Technische Universität Berlin

Bachelorarbeit

Äquilibrierter Fehlerschätzer für elliptische PDE's

Author:
Robert Gruhlke

Erstgutachter:
Dr. Kersten Schmidt
Zweitgutachter:
Dr. Joscha Gedicke

Berlin, 18. Februar 2013

Inhaltsverzeichnis

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und eigenhändig sowie ohne unerlaube fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Berlin, 18.02.2013

_____

1 Einleitung

Das approximative Lösen partieller Differentialgleichungen führt unmittelbar auf die Fragestellung nach der Güte der Approximation. In diesem Zusammenhang unterscheidet man *a-priori* und *a-posteriori* Abschätzungen. Im Kontext der a-posteriori Analyse wurden in den letzten Jahren eine Vielzahl von Verfahren vorgestellt, um den *Approximationsfehler* zu berechnen. Neben *hierarchischen-* und *Gradientglättungs*-Verfahren haben sich hierbei *residuale implizite Fehlerschätzer* auf Grundlage von Äquilibrierungstechniken zunehmend etabliert, da diese meist konstantenfreie Abschätzungen des exakten Fehlers bezüglich der Energienorm ermöglichen. Im Rahmen dieser Arbeit wird der äquilibrierte Fehlerschätzer nach M. Ainsworth und J.T.Oden ([1],[2]) vorgestellt sowie die Implementation dessen innerhalb der `C++`-Biliothek `CONCEPTS`.
Der Aufbau der Arbeit gestaltet sich dabei wie folgt. Zunächst sollen für die Theorie wichtige Grundlagen und Notationen eingeführt werden, auf denen aufbauend, ein Verfahren zur Konstruktion äquilibrierter Flüsse vorgestellt wird. Die zweite Hälfte der Arbeit dient der Beschreibung der Implementation des Fehlerschätzers und die Durchführung numerischer Experimente zur Bewertung der *Qualität* des Fehlerschätzers. Die Notation, Definitionen und Sätze sind dabei zu großen Teilen motiviert durch die Arbeiten in [1], [2],

2 Grundlagen und Notation

Dieser Abschnitt dient der Normierung der Notation innerhalb dieser Arbeit. Zunächst sollen grundlegende Notationen aufgelistet werden. Auf eine detaillierte Definition wird hierbei verzichtet. Auf der Grundlage des *Modellproblems* werden wir die Finite-Elemente-Methode auf regulären Viereckgittern nach ([1], [13],[14]) vorstellen, um so die Notationen der Approximationsräume, der sogenannten *Pull-Back*-Räume einzuführen. Eine wesentliche Rolle spielt hierbei die Gestalt der Basisfunktionen, die in der Herleitung des Fehlerschätzers von Bedeutung sein wird. Die Definitionen und Notationen des darauffolgenden Kapitels sind angelehnt an die Arbeit in [8]. Das Grundlagenkapitel wird durch die Einführung *impliziter Fehlerschätzer* nach [1] abgeschlossen.

Es folgt eine Übersicht der grundlegenden Notation.

$\mathbb{R}, \mathbb{R}^n$:	Menge der reellen Zahlen und deren n-faches kartesisches Produkt
$\mathbb{N}$:	Menge der natürlichen Zahlen
mod:	Modulu
$C^k(\hat{K}, K)$:	Raum der k-mal stetig differenzierbaren Funktionen von $\hat{K}$ nach K
$L^p(\Omega), L^p(K)$:	Lesbeque-Raum für $p \in [1, \infty]$ auf Ω bzw. K
$L^p(\gamma)$:	Lesbeque-Raum auf einer Kante γ
$H^k(\Omega), H^k(K)$:	Sobolev-Raum für $k \in \mathbb{N}$
$H_0^k(\Omega), H_0^k(K)$:	Sobolev-Raum für $k \in \mathbb{N}$ mit Nullspur auf dem Dirichletrand $\Gamma_D \subset \partial\Omega$ bzw. $\Gamma_D \cap \partial K$
$H^1(\Omega) \setminus \mathbb{R}$:	(Quotienten)-Sobolev-Raum ohne konstante Moden, d.h. $u = v + c$, $c \in \mathbb{R} \Rightarrow u = v$ in $H^1(\Omega) \setminus \mathbb{R}$
$\lVert\!\lvert .\rVert\!\rvert, \lVert\!\lvert .\rVert\!\rvert_K$:	Energienorm auf Ω bzw. K
$\lVert . \rVert$:	Norm
$\lVert . \rVert_{L^p(M)}$:	Lebesque-Norm für $M = \Omega, K, \gamma$
$\partial\Omega$, ∂K:	Rand des Gebietes Ω bzw. K
$\mathrm{d}x, \mathrm{d}s$:	Lebesque-Maß auf Gebieten bzw. Randelementen
$\mathbf{n}, \mathbf{n}_K$:	äußerer Normalenvektor bzgl. Ω oder K
$\frac{\partial u}{\partial \mathbf{n}}, \frac{\partial u}{\partial \mathbf{n}_K}$:	Normalenableitung einer Funktion u in Richtung $\mathbf{n}$, bzw $\mathbf{n}_K$
∇:	Gradient
Δ:	Laplace-Operator
conv:	konvexe Hülle
supp:	Träger, z.B. $\operatorname{supp} u$, Träger einer Funktion u
span:	linearer Aufspann
vol.	Volumen bzgl. $\mathrm{d}x$ oder $\mathrm{d}s$

Innerhalb der Arbeit bezeichnen wir mit C eine generische Konstante. Wird eine Funktion $v \in H^1(\Omega)$ auf einem Randstück $\gamma, \partial K$ oder $\partial\Omega$ betrachtet, so verstehen wir dies stets im Sinne einer Spur. Dies wird durch das folgende Lemma motiviert

Lemma 2.1. (**Spursätze**)
Sei K ein Lipschitz-Gebiet, dann existiert eine stetige Abbildung $\gamma_0 \colon H^1(K) \to L^2(\partial K)$, das heißt

$$\exists C = C(K) > 0 : \qquad \|\gamma_0 u\|^2_{L^2(\partial K)} \leq C \|u\|_{L^2(K)} \left(\|u\|_{L^2(K)} + \|\nabla u\|_{L^2(K)} \right), \quad \forall u \in H^1(K).$$

Sei $u \in H^1_\Delta(K) = \{v \in H^1(K) | \Delta v \in L^2(K)\}$, dann ist die Normalenspur von u stetig, das heißt es existiert ein $C > 0$, sodass

$$\left\| \frac{\partial u}{\partial \mathbf{n}_K} \right\|_{H^{-1/2}(\partial K)} \leq C \|u\|_{H^1(K)},$$

wobei $\mathbf{n}_K$ den äußeren Normalenvektor zu K bezeichnet.

Beweis. Beweis der Stetigkeit der Spuroperatoren befindet sich in [6] und [12]. □
Die Stetigkeit der Normalenspur werden wir bei der Anforderungsherleitung der äquilibrierten Flüsse benötigen. Eine weitere Ungleichung wird in der Analysis der Effizienz des Fehlerschätzers verwendet.

Lemma 2.2. Poincaré-Ungleichung Es gilt die Poincaré-Ungleichung:

$$\exists C > 0 : \quad \|u - \overline{u}\|_{L^2(K)} \leq C \|\nabla u\|_{L^2(K)}, \quad \overline{u} = \frac{1}{\mathrm{vol}(K)} \int_K u \mathrm{d}x$$

wobei vol das Volumen bezeichnet.

Beweis. Ein Beweis der Poincaré-Ungleichung findet sich in [7]. □
Desweiteren werden wir die diskrete und die stetige Variante der Cauchy-Schwarz-Ungleichung benötigen.

Lemma 2.3. (**Cauchy-Schwarz-Ungleichung**)
Sei $x, y \in \mathbb{R}^n$ mit $x = (x_1, \dots, x_n)^T, y = (y_1, \dots, y_n)^T$, dann gilt die diskrete Cauchy-Schwarz-Ungleichung

$$\left(\sum_{i=1}^{n} x_i \cdot y_i \right)^2 \leq \left(\sum_{i=1}^{n} x_i^2 \right) \cdot \left(\sum_{i=1}^{n} y_i^2 \right).$$

Sei $f, g \in L^2(\Omega)$, dann gilt die stetige Cauchy-Schwarz-Ungleichung

$$\left| \int_\Omega f g \mathrm{d}x \right|^2 \leq \left(\int_\Omega |f(x)|^2 \mathrm{d}x \right) \left(\int_\Omega |g(x)|^2 \mathrm{d}x \right).$$

Beweis. Ein Beweis dieser Ungleichungen befindet sich zum Beispiel in [19]. □

2.1 Modellproblem und Regularitätstheorie

Zunächst soll das Modellproblem beschrieben werden, auf dessen Grundlage im Rahmen dieser Arbeit die äquilibrierten Flüsse konstruiert werden.
Im Folgenden sei $\Omega \subset \mathbb{R}^2$ ein beschränktes Gebiet mit Lipschitzrand $\Gamma = \partial\Omega$. Auf diesem betrachten wir das elliptische, gemischte Randwertproblem

$$\begin{cases} -\Delta u + cu = f & \text{in} \quad \Omega, \\ \frac{\partial u}{\partial \mathbf{n}} = g & \text{auf} \quad \Gamma_N, \\ u = 0 & \text{auf} \quad \Gamma_D. \end{cases} \tag{1}$$

Hierbei bilden der Neumannrand Γ_N und der Dirichletrand Γ_D eine disjunkte Zerlegung des Randes Γ. Insbesondere ist auch $\Gamma_D = \emptyset$ oder $\Gamma_N = \emptyset$ zugelassen. Weiter sei $f \in L^2(\Omega)$, $g \in L^2(\Gamma_N)$, $c \in \mathbb{R}, c \geq 0$ und $\mathbf{n} \in (L^\infty(\Gamma))^2$, wobei $\mathbf{n}$ den äußeren Normalenvektor bezüglich Ω bezeichne. Die schwache Formulierung dieses Randwertproblems lautet:
Finde $u \in H_0^1(\Omega) = H_{\Gamma_D}^1(\Omega)$, welches die Gleichung

$$B(u, v) = L(v) \quad \forall v \in H_0^1(\Omega) \tag{2}$$

erfüllt. Dabei seien die Billinearform $B\colon H_0^1(\Omega) \times H_0^1(\Omega) \to \mathbb{R}$ sowie die Linearform $L\colon H_0^1(\Omega) \to \mathbb{R}$ definiert durch

$$B(w, v) = \int_\Omega \nabla w \cdot \nabla v \mathrm{d}x + c \int_\Omega wv \mathrm{d}x,$$
$$L(v) = \int_\Omega fv \mathrm{d}x + \int_{\Gamma_N} gv \mathrm{d}s.$$

Aus der Theorie der partiellen Differentialgleichungen (Lax-Milgram, [13]) wissen wir, dass eine eindeutige Lösung dieses Problems für $c > 0$ existiert. Weiter existiert eine eindeutige Lösung, falls $c = 0$ und $\mathrm{vol}(\Gamma_D) > 0$. Für den Fall $c = 0$ und $\Gamma_D = \emptyset$ ist B nicht mehr koerziv auf $H_0^1(\Omega) = H^1(\Omega)$, aber auf $H^1(\Omega) \setminus \mathbb{R}$ und damit (2) nicht mehr eindeutig lösbar. Eine notwendige Bedingung zur Lösbarkeit von reinen Neumannproblemen mit $c = 0$, ist die Kompatibilitätsbedingung der rechten Seiten von (2):

$$L(1) = \int_\Omega f \mathrm{d}x + \int_{\Gamma_N} g \mathrm{d}s = 0.$$

Dies folgt sofort aus der schwachen Formulierung, da $B(v, 1) = 0 \, \forall v \in H^1(\Omega)$ und $1 \in H^1(\Omega)$. Diese sei an dieser Stelle benannt, denn sie wird in der Herleitung der äquilibrierten Flüsse und der Lösbarkeit lokaler Probleme eine wichtige Rolle spielen.

2.2 Finite-Elemente-Methode auf regulären Viereckgittern

Im Allgemeinen lässt sich keine exakte Lösung u unseres Modellproblems bestimmen. Vielmehr wird numerisch eine Näherungslösung u_h bestimmt, die mit einer gewissen Güte die gesuchte Funktion approximiert. In der Praxis hat sich hierbei unter anderem die Methode der Finiten-Elemente, einem Galerkin-Verfahren ([14]), etabliert. Hierbei wird der Ansatz-und Testraum des Problems (2) auf einen endlich dimensionalen abgeschlossenen Unterraum $V_h^p \subset H_0^1(\Omega)$ mit $\dim(V_h^p) = N$ eingeschränkt. Das diskrete Problem lautet dann: Finde $u_h \in V_h^p$, welches die Gleichung

$$B(u_h, v_h) = L(v_h) \quad \forall v_h \in V_h^p \tag{3}$$

erfüllt. Die Wahl der Basisfunktionen $\{\phi_1, \dots, \phi_N\}$ von V_h^p spielt dabei eine entscheidene Rolle. Mit dem Grundprinzip der Lokalität werden diese in der Finite-Elemente-Methode so gewählt, dass ihr Träger möglichst klein ist, und weiter

$$B(\phi_i, \phi_j) = 0, \quad \text{falls} \quad \operatorname{vol}(\operatorname{supp}(\phi_i) \cap \operatorname{supp}(\phi_j)) = 0,$$

gilt. Dies hat einen wesentlichen Vorteil. Sei dazu $\alpha_i \in \mathbb{R}$, $i = 1, \dots, N$ und $u_h = \sum_{i=1}^{N} \alpha_i \phi_i$, dann ist (3) äquivalent zu

$$\sum_{i=1}^{N} \alpha_i B(\phi_i, \phi_j) = L(\phi_j), \quad \forall \phi_j, j = 1, \dots, N.$$

Hierbei handelt es sich um ein lineares Gleichungssystem der Form $A\alpha = b$. Die Systemmatrix A ist aufgrund der Wahl der Basisfunktionen dünn besetzt.
Um nun eine Klasse von Approximationsräumen zu charakterisieren, definieren wir zunächst eine reguläre Quadrangulierung eines Gebietes.

Definition 2.4. (**Partition mit Vierecken**)
Als Quandrangulierung (Partitionierung oder Zerlegung) von Ω in Vierecke $K \subset \Omega$ bezeichnen wir eine Menge von konvexen nicht-leeren offenen Vierecken $\mathcal{P} = \{K\}_{K \subset \Omega}$, für die gilt

(i) $\overline{\Omega} = \bigcup_{K \in \mathcal{P}} \overline{K}$ und

(ii) der nicht-leere Schnitt zweier verschiedener Vierecke $\overline{K}$ und $\overline{K'}$, $K, K' \in \mathcal{P}$ ist entweder eine einzelne gemeinsame Kante oder ein einzelner gemeinsamer Knoten.

Wir betrachten die Knoten n_1, n_2, n_3, n_4 eines Vierecks $K \in \mathcal{P}$. Betrachte die 4 Dreiecke $T_i, i = 1, \dots, 4$, die durch jeweils drei Knoten innerhalb eines Vierecks K (im Sinne der konvexen Hülle)

aufgespannt werden. Sei h_i der Durchmesser von T_i und

$$\rho_i = 2 \cdot \sup\{r \in \mathbb{R} | B(p,r) \subset T_i, p \in \mathbb{R}^2\}.$$

Weiter setzen wir

$$h_K = \max_{i=1,\dots,4} h_i \quad \text{und} \quad \rho_K = \max_{i=1,\dots,4} \rho_i \quad \text{sowie} \quad \kappa_K = \frac{h_K}{\rho_K}.$$

Definition 2.5. (**reguläre Quadrangulierung**)
Eine Sequenz von Quadrangulierungen $(\mathcal{P})_{i\in\mathbb{N}}$ heißt **regulär**, falls $\kappa > 0$ existiert mit

$$\kappa_K \leq \kappa, \quad \forall K \in \mathcal{P}_i, i \in \mathbb{N}.$$

Bemerkung. Sei $\mathcal{P}_{i\in\mathbb{N}}$ eine reguläre Quadrangulierung von Ω und $K \in \mathcal{P}_i, i \in \mathbb{N}$. Wir betrachten die Abbildung $\mathcal{F}_K \in \mathcal{C}(\hat{K}, K)$ mit dem Referenzviereck $\hat{K} = [0,1]^2$. Da die Partition regulär ist, gilt für die Determinante der Jakobi-Matrix $J_{\mathcal{F}_K}$ (vgl. [1])

$$\|\det(J_{F_K})\|_{L^\infty(\hat{K})} \leq C h_K, \tag{4}$$

mit einer von K unabhängigen Konstante. Wir bilden hier das Supremum der Determinante, da diese für beliebige Vierecke im Allgemeinen nicht konstant ist, falls $\mathcal{F}_K$ nicht affin linear ist. Diese Abschätzung werden wir im Kapitel zur Effizienz des Fehlerschätzers benötigen.

Definition 2.6. (**Knoten, Kanten, Patches**)

Wir bezeichnen mit

$$\mathcal{N} := \mathcal{N}(\Omega) := \left\{ n \in \mathbb{R}^2 : \exists K \in \mathcal{P}, \text{ so dass } n \text{ ein Knoten von K ist} \right\}$$

die Menge aller Eckpunkte (**Knoten**) und mit

$$\mathcal{E} := \mathcal{E}(\Omega) := \left\{ \gamma = \text{conv}(n_1, n_2) \subset \mathbb{R}^2 : n_1 \neq n_2 \text{ und } \exists K \in \mathcal{P} \text{ mit } \{n_1, n_2\} \subset \overline{K} \cap \mathcal{N} \right\}$$

die Menge aller **Kanten**. Mit $\mathcal{E}^I$ bezeichnen wir die Menge aller inneren Kanten, mit $\mathcal{E}^N$ die Menge aller Neumannkanten und mit $\mathcal{E}^D$ die Menge aller Dirichletkanten.
Für ein Viereck K definieren wir die Mengen

$$\begin{aligned} \mathcal{N}_K &:= \left\{ n \in \mathbb{R}^2 : \ n \text{ ist Knoten von K} \right\}, \\ \mathcal{E}_K &:= \left\{ \gamma \subset \mathbb{R}^2 : \ \gamma \text{ ist Kante von K} \right\} \end{aligned}$$

und entsprechend $\mathcal{E}_K^I = \mathcal{E}^I \cap \mathcal{E}_K, \mathcal{E}_K^N = \mathcal{E}^N \cap \mathcal{E}_K$ sowie $\mathcal{E}_K^D = \mathcal{E}^D \cap \mathcal{E}_K$.
Weiter sei $\mathcal{N}(\gamma)$ die Menge der Knoten der Kante $\gamma \in \mathcal{E}$. Zu einem Knoten $n \in \mathcal{P}$ definieren wir den sogenannten (Element-)**Patch** $\mathcal{P}_n$ durch

$$\mathcal{P}_n := \{K \in \mathcal{P} | n \in \mathcal{N}_K\}$$

sowie den **Kanten-Patch** $\mathcal{E}_n$ durch

$$\mathcal{E}_n := \{\gamma \in \mathcal{E} | n \in \mathcal{N}(\gamma)\}.$$

Weiter definieren wir zu einem Element $K \in \mathcal{P}$ die Menge aller umliegenden Elemente $\mathcal{P}(K)$ mit

$$\mathcal{P}(K) := \{K' \in \mathcal{P} | \overline{K} \cap \overline{K'} \neq \emptyset\}$$

sowie mit $\mathcal{E}^I(\mathcal{P}(K))$ die Menge der inneren Kanten von $\mathcal{P}(K)$, das heißt

$$\mathcal{E}^I(\mathcal{P}(K)) := \{\gamma \in \mathcal{E}^I \,|\, \gamma = \partial K' \cap \partial K'', K', K'' \in \mathcal{P}(K)\}$$

Definition 2.7. (**finites Referenzelement**)
Als finites Referenzelement (vgl. [2]) verstehen wir ein Triple $(\hat{K}, \hat{P}, \hat{\Sigma})$, bestehend aus einem Gebiet $\hat{K}$, einem Polynomraum $\hat{P}$ und einer unisolventen Menge $\hat{\Sigma}$ von Freiheitsgraden auf $\hat{P}$. Als Referenzviereck wählen wir $\hat{K} = [-1, 1]^2$ und als Polynomraum setzen wir für $p \in \mathbb{N}$

$$\hat{P} := \mathbb{P}_p := \text{span}\{x^i y^j \,|\, 0 \leq i, j \leq p\}.$$

Die Freiheitsgrade werden auf Knoten, Kanten und das Innere von $\hat{K}$ aufgeteilt.

Als mögliche Basis von $\hat{P}$ eignen sich insbesondere Tensorprodukte integrierter Legrendre-Polynome, die aufgrund ihrer Orthogonalitätseigenschaften zusätzlich die Struktur der Systemmatrix beeinflussen können. Schließlich wollen wir einen Finite-Elemente-Raum auf einer regulären Viereckszerlegung von Ω als Approximationsraum vorstellen.

Definition 2.8. (**Finite-Element-Raum**)
Sei $\mathcal{P} = \{K\}$ eine reguläre Viereckszerlegung von Ω. Sei weiter $K \in \mathcal{P}$ und $\mathcal{F}_K \in \mathcal{C}^1(\hat{K}, K)$ invertierbar. Der zum finiten Element (K, P_K, Σ_K) zugehörige *stückweise Pull-Back-Raum* P_K^p ist definiert durch

$$P_K^p := \{v = \hat{v} \circ \mathcal{F}_K^{-1}, \text{ mit } \hat{v} \in \hat{P} = \mathbb{P}_p\}.$$

Der globale Finite-Elemente-Raum V_h^p ist dann gegeben durch

$$V_h^p = \left\{v \in H_0^1(\Omega) \cap \mathcal{C}(\overline{\Omega}) : \forall K \in \mathcal{P} : v \in P_K^p\right\}.$$

Im Gegensatz zu Triangulierungen sind die Pull-Back-Räume bei einer Quadrangulierung im Allgemeinen keine Polynomräume. Das liegt an der Tatsache, dass für ein allgemeines Viereck die Abbildung $\mathcal{F}_K$ nicht mehr affin linear ist. Einen wichtigen Spezialfall der Finite-Elemente-Methode (FEM) erhalten wir für $p = 1$. Bei der sogenannten $P1$-FEM sind die Freiheitsgrade ausschließlich den Knoten der Zerlegung zugeordnet. Der FEM-Raum wird aufgespannt durch nodale Basisfunktionen

Definition 2.9. (**nodale Basisfunktionen**)
Sei $\mathcal{P} = \{K\}$ eine reguläre Viereckszerlegung von Ω und $\mathcal{N}$ die dadurch induzierte Menge von Knoten. Zu einem Knoten $n \in \mathcal{N}$ definieren wir die nodale Basisfunktion $\theta_n \in V_h^1$ durch

$$\theta_n|_K \in P_K^1 \text{ für alle } K \in \mathcal{P} \text{ mit } \theta_n(m) = \delta_{mn} \text{ für alle } m \in \mathcal{N}.$$

Hierbei handelt es sich um die *Lagrange-Basis* von V_h^1. Für den Träger von θ_n gilt

$$\operatorname{supp} \theta_n = \bigcup_{K \in \mathcal{P}_n} K.$$

Bei Finite-Elemente-Methoden hat sich eine typische Struktur der Basisfunktionen etabliert (vgl. [13]). Überlicherweise definiert man die Basisfunktionen auf einem Referenzelement $(\hat{K}, \hat{P}, \hat{\Sigma})$.

Definition 2.10. (**Basisfunktionen auf dem Referenzelement**)
Aufgrund der Aufteilung der Freiheitsgrade des Raumes $\mathbb{P}_p, p \in \mathbb{N}$ auf Knoten, Kanten und das Innere von $\hat{K}$ unterscheidet man drei Arten von Basisfunktionen:

(i) Es gibt 4 *Knotenbasisfunktionen*. Sie entsprechen der nodalen Basis.

(ii) Es gibt je $p-1$ *Kantenbasisfunktionen* pro Kante, die auf allen anderen Kanten von $\hat{K}$ verschwinden.

(iii) Es gibt $(p-1)^2$ *innere Basisfunktionen*, die auf dem Rand von $\hat{K}$ den Wert Null annehmen.

