A Local Projection Stabilization method with shock capturing and diagonal mass matrix for solving non-stationary transport dominated problems *

F. Schieweck† and P. Skrzypacz†

Abstract

We consider a time-dependent convection diffusion equation in the transport dominated case. As a stabilization method in space we propose a new variant of Local Projection Stabilization (LPS) which uses special enriched bubble functions such that L^2-orthogonal local basis functions can be constructed. L^2-orthogonal basis functions lead to a diagonal mass matrix which is advantageous for time discretization. We use the discontinuous Galerkin method of order one for the discretization in time. In order to avoid the remaining oscillations in the LPS-solution we add for each time step in the space discretization an extra shock capturing term which acts only locally on those mesh cells where an error-indicator is relatively large. The novelty in the shock capturing term is that the scaling factor in front of the additive diffusion term is computed from a low order post-processing error. As a result we obtain both, an oscillation-free discrete solution and the information about the local regions where this solution is still inaccurate due to some smearing. The latter information can be used to create in each time step an adaptively refined space mesh. Whereas the numerical experiments are restricted to one space dimension the proposed ideas work also in the multi-dimensional spatial case. The numerical tests show that the discrete solution with shock capturing is oscillation-free and of optimal accuracy in the regions outside of the shock.

Keywords: Local Projection Stabilization, discontinuous Galerkin time discretization, shock capturing, post-processing, error indicator

2010 Mathematics Subject Classification (MSC): 65N30, 65M60

1 Introduction

In this paper, we consider the numerical solution of a time-dependent convection diffusion equation in the case of dominant convection. For this problem, one needs both a stable discretization in time due to the stiffness caused by small mesh-sizes in space as well as a stable space discretization due to the transport dominant character of the problem. In order to

*appears as Preprint No. 02-2012, Fakultät für Mathematik, Otto-von-Guericke Universität Magdeburg

†Institut für Analysis und Numerik, Otto-von-Guericke-Universität Magdeburg, Postfach 4120, D-39016 Magdeburg, Germany

get a good numerical efficiency we want to apply higher order methods for both discretization parts.

Concerning the time discretization, we choose the discontinuous Galerkin method of order one (dG(1)-method) because of the following advantages. Applied to the semi-discretization in space, the dG(1)-method leads to a strongly A-stable (L-stable) scheme which is super-convergent of order three in the endpoints of the time intervals. Also in the case, when we start first with the discretization in time (Rothe's method), the dG(1)-method has nice stability properties and leads to a time-discrete solution in the Hilbert space which is of order three in the discrete time points (see e.g. [22]). In this paper, we will follow this last approach which leads on each time interval to a linear coupled system of partial differential equations (pde's) for two space variables. Although the numerical costs for solving this system approximately by means of a finite element discretization are higher compared to the usual backward Euler or Crank-Nicolson method, the advantage of having a time error of order three for dG(1) allows to work with a much larger time step to achieve the same accuracy in time. So these higher costs will pay off at the end when we add the costs over all time steps.

One new contribution of this paper is that we try to construct systematically a finite element space for the space discretization that has L^2-orthogonal basis functions with local support. The advantage of such a basis is that we get a diagonal mass matrix in the space-discretization of the pde-system on each time interval. This means that we need less memory to store the mass matrix and the multiplication with the inverse mass matrix is just a cheap row scaling operation. The last property can be exploited for the elimination of one block-unknown in the coupled discrete 2×2-block system that approximates the pde-system on the actual time interval.

Among the stable space-discretizations which are known to work well for the transport dominant case of the convection diffusion equation we choose the LPS-method [11, 1, 17, 18]. The reason is firstly that we have here only a small number of stabilization terms which are even symmetric in contrast to the streamline diffusion method and secondly that the coupling pattern in the system matrix is much more sparse compared to that of the Continuous Interior Penalty method [2, 3] or the discontinuous Galerkin method [10, 9, 4]. In addition to the two-level variant of LPS there is also a one-level variant based on an enriched finite element space which we will apply here. In order to construct an L^2-orthogonal set of basis functions we exploit the freedom in the choice of the enriched finite element space. The extra bubble functions that define the space enrichment are chosen such that the original finite element basis functions can be modified by taking suitable linear combinations with bubble functions. We present the idea to construct an L^2-orthogonal basis first in detail for the one-dimensional case and give then some comments for the generalization to the 2D- or 3D-case under the assumption that the mesh consists of parallelograms or parallelepipeds. Thus, we will construct three new enriched $\mathbb{Q}_r^+$-elements for $r = 1, 2, 3$.

Another new contribution of our paper is that we introduce a novel kind of shock capturing which is based on a low order post-processing error indicator. The shock capturing method is a technique to avoid remaining unphysical oscillations in a stabilized discretization which are mainly caused by unresolved sharp layers in the solution. In the literature, we find

shock capturing techniques proposed by Hughes and Johnson and coworkers [13, 14, 6] in combination with the streamline diffusion finite element method (SDFEM). The idea is to add artificial diffusion in crosswind direction depending on the gradient of the discrete solution. The drawback is that one gets a nonlinear stabilization term even if the original problem is linear. Some important references concerning shock capturing methods based on nonlinear artificial diffusion can be found in [20, p. 319 - 321]. In particular, there are methods that use the local residual of the solution to control the scaling factor for the artificial diffusion [15, 16]. Guermond and coworkers [7, 12] add artificial diffusion in all directions depending on the gradient of the fine scale modes in the subgrid viscosity method. They also get a nonlinear additive term in the discretization. Feistauer et al. [5] work with the discontinuous Galerkin method as the underlying stable discretization and try to avoid the remaining oscillations by reducing the polynomial order in the critical regions. These regions are detected by the criterion that the jumps at the element boundaries are high. In our approach, we add like Guermond artificial diffusion in all directions but only on those mesh cells where a post-processing error indicator applied to the discrete solution is high. Also the scaling factor in front of the artificial diffusion term on a mesh cell is new. We choose here the discrete maximum norm of the error between the original discrete solution and the low order post-processed solution. The computation of these error indicators is very cheap; the costs are proportional to the number of degrees of freedom in the finite element space. By numerical experiments for a model problem with moving shocks we show that our local shock capturing method really avoids oscillations in the numerical solution and that it maintains the optimal high order of accuracy in regions outside of the shocks. Moreover, we demonstrate that our low order post-processing indicator detects in the right way the region of the shock at each discrete time. Thus, the amount of smearing caused by the local artificial diffusion can be reduced in future by adaptive mesh refinement. Another feature of our shock capturing approach is that we avoid the nonlinearity of the additive shock capturing terms by choosing these terms only in dependence of the discrete solution at the beginning of the actual time interval. Of course we have then to guarantee that the shock position does not move too much during one time step in relation to the size of the mesh cell.

2 Model problem

Let $\Omega \subset \mathbb{R}^d$ ($d = 1, 2$ or 3), denote a bounded spatial domain with a polygonal boundary $\partial\Omega$. For a given final time $T > 0$, we consider the time-dependent partial differential equation

$$
\begin{aligned}
&\textit{Find } u : \Omega \times [0,T] \to \mathbb{R} \textit{ such that} \\
&\begin{aligned}
\partial_t u - \varepsilon \Delta u + b \cdot \nabla u + cu &= f \quad \text{in} \quad \Omega \times (0,T)\,, \\
u &= 0 \quad \text{on} \quad \partial\Omega \times [0,T]\,, \\
u &= u_0 \quad \text{on} \quad \Omega \times \{0\}\,,
\end{aligned}
\end{aligned}
\tag{1}
$$

where $0 < \varepsilon \ll 1$ is a small diffusion constant, $b : \Omega \to \mathbb{R}^d$ a convection field, $c : \Omega \to \mathbb{R}$ the reaction and $u_0 : \Omega \to \mathbb{R}$ the initial value of the solution. For simplicity, we have assumed

that b and c are time independent and that the prescribed Dirichlet boundary data are homogeneous. However, the source function $f : \Omega \times (0,T) \to \mathbb{R}$ may depend on time t. Note that our proposed methods can easily be generalized to the non-homogeneous time-dependent case. In this paper, we use the following standard notation. For a given domain $G \subset \mathbb{R}^d$, the space $H^m(G)$ denotes the set of $L^2(G)$-functions that have weak derivatives in $L^2(G)$ up to the order m. The subspace of functions from $H^1(G)$ having a zero boundary trace is denoted by $H^1_0(G)$ and the inner product of $L^2(G)$ by $(\cdot,\cdot)_G$. With $\|\cdot\|_{m,G}$ and $|\cdot|_{m,G}$ we denote the standard norm and semi-norm of $H^m(G)$ and we will omit the index G in the case $G = \Omega$.

Using the solution space $W := H^2(\Omega) \cap H^1_0(\Omega)$, the test space $L := L^2(\Omega)$ and the linear operator A defined by

$$A : W \to L, \qquad Au := -\varepsilon \triangle u + b \cdot \nabla u + cu,$$

the weak formulation of (1) reads as follows:

$$\begin{aligned} &\textit{Find } t \mapsto u(t) \in W \textit{ such that } u(0) = u_0 \textit{ and for all } t \in (0,T) \textit{ it holds} \\ &(d_t u(t), v) + (Au(t), v) \quad = \quad (f(t), v) \quad \forall v \in L\,. \end{aligned} \tag{2}$$

where we have assumed that $f(t) \in L$ for all $t \in (0,T)$. Note that the operator $A : W \to L$ is an isomorphism if Ω is a convex domain and the coefficients b and c are sufficiently smooth. The operator A can be extended in the usual way to an operator $\tilde{A} : H^1_0(\Omega) \to H^{-1}(\Omega)$ and the regularity of the initial value u_0 can be weakened to $u_0 \in L$.

