

## A group of immersed finite-element spaces for elliptic interface problems

RUCHI GUO AND TAO LIN\*

Department of Mathematics, Virginia Tech, Blacksburg, Virginia, USA

\*Corresponding author: tlin@vt.edu

[Received on 30 April 2017; revised on 25 September 2017]

We present a unified framework for developing and analyzing immersed finite-element (IFE) spaces for solving typical elliptic interface problems with interface-independent meshes. This framework allows us to construct a group of new IFE spaces with either linear, or bilinear, or the rotated- $Q_1$  polynomials. Functions in these IFE spaces are locally piecewise polynomials defined according to the subelements formed by the interface itself instead of its line approximation. We show that the unisolvence for these IFE spaces follows from the invertibility of the Sherman–Morrison matrix. A group of estimates and identities are established for the interface geometry and shape functions that are applicable to all of these IFE spaces. These fundamental preparations enable us to develop a unified multipoint Taylor expansion procedure for proving that these IFE spaces have the expected optimal approximation capability according to the involved polynomials.

*Keywords:* interface problems; discontinuous coefficients; finite-element spaces; Cartesian mesh.

### 1. Introduction

This article presents a unified framework for developing and analyzing a group of immersed finite-element (IFE) spaces that use interface-independent meshes (such as highly structured Cartesian meshes) to solve interface problems of the typical second-order elliptic partial differential equations:

$$-\nabla \cdot (\beta \nabla u) = f, \quad \text{in } \Omega^- \cup \Omega^+, \quad (1.1)$$

$$u = g, \quad \text{on } \partial\Omega, \quad (1.2)$$

where, without loss of generality, the domain  $\Omega \subseteq \mathbb{R}^2$  is separated by an interface curve  $\Gamma$  into two subdomains  $\Omega^+$  and  $\Omega^-$ , the diffusion coefficient  $\beta(X)$  is discontinuous such that

$$\beta(X) = \begin{cases} \beta^- & \text{if } X \in \Omega^-, \\ \beta^+ & \text{if } X \in \Omega^+, \end{cases}$$

where  $\beta^\pm$  are positive constants. In addition, the solution  $u$  is assumed to satisfy the jump conditions:

$$[u]_\Gamma = 0, \quad (1.3)$$

$$[\beta \nabla u \cdot \mathbf{n}]_\Gamma = 0, \quad (1.4)$$

where  $\mathbf{n}$  is the unit normal vector to the interface  $\Gamma$ , and for every piecewise function  $v$  defined as

$$v = \begin{cases} v^-(X) & \text{if } X \in \Omega^-, \\ v^+(X) & \text{if } X \in \Omega^+, \end{cases}$$

we adopt the notation  $[v]_\Gamma = v^+|_\Gamma - v^-|_\Gamma$ .

It is well known that the standard finite-element method can be applied to interface problems, provided that the mesh is formed according to the interface (see Babuška, 1970; Bramble & Thomas King, 1996; Chen & Zou, 1998, and references therein). Many efforts have been made to develop alternative finite-element methods based on unfitted meshes for solving interface problems. The advantages of using unfitted meshes are discussed in Bordas *et al.* (2010), Moumnassi *et al.* (2011, 2014) and Nadal *et al.* (2013). A variety of finite-element methods that can use interface-independent meshes to solve interface problems have been reported in the literature (see Barrett & Elliott, 1987; Melenk & Babuška, 1996; Babuška & Zhang, 1998; Dolbow *et al.*, 2001; Babuška *et al.*, 2003; Bastian & Engwer, 2009; Efendiev & Hou, 2009; Guyomarc'h *et al.*, 2009; Guzmán *et al.*, 2017) for a few examples. In particular, instead of modifying the shape functions on interface elements, which is an approach to be discussed in this article, methods in Burman *et al.* (2015, 2017), Hansbo & Hansbo (2002) and Wang *et al.* (2016) employ standard finite-element functions, but use the Nitsche's penalty along the interface in the finite-element schemes.

This article focuses on the IFE methods, whose basic idea was introduced by Li (1998), for those applications where it is preferable to solve interface problems with a mesh independent of the interface, e.g., the particle-in-cell method for simulating plasma particles (Lin & Wang, 2002; Kafafy *et al.*, 2005; Kafafy & Wang, 2007), the problems with moving interfaces (He *et al.*, 2013; Lin *et al.*, 2013a) and the electroencephalography forward problem (Vallaghé & Papadopoulo, 2010). IFE methods for interface problems of other types of partial differential equations can be found in Adjerid *et al.* (2015), He *et al.* (2013), Hou *et al.* (2012), Li & Yang (2005), Lin *et al.* (2011, 2013a,b), Moon (2016) and Yang *et al.* (2003).

The IFE spaces developed in this article are extended from the IFE spaces constructed with linear polynomials (Li *et al.*, 2003, 2004; Gong *et al.*, 2007/08; Kwak *et al.*, 2010), bilinear polynomials (Lin *et al.*, 2001; He *et al.*, 2008; He, 2009) and rotate- $Q_1$  polynomials (Zhang, 2013; Lin *et al.*, 2013b) using the standard Lagrange-type local degrees of freedom imposed either at element vertices as usual, or at midpoints of element edges in the Crouzeix–Raviart way (Crouzeix & Raviart, 1973). We note that the local linear IFE space on each triangular interface element constructed here with the Lagrange local degrees of freedom imposed at vertices is very similar to the one recently introduced in Guzmán *et al.* (2015). The IFE spaces in this article are new because, locally on each interface element, they are Hsieh–Clough–Tocher-type macro finite-element functions (Clough & Tocher, 1966; Braess, 2001) defined with subelements formed by the interface curve itself in contrast to those IFE spaces in the literature defined with subelements formed by a straight line approximating the interface curve.

Our research presented here is motivated by two issues. The first issue concerns the general  $\mathcal{O}(h^2)$  order accuracy for a line to approximate a curve which is a fundamental ingredient for the optimal approximation capability of those IFE spaces in the literature. We hope the study of IFE spaces based on curve subelements can shed light on the development of higher-degree IFE spaces for which the  $\mathcal{O}(h^2)$  order is not sufficient. For examples, those techniques incorporating the exact geometry for constructing basis functions (Sevilla *et al.*, 2008; Lian *et al.*, 2012; Nguyen *et al.*, 2015) may be considered. The second issue is the attempt to unify the fragmented framework for developing and analyzing the IFE spaces in the literature. For IFE spaces based on different meshes, different polynomials and different local degrees of freedom, we show that their unisolvence, i.e., the existence and uniqueness of IFE shape functions, can be established through a uniform procedure related with the invertibility of the Sherman–Morrison matrix. We have derived a group of identities for the interface geometry and shape functions that are applicable to all of these IFE spaces, and this enables us to derive error estimates for the interpolation in these new IFE spaces in a general unified multipoint Taylor expansion approach, in which IFE functions defined according to the given interface actually simplify the analysis because we only need to apply the

same arguments to two subelements formed by the interface, while the analysis for the IFE spaces in the literature has to use a different set of arguments to handle the subelements sandwiched between the interface curve and its approximate line. Also inspired by [Guzmán \*et al.\* \(2015\)](#), we have made an effort to show how the error bounds explicitly depend on the maximum curvature of the interface curve and the ratio between  $\beta^-$  and  $\beta^+$ , which are two important problem-dependent characteristics affecting the approximation capability of IFE spaces. We note that the dependence of constants in the error bounds on the ratio between  $\beta^-$  and  $\beta^+$  in our article is similar to the one discussed in [Chu \*et al.\* \(2010\)](#), and we think this coincidence follows from the fact that we analyze the approximation capability of finite-element spaces with Lagrange-type degrees of freedom.

This article consists of five additional sections. In the next section, we describe common notations and some basic assumptions used in this article. In Section 3, we derive estimates and identities associated with the interface and the jump conditions in an element. From these estimates, we can see how their bounds explicitly depend on curvature of the interface and the ratio between  $\beta^-$  and  $\beta^+$ , and how the mesh size  $h$  is subject to the interface curvature. In Section 4, we present generalized multipoint Taylor expansions for piecewise  $C^2$  functions in an interface element. Estimates for the remainders in these expansions are derived in terms of pertinent Sobolev norms. In Section 5, first, we establish the unisolvence of immersed finite-element functions constructed with linear, bilinear, Crouzeix–Raviart and rotated- $Q_1$  polynomials, i.e., we show that the standard Lagrange local degrees of freedom imposed at the nodes of an interface element can uniquely determine an IFE function that satisfies the interface jump conditions in a suitable approximate sense. Then, we show that the IFE shape functions have several desirable properties such as the partition of unity and the critical identities in [Theorem 5.11](#). Finally, with a unified analysis, we show that the IFE spaces have the expected optimal approximation capability. In Section 6, we demonstrate features of these IFE spaces by numerical examples.

## 2. Preliminaries

Throughout the article,  $\Omega \subset \mathbb{R}^2$  denotes a bounded domain as a union of finitely many rectangles. The interface curve  $\Gamma$  separates  $\Omega$  into two subdomains  $\Omega^+$  and  $\Omega^-$  such that  $\overline{\Omega} = \overline{\Omega^+} \cup \overline{\Omega^-} \cup \Gamma$ . For every measurable subset,  $\tilde{\Omega} \subseteq \Omega$ , let  $W^{k,p}(\tilde{\Omega})$  be the standard Sobolev spaces on  $\tilde{\Omega}$  associated with the norm  $\|\cdot\|_{k,p,\tilde{\Omega}}$  and the seminorm  $|v|_{k,p,\tilde{\Omega}} = \|D^\alpha v\|_{0,p,\tilde{\Omega}}$ , for  $|\alpha| = k$ . The corresponding Hilbert space is  $H^k(\tilde{\Omega}) = W^{k,2}(\tilde{\Omega})$ . When  $\tilde{\Omega}^s = \tilde{\Omega} \cap \Omega^s \neq \emptyset$ ,  $s = \pm$ , we let

$$PH_{int}^k(\tilde{\Omega}) = \{u : u|_{\tilde{\Omega}^s} \in H^k(\tilde{\Omega}^s), s = \pm; [u]_\Gamma = 0 \text{ and } [\beta \nabla u \cdot \mathbf{n}]_\Gamma = 0 \text{ on } \Gamma \cap \tilde{\Omega}\},$$

$$PC_{int}^k(\tilde{\Omega}) = \{u : u|_{\tilde{\Omega}^s} \in C^k(\tilde{\Omega}^s), s = \pm; [u]_\Gamma = 0 \text{ and } [\beta \nabla u \cdot \mathbf{n}]_\Gamma = 0 \text{ on } \Gamma \cap \tilde{\Omega}\}.$$

The norms and seminorms to be used on  $PH_{int}^k(\tilde{\Omega})$  are

$$\begin{aligned} \|\cdot\|_{k,\tilde{\Omega}}^2 &= \|\cdot\|_{k,\tilde{\Omega}^+}^2 + \|\cdot\|_{k,\tilde{\Omega}^-}^2, & |\cdot|_{k,\tilde{\Omega}} &= |\cdot|_{k,\tilde{\Omega}^+} + |\cdot|_{k,\tilde{\Omega}^-}, \\ \|\cdot\|_{k,\infty,\tilde{\Omega}} &= \max(\|\cdot\|_{k,\infty,\tilde{\Omega}^+}, \|\cdot\|_{k,\infty,\tilde{\Omega}^-}), & |\cdot|_{k,\infty,\tilde{\Omega}} &= \max(|\cdot|_{k,\infty,\tilde{\Omega}^+}, |\cdot|_{k,\infty,\tilde{\Omega}^-}). \end{aligned}$$

Let  $\mathcal{T}_h$  be a Cartesian triangular or rectangular mesh of the domain  $\Omega$  with the maximum length of edge  $h$ . An element  $T \in \mathcal{T}_h$  is called an interface element, provided the interior of  $T$  intersects with the interface  $\Gamma$ ; otherwise, we name it a noninterface element. We let  $\mathcal{T}_h^i$  and  $\mathcal{T}_h^n$  be the set of interface elements and noninterface elements, respectively. Similarly,  $\mathcal{E}_h^i$  and  $\mathcal{E}_h^n$  are sets of interface edges and

noninterface edges, respectively. In addition, as in He *et al.* (2009), we assume that  $\mathcal{T}_h$  satisfies the following hypotheses when the mesh size  $h$  is small enough:

- (H1) The interface  $\Gamma$  cannot intersect an edge of any element at more than two points, unless the edge is part of  $\Gamma$ .
- (H2) If  $\Gamma$  intersects the boundary of an element at two points, these intersection points must be on different edges of this element.
- (H3) The interface  $\Gamma$  is a piecewise  $C^2$  function, and the mesh  $\mathcal{T}_h$  is formed such that the subset of  $\Gamma$  in every interface element  $T \in \mathcal{T}_h^i$  is  $C^2$ .
- (H4) The interface  $\Gamma$  is smooth enough so that  $PC_{int}^2(T)$  is dense in  $PH_{int}^2(T)$  for every interface element  $T \in \mathcal{T}_h^i$ .

On an element  $T \in \mathcal{T}_h$ , we consider the local finite-element space  $(T, \Pi_T, \Sigma_T^P)$  with

$$\Pi_T = \begin{cases} \text{Span}\{1, x, y\}, & \text{for } P_1 \text{ or Crouzeix–Raviart (C–R) finite-element functions,} \\ \text{Span}\{1, x, y, xy\}, & \text{for } Q_1 \text{ (bilinear) finite-element functions,} \\ \text{Span}\{1, x, y, x^2 - y^2\}, & \text{for rotated-}Q_1 \text{ finite-element functions,} \end{cases} \quad (2.1)$$

$$\Sigma_T^P = \{\psi_T^P(M_i) : i \in \mathcal{I}, \forall \psi_T^P \in \Pi_T\}, \quad (2.2)$$

where  $\mathcal{I} = \{1, 2, \dots, DOF(T)\}$ ,  $DOF(T) = 3$  or  $4$  depending on whether  $T$  is triangular or rectangular,  $M_i, i \in \mathcal{I}$  are the local nodes to determine shape functions on  $T$  and the superscript  $P$  is to emphasize the Lagrange-type degrees of freedom imposed at the points  $M_i$ s. For  $P_1$  and  $Q_1$  finite elements,  $M_i = A_i, i \in \mathcal{I}$ , where  $A_i$ 's are vertices of  $T$ . For C-R and rotated- $Q_1$  finite elements,  $M_i$  is the midpoint of the  $i$ th edge of  $T$  for  $i \in \mathcal{I}$ . It is well known (Crouzeix & Raviart, 1973; Ciarlet, 1978; Rannacher & Turek, 1992; Brenner & Scott, 1994) that  $(T, \Pi_T, \Sigma_T^P)$  has a set of shape functions  $\psi_i^P(X), i \in \mathcal{I}$  such that

$$\psi_{i,T}^P(M_j) = \delta_{ij}, \quad \|\psi_{i,T}^P\|_{\infty,T} \leq C, \quad \|\nabla \psi_{i,T}^P\|_{\infty,T} \leq Ch^{-1}, \quad i, j \in \mathcal{I}, \quad (2.3)$$

where  $\delta_{ij}$  is the Kronecker delta function.

Throughout this article, without loss of generality, we assume that  $\beta^+ \geq \beta^-$  and let  $\rho = \beta^-/\beta^+ \leq 1$ . In addition, on any  $T \in \mathcal{T}_h^i$ , we use  $D, E$  to denote the intersection points of  $\Gamma$  and  $\partial T$ , and let  $l$  be the line connecting  $DE$ .

