A monotonicity preserving, nonlinear, finite element upwind method for the transport equation

Erik Burman[a]

[a] *Department of Mathematics, University College London, London, UK–WC1E 6BT, UK*

Abstract

We propose a simple upwind finite element method that is monotonicity preserving and weakly consistent of order $O(h^{\frac{3}{2}})$. The scheme is nonlinear, but since an explicit time integration method is used the added cost due to the nonlinearity is not prohibitive. We prove the monotonicity preserving property for the forward Euler method and for a second order Runge-Kutta method. The convergence properties of the Runge-Kutta finite element method is verified on a numerical example.

Keywords: stabilized finite element method, shock capturing, flux correction, monotonicity preserving, transport equation

1. Introduction

The design of robust and accurate finite element methods for first order hyperbolic equations or convection dominated convection-diffusion problems remains an active field of research. Indeed the task of designing a numerical scheme that is of higher order than one, in the zone where the exact solution is smooth, but preserves the monotonicity properties of the exact solution on the discrete level, is nontrivial. Since it is known that such a scheme necessarily must be nonlinear even for linear equations the typical strategy adopted when working with stabilized finite element methods is to add an additional nonlinear shock-capturing term, designed to make the method satisfy a discrete maximum principle [1, 2, 3]. These methods however often result in very ill-conditioned nonlinear equations and include parameters that may be difficult to tune and depend on the mesh geometry. Another approach is the so-called flux corrected finite element method [4, 5]. In this scheme the system matrix is manipulated so that it becomes a so called M-matrix, the inverse of which has positive coefficients which yields a maximum principle. This scheme is monotonicity preserving, but of first order. In order to improve the accuracy anti-diffusive mechanisms, or flux-limiter techniques, have been proposed that reduce the amount of dissipation in the smooth region by blending a low and a high order approximation [6, 5, 7].

In this paper we will discuss a method that is related to both the above mentioned classes in the sense that the method consists of the addition of a nonlinear dissipative term to the standard Galerkin formulation as for a shock capturing term, but similarly as in a flux corrected transport methods the nonlinear term uses the coefficients of the system matrix for its definition. The method is entirely derived from the finite element variational formulation and the guiding principle of the analysis has been to add the smallest perturbation to the centered standard Galerkin formulation that ensures that the method is monotonicity preserving. The salient features of the resulting method is that the optimal value of the the stabilization parameter can be traced in the analysis, the monotonicity does not require any acute condition of the mesh and the artificial dissipation term depends on the residual of the exact solution in the form of a linear combination of the jumps of directional derivatives over each node (c.f. the edge based limiters that were proposed in the eighties, see [6] and references therein, but also [8, 3]). Formally this leads to a method with $O(h^{\frac{3}{2}})$ artificial viscosity where the solution is smooth and we show in a numerical example that the expected $O(h^{\frac{3}{2}})$ convergence of the error in the L^2-norm, indeed holds on structured meshes.

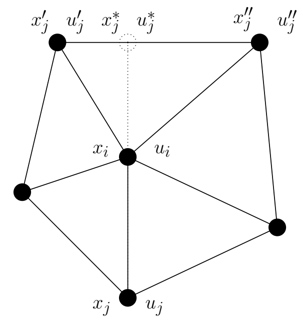

Figure 1: Illustration of the macro patch Ω_i and the points x_i, x_j and x_j^* with associated function values u_i, u_j and u_j^*.

