



Kernmethoden im Maschinellen Lernen

Bachelorarbeit

zur Erlangung des akademischen Grades
eines Bachelor of Science

durch die Fakultät für Mathematik und Informatik
der FernUniversität in Hagen

vorgelegt von

Olivier Seubert

geboren in Kaiserslautern

am 06. März 1988

2025

Erstgutachter: Prof. Dr. Delio Mugnolo
Zweitgutachterin: Dr. Bingying Lu

Erklärung

Ich erkläre, dass ich die Bachelorarbeit selbstständig und ohne unzulässige Inanspruchnahme Dritter verfasst habe. Ich habe dabei nur die angegebenen Quellen und Hilfsmittel verwendet und die aus diesen wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht. Die Versicherung selbstständiger Arbeit gilt auch für enthaltene Zeichnungen, Skizzen oder graphische Darstellungen. Die Bachelorarbeit wurde bisher in gleicher oder ähnlicher Form weder derselben noch einer anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht. Mit der Abgabe der elektronischen Fassung der endgültigen Version der Bachelorarbeit nehme ich zur Kenntnis, dass diese mit Hilfe eines Plagiatserkennungsdienstes auf enthaltene Plagiate geprüft werden kann und ausschließlich für Prüfungszwecke gespeichert wird.

Abstract

This Bachelor thesis examines the mathematical theory behind the kernel methods in machine learning on the basis of [27, 33]. To this aim, reproducing kernel Hilbert spaces (RKHS) [27], Moore's [2, 21] and Mercer's Theorems [18, 20] are addressed, i. e. their proofs are carried out thoroughly, including all necessary prerequisites.

In collaboration with and under great guidance by [5, 23, 26] this thesis was able to highlight that the version of Mercer's Theorem which is used for the reasoning in [33] and which was introduced by [18] in 1986, is unfortunately wrong in some parts and incorrectly proved in others. So far, this fact seems to only be known in outskirts of the mathematical community [5, 6, 19, 36]. The bulk of claims in [18] was salvageable – in particular with the help of [28] from 1969, which had already shown central partial claims. However, the main result in the version of Mercer's Theorem by [18] (namely the uniform convergence of the integration kernel's series representation) is unproved up to now and at the same time, a central element of the reasoning by [33]. Together with [23], the thesis was able to find an equivalent condition for above-described uniform convergence (called Mercer-Punkt-4-Zusatzbedingung in the thesis), which is weaker than the one used by [18]'s proof and which should be looked at in future works.

Due to the unproved claim in Mercer's Theorem by [18] the thesis discusses impacts on the conclusions in [33] and takes a closer look at special cases of Mercer's Theorem [1, 7, 10] as well as their practical utility in machine learning. It is being tried to generalise the special cases again, by introducing an error tolerance and a hypothesis is proposed concerning a construction rule for machine learning kernels. In closing, the strong relationship between RKHS and Mercer's Theorems is highlighted [27].

The main questions for future consideration resulting from this thesis are:

- Does Mercer's Theorem in the version by [18] hold nonetheless, if one removes the claim towards the uniformly bounded eigenfunctions?
- Is there another prerequisite instead of the Mercer-Punkt-4-Zusatzbedingung, which is simpler to check in machine learning applications and which also provides the uniform convergence in Mercer's theorem (Satz 6.5.1)?

Zusammenfassung

Diese Bachelorarbeit beleuchtet die mathematische Theorie hinter den Kernmethoden im Maschinellen Lernen auf Basis von [27, 33]. Dazu betrachtet sie Reproducing Kernel Hilbert Spaces (RKHS) [27], die Sätze von Moore [2, 21] und Mercer [18, 20] und führt deren Beweise inklusive aller nötigen Vorergebnisse aus.