### 3. Geometric properties of the interface

In this section, we discuss geometric properties on interface elements that are useful for developing and analyzing IFE spaces. Let  $T$  be an interface element. As illustrated in Fig. 1, for a point  $\tilde{X}$  on  $\Gamma \cap T$ , we let  $\mathbf{n}(\tilde{X}) = (\tilde{n}_x(\tilde{X}), \tilde{n}_y(\tilde{X}))$  be the normal of  $\Gamma$  at  $\tilde{X}$ , and we let  $\tilde{X}_\perp \in l$  be the orthogonal projection of  $\tilde{X} \in \Gamma \cap T$  onto the line  $l$ . Also, for the line  $l$ , we let  $\bar{\mathbf{n}} = (\bar{n}_x, \bar{n}_y)$  be its unit normal vector and, consequently,  $\bar{\mathbf{t}} = (\bar{n}_y, -\bar{n}_x)$  is the vector tangential to  $l$ . Without loss of generality, we assume the orientation of all the normal vectors are from  $T^-$  to  $T^+$ . In addition, we let  $\kappa$  be the maximum curvature of the curve  $\Gamma$ .

For  $T \in \mathcal{T}_h^i$ , without loss of generality, we can introduce a local coordinate system such that the point on  $\Gamma \cap T$  can be expressed as  $(\xi, \eta) = (\xi, \omega(\xi))$  for a suitable function  $\omega$  such that, in this local system,

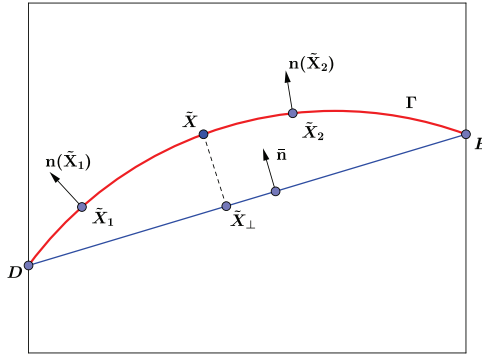


FIG. 1. The local system.

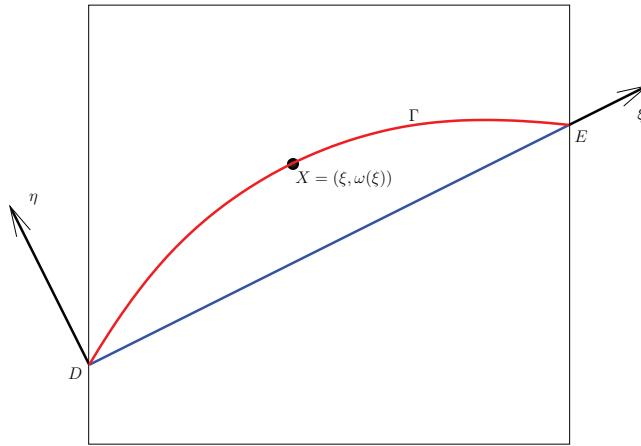


FIG. 2. The geometry of an interface element.

$D$  is origin and its  $x$ -axis is aligned with  $l$ , as shown in Fig. 2. We start from the following lemma, which extends similar results in Li *et al.* (2004):

LEMMA 3.1 Given any  $\epsilon \in (0, \frac{\sqrt{2}}{2})$ , assume  $h\kappa \leq \epsilon$ , then for any interface element  $T \in \mathcal{T}_h^i$ , there hold

$$|w(\xi)| \leq 2(1 - 2\epsilon^2)^{-3/2} \kappa h^2, \tag{3.1}$$

$$|w'(\xi)| \leq \sqrt{2}(1 - 2\epsilon^2)^{-3/2} \kappa h. \tag{3.2}$$

*Proof.* In the local system, let  $\xi_E$  be the coordinate of the point  $E$ . And by the mean value theorem, there is some  $\xi_0 \in [0, \xi_E]$  such that  $\omega'(\xi_0) = 0$ . Consider a function  $g$  as well as its derivative

$$g(\xi) = \frac{\omega'(\xi)}{\sqrt{1 + (\omega'(\xi))^2}} \quad \text{and} \quad g'(\xi) = \frac{\omega''(\xi)}{(1 + (\omega'(\xi))^2)^{3/2}}.$$

Note that  $g'(\xi)$  is the curvature of  $\Gamma$  at  $\xi$ ; hence, we have  $|g'(\xi)| \leq \kappa$ . Then, by  $g(\xi_0) = 0$ , we have  $|g(\xi)| = \left| \int_{\xi_0}^{\xi} g'(s) ds \right| \leq \int_{\xi_0}^{\xi} |g'(s)| ds \leq \sqrt{2}\kappa h = \sqrt{2}\epsilon$ , which implies  $|\omega'(\xi)| \leq \frac{\sqrt{2}\epsilon}{\sqrt{1-2\epsilon^2}}$ . By the definition of  $\kappa$ , we have  $|\omega''(\xi)| \leq (1 - 2\epsilon^2)^{-3/2}\kappa$ .

Now using the Taylor expansion for  $\omega$  around  $D$  leads to  $\omega(\xi) = \omega'(0)\xi + \frac{1}{2}\omega''(\tilde{\xi})\xi^2$  for some  $\tilde{\xi} \in [0, \xi_E]$ . Note that  $\omega(\xi_E) = 0$  shows  $\omega'(0) = -\frac{1}{2}\omega''(\tilde{\xi}_E)\xi_E$ . Thus, we have  $\omega(\xi) = -\frac{1}{2}\omega''(\tilde{\xi}_E)\xi_E\xi + \frac{1}{2}\omega''(\tilde{\xi})\xi^2$  and therefore,  $|\omega(\xi)| \leq 2\|\omega''\|_{\infty}h^2 \leq 2(1 - 2\epsilon^2)^{-3/2}\kappa h^2$ , which yields (3.1). Using the Taylor expansion again for  $\omega'$  around  $\xi_0$ , we have  $\omega'(\xi) = \omega''(\tilde{\xi})(\xi - \xi_0)$  for some  $\tilde{\xi}$  between  $\xi_0$  and  $\xi$ . Finally, we obtain  $|\omega'(\xi)| \leq \sqrt{2}\|\omega''\|h \leq \sqrt{2}(1 - 2\epsilon^2)^{-3/2}h$ .  $\square$

We note that the argument in the lemma above is similar to [Chu et al. \(2010, Assumption 3.14\)](#) with a minor difference that a local polar coordinate system is used on the interface element in [Chu et al. \(2010\)](#). The following lemmas provide estimates about various geometric quantities defined at points on  $\Gamma \cap T$ .

**LEMMA 3.2** Given any  $\epsilon \in (0, \frac{\sqrt{2}}{2})$ , assume  $h\kappa \leq \epsilon$ , then for any interface element  $T \in \mathcal{T}_h^i$  and any point  $\tilde{X} \in \Gamma \cap T$ , the following inequality holds:

$$\|\tilde{X} - \tilde{X}_{\perp}\| \leq 2(1 - 2\epsilon^2)^{-3/2}\kappa h^2, \tag{3.3}$$

and for any  $\tilde{X}_1, \tilde{X}_2 \in \Gamma \cap T$ , we have

$$\|\mathbf{n}(\tilde{X}_1) - \mathbf{n}(\tilde{X}_2)\| \leq \sqrt{2}(1 + (1 - 2\epsilon^2)^{-3/2})\kappa h, \tag{3.4a}$$

$$\mathbf{n}(\tilde{X}_1) \cdot \mathbf{n}(\tilde{X}_2) \geq 1 - 2(1 + (1 - 2\epsilon^2)^{-3/2})\kappa^2 h^2. \tag{3.4b}$$

*Proof.* Estimate (3.3) directly follows from (3.1). For (3.4a), we assume  $\tilde{X}_1 = (\xi_1, w(\xi_1))$  and  $\tilde{X}_2 = (\xi_2, w(\xi_2))$  in the local system, respectively. Then, we have

$$\mathbf{n}(\tilde{X}_1) = \frac{1}{\sqrt{1 + (w'(\xi_1))^2}} \begin{pmatrix} -w'(\xi_1) \\ 1 \end{pmatrix}, \quad \mathbf{n}(\tilde{X}_2) = \frac{1}{\sqrt{1 + (w'(\xi_2))^2}} \begin{pmatrix} -w'(\xi_2) \\ 1 \end{pmatrix}.$$

By the calculation in Lemma 3.1 and mean value theorem, there is some  $\tilde{\xi} \in [0, \xi_E]$  such that

$$\left| \frac{w'(\xi_1)}{\sqrt{1 + (w'(\xi_1))^2}} - \frac{w'(\xi_2)}{\sqrt{1 + (w'(\xi_2))^2}} \right| = \frac{|\omega''(\tilde{\xi})|}{(1 + (\omega'(\tilde{\xi}))^2)^{3/2}} |\xi_1 - \xi_2| \leq \sqrt{2}\kappa h,$$

and  $\tilde{\xi} \in [0, \xi_E]$  such that

$$\left| \frac{1}{\sqrt{1 + (w'(\xi_1))^2}} - \frac{1}{\sqrt{1 + (w'(\xi_2))^2}} \right| = \frac{|\omega'(\tilde{\xi})||\omega''(\tilde{\xi})|}{(1 + (\omega'(\tilde{\xi}))^2)^{3/2}} |\xi_1 - \xi_2| \leq 2(1 - 2\epsilon^2)^{-3/2}\epsilon\kappa h.$$

Then, (3.4a) follows by applying these estimates in the local coordinate forms of  $\mathbf{n}(\tilde{X}_1)$  and  $\mathbf{n}(\tilde{X}_2)$ . Furthermore, by (3.4a) and

$$\|\mathbf{n}(\tilde{X}_1) - \mathbf{n}(\tilde{X}_2)\|^2 = \|\mathbf{n}(\tilde{X}_1)\|^2 + \|\mathbf{n}(\tilde{X}_2)\|^2 - 2\mathbf{n}(\tilde{X}_1) \cdot \mathbf{n}(\tilde{X}_2) = 2 - 2\mathbf{n}(\tilde{X}_1) \cdot \mathbf{n}(\tilde{X}_2),$$

we have (3.4b). □

REMARK 3.3 Note that there exists a point  $\tilde{X}_1 = (\xi_1, w(\xi_1)) \in \Gamma \cap T$  such that  $w'(\xi_1) = 0$ , which means  $\mathbf{n}(\tilde{X}_1) = \bar{\mathbf{n}}$ . Hence, by Lemma 3.2, we have the following estimates for an arbitrary point  $\tilde{X} \in \Gamma \cap T$ :

$$\|\mathbf{n}(\tilde{X}) - \bar{\mathbf{n}}\| \leq \sqrt{2}(1 + (1 - 2\epsilon^2)^{-3/2})\kappa h, \tag{3.5a}$$

$$\mathbf{n}(\tilde{X}) \cdot \bar{\mathbf{n}} \geq 1 - 2(1 + (1 - 2\epsilon^2)^{-3/2})^2 \kappa^2 h^2. \tag{3.5b}$$

The two lemmas above have suggested a criteria about how small  $h$  should be according to the maximum curvature  $\kappa$  of  $\Gamma$  so that the related analysis is valid. Therefore, for all discussions from now on, we further assume that

- $h$  is sufficiently small such that for some fixed parameter  $\epsilon \in (0, \sqrt{2}/2)$  and  $\bar{\kappa} \in (0, 1]$  of one's own choice, there holds

$$h < \min \left\{ \frac{\sqrt{\bar{\kappa}}}{\sqrt{2}(1 + (1 - 2\epsilon^2)^{-3/2})\kappa}, \frac{\epsilon}{\kappa} \right\}. \tag{3.6}$$

Obviously  $\epsilon$  is the proportion by which we should choose the mesh size  $h$  according to the interface curvature  $\kappa$ . Also, by (3.6) and (3.5b), we have

$$\mathbf{n}(\tilde{X}) \cdot \bar{\mathbf{n}} \geq 1 - \bar{\kappa}, \tag{3.7}$$

which shows how much the angle between the normal of  $\Gamma \cap T$  and  $\bar{\mathbf{n}}$  can vary in an interface element  $T \in \mathcal{T}_h^i$ , a larger value of  $\bar{\kappa} \in (0, 1]$  allows  $\mathbf{n}(\tilde{X})$  to vary more from  $\bar{\mathbf{n}}$  up to, but not equal to, 90 degrees. Therefore, we will call  $\bar{\kappa}$  the angle allowance.

In the rest of this article, all the generic constants  $C$  are assumed to possibly depend only on the parameter  $\epsilon$  and  $\bar{\kappa}$ , but they are independent of the interface location,  $\beta^\pm$  and the curvature  $\kappa$ .

We now consider some matrices associated with the normal of interface  $\Gamma$  and the normal of  $l$ . First, for any  $\tilde{X} \in \Gamma \cap T$ , we use the normal  $\mathbf{n}(\tilde{X})$  to form two matrices:

$$N^s(\tilde{X}) = \begin{pmatrix} \tilde{n}_y(\tilde{X}) & -\tilde{n}_x(\tilde{X}) \\ \beta^s \tilde{n}_x(\tilde{X}) & \beta^s \tilde{n}_y(\tilde{X}) \end{pmatrix}, \quad s = \pm.$$

Since  $Det(N^s(\tilde{X})) = \tilde{\beta}^s, s = \pm$ , these matrices are nonsingular; therefore, we can define another two matrices at the point  $\tilde{X} \in \Gamma \cap T$ :

$$M^-(\tilde{X}) = (N^+(\tilde{X}))^{-1} N^-(\tilde{X}) = \begin{pmatrix} \tilde{n}_y^2(\tilde{X}) + \rho \tilde{n}_x^2(\tilde{X}) & (\rho - 1)\tilde{n}_x(\tilde{X})\tilde{n}_y(\tilde{X}) \\ (\rho - 1)\tilde{n}_x(\tilde{X})\tilde{n}_y(\tilde{X}) & \tilde{n}_x^2(\tilde{X}) + \rho \tilde{n}_y^2(\tilde{X}) \end{pmatrix}, \tag{3.8}$$

$$M^+(\tilde{X}) = (N^-(\tilde{X}))^{-1} N^+(\tilde{X}) = \begin{pmatrix} \tilde{n}_y^2(\tilde{X}) + 1/\rho \tilde{n}_x^2(\tilde{X}) & (1/\rho - 1)\tilde{n}_x(\tilde{X})\tilde{n}_y(\tilde{X}) \\ (1/\rho - 1)\tilde{n}_x(\tilde{X})\tilde{n}_y(\tilde{X}) & \tilde{n}_x^2(\tilde{X}) + 1/\rho \tilde{n}_y^2(\tilde{X}) \end{pmatrix}. \tag{3.9}$$

For matrices  $M^-(\tilde{X})$  and  $M^+(\tilde{X})$ , we recall from Li *et al.* (2004) the following results:

$$\nabla u^+(\tilde{X}) = M^-(\tilde{X})\nabla u^-(\tilde{X}), \quad \nabla u^-(\tilde{X}) = M^+(\tilde{X})\nabla u^+(\tilde{X}), \quad \forall \tilde{X} \cap T \in \Gamma, \quad \forall u \in PC_{int}^2(T). \tag{3.10}$$

In addition, for  $\tilde{X} \in \Gamma \cap T$ , we can use the normal vectors  $\mathbf{n}(\tilde{X})$  and  $\bar{\mathbf{n}}$  to form the following matrices:

$$\bar{N}^s(\tilde{X}) = \begin{pmatrix} \bar{n}_y & -\bar{n}_x \\ \beta^s \bar{n}_x(\tilde{X}) & \beta^s \bar{n}_y(\tilde{X}) \end{pmatrix}, \quad s = \pm.$$

