Technische Universität Chemnitz

Sonderforschungsbereich 393

Numerische Simulation auf massiv parallelen Rechnern

Gerd Kunert and Serge Nicaise

Zienkiewicz–Zhu error estimators on anisotropic tetrahedral and triangular finite element meshes

Preprint SFB393/01-20

Abstract

We consider *a posteriori* error estimators that can be applied to *anisotropic* tetrahedral finite element meshes, i.e. meshes where the aspect ratio of the elements can be arbitrarily large. Two kinds of Zienkiewicz–Zhu (ZZ) type error estimators are derived which are both based on some recovered gradient. Two different, rigorous analytical approaches yield the equivalence of both ZZ error estimators to a known residual error estimator. Thus reliability and efficiency of the ZZ error estimation is obtained. Particular attention is paid to the requirements on the anisotropic mesh. The analysis is complemented and confirmed by several numerical examples.

Keywords: Anisotropic mesh, error estimator, Zienkiewicz–Zhu estimator, recovered gradient

AMS: 65N30, 65N15, 65N50

Preprint-Reihe des Chemnitzer SFB 393

SFB393/01-20 **July 2001**

Contents

Gerd Kunert
TU Chemnitz, Fakultät für Mathematik, D-09107 Chemnitz, Germany
gkunert@mathematik.tu-chemnitz.de
http://www.tu-chemnitz.de/∼gku

Serge Nicaise
Université de Valenciennes et du Hainaut Cambrésis, MACS,
B.P. 311, 59304 Valenciennes Cedex, France
snicaise@univ-valenciennes.fr
http://www.univ-valenciennes.fr/macs/nicaise

1 Introduction

Several classes of boundary value problems intrinsically give rise to solutions that exhibit lower dimensional, *anisotropic* behaviour. Such anisotropic solutions show little variation in certain space directions but much variation otherwise. For example, singularly perturbed problems often result in solutions with boundary layers. Even the solution of the Poisson problem in three space dimensions is generically anisotropic along some concave edge, see also the numerical experiments of section 5. Within the finite element method, such anisotropic solutions can be favourably resolved with *anisotropic meshes*. By this we understand meshes with stretched elements which are characterized by an unbounded aspect ratio, i.e. the ratio of the diameters of the circumscribed and inscribed sphere can be arbitrarily large.

Our emphasis is on *error estimators* which form a basis of any adaptive solution algorithm. The theory of error estimation in nowadays well established for conventional, isotropic meshes (i.e. where the aspect ratio of the elements is bounded). The books [Ver96, AO00] provide a comprehensive and useful overview of the state of the art.

On *anisotropic meshes* the error estimation theory is much less developed. Recently the intensive research has led to several estimators that can be applied to different boundary value problems as well as norms, see [Sie96, Kun00, KV00, Kun01a, Kun01b, Kun01c, DGP99]. Exemplarily we mention residual error estimators and local problem error estimators for the Poisson problem or a singularly perturbed problem; the error can be estimates in the energy norm or in the L_2 norm.

There is one popular estimator for isotropic meshes that did not have yet a counterpart for anisotropic meshes. This so–called Zienkiewicz-Zhu (ZZ) estimator has been invented in [ZZ87] and later been improved in [ZZ92]; many more variants have been developed since. The basic idea consists in computing an improvement of the gradient of the numerical solution by some post–processing procedure. The difference between this so–called *recovered gradient* and the original gradient serves as error estimator. This idea of ZZ error estimation has been very appealing and popular in the finite element community since

- the estimator is comparatively cheap because a recovered gradient is often computed anyway,
- the estimator is astonishingly robust (in numerical experiments) for a wide range of problems, see e.g. [BSU94a, BSU$^+$94b].

Our work here is devoted to the extension of the ZZ estimator to *anisotropic meshes*. We start with a recapitulation of the existing isotropic analysis and discuss their suitability for anisotropic meshes. The theoretical approaches to ZZ error estimators (on isotropic meshes) can be divided roughly into three classes:

- proving equivalence to residual error estimators,
- utilizing superconvergence properties,
- minimization approach.

Each of these approaches will now be discussed briefly.

Equivalence to residual error estimator: Here the ZZ error estimator is proven to be equivalent to a residual error estimator, thus transferring reliability and efficiency to the first estimator. This approach goes back to [Rod94] and is repeated in [Ver96, section 1.5].

In our paper these ideas will be generalized to *anisotropic meshes*. Of course several modifications and extensions are necessary:

- Although some recovered gradient is still applied, it is now scaled with weights that depend on the stretching directions (i.e. the alignment) of the anisotropic elements.
- The anisotropic meshes have to meet additional requirements which are due to the anisotropy. These requirements roughly mean that the anisotropic meshes should not be totally unstructured but instead obey some 'sensible' geometrical principles. These demands also seem reasonable in the light of superconvergence properties discussed below.

Superconvergence approach: It forms the basis of most proofs concerning ZZ estimators. Exemplarily we refer to [AO00] and the citations therein. In suitable, specialized settings even asymptotic exactness of the (global) ZZ estimator can be shown. This requires

- consistence, localization, boundedness and linearity of the recovery operator and
- a superconvergence property of the finite element scheme.

Unfortunately the superconvergence approach inherits two drawbacks. Firstly, the theoretical analysis requires very specialized meshes which are rarely found in practice (e.g. in adaptive refinement procedures). Secondly local equivalences cannot be proven.

The application of such a superconvergence analysis to *anisotropic meshes* is unclear up to now. Superconvergence results are not known for general meshes but only for special Shishkin or Shishkin–type meshes, see [RL01, Zha98]. For example, [RL01] prove a certain kind of superconvergence for 2D Shishkin–type meshes consisting of axiparallel rectangles, bilinear finite elements, and a singularly perturbed reaction–convection–diffusion problem in the unit square. Most likely the results can be employed to define a ZZ estimator, even if this is not presented in the aforementioned work.

Summarizing, we do not pursue the superconvergence approach because of the high demands on the meshes which are hardly consistent with anisotropic solutions.

Minimization approach: A third kind of analysis utilizes a close relation between the ZZ estimator and a minimum formulation, cf. [CB01, BC01]. It allows to investigate general averaging operators which define the estimator, and it avoids superconvergence assumptions. The resulting error bounds involve so–called 'higher order terms' that contain the unknown solution. Hence these bounds can only be interpreted in an asymptotic sense. Moreover the constants in the reliability result depend on the shape of the finite elements.

After presenting different techniques to analyse ZZ estimators, we will consider from now on exclusively the first approach, namely the equivalence to a residual error estimator. As it has been explained, this analysis seems most promising for anisotropic meshes.

The outline of this paper is as follows. The model problem, some main notation as well as the assumptions on the mesh are introduced in section 2. In section 3 we first recall a known residual error estimator that is required for the subsequent analysis. Afterwards two kinds of ZZ error estimators are derived and rigorously analysed. The first estimator is based on a global projection property which corresponds to a particular choice of the underlying recovered gradient. The second ZZ estimator is an improvement because the recovered gradient can be defined with *arbitrary weights*. Our novel analysis additionally yields *local element–wise estimates*. Section 4 is devoted to a detailed examination of the mesh assumptions. The numerical examples of section 5 complement and confirm the theoretical analysis.

2 Model problem and notation

2.1 Model problem

Consider a Poisson model problem with homogeneous Dirichlet boundary conditions in a polyhedral domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$:

$$\left.\begin{array}{rcll} -\Delta u &=& f & \text{in } \Omega \\ u &=& 0 & \text{on } \Gamma_D := \partial\Omega. \end{array}\right\} \tag{1}$$

Our analysis is presented for three dimensional domains ($d = 3$); the application to two dimensional domains ($d = 2$) is readily possible. The corresponding variational formulation reads

$$\text{Find } u \in H^1_o(\Omega): \qquad \int_\Omega \nabla u\,\nabla v \;=\; \int_\Omega f\,v \qquad \forall v \in H^1_o(\Omega) \quad , \tag{2}$$

where $H^1_o(\Omega)$ denotes the usual Sobolev space of functions that vanish on Γ_D. For $f \in L_2(\Omega)$ problem (2) admits a unique solution due to the Lax–Milgram lemma.

In order to discretize (2), let $\mathcal{F} = \{\mathcal{T}_h\}$ be a family of triangulations $\mathcal{T}_h$ of Ω. We assume a conforming triangulation (cf. [Cia78, Chapter 2]) that consists of tetrahedra ($d = 3$) or triangles ($d = 2$). Let $V_h \subset H^1_o(\Omega)$ be the finite element space of piecewise affine linear functions on $\mathcal{T}_h$ that vanish on Γ_D. The finite element solution u_h is uniquely obtained via

$$\text{Find } u_h \in V_h: \qquad \int_\Omega \nabla u_h\,\nabla v_h \;=\; \int_\Omega f\,v_h \qquad \forall v_h \in V_h \quad . \tag{3}$$

2.2 Notation

The following paragraphs now introduce most of the notation required. For some domain $\omega \subset \mathbb{R}^2$ or $\omega \subset \mathbb{R}^3$ let $\|\cdot\|_\omega := \|\cdot\|_{L_2(\omega)}$ be the usual $L_2(\omega)$ norm. The space of polynomials of order at most k is denoted by $\mathbb{P}^k(\omega)$. For some (column) vectors $\underline{v}, \underline{w}$ let $(\underline{v}, \underline{w})$ be the Euclidean scalar product and $|\underline{v}| := (\underline{v}, \underline{v})^{1/2}$ be the Euclidean length. Instead of $x \leq c \cdot y$ or $c_1 y \leq x \leq c_2 y$ (with positive constants independent of x, y or $\mathcal{T}_h$) we use the shorthand notation $x \lesssim y$ or $x \sim y$, respectively.

The next paragraph presents notation that is related to the triangulation $\mathcal{T}_h$ and its elements. Tetrahedra are denoted by T, T' or T_i, faces are denoted by E, and nodes of $\mathcal{T}_h$ are denoted by x. Next, define nodal sets $\mathcal{N}_T, \mathcal{N}_E, \mathcal{N}_{\bar\Omega}$ that contain all nodes of a tetrahedron T, a face E, or of $\bar\Omega$ (i.e. including boundary nodes), respectively. Let $\mathcal{E}_\Omega$ be the set of all interior edges (2D) or faces (3D) of Ω. For a node x we introduce a local neighbourhood patch $\omega_x := \bigcup_{T:x\in\mathcal{N}_T} T \subset \mathbb{R}^3$ which is the union of all tetrahedra having this node. Similarly for some face E let $\omega_E \subset \mathbb{R}^3$ be the union of both tetrahedra having this face (with obvious boundary modifications). For a tetrahedron T, a face E or a patch ω_x set $|T| = \mathrm{meas}_3(T)$, $|E| = \mathrm{meas}_2(E)$ or $|\omega_x| = \mathrm{meas}_3(\omega_x)$, respectively (the distinction from the Euclidean vector length is obvious from the context).

The four vertices of an arbitrary but fixed tetrahedron $T \in \mathcal{T}_h$ are temporarily denoted by $P_0, \ldots, P_3$ such that P_0P_1 is the longest edge of T, $\mathrm{meas}_2(\triangle P_0P_1P_2) \geq \mathrm{meas}_2(\triangle P_0P_1P_3)$, and $\mathrm{meas}_1(P_1P_2) \geq \mathrm{meas}_1(P_0P_2)$. Additionally define three pairwise orthogonal vectors $\underline{p}_{i,T}$ with lengths $h_{i,T} := |\underline{p}_{i,T}|$, see figure 1. Observe $h_{1,T} > h_{2,T} \geq h_{3,T}$ and set

$$h_{min,T} := \min_{i=1\ldots3} h_{i,T} = h_{3,T} \qquad .$$

The matrix $C_T \in \mathbb{R}^{3\times3}$ is defined as

$$C_T := (\underline{p}_{1,T}, \underline{p}_{2,T}, \underline{p}_{3,T}) \qquad ,$$

and describes (roughly speaking) the anisotropic orientations of the tetrahedron T.

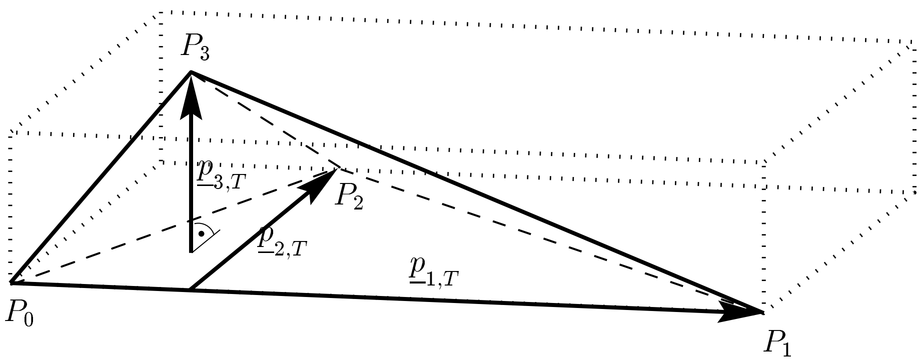

Figure 1: Notation of tetrahedron T

For a face E of a tetrahedron T let

$$h_{E,T} := 3|T|/|E|$$

be the length of the *height* over E in T. Note that $h_{E,T}$ is *not* the diameter of E, as in the usual convention.

The quantities $h_{min,T}$ and $h_{E,T}$ are associated with a *tetrahedron* T. Often it is more convenient to utilize equivalent data that are related to a *face* E or *node* x. To this end define averaged terms by

$$
\begin{aligned}
h_E &:= (h_{E,T_1} + h_{E,T_2})/2 && \text{for } E = T_1 \cap T_2 \\
h_{min,E} &:= (h_{min,T_1} + h_{min,T_2})/2 && \text{for } E = T_1 \cap T_2 \\
h_{i,x} &:= \frac{1}{n} \sum_{T \subset \omega_x} h_{i,T} \qquad h_{min,x} := \frac{1}{n} \sum_{T \subset \omega_x} h_{min,T} && ,
\end{aligned}
$$

where n is the number of elements T in ω_x. Note that $h_{min,E}$ is not the minimal dimension of the face E. For boundary faces $E \subset \partial\Omega$ modify $h_E := h_{E,T}$ and $h_{min,E} := h_{min,T}$, where $\partial T \supset E$.

Next consider an arbitrary interior face E. Let $\underline{n}_E$ be any of the two unit normal vectors for E, and keep it fixed from here on. For a piecewise continuous (scalar or vector valued) function v denote by $[\![v]\!]_E$ the jump of v across E in the direction $\underline{n}_E$. Let $\partial_{n_E} v := \underline{n}_E \cdot \nabla v$ be the (unitary) directional derivative. Note that the orientation of $\underline{n}_E$ influences terms like $[\![v]\!]_E$ but not $[\![\partial_{n_E} v]\!]_E$.

2.3 Mesh requirements

In addition to the usual conformity conditions of the mesh (see Ciarlet [Cia78, Chapter 2]) we demand the following assumptions.

(A1) The number of tetrahedra containing a node x is bounded uniformly.

(A2) The dimensions of adjacent tetrahedra must not change rapidly, i.e.

$$h_{i,T'} \sim h_{i,T} \qquad \forall\, T, T' \text{ with } T \cap T' \neq \emptyset\,,\; i = 1 \ldots d \quad .$$

(A3) For each node x there exists a matrix $C_x \in \mathbb{R}^{d \times d}$ such that

$$|C_x^{-1} \underline{v}| \;\sim\; |C_T^{-1} \underline{v}| \qquad \forall\, \underline{v} \in \mathbb{R}^d, \forall\, T \subset \omega_x \qquad .$$

(A4) An assumption on the shape of each element:

$$|C_T^{-1} \underline{n}_E| \;\sim\; h_{E,T}^{-1} \qquad \forall\, E \subset \partial T \qquad .$$

(A5) The L_2 projection is stable in the sense of [KV00, Section 4]. For self–containment we repeat the definition given there. Start with two (distinct) elements $T_1, T_2 \in \mathcal{T}_h$ and define their (topological) edge distance by

$$l(T_1, T_2) := 1 + \text{ minimal number of edges of any edge path connecting } T_1 \text{ and } T_2.$$

Set $l(T, T) := 0$. Note that in both the 2D and 3D case the *edges* count. Next, for a given element T introduce neighbourhood (ring) patches by

$$R_k(T) := \{T' : l(T', T) = k\}, \qquad k \in \mathbb{N} \qquad .$$

Then assumption (A5) is satisfied if there exist positive constants $c_1, c_2, \alpha, \beta, r$ such that

$$\left.\begin{array}{rcl} h_{min,T_1}/h_{min,T_2} & \leq & c_1 \cdot \alpha^{l(T_1,T_2)} \qquad \forall\, T_1, T_2 \in \mathcal{T}_h \\ \mathrm{card}(R_k(T)) & \leq & c_2 \cdot k^r\, \beta^k \qquad \forall\, T \in \mathcal{T}_h, \forall k \in \mathbb{N}_+ \\ \alpha \cdot \beta & < & \begin{cases} \sqrt{2}+\sqrt{3} \approx 3.146 & \text{if } d=2 \\ (3+\sqrt{5})/2 \approx 2.618 & \text{if } d=3. \end{cases} \end{array}\right\} \tag{4}$$

The mesh assumptions (A1) and (A2) imply several convenient equivalences.

$$\left.\begin{array}{rrcll} \text{(A2)} \Rightarrow & h_{i,x} & \sim & h_{i,T} & \forall\, T \subset \omega_x \\ \text{(A2)} \Rightarrow & h_{min,x} & \sim & h_{min,T} & \forall\, T \subset \omega_x \\ \text{(A2)} \Rightarrow & h_{min,x} & \sim & h_{min,E} & \forall\, E : x \in \mathcal{N}_E \\ \text{(A2)} \Rightarrow & h_E & \sim & h_{E,T} & \forall\, E \subset \partial T \\ \text{(A1)+(A2)} \Rightarrow & |T| & \sim & |\omega_x| & \forall\, T \subset \omega_x \end{array}\right\} \tag{5}$$

Furthermore, with the help of (A2) and (A3) we can rewrite assumption (A4) as

$$|C_x^{-1}\underline{n}_E| \sim h_E^{-1} \qquad \forall\, E : x \in \mathcal{N}_E \quad . \tag{6}$$

Remark 2.1 *The mesh assumptions are scrutinized in detail in section 4. Here some remarks may facilitate the understanding.*

Assumption (A3) roughly means that there exists a transformation C_x^{-1} which maps the patch ω_x onto an isotropic patch of size $\mathcal{O}(1)$.

