Leibniz Universität Hannover
Fakultät für Mathematik und Physik
Institut für Angewandte Mathematik

Diplomarbeit

Ein hierarchischer Fehlerschätzer für Hindernisprobleme

von Cornelius Rüther
Matr.-Nr.: 2517350

22. September 2014

Erstprüfer: Prof. Dr. Gerhard Starke
Zweitprüfer: Prof. Dr. Peter Wriggers

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Kapitel 1

Einleitung

- Thema (worum geht es?) → Fehlerabschätzung → analytische Lösung oftmals nicht bekannt und damit Fehlerschätzer interessant

→ in FEM soll Lösung genauer mit weniger Rechenzeit sein, daraus folgt Anwendung adaptiver Verfahren mit verschiedenen Fehlerschätzern

- Lücke zum neuen (Kontaktproblematik) füllen in dieser Arbeit

→ Übertragung unseres Fehlerschätzers auf Kontaktprobleme, wie und warum?! → möglicher Grund: Hindernisprobleme beinhalten Kontaktbereiche (später für Kapitel 4 interessant)

- Struktur der Arbeit

Kapitel 2

Grundlagen

In diesem Kapitel wollen wir uns mit grundlegender Theorie beschäftigen, die nicht im Anhang aufgeführt ist, zum Verständnis von den darauffolgenden Kapiteln jedoch notwendig ist.

2.1 Variationsformulierung

Stichpunkte für die Formulierung:

- Betrachte als Modellproblem Ausrenkung $u : \Omega \to \mathbb{R}$ einer in $\Omega \subset \mathbb{R}^d$ eingespannten Membran unter Kraft f
- mathematisch beschrieben wird dies durch das *Dirichlet-Problem*

$$\begin{aligned} -\Delta u &= f \text{ in } \Omega\,, \\ u &= g \text{ auf } \partial\Omega\,, \end{aligned} \tag{2.1}$$

- in der Praxis $d = 2, 3$ übliche Dimensionen
- Richtiger Punkt:

  Notation. der Einfachheit halber sei im Folgenden $d = 2$ und $\Omega \subset \mathbb{R}^2$ ein durch ein Polygonzug berandetes Gebiet, den Rand $\partial\Omega$ bezeichnen wir mit Γ.
- allgemeiner berandete Gebiete können durch polygonale beliebig genau approximiert werden
- Transformation: Sei $u_0 : \Omega \to \mathbb{R}$ eine zulässige Funktion, d.h. deren Regularität für (2.1) ausreichend ist, und für die $u_0 = g$ auf Γ gilt. Dann gilt für $\tilde{u} = u - u_0$

$$\begin{aligned} -\Delta \tilde{u} &= \tilde{f} \text{ in } \Omega\,, \\ \tilde{u} &= 0 \text{ auf } \Gamma \end{aligned} \tag{2.2}$$

  mit $\tilde{f} = f - \Delta u_0$.

- $\Rightarrow$ wir beschränken uns auf das *homogene Dirichlet-Problem* (2.2), d.h. sei $g \equiv 0$ in (2.1)
- Sei im Folgenden $H_0^1(\Omega)$ wie in Bemerkung A.8 der Raum der schwach differenzierbaren Funktionen, die am Rand Γ verschwinden im Sinne der Spur.
- für $v \in H_0^1(\Omega)$ gilt dann mit (2.1)

$$\int_\Omega -\Delta u \cdot v \, dx = \int_\Omega f v \, dx \, .$$

Betrachte also (2.1) im Mittel über das ganze Gebiet Ω. Durch Anwenden der 1. Green'schen Formel (bzw. Satz von Gauß) ergibt sich

$$\int_\Omega \nabla u \cdot \nabla v \, dx - \underbrace{\int_\Gamma v \partial_\nu u \, ds}_{=0, \text{ da } v|_\Gamma = 0} = \int_\Omega f v \, dx$$

$$\iff \qquad \int_\Omega \nabla u \cdot \nabla v \, dx = \int_\Omega f v \, dx \tag{2.3}$$

- kurz geschrieben ist (2.3) mit der Notation aus Satz A.5 (b)

$$(\nabla u, \nabla v)_0 = (f, v)_0 \, .$$

- wir definieren die Bilinearform $a : (H_0^1(\Omega))^2 \to \mathbb{R}, a(u, v) := (\nabla u, \nabla v)_0$ und $(f, v) := (f, v)_0$.

Definition 2.1. Eine Funktion $u \in H_0^1(\Omega)$ heißt *schwache Lösung* vom homogenen Dirichlet-Problem

$$\begin{aligned} -\Delta u &= f \text{ in } \Omega \, , \\ u &= 0 \text{ auf } \Gamma \, , \end{aligned} \tag{DP}$$

wenn die Gleichung

$$a(u, v) = (f, v) \quad \forall \, v \in H_0^1(\Omega) \tag{2.4}$$

gilt.

- Wir betrachten im folgenden alle Hilberträume über $\mathbb{R}$.
- Frage nach der Existenz und Eindeutigkeit einer schwachen Lösung für (DP) $\Rightarrow$ hierfür wird ein Hilbertraum benötigt (nachher im Beweis ersichtlich) $\rightarrow$ Lösung liefert der Satz von Lax-Milgram.
- zuvor noch eine Definition.

Definition 2.2. Sei H ein Hilbertraum. Die Bilinearform $a : H \times H \to \mathbb{R}$ heißt *stetig*, falls mit einem $c > 0$

$$|a(u,v)| \leq c \, \|u\|_H \|v\|_H \quad \forall \, u, v \in H$$

gilt. Sie heißt *H-elliptisch* (oder kurz *elliptisch* oder *koerziv*), falls es ein $\alpha > 0$ gibt, so dass

$$a(v,v) \geq \alpha \, \|v\|_H^2 \quad \forall \, v \in H$$

gilt.

- Bevor Existenz der Lösung gezeigt, betrachte Funktional $J(v) = \frac{1}{2}a(v,v) - F(v)$ genauer

-

Lemma 2.3. *Es sei H ein Hilbertraum. Das Funktional*

$$J : H \to \mathbb{R}\,, \quad J(v) := \frac{1}{2}a(v,v) - F(v)\,,$$

wobei $a : H \times H \to \mathbb{R}$ eine stetige bilineare koerzive und $F : H \to \mathbb{R}$ eine lineare Abbildung ist, ist konvex.

Beweis. Es seien $u, v \in H$, dann gilt $u + t(v-u) = (1-t)u + tv \in H$ (dies gilt auch, wenn wir den Satz auf eine konvexe Teilmenge $M \subset H$ beschränken). Damit folgt mit $t \in [0,1]$

$$\begin{aligned}
J((1-t)u + tv) =& \frac{1}{2}a((1-t)u + tv, (1-t)u + tv) - F((1-t)u + tv) \\
=& (1-t)\, J(u) + t\, J(v) + \frac{1}{2}a((1-t)u + tv, (1-t)u + tv) \\
& - \frac{1}{2}(1-t)\, a(u,u) - \frac{1}{2}t\, a(v,v) \\
=& (1-t)\, J(u) + t\, J(v) + \frac{1}{2}a(u,u) + t\, a(u, v-u) \\
& + \frac{t^2}{2}a(v-u, v-u) - \frac{1}{2}(1-t)\, a(u,u) - \frac{1}{2}t\, a(v,v) \\
=& (1-t)\, J(u) + t\, J(v) + \frac{t^2}{2}a(v-u, v-u) \\
& \underbrace{+t\, a(u,v) - \frac{1}{2}t\, a(u,u) - \frac{1}{2}t\, a(v,v)}_{= -\frac{1}{2}t\, a(v-u, v-u)} \\
=& (1-t)\, J(u) + t\, J(v) - \frac{1}{2}\underbrace{t\,(1-t)}_{\geq 0}\, \underbrace{a(v-u, v-u)}_{\geq \alpha \|v-u\|_H^2 \geq 0} \\
\leq& (1-t)\, J(u) + t\, J(v)\,.
\end{aligned}$$

Daraus folgt die Behauptung. □

- **Lemma 2.4.** *Sei H ein Hilbertraum. Das Funktional $J : H \to \mathbb{R}, J(v) = \frac{1}{2}a(v,v) - F(v)$ aus Lemma 2.3 ist Gâteaux-differenzierbar (s. Definition A.9).*

  Beweis. Wir rechnen einfach nach, dass der Grenzwert des Differenzenquotienten existiert und verwenden dabei die Bilinearität von a und Linearität von F. Seien $u, v \in H$, dann gilt

$$\begin{aligned}
\mathcal{D}_v J(u) &= \lim_{t\to 0} \frac{J(u+tv) - J(u)}{t} \\
&= \lim_{t\to 0} \frac{J(u) + t\,(a(u,v) - F(v)) + \frac{t^2}{2}a(v,v) - J(u)}{t} \\
&= \lim_{t\to 0} (a(u,v) - F(v)) + \frac{t}{2}a(v,v) \\
&= a(u,v) - F(v) < \infty\,,
\end{aligned}$$

  da a und F jeweils stetig sind und daher durch $\|u\|_H, \|v\|_H$ beschränkt sind. Damit folgt die Behauptung. □

- **Theorem 2.5.** (Lax-Milgram) *Es sei H ein Hilbertraum und $a : H \times H \to \mathbb{R}$ eine symmetrische, in H stetige, koerzive Bilinearform. Weiter sei $F : H \to \mathbb{R}$ ein stetiges lineares Funktional, d.h.*

$$|F(v)| \le c\,\|v\|_H \quad \forall\, v \in H$$

  mit einer Konstante $c > 0$. Dann gibt es eine eindeutige Lösung $u \in H$, für die

$$a(u,v) = F(v) \quad \forall\, v \in H\,.$$

  gilt. Diese minimiert den Ausdruck

$$J(v) = \frac{1}{2}a(v,v) - F(v)$$

  unter allen $v \in H$.

