On Discrete Maximum Principles for Discontinuous Galerkin Methods

Santiago Badia[a,b], Alba Hierro[a,b,*]

[a] *Centre Internacional de Mètodes Numèrics a l'Enginyeria (CIMNE), Parc Mediterrani de la Tecnologia, UPC, Esteve Terradas 5, 08860 Castelldefels, Spain*
[b] *Universitat Politècnica de Catalunya, Jordi Girona 1-3, Edifici C1, 08034 Barcelona, Spain.*

Abstract

The aim of this work is to propose a monotonicity-preserving method for discontinuous Galerkin (dG) approximations of convection-diffusion problems. To do so, a novel definition of discrete maximum principle (DMP) is proposed using the discrete variational setting of the problem, and we show that the fulfilment of this DMP implies that the minimum/maximum (depending on the sign of the forcing term) is on the boundary for multidimensional problems. Then, an artificial viscosity (AV) technique is designed for convection-dominant problems that satisfies the above mentioned DMP. The noncomplete stabilized interior penalty dG method is proved to fulfil the DMP property for the one-dimensional linear case when adding such AV with certain parameters. The benchmarks for the constant values to satisfy the DMP are calculated and tested in the numerical experiments section. Finally, the method is applied to different test problems in one and two dimensions to show its performance.

Keywords: discontinuous Galerkin, stabilized finite elements, shock capturing, nonlinear stabilization, convection-diffusion, convection-dominated flows

1. Introduction

It is well known that the operator L associated to an elliptic problem such as the convection-diffusion problem enjoys the maximum property, meaning that the maximum (resp., minimum) of the solution to the problem $Lu = f$ is achieved on the boundary of the domain if the source term, f, is negative (resp., positive). In particular, this property ensures that the solution of the problem will not show oscillations. It is well known that the solution of a convection-dominated problem may present sharp layers that may induce spurious oscillations in the discrete approximation of the solution. We are interested in finding a method that ensures a similar maximum property at the discrete discontinuous level in order to obtain a method that gives oscillation free solutions.

When the problem is discretized, this maximum property may be inherited by what is called *discrete maximum principle* (DMP). Several definitions of the DMP have been proposed in the literature for continuous discrete approximations (see [13, 16, 27, 7, 26]). Some of them are equivalent while some others are weaker or stronger. There is also a lot of literature about the conditions on the mesh for the Poisson problem to enjoy the DMP [16, 29, 17, 25] as well as discrete methods specially implemented to fulfil such property. Methods have been designed for linear finite differences [9] and continuous linear finite elements [8, 13, 23, 5, 7, 6]. These methods are implicit in sense, and usually based on the addition of AV to the problem at hand; they are traditionally called shock (or discontinuity)-capturing techniques, even though we favour the notation *nonlinear stabilization.* Some approaches to prove a DMP using piecewise higher order polynomials have been done [24, 25, 20, 29, 28, 31, 32] but only the Poisson problem has been proved to enjoy the DMP and only on certain one-dimensional (1D) meshes [29] and on very restrictive quadratic and cubic two dimensional meshes [22, 16]. When it comes to discontinuous methods, most of the shock capturing techniques are based on the concept of slope limiter, proposed by Cockburn and Shu for conservation laws [11, 10] and latter adapted to the convection-dominated convection-diffusion problem [12]. The same strategy can be applied to finite volume methods (see

[*] Corresponding author
Email addresses: sbadia@cimne.upc.edu (Santiago Badia), ahierro@cimne.upc.edu (Alba Hierro)

[33, 34, 35]). Again, these methods consist in a postprocess after the solution is computed and are designed for explicit methods. However, as far as we know, there are no works dealing with nonlinear stabilization and implicit DMP-preserving dG formulations. In fact, even the definition of what a DMP for dG means is open.

Concerning to the study of the DMP for the Poisson problem in the dG setting, there is only one work by Horváth and Mincsovics [17]; they analyse the fulfilment of certain condition on the stiffness matrix $\mathbf{K}$ that ensure the following property for the 1D interior penalty (IP) method:

$$\mathbf{K}\mathbf{u} \leq 0 \quad \Longrightarrow \quad \max \mathbf{u} \leq \max\{0, \max \mathbf{u}_{\partial\Omega}\}.$$

The aim of this work is twofold. On one side, we propose a new (variational) definition of the DMP for dG, and we prove that it is a sufficient condition to have the the minimum/maximum (depending on the sign of the forcing term) on the boundary for multidimensional problems. The new definition is stronger than the one given in [17] and it is, in some sense, closer to the one used in [4] for the 1D continuous Galerkin (cG) discretization of the Burgers' equation. On the other hand, we construct a multidimensional nonlinear stabilization based on AV for dG methods and prove that, when restricted to the 1D case, the nonlinear stabilization combined with an incomplete (or weighted) IP dG method with upwinding is capable to ensure our DMP for the discrete dG solution of (1). In any case, numerical experiments evindence that the method also satisfies the DMP in the multidimensional case.

The outline of the article is the following. In Section 2 we introduce the continuous convection-diffusion problems and its Galerkin discretization using finite elements. The IP dG method for the Laplacian is presented in Section 3. Our novel definition of the DMP for the dG scheme is proposed in Section 4 and some good properties derived from it are stated. Moreover, in Subsection 4.1, we prove that the IP method for the Laplacian enjoys the DMP in the 1D case. The extension of the IP method for the convection-diffusion problem is given in Section 5. In Section 6, an AV technique is proposed for the 1D case, and we prove that it satisfies the DMP. Further, we extend the method to the multidimensional case. Numerical experiments are included in Section 7. Finally, some conclusions are drawn in Section 8.

2. Weak Form and Notation

We will consider the convection-diffusion problem with Dirichlet boundary conditions:

$$\left\{ \begin{array}{rcll} Lu := -\nabla\cdot(\mu\nabla u) + \nabla\cdot(\beta u) & = & f & \text{in } \Omega, \\ u & = & g & \text{on } \partial\Omega. \end{array} \right. \tag{1}$$

We assume that $\mu \in L^2(\Omega)$ and $\beta \in H^1(\Omega) \cap C^0(\Omega)$ is solenoidal ($\nabla \cdot \beta = 0$). It is well known that the operator L associated to problem (1) enjoys the maximum principle (for proofs on maximum principles for elliptic problems see [14]).

Definition 1. We say that an operator L posseses the maximum principle if, for all $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\bar{\Omega})$, the following implication holds:

$$Lu \leq 0 \quad \text{in} \quad \Omega \quad \Longrightarrow \quad \max_S u \leq \max_{\partial S} u \quad \forall S \subset \Omega.$$

Before studying how to achieve a maximum principle at the discrete level for the convection-diffusion problem, we will focus on the rather simpler Poisson's equation:

$$\left\{ \begin{array}{rcll} -\Delta u & = & f & \text{in}\, \Omega \\ u & = & g & \text{on}\, \partial\Omega. \end{array} \right. \tag{2}$$

We denote by $(\cdot,\cdot)_K$ the $L^2(K)$ inner product for any $K \subset \Omega$ and by $(\cdot,\cdot)$ the $L^2(\Omega)$ product. We consider $(\cdot,\cdot)_h$ the $L^2(\Omega)$-scalar product evaluated using nodal quadrature (corresponding to the lumped mass matrix). The bilinear form $a(\cdot,\cdot)$ associated to the problem (2) is $a(u,v) = (\nabla u, \nabla v)$. So, the weak form of (2) reads as:

$$\text{Find } u \in H^1(\Omega) \text{ such that } a(u,v) = (f,v) \quad \forall v \in H^1(\Omega). \tag{3}$$

Let us consider partitions $\mathcal{T}_h^N = \{K\}$ of Ω formed by simplicial elements K of characteristic length h_K; we denote by h the characteristic size of the mesh. The corners of the mesh will be denoted by x_i, $i = 1, \cdots, N_h$ (N_h being the total number of corners), and the macroelement associated to x_i will be designated by $\Omega_i = \cup_{x_i \in K} K$. The discrete space considered henceforth is the discontinuous space of piecewise linear functions $V_h = \{v_h \,|\, v_h|_K \in \mathbb{P}_1(K)\ \forall K\}$. Let $\mathcal{E}_h = \cup_{K \in \mathcal{T}_h^N} \partial K$ be the set of the facets of the mesh and $\mathcal{E}_h^0 = \mathcal{E}_h \backslash \partial\Omega$. We define $T(\mathcal{E}_h) = \prod_{K \in \mathcal{T}_h^N} L^2(\partial K)$. The functions in $T(\mathcal{E}_h)$ are double-valued on $\mathcal{E}_h^0$ and single-valued on $\partial\Omega$; in particular, $V_h|_{\mathcal{E}_h} \subset T(\mathcal{E}_h)$. The functions $v_h \in V_h$ can be expressed as a linear combination of the basis $\{\varphi_i^K\}$ where φ_i^K is defined for all pairs $\{i, K\} \in \{1, \cdots, N_h\} \times \mathcal{T}_h^N$ such that $x_i \in K$. φ_i^K corresponds to the discontinuous function that is linear in K, with $\varphi_i^K(x_i) = 1$ and $\varphi_i^K(x_j) = 0$ for $x_j \in K$, $j \neq i$, and $\varphi_i^K = 0$ for $x \in \Omega \setminus K$. So, a function $v_h \in V_h$ would read as:

$$v_h(x) = \sum_{i=1}^{N_h} \sum_{K \subset \Omega_i} u_i^K \varphi_i^K(x), \qquad \forall x \in \Omega.$$

Moreover we can define the solution in a single element K as $u_h^K(x) = \sum_{x_i \in K} u_i^K \varphi_i^K(x)$, $\forall x \in K$, and its constant gradient $\nabla u_h^K = \sum_{x_i \in K} u_i^K \nabla \varphi_i^K|_K$. For any facet $F \in \mathcal{E}_h^0$ we know there are only two elements, say K_F^+ and K_F^-, such that $\partial K_F^+ \cap \partial K_F^- = F$. In addition, we can name n_F^+ and n_F^- the unitary normal to face F outside K_F^+ and K_F^-, respectively. Given $q \in T(\mathcal{E}_h)$, we can define the common concepts of average $\{\!\{\cdot\}\!\}$ and jump $[\![\cdot]\!]$ on an interior point x of a facet $F \in \mathcal{E}_h^0$ as follows:

$$\{\!\{q\}\!\}(x) = 0.5(q^{K_F^+}(x) + q^{K_F^-}(x)), \quad [\![q]\!](x) = q^{K_F^+}(x) n_F^+ + q^{K_F^-}(x) n_F^-.$$

We also define the harmonic average of q on x as $\langle q \rangle(x) = (2q^{K_F^+}(x) q^{K_F^-}(x))/(q^{K_F^+}(x) + q^{K_F^-}(x))$. On boundary points $x \in \partial\Omega$, we define $\{\!\{q\}\!\}(x) = q(x)$, $[\![q]\!](x) = q(x) n_{\partial\Omega}(x)$ and $\langle q \rangle(x) = q(x)$.

