TECHNISCHE UNIVERSITÄT CHEMNITZ

Fakultät für Mathematik

Professur Inverse Probleme

Diplomarbeit

Untersuchungen zu Konvergenzraten bei der Tikhonov-Regularisierung mit allgemeinen Abstandsfunktionalen

(korrigierte Fassung vom 15. Februar 2010)

Jens Geißler

geboren am 1. Oktober 1985 in Gera

Chemnitz, den 22. Juni 2009

verteidigt am 4. August 2009

Betreuer: Prof. Dr. Bernd Hofmann

Geißler, Jens
Untersuchungen zu Konvergenzraten bei der Tikhonov-Regularisierung
mit allgemeinen Abstandsfunktionalen
Diplomarbeit, Fakultät für Mathematik
Technische Universität Chemnitz, Juni 2009

Inhaltsverzeichnis

I. Analytische Untersuchungen

Der erste Teil dieser Arbeit soll sich einigen theoretischen Betrachtungen zur Konvergenz eines konkreten Lösungsverfahrens für schlecht-gestellte Operatorgleichungen, der Tikhonov-Regularisierung, widmen. Im ersten Kapitel werden wir kurz auf die Problematik inverser und schlecht-gestellter Probleme eingehen, da diese stets spezielle Lösungsverfahren erfordern. Im Anschluss führen wir in Kapitel 2 den Begriff des topologischen Raumes ein und legen damit die begriffliche Grundlage für die Betrachtungen in den folgenden Kapiteln. Die mathematische Symbolik wird in Kapitel 3 eingeführt. Zudem werden wir dort einige wesentliche Aussagen zu den Eigenschaften der Tikhonov-Regularisierung formulieren und beweisen. Kapitel 4 bildet den Hauptteil dieser Arbeit. Wir werden in diesem Kapitel ein neues Konzept zum Beweis von Aussagen über Konvergenzraten bei der Tikhonov-Regularisierung vorstellen, welches vollständig ohne Werkzeuge aus Hilbert- oder Banach-Räumen auskommt. In Kapitel 5 werden dann eingige aus der Literatur bekannte Resultate als Spezialfall des in dieser Arbeit beschriebenen Konzepts vorgestellt. Eine Sammlung von Anknüpfungspunkten für weitere Überlegungen werden wir in Kapitel 6 zusammenstellen.

1. Inverse und schlecht-gestellte Probleme

Von alters her besteht die Hauptarbeit der Naturwissenschaften darin, beobachtete Vorgänge durch (mathematische) Modelle zu beschreiben. Solche Modelle ermöglichen das theoretische Nachvollziehen von Zusammenhängen, aber auch die Vorhersage von zukünftigen Ereignissen anhand von Beobachtungen aus der Vergangenheit. Als wichtige Vertreter alter und neuer Anwendungsmöglichkeiten seien hier nur beispielhaft die Vorhersage der Mondbahn sowie von Sonnenfinsternissen und die computergestützte Simulation physikalischer Prozesse (z.B. Wärmeleitung) genannt. In neuerer Zeit werden mathematische Methoden und Modelle auch in anderen Wissenschaften verstärkt angewandt, sodass die im Folgenden aus naturwissenschaftlicher Sicht beschriebenen Fragestellungen zum Beispiel auch in den Wirtschaftswissenschaften (z.B. Optionspreismodelle) und im technischen Bereich (z.B. Bildverarbeitung) von Bedeutung sind.

Als Modell bezeichnen wir eine Vorschrift, die den Weg von einer Ursache zur entsprechenden Wirkung mathematisch beschreibt; d.h. jeder Ursache wird eine eindeutige Wirkung zugeordnet. Zusätzlich können in ein Modell verschiedene Parameter (z.B. Materialparameter) eingehen, die Einfluss auf diese Zuordnung haben. Die Parameter können gegebenenfalls auch zu den Ursachen gezählt werden. Als Beispiel betrachten wie ein einfaches Wärmeleitproblem: Gegeben sei ein quaderförmiger Körper, der an der Unterseite im Zeitintervall von $t = 0$ bis $t = T > 0$ erwärmt wird. Als Wirkung im Sinne des obigen Modellbegriffs sehen wir die Wärmeverteilung im Inneren des Körpers zum Endzeitpunkt $t = T$ an. Die Wärmeleitung im Körper kann mit Hilfe von partiellen Differentialgleichungen modelliert werden. Als Ursache sind hier zwei Dinge auszumachen: zum einen die Wärmezufuhr an der Unterseite sowie der Wärmeabfluss über die restlichen Flächen an die kältere Umgebung (Randbedingungen), wobei diese Randbedingungen innerhalb des vorgegebenen Zeitintervalls nicht konstant bleiben müssen, und zum anderen die Wärmeverteilung zum Zeitpunkt $t = 0$ im Inneren des Körpers (Anfangsbedingung). Ein in das Modell eingehender Parameter wird zum Beispiel der Wärmeleitkoeffizient des Körpers sein; dieser kann je nach Modell auch von Punkt zu Punkt variieren.

Mit dem Fortschritt der technischen Entwicklung wurde zunehmend die Frage interessant, wie man anhand mathematischer Modelle nicht nur eine Wirkung theoretisch vorhersagen, sondern auch Aussagen über andere Teile des Modells, also über die Ursachen, über Teile der Ursachen oder über die einfließenden Parameter gewinnen kann. Beim Wärmeleitmodell ist zum Beispiel der Wärmeleitkoeffizient, insbesondere wenn er nicht räumlich konstant ist, durch direkte Messung kaum zu bestimmen. Es bietet sich also an, Rand- und Anfangsbedingungen vorzugegeben und die Wärmeverteilung zum Zeitpunkt $t = T$, zumindest am Rand des Körpers, experimentell zu ermitteln, um aus diesen Daten mit Hilfe des Modells den Wärmeleitkoeffizienten zu bestimmen. Während man die Berechnung der Wirkung aus Ursache und Parametern als *direktes Problem* bezeichnet, spricht man bei entgegengesetzten Fragestellung, d.h. bei der Ermittlung der Ursache, von Teilen der Ursache oder von Modellparametern aus einer beobachteten Wirkung und weiteren bekannten Größen des Modells, von einem *inver-*

sen Problem. Wie das Beispiel zur Wärmeleitung zeigt, ist bei einem inversen Problem oft nicht nur die Wirkung bekannt, sondern es stehen auch Teile der Ursache oder der Modellparameter zur Rekonstruktion der fehlenden Informationen zur Verfügung.

Zur Vereinfachung der Darstellung reduzieren wir im Folgenden das direkte Problem darauf, aus einer gegebenen Größe (Ursache) eine dadurch eindeutig bestimmte zweite Größe (Wirkung) zu berechnen. Als inverses Problem betrachten wir die Ermittlung einer Ursache, die eine vorgegebene Wirkung nach sich zieht. Wir vernachlässigen also die Unterteilung in Ursachen und Parameter und betrachten nur noch die Größen des Modells, die bei der Lösung des direkten bzw. des inversen Problems gesucht sind. Alle weiteren Größen sehen wir als konstant an.

Abhängigkeiten zwischen zwei Größen bringen immer die Frage mit sich, wie empfindlich diese Abhängigkeiten sind: Führt eine geringfügige Änderung der einen Größe auch nur zu einer geringfügigen Änderung der anderen? Oder ist es möglich, dass sich die als abhängig betrachtete Größe beliebig stark ändern kann, obwohl die Änderung der vorgegebenen Größe nur sehr klein war? Im ersten Fall heißt die Abhängigkeit *stetig*, im zweiten *unstetig*. Bei der Untersuchung solcher Stetigkeitsfragen stellt sich meist heraus, dass das direkte Problem eine Glättungseigenschaft besitzt, d.h. sehr verschiedene Ursachen können zu ähnlichen oder sogar identischen Wirkungen führen. Dies ist der Grund für die Unstetigkeit vieler inverser Probleme.

Die einem inversen Problem meist innewohnende Unstetigkeit stellt die größte Schwierigkeit bei der mathematischen Behandlung solcher Probleme dar, da beliebig kleine Fehler bei der Beobachtung der Wirkung nahezu beliebig starke Fehler im Ergebnis hervorrufen können. Hinzu kommen noch die Fragen nach Existenz und Eindeutigkeit einer Lösung. Um ein inverses Problem mit geeigneten mathematischen Methoden lösen zu können, müssen somit zunächst die drei folgenden Fragen beantwortet werden:

- Gibt es im Rahmen des betrachteten Modells überhaupt eine Ursache, die zur beobachteten Wirkung führt?
- Gibt es *höchstens eine* Ursache, die zur beobachteten Wirkung führt?
- Hängt die ermittelte Ursache stetig von der beobachteten Wirkung ab?

Können alle drei Fragen, insbesondere die letzte, positiv beantwortet werden, so heißt das Problem *gut-gestellt* (korrekt) nach Hadamard. Muss mindestens eine dieser Fragen verneint werden, so ist das Problem *schlecht-gestellt* (inkorrekt) nach Hadamard.

Ziel der mathematischen Theorie der inversen und schlecht-gestellten Probleme ist die Entwicklung und Untersuchung so genannter Regularisierungsverfahren, mit deren Hilfe auch schlecht-gestellte Probleme stabil, d.h. ohne übermäßige Verstärkung von Datenfehlern, gelöst werden können. Erreicht wird dies durch den Übergang von exakten zu näherungsweisen Lösungen. Im weiteren Verlauf des ersten Teils dieser Arbeit werden wir ein solches Verfahren vorstellen und dessen Eigenschaften untersuchen.

Als allgemeine Einführung in das mathematische Gebiet der inversen und schlecht-gestellten Probleme sowie für einen Überblick über gängige Regularisierungsverfahren sei das Standardwerk [9] empfohlen.

2. Topologische Räume

Die meisten Aussagen im ersten Teil dieser Arbeit werden wir für recht allgemeine topologische Räume formulieren. Somit ist es notwendig, die oft nur in normierten oder metrischen Räumen betrachteten Grundbegriffe der mathematischen Analysis, wie Konvergenz, Stetigkeit und Kompaktheit, auch für topologische Räume einzuführen. Wir werden den Umfang der Ausführungen auf das Nötigste beschränken und auf sämtliche Beweise verzichten. Eine kurze Einführung in das Gebiet der topologischen Räume bietet [37, Anhang B.2]. Für umfassende Abhandlungen empfehlen wir [28] und [2].

2.1. Grundbegriffe

Im Folgenden bezeichnen wir mit $\mathcal{P}(W)$ die Potenzmenge, d.h. das System aller Teilmengen, einer nichtleeren Menge W.

Definition 2.1. Sei W eine beliebige nichtleere Menge und sei $\tau \subseteq \mathcal{P}(W)$ ein nichtleeres System von Teilmengen von W. Das Paar (W, τ) heißt *topologischer Raum* und τ heißt *Topologie* auf W, wenn gilt:

(i) $\emptyset \in \tau$ und $W \in \tau$,

(ii) aus $G_i \in \tau$ für $i \in I$ (beliebige Indexmenge) folgt $\bigcup_{i \in I} G_i \in \tau$,

(iii) aus $G_1, G_2 \in \tau$ folgt $G_1 \cap G_2 \in \tau$.

Definition 2.2. Sei (W, τ) ein topologischer Raum und sei $G \subseteq W$. Die Menge G heißt *offen*, wenn $G \in \tau$ gilt, und *abgeschlossen*, wenn $W \setminus G \in \tau$ gilt. Die Vereinigung aller in G enthaltenen offenen Mengen heißt *Inneres* von G und wird mit $\mathring{G}$ bezeichnet. Der Durchschnitt aller G umfassenden abgeschlossenen Mengen heißt *Abschluss* von G und wird mit $\overline{G}$ bezeichnet.

Definition 2.3. Sei W eine beliebige nichtleere Menge und seien τ_1 und τ_2 zwei Topologien auf W. Die Topologie τ_1 heißt *schwächer* oder *gröber* als τ_2, wenn $\tau_1 \subseteq \tau_2$ gilt. Die Topologie τ_2 heißt dann entsprechend *stärker* oder *feiner* als τ_1.

In einem topologischen Raum steht kein Mittel zur Verfügung, um den Abstand zwischen zwei Elementen zu messen (im Gegensatz zum metrischen Raum). Dennoch kann man im weitesten Sinne über Entfernungsbeziehungen sprechen. Dazu führen wir den Umgebungsbegriff ein.

Definition 2.4. Sei (W, τ) ein topologischer Raum und sei $w \in W$.

(i) Eine Menge $H \subseteq W$ heißt *Umgebung* von w, wenn eine offene Menge $G \in \tau$ mit $w \in G \subseteq H$ existiert.

(ii) Das System $\mathcal{U}(w) \subseteq \mathcal{P}(W)$ aller Umgebungen des Punktes w heißt *Umgebungssystem* von w.

(iii) Ein Teilsystem $\mathcal{B}(w) \subseteq \mathcal{U}(w)$ des Umgebungssystems von w heißt *Umgebungsbasis* von w, wenn zu jeder Menge $H \in \mathcal{U}(w)$ eine Menge $G \in \mathcal{B}(w)$ mit $G \subseteq H$ existiert.

Weiter unten werden wir den Begriff der kompakten Menge einführen. Dazu benötigen wir die folgende Definition.

Definition 2.5. Sei (W, τ) ein topologischer Raum und sei $\tilde{W} \subseteq W$. Dann heißt das System

$$\tilde{\tau} := \{G \cap \tilde{W} : G \in \tau\} \subseteq \mathcal{P}(\tilde{W})$$

Spurtopologie von W auf $\tilde{W}$ und $(\tilde{W}, \tilde{\tau})$ wird als *topologischer Teilraum* von (W, τ) bezeichnet.

Beispiel. Sei (W, d) ein metrischer Raum. Dann bildet die Menge aller bezüglich der Metrik d offenen Mengen eine Topologie auf W. Somit ist jeder metrische Raum auch ein topologischer Raum. Für $w \in W$ ist

$$\mathcal{B}(w) := \{B_{1/n}(w) : n \in \mathbb{N}\}$$

mit $B_r(w_0) := \{w \in W : d(w, w_0) < r\}$ eine (abzählbare) Umgebungsbasis des Punktes w. Ein Teileraum $\tilde{W} \subseteq W$ im Sinne des metrischen Raumes stimmt mit dem entsprechenden topologischen Teilraum überein, d.h. die Menge aller in $(\tilde{W}, d|_{\tilde{W}\times\tilde{W}})$ offenen Mengen stimmt gerade mit der Spurtopologie von W auf $\tilde{W}$ überein.

2.2. Konvergenz

Für die Arbeit im metrischen Raum ist der Begriff der Folgenkonvergenz vollkommen ausreichend. Im topologischen Raum zeigt sich jedoch, dass das Konvergenzkonzept allgemeiner formuliert werden muss.

Definition 2.6. Eine nichtleere Indexmenge I heißt *gerichtet*, wenn einen Relation „$\preceq$“ auf I existiert, sodass gilt:

(i) $i \preceq i$ für alle $i \in I$ (Reflexivität),

(ii) aus $i_1 \preceq i_2$ und $i_2 \preceq i_3$ folgt $i_1 \preceq i_3$ für alle $i_1, i_2, i_3 \in I$ (Transitivität),

(iii) zu beliebigen $i_1, i_2 \in I$ existiert ein $i_3 \in I$ mit $i_1 \preceq i_3$ und $i_2 \preceq i_3$.

Definition 2.7. Sei W eine beliebige nichtleere Menge und sei I eine gerichtete Indexmenge. Jede Abbildung $\phi : I \to W$ heißt *Netz* oder *Moore-Smith-Folge* in W. Statt ϕ schreiben wir $(w_i)_{i\in I}$ mit $w_i := \phi(i)$.

Definition 2.8. Sei (W, τ) ein topologischer Raum. Ein Netz $(w_i)_{i\in I}$ in W heißt *konvergent* gegen $w \in W$ (Schreibweise: $w_i \to w$), wenn zu jeder Umgebung $G \in \mathcal{U}(w)$ von w ein $i_0 \in I$ mit $w_i \in G$ für alle $i \succeq i_0$ existiert.

Beispiel. Die Menge $\mathbb{N}$ der natürlichen Zahlen ist mit der Relation „$\leq$“ eine gerichtete Menge. Somit ist jede Folge $(w_k)_{k\in\mathbb{N}}$ in einer Menge W ein Netz in W. Eine Folge $(w_k)_{k\in\mathbb{N}}$ in einem topologischen Raum (W,τ) heißt nach Definition 2.8 konvergent gegen $w \in W$, wenn zu jeder Umgebung $G \in \mathcal{U}(w)$ von w ein $k_0 \in \mathbb{N}$ mit $w_k \in G$ für alle $k \geq k_0$ existiert. Die aus metrischen Räumen bekannte Folgenkonvergenz ist also ein Spezialfall der Netzkonvergenz.

Satz 2.9. *Seien (W,τ) ein topologischer Raum, $w \in W$ beliebig und $\mathcal{B}(w)$ eine Umgebungsbasis von w. Dann konvergiert ein Netz $(w_i)_{i\in I}$ genau dann gegen w, wenn zu jeder Menge $G \in \mathcal{B}(w)$ ein $i_0 \in I$ mit $w_i \in G$ für alle $i \succeq i_0$ existiert.*

Konvergente Netze (auch konvergente Folgen) können mehrere Grenzelemente besitzen. Ein extremes Beispiel dafür stellt die so genannte *triviale* oder *chaotische Topologie* dar: Sei W eine beliebige nichtleere Menge und sei $\tau := \{\emptyset, W\}$. Dann ist τ eine Topologie auf W und für jedes Element $w \in W$ gilt $\mathcal{U}(w) = \{W\}$. Somit konvergiert jedes Netz in W gegen jedes Element in W. Einen Ausweg liefert die folgende Definition.

Definition 2.10. Ein topologischer Raum (W,τ) heißt *Hausdorff-Raum*, wenn für beliebige Elemente $w_1, w_2 \in W$ offene Mengen $G_1, G_2 \in \tau$ mit $w_1 \in G_1$, $w_2 \in G_2$ und $G_1 \cap G_2 = \emptyset$ existieren.

Bemerkung. Offensichtlich ist jeder metrische Raum ein Hausdorff-Raum.

Satz 2.11. *Ein topologischer Raum ist genau dann ein Hausdorff-Raum, wenn in ihm jedes konvergente Netz genau ein Grenzelement besitzt.*

In metrischen Räumen ist bekannt, dass alle wesentlichen Begriffe, wie zum Beispiel Abgeschlossenheit, Stetigkeit und Kompaktheit, mit Hilfe von Folgen charakterisiert werden können. Im topologischen Raum übernehmen Netze diese Charakterisierung.

Satz 2.12. *Eine Menge $G \subseteq W$ im topologischen Raum (W,τ) ist genau dann abgeschlossen, wenn die Grenzelemente jedes konvergenten Netzes in G ebenfalls in G liegen.*

Definition 2.13. Sei (W,τ) ein topologischer Raum. Eine Menge $G \subseteq W$ heißt *folgenabgeschlossen*, wenn die Grenzelemente jeder konvergenten Folge in G ebenfalls in G liegen.

Bemerkung. Offensichtlich folgt aus der Abgeschlossenheit die Folgenabgeschlossenheit. Die Umkehrung gilt im Allgemeinen nicht.

Definition 2.14. Ein topologischer Raum (W,τ) heißt *A_1-Raum*, wenn er das *erste Abzählbarkeitsaxiom* erfüllt, d.h. wenn zu jedem Element $w \in W$ eine abzählbare Umgebungsbasis existiert.

Bemerkung. Jeder metrische Raum ist ein A_1-Raum.

Satz 2.15. *In einen A_1-Raum ist eine Menge genau dann abgeschlossen, wenn sie folgenabgeschlossen ist.*

2.3. Stetigkeit

Seien (U, τ_U) und (V, τ_V) topologische Räume.

Definition 2.16. Eine Abbildung $f : U \to V$ heißt *stetig im Punkt* $u \in U$, wenn zu jeder Umgebung $H \in \mathcal{U}(f(u))$ eine Umgebung $G \in \mathcal{U}(u)$ mit $f(G) \subseteq H$ existiert. Eine Abbildung $f : U \to V$ heißt *stetig*, wenn sie in jedem Punkt von U stetig ist.

Satz 2.17. *Eine Abbildung $f : U \to V$ ist genau dann stetig im Punkt $u \in U$, wenn für jedes Netz $(u_i)_{i \in I}$ in U mit $u_i \to u \in U$ auch $f(u_i) \to f(u)$ gilt.*

Definition 2.18. Eine Abbildung $f : U \to V$ heißt *folgenstetig im Punkt* $u \in U$, wenn für jede Folge $(u_k)_{k \in \mathbb{N}}$ mit $u_k \to u \in U$ auch $f(u_k) \to f(u)$ gilt. Eine Abbildung $f : U \to V$ heißt *folgenstetig*, wenn sie in jedem Punkt von U folgenstetig ist.

Bemerkung. Offensichtlich folgt aus der Stetigkeit die Folgenstetigkeit.

Satz 2.19. *Sei U ein A_1-Raum. Dann ist eine Abbildung $f : U \to V$ genau dann stetig, wenn sie folgenstetig ist.*

2.4. Kompaktheit

Die Übertragung des Kompaktheitsbegriffes von metrischen auf topologische Räume bringt eine Reihe neuer Effekte mit sich, sodass mehrere verschiedene, im metrischen Raum zusammenfallende Kompaktheitsbegriffe eingeführt werden können. Wir wollen uns hier auf Kompaktheit und Folgenkompaktheit beschränken.

Im Folgenden verstehen wir unter einer (endlichen) offenen Überdeckung einer Menge $G \subseteq W$ im topologischen Raum (W, τ) stets eine Familie von (endlich vielen) offenen Mengen, deren Vereinigung die Menge G umfasst.

Definition 2.20. Sei (W, τ) ein topologischer Raum.

(i) (W, τ) heißt *kompakt*, wenn jede offene Überdeckung von W eine endliche offene Überdeckung von W enthält.

(ii) (W, τ) heißt *folgenkompakt*, wenn jede Folge in W eine konvergente Teilfolge besitzt.

(iii) Eine Menge $G \subseteq W$ heißt *kompakt* bzw. *folgenkompakt*, wenn G als Teilraum von (W, τ) kompakt bzw. folgenkompakt ist.

Satz 2.21. *Sei (W, τ) ein A_1-Raum. Dann ist jede kompakte Menge $G \subseteq W$ folgenkompakt.*

Um die Kompaktheit einer Menge in einem topologischen Raum mit Hilfe von Netzen zu charakterisieren, benötigt man den Begriff des Teilnetzes. Dieser ist keine Verallgemeinerung des Teilfolgenbegriffs und soll hier nicht erläutert werden.

Definition 2.22. Eine Menge $G \subseteq W$ im topologischen Raum (W, τ) heißt *relativ kompakt* bzw. *relativ folgenkompakt*, wenn ihr Abschluss $\overline{G}$ kompakt bzw. folgenkompakt ist.

3. Tikhonov-Regularisierung

In diesem Kapitel stellen wir ein sehr flexibles Regularisierungsverfahren, die Tikhonov-Regularisierung, vor. Wir wählen von Anfang an einen recht allgemeinen Ansatz und werden erst später auf wichtige Spezialfälle eingehen.

3.1. Das Tikhonov-Funktional

Inverse Probleme werden üblicherweise als Operatorgleichung formuliert, d.h. das direkte Problem wird durch eine Abbildung zwischen zwei Räumen beschrieben und das inverse Problem besteht darin, die entsprechende Operatorgleichung für ein vorgegebenes Element aus dem Bildraum zu lösen.

Wir betrachten einen linearen topologischen Raum U über dem Skalarkörper $\mathbb{R}$ mit Topologie τ_U, einen topologischen Raum V mit Topologie τ_V sowie einen im Allgemeinen nichtlinearen Operator $F : D(F) \subseteq U \to V$ mit Definitionsbereich $D(F)$. Die Operationen des linearen Raumes U, d.h Addition und Multiplikation mit Skalaren, müssen nicht stetig bezüglich τ_U sein. Die Räume U und V seien Hausdorff-Räume, d.h. das Grenzelement einer konvergenten Folge sei eindeutig bestimmt. Als Symbol für die Konvergenz im Sinne der Topologie verwenden wir „$\to$“.

Ziel ist es, für ein $v^0 \in V$ die Gleichung

$$F(u) = v^0 \tag{3.1}$$

zu lösen. Im Weiteren bezeichnen wir mit v^0 stets die exakte rechte Seite der Gleichung (3.1), die aufgrund von Beobachtungsfehlern jedoch nur näherungsweise in Form eines Elementes $v^\delta \in V$ bekannt ist. Das Datenfehlerniveau $\delta > 0$ beschreibt die Genauigkeit dieser Näherung. Um dies zu präzisieren, benötigen wir einen Abstandsbegriff im Raum V. Dazu geben wir uns ein nichtnegatives Funktional $\mathcal{S} : V \times V \to [0, \infty]$ vor, das jedem Paar von Elementen aus V einen Abstand zuordnet, der nicht notwendigerweise die Eigenschaften eines metrischen Abstandes besitzen muss. Die Näherung v^δ von v^0 soll nun

$$\mathcal{S}(v^0, v^\delta) \leq \delta \quad \text{und} \quad \mathcal{S}(v^\delta, v^0) \leq \delta \tag{3.2}$$

erfüllen.

Da wir Gleichung (3.1) aufgrund der fehlenden Kenntnis von v^0 nicht exakt lösen können, suchen wir eine Näherungslösung als Lösung des Minimierungsproblems

$$\mathcal{S}(F(u), v^\delta)^p \to \min_{u \in D(F)} \tag{3.3}$$

für ein $p > 0$. Die Minimierung der p-ten Potenz von $\mathcal{S}$ kann aus praktischer bzw. numerischer Sicht günstiger sein als die Minimierung des reinen Abstandsfunktionals $\mathcal{S}$. Allgemeiner kann man auch $f\big(\mathcal{S}(F(u), v^\delta)\big)$ mit einer Funktion $f : [0, \infty] \to [0, \infty]$ über $D(F)$ minimieren; einige Betrachtungen dazu findet man in [31].

Oft ist aufgrund der Eigenschaften des Operators F nicht einmal die Existenz von

Lösungen des Minimierungsproblems (3.3) gesichert; selbst wenn Minimierer existieren, hängen diese in der Regel nicht stetig von v^δ ab. Um dennoch eine stabile Lösungsmethode zu erhalten, ändern wir das Minimierungsproblem (3.3) ab. Dazu führen wir ein weiteres Funktional $\mathcal{R} : U \to [0, \infty]$ ein, welches die „Qualität“ eines Elementes $u \in U$ messen soll; je kleiner der Wert von $\mathcal{R}$ ist, desto „besser“ ist das Element. Im Weiteren sei $D(\mathcal{R}) := \{u \in U : \mathcal{R}(u) < \infty\}$ und $D := D(F) \cap D(\mathcal{R})$.

Für ein $\alpha > 0$ betrachten wir nun Näherungslösungen zu (3.1) in Form von Lösungen des Minimierungsproblems

$$\mathcal{T}_\alpha^\delta(u) := \mathcal{S}(F(u), v^\delta)^p + \alpha\mathcal{R}(u) \to \min_{u \in D}, \tag{3.4}$$

wobei der Parameter α zur Gewichtung zwischen beiden Funktionalen dient. Wenn α klein gewählt wird, dann kommen wir den gegebenen Daten v^δ sehr nahe, aber das Minimierungsproblem ist nur wenig stabil. Wird α hingegen groß gewählt, dann ist das Problem zwar stabil lösbar, jedoch ist der Abstand der Lösung zu den Daten groß. Der Parameter α muss also einen Kompromiss zwischen Datenapproximation und Stabilität herstellen.

Das Funktional $\mathcal{T}_\alpha^\delta$ wird als Tikhonov-Funktional bezeichnet. Das beschriebene Vorgehen, d.h. die Minimierung eines solchen Funktionals, zur Lösung einer Operatorgleichung nennt man Tikhonov-Regularisierung, wobei wir hier eine sehr allgemeine Variante dieses Regularisierungsverfahrens behandeln. Diese allgemeine Variante wird in der Literatur auch als Variationsregularisierung bezeichnet; Untersuchungen dazu beschränkten sich jedoch bisher auf Potenzen von Normen als Abstandsfunktional $\mathcal{S}$. Anwendungen mit konkreten Abstandsfunktionalen $\mathcal{S}$ findet man in [25].

An dieser Stelle sollte erwähnt werden, dass solch allgemeine Ansätze bereits vor mehr als 25 Jahren in Erwägung gezogen wurden (vgl. [29, S. 53]), jedoch erst in jüngster Zeit, motiviert durch neue Resultate für Tikhonov-Regularisierung in Banach-Räumen (z.B. [19] und [15]), genauer untersucht werden.

3.2. Existenz, Stabilität, Konvergenz

Nachdem die Methode der Tikhonov-Regularisierung formal eingeführt wurde, bleibt zu zeigen, dass auf diesem Weg Operatorgleichungen stabil gelöst werden können. Um Aussagen über das Lösungsverhalten von (3.1) bzw. (3.4) formulieren und beweisen zu können, müssen wir einige Annahmen treffen. Zunächst geben wir zur Vereinfachung der Notation eine Definition an. Angemerkt sei, dass wir bei der Konvergenz von Folgen $(w_k)_{k\in\mathbb{N}}$ gegen ein Element w nur „$w_k \to w$“ anstelle der exakten Formulierung „$w_k \to w$ für $k \to \infty$“ schreiben, sofern klar ist, bezüglich welchem Index der Grenzübergang vollzogen wird.

Definition 3.1. Wir sagen, dass eine Folge $(v_k)_{k\in\mathbb{N}}$ in V gegen ein Element $v \in V$ *im Sinne von* $\mathcal{S}$ *konvergiert*, und schreiben $v_k \xrightarrow{\mathcal{S}} v$, wenn

$$\mathcal{S}(v, v_k) \to 0 \quad \text{und} \quad \mathcal{S}(v_k, v) \to 0$$

gelten.

Voraussetzung 3.2.

(i) $F : D(F) \subseteq U \to V$ ist folgenstetig (vgl. Definition 2.18).

(ii) $D(F)$ ist folgenabgeschlossen (vgl. Definition 2.13).

(iii) Es existiert ein $u \in D$, sodass $F(u) = v^0$ gilt. Insbesondere ist also $D \neq \emptyset$.

(iv) Für $\mathcal{S} : V \times V \to [0, \infty]$ gilt (mit Folgen $(v_k)_{k\in\mathbb{N}}$ und $(\tilde{v}_k)_{k\in\mathbb{N}}$ in V sowie $v, \tilde{v} \in V$):

(a) $\mathcal{S}(v, \tilde{v}) = 0$ gilt genau dann, wenn $v = \tilde{v}$.

(b) Es existiert ein $s \geq 1$, sodass

$$\mathcal{S}(v_1, v_2) \leq s\mathcal{S}(v_1, v_3) + s\mathcal{S}(v_3, v_1)$$

für alle $v_1, v_2, v_3 \in V$ gilt.

(c) $\mathcal{S}$ ist unterhalb folgenstetig, d.h. aus $v_k \to v$ und $\tilde{v}_k \to \tilde{v}$ folgt

$$\mathcal{S}(v, \tilde{v}) \leq \liminf_{k\to\infty} \mathcal{S}(v_k, \tilde{v}_k).$$

(d) Aus $\mathcal{S}(v_k, v) \to 0$ folgt $v_k \to v$.

(e) Gelten $v_k \xrightarrow{\mathcal{S}} v$ und $\mathcal{S}(\tilde{v}, v) < \infty$, so folgt $\mathcal{S}(\tilde{v}, v_k) \to \mathcal{S}(\tilde{v}, v)$.

(v) $\mathcal{R} : U \to [0, \infty]$ ist unterhalb folgenstetig, d.h. aus $u_k \to u$ folgt

$$\mathcal{R}(u) \leq \liminf_{k\to\infty} \mathcal{R}(u_k).$$

(vi) Für jedes $\alpha > 0$ und jedes $c > 0$ sind die Niveaumengen

$$M_\alpha(c) := \{u \in D : \mathcal{T}_\alpha^0(u) \leq c\}$$

relativ folgenkompakt (vgl. Definition 2.22), d.h. jede Folge $(u_k)_{k\in\mathbb{N}}$ in $M_\alpha(c)$ besitzt eine in U konvergente Teilfolge.

Bemerkung. Einige Bemerkungen zu Voraussetzung 3.2:

- Für $s = 1$ folgt (iv)(e) aus (iv)(b), denn es gilt

$$-\underbrace{\mathcal{S}(v_k, v)}_{\to 0} \leq \mathcal{S}(\tilde{v}, v_k) - \mathcal{S}(\tilde{v}, v) \leq \underbrace{\mathcal{S}(v, v_k)}_{\to 0}.$$

- Aus den Punkten (iv)(a) und (iv)(b) folgt, dass das Grenzelement einer im Sinne von $\mathcal{S}$ konvergenten Folge $(v_k)_{k\in\mathbb{N}}$ eindeutig bestimmt ist. Denn hätte diese Folge

zwei verschiedene Grenzelemente $v \in V$ und $\tilde{v} \in V$, so würde

$$\mathcal{S}(v, \tilde{v}) \leq s\mathcal{S}(v, v_k) + s\mathcal{S}(v_k, \tilde{v}) \to 0$$

und somit $v = \tilde{v}$ folgen.

- Gilt $D(F) = U$, so ist $D(F)$ trivialerweise folgenabgeschlossen, d.h Punkt (ii) ist automatisch erfüllt.

Um die Formulierungen im weiteren Verlauf des ersten Teils dieser Arbeit nicht unnötig aufzublähen, werden wir statt folgenabgeschlossen nur abgeschlossen, statt folgenstetig nur stetig usw. schreiben, sofern dadurch keine Unklarheiten zu erwarten sind. Falls U und V metrische Räume (oder allgemeiner A_1-Räume) sind, so fallen diese Begriffe ohnehin zusammen.

Im Zusammenhang mit Punkt (iv)(b) von Voraussetzung 3.2 werden wir in diesem und auch in späteren Kapiteln des Öfteren die Ungleichung

$$(a + b)^p \leq c_p(a^p + b^p)$$

für reelle Zahlen $a \geq 0$, $b \geq 0$ und $p > 0$ mit

$$c_p := \begin{cases} 1 & \text{für } p \in (0, 1), \\ 2^{p-1} & \text{für } p \geq 1 \end{cases}$$

verwenden. Einen Beweis dieser Ungleichung findet man in [29].

Alle Sätze in diesem Abschnitt sind aus [25] übernommen. Die Beweise sollen hier nochmals angegeben werden, wobei wir kleinere Modifikationen vornehmen, da wir im Gegensatz zu [25] die Annahme (iv)(b) aus Voraussetzung 3.2 treffen und Punkt (vi) schwächer als in [25] formuliert haben.

Satz 3.3 (Existenz)**.** *Es sei Voraussetzung 3.2 erfüllt. Dann besitzt das Minimierungsproblem (3.4) eine Lösung.*

Beweis. Im Folgenden beziehen sich römische Zahlen stets auf Voraussetzung 3.2.

Setzen wir $c := \inf\{\mathcal{T}_\alpha^\delta(u) : u \in D\}$, so existiert wegen (iii) eine Folge $(u_k)_{k\in\mathbb{N}}$ in D mit $\mathcal{T}_\alpha^\delta(u_k) \to c$; insbesondere gilt $c < \infty$. Für hinreichend großes k gilt nun

$$\begin{aligned} \mathcal{T}^0_{c_p s^p \alpha}(u_k) &\leq c_p s^p \big(\mathcal{S}(F(u_k), v^\delta)^p + \mathcal{S}(v^\delta, v^0)^p + \alpha\mathcal{R}(u_k)\big) \\ &\leq c_p s^p \big(\delta^p + \mathcal{T}_\alpha^\delta(u_k)\big) \leq c_p s^p(\delta^p + c + 1) \end{aligned}$$

und damit $u_k \in M_{c_p s^p \alpha}(c_p s^p(\delta^p + c + 1))$. Nach (vi) besitzt (u_k) also eine in U konvergente Teilfolge, die wir wieder mit $(u_k)_{k\in\mathbb{N}}$ bezeichnen; das Grenzelement sei $\tilde{u} \in U$.

Wegen (ii) gilt $\tilde{u} \in D$ und aus (i) folgt $F(u_k) \to F(\tilde{u})$. Damit erhalten wir aus (iv)(c)

und (v)

$$\begin{aligned}\mathcal{T}_\alpha^\delta(\tilde{u}) &\le \liminf_{k\to\infty} \mathcal{S}(F(u_k), v^\delta)^p + \alpha \liminf_{k\to\infty} \mathcal{R}(u_k) \\ &\le \liminf_{k\to\infty} \big(S(F(u_k), v^\delta)^p + \alpha\mathcal{R}(u_k)\big) = \lim_{k\to\infty} \mathcal{T}_\alpha^\delta(u_k) = c,\end{aligned}$$

d.h. $\tilde{u}$ minimiert $\mathcal{T}_\alpha^\delta$ über D. □

Satz 3.4 (Stabilität)**.** *Es sei Voraussetzung 3.2 erfüllt und es sei $(v_k)_{k\in\mathbb{N}}$ eine Folge in V mit $v_k \xrightarrow{\mathcal{S}} v^\delta$. Dann besitzt jede Folge $(u_k)_{k\in\mathbb{N}}$ in U mit $u_k \in \operatorname{argmin}\{\mathcal{S}(F(u), v_k)^p + \alpha\mathcal{R}(u) : u \in D\}$ eine konvergente Teilfolge und das Grenzelement jeder konvergenten Teilfolge minimiert $\mathcal{T}_\alpha^\delta$ über D. Außerdem gilt für jede konvergente Teilfolge $(u_{k_l})_{l\in\mathbb{N}}$ von (u_k) mit Grenzelement $\tilde{u}$ die Beziehung $\mathcal{R}(u_{k_l}) \to \mathcal{R}(\tilde{u})$.*

Beweis. Im Folgenden beziehen sich römische Zahlen stets auf Voraussetzung 3.2.

