SIAM J. SCI. COMPUT.
Vol. 35, No. 6, pp. A2752–A2780


# STABILIZED FINITE ELEMENT METHODS FOR NONSYMMETRIC, NONCOERCIVE, AND ILL-POSED PROBLEMS. PART I: ELLIPTIC EQUATIONS*

ERIK BURMAN†

**Abstract.** In this paper we propose a new method to stabilize nonsymmetric indefinite problems. The idea is to solve a forward and an adjoint problem simultaneously using a suitable stabilized finite element method. Both stabilization of the element residual and of the jumps of certain derivatives of the discrete solution over element faces may be used. Under the assumption of well-posedness of the partial differential equation and its associated adjoint problem we prove optimal error estimates in $H^1$ and $L^2$ norms in an abstract framework. Some examples of problems that are neither symmetric nor coercive but that enter the abstract framework are given. First we treat indefinite convection-diffusion equations with nonsolenoidal transport velocity and either pure Dirichlet conditions or pure Neumann conditions and then a Cauchy problem for the Helmholtz operator. Some numerical illustrations are given.

**Key words.** stabilized finite element methods, noncoercive problems, Galerkin least squares, continuous interior penalty, compressible flow, Cauchy problem

**AMS subject classifications.** 65N12, 65N20, 65N30

**DOI.** 10.1137/130916862

## 1. Introduction.

The computation of indefinite elliptic problems often involves certain conditions on the mesh size $h$ for the system to be well-posed and for the derivation of error estimates. The first results on this problem are due to Schatz [19]. The conditions on the mesh parameter can be avoided if a stabilized finite element method is used. Such methods have been proposed by Bramble, Lazarov, and Pasciak [4] and Ku [16] or more recently the continuous interior penalty (CIP) method for the Helmholtz equation suggested by Wu [21], and Zhu, Burman, and Wu [20]. The method proposed herein has some common features with both these methods but appears to have a wider field of applicability. We may treat not only symmetric indefinite problems such as the (real valued) Helmholtz equation but also nonsymmetric indefinite problems such as convection-diffusion problems with nonsolenoidal convection velocity or the Cauchy problem. The latter problem is known to be ill-posed in general [1] and will mainly be explored numerically herein. For all these cases we show that if the primal and adjoint problems admit a unique solution with sufficient smoothness the proposed algorithm converges with optimal order. The case of hyperbolic problems is treated in the companion paper [5].

The idea of this work is to assume ill-posedness of the discrete form of the PDE and regularize it in the form of an optimization problem under constraints. Indeed we seek to minimize the size of the stabilization operator under the constraint of the discrete variational form. The regularization terms are then chosen from well-known stabilized methods respecting certain design criteria given in an abstract analysis. This leads to an extended method where simultaneously both a primal and a dual

*Submitted to the journal's Methods and Algorithms for Scientific Computing section April 15, 2013; accepted for publication (in revised form) September 18, 2013; published electronically December 3, 2013. This work was partially supported by EPSRC (award EP/J002313/1).
http://www.siam.org/journals/sisc/35-6/91686.html
†Department of Mathematics, University College London, London, WC1E 6BT, United Kingdom (e.burman@ucl.ac.uk).

problem are solved but where the exact solution of the dual problem is always trivial. The aim is to obtain a method where possible discrete nonuniqueness is alleviated by discrete regularization, with a nonconsistency that can be controlled so that optimal convergence for smooth solutions is obtained. The method is also a good candidate for cases where the solution to the continuous problem is nonunique, but that is beyond the scope of the present paper.

In spite of the lack of coercivity for the physical problem, the discrete problem has partial coercivity on the stabilization operator. A consequence of this is that depending on the kernel of the stabilization operator a unique discrete solution may often be shown to exist independently of the underlying partial differential equation. This can be helpful when exploring ill-posed problems numerically or when measurement errors in the data may lead to an ill-posed problem, although the true problem is well-posed.

An outline of the paper is as follows. In section 2 we propose an abstract method and prove that the method will have optimal convergence under certain assumptions on the bilinear form. Then in section 3 we discuss stabilized methods that satisfy the assumptions of the abstract theory with particular focus on the Galerkin least squares (GLS) method and the CIP method. Three examples of applications are given in section 4, two different noncoercive transport problems in compressible flow fields and one elliptic Cauchy problem. Finally in section 5 the accuracy and robustness of the proposed method are shown by some computations of solutions to the problems discussed in section 4. In particular we study the performance of the approach for some different Cauchy problems of varying difficulty.

## 2. Abstract formulation.

Let $\Omega$ be a polygonal/polyhedral subset of $\mathbb{R}^d$, $d = 2, 3$. The boundary of $\Omega$ will be denoted $\partial\Omega$ and its normal $n$. For simplicity we will reduce the scope to second order elliptic problems, but the methodology can readily be extended to indefinite elliptic problems of any order, providing the operator has a smoothing property. We will also describe the method mainly in the two-dimensional case, only mentioning the dimension when the two- and three-dimensional cases differ.

We let $V, W$ denote two subspaces of $H^1(\Omega)$. The abstract weak formulation of the continuous problem takes the following form: find $u \in V$ such that

$$a(u, w) = (f, w) \quad \forall w \in W. \tag{2.1}$$

The formal adjoint of (2.1) reads: find $z \in W$ such that

$$a(v, z) = (g, v) \quad \forall v \in V. \tag{2.2}$$

The bilinear form $a(\cdot, \cdot) : V \times W \to \mathbb{R}$ is assumed to be elliptic but neither symmetric nor coercive. We denote the forward problem on strong form $\mathcal{L}u = f$ and the adjoint problem on strong form $\mathcal{L}^* z = g$. Suitable boundary conditions are integrated either in the spaces $V, W$ or in the linear form.

We assume that both these problems are well-posed and that the geometry and data are such that the smoothing property holds,

$$|u|_{H^2(\Omega)} \le c_{a,\Omega} \|f\|, \quad |z|_{H^2(\Omega)} \le c_{a,\Omega} \|g\|. \tag{2.3}$$

We will frequently use the notation $a \lesssim b$ for $a \le Cb$ with $C$ a constant depending only on the mesh geometry and physical parameters giving an order one contribution. We will also use $a \sim b$ for $a \lesssim b$ and $b \lesssim a$. Indexed constants $c_{xy}$ will depend on the variables $xy$ but can differ at each occurrence.

The $L^2$-scalar product over some $X \subset \mathbb{R}^d$ is denoted $(\cdot,\cdot)_X$ and the associated norm $\|\cdot\|_X$, and the subscript is dropped whenever $X = \Omega$. We will also use $\langle \cdot,\cdot\rangle_Y$ to denote the $L^2$-scalar product over $Y \subset \mathbb{R}^{d-1}$ and $(\cdot,\cdot)_h$ the elementwise $L^2$-norm with the associated broken norm $\|\cdot\|_h$.

*Remark* 2.1. The above regularity assumptions are necessary to ensure optimal convergence for piecewise affine approximation spaces. If polynomial approximation of order $k$ is used we additionally need $u \in H^{k+1}(\Omega)$. If on the other hand the solution is less regular the convergence order is reduced in the standard way and in some cases the mesh constraints for well-posedness will be stronger. More precisely if $u \in H^s(\Omega)$ and $z \in H^t(\Omega)$ with $s,t \in (1,2)$ the analysis below leads to estimates on the form

$$\|u - u_h\| \lesssim h^{s+t-2}.$$

**2.1. Finite element discretization.** Let $\{\mathcal{T}_h\}_h$ denote a family of quasi-uniform, shape regular triangulations $\mathcal{T}_h := \{K\}$, indexed by the maximum triangle radius $h := \max_{K\in\mathcal{T}_h} h_K$, $h_K := \mathrm{diam}(K)$. The set of faces of the triangulation will be denoted by $\mathcal{F}$ and $\mathcal{F}_{int}$ denotes the subset of interior faces. Now let $X_h^k$ denote the finite element space of continuous, piecewise polynomial functions on $\mathcal{T}_h$,

$$X_h^k := \{v_h \in H^1(\Omega) : v_h|_K \in \mathbb{P}_k(K) \quad \forall K \in \mathcal{T}_h\}.$$

Here $\mathbb{P}_k(K)$ denotes the space of polynomials of degree less than or equal to $k$ on a triangle $K$.

We let $\pi_L$ denote the standard $L^2$-projection onto $X_h^k$ and $i_h : C^0(\bar{\Omega}) \mapsto X_h^k$ the standard Lagrange interpolant. Recall that for any function $u \in (V \cup W) \cap H^{k+1}(\Omega)$ there holds

$$\|u - i_h u\| + h\|\nabla(u - i_h u)\| + h^2\|D^2(u - i_h u)\|_h \le c_i h^{k+1}|u|_{H^{k+1}(\Omega)} \tag{2.4}$$

and under our assumptions on the mesh, similarly for $\pi_L$. We propose the following finite element method for the approximation of (2.1): find $u_h, z_h \in V_h \times W_h$ such that

$$\begin{aligned} a_h(u_h, w_h) + s_a(z_h, w_h) &= (f, w_h),\\ a_h(v_h, z_h) - s_p(u_h, v_h) &= -s_p(u, w_h) \end{aligned} \tag{2.5}$$

for all $v_h, w_h \in V_h \times W_h$. Note the appearance of $s_p(u, w_h)$ in the right-hand side of the second equation of (2.5). This means that only stabilizations for which $s_p(u, w_h)$ can be expressed using known data may be used. For residual-based stabilizations this typically is the case, but also for so-called observers that stabilize the computation using measured data. We will always assume that $u$ is sufficiently regular so that $s_p(u,\cdot)$ is well defined, i.e., the stabilization is strongly consistent. The analysis using weak consistency of the stabilization is a straightforward modification.

The bilinear form $a_h(\cdot,\cdot)$ is a discrete realization of $a(\cdot,\cdot)$, typically modified to account for the effect of boundary conditions, since in general $V_h \notin V$ and $W_h \notin W$. The penalty operators $s_a(\cdot,\cdot)$ and $s_p(\cdot,\cdot)$ are symmetric stabilization operators and associated with the adjoint and the primal equation, respectively.

The rationale of the formulation may be explained in an optimization framework. Assume that we want to solve the problem: find $u_h \in V_h$ such that

$$a_h(u_h, w_h) = (f, w_h) \quad \forall w_h \in W_h,$$

but that the system matrix corresponding to $a_h(u_h, w_h)$ has zero eigenvalues. The discrete system is ill-posed. This often reflects some poor stability properties of the underlying continuous problem. The idea is to introduce a selection criterion for the solution, in order to ensure discrete uniqueness, measured by some operator $s_p(u_h, v_h)$. This can include both stabilization (regularization) terms and the fitting of the computed solution to measurements. The formulation then is written as follows: find $u_h, z_h \in V_h \times W_h$ stationary point of the Lagrangian

$$\text{Ł}(u_h, z_h) := \frac{1}{2}s_p(u_h - u, u_h - u) - \frac{1}{2}s_a(z_h, z_h) - a_h(u_h, z_h) + (f, z_h). \tag{2.6}$$

The saddle point structure of the Lagrangian has been enhanced by the addition of the regularizing term $-\frac{1}{2}s_a(z_h, z_h)$. We may readily verify that

$$\frac{\partial \text{Ł}}{\partial u_h}(v_h) = s_p(u_h - u, v_h) - a_h(v_h, z_h)$$

and

$$\frac{\partial \text{Ł}}{\partial z_h}(w_h) = -a_h(u_h, w_h) - s_a(z_h, w_h) + (f, w_h).$$

It follows that (2.5) corresponds to the optimality conditions of (2.6).

Observe that the second equation of (2.5) is a finite element discretization of the dual problem (2.2) with data $g = 0$. Hence the solution to approximate is $z = 0$. The discrete function $z_h$ will most likely not be zero, since it is perturbed by the stabilization operator acting on the solution $u_h$, which in general does not coincide with the stabilization acting on $u$.

We will assume that the following strong consistency properties hold. If $u$ is the solution of (2.1), then

$$a_h(u, \varphi) = (\mathcal{L}u, \varphi) \ \forall \ \varphi \in W + W_h, \tag{2.7}$$

and if $z$ is the solution of (2.2), then

$$a_h(\phi, z) = (\phi, \mathcal{L}^* z) \ \forall \ \phi \in V + V_h. \tag{2.8}$$

As a consequence the following Galerkin orthogonalities hold:

$$a_h(u - u_h, v_h) = s_a(z_h, v_h) \text{ and } a_h(w_h, z - z_h) = s_p(u - u_h, w_h). \tag{2.9}$$

The bilinear forms $s_a(\cdot, \cdot)$, $s_p(\cdot, \cdot)$ are symmetric, positive semidefinite, weakly consistent stabilization operators. The seminorms on $V_h$ and $W_h$ associated to the stabilization are defined by

$$|x_h|_{S_y} := s_y(x_h, x_h)^{\frac{1}{2}}, \quad y = a, p.$$

The rationale of this formulation is that the following partial coercivity is obtained by taking $w_h = z_h$ and $v_h = u_h$:

$$|z_h|^2_{S_a} + |u_h|^2_{S_p} = (f, z_h) - s_p(u, u_h). \tag{2.10}$$

We assume that there are interpolation operators $\pi_V : V \to V_h$ and $\pi_W : W \to W_h$, satisfying (2.4) and also that the following continuity relations hold for all $v, w, y \in H^2(\Omega)$ and for all $v_h, x_h \in W_h$:

$$a_h(v - \pi_V v, x_h) \leq \|v - \pi_V v\|_+ (c_a |x_h|_{S_a} + \epsilon(h) \|x_h\|) \tag{2.11}$$

and

(2.12) $$a_h(v-v_h, y-\pi_W y) \le \|y-\pi_W y\|_*(c_a\|v-\pi_V v\|_{\mathcal{L}}+c_a|v_h-\pi_V v|_{S_p}+\epsilon(h)\|v-v_h\|).$$

We have introduced the notation $\|\cdot\|_+$, $\|\cdot\|_*$ and $\|\cdot\|_{\mathcal{L}}$ for seminorms to be defined. These norms, and those induced by the stabilization operators, will be assumed to satisfy the approximation estimates

(2.13)
$$\|v-\pi_V v\|_{\mathcal{L}} + \|v-\pi_V v\|_+ + |v-\pi_V v|_{S_p} \le c_{a,\gamma} h^k |v|_{H^{k+1}(\Omega)} \quad \forall v \in V \cap H^{k+1}(\Omega),$$

(2.14) $$\|w-\pi_W w\|_* + |w-\pi_W w|_{S_a} \le c_{a,\gamma} h^k |w|_{H^{k+1}(\Omega)} \quad \forall w \in W \cap H^{k+1}(\Omega)$$

and the additional upper bounds

(2.15)
$$|\pi_W w|_{S_a} \le c_{a,\gamma} h |w|_{H^2(\Omega)} \, \forall w \in W \cap H^2(\Omega), \; |\pi_V v|_{S_p} \le c_{a,\gamma} h |v|_{H^2(\Omega)} \, \forall v \in V \cap H^2(\Omega).$$

Here $c_{a,\gamma}$ depends on the form $a(\cdot,\cdot)$ and a stabilization parameter $\gamma$.

