



Fachbereich VI Informatik und Medien

Masterarbeit

von

Radmir Gesler

Zur Erlangung
des akademischen Grades
Master of Engineering (M.Eng.)

Im Studiengang
Technische Informatik

Thema:

Physics Informed Neural Networks (PINN) zur Lösung eines inversen Problems der
Wärmeleitungsgleichung

Betreuer: Prof. Dr. Frank Haußer
Betreuer/in extern: —
Gutachter: —

Eingereicht am: 20. August 2019

Inhaltsverzeichnis

1. Einführung	1
1.1. Kurze Zusammenfassung über Wärmeübertragung	1
1.1.1. Wärmeleitung	1
1.1.2. Wärmekonvektion	1
1.1.3. Wärmestrahlung	1
1.2. Das Problem	2
2. Wärmeleitungsproblem	3
2.1. Herleitung der Wärmeleitungsgleichung	3
2.1.1. Analytische Untersuchung der Wärmeleitungsgleichung	6
2.2. Inverses Problem der Wärmeleitung	6
2.3. Numerische Lösungsverfahren der Wärmeleitungsgleichung	9
2.3.1. Herleitung der Variationsformulierung der Wärmegleichung . .	9
2.4. Geometrie und Gitter definieren	9
2.4.1. Warum Geometrie und Gitter definieren?	10
2.4.2. Beispiel für ein rechteckiges Gebiet	10
2.5. Funktionsräume definieren	10
2.5.1. Was sind P_1 - und P_2 -Elemente?	10
2.5.2. Warum Funktionsräume definieren?	10
2.6. Variationsformulierung herleiten	10
2.6.1. Herleitung der schwachen Form	10
2.7. Zeitdiskretisierung	11
2.7.1. Warum Zeitdiskretisierung?	11
2.7.2. Backward-Euler-Verfahren	11
2.8. Schlüsselkonzepte	12
3. Deep Learning Methode	13
3.1. Statistische Lernverfahren: Grundlagen	13
3.1.1. Definition des statistischen Lernens	13
3.1.2. Lineare Modelle und Regression	14
3.1.3. Nichtlineare Modelle und neuronale Netze	14
3.2. Einführung in Physics-Informed Neural Networks	15
3.2.1. Motivation für PINNs	15
3.2.2. Definition von PINNs	15
3.2.3. Beispiel: Wärmeleitungsgleichung	15
3.3. Nichtlineare und gekoppelte Probleme	20
3.4. Das des Computertomographen	21
3.5. Die Radon Transformation	23
3.5.1. Kleiner Einschub über L^2 -Raum	26
3.5.2. Eigenschaften der Radon Transformation	28

A. Implementierung	35
A.1. Radon Transformation	35
A.2. Ungefilterte Rückprojektion	36
A.3. Gefilterte Rückprojektion	36
A.4. Erzeugung der Systemmatrix von \mathcal{R}	37
A.5. Iterative Rekonstruktion nach Kaczmarz	37
A.6. Rekonstruktion durch SWZ	38
B. Inhalt der CD-ROM	40
B.1. PDF-Dateien	40
B.2. MATLAB Funktionen	40
Quellenverzeichnis	41
Literatur	41
Online-Quellen	41

1. Einführung

1.1. Kurze Zusammenfassung über Wärmeübertragung

Die Beobachtung der Naturvorgänge, die mit dem Transport der Wärme zu tun haben, hat zu einer Feststellung geführt. Und zwar, falls ein Medium zwei Gebiete mit unterschiedlichen Temperaturen besitzt, bewegt sich die Wärme immer aus dem Gebiet mit der höheren Temperatur in das, wo eine niedrigere Temperatur herrscht. Dieser Vorgang der Wärmeübertragung kann auf drei Arten geschehen, durch sogenannte Wärmeleitung, durch Wärmekonvektion und durch Wärmestrahlung. Im folgendem wollen wir diese Begrifflichkeiten zuerst klären.

1.1.1. Wärmeleitung

Man stelle sich vor, wie ein Metallstab an einem Ende aufgeheizt wird. Nach einiger Zeit würde man an dem anderem Ende eine Temperaturerhöhung messen können. Im Umkehrschluss hieße es, dass die Wärme von dem Stabende, wo eine Wärmequelle angelegt wurde, zu dem anderem gewandert ist. Eine physikalische Erklärung dafür könnte so lauten, die Atome, wo eine Wärmequelle angelegt ist, fangen an stärker zu schwingen, also bekommen mehr Energie. Durch die Gitterstruktur im Metall regen die schon stärker schwingende Atome ihre Nachbarn an und so weiter, bis die Atome an dem anderem Ende auch stärker schwingen. Somit wird die Wärmeenergie in Form schwingender Atome durch einen Festkörper transportiert, in solchen Fällen spricht man von *Wärmeleitung*.

1.1.2. Wärmekonvektion

Bei der *Konvektion* handelt es sich um Transport der Stoffmengen. Dieser Prozess findet in festen, flüssigen und gasförmigen Materialzuständen statt. Zum Beispiel bei Flüssigkeiten und Gasen tritt eine Ausgleichsströmung durch den Dichteunterschied zwischen warmen und kühleren Gebieten auf. In solchen Fällen sind die wärmeren Gebiete spezifisch leichter und steigen in die Höhe, wo sie sich dann abkühlen. Der beschriebene Verlauf ist ein Fall für sogenannte *freie Konvektion*. Eine *erzwungene Konvektion* liegt dann vor, wenn eine äußere Einwirkung dem System hinzugefügt wird. Hierfür könnte eine Vulkaneruption als Beispiel dienen.

1.1.3. Wärmestrahlung

Im Vakuum, da wo keine Materie vorhanden ist, kann die Wärmeübertragung weder durch Wärmeleitung noch durch Konvektion stattfinden. Nehmen wir ein Beispiel der Wärmeübertragung von der Sonne zur Erde an. Trotz der Tatsache, dass der

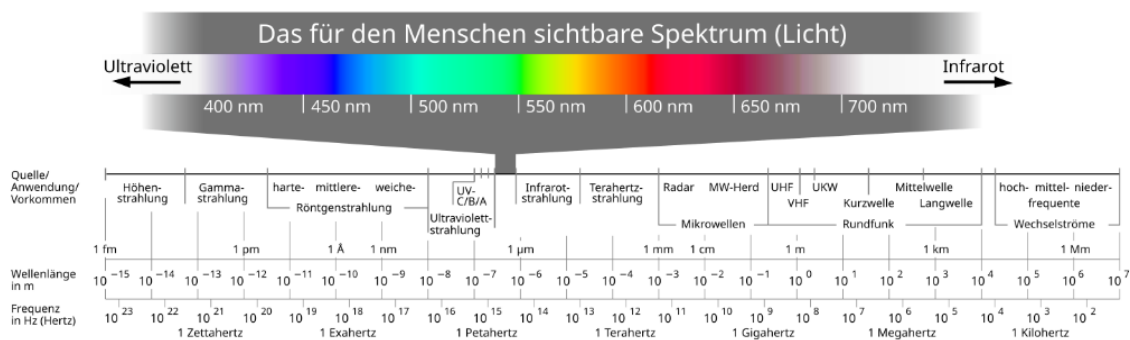


Abbildung 1.1.: Das Spektrum der elektromagnetischen Strahlung. Quelle: Wikipedia.

Raum zwischen der Erde und Sonne (fast) leer von Materie ist, findet die Wärmeübertragung von der Sonne auf die Erde statt. Es ist die *Wärmestrahlung*, die für die Energieübertragung sorgt. Im Allgemeinen werden solche Prozesse durch das Konzept der elektromagnetischen Strahlung erklärt. Fakt ist, dass die elektromagnetische Strahlung kein Medium braucht um sich im Raum auszubreiten. Strahlung an sich kann mittels Wellen mathematisch beschrieben werden. Wellen schwingen, alles was schwingt hat eine Frequenz und kann daher in einem Frequenzspektrum eingeordnet werden. So ist die Wärmestrahlung nur ein Teil des gesamten elektromagnetischen Spektrums, wobei die UV-Strahlung, das sichtbare Spektrum und die Infrarot-Strahlung die Wärmestrahlung ausmachen (Abb.1.1). Tritt die Wärmestrahlung auf Materie, so wird sie absorbiert, wodurch sich die Materie aufwärmt oder gar erhitzt.

1.2. Das Problem

Die vorangegangenen Überlegungen werden unten im Text herangezogen, um die theoretische Herangehensweise zum Problem der Wärmeausbreitung in Stoffen bereitzustellen ... bla bla ... was wird gemacht:

1. Überlegungen, die aus physikalischen Gegebenheiten auf eine Gleichung Führen, im Kapitel 2.
2. Gleichung benennen, als Wärmegleichung, ev. Kapitel so und so.
3. Mögliche Problemstellungen benennen, direkte und inverse ..., ev. Kapitel so und so
4. Das Vorhaben benennen. Mit NN inverses Problem lösen, bla bla
5. Die Struktur des Vorhabens schildern: FEM -> Lösung -> einige Werte der Lösung als Input für NN, , ev. Kapitel so und so
6. Vergleich der Ergebnisse, ev. Kapitel so und so

Bla Bla ... somit ist das Ziel der vorliegenden Arbeit festgelegt worden

2. Wärmeleitungsproblem

2.1. Herleitung der Wärmeleitungsgleichung

Im Allgemeinen beschreibt die Wärmegleichung die Ausbreitung der Wärme in einem Medium. Das Medium kann beispielsweise ein räumlicher Gegenstand sein, zum Beispiel eine rechteckige Metallplatte. Man stelle sich vor, die gegebene Platte wird in der Mitte kurz erhitzt. Diesen Prozess kann man mithilfe einer Wärmebildkamera sichtbar machen.

Würde man diese Situation über einen Zeitraum beobachten, so würde man feststellen, dass die Temperatur in der Mitte der Platte sich über die gesamte Fläche verteilt. Und irgendwann kommt es zum Abkühlen, also zur Gleichverteilung der Temperatur in der Platte.

Der beschriebene Prozess kann mithilfe sogenannter Wärmeleitungsgleichung modelliert werden, die wir zunächst herleiten möchten.

Wärmeausbreitung

Sei $\Omega \subset \mathbb{R}^3$ das Gebiet auf dem die Verteilung der Temperatur bestimmt werden soll.¹ Die gesuchte Temperaturverteilung auf Ω in einem Zeitintervall $T = [t_0, t_1] \subset \mathbb{R}$ bezeichnen wir mit u . u ordnet jedem Punkt $x \in \Omega$ zu jedem Zeitpunkt $t \in T$ eine Zahl $u(x, t) \in \mathbb{R}$ zu, also $u : \Omega \times T \rightarrow \mathbb{R}$. Zusätzlich wollen wir die Bewegung der Wärme in Ω auffassen, es ist der Wärmefluss j . Für jeden Punkt $x \in \Omega$ zu jedem Zeitpunkt $t \in T$ gibt der Wärmefluss die Richtung an, in der die Wärmeenergie transportiert wird, somit $j : \Omega \times T \rightarrow \mathbb{R}^3$. Damit sind nötigen *Variablen* für die Modellierung festgelegt.

Für die Herleitung der Gleichungen der Modellierung wird ein beliebiges Volumen $V \subset \Omega$ betrachtet. In einem homogenen Fall, also es gibt keine Wärmequellen, die in das Volumen V die Wärmeenergie hinein oder heraus transportieren, würde man keinen Wärmestrom am Rand dV detektieren, es gilt also

$$\int_{\partial V} \langle j, v \rangle dS = 0. \quad (2.1)$$

Mit der Gleichung 2.1 wird der Strom j in Richtung jeder Normalen v am Rand dV von V aufsummiert (Abb.2.1). Und mit der Annahme, dass es keine Quellen gibt, erwarten wir eine 0 Summe.

¹ Im Allgemeinen gilt $\Omega \subset \mathbb{R}^n$.

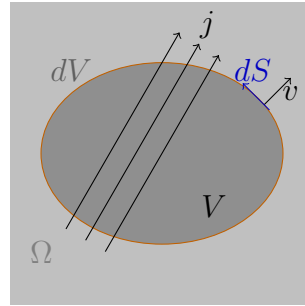


Abbildung 2.1.: Schematische Darstellung der Gleichung 2.1, mit V das Volumen eines Mediums, dV Rand (orange) von V , j Der Wärmestrom durch V , $v \in \mathbb{R}^3$ Normalenvektor an dV und dS Tangente an dV .

An dieser Stelle benötigen wir den *Gauß'schen Satz* um das Oberflächenintegral 2.1 in ein Volumenintegral umzuschreiben.

Definition 1. Der Gauß'sche Satz

Sei $V \subset \mathbb{R}^n$ kompakt mit abschnittsweise glattem Rand $S = \partial V$, der Rand ist orientiert durch äußeres Normaleneinheitsvektorfeld n . Das Vektorfeld F sei stetig differenzierbar auf einer offenen Menge Ω mit $V \in \Omega$. Dann gilt

$$\int_V \langle \nabla, F \rangle dV = \int_{\partial V} \langle F, n \rangle dS. \quad (2.2)$$

Temporäres DenkZettel

Definition 2. Kompakte Menge

\mathbb{R}^n ist metrischer Raum. Wenn es für jede **offene Überdeckung** von $A \subset \mathbb{R}^n$, eine endliche Überdeckung gibt, also

$$A \subset \bigcup_{i \in I} U_i, \quad \text{mit endlichem } I$$

Definition 3. Glatter Rand

Sei $A \in \mathbb{R}^n$ kompakt, A hat einen glatten Rand, wenn es zu jedem Punkt $x \in \partial A$ eine offene Umgebung $U \subseteq \mathbb{R}^n$ und $\psi : U \rightarrow \mathbb{R}$, $\psi \in C^n$ so gibt, dass

- (a) $A \cap U = \{x \in U : \psi(x) \leq 0\}$
- (b) $\forall x \in U : \psi'(x) \neq 0$

gilt.

