

WEIGHTING THE EDGE STABILIZATION*

ALEXANDRE ERN[†] AND JEAN-LUC GUERMOND[‡]

Abstract. A modification of the edge stabilization technique is proposed to improve the behavior of the method when solving conservation equations with nonsmooth data or nonsmooth solutions. The key ingredient is tempering the edge stabilization in regions of large gradients through appropriate weights. The new method is shown to preserve the convergence properties of the original method on smooth solutions, and numerical tests indicate that it performs better on nonsmooth solutions.

Key words. finite elements, conservation equations, edge stabilization, linear stabilization, nonlinear viscosity

AMS subject classifications. 65M12, 35L65, 65M60

DOI. 10.1137/120867482

1. Introduction. Linear stabilization techniques are known to be very effective for solving linear first-order PDEs like the transport equation. In particular, denoting h the meshsize and k the polynomial degree of the approximation, linear stabilization techniques yield the near optimal convergence rate $\mathcal{O}(h^{k+\frac{1}{2}})$ in the L^2 -norm for smooth solutions. The situation is not so clear when it comes to solving linear problems with nonsmooth data and nonlinear conservation equations with nonunique weak solutions. Linear stabilization methods generally promote the Gibbs phenomenon and introduce high-order dissipation that, in the case of nonlinear conservation equations, can lead to convergence to entropy-violating solutions. The objective of the present paper is to propose a nonlinear modification of a standard linear stabilization technique (edge stabilization) that alleviates these problems.

We focus our attention in this paper on the edge stabilization technique [8, 7] as a prototype of linear stabilization that is relatively easy to implement with H^1 -conforming finite elements. We show through numerical examples that edge stabilization promotes the Gibbs phenomenon and cannot select the proper entropy solution of some nonlinear conservation equations with nonconvex flux. We also show that the extra dissipation induced by edge stabilization can transform a convergent nonlinear viscosity method into a nonconvergent one when the edge stabilization is not properly scaled. The purpose of this paper is then to introduce and analyze a modified version of edge stabilization that does not suffer from the above problems. The main modification we propose consists of introducing a nonlinear mechanism that weakens the edge stabilization in the regions where the discrete solution exhibits large gradients. This may seem a bit counterintuitive at first glance, since the use of linear stabilization techniques is often motivated to counter spurious oscillations that are produced by large gradients. The proposed method is proved to deliver the near optimal convergence rate $\mathcal{O}(h^{k+\frac{1}{2}})$ in the L^2 -norm for smooth solutions. Numerical

*Received by the editors February 27, 2012; accepted for publication (in revised form) February 26, 2013; published electronically June 4, 2013. This material is based upon work supported by National Science Foundation grants DMS-1015984 and DMS-1217262, AFOSR grant FA99550-12-0358, and GNR MOMAS (CNRS/PACEN, ANDRA, BRGM, CEA, EDF, IRSN).

<http://www.siam.org/journals/sinum/51-3/86748.html>

[†]Université Paris-Est, CERMICS (ENPC), 77455 Marne-la-Vallée cedex 2, France (ern@cermics.enpc.fr).

[‡]Department of Mathematics, Texas A&M University, 3368 TAMU, College Station, TX 77843 (guermond@math.tamu.edu).

tests on the linear transport equation in one and two space dimensions show that the weighted edge stabilization performs as required when combined with a nonlinear viscosity method: it no longer promotes the Gibbs phenomenon and does not prevent the nonlinear viscosity method to converge to the correct entropy solution. In other words, the weighted edge stabilization does not antagonize the nonlinear viscosity method. Quite importantly, when combined with a nonlinear viscosity method and for polynomial orders larger than or equal to two, we observe that the weighted edge stabilization increases the convergence order of the nonlinear method in the regions where the solution is smooth. Thus, when combining the weighted edge stabilization with a nonlinear viscosity method, one improves the convergence order of the nonlinear viscosity method without sacrificing its weakened maximum principle property and its ability to properly converge to entropy solutions.

The paper is organized as follows. In section 2, we set the notation and present numerical experiments illustrating the main difficulties that are addressed herein. In section 3, we introduce and analyze the weighted edge stabilization method. The two key theoretical results of the paper are Theorem 3.1 and Theorem 3.2. In section 4, we present one- and two-dimensional tests that illustrate the improvements achieved by weighting the edge stabilization. Finally, we draw some conclusions in section 5.

2. Preliminaries. The objective of this section is twofold: (i) to set the notation and the model problems we are interested in and (ii) to present numerical experiments that identify the main difficulties that we want to address in the present work. The key conclusions of this section are summarized in section 2.5.

2.1. Formulation of the problem. We are interested in approximating the solution of scalar-valued conservation equations in the form

$$(2.1) \quad \partial_t u + \nabla \cdot \mathbf{f}(u) = 0, \quad u(x, 0) = u_0(x), \quad (x, t) \in \Omega \times \mathbb{R}_+,$$

where Ω is an open polyhedral domain in \mathbb{R}^d and $\mathbf{f} \in \mathcal{C}^1(\mathbb{R}; \mathbb{R}^d)$. For the sake of simplicity we assume that there are no issues with the boundary conditions. For instance, either we assume periodic boundary conditions or the initial data is compactly supported, and we are interested in the solution before the domain of dependence of u_0 reaches the boundary of Ω . We assume that (2.1) has a unique entropy solution satisfying the additional entropy inequalities $\partial_t E(u) + \nabla \cdot \mathbf{F}(u) \leq 0$ for all convex entropy $E \in \mathcal{C}^2(\mathbb{R}; \mathbb{R})$ and associated entropy flux $\mathbf{F} \in \mathcal{C}^2(\mathbb{R}; \mathbb{R})$ with $\mathbf{F}'_i(u) = E'(u)\mathbf{f}_i(u)$, $1 \leq i \leq d$.

In order to approximate the entropy solution of (2.1) with H^1 -conforming finite elements, we consider a mesh family $\{\mathcal{K}_h\}_{h>0}$ that we assume to be conforming (no hanging nodes) and shape-regular in the sense of Ciarlet. By convention, the elements in $\{\mathcal{K}_h\}_{h>0}$ are closed in \mathbb{R}^d . The reference element is denoted \widehat{K} , and the map between \widehat{K} and an arbitrary element $K \in \mathcal{K}_h$ is denoted $\Phi_K : \widehat{K} \rightarrow K$. We define the scalar-valued finite element approximation space

$$(2.2) \quad X_h = \{v \in \mathcal{C}^0(\Omega; \mathbb{R}); v|_{K \circ \Phi_K} \in \mathbb{P}_k \ \forall K \in \mathcal{K}_h\},$$

where $k \in \mathbb{N} \setminus \{0\}$ and \mathbb{P}_k denotes the set of multivariate polynomials of total degree at most k . For all $K \in \mathcal{K}_h$, h_K denotes the diameter of K divided by k . The purpose of the normalization $h_K = \text{diam}K/k$ is to make the parameters c_{ev} and c_{ed} introduced below for the entropy viscosity and edge stabilization independent of k . The quantity $h := \max_{K \in \mathcal{K}_h} h_K$ is the so-called meshsize.

Let $t_F > 0$ be the final simulation time. We call the Galerkin solution of (2.1) the function $u_h \in \mathcal{C}^1([0, t_F]; X_h)$ such that

$$(2.3) \quad \int_{\Omega} v \partial_t u_h \, d\Omega + \int_{\Omega} v \nabla \cdot \mathbf{f}(u_h) \, d\Omega = 0$$

for all $v \in X_h$ and all $t \in (0, t_F)$, and $u_h(0) = u_{0,h}$, where $u_{0,h}$ is an appropriate approximation of u_0 in X_h . In all the numerical tests reported herein, the time stepping is done by using either the strong stability-preserving third-order Runge–Kutta (SSP RK3) method, see, e.g., [15], or the standard fourth-order Runge–Kutta (RK4) method. The Butcher tableaux of both methods are reported in the appendix for the reader’s convenience; see (A.2). To avoid mixing the space and time discretization errors, we always perform our tests with a small CFL, say $\text{CFL} = 0.2$ or even less. We finally emphasize that the mass matrix is not lumped.

The Galerkin solution is known to be a poor approximation of the solution to (2.1) even when the flux is linear. One simple device to stabilize the approximation is to add first-order viscous dissipation. We call the viscous solution the function $u_h \in \mathcal{C}^1([0, t_F]; X_h)$ such that $u_h(0) = u_{0,h}$ and

$$(2.4) \quad \int_{\Omega} v \partial_t u_h \, d\Omega + \int_{\Omega} v \nabla \cdot \mathbf{f}(u_h) \, d\Omega + n_{\text{visc}}(u_h; v) = 0,$$

$$(2.5) \quad n_{\text{visc}}(w; v) := c_{\max} \sum_{K \in \mathcal{K}_h} h_K \|\mathbf{f}'(w)\|_{L^\infty(K)} \int_K \nabla w \cdot \nabla v \, dK$$

for all $v \in X_h$ and all $t \in (0, t_F)$. We take $c_{\max} = \frac{1}{2k}$ in one space dimension and $c_{\max} = \frac{1}{4k}$ in two space dimensions on triangular meshes. Note that $n_{\text{visc}}(w; v)$ is linear in w for linear transport since \mathbf{f}' is constant in this case. When using finite differences on one-dimensional uniform grids, setting $c_{\max} = \frac{1}{2}$ corresponds to replacing the centered differences by first-order upwind differences, and the resulting scheme is known to be monotone.

It is well-known that the viscous solution is only first-order accurate. The performance of the method can be greatly improved by substituting the first-order viscous dissipation by some linear or nonlinear stabilization mechanism. We are going to use in this paper the so-called entropy viscosity [18, 22, 17] as a nonlinear stabilization mechanism. We call the entropy viscosity solution the function $u_h \in \mathcal{C}^1([0, t_F]; X_h)$ such that $u_h(0) = u_{0,h}$ and

$$(2.6) \quad \int_{\Omega} v \partial_t u_h \, d\Omega + \int_{\Omega} v \nabla \cdot \mathbf{f}(u_h) \, d\Omega + n_{\text{ev}}(u_h; v) = 0,$$

$$(2.7) \quad n_{\text{ev}}(w; v) := \sum_{K \in \mathcal{K}_h} \nu_K(w) \int_K \nabla w \cdot \nabla v \, dK$$

for all $v \in X_h$ and all $t \in (0, t_F)$, where $\nu_K(\cdot) : \mathcal{C}^1([0, t_F]; X_h) \rightarrow \mathbb{P}_0(K)$ is a nonlinear viscosity functional defined as follows:

$$(2.8) \quad \nu_K(w) = \min(c_{\max} h_K \|\mathbf{f}'(w)\|_{L^\infty(K)}, c_{\text{ev}} h_K R_K(w)).$$

The functional $R_K(\cdot)$ is the so-called entropy residual and is defined in the appendix. Typically, the first argument of the minimum in the right-hand side is active near discontinuities, while the second argument prevails in smooth regions. Note that, unlike $n_{\text{visc}}(w; v)$, $n_{\text{ev}}(w; v)$ is nonlinear in w even for linear transport. In the appendix, we detail the construction of the nonlinear viscosity in the context of an explicit time-stepping Runge–Kutta scheme, where the entropy residual R_K has a more complex

dependency than just on w ; we keep the present notation to fix the general ideas at this stage. The entropy viscosity solution has reasonable convergence properties. For instance, the entropy viscosity solution has been numerically observed to satisfy a weakened maximum principle in the sense that for all $\epsilon > 0$, there is $h_0 > 0$ and there are uniform constants c and $\alpha > 0$ such that

$$(2.9) \quad \|u_h(t)\|_{L^\infty(\Omega)} \leq \|u_0\|_{L^\infty(\Omega)} + ch^\alpha \quad \forall h < h_0, \forall t > \epsilon.$$

This property is illustrated, for instance, in Table 2.1(a). No proof of this observation has been published yet.

