ϕ-FEM: a finite element method on domains defined by level-sets
Michel Duprez, Alexei Lozinski

To cite this version:
Michel Duprez, Alexei Lozinski. ϕ-FEM: a finite element method on domains defined by level-sets. SIAM Journal on Numerical Analysis, 2020, 58 (2), 10.1137/19M1248947. hal-02521111

HAL Id: hal-02521111
https://hal.science/hal-02521111
Submitted on 27 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract

We propose a new fictitious domain finite element method, well suited for elliptic problems posed in a domain given by a level-set function without requiring a mesh fitting the boundary. To impose the Dirichlet boundary conditions, we search the approximation to the solution as a product of a finite element function with the given level-set function, which is also approximated by finite elements. 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 consider the Poisson equation discretized with piecewise polynomial Lagrange finite elements of any order and prove the optimal convergence of our method in the $H^1$-norm. Moreover, the discrete problem is proven to be well conditioned, i.e. the condition number of the associated finite element matrix is of the same order as that of a standard finite element method on a comparable conforming mesh. Numerical results confirm the optimal convergence in both $H^1$ and $L^2$ norms.

1 Introduction

We consider the Poisson-Dirichlet problem

$$-\Delta u = f \text{ on } \Omega, \quad u = 0 \text{ on } \Gamma,$$

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

$$\Omega := \{ \phi < 0 \} \text{ and } \Gamma := \{ \phi = 0 \}. \quad (2)$$

Such a representation is a popular and useful tool to deal with problems with evolving surfaces or interfaces [17]. In the present article, the level-set function is supposed to be known on $\mathbb{R}^d$, smooth, and to behave near $\Gamma$ as the signed distance to $\Gamma$. We propose a finite element method for the problem above which is easy to implement, does not require a mesh fitted to $\Gamma$, and is guaranteed to converge optimally. Our basic idea is very simple: one cannot impose the Dirichlet boundary conditions in the usual manner since the boundary $\Gamma$ is not resolved by the mesh, but one can search for the approximation to $u$ as a product of a finite element function $w_h$ with the level-set $\phi$ itself: such a product obviously vanishes on $\Gamma$. In order to make this idea work, some stabilization should be added to the scheme as outlined below and explained in detail in the next section. We coin our method $\phi$-FEM in accordance with the tradition of denoting the level-sets by $\phi$.

More specifically, let us assume that $\Omega$ lies inside a simply shaped domain $\mathcal{O}$ (typically a box in $\mathbb{R}^d$) and introduce a quasi-uniform simplicial mesh $\mathcal{T}_h^\mathcal{O}$ on $\mathcal{O}$ (the background mesh). Let $\mathcal{T}_h$ be a submesh of $\mathcal{T}_h^\mathcal{O}$ obtained by getting rid of mesh elements lying entirely outside $\Omega$ (the definition of $\mathcal{T}_h$ will be slightly
changed afterwords). Denote by $\Omega_h$ the domain covered by the mesh $\mathcal{T}_h$ (so that typically $\Omega_h$ is only slightly larger than $\Omega$). Our starting point is the following formal observation: assuming that the right-hand side $f$ is actually well defined on $\Omega_h$, and the solution $u$ can be extended to $\Omega_h$ so that $-\Delta u = f$ on $\Omega_h$, we can introduce the new unknown $w \in H^1(\Omega_h)$ such that $u = \phi w$ and the boundary condition on $\Gamma$ is automatically satisfied. An integration by parts yields then

$$
\int_{\Omega_h} \nabla(\phi w) \cdot \nabla(\phi v) - \int_{\partial \Omega_h} \frac{\partial}{\partial n}(\phi w)\phi v = \int_{\Omega_h} f\phi v, \quad \forall v \in H^1(\Omega_h).
$$

Given a finite element approximation $\phi_h$ to $\phi$ on the mesh $\mathcal{T}_h$ and a finite element space $V_h$ on $\mathcal{T}_h$, one can then try to search for $w_h \in V_h$ such that the equality in (3) with the subscripts $h$ everywhere is satisfied for all the test functions $v_h \in V_h$ and to reconstruct an approximate solution $u_h$ to (1) as $\phi_h w_h$. These considerations are very formal and, not surprisingly, such a method does not work as is. We shall show however that it becomes a valid scheme once a proper stabilization in the vein of the Ghost penalty [3] is added. The details on the stabilization and on the resulting finite element scheme are given in the next section.

Our method shares many features with other finite elements methods on non-matching meshes, such as XFEM [15, 14, 18, 11] or CutFEM [5, 6, 7, 4]. Unlike the present work, the integrals over $\Omega$ are kept in XFEM or CutFEM discretizations, which is cumbersome in practice since one needs to implement the integration on the boundary $\Gamma$ and on parts of mesh elements cut by the boundary. The first attempt to alleviate this practical difficulty was done in [12] with method that does not require to perform the integration on the cut elements, but needs still the integration on $\Gamma$. In the present article, we fully avoid any non trivial numerical integration: all the integration in $\phi$-FEM is performed on the whole mesh elements, and there are no integrals on $\Gamma$. We also note that an easily implementable version of $\phi$-FEM is here developed for $P_k$ finite elements of any order $k \geq 1$. This should be contrasted with the situation in CutFEM where some additional terms should be added in order to achieve the optimal $P_k$ accuracy if $k > 1$, cf. [8]. An alternative approach avoiding non trivial quadrature is presented in a recent work on the shifted boundary method [13]. The optimal convergence with piecewise linear finite elements ($k = 1$) on a non-fitted mesh is achieved there by introducing a truncated Taylor expansion on the approximate boundary.

The article is structured as follows: our $\phi$-FEM method is presented in the next section. We also give there the assumptions on the level-set $\phi$ and on the mesh, and announce our main result: the a priori error estimate for $\phi$-FEM. We work with standard continuous $P_k$ 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.\footnote{Our approach can also be realized using $Q_k$ finite elements on a mesh consisting of rectangles/cubes. The convergence results and proofs can be straightforwardly passed over to this case.} The proofs of these estimates are the subject of Section 3. We contend ourselves with the error analysis pertinent to the $h$-refinement only, i.e. we do not attempt to track the dependence of the constants, appearing in our estimates, on the polynomial degree. In Section 4, we prove that the linear system produced by our method has the condition number of order $1/h^2$, the same as that of a standard finite element method. Numerical illustrations are given in Section 5, including a test case covered by our theory that confirms the theoretical predictions and other test cases going slightly beyond the theoretical framework. Finally, conclusions and perspectives are presented in Section 6.

## 2 Definitions, assumptions, description of $\phi$-FEM, and the main result

We recall that we work with a bounded domain $\Omega \subset \mathcal{O} \subset \mathbb{R}^d$ ($d = 2, 3$) with boundary $\Gamma$ given by a level-set $\phi$ as in (2). We assume that $\phi$ is sufficiently smooth and behaves near $\Gamma$ as the signed distance to $\Gamma$ after an appropriate change of local coordinates. More specifically, we fix an integer $k \geq 1$ and introduce the following...
Assumption 1. The boundary $\Gamma$ can be covered by open sets $\mathcal{O}_i$, $i = 1, \ldots, I$ and one can introduce on every $\mathcal{O}_i$ local coordinates $\xi_1, \ldots, \xi_d$ with $\xi_d = \phi$ such that all the partial derivatives $\partial^a \xi_1/\partial x^a$ and $\partial^a x/\partial \xi_1$ up to order $k + 1$ are bounded by some $C_0 > 0$. Moreover, $\phi$ is of class $C^{k+1}$ on $\mathcal{O}$ and $|\phi| \geq m$ on $\mathcal{O} \setminus \cup_{i=1,\ldots,I} \mathcal{O}_i$ with some $m > 0$.

Let $\mathcal{T}_h^O$ be a quasi-uniform simplicial mesh on $\mathcal{O}$ of mesh size $h$, meaning that $h_T \leq h$ and $\rho(T) \geq \beta h$ for all simplexes $T \in \mathcal{T}_h^O$ with some mesh regularity parameter $\beta > 0$ (here $h_T = \text{diam}(T)$ and $\rho(T)$ is the radius of the largest ball inscribed in $T$). Consider, for an integer $l \geq 1$, the finite element space

$$V_{h,O}^{(l)} = \{ v_h \in H^1(\mathcal{O}) : v_h|_T \in \mathbb{P}_l(T) \forall T \in \mathcal{T}_h^O \}$$

where $\mathbb{P}_l(T)$ stands for the space of polynomials in $d$ variables of degree $\leq l$ viewed as functions on $T$. Introduce an approximate level-set $\phi_h \in V_{h,O}^{(l)}$ by

$$\phi_h := I_{h,O}^{(l)}(\phi)$$

where $I_{h,O}^{(l)}$ is the standard Lagrange interpolation operator on $V_{h,O}^{(l)}$. We shall use this to approximate the physical domain $\Omega = \{ \phi < 0 \}$ with smooth boundary $\Gamma = \{ \phi = 0 \}$ by the domain $\{ \phi_h < 0 \}$ with the piecewise polynomial boundary $\Gamma_h = \{ \phi_h = 0 \}$. We employ $\phi_h$ rather than $\phi$ in our numerical method in order to simplify its implementation (all the integrals in the forthcoming finite element formulation will involve only the piecewise polynomials). This feature will also turn out to be crucial in our theoretical analysis.

