

# Average Interpolation Under the Maximum Angle Condition

Alexander Rand\*

September 24, 2018

## Abstract

Interpolation error estimates needed in common finite element applications using simplicial meshes typically impose restrictions on the both the smoothness of the interpolated functions and the shape of the simplices. While the simplest theory can be generalized to admit less smooth functions (e.g., functions in  $H^1(\Omega)$  rather than  $H^2(\Omega)$ ) and more general shapes (e.g., the maximum angle condition rather than the minimum angle condition), existing theory does not allow these extensions to be performed simultaneously. By localizing over a well-shaped auxiliary spatial partition, error estimates are established under minimal function smoothness and mesh regularity. This construction is especially important in two cases:  $L^p(\Omega)$  estimates for data in  $W^{1,p}(\Omega)$  hold for meshes without any restrictions on simplex shape, and  $W^{1,p}(\Omega)$  estimates for data in  $W^{2,p}(\Omega)$  hold under a generalization of the maximum angle condition which requires  $p > 2$  for standard Lagrange interpolation.

Interpolation error estimates for the standard finite element spaces typically involve two requirements: conditions on the smoothness of the function being interpolated (*data regularity*) and constraints on the geometry of the elements of the mesh (*shape regularity*). Beyond the simplest interpolation error estimates (stated precisely in Proposition 1.2), data or shape regularity can be substantially relaxed with some additional analysis. However the existing theory does not accept the weakest restrictions on both data and shape regularity simultaneously.

**Data Regularity and Average Interpolation** Lagrange interpolation involves point-wise function evaluation leading to data regularity restrictions associated with the requirements of the Sobolev embedding theorem. Alternatively, average interpolation can be used to handle less regular data [13, 38, 5, 37, 11]. Because average interpolation ‘smears’ local estimates over a neighborhood of simplices, error estimates rely more heavily on the geometry of the mesh, often needing meshes of bounded ply (defined in Section 1.3, see Figure 1) so global summation can be performed.

**Shape Regularity and Angle Conditions** For triangular meshes the classical error estimates are typically proved under a minimum angle restriction, but this condition is overly restrictive. Triangle geometry and interpolation error are actually related through the largest angle of the triangle [43, 6, 22]. (Convergence of the finite element method can occur for meshes with arbitrarily large angles [8, 19], but the known counterexamples are finite element spaces containing *subspaces*

---

\*Institute for Computational Engineering and Sciences, The University of Texas at Austin, 201 East 24th St, Stop C0200, Austin, Texas 78712 ([arand@ices.utexas.edu](mailto:arand@ices.utexas.edu))

associated with a shape regular mesh.) The maximum angle condition can be generalized to higher-dimensional simplices but existing analysis gives stronger data regularity requirements [22, 24, 39]. The linear tetrahedral element is especially important to this discussion. Sherk constructed a counterexample demonstrating that the expected Lagrange interpolation error estimates do not hold for a fairly innocuous family of narrow tetrahedra [39]; see Fig. 3(right). Analysis using an average interpolant eliminates this problem but again leads to new geometric restrictions on the ply of the mesh [1].

**Mesh Generation** Geometric restrictions needed to prove interpolation error estimates are essentially the output requirements for a mesh generator. In 2D many mesh generation techniques aim to produce bounded aspect ratio triangulations (although a notable exception [30] only removes large angles), but aspect ratio guarantees are impossible when the input (i.e., the boundary of the domain) contains small angles [40]. To admit arbitrary input several mesh generation strategies have been developed which allow a few poor quality triangles near small input angles [29, 33, 40, 36]. Guided by the interpolation theory the most commonly used strategy [29] provides provable upper bounds on the largest angle and the ply of the mesh (as long as the input does not include high degree vertices). This effective approach does not extend to 3D and guaranteed mesh generation algorithms do not yield a bound on the ply [25, 42, 41, 10, 35, 36]. Figure 2(left) demonstrates the high ply construction in the analogous 2D algorithm. Failures of these algorithms are difficult to observe in practice since the mesh ply only grows logarithmically in the (inverse of the) size of the smallest angle, but they represent a theoretical disconnect between output guarantees of the mesh generator and the input requirements of numerical methods. Beyond provably-correct mesh generation algorithms, simplified mesh generation algorithms are often practically successful, and allowing high ply meshes extends the range of inputs producing acceptable results. Uniform refinement of a structured grid in polar coordinates is perhaps the simplest situation yielding a high ply node; see Figure 2(right). Providing the loosest set of geometric requirements in the interpolation theory gives mesh generators the most flexibility in practical applications as well as new challenges to improve existing guarantees under weaker requirements.

**Outline** This paper demonstrates that there is no inherent trade-off between data regularity and shape regularity in interpolation of less regular Sobolev functions on simplicial meshes by developing a theory accepting minimal data and shape regularity requirements. The key insight is that error estimates cannot be localized over the meshes of high ply; rather a simpler auxiliary spatial partition must be used. Before describing this construction, Section 1 contains the necessary preliminary discussion including a more formal exposition of the prior results. Error estimates for the average interpolant under uniform size and quality meshes are shown in Section 2. A generalization to non-uniform meshes is given in Section 3 before some concluding remarks in Section 4.

## 1 Preliminaries

The statement  $x \lesssim y$  is used to mean there exists a positive constant  $C$  such that  $x \leq C y$  when the details of the constant are unimportant. For nearly all estimates discussed, this implied constant will be independent of simplex/mesh shape regularity; whenever this is not the case, it will be explicitly noted.

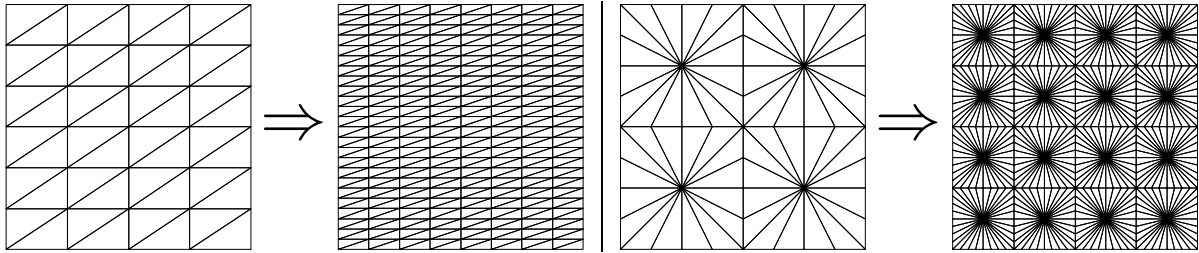


Figure 1: Two refinement strategies that do not produce bounded aspect ratio meshes. (left) A refinement strategy creating a mesh with bounded ply and maximum angle: all vertices have degree six and all angles are acute. (right) A refinement strategy creating a high ply mesh without large angles: all angles are smaller than 135 degrees: in the limiting case, the triangles in the corners contain angles very near  $135^\circ$ ,  $45^\circ$  and  $0^\circ$ .

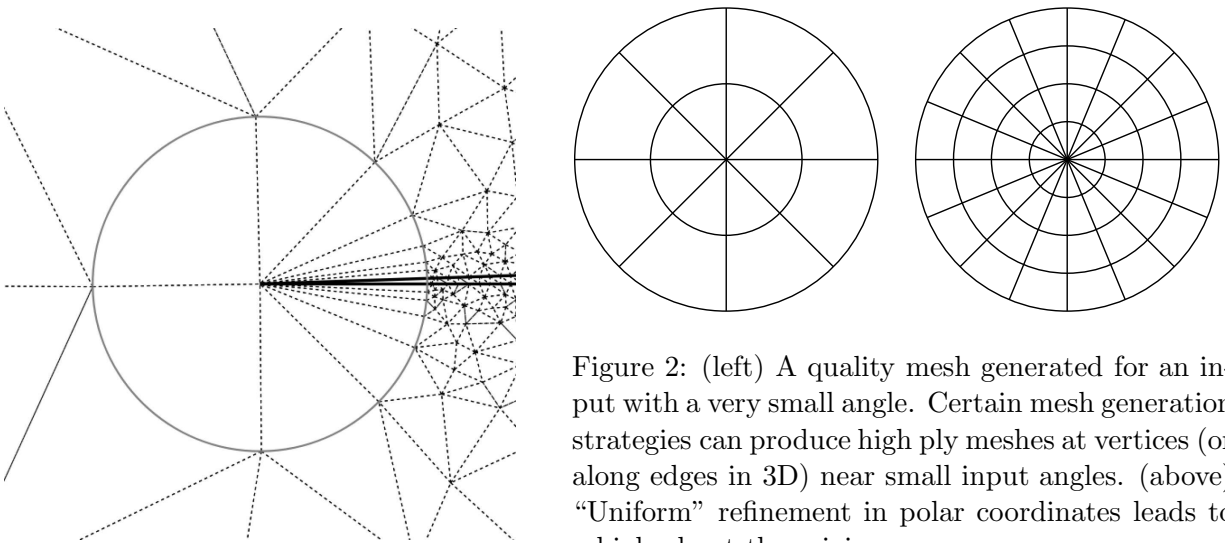


Figure 2: (left) A quality mesh generated for an input with a very small angle. Certain mesh generation strategies can produce high ply meshes at vertices (or along edges in 3D) near small input angles. (above) “Uniform” refinement in polar coordinates leads to a high ply at the origin.

## 1.1 Balls, Cubes, and Simplices

The *ball* of radius  $r$  centered at  $\mathbf{x} \in \mathbb{R}^d$  is the set of points with Euclidean distance from  $\mathbf{x}$  less than  $r$ :

$$B(\mathbf{x}, r) := \{\mathbf{y} : |\mathbf{x} - \mathbf{y}| < r\}.$$

A  $d$ -cube or *cube* is a closed set  $Q \subset \mathbb{R}^d$  of the form

$$Q = \left[ a_1 - \frac{h}{2}, a_1 + \frac{h}{2} \right] \times \left[ a_2 - \frac{h}{2}, a_2 + \frac{h}{2} \right] \times \dots \times \left[ a_d - \frac{h}{2}, a_d + \frac{h}{2} \right].$$

The point  $\mathbf{a} = (a_1, \dots, a_d)$  is called the *center* of  $Q$  and  $h$  is called the *size* of  $Q$ , denoted  $\text{size}(Q)$ . A *simplex* is the convex hull of  $d + 1$  non-coplanar points (called *vertices*),  $\{\mathbf{v}_i\}_{i=0}^d$ :

$$K = \left\{ \sum_{i=0}^d t_i \mathbf{v}_i : t_i > 0, \sum_{i=0}^d t_i = 1 \right\}.$$

We let  $h_K$  denote the length of the longest side of  $K$  and  $\rho_K$  denote the radius of the largest sphere inscribed in  $K$ . Following Jamet's shape regularity condition [22], the angle  $\theta_K$  is defined to characterize the quality of a simplex  $K$  by the formula,

$$\theta_K := \max_{|\boldsymbol{\xi}|=1} \min_i \arccos \boldsymbol{\xi} \cdot \mathbf{u}_i \tag{1}$$

where  $\{\mathbf{u}_i\}$  is the set of unit vectors parallel to the edges of the simplex  $K$ ; see Fig. 3. This angle measures how far from coplanar the set of edges is, so we call  $\theta_K$  the *coplanarity* measure. The subscript on  $h$ ,  $\rho$ , and  $\theta$  is often omitted when the simplex in question is apparent. In two dimensions  $\theta_K$  is half the largest angle of the triangle, and thus the set of triangles with coplanarity bounded away from  $\pi/2$  satisfy the maximum angle condition in the sense that all angles are bounded away from  $\pi$ . In three dimensions, a more intuitive generalization of the maximum angle condition defined by defined by Křížek [24] requires that both face and dihedral angles of the tetrahedron are bounded away from  $\pi$ . This is also equivalent to bounding coplanarity away from  $\pi/2$  as shown in the proposition below.

**Proposition 1.1.** *Let  $\psi_K$  denote the maximum of all face and dihedral angles of (three-dimensional) tetrahedron  $K$ . For tetrahedra, bounds on coplanarity and face/dihedral angles are equivalent, i.e., if  $\theta_K$  is bounded away from  $\pi/2$  then  $\psi_K$  is bounded away from  $\pi$  and vice versa.*

The proof, given in Appendix A, relies on a characterization of tetrahedra satisfying the maximum angle condition given in [2].

