Approximation properties of the Generalized Finite Element Method

C. Anitescu * U. Banerjee †

Abstract

In this paper, we have obtained an approximation result in the Generalized Finite Element Method (GFEM) that reflects the global approximation property of the Partition of Unity (PU) as well as the approximability of the local approximation spaces. We have considered a GFEM, where the underlying PU functions reproduce polynomials of degree l. With the space of polynomials of degree k serving as the local approximation spaces of the GFEM, we have shown, in particular, that the energy norm of the GFEM approximation error of a smooth function is $O(h^{l+k})$. Estimates in the W_p^t-norm have also been established. This result could not be obtained from the classical approximation result of GFEM, which does not reflect the global approximation property of the PU.

Keywords: Generalized finite element method; partition of unity; approximation; quasi-interpolation, error estimates.

1 Introduction

For last 50 years, the classical Finite Element Method (FEM) has been extensively used to approximate solutions of partial differential equations (PDE). From early nineties, there has been a growing interest in modifying the classical FEM, so that the modified methods either do not require a mesh or use a very simple mesh. As as consequence, these methods could be used to address problems with complicated geometry. One of these methods is called the Generalized Finite Element Method (GFEM), which was first introduced in [5] and further developed later in [6], [12].

In GFEM, one starts with a finite open cover $\{\omega_i\}_{i=1}^N$ of the underlying domain Ω of the PDE. The unknown solution of the PDE is accurately approximated locally in ω_i using local approximation spaces V_i, which contains

*Department of Mathematics, 215 Carnegie, Syracuse University, Syracuse, NY 13244. E-mail address: canitesc@syr.edu. This research was partially supported by the NSF Grant # DMS-0610778.

†Department of Mathematics, 215 Carnegie, Syracuse University, Syracuse, NY 13244. E-mail address: banerjee@syr.edu. This research was partially supported by the NSF Grant # DMS-0610778.

polynomials or non-polynomial functions, or both. Then these accurate local approximations, say $\xi_i \in V_i$, of the unknown solution, are "pasted together" using a Partition of Unity (PU), $\{\phi_i\}_{i=1}^N$, subordinate to $\{\omega_i\}_{i=1}^N$, to obtain the GFEM global approximation of the unknown solution, namely $\sum_{i=1}^N \phi_i \xi_i$. The proper choice of local approximation spaces V_i is a crucial aspect of GFEM. Functions in V_i are chosen using *a priori* available information about the unknown solution, such that they mimic the unknown solution locally in ω_i; for examples of V_i, we refer to [3], [6], [12].

It has been shown in the main approximation result of GFEM (see [6], [12], [3]) that the accuracy of the GFEM global approximation of the unknown solution of the PDE depends only on how accurately the functions $\xi_i \in V_i$ approximate the unknown solution locally in ω_i, i.e., the accuracy of the global approximation depends only on the accuracy of the local approximation in V_i. But often the PU functions $\{\phi_i\}_{i=1}^N$ have their own global approximation property, i.e., a linear combination of ϕ_i (which becomes a global function defined on Ω) may approximate a function defined on Ω with a certain accuracy. For example, the standard piecewise linear "hat functions" used in FEM form a PU subordinate to the "finite element stars" associated with a finite element triangulation of Ω, and it is well known that they have good global approximation property. The main approximation result of GFEM does not reflect the global approximation property of the PU functions.

It has been known as a "folklore" in the engineering community that the use of PU functions, with good global approximation property, enhances the global accuracy of the GFEM global approximation. In fact, this feature was stated as a mathematical statement in Theorem 3.6 in [1] without proof. To the best of our knowledge, the proof of this result is not available in the literature. In this paper, we present another approximation result for GFEM, which incorporates the global approximation property of PU functions. We have considered PU functions that "reproduce polynomials of degree l" (see (5)); this property yields good global approximation property of the PU functions $\{\phi_i\}_{i=1}^N$ (see Theorem 2.7). We have considered V_i to be the space of polynomials of degree k. Our approximation result for GFEM involves both l and k and provides an enhanced global accuracy of the GFEM global approximation, provided the approximated function has enough smoothness. This result could not be obtained from the main approximation result of GFEM, as given in [6], [12]. Our result is similar to the result stated in [1]. But we mention that PU functions in [1] satisfied a more general "quasi-reproducing" property; the class of PU functions considered in this paper form a sub-class of the class of "quasi-reproducing" PU functions.

The organization of the paper as follows: In Section 2, we have defined the preliminary notions in GFEM, the main approximation result for GFEM, the notion that the PU functions "reproduce polynomials of degree l", and the associated global approximation result for such PU functions. In Section 3, we presented our main approximation result (Theorem 3.6), which involved the construction of a quasi-interpolant of the approximated function. In Section 4, we presented numerical results illuminating the main result of this paper. We

have also commented on the regularity of the PU functions and the associated accuracy in the GFEM solution. Moreover, we have shown that for certain PU functions, the error could be extremely small. A technical result about the quasi-interpolant, which depends on a combinatorial argument, was presented in Section 5.

2 Preliminaries and Motivation

Let $\mathbf{s} := (s_1, s_2)$ be a multi-index, and in the following we will use the usual multi-index notation. Specifically, we define $|\mathbf{s}| = s_1 + s_2$, $\mathbf{x}^{\mathbf{s}} = x^{s_1}y^{s_2}$ for $\mathbf{x} := (x, y)$, $\mathbf{s}! = s_1!s_2!$, $\mathbf{s} - \mathbf{t} := (s_1 - t_1, s_2 - t_2)$ and for $\mathbf{t} := (t_1, t_2)$ another multi-index, we write $\mathbf{s} \leq \mathbf{t}$ if and only if $s_1 \leq t_1$ and $s_2 \leq t_2$. Also we write

$$D^{(\mathbf{s})}u := \frac{\partial^{|\mathbf{s}|}}{\partial x^{s_1} y^{s_2}}.$$

We will now describe the finite dimensional subspace used in GFEM to approximate solutions of partial differential equations. Let $\Omega \subset \mathbb{R}^2$ and for a given parameter $0 < h < 1$, and integer $N = O(h^{-2})$, we consider the overlapping convex sets ω_j, called *patches*, where $j = 1, 2....N$. We define $d_j :=\text{diam}(\omega_j) \leq 2h$, and require that the patches form an open cover of Ω, i.e.

$$\Omega \subset \tilde{\Omega} := \bigcup_j \omega_j.$$

We assume that the patches ω_j are *quasi-uniform*, specifically that there exists a constant C independent of h such that for all j, $0 < C < d_j/h$. Let

$$\gamma(\omega_i) := \{j : \omega_j \cap \omega_i \neq \emptyset\} \tag{1}$$

and for a given $\mathbf{x} \in \Omega$, we define

$$\gamma(\mathbf{x}) := \{j : \mathbf{x} \in \omega_j\}. \tag{2}$$

We further assume that

$$\max_{\mathbf{x}\in\Omega} \text{card}(\gamma(\mathbf{x})) \leq \max_i \text{card}(\gamma(\omega_i)) \leq \kappa, \tag{3}$$

where κ is independent of h. In other words, we require that each ω_i intersects at most κ of the other patches ω_j.

Associated with the patches $\{\omega_j\}$, let $\{\phi_j\}_{j=1}^N$ be a family of functions defined on Ω, having piecewise continuous first derivatives, satisfying

$$\phi_j(\mathbf{x}) = 0 \text{ for } \mathbf{x} \in \Omega\backslash\omega_j$$

$$\sum_{j=1}^{N} \phi_j(\mathbf{x}) = 1 \tag{4}$$

$$|\phi_j|_{W^s_\infty(\Omega)} \leq Ch^{-s}, s = 0, 1.$$

It is clear from (4) that $\{\phi_j\}$ form a PU on Ω.

Next, for each ω_j, we associate an m_j-dimensional space of functions V_j, defined on ω_j:

$$V_j := \text{span}\{\xi_{sj} : s = 1, ..., m_j\}.$$

The functions ξ_{sj} in V_j are chosen carefully, so that they approximate the unknown solution of the PDE efficiently in the patch ω_j. The only requirement is that V_j must contain constant functions.

We now define the GFEM space as

$$S^{GFEM} := \sum_{j=1}^{N} \phi_j V_j = \text{span}\{\phi_j \xi_{sj}, j = 1, ..., N, s = 1, ..., m_j\}.$$

The main approximation result of S^{GFEM} was derived in [6], [12], and is as follows:

Theorem 2.1. *Let $u \in H^1(\Omega)$. There exists $\xi^u \in S^{GFEM}$ such that*

$$||u - \xi^u||^2_{H^1(\Omega)} \leq C \sum_{j=1}^{N} \inf_{\xi \in V_j} ||u - \xi||^2_{H^1(\omega_j)}.$$

Remark 2.2. If $u \in H^1_0(\Omega)$, then for the index j such that $\omega_j \cap \partial\Omega \neq \emptyset$, the space V_j does not contain constants. Moreover, the functions in such V_j's have to satisfy a local Poincaré inequality, on which we do not elaborate in this paper.

Remark 2.3. It is clear from Theorem 2.1 that the global approximation property of S^{GFEM} is dictated by the local approximation property of the space V_j. In other words, if u could be approximated accurately on ω_j by a function in V_j (i.e. locally), then u could be accurately approximated globally in the domain Ω.

It is important to note that the approximation result of the theorem does not incorporate the global approximation property of the PU, and therefore may not provide precise information on the approximation. This phenomenon can be easily seen when the shape functions are hat functions with respect to a finite element triangulation of Ω, patches are finite element stars, and the local approximation spaces V_j are polynomials of degree k. It is well known that in this situation (see [3]), S^{GFEM} is exactly the finite element space based on piecewise polynomials of degree $k + 1$. From the approximation properties of the finite element space, it is well known ([8], [7]) that if $u \in H^{k+2}(\Omega)$, there exists a $\xi^u \in S^{GFEM}$ such that

$$||u - \xi^u||_{H^1(\Omega)} \leq Ch^{k+1}|u|_{H^{k+2}(\Omega)}.$$

But from Theorem 2.1, we get

$$||u - \xi^u||_{H^1(\Omega)} \leq C \sum_{j=1}^{N} \inf_{\xi \in V_j} ||u - \xi||_{H^1(\omega_j)} \leq Ch^k |u|_{H^{k+1}(\Omega)}.$$

Thus for $u \in H^{k+2}(\Omega)$, the Theorem 2.1 does not give the right order of convergence in this situation, as it does not incorporate the approximation property of the PU hat functions. In this paper, we will address this issue by proving an optimal convergence result that will incorporate the global approximation property of the PU functions that are more general than piecewise polynomials. We will prove this result when V_j's are the space of polynomials of degree k.

We will now discuss PU functions $\{\phi_j\}$ that have global approximation properties.

Definition 2.4. *Let ϕ_j be given functions with support in ω_j (for each j), and suppose the points $\mathbf{x}_j \in \Omega$ are associated with ω_j. We say ϕ_j are* reproducing of order l *(or that they* reproduce polynomials of total degree l*) if:*

$$\sum_{j=1}^{N} \mathbf{x}_j^{\mathbf{p}} \phi_j(\mathbf{x}) = \mathbf{x}^{\mathbf{p}} \text{ for all } |\mathbf{p}| \leq l, \text{ and } \mathbf{x} \in \Omega, \tag{5}$$

where $\mathbf{p} = (p_1, p_2)$ is a multi-index.

Remark 2.5. The points $\mathbf{x}_j$, called particles, defined above are usually inside ω_j but some of the particles could be outside Ω. Using $l = 0$ in (5), it is clear that $\{\phi_j\}$ form a PU. It can be seen that the standard hat functions, defined on a finite element triangulation of Ω, reproduce polynomials of degree 1, where $\mathbf{x}_j$'s are the finite element nodes.