Nun, der Gauß'sche Satz angewandt auf Gleichung 2.1 liefert folgendes Ergebnis

$$\int_V \langle \nabla, j \rangle dx = \int_{\partial V} \langle j, v \rangle dS = 0 \quad (2.3)$$

Da, das Volumen V nicht näher spezifiziert war, können wir aus 2.3 folgern, dass **Dahmen/Reusken: S.456, sinnvolle Argumentation nach dem Gaußschen Satz**

$$\langle \nabla, j \rangle = 0 \quad (2.4)$$

auf ganz Ω gilt.

Als Nächstes, wollen wir herausfinden, wie j und u von einander abhängen. Eine einfache Modellannahme gibt uns das *Fourier'sche Gesetz*:²

Die Wärmeenergie strömt immer von Wärmeren Regionen zu den kälteren. Die Flussrate des Stroms ist proportional zum Temperaturgradienten.

Eine weitere Annahme für das Modell ist, dass jedes Material eine Eigenschaft hat, die Wärme auf seine Art und Weise zu leiten. Diese Eigenschaft nennt man die *Wärmeleitfähigkeit* a . Sie ordnet jedem Punkt im Körper eine positive Zahl zu, also $a : \Omega \rightarrow \mathbb{R}_+$. Führen wir beide Überlegungen zusammen und erhalten

$$j = -a \nabla u. \quad (2.5)$$

Einsetzen der Gleichung 2.5 in 2.4 führt uns zum folgenden Ausdruck

$$-\langle \nabla, a \nabla u \rangle = 0 \quad (2.6)$$

Falls auf Ω noch zusätzliche Wärmequellen $f : \Omega \rightarrow \mathbb{R}$ vorhanden sind und die Wärmeverteilung *instationär* ist, erhalten wir insgesamt folgende Gleichung [6, S.5]

$$c\rho \partial_t u - \langle \nabla, a \nabla u \rangle = f \quad (2.7)$$

c spezifische Wärmekapazität [J/Kkg]
 ρ - Einheit [kg/m^3]

Einschub über die Klassifikation der DGL

Nie Gleichung 2.7 beschreibt ein abstraktes System. Die sogenannten Anfangs- und Randbedingungen würden eine mehr der Realität entsprechende Beschreibung des Wärmeleitungsproblems liefern. Jedes physikalische System besitzt zum Anfang der Messung oder der Beobachtung einen gewissen Anfangszustand u_0 . Meistens wird die Zeit t zu diesem Anfang auf den Wert 0 gesetzt, somit erhalten wir

$$u(x, 0) = u_0(x), \quad x \in \Omega \quad (2.8)$$

² J. B. Fourier, 1768-1830

Die Randbedingungen geben die Temperaturverteilung u am Rand $\partial\Omega$ des Gebiets Ω und können im Allgemeinen mit Sturm'schen Randbedingungen angegeben werden (oder gemischte Randbedingung?):

$$u(x, t) = g(x, t), \quad x \in \partial\Omega \quad (2.9)$$

Nun kann man die Gleichungen 2.7, 2.8 und 2.9 zusammen in einer Rand- und Anfangswertaufgabe (RAWP) der Wärmeleitung zusammenfassen, also:

Definition 4. Rand- und Anfangswertaufgabe der Wärmeleitung

Sei $\Omega \subset \mathbb{R}^3$ ein Gebiet mit glattem Rand $\partial\Omega$, der Rand ist orientiert durch äußeres Normaleneinheitsvektorfeld n . So beschreibt das Modell

$$\begin{aligned} \partial_t u - \langle \nabla, a \nabla u \rangle &= f \\ u(x, 0) &= u_0(x), \quad x \in \Omega \\ u(x, t) &= g(x, t), \quad x \in \partial\Omega \end{aligned} \quad (2.10)$$

die Rand- und Anfangswertaufgabe der Wärmeleitung für alle $t \in T \subset \mathbb{R}$.

Das Ergebnis dieses Kapitels ist in der obigen Definition 4 angegeben und wird als Grundmodell in folgenden Kapiteln dieser Arbeit vorausgesetzt.

2.1.1. Analytische Untersuchung der Wärmeleitungsgleichung

Eventuell ...

2.2. Inverses Problem der Wärmeleitung

Inverses Problem aus der Praktischen Sicht, also $u(x, y, t)$ wird gemessen und $a(x, y)$ dann bestimmt ...

Betrachten wir das Problem aus der Definition 4, insbesondere die Wärmeleitfähigkeit a . In vielen praktischen Problemstellungen hängt die Wärmeleitfähigkeit direkt von der Temperatur $u(x, t)$ ab ([3, S. 506]), also gilt $a : \mathbb{R} \rightarrow \mathbb{R}$.

Die DGL aus der Definition 4 bleibt bei dieser Betrachtung dieselbe, jedoch wird a jetzt explizit als Funktion von u betrachtet, also:

$$\frac{\partial u}{\partial t} - \nabla \cdot (a(u) \nabla u) = f \quad (2.11)$$

Um $a(u)$ zu isolieren, können wir folgendermaßen vorgehen:

1. Schreibe die Gleichung um, sodass der diffusive Term isoliert wird:

$$-\nabla \cdot (a(u) \nabla u) = f - \frac{\partial u}{\partial t}.$$

2. Verwende die Produktregel für den diffusen Term:

$$-\nabla \cdot (a(u) \nabla u) = -a(u) \nabla^2 u - \nabla a(u) \cdot \nabla u = -a(u) \Delta u - \nabla a(u) \cdot \nabla u.$$

3. Wenn wir annehmen, dass $a(u)$ nur von u abhängt (keine explizite Abhängigkeit von x, y), vereinfacht sich der Gradienten-Term:

$$\nabla a(u) = \frac{da(u)}{du} \nabla u.$$

4. Setze dies in die Gleichung ein:

$$-a(u) \nabla^2 u - \frac{da(u)}{du} (\nabla u \cdot \nabla u) = f - \frac{\partial u}{\partial t}.$$

5. Gruppiere die Terme, um $a(u)$ zu isolieren:

$$a(u) = -\frac{1}{\nabla^2 u} \left(f - \frac{\partial u}{\partial t} + \frac{da(u)}{du} (\nabla u \cdot \nabla u) \right).$$

die Wärmeleitungsgleichung im Gebiet $\Omega \subset \mathbb{R}^2$ mit zeitlicher Dimension $t \in T \subset \mathbb{R}$:

$$\frac{\partial u}{\partial t} - \nabla \cdot (a \nabla u) = f,$$

ergänzt durch Anfangsbedingungen

$$u(x, 0) = u_0(x, y) \quad \text{für } x \in \Omega,$$

und Dirichlet-Randbedingungen

$$u(x, t) = g(x, t) \quad \text{für } x \in \partial\Omega, t \in (0, T].$$

Das inverse Problem besteht darin, $a(u)$ zu finden, sodass die Differentialgleichung und die Randbedingungen erfüllt werden, wenn $u(x, t)$ und $f(x, t)$ bekannt sind.

2. Existenz und Eindeutigkeit

Satz 1: Existenz der Lösung

Sei $u(x, y, t)$ eine hinreichend glatte Lösung der Wärmeleitungsgleichung und $f(x, y, t)$ sowie $u_0(x, y)$ bekannt. Dann existiert unter geeigneten Regularitätsvoraussetzungen an die Daten eine Funktion $a(x, y)$, die das inverse Problem löst.

Beweisidee:

1. Schreibe die Wärmeleitungsgleichung in schwacher Form:

$$\int_{\Omega} \left(\frac{\partial u}{\partial t} v + a(x, y) \nabla u \cdot \nabla v \right) d\Omega = \int_{\Omega} f(x, y, t) v d\Omega,$$

für alle Testfunktionen $v \in H^1(\Omega)$.

2. Die Lösung $a(x, y)$ ergibt sich aus der schwachen Form, indem u und ∇u durch die Messdaten approximiert werden. Standardresultate aus der Variationsrechnung garantieren die Existenz von $a(x, y)$.

Satz 2: Eindeutigkeit der Lösung

Seien $u(x, y, t)$ und $f(x, y, t)$ eindeutig gegeben, und sei u hinreichend differenzierbar. Dann ist $a(x, y)$ eindeutig bestimmt, wenn $\nabla u \neq 0$ in Ω .

Beweis: Nehmen wir an, es existieren zwei Funktionen $a_1(x, y)$ und $a_2(x, y)$, die das inverse Problem lösen. Dann gilt:

$$\frac{\partial u}{\partial t} - \nabla \cdot (a_1 \nabla u) = f = \frac{\partial u}{\partial t} - \nabla \cdot (a_2 \nabla u).$$

Durch Subtraktion erhalten wir:

$$\nabla \cdot ((a_1 - a_2) \nabla u) = 0.$$

Da $\nabla u \neq 0$, folgt aus der Divergenzfreiheit, dass $a_1 = a_2$ fast überall in Ω . Dies zeigt die Eindeutigkeit.

3. Stabilität

Satz 3: Stabilität der Lösung

Sei $u_{\text{obs}}(x, y, t)$ eine Rauschbehaftung der exakten Lösung $u(x, y, t)$ mit einem maximalen Fehler $\epsilon > 0$, d. h.

$$\|u_{\text{obs}} - u\|_{L^2} \leq \epsilon.$$

Dann existiert eine Konstante $C > 0$, sodass der Fehler in $a(x, y)$ beschränkt ist durch

$$\|a_{\text{obs}} - a\|_{L^2} \leq C\epsilon.$$

Beweisidee:

1. Schreibe die Wärmeleitungsgleichung als Operatorgleichung:

$$\mathcal{A}(a) = f,$$

wobei $\mathcal{A}(a) = \frac{\partial u}{\partial t} - \nabla \cdot (a \nabla u)$.

2. Zeige, dass der Operator \mathcal{A} stetig und Lipschitz-stetig bezüglich a ist.
3. Wende den Satz von Lax-Milgram oder perturbative Stabilitätsresultate an, um die beschränkte Fehlerfortpflanzung von u auf a zu zeigen.

4. Regularisierung des inversen Problems

Da inverse Probleme oft schlecht gestellt sind, d. h. kleine Fehler in den Daten können zu großen Abweichungen in der Lösung führen, wird eine Regularisierung benötigt.

Definition (Regularisierte Zielfunktion)

Das inverse Problem wird als Minimierung einer modifizierten Zielfunktion formuliert:

$$J(a) = \int_{\Omega} \int_0^T \left| \frac{\partial u}{\partial t} - \nabla \cdot (a \nabla u) - f \right|^2 dt d\Omega + \lambda R(a),$$

wobei $R(a)$ eine Regularisierungsfunktion ist, typischerweise:

- $R(a) = \|a\|_{H^1(\Omega)}^2$ (Tikhonov-Regularisierung) oder
- $R(a) = \int_{\Omega} |\nabla a|^2 d\Omega$ (Glattheitsregularisierung).

Satz 4: Existenz der regularisierten Lösung

Sei $J(a)$ schwach unterer-halbstetig und sei $R(a)$ strikt konvex. Dann existiert mindestens eine minimierende Funktion $a \in H^1(\Omega)$, die das reguläre inverse Problem löst.

Beweis:

1. Schwache Halbstetigkeit und kompakte Einbettungen von Sobolev-Räumen garantieren die Existenz einer minimierenden Funktion.
2. Die Konvexität von $R(a)$ sorgt dafür, dass die Lösung eindeutig ist.

2.3. Numerische Lösungsverfahren der Wärmeleitungsgleichung

2.3.1. Herleitung der Variationsformulierung der Wärmeleitungsgleichung

Die Wärmeleitungsgleichung lautet:

$$\frac{\partial u(x, y, t)}{\partial t} - \nabla \cdot (a(u(x, y, t)) \nabla u(x, y, t)) = f(x, y, t), \quad (2.12)$$

und soll in die Variationsformulierung in schwacher Form umgeformt werden:

$$\int_{\Omega} \frac{\partial u}{\partial t} v dx + \int_{\Omega} a(u) \nabla u \cdot \nabla v dx = \int_{\Omega} f v dx. \quad (2.13)$$

Im Folgenden werden die Schritte zur Herleitung detailliert erläutert.

2.4. Geometrie und Gitter definieren

Die Geometrie und das Gitter definieren den Bereich Ω , in dem die Gleichung gelöst wird. Das Gitter teilt diesen Bereich in kleine Unterbereiche (Finite Elemente) auf.

2.4.1. Warum Geometrie und Gitter definieren?

- Die Methode der Finiten Elemente (FEM) basiert auf der Approximation der Lösung $u(x, y, t)$ auf einer diskreten Darstellung des Gebiets.
- Ein Gitter zerlegt Ω in einfachere Formen (z. B. Dreiecke oder Vierecke). Innerhalb jedes Elements wird die Lösung lokal approximiert.

2.4.2. Beispiel für ein rechteckiges Gebiet

$$\Omega = \{(x, y) \mid 0 \leq x \leq L_x, 0 \leq y \leq L_y\}. \quad (2.14)$$

Dieses Gebiet könnte mit Dreiecken oder Rechtecken diskretisiert werden.

2.5. Funktionsräume definieren

Ein Funktionsraum beschreibt, wie die Lösung u und die Testfunktionen v innerhalb der Finiten Elemente dargestellt werden.

2.5.1. Was sind P_1 - und P_2 -Elemente?

- P_1 : Lineare Basisfunktionen – u wird innerhalb eines Elements als lineares Polynom approximiert.
- P_2 : Quadratische Basisfunktionen – u wird innerhalb eines Elements als quadratisches Polynom approximiert.

Diese Basisfunktionen gewährleisten, dass u über das gesamte Gebiet hinweg stetig ist.

2.5.2. Warum Funktionsräume definieren?

Die Funktionsräume sorgen dafür, dass die approximierte Lösung u und die Testfunktionen v die notwendigen Kontinuitäts- und Randbedingungen der PDE erfüllen.

