

## OPTIMAL ORDER CONVERGENCE IMPLIES NUMERICAL SMOOTHNESS

SO-HSIANG CHOU

**Abstract.** It is natural to expect the following loosely stated approximation principle to hold: a numerical approximation solution should be in some sense as smooth as its target exact solution in order to have optimal convergence. For piecewise polynomials, that means we have to at least maintain numerical smoothness in the interiors as well as across the interfaces of cells or elements. In this paper we give clear definitions of numerical smoothness that address the across-interface smoothness in terms of *scaled* jumps in derivatives [9] and the interior numerical smoothness in terms of differences in derivative values. Furthermore, we prove rigorously that the principle can be simply stated as *numerical smoothness is necessary for optimal order convergence*. It is valid on quasi-uniform meshes by triangles and quadrilaterals in two dimensions and by tetrahedrons and hexahedrons in three dimensions. With this validation we can justify, among other things, incorporation of this principle in creating adaptive numerical approximation for the solution of PDEs or ODEs, especially in designing proper smoothness indicators or detecting potential non-convergence and instability.

**Key words.** Adaptive algorithm, discontinuous Galerkin, numerical smoothness, optimal order convergence.

### 1. Introduction

Consider the problem of approximating a function  $u$  defined on a domain in  $\mathbb{R}^n$  by a sequence of numerical solutions  $\{u_h\}$ . The target function  $u$  may be an exact solution of a second or higher order partial or ordinary differential equation, and the sequence may be piecewise polynomials from a discontinuous Galerkin method [7] or reconstructed polynomials  $u^R$  in an intermediate phase [8], and even post-processed finite element solutions to achieve superconvergence [14]. Although we had discontinuous Galerkin numerical solutions in mind, the source of the problem is not important for our purpose here, as it only puts the degree of smoothness of  $u$  in perspective. Now suppose that  $u$  is in  $W_s^{p+1}(\Omega)$  (standard notation for Sobolev spaces here, supindex for the order of derivative and subindex for the  $L^s$  based space). It is natural to expect that the approximation solutions should be as smooth (in some sense) in order to achieve optimal convergence rate. The purpose of this paper is to give clear and rigorous results on this simple minded principle.

Sun [9] showed in one dimension if the mesh is uniform and the function  $u$  has  $p + 1$  weak derivatives in  $L^s$ ,  $s = 1, 2, \infty$ , then a necessary condition can be formulated. In particular in the  $s = \infty$  case, the jumps of the  $k^{th}$  derivatives, (across a mesh point) of the approximation piecewise polynomial of degree  $p$  must be less than or equal to  $\mathcal{O}(h^{p+1-k})$ ,  $0 \leq k \leq p$ . This one dimensional result is perhaps not surprising, once one realizes the interpolation error behaves in a similar way: taking a derivative, one loses a power of  $h$ , assuming  $u \in W_\infty^{p+1}$ . In the appendix of this paper, the assertion is actually proved by comparing the derivatives of  $u$ , its continuous piecewise Lagrange interpolant  $u^I$ , and  $u^R$  at a mesh point. This short proof can even be carried over to higher dimensions. Unfortunately, it cannot be

---

Received by the editors December 30, 2014.

2000 *Mathematics Subject Classification.* 65M12, 65M15, 65N30.

extended to higher dimensions when  $s = 1, 2$  due to the restriction on continuity imposed by the Sobolev imbedding theorem (See Remark 4.1 in the appendix for other reasons). Since now one starts with a function  $u \in W_s^{p+1}$ ,  $s = 1, 2$ , there are always some  $k$  and up for which the  $k$ th derivative of  $u$  at a point of interest is not well defined. On the other hand, in hindsight an idea (Lemma 2.1 below) in the much lengthier and originally unfavored proof in [9] for one dimension can be distilled and generalized to prove the two and three dimensional versions of the same principle.

While Sun *et al.* [11, 12] have successfully applied it to the analysis of numerical methods for one dimensional nonlinear conservation laws, it is quite clear that this principle has a very broad scope of applications such as safeguarding divergence or negating optimal order convergence in designing new methods, let alone in creating smoothness indicators [11, 12] in an adaptive algorithm, and so on. Being motivated by its application potential in higher dimensions, in this paper we generalize the concept of numerical smoothness of a piecewise polynomial in [9] to higher dimensions and show that in order for the convergence of  $u_h$  to  $u$  to have optimal order  $p$  in  $W_s^{p+1}$ ,  $u_h$  must have  $W_s^{p+1}$ ,  $s = 1, 2, \infty$  numerical smoothness, provided that the domain can be meshed by quasi-uniform subdivisions into triangles or quadrilaterals in 2-D and tetrahedrons or hexahedrons (cubes) in 3-D. We accounted for both interior and across interface numerical smoothness. In § 3, we formerly define the across-interface numerical smoothness in Definition 3.1, which is well motivated by the theorems in § 2 and also define interior numerical smoothness in Definition 3.2. The main result that states optimal order convergence implies numerical smoothness is proved in Theorems 3.3 and 3.4. This section is written in such a way that the reader can go read it directly after the introduction section.

The organization of rest of this paper is as follows. In § 2, we first derive basic error estimates without imposing conditions on meshes other than the shape regularity. The main theorem is Theorem 2.11, now under the quasi-uniform condition on the mesh. Finally, in § 4 we give a short proof of the one dimensional version of Theorem 3.3 and explain why it cannot be extended to higher dimensions.

## 2. Basic Estimates for Numerical Smoothness

Let  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ ,  $\alpha_i \geq 0$ ,  $1 \leq i \leq n$  be a multi-index and  $|\alpha| = \sum_i \alpha_i$ . Some of the theorems in this section have their one dimensional counterparts in [9]. We are especially inspired by the central use of Lemma 2.2 in [9]. The next lemma is its higher dimensional version, which will be used after a scaling argument back to the master element of unity size. At a certain point  $x \in \mathbb{R}^n$  of interest, e.g., a mesh nodal point, a center of a simplex (edge or face), to measure the smoothness of a mesh function  $u_h$ , we will be examining all the jumps  $[[\partial^\alpha u_h]]_x$ ,  $|\alpha| = k$  in the partial derivatives of order  $k$  for  $0 \leq k \leq p$ . In this perspective, we now state and prove the next lemma. Denote by  $\mathbb{P}_p$  the space of polynomials of total degree at most  $p$ .

**Lemma 2.1.** *Let  $\Delta = (\Delta_0, \Delta_1, \dots, \Delta_p)$ , where each  $\Delta_k$  is a vector of a certain length (e.g., it has as many components as the number of partial derivatives of order  $k$ ). Let  $\hat{\Omega}_\pm$  be two open sets in  $\mathbb{R}^n$ . Define*

$$Q(\Delta) = \min_{\hat{v} \in \mathcal{P}} \left( \left\| \hat{v} + \frac{1}{2} \sum_{\alpha \in \mathcal{I}} \frac{\Delta_\alpha}{\alpha!} \xi^\alpha \right\|_{L^2(\hat{\Omega}_-)}^2 + \left\| \hat{v} - \frac{1}{2} \sum_{\alpha \in \mathcal{I}} \frac{\Delta_\alpha}{\alpha!} \xi^\alpha \right\|_{L^2(\hat{\Omega}_+)}^2 \right),$$

where the minimum is taken over  $\mathcal{P} = \mathbb{P}_p$  in  $\xi$  and the index set

$$(1) \quad \mathcal{I} = \mathcal{I}_{\mathbb{P}} := \{\alpha : |\alpha| = k, 0 \leq k \leq p\}.$$

Then  $Q(\Delta)$  is a positive definite quadratic form in  $\Delta$ , and there exists a constant  $C_p > 0$  such that

$$(2) \quad Q(\Delta) \geq C_p \|\Delta\|^2 = C_p \sum_{i=0}^p \|\Delta_i\|^2,$$

where  $\|\Delta_i\|$  is the spectral norm of vector  $\Delta_i$ .

*Proof.* The  $n = 1$  version is in [9]. Simply notice that the minimizer  $\sum_{\alpha} V_{\alpha} \xi^{\alpha}$  is such that each  $V_{\alpha}$  is a linear combination of  $\Delta_{\beta}$ 's. Non-degeneracy comes from the fact that  $V_{\alpha} + \frac{1}{2} \frac{\Delta_{\alpha}}{\alpha!} = 0$  and  $V_{\alpha} - \frac{1}{2} \frac{\Delta_{\alpha}}{\alpha!} = 0$  implies  $\Delta_{\alpha} = 0$ .  $\square$

We will write  $Q(\Delta)$  as  $Q(\Delta_0, \Delta_1, \dots, \Delta_p)$  if this longer notation can be accommodated in display.

