Erik Burman and Riccardo Puppi*

Two mixed finite element formulations for the weak imposition of the Neumann boundary conditions for the Darcy flow

https://doi.org/10.1515/jnma-2021-0042
Received April 03, 2021; revised June 08, 2021; accepted June 18, 2021

Abstract: We propose two different discrete formulations for the weak imposition of the Neumann boundary conditions of the Darcy flow. The Raviart–Thomas mixed finite element on both triangular and quadrilateral meshes is considered for both methods. One is a consistent discretization depending on a weighting parameter scaling as $\mathcal{O}(h^{-1})$, while the other is a penalty-type formulation obtained as the discretization of a perturbation of the original problem and relies on a parameter scaling as $\mathcal{O}(h^{-k-1})$, k being the order of the Raviart–Thomas space. We rigorously prove that both methods are stable and result in optimal convergent numerical schemes with respect to appropriate mesh-dependent norms, although the chosen norms do not scale as the usual L^2-norm. However, we are still able to recover the optimal a priori L^2-error estimates for the velocity field, respectively, for high-order and the lowest-order Raviart–Thomas discretizations, for the first and second numerical schemes. Finally, some numerical examples validating the theory are exhibited.

Keywords: Nitsche method, penalty, Darcy flow, mixed finite element, Neumann boundary conditions

MSC: 65M60

1 Introduction

We consider the finite element approximation for the weak imposition of the Neumann boundary conditions for the Poisson problem in its mixed formulation, also known as Darcy's law in the context of fluid dynamics.

Let us point out that the situation is dual with respect to the standard formulation of the Poisson problem: here the Neumann boundary conditions are essential and to the best of our knowledge it is not clear in the literature how to proceed in order to enforce them by manipulating the weak formulation rather than the functional spaces.

As far as the primary formulation is concerned, a wide variety of techniques have already been proposed and are now well-understood, the most prominent of which are undoubtedly the penalty method introduced in [4], the Lagrange multipliers approach of [3] and, of course, the Nitsche method developed in [15] and later promoted in [17], where the author relates it to the stabilized Lagrange multiplier method of [5].

In this work two different discrete formulations of the Darcy problem for the weak imposition of the Neumann boundary conditions are provided. Both of them are based on the Raviart–Thomas finite element discretization for triangular and quadrilateral meshes. Let us notice that the two schemes do not add any additional degrees of freedoms and moreover, for simplicity, all dimensionless parameters have been set to 1.

The first formulation is a consistent discretization of the Darcy system, hence it falls into the class of Nitsche-type methods. A weighting parameter scaling as $\mathcal{O}(h^{-1})$ needs to be introduced. Let us observe that this formulation had already appeared in the literature in [9] for the lowest-order Raviart–Thomas element, in the context of incompressible flows in fractured media.

The latter, which is inspired by [13] and based on a perturbed variational principle, belongs instead to the family of penalty methods. In this case the penalization parameter scales as $\mathcal{O}(h^{-k-1})$, k being the order

Erik Burman, Chair of Computational Mathematics, University College London, London, UK.
***Corresponding author: Riccardo Puppi,** Chair of Modelling and Numerical Simulation, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. Email: riccardo.puppi@epfl.ch

of the Raviart–Thomas discretization, entailing a much more severe ill-conditioning of the resulting stiffness matrix.

We are able to prove that both formulations are stable and give rise to optimal convergent schemes with respect to suitable mesh-dependent norms, which do not scale as the L^2-norm as it is customary for the Darcy problem. At this point we are able to demonstrate super convergence results that allow us to find an optimal a priori estimate of the velocity error with respect to the L^2-norm, respectively, for any higher order Raviart–Thomas discretization combined with the first method and for the lowest order element with the second formulation.

Note that this work should be considered as a preliminary step towards the much more involved situation of an underlying mesh which is not fitted with the boundary of the physical domain [7].

Let us briefly sketch the outline of the paper. In Sections 2 and 3 we introduce, respectively, the strong formulation of the Darcy problem and a singularly perturbed formulation of it, parametrized by $\varepsilon \geqslant 0^+$. For the latter we are able to show that its solution is stable independently of ε and that, for $\varepsilon \to 0^+$, it converges to the solution of the original problem under some extra regularity assumptions on the data and on the boundary. In Section 3 the Raviart–Thomas finite element is introduced together with our two discrete formulations, both depending on a mesh-dependent weighting parameter γ. As already mentioned, for the first one, $\gamma = h^{-1}$, while for the other $\gamma = h^{-(k+1)}$. In Section 4 we prove the desired stability estimates, with respect to different mesh-dependent norms, guaranteeing the well-posedness of the associated problems. Then, in Section 5, optimal a priori error estimates, in terms of the chosen norms, are demonstrated for the velocity and pressure fields. We demonstrate some super convergent results that enable us, for the two methods, to recover optimality for the L^2-error of the velocity field as well. Finally, some two-dimensional numerical examples are provided in order to corroborate the theory.

2 The Darcy problem and its variational formulation

We briefly introduce some useful notations for the forthcoming analysis. Let D be a Lipschitz-regular *domain* (subset, open, bounded, connected) of $\mathbb{R}^d$. Standard Sobolev spaces $H^s(D)$ for any $s \in \mathbb{R}$ and $H^t(\omega)$ for $t \in [-1, 1]$ are defined on the domain D and on a non-empty open subset of its boundary $\omega \subset \partial D$, see [1], with the convention $H^0(D) := L^2(D)$, $H^0(\omega) := L^2(\omega)$. Moreover, we introduce the following usual notations:

$$
\begin{aligned}
L^2_0(D) &:= L^2(D)/\mathbb{R}\\
H^s_{0,\omega}(D) &:= \{v \in H^s(D) : v\big|_\omega = 0\}\\
H^s_{f,\omega}(D) &:= \{v \in H^s(D) : v\big|_\omega = f\}\\
\boldsymbol{H}^s(D) &:= \big(H^s(D)\big)^d, \quad \boldsymbol{H}^s_{0,\omega}(D) := \big(H^s_{0,\omega}(D)\big)^d, \quad \boldsymbol{H}^s_{f,\omega}(D) := \big(H^s_{f,\omega}(D)\big)^d, \quad \boldsymbol{H}^t(\omega) := \big(H^t(\omega)\big)^d\\
\boldsymbol{H}(\mathrm{div}; D) &:= \{\mathbf{v} \in \boldsymbol{H}^0(D) : \mathrm{div}\,\mathbf{v} \in H^0(D)\}\\
\boldsymbol{H}_{0,\omega}(\mathrm{div}; D) &:= \{\mathbf{v} \in \boldsymbol{H}(\mathrm{div}; D) : \mathbf{v}\cdot\mathbf{n}\big|_\omega = 0\}\\
\boldsymbol{H}_{\sigma,\omega}(\mathrm{div}; D) &:= \{\mathbf{v} \in \boldsymbol{H}(\mathrm{div}; D) : \mathbf{v}\cdot\mathbf{n}\big|_\omega = \sigma\}\\
\boldsymbol{H}^0(\mathrm{div}; D) &:= \{\mathbf{v} \in \boldsymbol{H}(\mathrm{div}; D) : \mathrm{div}\,\mathbf{v} = 0\}
\end{aligned}
$$

where the divergence operator and the traces on ω are defined in the sense of distributions (see [14]). For the sake of convenience we are going to employ the same notation $|\cdot|$ for the volume (Lebesgue) and surface (Hausdorff) measures of $\mathbb{R}^d$.

We also denote as $\mathbb{Q}_{r,s,t}$ the vector space of polynomials of degree at most r in the first variable, at most s in the second and at most t in the third one (analogously for th case $d = 2$), $\mathbb{P}_u$ the vector space of polynomials of degree at most u. For the sake of simplicity of the notation we may write $\mathbb{Q}_k$ instead of $\mathbb{Q}_{k,k}$ or $\mathbb{Q}_{k,k,k}$.

Note that throughout this document C will denote generic constants that may change at each occurrence, but that are always independent of the local mesh size.

Let Ω be a Lipschitz-regular domain of $\mathbb{R}^d$, $d \in \{2,3\}$. We assume its boundary Γ to be partitioned into $\Gamma = \Gamma_N \cup \Gamma_D$ with $\Gamma_N \cap \Gamma_D = \varnothing$. Let us consider the following problem, often associated to a linearized model for the flow of groundwater through our domain Ω, here representing a saturated porous medium with permeability $\varkappa$. Given $\mathbf{f} \in L^2(\Omega;\mathbb{R}^d)$, $g \in L^2(\Omega)$, $u_N \in H^{-1/2}(\Gamma_N)$, $p_D \in H^{1/2}(\Gamma_D)$, we look for $(\mathbf{u}, p) \in H_{u_N,\Gamma_N}(\mathrm{div};\Omega) \times L^2(\Omega)$ such that

$$\begin{cases} \varkappa^{-1}\mathbf{u} - \nabla p = \mathbf{f} & \text{in } \Omega \\ \operatorname{div}\mathbf{u} = g & \text{in } \Omega \\ \mathbf{u}\cdot\mathbf{n} = u_N & \text{on } \Gamma_N \\ p = p_D & \text{on } \Gamma_D. \end{cases} \tag{2.1}$$

The unknowns $\mathbf{u}$ and p represent the seepage velocity and the pressure of the fluid, respectively. The first equation of (2.1) is called *Darcy law* relating the velocity and the pressure gradient of the fluid, the second one expresses *mass conservation*, the third and the fourth equations are, respectively, a *Neumann boundary condition* for the velocity field and a *Dirichlet boundary condition* for the pressure. Moreover, $\varkappa \in \mathbb{R}^{d\times d}$ is symmetric positive definite with eigenvalues λ_i such that $0 < \lambda_{\min} \leqslant \lambda_i \leqslant \lambda_{\max} < +\infty$, for every $i = 1,\dots,d$.

Remark 2.1. Contrary to the case of the Poisson problem, here Dirichlet boundary conditions for the pressure are *natural*, in the sense that they can be implicitly enforced in the weak formulation of the problem, while Neumann boundary conditions for the velocity are *essential*, i.e., they are imposed on the functional space. Moreover, let us observe that in the case of purely Neumann boundary conditions, in order to have well-posedness, we have to 'filter out' the constant pressures, i.e., the trial and test functions for the pressures are required to lie in $L^2_0(\Omega)$, and to impose a compatibility condition on the data: $\int_\Gamma u_N = \int_\Omega g$.

3 A perturbed formulation

Find $(\mathbf{u}^\varepsilon, p^\varepsilon) \in H(\mathrm{div};\Omega)\times L^2(\Omega)$ such that

$$\begin{cases} \varkappa^{-1}\mathbf{u}^\varepsilon - \nabla p^\varepsilon = \mathbf{f} & \text{in } \Omega \\ \operatorname{div}\mathbf{u}^\varepsilon = g & \text{in } \Omega \\ \varepsilon^{-1}\mathbf{u}^\varepsilon\cdot\mathbf{n} = \varepsilon^{-1}u_N - p^\varepsilon & \text{on } \Gamma_N \\ p^\varepsilon = p_D & \text{on } \Gamma_D. \end{cases} \tag{3.1}$$

Note that as $\varepsilon \to 0^+$ problem (3.1) formally degenerates to (2.1). In this sense (3.1) is a perturbation of problem (2.1).

In the subsequent analysis we are going to consider, for the sake of simplicity, $\varkappa = I$ the identity matrix.

Proposition 3.1. Let $(\mathbf{u}^\varepsilon, p^\varepsilon)$ and $(\mathbf{u}, p)$ be the solutions to (3.1) and (2.1), respectively, and assume that $\mathbf{f} \in \boldsymbol{H}(\mathrm{div};\Omega)$. Then there exists $C > 0$ such that

$$\|\mathbf{u} - \mathbf{u}^\varepsilon\|_{L^2(\Omega)} \leqslant C\varepsilon\left(\|\operatorname{div}\mathbf{f}\|_{L^2(\Omega)} + \|g\|_{L^2(\Omega)} + \|u_N\|_{H^{-1/2}(\Gamma_N)} + \|p_D\|_{H^{1/2}(\Gamma_D)}\right).$$

Proof. Let us observe that if $(\mathbf{u}, p)$ and $(\mathbf{u}^\varepsilon, p^\varepsilon)$ solve, respectively, the problems (2.1) and (3.1), then p and p^ε are the solutions of

$$\begin{cases} -\Delta p = -g + \operatorname{div}\mathbf{f} & \text{in } \Omega \\ \dfrac{\partial p}{\partial n} = u_N - \mathbf{f}\cdot\mathbf{n} & \text{on } \Gamma_N \\ p = p_D & \text{on } \Gamma_D \end{cases} \tag{3.2}$$

$$\begin{cases} -\Delta p^\varepsilon = -g + \operatorname{div}\mathbf{f} & \text{in } \Omega \\ \dfrac{\partial p^\varepsilon}{\partial n} + \varepsilon p^\varepsilon = u_N - \mathbf{f}\cdot\mathbf{n} & \text{on } \Gamma_N \\ p^\varepsilon = p_D & \text{on } \Gamma_D. \end{cases} \tag{3.3}$$

Let $\delta := p - p^\varepsilon$, with p and p^ε, respectively, the solutions of (3.2) and (3.3), then δ solves

$$\begin{cases} -\Delta\delta = 0 & \text{in } \Omega \\ \delta + \varepsilon^{-1}\dfrac{\partial\delta}{\partial n} = p & \text{on } \Gamma_N \\ \delta = 0 & \text{on } \Gamma_D. \end{cases} \tag{3.4}$$

We rewrite (3.4) in variational form. Find $\delta \in H^1_{0,\Gamma_D}(\Omega)$ such that

$$(\nabla\delta, \nabla\varphi)_\Omega + \varepsilon\langle\delta, \varphi\rangle_{\Gamma_N} = \varepsilon\langle p, \varphi\rangle_{\Gamma_N} \quad \forall\, \varphi \in H^1_{0,\Gamma_D}(\Omega). \tag{3.5}$$

Since δ vanishes on a part of the boundary, a Poincaré-like inequality entails $\|\delta\|_{H^1(\Omega)} \leqslant C_P \|\nabla\delta\|_{L^2(\Omega)}$, with $C_P > 0$ independent of ε. Now, by testing (3.5) with δ, we get

$$\|\nabla\delta\|^2_{L^2(\Omega)} + \varepsilon\, \|\delta\|^2_{L^2(\Gamma_N)} = \varepsilon\langle p, \delta\rangle_{\Gamma_N}.$$

The Cauchy–Schwarz, a standard trace inequality, and the Poincaré inequality above imply

$$\|\nabla\delta\|^2_{L^2(\Omega)} \lesssim \varepsilon\, \|p\|_{H^1(\Omega)}\, \|\nabla\delta\|_{L^2(\Omega)}\,.$$

On the other hand, p solves (3.2), hence

$$\|\nabla\delta\|^2_{L^2(\Omega)} \lesssim \varepsilon\left(\|\operatorname{div}\mathbf{f}\|_{L^2(\Omega)} + \|g\|_{L^2(\Omega)} + \|u_N\|_{H^{-1/2}(\Gamma_N)} + \|p_D\|_{H^{1/2}(\Gamma_D)}\right)\|\nabla\delta\|_{L^2(\Omega)}\,.$$

Since $\mathbf{u} - \mathbf{u}^\varepsilon = \nabla p + \mathbf{f} - (\nabla p^\varepsilon + \mathbf{f}) = \nabla(p - p^\varepsilon) = \nabla\delta$, then we are done. □

Remark 3.1. With the extra assumption $\mathbf{f}\cdot\mathbf{n} = 0$ on Γ_N, it turns out that problems (2.1) and (3.1) are equivalent to problems (3.2) and (3.3), respectively.