2.6. Variationsformulierung herleiten

Die Variationsformulierung (schwache Form) wird durch Multiplikation der PDE mit einer Testfunktion v und Integration über das Gebiet Ω hergeleitet. Dieser Schritt erlaubt es, auch Lösungen u zu berücksichtigen, die möglicherweise nicht überall differenzierbar sind.

2.6.1. Herleitung der schwachen Form

1. Starte mit der PDE:

$$\frac{\partial u}{\partial t} - \nabla \cdot (a(u) \nabla u) = f. \quad (2.15)$$

2. **Multipliziere mit einer Testfunktion v :** Multipliziere beide Seiten mit einer glatten Testfunktion $v \in V$, wobei V der Raum der zulässigen Testfunktionen ist:

$$v \left(\frac{\partial u}{\partial t} - \nabla \cdot (a(u) \nabla u) \right) = v f. \quad (2.16)$$

3. **Integriere über Ω :**

$$\int_{\Omega} v \frac{\partial u}{\partial t} dx - \int_{\Omega} v \nabla \cdot (a(u) \nabla u) dx = \int_{\Omega} v f dx. \quad (2.17)$$

4. **Wende den Divergenzsatz an:** Für den zweiten Term wende den Divergenzsatz an, um das Volumenintegral in ein Oberflächenintegral umzuwandeln:

$$\int_{\Omega} v \nabla \cdot (a(u) \nabla u) dx = \int_{\partial\Omega} v (a(u) \nabla u \cdot \mathbf{n}) dS - \int_{\Omega} \nabla v \cdot (a(u) \nabla u) dx. \quad (2.18)$$

Hier ist $\partial\Omega$ der Rand des Gebiets und \mathbf{n} der nach außen gerichtete Normalenvektor.

5. **Vereinfachung bei homogenen Randbedingungen:** Unter der Annahme homogener Dirichlet-Randbedingungen ($u = 0$ auf dem Rand) verschwindet der Oberflächenterm:

$$\int_{\Omega} v \nabla \cdot (a(u) \nabla u) dx = - \int_{\Omega} \nabla v \cdot (a(u) \nabla u) dx. \quad (2.19)$$

6. **Kombiniere die Terme:** Substituiere zurück in die Gleichung. Die schwache Form lautet:

$$\int_{\Omega} v \frac{\partial u}{\partial t} dx + \int_{\Omega} a(u) \nabla u \cdot \nabla v dx = \int_{\Omega} v f dx. \quad (2.20)$$

2.7. Zeitdiskretisierung

Die Zeitdiskretisierung ist notwendig, um die Lösung schrittweise für diskrete Zeitpunkte zu berechnen.

2.7.1. Warum Zeitdiskretisierung?

- Der Term $\frac{\partial u}{\partial t}$ ist eine partielle Ableitung nach der Zeit. FEM-Frameworks lösen jedoch räumliche Ableitungen.
- Die Zeitdiskretisierung ersetzt $\frac{\partial u}{\partial t}$ durch eine algebraische Differenz.

2.7.2. Backward-Euler-Verfahren

Das Backward-Euler-Verfahren approximiert $\frac{\partial u}{\partial t}$ zum Zeitpunkt t_{n+1} :

$$\frac{\partial u}{\partial t} \approx \frac{u^{n+1} - u^n}{\Delta t}. \quad (2.21)$$

Setze dies in die schwache Form ein:

$$\int_{\Omega} \frac{u^{n+1} - u^n}{\Delta t} v dx + \int_{\Omega} a(u^{n+1}) \nabla u^{n+1} \cdot \nabla v dx = \int_{\Omega} v f^{n+1} dx. \quad (2.22)$$

Dies führt zu einem System von Gleichungen, das für jeden Zeitschritt numerisch gelöst werden kann.

2.8. Schlüsselkonzepte

- **Testfunktion v :** Eine Testfunktion v gehört demselben Funktionsraum wie u und stellt sicher, dass die schwache Form für alle zulässigen Lösungen u gilt.
- **Warum schwache Form?**
 - Die schwache Form ermöglicht es, die PDE für Funktionen u zu lösen, die möglicherweise nicht überall differenzierbar sind (z. B. wenn ∇u nicht existiert).
 - Die Integration durch Teile verschiebt die Ableitung von u auf die Testfunktion v , die als glatt angenommen wird.

3. Deep Learning Methode

Kapitel X: Grundlagen des statistischen Lernens und Einführung in Physics-Informed Neural Networks (PINNs)

Traditionelle numerische Methoden wie Finite-Differenzen oder Finite-Elemente sind häufig mit hohen Rechenkosten verbunden. Eine vielversprechende Methode zur Lösung dieser Probleme stellen Physics-Informed Neural Networks (PINNs) dar, die maschinelles Lernen mit physikalischen Modellen kombinieren.

Dieses Kapitel beginnt mit einer detaillierten Einführung in die Grundlagen des statistischen Lernens, das eine mathematische Grundlage für maschinelles Lernen bildet. Anschließend wird die Theorie von PINNs eingeführt und detailliert ausgebaut.

3.1. Statistische Lernverfahren: Grundlagen

3.1.1. Definition des statistischen Lernens

Das Ziel des statistischen Lernens ist es, eine Abbildung $f : X \rightarrow Y$ zu finden, die die zugrunde liegende Verteilung der Daten möglichst gut approximiert. Hierbei sind X der Raum der Eingabedaten und Y der Raum der Zielgrößen (z. B. Labels). Formal lässt sich das Problem des Lernens als Minimierung eines Risikos beschreiben.

Sei (x_i, y_i) eine i.i.d. Stichprobe aus einer unbekannten gemeinsamen Verteilung $P(x, y)$. Ziel ist es, ein Modell f aus einem Hypothesenraum \mathcal{F} zu finden, das die zugehörige Verlustfunktion minimiert. Das Risiko-Funktional ist gegeben durch:

$$R(f) = \mathbb{E}_{(x,y) \sim P}[L(y, f(x))],$$

wobei $L : Y \times Y \rightarrow \mathbb{R}^+$ eine Verlustfunktion ist. Die Erwartung wird hierbei über die wahre Datenverteilung $P(x, y)$ genommen.

Da jedoch die wahre Verteilung $P(x, y)$ nicht bekannt ist, wird das Risiko durch den empirischen Risikofunktional approximiert:

$$R_n(f) = \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i)),$$

wobei (x_i, y_i) die beobachteten Trainingsdaten sind.

Die Zielsetzung des Lernens besteht also darin, den Schätzer f zu finden, der das Risiko minimiert:

$$\hat{f} = \arg \min_{f \in \mathcal{F}} R_n(f).$$

Eine zentrale Annahme des statistischen Lernens ist, dass das Modell f generalisierbar ist, d. h., es kann auf neue, unbekannte Daten angewendet werden.

3.1.2. Lineare Modelle und Regression

Eine der einfachsten und am häufigsten verwendeten Klassen von Modellen im statistischen Lernen sind lineare Modelle. Das grundlegende lineare Modell wird durch

$$f(x) = w^\top x + b$$

gegeben, wobei $w \in \mathbb{R}^d$ der Gewichtungsvektor und $b \in \mathbb{R}$ der Bias-Term sind. In der linearen Regression wird $f(x)$ so gewählt, dass der Fehler zwischen den Vorhersagen $f(x_i)$ und den tatsächlichen Werten y_i minimiert wird.

Die Verlustfunktion ist hierbei oft die quadratische Verlustfunktion:

$$L(y, f(x)) = \|y - f(x)\|^2.$$

Um die Parameter w und b zu optimieren, minimiert man das empirische Risiko:

$$R_n(w, b) = \frac{1}{n} \sum_{i=1}^n \|y_i - (w^\top x_i + b)\|^2.$$

Durch Anwendung der Methode der kleinsten Quadrate ergibt sich eine Lösung für w und b , die durch die Normalengleichung beschrieben wird:

$$\hat{w}, \hat{b} = \arg \min_{w, b} R_n(w, b).$$

Die Lösung dieser Gleichung lässt sich durch:

$$\hat{w} = (X^\top X)^{-1} X^\top y$$

für w (wobei X die Designmatrix ist) und \hat{b} durch den Mittelwert der Zielwerte berechnen.

3.1.3. Nichtlineare Modelle und neuronale Netze

Da lineare Modelle für viele Probleme nicht ausreichend sind, hat sich die Nutzung nichtlinearer Modelle als Standard etabliert. Ein prominentes nichtlineares Modell ist das neuronale Netz. Ein neuronales Netz ist eine Verknüpfung von linearen Transformationen und nichtlinearen Aktivierungsfunktionen. Ein einfaches Feedforward-Netzwerk besteht aus mehreren Schichten, wobei die Funktion eines neuronalen Netzes durch

$$f(x) = \sigma(W^{(L)} \sigma(W^{(L-1)} \dots \sigma(W^{(1)} x + b^{(1)}) \dots + b^{(L-1)}) + b^{(L)})$$

gegeben ist. Hierbei sind $W^{(l)}$ die Gewichtsmatrizen und $b^{(l)}$ die Bias-Vektoren der l -ten Schicht, und σ ist eine nichtlineare Aktivierungsfunktion wie ReLU oder Sigmoid.

Neuronale Netze werden häufig mit Gradientenabstieg optimiert, indem die Kreuzentropie-Verlustfunktion oder die Mean Squared Error (MSE)-Verlustfunktion minimiert wird:

$$L(y, f(x)) = \|y - f(x)\|^2.$$

Die Optimierung erfolgt durch Rückpropagation und den Gradientenabstiegsalgorithmus.

3.2. Einführung in Physics-Informed Neural Networks

3.2.1. Motivation für PINNs

In vielen physikalischen Problemen, wie der Wärmeleitungsgleichung, sind die zugrunde liegenden physikalischen Gesetze bekannt und können als Randbedingungen und PDEs formuliert werden. In vielen praktischen Anwendungen ist jedoch das inverse Problem von Interesse: Man möchte aus Messdaten, die an den Randbedingungen oder an einem Teil des Systems erfasst werden, die vollständige Lösung der PDE rekonstruieren.

3.2.2. Definition von PINNs

Ein Physics-Informed Neural Network (PINN) ist ein neuronales Netz, das nicht nur aus Daten lernt, sondern auch durch die Einbeziehung von physikalischen Gesetzen (wie PDEs) in den Optimierungsprozess trainiert wird.

Sei $u(x, t; \theta)$ die Funktion, die die Lösung des Problems approximiert, wobei θ die Parameter des neuronalen Netzes sind. Die Verlustfunktion eines PINNs ist eine Kombination aus zwei Termen:

- **Datenverlust** L_{Daten} : Dieser Term misst den Unterschied zwischen den durch das Netz vorhergesagten Werten und den gemessenen Werten.
- **Physikverlust** L_{Physik} : Dieser Term stellt sicher, dass die Lösung die physikalischen Gesetze (z. B. PDEs) erfüllt.

Die Gesamtkostenfunktion wird als Summe dieser beiden Komponenten formuliert:

$$L(\theta) = L_{\text{Daten}} + L_{\text{Physik}}.$$

3.2.3. Beispiel: Wärmeleitungsgleichung

Ein klassisches Beispiel für ein physikalisches Modell ist die Wärmeleitungsgleichung:

$$\frac{\partial u}{\partial t} - \alpha \Delta u = 0,$$

die die Temperaturverteilung $u(x, t)$ in einem Medium beschreibt, wobei α die Wärmeleitfähigkeit ist.

In einem PINN-Modell formuliert man den Physikverlust als:

$$L_{\text{Physik}} = \frac{1}{m} \sum_{j=1}^m \left\| \frac{\partial u}{\partial t}(x_j, t_j; \theta) - \alpha \Delta u(x_j, t_j; \theta) \right\|^2.$$

Das Ziel der vorliegenden Arbeit wird folgendermaßen in einzelne Punkte gegliedert

1. **Parameterschätzung** Ziel ist die Bestimmung der unbekannten Parameter $a(x, y)$ (Wärmeleitfähigkeit).
 - a) Aufbau eines PINN-Modells basierend auf der Wärmeleitungsgleichung.
 - b) Verwendung von synthetischen oder experimentellen Temperaturdaten.

- c) Optimierung des Modells zur Schätzung von $a(x, y)$.
- 2. **Identifikation von Wärmequellen** Ziel ist die Lokalisierung und Quantifizierung der unbekannten Wärmequellen $f(x, y, t)$.
 - a) Erweiterung des PINN-Modells um $f(x, y, t)$ als zusätzliche Unbekannte.
 - b) Kombination mit den geschätzten Parametern $a(x, y)$.
 - c) Validierung durch Vergleich von simulierten und beobachteten Temperaturprofilen.
- 3. **Optimale Steuerung** Ziel ist die Entwicklung einer optimalen Steuerungsstrategie für Wärmequellen oder Randbedingungen.
 - a) Formulierung eines Optimierungsproblems, z. B. Minimierung des Energieverbrauchs.
 - b) Verwendung des PINN-Modells zur Simulation und Bewertung von Steuerungsansätzen.
- 4. **Nichtlineare Wärmeleitung** Ziel ist die Erweiterung des Modells um temperaturabhängige Wärmeleitfähigkeit $a(T)$.
 - a) Einführung von nichtlinearen Termen in die Wärmeleitungsgleichung.
 - b) Anpassung des PINN-Modells für nichtlineare Effekte.
 - c) Validierung mit Szenarien extremer Temperaturen.
- 5. **Gekoppelte Multi-Physik-Probleme** Ziel ist die Kombination der Wärmeleitung mit anderen physikalischen Prozessen.
 - a) Erweiterung des Modells um zusätzliche Gleichungen, z. B. Navier-Stokes.
 - b) Verwendung von PINNs zur simultanen Lösung der gekoppelten Prozesse.
 - c) Untersuchung der Wechselwirkungen zwischen Wärmeleitung und anderen Prozessen.
- 6. **Zeitabhängige Randbedingungen** Ziel ist die Simulation dynamischer Randbedingungen und deren Einfluss auf das Temperaturfeld.
 - a) Integration zeitabhängiger Randbedingungen in das PINN-Modell.
 - b) Kombination mit den Ergebnissen aus den vorherigen Themen.
 - c) Validierung durch Szenarien mit realistischen, dynamischen Randbedingungen.
- 7. **Datengetriebenes Surrogat-Modell** Ziel ist die Erstellung eines datengetriebenen Modells für Echtzeit-Anwendungen.
 - a) Training eines schnellen Surrogat-Modells auf Basis der PINN-Ergebnisse.
 - b) Vergleich der Geschwindigkeit und Genauigkeit mit dem PINN-Modell.
 - c) Anwendung in Optimierungs- und Steuerungsaufgaben.