  Beweis. (i) Zunächst zeigen wir die Äquivalenz der beiden oberen Probleme.

„⇒“ Es sei $u \in H$, so dass $a(u,v) = F(v) \,\forall\, v \in H$. Für $t > 0$ und $v \in H$ gilt dann

$$\begin{aligned} J(u+tv) &= \frac{1}{2}a(u+tv, u+tv) - F(u+tv) \\ &= \frac{1}{2}a(u,u) + t\,a(u,v) + \frac{t^2}{2}a(v,v) - F(u) - t\,F(v) \\ &= \frac{1}{2}a(u,u) - F(u) + t\,\underbrace{(a(u,v) - F(v))}_{=0} + \frac{t^2}{2}\,\underbrace{a(v,v)}_{\substack{\geq 0,\text{ da } a \\ \text{koerziv}}} \\ &> \frac{1}{2}a(u,u) - F(u) = J(u)\,, \end{aligned}$$

also ist $u = \arg\min\limits_{v \in H} J(v)$.

„⇐“ Es sei $u \in H$ das Minimum von dem Problem

$$\min_{v \in H} J(v) = \frac{1}{2}a(v,v) - F(v)\,.$$

Da $J : H \to \mathbb{R}$ nach Lemma 2.3 ein konvexes Funktional ist und J nach Lemma 2.4 Gâteaux-differenzierbar, gilt mit Satz A.10 für alle $v \in H$

$$\begin{aligned} 0 = \mathcal{D}_v J(u) &= \frac{d}{dt}J(u+tv)\Big|_{t=0} \\ &= \frac{d}{dt}(J(u) + t\,(a(u,v) - F(v)) + \frac{t^2}{2}a(v,v))\Big|_{t=0} \\ &= a(u,v) - F(v) + t\,a(v,v)\Big|_{t=0} = a(u,v) - F(v) \end{aligned}$$

(ii) Eindeutigkeit: Es seien $u, \tilde{u} \in H$ Lösungen der Variationsungleichung, d.h.

$$a(u,v) = F(v) \,\wedge\, a(\tilde{u},v) = F(v) \quad \forall\, v \in H\,.$$

Damit folgt durch Subtraktion der beiden Gleichungen für alle $v \in H$

$$a(u,v) = a(\tilde{u},v) \Longleftrightarrow a(u-\tilde{u},v) = 0\,. \tag{2.5}$$

Da H ein Vektorraum ist, gilt auch $u - \tilde{u} \in H$. Ersetzen wir also in (2.5) $v = u - \tilde{u}$, dann ergibt sich

$$0 = a(u-\tilde{u}, u-\tilde{u}) \overset{a \text{ koerziv}}{\geq} \underbrace{\alpha}_{>0}\|u-\tilde{u}\|_H^2 \geq 0 \Longrightarrow \|u-\tilde{u}\|_H^2 = 0\,,$$

also folgt $u = \tilde{u}$.

(iii) Existenz: Die Existenz einer Lösung weisen wir über das Funktional nach.

$$\begin{aligned} J(v) &= \frac{1}{2}a(v,v) - F(v) \overset{\substack{a \text{ koerziv} \\ F \text{ linear}}}{\geq} \frac{1}{2}\alpha\|v\|_H^2 - c\|v\|_H \\ &= \frac{1}{2}\alpha\left(\|v\|_H^2 - \frac{2c}{\alpha}\|v\|_H\right) = \frac{1}{2}\alpha\left(\|v\|_H - \frac{c}{\alpha}\right)^2 - \frac{c^2}{2\alpha} \\ &\geq -\frac{c^2}{2\alpha} \end{aligned}$$

Folglich ist J nach unten beschränkt. Sei $\eta := \inf\{J(v) \mid v \in H\}$ und $(v_n)_{n\in\mathbb{N}}$ eine Folge mit $J(v_n) \to \eta$ für $n \to \infty$. Dann folgt mit der Koerzivität von a

$$\begin{aligned} \alpha\|v_n - v_m\|_H^2 \leq& a(v_n - v_m, v_n - v_m) \\ =& a(v_n, v_n) + a(v_m, v_m) - a(v_n, v_m) - a(v_m, v_n) \\ =& 2a(v_n, v_n) + 2a(v_m, v_m) \underbrace{-a(v_n, v_n + v_m) - a(v_m, v_n + v_m)}_{=-a(v_n+v_m, v_n+v_m)} \\ =& 2a(v_n, v_n) - 4F(v_n) + 2a(v_m, v_m) - 4F(v_m) \\ & - a(v_n + v_m, v_n + v_m) + 4F(v_n + v_m) \\ =& 4J(v_n) + 4J(v_m) - 4a\left(\frac{v_n + v_m}{2}, \frac{v_n + v_m}{2}\right) + 8F\left(\frac{v_n + v_m}{2}\right) \\ =& 4J(v_n) + 4J(v_m) - 8J\left(\frac{v_n + v_m}{2}\right) \\ \leq& 4J(v_n) + 4J(v_m) - 8\eta \xrightarrow[n,m\to\infty]{} 4\eta + 4\eta - 8\eta = 0\,, \end{aligned}$$

d.h. $(v_n)_{n\in\mathbb{N}}$ ist eine Cauchy-Folge. Da H ein Hilbertraum ist, gilt somit: $\exists\, u \in H : v_n \xrightarrow[n\to\infty]{} u$ mit $J(u) = \eta$. □

•

Satz 2.6. (Poincaré-Friedrich-Ungleichung) *Es sei Ω in einem d-dimensionalen Würfel der Kantenlänge $s > 0$ enthalten. Dann gilt*

$$\|v\|_0 \leq s\|\nabla v\|_0 \quad \forall\, v \in H_0^1(\Omega)\,,$$

wobei $\|\cdot\|_0$ die durch das Skalarprodukt $(\cdot,\cdot)_0$ induzierte Norm ist.

Beweis. Der Beweis ist in [Bra13] Kapitel II, §1 Sobolev-Räume, Satz 1.5 oder [Sta08] Satz 1.5 zu finden. □

- Greifen wieder die Frage auf, ob das Problem (2.4) mit $a : (H_0^1(\Omega))^2 \to \mathbb{R}, a(u,v) = (\nabla u, \nabla v)_0$ und $F : H_0^1(\Omega) \to \mathbb{R}, F(v) := (f, v)$ eine eindeutige Lösung hat.