In order to avoid technicalities, let us assume Ω to be a convex domain with a C^2 boundary and the Neumann data to be homogeneous.

Proposition 3.2. Let $(\mathbf{u}^\varepsilon, p^\varepsilon)$ be the solution of (3.1) and suppose that Ω is a convex domain with a Lipschitz polygonal boundary Γ, $\mathbf{f}\cdot\mathbf{n} = 0$ on Γ_N and $u_N = 0$. Then there exists $C > 0$, independent of ε, such that

$$\|\mathbf{u}^\varepsilon\|_{H^1(\Omega)} \leqslant C\left(\|\mathbf{f}\|_{H^1(\Omega)} + \|g\|_{L^2(\Omega)}\right). \tag{3.6}$$

Proof. For the sake of brevity, we are not giving the details of the proof. As above, observe that if $(\mathbf{u}^\varepsilon, p^\varepsilon)$ solves (3.1), then p_ε is the solution of (3.3). We first assume that Ω is a convex C^2-domain and, without loss of generality, we put ourselves in the pure Neumann case $\Gamma = \Gamma_N$. Then the statement follows as a particular case of the a priori inequality of Theorem 3.1.2.3 in [11], which itself is based on the special integration by parts identity of Theorem 3.1.1.1 in [11]. Finally, as observed in Remark 3.2.4.6, the statement can be generalized to a convex domain with Lipschitz polygonal boundary, since it can be approximated with a sequence of convex C^2-domains. □

4 The finite element discretization

Let $(\mathcal{T}_h)_{h>0}$ denote a family of triangular or quadrilateral meshes of Ω. It will be useful to partition the collection of edges (or faces if $d = 3$) $\mathcal{F}_h$ of $\mathcal{T}_h$ into three collections: the internal ones $\mathcal{F}_h^i$ and the ones lying on Γ_N and on Γ_D, grouped respectively in $\mathcal{F}_h^\partial(\Gamma_N)$ and $\mathcal{F}_h^\partial(\Gamma_D)$, $\mathcal{F}_h^\partial = \mathcal{F}_h^\partial(\Gamma_N) \cup \mathcal{F}_h^\partial(\Gamma_D)$. For every $K \in \mathcal{T}_h$, $h > 0$, let $h_K := \operatorname{diam}(K)$ and $h := \max_{K\in\mathcal{T}_h} h_K$. We assume the mesh to be *shape-regular*, i.e., there exists $\sigma > 0$, independent of h, such that $\max_{K\in\mathcal{T}_h} h_K/\rho_K \leqslant \sigma$, ρ_K being the diameter of the largest ball inscribed in K. Moreover, $\mathcal{T}_h$ is supposed to be *quasi-uniform* in the sense that there exists $\tau > 0$, independent of h, such that $\min_{K\in\mathcal{T}_h} h_K \geqslant \tau h$. We fix an orientation for the internal faces, given $f \in \mathcal{F}_h^i$ such that $f = \partial K_1 \cap \partial K_2$ we assume that the normal n_f points from K_1 toward K_2. Note that nothing said from here on will depend on this choice. Let $\varphi : \Omega \to \mathbb{R}$ be smooth enough so that for every $K \in \mathcal{T}_h$ its restriction $\varphi|_K$ can be extended up to the boundary ∂K. Then, for all $f \in \mathcal{F}_h^i$ and a.e. $x \in f$, we define the *jump* of φ as

$$[\varphi]_f(x) := \varphi|_{K_1}(x) - \varphi|_{K_2}(x)$$

where $f = \partial K_1 \cap \partial K_2$. We may remove the subscript f when it is clear from the context to which *facet* (edge if $d = 2$, face if $d = 3$) we refer to.

In order to discretize problem (2.1), we need to choose a suitable couple of subspaces $V_h \subset H(\mathrm{div}; \Omega)$ and $Q_h \subset L^2(\Omega)$. In the following, $\widehat{K}$ will be our *reference element*, and, according to the type of mesh employed, it will be either the unit d-simplex, i.e., the triangle of vertices $(0, 0)$, $(1, 0)$, $(0, 1)$, or the unit d-cube $[0, 1]^d$.

For the triangular meshes, the *Raviart–Thomas* finite element on $\widehat{K}$ is

$$\mathbb{RT}_k(\widehat{K}) := \left(\mathbb{P}_k(\widehat{K})\right)^d \oplus \boldsymbol{x}\widetilde{\mathbb{P}}_k(\widehat{K})$$

while, in the case of quadrilaterals, it reads as follows (see, e.g., [2]):

$$\mathbb{RT}_k(\widehat{K}) := \begin{cases} \mathbb{Q}_{k+1,k}(\widehat{K}) \times \mathbb{Q}_{k,k+1}(\widehat{K}), & d = 2 \\ \mathbb{Q}_{k+1,k,k}(\widehat{K}) \times \mathbb{Q}_{k,k+1,k}(\widehat{K}) \times \mathbb{Q}_{k,k,k+1}(\widehat{K}), & d = 3. \end{cases}$$

We map the reference element to a general $K \in \mathcal{T}_h$ via the *affine map* $F_K : \widehat{K} \to K$, $F_K(\hat{x}) := B_K\hat{x} + b_K$, where $B_K \in \mathbb{R}^{d\times d}$ is diagonal and invertible, and $b_K \in \mathbb{R}^d$. For $H(\mathrm{div}; \Omega)$, the natural way to transform functions from $\widehat{K}$ to K is through the *Piola transform*. Namely, given $\hat{\mathbf{u}} : \widehat{K} \to \mathbb{R}^d$, we define $\mathbf{u} = \mathcal{P}_K\hat{\mathbf{u}} : K \to \mathbb{R}^d$ by

$$\mathbf{u}(x) = \mathcal{P}_K\hat{\mathbf{u}}(x) := |\det(B_K)|^{-1} B_K\hat{\mathbf{u}}(\hat{x}), \qquad \hat{x} = B_K^{-1}(x - b_K).$$

For functions in $L^2(\Omega)$, we just compose them with the affine map, namely $\hat{q} : \widehat{K} \to \mathbb{R}$ is transformed to $q = \hat{q} \circ F_K^{-1} : K \to \mathbb{R}$. In this way, we can define the finite-dimensional subspaces:

$$\begin{aligned} V_h &:= \{\mathbf{v}_h \in H(\mathrm{div}; \Omega) : \mathbf{v}_h|_K \in \mathbb{RT}_k(K) \quad \forall K \in \mathcal{T}_h\} \\ Q_h &:= \{q_h \in L^2(\Omega) : q_h|_K \circ F_K \in \mathbb{P}_k(K) \quad \forall K \in \mathcal{T}_h\} \quad \text{for triangles} \\ Q_h &:= \{q_h \in L^2(\Omega) : q_h|_K \circ F_K \in \mathbb{Q}_k(K) \quad \forall K \in \mathcal{T}_h\} \quad \text{for quadrilaterals} \end{aligned}$$

where $\mathbb{RT}_k(K) := \{\mathcal{P}_K\widehat{\mathbf{w}}_h : \widehat{\mathbf{w}}_h \in \mathbb{RT}_k(\widehat{K})\}$. Remember that in the pure Neumann case, i.e., $\Gamma = \Gamma_N$, we have to filter out constant discrete pressures by imposing the zero average constraint to the space Q_h.

Let us construct the interpolation operator onto the discrete velocities $r_h : \prod_{K\in\mathcal{T}_h} \boldsymbol{H}^s(K^o) \to V_h$, by gluing together the local interpolation operators $r_K : \boldsymbol{H}^s(K^o) \to \mathbb{RT}_k(K)$, $s > 1/2$, K^o denoting the interior of K, and using the natural degrees of freedom of the Raviart–Thomas finite element. For every $\mathbf{v} \in \boldsymbol{H}^s(K^o)$, $s > 1/2$, r_K is uniquely defined by

$$\begin{cases} \langle r_K\mathbf{v} \cdot \mathbf{n}_e, q_h\rangle_f = \langle \mathbf{v} \cdot \mathbf{n}_e, q_h\rangle_f & \forall q_h \in \Psi_k(f) \\ (r_K\mathbf{v}, \mathbf{w}_h)_K = (\mathbf{v}, \mathbf{w}_h)_K & \forall \mathbf{w}_h \in \Psi_k(K), \qquad k > 0 \end{cases}$$

where, for triangles,

$$\Psi_k(K) := (\mathbb{P}_{k-1}(K))^d, \qquad \Psi_k(f) := \mathbb{P}_k(f)$$

and, for quadrilaterals,

$$\Psi_k(K) := \begin{cases} \mathbb{Q}_{k-1,k}(K) \times \mathbb{Q}_{k,k-1}(K), & d = 2 \\ \mathbb{Q}_{k-1,k,k}(K) \times \mathbb{Q}_{k,k-1,k}(K) \times \mathbb{Q}_{k,k,k-1}(K), & d = 3, \end{cases} \qquad \Psi_k(f) := \begin{cases} \mathbb{P}_k(f), & d = 2 \\ \mathbb{Q}_k(f), & d = 3 \end{cases}$$

for all facets f and triangles or quadrilaterals K of $\mathcal{T}_h$. The natural choice in order to interpolate onto Q_h is to employ an elementwise L^2-orthogonal projection, i.e., $\Pi_h : L^2(\Omega) \to Q_h$ such that for every $K \in \mathcal{T}_h$, $\Pi_h|_K := \Pi_K$, where for $\xi \in L^2(\Omega)$

$$(\Pi_K\xi, q_h)_K = (\xi, q_h)_K \quad \forall q_h \in Q_h$$

for every $K \in \mathcal{T}_h$.

Remark 4.1. It is worth mentioning that the following numerical analysis remains valid if we employ another $H(\mathrm{div})$-conforming discretization instead, such as the so-called Brezzi–Douglas–Marini mixed element [6].

Proposition 4.1. The following diagram commutes:

$$\begin{tikzcd} H(\mathrm{div};\Omega) \cap \prod_{K\in\mathcal{T}_h} \boldsymbol{H}^s(K^\circ) \arrow[r, "\mathrm{div}"] \arrow[d, "r_h"] & L^2(\Omega) \arrow[d, "\Pi_h"] \\ V_h \arrow[r, "\mathrm{div}"] & Q_h. \end{tikzcd}$$

In particular, it holds

$$\mathrm{div}\, V_h = Q_h.$$

Proof. For the commutative diagram note that, for every $\mathbf{v} \in H(\mathrm{div};\Omega) \cap \prod_{K\in\mathcal{T}_h} \boldsymbol{H}^s(K^\circ)$, it holds

$$\begin{aligned}(\Pi_h \,\mathrm{div}\,\mathbf{v}, \varphi_h)_K &= (\mathrm{div}\,\mathbf{v}, \varphi_h)_K = -(\mathbf{v}, \nabla\varphi_h)_K + \langle \varphi_h, \mathbf{v}\cdot\mathbf{n}\rangle_{\partial K} \\ &= -(r_h\mathbf{v}, \nabla\varphi_h)_K + \langle \varphi_h, r_h\mathbf{v}\cdot\mathbf{n}\rangle_{\partial K} = (\mathrm{div}\, r_h\mathbf{v}, \varphi_h)_K \quad \forall\, \varphi_h \in \mathbb{P}_k(K) \text{ (resp. } \mathbb{Q}_k(K)).\end{aligned}$$

A direct calculation readily shows the inclusion $\mathrm{div}\, V_h \subseteq Q_h$. Let us prove the other one. Let $q_h \in Q_h$, then, by the surjectivity of $\mathrm{div} : \boldsymbol{H}^1(\Omega) \to L^2(\Omega)$ [6], there exists $\mathbf{v} \in H(\mathrm{div};\Omega) \cap \prod_{K\in\mathcal{T}_h} \boldsymbol{H}^s(K^\circ) \subseteq \boldsymbol{H}^1(\Omega)$ such that $\mathrm{div}\,\mathbf{v} = q_h$. Let us define $\mathbf{v}_h := r_h\mathbf{v}$. Thanks to the commutativity diagram we have $\mathrm{div}\,\mathbf{v}_h = q_h$. □

We are now ready to introduce the discrete formulations we want to analyze.

4.1 First formulation

Find $(\mathbf{u}_h, p_h) \in V_h \times Q_h$ such that

$$\begin{cases} a_h(\mathbf{u}_h, \mathbf{v}_h) + b_1(\mathbf{v}_h, p_h) = (\mathbf{f}, \mathbf{v}_h)_\Omega + h^{-1}\langle u_N, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N} + \langle p_D, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_D} & \forall\, \mathbf{v}_h \in V_h \\ b_m(\mathbf{u}_h, q_h) = (g, q_h)_\Omega - m\langle q_h, u_N\rangle_{\Gamma_N} & \forall\, q_h \in Q_h \end{cases} \tag{4.1}$$

where $m \in \{0, 1\}$. Here,

$$\begin{aligned} a_h(\mathbf{w}_h, \mathbf{v}_h) &:= (\mathbf{w}_h, \mathbf{v}_h)_\Omega + h^{-1}\langle \mathbf{w}_h\cdot\mathbf{n}, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N} && \forall\, \mathbf{w}_h, \mathbf{v}_h \in V_h \\ b_m(\mathbf{w}_h, q_h) &:= (q_h, \mathrm{div}\,\mathbf{w}_h)_\Omega - m\langle q_h, \mathbf{w}_h\cdot\mathbf{n}\rangle_{\Gamma_N} && \forall\, \mathbf{w}_h \in V_h, \quad q_h \in Q_h. \end{aligned}$$

In what follows just the analysis for the symmetric case $m = 1$ will be presented, however numerical results will be provided for the case $m = 0$ as well.

