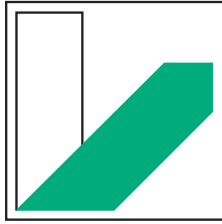


Universität Bayreuth



UNIVERSITÄT
BAYREUTH

Fakultät für Mathematik, Physik und Informatik

Lehrstuhl für Angewandte Mathematik

Diplomarbeit

Diskretisierungsfehler bei der Optimalsteuerung der linearen
Wärmeleitungsgleichung

Tobias Masching

Bayreuth, den 30. April 2012

Prüfer: Prof. Dr. Lars Grüne

Betreuer: Dipl.-Math. Nils Altmüller

Masching, Tobias

tobias.masching@googlemail.com

Diskretisierungsfehler bei der Optimalsteuerung der linearen Wärmeleitungsgleichung

Diplomarbeit, Fakultät für Mathematik, Physik und Informatik

Universität Bayreuth, April 2012

Zusammenfassung

Die vorliegende Diplomarbeit beschäftigt sich damit, den Diskretisierungsfehler bei der optimalen Steuerung der linearen Wärmeleitungsgleichung im stationären Fall für beschränkte Gebiete $\Omega \subset \mathbb{R}^N$, $N \leq 3$. Zur Diskretisierung wird die Finiten-Elemente-Methode gewählt. Das sich daraus ergebende Modellproblem der optimalen Temperaturquelle besitzt ein lineares quadratisches Zielfunktional und führt mit linearen Steuerbeschränkungen auf ein konvexes Optimierungsproblem in L^2 -Räumen. Für dieses wird, wie zunächst auch für die lineare Wärmeleitungsgleichung, die Existenz, bzw. im strikt konvexen Fall die Eindeutigkeit einer globalen Lösung gezeigt.

Des Weiteren werden hinreichende und notwendige Optimalitätsbedingungen zur Charakterisierung der eindeutigen optimalen Steuerung hergeleitet. Diese werden benutzt um den Diskretisierungsfehler zu bestimmen. Zur Diskretisierung wird die Finite Elemente Methode auf einer simplizialen Zerlegung von Ω mit dem diskreten Ansatzraum der stetigen und stückweise linearen bzw. stückweise konstanten Funktionen benutzt.

Neben dem Diskretisierungsfehler für die lineare Wärmeleitungsgleichung wird schlussendlich auch der Diskretisierungsfehler für das Modellproblem der optimalen Temperaturquelle in der $L^2(\Omega)$ -Norm abgeschätzt. Es wird gezeigt, dass sich dieser bei Verzicht einer Steuerungsdiskretisierung im Vergleich zur Voll-Diskretisierung um eine h -Potenz günstiger gestalten lässt. Der Diskretisierungsfehler der variationellen Diskretisierung hat somit keine schlechtere Ordnung als der, der durch Diskretisierung der linearen Wärmeleitungsgleichung entsteht.

Inhaltsverzeichnis

I Einführung und Motivation	1
§1 Die lineare Wärmeleitungsgleichung als Modellproblem	2
II Schwache Lösungen linearer elliptischer PDEs	6
§2 Normierte lineare Räume	6
§3 Lebesgue- und Sobolevräume	7
§3.1 Lebesgueräume	8
§3.2 Reguläre Gebiete	10
§3.3 Schwache Ableitung und Sobolevräume	13
§4 Stetige Lineare Operatoren und Funktionale	18
§5 Schwache Lösungen der Poissongleichung	22
III Die Existenz optimaler Steuerungen	28
§6 Schwache Konvergenz und Konvexität	28
§7 Existenz optimaler Steuerungen	32
IV Optimalitätsbedingungen erster Ordnung	39
§8 Differenzierbarkeit in normierten linearen Räumen	39
§9 Adjungierte Operatoren	43
§10 Notwendige und hinreichende Optimalitätsbedingungen	45
§10.1 Quadratische Optimierungsaufgabe im Hilbertraum	46
§10.2 Die adjungierte Gleichung	48
§10.3 Die Projektionsformel	53
V Diskretisierung elliptischer Optimalsteuerungsaufgaben	57
§11 Das Ritz-Galerkin-Verfahren	57
§12 Die Finite-Elemente-Methode und die Diskretisierung der linearer elliptischer PDEs	60
§13 Fehlerabschätzung für die Zustandsgleichung	69
§14 Diskretisierung des Optimalsteuerungsproblems	74
§14.1 Die variationelle Diskretisierung und der Diskretisierungsfehler	76
§14.2 Die Voll-Diskretisierung und der Diskretisierungsfehler	81

VI Zusammenfassung und Analyse	92
Literaturverzeichnis	95
Index	98

Abbildungsverzeichnis

Abb. 1: Beispiel für ein zweidimensionales Gebiet Ω mit Rand Γ , verteilter Steuerung u und Normalenvektor n	3
Abb. 2: Grafische Darstellung der Idee lokaler Koordinatensysteme.	11
Abb. 3: Beispiele für Lipschitzgebiete.	12
Abb. 4: Beispiele für die Verletzung von Lipschitzgebietseigenschaften.	12
Abb. 5: Stückweise lineare Funktion und ihre schwache Ableitung.	14
Abb. 6: Grobe und feine Zerlegung eines zweidimensionalen Gebiets.	61
Abb. 7: Lineare Ansatzfunktionen bei uniformer Triangulierung eines eindimensionalen Gebiets.	63
Abb. 8: Pyramidenfunktion auf gleichmäßig trianguliertem Einheitsquadrat.	64
Abb. 9: Spitzes Dreieck T mit Durchmesser $\rho(T)$ und Innendurchmesser $R(T)$	65
Abb. 10: Beispiel einer unzulässigen Zerlegung des Einheitsquadrats.	65
Abb. 11: Basisdarstellung für die Approximation durch stückweise konstante Funktionen im uniform triangulierten eindimensionalen Fall.	82

I Einführung und Motivation

In Naturwissenschaften und Technik werden zahlreiche praktisch relevante Prozesse durch partielle Differentialgleichungen (PDEs) beschrieben. Als Anwendungsgebiete seien etwa die Strömungsmechanik, Mikroelektronik, Kristallzüchtung, Gefäßchirurgie, Herzmedizin, das Verhalten quantenphysikalischer Teilchen und die für diese Arbeit in den Vordergrund gerückte Wärmeausbreitung genannt. Da Computer tendenziell immer leistungstärker werden, erweitert sich das Anwendungsgebiet in der Praxis kontinuierlich.

Zusätzlich zur numerischen Simulation zur Vorhersage oben genannter Prozesse, ist man etwa unter ökonomischen Gesichtspunkten an deren Optimierung interessiert, was folglich zu Optimierungsaufgaben mit PDEs führt. Stellt die dabei gesuchte Optimierungsgröße eine orts- und/oder zeitabhängige Funktion dar, die man im Prozess frei oder unter Einhaltung gewisser Grenzen wählen kann, so spricht man von einer Aufgabe der optimalen Steuerung.

Wie bereits angedeutet werden wir uns in dieser Arbeit vorrangig mit der Wärmeleitungsgleichung in stationärer, also zeitunabhängiger Form beschäftigen. Es handelt sich dabei um eine linear-elliptische PDE, was zu einem linear-quadratischen elliptischen Optimalsteuerungsproblem führen wird. Auch der Großteil der theoretischen Vorbereitung wird sich darauf beziehen. Für die numerische Lösung solcher Aufgaben ist eine Diskretisierung der Gleichungen und eventuell auch der Steuerung notwendig. Das endgültige Anliegen dieser Arbeit wird es sein, bei einer speziellen Diskretisierung, der Finiten-Elemente-Methode, den Diskretisierungsfehler abzuschätzen, womit man in der Praxis etwa die Möglichkeit schafft, a priori Diskretisierungen festzulegen, mit denen sich wiederum ein Optimalsteuerungsproblem mit vorgegebener Genauigkeit lösen lässt.

Auf dem Weg zur Approximationsfehleranalyse wird es einer der Hauptanliegen dieser Arbeit sein, notwendige und hinreichende Bedingungen für die optimale Lösung einer solchen Problemstellung herzuleiten, die dann in den theoretischen Überlegungen rund um den Diskretisierungsfehler berücksichtigt werden.

Bezüglich der für diese Arbeit verwendeten Literatur, seien besonders [Trö09], [Alt02], [Eva10] und [Bra07] herausgestellt. Der Aufbau dieser Arbeit spiegelt in großen Teilen die Vorgehensweise wieder, die sich in der aktuellen Forschungsliteratur, vgl. etwa [M⁺08], und in den hier herangezogenen Vorlesungsskripten, vgl. etwa [Mey11], [Her12], [Her11], [Oh108], findet.

§1 Die lineare Wärmeleitungsgleichung als Modellproblem

Wir folgen bei der Vorstellung unseres Modellproblems im Groben [Trö09, 1.2], sowie [Mey11, §1] und [Her12, §1.1].

Wir betrachten einen Körper, der durch eine Punktmenge $\Omega \subset \mathbb{R}^3$ beschrieben wird. Zur Vereinfachung treffen wir folgende Annahmen: Das Material ist isotrop und homogen, der Wärmeleitkoeffizient, eine temperaturabhängige Materialkonstante, sei $\kappa = 1$, zudem betrachten wir den stationären Fall.

Dann lautet die Gleichung der Wärmeleitung im linearen Fall:

$$-\Delta y(x) = u(x) \quad \forall x \in \Omega \quad (1.1)$$

mit $\Delta y(x) \stackrel{\text{def}}{=} \sum_{i=1}^3 \frac{\partial^2 y}{\partial x_i^2}$, dem sogenannten Laplace-Operator.

$y(x)$ bezeichne die Temperatur, $u(x)$ die Wärmeleitungsdichte an der Stelle x . Allgemein ist die Wärmeleitungsgleichung eine Folgerung aus dem ersten Hauptsatz der Thermodynamik.¹

I. A. hat eine solche Problemstellung unendlich viele Lösungen. Um die Temperatur im Körper aber eindeutig beschreiben zu können, benötigt man sogenannte Randbedingungen.

$\Gamma = \partial\Omega = \overline{\Omega} \setminus \text{int}(\Omega)$ sei der Rand von Ω . Die Bedingungen für die Temperatur auf Γ hängen von diversen physikalischen Gegebenheiten ab.

Für das weitere Vorgehen nehmen wir an, dass der Rand, etwa durch Eiswasser, konstant auf 0°C gehalten wird, also dass gilt:

$$y(x) = 0 \quad \forall x \in \Gamma. \quad (1.2)$$

Man spricht hier von einer homogenen Dirichlet-Randbedingung.

Zur Beschreibung andersartiger physikalischer Umstände sind auch Randbedingungen anderer Form denkbar. Es seien etwa die Neumann- oder Robin-Randbedingung genannt. Auch eine Kombination verschiedener Randbedingungen ist möglich.²

Wir beschränken uns in dieser Arbeit aber auf die Randbedingung (1.2) und erhalten zusammen mit (1.1) die sogenannte Poissonsgleichung.

Ist u bekannt, so kann die Temperaturverteilung im Körper für gewöhnlich numerisch anhand von (1.1) und (1.2) berechnet werden. Wir sind i. A. auf numerische Verfahren, die Approximationen der exakten Lösung liefern, angewiesen, da es oft schwierig oder gar unmöglich ist, Lösungen solcher partieller Differentialgleichungen analytisch in geschlossener Form anzugeben.

In der Praxis ist es jedoch so, dass u innerhalb gewisser Grenzen gewählt werden kann, um den Prozess zu steuern. Dies geschieht etwa durch Mikrowellenstrahler oder Heizkörper. u wird daher als Steuerung des Problems bezeichnet. y , als Lösung von (1.1) und (1.2), ist eine von

¹Siehe [Mes04, 5.2.3 und 5.4].

²Siehe dazu z. B. [Gro94, 2.1].

u abhängige Größe und wird als Zustand bezeichnet. Demzufolge nennt man die PDE selbst Zustandsgleichung.

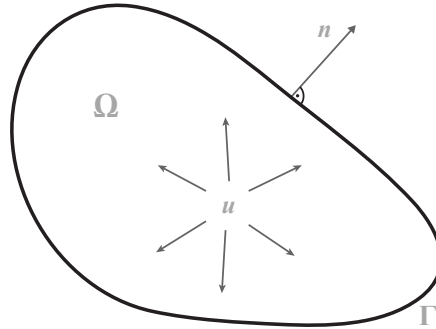


Abbildung 1: Beispiel für ein zweidimensionales Gebiet Ω mit Rand Γ , verteilter Steuerung u und Normalenvektor n .³

Für die Güte einer Wahl von u ist eine Bewertung der Steuerung und der dadurch erzeugten Temperaturverteilung y nötig. Diese geschieht durch ein Zielfunktional, wobei es in der Praxis üblich ist, dass kleinere Werte für eine bessere Steuerung stehen. Ein Beispiel für ein solches Zielfunktional wäre etwa

$$J(y, u) = \frac{1}{2} \int_{\Omega} |y(x) - y_d(x)|^2 dx. \quad (1.3)$$

y_d steht für die gewünschte Temperaturverteilung, an die möglichst nahe herangeführt werden soll.

Die Abweichung $|y - y_d|$ wird quadriert, weil wir im Folgenden nur differenzierbare Funktionale betrachten werden. Der Faktor $\frac{1}{2}$ kompensiert den Faktor 2, der später bei der Ableitung entstehen wird. Ein Zielfunktional der Form (1.3) genügt ohne weitere Bedingungen i. A. nicht um eine wohlgestellte Aufgabe zu erhalten. Man könnte bei (1.3) das vorgegebene Temperaturprofil y_d umso besser annähern, je größere Ausschläge die Steuerung u besitzt. Eine optimale Steuerung würde hier also unendlich große (sowohl positive als auch negative) Werte annehmen. Für (1.3) würde also keine optimale Steuerung existieren, sondern nur eine unbeschränkte Folge von Steuerungen $\{u_n\}$, deren Zielfunktionswerte $J(y_n, u_n)$ lediglich gegen das Infimum aller möglichen Zielfunktionswerte konvergieren. Optimale Steuerungen sind i. A. also nicht beschränkt. Wir müssen also dafür sorgen, dass dies in einer geeigneten Norm der Fall ist. Dazu führen wir zunächst eine punktweise Steuerbeschränkung ein. $u(x)$ muss folgendes erfüllen:

$$u_a(x) \leq u(x) \leq u_b(x) \quad \forall x \in \Omega.$$

³Vgl. [Trö09, S.4].

Dies wird etwa durch die Leistungsdichte von Mikrowellenstrahlern, oder die Möglichkeit nur bis zu einer bestimmten Temperatur zu heizen oder zu kühlen, motiviert.

Ebenso ist die Steuerung mit gewissen Kosten, etwa dem Verbrauch elektrischer Energie, verbunden. Daher wird das Zielfunktional um einen sogenannten Kontrollkostenterm erweitert:

$$J(y, u) = \frac{1}{2} \int_{\Omega} |y(x) - y_d(x)|^2 dx + \frac{\lambda}{2} \int_{\Omega} |u(x)|^2 dx. \quad (1.4)$$

$\lambda > 0$ nennt man den Kontrollkostenparameter. Er beschreibt eine relative Gewichtung der beiden Anteile im Zielfunktional. Wie wir noch sehen werden, führt das Funktional für $\lambda > 0$ auf ein wohl gestelltes Problem.

Die gesamten Vorüberlegungen führen uns nun zu einer Optimalsteuerungsaufgabe, unserem Modellproblem der optimalen Temperaturquelle:

$$\left. \begin{array}{l} \text{Minimiere } J(y, u) = \frac{1}{2} \int_{\Omega} |y(x) - y_d(x)|^2 dx + \frac{\lambda}{2} \int_{\Omega} |u(x)|^2 dx \\ \text{unter } \left\{ \begin{array}{l} -\Delta y(x) = u(x) \quad \forall x \in \Omega \\ y(x) = 0 \quad \forall x \in \Gamma \end{array} \right. \\ \text{und } u_a(x) \leq u(x) \leq u_b(x) \quad \forall x \in \Omega. \end{array} \right\} \quad (1.5)$$

Bei (1.5) spricht man von einer Aufgabe mit verteilter Steuerung, da $u(x)$ auf ganz Ω wirkt.

Um eine Aufgabenstellung wie (1.5) zu lösen und auch diverse Grundlagen zur Optimalsteuerung von PDEs zu erläutern und herzuleiten, gehen wir folgendermaßen vor:

- Diskussion der Zustandsgleichung im schwachen Sinne (siehe Kapitel II).
- Beweis der Existenz und Eindeutigkeit einer optimalen Lösung bzw. Steuerung (siehe Kapitel III).
- Herleitung von notwendigen und hinreichenden Optimalitätsbedingungen (siehe Kapitel IV).
- Methoden zur numerischen Lösung und Diskretisierung des Problems, sowie der Bestimmung des Diskretisierungsfehlers für die spezielle Diskretisierungsmethode der Finiten-Elemente-Methode (siehe Kapitel V).

Wie bereits angedeutet, beschäftigt sich diese Arbeit im Großen und Ganzen nur mit linearen elliptischen PDEs. Zahlreiche physikalische Prozesse sind nicht stationär und führen zu parabolischen oder hyperbolischen, oft auch nichtlinearen (etwa semi- oder quasilinearen) PDEs. Es treten auch Fälle auf, in denen die Steuerung nicht als verteilte Inhomogenität wie in (1.5) auftritt, sondern in den Randbedingungen oder als Koeffizienten im Differentialoperator. Manchmal

tritt sogar Ω selbst als Steuerung, wie etwa bei der Konstruktion von Flugzeugflügeln, auf. Dies führt zu sogenannten Shape- oder Formoptimierungsproblemen, welche sehr viel aufwändiger zu lösen sind. Viele Ergebnisse die wir im Laufe der Arbeit erhalten werden, sind auf solche Probleme nicht anwendbar, hierfür sei etwa auf [Trö09] und [H⁺09] sowie für theoretische Hintergründe auf [Lue69] hingewiesen.

Wir fassen also zusammen: Unser Modellproblem sowie das gesamte Vorgehen beruhen darauf, dass wir lineare elliptische PDEs betrachten. Das Zielfunktional ist quadratisch, die Steuerung ist auf dem ganzen Gebiet definiert. Dies führt wie im Falle (1.5) auf ein linear-quadratisches elliptisches Optimalsteuerungsproblem mit verteilter Steuerung.

II Schwache Lösungen linearer elliptischer PDEs

Wir beschränken uns zunächst bei der Diskussion von (1.5) auf die Zustandsgleichung, also die PDE (1.1) und (1.2). Wir beginnen mit einer Vorstellung von funktionalanalytischen Grundlagen, die im Folgenden nötig sind. Auch in den folgenden Kapiteln werden wir so verfahren. In diesem Kapitel werden wir im Groben [Trö09, 2.1 bis 2.4] und der dort genannten Literatur, sowie [Mey11, §3 bis §6], [Her12, §3 bis §5] und [Her11, §1 bis §4] folgen.

Alle betrachteten Räume seien von nun an über \mathbb{R} definiert.

§2 Normierte lineare Räume

Definition 2.1

Es sei X ein linearer Raum über \mathbb{R} . Eine Abbildung $\|\cdot\| : X \rightarrow \mathbb{R}$ heißt Norm in X , wenn folgende Eigenschaften $\forall x, y \in X$ und $\forall \lambda \in \mathbb{R}$ erfüllt sind:

- (i) $\|x\| \geq 0$ und $\|x\| = 0 \Leftrightarrow x = 0$.
- (ii) $\|x + y\| \leq \|x\| + \|y\|$ (Dreiecksungleichung).
- (iii) $\|\lambda x\| = |\lambda| \cdot \|x\|$ (positive Homogenität).

Ist $\|\cdot\|$ eine Norm in X , dann heißt $\{X, \|\cdot\|\}$ (reeller) normierter linearer Raum. □

Definition 2.2

Eine Folge $\{x_n\}_{n=1}^{\infty}$ von Elementen eines normierten Raumes $\{X, \|\cdot\|\}$ heißt konvergent, wenn ein Grenzelement $x \in X$ existiert, s.d.

$$\lim_{n \rightarrow \infty} \|x_n - x\| = 0.^4$$

Sie heißt Cauchyfolge, wenn es zu jedem $\varepsilon > 0$ eine natürliche Zahl $N = N(\varepsilon)$ gibt mit

$$\|x_n - x_m\| \leq \varepsilon \quad \forall n, m > N(\varepsilon).$$

Ein normierter linearer Raum $\{X, \|\cdot\|\}$ heißt vollständig, wenn in X jede Cauchyfolge konvergiert, also ein Grenzelement in X besitzt. Einen solchen Raum nennt man Banachraum. □

⁴Im Folgenden schreiben wir zur Vereinfachung auch oft $\lim_{n \rightarrow \infty} x_n = x$ oder $x_n \rightarrow x$.

Bemerkung:

Beachte zum letzten Teil obiger Def. 2.2, dass jede konvergente Folge auch eine Cauchyfolge ist, die Umkehrung gilt i. A. natürlich nicht.⁵

Definition 2.3

Ein linearer Raum H über \mathbb{R} heißt Innenproduktraum oder Prä-Hilbertraum, wenn eine Abbildung $H \times H \ni (u, v) \mapsto (u, v)_H \in \mathbb{R}$ mit folgenden Eigenschaften existiert:

- (i) $(u, u)_H \geq 0 \ \forall u \in H$ und $(u, u)_H = 0 \Leftrightarrow u = 0$.
- (ii) $(u, v)_H = (v, u)_H \ \forall u, v \in H$.
- (iii) $(u_1 + u_2, v)_H = (u_1, v)_H + (u_2, v)_H \ \forall u_1, u_2, v \in H$.
- (iv) $(\lambda u, v)_H = \lambda(u, v)_H \ \forall u, v \in H, \forall \lambda \in \mathbb{R}$.

Eine solche Abbildung bezeichnet man als Skalarprodukt. Sie erzeugt mittels

$$\|u\| \stackrel{def}{=} \sqrt{(u, u)_H} \tag{2.1}$$

eine Norm auf H . Wir werden von nun an stets diese Norm auf einem Innenproduktraum verwenden. Ebenso gilt die Cauchy-Schwarz'sche Ungleichung, welche einen Spezialfall der Hölder'schen Ungleichung aus noch folgendem Satz 3.1 darstellt:

$$|(u, v)_H| \leq \|u\| \cdot \|v\| \ \forall u, v \in H.$$

Bezüglich der letzten beiden Aussagen, betrachte etwa [Alt02, 0.2].

Ein Innenproduktraum heißt Hilbertraum, wenn er vollständig bezüglich der Norm (2.1) ist. \square

Bemerkung:

Im Folgenden werden wir den Index H weglassen, wenn klar ist in welchem Raum wir uns befinden. Ebenso werden wir einige wichtige Beispiele für normierte lineare Räume, Banach- und Hilberträume genauer betrachten.

§3 Lebesgue- und Sobolevräume

Nun werden noch Grundbegriffe aus der Theorie der Räume integrierbarer Funktionen und Sobolevräume bereitgestellt.

Von nun an sei $E \subset \mathbb{R}^N$ eine nichtleere, beschränkte und Lebesgue-messbare Menge, also mit $|E| < \infty$ bezüglich des N -dimensionalen Lebesgue-Maßes.⁶

⁵Betrachte dazu den Satz von Bolzano-Weierstraß [Kön04, 1.1].

⁶Siehe dazu etwa [Alt02, 1.7, 1.8 und A 1.2] und [Ada75, 1.32 bis 1.39].

§3.1 Lebesgueräume

Definition 3.1

Unter dem Lebesgueräum $L^p(E)$, $1 \leq p < \infty$ versteht man den Raum aller (Äquivalenzklassen von) messbaren Funktionen $y : E \rightarrow \mathbb{R}$ mit der Eigenschaft

$$\int_E |y(x)|^p dx < \infty.$$

$L^p(E)$ wird versehen mit der Norm

$$\|y\|_{L^p(E)} \stackrel{\text{def}}{=} \left(\int_E |y(x)|^p dx \right)^{1/p}.$$

□

$L^p(E)$ beschreibt also den Raum aller Äquivalenzklassen von Funktionen, s.d. ein beliebiger Repräsentant der Äquivalenzklasse zur p -ten Potenz integrierbar ist. Zwei Funktionen, die im Sinne des Lebesgue-Maßes auf E fast überall gleich sind, liegen in der gleichen Äquivalenzklasse.⁷

Definition 3.2

Mit dem Lebesgueräum $L^\infty(E)$ wird der Raum aller (Äquivalenzklassen von) fast überall gleichmäßig beschränkten und messbaren Funktionen $y : E \rightarrow \mathbb{R}$ bezeichnet. Er wird versehen mit der Norm

$$\|y\|_{L^\infty(E)} \stackrel{\text{def}}{=} \text{ess sup}_{x \in E} |y(x)| \stackrel{\text{def}}{=} \inf_{|F|=0} \left(\sup_{x \in E \setminus F} |y(x)| \right).$$

□

ess sup beschreibt also das essentielle bzw. wirkliche Supremum einer Funktion. Es werden etwa Maxima ausgeschlossen, die sich bereits durch Löschung einzelner, isolierter Punkte ändern und damit unwesentlich sind.

Als Beispiel betrachte man etwa die Funktion $y : [0, 1] \rightarrow \mathbb{R}$, die auf $[0, 1] \setminus \{0\}$ den Wert 0 und im Nullpunkt den Wert 1 annimmt. Ihr Maximum beträgt 1, ihr ess sup ist 0.

Für L^p -Räume, sowie unsere weiteren theoretischen Überlegungen, spielt die bereits erwähnte Hölder'sche Ungleichung eine ausgezeichnete Rolle:

Satz 3.1 (Hölder'sche Ungleichung)

Es sei $f \in L^p(E)$ und $g \in L^q(E)$, wobei p und q konjugierte Exponenten⁸ sind, d.h. $1 \leq p, q \leq \infty$ und $\frac{1}{p} + \frac{1}{q} = 1$. Dann ist das Produkt $fg \in L^1(E)$ und es gilt:

$$\int_E |fg| dx \leq \left(\int_E |f|^p dx \right)^{1/p} \left(\int_E |g|^q dx \right)^{1/q}$$

oder gemäß unserer Normdefinition aus Def. 3.1: $\|fg\|_{L^1(E)} \leq \|f\|_{L^p(E)} \|g\|_{L^q(E)}$.

□

⁷Siehe dazu [Alt02, 1.12].

⁸Im Grenzfall $p = \infty$ und $q = 1$ ist $\frac{1}{p}$ als Null zu verstehen.

BEWEIS Siehe [Alt02, 1.14]. ■

Bemerkung:

Die Räume $L^p(E)$ sind Banachräume für $1 \leq p \leq \infty$. Siehe dazu [Alt02, 1.13] und den Satz von Fischer-Riesz ([Alt02, 1.17]). Der Beweis dazu ist allerdings nicht sehr einfach und erfordert die Anwendung diverser funktionalanalytischer Grundlagen wie etwa die Hölder'sche Ungleichung (3.1), das Majoranten-Kriterium und die Minkowski-Ungleichung.⁹

Das Skalarprodukt

$$(u, v)_{L^2(E)} \stackrel{def}{=} \int_E u(x)v(x)dx$$

erzeugt auf $L^2(E)$ die oben definierte Norm, weshalb $L^2(E)$ sogar ein Hilbertraum nach Def. 2.3 ist.

Wie man bereits erahnen kann, wird uns obige Normdefinition noch nützlich sein, um das Modellproblem der optimalen Temeperaturquelle (1.5) kompakter zu schreiben.

Im weiteren Verlauf sei $\Omega \subset \mathbb{R}^N$, $N \geq 1$ eine beschränkte offene zusammenhängende Menge. Wir sagen dazu beschränktes Gebiet. $\bar{\Omega}$ sei der Abschluss, $\Gamma = \partial\Omega = \bar{\Omega} \setminus \text{int}(\Omega)$ der Rand von Ω .

Definition 3.3

Die Menge $\text{supp } v \stackrel{def}{=} \overline{\{x \in \Omega : v(x) \neq 0\}}$ heißt Träger, oder auch patch von v .

$C_0^k(\Omega)$, $0 \leq k \leq \infty$ ist die Menge aller k -mal stetig differenzierbaren Funktionen $v : \Omega \rightarrow \mathbb{R}$ mit kompaktem Träger in Ω . □

Für uns wird im Folgenden der Fall $k = \infty$, d.h. die Menge $C_0^\infty(\Omega)$, von besonderem Interesse sein. Funktionen aus $C_0^\infty(\Omega)$ haben auf Γ den Wert 0, s.d. die Randintegrale bei partieller Integration verschwinden. Dass sie zusätzlich nach Belieben differenzierbar sind, werden wir später bei der Einführung von Sobolevräumen benutzen. Mit $C^k(\Omega)$ bezeichnen wir von nun an die Menge der k -mal in Ω stetig differenzierbaren reellwertigen Funktionen. Wir benutzen hier den Begriff 'Menge' statt 'Raum' da wir an Topologisierungen nicht interessiert sind.

Definition 3.4

Einen Vektor nichtnegativer Zahlen

$$\alpha = (\alpha_1, \dots, \alpha_N) \in (\mathbb{N} \cup \{0\})^N$$

wird als Multiindex bezeichnet. Im Symbol D^α geben die Koordinaten von α an, wie oft nach welcher Raumkoordinate differenziert wird. Da wir lediglich hinreichend glatte Funktionen betrachten, ist die Reihenfolge der Differentiation nach dem Satz von Schwarz¹⁰ unwichtig. Die Ordnung von α wird als $|\alpha| = \sum_{i=1}^N \alpha_i$ definiert. Insbesondere gilt $D^{0, \dots, 0} y \stackrel{def}{=} y$. □

⁹Siehe [Alt02, 1.14 bis 1.16].

¹⁰Siehe [Kön04, 2.3].

Definition 3.5

$\Omega \in \mathbb{R}^N$ sei beschränkt. $C(\overline{\Omega})$ sei der lineare Raum aller auf $\overline{\Omega}$ stetigen reellwertigen Funktionen, $C^k(\overline{\Omega})$, $k \in \mathbb{N}$, der lineare Raum aller Funktionen aus $C^k(\Omega)$, die mit ihren partiellen Ableitungen bis zur Ordnung k stetig auf $\overline{\Omega}$ fortgesetzt werden können.

Werden beide versehen mit den Normen

$$\|y\|_{C(\overline{\Omega})} \stackrel{def}{=} \max_{x \in \overline{\Omega}} |y(x)| \quad \text{und} \quad \|y\|_{C^k(\overline{\Omega})} \stackrel{def}{=} \sum_{|\alpha| \leq k} \|D^\alpha y\|_{C(\overline{\Omega})},$$

so sind dies Banachräume.¹¹ □

Mit den hier eingeführten Räumen können wir nun das sogenannte Fundamentallemma der Variationsrechnung, welches wir später noch benötigen werden, betrachten:

Lemma 3.1 (Fundamentallemma der Variationsrechnung)

Sei $\Omega \subset \mathbb{R}^N$ wie beschrieben, $g \in L^1(\Omega)$, dann gilt:

$$\int_{\Omega} gv \, dx = 0 \quad \forall v \in C_0^\infty(\Omega) \Leftrightarrow g(x) = 0 \text{ fast überall in } \Omega. \quad \square$$

BEWEIS Siehe [Alt02, 2.21]. Der Beweis erfordert Kenntnis über die Faltung von Funktionen¹² sowie Dirac-Folgen¹³. ■

§3.2 Reguläre Gebiete

Eine Diskussion von PDEs wie wir sie vornehmen, erfordert Ortsgebiete Ω mit hinreichend glattem Rand. Wir betrachten dazu folgende Definition und verweisen auf [Alt02, A 6.2], [Gri85, 2.1] und v.a. [Wlo82, 2] für eine ausführliche Betrachtung dieser Klasse von Gebieten. Für unsere Zwecke genügt uns allerdings:

Definition 3.6

Es sei $\Omega \subset \mathbb{R}^N$, $N \geq 2$, ein beschränktes Gebiet mit Rand Γ . Das Gebiet Ω bzw. der Rand Γ gehören zur Klasse $C^{k,1}$, $k \in \mathbb{N} \cup \{0\}$, wenn endlich viele Koordinatensysteme S_1, \dots, S_M , Funktionen h_1, \dots, h_M , sowie Konstanten $a, b > 0$ existieren, s.d. folgende Eigenschaften erfüllt sind:

- (i) Alle h_i sind auf dem $(N - 1)$ -dimensionalen abgeschlossenen Würfel

$$\overline{Q}_{N-1} = \{y = (y_1, \dots, y_{N-1}) : |y_i| \leq a, i = 1, \dots, N - 1\}$$

k -mal differenzierbar mit Lipschitz-stetigen Ableitungen der Ordnung k .¹⁴

¹¹ Siehe [Alt02, 1.5].

¹² Siehe dazu [Alt02, 2.12].

¹³ Siehe dazu [Alt02, 2.13 und 2.14].

¹⁴ Zur Lipschitz-Stetigkeit siehe [Alt02, 1.6] und [Wlo82, 1.1]. Es werden die Begriffe der Metrik und metrischer Räume benötigt, siehe dazu [Alt02, 0.6]. Der Übersicht halber wird hier darauf verzichtet.

(ii) $\forall P \in \Gamma \exists i \in \{1, \dots, M\}$ s.d. P im Koordinatensystem S_i die Darstellung $P = (y, h_i(y))$, $y \in Q_{N-1}$ hat.

(iii) In S_i gilt:

$$\begin{aligned} (y, y_N) \in \Omega &\Leftrightarrow y \in \overline{Q_{N-1}}, h_i(y) < y_N < h_i(y) + b \\ (y, y_N) \notin \Omega &\Leftrightarrow y \in \overline{Q_{N-1}}, h_i(y) - b < y_N < h_i(y). \end{aligned}$$

□

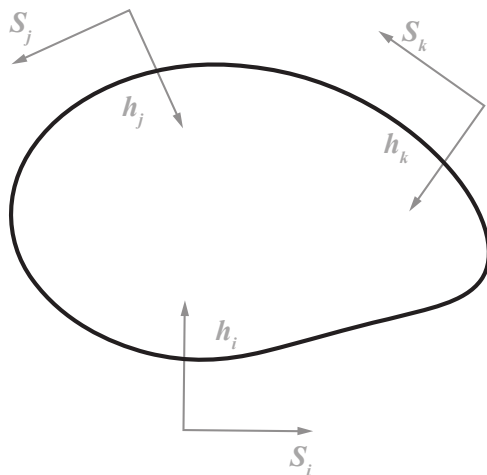


Abbildung 2: Grafische Darstellung der Idee lokaler Koordinatensysteme.¹⁵

Gebiete bzw. Ränder der Klasse $C^{0,1}$ heißen Lipschitz- oder reguläres Gebiet bzw. Lipschitzrand. Diese werden in unserem weiteren Vorgehen genügen. Für $N = 1$ ist $\Omega \subset \mathbb{R}$ notwendiger Weise ein offenes Intervall, s.d. sich die Frage der Glattheit hier gar nicht stellt.

Wir nehmen also mit:

- Γ kann in endlich viele Stücke Γ_i zerlegt werden, s.d. jedes Γ_i der Graph einer stetigen Funktion h_i in einem lokalen Koordinatensystem S_i ist.
- Die Ableitungen der h_i sind einschließlich zur Ordnung k Lipschitz-stetig.
- Das Gebiet liegt lokal nur auf einer Seite des Randes.

¹⁵Vgl. [Mey11, S.21].

Zur Veranschaulichung betrachten wir folgende Beispiele:

Lipschitzgebiete sind:

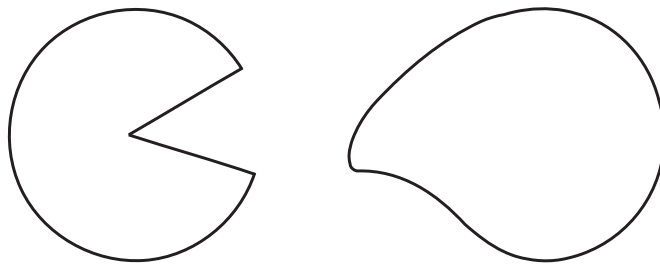


Abbildung 3: Beispiele für Lipschitzgebiete.¹⁶

Dagegen sind keine Lipschitzgebiete:

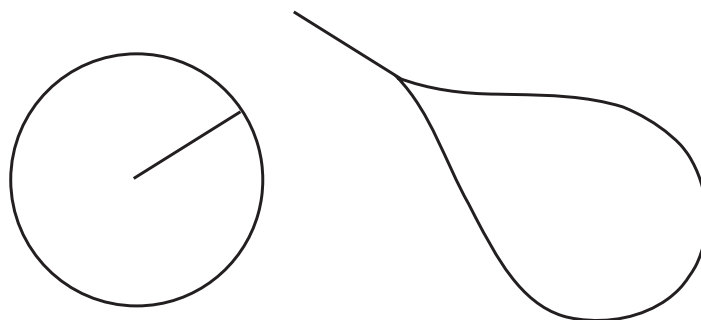


Abbildung 4: Beispiele für die Verletzung von Lipschitzgebietseigenschaften.¹⁷

Bemerkung:

Dadurch, dass man die Γ_i mit Graphen einer Funktion identifiziert, ergibt sich in natürlicher Weise das $(N - 1)$ -dimensionale Lebesgue-Maß auf dem Rand. Im Folgenden werden wir die Integration bezüglich des Randmaßes zur Verdeutlichung mit ds statt dx bezeichnen.

¹⁶Vgl. [Her12, S.24].

