Karlsruhe Institute of Technology

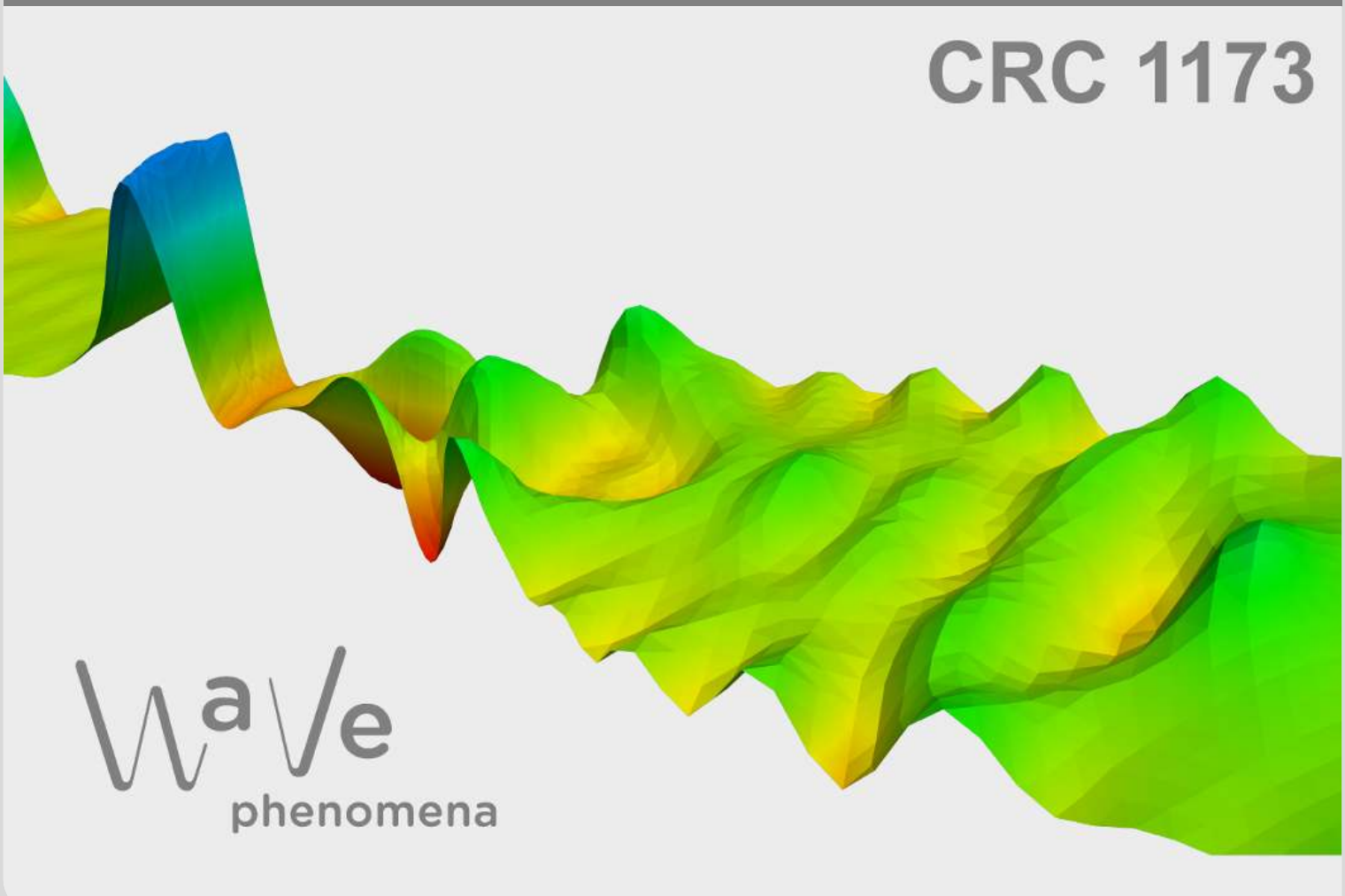
wave phenomena
analysis and numerics

Numerical approximation of planar oblique derivative problems in nondivergence form

Dietmar Gallistl

CRC Preprint 2017/30, November 2017

KARLSRUHE INSTITUTE OF TECHNOLOGY



Participating universities

Universität Stuttgart

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Funded by

and

Klaus Tschira Stiftung
gemeinnützige GmbH

ISSN 2365-662X

NUMERICAL APPROXIMATION OF PLANAR OBLIQUE DERIVATIVE PROBLEMS IN NONDIVERGENCE FORM

DIETMAR GALLISTL

Abstract. A numerical method for approximating a uniformly elliptic oblique derivative problem in two-dimensional simply-connected domains is proposed. The numerical scheme employs a mixed formulation with piecewise affine functions on curved finite element domains. The direct approximation of the gradient of the solution turns the oblique derivative boundary condition into an oblique direction condition. A priori and a posteriori error estimates as well as numerical computations on uniform and adaptive meshes are provided.

1. Introduction

This article deals with the approximation of elliptic nondivergence form partial differential equations (PDEs) in a planar domain $\Omega \subseteq \mathbb{R}^2$ subject to boundary conditions involving an oblique derivative. Such problems may arise in linearizations of fully nonlinear problems with transport boundary conditions. The linear model problem is to find a function u with $\int_\Omega u\,dx = 0$ such that

$$\begin{aligned} A : D^2 u := \sum_{j,k=1}^{2} A_{jk}\partial^2_{jk} u = f \qquad & \text{a.e. in } \Omega \\ \text{and} \qquad \nabla u \cdot \ell \ \text{ is constant } \quad & \text{on } \partial\Omega \end{aligned} \tag{1}$$

for a given unit vector field ℓ. This problem is easily generalized to an inhomogeneous oblique boundary condition of the type "$\nabla u \cdot \ell = g$ up to some constant" and the focus of this work is on (1), which corresponds to the choice $g = 0$. It is well known (and already relevant in standard Neumann problems) that for a boundary condition $\nabla u \cdot \ell = g$ the data f and g need to satisfy a compatibility condition. The formulation with equality up to some constant simply bypasses this technicality; the idea goes back to [21]. See [13] for the derivation of this problem from more general oblique derivative problems. The coefficient $A \in L^\infty(\Omega; \mathbb{R}^{2\times 2})$ is assumed to satisfy uniform ellipticity

$$0 < \alpha_1 = \inf_{\substack{\xi\in\mathbb{R}^2 \\ |\xi|=1}} \xi^* A\xi \le \sup_{\substack{\xi\in\mathbb{R}^2 \\ |\xi|=1}} \xi^* A\xi = \alpha_2 < \infty \quad \text{a.e. in } \Omega \tag{2}$$

and thereby (in this planar case) also the so-called *Cordes condition*

$$\frac{|A|^2}{(\operatorname{tr} A)^2} \le \frac{1}{1+\varepsilon} \qquad \text{for some } 0 < \varepsilon \le 1 \tag{3}$$

2010 *Mathematics Subject Classification.* 65N12, 65N15, 65N30,

Key words and phrases. oblique derivative problem, nondivergence form, Cordes coefficents, a priori error analysis, a posteriori error analysis.

Supported by the Deutsche Forschungsgemeinschaft (DFG) through CRC 1173.

where $|\cdot|$ is the Frobenius norm and tr denotes the trace. The Cordes condition is an algebraic condition on the coefficient that quantifies an appropriate closeness between A and the identity matrix. For the analysis of PDEs with discontinuous coefficients under the Cordes condition, the reader is referred to the monograph [13]. While in two space dimensions, (3) is implied by the classical condition (2) (see [13, 16]), it is an essential condition for problems in nondivergence form in higher space dimensions. The well-posedness of (1) requires additional conditions on the unit vector field ℓ and the domain Ω, which shall be specified in §§2–3 below.

The numerical analysis of elliptic equations in nondivergence form satisfying the Cordes condition started with the discontinuous Galerkin scheme proposed in [18] and was generalized in [19, 20] to stationary and parabolic Hamilton–Jacobi–Bellman problems. The work [11] analyzes a mixed discretization and derives a posteriori error estimates. For the numerical discretization of nondivergence form equations with continuous coefficients, see [9, 8, 12, 14].

This contribution builds upon the numerical scheme from [11] and focusses on the model problem (1) in simply-connected planar $\mathcal{C}^2$ domains; the mathematical analysis of the PDE can be found in [13]. A clear advantage of a mixed approach, where the gradient ∇u is approximated with an independent variable w, is that the oblique derivative boundary condition simplifies to the oblique direction boundary condition '$w \cdot \ell$ is constant', which can be easily incorporated in the finite element formulation. Still, the resulting scheme based on piecewise affines on curved finite elements turns out to be necessarily nonconforming in the sense that the boundary condition on the approximation of w is only enforced in the finite element vertices on the boundary.

The analysis of well-posedness of the discrete equations as well as the error analysis hinge on new generalizations of some existing estimates that bound the L^2 norm of the Laplacian of a function by the norm of the Hessian plus contributions on the boundary. Estimates of this type are sometimes referred to as *Miranda–Talenti* estimates in the literature and this nomenclature will be pursued throughout this work. The careful analysis of certain additional boundary terms is the key to the design of a stabilized finite element scheme. Furthermore, those tools are required for proving a novel discrete Poincaré–Friedrichs inequality, which is utilized in the analysis of the method. The error analysis comprises quasi-optimal a priori error estimates as well as the derivation of a reliable and efficient a posteriori error estimator. The practical performance of the scheme is studied in numerical experiments on uniform as well as on adaptive meshes.

The remaining parts of this paper are organized as follows. §2 clarifies the required notation and proves generalized Miranda–Talenti estimates. They bound the derivative of a vector field (in the L^2 norm) by its rotation, its divergence, and certain boundary terms. These estimates are required to design stabilized finite element schemes. §3 states the mixed formulation and proves its equivalence to the original problem. The finite element scheme is presented and analyzed in §4; §5 concludes with numerical computations. Some technical proofs can be found in Appendix A–C.

Standard notation on function spaces applies throughout this article. Lebesgue and Sobolev functions with values in $\mathbb{R}^n$ are denoted by $L^2(\Omega;\mathbb{R}^n)$ with $L^2(\Omega) :=$

$L^2(\Omega;\mathbb{R})$, $H^1(\Omega;\mathbb{R}^n)$ with $H^1(\Omega) := H^1(\Omega;\mathbb{R})$, etc. The subspace of $L^2(\Omega)$ consisting of functions with vanishing integral over Ω is denoted by $L^2_0(\Omega)$. Furthermore, denote $\widetilde{H}^1(\Omega) := H^1(\Omega) \cap L^2_0(\Omega)$. The $n \times n$ identity matrix is denoted by $I_{n\times n}$. The inner product of real-valued $n \times n$ matrices A, B is denoted by $A : B = \sum_{j,k=1}^n A_{jk}B_{jk}$. The Frobenius norm of an $n \times n$ matrix A is denoted by $|A| := \sqrt{A : A}$; the trace reads $\operatorname{tr} A$. For vectors, $|\cdot|$ refers to the Euclidean length. The notation $a \lesssim b$ denotes an inequality $a \leq Cb$ up to a multiplicative constant C that does not depend on the mesh-size.

2. Variants of the Miranda–Talenti estimate

This section briefly describes the setting of [13, Chapter 1.5] and proceeds with some new generalized Miranda–Talenti estimates. In order to precisely state the problem under consideration, some notation is introduced. Let $\Omega \subseteq \mathbb{R}^2$ be an open, simply-connected and bounded domain with $\mathcal{C}^2$ boundary so that $\partial\Omega$ is a closed planar curve of class $\mathcal{C}^2$. Assume that the boundary $\partial\Omega$ of Ω is parametrized by the arc length (i.e. in the natural parametrization) through the continuous curve $\boldsymbol{x} : [0, L] \to \mathbb{R}^2$

$$\boldsymbol{x}(\varphi) = \begin{pmatrix} \boldsymbol{x}_1(\varphi) \\ \boldsymbol{x}_2(\varphi) \end{pmatrix} \qquad \text{for } \varphi \in [0, L] \qquad \text{with } \boldsymbol{x}(L) = \boldsymbol{x}(0).$$

The assumption that $\partial\Omega$ is $\mathcal{C}^2$ regular means that $\boldsymbol{x} \in \mathcal{C}^2([0, L];\mathbb{R}^2)$. The derivative of a function $v : [0, L] \to \mathbb{R}$ with respect to the arc-length parameter φ is denoted by $\dot{v}$. Analogously, $\ddot{v}$ denotes the second derivative of v.