2. Model problem and finite element discretization

We will conside the pure transport equation in $\mathbb{R}^2$

$$\partial_t u + \boldsymbol{\beta} \cdot \nabla u = 0 \tag{1}$$

with $u(x, 0) = u_0(x)$ where $u_0(x)$ is some function with compact support in $\mathbb{R}^2$ and $\boldsymbol{\beta} \in [W^{1,\infty}(\mathbb{R}^2)]^2$. Let $\mathcal{T}_h := \{K\}$ denote a conforming, shape regular, triangulation of $\mathbb{R}^2$. The finite element space of piecewise affine continuous functions is defined on $\mathcal{T}_h$ as

$$V_h := \{v_h \in H^1(\mathbb{R}^2) : v_h|_K \in P_1(K), \forall K \in \mathcal{T}_h\}$$

where $P_1(K)$ denotes the polynomials of degree less than or equal to 1 over K. The nodal basis functions of V_h will be denoted φ_i, i.e. $\varphi_i(x_j) = \delta_{ij}$, with δ_{ij} the Kronecker delta function. Any function $v_h \in V_h$ is then defined by $\sum_i v_i \varphi_i$, where the v_i denotes the nodal values of the function. We denote by $\mathcal{N}_K$ the set of indices of the vertices x_i, of K. We also introduce the length of the edge e_{ij} between the nodes x_i and x_j, $h_{ij} := |x_i - x_j|$ and the unit vector pointing from x_j to x_i, $\tau_{ij} := (x_i - x_j)/h_{ij}$. To each node x_i of the mesh we associate the macro element $\Omega_i := \{K \in \mathcal{T}_h : x_i \in K\}$, with associated set of indices $\mathcal{N}_{\Omega_i}$ of the vertices $x_j \in \Omega_i$. For every node x_j in the boundary of Ω_i we associate a distance $h_{ij}^* > 0$ such that $x_j^* := x_i + h_{ij}^* \tau_{ij} \in \partial\Omega_i$ (see Fig. 1.) The value of the finite element solution at x_j^* will be denoted $u_j^* := u_h(x_j^*)$. If u_j' and u_j'' denotes the values of u_h in the nodes of the endpoints of the edge with x_j^* in its interior we see that there exists some $\alpha_j^* \in (0, 1)$ such that $u_j^* = \alpha_j^* u_j' + (1 - \alpha_j^*) u_j''$. By the shape regularity assumption we know that the number of points x_j^* in the interior of any edge in Ω_i is upper bounded by some $n_i^* \in \mathbb{N}$. Let $\underline{h}_K$ denote the radius of the largest circle inscribed in $K \in \Omega_i$, similarly let $\overline{h}_K$ denote the radius of the smallest circle circumscribing $K \in \Omega_i$ the maximum ratio of the two within one macroelement is denoted $\rho_i := \max_{K \in \Omega_i} \overline{h}_K / \min_{K \in \Omega_i} \underline{h}_K$. We also denote an extended patch, of two layers of elements around the node x_i by $\tilde{\Omega}_i := \cup_{j \in \mathcal{N}_{\Omega_i}} \Omega_j$. We define $(u_h, v_h) := \int_{\mathbb{R}^2} u_h v_h \, \mathrm{d}x$ and the discrete variant obtained by approximating the integral using nodal quadrature ("lumped mass") by $(u_h, v_h)_h := \sum_{K \in \mathcal{T}_h} \sum_{i \in \mathcal{N}_K} u_h(x_i) v_h(x_i) m_K/3$ where m_K denotes the area of the triangle K. Observe in particular that $(u_h, \varphi_i)_h := \frac{1}{3} \sum_{K \in \Omega_i} m_K \, u_h(x_i) = \tilde{m}_i u_h(x_i)$, with $\tilde{m}_i := \frac{1}{3} \sum_{K \in \Omega_i} m_K$.

Now consider the forward Euler finite element discretization of (1), find $u_h^n \in V_h$ such that

$$k^{-1}(u_h^n - u_h^{n-1}, v_h)_h + a(u_h^{n-1}, v_h) + s(u_h^{n-1}; u_h^{n-1}, v_h) = 0 \tag{2}$$

and $(u_h^0, z_h)_h = (u_0, z_h)_h$ for all $v_h, z_h \in V_h$. Here $k \in \mathbb{R}^+$ is the timestep and $a(u_h^{n-1}, v_h) := (\boldsymbol{\beta} \cdot \nabla u_h^{n-1}, v_h)$. In order to define the stabilization operator $s(\cdot; \cdot, \cdot)$ we introduce the upwind and downwind sets of nodes with respect to a node i. Let $\mathcal{N}_{\Omega_i}^+$ be the subsets of vertex indices j in $\mathcal{N}_{\Omega_i}$ such that $a(\varphi_j, \varphi_i) > 0$. Then define $\mathcal{N}_{\Omega_i}^- := \mathcal{N}_{\Omega_i} \setminus \mathcal{N}_{\Omega_i}^+$.