Wir wollen nun solche Basisfunktionen näher beleuchten und nehmen ab sofort an, dass V_h^p für gegebenes $p \in \mathbb{N}$ wie folgt aufgespannt wird:

Beispiel 2.11. (**Definition einer Basis $\mathcal{B}$ von V_h^p**)
Sei $\{\hat{N}_0, \dots, \hat{N}_p\}$ eine Basis des Polynomraumes $\Pi_{\leq p}([-1,1])$, sodass $\hat{N}_0$ und $\hat{N}_1$ affin lineare Funktionen sind mit $\hat{N}_0(-1) = \hat{N}_1(1) = 1$ und $\hat{N}_0(1) = \hat{N}_0(-1) = 0$ und $\hat{N}_i(-1) = \hat{N}_i(1) = 0, i = 2 \dots p$. Sei $\hat{N}_{ij}(x,y) := \hat{N}_i(x)\hat{N}_j(y)$, dann definiert $\hat{\mathcal{B}} := \{\hat{N}_{ij} : 0 \leq i,j \leq p\}$ eine Basis von $\mathbb{P}_p$ auf dem Referenzelement $\hat{K} = [-1,1]^2$ mittels Tensorproduktansatz.
Dabei bildet

$$\begin{aligned} \hat{\mathcal{B}}^{node} &= \{\hat{N}_{ij} | 0 \leq i,j \leq 1\}, \\ \hat{\mathcal{B}}^{edge} &= \{\hat{N}_{ij} | i \in \{0,1\} \dot{\vee} j \in \{0,1\}\}, \\ \hat{\mathcal{B}}^{inner} &= \{\hat{N}_{i,j} | i,j \geq 2\} \end{aligned}$$

die Menge der Knoten-, Kanten- und inneren Referenzbasisfunktionen. Sei K ein Element der Zelegung von Ω und $\mathcal{F}_K \in \mathcal{C}^1(\hat{K}, K)$ die dazugehörige invertierbare Transformation. Sei weiter P_K^p der lokale stückweise Pull-Back-Raum bzgl. dieses Referenzbasisansatzes.

Die Konstruktion einer (globalen) Basis $\mathcal{B}$ von V_h^p geschieht durch stetiges Zusammenfügen der Basisfunktionen der einzelnen P_K^p unter Beachtung der Orientierungen von Knoten-und Kantenbasisfunktionen und Handhabung der Dirichlet-Randstücke (vgl. [13]).
Aufgrund der Konstruktion der Referenzbasisfunktionen bietet es sich an, die globalen Basisfunktionen in ihrer Nummerierung den Knoten, Kanten und Vierecken einer Zerlegung $\mathcal{P}$ des Gebietes Ω zuzuordnen. So etwas lässt sich durch sogenannte T-Matrizen (vgl. [11], [13]) realisieren.

Beispiel 2.12. (Eigenschaften der globalen Basisfunktionen)
Wir wollen an dieser Stelle noch wichtige Eigenschaften der Basisfunktionen aus Beispiel 2.11 bezüglich ihrer Träger zusammentragen, da wir diese bei der Strukturfestlegung der äquilibrierenden Flüsse an späterer Stelle nutzen wollen.

- Die den Knoten zugeordneten Basen sind gerade die Lagrangebasisfunktionen aus Definition 2.9. Sei θ_n^{node} eine dem Knoten $n \in \mathcal{N}$ zugeordnete Basisfunktion. Es ist

$$\operatorname{supp} \theta_n^{\text{node}} = \bigcup_{K \in \mathcal{P}_n} K.$$

  Weiter gilt aufgrund der Charakterisierung der nodalen Basis aus Definition 2.9, dass θ_n^{node} auf allen Kanten ohne Endknoten n konstant Null ist:

$$\forall \gamma \in \mathcal{E}, n \notin \mathcal{N}(\gamma): \quad \theta_n^{\text{node}} \equiv 0.$$

- Sei $\theta_\gamma^{\text{edge}}$ eine der Kante $\gamma \in \mathcal{E}$ zugeordnete Basisfunktion. Es ist

$$\operatorname{supp} \theta_\gamma^{\text{edge}} = \bigcup_{\gamma \in \mathcal{E}_K} K.$$

  Aufgrund der regulären Zerlegung besteht der Träger aus höchstens zwei Elementen K, K' mit $\gamma \in \mathcal{E}_K \cap \mathcal{E}_{K'}$. Durch die Stetigkeit der Basisfunktionen folgt insbesondere, dass

$$\forall \mathcal{E} \setminus \{\gamma\}: \quad \theta_\gamma^{\text{edge}} \equiv 0,$$

  die Funktion also auf allen anderen Kanten verschwindet.

- Sei schließlich θ_K^{inner} eine dem Element $K \in \mathcal{P}$ zugeordnete Basisfunktion, dann gilt

$$\operatorname{supp} \theta_K^{\text{inner}} = K.$$

  Aufgrund der Stetigkeit der Basisfunktionen gilt

$$\theta_K^{\text{inner}} \equiv 0 \quad \text{auf } \partial K.$$

2.3 A-posteriori Fehlerschätzer und Approximationsfehler

2.3.1 A-posteriori Fehlerschätzer

Liegt nun eine Approximation $u_h \in V_h^p$ an die Lösung $u \in H_0^1(\Omega)$ von (2) vor, so stellt sich die Frage nach der Güte der Annäherung und damit nach der Berechnung des Fehlers $e = u - u_h \in H_0^1(\Omega)$ in einer geeigneten Norm $\|.\|$. Da im Allgemeinen die exakte Lösung u nicht bekannt ist, erwarten wir dasselbe auch für e. Die Aufgabe eines Fehlerschätzers ist es, $\|e\|$ möglichst gut anzunähern, um damit ein geeignetes Abbruchkriterium für das numerische Verfahren festlegen zu können. Dies wollen wir nun mathematisch formulieren. Dazu definieren wir vorerst einen a-posteriori Fehlerschätzer.

Definition 2.13. (**a-posteriori Fehlerschätzer**)
Unter einem a-posteriori Fehlerschätzer verstehen wir eine (berechenbare) Größe η, die sich nur aus vorhandenen Informationen, das sind f und g sowie die Näherungslösung u_h, bestimmen lässt.

Ist im Rahmen dieser Arbeit die Rede eines Fehlerschätzers, so ist damit ein a-posteriori Fehlerschätzer gemeint. Um die Qualität eines Fehlerschätzers zu bewerten, definieren wir zwei Eigenschaften.

Definition 2.14. (**Eigenschaften von Fehlerschätzern**)
Ein Fehlerschätzer η heißt zuverlässig, wenn eine Konstante $C_{zuv} > 0$ existiert mit

$$\|e\| \leq C_{zuv}\eta + hot.$$

Ein Fehlerschätzer η heißt effizient, wenn eine Konstante C_{eff} existiert, sodass

$$\eta \leq C_{eff}\|e\| + hot.$$

Dabei sind *hot* Terme höherer Ordnung (*higher order terms*). Sie konvergieren schneller gegen 0 als der eigentliche Fehler.

Ziel ist es nun Fehlerschätzer zu finden, welche zuverlässig und effizient sind. Sie ermöglichen die Konstruktion von Abbruchkriterien. Möchte man beispielsweise bei adaptiven FEM-Verfahren stoppen, wenn der exakte Fehler $\|e\|$ kleiner als eine gewisse Abbruchskonstante $TOL > 0$ ist, so kann man $\|e\|$ durch seine obere Zuverlässigkeitsschranke ersetzen und fordert

$$C_{zuv}\eta + hot \leq TOL.$$

Die Effizienzkonstante der unteren Schranke können wir dabei als Maß interpretieren, in wie fern der Fehler überschätzt wird. Die Konstanten von konstruierten Fehlerschätzern sind in der Regel unbekannt und stellen eine mögliche Fehlerquelle dar. Wird der berechnete Fehler

überschätzt, können Überverfeinerungen zu Ressourcenproblemen in der Berechnung führen. Wir er überschätzt kann ein adaptives Verfahren abbrechen, obwohl sich der tatsächliche Fehler nicht ignorieren lässt. In der Literatur finden sich zahlreiche theoretische und numerische Abschätzungen, welche das Verhalten der Konstanten charakterisieren.
Eine weitere wichtige Anwendung von Fehlerschätzern ist die Steuerung adaptiver numerischer Verfahren. Hierbei werden lokale Fehlerschätzer η_K verwendet, die, analog zum globalen Fall, Aussagen über den lokalen Fehler treffen. Hat man eine Familie von lokalen Fehlerschätzern $\{\eta_K\}_{K\in\mathcal{P}}$ für eine Zerlegung $\mathcal{P} = \{K\}$, so ist es üblich, einen globalen Schätzer η durch

$$\eta^2 = \sum_{K\in\mathcal{P}} \eta_K^2$$

festzulegen. Ist der exakte Fehler e bekannt, dann kann die Qualität eines Fehlerschätzers durch eine weitere Bewertungseinheit charakterisiert werden. Hierfür definieren wir

Definition 2.15. (**Effektivitätsindizes**)
Als globalen Effektiviätsindex ι verstehen wir den Quotienten

$$\iota = \frac{\eta}{\|e\|}.$$

Zu einem Viereck K definieren wir den lokalen Effektivitätsindex ι_K durch

$$\iota_K = \frac{\eta_K}{\|e\|_K}.$$

Ziel ist es, Fehlerschätzer zu konstruieren, dessen Effektivitätsindizes nahe der 1 liegen. Dies führt auf den Begriff asymptotischer Exaktheit.

Definition 2.16. (**asymptotische Exaktheit**)
Ein Fehlerschätzer η heißt asymptotisch exakt, falls er effizient und zuverlässig ist.

Ein asymptotisch exakter Fehlerschätzer verhält sich wie der exakte Fehler. Konvergiert der Fehler gegen Null, so konvergiert der Fehlerschätzer ebenso.

2.3.2 Approximationsfehler in der Energienorm

Wir wollen nun den Approximationsfehler $e = u - u_h$ in einer geeigneten Norm messen und ihn näher charakterisieren. Als Norm $\|.\|$ wählen wir die durch die Billinearform B induzierte Energienorm $|\!|\!|.|\!|\!|$, mit

$$|\!|\!|v|\!|\!|^2 := B(v,v) = \int_\Omega (\nabla v)^2 \mathrm{dx} + \int_\Omega (c^{1/2}v)^2 \mathrm{dx} = \|\nabla v\|^2_{L^2(\Omega)} + \left\|c^{1/2}v\right\|^2_{L^2(\Omega)}, \quad v \in H^1(\Omega).$$

Der folgende Satz fasst wichtige Eigenschaften des Approximationsfehlers e zusammen.

Satz 2.17. Sei $u \in H_0^1(\Omega)$ die exakte Lösung von (2) und $u_h \in V_h^p$ eine Approximation, dann gilt für den Fehler $e = u - u_h \in H_0^1$, dass

(i) $B(e,v) = L(v) - B(u_h, v)\ \forall v \in H_0^1(\Omega)$ und (Residuums-Gleichung)

(ii) $B(e, v_h) = 0$ für $v_h \in V_h^p$. (Galerkin-Orthogonalität)

(iii) Der Fehler e lässt sich in der Energienorm charakterisieren. Es gilt

$$\|\!|e|\!\| = \sup_{0 \neq v \in H_0^1} \frac{|B(e,v)|}{\|\!|v|\!\|}.$$

Beweis.

(i) Sei $v \in H_0^1(\Omega)$, dann gilt aufgrund der Billinearität von B

$$B(e,v) = B(u,v) - B(u_h,v) = L(v) - B(u_h,v).$$

(ii) Wegen (3) und $V_h^p \subset H_0^1(\Omega)$ folgt sofort $B(e, v_h) = 0$ für $v_h \in V_h^p$.

(iii) Betrachte das Residuum $\mathrm{R} \in H_0^1(\Omega)^*$ mit $\mathrm{R}(v) = B(e,v)$. Sei $\|\!|.|\!\|_*$ die Norm des Dualraumes, dann ist

$$\|\!|\mathrm{R}|\!\|_* := \sup_{0 \neq v \in H_0^1} \frac{|B(e,v)|}{\|\!|v|\!\|}.$$

Wir zeigen die Gleichheit durch Abschätzung in beide Richtungen.
Wir zeigen $\|\!|\mathrm{R}|\!\| \leq \|\!|e|\!\|$. Sei dazu $t \in \mathbb{R}$ und $v \in H_0^1(\Omega) \setminus \{0\}$, dann gilt

$$0 < B(e + tv, e + tv) = t^2 B(v,v) + 2tB(e,v) + B(e,e).$$

Das Polynom $t \mapsto B(e + tv, e + tv)$ besitzt damit keine reellen Nullstellen. Das ist gleichbedeutend mit der Tatsache, dass die Diskriminante des Polynoms negativ ist. Das heißt

$$4B(e,v)^2 - 4B(v,v)B(e,e) < 0$$

und damit

$$B(e,v)^2 < \|\!|v|\!\|^2 \, \|\!|e|\!\|^2 \quad \Leftrightarrow \quad \frac{|B(e,v)|}{\|\!|v|\!\|} < \|\!|e|\!\|\,.$$

Da diese Ungleichung für alle $v \in H_0^1(\Omega) \setminus \{0\}$ gilt, folgt beim Übergang zum Supremum $\|\mathrm{R}\|_* \leq \|e\|$.
Umgekehrt gilt für $v = e$

$$\mathrm{R}(e) = B(e,e) = \|e\|^2 \quad \Rightarrow \quad \|\mathrm{R}\|_* \geq \frac{\mathrm{R}(e)}{\|e\|} = \|e\|\,.$$

Also folgt insgesamt $\|\mathrm{R}\|_* = \|e\|$ und damit die Behauptung.

□

Die erste Aussage des Satzes 2.17 werden wir im Kapitel über implizite Fehlerschätzer und der dortigen Zerlegung des Fehlers zu lokalen Fehlern gebrauchen. Die zweite Aussage dient dazu, einen Vorteil einer später eingeführten Äquilibrierungsbedingung aufzuzeigen. Der Beweis zeigt, dass die Aussage in (iii) auch für beliebige Funktionen $v \in H_0^1(\Omega)$ gilt. Sie vererbt sich ebenfalls auf einzelne Vierecke K der Zerlegung $\mathcal{P}$ von Ω für Funktionen mit Träger auf K.

2.4 Implizite Fehlerschätzer

Das Grundprinzip impliziter Fehlerschätzer wird durch die Residuums-Gleichung in Satz 2.17 motiviert: Approximiere den Fehler e auf einem geeigneten Ansatzraum. Die naive Wahl wieder V_h^p als Testraum zu verwenden, ist aufgrund der Galerkin-Orthogonalität nicht geeignet. Die Konsequenz wäre die triviale Nulllösung. Folglich muss der Ansatzraum größer gewählt werden. Um einen geringeren Aufwand, als die erneute globale Berechnung einer Approximation, zu erreichen, zerlegt man das globale Problem zu einer Folge von lokalen und damit kleineren Randwertproblemen, dessen Lösung den Fehler lokal beschreiben soll. Als lokale Fehlerschätzer dienen dann die Normen der lokalen Fehlerapproximationen. Der globale Fehlerschätzer wird anschließend durch Summation jener lokalen Fehlerschätzer gewonnen.
Als preiswertes Verfahren hat sich hier die Element-Residual-Methode in ihren verschiedenen Varianten in der Praxis durchgesetzt. Hierbei werden lokale Randwertprobleme auf einem einzelnen Element $K \in \mathcal{P}$ gelöst. Aufgrund der Regularität der Approximation in jedem Element genügt der Fehler der partiellen Differentialgleichung

$$-\Delta e + ce = f + \Delta u_h - cu_h \quad \text{in } K.$$

Es stellt sich nun die Frage nach der Wahl zugehöriger Randbedingungen. Man unterscheidet hierbei drei Fälle:

- Wurde die Approximation so konstruiert, dass $u_h = 0$ auf dem Dirichletrand gilt, so wählt man essentielle Randbedingungen

$$e = 0 \quad \text{auf } \partial K \cap \Gamma_D.$$

- Auf einem Neumannrandstück wählt man natürliche Randbedingungen.

$$\frac{\partial e}{\partial \mathbf{n}_K} = g - \frac{\partial u_h}{\partial \mathbf{n}_K} \quad \text{auf } \delta K \cap \Gamma_N.$$

- Zuletzt bleibt der Fall, das eine Randkante γ von K eine innere Kante des Gitters ist. Die Wahl der (Neumann-)Randbedingungen unterscheidet dann die Art des impliziten Element-Residualen Fehlerschätzers.

Wir wollen an dieser Stelle zwei Möglichkeiten einer Wahl von Randbedingungen auf inneren Kanten kennenlernen und betrachten dazu die Gleichung

$$\frac{\partial e}{\partial \mathbf{n}_K} = \frac{\partial u}{\partial \mathbf{n}_K} - \left.\frac{\partial u_h}{\partial \mathbf{n}_K}\right|_K \quad \text{auf } \gamma,$$

Der exakte Fluss $\frac{\partial u}{\partial \mathbf{n}_K}$ der rechten Seite ist im Allgemeinen unbekannt und daher sucht man eine geeignete Näherung. Nach der klassischen Element-Residual-Methode nach Bank und Weiser [5] approximiert man den exakten Fluss durch eine Mittellungsstrategie. Seien $K, K' \in \mathcal{P}$ zwei Elemente mit $\gamma = K \cap K'$, dann mittelt man die Normalenableitungen der Approximierten bezüglich K und K':

$$\frac{\partial u}{\partial \mathbf{n}_K} \approx \frac{1}{2}\mathbf{n}_K \left(\nabla u_h|_K + \nabla u_h|_{K'}\right).$$

Dieser Ansatz liefert unter Umständen einen effizienten und verlässlichen impliziten Fehlerschätzer. Allerdings ist diese Aussage differenziert zu betrachten. Hierbei ist die Wahl des lokalen Ansatzraumes (durch die Wahl der Basisfunktionen) von großer Wichtigkeit. Wie eine detailliertere Analyse in [1] zeigt, beeinflussen diese maßgeblich die Qualität des Fehlerschätzers. Bei der Wahl eines falsches Ansatzraumes kann es mitunter vorkommen, dass der approximierte Fehler weit überschätzt wird oder aber auch gleich 0 ist (vgl. *Legendre-Basis*). In solchen Fällen ist ein Fehlerschätzer dieser Art ungeeignet.
Ein weiterer Nachteil tritt für den Spezialfall $c = 0$ auf. Hier kommt es vor, dass die lokalen (Neumann-)Probleme im Allgemeinen nicht mehr auf $H^1_0(K) = H^1(K)$ lösbar sind.
Der zweite Ansatz versucht genau diese letzte Schwierigkeit zu umgehen und wählt die Neumann-Randbedingungen so, dass die lokalen Probleme wohlgestellt sind. Dies geschieht durch sogenannte Äquilibrierungstechniken. Eine solche wird im Kapitel 3 detailliert vorgestellt.

2.4.1 Lokaler Fehler

Hat man auf einem Element $K \in \mathcal{P}$ eine Annäherung g_K an den exakten Fluss $\frac{\partial u}{\partial \mathbf{n_K}}\Big|_K$ auf $\partial K \setminus \Gamma_D$ gewählt, so ist man an der Lösung des lokalen (Fehlerapproximation-)Problems

$$\begin{cases} -\Delta \tilde{e}_K + c\tilde{e}_K = f + \Delta u_h - cu_h & \text{in} \quad K, \\ \frac{\partial \tilde{e}_K}{\partial \mathbf{n}_K} = g_K - \frac{\partial u_h}{\partial \mathbf{n}_K}\Big|_K & \text{auf} \quad \partial K \setminus \Gamma_D, \\ \tilde{e_K} = 0 & \text{auf} \quad \Gamma_D \cap \partial K, \end{cases} \tag{5}$$

interessiert. Da g_K nur eine Approximation darstellt, ist $\tilde{e}_K$ ebenfalls nur eine mögliche Approximation an den exakten Fehler $e|_K$. Die Notation der Tilde soll dies verdeutlichen. Wir wollen eine schwache Formulierung von (5) herleiten. Sei dafür $v \in H_0^1(K)$. Mithilfe der partiellen Integration erhalten wir die Gleichungen

$$\int_K -\Delta \tilde{e}_K v + c\tilde{e}_K v \mathrm{d}x = \int_K \nabla \tilde{e}_K \cdot \nabla v + c\tilde{e}_K v \mathrm{d}x - \int_{\partial K} \frac{\partial \tilde{e}_K}{\partial \mathbf{n}_K} v \mathrm{d}s$$

und

$$\int_K \Delta u_h v - cu_h v \mathrm{d}x = -\int_K \nabla u_h \cdot \nabla v + cu_h v \mathrm{d}x + \int_{\partial K} \frac{\partial u_h}{\partial \mathbf{n}_K}\Big|_K \mathrm{d}s.$$

Wir definieren die lokale Billinearform $B_K \colon H_0^1(K) \times H_0^1(K) \to \mathbb{R}$ durch

$$B_K(v, w) = \int_K \nabla v \cdot \nabla w \mathrm{d}x + c \int_K vw \mathrm{d}x \quad v, w \in H_0^1(K)$$

sowie das lokale (innere) Residuum $R_K \colon H_0^1(K) \to \mathbb{R}$ durch

$$R_K(v) = \int_K fv \mathrm{d}x - B_K(u_h, v), \quad v \in H_0^1(K).$$

Diese Definitionen nutzen wir und erhalten die folgende Äquivalenz für $v \in H_0^1(K)$

$$\begin{aligned} \int_K -\Delta \tilde{e}_K v + c\tilde{e}_K v \mathrm{d}x &= \int_K fv \mathrm{d}x + \int_K \Delta u_h v - cu_h v \mathrm{d}x \\ \Leftrightarrow \qquad B_K(\tilde{e}_K, v) &= R_K(v) + \int_{\partial K} \frac{\partial \tilde{e}_K}{\partial \mathbf{n}_K} v \mathrm{d}s + \int_{\partial K} \frac{\partial u_h}{\partial \mathbf{n}_K}\Big|_K \mathrm{d}s. \end{aligned}$$

Wegen der Beziehung $\frac{\partial \tilde{e}_K}{\partial \mathbf{n_K}} = g_K - \frac{\partial u_h}{\partial \mathbf{n_K}}\Big|_K$ folgt schließlich die schwache Formulierung von (5):

$$\text{Finde } \tilde{e}_K \in H_0^1(K): \quad B_K(\tilde{e}_K, v) = R_K(v) + \int_{\partial K} g_K v \mathrm{d}s, \quad \forall v \in H_0^1(K). \tag{6}$$

Wir wollen nun eine spezielle Äquibrilierungstechnik kennenlernen und mit ihrer Hilfe einen impliziten Fehlerschätzer konstruieren.

3 Äquilibrierter residualer a-posteriori Fehlerschätzer

Die Wahl der Flussapproximationen $\{g_K\}_{K\in\mathcal{P}}$ mit

$$g_K \approx \left.\frac{\partial u}{\partial \mathbf{n}_K}\right|_K \quad \text{auf } \partial K$$

ist nun von zentralem Interesse. Mit ihnen können wir die lokalen Probleme definieren, um den exakten Fehler möglichst genau zu approximieren. Die Notation und Herleitung sowie die Beweise des Abschnittes sind an die Arbeit in [1] und [2] angelehnt. Das im vorherigem Kapitel vorgestellte innere (lokale) Residuum soll zunächst in einer Definition festgehalten werden.

Definition 3.1. (**Inneres Residuum**)
Als inneres Residuum definieren wir die Abbildung $R \in H_0^1(\Omega)^*$ durch

$$R(v) = B(e, v), \quad v \in H_0^1(\Omega).$$

Sei $K \in \mathcal{P}$. Dann definieren wir analog das innere lokale Residuum $R_K \in H_0^1(K)^*$ durch

$$R_K(v) = \int_K f v \mathrm{d}x - B_K(u_h, v), \quad v \in H_0^1(K)$$

Dabei ist der lokale Sobolev-Raum $H_0^1(K)$ definiert als

$$H_0^1(K) = \{v \in H^1(K) \,|\, v = 0 \text{ auf } \partial K \cap \Gamma_D\}.$$

3.1 Anforderungen an g_K

Wir wollen zunächst Bedingungen an die Approximationen stellen, in dem wir Eigenschaften des exakten Flusses nachempfinden.