By Remark 3.3, we have

$$\text{Det}(\bar{N}^s(\tilde{X})) = \beta^s \mathbf{n}(\tilde{X}) \cdot \bar{\mathbf{n}} \geq \beta^s (1 - \bar{\kappa}), \quad s = \pm,$$

which means  $\bar{N}^s(\tilde{X})$  are nonsingular when  $h$  is small enough; hence, we can use them to form

$$\bar{M}^+(\tilde{X}) = (\bar{N}^-(\tilde{X}))^{-1} \bar{N}^+(\tilde{X}), \quad \bar{M}^-(\tilde{X}) = (\bar{N}^+(\tilde{X}))^{-1} \bar{N}^-(\tilde{X}). \tag{3.11}$$

LEMMA 3.4 For the mesh  $\mathcal{T}_h$  with  $h$  sufficiently small, there exists a constant  $C$  independent of interface location,  $\beta^\pm$  and  $\kappa$ , such that for two arbitrary points  $\tilde{X}_i, i = 1, 2$  on  $\Gamma \cap T$ , we have

$$\|\bar{M}^-(\tilde{X}_1)\| \leq C, \quad \|\bar{M}^+(\tilde{X}_1)\| \leq \frac{C}{\rho}, \quad \|M^-(\tilde{X}_2)\| \leq C, \quad \|M^+(\tilde{X}_2)\| \leq \frac{C}{\rho} \tag{3.12}$$

and

$$\|\bar{M}^-(\tilde{X}_1) - M^-(\tilde{X}_2)\| \leq C\kappa h, \quad \|\bar{M}^+(\tilde{X}_1) - M^+(\tilde{X}_2)\| \leq \frac{C\kappa}{\rho} h. \tag{3.13}$$

*Proof.* (3.12) can be verified directly. We only prove (3.13) for the  $-$  case and the arguments for the  $+$  case are similar. For simplicity, we denote  $\mathbf{n}(\tilde{X}_i) = (\tilde{n}_{ix}, \tilde{n}_{iy}), i = 1, 2$ . Then, by direct calculations, we have

$$\begin{aligned} \bar{M}^-(\tilde{X}_1) - M^-(\tilde{X}_2) &= \begin{pmatrix} \tilde{n}_{1y}\bar{n}_y - \tilde{n}_{2y}^2 + \rho(\tilde{n}_{1x}\bar{n}_x - \tilde{n}_{2x}^2) & (\rho - 1)(\tilde{n}_{1y}\bar{n}_x - \tilde{n}_{2x}\tilde{n}_{2y}) \\ (\rho - 1)(\tilde{n}_{1x}\bar{n}_y - \tilde{n}_{2x}\tilde{n}_{2y}) & \tilde{n}_{1x}\bar{n}_x - \tilde{n}_{2x}^2 + \rho(\tilde{n}_{1y}\bar{n}_y - \tilde{n}_{2y}^2) \end{pmatrix} \\ &+ \frac{1 - \bar{\mathbf{n}} \cdot \mathbf{n}(\tilde{X}_1)}{\bar{\mathbf{n}} \cdot \mathbf{n}(\tilde{X}_1)} \begin{pmatrix} \tilde{n}_{1y}\bar{n}_y + \rho\tilde{n}_{1x}\bar{n}_x & (\rho - 1)\tilde{n}_{1y}\bar{n}_x \\ (\rho - 1)\tilde{n}_{1x}\bar{n}_y & \tilde{n}_{1x}\bar{n}_x + \rho\tilde{n}_{1y}\bar{n}_y \end{pmatrix}. \end{aligned}$$

By the triangular inequality, (3.4a), (3.5b), (3.6) and  $\rho \leq 1$ , we can verify that  $\|\bar{M}^-(\tilde{X}_1) - M^-(\tilde{X}_2)\| \leq C\kappa h$  for a constant  $C$  independent of interface location,  $\beta^\pm$  and  $\kappa$ .  $\square$

The following lemmas provide a group of identities on interface elements.

LEMMA 3.5 For the mesh  $\mathcal{T}_h$  with  $h$  sufficiently small, the following results hold for all  $\tilde{X} \in \Gamma$ :

- $\bar{M}^-(\tilde{X})$  and  $\bar{M}^+(\tilde{X})$  are inverse matrices to each other, i.e.,

$$\bar{M}^-(\tilde{X})\bar{M}^+(\tilde{X}) = I, \quad \bar{M}^+(\tilde{X})\bar{M}^-(\tilde{X}) = I. \tag{3.14}$$



- Matrix  $(\bar{M}^-(\tilde{X}))^T$  has two eigenvalues 1 and  $\rho$  with the corresponding eigenvectors  $\bar{\mathbf{t}}$  and  $\mathbf{n}(\tilde{X})$ , i.e.,

$$(\bar{M}^-(\tilde{X}))^T \bar{\mathbf{t}} = \bar{\mathbf{t}}, \quad (\bar{M}^-(\tilde{X}))^T \mathbf{n}(\tilde{X}) = \rho \mathbf{n}(\tilde{X}). \tag{3.15}$$

- Similarly, matrix  $(\bar{M}^+(\tilde{X}))^T$  has two eigenvalues 1 and  $1/\rho$  with the corresponding eigenvectors  $\bar{\mathbf{t}}$  and  $\mathbf{n}(\tilde{X})$ , respectively, i.e.,

$$(\bar{M}^+(\tilde{X}))^T \bar{\mathbf{t}} = \bar{\mathbf{t}}, \quad (\bar{M}^+(\tilde{X}))^T \mathbf{n}(\tilde{X}) = \frac{1}{\rho} \mathbf{n}(\tilde{X}). \tag{3.16}$$

*Proof.* First it is easy to see that  $\bar{M}^-(\tilde{X})\bar{M}^+(\tilde{X}) = (\bar{N}^-(\tilde{X}))^{-1}\bar{N}^+(\tilde{X})(\bar{N}^+(\tilde{X}))^{-1}\bar{N}^-(\tilde{X}) = I$ . Next by direct calculation, we have

$$\bar{M}^-(\tilde{X}) = \frac{1}{\bar{n}_x\bar{n}_x(\tilde{X}) + \bar{n}_y\bar{n}_y(\tilde{X})} \begin{pmatrix} \bar{n}_y\bar{n}_y(\tilde{X}) + \rho\bar{n}_x\bar{n}_x(\tilde{X}) & -\bar{n}_x\bar{n}_y(\tilde{X}) + \rho\bar{n}_x\bar{n}_y(\tilde{X}) \\ -\bar{n}_y\bar{n}_x(\tilde{X}) + \rho\bar{n}_y\bar{n}_x(\tilde{X}) & \bar{n}_x\bar{n}_x(\tilde{X}) + \rho\bar{n}_y\bar{n}_y(\tilde{X}) \end{pmatrix}$$

from which we can easily verify that  $(\bar{M}^-(\tilde{X}))^T \bar{\mathbf{t}} = \bar{\mathbf{t}}$  and  $(\bar{M}^-(\tilde{X}))^T \mathbf{n}(\tilde{X}) = \rho \mathbf{n}(\tilde{X})$ . The results about  $(\bar{M}^+(\tilde{X}))^T$  follow from the fact  $(\bar{M}^-(\tilde{X}))^T(\bar{M}^+(\tilde{X}))^T = I$ .  $\square$

LEMMA 3.6 Let  $\mathcal{T}_h$  be a mesh with  $h$  sufficiently small. Let  $P \in \Omega$  and  $\tilde{X}$  be an arbitrary point on  $\Gamma \cap T$ . Then, the following vectors are independent of  $\bar{X} \in l$ :

$$(\bar{M}^+(\tilde{X}) - I)^T(P - \bar{X}) \quad \text{and} \quad (\bar{M}^-(\tilde{X}) - I)^T(P - \bar{X}).$$

*Proof.* For two arbitrary points  $\bar{X}_i \in l, i = 1, 2, \bar{X}_1 - \bar{X}_2$  is a scalar multiple of  $\bar{\mathbf{t}}$ . Hence, by Lemma 3.5,

$$(\bar{M}^-(\tilde{X}) - I)^T(P - \bar{X}_1) - (\bar{M}^-(\tilde{X}) - I)^T(P - \bar{X}_2) = (\bar{M}^-(\tilde{X}) - I)^T(\bar{X}_1 - \bar{X}_2) = 0,$$

which leads to  $(\bar{M}^-(\tilde{X}) - I)^T(P - \bar{X}_1) = (\bar{M}^-(\tilde{X}) - I)^T(P - \bar{X}_2)$ . Therefore,  $(\bar{M}^-(\tilde{X}) - I)^T(P - \bar{X})$  does not change when  $\bar{X} \in l$  varies. The result for  $(\bar{M}^+(\tilde{X}) - I)^T(P - \bar{X})$  can be proved similarly.  $\square$

#### 4. Multipoint Taylor expansions on interface elements

In this section, extending those in He (2009), He *et al.* (2008), Li *et al.* (2004), Xu (1982) and Zhang (2013), we derive multipoint Taylor expansions in more general formats for a function  $u \in PC_{int}^2(T)$  over an arbitrary interface element  $T \in \mathcal{T}_h^i$ , in which  $u(M_i), i \in \mathcal{I}$  is described in terms of  $u$  and its derivatives at  $x \in T^s, s = \pm$ . We also estimate the remainders in these expansions. As in He *et al.* (2008), we call a point  $X \in T$  an obscure point if one of the lines  $\bar{X}M_i, 1 \leq i \leq DOF(T)$  can intersect  $\Gamma$  more than once. To facilitate a clear expository presentation of the main ideas in our analysis, we carry out error estimation only for interface elements without any obscure points. For the case containing obscure points, we can use a first-order expansion for  $u$  and use the argument that the measure of obscure points is bounded by  $\mathcal{O}(h^3)$ .

First, we partition  $\mathcal{I}$  into two index sets:  $\mathcal{I}^+ = \{i : M_i \in T^+\}$  and  $\mathcal{I}^- = \{i : M_i \in T^-\}$  according to the locations of  $M_i, i \in \mathcal{I}$ . For every  $X \in T$ , we let  $Y_i(t, X) = tM_i + (1 - t)X$ . When  $X$  and  $M_i$  are on

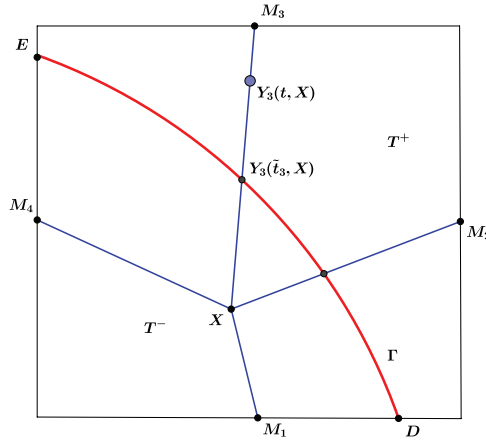


FIG. 3. The expansion of  $u$  in a rectangular interface element.

different sides of  $\Gamma$ , we let  $\tilde{t}_i = \tilde{t}_i(X) \in [0, 1]$  such that  $\tilde{Y}_i = Y_i(\tilde{t}_i, X)$  is on the curve  $\Gamma \cap T$ , see Fig. 3 for an illustration in which the rotated- $Q_1$  finite elements are considered. When  $M_i$  and  $X$  are on the same side of  $\Gamma$ , by the standard second-order Taylor expansion of  $u \in PC_{int}^2(T)$ , we have

$$u^s(M_i) = u^s(X) + \nabla u^s(X) \cdot (M_i - X) + R_i^s(X), \quad i \in \mathcal{I}^s, s = \pm \forall X \in T^s, \tag{4.1}$$

$$\text{with } R_i^s(X) = \int_0^1 (1-t) \frac{d^2}{dt^2} u^s(Y_i(t, X)) dt, \quad i \in \mathcal{I}^s \forall X \in T^s. \tag{4.2}$$

In the following discussion, we denote  $s = \pm, s' = \mp$ , i.e.,  $s$  and  $s'$  take opposite signs whenever a formula have them both. When,  $M_i$  and  $X$  are on different sides of  $\Gamma$ , the expansions in He (2009), He et al. (2008), Li et al. (2004) and Zhang (2013) can be generalized to the following format for  $u \in PC_{int}^2(T)$ :

$$u^{s'}(M_i) = u^s(X) + \nabla u^s(X) \cdot (M_i - X) + ((M^s(\tilde{Y}_i) - I) \nabla u^s(X)) \cdot (M_i - \tilde{Y}_i) + R_i^s(X), \quad i \in \mathcal{I}^{s'} \forall X \in T^s, \tag{4.3}$$

with

$$\begin{cases} R_i^s(X) = R_{i1}^s(X) + R_{i2}^s(X) + R_{i3}^s(X), \quad i \in \mathcal{I}^{s'}, X \in T^s, \\ R_{i1}^s(X) = \int_0^{\tilde{t}_i} (1-t) \frac{d^2 u^s}{dt^2}(Y_i(t, X)) dt, \quad R_{i2}^s(X) = \int_{\tilde{t}_i}^1 (1-t) \frac{d^2 u^{s'}}{dt^2}(Y_i(t, X)) dt, \quad i \in \mathcal{I}^{s'}, X \in T^s, \\ R_{i3}^s(X) = (1 - \tilde{t}_i) \int_0^{\tilde{t}_i} \frac{d}{dt} ((M^s(\tilde{Y}_i) - I) \nabla u^s(Y_i(t, X))) \cdot (M_i - X) dt, \quad i \in \mathcal{I}^{s'}, X \in T^s, \end{cases} \tag{4.4}$$

where  $M^s(\tilde{Y}_i)$  are from (3.8) and (3.9). We proceed to estimate remainders in (4.1) and (4.3).

LEMMA 4.1 Assume  $u \in PC_{int}^2(T)$ . Then, there exist constants  $C > 0$  independent of the interface location and  $\beta^\pm$  such that

$$\int_{T^s} (1-t)^2 |u_{d_1 d_2}(Y_i(t, X))|^2 dX \leq C |u|_{2, T^s}^2, \quad s = \pm \forall t \in [0, 1], \tag{4.5}$$

where  $d_1, d_2 = x$  or  $y$ .

*Proof.* Let  $M_i = (x_i, y_i), i \in \mathcal{I}$  and let  $\xi = tx_i + (1 - t)x$  and  $\eta = ty_i + (1 - t)y$ . For each fixed  $t \in [0, 1]$ , define

$$T^s(t) = \{tM_i + (1 - t)X \mid X \in T^s\}. \tag{4.6}$$

Since  $T^s(t) \subseteq T, s = \pm$ , we have

$$\int_{T^s} (1 - t)^2 |u_{d_1 d_2}(Y(t, X))|^2 dX = \int_{T^s(t)} (1 - t)^2 u_{d_1 d_2}^2(\xi, \eta) (1 - t)^{-2} d\xi d\eta \leq |u|_{2,T}^2,$$

which leads to (4.5). □

By a direct calculation, we have

$$\frac{d^2}{dt^2} u^s(Y_i(t, X)) = (M_i - X)^T H_u^s(Y_i(t, X))(M_i - X), \quad s = \pm, \tag{4.7}$$

where

$$H_u^s(Y_i(t, X)) = \begin{pmatrix} u_{xx}^s(Y_i(t, x)) & u_{xy}^s(Y_i(t, x)) \\ u_{yx}^s(Y_i(t, x)) & u_{yy}^s(Y_i(t, x)) \end{pmatrix}, \quad s = \pm,$$

is the Hessian matrix of  $u^s$ . We are now ready to derive bounds for the remainders in the following lemmas.