## 1.2 Sobolev Spaces

We will distinguish between the notation  $\frac{\partial}{\partial x_i}$  and  $\partial_i$ . The former will be used to denote differentiation with respect to a particular variable while the latter represents differentiation with respect to a function argument. For example, the following notation is used for an application of the chain rule:

$$\frac{\partial}{\partial x_i} [u(2\mathbf{x} + 3\mathbf{y})] = 2\partial_i u(2\mathbf{x} + 3\mathbf{y}).$$

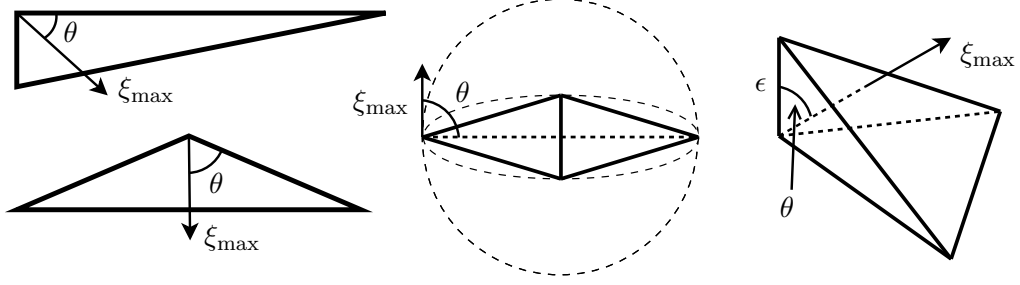


Figure 3: Coplanarity  $\theta$  for several simplices with vector  $\xi_{\max}$  which maximizes the expression in (1) labeled. (left) In 2D the coplanarity is half the largest angle. (center) The coplanarity of a nearly flat “sliver” tetrahedron is nearly  $\pi/2$ . (right) The simplices used in a counter-example due to Shenk [39] have coplanarity bounded away from  $\pi/2$  even as the edge length  $\epsilon$  approaches 0.

The Sobolev semi-norms and norms on functions over an open set  $\Omega$  are defined by

$$|u|_{W^{m,p}(\Omega)}^p := \int_{\Omega} \sum_{|\beta|=m} |\partial^{\beta} u(\mathbf{x})|^p dx \quad \text{and} \quad \|u\|_{W^{m,p}(\Omega)}^p := \sum_{0 \leq k \leq m} |u|_{W^{k,p}(\Omega)}^p,$$

where the first summation is taken over multi-indices  $\beta$  and differentiation with respect to multi-indices has the standard definition,  $\partial^{\beta} u = \partial^{\beta_1} \partial^{\beta_2} \dots u$ . The  $W^{0,p}$ -norm is simply the standard  $L^p$ -norm.

Throughout this paper,  $\Omega \subset \mathbb{R}^d$  is a bounded set and an extension domain for any Sobolev spaces used; i.e. there is a continuous, linear operator  $E_{\Omega} : W^{m,p}(\Omega) \rightarrow W^{m,p}(\mathbb{R}^d)$  such that  $u(x) = E_{\Omega} u(x)$  for all  $x \in \Omega$ ; see [3, p. 83] for technical results concerning extension domains. Without ambiguity, the functions  $u$  and  $E_{\Omega} u$  will not be distinguished: for  $x \notin \Omega$ ,  $u(x)$  denotes  $E_{\Omega} u(x)$ . Likewise, the estimate  $|u|_{W^{m,p}(\mathbb{R}^d)} \lesssim |u|_{W^{m,p}(\Omega)}$  is used freely.

### 1.3 Lagrange Interpolation on Simplicial Meshes

$\mathcal{T}$  always denotes a simplicial mesh of  $\Omega$ . The *ply* of the mesh  $\mathcal{T}$  is the maximum number of mesh simplices intersecting a single simplex. Ply restrictions are often implicit in interpolation error estimates since a mesh with bounded aspect ratio has a bounded ply. (In fact, the weaker restriction of bounded circumradius to shortest edge ratio is sufficient to ensure bounded ply [44, 21].)

Let  $\Pi_k$  denote the  $k$ -th degree Lagrange interpolant on a simplicial mesh,

$$\Pi_k u(\mathbf{x}) := \sum_i u(\mathbf{v}_i) \phi_i(\mathbf{x})$$

where  $\phi$  are the piecewise nodal basis functions on  $\mathcal{T}$ . While this paper only considers linear interpolants ( $k = 1$ ), the higher order cases are included in the statement of classical estimates for smooth functions.

### 1.4 Prior Interpolation Error Estimates

The Bramble-Hilbert Lemma [7, 14] along with scaling properties of Sobolev semi-norms leads to the following classical error estimate for the Lagrange interpolant; see [12, 15, 9] for complete details.

**Proposition 1.2.** *If  $k + 1 > \frac{d}{p}$  and  $0 \leq m \leq k$ , then for all simplices  $K$  and all  $u \in W^{k+1,p}(K)$ ,*

$$|u - \Pi_k u|_{W^{m,p}(K)} \lesssim \frac{h^{k+1}}{\rho^m} |u|_{W^{k+1,p}(K)}. \quad (2)$$

For simplices with bounded aspect ratio,  $h$  and  $\rho$  are proportional and thus the factor  $\frac{h^{k+1}}{\rho^m}$  reduces to  $h^{k+1-m}$ . The estimate in the  $L^2$ -norm (i.e., the  $m = 0$  case) is particularly nice: the term in the denominator disappears and the estimate is independent of the shape of the simplex.

The first generalization involves weakening the data regularity requirement, i.e., the restriction  $u \in W^{2,p}(\Omega)$ . A function  $u \in W^{1,p}(\Omega)$  (but not in  $W^{2,p}(\Omega)$ ) is not necessarily continuous so an alternative to Lagrange interpolation must be considered. Clément first described such an interpolant and proved an estimate which corresponds to the ( $k = 0$ )-case of (2) [13]. Scott and Zhang proved a similar estimate based on an average interpolant [38]: specifically interpolant values at mesh vertices are defined to be averages over an arbitrarily selected adjacent simplex. A distinct advantage of the Scott and Zhang interpolant is that coefficients of basis functions corresponding to boundary nodes only depend upon the boundary data. An alternative approach based on averaging over balls has been developed which extends to vector-valued functions in the discrete de Rham complex [5, 37, 11]; see Figure 4 for a comparison of the averaging regions. The resulting interpolant of Christiansen and Winther also matches homogeneous boundary data selecting an appropriate extension of the function outside the domain. In the simplest case, each of these constructions reduces to the following theorem.

**Theorem 1.3** (Specialized from [38, 11], etc.). *Suppose  $\mathcal{T}$  has a bounded aspect ratio and let  $\Pi_0$  be an average interpolant. Then for all  $u \in W^{1,p}(\Omega)$ ,*

$$\|u - \Psi_0 u\|_{L^p(K)} \lesssim h |u|_{W^{1,p}(\hat{K})}, \quad (3)$$

where  $\hat{K}$  is the union of  $K$  and its neighboring simplices.

Unlike the ( $m = 0, k = 1$ )-case of Proposition 1.2, Theorem 1.3 requires a bounded aspect ratio mesh ensuring two things. First,  $K$  and its neighboring simplices have roughly the same size so that  $\text{diam}(K) \approx \text{diam}(\hat{K})$ . Second,  $K$  has a bounded ply so these local estimates (3) over each simplex can be summed to a global estimate. Our aim is to remove these restrictions on simplex shape to match the  $m = 0$  case of the classical estimate (2) which is independent of simplex shape.

The second generalization of Proposition 1.2 weakens the shape regularity requirements. For this purpose, Jamet developed an improved estimate for sufficiently regular data.

**Theorem 1.4** (Jamet, [22]). *If  $k + 1 - m > \frac{d}{p}$ , then for all  $u \in W^{k+1,p}(K)$ ,*

$$|u - \Pi_k u|_{W^{m,p}(K)} \lesssim \frac{h^{k+1-m}}{(\cos \theta)^m} |u|_{W^{k+1,p}(K)}. \quad (4)$$

The restriction  $k + 1 - m > \frac{d}{p}$  prevents Jamet's result from applying in certain cases including (perhaps the most important cases of) linear interpolation on triangles and tetrahedra:  $k = m = 1$  when  $d = 2$  or  $d = 3$ . The analysis of Babuška and Aziz [6] yields the result for the linear triangular element (although care must be taken to derive the correct geometric factor  $1/\cos \theta$  using this approach [23, 18, 28, 34]). Shenk generalized the approach of Babuška and Aziz to

higher dimensions and, more importantly, demonstrated that (4) does not hold for the linear Lagrange interpolant on tetrahedra [39]. Shenk’s example involves particularly simple geometry: a family of tetrahedra depicted in Figure 3 with vertices  $(0, 0, 0)$ ,  $(1, 0, 0)$ ,  $(0, 1, 0)$ , and  $(0, 0, \epsilon)$  where  $\epsilon$  approaches zero. Demonstrating the error estimate for an analogous class of triangles (i.e., 2D simplices) is the crux of the argument of Babuška and Aziz.

An analysis by Acosta demonstrated that expected error estimates can be extended to certain narrow tetrahedra (including those in Shenk’s example) by using average (rather than Lagrange) interpolation [1]. This method does allow data and shape regularity to be weakened simultaneously, but the analysis is inherently anisotropic: locally it must be possible to affinely transform the mesh into a shape-regular mesh. (This is a standard requirement in an anisotropic framework; e.g., [4, 17, 20].) The resulting estimates expect the mesh to have a bounded ply so that local estimates can be summed to form global error estimates. Note that the meshes in Figure 2 do not satisfy the requirements of [1].

We will develop interpolation error estimates based on average interpolation under weakened data and shape regularity requirements. Emphasis will be placed on avoiding any unnecessary element shape requirements which, most notably, allows meshes with unbounded ply (as is the case for Lagrange interpolants in Theorem 1.4). The result will be an interpolation theory which extends Jamet’s minimal shape regularity requirements to less smooth data including the linear tetrahedral element.

## 2 Uniform Error Estimates

To avoid the limitations associated with pointwise function evaluation, an average interpolant is defined by mollifying the data before Lagrange interpolation:

$$\Xi_h u(\mathbf{x}) := \Pi_1 \mathcal{M}_h u(\mathbf{x}). \quad (5)$$

In the next subsection, the standard mollification procedure  $\mathcal{M}_h$  is defined and a number of needed properties are given.

*Remark 2.A.* This interpolant is used in [5, 37, 11], but in that context, the mollification radius was selected to be small enough so that the averaging region stays within the neighboring triangles. Since we do not assume aspect ratio bounds, we cannot use this property. Acosta uses essentially an anisotropic variant of this construction [1]; see Figure 4.

### 2.1 Mollification and Integration

Mollification is a common tool in analysis for producing smooth approximations of functions in  $L^p$  spaces. After defining the procedure, a few common facts about mollification are listed. The proofs of these and similar results can be found in a number of textbooks, but are given in the appendix for comparison to the later sections. For some examples of these kinds of technical details, especially those for more regular data and including explicit dependence on the mollification radius, see [32, p. 191], [26, p. 58], [27, p. 98], and [16, p. 201].