3. The Interior Penalty Method for the Poisson's Problem

There are numerous dG methods in the literature to approximate the Poisson problem. Many of them are contained in the unified analysis carried out by Arnold *et al.* in [1], where they conclude that any dG method approximating the second-order elliptic problem $-\Delta u = f$ uses the following bilinear form:

$$a_h(u_h, v_h) = \int_\Omega \nabla u_h \nabla v_h + \int_{\mathcal{E}_h} ([\![\tilde{u} - u_h]\!]\{\!\{\nabla v_h\}\!\} - \{\!\{\tilde{\sigma}\}\!\}[\![v_h]\!]) + \int_{\mathcal{E}_h^0} (\{\!\{\tilde{u} - u_h\}\!\}[\![\nabla v_h]\!] - [\![\tilde{\sigma}]\!]\{\!\{v_h\}\!\}),$$

where $\tilde{u} = \tilde{u}(u_h)$ and $\tilde{\sigma} = \tilde{\sigma}(u_h)$ are scalar numerical fluxes that approximate u and ∇u respectively on the boundaries of the elements. Different choices for these fluxes lead to different dG methods. We consider the IP method, which consists in taking

$$\tilde{u} = \{\!\{u_h\}\!\} + \xi n_K \cdot [\![u_h]\!], \qquad \tilde{\sigma} = \{\!\{\nabla u_h\}\!\} - C_1 [\![u_h]\!].$$

Given a facet F, the value $C_1(x) = c^{ip} \tilde{h}^{-1}$ for any $x \in F$, where $c^{ip}|_F = c_F^{ip}$ is a facet constant to be chosen and $\tilde{h}|_F = h_F := \min_{\bar{K} \supset F}\{h_K\}$. The parameter ξ can take values $\xi = 0, 0.5$ or 1, leading to the symmetric, incomplete, or nonsymmetric IP method, respectively:

$$a_h(u_h, v_h) = \int_\Omega \nabla u_h \nabla v_h - \int_{\mathcal{E}_h} ((1 - 2\xi)[\![u_h]\!]\{\!\{\nabla v_h\}\!\} + \{\!\{\nabla u_h\}\!\}[\![v_h]\!]) + \int_{\mathcal{E}_h} c^{ip} \tilde{h}^{-1} [\![u_h]\!][\![v_h]\!]. \quad (4)$$

According to the analysis performed in [17], the best option in order to guarantee the DMP is to choose $\xi = 0.5$.

4. Discrete Maximum Principle

We recall the definition of DMP given by Burman and Ern in [5, 4] for the linear cG method:

Definition 2 (DMP cG). We say that the semilinear form $a_h(u_h, v)$ has the DMP property if the following holds true: $\forall u_h \in V_h \cap \mathcal{C}(\Omega)$ and for all interior vertex x_i, if u_h is locally minimal (resp., maximal) on vertex x_i over a macroelement Ω_i (i.e., $u_h(x_i) \leq u_h(x)$, $\forall x \in \Omega_i$), there exists $\gamma_K > 0$ such that

$$a_h(u_h, \varphi_i) \leq -\sum_{K \subset \Omega_i} \gamma_K |\nabla u_h|_K|,$$

(resp., $a_h(u_h, \varphi_i) \geq \sum_{K \subset \Omega_i} \gamma_K |\nabla u_h|_K|$) where φ_i is the continuous shape function associated with the node x_i.

Basically, the previous definition ensures that, when $f \geq 0$, the solution to the discrete problem associated with the bilinear form has no local discrete minimum in the interior of the domain. As far as we know, there is no such a DMP definition for dG methods. So, we have to find out the properties that the dG method should enjoy in order to have a solution without local extrema. But even the definition of local extremum is not clear in dG. We have come up with the following definition of extremum:

Definition 3 (local discrete extremum). The function $u_h \in V_h$ has a local discrete minimum (resp., maximum) on node x_i in K if $u_i^K \leq u_h(x)$ (resp., $u_i^K \geq u_h(x)$) $\forall x \in \Omega_i$.

Remark 1. We use the adjective local to differentiate between the previous concept and a global minimum of the function u_h on x_i in K, what would mean that $u_i^K \leq u_h(x)$ for all $x \in \Omega$. The adjective discrete tries to emphasize that the definition is linked to the mesh provided in each case. Moreover, we will use strict local discrete extremum when the strict inequality holds.

Now, taking into account the definition of local discrete extremum we are ready to give our own definition of DMP for dG:

Definition 4 (DMP dG). We say that the bilinear form $a_h(u_h, v)$ has the DMP property if the following holds true: for all $u_h \in V_h$ and for all interior vertex x_i, if u_h is locally minimal (resp., maximal) on vertex x_i in K, then there exist $\gamma_F > 0$ and $\delta_K > 0$ such that

$$a_h(u_h, \varphi_i^K) \leq -\sum_{F \in K, F \ni x_i} \gamma_F h_F^{-1} \int_F |[\![u_h]\!]| - \delta_K h_K^{-1} \int_K |\nabla u_h^K|, \tag{5}$$

(resp., $a_h(u_h, \varphi_i^K) \geq \sum_F \gamma_F h_F^{-1} \int_F |[\![u_h]\!]| + \delta_K h_K^{-1} \int_K |\nabla u_h^K|$).

This definition implies the following interesting property for the solution of the method.

Lemma 1. *Let $a_h(u_h, v_h)$ be a bilinear form enjoying the DMP property. If we solve the problem $a_h(u_h, v_h) = (f, v_h)$ with $f \geq 0$ (resp., $f \leq 0$), the solution u_h has no strict local discrete minimum (resp., maximum) in any interior point. As a result, the global minimum (resp. maximum) is on the boundary.*

Proof. Suppose that u_h has a local discrete minimum on an interior node x_i in K. Then, $a_h(u_h, \varphi_i^K) \leq -\sum_F \gamma_F h_F^{-1} \int_F |[\![u_h]\!]| - \delta_K h_K^{-1} \int_K |\nabla u_h^K| \leq 0$. Since $(f, \varphi_i^K) \geq 0$, it implies that $a_h(u_h, \varphi_i^K) = 0$. Then, the right hand side of (5) must be zero, implying that $\nabla u_h^K = 0$ and $[\![u_h]\!] = 0$. Let $K' \subset \Omega_i$ be a finite element sharing a facet F with K. The previous result implies that $u_h^K(x) = u_h^{K'}(x)$ for any $x \in F$. In particular, $u_i^K = u_i^{K'}$, and using the definition of minimum, we infer that u_h has a local discrete minimum on x_i in K' too. By induction, $\nabla u_h = 0$ on Ω_i and $u_h|_{\Omega_i} = u_i^K$ is constant. Clearly the minimum is not strict. Since a global minimum on x_i would imply, in particular, a local discrete minimum, we could follow the same reasoning and deduce that the global minimum is shared by all the nodes in Ω_i. By induction, the function should be constant and the value of the function would be the same as in the boundary. So the global minimum must be on the boundary. □

Remark 2. The DMP introduced before would correspond to a strong maximum principle at the continuous level (see [14]).

Now let us consider the transient problem

$$\begin{cases} u_t - \Delta u = f & \text{in}\,\Omega, \\ u(x,0) = u_0(x) & x \in \Omega, \\ u(x,t) = g(x,t) & x \in \partial\Omega, \end{cases} \tag{6}$$

and discretise it (in space) as follows:

$$\text{Find}\, u_h \in V_h \,\text{such that}\, (\partial_t u_h, v_h)_h + a_h(u_h, v_h) = (f, v) \quad \forall v_h \in V_h, \tag{7}$$

almost everywhere in $[0, T]$. It is possible to prove, following the same reasoning as the one in [4], that the solution of the problem will enjoy the local extremum decreasing (LED) property. This property is defined in the following lemma:

Lemma 2. *Let u_h be the solution of (7) with $f = 0$ and with the bilinear form $a_h(\cdot,\cdot)$ satisfying the DMP property. Then, any interior local discrete extremum of $|u_h|$ is decreasing in time.*

Proof. Assume there is a local discrete maximum on node x_i in element K. Taking $v_h = \varphi_i^K$ in (7) and using the definition of lumped mass matrix, we have

$$\partial_t u_i^K(t) = -\left(\int_K \varphi_i^K\right)^{-1} a_h(u_h, \varphi_i^K).$$

By the DMP property we know that $a_h(u_h, \varphi_i^K) \geq 0$. Thus, $\partial_t u_i^K(t) \leq 0$ and so, the local discrete maximum is decreasing. The results for the minima follow the same fashion. □

