**Noname manuscript No.**
(will be inserted by the editor)

# A high-order augmented basis positivity-preserving discontinuous Galerkin method for a Linear Hyperbolic Equation

**Maurice S. Fabien**

Received: date / Accepted: date

**Abstract** This paper designs a high-order positivity-preserving discontinuous Galerkin (DG) scheme for a linear hyperbolic equation. The scheme relies on augmenting the standard polynomial DG spaces with additional basis functions. The purpose of these augmented basis functions is to ensure the preservation of a positive cell average for the unmodulated DG solution. As such, the simple Zhang and Shu limiter [30] can be applied with no loss of accuracy for smooth solutions, and the cell average remains unaltered. A key feature of the proposed scheme is its implicit generation of suitable augmented basis functions. Nonlinear optimization facilitates the design of these augmented basis functions. Several benchmarks and computational studies demonstrate that the method works well in two and three dimensions.

**Keywords** discontinuous Galerkin · High-order · Positivity-preserving

**Mathematics Subject Classification (2020)** MSC code1 · MSC code2 · more

## 1 Introduction and literature review

The main objective of this paper is to design a high-order positivity-preserving discontinuous Galerkin (DG) scheme for a linear hyperbolic equation (see equation (1)). The main idea is to enrich the standard finite element spaces with additional functions. The main purpose for these functions is to ensure that the resulting cell average from the augmented DG scheme preserves a positive cell average.

In [12], the authors developed a DG method for the same linear hyperbolic problem (with positive inflow conditions and source functions), which relied

Maurice S. Fabien
Department of Mathematics, University of Wisconsin, Madison, WI 53706, USA
E-mail: mfabien@wisc.edu

on augmenting the standard polynomial spaces $\mathcal{P}_1$ (the space of polynomials of total degree less than or equal to one) and $\mathcal{Q}_1$ (the space of polynomials of degree less than one in each space direction) with additional basis functions. By doing so, they demonstrated that the resulting scheme preserves a positive cell average for the unmodulated DG solution. Here, unmodulated DG solution refers to a DG solution is not modified in any way through post-processing (such as slope limiting, flux correct transport, or simple 'fix-up'/'clipping', see [27]). The present work is also an extension of [6], where similar ideas are explored. In particular, explicit non-polynomial augmented basis functions are defined based on the method of characteristics. Then, these augmented basis functions are used to enrich the standard polynomial spaces $\mathcal{P}_k$ and $\mathcal{Q}_k$. These basis functions are specifically constructed to ensure that the resulting scheme preserves a positive cell average for the DG solution without limiters. As such, the straightforward scaling limiter from [30] can be applied without any loss of accuracy for smooth solutions, and the cell average remains unaltered.

Bound-preserving schemes have received considerable attention in recent years. Various approaches have been developed to address this challenge. An explicit positivity-preserving scheme that utilizes Lax-Wendroff time stepping was developed in [21].

Operator splitting methods, particularly Strang splitting, have proven effective in the context of bound-preserving schemes. In [21], the authors combine Strang operator splitting with a discontinuous Galerkin spatial discretization to achieve high-order accuracy in both space and time for the Vlasov-Possion equations. Similarly, Zhang et al.[29] applied operator splitting with a discontinuous Galerkin discretization to a chemotaxis PDE model - by exploiting the splitting, they are able to retain bound-preservation. These operator splitting approaches are particularly valuable as they can achieve temporal accuracy beyond second order.

For parabolic and convection-diffusion systems, several methods have been proposed in [22,7,24]. Additionally, techniques that combine *hp*-aptive refinement, and composing different limiters has demonstrated to be effective for nonlinear problems on unstructured meshes [8].

Limiters have also been designed and applied to a variety of other linear and nonlinear problems. For instance, Vlasov-Boltzmann transport equations [4], extended magnetohydrodynamics equations [31], nonlinear shallow water equations [11], Gaussian closure equations [16], and radiative transfer equations [28], and compressible Navier–Stokes equations [28].

There also exist a number of works which utilize various optimization techniques to enforce bound constraints. For semi-implicit and operator splitting schemes, see [14,15,13]. An efficient active set semismooth Newton method was developed for the fully time-implicit DG discretization of nonlinear problems in [23]. In general, provable bound-preserving schemes for fully time-implicit DG methods still remains a challenge - even for linear PDEs.

We note that this list of references on limiters is far from complete. It is quite challenging to prove that the resulting schemes retain high-order accuracy, positivity, as well as preserve other desirable quantities (such as: momentum, energy, and so on). Time-dependent problems introduce additional challenges if implicit time-stepping is considered. In practice, implicit time-stepping gives rise to nonlinear systems that need to be resolved via iterations such as Newtons method. Imposing additional constraints on the nonlinear system can cause issues with convergence [17].

In this paper, we extend the results in [6,12] by considering the case of polynomial degree $k > 1$, and also develop a new strategy related to the work in [6]. The work done in [12] produces an explicitly formed defined augmented basis function, which when augmented to the standard DG approximation spaces, results in a positive cell average for the unmodulated DG solution. However, the results are only valid for $k = 1$. In [6], an explicitly defined augmented basis function is constructed. Here, specific non-polynomial augmented basis functions are designed using the method of characteristics. However, this work focuses on producing appropriate augmented basis functions on the fly via an optimization procedure. The advantage of this is that the augmented basis functions are guaranteed to be polynomials. In particular, we additionally augment the standard DG spaces $\mathcal{P}_k$ (the space of polynomials of total degree less than or equal to $k$) and $\mathcal{Q}_k$ (the space of polynomials of degree less than $k$ in each space direction) with additional basis functions so that the resulting scheme is high-order accurate and preserves a positive cell average for the unmodulated DG solution. The so-called serendipity elements (see [2]) are also examined. Our approach utilizes a numerical optimization procedure to generate suitable augmented basis functions automatically. This is done because it is difficult to a priori explicitly exhibit suitable augmented basis functions for $k > 1$. Some of the notable features of this approach are:

- The unmodulated DG solution for any $k > 0$ will have a positive cell average for any Courant–Friedrichs–Lewy (CFL) number [5,10].
- The resulting method can be used for non-negative variable coefficients.
- The algorithm is independent of the domain dimension, therefore, 2D and 3D meshes can be used.

We mention some other related works. It is emphasized that the following three works do not consider function space augmentation. Optimization techniques have been used to enforce that the DG approximation is non-negative everywhere, for instance [23,27]. Here, a constrained optimization problem is solved where the entire DG solution is constrained to be bound-preserving. The difference between the current work is that we merely request that the cell average remains positive, and not the entire DG approximation. This reduced the computational effort, and relaxes the complexity of the underlying optimization problem. Similarly, the works in [14,15,13] also leverage optimization procedures to obtain bound-preserving schemes, but they do not consider function space augmentation, which is the primary focus of the current work.

In [25] the authors consider under-integration (inexact quadrature), using quadrature rules with reduced precision. By doing this, they show good results and some theoretical evidence that under-integration can produce positive cell averages for the unmodulated unaugmented DG solver. Of course, the reduced precision of the quadrature effects the overall accuracy for smooth solutions.

The work in [26] explores new notions of conservation which are arguably more suitable for steady-state or implicit problems. It is found that the standard notion of conservation, preserving the cell average, may not be ideal. For instance, this may result in mass loss over long time scales. By redefining conservation, they are able to design high-order positivity-preserving DG schemes without the need for augmentation. The limitation is of course that the classical notion of conservation must be abandoned; and it is not immediately clear how to extend the results to nonlinear problems.

The current paper is organized as follows. Sections 2 and 3 describe the model problems and briefly introduce the DG discretization. The strategy for obtaining suitable augmented basis functions is introduced in Section 4. To ensure the overall DG approximation remains non-negative, the simple scaling limiter from [30] is rehashed in Section 5. Section 6 presents several examples motivating function space augmentation, as well as explicit basis functions. Numerical experiments are conducted in Section 7, which verify and validate the numerical optimization strategy. The paper is finalized with important findings and conclusions.

## 2 Model problem and discretization in (1+1) dimensions

The two dimensional linear hyperbolic equation considered in this paper has the form

$$\frac{\partial(\alpha u(x,y))}{\partial x} + \frac{\partial(\beta u(x,y))}{\partial y} + \gamma u(x,y) = f(x,y), \tag{1}$$

where $\alpha, \beta > 0$, $\gamma \geq 0$, $f(x,y) \geq 0$, and the domain is $\mathcal{D} = [a,b] \times [c,d] \subset \mathbb{R}^2$. If equation (1) is prescribed non-negative initial/inflow boundary conditions, then one can show that the solution $u$ will be non-negative (e.g., method of characteristics). The computational domain is partitioned into a rectangular mesh with cells $S_{ij} = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$, with $i = 1, \ldots, N_x$ and $j = 1, \ldots, N_y$. Define the traditional DG space $V_h^k$ as

$$V_h^k = \{v \in L^2(\mathcal{D}) : v|_{S_{ij}} \in \mathcal{X}_k(S_{ij}),\ \forall i = 1, \ldots, N_x,\ \forall j = 1, \ldots, N_y\},$$

where $\mathcal{X}_k$ is a standard polynomial space such as $\mathcal{P}_k$ (space of all polynomials up to degree $k$), $\mathcal{Q}_k$ (space of tensor-product polynomials of degree $k$), or the serendipity space $\mathcal{S}_k$ (for details, see [2]) with polynomial degree $k \geq 1$. The DG scheme for solving (1) then seeks $u \in V_h^k$ so that for all $v \in V_h^k$ we have

$$\mathcal{L}(u,v) = \mathcal{R}(v), \tag{2}$$

where the bilinear form $\mathcal{L}(\cdot,\cdot)$ and functional $\mathcal{R}(\cdot)$ are given by

$$\begin{aligned}\mathcal{L}(u,v) = &-\int_{S_{ij}} u(x,y)(\alpha v_x(x,y) + \beta v_y(x,y))\,dx\,dy \\ &+ \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \alpha u(x^-_{i+\frac{1}{2}}, y) v(x^-_{i+\frac{1}{2}}, y)\,dy \\ &+ \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \beta u(x, y^-_{j+\frac{1}{2}}) v(x, y^-_{j+\frac{1}{2}})\,dx + \int_{S_{ij}} \gamma u(x,y) v(x,y)\,dx\,dy,\end{aligned} \tag{3a}$$

$$\begin{aligned}\mathcal{R}(v) = &\int_{S_{ij}} f(x,y) v(x,y)\,dx\,dy + \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \alpha u(x^-_{i-\frac{1}{2}}, y) v(x^+_{i-\frac{1}{2}}, y)\,dy \\ &+ \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \beta u(x, y^-_{j-\frac{1}{2}}) v(x, y^+_{j-\frac{1}{2}})\,dx.\end{aligned} \tag{3b}$$

Equation (2) can be solved locally on each cell, with inflow boundary conditions. In other words, inflow conditions on a given cell are outflow conditions from an adjacent cell or inflow from the domain boundary.

The cell average of a function $u$ on cell $S_{ij}$ is denoted by

$$\overline{u}_{ij} = \frac{1}{|S_{ij}|} \int_{S_{ij}} u\,dx\,dy,$$

where $|S_{ij}|$ is the measure (area) of cell $S_{ij}$.

## 3 Model problem and discretization in (1+2) dimensions

Consider the (1+2) dimension problem:

$$\frac{\partial u(x,y,t)}{\partial t} + \alpha \frac{\partial u(x,y,t)}{\partial x} + \beta \frac{\partial u(x,y,t)}{\partial y} + \gamma u(x,y,t) = f(x,y,t), \tag{4}$$

with the space-time domain $(x,y,t) \in \mathcal{D} = \Omega \times [0,T]$ (for some final time $T > 0$), $\Omega \subset \mathbb{R}^2$. Similar to the model problem in Section 2, $\alpha, \beta > 0$ and $\gamma \geq 0$ are constants. The PDE (4) is equipped with non-negative inflow boundary conditions, and it is assumed that $f \geq 0$.

We employ a space-time DG method, so that the space-time slab is now a rectangular cuboid (instead of a rectangle). In other words, the space-time domain $\Omega \times [0,T]$ is partitioned into a total of $N_x * N_y * N_t$ rectangular cuboids $S_{ij\ell}$ ($N_x$ cuboids in the $x$-direction, $N_y$ cuboids in the $y$-direction, $N_t$ cuboids in the $t$-direction). The cuboid mesh then has cells $S_{ij\ell} = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}] \times [t_{\ell-\frac{1}{2}}, t_{\ell+\frac{1}{2}}]$, with $i = 1, \ldots, N_x$, $j = 1, \ldots, N_y$ and $\ell = 1, \ldots, N_t$.

The DG scheme for solving (4) then seeks $u \in V_h^k$ so that for all $v \in V_h^k$ we have

$$\begin{aligned}
&- \int_{S_{ij\ell}} u(\alpha v_x + \beta v_y + v_t - \gamma v)\, dx\, dy\, dt \\
&+ \alpha \int_{t_{j-\frac{1}{2}}}^{t_{j+\frac{1}{2}}} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} u(x^-_{i+\frac{1}{2}}, y, t) v(x^-_{i+\frac{1}{2}}, y, t)\, dy\, dt \\
&+ \beta \int_{t_{j-\frac{1}{2}}}^{t_{j+\frac{1}{2}}} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, y^-_{i+\frac{1}{2}}, t) v(x, y^-_{i+\frac{1}{2}}, t)\, dx\, dt \\
&+ \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, y, t^-_{i+\frac{1}{2}}) v(x, y, t^-_{i+\frac{1}{2}})\, dx\, dy \\
&= \int_{S_{ij\ell}} f v\, dx\, dy\, dt + \alpha \int_{t_{j-\frac{1}{2}}}^{t_{j+\frac{1}{2}}} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} u(x^-_{i-\frac{1}{2}}, y, t) v(x^+_{i-\frac{1}{2}}, y, t)\, dy\, dt \\
&+ \beta \int_{t_{\ell-\frac{1}{2}}}^{t_{\ell+\frac{1}{2}}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u(x, y^-_{j-\frac{1}{2}}, t) v(x, y^+_{j-\frac{1}{2}}, t)\, dx\, dt \\
&+ \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u(x, y, t^-_{\ell-\frac{1}{2}}) v(x, y, t^+_{\ell-\frac{1}{2}})\, dx\, dy.
\end{aligned} \tag{5}$$

The cell average of a function $u$ on cell $S_{ij\ell}$ is denoted by

$$\overline{u}_{ij\ell} = \frac{1}{|S_{ij\ell}|} \int_{S_{ij\ell}} u\, dx\, dy\, dt,$$

where $|S_{ij\ell}|$ is the measure (volume) of cell $S_{ij\ell}$.