- Kann nun mit Theorem 2.5 beantwortet werden. Es seien $u, v \in H_0^1(\Omega)$, dann gilt

$$\begin{aligned} a(v,v) &= \int_\Omega \nabla v \nabla v \, dx = \|\nabla v\|_0^2 \\ &\geq \frac{s^2+1}{(1+s)^2}\|\nabla v\|_0^2 \overset{\text{Satz 2.6}}{\geq} \frac{1}{(1+s)^2}(\|v\|_0^2 + \|\nabla v\|_0^2) \\ &= \frac{1}{(1+s)^2}\|v\|_1^2 \,. \end{aligned}$$

  Damit ist a mit $\alpha := \frac{1}{(1+s)^2}$ koerziv. Weiter rechnen wir nach:

$$\begin{aligned} |a(u,v)| &= \left| \int_\Omega \nabla u \nabla v \, dx \right| \leq \sum_{i=1}^d \int_\Omega |\partial_i u||\partial_i v| \, dx \\ &\overset{\text{CS}}{\leq} \sum_{i=1}^d \left(\int_\Omega |\partial_i u|^2 \, dx \right)^{\frac{1}{2}} \left(\int_\Omega |\partial_i v|^2 \, dx \right)^{\frac{1}{2}} \\ &\leq \left(\sum_{i=1}^d \int_\Omega |\partial_i u|^2 \, dx \right)^{\frac{1}{2}} \left(\sum_{i=1}^d \int_\Omega |\partial_i v|^2 \, dx \right)^{\frac{1}{2}} \\ &\leq \left(\int_\Omega |\nabla u|^2 \, dx + \int_\Omega u^2 \, dx \right)^{\frac{1}{2}} \left(\int_\Omega |\nabla v|^2 \, dx + \int_\Omega v^2 \, dx \right)^{\frac{1}{2}} \\ &= \|u\|_1 \, \|v\|_1 \,, \end{aligned}$$

  d.h. a ist stetig mit $c := 1$. Die Symmetrie von a ist trivial, also bleibt nur noch die Stetigkeit von F zu zeigen. Es sei $v \in H_0^1(\Omega)$, dann gilt

$$\begin{aligned} |F(v)| = |(f,v)| = \left| \int_\Omega fv \, dx \right| &\overset{\text{CS}}{\leq} \left(\int_\Omega |f|^2 \, dx \right)^{\frac{1}{2}} \left(\int_\Omega |v|^2 \, dx \right)^{\frac{1}{2}} \\ &\leq c \left(\int_\Omega |\nabla v|^2 + |v|^2 \, dx \right)^{\frac{1}{2}} = c \, \|v\|_1 \end{aligned}$$

  mit $0 < c := \int_\Omega |f|^2 \, dx < \infty$, wenn $f \in L_2(\Omega)$ ist. Damit ist F ein stetiges lineares Funktional und somit existiert nach Theorem 2.5 eine Eindeutige Lösung $u \in H_0^1(\Omega)$ für die schwache Formulierung des homogenen Dirichlet-Problems.

- *Bemerkung.* Die Stetigkeit vom Funktional F zeigt, welche Eigenschaft die Kraft f aus dem Dirichlet-Problem wenigstens quadratisch integrierbar sein muss, damit es eine schwache Lösung geben kann.

-

Bemerkung. (a) Mit H' bezeichnen wir den Dualraum zu einem Hilbertraum H.

(b) Den Dualraum zu $H^1(\Omega)$ bezeichnen wir mit $H^{-1}(\Omega)$.

- Hier noch eine Folgerung aus dem Satz von Lax-Milgram:

Satz 2.7. (Riesz'scher Darstellungssatz) *Es sei H ein Hilbertraum mit einem Skalarprodukt $(\cdot,\cdot)_H$. Es sei $F \in H'$, dann existiert genau ein $u \in H$, so dass*

$$(u,v)_H = F(v) \quad \forall\, v \in H\,.$$

Beweis. Dies ist eine direkte Folgerung aus dem Theorem 2.5. Die Abbildung $(\cdot,\cdot)_H : H \times H \to \mathbb{R}$ ist als Skalarprodukt bilinear, symmetrisch und positiv definit, damit auch bzgl. der auf H durch das Skalarprodukt induzierten Norm $\|v\|_H := \sqrt{(v,v)_H}$, koerziv. F ist als Element des Dualraumes H' eine lineare stetige Abbildung $F : H \to \mathbb{R}$ und damit folgt mit $a(\cdot,\cdot) := (\cdot,\cdot)_H$ aus dem Theorem von Lax-Milgram die Behauptung. □

2.2 Finite Elemente Methode

- FEM → einleitend ansprechen, dass analytische nicht immer lösbar
- Was ist Galerkin-Approximation und warum gibt es eine Lösung (hier ist Lax-Milgram auch anwendbar (warum?))
- Der für uns verwendete Finite Element Raum wird eingeführt (lineare Funktionen).
- Was ist eine Triangulierung (vgl. Braess auf Seite 58)?
- local-global node ordering zur Effizienzsteigerung

2.3 Adaptive Verfeinerungsstrategien

2.3.1 A posteriori Fehlerschätzer

- Fehlerschätzer → alle aufführen (s. Braess) → damit verbundene adaptive Verfeinerungsstrategien (wie arbeitet Matlab mit Verfeinerung und welche Verfeinerungen gibt es?)

2.4 Einführung in die Strukturmechanik

- Beschreibung der Kinematik: Referenz- bzw. Ausgangskonfiguration, Deformationsgradient, Verzerrungsmaße (Konti-Buch)
- Linearisierung der Verzerrungsmaße für unseren Fall (kleine Deformationen) mittels "Taylor"(siehe auch Gateaux-Ableitung - Seite 24 Konti Skript):
$$\boldsymbol{\varepsilon} = \frac{1}{2}(\nabla \boldsymbol{u} + \nabla^T \boldsymbol{u})$$
- Kinetik: Kräftegleichgewicht und äußere Kontaktlast
- Konzepte für ebene Spannungs- bzw. Verzerrungszustände
- Konstitutive Modelle (vor allem Materialgesetze) $\Rightarrow$ Hier vor allem Hooke:
$$\boldsymbol{\sigma} = \mathcal{C}\boldsymbol{\varepsilon} = 2\mu\boldsymbol{\varepsilon} + \lambda(\text{tr } \boldsymbol{\varepsilon})\boldsymbol{I}\,,$$
wobei λ, μ die Lamé-Konstanten sind (Materialabhängige Parameter). $\Rightarrow$ Hier noch mal den Zusammengang von Konstanten zu E, ν aufzeigen.
- falls Tensorrechnungen konkret benötigt werden, können diese im Anhang dargelegt werden

Kapitel 3

Variationsungleichungen

3.1 Ein Hindernisproblem

- Hindernisproblem: Auslenkung u einer Membran Ω unter Krafteinwirkung f, wobei die Membran durch ein Hindernis ψ behindert wird. Mathematische modelliert bedeutet dies:

$$\min_{u \in K} J(u) = \frac{1}{2} a(u,u) - (f,u) \tag{3.1}$$

mit $K := \{u \in H_0^1(\Omega) \mid u \geq \psi \text{ fast überall}\}$. $J(u)$ gibt wieder die Energie, die in der Membran gespeichert wird an. Neu ist also, dass die Lösung u nicht mehr in ganz $H_0^1(\Omega)$ liegt, sondern in einer Teilmenge.

3.1.1 Variationsformulierung für das Hindernisproblem

- Das Hindernisproblem zur Herleitung einer Variationsungleichung

$$a(u, v-u) \geq f(v-u) \, \forall \, v \in K$$

benutzen. Hierfür die Minimierungsaufgabe (3.1) unter Nebenbedingung optimieren. (Hierfür noch einmal in Nichtlineare Optimierung schauen.)

- Hier als Bemerkung vllt noch einmal anführen, dass die Variationsungleichung äquivalent zu der starken Formulierung

$$\begin{aligned} -\Delta u - f &\geq 0 \\ u - \psi &\geq 0 \\ (u-\psi)(-\Delta u - f) &= 0 \end{aligned}$$

ist. Beweis hierfür im Stephan-Skript (analog umzuschreiben).

3.1.2 Existenz und Eindeutigkeit der Lösung

- Kapitel 3 in [KO88] mit Theorem 3.1-3.4 (**Beweis vgl. NPDE I von Stephan Seite 39**, auch in Solution of Variational Inequalities in Mechanics (Theorem 1.1 Seite 4))

3.1.3 Lösung des Hindernisproblems mittels FEM

- Analog zum vorherigen Kapitel kann man auch im $\mathbb{R}^n$ Existenz und Eindeutigkeit der Lösung unter bestimmten Voraussetzungen zeigen. (vgl. Vug Skript Kapitel 2) $\Rightarrow$ Beachte hierfür auch den Fixpunktsatz von Brouwer.

3.2 Kontaktprobleme

3.2.1 Mathematische Modellierung von Kontaktproblemen

- Starke Formulierung (s. Wriggers Paper) für Kontaktproblem mit Signorini-Kontakt (ohne Reibung).

$$\operatorname{div} \boldsymbol{\sigma} + \boldsymbol{b} = \mathbf{0} \text{ in } \Omega \tag{3.2}$$
$$\boldsymbol{\sigma} - \mathcal{C}\boldsymbol{\varepsilon} = \mathbf{0} \text{ in } \Omega \tag{3.3}$$
$$\boldsymbol{\sigma} \cdot \boldsymbol{n} = \boldsymbol{t} \text{ auf } \Gamma_N \tag{3.4}$$
$$\boldsymbol{u} = \mathbf{0} \text{ auf } \Gamma_D \tag{3.5}$$
$$(\boldsymbol{u} \circ \chi - \boldsymbol{u}) \cdot \boldsymbol{n}_c + g \geq 0 \text{ auf } \Gamma_C \tag{3.6}$$

  sowie auf Γ_C muss $\sigma_n \leq 0$ (Normalenkraft $\sigma_n = \boldsymbol{n} \cdot (\boldsymbol{\sigma} \cdot \boldsymbol{n})$), $\boldsymbol{\sigma}_t = \mathbf{0}$ (keine Tangentialkraft, da keine Reibung – $\boldsymbol{\sigma}_t = \boldsymbol{\sigma} \cdot \boldsymbol{n} - \sigma_n \boldsymbol{n}$) und $((\boldsymbol{u} \circ \chi - \boldsymbol{u}) \cdot \boldsymbol{n}_c + g)\sigma_n = 0$, d.h. wenn kein Kontakt ist, ist die Normalkraft in den Punkten Null, also herrscht Kräftegleichgewicht.