$$s(u_h^{n-1}; u_h^{n-1}, v_h) := (\xi(u_h^{n-1}) u_h^{n-1}, v_h)_h - (\xi(u_h^{n-1}) u_h^{n-1}, v_h), \tag{3}$$

with the nonlinear upwind factor given by

$$\xi(u_h)|_K := \frac{6}{m_K} \max_{i \in \mathcal{N}_K} \left((n_i^* \rho_i + 1) \max_{j \in \mathcal{N}_{\Omega_i}} |a(\varphi_j, \varphi_i)| \frac{\underline{a}_i}{\overline{a}_i} \right) \tag{4}$$

where $\underline{a}_i := |\sum_{j \in \mathcal{N}^+_{\Omega_i}} h_{ij} [\![\nabla u_h \cdot \tau_{ij}]\!]_{x_i} a(\varphi_j, \varphi_i)|$ and $\overline{a}_i := \sum_{j \in \mathcal{N}^+_{\Omega_i}} h_{ij} \{|\nabla u_h \cdot \tau_{ij}|\}_{x_i} |a(\varphi_j, \varphi_i)|$. Here the jump and the average across the node x_i are defined by $[\![\nabla u_h \cdot \tau_{ij}]\!]_{x_i} := \lim_{\epsilon \to 0^+} (\nabla u_h(x_i - \epsilon \tau_{ij}) - \nabla u_h(x_i + \epsilon \tau_{ij})) \cdot \tau_{ij}$ and $\{|\nabla u_h \cdot \tau_{ij}|\}_{x_i} := \frac{1}{2} \lim_{\epsilon \to 0^+} (\nabla u_h(x_i - \epsilon \tau_{ij}) + \nabla u_h(x_i + \epsilon \tau_{ij})) \cdot \tau_{ij}$. Observe that the sum in the definition of $\overline{a}_i$ may also be taken over $\mathcal{N}_{\Omega_i}$ to improve the continuity of $s(u_h; u_h, \cdot)$. For the numerical examples presented below, this modification of $\overline{a}_i$ had no influence on the solution quality. If $\overline{a}_i = 0$ the (undefined) factor $\frac{\underline{a}_i}{\overline{a}_i}$ is replaced by zero (in practice the quotient is perturbed by adding a small positive coefficient to the denominator).

Below we will use the following abstract notation for the Euler step (2): $u_h^n = \mathbb{E} u_h^{n-1}$.

We end this section with a technical lemma, showing some properties of the stabilization operator $s(\cdot; \cdot, \cdot)$. First we show that the stabilization operator is mass conserving, linearity preserving and dissipative, then we give an expression for $s(\cdot; \cdot, \cdot)$ in terms of the local unknowns.

Lemma 2.1. *The stabilization operator defined by* (3) *satisfies*

$$s(u_h; u_h, 1) = 0, \quad s(v_h, u_h, \varphi_i) = 0, \forall v_h \in P_1(\tilde{\Omega}_i), \, i \in [1, dim(V_h)], \tag{5}$$

$$s(u_h; u_h, u_h) = \frac{1}{12} \sum_{K \in \mathcal{T}_h} \xi(u_h)|_K \sum_{i,j \in \mathcal{N}_K} (h_{ij} (\nabla u_h \cdot \tau_{ij})|_{e_{ij}})^2 \, m_K, \tag{6}$$

$$s(u_h; u_h, \varphi_i) = -\frac{1}{12} \sum_{K \in \Omega_i} \xi(u_h)|_K \sum_{j \in \mathcal{N}_K} (u_j - u_i) \, m_K, \quad i \in [1, dim(V_h)]. \tag{7}$$