4.2 Second formulation

Find $(\mathbf{u}_h, p_h) \in V_h \times Q_h$ such that

$$\begin{cases} a_\varepsilon(\mathbf{u}_h, \mathbf{v}_h) + b_0(\mathbf{v}_h, p_h) = (\mathbf{f}, \mathbf{v}_h)_\Omega + \langle \varepsilon^{-1} u_N, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N} + \langle p_D, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_D} & \forall\, \mathbf{v}_h \in V_h \\ b_0(\mathbf{u}_h, q_h) = (g, q_h)_\Omega & \forall\, q_h \in Q_h \end{cases} \tag{4.2}$$

where

$$a_\varepsilon(\mathbf{w}_h, \mathbf{v}_h) := (\mathbf{u}_h, \mathbf{v}_h)_\Omega + \varepsilon^{-1}\langle \mathbf{w}_h\cdot\mathbf{n}, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N} \qquad \forall\, \mathbf{w}_h, \mathbf{v}_h \in V_h.$$

Remark 4.2. Let us observe that the non-symmetric version of problem (4.1), i.e., with $m = 0$, and formulation (4.2), thanks to Proposition 4.1 allows for a *weakly divergence-free* numerical solution $\mathbf{u}_h$, namely, $\mathrm{div}\,\mathbf{u}_h = 0$ in the sense of L^2, provided that the right-hand side g vanishes.

Lemma 4.1. *Formulations* (4.1) *and* (4.2) *are consistent discretizations of* (2.1) *and* (3.1), *respectively.*

Proof. It is clear that (4.1) is a consistent discetization of (2.1). Let $(\mathbf{u}^\varepsilon, p^\varepsilon)$ be the solution to (3.1). Of course, we have

$$b_0(\mathbf{u}^\varepsilon, q_h) = (g, q_h) \quad \forall\, q_h \in Q_h.$$

By integrating by parts the first equation of (3.1), we obtain

$$(\mathbf{u}^\varepsilon, \mathbf{v}_h)_\Omega + b_0(\mathbf{v}_h, p^\varepsilon) - \langle p^\varepsilon, \mathbf{v}_h \cdot \mathbf{n}\rangle_{\Gamma_N} = (\mathbf{f}, \mathbf{v}_h)_\Omega + \langle p_D, \mathbf{v}_h \cdot \mathbf{n}\rangle_{\Gamma_D} \quad \forall\, \mathbf{v}_h \in V_h. \tag{4.3}$$

By performing *static condensation* of the multiplier from the boundary conditions, we obtain

$$p^\varepsilon = \varepsilon^{-1}\left(\mathbf{u}_N - \mathbf{u}^\varepsilon \cdot \mathbf{n}\right) \quad \text{on } \Gamma_N. \tag{4.4}$$

Substituting (4.4) back into (4.3), we obtain

$$a_\varepsilon(\mathbf{u}^\varepsilon, \mathbf{v}_h) + b_0(\mathbf{v}_h, p^\varepsilon) = (\mathbf{f}, \mathbf{v}_h)_\Omega + \langle \varepsilon^{-1}\mathbf{u}_N, \mathbf{v}_h \cdot \mathbf{n}\rangle_{\Gamma_N} + \langle p_D, \mathbf{v}_h \cdot \mathbf{n}\rangle_{\Gamma_D} \quad \forall\, \mathbf{v}_h \in V_h. \qquad \square$$

For the numerical analysis of (4.1), we endow the discrete spaces with the following mesh-dependent norms

$$\begin{aligned}
\|\mathbf{v}_h\|^2_{0,h} &:= \|\mathbf{v}_h\|^2_{L^2(\Omega)} + \sum_{f \in \mathcal{F}_h^\partial(\Gamma_N)} h^{-1}\,\|\mathbf{v}_h \cdot \mathbf{n}\|^2_{L^2(f)} \\
\|q_h\|^2_{1,h} &:= \sum_{K \in \mathcal{T}_h} \|\nabla q_h\|^2_{L^2(K)} + \sum_{f \in \mathcal{F}_h^i} h^{-1}\,\|[\![q_h]\!]\|^2_{L^2(f)} + \sum_{f \in \mathcal{F}_h^\partial(\Gamma_D)} h^{-1}\,\|q_h\|^2_{L^2(f)}
\end{aligned}$$

while for (4.2) we are going to employ:

$$\begin{aligned}
\|\mathbf{v}_h\|^2_{0,h,\varepsilon} &:= \|\mathbf{v}_h\|^2_{L^2(\Omega)} + \sum_{f \in \mathcal{F}_h^\partial(\Gamma_N)} \varepsilon^{-1}\,\|\mathbf{v}_h \cdot \mathbf{n}\|^2_{L^2(f)} \\
\|q_h\|^2_{1,h,\varepsilon} &:= \sum_{K \in \mathcal{T}_h} \|\nabla q_h\|^2_{L^2(K)} + \sum_{f \in \mathcal{F}_h^i} h^{-1}\,\|[\![q_h]\!]\|^2_{L^2(f)} + \sum_{f \in \mathcal{F}_h^\partial} h^{-1}\,\|q_h\|^2_{L^2(f)}
\end{aligned}$$

for every $\mathbf{v}_h \in V_h$ and $q_h \in Q_h$.

Remark 4.3. Informally speaking, the idea of both approaches is to unbalance the norms in order to go back to the elliptic case. Note that the natural functional setting for the mixed formulation of the Poisson problem is $H(\mathrm{div};\Omega) \times L^2(\Omega)$, but here we consider norms that induce the same topology as that of $[L^2(\Omega)]^d \times H^1(\Omega)$. Moreover, we observe that in both formulations (4.1) and (4.2) a superpenalty parameter γ is imposed in the flux variable. Indeed, the natural weight, mimicking the $H^{-1/2}$-scalar product, would be $h\langle \mathbf{u}_h \cdot \mathbf{n}, \mathbf{v}_h \cdot \mathbf{n}\rangle_{\Gamma_N}$. However such a weight does not lead to an optimally converging scheme. In addition, this is also what destroys the conditioning (see Section 7.2).

5 Stability estimates

In this section we carry on at the same time the proofs of the well-posedness of the two discrete formulations.

Proposition 5.1. There exist $M_{a_h}, M_{a_\varepsilon}, M_{b_m} > 0$, $m = 0, 1$, such that

$$\begin{aligned}
|a_h(\mathbf{w}_h, \mathbf{v}_h)| &\leqslant M_{a_h}\,\|\mathbf{w}_h\|_{0,h}\,\|\mathbf{v}_h\|_{0,h} && \forall\, \mathbf{w}_h, \mathbf{v}_h \in V_h \\
|a_\varepsilon(\mathbf{w}_h, \mathbf{v}_h)| &\leqslant M_{a_\varepsilon}\,\|\mathbf{w}_h\|_{0,h,\varepsilon}\,\|\mathbf{v}_h\|_{0,h,\varepsilon} && \forall\, \mathbf{w}_h, \mathbf{v}_h \in V_h \\
|b_1(\mathbf{v}_h, q_h)| &\leqslant M_{b_1}\,\|\mathbf{v}_h\|_{0,h}\,\|q_h\|_{1,h} && \forall\, \mathbf{v}_h \in V_h, \quad q_h \in Q_h \\
|b_0(\mathbf{v}_h, q_h)| &\leqslant M_{b_0}\,\|\mathbf{v}_h\|_{0,h,\varepsilon}\,\|q_h\|_{1,h,\varepsilon} && \forall\, \mathbf{v}_h \in V_h, \quad q_h \in Q_h.
\end{aligned}$$

Proof. Let $\mathbf{w}_h, \mathbf{v}_h \in V_h$, $q_h \in Q_h$ be arbitrary. It holds

$$|a_h(\mathbf{w}_h, \mathbf{v}_h)| \leqslant \|\mathbf{w}_h\|_{L^2(\Omega)} \|\mathbf{v}_h\|_{L^2(\Omega)} + h^{-1/2} \|\mathbf{v}_h \cdot \mathbf{n}\|_{L^2(\Gamma_N)} h^{-1/2} \|\mathbf{w}_h \cdot \mathbf{n}\|_{L^2(\Gamma_N)} \leqslant \|\mathbf{w}_h\|_{0,h} \|\mathbf{v}_h\|_{0,h}$$
$$|a_\varepsilon(\mathbf{w}_h, \mathbf{v}_h)| \leqslant \|\mathbf{w}_h\|_{L^2(\Omega)} \|\mathbf{v}_h\|_{L^2(\Omega)} + \varepsilon^{-1/2} \|\mathbf{v}_h \cdot \mathbf{n}\|_{L^2(\Gamma_N)} \varepsilon^{-1/2} \|\mathbf{w}_h \cdot \mathbf{n}\|_{L^2(\Gamma_N)} \leqslant \|\mathbf{w}_h\|_{0,h,\varepsilon} \|\mathbf{v}_h\|_{0,h,\varepsilon} .$$

By integration by parts, we get

$$\begin{aligned} b_1(\mathbf{v}_h, q_h) &= (q_h, \operatorname{div} \mathbf{v}_h)_\Omega - \langle q_h, \mathbf{v}_h \cdot \mathbf{n}\rangle_{\Gamma_N} = \sum_{K\in\mathcal{T}_h} (q_h, \operatorname{div} \mathbf{v}_h)_K - \sum_{f\in\mathcal{F}_h^\partial(\Gamma_N)} \langle q_h, \mathbf{v}_h \cdot \mathbf{n}\rangle_f \\ &= -\sum_{K\in\mathcal{T}_h} (\nabla q_h, \mathbf{v}_h)_K + \sum_{f\in\mathcal{F}_h^i} \langle [q_h], \mathbf{v}_h \cdot \mathbf{n}\rangle_f + \sum_{f\in\mathcal{F}_h^\partial(\Gamma_D)} \langle q_h, \mathbf{v}_h \cdot \mathbf{n}\rangle_f . \end{aligned}$$

Thus,

$$\begin{aligned} |b_1(\mathbf{v}_h, q_h)| \leqslant &\sum_{K\in\mathcal{T}_h} \|\nabla q_h\|_{L^2(K)} \|\mathbf{v}_h\|_{L^2(K)} + \sum_{f\in\mathcal{F}_h^i} h^{-1/2} \|[q_h]\|_{L^2(f)} h^{1/2} \|\mathbf{v}_h \cdot \mathbf{n}\|_{L^2(f)} \\ &+ \sum_{f\in\mathcal{F}_h^\partial(\Gamma_D)} h^{-1/2} \|q_h\|_{L^2(f)} h^{1/2} \|\mathbf{v}_h \cdot \mathbf{n}\|_{L^2(f)} . \end{aligned}$$

We recall some standard inverse inequalities, namely,

$$h^{1/2} \|\mathbf{v}_h \cdot \mathbf{n}\|_{L^2(f)} \lesssim \|\mathbf{v}_h\|_{L^2(K)} , \qquad f \in \mathcal{F}_h^\partial(\Gamma_D), \quad f \in \mathcal{F}_h^i, \quad f \subset \partial K. \tag{5.1}$$

In this way we obtain

$$|b_1(\mathbf{v}_h, q_h)| \lesssim \|\mathbf{v}_h\|_{0,h} \|q_h\|_{1,h} .$$

On the other hand,

$$\begin{aligned} b_0(\mathbf{v}_h, q_h) &= -\sum_{K\in\mathcal{T}_h} (\nabla q_h, \mathbf{v}_h)_K + \langle q_h, \mathbf{v}_h \cdot \mathbf{n}\rangle_{\partial K} = -\sum_{K\in\mathcal{T}_h} (\nabla q_h, \mathbf{v}_h)_K + \sum_{f\in\mathcal{F}_h^i} \langle [q_h], \mathbf{v}_h \cdot \mathbf{n}\rangle_f \\ &+ \sum_{f\in\mathcal{F}_h^\partial(\Gamma_N)} \langle q_h, \mathbf{v}_h \cdot \mathbf{n}\rangle_f + \sum_{f\in\mathcal{F}_h^\partial(\Gamma_D)} \langle q_h, \mathbf{v}_h \cdot \mathbf{n}\rangle_f . \end{aligned}$$

We have

$$\begin{aligned} |b_0(\mathbf{v}_h, q_h)| \leqslant &\sum_{K\in\mathcal{T}_h} \|\nabla q_h\|_{L^2(K)} \|\mathbf{v}_h\|_{L^2(K)} + \sum_{f\in\mathcal{F}_h^i} h^{-1/2} \|[q_h]\|_{L^2(f)} h^{1/2} \|\mathbf{v}_h \cdot \mathbf{n}\|_{L^2(f)} \\ &+ \sum_{f\in\mathcal{F}_h^\partial} h^{-1/2} \|q_h\|_{L^2(f)} h^{1/2} \|\mathbf{v}_h \cdot \mathbf{n}\|_{L^2(f)} \lesssim \|\mathbf{v}_h\|_{0,h,\varepsilon} \|q_h\|_{1,h,\varepsilon} \end{aligned}$$

having used again (5.1) and $h^{-1/2} < \varepsilon^{-1/2}$ for $\varepsilon \ll h$. □

Proposition 5.2. There exist $\alpha_{a_h}, \alpha_{a_\varepsilon} > 0$ such that

$$\begin{aligned} a_h(\mathbf{v}_h, \mathbf{v}_h) &\geqslant \alpha_{a_h} \|\mathbf{v}_h\|_{0,h}^2 \qquad \forall\, \mathbf{v}_h \in V_h \\ a_\varepsilon(\mathbf{v}_h, \mathbf{v}_h) &\geqslant \alpha_{a_\varepsilon} \|\mathbf{v}_h\|_{0,h,\varepsilon}^2 \qquad \forall\, \mathbf{v}_h \in V_h . \end{aligned}$$

Proof. Let us take $\mathbf{v}_h \in V_h$ arbitrary and compute

$$a_h(\mathbf{v}_h, \mathbf{v}_h) = \|\mathbf{v}_h\|_{L^2(\Omega)}^2 + h^{-1} \|\mathbf{v}_h \cdot \mathbf{n}\|_{L^2(\Gamma_N)}^2 = \|\mathbf{v}_h\|_{0,h}^2 .$$

The other coercivity estimate follows analogously. Hence, $\alpha_{a_h} = \alpha_{a_\varepsilon} = 1$. □

Proposition 5.3. There exist $\beta_m > 0$, $m \in \{0, 1\}$, such that

$$\begin{aligned} \inf_{q_h\in Q_h} \sup_{\mathbf{v}_h\in V_h} \frac{b_1(\mathbf{v}_h, q_h)}{\|\mathbf{v}_h\|_{0,h} \|q_h\|_{1,h}} &\geqslant \beta_1 \\ \inf_{q_h\in Q_h} \sup_{\mathbf{v}_h\in V_h} \frac{b_0(\mathbf{v}_h, q_h)}{\|\mathbf{v}_h\|_{0,h,\varepsilon} \|q_h\|_{1,h,\varepsilon}} &\geqslant \beta_0 . \end{aligned}$$