We now introduce the computational mesh $\mathcal{T}_h$ as the subset of $\mathcal{T}_h^O$ composed of the triangles/tetrahedrons having a non-empty intersection with the approximate domain $\{ \phi_h < 0 \}$. We denote the domain occupied by $\mathcal{T}_h$ by $\Omega_h$, i.e.

$$\mathcal{T}_h := \{ T \in \mathcal{T}_h^O : T \cap \{ \phi_h < 0 \} \neq \emptyset \} \quad \text{and} \quad \Omega_h = (\bigcup_{T \in \mathcal{T}_h} T)^c.$$ 

Note that we do not necessarily have $\Omega \subset \Omega_h$. Indeed some mesh elements can be cut by the exact boundary $\{ \phi = 0 \}$ but not by the approximate one $\{ \phi_h = 0 \}$. In such rare occasions, a mesh element containing a small portion of $\Omega$ will not be included into $\mathcal{T}_h$.

Fix an integer $k \geq 1$ (the same $k$ as in Assumption 1) and consider the finite element space

$$V_h^{(k)} = \{ v_h \in H^1(\Omega_h) : v_h|_T \in \mathbb{P}_k(T) \forall T \in \mathcal{T}_h \}.$$

The $\phi$-FEM approximation to (1) is introduced as follows: find $w_h \in V_h^{(k)}$ such that:

$$a_h(w_h, v_h) = l_h(v_h) \quad \text{for all } v_h \in V_h^{(k)},$$

where the bilinear form $a_h$ and the linear form $l_h$ are defined by

$$a_h(w, v) := \int_{\Omega_h} \nabla (\phi_h w) \cdot \nabla (\phi_h v) - \int_{\partial \Omega_h} \frac{\partial}{\partial n} (\phi_h w) \phi_h v + G_h(w, v)$$

and

$$l_h(v) := \int_{\Omega_h} f \phi_h v + G_{h}^{rhs}(v),$$

with $G_h$ and $G_{h}^{rhs}$ standing for

$$G_h(w, v) := \sigma h \sum_{E \in \mathcal{E}_h^I} \int_E \left[ \frac{\partial}{\partial n} (\phi_h w) \right] \left[ \frac{\partial}{\partial n} (\phi_h v) \right] + \sigma h^2 \sum_{T \in \mathcal{T}_h} \int_T \Delta (\phi_h w) \Delta (\phi_h v),$$

$$G_{h}^{rhs}(v) := -\sigma h^2 \sum_{T \in \mathcal{T}_h} \int_T f \Delta (\phi_h v),$$

3
where $\sigma > 0$ is an $h$-independent stabilization parameter, $T_h^T \subset T_h$ contains the mesh elements cut by the approximate boundary $\Gamma_h = \{ \phi_h = 0 \}$, i.e.

$$T_h^T = \{ T \in T_h : T \cap \Gamma_h \neq \emptyset \}, \quad \Omega_h^T := \left( \bigcup_{T^T \in T_h^T} T^T \right) \cap \Omega_h.$$  

and $\mathcal{F}_h^T$ collects the interior facets of the mesh $T_h$ either cut by $\Gamma_h$ or belonging to a cut mesh element

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

The brackets inside the integral over $E \in \mathcal{F}_h^T$ in the formula for $G_h$ stand for the jump over the facet $E$. The first part in $G_h$ actually coincides with the ghost penalty as introduced in [3] for $P_1$ finite elements.

We shall also need the following assumptions on the mesh $T_h$, more specifically on the intersection of elements of $T_h$ with the approximate boundary $\Gamma_h = \{ \phi_h = 0 \}$.

**Assumption 2.** The approximate boundary $\Gamma_h$ can be covered by element patches $\{ \Pi_i \}_{i=1,...,N_n}$ having the following properties:

- Each patch $\Pi_i$ is a connected set composed of a mesh element $T_i \in T_h \setminus T_h^T$ and some mesh elements cut by $\Gamma_h$. More precisely, $\Pi_i = T_i \cup \Pi_i^T$ with $\Pi_i^T \subset T_h^T$ containing at most $M$ mesh elements;
- $T_h^T = \bigcup_{i=1}^{N_n} \Pi_i^T$;
- $\Pi_i$ and $\Pi_j$ are disjoint if $i \neq j$.

Assumption 2 is satisfied for $h$ small enough, preventing strong oscillations of $\Gamma$ on the length scale $h$. It can be reformulated by saying that each cut element $T \in T_h^T$ can be connected to an uncut element $T' \in T_h \setminus T_h^T$ by a path consisting of a small number of mesh elements adjacent to one another; see [12] for a more detailed discussion and an illustration (Fig. 2).

In what follows, $\| \cdot \|_k, \mathcal{D}$ (resp. $| \cdot |_k, \mathcal{D}$) denote the norm (resp. the semi-norm) in the Sobolev space $H^k(\mathcal{D})$ with an integer $k \geq 0$ where $\mathcal{D}$ can be a domain in $\mathbb{R}^d$ or a $(d-1)$-dimensional manifold.

**Theorem 2.1.** Suppose that Assumptions 1 and 2 hold true, $l \geq k$, the mesh $T_h$ is quasi-uniform, and $f \in H^k(\Omega_h \cup \Omega)$. Let $u \in H^{k+2}(\Omega)$ be the solution to (1) and $w_h \in V_h^{(k)}$ be the solution to (5). Denoting $u_h := \phi_h w_h$, it holds

$$|u - u_h|_{1, \Omega \cap \Omega_h} \leq C h^k \| f \|_{k, \Omega \cap \Omega_h},$$

with a constant $C > 0$ depending on the $C_0$, $m$, $M$ in Assumptions 1, 2, on the maximum of the derivatives of $\phi$ of order up to $k+1$, on the mesh regularity, and on the polynomial degrees $k$ and $l$, but independent of $h$, $f$, and $u$.

Moreover, supposing $\Omega \subset \Omega_h$

$$\| u - u_h \|_{0, \Omega} \leq C h^{k+1/2} \| f \|_{k, \Omega_h},$$

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

### 3 Proof of the a priori error estimate

The proof of Theorem 2.1 is preceded with auxiliary lemmas in Sections 3.1 and 3.2, followed by the proof of coercivity of the form $a_h$ in Section 3.3.
3.1 A Hardy-type inequality

Lemma 3.1. We assume that the domain $\Omega$ is given by the level-set $\phi$, cf. (2), and satisfies Assumption 1. Then, for any $u \in H^{k+1}(\Omega)$ vanishing on $\Gamma$,

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

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

Proof. The proof is decomposed into three steps:

Step 1. We start in the one dimensional setting and adapt the proof of Hardy’s inequality from [16]. Let $u : \mathbb{R} \to \mathbb{R}$ be a $C^\infty$ function with compact support such that $u(0) = 0$. Set $w(x) = u(x)/x$ for $x \neq 0$ and $w(0) = u'(0)$. We shall prove that $w$ is a $C^\infty$ function on $\mathbb{R}$ and, for any integer $s \geq 0$,

$$
\left( \int_{-\infty}^{\infty} \left| w^{(s)}(x) \right|^2 \, dx \right)^{1/2} \leq C \left( \int_{-\infty}^{\infty} \left| u^{(s+1)}(x) \right|^2 \, dx \right)^{1/2}
$$

with $C$ depending only on $s$.

Observe, for any $x > 0$,

$$
w(x) = \frac{u(x)}{x} = \frac{1}{x} \int_0^x u'(t) \, dt = \int_0^1 u'(xt) \, dt.
$$

Hence,

$$
w^{(s)}(x) = \int_0^1 u^{(s+1)}(xt) t^s \, dt.
$$

It implies $\lim_{x \to 0^+} w^{(s)}(x) = u^{(s+1)}(0)/(s + 1)$. The same formula holds for the limit as $x \to 0^-$. This means that $w$ is continuous (the special case $s = 0$), and $w^{(s)}(0)$ exists for all $s \geq 1$.

We have now by (11) and the integral version of Minkowski’s inequality

$$
\left( \int_0^\infty \left| w^{(s)}(x) \right|^2 \, dx \right)^{1/2} = \left( \int_0^1 \left| \int_0^1 u^{(s+1)}(xt) t^s \, dt \right|^2 \, dx \right)^{1/2}
$$

$$
\leq \int_0^1 \left( \int_0^\infty \left| u^{(s+1)}(xt) \right|^2 \, dx \right)^{1/2} t^s \, dt = C \left( \int_0^\infty \left| u^{(s+1)}(x) \right|^2 \, dx \right)^{1/2}
$$

with $C = \int_0^1 t^{s-1/2} \, dt = 1/(s + 1/2)$. Applying the same argument to negative $x$, we get (10).

Step 2. Let now $u : \mathbb{R}^d \to \mathbb{R}$ be a compactly supported $C^\infty$ function vanishing at $x_d = 0$ and set $w = u/x_d$. We shall prove

$$
|w|_{k,\mathbb{R}^d} \leq C |u|_{k+1,\mathbb{R}^d}
$$

with $C$ depending only on $k$.

