Computers and Mathematics with Applications 95 (2021) 41–66

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

The double adaptivity paradigm
(How to circumvent the discrete inf–sup conditions of Babuška and Brezzi)

Leszek Demkowicz [a,*], Thomas Führer [b], Norbert Heuer [b], Xiaochuan Tian [c,1]

[a] Oden Institute, The University of Texas at Austin, United States of America
[b] Dept. of Mathematics, Pontificia Universidad Catolica de Chile, Chile
[c] Dept. of Mathematics, The University of Texas at Austin, United States of America

ARTICLE INFO

Article history:
Available online 3 November 2020

ABSTRACT

We present an efficient implementation of the double adaptivity algorithm of Cohen et al. (2012) within the setting of the Petrov–Galerkin method with optimal test functions. We apply this method to the ultraweak variational formulation of a general linear variational problem discretized with the standard Galerkin finite element method. As an example, we demonstrate the feasibility of the method in the context of the convection-dominated diffusion problem. The presented ideas, however, apply to virtually any well-posed system of first-order partial differential equations, including singular perturbation problems.

Published by Elsevier Ltd.

1. Introduction

The paper is devoted to an implementation of the double adaptivity algorithm of Cohen, Dahmen and Welper [1] within the setting of the ideal and practical Petrov–Galerkin (PG) methods with optimal test functions. The 'double adaptivity paradigm' refers to adaptive refinements of both trial and test spaces. Whereas adapting the trial space aims at improving the approximation of the solution, adapting the test space secures the stability of the underlying discrete problem. The main contribution of the paper is the use of the classical duality theory to estimate the difference between the solutions (and the corresponding residuals) delivered by the ideal and practical PG methods. This strategy represents a departure from the original methodology of Cohen et al. see Remark 3 for details.

Duality pairings. Let U, V be two Hilbert spaces. A bilinear (sesquilinear in the complex case) form $b(u, v)$, $u \in U$, $v \in V$, is called a *duality pairing* if the following relations hold:

$$\|u\|_U = \|b(u,\cdot)\|_{V'} = \sup_{v\in V} \frac{|b(u,v)|}{\|v\|_V} \quad \text{and} \quad \|v\|_V = \|b(\cdot,v)\|_{U'} = \sup_{u\in U} \frac{|b(u,v)|}{\|u\|_U}\,. \tag{1.1}$$

In particular, the duality pairing is *definite*, i.e.,

$$b(u,v) = 0 \quad \forall v \in V \quad \Rightarrow \quad u = 0 \qquad \text{and} \qquad b(u,v) = 0 \quad \forall u \in U \quad \Rightarrow \quad v = 0\,.$$

* Corresponding author.
E-mail address: leszek@oden.utexas.edu (L. Demkowicz).
1 Current address: Dept. of Mathematics, University of California at San Diego, United States of America.

https://doi.org/10.1016/j.camwa.2020.10.002
0898-1221/Published by Elsevier Ltd.

The definition is motivated with the standard duality pairing, where $V = U'$, $b(u, v) = \langle u, v\rangle := v(u)$, and the (induced) norm in the dual space is defined by:

$$\|v\|_{U'} := \sup_{u\in U} \frac{|\langle u, v\rangle|}{\|u\|_U} .$$

For non-trivial examples of duality pairings for trace spaces of the exact sequence energy spaces, see [2]. As in the case of the classical duality pairing, any definite bilinear (sesquilinear) form *can be made into a duality pairing* if we equip V with the norm induced by the norm on U or, vice versa, the space U with the norm induced by the norm on V. In other words, if we equip V with the norm induced by the norm $\|u\|_U$ and use it to induce a norm on U, we recover the original norm on U,

$$\|v\|_V := \sup_{u\in U} \frac{|b(u, v)|}{\|u\|_U} \qquad \Rightarrow \qquad \sup_{v\in V} \frac{|b(u, v)|}{\|v\|_V} = \|u\|_U .$$

Petrov–Galerkin (PG) method with optimal test functions. Consider an abstract variational problem:

$$\begin{cases} u \in U \\ b(u, v) = l(v) \quad v \in V \end{cases} \tag{1.2}$$

where U is a Hilbert trial space, and V is a Hilbert test space, $l \in V'$, and $b(u, v)$ is a bilinear (sesquilinear) form satisfying the inf–sup condition,

$$\inf_{u\in U} \sup_{v\in V} \frac{|b(u, v)|}{\|u\|_U \|v\|_V} \geq \gamma > 0 \qquad \Leftrightarrow \qquad \sup_{v\in V} \frac{|b(u, v)|}{\|v\|_V} \geq \gamma \|u\|_U \quad \forall u \in U .$$

For simplicity, we will assume that b is definite, i.e. ,

$$b(u, v) = 0 \quad \forall u \in U \qquad \Rightarrow \qquad v = 0 .$$

By the Banach–Babuška–Nečas Theorem (see [3], Thm. 6.6. (1) the problem is well-posed, i.e. it possesses a unique solution u that depends continuously on l.

The Petrov–Galerkin method assumes existence of finite-dimensional trial and test spaces,

$$U_h \subset U, \quad V_h \subset V, \quad \dim U_h = \dim V_h ,$$

and formulates a finite-dimensional *approximate problem*:

$$\begin{cases} u_h \in U_h \\ b(u_h, v_h) = l(v_h) \quad v_h \in V_h . \end{cases} \tag{1.3}$$

With the approximate trial and test spaces of equal dimension, the approximate problem translates into a system of linear algebraic equations to solve. If the choice of U_h is dictated by *approximability*, i.e. using functions that can approximate the exact solution well, the approximate test space is usually selected ad hoc. In particular, the selection of the test space is often completely uninformed about the choice of the test norm. This is especially the case when $V = U$, where the choice $V_h = U_h$, leading to the classical Bubnov–Galerkin method, is feasible.

The situation is completely different in the approach proposed by Cohen, Dahmen and Welper [1] where the original problem is embedded into a mixed problem,

$$\begin{cases} \psi \in V , u \in U \\ (\psi, v)_V + b(u, v) \quad = l(v) \quad v \in V \\ b(\delta u, \psi) \quad = 0 \quad \delta u \in U . \end{cases} \tag{1.4}$$

The function $\psi \in V$ is identified as the *Riesz representation of the residual*,

$$(\psi, v)_V = l(v) - b(u, v) \quad v \in V ,$$

and, on the continuous level, is zero. Clearly, both formulations (1.2) and (1.4) deliver the same solution u.

This is no longer true on the approximate level. The *Ideal Petrov–Galerkin Method with Optimal Test Functions* seeks an approximate solution $\widetilde{u}_h \in U_h$ along with the corresponding exact (Riesz representation of) residual $\psi^h \in V$ that solves the semi-discrete mixed problem:

$$\begin{cases} \psi^h \in V , \widetilde{u}_h \in U_h \\ (\psi^h, v)_V + b(\widetilde{u}_h, v) \quad = l(v) \quad v \in V \\ b(\delta u_h, \psi^h) \quad = 0 \quad \delta u_h \in U_h . \end{cases} \tag{1.5}$$

The name of the method refers to the fact that the mixed problem is equivalent to the original PG scheme with approximate test space $V_h = TU_h$ where

$$T : U \to V , \qquad (Tu, v)_V = b(u, v) \quad v \in V .$$

In other words, $T = R_V^{-1}B$ where $B : U \to V'$ is the linear operator generated by form $b(u, v)$, and R_V is the Riesz operator generated by the test inner product $(\psi, v)_V$. The ideal PG method with optimal test functions delivers the orthogonal projection $\widetilde{u}_h$ in the norm induced by the test norm (called the *energy norm* in the original contribution [4]). Note two critical points: (a) the approximate solution $\widetilde{u}_h$ comes with the residual ψ^h which provides a perfect a-posteriori error estimate enabling adaptivity in the trial space, (b) if we use the test norm induced by the trial norm, the ideal PG scheme will deliver the orthogonal projection in the trial norm that now coincides with the "energy norm" [5]. We shall call the test norm induced by the trial norm on U the *optimal test norm* and denote it by $\|v\|_{V_{\text{opt}}}$.

For obvious reasons, we cannot compute with the ideal PG method. We need to approximate space V with some finite-dimensional subspace $V_h \subset V$ as well. The ultimate approximate problem reads as follows.

$$\begin{cases} \psi_h \in V_h\,, u_h \in U_h \\ (\psi_h, v_h)_V + b(u_h, v_h) & = l(v_h) \quad v_h \in V_h \\ b(\delta u_h, \psi_h) & = 0 \quad\quad \delta u_h \in U_h\,. \end{cases} \tag{1.6}$$

This is the *Practical PG Method with Optimal Test Functions*. Brezzi's theory tells us that we have to satisfy now two discrete inf–sup conditions. The *inf–sup in kernel* is trivially satisfied because of the presence of the test inner product. The discrete Ladyzhenskaya–Babuška–Brezzi (LBB) condition,

$$\sup_{v_h \in V_h} \frac{|b(u_h, v_h)|}{\|v_h\|_V} \geq \gamma \|u_h\|_U\,, \quad u_h \in U_h\,,$$

coincides with the discrete Babuška condition for the original problem but *it is much easier now to satisfy as we can employ test spaces of larger dimension*:

$$\dim V_h \gg \dim U_h\,.$$

Remark 1. The *Discontinuous* Petrov–Galerkin method [4] employs variational formulations with *discontinuous* (broken, product) test spaces, and the standard way to guarantee the satisfaction of the discrete LBB condition is to use *enriched* test spaces with order $r = p + \Delta p$ where p is the polynomial order of the trial space. With broken test spaces, the Gram matrix corresponding to the inner product becomes block diagonal, and the (Riesz representation of) residual ψ_h can be eliminated at the element level. The satisfaction of the discrete LBB condition can be assured by constructing *local* Fortin operators [6–9]. For standard elliptic problems (and simplicial meshes), we arrive typically at the condition $\Delta p \geq N$ for N-dimensional problems. The local construction of the Fortin operator is equivalent to the satisfaction of the discrete LBB condition *on the element level* which in turn implies the satisfaction of the global condition. Alternatively, one can construct global Fortin operators that yield much sharper estimates for a minimum enrichment Δp but the proofs get much more technical [10]. ■

Double adaptivity. The revolutionary idea of Cohen, Dahmen and Welper [1] was to propose to determine an optimal discrete test space V_h using adaptivity. After all, the fully discrete mixed problem (1.6) is supposed to be an approximation of the semi-discrete mixed problem (1.5). Both problems share the same discrete trial space U_h and the task is to determine a good approximation $\psi_h \in V_h$ to the ideal $\psi^h \in V$ in terms of the test norm. This, as we will show later, will guarantee that the corresponding ultimate discrete solution $u_h \in U_h$ approximates well the ideal discrete solution $\widetilde{u}_h \in U_h$ as well. We now formulate the double adaptivity algorithm in the context of Finite Element (FE) discretizations.

Given error tolerances tol_U, tol_V for the trial and test mesh, proceed as follows.

```
Set the initial trial mesh U_h
do
  (re)set the test mesh V_h to coincide with the trial mesh U_h
  do
    solve the problem (1.6) on the current trial and test meshes
    estimate the error ||ψ^h − ψ_h||_V ≤ err_V and compute the norm ||ψ_h||_V
    if err_V/||ψ_h||_V < tol_V exit the inner (test) loop
    adapt the test mesh V_h using element contributions of err_V
  enddo
  compute the trial norm of the solution ||u_h||_U
  if ||ψ_h||_V/||u_h||_U < tol_U STOP
  use the element contributions to ||ψ_h||_V to refine the trial mesh
enddo
```

The main contribution of this paper is the development of a reliable a-posteriori error estimation technique for the inner (test) adaptivity loop based on the classical duality theory which enters the algorithm above as err_V. It is in the context of the duality-based error estimation that the *ultraweak variational formulation* distinguishes itself from other formulations as we hope to communicate in this paper.

Remark 2. By setting the test mesh to coincide with the trial mesh, we mean the mesh and the corresponding data structure. The corresponding trial and test energy spaces may be different, dependent upon the variational formulation. The implementation of the method is challenging. As the logic of the double adaptivity calls for two independent meshes, developing an adaptive code in this context seems to be very non-trivial. We have successfully resolved this problem by using *pointers* in our Fortran 90 *hp* FE codes. With a very little investment, the pointer technology allows for converting an adaptive code supporting one mesh into a code supporting two or more *independent meshes*. ■

Remark 3. Whereas we follow closely the paradigm of double adaptivity from [1], our idea of the error estimation for the inner adaptivity loop is slightly different. Cohen, Dahmen and Welper consider the Riesz representation of the residual corresponding to the solution u_h of the fully discrete problem (1.6),

$$\hat{\psi}_h := R_V^{-1}(l(\cdot) - b(u_h, \cdot)) . \tag{1.7}$$

Note that $\hat{\psi}_h$ does not coincide with our ψ^h corresponding to the semidiscrete problem (1.5). Lemma 3.3 in [1] shows that, if

$$\|\hat{\psi}_h - \psi_h\|_V \le \delta \|\psi_h\|_V ,$$

with a constant $\delta \in (0, 2)$, then the fully discrete problem (1.6) is stable with a stability constant $4(1-\frac{\delta}{2})^{-2}$. Consequently, the scheme advocated in [1] is based on an a-posteriori error estimation of $\hat{\psi}_h - \psi_h$, the error corresponding to the Galerkin approximation of problem (1.7). In contrary, we focus on estimating $\psi^h - \psi_h$ which involves comparing the semi-discrete mixed problem (1.5) and the fully discrete mixed problem (1.6). ■

Scope of this paper. We hope to convince the reader that the presented ideas are very general and can be implemented in any space dimension but, in this work, we illustrate them mostly with a model *convection dominated diffusion* (the *confusion*) problem. We do strive though for the range of small viscosities all the way to $\epsilon = 10^{-7}$, a value relevant for compressible Navier–Stokes equations. Building a fully adaptive methodology for this class of problems with a double precision (only) has always been a challenge. In Section 2 we introduce the model confusion problem and review the concept of the ultraweak variational formulation along with the optimal and quasi-optimal test norms. We develop then the main contribution of this paper — an a-posteriori error estimation for the inner adaptivity loop based on the duality theory. The corresponding 1D and 2D numerical examples are presented in Section 3, and we finish with a discussion in Section 4. The main body of the paper is complemented with two appendices. In Appendix A we present a 1D stability analysis for the first order system of Broersen and Stevenson [11,12], and in Appendix B we present an analogous theory with corresponding numerical experiments for the classical variational formulation of the 1D version of the model problem.

2. Ultraweak variational formulation and a duality–based a-posteriori error estimation for the inner adaptivity loop

In this section we develop a duality based error estimation methodology for the *ultraweak (UW)* variational formulation of any well-posed system of first-order Partial Differential Equations (PDEs). As mentioned in the Introduction, we shall use the convection-dominated diffusion (the 'confusion') problem as a motivating and illustrating example. Contrary to the classical formulation discussed in Appendix B, the ultraweak formulation allows for the computation of the optimal test norm in any space dimension.

The derivation of the UW formulation starts by rewriting the original problem as a system of first order equations. This involves introducing a new variable and can be done in more than one way. In these notes, we will use a formulation related to the one advocated by Broersen and Stevenson [11,12]. Consider a multidimensional version of the confusion problem,

$$\begin{cases} u &= 0 \quad \text{on } \Gamma \\ -\epsilon \Delta u + \beta \cdot \nabla u &= f \quad \text{in } \Omega . \end{cases} \tag{2.8}$$

We begin by rewriting the second order problem as a system of first order equations,

$$\begin{cases} u &= 0 \quad \text{on } \Gamma \\ \sigma - \epsilon^{\frac{1}{2}} \nabla u &= 0 \quad \text{in } \Omega \\ -\epsilon^{\frac{1}{2}} \operatorname{div} \sigma + \beta \cdot \nabla u &= f \quad \text{in } \Omega . \end{cases}$$

The second equation defines the auxiliary variable — a scaled viscous flux.