Proof. We start with $m = 1$. Let us fix $q_h \in Q_h$ arbitrary. We construct $\mathbf{v}_h$ by using the DOFs of the Raviart–Thomas space:

$$\langle \mathbf{v}_h \cdot \mathbf{n}, \varphi_h \rangle_f = h^{-1} \langle [q_h], \varphi_h \rangle_f \quad \forall f \in \mathcal{F}_h^i, \quad \varphi_h \in \Psi_k(f) \tag{5.2}$$

$$\langle \mathbf{v}_h \cdot \mathbf{n}, \varphi_h \rangle_f = 0 \quad \forall f \in \mathcal{F}_h^\partial(\Gamma_N), \quad \varphi_h \in \Psi_k(f) \tag{5.3}$$

$$\langle \mathbf{v}_h \cdot \mathbf{n}, \varphi_h \rangle_f = h^{-1} \langle q_h, \varphi_h \rangle_f \quad \forall f \in \mathcal{F}_h^\partial(\Gamma_D), \quad \varphi_h \in \Psi_k(f) \tag{5.4}$$

$$(\mathbf{v}_h, \boldsymbol{\psi}_h)_K = -(\nabla q_h, \boldsymbol{\psi}_h)_K \quad \forall K \in \mathcal{T}_h, \quad \boldsymbol{\psi}_h \in \Psi_k(K), \quad k > 0. \tag{5.5}$$

By using the definition of $\mathbf{v}_h$,

$$\begin{aligned} b_1(\mathbf{v}_h, q_h) = (q_h, \operatorname{div} \mathbf{v}_h)_\Omega - \langle q_h, \mathbf{v}_h \cdot \mathbf{n} \rangle_{\Gamma_N} &= \sum_{K \in \mathcal{T}_h} (q_h, \operatorname{div} \mathbf{v}_h)_K - \sum_{f \in \mathcal{F}_h^\partial(\Gamma_N)} \langle q_h, \mathbf{v}_h \cdot \mathbf{n} \rangle_f \\ &= -\sum_{K \in \mathcal{T}_h} (\nabla q_h, \mathbf{v}_h)_K + \langle q_h, \mathbf{v}_h \cdot \mathbf{n} \rangle_{\partial K} - \sum_{f \in \mathcal{F}_h^\partial(\Gamma_N)} \langle q_h, \mathbf{v}_h \cdot \mathbf{n} \rangle_f \\ &= -\sum_{K \in \mathcal{T}_h} (\nabla q_h, \mathbf{v}_h)_K + \sum_{f \in \mathcal{F}_h^i} \langle [q_h], \mathbf{v}_h \cdot \mathbf{n} \rangle_f + \sum_{f \in \mathcal{F}_h^\partial(\Gamma_D)} \langle q_h, \mathbf{v}_h \cdot \mathbf{n} \rangle_f \\ &= \sum_{K \in \mathcal{T}_h} \|\nabla q_h\|^2_{L^2(K)} + \sum_{f \in \mathcal{F}_h^i} h^{-1} \|[q_h]\|^2_{L^2(f)} + \sum_{f \in \mathcal{F}_h^\partial(\Gamma_D)} h^{-1} \|q_h\|^2_{L^2(f)} = \|q_h\|^2_{1,h}. \end{aligned}$$

Finally, let us show that $\|\mathbf{v}_h\|_{0,h} \leqslant C \|q_h\|_{1,h}$. Note that for every $f \in \mathcal{F}_h^\partial(\Gamma_N)$, since $\mathbf{v}_h \cdot \mathbf{n}\big|_f \in \mathbb{P}_k(f)$, (5.3) implies

$$\|\mathbf{v}_h \cdot \mathbf{n}\|^2_{L^2(f)} = \langle \mathbf{v}_h \cdot \mathbf{n}, \mathbf{v}_h \cdot \mathbf{n} \rangle_f = 0 \quad \Longrightarrow \quad \|\mathbf{v}_h \cdot \mathbf{n}\|_{L^2(f)} = 0.$$

Then, let us show $\|\mathbf{v}_h\|_{L^2(\Omega)} \leqslant C \|q_h\|_{1,h}$. From (5.2) it holds $\mathbf{v}_h \cdot \mathbf{n}\big|_f = h_K^{-1} \pi_{f,k} [q_h]\big|_f$ for every $f \in \mathcal{F}_h^i$ and from (5.5) we have $\pi_{K,k} \mathbf{v}_h\big|_K = -\pi_{K,k} \nabla q_h\big|_K$ for every $K \in \mathcal{T}_h$. Note that here $\pi_{K,k}$ denotes the L^2-orthogonal projection onto $\Psi_k(K)$. Similarly, $\pi_{f,k}$ is the L^2-projection onto $\Psi_k(f)$. From finite dimensionality it holds $\|\widehat{\mathbf{v}}_h\|^2_{L^2(\widehat{K})} \lesssim \|\pi_{\widehat{K},k} \widehat{\mathbf{v}}_h\|^2_{L^2(\widehat{K})} + \|\widehat{\mathbf{v}}_h \cdot \widehat{\mathbf{n}}\|^2_{L^2(\widehat{f})}$. Hence, $\|\mathbf{v}_h\|^2_{L^2(K)} \lesssim \|\nabla q_h\|^2_{L^2(K)} + h_K^{-1} \|[q_h]\|^2_{L^2(f)}$, f being a facet of K, which follows by a standard scaling argument (see Prop. 2.1 of [8]) and by construction of $\mathbf{v}_h$.

Let us now take $m = 0$ and $q_h \in Q_h$. We define $\mathbf{v}_h$ as follows:

$$\langle \mathbf{v}_h \cdot \mathbf{n}, \varphi_h \rangle_f = h^{-1} \langle [q_h], \varphi_h \rangle_f \quad \forall f \in \mathcal{F}_h^i, \quad \varphi_h \in \Psi_k(f) \tag{5.6}$$

$$\langle \mathbf{v}_h \cdot \mathbf{n}, \varphi_h \rangle_f = h^{-1} \langle q_h, \varphi_h \rangle_f \quad \forall f \in \mathcal{F}_h^\partial(\Gamma_N), \quad \varphi_h \in \Psi_k(f) \tag{5.7}$$

$$\langle \mathbf{v}_h \cdot \mathbf{n}, \varphi_h \rangle_f = h^{-1} \langle q_h, \varphi_h \rangle_f \quad \forall f \in \mathcal{F}_h^\partial(\Gamma_D), \quad \varphi_h \in \Psi_k(f) \tag{5.8}$$

$$(\mathbf{v}_h, \boldsymbol{\psi}_h)_K = -(\nabla q_h, \boldsymbol{\psi}_h)_K \quad \forall K \in \mathcal{T}_h, \quad \boldsymbol{\psi}_h \in \Psi_k(K), \quad k > 0 \tag{5.9}$$

$$\begin{aligned} b_0(\mathbf{v}_h, q_h) = (q_h, \operatorname{div} \mathbf{v}_h)_\Omega &= \sum_{K \in \mathcal{T}_h} (q_h, \operatorname{div} \mathbf{v}_h)_K = -\sum_{K \in \mathcal{T}_h} (\nabla q_h, \mathbf{v}_h)_K + \langle q_h, \mathbf{v}_h \cdot \mathbf{n} \rangle_{\partial K} \\ &= -\sum_{K \in \mathcal{T}_h} (\nabla q_h, \mathbf{v}_h)_K + \sum_{f \in \mathcal{F}_h^i} \langle [q_h], \mathbf{v}_h \cdot \mathbf{n} \rangle_f + \sum_{f \in \mathcal{F}_h^\partial(\Gamma_N)} \langle q_h, \mathbf{v}_h \cdot \mathbf{n} \rangle_f + \sum_{f \in \mathcal{F}_h^\partial(\Gamma_D)} \langle q_h, \mathbf{v}_h \cdot \mathbf{n} \rangle_f \\ &= \sum_{K \in \mathcal{T}_h} \|\nabla q_h\|^2_{L^2(K)} + \sum_{f \in \mathcal{F}_h^i} h^{-1} \|[q_h]\|^2_{L^2(f)} + \sum_{f \in \mathcal{F}_h^\partial} h^{-1} \|q_h\|^2_{L^2(f)} = \|q_h\|^2_{1,h,\varepsilon}. \end{aligned}$$

We refer to [13] for the inequality $\|\mathbf{v}_h\|_{0,h,\varepsilon} \leqslant C \|q_h\|_{1,h,\varepsilon}$. □

6 A priori error estimates

In this section we will prove a priori error estimates for the formulations (4.1) and (4.2).

We observe that all the constants appearing throughout this section and concerning the error bounds for the formulation (4.2) are independent of the parameter ε. This is due to the orthogonality properties of the interpolants along the boundary.

Lemma 6.1. *Let* $(\mathbf{u}, p)$ *be the solution of the continuous problem* (2.1) *and* $(\mathbf{u}_h, p_h) \in V_h \times Q_h$ *the one of the discrete problem* (4.1) *with* $m = 1$. *Then*

$$\|\mathbf{u}_h - r_h\mathbf{u}\|_{0,h} + \|p_h - \Pi_h p\|_{1,h} \lesssim \|\mathbf{u} - r_h\mathbf{u}\|_{L^2(\Omega)} + h^{1/2} \sum_{f\in\mathcal{F}_h^\partial(\Gamma_N)} \|p - \Pi_h p\|_{L^2(f)} .$$

Proof. The stability estimates previously shown for $a_h(\cdot,\cdot)$ and $b_1(\cdot,\cdot)$ with respect to $\|\cdot\|_{0,h}$ and $\|\cdot\|_{1,h}$ imply

$$|||\boldsymbol{\eta}_h, s_h|||_h \lesssim \sup_{(\mathbf{v}_h,q_h)} \frac{\mathcal{A}_h\left((\boldsymbol{\eta}_h, s_h), (\mathbf{v}_h, q_h)\right)}{|||\mathbf{v}_h, q_h|||_h} \qquad \forall\, (\boldsymbol{\eta}_h, s_h) \in V_h \times Q_h \tag{6.1}$$

where

$$\mathcal{A}_h\left((\boldsymbol{\eta}_h, s_h), (\mathbf{v}_h, q_h)\right) := a_h(\boldsymbol{\eta}_h, \mathbf{v}_h) + b_1(\mathbf{v}_h, s_h) + b_1(\boldsymbol{\eta}_h, q_h)$$

$$|||\boldsymbol{\eta}_h, s_h|||_h^2 := \|\boldsymbol{\eta}_h\|_{0,h}^2 + \|s_h\|_{1,h}^2 .$$

Using (6.1), for $(\mathbf{u}_h - r_h\mathbf{u}, p_h - \Pi_h p)$ there exists $(\mathbf{v}_h, q_h) \in V_h \times Q_h$ such that

$$\|\mathbf{u}_h - r_h\mathbf{u}_h\|_{0,h} + \|p_h - \Pi_h p\|_{1,h} \leqslant \sqrt{d}|||\mathbf{u}_h - r_h\mathbf{u}, p_h - \Pi_h p|||_h \lesssim \frac{\mathcal{A}_h\left((\mathbf{u}_h - r_h\mathbf{u}, p_h - \Pi_h p), (\mathbf{v}_h, q_h)\right)}{|||\mathbf{v}_h, q_h|||_h}.$$

Hence, we have

$$\begin{aligned}\mathcal{A}_h((\mathbf{u}_h - r_h\mathbf{u}, p_h - \Pi_h p), (\mathbf{v}_h, q_h)) = {} & (\mathbf{u}_h - \mathbf{u}, \mathbf{v}_h)_{L^2(\Omega)} + (\mathbf{u} - r_h\mathbf{u}, \mathbf{v}_h)_{L^2(\Omega)} + h^{-1}\langle(\mathbf{u}_h - \mathbf{u})\cdot\mathbf{n}, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N} \\ & + h^{-1}\langle(\mathbf{u} - r_h\mathbf{u})\cdot\mathbf{n}, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N} + b_0(\mathbf{v}_h, p_h - p) + b_0(\mathbf{v}_h, p - \pi_h p) \\ & - \langle p_h - p, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N} - \langle p - \Pi_h p, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N} \\ & + b_1(\mathbf{u}_h - \mathbf{u}, q_h) + b_1(\mathbf{u} - r_h\mathbf{u}, q_h).\end{aligned} \tag{6.2}$$

By construction of r_h and Π_h we have, respectively,

$$\begin{aligned} h^{-1}\langle(\mathbf{u} - r_h\mathbf{u})\cdot\mathbf{n}, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N} &= 0 \qquad \forall\, \mathbf{v}_h \in V_h \\ b_1(\mathbf{u} - r_h\mathbf{u}, q_h) = -\sum_{K\in\mathcal{T}_h} (\nabla q_h, \mathbf{u} - r_h\mathbf{u})_K + \sum_{f\in\mathcal{F}_h^i} \langle[q_h], (\mathbf{u} - r_h\mathbf{u})\cdot\mathbf{n}\rangle_f &= 0 \qquad \forall\, q_h \in Q_h \\ b_0(\mathbf{v}_h, p - \Pi_h p) &= 0 \qquad \forall\, \mathbf{v}_h \in V_h. \end{aligned}$$

By consistency, we have

$$\begin{aligned}(\mathbf{u}_h - \mathbf{u}, \mathbf{v}_h)_{L^2(\Omega)} + h^{-1}\langle(\mathbf{u}_h - \mathbf{u})\cdot\mathbf{n}, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N} + b_0(\mathbf{v}_h, p_h - p) + \langle p_h - p, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N} &= 0 \qquad \forall\, \mathbf{v}_h \in V_h \\ b_1(\mathbf{u}_h - \mathbf{u}, q_h) &= 0 \qquad \forall\, q_h \in Q_h.\end{aligned}$$

Hence, in (6.2) we are left with

$$\mathcal{A}_h\left((\mathbf{u}_h - r_h\mathbf{u}, p_h - \Pi_h p), (\mathbf{v}_h, q_h)\right) = (\mathbf{u} - r_h\mathbf{u}, \mathbf{v}_h)_{L^2(\Omega)} - \langle p - \Pi_h p, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N}.$$

We have

$$\begin{aligned}(\mathbf{u} - r_h\mathbf{u}, \mathbf{v}_h)_{L^2(\Omega)} - \langle p - \Pi_h p, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N} \leqslant {} & \|\mathbf{u} - r_h\mathbf{u}\|_{L^2(\Omega)}\, \|\mathbf{v}_h\|_{L^2(\Omega)} \\ & + \sum_{f\in\mathcal{F}_h^\partial(\Gamma_N)} h^{1/2}\, \|(p - \Pi_h p)\|_{L^2(f)}\, h^{-1/2}\, \|\mathbf{v}_h\cdot\mathbf{n}\|_{L^2(\Gamma_N)}\end{aligned}$$

and we can write

$$\begin{aligned}\|\mathbf{u}_h - r_h\mathbf{u}\|_{0,h} + \|p_h - \Pi_h p\|_{1,h} &\lesssim \frac{\left(\|\mathbf{u} - r_h\mathbf{u}\|_{L^2(\Omega)} + h^{1/2}\sum_{f\in\mathcal{F}_h^\partial(\Gamma_N)} \|p - \Pi_h p\|_{L^2(f)}\right) |||\mathbf{v}_h, 0|||_h}{|||\mathbf{v}_h, q_h|||_h} \\ &\lesssim \|\mathbf{u} - r_h\mathbf{u}\|_{L^2(\Omega)} + h^{1/2} \sum_{f\in\mathcal{F}_h^\partial(\Gamma_N)} \|p - \Pi_h p\|_{L^2(f)} .\end{aligned}$$