In Kooperation mit und unter großer Mithilfe von [5, 23, 26] konnte die Arbeit dabei erarbeiten, dass die durch [33] in ihrer Argumentation verwendete, sehr allgemeine Version des Satzes von Mercer – durch [18] in 1986 aufgestellt – in einigen Teilbehauptungen falsch und in anderen nicht korrekt bewiesen ist. Diese Tatsache scheint bisher nur in Randbereichen der mathematischen Gemeinschaft bekannt zu sein [5, 6, 19, 36]. Ein Großteil der Behauptungen in [18] konnte zwar gerettet werden – vor allem mithilfe von [28] aus dem Jahr 1969, da darin zentrale Teilbehauptungen bereits gezeigt waren. Trotzdem bleibt das Hauptergebnis der Version des Satzes von Mercer aus [18] (die gleichmäßige Konvergenz der Reihendarstellung des Integralkerns) zum jetzigen Zeitpunkt unbewiesen und ist gleichzeitig ein zentraler Bestandteil der Argumentation von [33]. Gemeinsam mit [23] ist es der Arbeit zumindest gelungen, eine zur oben beschriebenen gleichmäßigen Konvergenz äquivalente Bedingung zu finden (in der Arbeit als Mercer-Punkt-4-Zusatzbedingung bezeichnet), welche schwächer ist, als die im Beweis durch [18] verwendete und die es in Folgearbeiten weiter zu betrachten gilt.

Wegen der unbewiesenen gleichmäßigen Konvergenz im Satz von Mercer durch [18] diskutiert die Arbeit Auswirkungen auf die Argumentation von [33] und beleuchtet Spezialfälle des Satzes von Mercer [1, 7, 10] sowie ihre praktische Anwendbarkeit im Maschinellen Lernen. Sie versucht zudem durch Zulassen einer Fehlertoleranz die Spezialfälle leicht zu verallgemeinern und stellt eine Hypothese zu einer Konstruktionsvorschrift für Kerne im Maschinellen Lernen auf. Abschließend stellt sie die Verbindung zwischen RKHS und dem Satz von Mercer [27] heraus.

Die sich aus der Arbeit ergebenden Hauptfolgefragen sind:

- Gilt der Satz von Mercer in der Version durch [18] trotzdem, wenn man die uniforme Beschränktheit der Eigenfunktionen in dessen Behauptung entfernt?
- Gibt es eine für das Maschinelle Lernen besser zu prüfende Voraussetzung anstelle der Mercer-Punkt-4-Zusatzbedingung, welche die gleichmäßige Konvergenz im Satz 6.5.1 von Mercer liefert?

Inhaltsverzeichnis

Abstract	v
Zusammenfassung	vii
Symbolverzeichnis	xi
1 Einleitung	1
2 Ein Anwendungsfall – Motivation der Fragestellung	3
3 Reproducing Kernel Hilbert Spaces (RKHS)	5
3.1 Definition und erste Eigenschaften	5
3.2 Der Reproduzierende Kern (RK)	9
4 Der Satz von Moore(-Aronszajn)	13
4.1 Positiv semidefinite Funktionen	13
4.2 Satz von Moore(-Aronszajn)	15
5 Der Kernel-Trick	21
6 Der Satz von Mercer	25
6.1 Etwas Spektraltheorie	25
6.2 Die Wurzel eines positiven Operators	27
6.3 2-summierende Operatoren	30
6.4 $L^2(\Omega, \mu)$ liegt dicht in $L^1(\Omega, \mu)$	32
6.5 Der Satz von Mercer	33
6.6 Konsequenzen für das Maschinelle Lernen	41
6.6.1 Hypothese zum Satz von Mercer im Maschinellen Lernen	43

6.7	Spezialfälle des Satzes von Mercer	45
6.7.1	Stetige Reproduzierende Kerne (RK) auf kompaktem Raum	45
6.7.2	Satz von Mercer für stetige Integralkerne auf Kompaktum	48
6.7.3	Satz von Mercer für metrischen, regulären Borel-Raum	52
6.7.4	Satz von Mercer für Heat Kernels	54
6.8	Interessantes zum Abschluss	56
6.8.1	Notwendiges Kriterium für den Satz von Mercer	56
6.8.2	Der Satz von Mercer und RKHS	56
Literaturverzeichnis		61
Appendix		67
A Funktionalanalytische Grundlagen		67
A.1	Metrische Räume – Vollständige Räume	67
A.2	Normierte Räume – Banachräume	70
A.3	Euklidische/Unitäre Räume – Hilberträume	74
A.4	Operatoren – Funktionale	77
A.5	Spektraltheorie	81
A.6	Maß- und Integrationstheorie	83
A.7	Glättungskerne / Mollifier	87
A.8	Eine Winzigkeit Komplexes sowie Algebra	87
B Vervollständigungen von Vektorräumen		89
C Absolutsummierende Hilbert-Schmidt Operatoren		95
C.1	Absolutsummierende Operatoren	95
C.2	Die Rademacherschen Funktionen	96
C.3	Die Chintschinsche Ungleichung	98
C.4	Der Einbettungsoperator von $\ell^1(\mathbb{N})$ nach $\ell^2(\mathbb{N})$	100
C.5	Absolutsummierende Hilbert-Schmidt (Integral-)Operatoren	101
Index		105