Sei $u_k \in \operatorname{argmin}\{\mathcal{S}(F(u), v_k)^p + \alpha\mathcal{R}(u) : u \in D\}$ für $k \in \mathbb{N}$, sei $\bar{u} \in D$ so gewählt, dass $F(\bar{u}) = v^0$ gilt und sei $c > 0$ beliebig. Dann folgt aus (iv)(b) für hinreichend großes k

$$\begin{aligned}\mathcal{T}^0_{c_p s^p \alpha}(u_k) &\le c_p s^p \big(\mathcal{S}(F(u_k), v_k)^p + \alpha\mathcal{R}(u_k) + \mathcal{S}(v_k, v^0)^p\big) \\ &\le c_p s^p \big(\mathcal{S}(F(\bar{u}), v_k)^p + \alpha\mathcal{R}(\bar{u}) + \mathcal{S}(v_k, v^0)^p\big) \\ &\le c_p s^p \big(2c_p s^p \delta^p + c_p s^p \mathcal{S}(v^\delta, v_k)^p + c_p s^p \mathcal{S}(v_k, v^\delta)^p + \alpha\mathcal{R}(\bar{u})\big) \\ &\le 2c_p^2 s^{2p}(\delta^p + c) + c_p s^p \alpha\mathcal{R}(\bar{u})\end{aligned}$$

und damit $u_k \in M_{c_p s^p \alpha}\big(2c_p^2 s^{2p}(\delta^p + c) + c_p s^p \alpha\mathcal{R}(\bar{u})\big)$. Nach (vi) besitz (u_k) also eine in U konvergente Teilfolge.

Sei nun $(u_{k_l})_{l\in\mathbb{N}}$ eine beliebige konvergente Teilfolge von (u_k) mit Grenzelement $\tilde{u} \in U$ sowie $(v_{k_l})_{l\in\mathbb{N}}$ die entsprechende Teilfolge von (v_k). Wegen (ii) gilt $\tilde{u} \in D$ und aus (i) folgt $F(u_{k_l}) \to F(\tilde{u})$. Außerdem gilt $v_l \to v^\delta$ nach (iv)(d). In Verbindung mit (iv)(c), (v) und (iv)(e) folgt für $u \in D$ mit $\mathcal{S}(F(u), v^\delta) < \infty$ somit

$$\begin{aligned}\mathcal{T}_\alpha^\delta(\tilde{u}) &\le \liminf_{l\to\infty} \mathcal{S}(F(u_{k_l}), v_{k_l})^p + \alpha \liminf_{l\to\infty} \mathcal{R}(u_{k_l}) \le \liminf_{l\to\infty} \big(\mathcal{S}(F(u_{k_l}), v_{k_l})^p + \alpha\mathcal{R}(u_{k_l})\big) \\ &\le \limsup_{l\to\infty} \big(\mathcal{S}(F(u_{k_l}), v_{k_l})^p + \alpha\mathcal{R}(u_{k_l})\big) \le \limsup_{l\to\infty} \big(\mathcal{S}(F(u), v_{k_l})^p + \alpha\mathcal{R}(u)\big) \\ &= \lim_{l\to\infty} \mathcal{S}(F(u), v_{k_l})^p + \alpha\mathcal{R}(u) = \mathcal{S}(F(u), v^\delta)^p + \alpha\mathcal{R}(u),\end{aligned}$$

d.h. $\tilde{u}$ minimiert $\mathcal{T}_\alpha^\delta$ über D. Insbesondere erhalten wir mit $u := \tilde{u}$ die Beziehung

$$\mathcal{T}_\alpha^\delta(\tilde{u}) = \lim_{l\to\infty} \big(\mathcal{S}(F(u_{k_l}), v_{k_l})^p + \alpha\mathcal{R}(u_{k_l})\big). \tag{3.5}$$

Wir nehmen nun an, dass $\mathcal{R}(u_{k_l})$ nicht gegen $\mathcal{R}(\tilde{u})$ konvergiert. Dann gilt

$$\liminf_{l\to\infty} \mathcal{R}(u_{k_l}) \ne \mathcal{R}(\tilde{u}) \quad \text{oder} \quad \limsup_{l\to\infty} \mathcal{R}(u_{k_l}) \ne \mathcal{R}(\tilde{u})$$

und wegen

$$\mathcal{R}(\tilde{u}) \le \liminf_{l\to\infty} \mathcal{R}(u_{k_l}) \le \limsup_{l\to\infty} \mathcal{R}(u_{k_l})$$

folgt somit

$$\tilde{c} := \limsup_{l\to\infty} \mathcal{R}(u_{k_l}) > \mathcal{R}(\tilde{u}).$$

Nach Definition von $\tilde{c}$ besitzt (u_{k_l}) eine Teilfolge $(u_{l_m})_{m\in\mathbb{N}}$, die $\mathcal{R}(u_{l_m}) \to \tilde{c}$ erfüllt, und somit folgt aus (3.5)

$$\begin{aligned}\lim_{m\to\infty} \mathcal{S}(F(u_{l_m}), v_{l_m})^p &= \lim_{m\to\infty} \big(\mathcal{S}(F(u_{l_m}), v_{l_m})^p + \alpha\mathcal{R}(u_{l_m})\big) - \alpha \lim_{m\to\infty} \mathcal{R}(u_{l_m}) \\ &= \mathcal{S}(F(\tilde{u}), v^\delta)^p + \alpha\mathcal{R}(\tilde{u}) - \alpha\tilde{c} < \mathcal{S}(F(\tilde{u}), v^\delta)^p.\end{aligned}$$

Dies steht im Widerspruch zu (iv)(c). □

Zur Formulierung von Konvergenzaussagen ist es zweckmäßig, den Lösungsbegriff für Gleichung (3.1) enger zu fassen.

Definition 3.5. Ein Element $u^\dagger \in D$ heißt *$\mathcal{R}$-minimierende Lösung*, wenn $F(u^\dagger) = v^0$ und $\mathcal{R}(u^\dagger) = \inf\{\mathcal{R}(u) : u \in D, F(u) = v^0\}$ gelten.

Satz 3.6. *Es sei Voraussetzung 3.2 erfüllt. Dann existiert eine $\mathcal{R}$-minimierende Lösung.*

Beweis. Im Folgenden beziehen sich römische Zahlen stets auf Voraussetzung 3.2.

Setzen wir $c := \inf\{\mathcal{R}(u) : u \in D, F(u) = v^0\}$, so existiert wegen (iii) eine Folge $(u_k)_{k\in\mathbb{N}}$ in D mit $F(u_k) = v^0$ und $\mathcal{R}(u_k) \to c$; insbesondere gilt $c < \infty$. Für hinreichend großes k gilt nun

$$\mathcal{T}_1^0(u_k) = \mathcal{R}(u_k) \le c + 1$$

und damit $u_k \in M_1(c+1)$. Nach (vi) besitzt (u_k) also eine in U konvergente Teilfolge, die wir wieder mit $(u_k)_{k\in\mathbb{N}}$ bezeichnen; das Grenzelement sei $\tilde{u} \in U$. Wegen (ii) gilt $\tilde{u} \in D$, aus (i) folgt $F(\tilde{u}) = v^0$ und (v) liefert $\mathcal{R}(\tilde{u}) \le \liminf_{k\to\infty} \mathcal{R}(u_k) = c$. Somit ist $\tilde{u}$ eine $\mathcal{R}$-minimierende Lösung. □

Satz 3.7 (Konvergenz)**.** *Es sei Voraussetzung 3.2 erfüllt und es sei $(\delta_k)_{k\in\mathbb{N}}$ eine monoton fallend gegen Null konvergierende Folge in $\mathbb{R}$. Weiterhin seien $\alpha : (0, \delta_1] \to (0, \infty)$ eine Parameterwahl mit $\alpha(\delta) \to 0$ und $\frac{\delta^p}{\alpha(\delta)} \to 0$ für $\delta \to 0$ sowie $\alpha_k := \alpha(\delta_k)$, $v_k := v^{\delta_k}$ und $(u_k)_{k\in\mathbb{N}}$ eine Folge in U mit $u_k \in \operatorname{argmin}\{\mathcal{S}(F(u), v_k)^p + \alpha\mathcal{R}(u) : u \in D\}$. Dann besitzt (u_k) eine konvergente Teilfolge und jede konvergente Teilfolge von (u_k) konvergiert gegen eine $\mathcal{R}$-minimierende Lösung. Ist die $\mathcal{R}$-minimierende Lösung eindeutig, so konvergiert (u_k) gegen diese $\mathcal{R}$-minimierende Lösung.*

Beweis. Im Folgenden beziehen sich römische Zahlen stets auf Voraussetzung 3.2.

Sei $u_k \in \operatorname{argmin}\{\mathcal{S}(F(u), v_k)^p + \alpha\mathcal{R}(u) : u \in D\}$ für $k \in \mathbb{N}$, sei $\alpha_{\max} := \max\{\alpha_k : k \in \mathbb{N}\}$ und sei $c > 0$ beliebig. Dann gilt für hinreichend großes k

$$\begin{aligned}
\mathcal{T}^0_{c_p s^p \alpha_{\max}}(u_k) &\le c_p s^p \big(\mathcal{S}(F(u_k), v_k)^p + \mathcal{S}(v_k, v^0)^p + \alpha_k \mathcal{R}(u_k) + (\alpha_{\max} - \alpha_k)\mathcal{R}(u_k)\big) \\
&\le c_p s^p \big(\delta_k^p + \mathcal{S}(F(u^\dagger), v_k)^p + \alpha_k \mathcal{R}(u^\dagger) + (\alpha_{\max} - \alpha_k)\mathcal{R}(u_k)\big) \\
&\le c_p s^p \Big(2\delta_k^p + \alpha_k \mathcal{R}(u^\dagger) + \tfrac{\alpha_{\max} - \alpha_k}{\alpha_k}\big(\mathcal{S}(F(u_k), v_k)^p + \alpha_k \mathcal{R}(u_k)\big)\Big) \\
&\le c_p s^p \Big(2\delta_k^p + \alpha_k \mathcal{R}(u^\dagger) + \tfrac{\alpha_{\max} - \alpha_k}{\alpha_k}\big(\delta_k^p + \alpha_k \mathcal{R}(u^\dagger)\big)\Big) \\
&= c_p s^p \Big(\delta_k^p + \alpha_{\max}\big(\tfrac{\delta_k^p}{\alpha_k} + \mathcal{R}(u^\dagger)\big)\Big) \le c_p s^p \Big(c + \alpha_{\max}\big(c + \mathcal{R}(u^\dagger)\big)\Big)
\end{aligned}$$

und damit $u_k \in M_{c_p s^p \alpha_{\max}}\big(c_p s^p \big(c + \alpha_{\max}(c + \mathcal{R}(u^\dagger))\big)\big)$. Nach (vi) besitzt (u_k) also eine in U konvergente Teilfolge.

Sei nun $(u_{k_l})_{l\in\mathbb{N}}$ eine beliebige konvergente Teilfolge von (u_k) mit Grenzelement $\tilde{u} \in U$. Wegen (ii) gilt $\tilde{u} \in D$ und aus (i) folgt $F(u_{k_l}) \to F(\tilde{u})$, sodass wir

$$\begin{aligned}
\mathcal{S}(F(\tilde{u}), v^0)^p &\le \liminf_{l\to\infty} \mathcal{S}(F(u_{k_l}), v^0)^p \\
&\le \liminf_{l\to\infty} c_p s^p \big(\mathcal{S}(F(u_{k_l}), v_{k_l})^p + \mathcal{S}(v_{k_l}, v^0)^p + \alpha_{k_l}\mathcal{R}(u_{k_l})\big) \\
&\le \liminf_{l\to\infty} c_p s^p \big(2\delta_{k_l}^p + \alpha_{k_l}\mathcal{R}(u^\dagger)\big) = 0
\end{aligned}$$

erhalten, d.h. $F(\tilde{u}) = v^0$. Aus

$$\begin{aligned}
\mathcal{R}(\tilde{u}) &\le \liminf_{l\to\infty} \mathcal{R}(u_{k_l}) \le \limsup_{l\to\infty} \mathcal{R}(u_{k_l}) \le \limsup_{l\to\infty} \tfrac{1}{\alpha_{k_l}}\big(\mathcal{S}(F(u_{k_l}), v_{k_l})^p + \alpha_{k_l}\mathcal{R}(u_{k_l})\big) \\
&\le \limsup_{l\to\infty} \Big(\tfrac{\delta_{k_l}^p}{\alpha_{k_l}} + \mathcal{R}(u^\dagger)\Big) = \mathcal{R}(u^\dagger)
\end{aligned}$$

folgt somit, dass $\tilde{u}$ eine $\mathcal{R}$-minimierende Lösung ist.

Ist die $\mathcal{R}$-minimierende Lösung $u^\dagger$ eindeutig, so konvergiert nach dem eben Gezeigten jede konvergente Teilfolge von (u_k) gegen diese $\mathcal{R}$-minimierende Lösung. Angenommen (u_k) würde jedoch nicht gegen $u^\dagger$ konvergieren, dann gäbe es eine Teilfolge $(u_{k_l})_{l\in\mathbb{N}}$ von (u_k) und eine offene Menge $G \in \tau_U$, sodass $u_{k_l} \notin G$ für alle $l \in \mathbb{N}$ gilt. Die Folge (u_{k_l}) erfüllt aber ausgehend von der dazugehörigen Teilfolge $(\delta_{k_l})_{l\in\mathbb{N}}$ von (δ_k) die Voraussetzungen des Satzes, sodass (u_{k_l}) nach dem bereits Bewiesenen eine gegen $u^\dagger$ konvergierende Teilfolge besitzt. Dies ist ein Widerspruch zu $u_{k_l} \notin G$ für $l \in \mathbb{N}$. □

3.3. Zwei wichtige Spezialfälle

Die bisherigen Ausführungen zur Tikhonov-Regularisierung waren sehr allgemein gehalten. Hier wollen wir nun zwei wichtige Spezialfälle betrachten, die in der Literatur bereits früher als die allgemeine Situation von oben untersucht wurde (siehe z.B. [10] und [19]).

3.3.1. Spezialfall Banach-Raum

Seien U und V Banach-Räume. Als Topologien τ_U auf U und τ_V auf V wählen wir die entsprechenden schwachen Topologien. Dabei verstehen wir unter der schwachen Topologie die Topologie auf U bzw. V, die der schwachen Konvergenz im Banach-Raum U bzw. V entspricht. Um die schwache Topologie auf U bzw. V in Form einer Familie von Mengen zu beschreiben, benötigt man den Begriff der Initialtopologie bezüglich einer Familie von Abbildungen zwischen topologischen Räumen. Darauf wollen wir hier jedoch nicht näher eingehen.

In diesem Abschnitt bezeichnen wir die Konvergenz im Sinne der Topologie, d.h. die schwache Konvergenz, mit „$\rightharpoonup$“, um Verwechslungen mit der Normkonvergenz zu vermeiden. Für die Normkonvergenz selbst benötigen wir hier kein Symbol.

Das Abstandsfunktional $\mathcal{S}$ sei durch

$$\mathcal{S}(v_1, v_2) := \|v_1 - v_2\|$$

definiert und es sei $p > 0$, d.h. wir betrachten das Tikhonov-Funktional

$$\mathcal{T}_\alpha^\delta(u) = \|F(u) - v^\delta\|^p + \alpha\mathcal{R}(u).$$

Des Weiteren treffen wir die folgenden Voraussetzungen.

Voraussetzung 3.8.

(i) $F : D(F) \subseteq U \to V$ ist schwach stetig.

(ii) $D(F)$ ist schwach abgeschlossen.

(iii) Es existiert ein $u \in D$, sodass $F(u) = v^0$ gilt. Insbesondere ist also $D \neq \emptyset$.

(iv) $\mathcal{R} : U \to [0, \infty]$ ist unterhalb schwach stetig, d.h. aus $u_k \rightharpoonup u$ folgt

$$\mathcal{R}(u) \leq \liminf_{k\to\infty} \mathcal{R}(u_k).$$

(v) Für jedes $\alpha > 0$ und jedes $c > 0$ sind die Niveaumengen

$$M_\alpha(c) := \{u \in D : \mathcal{T}_\alpha^0(u) \leq c\}$$

relativ schwach kompakt, d.h. jede Folge $(u_k)_{k\in\mathbb{N}}$ in $M_\alpha(c)$ besitzt eine in U schwach konvergente Teilfolge.

Zur Vereinfachung des Beweises zum nächsten Satz formulieren wir ein Lemma. Eine Hilbert-Raum-Version davon findet man in [36, Satz 2.28].

Lemma 3.9. *Die Norm in einem Banach-Raum W ist unterhalb schwach stetig, d.h. für jede Folge $(w_k)_{k\in\mathbb{N}}$ in W mit $w_k \rightharpoonup w \in W$ gilt*

$$\|w\| \leq \liminf_{k\to\infty} \|w_k\|.$$

Beweis. Aus der Stetigkeit der Norm folgt $\|w_k\| \to \|w\|$. Wir wählen eine Teilfolge $(w_{k_l})_{l\in\mathbb{N}}$ von (w_k), die

$$\lim_{l\to\infty} \|w_{k_l}\| = \liminf_{k\to\infty} \|w_k\|$$

erfüllt, und erhalten für $\zeta \in W'$ (Dualraum bezüglich Normtopologie) mit $\|\zeta\| \leq 1$ aus der schwachen Konvergenz der Folge (w_{k_l}) die Beziehung

$$|\zeta(w)| = \lim_{l\to\infty} |\zeta(w_{k_l})| \leq \lim_{l\to\infty} \|\zeta\|\|w_{k_l}\| = \liminf_{k\to\infty} \|\zeta\|\|w_k\| \leq \liminf_{k\to\infty} \|w_k\|.$$

Für $w \in W$ gilt nach [37, Korollar III.1.7] nun

$$\|w\| = \sup\{|\zeta(w)| : \zeta \in W', \|\zeta\| \leq 1\},$$

sodass $\|w\| \leq \liminf_{k\to\infty} \|w_k\|$ folgt. □

Satz 3.10. *Mit der speziellen Wahl von U, V, τ_U, τ_V und $\mathcal{S}$ sowie unter Voraussetzung 3.8 ist Voraussetzung 3.2 erfüllt.*

Beweis. Die Punkte (i), (ii) und (iii) von Voraussetzung 3.2 folgen sofort aus den entsprechenden Punkten in Voraussetzung 3.8. Die Punkte (iv)(a) und (iv)(b) sind Normaxiome und sind somit trivialerweise erfüllt. In Lemma 3.9 wurde (iv)(c) gezeigt. Punkt (iv)(d) folgt aus

$$|\eta(v_k) - \eta(v)| = |\eta(v_k - v)| \leq \|\eta\|\|v_k - v\| \to 0 \quad \text{für alle } \eta \in V'$$

(V' sei der Dualraum bezüglich der Normtopologie). Für (iv)(e) genügt es, $\|\tilde{v} - v_k\| \to \|\tilde{v} - v\|$ zu zeigen; dies folgt analog zur Bemerkung im Anschluss an Voraussetzung 3.2 aus der Norm-Dreiecksungleichung. Die Punkte (v) und (vi) von Voraussetzung 3.2 entsprechen gerade den Punkten (iv) und (v) von Voraussetzung 3.8. □

Satz 3.11. *Sei U reflexiv und sei $\mathcal{R}$ durch*

$$\mathcal{R}(u) := \|u - u^*\|^q$$

mit einem Element $u^* \in U$ und $q \geq 1$ gegeben. Dann sind die Punkte (iv) und (v) von Voraussetzung 3.8 erfüllt.

Beweis. Punkt (iv) wurde in Lemma 3.9 gezeigt, wobei zu beachten ist, dass die Addition in Banach-Räumen schwach stetig ist und die Hintereinanderausführung einer schwach stetigen Abbildung und eines unterhalb schwach stetigen Funktionals wieder ein unterhalb schwach stetiges Funktional liefert. Zum Nachweis von Punkt (v) sei $u \in M_\alpha(c)$. Dann gilt

$$\|u\| \leq \|u - u^*\| + \|u^*\| \leq \left(\tfrac{1}{\alpha}\mathcal{T}_\alpha^0(u)\right)^{\frac{1}{q}} + \|u^*\| \leq \left(\tfrac{c}{\alpha}\right)^{\frac{1}{q}} + \|u^*\|;$$

also ist $M_\alpha(c)$ eine beschränkte Menge. Aus [37, Satz VIII.3.18] folgt, dass beschränkte Mengen im reflexiven Banach-Raum relativ schwach kompakt sind. □

3.3.2. Spezialfall Hilbert-Raum

Seien U und V Hilbert-Räume und seien τ_U und τ_V die entsprechenden schwachen Topologien. Das Abstandsfunktional $\mathcal{S}$ sei durch

$$\mathcal{S}(v_1, v_2) := \|v_1 - v_2\|$$

und das Regularisierungsfunktional durch

$$\mathcal{R}(u) := \|u - u^*\|^2$$

mit einem vorgegebenen Element $u^* \in U$ definiert. Es sei $p = 2$. Wir betrachten somit das Tikhonov-Funktional

$$\mathcal{T}_\alpha^\delta(u) = \|F(u) - v^\delta\|^2 + \alpha\|u - u^*\|^2.$$

Des Weiteren treffen wir die folgenden Voraussetzungen.

Voraussetzung 3.12.

(i) $F : D(F) \subseteq U \to V$ ist schwach stetig.

(ii) $D(F)$ ist schwach abgeschlossen.

(iii) Es existiert ein $u \in D(F)$, sodass $F(u) = v^0$ gilt.

Satz 3.13. *Mit der speziellen Wahl von U, V, τ_U, τ_V, $\mathcal{S}$ und $\mathcal{R}$ sowie unter Voraussetzung 3.12 ist Voraussetzung 3.2 erfüllt.*

Beweis. Die Behauptung folgt sofort mit $q = 2$ aus Satz 3.11. □

Proposition 3.14. *Sei $A : U \to V$ ein beschränkter, d.h. stetiger, linearer Operator. Dann sind die Punkte (i) und (ii) von Voraussetzung 3.12 erfüllt.*

Beweis. Punkt (ii) ist wegen $D(A) = U$ trivialerweise erfüllt. Zum Beweis von Punkt (i) sei $(u_k)_{k\in\mathbb{N}}$ eine Folge in U mit $u_k \rightharpoonup u \in U$. Dann gilt

$$\langle Au_k, v\rangle = \langle u_k, A^*v\rangle \to \langle u, A^*v\rangle = \langle Au, v\rangle$$

für alle $v \in V$, wobei $A^* : V \to U$ der Hilbert-Raum-Adjungierte zu A sei. Somit ist $Au_k \rightharpoonup Au$ und damit die schwache Stetigkeit von A gezeigt (man beachte dazu auch Proposition 5.1). □

4. Konvergenzraten

Im vorangegangenen Kapitel wurde gezeigt, dass die Methode der Tikhonov-Regularisierung zur Lösung von schlecht-gestellten Operatorgleichungen geeignet ist. Diese Feststellung beruhte darauf, dass sich die regularisierte Lösung stabil bestimmen lässt (vgl. Satz 3.4) und, einen hinreichend kleinen Datenfehler vorausgesetzt, die $\mathcal{R}$-minimierenden Lösungen der Operatorgleichung beliebig genau im Sinne der Topologie τ_U approximiert werden können (vgl. Satz 3.7). Über die erzielbare Konvergenzgeschwindigkeit der regularisierten Lösungen gegen eine $\mathcal{R}$-minimierende Lösung bei Verringerung des Datenfehlerniveaus δ macht Satz 3.7 jedoch keinerlei Aussagen. Diese Frage kann auch nicht direkt beantwortet werden, sondern erfordert ein sehr umsichtiges Vorgehen.

Aus praktischer Sicht bedeutet eine geringe Konvergenzgeschwindigkeit, dass eine Reduzierung des Datenfehlers, zum Beispiel durch die Anschaffung genauerer Messgeräte, kaum zu besseren Ergebnissen führen wird. Ist die Konvergenzgeschwindigkeit hingegen groß, so besteht Hoffnung auf eine Erhöhung der Ergebnisqualität.

4.1. Vorbetrachtungen

4.1.1. Bregman-Abstände

Will man über Konvergenzgeschwindigkeit sprechen, so benötigt man ein auf Zahlen beruhendes Maß für die Konvergenz. In Satz 3.7 hatten wir die Topologie τ_U als Konvergenzmaß herangezogen; diese ist nun nicht mehr geeignet. Deshalb führen wir ein spezielles Abstandsfunktional auf U ein. Dazu benötigen wir den Begriff des Subgradienten.

Definition 4.1. Sei $\mathcal{F} : U \to \mathbb{R} \cup \{+\infty\}$ ein konvexes Funktional auf U. Ein Element $\xi \in U'$ (Dualraum bezüglich τ_U) heißt *Subgradient* im Punkt $\tilde{u} \in U$, falls

$$\mathcal{F}(\tilde{u}) + \xi(u - \tilde{u}) \leq \mathcal{F}(u) \quad \text{für alle } u \in U$$

gilt. Die Menge $\partial\mathcal{F}(\tilde{u}) \subseteq U'$ aller Subgradienten im Punkt $\tilde{u}$ heißt *Subdifferential* im Punkt $\tilde{u}$.

Bemerkung. Als Dualraum W' eines linearen und topologischen Raumes (W, τ_W) bezeichnen wir die Menge aller stetigen linearen Funktionale auf W.

Voraussetzung 4.2.

(i) $\mathcal{R}$ ist konvex.

(ii) Es existiert eine $\mathcal{R}$-minimierende Lösung $u^\dagger \in D_B := \{u \in D : \partial\mathcal{R}(u) \neq \emptyset\}$.

Im Folgenden nehmen wir stets an, dass die Voraussetzungen 3.2 und 4.2 erfüllt sind.

Definition 4.3. Für $\tilde{u} \in D_B$ und $\xi \in \partial\mathcal{R}(\tilde{u})$ heißt das durch

$$\mathcal{B}_\xi^{\mathcal{R}}(u, \tilde{u}) := \mathcal{R}(u) - \mathcal{R}(\tilde{u}) - \xi(u - \tilde{u})$$

gegebene Funktional $\mathcal{B}_\xi^{\mathcal{R}}(\bullet, \tilde{u}) : U \to [0, \infty]$ *Bregman-Abstand* bezüglich $\mathcal{R}$, $\tilde{u}$ und ξ.

Das folgende Beispiel zeigt, dass der Bregman-Abstand eine Verallgemeinerung der Norm im Hilbert-Raum darstellt.

Beispiel. Sei U ein Hilbert-Raum, $u^* \in U$ und

$$\mathcal{R}(u) := \|u - u^*\|^2.$$

Für jedes $\tilde{u} \in U$ gilt dann $\partial\mathcal{R}(\tilde{u}) = \{\langle \bullet, 2(\tilde{u} - u^*)\rangle\}$ (man beachte dazu auch Proposition 5.1), sodass wir $\xi := \langle \bullet, 2(\tilde{u} - u^*)\rangle$ setzen. Wir erhalten für $u \in U$

$$\begin{aligned}\mathcal{B}_\xi^{\mathcal{R}}(u, \tilde{u}) &= \|u - u^*\|^2 - \|\tilde{u} - u^*\|^2 - \langle u - \tilde{u}, 2(\tilde{u} - u^*)\rangle \\ &= \|u\|^2 - 2\langle u - \tilde{u}, u^*\rangle - \|\tilde{u}\|^2 - 2\langle u - \tilde{u}, \tilde{u}\rangle + 2\langle u - \tilde{u}, u^*\rangle \\ &= \|u\|^2 - 2\langle u, \tilde{u}\rangle + \|\tilde{u}\|^2 = \|u - \tilde{u}\|^2.\end{aligned}$$

Zur bildlichen Vorstellung des Bregman-Abstandes $\mathcal{B}_\xi^{\mathcal{R}}(\bullet, \tilde{u})$ kann man diesen für $U = \mathbb{R}$ bzw. $U = \mathbb{R}^2$ als Differenz zwischen $\mathcal{R}$ und der Tangente bzw. der Tangentialebene an $\mathcal{R}$ in $\tilde{u}$ ansehen (vgl. Abbildung 4.1).

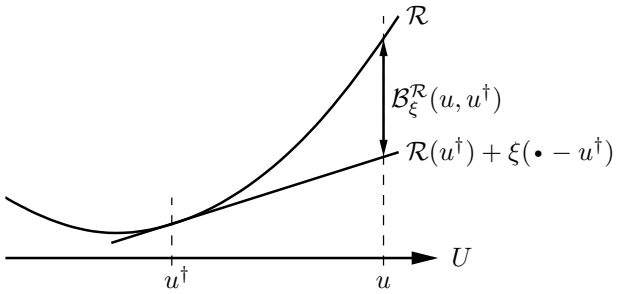

Abbildung 4.1.: Veranschaulichung des Bregman-Abstandes für $U = \mathbb{R}$.

4.1.2. Betrachtungen zur Eindeutigkeit der Lösung

Da wir nun über ein zahlenmäßiges Konvergenzmaß verfügen, können wir den nächsten Schritt in Richtung Konvergenzgeschwindigkeit gehen. Interessiert sind wir an der Konvergenz $\mathcal{B}_\xi^{\mathcal{R}}(u_{\alpha(\delta)}^\delta, u^\dagger) \to 0$ für $\delta \to 0$, wobei $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung ist und $\xi \in \partial\mathcal{R}(u^\dagger)$ gelten soll. Mit $u_\alpha^\delta \in D$ bezeichnen wir einen Minimierer von $\mathcal{T}_\alpha^\delta$ und $\delta \mapsto \alpha(\delta)$ sei eine Parameterwahl. Wir formulieren eine notwendige Bedingung für diese Konvergenz.

Proposition 4.4. *Seien $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung, $\xi \in \partial\mathcal{R}(u^\dagger)$ und $(u_k)_{k\in\mathbb{N}}$ eine Folge in D mit $\mathcal{B}_\xi^{\mathcal{R}}(u_k, u^\dagger) \to 0$. Jedes Element $\tilde{u} \in D$, zu dem eine Teilfolge $(u_{k_l})_{l\in\mathbb{N}}$ von (u_k) mit $u_{k_l} \to \tilde{u}$ existiert, muss dann $\mathcal{B}_\xi^{\mathcal{R}}(\tilde{u}, u^\dagger) = 0$ erfüllen.*

Beweis. Für $\varepsilon > 0$ und hinreichend großes $l \in \mathbb{N}$ folgt mit der Unterhalbstetigkeit von $\mathcal{R}$ die Abschätzung

$$\begin{aligned}\mathcal{B}_\xi^{\mathcal{R}}(\tilde{u}, u^\dagger) &= \mathcal{R}(\tilde{u}) - \mathcal{R}(u^\dagger) - \xi(\tilde{u} - u^\dagger) = \mathcal{R}(\tilde{u}) - \mathcal{R}(u^\dagger) - \xi(\tilde{u} - u_{k_l}) - \xi(u_{k_l} - u^\dagger) \\ &= \underbrace{\mathcal{B}_\xi^{\mathcal{R}}(u_{k_l}, u^\dagger)}_{\to 0} - \mathcal{R}(u_{k_l}) + \underbrace{\mathcal{R}(\tilde{u})}_{\leq \mathcal{R}(u_{k_l}) + \varepsilon} - \underbrace{\xi(\tilde{u} - u_{k_l})}_{\to 0} \leq 3\varepsilon\end{aligned}$$

und somit, da $\varepsilon > 0$ beliebig war, $\mathcal{B}_\xi^{\mathcal{R}}(\tilde{u}, u^\dagger) = 0$. □

Proposition 4.5. *Sei $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung, $\xi \in \partial\mathcal{R}(u^\dagger)$ und $\tilde{u} \in D$ mit $\mathcal{B}_\xi^{\mathcal{R}}(\tilde{u}, u^\dagger) = 0$. Dann gilt $\xi \in \partial\mathcal{R}(\tilde{u})$ und $\mathcal{B}_\xi^{\mathcal{R}}(u^\dagger, \tilde{u}) = 0$.*

Beweis. Für $u \in U$ gilt

$$\begin{aligned}\mathcal{R}(\tilde{u}) + \xi(u - \tilde{u}) &= \mathcal{R}(\tilde{u}) + \xi(u - u^\dagger) - \xi(\tilde{u} - u^\dagger) \\ &= \mathcal{B}_\xi^{\mathcal{R}}(\tilde{u}, u^\dagger) + \mathcal{R}(u^\dagger) + \xi(u - u^\dagger) \leq \mathcal{R}(u)\end{aligned}$$

und

$$\mathcal{B}_\xi^{\mathcal{R}}(u^\dagger, \tilde{u}) = -\mathcal{B}_\xi^{\mathcal{R}}(\tilde{u}, u^\dagger) = 0.$$

□

Im Kontext von Satz 3.7 besagt Proposition 4.4, dass es zur Erfüllung der Bedingung $\mathcal{B}_\xi^{\mathcal{R}}(u_k, u^\dagger) \to 0$ notwendig ist, die Menge der $\mathcal{R}$-minimierenden Lösungen, die Grenzelement einer konvergenten Teilfolge von (u_k) sein können, einzuschränken. Und zwar so, dass alle betroffenen $\mathcal{R}$-minimierenden Lösungen zueinander den Bregman-Abstand Null haben. Proposition 4.5 bringt etwas mehr Klarheit über den verschwindenden Bregman-Abstand: Alle fraglichen $\mathcal{R}$-minimierenden Lösungen besitzen einen gemeinsamen Subgradienten, der zur Definition des Bregman-Abstandes verwendet werden kann.

4.1.3. Quellbedingungen und Variationsungleichungen

Um Aussagen über die Konvergenzgeschwindigkeit der Minimierer $u^\delta_{\alpha(\delta)}$ von $\mathcal{T}^\delta_{\alpha(\delta)}$ treffen zu können, werden wir weitere Einschränkungen hinnehmen müssen, denn im Allgemeinen kann die Konvergenz $\mathcal{B}_\xi^{\mathcal{R}}(u^\delta_{\alpha(\delta)}, u^\dagger) \to 0$ bei $\delta \to 0$ für „ungünstige“ v^0 bzw. $u^\dagger$ beliebig langsam sein (siehe [9, Proposition 3.11] für eine entsprechende Aussage zu linearen Operatoren in Hilbert-Räumen). Es sind also gewisse Eigenschaften an die $\mathcal{R}$-minimierenden Lösungen und auch an den Operator F vorauszusetzen. Eigenschaften der Lösung können als so genannte Quellbedingung formuliert werden, d.h. die $\mathcal{R}$-minimierenden Lösungen sollen im Bild eines gewissen von F abhängigen Operators liegen. Strukturbedingungen an den Operator F kann mit von Elementen aus U oder aus speziellen Teilmengen von U zu erfüllenden Ungleichungen formulieren. Eine Kombination von Lösungs- und Operatoreigenschaften in einer einzelnen Bedingung stellt die in [19] eingeführte Variationsungleichung dar.

4.2. Eine Variationsungleichung

In diesem Abschnitt werden wir das erste Resultat zur Konvergenzgeschwindigkeit formulieren; den Beweis dazu verschieben wir auf den nächsten Abschnitt dieses Kapitels, um Wiederholungen zu vermeiden. Im Folgenden nehmen wir stets an, dass die Voraussetzungen 3.2 und 4.2 erfüllt sind.

4.2.1. Die Niveaumengen des Tikhonov-Funktionals

Eine wesentliche Rolle bei der Formulierung der im Rahmen dieser Arbeit untersuchten Variationsungleichungen spielen die Niveaumengen des Tikhonov-Funktionals.

Proposition 4.6. *Für $\alpha > 0$ und $c > 0$ sind die Mengen $M_\alpha(c)$ folgenabgeschlossen.*

Beweis. Aus der Stetigkeit von F und der Unterhalbstetigkeit von $\mathcal{S}$ folgt die Unterhalbstetigkeit von $\mathcal{S}(F(\bullet), v^0)^p$ und zusammen mit der Unterhalbstetigkeit von $\mathcal{R}$ die Unterhalbstetigkeit von $\mathcal{T}_\alpha^0$. Sei nun $(u_k)_{k\in\mathbb{N}}$ eine Folge in $M_\alpha(c)$, die gegen ein $\tilde{u} \in U$ konvergiert. Dann gilt

$$\mathcal{T}_\alpha^0(\tilde{u}) \leq \liminf_{k\to\infty} \mathcal{T}_\alpha^0(u_k) \leq c,$$

d.h. $\tilde{u} \in M_\alpha(c)$. Also ist $M_\alpha(c)$ folgenabgeschlossen. □

Proposition 4.7. *Sei $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung und sei $\varrho > 0$ beliebig. Für $0 < \alpha_1 \leq \alpha_2$ gilt dann*

(i) $M_{\alpha_1}(\varrho) \supseteq M_{\alpha_2}(\varrho)$,

(ii) $M_{\alpha_1}(\varrho\alpha_1) \subseteq M_{\alpha_2}(\varrho\alpha_2)$,

(iii) $\bigcap_{\alpha>0} M_\alpha(\varrho\alpha) = \{u \in D : F(u) = v^0, \mathcal{R}(u) \leq \varrho\}$.

Beweis.

(i) Sei $u \in M_{\alpha_2}(\varrho)$. Dann gilt

$$\mathcal{T}_{\alpha_1}^0(u) = \mathcal{S}(F(u), v^0)^p + \alpha_1\mathcal{R}(u) \leq \mathcal{S}(F(u), v^0)^p + \alpha_2\mathcal{R}(u) = \mathcal{T}_{\alpha_2}^0(u) \leq \varrho$$

und somit $u \in M_{\alpha_1}(\varrho)$.

(ii) Sei $u \in M_{\alpha_1}(\varrho\alpha_1)$. Dann gilt

$$\begin{aligned}\mathcal{T}_{\alpha_2}^0(u) &= \mathcal{S}(F(u), v^0)^p + \alpha_1\mathcal{R}(u) - (\alpha_1 - \alpha_2)\mathcal{R}(u) \leq \varrho\alpha_1 - (\alpha_1 - \alpha_2)\mathcal{R}(u) \\ &= \varrho\alpha_2 + \underbrace{(\alpha_1 - \alpha_2)}_{\leq 0}(\varrho - \underbrace{\mathcal{R}(u)}_{\leq \varrho}) \leq \varrho\alpha_2\end{aligned}$$

und somit $u \in M_{\alpha_2}(\varrho\alpha_2)$.

(iii) Die Inklusion „$\supseteq$“ ist offensichtlich und „$\subseteq$“ folgt, da für $u \in \bigcap_{\alpha>0} M_\alpha(\varrho\alpha)$ gilt:

$$\mathcal{S}(F(u), v^0)^p \leq \alpha(\varrho - \mathcal{R}(u)).$$

Der Grenzübergang $\alpha \to 0$ liefert wegen Voraussetzung 3.2 (iv)(a) die Gleichheit $F(u) = v^0$ und $\mathcal{R}(u) \leq \varrho$ ist offensichtlich erfüllt. □

Die Bedeutung der Niveaumengen $M_\alpha(\varrho\alpha)$ verdeutlicht die nächste Proposition.