Let $\nu = (\nu_1, \nu_2)$ denote the outward pointing unit normal to $\partial\Omega$ and let $t = \dot{\boldsymbol{x}}$ denote the unit tangent vector. Denote the curvature of $\partial\Omega$ at $\boldsymbol{x}(\varphi)$ by $\chi(\varphi)$. The orientation of the parametrization is chosen such that

$$\nu(\boldsymbol{x}(\varphi)) = \begin{bmatrix} \nu_1(\boldsymbol{x}(\varphi)) \\ \nu_2(\boldsymbol{x}(\varphi)) \end{bmatrix} = \begin{bmatrix} \dot{\boldsymbol{x}}_2(\varphi) \\ -\dot{\boldsymbol{x}}_1(\varphi) \end{bmatrix}, \quad \chi(\varphi) = \ddot{\boldsymbol{x}}_1(\varphi)\dot{\boldsymbol{x}}_2(\varphi) - \dot{\boldsymbol{x}}_1(\varphi)\ddot{\boldsymbol{x}}_2(\varphi). \tag{4}$$

Let $\ell : [0, L] \to \mathbb{R}^2$ with $\ell(0) = \ell(L)$ be a $\mathcal{C}^2$-regular unit vector field. Let $\vartheta(\varphi)$ denote the oriented angle (modulo 2π) between $\nu(\boldsymbol{x}(\varphi))$ and $\ell(\varphi)$. For an illustration see Figure 1. Clearly $\vartheta : [0, L] \to \mathbb{R}$ is of class $\mathcal{C}^1$. Then

$$\dot{\ell}_2\ell_1 - \dot{\ell}_1\ell_2 = \dot{\vartheta} - \chi. \tag{5}$$

Identity (5) is shown in [13, p. 48] and the proof is briefly repeated here for convenient reading. Since $\dot{\boldsymbol{x}}$ is a $\mathcal{C}^1$ unit vector field, there exists a function $\psi \in \mathcal{C}^1([0, L])$ such that

$$\dot{\boldsymbol{x}}_1(\varphi) = \cos\psi(\varphi) \quad \text{and} \quad \dot{\boldsymbol{x}}_2(\varphi) = \sin\psi(\varphi) \quad \text{for all } \varphi \in [0, L].$$

The function ψ is the oriented angle between the x_1 axis and the tangent vector t to $\partial\Omega$ as displayed in Figure 1. Since $\dot{\boldsymbol{x}} = t = (\cos\psi, \sin\psi)$, it holds that

$$-\chi = \dot{\boldsymbol{x}}_1(\varphi)\ddot{\boldsymbol{x}}_2(\varphi) - \ddot{\boldsymbol{x}}_1(\varphi)\dot{\boldsymbol{x}}_2(\varphi) = (\cos^2\psi + \sin^2\psi)d\psi/d\varphi = \dot{\psi}.$$

Let $\omega \in \mathcal{C}^1([0, L])$ denote the oriented angle (modulo 2π) between the x_1 axis and ℓ. Then, obviously, $\psi = \omega + \pi/2 - \vartheta$ as well as $\dot{\omega} = \dot{\vartheta} + \dot{\psi} = \dot{\vartheta} - \chi$. On the other hand, $\ell = (\cos\omega, \sin\omega)$ implies $\dot{\omega} = \dot{\ell}_2\ell_1 - \dot{\ell}_1\ell_2$, and comparing the two expressions obtained for $\dot{\omega}$ proves (5).

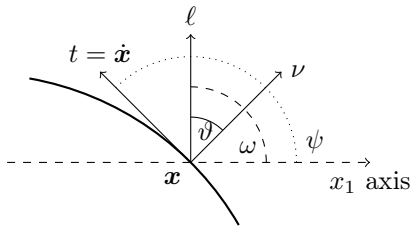


FIGURE 1. Illustration of the geometric setting.

In what follows, the vector field $\ell^\perp$ is defined as $\ell^\perp := (-\ell_2, \ell_1)$. For a vector field w, the notation $w_\ell = w \cdot \ell$ and $w_\perp = w \cdot \ell^\perp$ abbreviates

$$w_\ell(\varphi) = w_1(\boldsymbol{x}(\varphi))\ell_1(\varphi) + w_2(\boldsymbol{x}(\varphi))\ell_2(\varphi)$$
$$\text{and} \quad w_\perp(\varphi) = w_2(\boldsymbol{x}(\varphi))\ell_1(\varphi) - w_1(\boldsymbol{x}(\varphi))\ell_2(\varphi) \qquad \text{for all } \varphi \in [0, L].$$

Remark 1 (convention on notation)**.** Although ℓ is formally defined as function over the interval $[0, L]$, it can be identified with a function on $\partial\Omega$. Notation like $\ell(\boldsymbol{x}(\varphi))$ instead of $\ell(\varphi)$ will sometimes be used. This is often advantageous, for example when the product of ℓ with functions defined on $\partial\Omega$ is considered.

The following lemma generalizes [13, Lemma 1.5.5] which therein is crucial for the analysis of well-posedness of (1). The refined result is essential for the stabilization technique utilized in the numerical scheme.

Lemma 2. *Any $w \in H^1(\Omega; \mathbb{R}^2)$ that is piecewise smooth with respect to a given triangulation $\mathcal{T}$ (with curved elements) of $\overline{\Omega}$ satisfies*

$$2\int_\Omega (\partial_1 w_1 \partial_2 w_2 - \partial_2 w_1 \partial_1 w_2)\, dx = \int_0^L |w|^2(\dot{\vartheta} - \chi)\, d\varphi - 2\int_0^L \dot{w}_\ell w_\perp\, d\varphi.$$

Proof. The proof is postponed to Appendix A. □

The following result generalizes the classical Miranda–Talenti estimate, which reads $\|\Delta u\|_{L^2(\Omega)} \le \|D^2 u\|_{L^2(\Omega)}$ for all functions $u \in H^2(\Omega) \cap H_0^1(\Omega)$ on polytopes as well as on convex domains. The additional terms involve quantities related to ℓ and the curvature of $\partial\Omega$.

Corollary 3 (generalized Miranda–Talenti estimate)**.** *Let $v \in H^1(\Omega; \mathbb{R}^2)$ such that $v \cdot \ell$ is constant along $\partial\Omega$ and let $v_h \in H^1(\Omega; \mathbb{R}^2)$ be piecewise smooth with respect to a given triangulation $\mathcal{T}$ of $\overline{\Omega}$. Then $w := v + v_h$ satisfies*

$$\|Dw\|_{L^2(\Omega)}^2 = \|\operatorname{div} w\|_{L^2(\Omega)}^2 + \|\operatorname{rot} w\|_{L^2(\Omega)}^2 - \int_0^L |w|^2(\dot{\vartheta} - \chi)\, d\varphi + 2\int_0^L \dot{w}_\ell w_\perp\, d\varphi.$$

If $\dot{\vartheta} - \chi \ge 0$ on $[0, L]$, then

$$\|Dw\|_{L^2(\Omega)}^2 \le \|\operatorname{div} w\|_{L^2(\Omega)}^2 + \|\operatorname{rot} w\|_{L^2(\Omega)}^2 + 2\int_0^L \dot{w}_\ell w_\perp\, d\varphi.$$

Proof. For piecewise smooth w, this follows from combining the pointwise identity

$$|Dw|^2 = |\operatorname{div} w|^2 + |\operatorname{rot} w|^2 - 2(\partial_1 w_1 \partial_2 w_2 - \partial_2 w_1 \partial_1 w_2)$$

with Lemma 2. The general case follows from an approximation argument. Indeed, any $v \in H^1(\Omega; \mathbb{R}^2)$ such that $v \cdot \ell = c$ is constant can be approximated in the H^1

norm by piecewise affine and globally continuous functions $(\mu_n)_{n\geq 0}$ on a sequence of meshes $\mathcal{T}_n$ satisfying $(\mu_n \cdot \ell)(z) = c$ at all boundary vertices z of the triangulation $\mathcal{T}_n$ (as constructed in Proposition 17 below). The inverse and the trace inequality show that the boundary term vanishes in the limit. Such arguments are discussed in more detail in the proof of Lemma 4 below. This establishes the claimed estimate for functions of the type $w = v + v_h$. □

The subsequent result is a Poincaré–Friedrichs type inequality. It generalizes [13, Lemma 1.5.8] and, moreover, states how the involved constant depends on the data.

Lemma 4 (Poincaré–Friedrichs inequality)**.** *Let $v \in H^1(\Omega;\mathbb{R}^2)$ such that $v \cdot \ell$ is constant along $\partial\Omega$ and let $v_h \in H^1(\Omega;\mathbb{R}^2)$ be piecewise smooth with respect to a given triangulation $\mathcal{T}$ of $\overline{\Omega}$. If $\dot\vartheta - \chi > 0$ on $[0, L]$, then $w := v + v_h$ satisfies*

$$\|w\|^2_{L^2(\Omega)} \leq C(\Omega,\ell)\left(\|\operatorname{div} w\|^2_{L^2(\Omega)} + \|\operatorname{rot} w\|^2_{L^2(\Omega)} + 2\int_0^L \dot w_\ell w_\perp \, d\varphi\right)$$

for the constant

$$C(\Omega,\ell) := \frac{4\operatorname{diam}(\Omega)^2}{\min\{1, 2\operatorname{diam}(\Omega)\min_{\varphi\in[0,T]}(\dot\vartheta - \chi)\}}.$$

Proof. Assume, without loss of generality, that $0 \in \Omega$ so that $\max\{|x_1|, |x_2|\} \leq \operatorname{diam}(\Omega)$. Then, an elementary calculation reveals the pointwise relation

$$|w|^2 = \operatorname{div}\begin{pmatrix} x_1 w_1^2 \\ x_2 w_2^2\end{pmatrix} - 2(x_1 w_1 \partial_1 w_1 + x_2 w_2 \partial_2 w_2). \tag{6}$$

The fact that $\max\{|x_1|, |x_2|\} \leq \operatorname{diam}(\Omega)$ together with Young's inequality

$$2\operatorname{diam}(\Omega)ab \leq 2^{-1}a^2 + 2\operatorname{diam}(\Omega)^2 b^2 \qquad \text{for any } a, b \geq 0$$

show that

$$\begin{aligned} |2(x_1 w_1 \partial_1 w_1 + x_2 w_2 \partial_2 w_2)| &\leq 2\operatorname{diam}(\Omega)(|w_1||\partial_1 w_1| + |w_2||\partial_2 w_2|) \\ &\leq 2^{-1}|w|^2 + 2\operatorname{diam}(\Omega)^2 |Dw|^2. \end{aligned} \tag{7}$$

Integrating (6) over Ω and applying the divergence theorem and (7) results in

$$\begin{aligned} \frac{1}{2}\|w\|^2_{L^2(\Omega)} &\leq \int_{\partial\Omega}\begin{pmatrix} x_1 w_1^2 \\ x_2 w_2^2\end{pmatrix}\cdot \nu \, ds + 2\operatorname{diam}(\Omega)^2\|Dw\|^2_{L^2(\Omega)} \\ &\leq \operatorname{diam}(\Omega)\|w\|^2_{L^2(\partial\Omega)} + 2\operatorname{diam}(\Omega)^2\|Dw\|^2_{L^2(\Omega)}. \end{aligned}$$

Thus,

$$\|w\|^2_{L^2(\Omega)} \leq 2\operatorname{diam}(\Omega)\|w\|^2_{L^2(\partial\Omega)} + 4\operatorname{diam}(\Omega)^2\|Dw\|^2_{L^2(\Omega)}. \tag{8}$$

Let

$$\eta := \min\{1, 2\operatorname{diam}(\Omega)\min_{\varphi\in[0,T]}(\dot\vartheta - \chi)\}.$$

Then $\eta \leq 1$ and, therefore, Corollary 3 leads to

$$\begin{aligned}\|Dw\|^2_{L^2(\Omega)} &\leq \eta^{-1}\|Dw\|^2_{L^2(\Omega)} \\ &\leq \eta^{-1}\left(\|\operatorname{div} w\|^2_{L^2(\Omega)} + \|\operatorname{rot} w\|^2_{L^2(\Omega)} + 2\int_0^L \dot{w}_\ell w_\perp \, d\varphi\right) \\ &\qquad - \|w\|^2_{L^2(\partial\Omega)}\frac{\min_{\varphi\in[0,T]}(\dot{\vartheta}-\chi)}{\eta}.\end{aligned}$$

The definition of η implies that $\min_{\varphi\in[0,T]}(\dot{\vartheta}-\chi)/\eta \geq (2\operatorname{diam}(\Omega))^{-1}$, which shows

$$\|Dw\|^2_{L^2(\Omega)} \leq \eta^{-1}\left(\|\operatorname{div} w\|^2_{L^2(\Omega)} + \|\operatorname{rot} w\|^2_{L^2(\Omega)} + 2\int_0^L \dot{w}_\ell w_\perp \, d\varphi\right) - \frac{\|w\|^2_{L^2(\partial\Omega)}}{2\operatorname{diam}(\Omega)}.$$

The multiplication with $4\operatorname{diam}(\Omega)^2$ and the combination with (8) conclude the proof. □

3. The model problem and its variational formulation

The well-posedness result involves an additional condition on the winding number of ℓ, namely

$$\frac{\vartheta(L)-\vartheta(0)}{2\pi} = 0. \tag{9}$$

This means that, in total, ℓ does not perform a turn around the normal ν. The result from [13] reads as follows.

Proposition 5. *Let $A \in L^\infty(\Omega;\mathbb{R}^{2\times 2})$ satisfy the uniform ellipticity (2), let $\partial\Omega$ be of class $\mathcal{C}^2$ and ℓ of class $\mathcal{C}^1$. Let furthermore $\dot{\vartheta}-\chi > 0$ on $\partial\Omega$ and (9) be satisfied. Then, for given $f \in L^2(\Omega)$, problem (1) has a unique solution $u \in H^2(\Omega)\cap L^2_0(\Omega)$.*

Proof. For a proof, see [13, Proposition 1.5.13]. □

In view of Proposition 5 the following assumption on the data is made.

