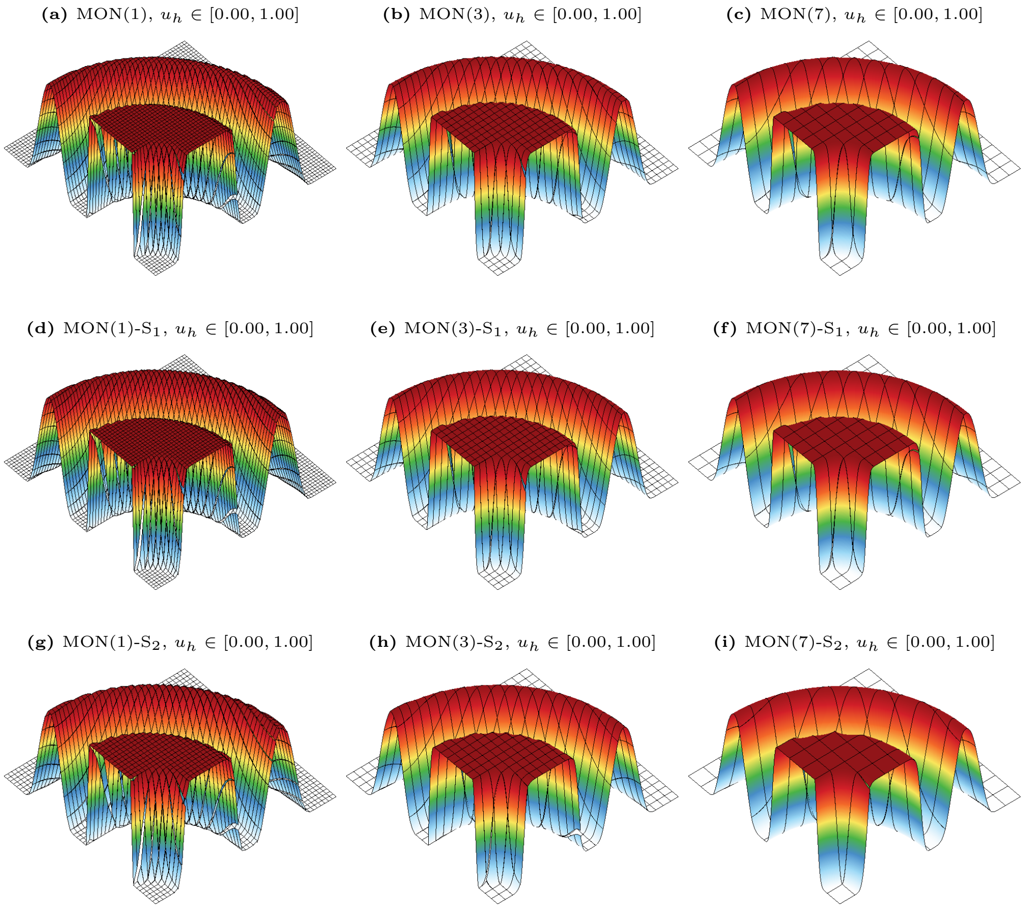
Table 6: $\|\cdot\|_{L^1(\Omega)}$ errors and corresponding EOC for MON-S$_2$ and $p \in \{1,\dots,4\}$.

solutions, in 1D as well as 2D scenarios. The steady-state MON, MON-S$_1$, and MON-S$_2$ results for Test 2 are shown in Fig. 4. It can be seen that the use of smoothness indicators improves the peak preservation capability of MON without generating spurious undershoots and/or overshoots in the neighborhood of discontinuities.

7.4. Application to advection based remap

The tests in this section are used to assess the behavior of the proposed methods in the framework of advection remap. As explained in Section 7 of [4], solving the transport equation on a fixed grid is similar to remapping fields on a moving mesh. A difference in the methodology is that the latter approach requires updating the mesh positions and recomputing mass and convective matrices corresponding to the moved mesh. The limiting procedure, however, is identical.