Erweiterung: Mathematische Theorie für das inverse Problem mit temperaturabhängiger Wärmeleitfähigkeit

1. Das zugrunde liegende Problem

Die modifizierte Wärmeleitungsgleichung lautet:

$$\frac{\partial u}{\partial t} - \nabla \cdot (a(u) \nabla u) = f(x, y, t),$$

wobei:

- $u(x, y, t)$ die Temperaturverteilung ist,
- $a(u)$ eine unbekannte temperaturabhängige Wärmeleitfähigkeit darstellt,
- $f(x, y, t)$ eine unbekannte Wärmequelle ist.

Das inverse Problem besteht darin, die Funktionen $a(u)$ und $f(x, y, t)$ basierend auf Messungen der Temperatur $u_{\text{obs}}(x, y, t)$ zu rekonstruieren.

2. Formulierung als Variationsproblem

Für das inverse Problem wird das Modell in ein Optimierungsproblem überführt. Die Zielfunktion besteht aus einem Datenanpassungsterm und physikalischen Residuen:

$$\mathcal{J}(u, a, f) = \mathcal{J}_{\text{Data}}(u) + \mathcal{J}_{\text{PDE}}(u, a, f),$$

wobei:

- **Datenanpassung:**

$$\mathcal{J}_{\text{Data}}(u) = \frac{1}{2} \int_0^T \int_{\Omega} (u(x, y, t) - u_{\text{obs}}(x, y, t))^2 \, d\Omega \, dt,$$

- **PDE-Residuum:**

$$\mathcal{J}_{\text{PDE}}(u, a, f) = \frac{1}{2} \int_0^T \int_{\Omega} \left(\frac{\partial u}{\partial t} - \nabla \cdot (a(u) \nabla u) - f(x, y, t) \right)^2 \, d\Omega \, dt.$$

Zur Stabilisierung wird eine Regularisierung hinzugefügt:

$$\mathcal{J}_{\text{Reg}}(a, f) = \alpha \|a(u)\|_{H^1}^2 + \beta \|f(x, y, t)\|_{L^2}^2,$$

wobei $\alpha, \beta > 0$ Regularisierungsparameter sind.

Die kombinierte Zielfunktion lautet:

$$\mathcal{J}(u, a, f) = \mathcal{J}_{\text{Data}}(u) + \mathcal{J}_{\text{PDE}}(u, a, f) + \mathcal{J}_{\text{Reg}}(a, f).$$

3. Existenz und Eindeutigkeit

Satz 1 (Existenz): Unter der Annahme, dass die Beobachtungsdaten u_{obs} ausreichend regelmäßig sind ($u_{\text{obs}} \in H^1(\Omega \times (0, T))$), existiert eine Lösung (u, a, f) des Optimierungsproblems.

Beweis (Skizze):

1. Zeige, dass die Zielfunktion \mathcal{J} nach unten beschränkt ist.
2. Zeige, dass die Menge der zulässigen Parameter (u, a, f) schwach kompakt ist.
3. Wende den Satz von Weierstraß auf das minimierende Funktional an.

Bemerkung: Die Eindeutigkeit ist im Allgemeinen nicht garantiert, da das Problem schlecht gestellt ist (mehrere Kombinationen von $a(u)$ und $f(x, y, t)$ können dieselben Messdaten erklären). Regularisierungen wie H^1 - und L^2 -Normen tragen dazu bei, die Lösung zu stabilisieren.

4. PINNs für das inverse Problem

Ein Physics-Informed Neural Network (PINN) wird verwendet, um sowohl die Temperatur $u(x, y, t)$ als auch die unbekannten Funktionen $a(u)$ und $f(x, y, t)$ zu approximieren. Die neuronalen Netze sind wie folgt definiert:

- $u(x, y, t; \theta_u)$: Neuronales Netz zur Approximation der Temperatur,
- $a(u; \theta_a)$: Neuronales Netz für die Wärmeleitfähigkeit,
- $f(x, y, t; \theta_f)$: Neuronales Netz für die Wärmequelle.

Die Verlustfunktion des PINN-Modells kombiniert mehrere Terme:

$$\mathcal{L} = \mathcal{L}_{\text{Data}} + \mathcal{L}_{\text{PDE}} + \mathcal{L}_{\text{Reg}}.$$

- **Datenanpassung:**

$$\mathcal{L}_{\text{Data}} = \frac{1}{N_{\text{obs}}} \sum_{i=1}^{N_{\text{obs}}} (u(x_i, y_i, t_i; \theta_u) - u_{\text{obs}}(x_i, y_i, t_i))^2,$$

- **PDE-Residuum:**

$$\mathcal{L}_{\text{PDE}} = \frac{1}{N_{\text{PDE}}} \sum_{i=1}^{N_{\text{PDE}}} \left(\frac{\partial u}{\partial t} - \nabla \cdot (a(u; \theta_a) \nabla u) - f(x, y, t; \theta_f) \right)^2,$$

- **Regularisierung:**

$$\mathcal{L}_{\text{Reg}} = \alpha \|\nabla_{\theta} a(u; \theta_a)\|^2 + \beta \|\nabla_{\theta} f(x, y, t; \theta_f)\|^2.$$

Das PINN wird durch Minimierung der Verlustfunktion trainiert:

$$\theta_u^*, \theta_a^*, \theta_f^* = \arg \min_{\theta_u, \theta_a, \theta_f} \mathcal{L}.$$

5. Gradientenbasierte Optimierung und automatische Differentiation

Die Ableitungen $\frac{\partial u}{\partial t}$ und ∇u werden direkt durch automatische Differentiation (AD) berechnet, was PINNs von traditionellen numerischen Verfahren unterscheidet. Dies ermöglicht eine nahtlose Integration der Differentialgleichung in den Trainingsprozess.

6. Eigenschaften und Herausforderungen

- **Kopplung von $a(u)$ und $f(x, y, t)$:** Die Rekonstruktion von $a(u)$ ist nichtlinear gekoppelt mit der Lösung $u(x, y, t)$, da a explizit von u abhängt. Diese Nichtlinearität führt zu Herausforderungen in der Konvergenz der Optimierung.
- **Ill-posedness:** Das inverse Problem bleibt empfindlich gegenüber Rauschen in den Beobachtungsdaten. Regularisierungen und sorgfältig gewählte Netzarchitekturen sind entscheidend.
- **Einschränkungen:** Die Lösungsgüte hängt stark von der Verfügbarkeit hochauflösender Messdaten u_{obs} ab, insbesondere für stark heterogene Materialien oder schnell variierende Wärmequellen.

7. Beispiel für das PINN-Inverse-Problem

Ein mögliches Szenario: Gegeben sei eine Messung der Temperatur in einem rechteckigen Gebiet, in dem $a(u)$ quadratisch von u abhängt ($a(u) = a_0 + a_1 u + a_2 u^2$), und $f(x, y, t)$ unbekannt ist. Das PINN-Training führt zur simultanen Rekonstruktion von:

- $a(u)$ durch das Netzwerk θ_a ,
- $f(x, y, t)$ durch θ_f ,
- $u(x, y, t)$ durch θ_u .

Die Anwendung solcher PINNs kann sowohl für direkte Echtzeitsimulationen als auch für die Materialdiagnostik genutzt werden.

PINNs als methodisches Fundament

Physics-Informed Neural Networks (PINNs) sind eine neuartige Klasse von maschinellen Lernmodellen, die physikalische Gesetzmäße direkt in den Trainingsprozess einbinden. Die Verlustfunktion eines PINN-Modells kombiniert die Residuen der Differentialgleichung mit Rand- und Anfangsbedingungen:

Hierbei wird ∇u und $\frac{\partial u}{\partial t}$ direkt durch automatische Differentiation berechnet. PINNs sind besonders geeignet für inverse Probleme, da sie flexibel auf Messdaten trainiert werden können.

Parameteridentifikation und Quellenerkennung

Eine zentrale Herausforderung bei der Modellierung der Wärmeleitung ist die Identifikation der unbekannten Parameter $a(x, y)$ und $f(x, y, t)$. Hierfür werden:

1. Beobachtungsdaten $u_{\text{obs}}(x, y, t)$ verwendet,
2. ein PINN-Modell so trainiert, dass die physikalischen Gesetze erfüllt sind und die Diskrepanz zwischen Simulation und Beobachtungen minimiert wird.

Durch die Einbindung von Messdaten können sowohl Materialeigenschaften wie Wärmeleitfähigkeit als auch externe Wärmequellen hochgenau identifiziert werden.

Dieses Vorgehen ist insbesondere für nicht-homogene oder zeitlich variierende Materialien relevant.

Optimierung und Steuerung

Die Optimierung von Steuerungsstrategien spielt eine wichtige Rolle, um gewünschte Temperaturprofile zu erreichen oder Energieverluste zu minimieren. Ein datengetriebenes Modell erlaubt es, die Steuerung in Echtzeit anzupassen. Beispiele umfassen:

- Optimierung der Heizungs- oder Kühlmuster,
- dynamische Anpassung von Randbedingungen basierend auf Sensorwerten.

PINNs können durch ihre schnelle Auswertung und Integration mit Steuerungsalgorithmen wie modellprädiktiver Regelung (MPC) besonders effizient eingesetzt werden.

3.3. Nichtlineare und gekoppelte Probleme

Die Erweiterung auf nichtlineare und gekoppelte Probleme ist ein natürlicher nächster Schritt. Beispiele umfassen:

- Temperaturabhängige Wärmeleitfähigkeit $a(u)$,
- Kopplung mit anderen physikalischen Phänomenen wie Feuchtigkeitsdiffusion oder chemischen Reaktionen.

Die PINN-Architektur erlaubt eine flexible Anpassung an diese komplexeren Szenarien, indem weitere physikalische Gesetzmäße in die Verlustfunktion eingebunden werden.

Erstellung eines datengetriebenen Surrogat-Modells

Ein Surrogat-Modell dient als schnelle Approximation der PINN-Ergebnisse. Es wird auf Basis von PINN-Simulationen trainiert und kann für Echtzeitanwendungen eingesetzt werden. Ziel ist es, die Genauigkeit beizubehalten, während die Rechenzeit erheblich reduziert wird. Typische Methoden umfassen:

- Reduktion der Eingabedimensionen durch Hauptkomponentenanalyse (PCA),
- Einsatz von einfachen neuronalen Netzen oder Gaussian Processes als Approximation.

Schlussfolgerung und Ausblick

Die Entwicklung eines datengetriebenen Echtzeitmodells für die Wärmeleitung zeigt das Potenzial moderner Methoden des maschinellen Lernens auf. Insbesondere PINNs bieten eine elegante Möglichkeit, physikalisches Wissen und datengetriebene Ansätze zu kombinieren. In Zukunft könnten Anwendungen in der industriellen Prozesssteuerung, der Klimamodellierung oder der medizinischen Bildgebung weiter vorangetrieben werden.

3.4. Das des Computertomographen

Das Wort Tomographie setzt sich aus zwei dem Altgriechischen entstammten Wörtern zusammen. $\tau\omicron\mu\eta$ [*tome*], bedeutet Schnitt und $\gamma\rho\alpha\varphi\epsilon\iota\nu$ [*graphien*] bezeichnet Schreiben. Zusammengefasst zu einem Wort *Tomographie* kann es als Schnittbild eines Objekts verstanden werden.

Die CT ist ein bildgebendes Verfahren, das in der medizinischen Radiologie und in der Materialforschung eingesetzt wird. Die Technologie des Verfahrens erlaubt eine zerstörungsfreie Materialuntersuchung. Das ausschlaggebende Wort ist hier *zerstörungsfrei*, damit ist die Untersuchung des inneren Aufbaus eines Objekts, ohne es zu öffnen, gemeint.

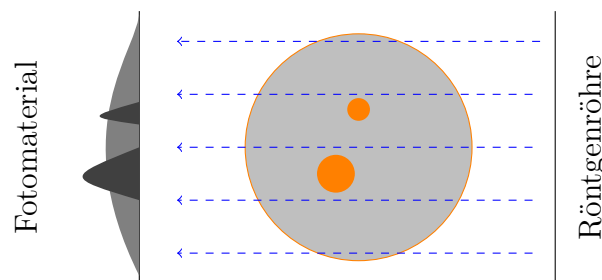


Abbildung 3.1.: Schematischer Aufbau einer Röntgenaufnahme im Draufsicht. Links ist das Schwächungsprofil nach Durchgang der Röntgenstrahlen durch ein Medium mit verschiedenen Dichten zu sehen (mit der Annahme, dass orangene Kreise höhere Dichte haben, als die graue Fläche).

Sicherlich wäre die Computertomographie ohne die Röntgenstrahlung nicht denkbar. Diese verdankt ihren Namen dem deutschen Physiker Wilhelm Conrad Röntgen (1845-1923). Im Jahre 1895 gelang es ihm zum ersten Mal eine elektromagnetische Strahlung zu erzeugen, deren Wellenlänge zwischen 10nm und 1pm lag. Solche kurze Wellenlängen haben nur energiereiche Strahlen. Das erlaubt ihnen die Durchdringung der Materie. Anzumerken ist, dass die Intensität hochenergetischer Strahlen beim Durchgehen der Materie mit der Eindringtiefe abfällt. Dies bedeutet, dass die Röntgenstrahlung beim Auftreffen auf ein Probestück eine höhere Intensität besitzt als beim Austreten. Bei den herkömmlichen Röntgenaufnahmen wurde diese Eigenschaft direkt ausgenutzt und man hat überlagerte Schatten von Objekten mit verschiedenen Dichten auf das Aufnahmematerial projiziert bekommen. Damit erschienen die dichteren Stellen des Objekts dunkler, weiche entsprechend heller. Für gewisse Situationen geben solche Projektionen genug Information her, jedoch nicht, wenn die räumliche Anordnung des Inneren eines Objekts diskutiert werden soll (Abb.??).