Assumption (A4) roughly demands for an anisotropic tetrahedron that small faces are almost perpendicular to long edges, depending on the aspect ratio.

Finally the stability assumption (A5) is only a sufficient condition to derive the residual error estimation. Recent research [Car01, BPS01, Ste01] suggests that the restrictions of (A5) can be weakened; some results of the aforementioned work already apply to our setting here.

2.4 Matching function

Reliability and *efficiency* are highly desirable properties in *a posteriori* error estimation. They basically mean that the error $\|u - u_h\|_*$ (in some suitable norm) can be bounded from above and below, respectively, with constants independent of u, u_h or $\mathcal{T}_h$.

Most standard error estimators on *isotropic* finite element meshes are reliable and efficient at the same time, cf. [AO00, Ver96]. Unfortunately the situation is much less obvious on *anisotropic* meshes. The analysis as well as numerical experiments strongly suggest that reliability and efficiency cannot be achieved simultaneously on *arbitrary* anisotropic meshes. However if the anisotropy of the solution is sufficiently well aligned with the anisotropy of the mesh then one can expect both properties at the same time. Intuitively all applications of anisotropic finite elements follow this concept: an element should be stretched in that direction where the function (or more precisely, its derivative) exhibit little variation.

In order to measure the alignment of an anisotropic mesh $\mathcal{T}_h$ with an anisotropic function v, a so–called *matching function* has been proposed by Kunert [Kun99, Kun00].

Definition 2.2 (Matching function) *Let $v \in H^1(\Omega)$, and $\mathcal{T}_h \in \mathcal{F}$ be a triangulation of Ω. Define the* matching function $m_1 : H^1(\Omega) \times \mathcal{F} \mapsto \mathbb{R}$ *by*

$$m_1(v, \mathcal{T}_h) \ := \ \Big(\sum_{T \in \mathcal{T}_h} h_{min,T}^{-2} \cdot \| C_T^T \, \nabla v \|_T^2 \Big)^{1/2} \Big/ \ \|\nabla v\|_\Omega \qquad . \tag{7}$$

The vital importance of the matching function for anisotropic error estimation can be seen in the error bounds (9), (10) and (14), (15) below.

The matching function is not central to our analysis here. Hence we refer to [Kun00] for a comprehensive discussion, and restrict ourselves to a brief discussion of basic features. Firstly the definition immediately implies $m_1(v, \mathcal{T}_h) \geq 1$. On *isotropic* meshes one obtains easily $m_1(v, \mathcal{T}_h) \sim 1$; then the matching function merges with other constants and becomes invisible. In contrast to this more care is necessary for *anisotropic* meshes. If the mesh is suitably aligned with the anisotropic solution one still achieves $m_1(v, \mathcal{T}_h) \sim 1$ and thus reliable and efficient error estimation. If however the anisotropic mesh is not aligned with the solution then $m_1(v, \mathcal{T}_h)$ can be arbitrarily large (cf. [Kun01a, Numerical experiment 2] or [Kun99, Remark 3.3]). Hence upper and lower error bounds may differ by an arbitrarily large factor; thus rendering the error estimation useless.

3 Error estimators

Here we start by presenting the residual error estimator of [KV00] which forms the basis of the subsequent analysis. Afterwards two kinds of ZZ error estimators are presented. The first kind follows the lines of [Rod94] (also described in [Ver96, section 1.5]). It is based on a recovered gradient ∇^{R_1} which satisfies a global projection property. Next, we improve this first approach by employing a much more flexible recovered gradient ∇^{R_2} to define the second ZZ error estimator. Accordingly a novel analysis is required (cf. lemma 3.12 and theorem 3.13) which is based on different techniques than for the first estimator.

Note that all estimators are given in several forms. The first representation is the one used in practice, and is related either to a face E or an element T. The other, equivalent representation is related to a node x, and is required for analytical purposes.

3.1 Residual error estimator

In [KV00] a face residual based error estimator is introduced for interior faces by

$$\eta_{R,E} \ := \ h_{min,E} h_E^{-1/2} \cdot \| [\![\partial_{n_E} u_h]\!]_E \|_E, \qquad E \in \mathcal{E}_\Omega \qquad . \tag{8}$$

The corresponding lower and upper error bounds are given in [KV00, theorem 5.1]. Provided that mesh assumptions (A1), (A2) and (A5) are satisfied, one has

$$\eta_{R,E} \lesssim \|\nabla(u-u_h)\|_{\omega_E} + \inf_{f_h \in V_h} h_{min,E}\, \|f-f_h\|_{\omega_E} \qquad \forall\, E \in \mathcal{E}_\Omega \tag{9}$$

$$\|\nabla(u-u_h)\|_\Omega \lesssim m_1(u-u_h,\mathcal{T}_h)\left(\sum_{E\in\mathcal{E}_\Omega} \eta^2_{R,E} + \inf_{f_h\in V_h} \sum_{T\in\mathcal{T}_h} h^2_{min,T}\|f-f_h\|^2_T\right)^{1/2}. \tag{10}$$

Clearly $\eta_{R,E}$ is associated with a face E. For our purposes, however, node related quantities are much better suited. Therefore we fix a patch ω_x and combine all its (interior) faces. The first expression below introduces the local, node related error estimator. The second definition introduces the global error estimator whereas the remaining definition facilitate our exposition later on.

Definition 3.1 (Residual error estimators) *The local and global residual error estimators are given by*

$$\eta^2_{R,x} := h^2_{min,x}|\omega_x| \sum_{E:x\in\mathcal{N}_E} h_E^{-2}[\![\partial_{n_E} u_h]\!]^2_E \tag{11}$$

$$\eta^2_R := \sum_{x\in\mathcal{N}_{\bar\Omega}} \eta^2_{R,x} \tag{12}$$

$$\eta^2_{\tilde R,x} := h^2_{min,x}|\omega_x| \sum_{E:x\in\mathcal{N}_E} |C_x^{-1}\,[\![\nabla u_h]\!]_E|^2 \qquad . \tag{13}$$

Lemma 3.2 *Let the mesh assumptions (A1), (A2) be satisfied. Then*

$$\eta^2_{R,x} \sim \sum_{E:x\in\mathcal{N}_E} \eta^2_{R,E} \qquad .$$

Proof: Recall that the dimensions of neighbouring elements must not change rapidly, cf. (5), which implies in particular

$$|E|\cdot h_E = d\cdot(|T_1|+|T_2|)/2 \overset{(5)}{\sim} |\omega_x| \qquad \text{with } E=T_1\cap T_2,\ \forall\, x\in\mathcal{N}_E \quad .$$

Combining now the error estimators $\eta_{R,E}$ for all the interior faces of ω_x yields

$$\begin{aligned}\sum_{E:x\in\mathcal{N}_E} \eta^2_{R,E} &= \sum_{E:x\in\mathcal{N}_E} h^2_{min,E}\, h_E^{-1}\cdot\|[\![\partial_{n_E} u_h]\!]_E\|^2_E \\ &\overset{(5)}{\sim} h^2_{min,x} \sum_{E:x\in\mathcal{N}_E} h_E^{-2}\cdot|E|\, h_E\cdot[\![\partial_{n_E} u_h]\!]^2_E \sim \eta^2_{R,x}\end{aligned}$$

which proves the assertion. ∎

The error estimation by means of the node related error estimator $\eta_{R,x}$ can now be derived easily.

Lemma 3.3 (Residual error estimation) *Assume that mesh assumptions (A1), (A2) and (A5) are satisfied. The error is bounded locally from below and globally from above.*

$$\eta_{R,x} \lesssim \|\nabla(u-u_h)\|_{\omega_x} + \inf_{f_h\in V_h} h_{min,x}\,\|f-f_h\|_{\omega_x} \qquad \forall\, x\in\mathcal{N}_{\bar\Omega} \tag{14}$$

$$\|\nabla(u-u_h)\|_\Omega \lesssim m_1(u-u_h,\mathcal{T}_h)\left(\eta_R^2 + \inf_{f_h\in V_h}\sum_{T\in\mathcal{T}_h} h^2_{min,T}\|f-f_h\|_T^2\right)^{1/2} . \tag{15}$$

Proof: The inequalities follow immediately from (9), (10) and lemma 3.2. ■

The next lemma presents a sufficient condition for the equivalence of $\eta_{R,x}$ and $\eta_{\check{R},x}$. This lemma will be essential for further analysis.

Lemma 3.4 *Let the mesh assumption (A2)–(A4) be satisfied, i.e. in particular (6) holds: $|C_x^{-1}\,\underline{n}_E| \sim h_E^{-1}$ for all faces E with $x\in\mathcal{N}_E$. Then the following equivalence holds:*

$$\eta_{R,x} \sim \eta_{\check{R},x} \qquad . \tag{16}$$

Proof: Consider an arbitrary face E, $x\in\mathcal{N}_E$, and any one of its two unit normal vectors $\underline{n}_E$. Then there exist two further unit vectors $\underline{\tau}_1,\underline{\tau}_2$ such that $(\underline{n}_E,\underline{\tau}_1,\underline{\tau}_2)$ forms an orthonormal vector system (note that $E\subset \operatorname{span}\{\underline{\tau}_1,\underline{\tau}_2\}$). Hence

$$\begin{array}{rrcl} & \underline{n}_E\,\underline{n}_E^T \;+\; \underline{\tau}_1\,\underline{\tau}_1^T \;+\; \underline{\tau}_2\,\underline{\tau}_2^T &=& I_{3\times 3} \\ \text{giving} & \underline{n}_E\cdot\partial_{n_E}u_h \;+\; \underline{\tau}_1\cdot\partial_{\underline{\tau}_1}u_h \;+\; \underline{\tau}_2\cdot\partial_{\underline{\tau}_2}u_h &=& \nabla u_h \qquad . \end{array}$$

Both terms $\partial_{\underline{\tau}_i}u_h$ are continuous across E; only $\partial_{n_E}u_h$ jumps. Thus we conclude

$$[\![\partial_{n_E}u_h]\!]_E\cdot\underline{n}_E = [\![\nabla u_h]\!]_E \qquad .$$

Together with assumptions (A2)–(A4) which imply (6) one obtains

$$\begin{aligned} &\sum_{E:x\in\mathcal{N}_E} |C_x^{-1}\,[\![\nabla u_h]\!]_E|^2 = \sum_{E:x\in\mathcal{N}_E} |C_x^{-1}\,[\![\partial_{n_E}u_h]\!]_E\cdot\underline{n}_E|^2 = \\ &= \sum_{E:x\in\mathcal{N}_E} |C_x^{-1}\,\underline{n}_E|^2\cdot[\![\partial_{n_E}u_h]\!]_E^2 \overset{(6)}{\sim} \sum_{E:x\in\mathcal{N}_E} h_E^{-2}\cdot[\![\partial_{n_E}u_h]\!]_E^2 \end{aligned}$$

which proves the assertion. ■

3.2 First ZZ error estimator

Let us first define the recovered gradient ∇^{R_1} by means of a projection with respect to a particular scalar product. For a precise definition of this inner product, let W_h be the space of piecewise linear vector fields on the triangulation, and set $V_h := W_h\cap C(\Omega,\mathbb{R}^d)$, cf. also [Ver96]. In order to shorten the notation we temporarily introduce the matrices

$$B_x := h_{min,x}\,C_x^{-1} \qquad \text{and} \qquad B_T := h_{min,T}\,C_T^{-1} \qquad .$$

On W_h, we introduce the mesh dependent inner product $(\cdot,\cdot)_h$ by

$$(\underline{v},\underline{w})_h := \sum_{T\in\mathcal{T}_h} |T| \sum_{x\in\mathcal{N}_T} (B_x\, \underline{v}_{|T}(x), B_x\, \underline{w}_{|T}(x)), \tag{17}$$

where

$$\underline{v}_{|T}(x) = \lim_{\substack{y\to x\\ y\in T}} \underline{v}(y).$$

From mesh assumptions (A2), (A3) we have concluded (5), i.e. $h_{min,x} \sim h_{min,T}$, $|C_x^{-1}\underline{v}| \sim |C_T^{-1}\underline{v}|$ for all $T\subset\omega_x$ and thus also $|B_x\underline{v}| \sim |B_T\underline{v}|$ for all $T\subset\omega_x$ and all vectors $\underline{v}\in\mathbb{R}^d$. For an arbitrary but fixed tetrahedron T and for any piecewise linear function $\underline{v}\in W_h$ we can further conclude

$$|T| \sum_{x\in\mathcal{N}_T} |B_x\, \underline{v}_{|T}(x)|^2 \overset{\text{(A3)}}{\sim} |T| \sum_{x\in\mathcal{N}_T} |B_T\, \underline{v}_{|T}(x)|^2 \;\sim\; \|B_T\underline{v}\|_T^2 \qquad .$$

Therefore the mesh assumptions (A2), (A3) imply

$$(\underline{v},\underline{v})_h \sim \sum_{T\in\mathcal{T}_h} \|B_T\underline{v}\|_T^2 \qquad . \tag{18}$$

This last result also shows that $(\cdot,\cdot)_h$ is a scalar product indeed since all B_T are regular matrices. Now the recovered gradient can be defined.

Definition 3.5 (First recovered gradient) *The* recovered gradient $\nabla^{\mathrm{R}_1} : W_h \to V_h$ *is defined as the projection of ∇u_h onto V_h with respect to the inner product $(\cdot,\cdot)_h$, i.e. $\nabla^{\mathrm{R}_1}u_h \in V_h$ is uniquely determined by the condition*

$$(\nabla^{\mathrm{R}_1}u_h - \nabla u_h, \underline{v}_h)_h \;=\; 0 \qquad \forall \underline{v}_h \in V_h \qquad . \tag{19}$$

The recovered gradient $\nabla^{\mathrm{R}_1}u_h$ is piecewise linear and continuous. Its nodal values can be computed locally and coincide with the usual recovered gradient as presented, for example, in [Ver96, equality (1.80)]. Detail are given in the next lemma.

Lemma 3.6 *The value of the recovered gradient at a node x can be determined locally by*

$$(\nabla^{\mathrm{R}_1}u_h)(x) = \sum_{T\subset\omega_x} \mu_T\, \nabla u_{h|T} \qquad \textit{with weight} \quad \mu_T := \frac{|T|}{|\omega_x|} \in \mathbb{R}, T\subset\omega_x \qquad . \tag{20}$$

Proof: The proof utilizes standard ideas as presented e.g. in [Ver96]. Fix the node x and apply the definition of the recovered gradient with $\underline{v}_h := \varphi_x \cdot \underline{e}_i$, where φ_x is the standard (piecewise linear) basis function of V_h for node x, and $\underline{e}_i \in \mathbb{R}^d$ is the ith unit vector. Then

$$\begin{aligned} 0 &= (\nabla^{\mathrm{R}_1}u_h - \nabla u_h \;,\; \varphi_x\cdot\underline{e}_i)_h \\ &= \sum_{T\in\mathcal{T}_h} |T| \sum_{x'\in\mathcal{N}_T} \Big(B_{x'}^T B_{x'}\,(\nabla^{\mathrm{R}_1}u_h(x') - \nabla u_{h|T}(x')) \;,\; \varphi_{x|T}(x')\,\underline{e}_i \Big) \\ &= \sum_{T\subset\omega_x} |T|\cdot \Big(B_x^T B_x\,(\nabla^{\mathrm{R}_1}u_h(x) - \nabla u_{h|T}(x)) \;,\; \underline{e}_i \Big) \end{aligned}$$

holds for $i = 1 \dots d$. Furthermore $B_x^T B_x$ is regular, and hence

$$\begin{aligned} \underline{0} &= \sum_{T \subset \omega_x} |T| \cdot (\nabla^{\mathrm{R}_1} u_h(x) - \nabla u_{h|T}(x)) \\ &= |\omega_x| \, \nabla^{\mathrm{R}_1} u_h(x) \; - \; \sum_{T \subset \omega_x} |T| \cdot \nabla u_{h|T}(x) \end{aligned}$$

which proves the assertion.

Note that the choice of the regular matrix B_x in the definition of the scalar product $(\cdot, \cdot)_h$ has no influence on the nodal value of the recovered gradient. ∎

Now we are ready to define our anisotropic version of the first ZZ estimator. Again, the first two terms are given in a form that can be used in practice. The third quantity is a node related term which can be utilized in further analysis.