Proof. For the inequalities of equation (5) first note that the mass conservation property is immediate by the fact that mass lumping integrates piecewise affine functions exactly. The right inequality follows by observing that if $v_h \in P_1(\tilde{\Omega}_i)$ then $\xi(v_h)_K = 0$ for all $K \in \Omega_i$, since $[\![\nabla u_h \cdot \tau]\!]_{x_i} = 0$ for all $i \in \mathcal{N}_{\Omega_i}$ and for all $\tau \in \mathbb{R}^2$. The results (6), (7) follow by straightforward integration. For a given node $x_i \in K$ we denote the two other nodes in K by x_i' and x_i'' and the associated coefficients $u_i' = u_h(x_i')$ and $u_i'' = u_h(x_i'')$

$$s(u_h; u_h, u_h) = \sum_{K \in \mathcal{T}_h} \frac{m_K}{3} \xi(u_h)|_K \sum_{i \in \mathcal{N}_K} u_i^2 - \sum_{K \in \mathcal{T}_h} \Big(\frac{m_K}{3} \xi(u_h)|_K \sum_{i \in \mathcal{N}_K} \frac{1}{2} (u_i^2 + \frac{1}{2} u_i u_i' + \frac{1}{2} u_i u_i'') \Big)$$
$$= \sum_{K \in \mathcal{T}_h} \Big(\frac{m_K}{3} \xi(u_h)|_K \sum_{i \in \mathcal{N}_K} \frac{1}{4} \Big((u_i - u_i')^2 + (u_i - u_i'')^2 \Big) \Big).$$

The first equality (6) follows after recalling that $(u_i - u_j)^2 = (h_{ij} (\nabla u_h \cdot \tau_{ij})|_{e_{ij}})^2$. For the second relation (7) first integrate using midpoint quadrature for the consistent mass

$$(\xi(u_h) \, u_h, \varphi_i) = \sum_{K \in \Omega_i} \Big(\frac{m_K}{3} \xi(u_h)|_K \frac{1}{2} \sum_{\substack{j \in \mathcal{N}_K \\ j \neq i}} (u_i + u_j)/2 \Big)$$

and then the nodal quadrature approximation, $(\xi(u_h) \, u_h, \varphi_i)_h = \sum_{K \in \Omega_i} \frac{m_K}{3} \xi(u_h)|_K \, u_i$. Finally take the difference of the two expressions. □

Remark 2.1. *The consistency of the scheme is expected to be of first order close to local extrema and of order* $O(h^{\frac{3}{2}})$ *where the solution is smooth. This is reflected in equation* (6) *by observing that for a triangle* *where none of the nodes in the associated macroelements have a local extremum we expect* $\xi(u_h)|_K = O(h^{-\frac{1}{2}})$ *leading to a dissipation of order* $O(h^{\frac{3}{2}})$ *whereas if there is a local extremum, then* $\xi(u_h)|_K = O(h^{-1})$*, leading to first order dissipation. Typically for linear stabilized methods* $O(h^{3/2})$ *diffusion is compatible with* $O(h^{3/2})$ *error estimates in the* L^2*-norm.*

3. Discrete maximum principle (DMP) for the forward Euler scheme

The nonlinear factor $\xi(u_h)$ has been designed so that $s(u_h; u_h, v_h)$ should make the scheme monotonicity preserving, while adding in some sense the smallest perturbation possible. We prove this property in the following main result of this note. Observe that this result holds for any bilinear form $a(\cdot,\cdot)$ such that $a(c, v_h) = 0$ for $c \in \mathbb{R}$ and for all $v_h \in V_h$, not only the transport operator.