## 4 Procedure for high-order positivity-preserving scheme

It is possible to augment standard DG spaces with additional basis functions so that high-order accuracy is retained and the cell average is guaranteed to be positive. In [12], the authors augment the standard $\mathcal{P}_1$ space by adding additional basis functions which guarantee that the cell average of the unmodulated DG solution remains positive. As long as the inflow conditions and source are also non-negative, the cell average of the unmodulated DG solution is guaranteed to be non-negative (which can be deduced from (3)).

The focus of this paper is to extend the results in [12]. More specifically, assuming positive inflow and sources, we augment the standard DG spaces ($k > 1$) so that the resulting DG scheme is high-order and preserves a positive cell average for the unmodulated DG solution. This is accomplished by utilizing an optimization algorithm which numerically searches for appropriate basis functions to augment the standard DG spaces.

Here we search for a *single* basis function to augment the standard DG space with. The augmented basis function (denoted by $\psi$) is treated as an ansatz (we select a polynomial of degree $r > k$), and search for coefficients of $\psi$ (via nonlinear programming) which ensure that the cell average of the unmodulated DG solution remains positive.

### 4.1 Augmenting polynomial basis

Here we provide an algorithm that generates a special basis function; such that, when it is augmented to the standard DG spaces, it will preserve a positive cell average for the unmodulated DG solution. Let $\mathcal{X}_k$ be a traditional polynomial space (e.g., $\mathcal{P}_k$, $\mathcal{Q}_k$, or the serendipity space $\mathcal{S}_k$). Following [12], for $k \geq 1$, we augment $\mathcal{X}_k$ with an additional basis function, denoted by $\psi$. The polynomial $\psi$ will belong to some higher-order space $\mathcal{X}_r(S_{ij})$ with $r > k$. That is, $\psi \in \mathcal{X}_r(S_{ij}) \backslash \mathcal{X}_k(S_{ij})$. The space $\mathcal{X}_k$ augmented with $\psi$ is denoted by $\widetilde{\mathcal{X}}_k = \mathcal{X}_k \cup \operatorname{span}(\psi)$. Let $\widetilde{V}_h^k$ be the augmented DG space:

$$\widetilde{V}_h^k = \{v \in L^2(\mathcal{D}) : v|_{S_{ij}} \in \widetilde{\mathcal{X}}_k(S_{ij}),\, \forall i = 1, \ldots, N_x,\, \forall j = 1, \ldots, N_y\}.$$

We search for some higher-order polynomial $\psi$ that will generate a test function $\widetilde{v} \in \widetilde{V}_h^k$ which simultaneously satisfies $\widetilde{v} \geq 0$ and

$$\mathcal{L}(u, \widetilde{v}) = \int_{S_{ij}} u \, dx \, dy, \tag{6}$$

for any $u \in \widetilde{V}_h^k$.

Let $\Phi_i$ be the $i$th basis function for $\mathcal{X}_r$, with $1 \leq i \leq \dim \mathcal{X}_r$. In addition, set $\psi = \sum_{i=1}^{\dim \mathcal{X}_r} \vec{d}_i \Phi_i$, and $\vec{d} = [d_1, d_2, \ldots, d_{\dim \mathcal{X}_r}]^T$. Then, we pose the following optimization problem: find a test function $\psi := \widetilde{v} \in \widetilde{\mathcal{X}}_k(S_{ij})$ (not necessarily unique) such that

$$\min_{\vec{d}} \quad \|\vec{d}\|_2^2 \tag{7a}$$

$$\text{subject to} \quad \widetilde{v}(x, y) \geq 0, \;\; (x, y) \in S_{ij}, \tag{7b}$$

$$\mathcal{L}(\widetilde{u}, \widetilde{v}) = \int_{S_{ij}} \widetilde{u} \, dx \, dy, \quad \forall \widetilde{u} \in \widetilde{\mathcal{X}}_k(S_{ij}), . \tag{7c}$$

In full generality, this optimization problem can be numerically resolved on each troubled cell locally by using nonlinear programming techniques such as the interior point method [3].

Of course, this is an ansatz, we *assume* that there is such a polynomial $\psi$ that satisfies Problem (7). In general, it is difficult to verify that Problem (7) has a solution (e.g., establishing the Karush–Kuhn–Tucker conditions [17]). Numerically we are able to obtain feasible solutions to Problem (7).

We next describe some steps of Problem (7). It should be noted that this procedure can be preprocessed, and applied to a reference cell (say, the square $[-1, 1]^2$ in 2D) instead of the physical cell $S_{ij}$. In practice, the DG scheme in (2) is typically resolved through quadrature; and the approximate solution to is only desired on some subset of $S_{ij}$, for instance quadrature or interpolation points. Let $\mathcal{G} \subset S_{ij}$ be this set of finitely many points (e.g., tensor product Gauss quadrature rules of sufficiently large precision). Each optimization step of Problem (7) starts with an initial guess for $\vec{d}$ (the coefficients for the candidate augmented basis function $\psi$). Next, the augmented space is formed $\widetilde{\mathcal{X}}_k(S_{ij}) = \mathcal{X}_k(S_{ij}) \cup \text{span}\{\psi\}$, and the linear system in (7c) (which has size $1 + \dim \mathcal{X}_k$) is assembled and solved for $\widetilde{v} \in \widetilde{\mathcal{X}}_k(S_{ij})$. If $\widetilde{v}$ is nonzero on the desired set of points $\mathcal{G}$, then the optimization step terminates. Otherwise, a new candidate for $\vec{d}$ is generated and the process repeats until $v \geq 0$ on the set $G$. For all numerical experiments and results in this paper, we take $\mathcal{G}$ to be the standard tensor product Gauss quadrature points [9], with enough points to ensure $\int_{S_{ij}} \widetilde{u}\widetilde{v}$ with $\widetilde{u}, \widetilde{v} \in \widetilde{\mathcal{X}}_k(S_{ij})$ to be integrated exactly.

Ideas that incorporate optimization in positivity-preserving DG schemes have been considered in [23] and [27]. These works use tools from optimization to design maximum-principle preserving limiters. Essentially, they solve a constrained optimization problem where the entire solution is constrained (or modified) to be nonnegative. In contrast, our approach seeks to maintain a positive cell average. The optimization procedure in Problem (7) searches for a suitable basis function to augment $\mathcal{X}_k$. With appropriate augmented basis functions for the standard DG space, we will have that the cell average of the unmodulated DG solution remains positive. Then, if the unmodulated DG solution is negative, one can apply the scaling limiter described in [30]. As such, the proposed approach has a provably high-order accurate limiter, and, the cell average remains unaltered.

Assuming a troubled cell ($\overline{u}_{ij} < 0$), the overall procedure is summarized below:

1. Pick an ansatz polynomial $\psi \in \mathcal{X}_r$ with degree $r > k$, $\psi = \sum_{i=1}^{|\mathcal{X}_r|} \vec{d}_i \Phi_i$.
2. Numerically solve Problem (7) to determine viable coefficients $\vec{d}$ for $\psi$.
3. Augment the standard $\mathcal{X}_k$ space with $\psi$ (denote the augmented space by $\widetilde{\mathcal{X}}_k$); replace $V_h^k$ with $\widetilde{V}_h^k$ in (1). Then, solve (1) for $u_{ij} \in \widetilde{V}_h^k$, where
$$\widetilde{V}_h^k = \{v \in L^2(\mathcal{D}) : v|_{S_{ij}} \in \widetilde{\mathcal{X}}_k(S_{ij}),\, \forall i = 1, \dots, N_x,\, \forall j = 1, \dots, N_y\}.$$
4. If $v < 0$, limit the approximation. Apply the scaling limiter [30], $\tilde{u}_{ij}(x, y) = \theta(u_{ij}(x, y) - \overline{u}_{ij}) + \overline{u}_{ij}$.

Through various numerical experiments, it is found that $r$ is not substantially larger than $k$. If $\psi$ is be to valid for only one cell $S_{ij}$, then $r = k + 1$ appears to suffice. In the case that it is desirable for a given $\psi$ to work for multiple cells (all cells obeying some Courant–Friedrichs–Lewy (CFL) upper/lower bound), then larger values of $r$ seem to be necessary. This paper utilizes Problem (7) as stated, but there are many ways to modify the search for suitable augmented functions $\psi$. For instance, one could use non-negative basis functions and enforce $\vec{d}_i \geq 0$. Other possibilities include (but are not limited to) modifying the objective function (e.g., log barrier/penalty methods for remote initial guesses), augmenting multiple basis functions to a standard DG space, or requesting $\psi$ posses certain properties (e.g., $\psi$ vanishes at the outflow boundaries).

This approach may seem prohibitively expensive, however, Problem 7 only needs to be invoked on troubled cells (where the unmodulated cell average is negative). In other words, we use an adaptive scheme, where the standard $\mathcal{X}_k$ space is used when the resulting cell average is non-negative, and Problem 7 is only activated when the computed cell average from $\mathcal{X}_k$ is negative. Typically this only occurs on a subset of the mesh. It should be also noted that (3) is usually approximated using quadrature rules. As such, we may weaken the constraints of Problem (7), so that $v$ is only non-negative at certain quadrature or control points on a cell (instead of non-negative over the entire cell). Moreover, due to the local nature of the DG scheme, Problem 7 can be leveraged on a single cell. We also demonstrate that it is possible to pre-process suitable $\psi$ that is valid for all cells obeying a CFL-type bound. Furthermore, Problem 7 is extendable to higher dimensional Cartesian meshes.

### 4.2 Positivity-preserving DG scheme in (1+2) dimensions

To arrive at a positivity-preserving DG scheme for (5), we use the approach from Section 4. Let $\mathcal{L}(u, v)$ denote the bilinear form associated with the left hand side of (5). The classical DG space is given by

$$V_h^k = \{v \in L^2(\mathcal{D}) : v|_{S_{ij\ell}} \in \mathcal{X}_k(S_{ij\ell}),\ 1 \leq i \leq N_x,\ 1 \leq j \leq N_y,\ 1 \leq \ell \leq N_t\}.$$

We seek some higher-order polynomial $\psi$ to augment to the space $V_h^k$ with. The resulting augmented space $\widetilde{V}_h^k$ will contain a test function $v \in \widetilde{V}_h^k$ that simultaneously satisfies $\widetilde{v} \geq 0$ and

$$\mathcal{L}(\widetilde{u}, \widetilde{v}) = \int_{S_{ij\ell}} \widetilde{u}\, dx\, dy\, dt, \tag{8}$$

for any $\widetilde{u} \in \widetilde{V}_h^k$. One can obtain an appropriate $\psi \in \mathcal{X}_r$ from Problem (7) applied to (1+2) dimensions: For $r > k$, find a test function $\psi := \widetilde{v} \in \widetilde{\mathcal{X}}_k(S_{ij\ell})$ (not necessarily unique) such that

$$\min_{\vec{d}} \quad \|\vec{d}\|_2^2 \tag{9a}$$

$$\text{subject to} \quad \widetilde{v}(x, y, t) \geq 0,\ (x, y, t) \in S_{ij\ell}, \tag{9b}$$

$$\mathcal{L}(\widetilde{u}, \widetilde{v}) = \int_{S_{ij\ell}} \widetilde{u}\, dx\, dy, \quad \forall \widetilde{u} \in \widetilde{\mathcal{X}}_k(S_{ij\ell}). \tag{9c}$$

Problem (9) only needs to be applied on troubled cells (where the cell average from the unmodulated DG scheme using $\mathcal{X}_k(S_{ij\ell})$ is negative). In practice, Problem (9) can be weakened to hold at quadrature points on $S_{ij\ell}$ (and not necessarily over the entire cell).

*Remark 1* Problems (7) and (9) searches for coefficients $\vec{d}$ such that the solution $\widetilde{v} \geq 0$ and satisfies $\mathcal{L}(\widetilde{u}, \widetilde{v}) = \overline{u}$ for all $\widetilde{u} \in \widetilde{\mathcal{X}}_k$. In general, determining these coefficients is not trivial - especially if nonnegativity is requested to hold for *all* admissible parameters $\alpha, \beta, \omega, \gamma, \Delta x, \Delta y, \Delta z$. That is, $\vec{d}$ are independent of the parameter space $\Omega_\delta = (\alpha, \beta, \omega, \gamma, \Delta x, \Delta y, \Delta z)$.

For a fixed set of parameters $\Omega_\delta$, Problems (7) and (9) is able to generate feasible candidates for $\vec{d}$. However, determining $\vec{d}$ that are feasible for any parameters in $\Omega_\delta$ is more challenging. Effectively, this amounts to multi-parametric (MP) programming problems [18] of the type

$$J^*(\vec{\theta}) = \min_{\vec{x} \in \mathbb{R}^n} f(\vec{x}, \vec{\theta})$$
$$\text{subject to} \quad g(\vec{x}, \vec{\theta}), \quad \vec{\theta} \in \Omega_\delta,$$

where the optimal value $J$ in general *depends* on the parameters $\vec{\theta}$. Obtaining $\vec{d}$ that are constant for all $\vec{\theta} \in \Omega_\delta$ can also be interpreted as a semi-infinite program (SIP) with finitely many variables and an infinite number of constraints [20]. Nonlinear variants of these optimization problems remain a highly active area of research. It is worth pointing out that these SIP and MP optimization techniques assume that $\Omega_\delta$ is a compact set, which automatically rules out non-compact sets such as $\{(\alpha, \beta\, \gamma) : \alpha > 0, \beta > 0, \gamma \geq 0, \}$.

From another perspective, the primary focus is to satisfy the constraints (nonnegativity), as such, the parameter search could also be interpreted as a feasibility problem (no objective function, only constraints).