Definition 3.7 (First anisotropic ZZ estimators) *The local and global ZZ estimators are given by*

$$\eta_{Z_1,T} := h_{min,T} \, \|C_T^{-1} \, (\nabla^{\mathrm{R}_1} u_h - \nabla u_h)\|_T \tag{21}$$

$$\eta_{Z_1}^2 := \sum_{T \in \mathcal{T}_h} \eta_{Z_1,T}^2 \tag{22}$$

$$\eta_{Z_1,x}^2 := h_{min,x}^2 \, |\omega_x| \left(\sum_{T \subset \omega_x} \frac{|T|}{|\omega_x|} |C_x^{-1} \, \nabla u_{h|T}|^2 \; - \; \left| \sum_{T \subset \omega_x} \frac{|T|}{|\omega_x|} C_x^{-1} \, \nabla u_{h|T} \right|^2 \right) \quad . \tag{23}$$

Similar to the residual error estimator we first establish a relation between the global estimator η_{Z_1} and the node related quantity $\eta_{Z_1,x}$. To achieve this, assume that mesh assumptions (A2) and (A3) hold which imply (18). Furthermore utilize the projection property (19), recall the definition of the matrices B_x, B_T and of the scalar product to obtain

$$\begin{aligned} \eta_{Z_1}^2 &= \sum_{T \in \mathcal{T}_h} h_{min,T}^2 \, \|C_T^{-1} \, (\nabla^{\mathrm{R}_1} u_h - \nabla u_h)\|_T^2 \\ &\overset{(18)}{\sim} (\nabla^{\mathrm{R}_1} u_h - \nabla u_h \, , \, \nabla^{\mathrm{R}_1} u_h - \nabla u_h)_h \\ &\overset{(19)}{=} (\nabla u_h \, , \, \nabla u_h)_h - (\nabla^{\mathrm{R}_1} u_h \, , \, \nabla^{\mathrm{R}_1} u_h)_h \\ &\overset{(17)}{=} \sum_{T \in \mathcal{T}_h} |T| \sum_{x \in \mathcal{N}_T} h_{min,x}^2 \, \left(|C_x^{-1} \, \nabla u_{h|T}|^2 \; - \; |C_x^{-1} \, \nabla^{\mathrm{R}_1} u_h(x)|^2 \right) \quad . \end{aligned}$$

Insert now the nodal value of $\nabla^{\mathrm{R}_1} u_h$ and change the summation order from $\sum_{T \in \mathcal{T}_h} \sum_{x \in \mathcal{N}_T}$ to $\sum_{x \in \mathcal{N}_{\bar{\Omega}}} \sum_{T \subset \omega_x}$ to conclude

$$\eta_{Z_1}^2 \sim \sum_{x \in \mathcal{N}_{\bar{\Omega}}} h_{min,x}^2 \sum_{T \subset \omega_x} |T| \cdot \left(|C_x^{-1} \, \nabla u_{h|T}|^2 \; - \; |C_x^{-1} \, \nabla^{\mathrm{R}_1} u_h(x)|^2 \right)$$

$$\stackrel{(20)}{=} \quad \sum_{x\in\mathcal{N}_{\bar\Omega}} h_{min,x}^2\,|\omega_x| \left(\sum_{T\subset\omega_x} \frac{|T|}{|\omega_x|} |C_x^{-1}\,\nabla u_{h|T}|^2 \;-\; \left| \sum_{T\subset\omega_x} \frac{|T|}{|\omega_x|} C_x^{-1}\,\nabla u_{h|T} \right|^2 \right) \quad .$$

Hence the following relation between the global estimator η_{Z_1} and the node related estimator $\eta_{Z_1,x}$ is obtained provided that the mesh assumptions (A2) and (A3) hold:

$$\eta_{Z_1}^2 \;\sim\; \sum_{x\in\mathcal{N}_{\bar\Omega}} \eta_{Z_1,x}^2 \qquad . \tag{24}$$

Let us start the analysis of the estimator with a general equivalence lemma which is already known from isotropic investigations.

Lemma 3.8 *Let mesh assumption (A1) be satisfied, and consider an arbitrary node x and the associated patch ω_x. Let $\underline{v}$ be a (scalar or vector valued) function defined on ω_x such that $\underline{v}_{|T} \in \mathbb{P}^0(T)$, i.e. $\underline{v}$ is piecewise constant. Let further $\mu_T, T\subset\omega_x$ be arbitrary positive weights such that all μ_T are uniformly bounded away from zero, $\mu_T \geq c > 0$, and that satisfy $\sum_{T\subset\omega_x} \mu_T = 1$. Define the ZZ averaged value by $\underline{v}_{ZZ} := \sum_{T\subset\omega_x} \mu_T \underline{v}_{|T}$. Then*

$$\sum_{E:x\in\mathcal{N}_E} |[\![\underline{v}]\!]_E|^2 \;\sim\; \sum_{T\subset\omega_x} \mu_T|\underline{v}_{|T}|^2 \;-\; |\underline{v}_{ZZ}|^2 \qquad . \tag{25}$$

Proof: For two dimensional domains ($d=2$) this lemma has been proven in [Rod94]; the proof is also repeated in [Ver96, section 1.5]. An extension to three dimensional domains ($d=3$) is readily possible with the ideas from the proof of lemma 3.12. ■

The main result follows now.

Theorem 3.9 (Equivalences with first ZZ estimator) *Let mesh assumptions (A1)–(A4) be satisfied. Then the residual error estimator and the first ZZ error estimator are equivalent:*

$$\eta_{R,x} \;\sim\; \eta_{Z_1,x} \tag{26}$$

$$\eta_R \;\sim\; \eta_{Z_1} \qquad . \tag{27}$$

Proof: We apply the previous lemma 3.8 with $\underline{v} := C_x^{-1}\nabla u_h$ and $\mu_T = |T|/|\omega_x|$ as well as lemma 3.4 to derive

$$\begin{aligned}
\eta_{R,x}^2 \;\stackrel{(16)}{\sim}\; \eta_{\check R,x}^2 \;&=\; h_{min,x}^2|\omega_x| \sum_{E:x\in\mathcal{N}_E} |C_x^{-1}\,[\![\nabla u_h]\!]_E|^2 \\
&\stackrel{(25)}{\sim}\; h_{min,x}^2|\omega_x| \sum_{T\subset\omega_x} \frac{|T|}{|\omega_x|}|C_x^{-1}\,\nabla u_{h|T}|^2 \;-\; \left|\sum_{T\subset\omega_x} \frac{|T|}{|\omega_x|} C_x^{-1}\,\nabla u_{h|T}\right|^2 \\
&\stackrel{(23)}{=}\; \eta_{Z_1,x}^2 \qquad .
\end{aligned}$$

Hence

$$\eta_R^2 = \sum_{x\in\mathcal{N}_{\bar\Omega}} \eta_{R,x}^2 \sim \sum_{x\in\mathcal{N}_{\bar\Omega}} \eta_{Z_1,x}^2 \overset{(24)}{\sim} \eta_{Z_1}^2 \quad .$$

■

Note that this is only an equivalence between the *global* estimators. An equivalence involving the *local* estimator $\eta_{Z_1,T}$ cannot be proven in this way since the projection property (19) is given globally. The procedure of the second ZZ error estimator avoids this drawback.

3.3 Second ZZ error estimator

A different approach to describe a ZZ error estimator is given now. It avoids the global projection property (19) at the cost of a refined analysis. As a consequence local element–wise relations can be derived. We start with the definition of an arbitrary recovered gradient.

Definition 3.10 (Arbitrary recovered gradient)
The arbitrary recovered gradient $\nabla^{\mathrm{R}_2} : W_h \to V_h$ *is defined by the nodal values*

$$(\nabla^{\mathrm{R}_2} u_h)(x) := \sum_{T\subset\omega_x} \mu_{T,x}\, \nabla u_{h|T} \tag{28}$$

where the non–negative weights $\mu_{T,x}$ can be chosen arbitrarily such that $\sum\limits_{T\subset\omega_x} \mu_{T,x} = 1$.

The corresponding second ZZ estimator is given next. Again the first two definitions describe the local (element related) estimator and its global counterpart. The third term is a node related quantity required for the subsequent analysis.

Definition 3.11 (Second anisotropic ZZ estimator) *The local and global ZZ estimators are given by*

$$\eta_{Z_2,T} := h_{min,T}\, \|C_T^{-1}\, (\nabla^{\mathrm{R}_2} u_h - \nabla u_h)\|_T \tag{29}$$

$$\eta_{Z_2}^2 := \sum_{T\in\mathcal{T}_h} \eta_{Z_2,T}^2 \tag{30}$$

$$\eta_{Z_2,x}^2 := h_{min,x}^2 |\omega_x| \sum_{T\subset\omega_x} |C_x^{-1}(\nabla^{\mathrm{R}_2} u_h(x) - \nabla u_{h|T}(x))|^2 \quad . \tag{31}$$

In order to establish a relation between the node related term $\eta_{Z_2,x}$ and the element related estimator $\eta_{Z_2,T}$, recall that $\nabla^{\mathrm{R}_2} u_h - \nabla u_h$ is linear on T. Together with mesh assumptions (A1)–(A3) we conclude

$$\begin{aligned} \eta_{Z_2,T}^2 \quad &\sim \quad h_{min,T}^2 |T| \sum_{x\in\mathcal{N}_T} |C_T^{-1}(\nabla^{\mathrm{R}_2} u_h(x) - \nabla u_{h|T}(x))|^2 \\ &\overset{(5),(A3)}{\sim} \sum_{x\in\mathcal{N}_T} h_{min,x}^2 |\omega_x| \cdot |C_x^{-1}(\nabla^{\mathrm{R}_2} u_h(x) - \nabla u_{h|T}(x))|^2 \quad . \end{aligned}$$

Note that equivalences (5),(A3) have been applied to switch from element related data $h_{min,T}, C_T^{-1}$ to node related data $h_{min,x}, C_x^{-1}$. This yields immediately the desired inequalities

$$\eta_{Z_2,x}^2 \lesssim \sum_{T\subset\omega_x} \eta_{Z_2,T}^2 \tag{32}$$

$$\eta_{Z_2,T}^2 \lesssim \sum_{x\in\mathcal{N}_T} \eta_{Z_2,x}^2 \tag{33}$$

provided that the mesh assumptions (A1)–(A3) are satisfied. Note that the sums on the right–hand side of (32), (33) are necessary because $\eta_{Z_2,x}$ depends on $u_{h|\omega_x}$ whereas $\eta_{Z_2,T}$ depends on u_h on $\bigcup_{x\in\mathcal{N}_T} \omega_x$.

The next lemma states a novel equivalence which is similar to the one of lemma 3.8. The main difference is that now the weights μ_T do not have to be bounded away from 0. The technique to prove this lemma seems to be new.

Lemma 3.12 *Let mesh assumption (A1) be satisfied, and consider an arbitrary node x and the associated patch ω_x. Let $\underline{v}$ be a (scalar or vector valued) function defined on ω_x such that $\underline{v}_{|T} \in \mathbb{P}^0(T)$, i.e. $\underline{v}$ is piecewise constant. Let further $\mu_T, T \subset \omega_x$, be arbitrary non–negative weights such that $\sum_{T\subset\omega_x} \mu_T = 1$. Define $\underline{v}_{ZZ}$ as in lemma 3.8. Then*

$$\sum_{E:x\in\mathcal{N}_E} |[\![\underline{v}]\!]_E|^2 \sim \sum_{T\subset\omega_x} \left|\underline{v}_{ZZ} - \underline{v}_{|T}\right|^2 \quad . \tag{34}$$

Proof: Note first that it suffices to prove (34) component wise, i.e. assume that $\underline{v} \equiv v$ is a scalar, piecewise constant function on ω_x. For simplicity of notation denote the elements of ω_x temporarily by $T_1 \dots T_n$. Accordingly set $\mu_i := \mu_{T_i}$ and $v^i := v_{|T_i}$. The mesh assumptions state that n is bounded from above uniformly on $\mathcal{T}_h$.

We start the proof for an interior node x. Consider first the left hand side of (34) which now reads

$$\sum_{E:x\in\mathcal{N}_E} |[\![\underline{v}]\!]_E|^2 = \sum_{\substack{i,j \\ x\in\mathcal{N}_E, E=T_i\cap T_j}} |v^i - v^j|^2$$

i.e. we sum over all elements T_i and T_j that share a common face E (in 3D) or a common edge (in 2D). The last sum can be written in matrix notation as

$$0 \le \sum_{\substack{i,j \\ x\in\mathcal{N}_E, E=T_i\cap T_j}} |v^i - v^j|^2 = (A\underline{w}, \underline{w}) \tag{35}$$

with

$$\begin{aligned} A &= (a_{i,j})_{i,j=1}^n \in \mathbb{R}^{n\times n} \\ a_{i,j} &= \begin{cases} d & \text{if } i = j \\ -1 & \text{if } T_i \text{ and } T_j \text{ share a common face (3D) or edge (2D)} \\ 0 & \text{otherwise} \end{cases} \\ \underline{w} &:= (v^1, v^2, \dots, v^n)^T \quad . \end{aligned}$$

Obviously $A = A^T$ is positively semidefinite and weakly diagonally dominant. From (35) we further conclude that A has exactly one eigenvalue 0 corresponding to the eigenvector $\underline{w} = \underline{1} := (1, 1, \ldots, 1)^T \in \mathbb{R}^n$; all other eigenvalues are positive. The matrix A depends solely on the *topology* of the patch ω_x but not on its geometry. Since the number of such topologies is finite (n is bounded because of mesh assumption (A1)), there is only a finite number of possibilities for the corresponding matrices A. Hence all *positive* eigenvalues of A are bounded from above and below (and away from 0). Note that in 2D the matrix A simplifies to a circulant tridiagonal matrix consisting of $(-1, 2, -1)$.

Consider next the right hand side of (34) which can be rewritten as

$$\begin{aligned}
\sum_{T\subset\omega_x} \left|\underline{v}_{ZZ} - \underline{v}_{|T}\right|^2 &= \sum_{i=1}^{n} \left| \left(\sum_{j=1}^{n} \mu_j v^j \right) - v^i \right|^2 = (B\underline{w}, \underline{w}) \qquad , \\
\text{with} \qquad B &= (b_{i,j})_{i,j=1}^{n} \in \mathbb{R}^{n\times n} \\
b_{i,j} &= \begin{cases} 1 - 2\mu_i + n\mu_i^2 & \text{for } i = j \\ n\mu_i\mu_j - \mu_i - \mu_j & \text{for } i \neq j \end{cases} \\
&= \delta_{ij} + n\mu_i\mu_j - \mu_i - \mu_j \qquad .
\end{aligned}$$

Introducing $\underline{\mu} := (\mu_1, \ldots, \mu_n)^T \in \mathbb{R}^n$ one derives

$$\begin{aligned}
B &= I + n\underline{\mu}\,\underline{\mu}^T - \underline{\mu}\underline{1}^T - \underline{1}\underline{\mu}^T \\
B - I &= \left(\frac{n}{2}\underline{\mu} - \underline{1}\right)\underline{\mu}^T + \underline{\mu}\left(\frac{n}{2}\underline{\mu} - \underline{1}\right)^T \\
&= \underline{\nu}\underline{\mu}^T + \underline{\mu}\underline{\nu}^T \\
\text{with} \qquad \underline{\nu} &:= \left(\frac{n}{2}\underline{\mu} - \underline{1}\right) \qquad .
\end{aligned}$$

Since $B - I$ is symmetric, it has a full system of eigenvectors. Because $B - I$ is of rank 2, it has $n - 2$ eigenvalues 0. For every other eigenvalue λ of $B - I$ the corresponding eigenvector is a linear combination of $\underline{\mu}$ and $\underline{\nu}$. A simple calculation reveals that then λ is also an eigenvalue of the matrix

$$\begin{bmatrix} \underline{\mu}^T\underline{\nu} & \underline{\mu}^T\underline{\mu} \\ \underline{\nu}^T\underline{\nu} & \underline{\nu}^T\underline{\mu} \end{bmatrix} = \begin{bmatrix} \frac{n}{2}\underline{\mu}^T\underline{\mu} - 1 & \underline{\mu}^T\underline{\mu} \\ \frac{n^2}{4}\underline{\mu}^T\underline{\mu} & \frac{n}{2}\underline{\mu}^T\underline{\mu} - 1 \end{bmatrix} \in \mathbb{R}^{2\times 2} \qquad ,$$

i.e. $\lambda_1 = -1$ and $\lambda_2 = n\underline{\mu}^T\underline{\mu} - 1$. Hence the eigenvalues of B are

$$\lambda(B) = \begin{cases} 0 & \text{single eigenvalue} \\ n\underline{\mu}^T\underline{\mu} & \text{single eigenvalue} \\ 1, \ldots, 1 & n - 2 \text{ times} \end{cases} \qquad .$$

The arithmetic quadratic mean inequality gives

$$1 \leq n\underline{\mu}^T\underline{\mu} = n\sum_{i=1}^{n} \mu_i^2 \leq n \qquad ,$$

hence all positive eigenvalues of B lie in the range $[1, n]$. The eigenvalue 0 is associated with the eigenvector $\underline{1}$.

Summarizing, A and B both have a single eigenvalue 0 corresponding to the same eigenvector $\underline{1}$. All other eigenvalues are positive and bounded from above and below. This implies

$$A \sim B \qquad \text{and} \qquad (A\underline{w}, \underline{w}) \sim (B\underline{w}, \underline{w}) \quad \forall\, \underline{w} \in \mathbb{R}^n$$

which proves the lemma for an interior node x.