**Theorem 2.2.** Suppose that  $u \in H^{p+1}(\Omega)$ ,  $\Omega = (a, b)$  and  $u^R$  is a piecewise polynomial of degree  $\leq p$  with respect to a subdivision  $\{\mathcal{T}_h\}$  of subintervals  $\{I_i = (x_i, x_{i+1})\}_{i=0}^N$ . Let  $h = \max_i |I_i|$  and  $h_{\min} = \min_i |I_i|$ . Then there exists a positive constant  $C_1$  independent of  $h, u$  and  $u^R$  such that

$$(3) \quad \|u - u^R\|_{L^2(\Omega)} \geq C_1 h^{p+1} \left( \sqrt{\sum_{i=1}^N h_{\min}^n \|\tilde{D}_i\|^2} - |u|_{H^{p+1}(\Omega)} \right),$$

where  $n = 1$  and the components of  $\tilde{D}_i$  are

$$(4) \quad \tilde{D}_i^{\alpha} = J_i^{(\alpha)} / (h^{p+1} h_{\min}^{-|\alpha|}), \quad J_i^{(\alpha)} = \llbracket \partial^{\alpha} u^R \rrbracket_{x_i}, \quad |\alpha| = k, 0 \leq k \leq p.$$

*Remark 2.3.* In the above statement, we used higher dimension notation for 1-D case so that its extension to higher dimensions is more clear and straightforward. We will do the same in the proof below.

*Proof.* Let  $\mathcal{T}_h$  be a partition  $x_0 = a < x_1 < x_2 < \dots < x_N < b = x_{N+1}$  on  $(a, b)$ . Define a new partition by adding the midpoint  $x_{i+1/2}$ , center of each element  $(x_i, x_{i+1})$ , to the old partition. We call  $(x_{i-1/2}, x_i) \cup (x_i, x_{i+1/2}) \cup \{x_i\}$  the covolume associated with  $x_i$ . The closure of all covolumes associated with interior nodes, together with the two boundary covolumes  $(x_0, x_{1/2})$  and  $(x_{N+1/2}, x_{N+1})$  form a dual mesh  $\{\mathcal{T}_h^*\}$  to  $\mathcal{T}_h$ , which we call dual covolume mesh. Let  $\mathbb{P}_h^*$  be the space of piecewise polynomials of degree  $\leq p$  with respect to this dual mesh. Since  $u \in H^{p+1}(\Omega)$ , there exists a  $u^I \in \mathbb{P}_h^*$  such that

$$(5) \quad \|u - u^I\|_{L^2(\Omega)} \leq C_2 h^{p+1} |u|_{H^{p+1}(\Omega)}.$$

On the other hand, by the triangle inequality

$$\begin{aligned} \|u - u^R\|_{L^2(\Omega)} &\geq \|u^I - u^R\|_{L^2(\Omega)} - \|u - u^I\|_{L^2(\Omega)} \\ &\geq \sqrt{\sum_{i=1}^N \|u^I - u^R\|_{L^2(\Omega_i^*)}^2} - C_2 h^{p+1} |u|_{H^{p+1}(\Omega)}, \end{aligned}$$

where  $\Omega_i^*$  is a smaller subset of the covolume associated with  $x_i$ . Now let's choose  $\Omega_i^*$ . At each  $x_i$ ,  $1 \leq i \leq N$ , we take a 1-D ball  $\Omega_i^* = \{|x - x_i| \leq \delta\}$  so small so that it is in  $(x_{i-1}, x_{i+1})$ . It is essential for the later use of Lemma 2.1 that  $\delta$  works for all  $i$ . We take  $\delta = \frac{1}{4} h_{\min}$ .

Let  $\{q\}$  denote the average of  $q^+$  and  $q^-$  and let  $\llbracket q \rrbracket = q^+ - q^-$  denote the jump. Then it is trivial that

$$(6) \quad q^+ - \{q\} = \frac{1}{2}\llbracket q \rrbracket \quad \text{and} \quad q^- - \{q\} = -\frac{1}{2}\llbracket q \rrbracket.$$

Applying this with  $q = q_i^k = \frac{d^k u^R}{dx^k}(x_i)$ , the discontinuous  $k$ th derivative of  $u^R$  at  $x_i$ , and letting  $w(x) = \sum_{k=0}^p \frac{\{q_i^k\}}{k!}(x - x_i)^k$ , we have with  $J_i^{(k)} := \llbracket \frac{d^k u^R}{dx^k} \rrbracket_{x_i}$  that

$$u^R - w = \frac{1}{2} \sum_{k=0}^p \frac{J_i^{(k)}}{k!} (x - x_i)^k, \quad \forall x \in \Omega_{i,+}^* := (x_i, x_i + \delta)$$

and

$$u^R - w = -\frac{1}{2} \sum_{k=0}^p \frac{J_i^{(k)}}{k!} (x - x_i)^k, \quad \forall x \in \Omega_{i,-}^* := (x_i - \delta, x_i).$$

Now using the change of variable  $\xi = (x - x_i)/h_{\min}$  below, we have

$$(7) \quad \begin{aligned} \|u^I - u^R\|_{L^2(\Omega_i^*)}^2 &\geq \min_{v \in \mathbb{P}_p} \|v - u^R\|_{L^2(\Omega_i^*)}^2 \\ &= \min_{v \in \mathbb{P}_p} \|v - (u^R - w)\|_{L^2(\Omega_i^*)}^2 \\ &= \min_{v \in \mathbb{P}_p} \left( \left\| v + \frac{1}{2} \sum_{k=0}^p \frac{J_i^{(k)}}{k!} (x - x_i)^k \right\|_{L^2(\Omega_{i,-}^*)}^2 + \left\| v - \frac{1}{2} \sum_{k=0}^p \frac{J_i^{(k)}}{k!} (x - x_i)^k \right\|_{L^2(\Omega_{i,+}^*)}^2 \right) \\ &= h_{\min}^n \min_{\hat{v} \in \hat{\mathbb{P}}_p} \left( \left\| \hat{v} + \frac{1}{2} \sum_{k=0}^p \frac{J_i^{(k)}}{k!} (\xi h_{\min})^k \right\|_{L^2(\hat{\Omega}_-)}^2 + \left\| \hat{v} - \frac{1}{2} \sum_{k=0}^p \frac{J_i^{(k)}}{k!} (\xi h_{\min})^k \right\|_{L^2(\hat{\Omega}_+)}^2 \right), \end{aligned}$$

where  $\hat{\Omega}_- = (-\frac{1}{4}, 0)$ ,  $\hat{\Omega}_+ = (0, \frac{1}{4})$

$$(8) \quad \begin{aligned} &= (h_{\min}^n h^{2p+2}) \min_{\hat{v} \in \hat{\mathbb{P}}_p} \left( \left\| \hat{v} + \frac{1}{2} \sum_{k=0}^p \frac{\tilde{D}_i^{(k)}}{k!} \xi^k \right\|_{L^2(\hat{\Omega}_-)}^2 + \left\| \hat{v} - \frac{1}{2} \sum_{k=0}^p \frac{\tilde{D}_i^{(k)}}{k!} \xi^k \right\|_{L^2(\hat{\Omega}_+)}^2 \right), \\ &\quad (n = 1) \\ &= h_{\min}^n h^{2p+2} Q(\tilde{D}_i^0, \tilde{D}_i^1, \dots, \tilde{D}_i^p). \end{aligned}$$

(Note that the minimization-range's change to  $\mathbb{P}_p$  in (7) was possible due to the fact that  $u^I \in \mathbb{P}_h^*$  is a single piece of polynomial on the covolume.)

Now invoking (2) on

$$(9) \quad \|u^I - u^R\|_{L^2(\Omega_i^*)}^2 \geq h_{\min}^n h^{2p+2} Q(\tilde{D}_i^0, \tilde{D}_i^1, \dots, \tilde{D}_i^p).$$

completes the proof.  $\square$

Next we move to the 2-D case. For a given  $\{\mathcal{T}_h\}$ , we will use a dual mesh  $\{\mathcal{T}_h^*\}$  consisting of covolumes as in Chou and Vassilevski [5]. Covolumes are obtained by adding a point to an old element in  $\{\mathcal{T}_h\}$  and connect it with the vertices of the element. The covolumes around an edge or associated with the midpoint of the edge is shown in Figure 1 (see the diamond  $STVU$  associated with  $x_i$ ). Note that a covolume is obtained by connecting vertices with an added point, which could be a circumcenter or a barycenter for triangular grids and intersection of diagonals for

quadrilateral grids. We can now summarize the most important ingredients of the 1-D proof.