Symbolverzeichnis

$\ \cdot\ _V$	Eine <u>Norm</u> auf dem <u>normierten Raum</u> V	5
$\langle \cdot, \cdot \rangle_V$	Ein <u>Skalarprodukt</u> auf dem <u>euklidischen</u> oder <u>unitären Vektorraum</u> V	5
$\langle \cdot, \cdot \rangle_{X, X'}$	Eine <u>Funktionalauswertung</u> (siehe Def. A.4.7).	30
$\oplus$	Die direkte Summe zweier Mengen.	9
$\cong$	Eine <u>isometrisch-isomorphe</u> Beziehung.	35
$\mathbf{1}_A$	Die (charakteristische) Indikatorfunktion für die Menge A	32
A^T	Die Transponierte der Matrix A	14
$A^\perp$	Das <u>orthogonale Komplement</u> der Teilmenge A eines <u>euklidischen</u> oder <u>unitären</u> Vektorraums.	9
$\mathcal{A}$	Eine <u>Banachalgebra</u> (in der Regel mit Einselement).	25
$C(\Omega)$	Der Raum der <u>stetigen</u> Funktionen von Ω nach $\mathbb{K}$ (siehe Bsp. A.2.10).	27
$\mathbb{C}$	Der Körper der komplexen Zahlen.	5
$\mathbb{C}[X], \mathbb{R}[X]$	Der Raum der komplexen bzw. reellen Polynome.	25
$c_0(\mathbb{N})$	Der Raum der Nullfolgen.	44
$\frac{d}{dt}$	Der (schwache) Differentialoperator (für einparametrische Funktionen).	54
$\dim(V)$	Die Dimension der <u>Vektorraums</u> V	34
$d, \hat{d}$	Eine <u>Metrik</u>	47
$(e^{tT})_{t \geq 0}$	Die <u>Operatoren-Halbgruppe</u> , welche das durch T generierte <u>abstrakte Cauchy-Problem</u> löst, i. d. R. eine <u>symmetrische Markovsche Halbgruppe</u>	54
E_x	Das <u>Auswertungsfunktional</u> an der Stelle x	5
$e, (e_i)_{i \in I}$	Eine/die zu λ bzw. $(\lambda_i)_{i \in I}$ zugehörige(n) (paarweise verschiedene(n)) <u>Eigenfunktion(en)</u> , i. d. R. <u>orthonormiert</u>	27