To keep things simple, we give here the proof for the case $d = 2$ only (the case $d = 3$ is similar but would involve more complicated notations). Take any integers $t, s \geq 0$ with $t + s = k$, apply (10) to $\frac{\partial^t w}{\partial x_1^t} = \frac{1}{x_2} \frac{\partial^t u}{\partial x_1^t}$ treated as a function of $x_2$ (note that $\frac{\partial^t u}{\partial x_2^t}$ vanishes at $x_2 = 0$) and then integrate with respect to $x_1$. This gives

$$
\left\| \frac{\partial^k w}{\partial x_1^k \partial x_2^s} \right\|_{0,\mathbb{R}^d} \leq C \left\| \frac{\partial^{k+1} u}{\partial x_1^{k+1} \partial x_2^s} \right\|_{0,\mathbb{R}^d}.
$$

Thus,

$$
|w|_{k,\mathbb{R}^d}^2 = \sum_{s=0}^k \left\| \frac{\partial^k w}{\partial x_1^s \partial x_2^s} \right\|_{0,\mathbb{R}^d}^2 \leq C^2 \sum_{s=0}^k \left\| \frac{\partial^{k+1} u}{\partial x_1^{k-s} \partial x_2^{s+1}} \right\|_{0,\mathbb{R}^d}^2 \leq C^2 |u|_{k+1,\mathbb{R}^d}^2
$$

5
so that (12) is proved.

Step 3. Consider finally the domains $\Omega \subset \mathcal{O}$ as announced in the statement of this Lemma, let $u$ be a $C^\infty$ function on $\mathcal{O}$ vanishing on $\Gamma$, and set $w = u/\phi$. Assume first that $u$ is compactly supported in $\mathcal{O}_1$, one of the sets forming the cover of $\Gamma$ as announced in Assumption 1. Recall the local coordinated $\xi_1, \ldots, \xi_d$ on $\mathcal{O}_1$ with $\xi_d = \phi$ and denote by $\hat{u}$ (resp. $\hat{w}$) the function $u$ (resp. $w$) treated as a function of $\xi_1, \ldots, \xi_d$. Since $\hat{w} = \hat{u}/\xi_d$, (12) implies $\|\hat{w}\|_{k,R^d} \leq C\|\hat{u}\|_{k+1,R^d}$. Passing from the coordinates $x_1, \ldots, x_d$ to $\xi_1, \ldots, \xi_d$ and backwards we conclude $\|w\|_{k,\mathcal{O}} \leq C\|u\|_{k+1,\mathcal{O}}$ with a constant $C$ that depends on the maximum of partial derivatives $\partial^\alpha x/\partial \xi^\alpha$ up to order $k$ and that of $\partial^\alpha \xi/\partial x^\alpha$ up to order $k + 1$. Introducing a partition of unity subject to the cover $\{\mathcal{O}_i\}$ we can now easily prove $\|w\|_{k,\mathcal{O}} \leq C\|u\|_{k+1,\mathcal{O}}$ noting that $1/\phi$ is of class $C^k$ outside $\cup_i \{\mathcal{O}_i\}$. This estimate holds also true for $u \in H^{k+1}(\mathcal{O})$ by density of $C^\infty$ in $H^{k+1}$.

3.2 Some technical lemmas

This Section regroups some technical results to be used later in the proofs of the coercivity of $a_h$ (Section 3.3) and the a priori error estimates (Sections 3.4 and 3.5). The most important contribution here is Lemma 3.3 which extends to finite elements of any degree a result from [12]. 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. It shows indeed that the $H^1$ norm of a finite element function on $\Omega^\mathcal{O}_h$ can be bounded by its norm on the whole computational domain $\Omega_h$ multiplied by a number strictly smaller than 1, modulo the addition of stabilization terms. We recall that our stabilization is strongly inspired by that of [3] but differs from it by some extra terms involving the Laplacian on mesh elements. The proof that such a stabilization is sufficient in Lemma 3.3 relies on a simple observation on polynomials, announced and proven in Lemma 3.2.

Lemma 3.2. Let $T$ be a triangle/tetrahedron, $E$ one of its sides and $p$ a polynomial on $T$ of degree $s \geq 0$ such that $p = \partial p/\partial n = 0$ on $E$ and $\Delta p = 0$ on $T$. Then $p = 0$ on $T$.

Proof. Let us consider only the 2D case (3D is similar). Without loss of generality, we can assume that $E$ lies on the $x$-axis in $(x, y)$ coordinates. Let $p = \sum p_{ij}x^iy^j$ with $i, j \geq 0, i + j \leq s$ as above. We shall prove by induction on $m = 0, 1, \ldots, t$ that $p_{im} = 0, \forall i$. Indeed, this is valid for $m = 0, 1$ since $p(x, 0) = \sum p_{i0}x^i = 0$ and $\partial p/\partial y(x, 0) = \sum p_{i1}x^i = 0$. Now, $\Delta p = 0$ implies for all indices $i, j \geq 0$

$$(i + 2)(i + 1)p_{i+2,j} + (j + 2)(j + 1)p_{i,j+2} = 0$$

so that $p_{im} = 0, \forall i$ implies $p_{i,m+2} = 0, \forall i$.

Lemma 3.3. Under Assumption 2, for any $\beta > 0$ and $s \in \mathbb{N}^+$ one can choose $0 < \alpha < 1$ depending only on the mesh regularity and such that, for each $v_h \in V^s_h$,

$$|v_h|_{1,\Omega^T_h}^2 \leq \alpha |v_h|_{1,\Omega_h}^2 + \beta h \sum_{E \in \mathcal{F}_h^T} \left\| \frac{\partial v_h}{\partial n} \right\|_{0,E}^2 + \beta h^2 \sum_{T \in \mathcal{T}_h^T} \|\Delta v_h\|_{0,T}^2.$$  

(13)

Proof. Choose any $\beta > 0$, consider the decomposition of $\Omega^T_h$ in element patches $\{\Pi_k\}$ as in Assumption 2, and introduce

$$\alpha := \max_{v_{h,k},v_{h,\mathcal{F}_h^T} \neq 0} \frac{|v_h|_{1,\Omega^T_h}^2 - \beta h \sum_{E \in \mathcal{F}_h^T} \left\| \frac{\partial v_h}{\partial n} \right\|_{0,E}^2 - \beta h^2 \sum_{T \in \Pi_k} \|\Delta v_h\|_{0,T}^2}{|v_h|_{1,\Pi_k}^2},$$  

(14)

where the maximum is taken over all the possible configurations of a patch $\Pi_k$ allowed by the mesh regularity and over all the piecewise polynomial functions on $\Pi_k$ (polynomials of degree $\leq s$). The subset $\mathcal{F}_h^T \subset \mathcal{F}_h^T_k$ gathers the edges internal to $\Pi_k$. Note that the quantity under the max sign in (14) is invariant under the scaling transformation $x \mapsto hx$ and is homogeneous with respect to $v_h$. Recall also that the
Proof. It is easy to see that the supremum leading, in combination with (15), to (17). The proof of (18) is similar.

\[ |v_h|^2 |\mathcal{T}_k + \beta h \sum_{E \in \mathcal{F}_k} \left\| \left[ \frac{\partial v_h}{\partial n} \right]_{0, E} \right\|^2 \leq \Delta v_h \|_{0,T} = 0 \]

since \(|v_h|^2 |\mathcal{T}_k = |v_h|^2 |\mathcal{T}_k + |v_h|^2 |\Pi_k\). This implies \(v_h = c = \text{const} \) on \(T_k\), \(\frac{\partial v_h}{\partial n} = 0\) on all \(E \in \mathcal{F}_k\), and \(\Delta v_h = 0\) on all \(T \subset \Pi_k\). Thus applying Lemma 3.2 to \(v_h - c\), we deduce that \(v_h = c\) on \(\Pi_k\), which contradicts \(|v_h|_{1,\Pi_k} = 1\).

This proves \(\alpha < 1\). We have thus

\[ |v_h|^2 |\mathcal{T}_k \leq \alpha |v_h|^2 |\mathcal{T}_k + \beta h \sum_{E \in \mathcal{F}_k} \left\| \left[ \frac{\partial v_h}{\partial n} \right]_{0, E} \right\|^2 + \beta h^2 \sum_{T \subset \Pi_k} \Delta v_h \|_{0,T} \]

for all \(v_h \in V_h\) and all the admissible patches \(\Pi_k\). Summing this over \(\Pi_k\), \(k = 1, \ldots, N_\Pi\) yields (13).

Lemma 3.4. For all \(v_h \in V_h^{(k)}\), it holds

\[ \| \phi_h v_h \|_{0,\Omega_k^c} \leq C h |\phi_h v_h|_{1,\Omega_k^c} \tag{15} \]

\[ |\phi_h v_h|_{0,\Omega_k \setminus \Omega} \leq C h |\phi_h v_h|_{1,\Omega_k} \tag{16} \]

with a constant \(C > 0\) depending only on the regularity of \(T_h\) and \(k\).

