ϕ-FEM, a finite element method on domains defined by level-sets: the Neumann boundary case

Michel Duprez, Vanessa Lleras, Alexei Lozinski

ϕ-FEM, a finite element method on domains defined by level-sets: the Neumann boundary case

Michel Duprez *and Vanessa Lleras †and Alexei Lozinski ‡

March 26, 2020

Abstract

We extend a fictitious domain-type finite element method, called ϕ-FEM and introduced in [7], to the case of Neumann boundary conditions. The method is based on a multiplication by the level-set function and does not require a boundary fitted mesh. Unlike other recent fictitious domain-type methods (XFEM, CutFEM), our approach does not need any non-standard numerical integration on cut mesh elements or on the actual boundary. We prove the optimal convergence of ϕ-FEM and the fact that the discrete problem is well conditioned inependently of the mesh cuts. The numerical experiments confirm the theoretical results.

1 Introduction

We consider a second order elliptic partial differential equation with Neumann boundary conditions

$$\begin{cases} -\Delta u + u = f & \text{in } \Omega, \\ \frac{\partial u}{\partial n} = g & \text{on } \Gamma \end{cases} \tag{1}$$

in a bounded domain $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ with smooth boundary Γ assuming that Ω and Γ are given by a level-set function ϕ:

$$\Omega := \{\phi < 0\} \text{ and } \Gamma := \{\phi = 0\}. \tag{2}$$

Such a representation is a popular and useful tool to deal with problems with evolving surfaces or interfaces [13]. In the present article, the level-set function is supposed known on $\mathbb{R}^d$, smooth, and to behave near Γ similar to the signed distance to Γ.

Our goal is to develop a finite element method for (1) using a mesh which is fitted to Γ, i.e. we allow the boundary Γ to cut the mesh cells in an arbitrary manner. Unlike the existing finite elements methods on non-matching meshes, such as XFEM [12, 11, 14, 8] or CutFEM [4, 5, 6, 3], we want to avoid the cumbersome integrals over the actual boundary Γ or on the parts of mesh cells cut by Γ. In the same time, we want to maintain the optimal rates of convergence, contrary to the classical fictitious domain/penalty methods.

In the recent article [7], we have proposed such a method for the Poisson problem with homogeneous Dirichlet boundary conditions $u = 0$ on Γ. The idea behind this method, baptised ϕ-FEM, was to put $u = \phi w$ where $\phi = 0$ on Γ by its definition and w is the new unknown. We then replace ϕ and w by the finite element approximations ϕ_h and w_h, substitute $u \approx \phi_h w_h$ into the appropriate variational formulation

*CEREMADE, Université Paris-Dauphine & CNRS UMR 7534, Université PSL, 75016 Paris, France. mduprez@math.cnrs.fr

†IMAG, Univ Montpellier, CNRS, Montpellier, France. vanessa.lleras@umontpellier.fr

‡Laboratoire de Mathématiques de Besançon, UMR CNRS 6623, Université Bourgogne Franche-Comté, 16, route de Gray, 25030 Besançon Cedex, France. alexei.lozinski@univ-fcomte.fr

of the problem and get an easily implementable finite element discretization in terms of the new unknown w_h. The boundary conditions are thus automatically taken into account thanks to the multiplication by ϕ (or, rather, its approximation ϕ_h). Such a simple idea cannot be used directly to discretize the Neumann boundary conditions as in (1). In the present article, we show that the announced above goals can still be achieved introducing some additional unknowns and thus making the construction slightly more complex.

We start by recalling that the outward-looking unit normal is given on Γ by

$$n = \frac{1}{|\nabla\phi|}\nabla\phi\,.$$

We then note that the boundary condition in (1) is automatically satisfied if we postulate

$$\nabla u \cdot \frac{1}{|\nabla\phi|}\nabla\phi = \frac{1}{h|\nabla\phi|}p\phi + \tilde{g}, \tag{3}$$

where $\tilde{g}$ is some sufficiently smooth lifting of g from Γ to a vicinity of Γ, and p is an auxiliary unknown. The equation above will be enforced on a strip around Γ, called Ω_h^Γ, consisting of the elements of the computational mesh cut by Γ. The prefactor $1/h|\nabla\phi|$ is introduced in front of p in eq. (3) in order to simplify some of the forthecoming formulas and to control the conditionning number of the resulting finite element method (h stands for the mesh size and we shall see that introduction of h here provides a correct scaling for the auxiliar variable p). In fact, in order to achieve the optimal accuracy of the finite element discretization, we shall have to introduce yet another vector-valued additional unknown y on Ω_h^Γ, setting $y = -\nabla u$. Assuming that both f and u can be extended slightly outside Ω (to $\Omega_h := \Omega \cup \Omega_h^\Gamma$), we replace (1) by an equivalent formulation

$$\begin{cases} -\Delta u + u = f & \text{in } \Omega_h, \text{ a domain slightly larger than } \Omega, \\ y + \nabla u = 0 & \text{on } \Omega_h^\Gamma, \text{ a strip around } \Gamma, \\ y \cdot \nabla\phi + \frac{1}{h}p\phi + \tilde{g}|\nabla\phi| = 0 & \text{on } \Omega_h^\Gamma. \end{cases} \tag{4}$$

A finite element approximation resulting from the reformulation above will be explained in detail in the next section. As in [7], we coin our method ϕ-FEM in accordance with the tradition of denoting the level-sets by ϕ. We emphasize once again that our method does not require any conformity between the computational mesh and the geometry of Ω. More specifically, the mesh is constructed as follows: we assume that Ω is embedded into a simply shaped domain $\mathcal{O}$ (typically a box in $\mathbb{R}^d$) and rely on a quasi-uniform simplicial mesh $\mathcal{T}_h^{\mathcal{O}}$ on $\mathcal{O}$ (the background mesh). We introduce then the computational mesh $\mathcal{T}_h$ as a submesh of $\mathcal{T}_h^{\mathcal{O}}$ obtained by getting rid of mesh cells lying entirely outside Ω (the precise definition of $\mathcal{T}_h$ given in the next Section will rely on an approximation ϕ_h of ϕ rather than on the level-set ϕ itself). The finite element approximation to (1) will be constructed by discretizing a variational formulation of (4) and adding some stabilization terms. Note that the additional unknowns y and p (and their discrete counterparts y_h and p_h) live on a small portion of the whole computational domain, so that the extra cost induced by their introduction is negligible as $h \to 0$.

The article is structured as follows: our ϕ-FEM method is presented in the next section. We also give there the assumptions on the level-set ϕ and on the mesh, and announce our main result: the *a priori* error estimate for ϕ-FEM. We work with standard continuous $\mathbb{P}_k$ finite elements on a simplicial mesh and prove the optimal order h^k for the error in the H^1 norm and the (slightly) suboptimal order $h^{k+1/2}$ for the error in the L^2 norm. The proofs of these estimates are the subject of Section 3. Moreover, we show in Section 4 that the associated finite element matrix has the condition number of order $1/h^2$, i.e. of the same order as that of a standard finite element method on a matching grid of comparable size. In particular, the conditioning of our method does not suffer from arbitrarily bad intersections of Γ with the mesh. Numerical illustrations are given in Section 5. Finally, some conclusions and perspectives are given in Section 6.

2 Definitions, assumptions, description of ϕ-FEM, and the main result

Assume $\Omega \subset \mathcal{O}$ and let $\mathcal{T}_h^{\mathcal{O}}$ be a quasi-uniform simplicial mesh on $\mathcal{O}$ with $h = \max_{T\in\mathcal{T}_h} \operatorname{diam} T$. Fix integers $k, l \geq 1$ and let ϕ_h be the FE interpolation of ϕ on $\mathcal{T}_h^{\mathcal{O}}$ by the usual continuous finite elements of degree l.[1] Let $\Gamma_h := \{\phi_h = 0\}$ and introduce then the computational mesh $\mathcal{T}_h$ (approximately) covering Ω, another auxiliary computational mesh $\mathcal{T}_h^\Gamma$ covering Γ_h, and the corresponding domains as

$$\begin{aligned} \mathcal{T}_h &= \{T \in \mathcal{T}_h^{\mathcal{O}} : T \cap \{\phi_h < 0\} \neq \varnothing\} && \text{and } \Omega_h = (\cup_{T\in\mathcal{T}_h} T)^\circ, \\ \mathcal{T}_h^\Gamma &= \{T \in \mathcal{T}_h : T \cap \Gamma_h \neq \varnothing\} && \text{and } \Omega_h^\Gamma = (\cup_{T\in\mathcal{T}_h^\Gamma} T)^\circ. \end{aligned}$$

Thus, $\mathcal{T}_h^\Gamma$ represents the cells of the mesh cut by the approximate boundary Γ_h. Note that Ω_h is typically slightly larger than Ω by a strip of the width of order h. In general, we do not assume however $\Omega \subset \Omega_h$ since, in some rare occasions, the boundary $\partial\Omega_h$ can be locally inside Ω due to the difference between ϕ and ϕ_h. We shall also denote by $\Omega_h^i = \Omega_h \setminus \Omega_h^\Gamma$ the domain of mesh elements completely covered by Ω, and by Γ_h^i the internal boundary of Ω_h^Γ, *i.e.* the ensemble of the facets separating Ω_h^Γ from the mesh elements inside Ω, so that $\partial\Omega_h^\Gamma = \partial\Omega_h \cup \Gamma_h^i$.

We now introduce the finite element spaces[2]

$$\begin{aligned} V_h^{(k)} &= \{v_h \in H^1(\Omega_h) : v_h|_T \in \mathbb{P}_k(T) \quad \forall T \in \mathcal{T}_h\}, \\ Z_h^{(k)} &= \{z_h \in H^1(\Omega_h^\Gamma)^d : z_h|_T \in \mathbb{P}_k(T)^d \quad \forall T \in \mathcal{T}_h^\Gamma\}, \\ Q_h^{(k)} &= \{q_h \in L^2(\Omega_h^\Gamma) : q_h|_T \in \mathbb{P}_{k-1}(T) \quad \forall T \in \mathcal{T}_h^\Gamma\}, \\ W_h^{(k)} &= V_h^{(k)} \times Z_h^{(k)} \times Q_h^{(k)} \end{aligned}$$

and the finite element problem: Find $(u_h, y_h, p_h) \in W_h^{(k)}$ such that

$$\begin{aligned} a_h(u_h, y_h, p_h; v_h, z_h, q_h) = &\int_{\Omega_h} f v_h + \gamma_{div} \int_{\Omega_h^\Gamma} f(\operatorname{div} z_h + v_h) \\ &- \frac{\gamma_2}{h^2} \int_{\Omega_h^\Gamma} \tilde{g}|\nabla\phi_h|(z_h \cdot \nabla\phi_h + \frac{1}{h} q_h \phi_h) \end{aligned} \tag{5}$$

for all $(v_h, z_h, q_h) \in W_h^{(k)}$ where

$$\begin{aligned} a_h(u, y, p; v, z, q) = &\int_{\Omega_h} \nabla u \cdot \nabla v + \int_{\Omega_h} uv + \int_{\partial\Omega_h} y \cdot nv + \gamma_{div} \int_{\Omega_h^\Gamma} (\operatorname{div} y + u)(\operatorname{div} z + v) \\ &+ \gamma_1 \int_{\Omega_h^\Gamma} (y + \nabla u) \cdot (z + \nabla v) + \sigma h \int_{\Gamma_h^i} \left[\frac{\partial u}{\partial n}\right] \left[\frac{\partial v}{\partial n}\right] + \frac{\gamma_2}{h^2} \int_{\Omega_h^\Gamma} (y \cdot \nabla\phi_h + \frac{1}{h} p\phi_h)(z \cdot \nabla\phi_h + \frac{1}{h} q\phi_h) \end{aligned}$$

with some positive numbers γ_{div}, γ_1, γ_2 and σ properly chosen in a manner independent of h. We have assumed here that f is well defined on Ω_h, rather than on Ω only. We also recall that $\tilde{g}$ is a lifting of g from Γ to Ω_h^Γ. Both f and $\tilde{g}$ are supposed sufficiently smooth as detailed in the forthcoming statement of our main theorem.

[1] The integer k is the degree of finite elements which will be used to approximate the principal unknown u while ϕ is approximated by finite elements of degree l. We shall require $l \geq k+1$ in our convergence Theorem 2.1. Note, that we cannot set $l = k$ unlike the Dirichlet case in [7]. This is essentially due to the fact that ϕ_h is used here to approximate the normal on Γ in addition to approximating Γ itself.

[2] We can also choose $Q_h^{(k)}$ as the continuous finite elements of degree $k-1$ on (Ω_h^Γ) if $k \geq 2$.

The finite element problem (5) is inspired by (4). Let us detail the derivation of (5) which is for the moment completely formal. We assume that the first line of (4) is indeed valid on Ω_h (rather than on Ω only) and we pursue by multiplying it by a test function v and integrating by parts. This gives

$$\int_{\Omega_h} \nabla u \cdot \nabla v + \int_{\Omega_h} uv - \int_{\partial\Omega_h} \nabla u \cdot nv = \int_{\Omega_h} fv.$$

Moreover, we introduce the additional unknowns y and p on Ω_h^Γ so that the second and third lines of (4) are satisfied there. This implies in particular $-\nabla u \cdot n = y \cdot n$ on $\partial\Omega_h$ and we use this in the boundary integral above which gives the corresponding terms in the definition of a_h. The equations from the second and third lines of (4) are then imposed in a least squares manner giving the terms multiplied by γ_1 and γ_2/h^2. Finally, the terms multiplied by σh and γ_{div} are added to stabilize the method. Both these terms are consistent with the governing equations since ∇u is assumed continuous across Γ_h^i and $\operatorname{div} y + u = -\Delta u + u = f$ on Ω_h^Γ.