Theorem 3.1. *Let* u_h^n *be the solution of* (2)*, computed under the CFL-condition*

$$k < \frac{1}{10}\Big(\max_i \Big[\frac{card(\mathcal{N}_{\Omega_i})}{\tilde{m}_i}(1 + n_i^* \rho_i) \max_{j \in \mathcal{N}_{\Omega_i}} |a(\varphi_j, \varphi_i)|\Big]\Big)^{-1}$$

then there holds for all nodes x_i *and all* $n > 0$*,*

$$\min_{x \in \Omega_i} u_h^{n-1} \le u_h^n(x_i) \le \max_{x \in \Omega_i} u_h^{n-1}.$$

Proof. By the linearity of $a(\cdot,\cdot)$ and the property $a(u_i, \varphi_i) = 0$ since $u_i \in \mathbb{R}$, it follows that

$$\begin{aligned} a(u_h, \varphi_i) &= \sum_{j \in \mathcal{N}_{\Omega_i}} (u_j - u_i) a(\varphi_j, \varphi_i) = \sum_{j \in \mathcal{N}^-_{\Omega_i}} (u_j - u_i) a(\varphi_j, \varphi_i) + \sum_{j \in \mathcal{N}^+_{\Omega_i}} (u_j - u_i) a(\varphi_j, \varphi_i) \\ &= \sum_{j \in \mathcal{N}^-_{\Omega_i}} (u_j - u_i) a(\varphi_j, \varphi_i) - h_{ij}/h_{ij}^* \sum_{j \in \mathcal{N}^+_{\Omega_i}} (u_j^* - u_i) a(\varphi_j, \varphi_i) - \sum_{j \in \mathcal{N}^+_{\Omega_i}} h_{ij} [\![\nabla u_h \cdot \tau_{ij}]\!]_{x_i} a(\varphi_j, \varphi_i). \end{aligned}$$

Consider now the scheme (2) tested with φ_i and apply the previous inequality and (7) to obtain

$$\begin{aligned} \tilde{m}_i u_h^n(x_i) = \tilde{m}_i u_i - k \sum_{j \in \mathcal{N}^-_{\Omega_i}} (u_j - u_i) a(\varphi_j, \varphi_i) + k \sum_{j \in \mathcal{N}^+_{\Omega_i}} h_{ij}/h_{ij}^* (u_j^* - u_i) a(\varphi_j, \varphi_i) \\ + k \sum_{j \in \mathcal{N}^+_{\Omega_i}} h_{ij} [\![\nabla u_h \cdot \tau_{ij}]\!]_{x_i} a(\varphi_j, \varphi_i) + k \frac{1}{12} \sum_{K \in \Omega_i} \xi(u_h)|_K \sum_{j \in \mathcal{N}_K} (u_j - u_i)\, m_K. \end{aligned}$$

Here we have dropped the superscript $n-1$ in the right hand side. Observe that to bound the first term of the second line we may use the equality $k|\sum_{j \in \mathcal{N}^+_{\Omega_i}} h_{ij} [\![\nabla u_h \cdot \tau_{ij}]\!]_{x_i} a(\varphi_j, \varphi_i)| = k\frac{a_i}{\overline{a}_i}\overline{a}_i$. Also observe that the $\overline{a}_i$ factor may be bounded by $\overline{a}_i \le \frac{1}{2} \max_{j \in \mathcal{N}_{\Omega_i}} |a(\varphi_j, \varphi_i)| \sum_{j \in \mathcal{N}_{\Omega_i}} 2(1 + n_i^* \rho_i)|u_j - u_i|$, where we used that $h_{ij}/h_{ij}^* \le \rho_i$. Expressing the last two terms of the first line using positive coefficients α_{ij}, satisfying the bound $0 \le \alpha_{ij} \le 2(1 + n_i^* \rho_i) \max_{j \in \mathcal{N}_{\Omega_i}} |a(\varphi_j, \varphi_i)|$, we have