Proof. It is easy to see that the supremum

\[ C = \sup_{p_h \neq 0, T} \frac{\| p_h \|_{0,T}}{h_T |p_h|_{1,T}} \]

over all the polynomials \(p_h \in \mathbb{P}_{k+1}(T)\) vanishing at a point of \(T\) and all the simplexes \(T\) satisfying the regularity assumption \(h_T/\rho(T) \geq \beta\) is attained so that \(C\) is finite. Taking any \(T \in \mathcal{T}_h^\Gamma\) and putting \(p_h = \phi_h v_h\), this implies \(\| \phi_h v_h \|_{0,T} \leq C h |\phi_h v_h|_{1,T}\) for any \(V_h \in V_h^{(k)}\). Summing over all \(T \in \mathcal{T}_h^\Gamma\), concludes the proof of (15). Estimate (16) is proven similarly, adding, if necessary, neighbour elements to \(T \in \mathcal{T}_h^\Gamma\).}

Lemma 3.5. For all \(v_h \in V_h^{(k)}\)

\[ \sum_{E \in \mathcal{F}_h^\Gamma} \| \phi_h v_h \|_{0,E} \leq C h |\phi_h v_h|_{1,\Omega_h} \tag{17} \]

and

\[ |\phi_h v_h|_{0,\Omega_h \setminus \Omega} \leq C h |\phi_h v_h|_{1,\Omega_h} \tag{18} \]

with a constant \(C > 0\) depending only on the regularity of \(T_h\).

Proof. Let \(E \in \mathcal{F}_h^\Gamma\). Recall the well-known trace inequality

\[ \|v\|_{0,E} \leq C \left( \frac{1}{h} \|v\|_{0,T} + h |v|_{1,T} \right) \]

for each \(v \in H^1(E)\). Summing this over all \(E \in \mathcal{F}_h^\Gamma\) gives

\[ \sum_{E \in \mathcal{F}_h^\Gamma} \| \phi_h v_h \|_{0,E} \leq C \left( \frac{1}{h} \| \phi_h v_h \|_{0,\Omega_h} + h |\phi_h v_h|_{1,\Omega_h^c} \right) \]

leading, in combination with (15), to (17). The proof of (18) is similar.
Lemma 3.6. Under Assumption 1, it holds for all \( v \in H^s(\Omega_h) \) with integer \( 1 \leq s \leq k + 1 \), \( v \) vanishing on \( \Omega \),
\[
\|v\|_{0,\Omega_h \setminus \Omega} \leq C h^s \|v\|_{s,\Omega_h \setminus \Omega}.
\] (20)

Proof. Consider the 2D case \((d = 2)\). For simplicity, we can assume that \( v \) is \( C^\infty \) regular and pass to \( v \in H^s(\Omega_h) \) by density. By Assumption 1, we can pass to the local coordinates \( \xi_1, \xi_2 \) on every set \( \mathcal{O}_k \) covering \( \Gamma \) assuming that \( \xi_1 \) varies between 0 and \( L \) and, for any \( \xi_1 \) fixed, \( \xi_2 \) varies on \( \Omega_h \setminus \Omega \) from 0 to some \( b(\xi_1) \) with \( 0 \leq b(\xi_1) \leq C h \). We observe using the bounds on the mapping \((x_1, x_2) \mapsto (\xi_1, \xi_2)\)
\[
\|v\|_{0,(\Omega_h \setminus \Omega) \cap \mathcal{O}_k}^2 \leq C \int_0^L \int_0^{b(\xi_1)} v^2(\xi_1, \xi_2) d\xi_2 d\xi_1
\]
(recall that \( \frac{\partial^p v}{\partial \xi^2_2}(\xi_1, 0) = 0 \) for \( \alpha = 0, \ldots, s-1 \) and \( b \leq C h \))
\[
= C \int_0^L \int_0^{b(\xi_1)} \left( \int_0^{\xi_2} (\xi_2 - t)^{s-1} \frac{\partial^p v}{\partial \xi^2_2}(\xi_1, t) dt \right)^2 d\xi_2 d\xi_1
\]
\[
\leq C \left( \int_0^L \int_0^{b(\xi_1)} \frac{\partial^p v}{\partial \xi^2_2}(\xi_1, t)^2 dt \right) d\xi_1
\]
\[
\leq C h^{2s} \|v\|_{s,(\Omega_h \setminus \Omega) \cap \mathcal{O}_k}^2.
\]
Summing over all neighbourhoods \( \mathcal{O}_k \) gives (20). The proof in the 3D case is the same up to the change of notations. \(\square\)

3.3 Coercivity of the bilinear form \( a_h \)

Lemma 3.7. Under Assumption 2, the bilinear form \( a_h \) is coercive on \( V_h^{(k)} \) with respect to the norm
\[
\|v_h\|_h := \sqrt{\|\phi_h v_h\|_{1,\Omega_h}^2 + G_h(v_h, v_h)}
\]
i.e. \( a_h(v_h, v_h) \geq c \|v_h\|_h^2 \) for all \( v_h \in V_h^{(k)} \) with \( c > 0 \) depending only on the mesh regularity and on the constants in Assumption 2.

Proof. Let \( v_h \in V_h^{(k)} \) and \( B_h \) be the strip between \( \Gamma_h \) and \( \partial \Omega_h \), i.e. \( B_h = \{ \phi_h > 0 \} \cap \Omega_h \). Since \( \phi_h v_h = 0 \) on \( \Gamma_h \),
\[
\int_{\partial \Omega_h} \frac{\partial (\phi_h v_h)}{\partial n} \phi_h v_h = \int_{\partial B_h} \frac{\partial (\phi_h v_h)}{\partial n} \phi_h v_h
\]
\[
= \sum_{T \in T_h^\Gamma} \int_{\partial (B_h \cap \partial T)} \frac{\partial (\phi_h v_h)}{\partial n} \phi_h v_h - \sum_{T \in T_h^\Gamma} \sum_{E \in F_h^{cut}(T)} \int_{B_h \cap E} \frac{\partial (\phi_h v_h)}{\partial n} \phi_h v_h,
\]
where \( T_h^\Gamma \) is defined in (7) and \( F_h^{cut}(T) \) regroups the facets of a mesh element \( T \) cut by \( \Gamma_h \). By divergence theorem,
\[
\int_{\partial \Omega_h} \frac{\partial (\phi_h v_h)}{\partial n} \phi_h v_h = \int_{B_h} |\nabla (\phi_h v_h)|^2 + \sum_{T \in T_h^\Gamma} \int_{B_h \cap T} \Delta (\phi_h v_h) \phi_h v_h
\]
\[
- \sum_{E \in F_h^T} \int_{E \cap B_h} \phi_h v_h \left[ \frac{\partial (\phi_h v_h)}{\partial n} \right].
\]
Substituting this into the definition of \( a_h \) yields

\[
a_h(v_h, v_h) = \int_{\Omega_h} |\nabla (\phi_h v_h)|^2 - \int_{\Omega_h} |\nabla (\phi_h v_h)|^2 - \sum_{T \in T_h^k} \int_{T} \Delta (\phi_h v_h) \phi_h v_h \\
+ \sum_{F \in F_h^k} \int_{F \cap T_h} \phi_h v_h \left[ \frac{\partial (\phi_h v_h)}{\partial n} \right] + \sigma h^2 \sum_{T \in T_h^k} \int_T |\Delta (\phi_h v_h)|^2 + \sigma h \sum_{E \in F_h^k} \int_E \left[ \frac{\partial (\phi_h v_h)}{\partial n} \right]^2. \tag{21}
\]

Since \( B_h \subset \Omega_h^1 \) (cf. (7)), applying Lemma 3.3 to \( \phi_h v_h \in V_h^{(k+1)} \) gives

\[
\int_{B_h} |\nabla (\phi_h v_h)|^2 \leq \alpha |\phi_h v_h|_{1, \Omega_h}^2 + \beta h \sum_{E \in F_h^k} \int_E \left[ \frac{\partial (\phi_h v_h)}{\partial n} \right]^2 + \beta h \sum_{T \in T_h^k} \int_T |\Delta (\phi_h v_h)|^2.
\]

Moreover, by Young inequality, (15) and (17), we obtain for any \( \epsilon > 0 \)

\[
\sum_{T \in T_h^k} \int_{B_h \cap T} \Delta (\phi_h v_h) \phi_h v_h \leq \frac{h^2}{2\epsilon} \sum_{T \in T_h^k} \int_T |\Delta (\phi_h v_h)|^2 + C\epsilon |\phi_h v_h|_{1, \Omega_h}^2
\]

and

\[
\sum_{F \in F_h^k} \int_{F \cap T_h} \phi_h v_h \left[ \frac{\partial (\phi_h v_h)}{\partial n} \right] \leq \frac{h}{2\epsilon} \sum_{E \in F_h^k} \int_E \left[ \frac{\partial (\phi_h v_h)}{\partial n} \right]^2 + C\epsilon |\phi_h v_h|_{1, \Omega_h}^2.
\]

Thus, putting the last 3 bounds into (21) we arrive at

\[
a(v_h, v_h) \geq (1 - \alpha - C\epsilon) |\phi_h v_h|_{1, \Omega_h}^2
+ \left( \sigma - \beta - \frac{1}{2\epsilon} \right) h \sum_{E \in F_h^k} \left[ \frac{\partial (\phi_h v_h)}{\partial n} \right]_{0, E} + \left( \sigma - \beta - \frac{1}{2\epsilon} \right) h^2 \sum_{T \in T_h^k} \int_T |\Delta (\phi_h v_h)|^2.
\]

This leads to the conclusion taking \( \epsilon \) sufficiently small and \( \sigma \) sufficiently big.

\[\square\]

### 3.4 Proof of the \( H^1 \) error estimate in Theorem 2.1

Since the bilinear form \( a_h \) is coercive, it remains to construct a good interpolant of the exact solution \( u \) in the form of a product of a function from \( V_h^{(k)} \) with \( \phi_h \). The details of such a construction are given below together with the appropriate interpolation estimates. An additional difficulty will come from the extra terms in the Galerkin orthogonality relation (24) with \( f \) resulting from the extension of \( u \) from \( \Omega \) to \( \Omega_h \). These terms turn out to be of optimal order since \( \tilde{f} \) differs from \( f \) only on a narrow strip of width \( \sim h \), cf. Lemma 3.6 and (26).