## 5 Positivity-preserving limiter

In this section we summarize the positivity-preserving scheme. The augmented spaces $\widetilde{\mathcal{X}}_k$ only ensure that the cell average from the unmodulated DG scheme is non-negative. To arrive at a solution that is non-negative everywhere, a limiter may be needed. We first state the following important conclusion, then discuss the limiter.

**Theorem 1** *Consider the model PDE* (1) *or* (4)*. Then, the cell average* $\overline{u}|_S$ *of the DG schemes in* (3) *and* (5) *with the augmented spaces* $\widetilde{\mathcal{X}}_k(S)$ *will remain non-negative provided that the source term and inflow conditions from the upstream cells (including the inflow from the boundary conditions) are non-negative.*

*Proof* Let $S \in \mathcal{T}_h$ and set the polynomail degree $k \geq 1$. Assume that the inflow conditions and source function $f$ are non-negative. Further, suppose that Problem (7) has generated an augmented basis function $\psi \in \mathcal{X}_r(S)$, for $r > k$ (see Appendix A for special cases). As such, there exists $\widetilde{v} \in \widetilde{X}_k(S) = \mathcal{X}_k(S) \cup \text{span}\{\psi\}$ that simultaneously satisfies $0 \leq \widetilde{v}$ and

$$\mathcal{L}(\widetilde{u}, \widetilde{v}) = \int_S \widetilde{u}\, dx\, dy, \tag{10}$$

for any $\widetilde{u} \in \widetilde{X}_k(S)$ (where $\mathcal{L}(\cdot,\cdot)$ is defined in (3)). However, from (2), we also have

$$\mathcal{R}(\widetilde{v}) = \mathcal{L}(\widetilde{u}, \widetilde{v}), \tag{11}$$

where $\mathcal{R}(\widetilde{v}) \geq 0$, since we assumed $f$ and the inflow conditions are both non-negative. Using (11) in (10) gives

$$0 \leq \mathcal{R}(\widetilde{v}) = \mathcal{L}(\widetilde{u}, \widetilde{v}) = \int_S \widetilde{u}\, dx\, dy.$$

It follows that $\overline{u}_S = \frac{1}{|S|} \int_S \widetilde{u}\, dx\, dy \geq 0$. ■

The limiter we use is applied at quadrature points. In practice, equation (6) is approximated by quadrature. This suggests we may weaken the condition which enforces that the test function $v$ from (7) is non-negative on an entire cell $S_{ij}$. Instead, one may request that $v$ from (7) is non-negative only at specific quadrature points (and not necessarily everywhere on $S_{ij}$). If the corresponding quadrature weights are positive, it then follows that the resulting quadrature will be non-negative (similar conclusions hold for the DG scheme in $(1+2)$ dimensions).

Let the set $G_{2D}$ contain the abscissa on cell $S_{ij}$ from a tensor product quadrature rule with non-negative weights (such as the Gauss-Lobatto quadrature [1,19]). Similarly, in 3D, the set $G_{3D}$ contains the abscissa on cell $S_{ij\ell}$ from a tensor product quadrature rule with non-negative weights. In 2D, the scaling limiter from [30] modifies the approximation $u_{ij}(x,y)$ such that

$$\widetilde{u}_{ij}(x,y) = \theta(u_{ij}(x,y) - \overline{u}_{ij}) + \overline{u}_{ij}, \quad \theta = \min_{(x,y)\in G_{2D}} \left\{1, \left|\frac{\overline{u}_{ij}}{u_{ij}(x,y) - \overline{u}_{ij}}\right|\right\}.$$

The 3D analog of the scaling limiter is

$$\widetilde{u}_{ij\ell}(\vec{x}) = \theta(u_{ij\ell}(\vec{x}) - \overline{u}_{ij\ell}) + \overline{u}_{ij\ell}, \quad \theta = \min_{\vec{x}\in G_{3D}} \left\{1, \left|\frac{\overline{u}_{ij\ell}}{u_{ij\ell}(\vec{x}) - \overline{u}_{ij\ell}}\right|\right\},$$

where $\vec{x} = [x, y, z]^T$. The limited polynomial solution $\widetilde{u}$ is non-negative and also retains high-order accuracy [30].

## 6 Motivating examples for problem (7)

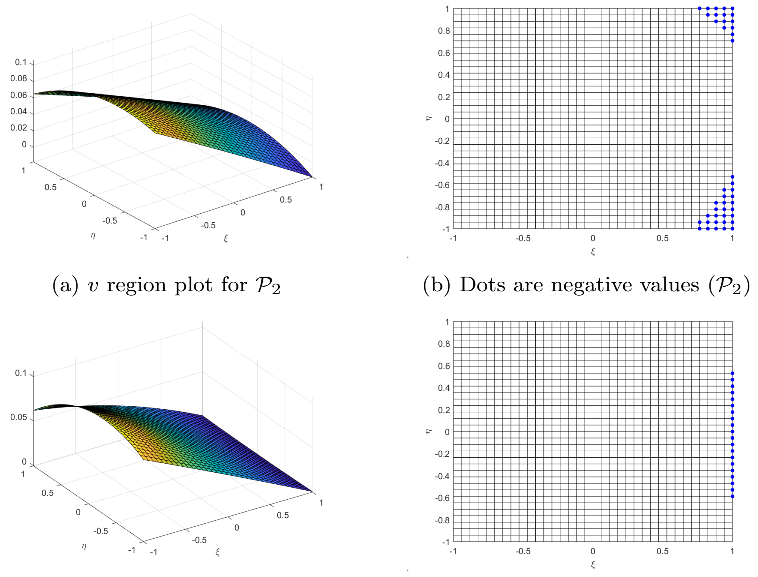
The purpose of this section is to showcase some features of Problem (7). In particular, the effectiveness of Problem (7) is demonstrated in the case of variable coefficients and polynomial degrees $k > 1$. Several motivating examples are also provided which illustrate the need for augmentation. This is done by examining the traditional unaugmented high-order DG schemes (e.g., with $\mathcal{P}_k$, $\mathcal{S}_k$, or $\mathcal{Q}_k$); to see whether or not they give rise to a test function $v$ that satisfies both $v \geq 0$ and (6). The condition $v \geq 0$ is enforced because we need the right hand side of (3) to be non-negative to guarantee a non-negative cell average. In more detail, we enforce that (6) holds, however, (3) gives

$$
\begin{aligned}
\int_{S_{ij}} u \, dx \, dy &= \mathcal{L}(u, v) \\
&= \int_{S_{ij}} f v \, dx \, dy + \alpha \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} u(x^-_{i-\frac{1}{2}}, y) v(x^+_{i-\frac{1}{2}}, y) \, dy \\
&+ \beta \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, y^-_{i-\frac{1}{2}}) v(x, y^+_{i-\frac{1}{2}}) \, dx.
\end{aligned}
$$

Since we assume that $f \geq 0$ and the inflow from $u$ is also non-negative, it follows that $v \geq 0$ ensures the cell average of $u$ will be non-negative.

### 6.1 The need for augmented spaces in (1+1) dimensions

We consider a single cell $S_{ij} = [-1, 1]^2$, and $\alpha = 1$, $\beta = 1/10$, $\gamma = 0$, $\Delta x_i = \Delta y_j = 0.1$. By solving (6) (replacing $\widetilde{V}_k$ with $V_k$) with $r = k = 2$, one finds that the resulting unique test function $v$ has a negative minimum (see Fig. 1).



(a) $v$ region plot for $\mathcal{P}_2$

(b) Dots are negative values ($\mathcal{P}_2$)

(c) $v$ region plot for $\mathcal{Q}_2$

(d) Dots are negative values ($\mathcal{Q}_2$)

Fig. 1: Plots of $v$ for $\mathcal{P}_2$ (Figure 1a) and $\mathcal{Q}_2$ (Figure 1c). Figure 1b and 1d display locations where $v$ is negative (indicated with dots).

One might hope that simply increasing the polynomial degree would satisfy the constraints, however, this does not appear to remedy the situation. The minimum of $v$ arising from the standard spaces $\mathcal{P}_k$ and $\mathcal{Q}_k$ for various $k > 1$ is displayed in Fig 2. We again restrict our study to a single cell, with the same parameters $\alpha = 1$, $\beta = 1/10$, $\gamma = 0$, $\Delta x_i = \Delta y_j = 0.1$. From Fig 2 it is clear that increasing the polynomial degree will not necessarily generate a non-negative test function $v$, therefore, this will not ensure a positive cell average.

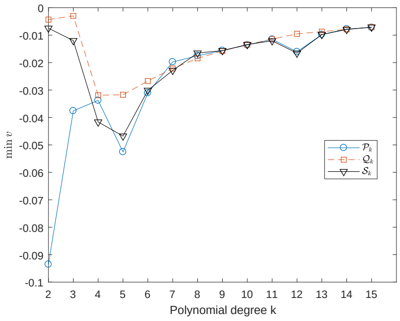


Fig. 2: Minimums for the basis function $v$ resulting from the unaugmented spaces $\mathcal{P}_k$, $\mathcal{Q}_k$, and $\mathcal{S}_k$ (Serendipity space) are plotted against their respective polynomial degree $k$. The minimums are strictly negative

### 6.2 Variable coefficients

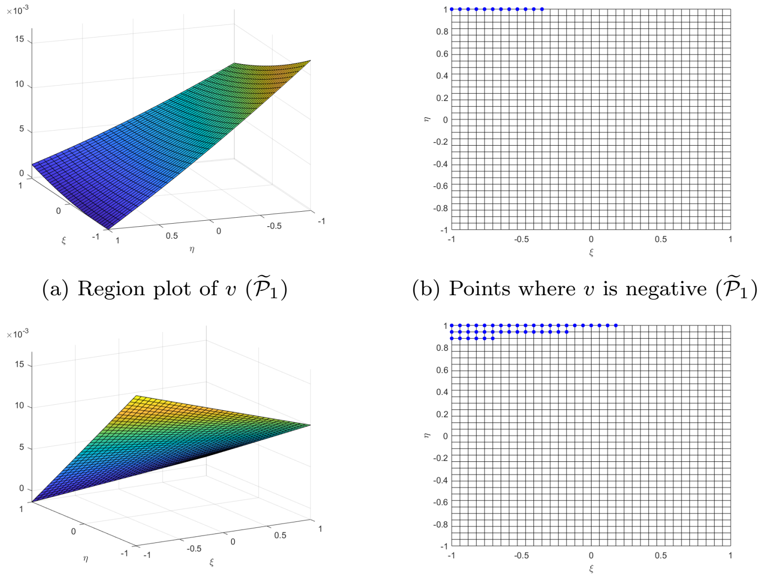
Here we highlight a convenient feature of Problem (7) - it can be used with non-negative variable coefficients. In the case of variable coefficients, the augmented functions for $k = 1$ suggested in [12] do not provide a non-negative test function $v$. In particular, they augment $\mathcal{P}_1$ with the function

$$\psi(x, y) = xy + (1/4)(x^2 + y^2) \tag{12}$$

However, this choice is only suitable for constant coefficients.

We provide a counterexample. Let $\alpha(x, y) = e^{-x-y}$, $\beta(x, y) = (1 + e^2) - e^{-x-y}$, and $\gamma(x, y) \equiv 0$. Set the domain to be the reference cell $[-1, 1]^2$, $\Delta x = \Delta y = 0.1$, and $x = \xi$, $y = \eta$. Note that $\alpha, \beta > 0$ on the reference cell, and the vector field $[\alpha(x, y), \beta(x, y)]^T$ is divergence-free.

Fig. 3 plots the test function $v$ resulting from using the augmented spaces $\widetilde{\mathcal{P}}_1$ or $\widetilde{\mathcal{Q}}_1$. We see that they do not remain non-negative. Thus, the augmented basis function (12) does not ensure a positive cell average. On the other hand, from Table 1, it is evident that Problem (7) is able to ensure that $v$ is non-negative by automatically finding a suitable augmented basis function.



(a) Region plot of $v$ ($\widetilde{\mathcal{P}}_1$) (b) Points where $v$ is negative ($\widetilde{\mathcal{P}}_1$)

(c) Region plot of $v$ ($\widetilde{\mathcal{Q}}_1$) (d) Points where $v$ is negative ($\widetilde{\mathcal{Q}}_1$)

Fig. 3: Basis function $v$ resulting from the augmented spaces ($\psi$ from (12), see [12]) $\widetilde{\mathcal{P}}_1$ or $\widetilde{\mathcal{Q}}_1$ are not non-negative. Filled circles indicate locations where $v$ is negative

| Augmented space | $\min_{(\xi,\eta)\in[-1,1]^2} v(\xi,\eta)$ |
|---|---|
| $\widetilde{\mathcal{P}}_1$ ($\psi$ from (12), see [12]) | $-1.268\text{E}-04$ |
| $\widetilde{\mathcal{Q}}_1$ ($\psi$ from (12), see [12]) | $-1.363\text{E}-03$ |
| $\widetilde{\mathcal{P}}_1$ (Problem (7)) | $1.843\text{E}-04$ |
| $\widetilde{\mathcal{Q}}_1$ (Problem (7)) | $1.155\text{E}-04$ |

Table 1: Minimums of the test function $v$ (acquired from different augmented spaces). Problem (7) automatically ensures $v$ is non-negative. The augmented spaces from [12] result in a negative test function $v$

## 6.3 Example augmented spaces

Although the primary feature of Problem (7) is to implicitly generate suitable test and trial spaces on the fly, it is possible to use it to pre-process them (as apposed to activating Problem (7) on troubled cells). In particular, in this section we explore explicit formulas for augmented basis functions which are valid for multiple cells (e.g, obeying CFL-type conditions).

We remark that the explicit augmented functions presented here are not unique or optimal, but are selected for their succinct representation. Problem (7) generates $\dim \mathcal{X}_r$ real coefficients (housed in $\vec{d}$) that allow us to form the augmented function candidate $\psi = \sum_{i=1}^{\dim \mathcal{X}_r} \vec{d}_i \Phi_i$. As $r > k$, not only will there be a proliferation of these coefficients, but their decimal expansions are unwieldy in general.

