

Technische Universität München  
Zentrum Mathematik

# Formoptimierung mit Variationsungleichungen als Nebenbedingung und eine Anwendung in der Kieferchirurgie

Jan Christoph Wehrstedt

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Bernd Simeon  
Prüfer der Dissertation: 1. Univ.-Prof. Dr. Stefan Ulbrich,  
Technische Universität Darmstadt  
2. Univ.-Prof. Dr. Martin Brokate  
3. Univ. Prof. Dr. Ekkehard Sachs,  
Universität Trier  
(schriftliche Beurteilung)

Die Dissertation wurde am 05.02.2007 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 03.07.2007 angenommen.



---

## Zusammenfassung

Diese Arbeit behandelt Formoptimierungsprobleme für Variationsungleichungen im Fall von Kontaktproblemen zwischen zwei elastischen Körpern ohne Reibung. Nach Untersuchung der Variationsungleichung auf Existenz und Eindeutigkeit der Lösung wird ein semiglattes Newton-Verfahren als schneller Löser vorgestellt. Dieser beruht auf einer regularisierten Gleichungsformulierung der Optimalitätsbedingungen, da die zueinander komplementären Variablen auf dem Kontaktrand  $\Gamma_C$  definiert sind.

Eine anschließende Untersuchung zeigt, dass die Lösung der Variationsungleichung Lipschitz-stetig vom Gebiet abhängt. Nun wird ein Formoptimierungsproblem behandelt, bei dem das zuvor behandelte Kontaktproblem eine Nebenbedingung bildet. Zur Lösung wird ein Kalkül, der auf die Lösung eines adjungierten Problems zurückgeht, zur Bestimmung von Subgradienten bezüglich der Formparameter vorgestellt, da ein differenzierbarer Zusammenhang im Allgemeinen nicht vorliegt.

Der Hauptteil der Arbeit besteht aus einem inexakten Bundle-Trust-Region-Verfahren zur Lösung des Formoptimierungsproblems. Die Lösung des Kontaktproblems und die des hierzu adjungierten Problems erfolgt iterativ. Dabei steuert das Bundle-Verfahren in jedem Schritt die Genauigkeit der Lösung des iterativen Löser in Abhängigkeit von den Genauigkeiten der aktuellen Iterierten der äußeren Optimierung. Das Verfahren verhält sich bei exakter Rechnung wie bekannte Verfahren und ist unabhängig von den verwendeten iterativen Lösern.

Die Diskussion einer Anwendung dieses Verfahrens in der Kieferchirurgie rundet die Arbeit ab.

## Abstract

We are dealing with a class of shape optimization problems, which have contact between two elastic bodies as a constraint. After discussing existence and uniqueness we are introducing a regularisation of the optimality conditions of the contact problem. This is necessary because of a complementarity on the contact boundary  $\Gamma_C$ . A semismooth Newton-Method is used to solve the contact problem.

After proving a Lipschitz continuous dependence between the solution of the contact problem and the shape, we are dealing with shape optimization problems with contact problems as a constraint. We are presenting an adjoint based method to compute subgradients, because there is usually no differential dependence on the shape parameters.

An inexact-bundle-trust-region-method is the main part of this thesis. We are using iterative solvers to find a solution of the contact problem and the adjoint problem. The accuracy of these iterative solvers is controlled by the bundle method in dependence of the exactness of the outer iteration. This method is independent of the iterative solvers used and behaves like well known methods if we are using exact solvers for the contact problem and the adjoint problem.

Using this method we discuss an application of it in the context of oral and maxillofacial surgery.



# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>1</b>
<b>2</b>	<b>Elastizität und Kontaktprobleme</b>	<b>5</b>
2.1	Lineare Elastizitätstheorie . . . . .	5
2.2	Lineare Kontaktprobleme . . . . .	8
2.3	Schwache Formulierung des Kontaktproblems . . . . .	11
2.4	Existenz der Lösung des Kontaktproblems . . . . .	13
<b>3</b>	<b>Semiglatte Newton-Verfahren</b>	<b>19</b>
3.1	Semiglatte Newton-Verfahren in $\mathbb{R}^n$ . . . . .	19
3.2	Newton-Verfahren für Komplementaritätsprobleme . . . . .	22
3.3	Semiglatte Newton-Verfahren für Kontaktprobleme . . . . .	24
3.4	Schnelle Löser für Kontaktprobleme . . . . .	28
3.5	Fehler bei iterativen Lösern . . . . .	33
<b>4</b>	<b>Formoptimierung</b>	<b>39</b>
4.1	Gebietsverformungen . . . . .	40
4.2	Formoptimierungsprobleme . . . . .	43
4.3	Berechnung von Formableitungen . . . . .	52
4.4	Beschreibung des Steuerungsraums . . . . .	64
4.5	Berechnung inexakter Formableitungen . . . . .	66
4.6	Semiglattheit der Formfunktion . . . . .	69
<b>5</b>	<b>Nichtglatte Optimierungsverfahren</b>	<b>71</b>
5.1	Bundle-Trust-Region-Verfahren für konvexe Funktionen . . . . .	71
5.2	Bundle-Trust-Region-Verfahren für nichtkonvexe Funktionen . . . . .	77
5.3	Inexakte Bundle-Trust-Region-Verfahren . . . . .	81

---

5.4	Ein neues inexaktes Bundle-Trust-Region-Verfahren . . . . .	81
5.4.1	Iterative Löser . . . . .	83
5.4.2	Die Optimierung . . . . .	84
5.4.3	Der Quasi-Normalschritt . . . . .	84
5.4.4	Berechnung eines genäherten Gradienten . . . . .	85
5.4.5	Bestimmung des Quasi-Tangentialschrittes . . . . .	86
5.4.6	Der Korrekturschritt . . . . .	88
5.4.7	Die innere Iteration . . . . .	88
5.4.8	Die Liniensuche . . . . .	91
5.4.9	Die äußere Iteration . . . . .	94
5.5	Konvergenz des neuen inexakten Bundle-Trust-Region-Verfahrens . .	95
5.6	Erweiterungen . . . . .	107
5.6.1	Reset des Bundles . . . . .	107
5.6.2	Inexaktes Bundle-Trust-Region-Verfahren und lineare Neben- bedingungen . . . . .	108
<b>6</b>	<b>Eine medizinische Anwendung und numerische Tests</b>	<b>111</b>
6.1	Eine medizinische Anwendung . . . . .	111
6.2	Ein Modellproblem . . . . .	112
6.2.1	Das Kontaktproblem . . . . .	113
6.2.2	Formoptimierung . . . . .	116
6.3	Das eigentliche Problem . . . . .	121
6.3.1	Das Kontaktproblem . . . . .	122
6.3.2	Formoptimierung . . . . .	123
<b>7</b>	<b>Ausblick</b>	<b>127</b>
<b>A</b>	<b>Lineare Algebra</b>	<b>129</b>
<b>B</b>	<b>Nichtglatte Analysis</b>	<b>131</b>
<b>C</b>	<b>Hilfsmittel aus der Funktionalanalysis</b>	<b>133</b>

# Abbildungsverzeichnis

2.1	Kontaktbedingung . . . . .	8
4.1	Teilgebiete der Strukturoptimierung . . . . .	39
4.2	Beispiel eines nichtglatten Zielfunktional . . . . .	45
5.1	Schnittebenenmodell für konvexe Funktionen . . . . .	73
5.2	Innere Iteration für konvexe Funktionen . . . . .	75
5.3	Innere Iteration für nichtkonvexe Funktionen . . . . .	80
5.4	Innere Iteration für inexaktes Bundle-Verfahren . . . . .	89
6.1	Drei Staples in situ eines Polyurethan-Kunststoffkiefers (sogenannter Sawbone). . . . .	112
6.2	Modellproblem . . . . .	113
6.3	Eine Kieferfraktur . . . . .	121



# Kapitel 1

## Einführung

Die vorliegende Arbeit behandelt Formoptimierungsprobleme für Variationsungleichungen im Fall von Kontaktproblemen zwischen zwei elastischen Körpern ohne Reibung.

In Kapitel Zwei geben wir eine kurze Einführung in die lineare Elastizitätstheorie. Darauf aufbauend modellieren wir dreidimensionale Kontaktprobleme. Hier gehen wir vergleichsweise wie in [KO88] vor. Zuerst wird die Variationsungleichung

$$\int_{\Omega} \sum_{i,j=1}^3 \tau_{ij}(u) \cdot \epsilon_{ij}(v - u) \, dx \geq \int_{\Omega} \langle f, v - u \rangle_3 \, dx + \int_{\Gamma_P} \langle P, v - u \rangle_3 \, ds \quad \forall v \in \mathbb{K},$$
$$\mathbb{K} := \{v \in H^1(\Omega), Bv - d \leq 0 \text{ auf } \Gamma_C, \gamma_D v = 0 \text{ auf } \Gamma_D\},$$

die die schwache Lösung des Kontaktproblems beschreibt, hergeleitet. Dabei steht links in der Ungleichung die Bilinearform der linearen Elastizität und rechts die Linearform der Last. Der Operator  $B$  ist ein Spuroperator, der die Relativverschiebung von Punkten auf den potentiellen Kontaktflächen im verschobenen Zustand misst, und die Spaltfunktion  $v$  gibt den Relativabstand im unverformten Fall an.

Anschließend werden Existenz und Eindeutigkeit der Lösung  $u \in \mathbb{K}$  diskutiert. Hierzu wenden wir klassische Methoden wie den Satz von Lax-Milgram an. Im Anschluß daran wird wie in [UU] eine Regularisierung durchgeführt. Diese ist für den Löser des Kontaktproblems nötig, weil zwei zueinander komplementäre Variablen auf dem Kontaktrand  $\Gamma_C$  auftreten und somit in  $H^{1/2}(\Gamma_C)$  beziehungsweise  $(H^{1/2}(\Gamma_C))^*$  definiert sind. Die Elemente von  $H^{1/2}(\Gamma_C)$  sind aufgrund der Einbettung in  $L^2(\Gamma_C)$  fast überall auswertbar, was für die Elemente in  $(H^{1/2}(\Gamma_C))^*$  nicht der Fall ist. Die Regularisierung führt die zueinander komplementären Funktionen in  $L^2(\Gamma_C)$ -Funktionen über. Das Optimalitätssystem mit dem Regularisierungsparameter  $\gamma > 0$  besitzt dann die Form

$$\begin{aligned} Au - F + B^*q &= 0 \text{ in } H^{-1}(\Omega) \\ Bu - d + z - \gamma q &= 0 \text{ in } L^2(\Gamma_C) \\ q \geq 0, z \geq 0, qz &= 0 \text{ fast überall in } L^2(\Gamma_C). \end{aligned}$$

Dabei sind  $A$ ,  $F$  und  $B$  die Riesz-Repräsentanten der Bilinearform und der Linearform sowie der obige Spuroperator. Mit  $u$ ,  $q$  und  $z$  bezeichnen wir die Verschiebung, die Kontaktkraft und den Spaltabstand des verformten Systems. Im Gegensatz zur ursprünglichen Formulierung können  $q$  und  $z$  jetzt punktweise fast überall ausgewertet werden.

In Kapitel Drei stellen wir semiglatte Newton-Verfahren vor. Dies erfolgt zunächst im Endlichdimensionalen. Wir führen dazu den Begriff der semiglatten Funktion ein [OKZ98]. Wir erarbeiten dann den Anwendungsvorteil dieser Verfahren bei der Bestimmung von Lösungen für nichtlineare Komplementaritätsprobleme (NCP). Anschließend stellen wir fest, dass die Optimalitätsbedingungen des Kontaktproblems aus Kapitel zwei, welche ein System aus Gleichungen und Ungleichungen sind, ein unendlich dimensionales nichtlineares Komplementaritätsproblem bilden. Wie formen es in eine Gleichung um. Dafür verwenden wir die Minimumsfunktion, die aus den Ungleichungen und Gleichungen in den Optimalitätsbedingungen eine Gleichung der Bauart

$$C(u, q, z) = \begin{pmatrix} Au - F + B^*q \\ Bu - d + z - \gamma q \\ \min(\gamma q, z) \end{pmatrix} = 0$$

konstruiert. Wir weisen nach, dass dieses System semiglatt im Sinne von [Ul02a] ist. Deshalb können wir ein semiglattes Newton-Verfahren als Löser benutzen, um eine Lösung dieser Gleichungsformulierung der Optimalitätsbedingungen zu finden. Wir wenden dieses sowohl im Funktionenraum als auch im diskretisierten Fall an.

Bei dem semiglatten Newton-Verfahren handelt es sich um einen iterativen Löser, mit dem in der Regel nur eine Näherungslösung bestimmt wird. Wir weisen im Fall strikter Komplementarität einen Zusammenhang zwischen exakter und Näherungslösung in Abhängigkeit des Residuums des Newton-Verfahrens nach.

In Kapitel Vier beschäftigen wir uns mit der Lösung der Variationsungleichung in Abhängigkeit des Gebietes. Hierzu betrachten wir ein Grundgebiet  $\Omega$  und eine Familie von Abbildungen  $T_p$  mit  $p \in P$ . Dabei beschreibt  $P$  die Indexmenge der zulässigen Formen. Wir betrachten nun solche Gebiete

$$\Omega(p) = \{y \in \mathbb{R}^n \mid y = x + T_p(x)\},$$

welche durch Verformung aus diesem Grundgebiet hervorgehen. Dabei ist die Abbildung  $x \mapsto x + T_p(x)$  bijektiv und die Abbildung und die Umkehrabbildung hinreichend glatt. Nach einer Parametrisierung des Gebietes wird gezeigt, dass nach Auswahl von geeigneten Räumen ein Lipschitz-stetiger Zusammenhang

$$\|u(p_1) - u(p_2)\| \leq L \|T_{p_1} - T_{p_2}\|$$

zwischen der Lösung  $u(p)$  des Kontaktproblems auf dem Gebiet  $\Omega(p)$  und den Formparametern vorliegt. Nun wird ein Formoptimierungsproblem aufgestellt, bei dem das zuvor behandelte Kontaktproblem eine Nebenbedingung bildet. Das Problem wird auf die Existenz von Lösungen untersucht. Anschließend wird ein Kalkül, der

---

auf die Lösung eines adjungierten Problems zurückgeht, zur Bestimmung von Formableitungen in Differenzierbarkeitspunkten vorgestellt. Nachdem die Abbildung der Gebiete auf die Lösung Lipschitz-stetig ist, liegt im Allgemeinen kein differenzierbarer Zusammenhang vor. Allerdings wissen wir, dass nach dem Satz von Rademacher Differenzierbarkeit fast überall vorliegt. Ansonsten kann zu jedem Funktionswert immerhin ein Subgradient berechnet werden. Für den Fall, dass bei der Lösung des Kontaktproblems strikte Komplementarität herrscht, können wir die Ableitung exakt berechnen.

Auch hier verwenden wir zur Lösung der adjungierten Gleichung wieder einen iterativen Löser, weshalb wir im weiteren Verlauf des Kapitels einen Zusammenhang zwischen der Näherungslösung eines Subgradienten und dem korrespondierenden exakten Subgradienten und dem Residuum des iterativen Löser nachweisen.

Zum Schluss zeigen wir, dass beim Übergang zur diskretisierten Formulierung ein semiglatte Zusammenhang zwischen den Formparametern und dem zu optimierenden Zielfunktional vorliegt.

Kapitel Fünf bildet den Hauptteil der Arbeit. Hier stellen wir ein neues inexaktes Bundle-Trust-Region-Verfahren zur Lösung des Formoptimierungsproblems vor. Bundle-Verfahren haben sich in der Vergangenheit zur Behandlung von nichtglatten Optimierungsproblemen etabliert. Wir stellen eine Erweiterung des nichtkonvexen Bundle-Trust-Region-Verfahrens aus [Sch89] vor. Nachdem wir die Lösung des Kontaktproblems und die des adjungierten Problems iterativ bestimmen, werden die Funktionswerte und die Subgradienten fehlerbehaftet berechnet. Dabei steuert das neu vorgestellte Verfahren in jedem Schritt die Fehlerschranken der Lösung der iterativen Löser in Abhängigkeit von den Genauigkeiten der aktuellen Iterierten der äußeren Optimierung. Das Verfahren verhält sich bei exakter Rechnung wie bekannte Verfahren und ist unabhängig von den verwendeten iterativen Lösern.

Wir weisen die Konvergenz des Verfahrens nach. Anschließend erläutern wir Variationen des Algorithmus, indem wir ein inexaktes Bundle-Verfahren mit linearen Nebenbedingungen an die Formparameter und Resets des Bundles zur Reduzierung des Speicher- und Rechenbedarfs als Erweiterungen vorstellen.

Abschließend wenden wir in Kapitel Sechs das Optimierungsverfahren auf ein Problem aus der Kieferchirurgie an. Dazu betrachten wir einen ebenen Bruch eines Unterkiefers. Dieser wird nach Relokalisierung der Fragmente mit Klammern fixiert. Wir berechnen eine Anordnung der Klammern, um eine möglichst stabile Fixierung zu gewährleisten. In diesem Zusammenhang diskutieren wir die numerischen Ergebnisse.

Wir schließen die Arbeit mit einem Ausblick auf Erweiterungen der Arbeit in Kapitel Sieben ab.

An dieser Stelle möchte ich Martin Brokate für die Möglichkeit danken, an seinem Lehrstuhl zu promovieren. Ihm ist es über die Dauer der Promotion immer wieder gelungen, alles Mögliche zu unternehmen, dass mein Vertrag verlängert wurde. Des Weiteren bin ich ihm dafür zu großem Dank verpflichtet, dass er mir ermöglicht hat,

durch die Teilnahme an zahlreichen Konferenzen und Workshops namhafte Wissenschaftler auf dem Gebiet der Optimierung von partiellen Differentialgleichungen kennenzulernen.

Des Weiteren möchte ich meinen Kollegen an der Lehr- und Forschungseinheit M6 danken. Besonders möchte ich Christian Clason und Robert Schreittmiller hervorheben, die mir jederzeit für Computerfragen und fachliche Diskussionen zur Verfügung standen.

Ich danke der Arbeitsgruppe Optimierung an der TU Darmstadt für ihre Gastfreundschaft bei meinen dortigen Forschungsaufenthalten und meinen aktuellen und ehemaligen Kollegen am Zentrum Mathematik für die wunderbare Zusammenarbeit.

Für ihre Informationen über Kieferfrakturen bin ich Robert Sader und Heinrich Schieferstein zu Dank verpflichtet.

Karin Mautner, Kathrin Ruf, Robert Schreittmiller und meiner Ehefrau Christine bin ich dankbar, dass sie die Arbeit Korrektur gelesen haben, und Thomas Stolte für die Unterstützung beim Drucken.

Ich danke den vielen Studenten, die ich in den letzten vier Jahren betreuen durfte. Ich habe die Lehraufgaben immer als einen angenehmen Kontrast zu den Forschungsaufgaben empfunden und hoffe, dass sie von der Zusammenarbeit mit mir genau so profitiert haben wie ich.

Außerdem möchte ich mich bei meiner Frau Christine für die Unterstützung in den letzten vier Jahren und besonders im letzten halben Jahr bedanken. Außerdem möchte ich an dieser Stelle meinen Sohn Frederic, meine Brüder Jens und Henning und meine Eltern grüßen und Letzteren für ihr Interesse an meiner Arbeit danken.

Schließlich möchte ich den noch unbekanntem Gutachtern und dem Prüfungsvorsitzenden meiner Dissertation danken, dass sie die Zeit und die Mühe für das Lesen dieser Arbeit und den Prüfungsprozess aufbringen.

Der größte Dank geht an Stefan Ulbrich. Viele Ideen in dieser Arbeit gehen direkt auf Diskussionen mit ihm zurück. Er hat sich immer Zeit für mich genommen und jeder Besuch in Darmstadt war so fruchtbar, dass die Entfernung zwischen den Orten vernachlässigbar ist.

# Kapitel 2

## Elastizität und Kontaktprobleme

In diesem Kapitel geben wir eine kurze Einführung in die lineare Elastizitätstheorie. Dabei werden die Begriffe und Zusammenhänge nur so weit erläutert, wie sie für diese Arbeit verwendet werden. Weitere Informationen findet man zum Beispiel in [Bra02, Cia88, LL66]. Anschließend beschäftigen wir uns mit dreidimensionalen Kontaktproblemen. Dabei werden die wichtigsten Zusammenhänge erklärt und Eigenschaften zitiert beziehungsweise hergeleitet.

### 2.1 Lineare Elastizitätstheorie

Wir lehnen uns hier an die Darstellung in [HM03] an. Ein Körper  $\Omega$ , der durch ein beschränktes Gebiet  $\Omega \subset \mathbb{R}^3$ , die sogenannte Referenzkonfiguration, beschrieben wird, werde durch Volumenkräfte  $f = (f_1, f_2, f_3)$  und Oberflächenkräfte  $P = (P_1, P_2, P_3)$  in einem Teilgebiet  $\Gamma_P$  des Randes  $\partial\Omega$  belastet. Wir nehmen an, dass der im Folgenden immer hinreichend glatt ist. Der restliche Teil des Randes  $\Gamma_D := \partial\Omega \setminus \Gamma_P$  mit positivem Maß wird fixiert. Ziel der Elastizitätstheorie ist es, einen Gleichgewichtszustand in  $\Omega$  zu finden. Dieses Gleichgewicht wird durch einen symmetrischen Spannungstensor

$$\tau : \mathbb{R} \rightarrow \mathbb{R}^{3 \times 3}, \quad \tau_{ij}(x) = \tau_{ji}(x) \quad (2.1)$$

charakterisiert, der im Gleichgewicht mit  $f$  und  $P$  die Differentialgleichung

$$\begin{aligned} \sum_{j=1}^3 \frac{\partial \tau_{ij}}{\partial x_j} + f_i &= 0 \quad \text{in } \Omega, \quad i = 1, 2, 3 \\ \sum_{j=1}^3 \tau_{ij} \nu_j &= P_i \quad \text{auf } \Gamma_P, \quad i = 1, 2, 3, \end{aligned} \quad (2.2)$$

erfüllt. Dabei bezeichnet  $x_i$  die  $i$ -te Komponente von  $x \in \Omega$  und  $\nu = (\nu_1, \nu_2, \nu_3)^T$  die äußere Flächennormale von  $\Gamma_P$  in  $x$ .

Wir interessieren uns für den aktuellen Zustand  $\phi(x)$  jedes Punktes  $x \in \bar{\Omega}$  der Referenzkonfiguration im verformten Gebiet. Wir setzen

$$\phi = Id + u$$

und bezeichnen  $u : \bar{\Omega} \rightarrow \mathbb{R}^3$  als Verschiebung. Ausgehend von den Verschiebungen interessiert man sich für die Verzerrung des Körpers. Sie gibt die Längenänderung eines Linienelements im deformierten Fall an und ist durch

$$E : \Omega \rightarrow \mathbb{R}^{3 \times 3}, \quad x \mapsto E_{ij} := \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) + \frac{1}{2} \sum_{k=1}^3 \frac{\partial u_i}{\partial x_k} \frac{\partial u_j}{\partial x_k}, \quad i, j = 1, 2, 3 \quad (2.3)$$

definiert. In der linearen Elastizitätstheorie werden die quadratischen Terme vernachlässigt, und es ergibt sich als Näherung die symmetrische Ableitung

$$\epsilon_{ij} := \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right), \quad i, j = 1, 2, 3. \quad (2.4)$$

Wir kennen nun die Verzerrungen und die Spannungen eines elastischen Systems. Um gegebenen Kräften Deformationen und Spannungen zuzuordnen, müssen die Gleichungen gekoppelt werden. Diese Kopplung ist materialabhängig.

Das einfachste Materialgesetz ist das lineare Hookesche Gesetz

$$\tau_{ij} = \sum_{k,l=1}^3 c_{ijkl} \cdot \epsilon_{kl}, \quad i, j = 1, 2, 3, \quad (2.5)$$

an deren Koeffizienten folgende Forderungen gestellt werden:

**Annahme 2.1.** Die Elastizitätskoeffizienten  $c_{ijkl}$  im Hookeschen Gesetz (2.5) erfüllen folgende Eigenschaften:

- (a) Für alle  $i, j, k, l = 1, 2, 3$  gilt  $c_{ijkl} \in L^\infty(\Omega)$ .
- (b) Für fast alle  $x \in \Omega$  gilt  $c_{ijkl}(x) = c_{jikl}(x) = c_{klij}(x)$ .
- (c) Es existiert ein  $q > 0$  mit  $\sum_{i,j,k,l=1}^3 c_{ijkl}(x) \xi_{ij} \xi_{kl} \geq q \sum_{i,j=1}^3 \xi_{ij} \cdot \xi_{ij}$  für fast alle  $x \in \Omega, \xi \in \mathbb{R}^{3 \times 3}$ .

Hiermit lässt sich nun eine erste Charakterisierung der Lösung des Elastizitätsproblems aufstellen.

**Problem 2.2** (Lineares Elastizitätsproblem). Suche eine Verschiebung  $u$ , so dass (2.2) von  $\tau(u)$  erfüllt wird, wobei  $\tau(u)$  wie in (2.5) beschrieben von  $\epsilon(u)$  abhängt. Zusätzlich muss  $u$  die homogenen Randbedingungen

$$u_i(x) = 0 \quad \text{für alle } x \in \Gamma_D, \quad i = 1, 2, 3, \quad (2.6)$$

erfüllen.

Zur Herleitung der schwachen Formulierung verwenden wir die Greensche Formel

$$\int_{\Omega} \sum_{i,j=1}^3 \tau(u)_{ij} \epsilon(v)_{ij} dx = - \int_{\Omega} \sum_{i,j=1}^3 \frac{\partial \tau_{ij}(u)}{\partial x_j} v_i dx + \int_{\partial \Omega} \sum_{i,j=1}^3 \tau_{ij}(u) \nu_j v_i ds, \quad (2.7)$$

die für hinreichend glatte vektorwertige Funktionen  $v$  und matrixwertige Funktionen  $\tau$  auf einem hinreichend glatten Gebiet  $\Omega$  gilt.

Sei  $\mathbb{V}(\Omega)$  der Raum der zulässigen Verschiebungen

$$\mathbb{V}(\Omega) = \{v \in (H^1(\Omega))^3 : v|_{\Gamma_D} = 0\}. \quad (2.8)$$

Falls eine Lösung  $u$  des Elastizitätsproblems (2.2) existiert, kann man mit (2.2), (2.7) und der Definition von  $\mathbb{V}(\Omega)$  sehen, dass  $u$  die folgende Integralgleichung

$$\int_{\Omega} \sum_{i,j=1}^3 \tau(u)_{ij} \epsilon(v)_{ij} dx = \int_{\Omega} f v dx + \int_{\Gamma_P} P v ds \quad \forall v \in \mathbb{V}(\Omega), \quad (2.9)$$

erfüllt. Diese Identität werden wir als Basis für die schwache Formulierung benutzen. Die schwache Formulierung des Elastizitätsproblems lautet:

**Problem 2.3** (Schwache Formulierung des Elastizitätsproblems). *Finde  $u \in \mathbb{V}(\Omega)$  so, dass*

$$a(u, v) = F(v) \quad \forall v \in \mathbb{V}(\Omega) \quad (2.10)$$

mit der stetigen Linearform

$$F(v) := \langle F, v \rangle_{((H^1)^*, H^1)} := \int_{\Omega} f v dx + \int_{\Gamma_P} P v ds \quad (2.11)$$

und der stetigen elliptischen Bilinearform

$$a(u, v) := \langle Au, v \rangle_{((H^1)^*, H^1)} := \int_{\Omega} \sum_{i,j=1}^3 \tau(u)_{ij} \epsilon(v)_{ij} dx \quad (2.12)$$

gilt.

Dabei bezeichnet  $\langle \cdot, \cdot \rangle_{((H^1)^*, H^1)}$  die duale Paarung,  $F$  einen Riesz-Repräsentanten von  $F$ , und  $A$  den Differentialoperator des Elastizitätsproblems.

Weil  $u$  aus  $\mathbb{V}$  ist, kann  $u$  auf der Nullmenge  $\partial\Omega$  nicht ausgewertet werden. Deshalb ist das zweite Integral auf der rechten Seite als

$$\int_{\Gamma_P} P u ds = \int_{\Gamma_P} P \gamma_{\Gamma_P} u ds$$

zu interpretieren. Dabei bezeichnet  $\gamma_{\Gamma_P}$  den Spuroperator, der  $\mathbb{V}$  auf  $H^{1/2}(\Gamma_P)$  abbildet. Wir nehmen nun an, dass  $f \in (L^2(\Omega))^3$  und  $g \in (L^2(\Gamma_P))^3$  sowie  $\Gamma_D$  nicht leer und offen in  $\partial\Omega$  ist. Aus Annahme 2.1 folgt, dass

$$a(v, v) \geq q \int_{\Omega} \epsilon(v) : \epsilon(v) dx = q \|\epsilon(v)\|_{0,\Omega}^2 \quad \forall v \in \mathbb{V}(\Omega)$$

gilt. Dabei bezeichnet  $\|\cdot\|_{0,\Omega}$  für alle  $v \in \mathbb{V}(\Omega)$  die Norm des Raumes  $H^{0,2}(\Omega)$ , welcher  $L^2(\Omega)$  entspricht.

Die  $\mathbb{V}(\Omega)$ -Elliptizität, die wichtigste Voraussetzung des Satzes von Lax-Milgram C.1, folgt aus der Kornschen Ungleichung (C.2)

$$\exists q_0 > 0 : \|\epsilon(v)\|_{0,\Omega}^2 \geq q_0 \|v\|_{1,\Omega}^2, \quad \forall v \in \mathbb{V}(\Omega).$$

Somit hat das schwache Problem (2.10) eine eindeutige Lösung in  $\mathbb{V}(\Omega)$ .

## 2.2 Lineare Kontaktprobleme

Wird ein Körper  $\Omega$  verformt, kann der Fall auftreten, dass er durch die Verformungen einen weiteren Körper berührt oder der Körper sich an einigen Stellen selbst berührt. In diesen Fällen spricht man von Kontaktproblemen. In der Literatur unterscheidet man bei Kontaktproblemen, ob an der Kontaktfläche Reibung auftritt oder nicht. Eine ausführliche Behandlung von Kontaktproblemen findet man in [KO88]. Bei Kontaktproblemen treten in der Regel zwei Probleme auf: sowohl die Kontaktfläche als auch die daraus resultierenden Kontaktkräfte sind in der Regel a priori nicht bekannt, so dass diese im Laufe der Rechnung identifiziert werden müssen.

Wir beschreiben in diesem Abschnitt eine Modellierung des Kontaktproblems ohne Reibung. Dabei gehen wir analog zur zweidimensionalen Beschreibung in [KO88] vor. Zur Verdeutlichung beachte man auch Abbildung 2.1.

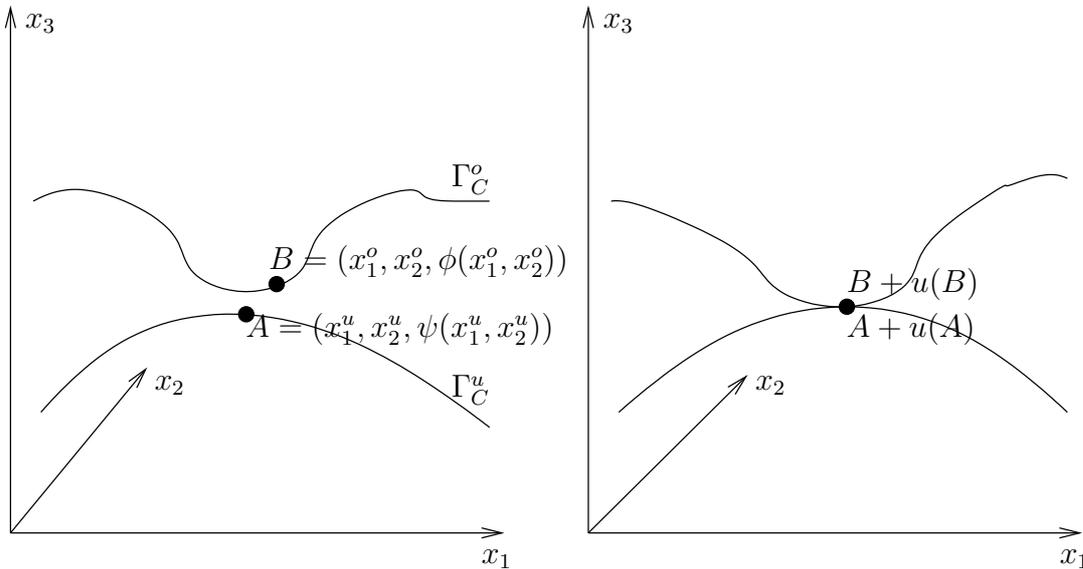


Abbildung 2.1: Die Abbildung zeigt die korrespondierenden Kontaktpunkte im be- und entlasteten Fall.

Wir nennen die mögliche obere Kontaktfläche  $\Gamma_C^o$  und die untere  $\Gamma_C^u$ . Wir nehmen an, dass sich die Kontaktflächen parametrisieren lassen zu

$$\begin{aligned} x_3 &= \phi(x_1, x_2) \quad \text{für } (x_1, x_2, x_3) \in \Gamma_C^o, \\ x_3 &= \psi(x_1, x_2) \quad \text{für } (x_1, x_2, x_3) \in \Gamma_C^u. \end{aligned}$$

Dabei sind  $\phi$  und  $\psi$  genügend glatte Funktionen. Den allgemeinen Fall kann man immer durch geeignete Koordinatentransformationen auf diesen Spezialfall zurückführen.

Sei nun  $x^o = (x_1^o, x_2^o, \phi(x_1^o, x_2^o))$  die Koordinate eines Punktes  $A$  auf der Kontaktfläche  $\Gamma_C^o$  und  $x^u = (x_1^u, x_2^u, \psi(x_1^u, x_2^u))$  der eindeutig bestimmte Punkt auf  $\Gamma_C^u$ , der

die Eigenschaften hat, dass ihre Bilder im deformierten Zustand die gleichen  $x_1$ - und  $x_2$ -Koordinaten haben. Also gelten punktweise die Kontaktbedingungen:

$$\begin{aligned} x_1^o + u_1(x_1^o, x_2^o, \phi(x_1^o, x_2^o)) &= x_1^u + u_1(x_1^u, x_2^u, \psi(x_1^u, x_2^u)), \\ x_2^o + u_2(x_1^o, x_2^o, \phi(x_1^o, x_2^o)) &= x_2^u + u_2(x_1^u, x_2^u, \psi(x_1^u, x_2^u)), \\ \phi(x_1^o, x_2^o) + u_3(x_1^o, x_2^o, \phi(x_1^o, x_2^o)) &\geq \psi(x_1^u, x_2^u) + u_3(x_1^u, x_2^u, \psi(x_1^u, x_2^u)). \end{aligned} \quad (2.13)$$

Dabei bezeichnet  $u(x) = (u_1(x), u_2(x), u_3(x))^T$  die Verschiebung des Punktes  $x \in \Omega$ . Unter der Annahme von kleinen Verschiebungen lassen sich die folgenden linearen Näherungen rechtfertigen: Für  $i = 1, 2$  gilt

$$\begin{aligned} u_i(x_1^u, x_2^u, \phi(x_1^u, x_2^u)) &= u_i(x_1^u - x_1^o + x_1^o, x_2^u - x_2^o + x_2^o, \phi(x_1^u, x_2^u) \\ &\quad - \phi(x_1^o, x_2^o) + \phi(x_1^o, x_2^o)) \\ &= u_i(x_1^o, x_2^o, \phi(x_1^o, x_2^o)) \\ &\quad + \frac{\partial u_i}{\partial x_1}(x_1^u - x_1^o) + \frac{\partial u_i}{\partial x_2}(x_2^u - x_2^o) \\ &\quad + \frac{\partial u_i}{\partial x_3} \left( \frac{\partial \phi}{\partial x_1}(x_1^u - x_1^o) + \frac{\partial \phi}{\partial x_2}(x_2^u - x_2^o) \right) \\ &\quad + \text{Terme höherer Ordnung} \\ &\approx u_i(x_1^o, x_2^o, \phi(x_1^o, x_2^o)). \end{aligned} \quad (2.14)$$

Weiter gilt in erster Näherung durch Linearisieren der Parametrisierung des Kontaktrandes:

$$\psi(x_1^u, x_2^u) = \psi(x_1^o, x_2^o) + \frac{\partial \psi}{\partial x_1}(x_1^o, x_2^o)(x_1^u - x_1^o) + \frac{\partial \psi}{\partial x_2}(x_1^o, x_2^o)(x_2^u - x_2^o).$$

Mit diesen Näherungen lässt sich die dritte Zeile der Kontaktbedingung (2.13) folgendermaßen umformen:

$$\begin{aligned} \phi(x_1^o, x_2^o) + u_3(x_1^o, x_2^o, \phi(x_1^o, x_2^o)) &\geq \psi(x_1^o, x_2^o) + \frac{\partial \psi}{\partial x_1}(x_1^o, x_2^o)(x_1^u - x_1^o) + \\ &\quad \frac{\partial \psi}{\partial x_2}(x_1^o, x_2^o)(x_2^u - x_2^o) + u_3(x_1^u, x_2^u, \psi(x_1^u, x_2^u)). \end{aligned} \quad (2.15)$$

Wir vereinfachen nun die Notation und schreiben  $u_i^u := u_i(x_1^u, x_2^u, \phi(x_1^u, x_2^u))$  und  $u_i^o := u_i(x_1^o, x_2^o, \phi(x_1^o, x_2^o))$  und so weiter. Damit wird (2.15) zu

$$\phi^o + u_3^o \geq \psi^o + \frac{\partial \psi^o}{\partial x_1}(x_1^u - x_1^o) + \frac{\partial \psi^o}{\partial x_2}(x_2^u - x_2^o) + u_3^u. \quad (2.16)$$

Wir bringen alle Terme auf die rechte Seite, verwenden (2.13), (2.14) und erhalten

$$\psi^o - \phi^o + \begin{pmatrix} \frac{\partial \psi^o}{\partial x_1} \\ \frac{\partial \psi^o}{\partial x_2} \\ -1 \end{pmatrix}^T \begin{pmatrix} u_1^R(x_1^o, x_2^o) \\ u_2^R(x_1^o, x_2^o) \\ u_3^R(x_1^o, x_2^o) \end{pmatrix} \leq 0 \quad (2.17)$$

mit dem relativen Verschiebungsvektor

$$u^R(x_1^o, x_2^o) := \begin{pmatrix} u_1(x_1^o, x_2^o, \psi(x_1^o, x_2^o)) - u_1(x_1^o, x_2^o, \phi(x_1^o, x_2^o)) \\ u_2(x_1^o, x_2^o, \psi(x_1^o, x_2^o)) - u_2(x_1^o, x_2^o, \phi(x_1^o, x_2^o)) \\ u_3(x_1^o, x_2^o, \psi(x_1^o, x_2^o)) - u_3(x_1^o, x_2^o, \phi(x_1^o, x_2^o)) \end{pmatrix}.$$

Wir führen den Einheitsnormalenvektor von  $\Gamma_C^u$

$$n = \frac{1}{\sqrt{1 + \left(\frac{\partial\psi}{\partial x}\right)^2 + \left(\frac{\partial\psi}{\partial y}\right)^2}} \begin{pmatrix} \frac{\partial\psi}{\partial x} \\ \frac{\partial\psi}{\partial y} \\ -1 \end{pmatrix}$$

ein und ersetzen (2.17) durch

$$n^T u^R - \frac{\phi^o - \psi^o}{\sqrt{1 + \left(\frac{\partial\psi}{\partial x}\right)^2 + \left(\frac{\partial\psi}{\partial y}\right)^2}} \leq 0. \quad (2.18)$$

Definieren wir nun die Spaltfunktion

$$d := \frac{\phi - \psi}{\sqrt{1 + \left(\frac{\partial\psi}{\partial x}\right)^2 + \left(\frac{\partial\psi}{\partial y}\right)^2}}$$

und den Normalenabstand der Kontaktpunkte  $u_n^R = n^T u^R$ , so erhält man aus (2.18) die Kontaktbedingung

$$u_n^R - d \leq 0 \quad \text{auf } \Gamma_C^o \text{ und } \Gamma_C^u. \quad (2.19)$$

Diese entspricht den Kontaktbedingungen des Signorini-Problems. Es gehen jedoch anstatt der absoluten die relativen Verschiebungen in die Formulierung ein.

Das Kontaktproblem lässt sich nun folgendermaßen charakterisieren:

**Problem 2.4** (Lineares Kontaktproblem). *Suche eine Verschiebung  $u$ , so dass (2.2) von  $\tau(u)$  erfüllt wird, wobei  $\tau(u)$  wie in (2.5) beschrieben von  $\epsilon(u)$  abhängt. Zusätzlich gelten für  $u$  die homogenen Randbedingungen*

$$u_i(x) = 0 \quad \text{für alle } x \in \Gamma_D, \quad i = 1, 2, 3 \quad (2.20)$$

und die Kontaktbedingung

$$u_n^R - d \leq 0 \quad \text{auf } \Gamma_C. \quad (2.21)$$

Bezeichnet  $\nu$  den Normalenvektor auf der oberen Kontaktfläche  $\Gamma_C$ , so gelten für die sogenannten Normalenspannungen

$$\tau_n(u) = \nu^T \tau(u) \nu \quad (2.22)$$

und die sogenannten Tangentialspannungen

$$\tau_t(u) = (\tau(u) - \tau_n(u) Id) \nu$$

auf dem möglichen oberen Kontaktrand  $\Gamma_C$

$$\tau_n(u) \leq 0, \quad \tau_n(u)(u_n^R - d) = 0 \quad \text{und} \quad \tau_t = 0. \quad (2.23)$$

**Bemerkung 2.5.** Für das Kontaktproblem 2.4 gilt:

- (a) In [Eck96] werden die Linearisierungsfehler bei der Kontaktbedingung für zweidimensionale Probleme angegeben.
- (b) Die Gleichung (2.23) lässt sich physikalisch folgendermaßen interpretieren: Berühren sich mögliche Kontaktpunkte nicht, so liegt an ihnen keine Normalspannung vor, wohingegen im Kontaktfall in jedem Kontaktpunkt eine Druckspannung senkrecht zur Fläche vorliegt, die eine Durchdringung vermeidet. Falls an der Kontaktfläche Reibung auftritt, dann liegen auch Tangentialspannungen vor.

## 2.3 Schwache Formulierung des Kontaktproblems

Wie wir bereits eine schwache Lösung des Elastizitätsproblems gesucht haben, werden wir uns nun mit der schwachen Formulierung des Kontaktproblems und ihrer Lösung befassen. Den Schlüssel zum Verständnis liefern Variationsungleichungen.

Wir führen zuerst den Raum  $\tilde{\mathbb{K}}$  der möglichen Verschiebungen

$$\tilde{\mathbb{K}} := \{v \in \mathbb{V} : v_n^R - d \leq 0 \text{ auf } \Gamma_C\} \quad (2.24)$$

ein. Dabei ist die Bedingung  $v_n^R - d \leq 0$  auf  $\Gamma_C$  zunächst wie im vorangegangenen Abschnitt punktweise zu verstehen. Im Laufe dieses Paragraphen wird sich zeigen, dass diese Auffassung nicht sinnvoll ist, da die schwache Lösung auf Sobolev-Räumen lebt und somit eine punktweise Definition keinen Sinn macht. Es ist leicht zu sehen, dass  $\tilde{\mathbb{K}}$  eine konvexe Teilmenge von  $\mathbb{V}$  ist.

Wir leiten nun einen schwachen Lösungsbegriff her. Sei dazu  $u$  eine klassische Lösung des Kontaktproblems. Dann gilt nach Definition des Standard-Skalarprodukts und der Definition von  $\epsilon$

$$\int_{\Omega} \sum_{i,j=1}^3 \tau_{ij}(u) \epsilon_{ij}(v-u) dx = \int_{\Omega} \sum_{i,j=1}^3 \tau_{ij}(u) \frac{\partial}{\partial x_i} (v_j - u_j) dx.$$

Wir nehmen nun an, dass  $\tau(u)$  und  $v-u$  die Voraussetzungen für die Greensche Formel (2.7) erfüllen und erhalten

$$\int_{\Omega} \sum_{i,j=1}^3 \tau_{ij}(u) \epsilon_{ij}(v-u) dx = \int_{\Omega} - \sum_{i,j=1}^3 \frac{\partial}{\partial x_i} \tau_{ij}(u) (v_j - u_j) dx + \int_{\partial\Omega} \sum_{i,j=1}^3 \tau(u) \nu_j (v_i - u_i) ds.$$

Wir zerlegen die Oberfläche  $\partial\Omega = \Gamma_D \cup \Gamma_P \cup \Gamma_C$  in die Teilflächen, in denen eine Dirichlet-Bedingung, die Oberflächenkraft  $P$  oder die Kontaktbedingung (2.21)

wirkt. Nachdem  $v_i$  und  $u_i$  auf  $\Gamma_D$  verschwinden und  $u$  eine Lösung des Kontaktproblems ist, folgt

$$\begin{aligned} \int_{\Omega} \sum_{i,j=1}^3 \tau_{ij}(u) \epsilon_{ij}(v-u) dx &= \int_{\Omega} \langle f, v-u \rangle_3 dx + \int_{\Gamma_P} \langle P, v-u \rangle_3 ds \\ &\quad + \int_{\Gamma_C} (v^R - u^R)^T \tau(u) \nu ds \quad \forall v \in \tilde{\mathbb{K}}, \end{aligned}$$

wobei  $\langle \cdot, \cdot \rangle_3$  das euklidische Skalarprodukt im  $\mathbb{R}^3$  ist. Zerlegt man die Oberflächenspannung wie in (2.22) und (2.23)

$$\tau(u) = \tau_n(u) \nu + \tau_t(u),$$

so erhält man auf  $\Gamma_C$

$$\begin{aligned} (v^R - u^R)^T \tau(u) \nu &= (v^R - u^R)^T (\tau_t(u) + \tau_n(u) \nu) \\ &\stackrel{(2.23)}{=} 0 + \tau_n(u) (v_n^R - u_n^R) \\ &= 0 + \tau_n(u) (v_n^R - u_n^R - d + d) \\ &\stackrel{(2.21)(2.23)}{=} \tau_n(u) (v_n^R - d) \geq 0. \end{aligned}$$

Also gilt für die Lösung  $u$  des Kontaktproblems die schwache Formulierung

$$\int_{\Omega} \sum_{i,j=1}^3 \tau_{ij}(u) \cdot \epsilon_{ij}(v-u) dx \geq \int_{\Omega} \langle f, v-u \rangle_3 dx + \int_{\Gamma_P} \langle P, v-u \rangle_3 ds \quad \forall v \in \tilde{\mathbb{K}}. \quad (2.25)$$

Hierbei handelt es sich um eine Variationsungleichung, die kompakt mit den Notationen (2.11), (2.12) folgendermaßen aufgeschrieben wird:

$$\text{Für die Lösung } u \text{ des Kontaktproblems gilt } a(u, v-u) \geq F(v-u) \quad \forall v \in \tilde{\mathbb{K}}.$$

Bislang sind wir bei der Herleitung der schwachen Formulierung davon ausgegangen, dass  $u$  eine Lösung des Kontaktproblems sei und  $u, v$  hinreichend glatt seien, so dass die Greenschen Formeln anwendbar sind. Nun fragen wir uns, wie Lösungen von (2.25), wenn sie existieren, mit Lösungen des Kontaktproblems 2.4 korrelieren, oder wie  $v_n^R$  oder

$$v_n^R - d \leq 0$$

zu interpretieren sind.

Da wir die schwache Lösung auf  $(H^1(\Omega))^3$  formuliert haben, sind die Elemente auf  $\Gamma_C$  und  $\Gamma_D$  nicht mehr punktweise auswertbar, da der Rand von  $\Omega$  eine Nullmenge im  $\mathbb{R}^3$  ist.

Falls der Rand von  $\Omega$  Lipschitz-stetig ist, existiert nach dem Spursatz [Eva98] ein Operator  $\gamma : (H^1(\Omega))^3 \rightarrow (L^2(\partial\Omega))^3$ , der linear und stetig ist. Außerdem gilt für  $\phi \in C(\bar{\Omega})$

$$\gamma\phi = \phi|_{\partial\Omega}.$$

Man nennt  $\gamma\phi$  die Spur von  $\phi$ , auch wenn  $\phi$  eine allgemeine Funktion in  $H^1(\Omega)$  ist.

Deshalb müssen wir die zunächst punktweisen Formulierungen  $v_{\Gamma_D} = 0$  in (2.8) und  $u_n^R - d \leq 0$  in (2.24) in  $H^{1/2}(\Gamma_D)$  beziehungsweise in  $H^{1/2}(\Gamma_C)$  auffassen.

**Bemerkung 2.6.** *Die Aussage  $a \geq 0$  in  $H^{1/2}(\Gamma)$  ist so zu verstehen, dass eine Folge positiver stetiger Funktionen  $a_n$  existiert, die in  $H^{1/2}(\Gamma)$  gegen  $a$  konvergiert.*

Wir suchen also eine Lösung der schwachen Formulierung des Kontaktproblems in

$$\mathbb{K} := \{v \in \mathbb{V} : Bv - d \leq 0 \text{ auf } \Gamma_C, \gamma_D v = 0 \text{ auf } \Gamma_D\}. \quad (2.26)$$

Dabei bezeichnet  $B : \mathbb{V} \rightarrow H := H^{1/2}(\Gamma_C)$  den Spuroperator  $v \mapsto \gamma_C(v_n^R)$ . Umgekehrt ist die Lösung der Variationsungleichung eine Lösung des Kontaktproblems im distributionellen Sinn, die Kontakt- und Randbedingungen sind bereits im Ansatzraum enthalten. Wählt man  $\phi \in C_0^\infty(\Omega)$ , und testet man in (2.25) mit  $v = u \pm \phi$ , so folgt

$$\int_{\Omega} \tau(u) : \epsilon(\phi) \, dx = \int_{\Omega} \langle f, \phi \rangle \, dx \quad \forall \phi \in C_0^\infty(\Omega).$$

Folglich gilt für die Lösung  $u \in \mathbb{K}$  der schwachen Formulierung des Kontaktproblems, falls sie existiert,

$$\sum_{j=1}^3 \frac{\partial \tau_{ij}}{\partial x_j} + f_i = 0 \text{ in } \Omega, \quad i = 1, 2, 3$$

im distributionellen Sinn.

Somit lässt sich die Lösung des Kontaktproblems als schwache Lösung von Problem 2.7 charakterisieren.

**Problem 2.7** (Schwache Formulierung des Kontaktproblems). *Finde  $u \in \mathbb{K}$  derart, dass  $u$  Lösung der Variationsungleichung*

$$a(u, v - u) \geq F(v - u) \quad \forall v \in \mathbb{K},$$

*mit der Bilinear- und Linearform aus Problem 2.3 ist.*

Falls die Lösung hinreichend glatt ist, stimmt sie mit der klassischen Lösung überein.

## 2.4 Existenz der Lösung des Kontaktproblems

Der konvexe Lösungsraum und die Stetigkeit der Bilinearform und der Linearform sowie die Elliptizität der Bilinearform garantieren mittels des Satzes von Lax-Milgram C.1 die eindeutige Lösung des Problems 2.7.

Wir stellen – ausgehend vom Satz von Lax-Milgram C.1 – eine weitere Formulierung zur Bestimmung der Lösung des Kontaktproblems auf, welche uns als Grundlage für die numerische Lösung dient. Wir können die Lösung der Variationsungleichung als Lösung eines Minimierungsproblems charakterisieren:

**Problem 2.8** (Kontaktproblem als Minimierungsaufgabe). *Wir suchen eine Funktion  $u \in \mathbb{K}$ , so dass das Funktional*

$$J(u) = \frac{1}{2}a(u, u) - F(u) \quad (2.27)$$

auf der Menge  $\mathbb{K}$  minimiert wird.

Das Problem lässt sich nach Einführung der Schlupfvariablen  $z \in H := H^{1/2}(\Gamma_C)$  in ein restringiertes Minimierungsproblem umschreiben:

$$\begin{aligned} \text{Minimiere} \quad & J(u) = \frac{1}{2}a(u, u) - F(u), \quad u \in \mathbb{V} \\ \text{u.d.N.} \quad & Bu - d + z = 0 \text{ in } H, \\ & z \geq 0 \text{ in } H. \end{aligned} \quad (2.28)$$

Die Sattelpunktformulierung von (2.28) lautet mit dem  $a(\cdot, \cdot)$  zugeordneten Operator  $A : \mathbb{V} \rightarrow \mathbb{V}^*$  und dem Lagrangemultiplikator  $q \in H^* = H^{1/2}(\Gamma_C)^*$

$$Au - F + B^*q = 0 \text{ in } \mathbb{V}^* \quad (2.29)$$

$$Bu - d + z = 0 \text{ in } H \quad (2.30)$$

$$q \in K^+, z \geq 0, \langle q, z \rangle_{H^*, H} = 0, \quad (2.31)$$

wobei

$$K^+ := \{q \in H^* : \langle q, z \rangle_{H^*, H} \geq 0 \forall z \in H, z \geq 0\}$$

der zu  $z \in H$  mit  $z \geq 0$  duale Kegel,  $B^*$  der zu  $B$  adjungierte Operator und  $\mathbb{V}^*$  der Dualraum von  $\mathbb{V}$  ist.

Wir wollen in Kapitel Drei Lösungsverfahren zur Lösung des Kontaktproblems entwickeln, die auf die Sattelpunktformulierung (2.29)–(2.31) aufbauen. Dabei ergibt sich das Problem, dass die Komplementaritätsbedingung (2.31) zwischen einem Element  $q \in H^{1/2}(\Gamma_C)^*$  und einer Funktion  $z \in H^{1/2}(\Gamma_C)$  nicht punktweise in der Form

$$q \geq 0, \quad z \geq 0, \quad qz = 0 \text{ fast überall}$$

geschrieben werden kann. Um dies zu umgehen, regularisieren wir die Gleichungen wie in [UU]. Weil eine punktweise Auswertung in  $L^2(\Gamma_C)$  möglich ist, und mit der Identifizierung  $L^2(\Gamma_C)^* = L^2(\Gamma_C)$

$$H = H^{1/2}(\Gamma_C) \subset L^2(\Gamma_C) = L^2(\Gamma_C)^* \subset H^* = H^{1/2}(\Gamma_C)^*$$

gilt, werden wir die Optimalitätsbedingungen folgendermaßen regularisieren und erhalten mit dem Regularisierungsparameter  $\gamma > 0$

$$Au^\gamma - F + B^*q^\gamma = 0 \text{ in } \mathbb{V}^*, \quad (2.32)$$

$$Bu^\gamma - d + z^\gamma - \gamma q^\gamma = 0 \text{ in } L^2(\Gamma_C), \quad (2.33)$$

$$q^\gamma \geq 0, z^\gamma \geq 0, q^\gamma z^\gamma = 0 \text{ fast überall in } L^2(\Gamma_C). \quad (2.34)$$

Dabei stammt  $q^\gamma$  jetzt aus dem regularisierten Kegel  $K^+ \cap L^2(\Gamma_C)$ , was gleichbedeutend ist zu

$$K_2^+ := K^+ \cap L^2(\Gamma_C) = \{q \in L^2(\Gamma_C), q \geq 0\}. \quad (2.35)$$

Das regularisierte System hat ebenfalls eine eindeutige Lösung.

Die Operatoren in der Gleichung (2.32) erfüllen die Voraussetzungen des Satzes von Lax-Milgram C.1 und somit kann (2.32) eindeutig nach  $u^\gamma = A^{-1}(F - B^*q^\gamma)$  aufgelöst werden. Wir setzen dies in (2.33) ein und erhalten nach Multiplikation der zweiten Gleichung das vereinfachte System

$$BA^{-1}B^*q^\gamma + \gamma q^\gamma + d - BA^{-1}F - z^\gamma = 0 \text{ in } L^2(\Gamma_C), \quad (2.36)$$

$$q^\gamma \geq 0, z^\gamma \geq 0, q^\gamma z^\gamma = 0 \text{ fast überall in } L^2(\Gamma_C). \quad (2.37)$$

Im endlichdimensionalen Fall entspricht diese Reduzierung einer, die durch Verwendung von Schur-Komplementen erhalten wird.

Das System (2.36)–(2.37) ist gleichzeitig das Optimalitätssystem des regularisierten Dualproblems von (2.28)

$$\begin{aligned} \min_{q^\gamma \in H^*} \frac{1}{2}(q^\gamma, BA^{-1}B^*q^\gamma)_{L^2(\Gamma_C)} + \frac{\gamma}{2}(q^\gamma, q^\gamma)_{L^2(\Gamma_C)} + (q^\gamma, d - BA^{-1}F)_{L^2(\Gamma_C)} \\ \text{u. d. N. } q^\gamma \in K_2^+. \end{aligned} \quad (2.38)$$

Für  $\gamma = 0$  erhalten wir das zu (2.28) duale Problem, falls die Lösung in  $L^2(\Gamma_C)$  liegt. Wir weichen hier von der Standardnominatur ab. Normalerweise ist das duale Problem zu einem Minimierungsproblem ein Maximierungsproblem. Dieses geht aus (2.38) durch Multiplikation mit  $-1$  hervor.

Nach [UU] ist  $BA^{-1}B^*$  elliptisch, falls  $B$  surjektiv ist und nach dem Satz von Lax-Milgram C.1 ist dann auch (2.36)–(2.37) eindeutig lösbar.

Wir werden nun wie in [UU] die Abhängigkeit der Lösungen des regularisierten Problems von der exakten Lösung und des Regularisierungsparameters  $\gamma$  untersuchen. Zunächst zeigen wir unter der Annahme, dass die Lösung  $\bar{q}$  des nicht regularisierten Problems in  $L^2(\Gamma_C)$  liegt, die Konvergenz von  $q^\gamma$  gegen  $\bar{q}$  bezüglich der  $L^2(\Gamma_C)$ -Norm.