¹⁷Vgl. [Her12, S.24].

§3.3 Schwache Ableitung und Sobolevräume

Nun geht es darum, den Begriff der schwachen Differenzierbarkeit von Funktionen, sowie geeignete Funktionenräume für die Lösungen von PDEs einzuführen, die schwächere Lösungen als die klassischen erlauben. Diese besitzen dann eine geringere Regularität. Der Raum, indem man die Lösung sucht, wird dadurch, ebenso wie die Chance sie zu finden, bedeutend größer. Dies geschieht durch die Vervollständigung klassischer Funktionenräume, in unserem Fall $L^2(\Omega)$, bezüglich von Integralsnormen. Es handelt sich dabei um die sogenannten Sobolevräume.

In beschränkten Lipschitzgebieten Ω kann der Satz von Gauß¹⁸ angewendet werden. Als direkte Folgerung daraus gilt für $y, v \in C^1(\bar{\Omega})$ die Formel der partiellen Integration¹⁹:

$$\int_{\Omega} v(x) D_i y(x) dx = \int_{\Gamma} v(x) y(x) n_i(x) ds - \int_{\Omega} y(x) D_i v(x) dx \quad (3.1)$$

wobei gilt:

- $n_i(x)$ ist die i -te Komponente des nach außen gerichteten Normaleneinheitsvektors $n(x)$ in $x \in \Gamma$ (siehe auch Abbildung 1)
- $D_i = \frac{\partial}{\partial x_i} = D^{0, \dots, 0, 1, 0, \dots, 0}$ (mit 1 an der i -ten Stelle)

Bemerkung:

Wie bereits angedeutet ist das Integral über Γ als Weg- bzw. Oberflächenintegral zu verstehen. Außerdem existiert $n(x)$ für Lipschitzgebiete fast überall, da Lipschitz-stetige Funktionen fast überall differenzierbar sind.²⁰

Für $v = 0$ auf Γ folgt aus (3.1) sofort

$$\int_{\Omega} v(x) D_i y(x) dx = - \int_{\Omega} y(x) D_i v(x) dx,$$

Der Randterm verschwindet also.

Allgemeiner ergibt sich für $y \in C^k(\bar{\Omega})$, $v \in C_0^k(\Omega)$ sowie Multiindizes α mit $|\alpha| \leq k$ nach mehrmaliger partieller Integration

$$\int_{\Omega} y(x) D^\alpha v(x) dx = (-1)^{|\alpha|} \int_{\Omega} v(x) D^\alpha y(x) dx. \quad (3.2)$$

Wir verallgemeinern nun den Ableitungsbegriff auf eine sehr große Funktionenklasse:

Im Folgenden bezeichne $L_{loc}^1(\Omega)$ die Menge der Funktionen $v : \Omega \rightarrow \mathbb{R}$, die für jede kompakte Teilmenge $K \subset \Omega$ Lebesgue-integrierbar sind.²¹

¹⁸Siehe [Alt02, A 6.8].

¹⁹Diese wird in der Literatur auch oft als Green'sche Formel bezeichnet.

²⁰Satz von Rademacher, siehe [Eva10, 5.8 Theorem 6].

²¹Es gilt also $v \in L^1(K)$.

Definition 3.7

Es sei $y \in L^1_{loc}(\Omega)$, α ein Multiindex. Existiert ein $w \in L^1_{loc}(\Omega)$ mit der Eigenschaft

$$\int_{\Omega} y(x) D^{\alpha} v(x) dx = (-1)^{|\alpha|} \int_{\Omega} v(x) w(x) dx \quad \forall v \in C_0^{\infty}(\Omega), \quad (3.3)$$

so heißt w schwache Ableitung oder Ableitung im Distributionensinn der Ordnung α von y . Man schreibt $w = D^{\alpha} y$.

Beide Integrale existieren, da $v \in C_0^{\infty}(\Omega)$. v und $D^{\alpha} v$ haben also kompakte Träger in Ω . □

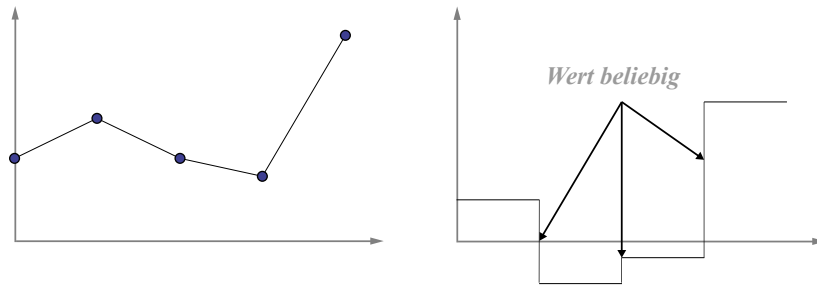


Abbildung 5: Stückweise lineare Funktion und ihre schwache Ableitung.²²

Der Vergleich von (3.3) und (3.2) macht deutlich, dass w also die schwache Ableitung von y ist, wenn w an der Stelle der klassischen Ableitung $D^{\alpha} y$ die Formel der partiellen Ableitung erfüllt. Zur Veranschaulichung unseres neuerworbenen Ableitungsbegriffes betrachten wir ein einfaches Beispiel:²³

Beispiel 3.1

$y(x) = |x|$ in $\Omega = (-1, 1)$ (Lipschitzgebiet nach Def.3.6) besitzt mit

$$y'(x) = w(x) = \begin{cases} -1 & \forall x \in (-1, 0) \\ 1 & \forall x \in (0, 1) \end{cases}$$

eine schwache Ableitung erster Ordnung, jedoch keine der Ordnung 2 in $L^1_{loc}(\Omega)$. □

²²Vgl. [Bas08, S.24].

²³Siehe auch [Trö09, S.22] oder [Mey11, Beispiel 4.9].

BEWEIS Sei $v \in C_0^\infty(-1, 1)$ beliebig. Mittels partieller Integration auf $(-1, 0)$ bzw $(0, 1)$ erhält man

$$\begin{aligned} \int_{\Omega} y(x)v'(x)dx &= \int_{-1}^1 |x|v'(x)dx \\ &= \int_{-1}^0 (-x)v'(x)dx + \int_0^1 xv'(x)dx \\ &= \int_{-1}^0 -v(x)dx + \int_0^1 v(x)dx \\ &= -\int_{\Omega} w(x)v(x)dx. \end{aligned}$$

w erfüllt also (3.3) und es gilt $w \in L_{loc}^1(-1, 1)$, eigentlich sogar $w \in L^\infty(-1, 1)$. Der erste Teil des Beispiels ist also bewiesen. Integriert man erneut partiell bezüglich $(-1, 0)$ bzw. $(0, 1)$, so ergibt sich

$$\begin{aligned} \int_{\Omega} w(x)v(x)dx &= \int_{-1}^0 (-1)v(x)dx + \int_0^1 v(x)dx \\ &= v(-1) - v(0) - v(0) + v(1) \\ &= -2v(0), \end{aligned}$$

da v kompakten Träger besitzt. Es ist also nicht möglich eine Funktion zu finden, s.d. (3.3) mit w anstatt y erfüllt ist. Die Betragsfunktion ist also nicht zweimal schwach differenzierbar. ■

Obiges Beispiel hat also gezeigt, dass eine L_{loc}^1 -Funktion (in unserem Falle y') keine schwache Ableitung besitzen muss, andererseits aber eine schwache Ableitung auch 'handlicheren' Räumen angehören kann als $L_{loc}^1(\Omega)$ (in unserem Fall $y' \in L^\infty(\Omega)$).²⁴

Definition 3.8

Seien $1 \leq p \leq \infty$ und $k \in \mathbb{N} \cup \{0\}$. Wir definieren

$$W^{k,p}(\Omega) \stackrel{def}{=} \{y \in L^p(\Omega) : D^\alpha y \in L^p(\Omega) \forall |\alpha| \leq k\}.$$

Für $1 \leq p < \infty$ versehen wir $W^{k,p}(\Omega)$ mit der Norm

$$\|y\|_{W^{k,p}(\Omega)} \stackrel{def}{=} \left(\sum_{|\alpha| \leq k} \int_{\Omega} |D^\alpha y(x)|^p dx \right)^{1/p} = \left(\sum_{|\alpha| \leq k} \|D^\alpha y\|_{L^p(\Omega)}^p \right)^{1/p}.$$

Für $p = \infty$ definieren wir die Norm

$$\|y\|_{W^{k,\infty}(\Omega)} \stackrel{def}{=} \max_{|\alpha| \leq k} \|D^\alpha y\|_{L^\infty(\Omega)}.$$

Diese Räume werden als Sobolewräume bezeichnet und spielen in unseren weiteren theoretischen Überlegungen eine entscheidende Rolle. Dass die Menge $W^{k,p}(\Omega)$ ein Vektorraum ist,

²⁴Siehe auch [Mey11, S.23] oder [Her11, S.11].

folgt sofort aus der Vektorraumeigenschaft von $L^p(\Omega)$, welche durch die Minkowski-Ungleichung gewährleistet ist und der Linearität der Ableitungsoperatoren D^α . Beachte, dass $W^{0,p}(\Omega) = L^p(\Omega)$ gilt. Außerdem sind alle Sobolewräume $W^{k,\infty}(\Omega)$ Banachräume.²⁵ \square

Im weiteren Verlauf wird für uns der Fall $p = 2$ von besonderem Interesse sein. In diesem Fall setzen wir

$$H^k(\Omega) \stackrel{\text{def}}{=} W^{k,2}(\Omega).$$

Der von nun an vorrangig benutzte Raum $H^1(\Omega)$ wird also definiert durch

$$H^1(\Omega) \stackrel{\text{def}}{=} \{y \in L^2(\Omega) : D_i y \in L^2(\Omega), i = 1, \dots, N\}$$

und wird versehen mit der Norm

$$\|y\|_{H^1(\Omega)} = \left(\int_{\Omega} |y|^2 + |\nabla y|^2 dx \right)^{1/2} = \left(\|y\|_{L^2(\Omega)}^2 + \|\nabla y\|_{L^2(\Omega)^N}^2 \right)^{1/2}$$

mit $|\nabla y|^2 = (D_1 y)^2 + \dots + (D_N y)^2$. Den Anteil $\|\nabla y\|_{L^2(\Omega)^N}$ bezeichnet man auch als H^1 -Seminorm²⁶. Durch die Einführung des Skalarprodukts

$$(u, v)_{H^1(\Omega)} = \int_{\Omega} uv \, dx + \int_{\Omega} \nabla u \cdot \nabla v \, dx \tag{3.4}$$

wird $H^1(\Omega)$ zu einem Hilbertraum.²⁷

Die Frage nach Randwerten von Funktionen in Sobolewräumen ist durchaus nicht trivial. Der Rand ist eine Menge vom Maß Null und Funktionen in Sobolewräumen sind auf Nullmengen nach Definition im Sinne von $L^p(\Omega)$ nicht unterscheidbar.

Definition 3.9

$W_0^{k,p}(\Omega)$ bezeichne den Abschluss der Menge $C_0^\infty(\Omega)$ bezüglich der Norm von $W^{k,p}(\Omega)$. Insbesondere setzen wir $H_0^k(\Omega) \stackrel{\text{def}}{=} W_0^{k,2}(\Omega)$. \square

Anmerkung 3.1

Damit ist $W_0^{k,p}(\Omega)$ ein abgeschlossener Unterraum des Banachraumes $W^{k,p}(\Omega)$, also selbst ein Banachraum mit der gleichen Norm. $H_0^1(\Omega)$ ist demzufolge ein Hilbertraum mit dem Skalarprodukt (3.4). Zudem ist $C_0^\infty(\Omega)$ dicht in jedem $W_0^{k,p}(\Omega)$. \square

²⁵Siehe [Alt02, 1.23] und [Ada75, 3.2].

²⁶Der einzige Unterschied zu einer Norm ist das Fehlen der Eigenschaft $\|y\| = 0 \Rightarrow y = 0$. Da konstante Funktionen $y \equiv c$ allerdings zu $H^1(\Omega)$ gehören und für diese $\nabla c \equiv 0$ gilt, ist dies hier aber der Fall.

²⁷Dies gilt bei obigem Vorgehen $\forall H^k(\Omega)$. Siehe dazu [Ada75, 3.5].

Funktionen aus $W_0^{k,p}(\Omega)$ können also als solche angesehen werden, bei denen die Randwerte aller Ableitungen bis zur Ordnung $k - 1$ verschwinden. Inhomogene Randwerte definiert man durch die sogenannte Spur von Funktionen aus $W^{k,p}(\Omega)$ auf dem Rand.²⁸ Dazu der folgende Satz:

Satz 3.2 (Spursatz)

Ω sei ein beschränktes Lipschitzgebiet und $1 \leq p \leq \infty$. Dann existiert eine lineare und stetige Abbildung $\tau : W^{1,p}(\Omega) \rightarrow L^p(\Gamma)$, die für alle $y \in W^{1,p}(\Omega) \cap C(\overline{\Omega})$ die Eigenschaft $(\tau y)(x) = y(x)$ fast überall auf Γ besitzt. □

BEWEIS Siehe [Alt02, A 6.6]. ■

Für stetige Funktionen fällt die Spurabbildung τy also mit den Randwerten $y|_{\Gamma}$ zusammen. Der sobolew'sche Einbettungssatz 4.3 wird für $p > N$ zeigen, dass Funktionen aus $W^{1,p}(\Omega)$ mit Funktionen aus $C(\overline{\Omega})$ identifiziert werden können. Daher ist τ in diesem Fall stetig von $W^{1,p}(\Omega)$ nach $C(\Gamma)$.²⁹

Definition 3.10

Das Element τy heißt Spur von y auf Γ , τ bezeichnet man auch als Spuroperator. □

Anmerkung 3.2

Die Stetigkeit des Spuroperators ist äquivalent zu seiner Beschränktheit, also zur Existenz einer Konstanten $c_{\tau} = c_{\tau}(\Omega, p)$, s.d. gilt:³⁰

$$\|y|_{\Gamma}\|_{L^p(\Gamma)} \leq c_{\tau} \|y\|_{W^{1,p}(\Omega)} \quad \forall y \in W^{1,p}(\Omega).$$

Für beschränkte Lipschitzgebiete gilt:

$$H_0^1(\Omega) = \{y \in H^1(\Omega) : y|_{\Gamma} = 0\}. \quad ^{31}$$

In $H_0^1(\Omega)$ führt man durch

$$\|y\|_{H_0^1(\Omega)} \stackrel{def}{=} \left(\int_{\Omega} |\nabla y|^2 dx \right)^{1/2}$$

eine Norm ein, die zur oben genannten in $H^1(\Omega)$ äquivalent ist. Mit $c_1, c_2 > 0$ gilt also:

$$c_1 \|y\|_{H_0^1(\Omega)} \leq \|y\|_{H^1(\Omega)} \leq c_2 \|y\|_{H_0^1(\Omega)} \quad \forall y \in H_0^1(\Omega), \text{ siehe auch den Beweis zu Satz 5.1.} \quad ^{32} \quad \square$$

²⁸Siehe auch [Trö09, S.23].

²⁹Siehe auch [Trö09, Satz 7.1] und [Ada75, V].

³⁰Siehe auch [Trö09, S.24].

³¹Siehe [Eva10, 5.5 Theorem 2].

³²Spezieller den Beweis zur V-Elliptizität unter Einbezug der Poincaré-Friedrich-Ungleichung aus Lemma 5.3.

§4 Stetige Lineare Operatoren und Funktionale

Nun stellen wir eine Übersicht über die letzten funktionalanalytischen Grundlagen zusammen, die wir als Rüstzeug benötigen, um die Frage nach der (eindeutigen) Lösbarkeit der Poisson-Gleichung zu diskutieren. Wie wir sehen werden, wird dies auch in einem allgemeineren Sinne geschehen.

U und V seien von nun an lineare normierte Räume über \mathbb{R} .

Definition 4.1

Eine Abbildung $A : U \rightarrow V$ heißt linearer Operator (oder auch linear) wenn $A(u + v) = A(u) + A(v)$ und $A(\lambda u) = \lambda A(u) \forall u, v \in U, \forall \lambda \in \mathbb{R}$ gilt. \square

Definition 4.2

Ein linearer Operator $A : U \rightarrow V$ heißt beschränkt, wenn ein $c \in \mathbb{R}$ existiert, s.d.

$$\|Au\|_V \leq c\|u\|_U \quad \forall u \in U$$

gilt. Die Zahl

$$\|A\| \stackrel{\text{def}}{=} \|A\|_{\mathcal{L}(U,V)} \stackrel{\text{def}}{=} \sup_{\|u\|_U=1} \|Au\|_V = \sup_{u \neq 0} \frac{\|Au\|_V}{\|u\|_U}$$

für gegebene Normen auf U und V heißt die OperatorNorm von A . \square

Satz 4.1

Ein linearer Operator ist beschränkt, genau dann wenn er stetig ist, d.h. wenn gilt: $u_n \rightarrow u \Rightarrow Au_n \rightarrow Au$. \square

BEWEIS Siehe [Alt02, 3.1]. \blacksquare

Definition 4.3

$\mathcal{L}(U, V)$ bezeichnet den Raum aller linearen und stetigen Operatoren von U nach V , versehen mit der obigen Operatornorm $\|\cdot\|_{\mathcal{L}(U,V)}$.³³

Es gilt, dass $\mathcal{L}(U, V)$ ein Banachraum ist, wenn V einer ist.³⁴ \square

Zu den bisher eingeführten Begriffen betrachten wir nun ein Beispiel:³⁵

Beispiel 4.1

Sei $z \in H^1(\Omega)$ und $A : H^1(\Omega) \rightarrow \mathbb{R}$ definiert durch

$$Ay = \int_{\Omega} \nabla z \cdot \nabla y \, dx.$$

³³Im Fall $U = V$ schreiben wir im Folgenden auch $\mathcal{L}(U)$.

³⁴Siehe [Alt02, 3.3].

³⁵Siehe auch [Mey11, Beispiel 5.5c) oder [Her12, Beispiel 4.2c)].

A ist beschränkt und es gilt:

$$\begin{aligned}
 |Ay| &= \left| \int_{\Omega} \nabla z \cdot \nabla y \, dx \right| = \left| \sum_{i=1}^N \int_{\Omega} \frac{\partial z}{\partial x_i} \frac{\partial y}{\partial x_i} \, dx \right| \\
 &\leq \sum_{i=1}^N \left| \int_{\Omega} \frac{\partial z}{\partial x_i} \frac{\partial y}{\partial x_i} \, dx \right| \quad (\text{Dreiecksungleichung}) \\
 &\leq \sum_{i=1}^N \left\| \frac{\partial z}{\partial x_i} \right\|_{L^2(\Omega)} \left\| \frac{\partial y}{\partial x_i} \right\|_{L^2(\Omega)} \quad (\text{Cauchy-Schwarz'sche-Ungleichung in } L^2(\Omega)) \\
 &\leq \left(\sum_{i=1}^N \left\| \frac{\partial z}{\partial x_i} \right\|_{L^2(\Omega)}^2 \right)^{1/2} \left(\sum_{i=1}^N \left\| \frac{\partial y}{\partial x_i} \right\|_{L^2(\Omega)}^2 \right)^{1/2} \quad (\text{Cauchy-Schwarz'sche-Ungleichung in } \mathbb{R}^N) \\
 &= \|\nabla z\|_{L^2(\Omega)^N} \|\nabla y\|_{L^2(\Omega)^N} \\
 &\leq \|\nabla z\|_{L^2(\Omega)^N} \|\nabla y\|_{H^1(\Omega)},
 \end{aligned}$$

denn: $\|\nabla y\|_{L^2(\Omega)^N} \leq \|\nabla y\|_{H^1(\Omega)} = \left(\|y\|_{L^2(\Omega)}^2 + \|\nabla y\|_{L^2(\Omega)^N}^2 \right)^{1/2}$

Aus Def. 4.2 folgt also $\|A\| \leq \|\nabla z\|_{L^2(\Omega)^N}$ □

Definition 4.4

Ein linearer Operator $A : U \rightarrow \mathbb{R}$ heißt lineares Funktional. Der Raum der beschränkten, also stetigen Funktionalen auf U heißt der Dualraum von U und wird mit $U^* = \mathcal{L}(U, \mathbb{R})$ definiert. Durch

$$\|f\|_{U^*} = \sup_{\|u\|_U=1} |f(u)|$$

wird auf U^* eine Norm definiert. Da \mathbb{R} vollständig ist, ist U^* stets ein Banachraum³⁶.

Für die Anwendung eines linearen stetigen Funktionalen $F \in U^*$ auf ein $u \in U$ schreibt man

$$\langle F, u \rangle_{U^*, U} \stackrel{def}{=} F(u).$$

Man bezeichnet dies als sogenannte duale Paarung. □

Dies führt uns zu folgendem wichtigen Satz, der aussagt, dass in Hilberträumen der Dualraum isomorph zum Raum selbst ist:

³⁶Siehe Def. 4.3.

Satz 4.2 (Riesz'scher Darstellungssatz)

Sei H Hilbertraum, dann kann jedes $F \in H^*$ eindeutig durch ein $f \in H$ dargestellt werden, s.d. gilt:

$$F(v) = \langle F, v \rangle_{H^*, H} = (f, v)_H \quad \forall v \in H.$$

Die Zuordnung $H^* \ni F \mapsto f \in H$ heißt Riesz-Isomorphismus. Es gilt $\|F\|_{H^*} = \|f\|_H$. Umgekehrt definiert $(f, v)_H \forall f \in H$ ein stetiges lineares Funktional F mit $\|F\|_{H^*} = \|f\|_H$, s.d. wir H und H^* (isometrisch) identifizieren können. □

BEWEIS Siehe [Alt02, 4.1]. ■

Zur Veranschaulichung der Aussage betrachten wir nun ein Beispiel.³⁷

Beispiel 4.2

Es sei $F \in L^2(\Omega)^*$. Nach dem Riesz'schen Darstellungssatz gibt es genau ein $f \in L^2(\Omega)$, s.d. die Anwendung von F auf ein beliebiges Element $v \in L^2(\Omega)$ als Skalarprodukt mit f dargestellt werden kann:

$$F(v) = \langle F, v \rangle_{L^2(\Omega)^*, L^2(\Omega)} = \int_{\Omega} f(x)v(x)dx \quad \forall v \in L^2(\Omega).$$

Die Cauchy-Schwarz'sche-Ungleichung liefert für die Norm von f :

$$\|f\|_{L^2(\Omega)} = \frac{1}{\|f\|} \int_{\Omega} f(x)^2 dx \leq \sup_{v \neq 0} \frac{1}{\|v\|} \left| \int_{\Omega} f(x)v(x)dx \right| = \|F\|,$$

s.d. $\|F\| = \|f\|$, was die Aussage des Riesz'schen Darstellungssatzes bestätigt. □

Bemerkung:³⁸

Für den normierten linearen Raum U betrachte man zu $u \in U$

$$U^* \ni f \mapsto F_u(f) \stackrel{def}{=} f(u) \in \mathbb{R}.$$

F_u ist eine lineare und stetige Abbildung von U^* nach \mathbb{R} , denn:

$$|F_u(f)| = |f(u)| \leq \|u\|_U \|f\|_{U^*}.$$

D.h. jedes $u \in U$ erzeugt ein sogenanntes Doppeldual $F_u \in U^{**}$. Die Identifikation von u mit F_u heißt die kanonische Einbettung $U \subset U^{**}$.

Definition 4.5

Ein normierter linearer Raum heißt reflexiv, wenn obige kanonische Einbettung surjektiv ist, also wenn gilt: $U = U^{**}$. Ein reflexiver Raum ist nach Definition stets ein Banachraum. Aus dem Riesz'schen Darstellungssatz folgt direkt, dass alle Hilberträume reflexiv sind.³⁹ □

³⁷Siehe auch [Her12, Beispiel 4.11] oder [Mey11, Beispiel 5.9c].

³⁸Siehe auch [Trö09, S.34].

³⁹Siehe dazu auch [Alt02, 6.10].

Bezüglich der Reflexivität gelten für L^p -Räume folgende Eigenschaften:

Beispiel 4.3

Für $1 < p < \infty$ sind $L^p(E)$ reflexiv, so auch etwa $L^p(\Omega)$ und $L^p(\Gamma)$. Der Dualraum zu $L^p(E)$ kann mit $L^q(E)$ (isometrisch) identifiziert werden, wobei q der zu p konjugierte Exponent ist. D.h. zu jedem $F \in L^p(E)^*$ existiert genau ein $f \in L^q(E)$ s.d. gilt:

$$F(v) = \int_E f v \, dx \quad \forall v \in L^p(E)$$

sowie $\|F\|_{L^p(\Omega)^*} = \|f\|_{L^q(\Omega)}$ gilt. Die Stetigkeit von F folgt aus der Hölder'schen Ungleichung 3.1.

Im Spezialfall $p = 1$ bzw. $p = \infty$ gilt zwar $L^1(E)^* \cong L^\infty(E)$, der Dualraum von $L^\infty(E)$ ist aber größer als $L^1(E)$.⁴⁰ Beide Räume sind also nicht reflexiv. □

BEWEIS Siehe [Alt02, 4.12]. ■

Zum Abschluss unserer theoretischen Überlegungen in diesem Kapitel gehen wir noch auf den bereits erwähnten Sobolew'schen Einbettungssatz ein, welcher im Laufe der Arbeit noch mehrmals benötigt wird. Eine zentrale Aussage aus diesen Überlegungen ist etwa, dass Funktionen aus $W^{k,p}(\Omega)$ für hinreichend große k, p sogar klassisch differenzierbar sind. Hierzu zunächst folgende Definition:

Definition 4.6

Es seien U, V normierte lineare Räume mit $U \subset V$, U heißt stetig eingebettet in V , wenn $i : U \ni u \mapsto v \in V$ stetig ist, d.h. wenn eine Konstante c existiert mit

$$\|u\|_V \leq c \|u\|_U \quad \forall u \in U.$$

Man schreibt $U \hookrightarrow V$. □

Satz 4.3 (Sobolew'scher Einbettungssatz)

Es sei $\Omega \subset \mathbb{R}^N$ beschränktes Lipschitzgebiet, $1 < p < \infty$ sowie k eine nichtnegative ganze Zahl. Dann sind folgende stetigen Einbettungen existent:

- falls $kp < N : W^{k,p}(\Omega) \hookrightarrow L^q(\Omega)$, falls $1 \leq q \leq \frac{Np}{N-kp}$,
- falls $kp = N : W^{k,p}(\Omega) \hookrightarrow L^q(\Omega)$, falls $1 \leq q < \infty$,
- falls $kp > N : W^{k,p}(\Omega) \hookrightarrow C(\overline{\Omega})$.

Also gilt in $\Omega \subset \mathbb{R}^2 : H^1(\Omega) \hookrightarrow L^q(\Omega) \forall 1 \leq q < \infty$ und für $\Omega \subset \mathbb{R}^3 : H^1(\Omega) \hookrightarrow L^q(\Omega) \forall 1 \leq q \leq 6$ □

Der Beweis hierfür ist sehr umfangreich. Für eine sehr ausführliche Version sei auf [Ada75, 5.6 bis 5.19] verwiesen.

⁴⁰ $L^\infty(E)^*$ enthält zusätzlich sogenannte endlich additive Maße

§5 Schwache Lösungen der Poissongleichung

Nun haben wir alles zusammen um den ersten Schritt zur Lösung des Optimalsteuerungsproblems (1.5) zu machen. Wir diskutieren nun die Frage nach der Existenz und Lösbarkeit der Modell-PDE (1.1) und (1.2).

Von nun an sei $\Omega \subset \mathbb{R}^N$, $N \geq 2$ ein beschränktes Lipschitzgebiet.

Unsere Modell-PDE, die Poissongleichung, hatte folgende Struktur:

$$\begin{aligned} -\Delta y &= f && \text{in } \Omega \\ y &= 0 && \text{auf } \Gamma \end{aligned} \tag{5.1}$$

mit gegebenen $f \in L^2(\Omega)$.

Ein Problem das sich darstellt ist, dass eine solche Funktion ohne weitere Bedingungen sehr irregulär sein kann. Man denke an die sogenannte Schachbrettfunktion auf dem schachbrettartig unterteilten Einheitsquadrat $\Omega \in \mathbb{R}$. Die dort definierte Funktion hat auf den schwarzen Feldern den Wert 1, ansonsten den Wert 0. Die Werte auf den inneren Rändern spielen keine Rolle da diese das Maß 0 haben und Funktionen aus $L^2(\Omega)$ dort nicht unterscheidbar sind. Man erkennt sofort, dass man für eine solche Funktion keine klassische Lösung $y \in C^2(\Omega) \cap C(\bar{\Omega})$ erwarten kann.⁴¹ An deren Stelle suchen wir nun die sogenannte schwache Lösung von (5.1) in $H_0^1(\Omega)$.

Zunächst nehmen wir an, dass f hinreichend glatt und $y \in C^2(\bar{\Omega})$ eine klassische Lösung ist. Wir multiplizieren (5.1) mit einer beliebigen aber festen Testfunktion $v \in C_0^\infty$ und integrieren über Ω . Wir erhalten also:

$$-\int_{\Omega} \Delta y v \, dx = \int_{\Omega} f v \, dx. \tag{5.2}$$

Durch anschließende partieller Integration erhalten wir daraus:

$$\int_{\Omega} \nabla y \cdot \nabla v \, dx - \int_{\Omega} \frac{\partial y}{\partial n} v \, ds = \int_{\Omega} f v \, dx.$$

Die partielle Integration auf jeden Summanden der linken Seite von (5.2) ist zulässig, da $D_i y, v \in C^1(\bar{\Omega})$ gilt.

Da $v \in C_0^\infty(\Omega)$ und somit $v(x) = 0 \, \forall x \in \Gamma$ gilt, folgt:

$$\int_{\Omega} \nabla y \cdot \nabla v \, dx = \int_{\Omega} f v \, dx \quad \forall v \in C_0^\infty(\Omega) \tag{5.3}$$

⁴¹Siehe auch [Trö09, S.24 f.].

Nun gilt es noch Folgendes zu zeigen:

Lemma 5.1

Seien $f \in L^2(\Omega)$ und $y \in H_0^1(\Omega)$. Falls (5.3) $\forall v \in C_0^\infty(\Omega)$ gilt, so gilt die Gleichung auch $\forall v \in H_0^1(\Omega)$. □

BEWEIS ⁴² Für $v \in H_0^1(\Omega)$ definieren wir folgendermaßen eine lineare Abbildung nach \mathbb{R} :

$$l(v) \stackrel{def}{=} \int_{\Omega} \nabla y \cdot \nabla v \, dx - \int_{\Omega} f v \, dx.$$

Analog zu Bsp. 4.1 gilt:

$$\begin{aligned} |l(v)| &\leq \sum_{i=1}^N \left\| \frac{\partial y}{\partial x_i} \right\|_{L^2(\Omega)} \left\| \frac{\partial v}{\partial x_i} \right\|_{L^2(\Omega)} + \|f\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} \\ &\leq \|\nabla y\|_{L^2(\Omega)^N} \|\nabla v\|_{L^2(\Omega)^N} + \|f\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} \\ &\leq \underbrace{\left(\|\nabla y\|_{L^2(\Omega)^N} + \|f\|_{L^2(\Omega)} \right)}_{\stackrel{def}{=} c < \infty} \underbrace{\left(\|\nabla v\|_{L^2(\Omega)^N}^2 + \|v\|_{L^2(\Omega)}^2 \right)^{1/2}}_{= \|v\|_{H^1(\Omega)}} = c \|v\|_{H^1(\Omega)}. \end{aligned}$$

Somit ist l eine stetige lineare Abbildung auf $H_0^1(\Omega)$ bezüglich der H^1 -Norm nach \mathbb{R} .⁴³ Nach Voraussetzung ist $l(v) = 0 \, \forall v \in C_0^\infty(\Omega)$. Zu zeigen ist noch: $l(v) = 0 \, \forall v \in H_0^1(\Omega)$:

Sei also $v \in H_0^1(\Omega)$ gegeben. Nach Def. 3.9 liegt $C_0^\infty(\Omega)$ dicht in $H_0^1(\Omega)$, daher gibt es eine Folge $\{v_n\} \subset C_0^\infty(\Omega)$, s.d. $\|v - v_n\|_{H_0^1(\Omega)} \rightarrow 0$. Zudem gilt nach Voraussetzung $l(v_n) = 0 \, \forall n \in \mathbb{N}$. Aus der Stetigkeit von l folgt somit $l(v) = 0$. ■

Unsere gesamten Vorüberlegungen führen uns zu folgender Definition:

Definition 5.1

$y \in H_0^1(\Omega)$ heißt schwache Lösung von (5.1), wenn die sogenannte schwache Formulierung (im Folgenden auch Variationsformulierung)

$$\int_{\Omega} \nabla y \cdot \nabla v \, dx = \int_{\Omega} f v \, dx \quad \forall v \in H_0^1(\Omega) \tag{5.4}$$

erfüllt ist. Dabei ist $y|_{\Gamma} = 0$ bereits durch die Wahl von $H_0^1(\Omega)$ als Lösungsraum enthalten (dies wird als wesentliche Randbedingung bezeichnet). □

⁴²Siehe auch [Mey11, Lemma 6.1] und [H⁺09, Lemma 7.1].

⁴³Siehe Def. 4.2 und Satz 4.1.

Dies macht den bereits erwähnten Vorteil einer schwachen Lösung deutlich:

Anmerkung 5.1

Man sieht, dass die schwache Lösung also nur Ableitungen erster Ordnung im schwachen Sinne besitzt. Dies ist durchaus bemerkenswert, da es sich ursprünglich doch um eine Gleichung zweiter Ordnung handelte. \square

Anmerkung 5.2

Aus Lemma 5.1 erhalten wir zudem, dass die Definition der schwachen Lösung äquivalent dazu ist, (5.4) lediglich $\forall v \in C_0^\infty(\Omega)$ zu fordern.⁴⁴

Im weiteren Verlauf dieser Arbeit beziehen wir uns nur noch auf die schwache Formulierung (5.4) bzw. deren Lösung in $H_0^1(\Omega)$. Die ursprüngliche 'starke' Formulierung (5.1) ist daher nur noch als formal anzusehen. \square

Für eine handlichere Darstellung, auch für allgemeinere Problemstellungen, führen wir nun den Raum

$$V = H_0^1(\Omega)$$

sowie die Bilinearform

$$a : V \times V \rightarrow \mathbb{R}, \quad a[y, v] \stackrel{\text{def}}{=} \int_{\Omega} \nabla y \cdot \nabla v \, dx \tag{5.5}$$

und ein lineares Funktional

$$F : V \rightarrow \mathbb{R}, \quad F(v) \stackrel{\text{def}}{=} \int_{\Omega} f v \, dx$$

ein. Dadurch nimmt die Variationsformulierung folgende Form an:

$$a[y, v] = F(v) \quad \forall v \in V. \tag{5.6}$$

In dieser Form gilt: $F \in V^*$, denn es gilt:⁴⁵

$$|F(v)| = \left| \int_{\Omega} f v \, dx \right| \leq \|f\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} \leq \|f\|_{L^2(\Omega)} \|v\|_{H^1(\Omega)} \tag{5.7}$$

und damit $\|F\|_{V^*} \leq \|f\|_{L^2(\Omega)}$.

Für die Existenz und Eindeutigkeit von Lösungen von elliptischen Randwertproblemen ist folgender Satz von erheblicher Bedeutung. Er stellt eine Folgerung des Riesz'schen Darstellungssatzes dar:

⁴⁴Siehe auch [Mey11, Bemerkung 6.3].

⁴⁵Siehe auch [Her12, S.30] oder [H⁺09, Lemma 7.1].

Lemma 5.2 (Lemma von Lax-Milgram)

V sei Hilbertraum und $a : V \times V \rightarrow \mathbb{R}$ eine Bilinearform mit den folgenden Eigenschaften: Es existieren Konstanten α_0 (Stetigkeitskonstante), β_0 (Koerzitivitätskonstante) > 0 s.d.

$$|a[y, v]| \leq \alpha_0 \|y\|_V \|v\|_V \text{ (Beschränktheit)} \tag{5.8}$$

$$a[y, y] \geq \beta_0 \|y\|_V^2 \text{ (V-Elliptizität/Koerzitivität)} \tag{5.9}$$

erfüllt sind.

Dann hat die Variationsformulierung (5.6) $\forall F \in V^*$ genau eine Lösung $y \in V$ und es gilt die a priori Abschätzung

$$\|y\|_V \leq \frac{1}{\beta_0} \|F\|_{V^*}. \tag{5.10}$$

□

BEWEIS Siehe [Alt02, 4.2].

■

Die Lösungsabbildung $V^* \ni F \mapsto y \in V$ ist aufgrund der Linearität der PDE ebenfalls linear und (5.10) liefert ihre Stetigkeit. In Kapitel III werden wir noch sehen, dass für den Fall dass f die Steuerung ist, wir so einen stetigen Steuerungs-Zustands-Operator bekommen.⁴⁶

Wir benötigen für die Anwendung von Lemma 5.2 auf homogene Dirichlet-Randbedingungen wie in (5.1) noch folgendes Lemma:

Lemma 5.3 (Poincaré-Friedrich-Ungleichung⁴⁷)

Ω sei ein beschränktes Lipschitzgebiet. Dann existiert eine von Ω abhängige Konstante $c = c_\Omega$, s.d. gilt:

$$\int_{\Omega} |y|^2 dx \leq c_\Omega \int_{\Omega} |\nabla y|^2 dx \quad \forall y \in H_0^1(\Omega).$$

□

BEWEIS Siehe [Alt02, 4.7].

