

# EIN GEMISCHTER FINITE ELEMENTE ANSATZ FÜR EIN NICHTLINEARES OPTIMAL-DESIGN-PROBLEM

Diplomarbeit

Humboldt-Universität zu Berlin  
Mathematisch-Naturwissenschaftliche Fakultät II  
Institut für Mathematik

eingereicht von: David Günther  
geb: 18.06.1980  
in: Berlin

1. Betreuer: Prof. Dr. Carsten Carstensen  
2. Betreuer: Dr. Soeren Bartels

Datum: 24. September 2007

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>5</b>
<b>2</b>	<b>Notation und elementare Resultate</b>	<b>8</b>
<b>3</b>	<b>Modellierung</b>	<b>10</b>
3.1	Der Schubmodul . . . . .	10
3.2	Die Torsionssteifheit . . . . .	11
3.3	Das Optimal-Design-Problem . . . . .	12
3.4	Die gemischte Formulierung . . . . .	14
<b>4</b>	<b>Regularisierung der Gemischten Formulierung</b>	<b>18</b>
4.1	Regularisierung konvexer Funktionen . . . . .	18
4.2	Regularisierung des Optimal-Design-Problems . . . . .	19
4.3	Die konjugierte Energiedichte . . . . .	22
<b>5</b>	<b>Finite Elemente Methode</b>	<b>24</b>
5.1	Die Triangulierung . . . . .	24
5.2	Das diskrete Problem . . . . .	25
<b>6</b>	<b>Existenz und Eindeutigkeit</b>	<b>28</b>
6.1	Die exakte Lösung . . . . .	28
6.2	Die diskrete Lösung . . . . .	28
<b>7</b>	<b>Fehlerabschätzungen</b>	<b>30</b>
7.1	A priori Fehlerabschätzung . . . . .	30
7.2	A posteriori Fehlerabschätzung . . . . .	31
<b>8</b>	<b>Die FEM-Schleife</b>	<b>35</b>
8.1	Markierung . . . . .	35
8.2	Verfeinerung . . . . .	38
8.3	Lösen . . . . .	39
8.4	Fehlerschätzung . . . . .	41

<b>9</b>	<b>Numerische Beispiele</b>	<b>42</b>
9.1	Definintion der Modellbeispiele . . . . .	42
9.2	Der Regularisierungsalgorithmus . . . . .	44
9.3	Das optimale Design der Modellbeispiele . . . . .	49
9.4	Weitere numerische Kenngrößen . . . . .	61
<b>10</b>	<b>Ein Zielorientierter Ansatz</b>	<b>65</b>
<b>11</b>	<b>Thesen</b>	<b>73</b>
<b>A</b>	<b>Dokumentation der Software</b>	<b>75</b>
A.1	Überblick über die verfügbaren FE-Methoden und Module . . . . .	75
A.2	Ein Flußdiagramm . . . . .	81
A.3	Ein Beispiel . . . . .	83
A.4	Auswertung der Lösung . . . . .	85
A.5	Der Verzeichnissbaum . . . . .	86
	<b>Literaturverzeichnis</b>	<b>89</b>

An dieser Stelle möchte ich Herrn Professor Dr. Carstensen für die Vergabe und die Betreuung dieser interessanten Diplomarbeit danken. Bei Dipl. Math. Hella Rabus möchte ich mich für ihre Hilfe und Betreuung während des Schreibens dieser Arbeit bedanken.

# Kapitel 1

## Einleitung

Ein klassisches Problem der Elastizitätstheorie betrachtet die Konstruktion eines Balken, welcher aus zwei verschiedenen Materialien bestehen soll. Beide Materialien besitzen eine unterschiedliche Steifheit und sollen zu bestimmten Anteilen in dem Balken vorkommen. Das Problem besteht nun darin, beide Materialien so im Balken zu verteilen, dass der Balken eine maximale Steifheit besitzt. Betrachtet man den Querschnitt  $\Omega \subset \mathbb{R}^2$  des Balken lässt sich dieses Problem mathematisch als Minimierungsproblem formulieren

$$J(v) = \int_{\Omega} \varphi(|\nabla v|) - v \, dx \longrightarrow \min!, \quad (1.1)$$

wobei die Funktion  $\varphi(t)$  gegeben ist durch

$$\varphi(t) = \begin{cases} \frac{1}{2}\mu_2 t^2 - \lambda\mu_2 & t^2 \leq 2\lambda \\ \frac{1}{2}\mu_1 t^2 - \lambda\mu_1 & t^2 \geq 2\lambda. \end{cases}$$

Die auftretenden Parameter  $\mu_1, \mu_2$  und  $\lambda$  sind durch die gegebenen Materialien und deren Anteil im Balken gegeben. (1.1) beschreibt ein nicht-konvexes Minimierungsproblem und damit existiert im Allgemeinen kein eindeutiger Minimierer. Da das Funktional zudem nicht schwach unterhalbstetig in  $H^1(\Omega)$  ist, führen auch die sogenannten direkten Methoden der Variationsrechnung nicht zu einem Minimum von (1.1). Um dieses Problem zu umgehen, reformuliert man (1.1) mittels der konvexen Hülle von  $\varphi$ . Das resultierende konvexe Optimierungsproblem ist nun eindeutig lösbar. Durch die Konvexifizierung (auch Relaxierung genannt) des Problems bildet sich aber eine Kontaktzone zwischen den Materialien, in der es zu einer Vermischung beider Materialien kommt. Die Fläche dieser Grenzschrift hängt stark von den Materialanteilen ab. Eine Äquivalenz des relaxierten Problems und (1.1) ist nur dann gegeben, wenn das Maß der Grenzschrift null ist.

Das Problem (1.1) wurde zunächst als Beispiel von Kohn und Strang in ihrer dreiteiligen Abhandlung *Optimal Design and Relaxation of Variational Problems* [21, 22, 23] 1986 formuliert. Fokus dieser Arbeit war es eine Verbindung zwischen Strukturoptimierung, Relaxierung und Homogenisierung (Vermischung) herzustellen und deren Lösungen zu beschreiben. Die ersten numerischen Experimente zu (1.1) wurden 1991 von Kawohl,

Stara und Wittum in ihrer Arbeit *Analysis and Numerical Studies of a Problem in Shape Design* [20] geführt. Zum einen wurden hier konkrete Existenz- und Eindeutigkeitsaussagen für das relaxierte Problem getroffen und die Regularität dessen Lösung bestimmt. Der zweite Teil der Arbeit beschäftigte sich mit den numerischen Experimenten. Mittels einer Finite Volumen Methode wurde das Optimal-Design Problem diskretisiert. Durch die mangelnde Rechenleistung der damaligen Zeit lag der Schwerpunkt der Experimente auf der Konstruktion eines Multigridl6ser. Dadurch waren sie in der Lage, das diskrete Problem auf feinen Triangulierungen zu l6sen. Und konnten somit die Kontaktzone relativ gut aufl6sen. Allerdings waren die zugrunde liegenden Triangulierungen stets uniform. Da der Verlauf der Grenzschicht a priori nicht bekannt ist, war eine lokal feinere Aufl6sung der Kontaktzone nicht m6glich. W6hrend der 90er Jahre wurden speziell adaptive Finite Elemente Methoden entwickelt, die die Konstruktion von problemspezifischen Triangulierungen erm6glichte. Mittels der diskreten L6sung konnte man nun Informationen gewinnen, wie man Triangulierungen lokal verfeinern muss, um den Fehler zum kontinuierlichen Problem optimal zu reduzieren. Nachdem zun6chst nur f6r lineare, haupts6chlich elliptische, Probleme Fehlersch6tzer entwickelt wurden, haben Bartels und Carstensen 2006 die ersten Fehlersch6tzer f6r das relaxierte Optimal-Design Problem (1.1) vorgestellt. In ihrer Arbeit *A convergent adaptive finite element method for an optimal design problem* [4] wurde mittels einer  $\mathcal{S}_1$ -Finite Elemente Methode das Optimal-Design Problem diskretisiert. Die Konvergenzanalyse dieser Methode motivierte Residual- bzw Average-Fehlersch6tzer, die eine Lokalisierung der Kontaktzone und eine optimale Fehlerreduktion der diskreten zur kontinuierlichen L6sung erm6glichten. Die Verschiebung  $u$ , welche die relaxierte Version von (1.1) l6st, wurde durch st6ckweise affine, global stetige Ansatzfunktionen approximiert. Die zugeh6rige diskrete Spannung  $\sigma_h = \varphi(|\nabla u_h|)$  wurde somit st6ckweise konstant approximiert. Betrachtet man den gegebenen Balken, so lassen sich durch die Spannungen  $\sigma(x)$  die Regionen feststellen, wo sehr grosse Kr6fte wirken. Diese Regionen sollten mit der Diskretisierung sehr fein aufgel6st werden. In [13] haben Carstensen und M6ller 2002 Eigenschaften der Spannung  $\sigma$  herausgearbeitet. Die Spannungen von (1.1) liegen in  $H(\text{div}; \Omega)$ . Ein  $H(\text{div}; \Omega)$ -konformer Ansatzraum f6r die Spannungen w6rde somit auf nat6rliche Weise dem Problem entsprechen.

In dieser Arbeit wollen wir mittels eines gemischten Finite Elemente Ansatz das Optimal-Design Problem diskretisieren. Gesucht ist eine Approximation der L6sung  $(\sigma, u) \in H(\text{div}; \Omega) \times L^2(\Omega)$  durch  $(\sigma_h, u_h) \in RT_0(\mathcal{T}) \times \mathcal{P}_0(\mathcal{T})$ , so dass das diskrete Optimal-Design Problem gel6st wird. Zun6chst werden wir in Kapitel 2 Notationen einf6hren und einige Vorbemerkungen f6hren, welche sp6ter Verwendung finden. Danach rekapitulieren wir in Kapitel 3 die Modellierung von Problem (1.1) und stellen die gemischte Formulierung auf. Da wir durch das gemischte Problem eine Regularit6tsordnung des nichtlinearen Funktionals verlieren, gl6tten wir dies, um den Schwierigkeiten des Subdifferentials in den sp6teren numerischen Experimenten zu entgehen. Dies geschieht in Kapitel 4. Hier formulieren wir die regularisierte Version der gemischten Methode. Mittels der Raviart-Thomas Finite Elemente Methode beschreiben wir in Kapitel 5 das diskrete Problem und zeigen

anschließend in Kapitel 6 die Existenz und Eindeutigkeit der Lösungen  $(\sigma, u)$  und  $(\sigma_h, u_h)$ . Die Güte des diskreten Problems wird in der a priori Analysis in Kapitel 7 festgestellt. Die anschließende a posteriori Analysis motiviert dann zwei Fehlerschätzer, die eine adaptive Gittergenerierung ermöglichen sollen.

Der zweite Teil dieser Arbeit beschäftigt sich mit der numerischen Realisierung des diskreten Optimal-Design Problems. Der Fokus liegt hierbei auf der Implementierung des Problems und den zugehörigen Algorithmen, und auf der Präsentation der numerischen Ergebnisse. Eine Verifikation der theoretischen Überlegungen steht hierbei im Mittelpunkt. Die Finite-Elemente Schleife, mittels der eine Approximation der kontinuierlichen Lösung berechnet wird, wird in Kapitel 8 beschrieben. Hier gehen wir besonders auf die adaptive Gittergenerierung ein. In Kapitel 9 definieren wir Modellbeispiele, anhand derer eine Verifikation der Theorie stattfinden soll. Der Regularisierungsalgorithmus wird ebenfalls in diesem Kapitel vorgestellt. Nachdem einige numerische Kenngrößen präsentiert und ausgewertet wurden, stellen wir in Kapitel 10 einen alternativen Ansatz zur adaptiven Gittergenerierung vor. Mit einem vorgegebenen Zielfunktional wird das Optimal-Design Problem in einem Minimierungsproblem mit PDE-Nebenbedingungen eingebettet. Mit Hilfe eines Lagrangeansatzes wird dann ein Fehlerindikator motiviert. An Hand der in Kapitel 9 vorgestellten Beispiele, werden dann die numerischen Ergebnisse des zielorientierten Ansatzes gezeigt. Abschließend findet sich im Anhang A die Dokumentation der Optimal-Design Software.

# Kapitel 2

## Notation und elementare Resultate

In diesem Kapitel wollen wir zunächst einige Notationen einführen, von denen im späteren Verlauf häufiger Gebrauch gemacht wird. Des Weiteren wollen wir einige elementare Aussagen aufführen, die vor allem in den Beweisen der Theoreme und Lemmata Gebrauch finden. Auf die Beweise der Aussagen wird hier verzichtet. Sie können in den einschlägigen Lehrbüchern wie z.B. in [18] gefunden werden.

Im Folgenden definieren wir einige Funktionenräume, die später Verwendung finden. Es sei  $\Omega \subset \mathbb{R}^2$  ein einfach zusammenhängendes Gebiet mit Lipschitzrand  $\partial\Omega$ .

$$\begin{aligned} L^2(\Omega) &= \{ \phi : \Omega \rightarrow \mathbb{R} \text{ meßbar} \mid \int_{\Omega} \phi(x)^2 dx < \infty \}, \\ H^1(\Omega) &= \{ \phi \in L^2(\Omega) \mid \int_{\Omega} |\nabla \phi(x)|^2 dx < \infty \}, \\ H_D^1(\Omega) &= \{ \phi \in H^1(\Omega) \mid \phi = \phi_D \text{ auf } \partial\Omega, \phi_D \in L^2(\partial\Omega) \}, \\ L_{\text{loc}}^2(\Omega) &= \{ \phi : \Omega \rightarrow \mathbb{R} \mid \text{für alle } D \subset\subset \Omega \text{ ist } \phi|_D \in L^2(D) \}, \\ H_{\text{loc}}^1(\Omega) &= \{ \phi \in L_{\text{loc}}^2(\Omega) \mid \text{für alle } D \subset\subset \Omega \text{ ist } \nabla \phi|_D \in L_{\text{loc}}^2(D) \}, \\ H(\text{div}; \Omega) &= \{ \phi \in L^2(\Omega)^2 \mid \text{div } \phi(x) \in L^2(\Omega) \}, \\ L^\infty(\Omega) &= \{ \phi : \Omega \rightarrow \mathbb{R} \mid \phi \text{ ist wesentlich beschränkt auf } \Omega \}. \end{aligned}$$

Die Differentialoperatoren  $\nabla$ , Curl und div sind für hinreichend glatte skalare Funktionen  $u = u(x, y)$  mit  $u_x := \frac{\partial}{\partial x} u$  und  $u_y := \frac{\partial}{\partial y} u$  wie folgt definiert

$$\nabla u = \begin{pmatrix} u_x \\ u_y \end{pmatrix}, \quad \text{Curl } u = \begin{pmatrix} u_y \\ -u_x \end{pmatrix}, \quad \text{div } u = u_x + u_y.$$

Für vektorwertige Funktionen  $\phi = \phi(x, y) = (\phi_1(x, y), \phi_2(x, y))$  sind die Operatoren curl und div gegeben durch

$$\text{curl } \phi = \phi_{2,x} - \phi_{1,y}, \quad \text{div } \phi = \phi_{1,x} + \phi_{2,y}.$$

Für  $A, B \in \mathbb{R}^2$  sei das Skalarprodukt definiert durch  $A \cdot B = A_1 B_1 + A_2 B_2$ , das Kreuzprodukt durch  $A \times B = A_1 B_2 - A_2 B_1$ . Für  $a, b \in \mathbb{R}$  bedeutet  $a \lesssim b$ , dass mit einer Konstanten  $c \in \mathbb{R}$  der Vergleich  $a \leq c b$  gilt.



Sei die  $L^2$ -Norm definiert durch  $\|u\|_{L^2(\Omega)}^2 = \int_{\Omega} u^2 dx$ , die  $H^1$ -Norm durch  $\|u\|_{H^1(\Omega)}^2 = \|u\|_{L^2(\Omega)}^2 + \|\nabla u\|_{L^2(\Omega)}^2$  und die  $H(\text{div})$ -Norm durch  $\|u\|_{H^1(\Omega)}^2 = \|u\|_{L^2(\Omega)}^2 + \|\text{div } u\|_{L^2(\Omega)}^2$ . Das  $L^2$ -Skalarprodukt ist definiert durch  $(u, v)_{L^2(\Omega)} = \int_{\Omega} uv dx$ . Die  $L^2$ -Orthogonalität  $(u, v)_{L^2(\Omega)} = 0$  werden wir gelegentlich mit  $u \perp_2 v$  abkürzen.

**Definition 2.1** (Gateaux-Differenzierbarkeit). *Seien  $V$  und  $W$  Banachräume. Sei  $U \subset V$  eine offene Teilmenge. Eine Funktion  $\phi$  ist am Punkt  $x \in U$  Gateaux-differenzierbar, wenn Ableitungen in allen Richtungen von  $x$  existieren. D.h., es existiert eine Funktion  $g : V \rightarrow W$ , so dass gilt*

$$g(y) = \lim_{t \rightarrow 0} \frac{f(x + ty) - f(x)}{t}. \quad (2.1)$$

**Definition 2.2** (Fréchet-Differenzierbarkeit). *Seien  $V$  und  $W$  Banachräume. Sei  $U \subset V$  eine offene Teilmenge. Eine Funktion  $\phi$  ist am Punkt  $x \in U$  Fréchet-differenzierbar, wenn es einen linearen beschränkten Operator  $A$  gibt, so dass gilt*

$$0 = \lim_{\|h\|_V \rightarrow 0} \frac{\|f(x + h) - f(x) - A(h)\|_W}{\|h\|_V}. \quad (2.2)$$

**Lemma 2.3** (Euler-Lagrange-Bedingung). *Für hinreichend glatte Funktionen  $v = v(x, y)$  betrachten wir das Variationsproblem*

$$\min_v \int_{\Omega} f(v, \nabla v, x, y) dx dy,$$

*welches bei  $u$  ein Minimum annimmt. Der Integrand  $f$  sei hinreichend oft differenzierbar. Dann muss der kritische Punkt  $u$ , die Euler-Lagrange-Bedingung erfüllen*

$$f_u - \frac{\partial}{\partial x} f_{u_x} - \frac{\partial}{\partial y} f_{u_y} = 0. \quad (2.3)$$

**Lemma 2.4** (Gauss'scher Integralsatz). *Für  $\phi \in H(\text{div}; \Omega)$  gilt*

$$(\text{div } \phi, 1)_{L^2(\Omega)} = (\phi, \nu)_{L^2(\partial\Omega)}, \quad (2.4)$$

*wobei  $\nu$  die Einheitsnormale an  $\partial\Omega$  ist.*

**Lemma 2.5** (Darstellungssatz von Riesz). *Sei  $X$  ein Hilbertraum. Für alle  $f \in X^*$  existiert ein eindeutiges  $u_f \in X$ , so dass*

$$f(v) = (u_f, v)_X \quad (2.5)$$

*für alle  $v \in X$  gilt. Die Abbildung  $i : X^* \rightarrow X$  mit  $i(f) = u_f$  beschreibt einen Normisomorphismus.*

# Kapitel 3

## Modellierung

In diesem Kapitel werden die primale und die duale Formulierung, sowie die Energieformulierung, zu dem Optimal-Design-Problem dargestellt, wobei wir uns auf die wesentlichen Schritte hierbei konzentrieren. Essentielle Arbeiten, welche wir hier aufgreifen, sind [21, 22, 23] und [4]. Weiterführende Details können dort nachgelesen werden.

### 3.1 Der Schubmodul

Wir betrachten ein 3-dimensionalen Balken, welcher aus zwei verschiedenen Materialien bestehen soll. Beide Materialien besitzen hierbei unterschiedliche Eigenschaften. Uns interessiert nun wie beide Materialien, unter vorgegebenen Anteilen, im Balken verteilt werden müssen, damit sich der Balken unter bestimmten Kräften möglichst steif verhält.

Auskunft über die linearen elastischen Verformungen in Folge einer zur Fläche  $A$  tangential wirkenden Kraft  $F$  gibt das Schubmodul  $\mu$ , welches nur vom jeweiligen Material abhängt. Hierbei gilt das Gleichgewicht  $|F| = \mu A \alpha$ , wobei der Winkel  $\alpha$  ein Maß für die Verzerrungen des Balkens infolge der Kraft  $F$  ist (Abbildung 3.1). Die Materialien sollen so verteilt sein, dass der Balken maximale Torsionssteifheit annimmt, und es somit nur zu minimalen Verzerrungen kommt.

Da sich dieses 3d-Problem invariant unter Translation verhält, können wir es auf ein 2d-Problem reduzieren. Sei  $\Omega \subset \mathbb{R}^2$ , welches man als Querschnitt unseres Balkens interpretieren kann, ein einfach zusammenhängendes Gebiet mit Lipschitzrand. Zu beiden Materialien, welche das Gebiet  $\Omega$  ausfüllen, seien die Schubmodule  $0 < \mu_1 < \mu_2 < \infty$  bekannt. Das Optimal-Design-Problem besteht nun darin, die Materialien so in  $\Omega$  zu verteilen, dass maximale Steifheit angenommen wird. Damit erhalten wir eine Zerlegung des Gebietes  $\Omega$  in  $\Omega_1$  und  $\Omega_2$ , so dass

$$\Omega = \Omega_1 \cup \Omega_2 \quad \text{und} \quad \Omega_1 \cap \Omega_2 = \emptyset$$

gilt. Die Anteile, zu welchen beide Materialien vorkommen, seien durch einen Parameter  $0 < \xi < 1$  vorgegeben. Da beide Materialien  $\Omega$  ausfüllen, sind die Flächen von  $\Omega_1$  und  $\Omega_2$

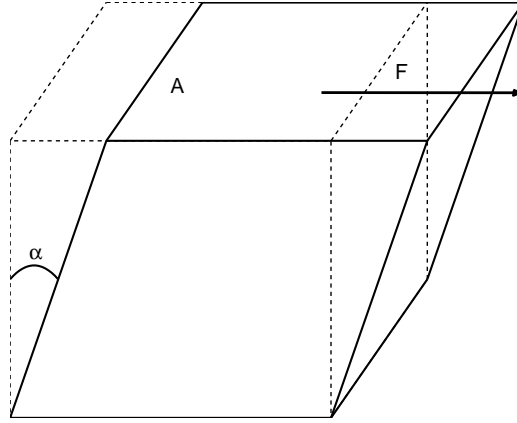


Abbildung 3.1: Eine Kraft  $F$  wirkt tangential auf der Oberfläche  $A$  eines Balkens. Der Winkel  $\alpha$  gibt in Abhängigkeit des Schubmoduls  $\mu$  an, wie stark sich der Balken unter dieser Kraft verformt.

gegeben durch

$$|\Omega_1| = \xi |\Omega| \quad \text{und} \quad |\Omega_2| = (1 - \xi) |\Omega|. \quad (3.1)$$

Man kann nun zeigen, dass die Torsionssteifheit nur von den Schubmodulen  $\mu(x)$  am Materialpunkt  $x \in \Omega$  abhängt. D.h., wir suchen eine Verteilung der Schubmodule in der Menge

$$\mathcal{M} := \{\mu \in L^\infty(\Omega) \mid \mu(x) \in \{\mu_1, \mu_2\} \text{ für fast alle } x \in \Omega\}.$$

Damit ist  $\Omega_j$  bis auf eine Nullmenge definiert durch  $\Omega_j := \{x \in \Omega \mid \mu(x) = \mu_j\}$ . Der Schubmodul  $\mu(x)$  am Materialpunkt  $x \in \Omega$  muss demzufolge bei einer optimalen Verteilung aus der Menge

$$\mathcal{M}_\xi := \{\mu \in \mathcal{M} \mid (\mu(x), 1)_{L^2(\Omega)} = \mu_\xi |\Omega|\}$$

mit  $\mu_\xi := \xi \mu_1 + (1 - \xi) \mu_2$  kommen. Die Menge  $\mathcal{M}_\xi$  definiert hierbei unsere Nebenbedingung zur optimalen Verteilung.

## 3.2 Die Torsionssteifheit

Für gegebenes Schubmodul  $\mu \in \mathcal{M}_\xi$  und einen Querschnitt  $\Omega$  ist die Torsionssteifheit  $T$  gesucht. Die Reziproke  $T^{-1}$  ist durch einen Minimierungsprozess über  $\Lambda := \{\sigma \in H(\text{div}; \mathbb{R}^2) \mid \sigma \text{ ist divergenzfrei und } \sigma \cdot \nu = 0 \text{ auf } \partial\Omega\}$  gegeben

$$T^{-1} = \inf_{\sigma \in \Lambda} \frac{E(\sigma)}{m(\sigma)^2}, \quad (3.2)$$

wobei die elastische Energie  $E$  und der Torsionsmoment  $m$  gegeben ist durch

$$E(\sigma) = \frac{1}{2} \int_{\Omega} \mu(x) |\sigma(x)|^2 dx \quad \text{und} \quad m(\sigma) = \int_{\Omega} \sigma \times x dx.$$

Da die Spannung  $\sigma$  divergenzfrei ist, existiert eine Verschiebung  $u \in V := H_0^1(\Omega)$ , so dass gilt

$$\sigma = -\text{Curl } u. \quad (3.3)$$

Substituiert man nun (3.3) im Torsionsmoment  $m$ , erhält man mit partieller Integration

$$m(\sigma) = (\sigma \times x, 1)_{L^2(\Omega)} = (\text{Curl } u \times x, 1)_{L^2(\Omega)} = 2(u, 1)_{L^2(\Omega)}. \quad (3.4)$$

Mit (3.3) und (3.4) ergibt sich für (3.2)

$$8T^{-1} = \inf_{v \in V} \frac{(\mu \nabla v, \nabla v)_{L^2(\Omega)}}{(v, 1)_{L^2(\Omega)}^2}. \quad (3.5)$$

Aus der Euler-Lagrange-Bedingung der Variationsrechnung folgt nach [4], dass für ein  $u$ , bei dem das Infimum in (3.5) angenommen wird, die erste Variation verschwinden muss. Daraus ergibt sich mit  $s(u) = (8T^{-1}, u)_{L^2(\Omega)}$

$$(\mu \nabla u, \nabla v)_{L^2(\Omega)} = (s(u), v)_{L^2(\Omega)} \quad \text{für alle } v \in V. \quad (3.6)$$

Bis auf Skalierung ist die Lösung  $u$  eindeutig bestimmt. Im Folgenden setzen wir die Skalierung auf  $s(u) = 1$ . Damit ist  $u$  die schwache Lösung der Differentialgleichung

$$-\text{div}(\mu \nabla u) = 1 \quad \text{in } \Omega \quad (3.7)$$

und nimmt das Minimum an bei

$$-\frac{T}{16} = \min_{v \in V} \left( \frac{1}{2} \int_{\Omega} (\mu |\nabla v|^2 - v) dx \right). \quad (3.8)$$

### 3.3 Das Optimal-Design-Problem

Für  $\mu \in \mathcal{M}_{\xi}$  und  $v \in V$  definieren wir folgendes Energiefunktional

$$F(u, \mu) = \frac{1}{2} \int_{\Omega} (\mu |\nabla v|^2 - v) dx$$

Das Optimal-Design-Problem besteht nun darin, das optimale Design  $\mu$  in

$$\inf_{\mu \in \mathcal{M}_{\xi}} \inf_{v \in V} F(v, \mu) \quad (3.9)$$

zu finden. Mit Hilfe eines Lagrange Ansatzes für die Nebenbedingung  $\mathcal{M}_{\xi}$  in (3.9) erhält man eine äquivalente Darstellung

$$\sup_{\lambda \in \mathbb{R}} \inf_{\mu \in \mathcal{M}} \inf_{v \in V} \left( F(v, \mu) - \lambda \int_{\Omega} (\mu - \mu_{\xi}) dx \right). \quad (3.10)$$

Vertauschung des Suprema und der Infima führt zu gegebenen  $\lambda \in \mathbb{R}$  und  $v \in V$  zu

$$\inf_{\mu \in \mathcal{M}_\xi} \left( F(v, \mu) - \lambda \int_{\Omega} (\mu - \mu_\xi) dx \right) = \int_{\Omega} (\tilde{\psi}_\lambda(|\nabla v|) - v) dx + \lambda \mu_\xi |\Omega|, \quad (3.11)$$

wobei  $\tilde{\psi}_\lambda$  gegeben ist durch

$$\tilde{\psi}_\lambda(t) := (0.5 t^2 - \lambda) \min \{\mu_1, \mu_2\}.$$

D.h., für  $0.5 |\nabla v|^2 - \lambda < 0$  ist  $\mu = \mu_2$  und für  $0.5 |\nabla v|^2 - \lambda > 0$  ist  $\mu = \mu_1$  die optimale Wahl. Im Fall  $0.5 |\nabla v|^2 - \lambda = 0$  ist jede Wahl von  $\mu$  optimal. Sei nun  $\lambda$  fixiert, dann reduziert sich (3.10) auf ein nicht konvexes Minimierungsproblem

$$\inf_{v \in V} \tilde{E}(\lambda, v) \quad \text{mit} \quad \tilde{E}(\lambda, v) = \int_{\Omega} (\tilde{\psi}_\lambda(|\nabla v|) - v) dx + \lambda \mu_\xi |\Omega|. \quad (3.12)$$

Da (3.12) weder streng konvex noch schwach unterhalbstetig in  $H^1(\Omega)$  ist ([20]), ist die Konstruktion eines eindeutigen Minimierers in Allgemeinen nicht möglich. Aus der Variationsrechnung ist bekannt [10], dass das Minimum der konvexen Hülle von  $\tilde{E}(\lambda, v)$  unter gewissen Voraussetzungen dem Infimum von  $\tilde{E}(\lambda, v)$  entspricht. Sei nun  $\psi_\lambda$  die konvexe Hülle von  $\tilde{\psi}_\lambda$ , definiert durch

$$\psi_\lambda(t) := \begin{cases} 0.5 \mu_2 t^2 + \lambda \xi(\mu_1 - \mu_2) & \text{for } t \leq t_1, \\ t_1 \mu_2 (t - 0.5 t_1) + \lambda \xi(\mu_1 - \mu_2) & \text{for } t_1 < t < t_2, \\ 0.5 \mu_1 t^2 + 0.5 \mu_1 t_2^2 - 0.5 t_1^2 \mu_2 + \lambda \xi(\mu_1 - \mu_2) & \text{for } t_2 \leq t, \end{cases} \quad (3.13)$$

mit Parametern  $t_1 := \sqrt{2\lambda(\mu_1/\mu_2)}$  und  $t_2 := (\mu_2/\mu_1)t_1$  und  $\mu_1 t_2 = \mu_2 t_1$ . Dann ist das zugehörige konvexe Problem gegeben durch: Finde  $(\sigma, u)$  mit  $\sigma = \psi_\lambda(|\nabla u|)$  für das Sattelpunktpunktproblem

$$\sup_{\lambda \in \mathbb{R}} \inf_{v \in V} E(\lambda, v) \quad \text{mit} \quad E(\lambda, v) = \int_{\Omega} (\psi_\lambda(|\nabla v|) - v) dx + \lambda \mu_\xi |\Omega|. \quad (3.14)$$

**Bemerkung 3.1.** *Durch die Konvexifizierung der Energiedichte  $\tilde{\psi}_\lambda$  entsteht eine Grenzschicht zwischen beiden Materialien. Hier kommt es zu einer Vermischung der Materialien, die im ursprünglichen Problem nicht gewünscht ist, und im Allgemeinen sind die Lösungen von (3.12) und (3.14) innerhalb der Kontaktzone nicht äquivalent ([20]). Ein Äquivalenz tritt erst ein, wenn das Maß der Fläche der Kontaktzone null ist.*

**Bemerkung 3.2.** *Für die Verschiebung  $u$  des Optimal-Design Problems gilt nach [20]*

- (a)  $u \in L^\infty(\Omega)$ ,
- (b)  $u \in W_{\text{loc}}^{1,\infty}(\Omega)$ ,
- (c)  $u \in C^\infty(\text{int } \Omega_1 \cup \text{int } \Omega_2)$ .

**Bemerkung 3.3.** *Für die Spannung  $\sigma \in H(\text{div}; \Omega)$  des Optimal-Design Problems gilt nach [13]*

- (a)  $f \in L^2(\Omega) \Rightarrow \sigma \in H_{\text{loc}}^1(\Omega)$ ,
- (b)  $f \in H_0^1(\Omega)$ ,  $\partial\Omega$  ist glatt  $\Rightarrow \sigma \in H^1(\Omega)$ .

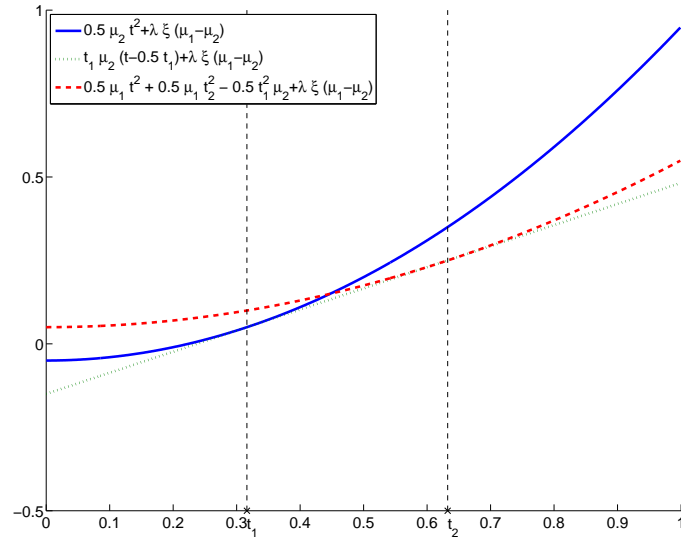


Abbildung 3.2: Darstellung der Energiedichte  $\psi_\lambda(t)$  nach (3.13) mit Parametern  $\mu_1 = 1$ ,  $\mu_2 = 2$ ,  $\xi = 0.5$  und  $\lambda = 0.1$ .

### 3.4 Die gemischte Formulierung

In diesem Abschnitt werden wir die gemischte Formulierung des Energiefunktional für ein fixiertes  $\lambda$  herleiten. Dazu benötigen wir zum einen den Begriff der konjugierten Funktion, welchen wir zunächst definieren, zum Zweiten die Euler-Lagrange-Bedingung. Anschliessend formulieren wir das gemischte Problem.

**Definition 3.4** (Die Konjugierte Funktion). *Sei  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  mit folgenden Eigenschaften gegeben*

- (a)  $\text{dom}(\varphi) := \{x \in \mathbb{R}^n \mid \varphi(x) < \infty\} \neq \emptyset$ ,
- (b) *es existiert eine affine Funktion, welche  $\varphi$  minorisiert.*

*Die konjugierte Funktion  $\varphi^*$  ist definiert durch*

$$\varphi^*(z) := \sup_{x \in \mathbb{R}^n} \{z \cdot x - \varphi(x)\}. \quad (3.15)$$

**Bemerkung 3.5.** *Unter den Voraussetzungen von Definition 3.4 gilt*

- (a)  $\varphi^{**} = \varphi \iff \varphi$  *ist konvex und unterhalbstetig.*
- (b)  $x \in \partial\varphi(y) \implies y \in \partial\varphi^*(x).$

In [3] werden diese Eigenschaften gezeigt.