For a boundary node x we can proceed in almost the same way. The only difference consists in a slight modification of the matrix A, namely, $a_{i,i} = d - k$ where k is the number of boundary faces (of the element T_i) that contain the node x. The properties of A and the remainder of the proof stay exactly the same as before. ■

Now we are able to prove equivalences with the second ZZ estimator (involving the arbitrary recovered gradient ∇^{R_2}).

Theorem 3.13 (Equivalences with second ZZ estimator) *Let the mesh assumptions (A1)–(A4) be satisfied. Then the following local and global relations hold (for all $x \in \mathcal{N}_{\bar{\Omega}}$ or $T \in \mathcal{T}_h$).*

$$\eta_{R,x} \sim \eta_{Z_2,x} \tag{36}$$
$$\eta_R \sim \eta_{Z_2} \tag{37}$$
$$\eta_{R,x}^2 \lesssim \sum_{T \subset \omega_x} \eta_{Z_2,T}^2 \tag{38}$$
$$\eta_{Z_2,T}^2 \lesssim \sum_{x \in \mathcal{N}_T} \eta_{R,x}^2 \quad . \tag{39}$$

Proof: To prove (36), fix an arbitrary node $x \in \mathcal{N}_{\bar{\Omega}}$ and consider $\underline{v} := C_x^{-1} \nabla u_h$ on the patch ω_x. Since $\underline{v}$ is piecewise constant on ω_x, lemma 3.12 can be applied which implies

$$\underline{v}_{ZZ} = \sum_{T \subset \omega_x} \mu_{T,x} C_x^{-1} \nabla u_{h|T} = C_x^{-1} \nabla^{\mathrm{R}_2} u_h(x) \quad .$$

In conjunction with lemma 3.4 this yields

$$\begin{aligned} \eta_{R,x}^2 \overset{(16)}{\sim} \eta_{\tilde{R},x}^2 &= h_{min,x}^2 |\omega_x| \sum_{E : x \in \mathcal{N}_E} |\, [\![C_x^{-1} \nabla u_h]\!]_E |^2 \\ &\overset{(34)}{\sim} h_{min,x}^2 |\omega_x| \sum_{T \subset \omega_x} | C_x^{-1} \nabla^{\mathrm{R}_2} u_h(x) - C_x^{-1} \nabla u_{h|T} |^2 = \eta_{Z_2,x}^2 \quad . \end{aligned}$$

Next, (38) is a direct consequence of (36) and (32):

$$\eta_{R,x}^2 = \eta_{R,x}^2 \overset{(36)}{\sim} \eta_{Z_2,x}^2 \overset{(32)}{\lesssim} \sum_{T \subset \omega_x} \eta_{Z_2,T}^2 \quad .$$

The converse relation (39) can be concluded similarly:

$$\eta_{Z_2,T}^2 \overset{(33)}{\lesssim} \sum_{x\in\mathcal{N}_T} \eta_{Z_2,x}^2 \overset{(36)}{\sim} \sum_{x\in\mathcal{N}_T} \eta_{R,x}^2 = \sum_{x\in\mathcal{N}_T} \eta_{R,x}^2 \quad .$$

Finally the global equivalence (37) can be proven via (38) and (39).

$$\begin{aligned} \eta_R^2 &= \sum_{x\in\mathcal{N}_{\bar\Omega}} \eta_{R,x}^2 \\ &\overset{(38)}{\lesssim} \sum_{x\in\mathcal{N}_{\bar\Omega}} \sum_{T\subset\omega_x} \eta_{Z_2,T}^2 = (d+1) \sum_{T\in\mathcal{T}_h} \eta_{Z_2,T}^2 = (d+1)\eta_{Z_2}^2 \\ &\overset{(39)}{\lesssim} \sum_{T\in\mathcal{T}_h} \sum_{x\in\mathcal{N}_T} \eta_{R,x}^2 \lesssim \sum_{x\in\mathcal{N}_{\bar\Omega}} \eta_{R,x}^2 = \eta_R^2 \quad . \end{aligned}$$

■

Note that the sums in (38) and (39) appear because $\eta_{R,x}$ is a node related term whereas $\eta_{Z_2,T}$ is an element related quantity.

Theorem 3.14 (ZZ error estimation) *Assume that mesh assumptions (A1)–(A5) are satisfied. Then the error is bounded locally from below and globally from above.*

$$\eta_{Z_2,x} \lesssim \|\nabla(u-u_h)\|_{\omega_x} + \inf_{f_h\in V_h} h_{min,x}\, \|f-f_h\|_{\omega_x} \qquad \forall\, x\in\mathcal{N}_{\bar\Omega} \tag{40}$$

$$\|\nabla(u-u_h)\|_\Omega \lesssim m_1(u-u_h,\mathcal{T}_h)\left(\eta_{Z_2}^2 + \inf_{f_h\in V_h} \sum_{T\in\mathcal{T}_h} h_{min,T}^2 \|f-f_h\|_T^2\right)^{1/2} . \tag{41}$$

Proof: These are immediate consequences of lemma 3.3 and theorem 3.13. ■

Corollary 3.15 (ZZ error estimation on *isotropic* meshes) *Assume that an* isotropic *mesh satisfies mesh assumption (A5). Then the ZZ error estimator $\eta_{Z_2,x}$ is reliable and efficient.*

This holds even when the corresponding recovered gradient ∇^{R_2} is defined with arbitrary *weights (non–negative with sum 1).*

As far as we know this result is new even for isotropic meshes. So far special weights had to be chosen for the recovered gradient in order to prove equivalence with the residual error estimator and, in turn, reliability *and local efficiency* of the ZZ error estimator, cf. [Ver96, Section 1.5]. Now there is the freedom to choose *arbitrary* weights.

Note that reliability alone for an arbitrary recovered gradient has been shown in [CB01]. *Global* efficiency (up to higher order terms) is obtained in the sequel [BC01].

4 The mesh assumptions revisited

As we have seen, the analysis of the ZZ error estimators required several mesh assumptions that were introduced in section 2.3. These assumptions are now discussed in more detail.

In section 4.1 it is shown that there exist meshes which satisfy all assumptions. Sections 4.2 and 4.3 are devoted to mesh assumption (A3) while section 4.4 investigates mesh assumption (A4). With that help we can prove in section 4.5 that the mesh assumptions are satisfied for another class of meshes.

In section 4.6 the role of the mesh assumptions is examined by showing that assumption (A4) is a necessary condition for error estimation. Finally section 4.7 summarizes the occurrences of the mesh assumptions in a compact way.

4.1 Rectangular tensor product type meshes satisfy the mesh assumptions

In this section we prove that the mesh assumptions (A1)–(A5) can be satisfied. To this end we consider rectangular tensor product type tetrahedral meshes. By this we understand that the tetrahedra of $\mathcal{T}_h$ can be grouped such that a set of six of them forms a rectangular hexahedron, cf. also figure 2. Assumption (A1) then clearly holds.

At this stage of generality, assumption (A2) obviously can be satisfied, so we assume that it holds. It states that the dimension of the tetrahedra (in each of the three anisotropic directions) must not change rapidly across neighbouring elements. This assumption is quite weak. It allows, for example, meshes that resolve boundary layers, see e.g. figure 2.

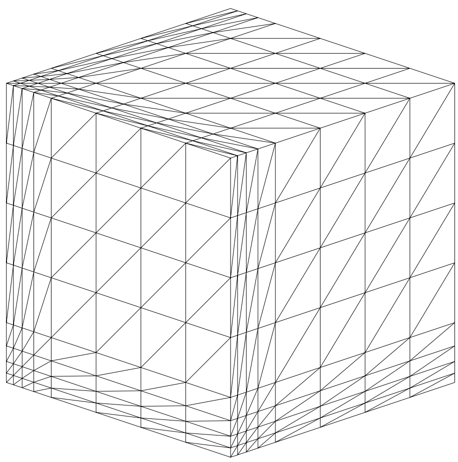

Figure 2: Anisotropic tensor product type mesh

For rectangular tensor product type meshes we now prove that (A1) and (A2) imply (A3)–(A5). Let us start with (A3), i.e. we construct a matrix C_x and show the corresponding properties. Our exposition describes the 3D case; the 2D analogies are straightforward.

Proof of assumption (A3)

Start with a node x of $\mathcal{T}_h$ and an arbitrary tetrahedron $T \subset \omega_x$. For this tetrahedron recall the definition of the matrix C_T, of the orthogonal vectors $\underline{p}_{i,T}$ and their length $h_{i,T}$, $i = 1, 2, 3$, cf. section 2.

Since we consider tensor product type meshes there exists a circumscribing rectangular brick (i.e. hexahedron) $B \supset T$. The three edge lengths of this brick B are denoted by

$$h_{1,B} \geq h_{2,B} \geq h_{3,B} \qquad .$$

Choose corresponding edge vectors $\underline{p}_{i,B}$, $i = 1, 2, 3$, i.e. such that $|\underline{p}_{i,B}| = h_{i,B}$. The orientation of these orthogonal vectors does not matter. The notation is visualized in figure 3.

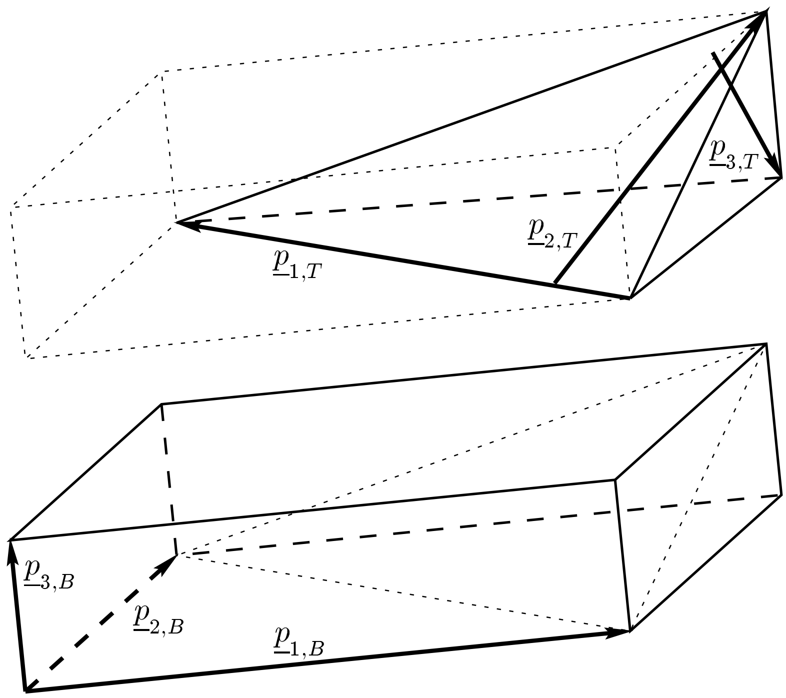

Figure 3: Tetrahedron T and vectors $\underline{p}_{i,T}$ (top)
Box B and vectors $\underline{p}_{i,B}$ (bottom)

Define next the matrix $C_B \in \mathbb{R}^{3\times 3}$ whose columns are formed by the vectors $\underline{p}_{i,B}$,

$$C_B := (\underline{p}_{1,B}, \underline{p}_{2,B}, \underline{p}_{3,B}) \qquad .$$

Finally recall the node related, averaged lengths $h_{i,x}$. Define three orthogonal vectors

$$\underline{p}_{i,x} := \frac{h_{i,x}}{h_{i,B}} \cdot \underline{p}_{i,B} = h_{i,x} \cdot \frac{\underline{p}_{i,B}}{|\underline{p}_{i,B}|} \qquad i = 1, 2, 3$$

which are oriented along the edges vectors $\underline{p}_{i,B}$ of B but which have a different length $|\underline{p}_{i,x}| = h_{i,x}$. Define the matrix $C_x \in \mathbb{R}^{3\times 3}$ by

$$C_x := (\underline{p}_{1,x}, \underline{p}_{2,x}, \underline{p}_{3,x}) \qquad .$$

This immediately implies $C_x^T C_x = \operatorname{diag}(h_{1,x}^2, h_{2,x}^2, h_{3,x}^2)$. Furthermore the geometric properties as well as relations (5) yield the equivalences

$$h_{i,T} \sim h_{i,B} \sim h_{i,x} \qquad i = 1,2,3 \qquad . \tag{42}$$

Now we are ready to prove assumption (A3) which reads

$$|C_x^{-1}\underline{v}| \sim |C_T^{-1}\underline{v}| \qquad \forall \underline{v} \in \mathbb{R}^3 \qquad .$$

Let us start with investigations of the linear transformations associated with C_B and C_T^{-1}. Recall first that $\underline{e}_i \in \mathbb{R}^3$ is the i th unit vector. Because of $C_B \underline{e}_i = \underline{p}_{i,B} \in \mathbb{R}^3$, the transformation via C_B maps the unit cube $[0,1]^3$ onto the brick B (or more precisely onto the corresponding brick at the origin of the coordinate system). Since the four vertices of $T \subset B$ are also vertices of B, the transformation via C_B thus maps

$$C_B : \tilde{T} \to T \qquad ,$$

where $\tilde{T} \subset [0,1]^3$ is a tetrahedron whose four vertices are also vertices of the unit cube $[0,1]^3$. Therefore the diameter $\varrho(\tilde{T})$ of the inscribed sphere of $\tilde{T}$ is of order $\mathcal{O}(1)$,

$$\varrho(\tilde{T}) \sim 1 \qquad .$$

Similarly the second transformation via C_T^{-1} is examined. It maps

$$C_T^{-1} : T \to \hat{T} \qquad ,$$

where the tetrahedron $\hat{T}$ has vertices $(0,0,0)^T$, $(1,0,0)^T$, $(x_2,1,0)^T$ and $(x_3,y_3,1)^T$, with $0 \le x_2, x_3 \le 1$ and $|y_3| \le 1$, cf. the definition of C_T or [Kun00, Section 1.2]. Thus the diameter $\operatorname{diam}(\hat{T})$ of the tetrahedron $\hat{T}$ satisfies

$$1 < \operatorname{diam}(\hat{T}) \le \sqrt{6} \qquad .$$

The combined transformation via $C_T^{-1} C_B$ now maps

$$C_T^{-1} C_B : \tilde{T} \to \hat{T} \qquad .$$

Hence the spectral norm of this matrix can be bounded from above by

$$\|C_T^{-1} C_B\| \le \frac{\operatorname{diam}(\hat{T})}{\varrho(\tilde{T})} \lesssim 1 \qquad .$$

This inequality can be used to derive the matrix bound

$$\begin{aligned}\|C_T^{-1}C_xC_x^TC_T^{-T}\| &= \|C_T^{-1}C_B\cdot C_B^{-1}C_xC_x^TC_B^{-T}\cdot C_B^TC_T^{-T}\| \le \|C_B^{-1}C_x\|^2\cdot\|C_T^{-1}C_B\|^2\\ &\lesssim \max_{i=1,2,3}\frac{h_{i,x}^2}{h_{i,B}^2}\cdot 1 \lesssim 1\end{aligned}$$

since $C_B^{-1}C_x = \mathrm{diag}(h_{1,x}/h_{1,B}\,,\,h_{2,x}/h_{2,B}\,,\,h_{3,x}/h_{3,B})$, and because of (42). The first matrix $M := C_T^{-1}C_xC_x^TC_T^{-T}$ is symmetric and positive definite. For such matrices the largest eigenvalue is $\lambda_{max}(M) = \|M\|$ and hence

$$\lambda_{max}(C_T^{-1}C_xC_x^TC_T^{-T}) \lesssim 1 \qquad .$$

In a completely analogous fashion one treats $M^{-1} = C_T^TC_x^{-T}C_x^{-1}C_T$ to obtain

$$\lambda_{max}(M^{-1}) = \|C_T^TC_x^{-T}C_x^{-1}C_T\| \lesssim 1 \qquad .$$

This implies $\lambda_{min}(M) = (\lambda_{max}(M^{-1}))^{-1} \gtrsim 1$, i.e. all eigenvalues of M are of order $\mathcal{O}(1)$. Since the eigenvalues of $M = C_T^{-1}C_xC_x^TC_T^{-T}$ and of $(C_x^{-T}C_x^{-1})^{-1}C_T^{-T}C_T^{-1}$ are the same, one further concludes

$$\underline{v}^TC_x^{-T}C_x^{-1}\underline{v} \sim \underline{v}^TC_T^{-T}C_T^{-T}\underline{v} \qquad \forall \underline{v}\in\mathbb{R}^3 \qquad .$$

This finally gives $|C_x^{-1}\underline{v}| \sim |C_T^{-1}\underline{v}|$ for all $\underline{v}\in\mathbb{R}^3$ which proves the remaining equivalence of (A3).

Proof of assumption (A4)

We now prove that (A2) also yields (A4). Thus let T be an arbitrary tetrahedron and E be any face thereof. Employ the notation of the previous paragraphs and consider the brick B that circumscribes T. Then C_B^{-1} maps T onto $\tilde{T}$ (see above). Next we consider the vector $h_{E,T}\,\underline{n}_E$ in a geometric way. If the unit vector $\underline{n}_E$ points inward (with respect to T) then $h_{E,T}\,\underline{n}_E$ points from the face E of T (or its plane) to the opposite vertex of T. If $\underline{n}_E$ is the outward vector then consider $-h_{E,T}\,\underline{n}_E$ instead.