Um das Problem der räumlichen Anordnung der Objekte zu lösen, ist folgendes Vorgehen sehr nützlich. Würde man das Schattenbild aus der Abbildung ?? zurück in die Schnittebene des Objekts projizieren, so wie es in der Abbildung 3.2 (a) gezeigt ist, erhält man etwas Information über die gegebene Objekte. Projiziert man mehrere Schattenbilder, die aus verschiedenen Winkeln aufgenommen worden sind, so erhält man als Summe der Rückprojektionen, die sogenannte Rückprojektion (Abb. 3.2(b)).

Später werden wir sehen, welcher mathematischer Aufwand sich hinter dieser Idee verbirgt, bevor ein annehmbares Resultat zustande kommt.

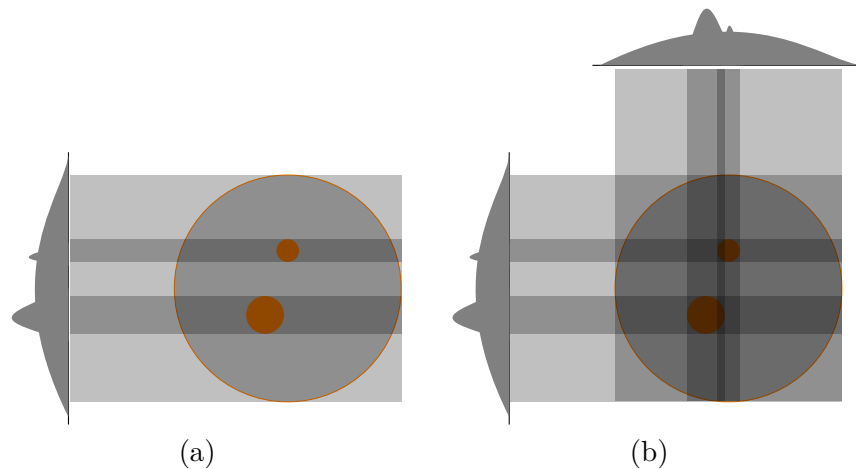


Abbildung 3.2.: Das Prinzip der Bildrekonstruktion aus CT-Daten mit einer Rückprojektion (a) und zwei Rückprojektionen (b).

Ein Pionier der Tomographie, A. M. Cormack³, veröffentlichte 1964 eine Arbeit über die mathematische Bestimmung der Dichten eines Objekts. Aus den Projektionsdaten konnte er die innere Dichteverteilung des Objekts bestimmen.

Der erste Prototyp eines CT-Scanners wurde vom G. N. Hounsfield⁴ im Jahr 1968 entwickelt. Seine Arbeit beruhte auf den Erkenntnissen von M. Cormack. Die Idee, der um das Objekt rotierenden Scanneinheit (Abb. 3.3) findet man hier wieder. Vorerst konnten nur anatomische Präparate vermessen werden. Aber schon im Jahre 1972 entstehen erste klinische Untersuchungen mit einem Schädelscanner durch Hounsfield und Ambrose. 1973 werden Berichte über die Weichteildiagnostik des Gehirns mithilfe eines CT-Scanners veröffentlicht. Hinsichtlich der Nützlichkeit des CT-Verfahrens in der Medizin, wurden A. M. Cormack und G. N. Hounsfield im Jahre 1979 mit dem Nobelpreis für Medizin ausgezeichnet.

In darauf folgenden Jahren entstand eine ganze Reihe von Generationen des CT-Scanners. Die Hauptunterschiede bestehen in der Komposition der Scanneinheit und dessen Technik, siehe auch [2, Kap. 3]. Heutzutage hat sich die dritte Generation am meisten durchgesetzt. Der prinzipielle Grundaufbau (Abb. 3.3) dieses Scanners besteht darin, dass die Röntgenröhre und Detektor fest auf einem Ringkörper, einander gegenüber installiert sind. In der Mitte des Rings ist ein fest installierter Probetisch, sodass die Scanneinheit um diesen rotieren kann. Mittels dieser Konstruktion ist es möglich *tomographische* Bilddaten aus allen Positionen um die Längsachse des Objekts aufzunehmen. Die oben beschriebene Rückprojektion aller tomographischen Daten (Projektionen) erzeugt ein Schnittbild. Erzeugt man mehrere Schnittbilder entlang der Längsachse des Objekts, so kann man diese der Reihe nach aufstellen und mithilfe programmierter technischer Werkzeuge ein dreidimensionales Bild dieses Objekts gewinnen (Abb. 3.4).

³ Allan M. Cormack (1924-1998) südafrikanischer Physiker.

⁴ Godfrey N. Hounsfield (1919 - 2004) englischer Ingenieur.

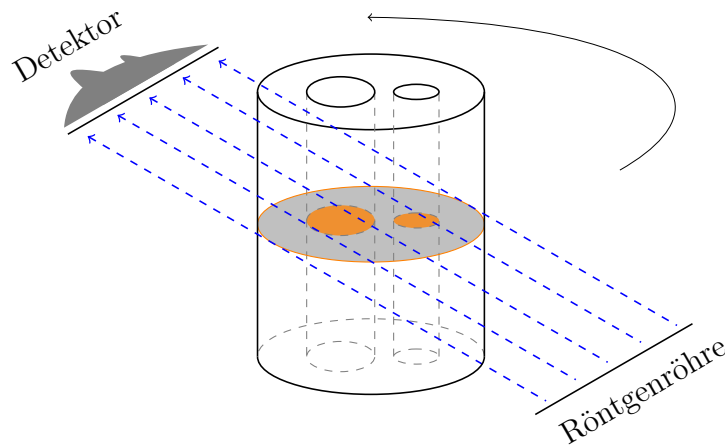


Abbildung 3.3.: Prinzipieller Aufbau eines CT-Scanners der dritten Generation.

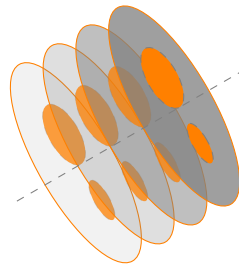


Abbildung 3.4.: Das Schichtmodell eines Objekts als Grundlage zur dreidimensionalen Rekonstruktion.

Im Folgenden gehen wir mehr auf die mathematisch-physikalischen Gegebenheiten der oben beschriebenen Situation ein.

3.5. Die Radon Transformation

Das Verständnis des mathematischen Problems der Bildrekonstruktion in der Computertomographie verlangt zunächst das Verständnis der Entstehung der tomographischen Bilddaten. Die Entstehung solcher Daten kann mithilfe eines mathematischen Modells nachvollzogen werden. Bekanntlich baut man das mathematische Modell auf physikalischen Annahmen auf. Die Abbildung 3.5 soll in diesem Fall als physikalisches Ausgangsmodell dienen. Wir nehmen an, dass der graue Kreis eine homogene Dichteverteilung besitzt. Unter weiteren Annahme sollen die Röntgenstrahlen sich geradlinig ohne Streuung in Richtung des Detektors ausbreiten.

Sei nun $f(x, y)$ die gesuchte Dichtefunktion, die auf einem Einheitskreis $\Omega = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$ definiert ist. Dass f auf einem Kreis liegen soll, folgern wir aus der physikalischen Situation. Das Objekt wird aus einem bestimmten Radius um den Mittelpunkt der Scanneinheit gescannt, daher möchten wir, dass ein Objekt in die CT-Scanneinheit passt, gegebenenfalls muss die Scanneinheit skaliert werden.

Die blau gestrichelte Linie in Abbildung 3.5 stelle einen Röntgenstrahl dar. Wir

Die rechte Seite in (3.5) liefert einen Summenwert von f entlang der Geraden L (Abb. 3.6), die wie folgt parametrisiert werden kann:

$$\begin{aligned} L(s, \theta, t) &= s \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} + t \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix}, \quad s, t \in \mathbb{R}; \quad \theta \in [0, \pi]. \\ &= s\omega(\theta) + t\omega^\perp(\theta) \end{aligned} \quad (3.6)$$

Betrachtet man alle Geraden, die den Einheitskreis schneiden, dann ergibt sich folgende Menge

$$G = \{L(s, \theta, t) \mid |s| \leq 1, \theta \in [0, \pi], t \in [-\gamma(s), \gamma(s)]\}. \quad (3.7)$$

Wobei $\gamma(s)$ den Parameter t für (3.6) festlegt und folgendermaßen definiert werden kann

$$\gamma(s) = \begin{cases} \sqrt{1-s^2} & : |s| \leq 1 \\ 0 & : \text{sonst} \end{cases}. \quad (3.8)$$

Die Gleichung (3.8) liefert nun die Integrationsgrenzen für (3.5). Da die Werte von $\pm\gamma(s)$ genau die Schnittpunkte von L mit Ω liefern und $\text{supp} f^5 \subseteq \Omega$, heißt es insgesamt, dass nur die Werte von $L \subset \text{supp} f$ bei der Integration von (3.5) berücksichtigt werden.

Kehren wir zur Gleichung (3.5) zurück und betrachten die Abbildung 3.6, werden wir feststellen, dass die Funktion f über den Parameter t integriert werden muss. Hierzu bedarf es einer Variablentransformation:

$$|(J(L))| = \left| \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \right| = 1$$

und

$$(3.9)$$

$$dL = \det(J(L))dt.$$

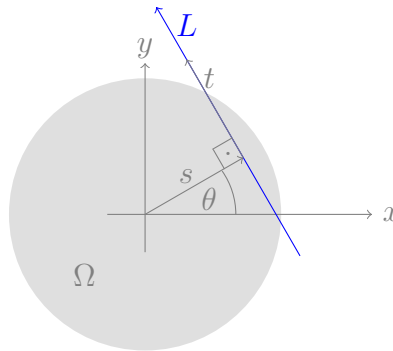


Abbildung 3.6.: Parametrisierung des Strahls L auf dem Gebiet Ω .

Mit den obigen Abmachungen haben wir nun alles Nötige in der Hand, um die Radon⁶ Transformation angeben zu können.

⁵ Mit $\text{supp} f$ (Support) ist der Träger einer Funktion f gemeint. Def.: $\text{supp} f := \overline{\{x \in \Omega \mid f(x) \neq 0\}}$

⁶ Johann Karl August Radon (1887-1956) österreichischer Mathematiker.

Definition 5. Sei nun $f : \Omega \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ eine integrierbare Funktion, dann ist ihre Radon Transformation durch

$$\mathcal{R}f(s, \theta) := \int_{-\gamma(s)}^{\gamma(s)} f(s\omega(\theta) + t\omega^\perp(\theta)) dt \quad (3.10)$$

definiert. Hier ist $\Omega = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$.

Bemerkung 1. \mathcal{R} wird als Integraloperator oder einfach als Operator bezeichnet.

Die Gleichung (3.10) trägt das Wesen der Computertomographie in sich und gehört in den Bereich der Integraltransformationen. In der Tat ist man aber an der Inversion von \mathcal{R} interessiert. Das wird klarer, wenn man (3.10) etwas anders auffasst:

$$p(s, \theta) = \mathcal{R}f(s, \theta). \quad (3.11)$$

Somit stellt $p(s, \theta)$ in (3.11) genau die Projektion einer unbekannten Dichteverteilung entlang eines Röntgenstrahls $L(s, \theta)$ dar. Da die Projektionen als Messwerte beim CT-Verfahren entstehen, möchte man f folgendermaßen wiedergewinnen

$$f(x) = \mathcal{R}^{-1}p(x). \quad (3.12)$$

Jedoch offenbaren sich einige mathematische Hürden bei der Inversion von (3.10), was die Rekonstruktion von f aus den Projektionsdaten schwierig macht. Um die dahinterstehende Problematik besser verstehen zu können, untersuchen wir zunächst die Eigenschaften von (3.10).

3.5.1. Kleiner Einschub über L^2 -Raum

Um den Operator \mathcal{R} untersuchen zu können, sollte man wissen auf welchen Räumen er operiert. Dieses Wissen erlaubt uns Aussagen über die Stetigkeit, Kompaktheit und auch über die Singulärwertzerlegung (SWZ) von \mathcal{R} zu treffen. Besonders wichtig ist die SWZ, denn anhand dieser ergibt sich eine Möglichkeit den Operator \mathcal{R} zu invertieren. Zudem wollen wir im nächsten Kapitel die Schlechtgestelltheit des computertomographischen Problems diskutieren, welche die Eigenschaften der SWZ benutzt.

In Bezug auf das hier vorgestellte Rekonstruktionsproblem der Computertomographie betrachten wir Funktionen auf dem Gebiet $\Omega \subset \mathbb{R}^2$. Wir setzen voraus, dass die Funktionen auf Ω integrierbar sind. Wie dieser Sachverhalt zu verstehen ist, wollen wir kurz und grob andeuten.

Zunächst klären wir den Begriff der Integrierbarkeit einer Funktion im Sinne von Lebesgue⁷. Das Lebesgue-Integral wird für Funktionen auf einem geeigneten Maßraum definiert. Ein Maßraum ist durch ein Tripel $(\Omega, \mathcal{A}, \lambda)$ gegeben.

Eine sogenannte σ -Algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$ auf Ω ist ein Mengensystem bestehend aus Teilmengen von Ω . \mathcal{A} besitzt folgende Eigenschaften: (i). Jede endliche oder abzählbar unendliche Vereinigung oder jeder Durchschnitt der Elemente aus \mathcal{A} liegen

⁷Henri Léon Lebesgue (1875-1941) französischer Mathematiker.

wieder in \mathcal{A} , sowie Ω und \emptyset . (ii). Die Komplemente der Elemente aus \mathcal{A} , also $\forall A \in \mathcal{A} : \Omega \setminus A = A^C \in \mathcal{A}$, liegen auch in \mathcal{A} .