A more general class of functions that reproduce polynomials of higher order are the Reproducing Kernel Particle (RKP) functions, or the Moving Least-Squares (MLS) particle functions, used in meshless methods. These functions are of the form

$$\phi_j(\mathbf{x}) := w_j(\mathbf{x}) \sum_{|\mathbf{t}|=0}^{l} (\mathbf{x} - \mathbf{x}_j)^{\mathbf{t}} b_{\mathbf{t}}(\mathbf{x}), \tag{6}$$

where w_j are given weight functions with support $\bar{\omega}_j$ and $b_{\mathbf{t}}(\mathbf{x})$ are chosen such that (5) is satisfied. We mention that, for each $\mathbf{x} \in \Omega$, $b_{\mathbf{t}}(\mathbf{x})$ is computed as a solution of a linear system. The construction of these functions can be found, for example, in [9], [10], [11]. Moreover, in this paper we will assume that $\{\phi_j(\mathbf{x})\}$ satisfy

$$|\phi_j|_{W^t_\infty(\Omega)} \leq Ch^{-t} \text{ for } t = 0, 1, ..., M. \tag{7}$$

Remark 2.6. M depends on the regularity of the approximated function and we will address this issue later in the paper. It has been shown in [10] (Theorem 4.7) that the condition (7) is satisfied by the RKP functions, provided the generating weight functions w_j are in $C^M(\omega_j)$.

The functions ϕ_j from (6) have the following *global approximation property*:

Theorem 2.7 ([10],[14]). *For a smooth v, there exists a linear combination Φ of $\{\phi_j\}$ such that*

$$||v - \Phi||_{H^t(\Omega)} \leq Ch^{l+1-t}, \quad t = 0, ..., l+1.$$

Next, for each ω_j we define the local approximation space

$$V_j := \mathcal{P}^k(\omega_j) = \text{span}\{(\mathbf{x} - \mathbf{x}_j)^{\mathbf{s}} : \mathbf{x} \in \omega_j,\ |\mathbf{s}| = 0...k\}$$

where $\mathbf{x} := (x, y)$, $\mathbf{s} = (s_1, s_2)$ and $(\mathbf{x} - \mathbf{x}_j)^{\mathbf{s}} := (x - x_j)^{s_1}(y - y_j)^{s_2}$. In other words, $\mathcal{P}^k(\omega_j)$, and therefore V_j, are the spaces of all polynomials of degree k restricted to ω_j. Then, as before, we define the space S^{GFEM} by

$$S^{GFEM} := \text{span}\{\phi_j(\mathbf{x})(\mathbf{x} - \mathbf{x}_j)^{\mathbf{s}} : j = 1, ..., N,\ |\mathbf{s}| = 0...k\}. \tag{8}$$

In the next section, we will show that there exists an "interpolation" operator $I^h : W^l_\infty(\tilde{\Omega}) \to S^h(\Omega)$ which *preserves all polynomials* of degree $k + l$ on Ω, i.e.

$$I^h[p] := I^h[p(\cdot)](\mathbf{x}) = p(\mathbf{x}) \quad \forall p \in \mathcal{P}^{k+l}.$$

3 Main approximation result

For $\mathbf{x} \in \Omega$ we define the interpolation operator $I^h : W^k_\infty(\tilde{\Omega}) \to S^{GFEM}(\Omega)$ by:

$$I^h[v(\cdot)](\mathbf{x}) := \sum_{j=1}^{N} \left[\sum_{|\mathbf{m}|=0}^{k} C_{\mathbf{m}} D^{(\mathbf{m})} v(\mathbf{x}_j)(\mathbf{x} - \mathbf{x}_j)^{\mathbf{m}} \right] \phi_j(\mathbf{x}), \tag{9}$$

where $C_{\mathbf{m}} := \frac{k!(k+l-|\mathbf{m}|)!}{\mathbf{m}!(k-|\mathbf{m}|)!(k+l)!}$ and $\mathbf{m} := (m_1, m_2)$ is also a multi-index. Since the inner term of the summation is a polynomial of degree k, it is clear from the definition of the S^{GFEM} that $I^h[v(\cdot)] \in S^{GFEM}$.

We further note that if $l = 0$ then the inner term in the summation is just the Taylor polynomial of degree k of $v(\mathbf{x})$ centered at $\mathbf{x}_j$. Since $\{\phi_j(\mathbf{x})\}$ form a PU, it can be seen easily that $I^h[v]$ preserves all polynomials of degree k. To derive an approximation property of the space S^{GFEM} that incorporates the approximation property of $\{\phi_j\}$, we will first show that I^h preserves all polynomials of degree $k + l$ on Ω, i.e.

$$I^h[p(\cdot)](\mathbf{x}) = p(\mathbf{x}), \quad \forall p \in \mathcal{P}^{k+l}, \quad \mathbf{x} \in \Omega.$$

Remark 3.1. Even though the interpolant defined in (9) is for a two-dimensional domain, we can define an interpolant using the same formula for the one-dimensional or three-dimensional cases.

In the following lemma, we derive an expression for $I^h[v(\cdot)](\mathbf{x})$ for v of the form $v = \mathbf{x}^{\mathbf{i}}$.

Lemma 3.2. *Let $v(\mathbf{x}) = \mathbf{x}^{\mathbf{i}}$ where $\mathbf{i} := (i_1, i_2)$ is a multi-index with $|\mathbf{i}| = 0, ..., k+l$. Then I^h can be written as*

$$I^h[v(\cdot)](\mathbf{x}) := \sum_{j=1}^{N} \sum_{\substack{0 \le \mathbf{t} \le \mathbf{i} \\ |\mathbf{t}| \le k}} c_{\mathbf{t}} \mathbf{x}^{\mathbf{t}} \mathbf{x}_j^{\mathbf{i}-\mathbf{t}} \phi_j(\mathbf{x}), \tag{10}$$

where

$$c_{\mathbf{t}} := \sum_{\substack{\mathbf{t} \le \mathbf{m} \le \mathbf{i} \\ |\mathbf{m}| \le k}} \frac{k!(k+l-|\mathbf{m}|)!\mathbf{i}!}{(k-|\mathbf{m}|)!(k+l)!(\mathbf{i}-\mathbf{m})!(\mathbf{m}-\mathbf{t})!\mathbf{t}!} (-1)^{|\mathbf{m}-\mathbf{t}|}. \tag{11}$$

Proof. Let $v(\mathbf{x}) = \mathbf{x}^{\mathbf{i}}$, where $\mathbf{i} := (i_1, i_2)$ is a multi-index with $|\mathbf{i}| = 0, ..., k+l$. Then

$$D^{(\mathbf{m})} v(\mathbf{x}_j) = \begin{cases} \frac{\mathbf{i}!}{(\mathbf{i}-\mathbf{m})!} \mathbf{x}_j^{\mathbf{i}-\mathbf{m}} & \text{if } \mathbf{m} \le \mathbf{i}; \\ 0 & \text{otherwise.} \end{cases}$$

Also from the binomial theorem,

$$\begin{aligned} (\mathbf{x} - \mathbf{x}_j)^{\mathbf{m}} &= \left[\sum_{t_1=0}^{m_1} \frac{m_1! x^{t_1} x_j^{m_1-t_1} (-1)^{m_1-t_1}}{(m_1-t_1)!t_1!} \right] \left[\sum_{t_2=0}^{m_2} \frac{m_2! y^{t_2} y_j^{m_2-t_2} (-1)^{m_2-t_2}}{(m_2-t_2)!t_2!} \right] \\ &= \sum_{0 \le \mathbf{t} \le \mathbf{m}} \frac{\mathbf{m}!}{(\mathbf{m}-\mathbf{t})!\mathbf{t}!} \mathbf{x}^{\mathbf{t}} \mathbf{x}_j^{\mathbf{m}-\mathbf{t}} (-1)^{|\mathbf{m}-\mathbf{t}|}, \end{aligned} \tag{12}$$

where the last summation is a double sum over $0 \le t_1 \le m_1$, and $0 \le t_2 \le m_2$. Thus from (9) and (12), we can write the interpolant as:

$$I^h[v(\cdot)](\mathbf{x}) := \sum_{j=1}^{N} f_j(\mathbf{x}) \phi_j(\mathbf{x})$$

where $f_j(\mathbf{x})$ is a polynomial in $\mathbf{x}$ given by

$$\begin{aligned} f_j(\mathbf{x}) &:= \sum_{|\mathbf{m}|=0}^{k} \frac{k!(k+l-|\mathbf{m}|)!}{\mathbf{m}!(k-|\mathbf{m}|)!(k+l)!} D^{(\mathbf{m})} v(\mathbf{x}_j)(\mathbf{x} - \mathbf{x}_j)^{\mathbf{m}} \\ &= \sum_{\substack{0 \le \mathbf{m} \le \mathbf{i} \\ |\mathbf{m}| \le k}} \left(\frac{k!(k+l-|\mathbf{m}|)!}{\mathbf{m}!(k-|\mathbf{m}|)!(k+l)!} \sum_{0 \le \mathbf{t} \le \mathbf{m}} \frac{\mathbf{i}!\mathbf{m}!}{(\mathbf{i}-\mathbf{m})!(\mathbf{m}-\mathbf{t})!\mathbf{t}!} \mathbf{x}^{\mathbf{t}} \mathbf{x}_j^{\mathbf{i}-\mathbf{t}} (-1)^{|\mathbf{m}-\mathbf{t}|} \right) \\ &= \sum_{\substack{0 \le \mathbf{t} \le \mathbf{i} \\ |\mathbf{t}| \le k}} \sum_{\substack{\mathbf{t} \le \mathbf{m} \le \mathbf{i} \\ |\mathbf{m}| \le k}} \frac{k!(k+l-|\mathbf{m}|)!\mathbf{i}!}{(k-|\mathbf{m}|)!(k+l)!(\mathbf{i}-\mathbf{m})!(\mathbf{m}-\mathbf{t})!\mathbf{t}!} (-1)^{|\mathbf{m}-\mathbf{t}|} \mathbf{x}^{\mathbf{t}} \mathbf{x}_j^{\mathbf{i}-\mathbf{t}} \\ &:= \sum_{\substack{0 \le \mathbf{t} \le \mathbf{i} \\ |\mathbf{t}| \le k}} c_{\mathbf{t}} \mathbf{x}^{\mathbf{t}} \mathbf{x}_j^{\mathbf{i}-\mathbf{t}}. \end{aligned}$$

The third equality above was obtained by changing the order of summation between $\mathbf{m}$ and $\mathbf{i}$. This shows that (10) holds with the coefficients $c_{\mathbf{t}}$ of $\mathbf{x}^{\mathbf{t}}\mathbf{x}_j^{\mathbf{i}-\mathbf{t}}$ defined as in (11). □

We will state some properties of the coefficients $\mathbf{c_t}$, namely that for a given $\mathbf{i}$ such that $0 \leq |\mathbf{i}| \leq k+l$,

$$c_{\mathbf{t}} = 0 \text{ if } |\mathbf{t}| < |\mathbf{i}| - l \text{ (i.e. } l < |\mathbf{i}| - |\mathbf{t}| = |\mathbf{i} - \mathbf{t}|) \tag{13}$$
$$\text{and } \sum_{\substack{0 \leq \mathbf{t} \leq \mathbf{i} \\ |\mathbf{t}| \leq k}} c_{\mathbf{t}} = 1.$$

These properties guarantee that terms with $\mathbf{x}^{\mathbf{i}-\mathbf{t}}$, for $l < |\mathbf{i} - \mathbf{t}|$, do not appear in the expansion (10), and that the sum of all coefficients is 1; we need this property to show that I^h preserves all monomials of degree $k+l$. We will illuminate this property in an example after the next lemma. Proving that (13) holds for arbitrary values of k and l requires a (somewhat lengthy) combinatorial argument. We have included a proof of (13) in the Appendix.