Ausgehend von (3.14) fixieren wir den Parameter  $\lambda$ . Sei weiterhin nun  $\psi := \psi_\lambda$ . Wir betrachten das zu minimierende Energiefunktional bei einer gegebenen stückweise konstanten Last  $f$

$$E(v) = \int_{\Omega} (\psi(|\nabla v|) - f v) dx + \lambda \mu_\xi |\Omega|. \quad (3.16)$$

Da nach der Euler-Lagrange-Bedingung die erste Variation am Minimum verschwinden muss, ist die starke Formulierung mit  $\varphi(\cdot) = \psi(|\cdot|)$  von (3.16) gegeben durch

$$\begin{cases} -\operatorname{div} D\varphi(\nabla u) = f & \text{in } \Omega, \\ u = 0 & \text{auf } \partial\Omega. \end{cases} \quad (3.17)$$

Betrachtet man die gemischte Formulierung erhält man,

$$\begin{cases} -\operatorname{div} \sigma = f & \text{in } \Omega, \\ \sigma = D\varphi(\nabla u) & \text{in } \Omega. \end{cases} \quad (3.18)$$

Da die Funktion  $\varphi$  stetig differenzierbar und konvex ist, können wir Definition 3.4 anwenden.

**Lemma 3.6.** *Die konjugierte Funktion  $\varphi^*$  zu  $\varphi$  ist gegeben durch*

$$\varphi^*(z) = \begin{cases} \frac{z^2}{2\mu_2} - \lambda\xi(\mu_1 - \mu_2) & \text{für } |z| < \mu_2 t_1, \\ \frac{\mu_2}{2} t_1^2 - \lambda\xi(\mu_1 - \mu_2) & \text{für } |z| = \mu_2 t_1 = \mu_1 t_2, \\ \frac{z^2}{2\mu_1} - \frac{\mu_1 t_2^2}{2} + \frac{\mu_2 t_1^2}{2} - \lambda\xi(\mu_1 - \mu_2) & \text{für } \mu_2 t_1 < |z|. \end{cases} \quad (3.19)$$

*Beweis.* Mit Definition 3.4 und  $h(x, z) := z \cdot x - \varphi(x)$  folgt

$$\varphi^*(z) = \sup_{x \in \mathbb{R}^n} h(x, z) \quad \text{und} \quad 0 = \frac{\partial}{\partial x} h(x, z) = z - \nabla \varphi(x)$$

Mit (3.13) wird das Supremum bei  $x = z/\mu_2$  für  $|z| \leq \mu_1 t_2$  und  $x = z/\mu_1$  für  $|z| \geq \mu_2 t_1$  angenommen.  $\square$

**Bemerkung 3.7.** *Im Gegensatz zur konvexen und stetig differenzierbaren Funktion  $\varphi$  ist die konjugierte Funktion  $\varphi^*$  nur konvex und stetig. Die Ableitung ist stückweise lipschitzstetig und besitzt bei  $|z| = \mu_1 t_2$  einen Sprung. An dieser Stelle ist  $\varphi^*$  nur subdifferenzierbar. Folglich kann man das Subdifferential  $\partial\varphi^*$  mit Hilfe der stückweisen Lipschitzstetigkeit durch*

$$|\partial\varphi^*(z_1) - \partial\varphi^*(z_2)| \lesssim C(z_1, z_2) + |z_1 - z_2| \quad (3.20)$$

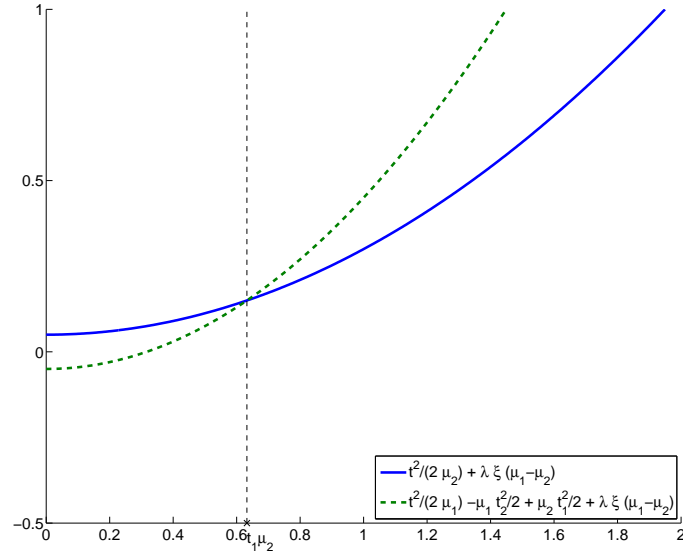


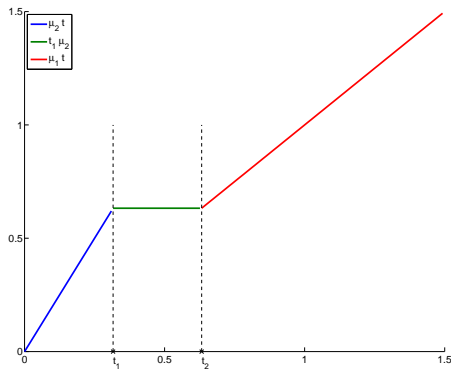
Abbildung 3.3: Darstellung der konjugierten Energiedichte  $\varphi^*(t)$  nach Lemma 3.6 mit Parametern  $\mu_1 = 1, \mu_2 = 2, \xi = 0.5$  und  $\lambda = 0.1$ .

kontrollieren, wobei die Konstante  $C(z_1, z_2)$  gegeben ist durch

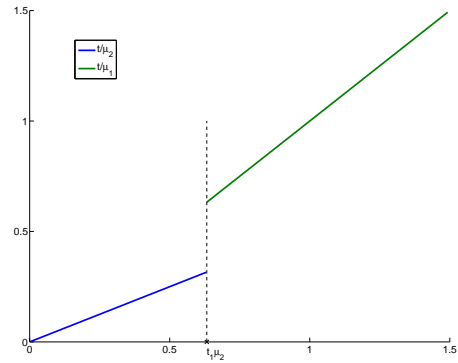
$$C(z_1, z_2) = \begin{cases} 0 & \text{für } |z_1|, |z_2| < \mu_2 t_1 \quad \text{oder} \quad |z_1|, |z_2| > \mu_2 t_1, \\ t_2 - t_1 & \text{sonst.} \end{cases}$$

Mit Hilfe der konjugierten Funktion und Bemerkung 3.5 lässt sich (3.18) schreiben als

$$\begin{cases} -\operatorname{div} \sigma = f & \text{in } \Omega, \\ \nabla u \in \partial \varphi^*(\sigma) & \text{in } \Omega. \end{cases} \quad (3.21)$$



(a) Energiedichte  $\varphi(t)$



(b) konjugierte Energiedichte  $\varphi^*(t)$

Abbildung 3.4: Vergleich der Ableitungen von  $\varphi$  und  $\varphi^*$  mit Parametern  $\mu_1 = 1, \mu_2 = 2, \xi = 0.5$  und  $\lambda = 0.1$ .



Damit ist die gemischte Energieformulierung gegeben durch

**Definition 3.8** (Gemischte Formulierung). Für eine stückweise konstante Last  $f$  finde  $\sigma \in Q_f := \{\tau \in H(\operatorname{div}; \Omega) \mid -\operatorname{div} \tau = f, f \in L^2(\Omega)\}$ , so dass das Minimum

$$\min_{\tau \in Q_f} \int_{\Omega} \varphi^*(\tau) \, dx \quad (3.22)$$

bei  $\sigma$  angenommen wird.

**Bemerkung 3.9.** Mit Hilfe eines Lagrange-Ansatzes lässt sich (3.22) als Sattelpunktproblem schreiben. Für eine stückweise konstante Last  $f$  finde  $(\sigma, \lambda) \in H(\operatorname{div}; \Omega) \times L^2(\Omega)$ , so dass der Sattelpunkt

$$\sup_{\kappa \in L^2(\Omega)} \inf_{\tau \in H(\operatorname{div}; \Omega)} \int_{\Omega} (\varphi^*(\tau) + \kappa(\operatorname{div} \tau + f)) \, dx \quad (3.23)$$

angenommen wird. Den Lagrangemultiplikator  $\lambda$  kann man als die Verschiebung  $u$  interpretieren.

# Kapitel 4

## Regularisierung der Gemischten Formulierung

Im Kapitel 3 wurde die gemischte Formulierung des Optimal-Design-Problems mit Hilfe der konjugierten Funktion dargestellt. Bei diesem Prozess hat sich die Regularitätseigenschaften im Vergleich zum ursprünglichen Problem verschlechtert. Die konjugierte Energiedichte  $\varphi^*$  ist zwar konvex und stetig aber nicht mehr stetig differenzierbar. In der späteren numerischen Realisierung verwenden wir die schwache Formulierung des gemischten Problems. Hier wird die Ableitung von  $\varphi^*$  eingehen. Um das Problem des Subdifferentials  $\partial\varphi^*$  in der Implementierung zu umgehen, approximieren wir  $\varphi^*$  mit einer konvexen stetig differenzierbaren Funktion  $\varphi_\theta^*$ , so dass mit  $\theta \rightarrow 0$  die Funktionen  $\varphi_\theta^* \rightarrow \varphi^*$  konvergieren.

Zunächst geben wir die allgemeine Definition der Yosida-Moreau-Regularisierung und einige Eigenschaften an. Wir folgen hier den Ausführungen von [3]. Anschließend bestimmen wir die Regularisierung von  $\varphi^*$  und betrachten einige Eigenschaften der konjugierten Energiedichte näher.

### 4.1 Regularisierung konvexer Funktionen

Im Folgenden sei  $X$  ein konvexer reflexiver Banachraum und  $\Phi : X \rightarrow (-\infty, \infty]$  eine konvexe Funktion die der Definition 3.4 genügt.

**Definition 4.1** (Yosida-Moreau-Regularisierung). *Sei  $\Phi$  wie oben gegeben. Die Yosida-Moreau-Regularisierung ist für jedes  $\theta > 0$  definiert durch*

$$\Phi_\theta(x) = \inf_{y \in X} \left\{ \Phi(y) + \frac{1}{2\theta} \|x - y\|_X^2 \right\}, \quad x \in X. \quad (4.1)$$

**Theorem 4.2.** *Sei  $\Phi$  wie oben gegeben. Wir setzen  $A = \partial\Phi$ . Dann ist  $\Phi_\theta$  Gateaux-differenzierbar auf  $X$  und es gilt  $A_\theta = \partial\Phi_\theta$  für  $\theta > 0$ . Des Weiteren gilt für alle  $x \in X$  und  $\theta > 0$ :*

- (a)  $\Phi_\theta = 0.5 \theta \|A_\theta x\|_X^2 + \Phi((Id + \theta A)^{-1}x),$
- (b)  $\lim_{\theta \rightarrow 0} \Phi_\theta(x) = \Phi(x),$
- (c)  $\Phi((Id + \theta A)^{-1}x) \leq \Phi_\theta(x) \leq \Phi(x).$

Hierbei ist  $Id$  der Identitätsoperator.

*Beweis.* Der Beweis wird in [3] geführt.

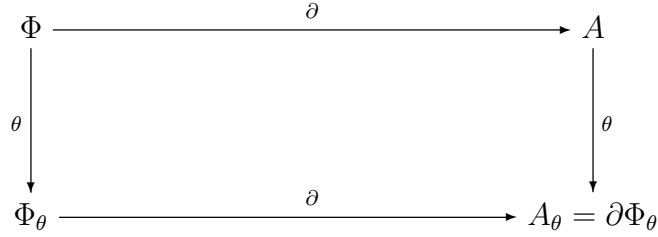


Abbildung 4.1: Kommutatives Diagramm für die Yosida-Regularisierung, bezeichnet mit  $\theta$ , und dem Subdifferential  $\partial$ .

**Bemerkung 4.3.** Gemäß der Definition 4.1 und Theorem 4.2 ist die Regularisierte  $\Phi_\theta$  konvex, Gateaux-differenzierbar und endlich auf  $X$ . Falls  $X$  ein Hilbertraum ist, so ist  $\Phi$  zudem Fréchet-differenzierbar und die Ableitung ist Lipschitzstetig.

## 4.2 Regularisierung des Optimal-Design-Problems

**Lemma 4.4.** Die Regularisierte der konjugierten Funktion (3.6) ist gegeben durch

$$\varphi_\theta^*(t) := \begin{cases} \frac{|t|^2}{2(\theta + \mu_2)} - \lambda\xi(\mu_1 - \mu_2) & \text{für } |t| < t_1(\theta + \mu_2), \\ \frac{\mu_2 t_1^2}{2} + \frac{1}{2\theta}(\mu_2 t_1 - t)^2 - \lambda\xi(\mu_1 - \mu_2) & \text{für } |t| \in t_1\mu_2 + \theta[t_1, t_2], \\ \frac{|t|^2}{2(\mu_1 + \theta)} - \frac{\mu_1 t_2^2}{2} + \frac{\mu_2 t_1^2}{2} - \lambda\xi(\mu_1 - \mu_2) & \text{für } t_2(\mu_1 + \theta) < |t|. \end{cases} \quad (4.2)$$

*Beweis.* Sei  $\varphi^*$  wie in (3.6) gegeben. Mit Definition 4.1 und  $h_\theta(t, z) := \varphi^*(z) + \frac{1}{2\theta}|z - t|^2$  folgt

$$\varphi_\theta^*(t) = \inf_{z \in \mathbb{R}^n} h_\theta(t, z) \quad \text{und} \quad 0 \in \frac{\partial}{\partial z} h_\theta(t, z) = \partial\varphi^*(z) + \theta^{-1}(z - t)$$

Daraus folgt, dass das Infimum bei

$$\begin{cases} z = \frac{\mu_2 t}{\theta + \mu_2} & \text{für } |t| < t_1(\theta + \mu_2), \\ z \in t - \theta[t_1, t_2] & \text{für } |t| \in \mu_2 t_1 + \theta[t_1, t_2], \\ z = \frac{\mu_1 t}{\theta + \mu_1} & \text{für } t_2(\theta + \mu_1) < |t| \end{cases}$$

angenommen wird. □

**Lemma 4.5.** *Der punktweise Fehler der Regularisierung  $|\varphi_\theta^*(t) - \varphi^*(t)|$  ist von der Ordnung  $o(\theta)$ .*

*Beweis.* Wir betrachten die konjugierte Energiedichte  $\varphi^*(t)$  und deren Regularisierung  $\varphi_\theta^*(t)$  für  $t \in \mathbb{R}_+$ . Es ergeben sich vier Fälle, die zu unterscheiden sind: (a)  $t < \mu_1 t_2$ , (b)  $t \in [\mu_2 t_1, t_1(\mu_2 + \theta)]$ , (c)  $t \in t_1 \mu_2 + \theta[t_1, t_2]$  und (d)  $t > t_1 \mu_2 + \theta t_2$ . Für die auftretenden Parameter gilt  $\theta > 0$ ,  $0 < \mu_1 < \mu_2 < \infty$ ,  $0 < t_1 < t_2 < \infty$  und  $\mu_1 t_2 = \mu_2 t_1$ .

(a) Sei  $t < \mu_1 t_2$  fixiert.

$$|\varphi_\theta^*(t) - \varphi^*(t)| = \left| \frac{t^2}{2(\theta + \mu_2)} - \frac{t^2}{2\mu_2} \right| = t^2 \frac{\theta}{2\mu_2(\theta + \mu_2)} \leq t_1^2 \theta = o(\theta)$$

(b) Sei  $t \in [\mu_2 t_1, t_1(\mu_2 + \theta)]$  fixiert.

$$|\varphi_\theta^*(t) - \varphi^*(t)| = \left| \frac{t^2}{2(\theta + \mu_2)} - \frac{t^2}{2\mu_1} + \frac{\mu_1 t_2^2}{2} - \frac{\mu_2 t_1^2}{2} \right| \leq \left| \frac{t_1^2(\mu_2 + \theta)}{2} - \frac{\mu_2 t_1^2}{2} \right| = o(\theta)$$

(c) Sei  $t \in t_1 \mu_2 + \theta[t_1, t_2]$  fixiert.

$$\begin{aligned} |\varphi_\theta^*(t) - \varphi^*(t)| &= \left| \frac{1}{2\theta}(t - \mu_2 t_1)^2 + \frac{\mu_2 t_1^2}{2} - \frac{t^2}{2\mu_1} + \frac{\mu_1 t_2^2}{2} - \frac{\mu_2 t_1^2}{2} \right| \\ &\leq \left| \frac{t_2^2 \theta}{2} - \frac{(\mu_1 t_2 + t_1 \theta)^2}{2\mu_1} + \frac{\mu_1 t_2^2}{2} \right| \\ &= \left| \frac{t_2^2 \theta}{2} - \theta t_2 t_1 - \theta^2 t_1^2 \right| = o(\theta) \end{aligned}$$

(d) Sei  $t > t_1 \mu_2 + \theta t_2$  fixiert.

$$\begin{aligned} |\varphi_\theta^*(t) - \varphi^*(t)| &= \left| \frac{t^2}{2(\mu_1 + \theta)} - \frac{\mu_1 t_2^2}{2} + \frac{\mu_2 t_1^2}{2} - \frac{t^2}{2\mu_1} + \frac{\mu_1 t_2^2}{2} - \frac{\mu_2 t_1^2}{2} \right| \\ &= t^2 \frac{\theta}{2\mu_1(\theta + \mu_1)} \leq t^2 \frac{\theta}{2\mu_1^2} = o(\theta) \end{aligned} \quad \square$$

Mit Hilfe von Lemma 4.4 können wir die starke Form des regularisierten Problems von (3.21) aufschreiben

$$\begin{cases} -\operatorname{div} \sigma_\theta = f & \text{in } \Omega, \\ \nabla u = D\varphi_\theta^*(\sigma_\theta) & \text{in } \Omega. \end{cases} \quad (4.3)$$

Multiplizieren mit entsprechenden Testfunktionen und partieller Integration liefert dann die schwache Formulierung des zu lösenden Sattelpunktproblems

**Definition 4.6** (Schwache Formulierung). *Seien eine stückweise konstante Last  $f \in L^2(\Omega)$  und homogene Randdaten  $u = 0$  auf  $\partial\Omega$  gegeben. Für  $\theta > 0$  finde  $(\sigma_\theta, u) \in H(\operatorname{div}; \Omega) \times L^2(\Omega)$ , so dass für alle  $(\tau, v) \in H(\operatorname{div}; \Omega) \times L^2(\Omega)$  gilt*

$$(\operatorname{div} \sigma_\theta + f, v)_{L^2(\Omega)} = 0 \quad \text{und} \quad (u, \operatorname{div} \tau)_{L^2(\Omega)} + (D\varphi_\theta^*(\sigma_\theta), \tau)_{L^2(\Omega)} = 0. \quad (4.4)$$

**Bemerkung 4.7.** Die regularisierte Energieformulierung ist dann analog zu (3.22) gegeben durch: Finde  $\sigma_\theta \in Q_f = \{\tau \in H(\text{div}; \Omega) \mid -\text{div } \tau = f, f \in L^2(\Omega)\}$ , so dass das Minimum

$$\min_{\tau \in Q_f} \int_{\Omega} \varphi_\theta^*(\tau) dx$$

angenommen wird.

**Bemerkung 4.8.** Die Verschiebung  $u$  wird durch die zweite Gleichung in (4.4) ebenfalls von der Regularisierung beeinflusst. Um die Lesbarkeit zu erleichtern, verzichten wir aber darauf, dies noch in einem Index an  $u$  deutlich zu machen. Wir beschränken uns darauf, die Abhängigkeit bei der regularisierten Spannung  $\sigma_\theta$  hervorzuheben.

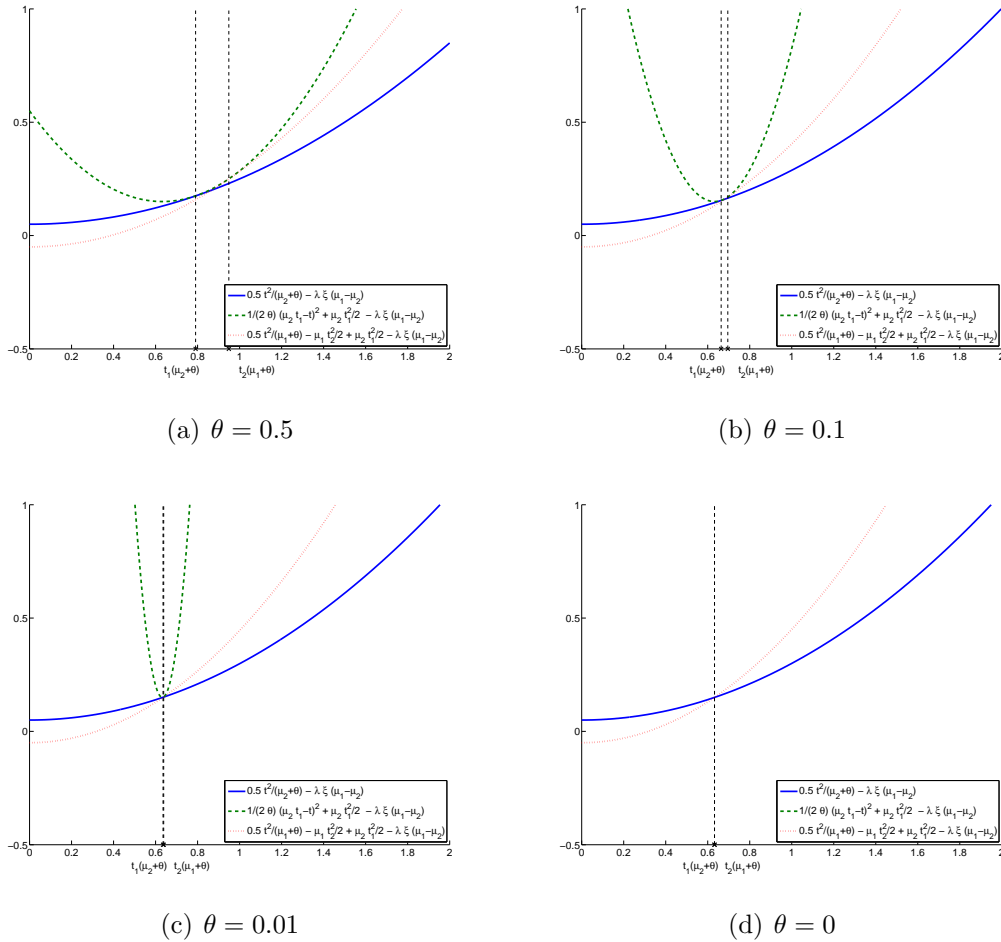


Abbildung 4.2: Darstellung der Regularisierung von  $\varphi^*$  nach Lemma 4.4 für verschiedene Werte des Regularisierungsparameters  $\theta$ . Die Materialparameter sind auf  $\mu_1 = 1, \mu_2 = 2, \xi = 0.5$  und  $\lambda = 0.1$  gesetzt.

### 4.3 Die konjugierte Energiedichte

Im Folgenden arbeiten wir verschiedene Eigenschaften von  $\varphi_\theta^*$  für  $\theta \geq 0$  heraus. Zur Vervollständigung wird hierfür noch mal die Definition von  $\varphi_\theta^*(t)$  genannt

$$\varphi_\theta^*(t) = \begin{cases} \frac{|t|^2}{2(\theta+\mu_2)} - \lambda\xi(\mu_1 - \mu_2) & \text{für } |t| < t_1(\theta + \mu_2), \\ \frac{\mu_2}{2}t_1^2 + \frac{1}{2\theta}(\mu_2 t_1 - t)^2 - \lambda\xi(\mu_1 - \mu_2) & \text{für } |t| \in t_1\mu_2 + \theta[t_1, t_2], \\ \frac{|t|^2}{2(\mu_1+\theta)} - \frac{\mu_1 t_2^2}{2} + \frac{\mu_2 t_1^2}{2} - \lambda\xi(\mu_1 - \mu_2) & \text{für } t_2(\mu_1 + \theta) < |t|. \end{cases} \quad (4.5)$$

**Lemma 4.9** (Strenge Monotonie). *Das Subdifferential der konjugierten Energiedichte  $\partial\varphi_\theta^*$  ist streng monoton für alle  $\theta \geq 0$ . D.h., für  $\tilde{a} \in \partial\varphi_\theta^*(a)$  und  $\tilde{b} \in \partial\varphi_\theta^*(b)$  mit  $C(\theta) = \frac{1}{\mu_2+\theta}$  gilt*

$$C(\theta)(a-b)^2 < (\tilde{a} - \tilde{b})(a-b) \quad (4.6)$$

*Beweis.* Sei  $\theta \geq 0$  fixiert. Für alle  $\tilde{a} \in \partial\varphi_\theta^*(a)$  und  $\tilde{b} \in \partial\varphi_\theta^*(b)$  existieren  $\alpha, \beta \in \left[\frac{1}{\mu_2+\theta}, \frac{1}{\mu_1+\theta}\right]$  mit  $\tilde{a} = \alpha a \in \partial\varphi_\theta^*(a)$  und  $\tilde{b} = \beta b \in \partial\varphi_\theta^*(b)$ . Zusätzlich gilt für  $|a| \leq |b|$  die Abschätzung  $|\tilde{a}| \leq |\tilde{b}|$ . Sei nun  $\gamma := \min\{\alpha, \beta\} \geq C(\theta)$ . Dann erhalten wir für  $|a| \neq |b|$  folgende zwei Ungleichungen

$$\begin{aligned} 0 &\leq 0.5(\beta - \gamma)(a-b)^2 \leq (\beta - \gamma)(b^2 - ab) && \text{für } |a| \leq |b|, \\ 0 &\leq 0.5(\alpha - \gamma)(a-b)^2 \leq (\alpha - \gamma)(a^2 - ab) && \text{für } |b| \leq |a|. \end{aligned}$$

Für  $\tilde{a} \in \partial\varphi_\theta^*(a)$  und  $\tilde{b} \in \partial\varphi_\theta^*(b)$  und  $|a| \neq |b|$  erhalten wir nun

$$\begin{aligned} C(\theta)(a-b)^2 &< \gamma(a-b)^2 + (\beta - \gamma)(b^2 - ab) + (\alpha - \gamma)(a^2 - ab) \\ &\leq \gamma(a^2 - 2ab + b^2 + ab - b^2 - a^2 + ab) + \beta(b^2 - ab) + \alpha(a^2 - ab) \\ &= (\alpha a - \beta b)(a-b) = (\tilde{a} - \tilde{b})(a-b). \end{aligned} \quad \square$$

**Bemerkung 4.10.** *Aus der strengen Monotonie folgt mit  $b = 0$  die Koerzivität von  $D\varphi_\theta^*$  für  $\theta > 0$ , d.h.*

$$\lim_{\|x\|_{H^1(\Omega)} \rightarrow \infty} \frac{(D\varphi_\theta^*(x), x)_{H^1(\Omega)}}{\|x\|_{H^1(\Omega)}} = \infty. \quad (4.7)$$

**Lemma 4.11** (Regularität I). *Die konjugierte Funktion  $\varphi_\theta^*(t) |_{\theta=0}$  ist stetig und strikt konvex.*

*Beweis.* Die Funktion  $\varphi^* = \varphi_\theta^* |_{\theta=0}$  ist gegeben durch

$$\varphi^*(t) = \begin{cases} \frac{|t|^2}{2\mu_2} - \lambda\xi(\mu_1 - \mu_2) & \text{für } |t| < t_1\mu_2, \\ \frac{\mu_2}{2}t_1^2 - \lambda\xi(\mu_1 - \mu_2) & \text{für } |t| = t_1\mu_2, \\ \frac{|t|^2}{2\mu_1} - \frac{\mu_1 t_2^2}{2} + \frac{\mu_2 t_1^2}{2} - \lambda\xi(\mu_1 - \mu_2) & \text{für } t_2\mu_1 < |t|. \end{cases}$$

Mit  $\mu_1 t_2 = \mu_2 t_1$  ergibt sich für  $t \uparrow t_1 \mu_2$  und  $t \downarrow t_2 \mu_1$  die Stetigkeit von  $\varphi^*$ . Das Subdifferential  $\partial \varphi^*$  ist gegeben durch

$$\partial \varphi^*(t) = \begin{cases} \frac{t}{\mu_2} & \text{für } |t| < t_1 \mu_2, \\ \frac{t}{t_1 \mu_2} [t_1, t_2] & |t| = t_1 \mu_2, \\ \frac{t}{\mu_1} & \text{für } t_2 \mu_1 < |t|. \end{cases}$$

Da das Subdifferential  $\partial \varphi^*$  nach (4.6) streng monoton ist, ist die konjugierte Energiedichte streng konvex.  $\square$

**Lemma 4.12** (Regularität II). *Die regularisierte konjugierte Funktion  $\varphi_\theta^*(t)$  ist für  $\theta > 0$  stetig differenzierbar und strikt konvex.*

*Beweis.* Aus der Definition von  $\varphi_\theta^*$  (4.5) folgt

$$\varphi^*(t) = \begin{cases} \frac{t_1^2(\theta + \mu_2)}{2} - \lambda \xi(\mu_1 - \mu_2) & \text{für } |t| \uparrow t_1(\theta + \mu_2) \\ \frac{\mu_2}{2} t_1^2 + \frac{t_1^2 \theta}{2} - \lambda \xi(\mu_1 - \mu_2) & \text{für } |t| \downarrow t_1(\theta + \mu_2) \end{cases}$$

und

$$\varphi^*(t) = \begin{cases} \frac{\mu_2}{2} t_1^2 + \frac{t_2^2 \theta}{2} - \lambda \xi(\mu_1 - \mu_2) & \text{für } |t| \uparrow t_2(\theta + \mu_1) \\ \frac{t_2^2(\theta + \mu_1)}{2} - \frac{\mu_1 t_2^2}{2} + \frac{\mu_2 t_1^2}{2} - \lambda \xi(\mu_1 - \mu_2) & \text{für } |t| \downarrow t_2(\theta + \mu_1). \end{cases}$$

Damit ist  $\varphi_\theta^*(t)$  stetig für alle  $\theta \geq 0$ . Die Ableitung  $D\varphi_\theta^*$  ist gegeben durch

$$D\varphi_\theta^*(t) = \begin{cases} \frac{t}{(\theta + \mu_2)} & \text{für } |t| < t_1(\theta + \mu_2), \\ \frac{t - \mu_2 t_1}{\theta} & \text{für } |t| \in t_1 \mu_2 + \theta [t_1, t_2], \\ \frac{t}{(\mu_1 + \theta)} & \text{für } t_2(\mu_1 + \theta) < |t|, \end{cases}$$

wodurch die Stetigkeit der ersten Ableitung direkt folgt. Damit ist  $\varphi_\theta^*$  stetig differenzierbar für alle  $\theta > 0$ . Aufgrund der strengen Monotonie von  $D\varphi_\theta^*$  folgte die strenge Konvexität von  $\varphi_\theta^*$ .  $\square$

# Kapitel 5

## Finite Elemente Methode

In diesem Kapitel betrachten wir die diskrete gemischte Formulierung von (4.4). Wir führen hierfür die Raviart-Thomas Räume ein. Es werden notwendige Begriffe wie Triangulierungen und Basisfunktionen erläutert und abschließend das diskrete Problem beschrieben.

### 5.1 Die Triangulierung

**Definition 5.1** (Dreieck). *Ein nicht entartetes Dreieck  $T \subset \mathbb{R}^2$  ist durch drei affin unabhängige Punkte  $P_1, P_2, P_3 \in \mathbb{R}^2$  durch*

$$T := \text{conv}\{P_1, P_2, P_3\} \quad (5.1)$$

*gegeben. Die Ecken  $P_j$  sind hierbei so durchnummeriert, so dass*

$$|T| := \frac{1}{2} \det \begin{pmatrix} 1 & 1 & 1 \\ P_1 & P_2 & P_3 \end{pmatrix} > 0, \quad (5.2)$$

*wobei  $|T|$  die Fläche des Dreiecks ist.*

**Bemerkung 5.2.** *Um die Bedingung (5.2) zu erfüllen, genügt es das Dreieck entgegen des Uhrzeigersinns durch zu nummerieren.*

Im Folgenden sei  $\Omega \subset \mathbb{R}^2$  eine beschränkte, offene und zusammenhängende Menge mit stückweisen affinem Rand, so dass  $\bar{\Omega}$  durch eine endliche Vereinigung nicht-entarteter Dreiecke beschrieben werden kann.

**Definition 5.3** (Triangulierung). *Für eine Menge von Dreiecke  $\{T_j\}$  ist eine Triangulierung  $\mathcal{T}$  von  $\Omega$  ist gegeben durch*

$$\begin{cases} \mathcal{T} := \{T_1, T_2, T_3, \dots, T_n\} \\ \bar{\Omega} = \cup \mathcal{T} := \cup_{j \in \mathbb{N}} T_j \\ |T_j \cap T_k| = 0 \text{ für } j \neq k. \end{cases} \quad (5.3)$$



Die Menge der Knoten  $\mathcal{N}$  und die Menge der Kanten  $\mathcal{E}$  einer Triangulierung  $\mathcal{T}$  werden mit  $\mathcal{N}$  bzw. mit  $\mathcal{E}$  bezeichnet. Die Kanten, welche den Rand von  $\Omega$  beschreiben, bilden eine Menge  $\mathcal{E}_D \subset \mathcal{E}$ . Der Durchmesser eines Dreiecks wird mit  $h_T$  bezeichnet. Die Gitterweite  $h_{\mathcal{T}}$  der Triangulierung  $\mathcal{T}$  ist dann gegeben durch  $h_{\mathcal{T}} := \max_{T \in \mathcal{T}} h_T$ .

**Bemerkung 5.4.** Aus der Definition 5.3 folgt, dass der Schnitt zweier Dreiecke  $T_1, T_2 \in \mathcal{T}$  mit  $T_1 \neq T_2$  entweder leer ist, ein Knoten  $N \in \mathcal{N}$  oder eine Kante  $E \in \mathcal{E}$  ist.

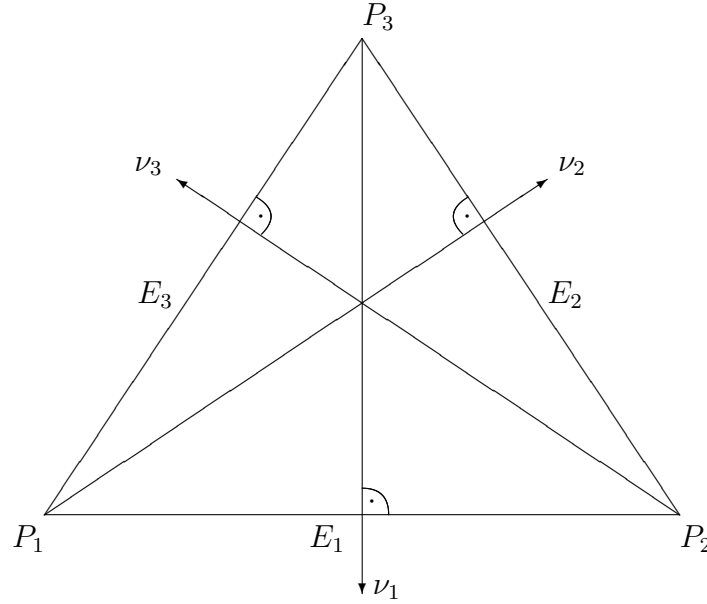


Abbildung 5.1: Dreieck  $T = \text{conv}\{P_1, P_2, P_3\}$  mit Knoten  $P_1, P_2, P_3$  und Kanten  $E_1 = \text{conv}\{P_1, P_2\}$ ,  $E_2 = \text{conv}\{P_2, P_3\}$ ,  $E_3 = \text{conv}\{P_3, P_1\}$  mit Längen  $|E_1|$ ,  $|E_2|$ ,  $|E_3|$ .