When $\mathbf{f}(u) = \beta u$ with \mathbb{R}^d -valued velocity field β , (2.1) reduces to the linear transport equation. Many linear (symmetric) stabilization techniques can be considered for solving the linear transport equation, e.g., subgrid viscosity [20], edge stabilization [8, 7, 4], and discontinuous Galerkin [25, 14, 23, 12]. All these methods are known to yield the near optimal convergence rate $\mathcal{O}(h^{k+\frac{1}{2}})$ in the L^2 -norm for smooth solutions. We focus herein on the edge stabilization technique, which we think is relatively simple to implement with H^1 -conforming finite elements. Let \mathcal{F}_h^i be the set of all mesh interfaces. By convention, interfaces are closed in \mathbb{R}^{d-1} . For all $F \in \mathcal{F}_h^i$, we denote h_F the diameter of F . Denoting $K_{1,F} \in \mathcal{K}_h$, $K_{2,F} \in \mathcal{K}_h$ the two (closed) cells so that $F = K_{1,F} \cap K_{2,F}$, we set $\Delta_F = K_{1,F} \cup K_{2,F}$ (see Figure 3.1). Moreover, letting v be a scalar-valued function defined over Δ_F and continuous over $K_{1,F}$ and $K_{2,F}$, we denote $\{v\}(x) = \frac{1}{2}(v|_{K_{1,F}}(x) + v|_{K_{2,F}}(x))$ the average of v at $x \in F$ and $\{\partial_n w\}$ the jump of the normal gradient across F , where ∂_n is the outward normal derivative. We call the edge stabilized solution the function $u_h \in \mathcal{C}^1([0, t_F]; X_h)$ such that $u_h(0) = u_{0,h}$ and

$$(2.10) \quad \int_{\Omega} v \partial_t u_h \, d\Omega + \int_{\Omega} v \nabla \cdot \mathbf{f}(u_h) \, d\Omega + n_{\text{ed}}(u_h; v) = 0,$$

$$(2.11) \quad n_{\text{ed}}(w; v) = c_{\text{ed}} \sum_{F \in \mathcal{F}_h^i} h_F^2 \| \mathbf{f}'(w) \|_{L^\infty(F)} \int_F \{ \partial_n w \} \{ \partial_n v \} \, dF$$

for all $v \in X_h$ and all $t \in (0, t_F)$. Note that $n_{\text{ed}}(w; v)$ is linear in w for linear transport. We take $c_{\text{ed}} = 0.05$ in our numerical experiments, unless stated otherwise.

Finally, we call the entropy viscosity solution augmented with edge stabilization the function $u_h \in \mathcal{C}^1([0, t_F]; X_h)$ such that $u_h(0) = u_{0,h}$ and

$$(2.12) \quad \int_{\Omega} v \partial_t u_h \, d\Omega + \int_{\Omega} v \nabla \cdot \mathbf{f}(u_h) \, d\Omega + n_{\text{ev}}(u_h; v) + n_{\text{ed}}(u_h; v) = 0$$

for all $v \in X_h$ and all $t \in (0, t_F)$.

The objective of this paper is to investigate some aspects of the stabilization properties of the semilinear form $n_{\text{ed}}(\cdot; \cdot)$. We want to show that, although this stabilization technique performs extremely well in smooth regions, it has counterproductive effects in regions of shocks and large gradients. The purpose of the rest of this section is to illustrate some of the negative effects of $n_{\text{ed}}(\cdot; \cdot)$ through numerical experiments. A weighting technique that cures these problems is proposed and analyzed in section 3.

2.2. One-dimensional transport. We consider the one-dimensional transport problem

$$(2.13) \quad \partial_t u + \partial_x u = 0, \quad u(x, 0) = \chi_{[0.4, 0.7]}, \quad t > 0, \quad x \in \Omega = (0, 1),$$

with periodic boundary conditions and where $\chi_{[0.4, 0.7]}$ is the characteristic function

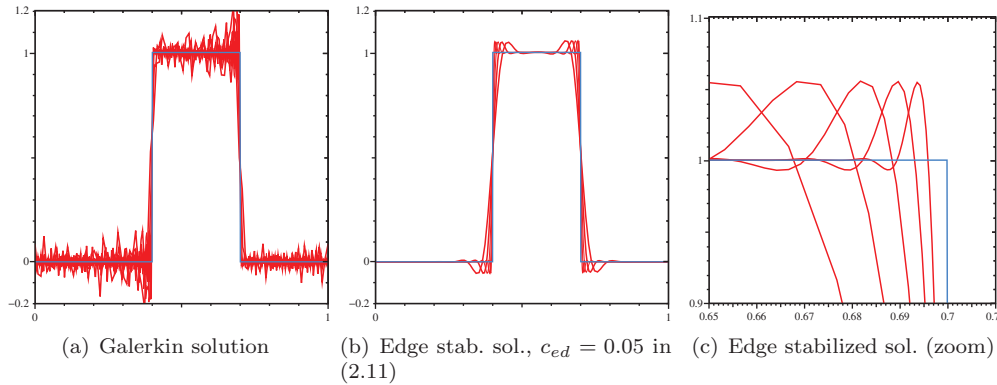


FIG. 2.1. One-dimensional linear transport. Results at $t_F = 1$ without and with edge stabilization on five uniform meshes composed of $100, 2 \times 100, \dots, 2^4 \times 100$ cells.

of $[0.4, 0.7]$. We compute the solution at $t_F = 1$ with continuous \mathbb{P}_1 finite elements on various uniform meshes.

2.2.1. Edge stabilization alone. We first compute the Galerkin solution on five uniform meshes composed of $100, 2 \times 100, \dots, 2^4 \times 100$ cells and we set $\text{CFL} = 0.2$. The graphs of the solutions are shown in Figure 2.1(a). We observe the familiar spurious oscillations that characterize the Galerkin technique. The graphs of the solutions computed with edge stabilization are shown in Figure 2.1(b). This test exemplifies at the same time the stabilizing capability of the edge stabilization and its inability to counter the so-called Gibbs phenomenon triggered in the vicinity of discontinuities and large gradients. Figure 2.1(c) displays details of the graphs of the solutions in the region $x \in [0.65, 0.71]$. Additional tests on eight refinement levels (not reported here) show that both indicators e_{Max} and e_{Min} (defined in (2.14) below) are bounded away from 0, i.e., mesh refinement does not help to satisfy the maximum principle.

2.2.2. Edge stabilization plus entropy viscosity. It is frequently advocated in the literature that linear stabilization must be supplemented with a shock capturing technique to handle properly shocks and large gradients [24, 26, 10, 5, 16]. We want to investigate the effects of combining the edge stabilization with the entropy viscosity.

We perform the following tests. We compute the entropy viscosity solution and the entropy viscosity solution augmented with edge stabilization for the one-dimensional transport equation (2.13) using \mathbb{P}_1 finite elements on various uniform meshes. To assess possible violations of the maximum principle, we compute the following indicators:

$$(2.14) \quad e_{\text{Max}} := \max_{x \in \Omega} u_h(x, 1) - 1, \quad e_{\text{Min}} := -\min_{x \in \Omega} u_h(x, 1).$$

The results for the entropy viscosity solution and the entropy viscosity solution augmented with edge stabilization are reported in Table 2.1(a) and Table 2.1(b), respectively. We observe that both indicators e_{Max} and e_{Min} for the entropy viscosity solution converge to zero with the meshsize, in agreement with the claim made in section 2.1 that the entropy viscosity solution satisfies a weakened maximum principle in the form (2.9). On the other hand, we observe in Table 2.1(b) that the weakened maximum principle is lost when edge stabilization is added to the entropy viscosity.

TABLE 2.1

One-dimensional transport. Maximum and minimum error indicators for entropy viscosity solution (left) and entropy viscosity + edge stabilized solution (right); \mathbb{P}_1 finite elements, $t_F = 1$.

(a) Entropy viscosity					(b) Entropy viscosity + edge stabilization			
h	e_{Min}	Rate	e_{Max}	Rate	e_{Min}	Rate	e_{Max}	Rate
2.500E-03	6.725E-03	—	6.715E-03	—	1.597E-02	—	1.597E-02	—
1.250E-03	5.441E-03	0.306	5.434E-03	0.305	1.600E-02	-0.003	1.600E-02	-0.003
6.250E-04	2.855E-03	0.930	2.854E-03	0.929	1.633E-02	-0.030	1.633E-02	-0.030
3.125E-04	2.235E-03	0.353	2.235E-03	0.353	1.626E-02	0.006	1.626E-02	0.006
1.563E-04	1.785E-03	0.324	1.785E-03	0.324	1.646E-02	-0.017	1.646E-02	-0.017

Thus, by adding edge stabilization to a method that satisfies a weakened maximum principle, one can obtain a method that does not satisfy the maximum principle, even in the weak sense defined above.

The above problem can be easily fixed by increasing the strength of the entropy viscosity. For instance, the weakened maximum principle can be recovered by using h_K^γ with $\gamma = \frac{1}{2}$ instead of h_K in the second argument of the minimum defining $\nu_K(w)$ in (2.8). The $\gamma = \frac{1}{2}$ fix is marginally satisfactory since it makes the method more dissipative and deteriorates its convergence properties. For instance, convergence tests (not shown here) on the one-dimensional transport problem (2.13) with \mathbb{P}_1 finite elements reveal that the convergence rate of the entropy viscosity method in the L^1 -norm is $\frac{3}{4}$ with $\gamma = 1$ and $\frac{2}{3}$ with $\gamma = \frac{1}{2}$.

2.3. One-dimensional nonconvex conservation equation. To test the edge stabilization technique with nonlinear conservation equations, we consider a problem with nonconvex flux proposed in [28]. We restrict ourselves to the one-dimensional domain $\Omega = (0, 1)$ and consider the following scalar flux and initial data:

$$(2.15) \quad f(u) = \begin{cases} \frac{1}{4}u(1-u) & \text{if } u < \frac{1}{2}, \\ \frac{1}{2}u(u-1) + \frac{3}{16} & \text{if } \frac{1}{2} \leq u, \end{cases} \quad u_0(x) = \begin{cases} 0, & x \in [0, 0.35], \\ 1, & x \in (0.35, 1]. \end{cases}$$

The entropy solution to this problem is a composite wave composed of a shock followed by a rarefaction wave. This problem is challenging since many second-order central schemes with compressive limiters are known to converge to entropy-violating weak solutions. For instance, it is demonstrated in [28] that the so-called central-upwind scheme using second-order piecewise linear reconstruction with either the superbee limiter or the so-called minmod2 limiter fails to converge to the entropy solution. We compute the solution at time $t_F = 1$ with continuous \mathbb{P}_1 finite elements. The solution at $t_F = 1$ is composed of a shock wave located at $x_s(1) = \frac{1}{4}\sqrt{6} - 0.15$ followed by a rarefaction wave. The left limit of the solution at the shock is $u_s^-(t_F) = 0$ and the right limit is $u_s^+(t_F) = \sqrt{\frac{3}{8}}$.

2.3.1. Edge stabilization alone. We show in Figure 2.2 the Galerkin and the edge stabilized solutions obtained at $t_F = 1$ on a uniform mesh composed of 1000 cells. We use the SSP RK3 scheme, and to avoid time discretization errors, the time step size is based on $\text{CFL} = 0.01$. The solution shown in the left panel of Figure 2.2 is the Galerkin solution and that shown in the right panel is the edge stabilized solution. The entropy solution is shown in dashed line. It is clear that none of these approximations converge to the entropy solution in any possible norm.