Let $\alpha, \beta, \gamma$ be constants. Proving that the augmented basis function $\psi$ gives rise to a viable augmented space is symbolically intensive, because $v$ is obtained a-posteriori via solving a linear system on the reference cell. State-of-the-art symbolic software such as Mathematica, Maxima, and Maple are unable to verify non-negativity analytically in all cases of admissible parameters. In this section we present numerical evidence that demonstrates $v(x, y) \geq 0$ and satisfies (6) (see Section 8, Appendix A, for analytic evidence).

Computations are focused on the reference cell, denoted by $\hat{S} = [-1, 1]^2$. To reduce the parameter space, we scale the equation in (6). Mapping (6) to the reference cell, and scaling the resulting equation by $(4/(\Delta x \Delta y))(\Delta x/(2\alpha))$ gives

$$\frac{\Delta x}{2\alpha} \iint_{\hat{S}} u \, d\xi d\eta = -\iint_{\hat{S}} u \left( v_\xi + \frac{\Delta x \beta}{\Delta y \alpha} v_\eta \right) d\xi d\eta + \frac{\Delta x \beta}{\Delta y \alpha} \int_{\eta=1} u(\xi, 1) v(\xi, 1) \, d\xi \tag{13a}$$

$$+ \int_{\xi=1} u(1, \eta) v(1, \eta) \, d\eta + \frac{\Delta x}{2\alpha} \iint_{\hat{S}} \gamma u v \, d\xi d\eta \tag{13b}$$

Define

$$B = \frac{\Delta x \beta}{\Delta y \alpha}, \quad G = \frac{\Delta x \gamma}{2\alpha}. \tag{14}$$

The scalar $B$ enables us to relate our computations to CFL-type conditions [5, 10]. Without a loss of generality, we assume that $\alpha = 1$, since it can be interpreted as scaling the model problem (1) by $1/\alpha$ and then applying the DG scheme to the scaled equation.

Therefore, with the above modifications and assumption on $\alpha$, the parameter space depends on the triplet $(\Delta x, B, G)$ (as opposed to 5 parameters $(\Delta x, \Delta y, \beta, \alpha, \gamma)$). In the following we assume $(\xi, \eta) \in \hat{S} = [-1, 1]^2$.

### Augmenting $\mathcal{Q}_k$ in (1+1) dimensions

*Case for $k = 2$, $B < 0.5$*

Using Problem (7), the following basis function is obtained

$$\psi_{1,\mathcal{Q}_2}(\xi, \eta) = \left( \xi \eta^2 (1 - \xi)^2 (\xi + 1)(1 - \eta^2)/8 \right)^2.$$

The associated augmented space $\widetilde{\mathcal{Q}}_2 = \mathcal{Q}_2 \cup \operatorname{span}(\psi_{1,\mathcal{Q}_2})$ is numerically found to ensure $v \geq 0$ for any $B \leq 0.5$, $G \geq 0$.

*Case for $k = 2,\ B > 2$*

The augmented basis function can be designed so that positivity is ensured for arbitrarily large CFL conditions. Using Problem (7), the following basis function is obtained

$$\psi_{2,\mathcal{Q}_2}(\xi,\eta) = \Big(\eta\xi^2(1-\eta)^2(\eta+1)(1-\xi^2)/8\Big)^2.$$

The associated augmented space $\widetilde{\mathcal{Q}}_2 = \mathcal{Q}_2 \cup \operatorname{span}(\psi_{2,\mathcal{Q}_2})$ is numerically found to ensure $v \geq 0$ for any $B \geq 2$, $G \geq 0$.

Table (2) exhibits explicit augmented basis functions for $\mathcal{Q}_k$, with various $k$. These augmented basis functions are found to be computationally valid for all cells satisfying a specific CFL type condition. The augmented functions in this table are not unique, but are selected for their somewhat straightforward expressions.

| Condition to enforce $\overline{u}_{ij} \geq 0$ | Augmented basis function $\psi$ | Original space |
|---|---|---|
| $(\beta/\alpha)\Delta x/\Delta y < \frac{1}{2}$ | $\psi_{1,\mathcal{Q}_2} = (xy^2(1-x)^2(x+1)(1-y^2)/8)^2$ | $\mathcal{Q}_2$ |
| $(\beta/\alpha)\Delta x/\Delta y > 2$ | $\psi_{2,\mathcal{Q}_2} = (yx^2(1-y)^2(y+1)(1-x^2)/8)^2$ | $\mathcal{Q}_2$ |
| $(\beta/\alpha)\Delta x/\Delta y < \frac{1}{4}$ | $\psi_{1,\mathcal{Q}_3}$ (see (22)) | $\mathcal{Q}_3$ |
| $(\beta/\alpha)\Delta x/\Delta y > 4$ | $\psi_{2,\mathcal{Q}_3}$ (see (23)) | $\mathcal{Q}_3$ |
| $(\beta/\alpha)\Delta x/\Delta y < \frac{1}{8}$ | $\psi_{1,\mathcal{Q}_4}$ (see (30)) | $\mathcal{Q}_4$ |
| $(\beta/\alpha)\Delta x/\Delta y > 8$ | $\psi_{2,\mathcal{Q}_4}$ (see (37)) | $\mathcal{Q}_4$ |

Table 2: Example augmented basis functions for $\mathcal{Q}_k$ generated by Problem 7. Appendix B has explicit formulas for select higher-order augmented functions ($\mathcal{Q}_k$)

The robustness of these augmented basis functions with respect to the CFL condition is examined. Consider a single cell $S_{ij}$. Random values are generated for $\Delta x_i, \Delta y_j, \alpha, \beta, \gamma$ to use in Problem (7). Define $B :=$ CFL $= (\beta/\alpha)(\Delta x_i/\Delta y_j)$. For several polynomial degrees, Fig. 4 plots the minimum of the test function $v$ obtained from Table (2) and Problem (7) against a wide range of CFL conditions. Our approach augments a single basis function to $\mathcal{Q}_2$. As such, the space $\widetilde{\mathcal{Q}}_2$ will have ten degrees of freedom, which is slightly more expensive than the unaugmented space. To arrive at a third order accurate scheme that has cost comparable to $\mathcal{Q}_2$ (and has a positive cell average), one could consider augmenting the eight-node serendipity element instead. This way one would have nine degrees of freedom (same as $\mathcal{Q}_2$), and the accuracy will be of third order [2].

Augmenting $\mathcal{S}_k$ in (1+1) dimensions

To reduce the computational cost, but retain $k+1$ order accuracy, one could consider augmenting the serendipity spaces $\mathcal{S}_k$. Table (3) displays several ex-

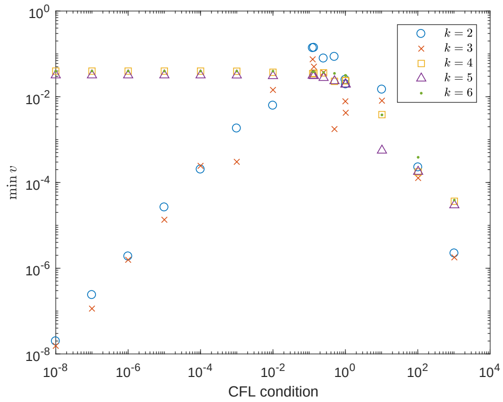


Fig. 4: Minimum of $v$ vs CFL number ($\widetilde{\mathcal{Q}}_k$). Log-log plot of the minimum of $v$ generated from Problem (7) vs CFL number. The test functions $v$ are non-negative for a wide range of CFL numbers

plicit augmented basis functions which are valid under certain CFL conditions. The spaces $\widetilde{\mathcal{S}}_k = \mathcal{S}_k \cup \operatorname{span}\{\psi_{\cdot,\mathcal{S}_k}\}$ enforce the existence of a non-negative test function under these CFL conditions. Fig. (5) demonstrates the effectiveness of

| Condition to enforce $\overline{u}_{ij} \geq 0$ | Augmented basis function $\psi$ | Original space |
|---|---|---|
| $(\beta/\alpha)\Delta x/\Delta y < \frac{1}{2}$ | $\psi_{1,\mathcal{S}_2} = (\xi\eta^2(\xi-1)^2(\eta-1)(\xi+1)(\eta+1)/8)^2$ | $\mathcal{S}_2$ |
| $(\beta/\alpha)\Delta x/\Delta y > 2$ | $\psi_{2,\mathcal{S}_2} = (\xi^2\eta(\xi-1)(\eta-1)^2(\xi+1)(\eta+1)/2)^2$ | $\mathcal{S}_2$ |
| $(\beta/\alpha)\Delta x/\Delta y < \frac{1}{8}$ | $\psi_{1,\mathcal{S}_3}$ (see (42)) | $\mathcal{S}_3$ |
| $(\beta/\alpha)\Delta x/\Delta y > 8$ | $\psi_{2,\mathcal{S}_3}$ (see (43)) | $\mathcal{S}_3$ |
| $(\beta/\alpha)\Delta x/\Delta y < \frac{1}{16}$ | $\psi_{1,\mathcal{S}_4}$ (see (48)) | $\mathcal{S}_4$ |
| $(\beta/\alpha)\Delta x/\Delta y > 16$ | $\psi_{2,\mathcal{S}_4}$ (see (49)) | $\mathcal{S}_4$ |

Table 3: Example augmented basis functions for the serendipity spaces $\mathcal{S}_k$ generated by Problem 7. Appendix C has explicit formulas for select higher-order augmented functions ($\mathcal{S}_k$)

the augmented basis functions. When augmented to the standard serendipity spaces, the minimum of $v$ remains non-negative for CFL conditions arbitrarily small or large. We emphasize that the augmented basis functions in Tables 2 and 3 are determined computationally, and theoretical evidence for validity for all parameters $\alpha, \beta, \gamma$ is lacking. However, Appendix A provides theoretical evidence for special cases, and numerical experiments give good results.

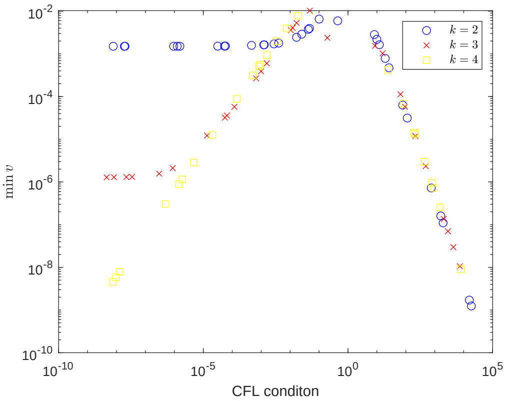


Fig. 5: Minimum of $v$ vs CFL number ($\widetilde{\mathcal{S}}_k$). Log-log plot of the minimum of $v$ generated from Problem (7) vs CFL number. The test functions $v$ are non-negative for a wide range of CFL numbers

Augmenting $\mathcal{P}_k$ in (1+1) dimensions

Problem 7 can successfully find augmented basis functions for $\mathcal{P}_k$ as well. Numerically we investigate the robustness of Problem 7 by generating random values for $\Delta x_i, \Delta y_j, \alpha, \beta, \gamma$. For each polynomial degree, we inspect the minimum of $v$ resulting from the associated augmented spaces. Fig. 6 plots

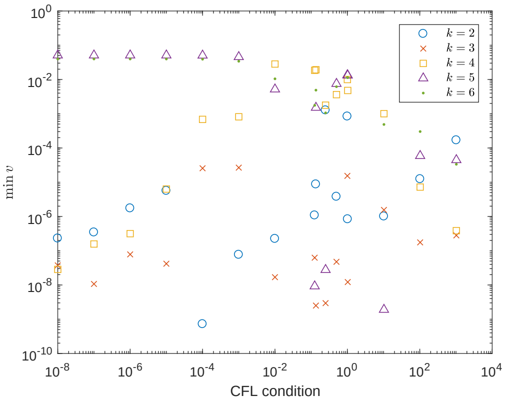


Fig. 6: Minimum of $v$ vs CFL number ($\widetilde{\mathcal{P}}_k$). Log-log plot of the minimum of $v$ generated from Problem (7) vs CFL number. The test functions $v$ are non-negative for a wide range of CFL numbers

the minimum of the test function $v$ against CFL numbers. It is evident that Problem 7 is creating non-negative test functions.

## 7 Numerical experiments

This section conducts several numerical experiments which verify and validate the proposed high-order accurate positivity-preserving DG scheme. Computational results are conducted in $(1+1)$ dimensions and $(1+2)$ dimensions. The numerical experiments provided here assume that it is understood from context that $x$ is replaced with $t$ and $y = x$ in (1) and (3). Similar test cases can be found in [12]. However, we remark that in [12] the backward Euler method is used to resolve the time derivative of (1), which reduces the overall accuracy to first order. Our method provided in this paper is $k+1$ order accurate in space and time.