- Anreißen von Kontaktproblem mit Tresca-Reibung (vgl. Numerik für Kontaktmechanik von Stephan und Vug von Starke) $\Rightarrow$ Herleitung der Variationsungleichung durch Ableitung nicht mehr möglich, da Reibungspotential nicht mehr differenzierbar.

3.2.2 Variationsformulierung für Kontaktprobleme

- Minimierung von Energiefunktional (vgl. [KO88] Seite 112 unten) mit $\boldsymbol{u} : \Omega \to \mathbb{R}^3$:

$$E(u) = \frac{1}{2} a(u,u) - f(u) \text{ mit}$$
$$a(u,u) = \int_\Omega \mathcal{C}\boldsymbol{\varepsilon}(\boldsymbol{u}) : \boldsymbol{\varepsilon}(\boldsymbol{u})\, d\Omega,\ f(u) = \int_\Omega \boldsymbol{b} \cdot \boldsymbol{u}\, d\Omega + \int_{\Gamma_N} \boldsymbol{t} \cdot \boldsymbol{u}\, d\Gamma$$

unter der Nebenbedingung $\boldsymbol{n} \cdot \boldsymbol{u} - g \leq 0$ auf Γ_C (siehe Vug Skript), bzw. $(\boldsymbol{u} \circ \chi - \boldsymbol{u}) \cdot \boldsymbol{n}_c + g \geq 0$ auf Γ_C (etwas allgemeiner, vgl. Wriggers Paper).

- Herleitung auch über starke Formulierung möglich, vgl. Stephan – Kontaktprobleme.
- Herleitung der Variationsformulierung: Finde $\boldsymbol{u} \in K$: $a(\boldsymbol{u}, \boldsymbol{v} - \boldsymbol{u}) \geq f(\boldsymbol{v} - \boldsymbol{u}) \, \forall \, \boldsymbol{v} \in K$ (s. auch Wriggers Paper) analog zum Hindernisproblem (nicht mehr ausführlich, wenn oben schon ausführlich).
- [KO88] Seite 113 für Bedingung für die Eindeutigkeit und Existenz der Lösung des Problems (hierfür wird Korn's Ungleichung benötigt $\Rightarrow$ vielleicht Anhang?).

3.2.3 Lösung des Kontaktproblems mittels FEM

- Beschreibe das diskrete Problem, was man bekommt mit: Finde $\boldsymbol{x}^* \in \mathbb{R}^N$ mit $B\boldsymbol{x}^* \geq c$, so dass

$$(A\boldsymbol{x}^* - \boldsymbol{b})^T (\boldsymbol{x} - \boldsymbol{x}^*) \geq 0 \, \forall \boldsymbol{x} \in \mathbb{R}^N \text{ mit } B\boldsymbol{x} \geq \boldsymbol{c} \,,$$

wobei

$$A = \left[\int_\Omega \mathcal{C}\boldsymbol{\varepsilon}(\boldsymbol{\Psi}_j) : \boldsymbol{\varepsilon}(\boldsymbol{\Psi}_i) \, d\Omega \right]_{1 \leq i,j \leq N}, \; \boldsymbol{b} = \left[\int_\Omega \boldsymbol{b} \cdot \boldsymbol{\Psi}_i \, d\Omega + \int_{\Gamma_N} \boldsymbol{t} \cdot \boldsymbol{\Psi}_i \, ds \right]_{1 \leq i \leq N}$$

$$B = [(\boldsymbol{\Psi}_j(\chi(\boldsymbol{x}_i)) - \boldsymbol{\Psi}_j(\boldsymbol{x}_i)) \cdot \boldsymbol{n}_c(\boldsymbol{x}_i)]_{\boldsymbol{x}_i \in \Gamma_c, 1 \leq j \leq N}, \; c = [-g(\boldsymbol{x}_i)]_{\boldsymbol{x}_i \in \Gamma_c}$$

Dieses Problem ist (wie vorher schon gezeigt) äquivalent zu einem quadratischen Problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^N} \frac{1}{2} \boldsymbol{x}^T A \boldsymbol{x} - \boldsymbol{b}^T \boldsymbol{x} \text{ s.t. } B\boldsymbol{x} \geq \boldsymbol{c} \,,$$

d.h. Lösbarkeit des quadratischen Programms sollte auch gezeigt sein (vgl. Vug Skript oder auch nichtlineare Optimierung).

Kapitel 4

Ein hierarchischer Fehlerschätzer für Hindernisprobleme

- Herleitung des Fehlerschätzers bei Hindernisproblemen (s. Mainpaper)
- Vergleich Hindernisprobleme zu Kontaktproblemen → warum gerade dieser Fehlerschätzer bei Hindernis- bzw. Kontaktproblemen

4.1 Herleitung eines a posteriori hierarchischen Fehlerschätzers

4.1.1 Diskretisierung

4.1.2 Lokaler Anteil des Fehlerschätzers

4.1.3 Oszillationsterme

4.1.4 Zuverlässigkeit des Fehlerschätzers

4.1.5 Effektivität des Fehlerschätzers

4.2 Ein adaptiver Algorithmus

4.3 Erfüllung einer Saturationseigenschaft

4.4 Übertragung des Fehlerschätzers auf Kontaktprobleme

Kapitel 5

Implementierung des Fehlerschätzers in Matlab

- Grundlegender Aufbau des Programms
- Gründe warum wo was.
- Warum Verwendung von Sparse, IPM und large scale?
- Berechnung der einzelnen lokalen Element-Steifigkeitsmatrizen bzw. Element-Vektoren (siehe hierfür auch die Berechnung für den Vektor ρ_S – hier ist die Berechnung durch lokalen Vektoren auch schneller gemacht worden).
- dokumentierter Quellcode ist im Anhang zu finden

Kapitel 6

Validierung

- numerisches Beispiel (Problemstellung) → vielleicht mit Kontakt und nur Hindernis
- Vergleich mit Analytischer Lösung?! (Tabelle mit Ergebnissen) → Ergebnisse diskutieren

6.1 Numerisches Beispiel zum Hindernisproblem

6.2 Numerisches Beispiel zum Kontaktproblem

Kapitel 7

Zusammenfassung und Ausblick

- kurz einleiten, worum es ging (Einleitung in einem Absatz zusammenfassen)
- Was ist rausgekommen?!
- Ausblick: Was ist noch offen geblieben, was kann man noch machen... In dieser Arbeit linearisierte Verzerrung verwendet; kann verallgemeinert werden durch allgemeine Verzerrungstensoren (bzgl. der jeweiligen Konfiguration).

Literaturverzeichnis

[BCH05] Bartels, S. ; Carstensen, C. ; Hecht, A.: 2D isoparametric FEM in MATLAB / Humboldt-Universität, Berlin. 2005. – Forschungsbericht

[BCH07] Braess, D. ; Carstensen, C. ; Hoppe, R.: Convergence analysis of a conforming adaptive finite element method for an obstacle problem. In: *Numerische Mathematik* 107 (2007), S. 455–471

[Bra05] Braess, Dietrich: A Posteriori Error Estimators for Obstacle Problems – Another Look / Faculty of Mathematics, Ruhr-University. 2005. – Forschungsbericht

[Bra13] Braess, Dietrich: *Finite Elemente – Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie.* 5. Auflage. Springer-Verlag, 2013

[CSW99] Carstensen, C. ; Scherf, O. ; Wriggers, P.: Adaptive finite elements for elastic bodies in contact. In: *SIAM J. Sci. Comput.* 20 (1999), Nr. 5, S. 1605–1626

[GRT09] Göpfert, A. ; Riedrich, T. ; Tammer, C.: *Angewandte Funktionalanalysis.* Vieweg und Teubner, 2009

[Joh92] Johnson, Claes: Adaptive finite element methods for the obstacle problem. In: *Math. Models Methods Appl. Sci.* 2 (1992), Nr. 4, S. 483–487