We can rewrite the system using the formalism of closed operators theory. We begin by introducing the first order operator:

$$\begin{aligned} &\mathsf{u} := (\sigma, u) \in D(A) := H(\operatorname{div}, \Omega) \times H_0^1(\Omega) \subset (L^2(\Omega))^N \times L^2(\Omega) \\ &A : D(A) \to (L^2(\Omega))^N \times L^2(\Omega) \\ &A\mathsf{u} = A(\sigma, u) := (\sigma - \epsilon^{\frac{1}{2}} \nabla u, -\epsilon^{\frac{1}{2}} \operatorname{div} \sigma + \beta \cdot \nabla u) . \end{aligned} \tag{2.9}$$

Note that domain $D(A)$ coincides with the subspace of the standard graph energy space, consisting of functions satisfying the boundary condition (BC),

$$D(A) = \{\mathsf{u} = (\sigma, u) \in (L^2(\Omega))^N \times L^2(\Omega) \, : \, A\mathsf{u} \in (L^2(\Omega))^N \times L^2(\Omega), \; u = 0 \text{ on } \Gamma\} \, .$$

The L^2-adjoint A^* of operator A is defined on the same domain, and it is given by:

$$\begin{array}{l} \mathsf{v} := (\tau, v) \in D(A^*) = D(A) \subset (L^2(\Omega))^N \times L^2(\Omega) \\ A^* \, : \, D(A^*) \to (L^2(\Omega))^N \times L^2(\Omega) \, , \\ A^*\mathsf{v} = A^*(\tau, v) = (\tau + \epsilon^{\frac{1}{2}} \nabla v, \epsilon^{\frac{1}{2}} \operatorname{div} \tau - \operatorname{div}(\beta v)) \end{array} \tag{2.10}$$

We shall not differentiate in notation between the formal adjoint of operator A and its actual L^2-adjoint whose definition involves identifying its domain. It should be clear from the context, whether by A^* we mean the formal adjoint or the actual L^2-adjoint operator.

We can write now the precise *strong formulation* of our problem:

$$\begin{cases} \mathsf{u} \in D(A) \\ A\mathsf{u} = \mathsf{f} \end{cases}$$

where $\mathsf{f} = (0, f)$.

Multiplying the equation with a test function $\mathsf{v} = (\tau, v) \in D(A^*)$ and integrating by parts, we obtain the UW formulation:

$$\begin{cases} \mathsf{u} \in (L^2(\Omega))^N \times L^2(\Omega) \\ (\mathsf{u}, A^*\mathsf{v}) = (\mathsf{f}, \mathsf{v}) \quad \mathsf{v} \in D(A^*) \, . \end{cases} \tag{2.11}$$

Computing the optimal test norm for the UW formulation is straightforward,

$$\|\mathsf{v}\|_{V_{\text{opt}}} = \sup_{\mathsf{u}} \frac{|(\mathsf{u}, A^*\mathsf{v})|}{\|\mathsf{u}\|} = \|A^*\mathsf{v}\| \, . \tag{2.12}$$

Above, and throughout the paper, the L^2-norms will be denoted with $\|\cdot\|$ without any additional symbol.

The optimal test norm coincides simply with the *adjoint norm*. Note that the result is independent of considered BCs. The ideal PG method with this norm delivers the L^2-projection of the exact solution. For reasons that will become clear in a moment, we will compute with a related *graph adjoint norm*:

$$\|\mathsf{v}\|^2_{V_{\text{qopt}}} = \|A^*\mathsf{v}\|^2 + \alpha \|\mathsf{v}\|^2 \tag{2.13}$$

with a scaling coefficient α. We have frequently called it the *quasi-optimal test norm*, hence the notation.

Remark 4. If the closed operator A defined in (2.9) whose definition depends upon ϵ, is bounded below with a constant β independent of ϵ,[2] the adjoint graph norm with $\alpha = 1$ is *robustly* equivalent with the adjoint norm. We no longer deliver the L^2-projection but a solution close to it, *uniformly in* ϵ. If β depends upon ϵ, we can still retain the robust equivalence of the adjoint and the adjoint graph norms by using a scaling constant α of order β^2. We have been able to show that, for the 1D confusion problem, the Broersen–Stevenson operator is indeed bounded below robustly in ϵ, see Appendix A. ■

2.1. Duality theory

We discuss now the main idea of this contribution — the a-posteriori error estimation and adaptivity for the inner loop problem based on the classical duality theory [13].

Let H_A, H_{A^*} be the energy graph spaces associated with operator A and its L^2-adjoint A^*,

$$\begin{array}{ll} H_A(\Omega) & := \{\mathsf{u} \in L^2(\Omega) \, : \, A\mathsf{u} \in L^2(\Omega)\} \\ H_{A^*}(\Omega) & := \{\mathsf{v} \in L^2(\Omega) \, : \, A^*\mathsf{v} \in L^2(\Omega)\} \, . \end{array}$$

Let C be the boundary operator resulting from integration by parts,

$$\begin{array}{l} (A^*\mathsf{v}, \mathsf{u}) = (\mathsf{v}, A\mathsf{u}) + \langle \mathsf{v}, C\mathsf{u} \rangle, \qquad \mathsf{v} \in D(A^*), \; \mathsf{u} \in H_A(\Omega), \\ D(A) = \{\mathsf{u} \in H_A(\Omega) \, : C\mathsf{u} = 0\} \, . \end{array}$$

In our case,

$$\langle \mathsf{v}, C\mathsf{u} \rangle = {}_{-\frac{1}{2}}\langle \tau_n, \epsilon^{\frac{1}{2}} u \rangle_{\frac{1}{2}}$$

where ${}_{-\frac{1}{2}}\langle \cdot, \cdot \rangle_{\frac{1}{2}}$ denotes the duality pairing between $H^{-1/2}(\Gamma)$ and $H^{1/2}(\Gamma)$, and $\tau_n = \tau \cdot n$ with n denoting the exterior normal unit vector.

[2] Recall that the Closed Range Theorem for Closed Operators implies that adjoint A^* is then bounded below with the same constant.

We begin by noticing that the semi-discrete mixed problem (1.5) for the ultraweak formulation with the quasi-optimal adjoint graph test norm, is equivalent to the constrained minimization (primal) problem:

$$\inf_{\substack{\psi\,\in D(A^*)\\ A^*\psi\,\in U_h^\perp}} \underbrace{\frac{1}{2}\|A^*\psi\|^2+\alpha\frac{1}{2}\|\psi\|^2-(\mathrm{f},\psi)}_{=:J(\psi)} \tag{2.14}$$

where $U_h^\perp$ denotes the $L^2(\Omega)$-orthogonal complement of the trial space U_h. Introducing the auxiliary variable,

$$\sigma=A^*\psi\,,$$

we observe that, due to the density of H_A in $(L^2(\Omega))^{N+1}$, we have:

$$\sup_{\phi\in H_A}(A^*\psi-\sigma,\phi)=\sup_{\phi\in(L^2(\Omega))^{N+1}}(A^*\psi-\sigma,\phi)=\begin{cases}0 & \text{if } A^*\psi=\sigma\\ +\infty & \text{otherwise.}\end{cases}$$

The minimization problem is thus equivalent to the saddle-point problem,

$$\begin{aligned}
&\inf_{\substack{\psi\,\in D(A^*)\\ A^*\psi\,\in U_h^\perp}} \underbrace{\frac{1}{2}\|A^*\psi\|^2+\alpha\frac{1}{2}\|\psi\|^2-(\mathrm{f},\psi)}_{J(\psi)}\\
&=\inf_{\substack{\sigma\,\in (L^2(\Omega))^{N+1}\\ \sigma\,\in U_h^\perp}}\ \inf_{\psi\in D(A^*)}\sup_{\phi\in H_A}\left\{\frac{1}{2}\|\sigma\|^2+\alpha\frac{1}{2}\|\psi\|^2-(\mathrm{f},\psi)+(A^*\psi-\sigma,\phi)\right\}\\
&=\inf_{\substack{\sigma\,\in (L^2(\Omega))^{N+1}\\ \sigma\,\in U_h^\perp}}\ \inf_{\psi\in D(A^*)}\sup_{\phi\in H_A}\left\{\frac{1}{2}\|\sigma\|^2+\alpha\frac{1}{2}\|\psi\|^2-(\mathrm{f},\psi)+(\psi,A\phi)+\langle\psi,C\phi\rangle-(\sigma,\phi)\right\}=(*)
\end{aligned} \tag{2.15}$$

At this point, we are ready to trade the inf sup for the sup inf,

$$(*)\quad\geq\sup_{\phi\in H_A}\inf_{\substack{\sigma\,\in (L^2(\Omega))^{N+1}\\ \sigma\,\in U_h^\perp}}\ \inf_{\psi\in D(A^*)}\left\{\frac{1}{2}\|\sigma\|^2+\alpha\frac{1}{2}\|\psi\|^2-(\mathrm{f},\psi)+(\psi,A\phi)+\langle\psi,C\phi\rangle-(\sigma,\phi)\right\}=(**)\,.$$

We plan to show *a posteriori* that, in fact, we still have the equality above. The whole point is now that we can compute the two minimization problems *explicitly*. Minimization in σ yields,

$$\sigma=\phi^\perp\quad\Rightarrow\quad\inf_{\substack{\sigma\,\in (L^2(\Omega))^{N+1}\\ \sigma\,\in U_h^\perp}}\ \frac{1}{2}\|\sigma\|^2-(\sigma,\phi)=-\frac{1}{2}\|\phi^\perp\|^2$$

where $\phi^\perp$ stands for the L^2-projection of ϕ onto the L^2-orthogonal complement of $U_h\subset U=L^2(\Omega)^{N+1}$.

Minimizing in $\psi\in D(A^*)$, we get,

$$\alpha\psi=\mathrm{f}-A\phi\quad\Rightarrow\quad\inf_{\psi\in D(A^*)}\left\{\frac{\alpha}{2}\|\psi\|^2-(\mathrm{f}-A\phi,\psi)+\langle\psi,C\phi\rangle\right\}=\begin{cases}-\frac{1}{2\alpha}\|\mathrm{f}-A\phi\|^2 & \text{if } C\phi=0\\ -\infty & \text{otherwise}\,.\end{cases}$$

Consequently, there is no chance for the equality of inf sup to sup inf, unless we restrict the maximization to $\phi\in D(A)$. We could have assumed that from the very beginning but our reasoning shows that the boundary conditions on ϕ defining domain $D(A)$, are a *must*. In the end, we obtain the dual problem:

$$(**)=\sup_{\phi\in D(A)}\underbrace{-\frac{1}{2}\|\phi^\perp\|^2-\frac{1}{2\alpha}\|\mathrm{f}-A\phi\|^2}_{=:J^*(\phi)}=-\inf_{\phi\in D(A)}\underbrace{\frac{1}{2}\|\phi^\perp\|^2+\frac{1}{2\alpha}\|\mathrm{f}-A\phi\|^2}_{=-J^*(\phi)}\,. \tag{2.16}$$

The strict convexity of the primal functional and the strict concavity of the dual functional imply that the minimizers of $J(\psi)$ and $-J^*(\phi)$ exist and are unique.

Simple algebra and one integration by parts show that,

$$2(J(\psi)-J^*(\phi))=\frac{1}{\alpha}\int_\Omega\{\alpha(A^*\psi-\phi^\perp)^2+(\alpha\psi-(\mathrm{f}-A\phi))^2\}\,, \tag{2.17}$$

for *any* ψ from the domain of the primal functional (2.14), and *any* ϕ from the domain of the dual functional (2.16), i.e. $\psi \in D(A^*), A^*\psi \in U_h^\perp$ and $\phi \in D(A)$. Indeed,

$$\begin{aligned} 2J(\psi) - 2J^*(\phi) &= \|A^*\psi\|^2 + \alpha\|\psi\|^2 - 2(\mathsf{f},\psi) + \|\phi^\perp\|^2 + \tfrac{1}{\alpha}\|\mathsf{f} - A\phi\|^2 \\ &= \|A^*\psi - \phi^\perp\|^2 + \underbrace{2(A^*\psi,\phi^\perp)}_{=2(A^*\psi,\phi)} + \tfrac{1}{\alpha}\|\alpha\psi - (\mathsf{f} - A\phi)\|^2 + 2(\psi, \mathsf{f} - A\phi) - 2(\mathsf{f},\psi) \\ &= \|A^*\psi - \phi^\perp\|^2 + \tfrac{1}{\alpha}\|\alpha\psi - (\mathsf{f} - A\phi)\|^2 . \end{aligned}$$

Notice that the two non-negative terms on the right-hand side correspond precisely to the consistency relations between the solutions to the primal and dual problems established above, i.e.,

$$A^*\psi = \sigma = \phi^\perp \quad \text{and} \quad \alpha\psi = \mathsf{f} - A\phi .$$

If $\psi = \psi^h$, the solution of the primal minimization problem, and $\phi = \phi^h$, the solution of the dual maximization problem, the right-hand side above is equal to zero, i.e. there is no duality gap on the continuous level (we will demonstrate it formally in a moment). This, of course, is essential for using the difference $2(J(\psi_h) - J^*(\phi_h))$ for an a-posteriori error estimate for the approximate solutions ψ_h, ϕ_h to the primal and dual problems.

The solution of the primal minimization problem satisfies the original semi-discrete mixed problem:

$$\begin{cases} \psi^h \in D(A^*),\ \widetilde{\mathsf{u}}_h \in U_h & \\ (A^*\psi^h, A^*\delta\psi) + \alpha(\psi^h,\delta\psi) \quad +(\widetilde{\mathsf{u}}_h, A^*\delta\psi) &= (\mathsf{f},\delta\psi) \quad \delta\psi \in D(A^*) \\ (A^*\psi^h, \delta\mathsf{u}_h) &= 0 \quad\quad\ \ \delta\mathsf{u}_h \in U_h \end{cases} \tag{2.18}$$

where $\widetilde{\mathsf{u}}_h \in U_h$ is the Lagrange multiplier corresponding to the constraint $A^*\psi \in U_h^\perp$ defining the set over which we solve the minimization problem.

The dual problem is equivalent to a double minimization problem:

$$\inf_{\phi\in D(A)} \frac{1}{2}\|A\phi - \mathsf{f}\|^2 + \frac{\alpha}{2}\|\phi^\perp\|^2 = \inf_{\phi\in D(A)} \inf_{\mathsf{w}_h\in U_h} \frac{1}{2}\|A\phi - \mathsf{f}\|^2 + \frac{\alpha}{2}\|\phi - \mathsf{w}_h\|^2$$

which in turn is equivalent to another 'mixed-like' problem,

$$\begin{cases} \phi^h \in D(A),\ \widetilde{\mathsf{w}}_h \in U_h & \\ (A\phi^h, A\delta\phi) + \alpha(\phi^h,\delta\phi) \quad -\alpha(\widetilde{\mathsf{w}}_h,\delta\phi) &= (\mathsf{f}, A\delta\phi) \quad \delta\phi \in D(A) \\ -\alpha(\phi^h,\delta\mathsf{w}_h) \quad\quad\quad\quad\quad\ +\alpha(\widetilde{\mathsf{w}}_h,\delta\mathsf{w}_h) &= 0 \quad\quad\quad\quad \delta\mathsf{w}_h \in U_h\,, \end{cases} \tag{2.19}$$

or, in the strong form,

$$A^*A\phi^h + \alpha(\phi^h - \widetilde{\mathsf{w}}_h) = A^*f \tag{2.20}$$

plus the BC:

$$C^*A\phi^h = C^*\mathsf{f} \quad \Rightarrow \quad \mathsf{f} - A\phi^h \in D(A^*) \tag{2.21}$$

where the boundary operator C^* defines the domain of the adjoint operator A^*.