We show that I^h preserves all polynomials of degree $k+l$ in the following lemma:

Lemma 3.3. *If $v \in \mathcal{P}^{k+l}$ and (13) holds, then $I^h[v]|_\Omega = v|_\Omega$, i.e. I^h preserves all polynomials of total degree less than or equal to $k+l$ inside Ω.*

Proof. Let $v(\mathbf{x}) = \mathbf{x}^{\mathbf{i}}$. Then from Lemma 3.2,

$$\begin{aligned} I^h[v(\cdot)](\mathbf{x}) &= \sum_{j=1}^{N} f_j(\mathbf{x})\phi_j(\mathbf{x}) \\ &= \sum_{j=1}^{N} \left[\sum_{\substack{0 \leq \mathbf{t} \leq \mathbf{i} \\ |\mathbf{t}| \leq k}} c_{\mathbf{t}} x^{\mathbf{t}} \mathbf{x}_j^{\mathbf{i}-\mathbf{t}} \right] \phi_j(\mathbf{x}), \end{aligned}$$

where $c_{\mathbf{t}}$ is given in (11). Using the properties of $c_{\mathbf{t}}$ in (13), changing the order of summation, and using the fact that $\{\phi_j\}$ reproduce all polynomials of degree up to l, we further have:

$$\begin{aligned} I^h[v(\cdot)](\mathbf{x}) &= \sum_{\substack{0 \leq \mathbf{t} \leq \mathbf{i} \\ |\mathbf{i}| - l \leq |\mathbf{t}| \leq k}} c_{\mathbf{t}} \mathbf{x}^{\mathbf{t}} \sum_{j=1}^{N} \mathbf{x}_j^{\mathbf{i}-\mathbf{t}} \phi_j(\mathbf{x}) \\ &= \sum_{\substack{0 \leq \mathbf{t} \leq \mathbf{i} \\ |\mathbf{t}| \leq k}} c_{\mathbf{t}} \mathbf{x}^{\mathbf{t}} \mathbf{x}^{\mathbf{i}-\mathbf{t}} \\ &= 1 \cdot \mathbf{x}^{\mathbf{i}} = \mathbf{x}^{\mathbf{i}} \end{aligned}$$

Therefore $I^h[v(\cdot)(\mathbf{x})] = v(\mathbf{x})$ for $v(\mathbf{x}) = \mathbf{x}^{\mathbf{i}}, |\mathbf{i}| \leq k+l$. □

We will illustrate the result in Lemma 3.3 through the following example, where we will also comment on the property (13):

Example 3.4. Consider the case when $k = 1$ and $l = 1$. For clarity, we will use the notation (x, y) instead of $\mathbf{x}$ for points in $\Omega \subset \mathbb{R}^2$. Then the interpolant is defined, as in (9), by:

$$I^h[v(\cdot)](x,y) = \sum_{j=1}^{N} \left[v(x_j, y_j) + \frac{1}{2} v_x(x_j, y_j)(x - x_j) + \frac{1}{2} v_y(x_j, y_j)(y - y_j) \right] \phi_j(x, y).$$

We will check that this reproduces all polynomials of degree up to $k + l = 2$.

If $v(x, y) = 1$ ($\mathbf{i} = (0, 0)$) then

$$I^h[v(\cdot)](x,y) = \sum_{j=1}^{N} 1 \cdot \phi_j(x, y) = 1.$$

Here $f_j(x, y) = 1$ and $c_{\mathbf{t}} = c_{(0,0)} = 1$.

If $v(x, y) = x$ ($\mathbf{i} = (1, 0)$) then

$$I^h[v(\cdot)](x,y) = \sum_{j=1}^{N} \left(x_j + \frac{1}{2} \cdot (x - x_j) \right) \phi_j(x, y) = \sum_{j=1}^{N} \left(\frac{1}{2} x + \frac{1}{2} x_j \right) \phi_j(x, y) =$$

$$= \frac{1}{2} x \sum_{j=1}^{N} \phi_j(x, y) + \frac{1}{2} \sum_{j=1}^{N} x_j \phi_j(x, y) = \frac{x}{2} + \frac{x}{2} = x.$$

Here $f_j(x, y) = \frac{1}{2} x + \frac{1}{2} x_j$ and $c_{(0,0)} = \frac{1}{2}$, $c_{(1,0)} = \frac{1}{2}$.

If $v(x, y) = y$ ($\mathbf{i} = (0, 1)$) then similarly

$$I^h[v(\cdot)](x,y) = \sum_{j=1}^{N} \left(y_j + \frac{1}{2} \cdot (y - y_j) \right) \phi_j(x, y) = \frac{y}{2} + \frac{y}{2} = y,$$

with $c_{(0,0)} = \frac{1}{2}$ and $c_{(0,1)} = \frac{1}{2}$. In all the cases so far, it is clear that the property (13) is satisfied.

If $v(x, y) = x^2$ ($\mathbf{i} = (2, 0)$) then

$$I^h[v(\cdot)](x,y) = \sum_{j=1}^{N} \left(x_j^2 + \frac{1}{2} \cdot 2x_j(x - x_j) \right) \phi_j(x, y) = \sum_{j=1}^{N} x x_j \phi_j(x, y) = x^2.$$

Here $c_{(0,0)} = 0, c_{(1,0)} = 1, c_{(2,0)} = 0$, and note that conditions (13) are satisfied, since $\sum c_{\mathbf{t}} = 1$ and $c_{(0,0)} = 0$ as $0 = |\mathbf{t}| < |\mathbf{i}| - l = 1$.

If $v(x, y) = xy$ ($\mathbf{i} = (1, 1)$) then

$$I^h[v(\cdot)](x,y) = \sum_{j=1}^{N} \left(x_j y_j + \frac{1}{2} y_j(x - x_j) + \frac{1}{2} x_j(y - y_j) \right) \phi_j(x, y) =$$

$$= \sum_{j=1}^{N} \left(\frac{1}{2} x y_j + \frac{1}{2} x_j y \right) \phi_j(x, y) = xy,$$

with $c_{(0,0)} = 0, c_{(1,0)} = \frac{1}{2}, c_{(0,1)} = \frac{1}{2}$.

Finally, if $v(x,y) = y^2$, then similarly to the case $v(x,y) = x^2$ we have:

$$I^h[v(\cdot)](x,y) = \sum_{j=1}^{N} \left(y_j^2 + \frac{1}{2} \cdot 2y_j(y - y_j) \right) \phi_j(x,y) = \sum_{j=1}^{N} y y_j \phi_j(x,y) = y^2,$$

and also $c_{(0,0)} = 0, c_{(0,1)} = 1, c_{(0,2)} = 0$. Again, it is clear that the property (13) is satisfied.

We will now prove the following approximation result, where we assume that the PU functions $\{\phi_j\}$ satisfy (7) with $M = k + l + 1$ (see Remark 2.6).

Lemma 3.5. *Let $B^i := B(\mathbf{x}_i, 4h)$, i.e. a ball of radius $4h$ centered at $\mathbf{x_i}$. If $v \in W_\infty^{k+l+1}(B^i)$, then for $t = 0, ..., M$ with $M = k + l + 1$ (as in Remark 2.6), and for any $1 \le i \le N$ we have*

$$||v - I^h[v]||_{W_\infty^t(\omega_i)} \le C h^{k+l+1-t} |v|_{W_\infty^{k+l+1}(B^i)}. \tag{14}$$

Note: Here, i and t are scalars and are not related to the multi-indices $\mathbf{i}$ and $\mathbf{t}$ in the first part of the section.

Proof. The ball B^i is chosen large enough so that $\mathbf{x}_j \in B^i$ for all $j \in \gamma(\omega_i)$, with $\gamma(\omega_i)$ as in (1). Also note that for certain values of i, B^i may contain points outside Ω, in which case it is necessary to extend v to an open ball $B(0,R)$, of radius R, where R is sufficiently large such that

$$\Omega \subset \bigcup_{j=1}^{N} B^j \subset B(0, R - 6h). \tag{15}$$

For the following argument we will assume that there exists an extension $E[v]$ that satisfies

$$\begin{gathered} E[v]|_\Omega = v \text{ and} \\ ||E[v]||_{W_\infty^{k+l+1}(B(0,R))} \le C||v||_{W_\infty^{k+l+1}(\Omega)} \end{gathered} \tag{16}$$

and we will identify $E[v]$ with v. The existence of the extension $E[v]$ is well known provided Ω has a Lipschitz boundary, see [7], [13].

Let Q_i be a $(k+l)$ degree polynomial approximation of v on B^i which satisfies the standard approximation estimate (see for example [7]):

$$||v - Q_i||_{W_\infty^t(\omega_i)} \le ||v - Q_i||_{W_\infty^t(B^i)} \le C h^{k+l+1-t} |v|_{W_\infty^{k+l+1}(B^i)} \tag{17}$$

For example, one could take Q_i to be the Taylor polynomial of v centered at x_i and restricted to B^i. Then:

$$||v - I^h[v]||_{W_\infty^t(\omega_i)} \le ||Q_i - I^h[Q_i]||_{W_\infty^t(\omega_i)} + ||(v - Q_i) - I^h[v - Q_i]||_{W_\infty^t(\omega_i)} \tag{18}$$

Since I^h is invariant over polynomials of degree $k+l$, the first term on the right hand side is zero. Also $v - Q_i$ is estimated by (17), so it remains to estimate $||I^h[v-Q_i]||_{W^t_\infty(B^i)}$. Using the definition of I^h (see (9)), we write

$$I^h[v-Q_i](\mathbf{x}) = \sum_{j=1}^{N} \sum_{|\mathbf{m}|=0}^{k} \left[C_{\mathbf{m}} D^{(\mathbf{m})}(v-Q_i)(\mathbf{x}_j)(\mathbf{x}-\mathbf{x}_j)^{\mathbf{m}} \right] \phi_j(\mathbf{x}), \tag{19}$$

where

$$C_m = \frac{k!(k+l-|\mathbf{m}|)!}{\mathbf{m}!(k-|\mathbf{m}|)!(k+l)!}.$$

Since each $\mathbf{x} \in \Omega$ belongs to at most κ of the sets ω_j, then for a fixed $\mathbf{x}$, $\phi_j(\mathbf{x}) = 0$, for $\mathbf{x} \notin \gamma(\mathbf{x})$. Furthermore, since $\mathbf{x}_j \in B^i$ for all $j \in \gamma(\omega_i)$, from the second inequality of (17) with $t = |\mathbf{m}|$ we have :

$$|D^{(\mathbf{m})}(v-Q_i)(\mathbf{x}_j)| \leq Ch^{k+l+1-|\mathbf{m}|}|v|_{W^{k+l+1}_\infty(B^i)} \text{ for } j \in \gamma(\omega_i). \tag{20}$$

Note that $\gamma(\omega_i)$ and κ were defined in (1) and (3) respectively.

We will now consider several different cases. First, suppose $t = 0$. Then:

$$\begin{aligned}
||I^h[v-Q_i]||_{W^0_\infty(\omega_i)} &= \left|\left| \sum_{j=1}^{N} \sum_{|\mathbf{m}|=0}^{k} \left[C_{\mathbf{m}} D^{(\mathbf{m})}(v-Q_i)(\mathbf{x}_j)(\mathbf{x}-\mathbf{x}_j)^{\mathbf{m}} \right] \phi_j(\mathbf{x}) \right|\right|_{W^0_\infty(\omega_i)} \\
&\leq \kappa \left[\sum_{|\mathbf{m}|=0}^{k} C_{\mathbf{m}} C h^{k+l+1-|\mathbf{m}|}|v|_{W^{k+l+1}_\infty(B^i)} h^{|\mathbf{m}|} \right] \max_{j \in \gamma(\omega_i)} ||\phi_j||_{W^0_\infty(\omega_i)} \\
&\leq Ch^{k+l+1}|v|_{W^{k+l+1}_\infty(B^i)}
\end{aligned}$$

where we have used (20), (7), and the fact that $|\mathbf{x}-\mathbf{x}_j|^{\mathbf{m}} \leq Ch^{|\mathbf{m}|}$, which is true by the quasiuniformity of the patches. Note here that the constant C depends on k but not on h.