□

Lemma 6.2. *Let* $(\mathbf{u}^\varepsilon, p^\varepsilon)$ *be the solution of the perturbed continuous problem* (3.1) *and* $(\mathbf{u}_h, p_h) \in V_h \times Q_h$ *the one of the discrete problem* (4.2). *Then*

$$\|\mathbf{u}_h - r_h\mathbf{u}^\varepsilon\|_{0,h,\varepsilon} + \|p_h - \Pi_h p^\varepsilon\|_{1,h,\varepsilon} \lesssim \|\mathbf{u}^\varepsilon - r_h\mathbf{u}^\varepsilon\|_{L^2(\Omega)}.$$

Proof. The stability estimates previously shown for $a_\varepsilon(\cdot,\cdot)$ and $b_0(\cdot,\cdot)$ with respect to $\|\cdot\|_{0,h,\varepsilon}$ and $\|\cdot\|_{1,h,\varepsilon}$ imply

$$|\!|\!|\boldsymbol{\eta}, s_h|\!|\!|_{h,\varepsilon} \lesssim \sup_{(\mathbf{v}_h,q_h)} \frac{\mathcal{A}_\varepsilon\left((\boldsymbol{\eta}_h, s_h), (\mathbf{v}_h, q_h)\right)}{|\!|\!|\mathbf{v}_h, q_h|\!|\!|_{h,\varepsilon}} \qquad \forall\, (\boldsymbol{\eta}_h, s_h) \in V_h \times Q_h$$

where

$$\begin{aligned}\mathcal{A}_\varepsilon\left((\boldsymbol{\eta}_h, s_h), (\mathbf{v}_h, q_h)\right) &:= a_\varepsilon(\boldsymbol{\eta}_h, \mathbf{v}_h) + b_0(\mathbf{v}_h, s_h) + b_0(\boldsymbol{\eta}_h, q_h)\\ |\!|\!|\boldsymbol{\eta}_h, s_h|\!|\!|^2 &:= \|\boldsymbol{\eta}_h\|^2_{0,h,\varepsilon} + \|s_h\|^2_{1,h,\varepsilon}.\end{aligned}$$

Hence, for $(\mathbf{u}_h - r_h\mathbf{u}^\varepsilon, p_h - \Pi_h p^\varepsilon)$ there exists $(\mathbf{v}_h, q_h) \in V_h \times Q_h$ such that

$$\|\mathbf{u}_h - r_h\mathbf{u}^\varepsilon\|_{0,h,\varepsilon} + \|p_h - \Pi_h p^\varepsilon\|_{1,h,\varepsilon} \leqslant \sqrt{d}\,|\!|\!|\mathbf{u}_h - r_h\mathbf{u}^\varepsilon, p_h - \Pi_h p^\varepsilon|\!|\!|_{h,\varepsilon} \lesssim \frac{\mathcal{A}_\varepsilon\left((\mathbf{u}_h - r_h\mathbf{u}^\varepsilon, p_h - \Pi_h p^\varepsilon), (\mathbf{v}_h, q_h)\right)}{|\!|\!|\mathbf{v}_h, q_h|\!|\!|_{h,\varepsilon}}.$$

Hence, we have

$$\begin{aligned}\mathcal{A}_\varepsilon((\mathbf{u}_h - r_h\mathbf{u}^\varepsilon, p_h - \Pi_h p^\varepsilon), (\mathbf{v}_h, q_h)) = {}& (\mathbf{u}_h - \mathbf{u}^\varepsilon, \mathbf{v}_h)_{L^2(\Omega)} + (\mathbf{u}^\varepsilon - r_h\mathbf{u}^\varepsilon, \mathbf{v}_h)_{L^2(\Omega)} + \varepsilon^{-1}\langle(\mathbf{u}_h - \mathbf{u}^\varepsilon)\cdot\mathbf{n}, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N}\\ &+ \varepsilon^{-1}\langle(\mathbf{u}^\varepsilon - r_h\mathbf{u}^\varepsilon)\cdot\mathbf{n}, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N} + b_0(\mathbf{v}_h, p_h - p^\varepsilon)\\ &+ b_0(\mathbf{v}_h, p^\varepsilon - \Pi_h p^\varepsilon) + b_0(\mathbf{u}_h - \mathbf{u}^\varepsilon, q_h) + b_0(\mathbf{u}^\varepsilon - r_h\mathbf{u}^\varepsilon, q_h).\end{aligned}$$

The following orthogonality relations hold by definition of r_h and Π_h:

$$\begin{aligned}\varepsilon^{-1}\langle(\mathbf{u}^\varepsilon - r_h\mathbf{u}^\varepsilon)\cdot\mathbf{n}, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N} &= 0 \qquad \forall\, \mathbf{v}_h \in V_h\\ b_0(\mathbf{v}_h, p^\varepsilon - \Pi_h p^\varepsilon) &= 0 \qquad \forall\, \mathbf{v}_h \in V_h\\ b_0(\mathbf{u}^\varepsilon - r_h\mathbf{u}^\varepsilon, q_h) = -\sum_{K\in\mathcal{T}_h}(\nabla q_h, \mathbf{u}^\varepsilon - r_h\mathbf{u}^\varepsilon)_K + \sum_{f\in\mathcal{F}_h^i}\langle[q_h], (\mathbf{u}^\varepsilon - r_h\mathbf{u}^\varepsilon)\cdot\mathbf{n}\rangle_f &\\ + \sum_{f\in\mathcal{F}_h^\partial(\Gamma_N)}\langle q_h, (\mathbf{u}^\varepsilon - r_h\mathbf{u}^\varepsilon)\cdot\mathbf{n}\rangle_f &= 0 \qquad \forall\, q_h \in Q_h.\end{aligned}$$

Moreover, by consistency, we have

$$\begin{aligned}(\mathbf{u}_h - \mathbf{u}^\varepsilon, \mathbf{v}_h)_{L^2(\Omega)} + \varepsilon\langle(\mathbf{u}_h - \mathbf{u}^\varepsilon)\cdot\mathbf{n}, \mathbf{v}_h\cdot\mathbf{n}\rangle_{\Gamma_N} + b_0(\mathbf{v}_h, p_h - p^\varepsilon) &= 0 \qquad \forall\, \mathbf{v}_h \in V_h\\ b_0(\mathbf{u}_h - \mathbf{u}^\varepsilon, q_h) &= 0 \qquad \forall\, q_h \in Q_h.\end{aligned}$$

Hence,

$$\mathcal{A}_\varepsilon\left((\mathbf{u}_h - r_h\mathbf{u}^\varepsilon, p_h - \Pi_h p^\varepsilon), (\boldsymbol{\tau}_h, q_h)\right) = (\mathbf{u}^\varepsilon - r_h\mathbf{u}^\varepsilon, \mathbf{v}_h)_{L^2(\Omega)}$$

and we can write

$$\|\mathbf{u}_h - r_h\mathbf{u}^\varepsilon\|_{0,h,\varepsilon} + \|p_h - \Pi_h p^\varepsilon\|_{1,h,\varepsilon} \lesssim \frac{\|\mathbf{u}^\varepsilon - r_h\mathbf{u}^\varepsilon\|_{L^2(\Omega)}\,|\!|\!|\mathbf{v}_h, 0|\!|\!|_{h,\varepsilon}}{|\!|\!|\mathbf{v}_h, q_h|\!|\!|_{h,\varepsilon}} \lesssim \|\mathbf{u}^\varepsilon - r_h\mathbf{u}^\varepsilon\|_{L^2(\Omega)}. \qquad \square$$

Proposition 6.1. Let $(\mathbf{u}, p) \in \boldsymbol{H}^{r+1}(\Omega) \times H^{t+1}(\Omega)$ be the solution of (2.1) and $(\mathbf{u}_h, p_h) \in V_h \times Q_h$ the one to (4.1) with $m = 1$. There exists $C > 0$ such that, for $s = \min\{r, t, k\}$,

$$\|\mathbf{u}_h - r_h\mathbf{u}\|_{0,h} + \|p_h - \Pi_h p\|_{1,h} \leqslant Ch^{s+1}\left(\|\mathbf{u}\|_{H^{r+1}(\Omega)} + \|p\|_{H^{t+1}(\Omega)}\right). \tag{6.3}$$

Proof. By Lemma 6.1, a multiplicative trace inequality for Sobolev functions and standard approximation results for the L^2-projection

$$\begin{aligned}\|\mathbf{u}_h - r_h\mathbf{u}\|_{0,h} + \|p_h - \Pi_h p\|_{1,h} &\lesssim \|\mathbf{u} - r_h\mathbf{u}\|_{L^2(\Omega)} + \sum_{f\in\mathcal{F}_h^\partial(\Gamma_N)} h^{1/2}\,\|p - \Pi_h p\|_{L^2(f)}\\ &\lesssim \|\mathbf{u} - r_h\mathbf{u}\|_{L^2(\Omega)} + \sum_{K\in\mathcal{T}_h} h^{1/2}\,\|p - \Pi_h p\|^{1/2}_{L^2(K)}\,\|\nabla(p - \Pi_h p)\|^{1/2}_{L^2(K)}\\ &\lesssim \|\mathbf{u} - r_h\mathbf{u}\|_{L^2(\Omega)} + h^{1/2}h^{(t+1)/2}\,\|p\|^{1/2}_{H^{t+1}(\Omega)}\,h^{t/2}\,\|p\|^{1/2}_{H^{t+1}(\Omega)}\\ &= \|\mathbf{u} - r_h\mathbf{u}\|_{L^2(\Omega)} + h^{t+1}\,\|p\|_{H^{t+1}(\Omega)}\,.\end{aligned}$$

By using Bramble–Hilbert/Deny–Lions Lemma [16], we get

$$\|\mathbf{u}_h - r_h\mathbf{u}\|_{0,h} + \|p_h - \Pi_h p\|_{1,h} \lesssim h^{r+1}\,\|\mathbf{u}\|_{H^{r+1}(\Omega)} + h^{t+1}\,\|p\|_{H^{t+1}(\Omega)}$$

with $0 \leqslant t \leqslant k$ and $0 \leqslant r \leqslant k$. □

Proposition 6.2. Let $(\mathbf{u}^\varepsilon, p^\varepsilon) \in \boldsymbol{H}^{r+1}(\Omega) \times H^{t+1}(\Omega)$ be the solution of the perturbed continuous problem (3.1) and $(\mathbf{u}_h, p_h) \in V_h \times Q_h$ the one to (4.2). There exists $C > 0$ such that, for $s = \min\{r, t, k\}$,

$$\left\|\mathbf{u}_h - r_h\mathbf{u}^\varepsilon\right\|_{0,h,\varepsilon} + \left\|p_h - \Pi_h p^\varepsilon\right\|_{1,h,\varepsilon} \leqslant Ch^{s+1}\,\left\|\mathbf{u}^\varepsilon\right\|_{H^{s+1}(\Omega)}\,. \tag{6.4}$$

Proof. By Lemma 6.2 and Bramble–Hilbert/Deny–Lions Lemma [16], we get

$$\left\|\mathbf{u}_h - r_h\mathbf{u}^\varepsilon\right\|_{0,h,\varepsilon} + \left\|p_h - \Pi_h p^\varepsilon\right\|_{1,h,\varepsilon} \lesssim h^{s+1}\,\left\|\mathbf{u}^\varepsilon\right\|_{H^{s+1}(\Omega)}\,.$$ □

Remark 6.1. Let us remark that the quantities $\|\mathbf{u}_h - r_h\mathbf{u}_h\|_{0,h}$, $\|p_h - \Pi_h p\|_{1,h}$ in (6.3) and $\left\|p_h - \Pi_h p^\varepsilon\right\|_{1,h,\varepsilon}$, $\left\|\mathbf{u}_h - r_h\mathbf{u}^\varepsilon\right\|_{0,h,\varepsilon}$ in (6.4), respectively, are super convergent.

Theorem 6.1. *Let $(\mathbf{u}, p) \in \boldsymbol{H}^{r+1}(\Omega) \times H^{t+1}(\Omega)$ be the solution to* (2.1) *and $(\mathbf{u}_h, p_h) \in V_h \times Q_h$ the one to* (4.1) *with $m = 1$. Then there exists $C > 0$ such that, for $s = \min\{r, t, k\}$,*

$$\|\mathbf{u} - \mathbf{u}_h\|_{L^2(\Omega)} \leqslant Ch^{s+1}\left(\|\mathbf{u}\|_{H^{r+1}(\Omega)} + \|p\|_{H^{t+1}(\Omega)}\right).$$

Proof. Let us proceed by triangular inequality:

$$\|\mathbf{u} - \mathbf{u}_h\|_{L^2(\Omega)} \leqslant \|\mathbf{u} - r_h\mathbf{u}\|_{L^2(\Omega)} + \|r_h\mathbf{u} - \mathbf{u}_h\|_{L^2(\Omega)}\,.$$

The first and the second terms in the right-hand side scale as $\mathcal{O}(h^{r+1})$ and $\mathcal{O}(h^{s+1})$, respectively, because of Bramble–Hilbert/Deny–Lions Lemma [16] and Proposition 6.1. □

Lemma 6.3. *Let $(\mathbf{u}^\varepsilon, p^\varepsilon) \in \boldsymbol{H}^{r+1}(\Omega) \times H^{t+1}(\Omega)$ be the solution to the perturbed continuous problem* (3.1) *and $(\mathbf{u}_h, p_h) \in V_h \times Q_h$ the one to* (4.2). *Then there exists $C > 0$ such that, for $s = \min\{r, k\}$,*

$$\left\|\mathbf{u}^\varepsilon - \mathbf{u}_h\right\|_{L^2(\Omega)} \leqslant Ch^{s+1}\,\left\|\mathbf{u}^\varepsilon\right\|_{H^{s+1}(\Omega)}\,.$$

Proof. Let us proceed by triangular inequality:

$$\left\|\mathbf{u}^\varepsilon - \mathbf{u}_h\right\|_{L^2(\Omega)} \leqslant \left\|\mathbf{u}^\varepsilon - r_h\mathbf{u}^\varepsilon\right\|_{L^2(\Omega)} + \left\|r_h\mathbf{u}^\varepsilon - \mathbf{u}_h\right\|_{L^2(\Omega)}\,.$$