**2.2. Convergence analysis for the abstract method.** We first prove that the stabilization seminorm of the discrete error is bounded by one term that converges to zero at an optimal rate and another nonessential perturbation that can be made small.

LEMMA 2.2. *Assume that that the solution of* (2.1) *is smooth and that the forms of* (2.5) *and the operators* $\pi_V$, $\pi_W$ *are such that* (2.9), (2.11), *and* (2.13) *are satisfied. Then for* $u_h, z_h$ *solution of* (2.5) *there holds*

$$|\pi_V u - u_h|_{S_p} + |\pi_W z - z_h|_{S_a} \lesssim c_{a,\gamma,\epsilon} h^k |u|_{H^{k+1}(\Omega)} + \epsilon(h)\|z_h\|,$$

*where* $c_{a,\gamma,\epsilon} = c_{a,\gamma}(1+c_a^2)^{\frac{1}{2}}$ *with* $c_a$ *and* $c_{a,\gamma}$ *defined by* (2.11) *and* (2.13), *respectively. Similarly, if* $s_p(u, w + w_h) = 0$, *for all* $w \in W$ *and* $w_h \in W_h$, *there holds*

$$|u_h|_{S_p} + |z_h|_{S_a} \lesssim (c_{a,\gamma} + c_{a,\gamma,\epsilon}) h^k |u|_{H^{k+1}(\Omega)} + \epsilon(h)\|z_h\|.$$

*Proof.* Let $\xi_h = \pi_V u - u_h$ and $\zeta_h = \pi_W z - z_h$. By definition (2.5) there holds

$$|\xi_h|_{S_p}^2 + |\zeta_h|_{S_a}^2 = s_p(\xi_h,\xi_h) + s_a(\zeta_h,\zeta_h) = a_h(\xi_h,\zeta_h) + s_a(\zeta_h,\zeta_h) - a_h(\xi_h,\zeta_h) + s_p(\xi_h,\xi_h).$$

Using now the Galerkin orthogonality of $a_h(\cdot,\cdot)$, (2.9), we have

$$|\xi_h|_{S_p}^2 + |\zeta_h|_{S_a}^2 = a_h(\pi_V u - u, \zeta_h) + s_a(\pi_W z, \zeta_h) - a_h(\xi_h, \pi_W z - z) + s_p(\pi_V u - u, \xi_h).$$

Observing that $z = \pi_W z = 0$ this reduces to

$$|\xi_h|_{S_p}^2 + |\zeta_h|_{S_a}^2 = a_h(\pi_V u - u, \zeta_h) + s_p(\pi_V u - u, \xi_h).$$

We conclude by applying the continuity (2.11)

$$|\xi_h|_{S_p}^2 + |\zeta_h|_{S_a}^2 \le \|u-\pi_V u\|_+(c_a|\zeta_h|_{S_a} + \epsilon(h)\|z_h\|) + |u-\pi_V u|_{S_p}|\xi_h|_{S_p}$$

followed by an arithmetic-geometric inequality and the approximability (2.13)

$$\begin{aligned} |\xi_h|_{S_p}^2 + |\zeta_h|_{S_a}^2 &\le (c_a^2+1)\|u-\pi_V u\|_+^2 + \epsilon(h)^2\|z_h\|^2 + |u-\pi_V u|_{S_p}^2 \\ &\le c_{a,\gamma}^2(1+c_a^2)h^{2k}|u|_{H^{k+1}(\Omega)}^2 + \epsilon(h)^2\|z_h\|^2. \end{aligned}$$

The second result follows by adding and subtracting $\pi_V u$, observing that here $\pi_W z = 0$, applying a triangle inequality and then (2.13) on $|\pi_V u|_{S_p} = |\pi_V u - u|_{S_p}$. ∎

We may now prove the main result which is optimal convergence in the $L^2$ and the $H^1$ norms.

THEOREM 2.3. *Assume that* (2.1) *and* (2.2) *are well-posed with exact solutions* $u$ *and* $z$ *satisfying* (2.3)*. Assume that the forms of* (2.5) *and the operators* $\pi_V$*,* $\pi_W$ *are such that* (2.9)–(2.15) *are satisfied and that* $h$ *is so small that*

$$c_{a,\gamma,\Omega,} h\, \epsilon(h) \leq \frac{1}{6}, \tag{2.16}$$

*where* $c_{a,\gamma,\Omega}$ *depends on the constants of the inequalities* (2.3) *and* (2.13)–(2.15)*. Then* (2.5) *admits a unique discrete solution* $u_h$*,* $z_h$ *that satisfies*

$$\|u - u_h\| + h\|\nabla(u - u_h)\| + \|z_h\| \leq C_{a,\Omega,\gamma} h^{k+1} |u|_{H^{k+1}(\Omega)}$$

*and in particular*

$$\|u - u_h\| + h\|\nabla(u - u_h)\| + \|z_h\| \leq C_{a,\Omega,\gamma} h^2 \|f\|. \tag{2.17}$$

*Proof.* Let $\varphi$ be the solution of (2.2) with $g = u - u_h$ and $\psi$ the solution of (2.1) with $f = z_h$. By (2.3) there holds

$$\|\varphi\|_{H^2(\Omega)} \leq c_{a,\Omega}\|u - u_h\| \text{ and } \|\psi\|_{H^2(\Omega)} \leq c_{a,\Omega}\|z_h\|.$$

By definition of the primal and dual problems and by (2.7), (2.8), (2.9), (2.11), and (2.12) there holds

$$\begin{aligned}
\|u - u_h\|^2 + \|z_h\|^2 &= (u - u_h, \mathcal{L}^*\varphi) + (\mathcal{L}\psi, z_h) = a_h(u - u_h, \varphi) + a_h(\psi, z_h)\\
&= a_h(u - u_h, \varphi - \pi_W\varphi) + s_a(z_h, \pi_W\varphi)\\
&\quad + a_h(\psi - \pi_V\psi, z_h) - s_p(u - u_h, \pi_V\psi)\\
&\leq (c_a\|u - \pi_V u\|_{\mathcal{L}} + c_a|u_h - \pi_V u|_{S_p} + \epsilon(h)\|u - u_h\|)\|\varphi - \pi_W\varphi\|_*\\
&\quad + (c_a|z_h|_{S_a} + \epsilon(h)\|z_h\|)\|\psi - \pi_V\psi\|_+\\
&\quad + |z_h|_{S_a}|\pi_W\varphi|_{S_a} + |u_h - u|_{S_p}|\pi_V\psi|_{S_p}.
\end{aligned}$$

First we observe that by (2.13), (2.14), and (2.3)

$$\begin{aligned}
&\epsilon(h)\|u - u_h\|\|\varphi - \pi_W\varphi\|_* + \epsilon(h)\|z_h\|\|\psi - \pi_V\psi\|_+\\
&\quad \leq c_{a,\gamma,\Omega} h(\epsilon(h)\|u - u_h\|^2 + \epsilon(h)\|z_h\|^2).
\end{aligned}$$

Then by Lemma 2.2 and the upper bounds (2.15) we have

$$\begin{aligned}
&|z_h|_{S_a}|\pi_W\varphi|_{S_a} + |u_h - u|_{S_p}|\pi_V\psi|_{S_p}\\
&\quad \lesssim ((c_{a,\gamma} + c_{a,\gamma,\epsilon})h^k|u|_{H^{k+1}(\Omega)} + \epsilon(h)\|z_h\|)c_{a,\gamma,\Omega} h(\|u - u_h\| + \|z_h\|).
\end{aligned}$$

Using the two previous bounds and an arithmetic-geometric inequality we have

$$(1 - 3c_{a,\gamma,\Omega} h\epsilon(h))(\|u - u_h\|^2 + \|z_h\|^2) \leq C_{a,\gamma} h^{k+1}|u|_{H^{k+1}(\Omega)}(|\varphi|_{H^2(\Omega)} + |\psi|_{H^2(\Omega)}).$$

Using (2.3), the result for the $L^2$-norm follows provided $h$ satisfies (2.16). The result for the $H^1$-norm follows using a global inverse inequality on the discrete error and then the $L^2$-norm error estimate.

$$\|\nabla(u - u_h)\| \leq \|\nabla(u - \pi_V u)\| + \|\nabla(\pi_V u - u_h)\| \lesssim h^k|u|_{H^{k+1}(\Omega)} + h^{-1}\|\pi_V u - u_h\|.$$

The existence of a unique solution to (2.5) is a consequence of (2.17). Well-posedness of (2.1) means that $f = 0$ implies $u = 0$, but then by (2.17) $u_h = z_h = 0$, which shows that the matrix is invertible. □

The optimal convergence of the stabilization terms follows.

COROLLARY 2.4. *Under the assumptions of Lemma* 2.2 *and Theorem* 2.3 *there holds*

$$|\pi_V u - u_h|_{S_p} + |\pi_W z - z_h|_{S_a} \lesssim c_{s,\epsilon} h^k |u|_{H^{k+1}(\Omega)} + O(h^{k+1}).$$

*Proof.* The proof is an immediate consequence of Lemma 2.2 and Theorem 2.3. □

*Remark* 2.5. The need to control a low order contribution of the dual solution $z_h$ above usually comes from oscillation of data, either in the form of stabilization terms that do not account for oscillation within the element or error in the numerical quadrature.

In case a Gårdings inequality holds for (2.5) and $s_a(\cdot,\cdot) \equiv s_p(\cdot,\cdot)$ the $H^1$-error can be recovered without using inverse inequalities as stated below.

COROLLARY 2.6. *Assume that for the bilinear form $a(\cdot,\cdot)$ there exists $\lambda \in \mathbb{R}$ such that*

$$\|\nabla v_h\|^2 - \lambda \|v_h\|^2 \lesssim a_h(v_h, v_h) + s_p(v_h, v_h)$$

*and that $s_a(\cdot,\cdot) \equiv s_p(\cdot,\cdot)$. Then*

$$\|\nabla(u - u_h)\| \lesssim h^k |u|_{H^{k+1}(\Omega)}.$$

*Proof.* The proof is similar to the proof of Lemma 2.2 and therefore is only sketched. Let $\xi_h := \pi_V u - u_h$. It follows by the Gårdings inequality that

$$\|\nabla \xi_h\|^2 \lesssim a_h(\xi_h, \xi_h) + \lambda \|\xi_h\|^2 + s_p(\xi_h, \xi_h).$$

Using Galerkin orthogonality we have

$$a_h(\xi_h, \xi_h) = a_h(\pi_V u - u, \xi_h) + s_a(z_h, \xi_h)$$

and the rest follows as in Lemma 2.2 by (2.11), (2.13), and using the known convergences of Lemma 2.2 and Theorem 2.3. □

## 3. Stabilization methods.

To fix the ideas let $\mathcal{L}$ be a second order elliptic operator on conservation form,

$$\mathcal{L}u := -\mu \Delta u + \nabla \cdot (\beta u) + cu. \tag{3.1}$$

Here $\mu \in \mathbb{R}^+$, $\beta \in [C^2(\Omega)]^2$ is a nonsolenoidal velocity vectorfield and $c \in C^1(\Omega)$. Formally, the corresponding bilinear form is written

$$a(u, v) := (\mu \nabla u, \nabla v) + (\nabla \cdot (\beta u) + cu, v), \quad u, v \in H^1(\Omega). \tag{3.2}$$

The continuities (2.11) and (2.12) suggest the following design criteria on the stabilization operators:

$$\inf_{w_h \in V_h} \|h(\mathcal{L} v_h - w_h)\|_h^2 + \|h^{\frac{1}{2}} \mu^{\frac{1}{2}} [\![\nabla v_h \cdot n_F]\!]\|_{\mathcal{F}_{int}}^2 \lesssim s_p(v_h, v_h), \tag{3.3}$$

$$\inf_{w_h \in V_h} \|h(\mathcal{L}^* x_h - w_h)\|_h^2 + \|h^{\frac{1}{2}} \mu^{\frac{1}{2}} [\![\nabla x_h \cdot n_F]\!]\|_{\mathcal{F}_{int}}^2 \lesssim s_a(x_h, x_h) \tag{3.4}$$

at least up to a nonessential low order perturbation. If we neglect terms due to boundary conditions we may apply an integration by parts in the left-hand side of (2.11), leading to

$$a_h(v - \pi_V v, x_h) = \langle u - \pi_V v, [\![\mu \nabla x_h \cdot n_F]\!] \rangle_{\mathcal{F}_{int}} + (v - \pi_V v, \mathcal{L}^* x_h)_h.$$