■

Mit den Lemmata 5.2 und 5.3 ist es uns nun möglich die Existenz und Eindeutigkeit der schwachen Lösung der Poissongleichung zu zeigen:

Satz 5.1 (Satz von Friedrich)

Ist Ω ein beschränktes Lipschitzgebiet, so besitzt (5.1) für alle $f \in L^2(\Omega)$ genau eine schwache Lösung $y \in H_0^1(\Omega)$. Es existiert zudem eine von f unabhängige Konstante c , s.d. gilt:

$$\|y\|_V \leq c \|f\|_{L^2(\Omega)}.$$

□

⁴⁶Siehe auch [Mey11, S.31].

⁴⁷Für die allgemeine Version sei auf [Alt02, 6.15] verwiesen.

BEWEIS ⁴⁸ Da $H_0^1(\Omega)$ nach Anmerkung 3.1 ein Hilbertraum ist, wenden wir das Lemma von Lax-Milgram mit $V = H_0^1(\Omega)$ an. Dazu müssen noch (5.8) und (5.9) nachgewiesen werden: Die Beschränktheit folgt mit Hilfe der Cauchy-Schwarz'schen Ungleichung aus folgender Abschätzung:

$$|a[y, v]| = \left| \int_{\Omega} \nabla y \cdot \nabla v \, dx \right| \leq \| \nabla y \|_{L^2(\Omega)^N} \| \nabla v \|_{L^2(\Omega)^N} \leq \| y \|_{H^1(\Omega)} \| v \|_{H^1(\Omega)},$$

vgl. auch den Beweis von Lemma 5.1. Beachte, dass $H_0^1(\Omega)$ Teilraum von $H^1(\Omega)$ ist. Daher ist es legitim mit der H^1 -Norm zu arbeiten.

Die V-Elliptizität folgt aus der Abschätzung:

$$\begin{aligned} a[y, y] &= \int_{\Omega} |\nabla y|^2 \, dx = \frac{1}{2} \int_{\Omega} |\nabla y|^2 \, dx + \frac{1}{2} \int_{\Omega} |\nabla y|^2 \, dx \\ &\geq \frac{1}{2} \int_{\Omega} |\nabla y|^2 \, dx + \frac{1}{2c_{\Omega}} \int_{\Omega} |y|^2 \, dx \quad (\text{Poincaré-Friedrich-Ungleichung}) \\ &\geq \frac{1}{2} \min\left\{1, \frac{1}{c_{\Omega}}\right\} \|y\|_{H^1(\Omega)}^2 \quad (\text{Definition } H^1\text{-Norm}). \end{aligned}$$

Die Beschränktheit von F ($F \in V^*$), also $\|F\|_{V^*} \leq \|f\|_{L^2(\Omega)}$, folgte bereits aus (5.7). Die Existenz von c mit $\|y\|_{H^1(\Omega)} \leq c\|F\|_{V^*} \leq c\|f\|_{L^2(\Omega)}$ folgt sofort aus dem Lemma von Lax-Milgram. ■

Wir haben also das Ziel des Kapitels erreicht. Abschließend kann man u. a. aus obigem Satz noch weitere Schlüsse ziehen und das Vorgehen dieses Paragraphen genauer beleuchten.⁴⁹

Anmerkung 5.3

Aus dem obigen Beweis kann man ersehen, dass die Poissongleichung auch eine eindeutige schwache Lösung in $H_0^1(\Omega)$ besitzt, wenn $F \in H^{-1}(\Omega) \stackrel{\text{def}}{=} H_0^1(\Omega)^*$ gilt.⁵⁰ Der Raum $H^{-1}(\Omega)$ umfasst im Gegensatz zu $L^2(\Omega)$ auch sogenannte Linienmaße:

$$F(v) = \int_{\Gamma^*} g(x)v(x)ds^*,$$

wobei Γ^* eine hinreichend glatte Kurve in Ω , $g \in L^2(\Gamma^*)$ und ds^* das Lebesgue-Maß auf Γ^* ist. Mit Hilfe der Cauchy-Schwarz'schen Ungleichung und dem Spursatz 3.2 zeigt man analog zu (5.7), dass $F \in H_0^1(\Omega)^*$ gilt. Auch darauf kann dann Lemma 5.2 angewendet werden.

Durch §3.3, speziell Anmerkung 3.1, wissen wir, dass $H^1(\Omega)$ und $H_0^1(\Omega)$ Hilberträume sind. D.h. nach Riesz'schem Darstellungssatz können sie also mit ihren Dualräumen identifiziert werden. Der Grund, warum wir dennoch explizit V^* bzw. $H^{-1}(\Omega)$ betrachten, und nicht Satz 4.2 benutzen, ist, dass z. B. die Identifizierung von V mit V^* nach dem Riesz'schen Darstellungssatz mit

⁴⁸Siehe auch [Trö09, Satz 2.4].

⁴⁹Siehe auch [Mey11, Bemerkungen 6.8 und 6.9].

⁵⁰Man spricht in dieser Form auch von negativen Sobolewräumen.

Hilfe eines Skalarprodukts in V erfolgt. Im Fall von $H_0^1(\Omega)$ heißt das, dass zu jedem $F \in H^{-1}(\Omega)^*$ ein eindeutiges $u \in H_0^1(\Omega)$ existiert, s.d.

$$(u, v)_{H_0^1(\Omega)} = \int_{\Omega} (\nabla u \cdot \nabla v + uv) dx = F(v) \quad \forall v \in H_0^1(\Omega).$$

was wiederum das erneute Lösen einer PDE notwendig macht. □

III Die Existenz optimaler Steuerungen

Nachdem wir uns also mit der Existenz und Eindeutigkeit der Zustandsgleichung beschäftigt haben, betrachten wir nun die eigentliche Optimalsteuerungsaufgabe (1.5). Zunächst diskutieren wir die Frage nach der Existenz und Eindeutigkeit einer optimalen Steuerung mit zugehörigem optimalen Zustand. Denn gäbe es in der Theorie eine solche nicht, so könnten in der Praxis numerische Näherungslösungen bei feiner werdender Diskretisierung gegen keinen aussagekräftigen Wert konvergieren. Wir werden im Verlaufe des Kapitels zunächst wieder von einer allgemeineren Aufgabenstellung ausgehen und schlussendlich Existenz und Eindeutigkeit für unser Modellproblem zeigen, wir folgen dabei im Groben [Trö09, 2.4 bis 2.5] sowie [Mey11, §7 und §8] und [Her12, §6 und §7]. Beginnen werden wir aber wieder mit Resultaten aus der Funktionalanalysis, die im Folgenden nötig sind.

§6 Schwache Konvergenz und Konvexität

Die Theorie rund um die Existenz optimaler Steuerungen wie auch das Untersuchen von Optimalitätsbedingungen (siehe Kapitel IV) ist im endlichdimensionalen Fall sehr viel einfacher als dies im Unendlichdimensionalen der Fall sein wird.⁵¹ Ein Grund dafür ist, dass wir dort den Satz von Heine-Borel anwenden können, wonach Mengen genau dann kompakt sind, wenn sie beschränkt und abgeschlossen sind.⁵² Auch im unendlichdimensionalen Fall brauchen wir aber solch eine Eigenschaft. Wir gehen hier den Weg über die sogenannte schwache Kompaktheit.

Definition 6.1

U sei normierter linearer Raum über \mathbb{R} . Eine Folge $\{u_n\} \subset U$ heißt schwach konvergent, wenn der sogenannte schwache Grenzwert $u \in U$ existiert, s.d.

$$f(u_n) \rightarrow f(u)$$

für alle stetigen linearen Funktionale $f \in U^*$ gilt. Wir schreiben im Folgenden $u_n \rightharpoonup u$.

Im endlichdimensionalen Fall sind schwache und starke Konvergenz äquivalent. □

Es gelten diesbezüglich folgende Eigenschaften:

Lemma 6.1

a) *Der schwache Grenzwert ist eindeutig, d.h.*

$$u_n \rightharpoonup u, u_n \rightharpoonup v \Rightarrow u = v.$$

⁵¹Für die Grundkonzepte im endlichdimensionalen Fall siehe [Trö09, 1.4].

⁵²Siehe [Kön04, 1.4].

b) Aus starker Konvergenz folgt schwache, d.h.

$$u_n \rightarrow u \Rightarrow u_n \rightharpoonup u.$$

c) Schwach konvergente Folgen sind stets beschränkt, d.h. $\exists c \in \mathbb{R}$ s.d. gilt:

$$u_n \rightharpoonup u \Rightarrow \|u_n\| \leq c.$$

d) Im Hilbertraum H ist $u_n \rightharpoonup u$ äquivalent zu

$$(v, u_n)_H \rightarrow (v, u)_H \quad \forall v \in H.$$

e) Es gilt:

$$u_n \rightharpoonup u \text{ in } U, f_n \rightarrow f \text{ in } U^* \Rightarrow \langle f_n, u_n \rangle_{U^*, U} \rightarrow \langle f, u \rangle_{U^*, U}.$$

f) Im Hilbertraum H gilt:

$$u_n \rightharpoonup u \text{ und } \|u_n\| \rightarrow \|u\| \Leftrightarrow u_n \rightarrow u. \quad \square$$

BEWEIS Siehe [Alt02, 6.3] und [Yos80, V.1.]. ■

Die schwache Konvergenz ist beim Beweis von Existenzsätzen, also bei theoretischen Überlegungen, von erheblicher Bedeutung wie wir noch sehen werden. Numerisch ist dies jedoch nicht der Fall wie das folgende Beispiel zeigen wird:⁵³

Beispiel 6.1

Im Hilbertraum $H = L^2(0, 2\pi)$ betrachten wir

$$u_n = \frac{1}{\sqrt{\pi}} \sin(nx).$$

Es gilt:

$$(u_n, u_m)_H = \frac{1}{\sqrt{\pi}} \int_0^{2\pi} \sin(nx) \sin(mx) dx = \delta_{nm} \stackrel{\text{def}}{=} \begin{cases} 1, & n = m, \\ 0, & n \neq m. \end{cases} \quad (6.1)$$

Die Folge stellt also ein Orthonormalitätssystem von H dar. Auf ein solches System kann man die Bessel'sche Ungleichung⁵⁴ anwenden:

$$\sum_{n=1}^{\infty} |(f, u_n)_H|^2 \leq \|f\|_H^2 \quad \forall f \in H.$$

⁵³Siehe auch [Trö09, S.34 f.] und [Mey11, Beispiel 7.4].

⁵⁴Siehe [Alt02, 7.6].

Für $f \in H$ ist also $\sum_{n=1}^{\infty} |(f, u_n)_H|^2 < \infty \Rightarrow (f, u_n)_H \rightarrow 0 = (f, 0)_H \Leftrightarrow u_n \rightarrow 0$ in H .
 Andererseits ist nach (6.1)

$$\|u_n - u_m\|_H^2 = (u_n, u_n)_H - 2(u_n, u_m)_H + (u_m, u_m)_H = 2 \text{ für } n \neq m.$$

$(u_n)_{n \in \mathbb{N}}$ ist somit keine Cauchy-Folge, also nicht konvergent, die Folge konvergiert nur im schwachen Sinne. \square

Man sieht also: Es gibt schwach konvergente Teilfolgen, deren Elemente alle auf der Einheitskugeloberfläche liegen ($\|u_n\| = 1$), der schwache Grenzwert ist jedoch Null. Schwache Konvergenz sagt also nur sehr wenig über das Konvergenzverhalten im herkömmlichen Sinne aus.

Definition 6.2

U und V seien normierte Lineare Räume.

a) $F : U \rightarrow V$ heißt schwach (folgen-)stetig, wenn gilt:

$$u_n \rightarrow u \Rightarrow F(u_n) \rightarrow F(u).$$

b) Eine Menge $M \subset U$ heißt schwach (folgen-)abgeschlossen, wenn gilt:

$$u_n \in M, u_n \rightarrow u \Rightarrow u \in M.$$

c) $M \subset U$ heißt relativ schwach folgenkompakt, wenn jede Folge $(u_n)_{n \in \mathbb{N}} \subset M$ eine in U schwach konvergente Teilfolge besitzt. M heißt schwach folgenkompakt, wenn sie zudem schwach folgenabgeschlossen ist. \square

Es ist ersichtlich, dass die schwache Abgeschlossenheit eine stärkere Forderung ist als die starke, also normale Abgeschlossenheit (es gibt viel mehr schwach konvergente als stark konvergente Folgen), d.h. jede schwach abgeschlossene Menge ist auch stark abgeschlossen (siehe dazu Lemma 6.1), die Umkehrung gilt i. A. nicht (siehe dazu die Einheitskugeloberfläche in Bsp. 6.1).⁵⁵ Zur Verdeutlichung betrachten wir folgendes Beispiel:⁵⁶

Beispiel 6.2

Die Normabbildung $u \mapsto \|u\|$ in unendlichdimensionalen normierten linearen Räumen ist i. A. nicht schwach stetig. Wie in Bsp. 6.1 betrachten wir $H = L^2(0, 2\pi)$ und u_n , welches schwach gegen 0 konvergiert:

$$u_n \rightarrow u = 0 \Rightarrow \underbrace{\|u_n\|}_{=1} \rightarrow \underbrace{\|u\|}_{=0}.$$

Also: Nicht jede stetige Abbildung ist schwach folgenstetig. \square

⁵⁵Siehe auch [Mey11, S.37].

⁵⁶Siehe auch [Trö09, S.36f.].

Wir werden nun Sätze betrachten, die uns aufzeigen warum die schwache Konvergenz ein wichtiges Konzept im Unendlichdimensionalen darstellt. Wie bereits angedeutet, wird die schwache Kompaktheit den üblichen Kompaktheitsbegriff im Unendlichdimensionalen ersetzen.

Der erste dieser Sätze zeigt uns die Wichtigkeit der schwachen Kompaktheit für unsere Anwendungen. In einem gewissen Sinne dient uns die relative schwache Folgenkompaktheit als Ersatz für die Präkompaktheit. Es folgt sogar aus dem Satz von Eberlein und Smuljan, dass folgende Eigenschaft reflexive Banachräume charakterisiert.⁵⁷

Satz 6.1

Jede beschränkte Menge eines reflexiven Banachraums ist relativ schwach folgenkompakt. □

BEWEIS Siehe [Alt02, 6.9], der Satz folgt aus dem Satz von Hahn-Banach [Alt02, 4.14] oder spezieller dem Satz von Hahn-Banach für lineare Funktionale [Alt02, 4.15]. ■

Definition 6.3

Sei U ein normierter linearer Raum. $C \subset U$ heißt konvex, wenn gilt:

$$u, v \in C \Rightarrow \alpha u + (1 - \alpha)v \in C \quad \forall \alpha \in [0, 1]. \quad \square$$

Satz 6.2

Jede konvexe und abgeschlossene Menge in einem normierten linearen Raum ist schwach folgenabgeschlossen. □

BEWEIS Siehe [Alt02, 6.12], folgt aus dem Trennungssatz [Alt02, 6.11]. ■

Die Sätze 6.1 und 6.2 liefern nun sofort einen Ersatz für den bereits angesprochenen Satz von Heine-Borel im Unendlichdimensionalen:

Korollar 6.1

Jede beschränkte, konvexe und abgeschlossene Menge in einem reflexiven Banachraum ist schwach folgenkompakt. □

Daraus folgt direkt das unendlichdimensionale Pendant zum Satz von Bolzano-Weierstrass⁵⁸.

Korollar 6.2

Jede beschränkte Folge in einem reflexiven Banachraum besitzt eine schwach konvergente Teilfolge. □

BEWEIS ⁵⁹ Ist $(u_n)_{n \in \mathbb{N}}$ beschränkt im reflexiven Banachraum U , dann existiert ein $c > 0$, s.d. $\|u_n\| \leq c \quad \forall n \in \mathbb{N}$. $M \stackrel{def}{=} \{u \in U : \|u\|_U \leq c\}$ ist beschränkt, konvex und abgeschlossen, nach Korollar 6.1 also schwach folgenkompakt. Somit existiert eine schwach konvergente Teilfolge von $(u_n)_{n \in \mathbb{N}}$. ■

⁵⁷Siehe auch [Trö09, S.37] und [Yos80, Appendix to Chapter V, 4.].

⁵⁸Für den endlichdimensionalen Fall siehe [Kön04, 1.1].

⁵⁹Siehe auch [Mey11, Korollar 7.12].

Wir erweitern nun den Konvexitätsbegriff auf Funktionale:

Definition 6.4

Ein Funktional $f : U \rightarrow \mathbb{R}$ in einem normierten linearen Raum heißt konvex, wenn gilt:

$$f(\alpha u + (1 - \alpha)v) \leq \alpha f(u) + (1 - \alpha)f(v) \quad \forall u, v \in U, \forall \alpha \in (0, 1). \quad (6.2)$$

Es heißt strikt konvex, wenn (6.2) $\forall u \neq v, \forall \alpha \in (0, 1)$ mit ' $<$ ' erfüllt ist. □

Satz 6.3

Jedes konvexe und stetige Funktional $f : U \rightarrow \mathbb{R}$ im normierten linearen Raum U ist schwach unterhalbstetig, d.h.

$$u_n \rightarrow u \Rightarrow \liminf_{n \rightarrow \infty} f(u_n) \geq f(u). \quad \square$$

BEWEIS Siehe [Wer11, Lemma III.5.9], folgt aus [Wer11, III.3.8]. ■

Daraus folgt:⁶⁰

Beispiel 6.3

Die Normabbildung ist schwach unterhalbstetig, denn sie ist stetig und konvex:

$$\|\alpha u_1 + (1 - \alpha)u_2\| \leq \alpha \|u_1\| + (1 - \alpha)\|u_2\|.$$

(Aufgrund der Dreiecksungleichung und der Homogenität der Norm.) □

Korollar 6.1 und Satz 6.3 unterstreichen die Wichtigkeit der Konvexität für die Behandlung von Optimalsteuerungsaufgaben in Funktionenräumen. Die Anwendung dieser Aussagen liefert eine leichtere Handhabung von konvexen Optimalsteuerungsaufgaben (wie unserem Modellproblem), welche u. a. dadurch wesentlich einfacher zu diskutieren als nicht konvexe.⁶¹

§7 Existenz optimaler Steuerungen

Wir haben wieder das theoretische Rüstzeug zusammen um nun eine der zentralen Fragen dieser Arbeit zu beantworten. Die nach der Existenz und der Eindeutigkeit optimaler Steuerungen, speziell für unser Modellproblem (1.5). Gelingt es für ein gegebenes Problem nicht, mit gängigen Methoden wie wir sie hier anwenden, die Existenz einer Lösung nachzuweisen, dann verbergen sich dahinter oft Modellierungsfehler, die auch zu numerischen Schwierigkeiten führen können. Wir beginnen mit einer sehr allgemeinen Aufgabe in einem reflexiven Banachraum und wenden unsere Ergebnisse auf unser Modellproblem an, um eine endgültige Aussage zu erhalten.

⁶⁰Siehe auch [Trö09, S.38].

⁶¹Siehe dazu etwa [Trö09, 3 ff.].

Definition 7.1

Sei U Banachraum, $f : u \rightarrow \mathbb{R}$ und $U_{ad} \subset U$ gegeben.⁶² Eine Steuerung $\bar{u} \in U_{ad}$ heißt global optimal, wenn gilt:

$$f(\bar{u}) \leq f(u) \quad \forall u \in U_{ad}. \quad \square$$

Wir beginnen nun die Ergebnisse aus §6 anzuwenden um unser Ziel zu erreichen:

Satz 7.1

Es sei U reflexiver Banachraum, $U_{ad} \subset U$ nichtleer, konvex, beschränkt und abgeschlossen. Sei $f : U \rightarrow \mathbb{R}$ konvexes und stetiges Funktional. Zudem existiere ein $c > -\infty$ s.d.

$$f(u) \geq c \quad \forall u \in U_{ad}.$$

Dann besitzt die Optimierungsaufgabe

$$\min_{u \in U_{ad}} f(u) \tag{7.1}$$

eine global optimale Lösung \bar{u} . Ist f strikt konvex, dann ist die optimale Lösung \bar{u} eindeutig. \square

BEWEIS ⁶³ Nach Voraussetzung gilt $f(u) \geq c \quad \forall u \in U_{ad}$, daher existiert

$$j = \inf_{u \in U_{ad}} f(u)$$

Daraus folgt: $\exists \{u_n\} \subset U_{ad} : f(u_n) \rightarrow j$.

Obwohl U_{ad} beschränkt und abgeschlossen, aber i. A. nicht kompakt ist, folgt aus Korollar 6.2 dass $\{u_n\}$ eine Teilfolge $\{u_{n_k}\}$ enthält, die im schwachen Sinne gegen ein $\bar{u} \in U$ konvergiert:

$$u_{n_k} \rightharpoonup \bar{u} \text{ in } U.$$

Nach Satz 6.2 ist U_{ad} schwach abgeschlossen, da es konvex und abgeschlossen ist, also gilt $\bar{u} \in U_{ad}$.

Damit ist \bar{u} zulässig. Es bleibt zunächst zu zeigen, dass es auch ein globales Minimum ist.

f ist zwar stetig, i. A. aber nicht schwach stetig. Daher ist es nicht möglich, $f(u_{n_k}) \rightarrow f(\bar{u})$ zu schließen. f ist jedoch nach Satz 6.3 schwach unterhalbstetig, da es nach Voraussetzung konvex ist, s.d.

$$j = \lim_{k \rightarrow \infty} f(u_{n_k}) = \liminf_{k \rightarrow \infty} f(u_{n_k}) \geq f(\bar{u}).$$

Da $\bar{u} \in U_{ad}$ und $j = \inf_{u \in U_{ad}} f(u)$ gilt $f(\bar{u}) = j$, d.h. \bar{u} ist global optimal.

Die Eindeutigkeit von \bar{u} für ein strikt konvexes f erhalten wir mittels Widerspruchsannahme:

⁶²Der Index 'ad' steht hierbei für 'admissible' (englisch für 'zulässig'). Hierauf wird später noch eingegangen.

⁶³Siehe auch [Mey11, Satz 8.2 und Bemerkung 8.3] oder [Trö09, Satz 2.14].

III Die Existenz optimaler Steuerungen

Sei $\tilde{u} \in U_{ad}$ eine weitere globale Lösung von (7.1), dann folgt aus der strikten Konvexität von f für ein beliebiges $\alpha \in (0, 1)$:

$$f(\alpha\bar{u} + (1-\alpha)\tilde{u}) < \alpha f(\bar{u}) + (1-\alpha)f(\tilde{u}) = \alpha j + (1-\alpha)j = j, \quad (7.2)$$

s.d. z. B. der Ausdruck $\frac{1}{2}(\bar{u} + \tilde{u})$ einen kleineren Funktionswert als das Infimum über U_{ad} liefern würde. Wegen der Konvexität von U_{ad} liegt $\frac{1}{2}(\bar{u} + \tilde{u})$ als konvexe Linearkombination in U_{ad} , was zu einem Widerspruch führt.

Man sieht, dass die Konvexität von f sogar nur auf U_{ad} und nicht auf ganz U gefordert werden müsste. Zudem erkennt man, für (7.2) mit ' \leq ' statt '<' und für den Fall eines konvexen Zielfunktionals, dass die Menge der globalen Optima eine konvexe Menge darstellt. ■

Auf die Beschränktheit von U_{ad} kann u.U. verzichtet werden:

Satz 7.2

Sei U Banachraum, $u \in U_{ad}$ nichtleer, konvex und abgeschlossen. Es seien die Voraussetzungen aus Satz 7.1 an ein $f : U \rightarrow \mathbb{R}$ erfüllt. Zudem gelte für f radiale Unbeschränktheit, d.h.

$$\|u\|_U \rightarrow \infty \Rightarrow f(u) \rightarrow \infty.$$

Dann existiert eine global optimale Lösung von 7.1, die für f strikt konvex eindeutig ist. □

BEWEIS ⁶⁴ Sei j das Infimum von f auf U_{ad} . Wegen der radialen Unbeschränktheit existiert ein $r > 0$, s.d. gilt:

$$f(u) > j + 1 \quad \forall u \in U_{ad} \text{ mit } \|u\|_U > r$$

Es reicht daher das Minimum auf der Menge

$$M \stackrel{\text{def}}{=} U_{ad} \cap \overline{B_r(0)} = \{u \in U_{ad} : \|u\|_U \leq r\}$$

zu suchen. M ist beschränkt und abgeschlossen und als Durchschnitt zweier konvexer Mengen selbst konvex. Satz 7.1 angewandt auf $\min_{u \in M} f(u)$ liefert dann die Aussage. ■

Lemma 7.1

U und H seien Hilberträume, $S \in \mathcal{L}(U, H)$, $y_d \in H$, $\lambda > 0$. Dann ist das Funktional

$$f(u) \stackrel{\text{def}}{=} \frac{1}{2} \|Su - y_d\|_H^2 + \frac{\lambda}{2} \|u\|_U^2$$

konvex und sogar strikt konvex für den Fall $\lambda > 0$. □

⁶⁴Siehe auch [Mey11, Satz 8.5].

BEWEIS ⁶⁵ Für die linke Seite in (6.2) (Definition konvexer Funktionale) erhalten wir hier:

$$f(\alpha u + (1 - \alpha)v) = \frac{1}{2} \|\alpha(Su - y_d) + (1 - \alpha)(Sv - y_d)\|_H^2 + \frac{\lambda}{2} \|\alpha u + (1 - \alpha)v\|_U^2. \quad (7.3)$$

Die zweite Norm kann wie Folgt abgeschätzt werden:

$$\begin{aligned} \|\alpha u + (1 - \alpha)v\|_U^2 &= \alpha^2 \|u\|_U^2 + 2\alpha(1 - \alpha)(u, v)_U + (1 - \alpha)^2 \|v\|_U^2 \\ &= \alpha \|u\|_U^2 + (1 - \alpha) \|v\|_U^2 - (\alpha - \alpha^2) \|u\|_U^2 + 2\alpha(1 - \alpha)(u, v)_U - (1 - \alpha - (1 - \alpha)^2) \|v\|_U^2 \\ &= \alpha \|u\|_U^2 + (1 - \alpha) \|v\|_U^2 - (\alpha - \alpha^2) \|u - v\|_U^2. \end{aligned} \quad (7.4)$$

Wegen $0 < \alpha < 1$ und dem daraus folgenden $\alpha^2 < \alpha$ gilt somit

$$\|\alpha u + (1 - \alpha)v\|_U^2 \leq \alpha \|u\|_U^2 + (1 - \alpha) \|v\|_U^2.$$

Schätzt man die erste Norm in (7.3) analog zu (7.4) ab, so folgt insgesamt

$$\begin{aligned} &\frac{1}{2} \|S(\alpha u + (1 - \alpha)v) - y_d\|_H^2 + \frac{\lambda}{2} \|\alpha u + (1 - \alpha)v\|_U^2 \\ &\leq \alpha \left(\frac{1}{2} \|Su - y_d\|_H^2 + \frac{\lambda}{2} \|u\|_U^2 \right) + (1 - \alpha) \left(\frac{1}{2} \|Sv - y_d\|_H^2 + \frac{\lambda}{2} \|v\|_U^2 \right), \end{aligned}$$

also die Konvexität von f .

Für den Beweis der strikten Konvexität von f sei nun $u \neq v$. Dann folgt aus (7.4):

$$\|\alpha u + (1 - \alpha)v\|_U^2 < \alpha \|u\|_U^2 + (1 - \alpha) \|v\|_U^2 \quad \forall \alpha \in (0, 1).$$

Schätzt man $\|\alpha Su + (1 - \alpha)Sv - y_d\|_H^2$ wie im obigen Fall ab, ergibt sich für den Fall $\lambda > 0$ also die strikte Konvexität. ■

Bemerkung:⁶⁶

Mann erkennt, dass f auch im Fall $\lambda = 0$ strikt konvex ist, falls S injektiv ist, denn analog zu (7.4) leitet man her:

$$\|S(\alpha u + (1 - \alpha)v) - y_d\|_H^2 = \alpha \|Su - y_d\|_V^2 + (1 - \alpha) \|Sv - y_d\|_H^2 - (\alpha - \alpha^2) \|S(u - v)\|_H^2. \quad (7.5)$$

Falls $u \neq v$, so folgt $S(u - v) \neq 0$ für ein injektives S und damit analog zu oben die strikte Konvexität von f .

Wir wenden nun die Ergebnisse aus Satz 7.1 und Lemma 7.1 auf unser Modellproblem (1.5) an.

⁶⁵Siehe auch [Mey11, Lemma 8.6].

⁶⁶Siehe auch [Mey11, Bemerkung 8.7].

Anmerkung 7.1

Im Folgenden sei $\Omega \subset \mathbb{R}^N$ beschränktes Lipschitzgebiet, $\lambda > 0$, $y_d, u_a, u_b \in L^2(\Omega)$ mit $u_a \leq u_b$ fast überall in Ω .

Um Satz 7.1 anwenden zu können wählen wir $U = L^2(\Omega)$. Die Menge der zulässigen Steuerungen sei definiert durch

$$U_{ad} \stackrel{\text{def}}{=} \{u \in L^2(\Omega) : u_a(x) \leq u(x) \leq u_b(x) \text{ fast überall in } \Omega\}. \quad \square$$

Wir müssen noch zeigen, dass U_{ad} die Voraussetzungen aus Satz 7.1 erfüllt. Da U_{ad} nach Definition konvex ist bleibt z.z.:

Lemma 7.2

U_{ad} ist beschränkt und abgeschlossen. □

BEWEIS ⁶⁷ Nach Definition gilt $|u(x)| \leq |u_a(x)| + |u_b(x)|$ fast überall $\forall u \in U_{ad}$ und damit $\|u\|_{L^2(\Omega)} \leq \|u_a\|_{L^2(\Omega)} + \|u_b\|_{L^2(\Omega)} \stackrel{\text{def}}{=} c < \infty$. U_{ad} ist also beschränkt. Zudem folgt aus $u_n \rightarrow u$ in $L^2(\Omega)$, dass eine Teilfolge u_{n_k} existiert, mit $u_{n_k}(x) \rightarrow u(x)$ fast überall in Ω ,⁶⁸ woraus folgt, dass U_{ad} abgeschlossen ist. Aus der Struktur von U_{ad} folgt dann trivialerweise die Konvexität. ■

Aufgrund der komponentenweisen Beschränktheit von U_{ad} bezüglich der Schranken u_a und u_b spricht man in diesem Zusammenhang auch von Box-Restriktionen.

Nach Satz 5.1 existiert $\forall u \in U_{ad}$ genau eine schwache Lösung $y \in H_0^1(\Omega)$ der Poissongleichung in (1.5). Diese wird von nun an als der zu u gehörende Zustand, der im Zustandsraum

$$Y \stackrel{\text{def}}{=} H_0^1(\Omega)$$

liegt, bezeichnet (im Folgenden schreiben wir $y(u)$).

Definition 7.2

Die Abbildung

$$G : L^2(\Omega) \rightarrow H_0^1(\Omega), G : u \mapsto y(u)$$

ist als Lösungsabbildung der PDE durch Satz 5.1 wohldefiniert und wird, wie bereits erwähnt, als Steuerungs-Zustands-Operator bezeichnet. □

G ist aufgrund der Linearität der PDE ebenso ein linearer Operator, zudem ist er stetig aufgrund der a priori Abschätzung in Satz 5.1 sowie der Aussage von Satz 4.1.

Es gilt: $y = Gu \in H_0^1(\Omega)$, dennoch werden wir im Zielfunktional lediglich $y \in L^2(\Omega)$ nutzen, Ableitungen des Zustands werden nicht betrachtet. Zur Vereinfachung definieren daher

$$S \stackrel{\text{def}}{=} E \circ G \tag{7.6}$$

⁶⁷Siehe auch [Mey11, Lemma 8.9].

⁶⁸Siehe [Alt02, 1.18].

mit $E : H_0^1(\Omega) \rightarrow L^2(\Omega)$, dem Einbettungsoperator, der jeder Funktion in $H_0^1(\Omega)$ dieselbe Funktion in $L^2(\Omega)$ zuordnet.⁶⁹

Dies ist gerechtfertigt dadurch, dass $H^1(\Omega)$ und damit auch sein Teilraum $H_0^1(\Omega)$ linear und stetig in $L^2(\Omega)$ eingebettet ist.⁷⁰ Da E wie G linear und stetig ist, ist es auch $S : L^2(\Omega) \rightarrow L^2(\Omega)$. Im weiteren Verlauf wird auch S gelegentlich als Steuerungs-Zustands-Operator bezeichnet.

Mit Hilfe der neugewonnenen Abbildungen definieren wir nun:

Definition 7.3

$\bar{u} \in U$ und der zugehörige Zustand $\bar{y} = G\bar{u} \in Y$ heißen global optimal, wenn für alle $u \in U_{ad}$ gilt:

$$J(S\bar{u}, \bar{u}) \leq J(Su, u),$$

mit dem Zielfunktional $J(y, u)$ aus (1.4). □

Wir können uns nun mit der Existenz und Eindeutigkeit einer Lösung für das Modellproblem beschäftigen:

Satz 7.3

(1.5) besitzt eine global optimale Steuerung \bar{u} , die eindeutig bestimmt ist für $\lambda > 0$. □

BEWEIS ⁷¹ Für den Beweis benötigen wir die Aussagen von Satz 7.1 sowie der Lemmata 7.1 und 7.2:

Sei $U = L^2(\Omega)$, wir definieren das sogenannte reduzierte Zielfunktional

$$f(u) \stackrel{def}{=} J(Su, u) = \frac{1}{2} \|Su - y_d\|_{L^2(\Omega)}^2 + \frac{\lambda}{2} \|u\|_{L^2(\Omega)}^2.$$

Per Definition ist f somit nach unten durch Null beschränkt. Die in Lemma 7.1 geforderte Struktur gilt mit $H = L^2(\Omega)$, woraus die Konvexität von f folgt. Die Stetigkeit von f folgt aus der Norm. Nach Lemma 7.2 ist U_{ad} zudem beschränkt, konvex und abgeschlossen, womit Satz 7.1 angewendet werden kann, sodass die Existenz global optimaler Steuerungen gegeben ist. Nach Lemma 7.1 ist f für $\lambda > 0$ strikt konvex nach Voraussetzung, s.d. nach Satz 7.1 die Eindeutigkeit der global optimalen Steuerung folgt. ■

Aufgrund von (7.5) folgt aus einer Injektivität von S die strikte Konvexität von f . Bei der Poissongleichung ist dies der Fall:

Aus der schwachen Formulierung (5.4) folgt für $y = 0$

$$\int_{\Omega} uv \, dx = 0 \quad \forall v \in H_0^1(\Omega).$$

⁶⁹Dass dies die weiteren Untersuchungen vereinfachen wird, wird sich bei der Theorie um den adjungierten Operator S^* zeigen. Dieser wird dann ebenfalls im Raum $L^2(\Omega)$ wirken. Hierauf wird in Kapitel IV ab §9 genauer eingegangen. Siehe auch [Trö09, S.39 f.].

⁷⁰Es gilt: $\|y\|_{L^2(\Omega)} \leq \|y\|_{H^1(\Omega)}$.

⁷¹Siehe auch [Trö09, Satz 2.15] oder [Mey11, Satz 8.12].

Nach dem Fundamentallemma der Variationsrechnung 3.1 ist dies äquivalent zu $u = 0$, woraus direkt die Injektivität von S folgt.

Für den Fall dass U_{ad} unbeschränkte Teilmenge von $L^2(\Omega)$ ist, etwa falls u_a und u_b als Beschränkungen fehlen, so wäre f für den Fall $\lambda > 0$ radial unbeschränkt, s.d aus Satz 7.2 folgt:

Korollar 7.3

Sei $\lambda > 0$, so besitzt 1.5 genau dann eine global optimale Steuerung, wenn U_{ad} nichtleer, konvex und abgeschlossen ist. □

Man muss jedoch beachten, dass f für $\lambda = 0$ nicht radial unbeschränkt ist: Aus $\|u\|_{L^2(\Omega)} \rightarrow \infty$ folgt i. A. nicht $\|y\|_{L^2(\Omega)} \rightarrow \infty$.

§5 hat gezeigt, dass L^2 -Regularität nicht nötig ist um eine schwache Lösung der Poissongleichung in $H_0^1(\Omega)$ zu erhalten (s. Anmerkung 5.3). Es genügt hierfür eine Inhomogenität aus $H^{-1}(\Omega)$. Im Zweidimensionalen sind dies u. a. Funktionen aus $L^p(\Omega)$, $p > 1$, im Dreidimensionalen $L^{6/5}(\Omega)$ -Funktionen.⁷² Es ist also möglich dass $\|y - y_d\|_{L^2(\Omega)} < \infty$ obwohl $\|u\|_{L^2(\Omega)} = \infty$, also obwohl für die zugehörige Inhomogenität gilt: $u \notin L^2(\Omega)$.⁷³

⁷²Siehe dazu [Eva10, 5.6].

⁷³Siehe auch [Mey11, Bemerkung 8.15].

IV Optimalitätsbedingungen erster Ordnung

In diesem Kapitel geht es nun um die Herleitung von Bedingungen, die die eindeutige Lösung des Modellproblems charakterisieren. Wir folgen dabei [Trö09, 2.6 bis 2.8], [Mey11, §9 bis §11] und [Her12, §8 bis §10]. In der Praxis benötigen wir diese um Schlüsse über die Form und Regularität optimaler Steuerungen zu ziehen und sie als Kriterium für die Optimalität numerischer Lösungen sowie als Grundlage für numerische Verfahren zu nutzen. Wir werden dafür v.a. die Ableitungen des Zielfunktional benutzen und beginnen daher das Kapitel mit der Verallgemeinerung des Ableitungsbegriffes in Funktionenräumen und stellen Grundlagen aus der nichtlinearen Funktionalanalysis zu Verfügung.