**Hilfssatz 2.9.** *Falls die Lösung des nicht regularisierten Problems  $\bar{q}$  in  $L^2(\Gamma_C)$  liegt, gilt*

$$\|q^\gamma\|_{L^2(\Gamma_C)} \leq \|\bar{q}\|_{L^2(\Gamma_C)}. \quad (2.39)$$

*Beweis:* Aufgrund der Minimaleigenschaften von  $q^\gamma$  und  $\bar{q}$  gilt

$$\begin{aligned}
J_\gamma(q^\gamma) &= \frac{1}{2} \langle q^\gamma, BA^{-1}Bq^\gamma \rangle_{L^2(\Gamma_C)} + \frac{\gamma}{2} (q^\gamma, q^\gamma)_{L^2(\Gamma_C)} + (q^\gamma, d - BA^{-1}F)_{L^2(\Gamma_C)} \\
&= J_0(q^\gamma) + \frac{\gamma}{2} (q^\gamma, q^\gamma)_{L^2(\Gamma_C)} \\
&\geq J_0(q^0) + \frac{\gamma}{2} (q^\gamma, q^\gamma)_{L^2(\Gamma_C)} \\
&= J_\gamma(q^0) - \frac{\gamma}{2} (q^0, q^0)_{L^2(\Gamma_C)} + \frac{\gamma}{2} (q^\gamma, q^\gamma)_{L^2(\Gamma_C)} \\
&\geq J_\gamma(q^\gamma) - \frac{\gamma}{2} (q^0, q^0)_{L^2(\Gamma_C)} + \frac{\gamma}{2} (q^\gamma, q^\gamma)_{L^2(\Gamma_C)}.
\end{aligned}$$

Also ist die Menge der regularisierten Lösungen in der  $L^2(\Gamma_C)$ -Norm beschränkt durch

$$\frac{\gamma}{2} (q^0, q^0)_{L^2(\Gamma_C)} \geq \frac{\gamma}{2} (q^\gamma, q^\gamma)_{L^2(\Gamma_C)}.$$

Wurzelziehen und Division durch  $\frac{\gamma}{2}$  liefern die Behauptung. ■

Wir untersuchen nun die Konvergenz von  $J_0(q^\gamma)$ .

**Hilfssatz 2.10.** *Falls die Lösung des nicht regularisierten Problems  $\bar{q}$  in  $L^2(\Gamma_C)$  liegt, dann gilt für die Lösungen des mit  $\gamma > 0$  regularisierten Problems  $q^\gamma \in L^2(\Gamma_C)$*

$$\lim_{\gamma \rightarrow 0, \gamma > 0} J_0(q^\gamma) = J_0(\bar{q}). \quad (2.40)$$

*Beweis:* Wir verwenden wieder die Minimalität der Lösungen der Optimierungsprobleme. Somit ist

$$\begin{aligned}
J_0(\bar{q}) &\leq J_0(q^\gamma) \\
&= J_\gamma(q^\gamma) - \frac{\gamma}{2} (q^\gamma, q^\gamma)_{L^2(\Gamma_C)} \\
&\leq J_\gamma(\bar{q}) - \frac{\gamma}{2} (q^\gamma, q^\gamma)_{L^2(\Gamma_C)} \\
&= J_0(\bar{q}) + \frac{\gamma}{2} (\bar{q}, \bar{q})_{L^2(\Gamma_C)} - \frac{\gamma}{2} (q^\gamma, q^\gamma)_{L^2(\Gamma_C)} \\
&\leq J_0(\bar{q}) + \frac{\gamma}{2} (\bar{q}, \bar{q})_{L^2(\Gamma_C)}.
\end{aligned} \quad (2.41)$$

Wenn wir nun überall  $J_0(\bar{q})$  subtrahieren, erhalten wir

$$0 \leq J_0(q^\gamma) - J_0(\bar{q}) \leq \frac{\gamma}{2} \|\bar{q}\|_{L^2(\Gamma_C)}^2.$$

■

**Satz 2.11.** *Falls die Lösung des nicht regularisierten Problems  $\bar{q}$  in  $L^2(\Gamma_C)$  liegt, dann gilt mit der Lösung  $q^\gamma$  des mit  $\gamma > 0$  regularisierten Problems*

$$\|\bar{q} - q^\gamma\|_{H^{-1/2}(\Gamma_C)^*} = o(\sqrt{\gamma}). \quad (2.42)$$

*Beweis:* Wir betrachten eine positive Nullfolge  $(\gamma_k)$ , dann ist die Folge  $(q^{\gamma_k})$  wegen (2.39) und  $q^{\gamma_k} \in K_2^+ \subset L^2(\Gamma_C)$  schwach Folgen-kompakt. Wir finden also eine gegen  $\hat{q} \in K_2^+$  konvergente Teilfolge, die wir wieder mit  $(q^{\gamma_k})$  bezeichnen.

Die Abbildung  $J_0 : L^2(\Gamma_C) \rightarrow \mathbb{R}$  ist stetig und konvex und somit schwach unterhalb stetig. Mit der Minimaleigenschaft von  $\bar{q}$  und (2.40) folgt

$$J_0(\bar{q}) \leq J_0(\hat{q}) \leq \limsup_{k \rightarrow \infty} J_0(q^{\gamma_k}) = \lim_{\gamma \rightarrow 0} J_0(q^\gamma) = J_0(\bar{q}). \quad (2.43)$$

Nachdem nun  $\bar{q}$  das eindeutig bestimmte Minimum des nicht regularisierten dualen Problems war, folgt  $\hat{q} = \bar{q}$ . Nachdem die Folge frei gewählt war, können wir nun schließen, dass auch  $q^\gamma$  schwach gegen  $\bar{q}$  konvergiert.

Die starke Konvergenz zeigen wir mit dem Argument, dass in Hilbert-Räumen Folgen, die schwach konvergieren und deren Normen konvergieren, stark konvergent sind.

Die Normabbildung  $y \in L^2(\Gamma_C) \mapsto \|y\|_{L^2(\Gamma_C)}$  ist ebenfalls konvex und stetig, folglich schwach unterhalb stetig. Somit folgt mit der schwachen Unterhalb-Stetigkeit und der Beschränktheit der  $\|q^\gamma\|_{L^2(\Omega)}$  nach Hilfssatz 2.9

$$\limsup_{\gamma \rightarrow 0, \gamma > 0} \|q^\gamma\|_{L^2(\Gamma_C)} \leq \|\bar{q}\|_{L^2(\Gamma_C)} \leq \liminf_{\gamma \rightarrow 0, \gamma > 0} \|q^\gamma\|_{L^2(\Gamma_C)}.$$

Somit folgt die Konvergenz der Norm, was die Konvergenz in der Norm impliziert.

Aus der Ungleichung (2.41) erhält man nach Subtraktion von  $J_0(\bar{q})$  für den Grenzwert  $\gamma \rightarrow 0, \gamma > 0$

$$0 \leq J_0(q^\gamma) - J_0(\bar{q}) \leq \frac{\gamma}{2} (\|\bar{q}\|_{L^2(\Gamma_C)} - \|q^\gamma\|_{L^2(\Gamma_C)}) = o(\gamma). \quad (2.44)$$

Wir nutzen wieder aus, dass  $\bar{q}$  das nicht regularisierte duale Problem löst. Damit gilt  $J_0(\bar{q}) \leq J_0(q^\gamma)$  für alle  $q^\gamma \in K_2^+ \subset K^+$ .

Somit folgt mit dem Gradienten bezüglich der  $L^2$  Norm

$$(\nabla J_0(\bar{q}), q^\gamma - \bar{q})_{L^2(\Gamma_C)} \geq 0. \quad (2.45)$$

Wenn wir  $J_0(q)$  um  $\bar{q}$  entwickeln, erhalten wir

$$\begin{aligned} J_0(q) &= J_0(\bar{q}) + (q - \bar{q}, BA^{-1}B^*\bar{q})_{L^2(\Gamma_C)} + (q - \bar{q}, d - BA^{-1}F)_{L^2(\Gamma_C)} \\ &\quad + \frac{1}{2} (q - \bar{q}, BA^{-1}B^*(q - \bar{q}))_{L^2(\Gamma_C)}. \end{aligned}$$

Somit ist mit  $\nabla J_0(\bar{q}) = BA^{-1}B^*\bar{q} + d - BA^{-1}F$  und (2.45)

$$\begin{aligned} J_0(q^\gamma) - J_0(\bar{q}) &= (q^\gamma - \bar{q}, BA^{-1}B^*\bar{q})_{L^2(\Gamma_C)} + (q^\gamma - \bar{q}, d - BA^{-1}F)_{L^2(\Gamma_C)} \\ &\quad + \frac{1}{2} (q^\gamma - \bar{q}, BA^{-1}B^*(q^\gamma - \bar{q}))_{L^2(\Gamma_C)} \\ &= (\nabla J_0(\bar{q}), q^\gamma - \bar{q}) + \frac{1}{2} (q^\gamma - \bar{q}, BA^{-1}B^*(q^\gamma - \bar{q}))_{L^2(\Gamma_C)} \\ &\geq \frac{1}{2} (q^\gamma - \bar{q}, BA^{-1}B^*(q - \bar{q}^\gamma))_{L^2(\Gamma_C)}. \end{aligned}$$

Zusammen mit der Elliptizität [UU] von  $BA^{-1}B^*$  folgt mit Elliptizitätskonstante  $\nu$  und (2.44)

$$\begin{aligned} \|q^\gamma - \bar{q}\|_{H^{-1/2}(\Gamma_C)} &\leq \nu \frac{1}{2} (q^\gamma - \bar{q}, BA^{-1}B^*(q - \bar{q}^\gamma))_{L^2(\Gamma_C)} \\ &\leq \nu (J_0(q_\alpha) - J_0(\bar{q})) \\ &\leq \nu \frac{\gamma}{2} (\|\bar{q}\|_{L^2(\Gamma_C)} - \|q^\gamma\|_{L^2(\Gamma_C)}) = o(\gamma). \end{aligned}$$

Wurzelziehen liefert die Behauptung. ■

**Folgerung 2.12.** *Für die Lösung des regularisierten Problems (2.32)–(2.34) gilt, falls  $\bar{q}$  in  $L^2(\Gamma_C)$  liegt,*

$$\|(u^\gamma, q^\gamma, z^\gamma) - (\bar{u}, \bar{q}, \bar{z})\|_{\mathbb{V} \times H^{-1/2}(\Gamma_C) \times L^2(\Gamma_C)} = o(\sqrt{\gamma}).$$

*Beweis:* Es ist

$$\begin{aligned} \|u^\gamma - \bar{u}\|_{\mathbb{V}} &= \|A^{-1}(F - B^*q^\gamma) - A^{-1}(F - B^*\bar{q})\|_{\mathbb{V}} \\ &\leq \|A^{-1}\|_{\mathbb{V}^*, \mathbb{V}} \|B^*(\bar{q} - q^\gamma)\|_{\mathbb{V}^*} \\ &\leq \|A^{-1}\|_{\mathbb{V}^*, \mathbb{V}} \|B^*\|_{H^{-1/2}(\Gamma_C), \mathbb{V}^*} \|\bar{q} - q^\gamma\|_{H^{-1/2}(\Gamma_C)} = o(\sqrt{\gamma}). \end{aligned}$$

Analog erhält man

$$\begin{aligned} \|\bar{z} - z^\gamma\|_{L^2(\Gamma_C)} &= \|d - Bu^\gamma - \gamma q^\gamma - d + B\bar{u}\|_{L^2(\Gamma_C)} \\ &\leq \|B(\bar{u} - u^\gamma)\|_{L^2(\Gamma_C)} + \gamma \|q^\gamma\|_{L^2(\Gamma_C)} \\ &\leq \|B\|_{\mathbb{V}, L^2(\Gamma_C)} \|(\bar{u} - u^\gamma)\|_{\mathbb{V}} + \sqrt{\gamma} \sqrt{\gamma} \|\bar{q}\|_{L^2(\Gamma_C)} = o(\sqrt{\gamma}). \end{aligned}$$

Dabei haben wir Hilfssatz 2.9 verwendet. ■

# Kapitel 3

## Semiglatte Newton-Verfahren

Zur Lösung des Kontaktproblems verwenden wir semiglatte Newton-Verfahren. Diese konvergieren superlinear und sind deshalb sehr gut zur Lösung des Kontaktproblems geeignet. Wir werden diese Klasse von Verfahren zunächst im Endlichdimensionalen vorstellen und den Begriff der Semiglattheit einführen. Danach werden wir dies für den unendlichdimensionalen Fall verallgemeinern.

Anschließend werden wir diese Verfahren auf unser Kontaktproblem anwenden und einen schnellen Löser vorstellen.

### 3.1 Semiglatte Newton-Verfahren in $\mathbb{R}^n$

Die Newton-artigen Verfahren zur Bestimmung der Nullstelle einer Funktion  $F$  basieren alle auf einer Verallgemeinerung des Newton-Verfahrens. Ihr Algorithmus lässt sich wie folgt beschreiben:

**Algorithmus 3.1** (Newton-artige Verfahren). *Sei  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  Lipschitzstetig.*

- (a) Wähle  $x_0 \in U$
- (b) Für  $k = 0, \dots$ 
  1. Falls  $F(x_k) = 0$ , stopp.
  2. Sonst wähle geeignetes  $M_k \in \mathbb{R}^{n \times n}$ .  
Löse  $M_k s_k = -F(x_k)$   
Setze  $x_{k+1} = x_k + s_k$ .

**Bemerkung 3.2.**

- (a) Das Verfahren ist durchführbar, solange  $x_k \in U$  und  $M_k$  invertierbar ist.
- (b) Beim gewöhnlichen Newton-Verfahren verwendet man  $M_k = F'(x_k)$ .

Im Folgenden wählen wir  $M_k$  aus einer nichtleeren mengenwertigen Abbildung

$$M_k \in M_F(x_k) \quad \forall k \in \mathbb{N}_0,$$

mit

$$M_F : U \subset \mathbb{R}^n \rightrightarrows \mathbb{R}^{n \times n},$$

das heißt

$$M_F(x) \subset \mathbb{R}^{n \times n}, \quad M_F(x) \neq \emptyset \quad \forall x \in U.$$

Der folgende Satz beschreibt die Konvergenz von Algorithmus 3.1:

**Satz 3.3** (Konvergenz des Newton-Verfahrens). *Sei  $\bar{x}$  eine Lösung von  $F(x) = 0$  und  $M_F$  eine mengenwertige Abbildung mit nichtleerem Bild. Weiter gebe es ein  $\eta > 0$ , so dass für alle  $x \in U$ ,  $\|x - \bar{x}\| < \eta$ , jedes  $M \in M_F(x)$  invertierbar ist.*

(a) *Gilt*

$$\sup_{M \in M_F(\bar{x}+s)} \|M^{-1} (F(\bar{x} + s) - F(\bar{x}) - Ms)\| = o(\|s\|) \text{ für } \|s\| \rightarrow 0, \quad (3.1)$$

*so gibt es ein  $\delta > 0$ , so dass für alle Startwerte  $x_0 \in U$  mit  $\|x_0 - \bar{x}\| < \delta$  der Algorithmus entweder terminiert, oder die Folge  $(x_k)$   $q$ -superlinear gegen  $\bar{x}$  konvergiert.*

(b) *Gilt*

$$\sup_{M \in M_F(\bar{x}+s)} \|M^{-1} (F(\bar{x} + s) - F(\bar{x}) - Ms)\| = O(\|s\|^2) \text{ für } \|s\| \rightarrow 0, \quad (3.2)$$

*so gibt es ein  $\delta > 0$ , so dass für alle Startwerte  $x_0 \in U$  mit  $\|x_0 - \bar{x}\| < \delta$  der Algorithmus entweder terminiert, oder die Folge  $(x_k)$   $q$ -quadratisch gegen  $\bar{x}$  konvergiert.*

*Beweis:* [Ulb02b] ■

Wir werden die Forderungen (3.1) und (3.2) wegen

$$\|M^{-1} (F(\bar{x} + s) - F(\bar{x}) - Ms)\| \leq \|M^{-1}\| \|F(\bar{x} + s) - F(\bar{x}) - Ms\|$$

durch die zwei zusammengenommen etwas stärkeren Bedingungen ersetzen, welche wir in den Definitionen 3.4 und 3.5 vorstellen.

**Definition 3.4** (Regularitätsbedingung). *Eine nichtleere mengenwertige Abbildung  $M_F : U \subset \mathbb{R}^n \rightrightarrows \mathbb{R}^{n \times n}$  erfüllt die Regularitätsbedingung in  $\bar{x}$ , falls es ein  $\delta > 0$  und  $C > 0$  gibt mit*

$$M \text{ nicht singular, } \|M^{-1}\| < C \quad \forall M \in M_F(x), \forall \|x - \bar{x}\| \leq \delta, x \in U.$$

**Definition 3.5** (Approximationseigenschaft). Sei  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  eine Lipschitz-stetige Abbildung,  $M_F : U \rightrightarrows \mathbb{R}^{n \times n}$  eine nichtleere mengenwertige Abbildung und  $\bar{x} \in U$ .

(a) Wir nennen  $M_F$  eine punktbasierete Approximation von  $F$  in  $\bar{x}$ , falls

$$\sup_{M \in M_F(\bar{x}+s)} \|F(\bar{x} + s) - F(\bar{x}) - Ms\| = o(\|s\|) \quad \text{für } \|s\| \rightarrow 0 \quad (3.3)$$

erfüllt ist.

(b) Wir nennen  $M_F$  eine punktbasierete Approximation der Ordnung 1 von  $F$  in  $\bar{x}$ , falls

$$\sup_{M \in M_F(\bar{x}+s)} \|F(\bar{x} + s) - F(\bar{x}) - Ms\| = O(\|s\|^2) \quad \text{für } \|s\| \rightarrow 0$$

gilt. In diesem Fall sprechen wir auch von einer quadratischen Approximationseigenschaft.

Wir können nun schließen, dass das Newton-Verfahren 3.1  $q$ -superlinear gegen die Nullstelle  $\bar{x}$  konvergiert, wenn der Startwert hinreichend nahe an  $\bar{x}$  liegt, die mengenwertige Abbildung  $M_F$  die Regularitätsbedingung erfüllt und eine punktweise Approximation von  $F$  in  $\bar{x}$  ist. Falls eine punktbasierete Approximation der Ordnung 1 vorliegt, konvergiert das Newton-Verfahren  $q$ -quadratisch.

Wir verwenden für die mengenwertige Abbildung das Clarkesche Subdifferential (siehe B.2) und bezeichnen es mit  $\partial^{Cl}$ .

Wir definieren nun die Klasse der semiglaten Funktionen. Das sind die Funktionen, für die das oben vorgestellte Newton-Verfahren konvergiert.

**Definition 3.6.** Eine Funktion  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  heißt semiglatt in  $x \in U$ , falls  $F$  in einer Umgebung von  $x$  Lipschitz-stetig ist und der Grenzwert

$$\lim_{\substack{V \in \partial^{Cl} F(x+th') \\ h' \rightarrow h, t \searrow 0}} Vh', \quad (3.4)$$

für alle  $h \in \mathbb{R}^n$  existiert.

Wir fassen für uns wichtige Eigenschaften von semiglaten Funktionen zusammen [Ul02b]:

**Satz 3.7** (Eigenschaften semiglatte Funktionen). Sei  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ , dann sind folgende Aussagen äquivalent:

(i)  $F$  ist semiglatt in  $x \in U$ .

(ii)  $\max_{M \in \partial F(x+s)} \|Ms - F'(x; s)\|_2 = o(\|s\|_2)$  für  $\|s\|_2 \rightarrow 0$ . Dabei bezeichnet  $F'(x; s)$  die Richtungsableitung von  $F$  in  $x$  in Richtung  $s$ .

(iii)  $F$  ist richtungsdifferenzierbar in  $x$  und es gilt

$$\max_{M \in \partial F(x+s)} \|F(x+s) - F(x) - Ms\|_2 = o(\|s\|_2) \text{ für } \|s\|_2 \rightarrow 0.$$

**Bemerkung 3.8.** Wir sind nun in der Lage zu zeigen, dass das verallgemeinerte Newton-Verfahren für einen Startwert  $x_0$  hinreichend nahe der Lösung  $\bar{x}$  konvergiert, wenn  $0 \notin \partial^{Cl} F(\bar{x})$  gilt und  $F$  semiglatt ist, denn die Regularität folgt aus der Oberhalb-Stetigkeit des Subdifferentials und die Approximationseigenschaft gilt nach Satz 3.7 (iii).

Die folgenden Bemerkungen gehen auf [Cla90] zurück:

**Bemerkung 3.9.**

- (a) Zur Klasse der semiglaten Funktionen gehören stückweise  $C^1$ -Funktionen und konvexe Funktionen.
- (b) Die Komposition, die Summe und das Produkt semiglaten Funktionen ist wiederum semiglatt.
- (c) Im Falle der Kompositionen semiglaten Funktionen gilt die Kettenregel im Allgemeinen nicht. Wohl aber gilt für  $f, g$  semiglatt

$$\partial^{Cl} f(g(x)) \subset \text{conv}(\partial^{Cl} f(g(x)) \partial^{Cl} g(x)).$$

## 3.2 Newton-Verfahren für Komplementaritätsprobleme

Semiglatte Newton-Verfahren eignen sich wegen ihrer Struktur besonders zur Lösung von Komplementaritätsproblemen. Zuerst definieren wir nichtlineare Komplementaritätsprobleme, kurz NCP:

**Definition 3.10.** Suche zu einer gegebenen Funktion  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  eine Lösung  $x_* \in \mathbb{R}^n$  des Systems von Gleichungen und Ungleichungen

$$x \geq 0, F(x) \geq 0, x^T F(x) = 0. \quad (3.5)$$

**Bemerkung 3.11.**

- (a) Die Bedingung (3.5) ist äquivalent zur komponentenweisen Komplementarität

$$x_i \geq 0, F_i(x) \geq 0, x_i F_i(x) = 0, \text{ für } 1 \leq i \leq n.$$

(b) *Beispiele für Komplementaritätsprobleme sind die KKT-Bedingungen bei endlichdimensionalen Optimierungsproblemen mit Ungleichungsnebenbedingungen, diskretisierte Hindernisprobleme oder die numerische Behandlung von Optionen vom amerikanischen Typ [WDH93]. Ebenso lassen sich in bestimmten Fällen Nash-Gleichgewichte auf Komplementaritätsprobleme zurückführen [GK02]. Beispiele für unendlichdimensionale Komplementaritätsprobleme, die wir später behandeln, treten bei den ursprünglichen Formulierungen der obigen diskretisierten Beispiele auf.*

Die verallgemeinerten Newton-Verfahren erlauben uns, nur Gleichungssysteme zu lösen, weshalb wir nun versuchen werden, die Ungleichungsbedingung (3.5) in eine Gleichung umzuformen. Den Schlüssel hierzu liefern NCP-Funktionen.

**Definition 3.12.** *Eine Funktion  $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$  heißt NCP-Funktion, falls*

$$\phi(a, b) = 0 \Leftrightarrow a \geq 0, b \geq 0, ab = 0$$

*gilt.*

Beispiele für NCP-Funktionen sind die

- (a) Minimums Funktion  $\phi(a, b) = \min(a, b)$ ,
- (b) Fischer-Burmeister-Funktion  $\phi(a, b) = a + b - \sqrt{a^2 + b^2}$ .

Diese Funktionen haben die Eigenschaft, dass sie alle nicht global differenzierbar sind. Man kann allgemein zeigen, dass die Ableitung einer differenzierbaren NCP-Funktion im Nullpunkt verschwindet. Somit sind differenzierbare NCP-Funktionen für ein Newton-Verfahren nicht geeignet, da die Regularitätsbedingung in der Nullstelle verletzt ist. Allerdings kann die Komplementaritätsbedingung mit Hilfe der eben erwähnten nicht differenzierbaren Funktionen umgeformt werden.

**Hilfssatz 3.13.** *Ein Vektor  $x_* \in \mathbb{R}^n$  ist genau dann Lösung des Komplementaritätsproblems (3.5), wenn mit einer NCP-Funktion  $\phi$  der Vektor  $x_*$  das nichtlineare Gleichungssystem  $\Phi(x) = 0$ , mit*

$$\Phi(x) = \begin{pmatrix} \phi(x_1, F_1(x)) \\ \phi(x_2, F_2(x)) \\ \vdots \\ \phi(x_n, F_n(x)) \end{pmatrix} \quad (3.6)$$

*löst.*

*Beweis:* Die Behauptung folgt direkt aus der Definition der NCP-Funktion. ■

**Satz 3.14.** *Sei  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  stetig differenzierbar und  $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$  eine der oben angegebenen NCP-Funktionen, dann gilt:*

- (a) Die Funktion  $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$  ist semiglatte.
- (b) Die Abbildung  $\Phi : U \rightarrow \mathbb{R}^n$ , mit  $\Phi$  definiert wie in (3.6), ist ebenfalls semiglatte.
- (c) Ist  $\bar{x}$  Lösung des Komplementaritätsproblems (3.5) und  $0 \neq \partial^{C^1}\Phi(\bar{x})$ , so gibt es eine Umgebung von  $\bar{x}$ , in der das semiglatte Newton-Verfahren gegen  $\bar{x}$  konvergiert.

*Beweis:* Die Aussage (a) kann man nachrechnen. Die Aussage (b) lässt sich beweisen, indem man zeigt, dass die Abbildung  $H : x \mapsto (x_i, F_i(x))$  für  $1 \leq i \leq n$  differenzierbar ist. Der Rest folgt dann mit der Kettenregel. Der dritte Teil folgt direkt aus Bemerkung 3.8 und Hilfssatz 3.13. ■

### 3.3 Semiglatte Newton-Verfahren für Kontaktprobleme

In diesem Abschnitt stellen wir ein semiglatte Newton-Verfahren in einem Funktionenraum vor. Dabei lehnen wir uns an die Arbeiten [Ul02a, Ul03] an.

Wir betrachten die Optimalitätsbedingungen des regularisierten Kontaktproblems (2.32)–(2.34). Diese formulieren wir wieder wie in Abschnitt 3.2 mit einer NCP-Funktion in ein nichtglattes Gleichungssystem

$$\hat{C}(u, q, z) = \begin{pmatrix} Au - F + B^*q \\ Bu - d + z - \gamma q \\ \min(\gamma q, z) \end{pmatrix} = 0_{H^{-1}(\Omega) \times L^2(\Gamma) \times L^2(\Gamma)} \quad (3.7)$$

um, welches wir wiederum mit einem nichtglatten Newton-Verfahren lösen möchten. Der Index an der Null gibt die Räume an, aus denen jeweils das Nullelement ausgewählt wird. Die ersten zwei Gleichungen sind affin linear, und somit differenzierbar. Also müssen wir für die letzte Gleichung eine geeignete semiglatte Formulierung finden, damit sich das System mit einem Newton-Verfahren lösen lässt.

**Definition 3.15** (Semiglattheit in Banach-Räumen). *Seien  $G : X \rightarrow Y$  ein stetiger Operator zwischen zwei unendlich dimensionalen Banach-Räumen und eine mengenwertige Abbildung  $\partial G : X \rightrightarrows L(X, Y)$  mit nichtleerem Bild gegeben, so nennen wir  $G$*

- (a)  $\partial G$ -semiglatte in  $x \in X$ , falls

$$\sup_{M \in \partial G(x+d)} \|G(x+d) - G(x) - Md\|_Y = o(\|d\|_X) \text{ für } \|d\|_X \rightarrow 0, \quad (3.8)$$

*gilt, und*

(b)  $\partial G$ -semiglatte der Ordnung  $\alpha > 0$  in  $x \in X$ , falls

$$\sup_{M \in \partial G(x+d)} \|G(x+d) - G(x) - Md\|_Y = O(\|d\|_X^{1+\alpha}) \text{ für } \|d\|_X \rightarrow 0, \quad (3.9)$$

gilt.

Wir nennen die mengenwertige Abbildung  $\partial G$  verallgemeinerte Ableitung.

**Bemerkung 3.16.**

- (a) Die Semiglattheit in Banach-Räumen unterscheidet sich von der Semiglattheit, wie sie in Satz (3.7) für endlichdimensionale Abbildungen charakterisiert wird, dadurch, dass auf die Existenz der Richtungsableitung verzichtet wurde. Für Newton-Verfahren spielt dieser Unterschied keine Rolle, da dort nur die Approximationseigenschaft verwendet wird. Allerdings werden wir im Kapitel Fünf die Eigenschaft der Richtungs-differenzierbarkeit benötigen, so dass wir Semiglatte im Endlichdimensionalen anders definieren als im Fall unendlichdimensionaler Banach-Räume.
- (b) Zur Unterscheidung bezeichnen wir das Clarkesche Differential im Endlichdimensionalen mit  $\partial^{Cl}$  und das verallgemeinerte Differential im Funktionenraum mit  $\partial$ .
- (c) Wie im Endlichdimensionalen sind stetig differenzierbare Abbildungen  $F$  semiglatte und für ihr verallgemeinertes Differential gilt  $\partial F = \{F'\}$  mit der Fréchet-Ableitung  $F'$ .
- (d) Produkt, Summe und Komposition semiglatte Funktionen sind wiederum semiglatte.

Wir lösen in (3.7) die zweite Gleichung nach  $z$  auf und setzen dies in die letzte Zeile ein. Somit erhalten wir das äquivalente System

$$\begin{aligned} \hat{C}(u, q, z) &= 0_{H^{-1}(\Omega) \times L^2(\Gamma) \times L^2(\Gamma)}, \\ \begin{pmatrix} Au - F + B^*q \\ Bu - d + z - \gamma q \\ \min(\gamma q, z) \end{pmatrix} &= 0_{H^{-1}(\Omega) \times L^2(\Gamma) \times L^2(\Gamma)}, \\ \begin{pmatrix} Au - F + B^*q \\ Bu - d + z - \gamma q \\ \gamma q + \min(0, -Bu + d) \end{pmatrix} &= 0_{H^{-1}(\Omega) \times L^2(\Gamma) \times L^2(\Gamma)}. \end{aligned} \quad (3.10)$$

Wir setzen

$$C(u, q, z) := \begin{pmatrix} Au - F + B^*q \\ Bu - d + z - \gamma q \\ \gamma q + \min(0, -Bu + d) \end{pmatrix}.$$

Nun untersuchen wir, ob die Abbildung  $u \mapsto \min(0, -Bu + d)$  semiglatt ist. Hierzu stellen wir, wie in [Ul02a, Ul03] beschrieben, für allgemeine Superpositionsoperatoren

$$\Psi : (L^p(\Omega))^m \rightarrow L^q(\Omega), \quad w = \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix} \mapsto \psi(w_1, w_2, \dots, w_m), \quad (3.11)$$

mit  $1 \leq p \leq q \leq \infty$ , für eine Lipschitz-stetige, semiglatte Funktion  $\psi : \mathbb{R}^m \rightarrow \mathbb{R}$  ein geeignetes Differential auf. Das Differential soll eine möglichst große Verwandtschaft mit dem Clarkeschen Differential aus B.2 haben und im Falle einer Diskretisierung der Gleichung mit ihm zusammenfallen. Deshalb führen wir eine fast überall punktweise Anwendung des Clarkeschen Differentials als Differential für den Superpositionsoperator ein.

**Definition 3.17** (Verallgemeinertes Differential für Superpositionsoperatoren). *Sei  $\psi : \mathbb{R}^m \rightarrow \mathbb{R}$  lokal Lipschitz-stetig und semiglatt. Dann wird für den Superpositionsoperator (3.11) das Differential durch*

$$\begin{aligned} \partial\Psi : L^p(\Omega)^m &\rightrightarrows L((L^p(\Omega))^m, L^q(\Omega)), \\ \partial\Psi(w) &= \{M, Mw = g^T w, g \in (L^\infty(\Omega))^m, g(x) \in \partial^{Cl}\psi(w(x)) \text{ für fast alle } x \in \Omega\} \end{aligned}$$

definiert.

Nachdem  $\Psi$  stetig von  $(L^p(\Omega))^m$  in  $L^q(\Omega)$  abbildet, würden wir erwarten, dass der Operator auch in diesem Falle semiglatt ist. Dies ist im Allgemeinen nicht der Fall, wie einfache Beispiele zeigen [HPUU05].

Die hier betrachteten Superpositionsoperatoren haben die Bauart  $\psi(G)$  mit einer stetig Fréchet-differenzierbaren Abbildung  $G$ . Für diese Bauart ist es möglich, Semiglattheit zu zeigen. Das Differential ergibt sich aus Anwendung der Kettenregel.

Wir zitieren folgenden Satz [Ul02a, Ul03]:

**Satz 3.18** (Semiglattheit von Superpositionsoperatoren). *Sei  $\Omega \subset \mathbb{R}^n$  messbar mit endlichem Maß,  $\psi : \mathbb{R}^m \rightarrow \mathbb{R}$  lokal Lipschitz-stetig und semiglatt,  $Y$  ein Banach-Raum und  $1 \leq q < p \leq \infty$ . Unter der Annahme, dass  $G : Y \rightarrow (L^q(\Omega))^m$  Fréchet-differenzierbar und  $G : Y \rightarrow L^p(\Omega)$  lokal Lipschitz-stetig ist, ist der Operator*

$$\Psi_G : Y \rightarrow L^q(\Omega), \quad \Psi_G(y)(x) = \psi(G(y)(x)),$$

$\partial\Psi_G$ -semiglatt mit dem Differential

$$\begin{aligned} \partial\Psi_G & : Y \rightrightarrows L(Y, L^q(\Omega)), \\ \partial\Psi_G(y) & = \{M, Mw = g^T(G'(y)v), g \in (L^\infty(\Omega))^m, \\ & \quad g(x) \in \partial^{Cl}\psi(G(y)(x)) \text{ für fast alle } x \in \Omega\}. \end{aligned}$$

**Bemerkung 3.19.** *Unter weiteren Annahmen kann man auch Semiglattheit in einer höheren Ordnung zeigen. Für unsere Zwecke begnügen wir uns mit einfacher Semiglattheit.*

Wir zeigen nun, dass die Abbildung  $C(u, q, z)$  semiglatt ist.

**Satz 3.20.** *Sei  $C : H^1(\Omega) \times L^2(\Gamma_C) \times L^2(\Gamma_C) \rightarrow H^{-1}(\Omega) \times L^2(\Gamma_C) \times L^2(\Gamma_C)$  gegeben durch*

$$C(u, q, z) = \begin{pmatrix} Au - F + B^*q \\ Bu - d + z - \gamma q \\ \gamma q + \min(0, -Bu + d) \end{pmatrix}, \quad (3.12)$$

mit  $A : H^1(\Omega) \rightarrow H^{-1}(\Omega)$ ,  $B : H^1 \rightarrow H^{1/2}(\Gamma_C)$ ,  $d \in L^\infty(\Gamma_C)$  und  $\gamma > 0$ . Dann ist  $C(u, q, z)$  semiglatt mit der verallgemeinerten Ableitung

$$\partial C = \begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -D(u)B & \gamma Id & 0 \end{pmatrix}. \quad (3.13)$$

Dabei bezeichnet  $D$  die verallgemeinerte Ableitung der Minimumsfunktion:

$$D(u) = \begin{cases} 1 & -Bu + d < 0, \\ 0 & -Bu + d > 0, \\ [0, 1] & \text{sonst.} \end{cases} \quad (3.14)$$

**Bemerkung 3.21.** *Die verallgemeinerte Ableitung  $\partial C$  ist nur im Fall strikter Komplementarität bis auf eine Nullmenge eine einelementige Menge.*

*Beweis:* Im Folgenden verwenden wir  $Y := H^1(\Omega) \times L^2(\Gamma_C) \times L^2(\Gamma_C)$  versehen mit der Produkttopologie.

Wir setzen  $G(u) := (0, -Bu + d)$  und bezeichnen mit  $D$  die Elemente des verallgemeinerten Differentials  $\partial \min(\cdot, \cdot)$ . Dann ist  $G : H^1(\Omega) \rightarrow H^{1/2}(\Gamma_C) \times H^{1/2}(\Gamma_C)$  affin linear. Mit  $H^{1/2}(\Gamma_C) \subset L^2(\Gamma_C)$  folgt, dass  $G$  Fréchet-differenzierbar ist. Ferner ist  $H^{1/2}(\Gamma_C)$  kompakt eingebettet in  $L^q(\Gamma_C)$  mit  $q \leq \frac{2}{1-\frac{1}{n}}$ , wobei  $n$  die Raumdimension ist [Ada75, ES98]. In unserem zweidimensionalen Fall also für  $q \leq 4$ . Somit erfüllt  $G$  die Anforderungen von Satz 3.18.

Sei  $\bar{y} = (\bar{u}, \bar{q}, \bar{z})^T \in Y$  fest. Dann gilt für  $\delta = (\delta_u, \delta_q, \delta_z) \in Y$  mit beliebigem  $C_y \in \partial C(\bar{y} + \delta)$

$$C(\bar{y} + \delta) - C(\bar{y}) - C_y \delta = \begin{pmatrix} 0 \\ 0 \\ \min(G(\bar{u} + \delta_u)) - \min(G(\bar{u})) + D(G(\bar{u} + \delta_u))B\delta_u \end{pmatrix}.$$

Dabei bezeichnet  $D$  das zu  $C_y$  gehörende Element aus dem in (3.14) definierten verallgemeinerten Differential. Da  $G$  die Anforderungen von Satz 3.18 erfüllt und die Minimumsfunktion semiglatt ist, folgt

$$\|C(\bar{y} + \delta) - C(\bar{y}) - C_y \delta\|_Y \leq o(\|\delta\|_Y).$$

Nachdem  $C_y$  aus der verallgemeinerten Ableitung beliebig war, können wir links zum Supremum über die verallgemeinerten Ableitungen übergehen und erhalten die  $\partial C$ -Semiglattheit der NCP-Formulierung des Variationsproblems. ■

Somit können wir das Kontaktproblem lösen, indem wir folgendes semiglatte Newton-Verfahren anwenden:

**Algorithmus 3.22** (Semiglattes Newton-Verfahren für Kontaktprobleme). *Gegeben sei  $C : U \subset Y := H^1(\Omega) \times L^2(\Gamma_C) \times L^2(\Gamma_C) \rightarrow H^{-1}(\Omega) \times L^2(\Gamma_C) \times L^2(\Gamma_C)$  semiglatt.*

(a) Wähle  $y_0 \in U$ .

(b)  $k = 0, \dots$

1. Falls  $C(y_k) = 0$ : Stopp.

2. Sonst wähle geeignetes  $M_k \in \partial C(y_k)$  mit  $\partial C$  wie in (3.13).

3. Löse

$$M_k s_k = -C(y_k). \quad (3.15)$$

4. Setze  $y_{k+1} = y_k + s_k$ .

5. Gehe zu 1.

In [Ulb02a] ist bewiesen, dass unter den gegebenen Voraussetzungen das semiglatte Newton-Verfahren superlinear gegen die Lösung des Komplementaritätsproblems konvergiert, falls die verallgemeinerten Ableitungen regulär, das heißt ihre Inversen beschränkt sind, und die Startwerte hinreichend nahe an der Lösung liegen.

**Bemerkung 3.23.** *Das hier vorgestellte Newton-Verfahren lässt sich auch über eine primal duale aktive Mengen-Strategie herleiten [HIK02].*

## 3.4 Schnelle Löser für Kontaktprobleme

In diesem Abschnitt beschäftigen wir uns mit der numerischen Bestimmung der Lösung des Kontaktproblems. Wir werden dazu einen iterativen Löser vorstellen und diesen charakterisieren. Bei der Lösung von Variationsproblemen in Funktionenräumen treten immer zwei Arten von Fehlern auf. Zum einen liegt ein Diskretisierungsfehler vor, der den Fehler zwischen der Lösung des Problems im Funktionenraum und der endlichdimensionalen Approximation beschreibt. Hierzu gibt es zahlreiche Forschungsergebnisse, auf die wir an dieser Stelle nicht weiter eingehen möchten. Eine Zusammenfassung findet man zum Beispiel in [Glo84, KO88].

Eine zweite Fehlerquelle resultiert in der nicht exakten Lösung des diskretisierten Problems. Dies liegt entweder daran, dass die exakte Lösung nicht in endlich vielen Schritten berechenbar ist oder daran, dass ein schneller Löser verwendet wird, der das Problem nur bis zu einer bestimmten Genauigkeit löst.

Wir diskretisieren den Funktionenraum nun mittels linearer Finiten Elemente im Ort und stückweise konstanter Elemente auf dem Kontaktrand. Die Genauigkeit, die durch die Diskretisierung auftritt, lässt sich über den Diskretisierungsparameter

$h > 0$  steuern. Wenn wir nun das in Abschnitt 3.3 entwickelte semiglatte Newton-Verfahren auf das diskretisierte Problem anwenden, können wir das so erhaltene Verfahren als eine Störung des unendlichdimensionalen Verfahrens auffassen.

Abstrakt gesehen lösen wir  $C(y) = 0$  mit dem semiglaten Newton-Verfahren (SNV) aus Algorithmus 3.22. Im diskretisierten Fall hingegen lösen wir  $C_h(y) = 0$  mit einem diskretisierten semiglaten Newton-Verfahren (SNV <sub>$h$</sub> ).

Das Newton-Verfahren gehört zur Klasse der iterativen Verfahren. In jedem Schritt muss im Endlichdimensionalen ein lineares Gleichungssystem und im Unendlichdimensionalen eine lineare Operatorgleichung gelöst werden. Darauf entfällt der Hauptrechenaufwand.

Deshalb schauen wir uns das System (3.15) genauer an. Hierzu setzen wir

$$\begin{aligned} \begin{pmatrix} s_k^u \\ s_k^q \\ s_k^z \end{pmatrix} &:= s_k, \\ r_k^1 &:= Au_k - F + B^*q_k, \\ r_k^2 &:= Bu_k - d - \gamma q_k + z_k, \\ r_k^3 &:= \gamma q_k + \min(0, -Bu_k + d). \end{aligned}$$

und erhalten  $\begin{pmatrix} r_k^1 \\ r_k^2 \\ r_k^3 \end{pmatrix} = -C(y_k)$ . Wir wählen

$$M_k = \begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -D(G(u_k))B & \gamma Id & 0 \end{pmatrix},$$

mit  $G$  und  $D$  wie im Beweis von Satz 3.20 aus dem verallgemeinerten Differential  $\partial C(y_k)$  aus. Somit können wir für (3.15)

$$\begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -D(G(u_k))B & \gamma Id & 0 \end{pmatrix} \begin{pmatrix} s_k^u \\ s_k^q \\ s_k^z \end{pmatrix} = - \begin{pmatrix} r_k^1 \\ r_k^2 \\ r_k^3 \end{pmatrix} \quad (3.16)$$

komponentenweise schreiben.

Wir führen an diesem linearen System eine Blockelimination durch, indem wir zur unteren Zeile das  $D(G(u_k))$ -fache der zweiten Zeile addieren und

$$\begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ 0 & \gamma(Id - D(G(u_k))) & D(G(u_k)) \end{pmatrix} \begin{pmatrix} s_k^u \\ s_k^q \\ s_k^z \end{pmatrix} = - \begin{pmatrix} r_k^1 \\ r_k^2 \\ r_k^3 - D(G(u_k))r_k^2 \end{pmatrix} \quad (3.17)$$

erhalten.

Nachdem wir bislang noch nicht darauf eingegangen sind, wie wir  $D$  aus der Menge der verallgemeinerte Ableitungen auswählen, wollen wir dies jetzt tun. Wir

werden die Elemente immer so auswählen, dass sie entweder den Wert Eins oder Null haben, also  $D(G(u_k)) = \chi_{-Bu+d \leq 0}$ . Dabei bezeichnet  $\chi$  die charakteristische Funktion. Also hat (3.17) die Form

$$\begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ 0 & \gamma \chi_{-Bu_k+d>0} & \chi_{-Bu_k+d \leq 0} \end{pmatrix} \begin{pmatrix} s_k^u \\ s_k^q \\ s_k^z \end{pmatrix} = - \begin{pmatrix} r_k^1 \\ r_k^2 \\ r_k^3 - \chi_{-Bu_k+d \leq 0} r_k^2 \end{pmatrix}. \quad (3.18)$$

Die letzte Gleichung liest sich als

$$\gamma \chi_{-Bu_k+d>0} s_k^q + \chi_{-Bu_k \leq 0} s_k^z = -r_k^3 - \chi_{-Bu_k+d \leq 0} r_k^2. \quad (3.19)$$

Wir zerlegen  $\Gamma_C$  in die zwei disjunkten Teilmengen  $\{x \in \Gamma_C, -Bu_k(x) + d(x) > 0\}$  und  $\{x \in \Gamma_C, -Bu_k(x) + d(x) \leq 0\}$ , auf denen die Nebenbedingung inaktiv beziehungsweise aktiv ist. Nun betrachten wir die Einschränkung von (3.19) auf diese zwei Mengen und erhalten

$$\begin{aligned} \gamma \chi_{-Bu_k+d>0} s_k^q &= \chi_{-Bu_k+d>0} (-r_k^3 - \chi_{-Bu_k+d \leq 0} r_k^2), \\ \gamma \chi_{-Bu_k+d \leq 0} s_k^z &= \chi_{-Bu_k+d \leq 0} (-r_k^3 - \chi_{-Bu_k+d \leq 0} r_k^2). \end{aligned}$$

Daraus ergeben sich

$$s_k^q|_{-Bu_k+d>0} = -\frac{r_k^3}{\gamma} = -q_k|_{-Bu_k+d>0}$$

und

$$s_k^z|_{-Bu_k+d \leq 0} = (-r_k^3 - \chi_{-Bu_k+d \leq 0} r_k^2)|_{-Bu_k+d \leq 0} = -z_k|_{-Bu_k+d \leq 0},$$

wobei der senkrechte Strich immer die Einschränkung auf die angegebene Teilmenge bezeichnet.

Wir setzen nun

$$s_k^q = -q_k|_{-Bu_k+d>0} + s_k^q|_{-Bu_k+d \leq 0}$$

und

$$s_k^z = -z_k|_{-Bu_k+d>0} + s_k^z|_{-Bu_k+d \leq 0}$$

oben ein und sortieren die bekannten Größen auf die rechte Seite. Somit ergibt sich

$$\begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} s_k^u \\ \chi_{-Bu_k+d \leq 0} s_k^q \\ \chi_{-Bu_k+d>0} s_k^z \end{pmatrix} = - \begin{pmatrix} r_k^1 - B^* \chi_{-Bu_k+d>0} q_k \\ Bu_k - d - \gamma q_k \chi_{-Bu_k+d \leq 0} + z_k \chi_{-Bu_k+d>0} \\ 0 \end{pmatrix}.$$

Um eine übersichtlichere Darstellung zu erlangen, setzen wir

$$\tilde{r}_k^1 := r_k^1 - B^* \chi_{-Bu_k+d>0} q_k,$$

$$\tilde{r}_k^2 := -\chi_{-Bu_k+d \leq 0} (Bu_k - d - \gamma q_k)$$

und

$$\tilde{r}_k^3 := -\chi_{-Bu_k+d>0} (Bu_k - d + z_k).$$

Außerdem benennen wir  $\tilde{s}_k^z := \chi_{-Bu_k+d>0} s_k^z$  und  $\tilde{s}_k^q := \chi_{-Bu_k+d\leq 0} s_k^q$ .

Nach Wegfall der dritten Gleichung und Aufspalten der zweiten Gleichung auf die Teilmengen  $\{x \in \Gamma_C, -Bu_k(x) + d(x) > 0\}$  und  $\{x \in \Gamma_C, -Bu_k(x) + d(x) \leq 0\}$  reduziert sich das System zu

$$\begin{pmatrix} A & B^*|_{-Bu+d\leq 0} & 0 \\ B|_{-Bu+d\leq 0} & -\gamma Id & 0 \\ B|_{-Bu+d>0} & 0 & Id \end{pmatrix} \begin{pmatrix} s_k^u \\ \tilde{s}_k^q \\ \tilde{s}_k^z \end{pmatrix} = - \begin{pmatrix} \tilde{r}_k^1 \\ \tilde{r}_k^2 \\ \tilde{r}_k^3 \end{pmatrix}. \quad (3.20)$$

Das System zerfällt in zwei Teilsysteme. Nach Lösen der ersten zwei Gleichungen kann die Lösung in die dritte eingesetzt und  $\tilde{s}_k^z$  durch  $\tilde{s}_k^z = -\tilde{r}_k^2 - B|_{-Bu+d>0} s_k^u$  bestimmt werden. Wir bezeichnen mit  $\hat{B} := B|_{-Bu+d\leq 0}$  und betrachten das reduzierte System

$$\begin{pmatrix} A & \hat{B}^* \\ \hat{B} & -\gamma Id \end{pmatrix} \begin{pmatrix} s_k^u \\ \tilde{s}_k^q \end{pmatrix} = - \begin{pmatrix} \tilde{r}_k^1 \\ \tilde{r}_k^2 \end{pmatrix}. \quad (3.21)$$

Hier können wir noch das  $\frac{1}{\gamma}\hat{B}^*$ -fache der zweiten Zeile zu der ersten addieren und erhalten

$$\left( A + \frac{1}{\gamma} \hat{B}^* \hat{B} \right) s_k^u = -\frac{1}{\gamma} \hat{B}^* \tilde{r}_k^2 - \tilde{r}_k^1. \quad (3.22)$$

Das System ist gleichmäßig elliptisch mit der gleichen Elliptizitätskonstante wie  $A$ , da zum gleichmäßig elliptischen Operator  $A$  ein positiv semidefiniter Operator  $\frac{1}{\gamma}\hat{B}^*\hat{B}$  addiert wird. Also lässt sich das System eindeutig lösen. Die Bestimmung von  $\tilde{s}_k^q$  folgt ohne großen Rechenaufwand aus der zweiten Gleichung.

Der Hauptrechenaufwand liegt in der Berechnung der Lösung von (3.22). Die restlichen Unbekannten lassen sich nahezu ohne große Rechnungen direkt bestimmen.

Die eben durchgeführten Umformungen gelten auch für das diskretisierte Problem.

Wir betrachten nun das diskretisierte System. Hierzu diskretisieren wir den Raum  $H^1(\Omega)$  mit linearen finiten Elementen und den Raum  $L^2(\Gamma_C)$  mit stückweise konstanten Elementen.

Seien  $\phi_i$  die Basiselemente der finiten Elemente Approximation von  $H^1(\Omega)$  und  $\psi_i$  die von  $L^2(\Omega)$ , so ist die Steifigkeitsmatrix  $(A_h)$  gegeben durch  $(A_h)_{ij} = a(\phi_i, \phi_j)$ , der Lastvektor  $F_h$  durch  $(F_h)_i = \langle F, \phi_i \rangle_{H^{-1}(\Omega), H^1(\Omega)}$ .

Wir werden den Spuroperator  $B$  so auswählen, dass er den relativen Verschiebungsvektor im Mittelpunkt  $x_S^{(i)}$  des  $i$ -ten Dreiecks  $\Delta^{(i)}$  auswertet. Somit ergibt sich für

$$\langle B^* \psi_i, \phi_j \rangle_{H^{-1}(\Omega), H^1(\Omega)} = (\psi_i, B\phi_j)_{L^2(\Omega)} = (\psi_i, B\phi_j)_{L^2(\Delta^{(i)})} = (B\phi_j)(x_S^{(i)}) |\Delta^{(i)}|,$$

wobei der Betrag die Fläche des Dreiecks  $\Delta^{(i)}$  bezeichnet. Wir bestimmen die Spurmatrix  $(B_h)_{ij} := (\psi_i, B\phi_j)_{L^2(\Delta^{(i)})}$  und führen unter Verwendung des Kroneckerdeltas  $\delta$  auf  $L^2(\Gamma_C)$  die Massenmatrix  $(M_h)_{ij} := \int_{\Omega} \psi_i \psi_j dx = \delta_{ij} |\Delta^{(i)}|$  ein.

Somit erhalten wir das nichtlineare Gleichungssystem

$$\begin{aligned} A_h u_h - F_h + B_h^T q_h &= 0, \\ B_h u_h - M_h d_h - \gamma M_h q_h + M_h z_h &= 0, \\ \gamma M_h q_h + \min(0, -B_h u_h + M_h d_h) &= 0, \end{aligned} \quad (3.23)$$

welches wir analog zu oben mit dem endlichdimensionalen semiglatten Newton-Verfahren lösen werden. Auch hier werden wir als Element der verallgemeinerten Ableitung der Minimumsfunktion komponentenweise die 0, falls  $-B_h u_h + M_h d_h \leq 0$  ist, und sonst die 1 wählen. Die Umformungen erfolgen alle analog. Wir lassen im weiteren Verlauf bei den Vektoren den Diskretisierungsindex wegfallen.

Der Hauptaufwand im diskreten System liegt in der Lösung des linearen Gleichungssystems

$$\left( A_h + \hat{B}_h^T \hat{B}_h \right) s_k^u = \left( -\frac{1}{\gamma} \hat{B}^* \tilde{r}_k^2 - \tilde{r}_k^1 \right)_h. \quad (3.24)$$

Das System ist zwar nicht notwendig dünn besetzt, aber zu der in der Regel dünn besetzten Steifigkeitsmatrix wird eine Niedrigrangstörung  $\hat{B}_h^T \hat{B}_h$  addiert, so dass  $A_h + \hat{B}_h^T \hat{B}_h$  eine einfache Struktur hat. Also kann zur Lösung von (3.24) ein iterativer Gleichungssystemlöser verwendet werden. Da es sich um ein symmetrisches positiv definites System handelt, kann man auf fast alle bekannten iterativen Löser für symmetrisch-positiv-definite Matrizen zurückgreifen, die zum Beispiel in [Hac94, Saa03] vorgestellt werden.

Im Folgenden werden wir zur Lösung ein präkonditioniertes konjugiertes Gradientenverfahren (pcg-Verfahren) einsetzen, wobei wir als Präkonditionierer den V-Zyklus eines Mehrgittersystems [Hac85] benutzen werden. Die Toleranz, die wir für die Lösung von (3.24) vorgeben, werden wir in Abhängigkeit des Residuums des Newton-Verfahrens verwenden.

Dabei gehen wir genauso vor, wie es in [GK99] für das gewöhnliche Newton-Verfahren beschrieben wird.

Die Beweise lassen sich eins zu eins übernehmen, weshalb wir an dieser Stelle nur den Algorithmus des inexakten semiglatten Mehrgitter-Newton-Verfahrens angeben, mit dem wir das Kontaktproblem lösen werden. Dabei verwenden wir die in Matlab gebräuchliche Matrizennotation

**Algorithmus 3.24** (Inexaktes semiglatte Mehrgitter-Newton-Verfahren).

- (a) Gegeben:  $A_h \in \mathbb{R}^{n \times n}$ ,  $B_h \in \mathbb{R}^{m \times n}$ ,  $M_h \in \mathbb{R}^{m \times m}$ ,  $\gamma > 0$ ,  $d \in \mathbb{R}^m$ ,  $\eta \in (0, 1)$ ,  $tol > 0$  und eine Nullfolge  $(\eta_k)_{k \in \mathbb{N}}$  mit  $0 < \eta_k < 1 \forall k \in \mathbb{N}$ .
- (b) Wähle  $u_0 \in \mathbb{R}^n$ ,  $q_0 \in \mathbb{R}^m$  und  $z_0 \in \mathbb{R}^m$ ; Setze  $k = 0$ .
- (c) Berechne  $r_k := \begin{pmatrix} r_k^1 \\ r_k^2 \\ r_k^3 \end{pmatrix} = - \begin{pmatrix} A_h u_k - F_h + B_h^T q_k \\ B_h u_k - M_h d - \gamma M_h q_k + M_h z_k \\ \gamma M_h q_k + \min(0, -B_h u_k + M_h d) \end{pmatrix}$ .
- (d) Falls  $\|r_k\| < tol$ ; Ende:  $(u_k, q_k, z_k)$  ist Näherungslösung zur Genauigkeit  $tol$ .

- (e) Setze  $\mathcal{I}_k := \{i \in \{1, \dots, m\} \mid -B(i, :)u + M_h(i, i)d(i) > 0\}$ ,  $\mathcal{A}_k := \{1, \dots, n\} \setminus \mathcal{I}_k$ .
- (f) Berechne  $s_k^q(\mathcal{A}_k) := -q_k(\mathcal{A})$ ;  $s_k^z(\mathcal{I}_k) := -z_k(\mathcal{I})$ .
- (g) Setze  $\tilde{r}_k^1 := r_k^1 - B_h(\mathcal{I}, :)^T s_k^q(\mathcal{I}) + B_h(\mathcal{A}, :)^T (M_h(\mathcal{A}))^{-1} (r_k^2(\mathcal{A}) - M_h(\mathcal{A})s_k^z(\mathcal{A}))$ .
- (h) Löse  $(A_h + \frac{1}{\gamma} B_h(\mathcal{A}, :)^T (M_h(\mathcal{A}))^{-1} B_h(\mathcal{A}, :)) s_k^u = \tilde{r}_k^1$  mit einem vorkonditionierten pcg-Verfahren bis auf die Genauigkeit  $\eta_k \|r_k\|$ , wobei ein Mehrgitterverfahren als Vorkonditionierer verwendet wird.
- (i) Löse  $s_k^q(\mathcal{A}) = -\frac{1}{\gamma} (M_h(\mathcal{A}))^{-1} (r_k^2(\mathcal{A}) - M_h(\mathcal{A})s_k^z(\mathcal{A}) - B_h(\mathcal{A}, :)s_k^u)$ .
- (j) Löse  $s_k^z(\mathcal{I}) := (M_h(\mathcal{I}))^{-1} (r_k^2(\mathcal{I}) + M_h(\mathcal{I})\gamma s_k^q(\mathcal{I}) - A_h s_k^u)$ .
- (k) Setze  $(u_{k+1}, q_{k+1}, z_{k+1}) := (u_k, q_k, z_k) + (s_k^u, s_k^q, s_k^z)$ ;  $k := k + 1$ .
- (l) Gehe zu (c).

**Bemerkung 3.25.**

- (a) Die Bezeichnungen  $\mathcal{A}$  und  $\mathcal{I}$  erinnern an die Strategien, die auf aktiven beziehungsweise inaktiven Mengen basieren.
- (b) Das diskrete Verfahren konvergiert, wenn es konvergiert, in endlich vielen Schritten, falls das Gleichungssystem (3.24) exakt gelöst wird. Falls das Verfahren nicht konvergiert, werden die Iterierten sich nach einem Vorlauf periodisch wiederholen. Dies liegt daran, dass nur endlich viele Kombinationen an aktiven und inaktiven Mengen vorkommen.