- (i) *Ensure that the optimal order estimate (5) in the  $L^2$  norm holds on the dual covolume mesh.*

In contrast, unlike in 1-D the dual mesh shape condition comes into play as well. In higher dimensions, the construction of the dual covolume mesh in a symmetric and smooth way causes the shape regularity of the primary mesh to be inherited. Furthermore the dual mesh of a triangular or quadrilateral mesh in the covolume construction is almost a quadrilateral mesh (boundary covolumes are triangles), but we will not need the boundary covolumes in our analysis below. In 3-D a covolume is the union of two tetrahedrons when the primary mesh is tetrahedral. Here the approximation order is usually achieved by the existence of a good local  $L^2$  projection-type interpolation operator. See Girault and Raviart [6, pp. 101-109] for such operators.

- (ii) *Ensure the choice of radius  $\delta$  is such that  $\delta/h_{min}$  is a constant, independent of  $h$ .*

This scales  $\Omega_i^*$  to unit size so that Lemma 2.1 can be used to extract a lower bound with the constant independent of  $h$ . Of course in this step the  $\Omega_i^*$ s are automatically non-overlapping due to the covolume construction.

- (iii) *The order of approximation on the dual mesh limits the extracting power.*

Note that what power of  $h$  in (8) to extract is determined by how well the approximation on the dual mesh can be done. The optimal case is  $h^{p+1}$ .

With this in mind, we now prove the corresponding theorem for triangular and quadrilateral meshes. We will denote by  $\mathbb{P}_p^h$  for the piecewise polynomial space associated with  $\mathbb{P}_p$ . The regularity of a family of quadrilateral subdivisions is defined as follows [6, p. 104]. Let  $Q$  be a quadrilateral with four vertices  $\mathbf{x}_i$  and denote by  $S_i$  the subtriangle of  $Q$  with vertices  $\mathbf{x}_{i-1}, \mathbf{x}_i$  and  $\mathbf{x}_{i+1}$  ( $\mathbf{x}_0 = \mathbf{x}_4$ ). Let  $h_Q$  be the diameter of  $Q$  and  $\rho_Q = 2 \min_{1 \leq i \leq 4} \{ \text{diameter of circle inscribed in } S_i \}$ . A family of quadrilateral partitions  $\{Q_h\}$  is said to be regular if there exists a positive constant  $\sigma$ , independent of  $h$ , such that

$$(10) \quad \frac{h_Q}{\rho_Q} \leq \sigma, \quad \forall Q \in Q_h, \forall Q_h \in Q.$$

There are equivalent definitions [4].

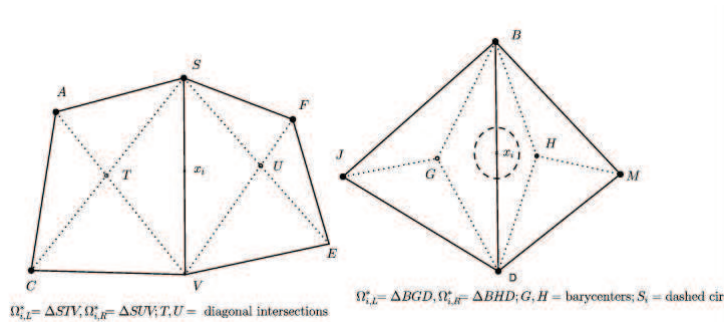


FIGURE 1. Quadrilateral mesh with its covolumes ( $STVU$ ) and Triangular mesh with its covolumes  $BGDH$  and selected disk.

**Theorem 2.4.** *Let  $u \in H^{p+1}(\Omega)$ ,  $\Omega \subset \mathbb{R}^2$ . In addition, let the following assumption hold.*

H1.  $u^R \in \mathbb{P}_p^h$  on a family of regular subdivisions  $\{\mathcal{T}_h\}$  by triangles or by quadrilaterals, i.e.,  $u^R \in \mathbb{P}_p$  on each element.

Then there exists a positive constant  $C_1$  independent of  $h, u$  and  $u^R$  such that

$$\|u - u^R\|_{L^2(\Omega)} \geq C_1 h^{p+1} \left( \sqrt{\sum_{i=1}^{N_e^\circ} h_{\min}^n \|\tilde{D}_i\|^2} - |u|_{H^{p+1}(\Omega)} \right),$$

where  $n = 2$  and the components of  $\tilde{D}_i$  are

$$(11) \quad \tilde{D}_i^\alpha = \frac{J_i^{(\alpha)}}{h^{p+1} h_{\min}^{-|\alpha|}}, \quad J_i^{(\alpha)} = \llbracket \partial^\alpha u^R \rrbracket_{x_i}, \quad |\alpha| = k, \quad 0 \leq k \leq p.$$

Here  $N_e^\circ$  is the number of interior edges and  $h_{\min}$  is the least edge length.

*Proof.* We will proceed as in 1-D case, pointing out the difference along the way.

*Case 1: Triangular mesh.*

For each  $\mathcal{T}_h$  we define a dual mesh  $\mathcal{T}_h^*$  formed as follows. With reference of Figure 1, in each element we connect the vertices (e.g.,  $B, M, D$ ) with a newly added point, which in this case is the barycenter (e.g.,  $H$ ), to create three new triangles. The two half-covolumes (e.g.  $BGD, BHD$ ) form a single covolume  $BGDH$  for the common edge  $BD$ . All covolumes form the dual mesh  $\mathcal{T}_h^*$ . We can find a  $u^I \in \mathbb{P}_h^*$  so that (5) holds under no regularity conditions on the dual mesh by quadrilaterals. The  $u^I$  is the local  $L^2$  projection and estimate (5) can be found in [6, p. 108].

Let  $x_i$  be the midpoint of an interior edge  $e_i$  common to two half-covolumes  $\Omega_{i,L}^*, \Omega_{i,R}^*$ . In Figure 1,  $\Omega_{i,L}^* = \triangle BGD, \Omega_{i,R}^* = \triangle BHD$ . Let us take  $\Omega_i^* = S_i$  to be an open disk with center  $x_i$  and a radius  $\delta$  small enough so that  $S_i$  is fully contained in the interior of  $\tilde{\Omega}_{i,L}^* \cup \tilde{\Omega}_{i,R}^*$ . The radius however has to work for all midpoints  $x_i$  on interior edges. It is well known that shape regularity is equivalent to the minimal angle condition, and consequently there is a constant  $\theta_0$  such that all interior angles  $\theta \geq \theta_0$  for all  $h$ . Without loss of generality, suppose the distance from  $x_i$  to the boundary of  $\tilde{\Omega}_{i,L}^* \cup \tilde{\Omega}_{i,R}^*$  is attained by  $|\overline{x_i F}|$ , where the foot  $F$  is on  $\overline{BH}$ . Then

$$|\overline{x_i F}| = |\overline{Bx_i}| \sin \frac{1}{2} \angle MBD \geq |\overline{Bx_i}| \sin \frac{\theta_0}{2} \geq \frac{1}{2} h_{\min} \sin \frac{\theta_0}{2},$$

where we have used the fact that the sine function is increasing on  $[0, \frac{\pi}{2}]$  and that  $BH$  is a bi-angle line. Thus it suffices to take  $\delta = \frac{1}{4} h_{\min} \sin \frac{\theta_0}{2}$  as the common radius. The rest of proof is just like 1-D case. For example, now

$$(12) \quad u^R - w = \frac{1}{2} \sum_{k=0}^p \sum_{|\alpha|=k} \frac{J_i^{(\alpha)}}{\alpha!} (x - x_i)^\alpha, \quad \forall x \in \Omega_{i,+}^*,$$

and

$$(13) \quad u^R - w = -\frac{1}{2} \sum_{k=0}^p \sum_{|\alpha|=k} \frac{J_i^{(\alpha)}}{\alpha!} (x - x_i)^\alpha, \quad \forall x \in \Omega_{i,-}^*.$$