4.1. DMP satisfaction for the 1D IP Method

In this section, we show that the IP method for the Poisson problem (4) enjoys the DMP property in the 1D case for large enough values of c^{ip}. In order to make compact the notation for the proof, we remark that in 1D the facets are the nodes x_i and the integral over the facets reduces to the simple evaluation of the value at that point, thus we define $\llbracket\cdot\rrbracket_i = \llbracket\cdot\rrbracket(x_i)$ and $\{\!\!\{\cdot\}\!\!\}_i = \{\!\!\{\cdot\}\!\!\}(x_i)$. Given a node x_i, we will denote by K_- and K_+ the elements $K_i = [x_{i-1}, x_i]$ and $K_{i+1} = [x_i, x_{i+1}]$ respectively; h_- and h_+ will be their corresponding lengths. Moreover the outside normals are simply $n_- = 1$ and $n_+ = -1$. There will be an abuse of notation in the proof of the proposition in which a binary parameter α is going to be used; it will be $\{-, +\}$ when used as a superscript of a node or subscript of an element and $\{-1, +1\}$ in the rest of the cases

Lemma 3. *The bilinear form (4) with $\xi = 0.5$ (incomplete) enjoys the DMP property if $c^{ip} > 0.5$.*

Proof. We will prove the DMP property assuming that there is a local discrete minimum on x_i in the element K_α either for $\alpha = -$ or $\alpha = +$. The proof for the local discrete maximum case is equivalent. Assuming that there is a minimum on x_i in K_α, we can compute the following jumps and means:

$$\llbracket u_h \rrbracket_i = \alpha |\llbracket u_h \rrbracket_i|, \qquad \llbracket \varphi_i^{K_\alpha} \rrbracket_i = -\alpha, \qquad \llbracket \varphi_{i,x}^{K_\alpha} \rrbracket_i = \frac{1}{h_\alpha}, \tag{8}$$

$$\{\!\!\{ u_{h,x} \}\!\!\}_i = \frac{1}{2} u_{h,x}^{K_{-\alpha}} + \frac{\alpha}{2} |\nabla u_{h,x}^{K_\alpha}|, \qquad \{\!\!\{ \varphi_i^{K_\alpha} \}\!\!\}_i = \frac{1}{2}, \qquad \{\!\!\{ \varphi_{i,x}^{K_\alpha} \}\!\!\}_i = -\frac{\alpha}{2h_\alpha}. \tag{9}$$

Moreover, knowing that $u_h^{K_\alpha}(x_i) \leq u_h(x)$ for any $x \in K_\alpha \cup K_{-\alpha}$, we can deduce that

$$\alpha u_{h,x}^{K_{-\alpha}} \leq |\llbracket u_h \rrbracket_i| h_{K_{-\alpha}}^{-1}.$$

In order to prove that $a_h(\cdot,\cdot)$ enjoys the DMP property we need to prove that there exist $\gamma_i > 0$ and $\delta_{K_\alpha} > 0$ such that $a_h(u_h, \varphi_i^{K_\alpha}) \leq -\gamma_i h_i |[\![u_h]\!]_i| - \delta_{K_\alpha} |u_{h,x}^{K_\alpha}|$. Substituting v_h by $\varphi_i^{K_\alpha}$ in (4) with $\xi = 0.5$:

$$\begin{aligned}
a_h(u_h, \varphi_i^{K_\alpha}) &= \int_{K_\alpha} u_{h,x} \varphi_{i,x}^{K_\alpha} dx - \left[-\alpha \frac{1}{2} u_{h,x}^{K_{-\alpha}} - \frac{1}{2} |u_{h,x}^{K_\alpha}| \right] - \frac{c_i^{ip}}{h_i} |[\![u_h]\!]_i| \\
&\leq -\, |u_{h,x}^{K_\alpha}| + \frac{1}{2h_{K_{-\alpha}}} |[\![u_h]\!]_i| + \frac{1}{2} |u_{h,x}^{K_\alpha}| - \frac{c_i^{ip}}{h_i} |[\![u_h]\!]_i| \\
&\leq -\, \frac{1}{2} |u_{h,x}^{K_\alpha}| - \left(c_i^{ip} - \frac{1}{2} \right) \frac{1}{h_{K_{-\alpha}}} |[\![u_h]\!]_i|.
\end{aligned}$$

Thus, if we define $\delta_{K_\alpha} = 0.5$ and $\gamma_i = (c_i^{ip} - 0.5) h_{K_{-\alpha}} h_i^{-1}$, it is clear that $\delta_K > 0$ and $\gamma_i > 0$ if $c_i^{ip} > 0.5$, as we wanted to prove. □

5. Convection-Diffusion Problem

Considering the original problem (1) we will have to combine the previous terms with the IP terms described in [3] to handle with the convective term $\nabla \cdot (\beta u)$, which basically consists in adding the term

$$a_h^\beta(u_h, v_h) = -\int_\Omega u_h \beta \cdot \nabla v_h + \int_{\mathcal{E}_h^+} \{\!\{\beta u_h\}\!\} [\![v_h]\!] + \int_{\mathcal{E}_h^0} c^{bms} |\beta| [\![u_h]\!] [\![v_h]\!] \tag{10}$$

to the bilinear form and subtracting the term $\int_{\partial\Omega^-} \beta \cdot n_{\partial\Omega} g v_h$ from the right hand side. The set $\mathcal{E}_h^+ = \mathcal{E}_h \backslash \partial\Omega^-$, where $\partial\Omega^- = \{x \in \partial\Omega \,|\, \beta \cdot n_{\partial\Omega}(x) < 0\}$ is the inflow boundary. We use $c_i^{bms} = 0.5$ in our computations, which is equivalent to use the upwind value of βu_h instead of $\{\!\{\beta u_h\}\!\}$ and $c^{bms} = 0$ in (10).

For convection-dominated problems, the solution may present sharp layers, i.e., small intervals in which the value of the solution changes abruptly. The IP method presented before can already control the global instabilities of the solution but it may still present local overshoots and undershoots around sharp layers. In particular, it means that the DMP is violated, so we would like to design a method that ensures a DMP in order to avoid this kind of problems; we will do so by means of an AV. That is, we will compute an extra AV, denoted by ε_h, in each K in such a way that it ensures the DMP; the explicit definition of the AV is introduced in the next section (see Eq. 13). Since the extra viscosity is not consistent, we will not add it in all the terms of the bilinear form, but only in those that are useful for the DMP to be fulfilled.

Putting together the methods described in (4) and (10) with $\xi = 0.5$, using a piecewise constant approximation of μ, given by $\mu_h|_K := h_K^{-1} \int_K \mu dx$, and taking the AV ε_h, we can define the dG problem that we want to solve:

$$\text{Find} \quad u_h \in V_h \quad \text{such that} \quad a_h(u_h, v_h) = l(v_h) \quad \forall v_h \in V_h, \tag{11}$$

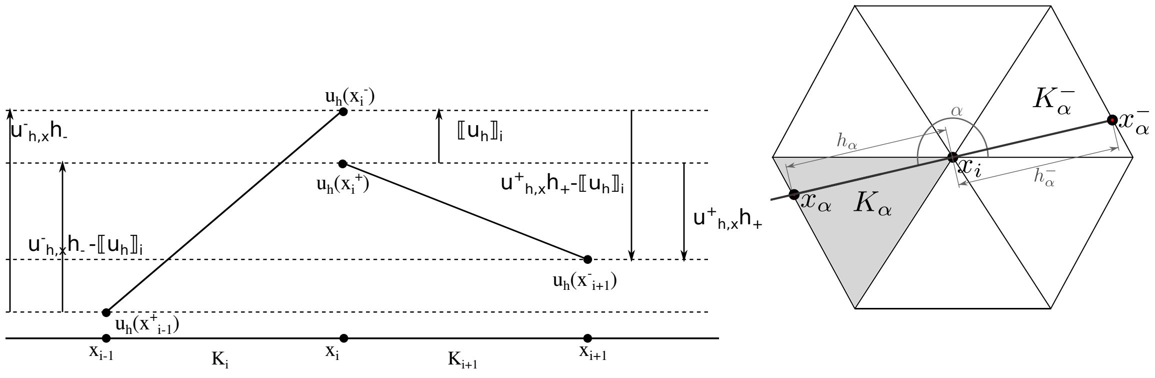
where

$$\begin{aligned}
a_h(u_h, v_h) = &\sum_{K \in \mathcal{T}_h^N} (\mu_h + \varepsilon_K(u_h))(\nabla u_h, \nabla v_h)_K - \sum_{K \in \mathcal{T}_h^N} (u_h, \beta \cdot \nabla v_h)_K \\
&- \int_{\mathcal{E}_h} \mu \{\!\{\nabla u_h\}\!\} [\![v_h]\!] + \int_{\mathcal{E}_h} c^{ip} \tilde{h}^{-1} \langle \mu + \varepsilon_h(u_h) \rangle [\![u_h]\!] [\![v_h]\!] \\
&+ \int_{\mathcal{E}_h^+} \{\!\{\beta u_h\}\!\} [\![v_h]\!] + c^{bms} \int_{\mathcal{E}^0} |\beta| [\![u_h]\!] [\![v_h]\!]
\end{aligned} \tag{12}$$

and

$$l(v_h) = \sum_{K \in \mathcal{T}_h^N} (f, v_h)_K - \int_{\partial\Omega^-} \beta \cdot n_{\partial\Omega} g v_h.$$

Notice that the piecewise μ_h is only used in the volumetric integral of the diffusion term. In the integrals over the facets either μ or $\langle \mu + \varepsilon \rangle$ are used (we recall that $\langle \cdot \rangle$ is the harmonic average defined at the end of section 2).



(a) 1D

(b) 2D

Figure 1:

6. The Artificial Viscosity technique

Now we are ready to design the AV in order to obtain a dG formulation satisfying the DMP property defined above. In particular, we will consider a piecewise constant AV function $\varepsilon_h = \varepsilon_h(u_h)$ such that, when added to μ_h, takes values in a bounded interval $\mu_K + \varepsilon_K := \mu_h|_K + \varepsilon_h|_K \in [0, \Lambda_K]$, where $\Lambda_K = \max\{\nu \|\beta\|_\infty h_K, \mu_K\}$ is the maximum amount of viscosity admitted in an element and $\nu > 0$ is a parameter to be fixed. We want that, if u_h has a local discrete extremum on x_i in K, then $\mu_K + \varepsilon_K = \Lambda_K$. This will be achieved by scaling the AV using a shock detector $s(u_h)$ that will take values in the interval $[0,1]$ with $s = 1$ in K if there is a local discrete extremum in the element. Notice that if $\mu_K + \varepsilon_K = \Lambda_K$ in every element K, the AV would correspond to the suboptimal isotropic diffusion introduced by Von Neumann and Richtmyer in [30] for the continuous case. We will start by designing the detector s for 1D and then we will extend the definition to the multidimensional case.