We now proceed with the detailed proof. Since \( f \in H^k(\Omega) \), the solution \( u \) of (1) belongs to \( H^{k+2}(\Omega) \) (see [10, p. 323]) and can be extended by a function \( \hat{u} \) in \( H^{k+2}(\Omega) \), cf. [10, p. 257], such that \( \hat{u} = u \) on \( \Omega \) and

\[
\|\hat{u}\|_{k+2, \Omega_h} \leq \|\hat{u}\|_{k+2, \Omega} \leq C\|u\|_{k+2, \Omega} \leq C\|f\|_{k, \Omega}. \tag{22}
\]

Let \( w = \hat{u}/\phi \). By Lemma 3.1,

\[
|w|_{k+1, \Omega_h} \leq C\|w\|_{k+2, \Omega} \leq C\|f\|_{k, \Omega}. \tag{23}
\]

Introduce the bilinear form \( \tilde{a}_h \), similar to \( a_h \) as defined in (6) but with \( \phi \) instead of \( \phi_h \) multiplying the trial function:

\[
\tilde{a}_h(w, v) = \int_{\Omega_h} \nabla (\phi w) \cdot \nabla (\phi_h v) - \int_{\partial \Omega_h} \frac{\partial (\phi w)}{\partial n} (\phi_h v) \\
+ \sigma h^2 \sum_{E \in F_h^k} \int_E \left[ \frac{\partial (\phi w)}{\partial n} \right] \left[ \frac{\partial (\phi_h v)}{\partial n} \right] + \sigma h \sum_{T \in T_h} \int_T \Delta (\phi w) \Delta (\phi_h v).
\]
Since \( \phi w = \tilde{u} \in H^2(\Omega_h) \), an integration by parts yields
\[
\bar{a}_h(w, v_h) = \int_{\Omega_h} f \psi_h v_h - \sigma h^2 \sum_{T \in T_h^c} \int_T \hat{f} \Delta(\phi_h v_h), \quad \forall v_h \in V_h
\]
with \( \hat{f} = -\Delta \tilde{u} \) on \( \Omega_h \). Hence,
\[
a_h(w, v_h) - \bar{a}_h(w, v_h) = \int_{\Omega_h} (f - \hat{f}) \psi_h v_h - \sigma h^2 \sum_{T \in T_h^c} \int_T (f - \hat{f}) \Delta(\phi_h v_h).
\]
(24)

Put
\[
v_h = w_h - I_h w \quad \text{and} \quad r_h = \phi w - \phi_h I_h w
\]
with the nodal interpolator \( I_h \). Eq. (24) can be rewritten as
\[
a_h(v_h, v_h) = \bar{a}_h(w, v_h) - a_h(I_h w, v_h)
\]
\[
+ \int_{\Omega_h} (f - \hat{f}) \psi_h v_h - \sigma h^2 \sum_{T \in T_h^c} \int_T (f - \hat{f}) \Delta(\phi_h v_h)
\]
\[
= \int_{\Omega_h} \nabla r_h \cdot \nabla(\phi_h v_h) - \int_{\partial \Omega_h} \frac{\partial r_h}{\partial n} \phi_h v_h
\]
\[
+ \sigma h \sum_{E \in F_h^c} \int_E \left[ \frac{\partial r_h}{\partial n} \left( \frac{\partial (\phi_h v_h)}{\partial n} \right) + \sigma h^2 \sum_{T \in T_h^c} \int_T \Delta r_h \Delta(\phi_h v_h)
\]
\[
+ \int_{\Omega_h} (f - \hat{f}) \psi_h v_h - \sigma h^2 \sum_{T \in T_h^c} \int_T (f - \hat{f}) \Delta(\phi_h v_h).
\]

By Lemma 3.7, Young inequality, and recalling \( f = \hat{f} \) on \( \Omega \), we now get
\[
\epsilon \|v_h\|_h^2 \leq \frac{1}{2\varepsilon} |r_h|_{T, \Omega_h}^2 + \frac{h}{2\varepsilon} \left\| \frac{\partial r_h}{\partial n} \right\|_{0, \partial \Omega_h}^2
\]
\[
+ \frac{\sigma^2 h}{2\varepsilon} \sum_{E \in F_h^c} \left\| \frac{\partial \phi_h}{\partial n} \right\|_{0, E}^2 + \frac{\sigma^2 h^2}{2\varepsilon} \sum_{T \in T_h^c} \|\Delta r_h\|_{0, T}^2 + \frac{(1 + \sigma^2)h^2}{2\varepsilon} \|f - \hat{f}\|_{0, \Omega_h \setminus \Omega}^2
\]
\[
+ \frac{\varepsilon}{2} \left( |\phi_h v_h|_{1, \Omega_h}^2 + \frac{1}{h} \|\phi_h v_h\|_{0, \partial \Omega_h}^2 + h \sum_{E \in F_h^c} \left\| \frac{\partial (\phi_h v_h)}{\partial n} \right\|_{0, E}^2
\]
\[
+ 2h^2 \sum_{T \in T_h^c} \|\Delta(\phi_h v_h)\|_{0, T}^2 + \frac{1}{h^2} \|\phi_h v_h\|_{0, \Omega_h \setminus \Omega}^2 \right). \]

The terms above multiplied by \( \varepsilon/2 \) can be absorbed by the left-hand side. Indeed, the first contribution
\( |\phi_h v_h|_{1, \Omega_h}^2 \) and the sums over \( F_h^c \) and \( T_h^c \) are evidently controlled by \( \|v_h\|_h^2 \). The remaining terms are controlled by \( |\phi_h v_h|_{1, \Omega_h}^2 \) and hence by \( \|v_h\|_h^2 \) thanks to (16) and (18). Taking \( \varepsilon \) small enough, we thus conclude
\[
\|v_h\|_h^2 \leq C \left( |r_h|_{1, \Omega_h}^2 + h \left\| \frac{\partial (\phi w - \phi_h I_h w)}{\partial n} \right\|_{0, \partial \Omega_h}^2
\]
\[
+ h^2 \sum_{T \in T_h^c} \|\Delta r_h\|_{0, T}^2 + h \sum_{E \in F_h^c} \left\| \frac{\partial r_h}{\partial n} \right\|_{0, E}^2 + h^2 \|f - \hat{f}\|_{0, \Omega_h \setminus \Omega}^2 \right). \quad (25)
\]
Similarly, we continue using the classical interpolation bounds (see for instance [2]).

We now estimate each term in the right-hand side of (25). By triangular inequality, we get

\[ |r_h|_{1, \Omega_h} \leq |(\phi - \phi_h)w|_{1, \Omega_h} + |\phi_h(w - I_h w)|_{1, \Omega_h} \]
\[ \leq \| \nabla (\phi - \phi_h) \|_{L^\infty(\Omega_h)} \| w \|_{0, \Omega_h} + \| \phi - \phi_h \|_{L^\infty(\Omega_h)} |w|_{1, \Omega_h} \]
\[ + \| \nabla \phi_h \|_{L^\infty(\Omega_h)} |w - I_h w|_{0, \Omega_h} + \| \phi_h \|_{L^\infty(\Omega_h)} |w - I_h w|_{1, \Omega_h}. \]

We continue using the classical interpolation bounds (see for instance [2])

\[ |r_h|_{1, \Omega_h} \leq Ch^k(\| \phi \|_{W^{k+1, \infty}(\Omega_h)} \| w \|_{0, \Omega_h} + \| \phi \|_{W^{k, \infty}(\Omega_h)} |w|_{1, \Omega_h} \]
\[ + \| \phi \|_{W^{1, \infty}(\Omega_h)} |w|_{k, \Omega_h} + \| \phi \|_{L^\infty(\Omega_h)} |w|_{k+1, \Omega_h}). \]

Similarly,

\[ \left( \sum_{T \in T_h} r_h^2_{2,T} \right)^{1/2} \leq Ch^k \| \phi \|_{W^{k+1, \infty}(\Omega_h)} \| w \|_{k+1, \Omega_h}. \]

and

\[ \left\| \frac{\partial r_h}{\partial n} \right\|_{0, \partial \Omega_h}^2 + \sum_{E \in F_h^k} \left\| \left[ \frac{\partial r_h}{\partial n} \right] \right\|_{0, E}^2 \leq Ch^{2k-1} \| \phi \|_{W^{k+1, \infty}(\Omega_h)} \| w \|_{k+1, \Omega_h}. \]

Finally, we get by Lemma 3.6 applied to \( f - \tilde{f} \) which vanishes on \( \Omega \),

\[ \| f - \tilde{f} \|_{0, \Omega_h} \leq Ch^k \| f - \tilde{f} \|_{k-1, \Omega_h} \leq Ch^k(\| f \|_{k-1, \Omega_h} + \| \tilde{u} \|_{k+1, \Omega_h}) \]

(26)

since \( \tilde{f} = -\Delta \tilde{u} \).