We now recall some technical assumptions on the domain and the mesh, the same as in [10, 7]. These assumptions hold true for smooth domains and sufficiently refined meshes.

Assumption 1. *There exists a neighborhood of Γ, a domain Ω^Γ, which can be covered by open sets $\mathcal{O}_i$, $i = 1, \dots, I$ and one can introduce on every $\mathcal{O}_i$ local coordinates $\xi_1, \dots, \xi_d$ with $\xi_d = \phi$ such that all the partial derivatives $\partial^\alpha \xi / \partial x^\alpha$ and $\partial^\alpha x / \partial \xi^\alpha$ up to order $k+1$ are bounded by some $C_0 > 0$. Thus, ϕ is of class C^{k+1} on Ω^Γ. Moreover, $|\nabla \phi| \geq m$ on Ω^Γ with some $m > 0$.*

Assumption 2. *$\Omega_h^\Gamma \subset \Omega^\Gamma$ and $|\nabla \phi_h| \geq \frac{m}{2}$ on all the mesh elements composing Ω_h^Γ.*

This assumption is clearly valid for h small enough. In particular, it guarantees that Γ_h is indeed a curve in 2D (a surface in 3D). Note that it could be in principle possible that ϕ vanished on all the interpolation points used for ϕ_h on an element T so that ϕ_h would vanish on the hole element T which would be thus included into Γ_h. Assuming that $|\nabla \phi_h| > 0$ excludes these situations (which are highly unlikely any way).

Assumption 3. *The approximate boundary Γ_h can be covered by element patches $\{\Pi_k\}_{k=1,\dots,N_\Pi}$ having the following properties:*

- *Each Π_k is composed of a mesh element T_k lying inside Ω and some elements cut by Γ, more precisely $\Pi_k = T_k \cup \Pi_k^\Gamma$ where $T_k \in \mathcal{T}_h$, $T_k \subset \bar{\Omega}$, $\Pi_k^\Gamma \subset \mathcal{T}_h^\Gamma$, and Π_k^Γ contains at most M mesh elements;*
- *Each mesh element in a patch Π_k shares at least a facet with another mesh element in the same patch. In particular, T_k shares a facet F_k with an element in Π_k^Γ;*
- *$\mathcal{T}_h^\Gamma = \cup_{k=1}^{N_\Pi} \Pi_k^\Gamma$ and $\Gamma_h^i = \cup_{k=1}^{N_\Pi} F_k$;*
- *Π_k and Π_l are disjoint if $k \neq l$.*

Assumption 3 is satisfied for h small enough, preventing strong oscillations of Γ on the length scale h. It can be reformulated by saying that each cut element $T \in \mathcal{T}_h^\Gamma$ can be connected to an uncut element $T' \subset \Omega_h^i$ by a path consisting of a small number of mesh elements adjacent to one another; see [10] for a more detailed discussion and an illustration (Fig. 2).

Theorem 2.1. *Suppose that Assumptions 1–3 hold true, $l \geq k+1$, the mesh $\mathcal{T}_h$ is quasi-uniform, and $f \in H^k(\Omega_h)$, $\tilde{g} \in H^{k+1}(\Gamma)$. Let $u \in H^{k+2}(\Omega)$ be the solution to (1) and $(u_h, y_h, p_h) \in W_h^{(k)}$ be the solution to (5). Provided γ_{div}, γ_1, γ_2, σ are sufficiently big, it holds*

$$|u - u_h|_{1,\Omega\cap\Omega_h} \leq C h^k (\|f\|_{k,\Omega\cup\Omega_h} + \|\tilde{g}\|_{k+1,\Omega_h^\Gamma}),$$

with $C > 0$ depending on the constants in Assumptions 1, 3 (and thus on the norm of ϕ in C^{k+1}), on the mesh regularity, on the polynomial degrees k and l, and on Ω, but independent of h, f, g and u.

Moreover, if $\Omega \subset \Omega_h$, then

$$\|u - u_h\|_{0,\Omega} \leq Ch^{k+1/2}(\|f\|_{k,\Omega_h} + \|\tilde{g}\|_{k+1,\Omega_h^\Gamma})$$

with a constant $C > 0$ of the same type.

3 Proof of the *a priori* error estimates

From now on, we shall use the letter C for positive constants (which can vary from one line to another) that depend only on the regularity of the mesh and on the constants in Assumptions 1–3.

We shall begin with some technical results, mostly adapted from [10] and [7] to be used later in the proofs of the coercivity of a_h (Section 3.2) and the *a priori* error estimates (Sections 3.3 and 3.4). The most important contribution here is Lemma 3.2 which extends to finite elements of any degree a result from [10]. This lemma will be the keystone of the proof of coercivity by allowing us to handle the non necessarily positive terms on the cut elements.

3.1 Technical lemmas

We recall first a lemma from [7]:

Lemma 3.1. *Let T be a triangle/tetrahedron, E one of its sides and p a polynomial on T such that $p = a$ on E for some $a \in \mathbb{R}$, $\frac{\partial p}{\partial n} = 0$ on E, and $\Delta p = 0$ on T. Then $p = a$ on T.*

We now adapt a lemma from [10]:

Lemma 3.2. *Let B_h be the strip between $\partial\Omega_h$ and Γ_h. For any $\beta > 0$, there exist $0 < \alpha < 1$ and $\delta > 0$ depending only on the mesh regularity and geometrical assumptions such that*

$$\left|\int_{B_h} z_h \cdot \nabla v_h\right| \leq \alpha |v_h|^2_{1,\Omega_h} + \delta \|z_h + \nabla v_h\|^2_{0,\Omega_h^\Gamma} + \beta h \left\| \left[\frac{\partial v_h}{\partial n}\right] \right\|^2_{0,\Gamma_h^i} + \beta h^2 \| \operatorname{div} z_h + v_h\|^2_{0,\Omega_h^\Gamma} + \beta h^2 \|v_h\|^2_{0,\Omega_h^\Gamma} \quad (6)$$

for all $v_h \in V_h^{(k)}, z_h \in Z_h^{(k)}$.

Proof. The boundary Γ can be covered by element patches $\{\Pi_k\}_{k=1,\dots,N_\Pi}$ as in Assumption 3. Choose any $\beta > 0$ and consider

$$\alpha := \max_{\Pi_k, (z_h, v_h) \neq (0,0)} F(\Pi_k, z_h, v_h) \quad (7)$$

with

$$F(\Pi_k, z_h, v_h) = \frac{\|z_h\|_{0,\Pi_k^\Gamma} |v_h|_{1,\Pi_k^\Gamma} - \beta \|z_h + \nabla v_h\|^2_{0,\Pi_k^\Gamma} - \beta h \left\| \left[\frac{\partial v_h}{\partial n}\right] \right\|^2_{0,F_k} - \frac{\beta}{2} h^2 \| \operatorname{div} z_h\|^2_{0,\Pi_k^\Gamma}}{\frac{1}{2}\|z_h\|^2_{0,\Pi_k^\Gamma} + \frac{1}{2}|v_h|^2_{1,\Pi_k}},$$

where the maximum is taken over all the possible configurations of a patch Π_k allowed by the mesh regularity and over all $v_h \in V_h^{(k)}$ and $z_h \in Z_h^{(k)}$ restricted to Π_k, i.e. $v_h \in H^1(\Pi_k)$, $v_h \in \mathbb{P}_k(T)$ (resp. $z_h \in H^1(\Pi_k^\Gamma)^d$, $z_h \in \mathbb{P}_k(T)^d$) for all mesh elements $T \in \Pi_k$ (resp. $T \in \Pi_k^\Gamma$). Note that $F(\Pi_k, z_h, v_h)$ is invariant under the scaling transformation $x \mapsto \frac{1}{h}x$, $v_h \mapsto \frac{1}{h}v_h$, $z_h \mapsto z_h$. This means that if we introduce $\mathcal{S}(x) = \frac{1}{h}x$ and set $\tilde{\Pi} = \mathcal{S}(\Pi_k)$, $\tilde{\Pi}^\Gamma = \mathcal{S}(\Pi_k^\Gamma)$, $\tilde{F} = \mathcal{S}(F_k)$, $\tilde{T} = \mathcal{S}(T)$, $\tilde{z} = z_h \circ \mathcal{S}^{-1}$, $\tilde{v} = \frac{1}{h}v_h \circ \mathcal{S}^{-1}$ then

$$F(\Pi_k, z_h, v_h) = \frac{\|\tilde{z}\|_{0,\tilde{\Pi}^\Gamma} |\tilde{v}|_{1,\tilde{\Pi}^\Gamma} - \beta \|\tilde{z} + \nabla \tilde{v}\|^2_{0,\tilde{\Pi}^\Gamma} - \beta \left\| \left[\frac{\partial \tilde{v}}{\partial n}\right] \right\|^2_{0,\tilde{F}} - \frac{\beta}{2} \| \operatorname{div} \tilde{z}\|^2_{0,\tilde{\Pi}^\Gamma}}{\frac{1}{2}\|\tilde{z}\|^2_{0,\tilde{\Pi}^\Gamma} + \frac{1}{2}|\tilde{v}|^2_{1,\tilde{\Pi}}}.$$

We can thus assume $h = 1$ when computing the maximum in (7). Moreover, $F(\Pi_k, z_h, v_h)$ is homogeneous with respect to v_h, z_h, meaning

$$F(\Pi_k, z_h, v_h) = F(\Pi_k, \mu z_h, \mu v_h)$$

for any $\mu \in \mathbb{R}$, $\mu \neq 0$. Thus, the maximum in (7) is indeed attained since it can be taken over a closed bounded set in a finite dimensional space (all the admissible patches on a mesh with $h = 1$ and all v_h, z_h such that $|v_h|^2_{1,\Pi_k} + \|z_h\|^2_{0,\Pi_k^\Gamma} = 1$).

Clearly, $\alpha \leq 1$. Supposing $\alpha = 1$ would lead to a contradiction. Indeed, if $\alpha = 1$, we can then take Π_k, v_h, z_h yielding this maximum (in particular, $|v_h|^2_{1,\Pi_k} + \|z_h\|^2_{0,\Pi_k^\Gamma} > 0$). We observe then

$$\frac{1}{2}|v_h|^2_{1,\Pi_k} - \|z_h\|_{0,\Pi_k^\Gamma}|v_h|_{1,\Pi_k^\Gamma} + \frac{1}{2}\|z_h\|^2_{0,\Pi_k^\Gamma} + \beta\|z_h + \nabla v_h\|^2_{0,\Pi_k^\Gamma} + \beta h \left\| \left[\frac{\partial v_h}{\partial n} \right] \right\|^2_{0,F_k} + \frac{\beta}{2} h^2 \| \operatorname{div} z_h\|^2_{0,\Pi_k^\Gamma} = 0$$

and consequently (recall $|v_h|^2_{1,\Pi_k} = |v_h|^2_{1,T_k} + |v_h|^2_{1,\Pi_k^\Gamma}$)

$$\frac{1}{2}|v_h|^2_{1,T_k} + \frac{1}{2}(|v_h|_{1,\Pi_k^\Gamma} - \|z_h\|_{0,\Pi_k^\Gamma})^2 + \beta\|z_h + \nabla v_h\|^2_{0,\Pi_k^\Gamma} + \beta h \left\| \left[\frac{\partial v_h}{\partial n} \right] \right\|^2_{0,F_k} + \frac{\beta}{2} h^2 \| \operatorname{div} z_h\|^2_{0,\Pi_k^\Gamma} = 0. \quad (8)$$

This implies

- $|v_h|_{1,T_k}$=0 so that $v_h = \text{const}$ on T_k.
- $\|z_h + \nabla v_h\|_{0,\Pi_k^\Gamma}$=0 so that $\nabla v_h = -z_h$ on Π_k^Γ. In particular, ∇v_h is continuous on Π_k^Γ.
- The jumps $\left[\frac{\partial v_h}{\partial n}\right]$ vanish at all the mesh facets inside Π_k. This is valid on the facet F_k separating T_k from Π_k^Γ directly by (8) and is also implied on the internal facets of Π_k^Γ by the continuity of ∇v_h on Π_k^Γ.
- $\operatorname{div} z_h = -\Delta v_h = 0$ on Π_k^Γ.

Combining all this with Lemma 3.1, starting from T_k and its neighbor in Π_k^Γ and then propagating to other elements of Π_k^Γ, we see that v_h =const on the whole Π_k. We have thus $\nabla v_h = 0$ on Π_k and $z_h = 0$ on Π_k^Γ, which is in contradiction with $|v_h|^2_{1,\Pi_k} + \|z_h\|^2_{0,\Pi_k^\Gamma} > 0$.