### 7.1 Positivity violation for the $\mathcal{S}_2$ DG scheme in (1+1) dimensions

In this section we examine an example which demonstrates that the $\mathcal{S}_2$ DG scheme will not preserve a positive cell average in general; even when given a non-negative source and inflow boundary conditions. The computational domain is a single cell for this experiment. In particular, the following space-time domain is selected: $(x,t) \in [0, \Delta x] \times [0, \Delta t]$. We take $\alpha = 2, \beta = 1, \gamma = 0$, and

$$u(x,t) = \begin{cases} 20t + x, & \text{for } \frac{487}{496} \le \xi \le 1, \ -\frac{17}{18} \le \eta \le \frac{60\xi - 67}{24\xi - 15} \\ 0, & \text{otherwise} \end{cases}$$

$$f(x,t) = \begin{cases} 41, & \text{for } \frac{487}{496} \le \xi \le 1, \ -\frac{17}{18} \le \eta \le \frac{60\xi - 67}{24\xi - 15} \\ 0, & \text{otherwise} \end{cases}.$$

Note that the forcing function $f$ and $u$ are defined by the mapping from the reference square $(\xi, \eta) \in [-1,1]^2$. For this example $(\beta/\alpha)\Delta t/\Delta x = 1/2$, and as such, we select the augmented basis function given in the second row of Table 3. From Table 4, we see that the unaugmented $\mathcal{S}_2$ DG scheme violates positivity of the cell average, even when the mesh spacing is reduced. On the other hand, the augmented $\widetilde{\mathcal{S}}_2$ DG scheme preserves a positive cell average.

| $\Delta t = \Delta x$ | Exact | $\mathcal{S}_2$ | $\widetilde{\mathcal{S}}_2$ |
|---|---|---|---|
| 1/8 | 9.4E − 04 | −2.5E − 05 | 1.9E − 05 |
| 1/16 | 4.7E − 04 | −1.2E − 05 | 9.8E − 06 |
| 1/32 | 2.3E − 04 | −6.4E − 06 | 4.9E − 06 |
| 1/64 | 1.1E − 04 | −3.2E − 06 | 2.4E − 06 |

Table 4: Cell averages $\overline{u}_{1,1}$ generated by the $\mathcal{S}_2$ DG and $\widetilde{\mathcal{S}}_2$ DG schemes

### 7.2 Positivity violation for the $\mathcal{Q}_2$ DG scheme in (1+1) dimensions

The numerical experiment in this section is analogous to the one described in Section 7.1. We demonstrate that the $\mathcal{Q}_2$ DG scheme will not preserve a positive cell average in general; even when given a non-negative source and inflow boundary conditions. The computational domain is again taken to be a single cell for this experiment. In particular, the following space-time domain is selected: $(x,t) \in [0, \Delta x] \times [0, \Delta t]$. We take $\alpha = \beta = 1$, $\gamma = 0$, and

$$u(x,t) = \begin{cases} 200x + t, & \text{for } \frac{487}{496} \leq \xi \leq 1, \quad -\frac{17}{18} \leq \eta \leq \frac{60\xi - 67}{24\xi - 15} \\ 0, & \text{otherwise} \end{cases}$$

$$f(x,t) = \begin{cases} 201, & \text{for } \frac{487}{496} \leq \xi \leq 1, \quad -\frac{17}{18} \leq \eta \leq \frac{60\xi - 67}{24\xi - 15} \\ 0, & \text{otherwise} \end{cases} .$$

Note the forcing function $f$ and $u$ are defined by the mapping from the reference square $(\xi, \eta) \in [-1,1]^2$. We select $\Delta t = \Delta x$. It is straightforward to verify that $(\beta/\alpha)\Delta t/\Delta x = 1$, so $\widetilde{\mathcal{Q}}_2$ results from augmenting $\mathcal{Q}_2$ with the associated $\psi$ from Table 2. Table 5 records the cell average $\overline{u}_{1,1}$ for both the $\mathcal{Q}_2$ DG and $\widetilde{\mathcal{Q}}_2$ DG schemes. It is evident that the $\mathcal{Q}_2$ DG scheme generates a negative cell average. However, the $\widetilde{\mathcal{Q}}_2$ DG scheme retains a positive cell average.

| $\Delta t = \Delta x$ | Exact | $\mathcal{Q}_2$ | $\widetilde{\mathcal{Q}}_2$ |
|---|---|---|---|
| 1/8 | 9.5E − 03 | −3.5E − 04 | 2.6E − 05 |
| 1/16 | 4.7E − 03 | −1.7E − 04 | 1.3E − 05 |
| 1/32 | 2.3E − 03 | −8.8E − 05 | 6.5E − 06 |
| 1/64 | 1.1E − 03 | −4.4E − 05 | 3.2E − 06 |
| 1/128 | 5.9E − 04 | −2.2E − 05 | 1.6E − 06 |

Table 5: Cell averages $\overline{u}_{1,1}$ generated by the $\mathcal{Q}_2$ DG and $\widetilde{\mathcal{Q}}_2$ DG schemes

### 7.3 Positivity violation for the $\mathcal{P}_2$ DG scheme in (1+1) dimensions

Higher-order $\mathcal{P}_k$ spaces also do not guarantee an unmodulated positive cell average. To see this, we present a counter example for $k = 2$. Take $\alpha = \beta = 1$, $\gamma = 0$ in (1), and put the exact solution as

$$u(x,t) = \left(\frac{31}{2}xt\right)^{13}, \quad (x,t) \in [0, 0.5]^2.$$

The computational domain $[0, 0.5]^2$ is partitioned uniformly, $N_x$ rectangles in the $x$ direction, and $N_t$ rectangles in the $t$ direction. Using Problem 7

| $N_x = N_t$ | Exact | $\mathcal{P}_2$ | $\widetilde{\mathcal{P}}_2$ |
|---|---|---|---|
| 2 | $3.37\text{E} - 03$ | $-1.09\text{E} - 04$ | $1.47\text{E} - 03$ |
| 4 | $5.03\text{E} - 11$ | $-1.62\text{E} - 12$ | $2.19\text{E} - 11$ |
| 8 | $7.49\text{E} - 19$ | $-2.42\text{E} - 20$ | $3.27\text{E} - 19$ |

Table 6: Cell averages $\overline{u}_{1,1}$ generated by the $\mathcal{P}_2$ DG and $\widetilde{\mathcal{P}}_2$ DG schemes

$(r = k + 2)$, we obtain $\psi(x, y) = (1/2)(xt)^2$, which is used to augment the space $\mathcal{P}_2$. We note that this specific $\psi$ is only valid for the CFL condition

$$\frac{(\beta/\alpha)\Delta t}{\Delta x} = 1.$$

We track the cell average in the bottom left cell, $S_{11} = [0, \Delta x] \times [0, \Delta t]$. Table 6 displays the cell average in the cell $S_{11}$ for the $\mathcal{P}_2$ DG and $\widetilde{\mathcal{P}}_2$ DG schemes. The exact cell average is also provided. It is clear that the $\mathcal{P}_2$ DG is unable to preserve a positive cell average. When the $\widetilde{\mathcal{P}}_2$ DG scheme is used, the cell average remains non-negative.

### 7.4 Accuracy test for the DG schemes in (1+1) dimensions

A manufactured solution is used to verify numerically that the correct convergence rates are obtained. We take $\alpha = \beta = \gamma = 1$, and $u(x, t) = \cos^4(x - t) + 10^{-14}$, with $\mathcal{D} = [0, 0.1] \times [0, 2\pi]$. The source $f(x, t)$ can be determined from (1) using the exact solution $u$. We see that $u \geq 0$ and $f \geq 0$. Therefore, the inflow boundary conditions $u(x, 0)$ and $u(0, t)$ are non-negative. The time domain $[0, 0.1]$ is divided into $N_t$ subintervals, and the space domain $[0, 2\pi]$ is divided into $N_x$ subintervals.

Several spaces $\mathcal{X}_k$ are verified with the DG scheme, including their augmented counterparts $\widetilde{\mathcal{X}}_k$. For reproducibility purposes, we use the augmented basis functions described in Appendix B and Appendix C. Errors and orders of accuracy are given in Table 7. The schemes give space-time orders of accuracy of $k + 1$ in the $L^2$ norm.

### 7.5 Step function propagation for the linear problem in (1+1) dimensions

This numerical experiment propagates a step function. The time interval is $[0, 0.1]$ (divided into $N_t = 40$ subintervals), and the space domain is $[0, 2\pi]$ (divided into $N_x = 40$ subintervals). We take $\alpha = \beta = 1$, $f \equiv 0$, and $\gamma = 0$. The initial condition is

$$u(x, 0) = \begin{cases} 1, & \text{for } x \in [3, 4] \\ 0, & \text{otherwise.} \end{cases}$$

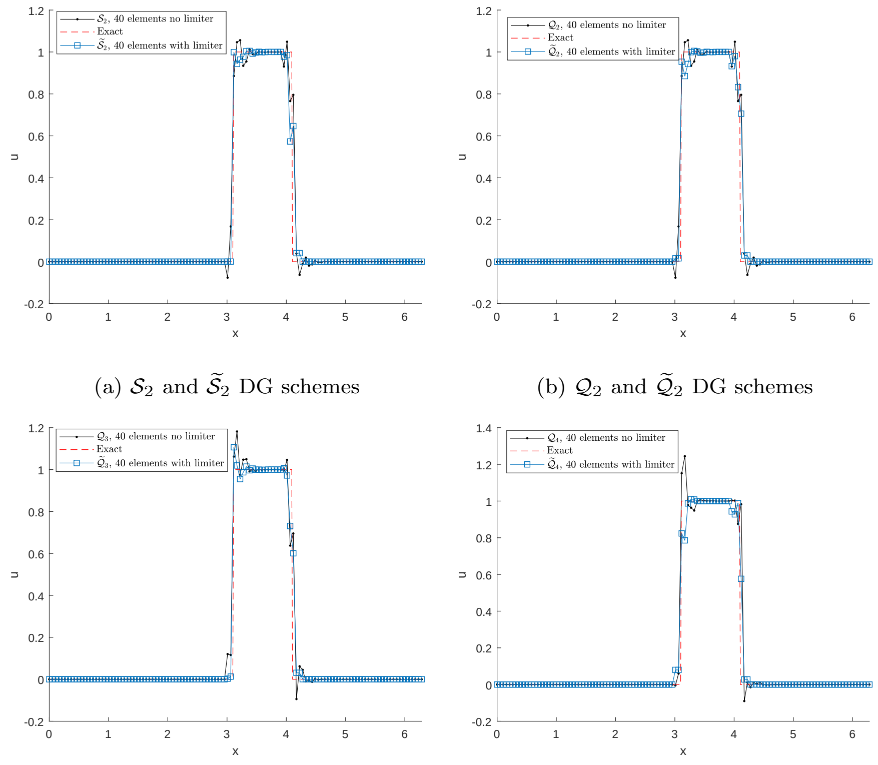
| | | $\mathcal{X}_k$ | | | $\widetilde{\mathcal{X}}_k$ | |
|---|---|---|---|---|---|---|
| | $N_x = N_t$ | $L^2$ error | $L^2$ order | $\min u_h$ | $L^2$ error | $L^2$ order |
| $\mathcal{X}_k = \mathcal{Q}_2$ | 10 | 2.01E − 03 | – | 2.74E − 04 | 2.02E − 03 | – |
| | 20 | 2.97E − 04 | 2.76 | −2.93E − 04 | 2.99E − 04 | 2.75 |
| | 40 | 4.29E − 05 | 2.79 | −1.76E − 05 | 4.32E − 05 | 2.79 |
| | 80 | 5.31E − 06 | 3.01 | −2.98E − 06 | 5.35E − 06 | 3.01 |
| | 160 | 6.63E − 07 | 3.00 | −1.87E − 07 | 6.69E − 07 | 2.99 |
| $\mathcal{X}_k = \mathcal{S}_2$ | 10 | 2.01E − 03 | – | 2.85E − 04 | 2.02E − 03 | – |
| | 20 | 2.97E − 04 | 2.76 | −2.93E − 04 | 2.97E − 04 | 2.76 |
| | 40 | 4.29E − 05 | 2.79 | −1.76E − 05 | 4.28E − 05 | 2.79 |
| | 80 | 5.31E − 06 | 3.01 | −2.98E − 06 | 5.29E − 06 | 3.01 |
| | 160 | 6.63E − 07 | 3.00 | −1.87E − 07 | 6.61E − 07 | 3.00 |
| $\mathcal{X}_k = \mathcal{P}_2$ | 10 | 2.02E − 03 | – | 4.61E − 04 | 2.00E − 03 | – |
| | 20 | 2.98E − 04 | 2.75 | −3.59E − 04 | 2.94E − 04 | 2.76 |
| | 40 | 4.29E − 05 | 2.79 | −2.32E − 05 | 4.23E − 05 | 2.79 |
| | 80 | 5.31E − 06 | 3.01 | −1.46E − 06 | 5.28E − 06 | 3.01 |
| | 160 | 6.64E − 07 | 3.00 | −9.17E − 08 | 6.62E − 07 | 2.99 |
| $\mathcal{X}_k = \mathcal{Q}_3$ | 10 | 1.68E − 04 | – | 2.63E − 04 | 1.67E − 04 | – |
| | 20 | 1.18E − 05 | 3.83 | −1.44E − 05 | 1.17E − 05 | 3.83 |
| | 40 | 8.32E − 07 | 3.82 | −8.58E − 07 | 8.31E − 07 | 3.82 |
| | 80 | 5.18E − 08 | 4.00 | −6.24E − 08 | 5.17E − 08 | 4.00 |
| | 160 | 3.26E − 09 | 3.98 | −4.60E − 09 | 3.26E − 09 | 3.98 |
| $\mathcal{X}_k = \mathcal{P}_4$ | 10 | 1.12E − 05 | – | 8.84E − 07 | 1.11E − 05 | – |
| | 20 | 3.94E − 07 | 4.83 | −8.77E − 08 | 3.89E − 07 | 4.83 |
| | 40 | 1.27E − 08 | 4.94 | −1.44E − 09 | 1.26E − 08 | 4.94 |
| | 80 | 4.01E − 10 | 4.99 | −4.27E − 11 | 3.99E − 10 | 4.98 |

Table 7: Manufactured solution example in 2D (Section 7.4). Errors and rates for the DG schemes with different spaces $\mathcal{X}_k$ and $\widetilde{\mathcal{X}}_k$

The boundary condition is $u(0,t) = 0$. Fig. 7 plots a comparison between the exact solution and the numerical approximations for the various higher-order DG schemes (with and without the positivity-preserving limiter). As expected, the DG schemes without the limiter produce approximations that have negative values. Non-negative approximations result when the limiter is applied. Fig. 8 has a zoom-in near the discontinuities to give the reader a better idea of the oscillations that occur due to the higher-order DG approximation.

### 7.6 Accuracy test for the DG schemes in (1+2) dimensions

A manufactured solution is used to verify numerically that the correct convergence rates are obtained. We take $\alpha = \beta = 8$ and $\gamma = 0.1$. The computational domain is the cube $[0, 0.5]^3$. The exact solution is set as $u(x, y, t) = xye^{t/8}e^{(x^2+y^2)/8}$.