Im späteren Verlauf werden häufig Polynomräume auftreten, die wir hier definieren wollen.

**Definition 5.5** (Polynomräume). Die stückweisen Polynomräume  $\mathcal{P}_k(\mathcal{T})$  und global stetigen Polynomräume  $\mathcal{S}_k(\mathcal{T})$  sind wie folgt definiert

$$\begin{aligned} \mathcal{P}_k(\mathcal{T}) &:= \{\phi \in L^2(\mathcal{T}) \mid T \in \mathcal{T} : \phi|_T \text{ ist ein Polynom vom Grad } k.\}, \\ \mathcal{S}_k(\mathcal{T}) &:= \mathcal{P}_k(\mathcal{T}) \cap C^0(\Omega). \end{aligned}$$

## 5.2 Das diskrete Problem

Nachdem nun der Begriff der Triangulierung definiert wurde, führen wir nun die Raviart-Thomas Räume und die Basisfunktionen ein, die wir später zur numerischen Realisierung benötigen.

**Definition 5.6** (Raviart-Thomas). *Sei eine Triangulierung  $\mathcal{T}$  von  $\Omega$  gegeben. Die Raviart-Thomas Räume sind definiert durch*

$$\begin{aligned} RT_0(\mathcal{T}) &:= \{ \tau \in H(\text{div}; \mathcal{T}) \mid \forall T \in \mathcal{T} \exists a \in \mathbb{R}^2 \exists b \in \mathbb{R} : \\ &\quad \forall x \in T \tau(x) = ax + b \}, \\ \mathcal{P}_0(\mathcal{T}) &:= \{ v \in L^2(\mathcal{T}) \mid v \text{ ist konstant auf } T \}. \end{aligned} \quad (5.4)$$

**Definition 5.7** (lokale Basisfunktionen). *Sei ein Dreieck  $T$  mit Kanten  $E_j$  gegeben. Der Knoten  $P = P(E_j)$  sei der gegenüberliegende Knoten der Kante  $E_j$ . Es seien  $\nu_{E_j}$  die Einheitsnormale der Kante  $E_j$  und  $\nu_j$  die äußere Einheitsnormale des Dreiecks  $T$  an der Kante  $E_j$ . Wir definieren dann mit  $\text{sgn}_j := \nu_{E_j} \cdot \nu_j$  die Basisfunktion  $\psi_{E_j}$  durch*

$$\psi_{E_j} := \text{sgn}_j \frac{|E_j|}{2|T|} (x - P). \quad (5.5)$$

**Bemerkung 5.8.** *Die Orientierung der Einheitsnormalen  $\nu_{E_j}$  in Definition 5.7 sei global fixiert.*

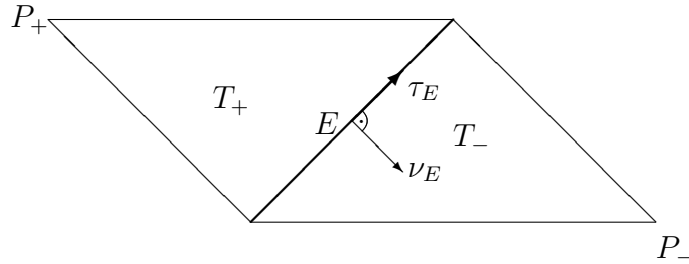


Abbildung 5.2: Zwei benachbarte Dreiecke  $T_+$  und  $T_-$ , die sich eine Kante  $E = \partial T_+ \cap \partial T_-$  teilen. Die Orientierung des Normalenvektors  $\nu_E$  und des Tangentialvektors  $\tau_E$  der Kante  $E$  ist global fixiert.

**Definition 5.9** (globale Basisfunktionen). *Für eine fixierte Kante  $E \in \mathcal{E}$ , sei  $\psi_E$  wie folgt definiert*

$$\psi_E := \begin{cases} \pm \frac{|E|}{2|T|} (x - P_{\pm}) & \text{für } x \in T_{\pm}, \\ 0 & \text{sonst.} \end{cases} \quad (5.6)$$

**Lemma 5.10.** *Für eine fixierte Kante  $E \in \mathcal{E}$  betrachten wir die Funktion  $\Psi_E$  nach Definition 5.9. Dann gilt*

$$(a) \quad \psi_E \cdot \nu_E = \begin{cases} 0 & \text{für alle Kanten in } \mathcal{E} \setminus E, \\ 1 & \text{entlang } E, \end{cases}$$

- (b)  $\psi_E \in H(\operatorname{div}; \Omega)$ ,
- (c)  $(\psi_E \mid E \in \mathcal{E})$  ist eine Basis von  $RT_0(\mathcal{T})$ ,
- (d)  $\operatorname{div} \psi_E = \begin{cases} \pm \frac{|E|}{|T|} & \text{auf } T_{\pm}, \\ 0 & \text{sonst.} \end{cases}$

*Beweis.* Der Beweis wird in [1] geführt.

**Bemerkung 5.11.** Aus der globalen Definition der Basisfunktionen folgt, dass für ein  $\tau \in RT_0(\mathcal{T})$  der  $\operatorname{curl} \tau$  und der Sprung von  $\tau$  über eine Kante  $E$  in Normalenrichtung verschwindet. Beachte der Sprung in Tangentialrichtung hingegen muss nicht verschwinden.

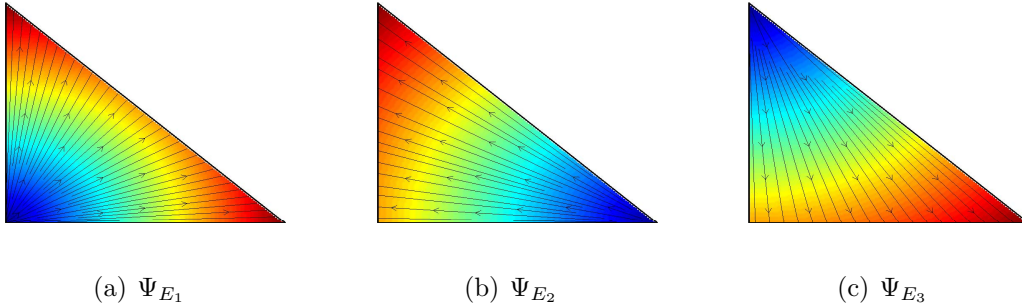


Abbildung 5.3: Visualisierung der Raviart-Thomas Basisfunktionen  $\Psi_{E_j}$  auf dem Referenzelement  $T_{ref}$ . Die Norm der Basisfunktionen wird durch den Farbverlauf wiedergegeben. Blau repräsentiert einen kleinen Normwert, während Rot einen hohen Normwert darstellt. Entnommen aus [9].

**Definition 5.12** (diskrete Formulierung). Das diskrete Problem zu (4.4) ist gegeben durch: Finde  $(\sigma_{\theta,h}, u_h) \in RT_0(\mathcal{T}) \times \mathcal{P}_0(\mathcal{T})$ , so dass für alle  $(\tau_h, v_h) \in RT_0(\mathcal{T}) \times \mathcal{P}_0(\mathcal{T})$  gilt

$$(\operatorname{div} \sigma_{\theta,h} + f, v_h)_{L^2(\Omega)} = 0 \quad \text{und} \quad (u_h, \operatorname{div} \tau_h)_{L^2(\Omega)} + (D\varphi_{\theta}^*(\sigma_{\theta,h}), \tau_h)_{L^2(\Omega)} = 0. \quad (5.7)$$

**Bemerkung 5.13.** Für stückweise konstante Lasten  $f$  wird die diskrete Energieerhaltung  $-\operatorname{div} \sigma_{\theta,h} = f$  lokal exakt erfüllt.

**Bemerkung 5.14.** Die diskrete Energieformulierung von (5.7) ist dann analog zu (3.22) gegeben durch: Finde  $\sigma_{\theta,h} \in RT_0(\mathcal{T}) \cap Q_{f_h}$ , so dass das Minimum

$$\min_{\tau \in RT_0(\mathcal{T}) \cap Q_{f_h}} \int_{\Omega} \varphi_{\theta}^*(\tau) \, dx \quad (5.8)$$

angenommen wird.

# Kapitel 6

## Existenz und Eindeutigkeit

In diesem Kapitel zeigen wir die Existenz und Eindeutigkeit der Lösung für (3.22) und (5.7).

### 6.1 Die exakte Lösung

**Theorem 6.1.** *Das Problem (3.22) besitzt eine eindeutige Lösung.*

*Beweis.* Die konjugierte Energiedichte  $\varphi^*$  ist streng konvex. Damit hat die Minimierungsaufgabe über den affinen Unterraum  $Q_f \subset H(\operatorname{div}; \Omega)$  in (3.22) einen eindeutigen Minimierer  $\sigma \in Q_f$ , so dass  $\varphi^*(\sigma) < \varphi^*(\tau)$  für alle  $\tau \in Q_f$  gilt.  $\square$

**Bemerkung 6.2.** *Sei  $\sigma$  die Lösung von (3.22). Dann muss  $\sigma$  folgende Bedingung erfüllen*

$$0 \in \partial(\varphi^*(\sigma) + \chi_{Q_f}(\sigma)) = \partial\varphi^*(\sigma) + \partial\chi_{Q_f}(\sigma).$$

### 6.2 Die diskrete Lösung

**Theorem 6.3.** *Das Problem (5.8) besitzt eine eindeutige Lösung.*

*Beweis.* Da der Divergenzoperator linear und beschränkt ist, ist der Unterraum  $Q_{f_h}$  als Urbild einer abgeschlossenen Teilmenge abgeschlossen. Der Schnitt  $Q_{f_h} \cap RT_0(\mathcal{T})$  ist ein abgeschlossener affiner endlichdimensionaler Unterraum und damit konvex. Da  $\varphi^*$  strikt konvex ist, existiert somit ein eindeutiger Minimierer.  $\square$

**Theorem 6.4.** *Sei  $\sigma_{\theta,h}$  Lösung von (5.8). Dann existiert ein eindeutiges  $u_h$ , so dass  $(\sigma_{\theta,h}, u_h)$  Lösung von (5.7) für  $\theta \geq 0$  ist.*

*Beweis.* Sei  $\sigma_{\theta,h}$  gegeben und  $M_h^0 := RT_0(\mathcal{T}) \cap Q_0$ . Da  $\sigma_{\theta,h}$  (5.8) minimiert, gilt  $\sigma_{\theta,h} \in Q_{f_h}$  und  $0 \in \partial(\varphi^*(\sigma_{\theta,h}) + \chi_{Q_{f_h}}(\sigma_{\theta,h}))$ . Folglich existiert ein  $\xi_h \in \partial\varphi^*(\sigma_{\theta,h})$  und  $-\xi_h \in \partial\chi_{Q_{f_h}}(\sigma_{\theta,h})$ .

(a) Es gilt:  $M_h^0 \subseteq \ker(\xi_h)$ .

Aus der Definition des Subdifferentials folgt, dass  $(-\xi_h, \tau - \sigma_{\theta,h})_{L^2(\Omega)} \leq 0$  für alle  $\tau \in Q_{f_h}$  gelten muss. Mit  $(2\sigma_{\theta,h} - \tau) \in Q_{f_h}$  folgt  $(-\xi_h, \tau - \sigma_{\theta,h})_{L^2(\Omega)} \geq 0$ , und somit  $\xi_h \perp_2 (\sigma_{\theta,h} - \tau)$  für alle  $\tau \in M_h^0$ . Folglich gilt  $\xi_h \perp_2 M_h^0$ .

(b) Es existiert ein eindeutiges  $u_h \in \mathcal{P}_0(\mathcal{T})$  mit  $(\xi_h, \cdot)_{L^2(\Omega)} = (u_h, \operatorname{div} \cdot)_{L^2(\Omega)} \in RT_0(\mathcal{T})^*$ .

Die Bilinearform  $b : RT_0(\mathcal{T}) \times \mathcal{P}_0(\mathcal{T}) \rightarrow \mathbb{R}$  definiert durch  $b(x, y) := (y, \operatorname{div} x)_{L^2(\Omega)}$  erfüllt die Brezzi-Bedingung, d.h.

$$\sup_{x \neq 0} \frac{b(x, y)}{\|x\|} \geq C \|y\| \text{ für } C > 0.$$

Die Operatoren  $B : (M_h^0)^\perp := RT_0(\mathcal{T}) / M_h^0 \rightarrow \mathcal{P}_0(\mathcal{T})$ , definiert durch  $Bx := b(x, \cdot)_{L^2(\Omega)} = (\cdot, \operatorname{div} x)_{L^2(\Omega)}$ , und  $B^* : \mathcal{P}_0(\mathcal{T}) \rightarrow ((M_h^0)^\perp)^*$ , definiert durch  $B^*y := b(\cdot, y)_{L^2(\Omega)} = (y, \operatorname{div} \cdot)_{L^2(\Omega)}|_{(M_h^0)^\perp}$  sind Isomorphismen.

Nach dem Darstellungssatz von Riesz existiert demzufolge ein eindeutiges  $u_h \in \mathcal{P}_0(\mathcal{T})$ , so dass für alle  $\xi_h \in (M_h^0)^\perp$  gilt:

$$B^*u_h(\tau_h) = (u_h, \operatorname{div} \tau_h)_{L^2(\Omega)} = (\xi_h, \tau_h)_{L^2(\Omega)}.$$

Schliesslich löst  $(\sigma_{\theta,h}, u_h)$  das Problem (5.7). □

# Kapitel 7

## Fehlerabschätzungen

In diesem Kapitel führen wir die a priori und die a posteriori Analysis. Mit Hilfe der a priori Analysis sind wir dann in der Lage zu sagen, was das optimale Konvergenzverhalten bzgl einer bestimmten Norm ist. Die a posteriori Analysis motiviert dann zwei zuverlässige Fehlerschätzer. Mit Hilfe von Fortin- und Clémentinterpolatoren werden die Abschätzungen bewiesen. Für die a posteriori Abschätzung wird zudem noch die Helmholtz-Zerlegung benötigt.

### 7.1 A priori Fehlerabschätzung

Für die a priori Abschätzung wird der Fortin Interpolator benötigt. Wir folgen hier den Ausführungen von [7].

**Definition 7.1** (Fortin Interpolation). *Sei  $\phi \in H(\text{div}; \Omega)$  ein gegebenes Vektorfeld und  $T \in \mathcal{T}$ . Die Fortin Interpolation  $I\phi$  ist definiert durch  $I : H(\text{div}; \Omega) \rightarrow RT_0(\mathcal{T})$  mit*

- (a)  $((I\phi - \phi) \cdot \nu_E, w_h)_{L^2(E)} = 0$  für alle  $w_h \in \mathcal{P}_1(E)$  und  $E \in \partial T$ ,
- (b)  $((I\phi - \phi), v_h)_{L^2(T)} = 0$  für alle  $v_h \in \mathcal{P}_0(T)$ .

**Lemma 7.2** (Eigenschaften des Fortin Interpolators). *Sei  $I$  der Fortin Interpolator und  $\phi \in H(\text{div}; \Omega)$ , dann gelten folgende Aussagen*

- (a)  $\|I\phi - \phi\|_{L^2(T)} \lesssim h_T |\phi|_{H^1(T)}$ ,
- (b) Sei  $\Pi : L^2(\Omega) \rightarrow \mathcal{P}_0(\mathcal{T})$  die  $L^2$ -Projektion. Es gilt:  $\text{div } I\phi = \Pi \text{div } \phi$ ,
- (c)  $\|\text{div}(I\phi - \phi)\|_{L^2(T)} \lesssim h_T |\text{div } \phi|_{L^2(T)}$ .

*Beweis.* Der Beweis wird in [7] geführt.

**Bemerkung 7.3.** Aus Definition 7.1 folgt für  $w_h \in \mathcal{P}_0(\mathcal{T})$  mit dem Gaußschen Integralsatz die  $L^2$ -Orthogonalität  $\text{div}(I\phi - \phi) \perp_2 \mathcal{P}_0(\mathcal{T})$ .

**Theorem 7.4.** *Sei die konjugierte Energiedichte  $\partial\varphi^*$  wie oben definiert. Für stückweise konstante Lasten  $f$  und homogenen Randbedingungen  $u = 0$  auf  $\partial\Omega$  gilt*

$$\|\sigma - \sigma_h\|_{L^2(\Omega)}^2 \lesssim h_{\mathcal{T}}^{3/2} |\sigma|_{H^1(\Omega)} + h_{\mathcal{T}}^2 |\sigma|_{H^1(\Omega)}^2. \quad (7.1)$$

*Beweis.* Gegeben die Fortin Interpolation  $I\sigma$ , eine stückweise konstante Last  $f$  und homogene Randdaten. Mit Hilfe der strengen Monotonie von  $\varphi^*$  und  $\operatorname{div}(I\sigma - \sigma) = f_h - f = 0$  erhält man

$$\begin{aligned} \|\sigma - \sigma_h\|_{L^2(\Omega)}^2 &\leq (\partial\varphi^*(\sigma) - \partial\varphi^*(\sigma_h), \sigma - \sigma_h)_{L^2(\Omega)} \\ &= (\partial\varphi^*(\sigma_h), I\sigma - \sigma)_{L^2(\Omega)} - (u_h, \operatorname{div}(\sigma_h - I\sigma))_{L^2(\Omega)} \\ &= (\partial\varphi^*(\sigma_h), I\sigma - \sigma)_{L^2(\Omega)} + (u, \operatorname{div}(I\sigma - \sigma))_{L^2(\Omega)} \\ &= (\partial\varphi^*(\sigma_h) - \partial\varphi^*(\sigma), I\sigma - \sigma)_{L^2(\Omega)}. \end{aligned}$$

Wir definieren  $g := |\partial\varphi^*(\sigma_h) - \partial\varphi^*(\sigma)|$ . Bemerkung 3.7, die Cauchy-Schwarz Ungleichung und die Eigenschaft des Fortin Interpolators führen dann zu

$$\begin{aligned} \|\sigma - \sigma_h\|_{L^2(\Omega)}^2 &\leq \|g\|_{L^2(\Omega)} \|I\sigma - \sigma\|_{L^2(\Omega)} \\ &\lesssim (\|C(\sigma, \sigma_h)\|_{L^2(\Omega)} + \|\sigma - \sigma_h\|_{L^2(\Omega)}) h_{\mathcal{T}} |\sigma|_{H^1(\Omega)}. \end{aligned}$$

Mit  $\|C(\sigma, \sigma_h)\|_{L^2(\Omega)} \leq h_{\mathcal{T}}^{1/2} (t_2 - t_1)$  erhält man

$$\|\sigma - \sigma_h\|_{L^2(\Omega)}^2 \lesssim h_{\mathcal{T}}^{3/2} |\sigma|_{H^1(\Omega)} + h_{\mathcal{T}} \|\sigma - \sigma_h\|_{L^2(\Omega)} |\sigma|_{H^1(\Omega)}.$$

Schliesslich führt die Youngsche Ungleichung,  $2ab \leq a^2 + b^2$ , zu

$$\|\sigma - \sigma_h\|_{L^2(\Omega)}^2 \lesssim h_{\mathcal{T}}^{3/2} |\sigma|_{H^1(\Omega)} + h_{\mathcal{T}}^2 |\sigma|_{H^1(\Omega)}^2,$$

womit die Behauptung bewiesen ist □.

## 7.2 A posteriori Fehlerabschätzung

Wir werden in diesem Abschnitt zwei Fehlerschätzer motivieren, die eine adaptive Gittergenerierung ermöglichen. Hierfür wird zum einen eine Helmholtz-Zerlegung benötigt. Hiermit kann man ein Vektorfeld in dessen Gradienten- und Rotationsfeld zerlegen. Um die Zuverlässigkeit der Fehlerschätzer zu zeigen, führen wir den Clément-Interpolator ein, der eine lokale Regularisierung der Funktion beschreibt.

**Definition 7.5** (Clément Interpolation). *Für  $z \in \mathcal{N}$  sei  $\omega_z := \{T \in \mathcal{T} \mid z \in T\}$  der Patch des Knoten  $z$  und  $\mathcal{S}_1(\mathcal{T})$  der Raum der stückweisen affinen, global stetigen Funktionen auf  $\mathcal{T}$ . Wir definieren die nodale Funktion  $\phi_z \in \mathcal{S}_1(\mathcal{T})$  durch  $\phi_z(z) = 1$  für*

$z \in \mathcal{N}$  und  $\phi_z(y) = 0$  für alle  $y \in \mathcal{N} \setminus \{z\}$ . Der Clément Interpolator  $J : H^1(\Omega) \rightarrow \mathcal{S}_1(\mathcal{T})$  ist dann wie folgt gegeben

$$J(\beta) := \sum_{z \in \mathcal{N}} \beta_z \phi_z \text{ mit } \beta_z := \begin{cases} \frac{1}{|\omega_z|} \int_{\omega_z} \beta \, dx & \text{für alle } z \in \mathcal{N} \setminus \partial\Omega \\ 0 & \text{für alle } z \in \mathcal{N} \cap \partial\Omega. \end{cases} \quad (7.2)$$

**Lemma 7.6** (Eigenschaften des Clément Interpolators). Für  $\beta \in H^1(\Omega)$  gilt

- (a)  $\|\nabla J(\beta)\|_{L^2(\Omega)} \lesssim \|\nabla \beta\|_{L^2(\Omega)},$
- (b)  $\|h_{\mathcal{T}}^{-1}(\beta - J(\beta))\|_{L^2(\Omega)} \lesssim \|\nabla \beta\|_{L^2(\Omega)},$
- (c)  $\|h_{\mathcal{E}}^{-1/2}(\beta - J(\beta))\|_{L^2(\cup \mathcal{E})} \lesssim \|\nabla \beta\|_{L^2(\Omega)}.$

*Beweis.* Der Beweis geht auf [16] zurück.

**Definition 7.7** (Sprungdefinition). Für ein hinreichend glattes Vektorfeld  $\phi$  definieren wir den Sprung von  $\phi$  in eine Richtung  $\tau \in \mathbb{R}^2$  über eine Kante  $E \in \mathcal{E} \setminus \mathcal{E}_D$  mit  $E \in T_1 \cap T_2$ ,  $T_1, T_2 \in \mathcal{T}$  wie folgt

$$[\phi]_E \cdot \tau := (\phi|_{T_1} - \phi|_{T_2}) \cdot \tau.$$

Der Sprung über Randkanten  $\mathcal{E}_D$  wird hierbei auf Null gesetzt.

**Bemerkung 7.8.** Da die Clément Interpolation global stetig ist, gilt für  $\beta \in H^1(\Omega)$

$$[\text{Curl } J\beta]_E \cdot \nu_E = [\nabla J\beta]_E \cdot \tau_E = 0.$$

Hierbei ist  $\nu_E$  der Normalenvektor und  $\tau_E$  der Tangentialvektor der Kante  $E \in \mathcal{E}$ .

**Theorem 7.9** (Helmholtz-Zerlegung). Sei  $\phi \in L^2(\Omega)^2$  ein gegebenes Vektorfeld. Dann existieren  $\alpha \in H_0^1(\Omega)$  und  $\beta \in H^1(\Omega)/\mathbb{R}$ , so dass gilt

$$\phi = \nabla \alpha + \text{Curl } \beta. \quad (7.3)$$

*Beweis.* Der Beweis wird in [6] bzw in [9] geführt. Weiterführende Details können auch in [19] nachgelesen werden.

**Bemerkung 7.10.** Die Eigenschaft  $\beta \in H^1(\Omega)/\mathbb{R}$  in (7.3) besagt, dass  $(\beta, 1)_{L^2(\Omega)} = 0$ . Mit der Helmholtz-Zerlegung gilt immer die  $L^2$ -Orthogonalität

$$(\nabla \alpha, \text{Curl } \beta)_{L^2(\Omega)} = 0.$$

**Theorem 7.11.** Seien  $\sigma_\theta$  und  $\sigma_{\theta,h}$  Lösungen von (4.4) bzw. von (5.7). Dann gilt für stückweise konstante Lasten  $f \in L^2(\Omega)$  und homogene Randdaten  $u = 0$  auf  $\partial\Omega$

$$\begin{aligned} & \|\sigma_\theta - \sigma_{\theta,h}\|_{L^2(\Omega)}^2 \\ & \leq \min_{v \in H_0^1(\Omega)} \|\text{D}\varphi_\theta^*(\sigma_{\theta,h}) - \nabla v\|_{L^2(\Omega)}^2 \end{aligned} \quad (7.4)$$

$$\lesssim \sum_{E \in \mathcal{E}} \left\| h_E^{1/2} [\text{D}\varphi_\theta^*(\sigma_{\theta,h})] \cdot \tau_E \right\|_{L^2(E)}^2 + \sum_{T \in \mathcal{T}} \|h_T \text{curl } \text{D}\varphi_\theta^*(\sigma_{\theta,h})\|_{L^2(T)}^2. \quad (7.5)$$



*Beweis.* Mit Hilfe der strengen Monotonie von  $\varphi_\theta^*$ , einer partiellen Integration,  $\nabla u = D\varphi_\theta^*(\sigma_\theta)$ , und  $\operatorname{div}(\sigma_\theta - \sigma_{\theta,h}) = f - f_h = 0$  für stückweise konstante Lasten  $f$  erhält man für beliebige  $v \in H_0^1(\Omega)$

$$\begin{aligned} \|\sigma_\theta - \sigma_{\theta,h}\|_{L^2(\Omega)}^2 &\leq (D\varphi_\theta^*(\sigma_\theta) - D\varphi_\theta^*(\sigma_{\theta,h}), \sigma_\theta - \sigma_{\theta,h})_{L^2(\Omega)} \\ &= (\nabla u - D\varphi_\theta^*(\sigma_{\theta,h}), \sigma_\theta - \sigma_{\theta,h})_{L^2(\Omega)} \\ &= (u, \operatorname{div}(\sigma_{\theta,h} - \sigma_\theta))_{L^2(\Omega)} - (D\varphi_\theta^*(\sigma_{\theta,h}), \sigma_\theta - \sigma_{\theta,h})_{L^2(\Omega)} \\ &= (\nabla v - D\varphi_\theta^*(\sigma_{\theta,h}), \sigma_\theta - \sigma_{\theta,h})_{L^2(\Omega)} \\ &\leq \|D\varphi_\theta^*(\sigma_{\theta,h}) - \nabla v\|_{L^2(\Omega)} \|\sigma_\theta - \sigma_{\theta,h}\|_{L^2(\Omega)}. \end{aligned}$$

Folglich erhalten wir

$$\|\sigma_\theta - \sigma_{\theta,h}\|_{L^2(\Omega)} \leq \min_{v \in H_0^1(\Omega)} \|D\varphi_\theta^*(\sigma_{\theta,h}) - \nabla v\|_{L^2(\Omega)}.$$

Im Folgenden sei  $v = \operatorname{argmin}_{\tilde{v} \in H_0^1(\Omega)} \|D\varphi_\theta^*(\sigma_{\theta,h}) - \nabla \tilde{v}\|_{L^2(\Omega)}$ . D.h.,  $v$  ist die Lösung von

$$\begin{cases} (\nabla v, \nabla w)_{L^2(\Omega)} = (D\varphi_\theta^*(\sigma_{\theta,h}), \nabla w)_{L^2(\Omega)} & \text{für alle } w \in H_0^1(\Omega), \\ v = 0 & \text{auf } \partial\Omega. \end{cases} \quad (7.6)$$

Die Helmholtz-Zerlegung (7.3) von  $D\varphi_\theta^*(\sigma_{\theta,h}) = \nabla\alpha + \operatorname{Curl}\beta$  und die Euler-Lagrange Bedingung für  $\min_{\tilde{v} \in H_0^1(\Omega)} \|D\varphi_\theta^*(\sigma_{\theta,h}) - \nabla \tilde{v}\|_{L^2(\Omega)}$  liefern  $\nabla\alpha = \nabla v$ . Somit erhält man

$$\|\sigma_\theta - \sigma_{\theta,h}\|_{L^2(\Omega)}^2 \leq \|D\varphi_\theta^*(\sigma_{\theta,h}) - \nabla v\|_{L^2(\Omega)}^2 = \|\operatorname{Curl}\beta\|_{L^2(\Omega)}^2.$$

Sei  $J\beta$  die Clément Interpolation von  $\beta$ . Mit der  $L^2$ -Orthogonalität  $\operatorname{Curl} J\beta \perp_{L^2} \nabla H_0^1(\Omega)$  und dem Gleichgewicht  $(D\varphi_\theta^*(\sigma_{\theta,h}), q_h) = (-u_h, \operatorname{div} q_h)$  erhält man

$$\begin{aligned} (\operatorname{Curl}\beta, \operatorname{Curl} J\beta)_{L^2(\Omega)} &= (D\varphi_\theta^*(\sigma_{\theta,h}) - \nabla\alpha, \operatorname{Curl} J\beta)_{L^2(\Omega)} \\ &= -(u_h, \operatorname{div} \operatorname{Curl} J\beta)_{L^2(\Omega)} = 0. \end{aligned}$$

Führt man nun alle Beobachtungen zusammen, erhält man mittels partieller Integration und den Eigenschaften des Clément Interpolators

$$\begin{aligned} \|\operatorname{Curl}\beta\|_{L^2(\Omega)}^2 &= (D\varphi_\theta^*(\sigma_{\theta,h}) - \nabla\alpha, \operatorname{Curl}(\beta - J\beta))_{L^2(\Omega)} \\ &= \sum_{T \in \mathcal{T}} -(\operatorname{curl} D\varphi_\theta^*(\sigma_{\theta,h}), \beta - J\beta)_{L^2(T)} \\ &\quad + (D\varphi_\theta^*(\sigma_{\theta,h}) \cdot \tau, \beta - J\beta)_{L^2(\partial T)} \\ &\leq \sum_{T \in \mathcal{T}} \|h_T \operatorname{curl} D\varphi_\theta^*(\sigma_{\theta,h})\|_{L^2(T)} \|h_T^{-1}(\beta - J\beta)\|_{L^2(T)} \\ &\quad + \sum_{E \in \mathcal{E}} \|h_E^{1/2} [D\varphi_\theta^*(\sigma_{\theta,h})] \cdot \tau_E\|_{L^2(E)} \|h_E^{-1/2}(\beta - J\beta)\|_{L^2(E)} \\ &\lesssim \left( \sum_{T \in \mathcal{T}} \|h_T \operatorname{curl} D\varphi_\theta^*(\sigma_{\theta,h})\|_{L^2(T)}^2 \right. \\ &\quad \left. + \sum_{E \in \mathcal{E}} \|h_E^{1/2} [D\varphi_\theta^*(\sigma_{\theta,h})] \cdot \tau_E\|_{L^2(E)}^2 \right)^{1/2} \|\nabla\beta\|_{L^2(\Omega)}. \end{aligned}$$

Mit  $\|\nabla\beta\|_{L^2(\Omega)} = \|\text{Curl}\beta\|_{L^2(\Omega)}$  ist die Behauptung bewiesen.  $\square$

**Bemerkung 7.12.** *Aus der Definition von  $\varphi_\theta^*$  lässt sich erkennen, dass  $D\varphi_\theta^*(\sigma_{\theta,h})$  ein Raviart-Thomas Element innerhalb eines jeden Materials ist. Folglich gilt dort, dass  $\text{curl } D\varphi_\theta^*(\sigma_{\theta,h}) = 0$  ist. D.h., nur die Elemente  $T$  in der Kontaktzone, in der eine Vermischung beider Materialien auftritt, haben einen Beitrag in  $\sum_{T \in \mathcal{T}} \|h_T \text{curl } D\varphi_\theta^*(\sigma_{\theta,h})\|_{L^2(T)}^2$ . Im Allgemeinen dominiert der Sprungterm in (7.5) den Fehlerschätzer.*

**Bemerkung 7.13.** *Um eine Approximation von (7.6) zu berechnen, genügt eine  $\mathcal{S}_1(\mathcal{T})$ -Diskretisierung.*

# Kapitel 8

## Die FEM-Schleife

In den bisherigen Kapiteln haben wir die gemischte Formulierung des Optimal-Design Problems modelliert und einige Eigenschaften näher betrachtet. Es wurde die Existenz und Eindeutigkeit der Lösungen gezeigt und die a priori und a posteriori Fehleranalyse geführt. In den nächsten beiden Kapiteln sollen nun numerische Experimente die theoretischen Aussagen verifizieren. Zunächst erklären wir die Finite-Elemente-Schleife, womit eine Folge von Triangulierungen  $\{\mathcal{T}_k\}$  erzeugt und die jeweiligen Lösungen  $(\sigma_{\theta,h}, u_h)$  berechnet werden. Wir erhalten so eine von der Gitterweite  $h_{\mathcal{T}}$  abhängige Approximation der exakten Lösung  $(\sigma_{\theta}, u)$ . Im darauffolgenden Kapitel betrachten wir vier Modellbeispiele. Wir untersuchen hier die Konvergenzraten des geschätzten Fehlers und des Energiefehlers, welche die a priori und a posteriori Fehleranalyse bestätigen sollen.

Im Allgemeinen werden vier Schritten während eines jeden Durchlaufs der FEM-Schleife ausgeführt

$$\text{Markierung} \implies \text{Verfeinerung} \implies \text{Lösen} \implies \text{Fehlerschätzung}. \quad (8.1)$$

Die Schleife selbst wird solange ausgeführt, bis ein vorgegebenes Abbrückkriterium erfüllt wird. Im Folgenden werden wir erklären, was in den einzelnen Schritten passiert. Häufig ist es aus Implementierungssicht günstig noch Zwischenschritte einzuführen, die z.B. Informationen aus den Triangulierungen herausziehen oder die berechnete Lösung noch weiterverarbeiten, um den Implementierungsaufwand in den einzelnen Schritten zu verringern. Wir verzichten hier aber auf eine nähere Ausführung davon.

### 8.1 Markierung

Im Markierungsschritt werden Kanten  $E$  einer Triangulierung  $\mathcal{T}$  markiert, die im nächsten Schritt verfeinert werden sollen. Wir fassen solche Kanten in der Menge  $\mathcal{M} := \{E \in \mathcal{E} \mid E \text{ ist markiert}\}$  zusammen. Im Falle uniformer Verfeinerung werden alle Kanten markiert  $\mathcal{M} = \mathcal{E}$ . Im Allgemeinen führt uniforme Verfeinerung allerdings nicht zu optimalen Konvergenzverhalten. Optimales Verhalten kann man nur erwarten, wenn das Gitter die Lösung passend repräsentiert. Für unser Optimal-Design Problem wäre es z.B.

natürlich die Kontaktzone beider Materialien stärker zu verfeinern. D.h., die Menge  $\mathcal{M}$  würde hauptsächlich aus Kanten  $E$  bestehen, welche sich in dieser Zone befinden. Des Weiteren kann es passieren, dass die Lösung Singularitäten besitzt. Die Kanten in der Nähe der Singularität sollten dann ebenfalls zur Menge  $\mathcal{M}$  hinzugefügt werden.