For $t = 1$ we have:

$$\begin{aligned}
\left|\left| \frac{\partial}{\partial x} \left(I^h[v-Q_i] \right) \right|\right|_{W^0_\infty(\omega_i)} &= \left|\left| \sum_{j=1}^{N} \sum_{|\mathbf{m}|=0}^{k} \frac{\partial}{\partial x} \left[C_{\mathbf{m}} D^{(\mathbf{m})}(v-Q_i)(\mathbf{x}_j)(\mathbf{x}-\mathbf{x}_j)^{\mathbf{m}} \phi_j(\mathbf{x}) \right] \right|\right|_{W^0_\infty(\omega_i)} \\
&= \left|\left| \sum_{j=1}^{N} \sum_{|\mathbf{m}|=0}^{k} C_{\mathbf{m}} D^{(\mathbf{m})}(v-Q_i)(x_j,y_j)(y-y_j)^{m_2} \right.\right. \\
&\quad \left.\left. \left[m_1(x-x_j)^{m_1-1}\phi_j(x,y) + (x-x_j)^{m_1} \frac{\partial}{\partial x}\phi_j(x,y) \right] \right|\right|_{W^0_\infty(\omega_i)} \\
&\leq \kappa C \sum_{|\mathbf{m}|=0}^{k} h^{k+l+1-|\mathbf{m}|}|v|_{W^{k+l+1}_\infty(B^i)} h^{m_2} \\
&\quad \left(m_1 h^{m_1-1} C + h^{m_1} C h^{-1} \right) \\
&\leq Ch^{k+l}|v|_{W^{k+l+1}_\infty(B^i)}.
\end{aligned}$$

Here we have again used that $|\mathbf{x}-\mathbf{x}_j|^{|\mathbf{m}|} \leq Ch^{|\mathbf{m}|}$ and condition (7). Similarly

$$\begin{aligned}
\left|\left|\frac{\partial}{\partial y}\left(I^h[v-Q_i]\right)\right|\right|_{W^0_\infty(\omega_i)} &= \left|\left|\sum_{j=1}^{N}\sum_{|\mathbf{m}|=0}^{k}\frac{\partial}{\partial y}\left[C_{\mathbf{m}}D^{(\mathbf{m})}(v-Q_i)(\mathbf{x}_j)(\mathbf{x}-\mathbf{x}_j)^{\mathbf{m}}\phi_j(\mathbf{x})\right]\right|\right|_{W^0_\infty(\omega_i)} \\
&= \left|\left|\sum_{j=1}^{N}\sum_{|\mathbf{m}|=0}^{k} C_{\mathbf{m}}D^{(\mathbf{m})}(v-Q_i)(x_j,y_j)(x-x_j)^{m_1}\right.\right. \\
&\quad \left.\left.\left[m_2(y-y_j)^{m_2-1}\phi_j(x,y)+(y-y_j)^{m_2}\frac{\partial}{\partial y}\phi_j(x,y)\right]\right|\right|_{W^0_\infty(\omega_i)} \\
&\leq \kappa C\sum_{|\mathbf{m}|=0}^{k} h^{k+l+1-|\mathbf{m}|}|v|_{W^{k+l+1}_\infty(B^i)}h^{m_1} \\
&\quad \left(m_2h^{m_2-1}C+h^{m_2}Ch^{-1}\right) \\
&\leq Ch^{k+l}|v|_{W^{k+l+1}_\infty(B^i)}.
\end{aligned}$$

Therefore we have,

$$|I^h[v-Q_i]|_{W^1_\infty(\omega_i)} \leq Ch^{k+l}|v|_{W^{k+l+1}_\infty(B^i).}$$

We note that in general, by repeated differentiation, we have

$$|(x-x_j)^{m_1}(y-y_j)^{m_2}\phi_j(x,y)|_{W^t_\infty(\omega_i)} \leq Ch^{|\mathbf{m}|-t},$$

where C may depend on k but is independent of h. Therefore, a similar argument shows that also for $t=2,3,...,k+l+1$:

$$|I^h[v-Q_i]|_{W^t_\infty(\omega_i)} \leq Ch^{k+l+1-t}|v|_{W^{k+l+1}_\infty(B^i)}.$$

Now we have

$$\begin{aligned}
||I^h[v-Q_i]||_{W^t_\infty(\omega_i)} &= \max_{0\leq\bar{t}\leq t}|I^h[v-Q_i]|_{W^{\bar{t}}_\infty(\omega_i)} \\
&\leq \max_{0\leq\bar{t}\leq t} Ch^{k+l+1-\bar{t}}|v|_{W^{k+l+1}_\infty(B^i)} \\
&\leq Ch^{k+l+1-t}|v|_{W^{k+l+1}_\infty(B^i)}
\end{aligned}$$

for h small enough. Thus, from (18), (17), and the above inequality, we get

$$||v-I^h[v]||_{W^t_\infty(\omega_i)} \leq ||v-Q_i||_{W^t_\infty(\omega_i)}+||I^h[v-Q_i]||_{W^t_\infty(\omega_i)} \leq Ch^{k+l+1}|v|_{W^{k+l+1}_\infty(B^i)},$$

which is the desired result. □

Finally, we will extend the result to the whole domain $\Omega \subset \mathbb{R}^2$.

Theorem 3.6. *If $v \in W^{k+l+1}_\infty(\Omega)$, then for $t=0,...,k+l+1$ and $1\leq p\leq\infty$, we have*

$$\inf_{\xi\in S^{GFEM}}||v-\xi||_{W^t_p(\Omega)} \leq Ch^{k+l+1-t}|v|_{W^{k+l+1}_\infty(\Omega)}, \tag{21}$$

where C depends on k, l, κ, p, t and Ω, but is independent of v.

Proof. We recall that

$$\Omega \subset \bigcup_{j=1}^{N} B^j \subset B(0,R),$$

where R was defined in (15). Therefore from Lemma 3.5 and (16), we have

$$\begin{aligned}
||v - I^h[v]||_{W^t_\infty(\Omega)} &= \max_{1\le j\le N} ||v - I^h[v]||_{W^{k+l+1}_\infty(\omega_j)} \\
&\le Ch^{k+l+1-t} \max_{1\le j\le N} |v|_{W^{k+l+1-t}_\infty(B^j)} \\
&\le Ch^{k+l+1-t}|v|_{W^{k+l+1-t}_\infty(B(0,R))} \\
&\le Ch^{k+l+1-t}|v|_{W^{k+l+1-t}_\infty(\Omega)}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\inf_{\xi\in S^{GFEM}} ||v-\xi||_{W^t_p(\Omega)} &\le ||v - I^h[v]||_{W^t_p(\Omega)} \\
&\le C||v - I^h[v]||_{W^t_\infty(\Omega)} \\
&\le Ch^{k+l+1-t}|v|_{W^{k+l+1-t}_\infty(\Omega)},
\end{aligned}$$

which completes the proof. □

4 Numerical Results

In this section, we will present numerical experiments to illuminate the result of Theorem 3.6.

For a domain $\Omega := [0,1]^2$, we will consider the following model problem:

$$\begin{cases} -\Delta u = f \text{ in } \Omega \\ \frac{\partial u}{\partial n} = g \text{ on } \partial\Omega \end{cases}, \tag{22}$$

where f and g must satisfy the compatibility condition

$$\int_\Omega f\, d\mathbf{x} + \int_{\partial\Omega} g\, ds = 0.$$

The variational formulation of (22) is:

$$\begin{aligned}
&\text{Find } u \in H^1(\Omega) \text{ such that} \\
&B(u,v) = F(v) \text{ for all } v \in H^1(\Omega),
\end{aligned} \tag{23}$$

where

$$\begin{aligned}
B(u,v) &= \int_\Omega \nabla u \cdot \nabla v\, d\mathbf{x} \text{ and} \\
F(v) &= \int_\Omega fv\, d\mathbf{x} + \int_{\partial\Omega} gv\, ds.
\end{aligned}$$

In this section, we will consider f and g such that the exact solution of (23) is $u(\mathbf{x}) := u(x,y) = e^{x+y} = e^{\mathbf{x}}$.

In the Generalized Finite Element Method (GFEM), which is a Galerkin method, we seek $u_h \in S^{GFEM}$ that satisfies:

$$B(u_h, \chi) = F(\chi) \quad \text{for all } \chi \in S^{GFEM}. \tag{24}$$

Note that the solution u_h is unique up to a constant, and unless additional conditions are imposed, the stiffness matrix of the resulting linear system is not invertible. We will discuss how to solve such linear systems later in the section.

The space S^{GFEM} used in (24) is defined, as in (8), by

$$S^{GFEM} := \text{span}\{\phi_{i_1,i_2}(\mathbf{x})(\mathbf{x} - \mathbf{x}_{i_1,i_2})^{\mathbf{s}} : 0 \leq i_1, i_2 \leq \bar{N}, |\mathbf{s}| = 0...k\},$$

with the particle $\mathbf{x}_{i_1,i_2} := (hi_1, hi_2)$, and $h = 1/\bar{N}$, where $\bar{N}$ is a positive integer. Note that here we are enumerating the PU functions ϕ_j and particles $\mathbf{x}_j$ differently than before, using the double index $0 \leq i_1, i_2 \leq \bar{N}$. A correspondence can be established to the single index $1 \leq j \leq N$ by taking $j = i_1(\bar{N}+1)+i_2+1$ and $N = (\bar{N}+1)^2$.

If ϕ_{i_1,i_2} used in the construction of $GFEM$, reproduce the polynomials of degree k, then from Theorem 3.6 it is clear that

$$|u - u_h|_{H^1(\Omega)} \leq \inf_{\xi \in S^{GFEM}} |u - \xi|_{H^1(\Omega)} \leq Ch^{l+k}|u|_{W_\infty^{k+l+1}(\Omega)}.$$

We will use the following two classes of PU functions in GFEM to compute u_h, namely RKP functions that reproduce polynomials of degree $l = 0$ and $l = 1$. The PU functions are centered at the nodes $\mathbf{x}_{i_1,i_2}$ and are of the form given in (6), with the double-index (i_1, i_2), (6) becomes

$$\phi_{i_1,i_2}(\mathbf{x}) := w_{i_1,i_2}(\mathbf{x}) \sum_{|\mathbf{t}|=0}^{l} (\mathbf{x} - \mathbf{x}_{i_1,i_2})^{\mathbf{t}} b_{\mathbf{t}}(\mathbf{x}). \tag{25}$$

Here we need to choose the weight functions w_{i_1,i_2} which generate the associated PU function; these associated functions will be referred to as RKP PU functions. One possible choice (see [4], [2] or [10]) is the conical weight function which, in 1 dimension, has the form:

$$\bar{w}(x) = \begin{cases} [1 - (x/R)^2]^L, & |x| \leq R \\ 0, & |x| > R, \end{cases} \tag{26}$$

where L is a parameter that controls the smoothness of the function (since $\bar{w} \in C^{L-1}$) and R is a parameter that determines the radius of support. For the following computations, we have chosen $R = 2$, together with $L = 2$ and $L = 4$. The function $\bar{w}(\mathbf{x})$ is scaled and translated to each node $\mathbf{x}_{i_1,i_2}$ and the two-dimensional weight functions $w_{i_1,i_2}(\mathbf{x})$ with support on the patches ω_{i_1,i_2} are obtained by the taking the tensor products, i.e.

$$w_{i_1,i_2}(\mathbf{x}) := w_{i_1,i_2}(x,y) = \bar{w}\left(\frac{x - hi_1}{h}\right) \bar{w}\left(\frac{y - hi_2}{h}\right) \text{ for } 0 \leq i_1, i_2, \leq N.$$

Note here that since the support of the function $\bar{w}(x)$ is $|x| \leq R$, the support of w_{i_1,i_2} is $[hi_1 - Rh, hi_1 + Rh] \times [hi_2 - Rh, hi_2 + Rh]$. Since it can be seen from the definition of ϕ_{i_1,i_2} in (25) that the support of w_{i_1,i_2} is the same as the support of ϕ_{i_1,i_2}, we define the patches ω_{i_1,i_2} by

$$\omega_{i_1,i_2} := \text{supp}(\phi_{i_1,i_2}) = \text{supp}(w_{i_1,i_2}) = [hi_1 - Rh, hi_1 + Rh] \times [hi_2 - Rh, hi_2 + Rh].$$

We have seen before, in Theorem 2.7, that the ϕ_{i_1,i_2} have global approximation properties. In the following, we will examine the approximation properties of ϕ_{i_1,i_2} together with local approximation spaces of degree k. Note that for the numerical results below, the integrals in the stiffness matrix and the load vector (corresponding to (24)), were computed using a 16×16-point Gauss quadrature rule.