Using a suitable weighting in $h$ and applying the Cauchy–Schwarz inequality justifies (3.4). The function $w_h$ may be included provided the interpolant $\pi_V$ has suitable orthogonality properties. It can be noted that one may also construct the interpolant with orthognality properties on the element faces, so that the influence of the gradient jump term may be reduced; we will not pursue this possibility herein. The choice $w_h = 0$ in the first term in the left-hand side of (3.3) results in a least squares term on the (homogeneous) residual over the element. It follows that the stabilization relies on two mechanisms: $L^2$-control of the element residual and $L^2$-control of the gradient jumps over element edges. If higher order differential equations are considered, jumps of higher derivatives must be added. The design criterion (3.3)–(3.4) makes it straightforward to adapt the analysis below to a range of stabilization methods, such as GLS, orthogonal subscales, CIP, or discontinuous Galerkin methods. For the discontinuous Galerkin method the penalty must act on the jump of $u_h$ itself and on the jump of the normal gradient. In all cases, however, the jumps of the gradient must be penalized, or an equivalent stabilization operator introduced. It therefore seems natural to consider two stabilizations in more detail, first the GLS stabilization combined with gradient penalty and then a CIP stabilization purely based on penalty on jumps of derivatives of the approximate solution. We introduce the stabilization operators

$$s_p(u_h, v_h) := s_{p,GLS}(u_h, v_h) + s_{cip}(u_h, v_h) \tag{3.5}$$

and

$$s_a(z_h, w_h) := s_{a,GLS}(z_h, w_h) + s_{cip}(u_h, v_h), \tag{3.6}$$

where

$$s_{p,GLS}(u_h, v_h) := (\gamma_{GLS} h^2 \mathcal{L} u_h, \mathcal{L} v_h)_h,$$
$$s_{a,GLS}(z_h, w_h) := (\gamma_{GLS} h^2 \mathcal{L}^* z_h, \mathcal{L}^* w_h)_h,$$

and

$$s_{cip}(u_h, v_h) := \sum_{F \in \mathcal{F}_{int}} \int_F (h_F \gamma_{1,F} [\![\nabla u_h]\!] \cdot [\![\nabla v_h]\!] + h_F^3 \gamma_{2,F} [\![\Delta u_h]\!] [\![\Delta v_h]\!]) \, dx. \tag{3.7}$$

Here $[\![\Delta u_h]\!]|_F$ denote the Laplacian, over the face $F$. Note that for smooth $u$, $s_{cip}(u, v_h) = 0$ and hence $s_p(u, v_h) = s_{p,GLS}(u, v_h) = (f, \gamma_{GLS} h^2 L v_h)_h$, showing that $s_p(u, v_h)$ is known. The abstract analysis typically holds for the parameter choices $\gamma_{GLS} > 0$, $\gamma_{1,F} > 0$, $\gamma_{2,F} = 0$ or $\gamma_{GLS} = 0$, $\gamma_{1,F} > 0$, $\gamma_{2,F} > 0$. Note that the matrix stencil for finite element methods remains the same for both approaches, and therefore the CIP method seems more appealing in this context. Eliminating the GLS term also reduces the computational effort since the same stabilization is used for the primal and adjoint solution. If on the other hand a $C^1$-continuous approximation space is used, the jumps of the gradients may be omitted and the GLS stabilization might prove competitive, since integrations on the faces may then be avoided. Below we will only consider the case where $V_h = W_h := X_h^k$ or some subset thereof, which will then be defined in each case.

**3.1. GLS stabilization.** The GLS method is one of the most popular stabilized methods. To fix the ideas we will assume that problems (2.1) and (2.2) are subject to homogeneous Dirichlet conditions and well-posed, with $f \in L^2(\Omega)$. For the reader's convenience we detail the Lagrangian (2.6) in this particular case,

$$\begin{aligned}(3.8)\qquad Ł(u_h, z_h) := \frac{1}{2}\|\tau^{\frac{1}{2}}(\mathcal{L}u_h - f)\|_h^2 + \frac{1}{2}s_{cip}(u_h, u_h)\\ - \frac{1}{2}\|\tau^{\frac{1}{2}}\mathcal{L}^* z_h\|_h^2 - \frac{1}{2}s_{cip}(z_h, z_h) - a(u_h, z_h) + (f, z_h).\end{aligned}$$

We let $V_h = W_h := X_h^k \cap H_0^1(\Omega)$. The optimality conditions are then written as follows: find $u_h, z_h \in V_h \times W_h$

$$\begin{aligned}(3.9)\qquad a(u_h, w_h) + s_a(z_h, w_h) = (f, w_h),\\ a(v_h, z_h) - s_p(u_h, v_h) = -s_p(u, v_h) = -(f, \tau\mathcal{L}v_h)_h\end{aligned}$$

for all $v_h, w_h \in V_h \times W_h$. Here $\gamma_{GLS}h^2 =: \tau > 0$ and $\gamma_{1,F} \sim \mu$, $\gamma_{2,F} = 0$. We assume that the physical parameters are all order unity for simplicity. Observe the nonstandard structure of the stabilization terms and that the formulation is consistent for $u$ the exact solution of (2.1) and $z = 0$. We will now prove that the assumptions of Proposition 2.2 and Theorem 2.3 are satisfied for formulation (3.9).

We define the following seminorms:

$$(3.10)\qquad \|v\|_+ := \|v\|_* := \|\tau^{-\frac{1}{2}}v\| + \|\mu^{\frac{1}{2}}h^{-\frac{1}{2}}v\|_{\mathcal{F}_{int}}$$

and

$$\|v\|_{\mathcal{L}} := |x|_{S_p} := \|\tau^{\frac{1}{2}}\mathcal{L}x\|_h + s_{cip}(x,x)^{\frac{1}{2}} \text{ and } |x|_{S_a} := \|\tau^{\frac{1}{2}}\mathcal{L}^*x\|_h + s_{cip}(x,x)^{\frac{1}{2}},$$

defined for $x \in H^2(\Omega) + V_h$. Let $\pi_V$ and $\pi_W$ be defined by the Lagrange interpolator $i_h$ (or any other $H^2$-stable interpolation operator that satisfies boundary conditions), and note that by (2.4) we readily deduce the approximation results for smooth enough functions $u$

$$\|u-\pi_V u\|_+ + \|u-\pi_V u\|_{\mathcal{L}} + |u-\pi_V u|_{S_p} + \|u-\pi_W u\|_* + |u-\pi_W u|_{S_a} \leq c_{a,\gamma}h^k|u|_{H^{k+1}(\Omega)}$$

and, by $H^2$-stability of the interpolation operator,

$$|\pi_V v|_{S_p} \leq c_{\gamma,a}h\|v\|_{H^2(\Omega)}, \quad |\pi_W w|_{S_a} \leq c_{\gamma,a}h\|w\|_{H^2(\Omega)} \quad \forall v, w \in H^2(\Omega).$$

This shows that (2.13) and (2.14) hold. It then only remains to show the continuities (2.11) and (2.12). First we show the inequality (2.11) For the second order elliptic problem we note that after an integration by parts and Cauchy–Schwarz inequality,

$$\begin{aligned}a_h(v - \pi_V v, x_h) &= \sum_{F \in \mathcal{F}_{int}} \langle u - \pi_V u, [\![\mu\nabla x_h \cdot n_F]\!]\rangle_F + (u - \pi_V u, \mathcal{L}^* x_h)_h\\ &\leq \|u - \pi_V u\|_+ |x_h|_{S_a}.\end{aligned}$$

Similarly, to prove (2.12) we integrate by parts in the opposite direction in the second order operator and obtain

$$\begin{aligned}a_h(u - u_h, y - \pi_W y) = &\sum_{K \in \mathcal{T}_h} (\mathcal{L}(u - u_h), y - \pi_W y)_K\\ &+ \sum_{F \in \mathcal{F}_{int}} \langle [\![\mu\nabla u_h \cdot n_F]\!], y - \pi_W y\rangle_F\\ &\times \|y - \pi_W y\|_*(\|u - \pi_V u\|_{\mathcal{L}} + |u_h - \pi_V u|_{S_p}).\end{aligned}$$

*Remark* 3.1. Note that for the GLS method $\epsilon(h) = 0$ in (2.11) and (2.12). This follows from the fact that the whole residual is considered in the stabilization term. This nice feature, however, holds only under exact quadrature. When the integrals are approximated, the quadrature error may give rise to oscillation terms from data that introduces a nonzero contribution to $\epsilon(h)$.

*Remark* 3.2. Since the exact adjoint solution is zero for the problems considered here one can use simplified forms of the adjoint stabilization, without sacrificing consistency. Observe however that the simplest form obtained by taking $s_a(\cdot,\cdot)$ to be the $H^1$-scalar product, does not satisfy (2.15) and hence does not allow for optimal error estimates.

**3.2. CIP.** Since in this case we must account for possible oscillation of the physical coefficients we postpone the detailed analysis to the examples below and here only discuss the general principle. In this case we use $\gamma_{GLS} = 0$, $\gamma_{1,F} > 0$, $\gamma_{2,F} > 0$ in the general expressions for the stabilization (3.5) and (3.6). The parameters $\gamma_{i,F}$, $i = 1, 2$, are stabilization coefficients, the form of which will be problem specific and will be given for each problem below. The key observation is that the following discrete approximation result holds for suitably chosen $\gamma_{i,F}$ in $s_{cip}(\cdot,\cdot)$ (see [6, 7]):

$$\|h^{\frac{1}{2}}(\beta_h \cdot \nabla u_h - I_{os}\beta_h \cdot \nabla u_h)\|^2 + \sum_K \|h\mu(\Delta u_h - I_{os}\Delta u_h)\|^2 \leq s_{cip}(u_h, u_h). \tag{3.11}$$

Here $\beta_h$ is some piecewise affine interpolant of the velocity vector field $\beta$ and $I_{os}$ is the quasi-interpolation operator defined in each node of the mesh as a straight average of the function values from simplices sharing that node (see [7]). For example,

$$(I_{os}\Delta u_h)(x_i) = N_i^{-1} \sum_{\{K: x_i \in K\}} \Delta u_h(x_i)|_K$$

with $N_i := \text{card}\{K : x_i \in K\}$. Using (3.11) one may prove that

$$\inf_{v_h \in V_h} \|h(\mathcal{L}u_h - v_h)\|_h \lesssim s_{cip}(u_h, u_h)^{\frac{1}{2}} + \epsilon(h)\|u_h\|. \tag{3.12}$$

It immediately follows that (3.3) and (3.4) are satisfied. We will leave the discussion of (2.11)–(2.14) and (3.12) to the applications below, giving the explicit form for $\epsilon(h)$ for each case. Here we instead proceed with an abstract analysis, assuming that all physical parameters are of order $O(1)$. We choose $\pi_V$ and $\pi_W$ as the $L^2$-projection in order to exploit orthogonality to "filter" the element residual. Let $\|\cdot\|_+$ and $\|\cdot\|_*$ have the same definition as in the GLS case and define

$$\|u\|_{\mathcal{L}} := \|h\mathcal{L}u\|_h + \|h^{\frac{1}{2}}\mu^{\frac{1}{2}}[\![\nabla u \cdot n_F]\!]\|_{\mathcal{F}_{int}} + \epsilon(h)\|u\|. \tag{3.13}$$

Then we proceed similarly as for GLS, but we use the orthogonality of the $L^2$-projection, ignoring here the contribution from boundary terms. It then follows using the orthogonality of the projection that formally

$$\begin{aligned} a_h(v - \pi_V v, x_h) &= \sum_{F \in \mathcal{F}_{int}} \langle u - \pi_V u, [\![\mu \nabla x_h \cdot n_F]\!] \rangle_F + (u - \pi_V u, \mathcal{L}^* x_h - w_h)_h \\ &\leq \|u - \pi_V u\|_+ (|x_h|_{S_a} + \epsilon(h)\|x_h\|). \end{aligned}$$

Similarly, to prove (2.12) we integrate by parts in the opposite direction in the second order operator and use the $L^2$-orthogonality to obtain

$$\begin{aligned} a_h(u-u_h, y-\pi_W y) &= \sum_{K\in\mathcal{T}_h} (\mathcal{L}(u-u_h), y-\pi_W y)_K \\ &\quad + \sum_{F\in\mathcal{F}_{int}} \langle [\![\mu\nabla u_h\cdot n_F]\!], y-\pi_W y\rangle_F \\ &= \sum_{K\in\mathcal{T}_h} (\mathcal{L}(u-\pi_V u) + \mathcal{L}(\pi_V u - u_h) - w_h, y-\pi_W y)_K \\ &\quad + \sum_{F\in\mathcal{F}_{int}} \langle [\![\mu\nabla u_h\cdot n_F]\!], y-\pi_W y\rangle_F \\ &\le \|y-\pi_W y\|_*(\|h\mathcal{L}(u-\pi_V u)\|_h \\ &\quad + |\pi_V u - u_h|_{S_p} + \epsilon(h)\|\pi_V u - u_h\|) \\ &\le \|y-\pi_W y\|_*(\|u-\pi_V u\|_{\mathcal{L}} + |u_h - \pi_V u|_{S_p} + \epsilon(h)\|u-u_h\|). \end{aligned}$$

The last inequality follows by adding and subtracting $u$ in the last norm in the right-hand side to obtain $\epsilon(h)\|\pi_V u - u + u - u_h\|$. This term is then split using a triangular inequality and the approximation error integrated in the $\|\cdot\|_{\mathcal{L}}$ term. To use the $L^2$-projection in this fashion we must impose the boundary conditions weakly so that the boundary degrees of freedom are included in $V_h$. For the GLS method one has the choice between weak and strong imposition of boundary condition. In the next section we will discuss how weakly imposed boundary conditions are included in the formulation (2.5).

**3.3. Imposition of boundary conditions.** To impose boundary conditions weakly in this framework we propose a Nitsche-type method. However, our formulation differs from the standard Nitsche boundary conditions in several ways:

- Both Dirichlet and Neumann conditions are imposed using a penalty.
- There is no lower bound of the parameter for the imposition of Dirichlet-type boundary conditions. This is related to the fact that the method never uses the coercivity of $a_h(\cdot,\cdot)$.
- Nitsche-type boundary terms are added to $a_h(\cdot,\cdot)$ in order to ensure consistency and adjoint consistency, but the penalty is added to the operators $s_p(\cdot,\cdot)$ and $s_a(\cdot,\cdot)$, allowing for a different boundary penalty for the primal and the adjoint. As we shall see below for some problems this is the only way to make the Nitsche formulation consistent.