Therefore $C_B^{-1}h_{E,T}\,\underline{n}_E$ is a vector that points from the face $\tilde{E} := C_B^{-1}E$ of $\tilde{T}$ to the opposite vertex of $\tilde{T}$. This results in

$$1 \sim \varrho(\tilde{T}) < |C_B^{-1}h_{E,T}\,\underline{n}_E| < \sqrt{3} \qquad \text{, i.e.} \qquad |C_B^{-1}\underline{n}_E| \sim h_{E,T}^{-1} \qquad .$$

Next recall that $C_x^{-1}C_B$ is a diagonal matrix. Apply the equivalence $h_{i,B}\sim h_{i,x}$ from above to conclude

$$\min_{i=1,2,3}\frac{h_{i,B}}{h_{i,x}}\cdot|C_B^{-1}\underline{n}_E| \le |C_x^{-1}\underline{n}_E| = |C_x^{-1}C_B\cdot C_B^{-1}\underline{n}_E| \le \max_{i=1,2,3}\frac{h_{i,B}}{h_{i,x}}\cdot|C_B^{-1}\underline{n}_E| \qquad .$$

In conjunction with (A3) one finally arrives at the desired equivalence

$$h_{E,T}^{-1} \sim |C_B^{-1}\underline{n}_E| \sim |C_x^{-1}\underline{n}_E| \overset{\text{(A3)}}{\sim} |C_T^{-1}\underline{n}_E| \qquad .$$

Proof of assumption (A5)

For (A5) to hold we have to specify assumption (A2) slightly more precisely, namely we demand

$$\frac{h_{min,T_1}}{h_{min,T_2}} < \alpha_d := \begin{cases} \sqrt{2}+\sqrt{3} & \approx \; 3.146 \qquad \text{if } d=2 \\ (3+\sqrt{5})/2 & \approx \; 2.618 \qquad \text{if } d=3 \end{cases} \qquad \forall T_1 \cap T_2 \neq \emptyset \quad .$$

This slightly more restrictive assumption on the change of $h_{min,T}$ across neighbouring elements immediately implies the first inequality of (4) in (A5).

In order to investigate the neighbourhood patches $R_k(T)$ observe first that $\bigcup_{l=0}^{k} R_l(T)$ contains $\mathcal{O}(k^d)$ elements. Hence $R_k(T)$ contains $\mathcal{O}(k^{d-1})$ elements, and the second inequality of (4) in (A5) holds with $r = d-1, \beta = 1$.

With these values of α_d and β the third inequality of (4) in (A5) is satisfied as well.

4.2 Assumption (A3) implies (A1) and (A2)

In this section it is proven that assumptions (A1) and (A2) are already consequences of assumption (A3).

First we show that (A3) implies a bounded number of elements in each patch ω_x (uniformly over $\mathcal{T}_h$). To this end consider a patch ω_x and an arbitrary element T thereof. In both the 2D and 3D case take an arbitrary *edge* of T and denote the corresponding edge vector by $\underline{v}$. Assumption (A3) yields for some matrix C_x

$$|C_x^{-1}\underline{v}| \;\sim\; |C_T^{-1}\underline{v}| \qquad .$$

Next consider the inverse mapping C_T^{-1} which maps T onto $\hat{T}$ (for simplicity we omit the transitional part of the mapping; see also section 2.2). Thus $\underline{\hat{v}} := C_T^{-1}\underline{v}$ is an edge of $\hat{T}$. Hence this mapped edge has a length $|\underline{\hat{v}}| = |C_T^{-1}\underline{v}| \sim \operatorname{diam}(\hat{T}) \sim 1$ which gives

$$|C_x^{-1}\underline{v}| \;\sim\; 1 \qquad .$$

Since this holds for all edges of ω_x, the transformation via C_x^{-1} maps ω_x onto a patch $C_x^{-1}\omega_x$ whose edges all have a length of $\mathcal{O}(1)$. Therefore $C_x^{-1}\omega_x$ is a patch consisting of *isotropic* elements. Clearly the number of elements in such a patch is bounded, and so is the number of elements in the original patch ω_x. Thus (A1) holds uniformly over $\mathcal{T}_h$.

Next we prove that assumption (A3) also implies (A2). Consider again a patch ω_x and two arbitrary elements T_1, T_2 thereof. Apply the transformation via C_x^{-1} to the patch ω_x which results in the transformed patch $C_x^{-1}\omega_x$ having the transformed elements $\tilde{T}_i := C_x^{-1}T_i, i = 1, 2$, cf. figure 4. Above we have proved that $\tilde{T}_i$ are *isotropic* elements of size $\mathcal{O}(1)$. Therefore we can scale $\tilde{T}_2$ with some factor $\alpha \sim 1$ such that

$$\alpha \cdot \tilde{T}_2 \subset \tilde{T}_1 \qquad .$$

In figure 4, $\tilde{T}_2$ and $\alpha \cdot \tilde{T}_2$ are depicted by the shaded triangles (the necessary translation is again omitted for ease of notation).

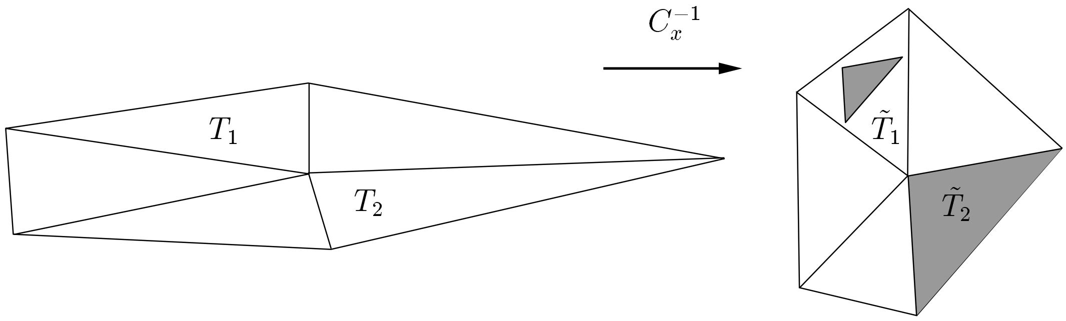

Figure 4: Patches ω_x (left) and $C_x^{-1}\omega_x$ (right)

The transformation via C_x back to the original domain yields

$$\alpha \cdot T_2 \subset T_1 \quad .$$

and thus

$$\begin{array}{rcl} \alpha \cdot |T_2| & \leq & |T_1| \\ \alpha \cdot h_{1,T_2} & \leq & \mathrm{diam}(T_1) \sim h_{1,T_1} \\ \varrho(\alpha T_2) \sim \alpha \cdot h_{min,T_2} & \leq & h_{min,T_1} \quad . \end{array}$$

Since T_1, T_2 are completely arbitrary and thus interchangeable, and because of a scaling factor $\alpha \sim 1$, one obtains

$$\begin{array}{rcl} |T_1| & \sim & |T_2| \\ h_{1,T_1} & \sim & h_{1,T_2} \\ h_{min,T_1} & \sim & h_{min,T_2} \quad . \end{array}$$

In the 2D case this already constitutes the desired assumption (A2). For the 3D case recall additionally that $6\,|T| = h_{1,T} \cdot h_{2,T} \cdot h_{3,T}$ which results in the remaining equivalence $h_{2,T_1} \sim h_{2,T_2}$.

Remark 4.1 *Assumption (A3) implies (A1) and (A2) but not vice versa, as a comparatively simple counterexample can show. Thus (A3) is a stronger assumption.*

4.3 Necessary and sufficient condition for mesh assumption (A3)

Here we state a geometrical condition which is necessary and sufficient for assumption (A3) on *unstructured* tetrahedral meshes. Recall that $\underline{p}_{i,T}$ are the three main anisotropic direction vectors of an element T, and that $h_{i,x}$ are the averaged lengths of a patch ω_x, $i = 1 \ldots d$, cf. section 2.2. The space dimension is either $d = 3$ or $d = 2$. We start with some technical equivalences.

Lemma 4.2 *The assumption (A3) is equivalent to the condition*

$$||C_x^{-1}C_T|| \sim 1 \qquad \textit{and} \qquad ||C_T^{-1}C_x|| \sim 1 \qquad \forall\, T \subset \omega_x \textit{ and all nodes } x \qquad . \tag{43}$$

Proof: $\Rightarrow$: Starting from (A3) and taking $\underline{v} := C_T\underline{w}$, we get

$$|C_x^{-1}C_T\underline{w}| = |C_x^{-1}\underline{v}| \overset{(A3)}{\sim} |C_T^{-1}\underline{v}| = |\underline{w}| \qquad \forall\, \underline{w} \in \mathbb{R}^d, \forall\, T \subset \omega_x \qquad .$$

This yields

$$||C_x^{-1}C_T|| = \max_{|\underline{w}|=1} |C_x^{-1}C_T\underline{w}| \sim 1.$$

We obtain similarly the second bound by taking $\underline{v} := C_x\underline{w}$.
$\Leftarrow$: Define the symmetric, positive definite matrix $M := C_T^{-1}C_xC_x^TC_T^{-T}$. Completely analogous to section 4.1 one concludes

$$\lambda_{max}(M) = \|C_T^{-1}C_xC_x^TC_T^{-T}\| \leq \|C_T^{-1}C_x\|^2 \sim 1$$

and

$$\lambda_{min}(M) = (\lambda_{max}(M^{-1}))^{-1} = \|C_T^TC_x^{-T}C_x^{-1}C_T\|^{-1} \geq \|C_x^{-1}C_T\|^{-2} \sim 1 \qquad .$$

Hence all eigenvalues of M are of order $\mathcal{O}(1)$. Following once more the arguments of section 4.1 yields

$$|C_T^{-1}\underline{v}| \;\sim\; |C_x^{-1}\underline{v}| \qquad \forall\, \underline{v} \in \mathbb{R}^d \qquad ,$$

which is nothing else than (A3). ■

Corollary 4.3 *The assumption (A3) is equivalent to the condition*

$$||C_{T_1}^{-1}C_{T_2}|| \lesssim 1 \qquad \forall\, T_1, T_2 \subset \omega_x \textit{ and all nodes } x. \tag{44}$$

Proof: For the necessity of the condition (44) apply lemma 4.2 and write

$$||C_{T_1}^{-1}C_{T_2}|| = ||C_{T_1}^{-1}C_xC_x^{-1}C_{T_2}|| \leq ||C_{T_1}^{-1}C_x|| \cdot ||C_x^{-1}C_{T_2}|| \sim 1 \qquad \forall\, T_1, T_2 \subset \omega_x \qquad .$$

The sufficiency of (44) follows directly by the choice $C_x := C_{T'}$ for an arbitrary element $T' \subset \omega_x$. ■

Theorem 4.4 (Equivalent formulation of (A3)) *Assume that for all patches ω_x and any two elements $T_1, T_2 \subset \omega_x$ the inequality*

$$\left|\cos\sphericalangle[\underline{p}_{i,T_1}, \underline{p}_{j,T_2}]\right| \lesssim \frac{h_{i,T_1}}{h_{j,T_2}} \qquad \forall\, 1 \leq i, j \leq d \tag{45}$$

is satisfied. Then we can fix an arbitrary element $T' \subset \omega_x$ and set $C_x := C_{T'}$. This choice implies assumption (A3), i.e.

$$|C_x^{-1}\underline{v}| \;\sim\; |C_T^{-1}\underline{v}| \qquad \forall\, \underline{v} \in \mathbb{R}^d, \forall\, T \subset \omega_x \qquad .$$

Conversely the assumption (A3) implies that (45) holds for all $T_1, T_2 \subset \omega_x$ and all nodes x.

Proof: Let us first derive an equivalent formulation of inequality (45). Fix an arbitrary patch ω_x and two arbitrary elements $T_1, T_2 \subset \omega_x$. Since the vectors $\underline{p}_{i,T_1}$ are mutually orthogonal, there exists a unique decomposition

$$\underline{p}_{j,T_2} = \sum_{i=1}^{d} \alpha_{ij} \cdot \underline{p}_{i,T_1} \qquad \forall\, j = 1 \ldots d \qquad .$$

The real coefficients α_{ij} satisfy

$$(\underline{p}_{j,T_2}, \underline{p}_{i,T_1}) = \alpha_{ij} \cdot (\underline{p}_{i,T_1}, \underline{p}_{i,T_1}) = \alpha_{ij} \cdot h_{i,T_1}^2 \qquad .$$

Utilizing the definition of h_{i,T_k} one obtains

$$\alpha_{ij} = \frac{h_{j,T_2} h_{i,T_1} \cdot \cos \sphericalangle[\underline{p}_{i,T_1}, \underline{p}_{j,T_2}]}{h_{i,T_1}^2} = \frac{h_{j,T_2}}{h_{i,T_1}} \cdot \cos \sphericalangle[\underline{p}_{i,T_1}, \underline{p}_{j,T_2}] \qquad .$$

Condition (45) of the theorem is thus equivalent to

$$|\alpha_{ij}| \lesssim 1 \qquad \forall\, 1 \le i, j \le d \qquad .$$

Recall next that the matrices C_{T_1}, C_{T_2} are formed by

$$C_{T_k} := (\underline{p}_{1,T_k}, \underline{p}_{2,T_k}, \underline{p}_{3,T_k}) \qquad ,$$

cf. section 2.2, which results in

$$\begin{aligned} C_{T_1}^{-1} \underline{p}_{j,T_2} &= C_{T_1}^{-1} \sum_{i=1}^{d} \alpha_{ij} \underline{p}_{i,T_1} = \sum_{i=1}^{d} \alpha_{ij} \underline{e}_i \\ C_{T_1}^{-1} C_{T_2} &= (\alpha_{ij})_{i,j=1}^{d} \\ \left\| C_{T_1}^{-1} C_{T_2} \right\| &\sim \max_{i,j=1\ldots d} |\alpha_{ij}| \qquad . \end{aligned}$$

Hence $\|C_{T_1}^{-1} C_{T_2}\| \lesssim 1$ is equivalent to $|\alpha_{ij}| \lesssim 1\ \forall\, i, j$ and to (45). From here we conclude the desired result thanks to Corollary 4.3. ∎

Remark 4.5 *The previous theorem provides the means for practical tests whether assumption (A3) is satisfied on a real mesh. For neighbouring elements one has to compute the angle between the main anisotropic direction vectors* $\underline{p}_{i,T_1}$ *and* $\underline{p}_{j,T_2}$ *and compare its cosine with the stretching ratio* $h_{i,T_1}/h_{j,T_2}$.

4.4 Necessary and sufficient condition for mesh assumption (A4)

In this section we give equivalent formulations of mesh assumption (A4), both of which are geometrically characterized.

Theorem 4.6 (Equivalent formulation of (A4)) *The assumption (A4) holds if and only if for all elements T and all faces $E \subset \partial T$ one has*

$$\max_{i=1,\cdots,d} \frac{|\cos \sphericalangle[\underline{p}_{i,T}, \underline{n}_E]|}{h_{i,T}} \lesssim h_{E,T}^{-1} \quad . \tag{46}$$

Proof: Fix an element T and a face $E \subset \partial T$. As before we may write in a unique way

$$\underline{n}_E = \sum_{i=1}^{d} \alpha_i \cdot \underline{p}_{i,T} \quad ,$$

with

$$(\underline{n}_E, \underline{p}_{i,T}) = \alpha_i \cdot h_{i,T}^2 \qquad \text{and} \qquad \alpha_i = \frac{\cos \sphericalangle[\underline{p}_{i,T}, \underline{n}_E]}{h_{i,T}} \quad .$$

Since $C_T^{-1}\underline{p}_{i,T} = \underline{e}_i$ we obtain $C_T^{-1}\underline{n}_E = (\alpha_1, \alpha_2, \alpha_3)^T$. From the equivalence of norms in $\mathbb{R}^d$ we conclude

$$||C_T^{-1}\underline{n}_E|| \sim \max_{i=1,\cdots,d} \frac{|\cos \sphericalangle[\underline{p}_{i,T}, \underline{n}_E]|}{h_{i,T}}$$

which finishes the proof. ∎

Next we derive a purely geometrical characterization of (A4). This assumption states

$$|C_T^{-1} h_{E,T}\underline{n}_E| \ \sim \ 1 \qquad \forall\, E \subset \partial T \quad .$$

Thus fix an arbitrary element T. Given a face $E \subset \partial T$, denote temporarily its opposite vertex by V_E. Let U_E be the orthogonal projection of V_E onto E (or the plane that contains E). Hence $\overrightarrow{U_E V_E}$ is the height of V_E onto the plane of E.

Next we have to define an appropriate neighbourhood of E. To this end denote by E^α the face E scaled by the real factor α with respect to the midpoint M_E of E. In vector notation this can be written as $E^\alpha := \{\vec{M}_E + \alpha \cdot (\underline{y} - \vec{M}_E) : \underline{y} \in E\}$. In other words, E^α is contained in the plane of E, and $E^1 \equiv E$.

With that definition we can reformulate (A4) as an equivalent geometrical condition.

Theorem 4.7 (Equivalent formulation of (A4))
(A4) holds if and only if $U_E \in E^\alpha$ is satisfied for all $E \subset \partial T$ with some $\alpha \lesssim 1$.