Table 1: H_1 seminorm of the error, $R = 2$, Conical Weight ($L = 2$)

	$\lvert u - u_h \rvert_{H^1(\Omega)}$			
h	$l = 1, k = 0$	$l = 1, k = 1$	$l = 2, k = 0$	$l = 2, k = 1$
1/10	3.89277E-02	6.31748E-04	3.72158E-03	9.86775E-05
1/15	2.60315E-02	2.90724E-04	1.67433E-03	3.31752E-05
1/20	1.95546E-02	1.66261E-04	9.47428E-04	1.47741E-05
1/25	1.56594E-02	1.07456E-04	6.08492E-04	7.78774E-06
1/30	1.30586E-02	7.51050E-05	4.23548E-04	4.58800E-06
1/35	1.11988E-02	5.54321E-05	3.11693E-04	2.92393E-06
1/40	9.80275E-03	4.25852E-05	2.38935E-04	1.97547E-06
1/45	8.71622E-03	3.37365E-05	1.88969E-04	1.39619E-06
	$O(h)$	$O(h^2)$	$O(h^2)$	$O(h^3)$

The values of $|u - u_h|_{H^1(\Omega)}$, with ϕ_{i_1,i_2} constructed using the conical weight functions with $L = 2$, are shown in Table 1. Here we observe that $|u - u_h|_{H^1(\Omega)}$ are $O(h^{k+l})$, where $k = 0, 1$ and $l = 1, 2$. This illuminates (21) of Theorem 3.6. In particular, for $l = 1$ and $k = 1$ we get the same order of convergence, $O(h^2)$ as for the case $l = 2$ and $k = 0$. However, we note that error is smaller for the column where where $l = 1$ and $k = 1$.

Using linear (or higher order) local approximation spaces (i.e., increasing k) may be more computationally efficient than using higher order RKP functions (increasing l). This is because higher order RKP functions become increasingly expensive to compute, as each point evaluation requires solving a $\frac{(l+1)(l+2)}{2} \times \frac{(l+1)(l+2)}{2}$ linear system. However, increasing k results in a larger global linear system, as the dimensions of the stiffness matrix are $\frac{N(k+1)(k+2)}{2} \times \frac{N(k+1)(k+2)}{2}$, where N is the number of patches.

Table 2: H_1 seminorm of the error, $R = 2$, Conical Weight ($L = 4$)

	$\lvert u - u_h \rvert_{H^1(\Omega)}$			
h	$l = 1, k = 0$	$l = 1, k = 1$	$l = 2, k = 0$	$l = 2, k = 1$
1/10	2.24036E-02	3.94090E-04	5.63202E-03	2.44373E-05
1/15	1.51859E-02	1.84814E-04	2.58490E-03	8.61940E-06
1/20	1.14841E-02	1.06457E-04	1.47621E-03	4.19781E-06
1/25	9.23300E-03	6.89850E-05	9.53123E-04	2.40869E-06
1/30	7.71968E-03	4.83827E-05	6.65702E-04	1.52436E-06
1/35	6.63253E-03	3.57650E-05	4.91067E-04	1.03006E-06
1/40	5.81376E-03	2.75148E-05	3.77101E-04	7.29970E-07
1/45	5.17492E-03	2.18241E-05	2.98645E-04	5.36507E-07
	$O(h)$	$O(h^2)$	$O(h^2)$	$O(h^3)$

Table 2 shows the values of $|u-u_h|_{H^1(\Omega)}$, with ϕ_{i_1,i_2} constructed using conical weight functions with $L = 4$. It can be seen from (26) and Remark 2.6 that that the PU functions generated by this choice of weight function are smoother than for the case $L = 2$. We observe that, in general, the smoother PU functions yield smaller values of the error $|u - u_h|_{H^1(\Omega)}$ with enrichment. However, this is not always the case for PU functions with no enrichment. Comparing the columns corresponding to $l = 2$ and $k = 0$ of Tables 1 and 2, we note that the semi-norm of the error is smaller for $L = 2$ than for $L = 4$.

We note that it is possible to determine *a priori* the relative approximation qualities of the GFEM shape functions. This can be done by examining the interpolation error for polynomials of degree $k + l + 1$ (since all polynomials of degree $k + l$ are reproduced exactly). Specifically, it was shown in [4], for the case $k = 0$ with $\Omega \subset \mathbb{R}^n$, that for $q > \frac{n}{2}$ when $n \geq 2$, and $q = 0$ for $n = 1$, we have

$$\sup_{u \in H^{l+2+q}(\Omega)} \lim_{h \to 0} \frac{||u - I_h[u]||^2_{H^1(\Omega)}}{h^{2l} Q_h(u)} = \bar{\lambda}, \tag{27}$$

where

$$Q_h(u) = |u|^2_{H^{l+1}(\Omega)} + h \sum_{|\alpha|=l+2} ||D^\alpha u||^2_{H^q(\Omega)}.$$

Here $\bar{\lambda}$ is the largest eigenvalue of the matrix

$$A_{ij} = \int_I \frac{1}{\alpha(i)!\alpha(j)!} \nabla \xi_{\alpha(i)} \cdot \nabla \xi_{\alpha(j)} \, d\mathbf{x},$$

where $\alpha = \alpha(i)$ is an enumeration of the multi-index α with $|\alpha(i)| = l + 1$, $I = [-1/2, 1/2]^n$ and $\xi_\alpha = \mathbf{x}^\alpha - I^h[\mathbf{x}^\alpha]$. From (27), we can conclude that smaller values of $\bar{\lambda}$ imply better approximation quality of the GFEM interpolant.

The values of $\bar{\lambda}$ for the GFEM spaces associated with a few choices of weight functions are given in Table 3. Comparing rows 1 and 2, we can see that the values of $\bar{\lambda}$ are smaller for conical weight with $L = 4$ for all cases except $l = 2, k = 0$. This indicates we expect better approximation properties from

Conical weight PU functions with $L = 4$ except in the case $l = 2, k = 0$, which is confirmed by the data in Tables 1 and 2 (columns 2,3 and 5). Note that the formula in (27) was only proven for $k = 0$; an extension for $k > 0$ will be addressed in a forthcoming paper.

Table 3: Values of $\bar{\lambda}$ for $R = 2$

	$\bar{\lambda}$, $R = 2$			
Weight function	$l = 1, k = 0$	$l = 1, k = 1$	$l = 2, k = 0$	$l = 2, k = 1$
Conical wt. ($L = 2$)	7.5704E-03	2.3952E-04	1.1030E-02	5.9658E-04
Conical wt. ($L = 4$)	2.7163E-03	1.0160E-04	1.9558E-02	3.3528E-04
Cubic spline wt.	0.0000E+00	0.0000E+00	1.6881E-02	3.1032E-04

Another choice of a weight function we will consider is the cubic spline weight function, which can be written as:

$$\bar{w}(x) := \begin{cases} \frac{2}{3} - 4|x|^2 + 4|x|^3 & \text{for } |x| \leq \frac{1}{2} \\ \frac{4}{3} - 4|x| + 4|x|^2 - \frac{4}{3}|x|^2 & \text{for } \frac{1}{2} < |x| \leq 1 \\ 0 & \text{for } |x| > 1 \end{cases} . \tag{28}$$

Note that this is a C^2 function. The values of $|u - u_h|_{H^1(\Omega)}$ are shown in Table 4.

Table 4: H_1 seminorm of the error, $R = 2$, Cubic Spline Weight

	$\lvert u - u_h \rvert_{H^1(\Omega)}$			
h	$l = 1, k = 0$	$l = 1, k = 1$	$l = 2, k = 0$	$l = 2, k = 1$
1/10	2.467210E-05	3.582517E-07	5.392829E-03	1.867585E-05
1/15	7.437905E-06	8.361688E-08	2.471138E-03	5.844202E-06
1/20	3.165037E-06	7.152803E-08	1.410205E-03	2.536195E-06
1/25	1.628884E-06	1.067760E-07	9.101249E-04	1.325887E-06
1/30	9.458857E-07	1.552849E-07	6.354979E-04	8.259678E-07
1/35	5.971220E-07	2.117555E-07	4.686974E-04	7.097027E-07
1/40	4.007614E-07	2.708601E-07	3.598728E-04	6.794109E-07
1/45	2.818711E-07	3.251667E-07	2.849709E-04	5.732386E-07
	-	-	$O(h^2)$	$O(h^3)$

Here it can be observed that the RKP functions with $l = 1$ achieve a higher convergence rate than expected (close to $O(h^3)$ vs. the expected $O(h)$ for $k = 0$, up to the point where roundoff errors affect the convergence rate). This is due to the *quasi-reproducing property* of the cubic spline weight function for $R = 2$ (see [2], [4]), on which we do not elaborate in this paper. This phenomenon does not occur for the RKP functions with $l = 2$ or with $R < 2$. However, enriching by linear functions improved the order of convergence by a power of h, up to machine precision.

We have mentioned before that the stiffness matrix for the Neumann problem (22) is not invertible and the resulting linear system will require special methods

to solve. For the computations presented above we have used an algorithm proposed in [15] and [3], which solves iteratively a perturbed system that is ill-conditioned but not singular.

5 Appendix

In this section, we will prove that the coefficients $c_{\mathbf{t}}$, defined in (11), satisfy the property (13) as mentioned in Section 3. We recall that for any multi-index $\mathbf{t} = (t_1, t_2)$, $c_{\mathbf{t}}$ is given by

$$c_{\mathbf{t}} := \sum_{\substack{\mathbf{t} \le \mathbf{m} \le \mathbf{i} \\ |\mathbf{m}| \le k}} \frac{k!(k+l-|\mathbf{m}|)!\mathbf{i}!}{(k-|\mathbf{m}|)!(k+l)!(\mathbf{i}-\mathbf{m})!(\mathbf{m}-\mathbf{t})!\mathbf{t}!}(-1)^{|\mathbf{m}-\mathbf{t}|}.$$

where k, l are non-negative integers and $0 \le |\mathbf{i}| \le k+l$. Using the change of variable $\bar{\mathbf{m}} = \mathbf{m} - \mathbf{t}$, $c_{\mathbf{t}}$ can be rewritten as:

$$c_{\mathbf{t}} := \frac{k!\mathbf{i}!}{(k+l)!\mathbf{t}!} \sum_{\substack{\mathbf{0} \le \bar{\mathbf{m}} \le \mathbf{i}-\mathbf{t} \\ |\bar{\mathbf{m}}| \le k-|\mathbf{t}|}} \frac{(k+l-|\bar{\mathbf{m}}+\mathbf{t}|)!}{(k-|\bar{\mathbf{m}}+\mathbf{t}|)!(\mathbf{i}-\bar{\mathbf{m}}-\mathbf{t})!(\bar{\mathbf{m}})!}(-1)^{|\bar{\mathbf{m}}|}.$$

Note that $c_{\mathbf{t}}$ depends on $\mathbf{i}$, k, and l, and therefore in this section, we will write $c_{\mathbf{t};\mathbf{i},k,l}$ for $c_{\mathbf{t}}$; moreover the index of summation will be $\mathbf{m}$ instead of $\bar{\mathbf{m}}$ as in the above expression.