3 dG(1) time discretization

We decompose the global time interval $I := [0,T]$ into disjoint time intervals I_n by means of discrete time levels t_n and time steps τ_n such that

$$0 = t_0 < t_1 < \ldots < t_N = T, \qquad I_n := (t_{n-1}, t_n], \qquad \tau_n := t_n - t_{n-1}.$$

Then, we approximate the exact solution $u(t)$ by means of a time-discontinuous piecewise linear function $u_\tau(t)$, i.e.,

$$u_\tau \in X_\tau := \{u \in L^2(I,W) : \; u|_{I_n} \in \mathbb{P}_1(I_n, W) \quad \forall n = 1, \ldots, N\}$$

where

$$\mathbb{P}_1(I_n, W) := \{u : I_n \to W : \exists\, U_0, U_1 \in W \quad \text{such that} \quad u(t) = U_0 + U_1 t \quad \text{for all } t \in I_n\}.$$

For the description of the dG(1)-method, we have to define the jump of the discrete solution u_τ at time level t_n by means of

$$[u]_n := u_n^+ - u_n^-, \qquad \text{where} \qquad u_n^\pm := \lim_{\varepsilon \to 0^\pm} u_\tau(t_n + \varepsilon).$$

Now, the exact dG(1)-method can be formulated as:

$$\begin{aligned}
&\textit{Find } u_\tau \in X_\tau \textit{ such that } u_\tau(0) = u_0^- = u_0 \textit{ and} \\
&\sum_{n=1}^{N} \int_{I_n} (d_t u_\tau(t), v_\tau(t))\; dt + \sum_{n=1}^{N} \left([u_\tau]_{n-1}, v_{n-1}^+\right) + \int_0^T (Au_\tau(t), v_\tau(t))\, dt \\
&= \int_0^T (f(t), v_\tau(t))\, dt \qquad\qquad \forall v_\tau \in Y_\tau,
\end{aligned} \tag{3}$$

where

$$Y_\tau := \{v \in L^2(I, L) :\; v|_{I_n} \in \mathbb{P}_1(I_n, L) \quad \forall n = 1, \ldots, N\}.$$

This problem can be solved by a time marching process where successively on each time interval I_n a pde-system in space for two functions $U_n^1, U_n^2 \in W$ has to be solved which determine the time-discrete solution $u_\tau|_{I_n} \in \mathbb{P}_1(I_n, W)$ by means of the ansatz

$$u_\tau|_{I_n}(t) = U_n^1 \phi_{n,1}(t) + U_n^2 \phi_{n,2}(t).$$

Here, $\phi_{n,j} : I_n \to \mathbb{R}$, $j = 1, 2$, are the two scalar basis functions of $\mathbb{P}_1(I_n, \mathbb{R})$ defined below. For the test function v_τ in (3), we take $v_\tau(t) = v\phi_{n,i}(t)$ where $v \in L$ is arbitrarily chosen and the basis function $\phi_{n,i}(t)$ is defined to be zero outside of the interval I_n. Therefore, only contributions from the interval I_n remain in (3). Furthermore, we apply the 2-point Gauß-Radau quadrature rule as numerical integration. This formula is exact if the function to be integrated is a time polynomial of degree less or equal to two. Thus, all integrals in (3) are approximated exactly except for the time integral containing $f(t)$. The two Gauß-Radau quadrature points on interval $I_n = (t_{n-1}, t_n]$ are $t_{n,1} = t_{n-1} + \tau_n/3$ and $t_{n,2} = t_n$ and the weights on the reference interval $(-1, 1]$ are $\hat{w}_1 = 3/2$ and $\hat{w}_2 = 1/2$. The basis functions $\phi_{n,j} \in \mathbb{P}_1(I_n, \mathbb{R})$ are defined by the conditions $\phi_{n,j}(t_{n,i}) = \delta_{i,j}$ where $\delta_{i,j}$ denotes the Kronecker delta. With all these choices we get from (3) the following *I_n-problem of the dG(1)-method*:

$$\begin{aligned}
&\textit{For given } U_n^0 := u_\tau(t_{n-1}), \textit{ find } U_n^1, U_n^2 \in W \textit{ such that} \\
&\{\tfrac{9}{8}U_n^1 + \tfrac{3\tau_n}{4} A U_n^1\} \;+\; \tfrac{3}{8}U_n^2 \;=\; \tfrac{3}{2}U_n^0 + \tfrac{3\tau_n}{4} f(t_{n,1}), \\
&-\tfrac{9}{8}U_n^1 \;+\; \{\tfrac{5}{8}U_n^2 + \tfrac{\tau_n}{4} A U_n^2\} \;=\; -\tfrac{1}{2}U_n^0 + \tfrac{\tau_n}{4} f(t_n).
\end{aligned} \tag{4}$$

Due to the choice of the basis functions $\phi_{n,j}$, the meaning of the unknown $U_n^j \in W$, $j = 1, 2$, is given by $U_n^j = u_\tau(t_{n,j})$.

4 Space discretization

Now we want to discretize the time-independent pde-system (4) in space by means of a finite element method. For the weak formulation of (4), we introduce the solution space $V := H_0^1(\Omega)$, the bilinear form $a : V \times V \to \mathbb{R}$ defined by

$$a(u, v) := \varepsilon(\nabla u, \nabla v) + (b \cdot \nabla u, v) + (cu, v), \qquad \forall\, u, v \in V$$

and the linear forms $\ell_i^n(U_n^0;\cdot)$ by

$$\ell_1^n(U_n^0;v) := \frac{3}{2}\left(U_n^0,v\right) + \frac{3\tau_n}{4}\left(f(t_{n,1}),v\right), \qquad \ell_2^n(U_n^0;v) := -\frac{1}{2}\left(U_n^0,v\right) + \frac{\tau_n}{4}\left(f(t_n),v\right).$$

Then, the weak formulation of (4) reads:

Find $U_n^1, U_n^2 \in V$ such that for all $v \in V$ it holds:

$$\begin{array}{lcll} \left\{\frac{9}{8}\left(U_n^1,v\right) + \frac{3\tau_n}{4}a(U_n^1,v)\right\} & + & \frac{3}{8}\left(U_n^2,v\right) & = \ \ell_1^n(U_n^0;v), \\ -\frac{9}{8}\left(U_n^1,v\right) & + & \left\{\frac{5}{8}\left(U_n^2,v\right) + \frac{\tau_n}{4}a(U_n^2,v)\right\} & = \ \ell_2^n(U_n^0;v). \end{array} \tag{5}$$

In the transport dominated case, we need a stable discretization of the operator A, i.e., we will replace in the discrete problem associated with (5) the bilinear form $a(\cdot,\cdot)$ by a stabilized bilinear form $a_h(\cdot,\cdot)$ defined as

$$a_h(U_{n,h}^0;u_h,v_h) := a(u_h,v_h) + a_{\mathrm{LPS}}\left(u_h,v_h\right) + a_{\mathrm{SC}}\left(U_{n,h}^0;u_h,v_h\right), \tag{6}$$

where $U_{n,h}^0 = u_{\tau,h}(t_{n-1})$ is the space-discrete approximation of $U_n^0 = u_\tau(t_{n-1})$, $a_{\mathrm{LPS}}(\cdot,\cdot)$ an additive bilinear form defined like in [17] by the idea of local projection on the basis of an enriched finite element space and $a_{\mathrm{SC}}(\cdot;\cdot,\cdot)$ an additive shock-capturing term similarly to [7, 12] but with a new scaling factor based on adaptive post-processing applied to $U_{n,h}^0$.

In the following sections, we will describe in detail the definition of a_{LPS} and a_{SC}.

4.1 LPS with enriched bubble functions

Let $\mathcal{T}_h$ be a shape-regular, admissible decomposition of Ω into intervals for $d=1$, quadrilaterals for $d=2$ or hexahedrons for $d=3$. The diameter of a cell $K \in \mathcal{T}_h$ is denoted by h_K and the global mesh size is defined as $h := \max\{h_K : K \in \mathcal{T}_h\}$. Let $\widehat{K} := (-1,1)^d$ denote the reference element, $F_K : \widehat{K} \to K$ the standard reference mapping and $\mathbb{Q}_r(\widehat{K})$ the space of all polynomials on $\widehat{K}$ with maximal degree $r \geq 1$ in each coordinate direction. Furthermore, we need an *enrichment space* $\mathbb{E}_r(\widehat{K})$ on the reference element defined by some basis functions as

$$\mathbb{E}_r(\widehat{K}) := \mathrm{span}\{\hat{\phi}_j^E : j = 1,\ldots,\hat{n}_E\}.$$

Then, the finite element function space $V_r(\widehat{K})$ on the reference element $\widehat{K}$ is defined as

$$V_r(\widehat{K}) := \mathbb{Q}_r(\widehat{K}) \oplus \mathbb{E}_r(\widehat{K}) \tag{7}$$

and the global finite element space V_h by

$$V_h := \{v_h \in H_0^1(\Omega) : v_h|_K \circ F_K \in V_r(\widehat{K}) \ \ \forall K \in \mathcal{T}_h\}. \tag{8}$$