[KO88] Kikuchi, N. ; Oden, J.T.: *Contact Problems in Elasticity: A Study of Variational Inequalities and Finite Element Methods.* SIAM, 1988

[KZ11] Kornhuber, Ralf ; Zou, Qingsong: Efficient and reliable hierarchical error estimates for the discretization error of elliptic obstacle problems. In: *Mathematics of Computation* 80 (2011), Nr. 273, S. 69–88

[NW06] Nocedal, Jorge ; Wright, Stephen J.: *Numerical Optimization.* 2. ed. New York, NY : Springer, 2006

[Sta08] Starke, Gerhard: Numerik partieller Differentialgleichungen / IFAM - Universität Hannover. 2008. – Vorlesungsskript

[Sta11] Starke, Gerhard: Variationsungleichungen / IFAM - Universität Hannover. 2011. – Vorlesungsskript

[Ste12] Stephan, Ernst P.: Numerik partieller Differentialgleichungen I / IFAM - Universität Hannover. 2012. – Vorlesungsskript

[Wal11] Walker, Christoph: Partielle Differentialgleichungen / IFAM - Universität Hannover. 2011. – Vorlesungsskript

[Zou11] Zou, Qingsong: Efficient and reliable hierarchical error estimates for an elliptic obstacle problem. In: *Applied Numerical Mathematics* 61 (2011), S. 344–355

[ZVKG11] Zou, Q. ; Veeser, A. ; Kornhuber, R. ; Gräser, C.: Hierarchical error estimates for the energy functional in obstacle problems. In: *Numerische Mathematik* (2011), Nr. 117, S. 653–677

Anhang A

Funktionalanalysis

A.1 Sobolev-Räume

Sei im Weiteren $\emptyset \neq \Omega \subset \mathbb{R}^n$. Wir definieren den Sobolev-Raum allgemein wie folgt (vgl. [Bra13] Kaptitel II, §2 und [Wal11] Kapitel 6).

Definition A.1. Seien $1 \leq p \leq \infty$ und $m \in \mathbb{N}$. Die Menge

$$W_p^m(\Omega) := \left(\{u \in L_p(\Omega) \mid \partial^\alpha u \in L_p(\Omega) \, \forall \, |a| \leq m\}, \|\cdot\|_{W_p^m}\right)$$

heißt *Sobolev-Raum* der Ordnung m. Dabei ist

$$\|u\|_{W_p^m} := \|u\|_{W_p^m(\Omega)} := \left(\sum_{|\alpha| \leq m} \|\partial^\alpha u\|_{L_p}^p\right)^{\frac{1}{p}},$$

wenn $1 \leq p < \infty$. Im Fall $p = \infty$ ist $\|u\|_{W_p^m} := \max_{|\alpha| \leq m} \|\partial^\alpha u\|_\infty$.

Weiterhin bezeichne $L_p(\Omega)$ den *Lebesgue-Raum*, d.h. den Raum der messbaren Funktionen, deren p-te Potenz Lebesgue-integrierbar über Ω ist, d.h.

$$L_p(\Omega) := \left(\{u : \Omega \to \mathbb{R} \mid f \text{ messbar}, \|\cdot\|_{L_p} < \infty\}, \|\cdot\|_{L_p}\right),$$

wobei $\|u\|_{L_p} := \|u\|_{L_p(\Omega)} = \|u\|_{W_p^0}$.

Definition A.2. Der Raum

$$\mathcal{D}(\Omega) := C_c^\infty(\Omega) = \{\varphi \in C^\infty(\Omega) \mid \operatorname{supp}(\varphi) \subset\subset \Omega\}$$

heißt der *Raum der Testfunktionen*, wobei $K \subset\subset \Omega :\Leftrightarrow \bar{K} \subset \Omega$ kompakt.

Bemerkung A.3. Seien $u \in W_p^m(\Omega)$, $\varphi \in \mathcal{D}(\Omega)$ und $\alpha \in \mathbb{N}^n$ mit $|\alpha| \leq m$. Dann bezeichnen wir $v = \partial^\alpha u$ als *schwache Ableitung* von u, wenn gilt

$$\int_\Omega v \cdot \varphi \, dx = (-1)^{|\alpha|} \int_\Omega u \cdot \partial^\alpha \varphi \, dx \,.$$

Beispiel A.4. Es sei $\Omega = (-1,1) \subset \mathbb{R}$ und $u(x) = |x| \in L_2(\Omega)$. Betrachten wir $v(x) = \operatorname{sign}(x)$, so ergibt sich für $\varphi \in \mathcal{D}(\Omega)$

$$\begin{aligned}\int_\Omega v \cdot \varphi \, dx &= \int_{-1}^{0} -1 \cdot \varphi(x)\, dx + \int_0^1 1 \cdot \varphi(x)\, dx \\ &= -x\varphi(x)\Big|_{-1}^{0} - \int_{-1}^{0} -x\varphi'(x)\, dx + x\varphi(x)\Big|_0^1 - \int_0^1 x\varphi'(x)\, dx \\ &= -\int_{-1}^{1} |x|\varphi'(x)\, dx = (-1)^1 \int_\Omega u \cdot \varphi' \, dx\,,\end{aligned}$$

da $\varphi(-1) = \varphi(1) = 0$. Also ist $v = \partial u$ und somit $u \in W_2^1(\Omega)$. Analog kann man nachrechnen, dass

$$\int_\Omega v \cdot \varphi' \, dx = -2\varphi(0)$$

ist und somit u nicht zweimal schwach ableitbar ist, d.h. $u \notin W_2^2(\Omega)$.

Wir wollen in der Theorie der Finiten Elemente Methode vor allem Sobolev-Räume über dem Raum $L_2(\Omega)$ betrachten, daher ist folgender Satz essentiell.

Satz A.5. *Es seien $1 \le p \le \infty$ und $m \in \mathbb{N}$. Dann gilt:*

(a) $W_p^m(\Omega)$ *ist ein Banachraum.*

(b) $H^m(\Omega) := W_2^m(\Omega)$ *ist ein Hilbertraum mit Skalarprodukt*

$$(u,v)_m := (u,v)_{H^m(\Omega)} := \sum_{|\alpha| \le m} (\partial^\alpha u, \partial^\alpha v)_0 \quad \forall\, u, v \in H^m(\Omega)\,,$$

wobei

$$(u,v)_0 := (u,v)_{L_2(\Omega)} := \int_\Omega uv \, dx\,.$$

Bemerkung A.6. (a) Die Norm auf $H^m(\Omega)$ ergibt sich analog zur Norm des allgemeinen Sobolev-Raumes durch das Skalarprodukt, d.h. $\|u\|_m := \|u\|_{H^m(\Omega)} := \|u\|_{W_2^m}$.

(b) Analog dazu definieren wir die Halbnorm $|\cdot|_m$ auf H^m wie folgt:

$$|u|_m := |u|_{H^m(\Omega)} := \left(\sum_{|\alpha|=m} \|\partial^\alpha u\|_{L_2}^2 \right)^{\frac{1}{2}}.$$

Definition A.7. Der Raum $H_0^m(\Omega)$ ist die Vervollständigung von $\mathcal{D}(\Omega)$ bzgl. der Norm $\|\cdot\|_m$.

Bemerkung A.8. Die Funktionen $u \in H_0^m(\Omega)$ können als die Funktionen $u \in H^m(\Omega)$ mit $u = 0$ auf $\partial\Omega$ aufgefasst werden.

A.2 Optimalitätskriterien

Zunächst definieren wir einen verallgemeinerten Begriff der Richtungsableitung, der auch auf unendlich dimensionalen Vektorräumen existiert.

Definition A.9. Es seien V ein Vektorraum, $M \subset V$ und W ein normierter Raum, sowie $F : M \to W$ eine Abbildung, $x_0 \in M$ und $v \in V$. Dann heißt F *Gâteaux-differenzierbar* (bzw. in Richtung v an der Stelle x_0 differenzierbar), falls es ein $\varepsilon > 0$ mit $[x_0 - \varepsilon v, x_0 + \varepsilon v] \subset M$ gibt und der Grenzwert

$$\mathcal{D}_v F(x_0) := \frac{d}{dt} F(x_0 + tv)\Big|_{t=0} := \lim_{t \to 0} \frac{F(x_0 + tv) - F(x_0)}{t} \tag{A.1}$$

in W existiert. $\mathcal{D}_v F(x_0)$ heißt dann *Gâteaux-Ableitung* von F an der Stelle x_0 in Richtung v.

Falls wir nur $[x_0, x_0 + \varepsilon v] \subset M$ voraussetzen, so können wir in (A.1) $\lim_{t \to 0}$ durch $\lim_{t \to +0}$ ersetzen. Dann nennen wir (A.1) die *rechtsseitige Gâteaux-Ableitung* und bezeichnen diese mit $\mathcal{D}_v^+ F(x_0)$.