Thus $\alpha < 1$ and

$$\|z_h\|_{0,\Pi_k^\Gamma}|v_h|_{1,\Pi_k^\Gamma} \leq \frac{\alpha}{2}\|z_h\|^2_{0,\Pi_k^\Gamma} + \frac{\alpha}{2}|v_h|^2_{1,\Pi_k} + \beta\|z_h + \nabla v_h\|^2_{0,\Pi_k^\Gamma} + \beta h \left\| \left[\frac{\partial v_h}{\partial n} \right] \right\|^2_{0,\partial T_k \cap \partial \Pi_k^\Gamma} + \frac{\beta}{2} h^2 \| \operatorname{div} z_h\|^2_{0,\Pi_k^\Gamma}$$

for all v_h, z_h and all admissible patches Π_k. We now observe

$$\begin{aligned}\left| \int_{B_h} z_h \cdot \nabla v_h \right| &\leq \sum_k \left| \int_{B_h \cap \Pi_k^\Gamma} z_h \cdot \nabla v_h \right| \leq \sum_k \|z_h\|_{0,\Pi_k^\Gamma}|v_h|_{1,\Pi_k^\Gamma} \\ &\leq \frac{\alpha}{2}\|z_h\|^2_{0,\Omega_h^\Gamma} + \frac{\alpha}{2}|v_h|^2_{1,\Omega_h} + \beta\|z_h + \nabla v_h\|^2_{0,\Omega_h^\Gamma} + \beta h \left\| \left[\frac{\partial v_h}{\partial n} \right] \right\|^2_{0,\Gamma_h^i} + \frac{\beta}{2} h^2 \| \operatorname{div} z_h\|^2_{0,\Omega_h^\Gamma}.\end{aligned}$$

We now use the Young inequality with any $\varepsilon > 0$ to obtain

$$\|z_h\|^2_{0,\Omega_h^\Gamma} = \|z_h + \nabla v_h\|^2_{0,\Omega_h^\Gamma} + \|\nabla v_h\|^2_{0,\Omega_h^\Gamma} - 2(z_h + \nabla v_h, \nabla v_h)_{0,\Omega_h^\Gamma}$$

$$\leq \left(1+\frac{1}{\varepsilon}\right)\|z_h+\nabla v_h\|^2_{0,\Omega_h^\Gamma}+(1+\varepsilon)|v_h|^2_{1,\Omega_h},$$

which leads to

$$\left|\int_{B_h} z_h\cdot\nabla v_h\right| \leq \alpha\left(1+\frac{\varepsilon}{2}\right)|v_h|^2_{1,\Omega_h}+\left(\beta+\frac{\alpha}{2}+\frac{\alpha}{2\varepsilon}\right)\|z_h+\nabla v_h\|^2_{0,\Omega_h^\Gamma}+\beta h\left\|\left[\frac{\partial v_h}{\partial n}\right]\right\|^2_{0,\Gamma_h^i}+\beta h^2\|\operatorname{div} z_h\|^2_{0,\Omega_h^\Gamma}.$$

Taking ε sufficiently small, redefining α as $\alpha\left(1+\frac{\varepsilon}{2}\right)$ and putting $\delta=\left(\beta+\frac{\alpha}{2}+\frac{\alpha}{2\varepsilon}\right)$ we obtain

$$\left|\int_{B_h} z_h\cdot\nabla v_h\right| \leq \alpha|v_h|^2_{1,\Omega_h}+\delta\|z_h+\nabla v_h\|^2_{0,\Omega_h^\Gamma}+\beta h\left\|\left[\frac{\partial v_h}{\partial n}\right]\right\|^2_{0,\Gamma_h^i}+\beta h^2\|\operatorname{div} z_h\|^2_{0,\Omega_h^\Gamma}.$$

This leads to (6) by the triangle inequality $\|\operatorname{div} z_h\|_{0,\Omega_h^\Gamma}\leq\|\operatorname{div} z_h+v_h\|_{0,\Omega_h^\Gamma}+\|v_h\|_{0,\Omega_h^\Gamma}$. □

Lemma 3.3. *For all $v\in H^1(\Omega_h^\Gamma)$,*

$$\|v\|_{0,\Omega_h^\Gamma}\leq C\left(\sqrt{h}\|v\|_{0,\Gamma_h^i}+h|v|_{1,\Omega_h^\Gamma}\right)$$

and for all $v\in H^1(\Omega_h\backslash\Omega)$,

$$\|v\|_{0,\Omega_h\backslash\Omega}\leq C\left(\sqrt{h}\|v\|_{0,\Gamma}+h|v|_{1,\Omega_h\backslash\Omega}\right).$$

We refer to [10] for the first inequality. The second one can be treated similarly.

The following lemma is borrowed from [7]. It's a partial generalization of Lemma 3.3 to derivatives of higher order.

Lemma 3.4. *Under Assumption 1, it holds for all $v\in H^s(\Omega_h)$ with integer $1\leq s\leq k+1$, v vanishing on Ω,*

$$\|v\|_{0,\Omega_h\backslash\Omega}\leq Ch^s\|v\|_{s,\Omega_h\backslash\Omega}.$$

Lemma 3.5. *For all piecewise polynomial (possibly discontinuous) functions v_h on $\mathcal{T}_h^\Gamma$*

$$\|v_h\|_{0,\Gamma_h}\leq\frac{C}{\sqrt{h}}\|v_h\|_{0,\Omega_h^\Gamma}$$

with a constant $C>0$ depending on the maximal degree of polynomials in v_h and on the constants in Assumptions 1–3.

Proof. Take any $T\in\mathcal{T}_h^\Gamma$ and denote $E=T\cap\Gamma_h$. Since ϕ_h is a polynomial of degree l on each cell T, a scaling argument gives $|\Gamma_h\cap T|\leq\frac{C_\Gamma}{h}|T|$ with some $C_\Gamma>0$ depending on l and the mesh regularity. Hence

$$\|v_h\|^2_{0,E}\leq\|v_h\|^2_{L^\infty(T)}|E|\leq\frac{C_\Gamma}{h}\|v_h\|^2_{L^\infty(T)}|T|.$$

Applying the inverse inequality $\|v_h\|^2_{L^\infty(T)}\leq C\|v_h\|^2_{0,T}/|T|$ yields

$$\|v_h\|^2_{0,E}\leq\frac{C}{h}\|v_h\|^2_{0,T}.$$

Summing over all $T\in\mathcal{T}_h^\Gamma$ concludes the proof. □

Finally, we recall a Hardy-type lemma from [7]. Unlike the Dirichlet case in [7], we need the level-set ϕ only in a neighborhood Ω^Γ of Γ. The lemma is adapted accordingly.

Lemma 3.6. *We assume that the domain Ω^Γ is a neighborhood of Γ, given by (2), and satisfies Assumption 1. Then, for any $u\in H^{s+1}(\Omega^\Gamma)$ vanishing on Γ, and an integer $s\in[0,k]$,*

$$\left\|\frac{u}{\phi}\right\|_{s,\Omega^\Gamma}\leq C\|u\|_{s+1,\Omega^\Gamma}$$

with $C>0$ depending only on the constants in Assumption 1 and on s.

3.2 Coercivity of the bilinear form a

It will be convenient to rewrite the bilinear form a_h in a manner avoiding the integral on $\partial\Omega_h$. To this end, we recall that B_h is the strip between $\partial\Omega_h$ and Γ_h and observe for any $y \in H^1(B_h)^d$, $v \in H^1(B_h)$, $q \in L^2(\Gamma_h)$:

$$\begin{aligned}\int_{\partial\Omega_h} y\cdot nv &= \int_{\partial\Omega_h} y\cdot nv - \int_{\Gamma_h}\frac{1}{|\nabla\phi_h|}(y\cdot\nabla\phi_h)v + \int_{\Gamma_h}\frac{1}{|\nabla\phi_h|}(y\cdot\nabla\phi_h+\frac{1}{h}q\phi_h)v_h\\ &= \int_{B_h}(v\,\mathrm{div}\, y + y\cdot\nabla v) + \int_{\Gamma_h}\frac{1}{|\nabla\phi_h|}(y\cdot\nabla\phi_h+\frac{1}{h}q\phi_h)v.\end{aligned}$$

Indeed, $\phi_h = 0$ on Γ_h and the unit normal to Γ_h, looking outward from B_h, is equal to $-\nabla\phi_h/|\nabla\phi_h|$. Thus,

$$\begin{aligned}a_h(u,y,p;v,z,q) = &\int_{\Omega_h}\nabla u\cdot\nabla v + \int_{\Omega_h} uv + \int_{B_h}(v\,\mathrm{div}\, y + y\cdot\nabla v) + \int_{\Gamma_h}\frac{1}{|\nabla\phi_h|}(y\cdot\nabla\phi_h+\frac{1}{h}q\phi_h)v\\ &+\gamma_{div}\int_{\Omega_h^\Gamma}(\mathrm{div}\, y+u)(\mathrm{div}\, z+v) + \gamma_1\int_{\Omega_h^\Gamma}(y+\nabla u)\cdot(z+\nabla v)\\ &+\sigma h\int_{\Gamma_h^i}\left[\frac{\partial u}{\partial n}\right]\left[\frac{\partial v}{\partial n}\right] + \frac{\gamma_2}{h^2}\int_{\Omega_h^\Gamma}(y\cdot\nabla\phi_h+\frac{1}{h}p\phi_h)(z\cdot\nabla\phi_h+\frac{1}{h}q\phi_h). \quad (9)\end{aligned}$$

Proposition 1. *Provided $\gamma_{div},\gamma_1,\gamma_2,\sigma$ are sufficiently big, there exists an h-independent constant $c>0$ such that*

$$a_h(v_h,z_h,q_h;v_h,z_h,q_h) \geq c|||v_h,z_h,q_h|||_h^2, \quad \forall (v_h,z_h,q_h)\in W_h^{(k)}$$

with

$$|||v,z,q|||_h^2 = \|v\|_{1,\Omega_h}^2 + \|\,\mathrm{div}\, z+v\|_{0,\Omega_h^\Gamma}^2 + \|z+\nabla v\|_{0,\Omega_h^\Gamma}^2 + h\left\|\left[\frac{\partial v}{\partial n}\right]\right\|_{0,\Gamma_h^i}^2 + \frac{1}{h^2}\|z\cdot\nabla\phi_h+\frac{1}{h}q\phi_h\|_{0,\Omega_h^\Gamma}^2.$$

Proof. Using the reformulation of the bilinear form a_h given by (9), we have for all $(v_h,z_h,q_h)\in W_h^{(k)}$,

$$\begin{aligned}a_h(v_h,z_h,q_h;v_h,z_h,q_h) = &\|v_h\|_{1,\Omega_h}^2 + \|v_h\|_{0,\Omega_h}^2 + \int_{B_h}(v_h\,\mathrm{div}\, z_h + z_h\cdot\nabla v_h)\\ &+\int_{\Gamma_h}\frac{1}{|\nabla\phi_h|}(z_h\cdot\nabla\phi_h+\frac{1}{h}q_h\phi_h)v_h + \gamma_{div}\|\,\mathrm{div}\, z_h+v_h\|_{0,\Omega_h^\Gamma}^2 + \gamma_1\|z_h+\nabla v_h\|_{0,\Omega_h^\Gamma}^2\\ &+\sigma h\left\|\left[\frac{\partial v_h}{\partial n}\right]\right\|_{0,\Gamma_h^i}^2 + \frac{\gamma_2}{h^2}\|z_h\cdot\nabla\phi_h+\frac{1}{h}q_h\phi_h\|_{0,\Omega_h^\Gamma}^2.\end{aligned}$$

Since $B_h\subset\Omega_h^\Gamma$, we remark that the integral of $v_h\,\mathrm{div}\, z_h$ can be combined with that of v_h on Ω_h^Γ to give

$$\|v_h\|_{0,\Omega_h^\Gamma}^2 + \int_{B_h} v_h\,\mathrm{div}\, z_h \geq \int_{B_h} v_h\,(\mathrm{div}\, z_h + v_h) \geq -\|v_h\|_{0,\Omega_h^\Gamma}\|\,\mathrm{div}\, z_h+v_h\|_{0,\Omega_h^\Gamma}.$$

We also use an inverse inequality from Lemma 3.5 and the fact that $1/|\nabla\phi_h|$ is uniformly bounded by Assumption 2, to estimate

$$\left|\int_{\Gamma_h}\frac{1}{|\nabla\phi_h|}(z_h\cdot\nabla\phi_h+\frac{1}{h}q_h\phi_h)v_h\right| \leq \frac{C}{h}\|z_h\cdot\nabla\phi_h+\frac{1}{h}q_h\phi_h\|_{0,\Omega_h^\Gamma}\|v_h\|_{0,\Omega_h^\Gamma}.$$

Applying the Young inequality (for any $\varepsilon > 0$) to the last two bounds and combining this with (6) yields

$$a_h(v_h, z_h, q_h; v_h, z_h, q_h) \geq (1-\alpha)|v_h|^2_{1,\Omega_h} + \|v_h\|^2_{0,\Omega_h^i} - (\varepsilon + \beta h^2)\|v_h\|^2_{0,\Omega_h^\Gamma}$$
$$+ \left(\gamma_{div} - \frac{1}{2\varepsilon} - \beta h^2\right) \|\operatorname{div} z_h + v_h\|^2_{0,\Omega_h^\Gamma} + (\gamma_1 - \delta)\|z_h + \nabla v_h\|^2_{0,\Omega_h^\Gamma} + (\sigma - \beta)h \left\| \left[\frac{\partial v_h}{\partial n} \right] \right\|^2_{0,\Gamma_h^i}$$
$$+ \left(\frac{\gamma_2}{h^2} - \frac{C^2}{2\varepsilon h^2}\right) \|z_h \cdot \nabla \phi_h + \frac{1}{h} q_h \phi_h\|^2_{0,\Omega_h^\Gamma}.$$

To bound further from below the first 3 terms we note, using Lemma 3.3 and the trace inverse inequality,

$$\|v_h\|^2_{0,\Omega_h^\Gamma} \leq C(h\|v_h\|^2_{0,\Gamma_h^i} + h^2|v_h|^2_{1,\Omega_h^\Gamma}) \leq C(\|v_h\|^2_{0,\Omega_h^i} + h^2|v_h|^2_{1,\Omega_h})$$

so that, introducing any $\kappa \geq 0$ and observing $h \leq h_0 := \operatorname{diam}(\Omega)$,

$$(1-\alpha)|v_h|^2_{1,\Omega_h} + \|v_h\|^2_{0,\Omega_h^i} - (\varepsilon + \beta h^2)\|v_h\|^2_{0,\Omega_h^\Gamma} \geq (1-\alpha)|v_h|^2_{1,\Omega_h} + \|v_h\|^2_{0,\Omega_h^i} + \kappa\|v_h\|^2_{0,\Omega_h^\Gamma}$$
$$-(\varepsilon + \beta h_0^2 + \kappa)\|v_h\|^2_{0,\Omega_h^\Gamma}$$
$$\geq (1 - \alpha - C(\varepsilon + \beta h_0^2 + \kappa)h_0^2)|v_h|^2_{1,\Omega_h} + (1 - C(\varepsilon + \beta h_0^2 + \kappa))\|v_h\|^2_{0,\Omega_h^i} + \kappa\|v_h\|^2_{0,\Omega_h^\Gamma}.$$