§8 Differenzierbarkeit in normierten linearen Räumen

Wir setzen voraus: U und V seien normierte lineare Räume, $F : U \rightarrow V$ eine beliebige Abbildung.

Definition 8.1

Existiert zu $u, \delta u \in U$ der Grenzwert

$$\delta F(u; \delta u) \stackrel{\text{def}}{=} \lim_{t \searrow 0} \frac{F(u + t\delta u) - F(u)}{t}.$$

in V , so wird dieser als die Richtungsableitung von F an der Stelle u in Richtung δu bezeichnet. F heißt dann richtungsdifferenzierbar an der Stelle u in Richtung δu . Nach Definition ist die Abbildung $\delta u \mapsto \delta F(u; \delta u)$ nicht zwingend linear, wohl aber positiv homogen, d.h.

$$\delta F(u; \lambda \delta u) = \lambda \delta F(u; \delta u) \quad \forall \lambda \geq 0 \quad \square$$

Definition 8.2

Existiert die Richtungsableitung in u für alle Richtungen $\delta u \in U$ und ist

$$\delta F(u; \delta u) = A\delta u$$

mit einem Operator $A \in \mathcal{L}(U, V)$, so bezeichnet man A als die sogenannte Gâteaux-Ableitung von F an der Stelle u und F heißt Gâteaux-differenzierbar an der Stelle u . Wir schreiben im Folgenden $A = F'_G(u)$. □

Bemerkung:

Ist $F : U \rightarrow \mathbb{R}$, so ist $F'_G(u) \in U^*$.

Man merke sich also: Gâteaux-differenzierbar \simeq richtungsdifferenzierbar mit linearer und stetiger Ableitung.⁷⁴

Als Beispiel betrachten wir ein nichtlineares Punktfunktional:⁷⁵

Beispiel 8.1

Seien $U = C[0, 1]$ und $f(u) = \sin(u(1)) \in V = \mathbb{R}$. Zunächst wird an einer beliebigen Stelle die Richtungsableitung in Richtung $\delta u \in C[0, 1]$ berechnet:

$$\begin{aligned} \lim_{t \searrow 0} \frac{f(u+t\delta u) - f(u)}{t} &= \lim_{t \searrow 0} \frac{\sin(u(1)+t\delta u(1)) - \sin(u(1))}{t} \\ &= \left. \frac{d}{dt} \sin(u(1) + t\delta u(1)) \right|_{t=0} \\ &= \cos(u(1) + t\delta u(1))\delta u(1) \Big|_{t=0} \\ &= \cos(u(1))\delta u(1). \end{aligned}$$

Die Richtungsableitung ist linear in δu und zudem stetig von $C[0, 1]$ nach \mathbb{R} , denn:

$$|\cos(u(1))\delta u(1)| \leq |\cos(u(1))| \cdot \|\delta u\|_{C[0,1]} = c\|\delta u\|_{C[0,1]}.$$

Daher ist f an jeder Stelle u Gâteaux-differenzierbar mit der Gâteaux-Ableitung

$$f'_G(u)\delta u = \cos(u(1))\delta u(1). \quad \square$$

Definition 8.3

F heißt Fréchet-differenzierbar an der Stelle $u \in U$, wenn es ein $A \in \mathcal{L}(U, V)$ und eine Abbildung $r : U \rightarrow V$ gibt mit der Eigenschaft

$$F(u + \delta u) = F(u) + A\delta u + r(\delta u), \tag{8.1}$$

wobei das Restglied die folgende Bedingung erfüllt:

$$\frac{\|r(\delta u)\|_V}{\|\delta u\|_U} \rightarrow 0 \text{ für } \delta u \rightarrow 0. \tag{8.2}$$

Für die Fréchet-Ableitung von F an der Stelle u schreiben wir im Folgenden $F'(u) = A$. Eine (oft nützlichere) alternative Definition für die Fréchet-Differenzierbarkeit findet sich z. B. in [Jeg79, 1.2.1 (15)]. Auf die Äquivalenz unserer und dieser Form wird in [Jeg79, 1.2.1 (16)] eingegangen. □

⁷⁴Linear im Sinne von: Die Ableitung ist ein linearer Operator bezüglich der Richtung δu .

⁷⁵Siehe auch [Trö09, S.45] oder [Her12, Beispiel 8.4c)].

Wir untersuchen nun den Zusammenhang zwischen den beiden neugewonnenen Ableitungsbe-
griffen:

Lemma 8.1

Ist $F : U \rightarrow V$ an einer Stelle $u \in U$ Fréchet-differenzierbar, so ist sie auch Gâteaux-differenzierbar und die beiden Ableitungen stimmen überein. \square

BEWEIS ⁷⁶ $\delta u \in U$ sei eine beliebige aber feste Richtung, $A \in \mathcal{L}(U, V)$ die Fréchet-Ableitung von F an der Stelle u . Mit Hilfe von (8.1) erhält man

$$\left\| \frac{F(u + t\delta u) - F(u)}{t} - A\delta u \right\|_V = \frac{\|r(t\delta u)\|_V}{t} = \frac{\|r(t\delta u)\|_V}{\|t\delta u\|_U} \|\delta u\|_U.$$

Aus (8.2) folgt $\frac{\|r(t\delta u)\|_V}{\|t\delta u\|_U} \rightarrow 0$ für $t \rightarrow 0$ und damit die Gâteaux-Differenzierbarkeit von F in u mit $F'_G(u) = A$. \blacksquare

Bemerkung:⁷⁷

Man merke sich also: Fréchet-Differenzierbarkeit \approx Gâteaux-Differenzierbarkeit plus Restgliedabschätzung.

Fréchet-Differenzierbarkeit beweist man durch das Suchen eines Kandidaten A für die Ableitung und zeigt

$$\frac{\|r(\delta u)\|_V}{\|\delta u\|_U} = \frac{\|F(u + \delta u) - F(u) - A\delta u\|_V}{\|\delta u\|_U} \rightarrow 0 \text{ für } \delta u \rightarrow 0.$$

Man bekommt A , indem man Richtungsableitungen bestimmt.⁷⁸ Ist $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, so handelt es sich bei $F'(u)$ um die Jacobi-Matrix.

Wir betrachten nun Beispiele für die Fréchet-Ableitung, die uns im Folgenden noch sehr nützlich sein werden:⁷⁹

Beispiel 8.2

a) Für jeden stetigen linearen Operator $A \in \mathcal{L}(U, V)$ gilt, dass er überall Fréchet-differenzierbar mit Restglied Null ist, denn

$$A(u + \delta u) = Au + A\delta u + 0.$$

Als Beispiel betrachten wir $U = H^1(\Omega)$ und $Au \stackrel{def}{=} \int_{\Omega} \nabla z \cdot \nabla u \, dx$ mit festem $z \in H^1(\Omega)$. Nach Bsp. 4.1 ist $A \in \mathcal{L}(H^1(\Omega), \mathbb{R}) = H^1(\Omega)^*$. Somit ergibt sich für die Fréchet-Ableitung

$$A'(u)\delta u = A\delta u = \int_{\Omega} \nabla z \cdot \nabla \delta u \, dx.$$

⁷⁶Siehe auch [Mey11, Lemma 9.5].

⁷⁷Siehe auch [Her12, S.39].

⁷⁸Vgl. auch Bsp. 8.1.

⁷⁹Siehe auch [Trö09, S.45 ff.], [Her12, Beispiel 8.4a) bis 8.4d)] oder [Mey11, Beispiel 9.6].

b) Wir greifen das Beispiel 8.1 erneut auf:

Die Gâteaux-Ableitung $f'_G(u)\delta u = \cos(u(1))\delta u$ ist sogar die Fréchet-Ableitung, denn:⁸⁰

$$\frac{|\sin(u(1) + \delta u(1)) - \sin(u(1)) - \cos(u(1))\delta u(1)|}{\|\delta u\|_{C[0,1]}} = \frac{1}{2} \frac{|\sin(\xi)(\delta u(1))^2|}{\|\delta u\|_{C[0,1]}}$$

ξ ist hierbei eine Zahl zwischen $u(1)$ und $u(1) + \delta u(1)$. Der \sin ist gleichmäßig beschränkt. Somit gilt:

$$\frac{1}{2} \frac{|\sin(\xi)(\delta u(1))^2|}{\|\delta u\|_{C[0,1]}} \leq \frac{1}{2} \frac{|\delta u(1)|^2}{\|\delta u\|_{C[0,1]}} \leq \frac{1}{2} \frac{\|\delta u\|_{C[0,1]}^2}{\|\delta u\|_{C[0,1]}} \rightarrow 0 \text{ für } \delta u \rightarrow 0 \text{ in } C[0,1].$$

c) $f(u) = \|u\|_H^2$ ist im Hilbertraum H stets Fréchet-Differenzierbar. Hierfür weisen wir zunächst die Gâteaux-Differenzierbarkeit nach:

$$\begin{aligned} \lim_{t \searrow 0} \frac{f(u+t\delta u) - f(u)}{t} &= \lim_{t \searrow 0} \frac{\|f(u+t\delta u) - f(u)\|_H^2 - \|u\|_H^2}{t} \\ &= \lim_{t \searrow 0} \frac{2t(u, \delta u)_H + t^2 \|\delta u\|_H^2}{t} \\ &= 2(u, \delta u)_H. \end{aligned}$$

Mit $\|f'_G(u)\|_{\mathcal{L}(H, \mathbb{R})} \leq 2\|u\|_H$ ist die Richtungsableitung linear und beschränkt, daher ist f Gâteaux-differenzierbar. Die Gâteaux-Ableitung ist nun auch eben unser Kandidat für die Fréchet-Ableitung. Die Restgliedabschätzung ergibt nun

$$\begin{aligned} \frac{|f(u+\delta u) - f(u) - f'(u)\delta u|}{\|\delta u\|_H} &= \frac{\|u+\delta u\|_H^2 - \|u\|_H^2 - 2(u, \delta u)_H}{\|\delta u\|_H} \\ &= \|\delta u\|_H \rightarrow 0 \text{ für } \delta u \rightarrow 0, \end{aligned}$$

weshalb f auch Fréchet-differenzierbar ist. Man muss hier beachten, dass die Fréchet-Ableitung ein Element aus $\mathcal{L}(H, \mathbb{R}) = H^*$ ist. Nach dem Riesz'schen Darstellungssatz 4.2 kann sie daher mit $\nabla f(u) = 2u \in H$ identifiziert werden. Somit gilt also $f'(u)\delta u = (2u, \delta u)_H$. \square

Lemma 8.2

Für die Fréchet-Ableitung gelten wie üblich Ketten- und Produktregel. \square

BEWEIS Siehe [Ruz04, 2.7]. \blacksquare

Folgendes Beispiel wird uns bei der weiteren Untersuchung von Optimalitätsbedingungen von großem Nutzen sein:⁸¹

⁸⁰Es gilt $\sin(x+h) = \sin(x) + \cos(x)h - \frac{1}{2}\sin(\xi)h^2$, mit $x = u(1)$, $h = \delta u(1)$.

⁸¹Siehe auch [Trö09, S.47], [Her12, Beispiel 8.4e)] oder [Mey11, Beispiel 9.8].

Beispiel 8.3

Es seien U und H Hilberträume, $S \in \mathcal{L}(U, H)$, $z \in H$ fest gegeben. Wir betrachten:

$$E(u) = \|Su - z\|_H^2.$$

E lässt sich als Verkettung von $G(v) = \|v\|_V^2$ (Fréchet-differenzierbar nach Bsp. 8.2c)) und $F(u) = Su - z$ (Fréchet-differenzierbar nach Bsp. 8.2a)) darstellen. Es gilt: $G'(v)\delta v = 2(v, \delta v)_H$ und $F'(u)\delta u = S\delta u$, da S linear ist.

Nach Lemma 8.2 ergibt sich dann:

$$E'(u)\delta u = G'(F(u))(F'(u)\delta u) = 2(F(u), F'(u)\delta u)_H = 2(Su - z, S\delta u)_H.$$

Im folgenden Paragraphen werden wir sehen, dass sich dieser Ausdruck mit Hilfe des sogenannte adjungierten Operators weiter vereinfachen lässt. □

Auf all dem aufbauend kann man noch die Differenzierbarkeit von bestimmten Superpositionsoperatoren untersuchen (den sogenannten Nemyzki-Operatoren), die v.a. bei der Diskussion von nichtlinearen Aufgabenstellungen eine entscheidende Rolle spielen, für unsere Zwecke aber nicht von Bedeutung sind. Es sei auf [Trö09, 4.3] verwiesen.

§9 Adjungierte Operatoren

Wie bereits angedeutet beschäftigen wir uns nun mit sogenannten adjungierten Operatoren, welche für eine handlichere Formulierung der folgenden notwendigen Optimalitätsbedingungen erfordert werden.⁸²

Es seien U, V normierte lineare Räume, $A : U \rightarrow V$ ein linearer und stetiger Operator, $f \in V^*$. $\langle Au, f \rangle_{V, V^*}$ liefert somit für alle $u \in U$ eine reelle Zahl, es handelt sich also um ein Funktional auf U , welches wir im Folgenden als g bezeichnen. g ist linear, da es f und A sind, außerdem ist g stetig, denn es ist beschränkt:

$$\|g\|_{\mathcal{L}(U, \mathbb{R})} = \sup_{\|u\|_U=1} |\langle Au, f \rangle_{V, V^*}| \leq \sup_{\|u\|_U=1} \|Au\|_V \|f\|_{V^*} \leq \|A\|_{\mathcal{L}(U, \mathbb{R})} \|f\|_{V^*},$$

daher gilt $g \in U^*$. Somit existiert eine Abbildung $A^* : f \mapsto g$ für die gilt:

$$\langle u, A^* f \rangle_{U, U^*} = \langle g, u \rangle_{U^*, U} = \langle Au, f \rangle_{V, V^*}.$$

Dies führt zu:

Definition 9.1

U, V seien normierte lineare Räume und $A \in \mathcal{L}(U, V)$. Eine Abbildung $A^* : U^* \rightarrow V^*$ heißt der zu A adjungierte Operator, wenn für alle $u \in U, f \in V^*$ gilt:

$$\langle Au, f \rangle_{V, V^*} = \langle u, A^* f \rangle_{U, U^*}. \quad \square$$

⁸²Siehe auch [Trö09, S.47 f.] oder [Mey11, S.51].

Folgendes Lemma liefert uns eine Aussage über die Existenz solcher adjungierter Operatoren:

Lemma 9.1

U, V seien normierte lineare Räume, $A \in \mathcal{L}(U, V)$ beliebig. Dann existiert ein adjungierter Operator $A^* \in \mathcal{L}(V^*, U^*)$ und es gilt:

$$\|A^*\|_{\mathcal{L}(V^*, U^*)} = \|A\|_{\mathcal{L}(U, V)}. \quad \square$$

BEWEIS Siehe [Alt02, 10.1] und [Lue69, 6.5 Theorem 1]. ■

Aus dem Riesz'schen Darstellungssatz 4.2 folgt für U, V Hilberträume folgende 'Abart' von Def. 9.1:

Definition 9.2

U, V seien Hilberträume, $A \in \mathcal{L}(U, V)$. Die Abbildung $A^{(*)} : V \rightarrow U$ heißt der zu A Hilbertraum-adjungierte Operator, wenn für alle $u \in U, v \in V$ gilt:

$$(Au, v)_V = (u, A^{(*)}v)_U.$$

Für $U = V$ bezeichnet man A als selbstadjungiert, wenn $A = A^{(*)}$.⁸³ □

Wir verdeutlichen den Zusammenhang zwischen adjungiertem und Hilbertraum-adjungiertem Operator um eine ähnliche Aussage zu Lemma 9.1 herzuleiten.⁸⁴

Sei $A \in \mathcal{L}(U, V)$ ein Operator in den Hilberträumen U und V mit adjungiertem Operator $A^* \in \mathcal{L}(V^*, U^*)$ und Hilbertraum-adjungiertem Operator $A^{(*)} : V \rightarrow U$. Wir bezeichnen für ein festes $v \in V$ das durch v induzierte Funktional mit $f_v \in V^*$. Es gilt dann nach Definition:

$$(u, A^{(*)}v)_U = (Au, v)_V = \langle Au, f_v \rangle_{V, V^*} = \langle u, A^* f_v \rangle_{U, U^*} \quad \forall u \in U.$$

$A^{(*)}v \in U$ ist somit der eindeutige Riesz-Repräsentant des Funktionals $A^* f_v \in U^*$. Nach dem Riesz'schen Darstellungssatz ergibt sich für die Norm von $A^{(*)}$

$$\begin{aligned} \|A^{(*)}\|_{\mathcal{L}(V, U)} &= \sup_{\|v\|_V=1} \|A^{(*)}v\|_U \\ &= \sup_{\|v\|_V=1} \|A^* f_v\|_{U^*} \\ &\leq \sup_{\|v\|_V=1} \|A^*\|_{\mathcal{L}(V^*, U^*)} \|f_v\|_{V^*} \\ &= \|A^*\|_{\mathcal{L}(V^*, U^*)} \quad (\text{wegen } \|f_v\|_{V^*} = \|v\|_V). \end{aligned} \quad (9.1)$$

⁸³In der Literatur wird der zu A Hilbertraum-adjungierte Operator meist mit A^* bezeichnet. Wegen Verwechslungsgefahr mit dem adjungierten Operator wird hier obige Notation verwendet.

⁸⁴Siehe auch [Mey11, S.52].

Andererseits, ist $f \in V^*$ gegeben, so existiert nach dem Riesz'schen Darstellungssatz ein Riesz-Repräsentant $v_f \in V$, s.d. $(v, v_f) = \langle v, f \rangle \forall v \in V$ und $\|v_f\|_V = \|f\|_{V^*}$. Analog zu (9.1) folgt aufgrund der Tatsache, dass $A^{(*)}v_f$ wieder Riesz-Repräsentant von A^*f ist

$$\|A^*\|_{\mathcal{L}(V^*, U^*)} = \sup_{\|f\|_{V^*}=1} \|A^*f\|_{U^*} = \sup_{\|f\|_{V^*}=1} \|A^{(*)}v_f\|_U \leq \|A^{(*)}\|_{\mathcal{L}(V, U)}.$$

Somit gilt also:

Lemma 9.2

U, V seien Hilberträume, $A \in \mathcal{L}(U, V)$. Dann existiert ein eindeutiger Hilbertraum-adjungierter Operator $A^{(*)} \in \mathcal{L}(U, V)$ mit

$$\|A^{(*)}\|_{\mathcal{L}(V, U)} = \|A^*\|_{\mathcal{L}(V^*, U^*)} = \|A\|_{\mathcal{L}(U, V)}. \quad \square$$

Die Eindeutigkeit ergibt sich bei den Lemmata 9.1 und 9.2 natürlich bereits aus den jeweiligen Definitionen. Im Sinne der Überlegungen nach Def. 9.2 unterscheiden wir von nun an nicht mehr zwischen den beiden adjungierten-Operator Begriffen und schreiben in beiden Fällen A^* .

Hierzu noch folgendes Beispiel um das ursprüngliche Vorhaben zu verdeutlichen:⁸⁵

Beispiel 9.1

Im Bsp. 8.3 hatten wir gezeigt:

Für $E(u) = \|Su - z\|_H^2$ gilt:

$$E'(u)\delta u = 2(Su - z, S\delta u)_H. \quad (9.2)$$

Mit Hilfe des adjungierten Operators kann man (9.2) nun umformen:

$$E'(u)\delta u = 2(S^*(Su - z), \delta u)_U.$$

Wir können also $E'(u) \subset U^*$ nach dem Riesz'schen Darstellungssatz 4.2 identifizieren mit

$$\nabla E(u) = 2S^*(Su - z) \in U.$$

Wie wir noch sehen werden, spielt diese Darstellung wegen der Isolierung von δu eine herausragende Rolle.

Als weitere Standardbeispiele für adjungierte Operatoren seien z. B. die Fredholm-⁸⁶ und die Volterra-Integraloperatoren⁸⁷ genannt. □

§10 Notwendige und hinreichende Optimalitätsbedingungen

Unser letzter großer Theorieblock ist nun beendet, sodass wir uns nun mit den bereits erwähnten Optimalitätsbedingungen beschäftigen können. Wie bereits bei der Existenz optimaler Lösungen betrachten wir erst eine allgemeinere Aufgabenstellung und benutzen diese Ergebnisse um Aussagen über das Modellproblem (1.5) machen zu können.

⁸⁵Siehe auch [Trö09, S.47].

⁸⁶Siehe z. B. [Lue69, 6.5 Example 2 und Example 3].

⁸⁷Siehe z. B. [Lue69, 6.5 Example 4].

§10.1 Quadratische Optimierungsaufgabe im Hilbertraum

Wir betrachten zunächst

$$\begin{aligned} & \text{Minimiere} && f(u) \\ & \text{unter} && u \in U_{ad} \end{aligned} \tag{10.1}$$

mit dem normierten linearen Raum U , $f : U \rightarrow \mathbb{R}$, $U_{ad} \subset U$ konvex.

Im Gegensatz zu Kapitel III, in dem wir global optimale Lösungen diskutiert haben, beschränken wir uns hier zunächst auf lokal optimale Lösungen.

Definition 10.1

$\bar{u} \in U_{ad}$ heißt lokal optimale Lösung für (10.1), wenn es ein $r > 0$ gibt, s.d. $f(\bar{u}) \leq f(u) \forall u \in U_{ad} \cap \overline{B_r(\bar{u})}$ gilt, d.h. $\forall u \in U_{ad}$ mit $\|u - \bar{u}\|_U \leq r$. □

Ein erstes großes Ergebnis liefert uns folgender Satz:

Satz 10.1 (Notwendige Optimalitätsbedingung)

U sei normierter linearer Raum, $U_{ad} \in U$ konvex und $\bar{u} \in U_{ad}$ lokal optimale Lösung von (10.1), $f : U_{ad} \rightarrow \mathbb{R}$ richtungsdifferenzierbar in \bar{u} auf U_{ad} , d.h. richtungsdifferenzierbar in \bar{u} in alle Richtungen δu , für die es ein $u \in U_{ad}$ gibt mit $\delta u = u - \bar{u}$. Dann gilt die Variationsungleichung

$$\delta f(\bar{u}; u - \bar{u}) \geq 0 \quad \forall u \in U_{ad}. \tag{10.2}$$

□

BEWEIS ⁸⁸ Für $u \in U_{ad}$ beliebig definieren wir die konvexe Linearkombination

$$u(t) \stackrel{\text{def}}{=} \bar{u} - t(u - \bar{u}).$$

Es gilt: $u(t) \in U_{ad} \forall t \in (0, 1]$, da U_{ad} konvex. Für hinreichend kleines t folgt zudem $u(t) \in \overline{B_r(\bar{u})}$, wobei $\overline{B_r(\bar{u})}$ die Umgebung von \bar{u} ist, in der \bar{u} lokal optimal ist. Für hinreichend kleine $t > 0$ folgt daraus dann

$$f(u(t)) \geq f(\bar{u}),$$

bzw.

$$\frac{1}{t} (f(\bar{u} + t(u - \bar{u})) - f(\bar{u})) \geq 0.$$

Nach Voraussetzung ist f richtungsdifferenzierbar in \bar{u} in alle Richtungen δu , daher können wir zur Grenze übergehen und erhalten somit (10.2). ■

Aus Def. 8.2 folgt: Wenn f Gâteaux-differenzierbar ist in \bar{u} , dann lautet (10.2):

$$f'_G(\bar{u})(u - \bar{u}) \geq 0 \quad \forall u \in U_{ad}.$$

In Satz 10.1 wurde explizit nicht gefordert, dass f konvex ist. Sollte dies der Fall sein, so ist die Variationsungleichung auch hinreichend für lokale Optimalität:

⁸⁸Siehe auch [Trö09, Lemma 2.21], [Mey11, Satz 11.2] oder [Her12, Satz 10.1].

Satz 10.2 (Hinreichende Optimalitätsbedingung)

U sei normierter linearer Raum, $U_{ad} \in U$ konvex, $f : U_{ad} \rightarrow \mathbb{R}$ in alle Richtungen richtungsdifferenzierbar auf U_{ad} . f sei konvex auf U_{ad} .⁸⁹ Dann gilt: Jedes $\bar{u} \in U_{ad}$, das (10.2) erfüllt, ist global optimal Lösung von (10.1). □

BEWEIS ⁹⁰ $\bar{u} \in U_{ad}$ erfülle die Variationsungleichung (10.2) und es sei $u \in U_{ad}$ beliebig. Aufgrund der Konvexität von f gilt für $t \in (0, 1]$

$$\begin{aligned} f(tu + (1-t)\bar{u}) &\leq tf(u) + (1-t)f(\bar{u}) \\ \Rightarrow \frac{f(\bar{u} + t(u-\bar{u})) - f(\bar{u})}{t} &\leq f(u) - f(\bar{u}) \end{aligned}$$

und durch den Grenzübergang $t \searrow 0$

$$f(u) - f(\bar{u}) \geq \delta f(\bar{u}; u - \bar{u}).$$

Da die rechte Seite nach Voraussetzung > 0 ist, folgt die Behauptung. ■

Um die Aussagen der Sätze 10.1 und 10.2 für unsere Zwecke nutzbar zu machen, konkretisieren wir (10.1) nun sukzessive. Zunächst betrachten wir folgendes Optimalitätsproblem im Hilbertraum:

$$\begin{aligned} \text{Minimiere} \quad & f(u) \stackrel{\text{def}}{=} \frac{1}{2} \|Su - y_d\|_H^2 + \frac{\lambda}{2} \|u\|_U^2 \\ \text{unter} \quad & u \in U_{ad} \end{aligned} \tag{10.3}$$

mit U, H Hilberträume, $S \in \mathcal{L}(U, H)$, $y_d \in H$. Wir bezeichnen f wieder als reduziertes Zielfunktional.

Satz 10.3

Seien U, H Hilberträume, U_{ad} konvex, $S \in \mathcal{L}(U, H)$, $y_d \in H$, $\lambda \geq 0$. $\bar{u} \in U_{ad}$ ist genau dann globales Minimum von (10.3) wenn die Variationsungleichung

$$(\nabla f(\bar{u}), u - \bar{u})_U = (S^*(S\bar{u} - y_d) + \lambda\bar{u}, u - \bar{u})_U \geq 0 \quad \forall u \in U_{ad}$$

erfüllt ist. □

BEWEIS ⁹¹ Die Fréchet-Ableitung von f ist nach Bsp. 9.1 gegeben durch

$$f'(u)\delta u = (S^*(Su - y_d), \delta u)_U + \lambda(u, \delta u)_U.$$

Durch die Sätze 10.1 und 10.2 bzw. die Identifizierung von $f'(u) \subset U^*$ mit

$$\nabla f(u) = 2S^*(Su - y_d) \in U$$

folgt dann sofort die Behauptung. ■

⁸⁹Es ist zu beachten, dass die Richtungs-differenzierbarkeit von f dann bereits aus deren Konvexität folgt.

⁹⁰Siehe auch [Trö09, Lemma 2.21], [Mey11, Satz 11.4] oder [Her12, Satz 10.1].

⁹¹Siehe auch [Her12, Satz 10.2].

Die Freistellung der sogenannten Variationsrichtung $u - \bar{u}$ wird noch sehr von Nutzen sein. Wir müssen allerdings noch klären, was S^* im Fall des Modellproblems ist. Darauf werden wir später noch sehr genau eingehen.

Aus Satz 10.3 folgt direkt:

Korollar 10.1

Falls $U_{ad} = U$, so ist unter den Voraussetzungen von Satz 10.3 $\bar{u} \in U$ genau dann globales Minimum, wenn gilt:

$$\nabla f(\bar{u}) = S^*(S\bar{u} - y_d) + \lambda\bar{u} = 0. \quad \square$$

Zur Verdeutlichung der bisherigen Ergebnisse betrachten wir folgendes Beispiel:⁹²

Beispiel 10.1 (Projektion im Hilbertraum)

Wir betrachten

$$\min_{v \in C} \frac{1}{2} \|v - u\|_H^2 \quad (10.4)$$

im Hilbertraum H , $C \subset H$ nichtleere, abgeschlossene und konvexe Teilmenge. Analog zum Beweis von Lemma 7.1 kann man zeigen, dass die Zielfunktion in (10.4) konvex (sogar strikt konvex) ist. Daraus folgt, dass nach Satz 7.2 eine eindeutige Lösung von (10.4) existiert. Es handelt sich dabei um die Projektion von u auf C ($\Pi_C u$). Nach Bsp. 8.2c) ist die Zielfunktion Fréchet-differenzierbar und aus den Sätzen 10.1 und 10.2 folgt wegen der Konvexität von C , dass $\Pi_C u \in C$ durch die folgende Variationsungleichung eindeutig charakterisiert wird:

$$(\Pi_C u - u, v - \Pi_C u)_H \geq 0 \quad \forall v \in C. \quad (10.5)$$

Für Aussagen über den sogenannten Projektionsoperator $\Pi_C u : H \rightarrow C$ sei auf [H⁺09, Lemma 1.10, Lemma 1.11 und Corollary 1.2] verwiesen. □

§10.2 Die adjungierte Gleichung

Wir kehren nun zum Modellproblem (1.5) zurück und schreiben es von nun an auf folgende Art und Weise:

$$\left. \begin{array}{l} \text{Minimiere } J(y, u) = \frac{1}{2} \|y - y_d\|_{L^2(\Omega)}^2 + \frac{\lambda}{2} \|u\|_{L^2(\Omega)}^2 \\ \text{unter } \left\{ \begin{array}{ll} -\Delta y = u & \text{in } \Omega \\ y(x) = 0 & \text{auf } \Gamma \end{array} \right. \\ \text{und } u_a \leq u \leq u_b \text{ fast überall in } \Omega. \end{array} \right\} \quad (10.6)$$

mit den Voraussetzungen aus Anmerkung 7.1.

⁹²Siehe auch [Mey11, Beispiel 11.9].

Mit dem Steuerungs-Zustands-Operator, dem Lösungsoperator der Poissongleichung, hat (10.6) die Form von (10.3) s.d. Satz 10.3 mit $U = H = L^2(\Omega)$ anwendbar ist. Für (10.6) ist eine Steuerung $\bar{u} \in U_{ad}$ somit genau dann global optimal, falls die Variationsungleichung

$$(S^*(S\bar{u} - y_d) + \lambda\bar{u}, u - \bar{u})_{L^2(\Omega)} \geq 0 \quad \forall u \in U_{ad} \quad (10.7)$$

erfüllt ist. Durch (10.7) wird das globale Optimum eindeutig charakterisiert, weil f strikt konvex ist.⁹³

Es bleibt die offene Frage nach Form und Bedeutung von S^* . Wir werden sehen, dass S^* wie auch S der Lösungsoperator einer PDE ist, der sogenannten adjungierten Gleichung, welche wir nun betrachten werden.

Für eine zunächst allgemeine Form sei V Hilbertraum, $a : V \times V \rightarrow \mathbb{R}$ eine beschränkte und koerzive Bilinearform. Für gegebenes $F \in V^*$ betrachten wir

$$a[y, v] = \langle F, v \rangle_{V^*, V} \quad \forall v \in V. \quad (10.8)$$

Das Lemma von Lax-Milgram 5.2 besagt, dass (10.8) für alle $F \in V^*$ eine eindeutige Lösung $y \in V$ besitzt. Im Fall der Poissongleichung galt $V = H_0^1(\Omega)$ und für a und F :

$$a[y, v] = \int_{\Omega} \nabla y \cdot \nabla v \, dx \quad \text{und} \quad \langle F, v \rangle_{V^*, V} = \int_{\Omega} f v \, dx. \quad (10.9)$$

Die adjungierte Gleichung zu (10.8) entsteht durch Vertauschen von Lösungs- und Testfunktion, s.d. sich

$$a[v, p] = \langle F, v \rangle_{V^*, V} \quad \forall v \in V \quad (10.10)$$

ergibt. Es stellt sich zunächst die Frage nach der Existenz einer Lösung für (10.10).

Lemma 10.1

Sei $a : V \times V \rightarrow \mathbb{R}$ koerzive und beschränkte Bilinearform, dann existiert für jedes $F \in V^$ genau eine Lösung $p \in V$ von (10.10) und es gilt $\|p\|_V \leq c\|F\|_{V^*}$.* □

Der Beweis hierzu ist sehr umfangreich und technisch aufwändig, wir geben hier eine Skizze hierfür an, welche die grundlegenden Ideen verdeutlichen soll:

BEWEIS ⁹⁴ Mit

$$\langle Ay, v \rangle_{V^*, V} = a[y, v] \quad \forall y, v \in V$$

ist es möglich einen linearen und stetigen Operator $A : V \rightarrow V^*$ zu definieren. Dies ist zulässig, da Ay jedem $v \in V$ so linear und stetig eine reelle Zahl zuordnet. Es gilt also $Ay \in V^*$. Die Stetigkeit von A folgt aus der Beschränktheit der Bilinearform. (10.8) entspricht somit der folgenden

⁹³Siehe auch [Mey11, S.56].

⁹⁴Siehe auch [Mey11, Lemma 11.10].

Operatorgleichung:

$$Ay = F \text{ in } V^*.$$

Aus dem Satz der stetigen Inversen⁹⁵ folgt aus der eindeutigen Lösbarkeit von (10.8) die Invertierbarkeit von A und $A^{-1} \in \mathcal{L}(V, V^*)$. Nach Def. 9.1 ist der adjungierte Operator $A^* : V^{**} = V \rightarrow V^*$ gegeben durch

$$\langle v, A^* p \rangle_{V, V^*} = \langle Av, p \rangle_{V, V^*} = a[v, p] \quad \forall p, v \in V.$$

$A^* p = F$ in V entspricht als Operatorgleichung dadurch also gerade der adjungierten Gleichung (10.10). Aus der stetigen Invertierbarkeit von A folgt die von A^* mit $(A^*)^{-1} = (A^{-1})^*$,⁹⁶ s.d. (10.10) für alle $F \in V^*$ eine eindeutige Lösung besitzt, die sich durch

$$\|p\|_V \leq \|(A^*)^{-1}\|_{\mathcal{L}(V^*, V)} \|F\|_{V^*} = \|A^{-1}\|_{\mathcal{L}(V^*, V)} \|F\|_{V^*}$$

abschätzen lässt. Mit $c = \|A^{-1}\|_{\mathcal{L}(V^*, V)}$ folgt die Behauptung. ■

Zwar benötigen wir die Aussage von Lemma 10.1 nicht für die Untersuchung des Modellproblems (wie man leicht sieht ist die Bilinearform der Poissongleichung symmetrisch, daher folgt die Lösbarkeit der adjungierten Gleichung aus Satz 5.1), doch werden wir die allgemeine Aussage im Laufe unserer Untersuchungen noch benötigen.⁹⁷

Sei nun $V \hookrightarrow L^2(\Omega)$ (etwa $V = H_0^1(\Omega)$). Es fällt auf, dass die rechte Seiten in (10.8) und (10.10) folgende speziellen Strukturen besitzen:

$$a[y, v] = \int_{\Omega} f v \, dx \quad \forall v \in V \tag{10.11}$$

$$a[v, p] = \int_{\Omega} g v \, dx \quad \forall v \in V \tag{10.12}$$

mit $f, g \in L^2(\Omega)$. Wegen $V \hookrightarrow L^2(\Omega)$ stellen die rechten Seiten Elemente aus V^* dar⁹⁸, daher existieren für alle f, g eindeutige Lösungen $y, p \in V$ für (10.11) und (10.12). Wie in Def. 7.2 ist es möglich lineare und stetige Lösungsoperatoren einzuführen:

Bei (10.11) wird dieser mit $G \in \mathcal{L}(L^2(\Omega), V)$, bei (10.12) mit $G_p \in \mathcal{L}(L^2(\Omega), V)$ bezeichnet.⁹⁹ Der Notation aus Kapitel III folgend sei $E : V \rightarrow L^2(\Omega)$ der Einbettungsoperator und $S = E \circ G$.

Dies führt uns nun zu folgender entscheidenden Aussage über S^* :

⁹⁵Siehe [Alt02, 5.8], folgt aus dem Satz von der offenen Abbildung [Alt02, 5.7].

⁹⁶Siehe [Lue69, 6.5 Proposition 1].

⁹⁷Siehe auch [Mey11, Bemerkung 11.11].

⁹⁸Siehe 5.7.

⁹⁹Im symmetrischen Fall der Poissongleichung wäre $G = G_p$.