- Da $f \in L^2(\Omega)$ und $u \in H_0^1(\Omega)$ ist $\Delta u \in L^2(\Omega)$, und damit $u \in H_\Delta^1(\Omega)$. Mit dem Lemma 2.1 folgt die Stetigkeit der Neumannspur, es gilt auf inneren Kanten für $K, K' \in \mathcal{P}$:

$$\left.\frac{\partial u}{\partial \mathbf{n_K}}\right|_K + \left.\frac{\partial u}{\partial \mathbf{n_{K'}}}\right|_{K'} = 0 \quad \text{auf } \partial K \cap \partial K'.$$

- Auf einem Neumann-Randstück ist der exakte Fluss bekannt, es gilt für $K \in \mathcal{P}$ mit $\partial K \cap \Gamma_N \neq \emptyset$

$$\left.\frac{\partial u}{\partial \mathbf{n_K}}\right|_K = g \quad \text{auf } \partial K \cap \Gamma_N$$

Deshalb fordern wir für die approximierten Flüsse $\{g_K\}_{K\in\mathcal{P}}$

$$\begin{aligned} g_K + g_K' &= 0 \qquad \text{auf } \partial K \cap \partial K', \\ g_K &= g \qquad \text{auf } \partial K \cap \Gamma_N. \end{aligned} \tag{7}$$

Aus den beiden Bedingungen (7) folgt für $v \in H_0^1(\Omega)$

$$\begin{aligned} \sum_{K\in\mathcal{P}} \int_{\partial K} g_K v \mathrm{d}s &= \sum_{K\in\mathcal{P}} \left(\sum_{\gamma\in\mathcal{E}_K^I} \int_\gamma g_K v \mathrm{d}s + \sum_{\gamma\in\mathcal{E}_K^D} \int_\gamma g_K v \mathrm{d}s + \sum_{\gamma\in\mathcal{E}_K^N} \int_\gamma g_K v \mathrm{d}s \right) \\ &= \sum_{\substack{K,K' \\ K\neq K'}} \sum_{\gamma\in\mathcal{E}_K^I\cap\mathcal{E}_{K'}^I} \left(\int_\gamma g_K v \mathrm{d}s + \int_\gamma g_{K'} v \mathrm{d}s \right) + \sum_{K\in\mathcal{P}} \sum_{\gamma\in\mathcal{E}_K^N} \int_\gamma g_K v \mathrm{d}s \\ &= \sum_{K\in\mathcal{P}} \sum_{\gamma\in\mathcal{E}_K^N} \int_\gamma g_K v \mathrm{d}s = \int_{\Gamma_N} g v \mathrm{d}s. \end{aligned}$$

Diese Tatsache motiviert und bestätigt den Ansatz des lokalen Fehlers in (5) bzw. (6), denn der globale exakte Fehler lässt sich zerlegen. Sei dazu $v \in H_0^1(\Omega)$, dann gilt

$$\begin{aligned} B(e,v) &= L(v) - B(u_h, v) + \int_{\Gamma_N} g v \mathrm{d}s = \sum_{K\in\mathcal{P}} R_K(v) + \sum_{K\in\mathcal{P}} \int_{\partial K} g_K v \mathrm{d}s \\ &= \sum_{K\in\mathcal{P}} \big(R_K(v) + \int_{\partial K} g_K v \mathrm{d}s \big). \end{aligned}$$

Hierbei ist $R_K(v) = B_K(e,v)$ das innere Residuum (vgl. Beweis Satz 2.17, *iii*). Die einzelnen Summanden lassen sich durch den in (6) gewählten Ansatz der schwachen Formulierung lokaler Fehlerprobleme lösen.

Bemerkung. Die schwache Formulierung in (6) wählt den lokalen Ansatz- und Testraum $H_0^1(K)$ und ist in diesem Sinne einer Obermenge stärker als hier gesehen mit $v \in H_0^1(\Omega)$. Das liegt an der Tatsache, dass der gebrochene Sobolevraum (engl. *broken Sobolev space*)

$$\hat{H}_0^1(\Omega) = \big\{ v \in L^2(\Omega) | \, \forall K \in \mathcal{P} : v|_K \in H_0^1(K) \big\}$$

eine Obermenge von $H_0^1(\Omega)$ bildet.

Die bis hierhin gestellten Bedingungen an die approximierten Flüsse werden nun um eine weitere wichtige Komponente ergänzt.
Wir betrachten dafür ein $K \in \mathcal{P}$ mit $\partial K \cap \Gamma_D = \emptyset$. Das lokale Problem auf K ist nun ein reines Neumannproblem auf $H_0^1(K) = H^1(K)$, dessen Lösbarkeit gesichert werden muss. Es sei

an dieser Stelle angemerkt, dass der lokale Lösungsraum konstante Funktionen enthält. Wir fordern die sogenannte *Äquilibrierungsbedingung* auf K:

$$R_K(1) + \int_{\partial K} g_K \mathrm{ds} = 0. \tag{8}$$

Sie wird durch zwei entscheidende Argumente motiviert:

- An die Kompatibilitätsbedingung angelehnt, stellt sie für den Fall $c = 0$ auf K eine äquivalente Bedingung zur Lösbarkeit des Problems dar. Man beachte, dass die Lösung nicht eindeutig auf $H_1(K)$ (jedoch auf $H_1(K) \setminus \mathbb{R}$) ist.

- Für $c \neq 0$ ist das lokale Problem zwar eindeutig lösbar auf $H^1(K)$, doch auch hier bietet sich ein positiver Nebeneffekt. Wegen Satz 2.17 gilt

$$\|\tilde{e}_K\|_K = \sup_{v \in H_0^1(K)\setminus\{0\}} \frac{|B_K(\tilde{e}_K, v)|}{\|v\|_K}.$$

  Wählt man nun $v = 1$, dann folgt wegen (4)

$$\|v\|_K^2 = \|1\|_K^2 = \int_K c \mathrm{d}x \approx C h_K^2.$$

  Daher gilt

$$\|\tilde{e}_K\|_K^2 \geq \frac{|B_K(\tilde{e}_K, 1)|}{\|1\|_K} = \frac{|R_K(1) + \int_{\partial K} g_K \mathrm{ds}|}{\|1\|_K}.$$

  Das Problem: Wäre die Äquilibrierungsbedingung nicht erfüllt, wäre der Zähler der rechten Seite positiv. Dahingegen ist der Nenner in der Größenordnung von Ch_K. Läuft nun bei $h_K \to 0$ der Zähler mit einer langsameren Ordnung gegen Null als der Nenner, so besteht die Gefahr, dass bei zunehmender h-Verfeinerung der lokale Fehler über alle Grenzen steigt. Die Äquilibrierungsbedingung stellt damit für den Fall $c > 0$ eine Qualitätssicherung der Fehlermessung dar.

Wir wollen die zusammengetragenen Bedingungen an die approximierenden Flüsse in einer Definition etwas spezieller zusammentragen.

Definition 3.2. (**Äquilibrierungseigenschaft nullten Grades**)
Als Äquilibrierungseigenschaft nullten Grades verstehen wir die Tatsache einer Familie approximierender Flüsses $\{g_K\}_{K\in\mathcal{P}}$ auf Elementen $K, K' \in \mathcal{P}$ mit $K \neq K'$ die folgenden Gleichungen zu erfüllen:

$$\begin{aligned} R_K(1) + \int_{\partial K} g_K \mathrm{d}s &= 0, \\ g_K + g_{K'} &= 0 \qquad \text{auf } \partial K \cap \partial K', \\ g_K &= g \qquad \text{auf } \partial K \cap \Gamma_N. \end{aligned}$$

Wir stellen fest, dass die Äquilibrierungsbedingung nun auch auf Vierecken gelten soll, die ein Dirichletrandstück haben und sehen, dass eine solche Familie approximierender Flüsse implizit voneinander abhängt.

3.2 (Nicht-)Eindeutigkeit der approximierenden Flüsse

Die Äquilibrierungseigenschaft nullten Grades (3.2) legt die approximierenden Flüsse nicht eindeutig fest. Dies wird am folgenden Beispiel deutlich.

Beispiel 3.3. Wir betrachten das gemischte Randwertproblem

$$\begin{cases} -\Delta u = 0 \quad \text{in} \quad \Omega = (0,1)^2, \\ \frac{\partial u}{\partial \mathbf{n}} = 0 \quad \text{auf} \quad \Gamma_N, \\ u = 0 \quad \text{auf} \quad \Gamma_D. \end{cases}$$

mit

$$\begin{aligned} \Gamma_N &= (\{0\} \times (0,1)) \cup (\{1\} \times (0,1)), \\ \Gamma_D &= ([0,1] \times \{0\}) \cup ([0,1] \times \{1\}) \end{aligned}$$

und exakter Lösung $u = 0$. Sei nun $K = \Omega$ und ein $g_K \colon K \to \mathbb{R}$, welches der Äquilibrierungseigenschaft nullten Grades genügt, gegeben. Wir setzen nun mit $h(x,y) = \cos(2\pi x)\sin(2\pi y)$

$$g_K^h = g_K + \frac{\partial h}{\partial \mathbf{n_K}}$$

dann besitzt g_K^h wegen $\frac{\partial h}{\partial \mathbf{n}_K} = 0$ auf Γ_N und

$$\int_{\Gamma_D} g_K^h \mathrm{d}s = \int_{\Gamma_D} \frac{\partial h}{\partial \mathbf{n}_K} \mathrm{d}s = 0$$

ebenfalls jene Eigenschaft. Das Problem liegt hierbei an der Freiheit an Konstruktionsmöglichkeiten auf dem Dirichletrand.

3.3 Struktur der approximierenden Flüsse

Bevor wir auf die Frage nach der Existenz eingehen, wollen wir zunächst die Beschaffenheit einer Familie approximierender Flüsse $\{g_K\}_{K\in\mathcal{P}}$ festlegen. Wir nehmen an, dass die exakte Lösung u durch ein $u_h \in V_h^p$ wie in Beispiel 2.11, für $p \in \mathbb{N}$ angenähert wurde. Sei nun $K \in \mathcal{P}$ und weiter $\gamma \in \mathcal{E}_K$. Wir betrachten den lokalen Pull-Back-Raum P_K^p mit Basis

$$\mathcal{B}_K := \{\theta_i|_K \,:\, \theta_i \in \mathcal{B}, K \subset \operatorname{supp}\theta_i\}. \tag{9}$$

Hierbei ist $\mathcal{B}$ die (globale) Basis von V_h^p aus Beispiel 2.11. Wir unterscheiden

- Die Kante γ ist eine Neumannkante, dann fordern wir im Sinne der Äquilibrierungseigenschaft, dass $g_K|_\gamma = g$.
- Die Kante γ ist keine Neumannkante. Dann wählen wir $g_K|_\gamma$ als Linearkombination der lokalen Basisfunktionen im Sinne von

  $$g_K|_\gamma \in \operatorname{span}\{\theta_i|_\gamma \,:\, \theta_i \in \mathcal{B}_K\}.$$

  Wir nutzen nun die Eigenschaften der Basisfunktionen bezüglich ihrer Funktionswerte aus Beispiel 2.12 aus. Es können nur Basisfunktionen, die einen Knoten $n \in \mathcal{N}(\gamma)$ oder der Kante γ selbst zugeordnet sind, Einfluss auf $g_K|_\gamma$ haben. Diese seien durch

  $$\mathcal{B}_\gamma^{\text{node}} = \{\theta_{n_l}, \theta_{n_r}\} \quad \text{und} \quad \mathcal{B}_\gamma^{\text{edge}} = \{\theta_{\gamma_1}, \ldots, \theta_{\gamma_{p-1}}\} \tag{10}$$

  gegeben und die Linearkombination vereinfacht sich auf jene Funktionen.

Die beiden Strukturansätze sollen an dieser Stelle zusammengefasst werden:

$$g_K|_\gamma \begin{cases} = & g, & \text{falls } \gamma \in \mathcal{E}^N, \\ \in & \operatorname{span}\{\theta_i|_\gamma : \theta_i \in \mathcal{B}_\gamma^{\text{edge}} \cup \mathcal{B}_\gamma^{\text{node}}\}, & \text{falls } \gamma \in \mathcal{E} \setminus \mathcal{E}^N. \end{cases} \tag{11}$$

Die Approximation wird also auf einer Kante $\gamma \in \mathcal{E} \setminus \mathcal{E}^N$ als Linearkombination von Pull-Back-Funktionen der Ordnung p festgelegt.

Bemerkung. Handelt es sich bei K um ein Parallelogramm, so ist $g_K|_\gamma, \gamma \in \mathcal{E}\setminus\mathcal{E}^N$ ein Polynom vom Grad p.

Zu einer FE-Lösung u_h p-ter Ordnung wollen wir nun noch eine weitere Einschränkung an die Flüsse einführen, indem wir die Äquilibrierungsbedingung nullten Grades verschärfen.

Definition 3.4. (**Äquilibrierungseigenschaft p-ten Grades**)
Sei $p \in \mathbb{N}$ und zu $K \in \mathcal{P}$ sei $\mathcal{B}_{\partial K}$ die Menge aller Basisfunktionen von V_h^p mit Einfluss auf dem Rand von K, das heißt

$$\mathcal{B}_{\partial K} = \bigcup_{\gamma \in \mathcal{E}_K} \mathcal{B}_\gamma^{\text{edge}} \cup \mathcal{B}_\gamma^{\text{node}}.$$

Eine Familie approximierender Flüsse $\{g_K\}_{K\in\mathcal{P}}$ auf Elementen $K, K' \in \mathcal{P}$ mit $K \neq K'$ besitzt die Äquilibrierungseigenschaft p-ten Grades, falls gilt

$$\left.\begin{aligned} R_K(\theta_i) + \int_{\partial K} g_K \theta_i \mathrm{ds} &= 0, \qquad \theta_i \in \mathcal{B}_{\partial K} \\ g_K + g_{K'} &= 0 \qquad \text{auf } \partial K \cap \partial K', \\ g_K &= g \qquad \text{auf } \partial K \cap \Gamma_N. \end{aligned}\right\} \tag{12}$$

Wir bemerken, dass für eine innere Basisfunktionen θ zu $K \in \mathcal{P}$ aufgrund der Galerkin-Orthogonalität und $\theta|_{\partial K} \equiv 0$, die erste Bedingung ebenfalls erfüllt ist:

$$R_K(\theta) + \int_{\partial K} g_K \theta \mathrm{ds} = R_K(\theta) + 0 = B(e, \theta) = 0.$$

Da solche Funktionen keinen Einfluss in der Struktur von g_K haben werden, sind diese in der Definition ausgenommen.
Es stellt sich nun die berechtige Frage:

> *Ist es möglich, eine Familie $\{g_K\}_{K\in\mathcal{P}}$, welche der Äquilibrierungseigenschaft p-ter Ordnung genügt, aus der FEM-Lösung u_h und den bekannten f, g zu konstruieren, ohne dabei ein erneutes globales und damit teures Problem zu lösen?*

Wie wir gesehen haben, hängen die Flüsse implizit voneinander ab. Man kann daher zumindest nicht erwarten, die Probleme lokal auf einem einzelnen Element lösen zu können. Auf die Frage der Existenz einer solchen Familie soll im Folgenden eingegangen werden.

3.4 Existenz von $\{g_K\}$ durch Momentenberechnung

Ziel dieses Kapitels ist es, auf konstruktive Weise eine Familie approximierender Flüsse nach dem Vorgehen in [1] herzuleiten. Dazu werden wir Bedingungen in den Äquilibrierungseigenschaften durch die Einführung sogenannter *Momente* umformulieren. Es wird sich zeigen, dass die Berechnung dieser Momente nur aus den gegebenen Daten f, g und u_h möglich ist. Der Ansatz der Momente wird sich als praktisch erweisen, die Berechnung derselben geschieht auf

lokalen (Patch-)Problemen. Die Existenz von $\{g_K\}_K \in \mathcal{P}$ ergibt sich dann durch die Bestimmung der Momente, denn die implizit voneinander abhängigen approximierten Flüsse lassen sich mithilfe dieser rekonstruieren, ohne je ein globales Problem gelöst zu haben.

3.4.1 Konstruktion der Momente

Ziel dieses Abschnittes ist es, eine Familie von Flüssen mit Äquilibrierungseigenschaft p-ter Ordnung zu erarbeiten. Sei dazu ein Viereck $K \in \mathcal{P}$ und $\gamma \in \mathcal{E}_K$ mit $\mathcal{N}(\gamma) = \{n_l, n_r\}$ gegebenen. Seien

$$\mathcal{B}_\gamma^{\text{node}} = \{\theta_{n_l}, \theta_{n_r}\} \quad \text{und } \mathcal{B}_\gamma^{\text{edge}} = \{\theta_{\gamma_1}, \dots, \theta_{\gamma_{p-1}}\}$$

die Mengen der zur Kante γ zu geordneten Basisfunktionen, wie in Abschnitt 3.3, dann besitzt g_K auf dieser Kante die Struktur

$$g_K|_\gamma = \lambda_l \theta_l + \lambda_r \theta_r + \sum_{i=1}^{p-1} \lambda_i \theta_{\gamma_i}, \quad \lambda_l, \lambda_r, \lambda_i \in \mathbb{R}, i = 1, \dots, p-1. \tag{13}$$

Durch Festlegung der Koeffizienten ist g_K auf γ damit eindeutig bestimmt. Wir definieren nun die erwähnten Momente der approximierenden Flüsse, mit deren Hilfe wir eine Möglichkeit zur Bestimmung solcher Koeffizienten bereitstellen.

Definition 3.5. (**Momente**)
Sei $K \in \mathcal{P}$ und $\gamma \in \mathcal{E}_K$. Wir definieren die Momente erster Ordnung durch

$$\mu_{K,n}^\gamma = \int_\gamma g_K \theta_n \mathrm{d}s, \quad \theta_n \in \mathcal{B}_\gamma^{\text{node}}.$$

Momente höherer Ordnung werden definiert als

$$\mu_{K,\gamma_i}^\gamma = \int_\gamma g_K \theta_{\gamma_i} \mathrm{d}s, \quad \theta_{\gamma_i} \in \mathcal{B}_\gamma^{\text{edge}}.$$

Wir wollen nun die Äquilibrierungseigenschaften p-ter Ordnung (12) mithilfe der Momente ausdrücken und betrachten die drei Bedingungen einzeln. Sei dazu zuächst γ eine Kante des Vierecks $K \in \mathcal{P}$.

- Zunächst gilt für eine nodale Basis $\theta_n \in \mathcal{B}_\gamma^{\text{node}}$

$$\int_{\partial K} g_K \, \theta_n \mathrm{d}s = \sum_{\gamma \in \mathcal{E}_K} \int_\gamma g_K|_\gamma \, \theta_n \mathrm{d}s = \sum_{\gamma \in \mathcal{E}_K \cap \mathcal{E}_n} \mu_{K,n}^\gamma. \tag{14}$$

Die letzte Gleichheit folgt aus der Eigenschaft einer nodalen Basisfunktion, auf allen Kanten, die keinen Endpunkt in n haben, zu verschwinden. Weiter gilt für eine der Kante zugeordnete Basisfunktion $\theta_{\gamma_i} \in \mathcal{B}_\gamma^{\text{edge}}$

$$\int_{\partial K} g_K\, \theta_{\gamma_i} \mathrm{d}s = \sum_{\gamma \in \mathcal{E}_K} \int_\gamma g_K|_\gamma\, \theta_{\gamma_i} \mathrm{d}s = \int_\gamma g_K|_\gamma\, \theta_{\gamma_i} \mathrm{d}s = \mu_{K,\gamma_i}^\gamma. \tag{15}$$

Dies liegt an der Tatsache, dass die Kanten-Basisfunktionen auf allen anderen Kanten den Wert Null annehmen.

- Aus der zweiten Forderung $g_K + g_{K'} = 0$ auf einer inneren Kante $\gamma = \partial K \cap \partial K'$ folgt

$$\begin{aligned} \int_\gamma g_K \theta_n \mathrm{d}s \;+\; \int_\gamma g_{K'} \theta_n \mathrm{d}s \;&=\; \mu_{K,n}^\gamma \;+\; \mu_{K',n}^\gamma \;=\; 0, \qquad \theta_n \in \mathcal{B}_\gamma^{\text{node}}, \\ \int_\gamma g_K \theta_{\gamma_i} \mathrm{d}s \;+\; \int_\gamma g_{K'} \theta_{\gamma_i} \mathrm{d}s \;&=\; \mu_{K,\gamma_i}^\gamma \;+\; \mu_{K',\gamma_i}^\gamma \;=\; 0, \qquad \theta_{\gamma_i} \in \mathcal{B}_\gamma^{\text{edge}}. \end{aligned} \tag{16}$$

- Die letzte Forderung $g_K|_\gamma = g$ auf $\gamma \in \mathcal{E}^N$ impliziert schließlich

$$\begin{aligned} \mu_{K,n}^\gamma &= \int_\gamma g \theta_n \mathrm{d}s \qquad \theta_n \in \mathcal{B}_\gamma^{\text{node}}, \\ \mu_{K,\gamma_i}^\gamma &= \int_\gamma g \theta_{\gamma_i} \mathrm{d}s, \qquad \theta_{\gamma_i} \in \mathcal{B}_\gamma^{\text{edge}}. \end{aligned} \tag{17}$$

Eine genaue Betrachtung der eben hergeleiteten Gleichungen für die Momente ergibt das folgende Resultat.

Satz 3.6. (**Eindeutigkeit höherer Momente**)
Die Momente höherer Ordnung sind eindeutig festgelegt. Sei $K \in \mathcal{P}$ und $\gamma \in \mathcal{E}_K$ mit zugeordneter Basis $\mathcal{B}_\gamma^{\text{edge}}$, dann gilt

$$\mu_{K,\gamma_i}^\gamma = -R_K(\theta_{\gamma_i}), \quad \forall \theta_{\gamma_i} \in \mathcal{B}_\gamma^{\text{edge}}. \tag{18}$$

Beweis. Wir unterscheiden die Art einer Kante $\gamma \in \mathcal{E}$. Sei dafür $\theta_{\gamma_i} \in \mathcal{B}_\gamma^{\text{edge}}$ eine beliebige aber feste Kantenbasisfunktion.

1. Falls $\gamma \in \mathcal{E}^I$ eine **innere** Kante ist, so existieren zwei Vierecke $K, K' \in \mathcal{P}$ mit $\gamma \in \mathcal{E}_K$ und $\gamma \in \mathcal{E}_{K'}$. Aus den Gleichungen in (12), (15) und (16) ergibt sich das lineare System

$$\left.\begin{aligned} \mu_{K,\gamma_i}^\gamma &= -R_K(\theta_{\gamma_i}), \\ \mu_{K',\gamma_i}^\gamma &= -R_{K'}(\theta_{\gamma_i}), \\ \mu_{K,\gamma_i}^\gamma \;+\; \mu_{K',\gamma_i}^\gamma &= 0. \end{aligned}\right\}$$

Dieses ist genau dann eindeutig lösbar, wenn die dritte Gleichung erfüllt ist. Wir rufen an dieser Stelle noch einmal die Trägereigenschaft der Kantenbasisfunktion auf inneren Kanten in Erinnerung:

$$\operatorname{supp}\theta_{\gamma_i} \subset K \cup K', \quad \gamma = \partial K \cap \partial K'.$$

Wegen dieser, Definition 3.1 und aufgrund der Galerkin-Orthogonalität folgt

$$\mu^{\gamma}_{K,\gamma_i} + \mu^{\gamma}_{K',\gamma_i} = -R_K(\theta_{\gamma_i}) - R_{K'}(\theta_{\gamma_i}) \overset{\text{supp}}{=} -R(\theta_{\gamma_i}) = B(u_h, \theta_{\gamma_i}) - L(\theta_{\gamma_i}) \overset{\text{Satz 2.17}}{=} 0.$$

2. Falls $\gamma \in \mathcal{E}^N$ eine **Neumannkante** ist, so existiert genau ein Viereck $K \in \mathcal{P}$ mit $\gamma \in \mathcal{E}_K$. Aus den Gleichungen (12) und (17) ergeben sich die zwei Bedingungen

$$\mu^{\gamma}_{K,\gamma_i} = -R_K(\theta_{\gamma_i}) \quad \text{und} \quad \mu^{\gamma}_{K,\gamma_i} = \int_{\gamma} g\theta_{\gamma_i} \mathrm{d}s.$$

Da es sich bei γ um eine Randkante handelt, ist $\operatorname{supp}\theta_{\gamma_i} \subset K$ und aufgrund der Galerkin-Orthogonalität folgt

$$\int_{\gamma} g\theta_{\gamma_i} \mathrm{d}s + R_K(\theta_{\gamma_i}) \overset{\text{supp}}{=} L(\theta_{\gamma_i}) - B(u_h, \theta_{\gamma_i}) \overset{2.17}{=} 0.$$

Also ist das Moment auch in diesem Fall eindeutig bestimmt.

3. Betrachten wir den letzten Fall einer Dirichletkante $\gamma \in \mathcal{E}_D$, dann vereinfachen sich die Gleichungen in (12) und (15) zu

$$\mu^{\gamma}_{K,\gamma_i} = -R_K(\theta_{\gamma_i})$$

und damit ist auch dieses Moment festgelegt.