We will prove the property (13) for the one-dimensional case. This will be helpful in understanding the arguments in the two-dimensional proof. We note that in one dimension, the multi-indices $\mathbf{t}, \mathbf{i}$, and $\mathbf{m}$ in the definition of $c_{\mathbf{t}}$ are replaced by non-negative integers t, i and m, respectively.

Lemma 5.1. *Let k and l be given and i be such that $0 \le i \le k+l$. For any non-negative integer t, let*

$$c_{t;i,k,l} = \begin{cases} \frac{k!\cdot i!}{(k+l)!\cdot t!} \sum_{m=0}^{min(k-t,i-t)} \frac{(k+l-m-t)!}{(k-m-t)!(i-m-t)!m!} \cdot (-1)^m, \text{ for } 0 \le t \le i \\ 0 \text{ for } t > i. \end{cases} \tag{29}$$

Note that it is possible that $c_{t;i,k,l} = 0$ even when $0 \le t \le i$, for example when $t > k$.

Then the following hold:

i) For $0 \le i \le k+l$,

$$c_{t;i,k,l+1} = \frac{k+l+1-t}{k+l+1} c_{t;i,k,l} + \frac{t+1}{k+l+1} c_{t+1;i,k,l} \tag{30}$$

ii) If $t < i-l$ then

$$c_{t;i,k,l} = 0 \tag{31}$$

iii) For $0 \le i \le k+l$,

$$\sum_{t=0}^{i} c_{t;i,k,l} = 1. \tag{32}$$

Note that (31) and (32) are the 1D equivalent of (13).

Proof. i) Let $M := \min(k-t, i-t)$. Then

$$c_{t;i,k,l+1} = \frac{k!i!}{(k+l+1)!t!} \sum_{m=0}^{M} \frac{(k+l+1-m-t)!}{(k-m-t)!(i-m-t)!m!} (-1)^m$$

$$= \frac{k!i!}{(k+l+1)(k+l)!t!} \sum_{m=0}^{M} \frac{[(k+l+1-t)-m](k+l-m-t)!(-1)^m}{(k-m-t)!(i-m-t)!m!}$$

$$= \frac{k+l+1-t}{k+l+1} c_{t;i,k,l} - \frac{k!i!}{(k+l+1)!t!} \sum_{m=1}^{M} \frac{(k+l-m-t)!(-1)^m}{(k-m-t)!(i-m-t)!(m-1)!}$$

Next, using the reindexing $\widetilde{m} := m-1$ in the above equality, we have:

$$\begin{aligned} c_{t;i,k,l+1} &= \frac{k+l+1-t}{k+l+1} c_{t;i,k,l} \\ &\quad - \frac{t+1}{k+l+1} \frac{k!i!}{(k+l)!(t+1)!} \sum_{\widetilde{m}=0}^{M-1} \frac{(k+l-\widetilde{m}-1-t)!(-1)^{\widetilde{m}+1}}{(k-\widetilde{m}-1-t)!(i-\widetilde{m}-1-t)!\widetilde{m}!} \\ &= \frac{k+l+1-t}{k+l+1} c_{t;i,k,l} + \frac{t+1}{k+l+1} c_{t+1;i,k,l}, \end{aligned}$$

where the second term of the last line was obtained using the definition of $c_{t+1;i,k,l}$ (see (29)). Note that in some cases M may be negative, in which case from the definition (29), $c_{t;i,k,l+1} = c_{t;i,k,l} = c_{t+1;i,k,l} = 0$, and therefore (30) holds trivially.

ii) First we consider the case $i = k+l$. We have:

$$\begin{aligned} c_{t;k+l,k,l} &= \frac{k!(k+l)!}{(k+l)!t!} \sum_{m=0}^{min(k-t,k+l-t)} \frac{(k+l-m-t)!}{(k-m-t)!(k+l-m-t)!m!} (-1)^m \\ &= \frac{k!}{t!} \sum_{m=0}^{k-t} \frac{(-1)^m}{(k-m-t)!m!} \\ &= \delta_{kt}, \end{aligned} \tag{33}$$

where δ_{kt} is the Kronecker delta. The last equality is obtained from the binomial identity

$$\sum_{\ell=0}^{p} \frac{p!}{(p-\ell)!\ell!} (-1)^\ell = \begin{cases} 0 \text{ for } p > 0 \\ 1 \text{ for } p = 0 \end{cases}, \tag{34}$$

with $\ell = m$, $p = k-t$, and dividing both sides by $(k-t)!$. Therefore $c_{t;k+l,k,l} = 0$ for $t < k = i-l$, which is the desired result (31). Also note that if $k = t$ then $c_{t;k+l,k,l} = c_{k;k+l,k,l} = 1$, which we will use in a later part of the proof.

Next, suppose $i < k+l$, and we prove (31) by induction on l. Let $l = 0$; we first show that $c_{t;i,k,0} = 0$ for $t < i-l = i$. From the definition of $c_{t;i,k,l}$ with

$l = 0$ (see (29)) we have

$$c_{t;i,k,0} = \frac{i!}{t!} \sum_{m=0}^{min(k-t,i-t)} \frac{1}{(i-m-t)!m!} (-1)^m.$$

Since $i < k + l = k$, it follows that $\min(k - t, i - t) = i - t$, and therefore

$$c_{t;i,k,0} = \frac{i!}{t!} \sum_{m=0}^{i-t} \frac{1}{(i-m-t)!m!} (-1)^m \tag{35}$$

$$= \delta_{ti}. \tag{36}$$

Here, we have used (34) with $\ell = m$, $p = i - t$ and have divided both sides by $(i - t)!$. Hence we conclude that (31) is true for $l = 0$.

Next suppose (31) is true for some $l \geq 0$, in other words

$$c_{t;i,k,l} = 0 \text{ for } t < i - l. \tag{37}$$

We will show that $c_{t;i,k,l+1} = 0$ for $t < i - l - 1$. Using $t + 1$ in the induction hypothesis (37), we have

$$c_{t+1;i,k,l} = 0 \text{ for } t < i - l - 1.$$

Hence from (37) and (30),

$$c_{t;i,k,l+1} = 0 \text{ for } t < i - l - 1.$$

Thus we have proven that (31) is true for $l + 1$, which completes the induction argument.

iii) Again, we first consider the case $i = k + l$. Previously, we have shown in (33) that $c_{t;k+l,k,l} = \delta_{kt}$. Therefore

$$\sum_{t=0}^{k+l} c_{t;k+l,k,l} = \sum_{t=0}^{k+l} \delta_{kt} = 1,$$

and hence (32) holds for $i = k + l$.

For $i < k+l$ we again use induction on l. The case $l = 0$ follows immediately from (36), since

$$\sum_{t=0}^{i} c_{t,i,k,0} = \sum_{t=0}^{i} \delta_{ti} = 1.$$

Next suppose (32), i.e. $\sum_{t=0}^{i} c_{t;i,k,l} = 1$, is true for some $l \geq 0$. We will show that $\sum_{t=0}^{i} c_{t;i,k,l+1} = 1$. Note that

$$\sum_{t=0}^{i} c_{t;i,k,l+1} = \sum_{t=0}^{i-1} c_{t;i,k,l+1} + c_{i;i,k,l+1}$$

and therefore from (30),

$$\sum_{t=0}^{i} c_{t;i,k,l+1} = \sum_{t=0}^{i-1} \left[\frac{(k+l+1)-t}{k+l+1}\, c_{t;i,k,l} + \frac{t+1}{k+l+1}\, c_{t+1;i,k,l} \right] + c_{i;i,k,l+1}$$
$$= \sum_{t=0}^{i-1} c_{t;i,k,l} - \sum_{t=0}^{i-1} \frac{t}{k+l+1}\, c_{t;i,k,l} + \sum_{t=0}^{i-1} \frac{t+1}{k+l+1}\, c_{t+1;i,k,l} + c_{i;i,k,l+1}$$

Now from the induction hypothesis (32), it follows that $\sum_{t=0}^{i-1} c_{t;i,k,l} = 1 - c_{i;i,k,l}$. Using this and the re-indexing $\bar{t} := t+1$ in the last summation term in the above equality, we have

$$\sum_{t=0}^{i} c_{t;i,k,l+1} = 1 - c_{i;i,k,l} - \sum_{t=0}^{i-1} \frac{t}{k+l+1}\, c_{t;i,k,l} + \sum_{\bar{t}=1}^{i} \frac{\bar{t}}{k+l+1}\, c_{\bar{t};i,k,l} + c_{i;i,k,l+1}$$
$$= 1 - c_{i;i,k,l} - \sum_{t=1}^{i-1} \frac{t}{k+l+1}\, c_{t;i,k,l}$$
$$+ \sum_{\bar{t}=1}^{i-1} \frac{\bar{t}}{k+l+1}\, c_{\bar{t};i,k,l} + \frac{i}{k+l+1}\, c_{i;i,k,l} + c_{i;i,k,l+1}$$
$$= 1 - c_{i;i,k,l} + \frac{i}{k+l+1}\, c_{i;i,k,l} + c_{i;i,k,l+1}. \quad (38)$$

Now from (30), we have

$$c_{i;i,k,l+1} = \frac{k+l+1-i}{k+l+1}\, c_{i;i,k,l} + \frac{i+1}{k+l+1}\, c_{i+1;i,k,l}$$
$$= \frac{k+l+1-i}{k+l+1}\, c_{i;i,k,l},$$

since $c_{i+1;i,k,l} = 0$ from the definition (29). Therefore from (38), we get

$$\sum_{t=0}^{i} c_{t;i,k,l+1} = 1 + \frac{i-k-l-1}{k+l+1}\, c_{i;i,k,l} + \frac{k+l+1-i}{k+l+1}\, c_{i;i,k,l}$$
$$= 1.$$

This establishes (32) by the induction argument. □

Remark 5.2. We recall that for a given k and l, we first proved (31) and (32) for $i = k+l$ directly. We then used an induction argument to prove (31) and (32) for $i < k+l$. We note that we cannot use induction on l (as we have used in the case $i < k+l$) to prove (31) and (32) for $i = k+l$; the proof employs the recursion relation (30) with $i = k+l+1$, but (30) is only true for $i \le k+l$. This idea of separating the case $i = k+l$ from the case $i < k+l$ in proving (31) and (32) will help us understand a similar result in two dimensions.

We will now prove the property (13) in two dimensions.