Using all the bounds above in (25), we get

\[ \| \phi_h(w_h - I_h w) \|_{1, \Omega_h} \leq \| v_h \|_{k+1, \Omega_h} \leq Ch^k(\| w \|_{k+1, \Omega_h} + \| f \|_{k-1, \Omega_h} + \| \tilde{u} \|_{k+1, \Omega_h}). \]

(27)

with a constant \( C > 0 \) that has absorbed \( \| \phi \|_{W^{k+1, \infty}(\Omega_h)} \). Applying again the triangle inequality and the interpolation bounds,

\[ |u - \phi_h w_h|_{1, \Omega \cap \Omega_h} \leq |\tilde{u} - \phi_h w_h|_{1, \Omega_h} \]
\[ \leq |(\phi - \phi_h)w|_{1, \Omega_h} + |\phi_h(w - I_h w)|_{1, \Omega_h} + |\phi_h(I_h w - w_h)|_{1, \Omega_h} \]
\[ \leq Ch^k(\| w \|_{k+1, \Omega_h} + \| f \|_{k-1, \Omega_h} + \| \tilde{u} \|_{k+1, \Omega_h}). \]

We have thus proven (8) taking into account the bounds (22) and (23).

3.5 Proof of the \( L^2 \) error estimate in Theorem 2.1

As usual, the \( L^2 \) error estimate will be proven here by Aubin-Nitsche trick. However, the discrepancy between \( \Omega \) and \( \Omega_h \), as well as between \( \phi \) and \( \phi_h \), gives rise to numerous terms, which should be bounded through rather tedious calculations. We shall skip some repetitive technical details as they are similar to those in the proof of the \( H^1 \) error estimate above. We also recall that we do not track explicitly the dependence of constants on the norms of \( \phi \).

Let \( z \in H^3(\Omega) \) be solution to

\[ -\Delta z = u - u_h \text{ on } \Omega, \quad z = 0 \text{ on } \Gamma. \]

Extend it to \( \Omega_h \) by \( \tilde{z} \in H^3(\Omega_h) \) using an extension operator bounded in the \( H^3 \) norm. Set \( y = \tilde{z}/\phi \). Then

\[ |y|_{1, \Omega_h} \leq C |\tilde{z}|_{3, \Omega_h} \leq C \| u - u_h \|_{1, \Omega} \text{ and } \| y \|_{1, \Omega_h} \leq C |\tilde{z}|_{2, \Omega_h} \leq C \| u - u_h \|_{0, \Omega}. \]

(28)
thanks to Lemma 3.1 and to the elliptic regularity estimate. We also have

By Lemma 3.1 from [12], we have for any \( v \in H^1(\Omega_h^k) \)

\[
\|v\|_{0,\Omega_h^k} \leq C \left( \sqrt{h} \|v\|_{0,\Gamma} + h |v|_{1,\Omega_h^k} \right). \tag{29}
\]

Putting \( v = \nabla \tilde{z} \), (29) gives

\[
\|\tilde{z}\|_{1,\Omega_h^k} \leq C \left( \sqrt{h} \|\nabla \tilde{z}\|_{0,\Gamma} + h |\tilde{z}|_{2,\Omega_h^k} \right) \leq C \sqrt{h} \|\tilde{z}\|_{2,\Omega_h} \leq C \sqrt{h} \|u - u_h\|_{0,\Omega}. \tag{30}
\]

By integration by parts,

\[
\|u - u_h\|_{0,\Omega}^2 = \int_{\Omega} (u - u_h)(-\Delta z) = -\int_{\Gamma} (u - u_h) \frac{\partial z}{\partial n} + \int_{\Omega} \nabla (u - u_h) \cdot \nabla z. \tag{31}
\]

To treat the first term in (31), we remark first

\[
-\int_{\Gamma} (u - u_h) \frac{\partial z}{\partial n} \leq \|u - u_h\|_{0,\Gamma} \left\| \frac{\partial z}{\partial n} \right\|_{0,\Gamma} \leq C \|u - u_h\|_{0,\Gamma} \left\| u - u_h \right\|_{0,\Gamma}.
\]

Furthermore, since the distance between \( \Gamma \) and \( \Gamma_h \) is of order \( h^{k+1} \), we have

\[
\|u - u_h\|_{0,\Gamma} \leq C(\|\tilde{u} - u_h\|_{0,\Gamma_h} + h^{(k+1)/2} |\tilde{u} - u_h|_{1,\Omega_h})
\]

\[
= C(\|\phi - \phi_h\|_{0,\Gamma_h} + h^{(k+1)/2} |\tilde{u} - u_h|_{1,\Omega_h})
\]

\[
\leq C(h^{k+1} \|w\|_{0,\Gamma_h} + h^{(k+1)/2 + k} \|f\|_{k,\Omega_h}).
\]

We have used here the already proven bound on \( |\tilde{u} - u_h|_{1,\Omega_h} \) and the interpolation error bound for \( \phi - \phi_h \).

We have thus thanks to Lemma 3.1,

\[
\|u - u_h\|_{0,\Gamma} \leq Ch^{k+1}(\|w\|_{1,\Omega_h} + \|f\|_{k,\Omega_h})
\]

\[
\leq Ch^{k+1}(\|\tilde{u}\|_{2,\Omega_h} + \|f\|_{k,\Omega_h}) \leq C h^{k+1}\|f\|_{k,\Omega_h}.
\]

Hence,

\[
-\int_{\Gamma} (u - u_h) \frac{\partial z}{\partial n} \leq Ch^{k+1}\|f\|_{k,\Omega_h} \left\| u - u_h \right\|_{0,\Omega}. \tag{32}
\]

The second term in (31) is treated by Galerkin orthogonality (24): for any \( y_h \in V_h^{(k)} \)

\[
\int_{\Omega} \nabla (u - u_h) \cdot \nabla z = \int_{\Omega_h} \nabla (\phi w - \phi_h w_h) \cdot \nabla (\phi y - \phi_h y_h)
\]

\[
- \int_{\Omega_h \setminus \Omega} \nabla (\phi w - \phi_h w_h) \cdot \nabla (\phi y) + \int_{\partial \Omega_h} \frac{\partial}{\partial n} (\phi w - \phi_h w_h)(\phi h y_h)
\]

\[
- \sigma h \sum_{E \in F_h^k} \int_{E} \left[ \frac{\partial}{\partial n}(\phi w - \phi_h w_h) \right] \left[ \frac{\partial}{\partial n}(\phi h y_h) \right] - \sigma h^2 \sum_{T \in T_h^k} \int_{T} \Delta(\phi w - \phi_h w_h) \Delta(\phi h y_h)
\]

\[
+ \int_{\Omega_h} (f - \tilde{f}) \phi_h y_h - \sigma h^2 \sum_{T \in T_h^k} \int_{T} (f - \tilde{f}) \Delta(\phi h y_h). \tag{33}
\]
We now estimate term by term the right-hand side of the above inequality taking \( y_h = \tilde{I}_h y \) with \( \tilde{I}_h \) the Clément interpolation operator on \( T_h \).