Proposition 4.8. *Seien $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung, $\bar\alpha > 0$ und $\varrho > c_p s^p \mathcal{R}(u^\dagger)$. Sei weiter $\delta \mapsto \alpha(\delta)$ eine Parameterwahl mit*

$$\alpha(\delta) \to 0, \quad \frac{\delta^p}{\alpha(\delta)} \to 0 \quad \text{für } \delta \to 0$$

und sei $u^\delta_{\alpha(\delta)} \in \operatorname{argmin}\{\mathcal{T}^\delta_{\alpha(\delta)}(u) : u \in D\}$ für $\delta > 0$. Dann existiert ein $\bar\delta > 0$, sodass $u^\delta_{\alpha(\delta)} \in M_{\bar\alpha}(\varrho\bar\alpha)$ für alle $\delta \in (0, \bar\delta]$ gilt.

Beweis. Wegen $\alpha(\delta) \to 0$ und $\frac{\delta^p}{\alpha(\delta)} \to 0$ für $\delta \to 0$ existiert ein $\bar\delta > 0$ mit $\alpha(\delta) \leq \bar\alpha$ und $\frac{\delta^p}{\alpha(\delta)} \leq \frac{\varrho}{2c_p s^p} - \frac{1}{2}\mathcal{R}(u^\dagger)$ für alle $\delta \in (0, \bar\delta]$. Der Kürze halber schreiben wir im Folgenden nur α statt $\alpha(\delta)$. Für $\delta \in (0, \bar\delta]$ gilt nun (in Analogie zu [15, S. 5])

$$\begin{aligned}
\mathcal{T}^0_{\bar\alpha}(u^\delta_\alpha) &\leq \big(s\mathcal{S}(F(u^\delta_\alpha), v^\delta) + s\delta\big)^p + \bar\alpha\mathcal{R}(u^\delta_\alpha) \\
&\leq c_p s^p \big(\mathcal{S}(F(u^\delta_\alpha), v^\delta)^p + \alpha\mathcal{R}(u^\delta_\alpha) + (\bar\alpha - \alpha)\mathcal{R}(u^\delta_\alpha) + \delta^p\big) \\
&\leq c_p s^p \Big(2\delta^p + \alpha\mathcal{R}(u^\dagger) + \tfrac{\bar\alpha - \alpha}{\alpha}\big(\mathcal{S}(F(u^\delta_\alpha), v^\delta)^p + \alpha\mathcal{R}(u^\delta_\alpha)\big)\Big) \\
&= c_p s^p \big(\delta^p + \tfrac{\bar\alpha}{\alpha}\delta^p + \bar\alpha\mathcal{R}(u^\dagger)\big) \leq c_p s^p \bar\alpha \big(2\tfrac{\delta^p}{\alpha} + \mathcal{R}(u^\dagger)\big) \leq \varrho\bar\alpha.
\end{aligned}$$

□

4.2.2. Definition und Hauptsatz

Wir formulieren nun eine Variationsungleichung. Für den Spezialfall $\kappa = 1$ wurde diese bereits in [19] (Banach-Räume und Norm als Abstandsfunktional $\mathcal{S}$) und in [25] (topologische Räume und allgemeines Abstandsfunktional $\mathcal{S}$) untersucht. Der Fall $\kappa \in (0, 1]$ wurde für eine Banach-Raum-Situation mit der Norm als Abstandsfunktional $\mathcal{S}$ in [15, Beweis zu Theorem 3.3] betrachtet.

Definition 4.9. Eine $\mathcal{R}$-minimierende Lösung $u^\dagger$ *erfüllt eine Variationsungleichung*, falls ein $\xi \in \partial\mathcal{R}(u^\dagger)$ und Konstanten $\varrho > c_p s^p \mathcal{R}(u^\dagger)$, $\bar\alpha > 0$, $\beta_1 \in [0, 1)$, $\beta_2 \geq 0$ und $\kappa \in (0, p)$ existieren, sodass

$$-\xi(u - u^\dagger) \leq \beta_1 \mathcal{B}^{\mathcal{R}}_\xi(u, u^\dagger) + \beta_2 \mathcal{S}(F(u), F(u^\dagger))^\kappa \tag{4.1}$$

für alle $u \in M_{\bar\alpha}(\varrho\bar\alpha)$ gilt.

Das folgende Resultat zur Konvergenzgeschwindigkeit werden wir später in einem allgemeineren Zusammenhang beweisen.

Satz 4.10. *Sei $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung, die eine Variationsungleichung im Sinne von Definition 4.9 erfüllt, und sei $\delta \mapsto \alpha(\delta)$ eine Parameterwahl mit $\underline{c}\delta^{p-\kappa} \leq \alpha(\delta) \leq \overline{c}\delta^{p-\kappa}$ für hinreichend kleine δ und Konstanten $\underline{c} > 0$, $\overline{c} > 0$. Die Minimierer $u^\delta_{\alpha(\delta)}$ von $\mathcal{T}^\delta_{\alpha(\delta)}$ erfüllen dann*

$$\mathcal{B}^{\mathcal{R}}_\xi(u^\delta_{\alpha(\delta)}, u^\dagger) = \mathcal{O}(\delta^\kappa) \quad \text{für } \delta \to 0.$$

Interessant ist der Zusammenhang der hier formulierten Variationsungleichung mit den Quellbedingungen, die allerdings nur in Hilbert- und, mit Einschränkungen, in Banach-Räumen formuliert werden können. Auf diese Frage werden wir weiter unten zurückkommen.

4.2.3. Bemerkungen zu den Konstanten

Die Variationsungleichung aus Definition 4.9 enthält die künstlich erscheinende Einschränkung $\beta_1 < 1$. Durch Umformung der Ungleichung (4.1) kann diese rein formell umgangen werden:

$$\begin{aligned} & -\xi(u-u^\dagger) \leq \beta_1 \mathcal{B}^{\mathcal{R}}_\xi(u,u^\dagger) + \beta_2 \mathcal{S}(F(u),F(u^\dagger))^\kappa \\ \Leftrightarrow \quad & -(1-\beta_1)\xi(u-u^\dagger) \leq \beta_1(\mathcal{R}(u)-\mathcal{R}(u^\dagger)) + \beta_2 \mathcal{S}(F(u),F(u^\dagger))^\kappa \\ \Leftrightarrow \quad & -\xi(u-u^\dagger) \leq \tfrac{\beta_1}{1-\beta_1}(\mathcal{R}(u)-\mathcal{R}(u^\dagger)) + \tfrac{\beta_2}{1-\beta_1}\mathcal{S}(F(u),F(u^\dagger))^\kappa. \end{aligned}$$

Die Einschränkungen $\beta_1 \in [0,1)$ und $\beta_2 \geq 0$ sind nun äquivalent zu $\frac{\beta_1}{1-\beta_1} \geq 0$ und $\frac{\beta_2}{1-\beta_1} \geq 0$.

Proposition 4.11. *Erfüllt $u^\dagger$ eine Variationsungleichung im Sinne von Definition 4.9 mit Konstante κ, so erfüllt $u^\dagger$ auch eine Variationsungleichung für jedes $\tilde{\kappa} \in (0,\kappa)$.*

Beweis. Wir setzen $\tilde{\alpha} := \min\{\frac{1}{\varrho}, \bar{\alpha}\}$. Dann gilt

$$\mathcal{S}(F(u),F(u^\dagger))^p \leq \mathcal{T}^0_{\tilde{\alpha}}(u) \leq \varrho\tilde{\alpha} \leq \min\{1, \varrho\bar{\alpha}\}$$

für alle $u \in M_{\tilde{\alpha}}(\varrho\tilde{\alpha})$. D.h. $M_{\tilde{\alpha}}(\varrho\tilde{\alpha}) \subseteq M_{\bar{\alpha}}(\varrho\bar{\alpha})$ und somit folgt

$$-\xi(u-u^\dagger) \leq \beta_1 \mathcal{B}^{\mathcal{R}}_\xi(u,u^\dagger) + \beta_2 \underbrace{\mathcal{S}(F(u),F(u^\dagger))^\kappa}_{\leq 1} \leq \beta_1 \mathcal{B}^{\mathcal{R}}_\xi(u,u^\dagger) + \beta_2 \mathcal{S}(F(u),F(u^\dagger))^{\tilde{\kappa}}$$

für alle $u \in M_{\tilde{\alpha}}(\varrho\tilde{\alpha})$. □

Eine weitere Beobachtung zu den Konstanten aus Definition 4.9 liefert die folgende Proposition, welche für eine speziellere Situation bereits in [16, Proposition 4.3] bewiesen wurde.

Proposition 4.12. *Sei $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung, die eine Variationsungleichung im Sinne von Definition 4.9 erfüllt. Existieren ein $q > 0$, ein $u \in U$ mit $\xi(u) < 0$ und ein $t_0 > 0$, sodass $u^\dagger + tu \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ für alle $t \in [0, t_0]$ gilt, die Grenzwerte*

$$L_{\mathcal{R}} := \lim_{t \to +0} \frac{\mathcal{R}(u^\dagger + tu) - \mathcal{R}(u^\dagger)}{t}, \qquad L_{\mathcal{S}} := \lim_{t \to +0} \frac{\mathcal{S}(F(u^\dagger + tu), F(u^\dagger))^q}{t},$$

d.h. die Richtungsableitungen im Punkt $u^\dagger$ in Richtung u von $\mathcal{R}$ und $\mathcal{S}(F(\bullet), F(u^\dagger))^q$, existieren und $L_{\mathcal{R}} = \xi(u)$ erfüllt ist, dann muss $\kappa \leq q$ gelten.

Beweis. Sei $\kappa > q$. Für jedes $t \in (0, t_0]$ folgt dann aus (4.1)

$$-\xi(tu) \leq \beta_1(\mathcal{R}(u^\dagger + tu) - \mathcal{R}(u^\dagger) - \xi(tu)) + \beta_2 \mathcal{S}(F(u^\dagger + tu), F(u^\dagger))^\kappa$$

und somit

$$-\xi(u) \leq \beta_1 \underbrace{\left(\tfrac{\mathcal{R}(u^\dagger+tu)-\mathcal{R}(u^\dagger)}{t} - \xi(u)\right)}_{\overset{t\to+0}{\longrightarrow} 0} + \beta_2 \underbrace{\left(\tfrac{\mathcal{S}(F(u^\dagger+tu),F(u^\dagger))^q}{t}\right)^{\frac{\kappa}{q}}}_{\overset{t\to+0}{\longrightarrow} L_{\mathcal{S}}^{\frac{\kappa}{q}}} \underbrace{t^{\frac{\kappa}{q}-1}}_{\overset{t\to+0}{\longrightarrow} 0}.$$

Der Grenzübergang $t \to +0$ liefert $\xi(u) \geq 0$; dies steht jedoch im Widerspruch zu $\xi(u) < 0$. □

Wie das nächste, etwas künstliche Beispiel zeigt, können auch Variationsungleichungen mit $\kappa > 1$ auftreten.

Beispiel. Wir wählen $U := V := \mathbb{R}$, $\mathcal{R}(u) := |u|^2$, $\mathcal{S}(v_1, v_2) := |v_1 - v_2|$ und $p > 0$. Für beliebiges $u^\dagger \in U$ und $\mu > \frac{1}{p}$ setzen wir $F(u) := |u - u^\dagger|^\mu$ und $v^0 := 0$. Dann ist $u^\dagger$ die eindeutige und somit eine $\mathcal{R}$-minimierende Lösung von $F(u) = v^0$. Mit $\xi := \mathcal{R}'(u^\dagger) = 2u^\dagger$ und $\kappa := \frac{1}{\mu} < p$ folgt nun

$$-\xi(u - u^\dagger) = -2u^\dagger(u - u^\dagger) \leq \underbrace{2|u^\dagger|}_{=:\beta_2} |u - u^\dagger| = \beta_2 |u - u^\dagger|^{\mu\kappa} = \beta_2 |F(u) - F(u^\dagger)|^\kappa$$

für alle $u \in U$. Wir erhalten aus Satz 4.10 somit für beliebiges $\kappa \in (0, p)$ die Kovergenzrate $\mathcal{O}(\delta^\kappa)$.

Im vorhergehenden Beispiel kann Proposition 4.12 nicht angewendet werden, da in $u^\dagger$ keine Richtungsableitungen existieren.

4.3. Approximative Variationsungleichungen

Variationsungleichungen im Sinne von Definition 4.9 sind nicht immer erfüllt. Um trotzdem Aussagen über die Konvergenzgeschwindigkeit, mit der die Minimierer des Tikhonov-Funktionals gegen eine $\mathcal{R}$-minimierende Lösung konvergieren, erhalten zu können,

wurde in [18] das Konzept der Quellbedingungen um so genannte approximative Quellbedingungen erweitert.

Grundidee dieses neuen Ansatzes ist es, die Verletzung einer vorgegebenen herkömmlichen Quellbedingung, welche in diesem Zusammenhang als Referenz-Quellbedingung bezeichnet wird, mit Hilfe einer Funktion $d : [0, \infty) \to \mathbb{R}$ zu messen und aus dem Verhalten dieser Funktion im Unendlichen auf Konvergenzraten zu schließen. Die so erhaltenen Konvergenzraten werden natürlich geringer ausfallen als bei erfüllter Referenz-Quellbedingung.

Anfänglich wurden approximative Quellbedingungen in einer Hilbert-Raum-Situation für lineare Operatoren betrachtet. In [15] wurde dieses Konzept schließlich auf Banach-Räume und nichtlineare Operatoren übertragen. Inspiriert von den Resultaten in [15] in Verbindung mit [19] und [25] werden wir nun ein im Rahmen dieser Arbeit entwickeltes Konzept vorstellen, welches den Gedanke der approximativen Quellbedingungen auf die hier betrachtete allgemeine Form der Tikhonov-Regularisierung überträgt.

Anstelle einer Referenz-Quellbedingung verwenden wir eine Referenz-Variationsungleichung (kurz: Referenzungleichung) im Sinne von Definition 4.9. Motiviert wird das weitere Vorgehen durch Formel (4.3) in [15], da auf dieser Abschätzung die wesentlichen Konvergenzratenresultate in [15] beruhen. Wie sich zeigen wird, führt unser recht naiver Ansatz zu akzeptablen Ergebnissen.

4.3.1. Definition und wesentliche Eigenschaften

Wir betrachten die Ungleichung (4.1) für festes $u^\dagger$ und festes ξ sowie für gegebene Konstanten ρ, $\bar{\alpha}$, β_1, β_2 und κ. Eine Verletzung der Ungleichung bedeutet, dass für manche $u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ die strenge Ungleichung

$$-\xi(u - u^\dagger) > \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(u, u^\dagger) + \beta_2 \mathcal{S}(F(u), F(u^\dagger))^\kappa$$

gilt; umgestellt also

$$-\xi(u - u^\dagger) - \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(u, u^\dagger) - \beta_2 \mathcal{S}(F(u), F(u^\dagger))^\kappa > 0.$$

Die „maximale Verletzung“ der Ungleichung (4.1) kann somit als

$$\sup_{u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})} \left(-\xi(u - u^\dagger) - \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(u, u^\dagger) - \beta_2 \mathcal{S}(F(u), F(u^\dagger))^\kappa\right)$$

ausgedrückt werden. Die Frage, ob man durch Vergrößern des Faktors vor dem Ausdruck $\mathcal{S}(F(\bullet), F(u^\dagger))^\kappa$ die Erfüllung der Referenzungleichung erreichen kann, führt uns zur Definition der approximativen Variationsungleichung.

Definition 4.13. Eine $\mathcal{R}$-minimierende Lösung $u^\dagger$ *erfüllt eine approximative Variationsungleichung* (kurz: approximative Ungleichung), falls ein $\xi \in \partial\mathcal{R}(u^\dagger)$ und Konstanten $\varrho > c_p s^p \mathcal{R}(u^\dagger)$, $\bar{\alpha} > 0$, $\beta_1 \in [0, 1)$, $\beta_2 \geq 0$, $\gamma \geq 0$ und $\kappa \in (0, p)$ existieren, sodass

für die durch

$$d(r) := -\min_{u \in M_{\bar\alpha}(\varrho\bar\alpha)} \left(\xi(u-u^\dagger) + \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(u,u^\dagger) + \beta_2 r^\gamma \mathcal{S}(F(u),F(u^\dagger))^\kappa\right)$$

gegebene Funktion $d : [0,\infty) \to \mathbb{R}$ gilt: $d(r) \to 0$ für $r \to \infty$.

Die Konstante γ in Definition 4.13 scheint zunächst überflüssig zu sein. Bei der Formulierung von Konvergenzraten wird sie jedoch explizit in Erscheinung treten. Die Funktion d bezeichnen wir in Anlehnung an [15] als Abstandsfunktion, da die in [15] entsprechend eingeführte Funktion tatsächlich gewisse Abstände repräsentiert.

Ziel bei der Entwicklung des hier vorgestellten Konzepts war es, die Idee der approximativen Quellbedingungen auf die Tikhonov-Regularisierung mit allgemeinem Abstandsfunktional $\mathcal{S}$ zu übertragen und dabei gleichzeitig die Entfernung zur Arbeit [15] gering zu halten. Deshalb formulieren wir in der folgenden Proposition wesentliche Eigenschaften von d, die auch von den Abstandsfunktionen in [15] erfüllt werden.

Proposition 4.14. *Sei $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung, die eine approximative Ungleichung im Sinne von Definition 4.13 erfüllt.*

(i) Es gilt $0 \le d(r) < \infty$ für jedes $r \ge 0$.

(ii) Das Minimum in der Definition von d wird angenommen.

(iii) d ist stetig.

(iv) d ist monoton fallend.

(v) Gilt $d(r) > 0$ für alle $r \ge 0$, so ist d streng monoton fallend.

Beweis.

(i) Wegen $\mathcal{T}_{\bar\alpha}^0(u^\dagger) = \bar\alpha \mathcal{R}(u^\dagger) \le \varrho\bar\alpha$ ist $u^\dagger \in M_{\bar\alpha}(\varrho\bar\alpha)$ und somit $d(r) \ge 0$.
Für jedes $r \ge 0$ gilt

$$\begin{aligned}
d(r) &= \sup_{u \in M_{\bar\alpha}(\varrho\bar\alpha)} \left(-\xi(u-u^\dagger) - \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(u,u^\dagger) - \beta_2 r^\gamma \mathcal{S}(F(u),F(u^\dagger))^\kappa\right) \\
&\le \sup_{u \in M_{\bar\alpha}(\varrho\bar\alpha)} \left|\xi(u-u^\dagger) + \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(u,u^\dagger) + \beta_2 r^\gamma \mathcal{S}(F(u),F(u^\dagger))^\kappa\right| \\
&= \sup_{u \in M_{\bar\alpha}(\varrho\bar\alpha)} \Big| \underbrace{(1-\beta_1)}_{\le 1} \underbrace{\xi(u-u^\dagger)}_{=\xi(u)-\xi(u^\dagger)} + \beta_1 (\underbrace{\mathcal{R}(u)}_{\le \varrho} - \mathcal{R}(u^\dagger)) + \beta_2 r^\gamma \underbrace{\mathcal{S}(F(u),F(u^\dagger))^\kappa}_{\le (\varrho\bar\alpha)^{\frac{1}{p}}} \Big| \\
&\le C(r,\mathcal{R}(u^\dagger)) + \sup_{u \in M_{\bar\alpha}(\varrho\bar\alpha)} |\xi(u)|
\end{aligned}$$

mit einer von r und $\mathcal{R}(u^\dagger)$ abhängigen Konstante $C < \infty$. Es existiere nun eine Folge $(u_k)_{k\in\mathbb{N}}$ in $M_{\bar\alpha}(\varrho\bar\alpha)$ mit $|\xi(u_k)| \to \infty$. Wegen Voraussetzung 3.2 (vi) existiert

eine in U konvergente Teilfolge $(u_{k_l})_{l\in\mathbb{N}}$ von (u_k); das Grenzelement sei $\tilde{u} \in U$. Aus der Stetigkeit von ξ folgt $|\xi(u_{k_l})| \to |\xi(\tilde{u})|$ und somit die Beschränktheit der Folge $(|\xi(u_{k_l})|)$. Dies ist ein Widerspruch zu $|\xi(u_k)| \to \infty$; also gilt

$$\sup_{u\in M_{\bar{\alpha}}(\varrho\bar{\alpha})} |\xi(u)| < \infty,$$

d.h. $d(r) < \infty$.

(ii) Wir setzen $g_r : U \to \mathbb{R} \cup \{+\infty\}$ durch

$$g_r(u) := \xi(u - u^\dagger) + \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(u, u^\dagger) + \beta_2 r^\gamma \mathcal{S}(F(u), F(u^\dagger))^\kappa.$$

Analog zum Beweis von Proposition 4.6 erhalten wir die Unterhalbstetigkeit von g_r. Es sei nun $(u_k)_{k\in\mathbb{N}}$ eine Folge in $M_{\bar{\alpha}}(\varrho\bar{\alpha})$ mit $g_r(u_k) \to \inf_{u\in M_{\bar{\alpha}}(\varrho\bar{\alpha})} g_r(u)$. Dann existiert eine konvergente Teilfolge $(u_{k_l})_{l\in\mathbb{N}}$ von (u_k) mit Grenzelement $\tilde{u} \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ und es gilt

$$g_r(\tilde{u}) \le \liminf_{l\to\infty} g_r(u_{k_l}) = \lim_{l\to\infty} g_r(u_{k_l}) = \inf_{u\in M_{\bar{\alpha}}(\varrho\bar{\alpha})} g_r(u)$$

und somit $g_r(\tilde{u}) = \inf_{u\in M_{\bar{\alpha}}(\varrho\bar{\alpha})} g_r(u)$.

(iii) Für $r \ge 0$ sei g_r wie im Beweis zu Punkt (ii) und $u_r \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ sei ein Minimierer von g_r. Dann gilt für alle $r, s \ge 0$

$$d(r) - d(s) = \underbrace{\min g_s}_{\le g_s(u_r)} - \underbrace{\min g_r}_{= g_r(u_r)} \le \beta_2(s^\gamma - r^\gamma) \underbrace{\mathcal{S}(F(u_r), F(u^\dagger))^\kappa}_{\le (\varrho\bar{\alpha})^{\frac{1}{p}}}$$

und

$$-(d(r) - d(s)) = \underbrace{\min g_r}_{\le g_r(u_s)} - \underbrace{\min g_s}_{= g_s(u_s)} \le \beta_2(r^\gamma - s^\gamma) \underbrace{\mathcal{S}(F(u_s), F(u^\dagger))^\kappa}_{\le (\varrho\bar{\alpha})^{\frac{1}{p}}},$$

also

$$|d(r) - d(s)| \le \beta_2(\varrho\bar{\alpha})^{\frac{\kappa}{p}} |r^\gamma - s^\gamma|,$$

woraus die Stetigkeit von d folgt.

(iv) Die Behauptung folgt sofort aus der Definition von d.

(v) Wäre $\beta_2 = 0$ oder $\gamma = 0$, so wäre d konstant. Dies ist aber wegen $d(r) > 0$ für jedes $r \ge 0$ und $d(r) \to 0$ für $r \to \infty$ nicht möglich; also gilt $\beta_2 > 0$ und $\gamma > 0$.

Wir nehmen an, es existiert ein $r \ge 0$, für das g_r (gesetzt wie in Punkt (ii) des Beweises) einen Minimierer $\tilde{u} \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ besitzt, der $F(\tilde{u}) = v^0$ erfüllt. Dann gilt

für beliebiges $s \geq 0$

$$\begin{aligned}
&\xi(\tilde{u} - u^\dagger) + \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(\tilde{u}, u^\dagger) \\
&\quad = \xi(\tilde{u} - u^\dagger) + \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(\tilde{u}, u^\dagger) + \beta_2 s^\gamma \mathcal{S}(F(\tilde{u}), F(u^\dagger))^\kappa \\
&\quad \geq \min_{u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})} \left(\xi(u - u^\dagger) + \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(u, u^\dagger) + \beta_2 s^\gamma \mathcal{S}(F(u), F(u^\dagger))^\kappa \right) \\
&\quad = -d(s)
\end{aligned}$$

und somit folgt aus $d(s) \to 0$ für $s \to \infty$, dass $\xi(\tilde{u} - u^\dagger) + \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(\tilde{u}, u^\dagger) \geq 0$ gilt. Dies steht jedoch im Widerspruch zu

$$\begin{aligned}
&\xi(\tilde{u} - u^\dagger) + \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(\tilde{u}, u^\dagger) \\
&\quad = \xi(\tilde{u} - u^\dagger) + \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(\tilde{u}, u^\dagger) + \beta_2 r^\gamma \mathcal{S}(F(\tilde{u}), F(u^\dagger))^\kappa \\
&\quad = \min_{u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})} \left(\xi(u - u^\dagger) + \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(u, u^\dagger) + \beta_2 r^\gamma \mathcal{S}(F(u), F(u^\dagger))^\kappa \right) \\
&\quad = -d(r) < 0.
\end{aligned}$$

Also gilt für jedes $r \geq 0$, dass jeder Minimierer $\tilde{u} \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ von g_r die Ungleichung $\mathcal{S}(F(\tilde{u}), F(u^\dagger)) > 0$ erfüllt. Somit gilt für $0 \leq s < r$

$$\begin{aligned}
d(r) &= -\min_{u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})} g_r(u) = -g_r(\tilde{u}) \\
&= -g_s(\tilde{u}) - \underbrace{\beta_2(r^\gamma - s^\gamma)}_{>0} \underbrace{\mathcal{S}(F(\tilde{u}), F(u^\dagger))^\kappa}_{>0} \\
&< -g_s(\tilde{u}) \leq -\min_{u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})} g_s(u) \\
&= d(s),
\end{aligned}$$

d.h. d ist streng monoton fallend. □

Proposition 4.15.

(i) Eine $\mathcal{R}$-minimierende Lösung erfüllt genau dann eine Variationsungleichung im Sinne von Definition 4.9, wenn sie eine approximative Ungleichung im Sinne von Definition 4.13 erfüllt und ein $r_0 \geq 0$ existiert, sodass $d(r_0) = 0$ gilt.

(ii) Erfüllt $u^\dagger$ eine approximative Ungleichung im Sinne von Definition 4.13 mit Konstante $\bar{\alpha}$, so ist auch für jedes $\alpha \in (0, \bar{\alpha}]$ eine approximative Ungleichung (mit den selben weiteren Konstanten) erfüllt.

Beweis.

(i) Erfüllt $u^\dagger$ eine Variationsungleichung, dann erfüllt $u^\dagger$ offensichtlich auch eine approximative Ungleichung mit $d(1) = 0$ und beliebigem $\gamma > 0$. Die Rückrichtung

folgt aus

$$-\xi(u-u^\dagger)-\beta_1\mathcal{B}_\xi^{\mathcal{R}}(u,u^\dagger)-\beta_2 r_0^\gamma\mathcal{S}(F(u),F(u^\dagger))^\kappa\le d(r_0)=0$$

für $u\in M_{\bar\alpha}(\varrho\bar\alpha)$.

(ii) Diese Behauptung ist offensichtlich (vgl. Proposition 4.7 (ii)). □

Definition 4.13 hat trotz Proposition 4.14 noch einen Schönheitsfehler: Wie wir später feststellen werden, tritt die Konstante $\bar\alpha$ aus der approximativen Ungleichung bei der Formulierung von Konvergenzraten nicht explizit in Erscheinung, sodass es aufgrund der Aussage von Proposition 4.15 (ii) „wünschenswert“ wäre, wenn die Funktion d oder wenigstens ihr Verhalten im Unendlichen nicht von $\bar\alpha$ abhängen würde. Ein Resultat in diese Richtung liefert die übernächste Proposition. Als Vorbereitung dazu zeigt die folgende Proposition, dass sich die Elemente, die das Minimum in der Definition von $d(r)$ realisieren, für $r\to\infty$ bei speziellen Elementen $\tilde u$ mit $F(\tilde u)=v^0$ „konzentrieren“.

Proposition 4.16. *Sei $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung, die eine approximative Ungleichung im Sinne von Definition 4.13 erfüllt. Weiter seien $(r_k)_{k\in\mathbb{N}}$ eine Folge in $(0,\infty)$ mit $r_k\to\infty$ und $(u_k)_{k\in\mathbb{N}}$ eine Folge von Elementen $u_k\in M_{\bar\alpha}(\varrho\bar\alpha)$, die das Minimum in der Definition der Werte $d(r_k)$ realisieren, sodass $u_k\to\tilde u$ für ein $\tilde u\in D$ gilt. Dann folgt*

$$F(\tilde u)=v^0,\quad \mathcal{R}(\tilde u)\le\varrho\quad\textit{und}\quad \xi(\tilde u-u^\dagger)=\frac{-\beta_1}{1-\beta_1}(\mathcal{R}(\tilde u)-\mathcal{R}(u^\dagger)).$$

Beweis. Aus der Definition von u_k und $d(r_k)$ folgt

$$-\beta_2 r_k^\gamma\mathcal{S}(F(u_k),F(u^\dagger))^\kappa=d(r_k)+\xi(u_k-u^\dagger)+\beta_1\mathcal{B}_\xi^{\mathcal{R}}(u_k,u^\dagger).$$

Aufgrund der Stetigkeit von ξ und der Unterhalbstetigkeit von $\mathcal{R}$ gilt für beliebiges $\varepsilon>0$ bei hinreichend großem $k\in\mathbb{N}$ somit

$$-\beta_2 r_k^\gamma\mathcal{S}(F(u_k),F(u^\dagger))^\kappa\ge d(r_k)+\xi(\tilde u-u^\dagger)+\beta_1\mathcal{B}_\xi^{\mathcal{R}}(\tilde u,u^\dagger)-\varepsilon,$$

also auch

$$\mathcal{S}(F(u_k),F(u^\dagger))^\kappa\le\underbrace{\tfrac{-1}{\beta_2 r_k^\gamma}}_{\to0}\big(\underbrace{d(r_k)}_{\to0}+\xi(\tilde u-u^\dagger)+\beta_1\mathcal{B}_\xi^{\mathcal{R}}(\tilde u,u^\dagger)-\varepsilon\big).$$

Der Grenzübergang $k\to\infty$ liefert

$$\mathcal{S}(F(u_k),F(u^\dagger))^\kappa\to0$$

und mit Voraussetzung 3.2 (iv)(d) folgt daraus $F(u_k)\to v^0$. Andererseits gilt wegen Voraussetzung 3.2 (i) auch $F(u_k)\to F(\tilde u)$ und damit $F(\tilde u)=v^0$.

Die zweite Behauptung folgt aus

$$\mathcal{R}(\tilde{u}) \leq \liminf_{k\to\infty} \mathcal{R}(u_k) \leq \liminf_{k\to\infty} \tfrac{1}{\bar{\alpha}} \mathcal{T}^0_{\bar{\alpha}}(u_k) \leq \varrho.$$

Zum Beweis der dritten und letzten Behauptung stellen wir zunächst

$$\begin{aligned} -\xi(\tilde{u}-u^\dagger) - \beta_1 \mathcal{B}^{\mathcal{R}}_\xi(\tilde{u},u^\dagger) &= -\xi(\tilde{u}-u^\dagger) - \beta_1 \mathcal{B}^{\mathcal{R}}_\xi(\tilde{u},u^\dagger) - \beta_2 r_k^\gamma \mathcal{S}(F(\tilde{u}),F(u^\dagger))^\kappa \\ &\leq d(r_k) \to 0 \end{aligned}$$

und damit

$$-\xi(\tilde{u}-u^\dagger) - \beta_1 \mathcal{B}^{\mathcal{R}}_\xi(\tilde{u},u^\dagger) \leq 0 \tag{4.2}$$

fest. Für beliebiges $\varepsilon > 0$ liefert die Stetigkeit von ξ und die Unterhalbstetigkeit von $\mathcal{R}$ bei hinreichend großem $k \in \mathbb{N}$

$$\begin{aligned} 0 &\geq -\beta_2 r_k^\gamma \mathcal{S}(F(u_k),F(u^\dagger))^\kappa = d(r_k) + \xi(u_k-u^\dagger) + \beta_1 \mathcal{B}^{\mathcal{R}}_\xi(u_k,u^\dagger) \\ &\geq \underbrace{d(r_k)}_{\to 0} + \xi(\tilde{u}-u^\dagger) + \beta_1 \mathcal{B}^{\mathcal{R}}_\xi(\tilde{u},u^\dagger) - \varepsilon. \end{aligned}$$

Aus dem Grenzübergang $k \to \infty$ folgt $\xi(\tilde{u}-u^\dagger) + \beta_1 \mathcal{B}^{\mathcal{R}}_\xi(\tilde{u},u^\dagger) \leq \varepsilon$ und da $\varepsilon > 0$ beliebig war, erhalten wir

$$-\xi(\tilde{u}-u^\dagger) - \beta_1 \mathcal{B}^{\mathcal{R}}_\xi(\tilde{u},u^\dagger) \geq 0. \tag{4.3}$$

Die Formeln (4.2) und (4.3) ergeben zusammen

$$-\xi(\tilde{u}-u^\dagger) = \beta_1 \mathcal{B}^{\mathcal{R}}_\xi(\tilde{u},u^\dagger)$$

und das Einsetzen der Definition des Bregman-Abstandes liefert die Behauptung. □

Proposition 4.17. *Sei $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung, die eine approximative Ungleichung im Sinne von Definition 4.13 erfüllt und sei d_α für $\alpha \in (0,\bar{\alpha}]$ die mit α statt $\bar{\alpha}$ analog zu d definierte Funktion. Existiert kein $u \in U$ mit $F(u) = v^0$, $\mathcal{R}(u) = \varrho$ und $\xi(u-u^\dagger) = \frac{-\beta_1}{1-\beta_1}(\varrho - \mathcal{R}(u^\dagger))$, so gilt:*

(i) Für jedes $\alpha \in (0,\bar{\alpha}]$ existiert ein $r_\alpha \geq 0$, sodass $d(r) = d_\alpha(r)$ für alle $r \geq r_\alpha$ erfüllt ist.

(ii) Für jedes $\alpha \in (0,\bar{\alpha}]$ existiert ein $r_\alpha \geq 0$, sodass für jedes $r \geq r_\alpha$ sämtliche Elemente aus $M_{\bar{\alpha}}(\varrho\bar{\alpha})$, für die das Minimum in der Definition von $d(r)$ angenommen wird, in $M_\alpha(\varrho\alpha)$ liegen.

Beweis. Behauptung (i) folgt sofort aus (ii). Behauptung (ii) beweisen wir indirekt. Wir nehmen also an, dass ein $\alpha \in (0,\bar{\alpha}]$ und eine Folge $(r_k)_{k\in\mathbb{N}}$ in $(0,\infty)$ mit $r_k \to \infty$ existieren, sodass es zu jedem r_k ein Element $u_k \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ gibt, welches das Minimum in der Definition von $d(r_k)$ realisiert und $u_k \notin M_\alpha(\varrho\alpha)$ erfüllt. Wegen Proposition 4.6

besitzt $(u_k)_{k\in\mathbb{N}}$ eine konvergente Teilfolge mit Grenzelement $\tilde{u} \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$. Diese Teilfolge bezeichnen wir wieder mit $(u_k)_{k\in\mathbb{N}}$.

Nach Proposition 4.16 gilt nun

$$F(\tilde{u}) = v^0, \quad \mathcal{R}(\tilde{u}) \leq \varrho \quad \text{und} \quad \xi(\tilde{u} - u^\dagger) = \frac{-\beta_1}{1-\beta_1}(\mathcal{R}(\tilde{u}) - \mathcal{R}(u^\dagger)). \tag{4.4}$$

Wegen $u_k \notin M_\alpha(\varrho\alpha)$ gilt außerdem

$$\mathcal{R}(u_k) > \varrho - \tfrac{1}{\alpha}\mathcal{S}(F(u_k), v^0)^p$$

für jedes $k \in \mathbb{N}$ und somit folgt aus $\mathcal{S}(F(u_k), F(u^\dagger)) \to 0$ (vgl. Beweis zu Proposition 4.16), dass für beliebiges $\varepsilon > 0$ bei hinreichend großem $k \in \mathbb{N}$ die Ungleichung $\mathcal{R}(u_k) > \varrho - \varepsilon$ gilt. Zusammen mit $\mathcal{R}(u_k) \leq \varrho$ liefert dies $\mathcal{R}(u_k) \to \varrho$. Damit können wir aus

$$0 \geq -\beta_2 r_k^\gamma \mathcal{S}(F(u_k), F(u^\dagger))^\kappa = \underbrace{d(r_k)}_{\to 0} + \xi(u_k - u^\dagger) + \beta_1\big(\mathcal{R}(u_k) - \mathcal{R}(u^\dagger) - \xi(u_k - u^\dagger)\big)$$

durch Grenzübergang die Beziehung

$$0 \geq (1-\beta_1)\xi(\tilde{u} - u^\dagger) + \beta_1(\varrho - \mathcal{R}(u^\dagger))$$

folgern und erhalten zusammen mit (4.4) die Beziehung

$$\tfrac{-\beta_1}{1-\beta_1}(\mathcal{R}(\tilde{u}) - \mathcal{R}(u^\dagger)) = \xi(\tilde{u} - u^\dagger) \leq \frac{-\beta_1}{1-\beta_1}(\varrho - \mathcal{R}(u^\dagger)) \leq \frac{-\beta_1}{1-\beta_1}(\mathcal{R}(\tilde{u}) - \mathcal{R}(u^\dagger));$$

d.h. insbesondere gilt $\mathcal{R}(\tilde{u}) = \varrho$. Einsetzen dieser Gleichheit in die Formeln (4.4) liefert einen Widerspruch zu den Voraussetzungen der Proposition. □

4.3.2. Hauptsatz

Die Vorbetrachtungen sind nun abgeschlossen, sodass wir mit dem nächsten Lemma den entscheidenden Schritt zum Beweis von Konvergenzraten gehen können. Dieses Lemma sowie auch der darauf folgende Satz sind im Wesentlichen aus [15] entnommen und wurden nur geringfügig an die allgemeinere Situation dieser Arbeit angepasst.