$$\begin{aligned} \tilde{m}_i u_h^n(x_i) \le \tilde{m}_i u_i + k \sum_{j \in \mathcal{N}_{\Omega_i}} \alpha_{ij}(u_j - u_i) + k(n_i^* \rho_i + 1) \max_{j \in \mathcal{N}_{\Omega_i}} |a(\varphi_j, \varphi_i)| \frac{a_i}{\overline{a}_i} \sum_{j \in \mathcal{N}_{\Omega_i}} |u_j - u_i| \\ + k \frac{1}{12} \sum_{K \in \Omega_i} \xi(u_h)|_K \sum_{j \in \mathcal{N}_K} (u_j - u_i)\, m_K. \end{aligned}$$

To exemplify the construction of the α_{ij} assume that there is only one $l \in \mathcal{N}^+_{\Omega_i}$ such that x_l^* is in one of the two edges adjacent to a node x_j, with $j \in \mathcal{N}^-_{\Omega_i}$ then $\alpha_{ij} = -a(\varphi_j, \varphi_i) + \alpha_l^* h_{il}/h_{il}^* a(\varphi_l, \varphi_i)$, with α_l^* the weight introduced in Section 2, such that $x_l' = x_j$. Using the definition of $\xi(u_h)$ the last two terms in the right hand side may be bounded as

$$\tilde{m}_i u_h^n(x_i) \le \tilde{m}_i u_i + k \sum_{j\in\mathcal{N}_{\Omega_i}} \alpha_{ij}(u_j - u_i) + k\frac{1}{3} \sum_{K\in\Omega_i} \xi(u_h)|_K \sum_{j\in\mathcal{N}_K} (u_j - u_i)_+ \, m_K$$

where $(x)_+ := \max(0, x)$. Introducing positive weights $\tilde{\alpha}_{ij} = \frac{1}{3}(\xi(u_h)|_{K'} m_{K'} + \xi(u_h)|_{K''} m_{K''})$ with $e_{ij} = K' \cap K''$ and satisfying the bounds $0 \le \tilde{\alpha}_{ij} \le \frac{2}{3} \max_{K\in\Omega_i}(m_K \xi(u_h)|_K) \le 8(1 + n_i^* \rho_i) \max_{j\in\mathcal{N}_{\Omega_i}} |a(\varphi_j, \varphi_i)|$ (recall that $\underline{a}_i/\overline{a}_i \le 2$) this may be written as

$$u_h^n(x_i) \le u_i + \frac{k}{\tilde{m}_i} \sum_{j\in\mathcal{N}_{\Omega_i}} \alpha_{ij}(u_j - u_i) + \frac{k}{\tilde{m}_i} \sum_{j\in\mathcal{N}_{\Omega_i}} \tilde{\alpha}_{ij}(u_j - u_i)_+.$$

Recalling the CFL-condition on k and the bounds on α_{ij} and $\tilde{\alpha}_{ij}$ we see that

$$\frac{k}{\tilde{m}_i} \sum_{j\in\mathcal{N}_{\Omega_i}} (\alpha_{ij} + \tilde{\alpha}_{ij}) \le 10\frac{k}{\tilde{m}_i} \operatorname{card}(\mathcal{N}_{\Omega_i})(1 + n_i^* \rho_i) \max_{j\in\mathcal{N}_{\Omega_i}} |a(\varphi_j, \varphi_i)| < 1.$$

We conclude that there exists weights $\alpha_j \in [0, 1]$, $j \in \mathcal{N}_{\Omega_i}$ such that $\sum_{j\in\mathcal{N}_{\Omega_i}} \alpha_j < 1$ and $u_h^n(x_i) \le \sum_{j\in\mathcal{N}_{\Omega_i}} \alpha_j u_j$. From this the upper bound follows. The proof of the lower bound is similar. □

4. Extension to second order in time

We consider Heun's method, which is an explicit second order in time Runge-Kutta method (RK2) defined by the following linear combination of two explicit Euler step: $w_h^n = \mathbb{E}u_h^{n-1}$, $\tilde{w}_h^n = \mathbb{E}w_h^n$ and $u_h^n = \frac{1}{2}(u_h^{n-1} + \tilde{w}_h^n)$. We now prove that the Runge-Kutta finite element method inherits the stability of the forward Euler finite element method. In this case the domain of dependence becomes one layer of elements wider.