The idea of the Local Projection Stabilization (LPS) method is to compute the projection of the gradient of finite element functions $v_h \in V_h$ into a discontinuous projection space of the *large scale modes* and to stabilize only the remaining so-called *fine scale modes* of the function

v_h for which the gradient cannot be represented by the projection space. As the projection space D_h we will use the discontinuous finite element space

$$D_h := \{v_h \in L^2(\Omega) : \ v_h|_K \in D_r(K) \ \ \forall K \in \mathcal{T}_h\},$$

where

$$D_r(K) := \{v \in L^2(K) : \ v \circ F_K \in \mathbb{Q}_{r-1}(\widehat{K})\}. \tag{9}$$

Let $\pi_K : L^2(K) \to D_r(K)$ denote the L^2-projection into the local projection space. Then, the so-called *fluctuation operator* $\kappa_h : L^2(\Omega) \to D_h$ which assigns to each local part $\phi|_K$ of a global function $\phi \in L^2(\Omega)$ its small scale part $(\kappa_h\phi)|_K$ is defined as

$$\kappa_h(\phi)|_K := \phi|_K - \pi_K\phi|_K \qquad \forall\ K \in \mathcal{T}_h.$$

The stabilizing additive LPS term in the discrete bilinear form is now chosen as the following sum of artificial diffusion parts applied only to the small scale modes of the finite element functions

$$a_{\mathrm{LPS}}(u_h, v_h) := \sum_{K\in\mathcal{T}_h} \gamma_0 h_K \sum_{j=1}^{d} \left(\kappa_h\left(\frac{\partial u_h}{\partial x_j}\right), \kappa_h\left(\frac{\partial v_h}{\partial x_j}\right)\right)_K, \tag{10}$$

where $\gamma_0 > 0$ is a user-defined parameter. By means of several numerical tests we found that the value $\gamma_0 = 0.1$ is a relatively good choice. We note that our stabilization term does not depend on the streamline derivative $b \cdot \nabla$ of the finite element functions like in [17]. In the case when the convection field b depends on the solution u, this is important to simplify the nonlinear iteration.

4.2 Construction of an L^2 orthogonal basis

The freedom that we have for the construction of the above described LPS method is the choice of the enrichment space $\mathbb{E}_r(\widehat{K})$. A theoretical condition to make the method work is (see [17])

$$\exists \beta_1 > 0 : \qquad \inf_{q_h\in D_r(K)} \sup_{v_h\in B_r(K)} \frac{(v_h, q_h)_K}{\|v_h\|_{0,K}\,\|q_h\|_{0,K}} \geq \beta_1 \qquad \forall\ K \in \mathcal{T}_h, \tag{11}$$

where β_1 is an h-independent constant and $B_r(K)$ is the local space of the element bubble functions defined by

$$B_r(K) := \{v \in L^2(K) : \ v \circ F_K \in V_r(\widehat{K}) \cap H_0^1(\widehat{K})\}. \tag{12}$$

In the following, we will choose the enrichment space $\mathbb{E}_r(\widehat{K})$ in such a way that the construction of a locally supported L^2-orthogonal basis is possible. The advantage of an L^2-orthogonal basis is that the arising mass matrix M with the entries $M_{i,j} := (\varphi_j, \varphi_i)$ is diagonal which can be exploited to create a simpler solver of the coupled space-discrete system on each time interval (see Sect. 4.4).

4.2.1 The 1D case

In this case, the reference mapping $F_K : \widehat{K} \to K$ is element-wise affine such that for any two basis functions on an element K, say $\varphi_i = \hat{\varphi}_i \circ F_K^{-1}$ and $\varphi_j = \hat{\varphi}_j \circ F_K^{-1}$, it holds

$$(\varphi_j, \varphi_i)_K = 0 \qquad \text{if and only if} \qquad (\hat{\varphi}_j, \hat{\varphi}_i)_{\widehat{K}} = 0. \tag{13}$$

Thus, it is sufficient to construct an $L^2(\widehat{K})$-orthogonal basis of the reference space $V_r(\widehat{K})$ defined in (7). In the following, we will present the idea of the construction for the order $r = 2$.

To this end, let us denote by $\tilde{\varphi}_i : \widehat{K} \to \mathbb{R}$, $i = 1, 2, 3$, the hierarchical basis functions of the space $\mathbb{Q}_2(\widehat{K})$ given by

$$\tilde{\varphi}_1(x) = \frac{1}{2}(1 - x)\,, \qquad \tilde{\varphi}_2(x) = \frac{1}{2}(1 + x)\,, \qquad \tilde{\varphi}_3(x) = 1 - x^2 \qquad \forall\, x \in \widehat{K}.$$

Here $\tilde{\varphi}_3$ is the only bubble function. The idea is now to choose the enrichment space as

$$\mathbb{E}_2(\widehat{K}) := \operatorname{span}\{\tilde{\varphi}_4\} \subset H_0^1(\widehat{K}), \qquad \text{where} \qquad \tilde{\varphi}_4(x) = (1 - x^2)x(1 + cx^2). \tag{14}$$

For each real parameter c, the odd function $\tilde{\varphi}_4$ is already $L^2(\widehat{K})$-orthogonal to the even bubble function $\tilde{\varphi}_3$. The first step is now to modify the standard hat functions $\tilde{\varphi}_1$, $\tilde{\varphi}_2$ into new functions $\hat{\varphi}_1$, $\hat{\varphi}_2$ such that they are orthogonal to the bubble functions $\tilde{\varphi}_3$ and $\tilde{\varphi}_4$. This can be simply achieved by means of a linear ansatz

$$\hat{\varphi}_i(x) = \tilde{\varphi}_i(x) + \alpha_{i,3}\,\tilde{\varphi}_3(x) + \alpha_{i,4}\,\tilde{\varphi}_4(x)\,, \quad i = 1, 2, \tag{15}$$

with the c-dependent coefficients

$$\alpha_{i,j} = -\frac{(\tilde{\varphi}_i, \tilde{\varphi}_j)_{\widehat{K}}}{(\tilde{\varphi}_j, \tilde{\varphi}_j)_{\widehat{K}}} \qquad i = 1, 2, \;\; j = 3, 4.$$

Finally, we determine the coefficient c by inserting these c-dependent $\alpha_{i,j}$ in the last remaining orthogonality condition

$$(\hat{\varphi}_1, \hat{\varphi}_2)_{\widehat{K}} = 0\,.$$

This is a nonlinear equation in c which can be solved exactly. We get

$$c = \frac{5\sqrt{385} - 154}{61} \approx -0.916277305709132\ldots \tag{16}$$

and

$$\alpha_{1,3} = \alpha_{2,3} = -\frac{5}{8}\,, \qquad \alpha_{1,4} = -\alpha_{2,4} = \frac{427 - 183\sqrt{385}}{-4016 + 48\sqrt{385}} \approx 1.02912891579759\ldots\,. \tag{17}$$

Then, the resulting reference space $V_2(\widehat{K})$ of our new $\mathbb{Q}_2^+$ element with its $L^2(\widehat{K})$-orthogonal basis functions $\hat{\varphi}_i$, $i = 1, \ldots, 4$ is given by

$$V_2(\widehat{K}) := \operatorname{span}\{\hat{\varphi}_1\,, \hat{\varphi}_2\,, \hat{\varphi}_3 := \tilde{\varphi}_3\,, \hat{\varphi}_4 := \tilde{\varphi}_4\}.$$

The above described idea for the construction of $L^2(\widehat{K})$-orthogonal basis functions can be applied in principle to an arbitrary order $r \geq 1$ leading to a new $\mathbb{Q}_r^+$-element. In Table 1, we present the set of the orthogonal basis functions for the cases $r = 1, 2, 3$. For the general case $r \geq 4$, there are still some questions open like, e.g., does the nonlinear equation for c admit a solution and how to get it or can we always achieve $\mathbb{Q}_r^+ \subset \mathbb{Q}_{r+2}$? However, these questions are beyond the scope of this paper and subject of a forthcoming paper.

Table 1: The new enriched $\mathbb{Q}_r(\widehat{K})$-LPS-elements, $r = 1, 2, 3$, with their $L^2(\widehat{K})$-orthogonal basis functions

	$\mathbb{Q}_1^+(\widehat{K}) \subset \mathbb{Q}_3(\widehat{K})$	
$\hat{\varphi}_1(x)$	$=$	$\frac{1}{2}(1-x) + \left(\frac{\sqrt{105}}{320}\sqrt{15} - \frac{9}{64}\sqrt{15}\right)\frac{\sqrt{7}}{84}(35x + 3\sqrt{15}\sqrt{7})(1-x^2)$
$\hat{\varphi}_2(x)$	$=$	$\frac{1}{2}(1+x) + \left(-\frac{\sqrt{105}}{320}\sqrt{15} - \frac{9}{64}\sqrt{15}\right)\frac{\sqrt{7}}{84}(35x + 3\sqrt{15}\sqrt{7})(1-x^2)$
$\hat{\varphi}_3(x)$	$=$	$\frac{\sqrt{7}}{84}(35x + 3\sqrt{15}\sqrt{7})(1-x^2)$
	$\mathbb{Q}_2^+(\widehat{K}) \subset \mathbb{Q}_5(\widehat{K})$	
$\hat{\varphi}_1(x)$	$=$	$\frac{1}{2}(1-x) - \frac{5}{8}(1-x^2) + \frac{427-183\sqrt{385}}{-4016+48\sqrt{385}}x(1 - \frac{154-5\sqrt{385}}{61}x^2)(1-x^2)$
$\hat{\varphi}_2(x)$	$=$	$\frac{1}{2}(1-x) - \frac{5}{8}(1-x^2) - \frac{427-183\sqrt{385}}{-4016+48\sqrt{385}}x(1 - \frac{154-5\sqrt{385}}{61}x^2)(1-x^2)$
$\hat{\varphi}_3(x)$	$=$	$1-x^2$
$\hat{\varphi}_4(x)$	$=$	$x(1 - \frac{154-5\sqrt{385}}{61}x^2)(1-x^2)$
	$\mathbb{Q}_3^+(\widehat{K}) \subset \mathbb{Q}_5(\widehat{K})$	
$\hat{\varphi}_1(x)$	$=$	$\frac{1}{2}(1-x) - \frac{5}{8}(1-x^2) + \frac{7}{8}x(1-x^2)$ $+\frac{-525+35\sqrt{33}}{512}\frac{\sqrt{33}}{1155}(231x^3 + 35\sqrt{33}x^2 - 77x - 5\sqrt{33})(1-x^2)$
$\hat{\varphi}_2(x)$	$=$	$\frac{1}{2}(1+x) - \frac{5}{8}(1-x^2) - \frac{7}{8}x(1-x^2)$ $-\frac{525+35\sqrt{33}}{512}\frac{\sqrt{33}}{1155}(231x^3 + 35\sqrt{33}x^2 - 77x - 5\sqrt{33})(1-x^2)$
$\hat{\varphi}_3(x)$	$=$	$1-x^2$
$\hat{\varphi}_4(x)$	$=$	$x(1-x^2)$
$\hat{\varphi}_5(x)$	$=$	$\frac{\sqrt{33}}{1155}(231x^3 + 35\sqrt{33}x^2 - 77x - 5\sqrt{33})(1-x^2)$