□

Es verbleibt also eine genauere Betrachtung der Momente erster Ordnung. Wir fassen die den nodalen Basisfunktionen zugeordneten Gleichungen in (14), (16) und (17) für ein $K \in \mathcal{P}$ noch einmal zusammen:

$$\begin{cases} \sum\limits_{\gamma \in \mathcal{E}_K \cap \mathcal{E}_n} \mu^{\gamma}_{K,n} &= -R_K(\theta_n), & \forall n \in \mathcal{N}_K, \\ \mu^{\gamma}_{K,n} + \mu^{\gamma}_{K',n} &= 0, & \forall n \in \mathcal{N}(\gamma), \gamma = \partial K \cap \partial K', \\ \mu^{\gamma}_{K,n} &= \int\limits_{\gamma} g\theta_n \mathrm{d}s, & \forall n \in \mathcal{N}(\gamma), \gamma = \partial K \cap \Gamma_N. \end{cases} \tag{19}$$

Nun wollen wir diese Gleichungen in Bezug auf einen Knoten $n \in \mathcal{N}$ mit zugehöriger nodaler Basis θ_n umsortieren. Sei $n \in \mathcal{N}$ und $\mathcal{P}_n$ der zugehörige Patch mit

$$\mathcal{P}_n = \{K_1, \dots, K_N\},$$

wobei K_i und K_{i+1} für $i = 1, \dots, N-1$ aneinandergrenzen. Wir unterscheiden nun, ob n ein innerer oder ein Randknoten, d.h. $n \in \partial\Omega$, ist.

- Sei n ein **innerer Knoten** und $\mathcal{E}_n = \{\gamma_1, \dots, \gamma_N\}$ der zugehörige Kanten-Patch. In diesem Fall grenzen K_1 und K_N ebenfalls aneinander. Die Abbildung 1 soll die Nummerierung der Vierecke und Kanten in diesem Fall verdeutlichen.

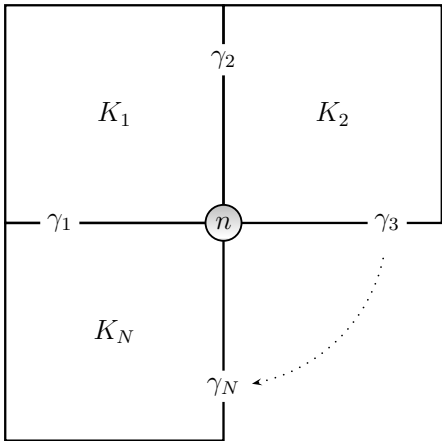

Abbildung 1: Patch $\mathcal{P}_n$ eines inneren Knoten n

In diesem Fall besteht die Menge $\mathcal{E}_{K_i} \cap \mathcal{E}_n$ für ein Viereck $K_i \in \mathcal{P}_n$ aus genau zwei Kanten. Es gilt

$$\mathcal{E}_{K_i} \cap \mathcal{E}_n = \{\gamma_i, \gamma_{\mathrm{mod}(i,N)+1}\}.$$

Da n ein innerer Knoten ist, brauchen wir nur die ersten beiden Bedingungen in (19) bei der Umformulierung berücksichtigen. Aus der ersten Bedingung ergeben sich damit die Gleichungen

$$\begin{aligned} \mu_{K_i,n}^{\gamma_i} &+ \mu_{K_i,n}^{\gamma_{i+1}} = -R_{K_i}(\theta_n), \qquad i = 1, \dots, N-1. \\ \mu_{K_N,n}^{\gamma_N} &+ \mu_{K_N,n}^{\gamma_1} = -R_{K_N}(\theta_n). \end{aligned}$$

Die zweite Bedingung bezieht sich auf angrenzende Elemente und daher ergeben sich weitere N Gleichungen

$$\begin{aligned} \mu_{K_1,n}^{\gamma_1} &+ \mu_{K_N,n}^{\gamma_1} = 0, \\ \mu_{K_i,n}^{\gamma_i} &+ \mu_{K_{i-1},n}^{\gamma_i} = 0, \qquad i = 2, \dots, N. \end{aligned}$$

- Sei n ein **Randknoten** und $\mathcal{E}_n = \{\gamma_1, \dots, \gamma_{N+1}\}$ der zugehörige Kanten-Patch. Die Kanten γ_1 und γ_{N+1} seien dabei Randkanten, die je nach Lage des Knotens n auf dem Neumannrand oder dem Dirichletrand liegen. Die Abbildung 2 soll die Nummerierung der Vierecke und Kanten auch in diesem Fall verdeutlichen.

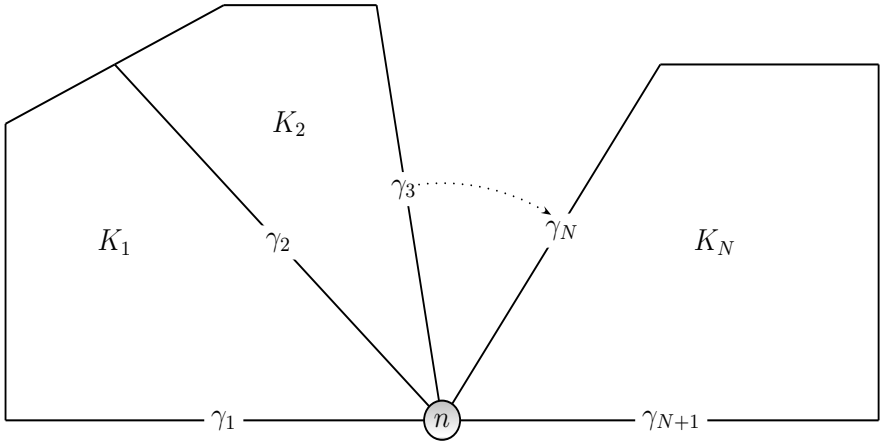

Abbildung 2: Patch $\mathcal{P}_n$ eines Randknoten n

Aus der ersten Momentengleichung in (19) folgen die Bedingungen

$$\mu_{K_i,n}^{\gamma_i} + \mu_{K_i,n}^{\gamma_{i+1}} = -R_K(\theta_n), \quad i = 1, \dots, N$$

und aus der zweiten ergeben sich die weiteren Einschränkungen

$$\mu_{K_{i-1},n}^{\gamma_i} + \mu_{K_i,n}^{\gamma_i} = 0, \quad i = 2, \dots, N$$

auf den inneren Kanten $\gamma_2, \dots, \gamma_N$. Um die letzten Bedingungen festhalten zu können, unterscheiden wir die Art der Randkanten.

1) Beide Kanten liegen auf dem **Dirichletrand**, das heißt $\gamma_1, \gamma_{N+1} \in \mathcal{E}^D$. In diesem Fall folgen keine weiteren Einschränkungen an die zugeordneten Momente.
2) Beide Kanten liegen auf dem **Neumannrand**, das heißt $\gamma_1, \gamma_{N+1} \in \mathcal{E}^N$. In diesem Fall liefert die dritte Gleichung in (19) die zwei weiteren Gleichungen

$$\mu_{K_1,n}^{\gamma_1} = \int_{\gamma_1} g\theta_n \mathrm{d}s \quad \text{und} \quad \mu_{K_N,n}^{\gamma_{N+1}} = \int_{\gamma_{N+1}} g\theta_n \mathrm{d}s.$$

3) Eine Kante liegt auf dem **Neumannrand**, die andere auf dem **Dirichletrand**. Analog zum ersten Fall ergeben sich für die Dirichletkante keine weiteren Einschränkungen. Falls $\gamma_1 \in \mathcal{E}^N$, dann folgt

$$\mu_{K_1,n}^{\gamma_1} = \int_{\gamma_1} g\theta_n \mathrm{d}s.$$

Umgekehrt folgt für $\gamma_{N+1} \in \mathcal{E}^N$

$$\mu_{K_N,n}^{\gamma_{N+1}} = \int_{\gamma_{N+1}} g\theta_n \mathrm{d}s.$$

Bemerkung. Die gegebenen Gleichungssysteme legen die Momente nicht eindeutig fest, wie das folgende Beispiel zeigt.

Beispiel 3.7. Betrachten wir den Fall eines inneren Knotens $n \in \mathcal{N}$ mit zugehörigem Patch $\mathcal{P}_n$ und Kantenpatch $\mathcal{E}_n$ mit

$$\mathcal{P}_n = \{K_1, K_2, K_3\} \quad \text{und} \quad \mathcal{E}_n = \{\gamma_1, \gamma_2, \gamma_3\}$$

mit Anordnung wie in Abbildung 1. Das lineare Gleichungssystem besitzt dann die Form

$$\underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}}_{:=A} \begin{pmatrix} \mu_{K_1,n}^{\gamma_1} \\ \mu_{K_1,n}^{\gamma_2} \\ \mu_{K_2,n}^{\gamma_2} \\ \mu_{K_2,n}^{\gamma_3} \\ \mu_{K_3,n}^{\gamma_3} \\ \mu_{K_3,n}^{\gamma_1} \end{pmatrix} = \begin{pmatrix} -R_{K_1}(\theta_n) \\ -R_{K_2}(\theta_n) \\ -R_{K_3}(\theta_n) \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Der Träger der Basisfunktion θ_n besteht in diesem Fall aus den drei Vierecken und deshalb summiert sich die rechte Seite aufgrund der Galerkin-Orthogonalität zu Null. Der Kern der Matrix ist nicht trivial und weiter ist $\ker(\mathrm{A}^{\mathrm{T}}) = \mathrm{span}\{(1,1,1,-1,-1,-1)^T\}$. Eine Lösung des Systems existiert, dank der Eigenschaft der rechten Seite, orthogonal auf $\ker\mathrm{A}^{\mathrm{T}}$ zu liegen. Dies liegt an der Eigenschaft, dass $\ker(A^T) = (\mathrm{Bild}(A))^{\perp}$ gilt und damit die rechte Seite in $\mathrm{Bild}(A)$ liegt. Eine Lösung ist aber aufgrund des nicht trivialen Kerns von A keineswegs eindeutig.

Dieses Verhalten werden wir im Abschnitt 3.4.2 über topologische Matrizen in ähnlicher Form erneut antreffen und dort genauer studieren. Die Idee ist es nun, durch eine Zusatzforderung die Eindeutigkeit (und Existenz) der Momente dennoch zu erreichen, um damit den Weg zu einem Algorithmus zur Berechnung dieser zu ebnen. In [1] wird dazu ein Minimierungsansatz gewählt, den wir nun folgend vorstellen wollen. Dazu definieren wir vorerst die approximierenden Momente.

Definition 3.8. (**approximierte Momente**)
Sei $K \in \mathcal{P}$, $\gamma \in \mathcal{E}_K$ und $n \in \mathcal{N}(\gamma)$ mit zugeordneter nodaler Basis $\theta_n \in \mathcal{B}_K^{\text{node}}$. Weiter sei u_h die approximierte FE-Lösung. Dann definieren wir das approximierte Moment $\tilde{\mu}_{K,n}^{\gamma}$ auf einer inneren oder Dirichletkante durch

$$\tilde{\mu}_{K,n}^{\gamma} = \int_{\gamma} \left.\frac{\partial u_h}{\partial \mathbf{n}_K}\right|_K \theta_n \mathrm{d}s = \int_{\gamma} \mathbf{n}_K \cdot \nabla u_h|_K \, \theta_n \mathrm{d}s$$

sowie auf einer Neumannkante durch

$$\tilde{\mu}_{K,n}^{\gamma} = \int_{\gamma} g\theta_n \mathrm{d}s.$$

Die Notation der Indizes ist dabei analog zur Notation der Momente erster Ordnung.

Wir minimieren den quadratischen Abstand zwischen den Momenten erster Ordnung und den approximierten Flüssen, sodass die hergeleiteten Bedingungen an die Momente erfüllt sind. Mit diesem Ansatz werden die Momente schließlich eindeutig definiert. Dieses Resultat wollen wir in einem Satz zusammenfassen.

Satz 3.9. (**Existenz und Eindeutigkeit der Momente erster Ordnung**)
Für einen Knoten $n \in \mathcal{N}$ ist das Minimierungsproblem

$$\min \quad \frac{1}{2} \sum_{K \in \mathcal{P}_n} \sum_{\gamma \in \mathcal{E}_n \cap \mathcal{E}_K} \left(\mu_{K,n}^{\gamma} - \tilde{\mu}_{K,n}^{\gamma} \right)^2$$

unter den Nebenbedingungen

$$\begin{aligned} \sum_{\gamma \in \mathcal{E}_n \cap \mathcal{E}_K} \mu_{K,n}^{\gamma} &= -R_K(\theta_n), && K \in \mathcal{P}_{\backslash}, \\ \mu_{K,n}^{\gamma} &= \int_{\gamma} g \theta_n \mathrm{d}s, && \gamma \in \mathcal{E}_n \cap \mathcal{E}^N, \\ \mu_{K,n}^{\gamma} + \mu_{K',n}^{\gamma} &= 0, && \gamma = \partial K \cap \partial K' \in \mathcal{E}_n, \end{aligned}$$

eindeutig lösbar, wobei für alle $K \in \mathcal{P}_n$ Lagrangemultiplikatoren $\sigma_{K,n} \in \mathbb{R}$ existieren, sodass

$$\mu_{K,n}^{\gamma} = \begin{cases} \frac{1}{2}(\sigma_{K,n} - \sigma_{K',n} + \tilde{\mu}_{K,n}^{\gamma} - \tilde{\mu}_{K',n}^{\gamma}), & \gamma = \partial K \cap \partial K' \in \mathcal{E}_n, \\ \int_{\gamma} g \theta_n \mathrm{d}s, & \gamma \in \mathcal{E}_n \cap \mathcal{E}_K \cap \mathcal{E}^N, \\ \sigma_{K,n} + \tilde{\mu}_{K,n}^{\gamma}, & \gamma \in \mathcal{E}_n \cap \mathcal{E}_K \cap \mathcal{E}^D. \end{cases} \tag{20}$$

Diese erfüllen dabei die Bedingungen

$$\frac{1}{2} \sum_{K' : \gamma = \partial K \cap \partial K'} (\sigma_{K,n} - \sigma_{K',n}) + \sum_{\gamma \in \mathcal{E}^D \cap \mathcal{E}_K} \sigma_{K,n} = \tilde{r}_K(\theta_n), \quad K \in \mathcal{P}_n, \tag{21}$$

mit

$$\tilde{r}_K(\theta_n) = -R_K(\theta_n) - \frac{1}{2} \sum_{\gamma = \partial K \cap \partial K'} (\tilde{\mu}_{K,n}^{\gamma} - \tilde{\mu}_{K',n}^{\gamma}) - \sum_{\gamma \in \mathcal{E}^D \cap \mathcal{E}_K} \tilde{\mu}_{K,n}^{\gamma} - \sum_{\gamma \in \mathcal{E}^N \cap \mathcal{E}_K} \tilde{\mu}_{K,n}^{\gamma}.$$

Beweis. Ein Beweis mithilfe von Lagrangemultiplikatoren befindet sich in ([1], S.123-127) □

Bemerkung. Das Gleichungssystem in (21) ist im Fall eines Neumann-Neumann Patches oder eines inneren Patches $\mathcal{P}_n$ nicht eindeutig lösbar. Die Summe über die Dirichletkanten entfällt hierbei. Im Kapitel über topologische Matrizen werden wir sehen, falls $\sigma = (\sigma_{K_1}, \dots, \sigma_{K_N})^T$, $K_i \in \mathcal{P}_n$ Lösung von (21) ist, so ist auch $\sigma + \alpha \mathbf{1}$ eine Lösung. Letztendlich sind wir an der Differenz der Komponenten von σ interessiert, daher ist die Wahl einer Lösung nicht entscheidend.

Einen weiteren Minimierungsansatz nach [9] fasst der folgende Satz zusammen.

Satz 3.10. (**gewichteter Minimierungsansatz**)
Minimiert man statt der Zielfunktion in Satz 3.9 die gewichtete Zielfunktion

$$\min \quad \frac{1}{2} \sum_{K \in \mathcal{P}_n} \sum_{\gamma \in \mathcal{E}_n \cap \mathcal{E}_K} \omega_\gamma^2 \left(\mu_{K,n}^\gamma - \tilde{\mu}_{K,n}^\gamma \right)^2$$

mit $\omega_\gamma = (\int_\gamma 1 \mathrm{d}s)^{-1}$, so sind die Momente eindeutig festgelegt durch

$$\mu_{K,n}^\gamma = \begin{cases} \frac{1}{2}\omega_\gamma^{-2}(\sigma_{K,n} - \sigma_{K',n}) + \frac{1}{2}(\tilde{\mu}_{K,n}^\gamma + \tilde{\mu}_{K',n}^\gamma), & \gamma = \partial K \cap \partial K' \in \mathcal{E}_n, \\ \int_\gamma g \theta_n \mathrm{d}s, & \gamma \in \mathcal{E}_n \cap \mathcal{E}_K \cap \mathcal{E}^N, \\ \omega_\gamma^{-2}\sigma_{K,n} + \tilde{\mu}_{K,n}^\gamma, & \gamma \in \mathcal{E}_n \cap \mathcal{E}_K \cap \mathcal{E}^D. \end{cases} \tag{22}$$

Die $\{\sigma_{K,n}\}$ erfüllen dabei die Bedingungen

$$\frac{1}{2} \sum_{\gamma = \partial K \cap \partial K'} \omega_\gamma^{-2}(\sigma_{K,n} - \sigma_{K',n}) + \sum_{\gamma \mathcal{E}^D \cap \mathcal{E}_K} \omega_\gamma^{-2}\sigma_{K,n} = \tilde{r}_K(\theta_n) \tag{23}$$

mit $\tilde{r}_K(\theta_n)$ aus Satz 3.9.

Beweis. Analog zum Beweis von Satz 3.9. □

Die Gewichtung der Momente führt mitunter zu besseren Effektivitätsindizes (vgl. [9]). Dieser Satz ist der Vollständigkeit halber an dieser Stelle erwähnt, er wird in unserer weiteren Analyse keine Rolle spielen.

Bemerkung. Für den Fall konstanter Kantenlängen, das heißt $\exists c \in \mathbb{R}$, sodass $\omega_\gamma^{-1} = c \; \forall \gamma \in \mathcal{E}$. Liefern die Sätze 3.9 und 3.10 die gleiche Familie von Momenten erster Ordnung.

Zu diesem Zeitpunkt haben wir die Existenz und Eindeutigkeit zweier Klassen von Momenten erster Ordnung sowie Momente höherer Ordnung erbracht. Zur Berechnung der Momente erster Ordnung müssen die linearen Gleichungssysteme in (21) bzw. (23) berechnet werden. Dies führt auf die Analyse topologischer Matrizen.

3.4.2 Topologische Matrizen und ihre Inversen

Ziel dieses Abschnittes ist die Gleichungssysteme in (21) effektiv zu lösen. Der Begriff der *topologischen Matrizen* und die Grundidee der Fixierung einer Lösung geht zurück auf [3]. In [9] wird eine Direktlösung der Differenzen $\sigma_{K,n} - \sigma_{K',n}$ für Gleichungssysteme zu inneren Knoten vorgeschlagen. Wir wollen hier, durch Angabe von Inversen, für jedes *Patchproblem* eine preiswerte Lösungsmethode beschreiben und damit einen anderen Ansatz wählen. Die Struktur des Gleichungssystems in (21) hängt von der Art des ihm zugeordneten Knotens $n \in \mathcal{N}$ mit

zugehörigem Patch $\mathcal{P}_n$ ab. Allgemein werden wir die Gleichungsysteme von der Form

$$\frac{1}{2}T\sigma = \tilde{r}_n \quad \text{oder} \quad \frac{1}{2}(T+F)\sigma = \tilde{r}_n \tag{24}$$

mit $T, F \in \mathbb{R}^{N,N}$, $\sigma = (\sigma_{K_1,n}, \dots, \sigma_{K_N,n})^T$ und $\tilde{r}_n = (\tilde{r}_{K_1}(\theta_n), \dots, \tilde{r}_{K_N}(\theta_n))^T$ für $K_i \in \mathcal{P}_n$, $i = 1 \dots, N$, mit einer symmetrischen Matrix T, betrachten. Wir definieren dazu diese sogenannte Fixierungsmatrix F mit

$$F := \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$$

Wir bezeichnen nun mit $T_{\mathrm{I}}, T_{\mathrm{DD}}, T_{\mathrm{NN}}, T_{\mathrm{ND}}, T_{\mathrm{DN}}$ die Systemmatrizen zu einem inneren-, Dirichlet-Dirichlet-,..., Dirichlet-Neumann-Knoten. In ([1], S.125-127) wird gezeigt, dass die Matrizen T_{DD} und T_{DN} invertierbar sind und dass die Matrizen T_{I} und T_{NN} den nichttrivialen Kern

$$\ker T_{\mathrm{I}} = \ker T_{\mathrm{NN}} = \operatorname{span}\{(1, \dots, 1) \in \mathbb{R}^N\}$$

besitzen. Aufgrund der Galerkin-Orthogonalität summiert sich in diesen Fällen die rechte Seite zu Null und daher sind die Gleichungssystem dennoch lösbar. Dies motiviert die Struktur der Matrix F: Sei $T = T_{\mathrm{I}}$ oder $T = T_{\mathrm{NN}}$ und $\tilde{\sigma} = \tilde{\sigma}^{\perp} + \alpha\mathbf{1}$, $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^N$, $\tilde{\sigma}^{\perp} \in (\ker T)^{\perp}$ eine Lösung von $\frac{1}{2}T\sigma = \tilde{r}_n$. Für die Wahl $\alpha = -\sum_{i=1}^{N} \tilde{\sigma}^{\perp}_{(i)}/N$ löst $\tilde{\sigma}$ die Gleichung $\frac{1}{2}(T+F)\sigma = \tilde{r}_n$, denn

$$\frac{1}{2}(T+F)(\tilde{\sigma}^{\perp} + \alpha\mathbf{1}) = \underbrace{\frac{1}{2}T\tilde{\sigma}^{\perp}}_{=b} + \frac{\alpha}{2}\underbrace{T\mathbf{1}}_{=0} + \frac{1}{2}\underbrace{(F\tilde{\sigma}^{\perp} + \alpha F\mathbf{1})}_{=0} = \tilde{r}_n.$$

Durch das Lösen von $\frac{1}{2}(T+F)\sigma = \tilde{r}_n$ fixiert man also eine Lösung von $\frac{1}{2}T\sigma = \tilde{r}_n$ für den inneren oder den Neumann-Neumann Fall. Die Matrizen $(T_{\mathrm{I}} + F)$ und $(T_{\mathrm{NN}} + F)$ sind positiv definit und damit invertierbar.
Sei rev: $\mathbb{R}^N \to \mathbb{R}^N$ der Reverseoperator mit $\operatorname{rev}(\sigma_1, \dots, \sigma_N) = (\sigma_N, \dots, \sigma_1)$. Löst $\tilde{\sigma}$ das System $T_{\mathrm{DN}}\sigma = \tilde{r}_n$, so löst $\operatorname{rev}(\tilde{\sigma})$ das System $T_{\mathrm{ND}}\sigma = \operatorname{rev}(\tilde{r}_n)$.
Es genügt also die Betrachtung der symmetrischen postiv definiten Matrizen $T_{\mathrm{I}} + F$, $T_{\mathrm{NN}} + F$, T_{DD} und T_{DN}. Bevor wir deren inverse Matrizen angeben, wollen wir, zur besseren Lesbarkeit, diese Matrizen noch einmal darstellen:

- Die Matrix $T_{\mathrm{I}}^F = T_{\mathrm{I}} + F$, ist eine *zirkuläre* Matrix, ihre Inverse ist damit auch zirkulär [10]. Sie ist eindeutig durch Angabe einer Zeile festgelegt. Dies werden wir bei der Inversenangabe ausnutzen. Die Matrix eines inneren Patchproblems wird aus geometrischen

Gründen nur für $N \geq 3$ relevant sein und ist gegeben durch

$$T_{\mathrm{I}}^F = \begin{bmatrix} 3 & 0 & 1 & \cdots & 1 & 0 \\ 0 & 3 & 0 & 1 & \cdots & 1 \\ & \vdots & & \to & \vdots & \\ 0 & 1 & \cdots & 1 & 0 & 3 \end{bmatrix}.$$

- Die Matrix T_{DD} ist für $N = 1, 2$ und $N \geq 3$ gegeben durch

$$T_{\mathrm{DD}} = 4, \qquad T_{\mathrm{DD}} = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}, \qquad T_{\mathrm{DD}} = \begin{bmatrix} 3 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 3 \end{bmatrix}.$$

- Die Matrix $T_{\mathrm{NN}}^F = T_{\mathrm{NN}} + F$ ist für $N = 1, 2$ und $N \geq 3$ gegeben durch

$$T_{\mathrm{NN}}^F = 2, \qquad T_{\mathrm{NN}}^F = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}, \qquad T_{\mathrm{NN}}^F = \begin{bmatrix} 2 & 0 & 1 & \cdots & 1 \\ 0 & 3 & 0 & \ddots & \vdots \\ 1 & \ddots & \ddots & \ddots & 1 \\ \vdots & \ddots & 0 & 3 & 0 \\ 1 & \cdots & 1 & 0 & 2 \end{bmatrix}.$$

- Schließlich ist die Matrix T_{DN} für $N = 1, 2$ und $N \geq 3$ gegeben durch

$$T_{\mathrm{DN}} = 2, \qquad T_{\mathrm{DN}} = \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix}, \qquad T_{\mathrm{DN}} = \begin{bmatrix} 3 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}.$$

Die angebenen Formeln der inversen Matrizen sind durch Polynominterpolation und scharfes Hinsehen entstanden. Wir definieren zunächst

Definition 3.11. (**Symmetrie bzgl. beider Diagonalen.**)
Eine Matrix $T^{-1} \in \mathbb{R}^{N,N}$ ist symmetrisch bezüglich beider Diagonalen, falls gilt

$$\left(T^{-1}\right)_{l,j} = \left(T^{-1}\right)_{j,l} = \left(T^{-1}\right)_{N-l,N-j} = \left(T^{-1}\right)_{N-j,N-l}$$

für $l = 1 \dots \lfloor \frac{N+1}{2} \rfloor$, $j = l \dots N - l + 1$.