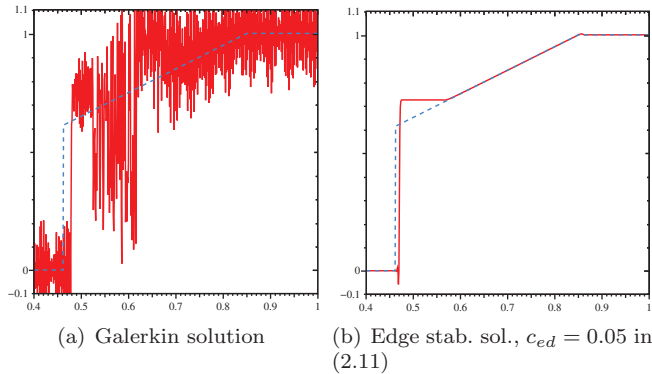


FIG. 2.2. *Nonconvex flux. Results at $t_F = 1$ on a uniform mesh composed of 1000 cells and $CFL = 0.01$.*

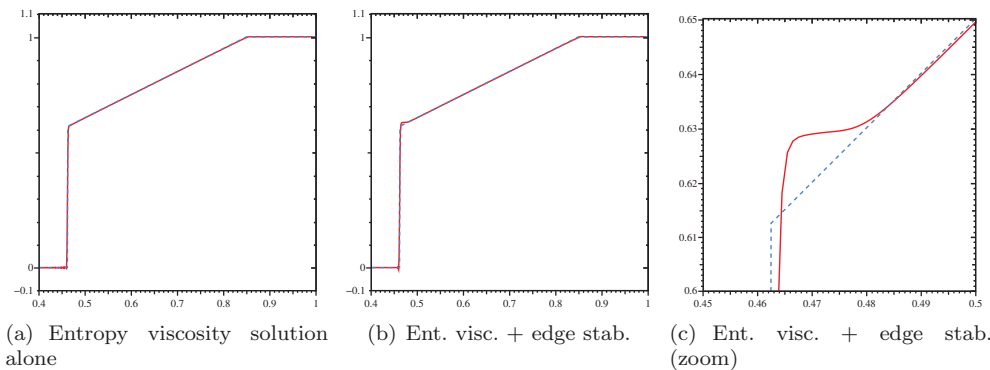


FIG. 2.3. *Nonconvex flux. Results at $t_F = 1$ with entropy viscosity without edge stabilization (left) and with edge stabilization on a uniform mesh composed of 1000 cells (center and right).*

The edge stabilized solution is almost free of spurious oscillations but converges to an entropy-violating weak solution.

2.3.2. Edge stabilization plus entropy viscosity. We show in Figure 2.3 the entropy viscosity solution and the entropy viscosity solution augmented with edge stabilization at $t_F = 1$ on a uniform mesh composed of 1000 cells. The entropy viscosity solution is shown in the left panel and the entropy viscosity solution with edge stabilization is shown in the center and right panels. The entropy solution is shown in dashed line. These tests show that the entropy viscosity solution converges to the entropy solution, whereas the entropy viscosity solution augmented with edge stabilization does not. Thus, by adding edge stabilization to a method that converges to the correct weak solution, one can obtain a method that converges to a wrong weak solution. This result is similar to what has been observed in [28] concerning the second-order piecewise linear reconstruction combined with either the superbee or the minmod2 limiter.

2.4. Edge stabilization plus first-order viscosity. In this section, we present two numerical examples showing that the edge stabilization can have adverse effects even when it is combined with the first-order viscosity method.

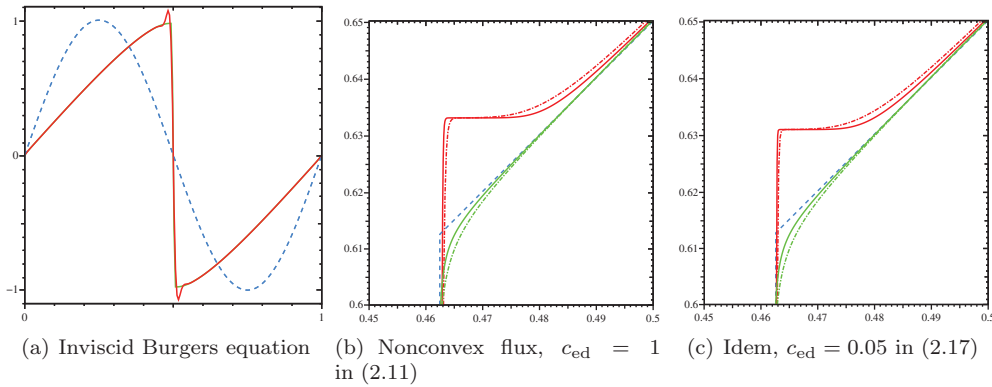


FIG. 2.4. Left: One-dimensional inviscid Burgers equation, $t_F = 0.25$, $CFL = 0.025$, uniform mesh composed of 200 cells; first-order viscous solution without (in green) and with edge stabilization (in red). Center and right: Nonconvex flux, $t_F = 1$, $CFL = 0.025$, uniform meshes composed of 4000 (dot-dash line) and 10,000 cells (solid line); the first-order viscous solution with edge stabilization (red) and the viscous solution (green) are shown; Center: $c_{ed} = 1$ in (2.11); right, $c_{ed} = 0.05$ in (2.17). The exact solution (blue) is shown in dashed line.

We first consider the one-dimensional inviscid Burgers equation

$$(2.16) \quad \partial_t u + \partial_x \left(\frac{1}{2} u^2 \right) = 0, \quad u(x, 0) = \sin(2\pi x),$$

over the interval $\Omega = (0, 1)$ with periodic boundary conditions. The solution is computed at $t_F = 0.25$ with continuous \mathbb{P}_1 finite elements on a mesh composed of 200 elements using SSP RK3 time stepping with $CFL = 0.025$. We compute the first-order viscous solution and the first-order viscous solution augmented with edge stabilization (whereby $n_{ev}(u_h; v)$ is replaced by $n_{visc}(u_h; v)$ in (2.12)). The coefficient c_{ed} in (2.11) is set to 1 and the coefficient c_{max} in (2.5) to $\frac{1}{2}$. The graph of the solution is shown in Figure 2.4(a). We observe overshoots and undershoots for the solution with edge stabilization, and the amplitude of these spurious features is constant as the mesh is refined. This is characteristic of the Gibbs phenomenon. These overshoots and undershoots can be tamed by increasing the viscous dissipation beyond what should normally be necessary. Numerical tests (not shown) reveal that $c_{max} = 2$ is the lower bound on c_{max} that makes the viscous dissipation strong enough to overcome the Gibbs phenomenon. Note however that increasing c_{max} from 0.5 to 2 requires decreasing the CFL fourfold in addition to increasing fourfold the numerical dissipation. These results indicate that edge stabilization (and possibly linear stabilization at large) tends to promote the Gibbs phenomenon.

The adverse effects of edge stabilization are even more dramatic on conservation equations with nonconvex flux. We consider again the test case of section 2.3. We show in Figure 2.4(b) a zoom of the graph of the solutions obtained with first-order viscosity ($c_{max} = \frac{1}{2}$) without edge stabilization and with edge stabilization ($c_{ed} = 1$) computed on two uniform meshes composed of 4000 and 10,000 cells. We observe that the viscous solution converges to the entropy solution (as expected), whereas the edge stabilized solution converges to a plateau between the expansion wave and the shock, which is wrong. The same effect is observed in Figure 2.4(c) with the choice $c_{ed} = 0.05$ in

$$(2.17) \quad n_{ed}(w; v) = c_{ed} \| \mathbf{f}'(w) \|_{L^\infty(\Omega)} \sum_{F \in \mathcal{F}_h^i} h_F^2 \int_F \{ \partial_n w \} \{ \partial_n v \} dF,$$

where the normalization is defined globally by using $\|\mathbf{f}'(w)\|_{L^\infty(\Omega)}$ instead of $\|\mathbf{f}'(w)\|_{L^\infty(F)}$. We have verified that the above results do not depend on the way the initial data is computed. Whether the initial data is the L^2 -projection of u_0 , the Lagrange interpolant thereof, or a smoothed version of the Lagrange interpolant (using some sliding averaging), the behaviors reported above are identical.

2.5. Conclusions from numerical tests. The first series of tests presented in section 2.2.1 shows that the linear edge stabilization does a great job at suppressing spurious oscillations in the regions where the solution is smooth, but this technique cannot get rid of the Gibbs phenomenon. The second series of tests reported in section 2.2.2 shows that the linear edge stabilization actually promotes the Gibbs phenomenon. The third series of tests on the one-dimensional nonlinear scalar conservation equation with a nonconvex flux in section 2.3.1 demonstrates that the linear edge stabilization again does a great job at removing the spurious oscillations plaguing the Galerkin solution, but it does not have the correct type of dissipation to make the approximate solution converge to the entropy solution. Finally, the tests reported in section 2.3.2 and section 2.4 show that not only does the linear edge stabilization not produce the right dissipation, but the type of dissipation that it produces can transform a convergent method (either first-order linear viscosity or nonlinear entropy viscosity) into a nonconvergent one. The authors conjecture that the above conclusions are not restricted to edge stabilization but can be extended to some (or most of the) other linear stabilization methods available in the literature.

3. Weighting the edge stabilization. We introduce in this section a weighting technique for the edge stabilization, and we prove that its convergence properties on the linear transport problem are identical to that of the original unweighted edge stabilization in the case of smooth solutions. We consider the linear transport equation

$$(3.1) \quad \partial_t u + \nabla \cdot (\boldsymbol{\beta} u) = 0, \quad u(x, 0) = u_0(x), \quad (x, t) \in \Omega \times \mathbb{R}_+,$$

in space dimension $d = 2$ or $d = 3$. We assume that $\boldsymbol{\beta}$ is Lipschitz and divergence-free in $\overline{\Omega}$. Recall that for simplicity we assume either periodic boundary conditions, compactly supported solutions, or $\boldsymbol{\beta} \cdot \mathbf{n}|_{\partial\Omega} = 0$.

In what follows, $a \lesssim b$ means that inequality $a \leq cb$ holds with a constant c independent of h (but possibly depending on the mesh-regularity, the polynomial degree k , and the regularity of the problem data and the exact solution). Without loss of generality, we assume $h \leq 1$. For any set $R \subset \Omega$ (a mesh element, a mesh face, or a collection thereof), we denote by $\|\cdot\|_{L^p(R)}$ the usual $L^p(R)$ -norm, $1 \leq p \leq \infty$, for scalar- or vector-valued functions.

3.1. Principle of the method. Consider the following discrete solution $u_h \in \mathcal{C}^1([0, t_F]; X_h)$ with X_h defined in (2.2), so that $u_h(t = 0) = u_{0,h}$ and

$$(3.2) \quad \int_{\Omega} v \partial_t u_h \, d\Omega + \int_{\Omega} v \nabla \cdot (\boldsymbol{\beta} u_h) \, d\Omega + n_{\text{lim,ed}}(u_h; u_h, v) = 0$$

for all $v \in X_h$ and all $t \in (0, t_F)$. The weighted edge stabilization semilinear form is defined by

$$(3.3) \quad n_{\text{lim,ed}}(z; w, v) = c_{\text{ed}} \sum_{F \in \mathcal{F}_h^i} \alpha(g_F(z)) h_F^2 |\boldsymbol{\beta}|_F \int_F \{\partial_n w\} \{\partial_n v\} \, dF$$

with the shorthand notation $|\beta|_F := \|\beta\|_{L^\infty(F)}$, and $g_F(z)$ is a measure of the gradient of z around F which we take in the form

$$(3.4) \quad g_F(z) = \ell^{-1} |\langle \nabla z \rangle_{\Delta_F}|,$$

where $\langle \phi \rangle_R := \text{meas}(R)^{-1} \int_R \phi \, dR$ denotes the average of a function ϕ over a set $R \subset \Omega$ and where the global scaling parameter ℓ is, e.g., set to $\ell := |\langle \nabla u_{0,h} \rangle_\Omega|$. The key ingredient in (3.3) is the function $\alpha : \mathbb{R}_+ \rightarrow (0, 1]$ which weights the amount of edge stabilization. The function α must be such that

- (i) α is nonincreasing;
- (ii) there is $\alpha_0 \in \mathbb{R}_+$ and $\lambda \in \mathbb{R}_+$ such that for all $r \geq 1$, $\alpha(r) \geq \alpha_0 r^{-\lambda}$.