It remains to show that our new $\mathbb{Q}_r^+$-elements satisfy the theoretical condition (11) which ensures that the LPS-method works.

Lemma 1 *The local inf-sup condition* (11) *is satisfied for the orders* $r = 1, 2, 3$.

Proof. At first one can show by standard arguments (transformation of norms between original and reference element) that condition (11) is equivalent to an analogous condition where all objects are defined on the reference element $\widehat{K} = [-1, 1]$. Using the equivalence of norms on finite dimensional spaces, this analogous condition is satisfied if the matrix B with the entries $B_{i,j} := (\hat{\varphi}_i^D, \hat{\varphi}_j^B)_{\widehat{K}}$ has positive singular values where $\{\hat{\varphi}_i^D\}$ denote the basis functions of $D_r(\widehat{K}) = \mathbb{Q}_{r-1}(\widehat{K})$ and $\{\hat{\varphi}_j^B\}$ the basis functions of $B_r(\widehat{K}) = V_r(\widehat{K}) \cap H_0^1(\widehat{K})$.

The singular values of this matrix can be computed exactly by means of a computer algebra program. In this way we can show for $r = 1, 2, 3$, that our $\mathbb{Q}_r^+$-element satisfies the condition (11). In the case $r = 2$, we can calculate matrix B simply by hand. Taking into account that $\varphi_1^B = \tilde{\varphi}_3$ and $\varphi_2^B = \tilde{\varphi}_4$ are even and odd functions, respectively, we easily calculate the matrix $B = \begin{pmatrix} \frac{4}{3} & 0 \\ 0 & \frac{4\sqrt{385}}{427} - \frac{4}{183} \end{pmatrix}$ which has obviously positive singular values. □

4.2.2 The 2D and 3D case

If we assume that we have an affine reference mapping $F_K : \widehat{K} \to K$ then the local $L^2(K)$-orthogonality is equivalent to the $L^2(\widehat{K})$-orthogonality, i.e. the equivalence (13) holds also in the 2D and 3D case. For the reference element $\widehat{K} = [-1, 1]^d$, $d = 2, 3$, the reference mapping F_K is affine if and only if the mesh cell K is a parallelogram or a parallelepiped. Therefore, we can generalize our 1D orthogonal basis functions easily to the 2D and 3D case if the mesh consists only of parallelograms or parallelepipeds. The idea is to define each reference basis function on $\widehat{K}$ in the d-dimensional case as a product of 1D reference basis functions in the single coordinates. In particular in the 2D case, the degrees of freedom on $\widehat{K} = [-1, 1]^2$ form a lattice consisting of $(r+2) \times (r+2)$ nodes since the one-dimensional $\mathbb{Q}_r^+$-element has $(r+2)$ nodes on the reference element $[-1, 1]$. Then, a node j from this lattice, which corresponds to a local number j_1 in the direction $\hat{x}_1$ and to a local number j_2 in the direction $\hat{x}_2$ with $j_1, j_2 \in \{1, \ldots, r+2\}$, is related to the 2D basis function

$$\hat{\varphi}_j^{2D}(\hat{x}_1, \hat{x}_2) := \hat{\varphi}_{j_1}^{1D}(\hat{x}_1)\hat{\varphi}_{j_2}^{1D}(\hat{x}_2),$$

where $\hat{\varphi}_{j_k}^{1D}(\cdot)$ denotes the reference basis function of the one-dimensional $\mathbb{Q}_r^+$-element associated with the local node number j_k. The drawbacks in the multi-dimensional case are the restriction to special mesh geometries and the fact that the number of enriched basis functions is relatively high compared to the number of basis functions for the original $\mathbb{Q}_r$-element.

4.3 Local shock capturing based on low order post-processing

We follow the idea of Guermond et al. [7, 12] and add artificial isotropic diffusion with a special scaling factor function $\gamma_K(\cdot)$, i.e., we use the additive bilinear form $a_{\mathrm{SC}}(\cdot, \cdot)$ defined by

$$a_{\mathrm{SC}}\left(u_h^0; u_h, v_h\right) := \sum_{K \in \mathcal{T}_h^*(u_h^0)} \gamma_K(u_h^0)\,(\nabla u_h, \nabla v_h)_K, \tag{18}$$

where $u_h^0 \in V_h$ is an approximation of the discrete solution containing the information about the intensity and position of the shocks. $\mathcal{T}_h^*(u_h^0)$ is the subset of the "bad cells" $K \in \mathcal{T}_h$, for which a local error indicator $\eta_K(u_h^0)$ is relatively large, and $\gamma_K(u_h^0)$ is the following scaling factor for the artificial diffusion on mesh cell $K \in \mathcal{T}_h$

$$\gamma_K(u_h^0) := \gamma_{\mathrm{SC}} h_K \eta_K(u_h^0). \tag{19}$$

The parameter $\gamma_{\rm SC}$ is a positive user-defined constant. The differences of our definition of $a_{\rm SC}$ compared to Guermond et al. are the special scaling factor $\gamma_K(u_h^0)$ and the introduction of the set of the "bad cells" $\mathcal{T}_h^*(u_h^0) \subset \mathcal{T}_h$.

In the following, we will explain, for an arbitrary order $r \geq 1$, the computation of the error indicators $\eta_K(u_h^0)$ for all cells $K \in \mathcal{T}_h$. Let $\pi_h^0 : V_h \to \mathbb{Q}_0^{dc}(\mathcal{T}_h)$ denote the operator of the $L^2(\Omega)$-projection into the discontinuous space $\mathbb{Q}_0^{dc}(\mathcal{T}_h)$ of piecewise constant functions. The computation of the projection $\bar{u}_h^0 := \pi_h^0 u_h^0$ for a given $u_h^0 \in V_h$ can be done element-wise with very low computational costs. Then, we apply to $\bar{u}_h^0$ a transfer (recovering) operator $R_h : \mathbb{Q}_0^{dc}(\mathcal{T}_h) \to \mathbb{Q}_1(\mathcal{T}_h)$ which maps into the continuous multi-linear finite element space $\mathbb{Q}_1(\mathcal{T}_h)$. This operator is known from [19] and is a special case of a general transfer operator from one finite element space into another [21]. For an interior vertex node a_j of the mesh $\mathcal{T}_h$, let $\mathcal{T}_h^j$ denote the set of those cells $K \in \mathcal{T}_h$ that have a_j as one of their vertices and n_h^j the number of cells in $\mathcal{T}_h^j$. Then, for $v_h \in \mathbb{Q}_0^{dc}(\mathcal{T}_h)$, the value of $R_h v_h$ at an interior node a_j is defined as

$$R_h v_h(a_j) := \frac{1}{n_h^j} \sum_{K \in \mathcal{T}_h^j} v_h|_K(a_j)$$

and at a boundary node as zero. The computation of $R_h v_h$ can be implemented efficiently with very low costs [21]. Now, our error indicator applied to a given function $u_h^0 \in V_h$ on a cell $K \in \mathcal{T}_h$ is defined as

$$\eta_K(u_h^0) := \| u_h^0 - R_h \left(\pi_h^0 u_h^0 \right) \|_{L^\infty(K)}, \tag{20}$$

where the exact $L^\infty(K)$-norm is replaced in practice by a discrete one defined as the maximum of absolute values over a set of discrete points. For the determination of the "bad-cell" subset $\mathcal{T}_h^*(u_h^0)$, we need the maximum value of the indicators $\eta_{\max}(u_h^0) := \max_{K \in \mathcal{T}_h} \eta_K(u_h^0)$ and define with a user-chosen threshold $s_0 \in (0,1)$ the subset

$$\mathcal{T}_h^*(u_h^0) := \{ K \in \mathcal{T}_h : \ \eta_K(u_h^0) \geq s_0 \eta_{\max}(u_h^0) \}.$$

The question is why the above defined error indicator detects the "bad cells" where the discrete function $u_h^0 \in V_h$ is not smooth? Let us assume that the cell K is located in a subdomain Ω_0 where u_h^0 is smooth in the sense that it is very close to the $\mathbb{Q}_1$-interpolant of a smooth function u. Then, one can show that $\eta_K(u_h^0) \leq Ch^2$ which is caused by superconvergence properties of the reconstruction operator R_h. On the other hand, if K belongs to a patch where u_h^0 is oscillating, the operator R_h cannot improve the low order approximation $\pi_h^0 u_h^0$ to a second order one such that $\eta_K(u_h^0) \geq Ch$. This heuristic argumentation shows that our indicator should be able to separate the "bad cells" from the good ones.