Advection based remap is performed by evolving the numerical solution in pseudotime $\tau \in [0,1]$. The parametrization limits $\tau \in \{0,1\}$ correspond to the initial and final mesh, respectively. Given a domain Ω and two corresponding meshes with the same topological structure, we define the velocity $\boldsymbol{\nu}(\mathbf{x}_i^e)$ of a mesh node $\mathbf{x}_i^e$ as the difference between its final and initial posi-



(a) MON(1), $u_h \in [0.00, 1.00]$

(b) MON(3), $u_h \in [0.00, 1.00]$

(c) MON(7), $u_h \in [0.00, 1.00]$

(d) MON(1)-S$_1$, $u_h \in [0.00, 1.00]$

(e) MON(3)-S$_1$, $u_h \in [0.00, 1.00]$

(f) MON(7)-S$_1$, $u_h \in [0.00, 1.00]$

(g) MON(1)-S$_2$, $u_h \in [0.00, 1.00]$

(h) MON(3)-S$_2$, $u_h \in [0.00, 1.00]$

(i) MON(7)-S$_2$, $u_h \in [0.00, 1.00]$

Fig. 4: Numerical solutions u_h obtained with MON(p) (top), MON(p)-S$_1$ (middle) and MON(p)-S$_2$ (bottom) for $p \in \{1, 3, 7\}$.

tions $\boldsymbol{\nu}(\mathbf{x}_i^e) = \mathbf{x}_i^e \mid_{\tau=1} - \mathbf{x}_i^e \mid_{\tau=0}$. The goal is to leave a function u, defined in Ω, unchanged during mesh movement. This requirement is expressed by the PDE system

$$\frac{\mathrm{d}u}{\mathrm{d}\tau} = \boldsymbol{\nu} \cdot \nabla u, \qquad \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\tau} = \boldsymbol{\nu} \qquad \text{in } \Omega \times [0,1]. \tag{50}$$

Since advection based remap fits in the framework of time dependent problems, it is worthwhile to solve (50) with FCT rather than MON. We present results for two remap configurations, where we start with uniform Cartesian meshes in either case. In the first test, the initial condition is identical to that of Section 7.2. The initial condition for the second test is

$$\begin{aligned} u(x,y) = \frac{1}{16} &\left[\xi(10(x-0.45))\, \xi(-10(x-0.45)) \right. \\ &\left. \xi(10(y-0.05))\, \xi(-10(y+0.45)) \right], \end{aligned}$$

where

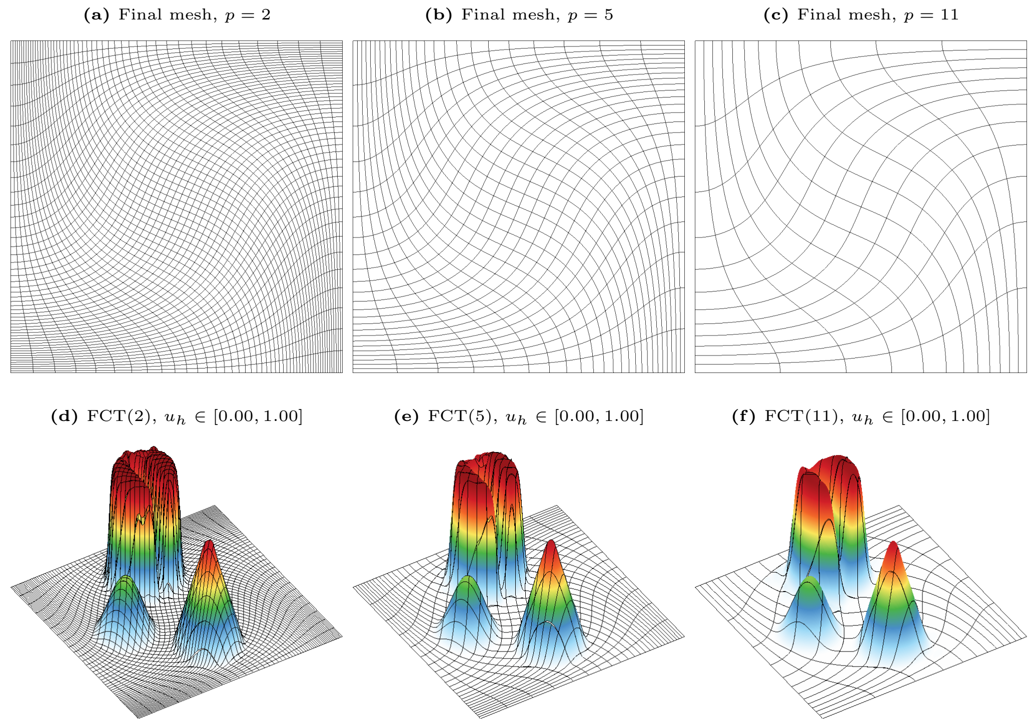
$$\xi(z) = \frac{1}{\sqrt{\pi}} \int_{-z}^{z} e^{-s^2} \mathrm{d}s$$

is the Gauss error function.