Ausgangspunkt für die Markierung kann ein Fehlerschätzer  $\eta$  sein, der jeder Kante einen bestimmten Wert, den geschätzten Fehler, zuweist. Eine Möglichkeit die Menge  $\mathcal{M}$  zu konstruieren, wäre es zu sagen,  $\mathcal{M}$  bestehe aus allen Kanten, die einen Wert größer als die Hälfte des maximalsten Wertes aller Kanten besitzen. Dieses Kriterium ist als das Maximumkriterium bekannt und führt zu optimalem Konvergenzverhalten. Allerdings erweist es sich in der Praxis häufig als ungeeignet. Die Fehler in der Kontaktzone und an den möglichen Singularitäten dominieren den Fehler im Inneren der Materialien. Dies führt dazu, dass nur wenige Kanten der Menge  $\mathcal{M}$  hinzugefügt werden. D.h., man benötigt viele Durchläufe der FEM-Schleife, um ein bestimmtes Abbruchkriterium zu erreichen.

Ein Kriterium, welches sich sowohl in der Theorie und in der Praxis bewährt hat, ist das sogenannte *bulk*-Kriterium. Die geschätzten Fehler der einzelnen Kanten  $\eta_E$  werden der Größe nach sortiert. Ausgehend vom maximalsten Wert werden dann Kanten der Menge  $\mathcal{M}$  hinzugefügt, bis das Kriterium

$$\frac{1}{2} \sum_{E \in \mathcal{E}} \eta_E^2 \leq \sum_{E \in \mathcal{M}} \eta_E^2 \quad (8.2)$$

erfüllt ist. Das Kriterium führt zu optimalem Konvergenzverhalten und die Anzahl der Schleifendurchläufe bis zur Erfüllung des Abbruchkriterium sind im Vergleich zum Maximumkriterium wesentlich geringer. Man kann sich das Maximumkriterium als Maximumnorm und das *bulk*-Kriterium als  $L^2$ -Norm vorstellen. Abbildung 8.1 illustriert diesen Vergleich.

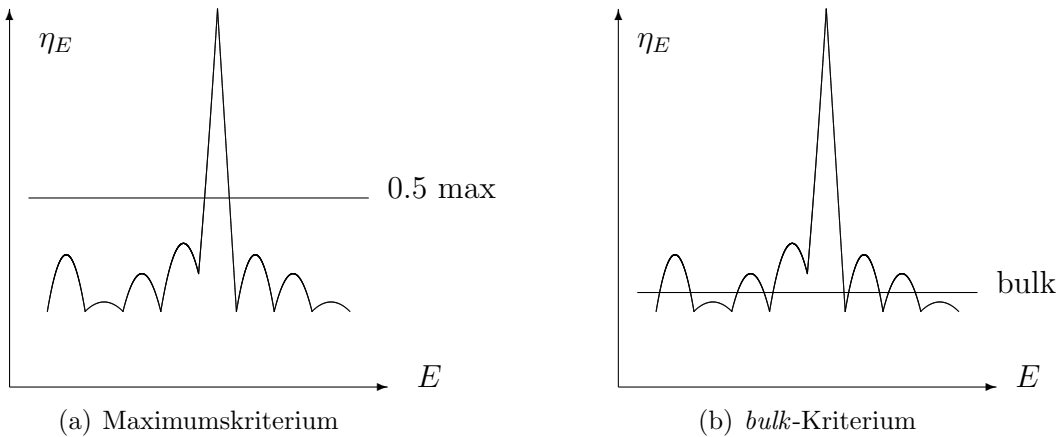


Abbildung 8.1: Vergleich der Markierungsstrategien. Entnommen aus [8].

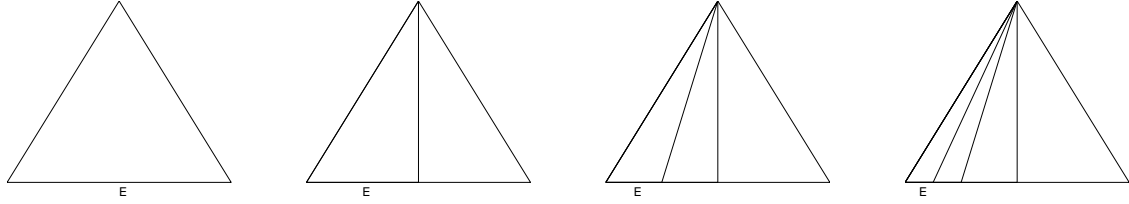


Abbildung 8.2: Verfeinerung eines Dreiecks  $T$ , wobei immer nur die Kante  $E$  markiert und halbiert wird.

Bei der späteren Verfeinerung wollen wir verhindern, dass unsere Triangulierungen  $\mathcal{T}$  degenerieren. Eine Degeneration hätte zur Folge, dass sich die Güte unseres Problems erheblich verschlechtert. Ein hohe Konditionszahl der auftretenden Matrizen wäre die Folge. Ein Kriterium, welches dieses Verhalten beschreibt, ist die Winkelbedingung.

**Definition 8.1** (Winkelbedingung). Sei  $\alpha_T := \{\alpha_1, \alpha_2, \alpha_3\}$  die Menge der Innenwinkel eines Dreiecks  $T$ . Eine Familie von Triangulierungen  $\{\mathcal{T}_k\}$  erfüllt die maximale respektive minimale Winkelbedingung, wenn gilt

$$\max \alpha_T < c_1 < \pi \quad \text{respektive} \quad 0 < c_2 < \min \alpha_T. \quad (8.3)$$

Hier sind  $c_1$  und  $c_2$  Konstanten, die von 0 bzw von  $\pi$  wegbeschränkt sind.

Durch die Markierungsstrategien, die wir oben beschrieben haben, kann es passieren, dass die Winkelbedingung verletzt wird. Abbildung 8.2 zeigt eine Verletzung der minimalen Winkelbedingung bei fortlaufender Verfeinerung.

Um eine Degenerierung zu verhindern, müssen zusätzliche Kanten markiert werden. Hierfür wird jedem Dreieck  $T$  eine Referenzkante  $E_{ref} \in \partial T$ , z.B. die längste Kante, zugeordnet. Der *closure*-Algorithmus 8.2 sorgt dann dafür, dass nach der *bulk*-Markierung noch zusätzliche Kanten markiert werden, die eine Degenerierung der Triangulierungen verhindern.

**Algorithmus 8.2** (Closure). Für alle Dreiecke  $T \in \mathcal{T}$ , bei denen eine Kante  $E \in \partial T$  mit  $E \in \mathcal{M}$  und  $E \neq E_{ref}$  existiert, wird die Referenzkante  $E_{ref} \in \partial T$  ebenfalls der Menge  $\mathcal{M}$  hinzugefügt.

Die Ausstrahlung des *closure*-Algorithmus ist amortisiert konstant. D.h., es kann passieren, dass aufgrund des Algorithmus in einer Triangulierung alle Kanten markiert werden und damit uniform verfeinert wird. Betrachtet man aber die Folge von Triangulierungen die aufgrund der Verfeinerung entstehen, kommen im Mittel nur konstant viele Kanten hinzu.

**Bemerkung 8.3.** Die Zuordnung der Referenzkanten wird nur in der Ausgangstriangulierung vorgenommen. Danach vererben sich die Referenzkanten durch die Verfeinerung. Dies wird in Abschnitt 8.2 erklärt.

## 8.2 Verfeinerung

Im Verfeinerungsschritt wird anhand einer Liste von markierten Kanten eine neue Triangulierung erstellt. Je nachdem wie viele Kanten eines Dreiecks markiert wurden, wird das Dreieck rot, grün oder blau verfeinert. Im Folgenden werden wir die drei Arten der Verfeinerung erklären und zeigen wie die Referenzkanten vererbt werden. In den Abbildungen 8.4 und 8.5 sind die Referenzkanten durch eine dicke Linie dargestellt. Die markierten Kanten werden durch Marker auf den Kanten dargestellt

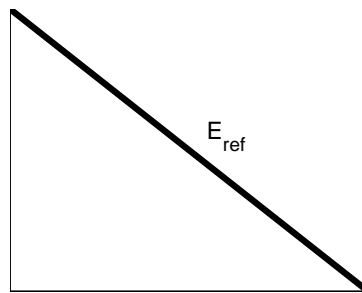
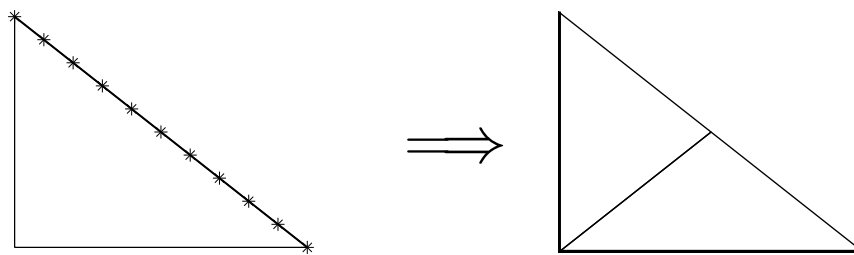
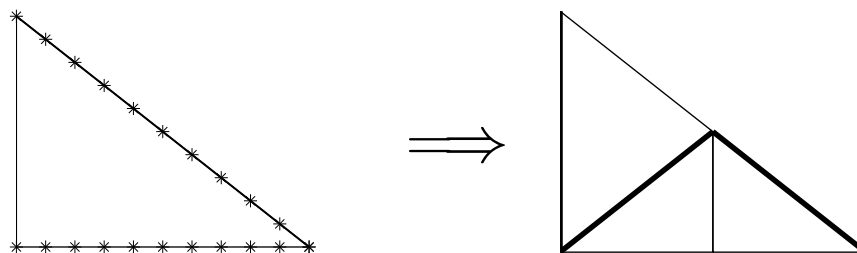


Abbildung 8.3: Ausgangsdreieck mit Referenzkante  $E_{ref}$



(a) Grünverfeinerung



(b) Blau-Rechts-Verfeinerung

Abbildung 8.4: Das Ausgangsdreieck wird zum einen Grün verfeinert und einmal Blau-Rechts verfeinert.

**Bemerkung 8.4.** Wenn eine Kante eines Dreiecks markiert wurde, dann ist auch die Referenzkante markiert. Und beide Kanten werden verfeinert. Bei der RGB-Verfeinerung treten keine hängende Knoten auf.

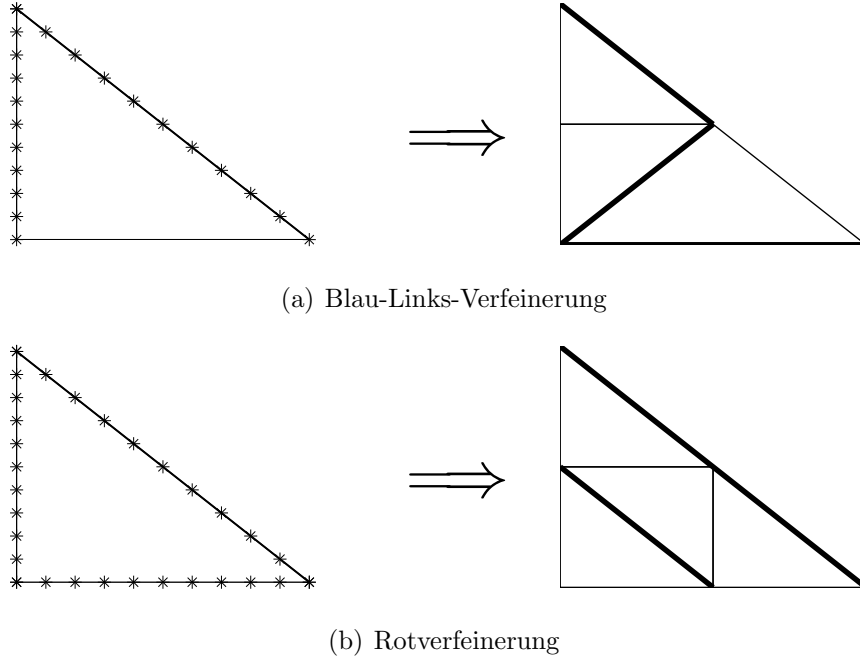


Abbildung 8.5: Das Ausgangsdreieck wird zum einen Blau-Links verfeinert und einmal Rot verfeinert.

### 8.3 Lösen

Im Lösungsschritt wird das nichtlineare System aufgebaut und mittels eines iterativen Verfahrens gelöst. Ausgehend von der diskreten schwachen Formulierung (5.7) wird mittels eines Gauss-Newton-Verfahrens für den Regularisierungsparameter  $\theta > 0$  die Nullstelle des Systems gesucht. Im Folgenden wird kurz erklärt wie das iterative Verfahren funktioniert und wie die benötigte Funktion und die Jacobiatrix definiert sind. Eine ausführliche Beschreibung des Verfahrens findet man in [5] bzw. [17].

Die numerische Realisierung von (5.7) findet in MATLAB statt. Wir verwenden die MATLAB-Funktion `fsolve`, welche für eine gegebene Funktion  $R$  mittels Gauss-Newton die Nullstelle  $x$  mit  $R(x) = 0$  findet. Sei  $R$  eine stetig differenzierbare Funktion. Um die Nullstelle zu finden, muss nun ein kleinstes Quadrate Problem

$$\min_x \frac{1}{2} \|R(x)\|^2 \quad (8.4)$$

gelöst werden. Das Gauss-Newton Verfahren linearisiert zunächst zu einem gegebenen Punkt  $x_k$  die Funktion  $R(x)$

$$r(x, x_k) = R(x_k) + DR(x_k)(x - x_k) \quad (8.5)$$

und minimiert dann die Norm der linearisierten Funktion

$$x_{k+1} = \operatorname{argmin}_x \frac{1}{2} \|r(x, x_k)\|^2. \quad (8.6)$$

Die Minimierung führt zu der Vorschrift

$$x_{k+1} = x_k - \left( DR(x_k) DR(x_k)^T \right)^{-1} DR(x_k) R(x_k). \quad (8.7)$$

**Bemerkung 8.5.** Die Richtung gegeben durch  $-(DR(x_k) DR(x_k)^T)^{-1} DR(x_k) R(x_k)$  ist eine Abstiegsrichtung, da  $DR(x_k) R(x_k)$  der Gradient der Funktion  $0.5\|R(x)\|^2$  an der Stelle  $x_k$  ist und die Matrix  $(DR(x_k) DR(x_k)^T)^{-1}$  positiv definit ist.

**Bemerkung 8.6.** Für den Fall, dass die Matrix  $(DR(x_k) DR(x_k)^T)$  singular ist, wird eine Modifizierung des Verfahrens vorgenommen. Es wird ein Stabilisierungsterm  $S_k$ , so dass  $(DR(x_k) DR(x_k)^T + S_k)$  positiv definit ist, eingeführt. Zum Beispiel kann  $S_k$  ein Vielfaches der Einheitsmatrix sein. Dieses modifizierte Verfahren ist als Levenberg-Marquardt Methode bekannt.

Für das Optimal-Design Problem ist die Funktion  $R(x)$ , das Residuum, durch die schwache diskrete Formulierung (5.7) gegeben. Sei  $x$  der Koeffizientenvektor, der die diskreten Funktionen  $\sigma_{\theta,h} \in RT_0(\mathcal{T})$  und  $u_h \in \mathcal{P}_0(\mathcal{T})$  beschreibt. Die Basisfunktionen des Raviart-Thomas Raums auf  $T$  seien  $\Psi_j$   $j \in \{1, 2, 3\}$ , die Basisfunktion der Verschiebungen auf  $T$  sei  $v \equiv 1$ . Dann ist das lokale Residuum auf jedem Dreieck  $T$  gegeben durch

$$R(x)|_T := \begin{pmatrix} (u_h(x), \operatorname{div} \Psi_j)_{L^2(T)} + D\varphi_{\theta}^*(\sigma_{\theta,h}(x))(\Psi_j) \\ (v, \operatorname{div} \sigma_{\theta,h}(x))_{L^2(T)} + (f, v)_{L^2(T)} \end{pmatrix} \in \mathbb{R}^4. \quad (8.8)$$

Die Freiheitsgrade des Raviart-Thomas Raums liegen auf den Kanten  $E$  der Triangulierung  $\mathcal{T}$ . Die Verschiebungen werden mit stückweisen konstanten Funktionen approximiert. D.h., die Freiheitsgrade der Verschiebungen sind die Dreiecke  $T$ . Damit ergibt sich für das globale Residuum die Dimension

$$R(x) \in \mathbb{R}^{|\mathcal{E}|+|\mathcal{T}|}. \quad (8.9)$$

Die Jacobimatrix  $DR(x)$  ist definiert durch

$$DR(x)|_T = \begin{pmatrix} D^2 \varphi_{\theta}^*(\sigma_{\theta,h}(x))(\Psi_j)(\Psi_k) & (v, \operatorname{div} \Psi_j)_{L^2(T)} \\ (v, \operatorname{div} \Psi_j)_{L^2(T)} & 0 \end{pmatrix} \in \mathbb{R}^{4 \times 4}. \quad (8.10)$$

Die globale Jacobimatrix ist dann von Dimension

$$DR(x) \in \mathbb{R}^{(|\mathcal{E}|+|\mathcal{T}|) \times (|\mathcal{E}|+|\mathcal{T}|)}. \quad (8.11)$$



**Definition 8.7.** Die für die Implementierung relevanten Formulierungen der Richtungsableitungen  $D\varphi_\theta^*(\sigma_{\theta,h})(\Psi_j)$  und  $D^2\varphi_\theta^*(\sigma_{\theta,h})(\Psi_j)(\Psi_k)$  sind gegeben durch

$$D\varphi_\theta^*(\sigma_{\theta,h})(\Psi_j) := \int_T \psi_\theta^{*'}(|\sigma_{\theta,h}|)/|\sigma_{\theta,h}| \sigma_{\theta,h} \cdot \Psi_j \, dx$$

und

$$\begin{aligned} D^2\varphi_\theta^*(\sigma_{\theta,h})(\Psi_j)(\Psi_k) &:= \int_T \psi_\theta^{*''}(|\sigma_{\theta,h}|)/|\sigma_{\theta,h}|^2 (\sigma_{\theta,h} \cdot \Psi_j)(\sigma_{\theta,h} \cdot \Psi_k) \\ &\quad + \psi_\theta^{*'}(|\sigma_{\theta,h}|)/|\sigma_{\theta,h}|^3 (\Psi_j \cdot \Psi_k |\sigma_{\theta,h}|^2 - (\sigma_{\theta,h} \cdot \Psi_j)(\sigma_{\theta,h} \cdot \Psi_k)) \, dx. \end{aligned}$$

## 8.4 Fehlerschätzung

Nachdem die Lösung für eine Triangulierung  $\mathcal{T}$  berechnet wurde, wird im letzten Schritt der Fehler der diskrete Lösung zur exakten Lösung geschätzt. Im Kapitel 7 wurden in der a posteriori Analysis zwei Fehlerschätzer motiviert

$$\eta_{(1)}^2 := \sum_{T \in \mathcal{T}} \eta_T^{(1)^2} \quad (8.12)$$

mit  $\eta_T^{(1)^2} := \|D\varphi_\theta^*(\sigma_{\theta,h}) - \nabla v_h\|_{L^2(T)}^2$  und

$$\eta_{(2)}^2 := \sum_{T \in \mathcal{T}} \eta_T^{(2)^2} + \sum_{E \in \mathcal{E}} \eta_E^{(2)^2} \quad (8.13)$$

mit  $\eta_T^{(2)^2} := \|h_T \operatorname{curl} D\varphi_\theta^*(\sigma_{\theta,h})\|_{L^2(T)}^2$  und  $\eta_E^{(2)^2} := \|h_E^{1/2} [D\varphi_\theta^*(\sigma_{\theta,h})] \cdot \tau_E\|_{L^2(E)}^2$ .

Unsere Verfeinerung ist kantenbasiert. D.h., die Terme  $\eta_T^{(1)}$  und  $\eta_T^{(2)}$ , welche jedem Dreieck einen Fehler zuordnen, müssen abschließend noch angepasst werden. Bei  $\eta_T^{(1)}$  wird das *bulk*-Kriterium auf die Menge der Dreiecke ausgeführt. Für jedes markierte Dreieck werden anschließend alle drei zugehörigen Kanten markiert. Bei  $\eta_T^{(2)}$  wird der Volumenfehler, welcher auf den Dreiecken lebt, auf die zugehörigen Kanten verteilt.

# Kapitel 9

## Numerische Beispiele

In diesem Kapitel betrachten wir numerische Beispiele, welche die Theorie verifizieren sollen. Zunächst werden vier Modellbeispiele und ein Algorithmus zur Berechnung des Materialparameters  $\lambda$  vorgestellt. Anschließend stellen wir ein Algorithmus vor, der die Regularisierung kontrolliert, und somit eine Approximation des Optimal-Design Problems für  $\theta \rightarrow 0$  realisiert. Der Algorithmus ist eine Adaption des in [14] präsentierten Algorithmus.

### 9.1 Definintion der Modellbeispiele

In den folgenden Beispielen benutzen wir folgende Parameterkonfiguration

- Die Schubmodule:  $\mu_1 = 1$  und  $\mu_2 = 2$ ,
- Der Massenanteil:  $\xi = 0.5$ ,
- Die Randdaten:  $u = 0$  auf  $\partial\Omega$ ,
- Die Last:  $f = 1$  in  $\Omega$ .

Die Gebiete sind gegeben durch das Quadrat  $\Omega_1 = [0, 1]^2$ , das L-shape Gebiet  $\Omega_2 = [-1, 1]^2 \setminus ((0, 1] \times (0, -1])$ , das Oktagon  $\Omega_3 \subset [-1, 1]^2$  und das Schlitzgebiet  $\Omega_4 = [-1, 1]^2 \setminus (0, 1]$ .

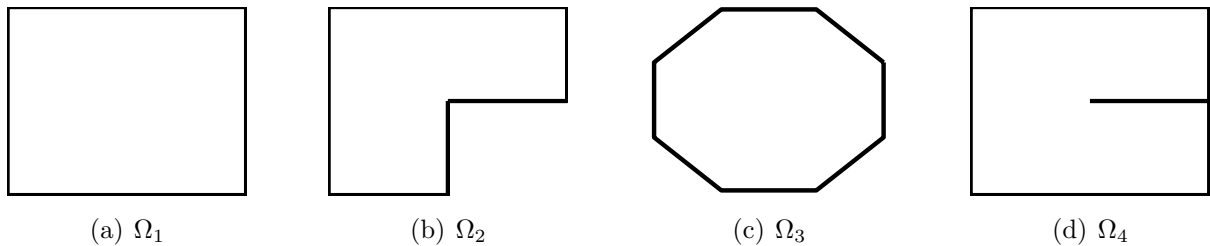


Abbildung 9.1: Geometrien der vier Modellbeispiele.

Die Energien der vier Beispiele ist durch die Formulierung (3.22) gegeben. Da die exakten Lösungen jeweils unbekannt sind, wird die  $RT_0(\mathcal{T}) - \mathcal{P}_0(\mathcal{T})$ -Approximation des Optimal-Design Problems mit Hilfe des Aitken'schen Delta Prozesses extrapoliert.

**Definition 9.1** (Aitken'scher Delta Prozess). *Sei  $(x_n)$  eine gegebene Folge. Dann ist die Aitken'sche Extrapolation  $(y_n)$  gegeben durch*

$$y_k = x_k - \frac{(x_{k+1} - x_k)^2}{x_{k+2} - 2x_{k+1} + x_k}. \quad (9.1)$$

*Die Konvergenz der Folge  $(y_n)$  gegen den Limes  $x$  erfolgt schneller als die der Folge  $(x_n)$ . Somit lässt sich der Limes  $x$  durch die Folge  $(y_n)$  approximieren.*

Die extrapolierten Energien  $E_A$  (9.4) der vier Modellbeispiele sind

Beispiel	$E_A$
$\Omega_1$	0.027983242756785
$\Omega_2$	0.160674590782529
$\Omega_3$	0.257317403528310
$\Omega_4$	0.247295558950406

Tabelle 9.1: Extrapolierte Energien der Modellbeispiele.

Um eine Lösung von (3.14) zu bestimmen, ist es notwendig den Lagrangemultiplikator  $\lambda$ , welcher die Nebenbedingung der Massenverteilung integriert, zu berechnen. Eine Approximation von  $\lambda$  lässt sich mit Algorithmus 9.2 bestimmen.

**Algorithmus 9.2** (Goldener Schnitt). *Sei eine Triangulierung  $\mathcal{T}$  fixiert und ein Startintervall  $[a, b]$  gegeben. Setze  $error = 1$  und  $P = (-1 + \sqrt{5})/2$ . Für eine vorgebene Toleranzschwelle  $Tol$  berechne: Solange  $error \geq Tol$*

- (a) *Bestimme die Energie  $I_1$  von (3.14) für  $\lambda = a$ .*
- (b) *Bestimme die Energie  $I_2$  von (3.14) für  $\lambda = b$ .*
- (c) *Für  $I_1 < I_2$  setze  $b = b - P(b - a)$ . Für  $I_2 < I_1$  setze  $a = a + P(b - a)$ .*
- (d) *Setze  $error = |I_2 - I_1|$ .*

**Bemerkung 9.3.** *Für das im Algorithmus 9.2 vorgegebene Startintervall  $[a, b]$  muss gelten, dass  $\lambda \in [a, b]$  liegt. Für die vier hier betrachteten Beispiele genügt es das Intervall  $[0, 1]$  zu nehmen.*

In der Arbeit von [4] wurden mit Hilfe einer Modifikation von Algorithmus 9.2 die Lagrangemultiplikatoren  $\lambda$  berechnet.

Beispiel	$\lambda$
$\Omega_1$	0.0084
$\Omega_2$	0.0143
$\Omega_3$	0.0284
$\Omega_4$	0.0168

Tabelle 9.2: Die Lagrangemultiplikatoren der vier Modellbeispiele.

## 9.2 Der Regularisierungsalgorithmus

In diesem Abschnitt präsentieren wir einen Algorithmus, der in Abhängigkeit der Gitterweite  $h_{\mathcal{T}}$  die Stärke der Regularisierung kontrolliert. Ziel des Algorithmus ist es, den Regularisierungsfehler sowie den Approximationsfehler in gleichem Maße für  $h \rightarrow 0$  zu reduzieren. Der Output ist eine Approximation des gemischten Optimal-Design Problems (3.21) und dessen Lösung  $(\sigma_{\theta,h}, u_h)$ .

**Algorithmus 9.4.** *Input: Es seien eine reguläre Triangulierung  $\mathcal{T}_0$ , ein Startwert  $x_0$ , Regularisierungsparameter  $\alpha > 0$  und  $\beta > 0$  und ein Abbruchkriterium  $TOL$  gegeben, und setze  $\eta = \infty$  und  $\ell = 1$ . Solange  $\eta \geq Tol$  berechne:*

- (a) *Erzeuge eine neue Triangulierung  $\mathcal{T}_\ell$  entsprechend dem Fehlerschätzer  $\eta$ .*
- (b) *Prolongiere die alte Lösung  $x_{\ell-1}$  auf das neue Gitter  $\mathcal{T}_\ell$ , um einen Startwert  $x_\ell^0$  zu erhalten.*
- (c) *Setze  $\theta|_T = \alpha h_T^\beta$ .*
- (d) *Berechne die Lösung  $x_\ell$  von (5.7) mittels einem Gauss-Newton-Verfahrens mit dem Startwert  $x_\ell^0$ .*
- (e) *Berechne den geschätzten Fehler  $\eta$  der Lösung  $x_\ell$  und setze  $\ell = \ell + 1$ .*

*Output: Wir bekommen eine Approximation der Nullstelle von (5.7) mit einem gitter-abhängigen Parameter  $\theta$ .*

**Bemerkung 9.5.** *Im Algorithmus 9.4 projizieren wir die Lösung  $x_{\ell-1}$  von  $\mathcal{T}_{\ell-1}$  auf  $\mathcal{T}_\ell$ . Wir erhalten somit einen Startwert in der Nähe von  $x_\ell$ . Für das Gauss-Newton Verfahren genügen dann in der Regel 3-5 Iterationsschritte, um die Lösung  $x_\ell$  zu finden. Ohne diese Prolongation, z.B. mit einem zufälligen Startwert, sind im Allgemeinen 10 und mehr Iterationsschritte, abhängig von der Gitterweite  $h_{\mathcal{T}}$ , nötig, um eine zulässige Lösung zu finden.*

**Bemerkung 9.6.** *Zur Darstellung der optimalen Verteilung der beiden Materialien folgen wir den Ausführungen von [4]. Die Massenverteilung  $\Lambda(|D\varphi_\theta^*(\sigma_{\theta,h})|)$  der beiden Materia-*

lien ist gegeben durch

$$\Lambda(t) := \begin{cases} 0 & \text{für } 0 \leq t \leq t_1, \\ \frac{t-t_1}{t_2-t_1} & \text{für } t_1 < t < t_2, \\ 1 & \text{für } t_2 \leq t. \end{cases} \quad (9.2)$$

In den numerischen Beispielen wird das „harte“ Material mit  $\Lambda(t) = 0$  blau und das „weiche“ Material mit  $\Lambda(t) = 1$  rot dargestellt.

Um den Algorithmus 9.4 nutzen zu können, müssen die Regularisierungsparameter  $\alpha$  und  $\beta$  bekannt sein. Im Folgenden betrachten wir ein numerisches Testbeispiel, um die optimalen Werte von  $\alpha$  und  $\beta$  in einer Fallstudie experimentell zu bestimmen. Hierfür geben wir eine analytische Funktion  $\tilde{u}$  mit homogenen Randdaten auf dem Einheitsquadrat  $\Omega = [0, 1]^2$  vor

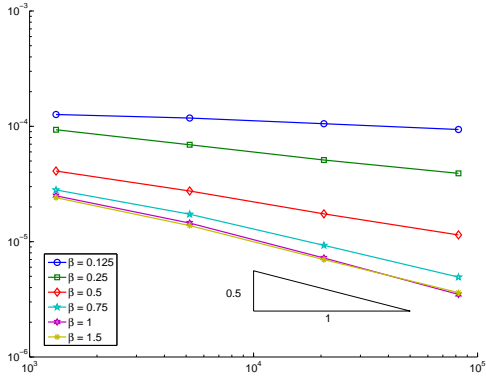
$$\tilde{u}(x, y) := x y (1 - x) (1 - y). \quad (9.3)$$

Die Materialparameter sind wie bei den Modellbeispielen vorgegeben. Die Last  $f$  ist dann analytisch durch  $f := -\operatorname{div} D\varphi(\nabla \tilde{u})$  definiert. Der Parameter  $\lambda$  ist für dieses Beispiel durch  $\lambda = 0.003101917488281$  gegeben.

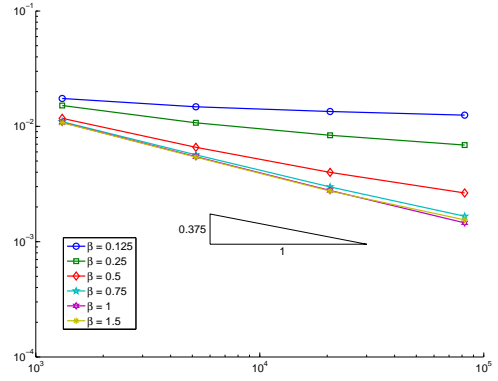
Zur experimentellen Bestimmung der Parameter  $\alpha$  und  $\beta$  betrachten wir für uniforme Verfeinerung die Fehlerschätzer  $\eta_{(1)}$  und  $\eta_{(2)}$ . Da dies nur geschätzte Fehler sind, untersuchen wir noch zusätzlich den exakten Spannungsfehler  $\|\sigma - \sigma_{\theta,h}\|_{L^2(\Omega)}$  und den exakten Fehler in der gemischten Energieformulierung. Nach (3.22) ist das gemischte Energiefunktional gegeben durch

$$E(\sigma) = \int_{\Omega} \varphi^*(\sigma) dx. \quad (9.4)$$

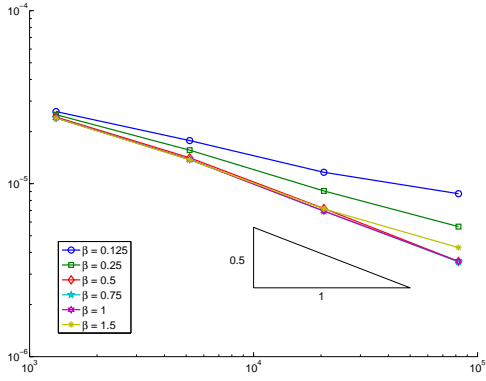
Gemäß der a priori Analysis (7.1) erwarten wir für die optimalen Parameter eine Konvergenzrate um 0.375 bzgl  $h$  im Spannungsfehler  $\|\sigma - \sigma_{\theta,h}\|_{L^2(\Omega)}$ . Steigungsdreiecke in den Abbildungen der Fehlergraphen zeigen einen Referenzanstieg an. Wir stellen den Fehler in Abhängigkeit der Freiheitsgrade an. Ein Abfall der Ordnung 0.5 entspricht hierbei linearer Konvergenz in  $h$ .



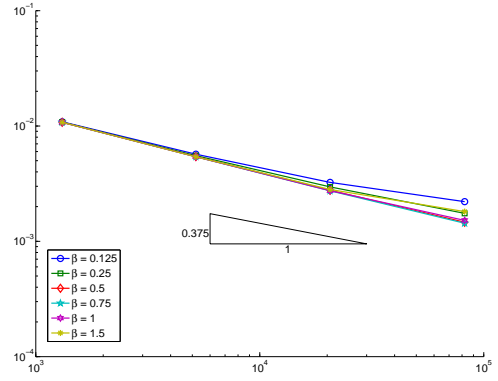
(a)  $E(\sigma_{\theta,h}) - E(\sigma)$  mit  $\alpha = 1$ .



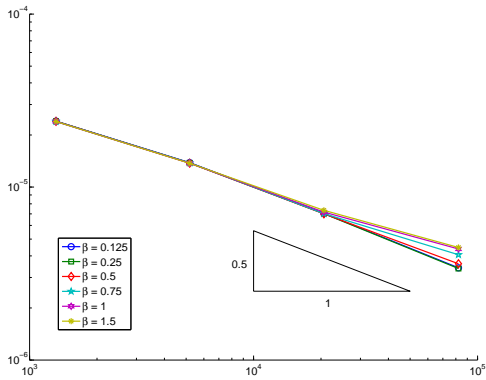
(b)  $\|\sigma - \sigma_{\theta,h}\|_{L^2(\Omega)}$  mit  $\alpha = 1$ .



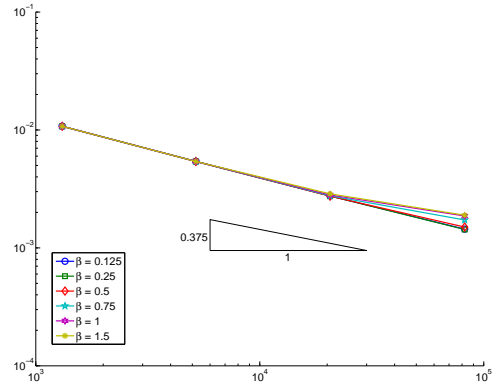
(c)  $E(\sigma_{\theta,h}) - E(\sigma)$  mit  $\alpha = 0.1$ .