In order to construct such a shock detector we need to come up with quantities that let us detect where there is a local discrete extremum. Following the same notation as in the proof of Lemma 3, a possible option is, for the point x_i, to consider the values of the jump $\llbracket u_h \rrbracket_i$, the derivatives in K_α and $K_{-\alpha}$, and the corresponding lengths h_α and $h_{-\alpha}$ of the elements. Using these values we can compute the shock detector function s:

$$s_\alpha(x_i) = \left(\frac{\left| u_{h,x}^{K_\alpha} h_\alpha - u_{h,x}^{K_{-\alpha}} h_{-\alpha} + 2\llbracket u_h \rrbracket_i \right|}{\left| u_{h,x}^{K_\alpha} h_\alpha \right| + \left| \llbracket u_h \rrbracket_i - u_{h,x}^{K_{-\alpha}} h_{-\alpha} \right| + \left| \llbracket u_h \rrbracket_i \right|} \right)^q$$

Remark 3. The parameter $q > 0$ is to be chosen. Low values of q improve the nonlinear convergence of the method since the value of $s_\alpha(x_i)$ changes smoothly between nonlinear iterations. On the other hand, high values of q improve the accuracy of the method, since, for $q \longrightarrow \infty$, the detector becomes binary and it only adds extra viscosity in the regions where there are local discrete extrema. Thus, the value of q can be modified during computation time, reducing q to ease nonlinear convergence, or increasing it to have sharper discontinuities at the expense of more CPU cost.

With the previous definition, it is easy to see that s fulfils the following property:

Lemma 4. *Given a node $x_i \in \mathcal{E}^0$ and $u_h \in V_h$, the detector $s_\alpha(x_i) = s_\alpha(u_h, x_i)$ takes values in the interval $[0,1]$ and $s_\alpha(x_i) = 1$ if and only if u_h has a local discrete extremum on x_i in K_α.*

Proof. First of all, it is obvious that $s_\alpha(x_i) \leq 1$. Then, we notice that $s_\alpha(x_i) = 1$ iff the sign of $u_{h,x}^{K_\alpha} h_\alpha$, $(\llbracket u_h \rrbracket_i - u_{h,x}^{K_{-\alpha}} h_{-\alpha})$ and $\llbracket u_h \rrbracket_i$ are the same. Observing Fig. 1(a) is easy to see that these three values correspond to $\alpha(u_h^{K_+}(x_{j+1}) - u_h^{K_\alpha}(x_i))$, $\alpha(u_h^{K_-}(x_{j-1}) - u_h^{K_\alpha}(x_i))$, and $\alpha(u_h^{K_{-\alpha}}(x_i) - u_h^{K_\alpha}(x_i))$, not necessarily in that order. So, by the definition of local discrete extremum, it is clear that these three values will have the same sign iff there is a local discrete extremum on x_i in K_α. □

For x_i and $K \subset \Omega_i$, let us define the value Γ_i^K to be such that $\Gamma_i^K = \alpha$ if K is the element K_α with respect to x_i. Then, we can define the AV of the problem as:

$$\varepsilon_K(u_h) = \max\{0, \nu\|\beta\|_\infty h_K \max_{x_i \in K}\{s_{\Gamma_i^K}(x_i)\} - \mu_K\}. \tag{13}$$

Theorem 5. *The semilinear form $a_h(\cdot,\cdot)$ described in (12) with ε_h as in (13) and $c^{bms} = 0.5$ enjoys the DMP property for any value of $q > 0$ if $\nu > 1$ and $c_i^{ip} > 0.5$.*

Proof. Let us use the same notation as in Lemma 3. We use the fact that if u_h has a local discrete extremum on x_i in K_α, then $\mu_{K_\alpha} + \varepsilon_{K_\alpha} = \Lambda_{K_\alpha}$. Using the identities in (8) and integrating by parts the convective term, we get:

$$\begin{aligned}
a_h(u_h, \varphi_i^{K_\alpha}) =& \Lambda_{K_\alpha} \int_{K_\alpha} u_{h,x}\varphi_{i,x}^{K_\alpha} dx + \int_{K_\alpha} \beta u_{h,x}\varphi_i^{K_\alpha} - \left[\beta u_h \varphi_i^{K_\alpha} \cdot n_{\partial K}\right]_{\partial K} \\
& - \mu\left[-\alpha\frac{1}{2}u_{h,x}^{K_{-\alpha}} - \frac{1}{2}|u_{h,x}^{K_\alpha}|\right] - \langle \mu + \varepsilon(u_h)\rangle_i \frac{c_i^{ip}}{h_i}|[\![u_h]\!]_i| \\
& - \frac{\alpha\beta(x_i)}{2}(u_h^{K_-}(x_i) + u_h^{K_+}(x_i)) - \frac{1}{2}|\beta(x_i)||[\![u_h]\!]_i| \\
\leq & -\Lambda_{K_\alpha}|u_{h,x}^{K_\alpha}| + \frac{1}{2}\|\beta\|_{\infty,K_\alpha} h_{K_\alpha}|u_{h,x}^{K_\alpha}| + \alpha\beta(x_i)u_h^{K_\alpha}(x_i) + \frac{\mu}{2h_{K_{-\alpha}}}|[\![u_h]\!]_i| \\
& + \frac{\mu}{2}|u_{h,x}^{K_\alpha}| - \frac{c_i^{ip}}{h_i}\mu|[\![u_h]\!]_i| - \frac{\alpha\beta(x_i)}{2}(u_h^{K_\alpha}(x_i) + u_h^{K_{-\alpha}}(x_i)) - \frac{1}{2}|\beta(x_i)||[\![u_h]\!]_i| \\
=& \left(-\Lambda_{K_\alpha} + \frac{1}{2}\|\beta\|_{\infty,K_\alpha} h_{K_\alpha} + \frac{\mu}{2}\right)|u_{h,x}^{K_\alpha}| + \left(\frac{\mu}{2h_{K_{-\alpha}}} - \frac{c_i^{ip}}{h_i}\mu - \frac{1}{2}|\beta(x_i)| - \frac{\alpha}{2}\beta(x_i)\right)|[\![u_h]\!]_i| \\
\leq & -(\nu - 1)\frac{1}{2}\|\beta\|_{\infty,K_\alpha} h_{K_\alpha}|u_{h,x}^{K_\alpha}| - \left(c_i^{ip} - \frac{1}{2}\right)\frac{1}{h_{K_{-\alpha}}}\mu|[\![u_h]\!]_i|.
\end{aligned}$$

Thus, if we define $\delta_K = 0.5\,(\nu - 1)\,\|\beta\|_{\infty,K_\alpha}$ and $\gamma_i = \left(c_i^{ip} - 0.5\right) h_i h_{K_{-\alpha}}^{-1}\langle\Lambda_K\rangle$, it is clear that $\delta_K > 0$ if $\nu > 1$ and $\gamma_i > 0$ if $c_i^{ip} > 0.5$, as we wanted to prove. □

If one is interested in recovering the symmetric or nonsymmetric form, it is possible to weight the extra term using the same shock capturing in such a way that the term vanishes in the facets around the elements with a local discrete extremum inside. For the symmetric term we define

$$\tilde{\xi}(x_i) = 0.5 \max_{K \subset \Omega_i} s_{\Gamma_i^K}$$

(resp., $\tilde{\xi}(x_i) = 1 - 0.5\max_{K\subset\Omega_i} s_{\Gamma_i^K}$ for the nonsymmetric term). Then, the weighted symmetric bilinear form would read:

$$\begin{aligned}
\tilde{a}_h(u_h, v_h) = & \sum_{K \in \mathcal{T}_h^N} (\mu_K + \varepsilon_K(u_h))(\nabla u_h, \nabla v_h)_K - \sum_{K \in \mathcal{T}_h^N} (u_h, \beta \cdot \nabla v_h)_K \\
& - \int_{\mathcal{E}_h} \mu \{\!\{\nabla u_h\}\!\}[\![v_h]\!] - \int_{\mathcal{E}_h} (1 - \tilde{\xi})\mu[\![u_h]\!]\{\!\{\nabla v_h\}\!\} + \int_{\mathcal{E}_h} c^{ip}\tilde{h}^{-1}\langle \mu + \varepsilon_h(u_h)\rangle[\![u_h]\!][\![v_h]\!] \\
& + \int_{\mathcal{E}_h^+} \{\!\{\beta u_h\}\!\}[\![v_h]\!] + c^{bms}\int_{\mathcal{E}^0} |\beta|[\![u_h]\!][\![v_h]\!].
\end{aligned} \tag{14}$$

This form is closer to the original symmetric scheme ($\xi = 1$), but it is not symmetric unless the shock detector is not activated.

Corollary 6. *The weighted semilinear form $\tilde{a}_h(\cdot,\cdot)$ described in (14), with ε_h as in (13) and $c^{bms} = 0.5$, enjoys the DMP property for any value of $q > 0$ if $\nu > 1$ and $c_i^{ip} > 0.5$.*

Proof. Since the term $(1 - \tilde{\xi}(x_j))$ nullifies for $x_j \in K$ if there is a maximum in K, the term $\int_{\mathcal{E}_h} (1 - \tilde{\xi})\mu [\![u_h]\!] \{\!\{\nabla \varphi_h\}\!\}$ does not add any contribution to $\tilde{a}_h(u_h, \varphi_i^K)$. Thus, the results hold from the proof of Theorem 5. □

The results of such technique are shown in the section 7.