Nun suchen wir eine Funktion, die allen Elementen A aus \mathcal{A} eine positive Zahl zuordnet, also $\lambda(A) \in [0, \infty]$. Somit beschreibt λ das Maß einer Menge in \mathcal{A} . Im Falle $\Omega \subset \mathbb{R}^2$ ordnet λ den Flächeninhalt einer Teilmenge von Ω zu. Des Weiteren hat λ folgende Eigenschaften: (i): $\lambda(\emptyset) = 0$ und (ii): die σ -Additivität, das heißt für paarweise disjunkte Mengen $(A_n)_{n \in \mathbb{N}}$ aus \mathcal{A} gilt

$$\lambda \left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \lambda(A_n).$$

Nun gilt per Definition *messbarer Abbildungen* zwischen zwei Maßräumen, dass die Urbilder messbarer Mengen messbar sind. Das heißt $f : (X_1, \mathcal{A}_1) \rightarrow (X_2, \mathcal{A}_2)$ ist messbar, wenn für alle $A_2 \in \mathcal{A}_2$ gilt $f^{-1}(A_2) \in \mathcal{A}_1$.

Um den Begriff der Integrierbarkeit im Sinne von Lebesgue einführen zu können, fehlt uns noch der Begriff der Borel⁸- σ -Algebra. Dazu benutzen wir für uns relevante Mengen \mathbb{R} und \mathbb{R}^2 . Das ist insofern klar, weil die gesuchten Dichtefunktionen in einem Punkt $x \in \Omega \subset \mathbb{R}^2$ einen skalaren Wert $f(x) \in \mathbb{R}$ liefern. Noch legen wir fest, dass die Werte der Funktionen auf Ω immer positiv sind, also $f(x) \in \mathbb{R}^+$, denn es gibt keine negativen Dichten. Betrachten wir den topologischen Raum $(\mathbb{R}, \mathcal{T})$, hier stellt \mathcal{T} ⁹ ein System von offenen Mengen von \mathbb{R} dar, dann heißt $\mathcal{B}(\mathbb{R})$ die von \mathcal{T} erzeugte σ -Algebra auf \mathbb{R} , die Borel- σ -Algebra. $\mathcal{B}(\mathbb{R})$ ist die kleinste Algebra, die \mathcal{T} enthält. Das bringt uns ein Vorteil bezüglich der Messbarkeit, denn somit erhalten wir die größtmögliche Menge messbarer Mengen von \mathbb{R} . Für \mathbb{R}^2 gilt das Obere auch.

Im letzten Schritt definieren wir die sogenannten positiven Treppenfunktionen $h^+ : (\Omega, \mathcal{B}(\Omega)) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, die endlich viele Werte $c_j \in \mathbb{R}_0^+$, $j \in [1, n]$, $n \in \mathbb{N}$ auf Ω annehmen. Auf diese Weise ist es möglich das Integral über h^+ folgendermaßen anzugeben:

$$\mathcal{I}(h^+) = \sum_{j=1}^n c_j \lambda(A_j), \quad \text{für } A_j \in \mathcal{B}(\Omega).$$

Jetzt können wir das Lebesgue-Integral einer messbaren Funktion $f : \Omega \rightarrow \mathbb{R}^+$ definieren, indem wir die Funktion f ‘von unten’ mit Treppenfunktionen approximieren:

Definition 6. Sei $f : \Omega \rightarrow \mathbb{R}^+$ messbar, dann ist sie integrierbar im Sinne von Lebesgue, wenn gilt

$$\int_{\Omega} f(x) = \sup \{ \mathcal{I}(h^+) \mid \forall x \in \Omega : h^+(x) \leq f(x) \} < \infty.$$

⁸Émile Borel (1871-1956) französischer Mathematiker.

⁹Sei X eine Menge, dann ist $\mathcal{T} \subset 2^X$ eine Topologie auf X , falls folgendes gilt:

- (i) $\emptyset, X \in \mathcal{T}$.
- (ii) $A, B \in \mathcal{T} \Rightarrow A \cap B \in \mathcal{T}$.
- (iii) $\{A_i\}_{i \in I} \in \mathcal{T} \Rightarrow \bigcup_{i \in I} A_i \in \mathcal{T}$

So nennt man das Paar (X, \mathcal{T}) ein topologischer Raum. Alle Elemente von \mathcal{T} sind offene Mengen.

Bemerkung 2. Im Falle einer Lebesgue-Integrierbarkeit spricht man meistens davon, dass eine Funktion integrierbar ist. Streng genommen müsste man das Lebesgue-Integral für beliebige Funktionen $f : \Omega \rightarrow \mathbb{R}$ mit einer geeigneten Fortsetzung einführen, um dem Begriff der Integrierbarkeit gerecht zu werden. Aber die Definition 6 reicht uns, um das Prinzip der Integration verstanden zu haben.

Nun sind wir im Stande den Raum der quadratintegrierbarer Funktionen folgendermaßen zu definieren:

$$\mathcal{L}^2(\Omega, \mathcal{A}, \lambda) = \{f : \Omega \rightarrow \mathbb{R} \mid f \text{ ist integrierbar und } \int_{\Omega} |f(x)|^2 dx \leq \infty\}. \quad (3.13)$$

Versehen wir den Raum $\mathcal{L}^2(\Omega, \mathcal{A}, \lambda)$ mit einer Norm

$$\|f\|_2 := \langle f, f \rangle^{\frac{1}{2}} = \left(\int |f(x)|^2 dx \right)^{\frac{1}{2}}, \quad x \in \Omega, \quad (3.14)$$

so erhalten wir den Raum $L^2(\Omega, \mathcal{A}, \lambda)$ (einfach $L^2(\Omega)$), also den *Hilbertraum der quadratintegrierbaren Funktionen*.

Im Allgemeinen ist $L^2(\Omega)$ ein unendlichdimensionaler Vektorraum über dem Körper der komplexen Zahlen, der bezüglich der durch das Skalarprodukt induzierten Norm vollständig ist.

Kehren wir zurück, zu den gesuchten Funktionen auf einer Kreisscheibe Ω , dann sei das Urbild des Operators \mathcal{R} genau die Menge $L^2(\Omega)$. Das Bild von \mathcal{R} liegt jedoch auf einem halben Zylinder $Z := [0, \pi] \times [-1, 1]$. Der Grund hierfür ist ein technischer, denn bei Projektionen über den ganzen Durchmesser einer Kreisfläche auf einem Intervall von 0 bis 2π , entstehen redundante Daten, also reichen die Projektionen von 0 bis π .

Des Weiteren wissen wir an dieser Stelle noch nicht wohin \mathcal{R} den Raum $L^2(\Omega)$ abbildet. Aber wie oben schon erwähnt wurde, wollen wir später die SWZ von \mathcal{R} angeben, dafür muss ein *adjungierter* Operator \mathcal{R}^* zu \mathcal{R} existieren¹⁰. Um das zu bewerkstelligen, muss man fordern, dass der Bildraum sowie der Kern von \mathcal{R} Orthogonalbasen besitzen sollen, was auch einen Skalarprodukt impliziert. Hieraus kann eine Vermutung aufgestellt werden, dass $\mathcal{R} : L^2(\Omega) \rightarrow L^2(Z)$ abbilden soll, weil eben L^2 Räume der Forderung gerecht sind. Somit erhalten wir insgesamt

$$\mathcal{R} : L^2(\Omega) \rightarrow L^2(Z). \quad (3.15)$$

Mit (3.15) kann man die Eigenschaften des Operators \mathcal{R} gut untersuchen.

3.5.2. Eigenschaften der Radon Transformation

In diesem Kapitel sollen einige Eigenschaften des Operators \mathcal{R} erarbeitet werden. Anhand der Eigenschaften von \mathcal{R} kann die Problematik der Gleichung (3.12) besser verstanden werden. Zunächst zeigen wir die Linearität von \mathcal{R} .

¹⁰ Der Grund weshalb \mathcal{R}^* existieren muss ist folgender: Es kann nicht vorausgesetzt werden, dass \mathcal{R} selbstadjungiert ist, wonach der *Spektralsatz für selbstadjungierte Operatoren* für \mathcal{R} nicht gelten würde. Aber für $\mathcal{R}\mathcal{R}^*$ funktioniert der Satz schon, da $\mathcal{R}\mathcal{R}^* = (\mathcal{R}\mathcal{R}^*)^* = \mathcal{R}^*\mathcal{R}$ gelte.

Linearität von \mathcal{R}

\mathcal{R} ist linear, denn für $f, g \in L^2(\Omega)$ beliebig, ergibt sich

$$\begin{aligned}
 \mathcal{R}(\alpha f + \beta g) &\stackrel{\text{Def.5}}{=} \int_{-\gamma(s)}^{\gamma(s)} (\alpha f + \beta g) dt \\
 &\stackrel{\text{Lin.}}{=} \alpha \int_{-\gamma(s)}^{\gamma(s)} f dt + \beta \int_{-\gamma(s)}^{\gamma(s)} g dt \\
 &= \alpha \mathcal{R}f + \beta \mathcal{R}g.
 \end{aligned} \tag{3.16}$$

Der Übersichtlichkeit halber wurden in (3.16) die Argumente weggelassen, also für alle $f \in \Omega$ gilt $f = f(s\omega(\theta) + t\omega^\perp(\theta))$.

Stetigkeit von \mathcal{R}

Mit dem obigen Ergebnis kann man die Theorie der linearen Operatoren in normierten Räumen auf \mathcal{R} anwenden. Nun wollen wir die Stetigkeit von \mathcal{R} überprüfen.

Lemma 1. *Der Operator $R : L^2(\Omega) \rightarrow L^2(Z)$ ist stetig.*

Beweis. Das Ziel ist zu zeigen, dass das Bild von \mathcal{R} beschränkt ist. Demnach ist jeder linearer beschränkter Operator stetig (s.a. [1, S.82]). Wir führen den Beweis in zwei Schritten durch. Zuerst geben wir eine Abschätzung für $|\mathcal{R}f(s, \theta)|^2$ und dann für $\|\mathcal{R}\|_2$.

- i. Sei nun $f \in L^2(\Omega)$ beliebig, dann gilt mit Benutzung der Cauchy-Schwarzschen Ungleichung für Integrale

$$\begin{aligned}
 |\mathcal{R}f(s, \theta)|^2 &= \left| \int_{-\gamma(s)}^{\gamma(s)} f(s\omega(\theta) + t\omega^\perp(\theta)) dt \right|^2 \\
 &\leq \int_{-\gamma(s)}^{\gamma(s)} 1 dt \int_{-\gamma(s)}^{\gamma(s)} |f(s\omega(\theta) + t\omega^\perp(\theta))|^2 dt \\
 &= 2\gamma(s) \int_{-\gamma(s)}^{\gamma(s)} |f(s\omega(\theta) + t\omega^\perp(\theta))|^2 dt.
 \end{aligned} \tag{3.17}$$

Aus der Gleichung (3.8) wissen wir, dass $\gamma(s) \leq 1$ ist, damit können wir (3.18) weiter abschätzen

$$\begin{aligned}
 |\mathcal{R}f(s, \theta)|^2 &\leq 2\gamma(s) \int_{-\gamma(s)}^{\gamma(s)} |f(s\omega(\theta) + t\omega^\perp(\theta))|^2 dt \\
 &\stackrel{(3.8)}{\leq} 2 \int_{-\gamma(s)}^{\gamma(s)} |f(s\omega(\theta) + t\omega^\perp(\theta))|^2 dt.
 \end{aligned} \tag{3.18}$$

ii. Mit (3.18) lässt sich das Skalarprodukt $\|\mathcal{R}f(s, \theta)\|_{L^2(Z)}^2$ abschätzen:

$$\begin{aligned}
\|\mathcal{R}f(s, \theta)\|_{L^2(Z)}^2 &= \int_0^\pi \int_{-1}^1 |\mathcal{R}f(s, \theta)|^2 ds d\theta \\
&\stackrel{(3.18)}{\leq} 2 \int_0^\pi \int_{-1}^1 \left(\int_{-\gamma(s)}^{\gamma(s)} |f(s\omega(\theta) + t\omega^\perp(\theta))|^2 dt \right) ds d\theta \\
&= 2 \int_0^\pi \left(\int_{-1}^1 \int_{-\gamma(s)}^{\gamma(s)} |f(s\omega(\theta) + t\omega^\perp(\theta))|^2 dt ds \right) d\theta \\
&= 2 \int_0^\pi \left(\int_\Omega |f(x)|^2 dx \right) d\theta \\
&= 2\pi \|f\|_{L^2(\Omega)}^2.
\end{aligned} \tag{3.19}$$

In der vorletzten Zeile von (3.19) wurde $s\omega(\theta) + t\omega^\perp(\theta) = x$ für ein festes θ gesetzt. Damit wurde der Integrationsbereich über t und s zu einem über x gefasst, um insgesamt über ganz Ω integrieren zu können.

Nach dem Wurzelziehen aus (3.19), kommt man zur Erkenntnis, dass \mathcal{R} beschränkt ist, nämlich $\|\mathcal{R}\|_2 \leq \sqrt{2\pi}$. Da \mathcal{R} linear und beschränkt ist, dürfen wir auf seine Stetigkeit schließen. \square

Ist \mathcal{R} linear und stetig so schreiben wir $\mathcal{R} \in \mathcal{L}(L^2(\Omega), L^2(Z))$ ¹¹.