The first and the second terms in the right-hand side scale as $\mathcal{O}(h^{s+1})$, respectively, because of Bramble–Hilbert/Deny–Lions Lemma [16] and Proposition 6.2. □

Theorem 6.2. *Let $(\mathbf{u}, p) \in \boldsymbol{H}^2(\Omega) \times H^{t+1}(\Omega)$ be the solution to the continuous* (2.1) *and $(\mathbf{u}_h, p_h) \in V_h \times Q_h$ the one to* (4.2) *with $\varepsilon = h$. Assume Ω to be a convex domain with a Lipschitz polygonal boundary Γ, $\mathbf{f} \in \boldsymbol{H}^1(\Omega)$, $\mathbf{f}\cdot\mathbf{n} = 0$ on Γ_N and $u_N = 0$. Then, there exists $C > 0$ such that*

$$\|\mathbf{u} - \mathbf{u}_h\|_{L^2(\Omega)} \leqslant Ch\left(\left\|\mathbf{f}\right\|_{H^1(\Omega)} + \|g\|_{L^2(\Omega)} + \|p_D\|_{H^{1/2}(\Gamma_D)}\right).$$

Proof. Let us proceed by triangular inequality:

$$\|\mathbf{u}-\mathbf{u}_h\|_{L^2(\Omega)} \leqslant \left\|\mathbf{u}-\mathbf{u}^\varepsilon\right\|_{L^2(\Omega)} + \left\|\mathbf{u}^\varepsilon-\mathbf{u}_h\right\|_{L^2(\Omega)} \lesssim \varepsilon\left(\|\operatorname{div}\mathbf{f}\|_{L^2(\Omega)} + \|g\|_{L^2(\Omega)} + \|p_D\|_{H^{1/2}(\Gamma_D)}\right) + h\left\|\mathbf{u}^\varepsilon\right\|_{H^1(\Omega)}$$
$$\lesssim \varepsilon\left(\|\operatorname{div}\mathbf{f}\|_{L^2(\Omega)} + \|g\|_{L^2(\Omega)} + \|p_D\|_{H^{1/2}(\Gamma_D)}\right) + h\left(\|\mathbf{f}\|_{H^1(\Omega)} + \|g\|_{L^2(\Omega)}\right).$$

We used Lemma 6.3, Proposition 3.1, and Proposition 3.2. Finally, let us choose $\varepsilon = h$. □

Remark 6.2. We observe that for both formulations, (4.1) and (4.2), all dimensionless parameters have been set for simplicity to 1, unlike for the standard Nitsche method for the Poisson problem [17], where the dimensionless parameter needs to be taken large enough.

7 Numerical examples

7.1 Convergence results

In this first set of numerical examples we verify that the optimal a priori error estimates of Theorems 6.1, 6.2. We also check that the result of Theorem 6.1 holds in the non-symmetric case $m = 0$, as already mentioned in Section 4. Moreover, we study the L^2-error of the pressure field, for which optimal convergence is observed in general and super convergence in the case of the lowest order Raviart–Thomas element and triangular meshes. Although Theorem 6.2 guarantees us optimal a priori error estimates for the discretization (4.2) only with the lowest order Raviart–Thomas element, numerical results show that we have optimal convergence rates also for higher orders.

7.1.1 Unit square with triangular meshes

We approximate the Darcy problem in the *unit square* $\Omega = (0,1)^2$ using a family of triangular meshes, with weakly enforced Neumann boundary conditions on the whole boundary, using as manufactured solutions

$$\mathbf{u}_{\text{ex}} = \begin{pmatrix} x\sin(x)\sin(y) \\ \sin(x)\cos(y) + x\cos(x)\cos(y) \end{pmatrix}, \quad p_{\text{ex}} = x^3y - 0.125.$$

Note that $\mathbf{u}_{\text{ex}}$ is divergence-free. The numerical results are presented in Fig. 1.

7.1.2 Unit circle with triangular meshes

Now, we consider the *unit circle* $\Omega = \{(x,y) \in R^2 : x^2 + y^2 \leqslant 1\}$ which is meshed using triangles. We weakly impose the essential boundary conditions on the boundary and consider the following reference solutions:

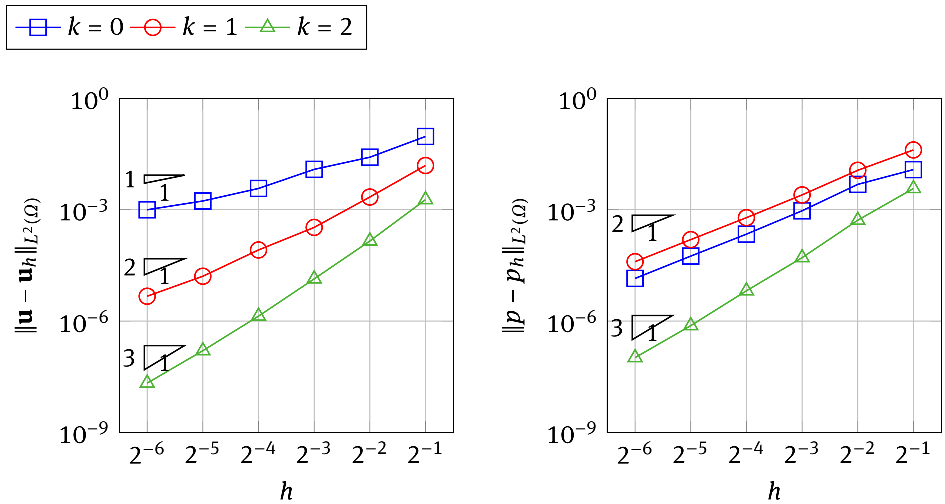
$$\mathbf{u}_{\text{ex}} = \begin{pmatrix} \frac{1}{10}\mathrm{e}^x\sin(xy) \\ x^4 + y^2 \end{pmatrix}, \quad p_{\text{ex}} = x^3\cos(x) + y^2\sin(x).$$

This time $\operatorname{div}\mathbf{u}_{\text{ex}} = 2y + \frac{1}{10}\left(\mathrm{e}^x\sin(xy) + y\mathrm{e}^x\cos(xy)\right)$. The numerical results are shown in Fig. 2.

7.1.3 Unit square with quadrilateral meshes

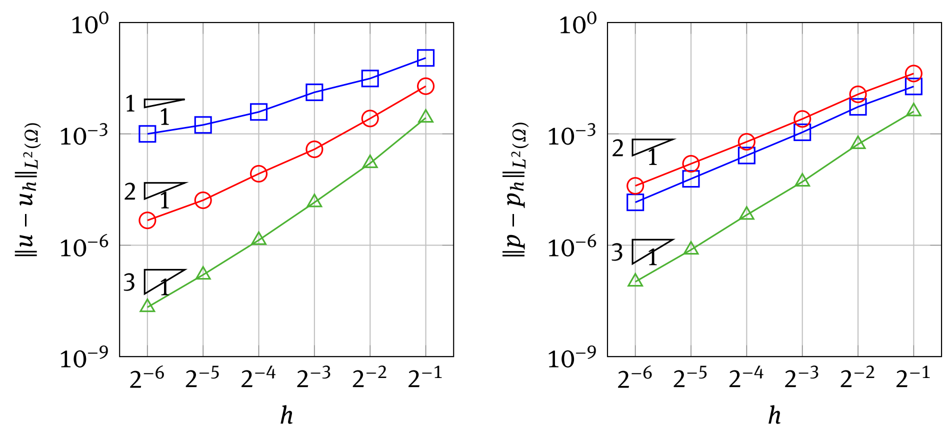
Let us consider the *unit square* $\Omega = (0,1)^2$ meshed using quadrilaterals. We impose natural boundary conditions on $\{(x,y) : 0 \leqslant x \leqslant 1, y = 0\}$ and essential boundary conditions everywhere else in a weak sense. The reference solutions are:

$$\mathbf{u}_{\text{ex}} = \begin{pmatrix} \cos(x)\cosh(y) \\ \sin(x)\cosh(y)) \end{pmatrix}, \quad p_{\text{ex}} = -\sin(x)\sinh(y) - (\cos(1)-1)(\cosh(1)-1).$$



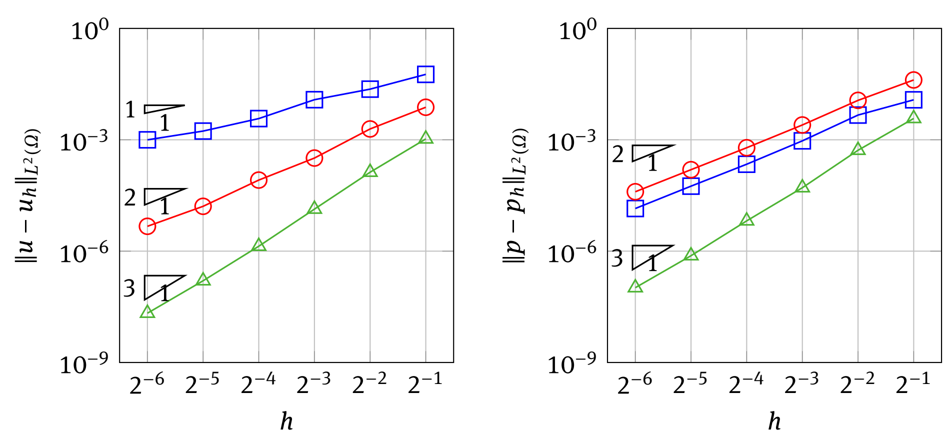
(a) Velocity error using (4.1) with $m = 1$.

(b) Pressure error using (4.1) with $m = 1$.



(c) Velocity error using (4.1) with $m = 0$.

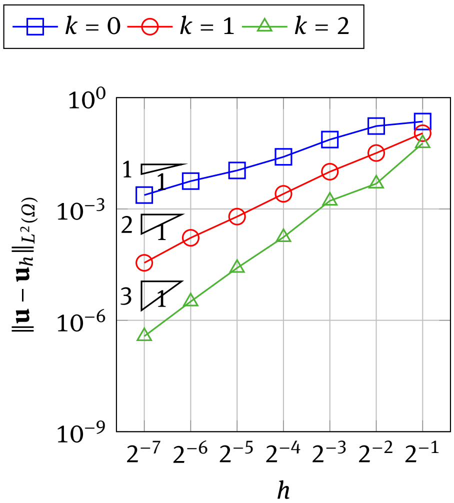
(d) Pressure error using (4.1) with $m = 0$.



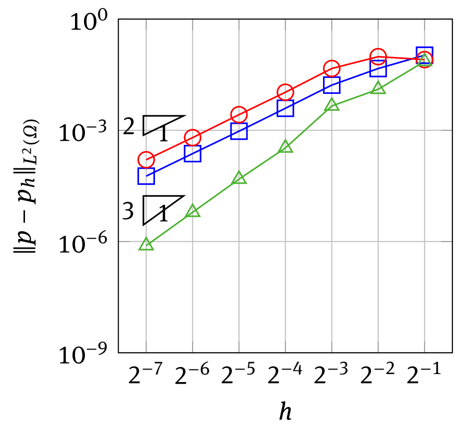
(e) Velocity error using (4.2).

(f) Pressure error using (4.2).

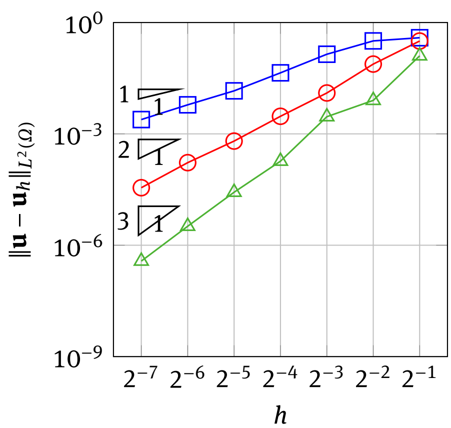
Fig. 1: Convergence rates in the *unit square* with triangular meshes.



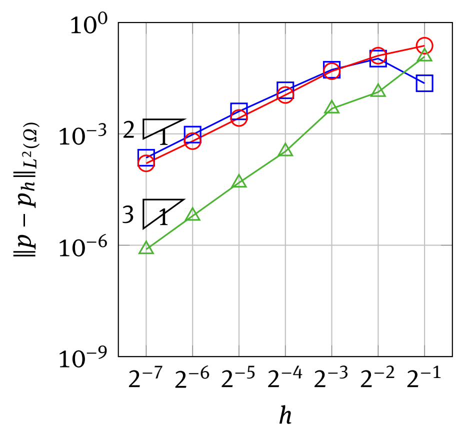
(a) Velocity error using (4.1) with $m = 1$.



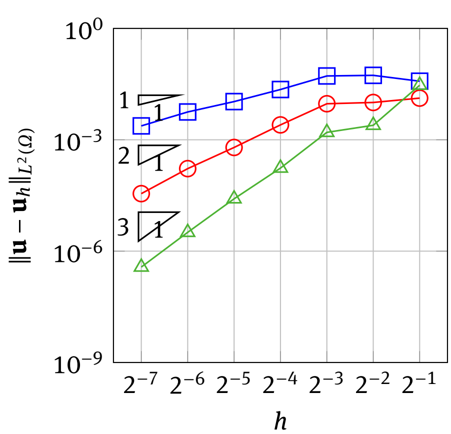
(b) Pressure error using (4.1) with $m = 1$.



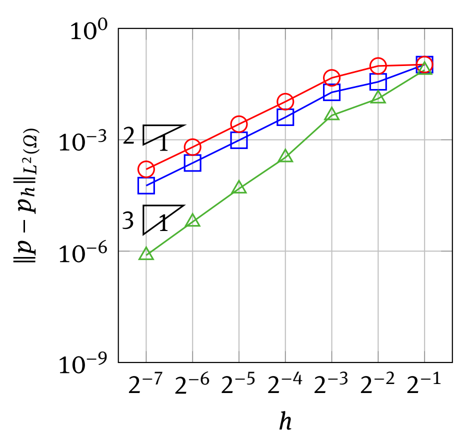
(c) Velocity error using (4.1) with $m = 0$.



(d) Pressure error using (4.1) with $m = 0$.