## 3.5 Fehler bei iterativen Lösern

Nachdem wir nun einen iterativen Löser zur Lösung des Kontaktproblems verwenden, können wir die Genauigkeit des Residuums steuern. Allerdings wissen wir nicht, wie sich die exakte Lösung des Kontaktproblems von der Näherungslösung unterscheidet.

Zur Untersuchung dieses Fehlers leiten wir in diesem Abschnitt eine Abschätzung her. Diese beruht auf folgenden Annahmen, die im weiteren Verlauf gelten sollen:

**Annahme 3.26.**

- (a) Es existiert eine positive stetige, streng monotone Abbildung  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  mit  $\phi(0) = 0$  derart, dass das Maß der Menge, in der die zueinander komplementären Variablen klein sind, durch diese beschränkt wird, also dass

$$\text{meas}(\{x \in \Gamma_C : |q(x)| + |z(x)| \leq t\}) \leq \phi(t)$$

gilt. Dies impliziert strikte Komplementarität.

(b) Die Funktion  $\tilde{y} = \begin{pmatrix} \tilde{u} \\ \tilde{q} \\ \tilde{z} \end{pmatrix}$  ist eine Näherung an  $y = \begin{pmatrix} u \\ q \\ z \end{pmatrix}$ , der Lösung des Komplementaritätsproblems.

(c) Die zweite Gleichung des Komplementaritätsproblems wird immer exakt gelöst. Das ist keine große Einschränkung, da dies entweder durch den Algorithmus geschehen kann, oder man betrachtet das reduzierte Problem, das man durch Elimination der Schlupfvariablen erhält. Alternativ kann man die Schlupfvariable nach jedem Optimierungsschritt so bestimmen, dass diese Bedingung erfüllt ist.

(d) Es gibt ein  $M > 0$ , so dass für die Lösung des Komplementaritätsproblems  $\|C_y(y)^{-1}\|_{Y^*, Y} < M$  gilt.<sup>1</sup>

**Hilfssatz 3.27.** Sei  $y$  die Lösung des Komplementaritätsproblems (3.7) und  $\tilde{y}$  eine algorithmisch bestimmte Näherung an  $y$ , dann existiert eine stetige positive Abbildung  $\Psi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_+$  mit  $\Psi(0, 0) = 0$ , so dass

$$\text{meas}(\{q < z, \tilde{q} \geq \tilde{z}\} \cup \{q > z, \tilde{q} < \tilde{z}\}) \leq \Psi(\|\tilde{q} - q\|_{L^2(\Gamma_C)}, \|\tilde{z} - z\|_{L^2(\Gamma_C)})$$

gilt. Dabei ist

$$\Psi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}, \quad (3.25)$$

$$(a, b) \mapsto 2 \left( 4\sqrt{a^2 + b^2} + \phi(\sqrt[4]{a^2 + b^2}) \right). \quad (3.26)$$

*Beweis:* Wir setzen  $\hat{\Gamma}_C^1 := \{x \in \Gamma_C : \tilde{q}(x) \geq \tilde{z}(x) \text{ und } q(x) < z(x)\}$ . Diese Menge ist bis auf eine Menge vom Maß Null eindeutig festgelegt. Es gilt also für fast alle  $x \in \hat{\Gamma}_C^1$   $\tilde{q}(x) \geq \tilde{z}(x)$  und  $q(x) < z(x)$ . Dann folgen

$$\begin{aligned} \tilde{q}(x) &\geq \tilde{z}(x), \\ \tilde{q}(x) - q(x) + q(x) &\geq \tilde{z}(x) - z(x) + z(x), \\ \tilde{q}(x) - q(x) + z(x) - \tilde{z}(x) &\geq z(x) - q(x). \end{aligned}$$

Nachdem  $q$  und  $z$  komplementär auf  $\Gamma_C$  zueinander sind, folgt  $q = 0$  und mit der Dreiecksungleichung des Absolutbetrags

$$|\tilde{q}(x) - q(x)| + |\tilde{z}(x) - z(x)| \geq z(x) \text{ für alle } x \in \hat{\Gamma}_C^1.$$

Also gilt aufgrund der Monotonie des Maßes

$$\text{meas}(\hat{\Gamma}_C^1) \leq \text{meas}(\{x \in \Gamma_C : |\tilde{q}(x) - q(x)| + |\tilde{z}(x) - z(x)| \geq z(x)\}).$$

<sup>1</sup>Diese Aussage gilt, falls  $C_y(y)$  regulär ist, trivialerweise. Wir werden diese Abschätzung in Kapitel Fünf verwenden. Dort werden wir Kontaktprobleme betrachten, die stetig von einem Parameter  $p \in P$  abhängen.  $P$  ist eine Menge, mit der wir uns intensiv in Kapitel Vier auseinandersetzen. In diesem Fall brauchen wir hier, dass  $\|C_y(y(p), p)^{-1}\| < M$  für alle  $p \in P$  gilt.

Für  $t > 0$  beliebig ist dann

$$\begin{aligned} & \text{meas}(\{x \in \Gamma_C : |\tilde{q}(x) - q(x)| + |\tilde{z}(x) - z(x)| \geq z(x)\}) \\ & \leq \text{meas}(\{x \in \Gamma_C : |\tilde{q}(x) - q(x)| + |\tilde{z}(x) - z(x)| \geq t\}) \\ & \quad + \text{meas}(\{x \in \Gamma_C : z(x) \leq t\}). \end{aligned}$$

Wir schätzen nun die Quadrate der Normen der Abweichungen von Lösung und Näherung ab durch

$$\begin{aligned} & \|\tilde{q} - q\|_{L^2(\Gamma_C)}^2 + \|\tilde{z} - z\|_{L^2(\Gamma_C)}^2 \\ & = \int_{\Gamma_C} |\tilde{q}(x) - q(x)|^2 + |\tilde{z}(x) - z(x)|^2 dx \\ & \geq \int_{\{x \in \Gamma_C : |\tilde{q}(x) - q(x)| + |\tilde{z}(x) - z(x)| \geq t\}} |\tilde{q}(x) - q(x)|^2 + |\tilde{z}(x) - z(x)|^2 dx \\ & \geq \int_{\{|\tilde{q}(x) - q(x)| + |\tilde{z}(x) - z(x)| \geq t\}} \left(\frac{t}{2}\right)^2 dx \\ & = \left(\frac{t}{2}\right)^2 \text{meas}\{x : |\tilde{q}(x) - q(x)| + |\tilde{z}(x) - z(x)| \geq t\}. \end{aligned}$$

Somit ist

$$\text{meas}(\{x : |\tilde{q}(x) - q(x)| + |\tilde{z}(x) - z(x)| \geq t\}) \leq \left(\frac{2}{t}\right)^2 \left(\|\tilde{q} - q\|_{L^2(\Gamma_C)}^2 + \|\tilde{z} - z\|_{L^2(\Gamma_C)}^2\right).$$

Zusammengefasst gilt

$$\text{meas}(\hat{\Gamma}_C^1) \leq \left(\frac{2}{t}\right)^2 \left(\|\tilde{q} - q\|_{L^2(\Gamma_C)}^2 + \|\tilde{z} - z\|_{L^2(\Gamma_C)}^2\right) + \phi(t).$$

Setzen wir  $t := \sqrt[4]{\|\tilde{q} - q\|_{L^2(\Gamma_C)}^2 + \|\tilde{z} - z\|_{L^2(\Gamma_C)}^2}$ , so erhalten wir

$$\begin{aligned} \text{meas}(\hat{\Gamma}_C^1) & \leq 4\sqrt{\|\tilde{q} - q\|_{L^2(\Gamma_C)}^2 + \|\tilde{z} - z\|_{L^2(\Gamma_C)}^2} \\ & \quad + \phi\left(\sqrt[4]{\|\tilde{q} - q\|_{L^2(\Gamma_C)}^2 + \|\tilde{z} - z\|_{L^2(\Gamma_C)}^2}\right). \end{aligned}$$

Analog kann für die Menge  $\hat{\Gamma}_C^2 := \{x \in \Gamma_C : q(x) > z(x), \tilde{q}(x) \leq z(x)\}$  gezeigt werden, dass

$$\begin{aligned} \text{meas}(\hat{\Gamma}_C^2) & \leq 4\sqrt{\|\tilde{q} - q\|_{L^2(\Gamma_C)}^2 + \|\tilde{z} - z\|_{L^2(\Gamma_C)}^2} \\ & \quad + \phi\left(\sqrt[4]{\|\tilde{q} - q\|_{L^2(\Gamma_C)}^2 + \|\tilde{z} - z\|_{L^2(\Gamma_C)}^2}\right) \end{aligned}$$

ist. Mit  $\Psi(a, b) = 2(4\sqrt{a^2 + b^2} + \phi(\sqrt[4]{a^2 + b^2}))$  folgt für

$$\begin{aligned} \hat{\Gamma}_C & := \{x \in \Gamma_C : q(x) < z(x), \tilde{q}(x) \geq \tilde{z}(x)\} \cup \{x \in \Gamma_C : q(x) > z(x), \tilde{q}(x) \leq \tilde{z}(x)\} \\ & = \hat{\Gamma}_C^1 \cup \hat{\Gamma}_C^2, \end{aligned} \tag{3.27}$$

dass

$$\text{meas}(\hat{\Gamma}_C) \leq \Psi (\|\tilde{q} - q\|_{L^2(\Gamma_C)}, \|\tilde{z} - z\|_{L^2(\Gamma_C)})$$

gilt. ■

Im Folgenden betrachten wir statt des vollen Komplementaritätsproblems das reduzierte, das man durch Auflösen der zweiten Gleichung nach der Schlupfvariablen  $z$  und Einsetzen dieser Lösung in die dritte Gleichung erhält, wie wir es bereits in (3.10) gemacht haben.

**Satz 3.28.** *Sei  $\bar{y} = (\bar{u}, \bar{q}, \bar{z})$  Lösung der Komplementaritätsbedingung  $C(u, q, z) = 0$  mit strikter Komplementarität und es gelte Annahme 3.26. Dann existieren  $\delta > 0$ ,  $\kappa > 0$ , so dass für eine Näherungslösung  $\tilde{y} = (\tilde{u}, \tilde{q}, \tilde{z})$  von  $C(\tilde{y}) = 0$  gilt*

$$\|\bar{y} - \tilde{y}\|_Y \leq \kappa \|C(\tilde{y})\|_{Y, Y^*} \quad (3.28)$$

für alle  $\tilde{y}$  mit  $\|\bar{y} - \tilde{y}\|_Y \leq \delta$ .

*Beweis:* Wir definieren eine Abbildung  $\tilde{C}_y$  mit der Eigenschaft

$$C(\tilde{y}) = \tilde{C}_y(\tilde{y} - \bar{y})$$

punktweise fast überall in  $\Omega \times \Gamma_C \times \Gamma_C$ . Hierzu betrachtet man

$$\begin{aligned} C(\tilde{y}) - C(\bar{y}) &= \begin{pmatrix} A\tilde{u} - F + B^*\tilde{q} \\ B\tilde{u} - d + \tilde{z} - \gamma\tilde{q} \\ \min(\gamma\tilde{q}, \tilde{z}) \end{pmatrix} - \begin{pmatrix} A\bar{u} - F + B^*\bar{q} \\ B\bar{u} - d + \bar{z} - \gamma\bar{q} \\ \min(\gamma\bar{q}, \bar{z}) \end{pmatrix} \\ &= \begin{pmatrix} A(\tilde{u} - \bar{u}) + B^*(\tilde{q} - \bar{q}) \\ B(\tilde{u} - \bar{u}) + (\tilde{z} - \bar{z}) - \gamma(\tilde{q} - \bar{q}) \\ \min(\gamma\tilde{q}, \tilde{z}) - \min(\gamma\bar{q}, \bar{z}) \end{pmatrix}. \end{aligned}$$

Nachdem die ersten zwei Zeilen der Matrix linear sind, können wir die zugehörigen linearen Abbildungen direkt in  $\tilde{C}_y$  übernehmen. Die letzte Zeile betrachten wir gesondert.

Hier verwenden wir zunächst die Annahme 3.26(c) und eliminieren  $\tilde{z}$  und  $\bar{z}$ . Somit erhalten wir

$$\begin{aligned} C(\tilde{y}) - C(\bar{y}) &= \begin{pmatrix} A(\tilde{u} - \bar{u}) + B^*(\tilde{q} - \bar{q}) \\ B(\tilde{u} - \bar{u}) + (\tilde{z} - \bar{z}) - \gamma(\tilde{q} - \bar{q}) \\ \min(\gamma\tilde{q}, \gamma\tilde{q} + d - B\tilde{u}) - \min(\gamma\bar{q}, \gamma\bar{q} + d - B\bar{u}) \end{pmatrix} \\ &= \begin{pmatrix} A(\tilde{u} - \bar{u}) + B^*(\tilde{q} - \bar{q}) \\ B(\tilde{u} - \bar{u}) + (\tilde{z} - \bar{z}) - \gamma(\tilde{q} - \bar{q}) \\ \gamma(\tilde{q} - \bar{q}) + \min(0, d - B\tilde{u}) - \min(0, d - B\bar{u}) \end{pmatrix}. \end{aligned}$$

Nachdem die Minimumsfunktion auf  $\mathbb{R}^2$  semiglatt ist, können wir unter Verwendung des Mittelwertsatzes für verallgemeinerte Ableitungen [OKZ98] fast überall punktweise schreiben, dass

$$\min(0, d - B\tilde{u}) - \min(0, d - B\bar{u}) = -\tilde{D}(\tilde{y}(x), \bar{y}(x))B(\tilde{u} - \bar{u})$$

gilt, wobei

$$\tilde{D}(\tilde{y}(x), \bar{y}(x)) := \begin{cases} 0 & -d(x) - (B\tilde{u})(x) > 0, -d(x) - (B\bar{u})(x) > 0, \\ 1 & -d(x) - (B\tilde{u})(x) < 0, -d(x) - (B\bar{u})(x) < 0, \\ \in [0, 1] & \text{sonst,} \end{cases}$$

ist. Wir setzen jetzt  $\tilde{C}_y := \begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -\tilde{D}B & \gamma Id & 0 \end{pmatrix}$  und erhalten wie oben gewünscht

$$C(\tilde{y}) = \tilde{C}_y(\tilde{y}, \bar{y})(\tilde{y} - \bar{y}).$$

Nach Multiplikation mit  $C_y(\bar{y})^{-1}$  von rechts und Addition einer Null erhalten wir unter Vernachlässigung der Argumente

$$\begin{aligned} C_y^{-1}C(\tilde{y}) &= C_y^{-1}\tilde{C}_y(\tilde{y} - \bar{y}) \\ &= C_y^{-1}(\tilde{C}_y - C_y + C_y)(\tilde{y} - \bar{y}) \\ &= (\tilde{y} - \bar{y}) - (C_y)^{-1}(C_y - \tilde{C}_y)(\tilde{y} - \bar{y}). \end{aligned}$$

Man kann jetzt zu den Normen übergehen und die linke Seite nach unten abschätzen durch

$$\|C_y^{-1}C(\tilde{y})\|_Y \geq \|\tilde{y} - \bar{y}\|_Y - \|C_y^{-1}\|_{Y^*, Y} \|\tilde{C}_y - C_y\|_{Y, Y^*} \|\tilde{y} - \bar{y}\|_Y. \quad (3.29)$$

Wir berechnen nun  $\|\tilde{C}_y - C_y\|_{Y^*, Y}$ . Dabei verwenden wir wieder die Darstellung (3.12) von  $C(y)$ , bei der die dritte Variable in der dritten Gleichung mit der zweiten Gleichung eliminiert wurde.

$$\begin{aligned} \tilde{C}_y - C_y &= \begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -\tilde{D}B & \gamma Id & 0 \end{pmatrix} - \begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -\chi_{\bar{z} \leq \gamma \bar{q}} B & \gamma Id & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ (-\tilde{D} + \chi_{\bar{z} \leq \gamma \bar{q}})B & 0 & 0 \end{pmatrix}. \end{aligned} \quad (3.30)$$

Nach der Definition der Operatornorm ist

$$\|\tilde{C}_y - C_y\|_{Y, Y^*} = \sup_{\|v\|_{Y^*} \leq 1} \|(\tilde{C}_y - C_y)v\|_{Y^*} = \sup_{\|v_1\|_{H^1(\Omega)} \leq 1} \|(-\tilde{D} + \chi_{\bar{z} \leq \gamma \bar{q}})Bv_1\|_{L^2(\Gamma_C)},$$

wobei  $v_1$  die erste Komponente von  $v$  bezeichnet. Wir setzen  $\omega = Bv_1$ , somit ist dann  $\omega \in H^{\frac{1}{2}}(\Gamma_C)$ . Weiterhin existiert nach den Einbettungssätzen für Sobolev-Räume [Ada75] ein  $s > 2$  mit  $\omega \in L^s(\Gamma_C)$ . Wir berechnen zunächst

$$\sup_{\|\omega\|_{L^s} \leq 1} \|(-D + \chi_{\bar{z} \leq \gamma \bar{q}})\omega\|_{L^2(\Gamma_C)} \leq 2\|\omega\|_{L^2(\hat{\Gamma}_C)}, \quad (3.31)$$

mit  $\hat{\Gamma}_C$  wie in(3.27), also ist die Abbildung stetig.

Mit der Hölder-Ungleichung gilt

$$\|\omega\|_{L^2(\hat{\Gamma}_C)}^2 = \|1 \cdot \omega^2\|_{L^1(\hat{\Gamma}_C)} \leq \|1\|_{p'} \|\omega^2\|_p$$

für  $\frac{1}{p'} + \frac{1}{p} = 1$ . Wir wählen  $p = \frac{s}{2}$  und erhalten

$$\begin{aligned} \|\omega\|_{L^2(\hat{\Gamma}_C)}^2 &\leq \|1\|_{\frac{s}{s-2}} \|\omega^2\|_{\frac{s}{2}} \\ &= \left(\text{meas}(\hat{\Gamma}_C)\right)^{\frac{s-2}{s}} \|\omega\|_{L^s(\hat{\Gamma}_C)}^2 \\ &\leq \left(\text{meas}(\hat{\Gamma}_C)\right)^{\frac{s-2}{2}} \|\omega\|_{L^s(\Gamma_C)}^2. \end{aligned}$$

Also folgt mit der Operatornorm  $\|B\|_{H^1, L^s}$  von  $B \in L(H^1, L^s)$  die Normabschätzung

$$\|\tilde{C}_y - C_y\|_{Y^*, Y} \leq 2 \left(\text{meas}(\hat{\Gamma}_C)\right)^{\frac{s-2}{2s}} \|B\|_{H^1, L^s}.$$

Verwendet man nun Hilfssatz 3.27, so gilt die Abschätzung

$$\begin{aligned} \|\tilde{C}_y - C_y\|_{Y, Y^*} &\leq 4 \left(\Psi(\|\tilde{q} - \bar{q}\|_{L^2(\Gamma_C)}, \|\tilde{z} - \bar{z}\|_{L^2(\Gamma_C)})\right)^{\frac{s}{s-2}} \|B\|_{H^1, L^s} \\ &=: \Phi(\|\tilde{q} - \bar{q}\|_{L^2(\Gamma_C)}, \|\tilde{z} - \bar{z}\|_{L^2(\Gamma_C)}). \end{aligned}$$

Nach Voraussetzung existiert ein  $M > 0$  mit  $\|C_y(y)^{-1}\|_{Y^*, Y} < M$ . Weil die Abbildung  $\Psi$  als Komposition stetiger Abbildung stetig ist, existiert ein  $\delta > 0$ , so dass für  $\|y\|_Y < \delta$

$$\Phi(\|\tilde{q} - \bar{q}\|_{L^2(\Gamma_C)}, \|\tilde{z} - \bar{z}\|_{L^2(\Gamma_C)}) < \frac{1}{2M}$$

gilt.

Wir setzen dies in (3.29) ein und erhalten unter Berücksichtigung der Norm

$$M\|C(\tilde{y})\| \geq \|\bar{y} - \tilde{y}\| - M\frac{1}{2M}\|\bar{y} - \tilde{y}\| \geq \frac{1}{2}\|\bar{y} - \tilde{y}\|. \quad (3.32)$$

Mit  $\kappa = 2M$  folgt die Behauptung. ■

# Kapitel 4

## Formoptimierung

Bislang haben wir uns damit beschäftigt, wann ein Kontaktproblem lösbar ist und wie die Lösung bestimmt werden kann. In diesem Kapitel wollen wir auf Formoptimierungsprobleme mit Kontaktproblemen als Nebenbedingungen eingehen.

Die Formoptimierung gehört zur Strukturoptimierung. Nach [BS03] gliedert sich die Strukturoptimierung in drei Teilgebiete (siehe Abbildung 4.1): Sizing-Optimization, Formoptimierung und die Topologieoptimierung. Bei der Sizing-Optimization (Abbildung 4.1a) wird die Struktur eines Bauteils vorgegeben und über die Größe der Teilkomponenten bezüglich eines Zielfunktional optimiert. Bei der Topologieoptimierung (Abbildung 4.1c) wird das Gebiet in Zellen unterteilt und während der Optimierung entschieden, ob diese Zellen mit Material gefüllt sind oder hohl bleiben, um ein gewünschtes Funktional zu minimieren. Hierbei handelt es sich um ein 0-1 Optimierungsproblem. Damit liegt eigentlich ein diskretes Optimierungsproblem vor; erst neuere Ansätze, wie zum Beispiel der SIMP-Ansatz [BS03], betten das Problem in ein kontinuierliches Setting ein.

Bei der Formoptimierung (Abbildung 4.1b), wie wir sie im Folgenden behandeln, wird die Topologie des Gebietes, das heißt die Anzahl und Lage von Löchern im Grundgebiet vorgegeben, aber die genaue Position und Gestalt der Löcher variiert.

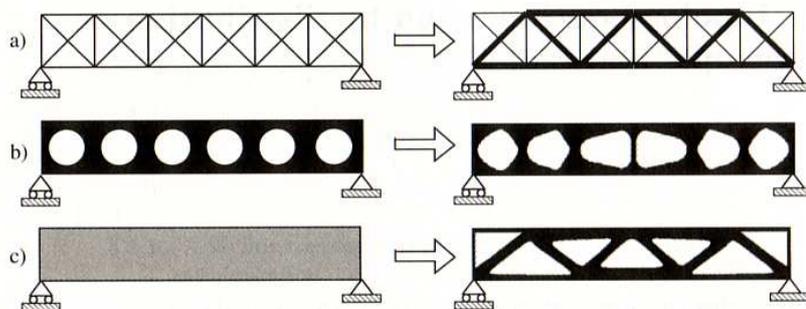


Abbildung 4.1: Die Strukturoptimierung gliedert sich in die Sizing-Optimization (a), die Formoptimierung (b) und die Topologieoptimierung (c) aus [BS03].

Man kann sich das Gebiet vorstellen, als sei es aus einem sehr dehnbaren Gummi so hergestellt, dass es sich verformen lässt, ohne dass sich in dem Gummi neue Löcher bilden oder bereits vorhandene Löcher verschwinden. Die Topologie bleibt somit erhalten.

In unserem Falle geht es darum, das Gebiet, auf dem das Kontaktproblem definiert ist, variabel zu halten und das Kontaktproblem in Abhängigkeit von dem Gebiet zu lösen. Dann sucht man das Gebiet, das ein gegebenes Funktional minimiert, aus der Menge der möglichen Gebiete.

Dabei handelt es sich um ein Bilevel-Optimierungsproblem. In unserem Fall handelt es sich um ein unendlichdimensionales MPEC<sup>1</sup>, also ein Optimierungsproblem mit Variationsungleichungen als Nebenbedingung.

Zunächst werden wir das Setting besprechen, dann zeigen wir die Existenz einer optimalen Lösung. Anschließend werden wir auf das Konzept der Formableitung eingehen. Es wird sich zeigen, dass das Formoptimierungsproblem nur Lipschitzstetig und somit nach dem Satz von Rademacher nur fast überall differenzierbar ist. Wir werden zeigen, dass das Formoptimierungsproblem differenzierbar ist, falls im Kontaktproblem strikte Komplementarität herrscht. In diesem Fall werden wir die Ableitung explizit berechnen. Dabei werden wir den Ideen folgen, die in [HCK86, HM03, HN96, SZ92b] vorgestellt werden.

## 4.1 Gebietsverformungen

In diesem Abschnitt beschreiben wir die zulässigen Gebiete. Hierzu gehen wir von einem beschränkten Grundgebiet  $\Omega \subset \mathbb{R}^3$  aus, welches zusammenhängend ist, und einer Parametermenge  $P \subset \mathbb{R}^n$ , die wir später genauer spezifizieren.

Wir bezeichnen die Menge der zulässige Gebiete mit

$$\mathcal{O} := \{\Omega(p) \mid p \in P\}. \quad (4.1)$$

Dabei ist

$$\Omega(p) := \{y \in \mathbb{R}^3 \mid y = x + \mathcal{V}_p(x), x \in \Omega\}$$

Wir schreiben hierfür auch  $\Omega(p) = T_p(\Omega)$  mit  $T_p(x) = x + \mathcal{V}_p(x)$  für  $x \in \Omega$  und  $p \in P$ .

Falls keine Missverständnisse auftreten, schreiben wir auch  $\Omega_p$  statt  $\Omega(p)$ .

Die Funktion  $\mathcal{V}_p : \Omega \rightarrow \mathbb{R}^3$  gibt die Verschiebung eines Punktes  $x$  aus der Referenzgeometrie  $\Omega$  in das Gebiet  $\Omega(p)$  an. Wir fordern  $\mathcal{V}_p \in W^{1,\infty}(\Omega)$ . Die Funktion lässt sich als das Geschwindigkeitsfeld deuten, unter dem sich das Gebiet verformt.

Ein vergleichbarer Zugang, der in der Literatur unter dem Namen „Method of Mapping“ bekannt ist, geht auf Murat und Simon zurück [MS76]. Dort werden mehr

---

<sup>1</sup>Mathematical Program with Equilibrium Constraints; für eine endlichdimensionale Behandlung verweisen wir auf [LPR96].

Annahmen an den Raum der Gebietsverformungen gestellt als bei uns, da wir die topologischen Eigenschaften auf die Parametermenge  $P$  zurückführen.

**Annahme 4.1.** *Sei  $P$  eine kompakte Menge und  $\Omega \subset \mathbb{R}^3$  beschränkt. Wir betrachten nun solche Abbildungen*

$$T_p : \Omega \rightarrow \Omega(p), x \mapsto x + \mathcal{V}_p(x), \quad (4.2)$$

wobei  $p \in P \mapsto T_p \in W^{1,\infty}(\Omega)$  eine stetige Abbildung ist in die Menge der Diffeomorphismen von  $\Omega \rightarrow \Omega(p)$ , welche stetig auf den Rand fortsetzbar sind. Wir nehmen an, dass die Einschränkung von  $T_p$  auf den Kontaktrand und auf den Teil des Randes, auf dem eine Neumann-Randbedingung gilt, die identische Abbildung ist. Außerdem nehmen wir an, dass ein  $\theta > 0$  existiert mit  $\frac{1}{\theta} > \|DT_p(x)^{-1}\|_F > \theta$  und  $|\det DT_p| > \theta$  für alle  $p \in P$  und  $x \in \bar{\Omega}$ .

**Hilfssatz 4.2.** *Sei  $C > 0$  und  $T_p, p \in P$ , eine Familie von Gebietsverformungen der Bauart (4.2) mit  $\|\mathcal{V}_p\|_{1,\infty} < C$ . Für  $u \in H^1(\Omega)$  setzen wir*

$$\epsilon_p(u) := (DuDT_p^{-1}) + (DuDT_p^{-1})^T.$$

Dann gilt unter Annahme 4.1:

(a) *Es existiert ein  $m > 0$  mit*

$$|\det DT_p(x)| < m \text{ für fast alle } x \in \Omega \text{ und } p \in P. \quad (4.3)$$

(b) *Es existiert ein  $L > 0$  mit*

$$|\det(DT_{p_1})(x) - \det(DT_{p_2})(x)| \leq L\|T_{p_1} - T_{p_2}\|_{1,\infty} \quad (4.4)$$

*fast überall in  $\bar{\Omega}$  für  $p_1, p_2 \in P$ .*

(c) *Es existiert ein  $L > 0$  mit*

$$\|DT_{p_1}(x) - DT_{p_2}(x)\|_F \leq L\|T_{p_1} - T_{p_2}\|_{1,\infty}$$

*und*

$$\|DT_{p_1}^{-1}(x) - DT_{p_2}^{-1}(x)\|_F \leq \frac{L}{\theta^2}\|T_{p_1} - T_{p_2}\|_{1,\infty}$$

*für fast alle  $x \in \Omega$  und alle  $p_1, p_2 \in P$ .*

(d) *Es gilt*

$$\|\epsilon_p(u(x))\|_F^2 \leq \frac{2}{\theta}\|Du(x)\|_F^2$$

*für fast alle  $x \in \Omega$  und  $p \in P$ .*

(e) *Es existiert ein  $L > 0$  mit*

$$\|\epsilon_{p_1}(u) - \epsilon_{p_2}(u)\|_{0,2} \leq L\|T_{p_1} - T_{p_2}\|_{1,\infty}\|u\|_{1,2}$$

Dabei sind die Lipschitz-Konstanten  $L$  jeweils verschieden.

*Beweis:*

- (a) Es ist  $\|DT_p(x)\| = \|Id + DV_p(x)\| \leq \sqrt{3}(C+1)$  fast überall. Dabei resultiert  $\sqrt{3}$  aus der Äquivalenz der Normen im  $\mathbb{R}^n$ . Weil die Determinantenabbildung eine stetige Abbildung ist, nimmt sie nach dem Satz von Weierstraß auf der kompakten Menge  $\mathcal{M} = \{M \in \mathbb{R}^{3 \times 3} : \|A\| \leq \sqrt{3}(C+1)\}$  ein Maximum  $m$  an. Somit folgt die Behauptung fast überall.
- (b) Die Determinantenabbildung ist als Summe und Produkt stetig differenzierbarer Abbildungen lokal Lipschitz-stetig. Die globale Lipschitz-Stetigkeit folgt dann aus der Kompaktheit von  $\bar{\Omega}$  und der Beschränktheit von  $\mathcal{V}_p$ .
- (c) Es ist  $\|DT_{p_1}(x) - DT_{p_2}(x)\|_F = \left( \sum_{i,j=1}^3 ((DT_{p_1}(x) - DT_{p_2}(x))_{ij})^2 \right)^{1/2}$  fast überall. Dies können wir nach oben abschätzen zu

$$\begin{aligned} \left( \sum_{i,j=1}^3 (DT_{p_1}(x) - DT_{p_2}(x))_{ij}^2 \right)^{1/2} &\leq 3 \max_{i,j} |(DT_{p_1}(x) - DT_{p_2}(x))_{ij}| \\ &\leq 3 \|T_{p_1} - T_{p_2}\|_{1,\infty} \end{aligned}$$

fast überall. Die andere Ungleichung folgt aus der ersten wegen

$$DT_{p_1}^{-1}(x) - DT_{p_2}^{-1}(x) = -DT_{p_1}^{-1}(x) (DT_{p_1}(x) - DT_{p_2}(x)) DT_{p_2}^{-1}(x)$$

und der Submultiplikativität der Normen.

- (d) Dies folgt aus

$$\begin{aligned} \|\epsilon_p(u)(x)\|_F &= \left\| Du(x)DT_p^{-1}(x) + (Du(x)DT_p^{-1}(x))^T \right\|_F \\ &\leq \|Du(x)DT_p^{-1}(x)\|_F + \left\| (Du(x)DT_p^{-1}(x))^T \right\|_F \\ &\leq 2 \|DT_p^{-1}(x)\|_F \|Du(x)\|_F \leq \frac{2}{\theta} \|Du(x)\|_F. \end{aligned}$$

(e)

$$\begin{aligned}
& \|\epsilon_{p_1}(u) - \epsilon_{p_2}(u)\|_{0,2}^2 \\
&= \int_{\Omega} \left\| DT_{p_1}^{-1} Du + (DT_{p_1}^{-1} Du)^T - DT_{p_2}^{-1} Du - (DT_{p_2}^{-1} Du)^T \right\|_F^2 dx \\
&= \int_{\Omega} \left\| (DT_1^{-1} - DT_{p_2}^{-1}) Du + ((DT_1^{-1} - DT_{p_2}^{-1}) Du)^T \right\|_F^2 dx \\
&\leq 2 \int_{\Omega} \left\| (DT_1^{-1} - DT_{p_2}^{-1}) Du \right\|_F^2 dx \\
&\stackrel{(c)}{\leq} 2 \int_{\Omega} L \|T_{p_1} - T_{p_2}\|_{1,\infty}^2 \|Du(x)\|_F^2 dx \\
&= 2L \|T_{p_1} - T_{p_2}\|_{1,\infty}^2 \int_{\Omega} \|Du(x)\|_F^2 dx \\
&\leq 2L \|T_{p_1} - T_{p_2}\|_{1,\infty}^2 \|u\|_{1,2}.
\end{aligned}$$

■

Nachdem wir die Menge der zulässigen Gebiete charakterisiert haben, werden wir zu den eigentlichen Formoptimierungsproblemen übergehen. Wir untersuchen dazu die Lösung eines Kontaktproblems in Abhängigkeit des zugrunde liegenden Gebietes. Hierzu werden wir im nächsten Abschnitt ein Formoptimierungsproblem einführen, das als Nebenbedingung ein Kontaktproblem enthält, und die Existenz einer Lösung diskutieren. In den darauf folgenden Abschnitten werden wir das Kalkül der Formableitungen vorstellen und auf unser Problem anwenden.

## 4.2 Formoptimierungsprobleme

In diesem Abschnitt stellen wir das Formoptimierungsproblem auf. Wie oben bereits erläutert wurde, geht es darum, dass wir aus der Menge der zulässigen Gebiete, das auswählen, welches ein bestimmtes Zielfunktional minimiert. Außerdem werden wir die Existenz einer optimalen Lösung zeigen. Dafür müssen wir die Menge der zulässigen Gebiete mit einer Topologie versehen.

**Definition 4.3.** Sei  $P$  eine Menge von Formparametern und  $\mathcal{O}$  wie in (4.1). Ferner gelte Annahme 4.1. Sei  $(\Omega(p_n))_{n \in \mathbb{N}} \subset \mathcal{O}$ , eine Folge von zulässigen Gebieten. Wir nennen die Folge  $(\Omega(p_n))_{n \in \mathbb{N}}$  konvergent gegen  $\Omega(p)$ , falls

$$\lim_{n \rightarrow \infty} p_n = p$$

ist. Hierfür schreiben wir

$$\lim_{n \rightarrow \infty} \Omega(p_n) = \Omega(p).$$

Falls keine Missverständnisse zu befürchten sind, schreiben wir  $\Omega_n$  statt  $\Omega(p_n)$ .

Um die Konvergenz zu verdeutlichen, betrachten wir eine beliebige geschlossene Kurve  $\gamma : [0, 1] \rightarrow \Omega$  im Referenzgebiet  $\Omega$  mit  $\gamma(0) = \gamma(1)$ . Wir setzen  $\gamma_{p_n} = T_{p_n} \circ \gamma$ . Dann ist  $\gamma_{p_n}$  eine geschlossene Kurve in  $\Omega_{p_n}$ . Sie ist stetig, da  $T_{p_n}$  in  $H^{1,\infty}(\Omega)$  liegt und somit Lipschitz-stetig ist.

Wir zeigen, dass  $\gamma_{p_n}(t) \rightarrow \gamma_p(t)$  für alle  $t \in [0, 1]$  konvergiert. Weil  $\gamma$  beliebig war, sagen wir, dass die Gebiete  $\Omega(p_n)$  gegen  $\Omega(p)$  konvergieren.

**Satz 4.4.** *Sei  $\gamma : [0, 1] \rightarrow \Omega$  und  $(p_n)_{n \in \mathbb{N}}$  eine gegen  $p$  konvergente Folge von Gebietsparametern. Ferner gelte Annahme 4.1. Sei  $\gamma_{p_n} = T_{p_n} \circ \gamma$  und  $\gamma_p = T_p \circ \gamma$ , dann ist*

$$\lim_{n \rightarrow \infty} \max_{t \in [0, 1]} \|\gamma_p(t) - \gamma_{p_n}(t)\|_\infty = 0.$$

*Beweis:* Nach Annahme 4.1 ist die Abbildung  $P \subset \mathbb{R}^n \rightarrow W^{1,\infty}(\Omega)$  stetig. Somit existiert zu  $\epsilon > 0$  ein  $\delta > 0$ , so dass  $\|T_{p_n} - T_p\|_{1,\infty} = \|\mathcal{V}_{p_n} - \mathcal{V}_p\|_{1,\infty} < \epsilon$  für  $\|p_n - p\| < \delta$ . Nachdem  $\mathcal{V}_{p_n}$  und  $\mathcal{V}_p$  in  $W^{1,\infty}(\Omega)$  liegen, wählen wir Lipschitz-stetige Repräsentanten. Sei  $t \in [0, 1]$  beliebig, dann ist

$$\|\gamma_p(t) - \gamma_{p_n}(t)\|_\infty \leq L \|\mathcal{V}_{p_n} - \mathcal{V}_p\|_{1,\infty}.$$

Weil  $t$  beliebig war, können wir links zum Maximum übergehen und wegen der Konvergenz von  $p_n$  gegen  $p$  existiert ein  $N_\epsilon \in \mathbb{N}$ , so dass

$$\max_{t \in [0, 1]} \|\gamma_p(t) - \gamma_{p_n}(t)\|_\infty < \epsilon$$

für  $n > N_\epsilon$  gilt. Damit folgt die Behauptung. ■

Aus der Definition folgt unmittelbar:

**Satz 4.5.** *Die Menge  $\mathcal{O}$  ist genau dann kompakt, wenn  $P$  kompakt ist*

Zu jedem  $\Omega_p \in \mathcal{O}$  assoziieren wir den Funktionenraum  $\mathbb{V}(\Omega_p)$ , den Raum der reellen vektorwertigen Funktionen über  $\Omega_p$ . Als nächstes definieren wir Konvergenz von Funktionen, die zu  $\mathbb{V}(\Omega_p)$  für verschiedene  $\Omega_p \in \mathcal{O}$  gehören.

**Definition 4.6.** *Sei  $(\Omega_{p_n})_{n \in \mathbb{N}}$  eine Folge von Gebieten in  $\mathcal{O}$ , die gegen  $\Omega_p$  konvergiert. Wir sagen,  $(y_n)_{n \in \mathbb{N}}$  mit  $y_n \in \mathbb{V}(\Omega_{p_n})$  konvergiert gegen  $y \in \mathbb{V}(\Omega_p)$ , falls*

$$\lim_{n \rightarrow \infty} \|y_n \circ T_{p_n} - y \circ T_p\|_{\mathbb{V}(\Omega)} = 0 \quad (4.5)$$

*gilt. Wir schreiben  $y_n \rightsquigarrow y$ , um eine Verwechslung von Funktionenfolgen in einem festen Funktionenraum zu vermeiden.*

Wir formulieren nun unser Formoptimierungsproblem:

**Problem 4.7** (Formoptimierungsproblem). Seien  $\mathcal{O}$  und  $\mathbb{V}(\Omega)$  für  $\Omega \in \mathcal{O}$  gegeben. Ferner sei  $J : (\Omega, y) \mapsto J(\Omega, y) \in \mathbb{R}$  mit  $\Omega \in \mathcal{O}$  und  $y \in \mathbb{V}(\Omega)$  eine Zielfunktion. Dann nennen wir  $\Omega^* \in \mathcal{O}$  Lösung des Formoptimierungsproblems, falls

$$J(\Omega^*, u(\Omega^*)) \leq J(\Omega, u(\Omega)) \quad \forall \Omega \in \mathcal{O} \quad (4.6)$$

gilt, wobei  $u = u(\Omega)$  die Lösung des Kontaktproblems (2.32)–(2.34) ist. Dabei bezeichnet  $\mathbb{K}(\Omega)$  eine geeignet gewählte konvexe Teilmenge von  $\mathbb{V}(\Omega)$ .

Als mögliche Zielfunktion treten in der Formoptimierung die Gesamtverschiebung, die Gesamtspannung, die Gesamtenergie, die Nachgiebigkeit oder die reziproke Energie häufig auf.

Wir untersuchen nun die Abhängigkeit der Lösungen des regularisierten Kontaktproblems

$$C^{\Omega(p)}(u, q, z) = \begin{pmatrix} Au - F + B^*q \\ Bu - d + z - \gamma q \\ \gamma q + \min(0, d - Bu) \end{pmatrix} = 0_{H^{-1}(\Omega(p)) \times L^2(\Gamma_C(p)) \times L^2(\Gamma_C(p))} \quad (4.7)$$

in Abhängigkeit des Gebiets  $\Omega_p$ . Dabei wollen wir eine Lipschitz-stetig Abhängigkeit der Lösungen von den Formparametern zeigen. Dass keine höhere Regularität auftritt, liegt an der Variationsungleichung, wie folgendes Beispiel zeigt:

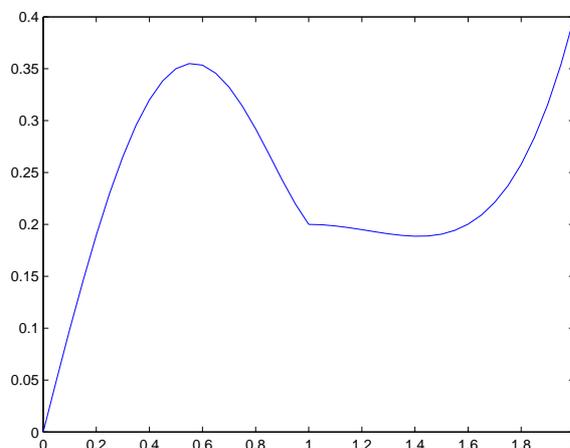


Abbildung 4.2: Graph des Zielfunktionalis aus Beispiel 4.8. Man erkennt den Knick bei  $\alpha = 1$ .

**Beispiel 4.8** (aus [HM03]). Wir betrachten das eindimensionale Formoptimierungsproblem

$$\begin{aligned} & \min_{\alpha \in [0.5, 2]} J(\alpha) := \int_0^\alpha (u_\alpha - 1)^2 dx \\ \text{mit} & \quad u_\alpha \in \mathbb{K}(\alpha), \\ & \int_0^\alpha u'_\alpha (v' - u'_\alpha) dx \geq \int_0^\alpha 2(v - u_\alpha) dx \quad \forall v \in \mathbb{K}(\alpha), \\ \text{für} & \quad \mathbb{K}(\alpha) = \{v \in H^1(]0, \alpha[) \mid v(0) = 0, v(\alpha) \leq 1\}. \end{aligned}$$

Die Lösung der Variationsungleichung lautet

$$u_\alpha(x) = \begin{cases} -x^2 + 2\alpha x & \alpha \leq 1, \\ -x^2 + \frac{\alpha^2+1}{\alpha}x, & \alpha \geq 1. \end{cases}$$

Setzt man diese in das Zielfunktional ein, so erhält man für  $J(\alpha)$  den analytischen Ausdruck

$$J(\alpha) = \begin{cases} \frac{8}{15}\alpha^5 - \frac{4}{3}\alpha^3 + \alpha, & \alpha < 1, \\ \frac{1}{30}\alpha^5 - \frac{1}{6}\alpha^3 + \frac{1}{3}\alpha, & \alpha \geq 1. \end{cases}$$

Das Zielfunktional (siehe auch Abbildung 4.2) ist zwar stetig, aber in 1 nicht differenzierbar. Weil die Abbildung

$$I : [0.5, 2] \times H^1(]0, \alpha[) \rightarrow \mathbb{R}, I(\alpha, y) \mapsto \int_0^\alpha (y - a)^2 dx$$

in beiden Argumenten stetig differenzierbar ist, muss die Abbildung  $\alpha \mapsto u_\alpha$  stetig, aber nicht differenzierbar sein.

Diese Situation tritt bei der parameterabhängigen Lösung von Variationsungleichungen häufig auf. Wir werden sehen, dass die Differenzierbarkeit genau an den Stellen verloren geht, an denen in den Optimalitätsbedingungen keine starke Komplementarität auftritt.

Wir zeigen nun, dass für die in dieser Arbeit behandelten Probleme die Lösung  $y(p)$  des Kontaktproblems Lipschitz-stetig von  $T_p$ ,  $p \in P$ , abhängt.

Hierzu untersuchen wir zunächst die Abhängigkeit der auftretenden Funktionale. Dabei betrachten wir zunächst das nicht regularisierte Kontaktproblem (2.29)–(2.31).

Wir zeigen nun, dass die Bilinearform  $a_p : \mathbb{V}(\Omega_p) \times \mathbb{V}(\Omega_p) \rightarrow \mathbb{R}$ , die durch

$$a_p(u_p, v_p) := \int_{\Omega_p} \tau(u_p) : \epsilon(v_p) dx_p, \quad u_p, v_p \in \mathbb{V}(\Omega_p), \quad (4.8)$$

gegeben ist, Lipschitz-stetig von  $T_p \in W^{1,\infty}(\Omega)$  abhängt. Dabei ist analog zu (2.8)

$$\mathbb{V}(\Omega_p) = \{v \in (H^1(\Omega_p))^3 : v|_{\Gamma_D} = 0\}.$$

Wir wissen, dass  $\Omega_p$  aus  $\Omega_0$  durch Anwendung der Transformation  $T_p$  hervorgeht. Mit der Transformationsformel für Integrale [Kön97] ist

$$\begin{aligned} a_p(u_p, v_p) &= \int_{\Omega_p} \tau(u_p) : \epsilon(v_p) dx_p \\ &= \int_{\Omega} \tau_{x_p}(u_p \circ T_p) : \epsilon_{x_p}(v_p \circ T_p) |\det DT_p| dx. \end{aligned} \quad (4.9)$$

Dabei bedeutet  $\epsilon_{x_p}$ , dass die Ableitungen bezüglich der alten Koordinaten  $x_p$  und nicht der neuen Koordinaten gebildet werden. Ferner sind  $u_p \circ T_p$  und  $v_p \circ T_p$  Elemente von  $\mathbb{V}(\Omega)$ , denn es gilt

$$\frac{1}{c} \|u\|_{1,\Omega_p}^2 \leq \|u \circ T_p\|_{1,\Omega}^2 \leq c \|u\|_{1,\Omega_p}^2 \quad (4.10)$$

mit einer von  $p$  unabhängigen Konstante  $c > 0$  nach [Dob06].

Wir setzen  $u^p := u_p \circ T_p$  und  $v^p := v_p \circ T_p$  und erhalten somit statt (4.9) die äquivalente Bilinearform

$$a^p(u^p, v^p) := \int_{\Omega} \tau_{x_p}(u^p) : \epsilon_{x_p}(v^p) | \det DT_p | dx \quad (4.11)$$

auf  $\mathbb{V}(\Omega)$ .

Mit der Kettenregel folgt

$$\epsilon_{x_p}(u^p) = \frac{1}{2} (Du^p DT_p^{-1} + (Du^p DT_p^{-1})^T). \quad (4.12)$$

Dies entspricht dem Ausdruck im Hilfssatz 4.2. Somit gilt für  $\tau$  nach (2.5)

$$(\tau_{x_p}(u^p))_{ij} = \sum_{k,l=1}^3 ((c_p)_{ijkl} \circ T_p) \cdot (\epsilon_{x_p}(u^p))_{kl}, \quad i, j = 1, 2, 3. \quad (4.13)$$

Wir nehmen an, dass die Punkte im verschobenen Gebiet die gleichen Materialeigenschaften haben wie die Referenzpunkte in  $\Omega$ . Deshalb gilt

$$(c_p)_{ijkl} \circ T_p = c_{ijkl} \text{ für alle } 1 \leq i, j, k, l \leq 3 \text{ und für alle } p \in P, \quad (4.14)$$

wobei  $c_{ijkl}$  wie in Annahme 2.1 in  $L^\infty$  sind. Also können wir

$$(\tau_{x_p}(u^p))_{ij} = \sum_{k,l=1}^3 c_{ijkl} (\epsilon_{x_p}(u^p))_{kl}, \quad i, j = 1, 2, 3 \quad (4.15)$$

schreiben. Wir nehmen an, dass (4.15) stetig ist, dass also ein  $M > 0$  existiert, so dass fast überall

$$\|\tau_{x_p}(u^p)(x)\|_F \leq M \|\epsilon_{x_p}(u^p)(x)\|_F \quad (4.16)$$

gilt und dass dieses  $M$  unabhängig von  $p$  ist.

Wir zeigen nun, dass für  $u$  und  $v$  mit  $\|u\|_{1,\Omega} \leq 1$  und  $\|v\|_{1,\Omega} \leq 1$  die Bilinearform  $a^p$  Lipschitz-stetig ist, dass also ein  $L > 0$  mit

$$|a^{p_1}(u, v) - a^{p_2}(u, v)| \leq L \|T_{p_1} - T_{p_2}\|_{1,\infty}, \quad \forall u, v \in \mathbb{V}; \|u\|_{1,\Omega} \leq 1, \|v\|_{1,\Omega} \leq 1$$

existiert.

Im Folgenden schreiben wir 1 statt  $p_1$  und 2 statt  $p_2$ . Außerdem verwenden wir für  $M_1, M_2 \in \mathbb{R}^{3 \times 3}$  den Ausdruck  $M_1 : M_2 = \sum_{i,j=1}^3 (M_1)_{ij} \cdot (M_2)_{ij}$  und zusätzlich für  $M_3 \in \mathbb{R}^{3 \times 3 \times 3 \times 3}$  die Identität  $M_1 : M_3 : M_2 = \sum_{ijkl=1}^3 (M_1)_{ij} \cdot (M_3)_{ijkl} \cdot (M_2)_{kl}$ .

Damit ist

$$\begin{aligned}
|a^1(u, v) - a^2(u, v)| &= \left| \int_{\Omega} \tau_1(u) : \epsilon_1(v) | \det DT_1| - \tau_2(u) : \epsilon_2(v) | \det DT_2| dx \right| \\
&\leq \left| \int_{\Omega} \tau_1(u) : \epsilon_1(v) | \det DT_1| - \tau_2(u) : \epsilon_1(v) | \det DT_1| dx \right| \\
&+ \left| \int_{\Omega} \tau_2(u) : \epsilon_1(v) | \det DT_1| - \tau_2(u) : \epsilon_2(v) | \det DT_1| dx \right| \\
&+ \left| \int_{\Omega} \tau_2(u) : \epsilon_2(v) | \det DT_1| - \tau_2(u) : \epsilon_2(v) | \det DT_2| dx \right|.
\end{aligned}$$

Dies lässt sich zusammenfassen zu

$$\begin{aligned}
|a^1(u, v) - a^2(u, v)| &\leq \int_{\Omega} |(\tau_1(u) - \tau_2(u)) : \epsilon_1(v)| \cdot |\det DT_1| dx \\
&+ \int_{\Omega} |\tau_2(u) : (\epsilon_1(v) - \epsilon_2(v))| \cdot |\det DT_1| dx \\
&+ \int_{\Omega} |\tau_2(u) : \epsilon_2(v)| \cdot \left| |\det DT_1| - |\det DT_2| \right| dx.
\end{aligned}$$

Wir schätzen nun die einzelnen Summanden ab.

Für den ersten Integranden des ersten Integrals gilt mit Satz A.1 (b) und (4.16)

$$|(\tau_1(u) - \tau_2(u)) : \epsilon_1(v)| \cdot |\det DT_1| \leq M \|\epsilon_1(u) - \epsilon_2(u)\|_F \|\epsilon_1(v)\|_F |\det DT_1|$$

fast überall. Wir integrieren dies und verwenden die Beschränktheit der Determinante nach Hilfssatz 4.2 und die Hölder-Ungleichung

$$\begin{aligned}
&\int_{\Omega} |(\tau_1(u) - \tau_2(u)) : \epsilon_1(v)| \cdot |\det DT_1| dx \\
&\leq M^* \left( \int_{\Omega} \|\epsilon_1(u) - \epsilon_2(u)\|_F^2 dx \right)^{1/2} \left( \int_{\Omega} \|\epsilon_1(v)\|_F^2 dx \right)^{1/2}.
\end{aligned}$$

Nach Hilfssatz 4.2 findet man nun eine Konstante  $\tilde{M}$  mit

$$\begin{aligned}
\int_{\Omega} |(\tau_1(u) - \tau_2(u)) : \epsilon_1(v)| |\det DT_1| dx &\leq \tilde{M} \|T_1 - T_2\|_{1,\infty} \|u\|_{1,2} \|v\|_{1,2} \\
&\stackrel{\|u\| \leq 1, \|v\| \leq 1}{\leq} \tilde{M} \|T_1 - T_2\|_{1,\infty}.
\end{aligned}$$

Den zweiten Integranden kann man fast überall analog abschätzen:

$$|\tau_2(u) : (\epsilon_1(v) - \epsilon_2(v))| \cdot |\det DT_1| \leq M \|\epsilon_2(u)\|_F \|\epsilon_1(v) - \epsilon_2(v)\|_F |\det DT_1|.$$

Somit folgt nach Integration und Anwendung der Hölder-Ungleichung wie oben

$$\int_{\Omega} |\tau_2(u) : (\epsilon_1(v) - \epsilon_2(v))| |\det DT_1| dx \leq L^* \|T_1 - T_2\|_{1,\infty}.$$

Nach Anwenden der umgekehrten Dreiecksungleichung gilt für den dritten Integranden fast überall

$$\begin{aligned} |\tau_2(u) : \epsilon_2(v)| \cdot \left| |\det DT_1| - |\det DT_2| \right| &\leq |\tau_2(u) : \epsilon_2(v)| \cdot |\det DT_1 - \det DT_2| \\ &\leq C^* \|\epsilon_2(u)\|_F \|\epsilon_2(v)\|_F \|T_{p_1} - T_{p_2}\|_{1,\infty}. \end{aligned}$$

Dabei haben wir die Stetigkeit von  $\tau$  ausgenutzt. Integration über  $\Omega$  und die Hölder-Ungleichung liefern

$$\begin{aligned} &\int_{\Omega} |\tau_2(u) : \epsilon_2(v)| \cdot \left| |\det DT_1| - |\det DT_2| \right| dx \\ &\leq C^* \int_{\Omega} \|\epsilon_2(u)\|_F \|\epsilon_2(v)\|_F \|T_{p_1} - T_{p_2}\|_{1,\infty} dx \\ &\leq C^* \|T_{p_1} - T_{p_2}\|_{1,\infty} \left( \int_{\Omega} \|\epsilon_2(u)\|_F^2 dx \right)^{1/2} \left( \int_{\Omega} \|\epsilon_2(v)\|_F^2 dx \right)^{1/2} \\ &\leq L^* \|T_{p_1} - T_{p_2}\|_{1,\infty}. \end{aligned} \tag{4.17}$$

Somit haben wir gezeigt, dass für festes  $\|u\|_{1,\Omega} \leq 1$  und  $\|v\|_{1,\Omega} \leq 1$  die Bilinearform  $a^p(u, v)$  Lipschitz-stetig in  $T_p$  ist.

Als nächstes betrachten wir die parametrisierte rechte Seite

$$F_p(u_p) := \int_{\Omega_p} f_p u_p dx + \int_{\Gamma_{F_p}} P_p u_p ds.$$

Auch hier transformieren wir auf  $\Omega$  und erhalten analog zu oben

$$F^p(u^p) = \int_{\Omega} f^p u^p |\det DT_p| dx + \int_{\Gamma_P} P^p u^p \sqrt{g_p} ds \tag{4.18}$$

mit der Gramschen Determinante  $g_p$  und

$$\begin{aligned} f^p : \Omega &\rightarrow \mathbb{R}^3, & x &\mapsto f^p(x) &:= f_p \circ T_p(x), \\ P^p : \Gamma_p &\rightarrow \mathbb{R}^3, & x &\mapsto P^p(x) &:= P_p \circ T_p(x). \end{aligned} \tag{4.19}$$

Nach Annahme 4.1 ist die Einschränkung von  $T_p$  auf den Neumann-Rand  $\Gamma_P$  die Identität  $T_p|_{\Gamma_P} = Id$ , also folgt

$$F^p(u^p) = \int_{\Omega} f^p u^p |\det DT_p| dx + \int_{\Gamma} P u^p ds, \tag{4.20}$$

wobei  $P$  für alle zulässigen Gebiete dieselbe Neumann-Bedingung beschreibt.

Falls  $f^p$  Lipschitz-stetig in  $T_p$  ist, dann liegt mit der obigen Glattheitsforderung an  $T_p$  eine Lipschitz-Abhängigkeit von  $F^p$  bezüglich  $T_p$  vor.

Nun zeigen wir die gleichmäßige Elliptizität von  $a^p$ . Dazu wählen wir  $u^p \in \mathbb{V}(\Omega)$  beliebig. Dann existiert ein  $u_p \in \mathbb{V}(\Omega_p)$  mit  $u^p := u_p \circ T_p$ . Also ist

$$a^p(u^p, u^p) = a_p(u_p, u_p) \geq q \|\epsilon_{x_p}(u_p)\|_{0,\Omega_p}^2 \geq qq_0 \|u_p\|_{1,\Omega_p}^2. \tag{4.21}$$

Dabei haben wir Annahme 2.1 und die Kornsche Ungleichung C.2 verwendet und das nicht-triviale Resultat ausgenutzt, dass die Konstante  $q_0$  aus der Kornschen Ungleichung gleichmäßig bezüglich aller Gebiete ist, die einer gleichmäßigen Kegelbedingung genügen [Nit81]. Somit ist die Konstante  $q_0$  unabhängig von  $p$ , falls alle zulässigen Gebiete einer gleichmäßigen Kegelbedingung genügen.