(a) $\mathcal{S}_2$ and $\widetilde{\mathcal{S}}_2$ DG schemes

(b) $\mathcal{Q}_2$ and $\widetilde{\mathcal{Q}}_2$ DG schemes

(c) $\mathcal{Q}_3$ and $\widetilde{\mathcal{Q}}_3$ DG schemes

(d) $\mathcal{Q}_4$ and $\widetilde{\mathcal{Q}}_4$ DG schemes

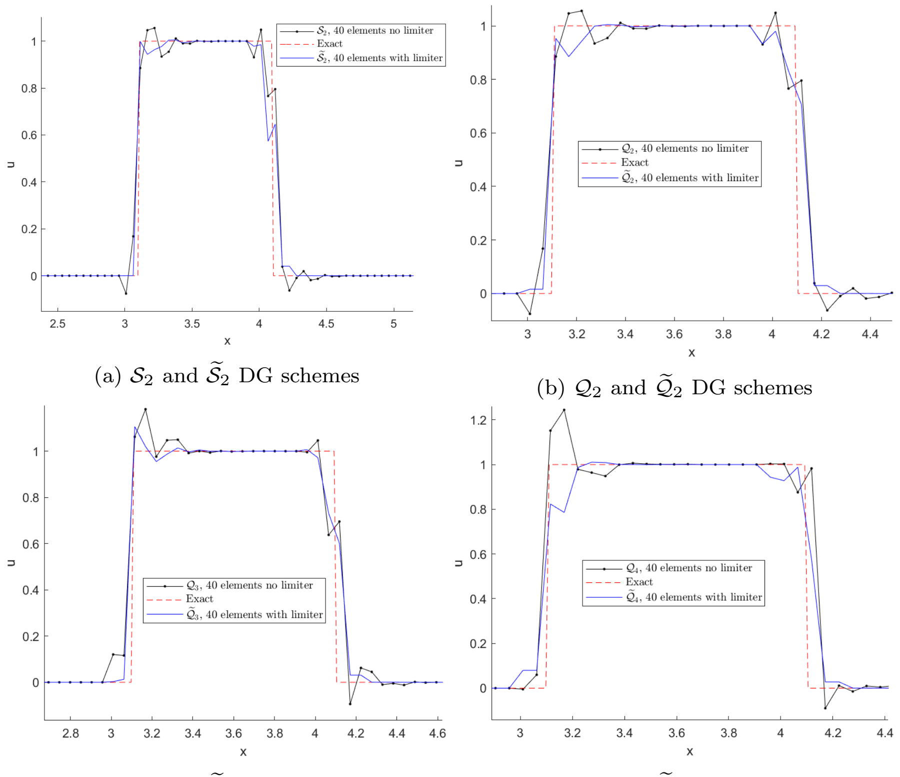
Fig. 7: Comparison of the DG schemes for various polynomial orders, with and without the positivity-preserving limiter

Table 8 has the errors and convergence rates for the various 3D DG schemes. The correct rate of $k+1$ in the $L^2$ norm is approximately observed for the unaugmented and augmented spaces.

### 7.7 Step function initial condition for the linear problem in (1+2) dimensions

This numerical experiment propagates a step function. The time interval is $[0,1]$ (divided into $N_t = 16$ subintervals), and the space domain is $[0,2\pi]^2$ (divided into $N_x = N_y = 32$ subintervals). We take $\alpha = \beta = 1$, $f \equiv 0$, and $\gamma = 0$. The initial condition is

$$u(x,y,0) = \begin{cases} 1, & \text{for } (x,y) \in [2,3] \\ 0, & \text{otherwise.} \end{cases}$$



(a) $\mathcal{S}_2$ and $\widetilde{\mathcal{S}}_2$ DG schemes

(b) $\mathcal{Q}_2$ and $\widetilde{\mathcal{Q}}_2$ DG schemes

(c) $\mathcal{Q}_3$ and $\widetilde{\mathcal{Q}}_3$ DG schemes

(d) $\mathcal{Q}_4$ and $\widetilde{\mathcal{Q}}_4$ DG schemes

Fig. 8: Zoom-in and comparison DG schemes for various polynomial orders, with and without the positivity-preserving limiter

The boundary condition is $u(0, 0, t) = 0$. Fig. 9 plots the exact solution and the numerical approximation for the $\mathcal{P}_2$ DG scheme without the positivity-preserving limiter. Without the limiter, negative undershoots appear in the approximation. Fig. 10 plots the numerical approximation for the $\widetilde{\mathcal{P}}_k$ DG schemes with the positivity-preserving limiter for $k = 2, 3, 4$. The limiter eliminates any negative undershoots.

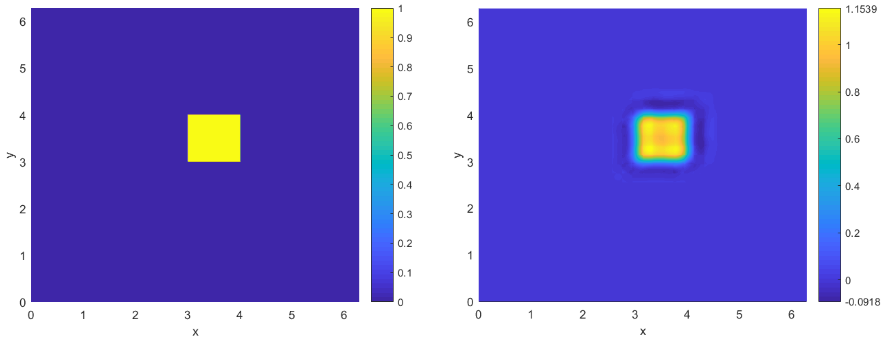
## 7.8 Computational performance

The previous examples focus on verification and validation of the optimization-based enrichment technique. In this section, we focus on the computational performance. We make a few remarks before presenting the test problem and results.

1. If the cell average from the unmodulated DG scheme using $\mathcal{X}_k(S_{ij\ell})$ is negative, we apply Problem (9) (or Problem (7)). Otherwise, for all untroubled cells, there is no need to apply the optimization procedure.

| | | $\mathcal{X}_k$ | | $\widetilde{\mathcal{X}}_k$ | |
|---|---|---|---|---|---|
| | $N_x = N_y = N_t$ | $L^2$ error | $L^2$ order | $L^2$ error | $L^2$ order |
| $\mathcal{X}_k = \mathcal{P}_1$ | 2 | $1.266\text{E}-03$ | $-$ | $1.283\text{E}-01$ | $-$ |
| | 4 | $3.205\text{E}-04$ | 1.98 | $3.462\text{E}-02$ | 1.88 |
| | 8 | $7.763\text{E}-05$ | 2.04 | $8.715\text{E}-03$ | 1.99 |
| | 16 | $1.879\text{E}-05$ | 2.04 | $2.123\text{E}-03$ | 2.03 |
| $\mathcal{X}_k = \mathcal{Q}_2$ | 2 | $3.6674\text{E}-05$ | $-$ | $3.654\text{E}-05$ | $-$ |
| | 4 | $5.4880\text{E}-06$ | 2.74 | $5.472\text{E}-06$ | 2.73 |
| | 8 | $5.9162\text{E}-07$ | 3.21 | $5.871\text{E}-07$ | 3.22 |
| | 16 | $6.5933\text{E}-08$ | 3.16 | $6.532\text{E}-08$ | 3.16 |
| $\mathcal{X}_k = \mathcal{Q}_3$ | 2 | $4.285\text{E}-06$ | $-$ | $4.285\text{E}-06$ | $-$ |
| | 4 | $2.408\text{E}-07$ | 4.14 | $2.411\text{E}-07$ | 4.15 |
| | 8 | $9.773\text{E}-09$ | 4.62 | $9.775\text{E}-09$ | 4.62 |
| | 16 | $4.309\text{E}-10$ | 4.50 | $4.303\text{E}-10$ | 4.50 |

Table 8: Manufactured solution example in 3D (Section 7.6). Errors and rates for the 3D DG schemes with different spaces $\mathcal{X}_k$ and $\widetilde{\mathcal{X}}_k$

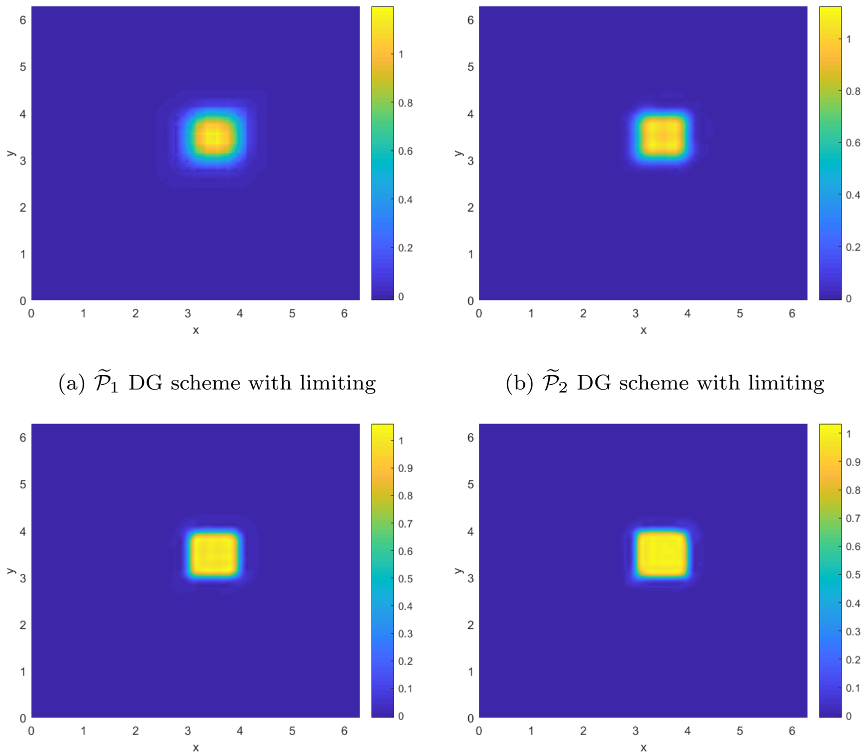


(a) Exact solution at $T = 1$

(b) $\mathcal{P}_2$ DG scheme without limiting

Fig. 9: Exact solution at $T = 1$, and the result of the $\mathcal{P}_2$ DG scheme without the positivity-preserving limiter. The approximation has negative undershoots

2. The model PDE has unidirectional flow, which is known before the simulation starts (for example, the parameters $\alpha, \beta$ from (4) are assumed to be given). This means we can pre-process suitable augmented basis functions sufficient for the expected range of parameters. For instance, in Table 3, several augmented basis functions have been pre-computed which are valid under a fixed CFL condition.
3. Since we know the CFL condition before the simulation begins, we do not require to run the optimization "on the fly" for each cell independently. Rather, similar to computing quadrature points on a reference element, we run numerically solve problem (9) once, at the start of the simulation, for



(a) $\widetilde{\mathcal{P}}_1$ DG scheme with limiting

(b) $\widetilde{\mathcal{P}}_2$ DG scheme with limiting

(c) $\widetilde{\mathcal{P}}_3$ DG scheme with limiting

(d) $\widetilde{\mathcal{P}}_4$ DG scheme with limiting

Fig. 10: Comparison of the augmented DG schemes for various polynomial orders, with the positivity-preserving limiter

a given CFL condition. We then augment the standard DG spaces with this function on troubled cells.

4. We do not focus on enhancing the performance the method. For example, the code neither leverages parallelism nor exploits any tensor product structures. Default options are passed into the numerical interior point optimizer. Additional attention to these details can further improve the performance.

Keeping the above remarks in mind, we treat the construction of the augmented basis function as a pre-processing step - not too dissimilar from mesh generation or quadrature construction on reference cells.

In all computational experiments, we solve the optimization problems (Problems (9) and (7)) approximately via interior point methods for nonlinear programming [3]. We focus on the reference cell $[-1,1]^2$ in 2D and $[-1,1]^3$ in 3D, as is standard in many finite element computations [19]. The parameter space $\Omega_\delta = (\alpha, \beta, \omega, \gamma, \Delta x, \Delta y, \Delta z)$ is sampled so that the augmented basis functions are valid for a wide range of CFL numbers (see Fig. 4). The computational results in 2D can be found in Table 9. We see that as the polynomial degree $k$ increases, the time to find a suitable basis function also increases. It should be emphasized that these timings are done once at the start of the simulation, and are of little overhead (compared to mesh generation, for instance). The augmented tensor product space $\widetilde{\mathcal{Q}}_k$ takes longer to compute than $\widetilde{\mathcal{P}}_k$ since it has more degrees of freedom (and thus, more parameters to search for, see the vector $\vec{d}$ in Problem (7)). On the other hand, if we use Problem (7)) "on the fly", meaning we invoke the optimization problem on each troubled cell, then the timings are significantly reduced - because the augmented function needs to only hold for a *fixed* parameter space $\Omega_\delta = (\alpha, \beta, \omega, \gamma, \Delta x, \Delta y, \Delta z)$. Table 10 displays the timings in this case.

| $k$ | $\widetilde{\mathcal{P}}_k$ | $\widetilde{\mathcal{Q}}_k$ |
|---|---|---|
| 1 | $7.5836E-01$ | $9.6983E-01$ |
| 2 | $9.9454E-01$ | $2.7070E+00$ |
| 3 | $1.8614E+00$ | $7.2597E+00$ |
| 4 | $4.3692E+00$ | $1.5475E+01$ |
| 5 | $1.0319E+01$ | $3.9193E+01$ |
| 6 | $1.9814E+01$ | $1.1344E+02$ |
| 7 | $5.3975E+01$ | $3.1415E+02$ |
| 8 | $8.3156E+01$ | $5.2043E+02$ |

Table 9: Computation times in seconds for $\widetilde{\mathcal{P}}_k$ and $\widetilde{\mathcal{Q}}_k$ in 2D for various polynomial degrees. Here, the augmented functions are enforced to be valid over a large parameter space $\Omega_\delta = (\alpha, \beta, \gamma, \Delta x, \Delta y)$.

| $k$ | $\widetilde{\mathcal{P}}_k$ | $\widetilde{\mathcal{Q}}_k$ |
|---|---|---|
| 1 | $6.524E-04$ | $8.6006E-04$ |
| 2 | $3.584E-03$ | $5.8656E-03$ |
| 3 | $4.441E-03$ | $8.6818E-03$ |
| 4 | $5.224E-03$ | $1.1875E-02$ |
| 5 | $8.226E-03$ | $2.1318E-02$ |
| 6 | $9.272E-03$ | $2.6980E-02$ |
| 7 | $1.961E-02$ | $6.3304E-02$ |
| 8 | $9.273E-02$ | $3.2886E-01$ |

Table 10: Computation times in seconds for $\widetilde{\mathcal{P}}_k$ and $\widetilde{\mathcal{Q}}_k$ in 2D for various polynomial degrees. Here, the augmented functions are enforced to be valid over a *fixed* parameter space $\Omega_\delta = (\alpha, \beta, \gamma, \Delta x, \Delta y)$. For simplicity, $\alpha = \beta = \gamma = 1$, $\Delta x = \Delta y = 2$.

