A ROBUST PETROV-GALERKIN DISCRETISATION OF CONVECTION-DIFFUSION EQUATIONS

DIRK BROERSEN AND ROB STEVENSON

Abstract. A Petrov-Galerkin discretization is studied of an ultra-weak variational formulation of the convection-diffusion equation in mixed form. To arrive at an implementable method, the truly optimal test space has to be replaced by its projection onto a finite dimensional test search space. To prevent that this latter space has to be taken increasingly large for vanishing diffusion, a formulation is constructed that is well-posed in the limit case of a pure transport problem. Numerical experiments show approximations that are very close to the best approximations to the solution from the trial space, uniformly in the size of the diffusion term.

1. Introduction

It is well-known that standard Galerkin discretisations of convection-diffusion equations fail to deliver good approximations for a vanishing diffusion term. In this paper, we study *Petrov-Galerkin* discretisations.

Unless the layers are resolved by the mesh, the H^1-errors of finite element approximations will be dominated by the errors in the layers. This holds also true for L_2-errors when conforming finite elements are applied due to the strong enforcement of Dirichlet boundary conditions. Therefore, we prefer to measure the errors in the L_2-norm, and to allow for discontinuous approximations. To this end, we consider an *ultra-weak* variational formulation of the convection-diffusion equation in mixed form. It is shown to define a boundedly invertible mapping $U \to V'$, with U and V being Hilbert spaces, where U is (essentially) a multiple copy of the L_2-space.

Building on the earlier works [BM84, DG11, CDW12], we equip V with the operator-dependent *optimal test norm*. Then given a finite dimensional trial space $U^h \subset U$, the Petrov-Galerkin discretisation with the *optimal test space* delivers the best approximation from U^h to the solution w.r.t. the norm on U.

To arrive at an implementable method, this truly optimal test space has to be replaced by its projection onto a finite dimensional *test search space*. With common variational formulations, the truly optimal test functions exhibit layers, and for vanishing diffusion, the test search space has to be chosen increasingly large to get satisfactory results.

Date: April 24, 2014.

2010 *Mathematics Subject Classification.* 65N12, 65N30, 35B25, 35J20.

Key words and phrases. Convection-diffusion equation, Petrov-Galerkin discretisation, optimal test space.

The first author has been supported by the Netherlands Organization for Scientific Research (NWO) under contract. no. 613.001.109.

In this paper, a non-standard variational formulation is constructed, such that for a zero diffusion term, the discrete system is a well-posed Petrov-Galerkin discretisation of the limiting transport problem. This can be seen as a necessary condition for the equations, which define the optimal test functions, not to be singularly perturbed.

Numerical experiments show that with a fixed test search space, the obtained approximations are very close to the *best* approximations to the solution from the trial space, uniformly in the size of the diffusion term.

This paper is organised as follows. In Sect. 2, we revisit the theory of Petrov-Galerkin discretizations with optimal test spaces. In Sect. 3, we apply it to convection-diffusion equations, and in Sect. 4 we present numerical results.

In this work, by $C \lesssim D$ we will mean that C can be bounded by a multiple of D, independently of parameters which C and D may depend on. Obviously, $C \gtrsim D$ is defined as $D \lesssim C$, and $C \eqsim D$ as $C \lesssim D$ and $C \gtrsim D$.

2. Some general theory

2.1. **Petrov-Galerkin discretizations with optimal test spaces.** For Hilbert spaces U and V over the scalar field $\mathbb{R}$, a bilinear form $b : U \times V \to \mathbb{R}$, let $(Bu)(v) := b(u, v)$ define a boundedly invertible mapping, i.e.,

$$B \in \mathcal{L}(U, V'), \quad B^{-1} \in \mathcal{L}(V', U). \tag{2.1}$$

Given $f \in V'$, we are interested in solving

$$Bu = f.$$

For defining our method, we will make use of $T \in \mathcal{L}(U, V)$ defined by

$$\langle Tu, v\rangle_V = b(u, v) \quad (u \in U,\, v \in V). \tag{2.2}$$

With the Riesz map $R_V \in \mathcal{L}(V, V')$ defined by $(R_V v)(z) = \langle v, w\rangle_V$ $(v, z \in V)$, it holds that $T = R_V^{-1} B$. Following [DG11], given a closed linear *trial space* $U^h \subset U$, we set the *optimal test space*

$$\operatorname{ran} T|_{U^h},$$

and consider the *Petrov-Galerkin* problem of finding $u^h \in U^h$ such that

$$b(u^h, v^h) = f(v^h) \quad (v^h \in \operatorname{ran} T|_{U^h}). \tag{2.3}$$

As will follow as a special case from Proposition 2.2, (2.3) has a unique solution, and it holds that

$$u^h = \underset{\bar{u}^h \in U^h}{\operatorname{argmin}} \|f - B\bar{u}^h\|_{V'},$$

so that actually the Petrov-Galerkin discretization with optimal test space is a least-squares method.

Only in cases where the dual norm $\|\cdot\|_{V'}$ can be evaluated exactly, this least-squares problem can be solved exactly. For this reason, in the following subsection we consider Petrov-Galerkin discretizations with *projected optimal test spaces*.

2.2. **Petrov-Galerkin with *projected* optimal test spaces.** Given a closed linear trial space $U^h \subset U$, let $V^h \subset V$ be a sufficiently large closed subspace, that we call *test search space*, such that

$$\gamma^h := \inf_{0 \neq w^h \in U^h} \sup_{0 \neq v^h \in V^h} \frac{b(w^h, v^h)}{\|w^h\|_U \|v^h\|_V} > 0. \tag{2.4}$$

Thanks to (2.1), in any case the latter is satisfied for $V^h = V$, with $\gamma^h \geq \|B^{-1}\|_{V' \to U}^{-1}$ (with equality when $U^h = U$).

Remark 2.1 (Fortin projector). From [For77], we recall that if there exists a projector $\Pi^h \in \mathcal{L}(V, V^h)$ with $b(w^h, \Pi^h v) = b(w^h, v)$ $(w^h \in U^h)$, then

$$\gamma^h \geq \inf_{0 \neq w^h \in U^h} \sup_{0 \neq v \in V} \frac{b(w^h, \Pi^h v)}{\|w^h\|_U \|\Pi^h v\|_V} \geq \frac{1}{\|\Pi^h\|_{V \to V} \|B^{-1}\|_{V' \to U}}$$

Conversely, if (2.4) is valid, then defining $\Pi^h v$ as the first component of the solution $(v^h, \lambda^h) \in V^h \times U^h$ of

$$\begin{aligned} \langle v^h, z^h \rangle_V + b(\lambda^h, z^h) &= \langle v, z^h \rangle_V \quad (z^h \in V^h), \\ b(w^h, v^h) &= b(w^h, v) \quad (w^h \in U^h), \end{aligned}$$

a projector as above is constructed, with $\|\Pi^h\|_{V \to V} \lesssim (\gamma^h)^{-1}$.

We define $T^h \in \mathcal{L}(U, V^h)$ by

$$\langle T^h u, v^h \rangle_V = b(u, v^h) \quad (u \in U,\, v \in V^h), \tag{2.5}$$

whose existence is guaranteed by Riesz' representation theorem.

Given a closed linear trial space $U^h \subset U$, we set the *projected optimal test space* by

$$\operatorname{ran} T^h|_{U^h},$$

and consider the *Petrov-Galerkin* problem of finding $\tilde{u}^h \in U^h$ such that

$$b(\tilde{u}^h, v^h) = f(v^h) \quad (v^h \in \operatorname{ran} T^h|_{U^h}). \tag{2.6}$$

Comparing (2.2) with (2.5), one infers that $\operatorname{ran} T^h|_{U^h}$ is the V-orthogonal projection of the optimal test space $\operatorname{ran} T|_{U^h}$ onto V^h.

In the following proposition, it will be shown that this latter Petrov-Galerkin problem has an equivalent formulation as a saddle point problem, that only involves the trial space U^h and test search space V^h, and so not explicitly the projected optimal test space $\operatorname{ran} T^h|_{U^h}$. This was the approach followed by Cohen, Dahmen and Welper in [CDW12].

Proposition 2.2. *Problems equivalent to* (2.6) *are finding* $\tilde{u}^h \in U^h$ *such that*

$$\langle T^h \tilde{u}^h, T^h w^h \rangle_V = f(T^h w^h) \quad (w^h \in U^h); \tag{2.7}$$

finding $(\tilde{y}^h, \tilde{u}^h) \in V^h \times U^h$ *such that*

$$\left\{ \begin{array}{lll} \langle \tilde{y}^h, v^h \rangle_V + b(\tilde{u}^h, v^h) & = & f(v^h) \quad (v^h \in V^h), \\ b(w^h, \tilde{y}^h) & = & 0 \qquad\;\; (w^h \in U^h); \end{array} \right. \tag{2.8}$$

and finally

$$\tilde{u}^h = \underset{\bar{u}^h \in U^h}{\operatorname{argmin}} \sup_{0 \neq v^h \in V^h} \frac{|f(v^h) - b(\bar{u}^h, v^h)|}{\|v^h\|_V}. \tag{2.9}$$

Under the assumption of $\gamma^h > 0$, (2.6)-(2.9) have a unique solution $\tilde{u}^h$ with $\|\tilde{u}^h\|_U \leq \|f\|_{V'}/\gamma^h$.

Proof. The equivalence of (2.6) and (2.7) follows from the definition of T^h.

With $R_{V^h} \in \mathcal{L}(V^h, (V^h)')$ defined by $(R_{V^h} v^h)(z^h) = \langle v^h, z^h \rangle_V$ $(v^h, z^h \in V^h)$, (2.7) can be written as

$$(R_{V^h} T^h|_{U^h} \tilde{u}^h)(T^h|_{U^h} w^h) = f(T^h|_{U^h} w^h) \quad (w^h \in U^h),$$

or, in operator form, as

$$(T^h|_{U^h})' R_{V^h} T^h|_{U^h} \tilde{u}^h = (T^h|_{U^h})' f,$$

and so by $T^h|_{U^h} = R_{V^h}^{-1} B^h$, with $B^h \in \mathcal{L}(U^h, (V^h)')$ being defined by $(B^h w^h)(v^h) = b(w^h, v^h)$ $(w^h \in U^h,\ v^h \in V^h)$, as

$$(B^h)' R_{V^h}^{-1} B^h \tilde{u}^h = (B^h)' R_{V^h}^{-1} f, \tag{2.10}$$

i.e., as finding $(\tilde{y}^h, \tilde{u}^h) \in V^h \times U^h$ such that

$$\begin{bmatrix} R_{V^h} & B^h \\ (B^h)' & 0 \end{bmatrix} \begin{bmatrix} \tilde{y}^h \\ \tilde{u}^h \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix}, \tag{2.11}$$

which in variational form reads as (2.8).