If the primal and the dual problems have a Dirchlet boundary condition on $\Gamma_D$ this is imposed by

$$a_h(u_h, v_h) := a(u_h, v_h) - \langle \mu\nabla u_h\cdot n, v_h\rangle_{\Gamma_D} - \langle \mu\nabla v_h\cdot n, u_h\rangle_{\Gamma_D},$$

where $a(\cdot,\cdot)$ is defined by (3.2) and by adding the boundary penalty term

$$\int_{\Gamma_D} \gamma_D \mu h^{-1} u_h v_h \,\mathrm{d}s \tag{3.14}$$

to $s_p(\cdot,\cdot)$ and $s_a(\cdot,\cdot)$ with $\gamma_D > 0$. In the nonhomogeneous case the suitable data is added to the right-hand side in the standard way. For Neumann conditions on $\Gamma_N$ in both the primal and the adjoint problems, these are introduced in the standard way in $a(u,v)$ with a suitable modification of the right-hand side of (2.1). No modification is introduced in $a_h(\cdot,\cdot)$ but the following penalty is added to $s_p(\cdot,\cdot)$ and $s_a(\cdot,\cdot)$ with $\gamma_N > 0$:

$$\int_{\Gamma_N} \gamma_N h \nabla u_h \cdot n \nabla v_h \cdot n \, \mathrm{d}s. \tag{3.15}$$

If the boundary conditions for $u$ are nonhomogeneous the usual data contributions are introduced in the right-hand side $-s_p(u, w_h)$.

As mentioned in the introduction the seminorm $|\cdot|_S$ can be a norm in certain situations so that the partial coercivity (2.10) implies the well-posedness of the linear system (2.5). In the following proposition we discuss some basic sufficient conditions for the matrix to be invertible in the case of piecewise affine approximation spaces. For particular cases other geometric arguments may prove fruitful, as we shall see in the second example below.

PROPOSITION 3.3. *The kernel of the linear system defined by* (2.5) *with the stabilization* (3.7) *has dimension at most* $2(d+1)$ *for* $k = 1$. *The system* (2.5) *admits a unique solution if the boundary conditions satisfy one of the following conditions:*

1. *two nonparallel polygon sides one subject to Dirichlet boundary conditions,*
2. *two nonorthogonal polygon sides one subject to a Dirichlet boundary condition and the other to a Neumann condition imposed using* (3.15)*,*
3. $d$ *nonparallel polygon sides subject to Neumann conditions imposed using* (3.15) *and either* $1 \notin V_h$ *or there exists* $v_h, w_h \in V_h$ *such that* $a_h(1, v_h) \neq 0$ *and* $a_h(w_h, 1) \neq 0$.

*Proof.* It is immediate from (2.10) that the kernel of the system matrix of (2.5) cannot be larger than the sum of the dimensions of the kernels of $s_p(\cdot,\cdot)$ and $s_a(\cdot,\cdot)$. For $s_{cip}(\cdot,\cdot)$ and $k = 1$ the kernel is identified as $[\mathbb{P}_1(\Omega)]^2$ with dimension $2(d+1)$.

To prove well-posedness of the linear system it is enough to prove uniqueness; we assume that $f = 0$ and prove that then $u_h \equiv z_h \equiv 0$.

If the Dirichlet boundary condition is imposed on a boundary, then the gradient must be zero in the tangential direction to this boundary; since the tangents of two boundaries span $\mathbb{R}^d$ we conclude that $f = 0$ in (2.5) implies $u_h = 0$ due to (2.10) and similarly $z_h = 0$ and the matrix is invertible.

In the second case, the function is zero on the Dirichlet boundary and the gradient is zero in the tangential directions of the Dirichlet boundary condition, eliminating $d$ elements in the kernel. The penalty on the Neumann boundary, being nonorthogonal to the Dirichlet boundary, cancels the remaining free gradient. The same argument leads to both $u_h = 0$ and $z_h = 0$.

For the third case we observe that the term (3.15) acting on $d$ nonparallel polygon sides implies that $\nabla u_h = 0$ and well-posedness is then immediate by the remaining conditions. ∎

*Remark* 3.4. Observe that Proposition 3.3 holds for any bilinear form $a(\cdot,\cdot)$, even strongly degenerate ones.

**4. Applications.** We will now give three examples of problems that enter the abstract framework. The first two problems we introduce below have well-posed primal and adjoint problems so that the above theory applies. For each method we will propose a formulation and prove that the relations (2.7), (2.8) hold. We only consider the CIP method in the examples below. In the first example we detail the dependence of physical parameters in all norms and coefficients and choose stabilization parameters to allow for high Péclet number flows. Due to the use of the duality argument, however, the present analysis is restricted to the case of moderate Péclet numbers.

In the later examples we assume that all physical parameters are unity and do not track the dependence. As suggested above we take $\pi_V \equiv \pi_W \equiv \pi_L$. In each case we will detail the form of $\epsilon(h)$. In the last case, the elliptic Cauchy problem, the stability properties of the problem strongly depend on the geometry of the problem and the assumption of well-posedness does not hold in general. We will nevertheless propose a method that satisfies the assumptions of the general theory and then study its performance numerically.

**4.1. Nonsymmetric indefinite elliptic problems.** Our first examples consist of a convection-diffusion-reaction problem with nonsolenoidal velocity field, as is the case for reactive transport in compressible flow. We first consider the case of homogeneous Dirichlet conditions where the analysis of [19] applies. Then we consider the case where failure of the coercivity is due also to the boundary condition; here we study a convection-diffusion equation with homogeneous Neumann boundary conditions. We will detail only how the analysis of this case differs from the Dirichlet case. For a detailed analysis of the well-posedness of the continuous problems we refer to [10, 12] and for a finite element analysis in the case of homogeneous Neumann conditions to [8]. Recent work on numerical methods for these problems has focused on finite volume methods [11, 9] or hybrid finite element/finite volume methods [15].

**4.1.1. Reactive transport in compressible flow: Dirichlet conditions.** In combustion problems, for example, it is important to accurately compute the transport of the reacting species in the compressible flow. We suggest a scalar model problem of convection-diffusion type with a linear reaction term $cu$, where the reaction can have arbitrary sign:

$$\begin{aligned} \mathcal{L}u &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{4.1}$$

The dual adjoint takes the form

$$\begin{aligned} \mathcal{L}^* z := -\mu\Delta z - \beta \cdot \nabla z + cz &= g \quad \text{in } \Omega, \\ z &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{4.2}$$

The variational formulation (2.1) is obtained by taking $V = W := H_0^1(\Omega)$ and $a(\cdot,\cdot)$ defined by (3.2). We assume that $f, g \in L^2(\Omega)$, that both (4.1) and (4.2) are well-posed in $H_0^1(\Omega)$ by the Fredholm alternative, and that the smoothing property (2.3) holds. See [10] for an analysis of existence and uniqueness under weaker regularity assumptions on $\beta$ and $c$ with $c \geq 0$. The below analysis can also be carried out assuming less regularity, but the constraints on the mesh size for the error estimate to hold will be stronger. Recall that the constants in the estimate (2.3) also depend on the regularity of the coefficients. The discrete form of the bilinear form is given by

$$a_h(u_h, v_h) := a(u_h, v_h) - \langle \nabla u_h \cdot n, v_h \rangle_{\partial\Omega} - \langle \nabla v_h \cdot n, u_h \rangle_{\partial\Omega} - \langle (\beta \cdot n)_- u_h, v_h \rangle_{\partial\Omega}, \tag{4.3}$$

where $(\beta \cdot n)_\pm := \frac{1}{2}(\beta \cdot n \pm |\beta \cdot n|)$. We define the approximation spaces $V_h = W_h := X_h^k$. The stabilization is chosen as

$$s_p(u_h, v_h) := s_{cip}(u_h, v_h) + s_{bc}^-(u_h, v_h) \tag{4.4}$$

and

$$s_a(z_h, v_h) := s_{cip}(z_h, v_h) + s_{bc}^+(z_h, v_h), \tag{4.5}$$

where $\gamma_{1,F} \sim (\mu+\|\beta_h \cdot n_F\|_{\infty,F} h_F)$ and $\gamma_{2,F} \sim \mu$ in (3.7) with $\beta_h$ the nodal interpolant of $\beta$ and

$$s_{bc}^{\pm}(x_h, v_h) := \langle \mu h^{-1} x_h, v_h \rangle_{\partial\Omega} + \langle |(\beta \cdot n)_{\pm}| x_h, v_h \rangle_{\partial\Omega} . \tag{4.6}$$

If only the low Péclet regime is considered the second term of (4.6) is always dominated by the first and may therefore be omitted.

PROPOSITION 4.1 (existence of discrete solutions). *Let $k = 1$. Then the formulation* (2.5) *with the bilinear form* (4.3) *and the stabilization* (4.4)–(4.5) *admits a unique solution $u_h \in V_h$.*

*Proof.* The proof is immediate by Proposition 3.3. □

It is well known that the bilinear form (4.3) satisfies the consistency relations (2.7) and (2.8) and that the stabilization (4.4)–(4.5) satisfies the upper bounds (2.13), (2.14), and (2.15). Now we define the norms by

$$\|v\|_+ := \|v\|_* := \|\mu^{\frac{1}{2}} h^{-\frac{1}{2}} v\|^2_{\mathcal{F}_{int}} + \|(\zeta_{Pe} + h^{-1})v\| + \|h^{\frac{1}{2}} \mu^{\frac{1}{2}} \nabla v\|_{\partial\Omega} + \|\beta_\infty^{\frac{1}{2}} v\|_{\partial\Omega}.$$

Here $\zeta_{Pe} := (\beta_\infty^{\frac{1}{2}} h^{-\frac{1}{2}} + \mu^{\frac{1}{2}} h^{-1} + c_\infty^{\frac{1}{2}})$ with $\beta_\infty = \|\beta\|_{L^\infty(\Omega)}$ and $c_\infty := \|c\|_{L^\infty(\Omega)}$. Also define

$$\|v\|_{\mathcal{L}} := \|\mu^{\frac{1}{2}} h \Delta v\|_h + \|\beta_\infty^{-\frac{1}{2}} h^{\frac{1}{2}} \beta \cdot \nabla v\| + \|c_\infty^{\frac{1}{2}} v\| + \|\mu^{\frac{1}{2}} h^{\frac{1}{2}} [\![\nabla v \cdot n_F]\!]\|_{\mathcal{F}_{int}} + \|(\mu^{\frac{1}{2}} h^{-\frac{1}{2}} + \beta_\infty^{\frac{1}{2}}) v\|_{\partial\Omega} + \epsilon(h)\|v\|.$$

It is straightforward to show that

$$\|u - \pi_V u\|_+ + \|u - \pi_W u\|_* \lesssim (\zeta_{Pe} + h^{-1}) h^{k+1} |u|_{H^{k+1}(\Omega)}$$

and (for simplicity with $\epsilon(h) = 0$)

$$\|u - \pi_V u\|_{\mathcal{L}} \lesssim \zeta_{Pe} h^{k+1} |u|_{H^{k+1}(\Omega)}.$$

It then only remains to prove the continuities (2.11) and (2.12) to conclude that Theorem 2.3 holds.

PROPOSITION 4.2. *The bilinear form* (4.3) *satisfies the continuities* (2.11) *and* (2.12) *with*

$$\epsilon(h) \sim h^2(|\beta|_{W^{2,\infty}} + |c|_{W^{1,\infty}}).$$

*Proof.* First we consider (2.11). After an integration by parts in $a(\cdot, \cdot)$ we have

$$a_h(u - \pi_V u, x_h) = \sum_{F \in \mathcal{F}_{int}} \langle u - \pi_V u, [\![\mu \nabla x_h \cdot n_F]\!] \rangle_F + (u - \pi_V u, \mathcal{L}^* x_h)_h - \langle u - \pi_V u, (\beta \cdot n)_+ x_h \rangle_{\partial\Omega} = \mathrm{I} + \mathrm{II} + \mathrm{III}.$$

Considering I–III we find using the Cauchy–Schwarz inequality

$$\mathrm{I} + \mathrm{III} \leq \|u - \pi_V u\|_+ |x_h|_{S_a}.$$

For II, using the discrete interpolation results (3.11), the discrete commutator property (see [2]), and the standard approximation followed by an inverse inequality in the last term,

$$\begin{aligned} \mathrm{II} &= (u - \pi_V u, -i_h \beta \cdot \nabla x_h + \mathrm{I}_{os}(i_h \beta_h \cdot \nabla x_h) - \mu \Delta x_h + \mathrm{I}_{os} \mu \Delta x_h)_h \\ &\quad + (u - \pi_V u, c x_h - i_h(c x_h)) + (u - \pi_V u, (\beta - i_h \beta) \cdot \nabla x_h) \\ &\leq \|u - \pi_V u\|_+ (c_{os} |x_h|_{S_a} + c_i h^2 (|\beta|_{W^{2,\infty}} + |c|_{W^{1,\infty}}) \|x_h\|). \end{aligned}$$

The second continuity follows in a similar fashion,

$$\begin{aligned} a_h(u-u_h, y-\pi_W y) &= (\mathcal{L}(u-u_h), y-\pi_W y)_h + \sum_{F\in\mathcal{F}_{int}} \langle [\![\mu\nabla u_h\cdot n_F]\!], y-\pi_W y\rangle_F \\ &\quad + \langle(\beta\cdot n)_- u_h, y-\pi_W y\rangle_{\partial\Omega} + \langle u_h, \mu\nabla(y-\pi_W y)\cdot n\rangle_{\partial\Omega} \\ &= \mathrm{I}+\mathrm{II}+\mathrm{III}+\mathrm{IV}. \end{aligned}$$