Also können wir mit (4.10) schließen, dass es unabhängig von  $p$  ein  $q > 0$  gibt mit

$$a^p(u^p, u^p) \geq q \|u^p\|_{1,\Omega}^2 \quad \text{für alle } u^p \in \mathbb{V}(\Omega). \quad (4.22)$$

Zuletzt zeigen wir noch, dass die Lösungen des nicht regularisierten Kontaktproblems in Abhängigkeit von den Gebieten beschränkt sind, falls  $\|F^p\|_{H^{-1}}$  für alle  $p \in P$  gleichmäßig beschränkt ist. Dann gilt mit (4.22) für  $p \in P$  und der Lösung  $u^p$  des Kontaktproblems

$$q \|u^p\|_{1,\Omega}^2 \leq a^p(u^p, u^p) = a^p(u^p, u^p - 0) \leq F^p(u^p - 0) \leq \|F^p\|_{H^{-1}} \|u^p\|_{1,\Omega}. \quad (4.23)$$

Dabei haben wir ausgenutzt, dass  $v = 0$  eine zulässig Verschiebung ist, da das nichtverformte Gebiet keine Selbstdurchdringung aufweist.

Nach diesen Vorarbeiten können wir nun den folgenden Satz beweisen:

**Satz 4.9.** *Sei  $P$  die Menge der Formparameter, und die Menge der zulässigen Gebiete  $\mathcal{O}$  erfülle eine gleichmäßige Kegelbedingung. Ferner gelten die Annahmen 4.1. Die Abbildungen  $f^p$  aus (4.19) seien Lipschitz-stetig und  $\|F^p\|_{H^{-1}}$  sei ebenso gleichmäßig beschränkt in  $p \in P$  wie (4.15). Dann ist die Lösungsabbildung des nicht regularisierten Kontaktproblems (2.29)–(2.31)*

$$\begin{aligned} u &: H^{1,\infty}(\Omega) \rightarrow \mathbb{V}(\Omega), \\ T_p &\mapsto u(p) \end{aligned}$$

*Lipschitz-stetig.*

*Beweis:* Seien  $p_1, p_2 \in P$  zwei Formparameter. Dann bezeichnen  $u^{p_1}$  und  $u^{p_2}$  aus  $\mathbb{V}(\Omega)$  die Lösungen des Kontaktproblems zu den Parametern. Somit sind die folgenden Variationsungleichungen erfüllt:

$$\begin{aligned} a^{p_1}(u^{p_1}, v - u^{p_1}) &\geq F^{p_1}(v - u^{p_1}) \quad \forall v \in \mathbb{K}(\Omega), \\ a^{p_2}(u^{p_2}, v - u^{p_2}) &\geq F^{p_2}(v - u^{p_2}) \quad \forall v \in \mathbb{K}(\Omega). \end{aligned}$$

Wir setzen in der ersten Gleichung  $v = u^{p_2}$  und in der zweiten Gleichung  $v = u^{p_1}$  und addieren die Gleichungen. Somit erhalten wir

$$a^{p_1}(u^{p_1}, u^{p_2} - u^{p_1}) + a^{p_2}(u^{p_2}, u^{p_1} - u^{p_2}) \geq F^{p_1}(u^{p_2} - u^{p_1}) + F^{p_2}(u^{p_1} - u^{p_2}).$$

Wir sortieren um und addieren auf beiden Seiten  $a^{p_1}(u^{p_2}, u^{p_2} - u^{p_1})$  und erhalten dann unter Verwendung der gleichmäßigen Elliptizität der Bilinearform und der Lipschitz-Stetigkeit von Linear- und Bilinearform

$$\begin{aligned} & a^{p_1}(u^{p_2} - u^{p_1}, u^{p_2} - u^{p_1}) \\ \leq & a^{p_1}(u^{p_2}, u^{p_2} - u^{p_1}) - a^{p_2}(u^{p_2}, u^{p_2} - u^{p_1}) - F^{p_1}(u^{p_2} - u^{p_1}) - F^{p_2}(u^{p_1} - u^{p_2}). \end{aligned}$$

Also ist

$$\begin{aligned}
q \|u^{p_1} - u^{p_2}\|_{1,\Omega}^2 &\leq a^{p_1}(u^{p_2} - u^{p_1}, u^{p_2} - u^{p_1}) \\
&\leq a^{p_1}(u^{p_2}, u^{p_2} - u^{p_1}) - a^{p_2}(u^{p_2}, u^{p_2} - u^{p_1}) \\
&\quad + F^{p_1}(u^{p_1} - u^{p_2}) + F^{p_2}(u^{p_1} - u^{p_2}) \\
&\leq \|T_{p_1} - T_{p_2}\|_{1,\infty} \|u^{p_1} - u^{p_2}\|_{1,\Omega} (L_a \|u^{p_2}\|_{1,\Omega} + L_F)
\end{aligned}$$

Weil  $\|u^{p_1}\|_{1,\Omega}$  in  $p$  wegen (4.23) gleichmäßig beschränkt ist, folgt die Behauptung. ■

Wir können nun zeigen, dass das Formoptimierungsproblem eine Lösung besitzt.

**Satz 4.10.** *Sei  $P$  eine kompakte Menge und die Abbildung  $p \mapsto T_p \in W^{1,\infty}(\Omega)$  stetig. Ferner erfüllen die zulässigen Gebiete eine gleichmäßige Kegelbedingung. Es gelte Annahme 4.1 und die Abbildungen  $f_p$  aus (4.19) seien Lipschitz-stetig und  $\|F^p\|_{H^{-1}}$  ebenso gleichmäßig beschränkt für  $p \in P$  wie (4.15). Falls das Zielfunktional  $J$  bezüglich  $\Omega_p = T_p(\Omega)$  und der Lösung  $u(p) \in \mathbb{V}(\Omega)$  des auf  $\Omega$  zurücktransformierten Kontaktproblems (2.29)–(2.31) stetig ist, dann besitzt das nicht regularisierte Formoptimierungsproblem eine Lösung.*

*Beweis:* Weil die Lösung  $u(p)$  des Kontaktproblems stetig von  $p$  abhängt, genügt es, eine minimierende Folge  $(p_n)_{n \in \mathbb{N}}$  in  $P$  zu betrachten. Nachdem  $P$  kompakt ist, besitzt sie eine konvergente Teilfolge, die wegen der Stetigkeit von  $J$  gegen eine Lösung des Kontaktproblems konvergiert. ■

**Bemerkung 4.11.** *Für die Existenz genügt die Betrachtung unterhalb stetiger Zielfunktionale anstatt stetiger Zielfunktionale.*

Betrachten wir nun das regularisierte Problem, so suchen wir Lösungen  $q$  der Variationsungleichung

$$a_\gamma^p(q, r - q) \geq f^p(r - q) \quad \forall r \in K_2^+$$

in  $K_2^+$  mit  $K_2^+ := \{q \in L^2(\Gamma_C) : q \geq 0\}$  mit der gebietsabhängigen regularisierten Bilinearform  $a_\gamma^p(q, r) := (q, (B(A^p)^{-1}B^* + \gamma Id)r)_{L^2(\Gamma_C)}$  und der ebenfalls gebietsabhängigen Linearform  $f(q) := (q, B(A^p)^{-1}F^p - d)_{L^2(\Gamma_C)}$ . Die Einbettungsoperatoren von  $H^{1/2}(\Gamma_C) \rightarrow L^2(\Gamma_C)$  wurden weggelassen. Wie oben können wir nun zeigen, dass die Lösung  $q(\Omega)$  des Formoptimierungsproblems lokal Lipschitz-stetig von den Gebietsverformungen abhängt. Hierzu muss zusätzlich zur bereits oben gezeigten Lipschitz-Stetigkeit von  $a^p(\cdot, \cdot)$ , aus der die lokale Lipschitz-Stetigkeit von  $A^{-1}$  folgt, noch die Lipschitz-Stetigkeit von  $B$ ,  $B^*$  und  $f$  gezeigt werden. Dies erfolgt analog für Gebietstransformationen, die Anforderung 4.1 erfüllen.

Hiermit erhält man den Satz:

**Satz 4.12.** *Sei  $P$  die Menge der Formparameter und die Menge der zulässigen Gebiete  $\mathcal{O}$  erfülle eine gleichmäßige Kegelbedingung. Ferner gelten die Annahmen 4.1.*

Die Abbildungen  $f^p$  aus (4.19) seien Lipschitz-stetig und  $\|F^p\|_{H^{-1}}$  ebenso gleichmäßig beschränkt in  $p \in P$  wie (4.15). Dann ist die Lösungsabbildung des regularisierten Kontaktproblems (4.7)

$$u : H^{1,\infty}(\Omega) \rightarrow \mathbb{V}(\Omega),$$

$$T_p \mapsto u(p)$$

Lipschitz-stetig.

*Beweis:* Analog zu oben beschrieben zeigt man, dass  $T_p \mapsto q(\Omega_p)$  lokal Lipschitz-stetig ist. Der Rest folgt aus der Lipschitz-Stetigkeit von  $a_\gamma^p$  und der Koerzitivität von  $a_\gamma^p$ . ■

Wir zeigen nun einen allgemeinen lokalen Lipschitz-stetigen Zusammenhang zwischen den Gebietsverformungen und den Lösungen des Formoptimierungsproblems

**Satz 4.13.** *Sei  $P$  eine kompakte Menge und die Abbildung  $p \mapsto T_p \in W^{1,\infty}(\Omega)$  stetig. Ferner erfüllen die zulässigen Gebiete eine gleichmäßige Kegelbedingung. Es gelte Annahme 4.1 und die Abbildungen  $f_p$  aus (4.19) seien Lipschitz-stetig und  $\|F^p\|_{H^{-1}}$  ebenso gleichmäßig beschränkt für  $p \in P$  wie (4.15). Falls das Zielfunktional  $J$  bezüglich  $\Omega_p = T_p(\Omega)$  und der Lösung  $u(p) \in \mathbb{V}(\Omega)$  des auf  $\Omega$  zurücktransformierten Kontaktproblems 4.7 stetig ist, dann besitzt das regularisierte Formoptimierungsproblem eine Lösung.*

*Beweis:* Analog zu Satz 4.10, da lokal Lipschitz-stetige Abbildungen stetig sind. ■

### 4.3 Berechnung von Formableitungen

In diesem Abschnitt möchten wir berechnen, wie sich das Zielfunktional  $J(\Omega, u(\Omega))$  in Abhängigkeit von  $\Omega$  verändert. Wir werden uns nicht direkt dem Kalkül der Shapeableitungen bedienen, wie er in [SZ92b] vorgestellt wird.

Wie bereits im vorangegangenen Abschnitten befassen wir uns mit Gebieten, die über eine  $W^{1,\infty}$ -Transformation  $T$  aus einem Referenzgebiet  $\Omega$  hervorgehen. Wir nehmen weiter an, dass  $T$  die Bauart  $T(x) = x + \mathcal{V}(x)$  hat.

Wir wollen nun untersuchen, inwiefern wir für verschiedene Funktionale eine Entwicklung der Bauart

$$E(T(\Omega)) = E(\Omega) + \delta E(\Omega) \cdot \mathcal{V} + o(\|\mathcal{V}\|_{1,\infty})$$

schreiben können. Diese Notation ist bisher nur angedeutet und daher mit Vorsicht zu genießen.

Wir werden untersuchen, inwiefern dieses  $\delta E$  von der Richtung der Gebietsvariation  $\mathcal{V}$  abhängt, also ob eine Fréchet- oder Gateaux-Differenzierbarkeit vorliegt.

Wir behandeln im Folgenden die Differenzierbarkeit ausgewertet an der Stelle des Gebietes  $\Omega$ . Falls man die Ableitungen in anderen Gebieten berechnen will, kann man dies durch Transformationen auf diesen Fall zurückführen.

Bei der Transformation von Gebieten treten die Ableitungen der Transformation und ihre Determinante auf. Wir untersuchen deshalb zunächst, wie sich diese in Abhängigkeit von der Transformation verhalten.

Wir entwickeln nun  $DT^{-1}$  und  $\det(DT)$  in eine Taylorreihe.

**Hilfssatz 4.14.** *Sei  $T$  eine  $C^1$ -Abbildung der Bauart  $T(x) = x + \mathcal{V}(x)$ , wobei  $\|\mathcal{V}\|_{W^{1,\infty}(\Omega)} < 1/2$  ist. Dann gilt*

$$(a) \quad \|DT^{-1} - Id + D\mathcal{V}\|_{L^\infty} = o_{L^\infty}(\|\mathcal{V}\|_{1,\infty}),$$

$$(b) \quad \left\| |\det(DT)| - 1 - \operatorname{div}(\mathcal{V}) \right\|_{L^\infty} = o_{L^\infty}(\|\mathcal{V}\|_{1,\infty}).$$

*Beweis:* (a) Die geometrische Reihe auf Banach-Algebren konvergiert für reelle quadratische Matrizen  $A$  mit Frobenius-Norm kleiner 1 [Kön97],

$$(Id - A)^{-1} = \sum_{n=0}^{\infty} A^n. \quad (4.24)$$

Also folgt

$$\|(Id - A)^{-1} - Id - A\|_F = \left\| \sum_{n=2}^{\infty} A^n \right\|_F \leq \frac{\|A\|_F^2}{1 - \|A\|_F}. \quad (4.25)$$

Betrachten wir nun  $DT = Id + D\mathcal{V}$ , so gilt fast überall

$$\begin{aligned} \|(DT(x))^{-1} - x + D\mathcal{V}(x)\|_F &= \|(Id + D\mathcal{V})^{-1}(x) - x + D\mathcal{V}(x)\|_2 \\ &\leq \frac{\|D\mathcal{V}(x)\|_2^2}{1 - \|D\mathcal{V}(x)\|_2}. \end{aligned}$$

Also gilt die Abschätzung auch, wenn wir auf beiden Seiten zum wesentlichen Supremum übergehen, und wir erhalten unter Berücksichtigung der Äquivalenz der Normen im  $\mathbb{R}^n$ , dass

$$\|DT^{-1} - Id + D\mathcal{V}\|_{L^\infty} = o(\|\mathcal{V}\|_{1,\infty}) \quad (4.26)$$

gilt.

(b) Wir zeigen zunächst, dass für eine Matrix  $A \in \mathbb{R}^{3 \times 3}$

$$|\det(Id + A) - 1 - \operatorname{spur}(A)| \leq \sum_{n=2}^3 \binom{3}{n} \|A\|_2^n$$

ist. Wir benennen die Spaltenvektoren von  $A$  mit  $A_i$  und die Einheitsvektoren mit  $e_i$ , dann ist unter Berücksichtigung der Multilinearität der Determinante

$$\begin{aligned} \det(Id + A) &= \det(e_1 + A_1 \ e_2 + A_2 \ e_3 + A_3) \\ &= \det(e_1 \ e_2 \ e_3) \\ &\quad + \det(A_1 \ e_2 \ e_3) + \det(e_1 \ A_2 \ e_3) + \det(e_1 \ e_2 \ A_3) \\ &\quad + \det(A_1 \ A_2 \ e_3) + \det(A_1 \ e_2 \ A_3) + \det(e_1 \ A_2 \ A_3) \\ &\quad + \det(A). \end{aligned}$$

Somit ist

$$\begin{aligned} \det(Id + A) - 1 - \text{spur}(A) &= \det(A_1 \ A_2, e_3) + \det(A_1 \ e_2, A_3) \\ &\quad + \det(e_1 \ A_2, A_3) + \det(A). \end{aligned}$$

Weil in  $\mathbb{R}^3$  die Darstellung  $\det(a_1 \ a_2 \ a_3) = (a_1 \times a_2) \cdot a_3$  mit dem Vektorprodukt  $\times$  und dem Skalarprodukt  $\cdot$  gilt, folgt die Abschätzung

$$|\det(Id + A) - 1 - \text{spur}(A)| \leq 3\|A\|_F^2 + \|A\|_F^3, \quad (4.27)$$

wobei  $\|A_i\|_2 \leq \|A\|_F$  für  $1 \leq i \leq 3$  verwendet wurde. Also gilt

$$|\det(Id + A) - 1 - \text{spur}(A)| \leq \|A\|_F^2 \sum_{k=2}^3 \binom{3}{k} \|A\|_F^{k-2}. \quad (4.28)$$

Für  $\|\mathcal{V}\|_{1,\infty}$  klein ist  $\det(Id + D\mathcal{V}(x))$  fast überall positiv, so dass man die Betragsstriche weglassen kann. Übergang zum Supremum in (4.28) fast überall liefert die Behauptung.

■

Wir berechnen nun zunächst die Veränderung des Komplementaritätsproblems in Abhängigkeit vom Gebiet. Zur Vereinfachung betrachten wir, wie bereits in Annahme 4.1 angegeben, nur solche Transformationen, die den Kontaktrand und den Rand, auf den die Kraft wirkt, auf sich selbst abbilden.

Für die Familie der Gebietstransformationen  $T_p = Id + \mathcal{V}_p$ ,  $p \in P$  schreiben wir im Folgenden  $T_{\mathcal{V}} = Id + \mathcal{V}$ ,  $\mathcal{V} \in \{\mathcal{V} \in H^{1,\infty}(\Omega), \mathcal{V} = \mathcal{V}_p, p \in P\}$ , falls keine Mißverständnisse zu befürchten sind. Ferner setzen wir  $Y = \mathbb{V} \times L^2(\Gamma_C) \times L^2(\Gamma_C)$  und bezeichnen den zugehörigen Dualraum mit  $Y^*$ .

Wir transportieren das regularisierte Kontaktproblem (4.7) auf  $\Omega_{\mathcal{V}}$  über auf ein Kontaktproblem im Referenzraum  $\Omega$ . Dies wird beschrieben durch die Abbildung

$$C : (u, q, z, \mathcal{V}) \in Y \times H^{1,\infty} \mapsto Y^*, \quad (4.29)$$

$$\begin{aligned} \langle C(u, q, z, \mathcal{V}), (v, r, t) \rangle_{Y^*, Y} &= a^{\mathcal{V}}(u, v) - F^{\mathcal{V}}(v) - (q, Bv)_{L^2(\Gamma_C)} \\ &\quad + (Bu - d - \gamma q + z, r)_{L^2(\Gamma_C)} \\ &\quad + (\gamma q + \min(0, d - Bu))_{L^2(\Gamma_C)} \end{aligned} \quad (4.30)$$

für alle  $(v, r, t) \in Y$ . Wir haben an dieser Stelle wieder den Buchstaben  $C$  verwendet, da es sich um eine Erweiterung der Funktion  $C$  aus Kapitel Drei handelt. Aus dem Zusammenhang sollte immer ersichtlich sein, welche Funktion gemeint ist, so dass wir hier keine Missverständnisse befürchten.

Falls nun  $(u, q, z) \in Y$  die Lösung von  $C(u, q, z, \mathcal{V}) = 0$  ist, so ist die transportierte Funktion  $(u \circ T_{\mathcal{V}}^{-1}, q \circ T_{\mathcal{V}}^{-1}, z \circ T_{\mathcal{V}}^{-1})$  Lösung des regularisierten Kontaktproblems auf  $\Omega_{\mathcal{V}}$ .

**Satz 4.15.** *Sei  $\Omega$  ein Referenzgebiet und  $P$  eine Menge von Gebietsparametern, so bezeichne  $\mathcal{O}$  die Menge der zulässigen Gebiete, die durch Transformation  $T_{\mathcal{V}} : Id + \mathcal{V}$  aus  $\Omega$  hervorgehen. Ferner existiere eine beschränkte Menge  $\hat{\Omega}$  mit  $\Omega_{\mathcal{V}} \subset \hat{\Omega}$  für alle  $\Omega_{\mathcal{V}} \in \mathcal{O}$ . Außerdem sei  $f$  bezüglich der  $W^{1,\infty}(\hat{\Omega})$ -Norm Taylor entwickelbar und es gelte Annahme 4.1. Seien zusätzlich  $(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}))$  die Lösung von  $C(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}), \mathcal{V}) = 0$  und  $(u(0), q(0), z(0))$  Lösungen von  $C(u(0), q(0), z(0), 0)$ . Dann existiert eine mengenwertige Abbildung  $\partial_{y,\mathcal{V}}C : \mathcal{V} \mapsto L(Y \times W^{1,\infty}(\Omega), Y^*)$  mit*

$$\begin{aligned} \sup_{C_{y,\mathcal{V}} \in \partial_{y,\mathcal{V}}C} \left\| \left\| C(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}), \mathcal{V}) - C(u(0), q(0), z(0), 0) - C_{y,\mathcal{V}} \begin{pmatrix} u(\mathcal{V}) - u(0) \\ q(\mathcal{V}) - q(0) \\ z(\mathcal{V}) - z(0) \\ \mathcal{V} \end{pmatrix} \right\| \right\|_{Y^*} \\ = o \left( \left\| \begin{pmatrix} u(\mathcal{V}) - u(0) \\ q(\mathcal{V}) - q(0) \\ z(\mathcal{V}) - z(0) \end{pmatrix} \right\|_Y \right) + o(\|\mathcal{V}\|_{1,\infty}), \end{aligned}$$

wobei  $C_{y,\mathcal{V}}$  die Bauart

$$C_{y,\mathcal{V}} = \begin{pmatrix} A^0 & B^* & 0 & \dot{A}(u(0), \cdot) + \dot{F}(\cdot) \\ B & -\gamma Id & Id & 0 \\ -D(Bu(p))B & \gamma Id & 0 & 0 \end{pmatrix} \quad (4.31)$$

hat. Dabei sind  $\dot{A}(u(0), \Omega)$  und  $\dot{F}(\Omega)$  die Gradienten der Bilinearform  $a$  und der Linearform  $F$  bezüglich der Gebietsverformung  $\mathcal{V}$ , welche im Beweis erläutert werden.  $D$  bezeichnet die Elemente aus dem in (3.13) definierten Differential.

*Beweis:* Wir betrachten zunächst

$$\begin{aligned} 0 &= C(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}), \mathcal{V}) - C(u(0), q(0), z(0), 0) \\ &= C(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}), \mathcal{V}) - C(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}), 0) \\ &\quad + C(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}), 0) - C(u(0), q(0), z(0), 0). \end{aligned}$$

Wir werden nun die Summanden einzeln abschätzen.

Für die zweite Differenz verwenden wir Satz 3.20. Weil  $C$  semiglatt ist, gilt

$$\begin{aligned}
& \|C(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}), 0) - C(u(0), q(0), z(0), 0)\|_Y \\
&= \left\| \begin{pmatrix} A^0 u(\mathcal{V}) - F^0 + B^* q(\mathcal{V}) - (A^0 u(0) - F^0 + B^* q(0)) \\ Bu(\mathcal{V}) - d - \gamma q(\mathcal{V}) + z(\mathcal{V}) - (Bu(0) - d - \gamma q(0) + z(0)) \\ \gamma q(\mathcal{V}) + \min(0, d - Bu(\mathcal{V})) - \gamma q(0) - \min(0, d - Bu(0)) \end{pmatrix} \right\|_{Y^*} \\
&= \left\| \begin{pmatrix} A^0 & B^* & 0 \\ B & -\gamma Id & Id \\ -D(Bu(\mathcal{V}))B & \gamma Id & 0 \end{pmatrix} \begin{pmatrix} u(\mathcal{V}) - u(0) \\ q(\mathcal{V}) - q(0) \\ z(\mathcal{V}) - z(0) \end{pmatrix} \right\|_{Y^*} \\
&\quad + o(\|u(\mathcal{V}) - u(0)\|_{H^1(\Omega)}).
\end{aligned}$$

Dabei bezeichnet  $D$  ein beliebiges Element aus dem in (3.13) definierten Differential. Das Element ist im Fall strikter Komplementarität eindeutig bestimmt.

Wir schätzen nun den ersten Summanden ab, dabei schreiben wir anstatt des Arguments  $\mathcal{V}$  einen Index  $\mathcal{V}$ . Für  $\mathcal{V} = 0$  lassen wir das Argument wegfallen und schreiben zum Beispiel  $u$  statt  $u(0)$ .

Sei im Folgenden  $(v, r, s) \in Y$  beliebig, aber fest. Dann wenden wir

$$C(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}), \mathcal{V}) - C(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}), 0)$$

auf  $(v, r, s)$  an und erhalten:

$$\begin{aligned}
& a^\mathcal{V}(u^\mathcal{V}, v) - F^\mathcal{V}(v) + (q^\mathcal{V}, Bv)_{L^2(\Gamma_C)} - a^0(u^\mathcal{V}, v) + F_0(v) - (q^\mathcal{V}, Bv)_{L^2(\Gamma_C)} \\
&= \int_\Omega \tau^\mathcal{V}(u^\mathcal{V}) : \epsilon^\mathcal{V}(v) \det(DT) dx - \int_\Omega f \circ T v \det(DT) dx + \int_\Gamma P v dS \\
&\quad - \left( \int_\Omega \tau(u^\mathcal{V}) : \epsilon(v) dx - \int_\Omega f v dx + \int_\Gamma P v dS \right) \\
&= \int_\Omega (Du^\mathcal{V} DT^{-1} + (Du^\mathcal{V} DT^{-1})^T) : C : (Dv DT^{-1} + (Dv DT^{-1})^T) \det(DT) \\
&\quad - (Du^\mathcal{V} + Du^\mathcal{V})^T : C : (Dv + (Dv)^T) dx \\
&\quad + \int_\Omega (f \circ T \det(DT) - f) v dx,
\end{aligned}$$

wobei wir daran erinnern möchten, dass das Funktional

$$C(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}), \mathcal{V}) - C(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}), 0)$$

die Bauart  $\begin{pmatrix} * \\ 0 \\ 0 \end{pmatrix}$  hat, Dabei haben wir die Forderung an  $T$  ausgenutzt, dass die Punkte auf der Kontaktfläche und der Neumann-Randbedingung nicht geändert

werden. Wir wenden nun Hilfssatz 4.14 an, und es folgt:

$$\begin{aligned}
& \int_{\Omega} (Du^{\mathcal{V}}DT^{-1} + (Du^{\mathcal{V}}DT^{-1})^T) : C : (DvDT^{-1} + (DvDT^{-1})^T) \det(DT) \\
& - (Du^{\mathcal{V}} + (Du^{\mathcal{V}})^T) : C : (Dv + (Dv)^T) dx \\
& + \int_{\Omega} (f \circ T \det(DT) - f) v dx \\
= & \int_{\Omega} (Du^{\mathcal{V}})(Id - D\mathcal{V} + o_{L^\infty}(\|\mathcal{V}\|_{1,\infty})) + (Du^{\mathcal{V}}(Id - D\mathcal{V} + o_{L^\infty}(\|\mathcal{V}\|_{1,\infty})))^T \\
& : C : (Dv(Id - D\mathcal{V} + o_{L^\infty}(\|\mathcal{V}\|_{1,\infty}))) + (Dv(Id - D\mathcal{V} + o_{L^\infty}(\|\mathcal{V}\|_{1,\infty})))^T \\
& (1 + \operatorname{div}(\mathcal{V}) + o_{L^\infty}(\|\mathcal{V}\|_{1,\infty})) \\
& - (Du^{\mathcal{V}} + (Du^{\mathcal{V}})^T) : C : (Dv + (Dv)^T) dx \\
& + \int_{\Omega} (f(x + \mathcal{V}(x)) (1 + \operatorname{spur}(D\mathcal{V}) + o_{L^\infty}(\|\mathcal{V}\|_{1,\infty}) - f(x)) v(x) dx. \quad (4.32)
\end{aligned}$$

Wir betrachten nun den Ausdruck

$$\int_{\Omega} Du^{\mathcal{V}}D\mathcal{V} : C : Dv dx = \int_{\Omega} DuD\mathcal{V} : C : Dv dx + \int_{\Omega} (Du^{\mathcal{V}} - Du)D\mathcal{V} : C : Dv dx.$$

Für den zweiten Summanden nutzen wir nun die lokale Lipschitz-Stetigkeit nach Satz 4.12 von  $u(\mathcal{V})$  aus. Dies liefert

$$\begin{aligned}
\left| \int_{\Omega} (Du^{\mathcal{V}} - Du)D\mathcal{V} : C : Dv dx \right| & \leq C \|D\mathcal{V}\|_{\infty} \|Du - Du^{\mathcal{V}}\|_{L^2} \|v\|_{L^2} \\
& \leq C \|\mathcal{V}\|_{1,\infty} \|u - u^{\mathcal{V}}\|_{\mathbb{V}} \|v\|_{\mathbb{V}} \\
& = o_{L^\infty}(\|\mathcal{V}\|_{1,\infty}).
\end{aligned}$$

Wir setzen dies in (4.32) ein und erhalten

$$\begin{aligned}
& a^{\mathcal{V}}(u^{\mathcal{V}}, v) - F^{\mathcal{V}}(v) + (q^{\mathcal{V}}, Bv)_{L^2(\Gamma_C)} - a^0(u^{\mathcal{V}}, v) + F_0(v) - (q^{\mathcal{V}}, Bv)_{L^2(\Gamma_C)} \\
= & \int_{\Omega} - (Du^{\mathcal{V}}D\mathcal{V} + (Du^{\mathcal{V}}D\mathcal{V})^T) : C : (Dv + (Dv)^T) \\
& + (DuD\mathcal{V} + (DuD\mathcal{V})^T) : C : (Dv + (Dv)^T) \\
& - (DuD\mathcal{V} + (DuD\mathcal{V})^T) : C : (Dv + (Dv)^T) \\
& - (Du^{\mathcal{V}} + (Du^{\mathcal{V}})^T) : C : (DvD\mathcal{V} + (DvD\mathcal{V})^T) \\
& + (Du^{\mathcal{V}} + (Du^{\mathcal{V}})^T) : C : (Dv + (Dv)^T) \operatorname{div}(\mathcal{V}) dx + o(\|\mathcal{V}\|_{1,\infty}) \\
& + \int_{\Omega} Df\mathcal{V}v + fv \operatorname{div}(\mathcal{V}) dx + o(\|\mathcal{V}\|_{1,\infty}) \\
= & \int_{\Omega} - (DuD\mathcal{V} + (DuD\mathcal{V})^T) : C : (Dv + (Dv)^T) \\
& - (Du^{\mathcal{V}} + (Du^{\mathcal{V}})^T) : C : (DvD\mathcal{V} + (DvD\mathcal{V})^T) \\
& + (Du^{\mathcal{V}} + (Du^{\mathcal{V}})^T) : C : (Dv + (Dv)^T) \operatorname{div}(\mathcal{V}) dx + o(\|\mathcal{V}\|_{1,\infty}) \\
& + \int_{\Omega} Df\mathcal{V}v + fv \operatorname{div}(\mathcal{V}) dx + o(\|\mathcal{V}\|_{1,\infty}).
\end{aligned}$$

Hierbei haben wir den Lebesgueschen Konvergenzatz zur Vertauschung der Landau-Symbole und der Integrale verwendet, dessen Voraussetzungen aufgrund der Endlichkeit des Maßes von  $\Omega$  erfüllt sind, denn für integrierbare Funktionen  $g \in L^1(\Omega)$  gilt  $|\int_{\Omega} o_{L^\infty} g dx| \leq \|o_{L^\infty}\|_{L^\infty} \|g\|_{L^1}$ .

Die obige Gleichung ist bis auf den klein  $o$ -Term linear in  $\mathcal{V}$ . Wir können also schreiben:

$$\begin{aligned} & a^\mathcal{V}(u^\mathcal{V}, v) - F^\mathcal{V}(v) + (q^\mathcal{V}, Bv)_{L^2(\Gamma_C)} - a^0(u^\mathcal{V}, v) + F_0(v) - (q^\mathcal{V}, Bv)_{L^2(\Gamma_C)} \\ &= A'(u, v)\mathcal{V} + F'(v)\mathcal{V} + o(\|\mathcal{V}\|_{1,\infty}) \end{aligned}$$

mit

$$\begin{aligned} A'(u, v)\mathcal{V} &:= \int_{\Omega} - (DuD\mathcal{V} + (DuD\mathcal{V})^T) : C : (Dv + (Dv)^T) \\ &\quad - (Du + (Du)^T) : C : (DvD\mathcal{V} + (DvD\mathcal{V})^T) \\ &\quad (Du + (Du)^T) : C : (Dv + (Dv)^T) \operatorname{div}(\mathcal{V}) dx \end{aligned} \quad (4.33)$$

und

$$F'(v)\mathcal{V} := \int_{\Omega} Df\mathcal{V}v + fv\operatorname{div}(\mathcal{V}) dx. \quad (4.34)$$

Für  $A'(u, v)\mathcal{V}$  kann man mit dem Skalarprodukt auf  $H^1(\Omega)$  und der Ableitung  $\dot{A}(u(0); \cdot) \in L(H^{1,\infty}, H^{-1}(\Omega))$  des Steifigkeitsoperators  $A$  auch  $(\dot{A}(u(0); \mathcal{V}), v)_{H^1(\Omega)}$  schreiben und mit der dualen Paarung und der Ableitung  $\dot{F}(\cdot) \in L(H^{1,\infty}, H^{-1}(\Omega))$  des Riesz-Repräsentanten des Lastvektors  $\langle \dot{F}\mathcal{V}, v \rangle_{H^{-1}(\Omega), H^1(\Omega)}$ . Wir bezeichnen mit

$$C_{\mathcal{V}}(u, q, z; \mathcal{V}) := \begin{pmatrix} \dot{A}(u(0), \cdot) + \dot{F}(\cdot) \\ 0 \\ 0 \end{pmatrix} \text{ den Formgradienten von } C \text{ nach } \mathcal{V}.$$

Somit ist zusammengefasst

$$\begin{aligned} & \left\| C(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}), \mathcal{V}) - C(u(0), q(0), z(0), 0) - C_{y,\mathcal{V}} \begin{pmatrix} u(\mathcal{V}) - u(0) \\ q(\mathcal{V}) - q(0) \\ z(\mathcal{V}) - z(0) \\ \mathcal{V} \end{pmatrix} \right\|_Y \\ &= o\left(\left\| \begin{pmatrix} u(\mathcal{V}) - u(0) \\ q(\mathcal{V}) - q(0) \\ z(\mathcal{V}) - z(0) \end{pmatrix} \right\|_Y\right) + o\left(\|\mathcal{V}\|_{1,\infty}\right), \end{aligned} \quad (4.35)$$

mit

$$\begin{aligned} C_{y,\mathcal{V}} &= C_{y,\mathcal{V}}(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V})) \\ &= \begin{pmatrix} A^0 & B^* & 0 & \dot{A}(u(0), \cdot) + \dot{F}(\cdot) \\ B & -\gamma Id & Id & 0 \\ -D(Bu(\mathcal{V}))B & \gamma Id & 0 & 0 \end{pmatrix}, \end{aligned}$$

wobei  $D(Bu(\mathcal{V}))$  das oben erwähnte Element aus dem verallgemeinerten Differential ist. Übergang zum Supremum liefert die Behauptung. ■

Wir betrachten nun, wie sich die Lösung des Komplementaritätsproblems in Abhängigkeit vom Gebiet entwickeln lässt. Hierbei unterscheiden wir, ob in der Lösung des Kontaktproblems strikte Komplementarität herrscht.

Wir können Folgendes annehmen:

**Annahme 4.16.**

- (a) Die zulässigen Formen werden durch  $T_{\mathcal{V}}$  parametrisiert.  
 (b) Die Elemente der verallgemeinerten Ableitung von  $C$  nach den Lösungen des Komplementaritätsproblems

$$\begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -DB & \gamma Id & 0 \end{pmatrix}$$

sind invertierbar und die Norm der Inversen beschränkt. Für  $A^0$  schreiben wir falls keine Missverständnisse zu befürchten sind nur  $A$ .

- (c)  $(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}))$  und  $(u(0), q(0), z(0))$  sind jeweils Lösungen des Komplementaritätsproblems.

Dann erhalten wir aus (4.35) nach Verwenden der lokalen Lipschitz-Stetigkeit der Lösung des Kontaktproblems bezüglich der Form nach Satz 4.12 .

**Satz 4.17.** Sei  $(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}))$  die Lösung des Kontaktproblems

$$C(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}), \mathcal{V}) = 0$$

zu dem Formparameter  $\mathcal{V} \in H^{1,\infty}$  und die Elemente des verallgemeinerten Differentials (3.13) invertierbar mit unabhängig von  $\mathcal{V}$  beschränkter Inversen. Dann hängt die Lösungsabbildung  $(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}))$  semiglatt von  $\mathcal{V}$  ab.

*Beweis:* Wir zeigen nur die Semiglattheit im Nullpunkt. Die anderen Fälle kann man durch eine Variablentransformation auf diesen Fall zurückführen. Wir setzen  $G(u) := \min(0, d - Bu)$ . Nach Satz 4.15 gilt unter Berücksichtigung der lokalen Lipschitz-Stetigkeit

$$\begin{aligned} & \sup_{D \in \partial(G(u(\mathcal{V})))} \left\| \begin{pmatrix} u(\mathcal{V}) - u(0) \\ q(\mathcal{V}) - q(0) \\ z(\mathcal{V}) - z(0) \end{pmatrix} + \begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -DB & \gamma Id & 0 \end{pmatrix}^{-1} \begin{pmatrix} \dot{A}(u(0); \mathcal{V}) + \dot{f}(\mathcal{V}) \\ 0 \\ 0 \end{pmatrix} \right\| \\ & \leq \sup_{D \in \partial(G(u(\mathcal{V})))} \left\| \begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -DB & \gamma Id & 0 \end{pmatrix}^{-1} \right\| \cdot \left\| C_{y,\mathcal{V}} \begin{pmatrix} u(\mathcal{V}) - u(0) \\ q(\mathcal{V}) - q(0) \\ z(\mathcal{V}) - z(0) \\ \mathcal{V} \end{pmatrix} \right\| \\ & = o\left( \left\| \begin{pmatrix} u(\mathcal{V}) - u(0) \\ q(\mathcal{V}) - q(0) \\ z(\mathcal{V}) - z(0) \end{pmatrix} \right\|_Y \right) + o(\|\mathcal{V}\|_{1,\infty}) = o(\|\mathcal{V}\|_{1,\infty}). \end{aligned} \quad (4.36)$$

■

Das zu  $\mathcal{V} = 0$  gehörende Subdifferential umfasst wegen der Oberhalb-Stetigkeit des Subdifferentials die Häufungspunkte von

$$- \begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -D(u(\mathcal{V}))B & \gamma Id & 0 \end{pmatrix}^{-1} \begin{pmatrix} \dot{A}(u(0), \cdot) + \dot{F}(\cdot) \\ 0 \\ 0 \end{pmatrix}, \quad (4.37)$$

wobei  $D$  wie in (3.13) aus dem Subdifferential für  $u = u(\mathcal{V})$  gewählt wird, für  $\|\mathcal{V}\|_{1,\infty} \rightarrow 0$ . Falls wir annehmen, dass strikte Komplementarität herrscht, das bedeutet,  $\{x \in \Gamma_C : q(0)(x) = z(0)(x)\}$  ist eine Nullmenge, und dass wie in Satz 3.28 eine von der Form unabhängige monotone stetige Funktion  $\phi$ , mit  $\phi(0) = 0$  und  $\text{meas}(\{x \in \Gamma_C : |q(p)(x)| + |z(p)(x)| < t\}) \leq \phi(t)$  existiert, sind die Elemente des Subdifferentials bis auf eine Nullmenge alle gleich. Somit ist die Lösung des Kontaktproblems im Fall strikter Komplementarität nach der Gebietsverformung differenzierbar.

**Satz 4.18.** *Sei  $(u, q, z)$  die strikt komplementäre Lösung des Komplementaritätsproblems für  $\mathcal{V} = 0$  und es gelte die Abschätzung (a) aus Annahme 3.26 unabhängig von der Form des Gebietes. Dann ist die Lösung des Kontaktproblems differenzierbar nach  $\mathcal{V} \in H^{1,\infty}$  und die Ableitung der Lösung der Zustandsgleichung lautet*

$$\begin{pmatrix} \dot{u} \\ \dot{q} \\ \dot{z} \end{pmatrix} = - \begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -D(u(0))B & \gamma Id & 0 \end{pmatrix}^{-1} \begin{pmatrix} \dot{A}(u(0); \cdot) + \dot{F}(\cdot) \\ 0 \\ 0 \end{pmatrix}. \quad (4.38)$$

*Beweis:* Es genügt zu zeigen, dass

$$\begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -D(u(\mathcal{V}))B & \gamma Id & 0 \end{pmatrix}$$

für  $\mathcal{V} \rightarrow 0$  gegen

$$\begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -\chi_{-Bu+d < 0} B & \gamma Id & 0 \end{pmatrix}$$

konvergiert. Sei  $(v, r, s) \in Y$  beliebig, dann ist

$$\begin{aligned} & \left\| \left( \begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -D(u(\mathcal{V}))B & \gamma Id & 0 \end{pmatrix} - \begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -\chi_{-Bu+d < 0} B & \gamma Id & 0 \end{pmatrix} \right) \begin{pmatrix} v \\ r \\ s \end{pmatrix} \right\| \\ &= \|(-D(u(\mathcal{V})) + \chi_{-Bu+d < 0})v\|_{L^2(\Gamma_C)}. \end{aligned}$$

Jetzt sind wir in exakt der gleichen Situation wie im Beweis von Satz 3.28 ab Gleichung (3.30).

Also konvergieren die obigen Matrizen fast überall gegen das gleiche Element unabhängig davon, welche Matrix aus dem verallgemeinerten Differential ausgewählt worden ist.

Die Darstellung der Formableitung in Richtung  $\mathcal{V}$  folgt durch Grenzübergang in (4.36). ■

Wir betrachten nun ein allgemeines Formoptimierungsproblem 4.7 mit einem Kontaktproblem als Nebenbedingung und untersuchen es auf Differenzierbarkeit. Hierzu werden wir ein Adjungiertenkalkül verwenden.

Wir gehen zunächst davon aus, dass wir uns in einem Differenzierbarkeitspunkt der Lösung des Kontaktproblems befinden.

Wir betrachten jetzt ein Zielfunktional der Bauart

$$J(\hat{\Omega}, \hat{u}(\hat{\Omega}), D\hat{u}(\hat{\Omega})) = \int_{\hat{\Omega}} g(x, \hat{u}(x), D\hat{u}(x)) dx \quad (4.39)$$

mit einer hinreichend glatten Funktion  $g$ . Wir transformieren wieder mit der Abbildung  $T_{\mathcal{V}}$  auf das Referenzgebiet und erhalten

$$J(\mathcal{V}, \hat{u}, D\hat{u}) = \int_{\Omega} g(T_{\mathcal{V}}x, \hat{u}(T_{\mathcal{V}}(x)), D(\hat{u}(T_{\mathcal{V}}(x)))DT_{\mathcal{V}}^{-1}(x)) \det DT_{\mathcal{V}} dx.$$

Wir setzen  $u_{\mathcal{V}} = \hat{u} \circ T_{\mathcal{V}}$  und erhalten

$$J(\mathcal{V}, u_{\mathcal{V}}, Du_{\mathcal{V}}) = \int_{\Omega} g(T_{\mathcal{V}}x, u_{\mathcal{V}}(x), Du_{\mathcal{V}}(x)DT_{\mathcal{V}}^{-1}(x)) \det DT_{\mathcal{V}} dx.$$

Wir untersuchen nun, unter welchen Voraussetzungen sich das Zielfunktional entwickeln lässt. Es ist

$$\begin{aligned} & J(\mathcal{V}, u_{\mathcal{V}}, Du_{\mathcal{V}}) - J(0, u, Du) \\ = & \int_{\Omega} g(T_{\mathcal{V}}x, u_{\mathcal{V}}(x), Du_{\mathcal{V}}(x)DT_{\mathcal{V}}^{-1}) \det DT_{\mathcal{V}} - g(x, u, Du) dx \\ = & \int_{\Omega} g(T_{\mathcal{V}}x, u_{\mathcal{V}}(x), Du_{\mathcal{V}}(x)DT_{\mathcal{V}}^{-1}) \det DT_{\mathcal{V}} - g(x, u_{\mathcal{V}}, DuDT_{\mathcal{V}}^{-1}) \det DT_{\mathcal{V}} dx \\ + & \int_{\Omega} g(x, u_{\mathcal{V}}(x), Du_{\mathcal{V}}(x)DT_{\mathcal{V}}^{-1}) \det DT_{\mathcal{V}} - g(x, u, Du_{\mathcal{V}}DT_{\mathcal{V}}^{-1}) \det DT_{\mathcal{V}} dx \\ + & \int_{\Omega} g(x, u(x), Du_{\mathcal{V}}(x)DT_{\mathcal{V}}^{-1}) \det DT_{\mathcal{V}} - g(x, u, Du_{\mathcal{V}}) \det DT_{\mathcal{V}} dx \\ + & \int_{\Omega} g(x, u(x), Du_{\mathcal{V}}(x)) \det DT_{\mathcal{V}} - g(x, u, Du) \det DT_{\mathcal{V}} dx \\ + & \int_{\Omega} g(x, u(x), Du(x)) \det DT_{\mathcal{V}} - g(x, u, Du) dx. \end{aligned}$$

Falls  $g$  Taylor entwickelbar ist, folgt weiter

$$\begin{aligned}
& J(\mathcal{V}, u_{\mathcal{V}}, Du_{\mathcal{V}}) - J(0, u, Du) \\
= & \int_{\Omega} g_x(x, u_{\mathcal{V}}, Du_{\mathcal{V}}DT_{\mathcal{V}}^{-1})\mathcal{V} \det DT_{\mathcal{V}} + o(\|\mathcal{V}\|_{1,\infty}) dx \\
& + \int_{\Omega} g_u(x, u, Du_{\mathcal{V}}DT_{\mathcal{V}}^{-1})(u_{\mathcal{V}} - u) \det DT_{\mathcal{V}} + o(\|u_{\mathcal{V}} - u\|_{0,2}) dx \\
& + \int_{\Omega} g_{Du}(x, u, Du_{\mathcal{V}})(Du_{\mathcal{V}}DT_{\mathcal{V}}^{-1} - Du_{\mathcal{V}}) \det DT_{\mathcal{V}} + o(\|(DT_{\mathcal{V}}^{-1}Du_{\mathcal{V}} - Du_{\mathcal{V}})\|_{0,2}) dx \\
& + \int_{\Omega} g_{Du}(x, u, Du)(Du_{\mathcal{V}} - Du) \det DT_{\mathcal{V}} + o(\|Du_{\mathcal{V}} - Du\|_{0,2}) dx \\
& + \int_{\Omega} g(x, u, Du)(\det DT_{\mathcal{V}} - 1) dx.
\end{aligned}$$

Verwenden wir nun, dass nach Satz 4.18 die Zustandsvariablen wiederum formdifferenzierbar sind und nach Satz 4.12 lokal Lipschitz-stetig von den Formparametern abhängen, so erhalten wir unter Berücksichtigung von Hilfssatz 4.14, der Beschränktheit von  $u$  und der stetigen Differenzierbarkeit von  $g$  wie oben nach dem Satz von Lebesgue:

$$\begin{aligned}
\dot{J}(0, u, Du; \mathcal{V}) &= \int_{\Omega} g_x(x, u, Du)\mathcal{V} + g_u(x, u, Du)\dot{u}(x; \mathcal{V}) - g_{Du}(x, u, Du)DuD\mathcal{V} \\
&\quad + g_{Du}(x, u, Du)D\dot{u}(x; \mathcal{V}) + g(x, u, Du)\operatorname{div}\mathcal{V} dx. \tag{4.40}
\end{aligned}$$

Wir zerlegen nun die Ableitung in einen Teil, der von  $u$  und einen, der von  $\dot{u}$  abhängt.

$$\dot{J}(0, u, Du; \mathcal{V}) = J_{\mathcal{V}}(\Omega, u, Du)\mathcal{V} + J_u(\Omega, u, Du; \mathcal{V})(\dot{u}) \tag{4.41}$$

mit der für festes  $u, Du$  stetigen Linearform

$$J_u : H^1(\Omega) \rightarrow \mathbb{R}, \lambda \mapsto \int_{\Omega} g_u(x, u, Du)\lambda + g_{Du}(x, u, Du)D\lambda dx \tag{4.42}$$

und der bei festem  $u, Du$  in  $\mathcal{V}$  stetigen Linearform

$$J_{\mathcal{V}}(0, u, Du)\mathcal{V} = \int_{\Omega} g_x(x, u, Du)\mathcal{V} - g_{Du}(x, u, Du)DuD\mathcal{V} + g(x, u, Du)\operatorname{div}\mathcal{V} dx. \tag{4.43}$$

Wir eliminieren  $J_u(\dot{u})$  aus (4.41) mittels eines adjungierten Kalküls. Sei nun  $\mu = (\mu_1, \mu_2, \mu_3) \in Y^{**}$ , dem Bidualraum von  $Y$ , so dass

$$\langle C_y(u(0), q(0), z(0)) \begin{pmatrix} \dot{u} \\ \dot{q} \\ \dot{z} \end{pmatrix}, \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} \rangle_{Y^*, Y^{**}} = J_u(\dot{u}) \tag{4.44}$$

mit der Abbildung  $C_y : Y \rightarrow Y^*$ , die durch

$$C_y(u, q, z) = \begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ -D(B(u))B & \gamma Id & 0 \end{pmatrix} \tag{4.45}$$

gegeben ist. Mit der Adjungierten  $C_y^*(u(0), q(0), z(0))$  von  $C_y(u(0), q(0), z(0))$  erhalten wir

$$\langle C_y^*(u(0), q(0), z(0)) \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \dot{u} \\ \dot{q} \\ \dot{z} \end{pmatrix} \rangle_{Y^*, Y} = J_u(\dot{u}). \quad (4.46)$$

Wir identifizieren nun  $Y^{**}$  wegen der Reflexivität der beteiligte Räume mit  $Y$ . Wenn wir nun  $\mu = (\mu_1, \mu_2, \mu_3) \in Y$  so bestimmen<sup>2</sup>, dass (4.46) für alle  $(\dot{u}, \dot{q}, \dot{z}) \in Y$  gilt, dann können wir die Lösung in (4.38) einsetzen und erhalten mit (4.33) und (4.34)

$$\langle C_y(y(0), q(0), z(0)) \begin{pmatrix} \dot{u} \\ \dot{q} \\ \dot{z} \end{pmatrix}, \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} \rangle_{Y^*, Y} = A'(u, \mu_1)\mathcal{V} + F'(\mu_1)\mathcal{V}. \quad (4.47)$$

Also können wir  $J_u(\dot{u}) = A'(u, \mu_1)\mathcal{V} + F'(\mu_1)\mathcal{V}$  in (4.41) einsetzen und erhalten

$$\dot{J}(0, u, Du; \mathcal{V}) = J_{\mathcal{V}}(0, u, Du)\mathcal{V} + A'(u, \mu_1)\mathcal{V} + F'(\mu_1)\mathcal{V}. \quad (4.48)$$

Somit ist die Ableitung von  $J$  nun linear in  $\mathcal{V}$  und nach dem oben gezeigten ist

$$J(\mathcal{V}, u_{\mathcal{V}}, Du_{\mathcal{V}}) = J(0, u, Du) + \dot{J}(0, u, Du; \mathcal{V}) + o(\|\mathcal{V}\|_{1, \infty}). \quad (4.49)$$

Zusammengefasst erhalten wir:

**Satz 4.19.** *Sei  $J$  ein Funktional der Bauart (4.39) und  $u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V})$  die Lösung des Komplementaritätsproblems  $C(u(\mathcal{V}), q(\mathcal{V}), z(\mathcal{V}), \mathcal{V}) = 0$ . Ferner gelte Annahme 4.1 und 3.26 unabhängig von  $\Omega \in \mathcal{O}$ . Dann ist  $J$  formdifferenzierbar und die Ableitung ist gegeben durch*

$$\dot{J}(0, u, Du; \mathcal{V}) = J_{\mathcal{V}}(0, u, Du)\mathcal{V} + A'(u, \mu_1)\mathcal{V} + F'(\mu_1)\mathcal{V}, \quad (4.50)$$

wobei  $(\mu_1, \mu_2, \mu_3)^T$  die Lösung der adjungierten Gleichung (4.46) ist.

Wir zeigen nun, dass die adjungierte Gleichung lösbar ist, indem wir gleichzeitig eine Lösungsmethode vorstellen.

**Satz 4.20.** *Die adjungierte Gleichung (4.46) besitzt für  $\gamma \neq 0$  eine eindeutige Lösung.*

*Beweis:* Die adjungierte Gleichung lautet

$$\begin{pmatrix} A & B^* & -DB^* \\ B & -\gamma Id & \gamma Id \\ 0 & Id & 0 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} J_u \\ 0 \\ 0 \end{pmatrix}. \quad (4.51)$$

Dann ist  $\mu_2 = 0$ . Setzt man dies in die ersten zwei Gleichungen ein, so erhält man nach Addition des  $DB^*/\gamma$ -fachen der zweiten Zeile zur ersten

$$(A + \frac{1}{\gamma}DB^*B)\mu_1 = J_u.$$

<sup>2</sup>Die Lösbarkeit zeigen wir in Satz 4.20.

Diese Gleichung ist eindeutig lösbar, weil  $A$  gleichmäßig elliptisch ist und dieser Ausdruck durch etwas positiv Semidefinites gestört wird. Zuletzt kann  $\mu_3$  aus der zweiten Gleichung berechnet werden. ■

Der Hauptaufwand beim Lösen der adjungierten Gleichung liegt in der Bestimmung von  $\mu_1$ . Bei der numerischen Lösung können hier die gleichen Löser wie bei der Lösung von (3.22) verwendet werden, da de facto die gleiche linke Seite vorliegt. Dies kann zur Verbesserung der Laufzeit ausgenutzt werden, da Manipulationen des Gleichungssystem wieder verwendet werden können. Sie müssen allerdings gespeichert werden.

## 4.4 Beschreibung des Steuerungsraums

Im letzten Abschnitt haben wir die Abhängigkeit der Lösung vom Gebiet diskutiert. Im Kapitel davor haben wir eine Strategie zur Lösung des Kontaktproblems vorgestellt, wobei wir einen Löser für das Problem in einen mit Finiten Elementen diskretisierten Zustandsraum eingeführt haben. Jetzt wollen wir den Steuerungsraum, also die Menge der zulässigen Gebiete, diskretisieren beziehungsweise parametrisieren.

Sei  $P \subset \mathbb{R}^n$  die Menge der Formparameter, dann ist die Menge der zulässigen Gebiete  $\mathcal{O} = \{\Omega(p) | p \in P\}$  gegeben durch  $\Omega(p) := \{y \in \mathbb{R}^3 | y = x + \mathcal{V}_p(x), x \in \Omega_0\}$ , wobei  $\mathcal{V}_p$  die Lösung von

$$-2\mu \operatorname{div} \epsilon(\mathcal{V}_p) - \lambda \operatorname{grad} \operatorname{div} \mathcal{V}_p = 0, \quad (4.52)$$

$$\gamma \mathcal{V}_p = \operatorname{rand}(p) \quad (4.53)$$

ist. Im Gegensatz zu (2.9) haben wir hier die starke Form der Elastizitätsgleichung angegeben. Die Zahlen  $\mu$  und  $\lambda$  sind materialabhängige Konstanten, die wir in diesem Fall beliebig wählen können. Des Weiteren ist  $\operatorname{rand}(p)$  eine differenzierbare Abbildung von  $P$  in den Raum  $H^{1/2}(\partial\Omega)$ . Die Abbildung  $\gamma : H^1(\Omega) \rightarrow H^{1/2}(\partial\Omega)$  gibt die benötigte Spurabbildung an. Wenn der Rand hinreichend glatt ist, dann existiert eine Lösung von (4.52)–(4.53), so dass die Abbildung  $T_p$ , welche die Abbildung von  $\Omega_0$  nach  $\Omega_p$  beschreibt, für festes  $p$  einen Diffeomorphismus zwischen  $\Omega$  und  $\Omega(p)$  bildet. Als nächstes untersuchen wir die Sensibilität von  $x + \mathcal{V}_p(x)$  in Abhängigkeit von  $p \in P$ . Hierzu formulieren wir folgenden Satz:

**Satz 4.21.** *Sei  $\Omega$  ein genügend glattes Referenzgebiet, und zur Menge  $P$  der Formparameter existiert eine genügend oft differenzierbare Abbildung  $\operatorname{rand}(p) : P \rightarrow H^{1/2}(\partial\Omega)$ , dann existiert ein Diffeomorphismus  $T_p : \Omega \rightarrow \Omega(p)$  der Bauart (4.2). Diese Abbildung ist nach  $p \in P$  differenzierbar und die Ableitung  $\mathcal{V}'_p$  ist gegeben durch die Lösung des Systems*

$$\begin{aligned} a(\mathcal{V}'_p, v) + b(v, \lambda'_p) &= 0 \quad \forall v \in H^1(\Omega) \\ b(\mathcal{V}'_p, \mu) &= \left\langle \frac{\partial \operatorname{rand}(p)}{\partial p}, \mu \right\rangle \quad \forall \mu \in H^{-1/2}(\partial\Omega), \end{aligned} \quad (4.54)$$

wobei  $a : H^1(\Omega) \times H^1(\Omega) \rightarrow \mathbb{R}$  die zu (4.52) und  $b : H^1(\Omega) \times H^{-1/2}(\partial\Omega) \rightarrow \mathbb{R}$  die zu (4.53) zugeordneten Bilinearformen sind.