For $\bar{u}^h \in U^h$, we have

$$\sup_{0 \neq v^h \in V^h} \frac{|f(v^h) - b(\bar{u}^h, v^h)|}{\|v^h\|_V} = \sup_{0 \neq v^h \in V^h} \frac{\langle R_{V^h}^{-1}(f - B^h \bar{u}^h), v^h \rangle_V}{\|v^h\|_V} = \|R_{V^h}^{-1}(f - B^h \bar{u}^h)\|_V,$$

Consequently, $\tilde{u}^h$ from (2.9) is the unique solution in U^h of

$$0 = \langle R_{V^h}^{-1}(f - B^h \bar{u}^h), R_{V^h}^{-1} B^h w^h \rangle_V = (f - B^h \tilde{u}^h)(T^h w^h) = f(T^h w^h) - b(\tilde{u}^h, T^h w^h)$$

for all $w^h \in U^h$. We conclude that (2.6) and (2.9) are equivalent.

From

$$\begin{aligned} \|T^h w^h\|_V &= \sup_{0 \neq v^h \in V^h} \frac{\langle T^h w^h, v^h \rangle_V}{\|v^h\|_V} \\ &= \sup_{0 \neq v^h \in V^h} \frac{b(w^h, v^h)}{\|v^h\|_V} \in \left[\gamma^h \|w^h\|_U, \|B\|_{U \to V'} \|w^h\|_U\right] \quad (w^h \in U^h), \end{aligned}$$

we conclude that (2.7) is a symmetric, bounded, coercive variational problem on U^h, that therefore is uniquely solvable.

Substituting $w^h = \tilde{u}^h$ in (2.7), we find that $\|T^h \tilde{u}^h\|_V^2 \leq \|f\|_{V'} \|T^h \tilde{u}^h\|_V$, and so $\gamma^h \|\tilde{u}^h\|_U \leq \|T^h \tilde{u}^h\|_V \leq \|f\|_{V'}$. $\square$

The formulations (2.6)-(2.7) are relevant for practical implementations when one has a basis for $V^h \subset V$ available such that the resulting mass matrix has a sparse inverse. This can be expected when V is an L_2-space, or a "broken" Sobolev space. In this setting one can reduce the saddle-point formulation to its Schur complement, or, alternatively, given a basis for U^h, one can find a basis for $\operatorname{ran} T^h|_{U^h}$ by applying T^h to each basis function for U^h by solving (2.5), and use this basis for solving (2.6), or the symmetric positive definite system (2.7). The latter approach has been advocated by Demkowicz and Gopalakrishnan and collaborators in a sequence of papers starting with [DG11].

In the following proposition, an error bound is given for the solution of the Petrov-Galerkin discretisation with projected optimal test space, in terms of the error of best approximation from the trial space and the inf-sup constant γ^h. In a slightly different form, it can be found in [GQ14, Thm. 2.1].

Proposition 2.3. *When $\gamma^h > 0$, the solution $\tilde{u}^h$ of* (2.6) *satisfies*

$$\|u - \tilde{u}^h\|_U \leq \frac{\|B\|_{U \to V'}}{\gamma^h} \inf_{w^h \in U^h} \|u - w^h\|_U.$$

Proof. The statement holds true when $U^h = \{0\}$, because of $\gamma^h \leq \|B\|_{U \to V'}$, as well as when $U^h = U$. Indeed, since $B : U \to V'$ is boundedly invertible, for $U^h = U$, the condition $\gamma^h > 0$ requires $V^h = V$, so that "$\tilde{u}^h$"$= u$.

Now let $\{0\} \subsetneq U^h \subsetneq U$. Let us denote the mapping $u \mapsto \tilde{u}^h$ by P^h. Clearly $P^h : U \to U$ is a projector onto U^h, and so for any $w^h \in U^h$, one has $u - \tilde{u}^h = (I - P^h)u = (I - P^h)(u - w^h)$. From the last statement of Proposition 2.2, it follows that P^h is bounded, with $\|P^h\|_{U \to U} \leq \frac{\|B\|_{U \to V'}}{\gamma^h}$. Since $P^h \neq 0$ and $P^h \neq I$ by our condition on U^h, it holds that $\|I - P^h\|_{U \to U} = \|P^h\|_{U \to U}$ (see [Kat60] or [XZ03, Lemma 5]), which completes the proof. ☐

Remark 2.4. The crucial constant γ^h can be monitored by computing it as the square root of the smallest eigenvalue of $\mathbf{M}_{U^h}^{-1} \mathbf{B}_h^T \mathbf{M}_{V^h}^{-1} \mathbf{B}_h$. Here, w.r.t. some bases Φ and Σ of U^h and V^h, respectively, the aforementioned mass and "stiffness" matrices are defined by $(\mathbf{M}_{U^h}\mathbf{x})^\top \mathbf{y} = \langle \mathbf{x}^\top \Phi, \mathbf{y}^\top \Phi \rangle_U$, $(\mathbf{M}_{V^h}\mathbf{x})^\top \mathbf{y} = \langle \mathbf{x}^\top \Sigma, \mathbf{y}^\top \Sigma \rangle_V$, and $(\mathbf{B}_h \mathbf{x})^\top \mathbf{y} = b(\mathbf{x}^\top \Phi, \mathbf{y}^\top \Sigma)$.

2.3. **The energy norm on U, and optimal test norm on V.** The factor $\|B\|_{U \to V'}/\gamma^h$ with which, in view of Proposition 2.3, the error in the Petrov Galerkin solution might be larger than the error in the best approximation from the trial space U^h, can be as large as the condition number $\|B\|_{U \to V'} \|B^{-1}\|_{V' \to U}$, even for $V^h = V$. With standard choices of the norms, for singularly perturbed problems this condition number tends to infinity for the singular perturbation parameter tending to its critical limit. An approach to control the condition number is to equip U with the operator-dependent *energy-norm*

$$\|B \cdot\|_{V'},$$

giving rise to a perfectly well-conditioned problem $Bu = f$, so with $\|B\|_{U \to V'} = 1 = \|B^{-1}\|_{V' \to U}$.

The following result, that extends upon [DHSW12, Thm.3.4], links in this setting the inf-sup condition (2.4) to the quality of best approximation of the truly optimal test space $\operatorname{ran} T|_{U^h}$ by elements from V^h.

Proposition 2.5. *For $\|\cdot\|_U := \|B \cdot\|_{V'}$, the constant γ^h defined in* (2.4) *satisfies*

$$\sup_{0 \neq z^h \in \operatorname{ran} T|_{U^h}} \inf_{v^h \in V^h} \frac{\|z^h - v^h\|_V}{\|z^h\|_V} = \sqrt{1 - (\gamma^h)^2}.$$

Proof. With $P^h : \operatorname{ran} T|_{U^h} \to V^h$ denoting the restriction to $\operatorname{ran} T|_{U^h}$ of the V-orthogonal projector onto V^h, the left hand side reads as $\|I - P^h\|_{V \to V}$.

It holds that

$$\|Tw^h\|_V = \sup_{0 \neq v \in V} \frac{\langle Tw^h, v \rangle_V}{\|v\|_V} = \sup_{0 \neq v \in V} \frac{b(w^h, v)}{\|v\|_V} = \sup_{0 \neq v \in V} \frac{(Bw^h)(v)}{\|v\|_V} = \|Bw^h\|_{V'},$$

and so

$$\inf_{0\neq w^h\in U^h} \frac{\|P^hTw^h\|_V}{\|Tw^h\|_V} = \inf_{0\neq w^h\in U^h} \sup_{0\neq v^h\in V^h} \frac{\langle P^hTw^h, v^h\rangle_V}{\|Tw^h\|_V\|v^h\|_V}$$

$$= \inf_{0\neq w^h\in U^h} \sup_{0\neq v^h\in V^h} \frac{\langle Tw^h, v^h\rangle_V}{\|Tw^h\|_V\|v^h\|_V} = \inf_{0\neq w^h\in U^h} \sup_{0\neq v^h\in V^h} \frac{b(w^h, v^h)}{\|Bw^h\|_{V'}\|v^h\|_V} = \gamma^h.$$

We infer that

$$\|I - P^h\|^2_{V\to V} = \sup_{0\neq w^h\in U^h} \frac{\|Tw^h - P^hTw^h\|^2_V}{\|Tw^h\|^2_V}$$

$$= \sup_{0\neq w^h\in U^h} \frac{\|Tw^h\|^2_V - \|P^hTw^h\|^2_V}{\|Tw^h\|^2_V} = 1 - \inf_{0\neq w^h\in U^h} \frac{\|P^hTw^h\|^2_V}{\|Tw^h\|^2_V} = 1 - (\gamma^h)^2. \qquad \square$$

Without making a proper choice of $\|\cdot\|_V$, the energy-norm $\|B\cdot\|_{V'}$ on U might not be the norm of interest in applications. Setting the *optimal test norm*

$$\|B'\cdot\|_{U'} \quad \text{on } V,$$

and equipping V' with the associated dual norm, for the energy-norm we have

$$\|Bw\|_{V'} = \sup_{0\neq v\in V} \frac{|(Bw)(v)|}{\|B'v\|_{U'}} = \sup_{0\neq v\in V} \frac{|(B'v)(w)|}{\|B'v\|_{U'}} = \sup_{0\neq g\in U'} \frac{|g(w)|}{\|g\|_{U'}} = \|w\|_U.$$

So equipping V with the optimal test norm, the resulting energy norm is *equal* to the original norm on U.

The use of this optimal test norm was proposed in [ZMD+11, DHSW12], but it can also already be found in [BM84].

Remark 2.6. To see the latter, note that if V is equipped with norm $\|B'\cdot\|_{U'}$, then $R_V = BR_U^{-1}B'$, with $R_U \in \mathcal{L}(U, U')$ defined by $(R_U w)(z) = \langle w, z\rangle_U$ $(w, z \in U)$, so that $T = (B')^{-1}R_U$, which is the inverse of the mapping "R_m" in [BM84, (2.1)] (there $U = V$ is considered). With this T, (2.3) reads as finding $u^h \in U^h$ such that $\langle u^h, w^h\rangle_U = f((B')^{-1}R_U w^h) = \langle B^{-1}f, w^h\rangle_U$ $(w^h \in U^h)$, which confirms that u^h is the best approximation from U^h to u w.r.t. $\|\cdot\|_U$.

A necessary condition to be able to *implement* the resulting Petrov-Galerkin discretization with a projected optimal test space using any of (2.6)-(2.9) is that the "optimal test inner product" $\langle B'\cdot, B'\cdot\rangle_{U'}$ on V can be evaluated. With the aim to be able to choose $U = L_2(\Omega)^K \simeq U'$ for some $K \in \mathbb{N}$, one therefore writes a boundary problem of second order as a first order system.