Let  $\rho \in C^\infty(\mathbb{R}^d)$  such that (i)  $\text{supp}(\rho) \subset \overline{B(0, 1)}$ , (ii)  $\int \rho(\mathbf{x}) \, d\mathbf{x} = 1$ , and (iii)  $\rho(\mathbf{x}) \geq 0$  for all  $\mathbf{x}$ . Let  $\rho_h(\mathbf{x}) = \frac{1}{h^d} \rho(\frac{\mathbf{x}}{h})$ . Then, (i)  $\text{supp}(\rho_h) \subset \overline{B(0, h)}$ , (ii)  $\int \rho_h(\mathbf{x}) \, d\mathbf{x} = 1$ , (iii)  $\rho_h(\mathbf{x}) \geq 0$  for all  $\mathbf{x}$ , and (iv)  $\|\nabla \rho_h\|_{L^1(\mathbb{R}^d)} \leq \frac{1}{h} \|\nabla \rho\|_{L^1(\mathbb{R}^d)}$ . The most commonly used mollifier is  $\rho(\mathbf{x}) = ce^{-1/(1-|\mathbf{x}|^2)}$  where the constant  $c$  is chosen so (ii) holds. We also require that  $\rho$  is *radially symmetric*. While

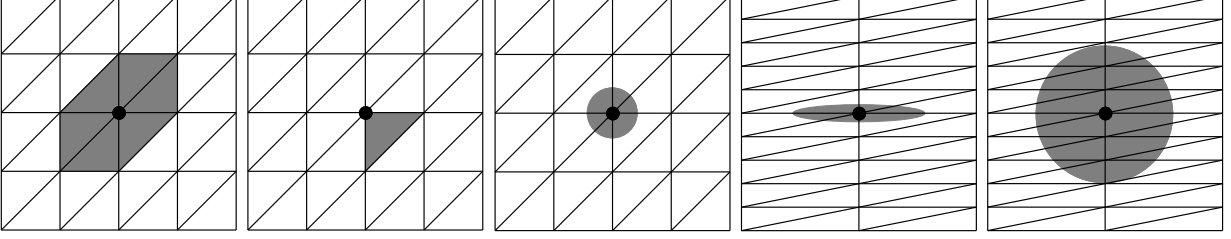


Figure 4: Averaging regions for the interpolants of (left to right) Clément, Scott and Zhang, Christiansen and Winther, Acosta, and this paper. The first three were developed for shape regular meshes. The final two are shown on an anisotropic mesh to highlight the differences in the construction. While the Acosta interpolant requires well-defined anisotropy in the mesh, our interpolant makes minimal assumptions on the mesh quality.

symmetry is not usually required, the common mollifier is symmetric and we explicitly use the property in one step of the upcoming arguments.

A function is mollified by convolving it with  $\rho_h$ : given  $u \in L^1(\mathbb{R}^d)$ , define

$$\mathcal{M}_h u(\mathbf{x}) := \int_{\mathbb{R}^d} u(\mathbf{x} - \mathbf{y}) \rho_h(\mathbf{y}) \, d\mathbf{y} = \int_{\mathbb{R}^d} u(\mathbf{x} - h\mathbf{z}) \rho(\mathbf{z}) \, d\mathbf{z}.$$

The two representations above are equivalent, resulting from a simple change of variables,  $\mathbf{y} = h\mathbf{z}$ , but the supports of the integrands are different. If a function is smooth enough (specifically, it belongs to  $W^{1,p}(\Omega)$  in the proposition below) then the difference between the function and its mollification can be bounded in terms of the mollification radius.

**Proposition 2.1.** *If  $p \geq 1$ , then for all  $u \in W^{1,p}(\Omega)$ ,*

$$\|u - \mathcal{M}_h u\|_{L^p(\mathbb{R}^d)} \lesssim h |u|_{W^{1,p}(\Omega)}.$$

The next proposition demonstrates how the smoothness of the mollifier depends on the mollification radius through a bound on the  $L^p$ -norm of a mollified function in terms of the  $L^q$ -norm of the original function.

**Proposition 2.2.** *If  $1 \leq q \leq p \leq \infty$  and  $Q$  is a bounded set, then for all  $u \in L^q(Q)$ ,*

$$\|\mathcal{M}_h u\|_{L^p(Q)} \lesssim h^{\frac{d(q-p)}{pq}} \|u\|_{L^q(Q_h)} \quad (6)$$

where  $Q_h := \bigcup_{x \in Q} B(x, h)$ .

*Remark 2.B.* Since  $\frac{\partial}{\partial x_i} (\mathcal{M}_h u) = \mathcal{M}_h \left( \frac{\partial u}{\partial x_i} \right)$ , Proposition 2.2 also applies for higher Sobolev norms; specifically, (6) implies that

$$|\mathcal{M}_h u|_{W^{k,p}(Q)} \lesssim h^{\frac{d(q-p)}{pq}} |u|_{W^{k,q}(Q_h)}. \quad (7)$$

Finally a standard estimate of the  $L^p$ -norm of a function in terms of the  $L^q$ -norm is stated for  $q$  larger than  $p$ . This can be proved with either Jensen's or Hölder's inequality.

**Proposition 2.3.** *If  $1 \leq p \leq q \leq \infty$  and  $Q$  is a bounded set, then for all  $u \in L^q(Q)$ ,*

$$\|u\|_{L^p(Q)} \leq |Q|^{\frac{q-p}{pq}} \|u\|_{L^q(Q)}.$$



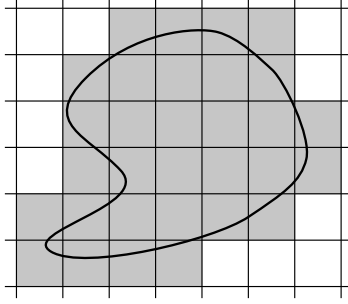


Figure 5: For an example domain, the cubes in  $\mathcal{Q}$  are shaded.

## 2.2 Error Estimate

Next we turn to removing the integrability condition in Jamet's estimate (4). Again for simplicity we first consider meshes in which simplex size and quality (measured by coplanarity, not aspect ratio) are uniformly bounded and then relax these assumptions in Section 3.

**Theorem 2.4.** *Suppose  $\mathcal{T}$  contain no edges longer than  $h$  and let  $\theta$  be the maximum simplex coplanarity. Then for  $k \in \{0, 1\}$  and for all  $u \in W^{k+1,p}(\Omega)$ ,*

$$|u - \Xi_h u|_{W^{k,p}(\Omega)} \lesssim \frac{h}{(\cos \theta)^k} |u|_{W^{k+1,p}(\Omega)}. \quad (8)$$

*Proof.* First the triangle inequality is applied using the intermediate function  $\mathcal{M}_h u$ ,

$$|u - \Xi_h u|_{W^{k,p}(\Omega)} \leq |u - \mathcal{M}_h u|_{W^{k,p}(\Omega)} + |\mathcal{M}_h u - \Xi_h u|_{W^{k,p}(\Omega)}. \quad (9)$$

The first term of (9) is independent of the triangulation  $\mathcal{T}$  and is estimated using Proposition 2.1:

$$|u - \mathcal{M}_h u|_{W^{k,p}(\Omega)} \lesssim h |u|_{W^{k+1,p}(\Omega)}. \quad (10)$$

Next the second term of (9) is addressed:

$$|\mathcal{M}_h u - \Xi_h u|_{W^{k,p}(\Omega)} = |\mathcal{M}_h u - \Pi_1 \mathcal{M}_h u|_{W^{k,p}(\Omega)}. \quad (11)$$

The estimate is established locally, but not over the individual simplices of  $\mathcal{T}$ . This is important since averaging 'smears' the estimate on one simplex over its neighbors. Since we allow for high ply meshes, it is not possible to sum these local estimates to yield the global result (8). Rather, the estimate is localized over a simple artificial (and bounded-ply) decomposition, a covering of  $\Omega$  by cubes.

In a lattice of cubes of side-length  $h$ , let  $\mathcal{Q} = \{Q_i\}$  be the subset which intersect  $\Omega$ ; see Figure 5. The term (11) is estimated on a single cube  $Q_i \in \mathcal{Q}$ . Let  $R_i$  and  $S_i$  denote the cubes of side-length  $3h$  and  $5h$ , respectively, centered around  $Q_i$ , and let  $\mathcal{K}_i$  be the set of simplices in the triangulation  $\mathcal{T}_h$  which intersect  $Q_i$ ; see Figure 6.

Next a value  $q > d$  is selected arbitrarily with the intention of applying Theorem 1.4 (since we are in the setting  $k + 1 - m = 1$ ). Applying Proposition 2.3 followed by the definition of  $\mathcal{K}_i$  yields,

$$\begin{aligned} |\mathcal{M}_h u - \Pi_1 \mathcal{M}_h u|_{W^{k,p}(Q_i)}^q &\lesssim |Q_i|^{\frac{q-p}{p}} |\mathcal{M}_h u - \Pi_1 \mathcal{M}_h u|_{W^{k,q}(Q_i)}^q \\ &\lesssim |Q_i|^{\frac{q-p}{p}} \sum_{K \in \mathcal{K}_i} |\mathcal{M}_h u - \Pi_1 \mathcal{M}_h u|_{W^{k,q}(K)}^q. \end{aligned}$$

Theorem 1.4 can now be applied to each term of the summation on the right hand side and Proposition 2.2 is used to return the estimate to the correct norm:

$$\begin{aligned} |\mathcal{M}_h u - \Pi_1 \mathcal{M}_h u|_{W^{k,p}(Q_i)}^q &\lesssim |Q_i|^{\frac{q-p}{p}} \sum_{K \in \mathcal{K}_i} \left( \frac{h}{(\cos \theta_K)^k} \right)^q |\mathcal{M}_h u|_{W^{k+1,q}(K)}^q \\ &\lesssim |Q_i|^{\frac{q-p}{p}} |h|^{d \frac{p-q}{p}} \left( \frac{h}{(\cos \theta)^k} \right)^q |u|_{W^{k+1,p}(S_i)}^q. \end{aligned}$$

Since  $|Q_i| = h^d$ , we conclude that

$$|\mathcal{M}_h u - \Pi_1 \mathcal{M}_h u|_{W^{k,p}(Q_i)} \lesssim \frac{h}{(\cos \theta)^k} |u|_{W^{k+1,p}(S_i)}. \quad (12)$$

Finally these estimates can be summed to get the global estimate:

$$\begin{aligned} |\mathcal{M}_h u - \Pi_1 \mathcal{M}_h u|_{W^{k,p}(\Omega)}^p &\leq \sum_i |\mathcal{M}_h u - \Pi_1 \mathcal{M}_h u|_{W^{k,p}(Q_i)}^p \\ &\lesssim \left( \frac{h}{(\cos \theta)^k} \right)^p \sum_i |u|_{W^{k+1,p}(S_i)}^p \lesssim \left( \frac{h}{(\cos \theta)^k} \right)^p |u|_{W^{k+1,p}(\Omega)}^p. \end{aligned} \quad (13)$$

The final inequality holds since each cube  $Q_i$  belongs to at most  $5^d$  cubes  $S_j$ .  $\square$

*Remark 2.C.* The essential difference between this analysis and that of Acosta [1] is the use of the artificial (and bounded-ply) mesh  $\mathcal{Q}$  for establishing the local estimate (12).

*Remark 2.D.* Looking closely at the inequality (13) reveals that the extension domain property has been used to estimate  $|u|_{W^{2,p}(\cup S_i)}$  by  $|u|_{W^{2,p}(\Omega)}$ .

*Remark 2.E.* In the  $L^p$  case of (8), i.e,  $k = 0$ , the estimate does not depend on simplex shape which agrees with the classical result for more regular data.

### 3 Locally Quasi-Uniform Error Estimates

Extending Theorem 2.4 to admit non-uniform meshes requires a precise characterization of the mesh size  $h$  at each point in the domain. We emphasize that the size function does not necessarily reflect the size of the triangles in the mesh. Rather it is only an upper bound. Because the average interpolant depends on a local neighborhood, the sizing function should not vary too rapidly so that there is roughly a unique size (up to a constant factor) associated with the neighborhood.