Beispiel 3.12. Sei $\dim(T_{\mathrm{DD}}) = 4$ und $\dim(T^F_{\mathrm{NN}}) = 3$, dann sind die Matrizen

$$\det(T_{\mathrm{DD}}) \left(\frac{1}{2} T_{\mathrm{DD}}\right)^{-1} = \begin{bmatrix} 7 & 5 & 3 & 1 \\ 5 & 15 & 9 & 3 \\ 3 & 9 & 15 & 5 \\ 1 & 3 & 5 & 7 \end{bmatrix} \quad \text{und} \quad \det(T^F_{\mathrm{NN}}) \left(\frac{1}{2} T^F_{\mathrm{NN}}\right)^{-1} = \begin{bmatrix} 12 & 0 & -6 \\ 0 & 6 & 0 \\ -6 & 0 & 12 \end{bmatrix}$$

symmetrisch bezüglich beider Diagonalen. Mit der Definition und der folgenden Struktur der Inversenformeln lässt sich die Lösung der Gleichungssysteme in (24) einfach implementieren.

Satz 3.13. (**Inversen der topologischen Matrizen**)

(i) Für die zirkuläre Matrix $T_{\mathrm{I}}^F \in \mathbb{R}^{N,N}$ gilt

$$\left(\frac{1}{2}T_{\mathrm{I}}^F\right)^{-1}_{1,j} = \frac{1}{\det(T_{\mathrm{I}}^F)} \cdot \frac{N^4 - 6(j-1)N^3 + (6(j-1)^2 - 1)N^2 + 12N}{6}, \quad j = 1 \dots N.$$

mit $\det(T_{\mathrm{I}}^F) = N^3$.

(ii) Die Inverse der Matrix $T_{\mathrm{DD}} \in \mathbb{R}^{N,N}$ ist symmetrisch bezüglich beider Diagonalen, mit

$$\left(\frac{1}{2}T_{\mathrm{DD}}\right)^{-1}_{l,j} = \frac{1}{\det(T_{\mathrm{DD}})}(2(l-1)+1)(2N - 2(l-1) - 1 - 2(j-l))$$

für $l = 1, \dots, \lfloor \frac{(N+1)}{2} \rfloor, j = l \dots N - l + 1$ und $\det(T_{\mathrm{DD}}) = 2N$.

(iii) Die Inverse der Matrix $T_{\mathrm{NN}}^F \in \mathbb{R}^{N,N}$ ist symmetrisch bezüglich beider Diagonalen, mit

$$\left(\frac{1}{2}T_{\mathrm{NN}}^F\right)^{-1}_{l,j} = \frac{1}{\det(T_{\mathrm{NN}}^F)} \left[\frac{2}{3}N^3 - (1 + 2(j-1))N^2 + \quad ((j-1)^2 + j - 1 + (l-1)^2 + l - 1 + 1/3)N + 2\right]$$

für $l = 1, \dots, \lfloor \frac{(N+1)}{2} \rfloor, j = l \dots N - l + 1$ und $\det(T_{\mathrm{NN}}^F) = N^2$.

(iv) Die Inverse der Matrix $T_{\mathrm{DN}} \in \mathbb{R}^{N,N}$ ist gegeben durch

$$\left(\frac{1}{2}T_{\mathrm{DN}}\right)^{-1}_{i,j} = 2\min(i,j) - 1, \quad i,j = 1 \dots, N.$$

Beweis. Ein Teil des Beweises befindet sich in Anhang A. □

3.4.3 Rekonstruktion der Flüsse

Wir sind zu diesem Zeitpunkt also in der Lage eine Familie von Momenten $\{\mu_{K,n}^{\gamma}\}$ zu berechnen, die der Momentenformulierung der Äquilibrierungseigenschaft p-ten Grades in (14)-(17) genügt. Es sei an dieser Stelle an den Strukturansatz der approximierten Flüsse in (11) und die Notation in (13) erinnert. Sei $K \in \mathcal{P}$ mit einer Kante $\gamma \in \mathcal{E}_{\mathcal{K}} \setminus \mathcal{E}_K^N$ gegeben, dann ist

$$g_K|_\gamma = \lambda_l \theta_l|_\gamma + \lambda_r \theta_r|_\gamma + \sum_{i=1}^{p-1} \lambda_i \theta_{\gamma_i}|_\gamma$$

für noch zu bestimmende $\lambda_l, \lambda_r, \lambda_i \in \mathbb{R}, i = 1 \dots p-1$. Die Definition der Momente wird nun ersichtlich, für $I := \{l, r, \gamma_1, \dots, \gamma_{p-1}\}$ und $n \in I$ gilt

$$\sum_{m \in I} \lambda_m (\theta_m, \theta_n)_{L^2(\gamma)} = \int_\gamma g_K|_\gamma \theta_n \mathrm{d}s = \mu_{K,n}^\gamma.$$

Dies ist ein lineares Gleichungssystem der Form

$$M_\gamma \lambda = \mu_K^\gamma,$$

wobei $M_\gamma \in \mathbb{R}^{p+1,p+1}$ die Kanten-Massematrix mit $(M_\gamma)_{m,n} = (\theta_m, \theta_n)_{L^2(\gamma)}$ und $\lambda, \mu_K^\gamma \in \mathbb{R}^{p+1}$, mit $\lambda = (\lambda_l, \lambda_r, \lambda_{\gamma_1}, \dots, \lambda_{\gamma_{p-1}})$ und $\mu_K^\gamma = (\mu_{K,l}^\gamma, \mu_{K,r}^\gamma, \mu_{K,\gamma_1}^\gamma, \dots, \mu_{K,\gamma_{p-1}}^\gamma)$, den Koeffizienten-und Momentevektor bezeichnet. Die Massematrix ist aufgrund der linearen Unabhängigkeit der Basisfunktionen invertierbar und somit lassen sich die Koeffizienten durch Lösen des Gleichungssystems eindeutig bestimmen.
Dass die so definierten Flüsse die Äquilibrierungseigenschaft p-ter Ordnung haben, folgt aus der Konstruktion der Momente. Die Existenz einer solchen Familie approximierender Flüsse ist damit gezeigt.

3.5 Qualität des Fehlerschätzers

Wir haben also einen impliziten Fehlerschätzer auf Basis äquilibrierter Flüsse konstruiert. Ziel dieses Abschnitts ist es nun, diesen Fehlerschätzer auf Verlässlichkeit und Effizienz zu untersuchen. Weiterhin werden wir für den Spezialfall von Rechteckgittern und genügend regulärer Lösung sogar asymptotische Exaktheit des Fehlerschätzers aufzeigen.

3.5.1 Zuverlässlichkeit

Zunächst werden wir zeigen, dass der angebene Fehlerschätzer eine *garantierte* obere Schranke mit $C_{zuv} = 1$ zum exakten Fehler darstellt. Diese Tatsache werden wir in einer etwas allgemeineren Form nachweisen und definieren dafür

Definition 3.14. (**äquilibrierte Flussfunktionale**)
Die Funktionale $\zeta_K \colon H_0^1(K) \to \mathbb{R}$, $K \in \mathcal{P}$ heißen äquilibrierte Flussfunktionale, falls gilt

(i) Die Familie $\{\zeta_K\}_{K\in\mathcal{P}}$ ist *konsistent*:

$$\sum_{K\in\mathcal{P}} \zeta_K(v) = \int_{\Gamma_N} gv, \quad v \in H_0^1(\Omega).$$

(ii) Die Familie $\{\zeta_K\}_{K\in\mathcal{P}}$ ist mit den lokalen inneren Residuen auf den lokalen Approximationsräumen *vollständig äquilibriert*:

$$R_K(v) + \zeta_K(v) = 0, \quad \forall v \in P_K^p, K \in \mathcal{P}.$$

Mithilfe dieser Definition erzielen wir das folgende Resultat

Satz 3.15. (**Zuverlässlichkeit**)
Sei $\{\zeta_K\}_{K\in\mathcal{P}}$ eine Familie äquilibrierter Flüsse und zu $K \in \mathcal{P}$ sei $\tilde{e}_K \in H_0^1(K)$ eine Lösung des lokalen residualen Problems

$$B_K(\tilde{e}_K, v) = R_K(v) + \zeta_K(v), \quad v \in H_0^1(K),$$

dann lässt sich der Fehler e in der Energienorm abschätzen durch

$$\|\!|e|\!\|^2 \leq \sum_{K\in\mathcal{P}} \|\!|\tilde{e}_K|\!\|_K^2 .$$

Beweis. Zunächst stellen wir fest, dass die lokalen Probleme lösbar sind. Dies folgt sofort für $c > 0$ oder $\Gamma_D \cap \partial K \neq \emptyset$. Für $c = 0$ stellt die vollständige Äquilibrierung der Flussfunktionale die Existenz einer Lösung sicher. Weiter gilt mithilfe der Konsistenzeigenschaft für $v \in H_0^1(\Omega)$

$$B(e, v) = L(v) - B(u_h, v) + \int_{\Gamma_N} gv\mathrm{d}s = \sum_{K\in\mathcal{P}} (R_K(v) + \zeta_K(v)) = \sum_{K\in\mathcal{P}} B_K(\tilde{e}_K, v).$$

Durch Anwendung der Cauchy-Schwarz-Ungleichungen erhalten wir die Abschätzung

$$|B(e, v)| \leq \sum_{K\in\mathcal{P}} |B_K(\tilde{e}_K, v)| \leq \sum_{K\in\mathcal{P}} \|\!|\tilde{e}_K|\!\|_K \, \|\!|v|\!\|_K \leq \left(\sum_{K\in\mathcal{P}} \|\!|\tilde{e}_K|\!\|^2\right)^{1/2} \underbrace{\left(\sum_{K\in\mathcal{P}} \|\!|v|\!\|_K\right)^{1/2}}_{=\|\!|v|\!\|} .$$

Wir Erinnerung uns an die Charakterisierung des Fehlers in Satz 2.17 und erhalten schließlich

$$\|\!|e|\!\| = \sup_{v\in H_0^1(\Omega)\setminus\{0\}} \frac{|B(e,v)|}{\|\!|v|\!\|} \leq \left(\sum_{K\in\mathcal{P}} \|\!|\tilde{e}_K|\!\|^2\right)^{1/2}.$$

□

Definieren wir nun für $K \in \mathcal{P}$,

$$\zeta_K = \int_{\partial K} g_K v \mathrm{d}s,$$

wobei $\{g_K\}_K$ die Äquilibrierungseigenschaft p-ten Grades erfüllt, so erhalten wir die Verlässlichkeit der Fehlerschätzers nach [1] und [9] :
Bei der Herleitung von Anforderungen an die bzw. bei der Strukurfestlegung von approximierten Flüssen haben wir bereits gesehen, dass die dadurch induzierten Flussfunktionale konsistent und vollständig äquilibriert sind.

Bemerkung. In Satz (3.15) wird die *exakte* Lösung der lokalen Probleme gefordert, um eine garantierte obere Schranke an den exakten Fehler zu erhalten. Im Allgemeinen können wir die lokalen Probleme wiederum nur approximieren und unterschätzen dadurch möglicherweise den Fehlerterm $\|\!|e|\!\|$.

3.5.2 Konsistenz

Der hier konstruierte Fehlerschätzer ist im folgenden Sinne konsistent. Wählen wir als Familie approximierter Flüsse $\{g_K\}$ die exakten Flüsse, das heißt

$$g_K = \left.\frac{\partial u}{\partial \mathbf{n}_K}\right|_K, \quad \forall K \in \mathcal{P},$$

so gilt in Satz 3.15 die Gleichheit. In diesem Fall gilt für den lokalen Fehler $\tilde{e}_K = e|_K$, $K \in \mathcal{P}$, denn analog wie im Abschnitt 2.4.1 über den lokalen Fehler erhalten wir für alle $v \in H_0^1(K)$

$$B_K(e|_K, v) = R_K(v) + \int_{\partial K} \left.\frac{\partial e}{\partial \mathbf{n}_K}\right|_K v \mathrm{d}s + \int_{\partial K} \left.\frac{\partial u_h}{\partial \mathbf{n}_K}\right|_K v \mathrm{d}s = R_K(v) + \int_{\partial K} \left.\frac{\partial u}{\partial \mathbf{n}_K}\right|_K \mathrm{d}s.$$

3.5.3 Effizienz

In diesem Kapitel werden wir zeigen, dass sich die lokale Fehlerapproximation durch den exakten Fehler und Terme höherer Ordnung abschätzen lässt. Dazu werden wir vorerst ein Stabilitätsresultat der hier konstruierten äquilibrierten Flüsse herleiten und mit dessen Hilfe schließlich die Effizienz des Fehlerschätzers zeigen. Wir benötigen zunächst einige Hilfslemmata.

Lemma 3.16. (**Stabilität der Lagrange-Multiplikatoren**)
Analog zur Notation der Gleichung (21) in Satz 3.9, sei $\{\sigma_{K,n}\}_{K\in\mathcal{P}_n}$ eine Lösung von

$$\frac{1}{2}\sum_{K'\,:\,\gamma=\partial K\cap\partial K'}(\sigma_{K,n}-\sigma_{K',n})+\sum_{\gamma\in\mathcal{E}^D\cap\mathcal{E}_K}\sigma_{K,n}=\tilde{r}_K(\theta_n),\quad K\in\mathcal{P}_n,$$

dann existiert ein $C=C(\#\mathcal{P}_n)>0$ mit

$$\Sigma_n:=\frac{1}{4}\sum_{K\,:\,\gamma=\partial K\cap\partial K'}(\sigma_{K,n}-\sigma_{K',n})^2+\sum_{K\,:\,\gamma=\partial K\cap\Gamma_D}\sigma_{K,n}^2\leq C\sum_{K\in\mathcal{P}_\backslash}|\tilde{r}_K(\theta_n)|^2\,.$$

Beweis. Sei $n\in\mathcal{N}$ und $\mathcal{P}_n=\{K_1,\dots,K_N\}$ der zugeordnete Element-Patch. Wir beweisen durch Fallunterscheidung.

- Sei $\mathcal{E}^D\cap\mathcal{E}_n$ nichtleer und $\mathcal{E}_n=\{\gamma_1,\dots,\gamma_{N+1}\}$. Ohne Beschränkung der Allgemeinheit sei $\gamma_1\in\mathcal{E}^D$. Sei $|.|_2$ die euklidische Norm auf dem $\mathbb{R}^N$. In Anlehnung an Σ_n definieren wir die Abbildung $|.|_1\colon\mathbb{R}^N\to\mathbb{R},v\mapsto|v|_1$ mit

$$|v|_1^2:=\frac{1}{4}\left(\sum_{i=1}^{N-1}(v_i-v_{i+1})^2+\sum_{i=2}^{N}(v_i-v_{i-1})^2\right)+v_1^2+\underbrace{\sum_{\gamma_{N+1}\in\mathcal{E}^D}v_N^2}_{=0,\text{ falls }\gamma_{N+1}\notin\mathcal{E}^D}\,.$$

  Sie definiert eine weitere Norm auf dem $\mathbb{R}^N$ (wegen der Addition mit v_1^2 ist die Abbildung positiv definit). Sei nun $\sigma_n=(\sigma_{K_1,n},\sigma_{K_2,n},\dots,\sigma_{K_N,n})^T\in\mathbb{R}^N$, dann existiert aufgrund der Normäquivalenz in endlich dimensionalen Räumen ein $C=C(\#\mathcal{P}_n)>0$ mit

$$|\sigma_n|_2\leq C|\sigma_n|_1=C\sqrt{\Sigma_n}.$$

  Weiter gilt

$$\begin{aligned}\sum_{K\in\mathcal{P}_n}\sigma_{K,n}\tilde{r}_K(\theta_n)&=\frac{1}{2}\sum_{K\in\mathcal{P}_n}\sigma_{K,n}\sum_{K'\,:\,\gamma=\partial K\cap\partial K'}(\sigma_{K,n}-\sigma_{K',n})+\sum_{K\in\mathcal{P}_n}\sigma_{K,n}\overbrace{\sum_{\gamma\in\mathcal{E}^D\cap\mathcal{E}_K}\sigma_{K,n}}^{=0,\text{ falls }\mathcal{E}^D\cap\mathcal{E}_K=\emptyset}\\&=\frac{1}{2}\cdot\frac{1}{2}\cdot\sum_{K\,:\,\gamma=\partial K\cap\partial K'}(\sigma_{K,n}^2-2\sigma_{K,n}\sigma_{K',n}+\sigma_{K',n}^2)+\sum_{K\,:\,\gamma\in\mathcal{E}^D\cap\mathcal{E}_K}\sigma_{K,n}^2\\&=\frac{1}{4}\sum_{K\,:\,\gamma=\partial K\cap\partial K'}(\sigma_{K,n}-\sigma_{K',n})^2+\sum_{K\,:\,\gamma\in\mathcal{E}^D\cap\mathcal{E}_K}\sigma_{K,n}^2\\&=\Sigma_n.\end{aligned}$$

Zusammen mit der Abschätzung der Normen und der Cauchy-Schwarz-Ungleichung erhalten wir für $\tilde{r}_n = (\tilde{r}_{K_1}(\theta_n), \tilde{r}_{K_2}(\theta_n), \dots, \tilde{r}_{K_N}(\theta_n))^T \in \mathbb{R}^N$

$$\Sigma_n = \sum_{K\in\mathcal{P}_n} \sigma_{K,n}\tilde{r}_K(\theta_n) \leq |\sigma_n|_2|\tilde{r}_n|_2 \leq C\sqrt{\Sigma_n}|\tilde{r}_n|_2$$

und damit die gesuchte Abschätzung in diesem Fall.

- Sei nun $\mathcal{E}^D \cap \mathcal{E}_n$ leer. Aufgrund der Galerkin-Orthogonalität gilt $\sum_{K\in\mathcal{P}_n} \tilde{r}_K(\theta_n) = 0$. Betrachte nun den Raum

$$\mathbb{R}^N_{\Sigma=0} = \left\{ v \in \mathbb{R}^n | \sum_{i=1}^N v_i = 0 \right\}.$$

Die euklidische Norm $|.|_2$ sowie die Abbildung $|.|_1 \colon \mathbb{R}^N \to \mathbb{R}$ mit

$$|v|_1^2 := \frac{1}{4}\left(\sum_{i=1}^{N-1}(v_i - v_{i+1})^2 + \sum_{i=2}^{N}(v_i - v_{i-1})^2 \right),$$

definieren wiederum zwei Normen auf $\mathbb{R}^N_{\Sigma=0}$ und sind daher äquivalent. Definiere nun $\tilde{\sigma}_n = (\tilde{\sigma}_{K_1,n}, \tilde{\sigma}_{K_2,n}, \dots, \tilde{\sigma}_{K_N,n}) \in \mathbb{R}^N_{\Sigma=0}$ durch Korrektur der Lösung $\{\sigma_{K,n}\}_{K\in\mathcal{P}_n}$:

$$\tilde{\sigma}_{K_i,n} = \sigma_{K_i,n} - \frac{1}{N}\sum_{K_i\in\mathcal{P}_n} \sigma_{K_i,n}, \quad i = 1, \dots, N.$$

Wir bemerken analog wie im ersten Fall und aufgrund der Definition von $\tilde{\sigma}_n$ den Zusammenhang

$$|\tilde{\sigma}_n|_1^2 = |\sigma_n|_1^2 = \Sigma_n = \sum_{K\in\mathcal{P}_n} \sigma_{K,n}\tilde{r}_K(\theta_n).$$

Wegen $\sum_{K\in\mathcal{P}_n} \tilde{r}_K(\theta_n) = 0$ und nach Anwendung der Cauchy-Schwarz-Ungleichung gilt weiter, dass

$$\Sigma_n = \sum_{K\in\mathcal{P}_n} \sigma_{K,n}\tilde{r}_K(\theta_n) = \sum_{K\in\mathcal{P}_n} \tilde{\sigma}_{K,n}\tilde{r}_K(\theta_n) \leq |\tilde{\sigma}_n|_2|\tilde{r}_n|_2.$$

Aufgrund der Normäquivalenz von $|.|_1$ und $|.|_2$ existiert wiederum ein $C > 0$, sodass

$$\Sigma_n \leq C|\tilde{\sigma}_n|_1|\tilde{r}_n|_2 = C\sqrt{\Sigma_n}|\tilde{r}_n|_2.$$

Damit haben wir auch in diesem Fall die Stabilität der Lagrange-Multiplikatoren nachgewiesen.

□

Zu einer Lösung u_h bezeichnen wir mit

$$\left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle = \begin{cases} \frac{1}{2}\left(\frac{\partial u_h}{\mathbf{n}_K}\Big|_K - \frac{\partial u_h}{\mathbf{n}_{K'}}\Big|_{K'} \right), & \gamma \in \mathcal{E}_n^I \cap \partial K \cap \partial K', \\ g, & \gamma \in \mathcal{E}_n^N \cap \partial K, \\ \frac{\partial u_h}{\partial \mathbf{n}}, & \gamma \in \mathcal{E}_n^D \cap \partial K. \end{cases}$$

für $K \in \mathcal{P}$ den verallgemeinerten Mittelwert der approximierten Flüsse.