Lemma 10.2

Der adjungierte Operator $S^* : L^2(\Omega) \rightarrow L^2(\Omega)$ ist gegeben durch

$$S^* = E \circ G_p,$$

d.h. für gegebenes $g \in L^2(\Omega)$ ist $p = S^*g$ die eindeutige schwache Lösung in V von (10.12). \square

BEWEIS ¹⁰⁰ Seien $f, g \in L^2(\Omega)$ beliebig. Wir betrachten nun die Gleichung, die zu $y = Sf$ gehört, also gerade (10.11). Setzen wir $p = G_p g \in V$ als Testfunktion ein, so erhalten wir

$$a[y, p] = \int_{\Omega} fp \, dx.$$

Aus (10.10) ergibt sich umgekehrt

$$a[y, p] = \int_{\Omega} gy \, dx$$

wenn man $y = Gf$ als Testfunktion einsetzt. Subtrahiert man beide Gleichungen, so erhält man

$$\int_{\Omega} gy \, dx = \int_{\Omega} fp \, dx.$$

Daraus folgt dann:

$$\begin{aligned} (S^*g, f)_{L^2(\Omega)} &= (g, Sf)_{L^2(\Omega)} \\ &= \int_{\Omega} gSf \, dx = \int_{\Omega} gy \, dx = \int_{\Omega} pf \, dx = ((E \circ G_p)g, f)_{L^2(\Omega)}. \end{aligned}$$

Wegen der Eindeutigkeit des adjungierten Operators nach Lemma 9.1 folgt daraus dann $S^* = E \circ G_p$. \blacksquare

Definition 10.2

Im Falle einer symmetrischen Bilinearform gilt wie bereits erwähnt $G_p = G$ und damit auch $S^* = S$. In diesem Fall bezeichnet man den Steuerungs-Zustands-Operator als selbstadjungiert. \square

Anmerkung 10.1

Wir haben S^* anhand einer allgemeinen Bilinearform hergeleitet, also völlig unabhängig von der Form der PDE. Die adjungierte Gleichung erhält man also unabhängig von der Problemstellung durch Vertauschung von Lösung und Testfunktion. Oft wird dies als Methode der Transposition bezeichnet. Alternativ sei auch auf die sogenannte formale Lagrange-Technik hingewiesen.¹⁰¹ \square

¹⁰⁰Siehe auch [Mey11, Lemma 11.12].

¹⁰¹Betrachte hierfür z. B. [Trö09, 2.10].

Kehren wir zum Modellproblem zurück, so sehen wir, dass durch die Form von (10.9) die adjungierte Gleichung (10.12) folgende Form annimmt:

$$\int_{\Omega} \nabla p \cdot \nabla v \, dx = \int_{\Omega} g v \, dx \quad \forall v \in H_0^1(\Omega).$$

Formal¹⁰² schreiben wir nun diese Gleichung in 'starker' Form:

$$\begin{aligned} -\Delta p &= g && \text{in } \Omega \\ p &= 0 && \text{auf } \Gamma \end{aligned}$$

und formulieren entsprechend die Variationsformulierung (10.7) um:

Definition 10.3

Die eindeutige schwache Lösung der adjungierten Gleichung

$$\begin{aligned} -\Delta p &= \bar{y} - y_d && \text{in } \Omega \\ p &= 0 && \text{auf } \Gamma \end{aligned} \tag{10.13}$$

heißt der zu \bar{y} gehörige adjungierte Zustand. □

Da $\bar{y} - y_d \in L^2(\Omega)$ gilt, existiert nach Satz 5.1 bzw. Lemma 10.1 genau eine eindeutige Lösung $p \in H_0^1(\Omega)$ der adjungierten Gleichung.¹⁰³ Wegen Lemma 10.2 gilt

$$S^*(S\bar{u} - y_d) = S^*(\bar{y} - y_d) = p.$$

Daher lautet die Variationsungleichung (10.7)

$$(p + \lambda \bar{u}, u - \bar{u})_{L^2(\Omega)} \geq 0 \quad \forall u \in U_{ad}.$$

Fassen wir die Ergebnisse dieses Kapitels zusammen, erhalten wir folgende weitreichende Aussage:

Aus Satz 10.3 folgen die notwendigen und hinreichenden Optimalitätsbedingungen für (10.6):

Satz 10.4 (Notwendige und hinreichende Optimalitätsbedingungen)

a) Ist $\bar{u} \in U_{ad}$ optimale Steuerung von (10.6) und \bar{y} zugehöriger optimaler Zustand, dann existiert genau eine schwache Lösung $p \in H_0^1(\Omega)$ der adjungierten Gleichung (10.13) und diese erfüllt die Variationsungleichung

$$\int_{\Omega} (p + \lambda \bar{u})(u - \bar{u}) \, dx \geq 0 \quad \forall u \in U_{ad}. \tag{10.14}$$

b) Erfüllt umgekehrt ein $\bar{u} \in U_{ad}$ mit zugehörigem Zustand \bar{y} und adjungiertem Zustand p die Ungleichung (10.14), so ist \bar{u} optimale Lösung von (10.6). □

¹⁰²Siehe Anmerkung 5.2.

¹⁰³Siehe auch [Mey11, S.59].

Es folgt also direkt:

Korollar 10.2

Eine Steuerung $\bar{u} \in U_{ad}$ ist genau dann optimal für (10.6), wenn sie gemeinsam mit $\bar{y}, p \in H_0^1(\Omega)$ das folgende Optimalitätssystem erfüllt:

$$\begin{aligned} -\Delta \bar{y} &= \bar{u} \quad , \quad -\Delta p = \bar{y} - y_d \quad \text{in } \Omega \\ \bar{y} &= 0 \quad , \quad p = 0 \quad \text{für } \Gamma \\ (p + \lambda \bar{u}, v - \bar{u})_{L^2(\Omega)} &\geq 0 \quad \forall v \in U_{ad}. \end{aligned} \tag{10.15}$$

□

§10.3 Die Projektionsformel

(10.14) bzw. (10.15) lassen sich zu zahlreichen äquivalenten Formulierungen umformen.¹⁰⁴ Wir gehen hier auf die sogenannte Projektionsformel näher ein, welche wir später noch benötigen werden und die eine punktweise Untersuchung der Variationsungleichung notwendig macht.

Lemma 10.3

(10.14) gilt genau dann, wenn für fast alle $x \in \Omega$ gilt:

$$\bar{u}(x) \begin{cases} = u_a(x) & , \text{ für } p(x) + \lambda \bar{u}(x) > 0 \\ \in [u_a(x), u_b(x)] & , \text{ für } p(x) + \lambda \bar{u}(x) = 0 \\ = u_b(x) & , \text{ für } p(x) + \lambda \bar{u}(x) < 0. \end{cases} \tag{10.16}$$

BEWEIS ¹⁰⁵ Wir kürzen zunächst ab:

$$z(x) \stackrel{def}{=} p(x) + \lambda \bar{u}(x)$$

und argumentieren per Widerspruchsannahme:

O.B.d.A. betrachten wir lediglich die untere Schranke. Für die obere Schranke argumentiert man analog. Es gibt also eine Menge $E \subset \Omega$ mit $|E| > 0$ für die gilt:

$$z(x) > 0 \text{ und } \bar{u}(x) > u_a(x) \text{ fast überall in } E.$$

Wir definieren

$$\tilde{u}(x) \stackrel{def}{=} \begin{cases} u_a(x) & , \text{ falls } x \in E \\ \bar{u}(x) & , \text{ falls } x \notin E. \end{cases}$$

Es gilt $\tilde{u} \in U_{ad}$ wegen $\bar{u} \in U_{ad}$. \tilde{u} ist daher zulässig für die Variationsungleichung (10.14). Andererseits berechnen wir

$$\int_{\Omega} z(x)(\tilde{u}(x) - \bar{u}(x))dx = \int_E \underbrace{z(x)}_{>0} \underbrace{(u_a(x) - \bar{u}(x))}_{<0} dx < 0$$

¹⁰⁴Siehe etwa das Maximumsprinzip [Trö09, 2.27].

¹⁰⁵Siehe auch [Trö09, Lemma 2.26] oder [Mey11, Lemma 11.19].

was einen Widerspruch zu (10.14) darstellt.

Es bleibt z.Z.: (10.16) \Rightarrow (10.14):

$x \in \Omega$ sei so gewählt, dass (10.16) gilt. Zudem sei $v \in [u_a(x), u_b(x)]$ beliebig.

Aus (10.16) folgt:

$$z(x)(v - \bar{u}(x)) = \begin{cases} z(x)(v - u_b(x)) > 0 & , \text{ für } z(x) < 0 \\ 0 & , \text{ für } z(x) = 0 \\ z(x)(v - u_a(x)) > 0 & , \text{ für } z(x) > 0. \end{cases}$$

Also ergibt sich für beliebiges $v \in [u_a(x), u_b(x)]$:

$$z(x)(v - \bar{u}(x)) \geq 0. \quad (10.17)$$

Diese Ungleichung ist fast überall erfüllt, da (10.16) in fast allen Punkten $x \in \Omega$ gilt.

Sei nun $u \in U_{ad}$ beliebig, dann ist $u(x) \in [u_a(x), u_b(x)]$ in fast allen Punkten $x \in \Omega$. Aufgrund von (10.17) folgt daraus:

$$z(x)(u(x) - \bar{u}(x)) \geq 0 \text{ fast überall in } \Omega.$$

(10.14) folgt schließlich aus der Monotonie des Integrals. ■

Aus dem Beweis folgt direkt:

Korollar 10.3

Die Variationsungleichung (10.14) ist äquivalent zu folgender punktweisen Variationsungleichung in \mathbb{R} :

$$(p(x) + \lambda \bar{u}(x))(v - \bar{u}(x)) \geq 0 \quad \forall v \in [u_a(x), u_b(x)]. \quad (10.18)$$

Bemerkung:¹⁰⁶

Für λ in (10.16) gibt es folgende Unterscheidungen:

a) Für $\lambda = 0$ liefert (10.16) keine Aussage auf

$$\Omega_0 = \{x \in \Omega : p(x) = 0\}.$$

Ist Ω_0 eine Nullmenge, so nimmt $\bar{u}(x)$ fast überall nur die Werte $\{u_a(x), u_b(x)\}$ an.¹⁰⁷

b) Für $\lambda > 0$ kann man für $p(x) + \lambda \bar{u}(x) = 0$ schließen:

$$\bar{u}(x) = -\frac{1}{\lambda} p(x).$$

Im Folgenden betrachten wir die Projektion von \mathbb{R} auf ein kompaktes Intervall $[a, b]$:

$$P_{[a,b]}(u) \stackrel{def}{=} \min\{b, \max\{a, u\}\}.$$

¹⁰⁶Siehe auch [Trö09, S.55] oder [Mey11, Bemerkung 11.21].

¹⁰⁷Dies führt zu einer sogenannten Bang-Bang-Steuerung, siehe dazu [Trö09, 2.8.2 und 2.9.1].

Satz 10.5

Sei $\lambda > 0$. Dann ist \bar{u} genau dann optimale Steuerung von (10.6), wenn mit dem zugehörigen Zustand \bar{y} und adjungiertem Zustand p die Projektionsformel

$$\bar{u}(x) = P_{[u_a(x), u_b(x)]} \left(-\frac{1}{\lambda} p(x) \right) \quad (10.19)$$

für fast alle $x \in \Omega$ erfüllt ist. □

BEWEIS ¹⁰⁸ Nach Konstruktion ist (10.19) äquivalent zu

$$\bar{u}(x) = \begin{cases} u_a(x), & \text{falls } -\frac{1}{\lambda} p(x) < u_a(x) \\ -\frac{1}{\lambda} p(x), & \text{falls } -\frac{1}{\lambda} p(x) \in [u_a(x), u_b(x)] \\ u_b(x), & \text{falls } -\frac{1}{\lambda} p(x) > u_b(x). \end{cases}$$

Dies entspricht gerade (10.16) und somit nach Lemma 10.3 den notwendigen und hinreichenden Bedingungen von (10.6). ■

Korollar 10.4

Im Fall $\lambda > 0$ und ohne Beschränkungen $U_{ad} = L^2(\Omega)$ folgt aus (10.18)

$$\bar{u} = -\frac{1}{\lambda} p \text{ fast überall in } \Omega.$$

Daraus folgt, dass man die Steuerung aus den notwendigen Bedingungen eliminieren kann, s.d. man das folgende sogenannte gekoppelte Randwertproblem¹⁰⁹ erhält:

$$\begin{aligned} -\Delta y = -\frac{1}{\lambda} p & \quad , \quad -\Delta p = y - y_d & \text{ in } \Omega \\ y = 0 & \quad , \quad p = 0 & \text{ für } \Gamma. \end{aligned}$$

□

Das Optimalitätssystem lässt Rückschlüsse auf die Regularität der optimalen Steuerung zu:

Satz 10.6 ($H^1(\Omega)$ -Regularität der optimalen Steuerung)

Zusätzlich zu den Annahmen an (10.6) seien $u_a, u_b \in H^1(\Omega)$ und $\lambda > 0$. Dann ist die optimale Steuerung von (10.6) eine Funktion aus $H^1(\Omega)$. □

¹⁰⁸Siehe auch [Trö09, Satz 2.28] oder [Mey11, Satz 11.22].

¹⁰⁹Man bezeichnet diese Formulierung auch als indirekte Methode. Wie wir noch sehen werden verwendet man sie als Basis numerischer Lösungsverfahren.

BEWEIS ¹¹⁰ Wir definieren

$$\Omega_a \stackrel{def}{=} \{x \in \Omega : u(x) = u_a(x)\} \text{ und } \Omega_b \stackrel{def}{=} \{x \in \Omega : u(x) = u_b(x)\}$$

sowie $\Omega_i \stackrel{def}{=} \Omega \setminus (\Omega_a \cup \Omega_b)$. Dann folgt aus (10.19):

$$\begin{aligned} \|u\|_{H^1(\Omega)}^2 &= \int_{\Omega_a} (|\nabla u_a|^2 + u_a^2) dx + \int_{\Omega_b} (|\nabla u_b|^2 + u_b^2) dx + \frac{1}{\lambda^2} \int_{\Omega_i} (|\nabla p|^2 + p^2) dx \\ &\leq \|u_a\|_{H^1(\Omega)}^2 + \|u_b\|_{H^1(\Omega)}^2 + \frac{1}{\lambda^2} \|p\|_{H^1(\Omega)}^2. \end{aligned}$$

Nach Satz 5.1 gilt nun $\|p\|_{H^1(\Omega)} \leq c\|\bar{u}\|_{L^2(\Omega)}$ und somit $\|\bar{u}\|_{H^1(\Omega)} < \infty$ und damit die höhere Regularität von \bar{u} . ■

Im Fall glatter Gebiete und Schranken ist es sogar möglich die Lipschitz-Stetigkeit der Steuerung zu beweisen. Regularitätsresultate dieser Art sind besonders wichtig für die Analyse von Approximationsfehlern bei einer Diskretisierung wie wir noch sehen werden.¹¹¹

Die Variationsungleichung (10.14) kann im Falle von Box-Restriktionen an die Elemente aus U_{ad} durch Einführung zusätzlicher Lagrange-Multiplikatoren als weitere Gleichung formuliert werden. Dies führt zu sogenannten Karush-Kuhn-Tucker-Systemen aus denen man direkt weitere notwendige Optimalitätsbedingungen herleiten kann. Dies erweist sich jedoch als äußerst komplex und spielt für unser weiteres Vorgehen keine Rolle. Es sei auf [Trö09, 2.8.2] für das Modellproblem und für die theoretischen Hintergründe auf [Trö09, 6.1] und [Lue69, 9.4] verwiesen.

¹¹⁰Siehe auch [Mey11, Satz 11.24] oder [Her12, Lemma 16.19].

¹¹¹Wir werden die sogenannte Finite-Elemente-Methode betrachten. Der Interpolationsfehler und damit die Glattheit der Lösung wird dabei eine herausragende Rolle spielen.

V Diskretisierung elliptischer Optimalsteuerungsaufgaben

Nun beschäftigen wir uns mit der Diskretisierung von Optimalsteuerungsaufgaben in Funktionenräumen, um sie für eine numerische Behandlung handhabbar zu machen. Wir folgen dabei [Mey11, §13 und §14], [Her12, §14.3 und §16] und [Ohl08, 2.1 bis 2.3].

Im Fall unseres Modellproblems beinhaltet diese die Diskretisierung von Zustand und Steuerung, der PDE, des Zielfunktional und der Ungleichungsnebenbedingung. Zusätzlich wird eine Strategie, die auf die Diskretisierung der Steuerung verzichtet, und auch deren Vor- und Nachteile betrachtet. Den aufwändigsten Teil der Diskretisierung, die der PDE, werden wir zunächst explizit mit Hilfe der Finiten-Elemente-Methode¹¹², welche einen Spezialfall der allgemeineren Ritz-Galerkin-Verfahren darstellt, inklusive einer anschaulichen Herleitung der verwendeten Räume, angehen. Alternativen hierfür wären z. B. die sogenannten Finite-Differenzen¹¹³ oder die Finite-Volumen-Methode¹¹⁴.

Den Abschluss des Kapitels werden wir damit machen, den Diskretisierungsfehler für das Modellproblem unter verschiedenen Umständen genauer zu untersuchen.

§11 Das Ritz-Galerkin-Verfahren

Wir stellen zunächst das Ritz-Galerkin-Verfahren als Überbegriff vor und verdeutlichen die Grundidee der folgenden Approximationsmethode von Variationsproblemen, die die Voraussetzungen des Lemmas von Lax-Milgram 5.2 erfüllen.

Vergleichbar zu (5.5) betrachten wir zunächst ein allgemeines Variationsproblem:

Definition 11.1

V sei Hilbertraum, $a : V \times V \rightarrow \mathbb{R}$ eine Bilinearform, die die Voraussetzungen des Lemmas von Lax-Milgram erfüllt, und $F \in V^*$. Dann bezeichne $y \in V$ die eindeutige Lösung des Variationsproblems

$$a[y, v] = F(v) \quad \forall v \in V. \quad (11.1)$$

□

¹¹²Die Grundidee dahinter ist die Energieminimierung in endlichdimensionalen Teilräumen.

¹¹³Hier werden Differentialoperatoren durch Differenzenquotienten approximiert, siehe z. B. [Gro94, 2.2].

¹¹⁴Hier wird ausgenutzt, dass unter bestimmten Umständen aus einem Grundprinzip der Kontinuumsmechanik, dem Erhaltungsprinzip, diskrete Verfahren abgeleitet werden können, siehe z. B. [Krö97, 3] und [BO04].

Das Ritz-Galerkin-Verfahren wird nun folgendermaßen definiert:

Definition 11.2

Es seien die Voraussetzungen aus Def. 11.1 erfüllt. Zudem sei $V_h \subset V$. Dann ist die Ritz-Galerkin-Approximation $y_h \in V_h$ definiert als die eindeutige Lösung von

$$a[y_h, v_h] = F(v_h) \quad \forall v_h \in V_h. \tag{11.2}$$

□

Bemerkung:¹¹⁵

Die Existenz und Eindeutigkeit von y bzw. y_h folgt unmittelbar aus dem Lemma von Lax-Milgram, da die Teilmenge V_h wieder ein Hilbertraum mit dem aus V geerbten Skalarprodukt ist.

Für die Wahl von V_h stehen nun verschiedene Möglichkeiten zur Diskussion. Betrachten wir ein elliptisches Problem mit $V = H_0^1(\Omega)$, wobei Ω ein offenes, zusammenhängendes und beschränktes Gebiet ist, so sind neben den Finite-Elemente-Verfahren, auf welche wir im Anschluss noch genauer eingehen werden, vor allem folgende Wahlen von Teilmengen gebräuchlich:¹¹⁶

- $V_h \stackrel{def}{=} \mathbb{P}_{k(h)}(\Omega) \cap \{v_h \in C^0(\bar{\Omega}) : v_h = 0 \text{ auf } \Gamma\}$, wobei $\mathbb{P}_{k(h)}(\Omega)$ der Raum der Polynome vom (meist sehr hohem) Grad $\leq k(h)$ über Ω ist. Die zugehörigen Verfahren bezeichnet man als Spektralverfahren.
- $V_h \stackrel{def}{=} \text{span}\{u_i \in V : \mathcal{A}u_i = \lambda_i u_i, i = 1, \dots, k(h)\}$, wobei u_i die i -te Eigenfunktion des zugrundeliegenden Differentialoperators \mathcal{A} ¹¹⁷ ist.
- $V_h \stackrel{def}{=} \text{span}\{u_i \in V : \Delta u_i = \lambda_i u_i, i = 1, \dots, k(h)\}$, wobei u_i die i -te Eigenfunktion des Laplace-Operators Δ ist.

Eines unserer Hauptziele bei der Anwendung der Finiten-Elemente-Methode wird es sein, aus (11.2) ein äquivalentes lineares Gleichungssystem herzuleiten und damit eine Näherungslösung für die Lösung y des ursprünglichen Variationsproblems zu bestimmen. Eine Voraussetzung dafür ist natürlich ein endlichdimensionaler Teilraum V_h , der V möglichst gut diskretisiert bzw. approximiert, und eine hierfür geeignete Basis. Im folgenden Paragraphen werden wir darauf explizit eingehen.

Unabhängig von der Wahl von V_h können wir der Untersuchung der Fehlerabschätzung für die Finite-Elemente-Methode vorgreifen und ein erstes Ergebnis, eine abstrakte Fehlerabschätzung für das allgemeine Ritz-Galerkin-Verfahren angeben:

¹¹⁵Siehe auch [Ohl08, Bemerkung 2.3].

¹¹⁶Siehe auch [Ohl08, Beispiel 2.6].

¹¹⁷Siehe dazu [Trö09, 2.3.3] oder ausführlicher [Wlo82, §10].

Satz 11.1 (Lemma von Céa/Galerkin-Orthogonalität)

Seien die Voraussetzungen aus den Definitionen 11.1 und 11.2 erfüllt. Dann gilt die abstrakte Fehlerabschätzung

$$\|y - y_h\|_V \leq c \inf_{v_h \in V_h} \|y - v_h\|_V \quad (11.3)$$

mit einer Konstanten $c > 0$.¹¹⁸

Zudem gilt die Galerkin-Orthogonalität

$$a[y - y_h, v_h] = 0 \quad \forall v_h \in V_h. \quad (11.4)$$

□

BEWEIS ¹¹⁹ Wir zeigen zunächst die Galerkin-Orthogonalität. Sei dazu $v_h \in V_h$ beliebig. Wegen $V_h \subset V$ folgt

$$a[y - y_h, v_h] = a[y, v_h] - a[y_h, v_h] = F(v_h) - F(v_h) = 0.$$

Für den Beweis des Lemmas von Céa sei $v_h \in V$ wieder beliebig. Aus der Koerzivität der Bilinearform folgt

$$\begin{aligned} \beta_0 \|y - y_h\|_V^2 &\leq a[y - y_h, y - y_h] \\ &= a[y - y_h, y - v_h] + \underbrace{a[y - y_h, v_h - y_h]}_{\stackrel{\in V_h}{(11.4)} 0} \\ &= a[y - y_h, y - v_h] \stackrel{(5.8)}{\leq} \alpha_0 \|y - y_h\|_V \|y - v_h\|_V. \end{aligned}$$

Nach Division mit $\beta_0 \|y - y_h\|_V$ folgt die Behauptung für $c = \frac{\alpha_0}{\beta_0}$ indem wir auf beiden Seiten zum Infimum übergehen. ■

Bemerkung:¹²⁰

Das Lemma von Céa liefert also eine Fehlerabschätzung in der Norm von V . Die Fehlerabschätzung wird zurückgeführt auf die Best-Approximation von y in V_h . Die Abschätzung (11.3) wird als quasi-optimal bezeichnet, weil man durch Lösen der diskreten Aufgabe (11.2) bis auf eine Konstante die überhaupt bestmögliche Approximation der kontinuierlichen Lösung y in V_h erhält. Die Galerkin-Orthogonalität (11.4) bedeutet im Falle einer symmetrischen Bilinearform a tatsächlich, dass der Fehler $y - y_h$ im von a erzeugten Skalarprodukt in der Tat senkrecht auf V_h steht. Die weitere numerische Analyse beruht somit also allein auf der Approximationstheorie wie wir auch noch sehen werden.

¹¹⁸Im Folgenden bezeichnen wir die Konstante c der Übersicht halber mit $c_{11.1}$.

¹¹⁹Siehe auch [Mey11, Lemma 13.9] oder [Her12, Satz 16.1].

¹²⁰Siehe auch [Her12, S.99 f.].

§12 Die Finite-Elemente-Methode und die Diskretisierung der linearer elliptischer PDEs

Finite-Elemente-Verfahren sind Spezialfälle von Ritz-Galerkin-Verfahren für eine bestimmte Klasse von Teilräumen $V_h \subset V$. Die Konstruktion von V_h beruht in diesem Fall auf einer Zerlegung von Ω in nichtüberlappende Teilgebiete, die selbst wiederum einfache geometrische Objekte sind. Wir werden im Folgenden die einfachste Klasse von Finiten Elementen, Lagrange-Elemente auf simplizialen Gittern, betrachten.

Hierzu definieren wir zunächst den sehr allgemeinen Begriff des Simplex' und konkretisieren ihn anschließend um ihn für unsere Zwecke leicht nutzbar zu machen. Eine wie folgt definierte simpliziale Zerlegung wird oft auch als Triangulierung bezeichnet.

Definition 12.1

Seien $s \in \{1, \dots, N\}$ und $a_0, \dots, a_s \in \mathbb{R}^N$ Punkte, s.d. $(a_j - a_0)_{j=1, \dots, s}$ linear unabhängig sind. Dann heißt

$$T \stackrel{def}{=} \{x \in \mathbb{R}^d : x = \sum_{i=0}^s \lambda_i a_i \text{ für } \lambda_i \geq 0, \sum_{i=0}^s \lambda_i = 1\}$$

nicht degeneriertes s -dimensionales Simplex im \mathbb{R}^N .

Die Punkte a_0, \dots, a_s heißen Ecken oder Knoten des Simplex. Ist $r \in \{1, \dots, s\}$ und $\tilde{a}_0, \dots, \tilde{a}_r \in \{a_0, \dots, a_s\}$, so heißt

$$\tilde{T} \stackrel{def}{=} \{x \in \mathbb{R}^N : x = \sum_{i=0}^r \lambda_i \tilde{a}_i \text{ für } \lambda_i \geq 0, \sum_{i=0}^r \lambda_i = 1\}$$

r -dimensionales Seitensimplex von T . Die eindimensionalen Seitensimplizes heißen Kanten, die nulldimensionalen Ecken und die $(s-1)$ -dimensionalen Flächen.

Wir bezeichnen mit T_0 den Simplex zu den Punkten $a_0 = 0, a_i = e_i, i = 1, \dots, N$. T_0 heißt N -dimensionaler Einheitssimplex.

Der Durchmesser von T ist gegeben durch

$$\rho(T) \stackrel{def}{=} \text{diam}(T) = \max_{i,j=1, \dots, s} |a_i - a_j|.$$

Mit $R(T) \stackrel{def}{=} 2 \sup\{r : B_r(x) \subset T\}$ bezeichnet man den Innenkegeldurchmesser von T . □

Bemerkung:¹²¹

Im \mathbb{R}^2 ist ein zweidimensionales Simplex etwa das Dreieck mit den Ecken a_0, a_1, a_2 , im \mathbb{R}^3

¹²¹ Siehe auch [Ohl08, Bemerkung 2.9].

ist es der Tetraeder mit den Ecken a_0, a_1, a_2, a_3 , sechs eindimensionalen und vier zweidimensionalen Seitensimplizes. Man erkennt leicht, dass ein s -dimensionales Seitensimplex $\binom{s+1}{r+1}$ r -dimensionale Seitensimplexe besitzt.

Die Größen $\lambda_0, \dots, \lambda_s \in [0, 1]$ sind Koordinaten, die dem Simplex besonders gut angepasst sind.

Wir gehen nun im Folgenden von einem allgemeinen Variationsproblem der Form (11.1) aus und ersetzen dieses durch (11.2) mit $\dim V_h < \infty$. V_h bezeichnen wir von nun an als diskreten Ansatzraum.

Entsprechend unserer Vorüberlegungen zerlegen wir das Rechengebiet Ω für die Wahl des endlichdimensionalen Unterraums in n_t disjunkte Simplizes, also Teilgebiete T_i , auch als Elemente bezeichnet, s.d.

$$\overline{\Omega} = \bigcup_{i=1}^{n_t} \overline{T_i}.$$

und veranschaulichen unser Vorhaben anhand von folgendem Beispiel:¹²²

Beispiel 12.1

Als Beispiel einer solchen Zerlegung betrachten wir also nun die Triangulierung eines zweidimensionalen Gebiets mittels eines Dreiecksgitters:

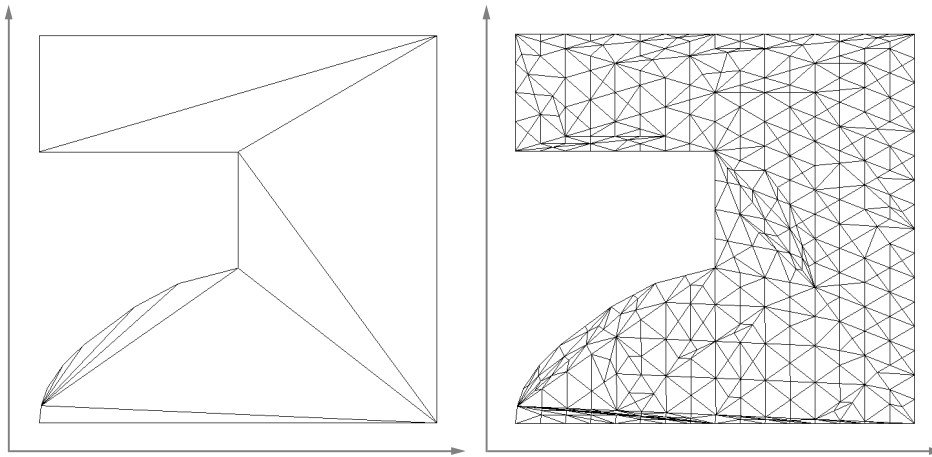


Abbildung 6: Grobe und feine (zulässige) Zerlegung eines zweidimensionalen Gebiets.

¹²²Siehe auch [Mey11, Beispiel 13.2].

n_p bezeichne von nun an die Anzahl der Punkte (Knoten), die Gitterweite h wird definiert durch

$$h = \max_{1 \leq i \leq n_p} \rho(T_i),$$

wobei $\rho(T_i)$ wieder der Durchmesser des Dreiecks T_i ist.

Eine andere Möglichkeit ist durch eine Zerlegung in Vierecke gegeben, für dreidimensionale Gebiete ist wie bereits erwähnt eine Zerlegung in Tetra- aber auch Hexaeder üblich. Eindimensionale Gebiete werden natürlich stets in Intervalle zerlegt. \square

Bemerkung:

Es ist klar, dass eine solche exakte Überdeckung von Ω nur möglich ist, falls Ω ein polygonal berandetes Gebiet ist. Andernfalls ist es nötig mit gekrümmten Elementen zu arbeiten, oder man macht einen zusätzlichen Fehler durch die Approximation des Randes durch einen Polygonzug.

Bei der Finiten-Elemente-Methode verwendet man als V_h den Raum der stetigen Funktionen, die auf jedem Element, d.h. stückweise, ein Polynom vom Grad $t \geq 1$ darstellen. Dass dadurch in der Tat ein Ritz-Galerkin-Verfahren entsteht, d.h. dass in unserem Fall $V_h \subset V = H_0^1(\Omega)$ gilt, verdankt man der Aussage des folgenden Satzes:

Satz 12.1

Sei $k \geq 1$ und Ω ein beschränktes, polygonal berandetes Gebiet. Eine stückweise beliebig oft differenzierbare Funktion $v : \overline{\Omega} \rightarrow \mathbb{R}$ gehört genau dann zu $H^k(\Omega)$, wenn $v \in C^{k-1}(\overline{\Omega})$ gilt. \square

BEWEIS Siehe [Bra07, II 5.2]. Der Beweis geht von einer zulässigen Zerlegung von Ω aus, welche in Def. 12.2 nachgereicht wird. Da wir später nur noch von solch einer Zerlegung ausgehen, ist dies hier durchaus legitim. \blacksquare

Wir werden also etwa Elemente aus $C^0(\Omega)$ betrachten, die nur in $H^1(\Omega)$ enthalten sind. Aus Satz 12.1, sowie dem Vorgehen aus Beispiel 3.1, in welchem wir die schwache Ableitung einer stetigen und stückweise glatten Funktion berechnet haben, folgern wir sofort:

Korollar 12.1

Sei $\Omega \subset \mathbb{R}^N, N \geq 1$, ein beschränktes Gebiet. Ist eine Funktion $v : \overline{\Omega} \rightarrow \mathbb{R}$ stetig und stückweise beliebig oft differenzierbar, dann gilt $v \in W^{1,\infty}(\Omega)$. \square

Es gilt also, dass die etwa dreiecksweise definierten und dort konstanten partiellen Ableitungen einer Funktion in V_h zusammen die schwache Ableitung der Funktion bilden. $V_h \subset V$ ist also erfüllt.

Wir machen uns nun daran, unser V_h , sowie die bereits erwähnten benötigten Lagrange-Elemente zu definieren.

Die durch unser Vorgehen entstehenden Unterräume sind endlichdimensional, es gilt also wie

angestrebt $\dim V \stackrel{def}{=} n < \infty$, es existiert also eine Basis $\{\varphi_i\}_{i=1}^n$, s.d. jedes $v_h \in V_h$ geschrieben werden kann als

$$v_h(x) = \sum_{i=1}^n v_i \varphi_i(x) \quad (12.1)$$

mit $\vec{v} \stackrel{def}{=} (v_1, \dots, v_n)^T \in \mathbb{R}^n$.

Wir betrachten nun folgendes Beispiel¹²³ zur Erstellung eines geeigneten endlichdimensionalen Unterraums V_h , sowie einer für unser weiteres Vorgehen geeigneten Basis dieses Raums.

Beispiel 12.2

Ein Beispiel für V_h , an dem wir im Folgenden auch festhalten wollen, ist wie bereits angedeutet etwa der Raum der stetigen und stückweise linearen Funktionen:

$$V_h = \{v \in C(\bar{\Omega}) : v|_{T_i} \in \mathbb{P}_1(T_i) \quad \forall i \in \{1, \dots, n_t\}\}, \quad (12.2)$$

wobei die T_i Simplexes im Sinne von Def. 12.1 darstellen. Die sogenannte nodale (oder auch knotenbasierende) Basis bildet auf diesen Räumen eine mögliche Basis:

$$\{\varphi_i\}_{i=1}^{n_p} \subset V_h, \quad \varphi_i(a_j) = \delta_{ij}. \quad (12.3)$$

Hierbei sind $(a_j)_{j=1}^{n_p}$ die Knoten der Triangulierung, die φ_i bezeichnet man als Ansatzfunktionen, die Anzahl der Knoten n_p stimmt mit $\dim V_h$ überein.

Im Fall eines eindimensionalen Gebiets haben die oben beschriebenen linearen Ansatzfunktionen folgende Form:

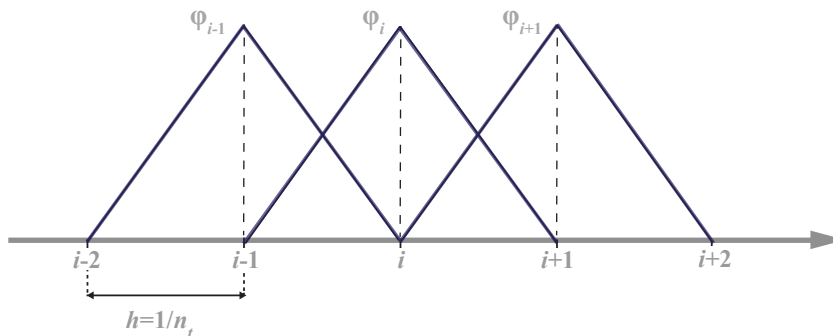


Abbildung 7: Lineare Ansatzfunktionen bei uniformer Triangulierung eines eindimensionalen Gebiets.¹²⁴

¹²³Siehe auch [Mey11, Beispiel 13.4].

¹²⁴Vgl. [Mey11, S.69].

Die Gitterweite beträgt hier stets $h = 1/n_t$, die Intervalle sind also stets gleich groß. Man spricht von einer uniformen Triangulierung. Aufgrund ihrer Form bezeichnet man die linearen Ansatzfunktionen in diesem Fall auch als Hutfunktionen. Im zweidimensionalen Fall nehmen diese erwartungsgemäß Pyramidenform an und werden dementsprechend auch als Pyramidenfunktionen bezeichnet. \square

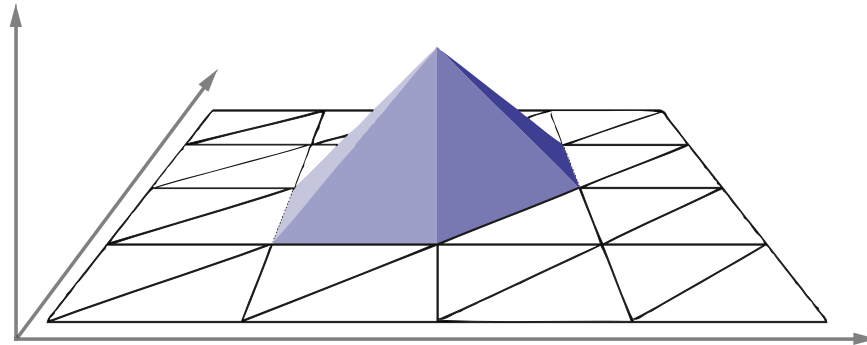


Abbildung 8: Pyramidenfunktion auf gleichmäßig (zulässig) trianguliertem Einheitsquadrat.