Taking $\varepsilon, \kappa, \beta$ sufficiently small and $\gamma_1, \gamma_2, \gamma_{div}$ sufficiently big, gives the announced lower bound for $a_h(v_h, z_h, q_h; v_h, z_h, q_h)$. □

3.3 Proof of the H^1 error estimate in Theorem 2.1

Under the Theorem's assumptions, the solution to (1) is indeed in $H^{k+2}(\Omega)$ and it can be extended to a function $\tilde{u} \in H^{k+2}(\Omega_h)$ such that $\tilde{u} = u$ on Ω and

$$\|\tilde{u}\|_{k+2,\Omega_h} \leq C(\|f\|_{k,\Omega} + \|g\|_{k+1/2,\Gamma}) \leq C(\|f\|_{k,\Omega} + \|\tilde{g}\|_{k+1,\Omega_h^\Gamma}). \tag{10}$$

Introduce $y = -\nabla\tilde{u}$ and $p = -\frac{1}{\phi}(y \cdot \nabla\phi + \tilde{g}|\nabla\phi|)$ on Ω_h^Γ. Then, $y \in H^{k+1}(\Omega_h^\Gamma)$ and $p \in H^k(\Omega_h^\Gamma)$ by Lemma 3.6. Moreover,

$$\|y\|_{k+1,\Omega_h^\Gamma} \leq C\|\tilde{u}\|_{k+2,\Omega_h} \leq C(\|f\|_{k,\Omega} + \|\tilde{g}\|_{k+1,\Omega_h^\Gamma}) \tag{11}$$

and

$$\|p\|_{k,\Omega_h^\Gamma} \leq C(\|y\|_{k+1,\Omega_h^\Gamma} + +\|\tilde{g}\|_{k+1,\Omega_h^\Gamma}) \leq C(\|f\|_{k,\Omega} + \|\tilde{g}\|_{k+1,\Omega_h^\Gamma}). \tag{12}$$

Clearly, $\tilde{u}$, y, p satisfy

$$a_h(\tilde{u}, y, p; v_h, z_h, q_h) = \int_{\Omega_h} \tilde{f} v_h + \gamma_{div} \int_{\Omega_h^\Gamma} \tilde{f}(\operatorname{div} z_h + v_h) + \frac{\gamma_2}{h^2} \int_{\Omega_h^\Gamma} (y \cdot \nabla\phi_h + \frac{1}{h} p\phi_h)(z_h \cdot \nabla\phi_h + \frac{1}{h} q_h\phi_h)$$
$$\forall (v_h, z_h, q_h) \in W_h^{(k)}$$

with $\tilde{f} := -\Delta\tilde{u} + \tilde{u}$. It entails a Galerkin orthogonality relation

$$a_h(\tilde{u} - u_h, y - y_h, p - p_h; v_h, z_h, q_h) = \int_{\Omega_h} (\tilde{f} - f) v_h + \gamma_{div} \int_{\Omega_h^\Gamma} (\tilde{f} - f)(\operatorname{div} z_h + v_h)$$
$$+ \frac{\gamma_2}{h^2} \int_{\Omega_h^\Gamma} (y \cdot \nabla\phi_h + \frac{1}{h} p\phi_h + \tilde{g}|\nabla\phi_h|)(z_h \cdot \nabla\phi_h + \frac{1}{h} q_h\phi_h), \quad \forall (v_h, z_h, q_h) \in W_h^{(k)}. \tag{13}$$

Introducing the standard nodal interpolation I_h or, if necessary, a Clément interpolation (recall that p is only in $H^1(\Omega_h^\Gamma)$ if $k=1$), we then have by Proposition 1,

$$\begin{aligned} c\,|||\, u_h - I_h\tilde{u}, y_h - I_h y, p_h - I_h p|||_h &\leq \sup_{(v_h,z_h,q_h)\in W_h^{(k)}} \frac{a_h(u_h - I_h\tilde{u}, y_h - I_h y, p_h - I_h p; v_h, z_h, q_h)}{|||v_h, z_h, q_h|||_h} \\ &\leq \sup_{(v_h,z_h,q_h)\in W_h^{(k)}} \frac{I - II - III}{|||v_h, z_h, q_h|||_h}, \end{aligned}$$

where

$$\begin{aligned} I &= a_h(e_u, e_y, e_p; v_h, z_h, q_h), \\ II &= \int_{\Omega_h} (\tilde{f} - f)v_h + \gamma_{div}\int_{\Omega_h^\Gamma} (\tilde{f} - f)(\operatorname{div} z_h + v_h), \\ III &= \frac{\gamma_2}{h^2}\int_{\Omega_h^\Gamma} (y\cdot\nabla\phi_h + \frac{1}{h}p\phi_h + \tilde{g}|\nabla\phi_h|)(z_h\cdot\nabla\phi_h + \frac{1}{h}q_h\phi_h), \end{aligned}$$

with

$$e_u = \tilde{u} - I_h\tilde{u},\ e_y = y - I_h\tilde{y},\ e_p = p - I_h\tilde{p}.$$

We now estimate each term separately. Recalling (9), we have

$$\begin{aligned} I \leq \|e_u\|_{1,\Omega_h}\|v_h\|_{1,\Omega_h} + \|\mathrm{div} e_y\|_{0,B_h}\|v_h\|_{0,B_h} + \|e_y\|_{0,B_h}|v_h|_{1,B_h} + \|\frac{1}{|\nabla\phi_h|}(e_y\cdot\nabla\phi_h + \frac{1}{h}e_p\phi_h)\|_{0,\Gamma_h}\|v_h\|_{0,\Gamma_h} \\ + \gamma_{div}\|\operatorname{div} e_y + e_u\|_{0,\Omega_h^\Gamma}\|\operatorname{div} z_h + u_h\|_{0,\Omega_h^\Gamma} + \gamma_1\|e_y + \nabla e_u\|_{0,\Omega_h^\Gamma}\|z_h + \nabla v_h\|_{0,\Omega_h^\Gamma} \\ + \sigma h\left\|\left[\frac{\partial e_u}{\partial n}\right]\right\|_{0,\Gamma_h^i}\left\|\left[\frac{\partial v_h}{\partial n}\right]\right\|_{0,\Gamma_h^i} + \frac{\gamma_2}{h^2}\|e_y\cdot\nabla\phi_h + \frac{1}{h}e_p\phi_h\|_{0,\Omega_h^\Gamma}\|z_h\cdot\nabla\phi_h + \frac{1}{h}q_h\phi_h\|_{0,\Omega_h^\Gamma}. \end{aligned}$$

Applying Lemma 3.5 to the L^2-norms on Γ_h, recalling that $1/|\nabla\phi_h|$ is uniformly bounded on Ω_h^Γ (cf. Assumption 2), and recombining the terms, we get

$$I \ \leq\ C\left(\|e_u\|^2_{1,\Omega_h} + \|e_y\|^2_{1,\Omega_h^\Gamma} + h\left\|\left[\frac{\partial e_u}{\partial n}\right]\right\|^2_{0,\Gamma_h^i} + \frac{1}{h^2}\|e_y\cdot\nabla\phi_h + \frac{1}{h}e_p\phi_h\|^2_{0,\Omega_h^\Gamma}\right)^{1/2} |||\ v_h, z_h, q_h\ |||_h\ .$$

The usual interpolation estimates give

$$\|e_u\|^2_{1,\Omega_h} + \|e_y\|^2_{1,\Omega_h^\Gamma} + h\left\|\left[\tfrac{\partial e_u}{\partial n}\right]\right\|^2_{0,\Gamma_h^i} \leq Ch^{2k}(\|\tilde{u}\|^2_{k+1,\Omega_h} + \|y\|^2_{k+1,\Omega_h^\Gamma}).$$

Moreover, recalling that $|\nabla\phi_h|$ and $|\phi_h|/h$ are uniformly bounded on Ω_h^Γ, we get

$$\frac{1}{h^2}\|e_y\cdot\nabla\phi_h + e_p\phi_h\|^2_{0,\Omega_h^\Gamma} \leq \frac{C}{h^2}\|e_y\|^2_{0,\Omega_h^\Gamma} + C\|e_p\|^2_{0,\Omega_h^\Gamma} \leq Ch^{2k}(|y|^2_{k+1,\Omega_h^\Gamma} + |p|^2_{k,\Omega_h^\Gamma}).$$

Thus, by regularity estimates (10)–(12),

$$I \leq Ch^k(\|f\|_{k,\Omega} + \|\tilde{g}\|_{k+1,\Omega_h^\Gamma})\ |||\ v_h, z_h, q_h\ |||_h\ .$$

We now estimate the second term

$$\begin{aligned} |II| &\leq C(\|\tilde{f} - f\|_{0,\Omega_h}\|v_h\|_{0,\Omega_h} + \|\tilde{f} - f\|_{0,\Omega_h^\Gamma}\|\operatorname{div} z_h + v_h\|_{0,\Omega_h^\Gamma}) \\ &\leq C\|\tilde{f} - f\|_{0,\Omega_h}\ |||\ v_h, z_h, q_h\ |||_h \\ &\leq Ch^k\|f\|_{k,\Omega\cup\Omega_h}\ |||\ v_h, z_h, q_h\ |||_h\ . \end{aligned}$$

Indeed, thanks to Lemma 3.4 and $f = \tilde{f}$ on Ω,

$$\|\tilde{f} - f\|_{0,\Omega_h} = \|\tilde{f} - f\|_{0,\Omega_h \setminus \Omega} \leq Ch^k \|\tilde{f} - f\|_{k,\Omega_h \setminus \Omega} \leq Ch^k \|f\|_{k,\Omega \cup \Omega_h}. \tag{14}$$

Finally,

$$|III| \leq \frac{C}{h} \|y \cdot \nabla \phi_h + \frac{1}{h} p\phi_h + \tilde{g}|\nabla \phi_h|\|_{0,\Omega_h^\Gamma} \,|||\, v_h, z_h, q_h \,|||_h$$

and, recalling $y \cdot \nabla \phi + p\phi + \tilde{g}|\nabla \phi| = 0$ on Ω_h^Γ,

$$\begin{aligned}
\frac{1}{h}\|y \cdot \nabla \phi_h + \frac{1}{h} p\phi_h + \tilde{g}|\nabla \phi_h|\|_{0,\Omega_h^\Gamma} &= \frac{1}{h}\|y \cdot \nabla(\phi_h - \phi) + \frac{1}{h} p(\phi_h - \phi) + \tilde{g}(|\nabla \phi_h| - |\nabla \phi|)\|_{0,\Omega_h^\Gamma} \\
&\leq \frac{1}{h}\left(\|y\|_{0,\Omega_h^\Gamma} \|\nabla(\phi_h - \phi)\|_\infty + \|p\|_{0,\Omega_h^\Gamma} \|\phi_h - \phi\|_\infty + \|\tilde{g}\|_{0,\Omega_h^\Gamma} \||\nabla \phi| - |\nabla \phi_h|\|_\infty \right) \\
&\leq Ch^k (\|y\|_{0,\Omega_h^\Gamma} + h\|p\|_{0,\Omega_h^\Gamma} + \|\tilde{g}\|_{0,\Omega_h^\Gamma}) \leq Ch^k (\|f\|_{k,\Omega} + \|\tilde{g}\|_{k+1,\Omega_h^\Gamma})
\end{aligned}$$

by regularity estimates (11)–(12). Note that the optimal order is achieved here since ϕ is approximated by finite elements of degree at least $k+1$.

Combining the estimate for the terms I–III leads to

$$|||u_h - I_h \tilde{u}, y_h - I_h y, p_h - I_h p|||_h \leq Ch^k (\|f\|_{k,\Omega \cup \Omega_h} + \|\tilde{g}\|_{k+1,\Omega_h^\Gamma}),$$

so that, by the triangle inequality together with interpolation estimate, we get

$$|||u_h - \tilde{u}, y_h - y, p_h - p|||_h \leq Ch^k (\|f\|_{k,\Omega \cup \Omega_h} + \|\tilde{g}\|_{k+1,\Omega_h^\Gamma}). \tag{15}$$

This implies the announced H^1 error estimate for $u - u_h$.