(d)  $\|\sigma - \sigma_{\theta,h}\|_{L^2(\Omega)}$  mit  $\alpha = 0.1$ .

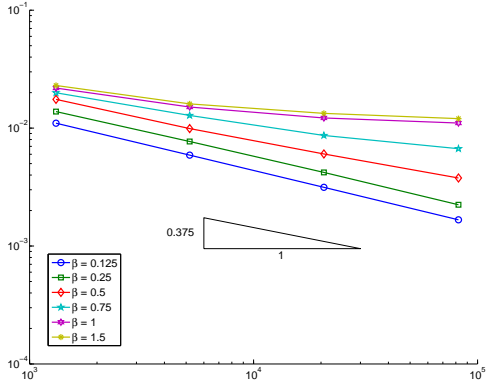


(e)  $E(\sigma_{\theta,h}) - E(\sigma)$  mit  $\alpha = 0.01$ .

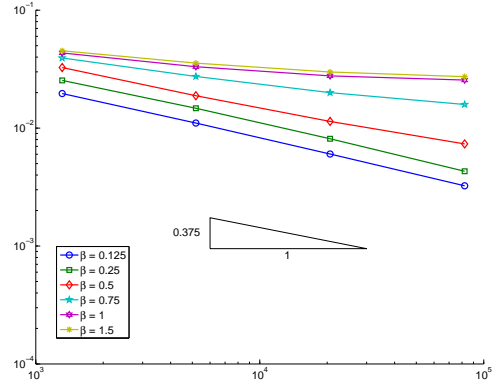


(f)  $\|\sigma - \sigma_{\theta,h}\|_{L^2(\Omega)}$  mit  $\alpha = 0.01$ .

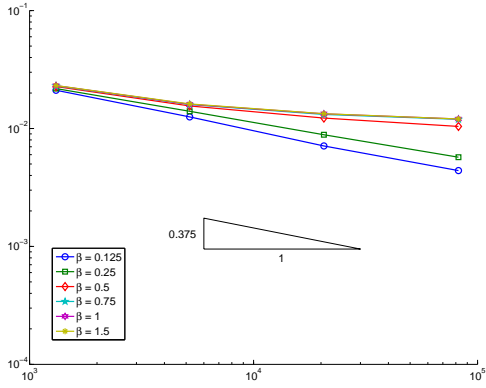
Abbildung 9.2: Vergleich des Konvergenzverhaltens des Fehlers in der konjugierten Energiedichte  $E(\sigma_{\theta,h}) - E(\sigma)$  (links) und des Spannungsfehlers  $\|\sigma - \sigma_{\theta,h}\|_{L^2(\Omega)}$  (rechts) mit verschiedenen Werten für  $\alpha$  und  $\beta$ .



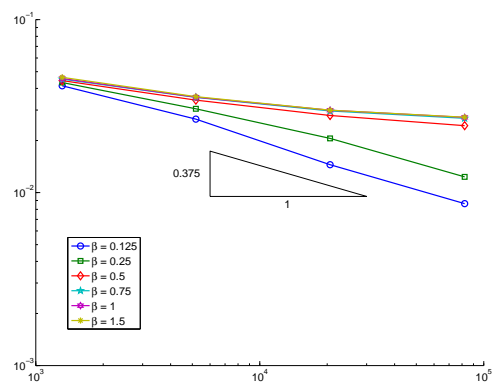
(a) Geschätzter Fehler  $\eta_{(1)}$  mit  $\alpha = 1$ .



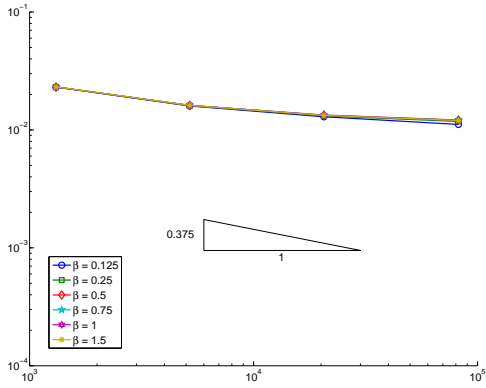
(b) Geschätzter Fehler  $\eta_{(2)}$  mit  $\alpha = 1$ .



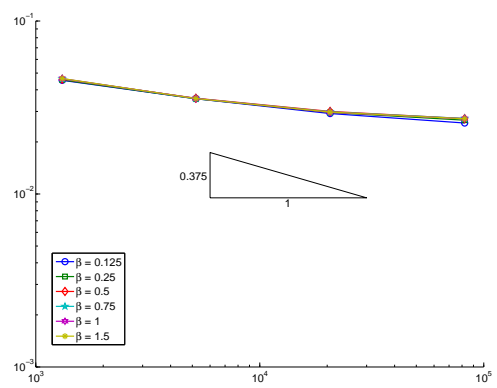
(c) Geschätzter Fehler  $\eta_{(1)}$  mit  $\alpha = 0.1$ .



(d) Geschätzter Fehler  $\eta_{(2)}$  mit  $\alpha = 0.1$ .



(e) Geschätzter Fehler  $\eta_{(1)}$  mit  $\alpha = 0.01$ .



(f) Geschätzter Fehler  $\eta_{(2)}$  mit  $\alpha = 0.01$ .

Abbildung 9.3: Vergleich des Konvergenzverhaltens der geschätzten Fehler  $\eta_{(1)}$  und  $\eta_{(2)}$  mit verschiedenen Werten für  $\alpha$  und  $\beta$ .

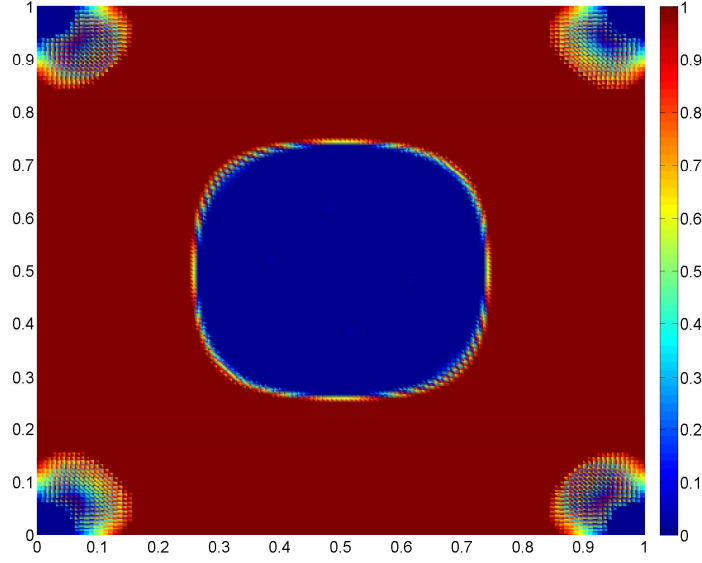
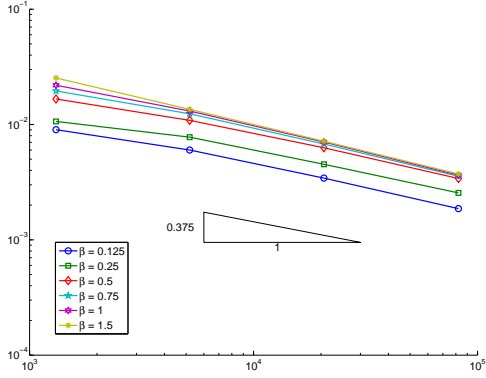


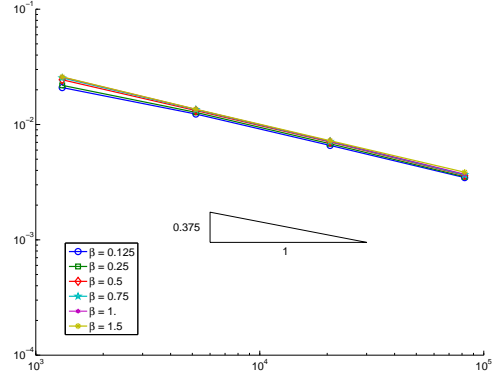
Abbildung 9.4: Massenverteilung  $\Lambda(t)$  für die Funktion  $\tilde{u}$ , gegeben in (9.3), für  $\alpha = 1$  und  $\beta = 0.5$ .

Das Konvergenzverhalten in Abbildung 9.2 vom exakten Energiefehler und dem exakten Spannungsfehler zeigen, dass sich bei zu starker Regularisierung (großes  $\alpha$  und kleines  $\beta$ ) die Konvergenzordnung erheblich verschlechtert, während wir bei geringer Regularisierung optimales Verhalten beobachten. Der Diskretisierungsfehler konvergiert zwar optimal gegen Null, was beide Fehlerschätzer in Abbildung 9.3 zeigen, der Regularisierungsfehler hingegen wird kaum reduziert, je stärker die Regularisierung ist. Ursache hierfür ist der Sprung in  $\partial\varphi^*$ . Treffen beide Materialien aufeinander, gibt es entlang der Kontaktlinie kein eindeutiges Materialverhalten. Betrachtet man den Sprungterm im Fehlerschätzer  $\eta_{(2)}$ , so kann der Term  $[D\varphi_\theta^*] \cdot \tau_E$  für  $h, \theta \rightarrow 0$  entlang dieser Kontaktlinie nicht verschwinden. Er nähert sich hier der Konstante  $C(\sigma, \sigma_{\theta,h})$  aus Bemerkung 3.7 an. Ein optimales Konvergenzverhalten des Fehlerschätzers ist somit nicht möglich. Betrachtet man nun den Fehlerschätzer  $\eta_{(2)}$  abzüglich des Sprungs, erhält man das Konvergenzverhalten von  $\eta_{(2)}$  im Inneren der beiden Materialien. Aus Abbildung 9.5 erkennen wir, dass der Fehler im Inneren optimal reduziert wird. Eine ähnliche Überlegung trifft auch auf den Fehlerschätzer  $\eta_{(1)}$  zu. Wir wissen, dass der Term  $D\varphi_\theta^*(\sigma_{\theta,h})$  im Inneren ein Raviart-Thomas Element ist und demzufolge verschwindet der  $\text{curl } D\varphi_\theta^*(\sigma_{\theta,h})$  dort. Aus der a posteriori Abschätzung (7.5) wissen wir, dass mit der Helmholtzzerlegung für  $D\varphi_\theta^*(\sigma_{\theta,h})$  der Schätzer  $\eta_{(1)}$  die  $L^2$ -Norm eines Rotationsfeldes berechnet. Demzufolge haben nur die Elemente entlang der Kontaktlinie einen Beitrag in  $\eta_{(1)}$ . Durch eine lokale  $H^1$ -semi Projektion, bezogen auf das Innere der beiden Materialien, anstelle einer globalen Projektion in (7.6) erhält man die Einschränkung des Fehlerschätzers auf die Materialien. Des Weiteren wird der Fehlerschätzer  $\eta_{(1)}$  durch  $\eta_{(2)}$  dominiert. Die lokale Projektion bzw. die Dominanz durch  $\eta_{(2)}$  führen zu optimalen Konvergenzverhalten von  $\eta_{(1)}$  im Inneren.

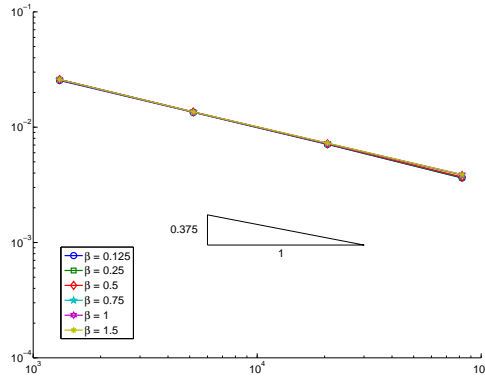




(a) Geschätzter Fehler  $\eta_{(2)}$  mit  $\alpha = 1$ .



(b) Geschätzter Fehler  $\eta_{(2)}$  mit  $\alpha = 0.1$ .



(c) Geschätzter Fehler  $\eta_{(2)}$  mit  $\alpha = 0.01$ .

Abbildung 9.5: Vergleich des Konvergenzverhaltens des geschätzten Fehlers  $\eta_{(2)}$  eingeschränkt auf das Innere beider Materialien mit verschiedenen Werten für  $\alpha$  und  $\beta$ .

Im nächsten Abschnitt bestimmen wir das optimale Design der vier Modellbeispiele. Hierfür setzen wir die Regularisierungsparameter auf  $\alpha = 1$  und  $\beta = 1$ . Mit dieser Konfiguration haben wir bei der vorangegangenen Fallstudie optimales Konvergenzverhalten sowohl im Fehlerschätzer als auch im Spannungsfehler beobachtet.

### 9.3 Das optimale Design der Modellbeispiele

In diesem Abschnitt berechnen wir für die vier Modellbeispiele das Optimal-Design Problem mit den in Abschnitt 9.1 und 9.2 gesetzten Parametern. Wir betrachten für die Fehlerschätzer (8.12) und (8.13) die generierten Gitter und das Konvergenzverhalten. Des Weiteren vergleichen wir den Energiefehler gegeben durch (9.4).

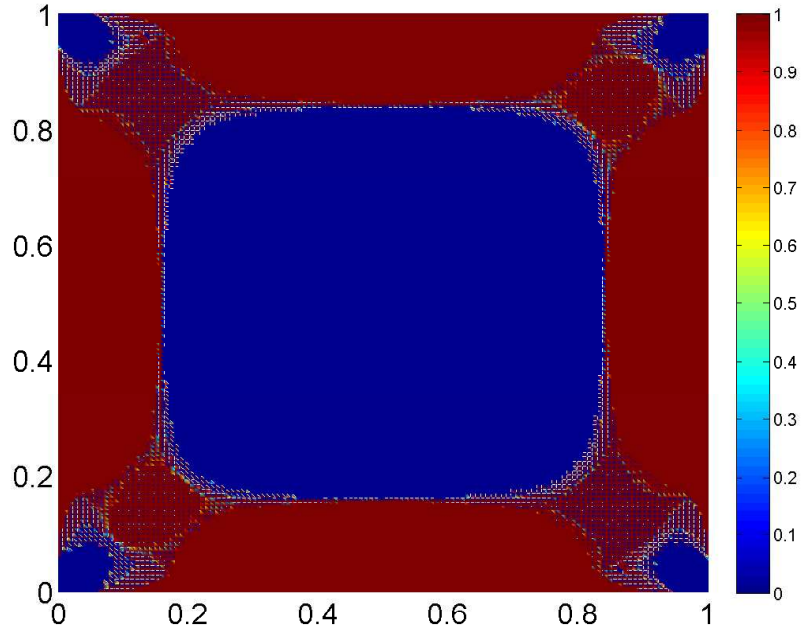


Abbildung 9.6: Massenverteilung bei uniformer Verfeinerung für das Quadrat.

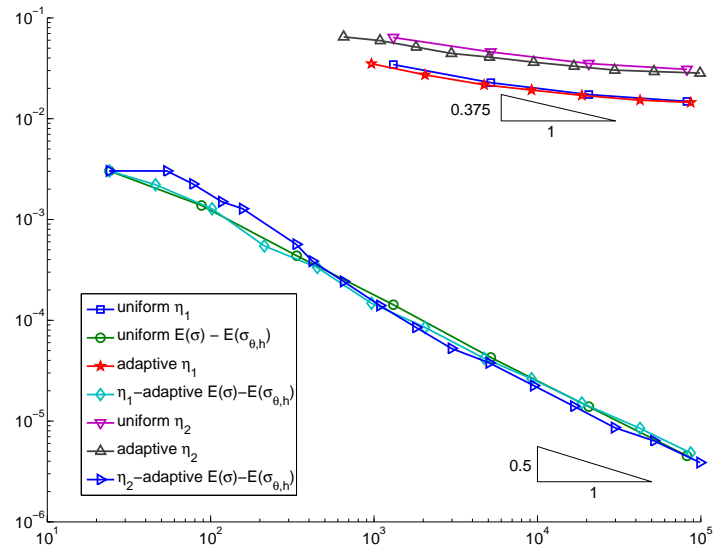


Abbildung 9.7: Vergleich des Konvergenzverhalten von den Fehlerschätzer  $\eta_{(1)}$  und  $\eta_{(2)}$  und des extrapolierten Energiefehlers  $E(\sigma) - E(\sigma_{\theta,h})$  für das Quadrat.

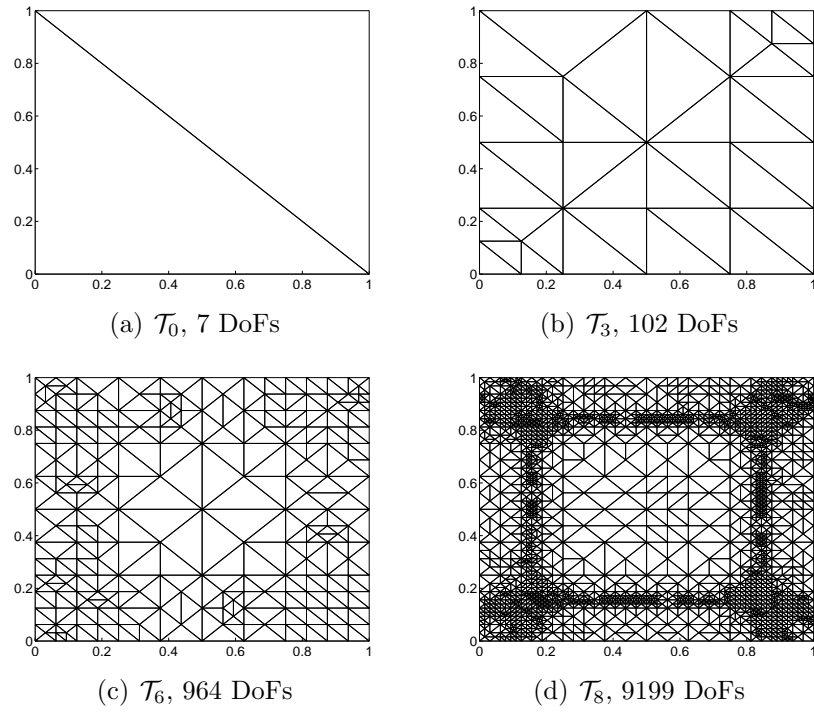


Abbildung 9.8: Folge von Triangulierungen generiert durch den Fehlerschätzer  $\eta_{(1)}$  für das Quadrat.

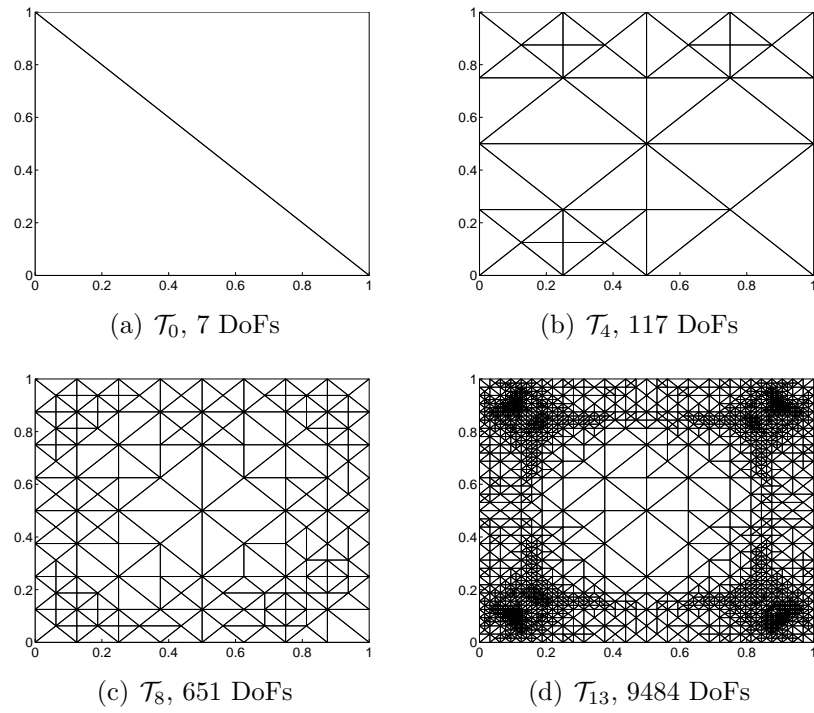


Abbildung 9.9: Folge von Triangulierungen generiert durch den Fehlerschätzer  $\eta_{(2)}$  für das Quadrat.

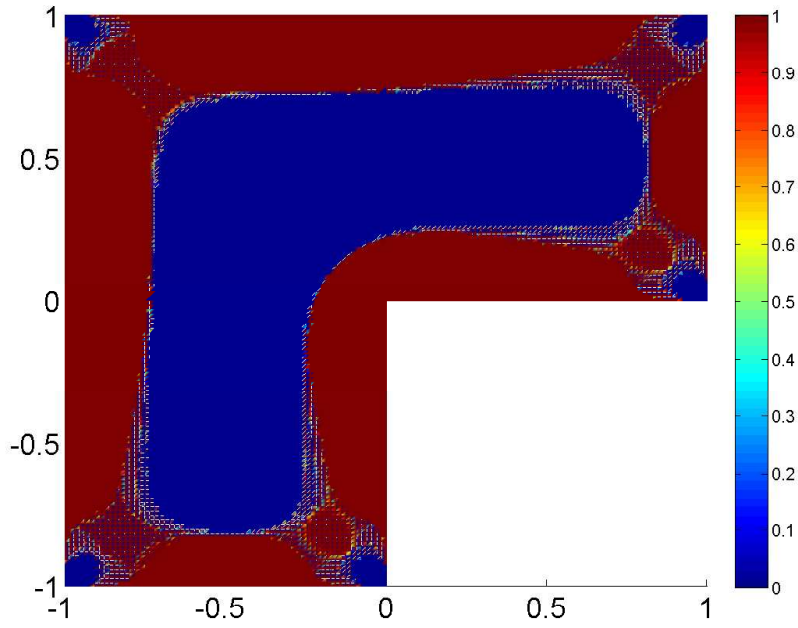


Abbildung 9.10: Massenverteilung bei uniformer Verfeinerung für das L-shape.

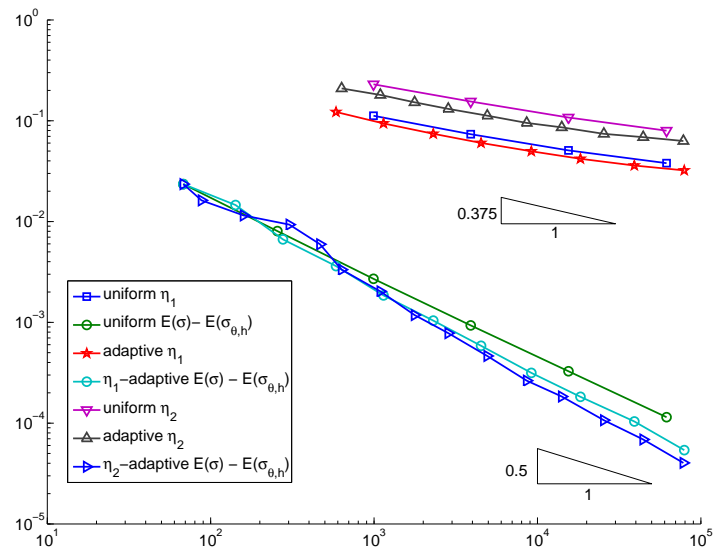


Abbildung 9.11: Vergleich des Konvergenzverhalten von den Fehlerschätzer  $\eta_{(1)}$  und  $\eta_{(2)}$  und des extrapolierten Energiefehlers  $E(\sigma) - E(\sigma_{\theta,h})$  für das L-shape.

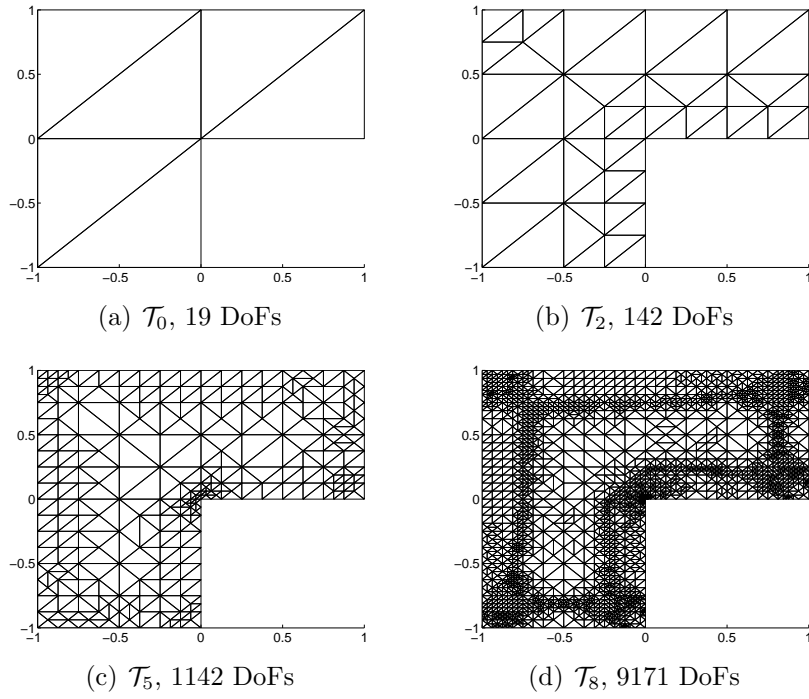


Abbildung 9.12: Folge von Triangulierungen generiert durch den Fehlerschätzer  $\eta_{(1)}$  für das L-shape.

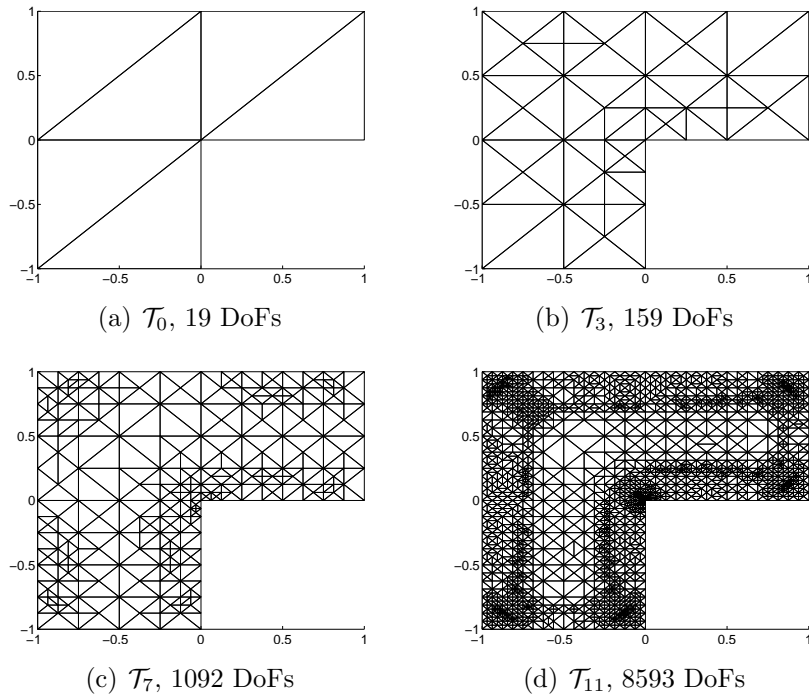


Abbildung 9.13: Folge von Triangulierungen generiert durch den Fehlerschätzer  $\eta_{(2)}$  für das L-shape.

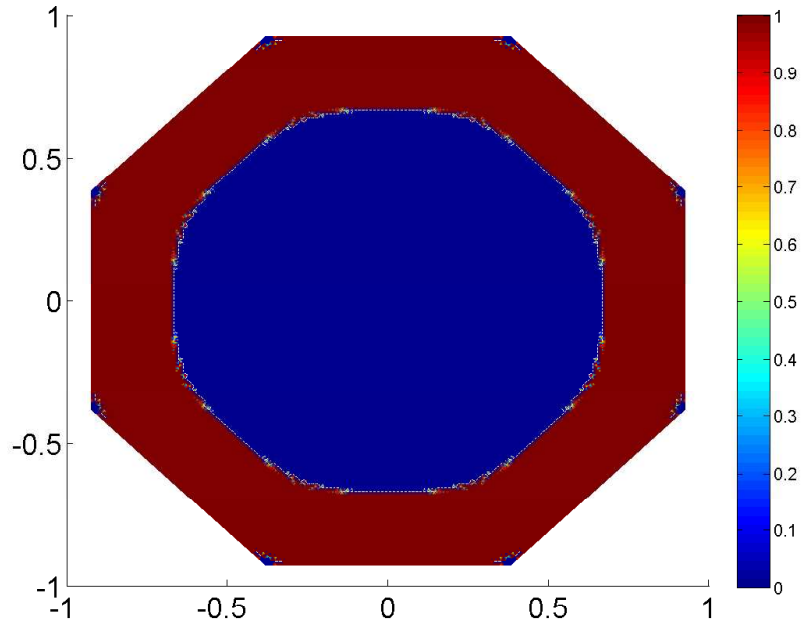


Abbildung 9.14: Massenverteilung bei uniformer Verfeinerung für das Oktagon.

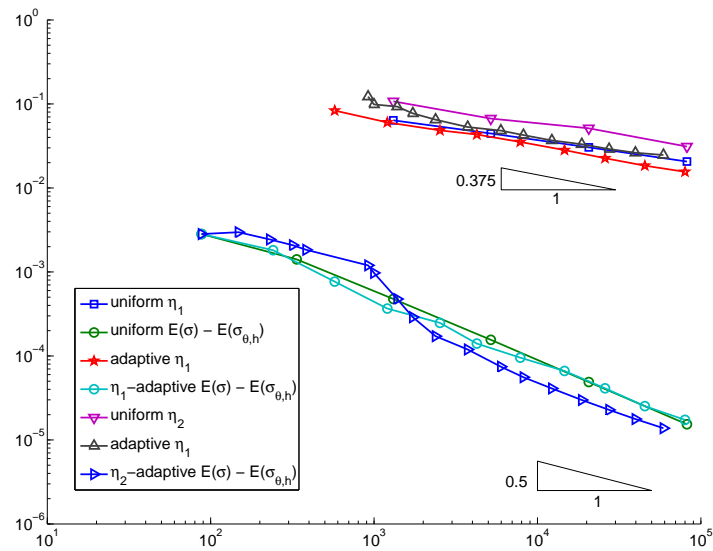


Abbildung 9.15: Vergleich des Konvergenzverhalten von den Fehlerschätzer  $\eta_{(1)}$  und  $\eta_{(2)}$  und des extrapolierten Energiefehlers  $E(\sigma) - E(\sigma_{\theta,h})$  für das Oktagon.

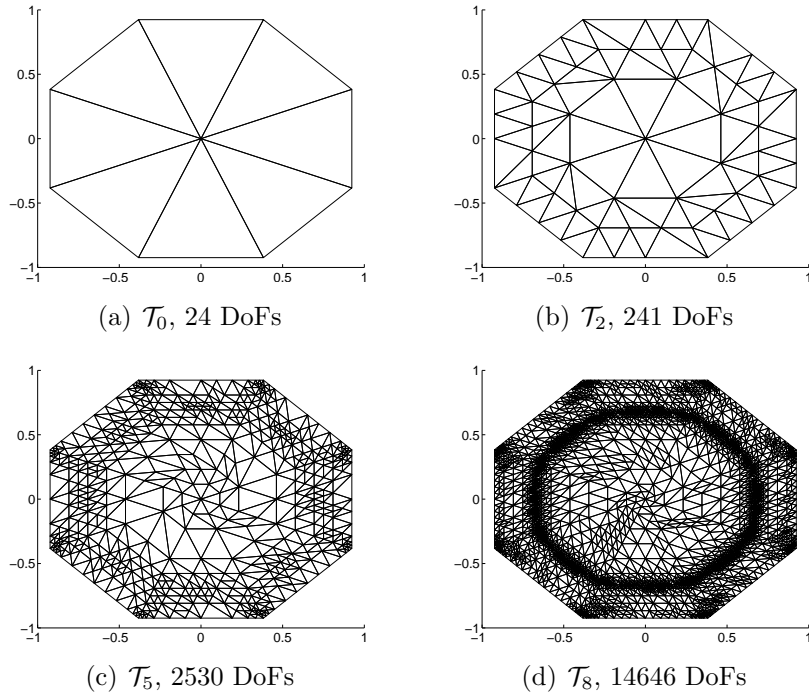


Abbildung 9.16: Folge von Triangulierungen generiert durch den Fehlerschätzer  $\eta_{(1)}$  für das Oktagon.

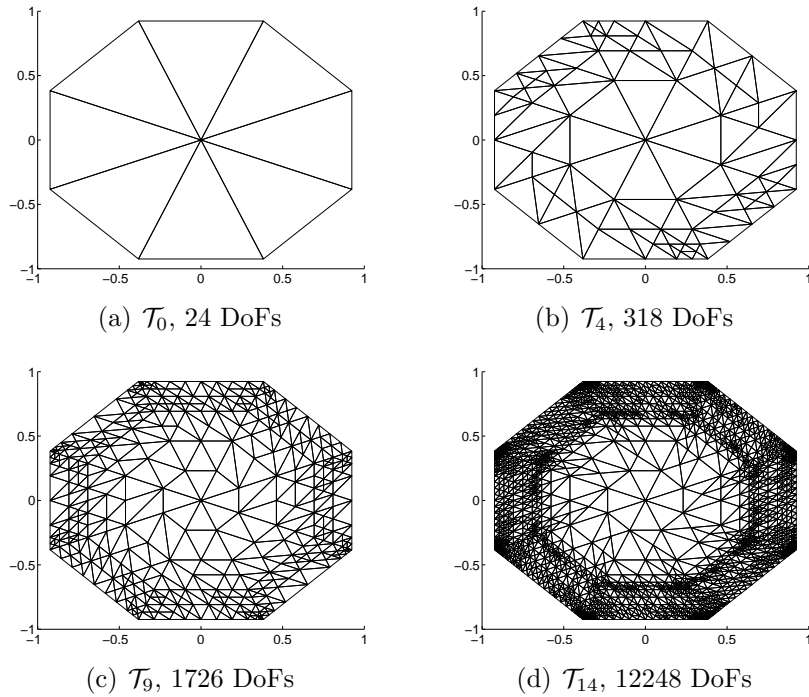


Abbildung 9.17: Folge von Triangulierungen generiert durch den Fehlerschätzer  $\eta_{(2)}$  für das Oktagon.

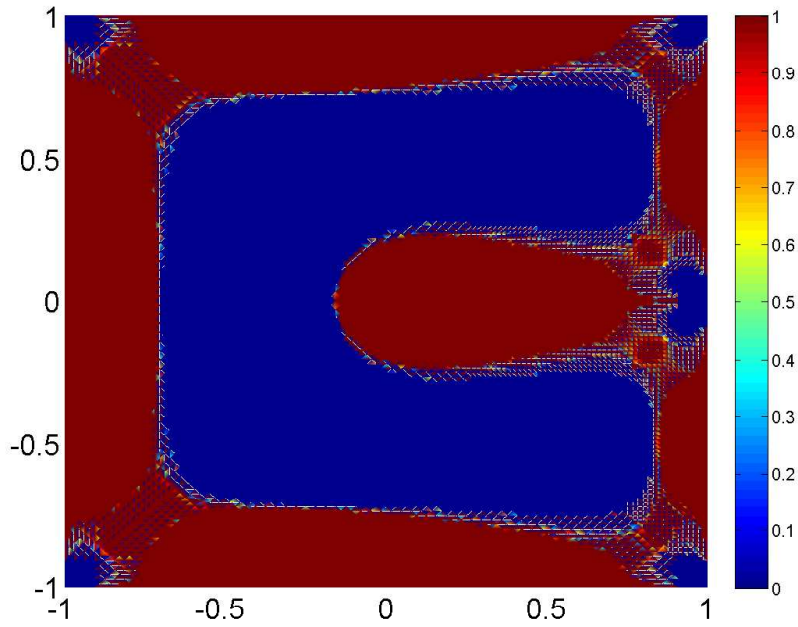


Abbildung 9.18: Massenverteilung bei uniformer Verfeinerung für das geschlitzte Quadrat.