Für die Variationsrechnung sind folgende zwei Sätze für uns von besonderer Bedeutung.

Satz A.10. (Charakterisierungssatz der konvexen Optimierung) *Es seien $M \subset V$ eine konvexe Menge, V ein Vektorraum und $F : M \to \mathbb{R}$ ein konvexes Funktional. Dann gilt für $x_0, x \in M$:*

x_0 ist Lösung von $\min\limits_{x \in M} F(x)$ genau dann, wenn für alle $x \in M$ gilt

$$\mathcal{D}_{x - x_0}^+ F(x_0) \geq 0 \,.$$

Beweis. Siehe [GRT09], Kapitel 3.3.3, Satz 3.34. □

Satz A.11. *Es sei $U \subset V$ ein (Unter-)Vektorraum, V ein Vektorraum und $F : U \to \mathbb{R}$ eine Gâteaux-differenzierbare konvexe Funktion. Dann ist $x_0 \in U$ genau dann Lösung von $\min\limits_{x \in U} F(x)$, wenn für alle $u \in U$ gilt*

$$\mathcal{D}_u F(x_0) = 0 \,.$$

Beweis. Siehe [GRT09], Kapitel 3.3.3, Satz 3.35. □

Anhang B

Optimierung

B.1 Quadratische Programmierung

Um im folgenden die Idee des Algorithmus zu verstehen, führen wir zunächst grundlegende Begriffe ein. Ein quadratisches Problem mit Gleichungs- und Ungleichungsnebenbedingungen ist von der Form

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{c} \\ \text{s.t.} \quad & \boldsymbol{a}_i^T\boldsymbol{x} = b_i\,, \quad i \in \mathcal{E}, \\ & \boldsymbol{a}_i^T\boldsymbol{x} \geq b_i\,, \quad i \in \mathcal{I}, \end{aligned} \tag{B.1}$$

wobei $\mathcal{E}$ und $\mathcal{I}$ die Indexmengen der Gleichungs- und Ungleichungsnebenbedingungen darstellen und $\boldsymbol{c}, \boldsymbol{x}, \boldsymbol{a}_i \in \mathbb{R}^n, b_i \in \mathbb{R}, i \in \mathcal{E} \cup \mathcal{I}$, sowie G eine symmetrische $(n \times n)$-Matrix ist, welche die Hesse-Matrix des Problems darstellt. Damit ist die Hesse-Matrix konstant und daher das Problem konvex, wenn G positiv semidefinit ist. (Ist G positiv definit, so nennen wir das Problem strikt konvex. Wenn G indefinit ist, ist (B.1) „nicht konvex“.)

Da sonst das quadratische Problem (und damit der Active-Set Algorithmus) zu kompliziert wird, betrachten wir hier nur den konvexen Fall. Für diesen Fall können wir leicht zeigen, dass eine Lösung $\boldsymbol{x}^*$, die die Bedingungen 1. Ordnung erfüllt, auch globale Lösung des Problems ist (s. Theorem B.1). Anschaulich kann es im indefiniten Fall mehrere optimale Punkte geben, die voneinander getrennt liegen, d.h. die Menge der optimalen Punkte ist nicht zusammenhängend, wodurch das Auffinden des globalen Minimums erschwert wird.

Die notwendigen Bedingungen 1. Ordnung sind die KKT-Bedingungen und können hier angewendet werden, da die Restriktionen und die Zielfunktion stetig differenzierbar sind. Die Lagrangefunktion $\mathcal{L}$ für das quadratische Problem ist

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{c} - \sum_{i \in \mathcal{I} \cup \mathcal{E}} \lambda_i(\boldsymbol{a}_i^T\boldsymbol{x} - b_i)\,. \tag{B.2}$$

Damit ergeben sich – vgl. [NW06], Theorem 12.1 – mit der Menge der aktiven Nebenbedingungen $\mathcal{A}(\boldsymbol{x}^*) = \{i \in \mathcal{E} \cup \mathcal{I} : \boldsymbol{a}_i^T \boldsymbol{x}^* = b_i\}$ die KKT-Bedingungen

$$\begin{aligned} \nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = G\boldsymbol{x}^* + \boldsymbol{c} - \sum_{i \in \mathcal{A}(\boldsymbol{x}^*)} \lambda_i^* \boldsymbol{a}_i = 0\,, \\ \boldsymbol{a}_i^T \boldsymbol{x}^* = b_i\,, \quad \forall\, i \in \mathcal{A}(\boldsymbol{x}^*), \\ \boldsymbol{a}_i^T \boldsymbol{x}^* \geq b_i\,, \quad \forall\, i \in \mathcal{I} \setminus \mathcal{A}(\boldsymbol{x}^*), \\ \lambda_i^* \geq 0\,, \quad \forall i \in \mathcal{I} \cap \mathcal{A}(\boldsymbol{x}^*). \end{aligned} \tag{B.3}$$

Hierbei ist $\boldsymbol{x}^*$ Lösung von (B.1) und erfüllt die LICQ-Bedingung; $\boldsymbol{\lambda}^*$ ist dazugehöriger optimaler Lagrange-Multiplikator. In (B.3) wird die Komplementaritätsbedingung $\lambda_i^* c_i(\boldsymbol{x}^*) = 0$ impliziert durch $\lambda_i^* = 0 \,\forall\, i \notin \mathcal{A}(\boldsymbol{x}^*)$.

Theorem B.1. *Wenn $\boldsymbol{x}^*$ die Bedingungen* (B.3) *erfüllt mit $\lambda_i^*, i \in \mathcal{A}(\boldsymbol{x}^*)$ und G ist positiv semidefinit, dann ist $\boldsymbol{x}^*$ eine globale Lösung von* (B.1).

Beweis. Wenn $\boldsymbol{x}$ ein beliebiger weiterer zulässiger Punkt für (1.1) ist, gelten die Restriktionen $\boldsymbol{a}_i^T \boldsymbol{x} = b_i, i \in \mathcal{E}$, sowie $\boldsymbol{a}_i^T \boldsymbol{x} \geq b_i, i \in \mathcal{I} \cap \mathcal{A}(\boldsymbol{x}^*)$ für $\boldsymbol{x}$ und damit gilt zusammen mit der ersten Bedingung von (B.3), dass

$$(\boldsymbol{x} - \boldsymbol{x}^*)^T (G\boldsymbol{x}^* + \boldsymbol{c}) = \sum_{i \in \mathcal{E}} \underbrace{\lambda_i^* \boldsymbol{a}_i^T (\boldsymbol{x} - \boldsymbol{x}^*)}_{\geq 0} + \sum_{i \in \mathcal{A}(x^*) \cap \mathcal{I}} \underbrace{\lambda_i^* \boldsymbol{a}_i^T (\boldsymbol{x} - \boldsymbol{x}^*)}_{\geq 0} \geq 0\,.$$

Dann drücken wir $q(\boldsymbol{x})$ durch $q(\boldsymbol{x}^*)$ aus und wenden die obere Ungleichung sowie die positive Semidefinitheit für G an, um zu zeigen, dass $q(\boldsymbol{x}) \geq q(\boldsymbol{x}^*)$ ist. Damit ist $\boldsymbol{x}^*$ globale Lösung des quadratischen Problems. □

Daher ist im positiv semidefiniten Fall gesichert, dass ein optimaler Punkt auch gleichzeitig globale Lösung ist.

B.2 Active Set-Methode für konvexe QPs

Wenn wir eine Lösung $\boldsymbol{x}^*$ für das Problem (B.1) kennen, so ist auch die Menge der aktiven Nebenbedingungen $\mathcal{A}(x^*)$ bekannt und wir können (B.1) vereinfachen zum Optimierungsproblem

$$\min_{\boldsymbol{x}} \quad q(\boldsymbol{x}) = \frac{1}{2} \boldsymbol{x}^T G \boldsymbol{x} + \boldsymbol{x}^T \boldsymbol{c}\,, \quad \text{s.t.} \quad \boldsymbol{a}_i^T \boldsymbol{x} = b_i\,, \quad i \in \mathcal{A}(\boldsymbol{x}^*)\,. \tag{B.4}$$

Dieses könnten wir dann beispielsweise mit direkten Verfahren wie der Schur-Komplement-Methode oder der Nullraum-Methode lösen. Natürlich ist die optimale Lösung zu Beginn noch nicht bekannt und damit auch nicht die aktiven Restriktionen. Jedoch können wir diese Idee für die Active-Set-Methode verwenden.