Proposition 4.1. *Let u_h^n be the solution of RK2 then for all nodes $x_i \in \mathcal{T}_h$ and all $n > 0$,*

$$\min_{x\in\tilde{\Omega}_i} u_h^{n-1}(x) \le u_h^n(x_i) \le \max_{x\in\tilde{\Omega}_i} u_h^{n-1}(x). \tag{8}$$

Proof. Observe that by Theorem 3.1 there holds $\min_{x\in\tilde{\Omega}_i} u_h^{n-1}(x) \le \min_{x\in\Omega_i} w_h^n \le \tilde{w}_h^n(x_i) \le \max_{x\in\Omega_i} w_h^n \le \max_{x\in\tilde{\Omega}_i} u_h^{n-1}(x)$. From this (8) follows since $u_h^n(x_i) = \frac{1}{2}\left(u_h^{n-1}(x_i) + \tilde{w}_h^n(x_i)\right)$. □

5. Numerical examples

The computations were carried out using FreeFEM++ [9]. We first consider an example in the bounded domain $\Omega = (0, 3) \times (0, 1)$ and solve the equation (1), on the time interval $[0, 1]$ for $\boldsymbol{\beta} = (1, 0)^T$, with $u_0 = (7r < \pi)(\cos(7r) + 1)/2$ where $r^2 = (x - 1.0)^2 + (y - 0.5)^2$ using the RK2 scheme, with $\mathbb{E}$ defined by (2) on a series of structured meshes consisting of right triangles with side $h = 0.025$, 0.0125, 0.00625, 0.003125 respectively. On the structured mesh with constant $\boldsymbol{\beta}$ we get $\xi(u_h)|_K = 4/h \max_{i\in\mathcal{N}_K}(\underline{a}_i/(\overline{a}_i + \varepsilon h))$ and chose $\varepsilon = 10^{-15}$. The timestep was set to $k = h/4$. We then considered the case of discontinuous initial data $u_0 = (7r < \pi)$. The errors in the L^2-norms with experimental convergence orders for both cases are reported in the columns marked (*) of Table 1. For the smooth solution convergence of order $O(h^{\frac{3}{2}})$ was observed. The maximum principle was respected to machine precision on all meshes. We then considered a computation in $\Omega = (0, 1) \times (0, 1)$, and solved the equation (1) on $[0, 1.5]$ with $\boldsymbol{\beta} = (\sin(\pi x)^2 \sin(2\pi y), -\sin(\pi y)^2 \sin(2\pi x))^T \cos(\pi t/T)$, $u0 = (12r < \pi)(\cos(12r) + 1)/2$ or $u0 = (12r < \pi)$, where $r^2 = (x - 0.35)^2 + (y - 0.5)^2$ using the same discretization parameters as above on structured

h	(*) smooth L^2	(*) rough L^2	(**S) smooth L^2	(**S) rough L^2	$\langle$DMP$\rangle$	(**U) smooth L^2	(**U) rough L^2	$\langle$DMP$\rangle$
0.025	0.11 (–)	0.27 (–)	0.25 (–)	0.34 (–)	$\langle 0.51 \rangle$	0.28 (–)	0.35 (–)	$\langle 1.6 \rangle$
0.0125	0.037 (1.8)	0.21 (0.36)	0.081 (1.6)	0.26 (0.38)	$\langle 0.78 \rangle$	0.092 (1.6)	0.26 (0.34)	$\langle 2.3 \rangle$
0.00625	0.011 (1.8)	0.17 (0.30)	0.017 (2.2)	0.20 (0.38)	$\langle 1.1 \rangle$	0.038 (1.3)	0.22 (0.24)	$\langle 2.1 \rangle$
0.003125	0.0038 (1.5)	0.13 (0.39)	0.0032 (2.4)	0.16 (0.32)	$\langle 1.6 \rangle$	0.010 (1.9)	0.18 (0.34)	$\langle 2.9 \rangle$