Lemma 4.18. *Sei $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung, die eine approximative Ungleichung im Sinne von Definition 4.13 erfüllt. Sei weiter $\alpha \mapsto \alpha(\delta)$ eine Parameterwahl wie in Proposition 4.8 und sei $\bar{\delta}$ die entsprechende Konstante aus Proposition 4.8. Dann existieren Konstanten $K_1 > 0$, $K_2 > 0$ und $K_3 > 0$, sodass*

$$\mathcal{B}_\xi^{\mathcal{R}}(u^\delta_{\alpha(\delta)}, u^\dagger) \leq K_1 \tfrac{\delta^p}{\alpha(\delta)} + K_2 \alpha(\delta)^{\frac{\kappa}{p-\kappa}} r^{\frac{\gamma p}{p-\kappa}} + K_3 d(r)$$

für alle $r \geq 0$ und alle $\delta \in (0, \bar{\delta}]$ gilt, wobei $u^\delta_{\alpha(\delta)}$ einen Minimierer von $\mathcal{T}^\delta_{\alpha(\delta)}$ bezeichnet.

Beweis. Wir schreiben der Kürze halber nur α statt $\alpha(\delta)$. Proposition 4.8 und die Definition von $d(r)$ liefern uns die Ungleichung

$$-\xi(u_\alpha^\delta - u^\dagger) \le \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(u_\alpha^\delta, u^\dagger) + \beta_2 r^\gamma \mathcal{S}(F(u_\alpha^\delta), F(u^\dagger))^\kappa + d(r)$$

für hinreichend kleines δ. Damit erhalten wir

$$\begin{aligned}
\alpha\mathcal{B}_\xi^{\mathcal{R}}(u_\alpha^\delta, u^\dagger) &= \mathcal{S}(F(u_\alpha^\delta), v^\delta)^p + \alpha\mathcal{R}(u_\alpha^\delta) - \alpha\mathcal{R}(u^\dagger) - \alpha\xi(u_\alpha^\delta - u^\dagger) - \mathcal{S}(F(u_\alpha^\delta), v^\delta)^p \\
&\le \delta^p - \alpha\xi(u_\alpha^\delta - u^\dagger) - \mathcal{S}(F(u_\alpha^\delta), v^\delta)^p \\
&\le \delta^p + \alpha\beta_1 \mathcal{B}_\xi^{\mathcal{R}}(u_\alpha^\delta, u^\dagger) + \alpha\beta_2 r^\gamma \mathcal{S}(F(u_\alpha^\delta), F(u^\dagger))^\kappa + \alpha d(r) - \mathcal{S}(F(u_\alpha^\delta), v^\delta)^p \\
&\le \delta^p + \alpha\beta_1 \mathcal{B}_\xi^{\mathcal{R}}(u_\alpha^\delta, u^\dagger) + \alpha\beta_2 r^\gamma s^\kappa c_\kappa \big(\mathcal{S}(F(u_\alpha^\delta), v^\delta)^\kappa + \delta^\kappa\big) \\
&\quad + \alpha d(r) - \mathcal{S}(F(u_\alpha^\delta), v^\delta)^p,
\end{aligned}$$

wobei wir mit c_κ analog zu c_p die durch

$$c_\kappa := \begin{cases} 1 & \text{für } \kappa \in (0,1), \\ 2^{\kappa-1} & \text{für } \kappa \ge 1 \end{cases}$$

gegebene Konstante bezeichnen. Somit gilt also

$$\begin{aligned}
\mathcal{B}_\xi^{\mathcal{R}}(u_\alpha^\delta, u^\dagger) \le \frac{1}{\alpha(1-\beta_1)} \Big(&2\delta^p + \alpha c_\kappa \beta_2 r^\gamma s^\kappa \delta^\kappa - \delta^p + \alpha c_\kappa \beta_2 r^\gamma s^\kappa \mathcal{S}(F(u_\alpha^\delta), v^\delta)^\kappa \\
&- \mathcal{S}(F(u_\alpha^\delta), v^\delta)^p + \alpha d(r) \Big).
\end{aligned}$$

Wir wenden nun die Beziehung

$$ab - \varepsilon a^{p_1} \le \frac{b^{p_2}}{(\varepsilon p_1)^{p_2/p_1} p_2},$$

wobei $a, b \ge 0$, $\varepsilon > 0$, $p_1, p_2 > 1$ und $\frac{1}{p_1} + \frac{1}{p_2} = 1$ gelten soll, einmal mit

$$a := \delta^\kappa, \quad b := \alpha c_\kappa \beta_2 r^\gamma s^\kappa, \quad \varepsilon := 1, \quad p_1 := \frac{p}{\kappa}, \quad p_2 := \frac{p}{p-\kappa}$$

und ein zweites Mal mit $\mathcal{S}(F(u_\alpha^\delta), v^\delta)$ statt δ an. Wir erhalten

$$\begin{aligned}
\mathcal{B}_\xi^{\mathcal{R}}(u_\alpha^\delta, u^\dagger) &\le \tfrac{1}{\alpha(1-\beta_1)} \Big(2\delta^p + 2(c_\kappa \beta_2 s^\kappa)^{\frac{p}{p-\kappa}} \big(\tfrac{\kappa}{p}\big)^{\frac{\kappa}{p-\kappa}} \tfrac{p-\kappa}{p} \alpha^{\frac{p}{p-\kappa}} r^{\frac{\gamma p}{p-\kappa}} + \alpha d(r) \Big) \\
&= \underbrace{\tfrac{2}{1-\beta_1}}_{=:K_1} \tfrac{\delta^p}{\alpha} + \underbrace{2(c_\kappa \beta_2 s^\kappa)^{\frac{p}{p-\kappa}} \big(\tfrac{\kappa}{p}\big)^{\frac{\kappa}{p-\kappa}} \tfrac{p-\kappa}{p(1-\beta_1)}}_{=:K_2} \alpha^{\frac{\kappa}{p-\kappa}} r^{\frac{\gamma p}{p-\kappa}} + \underbrace{\tfrac{1}{1-\beta_1}}_{=:K_3} d(r).
\end{aligned}$$

□

Beweis zu Satz 4.10. Nach Proposition 4.15 (i) erfüllt $u^\dagger$ eine approximative Ungle-

chung mit $d(r_0) = 0$ für ein $r_0 \geq 0$, sodass die Behauptung für $r := r_0$ aus Lemma 4.18 folgt. □

Satz 4.19. *Sei $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung, die eine approximative Ungleichung im Sinne von Definition 4.13 mit $d(r) > 0$ für alle $r \geq 0$ erfüllt. Für $r > 0$ setzen wir*

$$\Psi(r) := d(r)^{\frac{p-\kappa}{\kappa}} r^{-\frac{\gamma p}{\kappa}} \quad \textit{und} \quad \Phi(r) := d(r)^{\frac{1}{\kappa}} r^{-\frac{\gamma}{\kappa}}.$$

Weiterhin sei $\alpha \mapsto \alpha(\delta)$ eine Parameterwahl, die $\delta^p = \alpha(\delta) d\big(\Psi^{-1}(\alpha(\delta))\big)$ für hinreichend kleines $\delta > 0$ erfüllt. Dann gilt

$$\mathcal{B}_\xi^{\mathcal{R}}(u^\delta_{\alpha(\delta)}, u^\dagger) = \mathcal{O}\big(d(\Phi^{-1}(\delta))\big) \quad \textit{für } \delta \to 0,$$

wobei $u^\delta_{\alpha(\delta)}$ einen Minimierer von $\mathcal{T}^\delta_{\alpha(\delta)}$ bezeichnet.

Beweis. Der Kürze halber schreiben wir statt $\alpha(\delta)$ nur α. Da d streng monoton fallend ist, sind auch Ψ und Φ streng monoton fallend. Somit existieren die Umkehrfunktionen Ψ^{-1} und Φ^{-1}, welche ebenfalls streng monoton fallend sind.

Aus Lemma 4.18 mit $r := \Psi^{-1}(\alpha)$, also

$$\alpha^{\frac{\kappa}{p-\kappa}} r^{\frac{\gamma p}{p-\gamma}} = \Psi(r)^{\frac{\kappa}{p-\kappa}} r^{\frac{\gamma p}{p-\gamma}} = d(r),$$

erhalten wir

$$\mathcal{B}_\xi^{\mathcal{R}}(u^\delta_\alpha, u^\dagger) \leq K_1 \frac{\delta^p}{\alpha} + (K_2 + K_3) d(\Psi^{-1}(\alpha)) = (K_1 + K_2 + K_3) d(\Psi^{-1}(\alpha))$$

für alle hinreichend kleinen $\delta \leq \bar{\delta}$ und wegen

$$\begin{aligned} \Phi(\Psi^{-1}(\alpha)) &= d(\Psi^{-1}(\alpha))^{\frac{1}{\kappa}} \Psi^{-1}(\alpha)^{-\frac{\gamma}{\kappa}} = \left(\frac{\delta^p}{\alpha}\right)^{\frac{1}{\kappa}} \Psi^{-1}(\alpha)^{-\frac{\gamma}{\kappa}} \\ &= \delta^{\frac{p}{\kappa}} \underbrace{\left(\alpha^{\frac{\kappa}{p-\kappa}} \Psi^{-1}(\alpha)^{\frac{\gamma p}{p-\kappa}}\right)^{\frac{\kappa-p}{\kappa p}}}_{= d(\Psi^{-1}(\alpha)) = \frac{\delta^p}{\alpha}} \alpha^{\frac{1}{p}-\frac{1}{\kappa}} = \delta \end{aligned}$$

folgt $\Psi^{-1}(\alpha) = \Phi^{-1}(\delta)$ und damit die Behauptung. □

Bemerkung. Ist statt d nur eine streng monoton gegen Null fallende Majorante $\bar{d}$ zu d bekannt, so gelten Lemma 4.18 und Satz 4.19 auch mit $\bar{d}$ statt d.

4.3.3. Bemerkungen

Zunächst wollen wir festhalten, dass bei Erfüllung einer approximativen Ungleichung die Konvergenzraten aus Satz 4.19 nicht höher als bei Erfüllung der entsprechenden

Referenzungleichung sein können. Dieses Resultat entnehmen wir [15, Remark 4.5]; den Beweis geben wir nochmals mit der hier verwendeten Notation an.

Proposition 4.20. *Mit den Bezeichnungen aus Satz 4.19 gilt*

$$\delta^\kappa = o\big(d(\Phi^{-1}(\delta))\big) \quad \textit{für } \delta \to 0.$$

Beweis. Mit $r := \Phi^{-1}(\delta)$, also $\delta = \Phi(r)$, gilt

$$\frac{\delta^\kappa}{d(\Phi^{-1}(\delta))} = \frac{\Phi(r)^\kappa}{d(r)} = \frac{d(r)r^{-\gamma}}{d(r)} = r^{-\gamma} = \Phi^{-1}(\delta)^{-\gamma}.$$

Aus $\gamma > 0$ und $\Phi^{-1}(\delta) \to \infty$ für $\delta \to 0$ folgt $\Phi^{-1}(\delta)^{-\gamma} \to 0$ für $\delta \to 0$ und somit die Behauptung. □

Eine weitere Proposition klärt den Einfluss des Abklingverhaltens der Abstandsfunktion d auf die Konvergenzrate.

Proposition 4.21. *Mit den Bezeichnungen aus Satz 4.19 folgt aus*

$$d(r) = \mathcal{O}(\tilde{d}(r)) \quad \textit{für } r \to \infty$$

die Beziehung

$$d(\Phi^{-1}(\delta)) = \mathcal{O}\big(\tilde{d}(\tilde{\Phi}^{-1}(\delta))\big) \quad \textit{für } \delta \to 0,$$

wobei $\tilde{\Phi}$ analog zu Φ mit $\tilde{d}$ statt d definiert sei.

Beweis. Aus $d(r) = \mathcal{O}(\tilde{d}(r))$ für $r \to \infty$ folgt $\Phi(r) = \mathcal{O}(\tilde{\Phi}(r))$ für $r \to \infty$ und daraus $d(\Phi^{-1}(\delta)) = \mathcal{O}(\tilde{\Phi}^{-1}(\delta))$ für $\delta \to 0$. Mit $r := \Phi^{-1}(\delta)$, also $\delta = \Phi(r)$, gilt nun für hinreichend kleines $\delta > 0$ und für eine Konstante $c > 0$

$$d(\Phi^{-1}(\delta)) = \delta^\kappa \frac{d(\Phi^{-1}(\delta))}{\delta^\kappa} = \Phi(r)^\kappa \frac{d(r)}{d(r)r^{-\gamma}} = \delta^\kappa \Phi^{-1}(\delta)^\gamma \le c\delta^\kappa \tilde{\Phi}^{-1}(\delta)^\gamma = c\tilde{d}(\tilde{\Phi}^{-1}(\delta)).$$

□

Die folgende Proposition zeigt, dass die Konstante γ in der Definition der approximativen Ungleichung in bestimmten Situationen keinen Einfluss auf die erzielten Konvergenzraten hat. Dass solche Fälle tatsächlich auftreten können, werden wir im Beweis zur übernächsten Proposition sehen.

Proposition 4.22. *Sei $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung, die eine approximative Ungleichung im Sinne von Definition 4.13 mit $d(r) > 0$ für alle $r \ge 0$ erfüllt. Existiert zu d eine Majorante $\bar{d}$ der Form $\bar{d}(r) = ar^{-b\gamma}$ mit $a > 0$ und $b > 0$, so liefert Satz 4.19 die Parameterwahl*

$$\alpha(\delta) = c\delta^{p - \frac{b}{b+1}\kappa}$$

mit einer Konstante $c > 0$ und die Konvergenzrate

$$\mathcal{B}_\xi^{\mathcal{R}}(u_{\alpha(\delta)}^\delta, u^\dagger) = \mathcal{O}(\delta^{\frac{b}{b+1}\kappa}).$$

Beweis. Der Kürze halber schreiben wir statt $\alpha(\delta)$ nur α. Mit den Bezeichnungen aus Satz 4.19 und gewissen Konstanten $a_1, a_2, a_3 > 0$ erhalten wir

$$\Psi(r) = \bar{d}(r)^{\frac{p-\kappa}{\kappa}} r^{-\frac{\gamma p}{\kappa}} = a_1 r^{\frac{-b\gamma(p-\kappa)-\gamma p}{\kappa}},$$

$$\alpha d(\Psi^{-1}(\alpha)) = a_2 \alpha^{\frac{b\kappa}{b(p-\gamma)+p}+1} = a_2 \alpha^{\frac{p(b+1)}{p(b+1)+b\kappa}}$$

und aus $\alpha d(\Psi^{-1}(\alpha)) = \delta^p$ somit

$$\delta = a_3 \delta^{\frac{p(b+1)-b\kappa}{b+1}} = a_3 \delta^{p-\frac{b}{b+1}\kappa}.$$

Mit Konstanten $a_4, a_5 > 0$ gilt des Weiteren

$$\Phi(r) = \bar{d}(r)^{\frac{1}{\kappa}} r^{-\frac{\gamma}{\kappa}} = a_4 r^{\frac{-b\gamma-\gamma}{\kappa}}$$

und somit

$$d(\Phi^{-1}(\delta)) = a_5 \delta^{\frac{b\kappa}{b+1}} = a_5 \delta^{\frac{b}{b+1}\kappa}.$$

□

Zur Unterstützung der Plausibilität des in dieser Arbeit vorgestellten Konzepts der approximativen Variationsungleichungen ist es interessant, ob bei erfüllter Variationsungleichung auch eine approximative Ungleichung für eine Referenzungleichung mit größerem κ erfüllt ist; und wenn ja, stimmen dann die auf diesen zwei Wegen erhaltenen Konvergenzraten überein? Die nächste Proposition beantwortet diese Frage positiv. Ein entsprechendes Resultat für lineare Operatoren in Hilbert-Räumen und approximative Quellbedingungen findet man in [8].

Proposition 4.23. *Sei $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung, die eine Variationsungleichung im Sinne von Definition 4.9 erfüllt, und sei $\mu \in (\kappa, p)$ so gewählt, dass $u^\dagger$ keine Variationsungleichung mit μ anstelle von κ erfüllt. Dann erfüllt $u^\dagger$ eine approximative Ungleichung im Sinne von Definition 4.13 mit μ anstelle von κ und die aus der exakten Ungleichung mit κ und der approximativen Ungleichung mit μ erhaltenen Konvergenzraten stimmen überein.*

Beweis. Die Konstanten ϱ, $\bar{\alpha}$, β_1 und β_2 seien die Konstanten aus der von $u^\dagger$ mit κ erfüllten Variationsungleichung; $\gamma > 0$ sei beliebig. Für alle $u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ und alle $r > 0$ gilt dann

$$\begin{aligned} -\xi(u - u^\dagger) - \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(u, u^\dagger) - \beta_2 r^\gamma \mathcal{S}(F(u), F(u^\dagger))^\mu & \\ \leq \beta_2 \mathcal{S}(F(u), F(u^\dagger))^\kappa - \beta_2 r^\gamma \mathcal{S}(F(u), F(u^\dagger))^\mu & \end{aligned}$$

und aus

$$ab - \varepsilon a^{p_1} \leq \frac{b^{p_2}}{(\varepsilon p_1)^{p_2/p_1} p_2}$$

für $a, b \geq 0$, $\varepsilon > 0$, $p_1, p_2 > 1$ und $\frac{1}{p_1} + \frac{1}{p_2} = 1$ mit

$$a := \big(\beta_2 r^\gamma \mathcal{S}(F(u), F(u^\dagger))^\mu\big)^{\frac{\kappa}{\mu}}, \quad b := \beta_2^{\frac{\mu-\kappa}{\mu}} r^{-\frac{\gamma\kappa}{\mu}}, \quad \varepsilon := 1, \quad p_1 := \frac{\mu}{\kappa}, \quad p_2 := \frac{\mu}{\mu-\kappa}$$

folgt somit

$$\begin{aligned} d(r) &\leq \max_{M_{\bar\alpha}(\varrho\bar\alpha)} \big(-\xi(u-u^\dagger) - \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(u, u^\dagger) - \beta_2 r^\gamma \mathcal{S}(F(u), F(u^\dagger))^\mu\big) \\ &\leq \big(\tfrac{\mu}{\kappa}\big)^{\frac{\kappa}{\mu-\kappa}} \tfrac{\mu}{\mu-\kappa} \beta_2 r^{-\frac{\gamma\kappa}{\mu-\kappa}}, \end{aligned}$$

d.h. $u^\dagger$ erfüllt eine approximative Ungleichung und wir haben zur entsprechenden Abstandsfunktion d eine Majorante $\bar d$ der Form $\bar d(r) = a r^{-b\gamma}$ mit $a > 0$ und $b = \frac{\kappa}{\mu-\kappa}$ gefunden. Nach Proposition 4.22 mit μ anstelle von κ gilt nun

$$\mathcal{B}_\xi^{\mathcal{R}}(u^\delta_{\alpha(\delta)}, u^\dagger) = \mathcal{O}(\delta^{\frac{b}{b+1}\mu}) = \mathcal{O}(\delta^{\frac{\kappa}{\mu}\mu}) = \mathcal{O}(\delta^\kappa)$$

für die Parameterwahl

$$\alpha(\delta) = c\delta^{p-\frac{b}{b+1}\mu} = c\delta^{p-\frac{\kappa}{\mu}\mu} = c\delta^{p-\kappa}$$

mit einer Konstante $c > 0$. Dies ist genau die Konvergenzrate, die wir mit Satz 4.10 auch direkt aus der für κ erfüllten Variationsungleichung erhalten. □

In Proposition 4.12 wurde gezeigt, dass unter schwachen Voraussetzungen für Variationsungleichungen stets $\kappa \leq q$ mit einem gewissen $q > 0$ gelten muss. Auf den ersten Blick erscheint es denkbar, dass dies für approximative Ungleichungen nicht gelten muss, sodass über den Weg der approximativen Ungleichungen unter Umständen auch höhere Raten als $\mathcal{O}(\delta^q)$ erreicht werden können. Die folgende Proposition schließt zwar nicht aus, dass für approximative Ungleichungen auch $\kappa > q$ möglich ist, jedoch zeigt sie, dass die maximal erreichbare Konvergenzrate durch δ^q begrenzt wird.

Proposition 4.24. *Seien die Voraussetzungen von Proposition 4.12 erfüllt, wobei $u^\dagger$ jedoch eine approximative Ungleichung mit $d(r) > 0$ für alle $r \geq 0$ erfülle. Mit den Bezeichnungen aus Satz 4.19 gilt dann*

$$\delta^q = \mathcal{O}\big(d(\Phi^{-1}(\delta))\big) \quad \textit{für } \delta \to 0.$$

Beweis. Wir beweisen die Behauptung indirekt, d.h. wir nehmen

$$\frac{d(\Phi^{-1}(\delta))}{\delta^q} \to 0 \quad \text{für } \delta \to 0 \tag{4.5}$$

an. Wie im Beweis zu Proposition 4.12, wobei wir jedoch mit

$$-\xi(\tilde{u}-u^\dagger) \le \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(\tilde{u},u^\dagger) + \beta_2 r^\gamma \mathcal{S}(F(\tilde{u}),F(u^\dagger))^\kappa + d(r)$$

für $\tilde{u} \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ und $r \ge 0$ statt (4.1) beginnen, folgt für $t \in (0,t_0]$ und $r \ge 0$ zunächst

$$-\xi(u) \le \beta_1 \underbrace{\left(\tfrac{\mathcal{R}(u^\dagger+tu)-\mathcal{R}(u^\dagger)}{t} - \xi(u)\right)}_{\overset{t\to+0}{\longrightarrow} 0} + \beta_2 r^\gamma \underbrace{\left(\tfrac{\mathcal{S}(F(u^\dagger+tu),F(u^\dagger))^q}{t}\right)^{\frac{\kappa}{q}}}_{\overset{t\to+0}{\longrightarrow} L_{\mathcal{S}}^{\kappa/q}} t^{\frac{\kappa}{q}-1} + \frac{d(r)}{t}. \quad (4.6)$$

Wählen wir nun $r(t) := \Phi^{-1}(t^{\frac{1}{q}})$, d.h. es gilt $t = \Phi(r(t))^q$, so erhalten wir einerseits (unter Verwendung von (4.5))

$$\frac{d(r(t))}{t} = \frac{d(\Phi^{-1}(t^{\frac{1}{q}}))}{t} \to 0 \quad \text{für } t \to +0$$

und andererseits

$$r(t)^\gamma t^{\frac{\kappa}{q}-1} = \frac{r(t)^\gamma \Phi(r(t))^\kappa}{t} = \frac{d(r(t))}{t} \to 0 \quad \text{für } t \to +0.$$

Somit konvergieren alle Summanden in (4.6) gegen Null und wir erhalten $\xi(u) \ge 0$. Dies steht jedoch im Widerspruch zur Annahme $\xi(u) < 0$. □

5. Spezialfälle

In diesem Kapitel werden wir einige Konvergenzratenresultate aus der Literatur betrachten und zeigen, dass das im ersten Teil dieser Arbeit vorgestellte Konzept im entsprechenden Spezialfall die bekannten Konvergenzraten liefert.

5.1. Nichtlineare Operatoren zwischen Banach-Räumen

Wir betrachten die in [15] untersuchte Situation. Seien also U und V reflexive Banach-Räume mit den dazugehörigen schwachen Topologien τ_U und τ_V und sei $F : D(F) \subseteq U \to V$ ein im Allgemeinen nichtlinearer Operator. Die schwache Konvergenz in U bzw. V bezeichnen wir im weiteren Verlauf dieses Abschnitts mit „$\rightharpoonup$“; für die Normkonvergenz benötigen wir kein Symbol. Wir wählen $\mathcal{S}(v_1, v_2) := \|v_1 - v_2\|$ und $p > 0$, d.h. wir betrachten das Tikhonov-Funktional

$$\mathcal{T}_\alpha^\delta(u) = \|F(u) - v^\delta\|^p + \alpha\mathcal{R}(u)$$

mit $\|v^\delta - v^0\| \leq \delta$ für ein vorgegebenes $v^0 \in V$. Auf diesen Spezialfall waren wir bereits in Abschnitt 3.3.1 eingegangen. Im Weiteren sei stets Voraussetzung 3.8 erfüllt und $\mathcal{R}$ sei konvex.

Die Arbeit mit Subgradienten im Sinne von Definition 4.1, d.h. mit schwach stetigen linearen Funktionalen auf U wird durch die folgende Proposition erleichtert.

Proposition 5.1. *Seien W ein Banach-Raum und τ_W die schwache Topologie auf W. Ein lineares Funktional auf W ist genau dann normstetig, wenn es schwach stetig ist.*

Beweis. Die Behauptung ist trivialerweise erfüllt, da nach Definition der schwachen Konvergenz im Banach-Raum

$$w_k \rightharpoonup w \quad \Leftrightarrow \quad \zeta(w_k) \to \zeta(w) \text{ für alle } \zeta \in W'$$

gilt (W' sei der Dualraum bezüglich der Normtopologie), d.h. jedes normstetige lineare Funktional ist automatisch schwach stetig. Da in $\mathbb{R}$ schwache Konvergenz und Normkonvergenz übereinstimmen, folgt umgekehrt aus der schwachen Stetigkeit eines linearen Funktionals auch die Normstetigkeit. □

Sei im Weiteren $u^\dagger \in D_B = \{u \in D : \partial\mathcal{R}(u) \neq \emptyset\}$ stets die selbe $\mathcal{R}$-minimierende Lösung. Da relativ kompakte Teilmengen eines reflexiven Banach-Raumes beschränkt sind, existiert für jedes $\alpha > 0$ eine Konstante $K_\alpha > 0$, sodass

$$\|u - u^\dagger\| \leq K_\alpha \quad \text{für alle } u \in M_\alpha(\varrho\alpha)$$

gilt. Wir nehmen an, dass die folgenden Voraussetzungen von $u^\dagger$ erfüllt werden.

Voraussetzung 5.2. Es gilt:

(i) $D(F)$ ist sternähnlich bezüglich $u^\dagger$, d.h. für jedes $u \in D(F)$ existiert ein $t_0 > 0$, sodass $u^\dagger + t(u - u^\dagger) \in D(F)$ für jedes $t \in [0, t_0]$ gilt.

(ii) Es existiert ein beschränkter linearer Operator $F'(u^\dagger) : U \to V$, sodass

$$\left\| \frac{F(u^\dagger + t(u - u^\dagger)) - F(u^\dagger)}{t} - F'(u^\dagger)(u - u^\dagger) \right\| \to 0 \quad \text{für } t \to +0$$

für jedes $u \in D$ gilt.

Aus der Konvexität von $\mathcal{R}$ und der Sternähnlichkeit bezüglich $u^\dagger$ von $D(F)$ folgt die Sternähnlichkeit bezüglich $u^\dagger$ von D. Im Weiteren bezeichnen wir mit $F'(u^\dagger)^* : V' \to U'$ den adjungierten Operator zu $F'(u^\dagger)$. Dabei sind U' und V' die Dualräume von U und V bezüglich der Normtopologien.

Definition 5.3. Wir sagen, dass $u^\dagger$ eine *Quellbedingung* erfüllt, wenn ein $\xi \in \partial\mathcal{R}(u^\dagger)$ mit $\xi \in \operatorname{im} F'(u^\dagger)^*$ existiert. Wir sagen, dass $u^\dagger$ eine *approximative Quellbedingung* erfüllt, wenn ein $\xi \in \partial\mathcal{R}(u^\dagger)$ mit $\xi \in \overline{\operatorname{im} F'(u^\dagger)^*}$ existiert und definieren die zu $u^\dagger$ und ξ gehörende *Abstandsfunktion* $\tilde{d} : [0, \infty) \to [0, \infty)$ durch

$$\tilde{d}(r) := \min\{\|\xi - F'(u^\dagger)^*\eta\| : \eta \in V', \|\eta\| \le r\}.$$

Die Funktion $\tilde{d}$ aus Definition 5.3 ist nach [15] wohldefiniert, nicht negativ, endlich und monoton fallend. Erfüllt $u^\dagger$ eine Quellbedingung, so erfüllt $u^\dagger$ offensichtlich auch eine approximative Quellbedingung und es existiert ein $r_0 \ge 0$ mit $\tilde{d}(r) = 0$ für alle $r \ge r_0$. Im Fall $\xi \in \overline{\operatorname{im} F'(u^\dagger)^*} \setminus \operatorname{im} F'(u^\dagger)^*$ gilt stets $\tilde{d}(r) > 0$ für alle $r \ge 0$ und $\tilde{d}$ ist streng monoton fallend (vgl. [15]).

Die folgende Definition ist aus [15] übernommen, wobei wir die Handhabung der Niveaumengen des Tikhonov-Funktionals geringfügig an das Konzept dieser Arbeit angepasst haben.

Definition 5.4. Seien $c_1, c_2 \ge 0$. Der Operator F heißt nichtlinear vom Grad (c_1, c_2) bezüglich $\mathcal{R}$, $u^\dagger$ und $\xi \in \partial\mathcal{R}(u^\dagger)$, wenn Konstanten $\varrho > c_p\mathcal{R}(u^\dagger)$, $\bar{\alpha} > 0$ und $K > 0$ existieren, sodass

$$\|F(u) - F(u^\dagger) - F'(u^\dagger)(u - u^\dagger)\| \le K\|F(u) - F(u^\dagger)\|^{c_1}\mathcal{B}_\xi^{\mathcal{R}}(u, u^\dagger)^{c_2}$$

für alle $u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ gilt.

Das nächste Lemma findet man in ähnlicher Form ebenfalls in [15] (Lemma 3.1 und Formel (4.3)).

Lemma 5.5. *Die $\mathcal{R}$-minimierende Lösung $u^\dagger$ erfülle eine approximative Quellbedingung im Sinne von Definition 5.3 und F sei nichtlinear vom Grad (c_1, c_2) bezüglich $\mathcal{R}$, $u^\dagger$ und*

ξ mit $c_1 \in (0, 1 - c_2]$ und $c_2 \in [0, 1)$. Des Weiteren seien $\kappa := \frac{c_1}{1-c_2}$ und $r_0 > 0$. Dann existieren Konstanten $\beta_1 \in [0, 1)$, $\beta_2 \geq 0$ und $\gamma > 0$, sodass

$$-\xi(u - u^\dagger) \leq \beta_1 \mathcal{B}_\xi^{\mathcal{R}}(u, u^\dagger) + \beta_2 r^\gamma \|F(u) - F(u^\dagger)\|^\kappa + K_{\bar{\alpha}} \tilde{d}(r)$$

für alle $u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ und alle $r \geq r_0$ gilt. Dabei ist $\beta_1 = c_2$ und $\gamma = \frac{1}{1-c_2}$.

Beweis. Für $r \geq 0$ sei $\eta_r \in V'$ ein Element, für welches das Minimum in der Definition von $\tilde{d}(r)$ angenommen wird. Dann gilt für $u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ zunächst

$$\begin{aligned}
-\xi(u - u^\dagger) &\leq \big|\big(F'(u^\dagger)^*\eta_r + \xi - F'(u^\dagger)^*\eta_r\big)(u - u^\dagger)\big| \\
&= \big|\eta_r\big(F'(u^\dagger)(u - u^\dagger)\big) + \big(\xi - F'(u^\dagger)^*\eta_r\big)(u - u^\dagger)\big| \\
&\leq \underbrace{\|\eta_r\|}_{\leq r} \|F'(u^\dagger)(u - u^\dagger)\| + \underbrace{\|\xi - F'(u^\dagger)^*\eta_r\|}_{=\tilde{d}(r)} \underbrace{\|u - u^\dagger\|}_{\leq K_{\bar{\alpha}}} \\
&\leq r\|F(u) - F(u^\dagger) - F'(u^\dagger)(u - u^\dagger) + F(u^\dagger) - F(u)\| + K_{\bar{\alpha}} \tilde{d}(r) \\
&\leq Kr\|F(u) - F(u^\dagger)\|^{c_1} \mathcal{B}_\xi^{\mathcal{R}}(u, u^\dagger)^{c_2} + r\|F(u) - F(u^\dagger)\| + K_{\bar{\alpha}} \tilde{d}(r).
\end{aligned}$$

Wir müssen nun zwei Fälle unterscheiden:

- Fall $c_2 = 0$. Wir erhalten

$$\begin{aligned}
-\xi(u - u^\dagger) &\leq Kr\|F(u) - F(u^\dagger)\|^{c_1} + r\|F(u) - F(u^\dagger)\| + K_{\bar{\alpha}} \tilde{d}(r) \\
&= \big(Kr + r\|F(u) - F(u^\dagger)\|^{1-c_1}\big)\|F(u) - F(u^\dagger)\|^{c_1} + K_{\bar{\alpha}} \tilde{d}(r) \\
&\leq \big(K + (\varrho\bar{\alpha})^{\frac{1-c_1}{p}}\big) r\|F(u) - F(u^\dagger)\|^{c_1} + K_{\bar{\alpha}} \tilde{d}(r).
\end{aligned}$$

- Fall $c_2 \in (0, 1)$. Wir wenden die Ungleichung

$$ab \leq \frac{a^{p_1}}{p_1} + \frac{b^{p_2}}{p_2} \quad \text{für} \quad a, b \geq 0, \quad \frac{1}{p_1} + \frac{1}{p_2} = 1, \quad p_1, p_2 > 1$$

  mit

$$a := \mathcal{B}_\xi^{\mathcal{R}}(u, u^\dagger)^{c_2}, \quad b := Kr\|F(u) - F(u^\dagger)\|^{c_1}, \quad p_1 := \frac{1}{c_2}, \quad p_2 := \frac{1}{1 - c_2}$$

  an und erhalten

$$\begin{aligned}
-\xi(u - u^\dagger) &\leq Kr\|F(u) - F(u^\dagger)\|^{c_1} \mathcal{B}_\xi^{\mathcal{R}}(u, u^\dagger)^{c_2} + r\|F(u) - F(u^\dagger)\| + K_{\bar{\alpha}} \tilde{d}(r) \\
&\leq c_2 \mathcal{B}_\xi^{\mathcal{R}}(u, u^\dagger) + (1 - c_2) K^{\frac{1}{1-c_2}} r^{\frac{1}{1-c_2}} \|F(u) - F(u^\dagger)\|^{\frac{c_1}{1-c_2}} \\
&\quad + r\|F(u) - F(u^\dagger)\| + K_{\bar{\alpha}} \tilde{d}(r)
\end{aligned}$$

und weiter

$$
\begin{aligned}
&-\xi(u-u^\dagger)\\
&\quad= \Big((1-c_2)K^{\frac{1}{1-c_2}} r^{\frac{1}{1-c_2}} + r\|F(u)-F(u^\dagger)\|^{\frac{1-c_1-c_2}{1-c_2}}\Big)\|F(u)-F(u^\dagger)\|^{\frac{c_1}{1-c_2}}\\
&\qquad + c_2\mathcal{B}_\xi^{\mathcal{R}}(u,u^\dagger) + K_{\bar\alpha}\tilde d(r)\\
&\quad\le \Big((1-c_2)K^{\frac{1}{1-c_2}} + (\varrho\bar\alpha)^{\frac{1-c_1-c_2}{p(1-c_2)}} r_0^{\frac{-c_2}{1-c_2}}\Big) r^{\frac{1}{1-c_2}}\|F(u)-F(u^\dagger)\|^{\frac{c_1}{1-c_2}}\\
&\qquad + c_2\mathcal{B}_\xi^{\mathcal{R}}(u,u^\dagger) + K_{\bar\alpha}\tilde d(r). \qquad \square
\end{aligned}
$$

Satz 5.6. *Die $\mathcal{R}$-minimierende Lösung $u^\dagger$ erfülle eine approximative Quellbedingung im Sinne von Definition 5.3 und F sei nichtlinear vom Grad (c_1,c_2) bezüglich $\mathcal{R}$, $u^\dagger$ und ξ mit $c_1 \in (0, 1-c_2]$, $c_2 \in [0,1)$ und $\frac{c_1}{1-c_2} < p$. Dann erfüllt $u^\dagger$ eine approximative Ungleichung im Sinne von Definition 4.13 mit $\kappa = \frac{c_1}{1-c_2} \in (0,p)$ und es gilt*

$$d(r) \le K_{\bar\alpha}\tilde d(r) \quad \textit{für alle } r \ge r_0 > 0.$$

Beweis. Die Behauptung folgt sofort aus Lemma 5.5. $\square$

Erfüllt $u^\dagger$ eine Quellbedingung, so erfüllt $u^\dagger$ nach Satz 5.6 eine Variationsungleichung und Satz 4.10 und [15, Theorem 3.3] liefern die selbe Konvergenzrate. Analog gilt: Erfüllt $u^\dagger$ eine approximative Quellbedingung, so erfüllt $u^\dagger$ auch eine approximative Ungleichung. Gilt $d(r) > 0$ für alle $r \ge 0$, so liefern Satz 4.19 und [15, Theorem 4.3] abgesehen von eventuellen Unterschieden zwischen d und $\tilde d$ die selbe Konvergenzrate, d.h. die Struktur von Φ und Ψ in dieser Arbeit stimmt mit der Struktur von Φ und Ψ in [15] überein. Aus Proposition 4.21 folgt in Verbindung mit Satz 5.6 außerdem, dass die mit dem Konzept dieser Arbeit erhaltene Konvergenzrate mindestens so gut ist wie die Rate aus [15].

Bemerkung. In [19] und [15] wird gezeigt, dass für den Fall $c_1 = 0$ und $c_2 = 1$ aus der Quellbedingung $\xi = F'(u^\dagger)^*\eta$ mit $K\|\eta\| < 1$ eine Variationsungleichung mit $\kappa = 1$ folgt. Die Umkehrung, d.h. aus einer Variationsungleichung mit $\kappa = 1$ folgt die Quellbedingung, gilt, falls F und $\mathcal{R}$ in $u^\dagger$ Gâteaux-differenzierbar sind (siehe [29]).

Sind $\mathcal{R}$ und F in $u^\dagger$ Gâteaux-differenzierbar, so kann man zeigen, dass die Voraussetzungen von Proposition 4.12 bzw. Proposition 4.24 mit $q = 1$ erfüllt sind (vgl. [16]). Wir werden dieses Resultat weiter unten für beschränkte lineare Operatoren zwischen Hilbert-Räumen beweisen.

5.2. Nichtlineare Operatoren zwischen Hilbert-Räumen

Analog zum vorhergehenden Abschnitt können wir nichtlineare Operatoren auch zwischen Hilbert-Räumen U und V betrachten. Dies ermöglicht uns die Formulierung von allgemeineren Quellbedingungen: Erfülle F die Voraussetzung 5.2 und sei $F'(u^\dagger)^* : V \to$

U der Hilbert-Raum-Adjungierte zu $F'(u^\dagger)$. Dann verstehen wir unter einer Quellbedingung eine Darstellung der Form

$$\xi = \big(F'(u^\dagger)^* F'(u^\dagger)\big)^\mu \eta$$

für ein $\eta \in V$ und mit $\mu \in (0, 1]$. In [16] wird gezeigt, dass aus solch einer Quellbedingung eine Variationsungleichung folgt, wobei die Konstante κ von μ und dem Grad der Nichtlinearität von F abhängt. Für Details verweisen wir auf [16].