Adjungierter Operator \mathcal{R}

Da \mathcal{R} stetig von $L^2(\Omega)$ nach $L^2(Z)$ abbildet, existiert zu \mathcal{R} ein stetiger adjungierter Operator \mathcal{R}^* , d.h.

$$\forall f \in L^2(\Omega) \wedge g \in L^2(Z) : \langle \mathcal{R}f, g \rangle_{L^2(Z)} = \langle f, \mathcal{R}^*g \rangle_{L^2(\Omega)}. \tag{3.20}$$

Lemma 2. Der adjungierte Operator \mathcal{R}^* zu \mathcal{R} ist gegeben durch

$$\begin{aligned}
\mathcal{R}^* : L^2(Z) &\rightarrow L^2(\Omega) \\
\mathcal{R}^*g(x) &= \int_0^\pi g(x^T \omega(\theta), \theta) d\theta.
\end{aligned} \tag{3.21}$$

Bemerkung 3. Bevor wir (3.20) für (3.21) zeigen, wollen wir die letztere Gleichung etwas besser in unsere Vorstellung einprägen. Sei dazu die Abb. 3.7 betrachtet. Wir wählen einen beliebigen Punkt $x = (\tilde{x}_1, \tilde{x}_2)^T \in \Omega$ und einen Wert $s \in [-1, 1]$. So existiert folgende Beziehung zwischen x und s

$$x^T \omega(\theta) = (s \cos(\theta), s \sin(\theta)) \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} = s(\cos^2(\theta) + \sin^2(\theta)) = s. \tag{3.22}$$

¹¹ $\mathcal{L}(X, Y)$: Raum der stetigen linearen Abbildungen zwischen normierten Räumen X und Y .

Wobei wir gemäß der Abb.3.7 für $x = (s \cos(\theta), s \sin(\theta))^T$ schreiben dürfen.

(3.21) stellt eine Ungefilterte Rückprojektion dar. Man kann sich das folgendermaßen vorstellen. Für ein festes x in der Ebene werden entsprechende Projektionen an der Stelle s und Winkel θ zurück auf den Punkt x projiziert. Nach der Addition der Projektionen auf x über alle Winkel, bekommt man den Wert der ungefilterten Rückprojektion an dem Punkt x .

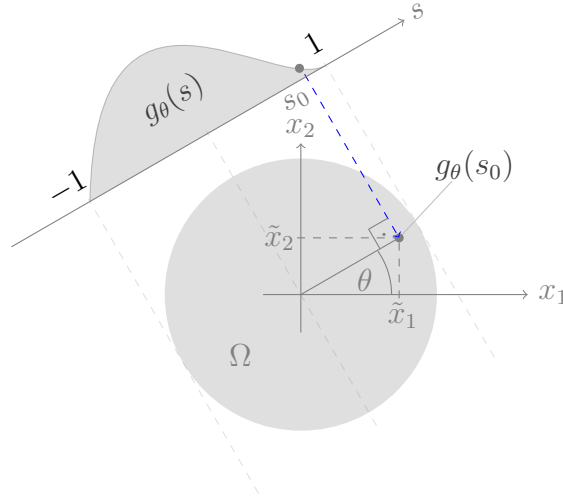


Abbildung 3.7.: Ein Beispiel der Rückprojektion eines Projektionswerts $g_\theta(s_0)$ auf einen Punkt $x \in \Omega$. Mit $g_\theta(s)$ ist hier eine Projektion bei einem festen Winkel θ bezeichnet, also $g_\theta(s) = g(s, \theta)$.

Beweis von Lemma 2. Seien $f \in L^2(\Omega)$ und $g \in L^2(Z)$ beliebig, dann gilt

$$\begin{aligned} \langle \mathcal{R}f, g \rangle_{L^2(Z)} &= \int_Z (\mathcal{R}f(s, \theta)) g(x^T \omega(\theta), \theta) \, ds d\theta \\ &\stackrel{(3.10)}{=} \int_0^\pi \int_{-1}^1 \left(\int_{-\gamma(s)}^{\gamma(s)} f(s\omega(\theta) + t\omega^\perp(\theta)) dt \right) g(x^T \omega(\theta), \theta) \, ds d\theta. \end{aligned} \quad (3.23)$$

An dieser Stelle machen wir eine Substitution in der rechten Seite von (3.23). Hier nutzen wir (3.22) aus und schreiben $x^T \omega(\theta) = \omega^T(\theta)x = s$. Dann bekommt man $g(x^T \omega(\theta), \theta) = g(\omega^T(\theta)x, \theta)$ und wir wollen die Funktion f im Punkt x betrachten, also $f(s\omega(\theta) + t\omega^\perp(\theta)) = f(x)$. Somit erhalten wir einen Integrationsbereich über

Ω . Nun führen wir (3.24) weiter:

$$\begin{aligned}
\langle \mathcal{R}f, g \rangle_{L^2(Z)} &= \int_0^\pi \int_{-1}^1 \left(\int_{-\gamma(s)}^{\gamma(s)} f(s\omega(\theta) + t\omega^\perp(\theta)) dt \right) g(x^T \omega(\theta), \theta) ds d\theta \\
&\stackrel{\text{Subst.}}{=} \int_0^\pi \left(\int_{-1}^1 \int_{-\gamma(s)}^{\gamma(s)} f(x) g(\omega^T(\theta)x, \theta) dt ds \right) d\theta \\
&= \int_0^\pi \int_\Omega f(x) g(\omega^T(\theta)x, \theta) dx d\theta \\
&= \int_\Omega f(x) \left(\int_0^\pi g(\omega^T(\theta)x, \theta) d\theta \right) dx \\
&= \langle f, \mathcal{R}^* g \rangle_{L^2(\Omega)}.
\end{aligned} \tag{3.24}$$

Somit ist \mathcal{R}^* adjungierter Operator der Radon Transformation und ist nach Satz von Schauder auch kompakt, wenn \mathcal{R} kompakt ist. ([4, S. 185]). \square

Kompaktheit von \mathcal{R}

Um die Kompaktheit von \mathcal{R} zeigen zu können, werden wir an dieser Stelle einen Umweg einschlagen. Dafür ziehen wir ein Ergebnis aus [5, S. 45] heran. Es handelt sich um einen zu \mathcal{R} adjungierten Operator $\tilde{\mathcal{R}}^*$ folgender Gestalt:

$$\tilde{\mathcal{R}}^* : L^2(Z, \frac{1}{\gamma(s)}) \rightarrow L^2(\Omega). \tag{3.25}$$

Hier wurde $\gamma(s)$ im Sinne der Gleichung (3.8) benutzt. Und der Raum $L^2(Z, \frac{1}{\gamma(s)})$ ist mit dem folgenden Skalarprodukt versehen

$$\langle g, f \rangle = \int_Z g \bar{f} \frac{1}{\gamma(s)} ds d\theta. \tag{3.26}$$

Wir können festhalten dass der Operator $\mathcal{R}' : L^2(\Omega) \rightarrow L^2(Z, \frac{1}{\gamma(s)})$ kompakt ist¹². Die Kompaktheit folgt eben aus der SWZ von \mathcal{R}' . Denn die Singulärwerte von \mathcal{R}'

$$\sigma_j = 2\sqrt{\frac{\pi}{j+1}}, \quad j \in \mathbb{N}$$

bilden eine Nullfolge, was nach einer längeren Rechnung aus [5, S. 43] folgt. Also schreiben wir $\mathcal{R}' \in \mathcal{K}(L^2(\Omega), L^2(Z, \frac{1}{\gamma(s)}))$ ¹³.

¹²Prinzipiell sind \mathcal{R} und \mathcal{R}' gleich, es ist derselbe Operator, der aber $L^2(\Omega)$ in einen anderen Raum abbildet. Die Bezeichnung \mathcal{R}' führen wir ein, um uns Verwechslungen zu ersparen.

¹³ $\mathcal{K}(X, Y) = \{A \in \mathcal{L}(X, Y) \mid A \text{ ist kompakt} \}$

Wir veranschaulichen die entstandene Situation in einem kommutativen Diagramm

$$\begin{array}{ccc}
 L^2(\Omega) & \xrightarrow{\mathcal{R}} & L^2(Z) \\
 \nwarrow \mathcal{R}' & & \nearrow \text{Id} \\
 & L^2(Z, \frac{1}{\gamma(s)}) &
 \end{array}
 \quad (3.27)$$

und stellen fest, dass \mathcal{R} als eine Komposition zweier Abbildungen darstellbar ist. Hier bedeutet Id die Identität auf $L^2(Z)$ dar. Der Unterschied zwischen $L^2(Z)$ und $L^2(Z, \frac{1}{\gamma(s)})$ besteht nur in der Definition des Skalarprodukts auf diesen beiden Räumen.

Insgesamt würde es heißen, dass wir die Komposition $\text{Id} \circ \mathcal{R}' = \mathcal{R}$ bezüglich der Kompaktheit beurteilen müssen. Dafür wird uns ein nützlicher Satz aus [5, S. 27] gute Dienste leisten (ohne Beweis) :

Satz 1. *Seien X, Y und Z normierte Räume. Falls $A \in \mathcal{L}(X, Y)$ und $B \in \mathcal{K}(Y, Z)$ oder $A \in \mathcal{K}(X, Y)$ und $B \in \mathcal{L}(Y, Z)$, dann gilt $BA \in \mathcal{K}(X, Z)$.*

Somit wollen wir zeigen, dass der Operator $\text{Id} : L^2(Z, \frac{1}{\gamma(s)}) \rightarrow L^2(Z)$ beschränkt ist, und somit auch stetig ist. Danach könnten wir den Satz 1 benutzen, um die Kompaktheit von \mathcal{R} zu zeigen.

Sei nun $g(s, \theta) \in L^2(Z)$ beliebig, dann können wir folgende Abschätzung durchführen:

$$\begin{aligned}
 \|\text{Id } g(s, \theta)\|_{L^2(Z)}^2 &= \int_Z |\mathcal{R}f(s, \theta)|^2 ds d\theta \\
 &\stackrel{(3.26)}{\leq} \int_Z |\mathcal{R}f(s, \theta)|^2 \frac{1}{\gamma(s)} ds d\theta \\
 &= \|g(s, \theta)\|_{L^2(Z, \frac{1}{\gamma(s)})}^2.
 \end{aligned}
 \quad (3.28)$$

Die letzte Abschätzung beruht auf der Kenntnis, dass $\gamma(s) \leq 1$ für alle $s \in (-1, 1)$, somit aber $\frac{1}{\gamma(s)} \geq 1$ gilt. Nun haben wir gezeigt, dass $\|\text{Id}\|_2 \leq 1$ beschränkt und demnach stetig ist. Also gilt $\text{Id} \in \mathcal{L}(L^2(Z, \frac{1}{\gamma(s)}), L^2(Z))$. Dieses Ergebnis ist für uns ausreichend um den Satz 1 benutzen zu können, d.h

$$\mathcal{R} \stackrel{(3.27)}{=} \text{Id} \circ \mathcal{R}' \stackrel{\text{Satz 1}}{\implies} \mathcal{R} \in \mathcal{K}(L^2(\Omega), L^2(Z)). \quad (3.29)$$

Nun sind wir am Ende dieses Kapitels angelangt und nach einigen Untersuchungen können wir festhalten, dass der lineare stetige Operator $\mathcal{R} : L^2(\Omega) \rightarrow L^2(Z)$ kompakt ist und einen kompakten adjungierten Operator \mathcal{R}^* hat. Diese Ergebnisse spielen eine zentrale Rolle bei der Singulärwertzerlegung (SWZ) eines kompakten Operators und beleuchten das Problem der Invertierung solcher Operatoren sehr gut.

Die Singulärwertzerlegung (SWZ) von \mathcal{R}

Im Folgenden wird die allgemeine Form der SWZ des Operators \mathcal{R} angegeben, eine konkrete wird in [5, Kap. 2.5] hergeleitet, jedoch aber nur für \mathcal{R}' . Um zu SWZ

von \mathcal{R} zu gelangen, ermittelt man zuerst die Eigenwerte und Eigenvektoren von $\mathcal{R}\mathcal{R}^* : L^2(Z) \rightarrow L^2(Z)$, die als Komposition kompakter Operatoren kompakt ist. Demnach gilt der *Spektralsatz für selbstadjungierte kompakte Operatoren* [5, S. 30] und man bekommt

$$\mathcal{R}\mathcal{R}^*g = \sum_{j=1}^{\infty} \lambda_j \langle g, v_j \rangle_{L^2(Z)} v_j, \quad \text{für alle } g \in L^2(Z). \quad (3.30)$$

Hier bildet $\{\lambda_j\}_{j \in \mathbb{N}} \in \mathbb{R}$ eine Folge mit der absteigenden Anordnung der Eigenwerte $|\lambda_1| \geq |\lambda_2| \geq \dots > 0$ und $\{v_j\}_{j \in \mathbb{N}} \in L^2(Z)$ ein Orthonormalsystem in $\text{Kern}(\mathcal{R})^\perp$. Eine wichtige Tatsache ist, dass die Folge $\{\lambda_j\}_{j \in \mathbb{N}}$ entweder abbricht oder gegen Null konvergiert.

Die Singulärwerte und das Orthonormalsystem $\{u_j\}_{j \in \mathbb{N}}$ des Bildes von \mathcal{R} bekommen wir aus

$$\begin{aligned} (i). \quad \sigma_j &:= \sqrt{\lambda_j} \\ (ii). \quad u_j &:= \sigma_j^{-1} \mathcal{R}v_j, \quad j \in \mathbb{N}. \end{aligned} \quad (3.31)$$

Mit den Ergebnissen aus (3.30) und (3.31) kann nun die SWZ der Radon Transformation angegeben werden

$$\mathcal{R}f = \sum_{j=1}^{\infty} \sigma_j \langle f, v_j \rangle_{L^2(\Omega)} u_j. \quad (3.32)$$

Das Tripel $\{(\sigma_j, v_j, u_j)\}$ nennt man *Singulärsystem* eines kompakten Operators und die Folgen $\{u_j\}_{j \in \mathbb{N}}$ und $\{v_j\}_{j \in \mathbb{N}}$ bilden folgende Orthonormalsysteme

$$\overline{\text{Span}\{u_j | j \in \mathbb{N}\}} = \overline{\text{Bild}(\mathcal{R})} \quad \text{und} \quad \overline{\text{Span}\{v_j | j \in \mathbb{N}\}} = \text{Kern}(\mathcal{R})^\perp. \quad (3.33)$$

Die Gleichung (3.32) wird im nächsten Kapitel bei der Klärung des Begriffs *schlecht gestellter Probleme* sehr hilfreich sein.