4.4 The space-discrete problem for one time interval

Now, having defined the bilinear forms $a_{\rm LPS}$ and $a_{\rm SC}$ for our stabilized bilinear form a_h given in (6), we come back to the discrete solution of the pde-system (5). To this end, we first approximate the linear functionals on the right hand side of (5). We assume that we have an approximation $U_{n,h}^0 \in V_h$ of the initial value $U_n^0 = u_\tau(t_{n-1}) \in V$ on the time interval

$I_n = (t_{n-1}, t_n]$. At the beginning, when $n = 1$, the approximation $U^0_{n,h}$ is just a projection of the given initial value $u_0 \in V$ into the finite element space V_h. For $n \geq 2$, $U^0_{n,h}$ is just the value from the previous time interval, i.e. $U^0_{n,h} = U^2_{n-1,h}$. Then, for $i = 1, 2$, we replace in (5) the exact linear functional $\ell^n_i(U^0_n; \cdot)$ on the right hand side by the functional $\ell^n_i(U^0_{n,h}; \cdot)$ and the exact bilinear form $a(\cdot, \cdot)$ by the stabilized discrete form $a_h(U^0_{n,h}; \cdot, \cdot)$ and get the following discrete problem:

Find $U^1_{n,h}, U^2_{n,h} \in V_h$ *such that for all* $v_h \in V_h$ *it holds:*

$$\begin{aligned} \left\{ \tfrac{9}{8} \left(U^1_{n,h}, v_h \right) + \tfrac{3\tau_n}{4} a_h(U^0_{n,h}; U^1_{n,h}, v_h) \right\} \quad + \quad \tfrac{3}{8} \left(U^2_{n,h}, v_h \right) \quad &= \quad \ell^n_1(U^0_{n,h}; v_h), \\ -\tfrac{9}{8} \left(U^1_{n,h}, v_h \right) \quad + \quad \left\{ \tfrac{5}{8} \left(U^2_{n,h}, v_h \right) + \tfrac{\tau_n}{4} a_h(U^0_{n,h}; U^2_{n,h}, v_h) \right\} \quad &= \quad \ell^n_2(U^0_{n,h}; v_h). \end{aligned} \tag{21}$$

Note that this is a linear problem although we apply shock capturing. However, this "linear version" of shock capturing only works if the shock position does not move too fast in relation to the mesh size in the shock region. Having solved this linear coupled system for $U^1_{n,h}, U^2_{n,h}$, we get the fully discrete solution $u_{\tau,h} \in \mathbb{P}_1(I_n, V_h)$ on the time interval I_n by

$$u_{\tau,h}(t) = U^1_{n,h} \phi_{n,1}(t) + U^2_{n,h} \phi_{n,2}(t) \qquad \forall\, t \in (t_{n-1}, t_n].$$

Due to the basis functions $\phi_{n,j}$, we have that $U^j_{n,h} = u_{\tau,h}(t_{n,j})$, $j = 1, 2$, where $t_{n,j}$ denote the Gauss-Radau points on $(t_{n-1}, t_n]$.

The discrete problem (21) is equivalent to the following 2×2 block-system for the nodal vectors $\underline{U}^j \in \mathbb{R}^{n_h}$, $j = 1, 2$, associated with the finite element functions $U^j_{n,h} \in V_h$:

$$\begin{bmatrix} \frac{9}{8} M + \frac{3\tau_n}{4} A & \frac{3}{8} M \\ -\frac{9}{8} M & \frac{5}{8} M + \frac{\tau_n}{4} A \end{bmatrix} \begin{bmatrix} \underline{U}^1 \\ \underline{U}^2 \end{bmatrix} = \begin{bmatrix} \underline{F}^1 \\ \underline{F}^2 \end{bmatrix}, \tag{22}$$

where $M \in \mathbb{R}^{n_h \times n_h}$ with the entries $M_{i,j} = (\varphi_j, \varphi_i)$ denotes the mass matrix, $A \in \mathbb{R}^{n_h \times n_h}$ with the entries $A_{i,j} = a_h(U^0_{n,h}; \varphi_j, \varphi_i)$ the stiffness matrix and $\underline{F}^j \in \mathbb{R}^{n_h}$ with the components $\underline{F}^j_i = \ell^n_j(U^0_{n,h}; \varphi_i)$ the right hand side vector.

In the following, we will explain how we can exploit the fact that we have a diagonal mass matrix M due to the L^2-orthogonality of the basis functions of our new $\mathbb{Q}^+_r$-element. We introduce the matrix $\tilde{A} := M^{-1} A$ which is just a row scaling of A if M is diagonal. Moreover, we define $R^j := M^{-1} \underline{F}^j$ for $j = 1, 2$. Then, multiplying the system (22) by M^{-1} we get

$$\underline{U}^1 = -\frac{8}{9} R^2 + \frac{1}{9} \left(5I + 2\tau_n \tilde{A} \right) \underline{U}^2$$

from the second equation and by inserting this into the first equation it follows

$$\left\{ I + \frac{2\tau_n}{3} \tilde{A} + \frac{\tau_n^2}{6} \tilde{A}^2 \right\} \underline{U}^2 = R^1 + R^2 + \frac{2\tau_n}{3} \tilde{A} R^2. \tag{23}$$

The remaining system (23) for $\underline{U}^2$ has only the half dimension of the original 2×2 block-system and can be solved iteratively in an efficient way.

5 Numerical results

5.1 Problem with a moving interior layer

We restrict our numerical tests to the one dimensional spatial case with the domain $\Omega = (-0.5, 2.5)$ which is partitioned by an equidistant grid with the mesh size $h_\ell = \frac{3}{80}\, 2^{1-\ell}$ where $\ell \in \{1, \ldots, 5\}$ denotes the refinement level. We investigate the performance of our $\mathbb{Q}_2^+$-element with an expected optimal convergence order of three in the L^2-norm. From the dG(1) time discretization we expect at the endpoints of the time intervals also order three in the L^2-norm. Therefore, we couple for each computation the uniform time step size τ_ℓ linearly to the actual mesh size h_ℓ, i.e., we define for refinement level ℓ the associated time step size as $\tau_\ell = 0.5 h_\ell$. As a test problem we consider problem (1) with the data $T = 0.8$, $b = 1$, $c = 0$, $f = 0$ and homogeneous Dirichlet boundary conditions. The difficulty in this problem comes from a non-smooth initial solution u_0 which already has a shock (discontinuity), i.e.,

$$u_0(x) = \begin{cases} \sin(\pi x/2)\,, & \text{for} \quad x \in [0,1], \\ 1.5\,, & \text{for} \quad x \in (1, 1.2), \\ 0\,, & \text{for} \quad x \in [-0.5, 0] \cup [1.2, 2.5]\,. \end{cases}$$

For the LPS parameter we choose $\gamma_0 = 0.1$ and for the shock capturing parameter $\gamma_{\mathrm{SC}} = 0.5$, respectively. From the structure of the exact solution we know that, for each time $t \in [0, T]$, the solution is arbitrary smooth in the subdomain $\Omega_0(t) = (0.3 + bt,\ 0.7 + bt)$. Therefore, we introduce for the error $e_{\tau,h} = u_{as} - u_{\tau,h}$ on a given mesh level the following local and global error measure over the range $[T_0, T]$ of time steps:

$$\|e_{\tau,h}\|_{\Omega_0} := \max_{T_0 \le t_n \le T} \|e_{\tau,h}(t_n)\|_{L^2(\Omega_0(t_n))}, \qquad \|e_{\tau,h}\|_{\Omega} := \max_{T_0 \le t_n \le T} \|e_{\tau,h}(t_n)\|_{L^2(\Omega)}.$$

Here, u_{as} is a highly accurate asymptotic approximation of the exact solution and $T_0 = 0$ in the case $\varepsilon = 0$ and $T_0 = 0.1$ in the case $\varepsilon = 10^{-4}$. Note that it is important to exclude the Gibbs phenomenon for the projection of the initial solution u_0 into the finite element space V_h. We have solved this problem by reducing the order of interpolation to one in the region of the discontinuity, see Figure 1. We let the shock capturing terms act only on those cells

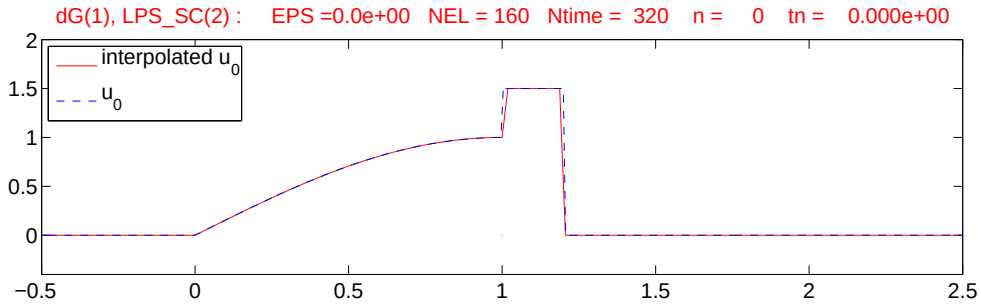

Figure 1: Initial solution u_0 and its projection into V_h on level 2.