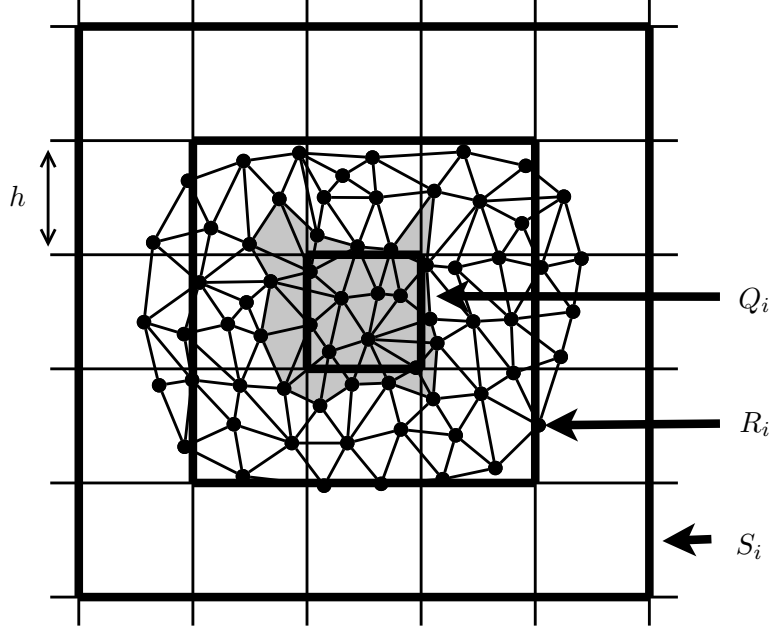


Figure 6: Diagram for the proof of Theorem 2.4. Triangles belonging to  $\mathcal{K}_i$  are shaded.

Before proving error estimates, we outline precise restrictions on these sizing functions. Then based on the sizing function, variable-radius mollification and compatible cubic coverings are defined. As before, the final interpolation operator is a composition of the (modified) mollification operation  $\overline{\mathcal{M}}_h$  with the usual Lagrange interpolation operator  $\Pi_1$ :

$$\overline{\Xi}_h u(\mathbf{x}) := \Pi_1 \overline{\mathcal{M}}_h u(\mathbf{x}). \quad (14)$$

### 3.1 Sizing Function

Using Lagrange interpolation for smooth (enough) functions, mesh size is captured locally by the length of the longest edge of a given triangle. For average interpolation, function values over neighboring triangles must be considered and thus the size of the averaging region also plays a role in the error estimate. On bounded aspect ratio meshes this is not an issue since the longest edges of any pair of neighboring triangles are within a constant factor [31]. To analyze average interpolants for meshes without aspect ratio bounds, we decouple the mesh sizing function from the actual size of the triangles. Rather, the sizing function will be a smooth (enough) function which provides an upper bound on the size of nearby triangles in the mesh. The result will be a sizing function which can be used in the error estimate despite admitting meshes with adjacent triangles with dramatically different sizes.

Letting  $C_{\text{sizing}}$  be a positive constant, we consider the class of sizing functions  $h : \mathbb{R}^d \rightarrow (0, \infty)$  with bounded first and second derivatives:

$$|\nabla h(\mathbf{x})| \leq \frac{1}{16\sqrt{d}}; \quad (15) \quad |\partial_i \partial_j h(\mathbf{x})| \leq C_{\text{sizing}}. \quad (16)$$

We call a triangulation  $\mathcal{T}$  of  $\Omega$  *compatible with sizing function*  $h$  if for all  $\mathbf{x} \in \Omega$ ,  $h(\mathbf{x}) \geq h_{T(\mathbf{x})}$  where

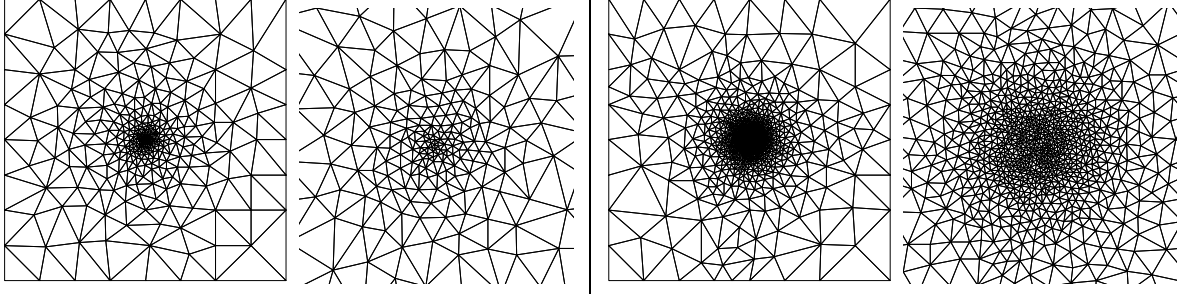


Figure 7: Comparison of two (bounded aspect ratio) meshes associated with sizing functions  $h(\mathbf{x}) = \frac{1}{2}|\mathbf{x}| + 0.01$  (left) and  $h(\mathbf{x}) = |\mathbf{x}|^2 + 0.01$  (right) over the square  $[-1, 1]^2$ . In each case, both the full mesh and the center region zoomed four times are shown. These meshes demonstrate the differences between conditions (15) and (16): meshes only satisfying the former condition can grade more quickly away from zero. Note: the meshes do not satisfy (15) with the particular choice of constant, but this exaggerated example is given to more clearly display the effect.

$T(\mathbf{x})$  denotes the triangle containing  $\mathbf{x}$ .<sup>1</sup> In this setting, we will establish uniform interpolation estimates for any sizing function with a compatible mesh as long as the sizing function satisfies (15) and (16) and the mesh satisfies the maximum angle condition (in the sense of coplanarity).

*Remark 3.A.* The particular choice of constant in (15) is not essential. However, this value is very convenient for localizing future estimates to one layer of neighboring cubes in a 2:1 balanced cubic partition described in Subsection 3.3. Selecting a larger constant in (15) will require analysis of more levels of the cubic partition.

*Remark 3.B.* Restriction (15) is implicitly present (although the constant is often weaker) in the context of shape regular meshes. For a shape-regular mesh  $\mathcal{T}$  and letting  $P(T)$  be the union of triangles in  $\mathcal{T}$  which intersect triangle  $T$ , we define

$$h(\mathbf{v}_i) = \max_{T \in \mathcal{T}, \mathbf{v}_i \in P(T)} h_T,$$

This definition of  $h$  can be extended to  $\Omega$  using linear interpolation over the triangulation and it immediately follows that the triangulation  $\mathcal{T}$  is compatible with  $h$ . Moreover, properties of well-spaced point sets ensure that  $|\nabla h|$  can be bounded depending only on the mesh shape regularity [31, 44, 21]. In other words, a shape regular mesh can be viewed as a mesh that is compatible with a Lipschitz sizing function, and thus (16) is the only *extra* condition on the sizing function we require to handle less shape-regular meshes.

*Remark 3.C.* Restrictions (15) and (16) do allow for non-uniform meshes. The sequence of sizing functions  $h_n(\mathbf{x}) = |\mathbf{x}|^2 + \frac{1}{2^n}$  can be paired with a sequence of meshes with progressively smaller elements near the origin. However, (16) does pose some additional limitations on how *quickly* the sizing function can approach zero: in the shape regular case common adaptive refinement matches a sizing function  $h_n(\mathbf{x}) = |\mathbf{x}| + \frac{1}{2^n}$  or  $h_n(\mathbf{x}) = \min\{|\mathbf{x}|, \frac{1}{2^n}\}$  which is not smooth enough for (16). See Figure 7.

*Remark 3.D.* Restrictions (15) and (16) are somewhat weaker than those used in [1] where there are restrictions essentially of the form  $h(\mathbf{x}) \leq \nabla h(\mathbf{x})$ . In that setting Gronwall's inequality prevents

<sup>1</sup>Without causing confusion this is a slight abuse of notation using  $h(\mathbf{x})$  to denote the sizing function and  $h_T$  to denote the length of the longest edge of triangle. This is done to emphasize the fact that the sizing function plays the role of the mesh size in the upcoming estimates.

such sizing functions from being truly non-uniform. Using the techniques of this paper, it should be possible to recover Acosta's anisotropic results in a similar relaxed setting.

### 3.2 Variable-Radius Mollification

We now consider the generalization of mollification by allowing the mollification radius to vary over the domain. Given  $u \in L^1(\mathbb{R}^d)$  and sizing function  $h$ , define

$$\overline{\mathcal{M}}_h u(\mathbf{x}) := \int_{\mathbb{R}^d} u(\mathbf{x} - \mathbf{y}) \rho_{h(\mathbf{x})}(\mathbf{y}) \, d\mathbf{y} = \int_{\mathbb{R}^d} u(\mathbf{x} + h(\mathbf{x})\mathbf{y}) \rho(\mathbf{y}) \, d\mathbf{y}.$$

The only difference between this definition and the standard mollifier is that the radius  $h(\mathbf{x})$  is spatially varying. We consider an isotropic mollification region (and thus produce isotropic error estimates). The prior construction of Acosta uses an anisotropic mollifier, i.e., the term  $(h_1(\mathbf{x})y_1, h_2(\mathbf{x})y_2, h_3(\mathbf{x})y_3)$  replaces  $h(\mathbf{x})\mathbf{y}$  in the above definition [1]. First we note a simple property of the variable radius mollifier.

**Proposition 3.1.** *Let  $p \geq 1$  and  $Q \subset \mathbb{R}^d$ , and define  $Q_h := \bigcup_{\mathbf{x} \in Q} B(\mathbf{x}, h(\mathbf{x}))$ . Then for all the sizing functions satisfying (15) and for all  $u \in L^p(Q_h)$ ,*

$$\int_Q |u(\mathbf{x} + h(\mathbf{x})\mathbf{y})|^p \, d\mathbf{x} \lesssim \|u\|_{L^p(Q_h)}^p. \quad (17)$$

*Proof.* This results from a simple change of variables,  $\mathbf{z} = \mathbf{x} + h(\mathbf{x})\mathbf{y}$ . The Jacobian of this transformation can be bounded by  $\frac{1}{(1 + \nabla h(\mathbf{x}) \cdot \mathbf{y})^d}$ . By assumption (15) and since  $\mathbf{y} \in B(\mathbf{0}, 1)$ ,  $1 + \nabla h(\mathbf{x}) \cdot \mathbf{y} \in [1 - 1/(16\sqrt{d}), 1 + 1/(16\sqrt{d})]$  producing (17).  $\square$

Extensions of Propositions 2.1 and 2.2 are necessary to prove the non-uniform error estimates. Since Remark 2.B no longer applies, i.e.,

$$\frac{\partial}{\partial x_i} (\overline{\mathcal{M}}_h u) \neq \overline{\mathcal{M}}_h \left( \frac{\partial}{\partial x_i} u \right),$$

we prove these lemmas directly in the correct Sobolev norms.

**Lemma 3.2.** *Let  $p \geq 1$  and  $Q$  be a cube. Then for  $k \in \{0, 1\}$  and for all  $u \in W^{k,p}(\Omega)$ ,*

$$|u - \overline{\mathcal{M}}_h u|_{W^{k,p}(Q)} \lesssim H_Q \|u\|_{W^{k+1,p}(Q_h)} \quad (18)$$

where  $H_Q = \max_{\mathbf{x} \in Q} h(\mathbf{x})$ , and  $Q_h = \bigcup_{\mathbf{x} \in Q} B(\mathbf{x}, h(\mathbf{x}))$ .

*Proof.* The  $k = 0$  case of inequality (18) follows from a very similar argument to the proof of Proposition 2.1. Letting  $w \in L^{p'}(\mathbb{R}^d)$  be an arbitrary function,

$$\begin{aligned} & \int_Q |(u(\mathbf{x}) - \overline{\mathcal{M}}_h u(\mathbf{x})) w(\mathbf{x})| \, d\mathbf{x} \\ &= \int_Q \left| \left( \int_{B(\mathbf{0},1)} \left( \int_0^1 \frac{d}{ds} (u(\mathbf{x} - sh(\mathbf{x})\mathbf{y})) \, ds \right) \rho(\mathbf{y}) \, d\mathbf{y} \right) w(\mathbf{x}) \right| \, d\mathbf{x} \\ &\leq \int_0^1 \int_{\mathbb{R}^d} \int_Q |h(\mathbf{x})\mathbf{y} \cdot \nabla u(\mathbf{x} - sh(\mathbf{x})\mathbf{y}) \rho(\mathbf{y}) w(\mathbf{x})| \, d\mathbf{x} \, d\mathbf{y} \, ds. \end{aligned}$$

Next we note that  $|h(\mathbf{x})\mathbf{y}| \leq H_Q$  and then bound the inner integral independent of  $\mathbf{y}$  by expanding the set  $Q$ :

$$\begin{aligned} & \int_Q |(u(\mathbf{x}) - \overline{\mathcal{M}}_h u(\mathbf{x}))w(\mathbf{x})| \, d\mathbf{x} \\ & \leq H_Q \int_0^1 \int_{\mathbb{R}^d} \rho(\mathbf{y}) \int_Q |\nabla u(\mathbf{x} - sh(\mathbf{x})\mathbf{y})w(\mathbf{x})| \, d\mathbf{x} \, d\mathbf{y} \, ds \\ & \lesssim H_Q \|u\|_{W^{1,p}(Q_h)} \|w\|_{L^{p'}(Q)}. \end{aligned}$$

Note: the construction  $\int \rho(\mathbf{y}) \, d\mathbf{y} = 1$  was used to eliminate the integral in the  $\mathbf{y}$  variable. Selecting  $w = (u(\mathbf{x}) - \overline{\mathcal{M}}_h u(\mathbf{x}))^{p-1}$  completes the result.