*Beweis:* Wir können mittels einer Sattelpunktformulierung [Bra02, Ste03] die Lösung von (4.52)–(4.53) charakterisieren:

$$\begin{aligned} A\mathcal{V}_p + \gamma^* \lambda_{\mathcal{V}} &= 0, \\ \gamma \mathcal{V}_p - \text{rand}(p) &= 0. \end{aligned}$$

Dabei bezeichnet  $A$  die zu (4.52) gehörende Steifigkeitsmatrix. Wir können  $(\mathcal{V}_p, \lambda_p, p)$  auch als Nullstelle von

$$F : H^1(\Omega) \times H^{-1/2}(\partial\Omega) \times P \rightarrow H^{-1}(\Omega) \times H^{1/2}(\partial\Omega), (\mathcal{V}_p, \lambda_p, p) \mapsto \begin{pmatrix} A\mathcal{V}_p + \gamma^* \lambda_{\mathcal{V}} \\ \gamma \mathcal{V}_p - \text{rand}(p) \end{pmatrix}$$

auffassen. Die Aussage des Satzes ist dann eine Anwendung des Satzes über implizite Funktionen, zum Beispiel [Wer00]. Die Ableitung von  $F$  nach  $(\mathcal{V}_p, \lambda_p)$  ist, da  $F$  für festes  $p$  eine lineare Abbildung ist, wieder  $F$  und somit ein Isomorphismus, weil das System (4.52)–(4.53) eindeutig lösbar ist. ■

Durch Hintereinanderschalten der Abbildungen  $p \rightarrow T_p, T_p : P \rightarrow \Omega(p)$  und dem Zielfunktional der Formoptimierung  $J : \Omega(p) \rightarrow \mathbb{R}$  erhalten wir eine Abbildung, die die Formparameter  $p \in P$  nach  $\mathbb{R}$  abbildet

$$J_p : p \in P \rightarrow \mathbb{R}, \quad p \mapsto J(\Omega(p), u(\Omega(p)), Du(\Omega(p))). \quad (4.55)$$

Diese ist als Komposition von lokal Lipschitz-stetigen Funktionen ebenfalls lokal Lipschitz-stetig und somit nach dem Satz von Rademacher fast überall differenzierbar [EG92]. Wir folgern nun aus (4.50) und (4.54) die Form der Ableitungen von  $J_p$ :

**Satz 4.22.** *Sei  $p \in P$  ein Formparameter und  $(u(p), q(p), z(p))$  die zugehörige Lösung des Kontaktproblems. Falls strikte Komplementarität herrscht und die Annahmen von Annahme 3.26 gelten, dann ist das  $J_p$  in  $p = 0$  differenzierbar und es gilt:*

$$\frac{\partial J}{\partial p} \cdot \delta p = J_{\mathcal{V}}(0, u, Du)\mathcal{V}' \cdot \delta p + A'(u, \mu_1)\mathcal{V}' \cdot \delta p + F'(\mu_1)\mathcal{V}' \cdot \delta p, \quad (4.56)$$

wobei  $\mu = (\mu_1, \mu_2, \mu_3)$  die Lösung der adjungierten Gleichung (4.46) ist.

*Beweis:* Der Beweis folgt durch Anwenden der Kettenregel. ■

Wie wir bereits oben bemerkt haben, bilden die Ausnahmestellen, an denen das Formoptimierungsproblem nicht nach den Formparametern  $p \in P$  differenzierbar ist, eine Nullmenge. Dadurch ist die Wahrscheinlichkeit, während eines iterativen Optimierungsalgorithmus auf einen Nichtdifferenzierbarkeitspunkt zu stoßen, gleich Null. Deshalb werden wir nicht weiter darauf eingehen.

Allerdings möchten wir kurz erwähnen, dass sich die verallgemeinerte Ableitung an einem Nichtdifferenzierbarkeitspunkt  $\tilde{p}$  nicht direkt berechnen lässt, da die Kettenregel für verallgemeinerte Ableitungen nur eine Inklusion angibt. Somit kann man

das Clarkesche Subdifferential zwar von außen approximieren, aber in der Regel liegt keine Gleichheit vor. In [OKZ98] werden für den endlichdimensionalen Fall Kriterien vorgestellt, wie man entscheiden kann, ob Elemente aus der äußeren Approximation Subgradienten sind beziehungsweise wann diese zusammenfallen.

## 4.5 Berechnung inexakter Formableitungen

Wir werden genauso iterative Löser zur Bestimmung der Lösung der adjungierten Gleichung verwenden, wie wir sie in Kapitel 3.5 zur Lösung des Kontaktproblems benutzt haben. Dort haben wir  $\tilde{y} = (\tilde{u}, \tilde{q}, \tilde{z})$  als Näherung an die Lösung  $y = (u, q, z)$  des Kontaktproblems bestimmt. Diese Lösungen hängen jeweils vom Gebiet ab, so dass wir  $\tilde{y}(p) = (\tilde{u}(p), \tilde{q}(p), \tilde{z}(p))$  und  $y(p) = (u(p), q(p), z(p))$  schreiben. Wir bestimmen nun  $(\tilde{\mu}_1(p), \tilde{\mu}_2(p), \tilde{\mu}_3(p))$  als Näherungslösung der adjungierten Gleichung (4.46)

$$\|C_y^*(\tilde{y}(0))(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3) - (J_u, 0, 0)\|_{Y^*} \leq \text{tol}_\mu. \quad (4.57)$$

Dabei machen wir zwei Fehler. Der eine Fehler resultiert darin, dass die Gleichung (4.46) nur näherungsweise gelöst wird, der andere darin, dass wir die auftretenden Funktionen und Operatoren nicht für die exakten Argumente  $(u, q, z)$ , sondern für die in 3.5 näherungsweise bestimmten Werte  $(\tilde{u}, \tilde{q}, \tilde{z})$  auswerten.

Wir untersuchen nun, wie sich der daraus resultierende Fehler auf einen Subgradienten angewendet auf  $\delta p$  auswirkt:

$$\frac{\partial \tilde{J}}{\partial p} \cdot \delta p = J_V(0, \tilde{u}, \tilde{D}u)\mathcal{V}' \cdot \delta p + A'(\tilde{u}, \tilde{\mu}_1)\mathcal{V}' \cdot \delta p + F'(\tilde{\mu}_1)\mathcal{V}' \cdot \delta p. \quad (4.58)$$

Dabei wurde  $\tilde{\mu}$  nach (4.57) bestimmt. Wir bezeichnen den Vektor im  $\mathbb{R}^n$ , der mit dem exakten Subgradienten  $\frac{\partial J}{\partial p}$  korrespondiert, mit  $g$ , und den, welcher mit dem gestörten korrespondiert, mit  $\tilde{g}$ .

**Satz 4.23.** *Sei  $\tilde{y}$  zu  $p \in P$  eine Approximation der Lösung  $y(p)$  der Zustandsgleichung, so dass (3.28) und strikte Komplementarität gelten. Ferner sei  $\tilde{\mu}$  eine Näherung an die Lösung der adjungierten Gleichung zur Genauigkeit  $\text{tol}_\mu$  nach (4.57). Ferner gelten die Annahmen 3.26 und 4.1 Dann existiert eine stetige positive Funktion  $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ , mit  $\Phi(0, 0) = 0$  und zu der nach (4.58) berechneten Näherung  $\tilde{g}$  an einen Subgradienten aus  $\partial_p f(y(p), p)$  ein Subgradient  $g \in \partial_p f(y(p), p)$ , so dass*

$$\|g - \tilde{g}\|_2 \leq \kappa (\kappa_\mu \text{tol}_\mu + \kappa_y \|\tilde{y} - y\|_Y + \Phi(\|\tilde{q} - q\|_{L^2(\Gamma_C)}, \|\tilde{z} - z\|_{L^2(\Gamma_C)}))$$

*gilt mit von  $u, y, \tilde{y}$  unabhängigen Konstanten  $\kappa > 0$ ,  $\kappa_\mu > 0$  und  $\kappa_{\tilde{y}} > 0$ .*

*Beweis:* Wir betrachten den  $\mathbb{R}^n$  mit dem euklidischen Skalarprodukt. Dann ist die Operatornorm

$$\|g - \tilde{g}\|_{\mathbb{R}^n, \mathbb{R}} = \sup_{\delta g \in \mathbb{R}^n \setminus \{0\}} \frac{|(g - \tilde{g}) \cdot \delta g|}{\|\delta g\|_2} = \|g - \tilde{g}\|_2.$$

Wir setzen  $J_y := \begin{pmatrix} J_u \\ 0 \\ 0 \end{pmatrix} \in Y^*$  mit  $J_u$  wie in (4.42). Es gilt

$$\langle \tilde{\mu}, C_V \mathcal{V}' \cdot \delta p \rangle_{Y, Y^*} = \langle C_y^{-*} J_y, C_V \mathcal{V}' \cdot \delta p \rangle_{Y, Y^*} = A'(\tilde{u}, \tilde{\mu}_1) \mathcal{V}' \cdot \delta p + F'(\tilde{\mu}_1) \mathcal{V}' \cdot \delta p. \quad (4.59)$$

Dann ist

$$\langle C_y^{-*} J_y, C_V \mathcal{V}' \cdot \delta p \rangle_{Y, Y^*} = \langle J_y, C_y^{-1} C_V \mathcal{V}' \cdot \delta p \rangle_{Y^*, Y}.$$

Wir wissen, dass  $f(y(p), p)$  fast überall in  $p$  differenzierbar ist. Gleiches gilt für  $f(\tilde{y}(p), p)$ , wobei  $(\tilde{y}(p))$  die eindeutige Lösung des Komplementaritätsproblems

$$C(\tilde{y}(p)) = \tilde{C}$$

mit festem  $\tilde{C}$  ist. Sei  $p_k$  eine Folge von Differenzierbarkeitspunkten von  $f(y(p), p)$  und  $f(\tilde{y}(p), p)$ , so dass  $g(p_k) \cdot \delta p \rightarrow \bar{g} \cdot \delta$  und  $\tilde{g}(p_k) \cdot \delta p \rightarrow \tilde{g} \cdot \delta p$  für eine Richtung  $\delta p$  konvergieren. Dabei bezeichnet  $g(p_k)$  den Gradienten von  $f(y(p), p)$  an der Stelle  $p_k$  und  $\tilde{g}(p_k)$  den Gradienten von  $f(\tilde{y}(p), p)$  an derselben Stelle. Dann sind wegen der Semiglattheit  $\bar{g}$  beziehungsweise  $\tilde{g}$  Subgradienten der entsprechenden Funktionen in  $\bar{p}$ .

Somit gilt wegen der Stetigkeit des Betrags

$$|\bar{g} \cdot \delta p - \tilde{g} \cdot \delta p| = \left| \lim_{k \rightarrow \infty} g_k \cdot \delta p - \lim_{k \rightarrow \infty} \tilde{g}_k \cdot \delta p \right| = \lim_{k \rightarrow \infty} |g_k \cdot \delta p - \tilde{g}_k \cdot \delta p|.$$

Wenn wir nun  $\|g_k - \tilde{g}_k\|$  für alle  $k$  nach oben abschätzen können, müssen wir nur noch zeigen, wie sich  $\tilde{g}(p_k)$  und seine Näherung, die durch nicht exaktes Lösen der adjungierten Gleichung entsteht, angewendet auf  $\delta p$  voneinander unterscheiden.

Somit ist

$$\begin{aligned} |(g(p_k) - \tilde{g}(p_k)) \cdot \delta p| &= |\langle J_y(y(p_k), p_k), (C_y(y(p_k), p_k))^{-1} C_V(y(p_k), p_k) \mathcal{V}' \cdot \delta p \rangle_{Y^*, Y} \\ &\quad + J_V(y(p_k), p_k) \mathcal{V}' \cdot \delta p \\ &\quad - \langle J_y(\tilde{y}(p_k), p_k), (C_y(\tilde{y}(p_k), p_k))^{-1} C_V(\tilde{y}(p_k), p_k) \mathcal{V}' \cdot \delta p \rangle_{Y^*, Y} \\ &\quad - J_V(\tilde{y}(p_k), p_k) \mathcal{V}' \cdot \delta p|. \end{aligned}$$

Einschieben von Nullen und Anwenden der Dreiecksungleichung liefert:

$$\begin{aligned} &|(g(p_k) - \tilde{g}(p_k)) \cdot \delta p| \\ &\leq |\langle J_y(y(p_k), p_k), (C_y(y(p_k), p_k))^{-1} (C_V(y(p_k), p_k) - C_V(\tilde{y}(p_k), p_k)) \mathcal{V}' \cdot \delta p \rangle_{Y^*, Y}| \\ &\quad + |\langle J_y(y(p_k), p_k), ((C_y(y(p_k), p_k))^{-1} - (C_y(\tilde{y}(p_k), p_k))^{-1}) C_V(\tilde{y}(p_k), p_k) \mathcal{V}' \cdot \delta p \rangle_{Y^*, Y}| \\ &\quad + |\langle J_y(y(p_k), p_k) - J_y(\tilde{y}(p_k), p_k), (C_y(\tilde{y}(p_k), p_k))^{-1} C_V(\tilde{y}(p_k), p_k) \mathcal{V}' \cdot \delta p \rangle_{Y^*, Y}| \\ &\quad + |J_V(y(p_k), p_k) \mathcal{V}' \cdot \delta p - J_V(\tilde{y}(p_k), p_k) \mathcal{V}' \cdot \delta p|. \end{aligned}$$

Bis auf den zweiten Summanden kann nun die Lipschitz-Stetigkeit ausgenutzt werden. Deshalb betrachten wir nun den Differenzterm im zweiten Summanden:

$$\|(C_y^{-1}(y(p_k), p_k) - C_y^{-1}(\tilde{y}(p_k), p_k))\|_{Y^*, Y}.$$

Wir nehmen an, dass  $\|C_y^{-1}\|_{Y^*,Y}$  gleichmäßig in der Norm beschränkt ist. Wegen

$$\begin{aligned} & \| (C_y^{-1}(y(p_k), p_k) - C_y^{-1}(\tilde{y}(p_k), p_k)) \|_{Y^*,Y} \\ &= \| C_y^{-1}(\tilde{y}(p_k), p_k)(C_y(\tilde{y}(p_k), p_k) - C_y(y(p_k), p_k))C_y^{-1}(y(p_k), p_k) \|_{Y^*,Y} \end{aligned}$$

genügt es, den Term  $\|(C_y(\tilde{y}(p_k), p_k) - C_y(y(p_k), p_k))\|_{Y,Y^*}$  nach oben abzuschätzen. Der Rest folgt dann aus der Submultiplikativität der Normen.

Wir wählen für  $C(y, p)$  die äquivalente Darstellung

$$C(y, p) = \begin{pmatrix} A(p)u - F + B(p)^*q \\ B(p)u - d + z - \gamma q \\ \min(\gamma q, \gamma q - B(p)u + d) \end{pmatrix} = 0_{H^{-1}(\Omega) \times L^2(\Gamma_c) \times L^2(\Gamma_c)} \quad (4.60)$$

mit  $y = \begin{pmatrix} u \\ q \\ z \end{pmatrix} \in H^1(\Omega) \times L^2(\Gamma_c) \times L^2(\Gamma_c)$ . Wir setzen ohne Einschränkung  $d = 0$

und lassen das Argument  $p$  weg. Dann können wir als Ableitung

$$C_y(y(p), p) = \begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ \chi_{z \leq \gamma q} & \gamma Id & 0 \end{pmatrix}$$

fast überall und

$$C_y(\tilde{y}(p), p) = \begin{pmatrix} A & B^* & 0 \\ B & -\gamma Id & Id \\ \chi_{\tilde{z} \leq \gamma \tilde{q}} & \gamma Id & 0 \end{pmatrix}$$

ebenfalls fast überall wählen.

Somit ist

$$\|(C_y(\tilde{y}(p_k), p_k) - C_y(y(p_k), p_k))\|_{Y,Y^*} = \left\| \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \chi_{z \leq \gamma q} - \chi_{\tilde{z} \leq \gamma \tilde{q}} & 0 & 0 \end{pmatrix} \right\|_{Y,Y^*}.$$

Nun kann man wie in Satz 3.28 eine stetige positive Funktion  $\Phi$  mit  $\Phi(0, 0) = 0$  konstruieren, so dass

$$\|\tilde{C}_y - C_y\|_{Y,Y^*} \leq \Phi(\|\tilde{z} - \bar{z}\|_{L^2(\Gamma_c)}, \|\tilde{q} - \bar{q}\|_{L^2(\Gamma_c)})$$

gilt.

Somit ist  $\|g(p_k) - \tilde{g}(p_k)\| \leq \kappa\|\tilde{y} - y\|_Y + \Phi(\|\tilde{y} - y\|_Y, \|\tilde{y} - y\|_Y)$  bei exakter Lösung der adjungierten Gleichung. Lösen wir die adjungierte Gleichung nur bis auf eine Genauigkeit von  $tol_\mu$ , dann ist

$$\begin{aligned} |(g(p_k) - \tilde{g}_\mu(p_k)) \cdot \delta p| &= |\langle (C_y(\tilde{y}(p_k), p_k))^{-*} J_y(\tilde{y}(p_k), p_k), C_V(\tilde{y}(p_k), p_k) \mathcal{V}' \cdot \delta p \rangle_{Y,Y^*} \\ &\quad + J_V(\tilde{y}(p_k), p_k) \mathcal{V}' \cdot \delta p \\ &\quad - \langle \tilde{\mu}, C_V(\tilde{y}(p_k), p_k) \mathcal{V}' \cdot \delta p \rangle_{Y^*,Y} - J_V(\tilde{y}(p_k), p_k) \mathcal{V}' \cdot \delta p| \\ &= |\langle (C_y(\tilde{y}(p_k), p_k))^{-*} J_y(\tilde{y}(p_k), p_k), C_V(\tilde{y}(p_k), p_k) \mathcal{V}' \cdot \delta p \rangle_{Y,Y^*} \\ &\quad - \langle res_\mu - J_y(\tilde{y}(p_k), p_k), C_V(\tilde{y}(p_k), p_k) \mathcal{V}' \cdot \delta p \rangle_{Y^*,Y}|. \end{aligned}$$

Unter der Annahme, dass  $\|J_{\mathcal{V}}\|$  und  $\|C_y\|_{Y, Y^*}$  beschränkt sind, existiert ein  $\kappa_\mu > 0$  mit  $\|g(p_k) - \tilde{g}_{\tilde{\mu}}(p_k)\| \leq \kappa_\mu \text{tol}_\mu$ . Somit ist

$$\begin{aligned} & |(g(p_k) - \tilde{g}_{\tilde{\mu}}(p_k)) \cdot \delta p| \\ \leq & |(g(p_k) - \tilde{g}(p_k)) \cdot \delta p| + |(\tilde{g}(p_k) - \tilde{g}_{\tilde{\mu}}(p_k)) \cdot \delta p| \\ \leq & (\kappa_\mu \text{tol}_\mu + \kappa_y \|\tilde{y} - y\|_Y + \Phi(\|\tilde{q} - q\|_{L^2(\Gamma_C)}, \|\tilde{z} - z\|_{L^2(\Gamma_C)})) \|\mathcal{V}\|_{\mathbb{R}^n, H^{1,\infty}} \|\delta p\|. \end{aligned}$$

Übergang zum Supremum über  $\|\delta p\|$  liefert mit der Beschränktheit von  $\|\mathcal{V}\|$  durch  $\kappa > 0$  die Behauptung. ■

## 4.6 Semiglattheit der Formfunktion

Die zusammengesetzte Formfunktion ist eine Abbildung von der Parametermenge  $P \subset \mathbb{R}^n$  nach  $\mathbb{R}$ . Wir haben in den vorangegangenen Kapiteln zwei Begriffe für Semiglattheit verwendet, nämlich den im Funktionenraum und den im Endlichdimensionalen. Diese sind nicht identisch, da im Funktionenraum die Richtungs-differenzierbarkeit nicht verlangt wird.

Im folgenden Kapitel werden wir ein numerisches Verfahren zur Lösung des Formoptimierungsproblems vorstellen, welches die Richtungs-differenzierbarkeit als Voraussetzung benötigt.

Deshalb wollen wir hier kurz erläutern, wann die diskretisierte Formulierung des Formoptimierungsproblems semiglat im Sinne der endlichdimensionalen Definition ist. Wir werden hier die Kriterien, die in [OKZ98] hergeleitet worden sind, vorstellen und verifizieren.

Die diskretisierte Version des Formoptimierungsproblems lautet:

**Problem 4.24** (Diskretisiertes Formoptimierungsproblem). *Gesucht sei  $p_0 \in P$  mit*

$$J_h(p_0), u_h(p_0) \leq J_h(p), u_h(p) \quad \forall p \in P, \quad (4.61)$$

wobei  $u_h(p)$  die Lösung des diskretisierten Kontaktproblems (3.23) zum Formparameter  $p$  ist,

$$\begin{aligned} A_h(p)u_h(p) - F_h(p) + B_h(p)^T q_h(p) &= 0, \\ B_h(p)u_h(p) - M_h(p)d_h(p) - \gamma M_h(p)q_h(p) + M_h(p)z_h(p) &= 0, \\ \gamma M_h(p)q_h(p) + \min(0, -B_h(p)u_h(p) + M_h(p)d_h(p)) &= 0, \end{aligned} \quad (4.62)$$

und  $J_h$  die formabhängige Diskretisierung des Zielfunktional.

Die Semiglattheit des diskretisierten Formoptimierungsproblems gilt nach dem folgenden Satz:

**Satz 4.25.** *Die Abbildung  $J_h : P \rightarrow \mathbb{R}$  aus Problem 4.24 ist semiglat.*

*Beweis:* Wir vernachlässigen den Index  $h$  und verwenden, dass nach Annahme 4.1  $B(p) = B$ ,  $d(p) = d$  und  $M(p) = M$  für alle  $p \in P$  gilt.

Wir reduzieren die Anzahl der Variablen, indem wir die zweite Gleichung in (4.62) vernachlässigen, da diese die Variable  $z_h$  definiert, welche nicht in das Formoptimierungsproblem eingeht und die erste Gleichung nach  $u_h(p)$  auflösen. Dies können wir dann in (4.61) und die dritte Gleichung aus (4.62) einsetzen. Wir erhalten somit das äquivalente reduzierte Optimierungsproblem

$$\begin{aligned} &\text{Minimiere } J(p, (A(p))^{-1}(F(p) - B^T q(p))), \\ &\text{mit } \min(\gamma M q(p), (\gamma M + B(A(p))^{-1} B^T) q(p) - B(A(p))^{-1} F(p) + M d) = 0. \end{aligned}$$

Nachdem  $M$  eine Diagonalmatrix mit positiven Einträgen ist, können wir die Nebenbedingung noch umskalieren zu

$$\min(\gamma q(p), q(p) + M^{-1} B(A(p))^{-1} B^T q(p) - M^{-1} B(A(p))^{-1} F(p) + d) = 0.$$

Anschließend führen wir  $\Phi(p, q) := - (M^{-1} B(A(p))^{-1} B^T q - M^{-1} B(A(p))^{-1} F + d)$  und  $F(q) := q$  ein und erhalten folgende kompakte Formulierung des diskretisierten Formoptimierungsproblems

$$\begin{aligned} &\text{Minimiere } J(p, (A(p))^{-1}(F(p) - B^T q)) \\ &\text{mit } \min(F(q), q - \Phi(p, q)) = 0. \end{aligned}$$

Nach [OKZ98] ist die Gleichung regulär, falls  $F$  und  $\Phi$  stetig differenzierbar sind,  $D_q F(q)$  positiv definit ist und  $D_q F(q)(D_y \Phi)^{-1}$  negativ semidefinit ist. Dies ist in unserem Fall mit  $D_q F(q) = Id$  und  $D_q F(q)(D_y \Phi)^{-1} = -M^{-1} B(A(p))^{-1} B^T q$  erfüllt, da  $A(p)$  und  $M$  ebenfalls positiv definit sind.

Weiterhin folgt nach [OKZ98], dass die Lösungsabbildung  $p \mapsto q(p)$  regulärer verallgemeinerter Gleichungen unter den eben erwähnten Voraussetzungen semiglatt ist.

Somit ist die Abbildung  $p \mapsto J_h(p, q(p))$  als Komposition semiglatte Abbildungen semiglatt. ■

Wir müssen an dieser Stelle darauf hinweisen, dass es nur möglich ist, eine äußere Approximation für das verallgemeinerte Differential der Lösungsabbildung anzugeben. Eine Mangasarian-Fromowitz Constraint Qualification, wie sie in [OKZ98] beschrieben wurde, konnte nicht nachgewiesen werden. Hierzu hätten Teilmatrizen einer Matrix, die sämtliche partiellen Ableitungen der auftretenden Funktionen enthält, auf lineare Abhängigkeit überprüft werden müssen, was aufgrund der Komplexität des Problems nicht durchgeführt werden kann.

Im folgenden Kapitel werden wir die Richtungs-differenzierbarkeit bei der Semiglattheit immer mitbehandeln, da sie für die diskretisierten Fälle, die letztendlich gerechnet werden, vorhanden ist.

# Kapitel 5

## Nichtglatte Optimierungsverfahren

Wir entwickeln ein Bundle-Verfahren zur Minimierung einer semiglatten Funktion unter einer nichtglatten Nebenbedingung. Im Laufe der Optimierung werden wir das Kontaktproblem und das adjungierte Problem nur näherungsweise lösen, was sowohl zu Fehlern im Zustand als auch zu Fehlern bei den Subgradienten führt. Solche Ansätze sind zur Zeit beim Lösen von Optimierungsproblemen mit partiellen Differentialgleichungen als Nebenbedingung Forschungsgegenstand, denn dadurch kann die Rechenzeit, die zum größten Teil auf das Lösen der Differentialgleichung fällt, stark reduziert werden. An dieser Stelle möchten wir auch auf die Arbeiten von [KS03] zu inexakten Newton-Verfahren und von [HV02, LS99, Ulb07] zu inexakten SQP-Verfahren hinweisen.

Bei der Konstruktion unseres Algorithmus orientieren wir uns an dem nichtkonvexen Bundle-Trust-Region-Verfahren, welches in [Sch89] vorgestellt wird.

Deshalb werden wir in den ersten beiden Abschnitten dieses Kapitels Bundle-Trust-Region-Verfahren für konvexe und nichtkonvexe nichtglatte Optimierungsprobleme vorstellen. Hierbei bedienen wir uns der Darstellung in [SZ92a]. Anschließend fassen wir den aktuellen Stand von nichtexakten Bundle-Verfahren zusammen und stellen ein neues inexaktes Bundle-Verfahren vor. Für dieses Verfahren führen wir im vorletzten Abschnitt einen Konvergenzbeweis und gehen dann auf Erweiterungen ein.

### 5.1 Bundle-Trust-Region-Verfahren für konvexe Funktionen

Wir betrachten zunächst das Minimierungsproblem

$$\min f(u), \text{ wobei } f : \mathbb{R}^n \rightarrow \mathbb{R}.$$

Dabei ist  $f$  eine nichtglatte Funktion, von der wir annehmen, dass sie zusätzlich lokal Lipschitz-stetig ist. Für diese Funktionen existiert das Clarkesche Subdifferential nach Definition B.2 an der Stelle  $u \in \mathbb{R}^n$

$$\partial f(u) := \text{conv} \left\{ g \in \mathbb{R}^n \mid g = \lim_{i \rightarrow \infty} \nabla f(u_i), u_i \rightarrow u, \nabla f(u_i) \text{ existiert und konvergiert} \right\}.$$

Im Folgenden nehmen wir an, dass wir zu jedem gegebenen  $u$  den Funktionswert  $f(u)$  und einen beliebigen Subgradienten  $g \in \partial f(u)$  bestimmen können.

In der nichtglatten Optimierung haben sich Bundle-Verfahren als Optimierungsverfahren durchgesetzt. Alle Bundle-Verfahren basieren auf zwei grundlegenden Ideen.

- (a) Im aktuellen Iterationspunkt  $u^k$  werden die vorangegangenen Informationen<sup>1</sup>  $(q^i, f(q^i), g^i)$ ,  $i \in J^k$  verwertet, um eine Modellfunktion (Näherung) für  $f$  in einer Umgebung von  $u^k$  aufzustellen. Die Punkte  $q^i$  sind Hilfspunkte, die teilweise mit  $u^i$  übereinstimmen. Für ein erstes Verständnis empfiehlt es sich, anzunehmen, dass die Hilfspunkte  $q^i$  mit den Iterationswerten  $u^i$  für  $i \in J^k$  übereinstimmen und  $J^k = \{1, \dots, k\}$  gilt.
- (b) Falls die Modellfunktion die zu optimierende Funktion  $f$  nicht adäquat beschreibt, werden weitere Informationen in Form von Subgradienten in der Nähe von  $u^k$  in das Bundle aufgenommen.

Wir beschreiben nun die Idee von Bundle-Verfahren für eine konvexe Zielfunktion  $f$ . Hierzu beachte man auch Abbildung 5.1.

Wir betrachten ein Schnittebenenmodell als Näherung von  $f$  an  $u^k$ :

$$\check{f} := \max_{i \in J^k} \{ (g^i)^T (u - q^i) + f(q^i) \}. \quad (5.1)$$

Gleichung (5.1) ist eine stückweise lineare Funktion. Sie stimmt in den Punkten  $q^i$  mit  $f(q^i)$  überein und liegt ansonsten unterhalb von  $f$ . Dies folgt direkt aus der Konvexität.

Wir betrachten nun die aufgrund der Konvexität von  $f$  positiven Linearisierungsfehler im Punkt  $u^k$ ,

$$\alpha_i^k := \alpha(u^k, q^i) := f(u^k) - (f(q^i) + (g^i)^T (u^k - q^i)). \quad (5.2)$$

Wir führen die neue Variable  $s := u - u^k \in \mathbb{R}^n$  ein und schreiben (5.1) in der kompakten Form

$$\check{f}(u^k, s) = \max_{i \in J^k} \{ (g^i)^T s - \alpha_i^k \} + f(u^k), \quad s \in \mathbb{R}^n.$$

Wir lassen den konstanten Summanden  $f(u^k)$  wegfallen und erhalten die Schnittebenenfunktion

$$f_{CP}(u^k, s) := \max_{i \in J^k} \{ (g^i)^T s - \alpha_i^k \}, \quad s \in \mathbb{R}^n, \quad (5.3)$$

<sup>1</sup>Diese nennt man Bündel, engl. Bundle.

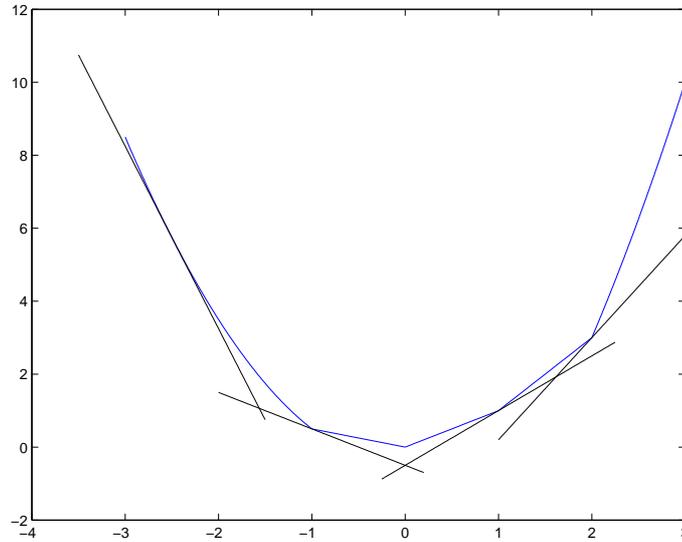


Abbildung 5.1: Die Abbildung zeigt die zu optimierende Funktion  $f$  (blau) und ein Schnittstellenmodell (schwarz) an  $f$ . Es stimmt an den Kontrollpunkten mit  $f$  exakt überein und liegt sonst unterhalb von  $f$ . Wir werden das punktweise Maximum über alle Schnittstellen als Modellfunktion verwenden

die wir im Folgenden als Trust-Region-Modell verwenden werden.

Wir bestimmen das Minimum von (5.3) im Trust-Region-Kreis  $\|s\|^2 \leq 2\rho$ . Ein Vergleich der KKT-Bedingungen zeigt, dass dies äquivalent ist zur Minimierung des penalisierten Schnittstellenmodells,

$$s(t) := \operatorname{argmin} \left\{ f_{CP}(u^k, s) + \frac{1}{2t} \|s\|^2 \quad : \quad s \in \mathbb{R}^n \right\}, \quad t > 0.$$

Dabei besteht zwischen dem Trust-Region-Radius  $\rho$  und dem Strafterm  $\frac{1}{2t}$  ein monotoner Zusammenhang. Wir werden daher im weiteren Verlauf das letzte Problem als Modellproblem ansehen. Dies ist ebenfalls äquivalent zu folgendem quadratischen Optimierungsproblem

$$(v(t), s(t)) = \operatorname{argmin} \left\{ v + \frac{1}{2t} \|s\|^2 \quad : \quad v \geq (g^i)^T s - \beta_i^k, i \in J^k \right\}, \quad (5.4)$$

das sich mit allen gängigen Verfahren für quadratische Optimierungsprobleme mit linearen Nebenbedingungen (siehe zum Beispiel [Alt02, JS04]) lösen lässt. Dabei gibt  $v(t)$  einen Schätzer für die Abstiegsdifferenz und  $s(t)$  die Abstiegsrichtung an.

Aus den KKT-Bedingungen für (5.4) folgt:

**Satz 5.1.** Für  $n \in \mathbb{N}$  bezeichne

$$\Lambda(n) := \left\{ \lambda \in \mathbb{R}^n : \lambda_i \geq 0, 1 \leq i \leq n, \quad \text{und} \quad \sum_{i=1}^n \lambda_i = 1 \right\}.$$

Im Minimum  $(v(t), s(t))$  von (5.4) existieren Multiplikatoren  $\lambda(t) \in \Lambda(|J^k|)$ , so dass

$$\lambda_i(t)(-v(t) + (g^i)^T s(t) - \alpha_i^k) = 0 \quad i \in J^k, \quad (5.5)$$

$$s(t) = -t \sum_{i \in J^k} \lambda_i(t) g^i, \quad (5.6)$$

$$v(t) = -t \left\| \sum_{i \in J^k} \lambda_i(t) g^i \right\|^2 - \sum_{i \in J^k} \lambda_i(t) \alpha_i^k \quad (5.7)$$

$$= -\frac{1}{t} \|s\|^2 - \sum_{i \in J^k} \lambda_i(t) \alpha_i^k \quad (5.8)$$

gelten.

Wir erwähnen einige Eigenschaften, die alle in [Alt04, Sch89] bewiesen sind.

**Bemerkung 5.2.**

- (a) Die Lösungen  $(v(t), s(t))$  hängen stetig von  $t$  ab.
- (b) Im Minimum von (5.4) gilt  $v(t) \leq 0$ .
- (c) Angenommen, zu  $\epsilon > 0$  existiere ein  $\lambda \in \Lambda(|J^k|)$  mit

$$\left\| \sum_{i \in K^k} \lambda_i g^i \right\| \leq \epsilon \quad \text{und} \quad \sum_{i \in J^k} \lambda_i \alpha_i^k \leq \epsilon.$$

Dann ist  $u^k$   $\epsilon$ -optimal, das heißt

$$f(u^k) \leq f(x) + \epsilon \|u - u^k\| + \epsilon \quad \text{für alle } u \in \mathbb{R}^n. \quad (5.9)$$

Für  $\epsilon = 0$  ist  $u^k$  dann das globale Minimum.

- (d) Das zu (5.4) duale Problem lautet

$$\min \left\{ \frac{1}{2} \left\| \sum_{i \in K^k} \lambda_i g^i \right\|^2 + \frac{1}{t} \sum_{i \in J^k} \lambda_i \alpha_i^k, \lambda \in \Lambda(|J^k|) \right\}. \quad (5.10)$$

Die Lösung des dualen Problems bezeichnen wir mit  $\omega^k$ . Einfache Dualitätsüberlegungen zeigen, dass die  $\lambda$  aus (5.10) mit denen aus Satz 5.1 korrespondieren.

Der Bundle-Trust-Region-Algorithmus zerfällt in zwei Schritte. In einer inneren Iteration wird ein geeignetes  $t$  für (5.4) bestimmt und entschieden, ob die Lösung zu einem hinreichenden Abstieg führt, oder ob das Bundle und damit die Modellfunktion verbessert werden muss.

Wir geben den Algorithmus der inneren Iteration an, der in der Abbildung 5.2 skizziert ist:

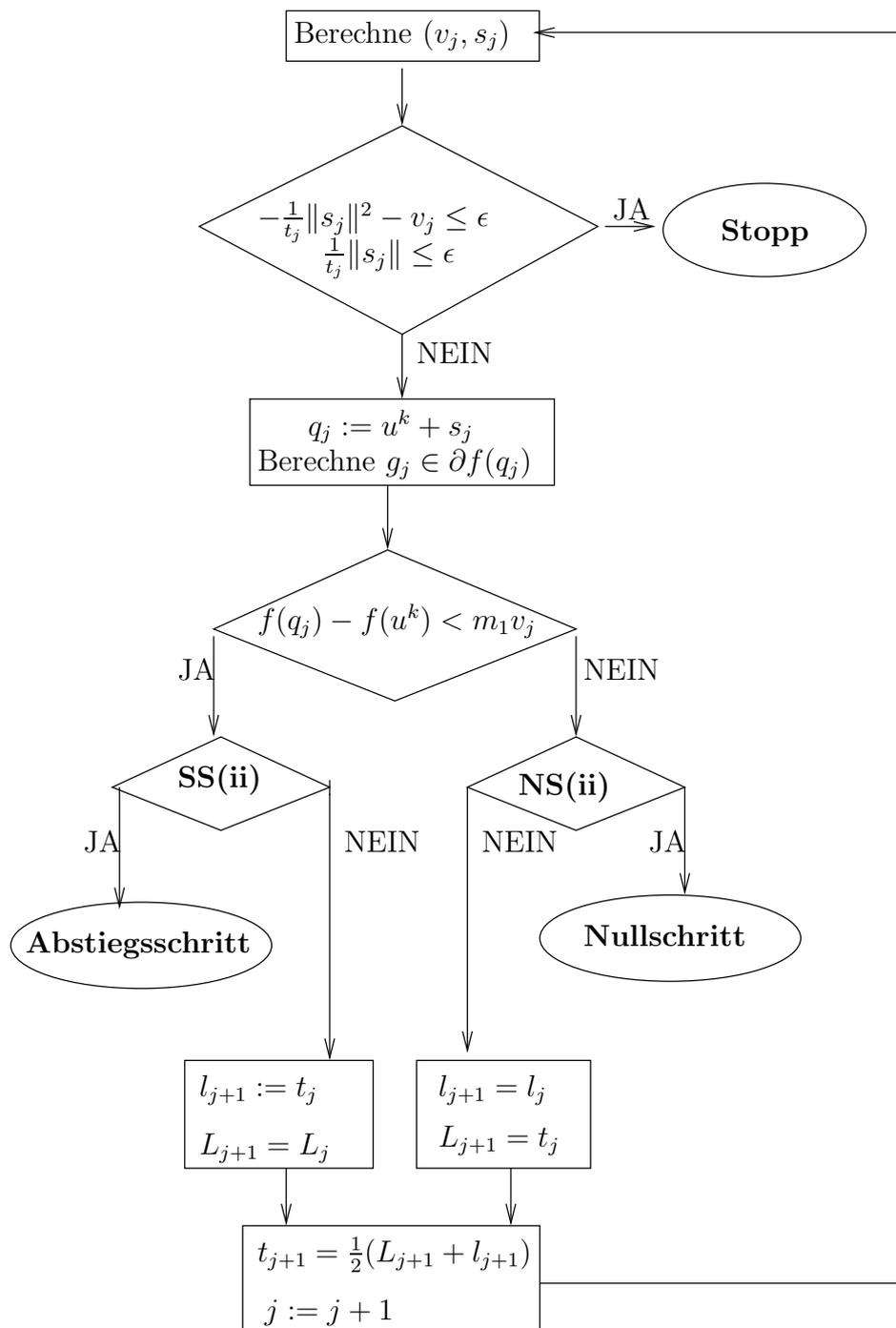


Abbildung 5.2: Flow Chart der inneren Iteration für konvexe Funktionen

**Algorithmus 5.3** (Innere Iteration für konvexe Funktionen). Gegeben seien eine obere Schranke  $T > 0$  für  $t$ , die Konstanten  $m_1, m_2, m_3$  mit  $0 < m_1 < m_2 < 1$  und  $0 < m_3$ , sowie eine kleine Zahl  $\nu > 0$  und ein Abbruchkriterium  $\epsilon$ . Der Index  $k$  ist während der inneren Iteration fixiert, wohingegen der Index  $j$  der Laufindex der Iteration ist.

- (0) Setze  $t_1 := t^{k-1}$ ,  $l_1 := 0$ ,  $L_1 := T$  und  $j := 1$ .
- (1) Berechne die Lösung  $(v_j, s_j) := (v_j(t_j), s_j(t_j))$  von (5.4). Falls  $(1/t_j)\|s_j\| \leq \epsilon$  und  $-(1/t_j)\|s_j\|^2 - v_j \leq \epsilon$ : Stopp,  $u^k$  ist  $\epsilon$ -optimal. Ansonsten setze  $q_j := u^k + s_j$  und berechne  $g_j \in \partial f(q_j)$ .
- (a) Gilt **SS(i)** und **SS(ii)**, so führe einen Abstiegschritt durch: Setze  $u^{k+1} := q^{k+1} := q_j$  und  $g^{k+1} := g_j$ ; stopp.
- (b) Gilt **SS(i)**, nicht aber **SS(ii)**, so setze  $l_{j+1} := t_j$ ,  $L_{j+1} := L_j$  und  $t_{j+1} := \frac{1}{2}(u_{j+1} + l_{j+1})$ . Erhöhe  $j$  um 1 und gehe zu (1).
- (c) Gilt **NS(i)** und **NS(ii)**, so führe einen Nullschritt durch: Setze  $u^{k+1} := u^k$ ,  $q^{k+1} := q_j$  und  $g^{k+1} := g_j$ ; stopp.
- (d) Gilt **NS(i)**, nicht aber **NS(ii)**, so setze  $L_{j+1} := t_j$ ,  $l_{j+1} := l_j$  und  $t_{j+1} := \frac{1}{2}(L_{j+1} + l_{j+1})$ . Erhöhe  $j$  um 1 und gehe zu (1).

Wir setzen  $v^k, s^k, \lambda^k$  und  $t^k$  gleich den Werten, bei denen wir die innere Iteration verlassen. Es gilt dann  $s^k = -t^k \sum_{i \in J^k} \lambda_i^k g^i$ . Mit diesen  $\lambda^k$  definieren wir für später

$$\zeta^k := \sum_{i \in J^k} \lambda_i^k g^i, \quad \sigma^k := \sum_{i \in J^k} \lambda_i^k \alpha_i^k.$$

Wir geben nun die Abstiegs- beziehungsweise die Nullschrittbedingungen an und diskutieren sie anschließend.

- SS: (i)**  $f(q_j) - f(u^k) < m_1 v_j$ ,  
**SS: (ii)**  $(g_j)^T s_j \geq m_2 v_j$  oder  $t_j \geq T - \nu$ ,  
**NS: (i)**  $f(q_j) - f(u^k) \geq m_1 v_j$ ,  
**NS: (ii)**  $\alpha(u^k, q_j) \leq m_3 \sigma^{k-1}$  oder  $|f(u^k) - f(q_j)| \leq \|\zeta^{k-1}\| + \sigma^{k-1}$

Bedingung **SS(i)** führt dazu, dass ein Abstieg nur dann ausgeführt wird, wenn die Abstiegsdifferenz mindestens  $m_1$ -mal den von (5.4) vorhergesagten Abstieg  $v^k$  beträgt. Die Bedingung **NS(i)** ist die Verneinung von **SS(i)**.

Die Bedingung **SS(ii)** stellt sicher, dass sich das Schnittebenenmodell wesentlich verändert, falls der Strafterm  $t$  seine obere Schranke  $T$  nicht überschreitet. Dies bedeutet, dass im nächsten Durchlauf der inneren Iteration die Lösungen  $(v^{k+1}, s^{k+1})$  von (5.4) sich von  $(v^k, s^k)$  unterscheiden. Diese Bedingung geht nicht in den Konvergenzbeweis ein, sie dient nur dazu, dass man keine zu kleinen Werte für  $t^k$  verwendet.

Die Bedingung **NS(ii)** stellt sicher, dass  $q^k$  nur dann in das Bündel aufgenommen wird, wenn  $q^k$  nahe bei  $u^k$  liegt und somit das Modell lokal verbessert. Gleichzeitig

führt **NS(i)** dazu, dass sich die Lösungen  $(v^{k+1}, s^{k+1})$  von (5.4) im nächsten Schritt von  $(v^k, s^k)$  unterscheiden, da das Bundle sonst redundant wird und keine weiteren Informationen mehr aufnehmen kann.

Es lässt sich zeigen [SZ92a], dass die innere Iteration nach endlich vielen Schritten terminiert.

Wir geben jetzt den Gesamtalgorithmus des Bundle-Trust-Region-Verfahrens nach [SZ92a] an. Dabei nehmen wir an, dass genügend Speicher vorliegt und das Bundle daher nicht zurückgesetzt werden muss (Reset).

**Algorithmus 5.4** (BT-Algorithmus für konvexe Optimierungsprobleme). *Gegeben seien Startpunkt  $u^1 \in \mathbb{R}^n$  und Parameter  $T > 0$ ,  $0 < m_1 < m_2 < 1$ ,  $0 < m_3 < 1$ ,  $\nu > 0$  und  $\epsilon \geq 0$*

- (0) Berechne  $f(u^1)$ ,  $g^1 \in \partial f(u^1)$ , setze  $J^1 := \{1\}$ ,  $\alpha_1^1 := 0$  und  $k := 1$ .
- (1) Berechne  $u^{k+1}$  und  $g^{k+1}$  nach Algorithmus 5.3 oder stopp, falls  $u^k$   $\epsilon$ -optimal
- (2) Falls die innere Iteration mit einem Abstiegschritt endet, setze

$$\begin{aligned}\alpha_i^{k+1} &:= \alpha_i^k + f(u^{k+1}) - f(u^k) - (g^i)^T s^k \text{ für } i \in J^k \\ \alpha_{k+1}^{k+1} &:= 0.\end{aligned}$$

Falls die innere Iteration mit einem Nullschritt endet, dann setze

$$\begin{aligned}\alpha_i^{k+1} &:= \alpha_i^k \text{ für } i \in J^k \\ \alpha_{k+1}^{k+1} &:= \alpha(u^k, q^{k+1})\end{aligned}$$

Setze  $J^{k+1} := J^k \cup \{k+1\}$  und erhöhe  $k$  um 1. Gehe zu (1).

Die Konvergenz liefert folgender Satz:

**Satz 5.5** (Konvergenz des Bundle-Trust-Region-Verfahrens). *Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  konvex. Außerdem sei die Lösungsmenge*

$$U^* = \{u^* \in \mathbb{R}^n \mid f(u^*) \leq f(u) \forall u \in \mathbb{R}^n\}$$

*nicht leer. Dann erzeugt das Bundle-Trust-Region-Verfahren 5.4 eine Folge  $(u^k)_{k \in \mathbb{N}}$ , so dass  $(f(u^k))_{k \in \mathbb{N}}$  gegen das Minimum von  $f$  konvergiert.*

*Beweis:* [SZ92a]. ■

## 5.2 Bundle-Trust-Region-Verfahren für nichtkonvexe Funktionen

Für nichtkonvexe  $f$  gilt die Subgradientenungleichung (B.1) nicht mehr, so dass die Linearisierungsfehler (5.2) negativ werden können. Als direkte Konsequenz ist  $f_{CP}$

nicht länger eine Approximation an  $f$  von unten in einer Umgebung von  $u^k$ . Es gilt sogar häufig  $f_{CP}(u^k, 0) = \max_i \{-\alpha_i^k\} \neq \lim_{h \rightarrow 0} f(u^k + h) - f(u^k) = 0$ . Deshalb wird eine Strategie, die auf Lemaréchal zurückgeht, angewendet. Wir ersetzen die  $\alpha_i^k$  durch

$$\beta_i^k := \beta(u^k, q^i) := \max \{ \alpha_i^k, c_0 \|u^k - q^i\|^2 \}; \quad (5.11)$$

hierbei ist  $c_0$  eine kleine positive Konstante. Setzt man  $c_0 = 0$  so stimmen die  $\alpha_i^k$  mit den  $\beta_i^k$  für konvexe  $f$  überein. Nach Konstruktion ist  $\beta_i^k \geq 0$ , und das neue Modell

$$f_{CP}(u^k, s) := \max_{i \in K^k} \{ (g^i)^T s - \beta_i^k \}$$

stimmt zumindest für  $s = 0$  mit  $f(u^k + s) - f(u^k)$  überein. Wenn wir nun in Abschnitt 5.1 die  $\alpha_i^k$  durch die  $\beta_i^k$  ersetzen, gilt weiterhin Satz 5.1 und die Probleme (5.4) und (5.10) bleiben zueinander dual.

Wie auch in der glatten nichtkonvexen Optimierung garantiert  $0 \in \partial f$  keine Optimalität, sondern nur einen stationären Zustand. Somit liefert

$$\left\| \sum_{i \in K^k} \lambda_i g^i \right\| \leq \epsilon \quad \text{und} \quad \sum_{i \in J^k} \lambda_i \beta_i^k \leq \epsilon \quad (5.12)$$

kein Kriterium für eine  $\epsilon$ -Optimalität (5.9) von  $u^k$ . Allerdings können wir aus (5.12) schließen, dass  $0$  bis auf einen Fehler  $\epsilon$  in der konvexen Hülle von gewissen Subgradienten  $g^i \in \partial f(q^i)$  liegt, wobei die  $q^i$  wegen  $\sum_{i \in J^k} \lambda_i \beta_i^k \leq \epsilon$  nicht weit von  $u^k$  entfernt sind. In Anlehnung an den glatten Fall nennen wir  $u^k$   $\epsilon$ -stationär.

Die Kriterien für einen Abstiegschritt beziehungsweise einen Nullschritt müssen ebenfalls korrigiert werden.

Wir werden auf das Kriterium **SS(ii)** verzichten und immer einen Abstiegschritt durchführen, wenn **SS(i)** erfüllt ist. Für einen Nullschritt müssen wir als zusätzliches Kriterium

$$\mathbf{NS(iii)} : g_j^T s_j - \beta_j^k \geq m_2 v_j$$

eingeführen, das uns eine Modelländerung garantiert. Weil im Fall  $t_j \rightarrow 0$  aus Stetigkeitsgründen  $v_j$  gegen  $0$  strebt, impliziert dies einen stationären Punkt. Also darf der Trust-Region-Parameter  $t$  bei nichtstationären Punkten nicht beliebig verkleinert werden. Deshalb schlagen [SZ92a] folgenden Ausweg vor: Falls die zweite Bedingung in **NS(ii)** erfüllt ist, also der Kandidat für den neuen Iterationspunkt nahe beim alten liegt, aber noch keine Modelländerung vorliegt, wird eine Liniensuche, die auf Lemaréchal [Lem81] zurückgeht, entlang der Suchrichtung  $s^k := s(t_j)$  durchgeführt. Um Konvergenz zu garantieren, benötigen wir die zusätzliche Forderung, dass  $f$  semiglatt ist.

Die Liniensuche ermittelt daher nach endlich vielen Schritten ein  $d^k \in (0, 1)$ , so dass in  $q^{k+1} := u^k + d^k s^k$  mit  $v^k := v(t_j)$  und  $g^{k+1} \in \partial f(q^{k+1})$  gilt:

$$\begin{aligned} \mathbf{SSS} : & f(q^{k+1}) - f(u^k) \leq m_1 s^k v^k \quad \text{und} \quad (g^{k+1})^T d^k \geq m_2 v^k, \\ \mathbf{NS} : & -\beta(u^k, q^{k+1}) + \langle g^{k+1}, d^k \rangle \geq m_2 v^k. \end{aligned} \quad (5.13)$$

Gilt ersteres, dann sagen wir, dass das Short-Serious-Schritt-Kriterium erfüllt ist, und wir führen einen verkürzten Abstiegschritt durch, im zweiten Fall führen wir einen Nullschritt durch.

Wir fassen noch einmal kurz die Abbruchkriterien für die innere Iteration zusammen und geben die innere Iteration an. Sie ist in dem Flow Chart, Abbildung 5.3, illustriert.

- SS:** (i)  $f(q_j) - f(u^k) < m_1 v_j$ ,  
**NS:** (i)  $f(q_j) - f(u^k) \geq m_1 v_j$ ,  
**NS:** (ii)  $\beta(u^k, q_j) \leq m_3 \sigma^{k-1}$  oder  $|f(u^k) - f(q_j)| \leq \|\zeta^{k-1}\| + \sigma^{k-1}$   
**NS:** (iii)  $g_j^T s_j - \beta_j^k \geq m_2 v_j$

**Algorithmus 5.6** (Innere Iteration für nichtkonvexe Funktionen). *Gegeben sei eine obere Schranke  $T > 0$  für  $t$ , Konstanten  $m_1, m_2, m_3$  mit  $0 < m_1 < m_2 < 1$  und  $0 < m_3$ , ebenso wie eine kleine Zahl  $\nu > 0$  und ein Abbruchkriterium  $\epsilon$ . Der Index  $k$  ist während der inneren Iteration fixiert, wohingegen der Index  $j$  der Laufindex der Iteration ist.*

- (0) Setze  $t_1 := t^{k-1}$ ,  $l_1 := 0$ ,  $L_1 := T$  und  $j := 1$ .
- (1) Berechne die Lösung  $(v_j, s_j) := (v_j(t_j), s_j(t_j))$  von (5.4), wobei die  $\alpha_i^k$  durch die  $\beta_i^k$  ersetzt wurden. Falls  $(1/t_j)\|s_j\| \leq \epsilon$  und  $-(1/t_j)\|s_j\|^2 - v_j \leq \epsilon$ : Stopp,  $u^k$  ist  $\epsilon$ -stationär. Ansonsten setze  $q_j := u^k + s_j$  und berechne  $g_j \in \partial f(q_j)$ .
- (a) Gilt **SS(i)**, so führe einen Abstiegschritt durch: Setze  $u^{k+1} := q^{k+1} := q_j$  und  $g^{k+1} := g_j$ ; stopp.
- (b) Gilt **NS(i)** und **NS(ii)** und **NS(iii)**, so führe einen Nullschritt durch: Setze  $u^{k+1} := u^k$ ,  $q^{k+1} := q_j$  und  $g^{k+1} := g_j$ ; stopp.
- (c) Gilt **NS(i)** und **NS(ii)**, nicht aber **NS(iii)**:
- i. Gilt der zweite Teil in **NS(ii)**, dann setze  $s^k := s_j$ ,  $v^k := v_j$  und führe eine Liniensuche entlang  $x^k + ds^k$ ,  $d > 0$  durch; stopp.
- ii. Sonst setze  $L_{j+1} := t_j$ ,  $l_{j+1} := l_j$  und  $t_{j+1} := \frac{1}{2}(L_{j+1} + l_{j+1})$ . Erhöhe  $j$  um 1 und gehe zu (1).
- (d) Gilt **NS(i)**, nicht aber **NS(ii)**, so setze  $L_{j+1} := t_j$ ,  $l_{j+1} := l_j$  und  $t_{j+1} := \frac{1}{2}(L_{j+1} + l_{j+1})$ . Erhöhe  $j$  um 1 und gehe zu (1).

Es lässt sich wieder zeigen [Sch89], dass die innere Iteration nach endlich vielen Schritten terminiert.

Der Algorithmus der äußeren Iteration entspricht Algorithmus 5.4, wobei die  $\alpha_i^k$  durch  $\beta_i^k$  ersetzt werden und der Short-Serious-Schritt wie ein Abstiegschritt behandelt wird.

Um Konvergenz des Algorithmus zu garantieren, müssen wir zusätzlich fordern, dass die Folge  $(u^k)_{k \in \mathbb{N}}$  beschränkt ist. Damit gelangen wir zu folgendem Konvergenzresultat:

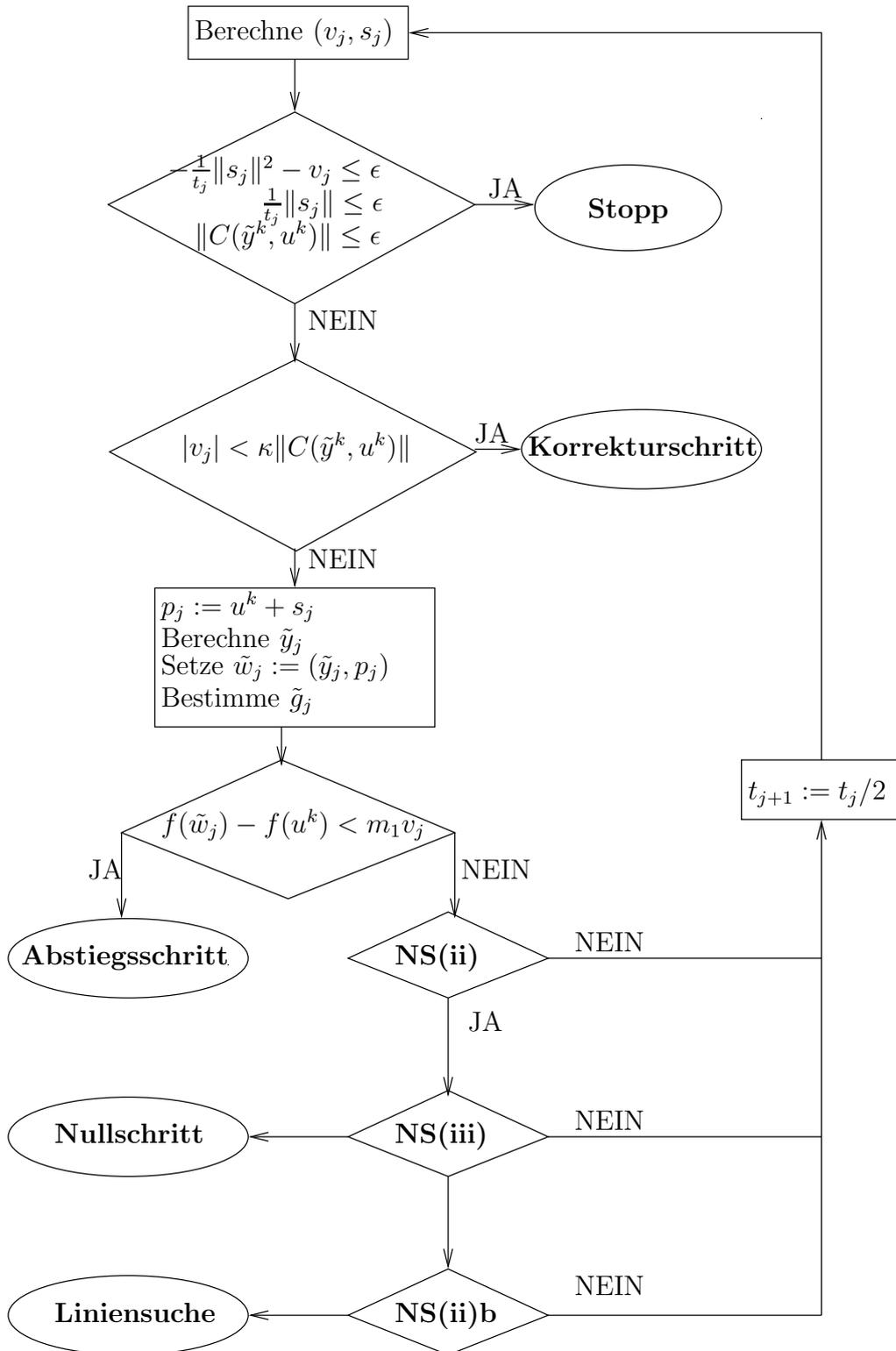


Abbildung 5.3: Flow Chart der inneren Iteration für nichtkonvexe Funktionen

**Satz 5.7** (Konvergenz des Bundle-Trust-Region-Verfahrens für nichtkonvexe Funktionen). *Sei  $f$  semiglatt und lokal Lipschitz-stetig auf  $\mathbb{R}^n$ . Wenn  $f$  von unten beschränkt ist und die Folge der Iterierten  $(p^k)_{k \in \mathbb{N}}$  des nichtkonvexen Bundle-Trust-Region-Verfahrens beschränkt ist, dann existiert ein Häufungspunkt  $\bar{p}$  der Folge  $(p^k)_{k \in \mathbb{N}}$ , so dass  $0 \in \partial f(\bar{p})$  gilt.*

*Beweis:* [Sch89]. ■

## 5.3 Inexakte Bundle-Trust-Region-Verfahren

Bei großen Optimierungsproblemen, wie sie in dieser Arbeit vorliegen, kann es sehr aufwändig sein, einen Subgradienten oder auch den Funktionswert exakt auszuwerten. Bevor wir in Kapitel 5.4 ein neues inexaktes Bundle-Trust-Region-Verfahren vorstellen, wollen wir bestehende Arbeiten kurz anreißen.