K for which the associated error indicator η_K is large in the sense $\eta_K \ge 0.1 \eta_{\max}$ where $\eta_{\max}$

denotes the maximum error indicator over all mesh cells. In the following we study separately the cases $\varepsilon = 0$ and $\varepsilon = 10^{-4}$.

5.1.1 Pure convection ($\varepsilon = 0$)

In this case, the exact solution is given by:

$$u(x,t) = u_{\text{as}}(x,t) = u_0(x - bt)\,.$$

Figures 2 and 3 show for the different discrete time levels $t_n = 0.4$ and $t_n = 0.8 = T$ that the pure LPS-solution ($\gamma_{\text{SC}} = 0$) exhibits oscillations whereas the corresponding LPS-SC-solution with shock capturing ($\gamma_{\text{SC}} = 0.5$) does not.

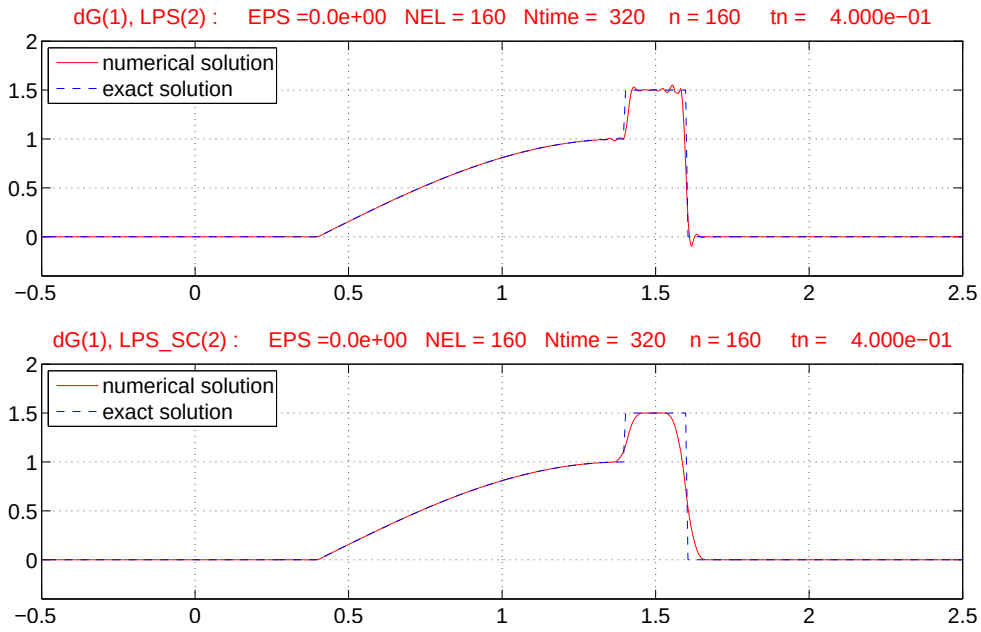

Figure 2: LPS- (top) and LPS-SC-solution (bottom) on level 2 for $\varepsilon = 0$ and $t_n = 0.4$.

Note that the amount of smearing in the LPS-SC-solution is restricted to a few mesh cells only near the position of the shock. The smearing can be decreased by taking a smaller value of γ_{SC}. However, it was not our aim to optimize this parameter.

As a quantitative measure for the intensity of spurious oscillations in a discrete function $u_h \in V_h$ we consider the following broken semi-norm

$$TV(u_h) := \sum_{K \in \mathcal{T}_h} \|\nabla u_h\|_{L^1(K)}. \tag{24}$$

In the one-dimensional case with piecewise linear finite elements, this semi-norm coincides with the standard total variation of the numerical solution, see [23]. The left plot of Figure 4 shows that $TV(u_{\tau,h}(t))$ for the LPS solution grows with the time t whereas for the LPS-SC solution this quantity is close to the exact value 3 of the total variation of the exact solution.

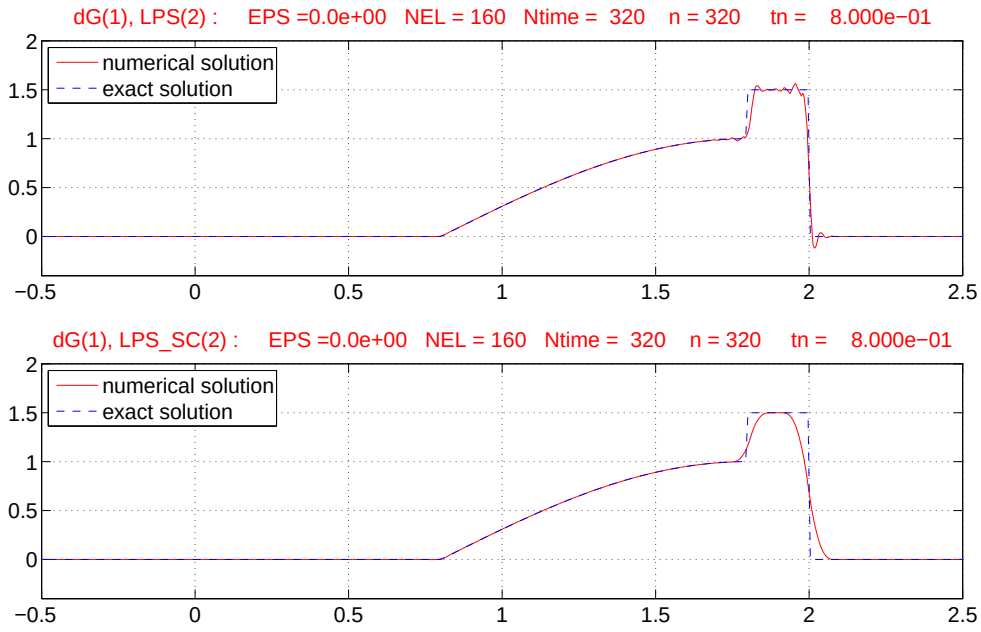

Figure 3: LPS- (top) and LPS-SC-solution (bottom) on level 2 for $\varepsilon = 0$ and $t_n = 0.8$.

In Table 2, we present the discretization errors in the L^2-norm on the global domain Ω and on the local shock-free subdomain $\Omega_0(t)$ for the LPS- and LPS-SC-method. We see that, in the local subdomain $\Omega_0(t)$ where the solution is smooth, both methods have the same optimal order of convergence. In the global domain, the error of the LPS-SC method is a little bit larger but the asymptotic rate of convergence seems to be more robust.

Table 2: L^2-norms of the discretization errors for $\varepsilon = 0$.

	LPS				LPS-SC			
level	$\|e_{\tau,h}\|_{\Omega_0}$	order	$\|e_{\tau,h}\|_{\Omega}$	order	$\|e_{\tau,h}\|_{\Omega_0}$	order	$\|e_{\tau,h}\|_{\Omega}$	order
1	3.498e-05		1.349e-01		2.313e-06		1.735e-01	
2	6.413e-07	5.769	9.005e-02	0.583	9.011e-08	4.682	1.334e-01	0.379
3	1.130e-08	5.826	7.760e-02	0.215	1.130e-08	2.995	1.026e-01	0.380
4	1.429e-09	2.984	5.050e-02	0.620	1.429e-09	2.984	7.893e-02	0.378
5	1.788e-10	2.999	4.029e-02	0.326	1.788e-10	2.999	6.080e-02	0.376

With Figure 5 we demonstrate that our low order post-processing error indicator applied to the pure LPS-solution works well for the detection of those mesh cells where the discrete solution exhibits oscillations. Figure 6 shows that the error indicator also works well as "bad-cell-detector" if it is applied to the LPS-SC-solution with shock-capturing. The height of each balk in the Figures 5 and 6 represents the value of the estimator on the corresponding mesh cell.

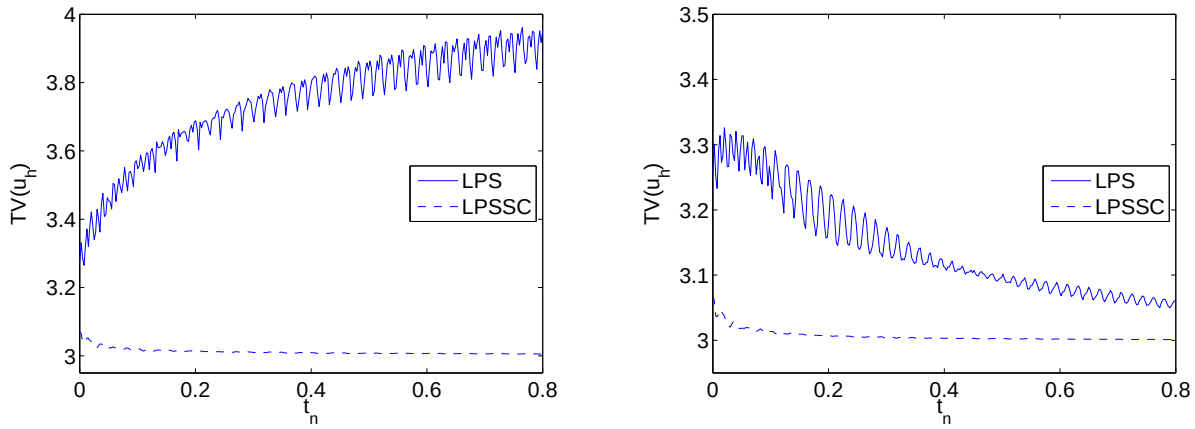

Figure 4: Time evolution of the TV-semi-norm (24) of the discrete LPS- or LPS-SC-solution on level 2 for $\varepsilon = 0$ (left) and $\varepsilon = 10^{-4}$ (right).