Proof: Obviously the vector $\overrightarrow{U_E V_E}$ equals

$$\overrightarrow{U_E V_E} = \pm h_{E,T}\underline{n}_E \quad .$$

Since C_T^{-1} maps T onto $\hat{T}$, the vector $\overrightarrow{U_E V_E}$ is mapped onto a vector from the point $\hat{U}_E := C_T^{-1}(U_E)$ of the the face $\hat{E} := C_T^{-1}(E)$ of $\hat{T}$ to the opposite vertex $\hat{V}_E := C_T^{-1}(V_E)$. Utilizing $\pm C_T^{-1} h_{E,T} \underline{n}_E = C_T^{-1}(\overrightarrow{U_E V_E}) = \overrightarrow{\hat{U}_E \hat{V}_E}$, assumption (A4) can be rewritten as

$$\mid \overrightarrow{\hat{U}_E \hat{V}_E} \mid \sim 1 \qquad .$$

Because $\hat{T}$ is an isotropic tetrahedron of size $\mathcal{O}(1)$ and $\hat{V}_E$ is a vertex thereof, this is equivalent to $\hat{U}_E \in \hat{E}^\alpha$ and $U_E \in E^\alpha$, with $\alpha \lesssim 1$. ■

4.5 Prismatic tensor product type meshes satisfy the mesh assumptions

In section 4.1 we have shown that the mesh assumptions are satisfied for tetrahedral meshes which are the tensor product of *three 1D meshes*. In this section we prove that the assumptions (A3)–(A5) hold also for anisotropic tensor product meshes of a *prismatic domain* $\Omega = G \times (a, b)$ with $a < b$, obtained using a 2D refined isotropic mesh of Raugel's type in G and a uniform mesh in the third direction. Examples of such meshes are given in the right part of figure 7 and in [Ape99].

We define families of meshes $\mathcal{T}_h$ of Ω by introducing in G the standard mesh grading for two-dimensional corner problems, see for example [OR79, Rau78]. Let $\mathcal{T}_G = \{K\}$ be a regular isotropic triangulation of G; the elements K are triangles. Let r_K be the distance of K to the corner,

$$r_K := \inf_{(x_1,x_2)\in K} (x_1^2 + x_2^2)^{1/2},$$

(note that Ω is scaled such that $r_K < 1$). With h being a global mesh parameter and $\mu \in (0, 1]$ being a grading parameter, we assume that the element size $h_K := \operatorname{diam} K$ satisfies

$$h_K \sim \begin{cases} h^{1/\mu} & \text{for } r_K = 0, \\ h r_T^{1-\mu} & \text{for } r_K > 0. \end{cases}$$

This graded two-dimensional mesh is now extended in the third dimension using the uniform mesh size h. In this way we obtain a pentahedral (i.e. prismatic) triangulation and, by dividing each pentahedron into three tetrahedra, we further get a tetrahedral triangulation $\mathcal{T}_h$ of Ω, see the right part of figure 7 for an illustration.

Proof of assumption (A3)

Each tetrahedron T is included in a prism $Q = K \times I$, where K is an isotropic element in G of diameter $h_K \lesssim h$ and I is a real interval of length $h_I \sim h$ (this notation will be used in the rest of the section). We now define the matrix

$$C_Q = \begin{pmatrix} 0 & \multicolumn{2}{c}{\multirow{2}{*}{B_K}} \\ 0 & & \\ h_I & 0 & 0 \end{pmatrix} \qquad ,$$

where B_K is a 2×2 matrix which maps the usual reference element $\hat{K}$ of $\mathbb{R}^2$ onto K. Then $||B_K|| \sim h_K$. Since C_Q^{-1} maps T into an element $\tilde{T}$ such that $\rho(\tilde{T}) \sim 1$ as well as $\operatorname{diam}\tilde{T} \sim 1$ and C_T^{-1} maps T into a reference element $\hat{T}$ such that $\rho(\hat{T}) \sim 1$ and $\operatorname{diam}\hat{T} \sim 1$, we get

$$||C_T^{-1}C_Q|| \leq \frac{\operatorname{diam}\hat{T}}{\rho(\tilde{T})} \sim 1,$$

and similarly

$$||C_Q^{-1}C_T|| \leq \frac{\operatorname{diam}\tilde{T}}{\rho(\hat{T})} \sim 1.$$

By lemma 4.2, we obtain

$$|C_T^{-1}\underline{v}| \ \sim \ |C_Q^{-1}\underline{v}| \qquad \forall\, \underline{v} \in \mathbb{R}^d \qquad .$$

For a node x we now define

$$C_x := \begin{pmatrix} 0 & h_{2,x} & 0 \\ 0 & 0 & h_{2,x} \\ h_{1,x} & 0 & 0 \end{pmatrix} \qquad .$$

For any prism $Q = K \times I$ that has x as a node, the construction of the mesh implies

$$h_{1,x} \sim h_I \qquad \text{and} \qquad h_{2,x} \sim h_{3,x} \sim h_K \qquad .$$

Subsequently one obtains

$$||C_Q^{-1}C_x|| \sim ||C_x^{-1}C_Q|| \sim 1 \qquad .$$

Applying lemma 4.2 one more, this yields for the node x of the prism Q

$$|C_Q^{-1}\underline{v}| \ \sim \ |C_x^{-1}\underline{v}| \qquad \forall\, \underline{v} \in \mathbb{R}^d \qquad .$$

In combination with $|C_T^{-1}\underline{v}| \ \sim \ |C_Q^{-1}\underline{v}|$ from above one ends up with the desired equivalence (A3).

Proof of assumption (A4)

To check this assumption we use the equivalent formulation given by theorem 4.6. Thus consider a tetrahedron T lying inside the prism $Q = K \times I$.

If $h_I \leq h_K$ then T is an 'outer' tetrahedron ($r_T \sim 1$), and one even has $h_I \sim h_K$. Hence T is an isotropic element for which assumption (A4) always holds.

For all other tetrahedra T one has $h_I > h_K$ and $h_{1,T} \sim h_I$, $h_{2,T} \sim h_{3,T} \sim h_K$. For such anisotropic elements we need to distinguish between 'large' faces and 'small' faces of T.

a) For a large face E, we have $|E| \sim h_{1,T} h_{3,T}$ and therefore $h_{E,T} \sim h_{3,T}$. Consequently for such a face the condition (46) always holds, i.e.

$$\max_{i=1,\cdots,d} \frac{|\cos \sphericalangle[\underline{p}_{i,T}, \underline{n}_E]|}{h_{i,T}} \lesssim h_{E,T}^{-1} \quad .$$

b) For small faces the situation is more delicate since $|E| \sim h_{3,T}^2$ and therefore $h_{E,T} \sim h_{1,T} \sim h_I$. This situation only occurs for a tetrahedron T having a face parallel to the x_1, x_2-plane. Then the small face E is such that $\underline{n}_E = \pm \underline{e}_3$. From

$$\cos \sphericalangle[\underline{p}_{i,T}, \underline{n}_E] = \pm \frac{(\underline{p}_{i,T})_3}{h_{i,T}} \quad ,$$

where $(p_{i,T})_k$ means the k^{th} component of the vector $p_{i,T}$, we see that

$$\frac{\left|\cos \sphericalangle[\underline{p}_{i,T}, \underline{n}_E]\right|}{h_{i,T}} = \frac{|(\underline{p}_{i,T})_3|}{h_{i,T}^2} \quad . \tag{47}$$

This directly yields (46) for $i = 1$ because $|(\underline{p}_{1,T})_3| \leq |\underline{p}_{1,T}| = h_{1,T}$.

For $i = 2$ or 3, we need to estimate $|(\underline{p}_{i,T})_3|$. The orthogonality relation $(p_{i,T}, p_{1,T}) = 0$ for $i = 2$ or 3 yields

$$(\underline{p}_{i,T})_3 (\underline{p}_{1,T})_3 = -\sum_{k=1}^{2} (\underline{p}_{i,T})_k (\underline{p}_{1,T})_k \quad .$$

The assumption $h_I > h_K$ implies that $\underline{p}_{1,T}$ is an edge common to two large faces of T and consequently

$$\begin{array}{rcll} & |(\underline{p}_{1,T})_3| & \sim & h_{1,T} \\ & |(\underline{p}_{1,T})_k| & \lesssim & h_{3,T} \qquad \forall k = 1, 2 \\ \text{as well as} & |(\underline{p}_{i,T})_k| & \leq & |\underline{p}_{i,T}| \sim h_{3,T} \qquad \forall i = 2, 3, \forall k = 1, 2, 3 \quad . \end{array}$$

Combining these inequalities and equivalences yields immediately

$$|(\underline{p}_{i,T})_3| \lesssim \frac{h_{3,T}^2}{h_{1,T}} \qquad \forall i = 2, 3 \quad .$$

Inserting this bound into (47) proves that (46) holds also for $i = 2, 3$; hence (A4) follows from theorem 4.6.

Proof of assumption (A5)

This assumption holds under exactly the same conditions as described for the rectangular tensor product type meshes of section 4.1. The completely analogous reasoning is thus omitted.

4.6 Mesh assumption (A4) is necessary for error estimation

In the previous sections we investigated what meshes satisfy the mesh assumptions. In contrast, this section sheds light on the *role* that the mesh assumptions play in error estimation.

Our main theorem 3.13 states that mesh assumptions (A1)–(A4) are *sufficient* to prove equivalences between the residual error estimator and the ZZ error estimator. Here we prove that mesh assumption (A4) is also a *necessary* condition. To this end we present a 2D counterexample where (A4) is violated and consequently the desired equivalences no longer hold.

Consider a criss–cross type mesh with nodal points located at

$$x_{ik} \;=\; \begin{pmatrix} i \cdot h_1 + k \cdot h_2 \\ k \cdot h_2 \end{pmatrix} \;=\; i \cdot h_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + k \cdot h_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad i, k \in \mathbf{Z} \qquad ,$$

where $0 < h_2 \ll h_1$ are fixed parameters, cf. figure 5.

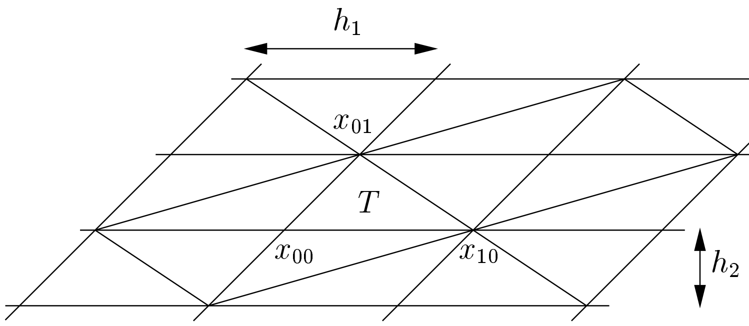

Figure 5: Mesh for the counterexample

This mesh clearly satisfies assumptions (A1) and (A2). It can be verified easily that assumption (A3) holds as well for the choice $C_x := \mathrm{diag}(h_1, h_2)$ for all nodes x of $\mathcal{T}_h$. Mesh assumption (A4) however is violated:

$$|C_T^{-1} \underline{n}_E| \;\sim\; h_{E,T}^{-1} \qquad \forall\, E \subset \partial T$$

does not hold anymore. This is mainly due to the fact that the triangles are not rectangular (or at least close to that).

We now prescribe a finite element solution $u_h := \max\{0, y - x, x - y - h_1\}$ which has in particular the nodal values

$$u_h(x_{ik}) \;=\; \begin{cases} 0 & \text{for } i = 0 \text{ or } i = 1 \\ h_1 & \text{for } i = 2 \text{ or } i = -1 \end{cases} \qquad ,$$

see figure 6. In the next paragraphs we compute the residual error estimator $\eta_{R,x}$ for the nodes x_{00}, x_{10}, x_{01} as well as the ZZ error estimator $\eta_{Z_2,T}$ for the triangle T that corresponds to the three aforementioned nodes. It will turn out that the desired equivalence of both error estimators does not hold anymore due to the violation of assumption (A4).

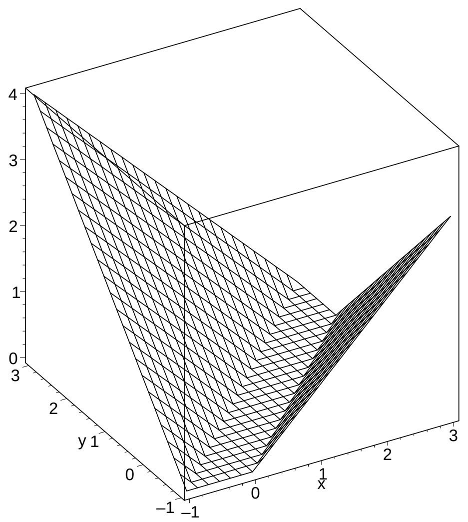

Figure 6: Finite element solution u_h for the counterexample

Residual error estimator $\eta_{R,x}$

A straightforward computation of the required terms results in

$$
\begin{array}{rcll}
h_{min,T} & \sim & h_2 & \forall\, T \in \mathcal{T}_h \\
h_{min,x} & \sim & h_2 & \forall\, x \in \mathcal{N}_{\bar\Omega} \\
|\omega_{x_{ik}}| & = & \left\{ \begin{array}{l} 2h_1h_2 \\ 4h_1h_2 \end{array} \right. & \begin{array}{l} \text{for } i+k \text{ even} \\ \text{for } i+k \text{ odd} \end{array} \\
[\![\partial_{n_E} u_h]\!]_E & = & \left\{ \begin{array}{l} \sqrt{2} \\ 0 \end{array} \right. & \begin{array}{l} \forall \text{ edges } E \text{ on the lines } y=x \text{ or } y = x-h_1 \\ \text{otherwise} \end{array} \\
h_E & = & \dfrac{1}{\sqrt{2}} h_1 & \forall \text{ edges } E \text{ parallel to the line } y = x \qquad .
\end{array}
$$

Altogether one obtains

$$
\begin{array}{rcl}
\dfrac{1}{2}\eta^2_{R,x_{10}} = \dfrac{1}{2}\eta^2_{R,x_{01}} = \eta^2_{R,x_{00}} & \stackrel{(11)}{=} & h^2_{min,x_{00}} |\omega_{x_{00}}| \displaystyle\sum_{E:x_{00}\in\mathcal{N}_E} h_E^{-2} [\![\partial_{n_E} u_h]\!]_E^2 \\
& \sim & h_2^2 \cdot 2h_1h_2 \cdot 2 \cdot \left(\dfrac{1}{\sqrt{2}} h_1\right)^{-2} \cdot \sqrt{2}^2 \\
& \sim & h_1^{-1} h_2^3
\end{array}
$$

and eventually

$$
\sum_{x\in\mathcal{N}_T} \eta^2_{R,x} \sim h_1^{-1} h_2^3 \qquad .
$$

ZZ error estimator $\eta_{Z_2,T}$

Here we utilize a recovered gradient with weights

$$\mu_{T,x_{ik}} := \frac{|T|}{|\omega_{x_{ik}}|} = \begin{cases} 1/4 & \text{for } i+k \text{ even} \\ 1/8 & \text{for } i+k \text{ odd} \end{cases} ,$$

cf. (28). The terms that are required to compute the ZZ error estimator $\eta_{Z_2,T}$ then become

$$\begin{aligned}
\nabla^{\mathrm{R}_2} u_h(x_{00}) \;=\; \nabla^{\mathrm{R}_2} u_h(x_{01}) \;=\; -\nabla^{\mathrm{R}_2} u_h(x_{10}) &= \frac{1}{2}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
\nabla^{\mathrm{R}_2} u_{h|T} &= \left(\frac{x-y}{h_1} - \frac{1}{2}\right) \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
\nabla u_{h|T} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
C_T &= \mathrm{diag}(h_1, h_2) \\
h_{min,T} &= h_2 \quad .
\end{aligned}$$

The ZZ error estimator now evaluates to

$$\eta^2_{Z_2,T} \overset{(29)}{=} h^2_{min,T} \, \|C_T^{-1} \, (\nabla^{\mathrm{R}_2} u_h - \nabla u_h)\|^2_T \;=\; h_2^2 \cdot \frac{h_1^2 + h_2^2}{24 h_1 h_2} \sim h_1 h_2 \qquad .$$

In conjunction with the results from above one concludes

$$\begin{aligned}
\eta^2_{Z_2,T} &\sim \frac{h_1^2}{h_2^2} \sum_{x \in \mathcal{N}_T} \eta^2_{R,x} \\
\eta^2_{Z_2,T} &\not\gtrsim \sum_{x \in \mathcal{N}_T} \eta^2_{R,x} \qquad ,
\end{aligned}$$

i.e. relation (39) of theorem 3.13 does not hold anymore. Consequently the local equivalence (36) is violated at certain nodes x of $\mathcal{T}_h$. This proves the necessity of the mesh assumption (A4).

4.7 Short summary of the Assumptions

The table below summarizes main ingredients and results together with the mesh assumptions required for corresponding proof.

Lemma			Requires assumption				
Thm	Eqn	Formula	(A1)	(A2)	(A3)	(A4)	(A5)
-	(6)	$\lvert C_x^{-1}\underline{n}_E\rvert \sim h_E^{-1}$		x	x	x	
3.2		$\eta_{R,x}^2 \sim \sum_{E:x\in\mathcal{N}_E} \eta_{R,E}^2$	x	x			
3.3	(14), (15)	Residual error estimation	x	x			x
3.4	(16)	$\eta_{R,x} \sim \eta_{\check{R},x}$		x	x	x	
-	(18)	$(\underline{v},\underline{v})_h \sim \sum_{T\in\mathcal{T}_h} \lVert B_T\underline{v}\rVert_T^2$		x	x		
3.6	(20)	$(\nabla^{\mathrm{R}_1} u_h)(x) = \sum_{T\subset\omega_x} \mu_T\,\nabla u_{h\mid T}$					
-	(24)	$\eta_{Z_1}^2 \sim \sum_{x\in\mathcal{N}_{\bar\Omega}} \eta_{Z_1,x}^2$		x	x		
3.8	(25)	Auxiliary lemma for $\eta_{Z_1,x}$	x				
3.9	(26), (27)	$\eta_{R,x} \sim \eta_{Z_1,x}$	x	x	x	x	
-	(32)	$\eta_{Z_2,x}^2 \lesssim \sum_{T\subset\omega_x} \eta_{Z_2,T}^2$	x	x	x		
-	(33)	$\eta_{Z_2,T}^2 \lesssim \sum_{x\in\mathcal{N}_T} \eta_{Z_2,x}^2$	x	x	x		
3.12	(34)	Auxiliary lemma for $\eta_{Z_2,x}$	x				
3.13	(36)–(39)	$\eta_{R,x} \sim \eta_{Z_2,x}$ etc.	x	x	x	x	
3.14	(40), (41)	ZZ error estimation with η_{Z_2}	x	x	x	x	x

5 Numerical experiments

The aims of the numerical experiments are threefold. Firstly we investigate the mesh assumptions. Secondly the main theoretical predictions are to be verified. Lastly the constants that are involved in most inequalities/equivalences are examined numerically, and the asymptotic behaviour is observed.