Lemma 5.3. *Let k and l be given non-negative integers, and $\mathbf{i}$ be such that $0 \le |\mathbf{i}| \le k+l$. For any multi-index $\mathbf{t} := (t_1, t_2)$ with non-negative components, let*

$$\mathcal{M} := \mathcal{M}_{\mathbf{t}} := \mathcal{M}_{(t_1,t_2)} := \{\mathbf{m} : 0 \le \mathbf{m} \le \mathbf{i} - \mathbf{t}, \text{ and } |\mathbf{m}| \le k - |\mathbf{t}|\},$$

where $\mathcal{M}$ depends on $\mathbf{t}$. Next we define define

$$c_{\mathbf{t};\mathbf{i},k,l} = \begin{cases} \frac{k!\,\mathbf{i}!}{(k+l)!\,\mathbf{t}!} \sum_{\mathbf{m}\in\mathcal{M}} \frac{(k+l-|\mathbf{m}+\mathbf{t}|)!}{(k-|\mathbf{m}+\mathbf{t}|)!(\mathbf{i}-\mathbf{m}-\mathbf{t})!\mathbf{m}!} (-1)^{|\mathbf{m}|} & \text{for } 0 \le \mathbf{t} \le \mathbf{i} \\ 0 \text{ otherwise .} \end{cases} \tag{39}$$

Note that it is possible that $c_{\mathbf{t};\mathbf{i},k,l} = 0$ even when $0 \le \mathbf{t} \le \mathbf{i}$, for example when $\mathcal{M}$ is the empty set.

We also assume the following:

a) If $|\mathbf{i}| = k+l$, then for all $\mathbf{t}$ such that $|\mathbf{t}| < k$, we have $c_{\mathbf{t};\mathbf{i},k,l} = 0$.

b) If $|\mathbf{i}| = k+l$, then

$$\sum_{0\le\mathbf{t}\le\mathbf{i}} c_{\mathbf{t};\mathbf{i},k,l} = 1.$$

Then the following hold:

i) For $0 \le |\mathbf{i}| \le k+l$,

$$c_{\mathbf{t};\mathbf{i},k,l+1} = \frac{k+l+1-|\mathbf{t}|}{k+l+1} c_{\mathbf{t};\mathbf{i},k,l} + \frac{t_1+1}{k+l+1} c_{(t_1+1,t_2);\mathbf{i},k,l} + \frac{t_2+1}{k+l+1} c_{(t_1,t_2+1);\mathbf{i},k,l}. \tag{40}$$

We note that the assumptions a) and b) for $|\mathbf{i}| = k+l$ stated above is not needed for this part.

ii) If $|\mathbf{t}| < |\mathbf{i}| - l$ then

$$c_{\mathbf{t};\mathbf{i},k,l} = 0 \tag{41}$$

iii) For $0 \le |\mathbf{i}| \le k+l$,

$$\sum_{0\le\mathbf{t}\le\mathbf{i}} c_{\mathbf{t};\mathbf{i},k,l} = 1. \tag{42}$$

Remark 5.4. The assumptions a) and b) can be easily verified for particular values of k and l. Also note that (41) and (42) are precisely the properties stated in (13). Here we do not write the constraint $|\mathbf{t}| \le k$ on the index of summation $\mathbf{t}$ in (42) (compare with (13)), since the set $\mathcal{M}_{\mathbf{t}}$ is empty for $|\mathbf{t}| > k$, and consequently $c_{\mathbf{t};\mathbf{i},k,l} = 0$ for $|\mathbf{t}| > k$.

Proof of Lemma 5.3. Using the definition of $c_{\mathbf{t};\mathbf{i},k,l}$ in (39),

$$\begin{aligned} c_{\mathbf{t};\mathbf{i},k,l+1} &= \frac{k!\mathbf{i}!}{(k+l+1)!\,\mathbf{t}!} \sum_{\mathbf{m}\in\mathcal{M}} \frac{(k+l+1-|\mathbf{m}+\mathbf{t}|)!}{(k-|\mathbf{m}+\mathbf{t}|)!(\mathbf{i}-\mathbf{m}-\mathbf{t})!\mathbf{m}!}\,(-1)^{|\mathbf{m}|} \\ &= \frac{k!\mathbf{i}!}{(k+l+1)(k+l)!\mathbf{t}!} \sum_{\mathbf{m}\in\mathcal{M}} \frac{[(k+l+1-|\mathbf{t}|)-|\mathbf{m}|](k+l-|\mathbf{m}+\mathbf{t}|)!(-1)^{|\mathbf{m}|}}{(k-|\mathbf{m}+\mathbf{t}|)!(\mathbf{i}-\mathbf{m}-\mathbf{t})!\mathbf{m}!} \\ &= \frac{k+l+1-|\mathbf{t}|}{k+l+1}\, c_{\mathbf{t};\mathbf{i},k,l} - \frac{k!\mathbf{i}!}{(k+l+1)!\mathbf{t}!} \sum_{\mathbf{m}\in\mathcal{M}} \frac{|\mathbf{m}|(k+l-|\mathbf{m}+\mathbf{t}|)!(-1)^{|\mathbf{m}|}}{(k-|\mathbf{m}+\mathbf{t}|)!(\mathbf{i}-\mathbf{m}-\mathbf{t})!\mathbf{m}!}. \end{aligned} \tag{43}$$

The last term of (43) can be written as:

$$\begin{aligned} &\frac{k!\mathbf{i}!}{(k+l+1)!\mathbf{t}!} \sum_{\mathbf{m}\in\mathcal{M}} \frac{(m_1+m_2)(k+l-|\mathbf{m}+\mathbf{t}|)!(-1)^{|\mathbf{m}|}}{(k-|\mathbf{m}+\mathbf{t}|)!(\mathbf{i}-\mathbf{m}-\mathbf{t})!m_1!m_2!} \\ &\qquad = \frac{k!\mathbf{i}!}{(k+l+1)!\mathbf{t}!} \left[\sum_{\substack{\mathbf{m}\in\mathcal{M} \\ m_1\neq 0}} \frac{(k+l-|\mathbf{m}+\mathbf{t}|)!(-1)^{|\mathbf{m}|}}{(k-|\mathbf{m}+\mathbf{t}|)!(\mathbf{i}-\mathbf{m}-\mathbf{t})!(m_1-1)!m_2!} \right. \\ &\qquad\qquad \left. + \sum_{\substack{\mathbf{m}\in\mathcal{M} \\ m_2\neq 0}} \frac{(k+l-|\mathbf{m}+\mathbf{t}|)!(-1)^{|\mathbf{m}|}}{(k-|\mathbf{m}+\mathbf{t}|)!(\mathbf{i}-\mathbf{m}-\mathbf{t})!m_1!(m_2-1)!} \right]. \end{aligned} \tag{44}$$

Here the terms with $m_1 = 0$ and $m_2 = 0$ can be discarded, since for example, $m_1/m_1! = 0/0! = 0$, so these terms do not contribute to the sum.

We will now examine each of the summation terms in the right hand side of (44) separately. We will use a re-indexing which shifts the indices m_1 and m_2 down by substituting

$$\bar{m}_1 := m_1 - 1 \text{ and } \bar{m}_2 = m_2 - 1.$$

We next define the set

$$\begin{aligned} \bar{\mathcal{M}}_1 := \{\bar{\mathbf{m}} := (\bar{m}_1, m_2) : \quad & 0 \le \bar{m}_1 \le i_1 - t_1 - 1, 0 \le m_2 \le i_2 - t_2, \\ & 0 \le \bar{m}_1 + m_2 \le k - t_1 - t_2 - 1\} \end{aligned}$$

and

$$\begin{aligned} \bar{\mathcal{M}}_2 := \{\bar{\bar{\mathbf{m}}} := (m_1, \bar{m}_2) : \quad & 0 \le m_1 \le i_1 - t_1, 0 \le m_2 \le i_2 - t_2 - 1, \\ & 0 \le m_1 + \bar{m}_2 \le k - t_1 - t_2 - 1\}. \end{aligned}$$

It is easy to check that

$$\bar{\mathcal{M}}_1 = \mathcal{M}_{(t_1+1,t_2)} \text{ and } \bar{\mathcal{M}}_2 = \mathcal{M}_{(t_1,t_2+1)}.$$

Now, with the notation $\mathbf{i} := (i_1, i_2)$ and using the definition of $c_{(t_1+1,t_2);\mathbf{i},k,l}$ (see (39), where the sum in this case is over the set $\mathcal{M}_{(t_1+1,t_2)}$), the first summation term of the right hand side of (44) can be written as:

$$\sum_{\substack{\mathbf{m}\in\mathcal{M}\\ m_1\neq 0}} \frac{(k+l-|\mathbf{m}+\mathbf{t}|)!(-1)^{|\mathbf{m}|}}{(k-|\mathbf{m}+\mathbf{t}|)!(\mathbf{i}-\mathbf{m}-\mathbf{t})!(m_1-1)!m_2!} =$$

$$= \sum_{\bar{\mathbf{m}}\in\bar{\mathcal{M}}_1} \frac{(k+l-\bar{m}_1-1-m_2-|\mathbf{t}|)!(-1)^{\bar{m}_1+1+m_2}}{(k-\bar{m}_1-1-m_2-|\mathbf{t}|)!(i_1-\bar{m}_1-1-t_1)!(i_2-m_2-t_2)!\bar{m}_1!m_2!}$$

$$= \sum_{\mathbf{m}\in\bar{\mathcal{M}}_1} \frac{(k+l-|\mathbf{m}|-t_1-1-t_2)!(-1)^{|\mathbf{m}|}}{(k-|\mathbf{m}|-t_1-1-t_2|)!(i_1-m_1-t_1-1)!(i_2-m_2-t_2)!\mathbf{m}!},$$

$$= -\frac{(k+l)!(t_1+1)!t_2!}{k!\mathbf{i}!}\, c_{(t_1+1,t_2);\mathbf{i},k,l} \tag{45}$$

Next, from the definition of $c_{(t_1,t_2+1);\mathbf{i},k,l}$ (see (39), where the sum in this case is over the set $\mathcal{M}_{(t_1,t_2+1)}$), the second summation term in the right hand side of (44) can similarly be written as:

$$\sum_{\substack{\mathbf{m}\in\mathcal{M}\\ m_2\neq 0}} \frac{(k+l-|\mathbf{m}+\mathbf{t}|)!(-1)^{|\mathbf{m}|}}{(k-|\mathbf{m}+\mathbf{t}|)!(\mathbf{i}-\mathbf{m}-\mathbf{t})!m_1!(m_2-1)!}$$

$$= -\frac{(k+l)!t_1!(t_2+1)!}{k!\mathbf{i}!}\, c_{(t_1,t_2+1);\mathbf{i},k,l}. \tag{46}$$

Finally, from (43) through (46), it follows that

$$c_{\mathbf{t};\mathbf{i},k,l+1} = \frac{k+l+1-|\mathbf{t}|}{k+l+1} c_{\mathbf{t};\mathbf{i},k,l} + \frac{t_1+1}{k+l+1} c_{(t_1+1,t_2);\mathbf{i},k,l} + \frac{t_2+1}{k+l+1} c_{(t_1,t_2+1);\mathbf{i},k,l}.$$

ii) The result (41) has been assumed for $|\mathbf{i}| = k+l$ in a). For $|\mathbf{i}| < k+l$, the proof is by induction on l. Suppose $l = 0$. Then from (39)

$$c_{\mathbf{t};\mathbf{i},k,0} = \frac{\mathbf{i}!}{\mathbf{t}!} \sum_{\mathbf{m}\in\mathcal{M}_\mathbf{t}} \frac{1}{(\mathbf{i}-\mathbf{m}-\mathbf{t})!\mathbf{m}!}\,(-1)^{|\mathbf{m}|}.$$

Since $|\mathbf{i}| < k+l = k$, the condition $\mathbf{m}\in\mathcal{M}_\mathbf{t}$ is equivalent to $0 \le \mathbf{m} \le \mathbf{i}-\mathbf{t}$ (in other words, the condition $|\mathbf{m}| < k-|\mathbf{t}|$ is automatically satisfied). Then

$$c_{\mathbf{t};\mathbf{i},k,0} = \frac{\mathbf{i}!}{\mathbf{t}!} \sum_{0\le\mathbf{m}\le\mathbf{i}-\mathbf{t}} \frac{1}{(\mathbf{i}-\mathbf{m}-\mathbf{t})!\mathbf{m}!}\,(-1)^{|\mathbf{m}|}$$

$$= \frac{\mathbf{i}!}{\mathbf{t}!} \sum_{0\le\mathbf{m}\le\mathbf{i}-\mathbf{t}} \frac{(-1)^{m_1+m_2}}{(i_1-t_1-m_1)!(i_2-t_2-m_2)!m_1!m_2!}$$

$$= \frac{\mathbf{i}!}{\mathbf{t}!} \left(\sum_{m_1=0}^{i_1-t_1} \frac{(-1)^{m_1}}{(i_1-t_1-m_1)!m_1!}\right)\left(\sum_{m_2=0}^{i_2-t_2} \frac{(-1)^{m_2}}{(i_2-t_2-m_2)!m_2!}\right). \tag{47}$$