Furthermore, from the validation of

$$\begin{aligned} \|u^I - u^R\|_{L^2(\Omega_i^*)}^2 &\geq \min_{v \in \mathbb{P}_p} \|v - u^R\|_{L^2(\Omega_i^*)}^2 \\ &= \min_{v \in \mathbb{P}_p} \|v - (u^R - w)\|_{L^2(\Omega_i^*)}^2 \end{aligned}$$

and use of  $\xi = (x - x_i)/h_{\min}$  and Lemma 2.1, we can derive as in the corresponding 1-D case that

$$\|u^I - u^R\|_{L^2(\Omega_i^*)}^2 \geq h_{\min}^n h^{2p+2} Q(\tilde{D}_i^0, \tilde{D}_i^1, \dots, \tilde{D}_i^p), \quad n = 2.$$

*Case 2: Quadrilateral mesh under assumption H1.*

First, optimal order estimate (5) holds with the local  $L^2$  projection as before. Equations (12)-(13) are valid as well. So the only concern is the choice of  $\delta$ . Since the local geometry in the left figure of Figure 1 is the same as in the triangular case, we would still need the minimal angle condition. However, it is shown in Theorem 4.1 of Chou and He [4], regularity of the quadrilateral meshes defined in [6] implies the minimal angle condition: all interior angles of quadrilaterals and the interior angles of the subtriangles  $S_i$  in (10) are bounded below, although the converse is not true. On the other hand, suppose the shortest distance from  $x_i$  to the covolume edges is attained by  $F$  on  $\overline{SU}$ . It should be clear that  $\angle VSU$  can never exceed 90 degrees as well. Thus, we can still take the same  $\delta$ .

This completes the proof.  $\square$

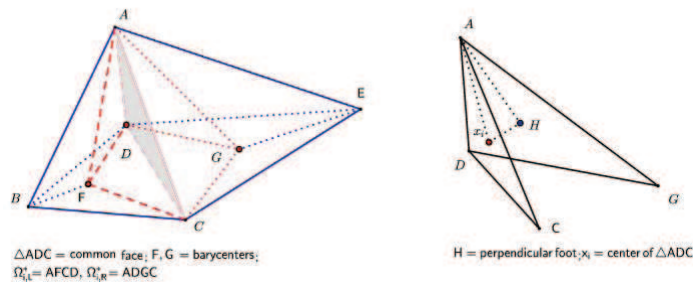


FIGURE 2. Tetrahedral mesh with a typical covolume (tetrahedron  $ADCG$  union tetrahedron  $ADCF$ ) and local configuration around a center point in common face.

We now prove the 3-D version of the previous theorem. In 3-D, regularity of tetrahedral subdivision is still defined in terms of the uniform boundedness of  $h_K/\rho_K$ , the ratio of the maximum diameter  $h_K$  to  $\rho_K$ , the diameter of the inscribed sphere. It is shown in Brandts *et al.* [2] that this condition is equivalent to the minimal angle condition, but now two types of angles are involved, angles between edges and between faces. We state this equivalence in the next lemma.

**Lemma 2.5.** *Let  $\{\mathcal{T}_h\}_{h>0}$  be a family of tetrahedral subdivisions for a domain in  $\mathbb{R}^3$ . Then  $\{\mathcal{T}_h\}_{h>0}$  being regular is equivalent to the minimal angle condition: there exists a constant  $\beta_0 > 0$  such that for any partition  $\mathcal{T}_h$ , any tetrahedron  $T \in \mathcal{T}_h$ , and any dihedral angle (angle between faces) or solid angle (angle between edges)  $\beta$  of  $T$ , we have*

$$(14) \quad \beta \geq \beta_0.$$

**Theorem 2.6.** *Let  $u \in H^{p+1}(\Omega)$ ,  $\Omega \subset \mathbb{R}^3$ . In addition, let the following assumption hold.*

- A1.  $u^R \in \mathbb{P}_p^h$  on a family of regular subdivisions  $\{\mathcal{T}_h\}$  by tetrahedrons or hexahedrons. i.e.,  $u^R \in \mathbb{P}_p$  on each element.

Then there exists a positive constant  $C_1$  independent of  $h, u$  and  $u^R$  such that

$$\|u - u^R\|_{L^2(\Omega)} \geq C_1 h^{p+1} \left( \sqrt{\sum_{i=1}^{N_f^o} h_{\min}^n \|\tilde{D}_i\|^2} - |u|_{H^{p+1}(\Omega)} \right),$$

where  $n = 3$  and the components of  $\tilde{D}_i$  are

$$(15) \quad \tilde{D}_i^\alpha = \frac{J_i^{(\alpha)}}{h^{p+1} h_{\min}^{-|\alpha|}}, \quad J_i^{(\alpha)} = \llbracket \partial^\alpha u^R \rrbracket_{x_i}, \quad |\alpha| = k, \quad 0 \leq k \leq p.$$

Here  $N_f^o$  is the number of interior faces and  $h_{\min}$  is the least edge length.

*Proof.* We consider the tetrahedral mesh case first. In Figure 2,  $\triangle ADC$  is a typical common tetrahedral element interface with the accompanying half-covolumes  $\Omega_{i,L}^* = AFCD$  and  $\Omega_{i,R}^* = AGCD$ . The barycenter of  $\triangle ADC$  is  $x_i$  (not labeled to avoid clustering). We need to choose the common radius  $\delta$  of the sphere centered at  $x_i$  that works for all  $i$ . Assume that the shortest distance from  $x_i$  to the boundary of covolumes is attained by the plane containing  $\triangle ADG$ . In the right side of Figure 2, a blowup of the situation is shown, the perpendicular foot from  $x_i$  is  $H$ . Let  $A_M$  be the midpoint of  $CD$  and let  $A_\perp$  be the foot on  $CD$  of the perpendicular from  $A$ , then

$$(16) \quad |\overline{AA_M}| \geq |\overline{AA_\perp}| = |\overline{AC}| \sin \angle ACD \geq h_{\min} \sin \beta_0,$$

where  $\beta_0$  is the common lower bound for angles between edges in the minimal angle condition.

Now let  $K$  be the perpendicular foot on  $\overline{AD}$  from  $x_i$  (see Figure 3 for the blowup version) and let  $\theta_s = \angle HKx_i$ . In general  $\theta_s$  is not equal to  $\theta_f$ , the (dihedral) angle between planes determined by  $ADC$  and  $ADG$  unless  $\overline{HK} \perp \overline{AD}$ . Since  $x_i \overline{H}$  is normal to plane  $ADG$  and  $x_i \overline{K}$  is normal to  $\overline{AD}$ , the two angles are expected to be related. This is indeed the case. In fact,

$$(17) \quad \cos \theta_f = \frac{1}{3} \cos \theta_s,$$

which will be proved in Lemma 2.7 below. Concentrating on  $\triangle Ax_i D$ , we see that

$$\begin{aligned} |\overline{Kx_i}| &= |\overline{Ax_i}| \sin \angle x_i AD = |\overline{Ax_i}| \sin \frac{1}{2} \angle CAD \\ &= \frac{2}{3} |\overline{AA_M}| \sin \frac{1}{2} \angle CAD \\ &\geq \frac{2}{3} h_{\min} \sin \beta_0 \sin \frac{\beta_0}{2}, \end{aligned}$$

where we used (16) in the last inequality.

With reference to Figure 4 (local blowup version of the left figure in Figure 1), let  $\theta_{ext} = \angle(ADC, ADE)$ , the angle between the planes  $ADC$  and  $ADE$ . Then there exists a constant  $\gamma > 1$  independent of  $h$  such that

$$(18) \quad \cos \theta_f \leq \gamma \cos \theta_{ext},$$

whose proof is given in Lemma 2.8 below.



Letting  $g(\theta_f) := \sin(\cos^{-1}(3 \cos \theta_f))$  (an increasing function on  $[0, \pi/2]$ ), and using (17)–(18), we have

$$\begin{aligned} \delta &= |\overline{x_i H}| = |\overline{K x_i}| \sin \theta_s \\ &\geq \frac{2}{3} h_{\min} \sin \frac{\beta_0}{2} \sin \beta_0 g(\theta_f) \\ &\geq \frac{2}{3} h_{\min} \sin \frac{\beta_0}{2} \sin \beta_0 g(\cos^{-1} \gamma \cos \theta_{ext}) \\ &\geq \frac{2}{3} h_{\min} \sin \frac{\beta_0}{2} \sin \beta_0 g(\cos^{-1} \gamma \cos \beta_0), \end{aligned}$$

where we have used the minimal angle condition. Thus we can take  $\delta$  to be half of the last expression in the above inequality. The rest of the proof is as before.