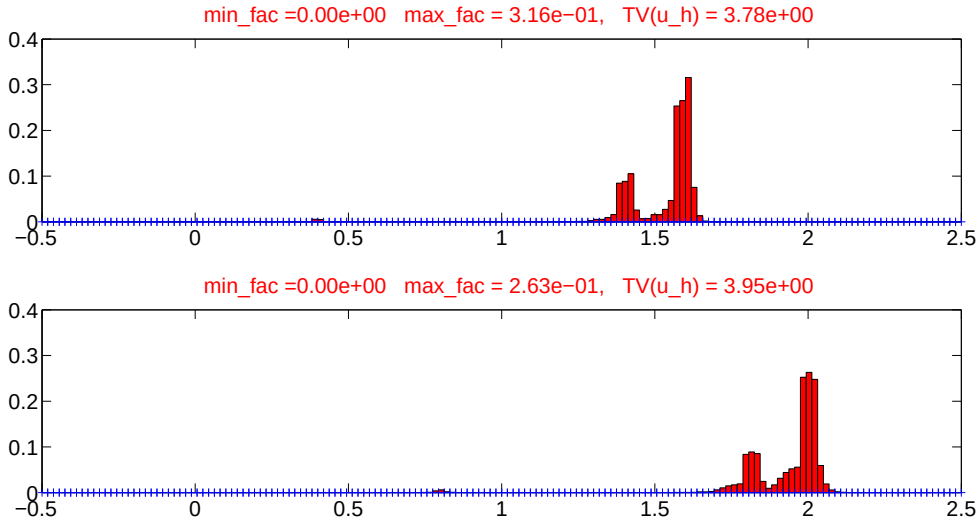

Figure 5: Error indicator for the LPS-solution on level 2 for $\varepsilon = 0$ and $t_n = 0.4$ (top) and $t_n = 0.8$ (bottom).

5.1.2 Convection and diffusion ($\varepsilon = 10^{-4}$)

In this case, the following approximation of the exact solution can be used

$$u(x,t) \approx u_{\text{as}}(x,t) = \frac{1}{\sqrt{4\pi\varepsilon t}} \int_{-0.5}^{2.5} u_0(y) \cdot \exp\Big\{-\frac{(x-t-y)^2}{4\varepsilon t}\Big\}\, dy\,. \tag{25}$$

One can easily verify that u_{as} satisfies the differential equation (1) with our special data. Moreover, it can be shown that $u_{\text{as}}(\cdot,t)$ converges to the initial solution $u_0(\cdot)$ for $t \to 0$. Only the homogeneous boundary conditions are violated by u_{as}. However, using the maximum

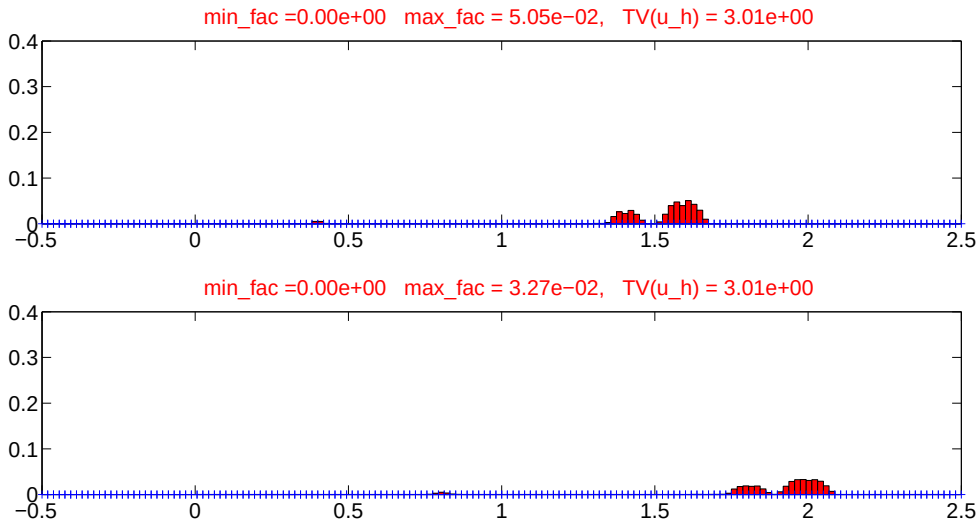

Figure 6: Error indicator for the LPS-SC-solution on level 2 for $\varepsilon = 0$ and $t_n = 0.4$ (top) and $t_n = 0.8$ (bottom).

principle for parabolic equations (see e.g. [8]), we get the estimate

$$\max_{x\in\Omega} |u_{\text{as}}(x,t) - u(x,t)| \le M_\varepsilon(t) := \max\{u_{\text{as}}(-0.5,t),\, u_{\text{as}}(2.5,t)\} \qquad \forall t \in (0,T]. \tag{26}$$

For all error measurements in this section, we have used u_{as} instead of the exact solution. If t is not too small, i.e., for $t \in [T_0, T]$ with $T_0 = 0.1$, the integral in (25) can be computed numerically in an accurate way using the adaptive Gauss-Kronrod quadrature (available in Matlab). By means of these numerical approximations we have found that, for $\varepsilon = 10^{-4}$, all the values $M_\varepsilon(t_n)$ have been zero up to machine accuracy which proves due to (26) the accuracy of u_{as}.

Table 3 shows the discretization errors in the L^2-norm on the global domain Ω and on the local shock-free subdomain $\Omega_0(t)$ for the LPS- and LPS-SC-method. In the local subdomain,

Table 3: L^2-norms of the discretization errors for $\varepsilon = 10^{-4}$.

	LPS				LPS-SC			
level	$\|e_{\tau,h}\|_{\Omega_0}$	order	$\|e_{\tau,h}\|_{\Omega}$	order	$\|e_{\tau,h}\|_{\Omega_0}$	order	$\|e_{\tau,h}\|_{\Omega}$	order
1	9.775e-06		9.170e-02		8.803e-07		1.245e-01	
2	4.706e-08	7.698	4.941e-02	0.892	4.683e-08	4.233	7.662e-02	0.700
3	5.699e-09	3.046	2.668e-02	0.889	5.699e-09	3.038	4.584e-02	0.741
4	6.784e-10	3.071	1.308e-02	1.028	6.784e-10	3.071	2.366e-02	0.954
5	7.687e-11	3.141	6.570e-03	0.994	7.687e-11	3.141	9.834e-03	1.267

both methods behave nearly the same and yield the optimal approximation order of 3. In the global domain, the error of the LPS-SC method is a little bit larger due to smearing effects, but the asymptotic rate of convergence seems to be more robust.

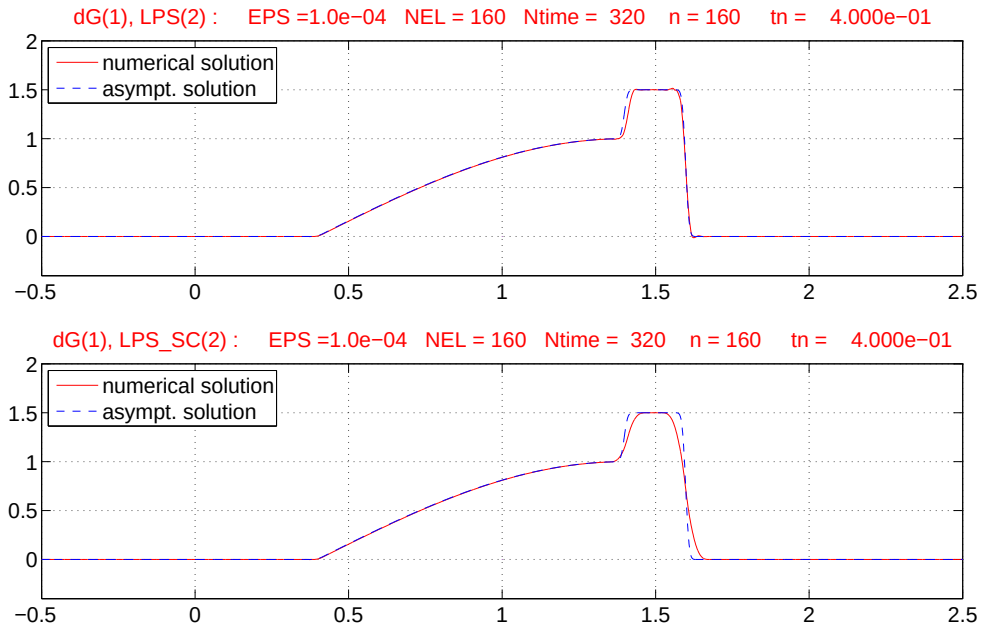

Figure 7: The LPS (top) and LPS-SC (bottom) solutions on level 2 for $\varepsilon = 10^{-4}$ and $t_n = 0.4$.