Using the binomial identity

$$(1+x)^n = \sum_{j=0}^{n} \frac{n!}{(n-j)!j!} x^j$$

with $n = i_2 - t_2$, $j = m_2$, $x = -1$ it follows that

$$\sum_{m_2=0}^{i_2-t_2} \frac{(-1)^{m_2}}{(i_2-t_2-m_2)!m_2!} = \begin{cases} 0 & \text{if } t_2 < i_2, \\ 1 & \text{if } t_2 = i_2 \end{cases}.$$

Similarly,

$$\sum_{m_1=0}^{i_1-t_1} \frac{(-1)^{m_1}}{(i_1-t_1-m_1)!m_1!} = \begin{cases} 0 & \text{if } t_1 < i_1, \\ 1 & \text{if } t_1 = i_1 \end{cases}.$$

Thus we conclude from (47) that:

$$c_{\mathbf{t};\mathbf{i},k,0} = \begin{cases} 0 & \text{if } t_1 < i_1 \text{ or } t_2 < i_2, \\ 1 & \text{if } t_1 = i_1 \text{ and } t_2 = i_2 \end{cases}. \tag{48}$$

This is equivalent to stating that $c_{\mathbf{t};\mathbf{i},k,0} = 0$ if $|\mathbf{t}| < |\mathbf{i}|$, which is precisely what we wanted to prove for the initial step $l = 0$.

Next suppose the proposition is true for $l \geq 0$ and we will show that it is also true for $l+1$. By the induction hypothesis,

$$c_{\mathbf{t};\mathbf{i},k,l} = 0 \text{ for } |\mathbf{t}| < |\mathbf{i}| - l.$$

Therefore for $\mathbf{t} = (t_1+1, t_2)$ and $\mathbf{t} = (t_1, t_2+1)$, we have

$$c_{(t_1+1,t_2);\mathbf{i},k,l} = 0 \text{ and } c_{(t_1,t_2+1);\mathbf{i},k,l} = 0 \text{ for } |\mathbf{t}| < |\mathbf{i}| - l - 1.$$

Hence, from (40),

$$c_{\mathbf{t},\mathbf{i},k,l+1} = 0 \text{ for } |\mathbf{t}| < |\mathbf{i}| - l - 1.$$

Thus we have proven that (41) is true for $l+1$, which completes the induction argument.

iii) The result (42) has been assumed for $\mathbf{i} = k+l$ in b). For $|\mathbf{i}| < k+l$, we will again use induction on l. When $l = 0$, we have seen in (48) that $c_{\mathbf{t};\mathbf{i},k,0} = 0$ for $\mathbf{t} \neq \mathbf{i}$ and $c_{\mathbf{t};\mathbf{i},k,0} = 1$ for $\mathbf{t} = \mathbf{i}$. Therefore $\sum_{0 \leq \mathbf{t} \leq \mathbf{i}} c_{\mathbf{t};\mathbf{i},k,0} = 1$ as desired.

Next suppose (42) is true for some $l \geq 0$. We want to show $\sum_{0 \leq \mathbf{t} \leq \mathbf{i}} c_{\mathbf{t};\mathbf{i},k,l+1} = 1$. Now using (40), we get

$$\sum_{0 \leq \mathbf{t} \leq \mathbf{i}} c_{\mathbf{t};\mathbf{i},k,l+1} = \sum_{0 \leq \mathbf{t} \leq \mathbf{i}} \frac{(k+l+1) - |\mathbf{t}|}{k+l+1} c_{\mathbf{t};\mathbf{i},k,l} +$$
$$+ \sum_{t_1=0}^{i_1} \sum_{t_2=0}^{i_2} \frac{t_1+1}{k+l+1} c_{(t_1+1,t_2);\mathbf{i},k,l} + \sum_{t_1=0}^{i_1} \sum_{t_2=0}^{i_2} \frac{t_2+1}{k+l+1} c_{(t_1,t_2+1);\mathbf{i},k,l}.$$

Now with the re-indexing:

$$\bar{t}_1 = t_1 + 1 \text{ (for the middle term) and } \bar{t}_2 = t_2 + 1 \text{ (for the last term)}$$

we get

$$\begin{aligned}
\sum_{0\le \mathbf{t}\le \mathbf{i}} c_{\mathbf{t};\mathbf{i},k,l+1} &= \sum_{t_1=0}^{i_1}\sum_{t_2=0}^{i_2} \frac{(k+l+1)-t_1-t_2}{k+l+1} c_{(t_1,t_2);\mathbf{i},k,l} + \\
&\quad + \sum_{\bar{t}_1=1}^{i_1+1}\sum_{t_2=0}^{i_2} \frac{\bar{t}_1}{k+l+1} c_{(\bar{t}_1,t_2);\mathbf{i},k,l} + \sum_{t_1=0}^{i_1}\sum_{\bar{t}_2=1}^{i_2+1} \frac{\bar{t}_2}{k+l+1} c_{(t_1,\bar{t}_2);j,k,l} \\
&= \sum_{t_1=0}^{i_1}\sum_{t_2=0}^{i_2} \frac{(k+l+1)-t_1-t_2}{k+l+1} c_{(t_1,t_2);\mathbf{i},k,l} + \\
&\quad + \sum_{\bar{t}_1=1}^{i_1}\sum_{t_2=0}^{i_2} \frac{\bar{t}_1}{k+l+1} c_{(\bar{t}_1,t_2);\mathbf{i},k,l} + \sum_{t_1=0}^{i_1}\sum_{\bar{t}_2=1}^{i_2} \frac{\bar{t}_2}{k+l+1} c_{(t_1,\bar{t}_2);\mathbf{i},k,l} \\
&\quad + \sum_{t_2=0}^{i_2} \frac{i_1+1}{k+l+1} c_{(i_1+1,t_2);(i_1,i_2),k,l} + \sum_{t_1=0}^{i_1} \frac{i_2+1}{k+l+1} c_{(t_1,i_2+1);(i_1,i_2),k,l}.
\end{aligned}$$

$$\begin{aligned}
&= \sum_{t_1=0}^{i_1}\sum_{t_2=0}^{i_2} c_{(t_1,t_2);\mathbf{i},k,l} - \sum_{t_1=0}^{i_1}\sum_{t_2=0}^{i_2} \frac{t_1}{k+l+1} c_{(t_1,t_2);\mathbf{i},k,l} - \sum_{t_1=0}^{i_1}\sum_{t_2=0}^{i_2} \frac{t_2}{k+l+1} c_{(t_1,t_2);\mathbf{i},k,l} \\
&\quad + \sum_{\bar{t}_1=1}^{i_1}\sum_{t_2=0}^{i_2} \frac{\bar{t}_1}{k+l+1} c_{(\bar{t}_1,t_2);\mathbf{i},k,l} + \sum_{t_1=0}^{i_1}\sum_{\bar{t}_2=1}^{i_2} \frac{\bar{t}_2}{k+l+1} c_{(t_1,\bar{t}_2);\mathbf{i},k,l} \\
&\quad + \sum_{t_2=0}^{i_2} \frac{i_1+1}{k+l+1} c_{(i_1+1,t_2);(i_1,i_2),k,l} + \sum_{t_1=0}^{i_1} \frac{i_2+1}{k+l+1} c_{(t_1,i_2+1);(i_1,i_2),k,l} \\
&= \sum_{0\le \mathbf{t}\le \mathbf{i}} c_{t;\mathbf{i},k,l} + \sum_{t_2=0}^{i_2} \frac{i_1+1}{k+l+1} c_{i_1+1,t_2;(i_1,i_2),k,l} + \sum_{t_1=0}^{i_1} \frac{i_2+1}{k+l+1} c_{t_1,i_2+1;(i_1,i_2),k,l}
\end{aligned} \tag{49}$$

Now from induction hypothesis, $\sum_{0\le \mathbf{t}\le \mathbf{i}} c_{\mathbf{t};\mathbf{i},k,l} = 1$. Also from the definition of $c_{\mathbf{t}}$ (see (39)), we have $c_{i_1+1,t_2;(i_1,i_2),k,l} = 0$ since $i_1 + 1 > i_1$, and $c_{t_1,i_2+1;(i_1,i_2),k,l} = 0$ since $i_2 + 1 > i_2$. Hence from (49), we conclude that

$$\sum_{0\le \mathbf{t}\le \mathbf{i}} c_{\mathbf{t};\mathbf{i},k,l+1} = 1,$$

which completes the induction argument to prove (42). □

Acknowledgement: The second author would like to thank Prof. I. Babuška of the University of Texas at Austin and Prof. J.E. Osborn of the

University of Maryland at College Park, for insightful discussions on the topic addressed in this paper.

References

[1] I. Babuška, U. Banerjee, and J. Osborn. Meshless and generalized finite element method: A survey of major results. In M. Griebel and M. A. Schweitzer, editors, *Meshfree Methods for Partial Differential Equations*, Lecture Notes in Computational Science and Engineering, pages 1–20. Springer, 2002.

[2] I. Babuška, U. Banerjee, and J. Osborn. Survey of meshless and generalized finite element methods. *Acta Numerica*, 12:1–125, 2003.

[3] I. Babuška, U. Banerjee, and J. Osborn. Generalized finite element methods: Main ideas, results, and perspective. *International Journal of Computational Methods*, 1(1):1–37, 2004.

[4] I. Babuška, U. Banerjee, and J. Osborn. On the approximability and the selection of particle shape functions. *Numer. Math.*, 96:601–640, 2004.

[5] I. Babuška, G. Caloz, and J. Osborn. Special finite element methods for a class of second order elliptic problems with rough coefficients. *SIAM J. Numer. Anal.*, 31:945–981, 1994.

[6] I. Babuška and J. M. Melenk. The partition of unity finite element method. *Int. J. Numer. Meth. Engng.*, 40:727–758, 1997.

[7] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer-Verlag, New York, 2007.

[8] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.

[9] J. Dolbow and T. Belytschko. Numerical integration of the Galerkin weak form in meshfree methods. *Computational Mechanics*, 23:219–230, 1999.

[10] W. Han and X. Meng. Error analysis of the reproducing kernel particle method. *Comput. Methods Appl. Mech. Engrg*, 190:6157–6181, 2001.

[11] W. K. Liu, Y. Chen, S. Jun, J. S. Chen, t. Belytschko, C. Pan, R. A. Uras, and C. T. Chang. Overview and applications of reproducing kernal particle methods. *Archives of Computational Methods in Engineering: State of the art reviews*, 3:3–80, 1996.

[12] J. M. Melenk and I. Babuška. The partition of unity finite element method: Theory and application. *Comput. Methods Appl. Mech. Engrg.*, 139:289–314, 1996.

[13] E. M. Stein. *Singular Integrals and Differentiability Properties of Functions.* Princeton Univ. Press, 1970.

[14] G. Strang and G. Fix. A fourier analysis of finite element variational method, in constructive aspects of functional analysis. *Edizioni Cremonese*, pages 795–840, 1973.

[15] T. Strouboulis, I. Babuška, and K. Copps. The design and analysis of the generalized finite element method. *Comput. Methods Appl. Mech. Engrg.*, 181:43–69, 2001.