Remark 5. We emphasize that, contrary to the primal problem, the mixed problem corresponding to the dual problem does not have the structure of a saddle problem but it is a minimization problem. ■

Lemma 1. *Let ψ^h and ϕ^h be the solutions to the primal and dual problem. Then, the right-hand side in identity (2.17) equals zero, i.e. there is no duality gap on the continuous level. Moreover, the Lagrange multiplier $\widetilde{\mathsf{u}}_h$ in (2.18) is equal to $\widetilde{\mathsf{w}}_h$ in (2.19).* ■

Proof. Let ϕ^h be the solution to the dual problem. Use one of the duality relations to define *a function* ψ^h,

$$\psi^h := \frac{1}{\alpha}(\mathsf{f} - A\phi^h) .$$

First of all, ψ^h satisfies the second duality relation. Indeed, Eq. (2.20) implies that

$$A^*\psi^h = \frac{1}{\alpha}(A^*\mathsf{f} - A^*A\phi^h) = \phi^h - \widetilde{\mathsf{w}}_h = (\phi^h)^\perp .$$

Secondly, BC (2.21) implies that $\psi^h \in D(A^*)$. Finally, plugging the function ψ^h and $\mathsf{u}_h := \widetilde{\mathsf{w}}_h$ into the variational formulation (2.18), we obtain,

$$\begin{aligned} (A^*\psi^h, A^*\delta\psi) + (\alpha\psi^h,\delta\psi) + (\widetilde{\mathsf{u}}_h, A^*\delta\psi) &= ((\phi^h)^\perp, A^*\delta\psi) + (\mathsf{f} - A\phi^h,\delta\psi) + (\widetilde{\mathsf{w}}_h, A^*\delta\psi) \\ &= (\phi^h, A^*\delta\psi) - (A\phi^h,\delta\psi) + (\mathsf{f},\delta\psi) \\ &= (\mathsf{f},\delta\psi) . \end{aligned}$$

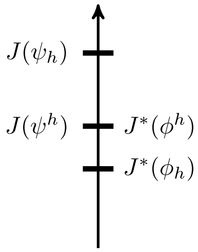

Fig. 1. Concept of the duality-based a-posteriori error estimation.

Consequently, uniqueness of the solution to the primal problem implies that the function ψ^h derived from the duality relations and $\widetilde{\mathrm{u}}_h := \widetilde{\mathrm{w}}_h$ are indeed the solution of the primal problem. ■

Remark 6. Vanishing of the duality gap holds for a general class of strictly convex minimization problems with linear constraints, see [13]. We hope that the elementary proof provided here will facilitate the reading for a non-specialist in convex minimization. Besides, we needed to prove the equality of $\widetilde{\mathrm{u}}_h = \widetilde{\mathrm{w}}_h$. ■

Duality-based a-posteriori error estimation. The main idea of using the difference $J(\psi_h) - J^*(\phi_h)$ for the a-posteriori error estimation is illustrated in Fig. 1. Minimization of the primal functional over a finite-dimensional subspace of the set

$$\{\psi \in D(A^*) \,:\, A^*\psi \in U_h^\perp\}$$

over which the primal minimization problem (2.14) is defined, yields a discrete minimizer ψ_h and the corresponding discrete minimum $J(\psi_h)$ above the exact one, $J(\psi_h) \geq J(\psi^h)$. Similarly, maximization of the dual functional over a finite-dimensional subspace of $D(A)$ over which the dual maximization problem (2.16) is defined, results in a discrete maximizer ϕ_h and the corresponding discrete maximum being smaller than the continuous one, $J^*(\phi_h) \leq J^*(\phi^h)$. Obviously,

$$\left.\begin{aligned} J(\psi_h) - J(\psi^h) \\ J^*(\phi^h) - J^*(\phi_h) \end{aligned}\right\} \leq J(\psi_h) - J^*(\phi_h)\,. \tag{2.22}$$

On the other side, it is well known that, for a quadratic minimization problem,

$$\inf_{u\in V} \underbrace{\frac{1}{2}b(u,u) - l(u)}_{=:\text{total energy}J(u)} \qquad \Leftrightarrow \qquad \begin{cases} u \in V \\ b(u,v) = l(v) \quad v \in V \end{cases} \tag{2.23}$$

where form b is symmetric and positive-definite, we have,

$$\|u - u_h\|_E^2 = 2(J(u_h) - J(u)) \tag{2.24}$$

where u is the minimizer, $u_h \in V$ an arbitrary element, and $\|u\|_E^2 := b(u,u)$ is the energy norm.

Now, the primal problem (2.14) is a quadratic minimization problem over $\psi \in D(A^*), A^*\psi \in U_h^\perp$ with total energy equal to the primal energy, and the energy norm (squared) is given by:

$$\|A^*\psi\|^2 + \alpha\|\psi\|^2\,.$$

Similarly, the dual problem (2.16) is also equivalent to a quadratic minimization problem with the energy norm (squared) given by:

$$\frac{1}{\alpha}\|A\phi\|^2 + \|\phi^\perp\|^2$$

and the total energy functional given by

$$\frac{1}{2}\{\frac{1}{\alpha}\|A\phi\|^2 + \|\phi^\perp\|^2\} - \frac{1}{\alpha}(\mathrm{f}, A\phi)\,.$$

Note this total energy equals minus times the functional defining the dual problem (2.16) modulo the shift $\frac{1}{2\alpha}\|\mathrm{f}\|^2$. Consequently, the difference of total energies in the estimate (2.24) equals $J^*(\phi^h) - J^*(\phi_h)$. In conclusion, putting together the estimate (2.22) and the energy norm estimates (2.24) for the primal and dual problems, we obtain the following a-posteriori error estimates for arbitrary elements $\psi \in D(A^*), A^*(\psi) \in U_h^\perp$ and $\phi \in D(A)$,

$$\left.\begin{aligned} \|A^*(\psi^h - \psi)\|^2 + \alpha\|\psi^h - \psi\|^2 \\ \|(\phi^h)^\perp - \phi^\perp\|^2 + \tfrac{1}{\alpha}\|A(\phi^h - \phi)\|^2 \end{aligned}\right\} \leq 2(J(\psi) - J^*(\phi))\,. \tag{2.25}$$

In particular, these relations hold for the approximate solutions $\psi = \psi_h$ and $\phi = \phi_h$ to the primal and dual problems. Both problems are solved using a conforming FE discretization, i.e. $\psi_h \in D(A^*)$, and $\phi_h \in D(A)$. For the confusion problem, this translates into the use of H^1- and $H(\mathrm{div})$-conforming elements for the two components of ψ_h and ϕ_h. Finally, we recall that the difference between the approximate primal and dual energies can be expressed as the integral of the non-negative consistency terms,

$$2(J(\psi_h) - J^*(\phi_h)) = \frac{1}{\alpha}\int_\Omega \alpha(A^*\psi_h - \phi_h^\perp)^2 + (\alpha\psi_h - (\mathsf{f} - A\phi_h))^2\,. \tag{2.26}$$

The non-negativeness of the terms makes the local, element contributions to estimate (2.26) good candidates for element error indicators.

Remark 7.

1. The idea of using the difference of the primal and dual energies to estimate the error in approximate solutions to the primal and dual problems is very old and known under the name of Prager–Synge *hypercircle method* in the engineering literature. The estimate is *reliable* but not necessarily *efficient*. If the $\inf_\psi J(\psi) = \sup_\phi J^*(\phi)$ is close to $J(\psi_h)$, the estimate may be a good estimate for the dual problem but hopeless for the primal one. Vice versa, if the exact values are close to the discrete maximum, the duality-based estimate will be efficient for the primal problem but not for the dual one. This results in various attempts to make the solution of the dual problem more accurate than that of the primal problem, e.g. by raising the polynomial order, or refining globally the mesh. As we saw no indication for the solution of the dual problem to be less regular than the solution of the primal one, we computed solutions of both problems with the same mesh. Finally, note that the duality-based estimate involves no auxiliary constants and, therefore, it is a good candidate for singular perturbation problems.
2. *Can we pass with $\alpha \to 0$* ? Clearly, for small α, the dual problem approaches the least squares method for the original problem, and the least squares term dominates the estimate. The two problems disconnect, and the duality-based estimate is no longer a meaningful estimate for neither the primal nor the dual problem. This is consistent with the well known fact that the duality theory for linear elastostatics requires the maximization over stress fields satisfying the equilibrium equations. In our case, we would need to maximize over ϕ_h satisfying the equation $A\phi_h = \mathsf{f}$. There is only one such a ϕ_h — the solution to our problem. In conclusion, *we have to compute with finite α*.
3. Is the formulation of the dual problem unique? Hard to say. In principle, we can derive multiple dual problems and use different dual energies for the error estimation. The practicality of the approach relies however on the possibility of deriving the dual problem *analytically* and the feasibility of its discretization. The best known example is the solution of a standard elliptic problem. If we do have a zero order positive term, the formulation and discretization of the dual problem is easy [14]. If we do not, the duality theory is much harder to implement. The same observations hold for the elastostatics problem mentioned above. In this context, the proposed dual problem seems to be rather natural. The trickiness of its derivation relies on the observation that the semi-discrete mixed problem (1.5) is in fact a minimization problem over a constrained set. This opened up the possibility of starting the duality reasoning based on the L^2-setting. This requires the use of an L^2 right-hand side in the ultraweak variational problem (2.11). ■

3. Numerical experiments

3.1. 1D experiments

We consider the following 1D model *confusion* problem.

$$\begin{cases} u(0) = u(1) = 0 \\ -\epsilon u'' + u' = f \quad \text{in } (0,1)\,. \end{cases} \tag{3.27}$$

In all numerical experiments, we use a single example with $f = 1$ for which the exact solution is readily available. All computations presented in this paper are performed with the double adaptivity algorithm outlined in the Introduction. For the ultraweak variational formulation we use the quasi-optimal test norm, i.e. the adjoint graph norm:

$$\|\mathsf{v}\|^2_{V_{\mathrm{qopt}}} := \|A^*\mathsf{v}\|^2 + \alpha\|\mathsf{v}\|^2$$

with $\alpha = 1$. For the error estimate err_V we use the duality-based a-posteriori error estimate (2.26) where ψ_h is the approximate solution of (2.18) and ϕ_h is the approximate solution of (2.19).

We start with a moderate value of $\epsilon = 10^{-2}$ to illustrate the algorithm. We use trial meshes of order p and test meshes of order $p+1$. This is because we use a classical frontal solver with no pivoting; for $p = 1$, and trial and test meshes of equal order, we encountered a zero pivot in the very first element, hence the use of test meshes of one order higher. Note that we are using the language of the exact sequence [15]. By the order of elements we mean always the polynomial degree for the H^1-conforming elements. This means that effectively, for $p = 3$, we approximate σ and u which live in L^2, with piece-wise quadratics. The two components of the residual ψ live in H^1 and, therefore, are approximated with piece-wise quartic elements.

Table 1
UW formulation, $\epsilon = 10^{-2}$. Column 1: Outer loop iteration number. Column 2: Error (residual) estimate for the "trusted" solution. Column 3 and next: the evolution of the inner loop a-posteriori error estimate.

1	50.6	162.4	76.8	35.9	23.6	14.4	9.0	5.6	4.7
2	27.8	106.3	34.3	20.1	10.3	7.2	5.0	2.7	
3	10.9	59.7	12.1	8.5	4.2				
4	2.6	31.0	4.6						
5	0.4	21.4	16.4	9.7	4.3				

Table 2
UW formulation. Number of inner loop iterations for the extreme values of the viscosity constant. Each column represents an outer loop iteration. The star indicates no convergence.

$\epsilon = 10^{-6}$	33	32	32	31	31	31	30	33	47	45	42	39	37	37	38	41	30	14		
$\epsilon = 10^{-7}$	42	41	41	40	40	40	40	40	40	52	57	56	53	50	47	46	45	46	48	*

Our original trial mesh consists of five elements, and the starting test mesh in the inner adaptivity loop is always (re)set to the trial mesh but with elements of one order higher. The tolerance for the outer, and the inner loop adaptivity, tol_U and tol_V, is set to 1 and 5 percent, respectively. We use the Dörfler refinement strategy with 1 and 25 percent factors. This means that we are very conservative with refinements of the trial mesh and try to accelerate the refinements of the test mesh. The inner loop iterations (a total of 9) corresponding to the first outer iteration, are presented in Figs. 2 and 3. Along with the trial meshes, we display the u_h-component of the solution $\mathsf{u}_h = (\sigma_h, u_h)$, the v-component of the solution $\psi_h = (\psi_h^\tau, \psi_h^v)$ of the primal problem (2.18) and the v-component of the solution $\phi_h = (\phi_h^\tau, \phi_h^v)$ of the dual problem (2.19). The solutions seem to change very little but the a-posteriori error estimate evolves from 162 to 4.7 percent of the error, see Table 1. Note that the ultimate discrete solution is *not* the L^2-projection of the exact solution. This is a consequence of using the adjoint graph norm rather than the adjoint norm. The shift between the exact and approximate solutions of the original, primal problem depends upon the value of the coefficient α in the quasi-optimal test norm. As expected, it diminishes with smaller values of α (results not shown). The evolution of the "trusted" trial solutions along with the corresponding resolved residual is shown in Fig. 4. By the "trusted" trial solution we mean the solution obtained with the ultimate test mesh resulting from the inner adaptivity. In order to solve the problem with the requested 1 percent of accuracy, the algorithm has performed five outer loop iterations. The corresponding evolution of the error and the inner loop duality error estimates is shown in Table 1.

Conclusions at this point? (1) The number of the inner loop iterations decreases with the outer loop iterations. (2) The residual for the unresolved solution has a significant variation not only in the boundary layer but also at the inflow. At the end, the residual around the inflow becomes insignificant, note the lack of refinements at the inflow in the last test mesh. Philosophically, we need to think of a new residual after each trial mesh refinement. If we decide to keep the test mesh from the previous inner loop iterations, we need to implement unrefinements as well.

Pushing the code. We have been able to solve the problem for $\epsilon = 10^{-6}$ but we failed for $\epsilon = 10^{-7}$. The number of the inner loop iterations increased significantly with smaller ϵ, and in the end, the inner loop iterations did not converge, see Table 2. We have implemented a number of energy identities which should be satisfied and the code stopped passing those tests. Note that the duality-based estimate *has to decrease* with any mesh refinements. This stopped being the case in the end of the last run. Clearly, we have lost the precision.

We were a bit more lucky using a continuation in ϵ. Starting with $\epsilon = 10^{-2}$, we ran the double adaptivity algorithm. Upon a convergence, we restarted the algorithm with $\epsilon_{\text{new}} = \epsilon_{\text{old}}/2$ and the initial trial mesh obtained from the previous run. Except for the last couple of cases, the number of inner loop iterations dramatically decreased (did not exceed 10) and, in the end, the smallest value of ϵ for which we were able to solve the problem, was $\epsilon = 3.81410^{-8}$.