Assumption 6. *The data $A \in L^\infty(\Omega;\mathbb{R}^{2\times 2})$ satisfies (2), The boundary $\partial\Omega$ is of class $\mathcal{C}^2$. The unit vector field ℓ on $\partial\Omega$ is of class $\mathcal{C}^2$. Furthermore $\dot{\vartheta}-\chi > 0$ on $\partial\Omega$ and (9) is satisfied.*

Remark 7. The well-posedness of the model merely requires $\mathcal{C}^1$ regularity of ℓ, see Proposition 5. The higher regularity of ℓ in Assumption 6 will be required for the analysis of the numerical method in §4.

As proposed in [11], (1) is replaced by an equivalent mixed problem. Define the space

$$W^\ell := \{v \in H^1(\Omega;\mathbb{R}^2) : v\cdot\ell \text{ is constant on } \partial\Omega\}. \tag{10}$$

With the same reasoning as in [10, 11] it is verified that ∇u is characterized as the unique field $w \in W^\ell$ with $\operatorname{rot} w = 0$ and $A : Dw = f$ in Ω. In order to state these relations in a mixed system, a suitable space Q of Lagrange multipliers is required. In [11] the boundary condition was such that the operator rot was surjective onto the space of all L^2 function with vanishing average. In the present case, the operator $\operatorname{rot} : W^\ell \to L^2(\Omega)$ is surjective onto the whole $L^2(\Omega)$.

Lemma 8. *There exists a constant $\beta_1 > 0$ (depending on ℓ) such that for any $q \in L^2(\Omega)$ there exists $v \in W^\ell$ with $\operatorname{rot} v = q$ and $\|Dv\|_{L^2(\Omega)} \le \beta_1^{-1}\|q\|_{L^2(\Omega)}$.*

Proof. The author believes that the proof is essentially known. For completeness, it is shown in Appendix B. □

The variational formulation of (1) is based on a proper choice of test functions. Define the test-function operators

$$\tau^{\mathrm{NS}}(\phi) := \gamma \operatorname{div}\phi \quad \text{and} \quad \tau^{\mathrm{LS}}(\phi) := A : D\phi \qquad \text{for any } \phi \in H^1(\Omega;\mathbb{R}^2). \tag{11}$$

for

$$\gamma := \frac{\operatorname{tr}(A)}{|A|^2}. \tag{12}$$

The operator τ^{NS} was proposed in [18] (NS abbreviates "nonsymmetric") while the choice τ^{LS} is from [11] (LS abbreviates "least squares"). These operators are used to state well-posed variational formulations on the continuous level.

Recall the definition of W^ℓ and define the space

$$Q := L^2(\Omega).$$

Define the bilinear forms $a_\tau : W^\ell \times W^\ell \to \mathbb{R}$ (for $\tau = \tau^{\mathrm{NS}}$ or $\tau = \tau^{\mathrm{LS}}$ defined in (11)) and $b : W^\ell \times Q \to \mathbb{R}$ by

$$\begin{aligned} a_\tau(v,z) &:= (A : Dv, \tau(z))_{L^2(\Omega)} && \text{for any } (v,z) \in W^\ell \times W^\ell, \\ b(v,q) &:= (\operatorname{rot} v, q)_{L^2(\Omega)} && \text{for any } (v,q) \in W^\ell \times Q. \end{aligned}$$

In view of Lemma 8, there exists a constant $\beta > 0$ such that the following inf-sup condition is satisfied

$$\beta \le \inf_{q \in L^2(\Omega)\setminus\{0\}} \sup_{v \in W^\ell \setminus \{0\}} \frac{b(v,q)}{\|Dv\|_{L^2(\Omega)}\|q\|_{L^2(\Omega)}}. \tag{13}$$

Recall the definition $\tilde{H}^1(\Omega) := H^1(\Omega) \cap L^2_0(\Omega)$. The general mixed formulation of (1) is to seek $(u,w,p) \in \tilde{H}^1(\Omega) \times W^\ell \times Q$ such that

$$(\nabla u, \nabla \eta)_{L^2(\Omega)} = (w, \nabla\eta)_{L^2(\Omega)} \qquad \text{for all } \eta \in \tilde{H}^1(\Omega), \tag{14a}$$

$$a_\tau(w,v) + b(v,p) = (f, \tau(v))_{L^2(\Omega)} \qquad \text{for all } v \in W^\ell, \tag{14b}$$

$$b(w,q) = 0 \qquad \text{for all } q \in Q. \tag{14c}$$

The structure of system (14) resembles that considered in [11]; the difference being that (14a) is a Neumann problem.

Remark 9. System (14) is more general than required for the treatment of the model problem (1). While it covers the case of more general fourth-order differential operators (see Remark 26 for comments), it will turn out below that for (1) the choice $Q = \{0\}$ and a suitable stabilization are sufficient.

Proposition 10. *Let Assumption 6 be satisfied. Then there exists a unique solution $(u,w,p) \in \tilde{H}^1(\Omega) \times W^\ell \times Q$ satisfying (14). Moreover, $w = \nabla u$ and u is the unique solution to (1).*

Proof. It is first shown that a_τ is coercive on the kernel of b, that is, on the subspace of W^ℓ satisfying $\operatorname{rot} = 0$. It is known [13, 18] that the Cordes condition (3) implies almost everywhere in Ω

$$|\gamma A - I_{d\times d}| \leq \sqrt{1-\varepsilon} \tag{15}$$

with γ from (12). Then, for $\tau = \tau^{\mathrm{NS}}$ and any $v \in W^\ell$ with $\operatorname{rot} v = 0$, the triangle and Cauchy inequalities together with Corollary 3 show

$$\begin{aligned}(A : Dv, \tau(v))_{L^2(\Omega)} &= (A : Dv, \gamma \operatorname{div} v)_{L^2(\Omega)} \\ &= \|\operatorname{div} v\|^2_{L^2(\Omega)} + ((\gamma A - I_{2\times 2}) : Dv, \operatorname{div} v)_{L^2(\Omega)} \\ &\geq (1-\sqrt{1-\varepsilon})\|\operatorname{div} v\|^2_{L^2(\Omega)} \geq (1-\sqrt{1-\varepsilon})\|Dv\|^2_{L^2(\Omega)}.\end{aligned}$$

Similarly, for $\tau = \tau^{\mathrm{LS}}$,

$$\|\gamma\|^2_{L^\infty(\Omega)}(A : D^2 v, \tau(\nabla v))_{L^2(\Omega)} \geq \|\gamma A : D^2 v\|^2_{L^2(\Omega)} \geq (1-\sqrt{1-\varepsilon})^2 \|D^2 v\|^2_{L^2(\Omega)}.$$

Thus, a_τ is coercive on the kernel of b and the well-posedness of (14b)–(14c) follows with standard arguments [2] from the theory of saddle-point problems because (13) is satisfied. Since $\operatorname{rot} w = 0$ and the domain Ω is simply-connected, w equals the gradient of an H^1 function which is unique up to an additive constant. The well-posed Neumann problem (14a) minimizes $\|\nabla v - w\|_{L^2(\Omega)}$ over all functions $v \in \tilde{H}^1(\Omega)$, whence $\nabla u = w$. Hence, the unique existence of a solution to (14) is shown. Moreover, $u \in \tilde{H}^1(\Omega) \cap H^2(\Omega)$. For $\tau = \tau^{\mathrm{NS}}$, equations (14b)–(14c) in particular imply

$$(A : D^2 u, \gamma\Delta\eta)_{L^2(\Omega)} = (f, \gamma\Delta\eta)_{L^2(\Omega)} \quad \text{for all } \eta \in \tilde{H}^1(\Omega) \cap H^2(\Omega)$$

(because in (14b) every test-function v in the kernel of b is the gradient of some $\eta \in \tilde{H}^1(\Omega) \cap H^2(\Omega)$). But since Proposition 5 with $A = I_{2\times 2}$ implies that Δ and so $\gamma\Delta$ is surjective from $\tilde{H}^1(\Omega) \cap H^2(\Omega)$ to $L^2(\Omega)$, the identity $A : D^2 u = f$ a.e. in Ω follows (such arguments were first utilized by [18]). The boundary condition "$w \cdot \ell$ is constant" implies that $\nabla u \cdot \ell$ is constant, and so u is the solution to (1), which is unique by Proposition 5. For $\tau = \tau^{\mathrm{LS}}$, it can be seen that $w = \nabla u$ and u minimizes $\|A : Du - f\|_{L^2(\Omega)}$ amongst all functions in $\tilde{H}^1(\Omega) \cap H^2(\Omega)$ with $\nabla u \cdot \ell =$ constant. □

As mentioned in Remark 9, the special structure of the right-hand side in (14b) enables an even simpler formulation without Lagrange multipliers. The mixed formulation (14) is based on a saddle-point formulation and covers the case of very general right-hand sides. In the present model problem, the right-hand side of the saddle-point problem has a very special structure and it enforces strong L^2 equality of the PDE. This is the reason why the multiplier p equals zero in this case. The proof of Proposition 10 revealed that the test function operator $\tau = \tau^{\mathrm{NS}}$ or $\tau = \tau^{\mathrm{LS}}$ is surjective onto $L^2(\Omega)$. Thus, in (14b), the multiplier p equals zero and (14b) is satisfied for all test functions $v \in W^\ell$. This proves that

$$(\nabla u, \nabla\eta)_{L^2(\Omega)} = (w, \nabla\eta)_{L^2(\Omega)}, \tag{16a}$$

$$a_\tau(w, v) + \sigma^\tau(\gamma, \varepsilon)^2(\operatorname{rot} w, \operatorname{rot} v)_{L^2(\Omega)} = (f, \tau(v))_{L^2(\Omega)} \tag{16b}$$

for all $\eta \in \tilde{H}^1(\Omega)$ and all $v \in W^\ell$ admits a unique solution $(u, w) \in \tilde{H}^1(\Omega) \times W^\ell$. This gives rise to numerical schemes with a positive definite formulation. Although both formulations may be discretized with similar techniques, the numerical scheme

presented here will rely on the positive definite formulation (16). For the more general case, the reader is referred to [11] for the Dirichlet problem and to the discussion on more general fourth-order problems in Remark 26 below. In order to filter out elements in the kernel, a stabilized or enriched bilinear form $\tilde{a}_\tau$ will be employed below. A second reason for utilizing a stabilization is that the scheme proposed here is nonconforming.

4. Finite element discretization

This section presents the numerical scheme for approximating (16).

4.1. **Curved finite elements.** There are various methods for approximating partial differential equations posed on curved domains with polygonal or isoparametric finite elements [3, 4]. In this article, curved finite elements are employed. The reason is that, for the class of discontinuous coefficients considered here, the solutions are not expected to exhibit the smoothness that would be required for controlling the error caused by the approximation of the domain. The planar domain Ω is regularly tesselated with a family $\mathcal{T}$ of triangles having at most one truly curved edge. It is assumed that the union of all curved edges is a subset of $\partial\Omega$ so that any triangle whose interior lies inside Ω has straight edges. A formal definition of a curved triangle is as follows.

Definition 11 (curved triangle)**.** A closed Lipschitz domain $T \subseteq \mathbb{R}^2$ is called a *curved triangle*, if the following is satisfied. There exist three points $(z_1, z_2, z_3) \in T^3$ that are not collinear (and so $\operatorname{conv}\{z_1, z_2, z_3\}$ is a triangle). There exists a planar curve $\Gamma \subseteq \mathbb{R}^2$ connecting z_2 with z_3 that can be represented as the graph of a $\mathcal{C}^2$ function over $E_{2,3} := \operatorname{conv}\{z_2, z_3\}$. The domain T is the bounded connectivity component of

$$\mathbb{R}^2 \setminus (\operatorname{conv}\{z_1, z_2\} \cup \Gamma \cup \operatorname{conv}\{z_3, z_1\}).$$

The points z_1, z_2, z_3 are called the *vertices* of T. The sets

$$\operatorname{conv}\{z_1, z_2\}, \quad \Gamma, \quad \operatorname{conv}\{z_3, z_1\}$$

are called the *edges* of T.

Let $\mathcal{T}$ denote a regular triangulation of (possibly curved) finite elements. The set of vertices is denoted by $\mathcal{N}$, and $\mathcal{N}(\Omega) := \mathcal{N} \cap \Omega$ is the set of interior vertices while $\mathcal{N}(\partial\Omega)$ denotes the vertices on the boundary. The set of faces reads $\mathcal{F}$; the set of faces that are subsets of the boundary reads $\mathcal{F}(\partial\Omega)$. For any $F \in \mathcal{F}$, let h_F denote its length. For any $T \in \mathcal{T}$, denote $h_T := \operatorname{diam}(T)$ and let $h := h_{\mathcal{T}}$ denote the piecewise constant mesh-size function with $h|_T = h_T$.

The triangulation $\mathcal{T}$ is assumed to satisfy the following admissibility conditions.

Definition 12 (admissible triangulation)**.** A collection $\mathcal{T}$ of curved triangles is said to be an *admissible triangulation* of Ω if it is a regular triangulation, each triangle is star-shaped with respect to a ball, and every edge E with $E \not\subseteq \partial\Omega$ is a straight line. An example triangulation is displayed in Figure 2.