Next, the  $k = 1$  case of inequality (18) is addressed. We begin by computing  $\frac{\partial}{\partial x_i} \overline{\mathcal{M}}_h u(\mathbf{x})$  directly:

$$\frac{\partial}{\partial x_i} \overline{\mathcal{M}}_h u(\mathbf{x}) = \overline{\mathcal{M}}_h \partial_i u(\mathbf{x}) + \int_{\mathbb{R}^d} \rho(\mathbf{y}) \partial_i h(\mathbf{x}) \nabla u(\mathbf{x} + h(\mathbf{x})\mathbf{y}) \cdot \mathbf{y} \, d\mathbf{y}. \quad (19)$$

Then letting  $g(\mathbf{x}) := \int_{\mathbb{R}^d} \rho(\mathbf{y}) \partial_i h(\mathbf{x}) \nabla u(\mathbf{x} + h(\mathbf{x})\mathbf{y}) \cdot \mathbf{y} \, d\mathbf{y}$ ,

$$\left\| \frac{\partial}{\partial x_i} (u - \overline{\mathcal{M}}_h u) \right\|_{L^p(Q)} \leq \|\partial_i u - \overline{\mathcal{M}}_h(\partial_i u)\|_{L^p(Q)} + \|g\|_{L^p(Q)}. \quad (20)$$

The first term can be estimated using the argument applied to (18) yielding,

$$\|\partial_i u - \overline{\mathcal{M}}_h(\partial_i u)\|_{L^p(Q)} \leq H_Q \|\partial_i u\|_{W^{1,p}(R)}. \quad (21)$$

It only remains to estimate  $\|g\|_{L^p(Q)}$ . Let  $\mathbb{R}_+^d$  denote the halfspace with positive first coordinate,  $x_i > 0$ . First  $g$  is rewritten in terms of second derivatives of  $u$ :

$$\begin{aligned} g(\mathbf{x}) &= \partial_i h(\mathbf{x}) \int_{\mathbb{R}^d} \rho(\mathbf{y}) \nabla u(\mathbf{x} + h(\mathbf{x})\mathbf{y}) \cdot \mathbf{y} \, d\mathbf{y} \\ &= \partial_i h(\mathbf{x}) \int_{\mathbb{R}_+^d} \rho(\mathbf{y}) (\nabla u(\mathbf{x} + h(\mathbf{x})\mathbf{y}) - \nabla u(\mathbf{x} - h(\mathbf{x})\mathbf{y})) \cdot \mathbf{y} \, d\mathbf{y} \\ &= \partial_i h(\mathbf{x}) \int_{\mathbb{R}_+^d} \rho(\mathbf{y}) \left( \int_{-1}^1 \frac{\partial}{\partial s} \nabla u(\mathbf{x} + sh(\mathbf{x})\mathbf{y}) \, ds \right) \cdot \mathbf{y} \, d\mathbf{y}. \end{aligned}$$

Our selection of a radially symmetric mollifier is important above as the property  $\rho(\mathbf{y}) = \rho(-\mathbf{y})$  has been applied.

$$\begin{aligned} \int |g(\mathbf{x})w(\mathbf{x})| \, d\mathbf{x} &\leq \int_Q \int_{\mathbb{R}_+^d} \int_{-1}^1 |\partial_i h(\mathbf{x})| \rho(\mathbf{y}) \left| \frac{\partial}{\partial s} \nabla u(\mathbf{x} + sh(\mathbf{x})\mathbf{y}) \right| |\mathbf{y}| |w(\mathbf{x})| \, ds \, d\mathbf{y} \, d\mathbf{x} \\ &\lesssim \int_{-1}^1 |s| \int_{\mathbb{R}_+^d} \rho(\mathbf{y}) \int_Q |h(\mathbf{x})| |D^2 u(\mathbf{x} + sh(\mathbf{x})\mathbf{y})| |w(\mathbf{x})| \, d\mathbf{x} \, d\mathbf{y} \, ds, \end{aligned}$$

where  $D^2 u$  denotes the Hessian of  $u$ , the matrix of second derivatives. Next, Hölder's inequality can be applied:

$$\int |g(\mathbf{x})w(\mathbf{x})| \, d\mathbf{x} \lesssim H_Q \|u\|_{W^{2,p}(Q_h)} \|w\|_{L^{p'}(Q)}. \quad (22)$$

Finally setting  $w(\mathbf{x}) = g(\mathbf{x})^{p-1}$ , and then combining (22) with (21) completes the result.  $\square$

The next step is to develop an extension of Proposition 2.2 for the variable-radius mollifier.

**Lemma 3.3.** *If  $1 \leq q \leq p \leq \infty$  and  $Q$  is a bounded set, then for all the sizing functions satisfying (15) and (16) and for all  $u \in W^{2,q}(Q)$ ,  $k \in \{0, 1, 2\}$ ,*

$$|\overline{\mathcal{M}}_h u|_{W^{k,p}(Q)} \lesssim h_Q^{-d} |Q_h|^{\frac{q-1}{q}} |Q|^{\frac{1}{p}} \|u\|_{W^{k,q}(Q_h)} \quad (23)$$

where  $h_Q = \min_{\mathbf{x} \in Q} h(\mathbf{x})$  and  $Q_h := \bigcup_{\mathbf{x} \in Q} B(\mathbf{x}, h(\mathbf{x}))$ .

*Proof.* We begin with the case  $k = 0$ . For all  $x \in Q$ ,

$$\begin{aligned} |\overline{\mathcal{M}}_h u(\mathbf{x})| &\leq \|\rho\|_{L^\infty(\mathbb{R}^d)} \int_{B(\mathbf{0},1)} |u(\mathbf{x} + h(\mathbf{x})\mathbf{y})| \, d\mathbf{y} \\ &\leq \|\rho\|_{L^\infty(\mathbb{R}^d)} \int_{Q_h} h(\mathbf{x})^{-d} |u(\mathbf{z})| \, d\mathbf{z}. \end{aligned}$$

The previous inequality results from a change of variables  $\mathbf{z} = \mathbf{x} + h(\mathbf{x})\mathbf{y}$ . (Note,  $\mathbf{x}$  is fixed in this transformation.) Hölder's inequality gives,

$$|\overline{\mathcal{M}}_h u(\mathbf{x})| = \|\rho\|_{L^\infty(\mathbb{R}^d)} h_Q^{-d} |Q_h|^{1/q'} \|u\|_{L^q(Q_h)}. \quad (24)$$

Integrating (24) in  $\mathbf{x}$  over  $Q$  produces (23) in the  $k = 0$  case.

Next, we turn to the case  $k = 2$ ; the  $k = 1$  case follows from very similar arguments.

$$\begin{aligned} \frac{\partial^2}{\partial x_i \partial x_j} \overline{\mathcal{M}}_h u(\mathbf{x}) &= \int_{\mathbb{R}^d} \rho(\mathbf{y}) \left[ \partial_i \partial_j u(\mathbf{x} + h(\mathbf{x})\mathbf{y}) + \right. \\ &\quad \partial_i h(\mathbf{x}) \nabla (\partial_j u(\mathbf{x} + h(\mathbf{x})\mathbf{y})) \cdot \mathbf{y} + \\ &\quad \partial_j h(\mathbf{x}) \nabla (\partial_i u(\mathbf{x} + h(\mathbf{x})\mathbf{y})) \cdot \mathbf{y} + \\ &\quad \partial_i h(\mathbf{x}) \partial_j h(\mathbf{x}) \mathbf{y} \cdot D^2 u(\mathbf{x} + h(\mathbf{x})\mathbf{y}) \mathbf{y} + \\ &\quad \left. \partial_i \partial_j h(\mathbf{x}) \nabla (u(\mathbf{x} + h(\mathbf{x})\mathbf{y})) \cdot \mathbf{y} \right] d\mathbf{y} \end{aligned} \quad (25)$$

The first term is identical to the  $k = 0$  case using the function  $\partial_i \partial_j u$  in the place of  $u$ . In the second, third, and fourth terms, the same argument holds after we recall (15) and the fact that  $|\mathbf{y}| \leq 1$ . While the argument is the same, the final term yields a few slight differences: (16) is required and the resulting estimate is in terms of  $|u|_{W^{1,p}(Q_h)}$  rather than  $|u|_{W^{2,p}(Q_h)}$  seen in all the other cases. This causes the right-hand side of (23) to contain the full Sobolev norm, rather than the semi-norm.  $\square$

*Remark 3.E.* In Lemma 3.3 (16) is only necessary for the case  $k = 2$ .

*Remark 3.F.* Lemmas 3.2 and 3.3 will be applied for cubes  $Q$  such that  $h_Q \approx H_Q \approx \text{size}(Q) \approx \text{size}(Q_h)$ . In this setting the estimates yield the same factors of 'h' as seen in the analog with fixed mollification radius.

---

**Algorithm 1** Cover  $\Omega$  by cubes proportional to  $h$ .

---

**Require:**  $\mathcal{Q} = \{Q\}$  where  $Q$  is a cube such that  $\text{size}(Q_i) > 4 \max_{\mathbf{x} \in \Omega} h(\mathbf{x})$

**while**  $\exists Q_i \in \mathcal{Q}$  such that  $\text{size}(Q_i) > 8 \max_{\mathbf{x} \in Q_i} h(\mathbf{x})$  **do**

    Remove  $Q_i$  from  $\mathcal{Q}$

    Subdivide  $Q_i$  into  $2^d$  new cubes of size  $\frac{\text{size}(Q_i)}{2}$ .

    Insert all of the new cubes which intersect  $\Omega$  into  $\mathcal{Q}$ .

**end while**

---

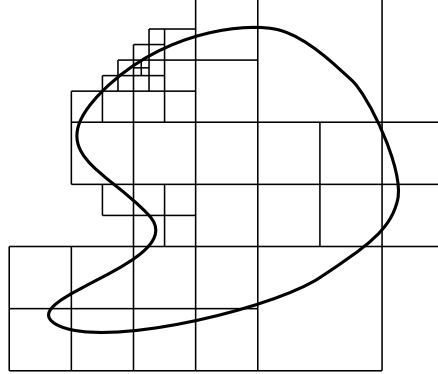


Figure 8: An example cubic covering of a domain for some non-uniform sizing function.

### 3.3 Cubic Covering of $\Omega$

An important idea in Section 2 was the use of a cubic grid of size  $h$  to cover the domain. In the non-uniform case we seek to cover  $\Omega$  with cubes such that each cube has side length proportional to  $h(\mathbf{x})$  for  $\mathbf{x} \in Q$ ; see Figure 8 for an example. This is possible due to restriction (15) and can be achieved through the simple algorithm given in Algorithm 1.

First we assert that any neighboring cubes produced have proportional sizes.