A. Implementierung

Die Implementierung aller Funktionen erfolgte im MATLAB R2016a. In diesem Abschnitt werden einige implementierte Funktionen aufgeführt und kurz erläutert. Wegen einer hohen Zeilenanzahl einiger Quellcodes, können sie nicht in voller Länge hier präsentiert werden. Für volle Quellcodes sei man hier auf die beigelegte CD verwiesen.

A.1. Radon Transformation

Die Radon Transformation wurde entsprechend der Gleichung (??) implementiert.

```
1 function [ sinogramm ] = radonTrafo( pic, angles, detWide )
2
3 ... % some argument preparation
4
5 sinogramm = zeros( detWide, length(angles) );
6
7 % loop over all angles
8 for i = 1:anglesNumber
9     % addition of 90 degrees prevents the phase shift of the sinogram
10    rotImg = imrotate( pic, angles(i)+90, 'bilinear', method );
11
12    % The difference is here, that for small picture sizes the summation
13    % has a better result than the trapez integration. Trapez is
14    % another case, which is for big size pictures better.
15    if( smallSize )
16        layer = sum( rotImg, 2 );
17    else
18        layer = trapz( rotImg, 2 );
19    end
20
21    ... % some index preparation for output
22
23    if( detWide < h )
24        sinogramm( :, i ) = layer(dist:h - b);
25    else
26        sinogramm( dist:h + dist-1, i ) = layer;
27    end
28 end % endFor i
29 end % end radonTrafo()
```

Betrachten wir die Zeile 10 in der Funktion `radonTrafo()`. Hier wird das Argument `pic` für jeden Winkel `angles(i)` rotiert. Dies entspricht der Durchstrahlung eines Objekts unter einem Winkel θ . Da hier nur parallelgeometrische Strahlengänge behandelt werden, kann somit das rotierte Bild `pic` entlang ihrer Zeilenpixel aufsummiert werden. Was auch der Summe in (??) gleichen würde. In der Zeile 29/31

wird die Spalte des Sinogramms `sinogramm` an der entsprechenden Stelle `i` mit dem Summenvektor `layer` ergänzt.

A.2. Ungefilterte Rückprojektion

Die Implementierung der ungefilterten Rückprojektion erfolgte entsprechend der Formel (3.21). Jedoch passiert die Rückprojizierung gleich für alle Pixel pro Winkel.

```

1 function [ output ] = backProjection( sinogramm, angels )
2
3     anglesNumber = length(angels);
4
5     % size for output
6     [ numberOfProjections, ~ ] = size( sinogramm );
7     output = zeros( numberOfProjections, numberOfProjections );
8
9     % discrete euklid coordinates, only one axis is needed, because
10    % the grid will be square.
11    [x] = meshgrid(ceil(-numberOfProjections/2):ceil(numberOfProjections/2-1));
12
13    % find the middle
14    middle = floor( numberOfProjections/2 ) + 1;
15
16    % back projection
17    for i = 1:anglesNumber
18        % rotate the coordinates and shift back
19        direction = round(middle + x*sind(angels(i)+90) + x'*cosd(angels(i)+90));
20
21        % indicate the pixels in which the backprojection will be written
22        indices = find(( direction > 0 ) & ( direction <= numberOfProjections ));
23
24        % sum of backprojections
25        output(indices) = output(indices) + ...
26            sinogramm(direction(indices),i)./anglesNumber;
27    end % endFor i
28 end % end backProjection()

```

Die Größe von `output` entspricht (??), also der Detektorbreite, was in diesem Falle die Zeilenanzahl des Sinogramms ist (Zeile 6, 7). In der Zeile 19 wird das erzeugte Koordinatensystem `x` (Zeile 11) gedreht und verschoben. Die Verschiebung der Koordinaten sorgt dafür, dass die Indizierung (Zeile 22) die Breite des Detektors bei Drehung beibehält. In der Zeile 25/26 wird `output` an den entsprechenden Stellen `indices` aufaddiert.

A.3. Gefilterte Rückprojektion

Der unten aufgeführter Codeabschnitt implementiert die gefilterte Rückprojektion. Die ausgelassenen Codeteile entsprechen genau dem Code der ungefilterten Rückprojektion von A.2.

```

1 function [ output ] = filteredBackProjection( sinogramm, angels, filter )
2

```

```

3  ... % some parameter preparations
4
5  % if no filter is given, than a ramp filter will be created
6  if( nargin < 3 )
7      filter = [floor(numberOfProjections/2):-1:0 1:ceil(numberOfProjections
          /2-1)]*2;
8  end
9
10 % back projection
11 for i = 1:anglesNumber
12
13     % filtering part before back projection
14     % -----
15     inpFT = ifftshift( fft( sinogram(:,anglesNumber-i+1)' ) );
16     % filtering in Fourier-Space is only multiplication:
17     filteredProjection = filter.*inpFT;
18     filteredProjection = real( ifft( fftshift( filteredProjection' ) ) );
19     % -----
20
21     ... % back projection of filteredProjection
22
23 end % endFor i
24 end % end backProjection()

```

In Zeile 7 wird eine kleine Absicherung eingeschaltet, falls kein Filter mitgeteilt wurde, so initialisiert die Funktion die Variable `filter` standardmäßig mit einem Rampenfilter. Zeile 15-18 ist die eigentliche Filterung der Projektionen. Ein kleiner Trick wurde hier angewendet. Die schnelle FT (`fft()`) liefert im MATLAB ein verschobenes Spektrum, sodass er mittels `ifftshift()` zurück verschoben werden muss. Nach der Filterung (Zeile 17) wird die schon gefilterte Projektion entsprechend verschoben und mit `ifft()` invers Fouriertransformiert. Danach passiert nur noch die Rückprojektion, wie in `backProjection()`.

A.4. Erzeugung der Systemmatrix von \mathcal{R}

Die Implementierung der Funktion `getSysMat()` (siehe CD-ROM) baut auf der Implementierung der Funktion `paralleltomo()` von [7] auf. Wobei die Funktion `paralleltomo()` für hier nötige Zwecke eingeschränkt und neu implementiert wurde.

A.5. Iterative Rekonstruktion nach Kaczmarz

Der iterative Algorithmus ist in der Funktion `iterativeReconstruction()` implementiert. Der Kernteil der Funktion (Zeile 12-32) läuft über eine *while*-Schleife, daher man eine Abbruchbedingung, die wie folgt definiert werden kann

$$\|Af_t - p\|_2 \leq \delta. \quad (\text{A.1})$$

Das heißt, wir erwarten eine Verbesserung von f_t nach jeder Iteration, so dass die neue Projektion Af_t immer näher an p liegen soll und daher sich die Fehlernorm verringern soll. Anschließend wird der Fehler mit einer Schranke δ verglichen. Diese

setzt sich aus dem Datenfehler $\|\tilde{p} - p\|_2 = \epsilon$ und einer Konstante $\alpha \geq 1$ zusammen, also $\delta = \alpha\epsilon$. Die Konstante α wird größer/gleich Eins gesetzt, weil man nicht erwarten kann, dass der Fehler in der Rekonstruktion kleiner als Datenfehler wird.

```

1 function [output, iter] = iterativeReconstruction( sinogram, angels, maxIterNr
    , delta, randPar, sysMat )
2
3 ... % some parameter preparation
4
5 while( stop > delta && maxIterNr > iter )
6     if(strcmp(randPar, 'random'))
7         % creates randomized index vector for iteration
8         iterVec = randperm(M);
9     else
10        iterVec = 1:M;
11    end
12    % here Kaczmarz algorithm
13    for i = iterVec
14        ai = sysMat(:,i);
15        output = output + relax*((projections(i)-ai'*output)/sum(ai.*ai,1))*ai;
16    end
17    % stop rule  $\|Af_i - p\|_2$ 
18    stop = norm(sysMat'*output - projections);
19    iter = iter + 1;
20 end % endWhile
21 % transform the output into a matrix
22 output = vec2mat( output, detectorSize );
23
24 ... % some output preparation
25
26 end % end iterativeReconstruction()

```

In den Zeilen 6-11 wird anhand des Parameters `randPar` die Art der Berechnung eingestellt, also ob es randomisiert oder vorwärts projektiv gerechnet werden soll. In Zeile 13-16 wird über alle Hyperebenen iteriert, der Ausdruck in Zeile 15 entspricht genau (?). In der Zeile 18 wird die Norm, wie oben beschrieben berechnet und für den Abbruch der *while*-Schleife benutzt. Die *while*-Schleife wird auch abgebrochen, falls die vorgegebene Anzahl der Iterationen überschritten wird (Zeile 5).

A.6. Rekonstruktion durch SWZ

Die Funktion `svdReconstruction()` mit oder ohne der SWZ der Systemmatrix ausgeführt werden. Falls auch die Systemmatrix nicht übergeben wird, so wird die Matrix mittels `getSysMat()` erzeugt und anschließend das singuläre System mit der MATLAB Funktion `svd()` berechnet.

In den Zeilen 6/7 werden zunächst die Singulärwerte extrahiert und anschließend in der Zeile 10 mittels (?) regularisiert. Für die Regularisierungswerte wird keine neue Matrix erzeugt, sie werden an die Stellen der Singulärwerte in die Matrix *S* reingeschrieben (Zeile 11). Die Pseudoinverse wird in der Zeile 14 mittels (?) erzeugt und damit schließlich die Gleichung (?) in der Zeile 17 gelöst.

```

1 function output = svdReconstruction(sinogram, angels, lambda, sysMat, U, S, V)

```

```
2
3  ... some parameter preparation
4
5  % preparation for regularization
6  ind = S ~= 0;
7  sigma = S( ind );
8
9  % thikonov regularization
10 sigma = (sigma)./( lambda^2 + sigma.^2 );
11 S( ind ) = sigma;
12
13 % take the pseudo inverse of sysMat
14 pseudoInvers = V*S'*U';
15
16 % calculate solution
17 output = pseudoInvers*projections;
18
19 ...some output preparation
20
21 end % end svdReconstruction()
```

B. Inhalt der CD-ROM

Format: CD-ROM

B.1. PDF-Dateien

- BAGESLER.pdf - gesamte Bachelorarbeit im PDF-Format

B.2. Matlab Funktionen

- Testskripte:
 1. FBPTTest
 2. svdTest
 3. iterTest
- Funktionen:
 1. radonTrafo()
 2. backProjection()
 3. filteredBackProjection()
 4. getSysMat()
 5. iterativeReconstruction()
 6. svdReconstruction()
- Filter für FBP:
 1. blackmannFilter()
 2. hammingFilter()
 3. shepLogan()
 4. ramLaksh()

Quellenverzeichnis

Literatur

- [1] Klemens Burg, Herbert Haf und Friedrich Wille. *Funktionalanalysis und partielle Differentialgleichungen*. 1. Aufl. Teubner, 1991 (siehe S. 29).
- [2] Thorsten M. Buzug. *Einführung in die Computertomographie*. 1. Aufl. Springer, 2004 (siehe S. 22).
- [3] S. Handrock-Meyer. *Inverse Probleme für die Wärmeleitungsgleichung*. Zeitschrift für Analysis und ihre Anwendungen, S. 503-510, 1988 (siehe S. 6).
- [4] Harro Heuser. *Funktionalanalysis*. 1. Aufl. B.G Teubner Stuttgart, 1975 (siehe S. 32).
- [5] Andres Rieder. *Keine Probleme mit inversen Problemen*. 1. Aufl. Vieweg, 2003 (siehe S. 32–34).
- [6] Ben Schweizer. *Partielle Differentialgleichungen*. 2. Aufl. Springer Spektrum, 2018 (siehe S. 5).

Online- Quellen

- [7] Per Christian Hansen, Jakob Sauer Jorgensen und Maria Saxild-Hansen. *A MATLAB Package of Algebraic Iterative Reconstruction Methods, AIR Tools II, Version 1.0*. URL: <http://people.compute.dtu.dk/pcha/AIRtoolsII/index.html> (besucht am 19.06.2019) (siehe S. 37).

Abbildungsverzeichnis

1.1.	Das Spektrum der elektromagnetischen Strahlung. Quelle: Wikipedia.	2
2.1.	Schematische Darstellung der Gleichung 2.1, mit V das Volumen eines Mediums, dV Rand (orange) von V , j Der Wärmestrom durch V , $v \in \mathbb{R}^3$ Normalenvektor an dV und dS Tangente an dV .	4
3.1.	Schematischer Aufbau einer Röntgenaufnahme im Draufsicht. Links ist das Schwächungsprofil nach Durchgang der Röntgenstrahlen durch ein Medium mit verschiedenen Dichten zu sehen (mit der Annahme, dass orangene Kreise höhere Dichte haben, als die graue Fläche).	21
3.2.	Das Prinzip der Bildrekonstruktion aus CT-Daten mit einer Rückprojektion (a) und zwei Rückprojektionen (b).	22
3.3.	Prinzipieller Aufbau eines CT-Scanners der dritten Generation.	23
3.4.	Das Schichtmodell eines Objekts als Grundlage zur dreidimensionalen Rekonstruktion.	23
3.5.	Intensitätsänderung der Röntgenstrahlung beim Passieren eines homogenen Objekts.	24
3.6.	Parametrisierung des Strahls L auf dem Gebiet Ω .	25
3.7.	Ein Beispiel der Rückprojektion eines Projektionswerts $g_\theta(s_0)$ auf einen Punkt $x \in \Omega$. Mit $g_\theta(s)$ ist hier eine Projektion bei einem festen Winkel θ bezeichnet, also $g_\theta(s) = g(s, \theta)$.	31