3.2. Controlling the error in u_h

The *Ideal PG Method with Optimal Test Functions* (equivalent to the semi-discrete mixed problem (1.5) with infinite-dimensional test space) inherits the inf–sup condition from the continuous level. In other words, the operator $B : U \to V'$ generated by the bilinear form $b(u, v)$ is bounded below. This implies that the error $\mathsf{u} - \widetilde{\mathsf{u}}_h$ is controlled by the residual,

$$\gamma \|\mathsf{u} - \widetilde{\mathsf{u}}_h\|_U \le \|l - B\widetilde{\mathsf{u}}_h\|_{V'} = \|\psi^h\|_V .$$

Once the residual converges to zero, so must the error, at the same rate. The inner adaptivity loop guarantees that we approximate the (Riesz representation of) residual ψ^h within a required tolerance with ψ_h. But coming with ψ_h is only the approximation u_h of $\widetilde{\mathsf{u}}_h$. How do we know that u_h converges to $\widetilde{\mathsf{u}}_h$? Can we estimate the error $\widetilde{\mathsf{u}}_h - \mathsf{u}_h$? The problem deals again with a mixed problem albeit somehow special — the space U_h is finite-dimensional. An attempt to use Brezzi's theory makes little sense as it calls for a discrete LBB inf–sup condition whose use we are trying to circumvent.

This is where the duality theory comes to the rescue again. Please recall the critical piece of information from Lemma 1: the ideal approximate solution $\widetilde{\mathsf{u}}_h$ coincides with the L^2-projection $\widetilde{\mathsf{w}}_h$ of ϕ^h- the solution of the dual problem onto U_h. The

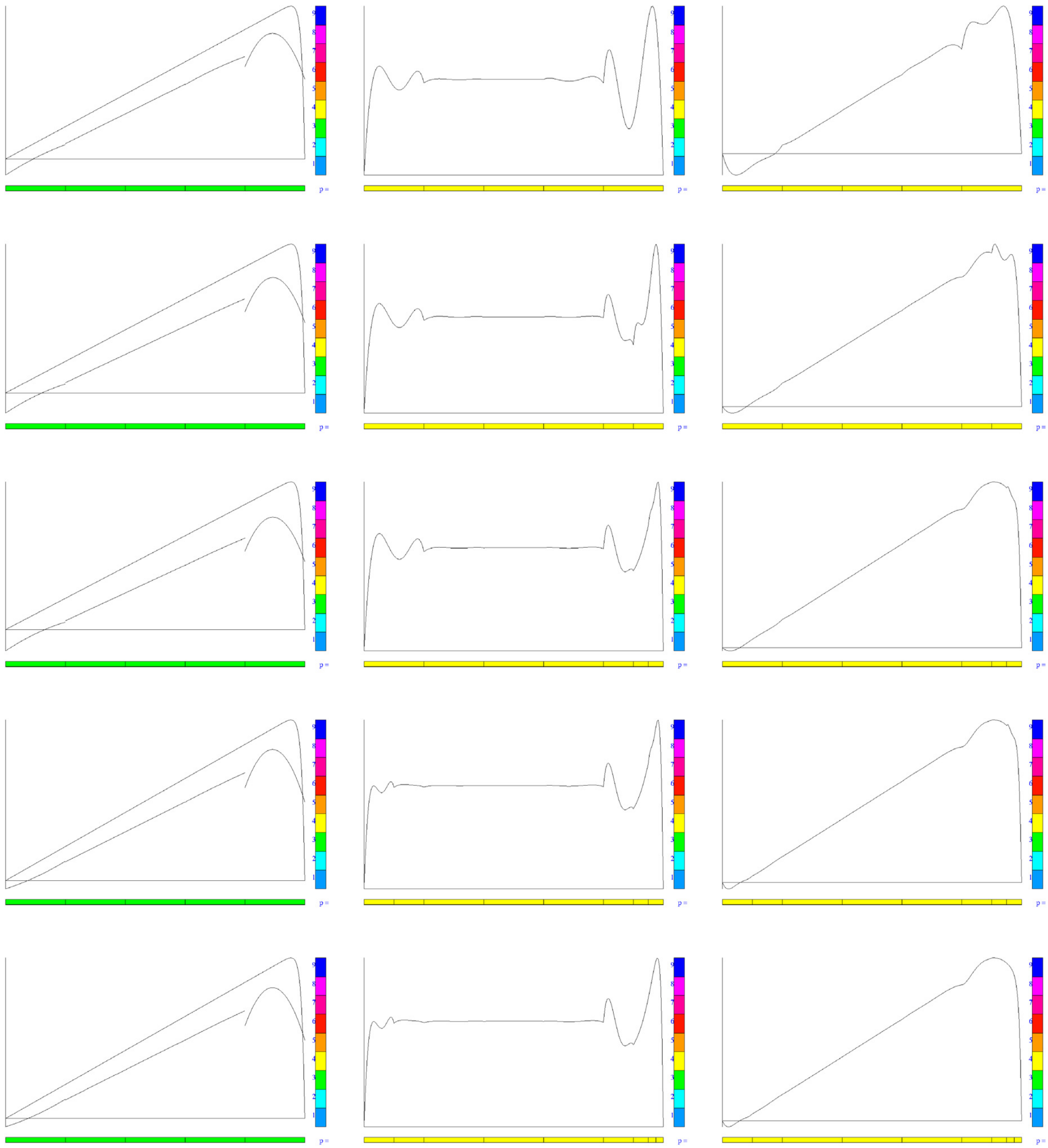

Fig. 2. UW formulation, $\epsilon = 10^{-2}$, inner iterations 1–5 inside the first outer iteration. Left: The evolution of the approximate solution u_h on a trial mesh of five cubic elements corresponding to different test meshes. Middle: The test mesh with the corresponding v component of the approximate residual ψ_h. Right: The test mesh with the corresponding v component of the approximate solution ϕ_h to the dual problem. The column on the right provides the color code for element polynomial orders, $p = 1, 2, \ldots, 9$.

primal problem is a standard (saddle point) mixed problem but the dual problem is a (double) minimization problem. The difference of the approximate primal and dual energies used to estimate the error in the solution to the primal problem, estimates also the error in the solution of the dual problem,

$$\frac{1}{\alpha}\|A(\phi^h - \phi_h)\|^2 + \|(\phi^h)^\perp - \phi_h^\perp\|^2 \le 2(J(\psi_h) - J^*(\phi_h)) =: \text{est}\,,$$

recall estimates (2.25). Operator A is bounded below,

$$\beta\|\phi^h - \phi_h\| \le \|A(\phi^h - \phi_h)\|$$

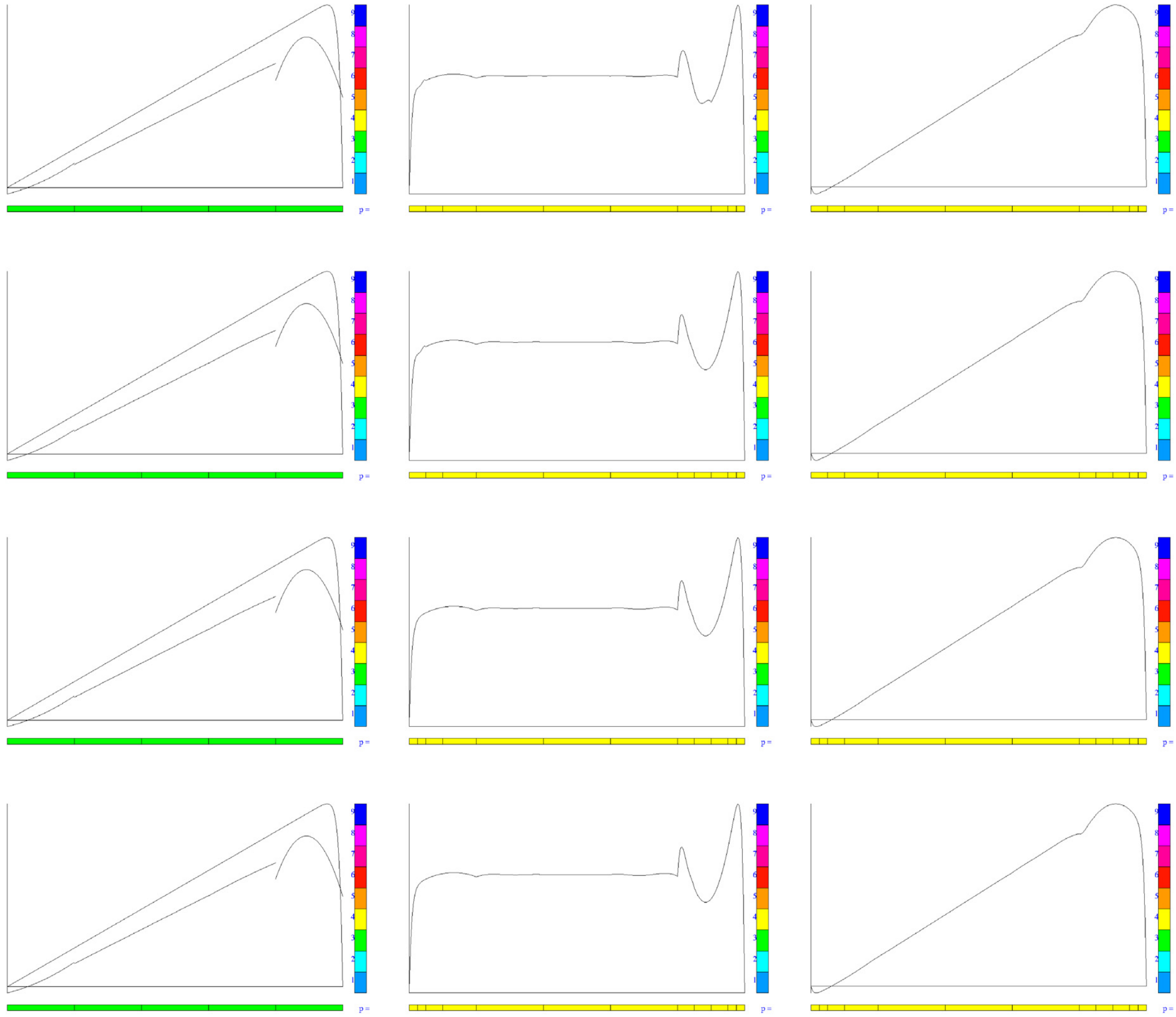

Fig. 3. UW formulation, $\epsilon = 10^{-2}$, inner iterations 6–9 inside the first outer iteration. Left: The evolution of the approximate solution u_h on a trial mesh of five cubic elements corresponding to different test meshes. Middle: The test mesh with the corresponding v component of the approximate residual ψ_h. Right: The test mesh with the corresponding v component of the approximate solution ϕ_h to the dual problem.

which implies that

$$\frac{\beta^2}{\alpha}\|\phi^h - \phi_h\|^2 \le \text{est}\,.$$

This implies the bound for the projection as well,

$$\frac{\beta^2}{\alpha}\|\widetilde{\mathsf{w}}_h - \mathsf{w}_h\| \le \frac{\beta^2}{\alpha}\|\phi^h - \phi_h\|^2 \le \text{est}\,. \tag{3.28}$$

In conclusion, if we believe in proofs, we should use w_h and not u_h as our final (numerical) solution of the problem. To illustrate the point, we present the approximate solution u_h and the projection w_h (second components) at the beginning and at the end of the first inner loop for the problem presented in this section, see Fig. 5. As we can see, with an unresolved residual, the two functions are significantly different. However, once the residual has been resolved (error tolerance = 5%), the two solutions are indistinguishable.

Remark 8. If the boundedness below constant β depends upon ϵ then, unfortunately, bound (3.28) is *not* robust in ϵ, even if we choose α to be of order β^2. ■

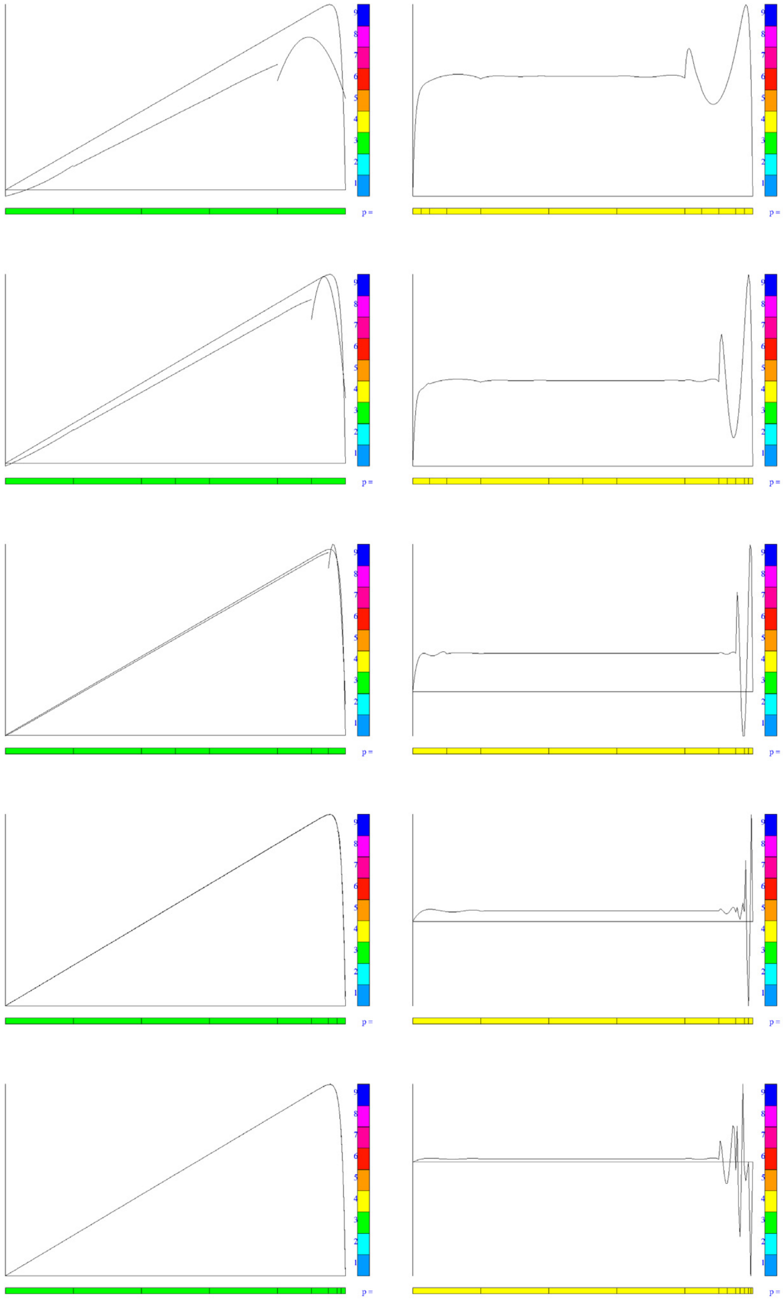

Fig. 4. UW formulation, $\epsilon = 10^{-2}$, outer loop, iterations 1–5. Left: The evolution of the approximate solution u_h. Right: The test mesh with the corresponding resolved v component of the approximate residual ψ_h.