5.3. Lineare Operatoren zwischen Hilbert-Räumen

Wir betrachten einen beschränkten linearen Operator $A : U \to V$ zwischen Hilbert-Räumen U und V und das Tikhonov-Funktional

$$\mathcal{T}_\alpha^\delta(u) = \|Au - v^\delta\|^2 + \alpha\|u\|^2,$$

wobei $\|v^\delta - v^0\| \leq \delta$ und $v^0 \in \operatorname{im} A$ gelten soll. In Abschnitt 3.3.2 hatten wir diese Situation bereits betrachtet und gezeigt, dass die Sätze über Existenz, Stabilität und Konvergenz der Minimierer des Tikhonov-Funktionals aus Abschnitt 3.2 anwendbar sind. Wir stellen zusätzlich fest, dass $\mathcal{R} = \|\bullet\|^2$ konvex ist und $D_B \neq \emptyset$ gilt. Im Beispiel in Abschnitt 4.1.1 hatten wir außerdem $\mathcal{B}_\xi^{\mathcal{R}}(u, \tilde{u}) = \|u - \tilde{u}\|^2$ gezeigt.

In der Literatur (z.B. [9]) werden zum Beweis von Konvergenzraten an eine $\mathcal{R}$-minimierende Lösung $u^\dagger$ Quellbedingungen der Form

$$u^\dagger \in \operatorname{im}(A^*A)^\mu \quad \text{mit } \mu \in (0, 1]$$

vorausgesetzt. Für den Fall $\mu = 1$ erhält man dann die Rate

$$\|u_{\alpha(\delta)}^\delta - u^\dagger\|^2 = \mathcal{O}(\delta^{\frac{4}{3}}).$$

Um diese Rate mit Hilfe einer Variationsungleichung zu erreichen, müsste $\kappa = \frac{4}{3}$ gelten. Dies ist aufgrund der folgenden Proposition jedoch nicht möglich.

Proposition 5.7. *Für den betrachteten Spezialfall eines linearen Operators zwischen Hilbert-Räumen sind die Voraussetzungen von Proposition 4.12 bzw. Proposition 4.24 mit $q = 1$ erfüllt.*

Beweis. Die $\mathcal{R}$-minimierende Lösung $u^\dagger$ erfülle also eine Variationsungleichung bzw. eine approximative Ungleichung. Wir setzen

$$u := -u^\dagger \qquad \text{und} \qquad t_0 := \min\big\{1, \tfrac{\sqrt{\bar{\alpha}}}{\|A\|}\big\} > 0$$

(bzw. $t_0 := 1$, falls $A = 0$). Dann gilt

$$\xi(u) = \langle u, 2u^\dagger \rangle = -2\|u^\dagger\|^2 < 0$$

und für $t \in (0, t_0]$ gilt

$$\begin{aligned}\mathcal{T}_{\bar\alpha}^0(u^\dagger + tu) = \mathcal{T}_{\bar\alpha}^0((1-t)u^\dagger) &= t^2\|Au^\dagger\|^2 + \bar\alpha(1-t)^2\|u^\dagger\|^2 \\ &\le \big(\tfrac{\|A\|^2}{\bar\alpha}t^2 + (1-t)^2\big)\bar\alpha\|u^\dagger\|^2 \le 2\bar\alpha\|u^\dagger\|^2 < \varrho\bar\alpha,\end{aligned}$$

d.h. $u^\dagger + tu \in M_{\bar\alpha}(\varrho\bar\alpha)$. Die Grenzwerte $L_{\mathcal{R}}$ und $L_{\mathcal{S}}$ aus Proposition 4.12 existieren offensichtlich und es gilt $L_{\mathcal{R}} = \langle u, 2u^\dagger\rangle$. □

Wir haben nun also eine Situation, für die das in dieser Arbeit vorgestellte Konzept nicht die beste Konvergenzrate liefert. In [16] wird jedoch gezeigt, dass aus einer Quellbedingung $u^\dagger \in \operatorname{im}(A^*A)^\mu$ mit $\mu \in (0, \frac{1}{2})$ eine Variationsungleichung mit $\kappa = \frac{4\mu}{1+2\mu}$ folgt. Für $\mu = \frac{1}{2}$ gilt dies ebenfalls, wie wir in Abschnitt 5.1 bereits festgestellt hatten (man beachte $\operatorname{im}(A^*A)^{\frac{1}{2}} = \operatorname{im} A^*$). Somit liefert unser Konzept wenigstens für $\mu \in (0, \frac{1}{2}]$ die selben Konvergenzraten wie klassische Quellbedingungen.

Wir wollen nun noch zeigen, wann die Voraussetzungen von Proposition 4.17 im betrachteten Spezialfall erfüllt sind.

Proposition 5.8. *Sei $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung, die eine approximative Ungleichung im Sinne von Definition 4.13 erfüllt. Gilt $\beta_1 > 0$, so existiert kein $u \in U$ mit $Au = v^0$, $\|u\|^2 = \varrho$ und $\langle u - u^\dagger, 2u^\dagger\rangle = \frac{-\beta_1}{1-\beta_1}(\varrho - \|u^\dagger\|^2)$. Gilt $\beta_1 = 0$, so existiert genau dann kein solches u, wenn A injektiv ist.*

Beweis. Zunächst stellen wir fest, dass $u^\dagger \in (\ker A)^\perp$ gilt: Denn wäre dies nicht so, so würden $u_1 \in \ker A \setminus \{0\}$ und $u_2 \in (\ker A)^\perp$ mit $u^\dagger = u_1 + u_2$ existieren und mit dem Satz von Pythagoras würde $\|u^\dagger\|^2 = \|u_1\|^2 + \|u_2\|^2$, d.h. $\|u_2\|^2 < \|u^\dagger\|^2$, folgen. Da $Au_2 = A(u_1 + u_2) = Au^\dagger$ gilt, ist dies ein Widerspruch zu der Annahme, dass $u^\dagger$ eine $\mathcal{R}$-minimierende Lösung ist. Es gilt nun also

$$Au = v^0 \quad\Rightarrow\quad A(u - u^\dagger) = 0 \quad\Rightarrow\quad u - u^\dagger \in \ker A \quad\Rightarrow\quad \langle u - u^\dagger, u^\dagger\rangle = 0.$$

Im Fall $\beta_1 > 0$ führt die Existenz eines Elements $u \in U$ entsprechend der Proposition wegen

$$0 = \langle u - u^\dagger, 2u^\dagger\rangle = \underbrace{\tfrac{-\beta_1}{1-\beta_1}}_{<0}\underbrace{(\varrho - \|u^\dagger\|^2)}_{>0} < 0$$

zu einem Widerspruch.

Gilt $\beta_1 = 0$ und ist A injektiv, so ist $u^\dagger$ die einzige Lösung. Wegen $\|u^\dagger\|^2 < \varrho$ existiert also kein $u \in U$ mit den in der Proposition formulierten Eigenschaften. Ist A nicht injektiv, so existiert ein $\tilde u \in \ker A$ mit $\tilde u \neq 0$. Setzen wir $t := \frac{1}{\|\tilde u\|}\sqrt{\varrho - \|u^\dagger\|^2}$ und $u := u^\dagger + t\tilde u$, so gilt $Au = v^0$, also $\langle u - u^\dagger, u^\dagger\rangle = 0$, und somit

$$\|u\|^2 = \|u^\dagger\|^2 + \|u - u^\dagger\|^2 = \|u^\dagger\|^2 + t^2\|\tilde u\|^2 = \varrho,$$

d.h. u erfüllt die Eigenschaften aus der Proposition. □

6. Ausblick

Abschließend wollen wir einige Anmerkungen und Gedankengänge zusammentragen, die im ersten Teil dieser Arbeit bisher keine oder nur wenig Beachtung gefunden haben.

- Bei der Arbeit mit topologischen Räumen fällt auf, dass die Folgenkonvergenz relativ unnatürlich ist, im Gegensatz zu metrischen Räumen. Aus rein theoretischem Interesse wäre denkbar, die Sätze 3.3 (Existenz von Minimierern), 3.4 (Stabilität des Minimierungsproblems) und 3.7 (Konvergenz bei $\delta \to 0$) auch für die in Kapitel 2 eingeführte Netzkonvergenz zu formulieren. Aus numerischer Sicht ist dies jedoch von keinerlei Bedeutung.

- Wie in Abschnitt 3.1 bereits erwähnt, kann man statt der p-ten Potenz von $\mathcal{S}(F(\bullet), v^\delta)$ im Tikhonov-Funktional auch $f(\mathcal{S}(F(\bullet), v^\delta))$ für eine gewisse Klasse von Funktionen $f : [0, \infty] \to [0, \infty]$ betrachten. Die in dieser Arbeit verwendeten Techniken lassen sich dann jedoch nicht ohne Weiteres auf diese allgemeinere Situation übertragen. In diesem Zusammenhang sollten die Resultate in [31] genauer untersucht werden.

- Von großem Interesse sind sogenannte Converse-Results, d.h. Aussagen die von einer Konvergenzrate des Bregman-Abstandes bei $\delta \to 0$ Schlüsse auf die Erfüllung einer Variationsungleichung zulassen. In [16] wurde ein solches Resultat für Hilbert-Raum-Situationen formuliert. Jedoch wurde dort der Umweg über Quellbedingungen gewählt. Offen ist, ob auf dem direkten Weg (sofern ein solcher existiert) bessere Ergebnisse, d.h. Variationsungleichungen mit größerem κ als in [16], erreicht werden können.

- Aus numerischer Sicht sind nur Exponenten $p \geq 1$ im Tikhonov-Funktional sinnvoll. Dies ist vermutlich der Grund, weshalb auch bei theoretischen Untersuchungen bisher stets $p \geq 1$ vorausgesetzt wurde. Im Rahmen dieser Arbeit stellte sich heraus, dass die verwendeten Techniken auch den Fall $p \in (0, 1)$ abdecken. Bei der Betrachtung von Konvergenzraten für $p \in (0, 1)$ tritt der bisher unbekannte Effekt auf, dass der Exponent p (im Banach-Raum-Fall) eine obere und nicht annehmbare Schranke für κ in der Variationsungleichung darstellt und somit die mit den Techniken dieser Arbeit erzielbaren Konvergenzraten geringer als $\mathcal{O}(\delta^p)$ sind. Für $p \geq 1$ wird die erzielbare Konvergenzrate hingegen durch Proposition 4.12 begrenzt. Eine tiefer gehende Schilderung und Interpretation dieser Beobachtung findet man in [17].

- Es wurde gezeigt, dass bei Betrachtung linearer Operatoren zwischen Hilbert-Räumen die mit Hilfe der in dieser Arbeit verwendeten Form von Variationsungleichungen maximale erreichbare Konvergenzrate bei $\mathcal{O}(\delta)$ für den Bregman-Abstand liegt (vgl. Proposition 5.7). Offen ist, ob mit einem ähnlichen Konzept unter Verwendung eines anderen Typs von Variationsungleichungen die maximale Rate $\mathcal{O}(\delta^{\frac{4}{3}})$ erreicht werden kann.

II. Numerische Studie

Im zweiten Teil dieser Arbeit soll die Methode der Tikhonov-Regularisierung, die wir bisher nur aus der theoretischen Perspektive betrachtet haben, anhand eines konkreten Problems aus der Praxis demonstriert werden.

Wir begeben uns dazu auf das Gebiet der Ultraschalltechnik und stellen ein Verfahren zur Untersuchung des von Ultraschallwandlern erzeugten Schallbildes vor, die Schlierentomografie.

Nachdem wir uns einen Überblick über die Thematik verschafft haben, werden wir zunächst einige Werkzeuge bereitstellen, um im Anschluss ein mathematisches Modell zu formulieren. Ausgehend von diesem Modell werden wir uns sukzessive der praktischen Umsetzung annähern. Den Abschluss dieses Teils der Arbeit bildet die Präsentation einiger numerischer Resultate.

7. Ultraschall und Ultraschallwandler

Ultraschalltechnik hat im Laufe des 20. Jahrhunderts eine Vielzahl von Anwendungsmöglichkeiten erschlossen. Nicht nur in der Medizin zur Visualisierung des Körperinneren eines Menschen oder zur Entfernung diverser Ablagerungen wie Gallensteinen, sondern zum Beispiel auch in der Industrie zur Bearbeitung oder Qualitätsprüfung von Werkstücken und in der Seefahrt zur Messung der Meerestiefe oder zum Auffinden von Fischschwärmen ist sie kaum noch wegzudenken.

Als Ultraschall bezeichnet man Schallwellen, deren Frequenz oberhalb der menschlichen Hörgrenze von 20 kHz liegt. Erzeugt wird er durch so genannte Ultraschallwandler (auch Ultraschall-Transducer genannt), die elektrischen Wechselstrom, ähnlich wie ein Radiolautsprecher, in Schallwellen umwandeln.

Schallwellen können sich nur in Materie, also zum Beipiel in Luft oder Wasser, ausbreiten. Dabei werden die Teilchen (z.B. Wassermoleküle) abwechselnd verdichtet und „auseinandergezogen“. Wesentliche Kenngröße von Ultraschallwellen ist neben der Frequenz die Amplitude (maximale Auslenkung) der Wellen. Physikalisch nicht ganz präzise werden wir im Weiteren die Auslenkung mit dem im Transportmedium erzeugten Druck, dem Schalldruck, gleichsetzen. Dieser wechselt also stets entsprechend der Frequenz des Ultraschalls zwischen positiven und negativen Werten.

Die wesentlichen technischen Entwicklungen auf dem Gebiet des Ultraschalls fanden in der Zeit des Zweiten Weltkriegs statt, doch auch heute wird aufgrund der zahlreichen Anwendungsmöglichkeiten noch itensiv an der Weiterentwicklung der Ultraschallwandler gearbeitet. Für einige Bemerkungen zur Geschichte der Ultraschalltechnik sei auf [24] verwiesen.

Ziel bei der Konstruktion von Ultraschallwandlern ist eine möglichst vorteilhafte Form des Schallbildes, d.h. eine für die entsprechende Anwendung geeignete Druckverteilung im Transportmedium des Schalls. Bei einfachen Ultraschallwandlern setzt sich das Schallbild üblicherweise aus mehreren keulenförmigen Bereichen unterschiedlichen Schalldrucks zusammen. In der so genannten Hauptkeule ist der Druck (genauer: der maximale Druck) am höchsten, dies ist der erwünschte Anteil am erzeugten Schall. Aufgrund physikalischer Effekte entstehen jedoch neben der Hauptkeule mehrere Nebenkeulen, in denen der Schalldruck zwar schwächer ist als in der Hauptkeule, die aber zu Störeffekten beim Einsatz des Schallwandlers führen können. So ist es zum Beispiel möglich, dass die Nebenkeulen bei auf Echoeffekten beruhenden Bildgebungsverfahren zu Bildfehlern führen. Kurze, auf das Wesentliche konzentrierte Erläuterungen zur Problematik der Nebenkeulen und allgemeiner zur Ausbreitung von Ultraschallwellen findet man in [14].

Seit einigen Jahren werden verstärkt Gruppenstrahler (auch als Phased-Array-Wandler bezeichnet) zur Erzeugung von Ultraschall eingesetzt. Dies sind Ultraschallwandler, die aus mehreren kleinen Wandlern zusammengesetzt sind. Durch geschickte, computergestützte Steuerung können mit Gruppenstrahlern komplexe Schallbilder erzeugt werden, die für viele Anwendungen besser geeignet sind als herkömmliche Ultraschallwandler, siehe [35].

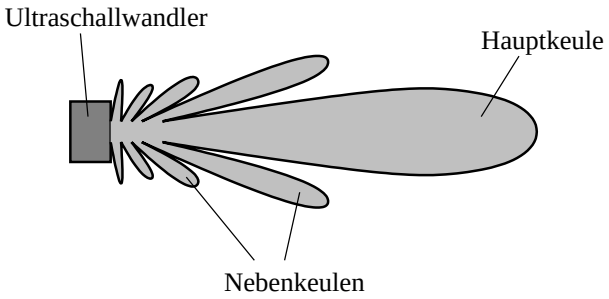

Abbildung 7.1.: Beispiel für das von einem Ultraschallwandler erzeugte Schallbild mit Hauptkeule und Nebenkeulen.

Neben den technischen Anforderungen müssen aus Sicherheitsgründen auch rechtliche Vorgaben bei der Konstruktion von Ultraschallwandlern beachtet werden. Nicht nur in der Medizin, sondern auch beim Einsatz in der Industrie halten sich Menschen längere Zeit in der Umgebung von Ultraschallquellen auf, sodass zur Vermeidung von Gesundheitsrisiken durch dauerhafte oder zu starke Bestrahlung mit Ultraschallwellen eine genaue Kenntnis des Schallbildes erforderlich ist.

In der Praxis werden die Konstruktionsparameter ausgehend vom gewünschten Schallbild durch Berechnungen bestimmt. Die Überprüfung des Schallbildes erfolgt dann üblicherweise durch Messung des Schalldrucks an jedem Punkt eines räumlichen Rasters. Dieses Verfahren ist sehr zeitaufwändig und fehleranfällig. Einen anderen, deutlich schnelleren und zugleich präziseren Weg zur Kontrolle und Bewertung von Ultraschallwandlern bietet die Schlierentomografie.

8. Prinzip der Schlierentomografie

Im Jahr 1864 entwickelte der deutsche Chemiker und Physiker August Toepler eine Fotografiemethode, die so genannte Schlierenfotografie, mit der Dichteunterschiede in Gasen und Flüssigkeiten sichtbar gemacht werden konnten (siehe [20]). Theoretisch unterlegt wurde dieses Verfahren 1935 durch die Raman-Nath-Theorie, die den Einfluss von Ultraschallwellen auf die Ausbreitung von Licht in Wasser beschreibt.

Im Wesentlichen besteht ein Schlierensystem aus einer punktförmigen Lichtquelle (z.B. Laser), einem zylindrischen, aufrecht stehenden Wassertank und einer Kamera, sowie einigen Linsen oder Hohlspiegeln. Das von der Lichtquelle abgegebene Licht wird durch eine Linse zu einem Bündel von parallelen Lichtstrahlen (wir verzichten hier zu Gunsten der Anschaulichkeit auf die physikalisch präzise Beschreibung) geformt, welches in horizontaler Richtung durch den Wassertank geleitet wird. Auf der Ober- oder Unterseite des Tanks ist ein Ultraschallwandler montiert; auf der gegenüberliegenden Seite eine Einrichtung zur Absorption der vom Wandler erzeugten Ultraschallwellen (vgl. Abbildung 8.1). Der Ultraschall verursacht Druck- und somit Dichteunterschiede im Wasser, die zur Beugung des durchlaufenden Lichts führen. Beim Austritt aus dem Wassertank wird das Licht durch eine Linse fokusiert und der nicht gebeugte Anteil, d.h. der Anteil des austretenden Lichts, der noch die selbe Richtung wie vor dem Durchlaufen des Tanks besitzt, herausgefiltert. Das verbleibende Licht trifft auf die Kamera und erzeugt das Schlierenbild. Je stärker das Licht in einem bestimmten Bereich des Wassertanks gebeugt wird, desto heller erscheint die entsprechende Region auf dem Bild. Zur präzisen Beschreibung der physikalischen Vorgänge sei auf [22] und [13] verwiesen.

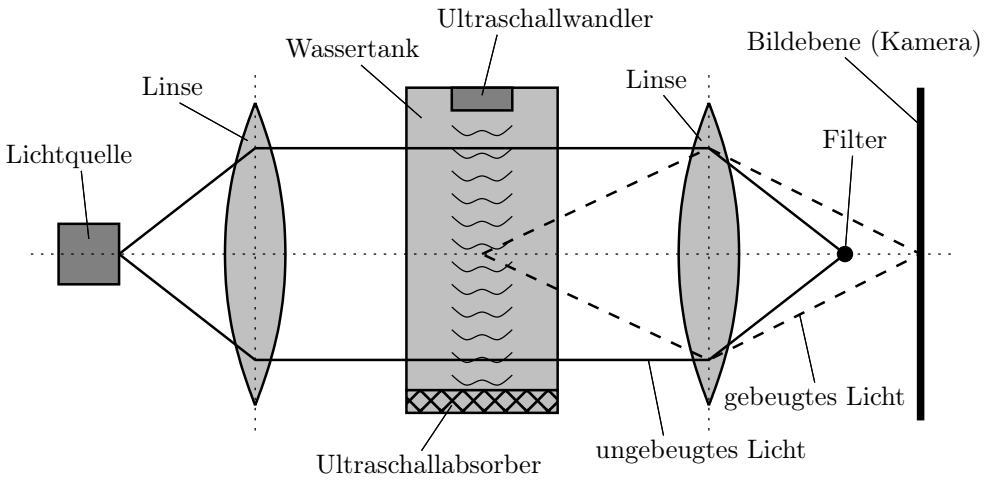

Abbildung 8.1.: Schematische Darstellung eines einfachen Schlierensystems.

Man kann zeigen, dass die Helligkeit eines Punktes auf dem Schlierenbild proportional zum Quadrat eines gewissen Integrals ist; integriert wird über den Druck entlang

des (nicht gebeugten) Lichtstrahls, der zum entsprechenden Bildpunkt gehört. Diese Beobachtung stellt die Grundlage für die Schlierentomografie dar, mit deren Hilfe man die Druckverteilung und damit die Ausbreitungseigenschaften der Ultraschallwellen im Wassertank rekonstruieren kann.

Analog zur Computertomografie, bei der (vereinfacht gesagt) Röntgen-Bilder aus verschiedenen Richtungen aufgenommen werden, um daraus ein dreidimensionales Bild des Körperinneren zu berechnen, werden bei der Schlierentomografie mehrere Schlierenfotografien aus verschiedenen Richtungen aufgenommen, um daraus die Druckverteilung im Tank zu ermitteln.

Die Berechnung der dreidimensionalen Druckverteilung (genauer: von zweidimensionalen Schnittbildern der dreidimensionalen Druckverteilung) aus den aufgenommenen Schlierenfotografien soll Gegenstand der folgenden Kapitel sein.

An dieser Stelle sei angemerkt, dass die beschriebene Technik die einfachste Ausführung eines Schlierensystems darstellt. In der Praxis verwendete Systeme enthalten komplexe elektronische Bauteile, die eine äußerst genaue Synchronisation von Lichtquelle, Ultraschallwandler und Kamera ermöglichen, um zum Beispiel Bilder von sehr kurzen Wellensequenzen oder so genannten Stoßwellen zu erzeugen.

9. Mathematische Modellierung

Wir werden nun ein Modell zur mathematischen Behandlung der Schlierentomografie aufstellen. Dabei beschränken wir uns auf die Rekonstruktion der Druckverteilung im Wassertank entlang einer horizontalen Ebene. Soll die Druckverteilung im gesamten Tank, d.h. in drei Dimensionen, berechnet werden, so muss das im Weiteren entwickelte Verfahren für mehrere horizontale Ebenen angewendet werden. Aus den Einzelergebnissen kann dann eine dreidimensionale Darstellung zusammengesetzt werden. Bei der Modellierung orientieren wir uns an [12].

Sei $\Omega := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$ die offene Einheitskreisscheibe im $\mathbb{R}^2$ und sei $H > 0$ die Höhe des Wassertanks. Wir beschreiben den Inhalt des Tanks als $\Omega \times (0, H) \subset \mathbb{R}^3$. Mit $h \in (0, H)$ bezeichnen wir das Niveau im Tank, für welches wir die Druckverteilung rekonstruieren möchten. Die Druckverteilung sei eine reellwertige Funktion u auf $\mathbb{R}^2$ mit $\operatorname{supp} u \subseteq \overline{\Omega}$. Zur Rekonstruktion seien $N \in \mathbb{N}$ Schlierenfotografien gegeben. Da wir uns hier auf eine horizontale Ebene einschränken, wird von jedem Bild nur die zur Höhe h gehörende „Zeile“ benötigt. Diese modellieren wir als reellwertige Funktion v_i, $i = 1, \ldots, N$, auf der reellen Achse mit $\operatorname{supp} v_i \subseteq [-1, 1]$. Mit $\sigma_i \in \mathbb{S}^1 := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ bezeichnen wir entsprechend einen Einheitsvektor, der parallel zur Bildebene verläuft (vgl. Abb. 9.1), und wir setzen $\sigma_i^\perp := (-\sigma_i^2, \sigma_i^1)$, wobei die oberen Indizes die Vektorkomponenten bezeichnen.

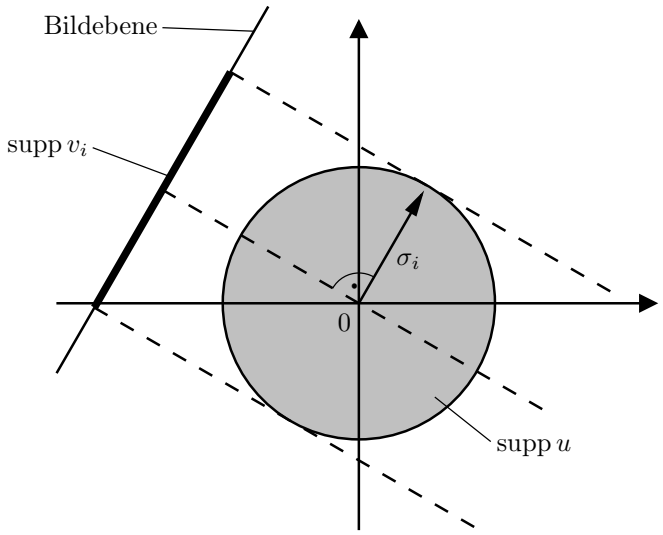

Abbildung 9.1.: Lagebeziehung zwischen dem Vektor σ_i und der Bildebene.

Setzen wir $u \in L^1(\Omega)$ voraus, so kann das direkte Problem der Schlierentomografie, d.h. die Bestimmung der v_i aus der Kenntnis von u, durch die folgende Familie von

Abbildungen beschrieben werden:

$$u \mapsto \left(\int_{-1}^{1} u(\bullet\sigma_i + q\sigma_i^{\perp}) \, \mathrm{d}q \right)^2, \quad i = 1, \ldots, N.$$

Nach dem Satz von Fubini existieren diese Integrale und die durch die rechte Seite gegebenen Funktionen besitzen einen in $[-1, 1]$ enthaltenen Träger.

Das inverse Problem besteht entsprechend darin, aus der Kenntnis der v_i die Druckverteilung u zu bestimmen. Bevor wir dies mathematisch exakt in geeigneten Unterräumen von $L^1(\Omega)$ formulieren, benötigen wir noch einige Werkzeuge.

Um Missverständnisse zu vermeiden, sei an dieser Stelle darauf hingewiesen, dass die oben eingeführten Bezeichnungen H und h nicht mehr benötigt werden und in späteren Kapiteln mit einer anderen Bedeutung belegt sein können.

10. Wavelets

In diesem Kapitel werden wir einige Aspekte aus der Theorie der Wavelets vorstellen. Dabei beschränken wir den Umfang der Ausführungen auf das für die numerische Lösung des betrachteten inversen Problems notwendige Minimum. Für tiefer gehende mathematische Abhandlungen der Wavelet-Theorie seien die Bücher [6] und [23] empfohlen. Anwendungsorientierte Einführungen bieten [1], [11] und [30].

Die Grundidee der Wavelet-Theorie ist, ähnlich zur Fourier-Analysis, die Zerlegung von Funktionen in kleine „Bausteine“, welche durch Verschiebung und Streckung aus einigen wenigen „Grundbausteinen“ entstehen. Mittels solcher Zerlegungen können Funktionen untersucht und zweckmäßig verändert werden. Letzteres spielt besonders in der Signal- und Bildverarbeitung eine wichtige Rolle.

Wir werden zunächst den Begriff des Wavelets, die kontinuierliche Wavelet-Transformation und orthogonale Wavelet-Basen einführen. Im Anschluss konstruieren wir die Familie der so genannten Daubechies-Wavelets und erläutern an diesem Beispiel den Algorithmus der schnellen Wavelet-Transformation. Nachdem die grundlegenden Gedanken im eindimensionalen Fall dargestellt wurden, werden wir diese auf zwei Dimensionen übertragen.

10.1. Grundbegriffe

Definition 10.1. Ein *Wavelet* (auch Mutter-Wavelet genannt) ist eine Funktion $\psi : \mathbb{R} \to \mathbb{R}$, die der Bedingung

$$\int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} \,\mathrm{d}\omega < \infty$$

genügt.

Dabei bezeichnet $\hat{\psi}$ die Fourier-Transformierte von ψ, d.h.

$$\hat{\psi}(\omega) := \int_{-\infty}^{\infty} \psi(x) e^{-\mathrm{i}x\omega} \,\mathrm{d}x.$$

Allgemeiner kann man auch komplexwertige Wavelets betrachten, wir beschränken uns hier jedoch auf den reellen Fall.

Beispiel. Die Funktion

$$\psi(x) = \begin{cases} 1, & x \in [0, \frac{1}{2}), \\ -1, & x \in [\frac{1}{2}, 1), \\ 0, & \text{sonst} \end{cases}$$

ist ein Wavelet, das *Haar-Wavelet*.

Durch Verschiebung und Streckung des Wavelets erhalten wir eine Vielzahl weiterer Funktionen. Für $s \neq 0$ und $t \in \mathbb{R}$ definieren wir die Funktionen $\psi_{(s,t)} : \mathbb{R} \to \mathbb{R}$ durch

$$\psi_{(s,t)}(x) := \sqrt{|s|}\psi(sx - t).$$

Das Wavelet wird also um t entlang der reellen Achse verschoben und anschließend um den Faktor s gestaucht. Der Faktor $\sqrt{|s|}$ dient der Normierung.

Definition 10.2. Sei $\psi \in L^2(\mathbb{R})$ ein Wavelet. Für $f \in L^2(\mathbb{R})$ heißt die durch

$$(W_\psi f)(s,t) := \int_{-\infty}^{\infty} f(x)\psi_{(s,t)}(x)\,\mathrm{d}x$$

definierte Funktion $W_\psi f : (\mathbb{R} \setminus \{0\}) \times \mathbb{R} \to \mathbb{R}$ *Wavelet-Transformierte* von f bezüglich ψ und die auf $L^2(\mathbb{R})$ definierte Abbildung W_ψ heißt *Wavelet-Transformation* bezüglich ψ.

Man kann zeigen, dass aus der Wavelet-Transformierten einer Funktion die ursprüngliche Funktion vollständig rekonstruiert werden kann (siehe [30]). Es genügt sogar die Kenntnis der Wavelet-Transformierten an geeignet gewählten, abzählbar vielen Punkten. Dies werden wir im weiteren Verlauf dieses Kapitels präzisieren.

Definition 10.3. Eine *Multi-Skalen-Analysis* in $L^2(\mathbb{R})$ ist ein Folge $V_0 \subseteq V_1 \subseteq V_2 \subseteq \cdots \subseteq L^2(\mathbb{R})$ von abgeschlossenen Unterräumen von $L^2(\mathbb{R})$, die die folgenden Bedingungen erfüllt:

(i) $\overline{\bigcup_{j\in\mathbb{N}_0} V_j} = L^2(\mathbb{R})$,

(ii) $f \in V_0 \Leftrightarrow f(\bullet - k) \in V_0$ für alle $k \in \mathbb{Z}$,

(iii) $f \in V_0 \Leftrightarrow f(2^j \bullet) \in V_j$ für alle $j \in \mathbb{N}_0$,

(iv) es existiert ein $\varphi \in V_0$, sodass $\{\varphi(\bullet - k) : k \in \mathbb{Z}\}$ eine Orthonormalbasis von V_0 ist.

Die Funktion φ heißt *Skalierungsfunktion* (oder auch Vater-Wavelet).

Bei der Definition der Multi-Skalen-Analysis stützen wir uns auf [34], um später die Charakterisierung der Besov-Räume mittels Wavelets ohne Änderungen übernehmen zu können. Die meisten Autoren betrachten allerdings doppelt unendliche Folgen von Teilräumen ($j \in \mathbb{Z}$); dies führt aber im Wesentlichen zu denselben Aussagen.

Setzen wir

$$\varphi_{j,k}(x) := 2^{\frac{j}{2}}\varphi(2^j x - k), \quad x \in \mathbb{R},$$

für $k \in \mathbb{Z}$ und $j \in \mathbb{N}_0$, so sieht man leicht, dass $\{\varphi_{j,k} : k \in \mathbb{Z}\}$ für jedes $j \in \mathbb{N}_0$ eine Orthonormalbasis von V_j ist.

Mit W_j bezeichnen wir im Folgenden stets das orthogonale Komplement von V_j in V_{j+1}. Kernpunkt der Wavelet-Theorie ist nun die Frage, ob es ein Wavelet $\psi \in W_0$ gibt, sodass $\{\psi(\bullet - k) : k \in \mathbb{Z}\}$ eine Orthonormalbasis in W_0 bildet. Dann wäre mit

$$\psi_{j,k}(x) := 2^{\frac{j}{2}} \psi(2^j x - k), \quad x \in \mathbb{R},$$

für $k \in \mathbb{Z}$ und $j \in \mathbb{N}_0$ durch $\{\psi_{j,k} : k \in \mathbb{Z}\}$ für jedes $j \in \mathbb{N}_0$ eine Orthonormalbasis in W_j gegeben. Diese Frage kann unter geeigneten Voraussetzungen an φ positiv beantwortet werden; wir erhalten dann die folgende Aussage.

Satz 10.4. *Für jedes $\underline{j} \in \mathbb{N}_0$ ist $\{\varphi_{\underline{j},k} : k \in \mathbb{Z}\} \cup \{\psi_{j,k} : k \in \mathbb{Z}, j \geq \underline{j}\}$ eine Orthonormalbasis in $L^2(\mathbb{R})$, d.h. für $f \in L^2(\mathbb{R})$ gilt*

$$f = \sum_{k \in \mathbb{Z}} c_k^{\underline{j}} \varphi_{\underline{j},k} + \sum_{j \geq \underline{j}} \sum_{k \in \mathbb{Z}} d_k^j \psi_{j,k} \tag{10.1}$$

mit

$$c_k^j = \int_{-\infty}^{\infty} f(x) \varphi_{j,k}(x) \, \mathrm{d}x \qquad \textit{und} \qquad d_k^j = \int_{-\infty}^{\infty} f(x) \psi_{j,k}(x) \, \mathrm{d}x.$$

Die Vorteile der Darstellung von Funktionen mittels orthonormaler Wavelet-Basen werden wir später im Zusammenhang mit der konkreten Formulierung des inversen Problems der Schlierentomografie diskutieren. An dieser Stelle soll zunächst nur auf die grundlegende Interpretation der Darstellung (10.1) eingegangen werden:

Der große Vorteil der Darstellung einer Funktion mit Hilfe einer Wavelet-Basis wie in Satz 10.4 ist die Zerlegung der Funktion in unterschiedlich große Details. Die Koeffizienten $c_k^{\underline{j}}$ repräsentieren eine sehr grobe Approximation der Funktion. Die Koeffizienten d_k^j hingegen spiegeln die Details, d.h. die lokalen Eigenschaften der Funktion, wider. Mit wachsendem j wird der Einfluss der entsprechenden d_k^j aufgrund des (für großes j) kleinen Trägers von ψ immer geringer, sodass eine endlichdimensionale Approximation in einer Wavelet-Basis (durch Weglassen der Details) deutlich bessere Ergebnisse erwarten lässt als eine endlichdimensionale Approximation in einer anderen Basis (z.B. Fourier-Basis).

10.2. Daubechies-Wavelets

In der Signal- und Bildverarbeitung sind besonders Wavelets von Bedeutung, die sowohl im Zeit- als auch im Frequenzbereich gut lokalisiert sind (siehe [30]). Eine ganze Familie von Wavelets mit überwiegend positiven Eigenschaften stellen die Daubechies-Wavelets dar, die in den 80er-Jahren von Ingrid Daubechies beschrieben wurden (siehe [6]). Diese Wavelet-Familie wird durch einen Parameter $r \in \mathbb{N}_0$ parametrisiert, der die Glattheit angibt. Das Daubechies-Wavelet der Ordnung $r \geq 1$ ist $(r-1)$-mal stetig differenzierbar und das Daubechies-Wavelet der Ordnung $r = 0$ ist das Haar-Wavelet, also stückweise konstant. Im Zeitbereich besitzen die Daubechies-Wavelets einen kompakten Träger.

Bevor wir zur Konstruktion der Daubechies-Wavelets kommen, wollen wir gleich den Zusammenhang zum Begriff der Multi-Skalen-Analysis formulieren.

Satz 10.5. *Seien φ und ψ die nachfolgend konstruierten Funktionen. Dann ist durch $V_0 := \overline{\operatorname{span}\{\varphi(\bullet - k) : k \in \mathbb{Z}\}}$ und die, entsprechend Definition 10.3, dadurch eindeutig bestimmten Unterräume V_j, $j \geq 1$, von $L^2(\mathbb{R})$ eine Multi-Skalen-Analysis mit Skalierungsfunktion φ gegeben und ψ ist ein passendes Wavelet zu dieser Multi-Skalen-Analysis, d.h. $\{\psi(\bullet - k) : k \in \mathbb{Z}\}$ ist eine Orthonormalbasis in W_0.*

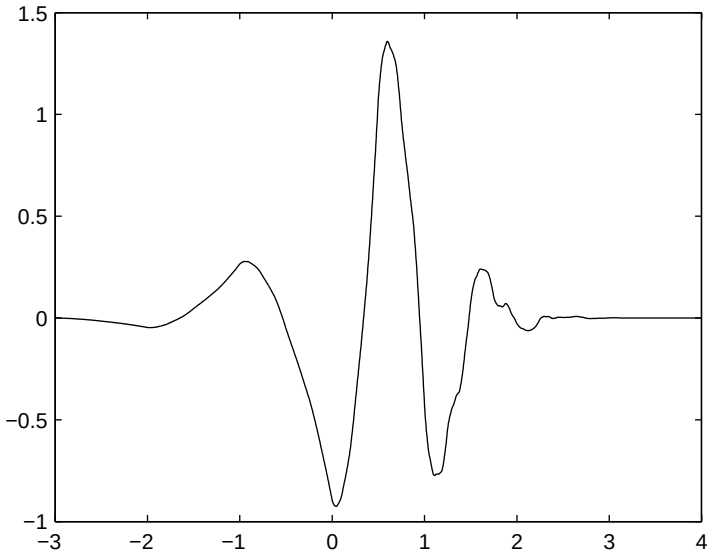

Abbildung 10.1.: Daubechies-Wavelet der Ordnung $r = 3$.