Definition 13 (shape regularity, cf. [4])**.** Let $\mathcal{T}$ be an admissible triangulation of Ω. The smallest constant $\rho > 0$ such that for any $T \in \mathcal{T}$

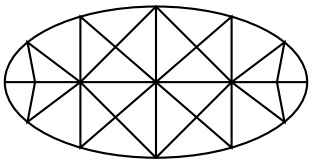


FIGURE 2. An admissible triangulation with curved elements of an ellipse.

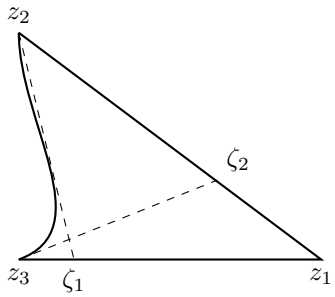


FIGURE 3. Maximal inscribed triangles; illustration to Definition 14.

(i) there exist concentric circular discs D_1, D_2 such that $D_1 \subseteq T \subseteq D_2$ and

$$\frac{\operatorname{diam}(D_2)}{\operatorname{diam}(D_1)} \leq \rho$$

(ii) any two edges F, F' of T satisfy

$$\frac{\operatorname{length}(F)}{\operatorname{length}(F')} \leq \rho$$

is called the *shape-regularity constant*.

Definition 14 (inscribed triangles)**.** Let T be a curved triangle with vertices z_1, z_2, z_3 and the two non-curved edges $\operatorname{conv}\{z_1, z_2\}$ and $\operatorname{conv}\{z_3, z_1\}$ (convention as in Definition 11). Let $\zeta_1 \in \operatorname{conv}\{z_3, z_1\}$ and $\zeta_2 \in \operatorname{conv}\{z_1, z_2\}$ be such that

$$K_1 := \operatorname{conv}\{z_1, z_2, \zeta_1\} \subseteq T \quad \text{and} \quad K_2 := \operatorname{conv}\{z_1, z_2, \zeta_1\} \subseteq T$$

and the area of K_1 and K_2 is maximal. Then K_1, K_2 are called the *maximal inscribed triangles* of T. See Figure 3 for an illustration.

Lemma 15 (stability of L^2 projection)**.** *Let T be a curved triangle, let $v \in H^1(\operatorname{int}(T))$ and let p_1 be the $L^2(T)$-best approximating affine function. There exists a constant that only depends on the shape regularity of T and its maximal inscribed triangles as well as on the chunkiness parameter of T, such that*

$$\|Dp_1\|_{L^2(T)} \leq C\|Dv\|_{L^2(T)}.$$

Proof. Since Dp_1 is constant over T, the shape regularity shows for any K^{in} of the maximal inscribed triangles that

$$\|Dp_1\|_{L^2(T)} \lesssim \|Dp_1\|_{L^2(K^{in})}.$$

Let $M_v := \fint_T v\,dx$ denote the mean of v over T. The well-known inverse estimate on triangles [4] with constants only dependent on the interior angles of K^{in} lead to

$$\begin{aligned}\|Dp_1\|_{L^2(K^{in})} &= \|D(p_1 - M_v)\|_{L^2(K^{in})} \\ &\lesssim h_T^{-1}\|p_1 - M_v\|_{L^2(K^{in})} \leq h_T^{-1}\|p_1 - M_v\|_{L^2(T)}.\end{aligned}$$

Since $p_1 - M_v$ is the $L^2(T)$-best approximation to $v - M_v$, the L^2 stability of the L^2 projection and the foregoing formulas result in

$$\|Dp_1\|_{L^2(T)} \lesssim h_T^{-1}\|v - M_v\|_{L^2(T)}.$$

The proof is concluded with the Poincaré inequality, whose constant is known to be only dependent on the domain's chunkiness parameter [4, Lemma 4.3.14]. □

Unlike in the case of parametric finite elements [3], the discrete spaces employed here are based on polynomials on the physical elements. Given a nonnegative integer k, the piecewise polynomial function over $\omega \subseteq \mathbb{R}^2$ of degree not larger than k are denoted by $P_k(\omega)$. The piecewise polynomial spaces read

$$\begin{aligned}P_k(\mathcal{T}) &:= \{v \in L^1(\Omega) : v|_T \in P_k(T) \text{ for any } T \in \mathcal{T}\},\\ S^k(\mathcal{T}) &:= P_k(\mathcal{T}) \cap H^1(\Omega)\end{aligned}$$

with the usual notation $P_k(\mathcal{T};\mathbb{R}^2)$, $S^k(\mathcal{T};\mathbb{R}^2)$, etc., for vector-valued functions. For simplicity, the discretization method proposed here is based on curvilinear P_1 finite elements. For more general versions of curved finite elements the reader is referred to [4] and [17]. The discrete space W_h^ℓ is the following subspace of the continuous and piecewise affine fields

$$W_h^\ell := \left\{ v \in S^1(\mathcal{T};\mathbb{R}^2) : \begin{array}{l}\text{there exists } c_v \in \mathbb{R} \text{ such that}\\ v(z)\cdot\ell(z) = c_v \text{ for all } z \in \mathcal{N}(\partial\Omega)\end{array} \right\}. \tag{17}$$

Recall that the set of boundary vertices is denoted by $\mathcal{N}(\partial\Omega)$. Note that the approximation W_h^ℓ of W^ℓ is generally be nonconforming in the sense that $W_h^\ell \not\subseteq W^\ell$. Indeed, the property that $v \cdot \ell$ is constant on the boundary will typically fail to hold. Approximation properties of W_h^ℓ can be derived through quasi-interpolation operators, which have been studied since the seminal work of [5]. For the present analysis, the Oswald quasi-interpolation operator [15] is chosen. The operator $J : W^\ell \to W_h^\ell$ is defined as follows. Given $v \in W^\ell$, let $\Pi v \in P_1(\mathcal{T};\mathbb{R}^2)$ denote its L^2 best-approximation by piecewise affine (possibly globally discontinuous) functions. For any interior vertex $z \in \mathcal{N}(\Omega)$ of the triangulation define

$$(Jv)(z) := \frac{1}{\operatorname{card}(\{T \in \mathcal{T} : z \in T\})} \sum_{\substack{T\in\mathcal{T}\\ \text{with } z\in T}} (\Pi v)|_T(z). \tag{18a}$$

For any vertex $z \in \mathcal{N}(\partial\Omega)$ on the boundary, let

$$\begin{aligned}&(Jv)(z)\cdot\ell^\perp(z) := \frac{1}{\operatorname{card}(\{T \in \mathcal{T} : z \in T\})} \sum_{\substack{T\in\mathcal{T}\\ \text{with } z\in T}} (\Pi v)|_T(z)\cdot\ell^\perp(z),\\ &\text{and}\quad (Jv)(z)\cdot\ell(z) := v(z)\cdot\ell(z)\end{aligned} \tag{18b}$$

(recall that $v\cdot\ell$ is constant on the boundary, which gives a meaning to the expression $v(z)\cdot\ell(z)$). In other words, J is the concatenation of Π with the averaging operator that assigns to each vertex the arithmetic mean of the corresponding function values of the neighbouring cells while enforcing the discrete boundary condition of W_h^ℓ.

Remark 16 (compactness arguments)**.** Some of the estimates derived in this section are not explicitly quantified in dependence of the geometry. This is the case when compactness arguments involving curved elements are utilized. Those steps will be clearly marked in the proofs.

The next result states local stability and approximation estimates for the operator J.

Proposition 17. *The operator $J : W^\ell \to W_h^\ell$ is a projection, i.e., $J \circ J = J$. There is a constant $C_J > 0$ such that any $v \in W$ and any $T \in \mathcal{T}$ satisfy*

$$h_T^{-1}\|v - Jv\|_{L^2(T)} + \|DJv\|_{L^2(T)} \leq C_J\|Dv\|_{L^2(\omega_T)}$$

and

$$\|Jv\|_{L^2(T)} \leq C_J\|v\|_{L^2(\omega_T)}$$

for the element patch ω_T, that is the interior of the union of all triangles of $\mathcal{T}$ having a nonempty intersection with T. Moreover, the following global best-approximation property holds

$$\|D(v - Jv)\|_{L^2(\Omega)} \lesssim \inf_{v_h \in W_h^\ell} \|D(v - v_h)\|_{L^2(\Omega)}.$$

Proof. The proof, although in principle known [4, 6], is briefly sketched in Appendix C to illustrate on which quantities the involved constants depend. □

Corollary 18 (approximation)**.** *Let $s \in [0, 1]$ and let $v \in W^\ell \cap H^{1+s}(\Omega; \mathbb{R}^2)$. Then*

$$\inf_{v_h \in W_h^\ell} \|D(v - v_h)\|_{L^2(\Omega)} \leq \|D(v - Jv)\|_{L^2(\Omega)} \lesssim \|h\|_\infty^s \|v\|_{H^{1+s}(\Omega)}$$

for the maximum mesh-size $\|h\|_\infty$.

Proof. For $s = 0$, the stability from Proposition 17 proves $\|D(v - Jv)\|_{L^2(\Omega)} \lesssim \|v\|_{H^1(\Omega)}$ and, for $s = 1$, the quasi-best approximation property of J and well known approximation error estimates (e.g., through nodal interpolation [4]) guarantee $\|D(v - Jv)\|_{L^2(\Omega)} \lesssim \|h\|_\infty \|v\|_{H^2(\Omega)}$. The dependence of the constant in the Bramble–Hilbert lemma on the chunkiness parameter is discussed in [4]. The claimed result then follows from standard operator interpolation theory [1, 4]. □

A further important tool in the analysis of the proposed scheme is the following discrete Poincaré–Friedrichs inequality.

Lemma 19 (discrete Poincaré–Friedrichs inequality)**.** *Any $v_h \in W_h^\ell$ satisfies*

$$\|v_h\|_{L^2(\Omega)} \lesssim \|Dv_h\|_{L^2(\Omega)} + \sqrt{\sum_{F \in \mathcal{F}(\partial\Omega)} h_F \|v_h\|_{L^2(T_F)}^2}$$

where T_F denotes the triangle adjacent to the boundary edge F. If the mesh is sufficiently fine, any $v_h \in W_h^\ell$ satisfies

$$\|v_h\|_{L^2(\Omega)} \lesssim \|Dv_h\|_{L^2(\Omega)}. \tag{19}$$

Proof. Lemma 4 shows

$$\begin{aligned}\|v_h\|_{L^2(\Omega)}^2 &\lesssim \|Dv_h\|_{L^2(\Omega)}^2 + \int_0^L \dot{v}_{h,\ell} v_{h,\perp}\, d\varphi \\ &= \|Dv_h\|_{L^2(\Omega)}^2 + \sum_{F \in \mathcal{F}(\partial\Omega)} \int_F \partial_s(v_h \cdot \ell) v_h \cdot \ell^\perp\, ds.\end{aligned}$$

Since the value of $v_h \cdot \ell$ coincides at all the boundary vertices, on any edge F one can subtract a constant and, after having used the Cauchy inequality, employ the Poincaré inequality for both resulting terms to deduce

$$\begin{aligned}\int_F \partial_s(v_h\cdot\ell)v_h\cdot\ell^\perp\,ds &= \int_F \partial_s(v_h\cdot\ell)\Big(v_h\cdot\ell^\perp - \fint_F v_h\cdot\ell^\perp\,ds\Big)\,ds\\ &\lesssim h_F^2\|\partial_s^2(v_h\cdot\ell)\|_{L^2(F)}\|\partial_s(v_h\cdot\ell^\perp)\|_{L^2(F)}.\end{aligned}$$

The product rule reveals

$$\begin{aligned}\|\partial_s^2(v_h\cdot\ell)\|_{L^2(F)} &\le \|(\partial_s^2 v_h)\cdot\ell\|_{L^2(F)} + 2\|\partial_s v_h\cdot\partial_s\ell\|_{L^2(F)} + \|v_h\cdot\partial_s^2\ell\|_{L^2(F)}\\ &\lesssim \|Dv_h\|_{L^2(F)} + \|v_h\|_{L^2(F)}\end{aligned}$$

where it has been used that $D^2 v_h = 0$. The involved constants depend on $\|\dot\ell\|_{L^\infty([0,T])}$ and $\|\ddot\ell\|_{L^\infty([0,T])}$ as well as on the maximal curvature $\|\ddot{\boldsymbol{x}}\|_{L^\infty([0,T])}$ of the boundary. Similarly,

$$\|\partial_s(v_h\cdot\ell^\perp)\|_{L^2(F)} \lesssim \|Dv_h\|_{L^2(F)} + \|v_h\|_{L^2(F)}.$$

Young's inequality therefore shows that

$$\int_F \partial_s(v_h\cdot\ell)v_h\cdot\ell^\perp\,ds \lesssim h_F^2\left(\|v_h\|_{L^2(F)}^2 + \|Dv_h\|_{L^2(F)}^2\right).$$

The trace inequality (note that Dv_h is piecewise constant) for the triangle T_F adjacent to F shows that this is bounded by

$$h_F\|v_h\|_{L^2(T)}^2 + (h_F^3 + h_F)\|Dv_h\|_{L^2(T_F)}^2 \lesssim h_T\big(\|v_h\|_{L^2(T_F)}^2 + \|Dv_h\|_{L^2(T_F)}^2\big).$$

Thus, the combination of the above estimates results in

$$\|v_h\|_{L^2(\Omega)}^2 \lesssim \|Dv_h\|_{L^2(\Omega)}^2 + \sum_{F\in\mathcal{F}(\partial\Omega)} h_F\|v_h\|_{L^2(T_F)}^2.$$

This concludes the proof. □

Remark 20 (on the constant in Lemma 19)**.** Given any triangulation $\mathcal{T}$ with three boundary vertices z_1, z_2, z_3 such that $\ell(z_j)$ for $j = 1, 2, 3$ are pairwise distinct, a compactness argument shows that an estimate of the form (19) is satisfied with a $\mathcal{T}$-dependent constant. In particular, if the considered class of triangulations consists of quasi-uniform refinements of a given initial mesh, only finitely many of those mesh-dependent constants arise until the meshes reach a certain fineness. Therefore, estimate (19) holds without mesh-size restrictions on families of quasi-uniform refinements for which ℓ evaluated at the boundary vertices points in three distinct directions.