In Table 11, we showcase the timings in 3D. It is evident that the optimization is more expensive. For larger polynomial degrees, the algorithm starts to stagnate more, due to poor initial guesses for the nonlinear optimizer. Also, each iteration is of course more expensive in 3D, which contributes to the poor scaling.

| $k$ | $\widetilde{\mathcal{P}}_k$ | $\widetilde{\mathcal{Q}}_k$ |
|---|---|---|
| 1 | $1.5167E+00$ | $1.9397E+00$ |
| 2 | $2.9836E+00$ | $8.1212E+00$ |
| 3 | $7.4456E+00$ | $2.9039E+01$ |
| 4 | $2.1846E+01$ | $7.7374E+01$ |
| 5 | $6.1914E+01$ | $2.3516E+02$ |
| 6 | $1.3869E+02$ | $5.7435E+02$ |
| 7 | $4.3180E+02$ | $7.1354E+02$ |
| 8 | $8.4840E+02$ | $1.5274E+03$ |

Table 11: Computation times in seconds for $\widetilde{\mathcal{P}}_k$ and $\widetilde{\mathcal{Q}}_k$ in 2D for various polynomial degrees. Here, the augmented functions are enforced to be valid over a large parameter space $\Omega_\delta = (\alpha, \beta, \omega, \gamma, \Delta x, \Delta y, \Delta z)$.

## 8 Conclusion

A high-order positivity-preserving DG scheme is presented in this paper. In [12], a second order accurate positivity-preserving DG scheme is studied. In the two dimensional case, they augment the traditional DG space $\mathcal{P}_1$ with an additional basis function, so that the cell average of the unmodulated DG space remains non-negative. Our method in this paper extends the results found in [12] to polynomial degrees greater than one. To do so, a similar premise is assumed. A standard space such $\mathcal{X}_k$ ($\mathcal{P}_k$, $\mathcal{Q}_k$, or $\mathcal{S}_k$) is augmented with an anzats function $\psi \in \mathcal{X}_r$ for some $r > k$. Then, we numerically search for a suitable function $\psi$ through a constrained nonlinear optimization problem. When $\psi$ is augmented to the standard DG space $\mathcal{X}_k$, it will preserve a positive cell average for the unmodulated DG solution. As the resulting cell average is guaranteed to be positive, the entire solution can be made positive by applying a simple scaling limiter [30]. This limiter scales the solution towards the cell average without reducing its formal accuracy [30].

We observe numerically that the method has good performance for any CFL condition (arbitrarily large or small). It is also feasible to obtain a single augmented basis function that is valid for a wide range of CFL conditions; several explicit basis functions were presented. The augmented DG space involves additional computational cost. Some of this cost may be mitigated by using an adaptive scheme, where $\mathcal{X}_k$ is used when the cell average is positive, and the augmented space is only used when the cell average from $\mathcal{X}_k$ is negative. Moreover, this technique may be used to preprocess suitable augmented basis functions, as opposed to using Problem (7) on the fly.

Numerical experiments are conducted in 2D and 3D, and they demonstrate that the method works as expected for polynomial degrees $k > 1$. Several standard DG spaces are also explored, such as $\mathcal{P}_k$, $\mathcal{Q}_k$, and the serendipity spaces $\mathcal{S}_k$. Additionally, careful convergence studies are provided.

In future work we plan to examine extension of the presented scheme for nonlinear problems, applications such as radiative transport, and unstructured meshes.

**Funding** This study was funded by the University of Wisconsin-Madison.

**Conflicts of interest/Competing interests** All authors have no conflicts of interest.

**Data Availability** Enquirers about data availability should be directed to the authors

## Declarations

**Conflict of interest** The author has no competing interests to declare that are relevant to the content of this article.

## Appendix A

Below (see (15)) are the tensor product Bernstein polynomials used as a basis for $\mathcal{Q}_2$.

$$\phi_1(\xi,\eta) = (\eta/2 - 1/2)^2(\xi/2 - 1/2)^2 \tag{15a}$$
$$\phi_2(\xi,\eta) = -(\eta/2 - 1/2)(\xi/2 - 1/2)^2(\eta + 1) \tag{15b}$$
$$\phi_3(\xi,\eta) = (\eta/2 + 1/2)^2(\xi/2 - 1/2)^2 \tag{15c}$$
$$\phi_4(\xi,\eta) = -(\eta/2 - 1/2)^2(\xi/2 - 1/2)(\xi + 1) \tag{15d}$$
$$\phi_5(\xi,\eta) = (\eta/2 - 1/2)(\xi/2 - 1/2)(\eta + 1)(\xi + 1) \tag{15e}$$
$$\phi_6(\xi,\eta) = -(\eta/2 + 1/2)^2(\xi/2 - 1/2)(\xi + 1) \tag{15f}$$
$$\phi_7(\xi,\eta) = (\eta/2 - 1/2)^2(\xi/2 + 1/2)^2 \tag{15g}$$
$$\phi_8(\xi,\eta) = -(\eta/2 - 1/2)(\xi/2 + 1/2)^2(\eta + 1) \tag{15h}$$
$$\phi_9(\xi,\eta) = (\eta/2 + 1/2)^2(\xi/2 + 1/2)^2 \tag{15i}$$

Case for $B < 0.5$, $\gamma = 0$

The purpose of this section is to provide analytic evidence that $\{\psi_{1,\mathcal{Q}_2}\}$ (see 2) augmented to $\mathcal{Q}_2$ ensures a non-negative cell average provided that $B < 1/2$ and $\gamma = 0$. Recall $\hat{S} = [-1,1]^2$ is the reference cell. Write $\widetilde{\mathcal{Q}}_2 = \mathcal{Q}_2 \cup \operatorname{span}\{\psi_{1,\mathcal{Q}_2}\}$. Express the solution to the linear system arising from

$$\mathcal{L}(u,v) = \iint_{\widehat{S}} u \, d\xi d\eta, \quad \forall u \in \widetilde{\mathcal{Q}}_2(\widehat{S}), \tag{16}$$

as

$$v(\xi,\eta)=\sum_{i=1}^{9}c_i\phi_i(\xi,\eta)+c_{10}\psi_{1,\mathcal{Q}_2}(\xi,\eta),$$

where $\psi_{1,\mathcal{Q}_2}$ is the augmented basis function. With the choice $\psi_{1,\mathcal{Q}_2}=(\xi\eta^2(1-\xi)^2(\xi+1)(1-\eta^2)/8)^2$, and $\phi_i$ are defined in (15), solving the linear system arising from (16) generates the following coefficients:

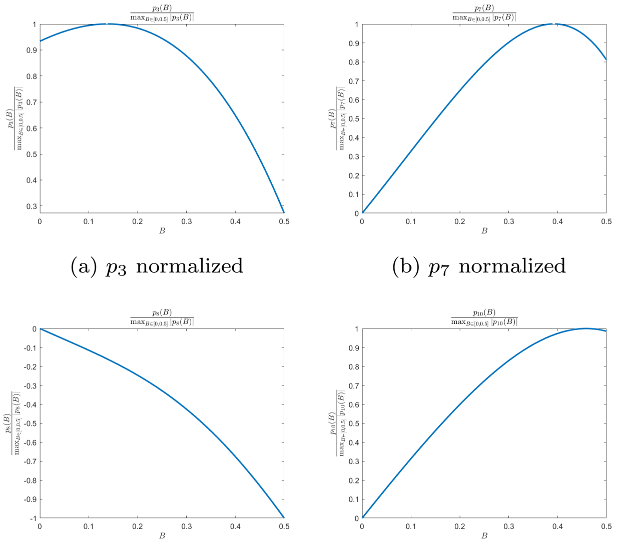
$$\begin{aligned}
c_1 &= \frac{\Delta x}{(B+1)\Lambda}(2020B^7+5718B^6+28920B^5+29148B^4+60834B^3+42181B^2+9265B+2350)\\
c_2 &= \frac{\Delta x}{(B+1)\Lambda}(1510B^7+4364B^6+11517B^5+12607B^4+30325B^3+42821B^2+23365B+2350)\\
c_3 &= \frac{\Delta x}{(B+1)\Lambda}(-190B^6+1944B^5-2174B^4-9774B^3-5889B^2+2215B+2350)\\
c_4 &= \frac{\Delta x}{(B+1)\Lambda}(2020B^7+20838B^6+37248B^5+38336B^4+29439B^3+2828B^2-955B+1175)\\
c_5 &= \frac{\Delta x}{(B+1)\Lambda}(1110B^7+5434B^6+22499B^5+53315B^4+46611B^3+26933B^2+9545B+1175)\\
c_6 &= \frac{\Delta x}{(B+1)\Lambda}(400B^6-3180B^5-2241B^4+7958B^3+6013B^2-955B+1175)\\
c_7 &= \frac{\Delta x}{(B+1)\Lambda}(2020B^7+658B^6-7744B^5-8846B^4-4256B^3+1595B^2+3675B)\\
c_8 &= \frac{\Delta x}{(B+1)\Lambda}(960B^7+5004B^6+5566B^5-5072B^4-5728B^3+815B^2-1575B)\\
c_9 &= \frac{\Delta x}{(B+1)\Lambda}(B(-110B^5+1556B^4+3632B^3+1640B^2+835B+3675))\\
c_{10} &= \frac{\Delta x}{4\Lambda}(-121275B(50B^5-70B^4+639B^3+317B^2+95B-525))\\
\Lambda &= (2020B^7+8228B^6+14544B^5+19129B^4+22737B^3+19926B^2+10440B+2350)
\end{aligned}$$

Using the assumption $B\leq 0.5$, we can show $c_i\geq 0$ for $i\neq 8$. Note that $\Lambda>0$, as it is a polynomial with positive coefficients. Therefore, the denominators of $c_i$ are positive. To inspect the sign of $c_i$, it suffices to examine the univariate polynomial terms (functions of $B$) in the numerator. Define these polynomials as follows: $p_i(B)=((B+1)\Lambda/\Delta x)c_i$ for $1\leq i\leq 9$ and $p_{10}(B)=(4\Lambda/\Delta x)c_{10}$.

It is straightforward to verify that $p_1,p_2,p_5$ are the sum of non-negative terms. The following lower bounds are immediate:

$$\begin{aligned}
p_4 &\geq -955B+1175\geq 0,\\
p_6 &\geq -3180B^5-2241B^4-955B+1175\geq 0,\\
p_9 &\geq B(3675-110B^5)\geq 0.
\end{aligned}$$

The remaining polynomials $p_3,p_7,p_8$, and $p_{10}$ are also non-negative (see Figure 16). Thus, $c_8$ is the only coefficient that is non-positive. With this result,



(a) $p_3$ normalized

(b) $p_7$ normalized

(c) $p_8$ normalized

(d) $p_{10}$ normalized

Fig. 11: Case of $B \leq 0.5$, $G = \gamma = 0$. Plots of $\frac{p_i(B)}{\max_{B\in[0,0.5]}|p_i(B)|}$ for $i = 3, 7, 8, 10$.

we have

$$\begin{aligned} v(\xi,\eta) &= \sum_{i=1}^{9} c_i\phi_i(\xi,\eta) + c_{10}\psi_{1,\mathcal{Q}_2}(\xi,\eta) \\ &\geq c_9\phi_9(\xi,\eta) + c_8\phi_8(\xi,\eta) \\ &= \frac{1}{4}(\xi+1)^2\left(c_9\frac{(1+\eta)(1+\eta)}{4} + c_8\frac{(1+\eta)(1-\eta)}{2}\right) \\ &= \frac{1}{4}(\xi+1)^2 w(\eta), \end{aligned} \tag{17}$$

where $w(y) = c_9\frac{(1+\eta)(1+\eta)}{4} + c_8\frac{(1+\eta)(1-\eta)}{2}$, and we used $c_i \geq 0$ for $i \neq 8$ and the non-negativity of the basis functions. The quadratic function $w$ is concave because $w''(\eta) = \frac{c_9}{2} - c_8 \geq 0$. Moreover, $w(\pm 1) \geq 0$, therefore, $w(\eta) \geq 0$. From (17), it is evident that $v(\xi,\eta) \geq 0$.