3.4 Proof of the L^2 error estimate in Theorem 2.1

Since $\Omega \subset \Omega_h$, we can introduce $w : \Omega \to \mathbb{R}$ such that

$$\begin{cases} -\Delta w + w = u - u_h & \text{in } \Omega, \\ \frac{\partial w}{\partial n} = 0 & \text{on } \Gamma. \end{cases}$$

By elliptic regularity, $\|w\|_{2,\Omega} \leq C\|u - u_h\|_{0,\Omega}$. Let $\tilde{w}$ be an extension of w from Ω to Ω_h preserving the H^2 norm estimate and set $w_h = I_h \tilde{w}$. We observe

$$\begin{aligned}
\|u - u_h\|_{0,\Omega}^2 &= \int_\Omega \nabla(u - u_h) \cdot \nabla(w - w_h) + \int_\Omega (u - u_h)(w - w_h) + \int_\Omega \nabla(u - u_h) \cdot \nabla w_h + \int_\Omega (u - u_h) w_h \\
&\leq Ch^{k+1} (\|f\|_{k,\Omega_h} + \|\tilde{g}\|_{k+1,\Omega_h^\Gamma}) |\tilde{w}|_{2,\Omega_h} + \left| \int_\Omega \nabla(u - u_h) \cdot \nabla w_h + \int_\Omega (u - u_h) w_h \right|
\end{aligned}$$

by the already proven H^1-error estimate and interpolation estimates for $I_h \tilde{w}$ (recall also $\Omega \subset \Omega_h$). Taking $v_h = w_h$, $z_h = 0$ and $q_h = 0$ in the Galerkin orthogonality relation (13), we obtain

$$a_h(\tilde{u} - u_h, y - y_h, p - p_h; w_h, 0, 0) = \int_{\Omega_h} (\tilde{f} - f) w_h + \gamma_{div} \int_{\Omega_h^\Gamma} (\tilde{f} - f) w_h.$$

By (9), this reads

$$\begin{aligned}
&\int_{\Omega_h} \nabla(\tilde{u} - u_h) \cdot \nabla w_h + \int_{\Omega_h} (\tilde{u} - u_h) w_h + \int_{B_h} (w_h \operatorname{div}(y - y_h) + (y - y_h) \cdot \nabla w_h) \\
&\quad + \int_{\Gamma_h} \frac{1}{|\nabla \phi_h|} ((y - y_h) \cdot \nabla \phi_h + \frac{1}{h}(p - p_h)\phi_h) w_h + \gamma_{div} \int_{\Omega_h^\Gamma} (\operatorname{div}(y - y_h) + \tilde{u} - u_h) w_h \\
&\quad + \gamma_1 \int_{\Omega_h^\Gamma} ((y - y_h) + \nabla(\tilde{u} - u_h)) \cdot \nabla w_h + \sigma h \int_{\Gamma_h^i} \left[\frac{\partial(\tilde{u} - u_h)}{\partial n} \right] \left[\frac{\partial w_h}{\partial n} \right] = (1 + \gamma_{div}) \int_{\Omega_h} (\tilde{f} - f) w_h.
\end{aligned}$$

Using the last relation in the bound for $\|u-u_h\|^2_{0,\Omega}$, we can further bound it as

$$\begin{aligned}
\|u-u_h\|^2_{0,\Omega} &\leqslant Ch^{k+1}(\|f\|_{k,\Omega_h}+\|\tilde{g}\|_{k+1,\Omega_h^\Gamma})|\tilde{w}|_{2,\Omega_h}\\
&\quad+\left|\int_{\Omega_h\setminus\Omega}\nabla(\tilde{u}-u_h)\cdot\nabla w_h+\int_{\Omega_h\setminus\Omega}(\tilde{u}-u_h)w_h\right|\\
&\quad+\left|\int_{B_h}(w_h\,\mathrm{div}(y-y_h)+(y-y_h)\cdot\nabla w_h)\right|\\
&\quad+\left|\int_{\Gamma_h}\frac{1}{|\nabla\phi_h|}((y-y_h)\cdot\nabla\phi_h+\frac{1}{h}(p-p_h)\phi_h)w_h\right|+\left|\gamma_{div}\int_{\Omega_h^\Gamma}(\mathrm{div}(y-y_h)+\tilde{u}-u_h)w_h\right|\\
&\quad+\left|\gamma_1\int_{\Omega_h^\Gamma}((y-y_h)+\nabla(\tilde{u}-u_h))\cdot\nabla w_h\right|+\left|\sigma h\int_{\Gamma_h^i}\left[\frac{\partial(\tilde{u}-u_h)}{\partial n}\right]\left[\frac{\partial w_h}{\partial n}\right]\right|\\
&\quad+(1+\gamma_{div})\left|\int_{\Omega_h}(\tilde{f}-f)w_h\right|\\
&\leqslant Ch^{k+1}(\|f\|_{k,\Omega_h}+\|\tilde{g}\|_{k+1,\Omega_h^\Gamma})|\tilde{w}|_{2,\Omega_h}+C\,|||\,\tilde{u}-u_h,y-y_h,p-p_h\,|||_h\\
&\quad\times\left(\|w_h\|_{1,\Omega_h\setminus\Omega}+\|w_h\|_{1,\Omega_h^\Gamma}+h\|w_h\|_{0,\Gamma_h}+\sqrt{h}\|[\nabla w_h]\|_{0,\Gamma_h^i}\right)\\
&\quad+C\|\tilde{f}-f\|_{0,\Omega_h\setminus\Omega}\|w_h\|_{1,\Omega_h\setminus\Omega}.
\end{aligned}$$

It remains to bound different norms of w_h featuring in the estimate above. By Lemma 3.3 and interpolation estimates

$$\begin{aligned}
\|w_h\|_{0,\Omega_h\setminus\Omega}&\leq\|\tilde{w}-I_h\tilde{w}\|_{0,\Omega_h\setminus\Omega}+\|\tilde{w}\|_{0,\Omega_h\setminus\Omega}\\
&\leq Ch^2|\tilde{w}|_{2,\Omega_h\setminus\Omega}+C\left(\sqrt{h}\|\tilde{w}\|_{0,\Gamma}+h|\tilde{w}|_{1,\Omega_h\setminus\Omega}\right)\leq C\sqrt{h}\|\tilde{w}\|_{2,\Omega_h}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\|\nabla w_h\|_{0,\Omega_h\setminus\Omega}&\leq\|\nabla(\tilde{w}-I_h\tilde{w})\|_{0,\Omega_h\setminus\Omega}+\|\nabla\tilde{w}\|_{0,\Omega_h\setminus\Omega}\\
&\leq Ch|\tilde{w}|_{2,\Omega_h\setminus\Omega}+C\left(\sqrt{h}\|\nabla\tilde{w}\|_{0,\Gamma}+h|\nabla\tilde{w}|_{1,\Omega_h\setminus\Omega}\right)\leq C\sqrt{h}\|\tilde{w}\|_{2,\Omega_h}.
\end{aligned}$$

Analogous estimates also hold for $\|w_h\|_{1,\Omega_h^\Gamma}$. Moreover, by interpolation estimates,

$$\|[\nabla w_h]\|_{0,\Gamma_h^i}=\|[\nabla(\tilde{w}-I_h\tilde{w})]\|_{0,\Gamma_h^i}\leqslant C\sqrt{h}|\tilde{w}|_{2,\Omega_h}$$

and, by Lemma 3.5,

$$h\|w_h\|_{0,\Gamma_h}\leq C\sqrt{h}\|w_h\|_{0,\Omega_h^\Gamma}\leq C\sqrt{h}\|\tilde{w}\|_{2,\Omega_h}.$$

Hence,

$$\begin{aligned}
\|u-u_h\|^2_{0,\Omega}&\leq Ch^{k+1}(\|f\|_{k,\Omega_h}+\|\tilde{g}\|_{k+1,\Omega_h^\Gamma})|\tilde{w}|_{2,\Omega_h}\\
&\quad+C\sqrt{h}(|||\tilde{u}-u_h,y-y_h,p-p_h\,|||_h+\|\tilde{f}-f\|_{0,\Omega_h\setminus\Omega})\|\tilde{w}\|_{2,\Omega_h}.
\end{aligned}$$

This implies, by (14) and (15),

$$\|u-u_h\|^2_{0,\Omega}\leq Ch^{k+\frac{1}{2}}(\|f\|_{k,\Omega_h}+\|\tilde{g}\|_{k+1,\Omega_h^\Gamma})\|\tilde{w}\|_{2,\Omega_h},$$

which entails the announced L^2-error estimate in $L^2(\Omega)$ since $\|\tilde{w}\|_{2,\Omega_h}\leq C\|u-u_h\|_{0,\Omega}$.

4 Conditioning

We are now going to prove that the condition number of the finite element matrix associated to the bilinear form a_h of ϕ-FEM with Neumann condition is of order $1/h^2$ on a quasi-uniform mesh of step h.

Theorem 4.1. *Under Assumptions 1–3 and recalling that the mesh $\mathcal{T}_h$ is quasi-uniform, the condition number defined by $\kappa(\mathbf{A}) := \|\mathbf{A}\|_2\|\mathbf{A}^{-1}\|_2$ of the matrix $\mathbf{A}$ associated to the bilinear form a_h on $W_h^{(k)}$ satisfies,*

$$\kappa(\mathbf{A}) \leq Ch^{-2}.$$

Here, $\|\cdot\|_2$ stands for the matrix norm associated to the vector 2-norm $|\cdot|_2$.

Proof. The proof is divided into 4 steps:

Step 1. We shall prove for all $q_h \in Q_h^{(k)}$

$$\|q_h\phi_h\|_{0,\Omega_h^\Gamma} \geq Ch\|q_h\|_{0,\Omega_h^\Gamma}. \tag{16}$$

First, take a simplex T of diameter h_T satisfying the regularity assumptions and prove

$$\min_{q_h\neq 0,\phi_h\neq 0} \frac{\|q_h\phi_h\|_{0,T}}{\|q_h\|_{0,T}\|\nabla\phi_h\|_{\infty,T}} \geqslant Ch_T, \tag{17}$$

where the minimum is taken over all polynomials q_h of degree $\leqslant k$ and all polynomials ϕ_h of degree $\leqslant l$ vanishing at least one point on T. Note that this excludes $\|\nabla\phi_h\|_{\infty,T} = 0$ because ϕ_h would vanish identically on T in such an occasion. The minimum in (17) is indeed attained since, by homogeneity, it can be taken over the compact set $\|q_h\|_{0,T} = \|\nabla\phi_h\|_{\infty,T} = 1$. We can see now that (17) is valid thanks to a scaling argument. In particular, $C > 0$. Applying (17) on any mesh element $T \in \mathcal{T}_h^\Gamma$ to any $q_h \in Q_h^{(k)}$ and ϕ_h approximation to ϕ satisfying Assumption 2 leads to

$$\|q_h\phi_h\|_{0,T} \geqslant Ch_T\frac{m}{2}\|q_h\|_{0,T}.$$

Taking the square on both sides and summing over all $T \in \mathcal{T}_h^\Gamma$ yields (16).

Step 2. We shall prove for all $(v_h, z_h, q_h) \in W_h^{(k)}$

$$a_h(v_h, z_h, q_h; v_h, z_h, q_h) \geq c\|v_h, z_h, q_h\|_0^2 \tag{18}$$

with

$$\|v_h, z_h, q_h\|_0^2 = \|v_h\|_{0,\Omega_h}^2 + \|z_h\|_{0,\Omega_h^\Gamma}^2 + \|q_h\|_{0,\Omega_h^\Gamma}^2.$$

Indeed, by Lemma 1,

$$\begin{aligned} a_h(v_h, z_h, q_h; v_h, z_h, q_h) &\geq c|||v_h, z_h, q_h|||_h^2 \\ &\geq c(||v_h||_{1,\Omega_h}^2 + \|z_h + \nabla v_h\|_{0,\Omega_h^\Gamma}^2 + \|z_h\cdot\nabla\phi_h + \frac{1}{h}q_h\phi_h\|_{0,\Omega_h^\Gamma}^2). \end{aligned}$$

We have assumed here (without loss of generality) $h \leq 1$. By Young's inequality with any $\epsilon_1 \in (0,1)$,

$$\|z_h + \nabla v_h\|_{0,\Omega_h^\Gamma}^2 = \|z_h\|_{0,\Omega_h^\Gamma}^2 + \|\nabla v_h\|_{0,\Omega_h^\Gamma}^2 + 2(z_h, \nabla v_h)_{0,\Omega_h^\Gamma} \geq (1-\epsilon_1)\|z_h\|_{0,\Omega_h^\Gamma}^2 - \frac{1-\epsilon_1}{\epsilon_1}\|\nabla v_h\|_{0,\Omega_h^\Gamma}^2. \tag{19}$$

Similarly, for any $\epsilon_2 \in (0,1)$, using that $\nabla\phi_h$ is uniformly bounded,

$$\|z_h\cdot\nabla\phi_h + \frac{1}{h}q_h\phi_h\|_{0,\Omega_h^\Gamma}^2 \geq \frac{1-\epsilon_2}{h^2}\|\phi_h q_h\|_{0,\Omega_h^\Gamma}^2 - C\frac{1-\epsilon_2}{\epsilon_2}\|z_h\|_{0,\Omega_h^\Gamma}^2. \tag{20}$$

Thus, combining (19), (20) and (16),

$$a_h(v_h,z_h,q_h;v_h,z_h,q_h) \geq c\left(\left(1-\frac{1-\epsilon_1}{\epsilon_1}\right)||v_h||^2_{1,\Omega_h} + \left(1-\epsilon_1 - C\frac{1-\epsilon_2}{\epsilon_2}\right)\|z_h\|^2_{0,\Omega_h^\Gamma} \right.$$
$$\left. + C(1-\epsilon_2)\|q_h\|^2_{0,\Omega_h^\Gamma}\right).$$

Taking ϵ_1, ϵ_2 close to 1, we get (18).