LEMMA 4.2 Assume  $u \in PC_{int}^2(T)$ , there exist constants  $C > 0$  independent of the location of the interface and  $\beta^\pm$  such that

$$\|R_i^s\|_{0,T^s} \leq Ch^2 |u|_{2,T}, \quad i \in \mathcal{I}^s, \quad s = \pm, \tag{4.8}$$

$$\|R_{i1}^s\|_{0,T^s} \leq Ch^2 |u|_{2,T}, \quad \|R_{i2}^s\|_{0,T^s} \leq Ch^2 |u|_{2,T}, \quad i \in \mathcal{I}^s, \quad s = \pm. \tag{4.9}$$

*Proof.* According to (4.2) and (4.7), for  $s = \pm$ , we have

$$\begin{aligned} \|R_i^s\|_{0,T^s} &= \left( \int_{T^s} \left( \int_0^1 (1 - t)(M_i - X) H_u^s(Y_i(t, X))(M_i - X) dt \right)^2 dX \right)^{\frac{1}{2}} \\ &\leq Ch^2 \int_0^1 \left( \int_{T^s} (1 - t)^2 (|u_{xx}^s(Y_i(t, X))|^2 + |u_{xy}^s(Y_i(t, X))|^2 + |u_{yy}^s(Y_i(t, X))|^2) dX \right)^{\frac{1}{2}} dt. \end{aligned}$$

Then (4.8) follows from Lemma 4.1. Estimate (4.9) can be derived similarly. □

Moreover, note that for  $i \in \mathcal{I}^s, X \in T^s$  and  $t \in [0, \tilde{t}_i(X)]$ , it can be verified that

$$\frac{d}{dt} \left( (M^s(\tilde{Y}_i) - I) \nabla u^s(Y_i(t, X)) \cdot (M_i - X) \right) = (M_i - X)^T H_u^s(Y_i(t, X))(M^s(\tilde{Y}_i) - I)^T (M_i - X). \tag{4.10}$$

LEMMA 4.3 Assume  $u \in PC_{int}^2(T)$ , there exist constants  $C > 0$  independent of the interface location and  $\beta^\pm$  such that

$$\|R_{i3}^-\|_{0,T^-} \leq Ch^2|u|_{2,T}, \quad i \in \mathcal{I}^+, \quad \|R_{i3}^+\|_{0,T^+} \leq \frac{C}{\rho}h^2|u|_{2,T}, \quad i \in \mathcal{I}^-. \tag{4.11}$$

*Proof.* We only provide the proof of  $\|R_{i3}^-\|_{0,T^-}$  and the argument for  $\|R_{i3}^+\|_{0,T^+}$  is similar. According to (4.10) and the fact  $0 \leq 1 - \tilde{t}_i(X) \leq 1 - t$  for any  $t \in [0, \tilde{t}_i(X)]$ , we have

$$\begin{aligned} \|R_{i3}^-\|_{0,T^-} &= \left( \int_{T^-} (1 - \tilde{t}_i(X))^2 \left( \int_0^{\tilde{t}_i(X)} (M_i - X)^T H_u^-(Y_i(t, X))(M^-(\tilde{Y}_i) - I)^T (M_i - X) dt \right)^2 dX \right)^{\frac{1}{2}} \\ &\leq Ch^2 \left( \int_{T^-} \left( \int_0^1 |1 - t| (|u_{xx}^-(Y_i(t, X))| + |u_{xy}^-(Y_i(t, X))| + |u_{yy}^-(Y_i(t, X))|) dt \right)^2 dX \right)^{\frac{1}{2}} \\ &\leq Ch^2 \int_0^1 \left( \int_{T^-} (1 - t)^2 (|u_{xx}^-(Y_i(t, X))|^2 + |u_{xy}^-(Y_i(t, X))|^2 + |u_{yy}^-(Y_i(t, X))|^2) dX \right)^{\frac{1}{2}} dt, \end{aligned}$$

where we use the fact  $\rho \leq 1$ . Then, applying the estimates in Lemma 4.1 to the above leads to (4.11).  $\square$

### 5. IFE spaces and their properties

In this section, we discuss IFE spaces constructed from the related finite elements  $(T, \Pi_T, \Sigma_T^p)$  for  $T \in \mathcal{T}_h$  described in (2.1) and (2.2). We will first address the unisolvence of the IFEs on interface elements. We will then present a few fundamental properties of IFE functions. Moreover, we will show that these IFE spaces have the optimal approximation capability according to the polynomials used to construct them.

#### 5.1 Local IFE spaces

First, on each element  $T \in \mathcal{T}_h$ , the standard finite element  $(T, \Pi_T, \Sigma_T^p)$  leads to the following local finite-element space:

$$S_h^p(T) = \text{Span}\{\psi_{i,T}^p : i \in \mathcal{I}\}, \tag{5.1}$$

where  $\psi_{i,T}^p, i \in \mathcal{I}$  are the shape functions satisfying (2.3). This local finite-element space is then naturally used as the local IFE space on every noninterface element  $T \in \mathcal{T}_h^n$ . Therefore, our effort here focuses on the local IFE space on interface elements. We will discuss the unisolvence, i.e., we will show that the local degrees of freedom  $\Sigma_T^p$  can uniquely determine an IFE function with a suitable set of interface jump conditions. The unisolvence guarantees the existence and uniqueness of IFE shape functions that can span the local IFE space on interface elements.

Let  $T \in \mathcal{T}_h^i$  be a typical interface element with vertices  $A_i, i \in \mathcal{I}$ . Without loss of generality, we assume

$$\begin{aligned} A_1 &= (0, 0), \quad A_2 = (h, 0), \quad A_3 = (h, h), & (T \text{ is a triangular element}), \\ A_1 &= (0, 0), \quad A_2 = (h, 0), \quad A_3 = (h, h), \quad A_4 = (0, h), & (T \text{ is a rectangular element}), \end{aligned} \tag{5.2}$$

and the edges of  $T$  are denoted as

$$\begin{aligned} b_1 = \overline{A_1A_2}, \quad b_2 = \overline{A_2A_3}, \quad b_3 = \overline{A_3A_1}, & \quad (T \text{ is a triangular element}), \\ b_1 = \overline{A_1A_2}, \quad b_2 = \overline{A_2A_3}, \quad b_3 = \overline{A_3A_4}, \quad b_4 = \overline{A_4A_1}, & \quad (T \text{ is a rectangular element}). \end{aligned} \tag{5.3}$$

On each interface element  $T$ , we consider IFE functions in the following piecewise polynomial format:

$$\phi_T^p(X) = \begin{cases} \phi_T^{p-}(X) = \phi_T^-(X) \in \Pi_T & \text{if } X \in T^-, \\ \phi_T^{p+}(X) = \phi_T^+(X) \in \Pi_T & \text{if } X \in T^+, \end{cases} \tag{5.4}$$

such that it can satisfy the jump conditions (1.3) and (1.4) in an approximate sense as follow:

$$\begin{cases} \phi_T^-|_l = \phi_T^+|_l, & (T \text{ is a triangular element}), \\ \phi_T^-|_l = \phi_T^+|_l, \quad d(\phi_T^-) = d(\phi_T^+), & (T \text{ is a rectangular element}), \end{cases} \tag{5.5}$$

$$\beta^- \nabla \phi_T^-(F) \cdot \mathbf{n}(F) = \beta^+ \nabla \phi_T^+(F) \cdot \mathbf{n}(F), \tag{5.6}$$

where  $d(p)$  denotes the coefficient in the second-degree term for  $p \in \Pi_T$  and  $F$  is an arbitrary point on  $\Gamma \cap T$ . For an IFE function  $\phi_T^p$  such that

$$\phi_T^p(M_i) = v_i, \quad i \in \mathcal{I}, \tag{5.7}$$

we can first expand  $\phi_T^p$  on the subelement with more degrees of freedom, i.e., on  $T^+$  with the assumption that  $|\mathcal{I}^+| \geq |\mathcal{I}^-|$  without loss of generality, and the condition (5.5) then implies that

$$\phi_T^p(X) = \begin{cases} \phi_T^{p-}(X) = \phi_T^{p+}(X) + c_0 L(X) & \text{if } X \in T^-, \\ \phi_T^{p+}(X) = \sum_{i \in \mathcal{I}^-} c_i \psi_{i,T}^p(X) + \sum_{i \in \mathcal{I}^+} v_i \psi_{i,T}^p(X) & \text{if } X \in T^+, \end{cases} \tag{5.8}$$

where the function

$$L(X) = \bar{\mathbf{n}} \cdot (X - D) \tag{5.9}$$

is such that  $L(X) = 0$  is the equation of the line  $l$  and  $\nabla L(X) = \bar{\mathbf{n}}$ .

Recall from Remark 3.3,  $\nabla L(F) \cdot \mathbf{n}(F) = \bar{\mathbf{n}} \cdot \mathbf{n}(F) \geq 1 - \bar{\kappa} > 0$ , when  $h$  is small enough; hence,

$$\mu = \left( \frac{1}{\rho} - 1 \right) \frac{1}{\bar{\mathbf{n}} \cdot \mathbf{n}(F)} \tag{5.10}$$

is well defined, and, by  $\rho \leq 1$ , we have

$$0 \leq \mu \leq \left( \frac{1}{\rho} - 1 \right) \frac{1}{1 - \bar{\kappa}}. \tag{5.11}$$

By condition (5.6), we then have

$$c_0 = \mu \left( \sum_{i \in \mathcal{I}^-} c_i \nabla \psi_{i,T}^P(F) \cdot \mathbf{n}(F) + \sum_{i \in \mathcal{I}^+} v_i \nabla \psi_{i,T}^P(F) \cdot \mathbf{n}(F) \right). \tag{5.12}$$

Putting this formula for  $c_0$  in formula (5.8) for  $\phi_T^P(X)$  and setting  $\phi_T^{P-}(M_j) = v_j$  for  $j \in \mathcal{I}^-$  leads to the following linear system for  $c_i, i \in \mathcal{I}^-$ :

$$\begin{aligned} & \sum_{i \in \mathcal{I}^-} (\psi_{i,T}^P(M_j) + \mu \nabla \psi_{i,T}^P(F) \cdot \mathbf{n}(F) L(M_j)) c_i \\ &= v_j - \sum_{i \in \mathcal{I}^+} (\psi_{i,T}^P(M_j) + \mu \nabla \psi_{i,T}^P(F) \cdot \mathbf{n}(F) L(M_j)) v_i, \quad j \in \mathcal{I}^-. \end{aligned} \tag{5.13}$$

Since  $\psi_{i,T}^P(M_j) = \delta_{ij}$ , for  $i, j \in \mathcal{I}^-$ , we can write the linear system (5.13) in the following matrix form:

$$(I + \mu \delta \boldsymbol{\gamma}^T) \mathbf{c} = \mathbf{b}, \tag{5.14}$$

where  $\mathbf{c} = (c_i)_{i \in \mathcal{I}^-}$ ,

$$\boldsymbol{\gamma} = (\nabla \psi_{i,T}^P(F) \cdot \mathbf{n}(F))_{i \in \mathcal{I}^-}, \quad \boldsymbol{\delta} = (L(M_i))_{i \in \mathcal{I}^-} \tag{5.15}$$

and

$$\mathbf{b} = \left( v_i - \mu L(M_i) \sum_{j \in \mathcal{I}^+} \nabla \psi_{j,T}^P(F) \cdot \mathbf{n}(F) v_j \right)_{i \in \mathcal{I}^-} \tag{5.16}$$

are all column vectors. We proceed to show that  $\phi_T^P(X)$ , i.e., its coefficients  $c_0, \mathbf{c}$  are uniquely determined. We need the following two lemmas. Let  $\tilde{\boldsymbol{\gamma}} = (\nabla \psi_{i,T}(F_\perp) \cdot \tilde{\mathbf{n}})_{i \in \mathcal{I}^-}$ .

LEMMA 5.1 For all the interface elements, we have  $\tilde{\boldsymbol{\gamma}}^T \boldsymbol{\delta} \in [0, 1]$ . And for the linear and bilinear  $\psi_{i,T}, i \in \mathcal{I}$ , there holds

$$|L(M_i)| < 2h \sqrt{\tilde{\boldsymbol{\gamma}}^T \boldsymbol{\delta}} \quad \forall i \in \mathcal{I}^-. \tag{5.17}$$

Furthermore, for the bilinear  $\psi_{i,T}, i \in \mathcal{I}$ , if  $F$  is chosen to be such that  $F_\perp$  is the midpoint of the line  $\overline{DE}$ , then

$$|L(M_i)| < 2h \tilde{\boldsymbol{\gamma}}^T \boldsymbol{\delta} \quad \forall i \in \mathcal{I}^-. \tag{5.18}$$

*Proof.* We only give the proof for the case in which  $\Pi_T$  is the rotated- $Q_1$  polynomial space, the interface element  $T$  is such that  $\mathcal{I}^- = \{1\}$  and  $\mathcal{I}^+ = \{2, 3, 4\}$  with  $D = (hd, 0)$  and  $E = (0, he)$  for some  $d \in [1/2, 1]$  and  $e \in [0, 1/2]$ . Similar arguments apply to all other cases. First,  $\tilde{\mathbf{n}} = (e, d)/\sqrt{d^2 + e^2}$ . Hence,  $F_\perp = (td, e(h-t))$  for some  $t \in (0, h)$ . By direct calculation, we have

$$\tilde{\boldsymbol{\gamma}}^T \boldsymbol{\delta} = \frac{1}{h(d^2 + e^2)} [h(e - 2d) - 2(2t - h)de] \left( \frac{1}{2} - d \right) e. \tag{5.19}$$

Note that

$$\begin{aligned}\bar{\boldsymbol{y}}^T \boldsymbol{\delta} &= \frac{e - 2d + 2de}{d^2 + e^2} \left( \frac{1}{2} - d \right) e \in [0, 1] \text{ if } t = 0 \text{ and} \\ \bar{\boldsymbol{y}}^T \boldsymbol{\delta} &= \frac{e - 2d - 2de}{d^2 + e^2} \left( \frac{1}{2} - d \right) e \in [0, 1] \text{ if } t = h,\end{aligned}$$

which leads to  $\bar{\boldsymbol{y}}^T \boldsymbol{\delta} \in [0, 1]$  because  $\bar{\boldsymbol{y}}^T \boldsymbol{\delta}$  is a linear function of  $t$  according to (5.19). And (5.17) and (5.18) follow from similar calculation.  $\square$

LEMMA 5.2 For  $h$  small enough, we have

$$1 + \mu \boldsymbol{y}^T \boldsymbol{\delta} \geq 1 - 4\mu\sqrt{\bar{\kappa}}. \quad (5.20)$$

And for the linear and Crouzeix–Raviart  $\psi_{i,T}, i \in \mathcal{I}$ , if  $F$  is chosen such that  $\mathbf{n}(F) = \bar{\mathbf{n}}$ , then

$$1 + \mu \boldsymbol{y}^T \boldsymbol{\delta} = 1 + \mu \bar{\boldsymbol{y}}^T \boldsymbol{\delta} \geq 1. \quad (5.21)$$

In addition, for the bilinear  $\psi_{i,T}, i \in \mathcal{I}$ , if  $F$  is chosen such that  $F_\perp$  is the midpoint of  $\overline{DE}$ , then

$$1 + \mu \boldsymbol{y}^T \boldsymbol{\delta} \geq 1 + \mu \bar{\boldsymbol{y}}^T \boldsymbol{\delta} (1 - 4\sqrt{2}\sqrt{\bar{\kappa}}). \quad (5.22)$$