As for hexahedral case, its local geometry around  $x_i$  is similar to the tetrahedral case. Similar comments made for quadrilateral case apply here too. This completes the proof.  $\square$

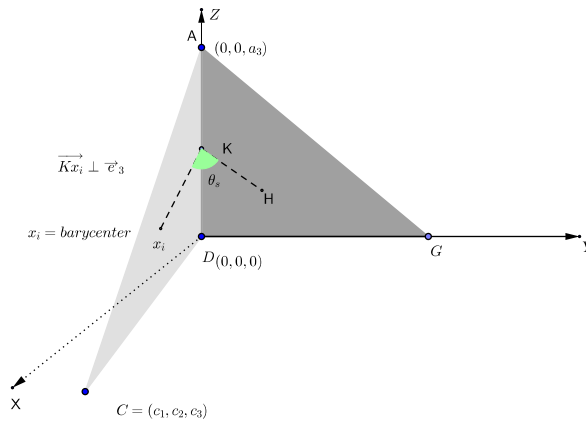


FIGURE 3. If  $\theta_f =$  dihedral angle between  $YZ$ -plane and  $\triangle ADC$ ,  $\theta_s = \angle HK x_i$ , then  $\cos \theta_f = 1/3 \cos \theta_s$ .

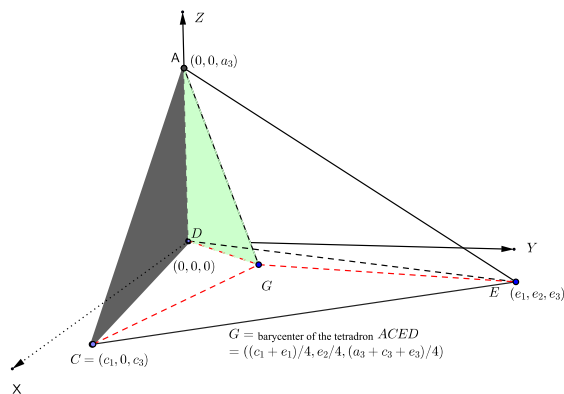


FIGURE 4. If  $\theta_f = \angle(ADC, ADG)$  and  $\theta_{ext} = \angle(ADC, ADE)$ , then  $\theta_f < \theta_{ext}$ .

**Lemma 2.7.** *With reference to Figure 3, let  $AD$  be the intersection of two triangles  $ADC$  and  $ADG$  in  $\mathbb{R}^3$  and denote by  $\theta_f = \angle(ADC, ADG)$ , the angle between them (angle between their normals). Let  $x = x_i$  be the barycenter of  $\triangle ADC$ ,  $K \in \overline{AD}$ ,  $\overline{xK} \perp \overline{AD}$  and  $H \in \triangle ADG$  such that  $\overline{xH} \perp \triangle ADG$ . Let  $\theta_s = \angle HKx$ . Then*

$$\cos \theta_f = \frac{1}{3} \cos \theta_s.$$

*Proof.* By the properties of  $x$ ,  $H$  and  $K$ , one derives

$$x = \left(\frac{c_1}{3}, \frac{c_2}{3}, \frac{c_3 + a_3}{2}\right), \quad H = \left(0, \frac{c_2}{3}, \frac{c_3 + 1}{3}\right), \quad K = \left(0, 0, \frac{a_3 + c_3}{3}\right).$$

Let  $n_1 = (1, 0, 0)$  and  $n_2 = \overline{xK} \times \overline{xC}$  normalized. Using  $\overline{KH}$  and  $\overline{Kx}$  to compute  $\cos \theta_s$  and using  $\cos \theta_f = n_1 \cdot n_2$ , after some computation we get  $\cos \theta_f = 1/3 \cos \theta_s$ .  $\square$

**Lemma 2.8.** *With reference to Figure 4, let  $AD$  be the intersection of two triangles  $ADC$  and  $ADG$  in  $\mathbb{R}^3$  and let*

$$\theta_f = \angle(ADC, ADG), \quad \theta_{ext} = \angle(ADC, ADE).$$

*Then there exists a constant  $\gamma > 0$  independent of  $h$  such that*

$$(19) \quad \cos \theta_{ext} < \cos \theta_f \leq \gamma \cos \theta_{ext}.$$

*Proof.* The lemma is stated and proved in the context of the main theorem and hence Figure 4 configuration is assumed. Since  $G$  is the barycenter, it is the average of four vertices of  $ADCE$ ,

$$G = \left(\frac{c_1 + e_1}{4}, \frac{e_2}{4}, \frac{a_3 + c_3 + e_3}{4}\right),$$

and three normal vectors (some not normalized) to the planes from left to right in Figure 4 are

$$n_1 = -\vec{j}, \quad n_2 = -e_2 \vec{i} + (c_1 + e_1) \vec{j}, \quad n_3 = e_2 \vec{i} - e_1 \vec{j}.$$

Thus

$$\cos \theta_f = \frac{c_1 + e_1}{\sqrt{(c_1 + e_1)^2 + e_2^2}}, \quad \cos \theta_{ext} = \frac{e_1}{\sqrt{e_1^2 + e_2^2}}.$$

In our figure, we have  $c_1 > 0$  and  $e_1 > 0$ , which is the case for the shortest distance case. Note that  $f(c) = \frac{c+e_1}{\sqrt{(c+e_1)^2 + e_2^2}}$  is increasing on  $[0, \infty]$  and so  $f(c_1) \geq f(0)$  implies  $\cos \theta_f \geq \cos \theta_{ext}$ . On the other hand,  $c_1$  is an edge size and so we can assume it is bounded by 1. Thus we use  $f(c_1) \leq f(1)$ . In turn since  $f(1)$  is comparable to  $\frac{e_1}{\sqrt{e_1^2 + e_2^2}}$ , we see that there exists a  $\gamma$  independent of  $h$  such that (19) holds.  $\square$

**Theorem 2.9.** *Suppose that*

$$H0: u \text{ is in } W_1^{p+1}(\Omega), \Omega \subset \mathbb{R}^n,$$

*where  $n = 2$  and that assumption H1 of Theorem 2.4 holds. Then there are constants  $C_1 > 0$  independent of  $h, u$  and  $u^R$  such that*

$$(20) \quad \|u - u^R\|_{L^1(\Omega)} \geq C_1 h^{p+1} \left[ \sum_{1 \leq i \leq N_e} h_{\min}^n \|\tilde{D}_i\| - |u|_{W_1^{p+1}(\Omega)} \right],$$

*where  $\tilde{D}_i$  has components*

$$\tilde{D}_i^\alpha = J_i^{(\alpha)} / (h^{p+1} h_{\min}^{-|\alpha|}), \quad J_i^{(\alpha)} = [[\partial^\alpha u^R]]_{x_i}, \quad |\alpha| = k, \quad 0 \leq k \leq p.$$

Suppose assumptions H0 and A1 of Theorem 2.6 holds for  $n = 3$ . Then (20) holds with  $N_e^\circ$  replaced by  $N_f^\circ$ .

*Proof.* As before, for  $u \in W_1^{p+1}(\Omega)$ , there is a  $u^I \in \mathbb{P}_h^*$  so that

$$\|u - u^I\|_{L^1(\Omega)} \leq Ch^{p+1}|u|_{W_1^{p+1}(\Omega)},$$

where  $C > 0$  is a constant independent of  $u$  and  $h$ . Now by the triangle inequality

$$\begin{aligned} \|u - u^R\|_{L^1(\Omega)} &\geq \|u^I - u^R\|_{L^1(\Omega)} - \|u - u^I\|_{L^1(\Omega)} \\ &\geq \sum_{i=1}^{N_e^\circ} \|u^I - u^R\|_{L^1(\Omega_i^*)} - C_2 h^{p+1} |u|_{W_1^{p+1}(\Omega)}. \end{aligned}$$

By the standard scaling argument, and (9) or the argument leading to it, we have

$$\|u^I - u^R\|_{L^1(\Omega_i^*)} \geq C_2 h_{\min}^{n/2} \|u^I - u^R\|_{L^2(\Omega_i^*)} \geq C_2 h_{\min}^n h^{p+1} \sqrt{Q(\tilde{D}_i^0, \tilde{D}_i^1, \dots, \tilde{D}_i^p)}.$$

Now invoking (2) and taking minimum of the constants completes proof for the two dimensional case.  $\square$

**Theorem 2.10.** *Suppose*

$$H_\infty : \quad u \in W_\infty^{p+1}(\Omega), \quad \Omega \subset \mathbb{R}^n,$$

where  $n = 2$ . In addition, H1 of Theorem 2.4 holds. Then there exists a constant  $C_\infty > 0$ , independent of  $h, u$  and  $u^R$ , such that

$$(21) \quad \|u - u^R\|_{L^\infty(\Omega)} \geq C_\infty h^{p+1} \left[ \left( \frac{h_{\min}}{h} \right)^{n/2} \max_{1 \leq i \leq N_e^\circ} \|\tilde{D}_i\| - |u|_{W_\infty^{p+1}(\Omega)} \right],$$

where  $n = 2$  and  $\tilde{D}_i$  has components

$$\tilde{D}_i^\alpha = J_i^{(\alpha)} / (h^{p+1} h_{\min}^{-|\alpha|}), \quad |\alpha| = k, \quad 0 \leq k \leq p.$$

Suppose assumptions H0 and A1 of Theorem 2.6 hold for  $n = 3$ . Then (21) holds. Now  $x_i$  is the face center and  $N_e^\circ$  replaced by  $N_f^\circ$ .