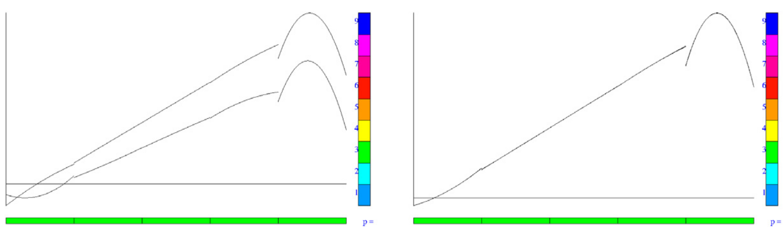

Fig. 5. UW formulation, $\epsilon = 10^{-2}$, first inner loop. Second components of the projection w_h (top), and the approximate solution u_h (bottom). Left: at the beginning of the inner loop. Right: at the end of the loop.

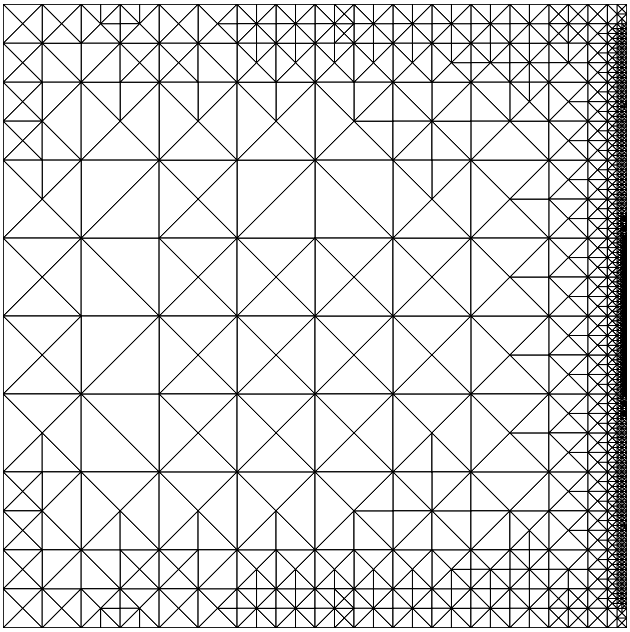

Fig. 6. 2D model problem, $\epsilon = 10^{-2}$. Final trial mesh with 1476 elements.

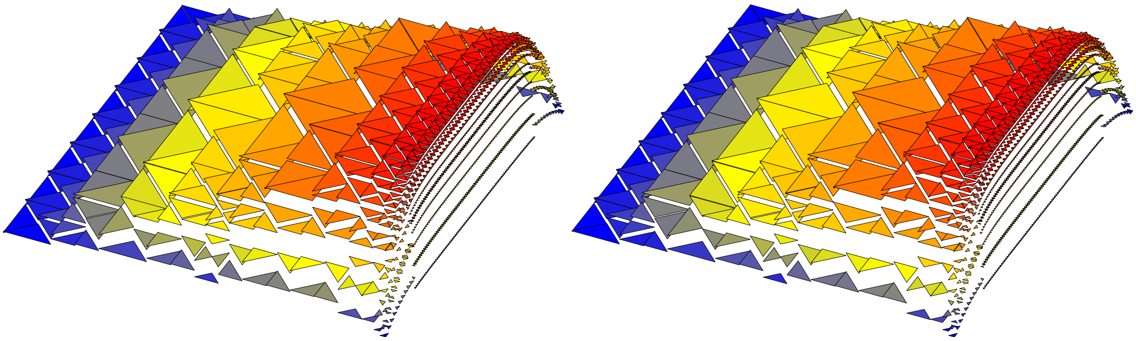

Fig. 7. 2D model problem, $\epsilon = 10^{-2}$. Solution u_h (left) compared to w_h (right).

3.3. Preliminary 2D results

We conclude with a few preliminary computational results in two space dimensions. We consider a 2D analogue of the 1D model problem. The problem is defined in a unit square domain, $\Omega = (0, 1)^2$, with constant advection vector $\beta = (1, 0)$, right-hand side $f = 1$ and homogeneous BCs for u. The problem was solved using the lowest order polynomial spaces, corresponding to the first Nedéléc sequence for $p = 1$ and triangular elements.

Fig. 6 presents the final trial mesh for $\epsilon = 0.01$. The corresponding u-component of the solutions u_h and w_h are presented in Fig. 7 confirming the claim that they converge to each other.

Figs. 8 and 9 present the analogous results for $\epsilon = 10^{-3}$.

Finally, Fig. 10 presents the evolution of the residual at the end of the inner adaptivity loop for different values of $\epsilon = 10^{-2}, \ldots, 10^{-5}$. As expected, a monotone convergence is observed with the asymptotic rates attained later for smaller values of ϵ.

4. Conclusions

The double adaptivity idea is a fascinating idea as it sends a clear message that the (uniform in h) discrete stability conditions of Babuška and Brezzi are sufficient but not necessary for convergence. This was emphasized for the first time by Bänsch, Morin and Nochetto [16] in the context of the Stokes problem and Uzawa's algorithm and, independently, by Cohen, Dahmen, DeVore, Dahlke and Urban, in the context of adaptive wavelet schemes [17,18]. Cohen, Dahmen and

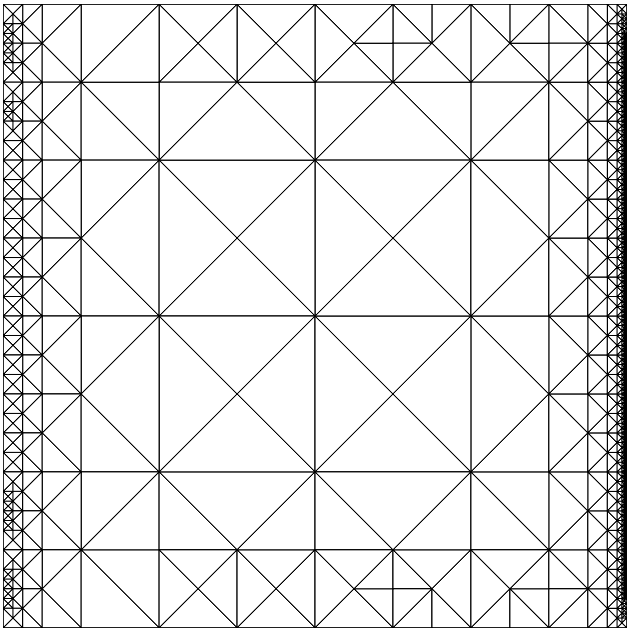

Fig. 8. 2D model problem, $\epsilon = 10^{-3}$. Final trial mesh with 2035 elements.

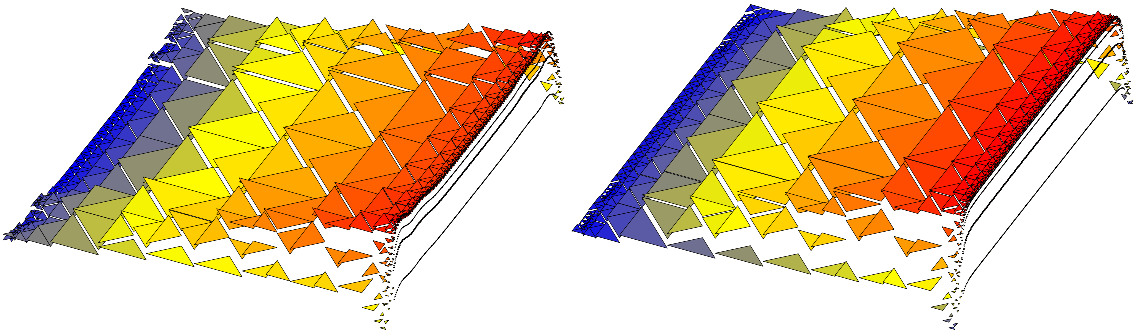

Fig. 9. 2D model problem, $\epsilon = 10^{-3}$. Solution u_h (left) compared to w_h (right).

Welper [1] developed the idea into the mixed problem stabilization of an arbitrary variational problem and applied it to the confusion problem. Similar ideas were used in [19] for the construction of optimal reduced bases.

In the presented numerical examples, the inf–sup condition is satisfied (otherwise, the discrete system of equations would have been singular), and we have good reasons to believe that the corresponding discrete inf–sup constants are independent of h. But, they are definitely *not independent of* ϵ, i.e. the ultimate test mesh delivered by the inner adaptivity loop does not provide a *robust* discretization.

Robustness of the discretization is the common goal in the singular perturbation problems community. For linear problems, it translates into the assumption that both the continuity and discrete inf–sup constants should be independent of ϵ. This is the case for the *ideal Petrov–Galerkin method* based on the UW formulation and properly scaled adjoint graph norm used for the test space, see Remark 4. The inner adaptivity loop delivers the discrete Riesz representation ψ_h of the residual, along with the discrete solution u_h that are sufficiently close to the exact Riesz representation of the residual ψ^h and the corresponding ideal solution $\widetilde{u}_h$, but *it does it for a particular right-hand side* f *only*. For a different right-hand side f, the corresponding residual is different, and the discrete test space obtained for the original f, is no longer optimal for the new f. A test space (mesh) with a robust inf–sup constant, would have delivered semi-optimal results for an arbitrary right-hand side. Consequently, the double adaptivity methodology clearly demonstrates that the robustness *is not necessary* for a successful discretization of singular perturbation problems.

The ultraweak variational formulation allows for determining the optimal and quasi-optimal test norms for *any well-posed* system of first order PDEs. In particular, it circumvents the need for an elaborate stability analysis that we performed for the confusion problem [20,21] that formed a foundation for our DPG methodology based on the so-called *robust test norms*.

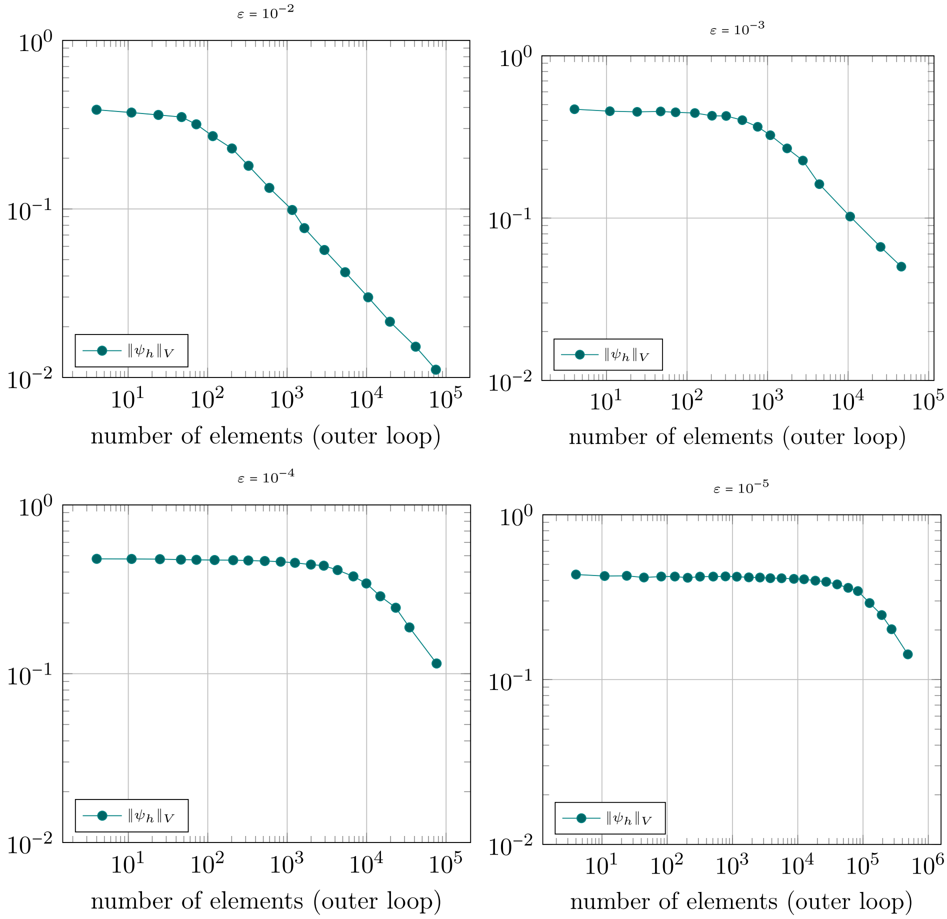

Fig. 10. 2D model problem. Convergence results for $\epsilon = 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$. Residual $\|\psi_h\|_V$ at the final inner iteration.

In the optimal test norm business, we are trading the solution of a convection-dominated problem with diffusion ϵ, for the solution of a reaction-dominated problem with diffusion ϵ^2 using standard Galerkin. Clearly, robustness of the a-posteriori error estimate is the main issue here, see e.g. [22,23] and the literature therein. We have used the duality for error estimation and adaptivity many years ago [14,24,25] but never in the context of singular perturbation problems. The use of duality theory arguments for the a-posteriori error estimation seems to be recently on a rise, see [26] and the literature therein. In this context, the fact that we have been able to solve our 1D reaction-dominated problem with $\epsilon^2 = 10^{-14}$ is really extraordinary.

Finally, we would like to comment on the role of the Riesz representation ψ^h of the residual. With every refinement of the trial mesh, the function ψ^h corresponding to the ideal PG method, changes, and its adaptive resolution leads to a very different ultimate discrete test mesh. In the considered example, for a coarse trial mesh, we observe a significant residual on the inflow that leads to the test mesh refinements there. As we refine the trial mesh in the boundary layer area, the residual on the inflow significantly decreases and there is no need for test mesh refinements there anymore. The moral of the story is that, after any trial mesh refinement, we have to restart the inner adaptivity loop with a brand new initial test mesh. In practice, we choose to restart it with an initial test mesh coinciding with the trial mesh.

On the coding side, this leads to the necessity of supporting two independent mesh data structures. This has been accomplished by using the technology of *pointers in Fortran 90*. Many thanks to Socratis Petrides and Stefan Henneking for teaching us how to use it to generalize our *hp* codes to support multiple meshes.

We hope to report on more multidimensional examples soon.

CRediT authorship contribution statement

Leszek Demkowicz: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Writing - review & editing. **Thomas Führer:** Conceptualization, Methodology, Software, Writing - review & editing. **Norbert Heuer:** Conceptualization, Methodology, Formal analysis, Writing - review & editing. **Xiaochuan Tian:** Conceptualization, Methodology, Formal analysis, Writing - review & editing.

Acknowledgments

L. Demkowicz was partially supported with NSF Grant No. 1819101, AFRL, United States Grant No. FA9550-17-1-0090 and Army Grant No. W911NF-15-1-0552. T. Führer and N. Heuer are supported by CONICYT, Chile through FONDECYT projects 11170050 and 1190009. X. Tian is partially supported with NSF Grant No. 1819233.

Appendix A. Stability analysis for a 1D system of Broersen and Stevenson

Independence of boundedness below constant β from ϵ implies that the quasi-optimal (adjoint graph) test norm with $\alpha = 1$ used in the ultraweak variational formulation is robustly equivalent to the optimal (adjoint) test norm. Consequently, the double adaptivity method with the quasi-optimal (adjoint graph) test norm delivers a projection onto the trial space that is robustly equivalent to the L^2-projection, see Remark 4. We show that, for the 1D problem version of the model confusion problem, the formulation of Broersen and Stevenson is indeed optimal.

Consider the following reformulation of the 1D model confusion problem,

$$\begin{cases} \sigma - \epsilon^{1/2}u' &= g \\ -\epsilon^{1/2}\sigma' + u' &= f \\ u(0) = u(1) &= 0\,. \end{cases} \tag{A.29}$$

We will show that $\|u\|$, $\|\sigma\|$ are bounded by $\|f\| + \|g\|$ robustly in ϵ.