*Proof.* By Lemma 3.2, (2.3) and Remark 3.3, we have

$$\begin{aligned}& |\nabla \psi_{i,T}^P(F) \cdot \mathbf{n}(F) - \nabla \psi_{i,T}^P(F_\perp) \cdot \bar{\mathbf{n}}| \\ & \leq |(\nabla \psi_{i,T}^P(F) - \nabla \psi_{i,T}^P(F_\perp)) \cdot \mathbf{n}(F)| + |\nabla \psi_{i,T}^P(F_\perp) \cdot (\mathbf{n}(F) - \bar{\mathbf{n}})| \\ & \leq \|(\nabla \psi_{i,T}^P(F) - \nabla \psi_{i,T}^P(F_\perp))\| \cdot \|\mathbf{n}(F)\| + \|\nabla \psi_{i,T}^P(F_\perp)\| \cdot \|(\mathbf{n}(F) - \bar{\mathbf{n}})\| \\ & \leq 4(1 + (1 - 2\epsilon^2)^{-3/2})\kappa, \quad i \in \mathcal{I},\end{aligned}$$

which implies  $\|\boldsymbol{y} - \bar{\boldsymbol{y}}\| \leq 2(1 + 2(1 - 2\epsilon^2)^{-3/2})\kappa$ . For all the types of finite elements considered in this article, by  $\|L\|_{\infty,T} \leq \sqrt{2}h$ , we have  $\|\boldsymbol{\delta}\| \leq \sqrt{2}h$  and therefore,

$$\mu \boldsymbol{y}^T \boldsymbol{\delta} = \mu \bar{\boldsymbol{y}}^T \boldsymbol{\delta} + \mu(\boldsymbol{y}^T - \bar{\boldsymbol{y}}^T)\boldsymbol{\delta} \geq \mu \bar{\boldsymbol{y}}^T \boldsymbol{\delta} - 4\sqrt{2}(1 + (1 - 2\epsilon^2)^{-3/2})\kappa\mu h,$$

which yields (5.20) by the Assumption (3.6). Furthermore, for linear finite elements, if  $F$  is chosen such that  $\mathbf{n}(F) = \bar{\mathbf{n}}$ , then  $\bar{\boldsymbol{y}} = \boldsymbol{y}$ ; thus Lemma 5.1 and (5.11) imply  $1 + \mu \bar{\boldsymbol{y}}^T \boldsymbol{\delta} \geq 1$ . For the bilinear finite elements, if  $F$  is chosen such that  $F_\perp$  is the midpoint of  $\overline{DE}$ , then by (5.18), we have

$$\begin{aligned}\mu \boldsymbol{y}^T \boldsymbol{\delta} &= \mu \bar{\boldsymbol{y}}^T \boldsymbol{\delta} + \mu(\boldsymbol{y}^T - \bar{\boldsymbol{y}}^T)\boldsymbol{\delta} \geq \mu \bar{\boldsymbol{y}}^T \boldsymbol{\delta} - \mu 4(1 + (1 - 2\epsilon^2)^{-3/2})\kappa(2h\bar{\boldsymbol{y}}^T \boldsymbol{\delta}) \\ &\geq \mu \bar{\boldsymbol{y}}^T \boldsymbol{\delta} (1 - 8(1 + (1 - 2\epsilon^2)^{-3/2})\kappa h),\end{aligned}$$

which leads to (5.22) by the Assumption (3.6).  $\square$

**THEOREM 5.3 (Unisolvence)** Let  $\mathcal{T}_h$  be a mesh satisfying (3.6) with  $\bar{\kappa}$  specified therein for linear and Crouzeix–Raviart  $\psi_{i,T}, i \in \mathcal{I}$  and

$$\sqrt{\bar{\kappa}} < \frac{1}{4\sqrt{2}}, \text{ for bilinear } \psi_{i,T}, i \in \mathcal{I}, \tag{5.23}$$

and for some  $\lambda \in (0, 1)$

$$\sqrt{\bar{\kappa}} \leq \frac{\rho(1 - \lambda)}{4 - (3 + \lambda)\rho}, \text{ for rotated-}Q_1 \psi_{i,T}, i \in \mathcal{I}. \tag{5.24}$$

In addition, we assume  $F$  and  $F_\perp$  are chosen such that estimates given by Lemma 5.2 hold. Then, given any vector  $v = (v_1, v_2, v_3, v_4) \in \mathbb{R}^4$  for the bilinear and rotated- $Q_1$  case (or  $v = (v_1, v_2, v_3) \in \mathbb{R}^3$  for the linear and Crouzeix–Raviart case), there exists one and only one IFE function  $\phi_T^p$  in the form of (5.4) satisfying (5.5)–(5.7).

*Proof.* For linear and Crouzeix–Raviart  $\psi_{i,T}, i \in \mathcal{I}$ , by (5.21) in Lemma 5.2, we have  $1 + \mu\boldsymbol{\gamma}^T\boldsymbol{\delta} \neq 0$ . For bilinear  $\psi_{i,T}, i \in \mathcal{I}$ , by (5.23), we have  $1 - 4\sqrt{2}\sqrt{\bar{\kappa}} > 0$ , which leads to  $1 + \mu\boldsymbol{\gamma}^T\boldsymbol{\delta} \neq 0$  because of (5.22) in Lemma 5.2. Similarly, for rotated- $Q_1 \psi_{i,T}, i \in \mathcal{I}$ , by (5.24), we have

$$1 - 4\mu\sqrt{\bar{\kappa}} \geq 1 - \frac{4\sqrt{\bar{\kappa}}}{1 - \bar{\kappa}} \left( \frac{1}{\rho} - 1 \right) \geq \lambda, \tag{5.25}$$

which, by (5.20) in Lemma 5.2, leads to  $1 + \mu\boldsymbol{\gamma}^T\boldsymbol{\delta} \neq 0$  again. Hence, by the well-known results about the Sherman–Morrison formula, the matrix in the linear system (5.14) is nonsingular, which together with (5.12) lead to the existence and uniqueness for coefficients  $c_i, i \in \mathcal{I}^-$  and  $c_0$  of  $\phi_T^p(X)$ .  $\square$

**REMARK 5.4** Theorem 5.3 provides guidelines on the choice for the angle allowance parameter  $\bar{\kappa}$  needed in (3.6) for bilinear and rotated- $Q_1 \psi_{i,T}, i \in \mathcal{I}$ . In the bilinear case, condition (5.23) suggests an upper bound for  $\bar{\kappa}$  which is nevertheless independent of  $\rho$ . In the rotated- $Q_1$  case, condition (5.24) leads to the following upper bound for  $\bar{\kappa}$ :  $\sqrt{\bar{\kappa}} < \frac{\rho}{4-3\rho}$  which depends on  $\rho$ , and this restriction on  $\bar{\kappa}$  becomes more severe when  $\rho$  approaches 0.

**REMARK 5.5** Under the conditions given in Theorem 5.3, we can apply the Sherman–Morrison formula to express the solution to (5.14) explicitly as

$$\mathbf{c} = \mathbf{b} - \mu \frac{(\boldsymbol{\gamma}^T \mathbf{b}) \boldsymbol{\delta}}{1 + \mu \boldsymbol{\gamma}^T \boldsymbol{\delta}} \tag{5.26}$$

which facilitates both analysis and computation for these IFE spaces.

On each interface element  $T$ , Theorem 5.3 guarantees the existence and uniqueness of the IFE shape functions  $\phi_{i,T}^p, i \in \mathcal{I}$  satisfying (5.4)–(5.6) such that

$$\phi_{i,T}^p(M_j) = \delta_{ij}, \quad i, j \in \mathcal{I}, \tag{5.27}$$



where  $\delta_{ij}$  is the Kronecker delta function. Therefore, we can define the local IFE space on each interface element  $T \in \mathcal{T}_h^i$  as

$$S_h^p(T) = \text{Span}\{\phi_{i,T}^p : i \in \mathcal{I}\}. \tag{5.28}$$

### 5.2 Properties of the IFE shape functions

In this section, we present some fundamental properties of IFE shape functions.

**THEOREM 5.6 (Bounds of IFE shape functions)** Under the conditions given in Theorem 5.3, we have the following estimates:

- For rotated- $Q_1$  and Crouzeix–Raviart  $\phi_{i,T}^p, i \in \mathcal{I} \forall T \in \mathcal{T}_h^i$ ,

$$|\phi_{i,T}^p|_{k,\infty,T} \leq \frac{C}{\rho} h^{-k}, k = 0, 1, \tag{5.29}$$

where  $C$  depends also on  $\lambda$  for rotated- $Q_1$  case.

- For linear  $\phi_{i,T}^p, i \in \mathcal{I} \forall T \in \mathcal{T}_h^i$ ,

$$|\phi_{i,T}^p|_{k,\infty,T^+} \leq \frac{C}{\sqrt{\rho}} h^{-k}, \quad |\phi_{i,T}^p|_{k,\infty,T^-} \leq \frac{C}{\rho} h^{-k}, \quad k = 0, 1. \tag{5.30}$$

- For bilinear  $\phi_{i,T}^p, i \in \mathcal{I} \forall T \in \mathcal{T}_h^i$ ,

$$|\phi_{i,T}^p|_{k,\infty,T^+} \leq Ch^{-k}, \quad |\phi_{i,T}^p|_{k,\infty,T^-} \leq \frac{C}{\rho} h^{-k}, \quad k = 0, 1. \tag{5.31}$$

*Proof.* For convenience, we let  $\mathbf{e} = (e_j)_{j \in \mathcal{I}}$  be the unit vector constructing the basis functions  $\phi_{i,T}^p$ , which could be  $(1, 0 \dots, 0), (0, 1 \dots, 0), \dots, (0, 0 \dots, 1)$  and  $\mathbf{e}^- = (e_j)_{j \in \mathcal{I}^-}$ . Let  $w = \sum_{j \in \mathcal{I}^+} \nabla \psi_{j,T}^p(F) \cdot \mathbf{n}(F) e_j$ . Then, (5.16) implies  $\mathbf{b} = \mathbf{e}^- - \mu w \delta$  and plugging it into the Sherman–Morrison formula (5.26) leads to

$$\mathbf{c} = \mathbf{e}^- - \mu \frac{(w + \boldsymbol{\gamma}^T \mathbf{e}^-) \delta}{1 + \mu \boldsymbol{\gamma}^T \delta} \tag{5.32}$$

and plugging (5.32) into (5.12) yields

$$c_0 = \frac{\mu(w + \boldsymbol{\gamma}^T \mathbf{e}^-)}{1 + \mu \boldsymbol{\gamma}^T \delta}. \tag{5.33}$$

Since  $\|\nabla \psi_{i,T}^p\|_{\infty,T} \leq Ch^{-1}, i \in \mathcal{I}, \|L\|_{\infty,T} < Ch, \|\nabla L\|_{\infty,T} < C$  for some constants  $C$  independent of the location of the interface and  $\beta^\pm$ , we have  $\|\boldsymbol{\gamma}\| \leq Ch^{-1}, \|\delta\| \leq Ch, \|\mathbf{b}\| \leq C$  and  $|w| \leq Ch^{-1}$ .

When  $\phi_{i,T}^p, i \in \mathcal{I}$  are rotated- $Q_1$  polynomials, we can apply (5.11) and (5.24) to (5.32) to obtain:

$$\|\mathbf{c}\| \leq \|\mathbf{e}^-\| + \frac{1}{1 - \bar{k}} \left( \frac{1}{\rho} - 1 \right) \frac{(|w| + \|\boldsymbol{\gamma}\| \|\mathbf{e}^-\|) \|\delta\|}{1 - \frac{4\sqrt{\bar{k}}}{1-\bar{k}} \left( \frac{1}{\rho} - 1 \right)} \leq \frac{C}{\rho}. \tag{5.34}$$

Applying similar arguments to (5.33), we have  $\|c_0\| < C/\rho$ . Constant  $C$  in these inequalities depends on  $\lambda$ . Finally, (5.29) follows from applying these bounds for  $\mathbf{c}$  and  $\mathbf{c}_0$  and the bounds for standard finite-element basis functions  $\psi_{i,T}$  to the formula of  $\phi_{i,T}^P$  given in (5.8). When  $\phi_{i,T}^P, i \in \mathcal{I}$  are Crouzeix–Raviart polynomials, we apply (5.11) and (5.21) to (5.32) and (5.33) to obtain:

$$\begin{aligned} \|\mathbf{c}\| &\leq \|\mathbf{e}^-\| + \frac{1}{1-\bar{\kappa}} \left(\frac{1}{\rho} - 1\right) (|w| + \|\boldsymbol{\gamma}\| \|\mathbf{e}^-\|) \|\boldsymbol{\delta}\| \leq \frac{C}{\rho}, \\ |c_0| &\leq \frac{1}{1-\bar{\kappa}} \left(\frac{1}{\rho} - 1\right) (|w| + \|\boldsymbol{\gamma}\| \|\mathbf{e}^-\|) \|\boldsymbol{\delta}\| \leq \frac{C}{\rho}. \end{aligned} \tag{5.35}$$

Then, (5.29) in this case follows from the same arguments used above for the rotated- $Q_1$  case.

When  $\phi_{i,T}^P, i \in \mathcal{I}$  are linear polynomials, using (5.17) and Lemma 5.2 in (5.32), we have

$$\|\mathbf{c}\| \leq C + C \frac{\mu\sqrt{\boldsymbol{\gamma}^T \boldsymbol{\delta}}}{1 + \mu\boldsymbol{\gamma}^T \boldsymbol{\delta}} \leq C + C \max(1, \sqrt{\mu}) \leq \frac{C}{\sqrt{\rho}}. \tag{5.36}$$

Also, estimate in (5.35) is valid. Then, estimates in (5.30) follow from applying (5.36) and (5.35) to (5.8).

Finally, when  $\phi_{i,T}^P, i \in \mathcal{I}$  are bilinear polynomials, we can apply (5.18) and Lemma 5.2 in (5.32) to have

$$\|\mathbf{c}\| < C + C \frac{\mu\bar{\boldsymbol{\gamma}}^T \boldsymbol{\delta}}{1 + (1 - 4\sqrt{2}\sqrt{\bar{\kappa}})\mu\bar{\boldsymbol{\gamma}}^T \boldsymbol{\delta}} < C, \tag{5.37}$$

because  $1 - 4\sqrt{2}\sqrt{\bar{\kappa}}$  is a positive constant by the condition (5.23). Also, the estimate for  $c_0$  is similar to (5.35). Then, (5.31) follows from applying these bounds for  $\mathbf{c}$  and  $c_0$  to (5.8). □

LEMMA 5.7 (Partition of Unity) On every interface element  $T \in \mathcal{T}_h^i$ , we have

$$\sum_{i \in \mathcal{I}} \phi_{i,T}^P(X) = 1 \quad \forall X \in T, \tag{5.38}$$

$$\nabla \left( \sum_{i \in \mathcal{I}} \phi_{i,T}^P(X) \right) = \sum_{i \in \mathcal{I}} \nabla \phi_{i,T}^P(X) = 0 \quad \forall X \in T. \tag{5.39}$$

*Proof.* Let  $p(X) = \sum_{i \in \mathcal{I}} \phi_{i,T}^P(X)$  and  $q(X) = 1$ . Obviously both  $p(X)$  and  $q(X)$  are in the format of (5.4) and they satisfy (5.5) and (5.6). Furthermore, it is easy to verify that  $p(M_i) = 1 = q(M_i), i \in \mathcal{I}$ . Hence,

$$\sum_{i \in \mathcal{I}} \phi_{i,T}^P(X) = p(X) = q(X) = 1$$

according to the unisolvence stated in Theorem 5.3. Property (5.39) follows from (5.38) directly. □

Now, on every  $T \in \mathcal{T}_h^i$ , choosing arbitrary points  $\bar{X}_i \in l, i \in \mathcal{I}$ , we can construct two vector functions as follows:

$$\mathbf{A}_1(X) = \sum_{i \in \mathcal{I}} (M_i - X)\phi_{i,T}^{p,+}(X) + \sum_{i \in \mathcal{I}^-} (\bar{M}^+(F) - I)^T (M_i - \bar{X}_i)\phi_{i,T}^{p,+}(X), \quad \text{if } X \in T^+, \quad (5.40a)$$

$$\mathbf{A}_2(X) = \sum_{i \in \mathcal{I}} (M_i - X)\phi_{i,T}^{p,-}(X) + \sum_{i \in \mathcal{I}^+} (\bar{M}^-(F) - I)^T (M_i - \bar{X}_i)\phi_{i,T}^{p,-}(X), \quad \text{if } X \in T^-. \quad (5.40b)$$

It follows from the Lemma 3.6 that the two functions in (5.40) are independent of the location of  $\bar{X}_i \in l, i \in \mathcal{I}$ . Furthermore, from the partition of unity stated in Lemma 5.7, we have

$$\mathbf{A}_1(X) = \sum_{i \in \mathcal{I}} M_i \phi_{i,T}^{p,+}(X) - X + \sum_{i \in \mathcal{I}^-} (\bar{M}^+(F) - I)^T (M_i - \bar{X}_i)\phi_{i,T}^{p,+}(X), \quad (5.41a)$$

$$\mathbf{A}_2(X) = \sum_{i \in \mathcal{I}} M_i \phi_{i,T}^{p,-}(X) - X + \sum_{i \in \mathcal{I}^+} (\bar{M}^-(F) - I)^T (M_i - \bar{X}_i)\phi_{i,T}^{p,-}(X), \quad (5.41b)$$

which imply that each component of  $\mathbf{A}_1(X)$  and  $\mathbf{A}_2(X)$  is a polynomial in  $\Pi_T$  because  $\phi_{i,T}^{p,s}(X) \in \Pi_T, s = \pm$ , for  $i \in \mathcal{I}$ . We consider two auxiliary vector functions

$$\mathbf{A}^+(X) = \mathbf{A}_1(X), \quad \text{and} \quad \mathbf{A}^-(X) = (\bar{M}^+(F))^T \mathbf{A}_2(X). \quad (5.42)$$

Let  $d(\mathbf{A}^s), s = \pm$  be the vector of the coefficients of the second-degree term in each component of  $\mathbf{A}^s$ .