3. Application to convection dominated convection-diffusion equations

We apply a Petrov-Galerkin discretisation with a projected optimal test space to convection-diffusion equations. To obtain satisfactory results also when the layers are not being resolved by the mesh, we approximate the solution in L_2-norm with discontinuous finite elements. To this end, we consider an ultra-weak variational formulation of the equation in mixed form.

Remark 3.1. Other than in [DG11], see also [BMS02], where such an ultra-weak formulation is derived by an *element-wise* integration-by-parts, we employ an integration-by-parts over the global domain. An advantage of the element-wise approach is that the optimal test functions are solutions of local variational problems on the individual elements. In such a setting, it seems easier to control the error due to the replacement of the truly optimal test functions by projected ones. On the other hand, by the element-wise integration-by-parts, additional solution components are introduced, that have the mesh skeleton as their domain. Approximation errors in these additional components may dominate the errors in the components of main interest, in particular because the functions on the skeleton are measured in intrinsically stronger norms.

To prevent that the test search space has to be taken increasingly large for vanishing diffusion, we construct a Petrov-Galerkin discretisation that in the limit of having no diffusion is a proper discretisation of the pure transport problem. We start with studying the transport problem.

3.1. **Transport equation.** For a domain $\Omega \subset \mathbb{R}^n$, $\mathbf{b} \in L_\infty(\Omega)^n$ with $\operatorname{div} \mathbf{b} \in L_\infty(\Omega)$, $c \in L_\infty(\Omega)$, and with

$$\Gamma_\pm := \{x \in \partial\Omega\colon\ \pm \mathbf{b}(x) \cdot \mathbf{n}(x) > 0 \text{ a.e.}\}, \quad \Gamma_0 := \partial\Omega \setminus (\Gamma_- \cup \Gamma_+),$$

for given f and g, consider the transport equation

$$\begin{cases} \mathbf{b} \cdot \nabla u + cu = f & \text{on } \Omega, \\ \qquad\quad u = g & \text{on } \Gamma_-. \end{cases}$$

Multiplying the equation with smooth test functions v that vanish at Γ_+, and using that $\mathbf{b} \cdot \mathbf{n}$ vanishes on Γ_0, by applying integration-by-parts we arrive at the variational formulation of finding $u \in L_2(\Omega)$ such that for all those v,

$$(Bu)(v) := \int_\Omega u(cv - \operatorname{div} v\mathbf{b}) = \int_\Omega fv - \int_{\Gamma_-} gv\mathbf{b} \cdot \mathbf{n}. \tag{3.1}$$

With

$$H(\mathbf{b};\Omega) := \{w \in L_2(\Omega)\colon \mathbf{b} \cdot \nabla w \in L_2(\Omega)\},$$

equipped with $\|w\|^2_{H(\mathbf{b};\Omega)} := \|w\|^2_{L_2(\Omega)} + \|\mathbf{b} \cdot \nabla w\|^2_{L_2(\Omega)}$, we set its closed subspace

$$H_{0,\Gamma_+}(\mathbf{b};\Omega) := \operatorname{clos}_{H(\mathbf{b};\Omega)}\{w \in C(\bar{\Omega})\colon w = 0 \text{ on } \Gamma_+\}.$$

For the case that $\mathbf{b} \in C^1(\Omega)^n$, or, for some constant $\kappa > 0$, $c - \frac{1}{2} \operatorname{div} \mathbf{b} \geq \kappa$ a.e., it is known ([Bar70, DHSW12]) that $B'\colon v \mapsto cv - \operatorname{div} v\mathbf{b} \in \mathcal{L}(H_{0,\Gamma_+}(\mathbf{b};\Omega), L_2(\Omega))$ is boundedly invertible, and so

$$B \in \mathcal{L}(L_2(\Omega), H_{0,\Gamma_+}(\mathbf{b};\Omega)'), \quad B^{-1} \in \mathcal{L}(H_{0,\Gamma_+}(\mathbf{b};\Omega)', L_2(\Omega)).$$

Now let a family of closed trial spaces $U^h \subset U := L_2(\Omega)$ be selected, with corresponding sufficiently large test search spaces $V^h \subset V := H_{0,\Gamma_+}(\mathbf{b};\Omega)$, and let V be equipped with the optimal test norm $\|B' \cdot\|_{U'}$. Then the Petrov-Galerkin discretizations with projected optimal test spaces will yield near-best approximations to u from the trial spaces in the $L_2(\Omega)$-norm, assuming that the infimum over the family of γ^h defined in (2.4) is strictly positive.

In [DHSW12, Sect. 5], numerical results for this approach were presented for U^h being the space of piecewise bilinears w.r.t. a uniform partition of a two-dimensional domain into squares, and V^h being the space of continuous piecewise quadratics w.r.t. a one- or two-times further dyadically refined partition. In addition, numerical results were given for adaptively refined partitions.

3.2. **Convection-diffusion-reaction equation.** We consider the boundary value problem

$$\left\{\begin{aligned} -\operatorname{div}\mathbf{A}\nabla u+\mathbf{b}\cdot\nabla u+cu&=f &&\text{on } \Omega,\\ u&=g &&\text{on } \partial\Omega,\end{aligned}\right. \tag{3.2}$$

where $\mathbf{A}\in L_\infty(\Omega)^{n\times n}$ is real, symmetric, and invertible with $\mathbf{A}^{-1}\in L_\infty(\Omega)^{n\times n}$, $\mathbf{b}\in L_\infty(\Omega)^n$ with $\operatorname{div}\mathbf{b}\in L_\infty(\Omega)$, and $c\in L_\infty(\Omega)$.

We assume that the standard, non-mixed variational formulation of (3.2) in case of homogeneous Dirichlet boundary conditions is well-posed, i.e.,

$$u\mapsto \Big(v\mapsto\int_\Omega \mathbf{A}\nabla u\cdot\nabla v+v\mathbf{b}\cdot\nabla u+cuv\Big)\in\mathcal{L}(H_0^1(\Omega),H_0^1(\Omega)'))\text{ is boundedly invertible.} \tag{3.3}$$

Writing $\mathbf{A}=\mathbf{A}_1\mathbf{A}_2$, where $\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_1^{-1},\mathbf{A}_2^{-1}\in L_\infty(\Omega)^{n\times n}$, and introducing $\boldsymbol{\sigma}=\mathbf{A}_2\nabla u$, we consider the reformulation of (3.2) as the first order div-grad mixed system

$$\left\{\begin{aligned} \boldsymbol{\sigma}-\mathbf{A}_2\nabla u&=0 &&\text{on } \Omega,\\ -\operatorname{div}\mathbf{A}_1\boldsymbol{\sigma}+\mathbf{b}\cdot\nabla u+cu&=f &&\text{on } \Omega,\\ u&=g &&\text{on } \partial\Omega.\end{aligned}\right. \tag{3.4}$$

We test the first equation with

$$\boldsymbol{\tau}\in H(\operatorname{div}\mathbf{A}_2^\top;\Omega):=\{\boldsymbol{\tau}\in L_2(\Omega)^n:\operatorname{div}\mathbf{A}_2^\top\boldsymbol{\tau}\in L_2(\Omega)\},$$

and the second one with $v\in H^1_{0,\Gamma_+}(\Omega)=\{v\in H^1(\Omega)\colon v|_{\Gamma_+}=0\}$. For some constant $\mu>0$, we introduce

$$\theta=-\mu^{-1}\mathbf{A}_1\boldsymbol{\sigma}|_{\partial\Omega\setminus\Gamma_+}\cdot\mathbf{n}$$

as an independent additional variable. We obtain the *ultra-weak* variational problem of finding $(\boldsymbol{\sigma},u)\in L_2(\Omega)^n\times L_2(\Omega)$, and θ from a space that will be specified later, such that

$$\left\{\begin{aligned} \int_\Omega\boldsymbol{\sigma}\cdot\boldsymbol{\tau}+u\operatorname{div}\mathbf{A}_2^\top\boldsymbol{\tau}&=\int_{\partial\Omega} g\mathbf{A}_2^\top\boldsymbol{\tau}\cdot\mathbf{n},\\ \int_\Omega\mathbf{A}_1\boldsymbol{\sigma}\cdot\nabla v-u\operatorname{div}v\mathbf{b}+cuv+\int_{\partial\Omega\setminus\Gamma_+}\mu\theta v&=\int_\Omega fv-\int_{\Gamma_-}gv\mathbf{b}\cdot\mathbf{n},\end{aligned}\right. \tag{3.5}$$

$(\boldsymbol{\tau}\in H(\operatorname{div}\mathbf{A}_2^\top;\Omega),\ v\in H^1_{0,\Gamma_+}(\Omega))$.

Remark 3.2. Note that for $f\in H_0^1(\Omega)'$ and $g\in H^{\frac12}(\partial\Omega)$, which conditions are necessary for the well-posedness of the standard non-mixed variational formulation, the linear functionals $\boldsymbol{\tau}\mapsto\int_{\partial\Omega}g\mathbf{A}_2^\top\boldsymbol{\tau}\cdot\mathbf{n}$ and $v\mapsto\int_\Omega fv-\int_{\Gamma_-}gv\mathbf{b}\cdot\mathbf{n}$ are in $H(\operatorname{div}\mathbf{A}_2^\top;\Omega)'$ and $H^1_{0,\Gamma_+}(\Omega)'$, respectively.

Remark 3.3. For $(\boldsymbol{\sigma},u,\theta)$ being the solution of (3.5), a reversed integration-by-parts of the first equation shows that $u\in H^1(\Omega)$ with $\nabla u=\mathbf{A}_2^{-1}\boldsymbol{\sigma}$, and $u=g$ on $\partial\Omega$. Now under the *additional condition* that $f\in L_2(\Omega)$, a reversed integration-by-parts of the second equation shows that $\mathbf{A}_1\boldsymbol{\sigma}\in H(\operatorname{div};\Omega)$, and $\theta=-\mu^{-1}\mathbf{A}_1\boldsymbol{\sigma}|_{\partial\Omega\setminus\Gamma_+}\cdot\mathbf{n}$.

Remark 3.4. Although Γ_+ has been given a precise meaning in the previous subsection, in the current subsection actually it could be read as *some* (measurable) subset of $\partial\Omega$, when reading the integral over Γ_- in the right-hand side of (3.5) as the integral over $\partial\Omega\setminus(\Gamma_0\cup\Gamma_+)$. The most obvious choice would be to take Γ_+ equal to the whole of $\partial\Omega$, with which the introduction of the additional variable

θ is avoided, and in which case $H^1_{0,\Gamma_+}(\Omega)$ reads as $H^1_0(\Omega)$. The next lemma deals with precisely this setting.

The motivation to take nevertheless Γ_+ as the outflow boundary of the corresponding transport problem will become clear in the next subsection, with the design of a Petrov-Galerkin discretization for solving the convection dominated convection-diffusion problem.