Considering first the term I we get, with $\xi_h = \pi_V u - u_h$,

$$\begin{aligned} \mathrm{I} &= (\mathcal{L}(u-\pi_V u) + \mathcal{L}\xi_h, y-\pi_W y) \\ &\lesssim \|y-\pi_W y\|_*(\|u-\pi_V u\|_{\mathcal{L}} + |\xi_h|_{S_p} + h^2(|\beta|_{W^{2,\infty}(\Omega)} + |c|_{W^{1,\infty}(\Omega)})\|u-u_h\|), \end{aligned}$$

where we used once again the inequalities

$$\begin{aligned} &(\mu\Delta\xi_h + \beta_h\nabla\xi_h, y-\pi_W y) \le c_{os}|\xi_h|_{S_p}\|y-\pi_W y\|_*, \\ &((\beta-i_h\beta)\cdot\nabla\xi_h, y-\pi_W y) \lesssim h^2|\beta|_{W^{2,\infty}}(\|u-\pi_V u\| + \|u-u_h\|)\|y-\pi_W y\|_* \end{aligned}$$

and, by the discrete commutator property,

$$\begin{aligned} &((\nabla\cdot\beta + c)\xi_h, y-\pi_W y) \\ &\quad = ((\nabla\cdot\beta + c)\xi_h - i_h((\nabla\cdot\beta + c)\xi_h), y-\pi_W y) \\ &\quad \lesssim h^2(|\beta|_{W^{2,\infty}} + |c|_{W^{1,\infty}})(\|u-\pi_V u\| + \|u-u_h\|)\|y-\pi_W y\|_*. \end{aligned}$$

For the second, third, and fourth terms we have using the Cauchy–Schwarz inequality, adding and subtracting $\pi_V u$ and recalling the form of the boundary penalty term,

$$\begin{aligned} \mathrm{II}+\mathrm{III}+\mathrm{IV} &\lesssim \big(\|\mu^{\frac12}h^{\frac12}[\![\nabla u_h\cdot n_F]\!]\|_{\mathcal{F}_{int}} \\ &\quad + \|\big(\mu^{\frac12}h^{-\frac12} + |(\beta\cdot n)_-|^{\frac12}\big)u_h\|_{\partial\Omega}\big)\|y-\pi_W y\|_* \\ &\le (\|u-\pi_V u\|_{\mathcal{L}} + |\pi_V u - u_h|_{S_p})\|y-\pi_W y\|_*. \end{aligned}$$

We conclude that the claim holds with $\epsilon(h) \sim h^2(|\beta|_{W^{2,\infty}} + |c|_{W^{1,\infty}})$. ∎

*Remark* 4.3. Note that if the physical parameters are constant, then the analysis holds without restrictions on the mesh size in contrast to the standard Galerkin analysis of [19]. In this case, for $k=1$ the estimate takes the simple form

$$\|u-u_h\| \lesssim c_{a,\Omega}\zeta_{Pe}^2 h^4 |u|_{H^2(\Omega)}.$$

Assuming that $\beta_\infty \sim O(1)$, $c_\infty = 0$ we get

$$\|u-u_h\| \lesssim c_{a,\Omega}\left(\frac{1}{h} + \frac{\mu}{h^2}\right) h^4 |u|_{H^2(\Omega)}.$$

Here the constant $c_{a,\Omega}$ typically is proportional to some negative power of $\mu$, making the estimate valid only for moderate Péclet numbers. If we assume that $c_{a,\Omega} = O(\mu^{-1})$ we see that the quasi-optimal convergence of order $h^{\frac32}$ is obtained when $h^{\frac32} < \mu$. A more precise estimate for the hyperbolic regime, showing that the estimate cannot degenerate further even for vanishing $\mu$, is the subject of the second part of this work [5].

**4.1.2. Transport in compressible flow: Pure Neumann conditions.** We will now consider the convection-diffusion equation with homogeneous Neumann conditions. The main difficulty in this problem compared to the previous one is that due to the homogeneous Neumann condition, the primal and dual problems have different boundary conditions. The nonsolenoidal $\beta$ imposes special compatibility conditions on $g$ leading to complications in the finite element analysis and additional stability issues for the discrete solution. For this example we will assume that all physical parameters are order one. After presenting the problem and the method we propose, we first show that the discrete problem is well-posed for all mesh sizes when piecewise affine approximation is used. Then we prove that the assumptions of Lemma 2.2 and Theorem 2.3 are satisfied. Optimal error estimates for the problem similar to that above are obtained after accounting for some minor modifications needed to accomodate the compatibility conditions particular to this problem. The problem reads

$$\begin{aligned} -\Delta u + \nabla \cdot (\beta u) &= f \quad \text{in } \Omega, \\ -\nabla u \cdot n + \beta \cdot n\, u &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{4.7}$$

The dual adjoint problem is formally written

$$\begin{aligned} -\Delta z - \beta \cdot \nabla z &= g \quad \text{in } \Omega, \\ -\nabla z \cdot n &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{4.8}$$

We assume that the following compatibility conditions hold:

$$\int_\Omega f \, \mathrm{d}x = 0, \quad \int_\Omega gm \, \mathrm{d}x = 0, \tag{4.9}$$

where $m \in H^2(\Omega)$, $m > 0$, is the unique solution to the homogeneous form of the primal problem

$$\begin{aligned} -\Delta m + \nabla \cdot (\beta m) &= 0 \quad \text{in } \Omega, \\ -\nabla m \cdot n + \beta \cdot n\, m &= 0 \quad \text{on } \partial\Omega \end{aligned} \tag{4.10}$$

under the additional constraint

$$|\Omega|^{-1} \int_\Omega m \, \mathrm{d}x = 1.$$

Then the problems (4.7) and (4.8) are both well-posed by the Fredholm alternative. Since we assume that the regularity estimate (2.3) holds, $m \in C^0(\bar{\Omega})$ and $\sup_{x\in\Omega} m =: M \in \mathbb{R}^+$ and since $m > 0$ we may introduce $m_{min} := inf_\Omega\, m > 0$ (see [8]).

The problem is cast in the form (2.1) by setting $V := H^1(\Omega) \cap L^2_0(\Omega)$, where $L^2_0(\Omega)$ denotes the set of functions with global average zero, and by taking

$$a(u, v) := (\nabla u, \nabla v) - (u, \beta \cdot \nabla v).$$

The finite element method (2.5) is obtained by setting $V_h = W_h := X^k_h \cap L^2_0(\Omega)$,

$$a_h(u_h, v_h) := a(u_h, v_h), \tag{4.11}$$

and the stabilization operators

$$s_x(\cdot,\cdot) := s_{cip}(\cdot,\cdot) + s_{bc,x}(\cdot,\cdot) \text{ with } x = a, p. \tag{4.12}$$

$s_{cip}(\cdot,\cdot)$ is given by (3.7) with $\gamma_{i,F} := 1$, $i = 1, 2$. The boundary operators finally are defined by

$$s_{bc,p}(u_h, v_h) := \int_\Omega h(\nabla u_h \cdot n - \beta \cdot n u_h)(\nabla v_h \cdot n - \beta \cdot n v_h)\ \mathrm{d}s$$

for $k \geq 2$ and

$$s_{bc,p}(u_h, v_h) := \int_\Omega h(\nabla u_h \cdot n - (i_h\beta) \cdot n u_h)(\nabla v_h \cdot n - (i_h\beta) \cdot n v_h)\ \mathrm{d}s$$

for $k = 1$, and $s_{bc,a}(\cdot,\cdot)$ finally is given by (3.15), with $\gamma_N \sim 1$. The boundary stabilization operator $s_{bc,p}(\cdot,\cdot)$ for $k = 1$ is only weakly consistent. It is straightforward to show that the inconsistency introduced by replacing $\beta$ by $i_h\beta$ is compatible with (2.13). We omit the details here, but similar arguments are used below to prove the continuity (2.12).

PROPOSITION 4.4 (existence of discrete solution). *Assume $k = 1$ in the definition of $V_h$. Then there exists a unique solution $u_h$ to the discrete problem* (2.5).

*Proof.* As before we assume $f = 0$ and observe that

$$s_p(u_h, u_h) + s_a(z_h, z_h) = 0.$$

This implies $z_h, u_h \in \mathbb{P}_1(\Omega)$. Since $\|\nabla z_h \cdot n\|_{\partial\Omega} = 0$ and $z_h$ has zero average, we conclude that $z_h = 0$. For $u_h$ there holds

$$\|\nabla u_h \cdot n + (i_h\beta \cdot n)u_h\|_{\partial\Omega} = 0.$$

Since $\nabla u_h \cdot n$ is constant on every polyhedral side $\Gamma$ of $\Omega$, so is $(i_h\beta \cdot n)u_h$. But since $(i_h\beta \cdot n)u_h|_\Gamma \in \mathbb{P}_2(\Gamma)$ we conclude that both $i_h\beta$ and $u_h$ must be constant. Since this is true for all faces $\Gamma$ of $\Omega$, $u_h$ is a constant globally. We conclude by recalling that zero average was imposed on the approximation space. ∎

In case $k \geq 2$, we let the norms $\|\cdot\|_+$, $\|\cdot\|_*$ be defined by (3.10) and $\|\cdot\|_{\mathcal{L}}$ by

$$\|v\|_{\mathcal{L}} := \|\mathcal{L}v\|_h + \|h^{\frac{1}{2}}[\![\nabla v \cdot n_F]\!]\|_{\mathcal{F}_{int}} + \|h^{\frac{1}{2}}\nabla v \cdot n\|_{\partial\Omega} + \beta_\infty\|h^{\frac{1}{2}}v\|_{\partial\Omega}.$$

When $k = 1$ we let the norm $\|\cdot\|_+$ be defined by (3.10) but define

$$\|v\|_* := \|h^{-1}v\| + \|h^{-\frac{1}{2}}v\|_{\mathcal{F}} + \epsilon(h)\|v\|_{\partial\Omega}$$

and, assuming $h < 1$,

$$\|v\|_{\mathcal{L}} := \|\mathcal{L}v\|_h + \|h^{\frac{1}{2}}[\![\nabla v \cdot n_F]\!]\|_{\mathcal{F}_{int}} + \|h^{\frac{1}{2}}\nabla v \cdot n\|_{\partial\Omega} + (1+\beta_\infty)\|v\|_{\partial\Omega} + \epsilon(h)\|v\|.$$

For the projection operators $\pi_V$ and $\pi_W$ we once again choose the $L^2$-projection.

PROPOSITION 4.5. *The bilinear form* (4.11) *satisfies the continuities* (2.11) *and* (2.12) *with*

$$\epsilon(h) \sim h^2|\beta|_{W^{2,\infty}(\Omega)}.$$

*Proof.* As before we integrate by parts in $a_h(\cdot,\cdot)$ to obtain

$$a_h(u - \pi_V u, x_h) = \sum_{F \in \mathcal{F}_{int}} \langle u - \pi_V u, [\![\nabla x_h \cdot n_F]\!]\rangle_F + (u - \pi_V u, \mathcal{L}^* x_h)_h$$
$$+ \langle u - \pi_V u, \nabla x_h \cdot n\rangle_{\partial\Omega} = \mathrm{I} + \mathrm{II} + \mathrm{III}.$$

The treatment of terms I and II is identical to the Dirichlet case. Term III is bounded using the Cauchy–Schwarz inequality, recalling that the Neumann condition is penalized in $s_a(\cdot,\cdot)$,

$$\mathrm{III} \leq \|u - \pi_V u\|_+ |x_h|_{S_a}.$$

The second continuity follows in a similar fashion. We write

$$\begin{aligned} a_h(u - u_h, y - \pi_W y) &= \sum_{K \in \mathcal{T}_h} (\mathcal{L}(u - u_h), y - \pi_W y)_K \\ &\quad + \sum_{F \in \mathcal{F}_{int}} \langle [\![\nabla u_h \cdot n_F]\!], y - \pi_W y \rangle_F \\ &\quad - \langle \nabla u_h \cdot n - (\beta \cdot n) u_h, y - \pi_W y \rangle_{\partial\Omega} \\ &= \mathrm{I} + \mathrm{II} + \mathrm{III} \end{aligned}$$

and observe that the treatment of terms I and II is analogous with the Dirichlet case. For term III, when $k \geq 2$ we recall that $\nabla u_h \cdot n - (\beta \cdot n) u_h$ is penalized in $s_p(\cdot,\cdot)$ and we may conclude as before using a Cauchy–Schwarz inequality

$$\mathrm{III} \lesssim \|u - \pi_V u\|_* (|u_h - \pi_V u|_{S_p} + \|u - \pi_V u\|_{\mathcal{L}}).$$

For the case $k = 1$ we must take care to handle the lack of consistency. Therefore we add and subtract $i_h \beta$ and use the boundary condition on $u$ to get

$$\begin{aligned} \mathrm{III} = &\langle \nabla(u_h - u) \cdot n - (i_h \beta \cdot n)(u_h - u), y - \pi_W y \rangle_{\partial\Omega} \\ &+ \langle (i_h \beta - \beta) \cdot n (u_h - u), y - \pi_W y \rangle_{\partial\Omega}. \end{aligned}$$

First we add and subtract $\pi_V u$ so that $u - u_h = u - \pi_V u + \xi_h$, $\xi_h := \pi_V u - u_h$ and split the scalar products with Cauchy–Schwarz inequality. For the first term we immediately have

$$\begin{aligned} &\langle \nabla(u_h - u) \cdot n - (i_h \beta \cdot n)(u_h - u), y - \pi_W y \rangle_{\partial\Omega} \\ &\quad \leq \|y - \pi_W y\|_* (\|u - \pi_V u\|_{\mathcal{L}} + |\xi_h|_{S_p}). \end{aligned}$$

For the $\pi_V u - u$ part of the second term we observe that

$$\langle (i_h \beta - \beta) \cdot n (\pi_V u - u), y - \pi_W y \rangle_{\partial\Omega} \lesssim \|y - \pi_W y\|_{\partial\Omega} h^2 |\beta|_{W^{2,\infty}(\Omega)} \|\pi_V u - u\|_{\partial\Omega}.$$

Applying an elementwise trace inequality in the $\xi_h$ part of the second term, we have

$$\langle (i_h \beta - \beta) \cdot n \, \xi_h, y - \pi_W y \rangle_{\partial\Omega} \lesssim \|y - \pi_W y\|_{\partial\Omega} \|i_h \beta - \beta\|_{L^\infty(\partial\Omega)} h^{-\frac{1}{2}} \|\xi_h\|.$$

Then we use the definition of the norms, in particular that $\|h^{-\frac{1}{2}}(y - \pi_W y)\|_{\partial\Omega} \leq \|y - \pi_W y\|_*$, to obtain

$$\mathrm{III} \lesssim \|y - \pi_W y\|_* (\|u - \pi_V u\|_{\mathcal{L}} + |\xi_h|_{S_p} + h^2 |\beta|_{W^{2,\infty}} \|\xi_h\|).$$