LEMMA 5.8  $\mathbf{A}^+$  and  $\mathbf{A}^-$  are such that  $d(\mathbf{A}^+) = d(\mathbf{A}^-)$ .

*Proof.* Let  $d(\phi_{i,T}^{p,+}) = d(\phi_{i,T}^{p,-}) = d_i, i \in \mathcal{I}$ . By the partition of unity,  $\sum_{i \in \mathcal{I}} d_i = 0$ . By using (3.14) given in Lemma 3.5 and using Lemma 3.6 to interchange  $\bar{X}_i, i \in \mathcal{I}$  with a fixed  $\bar{X} \in l$  and, we have

$$\begin{aligned} d(\mathbf{A}^-) &= \sum_{i \in \mathcal{I}} (\bar{M}^+(F))^T M_i d_i + \sum_{i \in \mathcal{I}^+} (I - \bar{M}^+(F))^T (M_i - \bar{X}) d_i \\ &= \sum_{i \in \mathcal{I}^-} (\bar{M}^+(F))^T M_i d_i + \sum_{i \in \mathcal{I}^+} M_i d_i - (I - \bar{M}^+(F))^T \bar{X} \sum_{i \in \mathcal{I}^+} d_i \\ &= \sum_{i \in \mathcal{I}} M_i d_i + \sum_{i \in \mathcal{I}^-} (\bar{M}^+(F) - I)^T M_i d_i + (I - \bar{M}^+(F))^T \bar{X} \sum_{i \in \mathcal{I}^-} d_i \\ &= \sum_{i \in \mathcal{I}} M_i d_i + \sum_{i \in \mathcal{I}^-} (\bar{M}^+(F) - I)^T (M_i - \bar{X}_i) d_i, \end{aligned}$$

which is exactly  $d(\mathbf{A}^+)$ . □

LEMMA 5.9  $\mathbf{A}^+$  and  $\mathbf{A}^-$  satisfy the condition (5.5), i.e.,  $\mathbf{A}^+(X)|_l = \mathbf{A}^-(X)|_l$ .

*Proof.* Since  $\Lambda^s(X), s = \pm$  are independent of points  $\bar{X}_i \in l, i \in \mathcal{I}$ , we can replace these points by an arbitrary  $\bar{X} \in l$  without changing values of  $\Lambda^s(X), s = \pm$ . Then, since  $\phi_{i,T}^p$  satisfies (5.5), applying (3.14), we have

$$\begin{aligned} \Lambda^+(\bar{X}) &= \sum_{i \in \mathcal{I}} (M_i - \bar{X}) \phi_{i,T}^{p-}(\bar{X}) + \sum_{i \in \mathcal{I}^-} (\bar{M}^+(F) - I)^T (M_i - \bar{X}) \phi_{i,T}^{p-}(\bar{X}) \\ &= \sum_{i \in \mathcal{I}^+} (M_i - \bar{X}) \phi_{i,T}^{p-}(\bar{X}) + (\bar{M}^+(F))^T \sum_{i \in \mathcal{I}^-} (M_i - \bar{X}) \phi_{i,T}^{p-}(\bar{X}) \\ &= (\bar{M}^+(F))^T \left( (\bar{M}^-(F))^T \sum_{i \in \mathcal{I}^+} (M_i - \bar{X}) \phi_{i,T}^{p-}(\bar{X}) + \sum_{i \in \mathcal{I}^-} (M_i - \bar{X}) \phi_{i,T}^{p-}(\bar{X}) \right) \\ &= (\bar{M}^+(F))^T \left( \sum_{i \in \mathcal{I}} (M_i - \bar{X}) \phi_{i,T}^{p-}(\bar{X}) + (\bar{M}^-(F) - I)^T \sum_{i \in \mathcal{I}^+} (M_i - \bar{X}) \phi_{i,T}^{p-}(\bar{X}) \right), \end{aligned}$$

which is exactly  $\Lambda^-(\bar{X})$ . □

LEMMA 5.10  $\Lambda^+$  and  $\Lambda^-$  satisfy the condition (5.6), i.e.,  $\beta^+ \nabla \Lambda^+(F) \cdot \mathbf{n}(F) = \beta^- \nabla \Lambda^-(F) \cdot \mathbf{n}(F)$ , where the gradient operator is understood as the gradient on each component.

*Proof.* Again, Lemma 3.6 allows us to exchange  $\bar{X}_i, i \in \mathcal{I}$  for an arbitrary point  $\bar{X} \in l$  in the discussion below. By (5.6), (5.39), (3.14) and (3.15), we have

$$\begin{aligned} \beta^+ \nabla \Lambda^+(F) \cdot \mathbf{n}(F) &= \sum_{i \in \mathcal{I}} M_i \beta^- \nabla \phi_{i,T}^{p-}(F) \cdot \mathbf{n}(F) + (\bar{M}^+(F) - I)^T \sum_{i \in \mathcal{I}^-} (M_i - \bar{X}) \beta^- \nabla \phi_{i,T}^{p-}(F) \cdot \mathbf{n}(F) - \beta^+ \mathbf{n}(F) \\ &= \sum_{i \in \mathcal{I}^+} M_i \beta^- \nabla \phi_{i,T}^{p-}(F) \cdot \mathbf{n}(F) + \sum_{i \in \mathcal{I}^-} \bar{X} \beta^- \nabla \phi_{i,T}^{p-}(F) \cdot \mathbf{n}(F) \\ &\quad + (\bar{M}^+(F))^T \sum_{i \in \mathcal{I}^-} (M_i - \bar{X}) \beta^- \nabla \phi_{i,T}^{p-}(F) \cdot \mathbf{n}(F) - \beta^+ \mathbf{n}(F) \\ &= \sum_{i \in \mathcal{I}^+} (M_i - \bar{X}) \beta^- \nabla \phi_{i,T}^{p-}(F) \cdot \mathbf{n}(F) + (\bar{M}^+(F))^T \sum_{i \in \mathcal{I}^-} (M_i - \bar{X}) \beta^- \nabla \phi_{i,T}^{p-}(F) \cdot \mathbf{n}(F) - \beta^+ \mathbf{n}(F) \\ &= (\bar{M}^+(F))^T \left( (\bar{M}^-(F))^T \sum_{i \in \mathcal{I}^+} (M_i - \bar{X}) \beta^- \nabla \phi_{i,T}^{p-}(F) \cdot \mathbf{n}(F) \right. \\ &\quad \left. + \sum_{i \in \mathcal{I}^-} (M_i - \bar{X}) \beta^- \nabla \phi_{i,T}^{p-}(F) \cdot \mathbf{n}(F) - \beta^+ (\bar{M}^-(F))^T \cdot \mathbf{n}(F) \right) \end{aligned}$$

$$\begin{aligned}
&= (\overline{M}^+(F))^T \left( (\overline{M}^-(F))^T \sum_{i \in \mathcal{I}^+} (M_i - \overline{X}) \beta^- \nabla \phi_{i,T}^{p,-}(F) \cdot \mathbf{n}(F) \right. \\
&\quad \left. + \sum_{i \in \mathcal{I}^-} (M_i - \overline{X}) \beta^- \nabla \phi_{i,T}^{p,-}(F) \cdot \mathbf{n}(F) - \beta^- \cdot \mathbf{n}(F) \right) = \beta^- \nabla \mathbf{\Lambda}^-(F) \cdot \mathbf{n}(F). \quad \square
\end{aligned}$$

**THEOREM 5.11** On every interface element  $T \in \mathcal{T}_h^i$ , we have

$$\sum_{i \in \mathcal{I}} (M_i - X) \phi_{i,T}^{p,-}(X) + \sum_{i \in \mathcal{I}^+} (\overline{M}^-(F) - I)^T (M_i - \overline{X}_i) \phi_{i,T}^{p,-}(X) = \mathbf{\Lambda}_1(X) = \mathbf{0} \quad \forall X \in T^-, \quad (5.43a)$$

$$\sum_{i \in \mathcal{I}} (M_i - X) \phi_{i,T}^{p,+}(X) + \sum_{i \in \mathcal{I}^-} (\overline{M}^+(F) - I)^T (M_i - \overline{X}_i) \phi_{i,T}^{p,+}(X) = \mathbf{\Lambda}_2(X) = \mathbf{0} \quad \forall X \in T^+, \quad (5.43b)$$

and

$$\sum_{i \in \mathcal{I}} (M_i - X) \partial_d \phi_{i,T}^{p,-}(X) + \sum_{i \in \mathcal{I}^+} \left[ (\overline{M}^-(F) - I)^T (M_i - \overline{X}_i) \partial_s \phi_{i,T}^{p,-}(X) \right] - \mathbf{e}_d = \mathbf{0} \quad \forall X \in T^-, \quad (5.44a)$$

$$\sum_{i \in \mathcal{I}} (M_i - X) \partial_d \phi_{i,T}^{p,+}(X) + \sum_{i \in \mathcal{I}^-} \left[ (\overline{M}^+(F) - I)^T (M_i - \overline{X}_i) \partial_s \phi_{i,T}^{p,+}(X) \right] - \mathbf{e}_d = \mathbf{0} \quad \forall X \in T^+, \quad (5.44b)$$

where  $d = 1, 2$ ,  $\partial_1 = \partial_x$ ,  $\partial_2 = \partial_y$  are partial differential operators, and  $\mathbf{e}_d$ ,  $d = 1, 2$  is the standard  $d$ th unit vector in  $\mathbb{R}^2$ .

*Proof.* We define a piecewise vector polynomial on  $T$  as

$$\mathbf{\Lambda}(X) = \begin{cases} \mathbf{\Lambda}^+(X) & \text{if } X \in T^+, \\ \mathbf{\Lambda}^-(X) & \text{if } X \in T^-. \end{cases}$$

First, the restriction of each component of  $\mathbf{\Lambda}$  to  $T^s$ ,  $s = \pm$  is a polynomial in  $\Pi_T$ . By Lemmas 5.8–5.10, the components of  $\mathbf{\Lambda}$  also satisfy (5.5) and (5.6). In addition, we can easily see that  $\mathbf{\Lambda}(M_i) = \mathbf{0}$ ,  $i \in \mathcal{I}$ . Therefore, by the unisolvence stated in Theorem 5.3, we have  $\mathbf{\Lambda}^+(X) = \mathbf{\Lambda}_1(X) \equiv \mathbf{0}$  and  $\mathbf{\Lambda}^-(X) = (\overline{M}^+(F))^T \mathbf{\Lambda}_2(X) \equiv \mathbf{0}$ . Since  $(\overline{M}^+(F))^T$  is nonsingular, we have  $\mathbf{\Lambda}_2(X) \equiv \mathbf{0}$ . Therefore, (5.43) is proved. The proof for (5.44) can be accomplished by differentiating (5.43) and applying (5.41).  $\square$

### 5.3 Optimal approximation capabilities of IFE spaces

As usual, the local IFE spaces on elements can be employed to define the IFE function space globally on  $\Omega$ . As an example, we consider

$$\begin{aligned}
S_h(\Omega) &= \{v \in L^2(\Omega) : v|_T \in S_h^p(T); \\
&\quad v|_{T_1}(M) = v|_{T_2}(M) \quad \forall M \in \mathcal{N}_h, \forall T_1, T_2 \in \mathcal{T}_h \text{ such that } M \in T_1 \cap T_2\}, \quad (5.45)
\end{aligned}$$

where  $\mathcal{N}_h$  is the set of nodes in the mesh  $\mathcal{T}_h$ , and this implies every IFE function in this space is continuous at every node in the mesh. However, as observed in Lin *et al.* (2015), every  $v \in S_h(\Omega)$  is usually discontinuous across the interface edges. When the interface is a generic curve, the discontinuity of

$v \in S_h(\Omega)$  also occurs along the interface curve because these IFE shape functions are defined according to the actual interface. These features ensure the  $H^1$  regularity of  $v \in S_h(\Omega)$  in the subdomain of  $\Omega$  minus the union of interface elements.

We proceed to show that these IFE spaces formed above by linear, bilinear, CR and rotated- $Q_1$  polynomials have the optimal approximation property from the point of view how well the interpolation of a function  $u$  in these IFE spaces can approximate  $u$ . First, we define the interpolation operator on an element  $T \in \mathcal{T}_h$  as the mapping  $I_{h,T} : C^0(T) \rightarrow S_h^p(T)$  such that

$$I_{h,T}^p u(X) = \begin{cases} \sum_{i \in \mathcal{I}} u(M_i) \psi_{i,T}^p(X), & \text{if } T \in \mathcal{T}_h^n, \\ \sum_{i \in \mathcal{I}} u(M_i) \phi_{i,T}^p(X), & \text{if } T \in \mathcal{T}_h^i. \end{cases} \tag{5.46}$$

Furthermore, the global IFE interpolation  $I_h^p : C^0(\Omega) \rightarrow S_h(\Omega)$  can be defined piecewisely as

$$(I_h^p u)|_T = I_{h,T}^p u \quad \forall T \in \mathcal{T}_h. \tag{5.47}$$

On every noninterface element  $T \in \mathcal{T}_h^n$ , the standard scaling argument (Ciarlet, 1978; Rannacher & Turek, 1992; Brenner & Scott, 1994) yields the following error estimate for the local interpolation  $I_{h,T}^p u$  on  $T$ :

$$\|I_{h,T}^p u - u\|_{0,T} + h|I_{h,T}^p u - u|_{1,T} \leq Ch^2 |u|_{2,T} \quad \forall u \in H^2(T). \tag{5.48}$$

However, how to use the scaling argument to derive an error bound for the interpolation on an interface element is unclear because the local IFE space  $S_h^p(T)$  is interface dependent, and it is not even a subspace of  $H^1(T)$  in general. Instead, we will use the multipoint Taylor expansion method (Xu, 1982; Li *et al.*, 2004; He *et al.*, 2008; He, 2009; Zhang, 2013) to derive estimates for the IFE interpolation error.