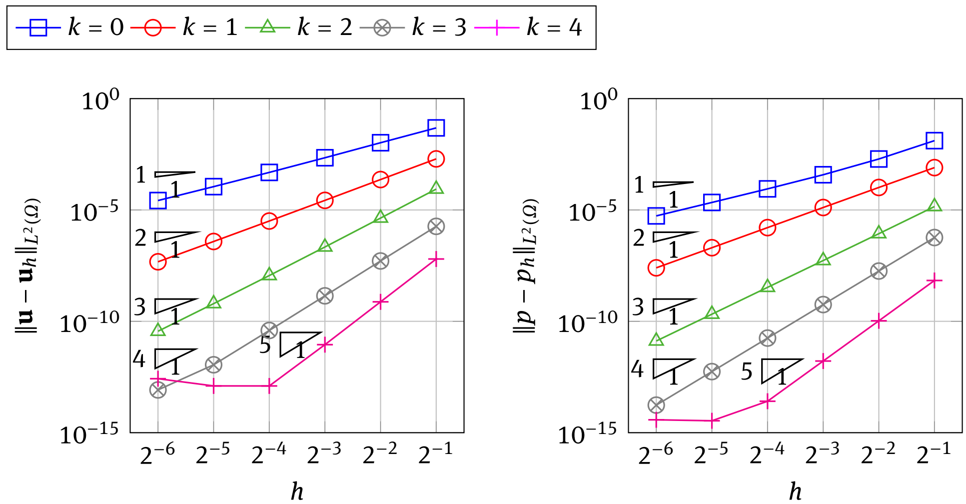


(e) Velocity error using (4.2).



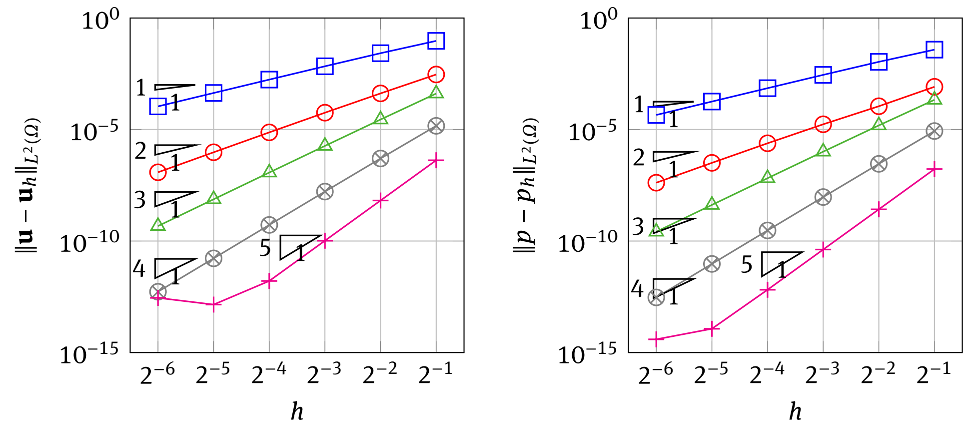
(f) Pressure error using (4.2).

Fig. 2: Convergence rates in the *unit circle* with triangular meshes.



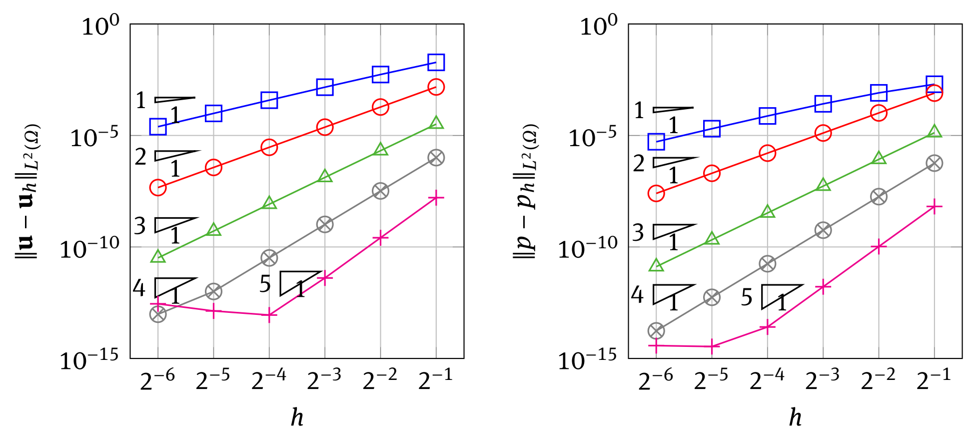
(a) Velocity error using (4.1) with $m = 1$.

(b) Pressure error using (4.1) with $m = 1$.



(c) Velocity error using (4.1) with $m = 0$.

(d) Pressure error using (4.1) with $m = 0$.

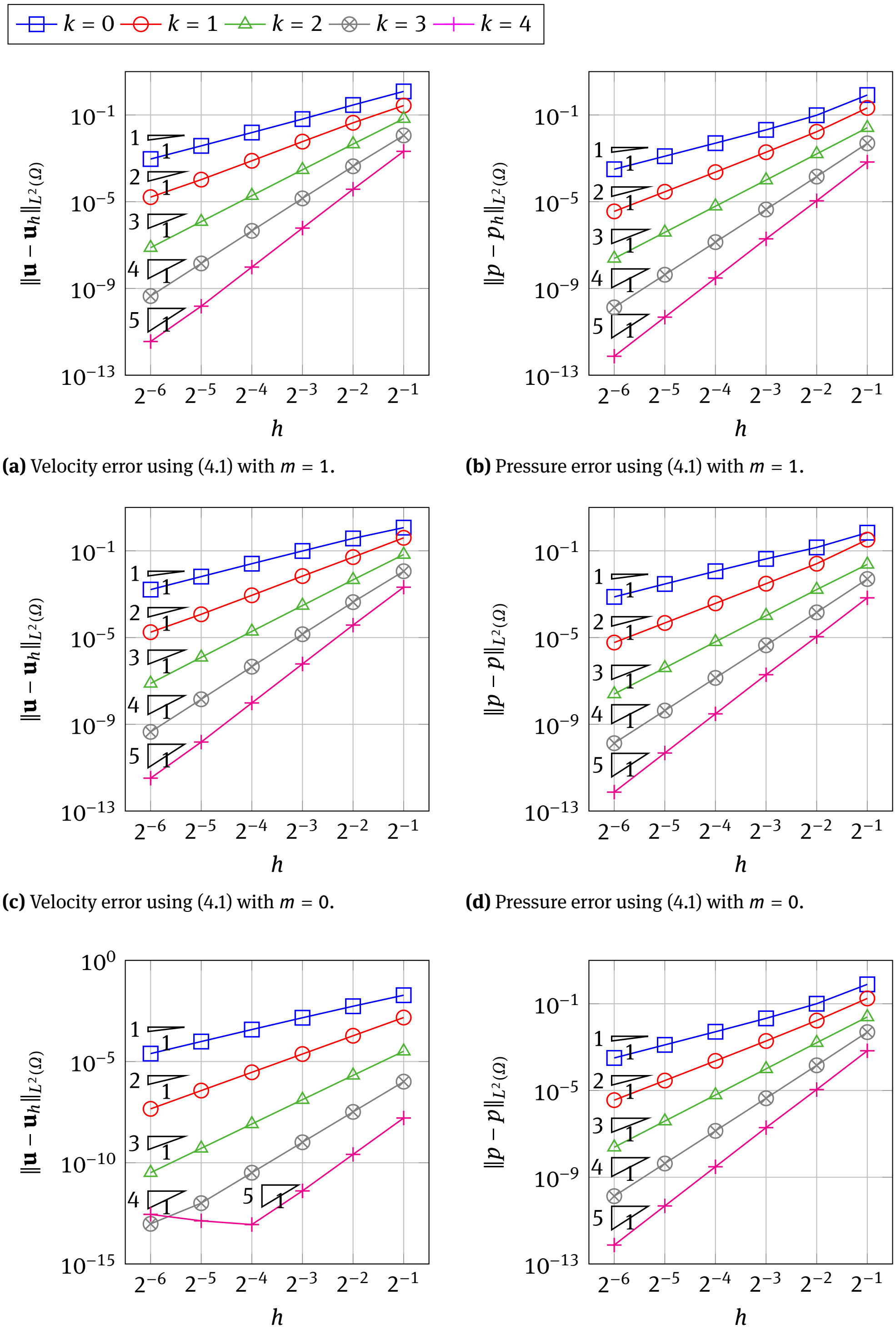


(e) Velocity error using (4.2).

(f) Pressure error using (4.2).

Fig. 3: Convergence rates in the *unit square* with quadrilateral meshes.

We have $\operatorname{div} \mathbf{u}_{\mathrm{ex}} = 0$. For the numerical results we refer to Fig. 3.



(a) Velocity error using (4.1) with $m = 1$.

(b) Pressure error using (4.1) with $m = 1$.

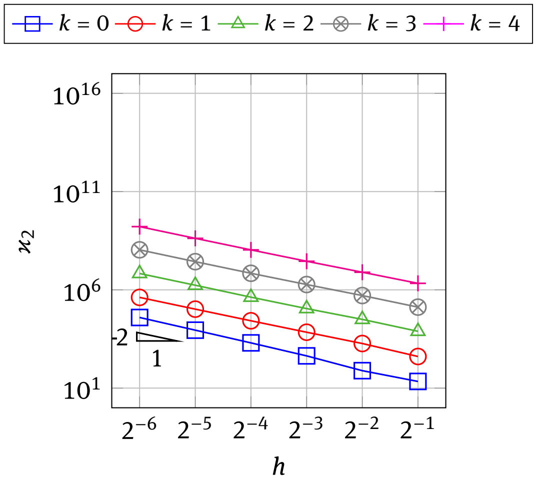
(c) Velocity error using (4.1) with $m = 0$.

(d) Pressure error using (4.1) with $m = 0$.

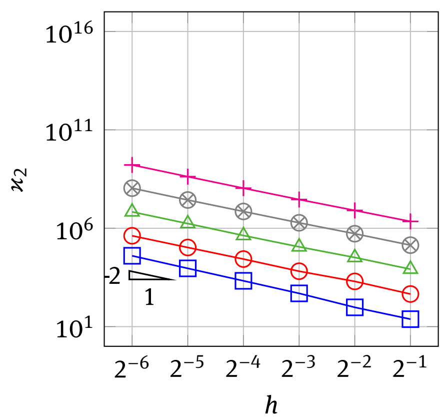
(e) Velocity error using (4.2).

(f) Pressure error using (4.2).

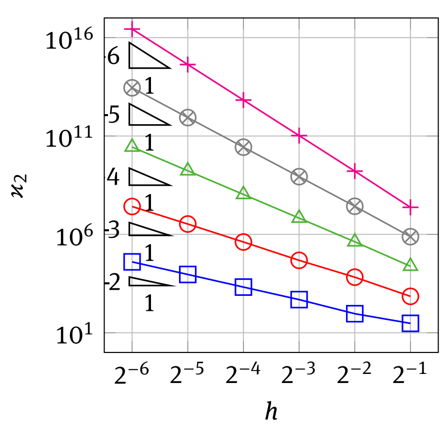
Fig. 4: A priori errors in the *quarter of annulus* with isoparametric quadrilateral meshes.



(a) Condition number using (4.1) with $m = 1$.

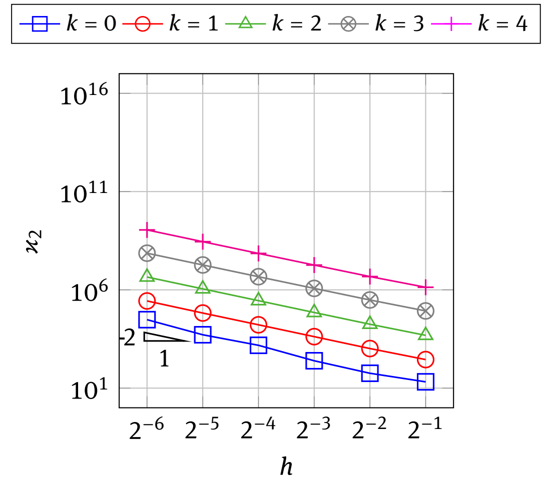


(b) Condition number using (4.1) with $m = 0$.

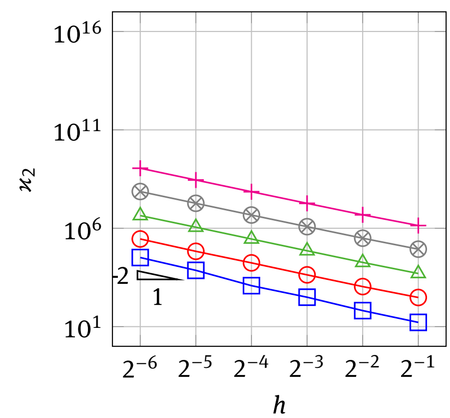


(c) Condition number using (4.2).

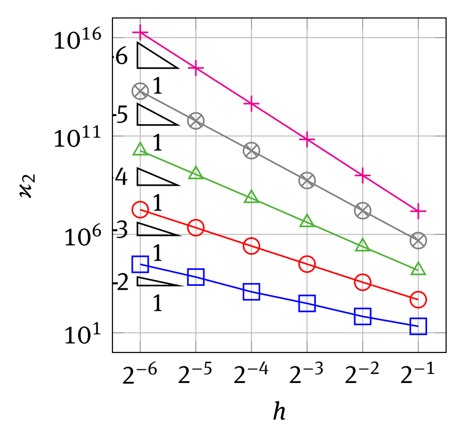
Fig. 5: Condition numbers in the *unit square* with quadrilateral meshes.



(a) Condition number using (4.1) with $m = 1$.

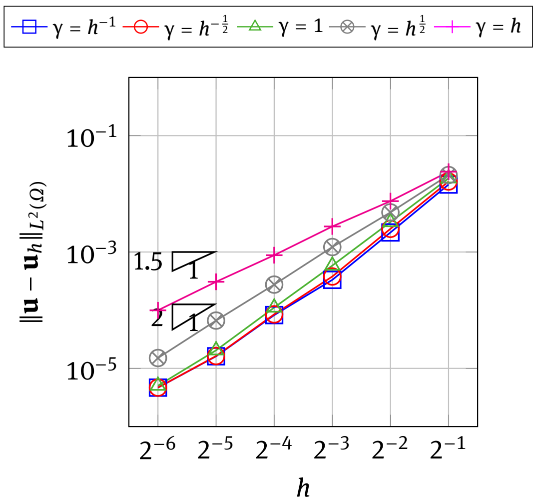


(b) Condition number using (4.1) with $m = 0$.

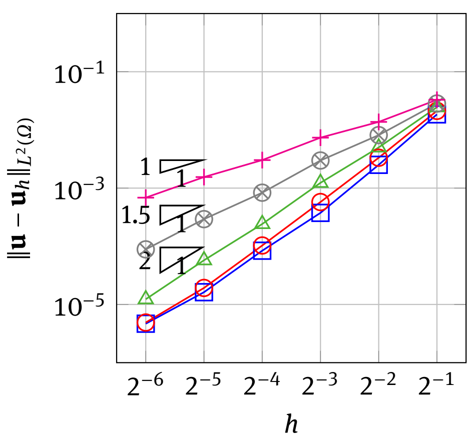


(c) Condition number using (4.2).