Step 3. We shall prove for all $(v_h, z_h, q_h) \in W_h^{(k)}$

$$a_h(v_h,z_h,q_h;v_h,z_h,q_h) \leq \frac{C}{h^2}\|v_h,z_h,q_h\|_0^2. \tag{21}$$

By definition of a_h (9) and Cauchy-Schwarz inequality,

$$a_h(v_h,z_h,q_h;v_h,z_h,q_h) \leq C\left(\|v_h\|^2_{1,\Omega_h} + \|z_h\|^2_{1,\Omega_h^\Gamma} + h\left\|\left[\frac{\partial v_h}{\partial n}\right]\right\|^2_{0,\Gamma_h^i} + \frac{1}{h^2}\|z_h\cdot\nabla\phi_h\|^2_{0,\Omega_h^\Gamma}\right.$$
$$\left. + \frac{1}{h^4}\|q_h\phi_h\|^2_{0,\Omega_h^\Gamma}\right)$$

so by inverse inequalities and the fact that both $\nabla\phi_h$ and $\frac{1}{h}\phi_h$ are uniformly bounded on Ω_h^Γ,

$$a_h(v_h,z_h,q_h;v_h,z_h,q_h) \leq C\left(\frac{1}{h^2}\|v_h\|^2_{0,\Omega_h} + \frac{1}{h^2}\|z_h\|^2_{0,\Omega_h^\Gamma} + \frac{1}{h^2}\|q_h\|^2_{0,\Omega_h^\Gamma}\right).$$

Step 4. Denote the dimension of $W_h^{(k)}$ by N and let us associate any $(v_h, z_h, q_h) \in W_h^{(k)}$ with the vector $\tilde{\mathbf{v}} \in \mathbb{R}^N$ containing the expansion coefficients of $(v_h, z_h, q_h)'$ in the standard finite element basis. Recalling that the mesh is quasi-uniform and using the equivalence of norms on the reference element, we can easily prove that

$$C_1 h^d |\tilde{\mathbf{v}}|_2^2 \leq \|v_h,z_h,q_h\|_0^2 \leq C_2 h^d |\tilde{\mathbf{v}}|_2^2. \tag{22}$$

The bounds (22) and (21) imply

$$\|\mathbf{A}\|_2 = \sup_{\tilde{\mathbf{v}}\in\mathbb{R}^N}\frac{(\mathbf{A}\tilde{\mathbf{v}},\tilde{\mathbf{v}})}{|\tilde{\mathbf{v}}|_2^2} = \sup_{\tilde{\mathbf{v}}\in\mathbb{R}^N}\frac{a_h(v_h,z_h,q_h;v_h,z_h,q_h)}{|\tilde{\mathbf{v}}|_2^2}$$
$$\leq Ch^d \sup_{(v_h,z_h,q_h\in W_h^{(k)}} \frac{a_h(v_h,z_h,q_h;v_h,z_h,q_h)}{\|v_h,z_h,q_h\|_0^2} \leq Ch^{d-2}.$$

Similarly, (22) and (18) imply

$$\|\mathbf{A}^{-1}\|_2 = \sup_{\tilde{\mathbf{v}}\in\mathbb{R}^N}\frac{|\tilde{\mathbf{v}}|_2^2}{(\mathbf{A}\tilde{\mathbf{v}},\tilde{\mathbf{v}})} = \sup_{\tilde{\mathbf{v}}\in\mathbb{R}^N}\frac{|\tilde{\mathbf{v}}|_2^2}{a_h(v_h,z_h,q_h;v_h,z_h,q_h)}$$
$$\leq \frac{C}{h^d}\sup_{(v_h,z_h,q_h)\in W_h^{(k)}}\frac{\|v_h,z_h,q_h\|_0^2}{a_h(v_h,z_h,q_h;v_h,z_h,q_h)} \leq \frac{C}{h^d}.$$

These estimates lead to the desired result. □

5 Numerical simulations

In this section, we illustrate ϕ-FEM on three different test cases, cf. Fig. 1, exploring the errors with respect to exact "manufactured" solutions. The numerical results for the 1st test case (in 2D) confirm the predicted theoretical estimates (in fact, better than theoretically predicted convergence rate is observed for the L^2-error). In the 2nd test case (also in 2D), we show that the optimal convergence is recovered even when the level-set function ϕ is less regular than assumed by the theory. Our method is also compared with CutFEM [5] in this case. Finally, a 3D example is given in the 3rd test case. The surrounding domains $\mathcal{O}$ are always chosen as boxes aligned with the Cartesian coordinates and the background meshes $\mathcal{T}_h^{\mathcal{O}}$ are obtained from uniform Cartesian grids, dividing the cells into the simplexes (semi-cross meshes in 2D) . We always use the numerical quadrature of a high enough order so that all the integrals in (5) are computed exactly.

We have implemented ϕ-FEM both in `FreeFEM` (a general purpose FEM software, see [9]) and in `multiphenics` [1], a python library that aims at providing tools in `FEniCS` [2] for an easy prototyping of multiphysics problems on conforming meshes. Both implementations give the same results in our 2D test cases and we present here only those obtained with `FreeFEM`. Unfortunately, it is not possible to fully implement ϕ-FEM in 3D using `FreeFEM` since it does not provide the tools to deal with the jumps on inter-element faces of a 3D mesh. That is why the numerical results for our 3rd test case were produced using `multiphenics` only. The implementation scripts can be consulted on GitHub.[3]

Remark 1. *Our method (5) features "mixed" terms, such as $\gamma_1 \int_{\Omega_h^\Gamma} (y_h + \nabla u_h) \cdot (z_h + \nabla v_h)$ for $u_h, v_h \in V_h$ and $y_h, z_h \in Z_h$, which contain finite element functions defined on two different meshes, namely $\mathcal{T}_h$ and $\mathcal{T}_h^\Gamma$. Of course, $\mathcal{T}_h^\Gamma$ is a submesh of $\mathcal{T}_h$ so that interpolating u_h, v_h from $\mathcal{T}_h$ to $\mathcal{T}_h^\Gamma$ amounts just to renumbering the degrees of freedom. However, this operation is not necessarily available in all the FEM softwares. This is the case of `FEniCS`, for example, where all the finite elements involved in a problem should be defined on the same mesh. Fortunately, the `multiphenics` library, although based on `FEniCS`, does not have such a restriction. It can be thus used to implement ϕ-FEM rather straightforwardly. As for `FreeFEM`, it naturally supports the interpolations on several meshes so that our method (5) can be implemented there exactly as it is written on paper. However, we have discovered that a straightforward implementation involving an implicit interpolation from $\mathcal{T}_h$ to $\mathcal{T}_h^\Gamma$ can lead to sub-optimal results and to some spurious oscillations in the error curves. This is certainly due to some imperfections of the interpolation algorithm. Much better results are obtained if we introduce explicitely the interpolation matrix I_h from $V_h^{(k)}$ to $V_h^{(k),\Gamma}$, the restriction of $V_h^{(k)}$ to $\mathcal{T}_h^\Gamma$, using the `FreeFEM` function `interpolate`. We thus compute the matrices of the "mixed" terms using the numerical integration on $\mathcal{T}_h^\Gamma$, putting there only the finite element spaces defined on $\mathcal{T}_h^\Gamma$, and then multiply these matrices by I_h. This trick leading to an implementation, which is both more efficient and more robust, has been used in all the simulations presented below.*

5.1 1st test case

Domain Ω (see Fig. 1 left) is defined by the level-set function ϕ given in the polar coordinates (r, θ) by

$$\phi(r, \theta) = r^4(5 + 3\sin(7(\theta - \theta_0) + 7\pi/36))/2 - R^4, \tag{23}$$

where $R = 0.47$ and $\theta_0 \in [0, 2\pi)$. The surrounding domain $\mathcal{O}$ is fixed to $(-0.5, 0.5)^2$. Varying the angle θ_0 results in a rotation of Ω, so that the boundary Γ cuts the triangles of the background mesh in a different manner, creating sometimes the "dangerous" situations when certain mesh triangles of $\mathcal{T}_h$ have only a tiny portion inside the physical domain Ω.

We use ϕ-FEM to solve numerically Poisson-Neumann problem (1) with the exact solution

$$u(x, y) = \sin(x)\exp(y). \tag{24}$$

[3] `https://github.com/michelduprez/PhiFEM-Neumann`

The Neumann boundary condition is extrapolated to a vicinity of Γ by

$$\tilde{g} = \frac{\nabla u \cdot \nabla \phi}{|\nabla \phi|} + u\phi. \tag{25}$$

The addition of $u\phi$ above does not perturb $\tilde{g}$ on Γ. Its purpose is to mimick the real life situation where g is known on Γ only and $\tilde{g}$ is some extension of g, not necessarily the natural one $\nabla u \cdot \nabla \phi / |\nabla \phi|$.

We report at Figs. 2 and 3 the evolution of the relative error under the mesh refinement for a fixed position of Ω ($\theta_0 = 0$), using finite element spaces $W_h^{(k)}$ with $k = 1$ ($\mathbb{P}_1$ FE for u_h) and $k = 2$ ($\mathbb{P}_2$ FE for u_h). We also try there different values of l, the degree of finite element used to approximate the level-set ϕ, recalling that it should be chosen as $k+1$ or greater. The experiments reported in these figures confirms the optimal convergence order of the method in both H^1 and L^2 norms (orders k and $k+1$ respectively). The convergence order in the L^2 norm is thus better than in theory. An interesting experimental observation comes from exploring the degree l: while the lowest possible value $l = k + 1$ ensures indeed the optimal convergence orders, it seems advantageous to increase the degree to $l = k + 2$, leading to more accurate results, especially in the L^2 norm. Another series of experiments is reported at Figs. 4 and 5. We explore there the errors with respect to the rotation of Ω over the background mesh (varying θ_0). We restrict ourselves here with finite elements degree $k = 1$ but compare two different values of l: $l = k + 1 = 2$ at Fig. 4 vs. $l = k + 2 = 3$ at Fig. 5. We observe again an advantage of the choice $l = k + 2$: the oscillations on any given background mesh become less important when increasing l and fade away under the mesh refinement in the case $l = k+2$ (this concerns mostly the L^2 errors; the H^1 errors are pretty much the same in both cases). The influence of the parameters σ, γ_{div}, γ_1, γ_2 on the accuracy of the method is explored by the numerical experiments reported at Figs. 6 and 7. Although a full assessment of the role of all the 4 parameters is difficult (we have chosen, somewhat arbitrarily, two scenarios of parameter variations out of endless other possibilities), the conclusion of our numerical experiments seems clear: the method is not sensible to variation of the parameters in the wide range from 10^{-6} to 10, and there is no need to take these parameters greater than 10. Finally, we report at Fig. 8 evolution of the condition number of the ϕ-FEM matrix under the mesh refinement and also its sensitivity with respect to the rotations of Ω. The theoretically predicted behaviour of $\sim 1/h^2$ is confirmed. The conditioning of the method is also found to be rather insensitive to the position of Ω over the mesh.

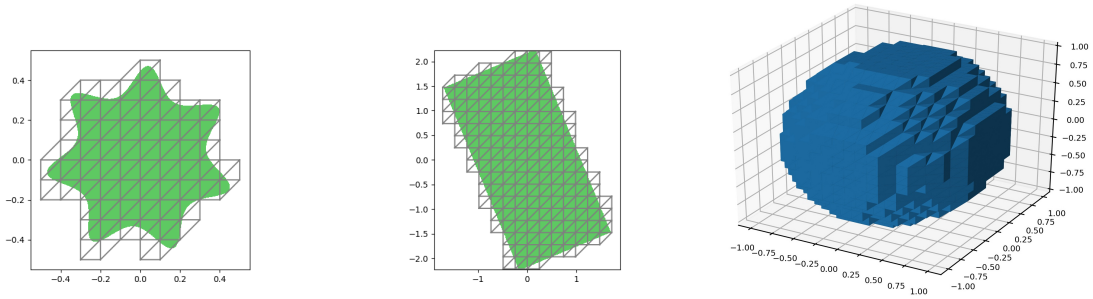

Figure 1: Domains and meshes considered in ϕ-FEM for the test case 1 (left), test case 2 (center) and test case 3 (right).

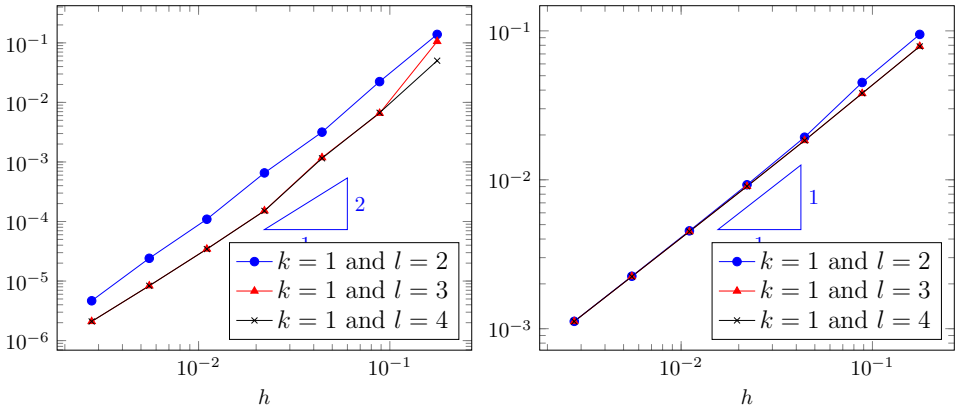

Figure 2: ϕ-FEM for the test case 1 (see (23)–(24)), $\theta_0 = 0$, $\sigma = 0.01$ and $\gamma_1 = \gamma_2 = \gamma_{\text{div}} = 10$, $k = 1$ and different values of l. Left: L^2 relative error $\|u - u_h\|_{0,\Omega_h^i}/\|u\|_{0,\Omega_h^i}$; Right: H^1 relative error $\|u - u_h\|_{1,\Omega_h^i}/\|u\|_{1,\Omega_h^i}$.