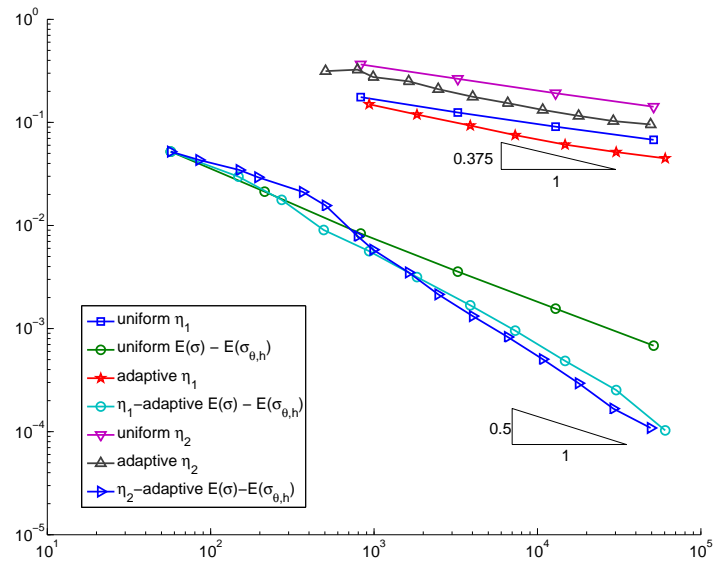
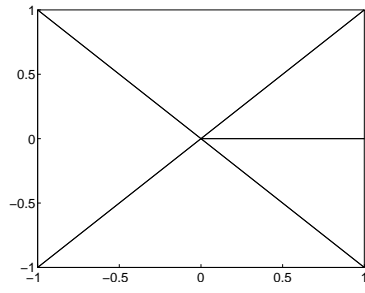
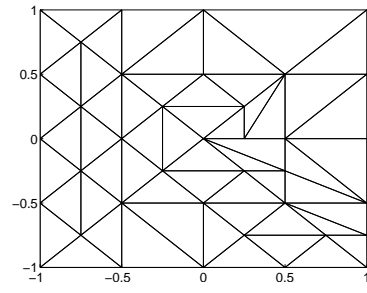


Abbildung 9.19: Vergleich des Konvergenzverhalten von den Fehlerschätzer  $\eta_{(1)}$  und  $\eta_{(2)}$  und des extrapolierten Energiefehlers  $E(\sigma) - E(\sigma_{\theta,h})$  für das geschlitzte Quadrat.

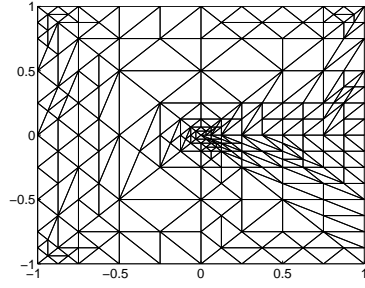




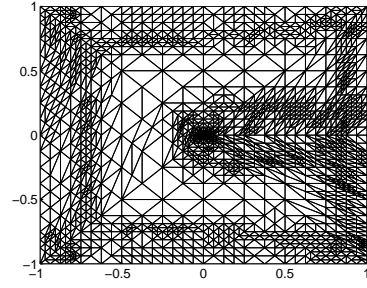
(a)  $\mathcal{T}_0$ , 16 DoFs



(b)  $\mathcal{T}_2$ , 148 DoFs

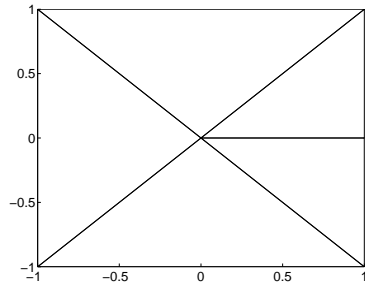


(c)  $\mathcal{T}_5$ , 934 DoFs

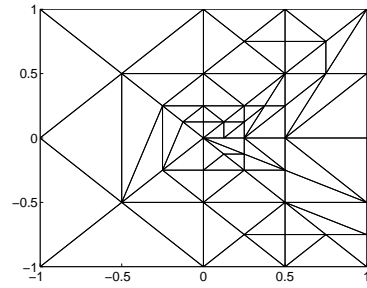


(d)  $\mathcal{T}_8$ , 7303 DoFs

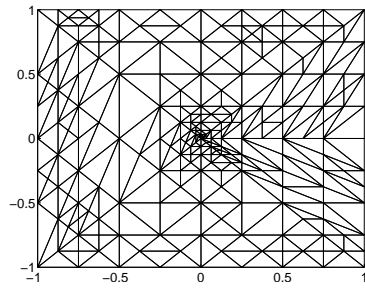
Abbildung 9.20: Folge von Triangulierungen generiert durch den Fehlerschätzer  $\eta_{(1)}$  für das geschlitzte Quadrat.



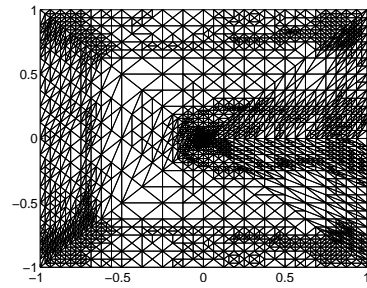
(a)  $\mathcal{T}_0$ , 16 DoFs



(b)  $\mathcal{T}_4$ , 195 DoFs



(c)  $\mathcal{T}_8$ , 988 DoFs



(d)  $\mathcal{T}_{13}$ , 10771 DoFs

Abbildung 9.21: Folge von Triangulierungen generiert durch den Fehlerschätzer  $\eta_{(2)}$  für das geschlitzte Quadrat.

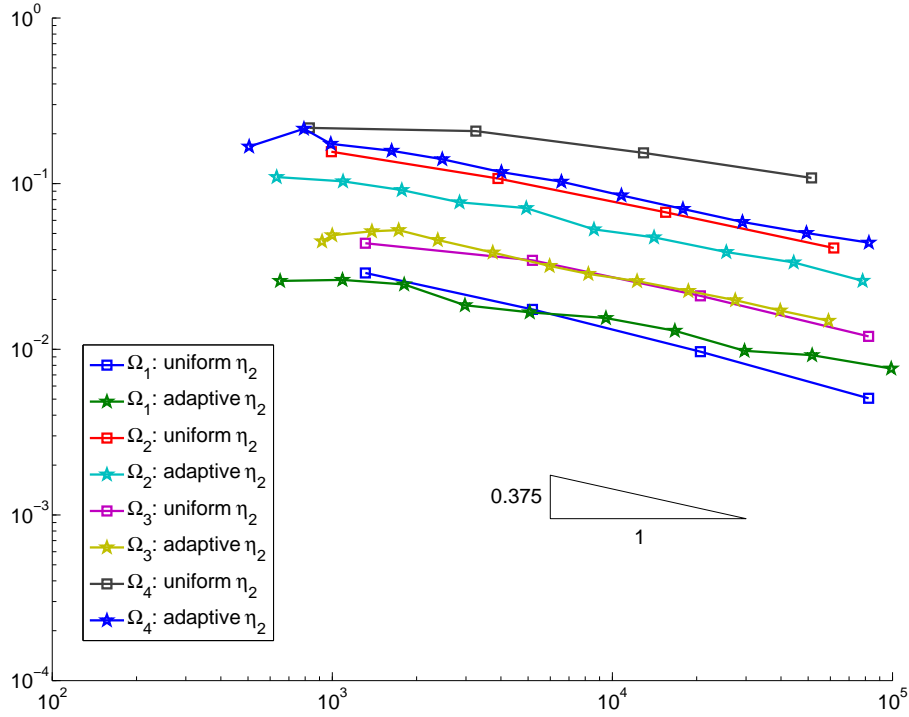


Abbildung 9.22: Konvergenzverhalten des Fehlerschätzers  $\eta_{(2)}$  eingeschränkt auf das Innere beider Materialien.

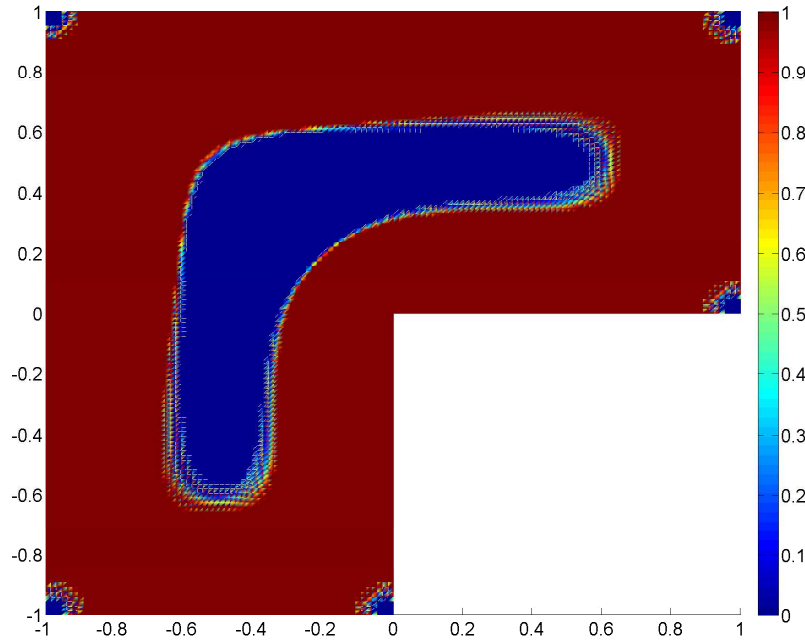
Bei allen vier Modellbeispielen lässt sich bei dem extrapolierten Energiefehler  $E(\sigma) - E(\sigma_{\theta,h})$  lineare bzw quadratische Konvergenz beobachten. Beim L-shape Gebiet und beim geschlitzten Quadrat ist eine deutliche Verbesserung der Konvergenzrate mit adaptiver Verfeinerung im Vergleich zur uniformen Verfeinerung zu erkennen. Bei den geschätzten Fehlern liegt die Konvergenzrate bei allen Beispielen ausser dem Oktagon-Beispiel deutlich unter 0.375. Im Vergleich zu den anderen drei Beispielen besitzt die Grenzschrift beim Oktagon, bei der beide Materialien zusammentreffen, scheinbar kein Inneres. Die Grenzschrift bildet einen Ring zwischen den Materialien, es findet keine Vermischung statt.

Die Modellbeispiele zeigen, dass, solange die Fläche der Grenzschrift klein genug ist, beide Fehlerschätzer effizient sind. In den Beispielen kommen beide Materialien zu 50% vor. Dies war zunächst eine willkürliche Festlegung. Das Oktagon-Beispiel zeigt, dass abhängig von der Geometrie des Gebiets die Anteile, gegeben durch den Parameter  $\xi$ , der Materialien passend gewählt werden müssen, wenn man ein effizientes Verhalten in den Fehlerschätzern erhalten möchte. Abbildung 9.23 zeigt die Massenverteilung und die geschätzten Fehler für das L-shape Gebiet. Der Parameter  $\xi$  ist hierbei auf  $\xi = 0.8$  gesetzt. Das „weiche“ Material kommt hierbei zu 80%, das „harte“ Material zu 20%, vor. Beide Fehlerschätzer zeigen ein optimales Konvergenzverhalten bei einem Materialanteil von  $\xi = 0.8$ . Ähnliche Beobachtungen zeigten auch auf die übrigen Modellbeispiele.

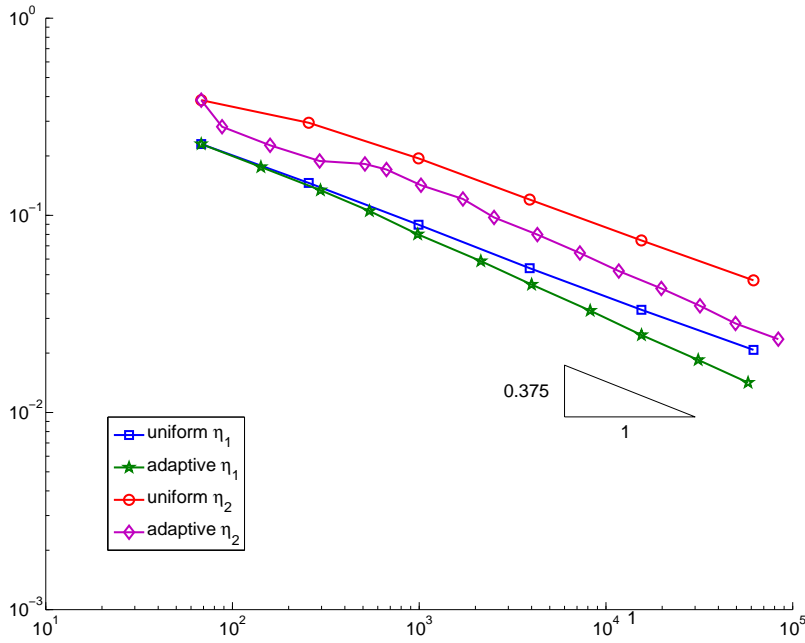
Beim Oktagon-Beispiel führt adaptive Verfeinerung wie beim L-shape Gebiet und beim geschlitzten Quadrat zu einer Verbesserung des Konvergenzverhaltens. Betrachtet man die durch die Fehlerschätzer  $\eta_{(1)}$  und  $\eta_{(2)}$  generierten Gitter, wird in allen vier Beispielen die Kontaktzone sehr fein aufgelöst, während das Innere der Materialien relativ grob verfeinert wurde. Die Konturen des Materialverlaufs lassen sich allerdings im Gitter, generiert durch  $\eta_{(1)}$ , deutlicher erkennen. Beim L-shape Beispiel wird ebenfalls zur einspringende Ecke hin stark verfeinert.

In Abbildung 9.22 haben wir das Konvergenzverhalten von  $\eta_{(2)}$  eingeschränkt auf das Innere beider Materialien dargestellt. Die Resultate bestätigen, die beim Testbeispiel (9.3) aufgestellten Beobachtungen. Beim Quadrat scheint es eine Anomalie zu geben. Während die uniforme Verfeinerung zu optimalem Verhalten führt, verschlechtert sich die Konvergenzrate bei adaptiver Verfeinerung. Bei allen anderen Beispielen beobachten wir für adaptive und uniforme Verfeinerung optimale Fehlerreduktion. Womöglich handelt es sich bei der Anomalie im Quadrat um ein vorasymptotisches Verhalten. Bei dem geschlitzten Quadrat ist zudem ein deutlicher Unterschied zwischen uniformer und adaptiver Verfeinerungen zu erkennen.

Zusammenfassend lässt sich feststellen, dass die durch beide Fehlerschätzer generierten Gitter zu optimaler Fehlerreduktion im konjugierten Energiefehler führen. Im Allgemeinen sind  $\eta_{(1)}$  und  $\eta_{(2)}$  aber nicht effizient. Erst wenn die Fläche der Grenzschicht klein genug ist, kommt es zu einem effizienten Verhalten.



(a) Massenverteilung beim L-shape Gebiet.



(b) Konvergenzhistorie mit uniformer und adaptiver Verfeinerung für die Fehlerschätzer  $\eta_{(1)}$  und  $\eta_{(2)}$ .

Abbildung 9.23: Massenverteilung und geschätzter Fehler beim L-shape Gebiet mit  $\xi = 0.8$ . Der Lagrangemultiplikator  $\lambda$  ist gegeben durch  $\lambda = 0.005157205592326$ .

## 9.4 Weitere numerische Kenngrößen

Nachdem wir im letzten Abschnitt gesehen haben, dass die Fehlerschätzer  $\eta_{(1)}$  und  $\eta_{(2)}$  im Allgemeinen nicht effizient sind, wollen wir in diesem Abschnitt weitere numerische Größen untersuchen. Ziel ist es, an Hand dieser Größen weitere numerische Belege für die Ergebnisse in Abschnitt 9.3 zu erhalten. Durch die Darstellung der Massenverteilung im letzten Abschnitt haben wir beobachtet, dass nur beim Modellbeispiel  $\Omega_3$ , dem Oktagon, die Fläche der Kontaktzone gegen Null geht. Bei allen anderen Beispielen scheint die Fläche einen positiven Wert anzunehmen. Dies ist die erste Größe, die wir hier untersuchen wollen.

**Definition 9.7** (Kontaktzone). *Sei  $C(z_1, z_2)$  wie in Bemerkung 3.7 definiert. Wir definieren die Kontaktzone  $\Gamma_C$  durch*

$$\Gamma_C(\mathcal{T}) := \{T \in \mathcal{T} \mid \nexists x, y \in T : C(\sigma_{\theta,h}(x), \sigma_{\theta,h}(y)) > 0\}. \quad (9.5)$$

Dann ist  $A_C$  gegeben durch  $A_C := |\Gamma_C|^{1/2}$ .

In der Arbeit [4] wurde dem Residualfehlerschätzer auch der Averagefehler untersucht. Dies ist unsere zweite Größe.

**Definition 9.8** (Average). *Sei  $\sigma_{\theta,h}$  Lösung von (5.7). Dann ist  $A\sigma_{\theta,h} \in \mathcal{S}_1(\mathcal{T})^2$  gegeben durch*

$$(A\sigma_{\theta,h})(z) := \frac{1}{|\omega_z|} \int_{\omega_z} \sigma_{\theta,h} dx \quad (9.6)$$

für  $z \in \mathcal{N} \setminus \partial\Omega$  mit dem Patch  $\omega_z$ , der alle Dreiecke  $T$  enthält, für die  $z \in T$  gilt. Der Averagefehler auf  $\mathcal{T}$  ist dann gegeben durch  $\|\sigma_{\theta,h} - A\sigma_{\theta,h}\|_{L^2(\mathcal{T})}$ .

Die dritte Größe ist durch die a priori Analysis (7.1) motiviert. Wir betrachten folgende Größe.

**Definition 9.9.** *Der Wert  $\eta_H$  ist definiert durch*

$$\eta_H := \sqrt{(\|\sigma_{\theta,h} - A\sigma_{\theta,h}\|_{L^2(\mathcal{T})} + A_C) \|\sigma_{\theta,h} - A\sigma_{\theta,h}\|_{L^2(\mathcal{T})}}. \quad (9.7)$$

Im Folgenden vergleichen wir die drei Größen für uniforme,  $\eta_{(1)}$ -adaptive und  $\eta_{(2)}$ -adaptive Verfeinerung an Hand der Modellbeispiele, wie sie in Abschnitt 9.1 definiert worden sind.

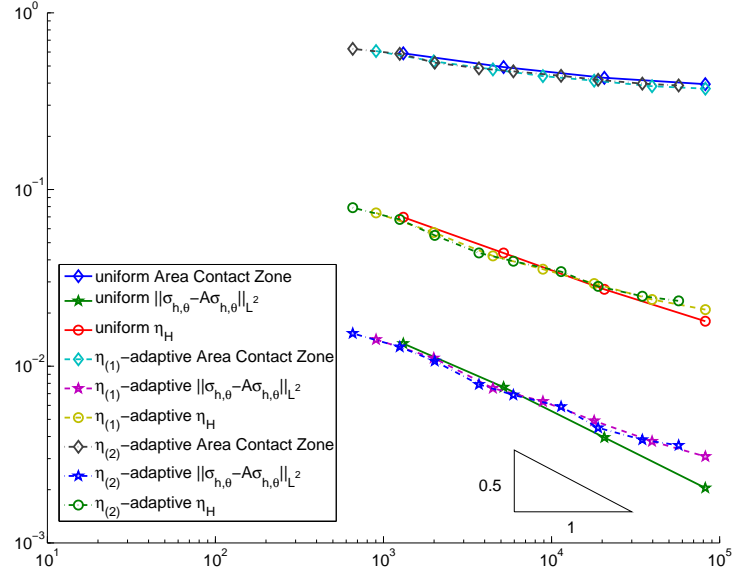


Abbildung 9.24: Konvergenzhistorie der numerischen Größen  $A_C$ ,  $\eta_H$  und  $\|\sigma_{\theta,h} - A\sigma_{\theta,h}\|_{L^2(\mathcal{T})}$  für das Modellbeispiel  $\Omega_1$  bei uniformer,  $\eta_{(1)}$ -adaptiver und  $\eta_{(2)}$ -adaptiver Verfeinerung.

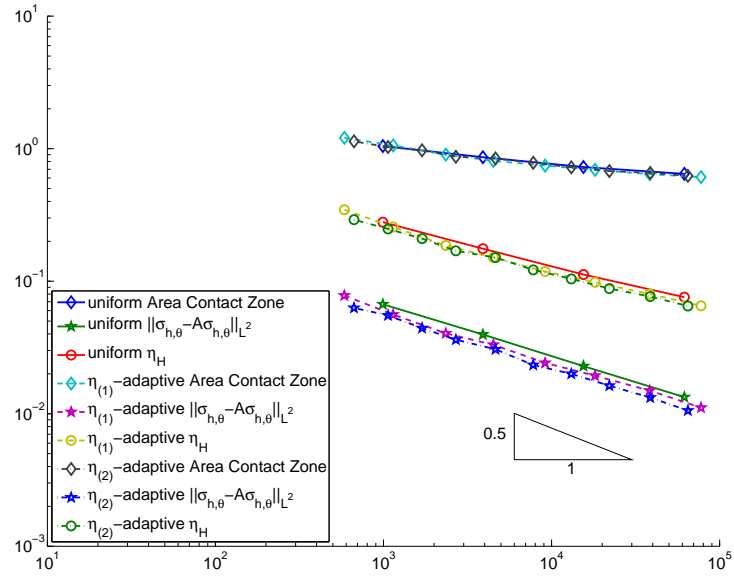


Abbildung 9.25: Konvergenzhistorie der numerischen Größen  $A_C$ ,  $\eta_H$  und  $\|\sigma_{\theta,h} - A\sigma_{\theta,h}\|_{L^2(\mathcal{T})}$  für das Modellbeispiel  $\Omega_2$  bei uniformer,  $\eta_{(1)}$ -adaptiver und  $\eta_{(2)}$ -adaptiver Verfeinerung.

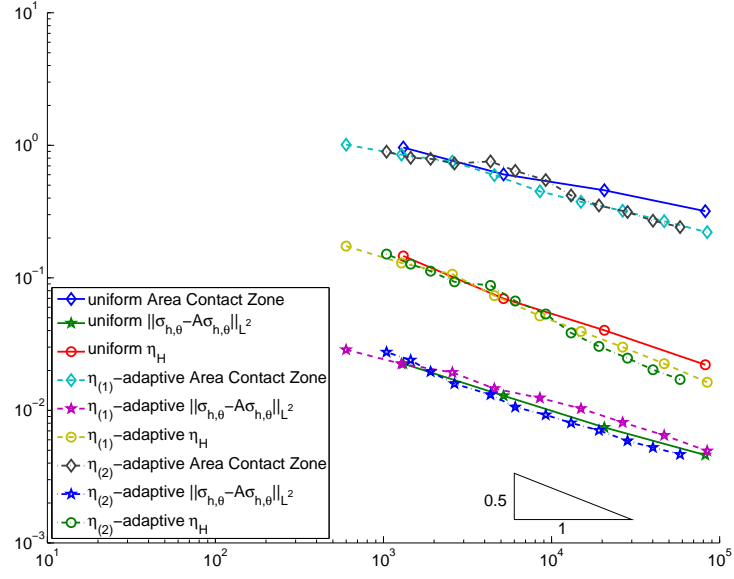


Abbildung 9.26: Konvergenzhistorie der numerischen Größen  $A_C$ ,  $\eta_H$  und  $\|\sigma_{\theta,h} - A\sigma_{\theta,h}\|_{L^2(\mathcal{T})}$  für das Modellbeispiel  $\Omega_3$  bei uniformer,  $\eta_{(1)}$ -adaptiver und  $\eta_{(2)}$ -adaptiver Verfeinerung.

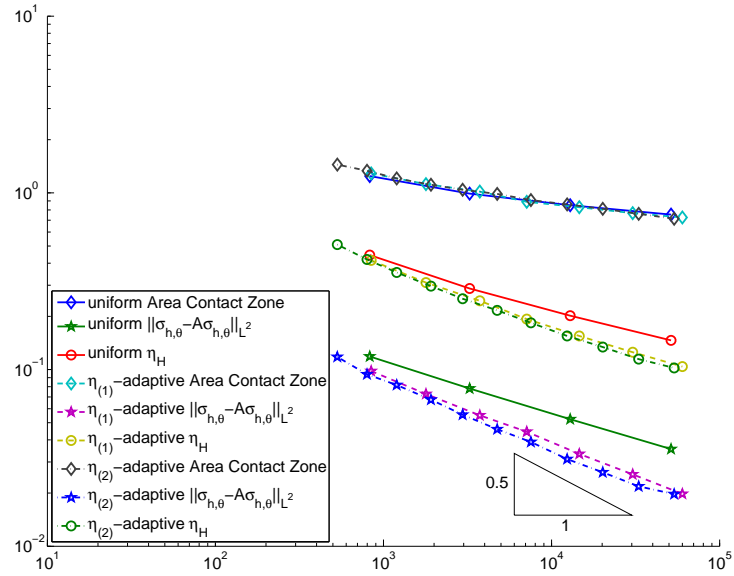


Abbildung 9.27: Konvergenzhistorie der numerischen Größen  $A_C$ ,  $\eta_H$  und  $\|\sigma_{\theta,h} - A\sigma_{\theta,h}\|_{L^2(\mathcal{T})}$  für das Modellbeispiel  $\Omega_4$  bei uniformer,  $\eta_{(1)}$ -adaptiver und  $\eta_{(2)}$ -adaptiver Verfeinerung.

Die Abbildungen 9.24-9.27 bestätigen, dass die Fläche der Kontaktzone nur bei dem Oktagonbeispiel gegen null geht. Bei allen anderen Beispielen nähert sie sich einem konstanten Wert an. Beim Oktagon bewirkt adaptive Verfeinerung eine Verbesserung der Reduktion des Wertes  $A_C$ . Betrachtet man den Averagefehler  $\|\sigma_{\theta,h} - A\sigma_{\theta,h}\|_{L^2(\mathcal{T})}$  erhalten wir bei allen vier Beispielen etwa lineare Konvergenz in  $h$ . Zwei Besonderheiten treten hier auf. Zum einen liefert adaptive Verfeinerung im Vergleich zur uniformen Verfeinerung eine Verbesserung des Konvergenzverhaltens beim Schlitzgebiet. Dies bestätigt die Beobachtungen aus Abschnitt 9.3. Das zweite Phänomen betrifft nur das Quadrat. Hier verschlechtert sich bei  $\eta_{(1)}$ -Adaptivität und  $\eta_{(2)}$ -Adaptivität das Konvergenzverhalten. Es scheint, dass die lokale Verfeinerung durch  $\eta_{(1)}$  und  $\eta_{(2)}$  nicht bei allen Beispielen für die Grösse  $\|\sigma_{\theta,h} - A\sigma_{\theta,h}\|_{L^2(\mathcal{T})}$  optimal ist. Die dritte Größe  $\eta_H$  als Kombination der anderen zwei Größen zeigt nur für das Oktagonbeispiel lineare Konvergenz in  $h$ . Bei den anderen drei Beispielen bewirkt, dass die Fläche der Kontaktzone kein Nullmaß besitzt, dass die Konvergenzrate für  $\eta_H$  um  $h^{1/2}$  liegt.

Insgesamt kann man sagen, dass solange die Fläche der Kontaktzone kein positives Maß annimmt, alle drei Größen lineare Konvergenz in  $h$  zeigen. Dies bestätigt die Beobachtungen aus Abschnitt 9.3, dass die Fehlerschätzer  $\eta_{(1)}$  und  $\eta_{(2)}$  effizient sind, wenn die Fläche der Kontaktzone gegen null geht. In Abschnitt 9.3 haben wir festgestellt, dass die Fläche der Kontaktzone von den Materialanteilen abhängt. Als wir den Materialanteil auf  $\xi = 0.8$  für das L-shape Gebiet gesetzt haben, konnten wir optimales Konvergenzverhalten in beiden Fehlerschätzern beobachten. In Abbildung 9.28 haben wir die Grössen  $A_C$ ,  $\eta_H$  und  $\|\sigma_{\theta,h} - A\sigma_{\theta,h}\|_{L^2(\mathcal{T})}$  für die Materialanteile  $\xi = 0.5$  und  $\xi = 0.8$  bei uniformer Verfeinerung miteinander verglichen. Wie man erkennt, geht die Fläche der Kontaktzone gegen null, die anderen beiden Größen konvergieren linear in  $h$  bei einem Materialanteil von  $\xi = 0.8$ .

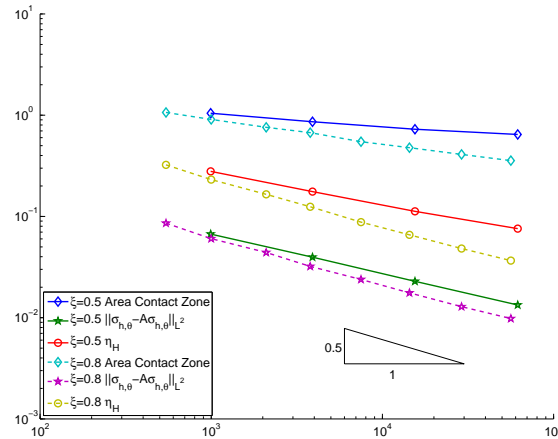


Abbildung 9.28: Vergleich von  $A_C$ ,  $\eta_H$  und  $\|\sigma_{\theta,h} - A\sigma_{\theta,h}\|_{L^2(\mathcal{T})}$  bei einem Materialanteil von  $\xi = 0.5$  und  $\xi = 0.8$  für das Modellbeispiel  $\Omega_2$  bei uniformer Verfeinerung.



# Kapitel 10

## Ein Zielorientierter Ansatz

In diesem Kapitel greifen wir die Idee aus [2] auf, mit Hilfe eines Zielfunktionals  $J$  eine adaptive Gittergenerierung zu ermöglichen. Das gemischte Optimal-Design Problem wird in eine Minimierungsaufgabe mit PDE-Nebenbedingungen eingebettet. Die Euler-Lagrange-Bedingung führt dann zu notwendigen Optimalitätsbedingungen, die die Definition eines Fehlerindikators ermöglichen. Abschließend betrachten wir die vier Modellbeispiele und vergleichen die numerischen Resultate.

Wir betrachten das nichtlineare gemischte Problem: Zu einer Last  $f \in L^2(\Omega)$  und Randdaten  $u_D \in L^2(\partial\Omega)$  bestimme  $(\sigma_\theta, u) \in H(\operatorname{div}, \Omega) \times L^2(\Omega)$ , so dass gilt

$$\begin{aligned} A(\sigma_\theta)(\tau) + B(u, \tau) &= G(\tau) \text{ für alle } \tau \in H(\operatorname{div}, \Omega) \\ B(v, \sigma_\theta) &= F(v) \text{ für alle } v \in L^2(\Omega) \end{aligned} \quad (10.1)$$

mit

$$\begin{aligned} A(\sigma_\theta)(\tau) &= \int_{\Omega} D_{\sigma} \varphi_{\theta}^*(\sigma_{\theta}) \cdot \tau \, dx, & B(v, \sigma_{\theta}) &= \int_{\Omega} v \operatorname{div} \sigma_{\theta} \, dx, \\ G(\tau) &= \int_{\partial\Omega} u_D (\tau \cdot \nu) \, ds_x, & F(v) &= \int_{\Omega} f v \, dx. \end{aligned}$$

In der zugehörigen diskreten Formulierung werden die Unterräume  $RT_0(\mathcal{T}) \subset H(\operatorname{div}, \Omega)$  und  $\mathcal{P}_0(\mathcal{T}) \subset L^2(\Omega)$  betrachtet. Das Problem ist dann gegeben durch:

Für eine Triangulierung  $\mathcal{T}$  von  $\Omega$  finde  $(\sigma_{\theta,h}, u_h) \in RT_0(\mathcal{T}) \times P_0(\mathcal{T})$ , so dass gilt

$$\begin{aligned} A(\sigma_{\theta,h})(\tau_h) + B(u_h, \tau_h) &= G(\tau_h) \text{ für alle } \tau_h \in RT_0(\mathcal{T}), \\ B(v_h, \sigma_{\theta,h}) &= F(v_h) \text{ für alle } v_h \in P_0(\mathcal{T}). \end{aligned} \quad (10.2)$$

Die Aufgabe besteht nun darin, zu einem gegebenen Zielfunktional  $J$ , Funktionen  $\sigma_\theta$  und  $u$  zu finden, so dass

$$J(\sigma_\theta, u) = \min_{\substack{v \in L^2(\Omega) \\ \tau \in H(\operatorname{div}, \Omega)}} J(\tau, v) \quad (10.3)$$

gilt, und dass  $(\sigma_\theta, u)$  das Problem (10.1) löst.

Für Lagrangemultiplikatoren  $\lambda_1 \in H(\operatorname{div}, \Omega)$  und  $\lambda_2 \in L^2(\Omega)$  ist die Lagrangefunktion  $\mathcal{L}$  für das Problem (10.3) unter den Nebenbedingungen (10.1) gegeben durch

$$\mathcal{L}(\sigma_\theta, u, \lambda_1, \lambda_2) := J(\sigma_\theta, u) - A(\sigma_\theta)(\lambda_1) - B(u, \lambda_1) + G(\lambda_1) - B(\lambda_2, \sigma_\theta) + F(\lambda_2). \quad (10.4)$$

Für einen stationären Punkt  $\{\sigma_\theta, u, \lambda_1, \lambda_2\} \in H(\operatorname{div}, \Omega) \times L^2(\Omega) \times H(\operatorname{div}, \Omega) \times L^2(\Omega)$  von  $\mathcal{L}$  muss für alle Richtungen  $\tau \in H(\operatorname{div}, \Omega)$  und  $v \in L^2(\Omega)$  die Euler-Lagrange-Bedingung gelten:

$$D\mathcal{L} = \begin{pmatrix} D_\sigma \mathcal{L} \\ D_u \mathcal{L} \\ D_{\lambda_1} \mathcal{L} \\ D_{\lambda_2} \mathcal{L} \end{pmatrix} = \begin{pmatrix} D_\sigma J(\sigma_\theta, u)(\tau) - D_\sigma A(\sigma_\theta)(\lambda_1, \tau) - B(\lambda_2, \tau) \\ D_u J(\sigma_\theta, u)(v) - B(v, \lambda_1) \\ -A(\sigma_\theta)(\tau) - B(u, \tau) + G(\tau) \\ -B(v, \sigma_\theta) + F(v) \end{pmatrix} = 0.$$

Entsprechendes gilt dann auch für das diskrete Analogon. Für einen stationären Punkt  $\{\sigma_{\theta,h}, u_h, \lambda_{1,h}, \lambda_{2,h}\} \in RT_0(\mathcal{T}) \times \mathcal{P}_0(\mathcal{T}) \times RT_0(\mathcal{T}) \times \mathcal{P}_0(\mathcal{T})$  von  $\mathcal{L}$  muss gelten, dass  $D\mathcal{L}(\sigma_{\theta,h}, u_h, \lambda_{1,h}, \lambda_{2,h})(\tau_h, v_h) = 0$  ist.