**Proposition 3.4.** *Let  $\mathcal{Q} = \{Q_i\}$  be the set of cubes resulting from Algorithm 1. If  $Q_i \cap Q_j \neq \emptyset$ , then  $\text{size}(Q_i) \leq 2 \text{size}(Q_j)$ .*

*Proof.* Suppose  $Q_i$  and  $Q_j$  are adjacent cubes such that  $\text{size}(Q_i) \leq \frac{\text{size}(Q_j)}{4}$ . Let  $\mathbf{x}_i \in Q_i \cap Q_j$  and let  $\mathbf{x}_j \in Q_j$ . Then,

$$h(\mathbf{x}_j) \leq h(\mathbf{x}_i) + \frac{1}{16\sqrt{d}} |\mathbf{x}_j - \mathbf{x}_i| \leq \frac{\text{size}(Q_i)}{4} + \frac{1}{16\sqrt{d}} \sqrt{d} \text{size}(Q_j) \leq \frac{1}{8} \text{size}(Q_j).$$

Thus  $Q_j$  is eligible to be split by Algorithm 1 and we conclude that upon termination of the algorithm the lemma holds.  $\square$

The next proposition shows that for a compatible triangulation the simplices which intersect a particular cube are contained well inside the cube and its neighboring cubes.

**Proposition 3.5.** *Let  $\mathcal{Q}$  be the set of cubes resulting from Algorithm 1. Then for all  $\mathbf{x} \in Q$ ,  $4h(\mathbf{x}) \leq \text{size}(Q) \leq 16h(\mathbf{x})$ .*



*Proof.* Let  $Q$  be any cube resulting from Algorithm 1 and let  $\hat{Q}$  be the cube which was subdivided to form  $Q$ . Then

$$2 \operatorname{size}(Q) = \operatorname{size}(\hat{Q}) > 8 \max_{\mathbf{x} \in \hat{Q}} h(\mathbf{x}) > 8 \max_{\mathbf{x} \in Q} h(\mathbf{x}). \quad (26)$$

There exists  $\mathbf{x} \in Q$  such that  $h(\mathbf{x}) > \frac{\operatorname{size}(Q)}{8}$  since  $Q$  is not split further by the algorithm. By requirement (15), for any  $\mathbf{y} \in Q$ ,

$$h(\mathbf{y}) \geq h(\mathbf{x}) - \frac{1}{16\sqrt{d}}|\mathbf{x} - \mathbf{y}| \geq \frac{\operatorname{size}(Q)}{16}. \quad (27)$$

Combining (26) and (27) yields the result.  $\square$

**Lemma 3.6.** *Let triangulation  $\mathcal{T}$  be compatible with  $h$  and let  $\mathcal{Q} = \{Q_i\}$  be the set of cubes resulting from Algorithm 1. Then for all  $Q \in \mathcal{Q}$ ,*

$$\bigcup_{K \in \mathcal{K}(Q)} \bigcup_{\mathbf{x} \in K} B(\mathbf{x}, h(\mathbf{x})) \subset S_i := \bigcup_{Q_j \cap Q \neq \emptyset} Q_j,$$

where  $\mathcal{K}(Q)$  denotes the set of simplices intersecting  $Q$ .

*Proof.* Let  $K \in \mathcal{K}(Q)$  and let  $\mathbf{x} \in K$ . Then  $\mathbf{x}$  belongs to  $Q$  or a cube neighboring  $Q$  since by Propositions 3.4 and 3.5  $\operatorname{dist}(\mathbf{x}, Q) \leq h_K \leq \operatorname{size}(Q)/4 \leq \operatorname{size}(\hat{Q})/2$  for any neighboring cube  $\hat{Q}$ . Let  $\hat{Q}$  denote the cube containing  $\mathbf{x}$  and let  $\mathbf{y} \in B(\mathbf{x}, h(\mathbf{x}))$ . Also let  $\mathbf{z}$  denote the nearest point in  $Q$  to  $\mathbf{x}$ . The result is shown in two cases depending on the size of  $\hat{Q}$ .

Case 1:  $\operatorname{size}(\hat{Q}) = \operatorname{size}(Q)$  or  $\operatorname{size}(\hat{Q}) = \operatorname{size}(Q)/2$ . Then

$$|\mathbf{y} - \mathbf{z}| \leq |\mathbf{y} - \mathbf{x}| + |\mathbf{x} - \mathbf{z}| \leq h(\mathbf{x}) + h_K \leq \frac{\operatorname{size}(Q)}{4} + \frac{\operatorname{size}(Q)}{4} = \frac{\operatorname{size}(Q)}{2}.$$

By the balance property (Proposition 3.4), this implies that  $\mathbf{y} \in \bigcup_{Q_j \cap Q \neq \emptyset} Q_j$ .

Case 2:  $\operatorname{size}(\hat{Q}) = 2 \operatorname{size}(Q)$ . Then

$$\begin{aligned} |\mathbf{y} - \mathbf{z}| &\leq |\mathbf{y} - \mathbf{x}| + |\mathbf{x} - \mathbf{z}| \leq h(\mathbf{x}) + h_K \\ &\leq h(\mathbf{z}) + \frac{|\mathbf{z} - \mathbf{x}|}{16\sqrt{d}} + \frac{\operatorname{size}(Q)}{4} \leq \operatorname{size}(Q) \left( \frac{1}{2} + \frac{1}{64\sqrt{d}} \right). \end{aligned}$$

The balance property (Proposition 3.4) ensures that  $\mathbf{y}$  must belong to a cube of size greater than or equal to  $\operatorname{size}(Q)$ . Since  $|\mathbf{y} - \mathbf{z}| \leq \operatorname{size}(Q)$ ,  $\mathbf{y} \in \bigcup_{Q_j \cap Q \neq \emptyset} Q_j$ .  $\square$

### 3.4 Error Estimate

Now we have the necessary tools to prove an analog of Theorem 2.4 for a non-uniform sizing function and compatible triangulation.

**Theorem 3.7.** *Let  $k \in \{0, 1\}$ . For sizing functions  $h$  satisfying (15) and (16) paired with any compatible mesh  $\mathcal{T}$ , let  $\mathcal{Q}$  be the set of cubes produced by Algorithm 1. Then for all  $u \in W^{1,p}(\Omega)$ ,*

$$\|u - \Xi_h u\|_{W^{k,p}(\Omega)} \lesssim \left( \sum_{Q_i \in \mathcal{Q}} \left( \frac{h_i}{(\cos \theta_i^*)^k} \right)^p \|u\|_{W^{k+1,p}(Q_i)}^p \right)^{\frac{1}{p}} \quad (28)$$

where  $h_i = \max_{\mathbf{x} \in Q_i} h(\mathbf{x})$  and  $\theta_i^* = \max_{T \cap S_i \neq \emptyset} \theta_T$ .

Note that this result is uniform over the class of sizing functions satisfying (15) and (16), i.e., the inequality depends only upon  $C_{\text{sizing}}$  and not the specific sizing function or mesh selected.

*Proof.* The proof structure is very similar to Theorem 2.4. First the estimate is divided using the triangle inequality and the intermediate function  $\Xi_h u$ .

$$|u - \Xi_h u|_{W^{k,p}(Q_i)} \leq |u - \overline{\mathcal{M}}_h u|_{W^{k,p}(Q_i)} + |\overline{\mathcal{M}}_h u - \Xi_h u|_{W^{k,p}(Q_i)}. \quad (29)$$

Let  $S_i$  denote the union of  $Q_i$  with its neighboring cubes. By Lemmas 3.2 and 3.6,

$$|u - \overline{\mathcal{M}}_h u|_{W^{k,p}(Q_i)} \leq h_i \|u\|_{W^{k+1,p}(S_i)}.$$

Next the second term in (29) is estimated. Select a value  $q > d$ . Proposition 2.3 and Theorem 1.4 are applied for the smooth function  $\overline{\mathcal{M}}_h u$ :

$$\begin{aligned} |\overline{\mathcal{M}}_h u - \Pi_1 \overline{\mathcal{M}}_h u|_{W^{k,p}(Q_i)}^q &\lesssim \text{size}(Q_i)^{\frac{q-p}{p}} \sum_{K \in \mathcal{K}_i} |\overline{\mathcal{M}}_h u - \Pi_1 \overline{\mathcal{M}}_h u|_{W^{k,q}(K)}^q \\ &\lesssim \text{size}(Q_i)^{\frac{q-p}{p}} \sum_{K \in \mathcal{K}_i} \left( \frac{h_K}{(\cos \theta_K)^k} \right)^q |\overline{\mathcal{M}}_h u|_{W^{k+1,q}(K)}^q. \end{aligned}$$

Lemma 3.6 is applied to combine terms in the summation and Lemma 3.3 is used to return the estimate to the correct norm. Letting  $\theta_i = \max_{T \cap Q_i \neq \emptyset} \theta_T$ ,

$$\begin{aligned} |\overline{\mathcal{M}}_h u - \Pi_1 \overline{\mathcal{M}}_h u|_{W^{k,p}(Q_i)}^q &\lesssim \text{size}(Q_i)^{\frac{q-p}{p}} \left( \frac{h_i}{(\cos \theta_i)^k} \right)^q |\overline{\mathcal{M}}_h u|_{W^{k+1,q}(\cup_{K \in \mathcal{K}_i} K)}^q \\ &\lesssim \text{size}(Q_i)^{\frac{q-p}{p}} \left( \frac{h_i}{(\cos \theta_i)^k} \right)^q h_i^{\frac{p-q}{p}} \|u\|_{W^{k+1,p}(S_i)}^q \\ &\lesssim \left( \frac{h_i}{(\cos \theta_i)^k} \right)^q \|u\|_{W^{k+1,p}(S_i)}^q. \end{aligned} \quad (30)$$

Inequality (30) follows from Proposition 3.5 which guarantees that  $h_i^{-1} \text{size}(Q_i) \lesssim 1$ . Finally summation over all cubes and again utilizing Proposition 3.5 finishes the result:

$$\begin{aligned} \sum_i \|u - \Xi_h u\|_{W^{k,p}(Q_i)}^p &\lesssim \sum_i \left( \frac{h_{Q_i}}{(\cos \theta_i)^k} \right)^p \|u\|_{W^{k+1,p}(S_i)}^p \\ &\lesssim \sum_i \left[ \sum_{j \text{ s.t. } Q_j \cap Q_i \neq \emptyset} \left( \frac{h_{Q_j}}{(\cos \theta_j)^k} \right)^p \right] \|u\|_{W^{1,p}(Q_i)}^p \\ &\lesssim 2^{(2d+1)p} \sum_i \left( \frac{h_{Q_i}}{(\cos \theta_i^*)^k} \right)^p \|u\|_{W^{1,p}(Q_i)}^p. \end{aligned}$$

The final inequality results from the fact that there are fewer than  $4^d$  cubes intersecting  $Q_i$  and for any such cube  $Q_j$ ,  $h_{Q_j} \leq 2h_{Q_i}$ .  $\square$

*Remark 3.G.* Due to Remark 3.E, the  $k = 0$  case of Theorem 3.7 does not depend upon (16). This means that  $L^p$  error estimates can be established for Lipschitz sizing functions (which, recalling Remark 3.C, result from adaptive refinement in many common settings).

*Remark 3.H.* Similarly to Theorem 2.4, the extension domain property is also important to Theorem 3.7. The usage is somewhat implicit: the continuous extension property is used to define  $u$  on  $Q_i \not\subset \Omega$ .

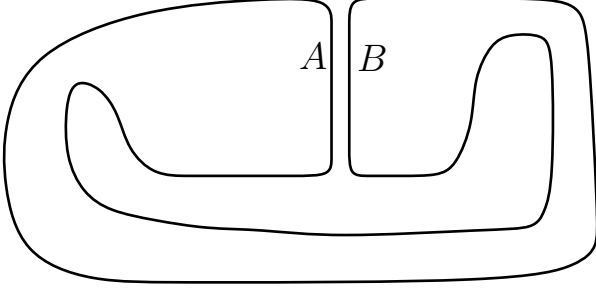


Figure 9: Requiring a globally defined sizing function can lead to restrictions associated with non-local portions of the boundary; e.g., a small mesh in region  $A$  will force a reasonably small mesh also in region  $B$ .

## 4 Final Remarks

To extend the maximum angle condition and its generalizations to less regular data Section 3 establishes ply-independent error estimates for a linear average interpolant. Broadly, this serves to demonstrate that requirements on shape regularity and data regularity can be weakened simultaneously without the inherent trade-off seen in previous analyses. Specifically, these results extend optimal error estimates under the maximum angle condition to several situations for which it was previously not shown: linear interpolation of functions in  $W^{1,p}(\Omega)$  and (when  $1 \leq d/p$ )  $W^{2,p}(\Omega)$ . The latter result is most important in three dimensions where the maximum angle condition does not hold for the Lagrange interpolant when  $p \leq 2$ .