Lemma 3.17. (**Stabilität der Momente erster Ordnung**)
Zu $n \in \mathcal{N}$ sei wie in der Notation von Satz 3.9 eine Familie $\{\mu_{K,n}^\gamma\}$ eindeutiger Momente erster Ordnung gegeben. Sei weiter $\mathcal{P}_n$ und $\mathcal{E}_n$ der zugeordnete Element- und Kantenpatch sowie θ_n die dem Knoten zugeordnete Lagrange-Basisfunktion. Dann existiert ein $C = C(\#\mathcal{P}_n) > 0$ mit

$$\sum_{K\in\mathcal{P}_n} \sum_{\gamma\in\mathcal{E}_n\cap\mathcal{E}_K} \left| \mu_{K,n}^\gamma - \int_\gamma \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \theta_n \mathrm{d}s \right|^2 \leq C \sum_{K\in\mathcal{P}_n} |\tilde{r}_K(\theta_n)|^2 .$$

Beweis. Zunächst erinnern wir uns an die Definition der approximierten Momente. Sei $\gamma \in \mathcal{E}_n$, dann gilt der Zusammenhang

$$\int_\gamma \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \theta_n \mathrm{d}s = \begin{cases} \frac{1}{2}\left(\tilde{\mu}_{K,n}^\gamma - \tilde{\mu}_{K',n}^\gamma\right), & \gamma \in \mathcal{E}_K^I \cap \mathcal{E}_{K'}^I \\ \tilde{\mu}_{K,n}^\gamma = \int_\gamma g\theta_n \mathrm{d}s, & \gamma \in \mathcal{E}_K^N, \\ \tilde{\mu}_{K,n}^\gamma, & \gamma \in \mathcal{E}_K^D. \end{cases}$$

Aufgrund der eindeutigen Darstellung der Momente in Satz 3.9 folgt daher

$$\mu_{K,n}^\gamma - \int_\gamma \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \theta_n \mathrm{d}s = \begin{cases} \frac{1}{2}(\sigma_{K,n} - \sigma_{K',n}), & \gamma \in \mathcal{E}_K^I \cap \mathcal{E}_{K'}^I, \\ 0, & \gamma \in \mathcal{E}_K^N, \\ \sigma_{K,n}, & \gamma \in \mathcal{E}_K^D. \end{cases}$$

Summieren wir über alle Kanten, erhalten wir damit, dass

$$\sum_{K\in\mathcal{P}_n} \sum_{\gamma\in\mathcal{E}_K} \left| \mu_{K,n}^\gamma - \int_\gamma \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \theta_n \mathrm{d}s \right|^2 \leq \frac{1}{4} \sum_{K\,:\,\gamma=\partial K\cap\partial K'} (\sigma_{K,n} - \sigma_{K',n})^2 + \sum_{K\,:\,\gamma=\partial K\cap\Gamma_D} \sigma_{K,n}^2$$

und durch Anwendung von Lemma 3.16 schließlich die Behauptung. □

Das nächste Ziel ist es, ein Stabilitätsresultat der hier konstruierten äquilibrierten Flüsse aufzuzeigen. Dazu führen wir einige Notationen ein. Auf einem Element K definieren wir durch

$$R^\circ|_K = f + \Delta u_h - c u_h \quad \text{in } K$$

das innere (starke) Residuum R° sowie mit

$$R^\partial|_K = \begin{cases} -\frac{1}{2}\left(\frac{\partial u_h}{\partial \mathbf{n}_K}\Big|_K + \frac{\partial u_h}{\partial \mathbf{n}_{K'}}\Big|_{K'}\right), & \text{auf} \quad \partial K \cap \partial K', \\ 0, & \text{auf} \quad \partial K \cap \partial \Gamma_D, \\ g - \frac{\partial u_h}{\partial \mathbf{n}}, & \text{auf} \quad \partial K \cap \partial \Gamma_N \end{cases}$$

das Residuum R^∂ auf dem Rand von K. Zu einer Kante $\gamma \in \mathcal{E}$ bezeichnen wir mit

$$\Pi_p^\gamma \colon L^2(\gamma) \to \operatorname{span}\{\theta_i|_\gamma | \theta_i \in \mathcal{B}_\gamma^{\text{edge}} \cup \mathcal{B}_\gamma^{\text{node}}\}$$

die orthogonale Projektion bezüglich $L^2(\gamma)$, wobei $\#(\mathcal{B}_\gamma^{\text{edge}} \cup \mathcal{B}_\gamma^{\text{node}}) = p+1$. Mit diesen Konstrukten sind wir in der Lage, den folgenden Satz zu formulieren.

Satz 3.18. (**Stabilität der äquilibrierten Flüsse**)
Sei $\mathcal{P}$ eine reguläre Partition von Ω und $\{g_K\}_{K\in\mathcal{P}}$ eine durch Momente rekonstruierte Familie äquilibrierter Flüsse p-ten Grades. Dann existiert für ein Viereck $K \in \mathcal{P}$ ohne gekrümmte Kanten ein $C = C(p) > 0$, sodass

$$\sum_{\gamma\in\mathcal{E}_K\setminus\mathcal{E}^N} h_\gamma \left\| g_K - \Pi_p^\gamma \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right\|_{L^2(\gamma)}^2 \leq C h_K^2 \sum_{K'\in\mathcal{P}(K)} \|R^\circ\|_{L^2(K')}^2 + C h_K \sum_{\gamma'\in\mathcal{E}^I(\mathcal{P}(K))} \left\|R^\partial\right\|_{L^2(\gamma')}^2,$$

wobei h_γ die Länge der Kante γ bezeichnet.

Beweis. Sei $K \in \mathcal{P}$ und $\gamma \in \mathcal{E}_K$ mit $\mathcal{N}(\gamma) = \{n_l, n_r\}$ mit der Kante zugeordneter Basis $\mathcal{B}_\gamma^{\text{node}} \cup \mathcal{B}_\gamma^{\text{edge}}$. Sei $I = \{n_l, n_r, \gamma_1, \dots, \gamma_{p-1}\}$, wie im Kapitel über die Rekonstruktion der Flüsse, die Menge der Indizes der Basisfunktionen. Wir definieren die Momente

$$\overline{\mu}_{K,n}^\gamma = \int_\gamma \Pi_p^\gamma \left(g_K - \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right) \theta_n|_\gamma \mathrm{d}s = \int_\gamma \left(g_K - \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right) \theta_n|_\gamma \mathrm{d}s, \quad n \in I.$$

Sei $\mathcal{B}_\gamma = \left\{\theta_i|_\gamma, i \in I | \theta_i \in \mathcal{B}_\gamma^{\text{node}} \cup \mathcal{B}_\gamma^{\text{edge}}\right\}$ die auf der Kante gegebene Basis, dann gilt per Definition

$$\Pi_p^\gamma \left(g_K - \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right)\Big|_\gamma \in \mathcal{B}_\gamma,$$

das heißt, es existieren $\alpha_i \in \mathbb{R}, i = 1, \dots, p+1$ mit

$$\Pi_p^\gamma \left(g_K - \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right)\Big|_\gamma = \sum_{i=1}^{p+1} \alpha_i \theta_i|_\gamma, \quad \theta_i|_\gamma \in \mathcal{B}_\gamma.$$

Sei M_γ die Massematrix bzgl. $\mathcal{B}_\gamma$, das heißt $(M_\gamma)_{i,j} = (\theta_i, \theta_j)_{L^2(\gamma)}$, $i, j \in I$. Wir erinnern uns an

die Rekonstruktion der Flüsse aus den Momenten und erhalten für

$$\alpha = (\alpha_1, \dots, \alpha_{p+1})^T, \overline{\mu}_K^\gamma = (\overline{\mu}_{K,n_l}^\gamma, \overline{\mu}_{K,n_r}^\gamma, \overline{\mu}_{K,\gamma_1}^\gamma, \dots, \overline{\mu}_{K,\gamma_{p-1}}^\gamma)^T \in \mathbb{R}^{p+1}$$

die folgende Beziehung:

$$\left\| \Pi_p^\gamma \left(g_K - \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right) \right\|_{L^2(\gamma)}^2 = \alpha^T M_\gamma \alpha = (\overline{\mu}_K^\gamma)^T M_\gamma^{-1} \overline{\mu}_K^\gamma.$$

Sei $\lambda_{\max}$ der maximale (von p) abhängige Eigenwert der inversen Referenzmassematrix $\hat{M}_{[0,1]}^{-1} = h_\gamma M_\gamma^{-1}$, wobei h_γ die Länge der Kante γ bezeichnet. Dann gilt wegen der Definition 3.5 der Momente $\{\overline{\mu}_{K,i}^\gamma\}_{i\in I}$, dass

$$(\overline{\mu}_K^\gamma)^T M_\gamma^{-1} \overline{\mu}_K^\gamma \leq \lambda_{\max} h_\gamma^{-1} \sum_{i\in I} (\overline{\mu}_{K,i}^\gamma)^2 = \lambda_{\max} h_\gamma^{-1} \sum_{i\in I} \left(\mu_{K,i}^\gamma - \int_\gamma \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \theta_i|_\gamma \mathrm{d}s \right)^2.$$

Wegen der Definition des verallgemeinerten Mittelwerts, gilt für die rechte Seite $\tilde{r}_K(\theta_n)$ der Lagrange-Multiplikator-Gleichung aus 21 bzgl. der nodalen Basisfunktion $\theta_n \in \mathcal{B}_\gamma^{\text{node}}$, dass

$$\tilde{r}_K(\theta_n) = -R_K(\theta_n) - \int_{\partial K} \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \theta_n \mathrm{d}s.$$

Wir erweitern $\tilde{r}_K$ analog auf Basisfunktionen höherer Ordnung. Wegen der eindeutigen Darstellung, $\mu_{K,\gamma_i}^\gamma = -R_K(\theta_{\gamma_i})$, $\gamma_i \in \mathcal{B}_\gamma^{\text{edge}}$, der Momente höherer Ordnung und dem Lemma (3.17) über die Stabilität der Momente erster Ordnung, können wir weiter abschätzen und erhalten

$$\left\| \Pi_p^\gamma \left(g_K - \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right) \right\|_{L^2(\gamma)}^2 \leq \lambda_{\max} h_\gamma^{-1} \left(\underbrace{\sum_{n=n_l,n_r} C_n \sum_{K'\in\mathcal{P}_n} |\tilde{r}_{K'}(\theta_n)|^2}_{\text{Lemma3.17}} + \sum_{i\in I\setminus\{n_l,n_r\}} |\tilde{r}_K(\theta_i)|^2 \right)$$

für $C_n > 0$. Zu $K' \in \mathcal{P}$ und $\theta_i \in \mathcal{B}_\gamma^{\text{node}} \cup \mathcal{B}_\gamma^{\text{edge}}$ liefert partielle Integration

$$\begin{aligned} \tilde{r}_{K'}(\theta_i) &= \int_{K'} \underbrace{(-\Delta u_h + c u_h + f)}_{=-R^\circ} \theta_i \mathrm{d}x + \int_{\partial K'} \underbrace{\left(\frac{\partial u_h}{\partial \mathbf{n}_{K'}} - \left\langle \frac{\partial u_h}{\partial \mathbf{n}_{K'}} \right\rangle \right)}_{=-R^\partial} \theta_i \mathrm{d}s \\ &= -(R^\circ, \theta_i)_{L^2(K')} - \int_{\partial K'} R^\partial \theta_i \mathrm{d}s. \end{aligned}$$

Weiter liefert die Cauchy-Schwarz-Ungleichung

$$\begin{aligned}|\tilde{r}_{K'}(\theta_i)| &\leq \|R^\circ\|_{L^2(K')} \|\theta_i\|_{L^2(K')} + \sum_{\gamma' \in \mathcal{E}_{K'}} \left\|R^\partial\right\|_{L^2(\gamma')} \|\theta_i\|_{L^2(\gamma')} \\ &\leq \hat{C}(p) \left[h_{K'} \|R^\circ\|_{L^2(K')} + \sum_{\gamma' \in \mathcal{E}_n \cap \mathcal{E}_{K'}} h_{K'}^{1/2} \left\|R^\partial\right\|_{L^2(\gamma')} \right].\end{aligned}$$

In der letzten Abschätzung wurde die Definition von $h_{K'}$ sowie die Abschätzung aus (4) benutzt. Die von p abhängige Konstante $\hat{C}(p)$ wird durch die Normen der Basisfunktionen auf Referenzebene beeinflusst. Dann existieren Konstanten C_1, C_2 mit

$$\begin{aligned}\sum_{n=n_l,n_r} C_n \sum_{K' \in \mathcal{P}_n} |\tilde{r}_{K'}(\theta_n)|^2 &\leq C_1 \left(\sum_{n=n_l,n_r} \sum_{K' \in \mathcal{P}_n} \left[h_{K'}^2 \|R^\circ\|_{L^2(K')}^2 + \sum_{\gamma' \in \mathcal{E}_n \cap \mathcal{E}_{K'}} h_{K'} \left\|R^\partial\right\|_{L^2(\gamma')}^2 \right] \right), \\ \sum_{i \in I \setminus \{n_l,n_r\}} |\tilde{r}_K(\theta_i)|^2 &\leq C_2 \left(\sum_{i \in I \setminus \{n_l,n_r\}} h_K^2 \|R^\circ\|_{L^2(K)}^2 + \sum_{\gamma' \in \mathcal{E}_n \cap \mathcal{E}_K} h_K \left\|R^\partial\right\|_{L^2(\gamma')}^2 \right).\end{aligned}$$

Bis hierhin, haben wir eine Kante betrachtet und wir summieren (mitunter mehrfach) über jedes Element aus $\mathcal{P}_{n_l} \cup \mathcal{P}_{n_r}$. Summieren wir nun über alle Kanten von K, so summieren wir (mitunter mehrfach) über alle Elemente aus $\bigcup_{n \in \mathcal{N}_K} \mathcal{P}_n = \mathcal{P}(K)$. Es existiert daher ein von $\lambda_{\max}, C_1, C_2$ abhängiges $\tilde{C} > 0$ mit

$$\begin{aligned}\sum_{\gamma \in \mathcal{E}_K} h_\gamma \left\| \Pi_p^\gamma \left(g_K - \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right) \right\|_{L^2(\gamma)}^2 &= \sum_{\gamma \in \mathcal{E}_K \setminus \mathcal{E}^N} h_\gamma \left\| g_K - \Pi_p^\gamma \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right\|_{L^2(\gamma)}^2 \\ &\leq \tilde{C} \sum_{K' \in \mathcal{P}(K)} h_{K'}^2 \|R^\circ\|_{L^2(K')}^2 + \tilde{C} \sum_{\gamma' \in \mathcal{E}^I(\mathcal{P}(K))} h_{K'} \left\|R^\partial\right\|_{L^2(\gamma')}^2.\end{aligned}$$

Wegen der Regularität der Partition können wir $h_{K'}$ durch ein Vielfaches von h_K abschätzen und erhalten für ein $C > 0$ schließlich die Abschätzung

$$\sum_{\gamma \in \mathcal{E}_K \setminus \mathcal{E}^N} h_\gamma \left\| g_K - \Pi_p^\gamma \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right\|_{L^2(\gamma)}^2 \leq C h_K^2 \sum_{K' \in \mathcal{P}(K)} \|R^\circ\|_{L^2(K')}^2 + C h_K \sum_{\gamma' \in \mathcal{E}^I(\mathcal{P}(K))} \left\|R^\partial\right\|_{L^2(\gamma')}^2.$$

□

Bemerkung. Wegen der Regularität der Partition können wir h_γ mit h_K vergleichen und es existiert ein $C = C(p) > 0$, sodass

$$h_K \sum_{\gamma \in \mathcal{E}_K \setminus \mathcal{E}^N} \left\| g_K - \Pi_p^\gamma \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right\|_{L^2(\gamma)}^2 \leq C h_K^2 \sum_{K' \in \mathcal{P}(K)} \|R^\circ\|_{L^2(K')}^2 + C h_K \sum_{\gamma' \in \mathcal{E}^I(\mathcal{P}(K))} \left\|R^\partial\right\|_{L^2(\gamma')}^2.$$

Wir benötigen zwei weitere Ungleichungen, die wir an dieser Stelle nur benennen wollen, die Herleitung dieser befindet sich in [1] und [17]. Sei $\Pi_p^{K'}\colon L^2(K') \to P_{K'}^p$ die orthogonale Projektion auf dem lokalen Approximationsraum. Für $K \in \mathcal{P}$ existiert $C = C(p) > 0$ mit

$$h_K^2 \left\| R^\circ \right\|_{L^2(K)}^2 \leq C \left(\left\| e \right\|_K^2 + h_K^2 \left\| f - \Pi_p^K f \right\|_{L^2(K)}^2 \right) \tag{25}$$

und

$$h_K \left\| R^\partial \right\|_{L^2(\partial K)}^2 \leq C \sum_{\substack{K' \in \mathcal{P} \\ \gamma \in \mathcal{E}_{K'} \cap \mathcal{E}_K}} \left\| e \right\|_{K'}^2 + C h_K \sum_{\gamma \in \mathcal{E}_K} \left\| \left\langle \frac{\partial u_h}{\partial \mathbf{n}_{K'}} \right\rangle - \Pi_p^\gamma \left\langle \frac{\partial u_h}{\partial \mathbf{n}_{K'}} \right\rangle \right\|_{L^2(\gamma)}^2 . \tag{26}$$

Wir sind nun in der Lage ein Effizienzresultat des hergeleiteten Fehlerschätzers zu formulieren.

Satz 3.19. (**Effizienz**) Sei $K \in \mathcal{P}$ und $\tilde{e}_K \in H_0^1(K)$ eine Lösung des lokalen Residuenproblems

$$B_K(\tilde{e}_K, v) = R_K(v) + \int_{\partial K} g_K v \mathrm{d}s, \quad \forall v \in H_0^1(K),$$

mit einem durch Momente rekonstruierten approximierten Fluss g_K, welcher der Äquilibrierungseigenschaft p-ter Ordnung genügt. Sei u_h die FEM-Approximation p-ten Grades, dann existiert $C = C(p) > 0$ mit

$$\frac{1}{C} \left\| \tilde{e}_K \right\|^2 \leq \sum_{K' \in \mathcal{P}(K)} \left[\left\| e \right\|_{K'}^2 + h_K^2 \left\| f - \Pi_p^{K'}(f) \right\|_{K'}^2 \right] + \sum_{\gamma \in \mathcal{E}_K \setminus \mathcal{E}^N} h_K \left\| \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle - \Pi_p^\gamma \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right\|_\gamma^2$$

Beweis. Im Folgenden sei C eine generische Konstante. Sei $K \in \mathcal{P}$ und $v \in H_0^1(K)$ und $\tilde{e}_K$ die Lösung des lokalen Problems. Mithilfe partieller Integration und den Definitionen des Randresiduums, des inneren Residuums und des verallgemeintern Mittelwerts erhalten wir

$$\begin{aligned}
\left| B_K(\tilde{e}_K, v) \right| &= \left| \int_K f v \mathrm{d}x - \int_K \nabla u_h \nabla v + c u_h v \mathrm{d}x + \int_{\partial K} g_K v \mathrm{d}s \right| \\
&= \left| \int_K R^\circ v \mathrm{d}x + \int_{\partial K} \left(g_K - \frac{\partial u_h}{\partial \mathbf{n}_K} \right) v \mathrm{d}s \right| \\
&= \left| \int_K R^\circ v \mathrm{d}x + \int_{\partial K} R^\partial v \mathrm{d}s + \int_{\partial K} \left(g_K - \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right) v \mathrm{d}s \right| \\
&\leq \left\| R^\circ \right\|_{L^2(K)} \left\| v \right\|_{L^2(K)} + \left\| R^\partial \right\|_{L^2(\partial K)} \left\| v \right\|_{L^2(\partial K)} + \left\| g_K - \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right\|_{L^2(\partial K)} \left\| v \right\|_{L^2(\partial K)}
\end{aligned}$$

Wegen den Bedingungen zur Äquilibrierung gilt $B(\tilde{e}_K, 1) = 0$ und daher ist

$$B(\tilde{e}_K, v) = B(\tilde{e}_K, v - \overline{v}), \quad \overline{v} = \frac{1}{\mathrm{vol}(K)} \int_K v \mathrm{d}x \in \mathbb{R}.$$

Mit der der Poincaré-Ungleichung erhält man $\|v - \overline{v}\|_{L^2(K)} \leq C h_K \|\nabla v\|_{L^2(K)}$. Zusammen mit der Stetigkeit des Spuroperators (vgl. Satz 2.1) ergibt sich daher

$$\begin{aligned}
\|v - \overline{v}\|^2_{L^2(\partial K)} &\leq C \|v - \overline{v}\|_{L^2(K)} \|v - \overline{v}\|_{H^1(K)} \\
&= C \|v - \overline{v}\|_{L^2(K)} \sqrt{\|v - \overline{v}\|^2_{L^2(K)} + \|\nabla(v - \overline{v})\|^2_{L^2(K)}} \\
&\leq C h_K \|\nabla v\|_{L^2(K)} \sqrt{(C h_K^2 + 1) \|\nabla v\|^2_{L^2(K)}} \\
&\leq C h_K \|\nabla v\|^2_{L^2(K)}.
\end{aligned}$$

Die beiden Abschätzungen zusammen ergeben dann mit der Äquivalenz der H^1-Norm und der Energienorm

$$|B_K(\tilde{e}_K, v)| \leq C \left(h_K \|R^\circ\|_{L^2(K)} + h_K^{1/2} \|R^\partial\|_{L^2(\partial K)} + h_K^{1/2} \left\| g_K - \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right\|_{L^2(\partial K)} \right) |\!|\!| v |\!|\!|_K$$

und für die Wahl $v = \tilde{e}_K$ mit anschließender Quadrierung

$$|\!|\!| \tilde{e} |\!|\!|^2_K \leq C \left(h_K^2 \|R^\circ\|^2_{L^2(K)} + h_K \|R^\partial\|^2_{L^2(\partial K)} + h_K \left\| g_K - \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right\|_{L^2(\partial K)} \right).$$

Weiter erhalten mit der Definition des verallgemeinerten Mittelwerts und der Dreiecksungleichung

$$\begin{aligned}
\left\| g_K - \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right\|_{L^2(\partial K)} &= \left\| g_K - \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right\|_{L^2(\partial K \setminus \Gamma_N)} = \sum_{\gamma \in \mathcal{E}_K \setminus \mathcal{E}^N} \left\| g_K - \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right\|_{L^2(\gamma)} \\
&\leq \sum_{\gamma \in \mathcal{E}_K \setminus \mathcal{E}^N} \left\| g_K - \Pi_p^\gamma \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right\|_{L^2(\gamma)} + \sum_{\gamma \in \mathcal{E}_K} \left\| \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle - \Pi_p^\gamma \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right\|_{L^2(\gamma)}.
\end{aligned}$$

Mithilfe des Stabilitätsresultates 3.18 der äquilibrierten Flüsse erhalten wir

$$\begin{aligned}
|\!|\!| \tilde{e} |\!|\!|^2_K \leq C \Bigg(& h_K^2 \sum_{K' \in \mathcal{P}(K)} \|R^\circ\|^2_{L^2(K')} + h_K \sum_{\gamma' \in \mathcal{E}^I(\mathcal{P}(K))} \|R^\partial\|^2_{L^2(\gamma')} \\
& + h_K \sum_{\gamma \in \mathcal{E}_K} \left\| \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle - \Pi_p^\gamma \left\langle \frac{\partial u_h}{\partial \mathbf{n}_K} \right\rangle \right\|_{L^2(\gamma)} \Bigg).
\end{aligned}$$

Die Anwendung der Ungleichungen in (25) und (26) liefert schließlich die gesuchte Ungleichung.

□

Durch Summierung über alle Elemente erhalten wir also eine Abschätzung des Fehlerschätzers durch den exakten Fehler und Terme höherer Ordnung.

Bemerkung. Ein Fehlerschätzer heißt *robust*, wenn die Effizienzkonstante nicht vom Grad p der Approximation abhängt. Der hier konstruierte Fehlerschätzer ist dementsprechend nicht notwendigerweise robust. Die vom Approximationsgrad abhängige Konstante kann die Qualität des Fehlerschätzers beeinflussen. Ihr Verhalten werden wir im Kapitel der numerischen Experimenten untersuchen.

3.5.4 asympotische Exaktheit

Während der numerischen Experimente fiel auf, dass der Fehlerschätzer für quasi-uniforme h-Verfeinerung durch Rechtecke in einigen Fällen gegen den exakten Fehler konvergierte. Dieses Phänomen wurde jüngst theoretisch erklärt und sei im folgenden Satz hier der Vollständigkeit halber zusammengefasst. Das Resultat geht zurück auf die Arbeit in [16].

Satz 3.20. (**asymptotische Exaktheit**, Lijun Yi , 2012)
Seien $\Omega_0 \subset \Omega_1 \subset \Omega$ Teilmengen von Ω. Die Partition $\mathcal{P}$ von Ω sei zu jeder Verfeinerungsstufe h auf Ω_1 quasi-uniform und bestehend aus Rechtecken. Für $p \in \mathbb{N}$ ungerade und $u \in H^{p+2}(\Omega_1)$ gilt unter den zusätzlichen Annahmen

$$\begin{aligned} \exists C_1(u) \in \mathbb{R}: \quad & \|\!|e|\!\|_{\Omega_0} \geq C_1(u) h^p, \\ \exists C_2(u) \in \mathbb{R}, \epsilon > 0: \quad & \|e\|_{L^2(\Omega_1)} \leq C_2(u) h^{p+\epsilon}, \end{aligned}$$

die asymptotische Exakheit des hier vorgestellten Fehlerschätzers η auf Ω_0, das heißt

$$\lim_{h \to 0} \frac{\eta|_{\Omega_0}}{\|\!|e|\!\|_{\Omega_0}} = 1,$$

wobei $\eta|_{\Omega_0}^2 = \sum_{K \in \mathcal{P}(\Omega_0)} \eta_K^2$.

4 Implementierung des Fehlerschätzers

Wir wollen uns nun der praktischen Seite widmen, konkret der Implementation des residualen impliziten äquilibrierten Fehlerschätzers. Zunächst soll ein genereller Ablauf in der Fehlerabschätzung mithilfe des bereits dargestellten Schätzers festgehalten werden. Der Hauptanteil dieser Arbeit lag insbesondere in der Implementation des Fehlerschätzers innerhalb der `C++` Bibliothek *CONCEPTS*. Eine Übersicht der Klassen mit ihren wichtigsten Methoden soll diesbezüglich einen Überblick geben. Darüber hinaus wird die Funktionsfähigkeit des Codes exemplarisch durch einen Beispiel-Code illustriert. Der Abschluss des Kapitels ist numerischen Experimenten gewidmet.