Wir wollen nun den Fehler im Zielfunktional  $J$  messen. Seien hierfür die Funktionale  $J$  und  $A$  hinreichend oft differenzierbar. Der Ausdruck  $D_\sigma^2 J$  steht hierbei für die zweimalige Ableitung von  $J$  in Richtung  $\sigma$ . Das Funktional  $A(\sigma_\theta)(\tau)$  sei in dem Argument  $\sigma_\theta$  nichtlinear und in  $\tau$  linear. Das Zielfunktional  $J(\sigma_\theta, u)$  kann in  $\sigma_\theta$  sowie in  $u$  nichtlinear sein.  $B(u, \sigma_\theta)$  ist in  $\sigma_\theta$  und in  $u$  linear.

**Theorem 10.1.** *Sei  $L(\cdot)$  ein hinreichend oft differenzierbares Funktional auf einem reellen Vektorraum  $X$ , welches bei  $x \in X$  einen stationären Punkt hat*

$$L'(x)(y) = 0 \text{ für alle } y \in X. \quad (10.5)$$

*Für einen endlich dimensionalen Unterraum  $X_h \subset X$  sei  $x_h$  die Galerkinapproximation, so dass gilt*

$$L'(x_h)(y_h) = 0 \text{ für alle } y_h \in X_h. \quad (10.6)$$

*Dann gilt*

$$L(x) - L(x_h) = 1/2 L'(x_h)(x - y_h) + \mathbb{R}_h \quad (10.7)$$

*mit einem Restterm  $\mathbb{R}_h$ , welcher kubisch im Fehler  $e^x := x - x_h$  ist:*

$$\mathbb{R}_h := 1/2 \int_0^1 L'''(x_h + s e^x)(e^x, e^x, e^x) s(s-1) ds.$$

Beweis: Der Beweis wird in [2] geführt.

**Theorem 10.2.** Für  $(\sigma_\theta, u)$  und  $(\sigma_{\theta,h}, u_h)$  seien die Bedingungen (10.5) bzw. (10.6) erfüllt. Dann gilt für beliebige Funktionen  $\tau_h, \tilde{\tau}_h \in H(\operatorname{div}, \Omega)$  und  $v_h, \tilde{v}_h \in L^2(\Omega)$

$$\begin{aligned} J(\sigma_\theta, u) - J(\sigma_{\theta,h}, u_h) &= 1/2 \rho_1(\sigma_{\theta,h}, u_h)(\lambda_1 - \tau_h) + 1/2 \rho_2(\sigma_{\theta,h}, u_h)(\lambda_2 - v_h) \\ &\quad + 1/2 \rho_1^*(\sigma_{\theta,h}, u_h)(\sigma_\theta - \tilde{\tau}_h) + 1/2 \rho_2^*(\sigma_{\theta,h}, u_h)(u - \tilde{v}_h) \quad (10.8) \\ &\quad + \mathbb{R}_h^{(1)}. \end{aligned}$$

Die Residuen sind hierbei wie folgt definiert:

$$\rho_1^*(\sigma_{\theta,h}, u_h)(\cdot) := D_\sigma J(\sigma_{\theta,h}, u_h)(\cdot) - D_\sigma A(\sigma_{\theta,h})(\lambda_{1h}, \cdot) - B(\lambda_{2h}, \cdot), \quad (10.9)$$

$$\rho_2^*(\sigma_{\theta,h}, u_h)(\cdot) := D_u J(\sigma_{\theta,h}, u_h)(\cdot) - B(\cdot, \lambda_{1h}), \quad (10.10)$$

$$\rho_1(\sigma_{\theta,h}, u_h)(\cdot) := -A(\sigma_{\theta,h})(\cdot) - B(u_h, \cdot) + G(\cdot), \quad (10.11)$$

$$\rho_2(\sigma_{\theta,h}, u_h)(\cdot) := -B(\cdot, \sigma_{\theta,h}) + F(\cdot). \quad (10.12)$$

*Beweis.* Um Theorem 10.1 anzuwenden, betrachten wir (10.4) mit  $L(x) := \mathcal{L}(\sigma_\theta, u, \lambda_1, \lambda_2)$  und  $x := \{\sigma_\theta, u, \lambda_1, \lambda_2\}$  (resp.  $x_h := \{\sigma_{\theta,h}, u_h, \lambda_{1h}, \lambda_{2h}\}$ ) auf dem Raum  $X := H(\operatorname{div}, \Omega) \times L^2(\Omega) \times H(\operatorname{div}, \Omega) \times L^2(\Omega)$ . Damit ergibt sich für Lösungen  $(\sigma_\theta, u)$  und  $(\sigma_{\theta,h}, u_h)$

$$\begin{aligned} J(\sigma_\theta, u) - J(\sigma_{\theta,h}, u_h) &= L(x) + A(\sigma_\theta)(\lambda_1) + B(u, \lambda_1) - G(\lambda_1) + B(\lambda_2, \sigma_\theta) - F(\lambda_2) \\ &\quad - L(x_h) - A(\sigma_{\theta,h})(\lambda_{1h}) - B(u_h, \lambda_{1h}) + G(\lambda_{1h}) \\ &\quad - B(\lambda_{2h}, \sigma_{\theta,h}) + F(\lambda_{2h}) \\ &= L(x) - L(x_h). \end{aligned}$$

Mit Theorem 10.1 erhalten wir für  $y_h \in X_h$  die Darstellung

$$J(\sigma_\theta, u) - J(\sigma_{\theta,h}, u_h) = 1/2 L'(x_h)(x - y_h) + \mathbb{R}_h,$$

wobei die Ableitung  $L'$  angewendet auf die Lagrangefunktion  $\mathcal{L}$  gegeben ist durch

$$\begin{aligned} L'(x_h)(x - y_h) &= D_\sigma \mathcal{L}(\sigma_{\theta,h}, u_h, \lambda_{1h}, \lambda_{2h})(\sigma_\theta - \tau_h) + D_u \mathcal{L}(\sigma_{\theta,h}, u_h, \lambda_{1h}, \lambda_{2h})(u - v_h) \\ &\quad + D_{\lambda_1} \mathcal{L}(\sigma_{\theta,h}, u_h, \lambda_{1h}, \lambda_{2h})(\lambda_1 - \tilde{\tau}_h) \\ &\quad + D_{\lambda_2} \mathcal{L}(\sigma_{\theta,h}, u_h, \lambda_{1h}, \lambda_{2h})(\lambda_2 - \tilde{v}_h) \\ &= \rho_1^*(\sigma_{\theta,h}, u_h)(\sigma_\theta - \tau_h) + \rho_2^*(\sigma_{\theta,h}, u_h)(u - v_h) \\ &\quad + \rho_1(\sigma_{\theta,h}, u_h)(\lambda_1 - \tilde{\tau}_h) + \rho_2(\sigma_{\theta,h}, u_h)(\lambda_2 - \tilde{v}_h). \end{aligned}$$

Eine explizite Darstellung des Restterms  $\mathbb{R}_h^{(1)}$  erhält man aus  $\mathbb{R}_h$  durch einsetzen von  $x = \{\sigma_\theta, u, \lambda_1, \lambda_2\}$ .  $\square$

**Theorem 10.3.** Für die in Theorem 10.2 definierten Residuen gilt:

$$\Delta \rho = \rho_1^* + \rho_2^* - \rho_1 - \rho_2 \quad (10.13)$$

mit

$$\begin{aligned}\Delta\rho := & \int_0^1 \left( D_\sigma^2 A(\sigma_{\theta,h} + se_\theta^\sigma)(\lambda_{1_h} + se^{\lambda_1}, e_\theta^\sigma, e_\theta^\sigma) - D_\sigma^2 J(\sigma_{\theta,h} + se^{\sigma_\theta}, u_h + se^u)(e_\theta^\sigma, e_\theta^\sigma) \right. \\ & - D_u^2 J(\sigma_{\theta,h} + se^{\sigma_\theta}, u_h + se^u)(e^u, e^u) \\ & \left. - 2D_\sigma D_u J(\sigma_{\theta,h} + se_\theta^\sigma, u_h + se^u)(e^u, e_\theta^\sigma) \right) ds.\end{aligned}$$

Die Funktionen  $(\sigma_\theta, u)$  und  $(\sigma_{\theta,h}, u_h)$  erfüllen die Bedingungen (10.5) bzw (10.6). Dann erhalten wir eine vereinfachte Darstellung des Fehlers im Zielfunktional.

$$J(\sigma_\theta, u) - J(\sigma_{\theta,h}, u_h) = \rho_1(\sigma_{\theta,h}, u_h)(\lambda_1 - \tau_h) + \rho_2(\sigma_{\theta,h}, u_h)(\lambda_2 - v_h) + \mathbb{R}_h^{(2)}, \quad (10.14)$$

wobei der Restterm  $\mathbb{R}_h^{(2)}$  definiert ist durch

$$\begin{aligned}\mathbb{R}_h^{(2)} := & \int_0^1 \left( -D_\sigma D_u J(\sigma_{\theta,h} + se_\theta^\sigma, u)(e^u, e_\theta^\sigma) + s \left( D_\sigma^2 A(\sigma_{\theta,h} + se_\theta^\sigma)(\lambda_1, e_\theta^\sigma, e_\theta^\sigma) \right. \right. \\ & - D_\sigma^2 J(\sigma_{\theta,h} + se^{\sigma_\theta}, u_h + se^u)(e_\theta^\sigma, e_\theta^\sigma) \\ & \left. \left. - D_u^2 J(\sigma_{\theta,h} + se^{\sigma_\theta}, u_h + se^u)(e^u, e^u) \right) \right) ds.\end{aligned}$$

*Beweis.* Wir definieren die skalare Funktion  $g : [0, 1] \rightarrow \mathbb{R}$

$$\begin{aligned}g(s) := & D_\sigma J(\sigma_{\theta,h} + se^{\sigma_\theta}, u_h + se^u)(e_\theta^\sigma) - D_\sigma A(\sigma_{\theta,h} + se_\theta^\sigma)(\lambda_{1_h} + se^{\lambda_1}, e_\theta^\sigma) \\ & - B(e^u, \lambda_{1_h} + se^{\lambda_1}) + D_u J(\sigma_{\theta,h} + se_\theta^\sigma, u_h + se^u)(e^u) - B(\lambda_{2_h} + se^{\lambda_2}, e_\theta^\sigma)\end{aligned}$$

mit den Fehlern  $e_\theta^\sigma := \sigma_\theta - \sigma_{\theta,h}$ ,  $e^u := u - u_h$ ,  $e^{\lambda_1} := \lambda_1 - \lambda_{1_h}$  und  $e^{\lambda_2} := \lambda_2 - \lambda_{2_h}$ . Es gilt

$$\begin{aligned}g(1) &= D_\sigma J(\sigma_\theta, u)(e_\theta^\sigma) - D_\sigma A(\sigma_\theta)(\lambda_1, e_\theta^\sigma) - B(\lambda_2, e_\theta^\sigma) \\ &\quad + D_u J(\sigma_\theta, u)(e^u) - B(e^u, \lambda_1) \\ &= 0, \\ g(0) &= D_\sigma J(\sigma_{\theta,h}, u_h)(e_\theta^\sigma) - D_\sigma A(\sigma_{\theta,h})(\lambda_{1_h}, e_\theta^\sigma) - B(\lambda_{2_h}, e^{\sigma_\theta}) \\ &\quad + D_u J(\sigma_{\theta,h}, u_h)(e^u) - B(e^u, \lambda_{1_h}) \\ &= \rho_1^*(\sigma_{\theta,h}, u_h)(e_\theta^\sigma) + \rho_2^*(\sigma_{\theta,h}, u_h)(e^u).\end{aligned}$$

Die Ableitung  $g'$  ist gegeben durch

$$\begin{aligned}g'(s) = & D_{\sigma_\theta}^2 J(\sigma_{\theta,h} + se^{\sigma_\theta}, u_h + se^u)(e_\theta^\sigma, e_\theta^\sigma) + D_u D_\sigma J(\sigma_{\theta,h} + se^{\sigma_\theta}, u_h + se^u)(e_\theta^\sigma, e^u) \\ & - D_{\sigma_\theta}^2 A(\sigma_{\theta,h} + se_\theta^\sigma)(\lambda_{1_h} + se^{\lambda_1}, e_\theta^\sigma, e_\theta^\sigma) - D_\sigma A(\sigma_{\theta,h} + se_\theta^\sigma)(e^{\lambda_1}, e_\theta^\sigma) \\ & - B(e^{\lambda_2}, e_\theta^\sigma) + D_u^2 J(\sigma_{\theta,h} + se_\theta^\sigma, u_h + se^u)(e^u, e^u) \\ & + D_\sigma D_u J(\sigma_{\theta,h} + se_\theta^\sigma, u_h + se^u)(e^u, e_\theta^\sigma) - B(e^u, e^{\lambda_1}).\end{aligned}$$

Mit dem Hauptsatz der Differential- und Integralrechnung folgt:

$$\begin{aligned}\rho_1^*(\sigma_{\theta,h}, u_h)(e_\theta^\sigma) + \rho_2^*(\sigma_{\theta,h}, u_h)(e^u) &= g(0) = g(0) - g(1) = - \int_0^1 g'(s) ds \\ &= \Delta\rho + \rho_1(\sigma_{\theta,h}, u_h)(e^{\lambda_1}) + \rho_2(\sigma_{\theta,h}, u_h)(e^{\lambda_2}).\end{aligned}$$

Damit haben wir (10.13) gezeigt. Aus der Definition von  $\mathbb{R}_h^{(2)}$ , partieller Integration und Galerkin-Orthogonalität folgt (10.14)

$$\begin{aligned}
\mathbb{R}_h^{(2)} &= \int_0^1 -D_\sigma D_u J(\sigma_{\theta,h} + se_\theta^\sigma, u)(e^u, e_\theta^\sigma) + s( D_\sigma^2 A(\sigma_{\theta,h} + se_\theta^\sigma)(\lambda_1, e_\theta^\sigma, e_\theta^\sigma) \\
&\quad - D_\sigma^2 J(\sigma_{\theta,h} + se^{\sigma_\theta}, u_h + se^u)(e_\theta^\sigma, e_\theta^\sigma) - D_u^2 J(\sigma_{\theta,h} + se^{\sigma_\theta}, u_h + se^u)(e^u, e^u) ) ds \\
&= D_u J(\sigma_{\theta,h}, u)(e^u) - D_u J(\sigma_\theta, u)(e^u) + D_u J(\sigma_\theta, u)(e^u) - B(e^u, \lambda_1) \\
&\quad + \int_0^1 s( D_\sigma^2 A(\sigma_{\theta,h} + se_\theta^\sigma)(\lambda_1, e_\theta^\sigma, e_\theta^\sigma) - D_\sigma^2 J(\sigma_{\theta,h} + se^{\sigma_\theta}, u_h + se^u)(e_\theta^\sigma, e_\theta^\sigma) \\
&\quad - D_u^2 J(\sigma_{\theta,h} + se^{\sigma_\theta}, u_h + se^u)(e^u, e^u) ) ds \\
&= D_\sigma J(\sigma_\theta, u)(e_\theta^\sigma) - D_\sigma A(\sigma_\theta)(\lambda_1, e_\theta^\sigma) - B(\lambda_2, e_\theta^\sigma) + D_u J(\sigma_{\theta,h}, u)(e^u) - B(e^u, \lambda_1) \\
&\quad + \int_0^1 s( D_\sigma^2 A(\sigma_{\theta,h} + se_\theta^\sigma)(\lambda_1, e_\theta^\sigma, e_\theta^\sigma) - D_\sigma^2 J(\sigma_{\theta,h} + se^{\sigma_\theta}, u_h + se^u)(e_\theta^\sigma, e_\theta^\sigma) \\
&\quad - D_u^2 J(\sigma_{\theta,h} + se^{\sigma_\theta}, u_h + se^u)(e^u, e^u) ) ds \\
&= -B(e^u, \lambda_1) - B(\lambda_2, e_\theta^\sigma) + \int_0^1 \{ -D_\sigma A(\sigma_{\theta,h} + se_\theta^\sigma)(\lambda_1, e_\theta^\sigma) \\
&\quad + D_\sigma J(\sigma_{\theta,h} + se_\theta^\sigma, u)(e_\theta^\sigma) + D_u J(\sigma_{\theta,h}, u_h + se^u)(e^u) \} ds \\
&= A(\sigma_{\theta,h})(\lambda_1) + B(u_h, \lambda_1) - A(\sigma_\theta)(\lambda_1) - B(u, \lambda_1) + B(\lambda_2, \sigma_{\theta,h}) - B(\lambda_2, \sigma_\theta) \\
&\quad + J(\sigma_\theta, u) - J(\sigma_{\theta,h}, u_h) \\
&= A(\sigma_{\theta,h})(\lambda_1) + B(u_h, \lambda_1) - G(\lambda_1) + B(\lambda_2, \sigma_{\theta,h}) - F(\lambda_2) + J(\sigma_\theta, u) - J(\sigma_{\theta,h}, u_h) \\
&= -\rho_1(\sigma_{\theta,h}, u_h)(\lambda_1 - \tau_h) - \rho_2(\sigma_{\theta,h}, u_h)(\lambda_2 - v_h) + J(\sigma_\theta, u) - J(\sigma_{\theta,h}, u_h)
\end{aligned}$$

für Funktionen  $\tau_h \in H(\text{div}, \Omega)$  und  $v_h \in L^2(\Omega)$ . □

Mit (10.8) und (10.14) haben wir zwei verschiedene Darstellungen des Fehlers im Ziel-funktional. Beide hängen allerdings von den schwachen Lösungen  $\sigma_\theta$ ,  $u$ ,  $\lambda_1$  und  $\lambda_2$  ab. Um eine adaptive Gittergenerierung zu ermöglichen, werden die schwachen Lösungen durch eine Glättung  $A$  der diskreten Lösungen ersetzt. Die geglätteten Lösungen sind stückweise lineare, global stetige Funktionen. Da die diskreten Funktionen  $u_h$  und  $\sigma_{\theta,h}$  stückweise konstant bzw Raviart-Thomas Funktionen sind, kann man die Glättung als Approximation höherer Ordnung interpretieren. Wenn man die Restterme vernachlässigt, erhält man so zwei approximative Darstellungen des Fehlers  $e$

$$\begin{aligned}
e_1(\sigma_{\theta,h}, u_h) &= 1/2 \rho_1(\sigma_{\theta,h}, u_h)(A\lambda_{1h} - \lambda_{1h}) + 1/2 \rho_2(\sigma_{\theta,h}, u_h)(A\lambda_{2h} - \lambda_{2h}) \\
&\quad + 1/2 \rho_1^*(\sigma_{\theta,h}, u_h)(A\sigma_{\theta,h} - \sigma_{\theta,h}) + 1/2 \rho_2^*(\sigma_{\theta,h}, u_h)(Au_h - u_h)
\end{aligned} \tag{10.15}$$

und

$$e_2(\sigma_{\theta,h}, u_h) = \rho_1(\sigma_{\theta,h}, u_h)(A\lambda_{1h} - \lambda_{1h}) + \rho_2(\sigma_{\theta,h}, u_h)(A\lambda_{2h} - \lambda_{2h}). \tag{10.16}$$

**Bemerkung 10.4.** Mit Hilfe der Gleichung (10.13) lassen sich die Restterme  $\mathbb{R}_h^{(1)}$  und

$\mathbb{R}_h^{(2)}$  ineinander überführen

$$\begin{aligned}
\mathbb{R}_h^{(2)} &= -\rho_1(\sigma_{\theta,h}, u_h)(\lambda_1 - \tau_h) - \rho_2(\sigma_{\theta,h}, u_h)(\lambda_2 - v_h) + J(\sigma_\theta, u) - J(\sigma_{\theta,h}, u_h) \\
&= -\rho_1(\sigma_{\theta,h}, u_h)(\lambda_1 - \tau_h) - \rho_2(\sigma_{\theta,h}, u_h)(\lambda_2 - v_h) + \\
&\quad \frac{1}{2}(\rho_1^*(\sigma_{\theta,h}, u_h)(\sigma_\theta - \tau_h) + \rho_2^*(\sigma_{\theta,h}, u_h)(u - v_h) + \\
&\quad \rho_1(\sigma_{\theta,h}, u_h)(\lambda_1 - \tilde{\tau}_h) + \rho_2(\sigma_{\theta,h}, u_h)(\lambda_2 - \tilde{v}_h)) + \mathbb{R}_h^{(1)} \\
&= \frac{1}{2}\Delta\rho + \mathbb{R}_h^{(1)}.
\end{aligned}$$

Somit können wir mit

$$\begin{aligned}
|\Delta\rho| &\approx |\rho_1^*(\sigma_{\theta,h}, u_h)(A\sigma_{\theta,h} - \sigma_{\theta,h}) + \rho_2^*(\sigma_{\theta,h}, u_h)(Au_h - u_h) - \\
&\quad \rho_1(\sigma_{\theta,h}, u_h)(A\lambda_{1_h} - \lambda_{1_h}) + \rho_2(\sigma_{\theta,h}, u_h)(A\lambda_{2_h} - \lambda_{2_h})|
\end{aligned} \tag{10.17}$$

den Linearisierungsfehler kontrollieren.

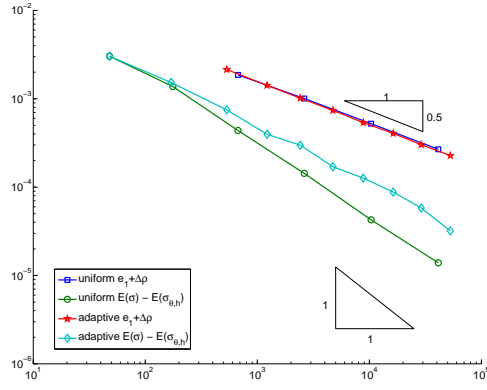
**Theorem 10.5.** Sei  $J(\sigma_{\theta,h}, u_h) = \int_\Omega \varphi^*(\sigma_{\theta,h}) dx$ . Dann gilt für stückweise konstante Lasten  $f$  und homogenen Dirichletdaten  $u_D$

$$\|\sigma_\theta - \sigma_{\theta,h}\|_{L^2(\Omega)}^2 \leq |J(\sigma_\theta, u) - J(\sigma_{\theta,h}, u_h)|. \tag{10.18}$$

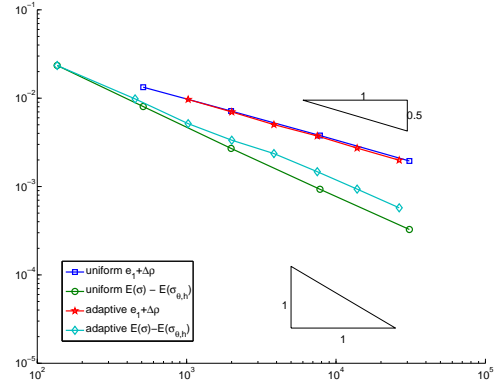
*Beweis.*

$$\begin{aligned}
\|\sigma_\theta - \sigma_{\theta,h}\|_{L^2(\Omega)}^2 &\leq \int_\Omega (D_\sigma \varphi^*(\sigma_\theta) - D_\sigma \varphi^*(\sigma_{\theta,h})) \cdot (\sigma_\theta - \sigma_{\theta,h}) dx \\
&= \int_\Omega \nabla u \cdot (\sigma_\theta - \sigma_{\theta,h}) - D_\sigma \varphi^*(\sigma_{\theta,h}) \cdot (\sigma_\theta - \sigma_{\theta,h}) dx \\
&= \int_{\partial\Omega} u_D(\sigma_\theta - \sigma_{\theta,h}) \cdot \nu ds_x - \int_\Omega u(f - f_h) dx \\
&\quad - \int_\Omega D_\sigma \varphi^*(\sigma_{\theta,h}) \cdot (\sigma_\theta - \sigma_{\theta,h}) dx \\
&= - \int_\Omega D_\sigma \varphi^*(\sigma_{\theta,h}) \cdot (\sigma_\theta - \sigma_{\theta,h}) dx \\
&\leq \int_\Omega \varphi^*(\sigma_{\theta,h}) - \varphi^*(\sigma_\theta) dx \\
&\leq |J(\sigma_\theta, u) - J(\sigma_{\theta,h}, u_h)|.
\end{aligned}$$

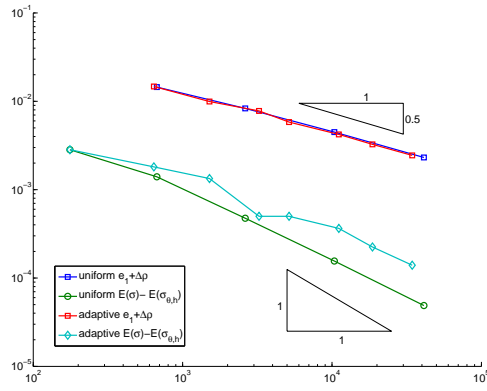
□



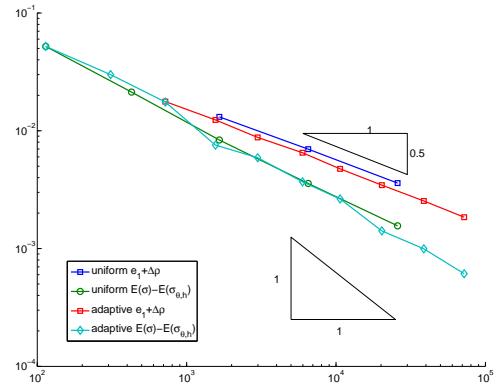
(a)  $\Omega_1$



(b)  $\Omega_2$



(c)  $\Omega_3$



(d)  $\Omega_4$

Abbildung 10.1: Darstellung des geschätzten Fehlers  $e_1 + \Delta\rho$  und des extrapolierten Energiefehlers  $E(\sigma) - E(\sigma_{\theta,h})$  für uniforme und adaptive Verfeinerung.

In Abbildung 10.1 ist die Konvergenzhistorie für die vier Modellbeispiele für den Fehlerschätzer  $e_1 + \Delta\rho$  und den extrapolierten Energiefehler  $E(\sigma) - E(\sigma_{\theta,h})$ . Als Zielfunktional wurde das Energiefunktional (9.4) verwendet. Sowohl für uniforme als auch adaptive Verfeinerung beobachten wir bei allen vier Modellbeispielen ungefähr lineare Konvergenz in  $h$  für den Fehlerschätzer. Zwar ist kein Unterschied zwischen Uniformität und Adaptivität zu erkennen, dennoch ist das Konvergenzverhalten besser als für die Fehlerschätzer  $\eta_{(1)}$  und  $\eta_{(2)}$  aus (8.12) und (8.13). Dort haben wir nur eine Konvergenz von höchstens  $h^{3/4}$  beobachten können unter der Annahme das die Fläche der Kontaktzone hinreichend klein ist. Für die Beispiele, die in Abbildung 10.1 gerechnet wurden, waren die Materialanteile auf  $\xi = 0.5$  gesetzt. Hierfür war in der Regel für  $\eta_{(1)}$  und  $\eta_{(2)}$  kein effizientes Verhalten zu beobachten im Gegenzug zum zielorientierten Ansatz.

Betrachtet man den extrapolierten Energiefehler  $E(\sigma) - E(\sigma_{\theta,h})$ , so lässt sich für das Quadrat, das L-shape und das Oktagon asypmtotisch kaum ein Unterschied im Konvergenzverhalten zwischen uniformer und adaptiver Verfeinerung beobachten. Wir erhalten

ungefähr quadratische Konvergenz in  $h$ . Beim Schlitzgebiet lässt sich eine kleine Verbesserung des Verhaltens bei adaptiver Verfeinerung beobachten. Diese Beobachtungen entsprechen den numerischen Ergebnissen für die Fehlerschätzer  $\eta_{(1)}$  und  $\eta_{(2)}$ . Hier war allerdings der Unterschied zwischen Uniformität und Adaptivität beim Schlitzgebiet deutlicher zu erkennen.

Zusammenfassend lässt sich sagen, dass der zielorientierte Ansatz sowohl keine Probleme mit der fehlenden Regularität des Energiefunktionals  $\varphi^*$  als auch mit einer positiven Fläche der Kontaktzone hat. Man beobachtet effizientes und zuverlässiges Verhalten des Fehlerschätzer bei einem optimalen Konvergenzverhalten. Es ließ sich kaum ein Unterschied in adaptiver und uniformer Verfeinerung feststellen. Nur beim Schlitzgebiet  $\Omega_4$  konnte man eine leichte Verbesserung mit Adaptivität erhalten. Vermutlich würde man nach umfangreicheren Berechnungen, über  $100k$  Freiheitsgrade, einen deutlicheren Unterschied beobachten. Die Rechnerkapazität hat dies leider nicht ermöglicht.



# Kapitel 11

## Thesen

- (I) Wir haben in dieser Arbeit die Modellierung des Optimal-Design Problems rekapituliert und anschließend die gemischte Formulierung aufgestellt. Beim Übergang in diese Formulierung haben wir eine Regularitätsordnung des nichtlinearen Funktionals verloren. Die konjugierte Energiedichte  $\varphi^*$  ist streng konvex und stetig, wogegen die Energiedichte  $\varphi$  streng konvex und stetig differenzierbar war.
- (II) Für die numerischen Ergebnisse benötigten wir die Ableitung der konjugierten Energiedichte. Um Schwierigkeiten mit dem Subdifferential zu umgehen, haben wir eine Yosida-Moreau-Regularisierung durchgeführt. Die resultierende regularisierte Energiedichte  $\varphi_\theta^*$  ist streng konvex und stetig differenzierbar. Die Ableitung  $D\varphi_\theta^*$  ist streng monoton.
- (III) Sowohl für die kontinuierliche als auch für die diskrete gemischte Formulierung haben wir Existenz und Eindeutigkeit der Lösungen  $(\sigma_\theta, u)$  respektive  $(\sigma_{\theta,h}, u_h)$  gezeigt. Hier hat sich ein Vorteil der gemischten Methode gezeigt. Für eine konforme Diskretisierung, wie z.B. mit  $\mathcal{S}_1(\mathcal{T})$ , haben wir nur Eindeutigkeit in der Spannung  $\sigma$ . Die Verschiebungen  $u$  können aufgrund der fehlenden strengen Monotonie von  $D\varphi$  mehrdeutig sein.
- (IV) In der a priori Analysis wurde die Konvergenzgüte der FE-Methode untersucht. Wir können für den Spannungsfehler  $\|\sigma_\theta - \sigma_{\theta,h}\|_{L^2(\Omega)}$  ein Konvergenzverhalten von maximal  $h^{3/4}$  erwarten. In der anschließenden a posteriori Analysis wurden zwei zuverlässige Fehlerschätzer  $\eta_{(1)}$  und  $\eta_{(2)}$  motiviert. Die numerische Ergebnisse schließen unter gewissen Voraussetzungen die Lücke zur fehlenden Effizienz. Ein alternativer Zugang zu einem Fehlerindikator wurde mit einem zielorientierten Ansatz ebenfalls präsentiert.
- (V) Wir haben einen Algorithmus vorgestellt, mit dem sich aus dem regularisierten Problem für  $\theta \rightarrow 0$  die diskrete Lösung berechnen lässt. Der Regularisierungsparameter  $\theta$  ist lokal von der Gitterweite  $h_T$  abhängig. In einem numerischen Testbeispiel haben wir diese Abhängigkeit untersucht, und festgestellt, dass  $\theta|_T = h_T$  die besten Ergebnisse liefert.

- (VI) In numerischen Beispielen wurden die theoretischen Überlegungen verifiziert. Es zeigte sich, dass beide Fehlerschätzer effizient sind, solange die Fläche der Kontaktzone beider Materialien das Maß null besitzt. In der Modellierung des Optimal-Design Problems mussten wir die nicht konvexe Energiedichte relaxieren, in dem wir zur konvexen Hülle übergegangen sind. Da die Lösungen des relaxierten Problems und des nicht relaxierten Problems nur äquivalent sind, wenn das Maß der Grenzsicht null ist, können wir dies als Voraussetzung an das Optimal-Design Problem sehen. Damit sind beide Fehlerschätzer  $\eta_{(1)}$  und  $\eta_{(2)}$  effizient. Die Fläche der Kontaktzone ist über die Materialanteile  $\xi$  kontrollierbar.
- (VII) Weitere numerische Größen, wie die Fläche der Kontaktzone  $A_C$ , der Averagefehler  $\|\sigma_{\theta,h} - A\sigma_{\theta,h}\|_{L^2(\Omega)}$  und ein aus der a priori Analysis motivierter Wert  $\eta_H$ , wurden untersucht, wobei sich optimales Verhalten auch nur bei einem Nullmaß der Fläche einstellte.
- (VIII) Für den zielorientierten Ansatz haben wir das konjugierte Energiefunktional als Zielfunktional betrachtet. Der resultierende Fehlerindikator zeigte lineare Konvergenz in der Gitterweite  $h_{\mathcal{T}}$  auch bei positiver Fläche der Kontaktzone. Im konjugierten Energiefehler beobachteten wir annähernd quadratische Konvergenz in  $h_{\mathcal{T}}$ .
- (IX) Im Rahmen dieser Arbeit ist eine Software entstanden, welche im Anhang dokumentiert wurde. Die Software wie auch die numerischen Daten befinden sich auf der beiliegenden CD.

# Anhang A

## Dokumentation der Software

In diesem Kapitel geben wir eine Einführung in die Software, mit Hilfe derer die numerischen Experimente geführt wurden. Die Software wurde in MATLAB v.R2007a geschrieben und basiert auf dem Softwarepaket F<sub>F</sub>W . Eine Dokumentation des F<sub>F</sub>W findet sich in [8]. Die Implementierung von grundlegende Algorithmen der Finite Elemente Theorie, wie z.B. Markierung und Verfeinerung von Gittern, werden dort beschrieben. In der Optimal-Design Software werden diese Algorithmen sowie Integrationsroutinen und allgemeine Hilfsfunktionen übernommen. Wir konzentrieren uns im Folgenden auf die Beschreibung der in dieser Arbeit präsentierten FE-Methoden und deren Algorithmen und wie man diese FE-Methoden für die numerischen Experimente nutzen kann. Wir benutzen in der Software den nichtlinearen MATLAB-Löser `fsolve`. Dieser Löser ist für vollbesetzte Matrizen konzipiert worden. Um ihn für unser Problem nutzen zu können, wurde die Optimierungstoolbox von MATLAB so angepasst, dass sie auch mit `sparse`-Matrizen effektiv arbeiten kann. In Abschnitt A.5 haben wir als Überblick für die in der Software auftretenden Dateien den Verzeichnissbaum aufgeführt.