Das Hauptziel der Active-Set-Methode ist, die Menge der aktiven Restriktionen bzgl. der optimalen Lösung zu finden, wobei wir hier die primale

Variante betrachten wollen, in der die Approximierte $\boldsymbol{x}_k$ zulässig bzgl. des primalen Problems ist.

Die Grundidee ist, ein quadratisches Teilproblem zu lösen, bei dem wir bestimmte Nebenbedingungen aus Problem (B.1) bzgl. $\mathcal{I}$ als aktiv annehmen. Die dadurch beschriebene Indexmenge der aktiven Restriktionen für $\boldsymbol{x}_k$ im k-ten Schritt heißt *working set* und kann wie folgt beschrieben werden

$$\mathcal{W}_k = \{i \,|\, \boldsymbol{a}_i^T \boldsymbol{x}_k = b_i, i \in \mathcal{E} \cup \mathcal{J}, \mathcal{J} \subset \mathcal{I}\}\,.$$

Hierbei muss vorausgesetzt werden, dass die Nebenbedingungen in $\mathcal{W}_k$ die LICQ-Bedingung erfüllen, selbst wenn diese bezogen auf alle Nebenbedingungen an der Stelle x_k nicht erfüllt wird.

Wir betrachten nun den k-ten Schritt mit der Approximierten $\boldsymbol{x}_k$ und dem working set $\mathcal{W}_k$. Wir berechnen die neue Iterierte $\boldsymbol{x}_{k+1}$, indem wir eine Richtung $\boldsymbol{p}$ finden, in der wir unter den Nebenbedingungen $\mathcal{W}_k$ die Funktion q minimieren. Hierfür betrachten wir $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{p}$ und setzen $\boldsymbol{x}_{k+1}$ in q ein:

$$\begin{aligned} q(\boldsymbol{x}_{k+1}) = q(\boldsymbol{x}_k + \boldsymbol{p}) &= \frac{1}{2}(\boldsymbol{x}_k + \boldsymbol{p})^T G(\boldsymbol{x}_k + \boldsymbol{p}) + (\boldsymbol{x}_k + \boldsymbol{p})^T \boldsymbol{c} \\ &= \frac{1}{2}\boldsymbol{x}_k^T G \boldsymbol{x}_k + \underbrace{\boldsymbol{x}_k^T G \boldsymbol{p}}_{\text{da } G \text{ symm.}} + \frac{1}{2}\boldsymbol{p}^T G \boldsymbol{p} + \boldsymbol{x}_k^T \boldsymbol{c} + \boldsymbol{p}^T \boldsymbol{c} \\ &= \frac{1}{2}\boldsymbol{p}^T G \boldsymbol{p} + \boldsymbol{g}_k^T \boldsymbol{p} + \rho_k\,, \end{aligned}$$

wobei $\boldsymbol{g}_k = G\boldsymbol{x}_k + \boldsymbol{c}$ und $\rho_k = \frac{1}{2}\boldsymbol{x}_k^T G \boldsymbol{x}_k + \boldsymbol{x}_k^T \boldsymbol{c}$. Da wir den Parameter $\boldsymbol{p}$ so wählen wollen, so dass $q(\boldsymbol{x}_{k+1})$ minimal wird, ist der Term ρ_k bzgl. des Problems konstant und kann somit für die Lösung jenes weggelassen werden. Da weiterhin auch $\boldsymbol{x}_{k+1}$ die aktiven Nebenbedingungen $\mathcal{W}_k$ erfüllen soll, gilt

$$\boldsymbol{a}_i^T \boldsymbol{p} = \boldsymbol{a}_i^T (\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) = \underbrace{\boldsymbol{a}_i^T \boldsymbol{x}_{k+1}}_{=b_i} - \underbrace{\boldsymbol{a}_i^T \boldsymbol{x}_k}_{=b_i} = 0 \quad \forall\, i \in \mathcal{W}_k\,.$$

Zusammengefasst müssen wir also im k-ten Schritt das Teilproblem

$$\begin{aligned} \min_{\boldsymbol{p}} \quad & \frac{1}{2}\boldsymbol{p}^T G \boldsymbol{p} + \boldsymbol{g}_k^T \boldsymbol{p}\,, \\ \text{s.t.} \quad & \boldsymbol{a}_i^T \boldsymbol{p} = 0\,, \quad \forall i \in \mathcal{W}_k \end{aligned} \tag{B.5}$$

lösen. Die Lösung im k-ten Schritt von (B.5) bezeichnen wir mit $\boldsymbol{p}_k$. Umgekehrt gilt damit, analog zur obigen Rechnung, natürlich auch, dass für alle $i \in \mathcal{W}_k$ die Restriktion aktiv bleibt für $\boldsymbol{x}_k + \alpha \boldsymbol{p}_k$ mit beliebigem α. Da G positiv definit ist, kann (B.5) nun – wie schon bei (B.4) erwähnt – mit Schur-Komplement-Methode oder Nullraum-Methode gelöst werden.

Wie wir schon wissen, ist die neue Iterierte $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{p}_k$ bzgl. $\mathcal{W}_k$ immer noch zulässig. Nun müssen wir jedoch feststellen, ob diese Iterierte

auch alle übrigen Restriktionen mit $i \notin \mathcal{W}_k$ erfüllt. Ist dies der Fall, so setzen wir $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{p}_k$, ansonsten suchen wir das größtmögliche $\alpha_k \in [0, 1]$, so dass

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$$

zulässig bleibt. Hierfür betrachten wir zwei Fälle.

Fall 1: Gilt für ein $i \notin \mathcal{W}_k$, dass $\boldsymbol{a}_i^T \boldsymbol{p}_k \geq 0$ ist, so folgt

$$\boldsymbol{a}_i^T (\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k) = \boldsymbol{a}_i^T \boldsymbol{x}_k + \underbrace{\alpha_k \boldsymbol{a}_i^T \boldsymbol{p}_k}_{\geq 0} \geq \boldsymbol{a}_i^T \boldsymbol{x}_k \geq b_i \,,$$

da $\alpha_k \geq 0$, d.h. für diese Nebenbedingungen müssen wir für die Wahl von α_k nichts beachten.

Fall 2: Existiert ein $i \notin \mathcal{W}_k$, für das $\boldsymbol{a}_i^T \boldsymbol{p}_k < 0$ ist, so gilt

$$\begin{aligned} & \boldsymbol{a}_i^T (\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k) \geq b_i \\ \iff \quad & \boldsymbol{a}_i^T \boldsymbol{x}_k + \alpha_k \boldsymbol{a}_i^T \boldsymbol{p}_k \geq b_i \\ \iff \quad & \alpha_k \underbrace{\boldsymbol{a}_i^T \boldsymbol{p}_k}_{<0} \geq b_i - \boldsymbol{a}_i^T \boldsymbol{x}_k \\ \iff \quad & \alpha_k \leq \frac{b_i - \boldsymbol{a}_i^T \boldsymbol{x}_k}{\boldsymbol{a}_i^T \boldsymbol{p}_k} \,. \end{aligned} \tag{B.6}$$

Damit folgt mit (B.6) und den vorherigen Überlegungen, dass zusammengefasst

$$\alpha_k = \min \left\{ 1, \min_{i \notin \mathcal{W}_k, \boldsymbol{a}_i^T \boldsymbol{p}_k < 0} \frac{b_i - \boldsymbol{a}_i^T \boldsymbol{x}_k}{\boldsymbol{a}_i^T \boldsymbol{p}_k} \right\} \tag{B.7}$$

gilt. Eine Restriktion $i \notin \mathcal{W}_k$, für die das Minimum für α_k angenommen wird, nennen wir *blocking constraint*; diese muss nicht eindeutig sein, da wir beispielsweise anschaulich auch von einer Ecke geblockt werden können. Ist $\alpha_k = 1$, so werden alle Restriktion außerhalb vom working set mit dem Schritt $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{p}_k$ erfüllt, d.h. es gibt keine blocking constraint. Gibt es eine Nebenbedingung $j \notin \mathcal{W}_k$, die aktiv ist, obwohl sie nicht zum working set gehört, so gilt

$$\begin{aligned} \alpha_k &= \min \left\{ 1, \min_{i \notin \mathcal{W}_k, \boldsymbol{a}_i^T \boldsymbol{p}_k < 0} \frac{b_i - \boldsymbol{a}_i^T \boldsymbol{x}_k}{\boldsymbol{a}_i^T \boldsymbol{p}_k} \right\} \\ &= \min \left\{ 1, \frac{b_j - \boldsymbol{a}_j^T \boldsymbol{x}_k}{\boldsymbol{a}_j^T \boldsymbol{p}_k} \right\} \\ &= \min \left\{ 1, \frac{b_j - b_j}{\boldsymbol{a}_j^T \boldsymbol{p}_k} \right\} = 0 \,. \end{aligned}$$