Table 1: Relative errors in the L^2-norm at the final time for the smooth and the rough solutions, experimental convergence orders in parenthesis. Violation of the DMP in % of $\|u\|_{L^\infty(\Omega)}$ for rough solutions

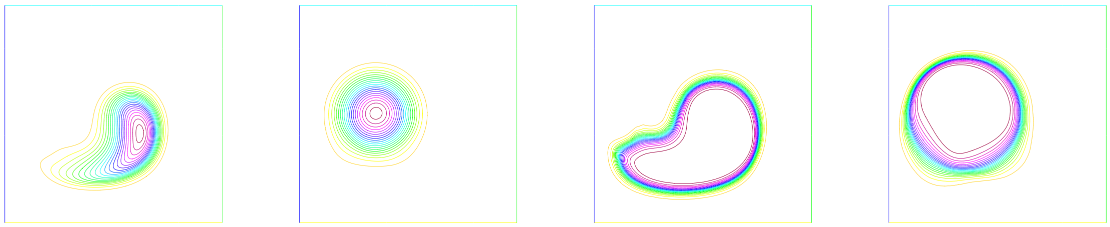

Figure 2: Contourplots at $T/2$ and T, for $h = 0.00625$. Left two plots (**S), smooth; right two plots (**S) rough.

and unstructured meshes. In this case only first order convergence was observed for smooth solutions. Increasing the regularization to $\varepsilon = 0.05$ on structured meshes and $\varepsilon = 0.1$ on unstructured improved the convergence orders. The results are reported in Table 1, in the columns marked (**S) for structured meshes and (**U) for unstructured. When this stronger regularization was used the maximum principle was violated by up to 2.9%. For comparison the standard Galerkin method violates the maximum principle by up to 70% for the rough cases. Example of contour plots of the solutions for maximum deformation at $T/2$ and at final time are presented in Fig. 2.

References

[1] A. Mizukami, T. J. R. Hughes, A Petrov-Galerkin finite element method for convection-dominated flows: an accurate upwinding technique for satisfying the maximum principle, Comput. Methods Appl. Mech. Engrg. 50 (2) (1985) 181–193.

[2] E. Burman, A. Ern, Stabilized Galerkin approximation of convection-diffusion-reaction equations: discrete maximum principle and convergence, Math. Comp. 74 (252) (2005) 1637–1652 (electronic).

[3] S. Badia, A. Hierro, On monotonicity-preserving stabilized finite element approximations of transport problems, SIAM J. Sci. Comput. 36 (6) (2014) A2673–A2697.

[4] R. Löhner, K. Morgan, J. Peraire, M. Vahdati, Finite element flux-corrected transport (FEM–FCT) for the Euler and Navier–Stokes equations, Int. J. Numer. Meths. Fluids 7 (10) (1987) 1093–1109.

[5] D. Kuzmin, S. Turek, Flux correction tools for finite elements, J. Comput. Phys. 175 (2) (2002) 525–558.

[6] P. R. M. Lyra, K. Morgan, A review and comparative study of upwind biased schemes for compressible flow computation. III., Arch. Comput. Methods Engrg. 9 (3) (2002) 207–256.

[7] J.-L. Guermond, M. Nazarov, B. Popov, Y. Yang, A second-order maximum principle preserving Lagrange finite element technique for nonlinear scalar conservation equations, SIAM J. Numer. Anal. 52 (4) (2014) 2163–2182.

[8] E. Burman, On nonlinear artificial viscosity, discrete maximum principle and hyperbolic conservation laws, BIT 47 (4) (2007) 715–733.

[9] F. Hecht, New development in FreeFem++, J. Numer. Math. 20 (3-4) (2012) 251–265.