Lemma 3.5. *The operator $\bar{B} \in \mathcal{L}(L_2(\Omega)^n \times L_2(\Omega), H(\operatorname{div} \mathbf{A}_2^\top; \Omega)' \times H^1_0(\Omega)')$, defined by*

$$(\bar{B}(\boldsymbol{\sigma}, u))(\boldsymbol{\tau}, v) := \int_\Omega \boldsymbol{\sigma} \cdot \boldsymbol{\tau} + u \operatorname{div} \mathbf{A}_2^\top \boldsymbol{\tau} + \mathbf{A}_1 \boldsymbol{\sigma} \cdot \nabla v - u \operatorname{div} v\mathbf{b} + cuv,$$

is boundedly invertible.

Proof. The boundedness of $\bar{B}$ follows easily. Thanks to the open mapping theorem, it remains to show that $\bar{B}$ is invertible, which by an application of the closed range theorem is equivalent to surjectivity of both $\bar{B}$ and $\bar{B}'$.

Given $(\mathbf{f}, g) \in H(\operatorname{div} \mathbf{A}_2^\top; \Omega)' \times H^1_0(\Omega)'$, consider the problem of finding $(\boldsymbol{\sigma}, u) \in L_2(\Omega)^n \times L_2(\Omega)$ such that

$$\int_\Omega \boldsymbol{\sigma} \cdot \boldsymbol{\tau} + u \operatorname{div} \mathbf{A}_2^\top \boldsymbol{\tau} = \mathbf{f}(\boldsymbol{\tau}) \quad (\boldsymbol{\tau} \in H(\operatorname{div} \mathbf{A}_2^\top; \Omega)),$$

$$\int_\Omega \mathbf{A}_1 \boldsymbol{\sigma} \cdot \nabla v - u \operatorname{div} v\mathbf{b} + cuv = g(v) \quad (v \in H^1_0(\Omega)).$$

An application of Riesz' representation theorem shows that there exists an $\mathbf{r} \in H(\operatorname{div} \mathbf{A}_2^\top; \Omega)$ such that $\mathbf{f}(\boldsymbol{\tau}) = \int_\Omega \mathbf{r} \cdot \boldsymbol{\tau} + \operatorname{div} \mathbf{A}_2^\top \mathbf{r} \operatorname{div} \mathbf{A}_2^\top \boldsymbol{\tau}$. Introducing $\underline{\boldsymbol{\sigma}} = \boldsymbol{\sigma} - \mathbf{r}$, and $\underline{u} = u - \operatorname{div} \mathbf{A}_2^\top \mathbf{r}$, the above system reads as

$$\int_\Omega \underline{\boldsymbol{\sigma}} \cdot \boldsymbol{\tau} + \underline{u} \operatorname{div} \mathbf{A}_2^\top \boldsymbol{\tau} = 0 \quad (\boldsymbol{\tau} \in H(\operatorname{div} \mathbf{A}_2^\top; \Omega)),$$

$$\int_\Omega \mathbf{A}_1 \underline{\boldsymbol{\sigma}} \cdot \nabla v - \underline{u} \operatorname{div} v\mathbf{b} + c\underline{u}v =$$
$$g(v) - \int_\Omega \mathbf{A}_1 \mathbf{r} \cdot \nabla v - \operatorname{div} \mathbf{A}_2^\top \mathbf{r} \operatorname{div} v\mathbf{b} + cv \operatorname{div} \mathbf{A}_2^\top \mathbf{r} \quad (v \in H^1_0(\Omega)).$$

Thanks to (3.3), we may define $\underline{u}$ as the solution in $H^1_0(\Omega)$ of

$$\int_\Omega \mathbf{A} \nabla \underline{u} \cdot \nabla v + v\mathbf{b} \cdot \nabla \underline{u} + c\underline{u}v = g(v) - \int_\Omega \mathbf{A}_1 \mathbf{r} \cdot \nabla v - \operatorname{div} \mathbf{A}_2^\top \mathbf{r} \operatorname{div} v\mathbf{b} + cv \operatorname{div} \mathbf{A}_2^\top \mathbf{r},$$

$(v \in H^1_0(\Omega))$, and take $\underline{\boldsymbol{\sigma}} = \mathbf{A}_2 \nabla \underline{u}$. Then both equations are satisfied, $\boldsymbol{\sigma} = \underline{\boldsymbol{\sigma}} + \mathbf{r} \in L_2(\Omega)^n$, and $u = \underline{u} + \operatorname{div} \mathbf{A}_2^\top \mathbf{r} \in L_2(\Omega)$, completing the proof of the surjectivity of $\bar{B}$.

To show surjectivity of $\bar{B}'$, given $(\mathbf{f}, g) \in L_2(\Omega)^n \times L_2(\Omega)$, consider the problem of finding $(\boldsymbol{\tau}, v) \in H(\operatorname{div} \mathbf{A}_2^\top; \Omega) \times H^1_0(\Omega)$ such that

$$\boldsymbol{\tau} + \mathbf{A}_1^\top \nabla v = \mathbf{f}, \quad \operatorname{div} \mathbf{A}_2^\top \boldsymbol{\tau} - \operatorname{div} v\mathbf{b} + cv = g.$$

Since bounded invertibility of the mapping guaranteed by (3.3) implies bounded invertibility of the adjoint mapping, we may define v as the solution in $H^1_0(\Omega)$ of

$$\int_\Omega \mathbf{A} \nabla v \cdot \nabla \tilde{v} + v\mathbf{b} \cdot \nabla \tilde{v} + cv\tilde{v} = \int_\Omega g\tilde{v} + \mathbf{A}_2^\top \mathbf{f} \cdot \nabla \tilde{v} \quad (\tilde{v} \in H^1_0(\Omega)),$$

and take $\boldsymbol{\tau} = \mathbf{f} - \mathbf{A}_1^\top \nabla v \in L_2(\Omega)^n$. From $\int_\Omega -\mathbf{A}_2^\top \boldsymbol{\tau} \cdot \nabla \tilde{v} = \int_\Omega (-\mathbf{A}_2^\top \mathbf{f} + \mathbf{A}\nabla v) \cdot \nabla \tilde{v} = \int_\Omega -v\mathbf{b} \cdot \nabla \tilde{v} - cv\tilde{v} + g\tilde{v} = \int_\Omega \tilde{v}(\operatorname{div} v\mathbf{b} - cv - g)$, we find that $\operatorname{div} \mathbf{A}_2^\top \boldsymbol{\tau} = \operatorname{div} v\mathbf{b} - cv - g \in L_2(\Omega)$, with which the proof is completed. $\square$

Well-posedness of the mixed variational formulation (3.5), thus for Γ_+ being the outflow boundary (or more generally, for any $\Gamma_+ \subset \partial\Omega$ with $|\Gamma_+| > 0$), is established next.

Theorem 3.6. *For the operator B defined by*

$$(B(\boldsymbol{\sigma}, u, \theta))(\boldsymbol{\tau}, v) := (\bar{B}(\boldsymbol{\sigma}, u))(\boldsymbol{\tau}, v) + \int_{\partial\Omega\setminus\Gamma_+} \mu\theta v,$$

it holds that $B \in \mathcal{L}(L_2(\Omega)^n \times L_2(\Omega) \times H_{00}^{\frac{1}{2}}(\partial\Omega \setminus \Gamma_+)', H(\operatorname{div} \mathbf{A}_2^\top; \Omega)' \times H_{0,\Gamma_+}^1(\Omega)')$ is boundedly invertible.

Proof. Recall that for a normed linear space H, and a closed subspace M, the quotient space H/M is equipped with $\|v\|_{H/M} = \inf_{\{\tilde{v}\in H: v-\tilde{v}\in M\}} \|\tilde{v}\|_H$. If H is a Banach (Hilbert) space, then so is H/M. With the annihilator $M^\circ := \{f \in H' : f(M) = \{0\}\}$, being a closed subspace of H', we have $(H/M)' \simeq M^\circ$.

The trace mapping $v \mapsto v|_{\partial\Omega\setminus\Gamma_+}$ is in $\mathcal{L}(H_{0,\Gamma_+}^1(\Omega), L_2(\partial\Omega\setminus\Gamma_+))$, with kernel being equal to $H_0^1(\Omega)$. As a mapping on $H_{0,\Gamma_+}^1(\Omega)/H_0^1(\Omega)$, this trace is injective, and its range equipped with the norm on $H_{0,\Gamma_+}^1(\Omega)/H_0^1(\Omega)$ is known as $H_{00}^{\frac{1}{2}}(\partial\Omega \setminus \Gamma_+)$. It is a densely embedded subspace of $L_2(\partial\Omega \setminus \Gamma_+)$.

Since $\bar{B}$ from Lemma 3.5 is also bounded as a mapping from $L_2(\Omega)^n \times L_2(\Omega)$ to $H(\operatorname{div} \mathbf{A}_2^\top; \Omega)' \times H_{0,\Gamma_+}^1(\Omega)'$, boundedness of B follows from $H_{0,\Gamma_+}^1(\Omega) \hookrightarrow H_{0,\Gamma_+}^1(\Omega)/H_0^1(\Omega)$.

To show that it is boundedly invertible, given $(\mathbf{f}, g) \in H(\operatorname{div} \mathbf{A}_2^\top; \Omega)' \times H_{0,\Gamma_+}^1(\Omega)'$, consider the unique $(\boldsymbol{\sigma}, u) \in L_2(\Omega)^n \times L_2(\Omega)$ such that $\bar{B}(\boldsymbol{\sigma}, u)(\boldsymbol{\tau}, v) = (\mathbf{f}(\boldsymbol{\tau}), g(v))$ for all $(\boldsymbol{\tau}, v) \in H(\operatorname{div} \mathbf{A}_2^\top; \Omega) \times H_0^1(\Omega)$. The difference $(\bar{B}(u, \boldsymbol{\sigma}))_2 - g \in H_{0,\Gamma_+}^1(\Omega)'$ vanishes on $H_0^1(\Omega)$, i.e., it is an element of the annihilator $\{h \in H_{0,\Gamma_+}^1(\Omega)' : h(H_0^1(\Omega)) = \{0\}\} \eqsim (H_{0,\Gamma_+}^1(\Omega)/H_0^1(\Omega))'$. We conclude that there exists a unique $\theta \in H_{00}^{\frac{1}{2}}(\partial\Omega \setminus \Gamma_+)'$ such that

$$B(\boldsymbol{\sigma}, u, \theta)(\boldsymbol{\tau}, v) = (\mathbf{f}(\boldsymbol{\tau}), g(v)) \quad (\boldsymbol{\tau} \in H(\operatorname{div} \mathbf{A}_2^\top; \Omega), v \in H_{0,\Gamma_+}^1(\Omega)).$$

This completes the proof of B being invertible, and thus, by the open mapping theorem, of having a bounded inverse. $\square$

Let a family of closed trial spaces

$$U^h \subset U := L_2(\Omega)^n \times L_2(\Omega) \times H_{00}^{\frac{1}{2}}(\partial\Omega \setminus \Gamma_+)'$$

be selected, with corresponding sufficiently large test search spaces

$$V^h = \mathbf{V}_1^h \times V_2^h \subset V := H(\operatorname{div} \mathbf{A}_2^\top; \Omega) \times H_{0,\Gamma_+}^1(\Omega),$$

and let V be equipped with the optimal test norm $\|B' \cdot\|_{U'}$. Then the Petrov-Galerkin discretizations with projected optimal test spaces of (3.5) will yield near-best approximations to $(\boldsymbol{\sigma}, u, \theta)$ from the trial spaces in the norm on U, assuming that the infimum over the family of γ^h defined in (2.4) is strictly positive.