*Proof.* Since  $u \in W_\infty^{p+1}(\Omega)$ , there exists a  $u^I \in \mathbb{P}_h^*$  such that

$$\|u - u^I\|_{L^\infty(\Omega)} \leq C_1 h^{p+1} |u|_{W_\infty^{p+1}(\Omega)}.$$

Now by the triangle inequality

$$\begin{aligned} \|u - u^R\|_{L^\infty(\Omega)} &\geq \|u^I - u^R\|_{L^\infty(\Omega)} - \|u - u^I\|_{L^\infty(\Omega)} \\ &\geq \max_i \|u^I - u^R\|_{L^\infty(\Omega_i^*)}^2 - C_2 h^{p+1} |u|_{W_\infty^{p+1}(\Omega)}. \end{aligned}$$

Let  $U_i = \|u^I - u^R\|_{L^\infty(\Omega_i^*)}$  and use (9) to derive

$$\begin{aligned} U_i^2 &= \frac{1}{|\Omega_i^*|} \int_{\Omega_i^*} U_i^2 dx \geq \frac{1}{|\Omega_i^*|} \int_{\Omega_i^*} (u^I - u^R)^2 dx \\ &\geq Ch^{-n} \|u^I - u^R\|_{L^2(\Omega_i^*)}^2 \geq (h_{\min}/h)^n h^{2p+2} Q(\tilde{D}_i^0, \tilde{D}_i^1, \dots, \tilde{D}_i^p). \end{aligned}$$

Invoking (2) and taking a common minimum constant completes the proof.  $\square$

Now we impose quasi-uniform conditions on the meshes to get the next theorem.

**Theorem 2.11.** *Suppose that*

$$H0. \quad u \in W_s^{p+1}(\Omega), \quad \Omega \subset \mathbb{R}^n, \quad n = 2.$$

*Suppose the following assumption holds.*

H1.  $u^R \in \mathbb{P}_p^h$  on a quasi-uniform family of subdivisions  $\{\mathcal{T}_h\}$  by triangles or by quadrilaterals, i.e.,  $u^R \in \mathbb{P}_p$  on each element.

Then

(i) in case  $s = 1, 2$ , there exists a positive constant  $C_1$  independent of  $h, u$  and  $u^R$  such that

$$(22) \quad \|u - u^R\|_{L^s(\Omega)} \geq C_1 h^{p+1} \left( \left( \sum_{i=1}^{N_e^\circ} h^n \|D_i\|^s \right)^{1/s} - |u|_{W_s^{p+1}(\Omega)} \right), n = 2,$$

(ii) in case  $s = \infty$ , there exists a constant  $C_\infty > 0$ , independent of  $h, u$  and  $u^R$ , such that

$$\|u - u^R\|_{L^\infty(\Omega)} \geq C_\infty h^{p+1} \left[ \max_{1 \leq i \leq N_e^\circ} \|D_i\| - |u|_{W_\infty^{p+1}(\Omega)} \right],$$

where the components of  $D_i$  are

$$D_i^\alpha = J_i^{(\alpha)} / (h^{p+1-|\alpha|}), \quad J_i^{(\alpha)} = \llbracket \partial^\alpha u^R \rrbracket_{x_i}, \quad |\alpha| = k, \quad 0 \leq k \leq p.$$

Suppose assumptions H0 and A1 of Theorem 2.6 hold for  $n = 3$ . Then assertions (i) and (ii) hold. Now  $x_i$  is a face center and  $N_e^\circ$  replaced by  $N_f^\circ$ .

*Proof.* Note that since  $\tilde{D}_i^\alpha = D_i^\alpha (\frac{h_{\min}}{h})^k$ ,  $\|\tilde{D}_i\|^2 = \sum_{k=0}^p (D_i^\alpha)^2 (\frac{h_{\min}}{h})^{2k}$ . By quasi-uniformness,  $h/h_{\min}$  is uniformly bounded above and we can replace all occurrences of  $h_{\min}$  by  $Ch$  in all the previous theorems. This completes the proof.  $\square$

### 3. Optimal order convergence implies numerical smoothness

We present this section independently from other sections so that the reader can read it directly. There might be some repetitions of the already introduced notations.

**Definition 3.1. Type A Numerical Smoothness.** Let  $u_h \in \mathbb{P}_p^h$  be a piecewise polynomial of degree  $\leq p$  with respect to a family of subdivisions  $\{\mathcal{T}_h\}$  on  $\Omega \subset \mathbb{R}^n$  by  $n$ -simplices and the alikes (quadrilateral, hexahedrons etc.). Let  $\{x_i\}_{i=1}^{N_e^\circ}$  be the set of all midpoints of interior edges for  $n = 2$  and barycenters of interior faces for  $n = 3$ . Then  $u_h$  is said to be  $W_s^{p+1}(\Omega)$ -smooth of Type A,  $s \geq 1$ , if there is a constant  $C_s$ , independent of  $h$  and  $u_h$ , such that

$$(23) \quad \sum_{i=1}^{N_e^\circ} h^n \|D_i\|^s \leq C_s,$$

and  $W_\infty^{p+1}(\Omega)$ -smooth, if there exists a constant  $C_\infty$  independent of  $h$  and  $u_h$  such that

$$(24) \quad \max_{1 \leq i \leq N_e^\circ} \|D_i\| \leq C_\infty,$$

where the components of  $D_i$  are the scaled jumps of partial derivatives

$$D_i^\alpha = J_i^{(\alpha)} / (h^{p+1-|\alpha|}), \quad J_i^{(\alpha)} = \llbracket \partial^\alpha u_h \rrbracket_{x_i}, \quad |\alpha| = k, \quad 0 \leq k \leq p.$$

The A in Type A stands for (a)cross the interface as opposed to the Type I (interior) smoothness below. As pointed out before, Definition 3.1 of numerical smoothness was first introduced in [9] for  $n = 1$ . It is also worth noting that for the  $n = 1$  case, several natural conditions for optimal convergence are already included. These include that the scaled functional value  $|D_i^0| \leq C$  for all  $i$  in the case of  $k = 0$ ,

and at the other end in the case of  $k = p$  that  $|D_i^p| \leq C$  or (23) with  $s = 1$  implies the piecewise constant function  $\frac{d^p u_h}{dx^p}$  has bounded variation.

Intuitively, the smoothness of a numerical solution  $u_h \in \mathbb{P}_p^h$  should be measured by the boundedness of partial derivatives  $\partial^\alpha u_h$ . On an element  $T \in \mathcal{T}_h$ , by Taylor expansion around any point  $x_m$  in  $T$ , e.g., the center of  $T$  or a point on the boundary of  $T$  using one-sided derivatives, we see that the quantities  $\partial^\alpha u_h(x_m)$  would be sufficient to give information on the interior smoothness. In other words, for this part of smoothness we need a constant  $M > 0$ , independent of  $h$ , such that

$$(25) \quad M_m^\alpha := |\partial^\alpha u_h(x_m)| \leq M, \quad \forall |\alpha| = k, \quad 0 \leq k \leq p.$$

On the other hand, the smoothness across the interface boundary of an element, by common sense, should be measured by the jumps of partials  $J_i^{(\alpha)}$ . The crucial part of Definition 3.1 is to point out that this intuition needs to be adjusted and that the quantities  $D_i^\alpha$  are what is needed to correctly measure numerical smoothness across the interface. Notice that the definition does not refer to any convergence to a target solution  $u$ . In an attempt to give a corresponding numerical smoothness of Type I (I for interior) we replace the  $D_i$  by  $F_i$ , which is the difference in the derivatives of  $u_h$  and  $u$  at  $x_m$ .