4.1 Algorithmus

Zur Bestimmung einer Fehlerapproximation mit einem wie hier vorgestellten Fehlerschätzer lässt sich das folgende algorithmische Vorgehen verwenden.

Algorithmus: Fehlerapproximation mit äquilibriertem Fehlerschätzer

I) **Berechnung der FEM-Lösung u_h p-ter Ordnung zu gegebenen rechten Seiten f und g.**

II) **Berechnung der Momente $\{\mu_{K,n}^{\gamma}\}$ mit u_h und den Daten f und g.**

II.a) Berechnung der lokalen Residuen $R_K(.)$ bzgl. der Kanten-und Knotenbasisfunktionen für jedes $K \in \mathcal{P}$.

II.b) Berechnung Momente erster Ordnung:

- Bestimmung des Kanten-Patches $\mathcal{E}_n$ und des Patches $\mathcal{P}_n$ für jeden Knoten $n \in \mathcal{N}$
- Berechnung der approximierten Momente $\{\tilde{\mu}_{K,n}^{\gamma}\}$ bzw. deren Differenzen, bzgl. der nodalen Basisfunktionen auf jedem Viereck K und jeder Kante $\gamma \in \mathcal{E}_K$,
- Bestimmung der Momente erster Ordnung aus $\{\tilde{\mu}_{K,n}^{\gamma}\}$ und $\{\sigma_{K,n}\}$. Die Lagrangemultiplikatoren $\{\sigma_{K,n}\}$ erhalten wir durch Lösen der Gleichungssysteme mit topologischen Matrizen mithilfe ihrer Inversen. Dazu Aufbau der rechten Seite durch Modifikation der inneren Residuen bzgl. nodaler Basis mit approximierten Momenten.

II.c) Berechnung Momente höherer Ordnung:

- Falls $p > 1$, Zuordnung der lokalen Residuen bzgl. der Kantenbasisfunktionen, zur Bestimmung der Momente höherer Ordnung.

III) **Rekonstruktion der äquilibrierten Flüsse $\{g_K\}$ aus $\{\mu_{K,n}^{\gamma}\}$ und g.**

III.a) Berechnung der Massematrizen bzgl. der Kanten und Berechnung der Koeffizienten.

IV) **Berechnung einer Fehlerabschätzung**

IV.a) Lösen der lokalen Fehlerprobleme mithilfe von $\{g_K\}$.

IV.b) Berechnung der lokalen Fehlerschätzungen durch Berechnung der Energienorm der lokalen FEM-Approximation.

IV.c) Summation der einzelnen Fehlerschätzungen zur Bestimmung der globalen Fehlerabschätzung.

4.2 Implementation in CONCEPTS

Durch die Umsetzung des Algorithmus wurde die `C++` Bibliothek `CONCEPTS` durch weitere Klassen ergänzt. An dieser Stelle werden wir nicht genauer auf die Funktionsfähigkeit von `CONCEPTS` eingehen und verweisen an dieser Stelle auf die Dissertationen [11] und [12]. Vielmehr sollen die durch die Implementation entstandenen wichtigen Hauptklassen in ihrer Struktur beschrieben werden:

- Zur Berechnung der lokalen Residuen dienen die Klassen `InnerResidual` sowie die Linearformen `LinInnerProd_0` und `LinInnerProd_1`.
- Die Verwaltung der Kanten-Patches und Element-Patches geschieht in der Klasse `VtxToPatchMaps`.
- Die approximierten Momente werden durch die Klasse `ApproxMoments` gehalten.
- Die Berechnung der Momente erster und höherer Ordnung geschieht in der Klasse `Moments`.
- Schließlich werden die äquilibrierten Flüsse innerhalb der Klasse `Fluxes` gehalten.

Neben diesen Hauptklassen wurden weitere Klassen, wie zum Beispiel `hp2D::NeumannTraceSpace` und `SingleElementSpace` sowie ein neuer Operator innerhalb der `hp1D::Value` für Elemente aus dem NeumannTraceSpace implementiert. Weiter wurde in Anlehnung an die Klasse `hp1D::Riesz` die Klasse `hp1D::LinInnerProd_0` implementiert, in der nur bezüglich der Basisfunktionen ersten Grades integriert wird. Diese Klassen finden in der Konstruktion der approximierten Flüsse ihre Hauptverwendung. Weiterhin wurde eine Mehrzahl an Unterklassen implementiert, auf welchen die Hauptklassen beruhen. Auf all diese Nebenklassen wollen wir

an dieser Stelle nicht genauer eingehen, da sie zunächst nur der Funktionalität innerhalb der Hauptklassen dienen. Wir werden einige der Vollständigkeit halber in den Übersichtsabbildungen der Klassen erwähnen, um die Abhängigkeiten zu verdeutlichen.

Bemerkung. Die Berechnung der Differenz approximierter Momente wurde komplett mithilfe des `hp2D::NeumannTraceSpace` im Jahr 2013/2014 ausgelagert und vereinfacht. Klassen, wie zum Beispiel `LinInnerProd_0` und `LinInnerProd_1` wurden beispielsweise durch Verallgemeinerung der Klassen `hpND::Riesz` und `hpND::GradLinearForm` redundant.

Es folgt eine Beschreibung der Hauptklassen.

Klassen zur Berechnung der lokalen Residuen

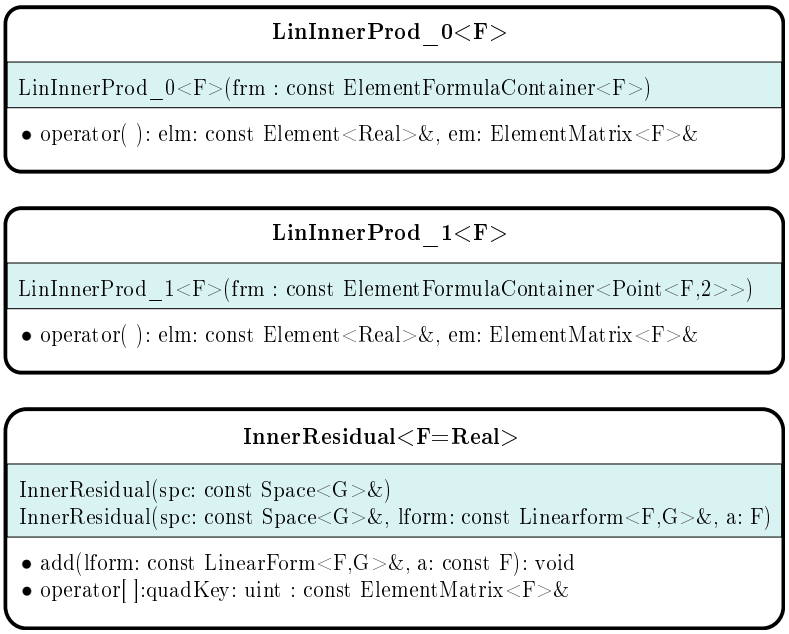

Abbildung 3: Übersicht wichtiger Klassen zur Berechnung der lokalen inneren Residuen.

- Die an die Linearform `hp2D::Riesz` angelehnte Klasse `Linearform_0` berechnet durch die `operator()`-Methode die L^2-Produkte der Inputfunktion frm mit allen kanten-und knotenbasierten Basisfunktionen auf einem Element $K =$ `elm` und speichert diese in einer `ElementMatrix<F>`. Sie dient der Berechnung von

$$\int_K f\theta_i \mathrm{d}x \quad \text{und} \quad \int_K u_h\theta_i \mathrm{d}x \quad \text{für } \theta_i \in \mathcal{B}_{\partial K}.$$

- Die an die Linearform `hp2D::GradLinearForm` angelehnte Klasse `Linearform_1` berechnet durch die `operator()`-Methode die L^2-Produkte der vektorwertigen Inputfunktion frm und der Ableitungen aller kanten-und knotenbasierten Basisfunktionen auf einem

Element $K =$ `elm` und speichert diese in einer `ElementMatrix<F>`. Sie dient der Berechnung von

$$\int_K \nabla u_h \cdot \nabla \theta_i \mathrm{d}x \quad \text{für } \theta_i \in \mathcal{B}_{\partial K}.$$

- Die Klasse `InnerResidual` verwaltet die lokalen Residuen $\{R_K(.)\}_{K\in\mathcal{P}}$. Diese werden durch die Linearform-Klassen `Linearform_0` und `Linearform_1` sowie einer `add`-Routine aufgebaut. Die Klasse selbst ruft die jeweilige `operator()`-Methode der Linearformen zur Berechnung der Residuen auf.

Klassen zur Verwaltung der Patches

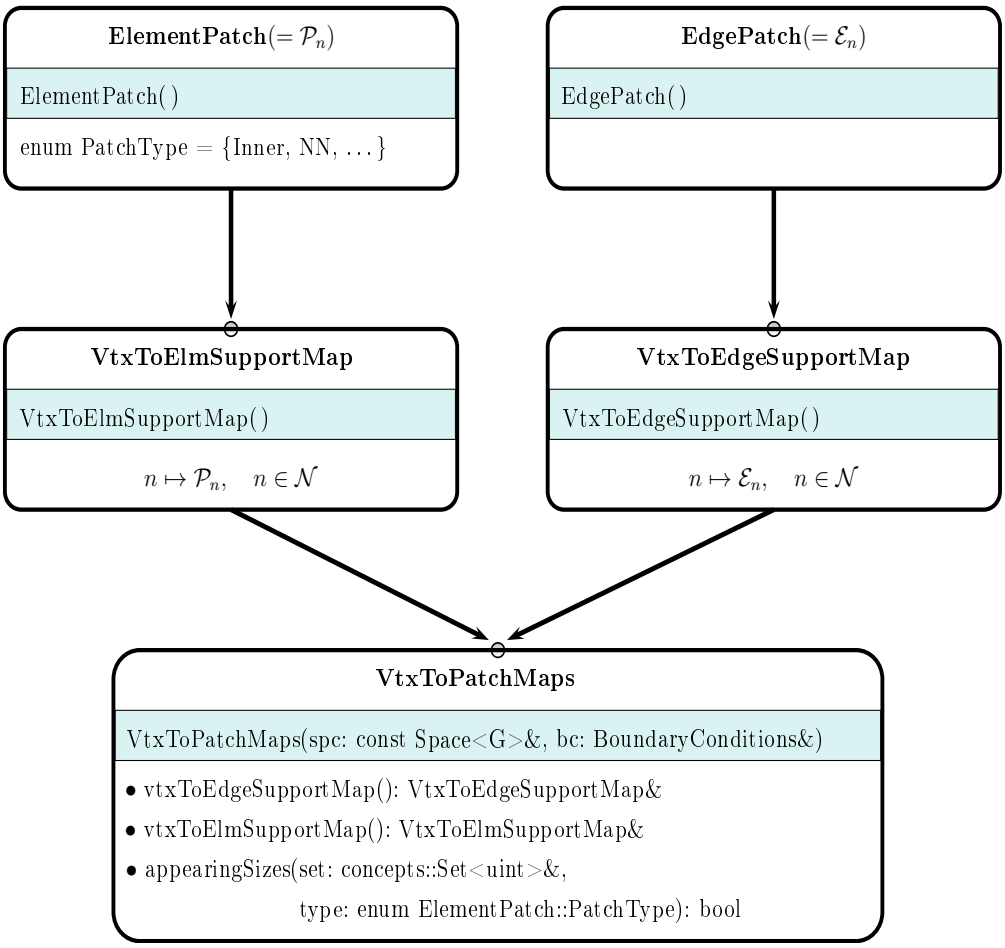

Abbildung 4: Übersicht der Klassen zur Bestimmung der Element- und Kantenpatches. Es werden nur ausgewählte Methoden und Strukturen abgebildet.

Die Hauptklasse `VtxToPatchMaps` hält die Element- und Kantenpatches in Form der Klassen `VtxToElmSupportMap` und `VtxToEdgeSupportMap` und sorgt für eine richtige Sortierung der Patches, wie in der Theorie vorgestellt. Mithilfe der übergebenen `BoundaryConditions` werden

die Patches aufgebaut und durch ein Objekt vom typ `PatchType` unterschieden. Die Klassen `ElmPatch` und `EdgePatch` sind Spezialisierungen der Klasse `std::vector<uint>`. Sie stellen benötigte Methoden zur Sortierung zur Verfügung.

Klasse zur Berechnung der approximierten Momente

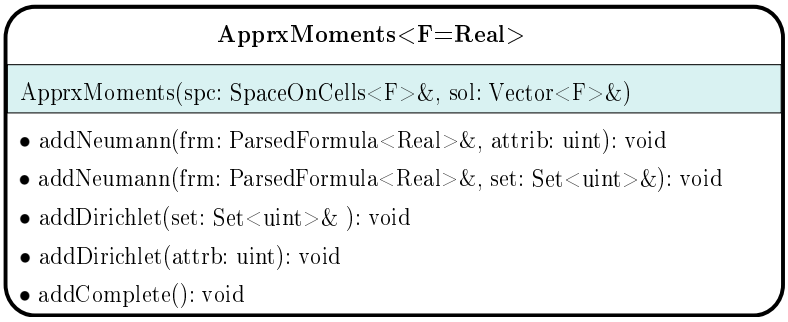

Abbildung 5: Übersicht der Klassen ApprxMoments mit seinen Methoden zur Berechnung der approximierten Momente aus der Lösung u_h und den Neumann-Daten.

Die Klasse `ApprxMoments<F>` verwaltet die approximierten Momente und ist abgeleitet von der Klasse `MomentsBase<F>`. Die Methoden `addNeumann` und `addDirichlet` dienen der Bereitstellung weiterer Informationen, wie den Attributen von Dirichlet- und Neumannkanten sowie dazugehöriger Neumann-Daten. Kernstück dieser Klasse ist die Methode `addComplete()`, die nach Abhandlung der `add`-Routinen aufgerufen wird und aus den gegebenen Informationen die approximierten Momente berechnet. Aus dem Lösungsvektor werden innerhalb der Klasse `ApprxMoments` auf einem Element $K \in \mathcal{P}$ die Neumannspuren $(\mathbf{n}_K \cdot \nabla u_h)|_{\partial K}$ der Lösung berechnet. Letzteres verwaltet die Klasse `NeumannTraceSpace`.

Klasse zur Berechnung der Momente

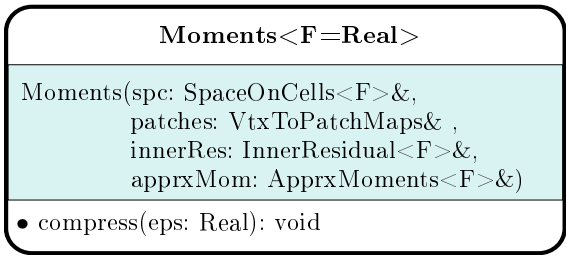

Abbildung 6: Übersicht der Klasse Moments, welche die Momente $\{\mu^{\gamma}_{K,n}\}$ erster und höherer Ordnung hält.

Die von `MomentsBase<F>` abgeleitete Klasse `Moments<F>` berechnet die Momente erster und höherer Ordnung aus den inneren Residuen sowie den approximierten Momenten. Die Klasse wird durch private Routinen verschiedener Funktionalität gesteuert, beispielsweise den Aufbau

der Topologie-Matrizen und das Lösen der dazugehörgen Gleichungssysteme, um die Lagrange-Multiplikatoren $\{\sigma_{K,n}\}$ zu berechnen. Die Inversen der Topologie-Matrizen werden mithilfe der `appearingSizes`-Routine des `patches`, mit den Formeln aus Satz 3.13, für die entsprechenden Patchgrößen vorberechnet.

Klasse der äquilibrierten Flüsse

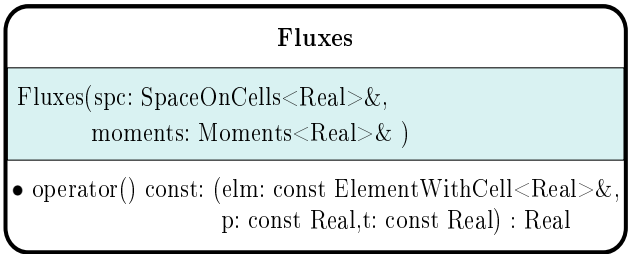

Abbildung 7: Übersicht der Klasse Fluxes.

Die von der Klasse `ElementFormula<Real, Real>` abgeleitete Klasse `Fluxes` hält die äquilibrierten Flüsse $\{g_K\}_{K\in\mathcal{P}}$ auf inneren Kanten. Der Konstruktor dieser Klasse erwartet einen kantenbasierten Raum, zum Beispiel einen `hp2D::TraceSpace`. Die Flüsse werden aus den übergebenen Momenten auf nicht gekrümmten Kanten berechnet. Dafür werden die benötigten Inversen der Kanten-Massematrizen auf Referenzebene einmal vorberechnet und innerhalb eines `operator()`-Aufrufes mit der Länge einer übergebenen Kante skaliert.

Anwendung des Fehlerschätzers

Bei der Integrierung des Fehlerschätzers in ein eigenes Programm innerhalb von `CONCEPTS` sollten einige Dinge beachtet werden:

- Löst man eine partielle Differentialgleichung mit $c = 0$, so sollten die lokalen Probleme auf einem Viereck $K \in \mathcal{P}$ mit $\partial K \cap \Gamma_D = \emptyset$ auf dem zu $H^1(K) \setminus \mathbb{R}$ isomorphen Raum

$$\left\{ v \in H^1(K) | \int_K v \mathrm{d}x = 0 \right\}$$

  gelöst werden. Für den Fall $c > 0$ entfällt diese Schwierigkeit und der globale und lokale Lösungsansatz funktioniert analog.

- Die Klasse `Fluxes` ist zum jetzigen Zeitpunkt nur für innere Kanten vorgesehen. Besitzt das lokale Problem einen Dirichlet-Rand, so sollte dies mithilfe von `BoundaryConditions` beim Aufbau der lokalen Räume kontrolliert werden. Besitzt das lokale Problem Neumannkanten, so muss die rechte Seite des lokalen Gleichungssystems mithilfe von Spurräumen und der exakten rechten Seite g aufgebaut werden.

Für einen einfach Fall mit $c > 0$ befindet sich in Anhang B ein Code-Beispiel, welches die Berechnung der Flüsse illustriert. Mit diesen können nun die lokalen Probleme formuliert und berechnet werden, wir erhalten lokale Fehlerapproximationen.

4.3 Numerik

Der implementierte Fehlerschätzer auf Viereckgittern soll nun numerisch getestet werden. Ziel ist eine Untersuchung der Qualität des Fehlerschätzers und insbesondere der Effizienzkonstante. Um eine garantiere obere Schranke und damit einen zuverlässigen Fehlerschätzer zu erhalten, müssen die lokalen Probleme exakt gelöst werden. Diese können in der Regel wiederum nur approximiert werden. Eine Betrachtung varierender lokaler Approximationsordnungen soll hierbei Aufschluss geben, ab wann der Fehlerschätzer, bei nicht exaktem Lösen, dennoch zuverlässig sein könnte. Anhand von drei Randwertproblemen werden wir das Verhalten des Fehlerschätzers nun folgend studieren.

4.3.1 Reguläres Problem

Die hergeleiteten Eigenschaften des Fehlerschätzers sollen an einem regulären Problem numerisch demonstriert werden. Wir betrachten das Randwertproblem

$$\begin{cases} -\Delta u + u = (2\pi^2+1)\sin(\pi x)\sin(\pi y) & \text{in} \quad \Omega = [0,1]^2, \\ u = 0 & \text{auf} \quad \Gamma_D = \partial\Omega \end{cases}$$

mit exakter Lösung $u \in \mathcal{C}^\infty(\Omega)$ mit $u(x,y) = \sin(\pi x)(\pi y)$. Wegen $c = 1$ sind die lokalen Problem stets eindeutig lösbar auf den zugeordneten Unterräumen. Das Gebiet Ω wird mit einem uniformen Gitter aus Quadraten zerlegt. Da $u \in H^k(\Omega)$ für alle $k \in \mathbb{N}$, erwarten wir aufgrund von Satz 3.20 asymptotische Exaktheit des Fehlerschätzers für eine ungerade Ordnung der Approximation. In Abbildung 8 beobachten wir, dass der Fehlerschätzer für eine gerade

	$\#\mathcal{P}$					
	16	64	256	1024	4096	16384
ι	1.001835	1.000566	1.000150	1.000039	1.000010	1.000002

Tabelle 1: Demonstration der asymptotischen Exaktheit bei h-Verfeinerung mit festem globalen Polynomgrad $p = 1$ und lokalem Polynomgrad $p_{loc} = 3$.

Approximationsordnung $p = 2, 4$ ebenfalls asymptotisch gegen $\|\!|e|\!\|$ konvergiert. Werden die lokalen Probleme mit einer Ordnung von $p+1$ gelöst beobachtet man Effektivitätsindizes kleiner 1. Dieses Verhalten wird bei weiterer Erhöhung des lokalen Polynomgrades p_{loc} nicht beobachet. Daher wird an dieser Stelle eine lokale Approximationordnung $p_{loc} = p + 2$ empfohlen.

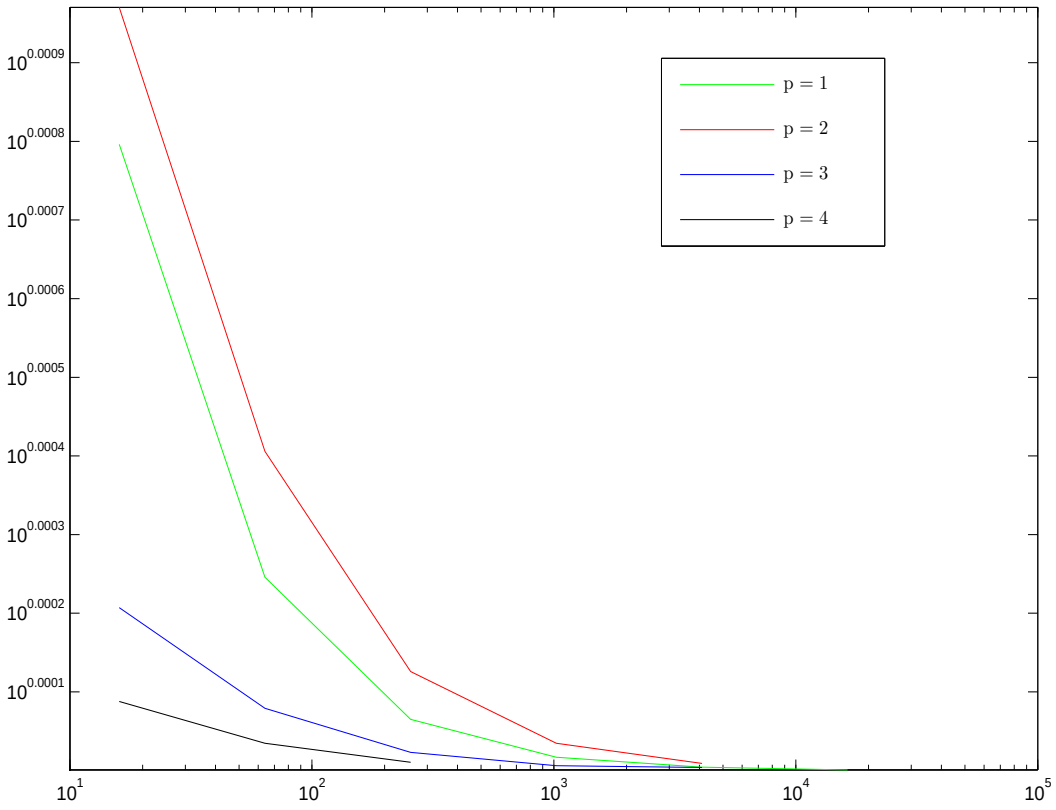

Abbildung 8: Darstellung des Effektivitätsindizes in Bezug auf die Anzahl der Freiheitsgrade bei varierendem globalen Approximationsgrad p zu verschiedenen Verfeinerungsstufen. Die lokalen Probleme wurden mit der Ordnung $p_{loc} = p + 2$ gelöst.