4.2. **Numerical scheme and error estimates.** Recall the operator τ from (11) and define the stabilization parameter

$$\sigma^\tau(\gamma,\varepsilon) := \begin{cases} 1/\sqrt{2} & \text{if } \tau = \tau^{\mathrm{NS}},\\ \frac{1}{\sqrt{2}\|\gamma\|_{L^\infty(\Omega)}}\sqrt{4-3\varepsilon-2\sqrt{1-\varepsilon}} & \text{if } \tau = \tau^{\mathrm{LS}}.\end{cases} \tag{20}$$

Define the following stabilized bilinear form $\tilde a_\tau$ on $(W^\ell + W_h^\ell)\times(W^\ell + W_h^\ell)$,

$$\tilde a_\tau(v,z) := (A : Dv, \tau(z))_{L^2(\Omega)} + \sigma^\tau(\gamma,\varepsilon)^2\left((\operatorname{rot} v, \operatorname{rot} z)_{L^2(\Omega)} + 2\int_0^L \dot v_\ell z_\perp\,d\varphi\right)$$

for any $v, z \in W^\ell + W_h^\ell$. Note that this is indeed well defined because elements $v \in W^\ell$ satisfy $\dot v_\ell = 0$ on the boundary.

Let $\tilde V_h \subseteq \tilde H^1(\Omega;\mathbb{R}^2)$ be a closed subspace. The numerical scheme seeks $u_h \in \tilde V_h$ and $w_h \in W_h^\ell$ such that

$$(\nabla u_h, \nabla \eta_h)_{L^2(\Omega)} = (w_h, \nabla \eta_h)_{L^2(\Omega)} \qquad \text{for all } \eta_h \in \tilde V_h, \tag{21a}$$

$$\tilde a_\tau(w_h, v_h) = (f, \tau(v_h))_{L^2(\Omega)} \qquad \text{for all } v_h \in W_h^\ell. \tag{21b}$$

Proposition 21 (well-posedness of the discrete problem)**.** *Let the mesh $\mathcal{T}$ be sufficiently fine such that* (19) *is satisfied. Then system* (21) *admits a unique solution* $(u_h, w_h) \in \tilde V_h \times W_h^\ell$.

Proof. It is enough to show that the form $\tilde a_\tau$ is coercive over $W_h^\ell \times W_h^\ell$. This is shown with the help of the Cordes condition (3) with arguments already utilized in [18, 10]. For $\tau = \tau^{\mathrm{NS}}$, the triangle and Cauchy inequalities together with (15) and Corollary 3 show

$$\begin{aligned}
(A : Dv, \tau(v))_{L^2(\Omega)} &= (A : Dv, \gamma \operatorname{div} v)_{L^2(\Omega)} \\
&= \|\operatorname{div} v\|_{L^2(\Omega)}^2 + ((\gamma A - I_{2\times 2}) : Dv, \operatorname{div} v)_{L^2(\Omega)} \\
&\geq \|\operatorname{div} v\|_{L^2(\Omega)}^2 - \sqrt{1-\varepsilon}\|\operatorname{div} v\|_{L^2(\Omega)}\|Dv\|_{L^2(\Omega)}. \\
&\geq \|\operatorname{div} v\|_{L^2(\Omega)}^2 \\
&\qquad - \sqrt{1-\varepsilon}\|\operatorname{div} v\|_{L^2(\Omega)}\sqrt{\|\operatorname{div} v\|_{L^2(\Omega)}^2 + \|\operatorname{rot} v\|_{L^2(\Omega)}^2 + 2\int_0^L \dot v_\ell v_\perp \, d\varphi}.
\end{aligned}$$

Note that, thanks to Corollary 3, the argument of the square root in the foregoing expression is nonnegative. The Young inequality bounds the right-hand side from below by

$$\big(1/2 - (1-\varepsilon)/2\big)\|\operatorname{div} v\|_{L^2(\Omega)}^2 - (1-\varepsilon)/2\,(\|\operatorname{rot} v\|_{L^2(\Omega)}^2 + 2\int_0^L \dot v_\ell v_\perp \, d\varphi).$$

After adding $\sigma^\tau(\gamma,\varepsilon)^2(\|\operatorname{rot} v\|_{L^2(\Omega)}^2 + 2\int_0^L \dot v_\ell v_\perp \, d\varphi)$, the coercivity

$$2^{-1}\varepsilon\|Dv\|_{L^2(\Omega)}^2 \leq \tilde a_\tau(v,v)$$

follows. The coercivity for $\tau = \tau^{\mathrm{LS}}$ is shown in a similar fashion following [11]. □

The following result states an a priori error estimate. Due to the nonconformity of the approximation, it does not directly follow from Céa's lemma. Instead, the proof employs techniques which are closely related to the classical Berger–Scott–Strang lemma [3, 4]. The error estimates are, however, independent of any regularity assumptions on the solution.

Theorem 22 (a priori error estimate)**.** *Let the mesh-size $h_{\mathcal{T}}$ of $\mathcal{T}$ be sufficiently small such that* (19) *is satisfied. The exact solution $(u,w) \in \tilde H^1(\Omega;\mathbb{R}^2) \times W^\ell$ to* (16) *and the discrete solution $(u_h, w_h) \in \tilde V_h \times W_h^\ell$ to* (21) *satisfy*

$$\|D(w - w_h)\|_{L^2(\Omega)} \lesssim \inf_{v_h \in W_h^\ell} \|D(w - v_h)\|_{L^2(\Omega)}$$

as well as

$$\|\nabla(u - u_h)\|_{L^2(\Omega)} \lesssim \inf_{z_h \in \tilde V_h} \|\nabla(u - z_h)\|_{L^2(\Omega)} + \inf_{v_h \in W_h^\ell} \|D(w - v_h)\|_{L^2(\Omega)}.$$

Proof. Let $\phi_h := Jw$. The proof departs from the split

$$\|D(w-w_h)\|_{L^2(\Omega)} \le \|D(w-\phi_h)\|_{L^2(\Omega)} + \|D(w_h-\phi_h)\|_{L^2(\Omega)}.$$

Let $e_h := w_h - \phi_h$. Due to the coercivity from the proof of Proposition 21, the second term on the right-hand side satisfies

$$\|D(w_h-\phi_h)\|^2_{L^2(\Omega)} \lesssim \tilde{a}_\tau(w_h-\phi_h, e_h).$$

The discrete solution property of w_h implies that $a_\tau(w_h, e_h) = a_\tau(w, e_h)$ because w solves $A : Dw = f$ and $\operatorname{rot} w = 0$ in Ω as well as $\dot{w}_\ell = 0$ on $\partial\Omega$. Thus, the definition of a_τ and the Cauchy inequality lead to

$$\|D(w_h-\phi_h)\|^2_{L^2(\Omega)} \lesssim \|D(w-\phi_h)\|_{L^2(\Omega)}\|De_h\|_{L^2(\Omega)} + \int_0^L \partial_s(w-\phi_h)_\ell (e_h)_\perp \, d\varphi.$$

Note that the derivative $\partial_s w_\ell$ is well-defined (and actually equals zero). Integration by parts on $[0, L]$ and the Cauchy inequality (on any edge $F \subseteq \partial\Omega$) lead to

$$\int_0^L \partial_s(w-\phi_h)_\ell (e_h)_\perp \, d\varphi \lesssim \sum_{F\in\mathcal{F}(\partial\Omega)} \|w-\phi_h\|_{L^2(F)}\|\partial_s e_{h,\perp}\|_{L^2(F)}.$$

For any edge $F \subseteq \partial\Omega$, the product rule reveals

$$\|\partial_s e_{h,\perp}\|_{L^2(F)} \lesssim \|De_h\|_{L^2(F)} + \|e_h\|_{L^2(F)}.$$

Recall that J is a projection and, thus, $w-\phi_h = (w-\phi_h) - J(w-\phi_h)$. Thus, the trace inequality, the properties from Proposition 17, and Young's inequality lead to

$$\begin{aligned}&\|w-\phi_h\|_{L^2(F)}\|\partial_s e_{h,\perp}\|_{L^2(F)}\\ &\quad\lesssim \|D(w-\phi_h)\|_{L^2(\omega_T)}((1+h_F)\|De_h\|_{L^2(T)} + \|e_h\|_{L^2(T)})\end{aligned}$$

for the triangle T adjacent to F. Note that the constant of the trance inequality (i.e. boundedness of the trace operator) may depend on the shape of the curved triangle T. Altogether, the Cauchy inequality proves

$$\int_0^L \partial_s(w-\phi_h)_\ell (e_h)_\perp \, d\varphi \lesssim \|D(w-\phi_h)\|_{L^2(\Omega)}(\|De_h\|_{L^2(\Omega)} + \|e_h\|_{L^2(\Omega)}).$$

It is the discrete Poincaré–Friedrichs inequality from Lemma 19 that allows to bound $\|e_h\|_{L^2(\Omega)} \lesssim \|De_h\|_{L^2(\Omega)}$ and, thus, to conclude

$$\|D(w_h-\phi_h)\|^2_{L^2(\Omega)} \lesssim \|D(w-\phi_h)\|_{L^2(\Omega)}\|De_h\|_{L^2(\Omega)}.$$

This proves $\|D(w-w_h)\|_{L^2(\Omega)} \lesssim \|D(w-\phi_h)\|_{L^2(\Omega)}$. The quasi-optimality of the quasi-interpolation from Proposition 17 concludes the proof of the first claimed estimate.

The second stated error estimate follows from standard error estimates for Neumann problems. The triangle inequality shows that the quasi-optimal Galerkin error is perturbed by a term $\|\nabla e_u\|_{L^2(\Omega)}$ for $e_u := u_h - u_h^w$, where u_h^w solves (16a) with the exact data w on the right-hand side. The solution properties show that

$$\|\nabla e_u\|^2_{L^2(\Omega)} = (w_h - w, \nabla e_u)_{L^2(\Omega)} \le \|w_h - w\|_{L^2(\Omega)}\|\nabla e_u\|_{L^2(\Omega)}.$$

The use of the projective quasi-interpolation J and the triangle inequality imply

$$\|w_h - w\|_{L^2(\Omega)} \le \|w - Jw\|_{L^2(\Omega)} + \|J(w-w_h)\|_{L^2(\Omega)}.$$

The proof is concluded by the discrete Poincaré–Friedrichs inequality of Lemma 19 and the properties of J from Proposition 17. □

Remark 23 (regularity assumptions)**.** The approximation properties from Corollary 18 allow to predict convergence rates in terms of the maximum mesh-size and the Sobolev regularity of w. Theorem 22 is valid without strong assumptions on the regularity of w. In contrast, it is easy to see that methods based on polygonal approximations of the domain also lead to linear convergence, but *qualitatively* require the assumption that $w \in H^2(\Omega;\mathbb{R}^2)$, which is unrealistic for the discontinuous coefficients under consideration. It is unclear whether a result as Theorem 22 may be verified for polygonal approximations without further restrictions. For a similar discussion, see [3, Remark to Thm. 1.5] where the author warns the reader that in general one should be aware whether smoothness assumptions enter *quantitatively* or *qualitatively* in a convergence proof.

Theorem 24 (a posteriori error estimate)**.** *Let the mesh-size $h_{\mathcal{T}}$ of $\mathcal{T}$ be sufficiently small such that* (19) *is satisfied. The error $e := w - w_h$ between the exact solution $w \in W^\ell$ to* (16b) *and the discrete solution $w_h \in W_h^\ell$ to* (21b) *satisfies the reliable a posteriori error estimate*

$$\|De\|_{L^2(\Omega)} \lesssim \|A : Dw_h - f\|_{L^2(\Omega)} + \|\operatorname{rot} w_h\|_{L^2(\Omega)} + \sqrt{\sum_{F\in\mathcal{F}(\partial\Omega)} h_F \|\partial_s(w_h\cdot\ell)\|^2_{L^2(\partial\Omega)}}$$

as well as, for any $T \in \mathcal{T}$ and any $F \in \mathcal{F}(\partial\Omega)$, the local efficiency estimates

$$\|A : Dw_h - f\|_{L^2(T)} + \|\operatorname{rot} w_h\|_{L^2(T)} \lesssim \|De\|_{L^2(T)}$$
$$h_F\|\partial_s(w_h\cdot\ell)\|^2_{L^2(F)} \lesssim \|De\|_{L^2(\omega_{K_F})} + \|e\|_{L^2(\omega_{K_F})}$$

for the triangle K_F adjacent to F.