Für unser weiteres Vorgehen benötigen wir eine bestimmte Art der Triangulierung:

Definition 12.2

Sei $\Omega \subset \mathbb{R}^N$ ein beschränktes Gebiet und

$$\mathcal{T}_h \stackrel{\text{def}}{=} \{T_j : j = 1, \dots, n_t, T_j \text{ ist } N\text{-dimensionaler Simplex im } \mathbb{R}^N\}.$$

\mathcal{T}_h heißt zulässige oder (geometrisch) konforme Zerlegung von Ω falls gilt:

- $\bar{\Omega} = \bigcup_{j=1}^{n_t} \bar{T}_j$.
- Für je zwei $T_1, T_2 \in \mathcal{T}_h$ mit $S \stackrel{\text{def}}{=} T_1 \cap T_2$ gilt: S ist entweder eine leere Menge, oder ein $(N - k)$ -dimensionaler Seitensimplex von T_1 und T_2 für ein $k \in \{0, \dots, N\}$, im Falle einer Dreieckszerlegung etwa also eine gemeinsame Kante oder Ecke.

Die Gitterweite von \mathcal{T}_h wird analog zu Beispiel 12.1 definiert durch

$$h = \max_{T \in \mathcal{T}_h} \rho(T).$$

Eine Familie von Zerlegungen $\{\mathcal{T}_h\}$ heißt quasiuniform oder shape-regular, wenn ein $\kappa > 0$ existiert, s.d. für alle \mathcal{T}_h gilt:

$$\frac{\rho(T)}{R(T)} \leq \kappa \quad \forall T \in \mathcal{T}_h, \tag{12.4}$$

wobei $R(T)$ wieder der Innendurchmesser von T ist. Man bezeichnet κ als Quasiuniformitätskonstante, den Quotienten $\frac{\rho(T)}{R(T)}$ als Aspektverhältnis von T . \square

Bemerkung:¹²⁵

Durch Quasiuniformität wird vermieden, dass etwa im zweidimensionalen Fall bei feinerer Gitterweite die Dreiecke nicht beliebig spitz werden, was sich negativ auf den Interpolationsfehler auswirken würde. Dies kann etwa durch wiederholte Dreieckshalbierung bei Gitterverfeinerung geschehen. Dreiecke folgender Form sollen dadurch also vermieden werden:

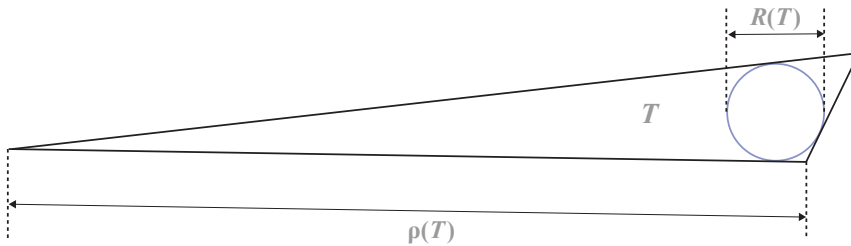


Abbildung 9: Spitzes Dreieck T mit Durchmesser $\rho(T)$ und Innendurchmesser $R(T)$.¹²⁶

Nach obiger Definition wäre eine unzulässige Zerlegung etwa gegeben durch:

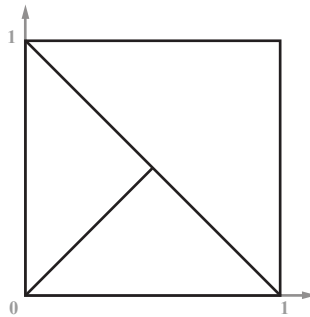


Abbildung 10: Beispiel einer unzulässigen Zerlegung des Einheitsquadrats.

Die Definition von Finite-Elemente-Räumen basierend auf einer Triangulierung \mathcal{T}_h erfordert lediglich die Angabe lokaler Funktionenräume auf den Simplizes. Zudem muss festgelegt werden, wie diese lokalen Funktionen dann global zusammengesetzt werden. Dazu betrachten wir zunächst folgende Definition und fassen anschließend unsere bisherigen Ergebnisse zusammen.

¹²⁵Siehe auch [Mey11, S.74].

¹²⁶Vgl. [Mey11, S.74].

Definition 12.3

Sei $T \subset \mathbb{R}^N$ ein N -dimensionaler Simplex. Sei $\mathcal{N} \stackrel{\text{def}}{=} \{a_i : i = 0, \dots, N\}$ die Menge der Ecken von T . Dann ist durch die Angabe von Werten in den Punkten $a_i \in \mathcal{N}$ eindeutig eine lineare Funktion $p \in \mathbb{P}_1(T)$ definiert. Durch $\Phi \stackrel{\text{def}}{=} \{\varphi_i : \varphi_i(a_i) = \delta_{ik}, i, k = 0, \dots, d\}$ ist eine nodale Basis von $\mathbb{P}_1(T_j)$ gegeben. Wir bezeichnen das Tripel (T, Φ, \mathcal{N}) als lineares simpliziales Lagrange-Element. Die Basisfunktionen werden oft auch Formfunktionen oder shape functions genannt, \mathcal{N} bezeichnet die Menge der nodalen Variablen. □

Bemerkung:¹²⁷

Die Wohldefiniertheit des linearen simplizialen Lagrange-Elements folgt unmittelbar aus der Form der Basisfunktionen und die daraus folgende Darstellung der Funktionen aus $\mathbb{P}_1(T)$.¹²⁸

Wir kehren nun zu unserem Vorgehen aus Beispiel 12.2 zurück und beschreiben nun entgültig unseren linearen Finite-Elemente-Raum V_h :

Definition 12.4

- (i) Sei $\Omega \subset \mathbb{R}^N$ gemäß den bisherigen Voraussetzungen und \mathcal{T}_h eine zulässige Triangulierung von Ω . Der Raum der linearen finite Elemente auf simplizialen Gittern sei definiert durch (12.2).
- (ii) Sind $a_j, j = 1, \dots, n_p$ die Knoten der Triangulierung \mathcal{T}_h , so ist eine Funktion $v_h \in V_h$ durch die Vorgabe von Funktionswerten in den Ecken $v_h(a_j)$ eindeutig definiert. Insbesondere gilt dadurch $\dim V_h = n_p$.
Eine Basis von V_h ist gegeben durch (12.3). Wie bereits erwähnt, bezeichnet man diese Basis als nodale oder Knotenbasis.
- (iii) Ist $(T_0, \tilde{\Phi}, \tilde{\mathcal{N}})$ das lineare Lagrange-Element auf dem Einheitssimplex T_0 und $v_h \in V_h$ gegeben durch

$$v_h(x) = \sum_{i=1}^{n_p} v_h(a_i) \varphi_i(x),$$

so gilt für beliebige Simplices $T \in \mathcal{T}_h$ mit Ecken $\tilde{a}_0, \dots, \tilde{a}_N$

$$v_h|_T(x) = \sum_{i=1}^N v_h(\tilde{a}_i) \tilde{\varphi}_i(F^{-1}(x)),$$

wobei $F : T_0 \rightarrow T$ die Referenzabbildung¹²⁹ und $\tilde{\varphi}_i \in \tilde{\Phi}$ die Formfunktion des Referenzelements T_0 ist. □

¹²⁷Siehe auch [Ohl08, Bemerkung 2.14].

¹²⁸Vergleiche auch Beispiel 12.2.

¹²⁹Zum Begriff der Referenzabbildung bzw. -element siehe z. B. [Bra07, II 5.12].

Um die Zulässigkeit obiger Definition zu verdeutlichen, beziehen wir unsere bisherigen Überlegungen mit ein.¹³⁰

Anmerkung 12.1

Wir stellen fest, dass $v_h|_T \in \mathbb{P}_1(T)$ durch die Vorgabe von Werten in den Eckpunkten der Triangulierung wohldefiniert ist. Die globale Stetigkeit von v_h folgt direkt daraus, dass für zwei Simplexe $T_1, T_2 \in \mathcal{T}_h$ mit einer gemeinsamen Fläche $S = T_1 \cap T_2$ gilt: $v_h|_S \in \mathbb{P}_1(S)$. Analog zum ersten Teil folgt, dass $v_h|_S$ durch die Vorgabe von Werten in den Eckpunkten der Triangulierung eindeutig bestimmt ist, was wiederum $v_h \in C(\bar{\Omega})$ entspricht. \square

Auf den bisherigen Ergebnissen aufbauend wollen wir nun, wie in §11 angedeutet, mit Hilfe der Finiten-Elemente-Methode, ein zu (11.2) äquivalentes lineares Gleichungssystem herleiten, um y_h berechnen zu können.

Durch Einsetzen von (12.1) wird (11.2) zu

$$\sum_{j=1}^n a[\varphi_i, \varphi_j] y_j = F(\varphi_i) \quad \forall i = 1, \dots, n \Leftrightarrow A \vec{y} = \vec{b}, \tag{12.5}$$

wobei $A \in \mathbb{R}^{n \times n}$ und $b \in \mathbb{R}^n$ mit

$$A_{ij} \stackrel{def}{=} a[\varphi_i, \varphi_j] \text{ und } b_i \stackrel{def}{=} F(\varphi_i). \tag{12.6}$$

Zur Lösbarkeit von (12.5) betrachten wir folgende, sehr an das Lemma von Lax-Milgram 5.2 erinnernde Aussage:

Lemma 12.1

Sei V Hilbertraum und $a: V \times V \rightarrow \mathbb{R}$ symmetrische, koerzive und beschränkte Bilinearform, $V_h \subset V$ endlichdimensionaler Unterraum mit Basis $\{\varphi_i\}_{i=1}^n$. Dann ist $A \in \mathbb{R}^{n \times n}$ aus (12.6) symmetrisch und positiv definit und damit invertierbar. \square

BEWEIS ¹³¹ Da a symmetrisch ist, gilt dies natürlich auch für A . Zum Beweis der positiven Definitheit sei $y_h \in V_h$ beliebig, $\vec{y} \in \mathbb{R}^n$ sei der Koordinatenvektor von y_h zur Basis $\{\varphi_i\}_{i=1}^n$. Aus der Basiseigenschaft folgt, dass $\vec{y} \neq 0 \Rightarrow y_h \neq 0$.

Aus der Koerzivität von a folgt:

$$\vec{y}^T A \vec{y} = \sum_{i,j=1}^n y_i a[\varphi_i, \varphi_j] y_j = a[y_h, y_h] \geq \beta_0 \|y_h\|_V^2 > 0 \quad \forall \vec{y} \neq 0,$$

also ist A positiv definit. ■

¹³⁰Siehe auch [Ohl08, Bemerkung 2.18].

¹³¹Siehe auch [Mey11, Lemma 13.6].

Beispiel 12.3 (Lineares Gleichungssystem für die Poissongleichung¹³²)

Im Fall unseres Modellproblems bzw. der schwachen Formulierung der Poissongleichung (5.4) nehmen A und \vec{b} die folgenden Formen an:

$$A_{ij} \stackrel{def}{=} \int_{\Omega} \nabla \varphi_i \cdot \nabla \varphi_j \, dx, \quad \vec{b} \stackrel{def}{=} \int_{\Omega} f \varphi_i \, dx \quad (12.7)$$

mit $f \in L^2(\Omega)$ gegeben. Dementsprechend ergibt sich

$$A \vec{y} = \vec{b} \Leftrightarrow \sum_{j=1}^n \int_{\Omega} \nabla \varphi_i \cdot \nabla \varphi_j \, dx y_j = \int_{\Omega} f \varphi_i \, dx \quad \forall i = 1, \dots, n. \quad (12.8)$$

Eine Matrix dieser Form nennt man Steifigkeitsmatrix, der Vektor \vec{b} wird als Lastvektor bezeichnet.¹³³ Nach dem Beweis des Satzes von Friedrich 5.1 ist die Bilinearform $a[y, v] = \int_{\Omega} \nabla y \cdot \nabla v \, dx$ beschränkt und koerziv in $H_0^1(\Omega)$, daher ist A nach Lemma 12.1 invertierbar, falls $\{\varphi_i\}_{i=1}^n$ eine Basis von $V_h \subset V = H_0^1(\Omega)$ bildet. Für die linearen Ansatzfunktionen aus Beispiel 12.2 bzw. Def. 12.4 ist dies etwa der Fall, falls nur Funktionen gewählt werden, die auf dem Rand verschwinden, also:

$$\varphi_i \in C(\overline{\Omega}), \quad \varphi_i|_{T_j} \in \mathbb{P}_1(T_j) \quad \forall j = 1, \dots, n_t, \quad \varphi_i(x_j) = \begin{cases} 1, & i = j, x_i \notin \Gamma \\ 0, & \text{sonst.} \end{cases} \quad (12.9)$$

Die Komponenten des Lösungsvektors \vec{y} , die zu $x_i \in \Gamma$ gehören, werden so in (12.8) nicht berücksichtigt, da ihre Ansatzfunktionen nicht Element der Basis sind. In diesem Fall gilt aber natürlich aufgrund der homogenen Dirichlet-Randbedingung $y_i = 0$, falls $x_i \in \Gamma$. Ist n_{Γ} die Anzahl der Knoten auf Γ , so ist die Dimension des Problems dann $n_p - n_{\Gamma}$.

Durch die Transformation auf ein Einheits- oder Referenzelement können die Einträge der Steifigkeitsmatrix im Fall von (12.9) sehr effizient berechnet werden.¹³⁴ Da sich die Träger der Basisfunktionen φ_i und ihre Ableitungen nur auf Elemente erstrecken, die x_i als Eckpunkt haben, ist A_{ij} nur dann ungleich Null, wenn i und j Eckpunkte gemeinsamer Elemente sind. Die sich ergebende Matrix ist also dünn besetzt, was vorteilhaft für die numerische lineare Algebra ist.

¹³²Siehe auch [Mey11, Beispiel 13.8].

¹³³Es handelt sich dabei um Begriffe aus der Festigkeitslehre.

¹³⁴Betrachte dazu [Bra07, II §8]. Für den Spezialfall unseres Modellproblems, allerdings mit anderen Randwerten, betrachte [Zul08, 4.1].

Geht man wie in Beispiel 12.2 vor, so ergibt sich die Tridiagonalmatrix

$$A = \frac{1}{h} \begin{pmatrix} 2 & -1 & 0 & \dots & \dots & 0 \\ -1 & \ddots & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & \ddots & -1 \\ 0 & \dots & \dots & 0 & -1 & 2 \end{pmatrix} \quad \square$$

Bemerkung:¹³⁵

I. A ist die Steifigkeitsmatrix nicht positiv definit, sondern lediglich positiv semidefinit. Die positive Definitheit von A aus unserem Beispiel folgt aus der homogenen Dirichlet-Randbedingung bzw. aus der Koerzivität von $a[y, v] = \int_{\Omega} \nabla y \cdot \nabla v \, dx$ in $H_0^1(\Omega)$ aufgrund der Poincaré-Friedrich Ungleichung 5.3.

Wir haben bereits festgestellt, dass Lemma 12.1 sehr an das Lemma von Lax-Milgram erinnert. Betrachtet man die Struktur beider Sätze, so erkennt man, dass die Koerzivität der positiven Definitheit im Unendlichdimensionalen entspricht. Positive Definitheit ist allerdings keinesfalls notwendig für die Invertierbarkeit von Matrizen.

Das Lemma von Lax-Milgram ist dementsprechend lediglich ein Spezialfall einer allgemeineren Aussage, der Fredholm'schen Alternative, die, für den Fall dass der Eigenwertfall nicht eintritt, die Invertierbarkeit von kompakten Operatoren sichert. Für den Fall elliptischer PDEs sei diesbezüglich auf [Eva10, 6.2.3] verwiesen.

Wie in §11 bereits angedeutet, gibt es noch zahlreiche Alternativen für Ansatzräume, bzw. -funktionen. Es ist möglich den Grad k der Ansatzfunktionen zu erhöhen und/oder die zusätzliche Stetigkeit der Ableitung zu fordern. Auch der Einsatz von nicht konformen Elementen¹³⁶ ist möglich. Dabei handelt es sich dann nicht mehr um ein Ritz-Galerkin-Verfahren nach Def. 11.2, ein Beispiel hierfür wären die Discontinuous-Galerkin-Verfahren.¹³⁷

§13 Fehlerabschätzung für die Zustandsgleichung

Um unser schlussendliches Ziel, die Bestimmung des Diskretisierungsfehlers für unser Modellproblem (10.6), zu erreichen, muss nun untersucht werden, inwiefern ein $y_h \in V_h$, das nach Lemma 12.1 für jedes lineare stetige $F \in H_0^1(\Omega)^*$ eine Lösung von (12.5) darstellt, gegen die exakte Lösung y konvergiert, wenn man die Triangulierung verfeinert und falls das der Fall ist, mit welcher Konvergenzordnung. Der Einfachheit halber beschränken wir uns von nun an auf zwei-

¹³⁵Siehe auch [Mey11, Beispiel 13.8 und Bemerkung 13.7].

¹³⁶D.h. $V_h \not\subset V$, siehe etwa [Gro94, 4.6].

¹³⁷Siehe hierfür etwa [G+07, 4.8] und [C+00].

bzw. zunächst auch noch zusätzlich dreidimensionale Gebiete. Dies entspricht auch unserem Problem der optimalen Temperaturquelle.

Mit Satz 11.1 haben wir bereits erste Ergebnisse zur Bestimmung eines Approximationsfehlers hergeleitet. Das Lemma von Céa und die Galerkin-Orthogonalität besitzen, wie bereits erwähnt, ausgehend von der allgemeinen Aufgabenstellung (11.1) und unabhängig von der Wahl des Ansatzraumes für jedes Ritz-Galerkin-Verfahren Gültigkeit.

Im Falle unseres Ansatzes der Finiten-Elemente-Methode erhalten wir für die Best-Approximation-Eigenschaft (11.3) der Näherungslösung y_h :

$$\|y - y_h\|_V \leq c_{11.1} \|y - v_h\|_V \quad \forall v_h \in V_h. \quad (13.1)$$

Da die Aufgabe der Best-Approximation i. A. schwer zu lösen ist, geht man zur Bestimmung der Konvergenzordnung folgendermaßen vor: Auf der rechten Seite von (13.1) ist es möglich, eine Interpolation von y zu verwenden, d.h. $v_h = I_h y \in V_h$ mit geeignetem Interpolationsoperator $I_h : V \rightarrow V_h$. Der sogenannte Finite-Elemente-Fehler $\|y - y_h\|_V$ kann dann durch den Interpolationsfehler abgeschätzt werden.¹³⁸

Zunächst brauchen wir jedoch noch einige zusätzliche Resultate, welche v.a. die Regularität von y und somit der Zulässigkeit des Zustands $I_h y$, wie wir ihn betrachten werden, betreffen.

Satz 13.1

Sei $\Omega \subset \mathbb{R}^N$, $N = 2, 3$, ein beschränktes Gebiet

- i) mit glattem Rand,¹³⁹ oder
- ii) zusätzlich konvex mit einem Polygonzug ($N = 2$) oder Polyeder ($N = 3$) als Rand

Dann gilt: Es existiert für alle $f \in L^2(\Omega)$ eine eindeutige Lösung $y \in H^2(\Omega) \cap H_0^1(\Omega)$ von

$$\int_{\Omega} \nabla y \cdot \nabla v \, dx = \int_{\Omega} f v \, dx \quad \forall v \in H_0^1(\Omega) \quad (13.2)$$

und es gilt die Abschätzung

$$\|y\|_{H^2(\Omega)} \leq c \|f\|_{L^2(\Omega)} \quad (13.3)$$

mit einer Konstanten $c > 0$, unabhängig von y und f .¹⁴⁰ □

BEWEIS Die Existenz einer eindeutigen Lösung in $H_0^1(\Omega)$ wurde bereits im Satz von Friedrich 5.1 gezeigt. Für die zusätzlichen Aussagen $y \in H^2(\Omega)$ und die Abschätzung (13.3) sei im Fall von i) auf [Eva10, 6.3.2 Theorem 4] und [Eva10, 6.2.3 Theorem 6], im Fall von ii) auf [Gri92, 2.7] verwiesen. Für eine ähnliche Aussage bezüglich beschränkter, konvexer Lipschitzgebiete sei auf [Gri85, Theorem 3.2.1.2] verwiesen. ■

¹³⁸Siehe auch [Mey11, S.72].

¹³⁹D.h. ein Gebiet der Klasse $C^{1,1}$.

¹⁴⁰Im Folgenden bezeichnen wir die Konstante c der Übersicht halber mit $c_{13.1}$.

Wir fassen die Bedeutung von Satz 13.1 für unser Vorgehen zusammen:¹⁴¹

Anmerkung 13.1

Wir haben bereits in Anmerkung 5.3 festgestellt, dass man keine Inhomogenitäten aus $L^2(\Omega)$ benötigt, um Lösungen in $H_0^1(\Omega)$ zu erhalten, man kommt mit Inhomogenitäten aus $H_0^1(\Omega)^*$ aus. Wenn jedoch wie in Satz 13.1 die rechte Seite aus $L^2(\Omega)$ stammt, so sind die Lösungen nach Abschätzung (13.3) 'besser'. Wir erhalten im obigen Fall der Poissongleichung sogar eine Aussage zur maximalen Regularität der Lösung y : Gilt für die schwache Lösung $y \in H^2(\Omega)$, dann erfüllt sie auch in der Tat $-\Delta y = f$ punktweise fast überall in Ω .¹⁴²

Für die Finite-Elemente-Methode ist der Fall ii) besonders wichtig, da solche Gebiete besonders einfach zu triangulieren sind. Ist das Gebiet jedoch nicht konvex, besitzt es etwa eine einspringende Ecke, so muss um diese herum das Gitter lokal verfeinert werden um die Konvergenzordnung zu erhalten.

Im Falle nicht glatter bzw. nicht konvexer Gebiete sei auf [A⁺07] verwiesen. □

Von nun an beschränken wir uns auf eine Zerlegung des Gebiets Ω in Tetraeder ($N = 3$) und v.a. in Dreiecke ($N = 2$) und den speziellen Ansatzraum

$$V_h = \text{span}\{\varphi_1, \dots, \varphi_{n_p}\}, \tag{13.4}$$

wobei $\{\varphi_i\}_{i=1}^{n_p}$ die linearen Ansatzfunktionen aus Beispiel 12.2 sind, d.h. es gilt

$$\varphi_i \in C(\overline{\Omega}), \varphi_i|_{T_j} \in \mathbb{P}_1(T_j), \varphi_i(x_j) = \delta_{ij} \quad \forall i = 1, \dots, n_p \quad \forall j = 1, \dots, n_t. \tag{13.5}$$

Bezüglich V_h ist die Lagrange-Interpolation von y gegeben durch

$$(I_h y)(x) \stackrel{\text{def}}{=} \sum_{i=1}^{n_p} y(x_i) \varphi_i(x),$$

welche wir im Folgenden als Interpolation benutzen werden.

Zur Wohldefiniertheit von $I_h y$ benötigen wir den Sobolew'schen Einbettungssatz 4.3:¹⁴³

Obwohl $y \in H^2(\Omega) \cap H_0^1(\Omega)$ als Lösung der PDE im schwachen Sinne nicht punktweise definiert ist, so wissen wir aus Satz 4.3 dass $H^2(\Omega) \hookrightarrow C(\overline{\Omega})$ für $N \leq 3$, was so zu verstehen ist, dass es in der Äquivalenzklasse von $y \in H^2(\Omega)$ genau einen Repräsentanten gibt, der stetig auf $\overline{\Omega}$ ist. Diesen benutzt man bei der Definition von $i : H^2(\Omega) \rightarrow C(\overline{\Omega})$ in Def. 4.6 bzw. bei der Definition von $I_h y$.

¹⁴¹Siehe auch [Mey11, Bemerkungen 13.12 und 13.13].

¹⁴²Man kann dies durch partielle Integration und anschließender Anwendung des verallgemeinerten Fundamentallemmas der Variationsrechnung (siehe [Alt06, 1-16]) beweisen.

¹⁴³Siehe auch [Mey11, Bemerkung 13.14].

Zur nun folgenden Abschätzung des Interpolationsfehlers $y - I_h y$ kehren wir zu einer Zulässigen Zerlegung nach Def. 12.2 zurück und gehen von nun an von solch einer Triangulierung des Gebiets Ω aus.

Satz 13.2 (Interpolationsfehler)

Seien $y \in H^2(\Omega)$ und $\{\mathcal{T}_h\}$ eine Familie von zulässigen, quasiuniformen Triangulierungen, dann gilt

$$\|y - I_h y\|_{H^1(\Omega)} \leq ch^{2-t} \|y\|_{H^2(\Omega)}$$

mit $t \in \{0, 1\}$ und einer Konstanten $c = c(\Omega, \kappa, t) > 0$, mit der Quasiuniformitätskonstante κ aus (12.4).¹⁴⁴ Wie schon vorher gilt $H^0(\Omega) = L^2(\Omega)$. □

BEWEIS Bei dem Satz handelt es sich um einen Spezialfall von [Bra07, II 6.4]. Für den Beweis spielt das Bramble-Hilbert-Lemma¹⁴⁵, sowie ein Transformationsatz auf ein Referenzelement, wodurch die Quasiuniformität ins Spiel kommt, eine ausgezeichnete Rolle. Um Beweis und Aussage komplett nachzuvollziehen betrachte [Bra07, II §6]. ■

Mit Hilfe obiger Aussage über den Interpolationsfehler ist es uns nun möglich ein erstes Resultat der Finite-Elemente-Approximation zu beweisen:

Satz 13.3 (H^1 -Fehler)

Sei $f \in L^2(\Omega)$ und Ω erfülle die Voraussetzungen aus Satz 13.1. Es sei zudem eine Familie $\{\mathcal{T}_h\}$ von zulässigen und quasiuniformen Triangulierungen gegeben. Außerdem sei V_h der diskrete Ansatzraum der linearen Ansatzfunktionen aus (12.9). Sind $y \in V = H_0^1(\Omega)$ und $y_h \in V_h$ die eindeutigen Lösungen von (13.2) bzw. (12.8), dann existiert eine Konstante $c > 0$, unabhängig von f und h ,¹⁴⁶ s.d.

$$\|y - y_h\|_{H^1(\Omega)} \leq ch \|f\|_{L^2(\Omega)}$$

gilt. □

BEWEIS ¹⁴⁷ Die Aussage ist eine direkte Folgerung der Sätze 11.1, 13.2 und 13.1. Mit ihnen gilt:

$$\|y - y_h\|_{H^1(\Omega)} \leq c_{11.1} \|y - I_h y\|_{H^1(\Omega)} \leq c_{11.1} c_{13.2} h \|y\|_{H^2(\Omega)} \leq c_{11.1} c_{13.2} c_{13.1} h \|f\|_{L^2(\Omega)}$$

Hierbei ist zu beachten, dass sich $I_h y$ auch mit der Basis aus (12.9) darstellen lässt: Aus den homogenen Dirichlet-Randwerten folgt $y(x_i) = 0$ für $x_i \in \Gamma$, also

$$(I_h y)(x) = \sum_{x_i \notin \Gamma} y(x_i) \varphi_i(x)$$

¹⁴⁴Im Folgenden bezeichnen wir die Konstante c der Übersicht halber mit $c_{13.2}$.

¹⁴⁵Siehe z. B. [Bra07, II 6.3].

¹⁴⁶Im Folgenden bezeichnen wir die Konstante c der Übersicht halber mit $c_{13.3}$.

¹⁴⁷Siehe auch [Mey11, Satz 13.17] oder [Her12, Satz 16.5]. ■

Wegen $H^1(\Omega) \hookrightarrow L^2(\Omega)$ folgt aus obigem Resultat sofort

$$\|y - y_h\|_{L^2(\Omega)} \leq c_{13.3} h \|f\|_{L^2(\Omega)}$$

Wir werden nun sehen, dass sich das Resultat noch um eine h -Potenz günstiger gestalten lässt, wenn der Fehler in der L^2 -Norm gemessen wird.¹⁴⁸ In der Literatur wird folgender Satz auch oft als Lemma von Aubin-Nitsche oder 'Nitsche-Trick' bezeichnet. Der Beweis wird mittels eines Dualitätsargument geschehen, das eine Folgerung des Riesz'schen Darstellungssatzes 4.2 ist.

Satz 13.4 (L^2 -Fehler)

Es seien die Voraussetzungen aus Satz 13.3 erfüllt. Dann gilt:

$$\|y - y_h\|_{L^2(\Omega)} \leq ch^2 \|f\|_{L^2(\Omega)}$$

mit einer Konstanten $c > 0$, unabhängig von f und h .¹⁴⁹ □

BEWEIS ¹⁵⁰ Wie bereits angedeutet, benutzen wir eine Folgerung des Riesz'schem Darstellungssatzes:

Da der Riesz-Isomorphismus eine Isometrie ist, d.h. dass die Norm von $y - y_h$ gleich der Norm des durch $y - y_h$ erzeugten Funktionals ist, folgt aus Satz 4.2:

$$\|y - y_h\|_{L^2(\Omega)} = \sup_{\substack{g \in L^2(\Omega) \\ g \neq 0}} \frac{|(y - y_h, g)_{L^2(\Omega)}|}{\|g\|_{L^2(\Omega)}}. \quad (13.6)$$

Wir betrachten nun die zu (11.1) adjungierte Gleichung, welche wie im Fall der Herleitung der Optimalitätsbedingung erster Ordnung durch (10.10), d.h. durch

$$a[v, p] = \int_{\Omega} gv \, dx \quad \forall v \in H_0^1(\Omega) \quad (13.7)$$

gegeben ist. Für jedes $g \in L^2(\Omega)$ existiert nach dem Lemma von Lax-Milgram 5.2 ein $p \in H_0^1(\Omega)$, welches (13.7) eindeutig löst. Für die Wahl von $y - y_h$ als Testfunktion in der adjungierten Gleichung erhält man

$$(y - y_h, g)_{L^2(\Omega)} = \int_{\Omega} g(y - y_h) dx = a[y - y_h, p] = a[y - y_h, p - I_h p].$$

Bei der letzten Umformung konnten wir aufgrund von $I_h p \in V_h$ die Galerkin-Orthogonalität (11.4) verwenden. Nach Satz 13.1 gilt $p \in H_0^1(\Omega)$ für jedes $g \in L^2(\Omega)$.¹⁵¹ Daher folgt aus Satz 13.2

$$\|p - I_h p\|_{H^1(\Omega)} \leq c_{13.2} h \|p\|_{H^2(\Omega)} \leq c_{13.2} c_{13.1} h \|g\|_{L^2(\Omega)}.$$

¹⁴⁸Man konnte dies bereits nach Satz 13.2 erahnen.

¹⁴⁹Im Folgenden bezeichnen wir die Konstante c der Übersicht halber mit $c_{13.4}$.

¹⁵⁰Siehe auch [Mey11, Satz 13.18] oder [Her12, Satz 16.7]. Vgl. auch [Bra07, II 7.6].

¹⁵¹Beachte dazu, dass (13.7) aufgrund der Symmetrie von a die Struktur der Poissongleichung (13.2) besitzt.

Aus der Beschränktheit von a , sowie Satz 13.3 erhalten wir also

$$\begin{aligned} |(y - y_h, g)_{L^2(\Omega)}| &= |a[y - y_h, p - I_h p]| \\ &\leq \alpha_0 \|y - y_h\|_{H^1(\Omega)} \|p - I_h p\|_{H^1(\Omega)} \\ &\leq c_{13.2} c_{13.1} c_{13.3} h^2 \|f\|_{L^2(\Omega)} \|g\|_{L^2(\Omega)}. \end{aligned}$$

Die Aussage erhält man dann durch Einsetzen in (13.6). ■

§14 Diskretisierung des Optimalsteuerungsproblems

Wir haben nun die Zustandsgleichung durch die Finite-Elemente-Methode diskretisiert und hierfür auch den Diskretisierungsfehler abgeschätzt. Wir wollen nun die restlichen Terme unseres Modellproblems (1.5) bzw. (10.6) für eine numerische Berechnung handhabbar machen und uns dann auch hier schlussendlich mit dem Diskretisierungsfehler beschäftigen.

Um zu einer endlichdimensionalen Aufgabe zu gelangen, sollten also noch Zielfunktional und Nebenbedingung diskretisiert werden, was insbesondere auch eine Diskretisierung der Steuerung impliziert. Wie bereits zu Beginn der Arbeit erwähnt, werden wir nun zunächst ein Vorgehen betrachten, bei dem die Steuerung explizit nicht diskretisiert wird, was die Approximationsfehler-Analyse sehr viel einfacher macht. Der Implementierungsaufwand erhöht sich dadurch allerdings, was uns aber nicht weiter interessieren wird, da für uns hier theoretische Ergebnisse Vorrang haben.

Für den weiteren Verlauf nehmen wir folgende Grundvoraussetzungen an, um die Diskussion und v.a. die Fehlerabschätzung zu vereinfachen. Im Bezug auf unser Modellproblem der optimalen Temperaturquelle sind diese größtenteils auch durchaus legitim:¹⁵²

- (1) Ω ist konvex und polygonal ($N = 2$) bzw. polyhedral ($N = 3$) berandet.
- (2) Für den Zielzustand y_d gilt: $y_d \in H^2(\Omega)$.
- (3) Die betrachteten Triangulierungen sind zulässig und quasiuniform.
- (4) Der diskrete Ansatzraum für die PDE ist der Raum der linearen finiten Elemente, d.h. durch (13.4) und (13.5) gegeben.
- (5) Die Schranken u_a und u_b in den Box-Restriktionen sind konstant.

(1) und (3) versichern uns, dass die Sätze 13.3 und 13.4 anwendbar sind. Der Lösungsoperator der Zustandsgleichung sei wieder $S : L^2(\Omega) \rightarrow L^2(\Omega)$.¹⁵³ Analog definieren wir als Lösungsoperator

¹⁵²Siehe auch [Her12, Voraussetzung 16.9 und S.108] oder [Mey11, S.76].

¹⁵³Siehe (7.6).

rator der diskreten Gleichung (12.8) mit $f = u$ durch:¹⁵⁴

$$S_h : u \mapsto y_h, S_h : L^2(\Omega) \rightarrow L^2(\Omega),$$

d.h.

$$\begin{aligned} \sum_{j=1}^{n_p} \int_{\Omega} \nabla \varphi_i \cdot \nabla \varphi_j \, dx \, y_j &= \int_{\Omega} u \varphi_i \, dx \quad \forall i = 1, \dots, n_p \\ \Leftrightarrow \int_{\Omega} \nabla y_h \cdot \nabla v_h \, dx &= \int_{\Omega} u v_h \, dx \quad \forall v_h \in V_h. \end{aligned} \tag{14.1}$$

wobei wir ausgenutzt haben, dass $y_h = \sum_i y_i \varphi_i(x)$ und $\{\varphi_i\}$ eine Basis von V_h ist. Die Steifigkeitsmatrix ist nach Lemma 12.1 invertierbar und es gilt $V_h \in L^2(\Omega)$, daher ist obiger Operator für alle $h > 0$ wohldefiniert. Zudem gilt:

Lemma 14.1

$S_h : L^2(\Omega) \rightarrow L^2(\Omega)$ ist linear und beschränkt, d.h. $S_h \in \mathcal{L}(L^2(\Omega))$ und es gilt:

$$\|S_h\|_{\mathcal{L}(L^2(\Omega))} \leq \frac{1}{\beta_0},$$

wobei β_0 die Koerzivitatskonstante der Poissongleichung ist. □

BEWEIS ¹⁵⁵ Aus der Linearitat von (14.1) folgt direkt die von S_h . Sei nun $u \in L^2(\Omega)$ beliebig und wir setzen $y_h = S_h u$. Aus der Koerzivitat der Bilinearform und (14.1) folgt dann

$$\begin{aligned} \|y_h\|_{L^2(\Omega)}^2 &\leq \|y_h\|_{H^1(\Omega)}^2 \\ &\leq \frac{1}{\beta_0} a[y_h, y_h] = \frac{1}{\beta_0} \int_{\Omega} u \, y_h \, dx \leq \frac{1}{\beta_0} \|u\|_{L^2(\Omega)} \|y_h\|_{L^2(\Omega)} \end{aligned}$$

und somit die Beschranktheit. ■

Die Norm von S_h ist unabhangig von h beschrankt. Zudem gilt fur S_h :

Korollar 14.1

Der diskrete Steuerungs-Zustands-Operator S_h erfullt

$$\|S - S_h\|_{\mathcal{L}(L^2(\Omega))} \leq ch^2$$

mit einer Konstanten $c > 0$, unabhangig von h .¹⁵⁶ □

BEWEIS ¹⁵⁷ Der Beweis folgt direkt aus Satz 13.4, wonach gilt:

$$\|(S - S_h)u\|_{L^2(\Omega)} \leq c_{13.4} h^2 \|u\|_{L^2(\Omega)}.$$

Aus der Definition der Operatornorm folgt dann die Behauptung. ■

¹⁵⁴Siehe auch [Mey11, S.76].

¹⁵⁵Siehe auch [Mey11, Lemma 14.1].

¹⁵⁶Im Folgenden bezeichnen wir die Konstante c der ubersicht halber mit $c_{14.1}$.

¹⁵⁷Siehe auch [Mey11, Korollar 14.2].

§14.1 Die variationelle Diskretisierung und der Diskretisierungsfehler

Wir werden nun zunächst also ein Optimalsteuerungsproblem betrachten, bei dem die Steuerung sowie die Ungleichungsnebenbedingungen nicht diskretisiert werden. Wir werden also nur den Zustand, also folglich die PDE, und den Zielzustand y_d diskretisieren. Dies geschieht folgendermaßen:

y_d wird durch $I_h y_d$ und die PDE durch die diskrete Finiten-Elemente-Gleichung ersetzt, s.d. wir mit Hilfe von (14.1) erhalten:

$$\left. \begin{aligned} \text{Minimiere } & J(y_h, u) = \frac{1}{2} \int_{\Omega} |y_h - I_h y_d|^2 dx + \frac{\lambda}{2} \int_{\Omega} |u|^2 dx \\ \text{unter } & \int_{\Omega} \nabla y_h \cdot \nabla v_h dx = \int_{\Omega} u v_h dx \quad \forall v_h \in V_h \\ \text{und } & u_a \leq u \leq u_b \text{ fastberallin } \Omega. \end{aligned} \right\} \quad (14.2)$$

Natürlich handelt es sich bei (14.2) immernoch um ein unendlichdimensionales Problem mit einer Lösung $\bar{u}^{(h)} \in L^2(\Omega)$. Dennoch ist es möglich Algorithmen zu entwerfen, um (14.2) mit dem Rechner zu lösen. Wir werden später noch darauf eingehen.