6.1. Extension to the multidimensional case

Let us consider the multidimensional convection-diffusion problem (1). It is possible to extend the nonlinear stabilization to the multidimensional case by generalising the computation of the AV. Even though it is unclear whether the multidimensional IP dG methods for the Poisson equation satisfy any DMP property, the underlying idea behind the multidimensional nonlinear stabilization design is similar to what was proposed in [2]. For each element K in the mesh we must compute the amount of AV $\varepsilon_K = \varepsilon|_K$ which will be scaled according to a shock detector $s \in [0, 1]$ that takes value $s_K = 1$ if there is a local discrete extremum in K.

First of all, we must extend the definition of the shock detector function $s_\alpha(x_i)$ to the multidimensional case. The definition will be done in two dimensions for simplicity, but it can be easily extended to any space dimension. As it can be observed in Fig. 1(b), in the two-dimensional case, α is not a binary parameter, but it corresponds to an angle, $\alpha \in [0, 2\pi)$, and it gives a certain direction, $r_\alpha = (\cos\alpha, \sin\alpha)$. Moreover the notation $\alpha^- = \alpha - \pi$ will be used to refer the opposite sense to α. The idea is to redefine the parameters used above to compute $s_\alpha(x_i)$ by projecting the solution in the direction r_α around the node x_i (see Fig. 1(b)).

In this sense, $K_\alpha = \{K \subset \Omega_i \,|\, \exists\, \delta > 0 \,:\, x_i + \delta r_\alpha \in K\}$ and $K_\alpha^- = K_{\alpha^-}$. Then, let $x_\alpha \in \partial K_\alpha$ and $h_\alpha > 0$ be such that $x_\alpha - x_i = h_\alpha r_\alpha$ (h_α^- and x_α^- defined similarly for α^-); see Fig. 1(b) for an illustration. These parameters are uniquely defined unless the direction r_α coincides with the direction of one of the edges of the mesh but these directions are not required in the definition of s_K below. Finally we define $[\![u_h]\!]_i^\alpha = u_h^{K_\alpha^-}(x_i) - u_h^{K_\alpha}(x_i)$. So, we can redefine $s_\alpha(x_i)$ as:

$$s_\alpha(x_i) = \left(\frac{\left| \nabla u_h^{K_\alpha} \cdot r_\alpha h_\alpha - \nabla u_h^{K_\alpha^-} \cdot r_\alpha h_\alpha^- + 2[\![u_h]\!]_i^\alpha \right|}{\left| \nabla u_h^{K_\alpha} \cdot r_\alpha h_\alpha \right| + \left| [\![u_h]\!]_i^\alpha - \nabla u_h^{K_\alpha^-} \cdot r_\alpha h_\alpha^- \right| + \left| [\![u_h]\!]_i^\alpha \right|} \right)^q.$$

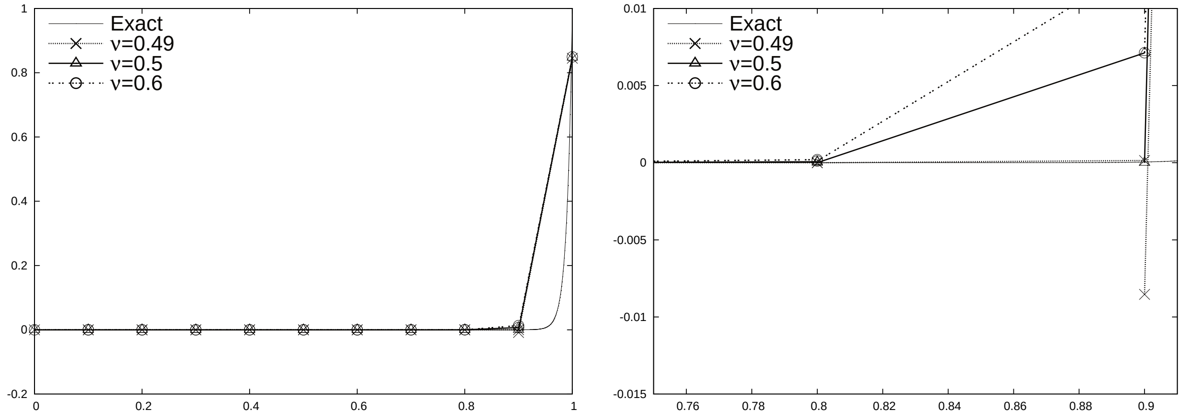
Following the proof of lemma 4 and noting that $u_h^{K_\alpha}(x_\alpha) - u_h^{K_\alpha}(x_i) = \nabla u_h^{K_\alpha} \cdot r_\alpha h_\alpha$, it can be proved that this shock detector takes values $s \in [0, 1]$ and that $s_\alpha(x_i) = 1$ if and only if u_h has a local discrete extremum on x_i in the direction r_α. Then, if we consider a node x_i and an element $K \subset \Omega_i$, we let Γ_i^K be the interval such that if $\alpha \in \Gamma_i^K$, $K_\alpha = K$. It is easy to see that if the function u_h has a local discrete extremum on x_i in K then $s_\alpha(x_i) = 1\ \forall \alpha \in \Gamma_i^K$. So it is possible to define the elemental shock detector as:

$$s_K = \max_{x_i \in K} \inf_{\alpha \in \Gamma_i^K} s_\alpha(x_i).$$

Notice that $s_\alpha(x_i) = 1\ \forall \alpha \in \Gamma_i^K$ does not necessarily imply that there is a local discrete extremum on x_i in K. This property is natural, since local instabilities can appear on nodes that are not local discrete extrema, e.g., on shock fronts.

On the other hand, given the cost of computing $\inf_{\alpha \in \Gamma_i^K} s(x_i^\alpha)$, we can avoid its computation by taking the minimum with respect to edge directions only (at both sides of the edge). This simplification leads to a very slightly different method, but the simplified definition still enjoys the property that $s_K = 1$ if u_h has a local discrete extremum in K.

Remark 4. We have designed a shock detector that ensures that there is the maximum amount of viscosity around a local discrete extremum. However, it is unclear how to prove the DMP for the multidimensional case, since it is not even available for the Laplacian problem. (It is due to the sign of the IP term, that cannot be determined.) In any case, looking at the results in Section 7, the DMP holds in practice for the same mesh conditions stated in [2].



(a) $\nu = 0.49, 0.5, 0.6$

(b) Detail of Fig. 2(a)

Figure 2:

7. Numerical results

For the following test, if nothing is said, the value of the parameters used will be $q = 10$, $c^{ip} = 10$ and $\nu = 0.5$. The choice of the last two parameters is explained in the first numerical experiment.

7.1. Sharpness of the parameters

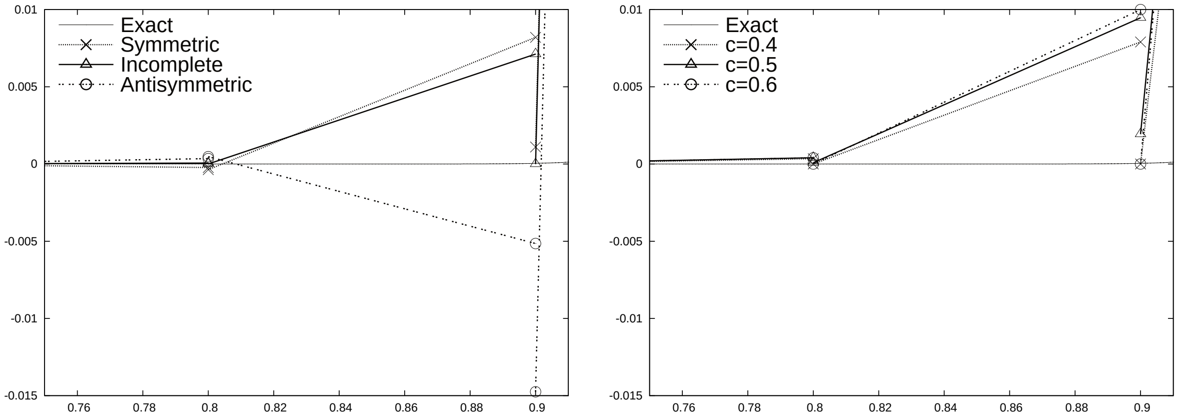
In order to check if the choice of the parameters is sharp, we will solve the problem:

$$\begin{cases} -0.01u_{xx} + u_x &= 0 \quad \text{in} \quad \Omega = (0,1) \\ u(0) &= 0 \\ u(1) &= 1 \end{cases}$$

using a mesh of $N = 10$ elements. The solution of the problem presents a sharp slope near the boundary $x = 1$. The solution with $N = 1000$ is plotted as a reference solution. Let us check the sharpness of the bound for ν such that the formulation satisfies a DMP (see Theorem 5). In previous works, using similar schemes with continuous finite element methods (see [2, 4]) the condition for the method to enjoy a DMP was $\nu \geq 0.5$. In this case, if we look at the proof of Theorem 5, and assuming that $\mu < \|\beta\|h$, we can sharpen the value of ν to be $\nu > 0.5 + \mu\|\beta\|^{-1}h^{-1}$ (instead of $\nu > 1$ as stated in the theorem) which in this particular case means $\nu > 0.55$. But, when varying the values of ν and testing the violation of the DMP, we observed that the threshold is still on $\nu = 0.5$ as it can be observed in Fig. 2. For $\nu = 0.49$ the DMP is violated while for $\nu = 0.5$ the violation is in the order of the machine precision. For this reason, for the next tests we will use the value $\nu = 0.5$. The results plotted in Fig. 2 are obtained using $c^{ip} = 10$.