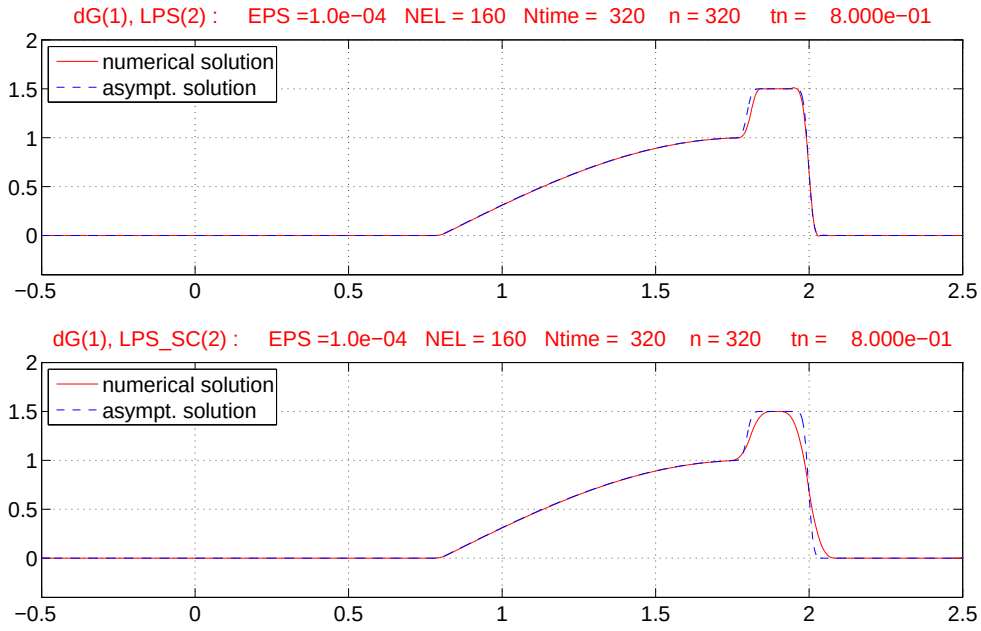

Figure 8: The LPS (top) and LPS-SC (bottom) solutions on level 2 for $\varepsilon = 10^{-4}$ and $t_n = 0.8$.

The right plot in Figure 4 compares the oscillation measure $TV(u_{\tau,h}(t))$ over the time between the pure LPS-method and the LPS-SC-method with shock-capturing. Again we see quantitatively that the additive shock-capturing term in the discretization leads to a much

more smoother discrete solution which cannot be seen so clearly when looking at the graphs of the solutions in Figures 7 and 8. We have omitted the analogous pictures to Figures 5 and 6 from the case $\varepsilon = 0$ about the action of the error indicator since they would tell us the same result that the indicator works well also in the case $\varepsilon = 10^{-4}$.

5.2 Problem with an exponential boundary layer

Let us consider an example where an exponential boundary layer will develop after some time. The domain $\Omega = (0, 1)$ is partitioned by an equidistant grid with the mesh size $h = \frac{1}{160}$. We consider problem (1) with the data

$$\varepsilon = 10^{-9}, \quad T = 0.3, \quad b(x) = 2 - x, \quad c = 0, \quad f = 0, \quad u_0(x) = 2x \sin(\pi x)$$

and homogeneous Dirichlet boundary conditions. In this problem, the unknown exact solution develops an exponential boundary layer at the outflow boundary $x = 1$. In Figure 9, we compare at time $t_N = 0.3$ the discrete LPS-solution (with parameter $\gamma_0 = 0.1$) and the discrete solution based on additional shock capturing (LPS-SC with parameter $\gamma_{SC} = 2.0$). We clearly see that the pure LPS solution generates oscillations with high amplitude in the

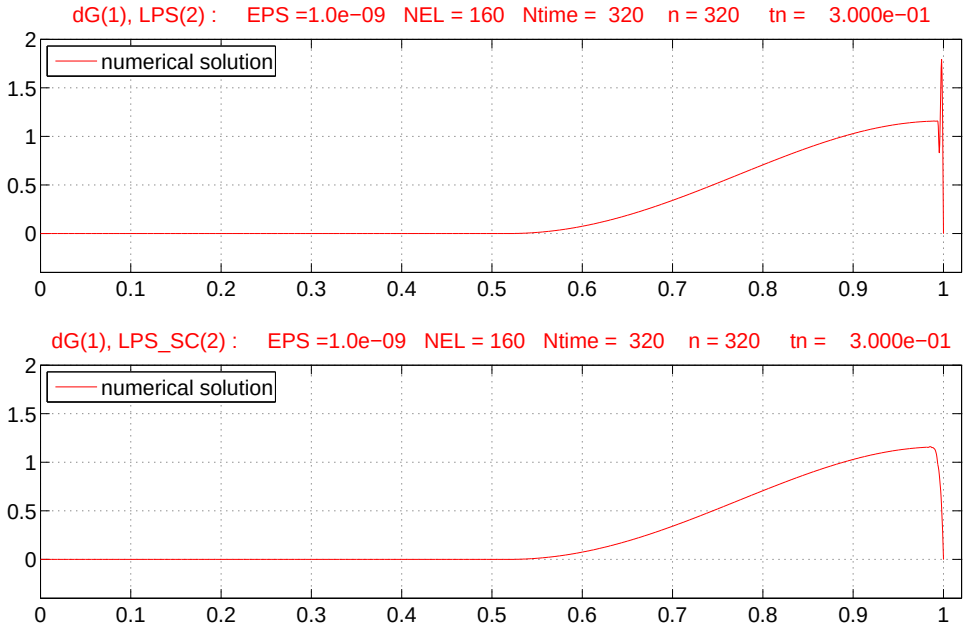

Figure 9: LPS- (top) and LPS-SC-solution (bottom) for $\varepsilon = 1e - 9$ and $t_n = 0.3$.

layer region whereas our shock capturing method creates an oscillation free solution.

References

[1] M. Braack and E. Burman. Local projection stabilization for the Oseen problem and its interpretation as a variational multiscale method. *SIAM J. Numer. Anal.*, 43(6):2544–2566 (electronic), 2006.

[2] E. Burman. A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty. *SIAM J. Numer. Anal.*, 43(5):2012–2033 (electronic), 2005.

[3] E. Burman and A. Ern. Continuous interior penalty *hp*-finite element methods for advection and advection-diffusion equations. *Math. Comp.*, 76(259):1119–1140 (electronic), 2007.

[4] V. Dolejší, M. Feistauer, V. Kučera, and V. Sobotíková. An optimal $L^\infty(L^2)$-error estimate for the discontinuous Galerkin approximation of a nonlinear non-stationary convection-diffusion problem. *IMA J. Numer. Anal.*, 28(3):496–521, 2008.

[5] V. Dolejší, M. Feistauer, and C. Schwab. On some aspects of the discontinuous Galerkin finite element method for conservation laws. *Math. Comput. Simulation*, 61(3-6):333–346, 2003. MODELLING 2001 (Pilsen).

[6] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson. *Computational differential equations*. Cambridge University Press, Cambridge, 1996.

[7] A. Ern and J.-L. Guermond. *Theory and practice of finite elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.

[8] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.

[9] M. Feistauer, J. Hájek, and K. Svadlenka. Space-time discontinuous Galerkin method for solving nonstationary convection-diffusion-reaction problems. *Appl. Math.*, 52(3):197–233, 2007.

[10] M. Feistauer and K. Švadlenka. Discontinuous Galerkin method of lines for solving nonstationary singularly perturbed linear problems. *J. Numer. Math.*, 12(2):97–117, 2004.

[11] J.-L. Guermond. Stabilization of Galerkin approximations of transport equations by subgrid modeling. *M2AN Math. Model. Numer. Anal.*, 33(6):1293–1316, 1999.

[12] J.-L. Guermond, A. Marra, and L. Quartapelle. Subgrid stabilized projection method for 2D unsteady flows at high Reynolds numbers. *Comput. Methods Appl. Mech. Engrg.*, 195(44-47):5857–5876, 2006.

[13] T. J. R. Hughes and M. Mallet. A new finite element formulation for computational fluid dynamics. IV. A discontinuity-capturing operator for multidimensional advective-diffusive systems. *Comput. Methods Appl. Mech. Engrg.*, 58(3):329–336, 1986.

[14] C. Johnson and A. Szepessy. On the convergence of a finite element method for a nonlinear hyperbolic conservation law. *Math. Comp.*, 49(180):427–444, 1987.

[15] T. Knopp, G. Lube, and G. Rapin. Stabilized finite element methods with shock capturing for advection-diffusion problems. *Comput. Methods Appl. Mech. Engrg.*, 191(27-28):2997–3013, 2002.

[16] G. Lube and G. Rapin. Residual-based stabilized higher-order FEM for a generalized Oseen problem. *Math. Models Methods Appl. Sci.*, 16(7):949–966, 2006.

[17] G. Matthies, P. Skrzypacz, and L. Tobiska. A unified convergence analysis for local projection stabilisations applied to the Oseen problem. *M2AN Math. Model. Numer. Anal.*, 41(4):713–742, 2007.

[18] G. Matthies, P. Skrzypacz, and L. Tobiska. Stabilization of local projection type applied to convection-diffusion problems with mixed boundary conditions. *Electron. Trans. Numer. Anal.*, 32:90–105, 2008.

[19] P. Oswald. On a BPX-preconditioner for P1 elements. *Computing*, 51(2):125–133, 1993.

[20] H.-G. Roos, M. Stynes, and L. Tobiska. *Robust numerical methods for singularly perturbed differential equations*, volume 24 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2008. Convection-diffusion-reaction and flow problems.

[21] F. Schieweck. A general transfer operator for arbitrary finite element spaces. Preprint 25/00, Otto-von-Guericke Universität Magdeburg, Fakultät für Mathematik, 2000. `http://www-ian.math.uni-magdeburg.de/home/schieweck`.

[22] V. Thomée. *Galerkin finite element methods for parabolic problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006.

[23] P. Wesseling. *Principles of computational fluid dynamics*, volume 29 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2001.