Typically, $\alpha(r) \rightarrow 0$ as $r \rightarrow +\infty$ so as to turn off edge stabilization in regions of large gradients. Condition (ii) then means that α must not decrease too quickly, so as to retain the optimal convergence properties of the method for smooth solutions.

3.2. Convergence analysis. We first prove a convergence result in two dimensions. In this situation, there is no restriction on the parameter λ controlling the decrease of the weighting function α at infinity.

THEOREM 3.1. *Let u and u_h be the solutions to (3.1) and to (3.2), respectively. Assume that $u \in C^1([0, t_F]; H^{k+1}(\Omega)) \cap C^0([0, t_F]; W^{k+1,\infty}(\Omega))$. Assume that the mesh family $\{\mathcal{K}_h\}_{h>0}$ is quasi-uniform, $d = 2$, and $\lambda > 0$. Then, the following holds for all $t \in [0, t_F]$:*

$$(3.5) \quad \|(u - u_h)(\cdot, t)\|_{L^2(\Omega)} + \left(\int_0^t n_{\text{lim,ed}}(u_h; u_h, u_h) \, d\tau \right)^{\frac{1}{2}} \lesssim h^{k+\frac{1}{2}}.$$

Proof. The proof is decomposed into three steps.

Step 1: Error equation. For all $t \in [0, t_F]$, let $w(\cdot, t)$ be the L^2 -orthogonal projection of the exact solution $u(\cdot, t)$ onto the discrete space X_h . (In the case of compactly supported solutions, we project onto $X_h \cap H_0^1(\Omega)$.) We define the quantities (the dependence with respect to t is now left implicit)

$$e := u_h - w, \quad \eta := u - w,$$

so that the approximation error is $u_h - u = e - \eta$. Observing that

$$\int_\Omega v \partial_t w \, d\Omega + \int_\Omega v \nabla \cdot (\beta w) \, d\Omega = - \int_\Omega v \partial_t \eta \, d\Omega - \int_\Omega v \nabla \cdot (\beta \eta) \, d\Omega \quad \forall v \in X_h$$

and subtracting this equation from (3.2), we infer that the following holds for all $v \in X_h$, $t \in [0, t_F]$:

$$(3.6) \quad \begin{aligned} & \int_\Omega v \partial_t e \, d\Omega + \int_\Omega v \nabla \cdot (\beta e) \, d\Omega + n_{\text{lim,ed}}(u_h; e, v) \\ &= \int_\Omega v \partial_t \eta \, d\Omega + \int_\Omega v \nabla \cdot (\beta \eta) \, d\Omega + n_{\text{lim,ed}}(u_h; \eta, v) \\ &=: \mathfrak{T}_1(\eta, v) + \mathfrak{T}_2(\eta, v) + \mathfrak{T}_3(u_h; \eta, v), \end{aligned}$$

where we have used the fact that $\{\partial_n u\} = 0$ for all $F \in \mathcal{F}_h^i$ for the exact solution u , so that

$$\begin{aligned} n_{\text{lim,ed}}(u_h; u_h, v) &= n_{\text{lim,ed}}(u_h; e, v) - n_{\text{lim,ed}}(u_h; \eta, v) + n_{\text{lim,ed}}(u_h; u, v) \\ &= n_{\text{lim,ed}}(u_h; e, v) - n_{\text{lim,ed}}(u_h; \eta, v). \end{aligned}$$

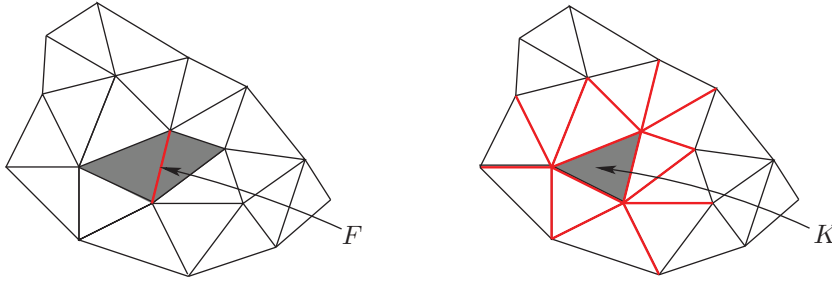


FIG. 3.1. Definition of Δ_F (grey triangles). FIG. 3.2. Definition of \mathfrak{F}_K (red thick lines).

Step 2: Basic estimates. Testing (3.6) with $v = e$ and using the conservativity property

$$\int_{\Omega} e \nabla \cdot (\beta e) \, d\Omega = 0,$$

which holds owing to the choice of boundary conditions and the fact that β is divergence-free, we infer that

$$\frac{1}{2} \frac{d}{dt} \|e\|_{L^2(\Omega)}^2 + n_{\text{lim,ed}}(u_h; e, e) \leq |\mathfrak{T}_1(\eta, e)| + |\mathfrak{T}_2(\eta, e)| + |\mathfrak{T}_3(u_h; \eta, e)|.$$

Moreover, since $u \in C^1([0, t_F]; H^{k+1}(\Omega))$, classical finite element interpolation properties [3, 13] yield

$$(3.7) \quad h_K \|\nabla \eta\|_{L^2(K)} + \|\eta\|_{L^2(K)} + \|\partial_t \eta\|_{L^2(K)} \lesssim h_K^{k+1} \quad \forall K \in \mathcal{K}_h.$$

Hence, by the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} |\mathfrak{T}_1(\eta, e)| &\lesssim h^{k+1} \|e\|_{L^2(\Omega)}, & |\mathfrak{T}_2(\eta, e)| &\lesssim h^k \|e\|_{L^2(\Omega)}, \\ |\mathfrak{T}_3(u_h; \eta, e)| &\leq n_{\text{lim,ed}}(u_h; e, e)^{\frac{1}{2}} n_{\text{lim,ed}}(u_h; \eta, \eta)^{\frac{1}{2}} \lesssim h^{k+\frac{1}{2}} n_{\text{lim,ed}}(u_h; e, e)^{\frac{1}{2}}, \end{aligned}$$

where the last bound results from the fact that $n_{\text{lim,ed}}(u_h; \eta, \eta)^{\frac{1}{2}} \leq n_{\text{ed}}(\eta, \eta)^{\frac{1}{2}}$ (since $\alpha_F(u_h) \leq 1$) and the classical bound $n_{\text{ed}}(\eta, \eta)^{\frac{1}{2}} \lesssim h^{k+\frac{1}{2}}$. Collecting the above estimates, using Young’s inequality, and recalling that $h \leq 1$, we arrive at

$$\frac{d}{dt} \|e\|_{L^2(\Omega)}^2 + n_{\text{lim,ed}}(u_h; e, e) \lesssim h^k \|e\|_{L^2(\Omega)} + h^{2k+1},$$

whence by using Gronwall’s lemma together with $h \leq 1$, we obtain the suboptimal estimate

$$(3.8) \quad \forall t \in [0, t_F], \quad \|e(\cdot, t)\|_{L^2(\Omega)} + \left(\int_0^t n_{\text{lim,ed}}(u_h; e, e) \, d\tau \right)^{\frac{1}{2}} \lesssim h^k.$$

Step 3: Improved estimate on $\mathfrak{T}_2(\eta, e)$. For any $K \in \mathcal{K}_h$, we define \mathfrak{F}_K to be the collection of all the mesh interfaces having a nonempty intersection with K (see Figures 3.1 and 3.2). We fix a time $t \in (0, t_F)$, we let $\epsilon \geq 0$ and $c_0 \geq 1$ (the value of these quantities is chosen later on), and we define the sets

$$\begin{aligned} \mathcal{F}_h^\sharp &:= \{F \in \mathcal{F}_h^i; g_F(u_h) \geq c_0 h^{-\epsilon}\}, \\ \mathcal{K}_h^\flat &:= \{K \in \mathcal{K}_h; \forall F \in \mathfrak{F}_K, F \notin \mathcal{F}_h^\sharp\}, \\ \mathcal{K}_h^\sharp &:= \{K \in \mathcal{K}_h; \exists F \in \mathfrak{F}_K, F \in \mathcal{F}_h^\sharp\} = \mathcal{K}_h \setminus \mathcal{K}_h^\flat. \end{aligned}$$

Let β_h be the continuous, piecewise affine interpolant of β on the mesh \mathcal{K}_h and set $z_h := \beta_h \cdot \nabla e$. Observe that z_h is a piecewise polynomial of degree $\leq k$, but it does not belong to X_h because it can jump across interfaces. We define $\mathcal{I}_{\text{av}}(z_h) \in X_h$ to be so that its value at any Lagrange node is equal to the arithmetic average of the values taken by z_h from the mesh elements sharing that node. It is well known, see [1, 27, 6], that this type of averaging leads to the estimate, for all $K \in \mathcal{K}_h$,

$$(3.9) \quad \|z_h - \mathcal{I}_{\text{av}}(z_h)\|_{L^2(K)} \lesssim \sum_{F \in \mathfrak{F}_K} h_F^{\frac{1}{2}} \|\llbracket z_h \rrbracket\|_{L^2(F)} \leq \sum_{F \in \mathfrak{F}_K} h_F^{\frac{1}{2}} |\beta|_F \|\{\partial_n e\}\|_{L^2(F)},$$

where $\llbracket \cdot \rrbracket$ denotes the jump across F and where we have used the continuity of β_h across F . Integrating by parts, accounting for boundary conditions, and using the fact that η is L^2 -orthogonal to X_h , $\mathfrak{T}_2(\eta, e)$ can be decomposed as

$$\begin{aligned} \mathfrak{T}_2(\eta, e) &= - \int_{\Omega} \eta(\beta \cdot \nabla e) \, d\Omega = - \int_{\Omega} \eta(\beta - \beta_h) \cdot \nabla e \, d\Omega - \int_{\Omega} \eta z_h \, d\Omega \\ &= - \int_{\Omega} \eta(\beta - \beta_h) \cdot \nabla e \, d\Omega - \int_{\Omega^{\flat}} \eta(z_h - \mathcal{I}_{\text{av}}(z_h)) \, d\Omega - \int_{\Omega^{\sharp}} \eta(z_h - \mathcal{I}_{\text{av}}(z_h)) \, d\Omega \\ &=: \mathfrak{T}_{2,0}(\eta, e) + \mathfrak{T}_{2,1}(\eta, e) + \mathfrak{T}_{2,2}(\eta, e), \end{aligned}$$

where $\Omega^{\flat} := \cup_{K \in \mathcal{K}_h^{\flat}} K$ and $\Omega^{\sharp} := \cup_{K \in \mathcal{K}_h^{\sharp}} K$. Since β is Lipschitz, the term $\mathfrak{T}_{2,0}(\eta, e)$ is readily bounded as follows:

$$|\mathfrak{T}_{2,0}(\eta, e)| \lesssim \|\eta\|_{L^2(\Omega)} h \|\nabla e\|_{L^2(\Omega)} \lesssim \|\eta\|_{L^2(\Omega)} \|e\|_{L^2(\Omega)} \lesssim h^{2k+1},$$

where we have used an inverse inequality and the estimate (3.8) on $\|e\|_{L^2(\Omega)}$. For the term $\mathfrak{T}_{2,1}(\eta, e)$, we use (3.9), the Cauchy–Schwarz inequality, $h_K \lesssim h_F$, and $|\beta|_F \lesssim 1$ to infer

$$\begin{aligned} |\mathfrak{T}_{2,1}(\eta, e)| &\lesssim \left(\sum_{K \in \mathcal{K}_h^{\flat}} [\alpha(g_K(u_h)) h_K]^{-1} \|\eta\|_{L^2(K)}^2 \right)^{\frac{1}{2}} \\ &\quad \left(\sum_{K \in \mathcal{K}_h^{\flat}} \alpha(g_K(u_h)) \sum_{F \in \mathfrak{F}_K} h_F^2 |\beta|_F \|\{\partial_n e\}\|_{L^2(F)}^2 \right)^{\frac{1}{2}} \end{aligned}$$

with $g_K(u_h) := \max_{F \in \mathfrak{F}_K} g_F(u_h)$. Since the function α is nonincreasing, $\alpha(g_K(u_h)) \leq \alpha(g_F(u_h))$ for all $F \in \mathfrak{F}_K$, so that this estimate implies

$$|\mathfrak{T}_{2,1}(\eta, e)| \lesssim \left(\sum_{K \in \mathcal{K}_h^{\flat}} [\alpha(g_K(u_h)) h_K]^{-1} \|\eta\|_{L^2(K)}^2 \right)^{\frac{1}{2}} n_{\text{lim,ed}}(u_h; e, e)^{\frac{1}{2}}.$$