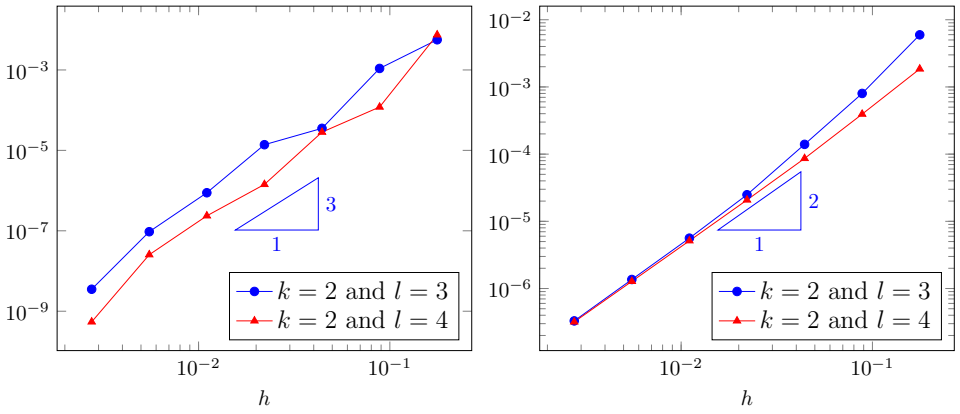

Figure 3: ϕ-FEM for the test case 1 (see (23)–(24)), $\theta_0 = 0$, $\sigma = 0.01$ $\gamma_1 = \gamma_2 = \gamma_{\text{div}} = 10$, $k = 2$ and different values of l. Left: L^2 relative error $\|u - u_h\|_{0,\Omega_h^i}/\|u\|_{0,\Omega_h^i}$; Right: H^1 relative error $\|u - u_h\|_{1,\Omega_h^i}/\|u\|_{1,\Omega_h^i}$.

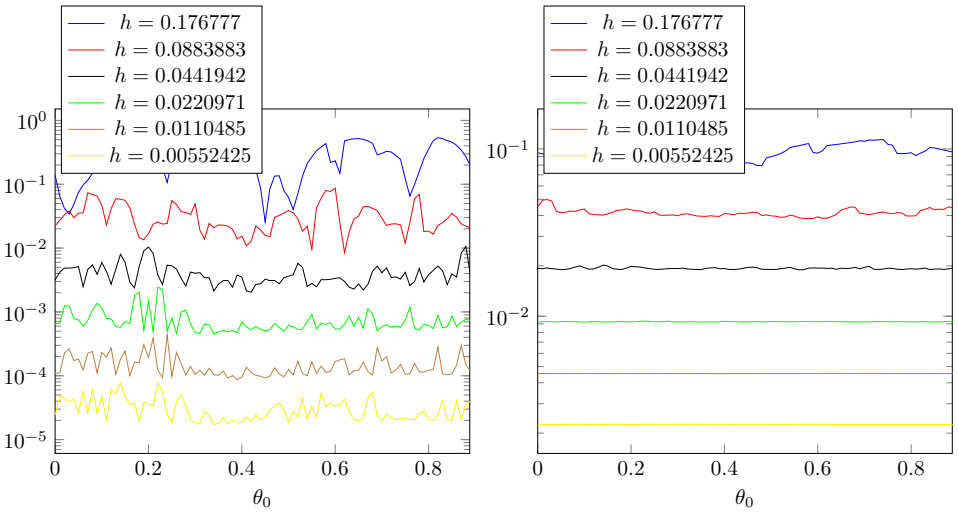

Figure 4: Sensitivity of the relative error with respect to θ_0 in ϕ-FEM for the test case 1 (see (23)–(24)), $\sigma = 0.01$ and $\gamma_1 = \gamma_2 = \gamma_{\text{div}} = 20$, $k = 1$ and $l = 2$. Left: L^2 relative error $\|u - u_h\|_{0,\Omega_h^i} / \|u\|_{0,\Omega_h^i}$; Right: H^1 relative error $\|u - u_h\|_{1,\Omega_h^i} / \|u\|_{1,\Omega_h^i}$.

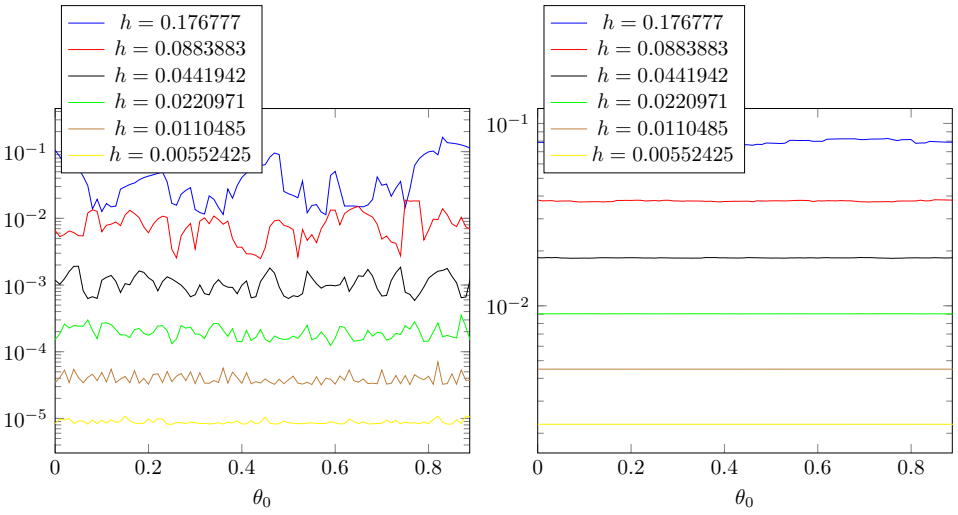

Figure 5: Sensitivity of the relative error with respect to θ_0 in ϕ-FEM for the test case 1 (see (23)–(24)), $\sigma = 0.01$ and $\gamma_1 = \gamma_2 = \gamma_{\text{div}} = 10$, $k = 1$ and $l = 3$. Left: L^2 relative error $\|u - u_h\|_{0,\Omega_h^i} / \|u\|_{0,\Omega_h^i}$; Right: H^1 relative error $\|u - u_h\|_{1,\Omega_h^i} / \|u\|_{1,\Omega_h^i}$.

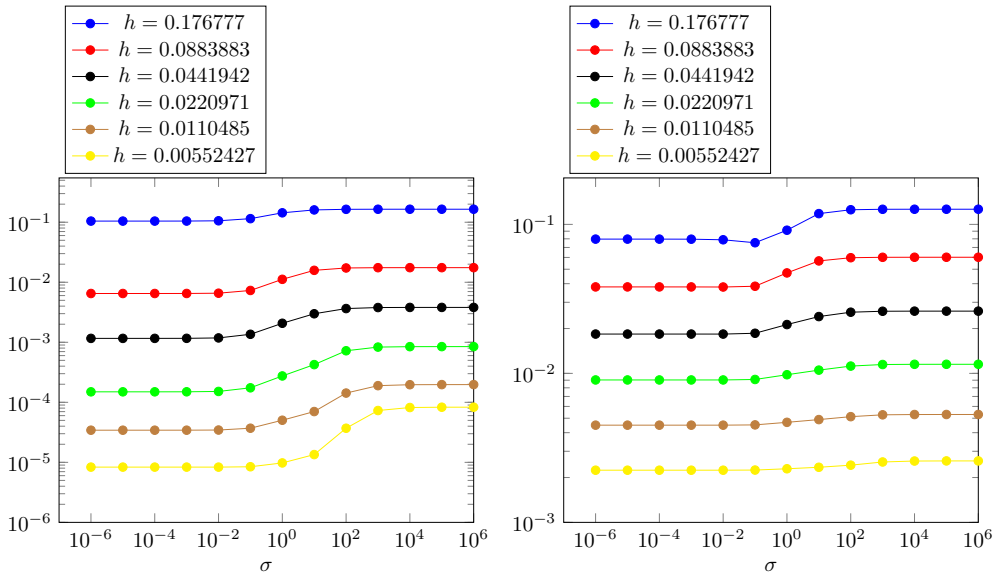

Figure 6: Sensitivity of the relative error in ϕ-FEM with respect to σ with $\gamma_1 = \gamma_2 = \gamma_{div} = 10$ being fixed for the test case 1 (see (23)–(24)), $\theta_0 = 0$, $k = 1$ and $l = 3$. Left: L^2 relative error; Right: H^1 relative error.

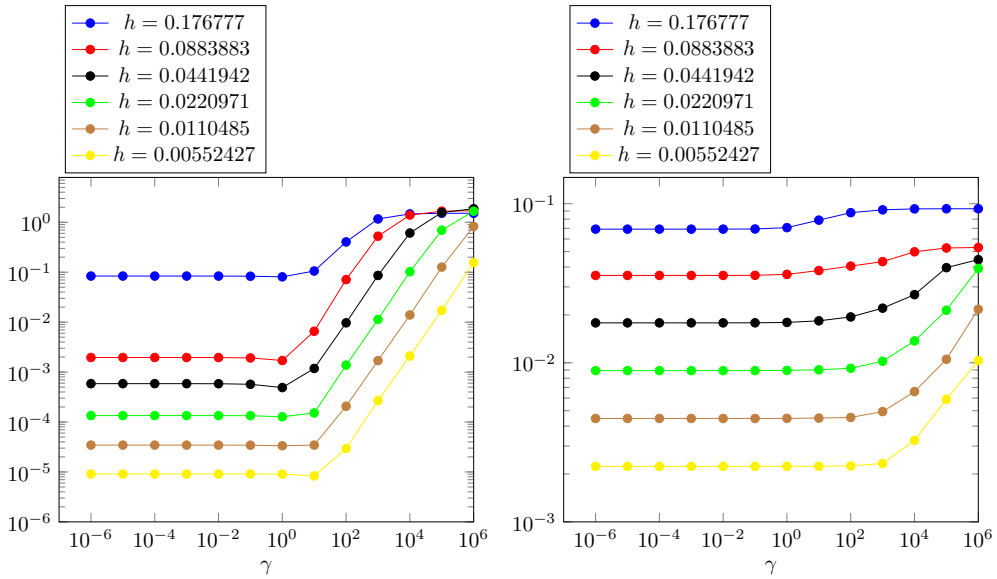

Figure 7: Sensitivity of the relative error in ϕ-FEM with respect to $\gamma_1 = \gamma_2 = \gamma_{div} = \gamma$ with $\sigma = 0.01$ fixed for the test case 1 (see (23)–(24)), $\theta_0 = 0$, $k = 1$ and $l = 3$. Left: L^2 relative error; Right: H^1 relative error.

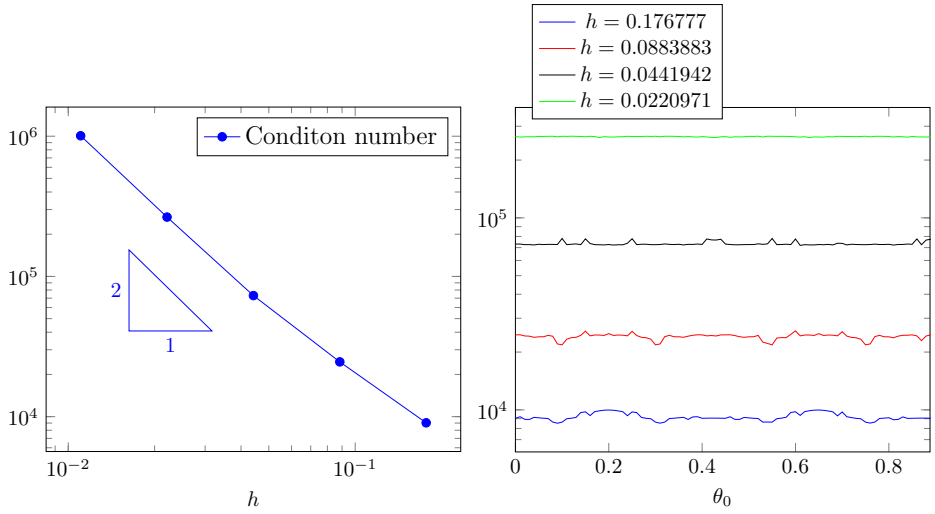

Figure 8: Condition number in ϕ-FEM for the test case 1 (see (23)–(24)), $\sigma = 0.01$, $\gamma_1 = \gamma_2 = \gamma_{\text{div}} = 10$, $\theta_0 = 0$, $k = 1$ and $l = k + 2$. Left: $\theta_0 = 0$; Right: different values of θ_0.

5.2 2nd test case

In this test case, the domain Ω is the rectangle $(-1, 1) \times (-2, 2)$ rotated by an angle θ_0 counter-clockwise around the origin. It is defined by the level-set function ϕ given by

$$\phi(x, y) = \Phi \circ \Pi(x, y), \tag{26}$$

with

$$\Phi(x, y) = \max(|x|, |y|/2) - 1 \text{ and } \Pi \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos(\theta_0) & -\sin(\theta_0) \\ \sin(\theta_0) & \cos(\theta_0) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The surrounding domain is taken as $\mathcal{O} = (-R, R)^2$, with $R = 1.1\sqrt{5}$, cf. Fig. 1 middle.

We use ϕ-FEM to solve numerically Poisson-Neumann problem (1) with the exact solution given by

$$u(x, y) = U \circ \Pi(x, y), \tag{27}$$

where

$$U(x, y) = \cos(\pi x) \cos(\pi y/2).$$

We have thus $g = 0$ for the Neumann boundary condition in (1) so that we can take the natural extension $\tilde{g} = 0$ in (5). The results are presented at Figs. 9 (left) and 10, first choosing a fixed inclination angle $\theta_0 = \pi/8$, and then varying θ_0 from 0 to $2\pi/7$. The numerical tests show again the optimal convergence of ϕ-FEM with $\mathbb{P}_1$ finite elements in the L^2 and H^1 norms, notwithstanding the fact that the level-set function ϕ is less regular than assumed in our theoretical results. Note that we have used here the FE of degree $l = 3$ to represent the level-set, which is higher than the minimal degree $k + 1 = 2$ suggested by the theory. The situation is here similar to that of the test case 1: the implementation using the lower degree $l = 2$ elements (not reported here) is also optimally convergent but turns out to be less robust than $l = 3$ wih respect to the placement of Ω over the mesh (higher oscillations, especially in the L^2 error, when varying θ_0).