Für den Fall, dass die Funktionswerte exakt ausgewertet werden und gestörte Subgradienten vorliegen, sogenannte  $\epsilon$ -Subgradienten<sup>2</sup>, schlägt Hintermüller [Hin01] ein Verfahren für konvexe Optimierungsprobleme vor, mit dem der Funktionswert bis auf einen Fehler von  $\epsilon$ , dem Fehler der Subgradienten, bestimmt werden kann. Das Verfahren verhält sich bei exakten Subgradienten wie herkömmliche Bundle-Verfahren.

Solodov [Sol03] hat ein Bundle-Verfahren zur Minimierung konvexer Funktionen entwickelt, in dem sowohl Inexaktheit im Subgradienten als auch in der Funktionsauswertung zugelassen sind. In diesem Verfahren kann man die Fehler so steuern, dass die Lösung gegen das Optimum des ungestörten Problems konvergiert. Leider müssen die fehlerbehafteten Auswertungen kleiner sein als die exakten Funktionswerte. Dies liegt daran, dass die Funktion  $f$  wieder durch Schnittebenen von unten approximiert wird.

## 5.4 Ein neues inexaktes Bundle-Trust-Region-Verfahren

Wir entwickeln ein Bundle-Verfahren zur Minimierung einer semiglatten Funktion unter einer nichtglatten Nebenbedingung. Wir werten die Nebenbedingung nichtexakt aus und verwenden nichtexakte verallgemeinerte Gradienten.

Dieses Verfahren soll angewendet werden, um das Formoptimierungsproblem aus Kapitel Vier zu lösen. Wir werden in diesem Kapitel den Lösungsraum der Zustandsgleichung mit  $Y$  bezeichnen. Die ist ein Sobolev-Raum. Den Steuerungsraum, das heißt den Raum der Formparameter, werden wir hier im Gegensatz zu Kapitel Vier mit  $U$  bezeichnen. Es gilt  $U \subset \mathbb{R}^n$ .

---

<sup>2</sup>Definition siehe B.5

Hierzu orientieren wir uns an dem Bundle-Trust-Region-Verfahren für nichtkonvexe Funktionen, welches in Kapitel 5.2 vorgestellt wurde.

Gegeben sei das Problem

$$\min_{y \in Y, u \in U} f(y, u) \quad \text{unter der Nebenbedingung } C(y, u) = 0. \quad (5.14)$$

Wir bezeichnen  $X := Y \times U$ .

Wir nehmen folgendes an:

**Annahme 5.8.**

(a) Die Abbildungen

$$\begin{aligned} f : X &\rightarrow \mathbb{R} & (y, u) &\mapsto f(y, u) \\ C : X &\rightarrow Y^* & (y, u) &\mapsto C(y, u) \end{aligned}$$

sind Lipschitz-stetig und semiglatt. Ihre verallgemeinerten Ableitungen sind gleichmäßig beschränkt und oberhalb Lipschitz-stetig.

(b) Für  $u \in U$  existiert eine eindeutige Lösung  $y(u) \in Y$  von  $C(y(u), u) = 0$ .

(c) Es gilt  $\|C(y, u)\| \rightarrow \infty \Leftrightarrow \|u\| + \|y\| \rightarrow \infty$ .

(d) Ein Element des reduzierten Subgradienten lässt sich immer wie folgt darstellen:

$$\langle g, \cdot \rangle_{\mathbb{R}^n} = \langle f_u, \cdot \rangle_{\mathbb{R}^n} + \langle \mu, C_u(y(u), u)(\cdot) \rangle_{Y, Y^*} \quad (5.15)$$

$$\text{mit } C_y(y(u), u)^* \mu - f_y(y(u), u) = 0. \quad (5.16)$$

Dabei sind  $f_u, f_y, C_u, C_y$  geeignete Elemente der verallgemeinerten Ableitungen. Wir bezeichnen mit  $\langle \cdot, \cdot \rangle$  die duale Paarung auf  $Y$ .<sup>3</sup>

(e) Die Elemente der verallgemeinerten Ableitung  $C_y(y, u)$  sind invertierbar und die Inverse ist gleichmäßig beschränkt für alle  $(y, u) \in X$ .

Im Folgenden werden wir  $y$  Zustandsvariable und  $u$  Steuerungsparameter nennen. Ferner bezeichnet  $C$  die Zustandsgleichung. Sie ist zum Beispiel die schwache Formulierung einer partiellen Differentialgleichung, ihre Diskretisierung, die Optimalitätsbedingungen einer Variationsungleichung und in unserem Fall die Gleichungsformulierung der Optimalitätsbedingungen des regularisierten Kontaktproblems mit einer NCP-Funktion.

Wir entwickeln einen iterativen Algorithmus, bei dem die Zustandsgleichung in jedem Iterationsschritt nur zu einer vorgegebenen Genauigkeit gelöst wird. Dadurch ist der Funktionswert an der Stelle  $(y, u)$  nur näherungsweise berechenbar. Außerdem wird in jedem Iterationsschritt ein Element des verallgemeinerten Gradienten

<sup>3</sup>Man beachte, dass die Darstellung (4.56) dieser Darstellung entspricht, mit  $f_u = J_{\mathcal{V}} \mathcal{V}'$  und  $\langle \mu, C_u(y(u), u) \cdot \delta u \rangle = A'(u, \mu_1) \mathcal{V}' \cdot \delta p + F'(\mu_1) \mathcal{V}' \delta p$ . Die adjungierte Gleichung ist durch (4.46) gegeben.

näherungsweise berechnet. Dabei wird die Rechengenauigkeit während eines Programmablaufs zunehmend heraufgesetzt.

Wir entwickeln ein Trust-Region-Verfahren, das (5.14) löst.

### 5.4.1 Iterative Löser

Im Laufe der Optimierung werden wir zu gegebenen  $u^k \in U$  die Zustandsgleichung  $C(y(u^k), u^k)$  nach  $y(u^k) \in Y$  lösen. Die Lösung bezeichnen wir mit  $y^k$ . Dies wird iterativ geschehen. Hierzu werden wir das Mehrgitter-Newton-Verfahren aus Kapitel 3.4 verwenden. Außerdem muss die adjungierte Gleichung (5.16) gelöst werden. Dies erfolgt mit einem Mehrgitterverfahren, das wir wie in Kapitel 4.3 beschrieben einsetzen.

Wir bezeichnen mit  $\tilde{y}^k(u^k)$  die Näherung an  $y(u^k)$ , welches  $(C(y(u^k), u^k))$  zu einer vorgegebenen Toleranz löst. Dabei lassen wir manchmal das Argument von  $\tilde{y}^k(u^k)$  weg und schreiben verkürzt  $\tilde{y}^k$ . Wir setzen  $\tilde{x}^k := (\tilde{y}^k(u^k), u^k)$ .

Wir stellen an die iterativen Löser die Anforderung, dass bei zunehmend genauere Lösung der Zustandsgleichung die Näherungslösungen und die Näherungen an die Subgradienten gegen die exakten Subgradienten konvergieren.

**Annahme 5.9.** *Seien  $y(u)$  Lösung der Zustandsgleichung und  $\tilde{y}$  eine Näherung an die exakte Lösung, dann gibt es ein  $\delta > 0$  und ein  $\kappa > 0$  mit*

$$\|\tilde{y} - y(u)\|_Y \leq \kappa \|C(\tilde{y}, u)\|_{Y^*} \quad (5.17)$$

für  $\|\tilde{y} - y(u)\|_Y < \delta$ .

Das in dieser Arbeit verwendete semiglatte Newton-Verfahren erfüllt nach Satz 3.28 diese Eigenschaft, falls die Inversen der verallgemeinerten Ableitungen beschränkt sind und  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  mit  $\phi(0) = 0$  derart existiert, dass für die auf dem Kontaktrand zueinander komplementären Variablen  $y_2$  und  $y_3$

$$\text{meas} \{x \in \Gamma_C : |y_3(x)| + |y_2(x)| \leq t\} \leq \phi(t)$$

gilt.

Auch die Elemente des Subdifferentials werden nur approximativ bestimmt. Die hierbei berechneten Näherungen beruhen auf zwei Fehlerquellen. Zum einen bestimmen wir das Subdifferential nicht an der Stelle  $(y(u), u)$ , sondern nur an der Stelle  $(\tilde{y}, u)$ , der Approximation an die Lösung der Zustandsgleichung. Zum anderen wird die mit  $\tilde{\mu}$  bezeichnete Approximation der Lösung der adjungierten Gleichung (5.16) ebenfalls nur iterativ bis auf eine vorgegebene Genauigkeit  $tol_\mu$  bestimmt werden.

Wir fordern für den näherungsweise berechneten Gradienten:

**Annahme 5.10.** *Sei  $u \in U$  und  $\tilde{y}$  eine Approximation der Lösung  $y(u)$  der Zustandsgleichung und  $\tilde{\mu}$  eine Näherung an die Lösung der adjungierten Gleichung*

zur Genauigkeit  $tol_\mu$ . Dann kann man den Fehler zwischen dem näherungsweise bestimmten Subgradienten  $\tilde{g}$  und dem exakten Subgradienten  $g$  abschätzen nach

$$\|\tilde{g} - g\|_2 \leq \kappa_y \|\tilde{y} - y(u)\|_Y + \kappa_\mu tol_\mu + \Phi(\|\tilde{y} - y\|_Y)$$

mit  $\kappa_\mu, \kappa_y > 0$  und einer stetigen positiven Funktion  $\Phi : \mathbb{R} \rightarrow \mathbb{R}$  mit  $\Phi(0) = 0$ .

Wir haben in Abschnitt 4.3 gezeigt, dass die in den vorangegangenen Kapiteln vorgestellten Algorithmen diese Eigenschaften erfüllen, falls die Inversen der verallgemeinerten Ableitungen beschränkt sind, und eine stetige monotone Funktion  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  mit  $\phi(0) = 0$  derart existiert, dass

$$\text{meas} \{x \in \Gamma_C : |y_3(x)| + |y_2(x)| \leq t\} \leq \phi(t)$$

für die auf dem Kontaktrand zueinander komplementären Variablen  $y_2$  und  $y_3$  gilt.

## 5.4.2 Die Optimierung

Das Verfahren gehört zu der Klasse der Composite-Trust-Region-Verfahren (siehe auch [CGT00]). Wir werden den Abstiegsschritt  $s$  zerlegen in einen Quasi-Tangentialschritt und einen Quasi-Normalschritt.

$$s = s^n + s^t.$$

Die vorkommenden Vektoren haben die Bauart

$$s^n = \begin{pmatrix} s_y \\ 0 \end{pmatrix}, \quad s^t = \begin{pmatrix} 0 \\ s_u \end{pmatrix}.$$

Der Quasi-Tangentialschritt liefert einen Kandidaten für die nächste Iterierte in den Steuerungsvariablen, wohingegen der Quasi-Normalschritt die Zulässigkeit des Problems verbessert.

Zur Bestimmung des Quasi-Tangentialschritts werden wir eine Modellreduktion durchführen. Hierzu stellen wir eine Modellfunktion vor, die das auf die Steuerungsvariablen reduzierte Optimierungsproblem approximiert. Wir minimieren die Modellfunktion und nennen ihren Minimierer Quasi-Tangentialschritt. Das Minimum verwenden wir als Schätzer für den Abstieg.

Anschließend berechnen wir einen Quasi-Normalschritt, der ausgehend vom Quasi-Tangentialschritt das Residuum im Zustand verändert.

## 5.4.3 Der Quasi-Normalschritt

Im Laufe der Iterationen verlangen wir, dass die Zulässigkeit verbessert wird. Dabei ist hinsichtlich der Verbesserung keine strikte Monotonie notwendig, sondern wir erlauben in dem Fall, dass die Zulässigkeit wesentlich besser als die Optimalität ist,

dass die Zulässigkeit in einem gewissen Rahmen verschlechtert werden darf. Hierbei gehen wir vergleichbar zu [Ulb07] vor.

In jedem Iterationsschritt berechnen wir ausgehend von der aktuellen Iterierten  $(\tilde{y}, u)$  und einem vorgegebenen Quasi-Tangentialschritt  $s^t$  eine Richtung  $s^n$ , so dass

$$R \geq \|C(\tilde{y}(u) + s^n, u + s^t)\| \quad (5.18)$$

gilt. Hierbei ist  $R$  ein Relaxationsparameter, der bestimmt wird gemäß:

**Algorithmus 5.11** (Bestimmung des Relaxationsparameters  $R$ ). *Gegeben seien eine Nullfolge  $(a_j)_{j \in \mathbb{N}}$ ,  $\frac{a_{j+1}}{a_j} \geq \alpha_0 > 0$ ,  $\alpha, \beta \in (0, \frac{1}{2})$ ,  $\kappa_c \in (0, 1)$ ,  $\kappa > 0$*

*Eingabe:*  $\xi, \|\tilde{s}^t\|, v, \|C(\tilde{x})\|$ .

*Ausgabe:*  $R, \xi$ .

*Falls  $\|C(\tilde{x})\| \geq \min\{\alpha a_\xi, \beta \|\tilde{s}\|\}$  setze  $\tilde{R} := \kappa_c \|C(\tilde{x})\|$ .*

*Sonst setze  $\tilde{R} := \min\{a_\xi, \|\tilde{s}\|\}$ ,  $\xi := \xi + 1$ .*

*Setze  $R := \min\{-\frac{1}{\kappa}v, \tilde{R}\}$ .*

Dabei entspricht  $\tilde{s}$  der aktuellen Suchrichtung und  $v$  einem Schätzer für den aktuellen Abstieg.

#### 5.4.4 Berechnung eines genäherten Gradienten

Zur Berechnung einer Näherung für einen reduzierten Gradienten betrachten wir die Gleichungen (5.15) und (5.16). Wir werten die auftretenden Matrizen an den gestörten Stellen aus und lösen die adjungierte Gleichung (5.16) nur bis auf eine vorgegebene Genauigkeit. Als Näherung an den Gradienten erhalten wir

$$\tilde{g}^k = f_u(\tilde{y}^k(u^k), u^k) + \langle \tilde{\mu}^k, C_u(\tilde{y}^k(u^k), u^k) \rangle, \quad (5.19)$$

wobei  $\tilde{\mu}^k$  so gegeben ist, dass

$$\|C_y(\tilde{y}^k(u^k), u^k) \tilde{\mu}^k - f_y(\tilde{y}^k(u^k), u^k)\| < \kappa_\mu \min\{t^{k-1}, \|\zeta^{k-1}\|, \|C(\tilde{x}^k)\|\}. \quad (5.20)$$

Dabei ist  $t^{k-1}$  der aktuelle Trust-Region-Parameter und  $\zeta^{k-1}$  ein Schätzer für einen aktuellen Subgradienten.

**Bemerkung 5.12.** *Der Einfluss von  $t^{k-1}$ ,  $\|\zeta^{k-1}\|$  und  $\|C(\tilde{x}^k)\|$  auf die Genauigkeit der Lösung des adjungierten Problems in (5.20) kann durch Multiplikation jeder einzelnen Größe mit einem beliebigen positiven Faktor beeinflusst werden. Zur Vereinfachung haben wir die Größen gleich stark gewichtet.*

### 5.4.5 Bestimmung des Quasi-Tangentialschrittes

Als Modellfunktion zur Berechnung des Quasi-Tangentialschrittes werden wir ein Bundle-Trust-Region-Verfahren verwenden. Dazu reduzieren wir das Optimierungsproblem (5.14) auf die Steuerungsvariable  $u$ :

$$f(u) = f(y(u), u). \quad (5.21)$$

Wir betrachten wie bei der Konstruktion des Bundles für exakte Bundle-Verfahren im  $k$ -ten Schritt wieder eine Folge von Iterationspunkten  $u_1, \dots, u_k \in U$  und eine Menge  $J^k$ . Für  $i \in J^k$  haben wir Hilfspunkte  $q^i \in U$ . Im weiteren Verlauf nehmen wir an, dass wir für jeden Wert  $u^k$  eine Näherung  $\tilde{y}^k(u^k) \in Y$  für  $y(u^k)$  und für jeden Wert  $q^i, i \in J^k$  eine Approximation  $\tilde{r}^i(q^i) \in Y$  an die Lösung  $r(q^i)$  des Komplementaritätsproblems  $C(r(q^i), q^i) = 0$  nach Abschnitt 3.4 und Abschnitt 5.4.1 haben.

Wir setzen  $\tilde{x}^k := (\tilde{y}^k, u^k)$  und  $\tilde{w}^i := (\tilde{r}^i, q^i)$ . Wie in den Abschnitten 5.4.1 und 5.4.4 erläutert, bestimmen wir jetzt noch eine Näherung an einen verallgemeinerten reduzierten Gradienten aus  $\partial f(\tilde{r}^i(q^i), q^i)$  und stellen analog zu oben ein Schnitt-ebenenmodell für  $f$  an der Stelle  $u^k$  auf.

$$f_{CP}(\tilde{x}^k; u) := f(\tilde{x}^k) + \max_{i \in J^k} \{(\tilde{g}^i)^T(u - u^k) - \beta_i^k\}. \quad (5.22)$$

Analog zu oben setzen wir

$$\begin{aligned} \beta_i^k &:= \beta(\tilde{x}^k, \tilde{w}^i) \\ &:= \max \left\{ \alpha_i^k, c_0 \left( \|u^k - q^i\|^2 + \left| \|C(\tilde{y}^k(u^k), u^k)\| - \|C(\tilde{r}^i(q^i), q^i)\| \right| \right) \right\}, \\ \text{mit } \alpha_i^k &:= f(\tilde{x}^k) - (f(\tilde{w}^i) + (\tilde{g}^i)^T(u^k - q^i)). \end{aligned} \quad (5.23)$$

Die  $\beta_i^k$  in (5.23) unterscheiden sich von den  $\beta_i^k$  in (5.11) dadurch, dass im Fall negativer Linearisierungsfehler  $\alpha$  nicht nur der Abstand der Stützstellen, sondern auch der Betrag der Differenz der Fehler bei der Auswertung der Nebenbedingung eingeht. Statt

$$c_0 \left| \|C(\tilde{y}^k(u^k), u^k)\|_{Y^*} - \|C(\tilde{r}^i(q^i), q^i)\|_{Y^*} \right|$$

hätte man auch

$$c_0 \|C(\tilde{y}^k(u^k), u^k) - C(\tilde{r}^i(q^i), q^i)\|_{Y^*}$$

als Strafterm wählen können, allerdings muss dann in jedem Iterationsschritt entweder  $C(\tilde{r}^i(q^i), q^i)$  oder  $\tilde{r}^i(q^i)$  gespeichert werden. Der in (5.23) gewählte Strafterm ist nach der umgekehrten Dreiecksungleichung eine untere Schranke für den eben beschriebenen Strafterm. Im Fall exakten LöSENS der Zustandsgleichung und der adjungierten Gleichung stimmen die Strafterme mit denen des nichtkonvexen Bundle-Verfahrens überein.

Zur Vereinfachung setzen wir wieder  $s = u - u^k$  und vernachlässigen den konstanten Summanden  $f(\tilde{x}^k)$ .

$$f_{CP}(u^k; s) = \max_{i \in J^k} \{(\tilde{g}^i)^T s - \beta_i^k\}. \quad (5.24)$$

Wir bestimmen nun für vorgegebendes  $t > 0$

$$s(t) := \operatorname{argmin} \left\{ f_{CP}(u^k; s) + \frac{1}{2t} \|s\|^2; \quad s \in U \right\}. \quad (5.25)$$

Dies ist äquivalent zu dem quadratischen Optimierungsproblem

$$(v(t), s(t)) := \operatorname{argmin} \left\{ v + \frac{1}{2t} \|s\|^2 \quad | \quad v \geq (\tilde{g}^i)^T s - \beta_i^k, \quad i \in J^k \right\}. \quad (5.26)$$

Aus den KKT-Bedingungen für (5.26) folgt

**Satz 5.13.** Für  $n \in \mathbb{N}$  bezeichne

$$\Lambda(n) := \left\{ \lambda \in \mathbb{R}^n : \lambda_i \geq 0, 1 \leq i \leq n, \text{ und } \sum_{i=1}^n \lambda_i = 1 \right\}.$$

Dann existiert im Minimum  $(v(t), s(t))$  von (5.26)  $\lambda(t) \in \Lambda(|J^k|)$ , so dass

$$\lambda_i(t)(-v(t) + (\tilde{g}^i)^T s(t) - \beta_i^k) = 0 \quad \forall i \in J^k, \quad (5.27)$$

$$s(t) = -t \sum_{i \in J^k} \lambda_i(t) \tilde{g}^i, \quad (5.28)$$

$$v(t) = -t \left\| \sum_{i \in J^k} \lambda_i(t) \tilde{g}^i \right\|^2 - \sum_{i \in J^k} \lambda_i(t) \beta_i^k \quad (5.29)$$

$$= -\frac{1}{t} \|s\|^2 - \sum_{i \in J^k} \lambda_i(t) \beta_i^k \quad (5.30)$$

gelten.

Wir führen die im Folgenden häufig verwendeten Variablen

$$\zeta^k := \sum_{i \in J^k} \lambda_i^k \tilde{g}^i, \quad (5.31)$$

$$\sigma^k := \sum_{i \in J^k} \lambda_i^k \beta_i^k \quad (5.32)$$

ein.

Das zu (5.26) duale Problem lautet

$$\min \left\{ \frac{1}{2} \left\| \sum_{i \in J^k} \lambda_i \tilde{g}^i \right\|^2 + \frac{1}{t} \sum_{i \in J^k} \lambda_i \beta_i^k \mid \lambda \in \Lambda(|J^k|) \right\}. \quad (5.33)$$

Den Optimalwert des dualen Problems bezeichnen wir mit  $\omega^k$ . Es gilt

$$\omega^k = \frac{1}{2} \|\zeta^k\|^2 + \frac{1}{t^k} \sigma^k. \quad (5.34)$$

### 5.4.6 Der Korrektorschritt

In jeder Iteration entscheiden wir aufgrund der Ungenauigkeit der Lösung der Variationsungleichung und dem Minimierer  $v$  der Modellfunktion, ob das Modell zuverlässig genug ist, oder ob die Genauigkeit der Lösung der Variationsungleichung verbessert werden muss. Hier gehen wir ähnlich vor wie in [CGT00] beschrieben. Gegeben sei eine Konstante  $\kappa > 0$ . In jedem Schritt vergleichen wir, ob der Fehler beim Lösen der Nebenbedingung (Variationsungleichung) im letzten Schritt im Vergleich zu  $\kappa$ -mal dem vom Modell vorgeschlagenen Abstieg  $-v$  ist.

Gilt  $|v_j| < \kappa \|C(\tilde{y}^k(u^k), u^k)\|$ , so wird ein Korrektorschritt durchgeführt. Unter einem Korrektorschritt verstehen wir, dass  $u^{k+1} := u^k$ , also  $s^t = 0$ , gesetzt und  $\tilde{y}^{k+1}(u^{k+1})$  mit einem Quasi-Normalschritt nach (5.18) bestimmt wird.

**Bemerkung 5.14.** *Wir verwenden das gleiche  $\kappa$  wie bei der Bestimmung des Relaxationsparameters im Algorithmus 5.11*

### 5.4.7 Die innere Iteration

In der inneren Iteration bestimmen wir das nächste Glied der Iterationsfolge, wie es in Abbildung 5.4 illustriert wird.

**Algorithmus 5.15** (Innere Iteration). *Gegeben:  $m_1, m_2$  mit  $0 < m_1 < m_2 < 1$ ,  $0 < m_3$ ,  $\kappa > 0$ ,  $\nu > 0$  Abbruchkriterium  $\epsilon$ .*

*Eingabe:  $J^k, \xi^k, \tilde{x}^k, t^{k-1}, \tilde{w}^i, \beta_i^k, \tilde{g}^i$  für  $i \in J^k$ .*

*Ausgabe:  $\xi^{k+1}, \tilde{w}^{k+1}, \zeta^k, v^k, \sigma^k, t^{k+1}, \tilde{g}^{k+1}, R^{k+1}$ .*

*Der Index  $k$  ist während der inneren Iteration fixiert, wohingegen der Index  $j$  der Laufindex der Iteration ist.*

(0) Setze  $t_1 := t^{k-1}$ ,  $j := 1$ ,  $\xi_0 = x^k$ .

(1) Berechne Optimierungsrichtung  $s_j^t := s(t_j)$  und Abstiegsschätzer  $v_j := v(t_j)$  als Lösung des Näherungsproblems (5.26).

(a) Falls  $(1/t_j)\|s_j\| \leq \epsilon$  und  $-(1/t_j)\|s_j\|^2 - v_j \leq \epsilon$  und  $\|C(\tilde{y}^k(u^k), u^k)\| < \epsilon$  gilt: Stopp,  $u^k$  ist  $\epsilon$ -stationär.

(b) Falls  $|v_j| < \kappa \|C(\tilde{y}^k(u^k), u^k)\|$  gilt, führe einen Korrektorschritt durch: Berechne  $R_j, \xi_j$  nach Algorithmus 5.11. Berechne  $y_j(u^k)$  durch Quasi-Normalschritt (5.18); setze  $\tilde{w}_j := (\tilde{y}_j(u^k), u^k)$  und berechne  $\tilde{g}_j$  nach (5.19) und (5.20) als Näherung an einen Subgradienten an der Stelle  $w_j$ ; gehe zu (2).

(c) Bestimme Relaxationsparameter  $R_j, \xi_j$  nach Algorithmus 5.11; berechne  $p_j := u^k + s_j^t$  und bestimme  $\tilde{y}_j(p_j)$  in einem Quasi-Normalschritt (5.18). Setze  $\tilde{w}_j := (\tilde{y}_j(p_j), p_j)$  und berechne  $\tilde{g}_j$  nach (5.19) und (5.20) als Näherung an einen Subgradienten an der Stelle  $w_j$ .

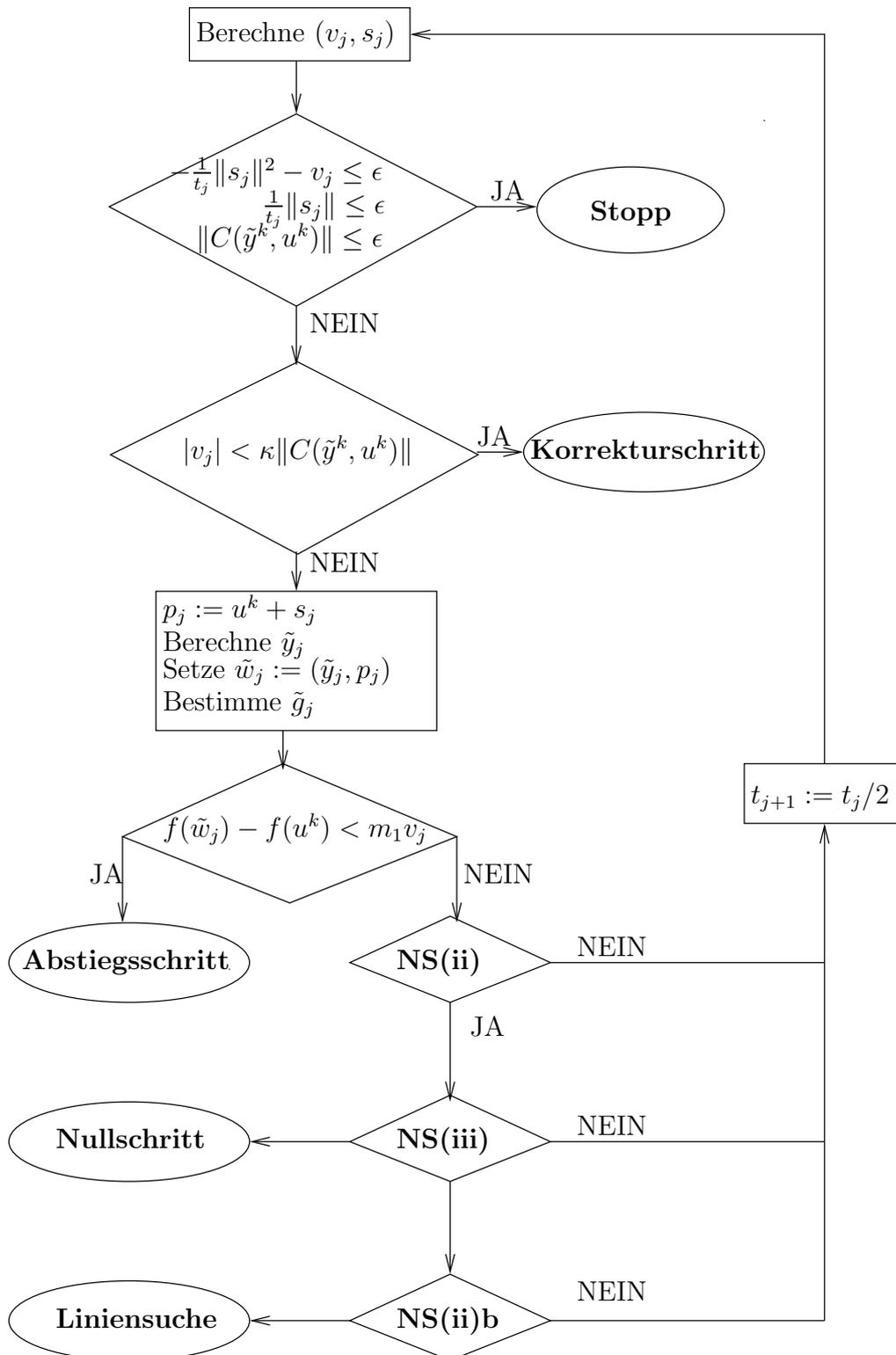


Abbildung 5.4: Flow Chart der inneren Iteration für ein nichtexaktes Bundle-Verfahren

- (d) Gilt das **SS**-Kriterium, dann führe einen Abstiegschritt durch; gehe zu (2).
- (e) Gelten **NS(i)**, **NS(ii)** und **NS(iii)**, so führe einen Nullschritt durch; gehe zu (2).
- (f) Gilt der zweite Teil in **NS(ii)** so führe eine Liniensuche durch; gehe zu (2).
- (g) Setze  $j := j + 1$  und  $t_{j+1} := t_j/2$  und gehe zu (1).
- (2) Setze  $\xi^{k+1} := \xi_j$ ,  $\tilde{w}^{k+1} := \tilde{w}_j$ ,  $v^k = v_j$ ,  $\tilde{g}^{k+1} := \tilde{g}_j$ ,  $R^{k+1} := R_j$ ,  $t^k := t_j$ ,  $v^k = v_j$ ; berechne  $\zeta^k$ ,  $\sigma^k$ ; stopp.

Dabei sind die Null- und Abstiegschrittkriterien wie oben gegeben durch

$$\begin{aligned} \text{SS:} & \quad f(\tilde{w}_j) - f(\tilde{x}^k) < m_1 v_j, \\ \text{NS: (i)} & \quad f(\tilde{w}_j) - f(\tilde{x}^k) \geq m_1 v_j, \\ \text{NS: (ii)} & \quad \beta(\tilde{x}^k, \tilde{w}_j) \leq m_3 \sigma^{k-1} \text{ oder } |f(\tilde{x}^k) - f(\tilde{w}_j)| \leq \|\zeta^{k-1}\| + \sigma^{k-1}, \\ \text{NS: (iii)} & \quad \tilde{g}_j^T s_j - \beta_j^k \geq m_2 v_j. \end{aligned}$$

Wir fassen noch einmal kurz zusammen, was wir unter einem Null-, einem Abstiegs- und einem Korrekturschritt verstehen, und erläutern die Bedeutung der Liniensuche.

Bei einem Abstiegschritt wird ein neuer Iterationspunkt in die Folge der Iterationspunkte eingeführt, dessen Funktionswert signifikant geringer als der des letzten Iterationspunktes ist.

Bei einem Nullschritt wird der alte Iterationspunkt zum neuen. Zusätzlich wird das Bundle durch einen genäherten Subgradienten vergrößert. Dies führt im nächsten Iterationsschritt zu einer anderen Suchrichtung.

Bei einem Korrekturschritt wird die Nebenbedingung mit einer höheren Genauigkeit als im letzten Schritt gelöst. Der neue Iterationspunkt setzt sich zusammen aus der Steuerungskomponente der vorangegangenen Iterierten und der verbesserten Näherungslösung des Zustandes.

Eine Liniensuche liefert entweder einen Nullschritt zur Verbesserung des Bundles oder einen Short-Serious-Schritt, welches ein verkürzter Abstiegschritt ist.

Wir zeigen nun, dass die innere Iteration mit einem Nullschritt, einem Abstiegschritt, einem Korrekturschritt oder einer Liniensuche abbricht.

**Satz 5.16.** *Die innere Iteration bricht nach endlich vielen Schritten ab.*

*Beweis:* Angenommen, die innere Iteration bricht nicht ab, dann gilt in jedem Schritt:

$$\begin{array}{ll} |v_j| \geq \kappa \|C(\tilde{x}^k)\| & \text{ist nach (b) erfüllt} \\ \text{SS} & \text{ist nach (d) nicht erfüllt} \\ \text{NS(i)} & \text{gilt nach (d)} \\ \text{NS(ii) und NS(iii)} & \text{gelten nicht gleichzeitig nach (e)} \\ \text{zweiter Teil in NS(ii)} & \text{gilt nicht nach (f)} \end{array}$$

Dann wird nach (g) in jedem Schritt  $t_j$  halbiert. Dadurch streben  $v_j$  und auch  $s_j^t$  gegen null. Wir betrachten 2 Fälle.

- (a) Im Fall  $\|C(\tilde{x}^k)\| = 0$  wird die Nebenbedingung für  $u^k$  exakt ausgerechnet. Wir können  $|f(\tilde{x}^k) - f(\tilde{w}_j)|$  nach oben mit der Dreiecksungleichung abschätzen durch

$$|f(\tilde{x}^k) - f(\tilde{w}_j)| \leq |f(\tilde{x}^k) - f(y(u^k + s_j^t), u^k + s_j^t)| + |f(y(u^k + s_j^t), u^k + s_j^t) - f(\tilde{w}_j)|. \quad (5.35)$$

Dabei bezeichnet  $y(u^k + s_j^t)$  die Lösung des Kontaktproblems an der Stelle  $u^k + s_j^t$ . Aufgrund der Lipschitz-Stetigkeit von  $f$  und  $y(u)$  konvergiert der erste Summand gegen Null. Die Relaxationsparameter  $R_j$  bilden eine Nullfolge, denn Algorithmus 5.11 liefert immer einen Relaxationsparameter, der kleiner als  $a_{\xi_j}$  ist und diese bilden eine Nullfolge. Somit folgt aus der Lipschitz-Stetigkeit von  $f$  und Annahme 5.9, dass der zweite Summand ebenfalls gegen Null strebt. Also ist die zweite Bedingung in **NS(ii)** irgendwann erfüllt, und die innere Iteration bricht ab.

- (b) Im Fall  $\|C(\tilde{x}^k)\| > 0$  folgt mit  $t_j \searrow 0$  auch  $-v(t_j) \searrow 0$ . Also existiert ein  $j \in \mathbb{N}$  mit  $-v(t_j) < \kappa \|C(\tilde{x}^k)\|$ . Somit ist die Korrekturschrittbedingung erfüllt und die innere Iteration wird verlassen.

■

Als nächstes werfen wir einen Blick auf die Liniensuche. Wir stellen den Algorithmus der Liniensuche auf und zeigen, dass er nach einer endlichen Anzahl an Schritten entweder mit einem Short-Serious Schritt oder einem Nullschritt abbricht. Hierzu gehen wir wie in [Lem81] vor.

### 5.4.8 Die Liniensuche

Wir führen auf der Mannigfaltigkeit

$$C(y(u), u) = C^k \quad (5.36)$$

eine Liniensuche durch. Dabei bezeichnet  $C^k$  das Residuum in  $u^k$ .

Die Liniensuche wird exakt durchgeführt. Dadurch verlassen wir unsere Philosophie der inexakten Rechnungen. Außerdem widerspricht die Liniensuche wie auch in [Sch89] der Idee der Trust-Region-Verfahren. Sie ist daher nur eine Notlösung. Bei der Liniensuche werden wir (5.36) exakt (sehr geringe Fehlerschranke) und die adjungierte mit zunehmender Genauigkeit ausrechnen.

Für die Liniensuche verwenden wir eine Variante des auf [Lem81] zurückgehenden

**Algorithmus 5.17** (Liniensuche). *Gegeben sind*  $0 < m_1 < m_2$ ,  $\delta_g > 0$ .

*Eingabe:*  $u^k$ ,  $s_j^t$ ,  $v_j$ ,  $j$ .

*Ausgabe:*  $\tilde{g}_j$ ,  $\tilde{w}_j$ .

- (0) Setze  $d_l := 0$  und  $d_r := 1$ ,  $\delta_n := \delta_g$ .
- (1) Löse (5.36) exakt für  $u := u^k + s_j^t$ . Falls **SS** erfüllt ist, führe einen Abstiegschritt mit  $\tilde{w}_j = (y(u), u)$  durch; gehe zu (3).
- (2) Setze  $n := 1$ .
- (a) Setze  $d_n := \frac{d_l + d_r}{2}$  und  $u_n := u^k + d_n s_j^t$ . Löse (5.36) exakt nach  $y_n := y(u_n)$ . Bestimme  $\tilde{g}_n := \tilde{g}(y(u_n), u_n)$  nach (5.19) mit Genauigkeit  $\delta_{g_n}$ .
- (b) Falls
- $$f(y(u_n), u_n) - f(\tilde{x}^k) < m_1 d_n v_j, \quad (5.37)$$
- $$\langle \tilde{g}_n, s_j^t \rangle \geq m_2 v_j \quad (5.38)$$
- gelten, dann führe einen Short-Serious-Schritt durch; gehe zu (3).
- (c) Falls
- $$-\beta(x^k, (y(u_n), u_n)) + \langle \tilde{g}_n, d_j^t \rangle \geq m_2 v_j \quad (5.39)$$
- gilt, führe einen Nullschritt durch; gehe zu (3).
- (d) Falls (5.37) gilt, so setze  $d_r^{n+1} := d_r^n$ ,  $d_l^{n+1} := d_n$ , sonst setze  $d_l^{n+1} := d_l^n$ ,  $d_r^{n+1} := d_n$ . Setze  $\delta_{n+1} := \delta_n/2$  und  $n := n + 1$ .
- (e) Gehe zu (a).
- (3) Setze  $\tilde{g}_j := \tilde{g}_n$ ,  $\tilde{w}_j := (y(u_n), u_n)$ ; stopp.

Wir bezeichnen mit  $g_n$  den exakten Subgradienten an der Stelle  $(y(u_n), u_n)$ . Wir zeigen:

**Satz 5.18.** *Die Liniensuche bricht entweder mit einem Short-Serious-Schritt, einem Abstiegschritt für  $d = 1$  oder einem Nullschritt ab.*

*Beweis:* Wir nehmen an, dass die Liniensuche nie abbricht. Dann unterscheiden wir drei Fälle.

- (a) Für alle  $n \in \mathbb{N}$  gilt  $d_r^n = 1$ .
- (b) Für alle  $n \in \mathbb{N}$  gilt  $d_l^n = 0$ .
- (c)  $(d_r^n)_{n \in \mathbb{N}}$  bildet eine fallende und  $(d_l^n)_{n \in \mathbb{N}}$  eine steigende Folge.

Zu (a): Wir bemerken zunächst, dass für  $d_n = 1$  die Bedingungen (5.37) und **SSS** übereinstimmen. Also ist nach Schritt (1) in Algorithmus 5.17 die Beziehung (5.37) für  $d_n = 1$  nicht erfüllt.

Die  $(d_l^n)_{n \in \mathbb{N}}$  bilden eine monoton wachsende Folge, die gegen 1 konvergiert. Weil aber  $f$  stetig ist, existiert eine Umgebung von  $d = 1$ , in der (5.37) nicht erfüllt ist.

Es existiert also ein  $N \in \mathbb{N}$ , so dass  $d^N$  (5.37) nicht erfüllt. Somit wird  $d_r^{N+1}$  gleich  $d^N$  gesetzt und wir sind in Fall (c).

Zu (b): Dieser Fall tritt auf, falls (5.37) nie erfüllt ist. Die  $d_r^n$  bilden eine Nullfolge und somit auch  $d_n$ . Aus (5.37) folgt somit zusammen mit der Richtungs-differenzierbarkeit von  $f$

$$f'(x^k, s_j^t) = \lim_{n \rightarrow \infty} \frac{f(y(u^k + d_n s_j^t), u^k + d_n s_j^t) - f(x^k)}{d_n} \geq m_1 v_j > m_2 v_j.$$

Weil  $f$  semiglatt ist, gilt

$$f'(x^k, s_j^t) = \lim_{n \rightarrow \infty} \langle g(y(u^k + d_n s_j^t), u^k + d_n s_j^t), s_j^t \rangle_{\mathbb{R}^n} > m_2 v_j.$$

Wir zeigen, dass die gestörte Richtungsableitung auch gegen die exakte Richtungsableitung konvergiert.

$$\begin{aligned} & |\langle g(y(u_n), u_n), s_j^t \rangle - \langle \tilde{g}(y(u_n), u_n), s_j^t \rangle_{\mathbb{R}^n} | \\ &= |\langle g(y(u_n), u_n) - \tilde{g}(y(u_n), u_n), s_j^t \rangle_{\mathbb{R}^n} | \\ &= |\langle f_u(y(u_n), u_n), s_j^t \rangle_{\mathbb{R}^n} + \langle \mu_n, C_u(y(u_n), u_n) \cdot s_j^t \rangle_{Y, Y^*} - \\ &\quad \langle f_u(y(u_n), u_n), s_j^t \rangle_{\mathbb{R}^n} - \langle \tilde{\mu}_n, C_u(y(u_n), u_n) \cdot s_j^t \rangle_{Y, Y^*} | \\ &\leq \|\mu - \tilde{\mu}\|_Y \|C_u(y(u_n), u_n) s_j^t\|_{Y^*} \\ &\leq \|C_y^{-*}(f_y(y(u_n), u_n) - (f_y(y(u_n), u_n) - res_n))\|_Y \|C_u(y(u_n), u_n) s_j^t\|_{Y^*} \\ &\leq \|res_n\| \|C_y^{-*}\| \|C_u(y(u_n), u_n) s_j^t\|_{Y^*}. \end{aligned}$$

Dabei bezeichnet  $res_n$  das Residuum, das im  $n$ -ten Schritt bei der Lösung der adjungierten Gleichung (5.20) auftritt. Also konvergiert die genäherte Richtungsableitung gegen die exakte Richtungsableitung. Da mit  $d_n \rightarrow 0$  auch  $\beta(x^k, (y(u_n), u_n)) \rightarrow 0$  gilt, ist (5.39) nach endlich vielen Schritten erfüllt.

Zu (c): Die  $d_r^n$  bilden eine absteigende Folge und die  $d_l^n$  eine aufsteigende Folge, die gegen den gleichen Grenzwert  $d^*$  konvergieren. Wegen

$$\begin{aligned} f(y(u^k + d_l^n s_j^t), u^k + d_l^n s_j^t) - f(\tilde{x}^k) &< m_1 d_l^n v_j, \\ f(y(u^k + d_r^n s_j^t), u^k + d_r^n s_j^t) - f(\tilde{x}^k) &\geq m_1 d_r^n v_j \end{aligned}$$

gilt

$$f(y(u^k + d^* s_j^t), u^k + d^* s_j^t) - f(\tilde{x}^k) = m_1 d^* v_j.$$

Also folgt

$$\begin{aligned} & f(y(u^k + d_l^n s_j^t), u^k + d_l^n s_j^t) - f(y(u^k + d^* s_j^t), u^k + d^* s_j^t) \\ &< f(x^k) + m_1 d_l^n v_j - (f(x^k) + m_1 d^* v_j) \\ &= m_1 (d_l^n - d^*) v_j. \end{aligned}$$

Dividiert man nun beide Seiten durch  $(d_l^n - d^*) < 0$  und betrachtet den Grenzwert  $n \rightarrow \infty$ , so erhält man

$$\begin{aligned}
& \lim_{n \rightarrow \infty} - \frac{f(y(u^k + d_l^n s_j^t), u^k + d_l^n s_j^t) - f(y(u^k + d^* s_j^t), u^k + d^* s_j^t)}{(d^* - d_l^n)} \\
&= -f'(y(u^k + d^* s_j^t), u^k + d^* s_j^t; -s_j^t) \\
&= - \lim_{n \rightarrow \infty} \langle g(y(u^k + d_l^n s_j^t), u^k + d_l^n s_j^t), -s_j^t \rangle_{\mathbb{R}^n} \\
&= - \lim_{n \rightarrow \infty} \langle \tilde{g}(y(u^k + d_l^n s_j^t), u^k + d_l^n s_j^t), -s_j^t \rangle_{\mathbb{R}^n} \\
&= \lim_{n \rightarrow \infty} \langle \tilde{g}(y(u^k + d_l^n s_j^t), u^k + d_l^n s_j^t), s_j^t \rangle_{\mathbb{R}^n} \\
&\geq m_1 v_j \\
&> m_2 v_j.
\end{aligned}$$

Also sind (5.37) und (5.38) erfüllt und die Liniensuche bricht nach endlich vielen Schritten ab. ■

### 5.4.9 Die äußere Iteration

Die äußere Iteration konstruiert eine Folge  $(\tilde{x}^k)_{k \in \mathbb{N}}$ , die gegen einen stationären Punkt  $\bar{x} = (\bar{y}, \bar{u})$  von  $f$  unter der Nebenbedingung  $C(\bar{y}, \bar{u}) = 0$  konvergiert. Gleichzeitig generiert sie eine Folge an Hilfspunkten  $(\tilde{w}^k, \tilde{g}^k, \beta_i^k)_{k \in \mathbb{N}, i \leq k}$ , die die Modellfunktion beschreiben. Wir geben nun den Algorithmus der äußeren Iteration an.

**Algorithmus 5.19** (Äußere Iteration). *Gegeben seien Startpunkt  $u^1 \in \mathbb{R}^n$  und Parameter  $0 < m_1 < m_2 < 1$ ,  $0 < m_3 < 1$ ,  $\kappa > 0$   $\epsilon \geq 0$ .*

- (a) *Berechne  $\tilde{y}^1(u^1)$  und  $\tilde{g}^1$  zu vorgegebenen Toleranzen. Setze  $\tilde{x}^1 := (\tilde{y}^1(u^1), u^1)$ ,  $q^1 := u^1$ ,  $\tilde{r}^1 = \tilde{y}^1$ ,  $\tilde{w}^1 := \tilde{x}^1$ ,  $\zeta^1 := g$ ,  $\sigma^1 := 1$ ,  $J^1 := \{1\}$ . Setze  $k := 1$ .*
- (b) *Führe eine innere Iteration durch.*
- (c) *Endet die innere Iteration mit  $\epsilon$ -stationärer Lösung: Stopp, stationärer Punkt bei  $\tilde{x}^k$ .*
- (d) *Endet die innere Iteration mit Korrekturschritt: Setze  $J^{k+1} := J^k \cup \{k+1\}$ ,  $\tilde{x}^{k+1} := (\tilde{y}^{k+1}, u^{k+1}) := (\tilde{w}^{k+1})$ ,  $k := k+1$ ; aktualisiere Strafterme nach (5.23). Gehe zu (b).*
- (e) *Endet die innere Iteration mit einem Abstiegs- oder Short-Serious-Schritt: Setze  $J^{k+1} := J^k \cup \{k+1\}$ ,  $\tilde{x}^{k+1} := (\tilde{y}^{k+1}, u^{k+1}) := (\tilde{w}^{k+1})$ ,  $k := k+1$ ; aktualisiere Strafterme nach (5.23). Gehe zu (b).*
- (f) *Endet die innere Iteration mit einem Nullschritt: Setze  $J^{k+1} := J^k \cup \{k+1\}$ ,  $\tilde{x}^{k+1} := (\tilde{y}^{k+1}, u^{k+1}) := (\tilde{x}^k)$ ,  $k := k+1$  und aktualisiere die Strafterme nach (5.23). Gehe zu (b).*

## 5.5 Konvergenz des neuen inexakten Bundle-Trust-Region-Verfahrens

Wir beweisen die Konvergenz unseres neuen BTR-Verfahrens. Dabei setzen wir wie [Sch89, Hin01, SZ92a] voraus, dass die Folge  $(u^k)_{k \in \mathbb{N}}$  beschränkt ist. Für den Beweis verwenden wir die Lösung des dualen Problems (5.33) mit  $\omega^k = \frac{1}{2} \|\zeta^k\|^2 + \frac{1}{t^k} \sigma^k$ .

**Hilfssatz 5.20.** *Die Folge  $(u^k)_{k \in \mathbb{N}}$  sei beschränkt. Dann sind ebenfalls die Folgen  $(x^k)_{k \in \mathbb{N}}$ ,  $(\tilde{x}^k)_{k \in \mathbb{N}}$ ,  $(q^k)_{k \in \mathbb{N}}$ ,  $(\tilde{w}^k)_{k \in \mathbb{N}}$ ,  $(g^k)_{k \in \mathbb{N}}$ ,  $(\tilde{g}^k)_{k \in \mathbb{N}}$ ,  $(\omega^k)_{k \in \mathbb{N}}$  und  $(\sigma^k)_{k \in \mathbb{N}}$  beschränkt.*

*Beweis:* Nach Konstruktion des Algorithmus gilt  $\|C(\tilde{y}^k, u^k)\| < \infty$ . Wegen der Annahme  $\|C(y, u)\| \rightarrow \infty \Leftrightarrow \|u\| + \|y\| \rightarrow \infty$  folgt die Beschränktheit von  $y(u^k)$ . Somit sind auch  $\tilde{y}^k$ ,  $\tilde{x}^k = (\tilde{y}^k(u^k), u^k)$  und  $f(\tilde{x}^k)$  beschränkt. Ferner gilt nach (5.33) und (5.34)

$$0 \geq -\omega^k = \frac{1}{t^k} (\max_{i \in J^k} \{-\beta_i^k + \langle \tilde{g}^i, s^{t,k} \rangle\}) + \frac{1}{2t^k} \|s^{t,k}\|^2. \quad (5.40)$$

Wir wählen  $j(k) \in J^k$  so, dass  $\beta_{j(k)}^k = 0$  ist. Dann folgt

$$0 \geq -\omega^k \geq \frac{1}{t^k} \left( \min_d \{ \langle \tilde{g}^{j(k)}, d \rangle + \frac{1}{2t^k} \|d\|^2 \} \right). \quad (5.41)$$

Mit  $s^{t,k} := -t^k \tilde{g}^{j(k)} = \operatorname{argmin}_d \{ \langle \tilde{g}^{j(k)}, d \rangle + \frac{1}{2t^k} \|d\|^2 \}$  folgt

$$0 \geq \omega^k \geq -\frac{1}{2} \|\tilde{g}^{j(k)}\|^2. \quad (5.42)$$

Die exakten Subgradienten sind wegen der Beschränktheit der Iterationswerte ebenfalls beschränkt. Somit folgt mit der Beschränktheit der adjungierten Variable die Beschränktheit der  $\tilde{g}^k$ . Also sind mit der Beschränktheit von  $t^k$  die Folgen  $v^k$ ,  $\zeta^k$  und  $\sigma^k$  beschränkt und damit auch  $p^{k+1} = u^k - t^k \zeta^k$ . Somit ist ebenso  $\tilde{w}^{k+1} = (\tilde{q}^{k+1}, u^k - t^k \zeta^{k+1})$  beschränkt.

■

**Hilfssatz 5.21.** *Falls die Folge  $(t^k)_{k \in \mathbb{N}}$  den Häufungspunkt 0 hat, dann gilt für jede Teilfolge  $t^{k(l)} \rightarrow 0$  auch  $v^{k(l)} \rightarrow 0$ .*

*Beweis:* Wir wissen nach (5.26), dass  $(v^k, s^k)$  die Lösung des quadratischen Optimierungsproblems

$$(v^k, s^k) = \operatorname{argmin} \left\{ v + \frac{1}{2t^k} \|s\|^2 \quad : \quad v \geq (\tilde{g}^i)^T s - \beta_i^k, i \in J^k \right\}$$

beschreibt. Wir wählen  $j(k) \in J^k$  so, dass  $\beta_{j(k)}^k = 0$  ist. Dann ist insbesondere die  $j(k)$ -te Nebenbedingung erfüllt. Es gilt also

$$v^k \geq (\tilde{g}^{j(k)})^T s^k - \beta_{j(k)}^k = (\tilde{g}^{j(k)})^T s^k. \quad (5.43)$$

Mit  $s^k = -t^k \sum_{i \in J^k} \lambda_k^i \tilde{g}^i$  nach (5.28) und der Beschränktheit der  $\tilde{g}$  durch  $c$  folgt unter Verwendung der Cauchy-Schwarzschen-Ungleichung

$$v^k \geq -t^k \left( \sum_{i \in J^k} \lambda_k^i \tilde{g}^i \right)^T (\tilde{g}^{j(k)}) \geq -t^k c^2. \quad (5.44)$$

Gibt es nun eine Teilfolge  $(t^{k(l)})_{l \in \mathbb{N}}$  von  $(t^k)_{k \in \mathbb{N}}$ , die gegen 0 konvergiert, so konvergiert  $v^{k(l)}$  ebenfalls gegen 0, da die  $v^k$  alle nicht positiv sind. ■

**Hilfssatz 5.22.** *Die Folge  $(u^k)_{k \in \mathbb{N}}$  sei beschränkt und 0 Häufungspunkt der Folge  $(t^k)_{k \in \mathbb{N}}$ . Dann existiert eine Teilfolge  $(k(l))_{l \in \mathbb{N}}$ , so dass gilt:*

$$\lim_{l \rightarrow \infty} \zeta^{k(l)} = 0 \text{ und } \lim_{l \rightarrow \infty} \sigma^{k(l)} \rightarrow 0. \quad (5.45)$$

*Beweis:* Wir führen einen Widerspruchsbeweis und nehmen dazu an, dass ein  $\delta > 0$  existiert mit

$$\|\zeta^k\| + \sigma^k \geq \delta \quad \forall k \in \mathbb{N}.$$

Nach der Konstruktion des Algorithmus ist die Folge  $(t^k)_{k \in \mathbb{N}}$  monoton fallend, also konvergiert sie gegen den Häufungspunkt 0. Also existiert wegen der Lipschitz-Stetigkeit von  $f$ , (5.17), (5.18), Algorithmus 5.11 und Hilfssatz 5.21 ein  $K \in \mathbb{N}$  mit

$$|f(\tilde{y}^k, u^k) - f(y(u^k), u^k)| \leq \frac{\delta}{3} \quad \forall k \geq K. \quad (5.46)$$

Im weiteren sei  $t > 0$  beliebig und  $s^k(t) = -t\zeta^k(t)$  für  $k \in \mathbb{N}$  die Lösung des Trust-Region-Problems

$$(v^k(t), s^k(t)) = \operatorname{argmin} \left\{ v + \frac{1}{2t} \|s\|^2 \quad : \quad v \geq (\tilde{g}^i)^T s - \beta_i^k, i \in J^k \right\}$$

mit Parameter  $t$  in der  $k$ -ten Iteration. Also folgt mit der lokalen Lipschitz-Stetigkeit von  $f$  bezüglich  $y$  nach Satz 4.12 und  $u$  sowie der Lipschitz-Stetigkeit von  $y(u)$  der Lösung des Kontaktproblems bezüglich  $u$ , dass es ein  $L > 0$  gibt mit

$$|f(y(u^k + s^k), u^k + s^k) - f(y(u^k), u^k)| \leq L \|u^k + s^k - u^k\| = L \|s^k\| = tL \|\zeta^k\|.$$

Mit der Beschränktheit der Folgen  $(\tilde{g}^k)_{k \in \mathbb{N}}$  nach Hilfssatz 5.20 existiert also ein  $c > 0$  mit

$$|f(y(u^k + s^k), u^k + s^k) - f(y(u^k), u^k)| \leq tLc.$$

Es existiert also ein  $T > 0$  mit  $|f(y(u^k + s^k), u^k + s^k) - f(y(u^k), u^k)| < \frac{\delta}{3}$  für  $t < T$ .

Weil im Algorithmus zur Bestimmung der Relaxationsparameter 5.11 der Parameter kleiner  $-\frac{1}{\kappa} v_j$  ist, gilt für jede Näherung  $\tilde{y}(u^k + s^k(t))$  an  $y(u^k + s^k(t))$  ebenfalls (5.46).