## Appendix B: $\mathcal{Q}_3$ and $\mathcal{Q}_4$ augmentation

This appendix contains examples of augmented basis functions for $\mathcal{Q}_k$ generated from Problem (7). Table 2 is duplicated here for convenience.

| Condition to enforce $\overline{u}_{ij} \geq 0$ | Augmented basis function $\psi$ | Original space |
|---|---|---|
| $(\beta/\alpha)\Delta x/\Delta y < \frac{1}{2}$ | $\psi_{1,\mathcal{Q}_2} = (xy^2(1-x)^2(x+1)(1-y^2)/8)^2$ | $\mathcal{Q}_2$ |
| $(\beta/\alpha)\Delta x/\Delta y > 2$ | $\psi_{2,\mathcal{Q}_2} = (yx^2(1-y)^2(y+1)(1-x^2)/8)^2$ | $\mathcal{Q}_2$ |
| $(\beta/\alpha)\Delta x/\Delta y < \frac{1}{4}$ | $\psi_{1,\mathcal{Q}_3}$ (see (22)) | $\mathcal{Q}_3$ |
| $(\beta/\alpha)\Delta x/\Delta y > 4$ | $\psi_{2,\mathcal{Q}_3}$ (see (23)) | $\mathcal{Q}_3$ |
| $(\beta/\alpha)\Delta x/\Delta y < \frac{1}{8}$ | $\psi_{1,\mathcal{Q}_4}$ (see (30)) | $\mathcal{Q}_4$ |
| $(\beta/\alpha)\Delta x/\Delta y > 8$ | $\psi_{2,\mathcal{Q}_4}$ (see (37)) | $\mathcal{Q}_4$ |

Table 12: Example augmented basis functions for $\mathcal{Q}_k$ generated by Problem 7.

$$r_1(\xi,\eta) = (\xi - \sqrt{5}/5)(\eta - \sqrt{5}/5)(\eta + \sqrt{5}/5)(\xi - 1)(\xi + 1)(\eta - 1) \quad (18)$$

$$r_2(\xi,\eta) = (\xi + \sqrt{5}/5)(\eta - \sqrt{5}/5)(\eta + \sqrt{5}/5)(\xi - 1)(\xi + 1)(\eta + 1) \quad (19)$$

$$r_3(\xi,\eta) = (\xi - \sqrt{5}/5)(\xi + \sqrt{5}/5)(\eta - \sqrt{5}/5)(\xi - 1)(\eta - 1)(\eta + 1) \quad (20)$$

$$r_4(\xi,\eta) = (\xi - \sqrt{5}/5)(\xi + \sqrt{5}/5)(\eta + \sqrt{5}/5)(\xi + 1)(\eta - 1)(\eta + 1) \quad (21)$$

$$\psi_{1,\mathcal{Q}_3}(\xi,\eta) = ((25\sqrt{5})^2 r_1(\xi,\eta) r_2(\xi,\eta)/8^4)^2 \quad (22)$$

$$\psi_{2,\mathcal{Q}_3}(\xi,\eta) = ((25\sqrt{5})^2 r_3(\xi,\eta) r_4(\xi,\eta)/8^4)^2 \quad (23)$$

Below is the augmented function ($\psi_{1,\mathcal{Q}_4}$, (30))) for $\mathcal{Q}_4$ with $B < 1/8$.

$$w_1(\xi) = \xi(\xi - 1)(\xi - (\sqrt{3}\sqrt{7})/7)(\xi + (\sqrt{3}\sqrt{7})/7) \quad (24)$$

$$w_2(\eta) = \eta(\eta - 1)(\eta - (\sqrt{3}\sqrt{7})/7)(\eta + (\sqrt{3}\sqrt{7})/7) \quad (25)$$

$$w_3(\xi) = \xi(\xi - 1)(\xi + 1)(\xi - (\sqrt{3}\sqrt{7})/7) \quad (26)$$

$$w_4(\eta) = \eta(\eta + 1)(\eta - (\sqrt{3}\sqrt{7})/7)(\eta + (\sqrt{3}\sqrt{7})/7) \quad (27)$$

$$r_1(\xi,\eta) = (49 w_1(\xi) w_2(\eta))/64 \quad (28)$$

$$r_2(\xi,\eta) = -(343 w_3(\xi) w_4(\eta))/192 \quad (29)$$

$$\psi_{1,\mathcal{Q}_4}(\xi,\eta) = (r_1(\xi,\eta) r_2(\xi,\eta))^2 \quad (30)$$

Below is the augmented function ($\psi_{2,\mathcal{Q}_4}$, (37)) for $\mathcal{Q}_4$ with $B > 8$.

$$w_1(\xi) = \xi(\xi - 1)(\xi - (\sqrt{3}\sqrt{7})/7)(\xi + (\sqrt{3}\sqrt{7})/7) \quad (31)$$

$$w_2(\eta) = \eta(\eta - 1)(\eta - (\sqrt{3}\sqrt{7})/7)(\eta + (\sqrt{3}\sqrt{7})/7) \quad (32)$$

$$w_3(\xi) = \xi(\xi + 1)(\xi - (\sqrt{3}\sqrt{7})/7)(\xi + (\sqrt{3}\sqrt{7})/7) \quad (33)$$

$$w_4(\eta) = \eta(\eta - 1)(\eta + 1)(\eta - (\sqrt{3}\sqrt{7})/7) \quad (34)$$

$$r_1(\xi,\eta) = (49 w_1(\xi) w_2(\eta))/64 \quad (35)$$

$$r_2(\xi,\eta) = -(343 w_3(\xi) w_4(\eta))/192 \quad (36)$$

$$\psi_{2,\mathcal{Q}_4}(\xi,\eta) = (r_1(\xi,\eta) r_2(\xi,\eta))^2 \quad (37)$$

## Appendix C: Augmentation for $\mathcal{S}_3$ and $\mathcal{S}_4$

This appendix contains examples of augmented basis functions for $\mathcal{S}_k$ generated from Problem (7). Table 3 is included here for convenience.

| Condition to enforce $\overline{u}_{ij} \geq 0$ | Augmented basis function $\psi$ | Original space |
|---|---|---|
| $(\beta/\alpha)\Delta x/\Delta y < \frac{1}{2}$ | $\psi_{1,\mathcal{S}_2} = (\xi\eta^2(\xi-1)^2(\eta-1)(\xi+1)(\eta+1)/8)^2$ | $\mathcal{S}_2$ |
| $(\beta/\alpha)\Delta x/\Delta y > 2$ | $\psi_{2,\mathcal{S}_2} = (\xi^2\eta(\xi-1)(\eta-1)^2(\xi+1)(\eta+1)/2)^2$ | $\mathcal{S}_2$ |
| $(\beta/\alpha)\Delta x/\Delta y < \frac{1}{8}$ | $\psi_{1,\mathcal{S}_3}$ (see (42)) | $\mathcal{S}_3$ |
| $(\beta/\alpha)\Delta x/\Delta y > 8$ | $\psi_{2,\mathcal{S}_3}$ (see (43)) | $\mathcal{S}_3$ |
| $(\beta/\alpha)\Delta x/\Delta y < \frac{1}{16}$ | $\psi_{1,\mathcal{S}_4}$ (see (48)) | $\mathcal{S}_4$ |
| $(\beta/\alpha)\Delta x/\Delta y > 16$ | $\psi_{2,\mathcal{S}_4}$ (see (49)) | $\mathcal{S}_4$ |

Table 13: Example augmented basis functions for the serendipity spaces $\mathcal{S}_k$ generated by Problem 7.

$$r_1(\xi,\eta) = (\xi - \sqrt{5}/5)(\eta - \sqrt{5}/5)^2(\xi-1)^2(\xi+1)^2(\eta-1)(\eta+1)^2 \tag{38}$$

$$r_2(\xi,\eta) = \sqrt{5}(\xi + \sqrt{5}/5)(\eta + \sqrt{5}/5) \tag{39}$$

$$r_3(\xi,\eta) = (\xi - \sqrt{5}/5)^2(\eta - \sqrt{5}/5)(\xi-1)(\xi+1)^2(\eta-1)^2(\eta+1)^2 \tag{40}$$

$$r_4(\xi,\eta) = \sqrt{5}(\xi + \sqrt{5}/5)(\eta + \sqrt{5}/5) \tag{41}$$

$$\psi_{1,\mathcal{S}_3}(\xi,\eta) = ((5^5/64^2)r_1(\xi,\eta)r_2(\xi,\eta))^2 \tag{42}$$

$$\psi_{2,\mathcal{S}_3}(\xi,\eta) = ((5^5/64^2)r_3(\xi,\eta)r_4(\xi,\eta))^2 \tag{43}$$

$$r_1(\xi,\eta) = \xi^2\eta^2(\xi-1)^2(\eta-1)(\xi - (\sqrt{3}\sqrt{7})/7)(\xi + (\sqrt{3}\sqrt{7})/7)^2 \tag{44}$$

$$r_2(\xi,\eta) = (\xi+1)(\eta+1)(\eta - (\sqrt{3}\sqrt{7})/7)^2(\eta + (\sqrt{3}\sqrt{7})/7)^2 \tag{45}$$

$$r_3(\xi,\eta) = \xi^2\eta(\xi-1)(\xi+1)^2(\eta-1)^2(\xi - (\sqrt{3}\sqrt{7})/7)^2 \tag{46}$$

$$r_4(\xi,\eta) = (\eta+1)(\xi + (\sqrt{3}\sqrt{7})/7)(\eta - (\sqrt{3}\sqrt{7})/7)^2(\eta + (\sqrt{3}\sqrt{7})/7)^2 \tag{47}$$

$$\psi_{1,\mathcal{S}_4}(\xi,\eta) = ((343/192)(49/64)r_1(\xi,\eta)r_2(\xi,\eta))^2 \tag{48}$$

$$\psi_{2,\mathcal{S}_4}(\xi,\eta) = ((343/192)(49/24)r_3(\xi,\eta)r_4(\xi,\eta))^2 \tag{49}$$

## References

1. Abramowitz, M., Stegun, I.: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Applied mathematics series. U.S. Government Printing Office (1965)
2. Arnold, D., Boffi, D., Falk, R.: Approximation by quadrilateral finite elements. Mathematics of computation **71**(239), 909–922 (2002)
3. Byrd, R.H., Gilbert, J.C., Nocedal, J.: A trust region method based on interior point techniques for nonlinear programming. Mathematical programming **89**(1), 149–185 (2000)

4. Cheng, Y., Gamba, I., Proft, J.: Positivity-preserving discontinuous Galerkin schemes for linear Vlasov-Boltzmann transport equations. Mathematics of Computation **81**(277), 153–190 (2012)
5. Courant, R., Friedrichs, K., Lewy, H.: On the partial difference equations of mathematical physics. IBM journal of Research and Development **11**(2), 215–234 (1967)
6. Fabien, M.S.: Positivity-preserving discontinuous galerkin scheme for linear hyperbolic equations with characteristics-informed augmentation. Results in Applied Mathematics **22**, 100460 (2024)
7. Guo, L., Yang, Y.: Positivity preserving high-order local discontinuous Galerkin method for parabolic equations with blow-up solutions. Journal of Computational Physics **289**, 181–195 (2015)
8. Kontzialis, K., Ekaterinaris, J.A.: High order discontinuous Galerkin discretizations with a new limiting approach and positivity preservation for strong moving shocks. Computers & Fluids **71**, 98–112 (2013)
9. Kovvali, N.: Theory and Applications of Gaussian Quadrature Methods. Synthesis digital library of engineering and computer science. Morgan & Claypool (2011)
10. LeVeque, R.J.: Numerical methods for conservation laws, vol. 3. Springer (1992)
11. Li, M., Guyenne, P., Li, F., Xu, L.: A positivity-preserving well-balanced central discontinuous Galerkin method for the nonlinear shallow water equations. Journal of Scientific Computing **71**, 994–1034 (2017)
12. Ling, D., Cheng, J., Shu, C.W.: Conservative high order positivity-preserving discontinuous Galerkin methods for linear hyperbolic and radiative transfer equations. Journal of Scientific Computing **77**(3), 1801–1831 (2018)
13. Liu, C., Buzzard, G.T., Zhang, X.: An optimization based limiter for enforcing positivity in a semi-implicit discontinuous galerkin scheme for compressible Navier–Stokes equations. Journal of Computational Physics **519**, 113440 (2024)
14. Liu, C., Hu, J., Taitano, W.T., Zhang, X.: An optimization-based positivity-preserving limiter in semi-implicit discontinuous galerkin schemes solving Fokker-Planck equations (2024). URL https://arxiv.org/abs/2410.19143
15. Liu, C., Riviere, B., Shen, J., Zhang, X.: A simple and efficient convex optimization based bound-preserving high order accurate limiter for Cahn–Hilliard–navier–stokes system. SIAM Journal on Scientific Computing **46**(3), A1923–A1948 (2024)
16. Meena, A.K., Kumar, H., Chandrashekar, P.: Positivity-preserving high-order discontinuous Galerkin schemes for ten-moment Gaussian closure equations. Journal of Computational Physics **339**, 370–395 (2017)
17. Nocedal, J., Wright, S.: Numerical optimization. Springer Science & Business Media (2006)
18. Pistikopoulos, E., Diangelakis, N., Oberdieck, R.: Multi-parametric Optimization and Control. Wiley Series in Operations Research and Management Science. Wiley (2020)
19. Quarteroni, A., Sacco, R., Saleri, F.: Numerical Mathematics. Texts in Applied Mathematics. Springer New York (2001)
20. Reemtsen, R., Rückmann, J.: Semi-Infinite Programming. Nonconvex Optimization and Its Applications. Springer US (2013)
21. Rossmanith, J.A., Seal, D.C.: A positivity-preserving high-order semi-Lagrangian discontinuous Galerkin scheme for the Vlasov–Poisson equations. Journal of Computational Physics **230**(16), 6203–6232 (2011)
22. Srinivasan, S., Poggie, J., Zhang, X.: A positivity-preserving high order discontinuous Galerkin scheme for convection–diffusion equations. Journal of Computational Physics **366**, 120–143 (2018)
23. van der Vegt, J.J.W., Xia, Y., Xu, Y.: Positivity preserving limiters for time-implicit higher order accurate discontinuous Galerkin discretizations. SIAM journal on scientific computing **41**(3), A2037–A2063 (2019)
24. Wen, X., Gao, Z., Don, W.S., Xing, Y., Li, P.: Application of positivity-preserving well-balanced discontinuous Galerkin method in computational hydrology. Computers & Fluids **139**, 112–119 (2016)
25. Xu, Z., Shu, C.W.: High order conservative positivity-preserving discontinuous Galerkin method for stationary hyperbolic equations. Journal of Computational Physics **466**, 111410 (2022)

26. Xu, Z., Shu, C.W.: On the conservation property of positivity-preserving discontinuous Galerkin methods for stationary hyperbolic equations. Journal of Computational Physics **490**, 112304 (2023)
27. Yee, B.C., Olivier, S.S., Haut, T.S., Holec, M., Tomov, V.Z., Maginot, P.G.: A quadratic programming flux correction method for high-order DG discretizations of SN transport. arXiv preprint arXiv:1910.02918 (2019)
28. Zhang, M., Cheng, J., Qiu, J.: High order positivity-preserving discontinuous Galerkin schemes for radiative transfer equations on triangular meshes. Journal of Computational Physics **397**, 108811 (2019)
29. Zhang, R., Zhu, J., Loula, A.F., Yu, X.: Operator splitting combined with positivity-preserving discontinuous Galerkin method for the chemotaxis model. Journal of Computational and Applied Mathematics **302**, 312–326 (2016)
30. Zhang, X., Shu, C.W.: On maximum-principle-satisfying high order schemes for scalar conservation laws. Journal of Computational Physics **229**(9), 3091–3120 (2010)
31. Zhao, X., Yang, Y., Seyler, C.E.: A positivity-preserving semi-implicit discontinuous Galerkin scheme for solving extended magnetohydrodynamics equations. Journal of Computational Physics **278**, 400–415 (2014)