We begin by eliminating σ to arrive at the second-order equation for u,

$$-\epsilon u'' + u' = f + \epsilon^{1/2}g'\,. \tag{A.30}$$

Following the proof in [20], Lemma 2.2, we introduce an auxiliary function $w(x) = e^{-x}u(x)$. A direct computation shows:

$$\begin{aligned} -\epsilon w'' + w' &= e^{-x}(-\epsilon u'' + u') - \epsilon e^{-x}(u - 2u') - e^{-x}u \\ &= e^{-x}(f + \epsilon^{1/2}g') + 2\epsilon w' - (1-\epsilon)w\,. \end{aligned}$$

Equivalently,

$$-\epsilon w'' + (1-2\epsilon)w' + (1-\epsilon)w = e^{-x}(f + \epsilon^{1/2}g')\,.$$

Multiplying by w, integrating over $(0,1)$, integrating the diffusion term by parts and using the BC on w to conclude vanishing of the convective term, we obtain,

$$\epsilon\|w'\|^2 + (1-\epsilon)\|w\|^2 = (e^{-x}(f + \epsilon^{1/2}g'), w)\,.$$

We now estimate the right-hand side. For the term with f, we have,

$$(e^{-x}f, w) \le \frac{1}{4\delta}\|f\|^2 + \delta\|w\|^2$$

with arbitrary $\delta > 0$. Furthermore, for the term with g,

$$\epsilon^{1/2}(e^{-x}g', w) = -\epsilon^{1/2}(g, (e^{-x}w)') = \epsilon^{1/2}(g, e^{-x}w) - \epsilon^{1/2}(g, e^{-x}w')\,.$$

The first term on the right-hand side estimates by,

$$\epsilon^{1/2}(g, e^{-x}w) = (g, e^{-x}\epsilon^{1/2}w) \le \frac{1}{2}\|g\|^2 + \frac{\epsilon}{2}\|w\|^2\,,$$

and the second one by,

$$-\epsilon^{1/2}(g, e^{-x}w') = -(g, e^{-x}\epsilon^{1/2}w') \le \frac{1}{2}\|g\|^2 + \frac{\epsilon}{2}\|w'\|^2\,.$$

Summing up, we obtain the estimate,

$$\frac{\epsilon}{2}\|w'\|^2 + (1 - \frac{3}{2}\epsilon - \delta)\|w\|^2 \le \frac{1}{4\delta}\|f\|^2 + \|g\|^2$$

Selecting sufficiently small δ, and assuming sufficiently small ϵ, we obtain robust estimates for $\|w\|^2$ and $\epsilon\|w'\|^2$,

$$\left.\begin{matrix} \epsilon\|w'\|^2 \\ \|w\|^2 \end{matrix}\right\} \lesssim \|f\|^2 + \|g\|^2\,.$$

Above, $\lesssim$ denotes a bound with a constant independent of ϵ. This, in turn, implies robust estimates for $\|u\|^2$ and $\epsilon\|u'\|^2$,

$$\left.\begin{matrix} \epsilon\|u'\|^2 \\ \|u\|^2 \end{matrix}\right\} \lesssim \|f\|^2 + \|g\|^2\,.$$

Finally, utilizing Eq. (A.29)$_1$, we obtain a robust estimate for $\|\sigma\|^2$ as well,

$$\|\sigma\|^2 \lesssim \|f\|^2 + \|g\|^2\,.$$

Appendix B. Classical variational formulation

In this section, we restart our discussion and develop a duality-based a-posteriori error estimate for the inner adaptivity loop for the classical variational formulation of the model confusion problem. Unfortunately, the presented results are valid only in one space dimension as only in 1D we can derive analytically the formula for the optimal test norm. This is a 'stand alone' presentation using its own notation.

We return to the 1D model *confusion* problem (3.27). Multiplying the equation with test functions v vanishing at the end-points, integrating over the domain, and integrating by parts, we arrive at the *classical variational formulation*:

$$\begin{cases} u \in H_0^1(0,1) \\ \epsilon(u',v') + (u',v) = (f,v) \quad v \in H_0^1(0,1)\,. \end{cases} \tag{B.31}$$

As usual, the parenthesis (u,v) denote the $L^2(0,1)$ inner product, with the corresponding L^2-norm denoted by $\|u\|$ without any symbol for the space.

Derivation of the optimal test norm. The formulation admits a symmetric functional setting, i.e. the trial and test spaces are the same. For the trial space we will employ the H_0^1-norm, i.e. the H^1-seminorm,

$$\|u\|_U := \|u'\|\,.$$

The reason for this choice is two-fold: (a) the orthogonal projection in this norm delivers *exact values at the vertex nodes*, i.e. solution $\widetilde{u}_h$ of semi-discrete problem (1.5) interpolates the exact solution u at vertex nodes, (b) for this trial norm, we can derive analytically the corresponding optimal test norm. Recall the definition of the optimal test norm,

$$\|v\|_{V_{\mathrm{opt}}} = \|b(\cdot,v)\|_{U'} = \|u_v\|_U$$

where u_v is the Riesz representation of functional $b(\cdot,v)$, i.e. it solves the variational problem:

$$\begin{cases} u_v \in H_0^1(0,1) \\ ((\delta u)',u_v') = (\delta u,u_v)_U = b(\delta u,v) = \epsilon((\delta u)',v') + ((\delta u)',v) \quad \delta u \in H_0^1(0,1)\,. \end{cases}$$

Consequently, u_v satisfies the equation:

$$-u_v'' = -\epsilon v'' - v'\,.$$

Integrating,

$$u_v' = \epsilon v' + v - C$$

and integrating over $(0,1)$, with the help of BCs, we obtain: $C = \int_0^1 v$. This leads to the formula for the optimal test norm,

$$\|v\|_{V_{\mathrm{opt}}}^2 = \|u_v'\|^2 = \|\epsilon v' + v - \int_0^1 v\|^2 = \epsilon^2\|v'\|^2 + \|v\|^2 - (\int_0^1 v)^2\,.$$

If we can use this test norm in the double adaptivity algorithm, the PG method with optimal test functions should deliver a vertex interpolant of u. Note that the optimal test norm includes the global term and therefore is not *localizable*, i.e. it cannot be used in the DPG setting where we work with a *broken* $H^1(\mathcal{T}_h)$ test space.

The concept of the optimal test norm and the importance of resolving the residual are illustrated in Fig. 11. In all cases we use the same trial mesh of five quadratic elements and refine the test mesh starting with test mesh coinciding with the trial mesh. Let $\hat{u}_h$ be the standard Bubnov–Galerkin solution. For the test mesh coinciding with the trial mesh, Galerkin orthogonality implies that the corresponding residual ψ_h vanishes, and solution u_h of the mixed problem coincides with $\hat{u}_h$. Note that this is true for *any test norm*. In other words, the choice of the test norm matters only if $\dim V_h > \dim U_h$. The second row shows the approximate solution obtained with a test mesh of elements of order $p = 4$, and the corresponding approximate residual ψ_h. Oscillations have disappeared but the solution is still far from the H_0^1-projection. The third row presents the analogous results for the test mesh with $p = 7$. With this test mesh, the exact residual has been resolved and the method delivers the H_0^1-projection, as promised. Finally, the last row shows the effect of refining globally the last test mesh. The residual is now better resolved but refining the test mesh has little effect on the approximate solution. Clearly, the extra refinement of the test mesh is not necessary.

Solution of the mixed problem. The algebraic structure of the mixed problem looks as follows.

$$\begin{pmatrix} G & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} \psi \\ u \end{pmatrix} = \begin{pmatrix} l \\ 0 \end{pmatrix} \tag{B.32}$$

where ψ and u denote now the vectors of degrees-of-freedom (d.o.f.), B is the rectangular stiffness matrix corresponding to the original problem, and G is the Gram matrix corresponding to the test inner product. Due to the presence of the non-local term in the optimal test product, the Gram matrix is fully populated (dense) which seems to necessitate the use of a dense matrix solver, a clear death sentence for the whole methodology. Fortunately, the global term represents a

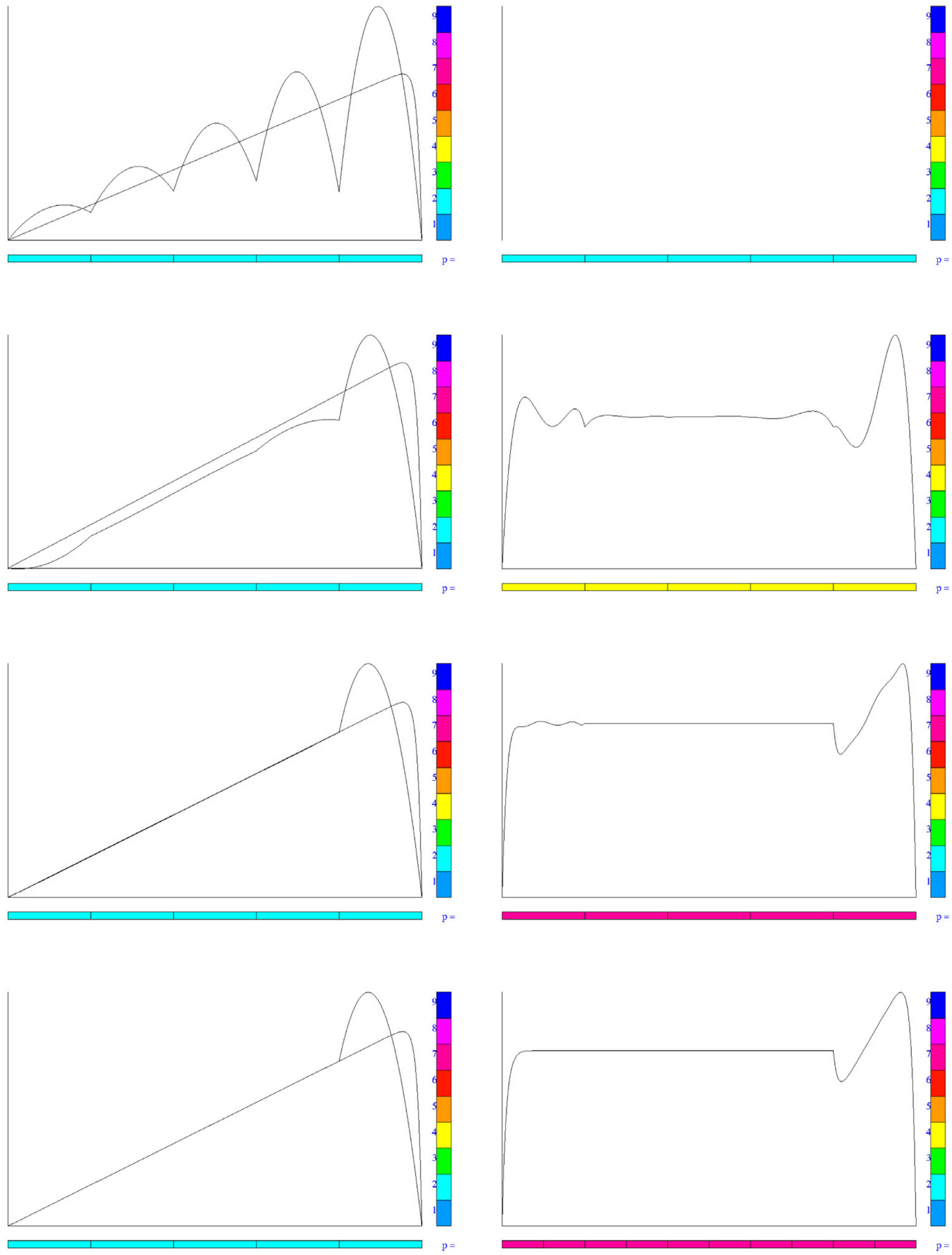

Fig. 11. $\epsilon = 10^{-2}$. Left: Evolution of the approximate solution on a trial mesh of five quadratic elements corresponding to different test meshes. Right: The test mesh with the corresponding approximate residual ψ_h.

rank one contribution only and the use of standard banded matrix solvers (including classical frontal solver w/o pivoting) is still possible. In order to see that, we rewrite first system (B.32) using indices,

$$\begin{cases} (\bar{G}_{ik} + d_i d_k)\psi_k + B_{il} u_l & = l_i \quad i = 1, \ldots, m \\ B_{kj}\psi_k & = 0 \quad j = 1, \ldots, n\,. \end{cases}$$

With e_k, $k = 1, \ldots, n$ and g_l, $l = 1, \ldots, m$ denoting the global basis functions for the trial and test spaces, the corresponding matrices are defined as follows.

$$\bar{G}_{ik} = \int_0^1 \epsilon^2 g_i' g_k' + g_i g_k$$

$$d_i = \int_0^1 g_i$$
$$B_{il} = \int_0^1 \epsilon e_l' g_i' + e_l' g_i$$
$$l_i = \int_0^1 f g_i \, .$$

Denote now $d_k \psi_k =: \lambda$ and solve the mixed system without the global term contribution to the Gram matrix twice, first for the original load, and second time with load d_i,

$$\begin{cases} \bar{G}_{ik}\psi_k^1 + B_{il}u_l^1 & = l_i \\ B_{kj}\psi_k^1 & = 0 \end{cases} \quad \text{and} \quad \begin{cases} \bar{G}_{ik}\psi_k^2 + B_{il}u_l^2 & = d_i \\ B_{kj}\psi_k^2 & = 0 \end{cases} \, . \tag{B.33}$$

Clearly, by superposition, the solution to the original problem is:

$$\psi_k = \psi_k^1 - \lambda\psi_k^2, \qquad u_l = u_l^1 - \lambda u_l^2 \, .$$

The constant λ is determined by solving the linear equation:

$$\lambda = d_k\psi_k = d_k(\psi_k^1 - \lambda\psi_k^2) \qquad \Rightarrow \qquad \lambda = \frac{d_k\psi_k^1}{1 + d_k\psi_k^2} \, .$$

The solution of the original system of equations is obtained thus by solving the band matrix mixed system (B.33) with two load vectors (using e.g. a frontal solver), computing constant λ, and using the superposition as explained above. The technique is known as *Sherman–Morrison formula.*

Other boundary conditions. The formula for the optimal test norm (and the presence of the global term) is a consequence of the particular boundary conditions (BCs). Consider another (physically meaningful) set of BCs:

$$\begin{array}{rll} -\epsilon u' + u & = 0 & \text{at} \quad x = 0 \\ u & = 0 & \text{at} \quad x = 1 \, . \end{array} \tag{B.34}$$

We have now,

$$U = V := \{ v \in H^1(0,1) \, : \, v(1) = 0 \}$$

and the bilinear form is given by:

$$b(u, v) = \epsilon(u', v') - (u, v') \, .$$

Note that the transport term has now been integrated by parts as well. Employing the same trial inner product, determination of the optimal test norm is reduced to the solution of the variational problem:

$$\begin{cases} u_v \in U \\ ((\delta u)', u_v') = (\delta u, u_v)_U = b(\delta u, v) = \epsilon((\delta u)', v') - (\delta u, v') \quad \delta u \in U \, . \end{cases}$$

This is equivalent to the following BVP,

$$\begin{array}{rll} -u_v'' & = -\epsilon v'' - v' & \text{in } (0,1) \\ -u_v' & = -\epsilon v' & \text{at } x = 0 \\ u_v & = 0 & \text{at } x = 1 \, . \end{array}$$

Integrating the first equation and utilizing the BC at $x = 0$, we obtain

$$-u_v' = -\epsilon v' - v + v(0) \, .$$

This leads to the following formula for the optimal test norm.

$$\|v\|_{V_{\text{opt}}}^2 = \|u_v'\|^2 = \|\epsilon v' + v - v(0)\|^2 \, . \tag{B.35}$$

The global term is now missing.