We now use assumption (ii) on the function α to infer that for all $K \in \mathcal{K}_h^{\flat}$ and all $F \in \mathfrak{F}_K$, $\alpha(g_F(u_h)) \geq \alpha(c_0 h^{-\epsilon}) \geq \alpha_0 c_0^{-\lambda} h^{\epsilon \lambda}$ since $c_0 h^{-\epsilon} \geq 1$ (because $c_0 \geq 1$, $\epsilon \geq 0$, and $h \leq 1$). Thus, we obtain $\alpha(g_K(u_h)) \geq \alpha_0 c_0^{-\lambda} h^{\epsilon \lambda}$, whence

$$\left(\sum_{K \in \mathcal{K}_h^{\flat}} [\alpha(g_K(u_h)) h_K]^{-1} \|\eta\|_{L^2(K)}^2 \right)^{\frac{1}{2}} \lesssim h^{k + \frac{1}{2} - \frac{1}{2} \lambda \epsilon},$$

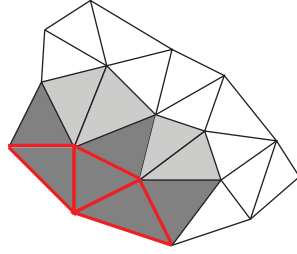


FIG. 3.3. The faces in thick red lines are in the set \mathcal{F}_h^\sharp ; the triangles filled in dark grey are in $\Omega^{\Delta\sharp}$; the triangles filled in dark or light grey are in Ω^\sharp (and thus belong to \mathcal{K}_h^\sharp); unfilled triangles belong to \mathcal{K}_h^\flat .

so that

$$|\mathfrak{T}_{2,1}(\eta, e)| \lesssim h^{k+\frac{1}{2}-\frac{1}{2}\lambda\epsilon} n_{\text{lim,ed}}(u_h; e, e)^{\frac{1}{2}}.$$

We now bound $\mathfrak{T}_{2,2}(\eta, e)$. Since $\|\eta\|_{L^\infty(\Omega)} \lesssim h^{k+1}$ owing to $u \in W^{k+1,\infty}(\Omega)$ and the approximation properties in $L^\infty(\Omega)$ of the L^2 -orthogonal projection for $d = 2$ and quasi-uniform meshes, see [11], we infer

$$|\mathfrak{T}_{2,2}(\eta, e)| \leq \|\eta\|_{L^\infty(\Omega^\sharp)} \text{meas}(\Omega^\sharp)^{\frac{1}{2}} \|z_h - \mathcal{I}_{\text{av}}(z_h)\|_{L^2(\Omega^\sharp)} \lesssim h^{2k} \text{meas}(\Omega^\sharp)^{\frac{1}{2}},$$

where we have used $\|z_h - \mathcal{I}_{\text{av}}(z_h)\|_{L^2(\Omega)} \lesssim \|z_h\|_{L^2(\Omega)} \lesssim \|\nabla e\|_{L^2(\Omega)} \lesssim h^{k-1}$. We now estimate $\text{meas}(\Omega^\sharp)$. This is achieved by proving that $\text{meas}(\Omega^\sharp) \lesssim \text{meas}(\Omega^{\Delta\sharp})$ and estimating $\text{meas}(\Omega^{\Delta\sharp})$. Here, $\Omega^{\Delta\sharp} := \cup_{F \in \mathcal{F}_h^\sharp} \Delta_F$, so that $\Omega^{\Delta\sharp}$ is the union of all the elements that have a face in \mathcal{F}_h^\sharp . Note that $\Omega^{\Delta\sharp} \subset \Omega^\sharp$; see Figure 3.3. Consider $K \in \mathcal{K}_h^\sharp$. By definition, there is $F \in \mathfrak{F}_K$ such that $F \in \mathcal{F}_h^\sharp$. By definition, also $\Delta_F \subset \Omega^{\Delta\sharp}$ and $K \cap \Delta_F \neq \emptyset$. (Recall that K and Δ_F are closed.) This means that there is $K' \in \Omega^{\Delta\sharp}$ such that $K \cap K'$ is nonempty, so that $\mathcal{K}_h^\sharp \subset \cup_{K' \in \Omega^{\Delta\sharp}} \{K'' \in \mathcal{K}_h; K' \cap K'' \neq \emptyset\}$. Hence, using mesh regularity, we infer

$$\text{meas}(\Omega^\sharp) \leq \sum_{K' \in \Omega^{\Delta\sharp}} \text{meas}(\{K'' \in \mathcal{K}_h; K' \cap K'' \neq \emptyset\}) \lesssim \sum_{K' \in \Omega^{\Delta\sharp}} \text{meas}(K') = \text{meas}(\Omega^{\Delta\sharp}).$$

The definition of \mathcal{F}_h^\sharp now implies that

$$\begin{aligned} \text{meas}(\Omega^{\Delta\sharp}) h^{-2\epsilon} &\leq \sum_{F \in \mathcal{F}_h^\sharp} \text{meas}(\Delta_F) h^{-2\epsilon} \leq (c_0 \ell)^{-2} \sum_{F \in \mathcal{F}_h^\sharp} \text{meas}(\Delta_F) |\langle \nabla u_h \rangle_{\Delta_F}|^2 \\ &\leq (c_0 \ell)^{-2} \sum_{F \in \mathcal{F}_h^\sharp} \int_{\Delta_F} |\nabla u_h|^2 \, d\Omega \leq 2(c_0 \ell)^{-2} \|\nabla u_h\|_{L^2(\Omega^{\Delta\sharp})}^2 \\ &\leq 6(c_0 \ell)^{-2} (\|\nabla e\|_{L^2(\Omega^{\Delta\sharp})}^2 + \|\nabla \eta\|_{L^2(\Omega^{\Delta\sharp})}^2 + \|\nabla u\|_{L^2(\Omega^{\Delta\sharp})}^2). \end{aligned}$$

Since $\|\nabla e\|_{L^2(\Omega^{\Delta\sharp})} \lesssim h^{k-1}$, $\|\nabla \eta\|_{L^2(\Omega^{\Delta\sharp})} \lesssim h^k$, and $\|\nabla u\|_{L^2(\Omega^{\Delta\sharp})} \leq \text{meas}(\Omega^{\Delta\sharp})^{\frac{1}{2}} \|\nabla u\|_{L^\infty(\Omega^{\Delta\sharp})}$, choosing $c_0 = \max(\sqrt{12} \ell^{-1} \|\nabla u\|_{L^\infty(\Omega \times (0, t_F))}, 1)$ allows us to hide $\|\nabla u\|_{L^2(\Omega^{\Delta\sharp})}^2$ in the left-hand side, so that $\text{meas}(\Omega^{\Delta\sharp}) \lesssim h^{2(k-1+\epsilon)}$, and hence

$$|\mathfrak{T}_{2,2}(\eta, e)| \lesssim h^{3k-1+\epsilon}.$$

Collecting the above estimates, we arrive at

$$\frac{d}{dt} \|e\|_{L^2(\Omega)}^2 + n_{\text{lim,ed}}(u_h; e, e) \lesssim h^{3k-1+\epsilon} + h^{2k+1-\lambda\epsilon}.$$

For $k \geq 2$, we can take $\epsilon = 0$ and obtain the desired estimate of order h^{2k+1} since $3k-1 \geq 2k+1$ in this case. In the case $k = 1$, the two terms in the right-hand side are balanced by choosing $\epsilon = \frac{1}{1+\lambda}$. This leads to the estimate (3.8) with the sharper bound $h^{1+\rho}$ with $\rho = \frac{1}{2}\epsilon$. We now use a bootstrap argument. Suppose that estimate (3.8) holds with bound $h^{1+\rho_n}$. Then, proceeding as above with a parameter ϵ_n to be chosen, the new bound for $\mathfrak{T}_{2,2}(\eta, e)$ becomes $h^{2+\epsilon_n+2\rho_n}$, while the contribution of $\mathfrak{T}_{2,1}(\eta, e)$ to the bound is still $h^{3-\lambda\epsilon_n}$. Balancing the two bounds leads to the choice $\epsilon_n = \frac{1}{1+\lambda}(1-2\rho_n)$, thus improving estimate (3.8) to $h^{1+\rho_{n+1}}$ with $\rho_{n+1} = \frac{1}{2}(1-\lambda\epsilon_n) = \frac{1}{2}(\frac{1}{1+\lambda} + \frac{2\lambda}{1+\lambda}\rho_n)$. This recursive relation shows that ρ_n converges to $\frac{1}{2}$ (for all λ), thereby completing the proof. \square

Remark 3.1. Theorem 3.1 holds in any space dimension under the assumption of locally quasi-uniform meshes provided the parameter $\delta = \sup_{h>0} \max_{K \in \mathcal{K}_h} \max_{K' \cap K \neq \emptyset} |1 - (h_{K'}/h_K)^2|$ is small enough, since the estimates $\|\eta\|_{L^\infty(\Omega)} \lesssim h^{k+1}$ and $\|h^{-\frac{1}{2}}\eta\|_{L^2(\Omega)} \lesssim h^{k+\frac{1}{2}}$ hold; see [2].

We now prove a convergence result in three dimensions (with a slightly less stringent regularity assumption on the exact solution). The main difference with the result of Theorem 3.1 is that the convergence result now requires that the parameter λ controlling the decrease of the weighting function α at infinity be not too large, at least for the lowest-order polynomials.

THEOREM 3.2. *Let u and u_h be the solutions to (3.1) and to (3.2), respectively. Assume that $u \in C^1([0, t_F]; H^{k+1}(\Omega)) \cap C^0([0, t_F]; W^{1,\infty}(\Omega))$. Assume that the mesh family $\{\mathcal{K}_h\}_{h>0}$ is quasi-uniform and $d = 3$. Then, the following holds for all $t \in [0, t_F]$:*

$$(3.10) \quad \begin{aligned} & \| (u - u_h)(\cdot, t) \|_{L^2(\Omega)} + \left(\int_0^t n_{\text{lim,ed}}(u_h; u_h, u_h) \, d\tau \right)^{\frac{1}{2}} \\ & \lesssim \begin{cases} h^{1+\frac{2-3\lambda}{4+\lambda}} & \text{if } k = 1 \text{ and } \lambda < \frac{2}{3}, \\ h^{2+\frac{2-\lambda}{4+\lambda}} & \text{if } k = 2 \text{ and } \lambda < 2, \\ h^{k+\frac{1}{2}} & \text{if } k \geq 3 \, \forall \lambda > 0. \end{cases} \end{aligned}$$

Proof. We proceed as in the proof of Theorem 3.1, up to the bound on $\mathfrak{T}_{2,2}(\eta, e)$ in Step 3 of the previous proof. Here, we no longer use approximation properties in $L^\infty(\Omega)$ of the L^2 -orthogonal projection. Instead, owing to the Sobolev embedding, we observe that there holds $\|\eta\|_{L^p(\Omega)} \lesssim h^k$ with $p = 6$. Then, $|\mathfrak{T}_{2,2}(\eta, e)| \lesssim h^{2k-1} \text{meas}(\Omega^\#)^{\frac{p-2}{2p}}$, and bounding $\text{meas}(\Omega^\#)$ as before yields

$$|\mathfrak{T}_{2,2}(\eta, e)| \lesssim h^{2k-1+\frac{p-2}{p}(k-1+\epsilon)}.$$

After collecting the above estimates, we arrive at

$$\frac{d}{dt} \|e\|_{L^2(\Omega)}^2 + n_{\text{lim,ed}}(u_h; e, e) \lesssim h^{2k-1+\frac{p-2}{p}(k-1+\epsilon)} + h^{2k+1-\lambda\epsilon}.$$