To this end we present three experiments. The first one features an *isotropic* solution on an *isotropic* mesh, and thus tells what can reasonably be expected. The second experiments exhibits an *anisotropic* solution on tensor product type, rectangular anisotropic mesh. We believe such structured meshes to be best suited for ZZ error estimation. Finally the third experiment involves an anisotropic solution on a more irregular anisotropic mesh (which is unstructured in the xy directions, cf. also section 4.5).

In section 5.1 we present the details of each experiment. Section 5.2 is devoted to the mesh assumptions (A3) and (A4). Finally in section 5.3 the main theoretical results are tested numerically. We restrict ourselves to the second ZZ error estimator η_{Z_2} because it is more general than the first ZZ estimator η_{Z_1}, and since the second estimator allows *local* equivalences/estimates.

The results are given both numerically in tables and graphically as figures. Experiment 1 is represented in all figures by the symbols –□–■–, experiment 2 by –○ ●–, and experiment 3 by - △ - ▲ -.

5.1 Description of the experiments

Experiment 1: Isotropic solution + uniform mesh

This experiment utilizes the most favourite settings; thus one can observe which results reasonably can be expected. Here we solve the Poisson problem

$$-\Delta u = f \quad \text{in } \Omega := (0,1)^3, \qquad u = u_D \quad \text{on } \partial\Omega \quad .$$

The exact *isotropic* solution u is prescribed to be

$$u \;=\; e^{-x} + e^{-y} + e^{-z} \quad ,$$

and the data f, u_D are chosen accordingly. We employ *isotropic*, uniform tetrahedral meshes $\mathcal{T}_l$, $l = 1 \dots 5$, which are the tensor product of three uniform 1D meshes of mesh size $h = 2^{-l}$. The table below displays some interesting information about mesh and solution.

Level l	Elements	$\|\nabla(u-u_h)\|_\Omega$	$\max_{T\in\mathcal{T}_l} h_{1,T}/h_{3,T}$	$m_1(u-u_h,\mathcal{T}_l)$
1	48	$1.61E-1$	2.45	1.71
2	384	$8.16E-2$	2.45	1.71
3	3 072	$4.10E-2$	2.45	1.71
4	24 576	$2.05E-2$	2.45	1.71
5	196 508	$1.03E-2$	2.45	1.71

Experiment 2: Anisotropic solution + structured anisotropic mesh

Here again the Poisson problem with inhomogeneous Dirichlet boundary conditions is solved in $\Omega := (0,1)^3$. The exact *anisotropic* solution u is here prescribed to be

$$u = e^{-x/\varepsilon} + e^{-y/\varepsilon} + e^{-z/\varepsilon}, \qquad \varepsilon := 10^{-2},$$

and thus exhibits sharp boundary layers along the planes $x = 0$, $y = 0$ and $z = 0$. The data f, u_D are chosen accordingly. We employ structured *anisotropic* meshes $\mathcal{T}_l$, $l = 1 \dots 5$, cf. left part of figure 7. These meshes are formed by the tensor product of three 1D Bakhvalov type meshes with transition point at $\tau = 2\varepsilon|\ln\varepsilon|$, see also [Kun01c] for a comprehensive description.

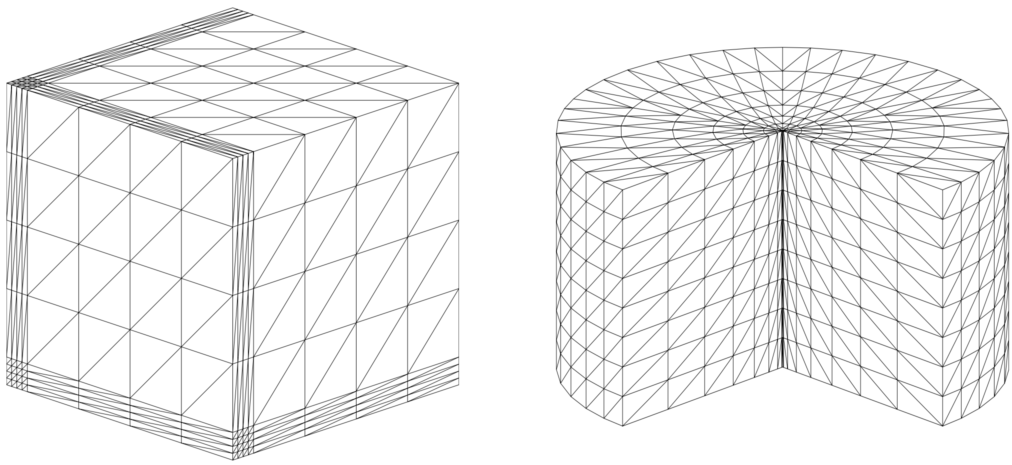

Figure 7: Meshes $\mathcal{T}_3$ of experiment 2 (left) and 3 (right)

In a similar fashion as before we present details of mesh and solution.

Level l	Elements	$\|\nabla(u-u_h)\|_\Omega$	$\max_{T\in\mathcal{T}_l} h_{1,T}/h_{3,T}$	$m_1(u-u_h,\mathcal{T}_l)$
1	48	$9.91E+0$	14.1	1.61
2	384	$8.82E+0$	14.3	1.71
3	3 072	$6.28E+0$	14.4	1.70
4	24 576	$3.67E+0$	14.5	1.67
5	196 508	$1.94E+0$	14.5	1.62

Note first that the problem is comparatively poorly resolved. This is mainly due to the right hand side $f = -\Delta u \equiv \varepsilon^{-1}u$ which has large and steep boundary layers (although still $f \in L_2(\Omega)$). Secondly, the maximum aspect ratio of the anisotropic meshes is about 1:15. These meshes are well suited to the anisotropic solution, as the small matching number $m_1(u-u_h,\mathcal{T}_l) \approx 1.7$ confirms (cf. also exp. 1).

Experiment 3: Anisotropic solution + semi–structured anisotropic mesh

The domain Ω here consists of 3/4 of a cylinder of height and radius 1, cf. the right part of figure 7. The exact *anisotropic* solution u is prescribed to be

$$u(r,\varphi,z) = r^{\lambda}\cdot\sin(\lambda\varphi)\cdot\begin{cases} 1+2z(2z-1) & \text{for } z\in(0,1/2) \\ 1+(3-4z)(2z-1) & \text{for } z\in(1/2,1) \end{cases}, \qquad \lambda = 1/3.$$

This function behaves anisotropically along the concave edge, and is piecewise quadratic in z direction. The data f, u_D are chosen accordingly.

The sequence of meshes $\mathcal{T}_l$, $l = 1\ldots5$, is constructed by first generating an isotropic, uniform mesh in the domain Ω. The subsequent nodal coordinate transformation

$$(x,y,z)^T := (\rho\cdot\hat{x}, \rho\cdot\hat{y}, \hat{z})^T \qquad \text{with } \rho := \{\hat{x}^2+\hat{y}^2\}^{(1-\mu)/2\mu}, \quad \mu = 0.4,$$

yields the final, *anisotropic* mesh, see right part of figure 7. Hence the semi–structured meshes $\mathcal{T}_l$ are the tensor product of an unstructured, graded 2D mesh in the xy plane, and a uniform 1D mesh in the z direction, as in section 4.5.

The details of the meshes and the solution are displayed below. The problem is well resolved, and all anisotropic meshes are well adapted to the solution, i.e. $m_1 < 2$.

Level l	Elements	$\|\nabla(u-u_h)\|_\Omega$	$\max_{T\in\mathcal{T}_l} h_{1,T}/h_{3,T}$	$m_1(u-u_h,\mathcal{T}_l)$
1	96	$1.59E+0$	5.4	1.91
2	768	$8.60E-1$	9.7	1.86
3	6 144	$4.50E-1$	27.3	1.83
4	49 152	$2.33E-1$	77.0	1.83
5	393 216	$1.21E-1$	217.7	1.83

5.2 Mesh Assumptions (A3) and (A4)

Here the mesh assumptions are investigated numerically.

Mesh Assumption (A3)

This assumption can be reformulated as

$$c_1\cdot|C_x^{-1}\underline{v}| \le |C_T^{-1}\underline{v}| \le c_2\cdot|C_x^{-1}\underline{v}| \qquad \forall\,\underline{v}\in\mathbb{R}^d, \forall\,T\subset\omega_x \qquad .$$

In order to investigate this condition numerically we have to specify the matrix C_x for a given node x. In view of theorem 4.4 choose that element $T\subset\omega_x$ that has the smallest aspect ratio $h_{1,T}/h_{3,T}$, and set $C_x := C_T$.

Table 1 gives the corresponding values of c_1, c_2 for all three experiments, and all meshes $\mathcal{T}_l$. Figure 8 presents the same results graphically (note the logarithmic y scale).

Level	Experiment 1 c_1	Experiment 1 c_2	Experiment 2 c_1	Experiment 2 c_2	Experiment 3 c_1	Experiment 3 c_2
1	0.500	1.856	0.082	11.908	0.451	6.169
2	0.500	1.856	0.078	11.730	0.323	7.328
3	0.500	1.856	0.078	11.640	0.379	8.178
4	0.500	1.856	0.078	13.444	0.372	8.336
5	0.500	1.856	0.078	13.419	0.353	8.357

Table 1: Values of c_1, c_2 for assumption (A3); all experiments

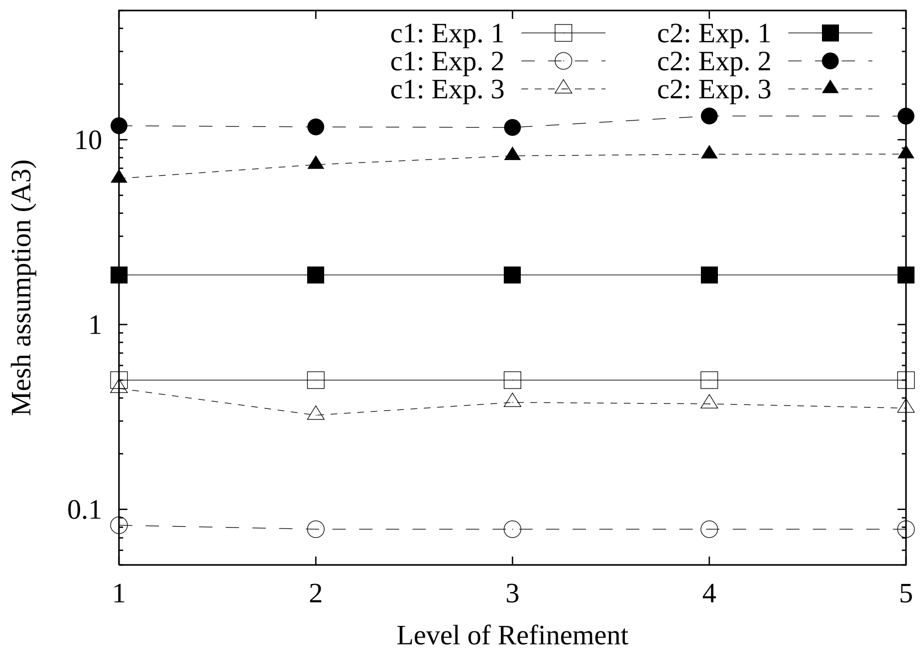

Figure 8: Values of c_1, c_2 for assumption (A3); all experiments

On isotropic meshes (experiment 1) one always has $c_1 \sim c_2 \sim 1$ which is confirmed by the moderate values. For the anisotropic mesh of experiment 2, the theoretical considerations of section 4.1 reveal that (A3) holds as well. The values of c_1, c_2, however, are less favourable than in the isotropic case. This mainly seems to be due to relatively large changes of the element sizes $h_{i,T}$ across neighbouring elements. This observation is strengthened by the results of experiment 3 which features a more steady change of the element sizes, and where the values of c_1, c_2 are more moderate.

Summarizing, a suitably graded mesh will be advantageous for (A3) to hold.

Mesh Assumption (A4)

The assumption (A4) on the shape of the elements can be rewritten as

$$c_3 \le |C_T^{-1}\underline{n}_E| \cdot h_{E,T} \le c_4 \qquad \forall T \in \mathcal{T}_l,\ \forall\ E \subset \partial T \qquad .$$

Utilizing the theory of section 4.4, we can apply theorem 4.7 to all three experiments which yields $c_3 \sim c_4 \sim 1$ (alternatively employ the results of section 4.1 for experiments 1 and 2, as well as the results of section 4.5 for experiment 3). This is verified impressively by the numerical results presented in table 2 and graphically in figure 9. Summarizing, (A4) does not cause problems for well shaped elements.

	Experiment 1		Experiment 2		Experiment 3	
Level	c_3	c_4	c_3	c_4	c_3	c_4
1	0.754	1.202	0.901	1.492	0.760	1.415
2	0.754	1.202	0.754	1.497	0.723	1.564
3	0.754	1.202	0.754	1.500	0.714	1.690
4	0.754	1.202	0.754	1.501	0.714	1.714
5	0.754	1.202	0.754	1.502	0.712	1.717

Table 2: Values of c_3, c_4 for assumption (A4); all experiments

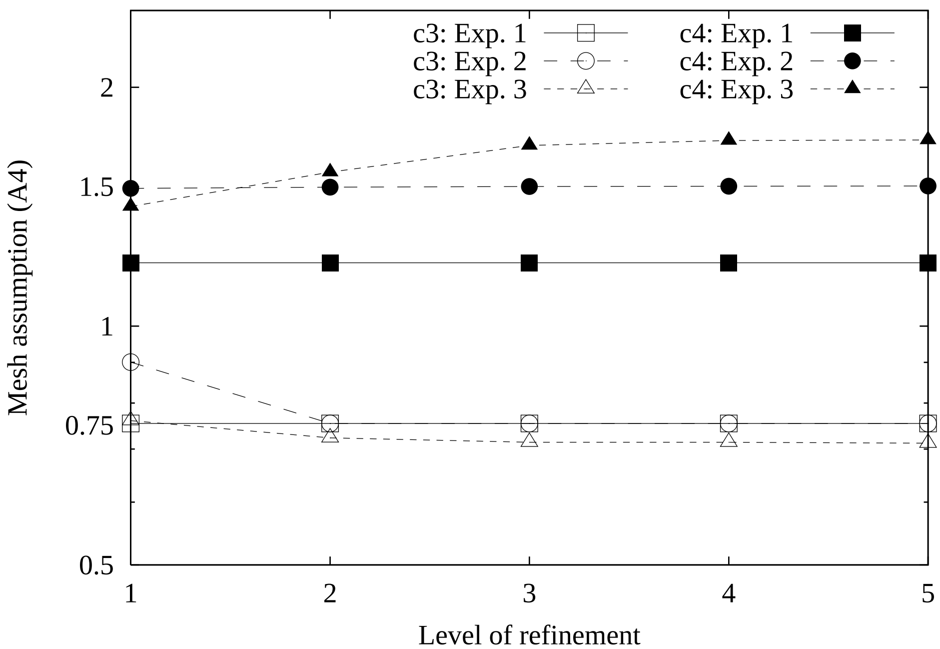

Figure 9: Values of c_3, c_4 for assumption (A4); all examples

5.3 Main numerical results

In this section the main theoretical results for the second ZZ error estimator are tested numerically. First we investigate relations (36), (37) of theorem 3.13 which state a local and global equivalence between the residual error estimator and the ZZ error estimator, respectively. Afterwards the results of the actual ZZ error estimation of theorem 3.14 are presented.

Results for theorem 3.13

The *local* equivalence (36) can be rewritten as

$$c_5 \cdot \eta_{Z_2,x} \leq \eta_{R,x} \leq c_6 \cdot \eta_{Z_2,x} \qquad \forall\, x \in \mathcal{N}_{\bar{\Omega}} \qquad .$$

The values of c_5, c_6 are given in table 3 and graphically in figure 10. One observes that the equivalence between both error estimators diminishes for anisotropic meshes but is still acceptable (note that c_5, c_6 describe only the worst cases over all $x \in \mathcal{N}_{\bar{\Omega}}$). The comparatively large values of c_6 in experiment 2 seem to be caused by (A3), see above,which underlines the importance of that mesh assumption.