### A.1 Überblick über die verfügbaren FE-Methoden und Module

Die Software lässt sich über das Startskript `startOptimalDesign.m`, welches sich im root-Verzeichnis befindet, starten. Hier werden die FE-Methode, die Markierungsstrategie, der nichtlineare Löser, die Problemdefinition und diverse Parameter definiert und anschließend die Berechnung der Optimal-Design Lösung gestartet. Im Folgenden geben wir eine Übersicht, welche Methoden und Probleme dem Benutzer zur Verfügung stehen. In den Klammern stehen die jeweiligen Variablennamen, welche in der Software Verwendung finden. Durch Ein- und Auskommentieren der jeweiligen Zeilen im Startskript lassen sich die Module ein- und ausschalten. Die numerischen Ergebnisse dieser Arbeit befinden sich im Verzeichniss `./results/` in Form von mat-Dateien, die die Datenstruktur `p` enthalten. Diese können in MATLAB nach dem Hinzufügen der Softwarepfade geladen werden. Die Pfade lassen sich durch einmaliges Ausführen des Startskriptes hinzufügen.

## FE-Methoden (pdeSolver)

- $\mathcal{S}_1(\mathcal{T})$ -Diskretisierung der konformen Formulierung (ODP1),
- $\mathcal{S}_2(\mathcal{T})$ -Diskretisierung der konformen Formulierung (ODP2),
- $RT_0(\mathcal{T})$ - $\mathcal{P}_0(\mathcal{T})$ -Diskretisierung der gemischten Formulierung (ODRT), siehe Abschnitt 5.2,
- $RT_0(\mathcal{T})$ - $\mathcal{P}_0(\mathcal{T})$ -Diskretisierung einer zielorientierter gemischten Formulierung, siehe Kapitel 10 (ODRTGOAL).

```
pdeSolver = 'ODP1';  
% pdeSolver = 'ODP2';  
% pdeSolver = 'ODRT';  
% pdeSolver = 'ODRTGOAL';
```

Abbildung A.1: Definition der FE-Methoden. Auszug aus dem Startskript `startOptimalDesign.m`

Die zugehörigen Dateien der Methoden befinden sich im Ordner `./PDEsolvers/optimalDesign/<Name>/`. Der Verzeichnisname und alle darin enthaltene Dateien besitzen das Prefix der Methode. Z.B. wird bei der  $\mathcal{S}_1(\mathcal{T})$ -Diskretisierung das Kürzel ODP1 vorangestellt. Danach kommt die Klasse des Problems, wenn es sich um ein Verzeichnis handelt, oder der Funktionsname bei einer Datei. Für die  $\mathcal{S}_1(\mathcal{T})$ -Diskretisierung ergibt sich somit als Ordnername `ODP1-OptimalDesign`. Der Fehlerschätzer für diese Methode trägt dann den Namen `ODP1estimate_Jump`. Für den zielorientierten Ansatz wurde das Energiefunktional (9.4) implementiert. Für die jeweiligen Diskretisierungsmethoden wurden die folgende Fehlerschätzer implementiert

## Fehlerschätzer (estimator)

- $\mathcal{S}_1(\mathcal{T})$ : Average-Fehlerschätzer (`estimate_Avg`) nach [12],  
Residual-Fehlerschätzer (`estimate_Jump`) nach [11],
- $\mathcal{S}_2(\mathcal{T})$ : Residual-Fehlerschätzer (`estimate`) nach [11],
- $RT_0(\mathcal{T})$ - $\mathcal{P}_0(\mathcal{T})$ : Projektionsfehlerschätzer  $\eta_1$  (`estimate_Proj`), Sprungfehlerschätzer  $\eta_2$  (`estimate_Jump`) gemäß (8.12) und (8.13),
- $RT_0(\mathcal{T})$ - $\mathcal{P}_0(\mathcal{T})$  zielorientiert: zielorientierte Fehlerschätzer (`estimate`) gemäß (10.16).

```

% ODP1
% estimator = 'estimate_Avg';
% estimator = 'estimate_Jump';
% ODP2 / ODRTGOAL
% estimator = 'estimate';
% ODRT
% estimator = 'estimate_Proj';
% estimator = 'estimate_Jump';

```

Abbildung A.2: Definition der verschiedenen Fehlerschätzer. Auszug aus dem Startskript `startOptimalDesign.m`

Die Problemdefinitionen, welche wir in Abschnitt 9.1 beschrieben haben, befinden sich im Verzeichnis `./problems/optimalDesign/<Name>/`. Die Namen der Probleme sind die aufgeführten Variablennamen. In den Problemdefinitionen werden die Daten, wie z.B. die rechte Seite  $f$  oder die Dirichlet Daten  $u_D$  definiert. Wenn es eine analytische Funktion gibt, welche das Problem beschreibt, dann wird diese ebenfalls hier unter der Variablenbezeichnung `u_exact` definiert. Die durch `u_exact` abgeleiteten Funktionen wie  $\sigma$ ,  $\nabla u$ ,  $f$ ,  $u_D$  etc. werden als function-handle in den Problemdefinitionen definiert. Im Ordner `./problems/geometries/<Name>/` werden die benötigten Geometrien des Problems gespeichert.

### Problemdefinitionen (problem)

- das Modellbeispiel Quadrat (`OptimalDesign_Square`),
- das Modellbeispiel Lshape (`OptimalDesign_Lshape`),
- das Modellbeispiel Oktagon (`OptimalDesign_Octagon`),
- das Modellbeispiel geschlitztes Quadrat (`OptimalDesign_SquareSlit`),
- das numerische Testbeispiel zur Bestimmung der Regularisierungsparameter (`OptimalDesign_Square_exact`).

```

problem = 'OptimalDesign_Square';
% problem = 'OptimalDesign_Lshape';
% problem = 'OptimalDesign_Octagon';
% problem = 'OptimalDesign_SquareSlit';
% problem = 'OptimalDesign_Square_exact';

```

Abbildung A.3: Definition der numerischen Beispiele. Auszug aus dem Startskript `startOptimalDesign.m`

Die Realisierung der Markierungsstrategien und der Gitterverfeinerung, welche wir im Kapitel 8 präsentiert haben, befinden sich in den Ordnern `./algorithms/mark/<Name>` bzw. `./algorithms/refine/<Name>`. Die Funktion `bulk.m` realisiert die Markierung mit dem *bulk*-Kriterium. In der Datei `redGreenBlue.m` befindet sich die Implementierung der RGB-Verfeinerung. Details zu den Implementierungen findet man in [8].

### Markierungsstrategien (mark)

- uniforme Verfeinerung (`uniform`),
- Verfeinerung mittels dem *bulk*-Kriterium (`bulk`),
- Verfeinerung mittels dem Maximumkriterium (`max`).

```
% mark = 'uniform'
mark = 'bulk'
% mark = 'max'
```

Abbildung A.4: Definition der Markierungsstrategien. Auszug aus dem Startskript `startOptimalDesign.m`

Die nichtlinearen Löser, wobei wir `ODeps_h_dependence` in Abschnitt 9.2 und `ODoptimalLambda` in Abschnitt 9.1 vorgestellt haben, befinden sich im Verzeichnis `./PDEsolvers/OptimalDesign/ODsolvers/<Name>`. Der Dateiname ist hier wieder der aufgeführte Variablenname.

### Nichtlineare Löser (solver)

- Bestimmung des optimalen Lagrangemultiplikators  $\lambda$  (`ODoptimalLambda`)
- direkte Nullstellensuche für die konformen Methoden (`ODdirectFsolve`),
- Nullstellenberechnung der gemischten Formulierung mit einem gitterabhängenden Regularisierungsparameter  $\theta$  (`ODeps_h_dependence`),
- Nullstellenberechnung der gemischten Formulierung, wobei das Optimal Design Problem durch eine Folge von Regularisierungen ( $\theta_k$ ) approximiert wird (`ODlimitEps`).

```
% solver = 'ODeps_h_dependence';
% solver = 'ODlimitEps';
% solver = 'ODoptimalLambda';
solver = 'ODdirectFsolve';
```

Abbildung A.5: Definition der nichtlinearen Löser. Auszug aus dem Startskript `startOptimalDesign.m`

Die Parameter wie auch die Namen der Module und die function-handle, welche benutzt werden sollen, um ein Problem zu rechnen, werden in einer globalen Struktur `p` gespeichert. Die Struktur dient als Speicher für die definierte Konfiguration des Problems und alle dafür berechneten Größen, sodass jede Funktion, welche in einer Berechnung aufgerufen wird, Zugriff auf alle benötigten Daten besitzt. Der Aufbau der Struktur `p` entspricht einem Verzeichnisbaum. Z.B. werden in `p.params.<Name>` alle benötigten Parameter gespeichert, in `p.problem.<Name>` ist das zu lösende Problem definiert und in `p.level(k)` werden die Daten der Lösung und alle zusätzlichen Information bezogen auf die  $k$ -te Verfeinerung gespeichert. In jedem einzelnen „Ordner“ in `p` können auch einzelnen Unterordner erstellt werden. So wird z.B. in `p.level(k).geom` die  $k$ -te Triangulierung definiert.

## Parameter

- Materialparameter  $\mu_1, \mu_2, t_1$  und  $t_2$  (`mu_1, mu_2, t1, t2`)
- Verteilungsparameter  $\xi$  (`distribParam`)
- Lagrangmultiplikator  $\lambda$  (`lambda`)
- Regularisierungsparameter  $\alpha$  und  $\beta$  aus Algorithmus 3 (`epsFactor, epsPower`)
- Integrationsgenauigkeit der rechten Seite (Angabe des Polynomgrades, der exakt integriert werden soll) (`rhsIntegrateExactDegree`)
- Integrationsgenauigkeit des nichtlinearen Systems (Angabe des Polynomgrades, der exakt integriert werden soll) (`nonLinearExactIntegrateDegree`)
- Maximale Anzahl der Freiheitsgrade als Abbruchkriterium (`maxNrDoF`)

```

% set up parameters
maxNrDoF = 5000;

% material parameters mu_1 and mu_2
mu1 = 1;
mu2 = 2;

% Lagrange multiplier for the model problems
%Square
% lambda = 0.0084;
%L-shape
% lambda = 0.0143;
%Octagon
% lambda = 0.0284;
%Square-Slit
% lambda = 0.0168;
% Lagrange multiplier for numerical test example
% Square_Exact
lambda = 0.003101917488281;

t1 = (2*lambda*mu1/mu2).^(1/2);
t2 = mu2/mu1*t1;

p.problem.mu1 = mu1;
p.problem.mu2 = mu2;

p.problem.t1 = t1;
p.problem.t2 = t2;
p.problem.lambda = lambda;

p.params.epsPower = 1;
p.params.epsFactor = 1;
p.params.rhsIntegrateExactDegree = 19;
p.params.nonLinearExactIntegrateDegree = 5;

```

Abbildung A.6: Definition einiger Parameter. Auszug aus dem Startskript startOptimalDesign.m



## A.2 Ein Flußdiagramm

In diesem Abschnitt wird der Datenfluß in der Software beschrieben. Dieser ist der FFW-Software ähnlich. Wir orientieren uns hier an dem in [8] beschriebenen Diagramm.

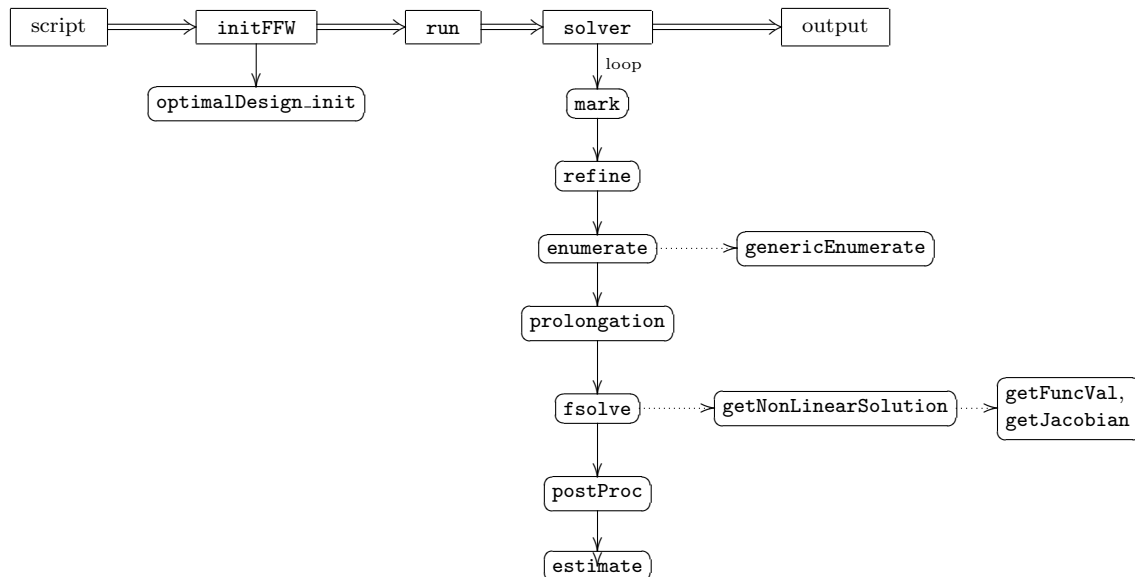


Abbildung A.7: Flußdiagramm der Optimal-Design-Software.

Funktionenname	Beschreibung
script	Kontroll-/Startskript, z.B. <code>startOptimalDesign</code> aus Abschnitt A.1
initFFW	Setzt Pfade, lädt Standardparameter, setzt die im Skript definierte problemspezifische Parameter und lädt Problemdefinitionen und deren Geometrien
run	Ruft <code>ODcomputesolution.m</code> und anschließend den in <code>script</code> spezifizierten Löser auf. Es erzeugt zudem methodenspezifische function-handle, wie z.B. die Basisfunktionen $\psi_k$ und die diskrete Funktion $u_h$ . Die function-handle der diskreten Funktionen werden in der Struktur <code>p</code> in <code>p.statics.&lt;Name&gt;</code> gespeichert.
optimalDesign_init	Es werden Optimal-Design-spezifische function-handle erzeugt, wie z.B. die diskrete Energie, die Massenverteilung und diverse Funktionen, die den Fehler zur exakten Lösung berechnen. Diese Funktionen stehen in <code>p.statics.&lt;Name&gt;</code> zur Verfügung.

<code>solver</code>	Der Löser des nichtlinearen Systems startet die FEM-Schleife und ruft die Funktionen <code>mark</code> , <code>refine</code> , <code>enumerate</code> , <code>prolong</code> , <code>fsolve</code> , <code>postProc</code> und <code>estimate</code> auf.
<code>mark</code>	Markiert Kanten und Dreiecke basierend auf einem gewählten Verfeinerungsindikator. Zusätzliche Kanten werden markiert, um die shape-regularity der Triangulierungen zu erhalten.
<code>refine</code>	Verfeinert die Liste der Kanten, die von <code>mark</code> bereitgestellt werden.
<code>enumerate</code>	Ruft <code>genericEnumerate</code> auf und berechnet methodenspezifische Datenstrukturen aus den Gitterinformationen. Z.B. werden jedem Element die globalen Freiheitsgrade zugeordnet.
<code>genericEnumerate</code>	Erzeugt nützliche Datenstrukturen, z.B. Normalen, Flächen, etc., aus der aktuellen Triangulierung [8].
<code>prolongation</code>	Setzt den Startwert für das erste Level und prolongiert die Lösung $x_{\ell-1}$ auf die Triangulierung $\mathcal{T}_\ell$ .
<code>fsolve</code>	Der MATLAB-Löser <code>fsolve.m</code> wird aufgerufen. Die Funktion <code>getNonlinearSolution</code> wird an <code>fsolve</code> übergeben.
<code>getNonlinearSolution</code>	Ruft die Funktionen <code>getFuncVal</code> und bei Bedarf auch die Funktion <code>getJacobian</code> auf.
<code>getFuncVal</code>	Stellt die Funktion (das Residuum) bereit, von der die Nullstelle gefunden werden soll.
<code>getJacobian</code>	Stellt die Jacobimatrix der Funktion bereit.
<code>postProc</code>	Aufbereitung der berechneten Lösung. Z.B. werden die Koeffizienten der diskreten Lösung und auftretende Lagrangemultiplikatoren aus dem Lösungsvektor voneinander getrennt.
<code>estimate</code>	Ausgehend von der diskreten Lösung wird deren Fehler durch einen vom Nutzer definierten Fehlerschätzer berechnet.
<code>output</code>	Rückgabe der Datenstruktur <code>p</code> . Anschließend kann durch den Benutzer eine Datenauswertung durchgeführt werden.

## A.3 Ein Beispiel

In diesem Abschnitt wollen wir eine Beispielimplementation zeigen, die veranschaulicht, wie jede Funktion aufgebaut ist. Da der Umfang der Software sehr gross ist, können wir nicht das ganze Programm abdrucken. Das vollständige Programm befindet sich auf der anliegenden CD. Hier beschränken wir uns auf die Funktion `ODRTgetFuncVal.m`, die die Funktion bereitstellt, von der die Nullstelle mittels `fsolve.m` berechnet wird. Die Funktion bekommt von `ODRTgetNonlinearSolution.m` die Struktur `p` übergeben. Zunächst werden alle benötigten Informationen aus der Struktur, z.B der aktuelle Wert  $x$ , extrahiert und als Variable gespeichert. Anschliessend wird der Funktionswert  $E(x)$  berechnet. Die Funktion ist über (5.7) definiert. Eine Schleife über die Elemente der Triangulierung berechnet den lokalen Funktionswert. Mittels der MATLAB-Funktion `accumarray` wird dann der globale Funktionswert berechnet. Abschließend werden noch die rechte Seite  $f$  und die Dirichlet-Daten im Funktionswert integriert. Der Funktionswert wird dann in der Struktur `p` gespeichert.

```
function p = ODRTgetFuncVal(p)
% author: David Guenther
%% INPUT
n4e = p.level(end).geom.n4e;
length4ed = p.level(end).enum.length4ed;
ed4e = p.level(end).enum.ed4e;
e4ed = p.level(end).enum.e4ed;
n4ed = p.level(end).enum.n4ed;
DbEd = p.level(end).enum.DbEd;
dofU4e = p.level(end).enum.dofU4e;
dofSigma4e = p.level(end).enum.dofSigma4e;
nrElems = p.level(end).nrElems;
nrEdges = p.level(end).nrEdges;
x = p.level(end).x0;
lvl = size(p.level,2);
degree = loadField('p.params','nonLinearExactIntegrateDegree',p,5);

%% compute the function value E(x)
BT = zeros(1,3,nrElems);
CT = zeros(1,3,nrElems);
DT = zeros(1,1,nrElems);

for curElem = 1:nrElems
    curEdges = ed4e(curElem,:);
    curEdgeLengths = length4ed(curEdges);
    curX4ed = x(curEdges);
    curX4e = x(nrEdges + curElem);

    signum = ones(1,3);
    signum(e4ed(curEdges,2) == curElem) = -1;
    %divergence of basis functions on T,e.g. div(q) = sign*|E|/|T|
    div_qh = signum.*curEdgeLengths';

    %calc int_T DW*(p_h)*q_h
    DW = integrate(n4e(curElem,:),lvl,degree,@integrand,p);

    BT(:, :, curElem) = DW;
    %calc int_T u_h*div(q_h)
    CT(:, :, curElem) = curX4e*div_qh;
```

```

        %calc int_T v_h*div(p_h)
        DT(:, :, curElem) = curX4ed'*div_qh';
    end

    I = dofSigma4e';
    B = accumarray(I(:), BT(:));
    C = accumarray(I(:), CT(:));
    D = accumarray(dofU4e(:), DT(:));

    boundary = zeros(nrEdges + nrElems, 1);
    boundary(DbEd) = integrate(n4ed(DbEd, :), lvl, degree, @integrandBoundary, p);

    b = -boundary;
    f4e = p.level(end).f4e;
    b(nrEdges+1:end) = f4e;

    funcVal = [B + C;
               D] + b;
%% OUTPUT
p.level(end).funcVal = funcVal;

```

Abbildung A.8: Auszug aus der Implementierung des Funktionswertes (5.7) für die gemischte Formulierung des Optimal-Design-Problems.

Um die lokalen Funktionswerte zu berechnen, muss der Ausdruck (5.7) integriert werden. Die Integration wird mittels Gausspunkten durchgeführt. Die Genauigkeit wird durch den Parameter `nonLinearExactIntegrateDegree` vorgegeben. Details hierzu findet man in [8]. Die Integranden werden als Subfunktionen bereitgestellt und werden dem Integrator `integrate.m` übergeben. In der in Abbildung A.8 gezeigten Funktion, wird der Ausdruck  $\int_T D\varphi^*(\sigma_{\theta,h}) \cdot \psi_k dx$  berechnet. Die MATLAB-Realisierung des Integranden hierfür wird in Abbildung A.9 gezeigt.

```

%% supply integrand DW^*(p_h)*q_h
function val = integrand(x,y,curElem,lvl,p)
% W'(|X|)/|X|*X*Y

grad_h = p.statics.grad_h;
stressBasis = p.statics.stressBasis;

evalGrad = grad_h(x,y,curElem,lvl,p);
evalGrad = reshape(evalGrad', [2 1 length(x)]);

evalBasis = stressBasis(x,y,curElem,lvl,p);

val = matMul(evalBasis, evalGrad);

```

Abbildung A.9: Implementierung des Integranden  $D\varphi^*(\sigma_{\theta,h}) \cdot \psi_k$  gegeben in `ODRTgetFuncVal.m`.

Die function-handle `stressbasis` und `grad_h` werden in der Datei `ODRTrun.m` definiert. Der Aufbau des Programms mit function-handle erlaubt eine einfache und fehlerfreie Auswertung der Funktionen. Als abschließendes Beispiel soll noch gezeigt werden, wie die

Funktion `grad_h` konstruiert wird. In `ODRTrun.m` wird die Funktion `ODRTgetPhFuncs.m` aufgerufen. Dort werden die zu  $\sigma_{\theta,h}$  gehörenden function-handle bereitgestellt.

```
p.statics.grad_h = @getGradU_h;

%% supply \grad u_{\epsilon,h}
function grad_h = getGradU_h(x,y,curElem,lv1,p)

sigma_h = p.statics.sigma_h;
conjNonLinearFuncDer = p.problem.conjNonLinearFuncDer;

evalSigmah = sigma_h(x,y,curElem,lv1,p);
absSigmah = (evalSigmah(:,1).^2 + evalSigmah(:,2).^2).^(1/2);

% \grad u = DW^*\epsilon(p_h) = W^*\epsilon(|p_h|)/|p_h|*p_h
evalConjNonLinear = conjNonLinearFuncDer(absSigmah,curElem,lv1,p);

grad_h = (evalConjNonLinear*[1,1]).*evalSigmah;
```

Abbildung A.10: Konstruktion des function-handle  $D\varphi^*(\sigma_{\theta,h})$ . Auszug aus `ODRTgetPhFuncs.m`.

## A.4 Auswertung der Lösung

Um die berechnete diskrete Lösung  $(\sigma_{\theta,h}, u_h)$  auszuwerten, stehen function-handle in `p.statics.sigmah` und in `p.statics.uh` zur Verfügung. Für den Fall, dass die analytische Spannung  $\sigma$  vorliegt, läßt sich der Spannungsfehler  $\|\sigma - \sigma_{\theta,h}\|_{L^2(\mathcal{T})}$  mittels den Funktionen `show.m` und `drawError.m` darstellen. Der Befehl

```
p = show('drawError_L2errorPhminusP0',p);
```

stellt den Spannungsfehler in einem loglog-Plot dar. Der Fehler wird in Abhängigkeit der Anzahl der Freiheitsgraden in einer logarithmischen Skalierung der Achsen dargestellt. Auf ähnlicheweise kann man sich beliebige Fehlernormen anzeigen lassen. In der Datei `optimalDesign_init.m` wird der Fehlerintegrand definiert und gespeichert und mit dem Befehl `show('drawError_<Name>',p)` berechnet und dargestellt. Der Name ist durch den Funktionsnamen des Fehlerintegranden gegeben. Falls keine exakte Lösung vorliegt lässt sich mit `showAitkenError.m` im Verzeichnis `./PDEsolvers/optimalDesign/ODmisc` der Limes einer Folge extrapolieren.

Ebenfalls in der Datei `optimalDesign_init.m` ist auch die Massenverteilung definiert. Durch die Funktion `./PDEsolvers/optimalDesign/ODshowVolumeFraction.m` lässt sich die Massenverteilung anzeigen. Der Syntax ist gegeben durch

```
ODshowVolumeFraction(p,<lv1>);
```

wobei `<lv1>` das jeweilige Verfeinerungslevel repräsentiert.

## A.5 Der Verzeichnissbaum

```

.
|-- PDEsolvers
|
|   '-- OptimalDesign
|
|       |-- ODP1-OptimalDesign
|
|           |   |-- ODP1enumerate.m
|
|           |   |-- ODP1estimate.Avg.m
|
|           |   |-- ODP1estimate.Jump.m
|
|           |   |-- ODP1getFuncVal.m
|
|           |   |-- ODP1getJacobian.m
|
|           |   |-- ODP1getNL4fminunc.m
|
|           |   |-- ODP1getNonLinearSolution.m
|
|           |   |-- ODP1getPhFuncs.m
|
|           |   |-- ODP1getUhFuncs.m
|
|           |   |-- ODP1postProc.m
|
|           |   |-- ODP1prolong.m
|
|           |   |-- ODP1run.m
|
|           |   '-- scripts
|
|           |       |-- ODP1drawGradU.m
|
|           |       '-- ODP1drawU.m
|
|       |-- ODP2-OptimalDesign
|
|           |   |-- ODP2enumerate.m
|
|           |   |-- ODP2estimate.m
|
|           |   |-- ODP2getFuncVal.m
|
|           |   |-- ODP2getJacobian.m
|
|           |   |-- ODP2getNL4fminunc.m
|
|           |   |-- ODP2getNonLinearSolution.m
|
|           |   |-- ODP2getPhFuncs.m
|
|           |   |-- ODP2getUhFuncs.m
|
|           |   |-- ODP2postProc.m
|
|           |   |-- ODP2prolong.m
|
|           |   |-- ODP2run.m
|
|           |   '-- scripts
|
|           |       |-- ODP2drawGradU.m
|
|           |       '-- ODP2drawU.m
|
|       |-- ODR1-OptimalDesign
|
|           |   |-- ODR1enumerate.m
|
|           |   |-- ODR1estimate.Jump.m
|
|           |   |-- ODR1estimate.Proj.m
|
|           |   |-- ODR1getFuncVal.m
|
|           |   |-- ODR1getJacobian.m

```

```

|         | |   |-- ODRTgetNL4fminunc.m
|         | |   |-- ODRTgetNonLinearSolution.m
|         | |   |-- ODRTgetPhFuncs.m
|         | |   |-- ODRTgetUhFuncs.m
|         | |   |-- ODRTpostProc.m
|         | |   |-- ODRTprolong.m
|         | |   |-- ODRTrun.m
|         | |   '-- scripts
|         |         '-- ODRTdrawU.m
|
|-- ODRTGOAL-OptimalDesign
|
|         |-- ODRTGOALenumerate.m
|         |-- ODRTGOALestimate.m
|         |-- ODRTGOALgetFuncVal.m
|         |-- ODRTGOALgetJacobian.m
|         |-- ODRTGOALgetLambda1Funcs.m
|         |-- ODRTGOALgetLambda2Funcs.m
|         |-- ODRTGOALgetNonLinearSolution.m
|         |-- ODRTGOALgetPhFuncs.m
|         |-- ODRTGOALgetUhFuncs.m
|         |-- ODRTGOALintegrateBoundary.m
|         |-- ODRTGOALpostProc.m
|         |-- ODRTGOALprolong.m
|         |-- ODRTGOALrun.m
|         '-- scripts
|
|         |-- ODRTGOALdrawGradU.m
|         '-- ODRTGOALdrawU.m
|
|-- ODcomputeSolution.m
|
|-- ODgenericNonLinear.m
|
|-- ODgetNonLinearRegularConj.m
|
|-- ODgetRegularConj.m
|
|-- ODshowVolumeFraction.m
|
|-- ODSolver
|
|         |-- ODDirectFminunc.m
|         |-- ODDirectFsolve.m
|         |-- ODDirectLineSearch.m
|         |-- ODEps_h.dependence.m
|         |-- ODlimitEps.m
|         |-- ODlimitEps4Fminunc.m
|         |-- ODlineSearch.m
|         '-- ODOptimalLambda.m
|
|-- misc
|
|         |-- degeneratedT.m

```

```

|         | |-- determOptimalParam.m
|         | |-- drawTriangle.m
|         | |-- evalExamples.m
|         | |-- latex.m
|         | |-- plotDiplomFigure.m
|         | |-- plotForce.m
|         | |-- plotGeom.m
|         | |-- rgb.m
|         | '-- showAitkenErrorEnergy.m
|
|         '-- optimalDesign_init.m
|-- algorithms
|
|     |-- enum
|
|         |-- getAngle4n.m
|
|         |-- getAngles4e.m
|
|         |-- getArea4e.m
|
|         |-- getArea4n.m
|
|         |-- getDbEdges.m
|
|         |-- getE4ed.m
|
|         |-- getE4n.m
|
|         |-- getEd4e.m
|
|         |-- getEd4n.m
|
|         |-- getLength4ed.m
|
|         |-- getMidPoint4e.m
|
|         |-- getMidPoint4ed.m
|
|         |-- getN4ed.m
|
|         |-- getNbEdges.m
|
|         |-- getNormals4DbEd.m
|
|         |-- getNormals4NbEd.m
|
|         |-- getNormals4e.m
|
|         |-- getNormals4ed.m
|
|         '-- getTangents4e.m
|
|     |-- integrate
|
|         |-- getCompositeGaussPoints.m
|
|         |-- getConProdGaussPoints.m
|
|         |-- getGaussPoints.m
|
|         '-- integrate.m
|
|     |-- mark
|
|         |-- bulk.m
|
|         |-- maximum.m
|
|         '-- uniform.m
|
|     |-- misc
|
|         |-- closure.m

```



```

| | | -- genericEnumerate.m
| | | -- localDoFtoGlobalDoF.m
| | -- refine
| | -- redGreenBlue.m
|-- helpers
| | -- Matlab4Maple.m
| | -- loadField.m
| | -- matMul.m
| | -- rowaddr.m
|-- initFFW.m
|-- problems
| | -- geometries
| | | -- Lshape
| | | | -- Lshape.Db.dat
| | | | -- Lshape.c4n.dat
| | | | -- Lshape.n4e.dat
| | | -- Octagon
| | | | -- Octagon.Db.dat
| | | | -- Octagon.c4n.dat
| | | | -- Octagon.n4e.dat
| | | -- Square
| | | | -- Square.Db.dat
| | | | -- Square.c4n.dat
| | | | -- Square.n4e.dat
| | | -- SquareSlit
| | | | -- SquareSlit.Db.dat
| | | | -- SquareSlit.c4n.dat
| | | | -- SquareSlit.n4e.dat
| | | -- Triangle
| | | | -- Triangle.Db.dat
| | | | -- Triangle.c4n.dat
| | | | -- Triangle.n4e.dat
| | -- optimalDesign
| | | -- OptimalDesign.Lshape.m
| | | -- OptimalDesign.Lshape.exact.m
| | | -- OptimalDesign.Octagon.m
| | | -- OptimalDesign.Square.m
| | | -- OptimalDesign.SquareSlit.m
| | | -- OptimalDesign.Square.exact.m
|-- results
-- startOptimalDesign.m

```

# Literaturverzeichnis

- [1] C. Bahriawati and C. Carstensen. Three matlab implementations of the lowest-order Raviart-Thomas MFEM with a posteriori error control. *Computational Methods in Applied Mathematics*, 5:333–361, 2005. 27
- [2] W. Bangerth and R. Rannacher. *Adaptive Finite Element Methods for Differential Equations*. Birkhauser Verlag, 2003. 65, 66
- [3] V. Barbu and T. Precupanu. *Convexity and Optimization in Banach Spaces*. Kluwer, 1986. 15, 18, 19
- [4] S. Bartels and C. Carstensen. A convergent adaptive finite element method for an optimal design problem. Preprint, 2006. 6, 10, 12, 43, 44, 61
- [5] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999. 39
- [6] D. Braess. *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, Cambridge, 2001. Second Edition. 32
- [7] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer, 1991. 30
- [8] A. Byfut, J. Gedicke, D. Günther, H. Mellmann, J. Reininghaus, and S. Wiedemann. Documentation FFW - in preparation. 36, 75, 78, 81, 82, 84
- [9] C. Carstensen. Five lectures on adaptive mixed finite element methods. 27, 32
- [10] C. Carstensen. Nonconvex energy minimization and relaxation in computational material science. *IUTAM Symposium on Computational Mechanics of Solid Materials at Large Strains*. Kluwer Academic Publishers, The Netherlands, pages 3–20, 2003. 13
- [11] C. Carstensen. A unifying theory of a posteriori finite element error control. *Numer. Math.*, 100(4):617–637, 2005. 76
- [12] C. Carstensen and S. Bartels. Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. I. Low order conforming, nonconforming, and mixed FEM. *Math. Comp.*, 71(239):945–969 (electronic), 2002. 76

- [13] C. Carstensen and S. Müller. Local stress regularity in scalar nonconvex variational problems. *SIAM J. Math. Anal.*, 34(2):495–509 (electronic), 2002. 6, 13
- [14] C. Carstensen and A. Prohl. Numerical analysis of relaxed micromagnetics by penalised finite elements. *Numer. Math.*, 90(1):65–99, 2001. 42
- [15] A. Cherkaev. *Variational Methods for Structural Optimization*. Springer, 2000.
- [16] P. Clément. Approximations by finite element functions using local regularization. *Sér. Rouge Anal.*, 2:77–84, 1975. 32
- [17] J. Dennis. Nonlinear least squares. *State of the Art in Numerical Analysis*, pages 269–312. 39
- [18] L. C. Evans. *Partial Differential Equations*. American Mathematical Society, Providence, 1998. 8
- [19] V. Girault and P. Raviart. Finite element methods for navier-stokes equations. theory and algorithms. *Computational Mechanics*, 5, 1986. 32
- [20] B. Kawohl, J. Stara, and G. Wittum. Analysis and numerical studies of a problem of shape design. *Arch. Rational Mech. Anal.*, 114:349–363, 1991. 6, 13
- [21] R. Kohn and G. Strang. Optimal design and relaxation of variational problems i. *Communications on Pure and Applied Mathematics*, XXXIX:113–137, 1986. 5, 10
- [22] R. Kohn and G. Strang. Optimal design and relaxation of variational problems ii. *Communications on Pure and Applied Mathematics*, XXXIX:139–182, 1986. 5, 10
- [23] R. Kohn and G. Strang. Optimal design and relaxation of variational problems iii. *Communications on Pure and Applied Mathematics*, XXXIX:353 – 377, 1986. 5, 10
- [24] E. Zeidler. *Nonlinear Functional Analysis and its Applications II/B*. Springer, 1985.
- [25] E. Zeidler. *Nonlinear Functional Analysis and its Applications III*. Springer, 1985.

### **Selbstständigkeitserklärung**

Ich erkläre, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 24. September 2007

David Günther

### **Einverständniserklärung**

Hiermit erkläre ich mich einverstanden, dass ein Exemplar meiner Diplomarbeit in der Bibliothek des Institutes für Mathematik der Humboldt Universität zu Berlin verbleibt.

Berlin, den 24. September 2007

David Günther