Wie bereits im vorhergehenden kontinuierlichen Fall betrachten wir die zu (14.2) äquivalente reduzierte Aufgabe¹⁵⁸:

$$\left. \begin{aligned} \text{Minimiere } & f(u) \stackrel{\text{def}}{=} \frac{1}{2} \|S_h u - I_h y_d\|_{L^2(\Omega)}^2 + \frac{\lambda}{2} \|u\|_{L^2(\Omega)}^2 \\ \text{unter } & u \in U_{ad} \stackrel{\text{def}}{=} \{u \in L^2(\Omega) : u_a \leq u(x) \leq u_b \text{ fast überall in } \Omega\}. \end{aligned} \right\} \quad (14.3)$$

Satz 14.1

(14.2) besitzt eine eindeutige Lösung $\bar{u}^{(h)} \in L^2(\Omega)$, die mit $\bar{y}^{(h)}$ und $p^{(h)}$ eindeutig durch die Lösung des folgenden Optimalitätssystems gegeben ist:

$$\left. \begin{aligned} \int_{\Omega} \nabla \bar{y}^{(h)} \cdot \nabla v_h dx &= \int_{\Omega} \bar{u}^{(h)} v_h dx \quad \forall v_h \in V_h \\ \int_{\Omega} \nabla p^{(h)} \cdot \nabla v_h dx &= \int_{\Omega} (\bar{y}^{(h)} - I_h y_d) v_h dx \quad \forall v_h \in V_h \\ (p^{(h)} + \lambda \bar{u}^{(h)}, v - u^{(h)})_{L^2(\Omega)} &\geq 0 \quad \forall v \in U_{ad}. \end{aligned} \right\} \quad (14.4)$$

□

BEWEIS ¹⁵⁹ Analog zum Vorgehen des Beweises von Satz 7.3 beweist man unter Ausnutzung von Satz 7.1 sowie der Lemmata 7.1 und 7.2, dass (14.2) bzw. (14.3) eine eindeutige Lösung $\bar{u}^{(h)} \in L^2(\Omega)$ besitzt. S_h ist ein linearer und stetiger Operator, weshalb auch die Zielfunktion in

¹⁵⁸Siehe auch (10.3).

¹⁵⁹Siehe auch [Mey11, Satz 14.4].

(14.3) differenzierbar ist. Daher ist Satz 10.3 anwendbar und wir erhalten als notwendige und hinreichende Optimalitätsbedingung:

$$(S_h^*(S_h \bar{u}^{(h)} - I_h y_d) + \lambda \bar{u}^{(h)}, v - \bar{u}^{(h)})_{L^2(\Omega)} \geq 0 \quad \forall v \in U_{ad}. \quad (14.5)$$

Es bleibt die Form von $S_h^* : L^2(\Omega) \rightarrow L^2(\Omega)$ nachzuweisen. Wir gehen analog zu Lemma 10.2 vor: Seien $f, g \in L^2(\Omega)$ beliebig. Wir betrachten die Gleichung für $y_h = S_h f$, d.h.

$$a[y_h, v_h] = \int_{\Omega} f v_h dx \quad \forall v_h \in V_h$$

und außerdem die Gleichung

$$a[v_h, p_h] = \int_{\Omega} g v_h dx \quad \forall v_h \in V_h \quad (14.6)$$

mit der eindeutigen Lösung $p_h \in S_h^*$. Setzt man nun p_h als Testfunktion in die erste und y_h in die zweite Gleichung ein, so liefert anschließende Subtraktion:¹⁶⁰

$$(S_h^* g, f)_{L^2(\Omega)} = (g, S_h f)_{L^2(\Omega)} = (g, y_h)_{L^2(\Omega)} = (p_h, f)_{L^2(\Omega)}.$$

Aus der Beliebigkeit von f und g folgt $p_h = S_h^* g$. Mit $p_h \stackrel{def}{=} S_h^*(\bar{y}^{(h)} - I_h y_d) = S_h^*(S_h \bar{u}^{(h)} - I_h y_d)$ folgt aus (14.5) die Aussage. ■

Für S_h^* , den diskreten Lösungsoperator von (14.6) gilt:

Lemma 14.2

Es existiert ein $c > 0$, unabhängig von h ,¹⁶¹ s.d.

$$\|S^* - S_h^*\|_{\mathcal{L}(L^2(\Omega))} \leq c h^2. \quad \square$$

BEWEIS ¹⁶² Mit Hilfe von Korollar 14.1 und Lemma 9.1, nachdem $\|S^*\| = \|S\|$ gilt, folgt direkt:

$$\|S^* - S_h^*\|_{\mathcal{L}(L^2(\Omega))} = \|(S - S_h)^*\|_{\mathcal{L}(L^2(\Omega))} = \|S - S_h\|_{\mathcal{L}(L^2(\Omega))} \leq c_{14.1} h^2. \quad \blacksquare$$

Mit Hilfe der bisher erarbeiteten Ergebnissen sowie den Variationsungleichungen von (10.6) und (14.2) können wir nun für den variationellen Fall den Approximationsfehler abschätzen.

Wie später im Fall der Voll-Diskretisierung handelt es sich hierbei um eine a priori Fehlerabschätzung. Wir schätzen den Approximationsfehler also ohne eine Berechnung der diskreten Lösung $\bar{u}^{(h)}$ ab.

¹⁶⁰Vergleiche auch den Beweis zu Lemma 10.2.

¹⁶¹Obwohl der Beweis zeigen wird, dass die Konstante c mit der aus Korollar 14.1 übereinstimmt, werden wir die Konstante c der Übersicht halber im Folgenden mit $c_{14.2}$ bezeichnen.

¹⁶²Siehe auch [Mey11, Lemma 14.5].

Satz 14.2 (Diskretisierungsfehler der variationellen Diskretisierung)

Sei $0 < \lambda \leq 1$, \bar{u} und $\bar{u}^{(h)}$ seien Lösungen von (10.6) und (14.2), \bar{y} und $\bar{y}^{(h)}$ die zugehörigen Zustände. Dann existiert eine Konstante $c > 0$, unabhängig von h und λ , s.d.

$$\|\bar{u} - \bar{u}^{(h)}\|_{L^2(\Omega)} + \|\bar{y} - \bar{y}^{(h)}\|_{H^1(\Omega)} \leq c \frac{1}{\lambda} h^2. \quad \square$$

BEWEIS ¹⁶³ Wir betrachten die Variationsungleichungen für (10.6) und (14.2)¹⁶⁴, d.h.

$$(S^*(S\bar{u} - y_d) + \lambda\bar{u}, v - \bar{u})_{L^2(\Omega)} \geq 0 \quad \forall v \in U_{ad} \quad (14.7)$$

$$(S_h^*(S_h\bar{u}^{(h)} - I_h y_d) + \lambda\bar{u}^{(h)}, v - \bar{u}^{(h)})_{L^2(\Omega)} \geq 0 \quad \forall v \in U_{ad}. \quad (14.8)$$

Da die Steuerung nicht diskretisiert wurde, sind beide zulässige Mengen der Variationsungleichungen durch U_{ad} gegeben. Dies ist der entscheidende Vorteil der variationellen Diskretisierung, wir können die kontinuierliche Lösung \bar{u} in (14.8) einsetzen.

Zudem kann man natürlich auch $\bar{u}^{(h)}$ in (14.7) einsetzen. Hierbei ist jeweils zu beachten, dass $\bar{u}, \bar{u}^{(h)} \in U_{ad}$ da die optimalen Lösungen natürlich zulässig sind. Die anschließende Addition der beiden Ungleichungen liefert uns:

$$\begin{aligned} 0 &\leq (S^*(S\bar{u} - y_d) + \lambda\bar{u}, \bar{u}^{(h)} - \bar{u})_{L^2(\Omega)} + (S_h^*(S_h\bar{u}^{(h)} - I_h y_d) + \lambda\bar{u}^{(h)}, \bar{u} - \bar{u}^{(h)})_{L^2(\Omega)} \\ &= \lambda(\bar{u}^{(h)} - \bar{u}, \bar{u} - \bar{u}^{(h)})_{L^2(\Omega)} + (S\bar{u} - y_d, S(\bar{u}^{(h)} - \bar{u}))_{L^2(\Omega)} + (S_h\bar{u}^{(h)} - I_h y_d, S_h(\bar{u} - \bar{u}^{(h)}))_{L^2(\Omega)} \\ &\quad + \underbrace{(S\bar{u} - y_d, S_h(\bar{u} - \bar{u}^{(h)} - (\bar{u} - \bar{u}^{(h)})))_{L^2(\Omega)}}_{=0} \\ &= -\lambda\|\bar{u}^{(h)} - \bar{u}\|_{L^2(\Omega)}^2 + (S_h\bar{u}^{(h)} - S\bar{u} + y_d - I_h y_d, S_h(\bar{u} - \bar{u}^{(h)}))_{L^2(\Omega)} \\ &\quad + (S\bar{u} - y_d, (S - S_h)(\bar{u}^{(h)} - \bar{u}))_{L^2(\Omega)} + \underbrace{(S_h\bar{u}^{(h)} - S\bar{u} + y_d - I_h y_d, S(\bar{u} - \bar{u}))_{L^2(\Omega)}}_{=0} \\ &= -\lambda\|\bar{u}^{(h)} - \bar{u}\|_{L^2(\Omega)}^2 + (S_h\bar{u}^{(h)} - S\bar{u}, S\bar{u} - S_h\bar{u}^{(h)})_{L^2(\Omega)} + (S_h\bar{u}^{(h)} - S\bar{u}, (S_h - S)\bar{u})_{L^2(\Omega)} \\ &\quad + (y_d - I_h y_d, S\bar{u} - S_h\bar{u}^{(h)})_{L^2(\Omega)} + (y_d - I_h y_d, (S_h - S)\bar{u})_{L^2(\Omega)} \\ &\quad + ((S - S_h)^*(S\bar{u} - y_d), \bar{u}^{(h)} - \bar{u})_{L^2(\Omega)}. \end{aligned}$$

Wegen $(S_h\bar{u}^{(h)} - S\bar{u}, S\bar{u} - S_h\bar{u}^{(h)})_{L^2(\Omega)} = -\|S_h\bar{u}^{(h)} - S\bar{u}\|_{L^2(\Omega)}^2$ folgt aus obiger Ungleichung

$$\begin{aligned} \lambda\|\bar{u}^{(h)} - \bar{u}\|_{L^2(\Omega)}^2 + \|S_h\bar{u}^{(h)} - S\bar{u}\|_{L^2(\Omega)}^2 &\leq (y_d - I_h y_d, (S_h - S)\bar{u})_{L^2(\Omega)} \\ &\quad + (S_h\bar{u}^{(h)} - S\bar{u}, (S_h - S)\bar{u})_{L^2(\Omega)} + (y_d - I_h y_d, S\bar{u} - S_h\bar{u}^{(h)})_{L^2(\Omega)} + ((S - S_h)^*(S\bar{u} - y_d), \bar{u}^{(h)} - \bar{u})_{L^2(\Omega)}. \end{aligned}$$

Die letzten drei Terme der rechten Seite können mit Hilfe der Young'schen Ungleichung¹⁶⁵ mit

¹⁶³Siehe auch [Mey11, Satz 14.6] und [Her12, Satz16.13].

¹⁶⁴Vgl. auch (10.7) und (14.5).

¹⁶⁵Es gilt für H Hilbertraum: $|(a, b)_H| \leq \|a\|_H \|b\|_H \leq \frac{1}{2\alpha} \|a\|_H^2 + \frac{\alpha}{2} \|b\|_H^2 \quad \forall a, b \in H, \alpha > 0$.

$\alpha = 1$ bzw. $\alpha = \lambda$ abgeschätzt werden:

$$\begin{aligned} & \lambda \|\bar{u}^{(h)} - \bar{u}\|_{L^2(\Omega)}^2 + \|S_h \bar{u}^{(h)} - S \bar{u}\|_{L^2(\Omega)}^2 \leq (y_d - I_h y_d, (S_h - S) \bar{u})_{L^2(\Omega)} \\ & + \frac{1}{2} \|S_h \bar{u}^{(h)} - S \bar{u}\|_{L^2(\Omega)}^2 + \frac{1}{2} \|(S_h - S) \bar{u}\|_{L^2(\Omega)}^2 \\ & + \frac{1}{2} \|S \bar{u} - S_h \bar{u}^{(h)}\|_{L^2(\Omega)}^2 + \frac{1}{2} \|y_d - I_h y_d\|_{L^2(\Omega)}^2 \\ & + \frac{\lambda}{2} \|\bar{u}^{(h)} - \bar{u}\|_{L^2(\Omega)}^2 + \frac{1}{2\lambda} \|(S - S_h)^*(S \bar{u} - y_d)\|_{L^2(\Omega)}^2. \end{aligned}$$

Somit erhalten wir auf diese Weise:

$$\begin{aligned} \frac{\lambda}{2} \|\bar{u}^{(h)} - \bar{u}\|_{L^2(\Omega)}^2 & \leq \|y_d - I_h y_d\|_{L^2(\Omega)} \|(S_h - S) \bar{u}\|_{L^2(\Omega)} + \frac{1}{2} \|(S_h - S) \bar{u}\|_{L^2(\Omega)}^2 \\ & + \frac{1}{2} \|y_d - I_h y_d\|_{L^2(\Omega)}^2 + \frac{1}{2\lambda} \|(S^* - S_h^*)(S \bar{u} - y_d)\|_{L^2(\Omega)}^2 \\ & \leq \|y_d - I_h y_d\|_{L^2(\Omega)} \|S_h - S\|_{\mathcal{L}(L^2(\Omega))} \|\bar{u}\|_{L^2(\Omega)} + \frac{1}{2} \|S_h - S\|_{\mathcal{L}(L^2(\Omega))}^2 \|\bar{u}\|_{L^2(\Omega)}^2 \\ & + \frac{1}{2} \|y_d - I_h y_d\|_{L^2(\Omega)}^2 + \frac{1}{2\lambda} \|S^* - S_h^*\|_{\mathcal{L}(L^2(\Omega))}^2 \|S \bar{u} - y_d\|_{L^2(\Omega)}^2. \end{aligned}$$

Zur weiteren Abschätzung benutzen wir nun Korollar 14.1 bzw. Lemma 14.2 für $\|S - S_h\|_{\mathcal{L}(L^2(\Omega))}$ bzw. $\|S^* - S_h^*\|_{\mathcal{L}(L^2(\Omega))}$, sowie, weil nach Voraussetzung $y_d \in H^2(\Omega)$ gilt¹⁶⁶, die Interpolationsfehlerabschätzung aus Satz 13.2¹⁶⁷ für $\|y_d - I_h y_d\|_{L^2(\Omega)}$. Außerdem folgt aus $\bar{u} \in U_{ad}$ bzw. der radialen Unbeschränktheit des Zielfunktional bei $\lambda > 0$ die gleichmäßige Beschränktheit von $\|\bar{u}\|_{L^2(\Omega)}$. Aus der Beschränktheit von S folgt dann auch die gleichmäßige Beschränktheit von $\|S \bar{u} - y_d\|_{L^2(\Omega)}$.

Zusammenfassend erhalten wir also:

$$\begin{aligned} \frac{\lambda}{2} \|\bar{u}^{(h)} - \bar{u}\|_{L^2(\Omega)}^2 & \leq c_{13.2} c_{14.1} \|y_d\|_{H^2(\Omega)} \|\bar{u}\|_{L^2(\Omega)} h^4 + \frac{1}{2} c_{14.1}^2 \|\bar{u}\|_{L^2(\Omega)}^2 h^4 \\ & + \frac{1}{2} c_{13.2}^2 \|y_d\|_{H^2(\Omega)}^2 h^4 + \frac{1}{2\lambda} c_{14.2}^2 \|S \bar{u} - y_d\|_{L^2(\Omega)}^2 h^4. \end{aligned}$$

Daraus können wir nun folgern: Es existiert eine Konstante $C_\alpha > 0$, abhängig von obigen Abschätzungen, unabhängig von h und λ , s.d. gilt:

$$\frac{\lambda}{2} \|\bar{u}^{(h)} - \bar{u}\|_{L^2(\Omega)}^2 \leq C_\alpha \left(1 + \frac{1}{\lambda}\right) h^4.$$

Multiplizieren wir beide Seiten mit λ , so folgt, wegen $\lambda \leq 1$ nach Voraussetzung: Es existiert eine Konstante $C_\beta = C_\beta(C_\alpha) > 0$, unabhängig von h und λ , s.d. gilt:

$$\|\bar{u}^{(h)} - \bar{u}\|_{L^2(\Omega)} \leq C_\beta \frac{1}{\lambda} h^2.$$

¹⁶⁶Bemerke, dass daraus auch die gleichmäßige Beschränktheit von $\|y_d\|_{H^2(\Omega)}$ folgt.

¹⁶⁷Für $t = 0$.

Für $\bar{y} - \bar{y}^{(h)}$ ergibt sich die Abschätzung aus

$$\begin{aligned} \|\bar{y}^{(h)} - \bar{y}\|_{H^1(\Omega)} &= \|S_h \bar{u}^{(h)} - S \bar{u}\|_{H^1(\Omega)} \\ &\leq \|S_h(\bar{u}^{(h)} - \bar{u})\|_{H^1(\Omega)} + \|(S_h - S)\bar{u}\|_{H^1(\Omega)} \\ &\leq \|S_h\|_{\mathcal{L}(L^2(\Omega), H^1(\Omega))} \|\bar{u}^{(h)} - \bar{u}\|_{L^2(\Omega)} + \|S_h - S\|_{\mathcal{L}(L^2(\Omega), H^1(\Omega))} \|\bar{u}\|_{L^2(\Omega)} \\ &\leq \frac{1}{\beta_0} C_\beta \frac{1}{\lambda} h^2 + c_{14.1} \|\bar{u}\|_{L^2(\Omega)} h^2. \end{aligned}$$

Somit folgt: Es existiert eine Konstante $C_\gamma > 0$, abhängig von obigen Abschätzungen, unabhängig von h und λ , s.d. gilt:

$$\|\bar{y} - \bar{y}^{(h)}\|_{H^1(\Omega)} \leq \frac{1}{\lambda} C_\gamma h^2.$$

Beachte dazu Lemma 14.1, Korollar 14.1, die Form des Lösungsoperators S_h nach Definition, obiges Ergebnis für $\bar{u} - \bar{u}^{(h)}$, sowie die bereits erwähnte gleichmäßige Beschränktheit von $\|\bar{u}\|_{L^2(\Omega)}$. Die Addition der beiden Abschätzungen liefert dann die Aussage für $c = C_\beta + C_\gamma$. ■

Aus Satz und Beweis kann man einige Schlüsse ziehen:¹⁶⁸

Anmerkung 14.1

Aus dem Beweis kann man ersehen dass gilt: Es existiert eine Konstante $c > 0$, unabhängig von h s.d. gilt:

$$\|\bar{u} - \bar{u}^{(h)}\|_{L^2(\Omega)} + \|\bar{y} - \bar{y}^{(h)}\|_{H^1(\Omega)} \leq ch^2$$

für den Fall, dass $\lambda > 1$, gilt. Im praktischen Fall spielt dies jedoch i. A. eine untergeordnete Rolle. Wichtiger ist der Fall $\lambda \ll 1$, weshalb die Abhängigkeit des Approximationsfehler für kleine λ explizit berücksichtigt wurde. Dennoch lässt sich erkennen, dass trotz einer quadratischen Konvergenz des Fehlers bezüglich h , dieser für kleine Werte von λ dennoch sehr groß werden kann. Außerdem fällt auf wie wichtig die notwendigen und hinreichenden Optimalitätsbedingungen in Form der Variationsungleichung aus §10 sind, da der Beweis von Satz 14.2 im Wesentlichen auf diesen beruht.

Es fällt außerdem auf, dass in die Konstante c Daten der exakten Lösung eingehen, was typisch für a priori Abschätzungen ist.

Zudem kann man erkennen, dass die Approximationsfehlerordnung ebenso gut ist, wie die des reinen Finiten-Elemente-Fehlers¹⁶⁹, dazu aber später mehr. □

Das Problem der Implementierung führt zu folgendem Gedankenspiel, bei dem vorherige Überlegungen, wieder v.a. aus §10, eine entscheidende Rolle spielen.¹⁷⁰ Eine Übersicht dazu findet sich auch in [H⁺09, 3], für ausführlichere Informationen sei auch auf [Hin05] hingewiesen.

Anmerkung 14.2

Da die Steuerung nicht diskretisiert wurde, sich dementsprechend also im Funktionenraum befindet, kann man auf (14.2) also nicht ohne Weiteres ein Verfahren der nichtlinearen Optimierung, wie etwa die Aktive-Mengen-Strategie¹⁷¹, anwenden. Wir können jedoch das Optimali-

¹⁶⁸Siehe auch [Her12, Bemerkung 16.14] oder [Mey11, Bemerkungen 14.7 und 14.8].

¹⁶⁹Siehe Satz 13.4 zum L^2 -Fehler.

¹⁷⁰Siehe auch [Her12, S.109] oder [Mey11, Bemerkung 14.9].

¹⁷¹Siehe dazu etwa [Trö09, 2.12.4].

tätssystem (14.4) benutzen um eine Lösung mit dem Rechner zu finden.

Ähnlich wie im kontinuierlichen Optimalitätssystem aus §10.2 lässt das System (14.4) eine punktweise Interpretation zu und kann so, analog zu (10.19), auf folgende Projektionsformel übergeführt werden:

$$\bar{u}^{(h)}(x) = P_{[u_a(x), u_b(x)]} \left(-\frac{1}{\lambda} p^{(h)}(x) \right).$$

Weil $p^{(h)}$ als Lösung der diskreten adjungierten Gleichung in V_h liegt gilt: Die variationell-diskretisierte Lösung $\bar{u}^{(h)}$ ist die punktweise Projektion einer Funktion auf V_h auf das in unserem Fall konstante Intervall $[u_a, u_b]$. Eine solche Funktion lässt sich dann im Rechner abbilden. Der erhebliche Mehraufwand bei der Implementierung kommt dann aber daher, dass die projizierte Funktion nicht mehr auf dem Triangulierungsgitter 'lebt'. \square

§14.2 Die Voll-Diskretisierung und der Diskretisierungsfehler

Zurückblickend könnte also die Diskretisierung der Steuerung den Vorteil haben die Implementierung zu vereinfachen und den Einsatz fertiger Software der endlichdimensionalen nichtlinearen Optimierung zu ermöglichen. Genau dies wird auch der Fall sein und so fordern wir nun, dass auch die Steuerung aus einem endlichdimensionalen U_h gewählt wird. Wir erhalten somit folgende Aufgabe:

$$\left. \begin{array}{l} \text{Minimiere } J(y_h, u_h) = \frac{1}{2} \int_{\Omega} |y_h - I_h y_d|^2 dx + \frac{\lambda}{2} \int_{\Omega} |u_h|^2 dx \\ \text{unter } u_h \in U_h \\ \text{und } \int_{\Omega} \nabla y_h \cdot \nabla v_h dx = \int_{\Omega} u_h v_h dx \quad \forall v_h \in V_h \\ \text{sowie } u_a \leq u_h \leq u_b \text{ fast überall in } \Omega. \end{array} \right\} \quad (14.9)$$

Zur Wahl von U_h betrachten wir von nun an zwei spezielle Beispiele:¹⁷²

Beispiel 14.1

Sei $\Omega \in \mathbb{R}^N$ erneut in n_t Elemente (Dreiecke oder Tetraeder für $N = 2$ oder $N = 3$) zerlegt. Als Möglichkeiten für U_h betrachten wir

- 1) $U_h^{(1)} \stackrel{\text{def}}{=} V_h = \{v \in C(\bar{\Omega}) : v|_{T_i} \in \mathbb{P}_1(T_i) \quad \forall i \in \{1, \dots, n_t\}\}.$
- 2) $U_h^{(0)} \stackrel{\text{def}}{=} \{v \in L^2(\Omega) : v|_{T_i} \in \mathbb{P}_0(T_i) \quad \forall i \in \{1, \dots, n_t\}\} = \{v \in L^2(\Omega) : v|_{T_i} = \text{konstant} \quad \forall i \in \{1, \dots, n_t\}\}.$

Es handelt sich in beiden Fällen um Finite-Elemente-Räume auf demselben Gitter wie V_h .

In 1) benutzen wir also wie bei der PDE den Raum der linearen finiten Elemente aus Def. 12.4. Wir wissen, dass dann $\dim U_h^{(1)} = n_p$ ¹⁷³ gilt, eine Basis ist durch (12.3) gegeben.

¹⁷²Siehe auch [Her12, S.109 f.] oder [Her12, Beispiel 14.10].

¹⁷³ n_p entspricht wieder der Anzahl der Knoten, siehe Beispiel 12.2.

In 2) approximieren wir die Steuerung durch stückweise konstante Funktionen. Durch

$$\{\psi_i\}_{i=1}^{n_t} \subset U_h^{(0)}, \psi_i(x) = \begin{cases} 1, & x \in T_i \\ 0, & x \notin T_i \end{cases} \quad (14.10)$$

ist, wie man leicht erkennt, eine Basis dieses Raumes gegeben. Außerdem gilt dann $\dim U_h^{(0)} = n_t$.

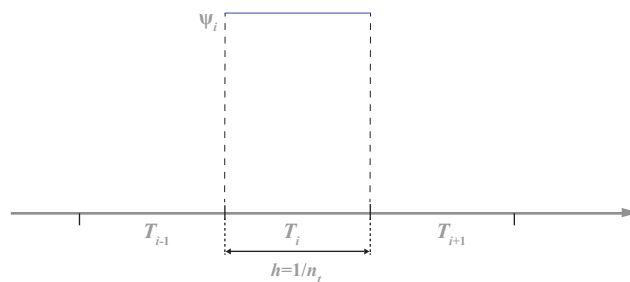


Abbildung 11: Basisdarstellung für die Approximation durch stückweise konstante Funktionen im uniform triangulierten eindimensionalen Fall.¹⁷⁴

Diese Funktionen sind nun nicht mehr schwach differenzierbar, weil sie offensichtlich nicht stetig sind.¹⁷⁵ Im Sinne eines Galerkin-Verfahrens, wäre dieser Raum für die Diskretisierung der PDE nicht zu gebrauchen. Dass wir die stückweise konstanten Ansatzfunktionen aber dennoch zur Diskretisierung der Steuerung benutzen können liegt daran, dass die Steuerung nur auf der rechten Seite der PDE auftaucht und somit nicht differenziert werden muss. \square

Es ist keinesfalls klar, dass es sich bei (14.9) um ein endlichdimensionales Optimalsteuerungsproblem handelt. Dies hängt davon ab, inwiefern sich die Wahl von U_h auf die Ungleichungsnebenbedingungen auswirkt.

Bei unseren Beispierräumen aus Beispiel 14.1 wollen wir dies nun genauer untersuchen.

Im ersten Fall, also für $U_h^{(1)}$ gilt

$$u_h(x) = \sum_{i=1}^{n_p} u_i \varphi_i(x). \quad 176$$

¹⁷⁴Vgl. [Mey11, S.82].

¹⁷⁵Vgl. auch Beispiel 3.1.

¹⁷⁶Siehe auch (12.1).

Analog zum Beweis des folgenden Lemmas überlegt man sich, da die Ansatzfunktionen φ_i für $i = 1, \dots, n_p$ nichtnegativ sind und

$$\varphi_i(x_j) = \delta_{ij}, \quad \sum_{i=1}^{n_p} \varphi_i(x) = 1, \quad \forall i = 1, \dots, n_p, \quad \forall x \in \Omega$$

erfüllen, dass folgende Äquivalenz gilt:

$$u_a \leq u_h(x) \leq u_b \text{ fast überall in } \Omega \Leftrightarrow u_a \leq u_i \leq u_b \quad \forall i \in \{1, \dots, n_p\}. \quad (14.11)$$

Da $u_h \in U_h$, wie wir noch sehen werden, in beiden Fällen durch endlich viele Box-Ungleichungen realisiert werden kann, ist dies für unsere Überlegungen von erheblicher Bedeutung, da uns dadurch ermöglicht wird, (14.9) in ein endlichdimensionales quadratisches Programm zu überführen:¹⁷⁷

Die sogenannte Massenmatrix $M \in \mathbb{R}^{n_p \times n_p}$ wird dann definiert durch

$$M_{ij} \stackrel{\text{def}}{=} \int_{\Omega} \varphi_i \varphi_j dx.$$

Damit zeigt sich dann bezüglich der diskreten PDE in (14.9) folgende Äquivalenz:

$$\sum_{j=1}^{n_p} \int_{\Omega} \nabla \varphi_i \cdot \nabla \varphi_j dx y_j = \sum_{j=1}^{n_p} \int_{\Omega} \varphi_i \varphi_j dx u_j \quad \forall i = 1, \dots, n \Leftrightarrow A \vec{y} = M \vec{u}$$

mit A , der Steifigkeitsmatrix aus (12.7), das Zielfunktional lässt sich dann ausdrücken durch

$$J(y_h, u_h) = \tilde{J}(\vec{y}, \vec{u}) = \frac{1}{2} (\vec{y} - \vec{y}_d)^T M (\vec{y} - \vec{y}_d) + \frac{\lambda}{2} \vec{u}^T M \vec{u}$$

mit $(\vec{y}_d)_i = (I_h y_d)(x_i)$, $i = 1, \dots, n_p$.

Um (14.9) nun schlussendlich in ein äquivalentes quadratisches Programm umzuwandeln definieren wir noch $\vec{z}, \vec{r}, \vec{q} \in \mathbb{R}^{2n_p}$ und $G \in \mathbb{R}^{n_p \times 2n_p}, Q, H \in \mathbb{R}^{2n_p \times 2n_p}$ durch:

$$\begin{aligned} \vec{z} &\stackrel{\text{def}}{=} \begin{pmatrix} \vec{y} \\ \vec{u} \end{pmatrix}, \quad Q \stackrel{\text{def}}{=} \begin{pmatrix} M & 0 \\ 0 & \lambda M \end{pmatrix}, \quad \vec{q} \stackrel{\text{def}}{=} \begin{pmatrix} M \vec{y}_d \\ 0 \end{pmatrix}, \\ G &\stackrel{\text{def}}{=} \begin{pmatrix} A & -M \end{pmatrix}, \quad H \stackrel{\text{def}}{=} \begin{pmatrix} 0 & -I \\ 0 & I \end{pmatrix}, \quad \vec{r} \stackrel{\text{def}}{=} \begin{pmatrix} -\vec{u}_a \\ \vec{u}_b \end{pmatrix}, \end{aligned} \quad (14.12)$$

wobei I die $(n_p \times n_p)$ -Einheitsmatrix ist und $\vec{u}_a, \vec{u}_b \in \mathbb{R}^{n_p}$, $\vec{u}_a \stackrel{\text{def}}{=} (u_a, \dots, u_a)$, \vec{u}_b wird analog definiert.

¹⁷⁷Siehe auch [Mey11, S.83] oder [Her12, S.110 ff.].

Dann ist folgendes quadratisches Programm äquivalent zu (14.9), wobei wir den konstanten Anteil des Zielfunktional $\frac{1}{2} \vec{y}_d^T M \vec{y}_d$ weglassen werden, da dieser bei der Optimierung keine Rolle spielt:

$$\left. \begin{array}{l} \text{Minimiere } \frac{1}{2} \vec{z}^T Q \vec{z} - \vec{q}^T \vec{z} \\ \text{unter } G \vec{z} = 0 \\ \text{und } H \vec{z} \leq \vec{r}. \end{array} \right\} \quad (14.13)$$

Nun betrachten wir noch den zweiten Fall der konstanten Ansatzfunktionen, d.h. $U_h = U_h^{(0)}$. Wir werden diesen Fall genauer betrachten da wir an diesem auch im Falle der Voll-Diskretisierung den Approximationsfehler abschätzen wollen.

Wie in (14.11) zeigen wir:

Lemma 14.3

Sei $u_h \in U_h^{(0)}$ gegeben, d.h. es existiert ein $\vec{u} \in \mathbb{R}^{n_t}$ mit

$$u_h = \sum_{i=1}^{n_t} u_i \psi_i(x).$$

Dann gilt:

$$u_a \leq u_h(x) \leq u_b \text{ fast überall in } \Omega \Leftrightarrow u_a \leq u_i \leq u_b \quad \forall i \in \{1, \dots, n_t\}. \quad \square$$

BEWEIS ¹⁷⁸ Exemplarisch betrachten wir die obere Schranke u_b , für u_a geschieht die Argumentation analog. Es sei $u_h \in U_h$ und $u_h(x) \leq u_b \quad \forall x \in \Omega$. Sei $i \in \{1, \dots, n_t\}$ beliebig. Nach (14.10) gilt somit $\forall x \in T_i$:

$$u_b \geq u_h(x) = \sum_{j=1}^{n_t} u_j \psi_j(x) = u_i \psi_i(x) = u_i.$$

Aus der Beliebigkeit von i folgt somit die Hinrichtung. Zum Beweis der Rückrichtung nehmen wir an, dass $u_i \leq u_b \quad \forall i \in \{1, \dots, n_t\}$. Da die ψ_i nichtnegativ sind folgt dann

$$u_h(x) = \sum_{i=1}^{n_t} u_i \psi_i(x) \leq \max_{1 \leq i \leq n_t} u_i \underbrace{\sum_{i=1}^{n_t} \psi_i(x)}_{=1 \text{ fast überall in } \Omega} \leq u_b$$

für fast alle $x \in \Omega$ und damit die Aussage. ■

¹⁷⁸Siehe auch [Her12, Lemma 16.15] oder [Mey11, Lemma 14.11].

Nach (14.10) gilt:

$$\sum_i \psi_i(x) = 0 \quad \forall x \in \partial T_i, \quad i \in \{1, \dots, n_t\}.$$

$\bigcup_i \partial T_i$ ist allerdings eine Nullmenge und spielt daher bei den Elementen von (14.9) keine Rolle. Mit Lemma 14.3 ist es nun auch im Fall Beispiel 14.12) möglich ein zu (14.9) äquivalentes quadratisches Programm herzuleiten:¹⁷⁹

Seien dazu die Einträge zweier Matrizen $N \in \mathbb{R}^{n_t \times n_t}$ und $R \in \mathbb{R}^{n_p \times n_t}$ wie folgt definiert:

$$N_{ij} \stackrel{\text{def}}{=} \int_{\Omega} \psi_i \psi_j dx, \quad R_{ij} \stackrel{\text{def}}{=} \int_{\Omega} \varphi_i \psi_j dx.$$

Somit gelten für diskrete PDE und Zielfunktional aus (14.9):

$$\sum_{i=1}^{n_p} \int_{\Omega} \nabla \varphi_i \cdot \nabla \varphi_j dx \, y_j = \sum_{i=1}^{n_t} \int_{\Omega} \varphi_i \psi_j dx \, u_j \quad \forall i = 1, \dots, n \Leftrightarrow A \vec{y} = R \vec{u},$$

$$J(y_h, u_h) = \tilde{J}(\vec{y}, \vec{u}) = \frac{1}{2}(\vec{y} - \vec{y}_d)^T M(\vec{y} - \vec{y}_d) + \frac{\lambda}{2} \vec{u}^T N \vec{u}.$$

Wir definieren nun $\vec{z}, \vec{q} \in \mathbb{R}^{n_p+n_t}$, $\vec{r} \in \mathbb{R}^{2n_t}$ und $H \in \mathbb{R}^{2n_t \times (n_p+n_t)}$ analog zu (14.12) sowie $Q \in \mathbb{R}^{(n_p+n_t) \times (n_p+n_t)}$ und $G \in \mathbb{R}^{n_p \times (n_p+n_t)}$ durch

$$Q \stackrel{\text{def}}{=} \begin{pmatrix} M & 0 \\ 0 & \lambda N \end{pmatrix}, \quad G \stackrel{\text{def}}{=} \begin{pmatrix} A & -R \end{pmatrix}$$

und erhalten auch dann ein quadratisches Programm der Form (14.13) das äquivalent zu (14.9) ist.

Wir fassen also zusammen:¹⁸⁰

Anmerkung 14.3

Mann kann also (14.9) sowohl im Fall von $U_h^{(0)}$ als auch im Fall von $U_h^{(1)}$ in ein äquivalentes endlichdimensionales quadratisches Programm umformulieren. Dies funktioniert aber in der Tat nicht immer so einfach. Im Fall quadratischer Ansatzfunktionen etwa gilt eine Äquivalenz wie in Lemma 14.3 nicht ohne Weiteres, da diese zum Teil negative Werte annehmen. In diesem Fall müssten weitere Approximationen vorgenommen werden (welche natürlich die Fehlerabschätzung zusätzlich beeinflussen) um ein endlichdimensionales quadratisches Programm zu erhalten. Ein solches quadratisches Programm kann dann jedenfalls mit Standardsoftware¹⁸¹ gelöst werden. I. A. darf dazu jedoch die Dimension, die wie beschrieben von der Gitterweite der Triangulierung abhängig ist, nicht zu groß sein. Allerdings muss die Gitterweite bei PDEs oft sehr klein gewählt werden um ein aussagekräftiges Ergebnis zu erhalten s.d. der Einsatz einer solchen Software zur Lösung von diskreten Optimalsteuerungsproblemen oft nur bei zweidimensionalen Gebieten und mit mäßiger Gitterweite möglich ist. □

¹⁷⁹Siehe auch [Mey11, S.84] oder [Her12, S.110 ff.].