Mit Ausnahme des Falles $r = 0$ sind die Daubechies-Wavelets nicht als geschlossener Ausdruck darstellbar. Man kann sie jedoch beliebig genau durch ein iteratives Verfahren approximieren. Der nachfolgend beschriebene Ansatz kann auch allgemeiner angewendet werden, wir werden uns hier jedoch auf Daubechies-Wavelets beschränken. Wir folgen der Beschreibung in [30].

Ausgangspunkt für die Konstruktion des Daubechies-Wavelets der Ordnung r ist eine doppelt unendliche Folge $h = (h_k)_{k\in\mathbb{Z}}$, die nur endlich viele Nicht-Null-Elemente besitzt. Diese Elemente werden wir aus einer Tabelle in [6] und zum Vergleich auch aus [11] entnehmen. Alternativ können die Folgenelemente als Nullstellen gewisser Polynome numerisch berechnet werden (siehe [1]). Der Vollständigkeit halber geben wir die Werte in Tabelle II.1 nochmals an. Zunächst wird aus der Folge h eine Approximation der Skalierungsfunktion φ berechnet und daraus anschließend eine Approximation des Wavelets ψ.

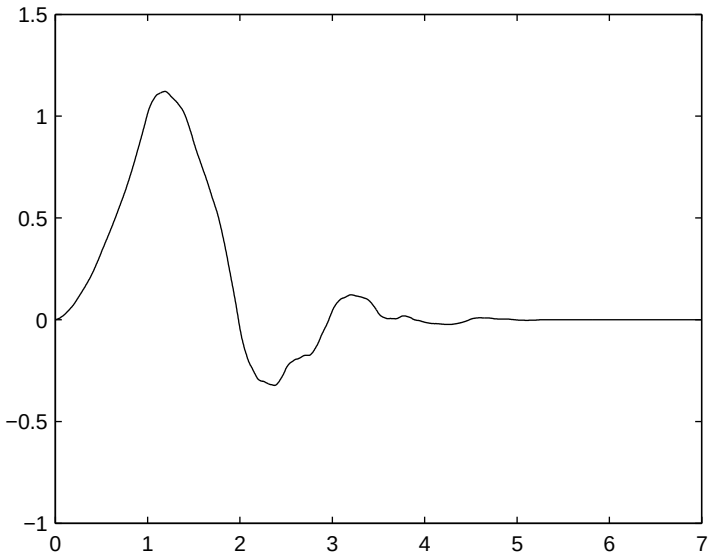

Abbildung 10.2.: Daubechies-Skalierungsfunktion der Ordnung $r = 3$.

Auf dem linearen Raum aller doppelt unendlichen Folgen führen wir den Operator H^* durch

$$H^*a := h * (\uparrow_2 a)$$

ein. Dabei ist $\uparrow_2$ der so genannte Up-Sampling-Operator, d.h.

$$(\uparrow_2 a)_k := \begin{cases} a_{\frac{k}{2}}, & k \text{ gerade}, \\ 0, & k \text{ ungerade}, \end{cases}$$

und $a * b$ mit

$$(a * b)_k := \sum_{l \in \mathbb{Z}} a_{k-l} b_l$$

bezeichnet die Faltung zweier Folgen a und b. Die Folge H^*a ist also durch

$$(H^*a)_k = \sum_{l \in \mathbb{Z}} h_{k-l} (\uparrow_2 a)_l = \sum_{l \in \mathbb{Z}} h_{k-2l} (\uparrow_2 a)_{2l} = \sum_{l \in \mathbb{Z}} h_{k-2l} a_l$$

gegeben. Die Konvergenz der Reihe ist gesichert, da h nur endlich viele Nicht-Null-Elemente besitzt.

$r = 0$		$r = 4$	
h_0	0,7071067811865475	h_0	0,1601023979741929
h_1	0,7071067811865475	h_1	0,6038292697971895
$r = 1$		h_2	0,7243085284377726
h_0	0,4829629131445341	h_3	0,1384281459013203
h_1	0,8365163037378077	h_4	−0,2422948870663823
h_2	0,2241438680420134	h_5	−0,0322448695846381
h_3	−0,1294095225512603	h_6	0,0775714938400459
$r = 2$		h_7	−0,0062414902127983
h_0	0,3326705529500825	h_8	−0,0125807519990820
h_1	0,8068915093110924	h_9	0,0033357252854738
h_2	0,4598775021184914	$r = 5$	
h_3	−0,1350110200102546	h_0	0,1115407433501095
h_4	−0,0854412738820267	h_1	0,4946238903984533
h_5	0,0352262918857095	h_2	0,7511339080210959
$r = 3$		h_3	0,3152503517091982
h_0	0,2303778133088964	h_4	−0,2262646939654400
h_1	0,7148465705529154	h_5	−0,1297668675672625
h_2	0,6308807679298587	h_6	0,0975016055873225
h_3	−0,0279837694168599	h_7	0,0275228655303053
h_4	−0,1870348117190931	h_8	−0,0315820393174862
h_5	0,0308413818355607	h_9	0,0005538422011614
h_6	0,0328830116668852	h_{10}	0,0047772575109455
h_7	−0,0105974017850690	h_{11}	−0,0010773010853085

Tabelle II.1.: Folgen h zur Beschreibung der Daubechies-Wavelets der Ordnungen 0 bis 5 (nicht angegebene Folgenelemente sind Null)

Wir berechnen aus der durch

$$e_k^0 := \begin{cases} 1, & k = 0, \\ 0, & k \neq 0 \end{cases}$$

gegebenen Startfolge e^0 nun iterativ die Folgen

$$e^n := \sqrt{2} H^* e^{n-1}, \quad n = 1, 2, \ldots,$$

und definieren entsprechend die stückweise konstanten Funktionen $\tilde{e}_n$ durch

$$\tilde{e}_n(x) := \sum_{k \in \mathbb{Z}} e_k^n \chi_{I_k^n}(x),$$

wobei $I_k^n := [\frac{k}{2^n} - \frac{1}{2^{n+1}}, \frac{k}{2^n} + \frac{1}{2^{n+1}})$ gesetzt wurde und $\chi_{I_k^n}$ die charakteristische Funktion des Intervalls I_k^n bezeichnet.

Die Funktionen $\tilde{e}_n$ sind offensichtlich in $L^2(\mathbb{R})$ und nach [6] konvergiert die Funktionenfolge $(\tilde{e}_n)_{n\in\mathbb{N}_0}$ in $L^2(\mathbb{R})$. Die Grenzfunktion ist die Daubechies-Skalierungsfunktion φ der entsprechenden Ordnung. Man kann sogar zeigen, dass diese Konvergenz für $r \geq 1$ auch eine punktweise Konvergenz ist. Für $r = 0$ (Haar-Wavelet) gilt punktweise Konvergenz in $\mathbb{R} \setminus \{0, 1\}$. Bereits für $n = 8$ erhält man eine hinreichend gute Approximation der Skalierungsfunktion (Länge der Intervalle I_k^n: $\frac{1}{256}$).

Aus der Skalierungsfunktion φ kann mittels

$$\psi(x) := \sqrt{2} \sum_{k\in\mathbb{Z}} g_k \varphi(2x - k) \approx \sqrt{2} \sum_{k\in\mathbb{Z}} g_k \tilde{e}_n(2x - k)$$

das Daubechies-Wavelet berechnet werden, wobei die Folge g durch

$$g_k := (-1)^k h_{1-k}, \quad k \in \mathbb{Z},$$

gegeben ist. Die Folge g besitzt wie h nur endlich viele Nicht-Null-Elemente.

Aus der beschrieben Konstruktion von φ und ψ kann man durch elementare Berechnungen leicht ableiten, dass

$$\operatorname{supp}\varphi \subseteq [0, 2r + 1] \qquad \text{und} \qquad \operatorname{supp}\psi \subseteq [-r, r + 1]$$

gelten.

10.3. Die schnelle Wavelet-Transformation

Seien φ und ψ die Skalierungsfunktion und das Wavelet der Ordnung $r \geq 0$ nach Daubechies. Die Funktionen $\varphi_{j,k}$, $\psi_{j,k}$ und f sowie die Koeffizienten c_k^j und d_k^j und der Index $\underline{j}$ seien wie in Abschnitt 10.1 eingeführt.

Wir sind nun für gegebenes f an der schnellen Berechnung der Koeffizientenfolgen $c^{\underline{j}}$ und d^j, $j \geq \underline{j}$, sowie an der schnellen Rekonstruktion von f aus diesen Koeffizienten interessiert.

Wir nehmen zur Vereinfachung $\operatorname{supp} f \subseteq [-1, 1]$ an. Dann sind in jeder Koeffizientenfolge nur noch endlich viele Elemente von Null verschieden, denn es gilt

$$\begin{aligned}
c_k^j \neq 0 \quad &\Rightarrow \quad \operatorname{meas}(\operatorname{supp} f \cap \operatorname{supp} \varphi_{j,k}) > 0 \\
&\Rightarrow \quad \operatorname{meas}\big([-1, 1] \cap [2^{-j}k, 2^{-j}(k + 2r + 1)]\big) > 0 \\
&\Leftrightarrow \quad \max\{-1, 2^{-j}k\} < \min\{1, 2^{-j}(k + 2r + 1)\} \\
&\Leftrightarrow \quad -1 < 2^{-j}(k + 2r + 1) \text{ und } 2^{-j}k < 1 \\
&\Leftrightarrow \quad -2^j - 2r - 1 < k < 2^j \quad \Leftrightarrow \quad -2^j - 2r \leq k \leq 2^j - 1
\end{aligned}$$

und analog

$$\begin{aligned} d_k^j \neq 0 \quad &\Rightarrow \quad \max\{-1, 2^{-j}(k-r)\} < \min\{1, 2^{-j}(k+r+1)\} \\ &\Leftrightarrow \quad -1 < 2^{-j}(k+r+1) \text{ und } 2^{-j}(k-r) < 1 \\ &\Leftrightarrow \quad -2^j - r - 1 < k < 2^j + r \quad \Leftrightarrow \quad -2^j - r \leq k \leq 2^j + r - 1. \end{aligned}$$

Desweiteren beschränken wir die Anzahl der Koeffizientenfolgen, indem wir ein $\overline{j} \geq \underline{j}$ vorgeben und annehmen, dass die Folgen d^j mit $j \geq \overline{j}$ Nullfolgen sind. In der Sprache der Multi-Skalen-Analysis setzen wir also $f \in V_{\overline{j}}$ voraus. Dies ist aus praktischen Gesichtspunkten zulässig, da die Koeffizientenfolgen d^j für große j kleine Details von f widerspiegel, die bei numerischen Berechnungen ohnehin nicht berücksichtigt werden können (vgl. Ende von Abschnitt 10.1).

10.3.1. Bestimmung der Koeffizienten

Um die gesuchten Koeffizientenfolgen $c^{\underline{j}}$ und d^j, $j = \underline{j}, \ldots, \overline{j} - 1$, zu erhalten, berechnen wir zunächst

$$c_k^{\overline{j}} = \int_{-\infty}^{\infty} f(x) 2^{\frac{\overline{j}}{2}} \varphi(2^{\overline{j}} x - k) \, \mathrm{d}x = 2^{-\frac{\overline{j}}{2}} \int_0^{2r+1} f\big(2^{-\overline{j}}(x+k)\big) \varphi(x) \, \mathrm{d}x$$

für $k = -2^{\overline{j}} - 2r, \ldots, 2^{\overline{j}} - 1$. Dabei haben wir die Kompaktheit des Trägers von φ verwendet. In der Praxis wird meist versucht, auf Integrale zu verzichten. In [1] findet man eine Diskussion, unter welchen Voraussetzungen entsprechende Vereinfachungen hinnehmbar sind. In allen Fällen führt der Verzicht auf die Berechnung der Integrale zu Fehlern, die wir hier vermeiden möchten, um die numerische Untersuchung des inversen Problems der Schlierentomografie nicht durch unnötig fehlerhafte „Nebenalgorithmen“ zur Handhabung der Wavelets zu verfälschen.

Für den nächsten Berechnungsschritt benötigen wir zwei Operatoren auf dem linearen Raum der doppelt unendlichen Folgen. Für eine Folge $a = (a_k)_{k \in \mathbb{Z}}$ bezeichnen wir zunächst mit a^{-1} die durch

$$a_k^{-1} := a_{-k}$$

definierte Folge und mit $\downarrow_2$ den durch

$$(\downarrow_2 a)_k := a_{2k}$$

gegebenen Down-Sampling-Operator. Wir führen die Operatoren H und G mittels

$$Ha := \downarrow_2 (h^{-1} * a) \qquad \text{und} \qquad Ga := \downarrow_2 (g^{-1} * a)$$

ein, wobei h und g die Folgen aus Abschnitt 10.2 sind. Wir haben also

$$(Ha)_k = (h^{-1} * a)_{2k} = \sum_{l \in \mathbb{Z}} h^{-1}_{2k-l} a_l = \sum_{k \in \mathbb{Z}} h_{l-2k} a_l$$

und eine analoge Formel für G.

Man kann nun zeigen, dass mit diesen beiden Operatoren die Beziehungen

$$c^{j-1} = Hc^j \qquad \text{und} \qquad d^{j-1} = Gc^j$$

für $j = \overline{j}, \ldots, \underline{j}+1$ gelten, d.h. wir können die gesuchten Koeffizienten sukzessive aus den $c^{\overline{j}}_k$ berechnen. Numerisch ist dies leicht umzusetzen, da alle beteiligten Folgen nur endlich viele Nicht-Null-Elemente besitzen. Die Faltung zweier Folgen kann zum Beispiel mit Hilfe der schnellen Fourier-Transformation berechnet werden (siehe [21, Befehl `conv`]).

10.3.2. Rekonstruktion der Funktion

Zur Rekonstruktion der Funktion f aus den Koeffizientenfolgen $c^{\underline{j}}$ und d^j, $j = \underline{j}, \ldots, \overline{j} - 1$, benötigen wir den im Abschnitt 10.2 eingeführten Operator H^* und den analog dazu definierten Operator G^*, also

$$(G^* a)_k = \sum_{l \in \mathbb{Z}} g_{k-2l} a_l.$$

Zur Bezeichnung der verwendeten Operatoren sei angemerkt, dass die Operatoren H^* und G^* bei Betrachtung auf Folgen aus $l^2(\mathbb{Z})$ gerade die adjungierten Operatoren (im Sinne des Hilbert-Raums) von H und G sind.

Die Funktion f erhalten wir nun durch sukzessive Anwendung der Beziehung

$$c^{j+1} = H^* c^j + G^* d^j$$

für $j = \underline{j}, \ldots, \overline{j} - 1$ und Berechnung von

$$f = \sum_{k \in \mathbb{Z}} c^{\overline{j}}_k \varphi_{\overline{j},k} = \sum_{k=-2^{\overline{j}}-2r}^{2^{\overline{j}}-1} c^{\overline{j}}_k \varphi_{\overline{j},k}.$$

10.3.3. Koeffizientenanzahl

Durch den beschriebenen Zerlegungsalgorithmus steigt die Maximalanzahl der Nicht-Null-Elemente in den Folgen $c^{\underline{j}}$ und d^j, $j = \underline{j}, \ldots, \overline{j} - 1$, im Vergleich zur Maximalanzahl der Nicht-Null-Elemente in der Folge $c^{\overline{j}}$ in Abhängigkeit von der Wavelet-Ordnung r und der Wahl von $\underline{j}$ und $\overline{j}$ leicht an. Man kann nachrechnen, dass gilt:

Maximalanzahl $c_k^{\overline{j}}$:	$2^{\overline{j}+1} + 2r$,
Maximalanzahl $c_k^{\underline{j}}$, d_k^j, $j = \underline{j}, \ldots, \overline{j} - 1$:	$2^{\overline{j}+1} + 2r + 2r(\overline{j} - \underline{j})$.

10.4. Zweidimensionale Wavelets

Um das inverse Problem der Schlierentomografie numerisch zu lösen, werden wir mit auf $\mathbb{R}^2$ definierten Funktionen arbeiten müssen. Somit ist eine Erweiterung der in den vorhergehenden Abschnitten dargestellten Konzepte auf den zweidimensionalen Fall nötig.

Seien φ und ψ die Skalierungsfunktion und das Wavelet der Ordnung $r \geq 0$ nach Daubechies. Wir führen Funktionen $\Phi, \Psi^h, \Psi^v, \Psi^d : \mathbb{R}^2 \to \mathbb{R}$ wie folgt ein:

$$\begin{aligned}
\Phi(x,y) &:= \varphi(x)\varphi(y),\\
\Psi^h(x,y) &:= \varphi(x)\psi(y),\\
\Psi^v(x,y) &:= \psi(x)\varphi(y),\\
\Psi^d(x,y) &:= \psi(x)\psi(y).
\end{aligned}$$

Dabei stehen die oberen Indizes h, v und d für „horizontal“, „vertikal“ und „diagonal“. Diese Bezeichnungen haben ihren Ursprung in der Bildverarbeitung und sollen die unterschiedliche „Empfindlichkeit“ der Funktionen gegenüber entsprechend gelagerten Kanten im Bild hervorheben. Für Details sei auf [30] verwiesen. Zur Vereinfachung der Notation setzen wir formal $T := \{h, v, d\}$ und schreiben dann Ψ^t mit $t \in T$.

Aus den Beobachtungen im eindimensionalen Fall erhält man sofort die Beziehungen

$$\begin{aligned}
\operatorname{supp} \Phi &\subseteq [0, 2r+1] \times [0, 2r+1],\\
\operatorname{supp} \Psi^h &\subseteq [0, 2r+1] \times [-r, r+1],\\
\operatorname{supp} \Psi^v &\subseteq [-r, r+1] \times [0, 2r+1],\\
\operatorname{supp} \Psi^d &\subseteq [-r, r+1] \times [-r, r+1].
\end{aligned}$$

Führen wir für $t \in T$, $j \in \mathbb{N}_0$ und $k = (k_1, k_2) \in \mathbb{Z}^2$ Funktionen $\Phi_{j,k}$ und $\Psi^t_{j,k}$ durch

$$\begin{aligned}
\Phi_{j,k}(x,y) &:= 2^j \Phi(2^j x - k_1, 2^j y - k_2),\\
\Psi^t_{j,k}(x,y) &:= 2^j \Psi^t(2^j x - k_1, 2^j y - k_2)
\end{aligned}$$

ein, so bildet die Menge $\{\Phi_{j,k} : k \in \mathbb{Z}^2\}$ für jedes $j \in \mathbb{N}_0$ ein Orthonormalsystem in $L^2(\mathbb{R}^2)$. Mit $V_j := \overline{\operatorname{span}\{\Phi_{j,k} : k \in \mathbb{Z}^2\}}$ für $j \in \mathbb{N}_0$ erhalten wir eine zur Multi-Skalen-Analysis analoge Konstruktion in $L^2(\mathbb{R}^2)$. Es stellt sich heraus, dass die Menge $\{\Psi^t_{j,k} : t \in T, k \in \mathbb{Z}^2\}$ eine Orthonormalbasis des orthogonalen Komplements von V_j in V_{j+1} ist. In diesem Sinne ist es also gerechtfertig, die Funktion Φ als (zweidimensionale) Skalierungsfunktion und die drei Funktionen Ψ^h, Ψ^v, Ψ^d als (zweidimensionale) Wavelets zu bezeichnen. Man kann zeigen, dass die folgende Aussage gilt.

Satz 10.6. *Für jedes $\underline{j} \in \mathbb{N}_0$ ist $\{\Phi_{\underline{j},k} : k \in \mathbb{Z}^2\} \cup \{\Psi^t_{j,k} : t \in T, k \in \mathbb{Z}^2, j \geq \underline{j}\}$ eine*

Orthonormalbasis in $L^2(\mathbb{R}^2)$*, d.h. für* $f \in L^2(\mathbb{R}^2)$ *gilt*

$$f = \sum_{k\in\mathbb{Z}^2} c_k^{\underline{j}} \Phi_{\underline{j},k} + \sum_{j\geq\underline{j}} \sum_{t\in T} \sum_{k\in\mathbb{Z}^2} d_k^{t,j} \Psi_{j,k}^t$$

mit

$$c_k^j = \int_{\mathbb{R}^2} f(x,y)\Phi_{j,k}(x,y)\,\mathrm{d}(x,y) \qquad \textit{und} \qquad d_k^{t,j} = \int_{\mathbb{R}^2} f(x,y)\Psi_{j,k}^t(x,y)\,\mathrm{d}(x,y).$$

Wir kommen nun zur Berechnung der Koeffizienten bei gegebenem f und zur Rekonstruktion der Funktion f aus den Koeffizienten. Zur Vereinfachung beschränken wir uns auf den Fall $\operatorname{supp} f \subseteq [-1,1]\times[-1,1]$. Dann erhält man analog zu den Überlegungen im Eindimensionalen

$$\begin{aligned}
c_k^j \neq 0 \quad &\Rightarrow \quad k \in [-2^j - 2r, 2^j - 1] \times [-2^j - 2r, 2^j - 1],\\
d_k^{h,j} \neq 0 \quad &\Rightarrow \quad k \in [-2^j - 2r, 2^j - 1] \times [-2^j - r, 2^j + r - 1],\\
d_k^{v,j} \neq 0 \quad &\Rightarrow \quad k \in [-2^j - r, 2^j + r - 1] \times [-2^j - 2r, 2^j - 1],\\
d_k^{d,j} \neq 0 \quad &\Rightarrow \quad k \in [-2^j - r, 2^j + r - 1] \times [-2^j - r, 2^j + r - 1].
\end{aligned}$$

Außerdem nehmen wir für gegebenes $\overline{j} \geq \underline{j}$ vereinfachend $d_k^{t,j} = 0$ für $k \in \mathbb{Z}^2$, $t \in T$ und $j \geq \overline{j}$ an, sodass insgesamt nur noch endlich viele Koeffizienten von Null verschieden sind.

10.4.1. Bestimmung der Koeffizienten

Zur Bestimmung der Koeffizienten $c_k^{\underline{j}}$, $k \in \mathbb{Z}^2$, und $d^{t,j}$, $k \in \mathbb{Z}^2$, $t \in T$, $j = \underline{j}, \ldots, \overline{j} - 1$, wird zuerst

$$\begin{aligned}
c_k^{\overline{j}} &= \int_{\mathbb{R}^2} f(x,y) 2^{\overline{j}} \Phi(2^{\overline{j}}x - k_1, 2^{\overline{j}}y - k_2)\,\mathrm{d}(x,y)\\
&= 2^{-\overline{j}} \int_{[0,2r+1]^2} f\big(2^{-\overline{j}}(x+k_1), 2^{-\overline{j}}(y+k_2)\big)\Phi(x,y)\,\mathrm{d}(x,y)
\end{aligned}$$

für $k_1, k_2 = -2^{\overline{j}} - 2r, \ldots, 2^{\overline{j}} - 1$ berechnet. Dabei haben wir $\operatorname{supp}\Phi \subseteq [0, 2r+1]\times[0, 2r+1]$ verwendet.

Wir fassen c^j und $d^{t,j}$, $t \in T$, nun als doppelt unendliche Matrizen auf und wenden die für den eindimensionalen Fall eingeführten Operatoren H und G zunächst zeilenweise und anschließend spaltenweise auf diese Matrizen an. Mit Hilfsvariablen a^j und b^j ist

also

$$\begin{aligned}
a^j_{(k_1,\bullet)} &:= Hc^j_{(k_1,\bullet)}, \quad k_1 \in \mathbb{Z},\\
b^j_{(k_1,\bullet)} &:= Gc^j_{(k_1,\bullet)}, \quad k_1 \in \mathbb{Z},\\
c^{j-1}_{(\bullet,k_2)} &= Ha^j_{(\bullet,k_2)}, \quad k_2 \in \mathbb{Z},\\
d^{h,j-1}_{(\bullet,k_2)} &= Ga^j_{(\bullet,k_2)}, \quad k_2 \in \mathbb{Z},\\
d^{v,j-1}_{(\bullet,k_2)} &= Hb^j_{(\bullet,k_2)}, \quad k_2 \in \mathbb{Z},\\
d^{d,j-1}_{(\bullet,k_2)} &= Gb^j_{(\bullet,k_2)}, \quad k_2 \in \mathbb{Z}
\end{aligned}$$

für $j = \overline{j}, \ldots, \underline{j} + 1$ zu berechnen.

10.4.2. Rekonstruktion der Funktion

Um f aus den Koeffizienten $c^{\underline{j}}_k$, $k \in \mathbb{Z}^2$, und $d^{t,j}$, $k \in \mathbb{Z}^2$, $t \in T$, $j = \underline{j}, \ldots, \overline{j} - 1$, zu rekonstruieren werden mit Hilfsvariablen $\tilde{c}^j$, $\tilde{d}^{h,j}$, $\tilde{d}^{v,j}$ und $\tilde{d}^{d,j}$ zunächst

$$\begin{aligned}
\tilde{c}^j_{(k_1,\bullet)} &:= H^* c^j_{(k_1,\bullet)}, \quad k_1 \in \mathbb{Z},\\
\tilde{d}^{h,j}_{(k_1,\bullet)} &:= H^* d^{h,j}_{(k_1,\bullet)}, \quad k_1 \in \mathbb{Z},\\
\tilde{d}^{v,j}_{(k_1,\bullet)} &:= G^* d^{v,j}_{(k_1,\bullet)}, \quad k_1 \in \mathbb{Z},\\
\tilde{d}^{d,j}_{(k_1,\bullet)} &:= G^* d^{d,j}_{(k_1,\bullet)}, \quad k_1 \in \mathbb{Z},\\
c^{j+1}_{(\bullet,k_2)} &= H^* c^j_{(\bullet,k_2)} + G^* d^{h,j}_{(\bullet,k_2)} + H^* d^{v,j}_{(\bullet,k_2)} + G^* d^{d,j}_{(\bullet,k_2)}, \quad k_2 \in \mathbb{Z}
\end{aligned}$$

für $j = \underline{j}, \ldots, \overline{j} - 1$ berechnet und anschließend

$$f = \sum_{k \in \mathbb{Z}^2} c^{\overline{j}}_k \Phi_{\overline{j},k} = \sum_{k_1,k_2=-2^{\overline{j}}-2r}^{2^{\overline{j}}-1} c^{\overline{j}}_k \Phi_{\overline{j},k}$$

gesetzt.

10.4.3. Koeffizientenanzahl

Wie im eindimensionalen Fall erhöht sich die Koeffizientenanzahl durch den Zerlegungsalgorithmus. Man rechnet leicht nach, dass gilt:

Maximalanzahl $c^{\overline{j}}_k$:	$(2^{\overline{j}+1} + 2r)^2$,
Maximalanzahl $c^{\underline{j}}_k$, $d^{t,j}_k$, $t \in T$, $j = \underline{j}, \ldots, \overline{j} - 1$:	$(2^{\overline{j}+1} + 2r)^2 + 12r^2(\overline{j} - \underline{j}) + 16r(2^{\overline{j}} - 2^{\underline{j}})$.

Bei der Anwendung der hier beschriebenen Algorithmen zur Lösung des Schlierentomografieproblems werden wir statt $\operatorname{supp} f \subseteq [-1,1] \times [-1,1]$ sogar $\operatorname{supp} f \subseteq \{(x,y) \in \mathbb{R}^2 : x^2+y^2 \leq 1\}$ voraussetzen können. Mit dieser zusätzlichen Eingrenzung des Trägers von f kann die Maximalanzahl der Nicht-Null-Koeffizienten noch um ca. 20 Prozent gesenkt werden. Die entsprechenden Berechnungen der Koeffizientenanzahlen sind zwar elementar, aber aufwändig und sollen deshalb hier nicht widergegeben werden. Bei der numerischen Umsetzung werden wir dies aber berücksichtigen, um die Rechenzeiten möglichst gering zu halten.

11. Besov-Räume

In diesem Kapitel führen wir eine Klasse von Räumen ein, mit deren Hilfe wir im nächsten Kapitel das inverse Problem der Schlierentomografie formulieren werden. Wir geben zunächst eine recht allgemeine Definition an, werden uns dann aber sofort auf einen Spezialfall zurückziehen.

Sei $n \in \mathbb{N}$ und $1 \leq p \leq \infty$. Für $f \in L^p(\mathbb{R}^n)$ und $M \in \mathbb{N}$ führen wir den Glattheitsmodul $\omega_M(f, \bullet) : \mathbb{R} \to \mathbb{R}$ mittels

$$\omega_M(f,t) := \sup_{h \in \mathbb{R}^n, |h| \leq t} \|\Delta_h^M f\|_{L^p(\mathbb{R}^n)}$$

ein, wobei

$$(\Delta_h^M f)(x) := \sum_{j=0}^{M} (-1)^{M-j} \binom{M}{j} f(x + jh)$$

gesetzt wurde. Letzteres ist im Wesentlichen der Vorwärtsdifferenzenquotient M-ter Ordnung zum Schritt h.

Definition 11.1. Sei $1 \leq p \leq \infty$, $1 \leq q < \infty$, $s > 0$ und $M := \lceil s \rceil$. Für $f \in L^p(\mathbb{R}^n)$ definieren wir

$$\|f\|_{B^{s,p,q}(\mathbb{R}^n)} := \|f\|_{L^p(\mathbb{R}^n)} + \left(\int_0^1 \left(\frac{\omega_M(f,t)}{t^s} \right)^q \frac{1}{t} \, \mathrm{d}t \right)^{\frac{1}{q}}$$

und setzen

$$B^{s,p,q}(\mathbb{R}^n) := \{ f \in L^p(\mathbb{R}^n) : \|f\|_{B^{s,p,q}(\mathbb{R}^n)} < \infty \}.$$

Der so definierte normierte lineare Raum heißt *Besov-Raum*.

Man kann zeigen, dass die eingeführten Besov-Räume stets Banach-Räume sind. Die Definition kann mittels Fourier-Techniken auch auf $0 < p \leq \infty$, $0 < q \leq \infty$, $s \in \mathbb{R}^n$ ausgedehnt werden, wobei dann aber nicht jeder Besov-Raum ein Banach-Raum ist. Für tief gehende Untersuchungen zu diesem allgemeinen Fall verweisen wir auf [32].

Von nun an beschränken wir unsere Betrachtungen auf den Fall $n = 2$, $p = q = s = 1$. Wir haben also

$$\|f\|_{B^{1,1,1}(\mathbb{R}^2)} = \|f\|_{L^1(\mathbb{R}^2)} + \int_0^1 \frac{1}{t^2} \sup_{h \in \mathbb{R}^2, |h| \leq t} \|f(\bullet) - 2f(\bullet + h) + f(\bullet + 2h)\|_{L^1(\mathbb{R}^2)} \, \mathrm{d}t.$$

Seien nun $\Phi_{j,k}$ und $\Psi_{j,k}^t$, $t \in T$, wie in Abschnitt 10.4 eingeführt. Die Ausdrücke

$$c_k^j := \int_{\mathbb{R}^2} f(x,y) \Phi_{j,k}(x,y) \, \mathrm{d}(x,y) \qquad \text{und} \qquad d_k^{t,j} := \int_{\mathbb{R}^2} f(x,y) \Psi_{j,k}^t(x,y) \, \mathrm{d}(x,y)$$

($t \in T, j \in \mathbb{N}_0, k \in \mathbb{Z}^2$) sind für $f \in B^{1,1,1}(\mathbb{R}^2)$ sinnvoll, da $B^{1,1,1}(\mathbb{R}^2)' = B^{-1,\infty,\infty}(\mathbb{R}^2)$ und $\Phi_{j,k}, \Psi^t_{j,k} \in B^{-1,\infty,\infty}(\mathbb{R}^2)$ für hinreichend große Daubechies-Ordnung r gezeigt werden kann, wobei wir den Raum $B^{-1,\infty,\infty}(\mathbb{R}^2)$ hier nicht eingeführt haben (siehe [34]). In [23] und [34] findet man den folgenden Satz, wobei der zweite Autor aufgrund der zum Beweis verwendeten Techniken die Ordnung r der Daubechies-Wavelets stärker einschränkt.

Satz 11.2. *Sei $\underline{j} \in \mathbb{N}_0$. Dann ist $\{\Phi_{j,k} : k \in \mathbb{Z}^2\} \cup \{\Psi^t_{j,k} : t \in T, k \in \mathbb{Z}^2, j \geq \underline{j}\}$ für $r \geq 3$ eine Basis in $B^{1,1,1}(\mathbb{R}^2)$, d.h. $f \in B^{1,1,1}(\mathbb{R}^2)$ kann als*

$$f = \sum_{k \in \mathbb{Z}^2} c^{\underline{j}}_k \Phi_{\underline{j},k} + \sum_{j \geq \underline{j}} \sum_{t \in T} \sum_{k \in \mathbb{Z}^2} d^{t,j}_k \Psi^t_{j,k}$$

dargestellt werden. Die auftretenden Reihen sind unbedingt konvergent.

Desweiteren findet man in [34] den folgenden Satz.

Satz 11.3. *Die durch*

$$|||f|||_{B^{1,1,1}(\mathbb{R}^2)} := \sum_{k \in \mathbb{Z}^2} |c^{\underline{j}}_k| + \sum_{j \geq \underline{j}} \sum_{t \in T} \sum_{k \in \mathbb{Z}^2} |d^{t,j}_k|$$

auf $B^{1,1,1}(\mathbb{R}^2)$ definierte Norm ist äquivalent zu $\|\bullet\|_{B^{1,1,1}(\mathbb{R}^2)}$.

12. Das inverse Problem

Nachdem wir in den vorhergehenden Kapiteln wesentliche Werkzeuge bereitgestellt haben, sind wir nun in der Lage das inverse Problem der Schlierentomografie in der üblichen Form, d.h. mit Hilfe eines Operators F zur Beschreibung des direkten Problems, mit konkreten Räumen für Daten und Lösung zu formulieren. Zunächst werden wir die benötigten Räume und den Operator F definieren, sowie wesentliche Eigenschaften von Räumen und Operator angeben. Anschließend soll diskutiert werden, worin Vor- und auch Nachteile der Raumwahl liegen. Aufgrund der Ähnlichkeit zwischen dem inversen Problem der Schlierentomografie und dem der Computertomografie ist es notwendig, den unscheinbaren, aber wesentlichen Unterschied zwischen beiden Problemen herauszustellen. Darauf werden wir am Ende dieses Kapitels eingehen.

Wie in Kapitel 9 setzen wir $\Omega := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$. Wir betrachten die Menge

$$B(\Omega) := \{u \in B^{1,1,1}(\mathbb{R}^2) : \operatorname{supp} u \subseteq \overline{\Omega}\}.$$

Mit der durch $\|u\|_{B(\Omega)} := \|u\|_{B^{1,1,1}(\mathbb{R}^2)}$ definierten Norm $\|\bullet\|_{B(\Omega)}$ ist dies offensichtlich ein normierter linearer Raum. Nach [34] ist $B(\Omega)$ sogar ein Banach-Raum. Mit $L^2(\Omega) = \{u \in L^2(\mathbb{R}^2) : \operatorname{supp} u \subseteq \overline{\Omega}\}$ folgt aus der stetigen Einbettung $B^{1,1,1}(\mathbb{R}^2) \hookrightarrow L^2(\mathbb{R}^2)$ (siehe [33]) die stetige Einbettung

$$B(\Omega) \hookrightarrow L^2(\Omega).$$

Man kann auch einen Raum $B^{1,1,1}(\Omega)$ definieren. In diesem kann jedoch nicht ohne Weiteres eine Wavelet-Basis angegeben werden, sodass wir hier mit dem Raum $B^{1,1,1}(\mathbb{R}^2)$ arbeiten und die Träger der Funktionen geeignet einschränken. Zum Thema Wavelet-Basen in Besov-Räumen über Gebieten Ω verweisen wir auf [34].

Wir werden den Operator F des direkten Problems aus einzelnen Operatoren F_i, $i = 1, \ldots, N$, zusammensetzen, wobei N die Anzahl der zur Verfügung stehenden Schlierenbilder ist (vgl. Kapitel 9). Die Operatoren F_i definieren wir mittels

$$(F_i(u))(s) := (R_i u)(s)^2, \quad s \in [-1, 1],$$

wobei R_i die zur Aufnahmerichtung σ_i des i-ten Schlierenbildes gehörende Radon-Transformation bezeichnet, d.h.

$$(R_i u)(s) := \int_{-\infty}^{\infty} u(s\sigma_i + q\sigma_i^{\perp})\,\mathrm{d}q, \quad s \in [-1, 1].$$

Wie bereits in Kapitel 9 bemerkt, ist das Integral für $u \in L^1(\Omega)$ nach dem Satz von Fubini für fast alle $s \in [-1, 1]$ wohldefiniert und endlich. Die folgenden Aussagen liefern weitere Erkenntnisse für $u \in B(\Omega)$.

Satz 12.1. *Für $u \in L^2(\Omega)$, also insbesondere für $u \in B(\Omega)$, gilt $F_i(u) \in L^1(-1, 1)$.*

Zudem sind die Abbildungen $F_i : L^2(\Omega) \to L^1(-1,1)$ und $F_i : B(\Omega) \to L^1(-1,1)$ stetig und besitzen in jedem Punkt u Richtungsableitungen in alle Richtungen. Die Richtungsableitung im Punkt u in Richtung h ist gegeben durch

$$\big(F_i'(u)h\big)(s) := 2(R_i u)(s)(R_i h)(s), \quad s \in [-1,1];$$

somit ist $F_i'(u)$ ein linearer Operator.

Beweis. Sei $u \in C_0^\infty(\Omega)$. Dann folgt aus der Cauchy-Schwarz'schen Ungleichung und dem Satz von Fubini

$$\begin{aligned}
\|F_i(u)\|_{L^1(-1,1)} &= \int_{-1}^{1} \left(\int_{-\infty}^{\infty} u(s\sigma_i + q\sigma_i^\perp)\,\mathrm{d}q \right)^2 \mathrm{d}s \\
&\le \int_{-1}^{1} \int_{-\infty}^{\infty} u(s\sigma_i + q\sigma_i^\perp)^2\,\mathrm{d}q \underbrace{\int_{-\infty}^{\infty} \chi_\Omega(s\sigma_i + q\sigma_i^\perp)^2\,\mathrm{d}q}_{\le 2}\,\mathrm{d}s \\
&\le 2 \int_\Omega u(s\sigma_i + q\sigma_i^\perp)^2\,\mathrm{d}(s,q) = 2\|u\|_{L^2(\Omega)}^2,
\end{aligned}$$

d.h. $F_i : (C_0^\infty(\Omega), \|\bullet\|_{L^2(\Omega)}) \to L^1(-1,1)$ ist stetig. Da $C_0^\infty(\Omega)$ dicht in $L^2(\Omega)$ ist, folgt daraus die Stetigkeit von $F_i : L^2(\Omega) \to L^1(-1,1)$ und somit aus $B(\Omega) \hookrightarrow L^2(\Omega)$ die Stetigkeit von $F_i : B(\Omega) \to L^1(-1,1)$.