Insgesamt gilt

$$\begin{aligned}
& |f(\tilde{y}(u^k + t\zeta^k(t)), u^k + t\zeta^k(t)) - f(y^k, u^k)| \\
\leq & |f(\tilde{y}(u^k + t\zeta^k(t)), u^k + t\zeta^k(t)) - f(y(u^k + t\zeta^k(t)), u^k + t\zeta^k(t))| \\
& + |f(y(u^k + t\zeta^k(t)), u^k + t\zeta^k(t)) - f(y(u^k), u^k)| \\
& + |f(y(u^k), u^k) - f(\tilde{y}^k, u^k)| \\
\leq & \frac{\delta}{3} + \frac{\delta}{3} + \frac{\delta}{3} = \delta \\
\leq & \|\zeta^{k-1}\| + \sigma^{k-1}
\end{aligned}$$

für  $k > K$  und  $t < T$ . Also ist für  $k > K$  und  $t < T$  der zweite Teil von **NS(ii)** immer erfüllt. Somit wird immer ein  $t > \frac{1}{10} \min\{T, t^K\}$  akzeptiert, was der Voraussetzung des Hilfssatzes widerspricht.

■

**Hilfssatz 5.23.** *Wenn der Algorithmus unendlich viele Korrekturschritte erzeugt und die Folge  $(t^k)_{k \in \mathbb{N}}$  durch ein  $\bar{t} > 0$  nach unten beschränkt ist, so existiert eine Teilfolge  $(k(l))_{l \in \mathbb{N}}$  mit*

$$\lim_{l \rightarrow \infty} \zeta^{k(l)} = 0 \text{ und } \lim_{l \rightarrow \infty} \sigma^{k(l)} \rightarrow 0.$$

*Beweis:* Wir unterscheiden, ob es ein  $K \in \mathbb{N}$  gibt, so dass im Algorithmus 5.11  $\xi^k = \xi^K$  für alle  $k \geq K$  gilt.

- (a) Es existiert ein  $K$  mit obiger Eigenschaft: Sei  $(k(l))_{l \in \mathbb{N}}$  die Folge der Schritte mit  $|v^{k(l)}| < \kappa \|C(\tilde{x}^{k(l)})\|$  für alle  $k(l) > K$ . Nach der Konstruktion bilden die  $\|C(\tilde{x}^k)\|$ ,  $k > K$ , eine (nicht streng!) monoton fallende Folge. Ferner gilt

$$\|C(\tilde{x}^{k(l)})\| \leq \kappa_c \|C(\tilde{x}^{k(l)-1})\|, \quad (5.47)$$

falls  $k(l) > K + 1$  ist. Wir setzen  $l_0 = \min\{l \in \mathbb{N} : k(l) > K\}$ .

Wir zeigen mit Induktion:  $\|C(\tilde{x}^{k(l)})\| \leq \kappa_c^{l-l_0} \|C(\tilde{x}^{k(l_0)})\|$  für alle  $l > l_0$ .

Mit  $k(l) \geq 2$  folgt die Verankerung unmittelbar aus

$$\|C(\tilde{x}^K)\| \geq \|C(\tilde{x}^{k(l_0)})\| \geq \frac{1}{\kappa_c} \|C(\tilde{x}^{k(l_0+1)})\|.$$

Also folgt für den Induktionsschritt

$$\|C(\tilde{x}^{k(l+1)})\| \leq \kappa_c \|C(\tilde{x}^{k(l+1)-1})\| \leq \kappa_c \|C(\tilde{x}^{k(l)})\| \leq \kappa_c^{l-l_0+1} \|C(\tilde{x}^{K+1})\|.$$

Also gilt  $|v^{k(l)}| < \kappa \|C(\tilde{x}^{k(l)})\| \leq \kappa_c^{l-l_0-1} \|C(\tilde{x}^{K+1})\|$ . Somit bilden die  $|v^{k(l)}|$  eine Nullfolge. Mit  $t^k > \bar{t} > 0$  und  $v^{k(l)} = -t^{k(l)} \|\zeta^{k(l)}\|^2 - \sigma^{k(l)}$  folgt die Behauptung.

- (b) Es existiert kein  $K$  mit obiger Eigenschaft. Sei  $(k(l))_{l \in \mathbb{N}}$  die Folge der Korrekturschritte, bei denen in Algorithmus 5.11  $\xi^{k(l)} \neq \xi^{k(l)-1}$  gilt. Dann folgt nach Konstruktion, dass

$$\|C(\tilde{x}^{k(l)})\| \leq a_{\xi^{k(l)-1}}$$

gilt. Wegen

$$|v^{k(l)}| < \kappa \|C(\tilde{x}^{k(l)})\| \leq \kappa a_{\xi^{k(l)-1}} \rightarrow 0 \text{ für } l \rightarrow \infty.$$

Also folgt mit  $t^k > \bar{t} > 0$  und  $v^{k(l)} = -t^{k(l)} \|\zeta^{k(l)}\|^2 - \sigma^{k(l)}$  wieder die Behauptung.

■

**Hilfssatz 5.24.** *Wenn der Algorithmus eine beschränkte Folge von Iterationspunkten erzeugt, die Folge  $(t^k)_{k \in \mathbb{N}}$  durch  $\bar{t} > 0$  nach unten beschränkt ist und nur endlich viele Korrekturschritte durchgeführt werden, so existiert im Fall unendlich vieler Abstiegsschritte eine Teilfolge  $(k(l))_{l \in \mathbb{N}}$  mit*

$$\lim_{l \rightarrow \infty} \zeta^{k(l)} = 0 \text{ und } \lim_{l \rightarrow \infty} \sigma^{k(l)} = 0.$$

*Beweis:* Weil nur endlich viele Korrekturschritte erfolgen, existiert nun ein  $K \in \mathbb{N}$  mit  $|v^{k+1}| \geq \kappa \|C(x^k)\|$  für alle  $k > K$ . Für  $k > K$  bildet  $f(\tilde{x}^k)$  eine monoton fallende Folge, weil nur noch Null-, Short-Serious- oder Abstiegsschritte durchgeführt werden. Nach Voraussetzung werden unendlich viele Abstiegsschritte ausgeführt.

Wir betrachten nun eine Teilfolge  $(k(l))_{l \in \mathbb{N}}$ , mit  $k(l) > K$ , für die das Abstiegschritt-Kriterium **SS** erfüllt ist. Somit gilt  $f(\tilde{x}^{k(l)+1}) - f(\tilde{x}^{k(l)}) < m_1 v^{k(l)}$ . Wir zeigen nun, dass  $\lim_{l \rightarrow \infty} v^{k(l)} = 0$  ist. Für jedes  $l$  gilt

$$f(\tilde{x}^{k(l)+1}) - f(\tilde{x}^{k(l)}) \leq m_1 v^{k(l)}.$$

Durch Aufsummieren erhalten wir

$$f(\tilde{x}^{k(l)+1}) - f(\tilde{x}^{k(2)}) \leq m_1 \sum_{j=2}^l v^{k(j)}.$$

Dabei haben wir ausgenutzt, dass  $f(\tilde{x}^{k+1}) \leq f(\tilde{x}^k)$  für  $k > K$  gilt, denn aufgrund möglicher Short-Serious-Schritte erhält man nicht unbedingt eine Teleskopsumme. Mit der Negativität der  $v^k$  führt dies zu

$$0 \leq - \sum_{j=2}^l v^{k(j)} \leq \frac{1}{m_1} (f(\tilde{y}^{k(2)}, u^{k(2)}) - f(\tilde{y}^{k(l)+1}, u^{k(l)+1})).$$

Aus der Beschränktheit der Iterationswerte und der Zielfunktion folgt für  $l \rightarrow \infty$

$$0 \leq - \sum_{j=2}^{\infty} v^{k(j)} < \infty \tag{5.48}$$

und somit gilt

$$\lim_{l \rightarrow \infty} v^{k(l)} = 0. \quad (5.49)$$

Aus  $v^{k(l)} = -t^{k(l)} \|\zeta^{k(l)}\|^2 - \sigma^{k(l)}$  folgt mit der Beschränktheit von  $(t^k)_{k \in \mathbb{N}}$  die Behauptung. ■

**Hilfssatz 5.25.** *Wenn der Algorithmus eine beschränkte Folge von Iterationspunkten und endlich viele Korrekturschritte erzeugt und die Folge  $(t^k)_{k \in \mathbb{N}}$  durch  $\bar{t} > 0$  nach unten beschränkt ist, so gilt, wenn unendlich viele Short-Serious-Schritte in den Iterationen  $k(l)$  durchgeführt werden,*

$$\lim_{l \rightarrow \infty} d^{k(l)} v^{k(l)} = 0, \quad (5.50)$$

$$\lim_{l \rightarrow \infty} \|u^{k(l)+1} - u^{k(l)}\| = 0. \quad (5.51)$$

*Beweis:* Analog zum Beweis in Hilfssatz 5.24 zeigen wir für den Fall, dass  $t^k > \bar{t} > 0$  und nur endlich viele Korrekturschritte durchgeführt werden, dass es nach (5.37) eine Folge  $(\tilde{x}^{k(l)})_{l \in \mathbb{N}}$  gibt, mit

$$f(\tilde{x}^{k(l)+1}) - f(\tilde{x}^{k(l)}) < m_1 d^{k(l)} v^{k(l)}.$$

Dazwischen finden bei genügend großem  $k(l)$  nur noch Abstiegs- und Nullschritte statt. Dann folgern wir

$$f(\tilde{x}^{k(l)+1}) - f(\tilde{x}^{k(2)}) \leq m_1 \sum_{j=2}^l d^{k(j)} v^{k(j)},$$

und somit folgt aus der Beschränktheit der Iterationspunkte und der Zielfunktion, dass

$$0 \leq - \sum_{j=2}^{\infty} d^{k(j)} v^{k(j)} < \infty.$$

Dabei haben wir verwendet, dass aus Stetigkeitsgründen  $f$  ausgewertet an den gestörten Stellen  $\tilde{x}^k$  ebenso beschränkt ist wie an den exakten  $x^k$ . Also gilt (5.50).

Aus (5.50) und der Multiplikation von (5.29) mit  $d^{k(l)}$  folgt unter Verwendung von  $\zeta^{k(l)} = \sum_{i \in J^{k(l)}} \lambda_i^{k(l)} \tilde{g}^i$  und der Beschränktheit der  $(\tilde{g}^k)_{k \in \mathbb{N}}$ , dass

$$\lim_{l \rightarrow \infty} d^{k(l)} t^{k(l)} \|\zeta^{k(l)}\|^2 = 0.$$

gilt Wegen  $u^{k(l)+1} - u^{k(l)} = d^{k(l)} t^{k(l)} \zeta^{k(l)}$  gilt also

$$\lim_{l \rightarrow \infty} \frac{\|u^{k(l)} - u^{k(l)+1}\|^2}{d^{k(l)} t^{k(l)}} = 0.$$

Mit  $d^{k(l)} t^{k(l)} > 0$  für alle  $l \in \mathbb{N}$  und der Beschränktheit der  $d^{k(l)} t^{k(l)}$  folgt die Behauptung. ■

**Hilfssatz 5.26.** *Wenn der Algorithmus eine beschränkte Folge von Iterationspunkten und nur endlich viele Korrektur- und Abstiegschritte erzeugt und ferner die Folge  $(t^k)_{k \in \mathbb{N}}$  durch  $\bar{t} > 0$  nach unten beschränkt ist, so existiert eine Teilfolge  $(k(l))_{l \in \mathbb{N}}$  mit*

$$\lim_{l \rightarrow \infty} \zeta^{k(l)} = 0 \text{ und } \lim_{l \rightarrow \infty} \sigma^{k(l)} = 0. \quad (5.52)$$

*Beweis:* Da nur endlich viele Korrektur- oder Abstiegschritte durchgeführt werden, existiert ein  $K \in \mathbb{N}$ , so dass im  $k$ -ten Schritt nur Nullschritte oder Short-Serious-Schritte durchgeführt werden. Außerdem gilt  $t^k \geq t^{k+1}$ . Also ist  $t^k$  monoton fallend und beschränkt, also konvergent gegen  $\bar{t}$ . Nach der Voraussetzung ist  $\bar{t} > 0$ .

- (a) Nach Hilfssatz 5.20 ist die Folge  $\omega^k$  beschränkt.
- (b) Im Weiteren sei der Index  $k$  fest gewählt und wird vernachlässigt. Es sei  $t = t^k$ ,  $t^+ = t^{k+1}$ ,  $v = v^k$ ,  $\omega = \omega^k$ ,  $\omega^+ = \omega^{k+1}$ ,  $\tilde{g} = \tilde{g}^k$  und  $\tilde{g}^+ = \tilde{g}^{k+1}$ . Weiter sei  $\zeta := \sum_{i \in J^k} \lambda_i^k \tilde{g}^i$ .

Mit  $\lambda_i^k$  bezeichnen wir die  $\lambda$ 's aus Satz 5.13 im  $k$ -ten Schritt.

Wurde von  $k$  nach  $k+1$  ein Nullschritt durchgeführt, so ist  $\beta(\tilde{x}^{k+1}, \tilde{w}^i) = \beta(\tilde{x}^k, \tilde{w}^i)$  für  $i \in J^k$  und  $\beta_{k+1}^{k+1} \geq 0$ .

Wurde dagegen ein Short-Serious-Schritt durchgeführt, so werden die Strafterme nach (5.23) geändert. Dann gilt  $\beta_i^{k+1} = \beta(\tilde{x}^{k+1}, \tilde{w}^i)$  für  $i \in J^k$  und  $\beta_{k+1}^{k+1} = 0$ .

Analog zu oben setzen wir  $\beta_i = \beta_i^k$ ,  $\beta_i^+ = \beta_i^{k+1}$  für  $i \in J^k$  und  $\sigma := \sum_{i \in J^k} \lambda_i^k \beta_i$ ,  $\sigma^+ := \sum_{i \in J^k} \lambda_i^k \beta_i^+$  sowie  $\beta_+^+ = \beta_{k+1}^{k+1}$ .

- (c) Wir zeigen jetzt, wie sich das Modell, genauer der Optimalwert  $\omega^k$  des zu (5.26) dualen Problem (5.33) bei einem Null- oder Short-Serious-Schritt ändert. Hierzu betrachten wir alle Konvexkombinationen von  $\zeta$  und dem zuletzt berechneten Subgradienten  $g^+$ .

Wir setzen

$$Q(\nu) := \frac{1}{2} \|(1-\nu)\zeta + \nu g^+\|^2 + (1-\nu) \frac{1}{t^+} \sigma + \nu \frac{1}{t^+} \beta_+^+ + (1-\nu) \frac{1}{t^+} (\sigma^+ - \sigma). \quad (5.53)$$

Falls in Schritt  $k$  ein Nullschritt stattgefunden hat, gilt

$$Q(\nu) := \frac{1}{2} \|(1-\nu)\zeta + \nu g^+\|^2 + (1-\nu) \frac{1}{t^+} \sigma + \nu \frac{1}{t^+} \beta_+^+$$

und im Falle eines Short-Serious-Schritts

$$Q(\nu) := \frac{1}{2} \|(1-\nu)\zeta + \nu g^+\|^2 + (1-\nu) \frac{1}{t^+} \sigma^+.$$

Damit ist  $\omega^+ \leq \min_{\nu \in [0,1]} Q(\nu) =: \tilde{\omega}$ . Wir schätzen  $\tilde{\omega}$  nun nach oben durch  $\omega$  ab. Da sowohl für die Nullschritte wie auch für die Short-Serious-Schritte

$$-\beta_+^+ + \langle g^+, -t\zeta \rangle \geq m_2 v = m_2 (-t \|\zeta\|^2 - \sigma)$$

gilt, erhalten wir

$$\langle g^+, \zeta \rangle \leq -\frac{1}{t} \beta_+^+ + m_2 (\|\zeta\|^2 + \frac{1}{t} \sigma). \quad (5.54)$$

Aus (5.53) erhalten wir mit (5.54)

$$\begin{aligned} Q(\nu) &= \frac{1}{2} \|(1-\nu)\zeta + \nu g^+\|^2 + (1-\nu) \frac{1}{t^+} \sigma + \nu \frac{1}{t^+} \beta_+^+ + (1-\nu) \frac{1}{t^+} (\sigma^+ - \sigma) \\ &= \frac{1}{2} \nu^2 \|\zeta - g^+\|^2 + \nu (\langle \zeta, g^+ \rangle - \|\zeta\|^2) + \frac{1}{2} \|\zeta\|^2 \\ &\quad + (1-\nu) \frac{1}{t^+} \sigma + \nu \frac{1}{t^+} \beta_+^+ + (1-\nu) \frac{1}{t^+} (\sigma^+ - \sigma) \\ &= \frac{1}{2} \nu^2 \|\zeta - g^+\|^2 + \nu (\langle \zeta, g^+ \rangle - \|\zeta\|^2) + \omega - \frac{1}{t} \sigma + \\ &\quad + (1-\nu) \frac{1}{t^+} \sigma + \nu \frac{1}{t^+} \beta_+^+ + (1-\nu) \frac{1}{t^+} (\sigma^+ - \sigma) \\ &= \frac{1}{2} \nu^2 \|\zeta - g^+\|^2 + \nu (\langle \zeta, g^+ \rangle - \|\zeta\|^2) + \omega \\ &\quad + \left( \frac{1}{t^+} - \frac{1}{t} \right) \sigma + \nu \frac{1}{t^+} (\beta_+^+ - \sigma) + (1-\nu) \frac{1}{t^+} (\sigma^+ - \sigma) \\ &\leq \frac{1}{2} \nu^2 \|\zeta - g^+\|^2 + \nu \left( -\frac{1}{t} \beta_+^+ + m_2 (\|\zeta\|^2 + \frac{1}{t} \sigma) - \|\zeta\|^2 \right) + \omega \\ &\quad + \left( \frac{1}{t^+} - \frac{1}{t} \right) \sigma + \nu \frac{1}{t^+} (\beta_+^+ - \sigma) + (1-\nu) \frac{1}{t^+} (\sigma^+ - \sigma) \\ &= \frac{1}{2} \nu^2 \|\zeta - g^+\|^2 - \nu(1-m_2) \left( \frac{1}{t} \sigma + \|\zeta\|^2 \right) - \frac{1}{t} (\beta_+^+ - \sigma) \nu + \omega \\ &\quad + \left( \frac{1}{t^+} - \frac{1}{t} \right) \sigma + \nu \frac{1}{t^+} (\beta_+^+ - \sigma) + (1-\nu) \frac{1}{t^+} (\sigma^+ - \sigma) \\ &= \frac{1}{2} \nu^2 \|\zeta - g^+\|^2 - \nu(1-m_2) \omega + \omega + \left( \frac{1}{t^+} - \frac{1}{t} \right) (\beta_+^+ - \sigma) \nu \\ &\quad + \left( \frac{1}{t^+} - \frac{1}{t} \right) \sigma + (1-\nu) \frac{1}{t^+} (\sigma^+ - \sigma) := q(\nu). \end{aligned} \quad (5.55)$$

Wir wählen nun eine Konstante  $\kappa_0 \geq \max\{\|\zeta\|, \|g^+\|, \frac{1}{t} \sigma, 1\}$  so, dass aus (5.55) mit  $\|\zeta - g^+\|^2 \leq (\|\zeta\| + \|g^+\|)^2$

$$\begin{aligned} Q(\nu) &\leq q(\nu) \\ &\leq 2\nu^2 \kappa_0^2 - \nu(1-m_2) \omega + \omega + \left( \frac{1}{t^+} - \frac{1}{t} \right) (\beta_+^+ - \sigma) \nu + \left( \frac{1}{t^+} - \frac{1}{t} \right) \sigma \\ &\quad + (1-\nu) \frac{1}{t^+} (\sigma^+ - \sigma) =: \bar{q}(\nu) \end{aligned} \quad (5.56)$$

für alle  $\nu \in [0, 1]$  folgt. Wir setzen  $\bar{\nu} = \frac{(1-m_2)\omega}{4\kappa_0^2}$ . Dann ist mit (5.34)

$$0 \leq \bar{\nu} = \frac{(1-m_2)\omega}{4\kappa_0} \leq \frac{1}{4\kappa_0} \frac{1}{2} \|\zeta\|^2 + \frac{1}{4\kappa_0} \frac{1}{t} \sigma \leq \frac{1}{8} + \frac{1}{4} < 1$$

und wir erhalten

$$\begin{aligned}
\omega_+ \leq \tilde{\omega} &= \omega - (1 - m_2)^2 \frac{\omega^2}{8\kappa_0^2} \\
&+ (1 - m_2) \frac{\omega}{4\kappa_0^2} \left( \frac{1}{t^+} - \frac{1}{t} \right) (\beta_+^+ - \sigma) \\
&+ \left( \frac{1}{t^+} - \frac{1}{t} \right) \sigma \\
&+ \left( 1 - (1 - m_2) \frac{\omega}{4\kappa_0^2} \right) \frac{1}{t^+} (\sigma^+ - \sigma). \tag{5.57}
\end{aligned}$$

- (d) Ab jetzt betrachten wir wieder alle Iterationen mit  $k \geq K$ : Nach Beweisschritt (c) wählen wir im  $k$ -ten Iterationsschritt eine Konstante  $\kappa_0^k$ , so dass

$$\kappa_0^k \geq \max \left\{ \left\| \sum_{i \in J^k} \lambda_i^k g^i \right\|, \|g^{k+1}\|, \frac{1}{t^k} \sum_{i \in J^k} \lambda_i^k \beta_i^k, 1 \right\}$$

gilt. Nach Hilfssatz 5.20 existiert mit der Beschränktheit der  $t^k$  durch  $\bar{t} > 0$  eine Konstante  $\bar{\kappa} \geq \kappa_0^k \geq 1$  für alle  $k \geq K$ . Aus (5.57) folgt

$$\begin{aligned}
\omega^{k+1} &\leq \omega^k - (1 - m_2)^2 \frac{(\omega^k)^2}{8\bar{\kappa}^2} \\
&+ (1 - m_2) \frac{\omega^k}{4(\kappa_0^k)^2} \left( \frac{1}{t^{k+1}} - \frac{1}{t^k} \right) (\beta_{k+1}^{k+1} - \sigma^k) \\
&+ \left( \frac{1}{t^{k+1}} - \frac{1}{t^k} \right) \sigma^k \\
&+ \left( 1 - (1 - m_2) \frac{\omega^k}{4(\kappa_0^k)^2} \right) \frac{1}{t^{k+1}} (\sigma^{k+1} - \sigma^k). \tag{5.58}
\end{aligned}$$

- (e) Um den Rest  $\sigma^{k+1} - \sigma^k$  abzuschätzen, untersuchen wir für  $k \geq K$  die Gewichte

$$\begin{aligned}
\beta_i^k &= \max \left\{ f(\tilde{x}^k) - f(\tilde{w}^i) - \langle \tilde{g}^i, u^k - q^i \rangle, \right. \\
&\quad \left. c_0 \|q^i - u^k\|^2 + c_0 \left| \|C(\tilde{w}^i)\| - \|C(\tilde{x}^k)\| \right| \right\}
\end{aligned}$$

und

$$\begin{aligned}
\beta_i^{k+1} &= \max \left\{ f(\tilde{x}^{k+1}) - f(\tilde{w}^i) - \langle \tilde{g}^i, u^{k+1} - q^i \rangle, \right. \\
&\quad \left. c_0 \|q^i - u^{k+1}\|^2 + c_0 \left| \|C(\tilde{w}^i)\| - \|C(\tilde{x}^{k+1})\| \right| \right\}.
\end{aligned}$$

Im Falle eines Nullschrittes gilt  $\beta_i^k = \beta_i^{k+1}$  und somit folgt wie bereits oben bemerkt  $\sigma^{k+1} = \sigma^k$ .

Deshalb werden im Folgenden nur noch Short-Serious-Schritte betrachtet. Wir erinnern zunächst, dass für  $a, b, c, d \in \mathbb{R}$

$$\max\{a, b\} - \max\{c, d\} \leq |a - c| + |b - d|$$

gilt.

Somit folgt für  $i \in J^k$

$$\begin{aligned}
\beta_i^k - \beta_i^{k+1} &\leq |f(\tilde{x}^k) - f(\tilde{w}^i) - \langle \tilde{g}^i, u^k - q^i \rangle \\
&\quad - (f(\tilde{x}^{k+1}) - f(\tilde{w}^i) - \langle \tilde{g}_i, u^{k+1} - q^i \rangle)| \\
&\quad + c_0 \left| \|q^i - u^k\|^2 - \|q^i - u^{k+1}\|^2 \right| \\
&\quad + \left| \|C(\tilde{w}^i)\| - \|C(\tilde{x}^k)\| \right| - \left| \|C(\tilde{w}^i)\| - \|C(\tilde{x}^{k+1})\| \right| \\
&\leq |f(\tilde{x}^k) - f(\tilde{x}^{k+1}) - \langle \tilde{g}_i, u^k - u^{k+1} \rangle| \\
&\quad + c_0 \left| \|u^k - u^{k+1}\|^2 - 2\langle q_i - u^{k+1}, u^k - u^{k+1} \rangle \right|,
\end{aligned}$$

weil bei Short-Serious-Schritten  $\|C(\tilde{x}^k)\| = \|C(\tilde{x}^{k+1})\|$  gilt.

Aufgrund der Beschränktheit von  $(u^k)_{k \in \mathbb{N}}$ ,  $(\tilde{x}^k)_{k \in \mathbb{N}}$ ,  $(\tilde{w}^k)_{k \in \mathbb{N}}$  und  $(\tilde{g}^k)_{k \in \mathbb{N}}$  existiert ein  $c \geq 0$ , so dass für alle  $i \in J^k$  unter Verwendung der Dreiecksungleichung und der Cauchy-Schwarz-Ungleichung

$$\begin{aligned}
\beta_i^k - \beta_i^{k+1} &\leq |f(\tilde{x}^k) - f(\tilde{x}^{k+1})| + c\|u^k - u^{k+1}\| \\
&\quad + c_0\|u^k - u^{k+1}\|^2 + 2c\|u^k - u^{k+1}\|
\end{aligned}$$

gilt. Nach Multiplikation mit  $\lambda_i^k$  und Aufsummieren über  $i \in J^k$  erhalten wir

$$\begin{aligned}
\sum_{i \in J^k} \lambda_i^k (\beta_i^k - \beta_i^{k+1}) &\leq |f(\tilde{x}^k) - f(\tilde{x}^{k+1})| + c\|u^k - u^{k+1}\| \\
&\quad + c_0\|u^k - u^{k+1}\|^2 + 2c\|u^k - u^{k+1}\|.
\end{aligned}$$

Da nur noch Short-Serious- und Nullschritte ausgeführt werden, folgt mit Hilfssatz 5.25 und der Stetigkeit von  $f$

$$\lim_{k \rightarrow \infty} \sigma^{k+1} - \sigma^k = 0.$$

Die Konvergenz von  $f(\tilde{x}^k) - f(\tilde{x}^{k+1}) \rightarrow 0$  gilt, da falls von  $k \rightarrow k+1$  ein Nullschritt durchgeführt wurde, Gleichheit und bei einem Short-Serious-Schritt bei exakter Rechnung nach Hilfssatz 5.25

$$\tilde{x}^k - \tilde{x}^{k+1} = (\tilde{y}^k(u^k), u^k) - (\tilde{y}^{k+1}(u^{k+1}), u^{k+1}) \rightarrow 0 \text{ für } k \rightarrow \infty$$

gilt.

- (f) Sei nun  $\bar{\omega}$  der größte Häufungspunkt der beschränkten Folge  $(\omega^k)_{k \in \mathbb{N}}$ . Dann ist  $\bar{\omega} \geq 0$ . Mit (5.58) und  $\lim_{k \rightarrow \infty} t^k = \bar{t} > 0$  existiert nach (c)-(e) ein Häufungspunkt  $\tilde{\omega} \geq 0$  von  $(\omega^k)_{k \in \mathbb{N}}$  mit

$$\bar{\omega} \leq \tilde{\omega} - (1 - m_2)^2 \frac{\tilde{\omega}^2}{8\bar{\kappa}_0^2}. \tag{5.59}$$

Da  $\bar{\omega}$  der größte Häufungspunkt ist, gilt

$$0 \leq -(1 - m_2)^2 \frac{\tilde{\omega}^2}{8\bar{\kappa}_0^2}. \quad (5.60)$$

Mit  $\omega^k \geq 0$ ,  $\bar{\kappa}_0 > 0$  und  $0 < m_2 < 1$  folgt, dass  $\tilde{\omega} = 0$  der einzige Häufungspunkt von  $(\omega^k)_{k \in \mathbb{N}}$  ist. Daraus folgt die Konvergenz der Folge  $\omega^k$  und damit die Konvergenz von  $\sigma^k$  und  $\zeta^k$ .

■

**Satz 5.27.** *Die Folge der Iterationspunkte sei beschränkt. Für die Folge  $(k(l))_{l \in \mathbb{N}}$  gelte*

$$\lim_{l \rightarrow \infty} \zeta^{k(l)} = 0 \text{ und } \lim_{l \rightarrow \infty} \sigma^{k(l)} = 0. \quad (5.61)$$

*Dann ist jeder Häufungspunkt der Folge  $(\tilde{x}^{k(l)})$  zulässig und stationär.*

*Beweis:* Sei  $\bar{x}$  Häufungspunkt von  $(\tilde{x}^{k(l)})_{l \in \mathbb{N}}$ . Dann existiert eine gegen  $\bar{x} =: (\bar{y}, \bar{u})$  konvergente Teilfolge, die wir ebenfalls mit  $(\tilde{x}^{k(l)})_{l \in \mathbb{N}}$  bezeichnen.

- (a) Wir zeigen zunächst die Zulässigkeit von  $\bar{x}$ . Hierzu unterscheiden wir, ob es ein  $L \in \mathbb{N}$  gibt, so dass für  $l > L$  nur noch Korrekturschritte durchgeführt werden.

Existiert so ein  $L$  nicht, dann gilt für alle Iterationen  $k(l)$ , in denen kein Korrekturschritt stattgefunden hat,

$$\|C(\tilde{x}^{k(l)})\| \leq \frac{1}{\kappa} |v^{k(l)}| \leq \frac{1}{\kappa} \left( t^{k(l)} \|\zeta^{k(l)}\|^2 + \sigma^{k(l)} \right),$$

nach (1)(b) aus Algorithmus 5.15 und  $C(\tilde{x}^{k(l)}) \rightarrow 0$  für  $l \rightarrow \infty$ , da alle Teilfolgen einer konvergenten Folge gegen den Grenzwert der Folge konvergieren.

Nun betrachten wir den Fall, dass für  $l > L$  nur noch Korrekturschritte durchgeführt werden. Dann gilt nach der letzten Zeile aus Algorithmus 5.11

$$\|C(\tilde{x}^{k(l)+1})\| \leq \frac{1}{\kappa} \left( t^{k(l)} \|\zeta^{k(l)}\|^2 + \sigma^{k(l)} \right). \quad (5.62)$$

Somit bilden die  $(\|C(\tilde{x}^{k(l)+1})\|)_{l \in \mathbb{N}}$  eine Nullfolge. Wir interessieren uns aber für die  $(\|C(\tilde{x}^{k(l)})\|)_{l \in \mathbb{N}}$ .

Wir versuchen  $\|C(\tilde{x}^{k(l)})\|$  abzuschätzen. Nach Algorithmus 5.11 gilt entweder  $\|C(\tilde{x}^{k(l)+1})\| \leq R^{k(l)+1} \leq \|C(\tilde{x}^{k(l)+1})\|$  oder es existiert ein Folgenglied  $a_{\zeta^{k^*}}$  der Folge  $(a_j)_{j \in \mathbb{N}}$  aus Algorithmus 5.11 mit  $k(l) + 1 \leq k^* \leq k(l + 1)$  und  $\|C(\tilde{x}^{k(l)+1})\| \leq a_{\zeta^{k^*}}$ .

Tritt der zweite Fall unendlich oft auf, so existiert eine Teilfolge von  $\|C(\tilde{x}^{k(l)})\|$ , die wie  $(a_j)$  gegen Null konvergiert. Somit konvergiert mit den gleichen Argumenten wie oben die Folge gegen Null.

Tritt hingegen der zweite Fall nur endlich oft auch, so tritt der erste Fall unendlich oft auf und die Behauptung folgt aus (5.62).

Zusammengefasst gilt also für jede Teilfolge  $(\tilde{x}^{k(l)})_{l \in \mathbb{N}}$  von  $(\tilde{x}^k)_{k \in \mathbb{N}}$ , die gegen einen Häufungspunkt  $\bar{x}$  konvergiert und für die (5.61) gilt, dass

$$\lim_{l \rightarrow \infty} \|C(\tilde{x}^{k(l)})\| = \|C(\bar{x})\| = 0.$$

Außerdem folgt, dass  $\bar{y} = y(\bar{u})$  die Lösung des Kontaktproblems zu den Formparametern  $\bar{u}$  ist.

- (b) Jetzt zeigen wir die Stationarität. Dazu setzen wir  $\bar{\zeta}^k := (\sigma^k, \zeta^k)^T \in \mathbb{R}^{n+1}$  für alle  $k \in \mathbb{N}$ . Wegen (5.61) gilt

$$\lim_{l \rightarrow \infty} \bar{\zeta}^{k(l)} = 0_{\mathbb{R}^{n+1}}. \quad (5.63)$$

Nach dem Satz von Caratheodory lässt sich  $\bar{\zeta}^k$  als Konvexkombination von höchstens  $n+2$  Elementen der Bauart  $(\beta_i^k, \tilde{g}^i)^T$  mit  $i \in \mathbb{N}$  darstellen als

$$\bar{\zeta}^k = \sum_{i=1}^{n+2} \mu_i^k (\beta(\tilde{x}^k, \tilde{w}^{\phi_k(i)}) (\tilde{g}^{\phi_k(i)})^T)$$

beziehungsweise

$$\bar{\zeta}^{k(l)} = \sum_{i=1}^{n+2} \mu_i^{k(l)} (\beta(\tilde{x}^{k(l)}, \tilde{w}^{\phi_{k(l)}(i)}) (\tilde{g}^{\phi_{k(l)}(i)})^T. \quad (5.64)$$

Dabei ist  $\phi_k$  eine Auswahl von  $n+2$  Elementen aus  $J^k$ . Nachdem die  $(\mu_i^{k(l)})_{l \in \mathbb{N}}$  für jedes  $i$  zwischen 0 und 1 liegen, existiert für  $1 \leq i \leq n+2$  eine konvergente Teilfolge  $(\mu_i^{k(l(m))})_{m \in \mathbb{N}}$  von  $(\mu_i^{k(l)})_{l \in \mathbb{N}}$  mit

$$\lim_{m \rightarrow \infty} \mu_i^{k(l(m))} = \mu_i, \quad \mu_i \geq 0 \text{ für } 1 \leq i \leq n+2.$$

Außerdem gilt

$$\sum_{i=1}^{n+2} \mu_i = 1.$$

Wir fassen nun die Indices der  $\mu_i$  zu einer Menge  $I$  nach dem Prinzip zusammen, dass wenn  $\mu_i > 0$  ist, dann wird der Index  $i$  der Menge  $I$  zugeführt.

Aus der ersten Komponente von (5.63) folgt mit (5.64)

$$\lim_{m \rightarrow \infty} \beta(\tilde{x}^{k(l(m))}, \tilde{w}^{\phi_{k(l(m))}(i)}) = 0 \quad \text{für } i \in I. \quad (5.65)$$

Aus der Definition der  $\beta(\tilde{x}^k, \tilde{w}^i)$  folgt mit (5.65)

$$\lim_{m \rightarrow \infty} \left| \|C(\tilde{x}^{k(l(m))})\| - \|C(\tilde{w}^{\phi_{k(l(m))}(i)})\| \right| = 0, \quad i \in I$$

und

$$\lim_{m \rightarrow \infty} \|u^{k(l(m))} - q^{\phi_{k(l(m))}(i)}\| = 0, \quad i \in I.$$

Aus der Konvergenz der Teilfolge  $(\tilde{x}^{k(l)})$  gegen  $\bar{x} = (\bar{y}, \bar{u})$  folgt für  $i \in I$  die Konvergenz von  $q^{\phi_{k(l(m))}(i)}$  gegen  $\bar{u}$ . Aus der Zulässigkeit von  $\bar{x}$  nach (a) folgt der Grenzwert  $\|C(\tilde{w}^{\phi_{k(l(m))}(i)})\| \rightarrow 0$ . Somit konvergiert  $\tilde{w}^{\phi_{k(l(m))}(i)}$  ebenfalls gegen  $\bar{x}$  für  $i \in I$ .

Wir zeigen nun, dass die Häufungspunkte von  $\tilde{g}^{\phi_{k(l(m))}(i)}$  in  $\partial_u f(\bar{x})$  liegen. Die Existenz von Häufungspunkten folgt aufgrund der Beschränktheit von  $(\tilde{g}^k)$  in  $\mathbb{R}^n$  und dem Satz von Bolzano Weierstraß.

Zu  $\tilde{g}^k$  existiert nach Annahme 5.10 und (5.20) ein  $g^k \in \partial_u f(y(q^k), q^k)$  mit

$$\|\tilde{g}^k - g^k\| \leq \kappa_g \|C(\tilde{w}^k)\|. \quad (5.66)$$

Wir wählen nun eine Teilfolge  $k(l(m(j)))$  der Folge  $k(l(m))$  so, dass für jedes  $i \in I$  die Folge  $\tilde{g}^{\phi_{k(l(m(j))}(i)}$  gegen  $\bar{g}^i$  konvergiert. Dann existieren Subgradienten  $\tilde{g}^{\phi_{k(l(m(j))}(i)}$ , so dass wegen

$$\begin{aligned} \tilde{g}^{\phi_{k(l(m(j))}(i)} &= \tilde{g}^{\phi_{k(l(m(j))}(i)} + (-g^{\phi_{k(l(m(j))}(i)} + g^{\phi_{k(l(m(j))}(i)}) \\ &= (\tilde{g}^{\phi_{k(l(m(j))}(i)} - g^{\phi_{k(l(m(j))}(i)}) + g^{\phi_{k(l(m(j))}(i)} \end{aligned}$$

$g^{\phi_{k(l(m(j))}(i)}$  ebenfalls gegen  $\bar{g}^i$  konvergiert. Da der Subgradient oberhalb stetig ist, folgt  $\bar{g}^i \in \partial_u f(\bar{x})$  und wir schreiben  $g^i$  anstatt  $\bar{g}^i$ .

Dann ergibt sich

$$\begin{aligned} \sum_{i \in I} \mu_i g^i &= \sum_{i \in I} \lim_{j \rightarrow \infty} \mu_i^{k(l(m(j)))} \tilde{g}^{\phi_{k(l(m(j))}(i)} \\ &= \sum_{i=1}^{n+2} \lim_{j \rightarrow \infty} \mu_i^{k(l(m(j)))} \tilde{g}^{\phi_{k(l(m(j))}(i)} \\ &= \lim_{j \rightarrow \infty} \sum_{i=1}^{n+2} \mu_i^{k(l(m(j)))} \tilde{g}^{\phi_{k(l(m(j))}(i)} \\ &= \lim_{j \rightarrow \infty} \sum_{i=1}^{n+2} \mu_i^{k(l(m(j)))} \tilde{g}^{\phi_{k(l(m(j))}(i)} \\ &= 0. \end{aligned} \quad (5.67)$$

Die letzte Gleichheit gilt nach (5.63).

Somit folgt  $0 \in \partial_u f(\bar{x})$ , welches einen stationären Punkt impliziert.

■

## 5.6 Erweiterungen

### 5.6.1 Reset des Bundles

Das Bundle-Verfahren kann so erweitert werden, dass das Bundle regelmäßig resetet wird. Dies hat den Vorteil, dass das quadratische Optimierungsproblem bei einem kleineren Bundle weniger aufwendig zu lösen ist und dass der Speicher der Zwischenwerte nicht zu groß wird.

Wir lehnen uns an die Resetstrategien an, wie sie in [Sch89], [Hin01] verwendet werden, welche auf [Kiw85] zurückgehen.

Dabei behalten wir die innere Iteration nach Algorithmus 5.15 bei, führen aber folgende neue äußere Iteration ein:

**Algorithmus 5.28** (Äußere Iteration mit Reset). *Gegeben seien Startpunkt  $u^1 \in \mathbb{R}^n$ , eine maximale Bundlegröße  $j_{max} \geq 3$  und Parameter  $0 < m_1 < m_2 < 1$ ,  $0 < m_3 < 1$ ,  $\kappa > 0$ ,  $\epsilon \geq 0$ .*

- (a) *Berechne  $\tilde{y}^1(u^1)$  und  $\tilde{g}^1$  zu vorgegebenen Toleranzen. Setze  $\tilde{x}^1 := (\tilde{y}^1(u^1), u^1)$ ,  $q^1 := u^1$ ,  $\tilde{r}^1 = \tilde{y}^1$ ,  $\tilde{w}^1 := \tilde{x}^1$ ,  $\zeta^1 := g$ ,  $\sigma^1 := 1$ ,  $J^1 := \{1\}$ . Setze  $k := 1$ .*
- (b) *Führe eine innere Iteration durch.*
- (c) *Endet die innere Iteration mit  $\epsilon$ -stationärer Lösung, stopp: stationärer Punkt bei  $\tilde{x}^k$ .*
- (d) *Falls  $|J^k| = j_{max}$  gehe zu (e), sonst setze  $J = J^k$  und gehe zu (f).*
- (e) *Reset: Wähle  $J \subset J^k$  mit  $|J| \leq j_{max} - 2$  und  $\max\{i | i \in J^k, \beta_i^k = 0\} \in J$ . Wähle einen Index  $\tilde{k}$  und setze*

$$\tilde{g}_{\tilde{k}} = \zeta^k, \beta_{\tilde{k}}^k = \sigma^k, J = J \cup \{\tilde{k}\}.$$

- (f) *Endet die innere Iteration mit Korrekturschritt. Setze  $J^{k+1} := J \cup \{k+1\}$ ,  $\tilde{x}^{k+1} := (\tilde{y}^{k+1}, u^{k+1}) := (\tilde{w}^{k+1})$ ,  $k := k+1$ ; aktualisiere Strafterme nach (5.23). Gehe zu (b).*
- (g) *Endet die innere Iteration mit einem Abstiegs- oder Short-Serious-Schritt, setze  $J^{k+1} := J \cup \{k+1\}$ ,  $\tilde{x}^{k+1} := (\tilde{y}^{k+1}, u^{k+1}) := (\tilde{w}^{k+1})$ ,  $k := k+1$ ; aktualisiere Strafterme nach (5.23). Gehe zu (b).*
- (h) *Endet die innere Iteration mit einem Nullschritt, setze  $J^{k+1} := J \cup \{k+1\}$ ,  $\tilde{x}^{k+1} := (\tilde{y}^{k+1}, u^{k+1}) := (\tilde{x}^k)$ ,  $k := k+1$  und aktualisiere die Strafterme nach (5.23). Gehe zu (b).*

Wir können zeigen:

**Satz 5.29.** *Unter den gleichen Voraussetzungen, wie beim nichtexakten Bundle-Verfahren ohne Reset gilt, falls es eine Teilfolge  $(k(l))_{l \in \mathbb{N}}$  mit der Eigenschaft*

$$\lim_{l \rightarrow \infty} \zeta^{k(l)} = 0 \text{ und } \lim_{l \rightarrow \infty} \sigma^{k(l)} = 0 \quad (5.68)$$

*gibt, dass jeder Häufungspunkt der Folge  $(\tilde{x}^{k(l)})$  zulässig und stationär ist.*

*Beweis:* Der Beweis läuft analog zu oben. Wir müssen nur untersuchen, was in der Argumentation im Falle eines Resets passiert. Durch die Wahl der  $\beta_k^k = 0$  wird erreicht, dass (5.41) erfüllt bleibt, und durch die Speicherung des aggregierten Subgradienten  $\tilde{g}_k$  wird im Beweis von Hilfssatz 5.26 in Teil (c) sichergestellt, dass die Abschätzung in (5.53) gültig bleibt, denn der zuletzt berechnete Subgradient  $g^+$  ist entweder noch aus den Elementen des Bundles als Konvexkombination darstellbar, da kein Reset stattgefunden hat, oder dieser Wert wurde durch Aufnahme von  $\tilde{g}_k$  explizit ins Bundle aufgenommen. An weiteren Stellen geht der Reset nicht in den Beweis ein. ■

## 5.6.2 Inexaktes Bundle-Trust-Region-Verfahren und lineare Nebenbedingungen

Wir betrachten nun das Optimierungsproblem

$$\min_{y \in Y, u \in U} f(y, u) \quad \text{unter den Nebenbedingungen } C(y, u) = 0 \text{ und } Au \leq b, \quad (5.69)$$

mit  $A \in \mathbb{R}^{m \times n}$  und  $b \in \mathbb{R}^m$ , wobei die Ungleichung komponentenweise zu verstehen ist und  $A$  vollen Spaltenrang hat.

Für einen stationären Punkt gilt dann:

**Satz 5.30.** *Sei  $\bar{u}$  zulässig und stationär für (5.69), und für jedes  $u$  existiert ein eindeutiges  $y(u) \in Y$  mit  $C((y(u), u))$ , dann existiert ein  $w \in \mathbb{R}^m$  mit*

$$\begin{aligned} w_i &\geq 0, & 1 \leq i \leq m, \\ 0 &\in \partial f(y(\bar{u}), \bar{u}) + A^T w, \\ w^T (A\bar{u} - b) &= 0. \end{aligned} \quad (5.70)$$

Im Optimierungsalgorithmus nehmen wir wie in [Sch89] für konvexe Optimierungsprobleme die Nebenbedingung in der Inneren Iteration mit auf, indem wir statt (5.26) die Lösung von

$$(v(t), s(t)) := \operatorname{argmin} \left\{ v + \frac{1}{2t} \|s\|^2 \cdot \begin{array}{l} v \geq (\tilde{g}^i)^T s - \beta_i^k, i \in J^k \\ As \leq b - Au^k \end{array} \right\} \quad (5.71)$$

bestimmen.

Die optimale Lösung von (5.71) ist charakterisiert durch

**Satz 5.31.** Für  $n \in \mathbb{N}$  bezeichne

$$\Lambda(n) := \left\{ \lambda \in \mathbb{R}^n : \lambda_i \geq 0, 1 \leq i \leq n, \text{ und } \sum_{i=1}^n \lambda_i = 1 \right\}.$$

Im Minimum  $(v(t), s(t))$  von (5.71) existieren  $\lambda(t) \in \Lambda(|J^k|)$  und  $\mu \in \mathbb{R}^m$ , mit  $\mu_i \geq 0$  für  $1 \leq i \leq m$ , so dass

$$\begin{aligned} \lambda_i(t)(-v(t) + (g^i)^T s(t) - \beta_i^k) &= 0 \quad i \in J^k, \\ \mu_i(As(t) + Au^k - b)_i &= 0 \quad 1 \leq i \leq m, \\ s(t) &= -t \left( \sum_{i \in J^k} \lambda^i(t) g^i + A^T \mu \right), \\ v(t) &= -t \left\langle \sum_{i \in J^k} \lambda_i^k g^i, \sum_{i \in J^k} \lambda_i^k g^i + A^T \mu \right\rangle_{\mathbb{R}^n} - \sum_{i \in J^k} \lambda_i(t) \beta_i^k \end{aligned}$$

gelten.

Wir lösen das restringierte nichtkonvexe Optimierungsproblem mit einem inexakten Bundle-Verfahren, indem wir das Bundle-Verfahren aus Kapitel 5.4 verwenden, wobei wir (5.71) anstatt (5.26) als Trust-Region-Problem verwenden und die Optimalitätsbedingungen in Algorithmus 5.4 durch

$$-v(t) \leq \epsilon$$

mit  $v(t)$  aus Satz 5.31 ersetzen. Denn aus  $v(t) \rightarrow 0$  folgt  $(1/t_j) \|s_j\| \rightarrow 0$  und  $-(1/t_j) \|s_j\|^2 - v_j \rightarrow 0$ .

Die Konvergenz lässt sich vergleichbar zu oben zeigen.



# Kapitel 6

## Eine medizinische Anwendung und numerische Tests

### 6.1 Eine medizinische Anwendung

Wir betrachten die Behandlung von Frakturen am menschlichen Unterkiefer durch Osteosynthesen, die dazu dienen, die Fragmente in optimaler Position zu fixieren. Dabei ist das Konzept der Repositionierung der Frakturflächen von hoher Bedeutung, denn die Gebirgsstrukturen des Bruches können ineinander greifen und so je nach Anordnung zum Funktionieren des Verbundes aus Knochen und Osteosyntheseimplantat beitragen. Dabei soll die Funktion des Verbundes nach Möglichkeit nicht oder wenig beeinträchtigt werden, Blutgefäße und Nerven dürfen beispielsweise nicht durch Schrauben und Löcher beschädigt werden [Hor97, Sch03].

In Studien zeigen sich SMA-Klammern<sup>1</sup>, wie sie in Abbildung 6.1 zu sehen sind, ähnlich stabil wie gängige Titanosteosynthesysteme. Die Klammern (Staples) bestehen aus einem Stück: sie verfügen über zwei Haken, welche durch Bohrungen im Knochen greifen. Zur Implantation wird die Klammer im unterkühlten, das bedeutet martensitischen Zustand aufgebogen, damit die Haken in eine parallele Lage gebracht werden können, die es ihnen ermöglicht, in die Bohrungen eingeführt zu werden. Durch den Kontakt mit dem Knochen kommt es zum Wärmeaustausch, der die Klammer unmittelbar nach Implantation wieder durch Erwärmung in den Anfangszustand, den austenitischen Zustand bringt. Dies führt zu einem Halt der Klammern und zum anderen zur Kompression der Bruchfragmente.

Entscheidend für die Stabilität der Frakturversorgung ist in jedem Fall die Art der Anbringung der Staples. Bei den bis jetzt gemachten Vorversuchen wurden drei Staples nahezu parallel angebracht [Abe05]. Hier konnte eine ähnlich gute Stabilität wie mit zwei Titanplatten erreicht werden. Eine optimierte Anbringungsweise kann diese Stabilität weiter erhöhen oder sogar eine Anwendung von zwei Staples ermöglichen.

---

<sup>1</sup>SMA: Shape Memory Alloy

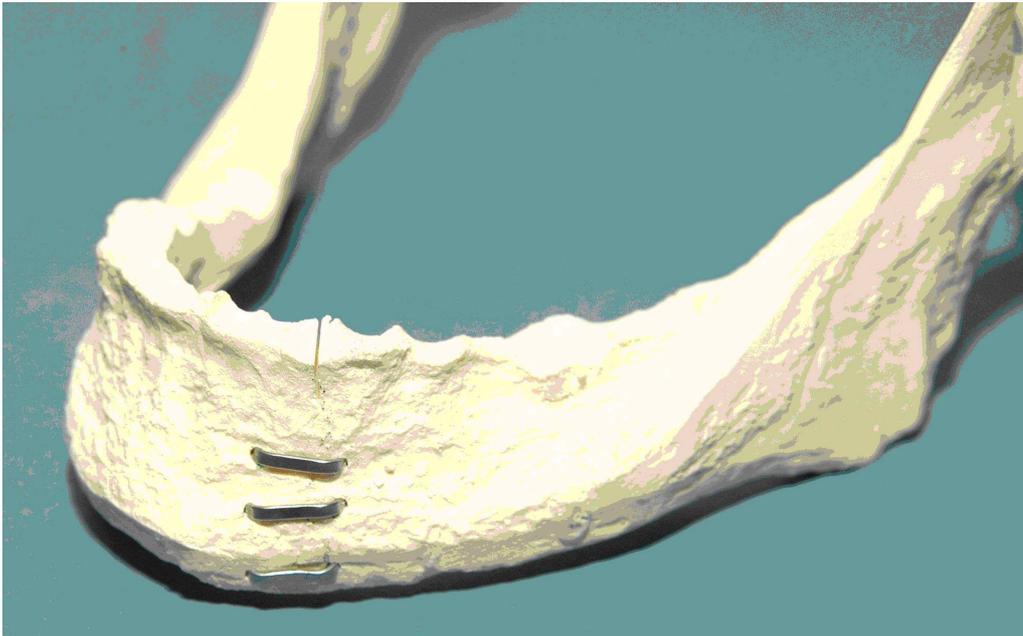


Abbildung 6.1: Drei Staples in situ eines Polyurethan-Kunststoffkiefers (sogenannter Sawbone).

Wir erläutern im weiteren Verlauf dieses Kapitels, wie die in dieser Arbeit vorgestellten Verfahren zur Bestimmung einer optimalen Anordnung der Klammern verwendet werden können. Hierzu testen wir die Methoden zunächst an einem Modellproblem, dann an einer tatsächlichen Geometrie. Wir möchten darauf hinweisen, dass alle Rechnungen nur zu Testzwecken durchgeführt wurden. Die verwendeten Daten stimmen nicht mit Experimenten überein, sondern wurden willkürlich gewählt. Die Verfahren sind allerdings mit kleineren Modifikationen geeignet, die medizinischen und biomechanischen Untersuchungen zu unterstützen.

## 6.2 Ein Modellproblem

Wir beginnen zunächst mit einem Modellproblem. Wir betrachten zwei Quader, die sich an der Stirnseite berühren und mittels zweier Klammern verbunden sind. Hierzu beachte auch Abbildung 6.2. Wir nehmen an, dass der Verbund zwischen den Klammern und den Quader starr ist und dass zwischen beiden Quadern eine Kontaktbedingung gilt. Die Klammern sind so hoch angebracht, dass sie die Quader an der Oberseite nicht berühren. Wir betrachten eine Kraft, welche auf die Unterseite der Quader-Klammer-Konfiguration wirkt und fixieren die Konfiguration auf beiden Stirnseiten.

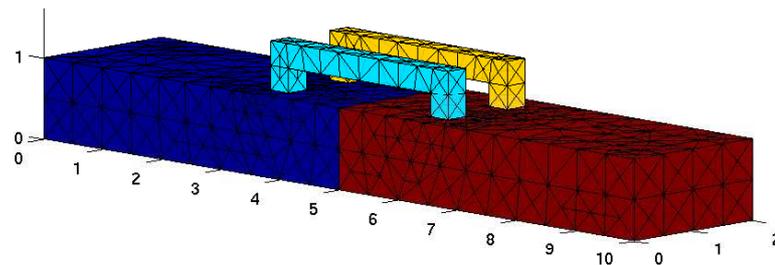


Abbildung 6.2: Das Modellproblem bestehend aus zwei Körpern, die durch Klammern verbunden sind.

### 6.2.1 Das Kontaktproblem

Wir lösen das Kontaktproblem auf dem Modellgebiet. Die Rechnungen werden unter Matlab durchgeführt, wobei einige rechenintensive Funktionen wie zum Beispiel das Aufstellen der Steifigkeitsmatrizen als MEX-Files ausgelagert wurden. Die Diskretisierung, die unter Comsol berechnet wird, ist so gewählt, dass die potentiellen Kontaktpunkte paarweise – jeweils einmal auf jeder Kontaktfläche – auftreten.

Weil wir in dieser Arbeit nur den Algorithmus, nicht aber die medizinische Anwendung diskutieren, nehmen wir an, dass die Klammern und die Quader aus demselben Material bestehen.

Wir untersuchen nun, wie sich die Lösung in Abhängigkeit vom Gitter und dem Regularisierungsparameter  $\gamma$  verhält. Die Tabellen 6.1 – 6.7 zeigen die Residuen der einzelnen Newtonschritte bei der Lösung des Kontaktproblems zu den Regularisierungsparametern  $\gamma = 1$  bis  $\gamma = 10^{-6}$  für Gitter mit 600 bis 40000 Knoten. In den Tabellen steht die Anzahl der Knoten immer in der linken Spalte. Dabei wurde in jedem Newton-Schritt das lineare System exakt gelöst. Die Rechnungen wurden mit den Matlab eigenen Lösern durchgeführt.

Wir verwenden die willkürlich gewählten Materialparameter  $E = 3.4 \cdot 10^5$ ,  $\mu = 0.3$  sowohl für die Klammern als auch für die Quader. Die äußere Kraft  $F = (0, 0, 100)^T$  wirkt auf die Unterseite. An den freien Stirnseiten der Quader wirkt jeweils eine

Dirichlet-Bedingung. Wir verwenden als Norm

$$\|y\| = \sqrt{\|y_1\|_{L^2(\Omega)}^2 + \|y_2\|_{L^2(\Gamma_C)}^2 + \|y_3\|_{L^2(\Gamma_C)}^2}$$

für  $y \in H^{-1}(\Omega) \times L^2(\Gamma_C) \times L^2(\Gamma_C)$ . Allerdings berechnen wir ab dem zweiten Schritt nur noch die Norm in der letzten Komponente, da bei exakter Rechnung der Beitrag der ersten zwei Komponenten verschwindet. Der Rechenfehler, der dabei wirklich auftritt, wird vernachlässigt. Wir starten immer mit dem Nullvektor. Wir wählen  $10^{-8}$  als Abbruchkriterium für den Newton-Löser.

#	$\ \cdot\ $	$\ \cdot\ $	$\ \cdot\ $	$\ \cdot\ $
659	1.4E3	6.1E-3	1.3E-7	0.0E-8
1717	1.7E3	6.0E-3	1.4E-7	0.0E-8
4784	2.1E3	4.3E-3	1.0E-7	0.0E-8
13516	2.5E3	3.4E-3	8.5E-8	0.0E-8
39705	3.0E3	2.4E-3	6.1E-8	0.0E-8

Tabelle 6.1: Die Residuen des Newton-Verfahrens zum Regularisierungsparameter  $\gamma = 1$ .

#	$\ \cdot\ $	$\ \cdot\ $	$\ \cdot\ $	$\ \cdot\ $
659	1.4E+3	6.0E-3	1.3E-6	0.0E-8
1717	1.7E+3	6.0E-3	1.3E-6	0.0E-8
4784	2.1E+3	4.3E-3	1.0E-6	0.0E-8
13516	2.5E+3	3.4E-3	8.5E-7	0.0E-8
39705	3.0E+3	2.4E-3	6.1E-7	0.0E-8

Tabelle 6.2: Die Residuen des Newton-Verfahrens zum Regularisierungsparameter  $\gamma = 1E - 1$ .