We will develop two dual problems that provide a basis for the a-posteriori error estimation and inner loop adaptivity. As we will see, developing the dual formulations is much less straightforward than for the UW formulation and the adjoint graph norm.

We begin by noticing that the semi-discrete mixed problem is equivalent to the constrained minimization problem:

$$\inf_{\psi \in V_0} \frac{1}{2}\|\epsilon\psi'\|^2 + \frac{1}{2}\|\psi - \int_0^1 \psi\|^2 - (f, \psi) \, , \tag{B.36}$$

where

$$V_0 := \{ v \in V : b(\delta u_h, v) = 0 \quad \forall \delta u_h \in U_h \}.$$

Note that through integration by parts,

$$\epsilon(u_h', v') + (u_h', v) = 0 \quad \Leftrightarrow \quad (\epsilon u_h' - u_h, v') = 0 \quad \forall u_h \in U_h.$$

The orthogonality of v to the trial space through the bilinear form can thus be interpreted in terms of orthogonality of v' to a special finite dimensional space. We will see in the next two different dual formulations slightly different interpretations of this orthogonality.

B.1. Dual formulation I

Introducing the finite dimensional space

$$W_h = \{\epsilon u_h' - u_h \,:\, u_h \in U_h\}\,, \tag{B.37}$$

and the notation $\bar{\psi}$ for $\int_0^1 \psi$, we reformulate the semi-discrete mixed problem as the constrained minimization problem,

$$\inf_{\substack{\psi \in H_0^1(0,1)\\ \psi' \in W_h^\perp}} \underbrace{\frac{1}{2}\|\epsilon\psi'\|^2 + \frac{1}{2}\|\psi - \bar{\psi}\|^2 - (f, \psi)}_{=:J(\psi)}, \tag{B.38}$$

where $W_h^\perp$ denotes the L^2-orthogonal complement of W_h. The derivation follows now similar lines to those for the UW formulation. First, we turn the minimization problem into an inf–sup problem,

$$\inf_{\psi\in H_0^1(0,1)} \inf_{\substack{\sigma \in L^2(0,1)\\ \sigma \in W_h^\perp}} \sup_{\phi\in H^1(0,1)} \frac{1}{2}\|\sigma\|^2 + \frac{1}{2}\|\psi - \bar{\psi}\|^2 - (f, \psi) + (\epsilon\psi' - \sigma, \phi).$$

We replace now inf sup with sup inf, and perform minimizations in σ and ψ, leading to the duality relations:

$$\sigma = \phi^\perp, \quad \psi - \bar{\psi} = f + \epsilon\phi'.$$

Now since the average of $\psi - \bar{\psi}$ is zero, this forces ϕ to satisfy the periodic boundary condition:

$$\phi(1) - \phi(0) = -\frac{1}{\epsilon}\bar{f}.$$

The dual problem is thus given by

$$\sup_{\substack{\phi \in H^1(0,1)\\ \phi|_0^1 = -\frac{1}{\epsilon}\bar{f}}} \underbrace{-\frac{1}{2}\|\phi^\perp\|^2 - \frac{1}{2}\|f + \epsilon\phi'\|^2}_{=:J^*(\phi)}. \tag{B.39}$$

A simple algebra combined with integration by parts shows that

$$2(J(\psi) - J^*(\phi)) = \int_0^1 (\epsilon\psi' - \phi^\perp)^2 + (\psi - \bar{\psi} - (f + \epsilon\phi'))^2 \tag{B.40}$$

where $\psi \in H_0^1(0,1)$, $\psi' \in W_h^\perp$ and ϕ is taken from $H^1(0,1)$ with periodic constraint $\phi(1) - \phi(0) = -\frac{1}{\epsilon}\bar{f}$.

The solution of the primal problem satisfies the mixed problem:

$$\begin{cases} \psi^h \in H_0^1(0,1),\ \widetilde{u}_h \in U_h \\ \epsilon^2((\psi^h)', \delta\psi') + (\psi^h - \overline{\psi^h}, \delta\psi) \quad + b(\widetilde{u}_h, \delta\psi) & = (f, \delta\psi) \quad \delta\psi \in H_0^1(0,1) \\ b(\delta u_h, \psi^h) & = 0 \qquad\quad \delta u_h \in U_h \end{cases} \tag{B.41}$$

where $\widetilde{u}_h \in U_h$ is the corresponding Lagrange multiplier.

The dual problem (B.39) is equivalent to the double minimization problem:

$$\inf_{\substack{\phi \in H^1(0,1)\\ \phi(1) - \phi(0) = -\frac{1}{\epsilon}\bar{f}}} \inf_{w_h\in W_h} \frac{1}{2}\|\phi - w_h\|^2 + \frac{1}{2}\|f + \epsilon\phi'\|^2$$

where W_h is defined by (B.37). This leads to the mixed problem:

$$\begin{cases} \phi^h \in H^1(0,1),\ \phi^h|_0^1 = -\frac{1}{\epsilon}\bar{f},\ \widetilde{w}_h \in W_h \\ \epsilon^2((\phi^h)', \delta\phi') + (\phi^h, \delta\phi) - (\widetilde{w}_h, \delta\phi) & = -\epsilon(f, \delta\phi') \quad \delta\phi \in H^1(0,1),\ \delta\phi|_0^1 = 0 \\ -(\phi^h, \delta w_h) + (\widetilde{w}_h, \delta w_h) & = 0 \qquad\qquad \delta w_h \in W_h \end{cases} \tag{B.42}$$

which has the strong form:

$$-\epsilon^2(\phi^h)'' + \phi^h - \widetilde{w}_h = \epsilon f', \quad \phi^h - \widetilde{w}_h \in W_h^\perp$$

plus the boundary condition:

$$(\epsilon(\phi^h)' + f)\big|_0^1 = 0.$$

Lemma 2. *Let ψ^h and ϕ^h be the solutions to primal problem (B.41) and dual problem (B.42). Then, the right-hand side in identity (B.40) equals zero, i.e. there is no duality gap on the continuous level. Moreover, the Lagrange multiplier $\widetilde{u}_h$ in (B.41) and $\widetilde{w}_h$ in (B.42) satisfy the relation:*

$$\epsilon\widetilde{u}_h' - \widetilde{u}_h = \widetilde{w}_h\,. \quad \blacksquare \tag{B.43}$$

Proof. Assume that ϕ^h is a solution to the dual problem, then define $\psi^h = f + \epsilon(\phi^h)' + C$, with a constant C to be determined. With the boundary conditions, we have the following relations:

$$\psi^h - \overline{\psi^h} = f + \epsilon(\phi^h)', \quad \psi^h(1) - \psi^h(0) = 0\,. \tag{B.44}$$

Notice that there are infinitely many ψ^h's that satisfy the above relations. In fact, any constant shift of a particular ψ^h still satisfies the relations. We can thus pick ψ^h such that $\psi^h(0) = 0$, so that $\psi^h \in H_0^1(0,1)$. First, we see that ψ^h and ϕ^h defined in this way satisfy $2(J(\psi^h) - J^*(\phi^h)) = 0$, because in addition to (B.44), we have

$$(\psi^h)' = f' + \epsilon(\phi^h)'' = \frac{1}{\epsilon}(\phi^h - \widetilde{w}_h) = \frac{1}{\epsilon}(\phi^h)^\perp\,.$$

Next, let $\widetilde{u}_h$ satisfy relation (B.43). We claim that ψ^h and $\widetilde{u}_h$ satisfy primal problem (B.41). This is true because, for all $\delta\psi \in H_0^1(0,1)$, we have

$$\begin{aligned}
&\epsilon^2((\psi^h)', \delta\psi') + (\psi^h - \overline{\psi^h}, \delta\psi) + \epsilon(\widetilde{u}_h', \delta\psi') + (\widetilde{u}_h', \delta\psi)\\
=&\epsilon(\phi^h - \widetilde{w}_h, \delta\psi') + (f + \epsilon(\phi^h)', \delta\psi) + (\epsilon\widetilde{u}_h' - \widetilde{u}_h', \delta\psi')\\
=&\epsilon(\phi^h, \delta\psi') + (f + \epsilon(\phi^h)', \delta\psi) = 0\,,
\end{aligned}$$

and

$$b(\delta u_h, \psi^h) = (\epsilon\delta u_h' - \delta u_h, (\psi^h)') = (\delta w_h, (\phi^h)^\perp) = 0.$$

This shows that the function ψ^h derived from the duality relations indeed is the solution of the primal problem. Therefore, there is no duality gap at the continuous level. $\blacksquare$

B.2. Dual formulation II

The second dual formulation is obtained by making a slight change in the definition of the finite-dimensional space W_h. Instead of (B.37), we now define

$$W_h = \{\epsilon u_h' - u_h \,:\, u_h \in U_h\} \oplus \mathbb{R}\,. \tag{B.45}$$

Noticing that the homogeneous Dirichlet boundary condition for u implies that u' is L^2-orthogonal to constants, we can rewrite the semi-discrete mixed problem into

$$\inf_{\substack{\psi \in H^1(0,1),\, \psi(0)=1\\ \psi' \in W_h^\perp}} \underbrace{\frac{1}{2}\|\epsilon\psi'\|^2 + \frac{1}{2}\|\psi - \bar{\psi}\|^2 - (f,\psi)}_{=:J(\psi)}. \tag{B.46}$$

Now turning the minimization problem into a inf–sup problem, we have

$$\inf_{\substack{\psi \in H^1(0,1)\\ \psi(0)=0}} \inf_{\substack{\sigma \in L^2(0,1)\\ \sigma \in W_h^\perp}} \sup_{\substack{\phi \in H^1(0,1)\\ \phi(1)=0}} \frac{1}{2}\|\sigma\|^2 + \frac{1}{2}\|\psi - \bar{\psi}\|^2 - (f,\psi) + (\epsilon\psi' - \sigma, \phi).$$

Following the reasoning for the first dual formulation, we arrive at the second dual formulation:

$$\sup_{\substack{\phi \in H^1(0,1)\\ \phi(0) = \frac{1}{\epsilon}\bar{f},\ \phi(1)=0}} \underbrace{-\frac{1}{2}\|\phi^\perp\|^2 - \frac{1}{2}\|f + \epsilon\phi'\|^2}_{=:J^*(\phi)}\,. \tag{B.47}$$

where $\phi^\perp$ is again the L^2-projection of ϕ onto the L^2-orthogonal component of W_h, with the slightly different definitions of W_h in the two dual formulations.

The proof for the second dual formulation is quite similar, except for dealing with different boundary conditions. Formula (B.40) for the difference between the approximate primal and dual energies still holds, but we now take

Table 3
Primal formulation. Number of inner loop iterations for different values of the viscosity constant.

Outer iter	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
$\epsilon = 10^{-2}$	8	7	6	4	4	4	3	3	3																			
$\epsilon = 10^{-3}$	20	20	18	15	14	12	8	7	6	5	5	4	4															
$\epsilon = 10^{-4}$	38	37	36	35	33	32	31	29	24	19	10	10	9	8	8	7												
$\epsilon = 10^{-5}$	58	57	56	55	54	53	52	50	49	48	46	43	35	27	15	13	12	11	10									
$\epsilon = 10^{-6}$	78	77	72	71	66	65	60	59	54	53	48	47	46	44	43	42	41	40	38	36	30	22	12	11	10	10	9	8

$\psi \in H^1(0,1)$ with $\psi(0) = 0$, and $\phi \in H^1(0,1)$ with $\phi(0) = \frac{1}{\epsilon}\bar{f}$ and $\phi(1) = 0$. The mixed formulation of the dual problem reads as follows.

$$\begin{cases} \phi^h \in H^1(0,1),\ \phi^h(0) = \dfrac{1}{\epsilon}\bar{f},\ \phi^h(1) = 0,\ \widetilde{w}_h \in W_h & \\ \epsilon^2((\phi^h)', \delta\phi') + (\phi^h, \delta\phi) - (\widetilde{w}_h, \delta\phi) = -\epsilon(f, \delta\phi') & \delta\phi \in H^1_0(0,1) \\ -\ (\phi^h, \delta w_h) + (\widetilde{w}_h, \delta w_h) = 0 & \delta w_h \in W_h \end{cases} \tag{B.48}$$

where W_h is defined by (B.45). If ϕ^h is the solution to (B.48), then it satisfies in the strong form

$$-\epsilon^2(\phi^h)'' + \phi^h - \widetilde{w}_h = \epsilon f', \quad \phi^h - \widetilde{w}_h \in W_h^\perp.$$

Note that in this case we do not have additional boundary conditions.

Lemma 3. *Let ψ^h and ϕ^h be the solutions to primal problem (B.41) and dual problem (B.48). Then, the right-hand side in identity (B.40) equals zero, i.e. there is no duality gap on the continuous level. Moreover, the Lagrange multiplier $\widetilde{u}_h$ in (B.41) and $\widetilde{w}_h$ in (B.48) satisfy relation (B.43).* ■

Proof. Define again $\psi^h = f + \epsilon(\phi^h)' + C$, with constant C to be determined. Then from the boundary condition on ϕ^h we have, $C = \overline{\psi^h}$, so we arrive at $\psi^h - \overline{\psi^h} = f + \epsilon(\phi^h)'$. For the same reason as before, we can select ψ^h satisfying $\psi^h(0) = 0$. Now, the other boundary condition for ψ^h comes from the fact that

$$(\psi^h)' = f' + \epsilon(\phi^h)'' = \frac{1}{\epsilon}(\phi^h - \widetilde{w}_h) = \frac{1}{\epsilon}(\phi^h)^\perp,$$

and since this time $(\phi^h)^\perp$ is orthogonal to constants by the definition of W_h, so we have $\psi^h(1) - \psi^h(0) = 0$. The rest of the proof is identical with the proof of Lemma 2. ■

B.3. Numerical experiments

We present numerical results for the first dual problem only. All results were obtained with the same tolerances as for the UW formulation: 5% for the inner loop, and 1% error for the outer loop. As the trial norm is now stronger, these tolerances seem perhaps to be a bit too strict. Visually, one observes perfect resolution of the solution already for bigger errors. Figs. 12 and 13 present the outer loop iterations for $\epsilon = 10^{-2}$. As we can see, at least visually, we could have stopped after just four iterations. Notice how the (resolved!) residual changes with the trial mesh. Note also that the solution of the dual problem is rather smooth and gets smoother with iterations. This indicates that the duality-based estimate is a perfect estimate for the error corresponding to the primal problem.

Table 3 presents the number of inner loop iterations for different values of ϵ. The smallest value for which we have been able to solve the problem, was $\epsilon = 10^{-6}$ and it was clearly on the round-off limit as the inner loop convergence was not always monotone. As expected, the number of inner loop iterations grows with smaller ϵ as one has to resolve smaller scales in the residual. Also, as the trial mesh gets refined, the number of inner loop iterations *monotonically decreases*. Note that we did not observe such a systematic decrease for the UW formulation.