For $k \geq 4$, we can take $\epsilon = 0$ so that with $p = 6$, $2k - 1 + \frac{p-2}{p}(k - 1 + \epsilon) = 2k - 1 + \frac{2}{3}(k - 1) \geq 2k + 1$, yielding the desired estimate. In the case $k \leq 3$, we

use a bootstrap argument. Suppose that the error estimate (3.8) holds with bound $h^{k+\rho_n}$. Then, the new bound for $\mathfrak{T}_{2,2}(\eta, e)$ is $h^{2k-1+\rho_n+\frac{p-2}{p}(k-1+\epsilon_n+\rho_n)}$, the bound on $\mathfrak{T}_{2,1}(\eta, e)$ remaining unchanged. The two terms are balanced by the choice

$$\epsilon_n = \frac{(3-k)p + 2(k-1)}{(1+\lambda)p - 2} - \frac{2(p-1)}{(1+\lambda)p - 2} \rho_n,$$

and this leads to the new error estimate with bound $h^{k+\rho_{n+1}}$, where $\rho_{n+1} = \frac{1}{2}(1-\lambda\epsilon_n)$. Hence, we obtain a recursive relation of the form $\rho_{n+1} = a + b\rho_n$, where using $p = 6$ we have

$$a = \frac{1}{2} \left(1 - 2\lambda \frac{4-k}{3\lambda+2} \right), \quad b = \frac{5\lambda}{6\lambda+4}.$$

The recursive relation has a fixed point, $\rho_\infty = \frac{a}{1-b} = \frac{2+\lambda(2k-5)}{4+\lambda}$, since $0 \leq b < 1$. The quantity ρ_∞ is positive for all λ if $k = 3$, all $\lambda < 2$ if $k = 2$, and all $\lambda < \frac{2}{3}$ if $k = 1$. The fixed-point ρ_∞ is larger than $\frac{1}{2}$ only for $k = 3$. This completes the proof. \square

Remark 3.2. Sharper results can be derived by using the advective derivative in the weighting function. Following [21], the proof of the error estimate then hinges on discrete inf-sup stability, assuming additional regularity in time of the exact solution. As a result, the desired $h^{k+\frac{1}{2}}$ error estimate is achieved for any polynomial degree, any value for λ , and shape-regular meshes.

4. Numerical experiments. The objective of this section is to illustrate the efficiency of the weighted edge stabilization introduced and analyzed in section 3. The weighting function α in (3.3) is chosen to be $\alpha(r) = 2(1+r^2)^{-1}$ in all the test cases reported below.

4.1. One-dimensional tests. We report in this section one-dimensional tests over the periodic domain $\Omega := (0, 1)$ using \mathbb{P}_1 finite elements. The time stepping uses the SSP RK3 scheme with CFL = 0.2. The parameters for the first-order viscosity, entropy viscosity, and edge stabilization are $c_{\max} = \frac{1}{2}$, $c_{\text{ev}} = \frac{1}{2}$, and $c_{\text{ed}} = 0.05$, respectively.

4.1.1. Smooth transport. We consider the one-dimensional transport equation $\partial_t u + \partial_x u = 0$ with initial data $u_0(x) = \sin(2\pi x)$ to compare the convergence properties of the linear (unweighted) edge stabilization with those of the weighted edge stabilization. The computations are performed up to $t_F = 1$ on various meshes. The meshes are nonuniform to avoid superconvergence effects. The size of each cell is chosen randomly with a maximum deformation factor equal to 3, that is, the size ratio between two neighboring cells is at most 3.

The L^1 - and L^2 -norms of the error at $t_F = 1$ are reported in Table 4.1 for seven meshes. This test shows that the convergence properties of the linear edge stabilization and those of the weighted edge stabilization are identical for smooth solutions.

4.1.2. Nonsmooth transport. We now test the convergence properties of the entropy viscosity method with the weighted edge stabilization. We consider the one-dimensional transport equation with nonsmooth initial data (2.13) and periodic boundary conditions. The computations are performed with \mathbb{P}_1 finite elements up to $t_F = 1$ on various meshes. The L^1 - and L^2 -norms of the error are reported in Table 4.2 for seven uniform (left table) and seven nonuniform meshes (right table).

TABLE 4.1

Linear transport with smooth solution. L^1 - and L^2 -norms of error, \mathbb{P}_1 finite elements with edge stabilization (left) and weighted edge stabilization (right), $t_F = 1$, nonuniform meshes.

(a) Edge stabilization					(b) Weighted edge stabilization			
h	L^1 -norm	Rate	L^2 -norm	Rate	L^1 -norm	Rate	L^2 -norm	Rate
1.000E-02	3.309E-04	–	4.095E-04	–	3.442E-04	–	5.233E-04	–
5.000E-03	7.845E-05	2.077	9.699E-05	2.078	7.837E-05	2.135	1.068E-04	2.292
2.500E-03	1.875E-05	2.065	2.342E-05	2.050	1.866E-05	2.070	2.448E-05	2.126
1.250E-03	4.630E-06	2.018	5.788E-06	2.017	4.625E-06	2.013	5.954E-06	2.039
6.250E-04	1.158E-06	1.999	1.439E-06	2.008	1.157E-06	1.999	1.467E-06	2.021
3.125E-04	2.874E-07	2.011	3.572E-07	2.010	2.875E-07	2.009	3.625E-07	2.017
1.563E-04	7.081E-08	2.021	8.796E-08	2.022	7.114E-08	2.015	8.939E-08	2.020

TABLE 4.2

Linear transport with nonsmooth solution. Entropy viscosity method with the weighted edge stabilization. L^1 - and L^2 -norms of error on uniform meshes (left) and nonuniform meshes (right), \mathbb{P}_1 finite elements, $t_F = 1$.

(a) Uniform meshes					(b) Nonuniform meshes			
h	L^1 -norm	Rate	L^2 -norm	Rate	L^1 -norm	Rate	L^2 -norm	Rate
1.000E-02	4.694E-02	–	1.194E-01	–	4.897E-02	–	1.219E-01	–
5.000E-03	2.720E-02	0.787	9.094E-02	0.393	2.873E-02	0.769	9.320E-02	0.388
2.500E-03	1.590E-02	0.774	6.965E-02	0.385	1.685E-02	0.770	7.132E-02	0.386
1.250E-03	9.342E-03	0.768	5.350E-02	0.381	1.002E-02	0.750	5.511E-02	0.372
6.250E-04	5.504E-03	0.763	4.115E-02	0.379	5.954E-03	0.751	4.240E-02	0.378
3.125E-04	3.254E-03	0.758	3.168E-02	0.378	3.552E-03	0.745	3.269E-02	0.375
1.563E-04	1.930E-03	0.754	2.439E-02	0.377	2.115E-03	0.748	2.522E-02	0.374

TABLE 4.3

Weakened maximum principle for linear transport with nonsmooth solution on uniform meshes (left) and nonuniform meshes (right), \mathbb{P}_1 finite elements, $t_F = 1$.

(a) Uniform meshes					(b) Nonuniform meshes			
h	e_{Min}	Rate	e_{Max}	Rate	e_{Min}	Rate	e_{Max}	Rate
1.000E-02	6.498E-03	–	6.515E-03	–	6.655E-03	–	6.354E-03	–
5.000E-03	4.470E-03	0.540	4.470E-03	0.544	4.452E-03	0.580	4.462E-03	0.510
2.500E-03	2.966E-03	0.592	2.966E-03	0.592	3.171E-03	0.490	3.153E-03	0.501
1.250E-03	2.021E-03	0.553	2.021E-03	0.553	2.124E-03	0.578	2.177E-03	0.534
6.250E-04	1.345E-03	0.588	1.345E-03	0.588	1.474E-03	0.527	1.473E-03	0.564
3.125E-04	8.992E-04	0.580	8.992E-04	0.580	1.013E-03	0.541	1.022E-03	0.527
1.563E-04	6.091E-04	0.562	6.091E-04	0.562	7.544E-04	0.425	7.189E-04	0.508

The convergence orders in the L^1 - and L^2 -norm appear to be $\frac{3}{4}$ and $\frac{3}{8}$, respectively. Considering that the initial data is in $BV(\Omega)$ (almost $W^{1,1}(\Omega)$) and in $B_{\infty,2}^{\frac{1}{2}}(\Omega)$ (almost $H^{\frac{1}{2}}(\Omega) := W^{\frac{1}{2},2}(\Omega)$), these rates are compatible with the estimates $\frac{k+\frac{1}{2}}{k+1}$ and $\frac{1}{2} \frac{k+\frac{1}{2}}{k+1}$ obtained by interpolation using the rate $(k + \frac{1}{2})$ for smooth solutions.

We now verify that the weighted edge stabilization does preserve the weakened maximum principle (see (2.9)) of the entropy viscosity, contrary to the linear (unweighted) edge stabilization; see section 2.2.2. We show in Table 4.3 the indicators e_{Max} and e_{Min} defined by (2.14) for seven uniform and nonuniform meshes. By comparing Table 2.1 with Table 4.3, we observe that the convergence rate on e_{Max} and e_{Min} for the entropy viscosity solution with weighted edge stabilization is larger than that

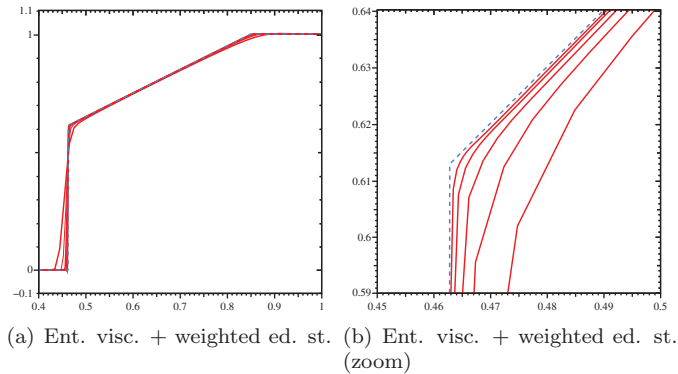


FIG. 4.1. *Nonconvex flux. Results at $t_F = 1$ with entropy viscosity with weighted edge stabilization on five uniform meshes composed of 100, 200, 400, 800, and 1600 cells.*

for the entropy viscosity solution alone, so that the entropy viscosity and the weighted edge stabilization are now working together instead of antagonizing each other.

4.1.3. One-dimensional nonconvex conservation equation. We finish this series of one-dimensional tests by testing again the nonlinear scalar conservation equation with nonconvex flux and initial data specified in (2.15). The computations are performed on five uniform meshes composed of 100, 200, 400, 800, and 1600 cells using the entropy viscosity method with weighted edge stabilization. The graphs of the solutions obtained at $t_F = 1$ are shown in Figure 4.1. We observe that, contrary to what is shown in Figure 2.2(b), the method now converges to the correct entropy solution thereby confirming again that weighting the edge stabilization is a good idea.

4.2. Two-dimensional tests. We illustrate the method on two-dimensional problems.

4.2.1. Smooth transport. Let us consider the domain $\Omega = \{\mathbf{x} \in \mathbb{R}^2, \|\mathbf{x}\| < 1\}$, the rotating velocity field $\boldsymbol{\beta}(\mathbf{x}) = 2\pi(-y, x)$, where $\mathbf{x} = (x, y)$, and the two-dimensional transport problem

$$(4.1) \quad \partial_t u + \nabla \cdot (\boldsymbol{\beta} u) = 0, \quad u(\mathbf{x}, 0) = \frac{1}{2} \left(1 - \tanh \left(\frac{(\mathbf{x} - \mathbf{r}_0)^2}{a^2} - 1 \right) \right)$$

with $a = 0.3$ and $\mathbf{r}_0 = (0.4, 0)$. Note that $\nabla \cdot \boldsymbol{\beta} = 0$ and $\boldsymbol{\beta} \cdot \mathbf{n}|_{\partial\Omega} = 0$.