**Definition 3.2. Type I Numerical Smoothness.** Let  $u \in C^{p+1}(\Omega)$ ,  $\Omega \subset \mathbb{R}^n$  and let  $u_h \in \mathbb{P}_p^h$  be a piecewise polynomial of degree  $\leq p$  with respect to a family of subdivisions  $\{\mathcal{T}_h\}$  of  $\Omega$ . Let  $\{x_i\}_{i=1}^{N^T}$  be a collection of points  $x_i \in T_i \in \mathcal{T}_h$ ,  $1 \leq i \leq N^T$ , where  $N^T$  is the cardinality of  $\mathcal{T}_h$ . Then  $u_h$  is said to be  $W_s^{p+1}(\Omega)$ -smooth of Type I,  $s \geq 1$ , if there is a constant  $C_s$ , independent of  $h$  and  $u_h$ , such that

$$(26) \quad \sum_{i=1}^{N^T} h^n \|F_i\|^s \leq C_s,$$

and  $W_\infty^{p+1}(\Omega)$ -smooth of Type I, if there exists a constant  $C_\infty$  independent of  $h$  and  $u_h$  such that

$$(27) \quad \max_{1 \leq i \leq N^T} \|F_i\| \leq C_\infty,$$

where the components of  $F_i$  are the scaled differences between partial derivatives

$$F_i^\alpha = \partial^\alpha (u - u_h)(x_i) / (h^{p+1-|\alpha|}), \quad |\alpha| = k, \quad 0 \leq k \leq p.$$

Now in creating a sound smoothness indicator that can account for interior and boundary smoothness, ideally one must incorporate the  $F^\alpha$  (more practically, the computable  $M^\alpha = \partial^\alpha u_h$ ) and  $D^\alpha$  quantities. In other words, in an adaptive algorithm, a computable bound ideally should include all or some of them in a proper expression, so long as the cost effect is not too much of a concern. On the other hand, our next theorem concerns a necessary condition for convergence. So we only use the  $D^\alpha$  quantities to describe it. The one using  $F^\alpha$  will come after that. In this perspective we can say that these theorems put the statement “a numerical approximate solution ought to be as smooth as its targeted exact solution.” on a rigorous footing.

**Theorem 3.3.** Suppose that  $u \in W_s^{p+1}(\Omega)$ ,  $s = 1, 2, \infty$ ,  $\Omega \subset \mathbb{R}^n$  and that  $u^R$  is in  $\mathbb{P}_p^h$  on a quasi-uniform family of meshes on  $\Omega$ , be it made of triangles or quadrilaterals ( $n = 2$ ) or tetrahedrons or hexahedrons ( $n = 3$ ). Then a necessary condition for

$$\|u - u^R\|_{L^s(\Omega)} \leq \mathcal{O}(h^{p+1})$$

is for  $u^R$  to be  $W_s^{p+1}$  smooth. In particular, for

$$\|u - u^R\|_{L^\infty(\Omega)} \leq \mathcal{O}(h^{p+1})$$

a necessary condition is that all jumps in the  $k^{\text{th}}$  partial derivatives at midpoints  $x_i$  ( $n=2$ ) and face centers  $x_i$  ( $n=3$ ) satisfy

$$|[\partial^\alpha u^R]_{x_i}| = \mathcal{O}(h^{p+1-k}), \quad |\alpha| = k, \quad 0 \leq k \leq p.$$

Here all smoothness refers to Type A smoothness.

*Proof.* Suppose  $\|u - u^R\|_{L^s(\Omega)} \leq Ch^{p+1+\sigma}$ ,  $\sigma \geq 0$ . Applying this to inequality (22) deduces the result. Other assertions follow in a similar way.  $\square$

Note that all  $D_i^\alpha$  need to be bounded for convergence as a consequence of this theorem.

**Theorem 3.4.** *Suppose that  $u \in C^{p+1}(\Omega)$ ,  $s = 1, 2, \infty$ ,  $\Omega \subset \mathbb{R}^n$ ,  $1 \leq n \leq 3$  and that  $u^R$  is in  $\mathbb{P}_p^h$  on a quasi-uniform family  $\{\mathcal{T}_h\}$  of meshes on  $\Omega$ , be it made of triangles or quadrilaterals ( $n=2$ ) or tetrahedrons or hexahedrons ( $n=3$ ). Then a necessary condition for*

$$\|u - u^R\|_{L^s(\Omega)} = \mathcal{O}(h^{p+1})$$

is for  $u^R$  to be  $W_s^{p+1}$  smooth. In particular, for

$$\|u - u^R\|_{L^\infty(\Omega)} = \mathcal{O}(h^{p+1})$$

a necessary condition is that all the  $k^{\text{th}}$  partial derivatives  $x_i \in T$  satisfy

$$(28) \quad |\partial^\alpha u^R(x_i)| = \mathcal{O}(1), \quad |\alpha| = k, \quad 0 \leq k \leq p.$$

Here all smoothness refers to Type I smoothness and  $\{x_i\}$  is any collection of points, one from each element.

*Proof.* Since there is no essential difference between the proof for 1-D and those for higher dimensions, we will just give a 1-D version. Let  $\mathcal{T}_h = \cup_{i=0}^N [x_i, x_{i+1}]$  be a quasi-uniform subdivision on  $\Omega = (a, b)$ , and let  $u \in W_\infty^{p+1}(\Omega)$  and  $u^I \in C(\Omega) \cap \mathbb{P}_p^h$  such that  $u^I$  restricted to  $T_i = (x_i, x_{i+1})$  is the Lagrange nodal interpolant of degree  $\leq p$ . Let  $u^R \in \mathbb{P}_p^h$  be given and to simplify the presentation,  $u_R^{(k)} = \frac{d^k}{dx^k} u^R$ ,  $u_I^{(k)} = \frac{d^k}{dx^k} u^I$ , and  $u^{(k)} = \frac{d^k}{dx^k} u$ . At a typical point  $x_m \in T_i$ , we have the difference in derivatives

$$\begin{aligned} |\tilde{F}_i^{(k)}| &:= |u_R^{(k)}(x_m) - u^{(k)}(x_m)| \\ &\leq |u_R^{(k)}(x_m) - u_I^{(k)}(x_m)| + |u_I^{(k)}(x_m) - u^{(k)}(x_m)| \\ &= I_1 + I_2. \end{aligned}$$

On the one hand

$$(29) \quad I_2 \leq Ch^{p+1-k} |u|_{W_\infty^{p+1}(T_i)},$$

and on the other hand

$$(30) \quad I_1 \leq Ch^{-k} \|u^R - u^I\|_{L^\infty(T_i)},$$

where we have used quasi-uniformness of the mesh. In addition

$$\|u^R - u^I\|_{L^\infty(T_i)} \leq \|u^R - u\|_{L^\infty(T_i)} + \|u^I - u\|_{L^\infty(T_i)}.$$

Combining all the related estimates, we have

$$(31) \quad |\tilde{F}_i^{(k)}| \leq Ch^{-k} \left( h^{p+1} |u|_{W_\infty^{p+1}(\Omega)} + \|u_R - u\|_{L^\infty(\Omega)} \right),$$

which stated in a more practical manner is (28).

As for the  $W_s^{p+1}$ ,  $s = 1, 2$  smoothness estimates, we proceed as before, but now

$$I_2 \leq Ch^{p+1-k} |u|_{W_s^{p+1}(T_i)}.$$

Using a standard scaling argument on (30), we have

$$I_1 \leq Ch^{-k} h^{-\frac{1}{s}} \|u^R - u^I\|_{L^s(T_i)}.$$

and

$$\|u^R - u^I\|_{L^s(T_i)} \leq \|u^R - u\|_{L^s(T_i)} + \|u^I - u\|_{L^s(T_i)},$$

Thus

$$(32) \quad \begin{aligned} |\tilde{F}_i^k| &\leq Ch^{-k} h^{p+1} |u|_{W_s^{p+1}(T_i)} \\ &\quad + Ch^{-k} h^{-\frac{1}{s}} (\|u^R - u\|_{L^s(T_i)} + \|u^I - u\|_{L^s(T_i)}). \end{aligned}$$

Combining all the related estimates, we have

$$|\tilde{F}_i^k| \leq Ch^{-k} \left( h^{p+1} (1 + h^{-\frac{1}{s}}) |u|_{W_s^{p+1}(T_i)} + h^{-\frac{1}{s}} \|u^R - u\|_{L^s(T_i)} \right).$$

Hence

$$|F_i^k| \leq C \left( 1 + h^{-1/s} |u|_{W_s^{p+1}(T_i)} + h^{-(p+1+1/s)} \|u^R - u\|_{L^s(T_i)} \right).$$

Summing appropriately completes the proof.  $\square$

*Remark 3.5.* An immediate consequence of this theorem is that if a single  $D_i^\alpha$  is bigger than  $\mathcal{O}(1/h)$ , then  $\|u - u^R\|_{L^2}$  must be bigger than  $\mathcal{O}(h^{p+1})$ . Note also that the boundedness of the computable  $M^\alpha = \partial^\alpha u^R(x_i)$  is necessary. Again it advocates that  $M^\alpha$  should be incorporated one way or another in a computable bound of an adaptive algorithm. The success of using smoothness indicator of the type  $S = (D, M)$  can be found in [11, 12]. It insisted in the control that  $D$  and  $M$  must be bounded and our two theorems justify that approach.