Proof. Let $w_h \in W_h^\ell$ solve (21b). The coercivity of $\tilde{a}_\tau$ together with $A : Dw = f$, $\operatorname{rot} w = 0$ and $\dot{w}_\ell = 0$ on $\partial\Omega$ lead to

$$\|De\|^2_{L^2(\Omega)} \lesssim (A : Dw_h - f, \tau(e))_{L^2(\Omega)} + \|\operatorname{rot} w_h\|^2_{L^2(\Omega)} + \int_0^L \partial_s(w_h\cdot\ell)e_\perp \, d\varphi. \tag{22}$$

Since $w_h\cdot\ell$ takes the same value at every boundary vertex, the last integral can be rewritten for arbitrary constants $(c_F)_{F\in\mathcal{F}(\partial\Omega)} \in \mathbb{R}^{\operatorname{card}(\mathcal{F}(\partial\Omega))}$ as

$$\int_0^L \partial_s(w_h\cdot\ell)e_\perp \, d\varphi = \sum_{F\in\mathcal{F}(\partial\Omega)} \int_F \partial_s(w_h\cdot\ell)(e_\perp - c_F)\, ds.$$

Recall the (projective) quasi-interpolation operator J, which satisfies $Je = Jw - w_h$. For every boundary edge $F \in \mathcal{F}(\partial\Omega)$, the Cauchy and triangle inequalities reveal

$$\begin{aligned}\int_F \partial_s(w_h\cdot\ell)(e_\perp - c_F)\, ds &\le \|\partial_s(w_h\cdot\ell)\|_{L^2(F)}\|e_\perp - c_F\|_{L^2(F)} \\ &\le \|\partial_s(w_h\cdot\ell)\|_{L^2(F)}\big(\|w - Jw\|_{L^2(F)} + \|(Je)_\perp - c_F\|_{L^2(F)}\big).\end{aligned}$$

The trace inequality and the approximation and stability properties from Proposition 17 show for the triangle T_F adjacent to F that

$$\|w - Jw\|_{L^2(F)} \lesssim h_F^{1/2}\|D(w - Jw)\|_{L^2(\omega_{T_F})}.$$

The choice $c_F = \fint_F (Je)_\perp \, ds$, the Poincaré inequality, and the product rule show that

$$\|(Je)_\perp - c_F\|_{L^2(F)} \lesssim h_F \|\partial_s (Je)_\perp\|_{L^2(F)} \lesssim h_F \big(\|Je\|_{L^2(F)} + \|DJe\|_{L^2(F)} \big).$$

The trace inequality therefore proves (recall that DJe is piecewise constant)

$$\|(Je)_\perp - c_F\|_{L^2(F)} \lesssim h_F^{1/2} \big(\|Je\|_{L^2(T_F)} + (1+h_F) \|DJe\|_{L^2(T_F)} \big).$$

The combination of the foregoing five displayed estimates results in

$$\begin{aligned} \int_0^L \partial_s (w_h \cdot \ell) e_\perp \, d\varphi \lesssim \sum_{F \in \mathcal{F}(\partial\Omega)} h_F^{1/2} \|\partial_s (w_h \cdot \ell)\|_{L^2(F)} \big(& \|D(w - Jw)\|_{L^2(\omega_{T_F})} \\ & + \|Je\|_{L^2(T_F)} + \|DJe\|_{L^2(T_F)} \big). \end{aligned}$$

With the Cauchy inequality in $\mathbb{R}^{\operatorname{card} \mathcal{F}(\partial\Omega)}$, the finite overlap of patches, and the discrete Poincaré–Friedrichs inequality from Lemma 19 one concludes

$$\begin{aligned} & \int_0^L \partial_s (w_h \cdot \ell) e_\perp \, d\varphi \\ & \lesssim \sqrt{\sum_{F \in \mathcal{F}(\partial\Omega)} h_F \|\partial_s (w_h \cdot \ell)\|_{L^2(F)}^2} \big(\|D(w - Jw)\|_{L^2(\Omega)} + \|Je\|_{L^2(\Omega)} + \|DJe\|_{L^2(\Omega)} \\ & \lesssim \sqrt{\sum_{F \in \mathcal{F}(\partial\Omega)} h_F \|\partial_s (w_h \cdot \ell)\|_{L^2(F)}^2} \big(\|D(w - Jw)\|_{L^2(\Omega)} + \|DJe\|_{L^2(\Omega)} \big). \end{aligned}$$

The stability and quasi-optimality of J from Proposition 17 show that $\|D(w - Jw)\|_{L^2(\Omega)} + \|DJe\|_{L^2(\Omega)} \lesssim \|De\|_{L^2(\Omega)}$, whence

$$\int_0^L \partial_s (w_h \cdot \ell) e_\perp \, d\varphi \lesssim \sqrt{\sum_{F \in \mathcal{F}(\partial\Omega)} h_F \|\partial_s (w_h \cdot \ell)\|_{L^2(F)}^2} \|De\|_{L^2(\Omega)}.$$

The combination with (22) leads to the claimed reliability estimate because

$$\|\operatorname{rot} w_h\|_{L^2(\Omega)}^2 \leq \|\operatorname{rot} w_h\|_{L^2(\Omega)} \|De\|_{L^2(\Omega)}$$

and the term $\|De\|_{L^2(\Omega)}$ can be absorbed.

Because of the L^2 identities $A : Dw = f$ and $\operatorname{rot} w = 0$, the efficiency needs only to be shown for the boundary terms. Let $F \in \mathcal{F}(\partial\Omega)$ be a boundary edge and recall that $\partial_s (w \cdot \ell) = 0$. Thus, the inverse inequality implies

$$\|\partial_s ((Jw - w) \cdot \ell)\|_{L^2(F)} \lesssim h_F^{-1} \|(Jw - w) \cdot \ell\|_{L^2(F)} \leq h_F^{-1} \|Jw - w\|_{L^2(F)}.$$

It is a truly discrete argument because $w \cdot \ell$ equals a constant. The reader should nevertheless be aware that the constant in the inverse inequality depends on the local data configuration, in particular on the local oscillations of the boundary and of ℓ. The triangle inequality, the aforementioned inverse estimate, and the product rule show

$$\begin{aligned} h_F \|\partial_s (w_h \cdot \ell)\|_{L^2(F)}^2 & \lesssim h_F \|\partial_s ((Jw - w) \cdot \ell)\|_{L^2(F)}^2 + h_F \|\partial_s (Je \cdot \ell)\|_{L^2(F)}^2 \\ & \lesssim h_F^{-1} \|Jw - w\|_{L^2(F)}^2 + h_F (\|DJe\|_{L^2(F)}^2 + \|Je\|_{L^2(F)}^2). \end{aligned}$$

The trace inequality (with adjacent triangle K) leads to

$$\begin{aligned} &h_F\|\partial_s(w_h\cdot\ell)\|^2_{L^2(F)}\\ &\quad\lesssim h_F^{-2}\|Jw-w\|^2_{L^2(K)}+\|D(Jw-w)\|^2_{L^2(K)}+\|DJe\|^2_{L^2(K)}+\|Je\|^2_{L^2(K)}\\ &\quad= h_F^{-2}\|Je-e\|^2_{L^2(K)}+\|D(Je-e)\|^2_{L^2(K)}+\|DJe\|^2_{L^2(K)}+\|Je\|^2_{L^2(K)} \end{aligned}$$

where the projection property of J has been used. The approximation and stability properties of J from Proposition 17 therefore imply

$$h_F\|\partial_s(w_h\cdot\ell)\|^2_{L^2(F)}\lesssim\|De\|^2_{L^2(\omega_K)}+\|e\|^2_{L^2(\omega_K)}.$$

This concludes the efficiency proof. □

Remark 25 (convergence of adaptive methods)**.** A convergence proof of an adaptive algorithm as in [11] is also possible for the oblique derivative problem. The details are very similar to the proof in [11] and, thus, omitted.

Remark 26 (more general fourth-order problems)**.** If one wishes to consider more general fourth-order problems with oblique derivative boundary condition whose right-hand side has a structure different than in (14), a reformulation with (16) is not possible. In this case it is necessary to invoke the Stokes-like saddle-point problem (14) and the stability from Lemma 8. The methods from this work still apply to this case. Instead of the classical MINI element, a curvilinear version can be employed. This is briefly outlined in the following (the generalization to many other Stokes elements is straightforward). Let, for any $T\in\mathcal{T}$ the affine nodal basis functions be denoted by $\lambda_1,\lambda_2,\lambda_3$. For the vertices z_1,z_2,z_3 of T they satisfy $\lambda_j(z_k)=\delta_{jk}$. Let b_T denote a multiple of the generalized bubble function $(\lambda_1\lambda_2\lambda_3)^+$ (the + denotes the nonnegative part) such that $\fint_T b_T\,dx=1$. The space spanned by these functions is denoted by $\mathcal{B}_3(\mathcal{T})$. The set of boundary vertices is denoted by $\mathcal{N}(\partial\Omega)$. The discrete spaces read

$$W_h^{mini}:=W_h\oplus\mathcal{B}_3(\mathcal{T})^2\quad\text{and}\quad Q_h:=S^1(\mathcal{T}).$$

Not only the approximation properties of W_h but also the stability of the pairing (W_h,Q_h) can are quantified with the help of the quasi-interpolation operator J. An important point is that the shape regularity shows that the crucial scaling properties of the bubble function are independent of the curvature of the elements. The stability constant is therefore determined by the stability constant of J. The stability and error analysis follow from the usual saddle-point theory [2].

5. Numerical results

This section presents numerical computations in planar domains for the choice $\tau=\tau^{\mathrm{LS}}$. The space $\tilde{V}_h$ in (21a) is chosen as

$$\tilde{V}_h=S^2(\mathcal{T})\cap L^2_0(\Omega)=S^2(\mathcal{T})\cap\tilde{H}^1(\Omega)=P_2(\mathcal{T})\cap\tilde{H}^1(\Omega),$$

i.e. the space of globally continuous and piecewise quadratic functions over $\mathcal{T}$ with vanishing global average. The adaptive mesh-refining algorithm is based on the local error estimator contributions described in Theorem 24 and Dörfler marking [7, 22] with bulk parameter $\theta=0.3$.

The unit vector field ℓ is the rotated normal field

$$\ell=\sqrt{1/2}\begin{pmatrix}1&-1\\1&1\end{pmatrix}\nu$$

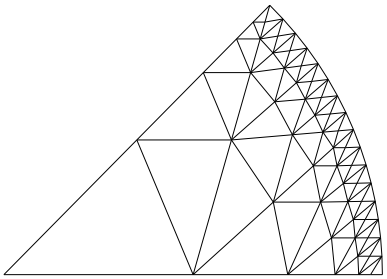


FIGURE 4. Sub-triangulation of a triangle for evaluation of composite Gauss-quadrature over the curved elements.

ndof	$\|u-u_h\|_{L^2_0(\Omega)}$	$\|\nabla u - w_h\|_{L^2(\Omega)}$	$\|D^2u - Dw_h\|_{L^2(\Omega)}$	η
10	3.5575-16	3.8062e-16	9.0124e-16	2.5381e-15
26	3.2651-16	4.0055e-16	1.3825e-15	3.3342e-15
82	3.6040-16	5.5379e-16	1.7303e-15	3.8420e-15
290	1.0194-15	1.2116e-15	4.5058e-15	9.0520e-15

TABLE 1. Discretization errors for the first experiment on the circle with constant coefficient.

so that the rotation angle is $\vartheta = \pi/4$.

Remark 27 (quadrature on curved elements)**.** In the numerical implementation, integrals over curved elements are approximately evaluated with a composite Gaussian quadrature over a locally refined sub-mesh of a polygonal approximation to the triangle as displayed in Figure 4.

5.1. **Experiment 1.** Let $\Omega := \{|x| < 1\}$ be the unit disk, let f be the constant function $f = 4$ and and let A be the constant coefficient $A := \left(\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right)$. Since A is a multiple of the unit matrix, the resulting problem is the diffusion equation $\Delta u = 2$ whose exact solution is given by

$$u(x) = -\frac{1}{2}(1 - x_1^2 - x_2^2) + \frac{1}{4}. \tag{23}$$

Then $u \in \tilde{H}^1(\Omega)$ and $\nabla u \in W^\ell$ because $\nabla u \cdot t$ as well as $\nabla u \cdot \nu$ are constant along $\partial\Omega$. Moreover, since u is quadratic and ∇u is affine, the exact solution belongs to the discrete space. Indeed, the errors as well as the error estimator displayed in Table 1 are in the range of machine precision. Although the method is not a Galerkin scheme, this observation is theoretically supported by Theorem 22 and Remark 20.