In both tests, we assess the ability of the scheme to prevent violations of maximum principles for Bernstein polynomials of degree $p \in \{2, 5, 11\}$ using a fixed total number of unknowns. All experiments are run using third order curved meshes. The final mesh configurations and remapped solutions are shown in Figs. 5–6. We observe that the initial bounds and shapes of the initial conditions are preserved despite substantial mesh deformations.

8. Conclusions

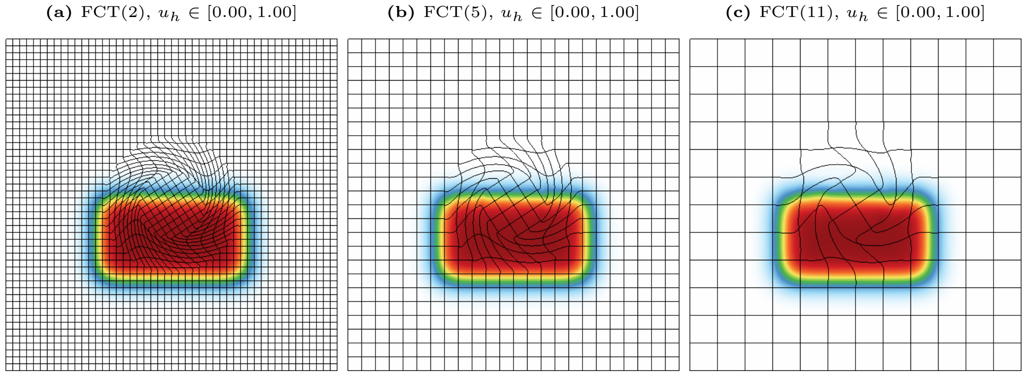
The theoretical and numerical studies presented in this work indicate that a properly designed RD scheme for high-order finite elements can suppress undershoots/overshoots without degrading the order of accuracy in smooth regions. However, the design of accuracy-preserving LED approximations is more difficult than for $\mathbb{P}_1$ and $\mathbb{Q}_1$ elements. Instead of modifying the bound-violating element contributions of a high-order Bernstein finite element approximation directly, discrete maximum principles can be enforced using *hp*-adaptive schemes in which algebraic fixes are restricted to $\mathbb{P}_1/\mathbb{Q}_1$



(a) Final mesh, $p = 2$ **(b)** Final mesh, $p = 5$ **(c)** Final mesh, $p = 11$

(d) FCT(2), $u_h \in [0.00, 1.00]$ **(e)** FCT(5), $u_h \in [0.00, 1.00]$ **(f)** FCT(11), $u_h \in [0.00, 1.00]$

Fig. 5: Mesh (top) and solution u_h (bottom) at $\tau = 1$ obtained using FCT(p) with $p \in \{2, 5, 11\}$ for the Taylor-Green vortex.



(a) FCT(2), $u_h \in [0.00, 1.00]$ **(b)** FCT(5), $u_h \in [0.00, 1.00]$ **(c)** FCT(11), $u_h \in [0.00, 1.00]$

Fig. 6: FCT(p) results of u_h at $\tau = 1$ for $p \in \{2, 5, 11\}$ for the Gresho vortex.

subdomains. In the context of stationary elliptic problems (including singularly perturbed advection-diffusion equations), novel *hp*-FEM approximations of this kind were recently proposed in [20]. Their extension to time-dependent conservation laws would further advance the state of the art in the field of bound-preserving high-resolution finite element schemes.

Acknowledgments

This work was performed under the auspices of the U.S. Department of Energy under Contract DE-AC52-07NA27344, LLNL-JRNL-796580. The first two authors were supported by the German Research Association (DFG) under grant KU 1530/23-1. The work of I. Tomas and J. N. Shadid was partially supported by the U.S. Department of Energy, Office of Science, Office of Applied Scientific Computing Research. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Disclaimer. This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

In memoriam

This paper is dedicated to the memory of Dr. Douglas Nelson Woods (*January 11th 1985 - †September 11th 2019), promising young scientist and postdoctoral 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. Abgrall, A review of residual distribution schemes for hyperbolic and parabolic problems: The July 2010 state of the art. *Commun. Comput. Phys.* **11** (2012) 1043–1080.

[2] R. Abgrall and S. Tokareva, Staggered grid residual distribution scheme for Lagrangian hydrodynamics. *SIAM J. Sci. Comput.* **39** (2017) A2317–A2344.