We have also compared our method with CutFEM [5]: Find $u_h \in V_h^{(k)}$ s.t.

$$\int_\Omega \nabla u_h \cdot \nabla v_h + \int_\Omega u_h v_h + \sigma h \sum_{E \in \mathcal{F}^\Gamma} \int_E \left[\frac{\partial u_h}{\partial n}\right] \left[\frac{\partial v_h}{\partial n}\right] = \int_\Omega f v_h + \int_\Gamma g v_h \quad \forall\ v_h \in V_h^{(k)},$$

where

$$\mathcal{F}^{\Gamma} = \{E(\text{internal facet of } \mathcal{T}_h) \text{ such that } \exists T \in \mathcal{T}_h : T \cap \Gamma \neq \varnothing \text{ and } E \in \partial T\}.$$

The results are reported at Figs. 9 (right, the simulation at fixed inclination angle θ_0) and 11 (simulations with the rotating domain Ω). Comparing two parts of Fig. 9, we conclude that ϕ-FEM and CutFEM are both optimally convergent and produce very similar results. However, looking closer at Figs. 10 and 11, we can point out an advantage of the ϕ-FEM over the CutFEM: the former seems more robust with respect to the position of Ω over the background mesh, the oscillations of the L^2-errors with rotating the domain are more pronounced for the latter method (the H^1 errors are almost the same in both cases).

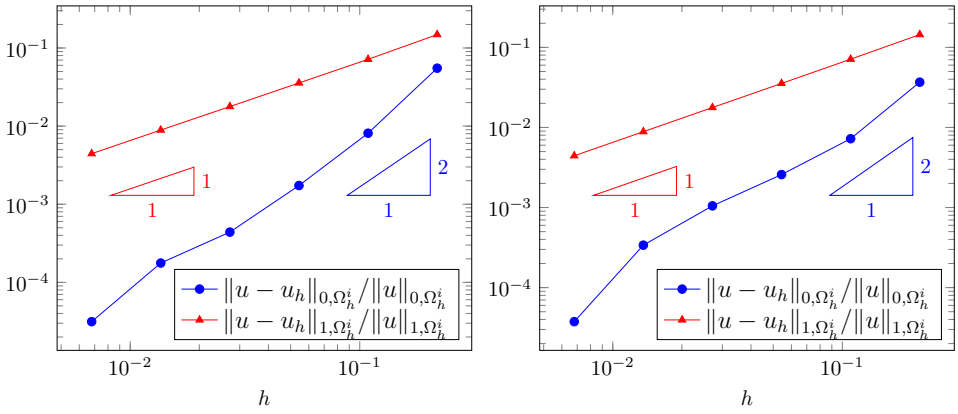

Figure 9: L^2 and H^1 relative error for the test case 2 (see (26)-(27)). Left: ϕ-FEM with $\sigma = 0.01$, $\gamma_1 = \gamma_2 = \gamma_{\text{div}} = 10$, $k = 1$ and $l = 3$; Right: CutFEM , $\theta_0 = \pi/8$, $\sigma = 0.01$.

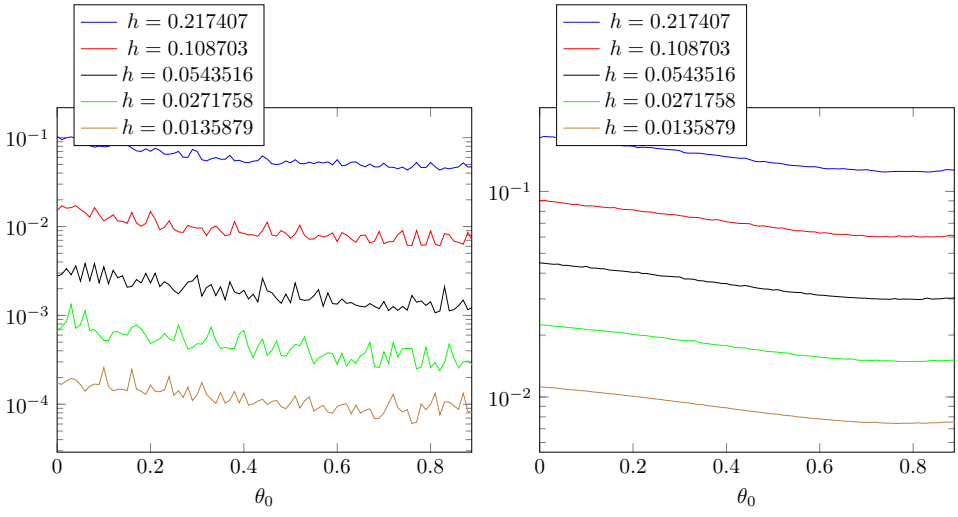

Figure 10: Sensitivity of the relative error with respect to θ_0 in ϕ-FEM for the test case 2 (see (26)-(27)), $\sigma = 0.01$, $\gamma_1 = \gamma_2 = \gamma_{\text{div}} = 10$, $k = 1$ and $l = 3$. Left: L^2 relative error $\|u-u_h\|_{0,\Omega_h^i}/\|u\|_{0,\Omega_h^i}$; Right: H^1 relative error $\|u-u_h\|_{1,\Omega_h^i}/\|u\|_{1,\Omega_h^i}$.

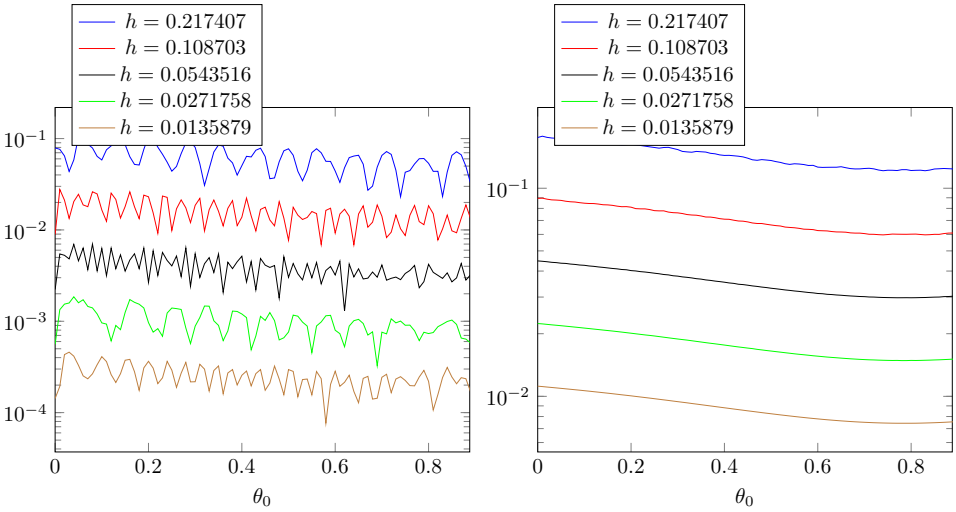

Figure 11: Sensitivity of the relative error with respect to θ_0 in CutFEM for the test case 2 (see (26)-(27)), $\sigma = 0.01$. Left: L^2 relative error $\|u - u_h\|_{0,\Omega_h^i}/\|u\|_{0,\Omega_h^i}$; Right: H^1 relative error $\|u - u_h\|_{1,\Omega_h^i}/\|u\|_{1,\Omega_h^i}$.

5.3 3rd test case

We here take $\Omega \subset \mathbb{R}^3$ as the ball of radius $R = 0.75$ centered at the origin encapsulated into the box $\mathcal{O} = (-1, 1)^3$. Ω is defined by the level-set function

$$\phi(x, y, z) = x^2 + y^2 + z^2 - R^2. \tag{28}$$

Fig. 1 right gives an example of mesh $\mathcal{T}_h$ for this test case. We choose the exact solution as

$$u(x, y, z) = \cos\left(\sqrt{x^2 + y^2 + z^2}\right). \tag{29}$$

The Neumann boundary condition is extrapolated to a vicinity of Γ by (25). Again, we observe in Fig. 12 the optimal orders of convergence for the L^2 and H^1 errors and the expexted behaviour of the condition number $\sim 1/h^2$.

6 Conclusions and outlook

The numerical results from the last section confirm the theoretically predicted optimal convergence of ϕ-FEM in the H^1 semi-norm. The convergence in the L^2 norm turns out to be also optimal, which is better than the theoretical prediction. We have thus an easily implementable optimally convergent finite element method for the Poisson problem with Neumann boundary conditions, suitable for non-fitted meshes and robust with respect to the cuts of the mesh by the domain boundary. The method is straightforward to generalize to Robin boundary conditions. We also recall that the case of homogeneous Dirichlet boundary conditions was treated in [7] via the ansatz $u \approx \phi_h w_h$ with ϕ_h the same as in the present paper and w_h the new unknown searched in a finite element space on $\mathcal{T}_h$. Inspired by the ideas of the present contribution, we could now propose an alternative treatment of non-homogeneous Dirichlet boundary conditions $u = g$ by searching an approximation u_h directly in a finite element space on $\mathcal{T}_h$, introducing an auxiliary unknown p_h in a finite element space on $\mathcal{T}_h^\Gamma$, and enforcing $u_h \approx \phi_h p_h + \tilde{g}$ on $\mathcal{T}_h^\Gamma$. This idea could be also applied to problems with mixed boundary conditions (Dirichlet on a portion of the boundary, Neumann or Robin on the remaining part).

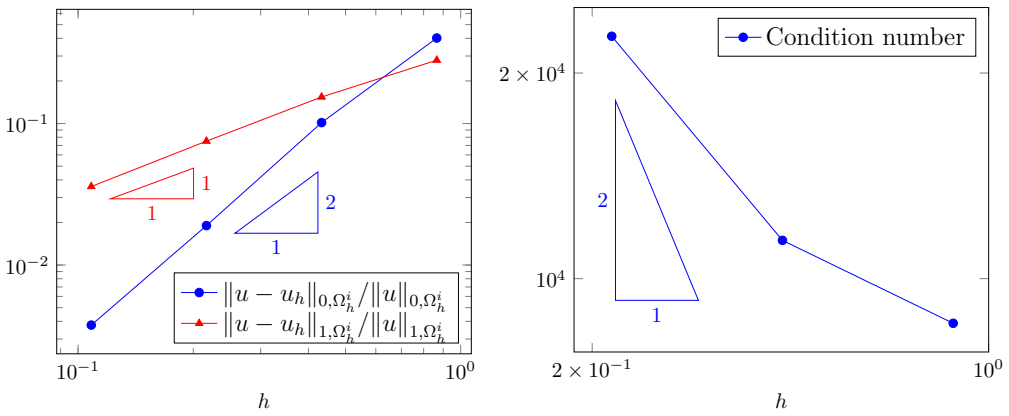

Figure 12: ϕ-FEM for the test case 3 (see (28)-(29)), $\sigma = 0.01$, $\gamma_1 = \gamma_2 = \gamma_{\text{div}} = 10$, $k = 1$ and $l = 3$. Left: L^2 and H^1 relative error; Right: condition number.

Future endeavors should be devoted to more complicated governing equations: time-dependent and non-linear problems, fluid-structure interaction, etc. Another important open question about ϕ-FEM is the effect of "under-integrating", i.e. lowering the quadrature order, on the accuracy of the method. Indeed, the polynomial degrees involved in ϕ-FEM are higher than those in the standard FEM. The quadrature rules of high enough order have been used in the implementation presented in this paper. However, such quadrature rules are not necessarily available in all the FEM libraries.

References

[1] multiphenics. `https://mathlab.sissa.it/multiphenics`.

[2] L. Anders, M. Kent-Andre, G. N. Wells, and al. *Automated Solution of Differential Equations by the Finite Element Method*. Springer, 2012.

[3] E. Burman, S. Claus, P. Hansbo, M. Larson, and A. Massing. Cutfem: discretizing geometry and partial differential equations. *International Journal for Numerical Methods in Engineering*, 104(7):472–501, 2015.

[4] E. Burman and P. Hansbo. Fictitious domain finite element methods using cut elements: I. A stabilized Lagrange multiplier method. *Computer Methods in Applied Mechanics and Engineering*, 199(41):2680–2686, 2010.

[5] E. Burman and P. Hansbo. Fictitious domain finite element methods using cut elements: Ii. A stabilized Nitsche method. *Applied Numerical Mathematics*, 62(4):328–341, 2012.

[6] E. Burman and P. Hansbo. Fictitious domain methods using cut elements: Iii. a stabilized nitsche method for stokes' problem. *ESAIM: Mathematical Modelling and Numerical Analysis*, 48(3):859–874, 2014.

[7] M. Duprez and A. Lozinski. ϕ-FEM: a finite element method on domains defined by level-sets. *SIAM J. Num. Anal., to appear; preprint on arXiv:1903.03703 [math.NA]*, 2020.

[8] J. Haslinger and Y. Renard. A new fictitious domain approach inspired by the extended finite element method. *SIAM Journal on Numerical Analysis*, 47(2):1474–1499, 2009.

[9] F. Hecht. New development in FreeFem++. *J. Numer. Math.*, 20(3-4):251–265, 2012.

[10] A. Lozinski. CutFEM without cutting the mesh cells: a new way to impose Dirichlet and Neumann boundary conditions on unfitted meshes. *Comput. Methods Appl. Mech. Engrg.*, 356:75–100, 2019.

[11] N. Moës, E. Béchet, and M. Tourbier. Imposing dirichlet boundary conditions in the extended finite element method. *International Journal for Numerical Methods in Engineering*, 67(12):1641–1669, 2006.

[12] N. Moës, J. Dolbow, and T. Belytschko. A finite element method for crack growth without remeshing. *International journal for numerical methods in engineering*, 46(1):131–150, 1999.

[13] S. Osher and R. Fedkiw. *Level set methods and dynamic implicit surfaces*, volume 153 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2003.

[14] N. Sukumar, D. Chopp, N. Moës, and T. Belytschko. Modeling holes and inclusions by level sets in the extended finite-element method. *Computer methods in applied mechanics and engineering*, 190(46-47):6183–6200, 2001.