Using the triangular inequality $\|\xi_h + u - u\| \leq \|\pi_V u - u\| + \|u - u_h\|$, we obtain

$$\mathrm{III} \lesssim \|y - \pi_W y\|_* (\|u - \pi_V u\|_{\mathcal{L}} + |\xi_h|_{S_p} + \epsilon(h) \|u - u_h\|)$$

with $\epsilon \sim h^2 |\beta|_{W^{2,\infty}(\Omega)}$. The proof is complete. ∎

*Remark* 4.6. It is straightforward to verify that Lemma 2.2 holds. The assumptions of Theorem 2.3, however, still are not satisfied since we want to use the solutions of the problems $\mathcal{L}^*\varphi = u - u_h$ and $\mathcal{L}\psi = z_h$, but the solution $\varphi$ will in general not exist since $u-u_h$ does not satisfy the second compatibility condition of (4.9). Instead we will use $m$, the solution of (4.10) as weight, as suggested in [8], and solve the well-posed problem

$$\mathcal{L}^*\varphi = (u - u_h)/m.$$

We may then write

$$\|(u-u_h)m^{-\frac{1}{2}}\|^2 + \|z_h\|^2 = (u-u_h, \mathcal{L}^*\varphi) + (\mathcal{L}\psi, z_h)$$

and proceed as in Theorem 2.3, now using the stability estimate

$$|\varphi|_{H^2(\Omega)} \le c_{a,\Omega}\|(u-u_h)m^{-1}\| \le c_{a,\Omega}/m_{min}^{\frac{1}{2}}\|(u-u_h)m^{-1/2}\|$$

to obtain

$$\|(u-u_h)m^{-\frac{1}{2}}\| + \|z_h\| \lesssim h^{k+1}|u|_{H^{k+1}(\Omega)}.$$

For an estimate in the unweighted $L^2$-norm we observe that

$$M^{-\frac{1}{2}}\|u-u_h\| \le \|(u-u_h)m^{-\frac{1}{2}}\|.$$

Convergence follows by Lemma 2.2 and the modified Theorem 2.3. Observe that the constants in $\epsilon(h)$ now depend on the (unknown) minimum value of $m$.

*Remark* 4.7. In practice the zero average condition can be imposed using Lagrange multipliers. The above analysis holds for that case after minor modifications.

**4.2. The Cauchy problem.** We consider the case of a Helmholtz-type problem where both the solution itself and its normal gradient are specified on one portion of the domain and the other portion is free. We let $\Gamma_V$ and $\Gamma_W$ be connected subsets of $\partial\Omega$ such that $\partial\Omega := \bar{\Gamma}_V \cup \bar{\Gamma}_W$ and $\Gamma_V \cap \Gamma_W = \emptyset$. We will consider the problem, $\kappa \in \mathbb{R}$,

$$\begin{aligned} -\Delta u + \kappa u &= f \text{ in } \Omega, \\ \nabla u \cdot n = u &= 0 \text{ on } \Gamma_V, \end{aligned} \tag{4.13}$$

with dual problem

$$\begin{aligned} -\Delta z + \kappa z &= g \text{ in } \Omega, \\ \nabla z \cdot n = z &= 0 \text{ on } \Gamma_W. \end{aligned} \tag{4.14}$$

The weak formulations (2.1) and (2.2) are obtained by setting

$$V := \{v \in H^1(\Omega) : v|_{\Gamma_V} = 0\}$$

and

$$W := \{v \in H^1(\Omega) : v|_{\Gamma_W} = 0\}$$

and defining

$$a(u,v) := (\nabla u, \nabla v) + \kappa(u,v) \quad \forall u \in V,\, v \in W.$$

Note that both symmetry and the Gårdings inequality fail in this case because the functions in the bilinear form have to be taken in different spaces and hence the choice $v = u$ is prohibited.

To design a suitable discrete formulation (2.5) for this problem we generalize the ideas of the Nitsche-type weak imposition of boundary condition. Observe that in this case boundary conditions imposed using a penalty in the standard fashion cannot be consistent for both the primal and the adjoint problem, since the primal and dual solutions are zero on different parts of the boundary. It is therefore important in this case that two stabilization operators are used, one for the primal and one for the adjoint. We define the approximation spaces $V_h = W_h := X_h^k$. We propose the bilinear form

$$a_h(u_h, v_h) := (\nabla u_h, \nabla v_h) + \kappa(u_h, v_h) - \langle \nabla v_h \cdot n, u_h \rangle_{\Gamma_V} - \langle \nabla u_h \cdot n, v_h \rangle_{\Gamma_W} \tag{4.15}$$

and for the stabilization we use

$$s_x(u_h, v_h) := s_{cip}(u_h, v_h) + s_{bc,x}(u_h, v_h), \quad x = a, p, \tag{4.16}$$

where $s_{cip}(\cdot,\cdot)$ is given by (3.7), with $\gamma_{F,i} \sim 1$, $i = 1, 2$, and

$$s_{bc,x}(u_h, v_h) := \int_X (h^{-1} u_h\, v_h + h \nabla u_h \cdot n \nabla v_h \cdot n)\, \mathrm{d}s,$$

where $X = \Gamma_V$ for $x = p$ and $X = \Gamma_W$ for $x = a$. If some part of the boundary is equipped with Dirichlet or Neumann boundary conditions this is imposed as described in section 3.3.

PROPOSITION 4.8 (existence of discrete solution for $k = 1$). *Define* (2.5) *by the bilinear forms* (4.15) *and* (4.16). *Let* $k = 1$ *in* $V_h$. *Then there exists a unique solution* $(u_h, z_h) \in [V_h]^2$ *to* (2.5).

*Proof.* Let $f = 0$. By (2.10) there holds, $u_h, z_h \in \mathbb{P}_1(\Omega)$ and $u_h|_{\Gamma_V} = \nabla u_h \cdot n|_{\Gamma_V} = 0$ as well as $z_h|_{\Gamma_W} = \nabla z_h \cdot n|_{\Gamma_W} = 0$, by which we conclude that the matrix is invertible using case 2 of Proposition 3.3. ∎

For the error analysis we once again choose the interpolants $\pi_V$ and $\pi_W$ to be the standard $L^2$-projection $\pi_L$. We will now prove that the assumptions (2.7)–(2.8) and (2.11)–(2.12) are satisfied.

LEMMA 4.9 (consistency of bilinear form). *The bilinear form* (4.15) *satisfies* (2.7) *and* (2.8).

*Proof.* By an integration by parts we see that for $u$ solution of (4.13)

$$\begin{aligned}(-\Delta u + \kappa u, v + v_h) &= (\nabla u, \nabla(v + v_h)) + (\kappa u, v + v_h) \\ &\quad - \underbrace{\langle \nabla u \cdot n, v + v_h \rangle_{\Gamma_W}}_{\text{since } \nabla u \cdot n = 0 \text{ on } \Gamma_V} - \underbrace{\langle \nabla(v + v_h) \cdot n, u \rangle_{\Gamma_V}}_{\text{since } u = 0 \text{ on } \Gamma_V} = a_h(u, v + v_h).\end{aligned}$$

Similarly for $z$ solution of (4.14) consistency follows by observing that

$$(v + v_h, -\Delta z) = (\nabla(v + v_h), \nabla z) - \underbrace{\langle \nabla z \cdot n, v + v_h \rangle_{\Gamma_V}}_{\text{since } \nabla z \cdot n = 0 \text{ on } \Gamma_W} - \underbrace{\langle \nabla(v + v_h) \cdot n, z \rangle_{\Gamma_W}}_{\text{since } z = 0 \text{ on } \Gamma_W}. \qquad \square$$

We define the norms $\|\cdot\|_+$ and $\|\cdot\|_*$ by

$$\|v\|_+ := \|h^{-\frac{1}{2}}v\|_{\mathcal{F}_{int}} + \|h^{-1}v\| + \|h^{-\frac{1}{2}}v\|_{\Gamma_W} + \|h^{\frac{1}{2}}\nabla v\cdot n\|_{\Gamma_W},$$
$$\|v\|_* := \|h^{-\frac{1}{2}}v\|_{\mathcal{F}_{int}} + \|h^{-1}v\| + \|h^{-\frac{1}{2}}v\|_{\Gamma_V} + \|h^{\frac{1}{2}}\nabla v\cdot n\|_{\Gamma_V},$$

and

$$\|v\|_{\mathcal{L}} := \|h\mathcal{L}v\|_h + \|h^{\frac{1}{2}}[\![\nabla v\cdot n_F]\!]\|_{\mathcal{F}_{int}} + \|h^{-\frac{1}{2}}v\|_{\Gamma_V} + \|h^{\frac{1}{2}}\nabla v\cdot n\|_{\Gamma_V}.$$

It is straightforward to show (2.13) and (2.14).

Proposition 4.10. *For $a_h(\cdot,\cdot)$ defined by* (4.15), *the continuities* (2.11) *and* (2.12) *hold with $\epsilon(h) = 0$.*

*Proof.* We proceed as before using an integration by parts in (4.15) to obtain

$$\begin{aligned} a_h(v-\pi_V v, x_h) &= \sum_{F\in\mathcal{F}_{int}} \langle v-\pi_V v, [\![\nabla x_h\cdot n_F]\!]\rangle_F + (v-\pi_V v, -\Delta x_h + \kappa x_h)_h \\ &\quad + \langle v-\pi_V v, \nabla x_h\cdot n\rangle_{\Gamma_W} - \langle \nabla(v-\pi_V v)\cdot n, x_h\rangle_{\Gamma_W} \\ &= \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}. \end{aligned}$$

The first sum I is upper bounded as before using the Cauchy–Schwarz inequality, and for the second sum, we use the orthogonality of the $L^2$-projection, $(v-\pi_V v, \kappa x_h) = 0$, and the discrete interpolation inequality (3.11) leading to

$$\mathrm{I} + \mathrm{II} \lesssim \|u-\pi_V u\|_+ |x_h|_{S_a}.$$

For the terms III and IV we note that by the definition of $\|\cdot\|_+$ and $|\cdot|_{S_a}$ there also holds

$$\mathrm{III} + \mathrm{IV} \le \|u-\pi_V u\|_+ |x_h|_{S_a}.$$

This ends the proof of (2.11). The proof of (2.12) is similar. Using integration by parts in the other direction we have

$$\begin{aligned} a_h(u-u_h, y-\pi_W y) &= (-\Delta(u-u_h) + \kappa(u-u_h), y-\pi_W y)_h \\ &\quad + \sum_{F\in\mathcal{F}_{int}} \langle [\![\nabla u_h\cdot n_F]\!], y-\pi_W y\rangle_F + \langle u-u_h, \nabla(y-\pi_W y)\cdot n\rangle_{\Gamma_V} \\ &\quad + \langle \nabla(u-u_h)\cdot n, y-\pi_W y\rangle_{\Gamma_V} = \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}. \end{aligned}$$

Using the same arguments as before, adding and subtracting $\pi_V u$ in all the terms in the left slot we have for the term I, using $\xi_h = \pi_V u - u_h$,

$$\begin{aligned} \mathrm{I} &= (-\Delta(u-\pi_V u) + \kappa(u-\pi_V u), y-\pi_W y)_h - (\Delta\xi_h - \mathrm{I}_{os}\Delta\xi_h, y-\pi_W y)_h \\ &\lesssim (\|u-\pi_V u\|_{\mathcal{L}} + |\xi_h|_{S_p})\|y-\pi_W y\|_*. \end{aligned}$$

By the definition of the stabilization operator and the fact that $u = \nabla u\cdot n = 0$ on $\Gamma_V$ we may once again add and subtract $\pi_V u$ in the terms II and III to obtain

$$\begin{aligned} \mathrm{II} + \mathrm{III} + \mathrm{IV} &= \langle [\![\nabla u_h\cdot n_F]\!], y-\pi_W y\rangle_{\mathcal{F}_{int}} \\ &\quad - \langle u_h, \nabla(y-\pi_W y)\cdot n\rangle_{\Gamma_V} - \langle \nabla u_h\cdot n, y-\pi_W y\rangle_{\Gamma_V} \\ &\le \|y-\pi_W y\|_*(\|u-\pi_V u\|_{\mathcal{L}} + |\xi_h|_{S_p}), \end{aligned}$$

by which we conclude. $\square$

COROLLARY 4.11. *Assume that the problems* (4.13) *and* (4.14) *admit unique solutions for which* (2.3) *holds. Then the conclusions of Theorem* 2.3 *hold for* (2.5) *defined by* $V_h$*,* (4.15)*, and* (4.16)*.*

*Proof.* In Lemma 4.9 we verified the consistencies (2.7) and (2.8). In Proposition 4.10 we verified the continuities (2.11) and (2.12). It is straightforward to verify that (2.13)–(2.14) hold for $\pi_V = \pi_W = \pi_L$ and $s_p(\cdot,\cdot)$, $s_a(\cdot,\cdot)$ defined by (4.16) under the assumptions on the mesh and the regularity assumptions on the solution. ∎

*Remark* 4.12. Admittedly Corollary 4.11 is of purely academic interest since the Cauchy problem under consideration in general is ill-posed, with very weak stability properties. As we shall see in the numerical section the method nevertheless returns useful approximations. An example of a sufficient condition for Theorem 2.3 to result in a convergence estimate, if $u$ is smooth, is that there exists $M \in \mathbb{R}^+$ and $s \in \mathbb{R}$, with $s > -k$ such that $\|\varphi - \pi_V \varphi\|_* \le M h^s$ for all $u - u_h$, with $\varphi$ the solution of (2.2), with $g = u - u_h$. The expected convergence rate in that case would be

$$\|u - u_h\| \lesssim M^{\frac{1}{2}} h^{(k+s)/2} |u|^{\frac{1}{2}}_{H^{k+1}(\Omega)}. \tag{4.17}$$

Unfortunately, no such stability estimates are known for the Cauchy problem and a regularized adjoint would have to be considered. We refer to [3] for conditional stability estimates for the problem (4.13) in general Lipschitz domains, leading to logarithmic estimates and to [13, 18] and [17] for other work on finite element methods on the Cauchy problem and some stability results under special geometrical assumptions.