**THEOREM 5.12** Let  $T \in \mathcal{T}_h^i$  assume  $u \in PC_{int}^2(T)$ . Then, for any  $\bar{X}_i \in l, i \in \mathcal{I}$ , we have

$$I_{h,T}^p u(X) - u(X) = \sum_{i \in \mathcal{I}^s} (E_i^s + F_i^s) \phi_{i,T}^p(X) + \sum_{i \in \mathcal{I}} R_i^s \phi_{i,T}^p(X) \quad \forall X \in T^s, s = \pm, \tag{5.49a}$$

$$\partial_d (I_{h,T}^p u(X) - u(X)) = \sum_{i \in \mathcal{I}^s} (E_i^s + F_i^s) \partial_d \phi_{i,T}^p(X) + \sum_{i \in \mathcal{I}} R_i^s \partial_d \phi_{i,T}^p(X) \quad \forall X \in T^s, s = \pm, \tag{5.49b}$$

where  $d = 1$  or  $2, R_i^s$  are given by (4.2) and (4.4), and

$$\begin{aligned} E_i^s &= \left( (M^s(\tilde{Y}_i) - \bar{M}^s(F)) \nabla u^s(X) \right) \cdot (M_i - \tilde{Y}_i), \quad s = \pm, i \in \mathcal{I}^s, \\ F_i^s &= - \left( (\bar{M}^s(F) - I) \nabla u^s(X) \right) \cdot (\tilde{Y}_i - \bar{X}_i), \quad s = \pm, i \in \mathcal{I}^s. \end{aligned} \tag{5.50}$$

*Proof.* For  $X \in T^s, s = \pm$ , substituting the expansion (4.1) and (4.3) into the IFE interpolation (5.46), we have

$$\begin{aligned} I_{h,T}^p u(X) &= u^s(X) \sum_{i \in \mathcal{I}} \phi_{i,T}^p(X) + \nabla u^s(X) \cdot \sum_{i \in \mathcal{I}} (M_i - X) \phi_{i,T}^p(X) \\ &+ \sum_{i \in \mathcal{I}^s} \left( (M^s(\tilde{Y}_i) - I) \nabla u^s(X) \right) \cdot (M_i - \tilde{Y}_i) \phi_{i,T}^p(X) + \sum_{i \in \mathcal{I}} R_i^s \phi_{i,T}^p(X), \quad X \in T^s, s = \pm. \end{aligned} \tag{5.51}$$

From Theorem 5.11, we have

$$\sum_{i \in \mathcal{I}} (M_i - X) \phi_{i,T}^P(X) = - \sum_{i \in \mathcal{I}'} (\bar{M}^s(F) - I)^T (M_i - \bar{X}_i) \phi_{i,T}^P(X) \quad \forall X \in T^s, s = \pm. \tag{5.52}$$

Then, applying (5.52) and the partition of unity to (5.51), leads to

$$\begin{aligned} I_{h,T}^P u(X) &= u^s(X) - \sum_{i \in \mathcal{I}'} \left( (\bar{M}^s(F) - I) \nabla u^s(X) \right) \cdot (M_i - \bar{X}_i) \phi_{i,T}^P(X) \\ &+ \sum_{i \in \mathcal{I}'} \left( (M^s(\tilde{Y}_i) - I) \nabla u^s(X) \right) \cdot (M_i - \tilde{Y}_i) \phi_{i,T}^P(X) + \sum_{i \in \mathcal{I}} R_i^s \phi_{i,T}^P(X), \quad X \in T^s, s = \pm \end{aligned} \tag{5.53}$$

from which (5.49a) follows by using  $M_i - \bar{X}_i = (M_i - \tilde{Y}_i) + (\tilde{Y}_i - \bar{X}_i)$ . For (5.49b), applying the expansions (4.1) and (4.3) in  $\partial_d I_{h,T}^P u(X) = \sum_{i \in \mathcal{I}} u(M_i) \partial_d \phi_{i,T}^P(X)$ ,  $d = 1, 2$ , we have

$$\begin{aligned} \partial_d I_{h,T}^P u(X) &= u^s(X) \sum_{i \in \mathcal{I}} \partial_d \phi_{i,T}^P(X) + \nabla u^s(X) \cdot \sum_{i \in \mathcal{I}} (M_i - X) \partial_d \phi_{i,T}^P(X) \\ &+ \sum_{i \in \mathcal{I}'} \left( (M^s(\tilde{Y}_i) - I) \nabla u^s(X) \right) \cdot (M_i - \tilde{Y}_i) \partial_d \phi_{i,T}^P(X) + \sum_{i \in \mathcal{I}} R_i^s \partial_d \phi_{i,T}^P(X) \quad \forall X \in T^s, s = \pm. \end{aligned} \tag{5.54}$$

Then, applying (5.39) and Theorem 5.11 to (5.54) we have

$$\begin{aligned} \partial_d I_{h,T}^P u(X) &= \nabla u^s(X) \cdot \mathbf{e}_d - \sum_{i \in \mathcal{I}'} \left( (\bar{M}^s(F) - I) \nabla u^s(X) \right) \cdot (M_i - \bar{X}_i) \partial_d \phi_{i,T}^P(X) \\ &+ \sum_{i \in \mathcal{I}'} \left( (M^s(\tilde{Y}_i) - I) \nabla u^s(X) \right) \cdot (M_i - \tilde{Y}_i) \partial_d \phi_{i,T}^P(X) + \sum_{i \in \mathcal{I}} R_i^s \partial_d \phi_{i,T}^P(X), \quad \forall X \in T^s, s = \pm \end{aligned}$$

which leads to (5.49b) because  $\nabla u^s(X) \cdot \mathbf{e}_d = \partial_d u^s(X)$  and  $M_i - \bar{X}_i = (M_i - \tilde{Y}_i) + (\tilde{Y}_i - \bar{X}_i)$ . □

By an argument similar to that used in Zhang (2013), we can estimate  $E_i^s$  and  $F_i^s$  in (5.50) by geometric properties established in Section 3.

**LEMMA 5.13** There exist constants  $C > 0$  independent of the interface location and  $\beta^\pm$  such that the following estimates hold for every  $T \in \mathcal{T}_h^i$  and  $u \in PC_{int}^2(T)$ :

$$\|E_i^-\|_{0,T^-} \leq C \kappa h^2 |u|_{1,T^-}, \quad i \in \mathcal{I}^+, \quad \|E_i^+\|_{0,T^+} \leq C \frac{\kappa}{\rho} h^2 |u|_{1,T^+}, \quad i \in \mathcal{I}^-. \tag{5.55a}$$

$$\|F_i^-\|_{0,T^-} \leq C h^2 |u|_{1,T^-}, \quad i \in \mathcal{I}^+, \quad \|F_i^+\|_{0,T^+} \leq \frac{C}{\rho} h^2 |u|_{1,T^+}, \quad i \in \mathcal{I}^-. \tag{5.55b}$$

*Proof.* By  $\|M_i - \tilde{Y}_i\| \leq h$  and Lemma 3.4, we have

$$\|E_i^s\|_{0,T^s} \leq \|(M^s(\tilde{Y}_i) - \bar{M}^s(F))\| \|\nabla u^s\|_{0,T^s} \|M_i - \tilde{Y}_i\| \tag{5.56}$$

which implies the (5.55a) by (3.13) in Lemma 3.4. By Lemma 3.6,  $F_i^s$  is independent of the choice of  $\bar{X}_i \in l$ . Hence, by taking  $\bar{X}_i = \tilde{Y}_{i\perp}$  in  $F_i^s$  and applying Lemma 3.2, we have

$$\|F_i^s\|_{0,T^s} \leq \|\bar{M}^s(F) - I\| \|\nabla u^s\|_{0,T^s} \|\tilde{Y}_i - \tilde{Y}_{i\perp}\|, \tag{5.57}$$

which establishes (5.55b) by (3.12). □

Now we are ready to prove the main result in this section.

**THEOREM 5.14** Assume all the conditions required by Theorem 5.6 hold and  $u \in PH_{int}^2(T)$ . Then, on every  $T \in \mathcal{T}_h^i$  the following hold.

- For rotated- $Q_1$  and Crouzeix–Raviart finite elements,

$$\|I_{h,T}^P u - u\|_{0,T^-} + h|I_{h,T}^P u - u|_{1,T^-} \leq C \frac{1+\kappa}{\rho} h^2 (|u|_{1,T} + |u|_{2,T}), \tag{5.58a}$$

$$\|I_{h,T}^P u - u\|_{0,T^+} + h|I_{h,T}^P u - u|_{1,T^+} \leq C \frac{1+\kappa}{\rho^2} h^2 (|u|_{1,T} + |u|_{2,T}), \tag{5.58b}$$

where  $C$  depends on  $\lambda$  chosen for rotated- $Q_1$  case in (5.24).

- For linear finite elements,

$$\|I_{h,T}^P u - u\|_{0,T^-} + h|I_{h,T}^P u - u|_{1,T^-} \leq C \frac{1+\kappa}{\rho} h^2 (|u|_{1,T} + |u|_{2,T}), \tag{5.59a}$$

$$\|I_{h,T}^P u - u\|_{0,T^+} + h|I_{h,T}^P u - u|_{1,T^+} \leq C \frac{1+\kappa}{\rho^{3/2}} h^2 (|u|_{1,T} + |u|_{2,T}). \tag{5.59b}$$

- For bilinear finite elements,

$$\|I_{h,T}^P u - u\|_{0,T} + h|I_{h,T}^P u - u|_{1,T} \leq C \frac{1+\kappa}{\rho} h^2 (|u|_{1,T} + |u|_{2,T}). \tag{5.60}$$

*Proof.* On each  $T \in \mathcal{T}_h^i$ , by Theorem 5.12, for every  $u \in PC_{int}^2(T)$ , we have

$$|I_{h,T}^P u - u|_{k,T^s} \leq \sum_{i \in \mathcal{I}^s} (\|E_i^s\|_{0,T^s} + \|F_i^s\|_{0,T^s}) |\phi_{i,T}^P|_{k,\infty,T^s} + \sum_{i \in \mathcal{I}} \|R_i^s\|_{0,T^s} |\phi_{i,T}^P|_{k,\infty,T^s}, \tag{5.61}$$

where  $k = 0, 1$ . Then, applying Lemmas 4.2, 4.3, 5.13 and Theorem 5.6 for corresponding IFE functions to the inequality above, we obtain (5.58)–(5.60) for  $u \in PC_{int}^2(T)$ . Finally, the density hypothesis **(H4)** shows that (5.58)–(5.60) also hold for any  $u \in PH_{int}^2(T)$ . □

The local estimate in Theorem 5.14 leads to the following global estimate for the IFE interpolation directly.



**THEOREM 5.15** For any  $u \in PH_{int}^2(\Omega)$ , the following estimate of interpolation error holds

$$\|I_h^p u - u\|_{0,\Omega^s} + h|I_h^p u - u|_{1,\Omega^s} \leq \tilde{C}^s h^2 (|u|_{1,\Omega} + |u|_{2,\Omega}), \quad s = \pm. \tag{5.62}$$

The constants  $\tilde{C}$  depending on  $\kappa$  and  $\rho$  are specified as the following:

- for the rotated- $Q_1$  and Crouzeix–Raviart IFE space,

$$\tilde{C}^- = C \frac{1 + \kappa}{\rho}, \quad \tilde{C}^+ = C \frac{1 + \kappa}{\rho^2}, \tag{5.63}$$

where  $C$  depends on  $\lambda$  for rotated- $Q_1$  case;

- for the linear IFE space,

$$\tilde{C}^- = C \frac{1 + \kappa}{\rho}, \quad \tilde{C}^+ = C \frac{1 + \kappa}{\rho^{3/2}}; \tag{5.64}$$

- for the bilinear IFE space,

$$\tilde{C}^- = \tilde{C}^+ = C \frac{1 + \kappa}{\rho}. \tag{5.65}$$

*Proof.* Estimate (5.62) follows directly from combining the estimates (5.58) to (5.60) and (5.48).  $\square$

### 6. Numerical examples

In this section, we use numerical examples to demonstrate the approximation capability of the IFE spaces by IFE interpolation and IFE solutions. In generating numerical results, all computations involving integrations on interface elements, such as the assemblage of local matrices and vectors with IFE shape functions or assessing the errors with integral norms, are handled by the numerical quadratures based on the transfinite mapping between the reference straight edge triangle/square and the physical curved edge triangles/quadrilaterals. More details about quadrature techniques on curved-edge domains can be found in Kopriva (2009) and Sevilla & Fernández-Méndez (2011).

All the numerical results to be presented are generated in the domain  $\Omega = (-1, 1) \times (-1, 1)$  in which the interface curve  $\Gamma$  is a circle with radius  $r_0 = \pi/6.28$ , which divides  $\Omega$  into two subdomains  $\Omega^-$  and  $\Omega^+$  with

$$\Omega^- = \{(x, y) : x^2 + y^2 < r_0^2\}.$$

The function to be approximated is

$$u(x, y) = \begin{cases} \frac{1}{\beta^-} r^\alpha, & (x, y) \in \Omega^-, \\ \frac{1}{\beta^+} r^\alpha + \left(\frac{1}{\beta^-} - \frac{1}{\beta^+}\right) r_0^\alpha, & (x, y) \in \Omega^+, \end{cases} \tag{6.1}$$

where  $r = \sqrt{x^2 + y^2}$  and  $\alpha = 5$ . It is easy to verify that  $u$  satisfies the interface jump condition (1.3) and (1.4). We note that this is the same interface problem for the numerical examples given in Lin *et al.* (2015). Numerical examples presented here are generated with the bilinear IFE space developed in Section 5,

TABLE 1 *Interpolation errors and rates for the bilinear IFE function,  $\beta^- = 1$  and  $\beta^+ = 10000$* 

$h$	$\ u - I_h u\ _{0,\Omega}$	rate	$ u - I_h u _{1,\Omega}$	rate
1/40	2.7681E-4		1.4482E-2	
1/80	7.2447E-4	1.9339	7.4468E-3	0.9596
1/160	1.8580E-5	1.9632	3.7827E-3	0.9772
1/320	4.7122E-6	1.9793	1.9061E-3	0.9888
1/640	1.1858E-6	1.9906	9.5723E-4	0.9937
1/1280	2.9744E-7	1.9952	4.7965E-4	0.9969

TABLE 2 *Interpolation errors and rates for the bilinear IFE function,  $\beta^- = 10000$  and  $\beta^+ = 1$* 

$h$	$\ u - I_h u\ _{0,\Omega}$	rate	$ u - I_h u _{1,\Omega}$	rate
1/40	9.0663E-3		4.3850E-1	
1/80	2.2680E-3	1.9991	2.1939E-1	0.9991
1/160	5.6711E-4	1.9997	1.0971E-1	0.9998
1/320	1.4179E-4	1.9999	5.4859E-2	0.9999
1/640	3.5447E-5	2.0000	2.7430E-2	1.0000
1/1280	8.8618E-6	2.0000	1.3715E-2	1.0000

and we note that numerical results with other IFE spaces developed in Section 5 are similar, which are therefore not presented to avoid redundancy.

Note that the curvature of the interface in this interface problem is  $\kappa \approx 2$ . Condition (5.23) allows us to use  $\bar{\kappa} \approx 0.031$ . Then, using  $\epsilon \approx 0.4$  in (3.6) leads to a suggested bound for the mesh size  $h \approx 0.0273$ . Therefore, our numerical experiments presented in this section are all on meshes whose sizes are not larger than  $1/40 = 0.025$ , which can sufficiently satisfy the conditions in the error estimation in the previous section.