Besides the issue of how to guarantee the latter, in view of

$$B'(\boldsymbol{\tau}, v) = (\boldsymbol{\tau} + \mathbf{A}_1^\top \nabla v, \operatorname{div} \mathbf{A}_2^\top \boldsymbol{\tau} - \operatorname{div} v\mathbf{b} + cv, \mu v|_{\partial\Omega\setminus\Gamma_+}),$$

and $U' = L_2(\Omega)^n \times L_2(\Omega) \times H^{\frac{1}{2}}_{00}(\partial\Omega \setminus \Gamma_+)$, we have to discuss the replacement of the norm on $H^{\frac{1}{2}}_{00}(\partial\Omega \setminus \Gamma_+)$ by a computable one that is (uniformly) equivalent on the subspace $V_2^h|_{\partial\Omega\setminus\Gamma_+}$.

Using Σ as a shorthand notation for $\partial\Omega \setminus \Gamma_+$, an option that is not so attractive is to use that for $z \in H^{\frac{1}{2}}_{00}(\Sigma)$, $\|z\|^2_{H^{\frac{1}{2}}_{00}(\Sigma)} \eqsim \int_\Sigma (Wz)(s)z(s)ds$, where W is the hypersingular integral operator.

Using that $H^{\frac{1}{2}}_{00}(\Sigma) \simeq [L_2(\Sigma), H^1_0(\Sigma)]_{\frac{1}{2}}$, see [LM72, pp. 64-66 & 98-99], another option is to construct a (wavelet) Riesz basis for $H^{\frac{1}{2}}_{00}(\Sigma)$. Then $\|z\|_{H^{\frac{1}{2}}_{00}(\Sigma)} \eqsim \|\mathbf{z}\|_{\ell_2}$ uniformly in $z \in H^{\frac{1}{2}}_{00}(\Sigma)$, where $\mathbf{z}$ is the coefficient vector of z w.r.t. the basis. Since it cannot be expected that $z \in V_2^h|_\Sigma$ is given as a linear combination of these wavelets, it is preferable that the corresponding dual wavelets are locally supported, so that the coefficient vector $\mathbf{z}$ can be computed in linear complexity. For Σ being a two- (or one-) dimensional manifold, a suitable continuous piecewise linear finite element wavelet basis has been constructed in [CES00].

If one prefers to avoid the use of wavelets, then one cannot resort to the BPX "preconditioner" ([Xu92]). The union over all levels of the, properly scaled, nodal basis functions give rise to frames for Sobolev spaces with positive smoothness indices only, whereas here a frame for $H^{\frac{1}{2}}_{00}(\Sigma)'$ is needed. An attractive alternative is to apply the optimal multi-level "preconditioner" from [BPV00], that involves an efficient computation of an approximately orthogonal multi-level decomposition.

In our experiments, we will consider $\Omega = (0,1)^2$, and $\emptyset \subsetneq \Gamma_+ \subsetneq \partial\Omega$, and so $\Sigma = \partial\Omega \setminus \Gamma_+$, will be a connected union of sides of Ω. The space $Z^h := V_2^h|_\Sigma$ will be a space of continuous piecewise *cubics*, zero at $\partial\Sigma$, w.r.t. a uniform mesh with mesh-size h, which is constructed by dyadic refinements.

It is well known, see e.g. [Cao97], that the *hierarchical basis* for the space $\breve{Z}_h$ of continuous piecewise *linears* w.r.t. this mesh, zero at $\partial\Sigma$, is nearly stable in $H^{\frac{1}{2}}_{00}(\Sigma)$. Indeed, denoting this basis as Φ^h, formally viewed as a column vector, one has $\|\mathbf{c}^\top\Phi^h\|^2_{H^{\frac{1}{2}}_{00}(\Sigma)} \lesssim \|\mathbf{c}\|^2_{\ell_2} \lesssim |\log h|^2 \|\mathbf{c}^\top\Phi^h\|^2_{H^{\frac{1}{2}}_{00}(\Sigma)}$. Since for the linear interpolant $I^h \in \mathcal{L}(Z^h, \breve{Z}^h)$, it holds that $\|(I-I^h)z^h\|_{H^{\frac{1}{2}}_{00}(\Sigma)} \lesssim h^{-\frac{1}{2}}\|(I-I^h)z^h\|_{L_2(\Sigma)} \lesssim \|z^h\|_{H^{\frac{1}{2}}_{00}(\Sigma)}$, one infers that the extension of Φ^h by a uniformly $L_2(\Omega)$-stable basis for $\operatorname{ran}(I-I^h)$, scaled by a factor $h^{\frac{1}{2}}$, yields a basis for Z^h with qualitatively the same properties as that of Φ^h. That is, its condition number w.r.t. $\|\cdot\|_{H^{\frac{1}{2}}_{00}(\Sigma)}$ is proportional to $|\log h|^2$.

For any $z^h \in Z^h$, we will replace $\|z^h\|_{H^{\frac{1}{2}}_{00}(\Sigma)}$ by the ℓ_2-norm of its representation w.r.t. this basis. We envisage that the fact that the latter norm is thus not truly equivalent to the norm on $H^{\frac{1}{2}}_{00}(\Sigma)$ has only a marginal effect on the numerical results.

Alternatively, with some small additional effort, one could also apply a "coarse-grid" correction to the hierarchical basis to yield the fully stable basis constructed in [CES00].

3.3. **Convection dominated convection-diffusion equation.** For $\varepsilon > 0$, $\mathbf{b} \in L_\infty(\Omega)^n$ with $\operatorname{div}\mathbf{b} \in L_\infty(\Omega)$, and $c \in L_\infty(\Omega)$, we consider the boundary value problem

$$\left\{ \begin{aligned} -\varepsilon\Delta u + \mathbf{b}\cdot\nabla u + cu &= f \quad \text{on } \Omega, \\ u &= g \quad \text{on } \partial\Omega, \end{aligned} \right. \tag{3.6}$$

that is, (3.2) with $\mathbf{A} = \varepsilon I$.

We consider the mixed formulation (3.5) with $\mathbf{A}_1 = \mathbf{A}_2 = \sqrt{\varepsilon}\, I$ and $\mu = \sqrt{\varepsilon}$, i.e., the ultra-weak variational problem of finding $(\boldsymbol{\sigma}, u, \theta) \in U = L_2(\Omega)^n \times L_2(\Omega) \times H_{00}^{\frac{1}{2}}(\partial\Omega \setminus \Gamma_+)'$, such that

$$\left\{ \begin{aligned} \int_\Omega \boldsymbol{\sigma}\cdot\boldsymbol{\tau} + \sqrt{\varepsilon}\, u \operatorname{div}\boldsymbol{\tau} &= \int_{\partial\Omega} \sqrt{\varepsilon}\, g \boldsymbol{\tau}\cdot\mathbf{n}, \\ \int_\Omega \sqrt{\varepsilon}\,\boldsymbol{\sigma}\cdot\nabla v - u \operatorname{div} v\mathbf{b} + cuv + \int_{\partial\Omega\setminus\Gamma_+} \sqrt{\varepsilon}\,\theta v &= \int_\Omega fv - \int_{\Gamma_-} gv\mathbf{b}\cdot\mathbf{n}, \end{aligned} \right. \tag{3.7}$$

$((\boldsymbol{\tau}, v) \in V = H(\operatorname{div};\Omega) \times H^1_{0,\Gamma_+}(\Omega))$.

We equip V with (squared) optimal test norm, that reads, for $(\boldsymbol{\tau}, v) \in V$, as

$$\|\boldsymbol{\tau} + \sqrt{\varepsilon}\,\nabla v\|^2_{L_2(\Omega)^n} + \|\sqrt{\varepsilon}\, \operatorname{div}\boldsymbol{\tau} - \operatorname{div}\mathbf{b}v + cv\|^2_{L_2(\Omega)} + \|\sqrt{\varepsilon}\, v|_{\partial\Omega\setminus\Gamma_+}\|^2_{H_{00}^{\frac{1}{2}}(\partial\Omega\setminus\Gamma_+)}, \tag{3.8}$$

where, as explained in the previous subsection, we will replace the evaluation of the last norm applied to a finite element function by a (nearly) equivalent expression.

Although at the continuous level, we have introduced θ as an additional unknown, cf. Remark 3.4, below we will eliminate it at the discrete level as an independent function. The reason for doing this is that as an independent function, it would be undetermined for $\varepsilon = 0$, meaning that the discrete system would be singular.

One may wonder why θ has been introduced at all, cf. Remark 3.4. The reason is that its introduction induced the enlargement of the test space with functions that do not vanish at $\partial\Omega \setminus \Gamma_+$, test functions which are needed for the well-posedness of the limiting transport problem.