It is also possible to see how different choices of the parameter ξ lead to a violation of the DMP. We recall that ξ could take values $\{0, 0.5, 1\}$, corresponding to the symmetric, incomplete, and nonsymmetric scheme. Without weighting, our analysis works for the incomplete scheme only, and, as it can be checked in Fig. 3(a), this is the only scheme that ensures the DMP. Finally, we vary the value of c^{ip}. Theorem 5 states that $c^{ip} > 0.5$ for the method to enjoy the DMP, but the method still satisfies the DMP for $c^{ip} = 0.4$ (see Fig. 3(b)). In any case, we prefer the use of $c^{ip} = 10$ which is common in the literature.

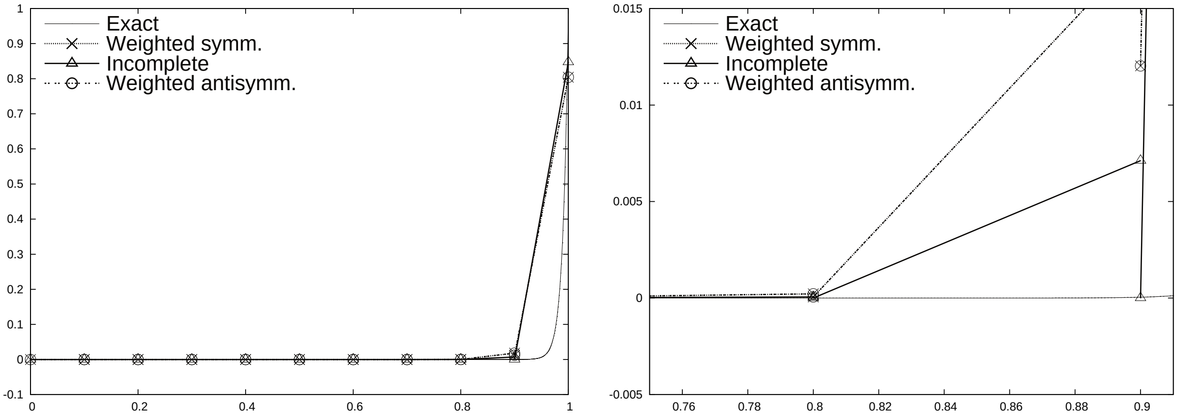
As it was proved in the Corollary 6, it is possible to consider a weighted symmetric/antisymmetric IP formulation by weighting the value of ξ with the shock detector. The results obtained with such method are plotted in Fig. 4. It can be appreciated how the DMP is effectively enjoyed by all the methods, being the incomplete the most accurate one. Moreover, the nonlinear convergence is much faster in the incomplete case since the weighting introduce extra nonlinearity to the problem.



(a) $\xi = 0, 0.5, 1$

(b) $c^{ip} = 0.4, 0.5, 0.6$

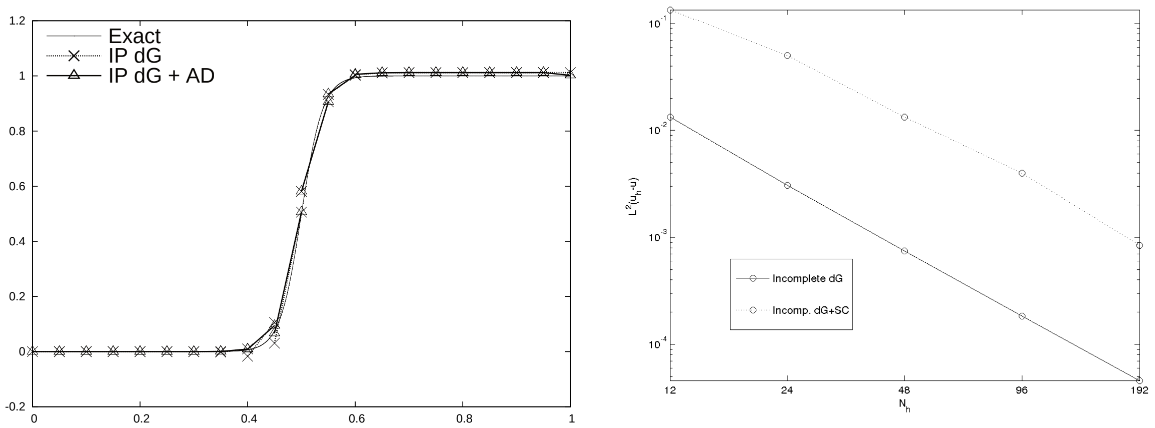
Figure 3:



(a)

(b) detail of Fig. 4(a)

Figure 4: $\xi(x_i) = 0.5\xi \max_K s_{\Gamma_i^K}$



(a) Solution of (15). $N = 20$.

(b) Smooth convergence

Figure 5:

7.2. 1D advection case

The next problem is a 1D first-order diferential equation ($\mu = 0$) inspired by the numerical example in [15]:

$$u_x = \frac{1}{2\epsilon}\left(1 - \tanh^2\left(\frac{x-0.5}{\epsilon}\right)\right), \qquad \text{in}\,\Omega = (0,1), \tag{15}$$

with $u(0) = 0$. The solution to this problem is $u = 0.5\,(\tanh\left((x-0.5)/\epsilon\right) + 1)$; it shows a sharp layer around $x = 0.5$. Since the source is positive in the whole domain, we expect the stabilized method not to have any local discrete minimum in Ω. In Fig. 5(a), it can be appreciated how the DMP is violated for the IP dG method without extra viscosity while it is not when the appropiate viscosity is added. Moreover the layer is captured with the same amount of elements in both cases, so the method is not too much diffusive.

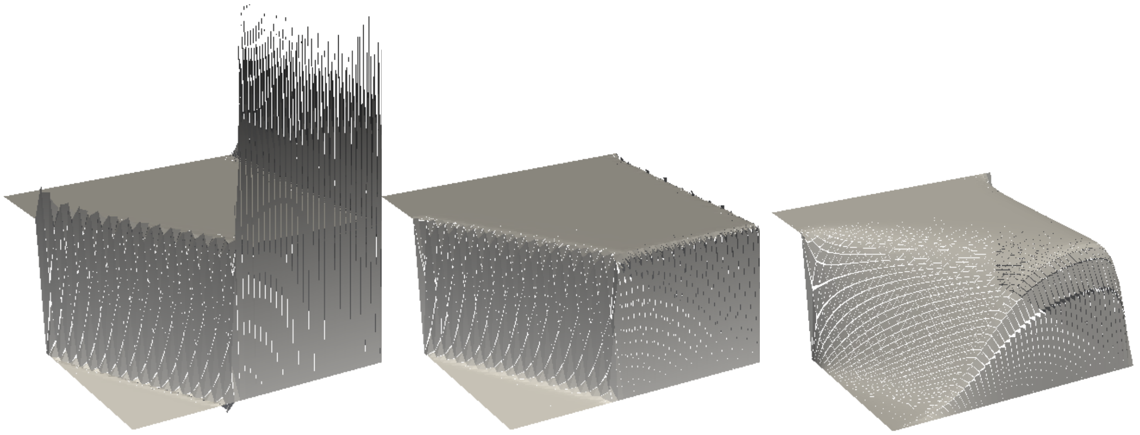
7.3. Convergence of a smooth solution

We would like to see that the L^2 convergence of the method towards an smooth solution is not affected by the activation of the AV. To test so, we consider the equation (1) with $\mu = 0$, $\beta = (1,1)$, $f(x,y) = \cos(2\pi x)$ in $\Omega = [0,1]^2$ and boundary conditions on the inflow boundary given by $u(x,y) = \sin(2\pi x)$, which is the exact solution. Since this solution presents maxima and minima in the x direction on the lines $x = 0.25$ and $x = 0.75$, the shock detector is activated and AV is added. The problem is solved in triangular meshes of $N_h \times N_h(\times 2)$ elements with size $N_h = 12, 24, 48, 96, 192$. Even though the presence of the AV increases the error in $L^2(\Omega)$ norm (as expected), the convergence of order 2 is maintained, as it can be appreciated in Fig. 5(b)

7.4. Propagation of discontinuities

A typical steady-state test is to study the propagation of a boundary discontinuity for a convection-dominated equation. To do so, problem (11) is solved in $\Omega = [0,1]^2$ with $\mu = 10^{-8}$, $\beta = (\cos(-\pi/3), \sin(-\pi/3))$, $f = 0$ and the boundary conditions $g = 1$ on $y = 1$ and on $\{x = 0\} \cap \{y \geq 0.7\}$ and $g = 0$ on the rest of the boundary. The solution of this problem has two boundary layers (on $x = 1$ and $y = 0$) and an interior layer of width proportional to the value of μ. The mesh used in this case consists of $48 \times 48(\times 2)$ triangles.

The AV definition depends on the solution, ending up with a nonlinear stabilization term. For transient problems, it can be considered in a semi-implicit way (taking the value from the previous time step without nonlinear iterations), but for the steady state case nonlinear iterations are needed. In order to improve convergence we have used the damping parameter $\omega \in [0.01, 1]$ proposed in [18]. In any case,



(a) Incomplete dG scheme (b) dG+AD, $q = 10$ (semi-implicit) (c) dG+AD, $q = 0.1$

Figure 6: Propagation of a discontinuity. $48 \times 48(\times 2)$ triangular mesh. $c^{ip} = 10$, $c^{bms} = 0.5$, $\nu = 0.5$

for $q = 10$, the nonlinear error got stuck in values of order 10^{-2}-10^{-3}. However, the results in Fig. 6(b) are very good, with a violation of the DMP of order 10^{-3}, clearly damping the initial sharp oscillations obtained by the original method (see Fig. 6(a)).

This nonconvergence problem is common to AV due to the fact that, for high values of q, the shock detector values are almost binary (very close to 0 or 1) and the activation and deactivation of the shock detector s may become cyclic. That is what happens in this case when using $q = 10$.