We have not addressed the question of boundary conditions. For error estimates in the  $L^p$ -norm, the function can be explicitly adjusted to match a homogeneous boundary condition. In [11] averaging regions near the boundary are shifted outside the domain where an explicit extension by zero is used. When only considering the more straightforward scalar functions, an explicit cutoff can also accomplish the task [34]. However, these manipulations corrupt error estimates of the derivatives and fully generalizing (28) to preserve boundary data requires an alternative approach. Ideally a mollification procedure analogous to the Scott and Zhang interpolant [38] can be constructed which averages over a lower dimensional manifold so that only boundary data is taken into account. Naïve approaches in this direction have proven unsuccessful.

The extension domain property has been used somewhat implicitly throughout the arguments presented posing some limitations on the set of admissible meshes. Since the sizing function is defined globally, mesh size in non-local portions of the domain can impact the maximum mesh size; see Figure 9. These limitations are often ignored in the analysis as asymptotically (as  $h$  becomes small) the impact disappears since the domain is fixed. Other mollification-based average interpolants face similar limitations. For example, the construction of Christiansen and Winther explicitly construct a small extension region outside the domain boundary and mesh elements are assumed to be smaller than the size of this extension region. Ideally the domain could be embedded in a manifold in which the non-local portions are far away (or more technically the extension operator has a much smaller constant). While ad hoc methods for doing this are apparent for simple examples, a complete theory for general domains is not.

In some cases, a higher-order variant of Theorem 3.7 may be necessary. Recall that for the expected convergence rate in the  $W^{m,p}$ -norm, Jamet's theorem requires  $k + 1 - m > \frac{d}{p}$ , where  $k + 1$  is the order of interpolation used and  $d$  is the spatial dimension. Thus average interpolation is needed when  $d$  is large,  $p$  is near 1, or  $m$  is near  $k$ , i.e., estimates on higher derivatives are desired. Extending Theorem 3.7 to cover these cases requires a higher-order mollification procedure. While this mollifier can be defined, precise technical requirements under which the interpolation error

estimate holds are still unknown.

## Acknowledgments

The author gratefully acknowledges Noel Walkington for many useful discussions and the anonymous reviewers for several insightful comments.

## References

- [1] G. ACOSTA, *Lagrange and average interpolation over 3D anisotropic elements*, J. Comput. Appl. Math., 135 (2001), pp. 91–109.
- [2] G. ACOSTA, T. APEL, R. G. DURÁN, AND A. L. LOMBARDI, *Error estimates for Raviart-Thomas interpolation of any order on anisotropic tetrahedra*, Math. Comput., 80 (2011), pp. 141–163.
- [3] R. A. ADAMS, *Sobolev Spaces*, vol. 140 of Pure and Applied Mathematics, Academic Press, Oxford, second ed., 2003.
- [4] T. APEL, *Anisotropic Finite Elements: Local Estimates and Applications*, Advances in Numerical Mathematics, Teubner, Stuttgart, 1999.
- [5] D. N. ARNOLD, R. S. FALK, AND R. WINTHER, *Finite element exterior calculus, homological techniques, and applications*, Acta Numer., 15 (2006), pp. 1–156.
- [6] I. BABUŠKA AND A. K. AZIZ, *On the angle condition in the finite element method*, SIAM J. Numer. Anal., 13 (1976), pp. 214–226.
- [7] J. H. BRAMBLE AND S. R. HILBERT, *Estimation of linear functionals on Sobolev spaces with application to Fourier transforms and spline interpolation*, SIAM J. Numer. Anal., 7 (1970), pp. 112–124.
- [8] J. BRANDTS, A. HANNUKAINEN, S. KOROTOV, AND M. KŘÍŽEK, *On angle conditions in the finite element method*, SeMA J., (2011), pp. 81–95.
- [9] S. C. BRENNER AND L. R. SCOTT, *The Mathematical Theory of Finite Element Methods*, vol. 15 of Texts in Applied Mathematics, Springer, New York, third ed., 2008.
- [10] S.-W. CHENG, T. K. DEY, AND J. A. LEVINE, *A practical Delaunay meshing algorithm for a large class of domains*, in Proc. 16th Int. Meshing Roundtable, 2007, pp. 477–494.
- [11] S. H. CHRISTIANSEN AND R. WINTHER, *Smoothed projections in finite element exterior calculus*, Math. Comput., 77 (2008), p. 813.
- [12] P. G. CIARLET, *The Finite Element Method for Elliptic Problems*, vol. 40 of Classics in Applied Mathematics, SIAM, Philadelphia, PA, second ed., 2002.
- [13] P. CLÉMENT, *Approximation by finite element functions using local regularization*, RAIRO Anal. Numér., 9 (1975), pp. 77–84.

- [14] S. DEKEL AND D. LEVIATAN, *The Bramble-Hilbert lemma for convex domains*, SIAM J. Math. Anal., 35 (2004), pp. 1203–1212.
- [15] A. ERN AND J.-L. GUERMOND, *Theory and Practice of Finite Elements*, vol. 159 of Applied Mathematical Sciences, Springer-Verlag, New York, 2004.
- [16] I. FONSECA AND G. LEONI, *Modern Methods in the Calculus of Variations:  $L^p$  Spaces*, Monographs in Mathematics, Springer, New York, 2007.
- [17] L. FORMAGGIA AND S. PEROTTO, *New anisotropic a priori error estimates*, Numer. Math., 89 (2001), pp. 641–67.
- [18] S. GUATTERY, G. L. MILLER, AND N. WALKINGTON, *Estimating interpolation error: a combinatorial approach*, in Proc. 10th Symp. Discrete Algorithms, 1999, pp. 406–413.
- [19] A. HANNUKAINEN, S. KOROTOV, AND M. KRÍŽEK, *The maximum angle condition is not necessary for convergence of the finite element method*, Numer. Math., 120 (2012), pp. 79–88.
- [20] U. HETMANIUK AND P. KNUPP, *Local anisotropic interpolation error estimates based on directional derivatives along edges*, SIAM J. Numer. Anal., 47 (2008), pp. 575–595.
- [21] B. HUDSON, G. L. MILLER, AND T. PHILLIPS, *Sparse Voronoi refinement*, in Proc. 15th Int. Meshing Roundtable, 2006, pp. 339–356.
- [22] P. JAMET, *Estimations d’erreur pour des éléments finis droits presque dégénérés*, RAIRO Anal. Numér., 10 (1976), pp. 43–61.
- [23] M. KRÍŽEK, *On semiregular families of triangulations and linear interpolation*, Appl. Math., 36 (1991), pp. 223–232.
- [24] ———, *On the maximum angle condition for linear tetrahedral elements*, SIAM J. Numer. Anal., 29 (1992), pp. 513–520.
- [25] X. Y. LI, *Generating well-shaped  $d$ -dimensional Delaunay meshes*, Theo. Comput. Sci., 296 (2003), pp. 145–165.
- [26] E. LIEB AND M. LOSS, *Analysis*, vol. 14 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, second ed., 2001.
- [27] A. MAJDA AND A. BERTOZZI, *Vorticity and Incompressible Flow*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 2002.
- [28] S. MAO AND Z. SHI, *Error estimates of triangular finite elements under a weak angle condition*, J. Comput. Appl. Math., 230 (2009), pp. 329–332.
- [29] G. MILLER, S. PAV, AND N. WALKINGTON, *When and why Ruppert’s algorithm works*, in Proc. 12th Int. Meshing Roundtable, 2003, pp. 91–102.
- [30] G. L. MILLER, T. PHILLIPS, AND D. SHEEHY, *Size competitive meshing without large angles*, in 34th Int. Coll. Automata Lang. Prog., 2007.

- [31] S. A. MITCHELL, *Cardinality bounds for triangulations with bounded minimum angle*, in Proc. 6th Canadian Conf. Comput. Geom., 1994, pp. 326–331.
- [32] R. NARASIMHAN, *Analysis on Real and Complex Manifolds*, vol. 35 of North-Holland Mathematical Library, Elsevier Publishing Company, Inc., New York, second ed., 1973.
- [33] S. E. PAV AND N. J. WALKINGTON, *Delaunay refinement by corner lopping*, in Proc. 14th Int. Meshing Roundtable, 2005, pp. 165–181.
- [34] A. RAND, *Delaunay Refinement Algorithms for Numerical Methods*, PhD thesis, Carnegie Mellon University, Pittsburgh, PA, 2009.
- [35] A. RAND AND N. WALKINGTON, *3D Delaunay refinement of sharp domains without a local feature size oracle*, in Proc. 17th Int. Meshing Roundtable, 2008, pp. 37–54.
- [36] A. RAND AND N. WALKINGTON, *Collars and intestines: Practical conforming Delaunay refinement*, in Proc. 18th Int. Meshing Roundtable, 2009, pp. 481–497.
- [37] J. S. SCHÖBERL, *A posteriori error estimates for Maxwell equations*, Math. Comput., 77 (2008), pp. 633–649.
- [38] L. R. SCOTT AND S. ZHANG, *Finite element interpolation of nonsmooth functions satisfying boundary conditions*, Math. Comput., 54 (1990), pp. 483–493.
- [39] N. A. SHENK, *Uniform error estimates for certain narrow Lagrange finite elements*, Math. Comput., 63 (1994), pp. 105–119.
- [40] J. SHEWCHUK, *Delaunay refinement algorithms for triangular mesh generation*, Comput. Geom., 22 (2002), pp. 86–95.
- [41] H. SI, *On refinement of constrained Delaunay tetrahedralizations*, in Proc. 15th Int. Meshing Roundtable, 2006, pp. 510–528.
- [42] H. SI AND K. GARTNER, *Meshing piecewise linear complexes by constrained Delaunay tetrahedralizations*, in Proc. 14th Int. Meshing Roundtable, 2005, pp. 147–163.
- [43] J. L. SYNGE, *The Hypercircle in Mathematical Physics*, Cambridge University Press, Cambridge, 1957.
- [44] D. TALMOR, *Well-spaced points for numerical methods*, PhD thesis, Carnegie Mellon University, Pittsburgh, PA, 1997.