4.3.2 Oszillierendes Problem

An dieser Stelle untersuchen wir das Verhalten des Schätzers bezüglich der Differentialgleichung:

$$\begin{cases} -\Delta u + u = (2k^2\pi + 1)\sin(k\pi x)\sin(k\pi y) & \text{in} \quad \Omega = [0,1]^2, \\ u = 0 & \text{auf} \quad \Gamma_D = \partial\Omega \end{cases}$$

für $k = 10, 50, 100$ mit exakter Lösung $u(x,y) = \sin(k\pi x)\sin(k\pi y)$. Aufgrund der (von k ab-

p	p_{loc}	ι	$\|\|\|e\|\|\|$
2	3	1.003898	$7.954870e-01$
3	5	1.001434	$6.579774e-02$
4	6	1.000616	$4.067477e-03$
5	7	1.000300	$2.007187e-04$

Tabelle 2: Darstellung der Wahl des lokalen Polynomgrades p_{loc} für eine Oszillationsstärke $k = 10$ bei uniformer Verfeinerung mit $\#\mathcal{P} = 1024$, sodass der Fehlerschätzer eine obere Schranke garantiert.

hängigen) Oszillation der Lösung erwarten wir erst bei einem höheren Verfeinerungs- oder Approximationsgrad, dass $\|\!|e|\!\| \leq 1$. Weiterhin sind lokale Approximationen aufgrund der Oszillation als grob einzustufen. Die Tabellen 2-4 illustrieren das Verhalten des Fehlerschätzers in den drei Fällen. Wir beobachten, dass um so stärker die Oszillation ist, um so höher muss der

p	p_{loc}	ι	$\|\!\|e\|\!\|$
2	4	1.009997	$7.583280e+01$
3	5	1.033168	$3.380772e+01$
4	6	1.010843	$1.069408e+01$
5	7	1.005381	$2.675344e+00$

Tabelle 3: Darstellung der Wahl des lokalen Polynomgrades p_{loc} für eine Oszillationsstärke $k = 50$ bei uniformer Verfeinerung mit $\#\mathcal{P} = 1024$, sodass der Fehlerschätzer eine obere Schranke garantiert.

lokale Approximationsgrad gewählt werden. Zur Sicherstellung einer oberen Schranke an den exakten Fehler genügt die Wahl $p_{loc} = p + 3$ in fast allen Fällen. Für $k \leq 50$ sichert ein lokaler Approximationsgrad von $p_{loc} = p + 2$ die Zuverlässigkeit des Fehlerschätzers. Da die Lösung

p	p_{loc}	ι	$\|\!\|e\|\!\|$
2	6	1.059110	$2.220853e+02$
3	6	1.031641	$2.169203e+02$
4	7	1.028913	$1.640779e+02$
5	7	1.014292	$9.606787e+01$

Tabelle 4: Darstellung der Wahl des lokalen Polynomgrades p_{loc} für eine Oszillationsstärke $k = 100$ bei uniformer Verfeinerung mit $\#\mathcal{P} = 1024$, sodass der Fehlerschätzer eine obere Schranke garantiert.

für alle $k \in \mathbb{N}$ glatt ist, erwarten wir in diesem Beispiel ebenfalls für einen ungeraden Approximationsgrad asymptotische Exaktheit. Die Effizienzkonstante bewirkt in diesem Beispiel keine sichtbare Minderung der Qualität des Fehlerschätzers. Für eine Approximationsordnung mit $\mathrm{mod}(p, 2) = 0$ beobachten wir wiederum asymptotische Exaktheit.

4.3.3 Nicht-Reguläres Problem

Wir betrachten die partielle Differentialgleichung

$$\left\{ \begin{array}{rcl} -\Delta u &=& 1 \quad \text{in} \quad \Omega \quad = \quad [-1,1]^2 \setminus [-1,0]^2, \\ u &=& 0 \quad \text{auf} \quad \Gamma_D \quad = \quad \partial\Omega \end{array} \right.$$

auf dem L-förmigen Gebiet Ω. Die Lösung des Problems ist unbekannt und weist im Nullpunkt eine Singularität auf. Der Fehlerschätzer wird nun bezüglich einer h-Verfeinerung auf uniformen und quasiuniformen Gittern sowie einer p-Verfeinerung genauer betrachtet. Abschließend wird eine uniforme h und p Verfeinerung betrachtet. Die Qualität des Fehlerschätzers wollen wir mit dem globalen Effektivitätsindizes ι messen, dafür benötigen wir den exakten Fehler. In Ermangelung an Kenntnis dieses Fehlers nutzen wir eine Referenzlösung $\overline{u} \approx u$ mit $\|\overline{u}\| = 0.462683262847194$. Sie ist durch FEM-Approximation auf einem $(\beta = 4, N = 30)$-graduierten Gitter [18] entstanden mit 2700 Elementen und Polynomgrad 5. Als Referenzfehler betrachten wir dann $\|\overline{e}\|^2 = \|\overline{u}\|^2 - \|u_h\|^2$.

Uniformes Gitter: Wir betrachten zunächst eine uniforme Verfeinerung des Gebietes, ausgehend von einer Zerlegung in 3 Quadrate. Approximiert wird durch Elemente erster Ordnung. Das Verhalten des Fehlerschätzers wird in Abbildung 9 illustriert.

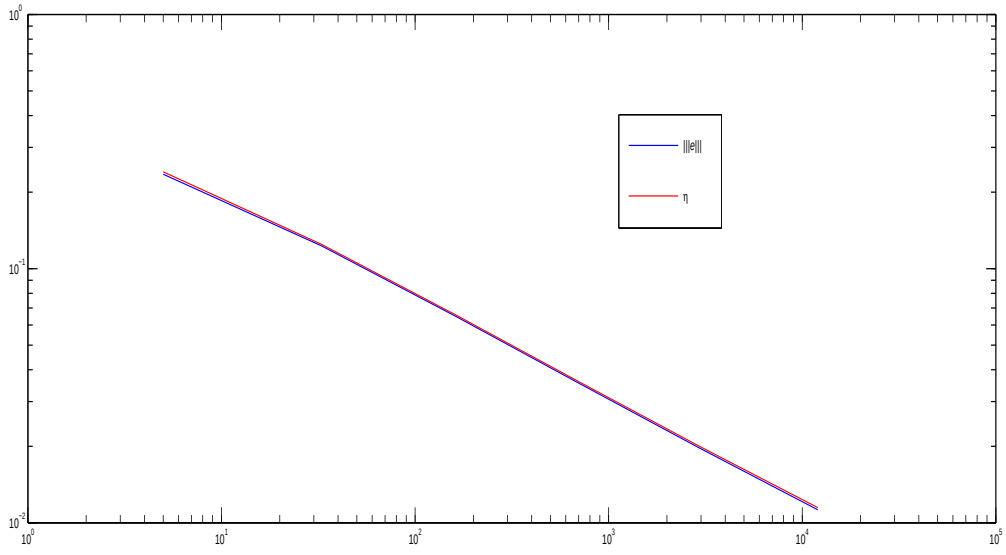

Abbildung 9: Darstellung des Fehlerschätzers (rot) und des Referenzfehlers (blau) in Abhängigkeit zur Anzahl der Freiheitsgrade. Das globale Problem wurde mit $p = 1$ und die lokalen Probleme wurden mit $p_{loc} = 3$ gelöst.

Der Fehlerschätzer liefert in diesem Fall einer sehr gute Annäherung an den exakten Fehler. Der Referenzfehler von $\|\overline{e}\| \approx 0.019678$ wird mit 2945 Freiheiheitsgraden erreicht, der Effektivitätsindex liegt hier bei ca. 1.0167. Es werden Effektivitätsindizes nahe der 1 beobachtet.

Da die Effektitätsindizes um den Wert 1.0167 schwanken, lässt wird hierbei nicht von einer asymptotisch exakten Verhalten ausgegangen. Dies stellt keinen Widerspruch zur Theorie dar, da diese nur für reguläre Lösungen aus $H_0^3(\Omega)$ in diesem Fall eine Aussage treffen würde, aber $u \notin H_0^2(\Omega)$.

Graduiertes Gitter: Das Gebiet wird nun durch ein graduiertes (β, N)-Gitter [18] mit $\beta = 2$ zur Singularität verfeinert. In Tabelle 5 beobachten wir Effektivitätsindizes nahe der 1 und für $p = 1$ ein monotones Verhalten gegen 1. Für höhere Approximationsgrade stellen wir fest, dass die Indizes sogar kleiner als Eins sind. Dies wird durch Erhöhung des lokalen Approximationsgrades um 1 behoben.

	$\#\mathcal{P}$				
p/p_{loc}	12	48	192	768	3072
1/2	1.032825	1.013097	1.005015	1.001863	1.000720
1/3	1.043508	1.020087	1.008415	1.003478	1.001535
	0.2979760	0.1625889	0.08383603	0.04234129	0.02123923
2/3	1.006947	1.007997	1.004500	0.998795	0.9935783
2/4	1.117896	1.097466	1.100856	1.112982	1.128084
	0.04510408	0.01833349	0.006505027	0.002252606	0.0008004975
3/4	0.9555775	0.9525579	0.9524379	0.9528503	0.9547542
3/5	1.141104	1.147761	1.159896	1.168598	1.175287
	0.01815304	0.006937769	0.002640900	0.001021102	$3.990802e-04$

Tabelle 5: Effektivitätsindizes zu verschiedenen Verfeinerungsstufen des L-Gebietes mit einer graduierten Gitterstrategie mit $\beta = 2$. Die FEM-Lösung ist von der Ordnung p. Die lokalen Fehlerprobleme werden mit einer erneuten FEM-Approximation der Ordnung p_{loc} gelöst. In ■ wird der absolute Fehlerwert der globalen Approximation dargestellt.

Zu fester Verfeinerungsvorgabe soll nun die Effizienz in Abhängigkeit der globalen Approximationsordnung p untersucht werden. Die Tabelle 6 zeigt, dass im untersuchten Szenario die Effizienzkonstante nicht wesentlich durch den varierenden Polynomgrad beeinflusst wird. Auch in diesem Fall wird ein lokalen Approximationsgrad $p_{loc} = p + 2$ empfohlen.

p	p_{loc}	$\|\|\|\bar{e}\|\|\|$	η	ι
1	3	4.234129e-02	4.248856e-02	1.003478
2	4	2.252606e-03	2.507110e-03	1.112982
3	5	1.021102e-03	1.193257e-03	1.168598
4	6	7.120689e-04	8.396069e-04	1.179109
5	7	5.447527e-04	6.393769e-04	1.173701
6	8	4.373111e-04	5.122979e-04	1.171472
7	9	3.624996e-04	4.232809e-04	1.167673
8	10	3.075955e-04	3.583318e-04	1.164945
9	11	2.657200e-04	3.088423e-04	1.162285
10	12	2.328146e-04	2.701806e-04	1.160497
11	13	2.063320e-04	2.391839e-04	1.159219

Tabelle 6: Verhalten des Fehlerschätzers bei p-Adaptivität auf einem graduierten Gitter mit $(N, \beta) = (16, 2)$ mit $\#\mathcal{P} = 768$.

5 Fazit und Ausblick

Im Ergebnis dieser Arbeit wurde der äquilibrierte Fehlerschätzer nach [1] für reguläre Gitter implementiert. Auf Viereckgittern überschätzt dieser den exakten Fehler bei geeigneter Wahl des lokalen Approxmiationsgrades nur geringfügig. In den Beispielproblemen wurde kein signifikanter Einfluss der Approximationsordnung auf die Effizienzkonstante beobachtet. Im Vergleich zur Implementation des Fehlerschätzers auf Dreiecksgittern und den damit verbundenen numerischen Experimenten in [3], erzeugt der Fehlerschätzer im Beispiel des nicht regulären Problems auf Viereckgittern besserere Resultate. Für die Wahl des lokalen Approximationsgrades wird aufgrund der Beobachtungen in den numerischen Beispielen $p_{loc} = p+2$ empfohlen. Der Fehlerschätzer wurde im Rahmen dieser Arbeit auf quasi-uniformen Gittern bestehend aus Rechtecken untersucht. Die Theorie bestätigt in diesem Fall die Qualität des Fehlerschätzers in seinem asymptotischen Verhalten für reguläre Problemstellungen.
Die hier vorgestellte Theorie überträgt sich analog auf den Fall von Dreiecksgittern oder hybriden Gittern und lässt sich auf Gitter mit hängenden Knoten durch die Einführung sogenannter *Makro-Elemente* erweitern. Die topologischen Matrizen werden in diesem Fall verallgemeinert und werden durch Analyse sogenannter *Makro-Patches* aufgebaut. Die Theorie hierfür wird ebenfalls in [1],[3] und [2] bereitgestellt. Die Makro-Patches entsprechen wiederum dem Träger der nodalen Lagrange-Basisfunktionen. In `CONCEPTS` wird dieser im Allgemeinen größer gewählt wodurch zunächst eine genauere Betrachtung der Theorie für hängende Knoten oder eine Anpassung der Prolongation (vgl. *S-Matrizen* [11]) von Nöten ist. Weiterhin kann der Fehlerschätzer durch Einführung weiterer Linearformen bzgl. kanten- und knotenbasierter Basisfunktionen oder durch Verallgemeinerung von c zu einer positiv-wertigen Funktion auf eine größere Problemklasse erweitert werden. Ein allgemeiner theoretischer Zugang dafür befindet sich in [4]. Desweiteren könnte durch die Berechnung der Momente mit Minimierungansatz aus Satz (3.10) ein Fehlerschätzer besserer Qualität gewonnen werden, auch wenn dies theoretisch nicht belegt

ist. Die numerischen Experimente in [9] zeigen hierbei mitunter eine bessere Resultate eines solchen Fehlerschätzers auf Dreiecksgittern. Hierfür ist nur eine Anpassung der topologischen Matrizen von Nöten. Weiter kann der vorgestellte Fehlerschätzer für gekrümmte Kannten implementiert werden. Dafür wäre nur eine Modifikation in der Klasse `Fluxes` von Nöten durch Verallgemeinerung der approximierten Flüsse mithilfe der Parametrisierung des zugeordneten Elementes. Die Implementation weiterer Fehlerschätzer in `CONCEPTS`, zum Beispiel mit Äquilibrierungstechniken durch *Raviart-Thomas*-Elemente oder mit *hierarchischen* Ansätzen würde eine Vergleichbarkeit zwischen Fehlerschätzern innerhalb von `CONCEPTS` ermöglichen und zum Zwecke der Anwendung von Interesse sein.

Literatur

[1] M. Ainsworth and J.T. Oden, *A posteriori error estimation in finite elemnt analysis*, Pure and Applied Mathematics, Wiley-Interscience , New York, 2000

[2] M. Ainsworth and L. Demkowicz and C.-W. Kim *Analysis of the equilibrated residual method for a posteriori error estimation on meshes with hanging nodes*, Computer methods in applied mechanics and engineering 196.37 (2007): 3493-3507

[3] M. Ainsworth and J.T. Oden, *A Posteriori Error Estimator for Second Order Elliptic Systems Part 2. An Optimal Order Process for Calculating Self-Equilibrating Fluxes*, Computers Math. Appl. Vol. 26, No. 9, pp. 75-87, 1993

[4] M. Ainsworth and J.T. Oden *A Posteriori Error Estimator for Second Order Elliptic Systems Part 1. Theoretical Foundations and a Posteriori Error Analysis* , Computers Math. Applic. Vol. 25, No. 2, pp. 101 – 113, 1993

[5] R.E. Bank and A. Weiser, *Some a posteriori error estimators for elliptic partial differential equations*, Math. Comp. 44, 283 (1985).

[6] J. Wloka, *Partial differential equations* Cambridge University Press, Cambridge, UK, 1987.

[7] H.W. Alt, *Lineare Funktionalanalysis: Eine anwendungsorientierte Einführung*, Vol. 5, Springer, 2006.

[8] C. Merdon, *A posteriori Fehlerschätzer für elliptische partielle Differentialgleichungen*, HU Berlin, Diplomarbeit, 2009

[9] Xiaoliang Wan, *Some improvements to the flux-type a posteriori error estimators*, Division of Applied Mathematics, Brown University, Providence, RI 02912, USA, 2007

[10] R.M. Gray *Toeplitz and Circulant Matrices: A review*, Department of Electrical Engineering, Stanford University, USA

[11] P. Frauenfelder, *hp-Finite Element Methods on Anisotropically, Locally Refined Meshes in Three Dimensions with Stochastic Data*, Doktorarbeit, ETH Zürich, Schweiz 2004

[12] K. Schmidt, *High-order numerical modelling of highly conductive thin sheets*, Doktorarbeit, ETH Zürich, Schweiz, 2008

[13] K. Schmidt *Numerics of Partial Differential Equations* Vorlesungsmanuskript, SS 2011.

[14] P. Deuflhard and M. Weiser *Numerische Mathematik 3. Adaptive Lösung partieller Differentialgleichungen* Vol. 3. Walter de Gruyter, 2011.

[15] R. Luce and B.I. Wohlmuth. *A local a posteori estimator based on equilibrated fluxes*, 2004

[16] Lijun Yi, *On the asymptotic exactness of error estimators based on the equilibrated residual method for quadrilateral finite elements*, Department of Mathematics, Shanghai Normal University, China, 2012

[17] R. Verfurth, *A review of a posteriori error estimation and adaptive mesh refinement techniques*, Wiley-Teubner, 1996

[18] C. Großmann and H.-G. Roos. *Numerische Behandlung partieller Differentialgleichungen*, Vol 3, Teubner, 2005.

[19] M. Aigner and G.M. Ziegler, *Das BUCH der Beweise*, Springer, 2009.

Anhang A

Beweis von Satz (24) über die Inversen der topologischen Matrizen.

(i) Wir definieren

$$f(j) = \frac{1}{N^3} \cdot \frac{N^4 - 6(j-1)N^3 + (6(j-1)^2 - 1)N^2 + 12N}{6}, \quad j = 1 \dots N.$$

Es ist

$$f(j) = \frac{N}{6} - (j-1) + \frac{(j-1)^2}{N} - \frac{1}{6N} + \frac{2}{N^2}$$

und durch Indexverschiebung und Anwendung der Summenformeln $\sum_{k=0}^{m} i = \frac{m(m+1)}{2}$ und $\sum_{k=0}^{m} i^2 = \frac{m(m+1)(2m+1)}{6}$ erhält man

$$\sum_{j \in \{1,\dots,N\}} f(j) = \sum_{j=1}^{N} f(j) = \frac{2}{N}.$$

Sei nun $a \in \{1, \dots, N\}$ und

$$b = \begin{cases} a+1, & a < N \\ 1, & a = N \end{cases}, \qquad c = \begin{cases} a-1, & a > 1 \\ N, & a = 1 \end{cases},$$

dann zeigt man durch Nachrechnen

$$2f(a) - f(b) - f(c) = \begin{cases} -\frac{2}{N} + 2, & a = 1, b = 2, c = N, \\ -\frac{2}{N}, & \text{sonst.} \end{cases}$$

Wir betrachten nun die Multiplikation von $\frac{1}{2}T_I^F$ und der zirkulären Matrix A mit erster Zeile $A_{1,:} = [f(1), f(2), \dots, f(N)]$. Wegen der Besetzung von T_I^F erhält man folgende Struktur

$$\begin{aligned} (\frac{1}{2}AT_I^F)_{i,k} &= \frac{1}{2}\sum_{j=1}^{N}(A)_{ij} \cdot (T_I^F)_{j,k} \\ &= \frac{1}{2}\begin{cases} \sum\limits_{j \in \{1,\dots,N\}} f(j) + 2f(1) - f(2) - f(N), & i = k \\ \sum\limits_{j \in \{1,\dots,N\}} f(j) + 2f(a) - f(b) - f(c) \text{ mit } a \neq 1, & i \neq k \end{cases} \\ &= \frac{1}{2}\begin{cases} 2, & i = k \\ 0, & i \neq k \end{cases} \end{aligned}$$

Damit ist A die Inverse zu $\frac{1}{2}T_I^F$.

Die anderen Aussagen folgen, ebenfalls durch nachrechnen und werden hier nicht aufgelistet.

Anhang B

Anhand des Randwertproblems

$$\begin{cases} -\Delta u + u = f \text{ in } \Omega = [0,1]^2, \\ \frac{\partial u}{\partial n} = g \text{ auf } \Gamma_N = [0,1]\times\{1\}, \\ u = 0 \text{ auf } \Gamma_D = \partial\Omega \setminus \Gamma_N. \end{cases} \tag{27}$$

soll ein Ablauf zur Berechnung äquilibrierter Flüsse in `CONCEPTS` illustriert werden. Um ein vollständiges Bild dessen zu erhalten, wird im Vorfeld die Konstruktion der FEM-Lösung mithilfe der `C++`-Bibiliothek im Listing 1 abgebildet.

```
//Aufbau des Gebietes
concepts::Array<uint> arr(concepts::makeArray<uint>(4, 1, 1, 2, 1));
concepts::Square msh(1.0, 1.0, arr);

//Aufbau des Approximationsraumes
concepts::BoundaryConditions bc;
bc.add(concepts::Attribute(1),
       concepts::Boundary(concepts::Boundary::DIRICHLET));
bc.add(concepts::Attribute(2),
       concepts::Boundary(concepts::Boundary::NEUMANN));
hp2D::hpAdaptiveSpaceH1 spc(msh, h, p, &bc);
spc.rebuild();

//Definieren der rechten Seiten
concepts::ParsedFormula<Real> f("((2*pi*pi+1)*sin(pi*x)*sin(pi*y))");
concepts::ParsedFormula<Real> g("(pi*sin(pi*x)*cos(pi*y))");

//Aufbau eines Spurraums auf Neumannkanten
concepts::Set<uint> attrNeumann(static_cast<uint> (2));
hp2D::TraceSpace NTspc(spc, attrNeumann);

//Linearformen der rechten Seite
hp2D::Riesz<Real> lform_f(f);
hp1D::Riesz<Real> lform_g(g);

//Bilinearformen
hp2D::Laplace<Real> la;
hp2D::Identity<Real> id;

//Berechnung der Integrale der rechten Seite...
concepts::Vector<Real> rhs_f(spc, lform);
concepts::Vector<Real> rhs_g(NTspc, lform);
```

```
33
34 //... und Assemblierung der Systemmatrix
35 concepts::SparseMatrix<Real> A(spc, la);
36 concepts::SparseMatrix<Real> M(spc, id);
37 M.addInto(A, 1.0);
38 A.compress();
39
40 //Berechnung der FEM-Loesung
41 concepts::SuperLU<Real> solver(A);
42 concepts::Vector<Real> sol(spc);
43 solver(rhs_f+rhs_g,sol);
```

Listing 1: Berechnung einer FEM-Approximation des Problems (27) in CONCEPTS zu vorgegebenem Verfeinerungsgrad h und geforderter Approximationsordnung p.

Die Berechnung der äquilibrierten Flüsse mithilfe der in dieser Arbeit vorgestellten Klassen wird im Listing 2 dargestellt.

```
1  //Darstellung der FEM-Loesung und deren Gradient
2  concepts::ElementFormulaVector<1> u_h(spc, sol, hp2D::Value<Real>());
3  concepts::ElementFormulaVector<2> gradu_h(spc, sol ,hp2D::Grad<Real>());
4
5  //Berechnung der Element- und Kantenpatches
6  concepts::VtxToPatchMaps patchMaps(spc, bc);
7
8  //Linearformen zur Bestimmung des Residuums der rechten Seite
9  hp2D::LinInnerProd_0<Real> lformu_h(u_h);
10 hp2D::LinInnerProd_0<Real> rhsLinform(f);
11 hp2D::LinInnerProd_1<Real> lformgradu_h(gradu_h);
12
13 //Aufbau des Residuums
14 hp2D::InnerResidual<Real> innerRes(spc);
15 innerRes.add(lformgradu_h);
16 innerRes.add(lformu_h);
17 innerRes.add(rhsLinform, -1);
18
19 //Berechnung der approximierenden Momente
20 hp2D::ApprxMoments<Real> apprxMoments(spc, sol);
21 //Dirichletkanten haben das Attribut 1
22 apprxMoments.addDirichlet(1);
23 //Neumannkanten haben das Attribut 2
24 apprxMoments.addNeumann(g, attrNeumann);
25 apprxMoments.addComplete();
26
27 //Berechnung der Momente erster (und hoeherer) Ordnung
28 hp2D::Moments<Real> moments(spc, patchMaps, innerRes, apprxMoments);
29
```

```
30 //Aufbau eines Spurraums auf inneren Kanten mit Attribut 0
31 concepts::Set<uint> innerEdgeSet(static_cast<uint>(0));
32 hp2D::TraceSpace ITspc(spc, tspcSet);
33
34 //Konstruktion der Fluesse auf inneren Kanten
35 hp2D::Fluxes flux(tspc, moments);
```

Listing 2: Berechnung äquilibrierter Flüsse mithilfe der in Listing 1 definierten rechten Seiten und der berechneten FEM-Lösung.