5.2. **Experiment 2.** As in the first experiment, the domain is the unit disk and $f = 4$. The coefficient A is given by

$$A = \tilde{A} \circ \varphi$$

for

$$\tilde{A}(x) = \begin{bmatrix} 2 & \frac{x_1 x_2}{|x_1|\,|x_2|} \\ \frac{x_1 x_2}{|x_1|\,|x_2|} & 2 \end{bmatrix} \quad \text{and} \quad \varphi(x) = \begin{bmatrix} x_1 + \frac{1}{3} \\ x_2 - \frac{1}{3} + (x_1 + \frac{1}{3})^{1/3} \end{bmatrix}.$$

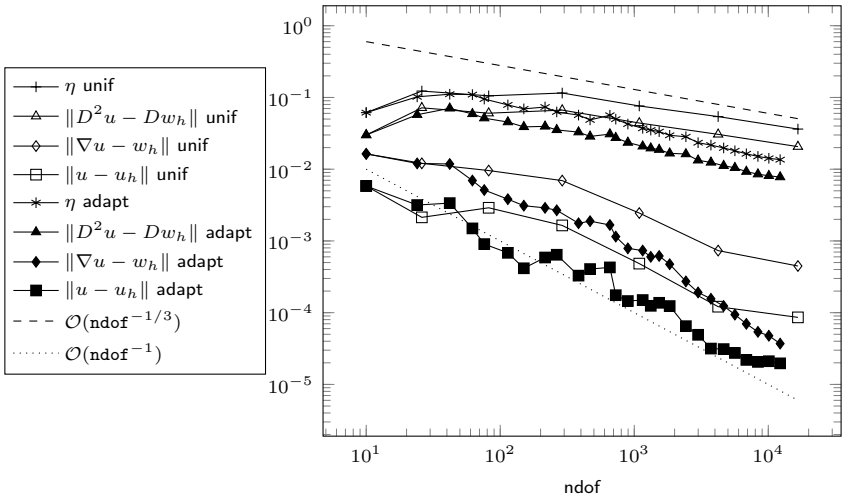


FIGURE 5. Convergence history for Experiment 2.

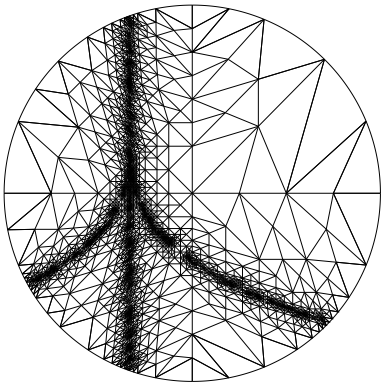


FIGURE 6. Adaptive mesh for Experiment 2; 2 818 vertices, 5 636 degrees of freedom, level 23.

Again, the exact solution is given by the polynomial in (23). The coefficient is, however, not resolved by the quadrature rules on the given sequence of meshes. Thus, the computed solution is different from u. The convergence history is displayed in Figure 5. The convergence rates on uniformly refined meshes are sub-optimal. Adaptive mesh refinement leads to improved convergence rates. A mesh from the sequences of adaptive meshes is displayed in Figure 6. A strong refinement along the discontinuity of the coefficient is observable.

5.3. **Experiment 3.** Let $\Omega := \{\frac{1}{4}|x_1|^2 + |x_2|^2 < 1\}$ be an ellipse with semi-axes of length 2 and 1 and let $f = \operatorname{sign}(x_1)\operatorname{sign}(x_2)$. The coefficient A is the same as in Experiment 2. For this example, the exact solution is unknown. Therefore, only the error estimator is plotted in Figure 7. The convergence rate on uniform meshes is observed to be $\mathcal{O}(\mathtt{ndof}^{-1/4})$. The convergence rate can be improved through adaptive mesh refinement, which shows the optimal convergence rate of 1/2. An

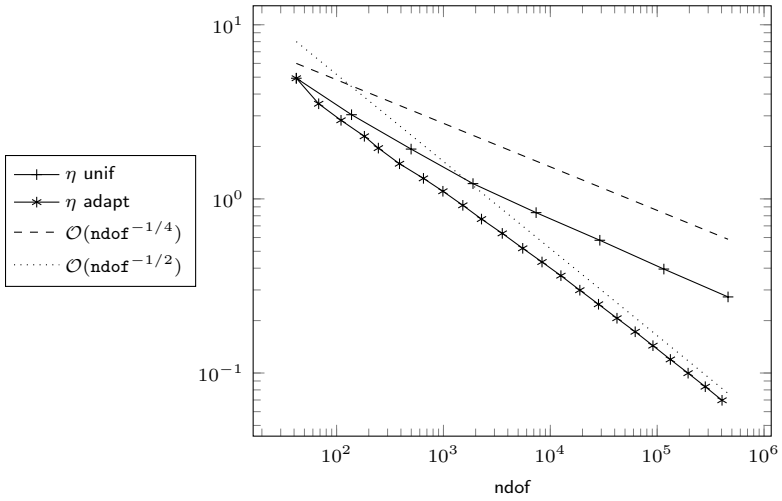


FIGURE 7. Convergence history for Experiment 3.

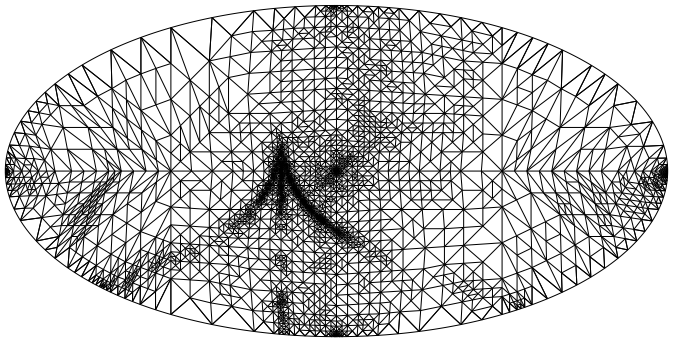


FIGURE 8. Adaptive mesh for Experiment 3; 2 759 vertices, 5 518 degrees of freedom, level 12.

adaptive mesh is shown in Figure 8. It shows strong refinement along the discontinuity of the coefficient A as well as at certain points where f is discontinuous: the origin $(0,0)$ and points near the boundary where f is discontinuous. Figure 9 displays the computed solution.

APPENDIX A. PROOF OF LEMMA 2

Proof of Lemma 2. The proof closely follows the presentation in [13]. The field H^1-regular vector field w is assumed to be piecewise smooth and all higher-order derivatives appearing in this proof are understood to be applied piecewise in the domain. It is straightforward to verify the following identity

$$2(\partial_1 w_1 \partial_2 w_2 - \partial_2 w_1 \partial_1 w_2) = \operatorname{div} \begin{bmatrix} w_1 \partial_2 w_2 - w_2 \partial_2 w_1 \\ -(w_1 \partial_1 w_2 - w_2 \partial_1 w_1) \end{bmatrix}.$$

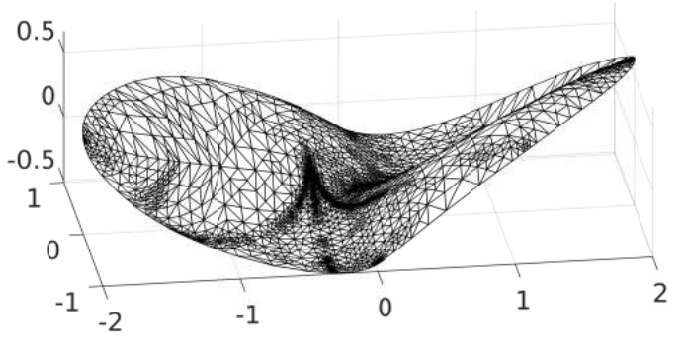


FIGURE 9. Discrete solution for Experiment 3; 2 759 vertices, 5 518 degrees of freedom, level 12.

Thus, the divergence theorem gives

$$2\int_\Omega (\partial_1 w_1 \partial_2 w_2 - \partial_2 w_1 \partial_1 w_2)\,dx = \int_0^L \begin{bmatrix} w_1\partial_2 w_2 - w_2 \partial_2 w_1 \\ -(w_1 \partial_1 w_2 - w_2 \partial_1 w_1)\end{bmatrix} \cdot \nu \, d\varphi. \tag{24}$$

Note that no interior jumps arise because $(\partial_2 w_k, -\partial_1 w_k)\cdot \nu$ for $k \in \{1,2\}$ and the normal ν of an interior interface coincides with the tangential derivative and, thus, is continuous. The definitions of w_ℓ and $w_\perp$ result, for any $\varphi \in [0, L]$, in the system

$$\begin{aligned} w_1 \ell_1 + w_2 \ell_2 &= w_\ell \\ -w_1 \ell_2 + w_2 \ell_1 &= w_\perp \end{aligned} \tag{25}$$

where the dependence on φ and $\boldsymbol{x}(\varphi)$ has been suppressed in the notation. System (25) has the unique solution

$$w_1 = -w_\perp \ell_2 + w_\ell \ell_1, \qquad w_2 = w_\perp \ell_1 + w_\ell \ell_2. \tag{26}$$

Substituting w_1 and w_2 in (24) by these values results in

$$\begin{aligned} &2\int_\Omega (\partial_1 w_1 \partial_2 w_2 - \partial_2 w_1 \partial_1 w_2)\,dx \\ &\quad = \int_0^L w_\perp \big(- \ell_2 \partial_2 w_2 \nu_1 - \ell_1 \partial_2 w_1 \nu_1 + \ell_2 \partial_1 w_2 \nu_2 + \ell_1 \partial_1 w_1 \nu_2\big)\, d\varphi \\ &\qquad + \int_0^L w_\ell \big(\ell_1 \partial_2 w_2 \nu_1 - \ell_2 \partial_2 w_1 \nu_1 - \ell_1 \partial_1 w_2 \nu_2 + \ell_2 \partial_1 w_1 \nu_2\big)\, d\varphi. \end{aligned} \tag{27}$$

Differentiation of (25) with the chain rule and (4) together with elementary rearrangements leads to

$$\begin{aligned} -\ell_2 \partial_2 w_2 \nu_1 - \ell_1 \partial_2 w_1 \nu_1 + \ell_2 \partial_1 w_2 \nu_2 + \ell_1 \partial_1 w_1 \nu_2 &= -\dot w_\ell + \dot \ell_1 w_1 + \dot \ell_2 w_2 \\ \ell_1 \partial_2 w_2 \nu_1 - \ell_2 \partial_2 w_1 \nu_1 - \ell_1 \partial_1 w_2 \nu_2 + \ell_2 \partial_1 w_1 \nu_2 &= \dot w_\perp + \dot \ell_2 w_1 - \dot \ell_1 w_2. \end{aligned}$$

Therefore, the right-hand side of (27) simplifies to

$$\int_0^L w_\perp \big(- \dot w_\ell + \dot \ell_1 w_1 + \dot \ell_2 w_2\big)\, d\varphi + \int_0^L w_\ell \big(\dot w_\perp + \dot \ell_2 w_1 - \dot \ell_1 w_2\big)\, d\varphi.$$

Using identities (26) and $w_\ell^2 + w_\perp^2 = |w|^2$, one eventually obtains

$$\begin{aligned} 2\int_\Omega (\partial_1 w_1 \partial_2 w_2 - \partial_2 w_1 \partial_1 w_2)\,dx \\ = \int_0^L |w|^2(\ell_1 \dot\ell_2 - \dot\ell_1 \ell_2)\,d\varphi + \int_0^L (-w_\perp \dot w_\ell + \dot w_\perp w_\ell)\,d\varphi. \end{aligned}$$

Employing (5) for the first term and integration by parts for the second term on the right-hand side concludes the proof. □

Appendix B. Proof of Lemma 8

Proof of Lemma 8. The function q is split as $q = q_0 + c$ for $c := \fint_\Omega q\,dx$ and $q_0 := q - c$. Then, $\int_\Omega q_0\,dx = 0$ and, hence, by the classical inf-sup condition of the divergence [4] applied to $\left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right] q_0$ there exists $v_0 \in H^1_0(\Omega;\mathbb{R}^2) \subseteq W^\ell$ such that $\operatorname{rot} v_0 = q_0$ and $\|Dv_0\|_{L^2(\Omega)} \le \beta_0^{-1}\|q_0\|_{L^2(\Omega)} \le \beta_0^{-1}\|q\|_{L^2(\Omega)}$ for a positive constant β_0 independent of v_0. Let $\varphi \in W^\ell$ be a function such that

$$\operatorname{meas}(\Omega) = \int_{\partial\Omega} \varphi \cdot t\,ds = \int_\Omega \operatorname{rot}\varphi\,dx.$$

Again the classical inf-sup condition shows that there exists some $\varphi_0 \in H^1_0(\Omega;\mathbb{R}^2) \subseteq W^\ell$ with $\operatorname{rot}\varphi_0 = \operatorname{rot}\varphi - \fint_\Omega \operatorname{rot}\varphi\,dx$ and $\|D\varphi_0\|_{L^2(\Omega)} \le \beta_0^{-1}\|\operatorname{rot}\varphi\|_{L^2(\Omega)}$. Thus, $v_1 := \varphi - \varphi_0 \in W^\ell$ satisfies

$$\begin{aligned} \operatorname{rot} v_1 = \fint_\Omega \operatorname{rot}\varphi\,dx = 1 \\ \text{and}\quad \|Dv_1\|_{L^2(\Omega)} \le \beta_0^{-1}\|\operatorname{rot}\varphi\|_{L^2(\Omega)} + \|D\varphi\|_{L^2(\Omega)} \le (1+\beta_0^{-1})\|D\varphi\|_{L^2(\Omega)}. \end{aligned}$$

Define $v := v_0 + cv_1$. This function satisfies $v \in W^\ell$, $\operatorname{rot} v = q$ and, with $c = \fint_\Omega q\,dx$,

$$\|Dv\|_{L^2(\Omega)} \le \beta_0^{-1}\|q\|_{L^2(\Omega)} + c(1+\beta_0^{-1})\|D\varphi\|_{L^2(\Omega)} \le \beta_1^{-1}\|q\|_{L^2(\Omega)}$$

for the constant $\beta_1 := (\beta_0^{-1} + (1+\beta_0^{-1}\operatorname{meas}(\Omega)^{-1/2})\|D\varphi\|_{L^2(\Omega)})^{-1}$. □

Appendix C. Proof of Proposition 17

Proof of Proposition 17. The projection property of J is immediately verified because Π and the averaging (18) act as the identity on the continuous piecewise affines and W^ℓ_h, respectively. The proof of approximation and stability properties is well known in the case of classical finite elements and its generalization to the present setting is briefly shown here. Let $v \in W^\ell$ and $T \in \mathcal{T}$. The triangle inequality reads

$$\|v - Jv\|_{L^2(T)} \le \|v - \Pi v\|_{L^2(T)} + \|\Pi - Jv\|_{L^2(T)}.$$

The first term is estimated with the Poincaré inequality

$$\|v - \Pi v\|_{L^2(T)} \lesssim h_T \|Dv\|_{L^2(T)}. \tag{28}$$

For any vertex $z \in \mathcal{N}$, $\lambda_z \in S^1(\mathcal{T})$ denotes the piecewise affine basis function that takes the value 1 at z and 0 at all remaining vertices. Denote $c_v := v \cdot \ell|_{\partial\Omega}$.