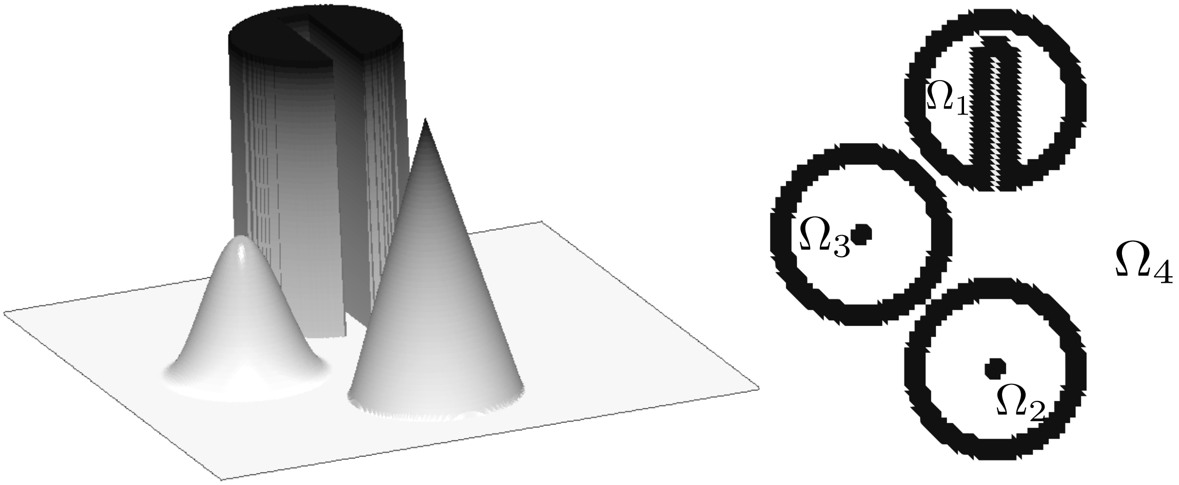
This behavior is explained by the fact that, even though the shock capturing term proposed herein is Lipschitz continuous, the Lipschitz constant blows up as $q \to \infty$. However, a main difference with respect to the popular non-differentiable flux correction transport schemes (see [19]) is that we can control this Lipschitz constant by decreasing the value of q. As an example, taking $q = 10\omega$ (we recall that $\omega \in [0.01, 1]$ is the damping parameter used to improve the nonlinear convergence and it is recalculated in each iteration), it is possible to converge with tolerance 10^{-8} with respect to the norm of the increment between iterations. In this last case we have lost accuracy in order to improve nonlinear convergence, but the method does fulfil the DMP exactly. The drawback of reducing q are less sharp fronts. In fact, in the limit case $q = 0$ we recover the sub-optimal Von Neumann-Richtmyer isotropic AV in the whole domain (see Fig. 6(c)). Ellaborated q-adaptive nonlinear algorithms could be considered, based on these observations.

7.5. Multidimensional transport problem

Finally, a pure convection problem will be used in order to show the performance of the method; let us note that the present numerical analysis does apply for the case $\mu = 0$. The test consists in solving the two dimensional transport problem $\partial_t u + \nabla \cdot (\beta u) = 0$ in $[0, 1]^2 \times [0, T]$, $\beta = (-2\pi(y - 0.5), 2\pi(x - 0.5))$. The solution of this equation is a displacement of the initial solution around the center of the square. The solution is computed at $T = 1$ after a complete cycle. The initial solution is given in [21] and its interpolation in a mesh of 250×250 bilinear elements is displayed in Fig. 7(a).

The solution is discretised with a $100 \times 100(\times 2)$ triangular mesh and the integration in time is performed using Crank-Nicolson with time step $\Delta t = 2.5 \cdot 10^{-4}$ and without mass lumping. The result can be compared with the original dG method and also with the results obtained in a finer mesh in the continuous case, using a similar shock capturing designed in [2] and denoted as boundary gradient jump viscosity (bGJV). If the reader is interested in comparing the results with some other state-of-art methods, like residual-based and entropy-viscosity methods, we refer to [2], where the results for the same tests are provided.

For the continuous case, the mesh consists of $250 \times 250(\times 2)$ elements and more than $60,000$ nodes, which is the number of nodes of the discontinuous case. The AV is computed semi-implicitly. The results



(a) Nodally exact initial solution

(b) Error regions

Figure 7: Multidimensional transport problem in a 250 × 250(×2) mesh

Method	Ω_1	Ω_2	Ω_3	Ω_4
Incomplete dG	6.672e-002	2.331e-003	1.329e-004	1.027e-002
Incomplete dG + shock capturing	7.452e-002	4.355e-003	1.480e-003	1.712e-002

Table 1: $\|u - u_h\|_{\Omega_i}$

are displayed in Fig. 8 and it can be observed that, in comparison with the original result, the AV does a good job removing the oscillations on the top of the cylinder. On the other hand, it damps the solution around the top of the cone and the hump by activating unnecessarily the viscosity. This problem is avoided by refining the mesh as it is the case for the continuous method (see Fig. 8(d)).

The L^2 error after one cycle has been computed in 4 different regions, namely Ω_1, Ω_2, Ω_3, Ω_4, corresponding to the ones plotted in Fig. 7(b). The first three regions correspond to the elements of the mesh such that the centroid of their Gauss points are at distance $2 \cdot 10^{-2}$ or less to a discontinuity or change of gradient of the exact solution. The region Ω_4 is simply the rest of the domain.

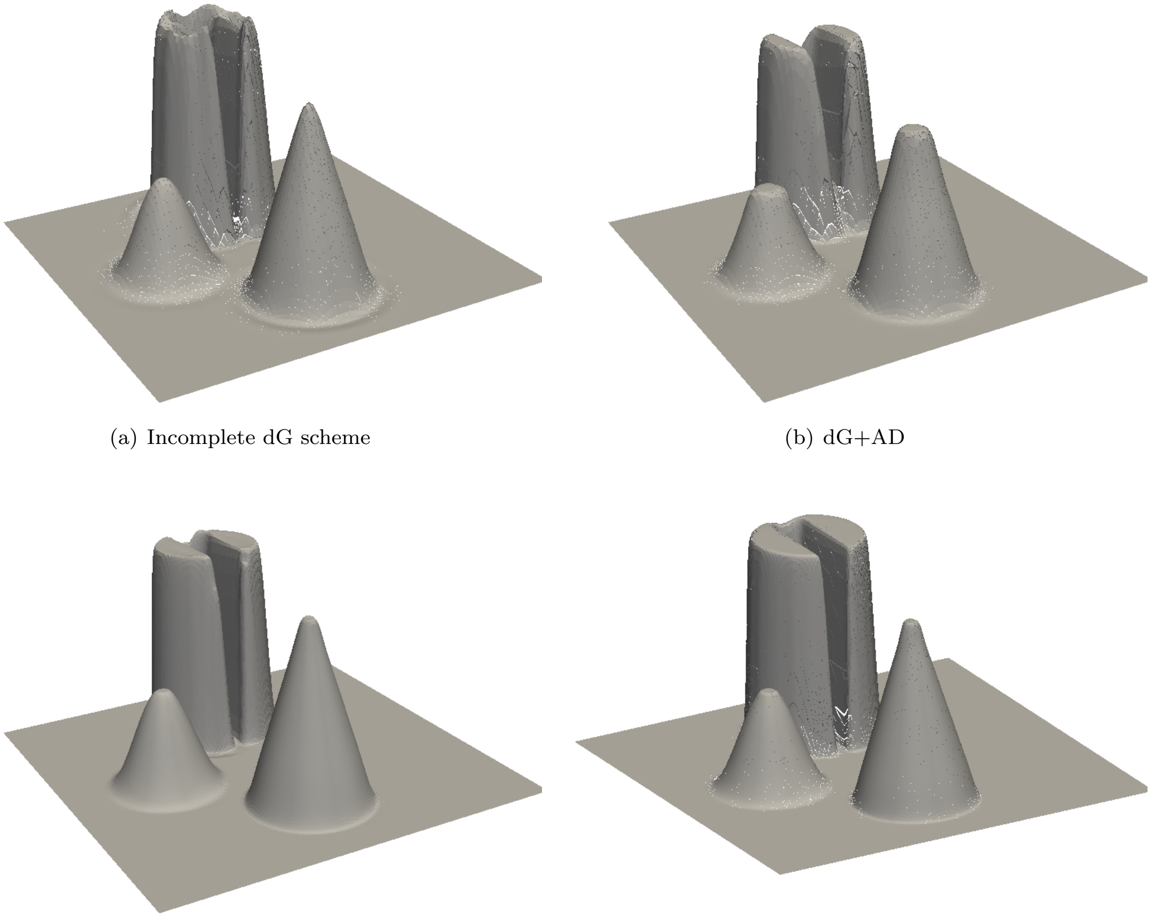
The results reported in Fig. 8 are at the final stage of the computation, i.e., $t = 1$, and the oscillations have already been smoothed out. In order to better evaluate how the different methods succeed eliminating oscillations, we introduce the oscillation function

$$\text{osc}(t) = \max_{(x,y)\in\Omega} \{0, u_h(x,y,t) - 1, -u_h(x,y,t)\}.$$

We compute the mean value of the parameter osc(t) in bunches of 50 time steps and the time evolution of this quantity for the different methods is plotted in Fig. 9. Clearly, the nonlinearly stabilized dG formulation defined herein beats by far the other two methods in terms of DMP violation; after the first 50 steps the method has already reduced the oscillation below 10^{-3}.

8. Conclusions

When considering dG methods for steady problems or transient problems via (semi-)implicit time integration, the use of traditional limiting techniques is not suitable in many instances. In this work, we propose dG formulations that satisfy some kind of monotonicity properties based on implicit nonlinear stabilization (AV-type terms). The analysis of monotonicity properties and discrete maximum principles in the frame of dG formulations (without additional postprocessing) is a quite unexplored area; as far as we know there are only some attempts to prove a DMP for the Laplacian problem in 1D (see [17]). For this reason, we have started our work defining the notion of local discrete extrema in dG; since



(a) Incomplete dG scheme

(b) dG+AD

(c) cG+SUPG+bGJV

(d) dG+AD finer mesh

Figure 8: Results at $t = 1.0$. Space discretization with $p = 1$ and ≈ 60000 DOFs ($100 \times 100(\times 2)$ and $250 \times 250(\times 2)$ triangular mesh). Time discretization with Crank-Nicolson, $\Delta t = 2.5 \cdot 10^{-4}$ and without mass lumping. Results for different schemes at $t = 1$.