¹⁸⁰Siehe auch [Her12, Bemerkung 16.16] oder [Mey11, Bemerkungen 14.12 und 14.13].

¹⁸¹Etwa quadprog.m in MATLAB.

Wie bei der variationellen Diskretisierung wollen wir nun auch im Fall der Voll-Diskretisierung den Approximationsfehler genauer untersuchen. Wir beschränken uns hierbei, wie bereits erwähnt, auf den zweiten Beispielunterraum, den Fall der stückweise konstanten Ansatzfunktionen, also $U_h = U_h^{(0)}$ und beginnen mit der Definition des reduzierten, diskreten Problems:¹⁸²

$$\begin{aligned} \text{Minimiere } f_h(u_h) &\stackrel{\text{def}}{=} \frac{1}{2} \|S_h u_h - I_h y_d\|_{L^2(\Omega)}^2 + \frac{\lambda}{2} \|u_h\|_{L^2(\Omega)}^2 \\ \text{unter } u &\in U_{ad}^{(h)} \stackrel{\text{def}}{=} \{u \in U_h \subset L^2(\Omega) : u_a \leq u(x) \leq u_b \text{ fast überall in } \Omega\}. \end{aligned} \quad (14.14)$$

Dies führt uns zu folgender Aussage:

Satz 14.3

Die Aufgabe (14.14) bzw. (14.9) besitzt eine eindeutige Lösung $\bar{u}_h \in U_h$ die folgender Variationsungleichung genügt:

$$(S_h^*(S_h \bar{u}_h - I_h y_d) + \lambda \bar{u}_h, v - \bar{u}_h)_{L^2(\Omega)} \geq 0 \quad \forall v \in U_{ad}^{(h)}. \quad (14.15)$$

BEWEIS ¹⁸³ Unser erstes Ziel wird es sein, die Lösbarkeit von (14.14) mit dem allgemeinen Satz 7.1 zu zeigen. Hierfür müssen wir lediglich nachweisen, dass $U_{ad}^{(h)} \subset L^2(\Omega)$ beschränkt, abgeschlossen und konvex ist:

Die Beschränktheit von $U_{ad}^{(h)}$ folgt sofort aus der Beschränktheit von U_{ad} , zudem ist $U_h = U_h^{(0)}$ als Unterraum von $L^2(\Omega)$ abgeschlossen und konvex. Somit ist $U_{ad}^{(h)} = U_{ad} \cap U_h$ als Durchschnitt zweier abgeschlossener und konvexer Mengen selbst abgeschlossen und konvex und erfüllt daher die Voraussetzungen von Satz 7.1.

$f_h : L^2(\Omega) \rightarrow \mathbb{R}$ ist nach Lemma 7.1 streng konvex¹⁸⁴, woraus die Eindeutigkeit der Lösung von (14.14) bzw. (14.9) folgt. Aus Satz 10.3 folgt aufgrund der Konvexität von $U_{ad}^{(h)}$ zuletzt sofort die Form der Variationsungleichung (14.15). ■

Anmerkung 14.4

Satz 14.3 ist mit geeigneter Vorarbeit ebenso auf den Fall $U_h = U_h^{(1)}$ übertragbar, d.h. auch in diesem Fall würde man eine eindeutige Lösung erhalten. □

Dies führt uns zu folgender Aussage über den Approximationsfehler, wobei es sich wie bereits im Fall der variationellen Diskretisierung um eine a priori Fehlerabschätzung handelt:

Satz 14.4 (Diskretisierungsfehler der Voll-Diskretisierung)

Es seien die Grundvoraussetzungen dieses Abschnitts erfüllt und $0 < \lambda \leq 1$. \bar{u} und $\bar{u}^{(h)}$ bezeichnen die Lösungen von (10.6) und (14.9) mit $U_h = U_h^{(0)}$ und zugehörigen Zuständen \bar{y} und \bar{y}_h . Dann existiert eine Konstante $c > 0$, unabhängig von h und λ , s.d.

$$\|\bar{u} - \bar{u}_h\|_{L^2(\Omega)} + \|\bar{y} - \bar{y}_h\|_{H^1(\Omega)} \leq c \frac{1}{\lambda} h. \quad \square$$

¹⁸²Vgl. auch (14.3).

¹⁸³Siehe auch [Mey11, Satz 14.14].

¹⁸⁴Siehe auch den Beweis zu Satz 14.1.

BEWEIS ¹⁸⁵ Im Prinzip verläuft der Beweis wie der von Satz 14.2:

Wir beginnen erneut mit den Variationsungleichungen für (10.6) und (14.9):

$$(S^*(S\bar{u} - y_d) + \lambda\bar{u}, v - \bar{u}) \geq 0 \quad \forall v \in U_{ad} \quad (14.16)$$

$$(S_h^*(S_h\bar{u}_h - I_h y_d) + \lambda\bar{u}_h, v_h - \bar{u}_h) \geq 0 \quad \forall v_h \in U_{ad}^{(h)}. \quad (14.17)$$

Im Unterschied zur variationellen Diskretisierung ist die kontinuierliche Lösung \bar{u} nun nicht mehr zulässig als Testfunktion für (14.17), denn i. A. gilt $\bar{u} \notin U_h^{(0)}$ und somit $\bar{u} \notin U_{ad}^{(h)}$, die exakte Lösung ist nicht stückweise konstant.

Der erste Schritt des Beweises wird also die Wahl, bzw. die Konstruktion, einer geeigneten Testfunktion sein, die nahe bei \bar{u} und gleichzeitig in $U_h^{(0)}$ liegt.

Wir benutzen dazu als Testfunktion

$$\bar{v}_h \stackrel{def}{=} \Pi_{U_h} \bar{u}.$$

Es handelt sich dabei um die orthogonale Projektion von $\bar{u} \in U = L^2(\Omega)$ auf den Unterraum $U_h = U_h^{(0)}$. $\Pi_{U_h} : L^2(\Omega) \rightarrow U_h$ ist also der Projektionsoperator auf U_h ¹⁸⁶, d.h. \bar{v}_h ist die eindeutige Lösung der Aufgabe

$$\min_{u_h \in U_h} \frac{1}{2} \|u_h - \bar{u}\|_{L^2(\Omega)}^2.$$

Nach (10.5) wird diese eindeutig charakterisiert durch die Variationsungleichung

$$(\bar{v}_h - \bar{u}, u_h - \bar{v}_h)_{L^2(\Omega)} \geq 0 \quad \forall u_h \in U_h. \quad 187$$

Da $\bar{v}_h \in U_h$ gilt, können wir dies in Koordinatenvektoren bezüglich der Basis $\{\psi_i\}_{i=1}^{n_t}$ ausdrücken und erhalten somit folgende Äquivalenz:

$$\begin{aligned} & (\bar{v}_h - \bar{u}, u_h - \bar{v}_h)_{L^2(\Omega)} \geq 0 \quad \forall u_h \in U_h \\ \Leftrightarrow & \int_{\Omega} \left(\sum_{j=1}^{n_t} (\bar{v}_h)_j \psi_j - \bar{u} \right) \left(u_h - \sum_{j=1}^{n_t} (\bar{v}_h)_j \psi_j \right) dx \geq 0 \quad \forall u_h \in U_h. \end{aligned} \quad (14.18)$$

Wählen wir nun speziell $u_h = \pm \psi_i + \bar{v}_h \in U_h$, $i \in \{1, \dots, n_t\}$ beliebig, als Testfunktion, erhalten wir somit aus (14.18):¹⁸⁸

$$0 \leq \int_{\Omega} \left(\sum_{j=1}^{n_t} (\bar{v}_h)_j \psi_j - \bar{u} \right) (\pm \psi_i) dx = \pm \int_{T_i} ((\bar{v}_h)_i - \bar{u}) dx, \quad (14.19)$$

¹⁸⁵Siehe auch [Mey11, Satz 14.15] und [Her12, Satz 16.18].

¹⁸⁶Vgl. auch Bsp. 10.1.

¹⁸⁷Bemerke, dass, da U_h linearer Raum ist, natürlich sogar ' $= 0$ ' gilt.

¹⁸⁸Beachte: $(\bar{v}_h)_i = \text{konst.}$, $\int_{T_i} ((\bar{v}_h)_i - \bar{u}) dx = 0 \Leftrightarrow (\bar{v}_h)_i = \frac{\int_{T_i} \bar{u} dx}{\int_{T_i} dx}$.

denn es gilt: $\psi_i|_{T_i} = 1$, $\psi_i|_{\Omega \setminus T_i} = 0$ und $\bar{v}_h|_{T_i} = (\bar{v}_h)_i$.
 Aus (14.19) folgt dann wiederum direkt:

$$(\bar{v}_h)_i = \frac{1}{|T_i|} \int_{T_i} \bar{u} \, dx. \quad (14.20)$$

Man kann es also so ausdrücken: Die $L^2(\Omega)$ -orthogonale Projektion auf U_h ergibt sich durch die Mittelwertbildung über die Elemente der Triangulierung.

Es bleibt die Zulässigkeit von \bar{v}_h , d.h. $\bar{v}_h \in U_{ad}^{(h)}$ zu überprüfen:

Exemplarisch untersuchen wir die obere Schranke u_b , für die untere Schranke u_a argumentiert man analog. Wegen $\bar{u} \in U_{ad}$ folgt $\bar{u}(x) \leq u_b$ fast überall in Ω , mit Hilfe von (14.20) erhalten wir also:

$$(\bar{v}_h)_i = \frac{1}{|T_i|} \int_{T_i} \bar{u} \, dx \leq \frac{1}{|T_i|} \int_{T_i} u_b \, dx \quad \forall i \in \{1, \dots, n_t\}.$$

Nach Lemma 14.3 ist dies äquivalent zu $\bar{v}_h \in U_{ad}$ und somit gilt auch $\bar{v}_h \in U_{ad}^{(h)}$.

Als nächstes können wir uns daran machen die Testfunktionen einzusetzen:

Wir setzen $\bar{v}_h = \Pi_{U_h} \bar{u}$ als Testfunktion in (14.17) und \bar{u}_h in (14.16) ein.¹⁸⁹ Die Addition der entstehenden Ungleichungen ergibt dann:

$$\begin{aligned} 0 &\leq (S^*(S\bar{u} - y_d) + \lambda\bar{u}, \bar{u}_h - \bar{u})_{L^2(\Omega)} + (S_h^*(S_h\bar{u}_h - I_h y_d) + \lambda\bar{u}_h, \bar{v}_h - \bar{u}_h)_{L^2(\Omega)} \\ &= (S^*(S\bar{u} - y_d) + \lambda\bar{u}, \bar{u}_h - \bar{u})_{L^2(\Omega)} + (S_h^*(S_h\bar{u}_h - I_h y_d) + \lambda\bar{u}_h, \bar{u} - \bar{u}_h)_{L^2(\Omega)} \\ &\quad + (S_h^*(S_h\bar{u}_h - I_h y_d) + \lambda\bar{u}_h, \bar{v}_h - \bar{u})_{L^2(\Omega)}. \end{aligned}$$

Man kann erkennen, dass die ersten beiden Summanden mit denen aus dem Beweis von Satz 14.2 übereinstimmen, man kann sie also analog abschätzen. Da zudem $\lambda > 0$ gilt, können wir folgern:

Es existiert eine Konstante $C_1 > 0$, unabhängig von h und λ , s.d. gilt:

$$\lambda^2 \|\bar{u} - \bar{u}_h\|_{L^2(\Omega)}^2 \leq C_1 h^4 + 2\lambda (S_h^*(S_h\bar{u}_h - I_h y_d) + \lambda\bar{u}_h, \bar{v}_h - \bar{u})_{L^2(\Omega)}. \quad (14.21)$$

Zur Abschätzung des letzten Summanden von (14.21) setzen wir nun

$$p_h \stackrel{\text{def}}{=} -S_h^*(S_h\bar{u}_h - I_h y_d)$$

und erhalten

$$(S_h^*(S_h\bar{u}_h - I_h y_d) + \lambda\bar{u}_h, \bar{v}_h - \bar{u})_{L^2(\Omega)} = - \underbrace{\int_{\Omega} p_h (\bar{v}_h - \bar{u}) \, dx}_{\stackrel{\text{def}}{=} I_1} + \lambda \underbrace{\int_{\Omega} \bar{u}_h (\bar{v}_h - \bar{u}) \, dx}_{\stackrel{\text{def}}{=} I_2}.$$

¹⁸⁹Dies ist zulässig, da $U_{ad}^{(h)} \subset U_{ad}$ gilt.

Es müssen nun also die beiden Terme I_1 und I_2 abgeschätzt werden:

Wir beginnen mit I_1 :

Nach Konstruktion gilt $p_h \in V_h \subset H^1(\Omega)$ und daher:¹⁹⁰

$$I_1 \leq \|p_h\|_{H^1(\Omega)} \|\bar{v}_h - \bar{u}\|_{H^1(\Omega)^*}.$$

Für die Abschätzung des ersten Faktors nutzen wir aus, dass wegen $u_a, u_b \in \mathbb{R}$ U_{ad} eine beschränkte Teilmenge von $L^2(\Omega)$ ist, weshalb $\|\bar{u}_h\|_{L^2(\Omega)}$ für $\bar{u}_h \in U_{ad}$ gleichmäßig beschränkt ist, was zudem natürlich auch für $\|I_h y_d\|_{L^2(\Omega)}$ gilt. Aus all dem folgt also, dass eine Konstante $c_1 > 0$, unabhängig von h und λ existiert, s.d. gilt:

$$\|p_h\|_{H^1(\Omega)} \leq \underbrace{\|S_h^*\|_{\mathcal{L}(L^2(\Omega), H^1(\Omega))}}_{\leq \frac{1}{\beta_0}} \left(\underbrace{\|S_h\|_{\mathcal{L}(L^2(\Omega))}}_{\leq \frac{1}{\beta_0}} \|\bar{u}_h\|_{L^2(\Omega)} + \|I_h y_d\|_{L^2(\Omega)} \right) \leq c_1.$$

Vgl. auch die Lemmata 9.1 und 14.1 sowie die Formen der Lösungsoperatoren S_h und S_h^* .

Nach der $H^1(\Omega)$ -Regularität der optimalen Steuerung nach Satz 10.6 und dem folgenden Lemma 14.4, gilt für den zweiten Faktor die Projektionsfehlerabschätzung

$$\|\bar{v}_h - \bar{u}\|_{H^1(\Omega)^*} \leq c_2 h^2 \|\bar{u}\|_{H^1(\Omega)}$$

für eine Konstante $c_2 > 0$, unabhängig von h , und somit insgesamt für I_1 nach Addition obiger Abschätzungen und wegen der nach den Vorbemerkungen trivialerweise erfüllten gleichmäßigen Beschränktheit von $\|\bar{u}\|_{H^1(\Omega)}$: Es existiert eine Konstante $C_2 > 0$, abhängig von obigen Abschätzungen, unabhängig von h und λ , s.d. gilt:

$$I_1 \leq C_2 h^2.$$

Im Fall von I_2 kann man so nicht vorgehen, da $\bar{u}_h \in U_h^{(0)} \not\subset H^1(\Omega)$ gilt. Stattdessen beginnen wir mit der Addition einer produktiven Null:

$$\begin{aligned} I_2 &= \int_{\Omega} \bar{u}_h (\bar{v}_h - \bar{u}) dx \\ &= \int_{\Omega} \bar{u} (\bar{v}_h - \bar{u}) dx + \int_{\Omega} (\bar{u}_h - \bar{u}) (\bar{v}_h - \bar{u}) dx \\ &\leq \|\bar{u}\|_{H^1(\Omega)} \|\bar{v}_h - \bar{u}\|_{H^1(\Omega)^*} + \|\bar{u}_h - \bar{u}\|_{L^2(\Omega)} \|\bar{v}_h - \bar{u}\|_{L^2(\Omega)}. \end{aligned}$$

Mit Hilfe der Young'schen Ungleichung¹⁹¹, die wir bereits im Beweis zu Satz 14.2 benutzt haben, sowie den Projektionsfehlerabschätzungen (14.22) und (14.23) aus Lemma 14.4 können wir folgern: Es existieren Konstanten $c_3, c_4 > 0$, abhängig von obigen Abschätzungen, unabhängig von h und λ , s.d. gilt:

$$\begin{aligned} I_2 &\leq \|\bar{u}\|_{H^1(\Omega)} \|\bar{v}_h - \bar{u}\|_{H^1(\Omega)^*} + \frac{1}{4} \|\bar{u}_h - \bar{u}\|_{L^2(\Omega)}^2 + \|\bar{v}_h - \bar{u}\|_{L^2(\Omega)}^2 \\ &\leq c_3 h^2 + \frac{1}{4} \|\bar{u}_h - \bar{u}\|_{L^2(\Omega)}^2 + c_4 h^2 \end{aligned}$$

¹⁹⁰Bemerke, dass das L^2 -Skalarprodukt in I_1 ein lineares und stetiges Funktional auf $H^1(\Omega)$ darstellt.

¹⁹¹Mit $\alpha = 2$.

und somit für $C_3 = c_3 + c_4$:

$$I_2 \leq C_3 h^2 + \frac{1}{4} \|\bar{u}_h - \bar{u}\|_{L^2(\Omega)}^2.$$

Wir fassen nun die bisherigen Ergebnisse zusammen und setzen die Abschätzungen in (14.21) ein und erhalten:

$$\begin{aligned} \lambda^2 \|\bar{u} - \bar{u}_h\|_{L^2(\Omega)}^2 &\leq C_1 h^4 + 2\lambda I_1 + 2\lambda^2 I_2 \\ &\leq C_1 h^4 + 2\lambda C_2 h^2 + 2\lambda^2 C_3 h^2 + \frac{\lambda^2}{2} \|\bar{u}_h - \bar{u}\|_{L^2(\Omega)}^2 \\ &\leq C_1 h^4 + 2C_2 h^2 + 2C_3 h^2 + \frac{\lambda^2}{2} \|\bar{u}_h - \bar{u}\|_{L^2(\Omega)}^2 \quad (\text{da } \lambda \leq 1). \end{aligned}$$

Da $h \leq \text{diam}(\Omega)$ folgt mit $C_4 = 2(C_2 + C_3)$:

$$\begin{aligned} \frac{\lambda^2}{2} \|\bar{u} - \bar{u}_h\|_{L^2(\Omega)}^2 &\leq C_1 h^4 + C_4 h^2 \\ &\leq C_1 h^2 + C_4 h^2 \\ &\leq C_\alpha h^2 \quad (\text{mit } C_\alpha = C_1 + C_4) \end{aligned}$$

und somit insgesamt: Es existiert eine Konstante $C_\beta = C_\beta(C_\alpha) > 0$, unabhängig von h und λ s.d. gilt:

$$\|\bar{u} - \bar{u}_h\|_{L^2(\Omega)}^2 \leq C_\beta \frac{1}{\lambda} h^2.$$

Analog zum abschließenden Vorgehen der Abschätzung für $\bar{y} - \bar{y}_h$ aus Satz 14.2 unter Einbezug von $h \leq \text{diam}(\Omega)$ und anschließender Addition beider Ergebnisse folgt dann die Behauptung. ■

Für die benutzten Projektionsfehlerabschätzungen im obigen Beweis kommt nun noch folgender Nachtrag:

Lemma 14.4

Sei $u \in H^1(\Omega)$ und $\{\mathcal{T}_h\}$ eine Familie von zulässigen, quasiuniformen Triangulierungen auf Ω . Dann gilt für die $L^2(\Omega)$ -orthogonale Projektion auf $U_h^{(0)}$: Es existieren Konstanten $c_1, c_2 > 0$, unabhängig von h , s.d.

$$\left\| \Pi_{U_h^{(0)}} u - u \right\|_{L^2(\Omega)} \leq c_1 h \|u\|_{H^1(\Omega)} \tag{14.22}$$

$$\left\| \Pi_{U_h^{(0)}} u - u \right\|_{H^1(\Omega)^*} \leq c_2 h^2 \|u\|_{H^1(\Omega)}. \tag{14.23}$$

□

BEWEIS Für Abschätzung (14.22) betrachte [G⁺07, Theorem 4.28] mit $r = k = 0$. (14.23) folgt dann, ähnlich wie im Beweis zu Satz 13.4 mit einem Dualitätsargument:

$$\left\| \Pi_{U_h^{(0)}} u - u \right\|_{H^1(\Omega)^*} = \sup_{\substack{v \in H^1(\Omega) \\ v \neq 0}} \frac{\left(\Pi_{U_h^{(0)}} u - u, v \right)_{L^2(\Omega)}}{\|v\|_{H^1(\Omega)}}.$$

Für beliebiges stückweise konstantes v_h erhalten wir aufgrund der Optimalität der Projektion:

$$\begin{aligned} \sup_{\substack{v \in H^1(\Omega) \\ v \neq 0}} \frac{\left(\Pi_{U_h^{(0)}} u - u, v \right)_{L^2(\Omega)}}{\|v\|_{H^1(\Omega)}} &= \sup_{\substack{v \in H^1(\Omega) \\ v \neq 0}} \frac{\left(\Pi_{U_h^{(0)}} u - u, v - v_h \right)_{L^2(\Omega)}}{\|v\|_{H^1(\Omega)}} \\ &= \|\Pi_{U_h^{(0)}} u - u\|_{L^2(\Omega)} \|v - v_h\|_{L^2(\Omega)}. \end{aligned}$$

Wählt man nun $v_h = \Pi_{U_h^{(0)}} v$ und benutzt für beide Terme die Abschätzung (14.22), so folgt die Behauptung. ■

Anmerkung 14.5

Der obige Beweis zeigt, dass Approximationsfehlerabschätzungen für Optimalsteuerungsprobleme selbst unter vereinfachten Voraussetzungen sehr umfangreich werden können. Auch die ganze theoretische Vorarbeit, etwa Optimalitätsbedingungen in Form der zusätzlichen $H^1(\Omega)$ -Regularität der optimalen Steuerung aus Satz 10.6, gehen hier, wie wir bereits bei der variationellen Diskretisierung bemerkt haben, in unsere Überlegungen ein.

Zudem kann man erkennen, dass sich im Fall der Voll-Diskretisierung die Approximationsfehlerordnung im Vergleich zur variationellen Diskretisierung um eine h -Potenz ungünstiger gestaltet. Wir werden darauf in einer abschließenden Analyse kurz eingehen. □

VI Zusammenfassung und Analyse

Bevor wir uns einem abschließenden Vergleich der beiden vorgestellten Diskretisierungsmethoden widmen, fassen wir das Vorgehen dieser Arbeit zusammen und gehen nochmals auf die wichtigsten Ergebnisse ein.

Ausgehend von der Poissongleichung und dem sich daraus ergebenden Optimalsteuerungsproblem der optimalen Temperaturquelle (1.5) bzw. (10.6) haben wir in Kapitel II zunächst die Lösbarkeit von linearen elliptischen PDEs untersucht. Dabei wurde nicht ein potentialtheoretischer Ansatz, sondern ein funktionalanalytische Zugang gewählt. Es wurde also nicht die Existenz einer klassischen Lösung untersucht, was etwa mit Hilfe von sogenannten Green'schen Funktionen, die als Fundamentallösung des Differentialoperators betrachtet werden können, geschehen könnte.¹⁹² Stattdessen haben wir einen Zugang gewählt, der flexibler und enger mit der numerischen Untersuchung verbunden ist. Die Frage nach der Existenz einer Lösung wurde somit ersetzt durch die Frage, in welchem Sinn und in welchen Funktionenräumen die Existenz einer Lösung gezeigt werden kann. Dies führte uns zum Begriff der schwachen Lösung in Sobolewräumen. Anschließend gelang es mit dem Lemma von Lax-Milgama 5.2 und der Poincaré-Friedrich-Ungleichung 5.3 die eindeutige Lösbarkeit der Poissongleichung im Satz von Friedrich 5.1 im schwachen Sinne zu zeigen.

Im Anschluss daran haben wir uns in Kapitel III mit der Lösbarkeit des Modellproblems als Ganzes beschäftigt. Die Einführung konvexer Räume und (strikt-)konvexer Funktionale führte uns, unter Berücksichtigung der weiteren Annahmen an das Modellproblem aus Anmerkung 7.1, zum Begriff des Steuerungs-Zustands-Operators, der die Lösungsabbildung der PDE darstellt. Darauf aufbauend konnten wir das Modellproblem der optimalen Temperaturquelle in Form eines reduzierten Zielfunktional schreiben und dessen eindeutige Lösbarkeit in Satz 7.3 im globalen Sinne zeigen.

Kapitel IV beschäftigte sich mit der Herleitung von Optimalitätsbedingungen, die die eindeutige Lösung des Modellproblems charakterisieren. Diese sind von erheblicher Bedeutung für die Herleitung des Diskretisierungsfehlers. Ausgehend von einer konvexen Menge zulässiger Steuerungen wurde als notwendige Optimalitätsbedingung lokal optimaler Steuerungen in Satz 10.1 eine Variationsungleichung hergeleitet. Diese konnte für ein konvexes Zielfunktional in Satz 10.2 zu einer hinreichenden Optimalitätsbedingung global optimaler Steuerungen erweitert werden. Mit Hilfe des Steuerungs-Zustands-Operators und dessen Adjunktion, welcher durch die adjungierte Gleichung hergeleitet wurde, konnte die Variationsungleichung in Form von Satz

¹⁹²Für den Fall der Wärmeleitungsgleichung sei etwa auf [Eva10, 2.3] hingewiesen.

10.3 als eindeutiges Charakteristikum der globalen Lösung des Modellproblems der optimalen Temperaturquelle beschrieben werden. Zudem wurde durch den Beweis der $H^1(\Omega)$ -Regularität der optimalen Steuerung in Satz 10.6 ein weiteres Hilfsmittel zur Bestimmung des Diskretisierungsfehlers gewonnen.

Die Diskretisierung der Aufgabe wurde dann in Kapitel V mit der Finiten-Elemente-Methode angegangen. Durch Definition eines zweckmäßigen endlichdimensionalen Ansatzraumes war es uns bei geeigneter Zerlegung des zugrunde liegenden Gebiets möglich, die Zustandsgleichung zu diskretisieren und gemäß Beispiel 12.3 in ein endlichdimensionales lineares Gleichungssystem zu überführen. Der Approximationsfehler wurde anschließend in den Sätzen 13.3 und 13.4 mit Hilfe des Interpolationsfehlers nach Satz 13.2 zunächst in der $H^1(\Omega)$ -Norm und schlussendlich in der $L^2(\Omega)$ -Norm bestimmt. Es zeigte sich, dass man dadurch das Resultat um eine h -Potenz günstiger gestalten konnte, wir erhielten also eine quadratische Ordnung des Fehlers bezüglich h .

Der Diskretisierungsfehler des Modellproblems wurde im Anschluss unter zwei verschiedenen Aspekten abgeschätzt. Bei der variationellen Diskretisierung wurde auf eine Diskretisierung der Steuerung verzichtet, die Voll-Diskretisierung bezog diese mit ein. Dadurch entstand ein endlichdimensionales Problem. Da hierbei aber die Regularität der Steuerung betrachtet werden muss, steigerte sich die Komplexität der benötigten Analysis.

Abschließend vergleichen wir nun in knapper Form die Resultate aus den Sätzen 14.2 und 14.4 unter Einbezug des Finiten-Elemente-Fehlers:¹⁹³

Im Fall der variationellen Diskretisierung fällt auf, dass die Approximationsfehlerabschätzung aus Satz 14.2 eine genau so gute Ordnung besitzt, wie die des reinen Finiten-Elemente-Fehlers aus Satz 13.4, es kommt kein zusätzlicher Fehler mit schlechterer h -Potenz hinzu. Durch die variationelle Diskretisierung verliert man also im Vergleich zur PDE-Diskretisierung nichts.

Im Vergleich dazu fällt uns bei Satz 14.4 auf, dass die Konvergenzordnung im Fall der Voll-Diskretisierung schlechter ist (h im Vergleich zu h^2). In diesem Fall ist die Ordnung des Diskretisierungsfehlers also nicht mehr so gut wie die des reinen Finite-Elemente-Fehler. Allerdings kann die Konvergenzordnung noch immer als optimal angesehen werden, da sie der Ordnung des Interpolations- bzw. Projektionsfehlers entspricht, der sich mit stückweise konstanten Ansatzfunktionen $u_h \in U_h^{(0)}$ erreichen lässt.¹⁹⁴

Wie bereits erwähnt, haben wir in dieser Arbeit lediglich a priori Fehlerabschätzung betrachtet. Als Anmerkung sei jedoch hinzugefügt, dass es auch möglich wäre mit a posteriori Fehlerabschätzungen zu arbeiten, die den Fehler mit Hilfe einer berechneten diskreten Lösung quantifizieren würden.¹⁹⁵ Diese Abschätzungen könnten dann für eine adaptive Gitterverfeinerung

¹⁹³Siehe auch [Her12, Bemerkung 16.21] und [Mey11, Bemerkungen 14.17 und 14.18].

¹⁹⁴Siehe Lemma 14.4, Abschätzung 14.22.

¹⁹⁵Im Falle einer Voll-Diskretisierung sei diesbezüglich etwa auf [RW10] verwiesen.

benutzt werden, die überall dort angewendet werden könnte, wo der Fehler eine bestimmte vorgegebene Toleranz überschreitet.

Literaturverzeichnis

- [A⁺07] Thomas Apel et al. Optimal control in non-convex domains: a priori discretization error estimates. *Calcolo*, 44:137–158, 2007.
- [Ada75] Robert A. Adams. *Sobolev Spaces*. Academic Press, 1975.
- [Alt02] Hans Wilhelm Alt. *Lineare Funktionalanalysis*. Springer-Verlag, 2002.
- [Alt06] Hans Wilhelm Alt. *Vorlesungsskript Analysis IV, SS2002*. Universität Bonn, 2006. http://www.iam.uni-bonn.de/~alt/ss2002/HTML/analysis4-hyp_0.html. Zuletzt geprüft am 27.03.2012.
- [Bas08] Peter Bastian. *Vorlesungsskript Finite-Elemente-Verfahren und schnelle Löser, SS2008*. Universität Stuttgart, 2008. <http://conan.iwr.uni-heidelberg.de/teaching/scripts/fesarticle.pdf>. Heruntergeladen am 12.12.2010.
- [BO04] Timothy Barth and Mario Ohlberger. *Finite Volume Methods: Foundation and Analysis*. Wiley, 2004.
- [Bra07] Dietrich Braess. *Finite Elemente*. Springer-Verlag, 2007.
- [C⁺00] Bernardo Cockburn et al. *Discontinuous Galerkin Methods: Theory, Computation and Applications*. Springer-Verlag, 2000.
- [Eva10] Lawrence C. Evans. *Partial Differential Equations*. American Mathematical Society, 2010.
- [G⁺07] Christian Grossmann et al. *Numerical treatment of partial differential equations*. Springer-Verlag, 2007.
- [Gri85] Pierre Grisvard. *Elliptic Problems in Nonsmooth Domains*. Pitman Publishing, 1985.
- [Gri92] Pierre Grisvard. *Singularities in Boundary Value Problems*. Masson, 1992.
- [Gro94] Christian Grossmann. *Numerik partieller Differentialgleichungen*. Teubner-Verlag, 1994.
- [H⁺09] Michael Hinze et al. *Optimization with PDE Constrains*. Springer-Verlag, 2009.

- [Her11] Roland Herzog. *Vorlesungsskript Einführung in Sobolevräume, SS2010*. Technische Universität Chemnitz, 2011. http://www.tu-chemnitz.de/mathematik/part_dgl/teaching/WS2011_Optimale_Steuerung_PDEs/Skript_Sobolevr%C3%A4ume.pdf. Heruntergeladen am 22.11.2011.
- [Her12] Roland Herzog. *Vorlesungsskript Optimale Steuerung partieller Differentialgleichungen, WS2011/12*. Technische Universität Chemnitz, 2012.
- [Hin05] Michael Hinze. A variational discretization concept in control constrained optimization: The linear-quadratic case. *Computational Optimization and Applications*, 30(1):45–63, 2005.
- [Jeg79] Hansgeorg Jeggler. *Nichtlineare Funktionalanalysis*. Teubner-Verlag, 1979.
- [Kön04] Konrad Königsberger. *Analysis 2*. Springer-Verlag, 2004.
- [Krö97] Dietmar Kröner. *Numerical Schemes for Conservation Laws*. Wiley/Teubner, 1997.
- [Lue69] David G. Luenberger. *Optimization By Vector Space Methods*. Wiley, 1969.
- [M⁺08] Sandra May et al. Error analysis for a finite element approximation of elliptic dirichlet boundary control problems. *Vorabdruck-Nr.: SPP1253-02-03, DFG Schwerpunktprogramm 1253*, 2008. <http://www.am.uni-erlangen.de/home/spp1253/wiki/images/4/4e/Preprint-spp1253-02-03.pdf>. Heruntergeladen am 03.04.2012.
- [Mes04] Dieter Meschede. *Gerthsen Physik*. Springer-Verlag, 2004.
- [Mey11] Christian Meyer. *Vorlesungsskript Optimale Steuerung partieller Differentialgleichungen, SS2011*. Technische Universität Dortmund, 2011. Heruntergeladen von <http://www.graduate-school-ce.de/index.php?id=367> am 13.08.2011.
- [Ohl08] Mario Ohlberger. *Vorlesungsskript Numerik partieller Differentialgleichungen, WS2008/09*. Westfälische Wilhelms-Universität Münster, 2008. http://wwwmath.uni-muenster.de/num/Vorlesungen/PDEI_WS08/Skript/skriptum_beta.pdf. Heruntergeladen am 22.11.2011.
- [Ruz04] Michael Ruzicka. *Nichtlineare Funktionalanalysis*. Springer-Verlag, 2004.
- [RW10] Arnd Rösch and Daniel Wachsmuth. A-posteriori error estimates for optimal control problems with state and control constraints. *Vorabdruck-Nr.: SPP1253-110, DFG Schwerpunktprogramm 1253*, 2010. <http://www.am.uni-erlangen.de/home/spp1253/wiki/images/b/bf/Preprint-SPP1253-110.pdf>. Heuntergeladen am 22.11.2011.
- [Trö09] Fredi Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen*. Vieweg + Teubner, 2009.

- [Wer11] Dirk Werner. *Funktionalanalysis*. Springer-Verlag, 2011.
- [Wlo82] Joseph Wloka. *Partielle Differentialgleichungen*. Teubner-Verlag, 1982.
- [Yos80] Kosaku Yosida. *Functional Analysis*. Springer-Verlag, 1980.
- [Zul08] Walter Zulehner. *Numerische Mathematik. Eine Einführung anhand von Differentialgleichungsproblemen. Band 1: Stationäre Probleme*. Birkhäuser-Verlag, 2008.

- Lagrange-Interpolation, 71
- Lebesgueraum, 8
- Lemma von
 - Céa, 59
 - Lax-Milgram, 25
- lineare Wärmeleitungsgleichung, 2
- Lipschitzgebiet, 11
- Lipschitzrand, 11
- Norm, 6
- Operator
 - adjungierter, 43, 51
 - Hilbertraum-adjungierter, 44
 - linearer, 18
 - selbstadjungierter, 44
 - Steuerungs-Zustands-, 36
 - diskreter, 75
- Operatornorm, 18
- optimale Temperaturquelle, 4
- Optimalitätsbedingung
 - hinreichende, 47, 52
 - notwendige, 46, 52
- Poissongleichung, 2
- Projektion, 48
 - $L^2(\Omega)$ -orthogonale, 87
- Projektionsformel, 55
- quadratisches Programm, 83
- quasiuniform, 64
- radial unbeschränkt, 34
- Raum
 - Finite-Elemente-, 66
 - normierter linearer, 6
 - reflexiver, 20
 - vollständiger, *siehe* Banachraum
- Riesz'scher Darstellungssatz, 20
- Riesz-Isomorphismus, 20
- Ritz-Galerkin-Verfahren, 58
- Satz von Friedrich, 25
- schwach
 - fogenabgeschlossen, 30
 - folgenkompakt, 30
 - folgenstetig, 30
 - unterhalbstetig, 32
- schwache
 - Formulierung, *siehe* Variationsformulierung
 - Lösung, 23
- Simplex, 60
- Skalarprodukt, 7
- Sobolew'scher Einbettungssatz, 21
- Sobolewraum, 15
- Spur, 17
- Spuroperator, 17
- Spursatz, 17
- Steuerung, 2
 - global optimale, 33, 37
 - verteilte, 4
- Teilraum
 - konvexer, 31
- Träger, 9
- Ungleichung
 - Cauchy-Schwarz'sche, 7
 - Hölder'sche, 8
 - von Poincaré-Friedrich, 25
 - Young'sche, 78
- Variationsformulierung, 23
- Variationsungleichung, 46, 78, 86
- Zerlegung
 - simpliziale, 60
 - zulässige, 64
- Zielfunktional, 3
 - reduziertes, 37
- Zustand, 3
 - adjungierter, 52
 - global optimaler, 37
- Zustandsgleichung, 3

Selbständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig angefertigt, nicht anderweitig zu Prüfungszwecken vorgelegt und keine anderen als die angegebenen Hilfsmittel verwendet habe.

Bayreuth, den 30. April 2012

Tobias Masching