Es sei $j \notin \mathcal{W}_k$ nun ein Index einer blocking constraint. Dann ist

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k = \boldsymbol{x}_k + \frac{b_j - \boldsymbol{a}_j^T \boldsymbol{x}_k}{\boldsymbol{a}_j^T \boldsymbol{p}_k} \boldsymbol{p}_k \,.$$

Setzen wir $\boldsymbol{x}_{k+1}$ in die j-te Restriktion ein, so erhalten wir

$$\begin{aligned} \boldsymbol{a}_j^T \boldsymbol{x}_{k+1} &= \boldsymbol{a}_j^T \left(\boldsymbol{x}_k + \frac{b_j - \boldsymbol{a}_j^T \boldsymbol{x}_k}{\boldsymbol{a}_j^T \boldsymbol{p}_k} \boldsymbol{p}_k \right) = \boldsymbol{a}_j^T \boldsymbol{x}_k + \frac{b_j - \boldsymbol{a}_j^T \boldsymbol{x}_k}{\cancel{\boldsymbol{a}_j^T \boldsymbol{p}_k}} \cdot \cancel{\boldsymbol{a}_j^T \boldsymbol{p}_k} \\ &= \boldsymbol{a}_j^T \boldsymbol{x}_k + b_j - \boldsymbol{a}_j^T \boldsymbol{x}_k = b_j \,, \end{aligned}$$

d.h. die blocking constraint ist für die neue Iterierte $\boldsymbol{x}_{k+1}$ nach Konstruktion aktiv. Daher setzen wir als neues working set $\mathcal{W}_{k+1} = \mathcal{W}_k \cup \{j\}$.

Das oben beschriebene Vorgehen wiederholen wir so lange, bis wir das working set $\hat{\mathcal{W}}$ mit dem Minimum des quadratischen Problems $\hat{\boldsymbol{x}}$ gefunden haben. Dies ist leicht zu erkennen, da wir (B.1) auf $\mathcal{W}_k$ nicht weiter minimieren können, sobald es keinen Schritt p gibt, in dessen Richtung wir q verringern können, d.h. wenn $\boldsymbol{p} = \boldsymbol{0}$ die Lösung für das Teilproblem (B.5) ist. Dann ist der optimale Punkt $\hat{\boldsymbol{x}}$ bzgl. des working sets $\hat{\mathcal{W}} \subset \mathcal{A}(\hat{\boldsymbol{x}})$ gefunden.

Wir müssen jetzt überprüfen, ob $\hat{\boldsymbol{x}}$ die KKT-Bedingungen erfüllt. Wir wissen, dass für $\boldsymbol{p} = \boldsymbol{0}$ die KKT-Bedingungen für (B.5)

$$\begin{pmatrix} G & A^T \\ A & 0 \end{pmatrix} \cdot \begin{pmatrix} -\boldsymbol{p} \\ \hat{\boldsymbol{\lambda}} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{g}} \\ \hat{\boldsymbol{h}} \end{pmatrix}$$

mit $\hat{\boldsymbol{g}} = \boldsymbol{c} + G\hat{\boldsymbol{x}}, \boldsymbol{h} = A\hat{\boldsymbol{x}} + \boldsymbol{b}$ und $\boldsymbol{p} = \boldsymbol{0}$ erfüllt. Daraus folgt

$$\begin{aligned} A^T \hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{g}} \quad &\iff \quad \sum_{i \in \hat{\mathcal{W}}} \boldsymbol{a}_i \hat{\lambda}_i = G\hat{\boldsymbol{x}} + \boldsymbol{c} \,, \\ \boldsymbol{0} = \hat{\boldsymbol{h}} \quad &\iff \quad A\hat{\boldsymbol{x}} = \boldsymbol{b} \,, \end{aligned}$$

wobei A die Gradienten $\boldsymbol{a}_i^T$ der aktiven Restriktionen $\hat{\mathcal{W}}$ zeilenweise enthält. Damit werden die ersten beiden KKT-Bedingungen aus (B.3) erfüllt. Da die Schrittlänge α_k mit (B.6) so gewählt ist, dass die übrigen Restriktionen erfüllt bleiben, gilt auch die dritte Bedingung aus (B.3). Es bleibt zu überprüfen, ob die Lagrange-Multiplikatoren $\hat{\lambda}_i \geq 0$ sind.

Gilt $\hat{\lambda}_i \geq 0$ für alle $i \in \hat{\mathcal{W}} \cap \mathcal{I}$, so sind alle KKT-Bedingungen erfüllt und damit $\boldsymbol{x}^* = \hat{\boldsymbol{x}}$. Existiert allerdings ein $j \in \hat{\mathcal{W}} \cap \mathcal{I}$, so dass $\hat{\lambda}_j < 0$ ist, so können wir den Wert von q noch weiter verringern, indem wir die j-te Restriktion wegfallen lassen (vlg. [NW06], Kapitel 12.3). Dies zeigt das folgende Theorem.

Theorem B.2. *Der Punkt $\hat{\boldsymbol{x}}$ erfülle die notwendigen Bedingungen 1. Ordnung für das Teilproblem* (B.5) *auf $\hat{\mathcal{W}}$. Weiter seien die Gradienten $\boldsymbol{a}_i, i \in$*

$\hat{\mathcal{W}}$, linear unabhängig (LICQ) und es gebe einen Index $j \in \mathcal{W}$ mit $\hat{\lambda}_j < 0$. Es sei $\boldsymbol{p}$ die Lösung vom Teilproblem (B.5) *ohne die Restriktion j, d.h.*

$$\begin{aligned} \min_{\boldsymbol{p}} \quad & \frac{1}{2}\boldsymbol{p}^T G \boldsymbol{p} + (G\hat{\boldsymbol{x}} + \boldsymbol{c})^T \boldsymbol{p}\,, \\ \text{s.t.} \quad & \boldsymbol{a}_i^T \boldsymbol{p} = 0\,, \quad \forall\, i \in \hat{\mathcal{W}} \setminus \{j\}\,. \end{aligned}$$

Dann ist p eine zulässige Richtung für die Nebenbedingung j, d.h. $\boldsymbol{a}_j^T \boldsymbol{p} \geq 0$. Weiterhin gilt sogar $\boldsymbol{a}_j^T \boldsymbol{p} > 0$ und p ist eine Abstiegsrichtung für q, wenn $\boldsymbol{p}$ die hinreichenden Bedingungen 2. Ordnung erfüllt.

Da wir zeigen können, dass der erzielte Abstieg für q durch das Weglassen einer Nebenbedingung mit negativem Lagrange-Multiplikator λ_i proportional zu $|\lambda_i|$ ist, eliminieren wir gerade die Restriktion mit kleinstem Langrange-Multiplikator. Es kann allerdings sein, dass der folgende zu berechnende Schritt $\boldsymbol{p}$ aufgrund einer blocking constraint kurz ist, wodurch nicht garantiert ist, dass q den größtmöglichen Abstieg erfährt.

B.3 Algorithmus

Algorithm B.3.1 Active-Set-Methode für konvexe quadratische Probleme

Gegeben sei ein zulässiger Startpunkt $\boldsymbol{x}_0$ für (B.1) und definiere $\mathcal{W}_0$ z.B. mit allen aktiven Restriktionen bzgl. $\boldsymbol{x}_0$.

for k = 0,1,2,... **do**
  Löse (B.5) zur Berechnung von $\boldsymbol{p}_k$;
  if $\boldsymbol{p}_k = \mathbf{0}$ **then**
    Berechne die Lagrange-Multiplikatoren mittels (2.5a)
      und setze $\hat{\mathcal{W}} = \mathcal{W}_k$;
    if $\hat{\lambda}_i \geq 0 \,\forall\, i \in \hat{\mathcal{W}} \cap \mathcal{I}$ **then**
      stop mit der Lösung $\boldsymbol{x}^* = \hat{\boldsymbol{x}}$;
    else
      $j \leftarrow \arg\min_{j \in \mathcal{W}_k \cap \mathcal{I}} \hat{\lambda}_j$;
      $\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k, \mathcal{W}_{k+1} \leftarrow W_k \setminus \{j\}$;
    end if
  else ($\boldsymbol{p}_k \neq \mathbf{0}$)
    Berechne α_k mit (B.7);
    $\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$;
    if $\alpha_k < 1$ (blocking constraint existiert) **then**
      Bestimme blocking constraint j und setze $\mathcal{W}_{k+1} \leftarrow \mathcal{W}_k \cup \{j\}$
    else
      $\mathcal{W}_{k+1} \leftarrow \mathcal{W}_k$
    end if
  end if
end for

Anhang C

Quellcode

C.1 Implementierung des Fehlerschätzers für das Hindernisproblem

Index