We perform tests on triangular isoparametric finite elements. The meshes are quasi-uniform and based on a Delaunay triangulation technique [29]. The time stepping uses the RK4 scheme, and the CFL used in all the tests is $\text{CFL} = 0.25$. The parameters for the first-order viscosity, entropy viscosity, and edge stabilization are $c_{\max} = \frac{1}{4k}$, $c_{\text{ev}} = 0.1$, and $c_{\text{ed}} = 0.025$, respectively.

Convergence tests on the above smooth problem, not reported here, have shown that, when used alone, the weighted edge stabilization performs as well as the original edge stabilization technique both for \mathbb{P}_1 and \mathbb{P}_2 finite elements. When combined with the weighted edge stabilization, the converge rate of the entropy viscosity method with \mathbb{P}_1 finite elements, which is already second-order in the L^2 -norm, is not increased. The story is a little different for higher-order finite elements. The entropy viscosity alone has been observed to be second-order for \mathbb{P}_1 finite elements and to be of order $(k + \epsilon(k))$ for higher-order polynomial degrees; see [17]. Augmenting the entropy

TABLE 4.4

Linear transport with smooth solution (4.1). L^1 - and L^2 -norms of error with the entropy viscosity method (left) and the entropy viscosity with weighted edge stabilization (right), \mathbb{P}_2 finite elements, $t_F = 1$.

(a) Entropy viscosity					(b) Ent. visc. + weighted ed. st.			
h	L^1	Rate	L^2	Rate	L^1	Rate	L^2	Rate
2.00E-01	1.102E-01	–	8.410E-02	–	1.104E-01	–	8.851E-02	–
1.00E-01	2.317E-02	2.251	1.700E-02	2.306	1.783E-02	2.630	1.675E-02	2.401
5.00E-02	3.659E-03	2.663	2.268E-03	2.906	1.362E-03	3.710	1.304E-03	3.684
2.50E-02	8.099E-04	2.176	4.579E-04	2.308	1.121E-04	3.603	1.137E-04	3.520
1.25E-02	2.142E-04	1.919	1.173E-04	1.965	9.368E-06	3.581	1.019E-05	3.479
1.00E-02	1.337E-04	2.109	7.370E-05	2.083	4.249E-06	3.543	4.733E-06	3.439

TABLE 4.5

Linear transport with nonsmooth solution. L^1 - and L^2 -norms of error with entropy viscosity with weighted edge stabilization for \mathbb{P}_1 finite elements (left) and \mathbb{P}_2 finite elements (right), at $t_F = 1$.

(a) Ent. visc. + weighted ed. st. \mathbb{P}_1					(b) Ent. visc. + weighted ed. st. \mathbb{P}_2			
h	L^1	Rate	L^2	Rate	L^1	Rate	L^2	Rate
2.00E-01	1.381E+00	–	7.374E-01	–	7.464E-01	–	5.058E-01	–
1.00E-01	9.170E-01	0.591	5.544E-01	0.411	4.562E-01	0.710	3.779E-01	0.420
5.00E-02	5.454E-01	0.750	4.172E-01	0.410	2.605E-01	0.808	2.794E-01	0.436
2.50E-02	3.236E-01	0.753	3.158E-01	0.402	1.490E-01	0.806	2.114E-01	0.402
1.25E-02	1.904E-01	0.765	2.411E-01	0.389	8.594E-02	0.794	1.601E-01	0.401
1.00E-02	1.607E-01	0.759	2.214E-01	0.383	7.211E-02	0.786	1.466E-01	0.394
6.25E-03	1.124E-01	0.760	1.851E-01	0.381				

viscosity method with weighted edge stabilization improves the convergence order. We show in Table 4.4 convergence tests on the above smooth transport problem (4.1) with \mathbb{P}_2 finite elements. We compare the convergence rate of the entropy viscosity method alone and the entropy viscosity method augmented with weighted edge stabilization. Adding the weighted edge stabilization clearly improves the convergence rate of the entropy viscosity method. Some superconvergence effect is observed. (The errors are estimated by using a Gaussian quadrature exact on \mathbb{P}_5 polynomials.)

4.2.2. Nonsmooth transport. We solve again the linear transport problem (4.1), but this time we consider the following nonsmooth initial data $u(\mathbf{x}, 0) = 1$ if $\|\mathbf{x} - \mathbf{r}_0\| < a$ and $u(\mathbf{x}, 0) = 0$ otherwise. The meshes and the time stepping are the same as before.

We report in Table 4.5 convergence tests in the L^1 - and L^2 -norms for various meshes using the entropy viscosity method augmented with weighted edge stabilization. The results for \mathbb{P}_1 elements are shown in the left table and those for \mathbb{P}_2 elements are shown in the right table. The convergence orders in the L^1 - and L^2 -norms appear to be $\frac{3}{4}$ and $\frac{3}{8}$ for \mathbb{P}_1 elements and $\frac{4}{5}$ and $\frac{2}{5}$ for \mathbb{P}_2 elements, respectively. As above, these rates are compatible with the estimates $\frac{k+\frac{1}{2}}{k+1}$ and $\frac{1}{2} \frac{k+\frac{1}{2}}{k+1}$ obtained by interpolation using the rate $(k + \frac{1}{2})$ for smooth solutions.

We now verify the weakened maximum principle by computing the indicators e_{Max} and e_{Min} defined by (2.14) for each mesh. The results are reported in Table 4.6; those for \mathbb{P}_1 finite elements are shown in the left table and those for \mathbb{P}_2 finite elements in the right table. We observe that both indicators e_{Max} and e_{Min} converge to zero with the

TABLE 4.6

Linear transport with nonsmooth solution. Weakened maximum principle at $t_F = 1$ for entropy viscosity with weighted edge stabilization with \mathbb{P}_1 finite elements (left) and \mathbb{P}_2 finite elements (right).

(a) Ent. visc. + weighted ed. st. \mathbb{P}_1					(b) Ent. visc. + weighted ed. st. \mathbb{P}_2			
h	e_{Min}	Rate	e_{Max}	Rate	e_{Min}	Rate	e_{Max}	Rate
2.00E-01	4.264E-02	—	5.354E-01	—	2.003E-02	—	1.626E-01	—
1.00E-01	5.337E-02	-0.324	2.037E-01	1.394	1.250E-02	0.680	1.015E-02	4.001
5.00E-02	5.313E-02	0.006	3.546E-02	2.522	8.397E-03	0.574	7.904E-03	0.361
2.50E-02	1.911E-02	1.475	1.283E-02	1.467	7.900E-03	0.088	6.943E-03	0.187
1.25E-02	6.362E-03	1.587	7.776E-03	0.722	6.131E-03	0.366	5.953E-03	0.222
1.00E-02	5.713E-03	0.482	6.798E-03	0.603	5.150E-03	0.782	5.211E-03	0.596
6.25E-03	4.885E-03	0.333	5.461E-03	0.466				

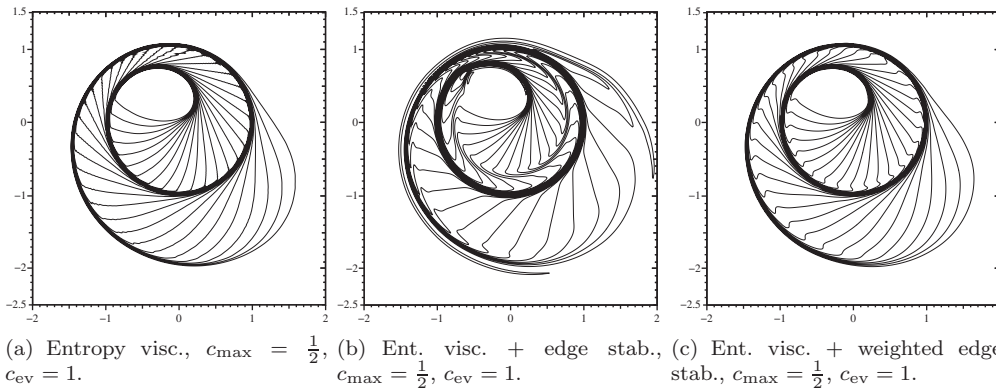


FIG. 4.2. Two-dimensional conservation equation with nonconvex flux, $t_F = 1$. Entropy viscosity solution without edge stabilization (left), with unweighted edge stabilization (center), and with weighted edge stabilization (right) on a uniform mesh composed of 118850 \mathbb{P}_1 nodes, $CFL = 0.025$.

meshsize, thus indicating that the entropy viscosity with weighted edge stabilization satisfies a weakened maximum principle.

4.3. Nonconvex conservation equation. Consider the following two-dimensional scalar conservation equation with nonconvex fluxes:

$$(4.2) \quad \partial_t u + \nabla \cdot \mathbf{f}(u) = 0, \quad \mathbf{f}(u) = (\sin u, \cos u), \quad u(x, y, 0) = \begin{cases} 3.5\pi, & x^2 + y^2 < 1, \\ 0.25\pi, & \text{otherwise.} \end{cases}$$

The local velocity is $\mathbf{f}'(u) = (\cos u, -\sin u)$. This is a Cauchy problem in \mathbb{R}^2 . This test was proposed in [28] and is challenging to many high-order numerical schemes because the solution has a two-dimensional composite wave structure; see [28] for details.

We perform computations on a mesh composed of triangular finite elements. The mesh family is quasi-uniform. The solution is computed at $t_F = 1$. The time stepping uses the standard RK4 scheme with $CFL = 0.025$. Three computations are performed with the entropy viscosity method (with coefficients $c_{\text{max}} = \frac{1}{2}$, $c_{\text{ev}} = 1$): one without edge stabilization, one with unweighted edge stabilization (with $c_{\text{ed}} = 1$), and one with weighted edge stabilization (with $c_{\text{ed}} = 1$ and the weighting function α specified at the beginning of section 4). The graph of the \mathbb{P}_1 solution without edge stabilization is shown in Figure 4.2(a). The solution shows the expected rotating composite wave structure composed of a shock and an expansion. This solution is an accurate

approximation of the entropy solution. The graph of the solution with unweighted edge stabilization is shown in Figure 4.2(b). We observe that this solution exhibits a nonphysical layer between the shock and the expansion wave. We have verified by mesh refinement that this approximation does not converge to the entropy solution. Finally, the graph of the solution with weighted edge stabilization is shown in Figure 4.2(c). We observe that the unphysical layer has been pushed back to the shock, thereby recovering the correct physical behavior.

5. Conclusions. To conclude we would like to offer some thoughts on the relative importance of nonlinear viscosities (also called shock capturing in the literature) and linear stabilization techniques. Substantial efforts have been devoted to the construction of linear stabilization techniques (GaLS, RFB, SUPG, VMS, streamline diffusion, subgrid viscosity, edge stabilization, local projection, etc.), and it is often implicitly understood that linear stabilization is the workhorse whereas shock capturing is only meant to remove remaining spurious oscillations. As a result, the amount of research dedicated to constructing and analyzing shock capturing techniques is comparatively smaller than that dedicated to linear stabilization. As shown in the present paper, linear stabilization techniques can promote the Gibbs phenomenon and can transform a converging method into a nonconverging one. We think that nonlinear stabilization methods (shock capturing, nonlinear viscosities, etc.) deserve more attention. These methods should be considered the workhorses that cure the Gibbs phenomenon and ensure convergence to the physically correct weak solution; linear stabilization techniques should be seen as auxiliary tools whose job is to improve the convergence of an already converging nonlinear stabilization method. We refer to [9] for a recent step in this direction.