### 6.1 The convergence of IFE interpolation

Tables 1 and 2 present interpolation error  $u - I_h u$  in both the  $L^2$  and the semi- $H^1$  norms over a sequence of meshes whose mesh size is  $h$ . In these tables, the rate is the estimated values of  $r$  such that  $\|u - I_h u\|_{0,\Omega} \approx Ch^r$  or  $|u - I_h u|_{1,\Omega} \approx Ch^r$  with numerical results generated on two consecutive meshes. The estimated values for  $r$  clearly demonstrate the optimal convergence of  $I_h u$ . We note that this example involves a coefficient  $\beta$  with a jump quite large. Our numerical experiments show that these IFE spaces converge optimally also when  $\beta$  has a moderate jump such as  $\beta^- = 1, \beta^+ = 10$ .

### 6.2 The convergence of the IFE solution

Let  $u_h$  be the IFE solution generated by the bilinear IFE space applied in the partially penalized method in Lin *et al.* (2015) for the interface problem (1.1)–(1.4), where  $f$  and  $g$  are chosen such that  $u$  given by (6.1)

TABLE 3 *Errors in the bilinear IFE solution generated by the symmetric partially penalized IFE method,  $\beta^- = 1$ ,  $\beta^+ = 10000$* 

	SPP IFE		SPP IFE	
$h$	$\ u - u_h\ _{0,\Omega}$	rate	$ u - u_h _{1,\Omega}$	rate
1/40	3.7917E-4		1.5276E-2	
1/80	1.0409E-4	1.8650	7.9599E-3	0.9405
1/160	2.5628E-5	2.0220	3.9096E-3	1.0257
1/320	6.6828E-6	1.9392	1.9501E-3	1.0035
1/640	1.7806E-6	1.9081	9.7745E-4	0.9964
1/1280	4.0278E-7	2.1443	4.8374E-4	1.0148

TABLE 4 *Errors in the bilinear IFE solution generated by the symmetric partially penalized IFE method,  $\beta^- = 10000$ ,  $\beta^+ = 1$* 

	SPP IFE		SPP IFE	
$h$	$\ u - u_h\ _{0,\Omega}$	rate	$ u - u_h _{1,\Omega}$	rate
1/40	1.0734E-2		4.4052E-1	
1/80	2.5715E-3	2.0616	2.1966E-1	1.0040
1/160	6.2918E-4	2.0310	1.0974E-1	1.0012
1/320	1.5709E-4	2.0019	5.4864E-2	1.0001
1/640	4.0137E-5	1.9686	2.7431E-2	1.0000
1/1280	9.8101E-6	2.0326	1.3715E-2	1.0000

is its exact solution. The errors in the bilinear IFE solution generated by the symmetric partially penalized IFE (SPP IFE) method on a sequence of meshes are listed in Tables 3 and 4. The values of numerically estimated rate  $r$  in these tables clearly indicate the optimal convergence of the bilinear IFE solution gauged in either the  $L^2$ -norm or  $H^1$ -norm. We also have carried out extensive numerical experiments by applying the IFE spaces developed in Section 5 to the partially penalized IFE methods in Lin *et al.* (2015) with all the popular penalties, and we have observed similar optimal convergence in the related IFE solution for this interface problem.

### Funding

National Science Foundation (in part) (DMS 1016313).

### REFERENCES

- ADJERID, S., CHAABANE, N. & LIN, T. (2015) An immersed discontinuous finite element method for stokes interface problems. *Comput. Methods Appl. Mech. Engrg.*, **293**, 170–190.
- BABUŠKA, I. (1970) The finite element method for elliptic equations with discontinuous coefficients. *Computing (Arch. Elektron. Rechnen)*, **5**, 207–213.
- BABUŠKA, I., BANERJEE, U. & OSBORN, J. E. (2003) Survey of meshless and generalized finite element methods: a unified approach. *Acta Numer.*, **12**, 1–125.

- BABUŠKA, I. & ZHANG, Z. (1998) The partition of unity method for the elastically supported beam. *Comput. Methods Appl. Mech. Engrg.*, **152**, 1–18 (Symposium on Advances in Computational Mechanics, vol. 5. Austin, TX, 1997).
- BARRETT, J. W. & ELLIOTT, C. M. (1987) Fitted and unfitted finite-element methods for elliptic equations with smooth interfaces. *IMA J. Numer. Anal.*, **7**, 283–300.
- BASTIAN, P. & ENGWER, C. (2009) An unfitted finite element method using discontinuous Galerkin. *Internat. J. Numer. Methods Engrg.*, **79**, 1557–1576.
- BORDAS, S. P. A., RABCZUK, T., RODENAS, J. J., KERFRIDEN, P., MOUMNASSI, M. & BELOUETTAR, S. (2010) Recent advances towards reducing the meshing and re-meshing burden in computational sciences. *Comput. Technol. Rev.*, **2**, 51–82.
- BRAESS, D. (2001) *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*, 2nd edn. Cambridge: Cambridge University Press (Translated from the 1992 German edition by Larry L. Schumaker).
- BRAMBLE, J. H. & THOMAS KING, J. (1996) A finite element method for interface problems in domains with smooth boundaries and interfaces. *Adv. Comput. Math.*, **6**, 109–138.
- BRENNER, S. C. & SCOTT, L. R. (1994) *The Mathematical Theory of Finite Element Methods*. Texts in Applied Mathematics, vol. 15. New York: Springer, 1994.
- BURMAN, E., CLAUS, S., HANSBO, P., LARSON, M. G. & MASSING, A. (2015) Cutfem: Discretizing geometry and partial differential equations. *Internat. J. Numer. Methods Engrg.*, **104**, 472–501.
- BURMAN, E., GUZMÁN, J., SÁNCHEZ, M. A. & SARKIS, M. (2017) Robust flux error estimation of an unfitted nitsche method for high-contrast interface problems. *IMA J. Numer. Math.*, doi: 10.1093/imanum/drx017.
- CHEN, Z. & ZOU, J. (1998) Finite element methods and their convergence for elliptic and parabolic interface problems. *Numer. Math.*, **79**, 175–202.
- CHU, C.-C., GRAHAM, I. G. & HOU, T.-Y. (2010) A new multiscale finite element method for high-contrast elliptic interface problems. *Math. Comp.*, **79**, 1915–1955.
- CIARLET, P. G. (1978) *The Finite Element Method for Elliptic Problems*. Studies in Mathematics and its Applications, vol. 4. Amsterdam-New York-Oxford: North-Holland Publishing Co.
- CLOUGH, R. W. & TOCHER, J. L. (1966) Finite element stiffness matrices for analysis of plate bending. *Matrix Methods in Structural Mechanics* (J. S. Przemieniecki, R. M. Bader, W. F. Bozich, J. R. Johnson & W. J. Mykytow eds), Washington: Air Force Flight Dynamics Laboratory, pp. 515–545.
- CROUZEIX, M. & RAVIART, P.-A. (1973) Conforming and nonconforming finite element methods for solving the stationary Stokes equations. I. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge*, **7**, 33–75.
- DOLBOW, J., MOËS, N. & BELYTSCHKO, T. (2001) An extended finite element method for modeling crack growth with frictional contact. *Comput. Methods Appl. Mech. Engrg.*, **190**, 6825–6846.
- EFENDIEV, Y. & HOU, T. Y. (2009) *Multiscale Finite Element Methods*. Surveys and Tutorials in the Applied Mathematical Sciences, vol. 4. New York: Springer (Theory and applications).
- GONG, Y., LI, B. & LI, Z. (2007/08) Immersed-interface finite-element methods for elliptic interface problems with nonhomogeneous jump conditions. *SIAM J. Numer. Anal.*, **46**, 472–495.
- GUYOMARC'H, G., LEE, C.-O. & JEON, K. (2009) A discontinuous Galerkin method for elliptic interface problems with application to electroporation. *Comm. Numer. Methods Engrg.*, **25**, 991–1008.
- GUZMÁN, J., SÁNCHEZ, M. A. & SARKIS, M. (2015) On the accuracy of finite element approximations to a class of interface problems. *Math. Comp.*, **85**, 2071–2098.
- GUZMÁN, J., SÁNCHEZ, M. A. & SARKIS, M. (2017) A finite element method for high-contrast interface problems with error estimates independent of contrast. *J. Scient. Comp.*, **73**, 330–365.
- HANSBO, A. & HANSBO, P. (2002) An unfitted finite element method, based on Nitsche's method, for elliptic interface problems. *Comput. Methods Appl. Mech. Engrg.*, **191**, 5537–5552.
- HE, X. (2009) Bilinear immersed finite elements for interface problems. *Ph.D. Thesis*, Virginia Polytechnic Institute and State University.
- HE, X., LIN, T. & LIN, Y. (2008) Approximation capability of a bilinear immersed finite element space. *Numer. Methods Partial Differential Equations*, **24**, 1265–1300.

- HE, X., LIN, T. & LIN, Y. (2009) A bilinear immersed finite volume element method for the diffusion equation with discontinuous coefficient. *Commun. Comput. Phys.*, **6**, 185–202.
- HE, X., LIN, T., LIN, Y. & ZHANG, X. (2013) Immersed finite element methods for parabolic equations with moving interface. *Numer. Methods Partial Differential Equations*, **29**, 619–646.
- HOU, S., LI, Z., WANG, L. & WANG, W. (2012) A numerical method for solving elasticity equations with interfaces. *Commun. Comput. Phys.*, **12**, 595–612.
- KAFAY, R., WANG, J. & LIN, T. (2005) A hybrid-grid immersed-finite-element particle-in-cell simulation model of ion optics plasma dynamics. *Dyn. Contin. Discrete Impuls. Syst. Ser. B Appl. Algorithms*, **12**, 1–16.
- KAFAY, R. & WANG, J. (2007) Whole ion optics gridlet simulations using a hybrid-grid immersed-finite-element particle-in-cell code. *J. Propulsion Power*, **23**, 59–68.
- KOPRIVA, D. A. (2009) *Implementing Spectral Methods for Partial Differential Equations: Algorithms for Scientists and Engineers*. Dordrecht: Springer.
- KWAK, D. Y., WEE, K. T. & CHANG, K. S. (2010) An analysis of a broken  $P_1$ -nonconforming finite element method for interface problems. *SIAM J. Numer. Anal.*, **48**, 2117–2134.
- LI, Z. (1998) The immersed interface method using a finite element formulation. *Appl. Numer. Math.*, **27**, 253–267.
- LI, Z., LIN, T., LIN, Y. & ROGERS, R. C. (2004) An immersed finite element space and its approximation capability. *Numer. Methods Partial Differential Equations*, **20**, 338–367.
- LI, Z., LIN, T. & WU, X. (2003) New Cartesian grid methods for interface problems using the finite element formulation. *Numer. Math.*, **96**, 61–98.
- LI, Z. & YANG, X. (2005) An immersed finite element method for elasticity equations with interfaces. *Recent Advances in Adaptive Computation* (Z.-C. Shi, Z. Chen, T. Tang & D. Yu eds), Contemporary Mathematics, vol. 383. Providence, RI: Amer. Math. Soc., pp. 285–298.
- LIAN, H., BORDAS, S. P. A., SEVILLA, R. & SIMPSON, R. N. (2012) Recent developments in cad/analysis integration. *Comput. Techn. Reviews*, **6**, 1–36.
- LIN, T., LIN, Y., ROGERS, R. & LYNNE RYAN, M. (2001) A rectangular immersed finite element space for interface problems. *Scientific Computing and Applications (Kananaskis, AB, 2000)* (P. Mineev & Y. Lin eds), Advances in Computation: Theory And Practice, vol. 7. Huntington, NY: Nova Sci. Publ., pp. 107–114.
- LIN, T., LIN, Y., SUN, W.-W. & WANG, Z. (2011) Immersed finite element methods for 4th order differential equations. *J. Comput. Appl. Math.*, **235**, 3953–3964.
- LIN, T., LIN, Y. & ZHANG, X. (2013a) A method of lines based on immersed finite elements for parabolic moving interface problems. *Adv. Appl. Math. Mech.*, **5**, 548–568.
- LIN, T., SHEEN, D. & ZHANG, X. (2013b) A locking-free immersed finite element method for planar elasticity interface problems. *J. Comput. Phys.*, **247**, 228–247.
- LIN, T., LIN, Y. & ZHANG, X. (2015) Partially penalized immersed finite element methods for elliptic interface problems. *SIAM J. Numer. Anal.*, **53**, 1121–1144.
- LIN, T. & WANG, J. (2002) An immersed finite element electric field solver for ion optics modeling. In *Proceedings of 38th AIAA Joint Propulsion Conference, Indianapolis, IN, July, 2002*, pages AIAA 2002–4263, Reston, VA, 2002. AIAA.
- MELENK, J. M. & BABUŠKA, I. (1996) The partition of unity finite element method: basic theory and applications. *Comput. Methods Appl. Mech. Engrg.*, **139**, 289–314.
- MOON, K. (2016) Immersed discontinuous Galerkin methods for acoustic wave propagation in inhomogeneous media. *Ph. D. Thesis*, Virginia Tech, 2016.
- MOUMNASSI, M., BELOUETTAR, S., BÉCHET, É., BORDAS, S. P. A., QUOIRIN, D. & POTIER-FERRY, M. (2011) Finite element analysis on implicitly defined domains: an accurate representation based on arbitrary parametric surfaces. *Comput. Methods Appl. Mech. Engrg.*, **200**, 774–7965E.
- MOUMNASSI, M., BORDAS, S. P. A., FIGUERO, R. & SANSEN, P. (2014) Analysis using higher-order xfem: implicit representation of geometrical features from a given parametric representation. *Mech. Ind.*, **15**, 443–448.

- NADAL, E., RÓDENAS, J. J., ALBELDA, J., TUR, M., TARANCÓN, J. E. & FUENMAYOR, F. J. (2013) Efficient finite element methodology based on cartesian grids: Application to structural shape optimization. *Abstr. Appl. Anal.*, **2013**, 19 pages.
- NGUYEN, V. P., ANITESCU, C., BORDAS, S. P. A. & RABCZUK, T. (2015) Isogeometric analysis: an overview and computer implementation aspects. *Math. Comput. Simulation*, **117**, 89–116.
- RANNACHER, R. & TUREK, S. (1992) Simple nonconforming quadrilateral Stokes element. *Numer. Methods Partial Differential Equations*, **8**, 97–111.
- SEVILLA, R. & FERNÁNDEZ-MÉNDEZ, S. (2011) Numerical integration over 2d nurbs-shaped domains with applications to nurbs-enhanced fem. *Finite Elem. Anal. Des.*, **47**, 1209–1220.
- SEVILLA, R., FERNÁNDEZ-MÉNDEZ, S. & HUERTA, A. (2008) Nurbs-enhanced finite element method (nefem). *Int. J. Numer. Meth. Eng.*, **76**, 56–83.
- VALLAGHÉ, S. & PAPADOPOULOU, T. (2010) A trilinear immersed finite element method for solving the electroencephalography forward problem. *SIAM J. Sci. Comput.*, **32**, 2379–2394.
- WANG, F., XIAO, Y. & XU, J. (2016) High-order extended finite element methods for solving interface problems. arXiv:1604.06171.
- XU, J. (1982) Estimate of the convergence rate of the finite element solutions to elliptic equation of second order with discontinuous coefficients. *Nat. Sci. J. Xiangtan Univ.*, **1**, 1–5.
- YANG, X., LI, B. & LI, Z. (2003) The immersed interface method for elasticity problems with interfaces. *Dyn. Contin. Discrete Impuls. Syst. Ser. A Math. Anal.*, **10**, 783–808 (Progress in Partial Differential Equations. Pullman, WA, 2002).
- ZHANG, X. (2013) Nonconforming immersed finite element methods for interface problems. *Ph. D. Thesis*, Virginia Polytechnic Institute and State University.