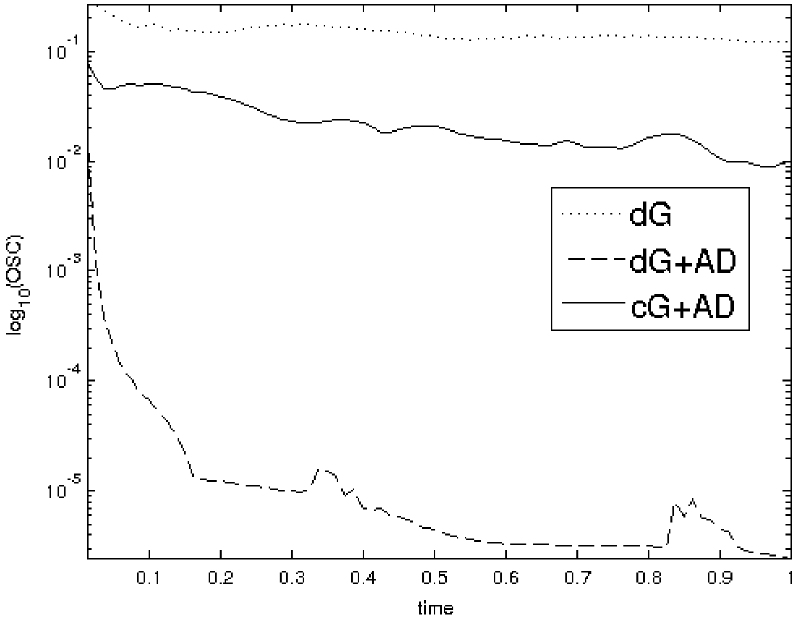


Figure 9: OSC evolution

the numerical solution is discontinuous on nodes, this concept is somehow open. Next, we propose a definition of DMP property for dG, and show that when the dG formulation enjoys this property, the maximum/minimum is on the boundary (given a negative/positive forcing term) for steady problems in the multidimensional case. Further, the method is LED for transient problems. Further we show that for the 1D Poisson problem, the incomplete IP dG formulation satisfies the DMP property. In order to make symmetric/antisymmetric IP versions to enjoy the DMP, a weighted version of these formulations is also proposed.

Next, we tackle convection-diffusion and transport problems. The dG formulation we consider is the IP method (see [1, 17]) for the viscosity term together with the advection stabilization proposed in [3]. On top of this dG formulation, we add a novel nonlinear stabilization (shock capturing) term, based on jumps of the unknown and its derivatives. As soon as the dG discretization of the Laplacian term satisfies the DMP (see above), we prove that the resulting dG method also satisfies the DMP property in 1D. It implies no overshoots/undershoots around sharp layers or discontinuities.

The formulation is extended to multi-dimensional problems, and applied to different test problems. Out of these results, we show that we have the monotonic properties predicted by the theory in 1D. In multi-dimension, the method does an excellent job reducing local oscillations, as expected. For time-dependent problems, we have considered semi-implicit formulations (computing the AV with the solution of the previous time step). As other shock capturing techniques, when the shock sensor is very sensitive, i.e., it acts in an almost binary fashion, nonlinear convergence is hard to get. However, the definition of the shock-capturing proposed herein includes a numerical parameter, q, that allows one to control the Lipschitz constant, improving nonlinear convergence by reducing q.

Acknowledgments

This work has been partially funded by the European Research Council under the FP7 Programme Ideas through the Starting Grant No. 258443 - COMFUS: Computational Methods for Fusion Technology. A. Hierro gratefully acknowledges the support received from the Catalan Government through a FI fellowship.

References

[1] D. N. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM Journal on Numerical Analysis*, 39(5):1749–1779, 2002.

[2] S. Badia and A. Hierro. On monotonicity-preserving stabilized finite element approximations of transport problems. *SIAM Journal on Scientific computing*, in press.

[3] F. Brezzi, L. D. Marini, and E. Süli. Discontinuous Galerkin methods for first-order hyperbolic problems. *Mathematical Models and Methods in Applied Sciences*, 14(12):1893–1903, 2004.

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

[5] E. Burman and A. Ern. Discrete maximum principle for Galerkin approximations of the Laplace operator on arbitrary meshes. *Comptes Rendus Mathematique*, 338(8):641–646, 2004.

[6] E. Burman and A. Ern. Stabilized Galerkin approximation of convection-diffusion-reaction equations: discrete maximum principle and convergence. *Mathematics of computation*, 74(252):1637–1652, 2005.

[7] E. Burman and P. Hansbo. Edge stabilization for Galerkin approximations of convection-diffusion-reaction problems. *Computer Methods in Applied Mechanics and Engineering*, 193(1516):1437–1453, 2004.

[8] P. Ciarlet and P.-A. Raviart. Maximum principle and uniform convergence for the finite element method. *Computer Methods in Applied Mechanics and Engineering*, 2(1):17–31, 1973.

[9] P. G. Ciarlet. Discrete maximum principle for finite-difference operators. *Aequationes mathematicae*, 4(3):338–352, 1970.

[10] B. Cockburn, S. Hou, and C.-W. Shu. The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws IV: The multidimensional case. *Mathematics of Computation*, 54(190):545–581, 1990.

[11] B. Cockburn and C.-W. Shu. The Runge-Kutta discontinuous Galerkin method for conservation laws V: Multidimensional systems. *Journal of Computational Physics*, 141(2):199–224, 1998.

[12] B. Cockburn and C.-W. Shu. Runge-Kutta discontinuous Galerkin methods for convection-dominated problems. *Journal of Scientific Computing*, 16(3):173–261, 2001.

[13] R. Codina. A discontinuity-capturing crosswind-dissipation for the finite element solution of the convection-diffusion equation. *Computer Methods in Applied Mechanics and Engineering*, 110(34):325–342, 1993.

[14] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*, volume 224. springer, 2001.

[15] J.-L. Guermond. Subgrid stabilization of Galerkin approximations of linear monotone operators. *IMA Journal of Numerical Analysis*, 21(1):165–197, 2001.

[16] W. Höhn and H. D. Mittelmann. Some remarks on the discrete maximum-principle for finite elements of higher order. *Computing*, 27(2):145–154, 1981.

[17] T. L. Horváth and M. E. Mincsovics. Discrete maximum principle for interior penalty discontinuous Galerkin methods. *Central European Journal of Mathematics*, 11(4):664–679, 2013.

[18] V. John and P. Knobloch. On spurious oscillations at layers diminishing (SOLD) methods for convection-diffusion equations: Part {II} - Analysis for and finite elements. *Computer Methods in Applied Mechanics and Engineering*, 197(2124):1997 – 2014, 2008.

[19] D. Kuzmin. On the design of general-purpose flux limiters for finite element schemes. I. scalar convection. *Journal of Computational Physics*, 219(2):513 – 531, 2006.

[20] D. Kuzmin. On the design of algebraic flux correction schemes for quadratic finite elements. *Journal of Computational and Applied Mathematics*, 218(1):79–87, 2008.

[21] D. Kuzmin. A guide to numerical methods for transport equations. 2010.

[22] D. J. Lorenz. Zur inversmonotonie diskreter probleme. *Numerische Mathematik*, 27(2):227–238, 1977.

[23] A. Mizukami and T. J. Hughes. A Petrov-Galerkin finite element method for convection-dominated flows: An accurate upwinding technique for satisfying the maximum principle. *Computer Methods in Applied Mechanics and Engineering*, 50(2):181–193, 1985.

[24] H. Nagarajan and K. B. Nakshatrala. Enforcing the non-negativity constraint and maximum principles for diffusion with decay on general computational grids. *International Journal for Numerical Methods in Fluids*, 67(7):820847, 2011.

[25] G. Payette, K. Nakshatrala, and J. Reddy. On the performance of high-order finite elements with respect to maximum principles and the nonnegative constraint for diffusion-type equations. *International Journal for Numerical Methods in Engineering*, 91(7):742771, 2012.

[26] H.G. Roos, M. Stynes and L. Tobiska. Robust numerical methods for singularly perturbed differential equations *Springer Ser. Comput. Math*, 24,2008.

[27] R. S. Varga. On a discrete maximum principle. *SIAM Journal on Numerical Analysis*, 3(2):355–359, 1966.

[28] T. Vejchodský. Higher-order discrete maximum principle for 1D diffusion-reaction problems. *Applied Numerical Mathematics*, 60(4):486–500, 2010.

[29] T. Vejchodský and P. Šolín. Discrete maximum principle for a 1D problem with piecewise-constant coefficients solved by hp-FEM. *Journal of Numerical Mathematics*, 15(3), 2007.

[30] J. Von Neumann and R. D. Richtmyer A method for the numerical calculation of hydrodynamic shocks *Journal of applied physics*, 21(3):232–237, 1950.

[31] E. Yanik. A discrete maximum principle for collocation methods. *Computers & Mathematics with Applications*, 14(6):459–464, 1987.

[32] E. Yanik. Sufficient conditions for a discrete maximum principle for high order collocation methods. *Computers & Mathematics with Applications*, 17(11):1431–1434, 1989.

[33] X. Zhang and C.-W. Shu. On maximum-principle-satisfying high order schemes for scalar conservation laws. *Journal of Computational Physics*, 229(9):3091–3120, 2010.

[34] X. Zhang and C.-W. Shu. Maximum-principle-satisfying and positivity-preserving high-order schemes for conservation laws: Survey and new developments. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 467(2134):2752–2776, 2011.

[35] X. Zhang, Y. Xia, and C.-W. Shu. Maximum-principle-satisfying and positivity-preserving high order discontinuous Galerkin schemes for conservation laws on triangular meshes. *Journal of Scientific Computing*, 50(1):29–62, 2012.