## 5. Numerical investigations.

We will present numerical examples of convergence for a smooth exact solution of the applications given above. For the computations we have used FreeFEM++ [14]. All problems will be set in $\Omega := (0,1) \times (0,1)$. We use unstructured meshes with $2^N$ elements on each side, $N = 3, \dots, 8$, and drawing on our previous experience of the CIP method we fix the stabilization parameters to be $\gamma_{1,F} = 0.01$ for piecewise affine approximation and $\gamma_{i,F} = 0.001$, $i = 1, 2$, for piecewise quadratic approximation. The boundary penalty parameter is chosen to be $\gamma_{bc} = 10$ for both cases and for both Dirichlet and Neumann penalty terms. Let us remark that in particular for the ill-posed Cauchy problem, an optimal choice of the stabilization parameter can have a big impact on the error on a fixed mesh but does not appear to influence the convergence behavior. For each example we plot the error quantities estimated in Lemma 2.2 and Theorem 2.3. When appropriate we indicate the experimental convergence order in parentheses. We report the computational mesh for $N = 5$ in the left plot of Figure 5.1.

### 5.1. Convection-diffusion problems.

We consider an example given in [9], where, in (4.1), the physical parameters are chosen as $\mu = 1$, $c = 0$,

$$\beta := -100 \begin{pmatrix} x + y \\ y - x \end{pmatrix}$$

(see the right plot of Figure 5.1) and the exact solution is given by

$$u(x, y) = 30x(1 - x)y(1 - y). \tag{5.1}$$

This function satisfies homogeneous Dirichlet boundary conditions and has $\|u\| = 1$. Note that $\|\beta\|_{L^\infty} = 200$ and $\nabla \cdot \beta = -200$, making the problem strongly noncoercive with a medium high Péclet number. The right-hand side is then chosen as $\mathcal{L}u$ and in the case of (nonhomogeneous) Neumann conditions, a suitable right-hand side is introduced to make the boundary penalty term consistent. The optimal convergence rate for the stabilizing terms given in Lemma 2.2 is verified in all the numerical examples.

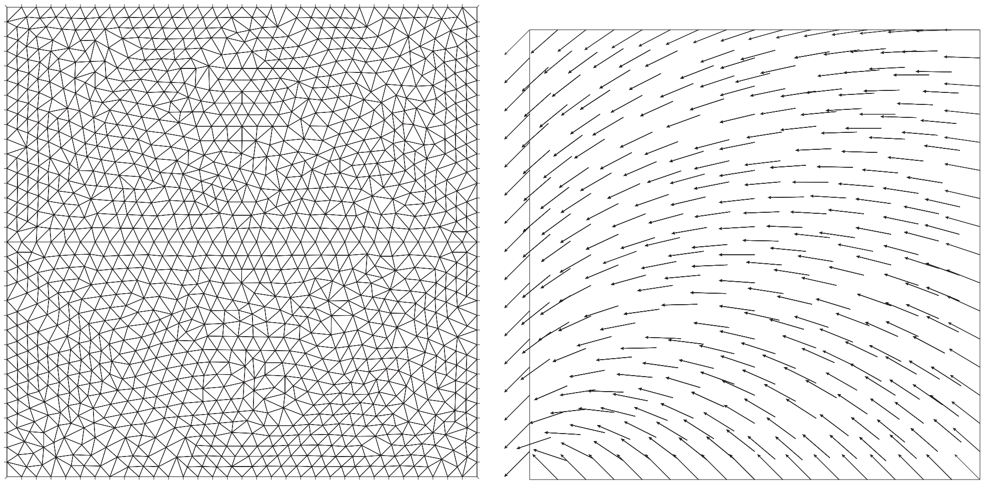

FIG. 5.1. *Left: example of unstructured mesh, $N = 5$. Right: plot of the velocity vector field.*

TABLE 5.1
*Convergence orders of estimated quantities for the Dirichlet problem approximated using piecewise affine elements.*

| $N$ | $\|u - u_h\|$ | $\|z_h\|$ | $\|u_h\|_{S_p} + \|z_h\|_{S_a}$ |
|---|---|---|---|
| 3 | 0.038 (–) | 0.024 | 0.57 |
| 4 | 0.012 (1.7) | 0.0017 | 0.24 |
| 5 | 0.0024 (2.3) | 0.00043 | 0.11 |
| 6 | 0.00043 (2.5) | 0.00012 | 0.052 |
| 7 | 0.00010 (2.1) | 2.5E-05 | 0.025 |
| 8 | 2.3E-05 (2.1) | 5.3E-06 | 0.012 |

TABLE 5.2
*Convergence orders of estimated quantities for the Dirichlet problem approximated using piecewise quadratic elements.*

| $N$ | $\|u - u_h\|$ | $\|z_h\|$ | $\|u_h\|_{S_p} + \|z_h\|_{S_a}$ |
|---|---|---|---|
| 3 | 0.0014 (–) | 0.00041 | 0.024 |
| 4 | 0.00012 (3.5) | 4.6E-05 | 0.0044 |
| 5 | 8.8E-06 (3.8) | 4.6Ee-06 | 0.00081 |
| 6 | 8.0E-07 (3.5) | 6.6E-07 | 0.00017 |
| 7 | 8.3E-08 (3.3) | 8.2E-08 | 3.7E-05 |

**5.1.1. Dirichlet boundary conditions.** In Table 5.1 we show the result of the computation when Dirichlet boundary conditions are applied and piecewise affine approximation is used on a sequence of unstructured meshes. We observe that the solution exhibits the preasymptotic convergence rate $h^{\frac{3}{2}}$ under one refinement before achieving the full second order convergence rate in $L^2$.

In Table 5.2 similar data for second order polynomials are presented. Here the asymptotic regime with full convergence is obtained from the first refinement.

**5.1.2. Neumann boundary conditions.** We consider the same differential operator but with (nonhomogeneous) Neumann boundary conditions. This is exactly the problem considered in [9]. The average values of the approximate solutions have

Table 5.3

*Convergence orders of estimated quantities for the Neumann problem approximated using piecewise affine elements.*

| $N$ | $\|u-u_h\|$ | $\|z_h\|$ | $\|u_h\|_{S_p}+\|z_h\|_{S_a}$ |
|---|---|---|---|
| 3 | 0.028 (–) | 0.028 (–) | 0.82 |
| 4 | 0.0066 (2.1) | 0.016 (0.8) | 0.32 |
| 5 | 0.0016 (2.0) | 0.0058 (1.5) | 0.13 |
| 6 | 0.00039 (2.0) | 0.0015 (2.0) | 0.060 |
| 7 | 9.7E-05 (2.0) | 0.00031 (2.3) | 0.028 |
| 8 | 2.3E-05 (2.1) | 6.5E-05 (2.3) | 0.013 |

Table 5.4

*Convergence orders of estimated quantities for the Neumann problem approximated using piecewise quadratic elements.*

| $N$ | $\|u-u_h\|$ | $\|z_h\|$ | $\|u_h\|_{S_p}+\|z_h\|_{S_a}$ |
|---|---|---|---|
| 3 | 0.00061 (–) | 0.0020 (–) | 0.030 |
| 4 | 6.6E-05 (3.2) | 0.00040 (2.3) | 0.0054 |
| 5 | 6.5E-06 (3.3) | 2.5E-05 (4.0) | 0.00099 |
| 6 | 7.1E-07 (3.2) | 1.7E-06 (3.9) | 0.00020 |
| 7 | 7.9E-08 (3.2) | 1.4E-07 (3.6) | 4.2E-05 |

been imposed using Lagrange multipliers. In Tables 5.3 and 5.4 we observe optimal convergence rates once again as predicted by theory. Observe that in the case of piecewise affine approximation the dual solution $z_h$ comes into the asymptotic regime only on the finer meshes.

**5.2. A Cauchy problem.** Since we have no complete theory for the ill-posed Cauchy problem we will proceed with a more thorough numerical investigation. First we consider the Cauchy problem obtained by taking $\kappa = 0$ in (4.13). Then we consider a Cauchy problem using the convection-diffusion operator of (4.7) in two different boundary configurations. For all test cases we use the exact solution (5.1) and the stabilization parameters given above. We present the data for the quantities estimated in Lemma 2.2 and Theorem 2.3, but also the error in the total diffusive flux in the discrete $H^{-1/2}(\partial\Omega)$ norm on the boundary.

$$\|\nabla(u-u_h)\cdot n\|^2_{-\frac{1}{2},h,\partial\Omega} := \int_{\partial\Omega} h(\nabla(u-u_h)\cdot n)^2 \,\mathrm{d}s.$$

**5.2.1. Poisson's equation.** Here we consider the problem with $\kappa = 0$ in (4.13). We impose the Cauchy data, i.e., both Dirichlet and Neumann data, on boundaries $x = 0$, $0 < y < 1$, and $y = 1$, $0 < x < 1$. In Table 5.5 we show the obtained errors

Table 5.5

*Convergence orders of estimated quantities for the Poisson Cauchy problem approximated using piecewise affine elements.*

| $N$ | $\|u-u_h\|$ | $\|z_h\|$ | $\|u_h\|_{S_p}+\|z_h\|_{S_a}$ | $\|\nabla(u-u_h)\cdot n\|_{-\frac{1}{2},h,\partial\Omega}$ |
|---|---|---|---|---|
| 3 | 0.070 (–) | 0.59 (–) | 2.0 (–) | 2.7 (–) |
| 4 | 0.074 (–) | 0.42 (0.49) | 0.79 (1.3) | 1.3 (1.1) |
| 5 | 0.037 (1.0) | 0.30 (0.49) | 0.30 (1.4) | 0.75 (0.80) |
| 6 | 0.029 (0.35) | 0.26 (0.2) | 0.13 (1.2) | 0.51 (0.56) |
| 7 | 0.024 (0.27) | 0.20 (0.37) | 0.054 (1.3) | 0.33 (0.62) |
| 8 | 0.020 (0.26) | 0.16 (0.32) | 0.022 (1.3) | 0.21 (0.65) |

TABLE 5.6
*Convergence orders of estimated quantities for the Poisson Cauchy problem approximated using piecewise quadratic elements,* $\gamma = 0.001$, $\gamma_{bc} = 10$.

| $N$ | $\lVert u - u_h \rVert$ | $\lVert z_h \rVert$ | $\lvert u_h \rvert_{S_p} + \lvert z_h \rvert_{S_a}$ | $\lVert \nabla(u - u_h) \cdot n \rVert_{-\frac{1}{2},h,\partial\Omega}$ |
|---|---|---|---|---|
| 3 | 0.031 (–) | 0.062 (–) | 0.073 (–) | 0.92 (–) |
| 4 | 0.022 (0.49) | 0.025 (1.3) | 0.014 (2.4) | 0.48 (0.94) |
| 5 | 0.013 (0.76) | 0.014 (0.84) | 0.0025 (2.5) | 0.24 (1.0) |
| 6 | 0.0088 (0.56) | 0.011 (0.35) | 0.00047 (2.4) | 0.13 (0.88) |
| 7 | 0.0069 (0.35) | 0.0067 (0.72) | 8.8E-05 (2.8) | 0.080 (0.70) |

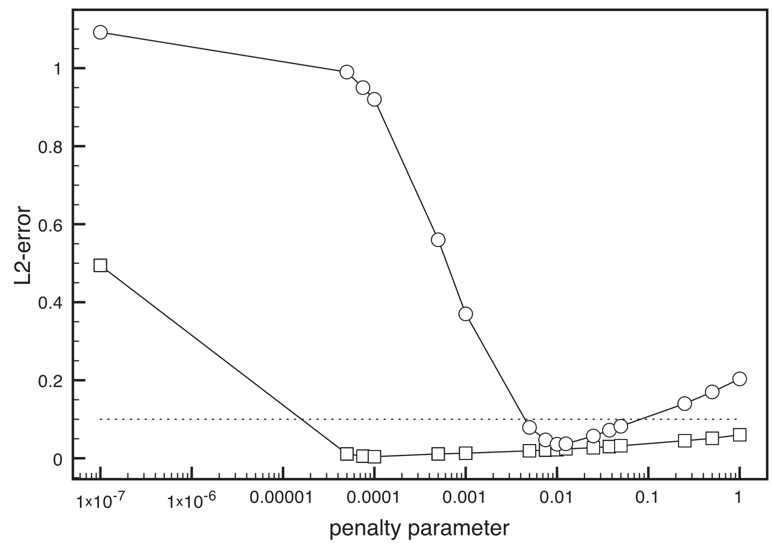


FIG. 5.2. *Study of the $L^2$-norm error under variation of the stabilization parameter. Circles: affine elements; squares: quadratic elements.*

when piecewise affine approximation is used and in Table 5.6 we show the results for piecewise quadratic approximation.

First note that in both cases one observes the optimal convergence of the stabilization terms predicted by Lemma 2.2. For the $L^2$-norm of the error we observe experimental convergence orders $h^\alpha$ with typically $\alpha \sim 0.25$ for piecewise affine approximation and $\alpha \sim 0.5$ for quadratic approximation. Higher convergence orders were obtained in both cases for the normal diffusive flux. In Figure 5.2, we present a study of the $L^2$-norm error under variation of the stabilization parameter. The computations are made on one mesh, with 32 elements per side, and the Cauchy problem is solved with $k = 1, 2$ and different values for $\gamma_{F,1} = \gamma_{F,2}$ with $\gamma_{bc} = 10$ fixed. The level of 10% relative error is indicated by the horizontal dotted line. Observe that the robustness with respect to stabilization parameters is much better for quadratic approximation. Indeed the 10% error level is met for all parameter values $\gamma_{i,F} \in [2.0E-5, 1]$, whereas in the case of piecewise affine approximation one has to take $\gamma_{1,F} \in [0.003, 0.05]$ approximately. Similar results for the boundary penalty parameter not reported here showed that the method was even more robust under perturbations of $\gamma_{bc}$. In the left plot of Figure 5.3 we present the contour plot of the error $u - u_h$ and in the right we show the contour plot of $z_h$. In both cases the error is concentrated on the boundary where no boundary conditions are imposed.