Die Behauptungen über die Richtungsableitungen folgen aus

$$\begin{aligned}
&\left\| \frac{F_i(u+\varepsilon h) - F_i(u)}{\varepsilon} - F_i'(u)h \right\|_{L^1(-1,1)} \\
&= \left\| \frac{F_i(u) + \varepsilon F_i'(u)h + \varepsilon^2 F_i(h) - F_i(u)}{\varepsilon} - F_i'(u)h \right\|_{L^1(-1,1)} \\
&= |\varepsilon|\, \|F_i(h)\|_{L^1(-1,1)} \overset{\varepsilon \to +0}{\longrightarrow} 0
\end{aligned}$$

für alle $u \in L^2(\Omega)$ und alle $h \in L^2(\Omega)$. □

Bemerkung. In Satz 12.1 wird festgestellt, dass die Richtungsableitungen in $u \in B(\Omega)$ durch einen linearen Operator $F_i'(u)$ gegeben sind. Wünschenswert wäre die Gâteaux-Differenzierbarkeit von F_i in u, d.h. $F_i'(u)$ sollte beschränkt sein. Diese Eigenschaft konnte jedoch im Rahmen dieser Arbeit weder nachgewiesen noch widerlegt werden. Weiter unten werden wir noch sehen, dass selbst die Gâteaux-Differenzierbarkeit für die numerische Analysis oft nicht ausreicht. Stattdessen benötigt man meist die Fréchet-Differenzierbarkeit des Operators, d.h. der Operator muss sich durch eine Taylor-Entwicklung erster Ordnung approximieren lassen.

Wir definieren nun den Operator $F : B(\Omega) \to (L^1(-1,1))^N$ durch

$$F(u) := \big(F_1(u), \ldots, F_N(u)\big).$$

Den Produktraum $(L^1(-1,1))^N$ statten wir dabei mit der üblichen Norm aus, d.h. für $v = (v_1, \ldots, v_N)$ gilt

$$\|v\|_{(L^1(-1,1))^N} = \|v_1\|_{L^1(-1,1)} + \cdots + \|v_N\|_{L^1(-1,1)}.$$

Aus obigem Satz folgt sofort die Stetigkeit von F und die Existenz sämtlicher Richtungsableitungen.

In der Praxis können keine exakten Daten $v \in (L^1(-1,1))^N$ gemessen werden, sondern nur gestörte Daten $v^\delta \in (L^1(-1,1))^N$, von denen wir annehmen, dass sie für $\delta > 0$ die Bedingung

$$\|v_i - v_i^\delta\|_{L^1(-1,1)} \leq \frac{\delta}{N}, \quad i = 1, \ldots, N,$$

erfüllen. Dann gilt automatisch $\|v - v^\delta\|_{(L^1(-1,1))^N} \leq \delta$.

Das inverse Problem der Schlierentomografie besteht nun darin, die Operatorgleichung

$$F(u) = v$$

für $v \in (L^1(-1,1))^N$ zu lösen, wobei die rechte Seite nicht bekannt ist, sondern nur ein v^δ mit $\|v - v^\delta\|_{(L^1(-1,1))^N} \leq \delta$ vorliegt. Die Lösung u soll in $B(\Omega)$ liegen, wobei sofort klar ist, dass die Lösung nicht eindeutig sein kann, da mit u auch $-u$ eine Lösung ist.

Der Operator F wurde in [12] bereits für $u \in H_0^1(\Omega)$ und $v \in L^2(\Omega)^N$ untersucht. Untersuchungen zum inversen Problem der Schlierentomografie mit etwas anderer Modellierung findet man auch in [29].

Wir wollen nun noch die Wahl des Raumes $B(\Omega)$ als Lösungsraum und den Ansatz zur numerischen Lösung des Schlierentomografieproblems mittels Tikhonov-Regularisierung begründen.

In der Bildverarbeitung haben Besov-Räume einen festen Platz. Besonders interessant ist der Raum $B^{1,1,1}(\mathbb{R}^2)$, da er (zumindest für beschränkte Gebiete Ω statt $\mathbb{R}^2$) „kleiner“ ist als der ebenfalls gern genutzte Raum der Funktionen beschränkter Variation (siehe [5]), aber immer noch „groß“ genug ist, um „natürliche“ Bilder, d.h. Bilder mit Kanten und strukturierten Flächen, zuzulassen. Einige Beispiele für in $B^{1,1,1}(\mathbb{R}^2)$ enthaltene Bilder werden in [7] diskutiert. Eine kurze Zusammenfassung von Nachteilen der Besov-Räume für die Bildverarbeitung findet man in [5]. Diese Nachteile betreffen im Wesentlichen so genannte „Real-World-Bilder“ (dies sind Bilder, die Ausschnitte der mit bloßem Auge wahrnehmbaren Welt abbilden, wie zum Beispiel Landschaftsfotografien) und sollen deshalb hier nicht näher erläutert werden. Bei der Schlierentomografie auftretende Bilder bestehen aufgrund der Darstellung von Druckverteilungen aus verschiedenen Bereichen konstanter Intensität, die durch mehr oder weniger scharfe Kanten begrenzt werden; somit fallen Schlierenbilder nicht in die Klasse der deutlich komplexeren Real-World-Bilder.

Ein zweiter wichtiger Vorteil der Besov-Räume ist die Möglichkeit Funktionen mittels Wavelet-Basen darzustellen. Die Wavelet-Koeffizienten verschiedener Skalierungsstufen (vgl. Multi-Skalen-Analysis in Abschnitt 10.1) entsprechen unterschiedlichen Detailstufen des Bildes. Ein Bild wird also in eine grobe Approximation (Koeffizienten der Skalierungsfunktion) und in verschiedene Detailstufen zerlegt. Man kann nun zeigen, dass die Wavelet-Koeffizienten unter geeigneten Voraussetzungen mit steigender Skalierungsstufe gegen Null fallen. Zudem sind die Wavelet-Koeffizienten, die zu Bildbereichen konstanter Intensität gehören, identisch Null. Die Darstellung von Bildern mit Hilfe von Wavelet-Basen führt also dazu, dass nur wenige Koeffizienten von Null verschieden sind. Eine Diskussion des Verhaltens von Wavelet-Koeffizienten findet man in [4] und auch in [7].

In Zusammenhang mit der Tikhonov-Regularisierung liefert uns das Verschwinden der Wavelet-Koeffizienten und die Arbeit im Besov-Raum wegen Satz 11.3 so genannte Sparsity-Constraints (wir bleiben hier bei der englischen Bezeichnung), die uns die Verwendung spezieller Minimierungsalgorithmen zur Minimierung des Tikhonov-Funktionals erlauben. Dazu kommen wir weiter unten noch genauer.

Neben diesen praktischen Erwägungen spielt bei der Raumwahl auch die Tatsache eine Rolle, dass das Schlierentomografieproblem für $u \in H_0^1(\Omega)$ und $u \in BV(\Omega)$ bereits in [29] mittels Tikhonov-Regularisierung numerisch gelöst wurde und in dieser Arbeit ein bisher noch nicht in Zusammenhang mit der Schlierentomografie erprobtes Verfahren auf seine Eigenschaften untersucht werden soll.

Zum Abschluss dieses Kapitels soll geklärt werden, wieso die bekannten und weit entwickelten Methoden der Computertomografie nicht zur Lösung des Schlierentomografieproblems geeignet sind. Auf den ersten Blick kann der Eindruck entstehen, dass man aus den Daten v bzw. v^δ die Wurzel ziehen kann und dann nur noch die Radon-Transformation zu invertieren hat. Bei genauerer Betrachtung stellt sich jedoch heraus, dass dies nicht korrekt ist. Der Grund dafür ist die Notwendigkeit, auch negative Werte für die Druckverteilung u zuzulassen. Die Schlierenbilder geben nur Auskunft über den Betrag der in Projektionsrichtung gemittelten Druckverteilung, sodass die üblichen Algorithmen für die Computertomografie (z.B. gefilterte Rückprojektion) zu falschen Ergebnissen führen können. Ein Beispiel für dieses „Fehlverhalten“ findet man in [12]. Das inverse Problem der Schlierentomografie muss also als nichtlineares inverses Problem behandelt werden, um eine Chance auf Rekonstruktion von negativen Funktionswerten zu haben. Ebenfalls in [12] wird ein iteratives Verfahren vorgestellt, welches dies leistet. Ob der hier verwendete Ansatz mittels Tikhonov-Regularisierung diesbezüglich ebenfalls gute Ergebnisse liefert, muss anhand von Testrechnungen untersucht werden.

Die Wichtigkeit, positive und negative Druckanteile im erzeugten Bild unterscheiden zu können, liegt in der Anwendung von Ultraschall für medizinische Zwecke begründet. Untersuchungen haben gezeigt, dass die negativen Druckanteile nachweislich zu Schädigungen des Gewebes durch Bildung von Gasbläschen (Kavitation) führen können. Man ist also bestrebt, negative Druckmaxima zu minimieren. Physikalisch ist dies durchaus möglich. Unterschiedlich große Druckminima und -maxima treten zum Beispiel bei Ultraschallstoßwellen zur Zerstörung von Gallensteinen auf. Genaueres findet man in [24].

Inzwischen werden auch Verfahren erforscht, um nicht nur den Betrag des entlang der

Projektionsrichtung gemittelten Drucks, sondern auch das Vorzeichen dieses Wertes zu messen. Eine Möglichkeit dazu wird in [27] beschrieben.

13. Numerische Realisierung

In diesem Kapitel soll ein Verfahren zur numerischen Lösung des inversen Problems der Schlierentomografie vorgestellt werden. Der Übergang vom analytisch formulierten Problem zum numerischen Algorithmus wird nicht ohne „Verluste“ vonstatten gehen können. Insbesondere die Raumwahl für Daten und Lösung wird sich als schwierig erweisen, sodass an mehreren Stellen Kompromisse zwischen analytischer Korrektheit und numerischer Umsetzbarkeit eingegangen werden müssen.

Auf die Untersuchung von Stabilität und Konvergenz werden wir verzichten müssen, um im Rahmen dieser Arbeit überhaupt zu numerischen Ergebnissen zu gelangen. Ziel soll vorrangig die „Beobachtung“ des im Weiteren vorgestellten Verfahrens im Sinne einer ersten Eignungsprüfung sein, nicht jedoch der mathematische Beweis seiner Eigenschaften.

13.1. Das Tikhonov-Funktional

Im vorangegangenen Kapitel wurde das inverse Problem der Schlierentomografie mit Hilfe des Operators $F : B(\Omega) \to (L^1(-1,1))^N$ und gestörten Daten $v^\delta \in (L^1(-1,1))^N$ formuliert. Die Wahl der Räume erfolgte dabei zunächst ohne Beachtung der Möglichkeiten zur numerischen Lösung. Ausschlaggebend war lediglich das Interesse an Lösungen im Besov-Raum $B^{1,1,1}(\mathbb{R}^2)$, um neben den in [29] betrachteten Räumen $H_0^1(\Omega)$ und $BV(\Omega)$ eine weitere Lösungsmethode zu erproben. Den Abbildungseigenschaften des Operators F geschuldet, musste als Datenraum daher $(L^1(-1,1))^N$ gewählt werden.

Zur Lösung mittels Tikhonov-Regularisierung kann man nun zum Beispiel das Funktional

$$\mathcal{T}_\alpha^\delta(u) = \|F(u) - v^\delta\|_{(L^1(-1,1))^N} + \alpha |||u|||_{B^{1,1,1}(\mathbb{R}^2)} \tag{13.1}$$

über $B(\Omega)$ minimieren. Der erste Summand ist nicht konvex und der zweite nicht differenzierbar. Zudem sind $B(\Omega)$ und $(L^1(-1,1))^N$ keine Hilbert-, sondern nur Banach-Räume.

Der Regularisierungsterm $|||u|||_{B^{1,1,1}(\mathbb{R}^2)}$ stellt ein Sparsity-Constraint dar, d.h. er bestraft die Größe und die Anzahl der Koeffizienten von u bezüglich einer Basis, die von Null verschieden sind. Dies sieht man sofort anhand der Definition von $|||\bullet|||_{B^{1,1,1}(\mathbb{R}^2)}$ (vgl. Satz 11.3).

Es stellt sich als schwieriges Problem heraus, einen geeigneten Algorithmus zur Minimierung des Tikhonov-Funktionals (13.1) zu finden. Die Minimierung unter Sparsity-Constraints ist ein sehr aktives Forschungsfeld, sodass in naher Zukunft die Veröffentlichung von Algorithmen zu erwarten ist, die auch Banach-Räume als Daten- bzw. Lösungsraum zulassen. Im Rahmen dieser Arbeit muss jedoch auf Hilbert-Raum-Algorithmen zurückgegriffen werden. Dies macht eine Anpassung des Tikhonov-Funktionals (13.1) und der zugrunde liegenden Räume nötig. Ziel soll es also sein, den L^1-Abstand in (13.1) durch einen L^2-Abstand zu ersetzen und dabei den Regularisierungsterm $|||u|||_{B^{1,1,1}(\mathbb{R}^2)}$ unverändert zu lassen.

Um den Übergang von $(L^1(-1,1))^N$ zu $(L^2(-1,1))^N$ zu realisieren, schränken wir den Definitionsbereich $D(F) \subseteq B(\Omega)$ so weit ein, dass im $F \subseteq (L^2(-1,1))^N$ gilt. In [12] wurde gezeigt, dass für $u \in H_0^1(\Omega)$ das Bild $F(u)$ in $(L^2(-1,1))^N$ liegt. Als Definitionsbereich von F bietet sich also $B(\Omega) \cap H_0^1(\Omega)$ an. Die Stärke dieser Einschränkung hält sich in Grenzen, da man zumindest $B(\Omega) \hookrightarrow W^{1,1}(\Omega)$ zeigen kann (vgl. [34]), d.h die verallgemeinerten ersten partiellen Ableitungen einer $B(\Omega)$-Funktion liegen in $L^1(\Omega)$.

Wir minimierten also das Tikhonov-Funktional

$$T_\alpha^\delta = \|F(u) - v^\delta\|^2_{(L^2(-1,1))^N} + \alpha |||u|||_{B^{1,1,1}(\mathbb{R}^2)} \tag{13.2}$$

über $\{u \in B(\Omega) : F(u) \in (L^2(-1,1))^N\}$. Dabei sollen die gestörten Daten v^δ die Abschätzung $\|v - v^\delta\|_{(L^2(-1,1))^N} \leq \delta$ erfüllen.

In [3] und [26] werden Algorithmen vorgeschlagen, um Tikhonov-Funktionale mit nichtlinearen Operatoren und Sparsity-Constraints in Hilbert-Räumen zu minimieren. In ersterem Paper wird jedoch vorausgesetzt, dass Daten- und Lösungsraum übereinstimmen. Daher werden wir uns für den in [26] vorgeschlagenen Algorithmus entscheiden.

Wesentliche Voraussetzung in [26] ist die Fréchet-Differenzierbarkeit des Operators F. Diese ist zum Beispiel für $F : H_0^1(\Omega) \to (L^2(-1,1))^N$ gegeben (siehe [12]). In unserem Fall, d.h. im Fall $F : D(F) \subseteq L^2(\Omega) \to (L^2(-1,1))^N$ mit hinreichend eingeschränktem Definitionsbereich $D(F) \subseteq B(\Omega)$, ist dies nicht gewährleistet. In Satz 13.2 werden wir nur zeigen, dass die Richtungsableitungen von F in einem Punkt u durch einen linearen Operator $F'(u)$ beschrieben werden. Um die Konvergenz des in [26] vorgestellten Verfahrens und die Regularisierungseigenschaften zu zeigen, sind weitere Voraussetzungen zu erfüllen.

13.2. Minimierung des Tikhonov-Funktionals

Wir stellen nun den in [26] vorgeschlagenen Algorithmus zur Lösung nichtlinearer inverser Probleme unter Sparsity-Constraints am konkreten Beispiel vor.

Zunächst wollen wir die Notation für die Darstellung einer Funktion u mit $\operatorname{supp} u \subseteq \overline{\Omega}$ in einer Wavelet-Basis vereinfachen. Wir hatten am Ende von Abschnitt 10.4 bereits angemerkt, dass aufgrund der Trägerbeschränkung viele Koeffizienten verschwinden. Mit Λ bezeichnen wir nun die Menge all der in der Wavelet-Darstellung auftretenden Indexkombinationen $(j,t,k) \in \mathbb{N}_0 \times (T \cup \{s\}) \times \mathbb{Z}^2$, für die der zugehörige Koeffizient nicht für jede Funktion u mit $\operatorname{supp} u \subseteq \overline{\Omega}$ verschwindet. Das zusätzliche Element s steht dabei analog zu $h, v, d \in T$ für die Koeffizienten der Skalierungsfunktion; zur Indexkombination $(j, s, (0,0))$ gehört also zum Beispiel der Koeffizient $c^j_{(0,0)}$. Mit ϕ_λ bezeichnen wir entsprechend das zur Indexkombination $\lambda \in \Lambda$ gehörige Basiselement und mit b_λ den dazugehörigen Koeffizient. Wir erhalten mit diesen Bezeichnungen die Darstellung

$$u = \sum_{\lambda \in \Lambda} b_\lambda \phi_\lambda.$$

Es ist notwendig, die Wahl des Definitionsbereichs von F anzupassen bzw. zu prä-

zisieren, um die Vorgehensweise aus [26] übernehmen zu können. Es sei $\rho > 0$ so groß gewählt, dass

$$\overline{\bigcup_{\lambda\in\Lambda} \operatorname{supp}\phi_\lambda} \subseteq \Omega_\rho$$

mit

$$\Omega_\rho := \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 < \rho^2\}$$

gilt. Für u mit $\operatorname{supp} u \subseteq \Omega_\rho$ gilt nun $\operatorname{supp} F_i(u) \subseteq [-\rho,\rho]$, $i = 1,\dots,N$, sodass wir

$$D(F) := \left\{u \in \overline{\operatorname{span}\{\phi_\lambda : \lambda\in\Lambda\}}^{B^{1,1,1}(\mathbb{R}^2)} : F(u) \in (L^2(-\rho,\rho))^N\right\}$$

setzen können.

Lemma 13.1. *$D(F)$ ist ein linearer Unterraum von $L^2(\Omega_\rho)$, der die Menge $\{u \in B(\Omega) : F(u) \in (L^2(-1,1))^N\}$ enthält.*

Beweis. $D(F) \subseteq L^2(\Omega_\rho)$ ist klar, da analog zu $B(\Omega) \hookrightarrow L^2(\Omega)$ die Einbettung $B(\Omega_\rho) \hookrightarrow L^2(\Omega_\rho)$ gilt und $D(F)$ eine Teilmenge von $B(\Omega_\rho)$ ist. Um die Abgeschlossenheit von $D(F)$ bezüglich der Addition zu zeigen, seien $u_1, u_2 \in D(F)$. Dann gilt zunächst

$$\begin{aligned}\|R_i(u_1)R_i(u_2)\|^2_{L^2(-\rho,\rho)} &= \int_{-\rho}^{\rho} (R_iu_1)(s)^2(R_iu_2)(s)^2\,\mathrm{d}s = \langle F_i(u_1), F_i(u_2)\rangle_{L^2(-\rho,\rho)}\\ &\le \|F_i(u_1)\|_{L^2(-\rho,\rho)}\|F_i(u_2)\|_{L^2(-\rho,\rho)} < \infty\end{aligned}$$

und somit

$$F_i(u_1+u_2) = R_i(u_1+u_2)^2 = R_i(u_1)^2 + 2R_i(u_1)R_i(u_2) + R_i(u_2)^2 \in L^2(-\rho,\rho),$$

d.h. $F(u_1+u_2) \in (L^2(-\rho,\rho))^N$. Die Abgeschlossenheit von $D(F)$ bezüglich der Multiplikation mit reellen Zahlen ist offensichtlich und die behauptete Inklusion folgt aus der Wahl von Λ und ρ. □

Nach Definition von $D(F)$ ist $\{\phi_\lambda : \lambda \in \Lambda\}$ eine Basis von $\big(D(F), \|\bullet\|_{B^{1,1,1}(\mathbb{R}^2)}\big)$ und offensichtlich auch von $\big(D(F), \|\bullet\|_{L^2(\Omega_\rho)}\big)$ (vgl. Satz 10.4).

Der folgende Satz liefert zwei für den Algorithmus benötigte Operatoren.

Satz 13.2.

(i) Die Richtungsableitung von $F : D(F) \subseteq L^2(\Omega_\rho) \to (L^2(-\rho,\rho))^N$ im Punkt $u \in D(F)$ in Richtung $h \in D(F)$ ist durch

$$F'(u)h = \big(F_1'(u)h, \dots, F_N'(u)h\big) \quad \textit{mit} \quad F_i'(u)h = 2R_i(u)R_i(h)$$

gegeben.

(ii) Der Operator $F'(u) : D(F) \subseteq L^2(\Omega_\rho) \to (L^2(-\rho,\rho))^N$ *ist linear und für* $g = (g_1,\dots,g_N) \in (L^2(-\rho,\rho))^N$ *gilt*

$$F'(u)^* g = \sum_{i=1}^{N} F_i'(u)^* g_i \quad \textit{mit} \quad \big(F_i'(u)^* g_i\big)(x,y) = 2g_i(x\sigma_i^1 + y\sigma_i^2)(R_i u)(x\sigma_i^1 + y\sigma_i^2).$$

Beweis. Aussage (i) folgt aus

$$\begin{aligned}
&\left\| \frac{F_i(u+\varepsilon h) - F_i(u)}{\varepsilon} - F_i'(u)h \right\|_{L^2(-\rho,\rho)} \\
&\quad = \left\| \frac{F_i(u) + \varepsilon F_i'(u)h + \varepsilon^2 F_i(h) - F_i(u)}{\varepsilon} - F_i'(u)h \right\|_{L^2(-\rho,\rho)} \\
&\quad = |\varepsilon| \, \|F_i(h)\|_{L^2(-\rho,\rho)} \overset{\varepsilon\to+0}{\longrightarrow} 0
\end{aligned}$$

für alle $u \in D(F)$ und alle $h \in D(F)$. Aus

$$\begin{aligned}
\|F_i'(u)h\|_{L^2(-\rho,\rho)}^2 &= 4\|R_i(u)R_i(h)\|_{L^2(-\rho,\rho)}^2 = 4\int_{-\rho}^{\rho} (R_i u)(s)^2 (R_i h)(s)^2 \,\mathrm{d}s \\
&= 4\langle F_i(u), F_i(h)\rangle_{L^2(-\rho,\rho)} \le 4\|F_i(u)\|_{L^2(-\rho,\rho)} \|F_i(h)\|_{L^2(-\rho,\rho)} < \infty
\end{aligned}$$

und

$$\langle F'(u)h, g\rangle_{(L^2(-\rho,\rho))^N} = \sum_{i=1}^{N} \langle F_i'(u)h, g_i\rangle_{L^2(-\rho,\rho)} = \left\langle h, \sum_{i=1}^{N} F_i'(u)^* g_i \right\rangle_{L^2(B_\rho(0))}$$

sowie

$$\begin{aligned}
\langle F_i'(u)h, g_i\rangle_{L^2(-\rho,\rho)} &= \int_{-\rho}^{\rho} \left(2(R_i u)(s) \int_{-\infty}^{\infty} h(s\sigma_i + q\sigma_i^\perp) \,\mathrm{d}q \right) g_i(s) \,\mathrm{d}s \\
&= 2\int_{-\rho}^{\rho}\int_{-\rho}^{\rho} (R_i u)(s) h(s\sigma_i + q\sigma_i^\perp) g_i(s) \,\mathrm{d}q \,\mathrm{d}s \\
&= 2\int_{-\rho}^{\rho}\int_{-\rho}^{\rho} (R_i u)(x\sigma_i^1 + y\sigma_i^2) h(x,y) g_i(x\sigma_i^1 + y\sigma_i^2) \,\mathrm{d}y \,\mathrm{d}x \\
&= 2\int_{B_\rho(0)} h(x,y) g_i(x\sigma_i^1 + y\sigma_i^2)(R_i u)(x\sigma_i^1 + y\sigma_i^2) \,\mathrm{d}(x,y)
\end{aligned}$$

erhalten wir Aussage (ii). □

Angemerkt sei, dass $F'(u)^* g$ für $g \in (L^2(-\rho,\rho))^N$ nicht in $D(F)$ liegen muss.

Wir benötigen noch den Operator $W : L^2(\Omega_\rho) \to l^2(\Lambda)$, der durch

$$Wu := (b_\lambda)_{\lambda\in\Lambda}$$

gegeben ist und die Wavelet-Transformation realisiert. Der Adjungierte $W^* : l^2(\Lambda) \to L^2(\Omega_\rho)$ ist durch

$$W^*(b_\lambda)_{\lambda\in\Lambda} = \sum_{\lambda\in\Lambda} b_\lambda \phi_\lambda$$

definiert, denn es gilt

$$\begin{aligned}\langle \tilde{u}, W^*(b_\lambda)_{\lambda\in\Lambda}\rangle_{L^2(\Omega_\rho)} &= \langle \tilde{u}, \sum_{\lambda\in\Lambda} b_\lambda\phi_\lambda\rangle_{L^2(\Omega_\rho)} = \sum_{\lambda\in\Lambda} b_\lambda \langle \tilde{u}, \phi_\lambda\rangle_{L^2(\Omega_\rho)} \\ &= \langle W\tilde{u}, (b_\lambda)_{\lambda\in\Lambda}\rangle_{l^2(\Lambda)}.\end{aligned}$$

Wir können nun den Algorithmus zur Minimierung des Tikhonov-Funktionals angeben. Für die Herleitung des Verfahrens sowie für die Konvergenzvoraussetzungen und -aussagen verweisen wir auf [26].

Die Grundidee besteht in der Ersetzung des Tikhonov-Funktionals durch eine Folge anderer Funktionale, die sich einfacher minimieren lassen. Aus den notwendigen Optimalitätsbedingungen der Ersatzfunktionale ergibt sich je eine Fixpunktiteration (innere Iteration), mit deren Hilfe eine neue Iterierte (äußere Iteration) zur Approximation eines stationären Punktes des eigentlichen Tikhonov-Funktionals berechnet wird.

In den Algorithmus fließt eine Konstante $C > 0$ ein, die aus den Glattheitseigenschaften des Operators F resultiert und das Verhalten des Algorithmus', insbesondere das Verhältnis zwischen der Anzahl äußerer und der Anzahl innerer Iterationen, beeinflusst. Das Verfahren arbeitet direkt mit den Koeffizientenfolgen $(b_\lambda)_{\lambda\in\Lambda}$. Die Iterierten der äußeren Iteration bezeichnen wir mit $b^\mu = (b_\lambda^\mu)_{\lambda\in\Lambda}$, $\mu = 0, 1, 2, \ldots$, die der μ-ten inneren Iteration mit $b^{\mu,\nu} = (b_\lambda^{\mu,\nu})_{\lambda\in\Lambda}$, $\nu = 0, 1, 2, \ldots$

Die Fixpunktiteration wird mittels des Operators $G : l^2(\Lambda) \times l^2(\Lambda) \to l^2(\Lambda)$, gegeben durch

$$G(\tilde{b}, b) := \frac{\alpha}{2C} P\left(\frac{2}{\alpha} W F'(W^*\tilde{b})^*(v^\delta - F(W^*b)) + \frac{2C}{\alpha} b\right),$$

realisiert, wobei $P : l^2(\Lambda) \to l^2(\Lambda)$ durch

$$\big(P(b)\big)_\lambda := (\operatorname{sgn} b_\lambda)\max\{0, |b_\lambda| - 1\}$$

definiert ist.

Der Algorithmus sieht wie folgt aus:

Konstanten $\tau_1 > 0$, $\tau_2 > 0$ und $C > 0$ wählen
Startfolge b^0 wählen
$\mu \leftarrow 0$
do
  $\mu \leftarrow \mu + 1$
  $b^{\mu-1,0} \leftarrow b^{\mu-1}$
  $\nu \leftarrow 0$
  do
    $\nu \leftarrow \nu + 1$
    $b^{\mu-1,\nu} \leftarrow G(b^{\mu-1,\nu-1}, b^{\mu-1})$
  while $\dfrac{\|b^{\mu-1,\nu-1} - b^{\mu-1,\nu}\|_{l^2(\Lambda)}}{\|b^{\mu-1,\nu-1}\|_{l^2(\Lambda)}} > \tau_2$
  $b^{\mu} \leftarrow b^{\mu-1,\nu}$
while $\dfrac{\|b^{\mu-1} - b^{\mu}\|_{l^2(\Lambda)}}{\|b^{\mu-1}\|_{l^2(\Lambda)}} > \tau_1$
$u \leftarrow W^* b^{\mu}$.

In [26] wird gezeigt, dass bei hinreichend groß gewähltem C und unter geeigneten weiteren Voraussetzungen

$$\|b^{\mu,\nu-1} - b^{\mu,\nu}\|_{l^2(\Lambda)} \to 0 \quad \text{für } \nu \to \infty, \quad \mu = 0, 1, 2, \ldots,$$

und

$$\|b^{\mu-1} - b^{\mu}\|_{l^2(\Lambda)} \to 0 \quad \text{für } \mu \to \infty$$

gelten, sodass der Algorithmus terminiert.

Als wesentliche Verletzungen der in [26] getroffenen Voraussetzungen für den Beweis von Konvergenz- und Regularisierungsaussagen stellen sich heraus:

- $D(F)$ ist kein Hilbert-Raum (nicht vollständig).
- F als Operator auf $D(F)$ ist nicht Fréchet-differenzierbar.
- $F'(u)^* g$ muss nicht für jedes $g \in (L^2(-\rho, \rho))^N$ in $L^2(\Omega_\rho)$ und erst recht nicht in $D(F)$ liegen.

Somit sind die Konvergenzresultate aus [26] nicht anwendbar. Da im Rahmen dieser Arbeit der Schwerpunkt der numerischen Untersuchungen ohnehin auf den konkreten Berechnungsergebnissen liegt, soll uns der Mangel an mathematischer Korrektheit hier

nicht weiter beschäftigen. Durch die Diskretisierung sind bei der praktischen Umsetzung des Verfahrens keine Probleme zu erwarten; gegebenenfalls müssen wir niedrigere Konvergenzgeschwindigkeiten als bei Erfüllung aller Voraussetzungen hinnehmen.

13.3. Diskretisierung

Es verbleibt die Diskretisierung des oben vorgestellten Verfahrens. Die Verwaltung einer Vielzahl von Indizes, insbsondere der Indexmenge Λ ist recht aufwändig. Einerseits besteht der Wunsch nach möglichst einfacher Indizierung, andererseits soll auf das Speichern von Vektor- und Matrixeinträgen, die ohnehin gleich Null sind, weitestgehend verzichtet werden.

Wir werden in diesem Abschnitt die diskrete Version einer Funktion über $\mathbb{R}$ als doppelt unendliche Folge auffassen, obwohl in allen Fällen nur endlich viele Einträge von Null verschieden sind. Die konkreten Grenzen für die Indizes der Nicht-Null-Einträge können elementar berechnet werden und sollen hier nicht angegeben werden.

Die abzählbare Indexmenge Λ ersetzten wir durch eine endliche Indexmenge, indem wir für vorgegebenes $\underline{j} \geq 0$ und vorgegebenes $\overline{j} > \underline{j}$ die Menge

$$\Lambda_{\underline{j},\overline{j}} := \{(j,t,k) \in \Lambda : \underline{j} \leq j < \overline{j}\}$$

einführen. Im Folgenden nehmen wir an, dass die gesuchte Lösung des Schlierentomografieproblems eine Darstellung

$$\sum_{\lambda \in \Lambda_{\underline{j},\overline{j}}} b_\lambda \phi_\lambda$$

besitzt, wobei die Skalierungsfunktion Φ und die Wavelets Ψ^h, Ψ^v, Ψ^d entsprechend Abschnitt 10.2 bzw. 10.4 durch stückweise konstante (auf „Intervallen“ im $\mathbb{R}^2$) Funktionen approximiert werden.

Die Diskretisierung der Wavelet-Transformation hatten wir in Abschnitt 10.3 bereits besprochen. Es sind lediglich die zur Berechnung der Koeffizienten $c_k^{\overline{j}}$ benötigten Integrale numerisch zu approximieren.

Für eine Funktion $f \in L^2(\mathbb{R})$ bezeichnen wir mit $\underline{f}$ im Folgenden eine Approximation von f, die auf den Intervallen $[\frac{l}{n}, \frac{l+1}{n})$, $l \in \mathbb{Z}$, für vorgegebenes $n \in \mathbb{N}$ konstant ist. Mit $\underline{\underline{f}}$ bezeichnen wir entsprechend die doppelt unendliche Folge, deren l-tes Element gerade den Wert von $\underline{f}$ im Intervall $[\frac{l}{n}, \frac{l+1}{n})$ angibt. Analoges gelte für $f = (f_1, \ldots, f_N) \in L^2(\mathbb{R})^N$, d.h. es seien $\underline{f} := (\underline{f_1}, \ldots, \underline{f_N})$ und $\underline{\underline{f}} := (\underline{\underline{f_1}}, \ldots, \underline{\underline{f_N}})$.

Der zweite Schwerpunkt der Diskretisierung, neben der Handhabung der Wavelet-Transformation, liegt bei der Diskretisierung der Radon-Transformierten $R_i u$ (für $i = 1, \ldots, N$) einer Funktion $u = \sum_{\lambda \in \Lambda_{\underline{j},\overline{j}}} b_\lambda \phi_\lambda = W^* b$. Hier machen wir uns die Linearität

von R_i zunutze, d.h. wir verwenden die Beziehung

$$R_i\Big(\sum_{\lambda\in\Lambda_{\underline{j},\overline{j}}} b_\lambda\phi_\lambda\Big) = \sum_{\lambda\in\Lambda_{\underline{j},\overline{j}}} b_\lambda R_i\phi_\lambda.$$

Die Berechnung einer stückweise konstanten Approximation $\underline{R_i\phi_\lambda}$ von $R_i\phi_\lambda$ wird mittels numerischer Integration realisiert. Die diskrete Version der Radon-Transformierten erhalten wir somit aus

$$\underline{\underline{R_iu}} = \sum_{\lambda\in\Lambda_{\underline{j},\overline{j}}} b_\lambda \underline{\underline{R_i\phi_\lambda}}.$$

Die diskrete Version $\underline{\underline{F_i(W^*b)}}$ des Ausdrucks $F_i(W^*b)$ bekommt man nun einfach durch elementweises Quadrieren von $\underline{\underline{R_iu}}$.

Zur numerischen Berechnung von $\overline{WF'(W^*b)^*g}$ für eine Funktion $g = (g_1,\dots,g_N) \in (L^2(-\rho,\rho))^N$ sind zwei Schritte notwendig. Zunächst wird das elementweise Produkt von $\underline{\underline{g_i}}$ und $\underline{\underline{R_iW^*b}}$ für $i = 1,\dots,N$ gebildet; die entstehenden Folgen bezeichnen wir mit $\underline{\underline{w_1}},\dots,\underline{\underline{w_N}}$. Anschließend muss noch die Wavelet-Transformation W (die aufgrund der Ersetzung der Indexmenge Λ durch $\Lambda_{\underline{j},\overline{j}}$ nur noch endlich viele Nicht-Null-Koeffizienten liefert) auf die durch

$$z(x,y) := 2\sum_{i=1}^{N} \underline{w_i}(x\sigma_i^1 + y\sigma_i^2)$$

definierte Funktion z angewendet werden.

Die den gestörten Daten v^δ zugrunde liegenden exakten Daten $v = (v_1,\dots,v_N) \in (L(-1,1))^N$ werden ebenfalls stückweise konstant durch $\underline{v}$ approximiert. Auf die Realisierung der Datenstörung werden wir im nächsten Abschnitt eingehen.

Die Umsetzung aller weiteren Schritte in obigem Algorithmus ist elementar und bedarf lediglich aus technischer Sicht einiger Überlegungen, da die Berechnungen sowohl rechenzeit- als auch speicherintensiv sind. Auf die Feinheiten der im Rahmen dieser Arbeit entstandenen Implementierung wollen wir hier nicht näher eingehen.

13.4. Versuchsaufbau

Nach der Vorstellung des Minimierungsalgorithmus' für das Tikhonov-Funktional (13.2) und der Beschreibung der wesentlichen Diskretisierungsschritte bleiben noch einige Fragen offen, die den konkreten Ablauf der numerischen Experimente betreffen.

Als Erstes stellt sich die Frage, welche Startfolge b^0 für den Algorithmus gewählt werden sollte. Um dieses Problem zu vereinfachen, werden wir uns eine stückweise konstante Funktion u_0 auf $\mathbb{R}^2$ mit $\operatorname{supp} u_0 \subseteq \Omega$ in Form einer Matrix aus Grauwerten mit Werten im Bereich von -1 (schwarz) bis 1 (weiß) vorgeben und die Wavelet-Koeffizienten von u_0 berechnen. Die Folge dieser Wavelet-Koeffizienten wird die Startfolge b^0 sein. Günstige Wahlen der Startfunktion bzw. Startmatrix werden wir im Experiment untersuchen.

Die nächste offene Frage ist die der Simulation der Daten v bzw. v^δ. Als Ausgangspunkt geben wir analog zur Startfunktion eine stückweise konstante Funktion u auf $\mathbb{R}^2$ mit $\operatorname{supp} u \subseteq \Omega$ in Form einer Grauwertmatrix vor. Durch möglichst genaue numerische Integration berechnen wir daraus die Radon-Transformierten $R_i u$ an den Stellen $\frac{l}{n} + \frac{1}{2n}$, $l = -n, \ldots, n-1$, und erhalten somit durch Quadrieren dieser Werte eine diskrete Version $\underline{v} = (\underline{\underline{v_1}}, \ldots, \underline{\underline{v_N}})$ der exakten Daten v (die Notation mit ein- und zweifacher Unterstreichung wurde im vorhergehenden Abschnitt eingeführt).