Proposition 3.7. *For $f \in L_2(\Omega)$, and with $(\boldsymbol{\sigma}, u, \theta)$ being the solution of (3.7), for $\boldsymbol{\Sigma}^h \subset H(\operatorname{div};\Omega)$ it holds that*

$$\inf_{\boldsymbol{\sigma}^h\in\boldsymbol{\Sigma}^h} \|(\boldsymbol{\sigma},\theta) - (\boldsymbol{\sigma}^h, -\boldsymbol{\sigma}^h|_{\partial\Omega\setminus\Gamma_+}\cdot\mathbf{n})\|_{L_2(\Omega)^n\times H_{00}^{\frac{1}{2}}(\partial\Omega\setminus\Gamma_+)'} \lesssim \inf_{\boldsymbol{\sigma}^h\in\boldsymbol{\Sigma}^h} \|\boldsymbol{\sigma} - \boldsymbol{\sigma}^h\|_{H(\operatorname{div};\Omega)}.$$

Proof. From $f \in L_2(\Omega)$, we know that $\theta = -\boldsymbol{\sigma}|_{\partial\Omega\setminus\Gamma_+}\cdot\mathbf{n}$ (see Remark 3.3). Now the result follows from $\boldsymbol{\sigma} \mapsto -\boldsymbol{\sigma}|_{\partial\Omega\setminus\Gamma_+}\cdot\mathbf{n} \in \mathcal{L}(H(\operatorname{div};\Omega), H_{00}^{\frac{1}{2}}(\partial\Omega\setminus\Gamma_+)')$ ([BS13, Lemma 3.3]). $\square$

In view of Proposition 3.7, for $P^h \subset L_2(\Omega)$ and $\boldsymbol{\Sigma}^h \subset H(\operatorname{div};\Omega)$ we take

$$U^h = \{(\boldsymbol{\sigma}^h, u^h, -\boldsymbol{\sigma}^h|_{\partial\Omega\setminus\Gamma_+}\cdot\mathbf{n})\colon \boldsymbol{\sigma}^h \in \boldsymbol{\Sigma}^h,\ u^h \in P^h\}. \tag{3.9}$$

Since V has been equipped with the optimal test norm, the Petrov-Galerkin discretization with the optimal test space delivers the best approximation to $(\boldsymbol{\sigma}, u, \theta)$ from this trial space U^h. An application of Proposition 2.3 shows that the Petrov-Galerkin solution $(\tilde{\boldsymbol{\sigma}}^h, \tilde{u}^h, \tilde{\theta}^h)$ with the application of a projected optimal test space satisfies

$$
\begin{aligned}
&\|\boldsymbol{\sigma}-\tilde{\boldsymbol{\sigma}}^h\|_{L_2(\Omega)^n}^2+\|u-\tilde{u}^h\|_{L_2(\Omega)}^2+\|\theta-\tilde{\theta}^h\|^2_{H_{00}^{\frac{1}{2}}(\partial\Omega\setminus\Gamma_+)} \\
&\leq (\gamma^h)^{-2}\inf_{(\boldsymbol{\sigma}^h,u^h,\theta^h)\in U^h}\|\boldsymbol{\sigma}-\boldsymbol{\sigma}^h\|_{L_2(\Omega)^n}^2+\|u-u^h\|_{L_2(\Omega)}^2+\|\theta-\theta^h\|^2_{H_{00}^{\frac{1}{2}}(\partial\Omega\setminus\Gamma_+)} \\
&\lesssim (\gamma^h)^{-2}\inf_{\boldsymbol{\sigma}^h\in\boldsymbol{\Sigma}^h,\, u^h\in P^h}\|\boldsymbol{\sigma}-\boldsymbol{\sigma}^h\|_{H(\mathrm{div};\Omega)}^2+\|u-u^h\|_{L_2(\Omega)}^2,
\end{aligned}
$$

by Proposition 3.7.

Remark 3.8. In view of the fact that the error in u is measured in $L_2(\Omega)$-norm, it is not very pleasant that the error in $\boldsymbol{\sigma}$, being a multiple of ∇u, now is measured in a norm that is even stronger than that on $L_2(\Omega)^n$, cf. the discussion in [BS13]. Fortunately, on the other hand it is advantageous that $\boldsymbol{\sigma}=\sqrt{\varepsilon}\,\nabla u$, so with a factor $\sqrt{\varepsilon}$ that vanishes for $\varepsilon\downarrow 0$.

For ε small we cannot expect much from the approximation of $\frac{\tilde{\boldsymbol{\sigma}}^h}{\sqrt{\varepsilon}}$ to $\nabla u=\frac{\boldsymbol{\sigma}}{\sqrt{\varepsilon}}$. Indeed, even assuming $(\gamma^h)^{-1}$ being uniformly bounded, one term in the above upper bound for $\|\frac{\boldsymbol{\sigma}}{\sqrt{\varepsilon}}-\frac{\tilde{\boldsymbol{\sigma}}^h}{\sqrt{\varepsilon}}\|_{L_2(\Omega)^n}$ reads as $\inf_{u^h\in P^h}\frac{1}{\sqrt{\varepsilon}}\|u-u^h\|_{L_2(\Omega)}$. Because for $\varepsilon=0$ membership of ∇u in $L_2(\Omega)^n$ is not guaranteed, note that any numerical approximation scheme of ∇u in $L_2(\Omega)^n$ is doomed to degenerate for $\varepsilon\downarrow 0$.

Fixing U^h and a test search space V^h, generally the inf-sup constant γ^h depends on ε. If the discrete system is singular for $\varepsilon=0$, as with the common formulations, then *necessarily* $\lim_{\varepsilon\downarrow 0}\gamma^h(\varepsilon)=0$. Typically, the truly optimal test functions develop boundary layers for $\varepsilon\downarrow 0$, and so are increasingly more difficult to approximate.

With the variational formulation (3.7), the Petrov-Galerkin discretization with trial space U^h from (3.9) and test search space

$$V^h=\mathbf{V}_1^h\times V_2^h\subset V \tag{3.10}$$

reads, for $\varepsilon=0$, as the Petrov-Galerkin discretization, with trial space $\boldsymbol{\Sigma}^h\times P^h$ and test search space V^h, of the decoupled system of finding $(\boldsymbol{\sigma},u)\in L_2(\Omega)^n\times L_2(\Omega)$ such that

$$
\begin{cases}
\displaystyle\int_\Omega \boldsymbol{\sigma}\cdot\boldsymbol{\tau}=0 & (\boldsymbol{\tau}\in L_2(\Omega)^n),\\
\displaystyle\int_\Omega u(cv-\operatorname{div} v\mathbf{b})=\int_\Omega fv-\int_{\Gamma_-} gv\mathbf{b}\cdot\mathbf{n} & (v\in H_{0,\Gamma_+}(\mathbf{b};\Omega)),
\end{cases} \tag{3.11}
$$

with $L_2(\Omega)^n\times H_{0,\Gamma_+}(\mathbf{b};\Omega)$ equipped with the optimal (squared) test norm $\|\boldsymbol{\tau}\|^2_{L_2(\Omega)^n}+\|cv-\operatorname{div} v\mathbf{b}\|^2_{L_2(\Omega)}$. To see this, consider the saddle-point formulation (2.8) of the Petrov-Galerkin discretization, and substitute $\varepsilon=0$ in (3.7) and (3.8).

Although useless, the first equation in (3.11) is well-posed, and the second equation is the well-posed formulation (3.1) of the limiting transport problem. When $\mathbf{V}_1^h\supseteq\boldsymbol{\Sigma}^h$ and $V_2^h\subset H^1_{0,\Gamma_+}(\Omega)\subset H_{0,\Gamma_+}(\mathbf{b};\Omega)$ be sufficiently large in relation to P^h, the Petrov-Galerkin discretization for $\varepsilon=0$ will yield a near-best approximation from U^h to the solution u of the transport problem in the $L_2(\Omega)$-norm.

We conclude that in any case a *necessary* condition has been fulfilled such that our Petrov-Galerkin discretization with trial space U^h from (3.9), and a sufficiently large, but ε-*independent* choice of the test search space V^h, yields a near-best approximation to $(\boldsymbol{\sigma}, u, \theta)$ from U^h in the norm on U *uniformly in* $\varepsilon \in [0, M]$, for some arbitrary constant $M > 0$.

The numerical results presented in the next section seem to suggest that such near-best approximations are found.

Remark 3.9. Continuing the discussion in Remark 3.4 and in the lines preceding Proposition 3.7, well-posedness of the Petrov-Galerkin discretisation for $\varepsilon = 0$ would *not* hold with the obvious variant of (3.7), where v runs over $H_0^1(\Omega)$, and the function θ does not appear. For small ε, numerical results with that formulation are incomparably worse.

Similarly important are the factorisation of $\mathbf{A} = \varepsilon I$ into two factors $\mathbf{A}_1$ and $\mathbf{A}_2$ that both vanish for $\varepsilon = 0$, the selection of μ dependent of ε such that it vanishes for $\varepsilon = 0$, and the choice of θ^h as a function of $\boldsymbol{\sigma}^h$.

4. Numerical results

We have tested the Petrov-Galerkin discretization with projected optimal test space of the ultra-weak mixed formulation (3.7) of the convection-diffusion problem for $\Omega = (0,1)^2$, $c = 0$ and $g = 0$. We equipped the test space $V = H(\mathrm{div}; \Omega) \times H^1_{0,\Gamma_+}(\Omega)$ with the (squared) optimal test norm (3.8) –where we approximated the arising $H_{00}^{\frac{1}{2}}(\partial\Omega \setminus \Gamma_+)$-norm in the way as explained at the end of Subsect. 3.2–, and considered trial spaces of the form (3.9). We used the formulation of the Petrov-Galerkin system as the saddle-point problem (2.8), and solved this system directly using the built-in `matlab` solver. In doing so, we did not encounter any instabilities due to ill-conditioning.

Remark 4.1. By equipping V with the optimal test norm, the bilinear form $b : U \times V \to \mathbb{R}$ is bounded with constant equal to 1. Obviously, $\langle \cdot, \cdot \rangle_V$ is bounded and coercive on $V \times V$ with constants equal to 1. So if $\gamma^h > 0$ uniformly in h and ε, then (2.8) defines a uniformly boundedly invertible linear mapping $V^h \times U^h \to (V^h \times U^h)'$. An optimal *iterative* solution method for (2.8) would now require uniformly stable bases for $U^h \subset U$ and $V^h \subset V$, or, equivalently, optimal preconditioners for resulting mass matrices.

For $\boldsymbol{\Sigma}^h$ and P^h, we took the Raviart-Thomas space $\mathrm{RT1}_h$ and the space of discontinuous piecewise linears, both w.r.t. the partition Ω^h, being the uniform partition of Ω into isosceles right-angled triangles with legs of length $h = 2^{-\ell}$ ($\ell \in \mathbb{N}_0$) and hypothenuses parallel to the vector $[1\ 1]^\top$.

For the test search space components $\mathbf{V}_1^h$ and V_2^h (cf. (3.10)), we took the Raviart-Thomas space $\mathrm{RT1}_{h/2}$ and the space of continuous piecewise cubics, zero at Γ_+, both w.r.t. the refined partition $\Omega_{h/2}$. Smaller test search spaces give less accurate results, whereas the improvements with larger spaces are marginal.

Although for our convenience, so far we only performed numerical experiments with uniform meshes, for completeness we emphasise that the application of our Petrov-Galerkin method is not restricted to such meshes.