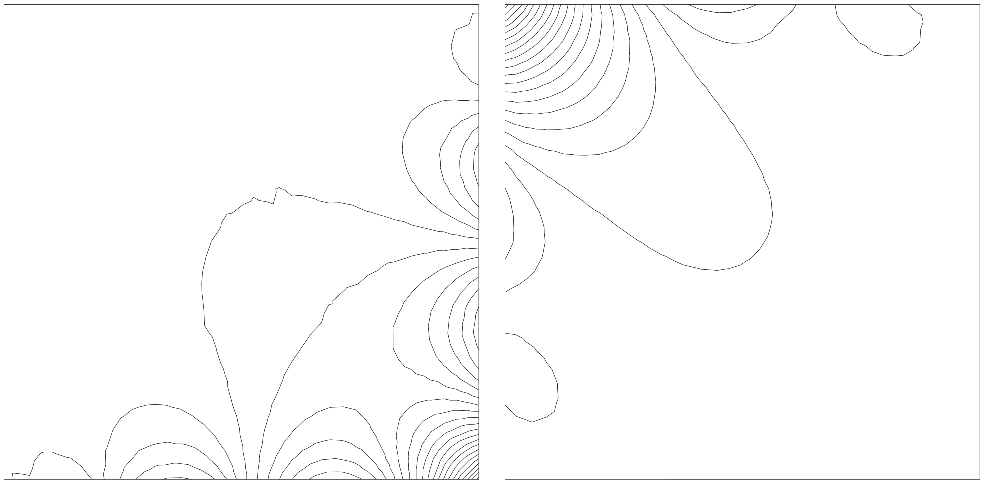

FIG. 5.3. *Contour plots of the error $u - u_h$ (left plot) and the error in the dual variable $z_h$ (right plot).*

**5.2.2. The noncoercive convection-diffusion equation.** As a last example we consider the Cauchy problem using the noncoercive convection-diffusion operator (3.1). The stability of the problem depends strongly on where the boundary conditions are imposed in relation to the inflow and outflow boundaries. To illustrate this we propose two configurations. Recalling the right plot of Figure 5.1 we observe that the flow enters along the boundaries $y = 0$, $y = 1$, and $x = 1$ and exits on the boundary $x = 0$. Note that the strongest inflow takes place on $y = 0$ and $x = 1$, the flow being close to parallel to the boundary in the right half of the segment $y = 1$. We propose the two different Cauchy problem configurations:

*Case* 1. We impose Dirichlet and Neumann data on the two mixed boundaries $x = 0$ and $y = 1$.

*Case* 2. We impose Dirichlet and Neumann data on the two inflow boundaries $y = 0$ and $x = 1$.

In the first case the outflow portion or the inflow portion of every streamline is included in the Cauchy boundary, whereas in the second case the main part of the inflow boundary is included. This highlights two different difficulties for Cauchy problems for the convection-diffusion operator. In Case 1 we must solve the problem backward along the characteristics, essentially solving a backward heat equation, whereas in Case 2 the crosswind diffusion must reconstruct missing boundary data.

In Tables 5.7–5.10, we report the results on the same sequence of unstructured meshes used in the previous examples for piecewise affine and piecewise quadratic approximations and the two problem configurations. First note that in all cases the result of Lemma 2.2 holds as expected. Otherwise the method behaves very differently in the two cases. For Case 1 we observe better convergence orders than in the case of the pure Poisson problem, typically $h^{\frac{1}{2}}$ for affine elements and $h$ for quadratic elements in the $L^2$-norm. Even higher orders are obtained for the global diffusive flux in the discrete $H^{-\frac{1}{2}}$ norm. The dual variable $z_h$ on the other hand has very poor convergence, although it is quite small on all meshes in the case of quadratic approximation. Case 2 (control on main part of the inflow) is clearly much more difficult. Convergence orders for both the affine case and the quadratic case are poor

TABLE 5.7
*Convergence orders of estimated quantities for the convection-diffusion Cauchy problem approximated using piecewise affine elements (Case 1).*

| $N$ | $\|u-u_h\|$ | $\|z_h\|$ | $\|u_h\|_{S_p}+\|z_h\|_{S_a}$ | $\|\nabla(u-u_h)\cdot n\|_{-\frac{1}{2},h,\partial\Omega}$ |
|---|---|---|---|---|
| 3 | 0.032 (–) | 0.044 (–) | 1.6 (–) | 0.35 (–) |
| 4 | 0.010 (1.7) | 0.020 (1.1) | 0.61 (1.4) | 0.13 (1.4) |
| 5 | 0.0045 (1.2) | 0.034 (–) | 0.24 (1.3) | 0.048 (1.4) |
| 6 | 0.0035 (0.36) | 0.052 (–) | 0.10 (1.3) | 0.018 (1.4) |
| 7 | 0.0039 (–) | 0.056 (–) | 0.045 (1.2) | 0.0074 (1.3) |
| 8 | 0.0026 (0.58) | 0.059 (–) | 0.020 (1.2) | 0.0031 (1.3) |

TABLE 5.8
*Convergence orders of estimated quantities for the convection-diffusion Cauchy problem approximated using piecewise affine elements (Case 2).*

| $N$ | $\|u-u_h\|$ | $\|z_h\|$ | $\|u_h\|_{S_p}+\|z_h\|_{S_a}$ | $\|\nabla(u-u_h)\cdot n\|_{-\frac{1}{2},h,\partial\Omega}$ |
|---|---|---|---|---|
| 3 | 0.13 (–) | 0.032 (–) | 1.74 (–) | 0.44 (–) |
| 4 | 0.097 (0.42) | 0.012 (1.4) | 0.63 (1.5) | 0.23 (0.94) |
| 5 | 0.075 (0.37) | 0.010 (0.26) | 0.24 (1.4) | 0.11 (1.1) |
| 6 | 0.067 (0.16) | 0.010 (–) | 0.10 (1.3) | 0.070 (0.65) |
| 7 | 0.063 (0.089) | 0.0097 (0.044) | 0.043 (1.2) | 0.047 (0.57) |
| 8 | 0.056 (0.17) | 0.0082 (0.24) | 0.018 (1.3) | 0.030 (0.65) |

TABLE 5.9
*Convergence orders of estimated quantities for the convection-diffusion Cauchy problem approximated using piecewise quadratic elements (Case 1).*

| $N$ | $\|u-u_h\|$ | $\|z_h\|$ | $\|u_h\|_{S_p}+\|z_h\|_{S_a}$ | $\|\nabla(u-u_h)\cdot n\|_{-\frac{1}{2},h,\partial\Omega}$ |
|---|---|---|---|---|
| 3 | 0.0022 (–) | 0.0037 (–) | 0.096 (–) | 0.033 (–) |
| 4 | 0.00054 (2.0) | 0.00089 (2.1) | 0.020 (2.3) | 0.0091 (1.9) |
| 5 | 0.00024 (1.2) | 0.0013 (–) | 0.0041 (2.3) | 0.0021 (2.1) |
| 6 | 0.00012 (1.0) | 0.00078 (0.74) | 0.00096 (2.1) | 0.00047 (2.2) |
| 7 | 5.6E-05 (1.1) | 0.00048 (0.70) | 0.00022 (2.1) | 0.00015 (1.6) |

TABLE 5.10
*Convergence orders of estimated quantities for the convection-diffusion Cauchy problem approximated using piecewise quadratic elements (Case 2).*

| $N$ | $\|u-u_h\|$ | $\|z_h\|$ | $\|u_h\|_{S_p}+\|z_h\|_{S_a}$ | $\|\nabla(u-u_h)\cdot n\|_{-\frac{1}{2},h,\partial\Omega}$ |
|---|---|---|---|---|
| 3 | 0.020 (–) | 0.0014 (–) | 0.074 (–) | 0.12 (–) |
| 4 | 0.034 (–) | 0.00028 (2.3) | 0.013 (2.5) | 0.11 (0.12) |
| 5 | 0.026 (0.39) | 0.00011 (1.4) | 0.0025 (2.4) | 0.065 (0.76) |
| 6 | 0.024 (0.12) | 8.3E-05 (0.4) | 0.00046 (2.4) | 0.043 (0.60) |
| 7 | 0.023 (0.06) | 3.6E-05 (1.2) | 8.7E-05 (2.4) | 0.029 (0.57) |

(around $\sim h^{\frac{1}{5}}$) and uneven. The diffusive fluxes on the boundary nevertheless still converge approximately as $h^{\frac{1}{2}}$ in both cases. We conclude that the Cauchy convection-diffusion problem is much less ill-posed if for each streamline *either the inflow part or the outflow part lies in the controlled zone*. The fact that we in Case 2 control more of the inflow boundary is unimportant compared to the fact that both the inflow and the outflow are unknown in the boundary portion around the corner $(0, 1)$.

**6. Concluding remarks.** We have proposed a framework for the design of stabilized finite element methods for noncoercive and nonsymmetric problems. The fundamental idea is to use an optimization framework to select the discrete solution on each mesh. This also opens new venues for inverse problems or boundary control problems, where Tichonov regularization can be introduced in the form of a stabilization operator with optimal weak consistency properties, eliminating the need to match a penalty parameter and the mesh size to obtain optimal performance. The method has some other interesting features. In particular for piecewise affine approximation spaces the discrete solution can be shown to exist under very mild assumptions. Both symmetric stabilization methods and the GLS methods are considered in the analysis. Convergence of the method is obtained formally under abstract assumptions on the bilinear form that are shown to hold for three nontrivial examples. The actual performance of the method in practice depends crucially on the stability properties of the underlying PDE and when these are unknown must be investigated numerically. Sometimes observed convergence orders are unlikely to match those predicted in Theorem 2.3 (except possibly for very small $h$), due to huge stability constants in the bound (2.3) (cf. the Helmholtz equation for large wave numbers), or more generally ill-posedness of the dual problem (cf. the Cauchy problem for Poisson's equation). Another problem that may arise when ill-conditioned problems are considered is poor conditioning of the system matrix. Even in the case of piecewise affine approximation the stabilization corresponds to a very weak norm, and in case the underlying problem is ill-posed this must be expected to show in the condition number. Clearly preconditioners for the linear systems arising is an important open problem. Other subjects for future work concern the inclusion of hyperbolic problems in the framework (see [5]) and the application of the method to data assimilation and boundary control.

## REFERENCES


[1] F. B. Belgacem, *Why is the Cauchy problem severely ill-posed?* Inverse Problems., 23 (2007), pp. 823–836.

[2] S. Bertoluzza, *The discrete commutator property of approximation spaces*, C. R. Acad. Sci. Paris Ser. I Math., 329 (1999), pp. 1097–1102.

[3] L. Bourgeois and L. J. Dard, *About stability and regularisation of ill-posed elliptic Cauchy problems: The case of Lipschitz domains*, Appl. Anal., 89 (2010), pp. 1745–1768.

[4] J. Bramble, R. Lazarov, and J. E. Pasciak, *Least-squares for second-order elliptic problems*, Comput. Methods Appl. Mech. Engrg., 152(1–2) (1998), pp. 195–210.

[5] E. Burman, *A stabilized finite element method for nonsymmetric, noncoercive, and ill-posed problems. Part* II: *Hyperbolic equations*, arXiv:1308.0529, 2013.

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

[7] E. Burman, M. A. Fernández, and P. Hansbo, *Continuous interior penalty finite element method for Oseen's equations*, SIAM J. Numer. Anal., 44 (2006), pp. 2012–2033.

[8] I. Capuzzo-Dolcetta and I. S. Finzi Vita, *Finite element approximation of some indefinite elliptic problems*, Calcolo, 25 (1988), pp. 379–395.

[9] C. Chainais-Hillairet and J. Droniou, *Finite-volume schemes for noncoercive elliptic problems with Neumann boundary conditions*, IMA J. Numer. Anal., 31 (2011), pp. 61–85.

[10] J. Droniou, *Non-coercive linear elliptic problems.* Potential Anal., 17 (2002), pp. 181–203.

[11] J. Droniou, T. Gallouët, and R. Herbin, *A finite volume scheme for a noncoercive elliptic equation with measure data*, SIAM J. Numer. Anal., 41 (2003), pp. 1997–2031.

[12] J. Droniou and J.-L. Vàzquez, *Noncoercive convection-diffusion elliptic problems with Neumann boundary conditions*, Calc. Var. Partial Differential Equations, 34 (2009), pp. 413–434.

[13] R. S. Falk and P. B. Monk, *Logarithmic convexity for discrete harmonic functions and the approximation of the Cauchy problem for Poisson's equation*, Math. Comp., 47 (1986), pp. 135–149.

[14] F. Hecht, O. Pironneau, A. Le Hyaric, and K. Ohtsuka, *FreeFEM++ v.* 2.11. *User's Manual*, University of Paris 6, 2005.

[15] K. Kavaliou and L. Tobiska, *A finite element method for a noncoercive elliptic problem with Neumann boundary conditions*, Comput. Methods Appl. Math., 12 (2012), pp. 168–183.

[16] J. Ku, *A least-squares method for second order noncoercive elliptic partial differential equations*, Math. Comp., 76 (2007), pp. 97–114.

[17] W. Lucht, *A finite element method for an ill-posed problem*, Appl. Numer. Math., 18 (1995), pp. 253–266.

[18] H.-J. Reinhardt, H. Han, and N. H. Dinh, *Stability and regularization of a discrete approximation to the Cauchy problem for Laplace's equation*, SIAM J. Numer. Anal., 36 (1999), pp. 890–905.

[19] A. Schatz, *An observation concerning Ritz-Galerkin methods with indefinite bilinear forms,* Math. Comp., 28 (1974), pp. 959–962.

[20] L. Zhu, E. Burman, and H. Wu, *Continuous Interior Penalty Finite Element Method for Helmholtz Equation with High Wave Number: One Dimensional Analysis*, arXiv:1211.1424, 2012.

[21] H. Wu, *Continuous Interior Penalty Finite Element Methods for the Helmholtz Equation with Large Wave Number*, arXiv:1106.4079, 2011.