Standard arguments from the analysis of averaging a posteriori error estimators with the triangle inequality prove

$$\begin{aligned}
&\|\Pi v - Jv\|_{L^2(T)} = \\
&\Big\| \sum_{z\in\mathcal{N}(T)} \sum_{\substack{K\in\mathcal{T}\\ \text{with } z\in K}} \frac{\lambda_z}{\operatorname{card}(\{K\in\mathcal{T}: z\in K\})} \Big((\Pi v)|_T(z) - (\Pi v)|_K(z) \Big) \Big\|_{L^2(T)} \\
&\quad \lesssim h_T \sum_{z\in\mathcal{N}(T)} \sum_{\substack{K\in\mathcal{T}\\ \text{with } z\in K}} \Big| (\Pi v)|_T(z) - (\Pi v)|_K(z) \Big| \qquad (29)\\
&\qquad + h_T \sum_{z\in\mathcal{N}(T)\cap\partial\Omega} \Big| (\Pi v)|_T(z)\cdot \ell(z) - c_v \Big|.
\end{aligned}$$

The first term on the right-hand side of (29) can be controlled with standard equivalence-of-norms arguments by

$$h_T^{1/2} \sum_{z\in\mathcal{N}(T)} \sum_{F\in\mathcal{F}(\Omega,z)} \|[\Pi v]_F\|_{L^2(F)}.$$

Here, $\mathcal{F}(\Omega,z) := \{F\in\mathcal{F} : F \not\subseteq \partial\Omega \text{ and } z\in F\}$ denotes the set of interior (in particular straight) edges containing the vertex z and the square bracket $[\cdot]_F$ denotes the jump across the edge F. In order to bound the second term on the right-hand side of (29), let $z\in\mathcal{N}(T)\cap\partial\Omega$ be a boundary vertex of T and let F denote the curved edge of T. The space

$$\Big\{ \varphi\in L^2(F) : \text{there exist } c\in\mathbb{R} \text{ and } p_1\in P_1(T;\mathbb{R}^2) \text{ such that } \varphi = c + p_1|_F\cdot\ell \Big\}$$

has dimension at most seven. Thus, compactness of the unit sphere and a scaling argument show that there exists a constant $C>0$ (independent of v, but possibly dependent on the oscillations of F and ℓ) such that

$$\Big| (\Pi v)|_T(z)\cdot\ell(z) - c_v \Big| \le C h_T^{-1/2} \|(\Pi v)\cdot\ell - c_v\|_{L^2(F)} \le C h_T^{-1/2} \|v - \Pi v\|_{L^2(F)}.$$

The trace and Poincaré inequalities together with Lemma 15 therefore prove

$$\Big| (\Pi v)|_T(z)\cdot\ell(z) - c_v \Big| \lesssim \|Dv\|_{L^2(T)}.$$

Note that the constant in the trace inequality may depend on the geometry of T while the Poincaré constant is determined by the chunkiness parameter. The combination of the foregoing estimates with (29) results in

$$\|\Pi v - Jv\|_{L^2(T)} \lesssim h_T^{1/2} \sum_{z\in\mathcal{N}(T)} \sum_{F\in\mathcal{F}(\Omega,z)} \|[\Pi v]_F\|_{L^2(F)} + h_T \|Dv\|_{L^2(T)}. \qquad (30)$$

The same calculation with $\|\nabla\lambda_z\|_{L^\infty} \lesssim h_T^{-1}$ proves

$$\|D(\Pi v - Jv)\|_{L^2(T)} \lesssim h_T^{-1/2} \sum_{z\in\mathcal{N}(T)} \sum_{F\in\mathcal{F}(\Omega,z)} \|[\Pi v]_F\|_{L^2(F)} + \|Dv\|_{L^2(T)}. \qquad (31)$$

Let $F\in\mathcal{F}(\Omega,z)$ for some $z\in\mathcal{N}(T)$ be an interior edge. Let K_1, K_2 denote the adjacent triangles with $\omega_F := \operatorname{int}(K_1\cup K_2)$ and let K_1^{in}, K_2^{in} denote the two adjacent maximal inscribed triangles (in the sense of Definition 14). The trace inequality

(with a constant only dependent on the shape regularity of the maximal inscribed triangles) shows

$$\begin{aligned}\|[\Pi v]_F\|_{L^2(F)} &= \|[v-\Pi v]_F\|_{L^2(F)}\\ &\lesssim \sum_{j=1}^{2}\big(h_T^{-1/2}\|v-\Pi v\|_{L^2(K_j^{in})}+h_T^{1/2}\|D(v-\Pi v)\|_{L^2(K_j^{in})}\big).\end{aligned}$$

The Poincaré inequality reveals for any $j\in\{1,2\}$ that

$$\|v-\Pi v\|_{L^2(K_j^{in})}\le\|v-\Pi v\|_{L^2(K_j)}\lesssim h_T^{1/2}\|Dv\|_{L^2(K_j)}$$

The combination of the two foregoing estimates results in

$$\|[\Pi v]_F\|_{L^2(F)}\lesssim h_T^{1/2}\big(\|Dv\|_{L^2(\omega_F)}+\|D\Pi v\|_{L^2(K_1^{in})}+\|D\Pi v\|_{L^2(K_2^{in})}\big). \tag{32}$$

For any K_j^{in} for $j\in\{1,2\}$, Lemma 15 shows

$$\|D\Pi v\|_{L^2(K_j^{in})}\le\|D\Pi v\|_{L^2(K_j)}\lesssim\|Dv\|_{L^2(K_j)}. \tag{33}$$

The combination of (30), (31), (32), (33) shows

$$h_T^{-1}\|\Pi v-Jv\|_{L^2(T)}+\|D(\Pi v-Jv)\|_{L^2(T)}\lesssim\|Dv\|_{L^2(\omega_T)}.$$

Estimate (28) and Lemma 15 imply

$$h_T^{-1}\|v-\Pi v\|_{L^2(T)}+\|D\Pi v\|_{L^2(T)}\lesssim\|Dv\|_{L^2(T)}.$$

The combination of the foregoing two displayed formulas proves the asserted local stability and approximation properties.

The L^2 stability follows with similar arguments. Indeed, as in (29), any $T\in\mathcal{T}$ satisfies

$$\|Jv\|_{L^2(T)}\lesssim h_T\sum_{z\in\mathcal{N}(T)}\sum_{\substack{K\in\mathcal{T}\\ \text{with } z\in K}}|(\Pi v)|_K(z)|\lesssim\|\Pi v\|_{L^2(\omega_T)}\le\|v\|_{L^2(\omega_T)}$$

where the involved constants depend on the maximal inscribed triangles to T and its neighbours.

The stated best-approximation property follows because J is a stable projection. Indeed, for any $\varphi_h\in W_h^\ell$,

$$\|D(w-Jw)\|_{L^2(\Omega)}\le\|D(w-\varphi_h)\|_{L^2(\Omega)}+\|DJ(w-\varphi_h)\|_{L^2(\Omega)}\lesssim\|D(w-\varphi_h)\|_{L^2(\Omega)}.$$

This concludes the proof. □

Acknowledgement

Main parts of this work were written while the author enjoyed the kind hospitality of the Hausdorff Institute for Mathematics in Bonn (February–April 2017).

References

1. Jöran Bergh and Jörgen Löfström, *Interpolation spaces. An introduction*, Grundlehren der Mathematischen Wissenschaften, No. 223, Springer-Verlag, Berlin-New York, 1976.
2. Daniele Boffi, Franco Brezzi, and Michel Fortin, *Mixed finite element methods and applications*, Springer Series in Computational Mathematics, vol. 44, Springer, Heidelberg, 2013.
3. Dietrich Braess, *Finite elements. Theory, fast solvers, and applications in elasticity theory*, third ed., Cambridge University Press, Cambridge, 2007.
4. Susanne C. Brenner and L. Ridgway Scott, *The mathematical theory of finite element methods*, third ed., Texts in Applied Mathematics, vol. 15, Springer, New York, 2008.

5. Ph. Clément, *Approximation by finite element functions using local regularization*, Rev. Française Automat. Informat. Recherche Operationnelle **9** (1975), no. R-2, 77–84.
6. Daniele Antonio Di Pietro and Alexandre Ern, *Mathematical aspects of discontinuous Galerkin methods*, Mathématiques & Applications (Berlin), vol. 69, Springer, Heidelberg, 2012.
7. Willy Dörfler, *A convergent adaptive algorithm for Poisson's equation*, SIAM J. Numer. Anal. **33** (1996), no. 3, 1106–1124.
8. Xiaobing Feng, Lauren Hennings, and Michael Neilan, *Finite element methods for second order linear elliptic partial differential equations in non-divergence form*, Math. Comp. **86** (2017), no. 307, 2025–2051.
9. Xiaobing Feng, Michael Neilan, and Stefan Schnake, *Interior penalty discontinuous Galerkin methods for second order linear non-divergence form elliptic PDEs*, arXiv e-prints **1605.04364** (2016).
10. D. Gallistl, *Stable splitting of polyharmonic operators by generalized Stokes systems*, Math. Comp. **86** (2017), no. 308, 2555–2577.
11. ———, *Variational formulation and numerical analysis of linear elliptic equations in non-divergence form with Cordes coefficients*, SIAM J. Numer. Anal. **55** (2017), no. 2, 737–757.
12. Omar Lakkis and Tristan Pryer, *A finite element method for second order nonvariational elliptic problems*, SIAM J. Sci. Comput. **33** (2011), no. 2, 786–801.
13. Antonino Maugeri, Dian K. Palagachev, and Lubomira G. Softova, *Elliptic and parabolic equations with discontinuous coefficients*, Wiley-VCH Verlag Berlin GmbH, Berlin, 2000.
14. Ricardo H. Nochetto and Wujun Zhang, *Discrete ABP estimate and convergence rates for linear elliptic equations in non-divergence form*, arXiv e-prints **1411.6036** (2014), Preprint.
15. Peter Oswald, *Multilevel finite element approximation*, Teubner Skripten zur Numerik., B. G. Teubner, Stuttgart, 1994.
16. Mikhail V. Safonov, *Nonuniqueness for second-order elliptic equations with measurable coefficients*, SIAM J. Math. Anal. **30** (1999), no. 4, 879–895.
17. Ridgway Scott, *Interpolated boundary conditions in the finite element method*, SIAM J. Numer. Anal. **12** (1975), 404–427.
18. Iain Smears and Endre Süli, *Discontinuous Galerkin finite element approximation of nondivergence form elliptic equations with Cordès coefficients*, SIAM J. Numer. Anal. **51** (2013), no. 4, 2088–2106.
19. ———, *Discontinuous Galerkin finite element approximation of Hamilton-Jacobi-Bellman equations with Cordes coefficients*, SIAM J. Numer. Anal. **52** (2014), no. 2, 993–1016.
20. ———, *Discontinuous Galerkin finite element methods for time-dependent Hamilton-Jacobi-Bellman equations with Cordes coefficients*, Numer. Math. **133** (2016), no. 1, 141–176.
21. Giorgio Talenti, *Problemi di derivata obliqua per equazioni ellittiche in due variabili*, Boll. Un. Mat. Ital. (3) **22** (1967), 505–526.
22. Rüdiger Verfürth, *A posteriori error estimation techniques for finite element methods*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2013.

(D. Gallistl) Institut für Angewandte und Numerische Mathematik, Karlsruher Institut für Technologie, 76128 Karlsruhe, Germany; and Universtität Heidelberg, Im Neuenheimer Feld 205 , 69120 Heidelberg, Germany