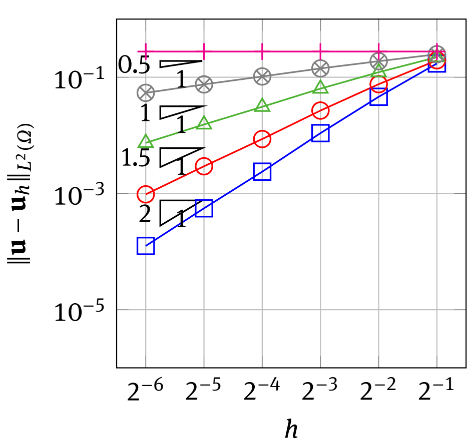
Fig. 6: Condition numbers in the *quarter of annulus* with isoparametric quadrilateral elements.



(a) Velocity error using (4.1) with $m = 1$.

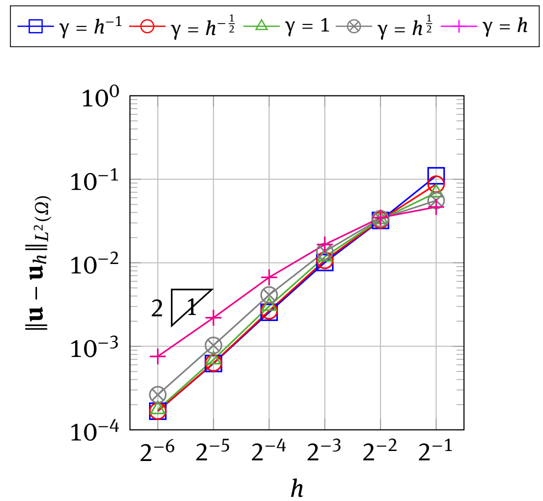


(b) Velocity error using (4.1) with $m = 0$.

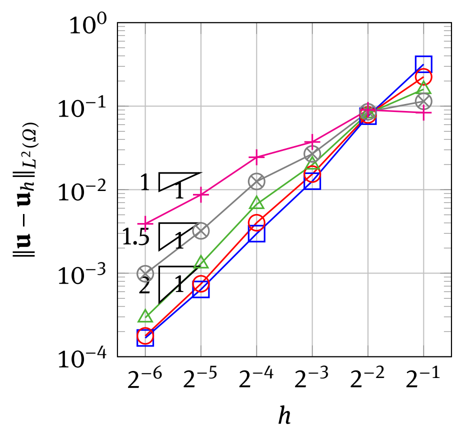


(c) Velocity error using (4.2).

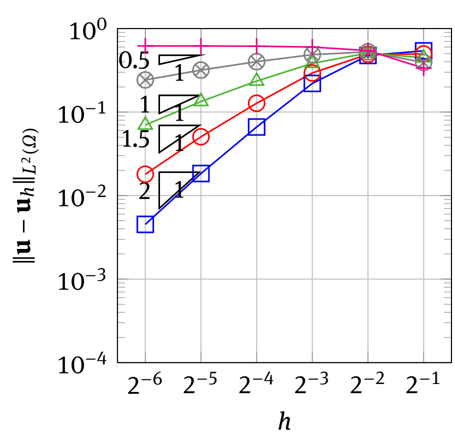
Fig. 7: Compare L^2-errors for the velocity in the *unit square* with respect to different values of the penalty parameter γ with triangular meshes.



(a) Velocity error using (4.1) with $m = 1$.

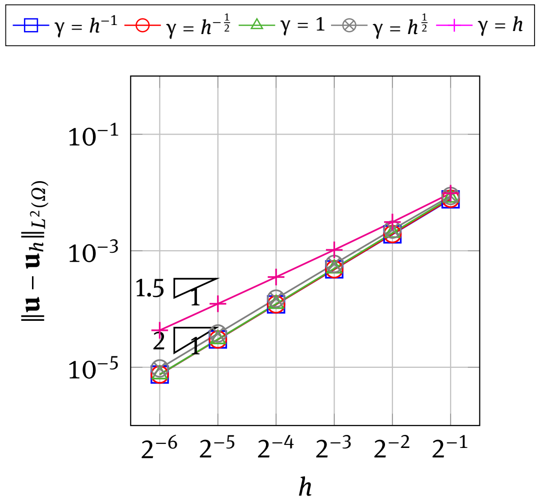


(b) Velocity error using (4.1) with $m = 0$.

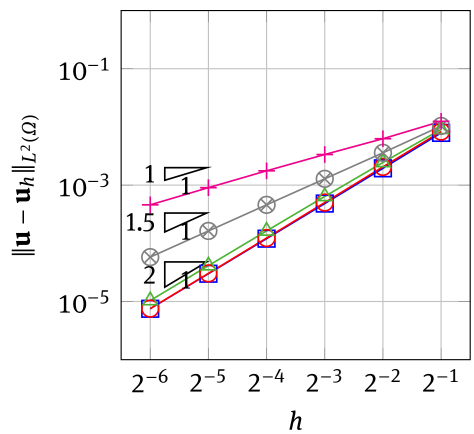


(c) Velocity error using (4.2).

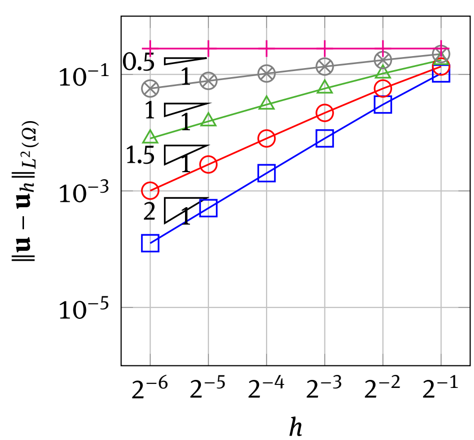
Fig. 8: Compare L^2-errors for the velocity in the *unit circle* with respect to different values of the penalty parameter γ with triangular meshes.



(a) Velocity error using (4.1) with $m = 1$.

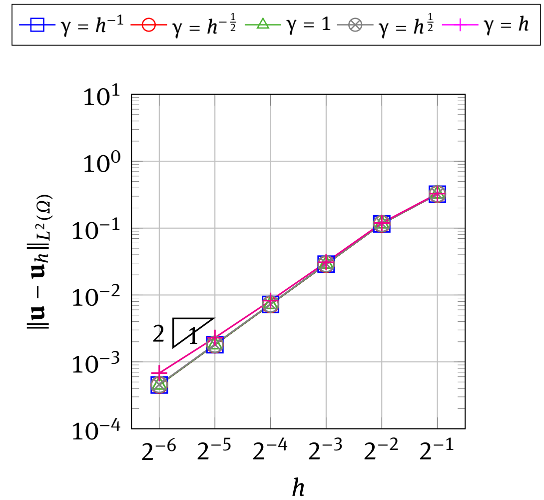


(b) Velocity error using (4.1) with $m = 0$.

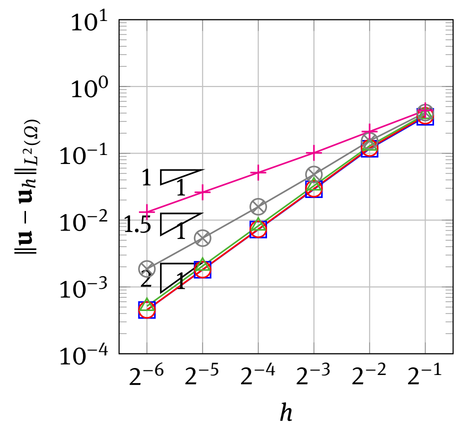


(c) Velocity error using (4.2).

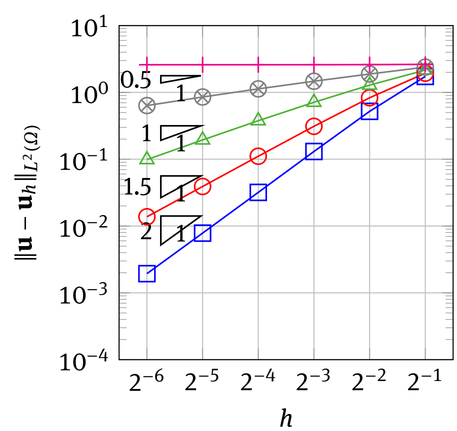
Fig. 9: Compare L^2-errors for the velocity in the *unit square* with respect to different values of the penalty parameter γ with quadrilateral meshes.



(a) Velocity error using (4.1) with $m = 1$.



(b) Velocity error using (4.1) with $m = 0$.



(c) Velocity error using (4.2).

Fig. 10: Compare L^2-errors for the velocity in the *quarter of annulus* with respect to different values of the penalty parameter γ with isoparametric quadrilateral elements.

7.1.4 Quarter of annulus with quadrilateral isoparametric elements

Let us consider the *quarter of annulus* centered in the origin with inner and outer radii, respectively, $r = 1$ and $R = 2$, discretized using quadrilateral isoparametric elements [12]. We impose natural boundary conditions on the straight edges $\{(x, y) : 1 \leqslant x \leqslant 2, y = 0\}$ and $\{(x, y) : x = 0, 1 \leqslant y \leqslant 2\}$ and weak essential boundary conditions on the curved ones. The manufactured solutions are:

$$\mathbf{u}_{\text{ex}} = \begin{pmatrix} -xy^2 \\ -x^2y - \frac{3}{2}y^2 \end{pmatrix}, \quad p_{\text{ex}} = \frac{1}{2}\left(x^2y^2 + y^3\right)$$

with $\operatorname{div} \mathbf{u}_{\text{ex}} = -x^2 - y^2 - 3y$. The numerical results are presented in Fig. 4.

7.2 A remark about the condition numbers

Proceeding as in [10] it would be possible to prove that the ℓ^2-condition number of the stiffness matrix arising from the discretizations (4.1), for both $m \in \{0, 1\}$, scales as h^{-2}, as Figures 5a, 5b, 6a, and 6b confirm. The penalty parameter for the weak imposition of the Neumann boundary conditions is the responsible of the deterioration of the conditioning with respect to the standard mixed finite element discretization of the Poisson problem, for which the condition number scales as h^{-1}. An even worse situation occurs when formulation (4.2) is employed. In this case the condition number scales as $h^{-(s+2)}$, $s = \min\{r, k\}$, r being the Sobolev regularity of the exact solution for the pressure field and k the polynomial degree of the Raviart–Thomas discretization, as confirmed by Figs. 5c and 6c.

7.3 The optimality of the penalty parameter

We want to analyze the optimality of the penalty parameter, denoted through this section as γ, for both numerical schemes. Let us observe indeed that in order for formulations (4.1) and (4.2) to be extend in the unfitted case and provide an optimal convergence scheme, we would expect γ to scale as $\mathcal{O}(h)$. Let us consider the Raviart–Thomas element of order $k = 1$ and compare the numerical results for the L^2-error of the velocity field with respect to different powers of the mesh-size as penalty parameter. The first set of numerical experiences is performed using triangular meshes, then we move to quadrilaterals.

To obtain Fig. 7 the same setting of Section 7.1.1 is employed. Then, in Fig. 8, we move to the configuration of Section 7.1.2. Finally, in Figs. 9 and 10 we use, respectively, the settings of Sections 7.1.3 and 7.1.4.

In all numerical experiments, we do not detect any particular sensitivity of the convergence of the error of the velocities with respect to γ in the case of method (4.1), see, e.g., Fig. 8a. On the other hand, for the formulation (4.2) strong influence of varying γ is clearly seen in Figs. 7c, 8c, 9c, and 10c. Let us observe indeed that for formulations (4.1) and (4.2) to be extended in the unfitted case and provide an optimal convergence scheme, we would expect γ to scale as $\mathcal{O}(h)$. In this sense, method (4.1) seems a more promising approach.

Funding: Erik Burman was partially supported by the EPSRC grants EP/P01576X/1 and EP/T033126/1. Riccardo Puppi was partially supported by ERC AdG project CHANGE No. 694515.

References

[1] R. Adams and J. Fournier, *Sobolev Spaces, 2nd ed.*, Pure and Applied Mathematics, Academic Press, New York, 2003.
[2] D. N. Arnold, D. Boffi, and R. S. Falk, Quadrilateral H(div) finite elements, *SIAM J. Numer. Anal.* **42** (2004), No. 6, 2429–2451.
[3] I. Babuška, The finite element method with Lagrangian multipliers, *Numer. Math.* **20** (1973), No. 3, 179–192.
[4] I. Babuška, The finite element method with penalty, *Math. Comput.* **27** (1973), No. 122, 221–228.

[5] H. J. C. Barbosa and T. J. R. Hughes, The finite element method with Lagrange multipliers on the boundary: circumventing the Babuška–Brezzi condition, *Comp. Methods Appl. Mech. Engrg.* **85** (1991), No. 1, 109 – 128.

[6] D. Boffi, F. Brezzi, and M. Fortin, *Mixed Finite Element Methods and Applications*, Springer Series in Computational Mathematics, 2013.

[7] E. Burman, S. Claus, P. Hansbo, M. G. Larson, and A. Massing, CutFEM: Discretizing geometry and partial differential equations, *Int. J. Numer. Meth. Engrg.* **104** (2015), No. 7, 472–501.

[8] E. Burman, M. G. Larson, and L. Oksanen, Primal–dual mixed finite element methods for the elliptic Cauchy problem, *SIAM J. Numer. Anal.* **56** (2018), No. 6, 3480–3509.

[9] C. D'Angelo and A. Scotti, A mixed finite element method for Darcy flow in fractured porous media with non-matching grids, *ESAIM: Math. Modelling Numer. Anal.* **46** (2011), No. 2, 465–489.

[10] A. Ern and J.-L. Guermond, Evaluation of the condition number in linear systems arising in finite element approximations, *ESAIM: Math. Modelling Numer. Anal.* **40** (2006), No. 1, 29–48.

[11] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Society for Industrial and Applied Mathematics, 2011.

[12] C. Johnson, *Numerical Solution of Partial Differential Equations by the Finite Element Method*, Cambridge Cambridge University Press, 1987.

[13] J. Könnö, D. Schötzau, and R. Stenberg, Mixed Finite Element Methods for Problmes with Robin Boundary Conditions, *SIAM J. Numer. Anal.* **49** (2011), No. 1/2, 285–308.

[14] P. Monk, *Finite Element Metods for Maxwell's Equations*, Oxford University Press, 2003.

[15] J. Nitsche, Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind, *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg* **36** (1971), No. 1, 9–15.

[16] A. Quarteroni and A. Valli, *Numerical Approximation of Partial Differential Equations*, Springer, Berlin–Heidelberg, 1994.

[17] R. Stenberg, On some techniques for approximating boundary conditions in the finite element method, *J. Comp. Appl. Math.* **63** (1995), No. 1, 139–148.