#	$\ \cdot\ $	$\ \cdot\ $	$\ \cdot\ $	$\ \cdot\ $
659	1.4E+3	2.2E-2	4.8E-5	0.0E-8
1717	1.7E+3	2.6E-2	5.9E-5	0.0E-8
4784	2.1E+3	2.9E-2	6.9E-5	0.0E-8
13516	2.5E+3	3.0E-2	7.5E-5	0.0E-8
39705	3.0E+3	3.2E-2	8.0E-5	0.0E-8

Tabelle 6.3: Die Residuen des Newton-Verfahrens zum Regularisierungsparameter  $\gamma = 1E - 2$ .

#	$\ \cdot\ $	$\ \cdot\ $	$\ \cdot\ $	$\ \cdot\ $
659	1.4E+3	2.1E-2	4.5E-4	0.0E-8
1717	1.7E+3	2.6E-2	5.6E-4	0.0E-8
4784	2.1E+3	2.8E-2	6.6E-4	0.0E-8
13516	2.5E+3	3.0E-2	7.1E-4	0.0E-8
39705	3.0E+3	3.2E-2	7.6E-4	0.0E-8

Tabelle 6.4: Die Residuen des Newton-Verfahrens zum Regularisierungsparameter  $\gamma = 1E - 3$ .

#	$\ \cdot\ $					
659	1.4E+3	1.8E-2	3.2E-3	0.9E-4	0.0E-8	0.0E-8
1717	1.7E+3	2.1E-2	3.9E-3	1.4E-4	4.7E-6	0.0E-8
4784	2.1E+3	2.3E-2	4.4E-3	0.0E-8	0.0E-8	0.0E-8
13516	2.5E+3	2.4E-2	4.7E-3	0.0E-8	0.0E-8	0.0E-8
39705	3.0E+3	2.5E-2	5.9E-3	2.1E-4	0.0E-8	0.0E-8

Tabelle 6.5: Die Residuen des Newton-Verfahrens zum Regularisierungsparameter  $\gamma = 1E - 4$ .

#	$\ \cdot\ $						
659	1.4E3	1.3E-3	5.2E-3	1.1E-3	1.5E-4	0.0E-8	0.0E-8
1717	1.7E3	1.3E-3	5.6E-3	2.3E-3	0.0E-8	0.0E-8	0.0E-8
4784	2.1E3	1.3E-3	5.9E-3	2.3E-3	4.8E-4	0.0E-8	0.0E-8
13516	2.5E3	1.3E-3	6.1E-3	2.4E-3	4.6E-4	3.5E-5	0.0E-8
39705	3.0E3	1.3E-3	6.3E-3	2.4E-3	3.8E-4	2.7E-5	0.0E-8

Tabelle 6.6: Die Residuen des Newton-Verfahrens zum Regularisierungsparameter  $\gamma = 1E - 5$ .

#	$\ \cdot\ $								
659	1.4E3	1.3E-3	1.2E-3	1.7E-3	0.9E-3	2.7E-4	0.0E-8	0.0E-8	0.0E-8
1717	1.7E3	1.3E-3	1.2E-3	1.7E-3	0.9E-3	6.6E-4	0.0E-8	0.0E-8	0.0E-8
4784	2.1E3	1.3E-3	1.2E-3	1.7E-3	1.2E-3	5.2E-4	0.3E-4	0.0E-8	0.0E-8
13516	2.5E3	1.3E-3	1.2E-3	1.7E-3	1.2E-3	7.1E-4	2.8E-4	0.0E-8	0.0E-8
39705	3.0E3	1.3E-3	1.2E-3	1.7E-3	1.2E-3	6.7E-4	2.4E-4	3.9E-5	0.0E-8

Tabelle 6.7: Die Residuen des Newton-Verfahrens zum Regularisierungsparameter  $\gamma = 1E - 6$ .

Es fällt auf, dass die Iterationen zu jedem festen Regularisierungsparameter  $\gamma$  unabhängig von der Gittergröße in etwa immer dieselbe Anzahl an Schritten brauchen. Diese Eigenschaft nennt man gitterunabhängig. Die Gitterunabhängigkeit des semiglatten Newton-Verfahrens wurde in [HU04] gezeigt.

Für kleinere Regularisierungsparameter als  $\gamma = 10^{-6}$  liegt der Nullvektor nicht mehr im Konvergenzgebiet der Lösung beziehungsweise es treten numerische Instabilitäten auf, so dass die Gitterunabhängigkeit nicht weiter untersucht werden kann. Allerdings eignen sich die Lösungen zu einem größeren Regularisierungsparameter als Startwert zu einem kleineren. Somit kann man auch Lösungen zu kleinen Regularisierungsparametern berechnen. Bei größeren Gittern konnten wir keinen direkten Löser für die Systemmatrix verwenden. Deshalb vergleichen wir die Werte an dieser Stelle nicht.

## 6.2.2 Formoptimierung

In diesem Abschnitt bestimmen wir die optimale Position der Klammern. Wir führen als Variablen die  $(x, y)$ -Koordinaten des Schwerpunkts jeder Klammer und den Schnittwinkel der Klammer mit der  $y$ - $z$ -Ebene ein. Dies gibt uns im Fall von zwei Klammern sechs Formparameter<sup>2</sup>. In unserem Beispiel suchen wir die Parameter, für die die Gesamtauslenkung

$$J(\Omega) := \int_{\Omega} \|u(\Omega)\|_2^2 dx$$

des Quader-Klammer-Komplexes unter der Kraft  $F$  minimal ist.

Zur Erzeugung eines Gebietes berechnen wir, wie es in 4.4 beschrieben ist, die neuen Koordinaten und ihre Abhängigkeit von den Formvariablen durch Lösen eines Elastizitätsproblems. Wir stellen das Elastizitätsproblem auf einem groben Gitter auf und lösen es mit einem direkten Verfahren exakt. Anschließend werden die Werte auf das feinere Gitter, auf dem die Formoptimierung durchgeführt wird, prolongiert.

Wir gehen im weiteren davon aus, dass alle Voraussetzungen, die wir in den vorangegangenen Kapiteln gestellt haben, für das semiglatte Newton-Verfahren und das Bundle-Verfahren erfüllt sind.

Wir wählen folgende Parameter für das Bundle-Verfahren:  $\gamma_1 = 0.3$ ,  $\gamma_2 = 0.7$ ,  $\gamma_3 = 0.5$ ,  $c_0 = 0.001$  und  $\kappa = 10$ .

Wir wählen als Nebenbedingung zum einen das regularisierte Kontaktproblem mit Regularisierungsparameter  $\gamma = 0.1$  und für die Formparameter  $p$  die Nebenbedingung

$$\begin{aligned} -0.25 &\leq p_1 \leq 0.25, & -0.05 &\leq p_2 \leq 0.05, \\ -0.01 &\leq p_3 \leq 0.01, & -0.25 &\leq p_4 \leq 0.25, \\ -0.05 &\leq p_5 \leq 0.05, & -0.01 &\leq p_6 \leq 0.01. \end{aligned}$$

<sup>2</sup>Die ersten drei Parameter beschreiben die Längs- und die Querverschiebung der ersten Klammer sowie ihren Drehwinkel, die letzten drei Parameter dasselbe für die zweite Klammer.

Wir wählen  $(a_j)_{j \in \mathbb{N}} = 2^{-j}/1000$  für die Folge zur Bestimmung des Relaxationsparameters.

Wir initialisieren den Trust-Region-Parameter mit  $t = 100$ . Falls im Laufe der Iteration zwei aufeinander folgende Abstiegschritte durchgeführt werden, so verdoppeln wir den Trust-Region-Parameter, solange er kleiner als 1000 ist. Im Falle von zwei aufeinander folgenden Nullschritten halbieren wir den Parameter.

Wir lösen die linearen Gleichungssysteme des Kontaktproblems (3.16) und der diskretisierten Form des adjungierten Problems (4.46) mit einem vorkonditionierten CG-Verfahren. Als Vorkonditionierer verwenden wir den V-Zyklus eines Mehrgitterverfahrens, wobei wir vier Vor- und vier Nachglättungsschritte verwenden. Dabei verwenden wir immer die Lösung des letzten Schrittes als Startwert in der nächsten Iteration.

Schritt	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$	$F(p)$
1	0.00	0.00	0.00	0.00	0.00	0.00	0.21326
2	-0.03	0.04	0.01	0.02	-0.04	-0.01	0.21318
3	0.09	0.05	0.01	0.08	-0.05	-0.01	0.21305
4	-0.16	0.05	0.01	0.14	-0.05	-0.01	0.21283
5	-0.22	0.05	0.01	0.19	-0.05	-0.01	0.21253
6	-0.25	0.05	0.01	0.25	-0.05	-0.01	0.21229
7	-0.25	0.05	0.01	0.25	-0.05	-0.01	0.21229
8	-0.25	0.05	0.01	0.25	-0.05	-0.01	0.21229

Tabelle 6.8: Die Iterierten des Formoptimierungsproblems bei inexakter Rechnung zum Startwert 0.

Schritt	It.	$t$	$\ C\ $	$ v $	# Kont.	# Adj.
1	-	-	1.7E-5	-	70	32
2	SS	100	2.0E-4	7.1E-5	53	5
3	SS	100	5.6E-5	7.8E-5	51	7
4	SS	200	5.8E-5	1.6E-4	54	7
5	SS	400	1.7E-5	2.5E-4	65	7
6	SS	800	2.2E-5	2.1E-4	66	16
7	KS	800	7.4E-9	0	66	72
8	opt	800	-	0	-	-

Tabelle 6.9: Simulationsprotokoll des Formoptimierungsproblems bei inexakter Rechnung zum Startwert 0.

Die Tabellen 6.8 und 6.9 zeigen für den Startwert  $p = (0, 0, 0, 0, 0, 0)$  und das Abbruchkriterium  $|v| \leq 10^{-8}$  den Verlauf der Simulation. Wir dokumentieren nur  $v$ . Wir rechnen auf dem Gitter mit 13516 Knoten und verwenden ein Mehrgitterverfahren über drei Gitterebenen.

In Tabelle 6.8 sieht man die Iterierten und die zugehörigen Funktionswerte. Tabelle 6.9 gibt ein Iterationsprotokoll an. In der zweiten Spalte steht, ob die innere Iteration mit einem Abstiegschritt (SS), einem Nullschritt (NS), einem Short-Serious-Schritt (SSS), einem Korrekturschritt (KS) oder mit dem Hinweis auf die optimale Lösung (opt) verlassen wurde. In der dritten Spalte wird der Trust-Region-Parameter  $t$  protokolliert. Man erkennt deutlich, wie er ab Schritt drei nach jedem Abstiegschritt erhöht wird. Die vierte Spalte gibt die euklidische Norm des Residuums im aktuellen Iterationsschritt an, Spalte fünf den vorhergesagten Abstieg. Man erkennt, dass das Residuum zunächst einiges größer ist als im Optimum. Dadurch wird der Rechenaufwand reduziert. Dies erkennt man an den letzten zwei Spalten, wo die Anzahl der Aufrufe des Mehrgitter-Verfahrens zur Bestimmung der Lösung des Kontaktproblems (Spalte 6) und des adjungierten Problems (Spalte 7) angegeben sind. Außerdem sieht man am Residuum, dass dies nicht notwendigerweise in jedem Schritt reduziert wird, sondern, dass es, falls der Abstiegschätzer  $v$  dies zulässt, vergrößert werden kann.

Schritt	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$	$F(p)$
1	0.181	0.029	0.006	-0.051	0.022	0.006	0.21320
2	0.250	0.050	-0.010	-0.250	-0.050	-0.010	0.21259
3	0.250	0.050	-0.010	-0.250	-0.050	-0.010	0.21259
4	0.250	0.050	-0.010	-0.250	-0.050	-0.010	0.21559
5	0.250	0.050	-0.010	-0.250	-0.050	0.009	0.21255
6	0.250	0.050	-0.010	-0.250	-0.050	0.009	0.21255
7	0.250	0.050	-0.010	-0.250	-0.050	0.010	0.21254
8	0.250	0.050	-0.010	-0.250	-0.050	0.010	0.21254
9	0.250	0.050	-0.010	-0.250	-0.050	0.010	0.21254

Tabelle 6.10: Die Iterierten des Formoptimierungsproblems bei inexakter Rechnung mit zufälligem Startvektor.

Schritt	It.	$t$	$\ C\ $	$ v $	# Kont.	# Adj.
1	-	-	3.4E-3	-	50	10
2	SS	200	1.3E-3	4.1E-4	62	20
3	KS	200	2.2E-4	8.1E-5	32	2
4	NS	200	2.2E-4	8.2E-5	63	33
5	SS	200	6.6E-5	6.0E-5	64	20
6	KS	200	4.1E-6	9.2E-7	29	2
7	SS	200	3.7E-6	9.2E-7	70	38
8	KS	200	7.0E-10	0	80	19
9	opt	200	-	0	-	-

Tabelle 6.11: Simulationsprotokoll des Formoptimierungsproblems bei inexakter Rechnung mit zufälligem Startwert.

Die Tabellen 6.10 und 6.11 zeigen den Simulationsverlauf zu einem zufälligen Startvektor. Die Nebenbedingung wird im ersten Schritt weniger genau ausgewertet als in der Simulation davor. Außerdem wird der Trust-Region-Parameter  $t$  anders initialisiert. Die Simulation zeigt, dass das Modell im Laufe der Simulation durch Null- und Korrekturschritte verbessert werden muss. Es wird ein anderes lokales Extremum gefunden. In beiden Lösungen liegen die Klammern antisymmetrisch zueinander. Dies entspricht der Symmetrie des Problems.

Schritt	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$	$F(p)$
1	0.000	0.000	0.000	0.000	0.000	0.000	0.191214
2	0.001	0.002	-0.004	0.003	-0.003	-0.006	0.191211
3	0.001	0.003	0.001	0.005	-0.004	-0.002	0.191211
4	0.002	0.004	-0.002	0.007	-0.006	-0.006	0.191209
5	0.003	0.006	0.003	0.012	-0.009	-0.003	0.191207
6	0.003	0.010	0.000	0.023	-0.014	-0.010	0.191201
7	-0.002	0.017	0.010	0.048	-0.022	-0.007	0.19118
8	-0.042	0.029	0.010	0.120	-0.031	-0.010	0.19108
9	-0.069	0.050	0.010	0.225	-0.050	-0.008	0.19082
10	-0.149	0.050	0.010	0.250	-0.050	-0.008	0.19054
11	-0.0249	0.050	0.010	0.250	-0.050	-0.009	0.19023
12	-0.250	0.050	0.010	0.250	-0.050	-0.010	0.19022
13	-0.250	0.050	0.010	0.250	-0.050	-0.010	0.19022

Tabelle 6.12: Die Iterierten des Formoptimierungsproblems bei exakter Rechnung mit Startvektor 0.

Schritt	It.	$t$	$\ C\ $	$ v $	# Kont.	# Adj.
1	-	-	4.7E-9	-	100	70
2	SS	10	2.2E-9	5.4E-6	93	67
3	SS	10	4.9E-9	4.7E-6	81	61
4	SS	20	4.8E-9	3.7E-6	90	64
5	SS	40	3.4E-9	5.0E-6	80	63
6	SS	80	4.8E-9	9.2E-6	96	65
7	SS	160	1.4E-9	1.8E-5	98	60
8	SS	320	3.5E-9	6.0E-5	85	68
9	SS	640	3.2E-9	1.5E-4	91	73
10	SS	1280	1.9E-9	2.4E-4	96	62
11	SS	1280	3.7E-9	2.9E-4	89	63
12	SS	1280	2.6E-9	4.2E-6	80	55
13	opt	1280	-	0	-	-

Tabelle 6.13: Simulationsprotokoll des Formoptimierungsproblems bei exakter Rechnung mit Startwert 0.

Wir vergleichen das inexakte Bundle-Trust-Region-Verfahren mit dem exakten Bundle-Verfahren, wie wir es in Kapitel 5.2 eingeführt haben. Wir wählen in beiden Verfahren die gleichen Parameter. Für das exakte Verfahren lösen wir das adjungierte (4.46) und das Kontaktproblem (3.16) jeweils mit einer Toleranz von  $10^{-8}$ . Wir rechnen auf einem Gitter mit 4784 Knoten. Wir wählen ein Bilevel-Mehrgitterverfahren als Vorkonditionierer.

Schritt	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$	$F(p)$
1	0.000	0.000	0.000	0.000	0.000	0.000	0.19121
2	0.000	0.000	0.000	0.000	0.000	0.000	0.19121
3	0.001	0.002	-0.004	0.003	-0.003	-0.006	0.19121
4	0.001	0.003	0.000	0.005	-0.004	-0.003	0.19121
5	0.003	0.005	-0.003	0.010	-0.008	-0.008	0.19121
6	0.005	0.010	0.005	0.019	-0.013	-0.005	0.19120
7	0.003	0.017	-0.002	0.043	-0.021	-0.010	0.19119
8	-0.021	0.030	0.008	0.102	-0.031	-0.010	0.19112
9	-0.063	0.050	0.007	0.218	-0.0500	-0.010	0.19085
10	-0.149	0.050	0.007	0.250	-0.0500	-0.010	0.19055
11	-0.250	0.050	0.007	0.250	-0.0500	-0.010	0.19023
12	-0.250	0.050	0.010	0.250	-0.0500	-0.010	0.19022
13	-0.250	0.050	0.010	0.250	-0.0500	-0.010	0.19022
14	-0.250	0.050	0.010	0.250	-0.0500	-0.010	0.19022

Tabelle 6.14: Die Iterierten des Formoptimierungsproblems bei inexakter Rechnung mit Startvektor 0 im Vergleich zur exakten Rechnung.

Schritt	It.	$t$	$\ C\ $	$ v $	# Kont.	# Adj.
1	-	-	3.1E-4	-	55	23
2	KS	10	5.0E-5	6.5E-6	6	2
3	SS	10	1.9E-3	6.5E-6	55	13
4	SS	20	2.3E-5	5.3E-6	48	12
5	SS	40	1.6E-5	3.5E-6	55	13
6	SS	80	2.4E-5	1.0E-5	45	13
7	SS	160	1.3E-5	1.5E-3	64	21
8	SS	320	5.7E-6	4.3E-5	63	24
9	SS	640	2.0E-5	.13E-4	57	24
10	SS	1280	1.1E-5	2.3E-4	57	21
11	SS	1280	3.5E-6	2.8E-4	66	27
12	SS	1280	7.6E-6	5.4E-6	45	15
13	KS	1280	5.9E-10	0	47	26
14	opt	1280	-	0	-	-

Tabelle 6.15: Simulationsprotokoll des Formoptimierungsproblems bei inexakter Rechnung mit Startwert 0 im Vergleich zur exakten Rechnung.

Die Tabellen 6.12 – 6.13 und 6.14 – 6.15 zeigen jeweils die Iterationsverläufe. Es wird in beiden Simulationen dasselbe Minimum gefunden. Dabei wird die Lösung in beiden Verfahren mit der gleichen Genauigkeitsschranke angegeben. Dabei benötigen beide Verfahren jeweils eine ähnliche Anzahl von Schritten. Allerdings benötigt das neue inexakte Bundle-Trust-Region-Verfahren etwas weniger als die Hälfte an Mehrgitterschritten (1850:897), was zu einer deutlichen Geschwindigkeitsverbesserung führt. Hier liegen die Vorteile des neuen Verfahrens gegenüber seinem Vorgänger. Das liegt daran, dass die Rechengenauigkeit erst am Ende stark erhöht wird.

### 6.3 Das eigentliche Problem

In diesem Abschnitt verwenden wir ein Gitter, das aus CT-Daten eines Unterkiefers mittels Amira generiert wurde. Die Gitterverfeinerung erfolgte wiederum unter Comsol. Wir betrachten wie schon bei dem Modellproblem den Fall, dass die Klammern fest mit dem Kiefer verbunden sind. Der Kiefer ist an den Kondylen, die in

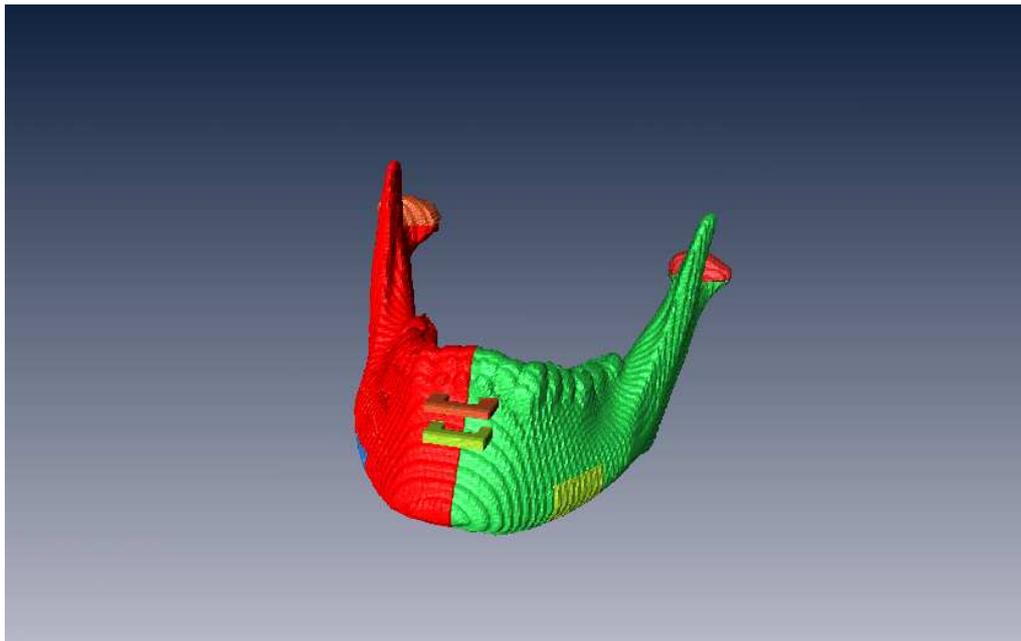


Abbildung 6.3: Die diskretisierte Kiefer-Klammer-Konfiguration: An den Kondylen, die im Bild hinten links hellrot und hinten rechts rot eingefärbt sind, wirken Dirichlet-Bedingungen, auf die blaue und gelbe Fläche eine Kraft.

der Abbildung 6.3 hinten links hellrot und hinten links rot eingefärbt sind, fest eingespannt. Dies beschreibt eine Dirichlet-Bedingung. Auf die blaue und gelbe Fläche wirkt eine äußere Kraft. Wir modellieren sie durch eine Neumann-Randbedingung. Volumenkräfte werden vernachlässigt. Zwischen den beiden Kieferhälften ist an der Frakturfläche eine Kontaktbedingung modelliert. Die Klammern sind so angebracht, dass sie den Knochen an der Kieferoberfläche auch im deformierten Zustand nicht berühren.

### 6.3.1 Das Kontaktproblem

Hier untersuchen wir wie in 6.2.1 wieder, wie die Lösung des Kontaktproblems vom Gitter und den Regularisierungsparametern abhängt. Wir wählen für Knochen und Klammer dieselben Materialdaten ( $E = 3.4 \cdot 10^5$ ,  $\mu = 0.1$ ). Auf beide Flächen wirkt dieselbe Kraft  $F = (0, 33, 10)$ . Wir rechnen auf Gittern mit einer Größe von 2500–25000 Knoten.

Die Tabellen 6.16 – 6.23 zeigen wieder die Residuen des Kontaktproblems in Abhängigkeit von den Iterationszahlen und dem Regularisierungsparameter  $\mu$ .

#	$\ \cdot\ $	$\ \cdot\ $	$\ \cdot\ $
2585	3.6E2	4.3E-4	0.0E-8
8600	4.5E2	1.8E-4	0.0E-8
25883	5.4E2	7.8E-5	0.0E-8

Tabelle 6.16: Die Residuen des Newton-Verfahrens zur Lösung auf dem Kiefernitter zum Regularisierungsparameter  $\gamma = 1$ .

#	$\ \cdot\ $	$\ \cdot\ $	$\ \cdot\ $
2585	3.6E2	4.3E-4	0.0E-8
8600	4.5E2	1.8E-4	0.0E-8
25883	5.4E2	7.8E-5	0.0E-8

Tabelle 6.17: Die Residuen des Newton-Verfahrens zur Lösung auf dem Kiefernitter zum Regularisierungsparameter  $\gamma = 1E - 1$ .

#	$\ \cdot\ $	$\ \cdot\ $	$\ \cdot\ $
2585	3.6E2	4.4E-4	0.0E-8
8600	4.5E2	1.8E-4	0.0E-8
25883	5.4E2	7.8E-5	0.0E-8

Tabelle 6.18: Die Residuen des Newton-Verfahrens zur Lösung auf dem Kiefernitter zum Regularisierungsparameter  $\gamma = 1E - 2$ .

#	$\ \cdot\ $	$\ \cdot\ $	$\ \cdot\ $	$\ \cdot\ $
2585	3.6E2	4.4E-4	6.5E-7	0.0E-8
8600	4.5E2	1.8E-4	8.4E-8	0.0E-8
25883	5.4E2	7.8E-5	2.3E-8	0.0E-8

Tabelle 6.19: Die Residuen des Newton-Verfahrens zur Lösung auf dem Kiefernitter zum Regularisierungsparameter  $\gamma = 1E - 3$ .

#	$\ \cdot\ $	$\ \cdot\ $	$\ \cdot\ $	$\ \cdot\ $
2585	3.6E2	4.3E-4	6.2E-6	0.0E-8
8600	4.5E2	1.8E-4	8.3E-7	0.0E-8
25883	5.4E2	7.8E-5	2.3E-7	0.0E-8

Tabelle 6.20: Die Residuen des Newton-Verfahrens zur Lösung auf dem Kiefernitter zum Regularisierungsparameter  $\gamma = 1E - 4$ .

#	$\ \cdot\ $				
2585	3.6E2	3.8E-4	4.5E-5	1.1E-6	0.0E-8
8600	4.5E2	1.7E-4	7.5E-5	0.0E-8	0.0E-8
25883	5.4E2	7.6E-5	2.1E-6	0.0E-8	0.0E-8

Tabelle 6.21: Die Residuen des Newton-Verfahrens zur Lösung auf dem Kiefernitter zum Regularisierungsparameter  $\gamma = 1E - 5$ .

#	$\ \cdot\ $					
2585	3.6E2	1.9E-4	9.4E-5	1.0E-5	1.1E-6	0.0E-8
8600	4.3E2	1.2E-6	3.6E-5	1.1E-6	4.4E-8	0.0E-8
25883	5.4E2	6.4E-5	1.3E-5	6.2E-7	0.0E-8	0.0E-8

Tabelle 6.22: Die Residuen des Newton-Verfahrens zum Regularisierungsparameter  $\gamma = 1E - 6$ .

#	$\ \cdot\ $								
2585	3.6E2	5.0E-5	3.7E-5	1.0E-5	3.9E-6	5.7E-7	3.1E-8	0.0E-8	0.0E-8
8600	4.5E2	4.0E-5	2.8E-5	5.3E-6	3.2E-6	2.7E-7	1.9E-7	7.7E-8	0.0E-8
25883	5.4E2	2.9E-5	1.8E-5	2.3E-6	5.6E-7	3.6E-8	0.0E-8	0.0E-8	0.0E-8

Tabelle 6.23: Die Residuen des Newton-Verfahrens zur Lösung auf dem Kiefernitter zum Regularisierungsparameter  $\gamma = 1E - 7$ .

Auch hier kann man wieder die Gitterunabhängigkeit des Verfahrens beobachten. Für kleinere Regularisierungsparameter läuft das Verfahren nicht mehr stabil.

### 6.3.2 Formoptimierung

Hier bestimmen wir wie in Absatz 6.2.2 eine optimale Position der Klammern. Wir verwenden die horizontale und vertikale Koordinate des Schwerpunktes und die Ausrichtung jeder Klammer als Parameter. Als Zielfunktional betrachten wir wie oben

das Quadrat über die Verschiebung der Konfiguration

$$J(\Omega) := \int_{\Omega} \|u(\Omega)\|_2^2 dx.$$

Das Bundle-Verfahren wird mit folgenden Parametern verwendet:  $\gamma_1 = 0.3$ ,  $\gamma_2 = 0.7$ ,  $\gamma_3 = 0.5$ ,  $c_0 = 0.001$ ,  $\kappa = 10$ .

Wir wählen als Nebenbedingung das regularisierte Kontaktproblem mit Regularisierungsparameter  $\gamma = 0.1$  und für die die Formparameter  $p$  die Nebenbedingung

$$\begin{aligned} -0.15 &\leq p_1 \leq 0.15, \\ -0.20 &\leq p_2 \leq 0.05, \\ -0.01 &\leq p_3 \leq 0.01, \\ -0.15 &\leq p_4 \leq 0.15, \\ -0.05 &\leq p_5 \leq 0.005, \\ -0.01 &\leq p_6 \leq 0.01. \end{aligned}$$

Für die Folge zur Bestimmung des Relaxationsparameters wählen wir  $(a_j)_{j \in \mathbb{N}} = 2^{-j}/1000$ .

Wir initialisieren den Trust-Region-Parameter mit  $t = 1$ . Falls im Laufe der Iteration zwei aufeinander folgende Abstiegschritte durchgeführt werden, so verdoppeln wir den Trust-Region-Parameter, im Falle von zwei aufeinander folgenden Nullschritten halbieren wir den Parameter. Wir rechnen auf dem Gitter mit 8600 Knoten, wobei wir das Gitter mit 2585 sowohl als zweites Gitter für das Mehrgitterverfahren verwenden, als auch zur Bestimmung der Gebietstransformationen verwendet haben. Wie oben benutzen wir jeweils vier Vor- und Nachglättungsschritte. Wir verwenden beim Kontaktproblem und adjungierten Problem wieder die Lösung aus dem vorherigen Schritt als Startwert für die neue Iteration.

Wir starten nun die Simulation mit dem Nullvektor. Die Iterationsschritte sind dabei in Tabelle 6.24 angegeben. Tabelle 6.25 zeigt das Simulationsprotokoll, das genauso aufgebaut ist wie in Abschnitt 6.2.2. Man erkennt wieder, dass die Genauigkeit der Lösung des Kontaktproblems zu Beginn stark variiert und erst am Schluß die volle Genauigkeit erreicht wird.

Schritt	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$	$F(p)$
1	0.000	0.000	0.000	0.000	0.000	0.000	4.0134
2	-0.008	0.005	0.0023	-0.009	0.024	-0.005	4.0125
3	-0.018	0.005	0.0046	-0.18	0.048	-0.009	4.0117
4	-0.027	0.005	0.0068	-0.027	0.050	-0.010	4.0116
5	-0.045	0.005	0.01	-0.043	0.050	-0.010	4.0113
6	-0.081	0.005	0.01	-0.075	0.050	-0.010	4.0107
7	-0.150	0.005	0.01	-0.136	0.050	-0.010	4.0098
8	-0.150	0.005	0.01	-0.150	0.050	-0.010	4.0096
9	-0.150	0.005	0.01	-0.150	0.050	-0.010	4.0096
10	-0.150	0.005	0.01	-0.150	0.050	-0.010	4.0096

Tabelle 6.24: Die Iterierten des Formoptimierungsproblems auf dem Kiefergitter bei inexakter Rechnung zum Startwert 0.

Schritt	It.	$t$	$\ C\ $	$ v $	# Kont.	# Adj.
1	-	-	3.6E-5	-	111	34
2	SS	1	3.8E-4	8.2E-4	69	8
3	SS	1	1.1E-4	7.8E-4	79	12
4	SS	2	2.8E-5	2.0E-4	63	20
5	SS	4	1.4E-4	3.1E-4	54	3
6	SS	8	5.9E-5	5.9E-4	61	11
7	SS	16	2.6E-5	1.1E-3	71	27
8	SS	32	7.06E-5	9.8E-5	67	3
9	KS	32	4.5E-9	0	80	105
10	opt	32	-	0	-	-

Tabelle 6.25: Simulationsprotokoll des Formoptimierungsproblems auf dem Kiefergitter bei inexakter Rechnung zum Startwert 0.

Nun starten wir die Simulation mit einem zufälligen zulässigen Startvektor und finden den gleichen stationären Punkt. Die Simulationen werden in den Tabelle 6.26 und 6.27 beschrieben. Wir brauchen zwar mehr Iterationsschritte als beim ersten Mal, allerdings unterscheidet sich die Anzahl der Aufrufe des Mehrgitter-Verfahrens, worin der Hauptrechenaufwand liegt, nur um eine Iteration.

Schritt	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$	$F(p)$
1	-0.013	-0.005	0.006	-0.017	0.001	0.006	4.0131
2	-0.021	0.003	0.008	-0.026	0.026	0.001	4.0123
3	-0.031	0.005	0.01	-0.033	0.050	-0.002	4.0116
4	-0.039	0.005	0.01	-0.042	0.050	-0.004	4.0114
5	-0.058	0.005	0.01	-0.058	0.050	-0.010	4.0111
6	-0.094	0.005	0.01	-0.090	0.050	-0.010	4.0105
7	-0.150	0.005	0.01	-0.150	0.050	-0.010	4.0096
8	-0.150	0.005	0.01	-0.150	0.050	-0.010	4.0096
9	-0.150	0.005	0.01	-0.150	0.050	-0.010	4.0096

Tabelle 6.26: Die Iterierten des Formoptimierungsproblems auf dem Kiefernitter bei inexakter Rechnung zum zufälligen Startwert.

Schritt	It.	$t$	$\ C\ $	$ v $	# Kont.	# Adj.
1	-	1	2.95E-5	-	111	44
2	SS	1	2.41E-4	8.6E-4	75	8
3	SS	1	8.4E-5	7.6E-4	77	12
4	SS	2	4.3E-5	1.6E-4	47	12
5	SS	4	6.6E-5	3.1E-4	51	4
6	SS	8	2.5E-8	5.9E-4	70	18
7	SS	16	4.2E-5	9.6E-4	75	11
8	KS	16	4.3E-9	0	111	153
9	opt	16	-	0	-	-

Tabelle 6.27: Simulationsprotokoll des Formoptimierungsproblems auf dem Kiefernitter bei inexakter Rechnung zum zufälligem Startwert.

Zusammenfassend haben wir in diesem Kapitel gezeigt, wie das semiglatte Newton-Verfahren als gitterunabhängiger Löser für das Kontaktproblem funktioniert und die Anwendung des neuen inexakten Bundle-Trust-Region-Verfahrens auf Formoptimierungsprobleme mit Kontaktproblemen als Nebenbedingung vorgeführt. Dabei konnten wir gegenüber einem exakten Verfahren eine deutliche Geschwindigkeitssteigerung feststellen, da die Anzahl der Mehrgitterschritte um die Hälfte reduziert wurde.

# Kapitel 7

## Ausblick

In dieser Arbeit haben wir einen iterativen Algorithmus zum Lösen von Formoptimierungsproblemen mit Kontaktproblemen als Nebenbedingungen vorgestellt. Außerdem haben wir eine mögliche Anwendung beschrieben.

Dadurch, dass das vorgestellte Bundle-Verfahren und das iterative Lösungsverfahren für das Kontaktproblem unabhängig sind, bietet sich auch die Kombination mit anderen iterativen Lösern an, wodurch neue Lösungsverfahren für zahlreiche Anwendungsverfahren bereitgestellt werden können.

Wir haben in Abschnitt 6.2.2 gesehen, dass das inexakte Bundle-Trust-Region-Verfahren dem exakten dadurch überlegen ist, dass die Anzahl der Mehrgitterschritte während der Lösung des adjungierten oder des Kontaktproblems deutlich niedriger ausfällt. Dies führte zu einer starken Reduzierung der Simulationsdauer, da auf den Mehrgitter-Vorkonditionierer die meiste Rechenzeit fällt. Allerdings wird eine geschicktere Wahl des Vorkonditionierers zu einer weiteren Performanceverbesserung führen.

Das hier vorgestellte inexakte Bundle-Trust-Region-Verfahren hat den Nachteil, dass wie bereits auch schon im Bundle-Trust-Region-Verfahren [Sch89] die Trust-Region-Philosophie nicht vollständig umgesetzt werden konnte, da die Liniensuche diese verletzt. Gleiches gilt in unserem Verfahren ebenfalls für die Inexaktheit, die in der Liniensuche geopfert werden musste. Gegenstand zukünftiger Überlegungen ist also die Algorithmen so zu entwerfen, dass die Trust-Region- oder die Inexaktheitsphilosophie bewahrt bleibt.

Bislang haben wir nur Kontaktprobleme ohne Reibung behandelt. Für die Anwendung ist die Hinzunahme von Reibung wünschenswert, da sie in der Realität auftritt. Außerdem sind die Kontaktflächen in der Regel nicht glatt, was zum Beispiel auch durch Reibung modelliert werden kann. Eine zukünftige Erweiterung des Newton-Verfahrens für Kontaktprobleme mit Reibung bietet sich an.

Die medizinische Anwendung und ein Lösungsansatz wurden in dieser Arbeit nur kurz angerissen. Das hier vorgestellte Verfahren bietet Potenzial, neue Verfahren in der Osteosynthese zu validieren. Hierzu ist ein Vergleich der numerisch berechne-

ten Lösung mit Experimenten dringend erforderlich. Langfristig kann unser Algorithmus zu einer individuelleren Behandlung von Kieferfrakturen beitragen. Auch müssen geeignete Zielfunktionale diskutiert werden. Wir haben zur Vereinfachung nur die  $L^2$ -Norm der Gesamtverschiebung minimiert, in der konkreten Anwendung sind andere Zielfunktionale wie zum Beispiel die Gesamtspannung, der Druck auf die Kontaktfläche oder die Verschiebung aus einer Referenzlage vorzuziehen.

# Anhang A

## Lineare Algebra

Wir fassen hier einige Eigenschaften der Frobeniusnorm für Matrizen zusammen, die wir in der Arbeit verwendet haben.

**Satz A.1** (Die Frobeniusnorm). *Sei  $A \in \mathbb{R}^{n \times n}$  eine beliebige Matrix, dann wird durch*

$$\|A\|_F := \left( \sum_{i,j=1}^n A_{i,j}^2 \right)^{1/2}$$

*eine Norm mit den Eigenschaften*

(a)  $\|AB\|_F \leq \|A\|_F \|B\|_F$  für  $A, B \in \mathbb{R}^{n \times n}$ .

(b) Für zwei  $3 \times 3$  Matrizen gilt

$$|A : B| \leq \|A\|_F \|B\|_F$$

*mit der Frobeniusnorm  $\|\cdot\|_F$ .*

*Beweis:* zu b)

Wir fassen die Matrizen  $A$  und  $B$  spaltenweise gelesen als Vektoren  $\tilde{A}$  und  $\tilde{B}$  der Dimension 9 auf. Dann entspricht  $A : B$  dem Skalarprodukt dieser Vektoren. Also folgt mit der Cauchy-Schwarzschen-Ungleichung

$$|A : B| = |\langle \tilde{A}, \tilde{B} \rangle| \leq \|\tilde{A}\|_2 \|\tilde{B}\|_2 = \|A\|_F \|B\|_F.$$

■

**Satz A.2.** *Sei  $A, B \in \mathbb{R}^{3 \times 3}$ , dann ist*

$$\|AB\|_F^2 \leq 9 \|B\|_\infty^2 \|A\|_F^2.$$

*Dabei bezeichnet  $\|\cdot\|_F$  die Frobeniusnorm und  $\|B\|_\infty$  das betragsgrößte Element von  $B$ .*

*Beweis:* Bezeichne  $A = (a_{ij})_{1 \leq i, j \leq 3}$  und  $B = (b_{ij})_{1 \leq i, j \leq 3}$ , dann gilt

$$\begin{aligned} \|AB\|_F^2 &= \sum_{i=1}^3 \sum_{j=1}^3 \left( \sum_{k=1}^3 a_{i,k} b_{k,j} \right)^2 \\ &= \sum_{i=1}^3 \sum_{j=1}^3 (a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j})^2 \\ &\leq \sum_{i=1}^3 \sum_{j=1}^3 (|a_{i1}b_{1j}|^2 + |a_{i2}b_{2j}|^2 + |a_{i3}b_{3j}|^2 + 2|a_{i1}b_{1j}a_{i2}b_{2j}| \\ &\quad + 2|a_{i1}b_{1j}a_{i3}b_{3j}| + 2|a_{i3}b_{3j}a_{i2}b_{2j}|). \end{aligned}$$

Unter Verwendung von  $2|ab| \leq a^2 + b^2$  ergibt sich

$$\begin{aligned} \|AB\|_F^2 &\leq 3 \sum_{i=1}^3 \sum_{j=1}^3 (|a_{i1}b_{1j}|^2 + |a_{i2}b_{2j}|^2 + |a_{i3}b_{3j}|^2) \\ &\leq 3 \max_{i,j} |b_{i,j}|^2 \sum_{i=1}^3 \sum_{j=1}^3 (|a_{i1}|^2 + |a_{i2}|^2 + |a_{i3}|^2) \\ &= 9 \|B\|_\infty^2 \|A\|_F^2. \end{aligned}$$

■

# Anhang B

## Nichtglatte Analysis

In diesem Abschnitt fassen wir einige Definitionen und Sätze zusammen, die aus der Theorie der konvexen Analysis beziehungsweise der Analysis von Lipschitz-stetigen Funktionen stammen. Konvexe Funktionen, die auf offenen Teilmengen des  $\mathbb{R}^n$  definiert sind, sind Lipschitz-stetig, dadurch werden sie im Folgenden, wenn nicht nötig, nicht extra erwähnt.

Der zentrale Satz bezüglich der Differenzierbarkeit ist der Satz von Rademacher:

**Satz B.1** (Rademacher). *Sei  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  lokal Lipschitz-stetig, dann ist  $F$  fast überall differenzierbar.*

*Beweis:* [EG92] ■

Aufbauend auf diesen Satz kann man nun einen verallgemeinerten Ableitungsbegriff definieren.

**Definition B.2** (Verallgemeinerte Ableitungen). *Sei  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  lokal Lipschitz-stetig und  $U_d \subset U$  die Menge der Punkte, an denen  $F$  differenzierbar ist, so wird Folgendes definiert:*

- *Das Bouligand oder B-Subdifferential*

$$\partial^B F(x) := \{M \in \mathbb{R}^{m \times n}; \exists x_k \rightarrow x; x_k \in U_d; F'(x_k) \rightarrow M\}$$

- *Clarkes verallgemeinerte Ableitung*

$$\partial^{Cl} F(x) := \text{conv}(\partial^B F(x))$$

**Satz B.3.** *Sei  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  lokal Lipschitz-stetig mit Lipschitz-Konstante  $L$  und  $\partial^{Cl} F$  das Clarkesche Differential von  $F$ , dann gilt:*

- $\partial^{Cl} F(x)$  ist nichtleer, konvex und kompakt und es gilt  $\partial F(x) \subset LB(0, 1)$ .*
- Die mengenwertige Abbildung  $x \mapsto \partial^{Cl} F(x)$  ist oberhalb stetig.*

*Beweis:* [OKZ98] ■

**Satz B.4.** Sei  $f : \mathbb{R}^n \subset \mathbb{R}^n \rightarrow \mathbb{R}$  konvex, dann gilt:

- (a)  $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$  für alle  $x, y \in \mathbb{R}^n$  und  $\lambda \in [0, 1]$ .  
 (b) Für jedes  $x \in \mathbb{R}^n$  existiert ein  $g \in \mathbb{R}^n$ , so dass für alle  $y \in \mathbb{R}^n$  gilt

$$f(y) \geq f(x) + g^T(y - x). \quad (\text{B.1})$$

Wir nennen  $g$  einen Subgradienten von  $f$  an der Stelle  $x$ . Die Menge aller Subgradienten heißt Subdifferential. Dies stimmt mit dem Clarkeschen Differential aus Definition B.2 überein.

*Beweis:* [Alt04] ■

Anschaulich bedeutet dies, dass die Tangentialebene durch  $(x, f(x))$  mit Normale  $g$  an  $f$  unterhalb von  $f$  verläuft. Diese Eigenschaft nennt man auch Stützebeneeigenschaft.

Es kann vorkommen, dass man die Subgradienten nur näherungsweise kennt, sondern nur einen Näherungswert an sie. Die bekanntesten Näherungen an Subgradienten sind die sogenannten  $\epsilon$ -Subgradienten.

**Definition B.5** ( $\epsilon$ -Subgradienten). Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  und  $\epsilon > 0$ . Dann heißt  $g_\epsilon \in \mathbb{R}^n$  approximativer Subgradient oder  $\epsilon$ -Subgradient in  $x$ , falls

$$f(y) \geq f(x) + g_\epsilon^T(y - x) - \epsilon \quad \forall y \in \mathbb{R}^n \quad (\text{B.2})$$

*gilt.*

Anschaulich bedeutet dies, dass die Ebene durch  $(x, f(x) - \epsilon)$  mit Flächennormale  $g_\epsilon$  unterhalb von  $f$  verläuft. Sie ist in der Regel keine Tangentialebene mehr.

# Anhang C

## Hilfsmittel aus der Funktionalanalysis

Wir fassen hier wichtige verwendete Sätze der Funktionalanalysis zusammen:

**Satz C.1** (Satz von Lax Milgram). *Sei  $(V, \|\cdot\|)$  ein reeller Hilbert-Raum und  $K \subset V$  eine nichtleere konvexe Teilmenge von  $V$ . Gegeben seien eine Bilinearform  $a : V \times V \rightarrow \mathbb{R}$  und eine Linearform  $L : V \rightarrow \mathbb{R}$ . Für  $a$  und  $L$  gelte*

$$\begin{aligned}\exists M > 0 : |a(u, v)| &\leq M\|u\|\|v\| \quad \forall u, v \in V, \\ \exists \alpha > 0 : a(v, v) &\geq \alpha\|v\|^2 \quad \forall v \in V, \\ \exists C > 0 : |L(v)| &\leq C\|v\| \quad \forall v \in V.\end{aligned}$$

Dann hat die Variationsungleichung

$$\text{Finde } u \in K : a(u, v - u) \geq L(v - u) \quad \forall v \in K \quad (\text{C.1})$$

genau eine Lösung und es gilt

$$\|u\| \leq \frac{1}{\alpha} \|L\|.$$

Falls  $a$  symmetrisch ist, also  $a(u, v) = a(v, u)$  für alle  $u, v \in V$  gilt, dann ist das Lösen der Variationsungleichung äquivalent zur Bestimmung der Lösung für folgendes Minimierungsproblem:

$$\text{Finde } u \in K : J(u) = \min_{u \in K} J(v),$$

wobei

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

ist. Falls  $K = V$  gilt, wird aus der Variationsungleichung (C.1) folgende Variationsgleichung:

$$\text{Finde } u \in K : a(u, v) = L(v) \quad \forall v \in V.$$

**Satz C.2** (Kornsche Ungleichung). *Sei  $\Omega$  eine beschränkte offene Teilmenge im  $\mathbb{R}^n$  mit stückweise glattem Rand. Ferner habe  $\Gamma_D \subset \partial\Omega$  ein positives zweidimensionales Maß. Dann gibt es eine Zahl  $c = c(\Omega, \Gamma_D) > 0$ , so dass*

$$\int_{\Omega} \epsilon(u) : \epsilon(v) \, dx \geq c \|v\|_{1,\Omega}^2 \quad \forall v \in (H_{\Gamma}^1(\Omega))^n$$

*gilt. Dabei ist  $H_{\Gamma}^1(\Omega)$  der Abschluss von  $\{v \in C^{\infty}(\Omega)^3 : v(x) = 0 \forall x \in \Gamma_D\}$  bezüglich der  $\|\cdot\|_{1,\Omega}$ -Norm.*

# Literaturverzeichnis

- [Abe05] ABELN, A.: *Primärstabilität von Osteosynthese-Staples aus Formgedächtnismaterial vs. Miniplattenosteosynthese bei der operativen Versorgung der Unterkieferfraktur – eine biomechanische Studie*, TU München, Diss., 2005  
Zitiert auf S. 111.
- [Ada75] ADAMS, R. A.: *Sobolev spaces*. Academic Press, 1975  
Zitiert auf S. 27, 37.
- [Alt02] ALT, W.: *Nichtlineare Optimierung*. Vieweg, 2002  
Zitiert auf S. 73.
- [Alt04] ALT, W.: *Numerische Verfahren der konvexen, nichtglatten Optimierung*. Teubner, 2004  
Zitiert auf S. 74, 132.
- [Bra02] BRAESS, D.: *Finite Elemente*. Springer, 2002  
Zitiert auf S. 5, 65.
- [BS03] BENDSØE, M. P. ; SIGMUND, O.: *Topology optimization*. Springer, 2003  
Zitiert auf S. 39.
- [CGT00] CONN, A. R. ; GOULD, N. I. M. ; TOINT, P. L.: *Trust-region methods*. SIAM, 2000  
Zitiert auf S. 84, 88.
- [Cia88] CIARLET, P.: *Mathematical Elasticity*. Bd. 1 Three Dimensional Elasticity. Elsevier, 1988  
Zitiert auf S. 5.
- [Cla90] CLARKE, F. H.: *Optimization and nonsmooth analysis*. SIAM, 1990  
Zitiert auf S. 22.
- [Dob06] DOBROWOLSKI, M.: *Angewandte Funktionalanalysis*. Springer, 2006  
Zitiert auf S. 47.
- [Eck96] ECK, C.: *Existenz und Regularität der Lösung für Kontaktprobleme mit Reibung*, Universität Stuttgart, Diss., 1996  
Zitiert auf S. 11.

- [EG92] EVANS, L. C. ; GARIEPY, R. F.: *Measure theory and fine properties of functions*. CRC Press, 1992  
Zitiert auf S. 65, 131.
- [ES98] EGOROV, Y. V. ; SHUBIN, M. A.: *Foundations of the classical theory of partial differential equations*. Springer, 1998  
Zitiert auf S. 27.
- [Eva98] EVANS, L. C.: *Partial differential equations*. American Mathematical Society, 1998  
Zitiert auf S. 12.
- [GK99] GEIGER, C. ; KANZOW, C.: *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*. Springer, 1999  
Zitiert auf S. 32.
- [GK02] GEIGER, C. ; KANZOW, C.: *Theorie und Numerik restringierter Optimierungsaufgaben*. Springer, 2002  
Zitiert auf S. 23.
- [Glo84] GLOWINSKI, R.: *Numerical methods for nonlinear variational problems*. Springer-Verlag, 1984  
Zitiert auf S. 28.
- [Hac85] HACKBUSCH, W.: *Multi-grid methods and applications*. Springer, 1985  
Zitiert auf S. 32.
- [Hac94] HACKBUSCH, Wolfgang: *Iterative solution of large sparse systems of equations*. Springer, 1994  
Zitiert auf S. 32.
- [HCK86] HAUG, E.J. ; CHOI, K.K. ; KOMKOV, V.: *Design Sensitivity of Structural Systems*. Academic Press, 1986  
Zitiert auf S. 40.
- [HIK02] HINTERMÜLLER, M. ; ITO, K. ; KUNISCH, K.: The primal-dual active set strategy as a semismooth Newton method. In: *SIAM J. Optim.* 13 (2002), Nr. 3, S. 865–888  
Zitiert auf S. 28.
- [Hin01] HINTERMÜLLER, M.: A proximal bundle method based on approximate subgradients. In: *Comput. Optim. Appl.* 20 (2001), S. 245–266  
Zitiert auf S. 81, 95, 107.
- [HM03] HASLINGER, J. ; MÄKINEN, R. A. E.: *Introduction to shape optimization*. SIAM, 2003  
Zitiert auf S. 5, 40, 45.

- [HN96] HASLINGER, J. ; NEITTAANMAKI, P.: *Finite element approximation for optimal shape, material and topology design*. John Wiley & Sons Ltd., 1996  
Zitiert auf S. 40.
- [Hor97] HORCH, H.-H. (Hrsg.): *Praxis der Zahnheilkunde 10/1: Mund-Kiefer-Gesichtschirurgie I*. Urban und Schwarzenberg, 1997  
Zitiert auf S. 111.
- [HPUU05] HINZE, M. ; PINNAU, R. ; ULBRICH, M. ; ULBRICH, S.: *Lecture Notes of the Autumn School Modelling and Optimization with Partial Differential Equations*. 2005  
Zitiert auf S. 26.
- [HU04] HINTERMÜLLER, M. ; ULBRICH, M.: A mesh-independence result for semismooth Newton methods. In: *Mathematical Programming* (2004)  
Zitiert auf S. 116.
- [HV02] HEINKENSCHLOSS, M. ; VICENTE, L.N.: Analysis of inexact trust-region SQP algorithms. In: *SIAM J. Optim.* 12 (2001/02), Nr. 2, S. 283–302  
Zitiert auf S. 71.
- [JS04] JARRE, F. ; STOER, J.: *Optimierung*. Springer, 2004  
Zitiert auf S. 73.
- [Kiw85] KIWIEL, K.C.: *Methods of Descent for Nondifferentiable Optimization*. Springer, 1985  
Zitiert auf S. 107.
- [KO88] KIKUCHI, N. ; ODEN, J. T.: *Contact problems in elasticity: a study of variational inequalities and finite element methods*. SIAM, 1988  
Zitiert auf S. 1, 8, 28.
- [Kön97] KÖNIGSBERGER, K.: *Analysis 2*. Springer, 1997  
Zitiert auf S. 46, 53.
- [KS03] KELLEY, C.T. ; SACHS, E.W.: Truncated Newton methods for optimization with inaccurate functions and gradients. In: *J. Optim. Theory Appl.* 116 (2003), Nr. 1, S. 83–98  
Zitiert auf S. 71.
- [Lem81] LEMARÉCHAL, C.: A view of Line Searches. In: OETTLI, W. (Hrsg.) ; STOER, J. (Hrsg.): *Optimization and Optimal Control*, Springer, 1981 (Lecture Notes in Control and Information Science)  
Zitiert auf S. 78, 91.
- [LL66] LANDAU, L.D. ; LIFSCHITZ, E.M.: *Lehrbuch der theoretischen Physik*. Bd. 7 Elastizität. Akademie Verlag, 1966  
Zitiert auf S. 5.

- [LPR96] LUO, Z.-Q. ; PANG, J.-S. ; RALPH, D.: *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, 1996  
*Zitiert auf S. 40.*
- [LS99] LEIBFRITZ, F. ; SACHS, E.W.: Inexact SQP interior point methods and large scale optimal control problems. In: *SIAM J. Control. Optim.* 38 (1999), Nr. 1, S. 272–293  
*Zitiert auf S. 71.*
- [MS76] MURAT, F. ; SIMON, J.: Studies of optimal shape design problems. In: *Lecture Notes in Computer Science 41* (1976)  
*Zitiert auf S. 40.*
- [Nit81] NITSCHKE, J.A.: On Korn's second inequality. In: *RAIRO Anal. Numér.* 15 (1981), S. 237–248  
*Zitiert auf S. 50.*
- [OKZ98] OUTRATA, Jiří ; KOČVARA, Michal ; ZOWE, Jochem: *Nonsmooth approach to optimization problems with equilibrium constraints*. Kluwer Academic Publishers, 1998  
*Zitiert auf S. 2, 36, 66, 69, 70, 132.*
- [Saa03] SAAD, Y.: *Iterative methods for sparse linear systems*. SIAM, 2003  
*Zitiert auf S. 32.*
- [Sch89] SCHRAMM, H.: *Eine Kombination von Bundle- und Trust-Region Verfahren zur Lösung nichtdifferenzierbarer Optimierungsprobleme*, Universität Bayreuth, Diss., 1989  
*Zitiert auf S. 3, 71, 74, 79, 81, 91, 95, 107, 108, 127.*
- [Sch03] SCHIEFERSTEIN, H.: *Experimentelle Analyse des menschlichen Kauesystems*, TU München, Diss., 2003  
*Zitiert auf S. 111.*
- [Sol03] SOLODOV, M. V.: On approximations with finite precision in bundle methods for nonsmooth optimization. In: *J. Optim. Theory Appl.* 119 (2003), S. 151–165  
*Zitiert auf S. 81.*
- [Ste03] STEINBACH, O.: *Numerische Näherungslösungen für elliptische Randwertprobleme*. Teubner, 2003  
*Zitiert auf S. 65.*
- [SZ92a] SCHRAMM, H. ; ZOWE, J.: A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results. In: *SIAM J. Optim.* 2 (1992), Nr. 1, S. 121–152  
*Zitiert auf S. 71, 77, 78, 95.*

- [SZ92b] SOKOLOWSKI, J. ; ZOLÉSIO, J.-P.: *Introduction to shape optimization*. Springer, 1992  
*Zitiert auf S. 40, 52.*
- [Ul02a] ULBRICH, M.: *Nonsmooth Newton-like Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*. Habilitation, TU München, 2002  
*Zitiert auf S. 2, 24, 26, 28.*
- [Ul02b] ULBRICH, M.: *Skriptum zur nichtglatten Optimierung*. 2002. – Vorlesung TU München  
*Zitiert auf S. 20, 21.*
- [Ul03] ULBRICH, M.: Semismooth Newton methods for operator equations in function spaces. In: *SIAM Journal of Optimization* 13 (2003), Nr. 3, S. 805–841  
*Zitiert auf S. 24, 26.*
- [Ul07] ULBRICH, S.: Generalized SQP-Methods with Parareal Time-Domain Decomposition for Time-dependent PDE-constrained Optimization. In: BIEGLER, L. (Hrsg.) ; GHATTAS, O. (Hrsg.) ; HEINKENLSCHLOSS, M. (Hrsg.) ; KEYES, D. (Hrsg.) ; BLOEMEN WAANDERS, B. van (Hrsg.): *Real-Time PDE-Constrained Optimization*, SIAM, Philadelphia, in press, 2007  
*Zitiert auf S. 71, 85.*
- [UU] ULBRICH, M. ; ULBRICH, S.: A multigrid semismooth Newton method for contact problems in linear elasticity. – Forschungsbericht  
*Zitiert auf S. 1, 14, 15, 18.*
- [WDH93] WILMOTT, P. ; DEWYNNE, J. ; HOWISON, S.: *Option Pricing: Mathematical Models and Computation*. Oxford Financial Press, 1993  
*Zitiert auf S. 23.*
- [Wer00] WERNER, D.: *Funktionalanalysis*. Springer, 2000  
*Zitiert auf S. 65.*