**Term I:** by Cauchy-Schwartz, the already proven bound on \( |\tilde{u} - u_h|_{1,\Omega_h} \), and (28),

\[
|I| \leq C|\tilde{u} - u_h|_{1,\Omega_h} |\phi y - \phi_h y_h|_{1,\Omega_h} \leq C h^{k+1} \|f\|_{k,\Omega_h} \|y\|_{2,\Omega_h} \leq C h^{k+1} \|f\|_{k,\Omega_h} \|u - u_h\|_{1,\Omega}.
\]

**Term II:** using (30) for \( \tilde{z} = \phi y \),

\[
|II| \leq |\tilde{u} - u_h|_{1,\Omega_h} |\tilde{z}|_{1,\Omega_h \setminus \Omega} \leq C h^{k+1/2} \|f\|_{k,\Omega_h} \|u - u_h\|_{0,\Omega}.
\]

**Term III:** applying the trace inequality (19) on the mesh elements adjacent to \( \partial \Omega_h \) yields

\[
|III| \leq C \left( \sum_{T \in \mathcal{T}_h} \left( \frac{1}{h} |\tilde{u} - u_h|_{1,T}^2 + |\tilde{u} - u_h|_{2,T}^2 \right) \right)^{1/2} \|\phi_h y_h\|_{0,\partial \Omega_h}.
\]

The term in parentheses can be further bounded using the triangle inequality, interpolation estimates, and the bound (27) on \( v_h = \phi_h (w_h - I_h w) \) as

\[
(\cdots)^{1/2} \leq \left( \frac{1}{h} \left| \tilde{u} - \phi_h I_h w \right|_{1,\Omega_h}^2 + h |\tilde{u} - \phi_h I_h w|_{2,T}^2 \right)^{1/2} + \frac{1}{\sqrt{h}} \|v_h\|_h \leq C h^{k-1/2} \|f\|_{k,\Omega_h}.
\]

Moreover, since \( \Omega_h^\varepsilon \) is a strip around \( \Gamma_h \) of width \( \sim h \), we have

\[
\|\phi_h\|_{L^\infty(\Omega_h^\varepsilon)} \leq C h \|\nabla \phi_h\|_{L^\infty(\partial \Omega_h)} \leq C h \tag{34}
\]

and, by (28),

\[
\|\phi_h y_h\|_{0,\partial \Omega_h} \leq C h \|y\|_{1,\Omega_h} \leq C h \|u - u_h\|_{0,\Omega}
\]

so that

\[
|III| \leq C h^{k+1/2} \|f\|_{k,\Omega_h} \|u - u_h\|_{0,\Omega}.
\]

**Term IV:** by Cauchy-Schwartz and trace inequalities, together with the interpolation estimates,

\[
|IV| \leq (C h^k \|f\|_{k,\Omega_h} + \|v_h\|_h) G_h(y_h, y_h)^{1/2} \leq C h^k \|f\|_{k,\Omega_h} G_h(y_h, y_h)^{1/2}
\]

and by (30) and (34),

\[
G_h(y_h, y_h)^{1/2} \leq \frac{C}{h} \|\phi_h y_h\|_{0,\Omega_h^\varepsilon} \leq C \|y\|_{0,\Omega_h^\varepsilon} \leq C |\tilde{z}|_{1,\Omega_h^\varepsilon} \leq C \sqrt{h} \|u - u_h\|_{0,\Omega}. \tag{35}
\]

Hence,

\[
|IV| \leq C h^{k+1/2} \|f\|_{k,\Omega_h} \|u - u_h\|_{0,\Omega}.
\]

**Term V:** by an inverse inequality and (26),

\[
|V| \leq C \|f - \tilde{f}\|_{0,\Omega_h^\varepsilon} \|\phi_h y_h\|_{0,\Omega_h^\varepsilon} \leq C h^{-1} \|f\|_{k,\Omega_h} \|\phi_h y_h\|_{0,\Omega_h^\varepsilon} \leq C h^{k-1} \|f\|_{k,\Omega_h} \|\phi_h y_h\|_{0,\Omega_h^\varepsilon} \|u - u_h\|_{0,\Omega}.
\]

Proceeding as in (35) we conclude

\[
|V| \leq C h^{k+1/2} \|f\|_{k,\Omega_h} \|u - u_h\|_{0,\Omega}.
\]
Combining the bounds for the terms I–V in (33) with (32) and putting all this into (31), we obtain by Young inequality
\[ \|u - u_h\|_{0,\Omega}^2 \leq C h^{k+1} \|f\|_{k,\Omega_h} \|u - u_h\|_{1,\Omega} + C h^{k+1/2} \|f\|_{k,\Omega_h} \|u - u_h\|_{0,\Omega} \]
\[ \leq C h^{k+1} \|f\|_{k,\Omega_h}^2 + \varepsilon h \|u - u_h\|_{1,\Omega}^2 + \varepsilon \|u - u_h\|_{0,\Omega}^2. \]

By the already established estimate for \( |u - u_h|_{1,\Omega} \),
\[ \|u - u_h\|_{0,\Omega}^2 \leq C \left( \frac{1}{\varepsilon} + \varepsilon \right) h^{2k+1} \|f\|_{k,\Omega_h}^2 + (\varepsilon + \varepsilon h) \|u - u_h\|_{0,\Omega}^2 \]
which proves (9) taking sufficiently small \( \varepsilon \).

4 Conditioning of the system matrix

We are now going to prove that the condition number of the finite element matrix associated to the bilinear form \( a_h \) of \( \phi \)-FEM does not suffer from the introduction of the multiplication by \( \phi_h \): it is of order \( 1/h^2 \) on a quasi-uniform mesh of step \( h \), similar to the standard FEM on a fitted mesh.

**Theorem 4.1 (Conditioning).** Under Assumptions 1 and 2 and recalling that the mesh \( T_h \) is supposed to be quasi-uniform, the condition number defined by \( \kappa(A) := \|A\|_2 \|A^{-1}\|_2 \) of the matrix \( A \) associated to the bilinear form \( a_h \) on \( V_h^{(k)} \), as in (6), satisfies
\[ \kappa(A) \leq C h^{-2}. \]

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

**Proof.** **Step 1** (a lower bound on \( a_h \)). We shall prove for all \( w_h \in V_h^{(k)} \)
\[ a_h(w_h, w_h) \geq C |w_h|_{0,\Omega_h}^2. \]

By Lemma 3.7, it holds for each \( w_h \in V_h^{(k)} \)
\[ a_h(w_h, w_h) \geq c |w_h|_{0,\Omega_h}^2 \geq c |\phi_h w_h|_{1,\Omega_h}^2. \]

We now denote \( u_h = \phi_h w_h \) and apply Lemma 3.1 with \( k = 0 \) and \( \phi_h \) instead of \( \phi \) to \( w_h = u_h/\phi_h \):
\[ |w_h|_{0,\Omega_h} \leq C |u_h|_{1,\Omega_h}. \]

This is justified by a possible relaxation of the hypotheses of Lemma 3.1. The constant in (37) depends on \( |\phi_h|_{W^{1,\infty}(\Omega_h)} \) which is bounded uniformly in \( h \). Moreover, the local coordinates around \( \Gamma \) evoked in Assumption 1 can be reused to build the same around \( \Gamma_h \).

Applying Poincaré inequality on the domain \( \Omega_h^{\text{in}} := \{ \phi_h < 0 \} \) yields, as \( u_h = 0 \) on \( \Gamma_h = \partial \Omega_h^{\text{in}} \),
\[ |u_h|_{0,\Omega_h} \leq C |u_h|_{1,\Omega_h^{\text{in}}} \]
with a constant that depends only on the diameter of \( \Omega_h^{\text{in}} \) and can be thus assumed \( h \)-independent. Moreover, invoking Lemma 3.4 and observing \( \Omega_h \setminus \Omega_h^{\text{in}} \subset \Omega_h^{\text{in}} \) we conclude that
\[ |u_h|_{0,\Omega_h} \leq C |u_h|_{1,\Omega_h}. \]

Combining this with (37) we finish the proof of (36) as follows:
\[ a_h(w_h, w_h) \geq c |u_h|_{1,\Omega_h}^2 \geq C |u_h|_{0,\Omega_h}^2 \geq C |w_h|_{0,\Omega_h}^2. \]
\textbf{Step 2} (an upper bound on \(a_h\)). We shall prove for all \(w_h \in V_h^{(k)}\)
\[
a_h(w_h, w_h) \leq C \frac{1}{h^2} \|w_h\|_{0, \Omega_h}^2. \tag{38}
\]
By definition of \(a_h\) and Lemma 3.5,
\[
a_h(w_h, w_h) \leq C|\phi_h w_h|_{1, \Omega_h}^2 + C \sqrt{h} \left\| \frac{\partial (\phi_h w_h)}{\partial n} \right\|_{0, \partial \Omega_h} + C h^2 \sum_{T \in T_h} |\phi_h w_h|_{2, T}^2.
\]
Using the inverse inequalities on \(V_h^{(k+l)}\)
\[
\left\| \frac{\partial (\phi_h w_h)}{\partial n} \right\|_{0, \partial \Omega_h} \leq \frac{C}{\sqrt{h}} \|\phi_h w_h\|_{0, \Omega_h}, \quad |\phi_h w_h|_{1, \Omega_h} \leq \frac{C}{h} \|\phi_h w_h\|_{0, \Omega_h},
\]
and \(|\phi_h w_h|_{2, T} \leq \frac{C}{h^2} \|\phi_h w_h\|_{0, T}\) yields
\[
a_h(w_h, w_h) \leq C \|\phi_h w_h\|_{0, \Omega_h}^2,
\]
which entails (38) since \(\phi_h\) is bounded uniformly in \(h\).