Appendix A. Entropy viscosity. The purpose of this appendix is to briefly recall the principles of the so-called entropy viscosity method. The specificities of the method are unrelated to the content of this paper, but we nevertheless describe how it is implemented for the sake of completeness. Full descriptions of the technique can be found in [18, 17]. The time approximation of (2.6) uses an explicit Runge–Kutta method defined by a Butcher tableau

$$(A.1) \quad \begin{array}{c|cccc} 0 & & & & \\ c_2 & a_{21} & & & \\ c_3 & a_{31} & a_{32} & & \\ \vdots & \vdots & & \ddots & \\ c_r & a_{r1} & a_{r2} & \cdots & a_{r,r-1} \\ \hline & b_1 & b_2 & \cdots & b_{r-1} & b_r \end{array}$$

We consider the SSP RK3 and RK4 schemes for which the Butcher tableaux are

$$(A.2) \quad \begin{array}{c|ccc} 0 & & & \\ 1 & 1 & & \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & \\ \hline & \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \end{array} \quad \begin{array}{c|ccc} 0 & & & \\ \frac{1}{2} & \frac{1}{2} & & \\ \frac{1}{2} & 0 & \frac{1}{2} & \\ 1 & 0 & 0 & 1 \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array}$$

The algorithm is initialized by setting $\nu_h^0 = 0$ and $u_h^0 = u_{0,h}$. The viscosity field is piecewise constant over the mesh \mathcal{K}_h . Assume that the current time is t_n and let u_h^n

be the approximation of $u_h(t_n)$. The next time increment is defined by

$$(A.3) \quad \Delta t_n = \text{CFL} \min_{K \in \mathcal{K}_h} \frac{h_K}{|\mathbf{f}'(u_h^n|_K)|}.$$

The approximation u_h^{n+1} and the viscosity field ν_h^{n+1} are evaluated at time $t_{n+1} = t_n + \Delta t_n$ as

$$(A.4) \quad u_h^{n+1} = u_h^n - \Delta t_n (b_1 k_1 + b_2 k_2 + \dots + b_r k_r), \quad \nu_h^{n+1} = \nu_r,$$

where the functions $\{k_l\}_{1 \leq l \leq r}$ and the viscosity fields $\{\nu_l\}_{1 \leq l \leq r}$ are recursively computed as follows:

$$(A.5) \quad \left\{ \begin{array}{l} w_1 = u_h^n, \\ \nu_1 = \nu_h^n, \quad \int_{\Omega} k_1 v \, d\Omega = \int_{\Omega} v \nabla \cdot \mathbf{f}(w_1) + b_{\text{ev}}(\nu_1, w_1, v) \quad \forall v \in X_h, \\ w_2 = u_h^n - \Delta t_n a_{21} k_1, \\ \nu_2 = H_2(w_1, w_2), \quad \int_{\Omega} k_2 v \, d\Omega = \int_{\Omega} v \nabla \cdot \mathbf{f}(w_2) + b_{\text{ev}}(\nu_2, w_2, v) \quad \forall v \in X_h, \\ \vdots \\ w_r = u_h^n - \Delta t_n (a_{r1} k_1 + a_{r2} k_2 + \dots + a_{r,r-1} k_{r-1}), \\ \nu_r = H_r(w_1, w_2, \dots, w_r), \quad \int_{\Omega} k_r v \, d\Omega = \int_{\Omega} v \nabla \cdot \mathbf{f}(w_r) + b_{\text{ev}}(\nu_r, w_r, v) \quad \forall v \in X_h, \end{array} \right.$$

where $b_{\text{ev}}(\nu, w, v) := \sum_{K \in \mathcal{K}_h} \nu|_K \int_K \nabla w \cdot \nabla v \, dK$.

The functions H_2, \dots, H_r are piecewise constant over the mesh \mathcal{K}_h and are defined as follows. First, we define the so-called entropy functional $E \in C^1(\mathbb{R}; \mathbb{R})$. In all our numerical tests, we use the quadratic entropy for the functional E . Second, we evaluate the entropy residual associated with two given states $c_h, d_h \in X_h$ as follows. Let δ be the time increment separating the states c_h and d_h . The entropy residual in the cell K at $\mathbf{x} \in K$ associated with the states c_h, d_h and the time increment δ is defined by

$$(A.6) \quad R_K(c_h, d_h, \delta)(\mathbf{x}) = \frac{E(d_h(\mathbf{x})) - E(c_h(\mathbf{x}))}{\delta} + \mathbf{f}'(d_h(\mathbf{x})) \cdot \nabla(E(d_h(\mathbf{x}))).$$

We now define the entropy residual associated with the mesh interfaces. For all $F \in \mathcal{F}_h^i$ and for all $\mathbf{x} \in F$, we set

$$(A.7) \quad J_F(d_h)(\mathbf{x}) = (\mathbf{f}'(d_h(\mathbf{x})) \cdot \mathbf{n}(\mathbf{x})) [\partial_n E(d_h(\mathbf{x}))].$$

The total entropy residual in the cell K associated with the states c_h, d_h and the time increment δ is then defined to be

$$(A.8) \quad D_K(c_h, d_h, \delta) = \max(\|R_K(c_h, d_h, \delta)\|_{L^\infty(K)}, \max_{F \in \partial K \cap \mathcal{F}_h^i} \|J_F(d_h)\|_{L^\infty(F)}).$$

The maximum local wave speed in the cell K is estimated by

$$(A.9) \quad \beta_K(d_h) = \max_{\mathbf{x} \in K} |\mathbf{f}'(d_h(\mathbf{x}))|.$$

Then, the value over K of the viscosity function H_l , $2 \leq l \leq r$, is defined as follows:

$$(A.10) \quad H_l(w_1, \dots, w_l)|_K = \min(c_{\max}\beta_K(w_l)h_K, c_{\text{ev}}h_K^2 D_K(w_1, w_l, c_l \Delta t_n)),$$

where the local meshsize is defined to be $h_K = \text{diam}(K)/k$.

Note that each step of (A.5) requires solving a mass matrix linear system. We stress again that we do not recommend to use the lumped mass matrix instead of the consistent mass matrix since mass lumping induces large dispersion errors [19].

REFERENCES

- [1] Y. ACHDOU, C. BERNARDI, AND F. COQUEL, *A priori and a posteriori analysis of finite volume discretizations of Darcy's equations*, Numer. Math., 96 (2003), pp. 17–42.
- [2] M. BOMAN, *Estimates for the L_2 -projection onto continuous finite element spaces in a weighted L_p -norm*, BIT, 46 (2006), pp. 249–260.
- [3] S.C. BRENNER AND R.L. SCOTT, *The Mathematical Theory of Finite Element Methods*, Texts Appl. Math. 15, Springer, New York, 1994.
- [4] E. BURMAN, A. ERN, AND M.A. FERNÁNDEZ, *Explicit Runge-Kutta schemes and finite elements with symmetric stabilization for first-order linear PDE systems*, SIAM J. Numer. Anal., 48 (2010), pp. 2019–2042.
- [5] E. BURMAN AND A. ERN, *Nonlinear diffusion and discrete maximum principle for stabilized Galerkin approximations of the convection–diffusion–reaction equation*, Comput. Methods Appl. Mech. Engrg., 191 (2002), pp. 3833–3855.
- [6] E. BURMAN AND A. ERN, *Continuous interior penalty hp-finite element methods for advection and advection-diffusion equations*, Math. Comp., 76 (2007), pp. 1119–1140.
- [7] E. BURMAN AND P. HANSBO, *Edge stabilization for Galerkin approximations of convection-diffusion-reaction problems*, Comput. Methods Appl. Mech. Engrg., 193 (2004), pp. 1437–1453.
- [8] E. BURMAN, *A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty*, SIAM J. Numer. Anal., 43 (2005), pp. 2012–2033.
- [9] E. BURMAN, *On nonlinear artificial viscosity, discrete maximum principle and hyperbolic conservation laws*, BIT, 47 (2007), pp. 715–733.
- [10] R. CODINA, *A discontinuity-capturing crosswind-dissipation for the finite element solution of the convection-diffusion equation*, Comput. Methods Appl. Mech. Engrg., 110 (1993), pp. 325–342.
- [11] M. CROUZEIX AND V. THOMÉE, *The stability in L_p and W_p^1 of the L_2 -projection onto finite element function spaces*, Math. Comp., 48 (1987), pp. 521–532.
- [12] D.A. DI PIETRO AND A. ERN, *Mathematical aspects of discontinuous Galerkin methods*, Mathématiques et Applications 69, Springer, Heidelberg, 2012.
- [13] A. ERN AND J.-L. GUERMOND, *Theory and practice of finite elements*, Appl. Math. Sci. 159, Springer-Verlag, New York, 2004.
- [14] A. ERN AND J.-L. GUERMOND, *Discontinuous Galerkin methods for Friedrichs' systems. I. General theory*, SIAM J. Numer. Anal., 44 (2006), pp. 753–778.
- [15] S. GOTTLIEB, C.-W. SHU, AND E. TADMOR, *Strong stability-preserving high-order time discretization methods*, SIAM Rev., 43 (2001), pp. 89–112.
- [16] J.-L. GUERMOND, A. MARRA, AND L. QUARTAPELLE, *Subgrid stabilized projection method for 2D unsteady flows at high Reynolds number*, Comput. Methods Appl. Mech. Engrg., 195 (2006), pp. 5857–5876.
- [17] J.-L. GUERMOND, R. PASQUETTI, AND B. POPOV, *Entropy viscosity method for nonlinear conservation laws*, J. Comput. Phys., 230 (2011), pp. 4248–4267.
- [18] J.-L. GUERMOND AND R. PASQUETTI, *Entropy-based nonlinear viscosity for Fourier approximations of conservation laws*, C. R. Math. Acad. Sci. Paris, 346 (2008), pp. 801–806.
- [19] J.-L. GUERMOND AND R. PASQUETTI, *A correction technique for the dispersive effects of mass lumping for transport problems*, CMAME, 253 (2012), pp. 186–198.
- [20] J.-L. GUERMOND, *Stabilization of Galerkin approximations of transport equations by subgrid modeling*, M2AN Math. Model. Numer. Anal., 33 (1999), pp. 1293–1316.
- [21] J.-L. GUERMOND, *Subgrid stabilization of Galerkin approximations of linear contraction semi-groups of class C^0 in Hilbert spaces*, Numer. Methods Partial Differential Equations, 17 (2001), pp. 1–25.
- [22] J.-L. GUERMOND, *On the use of the notion of suitable weak solutions in CFD*, Internat. J. Numer. Methods Fluids, 57 (2008), pp. 1153–1170.

- [23] J.S. HESTHAVEN AND T. WARBURTON, *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications*, Texts Appl. Math. 54, Springer, New York, 2008.
- [24] T.J.R. HUGHES AND M. MALLET, *A new finite element formulation for computational fluid dynamics. IV: A discontinuity-capturing operator for multidimensional advective-diffusive systems*, Comput. Methods Appl. Mech. Engrg., 58 (1986), pp. 329–336.
- [25] C. JOHNSON AND J. PITKÄRANTA, *An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation*, Math. Comp., 46 (1986), pp. 1–26.
- [26] C. JOHNSON AND A. SZEPESSY, *On the convergence of a finite element method for a nonlinear hyperbolic conservation law*, Math. Comp., 49 (1987), pp. 427–444.
- [27] O.A. KARAKASHIAN AND F. PASCAL, *A posteriori error estimates for a discontinuous Galerkin approximation of second-order elliptic problems*, SIAM J. Numer. Anal., 41 (2003), pp. 2374–2399.
- [28] A. KURGANOV, G. PETROVA, AND B. POPOV, *Adaptive semidiscrete central-upwind schemes for nonconvex hyperbolic conservation laws*, SIAM J. Sci. Comput., 29 (2007), pp. 2381–2401 (electronic).
- [29] S. REBAY, *Efficient unstructured mesh generation by means of Delaunay triangulation and Bowyer-Watson algorithm*, J. Comput. Phys., 106 (1993), pp. 125–138.