	Experiment 1		Experiment 2		Experiment 3	
Level	c_5	c_6	c_5	c_6	c_5	c_6
1	0.855	1.309	0.197	3.420	0.844	4.354
2	0.826	1.309	0.843	14.854	0.848	5.512
3	0.817	1.309	0.562	15.576	0.859	5.607
4	0.815	1.309	0.541	14.546	0.797	5.500
5	0.815	1.309	0.598	13.833	0.725	5.440

Table 3: Values of c_5, c_6 for equivalence (36); all experiments

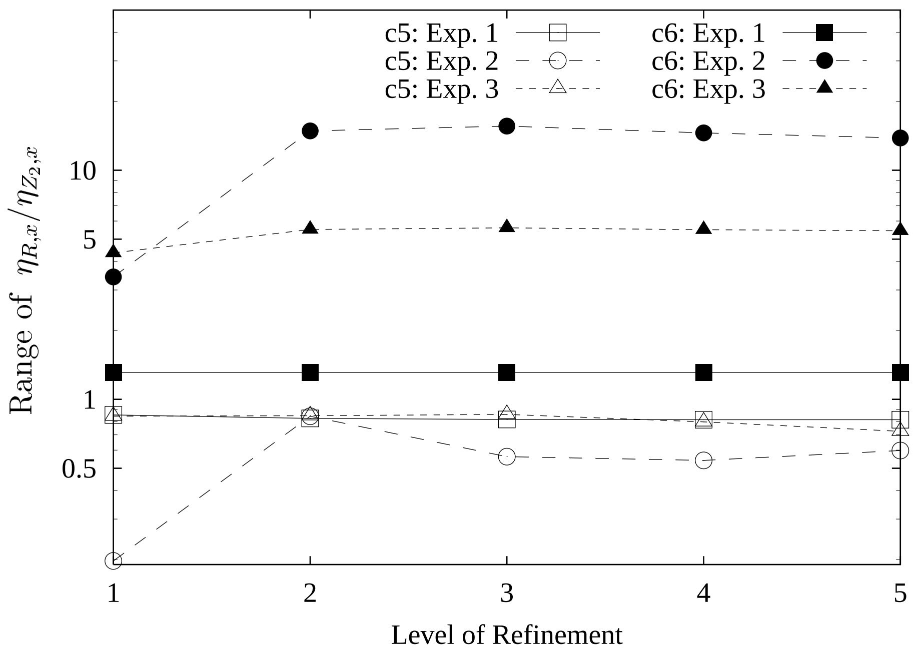

Figure 10: Values of c_5, c_6 for equivalence (36); all experiments

The *global* equivalence (37) between the residual estimator and the ZZ estimator reads

$$\eta_R \sim \eta_{Z_2} \qquad .$$

Thus we present η_R/η_{Z_2} for all meshes and experiments. The results of table 4 and figure 11 confirm the theoretically proven equivalence. Note that the comparatively large values of η_R/η_{Z_2} are mainly due to the different range of the sums, cf. (12) and (30). Furthermore the summand $\eta_{R,x}$ is contains the factor $|\omega_x|$ while $\eta_{Z_2,T}$ is related to $|T|$.

	η_R/η_{Z_2}		
Level	Experiment 1	Experiment 2	Experiment 3
1	25.1	39.9	43.7
2	28.7	35.1	36.7
3	31.2	28.2	39.5
4	32.8	30.5	43.0
5	33.8	33.6	45.8

Table 4: Equivalence (37); all experiments

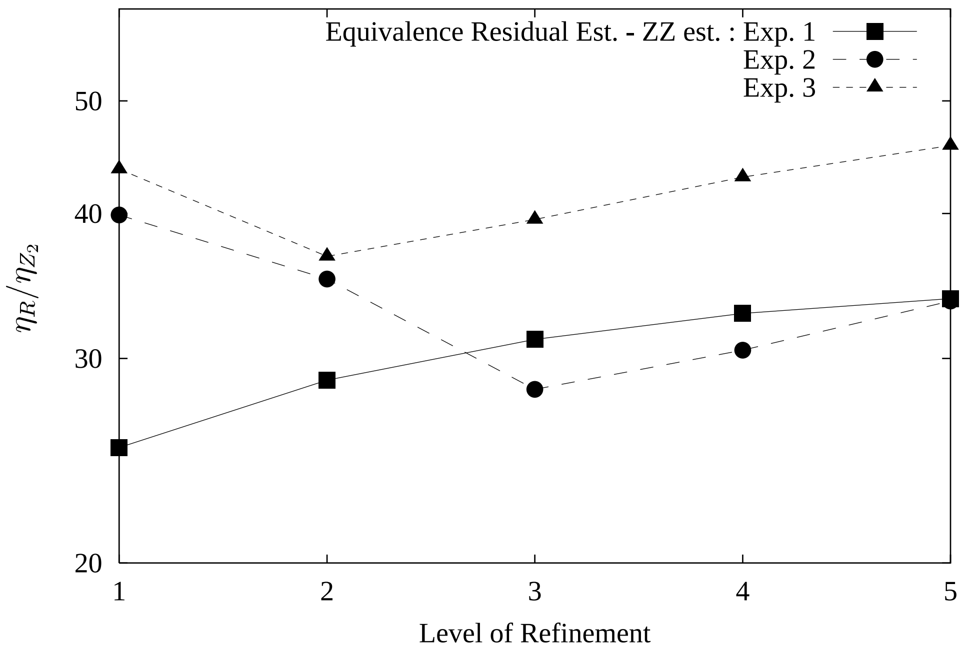

Figure 11: Equivalence (37); all experiments

Results for theorem 3.14

In order to present the results of the ZZ error estimation clearly, let us denote the data approximation term of theorem 3.14 by

$$\zeta_x := h_{min,x} \, \|f - L_h f\|_{\omega_x} \qquad \zeta^2 := \sum_{T \in \mathcal{T}_h} h^2_{min,T} \|f - L_h f\|^2_T \sim \sum_{x \in \mathcal{N}_{\bar{\Omega}}} \zeta_x^2 \quad ,$$

with L_h being the linear Lagrange interpolation operator. Then inequalities (40), (41) of theorem 3.14 can be reformulated as

$$\begin{aligned} \frac{\eta_{Z_2,x}}{\|\nabla(u-u_h)\|_{\omega_x} + \zeta_x} &\lesssim 1 \qquad \forall\, x \in \mathcal{N}_{\bar{\Omega}} \\ \frac{\|\nabla(u-u_h)\|_{\Omega}}{m_1(\eta^2_{Z_2} + \zeta^2)^{1/2}} &\lesssim 1 \qquad , \end{aligned}$$

i.e. both ratios have to be bounded from above. These theoretical predictions are confirmed by the numerical results presented in table 5 and figures 12 and 13. We note that the global ZZ error estimator η_{Z_2} is fairly close to the true error as soon as the solution features are well resolved.

Lastly, the comparatively poor results in experiment 2 on the coarse meshes (level 1–3) are mainly due to the poor resolution of the problem. There the data approximation terms ζ_x and ζ dominate the error terms by far.

	Lower error bound $\max_{x \in \mathcal{N}_{\bar{\Omega}}} \frac{\eta_{Z_2,x}}{\|\nabla(u-u_h)\|_{\omega_x} + \zeta_x}$			Upper error bound $\frac{\|\nabla(u-u_h)\|_{\Omega}}{m_1(\eta^2_{Z_2} + \zeta^2)^{1/2}}$		
Level	Exp. 1	Exp. 2	Exp. 3	Exp. 1	Exp. 2	Exp. 3
1	7.094	0.379	2.990	1.819	0.015	0.110
2	7.968	4.180	6.567	1.935	0.048	0.167
3	8.235	5.462	7.866	1.864	0.180	0.246
4	8.302	10.744	8.265	1.834	0.678	0.356
5	8.319	8.201	8.382	1.823	1.479	0.505

Table 5: Lower and upper ZZ error bounds of theorem 3.14; all experiments

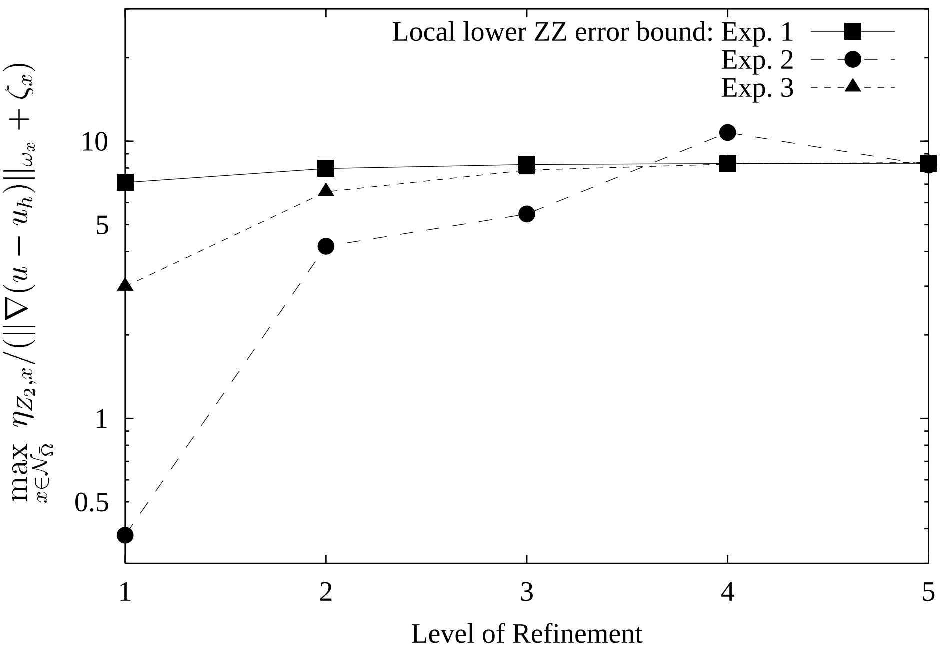

Figure 12: Lower ZZ error bound (40); all experiments

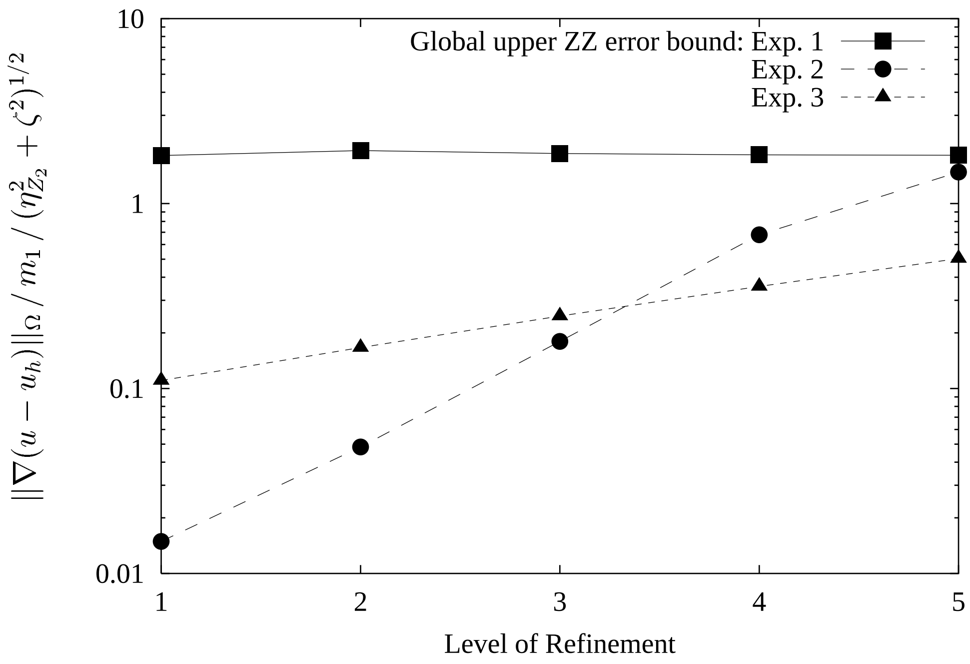

Figure 13: Upper ZZ error bound (41); all experiments

6 Summary

Zienkiewicz–Zhu error estimators are popular because of their cheap implementation and their astonishing robustness. We have proposed and rigorously analysed two kinds of ZZ error estimators that can be applied to *anisotropic* tetrahedral finite element meshes. Both estimators have been defined by scaling the components of the original gradient ∇u_h and some recovered gradient $\nabla^{\mathrm{R}} u_h$.

While our first ZZ estimator is related to a particular choice of the recovered gradient, our second ZZ estimator is much more flexible because *arbitrary weights* can be employed to define the recovered gradient. Hence our novel analysis proves that each averaging technique yields *reliable* and *efficient* error control.

Further emphasis has been given to the requirements on the anisotropic mesh. The analysis has been complemented and confirmed by several numerical examples.

References

[AO00] M. Ainsworth and J.T. Oden. *A posteriori error estimation in finite element analysis*. Wiley, 2000.

[Ape99] Th. Apel. *Anisotropic finite elements: Local estimates and applications*. Advances in Numerical Mathematics. Teubner, Stuttgart, 1999. Habilitationsschrift.

[BC01] S. Bartels and C. Carstensen. Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. Part II: High order FEM. *Math. Comp., to appear*, 2001.

[BPS01] J. H. Bramble, J. E. Pasciak, and O. Steinbach. On the stability of the L_2-projection in $H^1(\omega)$. *Math. Comp.*, 2001. published online.

[BSU94a] I. Babuška, T. Strouboulis, and C. S. Upadhyay. A model study of the quality of a posteriori error estimatorsfor linear elliptic problems. Error estimation in the interior of patchwise uniform grids of triangles. *Comput. Methods Appl. Mech. Eng.*, 114:307–378, 1994.

[BSU+94b] I. Babuška, T. Strouboulis, C. S. Upadhyay, S. K. Gangaraj, and K. Copps. Validation of a posteriori error estimators by numerical approach. *Int. J. Numer. Methods Eng.*, 37(7):1073–1123, 1994.

[Car01] C. Carstensen. Merging the Bramble-Pasciak-Steinbach and the Crouzeix-Thomée criterion for H^1-stability of the L^2-projection onto finite element spaces. *Math. Comp.*, 2001. published online.

[CB01] C. Carstensen and S. Bartels. Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. Part I: Low order conforming, nonconforming, and mixed FEM. *Math. Comp., to appear*, 2001.

[Cia78] P. G. Ciarlet. *The finite element method for elliptic problems*. North-Holland, Amsterdam, 1978.

[DGP99] M. Dobrowolski, S. Gräf, and C. Pflaum. On a posteriori error estimators in the finite element method on anisotropic meshes. *Electronic Transactions Num. Anal.*, 8:36–45, 1999.

[Kun99] G. Kunert. *A posteriori error estimation for anisotropic tetrahedral and triangular finite element meshes*. Logos Verlag, Berlin, 1999. Also PhD thesis, TU Chemnitz, http://archiv.tu-chemnitz.de/pub/1999/0012/index.html.

[Kun00] G. Kunert. An a posteriori residual error estimator for the finite element method on anisotropic tetrahedral meshes. *Numer. Math.*, 86(3):471–490, 2000. DOI 10.1007/s002110000170.

[Kun01a] G. Kunert. A local problem error estimator for anisotropic tetrahedral finite element meshes. *SIAM J. Numer. Anal.*, 39(2):668–689, 2001.

[Kun01b] G. Kunert. A posteriori l_2 error estimation on anisotropic tetrahedral finite element meshes. *IMA J. Numer. Anal.*, 21(2):503–523, 2001.

[Kun01c] G. Kunert. Robust a posteriori error estimation for a singularly perturbed reaction–diffusion equation on anisotropic tetrahedral meshes. *to appear in Adv. Comp. Math.*, 2001.

[KV00] G. Kunert and R. Verfürth. Edge residuals dominate a posteriori error estimates for linear finite element methods on anisotropic triangular and tetrahedral meshes. *Numer. Math.*, 86(2):283–303, 2000. DOI 10.1007/s002110000152.

[OR79] L. A. Oganesyan and L. A. Rukhovets. *Variational-difference methods for the solution of elliptic equations.* Izd. Akad. Nauk Armyanskoi SSR, Jerevan, 1979. In Russian.

[Rau78] G. Raugel. Résolution numérique par une méthode d'éléments finis du problème de Dirichlet pour le Laplacien dans un polygone. *C. R. Acad. Sci. Paris, Sér. A*, 286(18):A791–A794, 1978.

[RL01] H. G. Roos and T. Linss. Gradient recovery for singularly perturbed boundary value problems II: Two-dimensional convection-diffusion. *to appear in Math. Models Meth. Appl. Sci.*, 2001.

[Rod94] R. Rodriguez. Some remarks on the Zienkiewicz-Zhu estimator. *Numer. Meth. PDE*, 10:625–635, 1994.

[Sie96] K. G. Siebert. An a posteriori error estimator for anisotropic refinement. *Numer. Math.*, 73(3):373–398, 1996.

[Ste01] O. Steinbach. On the stability of the L_2-projection in fractional Sobolev spaces. *Numer. Math.*, 88(2):367–379, 2001.

[Ver96] R. Verfürth. *A review of a posteriori error estimation and adaptive mesh–refinement techniques.* Wiley-Teubner, Chichester; Stuttgart, 1996.

[Zha98] Zh. Zhang. Superconvergent finite element method on a Shishkin mesh for convection-diffusion problems. Report 98-006, Texas Tech University, 1998.

[ZZ87] O. C. Zienkiewicz and J. Z. Zhu. A simple error estimator and adaptive procedure for practical engineering analysis. *Internat. J. Numer. Methods Engrg.*, 24:337–357, 1987.

[ZZ92] O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery (SPR) and adaptive finite element refinement. *Comput. Methods Appl. Mech. Eng.*, 101(1-3):207–224, 1992.