[3] R. Abgrall and J. Trefilík, An example of high order residual distribution scheme using non-Lagrange elements. *J. Sci. Comput.* **45** (2010) 3–25.

[4] 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.

[5] S. Badia and J. Bonilla, Monotonicity-preserving finite element schemes based on differentiable nonlinear stabilization. *Computer Methods Appl. Mech. Engrg.* **313** (2017) 133-158.

[6] S. Badia, J. Bonilla and A. Hierro, Differentiable monotonicity-preserving schemes for discontinuous Galerkin methods on arbitrary meshes. *Computer Methods Appl. Mech. Engrg.* **320** (2017) 582-605.

[7] G. Barrenechea, V. John and P. Knobloch, Analysis of algebraic flux correction schemes. *SIAM J. Numer. Anal.* **54** (2016) 2427–2451.

[8] 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.

[9] S. Diot, S. Clain and R. Loubére, Improved detection criteria for the Multi-dimensional Optimal Order Detection (MOOD) on unstructured meshes with very high-order polynomials. *Comput. Fluids* 64 (2012) 43–63.

[10] S. Diot, R. Loubére and S. Clain, The Multidimensional Optimal Order Detection method in the three-dimensional case: very high-order finite volume method for hyperbolic systems. *Int. J. Numer. Methods Fluids* **73** (2013) 362–392.

[11] V. Dobrev, Tz. Kolev, D. Kuzmin, R. Rieben and V. Tomov, Sequential limiting in continuous and discontinuous Galerkin methods for the Euler equations. *J. Comput. Phys.* **356** (2018) 372–390.

[12] M. Dumbser, O. Zanotti, R. Loubère and S. Diot, A posteriori subcell limiting of the discontinuous Galerkin finite element method for hyperbolic conservation laws. *J. Comput. Phys.* **278** (2014) 47–75.

[13] T. N. T. Goodman, Further variation diminishing properties of Bernstein polynomials on triangles. *Constructive Approximation* **3** (1987) 297–305.

[14] 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.

[15] 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. Comput.* **40** (2018) A3211–A3239.

[16] H. Hajduk, D. Kuzmin, Tz. Kolev and R. Abgrall, Matrix-free subcell residual distribution for Bernstein finite element discretizations of linear advection equations. Submitted to *Computer Methods Appl. Mech. Engrg.* Preprint: *Ergebnisber. Inst. Angew. Math.* **598** TU Dortmund, 2019.

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

[18] D. Kuzmin, Gradient-based limiting and stabilization of continuous Galerkin methods. *Ergebnisber. Angew. Math.* **589**, TU Dortmund University, 2018. To appear in: G. Rozza et al. (eds), LNCSE Springer FEF 2017 Special Volume (2019).

[19] D. Kuzmin and M. Möller, Algebraic flux correction I. Scalar conservation laws. In: D. Kuzmin, R. Löhner, S. Turek (eds), *Flux-Corrected Transport: Principles, Algorithms, and Applications.* Springer, 1st edition: 2005, pp. 155–206, 2nd edition: 2012, pp. 145–192.

[20] D. Kuzmin, M. Quezada de Luna and C. Kees, A partition of unity approach to adaptivity and limiting in continuous finite element methods. *Computers & Mathematics with Applications*, Available online 20 March 2019, https://doi.org/10.1016/j.camwa.2019.03.021

[21] D. Kuzmin and F. Schieweck, A parameter-free smoothness indicator for high-resolution finite element schemes. *Central European Journal of Mathematics* **11** (2013) 1478–1488.

[22] R.J. LeVeque, High-resolution conservative algorithms for advection in incompressible flow. *SIAM Journal on Numerical Analysis* **33**, (1996) 627–665.

[23] C. Lohmann, *Physics-Compatible Finite Element Methods for Scalar and Tensorial Advection Problems.* PhD thesis, TU Dortmund University, 2019.

[24] 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.

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

[26] MFEM: Modular finite element methods. https://mfem.org.

[27] F. Vilar, A posteriori correction of high-order discontinuous Galerkin scheme through subcell finite volume formulation and flux reconstruction. *J. Comput. Phys.* **387** (2019) 245–279.

[28] X. Zhang and C.-W. Shu, On maximum-principle-satisfying high order schemes for scalar conservation laws. *J. Comput. Phys.* **229** (2010) 3091–3120.

[29] X. Zhang and C.-W. Shu, Maximum-principle-satisfying and positivity-preserving high-order schemes for conservation laws: survey and new developments. *Proc. R. Soc. A* **467** (2011) 2752–2776.