## A Coplanarity vs. the Maximum Angle Condition

*Proof.* (Proposition 1.1) For this proof we omit the subscript  $K$  on  $\theta_K$  and  $\psi_K$  without ambiguity. First, we will bound  $\psi$  from above by a function of  $\theta$  to establish that if  $\theta$  is bounded away from  $\pi/2$ , then  $\psi$  is bounded away from  $\pi$ . This will be shown by direct construction: for a tetrahedron with maximum angle  $\psi$ , we will construct a particular vector  $\hat{\xi}$  which ensures that  $\theta$  is sufficiently large. Construction is divided into two cases. First we suppose that  $\psi$  is realized by a dihedral angle of the tetrahedron. Considering this largest dihedral angle, all vertices of the tetrahedron are contained

in two planes that meet at angles of  $\psi$  and  $\pi - \psi$ . Selecting a vector that points “away” from these planes demonstrates

$$\theta \geq \psi/2. \quad (31)$$

Next we suppose that  $\psi$  is realized by an angle in one of the faces of the tetrahedron. Let  $\mathbf{v}_1$  be the vertex at which the angle  $\psi$  occurs, let  $\mathbf{v}_2$  and  $\mathbf{v}_3$  denote the other two vertices of that triangle and let  $\mathbf{v}_4$  denote the last vertex; see Fig. 10. Moreover, without loss of generality we can assume that  $|\mathbf{v}_2 - \mathbf{v}_4| \leq |\mathbf{v}_3 - \mathbf{v}_4|$ , i.e.,  $\mathbf{v}_4$  is nearer to  $\mathbf{v}_2$  than  $\mathbf{v}_3$ . Let  $\hat{\boldsymbol{\xi}}$  be the unit normal vector to the triangle with vertices  $\mathbf{v}_1$ ,  $\mathbf{v}_2$  and  $\mathbf{v}_4$ . To bound  $\theta$ , first observe that  $\hat{\boldsymbol{\xi}}$  is orthogonal to three of the six lines containing edges of the tetrahedron. Two other edges belong to the triangle with vertices  $\mathbf{v}_1$ ,  $\mathbf{v}_2$ ,  $\mathbf{v}_3$ , i.e., the triangle with a largest angle  $\psi$ . This large angle ensures that for any unit vector  $\mathbf{u}_i$  in the direction of either of these two lines,  $\hat{\boldsymbol{\xi}} \cdot \mathbf{u}_i \geq \alpha - \pi/2$ . Lastly, we consider unit vectors in the direction of the line through  $\mathbf{v}_3$  and  $\mathbf{v}_4$ . The triangle inequality implies

$$\left| \hat{\boldsymbol{\xi}} \cdot (\mathbf{v}_3 - \mathbf{v}_4) \right| \leq \left| \hat{\boldsymbol{\xi}} \cdot (\mathbf{v}_3 - \mathbf{v}_2) \right| + \left| \hat{\boldsymbol{\xi}} \cdot (\mathbf{v}_2 - \mathbf{v}_4) \right| = \left| \hat{\boldsymbol{\xi}} \cdot (\mathbf{v}_3 - \mathbf{v}_2) \right|,$$

where the second term in the sum is zero by construction of  $\hat{\boldsymbol{\xi}}$  orthogonal to the a plane containing  $\mathbf{v}_2$  and  $\mathbf{v}_4$ . Recalling our earlier argument,  $\left| \hat{\boldsymbol{\xi}} \cdot (\mathbf{v}_3 - \mathbf{v}_2) \right| \leq |\mathbf{v}_3 - \mathbf{v}_2| \cos(\psi - \pi/2)$ . Thus we can estimate the angle between  $\hat{\boldsymbol{\xi}}$  and the line containing  $\mathbf{v}_3$  and  $\mathbf{v}_4$ .

$$\frac{\left| \hat{\boldsymbol{\xi}} \cdot (\mathbf{v}_3 - \mathbf{v}_4) \right|}{|\mathbf{v}_3 - \mathbf{v}_4|} \leq \frac{|\mathbf{v}_3 - \mathbf{v}_2| \cos(\psi - \pi/2)}{|\mathbf{v}_3 - \mathbf{v}_4|} \leq 2 \cos(\psi - \pi/2). \quad (32)$$

The final inequality above follows from our definition of  $\mathbf{v}_3$ :  $|\mathbf{v}_3 - \mathbf{v}_2| \leq |\mathbf{v}_3 - \mathbf{v}_4| + |\mathbf{v}_4 - \mathbf{v}_2| \leq 2|\mathbf{v}_3 - \mathbf{v}_4|$ . We thus conclude that

$$\theta \geq \min_i \hat{\boldsymbol{\xi}} \cdot \mathbf{u}_i \geq \arccos(2 \cos(\psi - \pi/2)). \quad (33)$$

Considering the two expressions (31) and (33), we see that if  $\theta$  is bounded away from  $\pi/2$ , then  $\psi$  is bounded away from  $\pi$ .

Next we establish a reverse inequality. The crux of the argument is contained in an inequality from [2]. Defining  $m := \sin \frac{\pi - \psi}{2}$ , Acosta et al. assert that there are unit vectors  $\mathbf{u}_1$ ,  $\mathbf{u}_2$ ,  $\mathbf{u}_3$  in the directions of the edges of the tetrahedron such that

$$|(\mathbf{u}_1 \times \mathbf{u}_2) \cdot \mathbf{u}_3| \geq m^3. \quad [2, \text{p. 147}] \quad (34)$$

Letting  $\boldsymbol{\xi}$  be an arbitrary unit vector, we consider the decomposition  $\boldsymbol{\xi} = a(\mathbf{u}_1 \times \mathbf{u}_2) + b\mathbf{u}_{12}$  where  $\mathbf{u}_{12}$  is a unit vector in the plane spanned by  $\mathbf{u}_1$  and  $\mathbf{u}_2$ . Since these are orthogonal subspaces,  $a^2 + b^2 = 1$  and thus  $a \geq 1/\sqrt{2}$  or  $b \geq 1/\sqrt{2}$ . If  $a \geq 1/\sqrt{2}$ , then

$$|\boldsymbol{\xi} \cdot \mathbf{u}_3| \geq a |(\mathbf{u}_1 \times \mathbf{u}_2) \cdot \mathbf{u}_3| \geq \frac{m^3}{\sqrt{2}}. \quad (35)$$

Otherwise  $b \geq 1/\sqrt{2}$ . Letting  $\gamma$  denote the *non-acute* angle between the lines containing  $\mathbf{u}_1$  and  $\mathbf{u}_2$ , we observe that (34) implies that  $\sin(\pi - \gamma) \geq m^3$  or  $\gamma \leq \pi - \arcsin(m^3)$ . Also the angle between  $\mathbf{u}_{12}$  and the line in the direction of either  $\mathbf{u}_1$  or  $\mathbf{u}_2$  is no larger than  $\gamma/2$ . Thus,

$$\max_{i \in \{1,2\}} |\boldsymbol{\xi} \cdot \mathbf{u}_i| \geq b \cos\left(\frac{\gamma}{2}\right) \geq \frac{1}{\sqrt{2}} \cos\left(\frac{\pi - \arcsin(m^3)}{2}\right). \quad (36)$$

Combining (35) and (36),

$$\theta = \max_{|\boldsymbol{\xi}|=1} \min_i \arccos \boldsymbol{\xi} \cdot \mathbf{u}_i \leq \arccos \left( \max \left\{ \frac{m^3}{\sqrt{2}}, \frac{1}{\sqrt{2}} \cos \left( \frac{\pi - \arcsin(m^3)}{2} \right) \right\} \right).$$

Despite the complex form, this is the desired inequality. Specifically, if  $\psi$  is bounded away from  $\pi$ , then  $m$  is bounded away from zero. Since all the functions we consider are continuous, it follows that the argument of  $\arccos$  is also bounded away from zero. Thus  $\theta$  is bounded away from  $\pi/2$ .  $\square$

## B Mollification

Proofs of several propositions in Section 2.1 are provided here.

*Proof.* (Proposition 2.1) Define  $p'$  by  $\frac{1}{p} + \frac{1}{p'} = 1$  and let  $w \in L^{p'}(\mathbb{R}^d)$  be an arbitrary function. If  $u \in C_c^\infty(\mathbb{R}^d)$  then

$$\begin{aligned} & \int_{\mathbb{R}^d} |(u(\mathbf{x}) - \mathcal{M}_h u(\mathbf{x}))w(\mathbf{x})| \, d\mathbf{x} \\ &= \int_{\mathbb{R}^d} \left| \left( u(\mathbf{x}) - \int_{\mathbb{R}^d} u(\mathbf{x} - \mathbf{y}) \rho_h(\mathbf{y}) \, d\mathbf{y} \right) w(\mathbf{x}) \right| \, d\mathbf{x} \\ &= \int_{\mathbb{R}^d} \left| \left( \int_{\mathbb{R}^d} (u(\mathbf{x}) - u(\mathbf{x} - \mathbf{y})) \rho_h(\mathbf{y}) \, d\mathbf{y} \right) w(\mathbf{x}) \right| \, d\mathbf{x} \\ &= \int_{\mathbb{R}^d} \left| \left( \int_{\mathbb{R}^d} \left( \int_0^1 \frac{d}{ds} u(\mathbf{x} - s\mathbf{y}) \, ds \right) \rho_h(\mathbf{y}) \, d\mathbf{y} \right) w(\mathbf{x}) \right| \, d\mathbf{x}. \end{aligned}$$

The final equality results from the fundamental theorem of calculus. In the support of mollifier  $\rho_h$ ,  $|\mathbf{y}| \leq h$  so  $|\mathbf{y}\rho_h(\mathbf{y})| \leq h\rho_h(\mathbf{y})$ . Then Hölder's inequality can be applied:

$$\begin{aligned} & \int_{\mathbb{R}^d} |(u(\mathbf{x}) - \mathcal{M}_h u(\mathbf{x}))w(\mathbf{x})| \, d\mathbf{x} \\ &= \int_{\mathbb{R}^d} \left| \left( \int_{\mathbb{R}^d} \left( \int_0^1 -\mathbf{y} \cdot \nabla u(\mathbf{x} - s\mathbf{y}) \, ds \right) \rho_h(\mathbf{y}) \, d\mathbf{y} \right) w(\mathbf{x}) \right| \, d\mathbf{x} \\ &\leq \int_0^1 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |-\mathbf{y} \cdot \nabla u(\mathbf{x} - s\mathbf{y}) \rho_h(\mathbf{y}) w(\mathbf{x})| \, d\mathbf{x} \, d\mathbf{y} \, ds \\ &\leq \int_0^1 \int_{\mathbb{R}^d} h \rho_h(\mathbf{y}) \left( \int_{\mathbb{R}^d} |\nabla u(\mathbf{x} - s\mathbf{y})|^p \, d\mathbf{x} \right)^{\frac{1}{p}} \left( \int_{\mathbb{R}^d} |w(\mathbf{x})|^{p'} \, d\mathbf{x} \right)^{\frac{1}{p'}} \, d\mathbf{y} \, ds \\ &= h \|\nabla u\|_{L^p(\mathbb{R}^d)} \|w\|_{L^{p'}(\mathbb{R}^d)}. \end{aligned}$$

By selecting  $w = (u - \mathcal{M}_h u)^{p-1}$ , the left-hand side of the above estimate becomes  $\|u - \mathcal{M}_h u\|_{L^p(\mathbb{R}^d)}^p$ . Since  $p'(p-1) = p$ ,  $\|w\|_{L^{p'}(\mathbb{R}^d)} = \|u - \mathcal{M}_h u\|_{L^p(\mathbb{R}^d)}^{p-1}$ . Dividing both sides by  $\|w\|_{L^{p'}(\mathbb{R}^d)}$  gives the desired inequality. Density of smooth functions in  $W^{1,p}(\Omega)$  is used to extend this result to the full space.  $\square$



*Proof.* (Proposition 2.2) Let  $\hat{u}(x) := u(x)\chi_{Q_h}(x)$  where  $\chi_{Q_h}$  is the characteristic function of  $Q_h$ . Applying Young's inequality to  $\mathcal{M}_h\hat{u}$  gives

$$\|\mathcal{M}_h\hat{u}\|_{L^p(\mathbb{R}^d)} \lesssim \|\rho_h\|_{L^{\frac{pq}{pq+q-p}}(\mathbb{R}^d)} \|\hat{u}\|_{L^q(\mathbb{R}^d)}.$$

Since  $\mathcal{M}_h\hat{u} \equiv \mathcal{M}_hu$  on  $Q$ ,  $\|\mathcal{M}_hu\|_{L^p(Q)} = \|\mathcal{M}_h\hat{u}\|_{L^p(Q)} \leq \|\mathcal{M}_h\hat{u}\|_{L^p(\mathbb{R}^d)}$ . Then recognizing that  $\|\hat{u}\|_{L^q(\mathbb{R}^d)} = \|\hat{u}\|_{L^q(Q_h)}$  and computing the norm of  $\rho_h$  yields the desired estimate.  $\square$

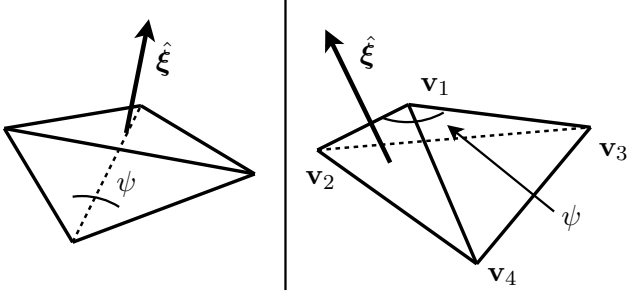


Figure 10: In the proof of Proposition 1.1, when bounding  $\psi$  in terms of  $\theta$ , a specific  $\hat{\xi}$  can be constructed explicitly when  $\psi$  is realized by a large dihedral angle (left) or a large angle in a face (right).