Zur Festlegung der Aufnahmerichtungen geben wir deren Anzahl $N \in \mathbb{N}$ vor und setzen

$$\zeta_i := \left(\frac{i}{N} - \frac{1}{2N}\right)\pi, \quad i = 1, \ldots, N.$$

Die Richtungen σ_i definieren wir nun als

$$\sigma_i := (\cos\zeta_i, \sin\zeta_i), \quad i = 1, \ldots, N.$$

Wird N gerade gewählt, so ist gewährleistet, dass $\cos\zeta_i$ und $\sin\zeta_i$ von Null verschieden sind. Dies Vereinfacht die Herleitung einiger Integrationsgrenzen für die numerische Integration bei der Berechnung der Wavelet-Koeffizienten.

Zur Erzeugung der gestörten Daten $\underline{v^\delta}$ geben wir den gewünschten relativen Datenfehler $\delta_{\text{rel}} > 0$ vor und erzeugen N (theoretisch) doppelt unendliche Folgen, deren Einträge mit Index $-n$ bis $n-1$ Realisierungen einer normalverteilten Zufallsgröße mit Erwartungswert 0 und Varianz 1 sind. Alle anderen Einträge sollen Null sein. Setzen wir

$$\delta := \delta_{\text{rel}} \|\underline{v}\|_{(L^2(-1,1))^N} = \delta_{\text{rel}} \left(\frac{1}{n}\sum_{i=1}^{N}\sum_{l=-n}^{n-1} \underline{\underline{v_i}}_l^2\right)^{\frac{1}{2}}$$

und

$$\underline{\underline{v_i^\delta}} := \underline{\underline{v_i}} + \frac{\delta}{\sqrt{N}} \frac{\underline{\underline{e_i}}}{\|\underline{e_i}\|_{L^2(-1,1)}} = \underline{\underline{v_i}} + \frac{\delta}{\sqrt{N}} \frac{\underline{\underline{e_i}}}{\left(\frac{1}{n}\sum_{l=-n}^{n-1} \underline{\underline{e_i}}_l^2\right)^{\frac{1}{2}}},$$

so gilt

$$\|\underline{v_i} - \underline{v_i^\delta}\|_{L^2(-1,1)} = \left(\frac{1}{n}\sum_{\tilde{l}=-n}^{n-1} \frac{\delta^2}{N} \frac{\underline{\underline{e_i}}_{\tilde{l}}^2}{\frac{1}{n}\sum_{l=-n}^{n-1} \underline{\underline{e_i}}_l^2}\right)^{\frac{1}{2}} = \frac{\delta}{\sqrt{N}}$$

und damit

$$\|\underline{v} - \underline{v^\delta}\|_{(L^2(-1,1))^N} = \left(\sum_{i=1}^{N} \|\underline{v_i} - \underline{v_i^\delta}\|_{L^2(-1,1)}\right)^{\frac{1}{2}} = \delta.$$

Die kritischste und letzte offene Frage ist die nach der Wahl des Regularisierungs-

parameters α. Hier gibt es zahlreiche Möglichkeiten. Wir entscheiden uns aufgrund der einfachen Implementierung für das Diskrepanzprinzip. Dieses läuft wie folgt ab.

$\alpha_0 > 0$ groß genug wählen
$q \in (0,1)$ und $\tau > 0$ wählen
$l \leftarrow 0$
do
    $l \leftarrow l + 1$
    Minimierer u_{l-1} von $T^{\delta}_{\alpha_{l-1}}$ berechnen
    $\alpha_l \leftarrow q\alpha_{l-1}$
while $\|F(u_{l-1}) - v^{\delta}\|_{(L^2(-\rho,\rho))^N} \geq \tau\delta$
$u \leftarrow u_l$.

Zu beachten ist, dass α_0 so groß gewählt werden sollte, dass $\|F(u_0) - v^{\delta}\|_{(L^2(-\rho,\rho))^N} \geq \tau\delta$ gilt. Die Wahl von τ ist in der Praxis schwierig, da für zu kleines τ die Abbruchbedingung unter Umständen nie erfüllt sein wird und für zu großes τ der mittels Diskrepanzprinzip gewählte Regularisierungsparameter zu einer zu glatten Lösung führen kann.

14. Ergebnisse

Anhand der Betrachtungen in den vorangegangenen Kapiteln wurde eine Implementierung mit Hilfe von Matlab erstellt. Wie weiter oben bereits angemerkt wurde, sind die Berechnungen sehr resourcenintensiv, sodass wir die Anzahl der Wavelet-Koeffizienten, d.h. die Anzahl der freien Variablen des Minimierungsproblems, gering halten müssen. Dies kann auf zwei Wegen erfolgen, von denen wir beide umsetzen werden müssen. Zum einen kann die feinste Skalierungsstufe $\overline{j}$ klein gewählt werden; dies führt zu grob gerasterten bzw. unscharfen Ergebnissen. Zum anderen wirkt sich eine große Daubechies-Ordnung r negativ auf die Koeffizientenanzahl aus. Erste Versuche haben gezeigt, dass $\overline{j} = 5$ und $r = 0$ einen vertretbaren Kompromiss zwischen Ergebnisqualität und Resourcennutzung darstellen. Die Arbeit mit größerem r ist zwar wünschenswert, bringt aber die verfügbare Rechentechnik (3-GHz-Prozessor, 2 GB Arbeitsspeicher) an ihre Grenzen. Gegebenenfalls würde eine Implentierung in der Sprache C hier mehr Freiraum bieten.

Für die Diskretisierung im Datenraum wählen wir die Parameter $N = 40$ (Anzahl der Aufnahmerichtungen) und $n = 20$ (Stützstellenanzahl in Intervall der Länge 1). Die gröbste Skalierungsstufe der Wavelet-Zerlegung legen wir auf $\underline{j} = 0$ fest. Als Abbruchschranken für den Minimierungsalgorithmus wählen wir $\tau_1 = 10^{-3}$ und $\tau_2 = 10^{-5}$; der Parameter C wird auf 50 gesetzt. Die Wahl dieser Werte erfolgte anhand zahlreicher Testrechnungen. Insbesondere hat sich gezeigt, dass der Parameter C auch größer gewählt werden kann (z.B. $C = 100$); dies wirkt sich nicht auf das Ergebnis aus, sondern nur auf die Iterationsanzahl, genauer auf das Verhältnis der Anzahl innerer und äußerer Iterationen. Wird C zu klein gewählt, so konvergiert die innere Iteration unter Umständen nicht mehr. In den folgenden Versuchen wurden stets ca. 100 äußere Iterationen und durchschnittlich 4 bis 5 innere Iterationen pro äußerer Iteration benötigt.

14.1. Abhängigkeit vom Startwert

Bei der Modellierung des Schlierentomografieproblems hatten wir bereits festgestellt, dass mit u auch $-u$ stets eine Lösung der Gleichung $F(u) = v$ ist. Somit stehen wir vor dem Problem, zu entscheiden, welche der beiden Lösungen uns der vorgeschlagene Algorithmus liefert. In [12] wurde für ein anderes Verfahren bereits beobachtet, dass diese Frage vom für den Algorithmus verwendeten Startwert u_0 bzw. von der entsprechenden Startfolge b^0 abhängt: ein positiver konstanter Startwert liefert die Lösung mit dem korrekten Vorzeichen, ein negativer konstanter Startwert liefert die Lösung mit dem falschen Vorzeichen.

Für die Untersuchung dieser Fragestellung verwenden wir eine einfache Druckverteilung, wie in Abbildung 14.1 (b) dargestellt. Als Startwert wählen wir die mit einem Skalierungsfaktor versehene charakteristische Funktion χ_Ω des Gebietes Ω (Abbildung 14.1 (a)). Als Faktor wählen wir 0,01 und $-0{,}01$. Der Datenfehler δ sei Null; dadurch können wir zur Wahl des Regularisierungsparameters α nicht mehr das Diskrepanzprinzip anwenden. Stattdessen geben wir uns eine fallende Folge von Parametern vor und betrachten für jeden dieser Parameter die mit der entsprechenden Lösung u des Mi-

nimierungsproblems erzielte Diskrepanz $\|F(u) - v\|_{(L^2(-\rho,\rho))^N}$. Wir wählen denjenigen Parameter aus der Folge aus, der die Diskrepanz minimiert. Für beide Rekonstruktionen wird dieses Vorgehen den Regularisierungsparameter $\alpha = 10^{-8}$ ergeben. Da es an dieser Stelle nicht so sehr auf die Genauigkeit der Rekonstruktion ankommt, reicht uns hier $\overline{j} = 4$.

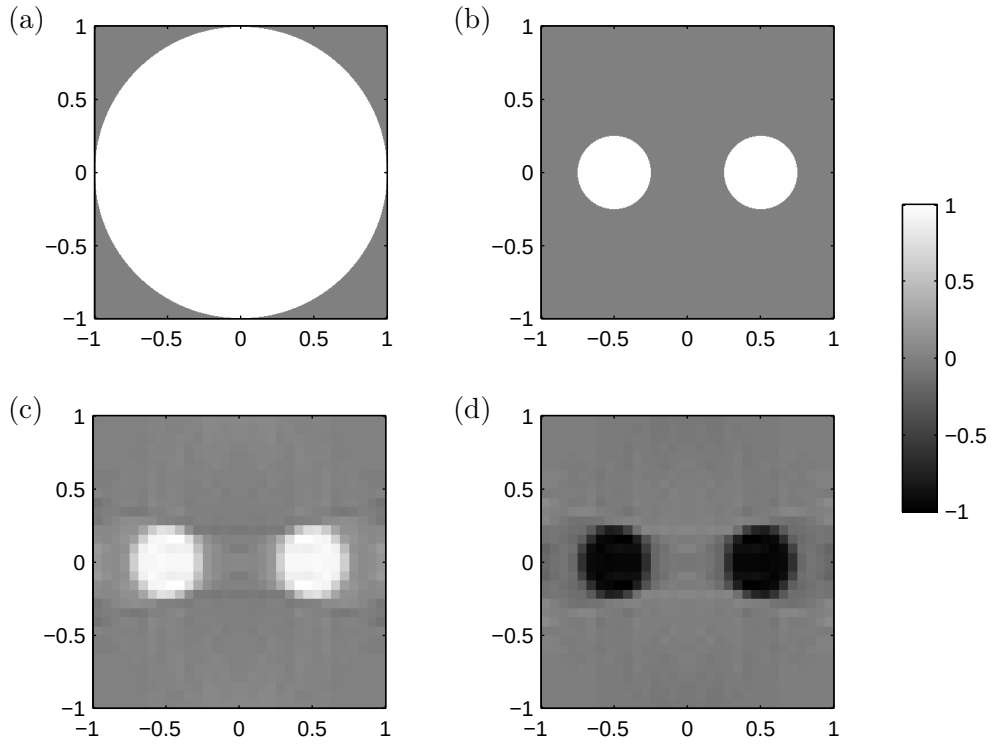

Abbildung 14.1.: (a) Funktion χ_Ω, (b) exakte Lösung u, (c) rekonstruierte Lösung mit Startwert $0{,}01\chi_\Omega$, (d) rekonstruierte Lösung mit Startwert $-0{,}01\chi_\Omega$.

Die Ergebnisse in Abbildung 14.1 (c) und (d) zeigen, dass das hier vorgestellte Verfahren das gleiche Verhalten wie in [12] beschrieben zeigt, d.h. das Vorzeichen des (konstanten) Startwerts bestimmt das Vorzeichen der Lösung. Weitere Berechnungen haben dies bestätigt.

14.2. Rekonstruktion negativer Funktionswerte

Die interessanteste Frage in Hinblick auf Verfahren zur numerischen Lösung des inversen Problems der Schlierentomografie ist die nach der Möglichkeit zur Rekonstruktion von negativen Werten der Druckverteilung u. Es stellt sich heraus, dass die Beantwortung dieser Frage stark von der Wahl der Startfolge b^0 für den Minimierungsalgorithmus

abhängt.

Zur experimentellen Beantwortung dieser Frage ziehen wir die in Abbildung 14.2 (a) dargestellte Druckverteilung heran. Den Startwert wählen wir wie im vorhergehenden Abschnitt als $0{,}01\chi_\Omega$. Da wir $\delta = 0$ setzen, müssen wir die Parameterwahl ebenfalls wie im vorhergehenden Abschnitt durchführen. Dieses Vorgehen führt zu $\alpha = 10^{-7}$.

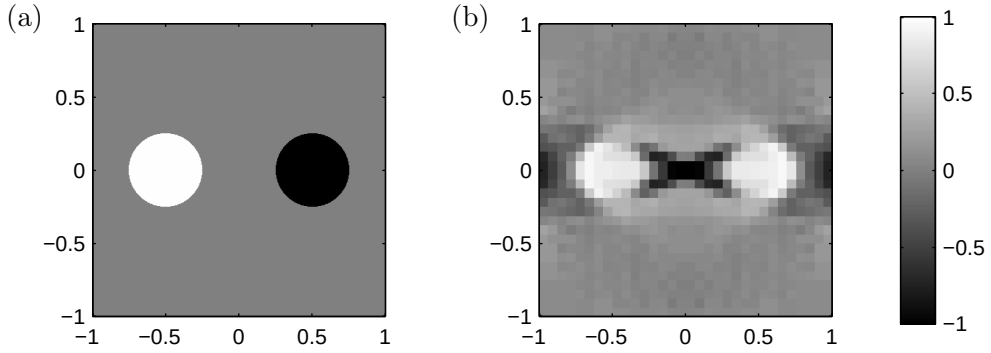

Abbildung 14.2.: (a) exakte Lösung, (b) Rekonstruktion.

In Abbildung 14.2 (b) erkennt man deutlich, dass die Druckverteilung nicht korrekt rekonstruiert wurde. Insbesondere sind die negativen Funktionswerte bei der rekonstruierten Lösung nun positiv.

Bei Verwendung eines Startwertes (siehe Abbildung 14.3 (a)), der die Verteilung von positiven und negativen Funktionswerten der exakten Lösung in gewisser Weise widerspiegelt, erfolgt eine korrekte Rekonstruktion. Die rekonstruierte Lösung ist in Abbildung 14.3 (b) dargestellt.

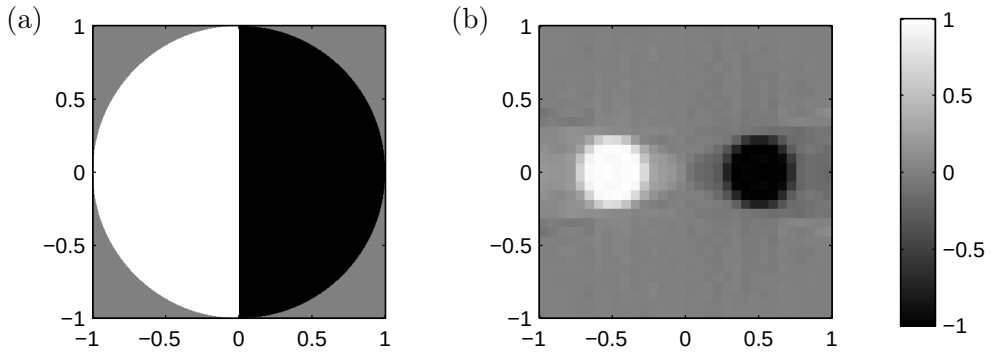

Abbildung 14.3.: (a) Startwert (mit Faktor 100 multipliziert), (b) Rekonstruktion.

Weitere Berechnungen haben die Beobachtungen bestätigt. Als Fazit ist also fest-

zuhalten, dass das hier verwendete Verfahren im Allgemeinen nicht in der Lage ist, Druckverteilungen mit negativen Funktionswerten korrekt zu rekonstruieren. Es gibt jedoch geeignete Startwerte, die zu einer korrekten Rekonstruktion führen.

14.3. Abhängigkeit vom Datenfehler

Da wir im vorhergehenden Abschnitt feststellen mussten, dass das verwendete Verfahren nicht in der Lage ist, negative Funktionswerte ordnungsgemäß zu rekonstruieren, werden wir in den folgenden Experimenten nur noch Druckverteilungen heranziehen, die keine negativen Werte aufweisen.

Wir wollen nun das Verhalten des Verfahrens bei gestörten Daten untersuchen. Dazu verwenden wir die in Abbildung 14.4 (a) dargestellte Druckverteilung. Da wir $\overline{j} = 5$ setzen, was aus praktischer Sicht ein recht niedriger Wert ist, projizieren wir diese Funktion jedoch zunächst in den durch die Basiselemente aufgespannten Raum, d.h. wir approximieren sie durch die in Abbildung 14.4 (b) dargestellte Druckverteilung. Dies hat den Vorteil, dass nun eine exakte Rekonstruktion theoretisch möglich ist, falls $\delta = 0$ gewählt wird.

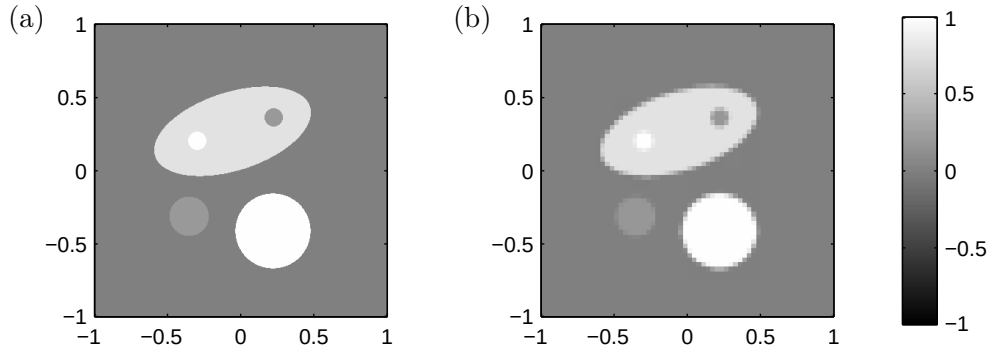

Abbildung 14.4.: (a) exakte Druckverteilung, (b) in Wavelet-Raum projizierte Druckverteilung.

Das Sinogramm der in den Wavelet-Raum projizierten Druckverteilung ist in Abbildung 14.5 (a) dargestellt (ein Sinogramm ist die spaltenweise Anordnung der Projektionen zu den verschiedenen Richtungen). Abbildung 14.5 (b) zeigt das mit $\delta_{\text{rel}} = 0{,}2$ gestörte Sinogramm.

In Abbildung 14.6 (a) bis (c) sind Rekonstruktionen für verschieden große Datenfehler dargestellt. Abbildung 14.6 (d) zeigt das Sinogramm der rekonstruierten Druckverteilung bei $\delta_{\text{rel}} = 0{,}2$. Für die Parameterwahl mittels Diskrepanzprinzip wurde $\tau = 1{,}2$ gewählt. Der Startwert für den Minimierungsalgorithmus wurde wie in Abschnitt 14.1 als $0{,}01\chi_\Omega$ gewählt.

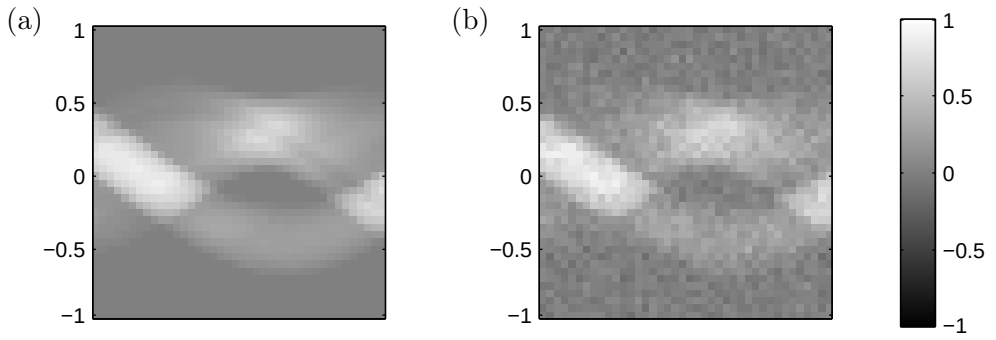

Abbildung 14.5.: (a) Sinogramm der in den Wavelet-Raum projizierten Druckverteilung, (b) Sinogramm mit Datenfehler $\delta_{\text{rel}} = 0{,}2$.

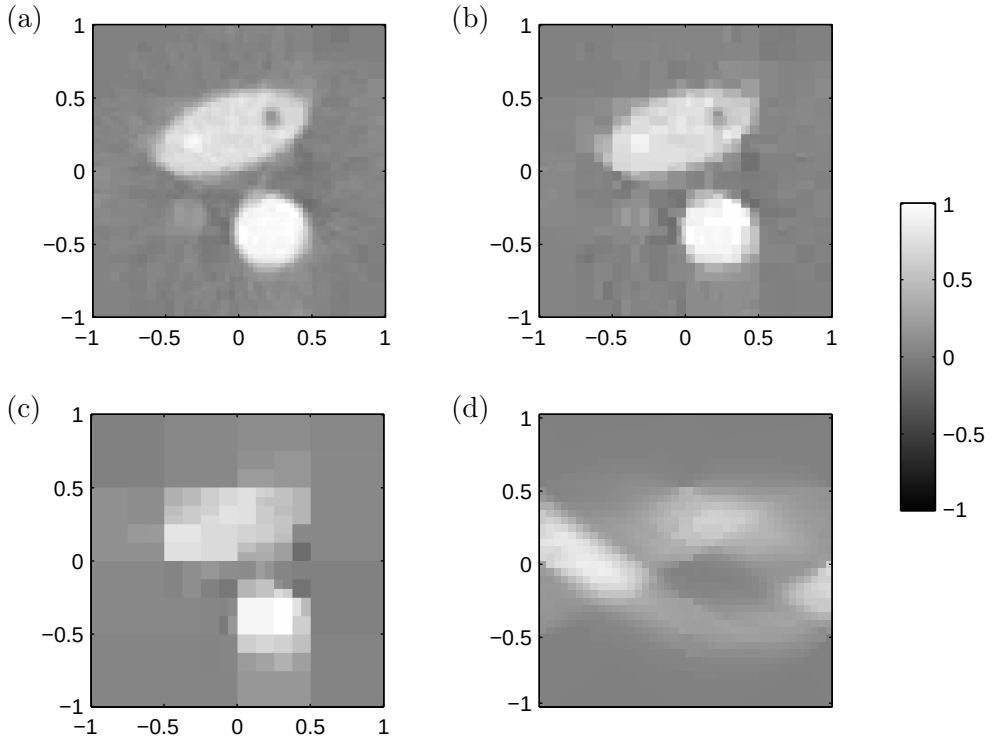

Abbildung 14.6.: (a) Rekonstruktion mit $\delta_{\text{rel}} = 0{,}05$ ($\alpha = 10^{-3}$), (b) Rekonstruktion mit $\delta_{\text{rel}} = 0{,}1$ ($\alpha = 10^{-2}$), (c) Rekonstruktion mit $\delta_{\text{rel}} = 0{,}2$ ($\alpha = 10^{-3}$), (d) Sinogramm der Rekonstruktion mit $\delta_{\text{rel}} = 0{,}2$.

Die Abbildungen zeigen, dass das Verfahren trotz großer Datenfehler in der Lage ist, die ursprüngliche Druckverteilung in gewissem Rahmen zu rekonstruieren. Auffällig ist, dass bei höherem Datenfehler die Rekonstruktion aus wenigen großen Quadraten zusammengesetzt ist. Die Quadrate entstehen aufgrund der Wahl $r = 0$ (Haar-Wavelets). Die geringe Anzahl kommt daher, dass der Regularisierungsparameter α recht groß gewählt wurde. Für höhere Daubechies-Ordnung r sind ähnliche Resultate zu erwarten, wobei die scharfen Kanten der Quadrate durch sehr weiche Übergänge ersetzt werden.

14.4. Konvergenzraten

Im letzen numerischen Experiment soll versucht werden, Konvergenzraten für $\delta \to 0$ zu berechnen. D.h. wir betrachten $\varepsilon(\delta) := \|u - \tilde{u}\|_{L^2(\Omega)}$ in Abhängigkeit von δ, wobei u die in den Wavelet-Raum projizierte exakte Druckverteilung ist (vgl. vorhergehenden Abschnitt) und $\tilde{u}$ die rekonstruierte Druckverteilung bezeichnet.

Wir berechnen $\varepsilon(\delta)$ für $\delta_{\text{rel}} = 0{,}1 \cdot 2^{-l}$ mit $l = 0, 1, 2, \ldots$ bzw. für die entsprechenden absoluten Datenfehler δ. Alle weiteren Parameter wählen wir wie im vorhergehenden Abschnitt. Insbesondere verwenden wir die Druckverteilung aus Abbildung 14.4 (b) und die Funktion $0{,}01\chi_\Omega$ als Startwert.

Es stellt sich heraus, dass für $\delta_{\text{rel}} \leq 0{,}0125$ die Abbruchbedingung für die Parameterwahl mittels Diskrepanzprinzip für kein α erfüllt wird, da sich die erzielte Diskrepanz für kleiner werdendes α kaum noch verringert. Die Werte $\varepsilon(\delta)$ sind für einige δ in Abbildung 14.7 dargestellt, wobei der Regularisierungsparameter α, falls möglich, mittels Diskrepanzprinzip gewählt wurde; ansonsten wurde α so gewählt, dass die Diskrepanz unter allen betrachtenen Werten für α minimal wird.

Die Abbildung zeigt deutlich, dass sich $\varepsilon(\delta)$ einem von Null verschiedenen Wert bei $\delta \to 0$ annähert. Zu Erwarten war eigentlich $\varepsilon(\delta) \to 0$. Es liegt die Vermutung nahe, dass dieses Verhalten durch eine zu kleine Wahl von n und N hervorgerufen wird. Die Wiederholung der Berechnungen mit verdoppeltem n brachte jedoch das selbe Ergebnis.

Eine andere Ursache für das ungewöhnliche Verhalten des Fehlers $\varepsilon(\delta)$ kann in der Parameterwahl mittels Diskrepanzprinzip vermutet werden. Bei genauerer Betrachtung der Berechnungsergebnisse hat sich jedoch gezeigt, dass für sehr kleine Parameter α nicht nur die Diskrepanz nahezu unverändert bleibt, sondern auch die Lösung für kleiner werdendes α kaum Veränderungen erfährt. Dieses Phänomen konnte trotz zahlreicher weiterer, hier nicht widergegebener Berechnungen im Rahmen dieser Arbeit nicht aufgeklärt werden.

14.5. Fazit

Die durchgeführten Experimente haben gezeigt, dass das in dieser Arbeit vorgestellte Verfahren zur numerischen Lösung des inversen Problems der Schlierentomografie nur bedingt in der Lage ist, die gestellten Anforderungen, insbesondere in Hinblick auf die Rekonstruktion von negativen Funktionswerten, zu erfüllen. Für den praktischen Einsatz

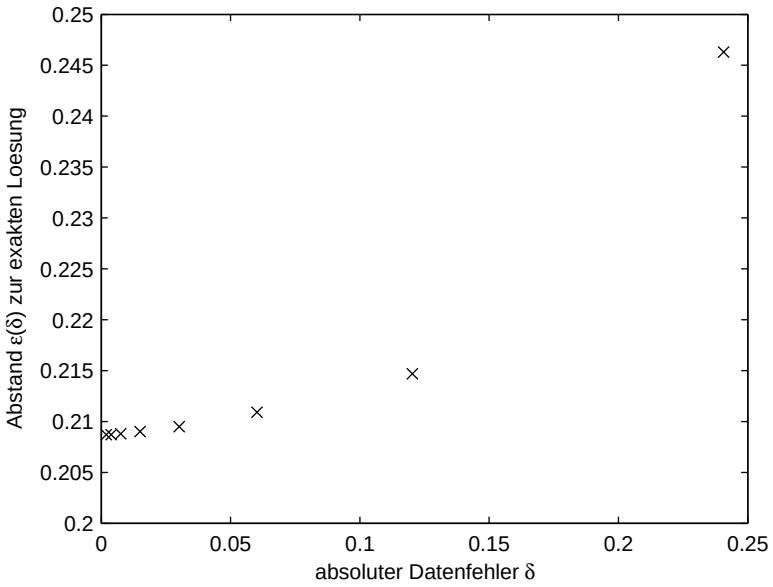

Abbildung 14.7.: Abhängigkeit des Lösungsfehlers vom Datenfehler.

ist das Verfahren nicht grundsätzlich ungeeignet, es müssen jedoch Einschränkungen in Kauf genommen werden.

Die Implementierung ist aufgrund der Arbeit mit Wavelets im Vergleich zu anderen Verfahren (z.B. Gradientenverfahren bei differenzierbarem Tikhonov-Funktional) sehr aufwändig und somit auch fehleranfällig. Um den hohen Resourcenbedarf zu verringern, ist eventuell eine Implementierung mit Hilfe einer geeigneteren Programmiersprache (z.B. C) vorteilhaft. Zudem bestehen noch Einsparmöglichkeiten bei der Speichernutzung durch eine konsequentere Umsetzung des Sparsity-Konzepts, d.h. durch den Verzicht auf die Speicherung zahlreicher Null-Koeffizienten. Der Resourcenbedarf soll hier also nicht zu negativ bewertet werden.

Zusammenfassend ist somit festzuhalten, dass der in dieser Arbeit eingeschlagene Weg zwar prinzipiell zum Ziel führt, allerdings weniger geeignet ist als andere Ansätze zur Lösung des Schlierentomografieproblems; in diesem Zusammenhang sei nochmals auf [29] und [12] verwiesen.

Danksagung

An dieser Stelle möchte ich mich bei der Arbeitsgruppe um Prof. Dr. Otmar Scherzer an der Universität Innsbruck für die Ermöglichung eines mehrtägigen Aufenthalts an ihrer Einrichtung bedanken. Insbesondere gilt mein Dank Dr. Christiane Pöschl für die herzliche Betreuung während des Aufenthalts, Prof. Dr. Otmar Scherzer für seine Bemerkungen zu den frühen Ideen zum ersten Teil der Arbeit, sowie Prof. Dr. Otmar Scherzer, Dr. Richard Kowar und Dr. Markus Haltmeier für die Anregung und Unterstützung zur Bearbeitung des Schlierentomografieproblems im zweiten Teil der Arbeit.

Besonders bedanken möchte ich mich bei dem Betreuer der Arbeit Prof. Dr. Bernd Hofmann für die sehr intensive, anregende und motivierende Betreuung dieser Diplomarbeit sowie für seine stete Unterstützung meines persönlichen und fachlichen Fortkommens über die letzten drei Jahre.

Symbolverzeichnis

An dieser Stelle werden alle in der Arbeit verwendeten Symbole aufgeführt, die aufgrund ihrer Verbreitung in der Mathematik im Text nicht explizit eingeführt wurden.

$\mathbb{N}$	Menge der natürlichen Zahlen $1, 2, 3, \ldots$
$\mathbb{N}_0$	Menge der natürlichen Zahlen mit Null, also $0, 1, 2, 3, \ldots$
$\mathbb{Z}$	Menge der ganzen Zahlen $\ldots, -2, 1, 0, 1, 2, \ldots$
$\mathbb{R}$	Menge der reellen Zahlen
$C_0^\infty(M)$	Menge aller auf der offenen Menge M beliebig oft differenzierbaren, reellwertigen Funktionen mit kompaktem Träger
$L^p(M)$	Menge aller reellwertigen Funktionen auf M, deren p-te Potenz auf M Lebesgue-integrierbar ist
$H_0^1(M)$	Abschluss von $C_0^\infty(M)$ in $L^2(M)$
$W^{1,1}(M)$	Menge aller Funktionen aus $L^1(M)$, deren erste verallgemeinerte Ableitung ebenfalls in $L^1(M)$ liegt

Literaturverzeichnis

[1] Bäni, W.: *Wavelets. Eine Einführung für Ingenieure.* Oldenbourg Verlag München Wien, 2002.

[2] Bartsch, R.: *Allgemeine Topologie I.* Oldenbourg Verlag München Wien, 2007.

[3] Bonesky, T., Bredies, K., Lorenz, D. A. und Maass, P.: *A generalized conditional gradient method for nonlinear operator equations with sparsity constraints.* Inverse Problems, 23(5):2041–2058, 2007.

[4] Chan, T. F. und Shen, J.: *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods.* Society for Industrial and Applied Mathematics Philadelphia, 2005.

[5] Choi, H. und Baraniuk, R. G.: *Wavelet statistical models and Besov spaces.* In: Unser, M. A., Aldroubi, A. und Laine, A. F. (Herausgeber): *Wavelet Applications in Signal and Image Processing VII*, Band 3813 der Reihe *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Seiten 489–501, 1999.

[6] Daubechies, I.: *Ten Lectures on Wavelets.* Nummer 61 in *CBMS-NSF Regional Conference Series in Applied Mathematics.* Society for Industrial and Applied Mathematics Philadelphia, 1992.

[7] Daubechies, I., Defrise, M. und de Mol, C.: *An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint.* Communications on Pure and Applied Mathematics, 57:1413–1457, 2004.

[8] Düvelmeyer, D., Hofmann, B. und Yamamoto, M.: *Range inclusions and approximate source conditions with general benchmark functions.* Numerical Functional Analysis and Optimization, 28(11-12):1245–1261, 2007.

[9] Engl, H. W., Hanke, M. und Neubauer, A.: *Regularization of Inverse Problems.* Kluwer Academic Publishers Dordrecht Boston London, 1996.

[10] Engl, H. W., Kunisch, K. und Neubauer, A.: *Convergence rates for Tikhonov regularization of non-linear ill-posed problems.* Inverse Problems, 5(4):523–540, 1989.

[11] Goswami, J. C. und Chan, A. K.: *Fundamentals of Wavelets.* Wiley-Interscience, 1999.

[12] Haltmeier, M., Kowar, R., Leitão, A. und Scherzer, O.: *Kaczmarz methods for regularizing nonlinear ill-posed equations II: Applications.* Inverse Problems and Imaging, 1(3):507–523, 2007.

[13] Hanafy, A. und Zanelli, C. I.: *Quantitative real-time pulsed Schlieren imaging of ultrasonic waves.* In: *Ultrasonics Symposium, 1991. Proceedings., IEEE 1991*, Band 2, Seiten 1223–1227, Dezember 1991.

[14] Hefter, A.: *Wie groß ist die Schallkeule bei Sonar-Näherungsschaltern?* Siemens AG, 2005. Produktinformationsbroschüre.

[15] Hein, T. und Hofmann, B.: *Approximate source conditions for nonlinear ill-posed problems – chances and limitations.* Inverse Problems, 25(3):035003 (16pp), 2009.

[16] Hofmann, B.: *Interplay of source conditions and variational inequalities for nonlinear ill-posed problems.* Preprint 6/2009, Preprintreihe der Fakultät für Mathematik Technische Universität Chemnitz.

[17] Hofmann, B.: *A note on convergence rates for variational regularization with nonconvex residual term.* Preprint 11/2009, Preprintreihe der Fakultät für Mathematik Technische Universität Chemnitz.

[18] Hofmann, B.: *Approximate source conditions in Tikhonov-Phillips regularization and consequences for inverse problems with multiplication operators.* Mathematical Methods in the Applied Sciences, 29(3):351–371, 2006.

[19] Hofmann, B., Kaltenbacher, B., Pöschl, C. und Scherzer, O.: *A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators.* Inverse Problems, 23(3):987–1010, 2007.

[20] Krehl, P. und Engemann, S.: *August Toepler – The first who visualized shock waves.* Shock Waves, 5(1-2):1–18, 1995.

[21] The MathWorks: *MATLAB Documentation*, 2004.

[22] Messerschmid, R.: *Zerstäubung von Flüssigkeiten in einer nicht angepassten Lavaldüse.* Doktorarbeit, Universität Bonn, 2004.

[23] Meyer, Y.: *Wavelets and operators.* Cambridge University Press, 1992.

[24] Morneburg, H. (Herausgeber): *Bildgebende Systeme für die medizinische Diagnostik.* Publicis MCD Verlag, Erlangen, 3. Auflage, 1995.

[25] Pöschl, C.: *Tikhonov Regularization with General Residual Term.* Doktorarbeit, Leopold Franzens Universität Innsbruck, 2008.

[26] Ramlau, R. und Teschke, G.: *A Tikhonov-based projection iteration for nonlinear Ill-posed problems with sparsity constraints.* Numerische Mathematik, 104(2):177–203, 2006.

[27] Reichel, E. K., Schneider, S. C. und Zagar, B. G.: *Characterization of Ultrasonic Transducers Using the Schlieren-Technique.* In: *Instrumentation and Measurement Technology Conference, 2005. IMTC 2005. Proceedings of the IEEE*, Band 3, Seiten 1956–1960, Mai 2005.

[28] Rinow, W.: *Lehrbuch der Topologie.* VEB Deutscher Verlag der Wissenschaften Berlin, 1975.

[29] Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M. und Lenzen, F.: *Variational Methods in Imaging.* Nummer 167 in *Applied Mathematical Sciences.* Springer New York, 2009.

[30] Stark, H.-G.: *Wavelets and Signal Processing.* Springer Berlin Heidelberg New York, 2005.

[31] Tikhonov, A. N., Leonov, A. S. und Yagola, A. G.: *Nonlinear Ill-Posed Problems*, Band 1 und 2. Chapman Hall London, 1998.

[32] Triebel, H.: *Theory of Function Spaces II.* Birkhäuser Verlag Basel Boston Berlin, 1992.

[33] Triebel, H.: *The Structure of Functions.* Birkhäuser Verlag Basel Boston Berlin, 2001.

[34] Triebel, H.: *Theory of Function Spaces III.* Birkhäuser Verlag Basel Boston Berlin, 2006.

[35] Waldron, B. und Kauth, G.: *Ultraschall-Array-Technik für industrielle Anwendungen.* ZfP-Zeitung, 69, März 2000.

[36] Weidmann, J.: *Lineare Operatoren in Hilberträumen.* B. G. Teubner Stuttgart Leipzig Wiesbaden, 2000.

[37] Werner, D.: *Funktionalanalysis.* Springer Berlin Heidelberg New York, 6. Auflage, 2007.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig angefertigt, nicht anderweitig zu Prüfungszwecken vorgelegt und keine anderen als die angegebenen Hilfsmittel verwendet habe. Sämtliche wissentlich verwendete Textausschnitte, Zitate oder Inhalte anderer Verfasser wurden ausdrücklich als solche gekennzeichnet.

Chemnitz, den 22. Juni 2009

Jens Geißler