Controlling the solution error. Controlling the solution error for the classical formulation follows the arguments from Section 3.2 although this is again a bit more complicated than for the UW formulation. We shall discuss the first dual problem only. The key starting point is again the relation between the solution to the ideal PG problem $\widetilde{u}_h$ and the auxiliary variable $\widetilde{w}_h$ in the exact dual problem,

$$\epsilon\widetilde{u}_h' - \widetilde{u}_h = \widetilde{w}_h\,.$$

This relation defines also the discrete space W_h present in the dual formulation. Recall that $\widetilde{w}_h$ is the L^2-projection of the exact solution ϕ^h to the dual problem onto the discrete space W_h. Similarly, after discretization of the dual problem, w_h is the L^2-projection of the approximate solution ϕ_h onto space W_h. The duality gap estimate,

$$\|(\phi^h)^\perp - \phi_h^\perp\|^2 + \|\epsilon(\phi^h - \phi_h)'\|^2 = 2(J^*(\phi^h) - J^*(\phi_h)) \le \text{est}\,,$$

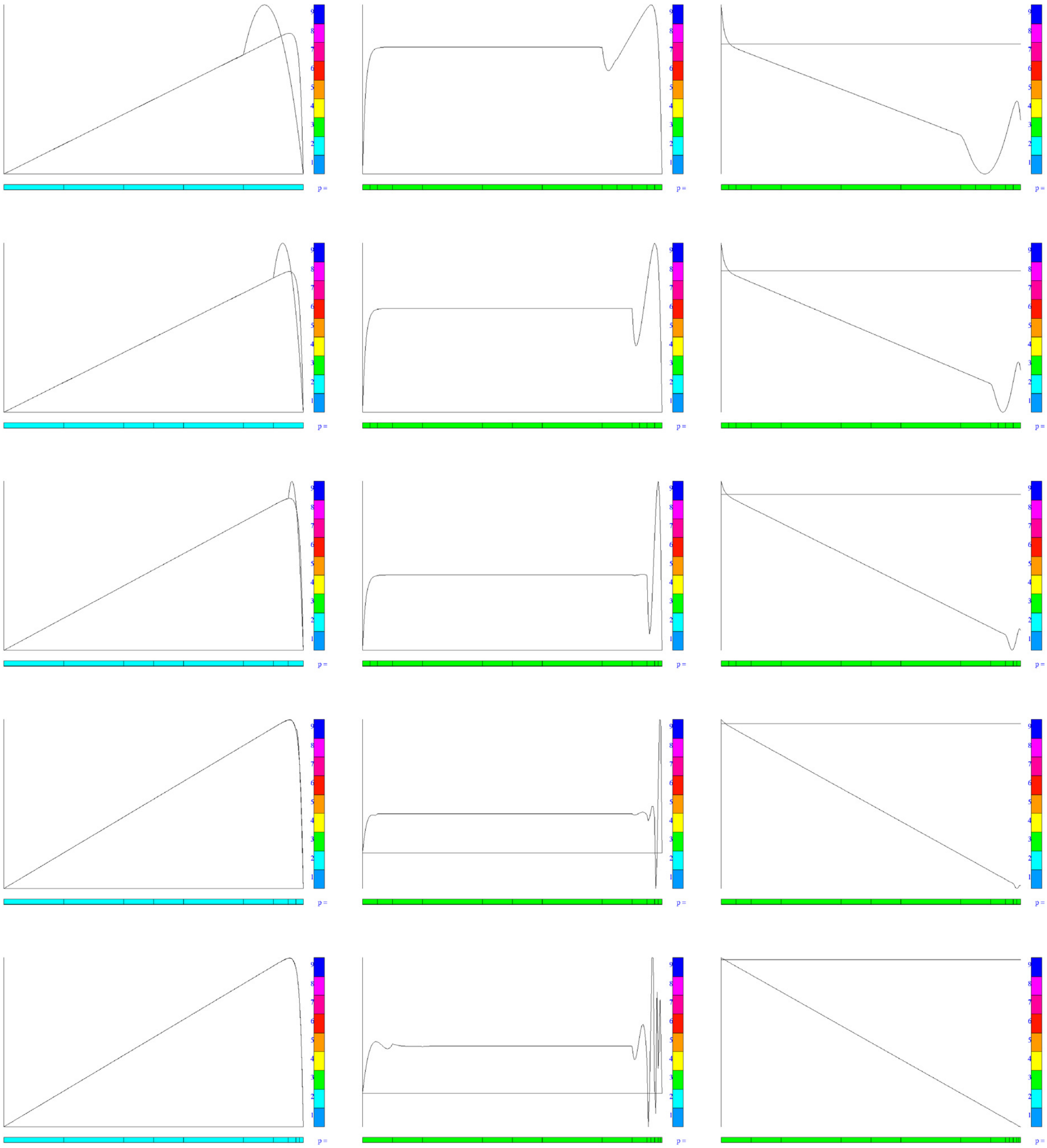

Fig. 12. Primal formulation, $\epsilon = 10^{-2}$, outer loop, iterations 1–5. Left: Evolution of the approximate solution u_h. Middle column: the test mesh with the corresponding resolved residual. Right: The test mesh with the corresponding solution of the dual problem.

and Poincaré inequality,

$$C\|\phi^h\| \le \epsilon\|(\phi^h)'\|, \qquad \phi^h \in H^1(0,1),\ \phi^h(0) = \phi^h(1),$$

imply that we control the error in ϕ^h,

$$C^2\|\phi^h - \phi_h\| \le \text{est}.$$

The Poincaré constant C depends linearly upon ϵ, so the estimate is not robust in ϵ. Control of the error $\|\phi^h - \phi_h\|$ translates into control of the error $\|\widetilde{w}_h - w_h\|$,

$$\|\widetilde{w}_h - w_h\| \le \|\phi^h - \phi_h\|.$$

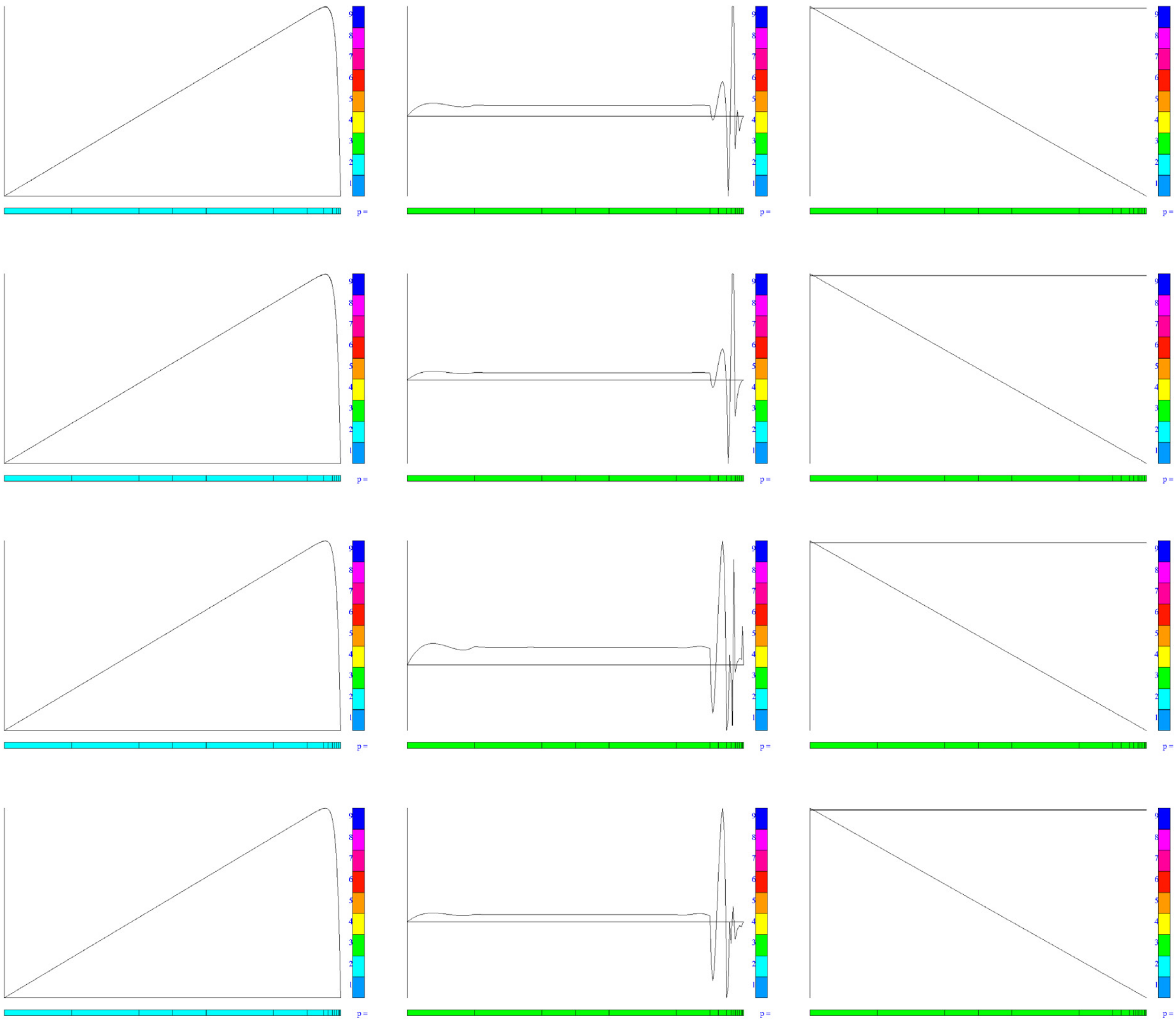

Fig. 13. Primal formulation, $\epsilon = 10^{-2}$, outer loop, iterations 6–9. Left: Evolution of the approximate solution u_h. Middle column: the test mesh with the corresponding resolved residual. Right: The test mesh with the corresponding solution of the dual problem.

Finally, let $u_h \in U_h$ be the unique function defining w_h - the approximation of $\widetilde{w}_h$. We have the relation,

$$\epsilon(\widetilde{u}_h - u_h)' - (\widetilde{u}_h - u_h) = \widetilde{w}_h - w_h \,.$$

Multiply both sides with $-(\widetilde{u}_h - u_h)$, integrate over the interval $(0, 1)$ and use periodic BC to obtain,

$$\|\widetilde{u}_h - u_h\|^2 = -\int_0^1 (\widetilde{w}_h - w_h)(\widetilde{u}_h - u_h)$$

which implies the final estimate,

$$\|\widetilde{u}_h - u_h\| \le \|\widetilde{w}_h - w_h\| \,.$$

Remark 9. Note that we actually need a different Poincaré-like bound,

$$C^2\|\phi\|^2 \le \|\phi^\perp\|^2 + \|\epsilon\phi'\|^2 \,.$$

So the linear dependence of C upon ϵ may be very pessimistic. At the moment, we do not know though how to improve the estimate. ■

References

[1] A. Cohen, W. Dahmen, G. Welper, Adaptivity and variational stabilization for convection–diffusion equations, ESAIM Math. Model. Numer. Anal. 46 (5) (2012) 1247–1273.

[2] L. Demkowicz, Lecture Notes on Energy Spaces, Technical Report 13, ICES, 2018.
[3] J.T. Oden, L.F. Demkowicz, Applied Functional Analysis for Science and Engineering, third ed., Chapman & Hall/CRC Press, Boca Raton, 2018.
[4] L. Demkowicz, J. Gopalakrishnan, A class of discontinuous Petrov–Galerkin methods. Part II: Optimal test functions, Numer. Methods Partial Differential Equations 27 (2011) 70–105, See also ICES Report 2009-16.
[5] J. Zitelli, I. Muga, L. Demkowicz, J. Gopalakrishnan, D. Pardo, V. Calo, A class of discontinuous Petrov-Galerkin methods. Part IV: Wave propagation problems, J. Comput. Phys. 230 (2011) 2406–2432.
[6] J. Gopalakrishnan, W. Qiu, An analysis of the practical DPG method, Math. Comp. 83 (286) (2014) 537–552.
[7] S. Nagaraj, S. Petrides, L. Demkowicz, Construction of DPG Fortin operators for second order problems, Comput. Math. Appl. 74 (8) (2017) 1964–1980.
[8] C. Carstensen, L. Demkowicz, J. Gopalakrishnan, Breaking spaces and forms for the DPG method and applications including Maxwell equations, Comput. Math. Appl. 72 (3) (2016) 494–522.
[9] T. Führer, N. Heuer, Fully discrete DPG methods for the Kirchhoff-Love plate bending model, Comput. Methods Appl. Mech. Engrg. 343 (2019) 550–571.
[10] C. Carstensen, F. Hellwig, Low order discontinuous Petrov-Galerkin finite element methods for linear elasticity, SIAM J. Numer. Anal. 54 (6) (2016) 3388–3410.
[11] D. Broersen, R.P. Stevenson, A robust Petrov-Galerkin discretization of convection–diffusion equations, Comput. Math. Appl. 68 (11) (2014) 1605–1618.
[12] D. Broersen, R.P. Stevenson, A Petrov-Galerkin discretization with optimal test space of a mild-weak formulation of convection–diffusion equations in mixed form, IMA J. Numer. Anal. 35 (1) (2015) 39–73.
[13] I. Ekeland, R. Témam, Convex Analysis and Variational Problems, North Holland, 1976.
[14] J.T. Oden, L. Demkowicz, R. Rachowicz, T.A. Westermann, Toward a universal *hp* adaptive finite element strategy. Part 2: A posteriori error estimation, Comput. Methods Appl. Mech. Engrg. 77 (1989) 113–180.
[15] L. Demkowicz, Computing with *hp* Finite Elements. I.One- and Two-dimensional Elliptic and Maxwell Problems, Chapman & Hall/CRC Press, Taylor and Francis, Boca Raton, 2006.
[16] E. Bänsch, P. Morin, R.H. Nochetto, An adaptive Uzawa FEM for the Stokes problem: Convergence without the inf-sup condition, SIAM J. Numer. Anal. 40 (4) (2002) 1207–1229.
[17] A. Cohen, W. Dahmen, R. De Vore, Adaptive wavelet methods II - beyond the elliptic case, Found. Comput. Math. 2 (2002) 203–245.
[18] S. Dahlke, W. Dahmen, K. Urban, Adaptive wavelet methods for saddle point probelms – Convergence rates, SIAM J. Numer. Anal. 40 (4) (2002) 1230–1262.
[19] W. Dahmen, C. Plesken, G. Welper, Reduced basis methods for transport dominated problems, ESAIM Math. Model. Numer. Anal. 48 (3) (2014) 623–663.
[20] L. Demkowicz, N. Heuer, Robust DPG method for convection-dominated diffusion problems, SIAM J. Numer. Anal 51 (2013) 2514–2537, see also ICES Report 2011/13.
[21] J. Chan, N. Heuer, B. Tan Bui-Thanh, L. Demkowicz, A robust DPG method for convection-dominated diffusion problems II: Adjoint boundary conditions and mesh-dependent test norms, Comput. Math. Appl. 67 (4) (2014) 771–795.
[22] N. Heuer, M. Karkulik, A robust DPG method for singularly perturbed reaction–diffusion problems, SIAM J. Numer. Anal. 55 (3) (2017) 1218–1242.
[23] I. Smears, M. Vohralík, Simple and robust equilibrated flux a posteriori estimates for singularly perturbed reaction–diffusion problems, ESAIM Math. Model. Numer. Anal. 54 (6) (2020) 1951–1973.
[24] L. Demkowicz, On some results concerning the reciprocal formulation for the signorini's problem, Comput. Math. Appl. 8 (1) (1982) 57–74.
[25] L. Demkowicz, M. Swierczek, An adaptive finite element method for a class of variational inequalities, in: W. Kosinsky, et al. (Eds.), Selected Problems of Modern Continuum Theory, Pitagora Editrice, Bologna, 1987, Proceedings of the Italian-Polish Meeting in Bologna, June 3–6, 1987.
[26] S. Bartels, M. Milicevic, Primal–dual gap estimators for a posteriori error analysis of nonsmooth minimization problems, 2019, math.NA.