\textbf{Step 3}. Denote the dimension of \(V_h^{(k)}\) by \(N\) and let us associate any \(v_h \in V_h^{(k)}\) with the vector \(v \in \mathbb{R}^N\) containing the expansion coefficients of \(v_h\) 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
\[
C_1 h^{d/2} |v|_2 \leq \|v_h\|_{0, \Omega_h} \leq C_2 h^{d/2} |v|_2. \tag{39}
\]
The bounds (39) and (38) imply
\[
\|A\|_2 = \sup_{v \in \mathbb{R}^N} \frac{(Av, v)}{|v|_2^2} = \sup_{v \in \mathbb{R}^N} \frac{a(v_h, v_h)}{|v|_2^2} \leq C h^d \sup_{v_h \in V_h} \frac{a(v_h, v_h)}{|v_h|_0^2} \leq C h^{d-2}.
\]
Similarly, (39) and (36) imply
\[
\|A^{-1}\|_2 = \sup_{v \in \mathbb{R}^N} \frac{|v|_2^2}{(Av, v)} = \sup_{v \in \mathbb{R}^N} \frac{|v_h|_2^2}{a(v_h, v_h)} \leq \frac{C}{h^d} \sup_{v_h \in V_h} \frac{|v_h|_0^2}{a(v_h, v_h)} \leq \frac{C}{h^d}.
\]
These estimates lead to the desired result. \(\Box\)

5 \ Numerical results

We have implemented \(\phi\)-FEM in FEniCS Project [1] and report here some results using uniform Cartesian meshes on a rectangle \(O\) as the background mesh \(T_h^O\). The level-sets \(\phi\) are approximated in all the tests by the same finite elements as \(w_h\), \textit{i.e.} we take \(l = k\) in (4). All the integrals in \(a_h\) are computed exactly by using quadrature rules of sufficiently high order.

\textbf{1st test case}

Let \(\Omega\) be the circle of radius \(\sqrt{2}/4\) centered at the point \((0.5, 0.5)\) with \(\phi(x, y) = 1/8 - (x - 1/2)^2 - (y - 1/2)^2\) and the surrounding domain \(O = (0, 1)^2\). We use \(\phi\)-FEM to solve numerically Poisson-Dirichlet problem (1) with the exact solution given by \(u(x, y) = \phi(x, y) \times \exp(x) \times \sin(2\pi y)\). The errors in \(L^2\) and \(H^1\) norms for \(\phi\)-FEM with \(P_1\) finite elements are reported in Fig. 1. Taking the stabilization parameter \(\sigma = 20\), the numerical results confirm the theoretically predicted optimal convergence orders (in fact, the convergence
We now set \( \Omega \) as the rectangle with corners 

\[
\left( \frac{2\pi^2}{\pi^2 + 1}, \frac{\pi^3 - \pi}{\pi^2 + 1} \right), (0, \pi), \left( -\frac{2\pi^2}{\pi^2 + 1}, -\frac{\pi^3 - \pi}{\pi^2 + 1} \right), (0, -\pi) \]

and \( \sigma \geq 0 \) with \( P \) finite elements. The results with higher order \( P_k \) finite elements, \( k = 2, 3 \) are reported in Fig. 3. The optimal convergence orders under the mesh refinement are again observed (with the order \((k+1)\) in the \( L^2 \) norm, better than in theory). The influence of the stabilization parameter \( \sigma \) on the accuracy of \( \phi \)-FEM with \( P_1 \) and \( P_2 \) finite elements is investigated at Fig. 4: the method is robust with respect to the value of \( \sigma \) at least in the range \([0.1, 20]\).

**2nd test case**

We now set \( \Omega \) as the rectangle with corners 

\[
\left( \frac{2\pi^2}{\pi^2 + 1}, \frac{\pi^3 - \pi}{\pi^2 + 1} \right), (0, \pi), \left( -\frac{2\pi^2}{\pi^2 + 1}, -\frac{\pi^3 - \pi}{\pi^2 + 1} \right), (0, -\pi) \]

and \( \sigma \geq 0 \) with \( P \) finite elements. The results with higher order \( P_k \) finite elements, \( k = 2, 3 \) are reported in Fig. 3. The optimal convergence orders under the mesh refinement are again observed (with the order \((k+1)\) in the \( L^2 \) norm, better than in theory). The influence of the stabilization parameter \( \sigma \) on the accuracy of \( \phi \)-FEM with \( P_1 \) and \( P_2 \) finite elements is investigated at Fig. 4: the method is robust with respect to the value of \( \sigma \) at least in the range \([0.1, 20]\).

**3rd test case**

To get further outside of the theoretically comfortable playground, we consider the problem

\[
-\text{div}(A \nabla u) + u = f \quad \Omega, \quad u = g \quad \Gamma,
\]

with a smooth positive coefficient \( A \), assumed to be known on \( \mathcal{O} \supset \Omega \). In order to apply the \( \phi \)-FEM idea to the non-homogeneous boundary conditions in (40), we assume that \( g \) is actually defined and is sufficiently smooth on \( \mathcal{O} \). We consider then the ansatz

\[
\begin{align*}
\phi_h &= \phi_h w_h + g_h, \quad g_h \in V_h^{(k)} \quad \text{an interpolant of} \quad g \quad \text{on} \quad T_h, \\
w_h &\in V_h^{(k)} \quad \text{is the solution to the following problem generalizing (5)}
\end{align*}
\]

\[
\begin{align*}
\int_{\Omega} [A \nabla (\phi_h w_h + g_h) \cdot \nabla (\phi_h v_h) + (\phi_h w_h + g_h) \phi_h v_h] - \int_{\partial \Omega} A \frac{\partial}{\partial n} (\phi_h w_h + g_h) \phi_h v_h \\
+ \sigma h \sum_{E \in \mathcal{T}_h} \int_E \left[ \frac{\partial}{\partial n} (\phi_h w_h + g_h) \right] \left[ \frac{\partial}{\partial n} (\phi_h v_h) \right] + \sigma h^2 \sum_{T \in \mathcal{T}_h} \int_T \mathcal{L}(\phi_h w_h + g_h) \mathcal{L}(\phi_h v_h) \\
= \int_{\Omega} f \phi_h v_h - \sigma h^2 \sum_{T \in \mathcal{T}_h} \int_T f \mathcal{L}(\phi_h v_h), \quad \forall v_h \in V_h^{(k)}
\end{align*}
\]

with \( \mathcal{L}(v) = -\text{div}(A \nabla v) + v \).

We use \( \phi \)-FEM (41) to solve numerically Problem (40) on the domain \( \Omega \) defined by the level-set function \( \phi \) given in the polar coordinates \((r, \theta)\) by \( \phi(r, \theta) = r^4(5 + 3 \sin(7\theta + 7\pi/36))/2 - 0.47^4 \), taking \( \mathcal{O} = (-1, 1)^2 \) as the surrounding domain (see [11, 12] for pictures of \( \Omega \)). We choose \( A(x, y) = (1 + x^2 + y^2) \), adjust \( f \) so that the exact solution is given by \( u(x, y) = \sin(x) \exp(y) \), and take the Dirichlet data as \( g(x, y) = \phi(x, y) \exp(x) \sin(y) + u(x, y) \) so that \( u = g \) on \( \Gamma \) only. The results are presented in Figure 5. We observe that the Ghost-penalty part is essential and ensures the optimal convergence of the method.

**6 Conclusions and outlook**

The numerical results from the last section confirm the theoretically predicted optimal convergence of \( \phi \)-FEM in the \( H^1 \) semi-norm. The convergence in the \( L^2 \) norm turns out to be also optimal, which is
Figure 1: Relative errors of $\phi$-FEM for the 1st test case and $k = 1$. Left: $\phi$-FEM with ghost penalty $\sigma = 20$; Right: $\phi$-FEM without ghost penalty ($\sigma = 0$).

Figure 2: Condition numbers for $\phi$-FEM in the 1st test case and $k = 1$. Left: $\phi$-FEM with ghost penalty $\sigma = 20$; Right: $\phi$-FEM without ghost penalty ($\sigma = 0$).
Figure 3: Relative errors of $\phi$-FEM for the 1st test case. Left: $k = 2$; Right: $k = 3$.

Figure 4: Influence of the ghost penalty parameter $\sigma$ on the relative error $|u - u_h|_{1,\Omega} / |u|_{1,\Omega}$ for $\phi$-FEM in the 1st test case. Left: $k = 1$; Right: $k = 2$.

Figure 5: Relative errors of $\phi$-FEM for the 2nd test case. Left: $k = 1$; Right: $k = 2$. The reference solution $u_{ref}$ is computed by a standard FEM on a sufficiently fine fitted mesh on $\Omega$. 
better than the theoretical prediction. We have thus an easily implementable optimally convergent finite element method suitable for non-fitted meshes and robust with respect to the cuts of the mesh with the domain boundary. This comes at the expense of augmenting the polynomial degrees in the finite element formulation in comparison with the standard FEM and thus necessitating higher order quadrature rules. It would be interesting to investigate the effect of “under-integrating”, i.e. lowering the quadrature order, on the accuracy of the method.

Of course, the scope of the present article is very limited and academic: we only consider here the Poisson equation with homogeneous boundary conditions. An extension to non-homogeneous Dirichlet condition $u = g$ on $\Gamma$ and to a more general second order equation (40) is straightforward. It is presented (without any theoretical analysis) in the 3rd test case above. On the other hand, treating Neumann or Robin boundary conditions is a completely different matter. We announce here an ongoing work [9], where a Neumann problem is discretized in the $\phi$-FEM manner by introducing some auxiliary unknowns on $\Omega^\Gamma_h$. Future endeavors should also be devoted to more complicated governing equations and boundary conditions.

References