Several concluding remarks concerning future extensions are in order here. Our results here pertain to a coordinate free approach in the sense that our approximating functions  $u_h$  are polynomial on each element. The case of the pullbacks being polynomials works as well, but it is more subtle to handle due to some complications including some tensor product polynomial approximation issues on quadrilateral meshes (Arnold *et al.* [1]). We will report it in a separate paper.

#### 4. Appendix: $W_s^{p+1}(\Omega)$ smoothness estimates for 1-D

In this section, we give a short proof of Theorem 3.3 for  $n = 1$ . The  $s = \infty$  case was communicated to me by T. Sun [10].

##### 4.1. Limitation of a short proof in 1-D.

*Proof.* Let  $\{\mathcal{T}\}_h = \cup_{i=0}^N [x_i, x_{i+1}]$  be a quasi-uniform subdivision on  $\Omega = (a, b)$ , and let  $u \in W_\infty^{p+1}(\Omega)$  and  $u^I \in C(\Omega) \cap \mathbb{P}_p^h$  such that  $u^I$  restricted to  $(x_i, x_{i+1})$  is the Lagrange nodal interpolant of degree  $\leq p$ . Let  $u^R \in \mathbb{P}_p^h$  be given and to simplify the presentation, first let  $\Omega_i^+ = (x_i, x_{i+1})$ ,  $\Omega_i^- = (x_{i-1}, x_i)$ ,  $\Omega_i = (x_{i-1}, x_{i+1})$ ,  $u_R^{(k)} = \frac{d^k}{dx^k} u^R$  and  $u_I^{(k)} = \frac{d^k}{dx^k} u^I$ .

Now

$$\begin{aligned}
 |J_i^{(k)}| &:= |u_R^{(k)}(x_i^+) - u_R^{(k)}(x_i^-)| \\
 (33) \quad &\leq |u_R^{(k)}(x_i^+) - u^{(k)}(x_i)| + |u_R^{(k)}(x_i^-) - u^{(k)}(x_i)| \\
 &\leq |u_R^{(k)}(x_i^+) - u_I^{(k)}(x_i^+)| + |u_I^{(k)}(x_i^+) - u^{(k)}(x_i)| \\
 &\quad + |u_R^{(k)}(x_i^-) - u_I^{(k)}(x_i^-)| + |u_I^{(k)}(x_i^-) - u^{(k)}(x_i)| \\
 &= I_1 + I_2 + I_3 + I_4.
 \end{aligned}$$

We have

$$(34) \quad I_2 + I_4 \leq Ch^{p+1-k} |u|_{W_\infty^{p+1}(\Omega)},$$

and

$$(35) \quad I_1 + I_3 \leq Ch^{-k} \left( \|u^R - u^I\|_{L^\infty(\Omega_i^+)} + \|u^R - u^I\|_{L^\infty(\Omega_i^-)} \right),$$

where we have used quasi-uniformness of the mesh. In addition

$$\|u^R - u^I\|_{L^\infty(\Lambda)} \leq \|u^R - u\|_{L^\infty(\Lambda)} + \|u^I - u\|_{L^\infty(\Lambda)}, \quad \Lambda = \Omega_i^\pm.$$

Combining all the related estimates, we have

$$(36) \quad |J_i^{(k)}| \leq Ch^{-k} \left( h^{p+1} |u|_{W_\infty^{p+1}(\Omega)} + \|u_R - u\|_{L^\infty(\Omega)} \right).$$

This completes the proof.  $\square$

This proof works even for higher dimensions. In fact, recalling  $W_\infty^1(\Lambda) = \text{Lip}(\Lambda)$ , convex  $\Lambda \subset \mathbb{R}^n$ ,  $n \geq 1$  ([3], p 33), we see that  $u \in W_\infty^{p+1}(\Omega)$  implies that the term  $\partial^\alpha u(x_i)$ ,  $|\alpha| = k$ ,  $0 \leq k \leq p$  in (33) makes sense for higher dimensions and the proof carries over with proper adjustment on the choice of  $u^I$ .

It is easy to see similar results can be obtained for  $W_s^{p+1}$ -smoothness estimates,  $s = 1, 2$  in 1-D case.

*Remark 4.1.* Unfortunately, this simple proof cannot be generalized to higher dimensions for several reasons. First, since we used point value  $u^{(k)}(x_i)$  in (33), and by the Sobolev imbedding theorem [13] for a  $W_s^{p+1}$  function this would require  $(p+1-k)s \geq n$  for  $s = 1$  and  $(p+1-k)s > n$  for  $s = 2$ . That means for  $s = 1$  we must conclude that in general if  $p \geq k > (p+1-n)$  the proof cannot be carried over to higher dimensions. For  $s = 2$ , the proof cannot be used if  $p \geq k \geq (p+1-n/2)$ . For  $n = 2, 3$  this involves derivatives of order  $p$  or  $p-1$ . Second, the existence of  $u^I$  depends on point values and in high dimensions the mesh point  $x_i$  is replaced by a center of an  $n-1$  simplex. The local geometry is different and it is hard to find or awkward to describe such an approximation  $u^I$  based on interpolation on the dual mesh. This problem is overcome by the use of covolumes in our main approach and point value based interpolants are replaced by projection type interpolants when necessary.

## References

- [1] D. Arnold, D. Boffi and R. Falk, Approximation by quadrilateral finite elements, *Math. Comp.*, 71 (2002) 909-922.
- [2] J. Brandts, S. Korotov, and M. Křížek, On the equivalence of regularity criteria for triangular and tetrahedral finite element partitions, *Comput. Math. Appl.*, 55 (2008) 2227-2333.
- [3] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, 3rd Ed., Springer, 2008.
- [4] S. H. Chou and S. He, On the regularity and uniformness conditions on quadrilateral grids, *Comput. Methods Appl. Mech. Engrg.*, 191 (2002) 5149-5158.



- [5] S. H. Chou and P. Vassilevski, A general mixed covolume framework for constructing conservative schemes for elliptic problems, *Math. Comp.*, 68 (1999) 991–1011.
- [6] V. Girault and P. Raviart, *Finite Element Methods for Navier-Stokes Equations*, Springer-Verlag, 1986.
- [7] R. Hartmann, *Numerical Analysis of Higher Order Discontinuous Galerkin Finite Element Method*, CFD - ADIGMA Course on Very High Order Discretization Methods, Von Karman Institute for Fluid Dynamics, Rhode Saint Genèse, Belgium, 2008.
- [8] R. LeVeque, *Finite Volume Methods for Hyperbolic Problems*, Cambridge University Press, 2002.
- [9] T. Sun, Necessity of numerical smoothness, *Internat. J. Infor. Sys. Sci.*, 1 (2012) 1-6. arXiv: 1207.3026 v1[math NA].
- [10] T. Sun, Private Communication, 2013.
- [11] T. Sun, Numerical smoothness and error analysis for WENO on nonlinear conservation laws, to appear in *Numer. Methods PDE*, (2013).
- [12] T. Sun and D. Rumsey, Numerical smoothness and error analysis for RKDG on the scalar nonlinear conservation laws, *J. Comput. Appl. Math.*, 241 (2013) 68-83.
- [13] J. Trangenstein, *Numerical Solution of Elliptic and Parabolic Partial Differential Equations*, Cambridge, University Press, 2013.
- [14] L. Wahlbin, *Superconvergence in Galerkin Finite Element Methods*, Springer, 1995.

Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH, 43403-0221, USA

*E-mail:* `chou@bgsu.edu`