Example 4.2. We took $\mathbf{b} = [2\ 1]^\top$, and right-hand side f such that the exact solution is

$$u(x,y) = [x + (e^{b_1 x/\varepsilon} - 1)/(1 - e^{b_1/\varepsilon})] \cdot [y + (e^{b_2 y/\varepsilon} - 1)/(1 - e^{b_2/\varepsilon})], \tag{4.1}$$

which has typical boundary layers at the top and right outflow boundaries. Comparisons of the solutions obtained by our Petrov-Galerkin discretisation and those

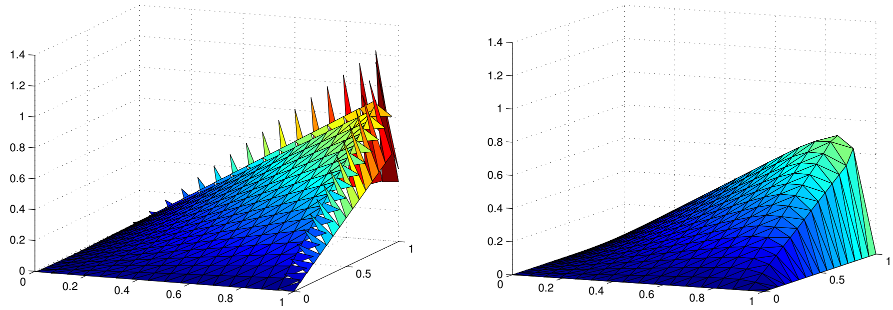

FIGURE 1. Approximate solution for Example 4.2, $\varepsilon = 10^{-2}$ and $h = 1/16$ with our Petrov-Galerkin method (left) and SUPG (right).

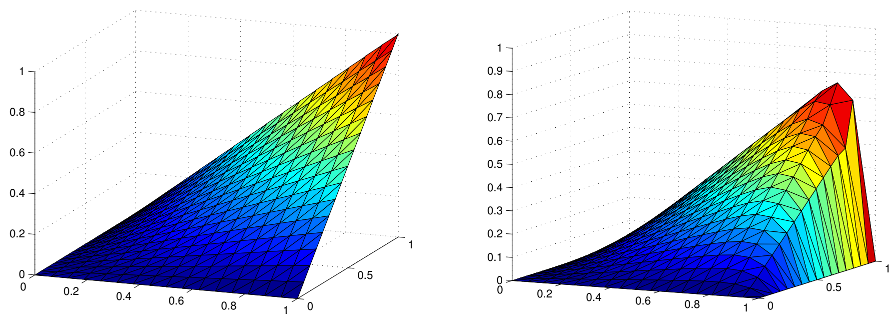

FIGURE 2. Approximate solution for Example 4.2, $\varepsilon = 10^{-6}$ and $h = 1/16$ with our Petrov-Galerkin method (left) and SUPG (right).

with the streamline upwind diffusion Petrov-Galerkin method (SUPG) (see [BH82]) are given in Figures 1 and 2. As recommended in a preprint version of [ES10], for the SUPG method we took as stabilisation parameter $\alpha = \max(h|\mathbf{b}| - 2\varepsilon, 0)/|\mathbf{b}|^2$, which turned out to give the best results.

In Figure 3, the $L_2((0,1)^2)$-errors in the computed u_h by our Petrov-Galerkin method are compared to those in the $L_2(\Omega)$-orthogonal projection of the exact solution onto the trial space P^h. It turns out that they are very close. Consequently, the plots of the approximate solution obtained by our Petrov-Galerkin method cannot be distinguished from those of the $L_2(\Omega)$-orthogonal projection of the exact solution onto the the trial space. This holds also true in the other examples where we know the exact solution.

For any $h \gtrapprox \varepsilon$, the difference of u and its $L_2(\Omega)$-best approximation is $\eqsim 1$ in the boundary layer of width $\eqsim \varepsilon$. This causes an $L_2(\Omega)$-error of order $\sqrt{\varepsilon}$, so essentially independent of $h \gtrapprox \varepsilon$. Only after having resolved the layer, i.e., for $h \lessapprox \varepsilon$ in case of uniform meshes, one may expect a quadrupling of the accuracy when halving the mesh size.

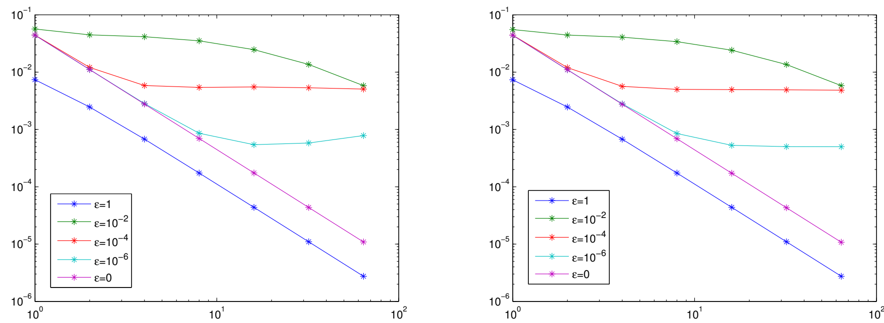

FIGURE 3. L_2-error vs. $1/h$ in u^h from Example 4.2 computed by our Petrov-Galerkin method (left), and in the $L_2(\Omega)$-orthogonal projection of the exact solution onto the space of discontinuous piecewise linears (right).

In both this and the next example, a numerically stable computation of the $L_2(\Omega)$-orthogonal projection onto the trial space P^h and the norm of its error turned out to be rather troublesome for $\varepsilon = 10^{-6}$. After having solved this, the squared norm of the error in the Petrov-Galerkin solution was computed as the sum of the squared norm of the error between the exact solution and its projection, and the squared norm of the difference between the projection and the Petrov-Galerkin solution.

Example 4.3. Following the example in [DH13, Sect. 3.1], we took $\mathbf{b} = [1\ 0]^\top$, and right-hand side f such that the exact solution is

$$u(x,y) = \Big(\frac{e^{r_1(x-1)} - e^{r_2(x-1)}}{e^{-r_1} - e^{-r_2}} + x - 1\Big)\sin \pi y,$$

where $r_{1,2} = \frac{-1 \pm \sqrt{1+4\varepsilon^2\pi^2}}{-2\varepsilon}$.

We compared u^h obtained by our Petrov-Galerkin method to the approximation to u produced by SUPG, the $L_2(\Omega)$-orthogonal projection of u onto P^h, and the approximation to u produced by the method by Demkowicz & Heuer in [DH13].

The latter method is a Petrov-Galerkin method with a projected optimal test space for an ultra-weak formulation of the mixed system (3.4), that is obtained by an *element-wise* partial integration of both equations w.r.t. the finite element partition Ω^h. Consequently, the test space is the product of spaces of "broken" $H(\mathrm{div})$- and H^1-functions. Apart from the trial functions $\boldsymbol{\sigma}^h$ and u^h, and the test functions $\boldsymbol{\tau}^h$ and v^h, by this procedure two additional trial functions $\hat{u}^h$ and $\hat{\sigma}_n^h$ are introduced that have the mesh skeleton as their domain.

For the method from [DH13], as trial and test search spaces we took discontinuous piecewise linears for u^h and $\boldsymbol{\sigma}^h$, and discontinuous cubics for $\boldsymbol{\tau}^h$ and v^h. As trial space for $\hat{u}^h$, we took the restriction to the skeleton of the continuous piecewise quadratics that vanish at $\partial\Omega$. As trial space for $\hat{\sigma}_n^h$, we took discontinuous piecewise linears on all edges. We equipped the test space with the weighted test norm $\|(v,\boldsymbol{\tau})\|_{V,2}$ as was recommended for this example.

Approximate solutions produced by our Petrov-Galerkin method and that from [DH13] are illustrated in Figure 4.

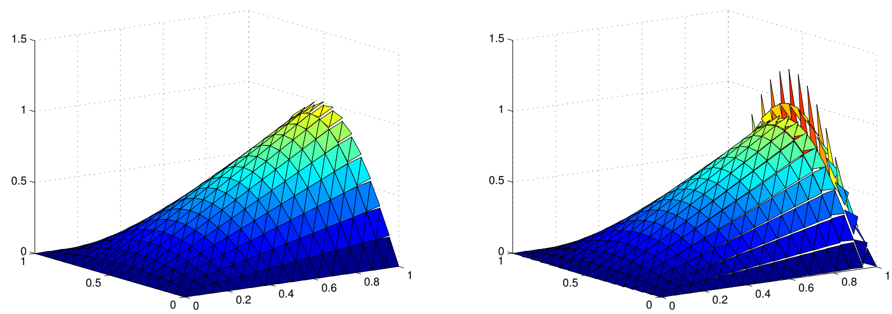

FIGURE 4. Approximate solution for Example 4.3, $\varepsilon = 10^{-4}$ and $h = 1/16$ with our Petrov-Galerkin method (left) and the method of Demkowicz and Heuer (right).

In Figure 5, the $L_2((0,1)^2)$-errors in the computed approximations to u are compared. The $L_2(\Omega)$-norm of the error in u^h obtained by our Petrov-Galerkin method is very close to that in the $L_2(\Omega)$-orthogonal projection of the solution u onto the space of discontinuous piecewise linears.

Example 4.4 (internal layer, not aligned with the mesh). In this example, $\mathbf{b} = [2\ 1]^\top$, $f(x,y) = \begin{cases} 1-x & y > x/2+1/4, \\ 0 & y < x/2+1/4, \end{cases}$ so for $\varepsilon = 0$, the solution is given by $u(x,y) = \begin{cases} x/2 - x^2/4 & y > x/2+1/4, \\ 0 & y < x/2+1/4. \end{cases}$

In Figure 6, for $\varepsilon = 0$ the *element-wise* L_2-error in u^h obtained by our Petrov-Galerkin discretisation is compared to that in the $L_2(\Omega)$-orthogonal projection of the solution u onto the space of discontinuous piecewise linears.

Solutions for $\varepsilon = 10^{-6}$ and $h = 1/16$ obtained by our Petrov-Galerkin discretisation and with SUPG are illustrated in Figure 7.

Example 4.5 (two internal layers, aligned with the mesh). In this example, $\mathbf{b} = [1\ 1]^\top$, $f(x,y) = \begin{cases} 1-x & y > x-1/4, \\ 0 & y < x-1/4, \end{cases}$ and so for $\varepsilon = 0$, the solution is given by $u(x,y) = \begin{cases} -\frac{1}{2}x^2 + x & y > x, \\ \frac{1}{2}y^2 - xy + y & x+1/4 < y < x, \\ 0 & y < x+1/4, \end{cases}$ meaning that it is discontinuous at $y = x - \frac{1}{4}$, and continuous at $x = y$ but with a normal derivative that has a jump at this curve.

Approximate solutions for $h = 1/16$ and $\varepsilon = 10^{-4}$ or $\varepsilon = 0$ obtained by our Petrov-Galerkin discretisation are illustrated in Figure 8.

5. Conclusion

We studied an ultra-weak variational formulation of the convection-diffusion equation $\begin{cases} -\varepsilon\Delta u + \mathbf{b}\cdot\nabla u = f & \text{on } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases}$ in mixed form. It was shown to define a boundedly invertible operator between spaces U and V', where U is essentially the product of the space $L_2(\Omega)$ for u, being the variable of our main interest, and

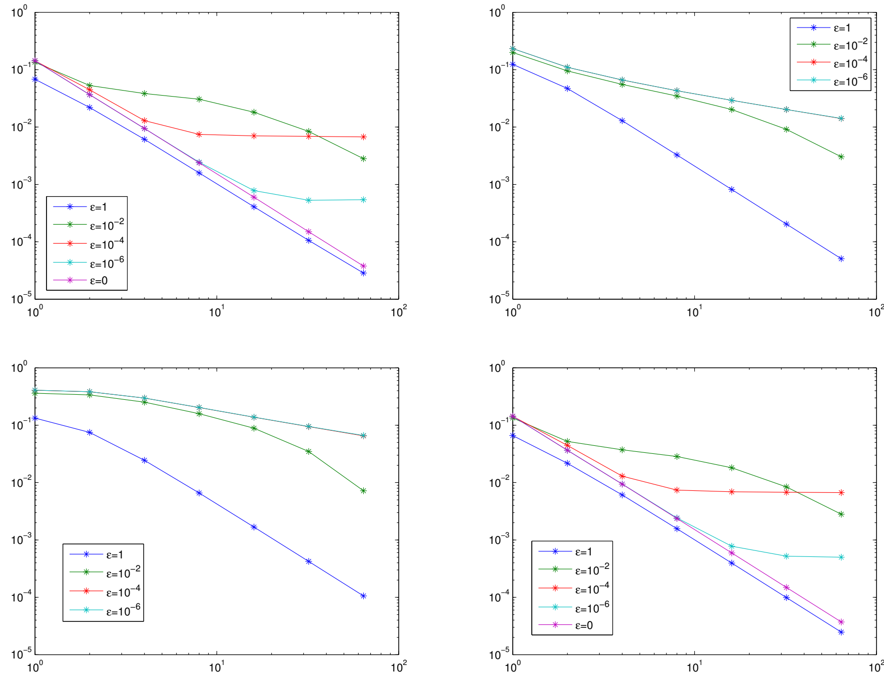

FIGURE 5. L_2-error vs. $1/h$ in u^h from Example 4.3 computed by our Petrov-Galerkin method (top left), by the method of Demkowicz & Heuer (top right), by SUPG (bottom left), and in the $L_2(\Omega)$-orthogonal projection of the solution u onto the space of discontinuous piecewise linears (bottom right). For both the SUPG and the method of Demkowicz & Heuer, the curves for $\varepsilon = 10^{-4}$ and $\varepsilon = 10^{-6}$ coincide as the upper curves in these figures.

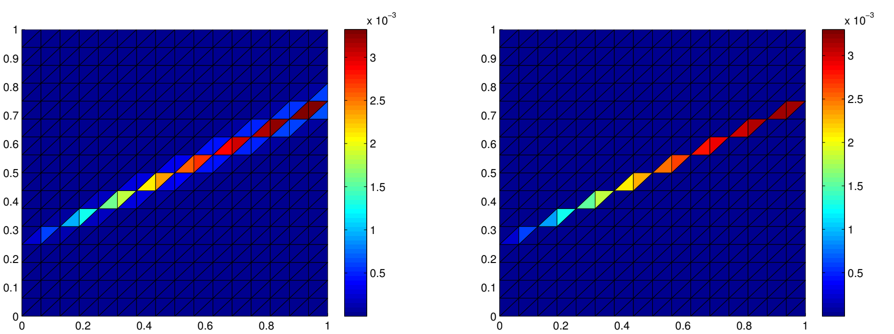

FIGURE 6. The element-wise L_2-error for $\varepsilon = 0$ and $h = 1/16$ in u^h for Example 4.4 obtained by our Petrov-Galerkin discretisation, and that in the $L_2(\Omega)$-orthogonal projection of the solution u onto the space of discontinuous piecewise linears.

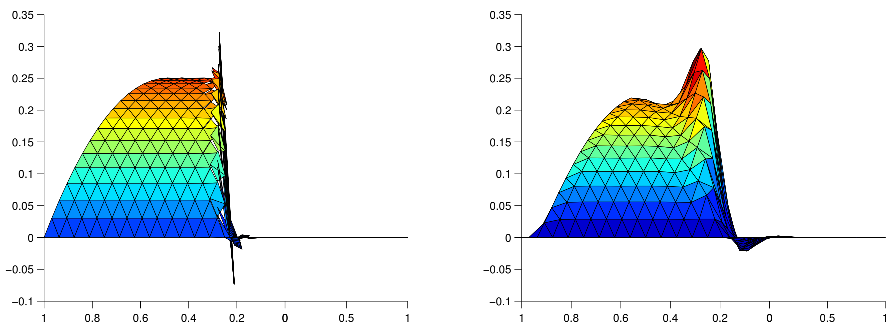

FIGURE 7. Approximate solution for Example 4.4, $\varepsilon = 10^{-6}$ and $h = 1/16$ with our Petrov-Galerkin method (left), and with SUPG (right).

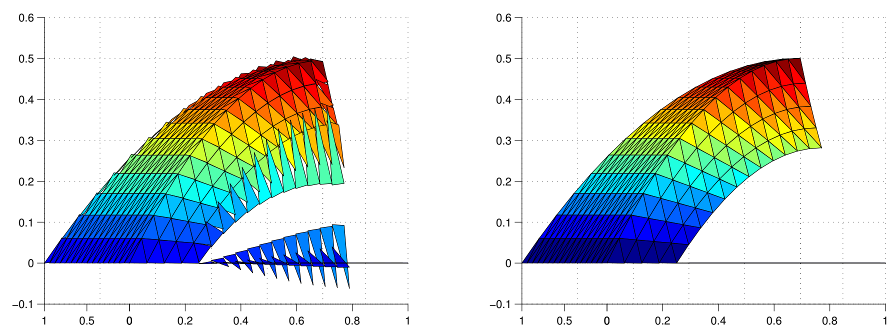

FIGURE 8. Approximate solution by our Petrov-Galerkin method for Example 4.5, $h = 1/16$ and $\varepsilon = 10^{-4}$ (left) or $\varepsilon = 0$ (right).

$L_2(\Omega)^n$ for the auxiliary variable $\boldsymbol{\sigma} := \sqrt{\varepsilon}\nabla u$. Equipping V with the optimal test norm, given $U^h \subset U$ the Petrov-Galerkin discretisation with optimal test space V^h yields the best approximation to the solution from U^h in the norm on U. We arrived at an implementable method by replacing this V^h by its V-orthogonal projection onto an ε-independent choice of a test search space, that has a dimension that is proportional to that of U^h. The numerical approximations for u turn out to be very close to the $L_2(\Omega)$-best approximations from U^h, uniformly in $\varepsilon \geq 0$. Our Petrov-Galerkin discretization requires solving a saddle-point system of a size that is a (moderate) multiple of the the dimension of the trial space, and so it is more costly to implement than a classical method as SUPG. For small ε, the much more accurate results justify the additional cost.

REFERENCES

[Bar70] C. Bardos. Problèmes aux limites pour les équations aux dérivées partielles du premier ordre à coefficients réels; théorèmes d'approximation; application à l'équation de transport. *Ann. Sci. École Norm. Sup. (4)*, 3:185–233, 1970.

[BH82] A.N. Brooks and Th. J. R. Hughes. Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations. *Comput. Methods Appl. Mech. Engrg.*, 32(1-3):199–259, 1982. FENOMECH '81, Part I (Stuttgart, 1981).

[BM84] J. W. Barrett and K. W. Morton. Approximate symmetrization and Petrov-Galerkin methods for diffusion-convection problems. *Comput. Methods Appl. Mech. Engrg.*, 45(1-3), 1984.

[BMS02] C.L. Bottasso, S. Micheletti, and R. Sacco. The discontinuous Petrov-Galerkin method for elliptic problems. *Comput. Methods Appl. Mech. Engrg.*, 191(31):3391–3409, 2002.

[BPV00] J.H. Bramble, J.E. Pasciak, and P.S. Vassilevski. Computational scales of Sobolev norms with application to preconditioning. *Math. Comp.*, 69(230):463–480, 2000.

[BS13] D. Broersen and R.P. Stevenson. A Petrov-Galerkin discretization with optimal test space of a mild-weak formulation of convection-diffusion equations in mixed form. Technical report, KdVI, UvA Amsterdam, June 2013. Accepted for publication in *IMA J. Numer. Anal.*

[Cao97] T. Cao. Hierarchical basis methods for hypersingular integral equations. *IMA J. Numer. Anal.*, 17(4):603–619, 1997.

[CDW12] A. Cohen, W. Dahmen, and G. Welper. Adaptivity and variational stabilization for convection-diffusion equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46:1247–1273, 2012.

[CES00] A. Cohen, L.M. Echeverry, and Q. Sun. Finite element wavelets. Technical report, Laboratoire d'Analyse Numérique, Université Pierre et Marie Curie, 2000.

[DG11] L. Demkowicz and J. Gopalakrishnan. A class of discontinuous Petrov-Galerkin methods. II. Optimal test functions. *Numer. Methods Partial Differential Equations*, 27(1):70–105, 2011.

[DH13] L. Demkowicz and N. Heuer. Robust DPG method for convection-dominated diffusion problems. *SIAM J. Numer. Anal.*, 51(5):2514–2537, 2013.

[DHSW12] W. Dahmen, C. Huang, Ch. Schwab, and G. Welper. Adaptive Petrov-Galerkin methods for first order transport equations. *SIAM J. Numer. Anal.*, 50(5):2420–2445, 2012.

[ES10] H. Egger and J. Schöberl. A hybrid mixed discontinuous Galerkin finite-element method for convection-diffusion problems. *IMA J. Numer. Anal.*, 30(4):1206–1234, 2010.

[For77] M. Fortin. An analysis of the convergence of mixed finite element methods. *RAIRO Anal. Numér.*, 11(4):341–354, iii, 1977.

[GQ14] J. Gopalakrishnan and W. Qiu. An analysis of the practical DPG method. *Math. Comp.*, 83(286):537–552, 2014.

[Kat60] T. Kato. Estimation of iterated matrices, with application to the von Neumann condition. *Numer. Math.*, 2:22–29, 1960.

[LM72] J.-L. Lions and E. Magenes. *Non-homogeneous boundary value problems and applications. Vol. I.* Springer-Verlag, New York, 1972. Translated from the French by P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 181.

[Xu92] J. Xu. Iterative methods by space decomposition and subspace correction. *SIAM Rev.*, 34:581–613, 1992.

[XZ03] J. Xu and L. Zikatanov. Some observations on Babuška and Brezzi theories. *Numer. Math.*, 94(1):195–202, 2003.

[ZMD+11] J. Zitelli, I. Muga, L. Demkowicz, J. Gopalakrishnan, D. Pardo, and V. M. Calo. A class of discontinuous Petrov-Galerkin methods. Part IV: the optimal test norm and time-harmonic wave propagation in 1D. *J. Comput. Phys.*, 230(7):2406–2432, 2011.

Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands

E-mail address: dirkbroersen@gmail.com, r.p.stevenson@uva.nl