$\mathcal{F}(X, \mathbb{K})$	Der <u>Vektorraum</u> der Funktionen/Abbildungen aus dem Definitionsbereich X in den Zahlkörper $\mathbb{K}$	5
$\mathcal{G}(\mathcal{A})$	Die Gruppen der <u>invertierbaren Elemente</u> der <u>Banachalgebra</u> $\mathcal{A}$	26
$\mathcal{H}, \tilde{\mathcal{H}}, \hat{\mathcal{H}}$	Ein <u>Hilbertraum</u>	5
$\mathcal{H}_k, \tilde{\mathcal{H}}_k$	Ein <u>Reproducing Kernel Hilbert Space (RKHS)</u> mit <u>reproduzierendem Kern</u> k	9
$\mathcal{H}'$	Der Dualraum $\mathcal{L}(\mathcal{H}, \mathbb{K})$ des <u>Hilbertraums</u> $\mathcal{H}$	5
I	Ein Indexmenge, i. d. R. abzählbar, oft $I = \mathbb{N}$	33
ι	Der <u>Einbettungsoperator</u> von $L^\infty(\Omega, \mu)$ nach $L^2(\Omega, \mu)$	31
Id	Das <u>Einselement</u> einer <u>Banachalgebra</u>	26
$\inf X$	Das <u>Infimum</u> der Menge X	77
$J, \tilde{J}, J_\varepsilon$	Ein <u>Glättungskern</u> / <u>Mollifier</u>	45
$\ker T$	Der <u>Kern</u> des <u>Operators</u> T	18
$\mathbb{K}$	Sammelsymbol für den Körper $\mathbb{R}$ (oder $\mathbb{C}$) der reellen (oder komplexen) Zahlen.	5
k	Im Fall von <u>RKHS</u> : Der <u>reproduzierende Kern</u>	9
k	Im Fall von <u>Hilbert-Schmidt Integraloperatoren</u> : Der <u>Integralkern</u>	33
$L^p(\Omega, \mu)$	Der Funktionenraum, dessen Elemente in der p -ten Potenz <u>integrierbar</u> sind ($p \in [1, \infty]$, siehe Bsp. A.2.14).	30
$\ell^p(\mathbb{N})$	Der Folgenraum, für den die Folgenglieder seiner Elemente in der p -ten Potenz <u>absolutsummierbar</u> sind ($p \in [1, \infty]$, siehe Bsp. A.2.11).	10
$\lambda, (\lambda_i)_{i \in I}$	Ein/die (mit Multiplizität gezählten) <u>Eigenwert(e)</u> eines zugehörigen <u>Operators</u>	27
$\mathcal{L}(X, Y)$	Der Raum der <u>beschränkten Operatoren</u> , die aus dem <u>normierten Raum</u> X in den normierten Raum Y abbilden ($\mathcal{L}(X) := \mathcal{L}(X, X)$).	5
$M_{n,n}(\mathbb{K})$	Der Raum der $n \times n$ -Matrizen über dem Körper $\mathbb{K}$ ($n \in \mathbb{N}$).	14
μ	Ein <u>Maß</u>	30
$N, \tilde{N}, \bar{N}, \hat{N}$	Eine <u>Nullmenge</u>	33
$\mathbb{N}$	Die (positiven) natürlichen Zahlen.	6
p	Ein <u>Polynom</u>	25
$(r_n)_{n \in \mathbb{N}}$	Die <u>Rademacherschen Funktionen</u>	96

$\mathbb{R}$	Der Körper der reellen Zahlen.....	5
$\text{ran } T$	Das <u>Bild</u> des <u>Operators</u> T	77
S	Die <u>Wurzel</u> eines <u>Operators</u>	27
$\sigma(a)$	Das <u>Spektrum</u> von a	25
$\sup X$	Das <u>Supremum</u> der Menge X	28
$\text{span } X$	Die <u>lineare Hülle</u> / der <u>Aufspann</u> der Menge X	6
$(T(t))_{t \geq 0}$	Eine <u>Operatoren-Halbgruppe</u>	54
T	Ein <u>Operator</u>	27
T'	Der <u>adjungierte Operator</u> zu T	32
T_k	Ein <u>Hilbert-Schmidt Integraloperator</u> mit <u>Integralkern</u> k	33
V	Ein <u>Vektorraum</u>	10
$\Phi, \tilde{\Phi}$	In der Arbeit: Eine Abbildung aus der Rohdatenmenge X in den zugehörigen <u>RKHS</u> / <u>Hilbertraum</u>	6
Φ	Im Anhang: Eine <u>Isometrie</u> zwischen dem <u>metrischen Raum</u> (X, d) und $(\hat{X}, \hat{d})$, seiner <u>Vervollständigung</u>	89
X	Oft: Eine (Rohdaten-)Menge. Gelegentlich: Ein <u>metrischer Raum</u>	5
$ X $	Die Anzahl der Elemente der Menge X	10
$\overline{X}$	Der <u>Abschluss</u> der Teilmenge X eines <u>metrischen Raums</u>	9
X°	Das <u>Innere</u> der Teilmenge X eines <u>metrischen Raums</u>	47
∂X	Der <u>Rand</u> der Teilmenge X eines <u>metrischen Raums</u>	69
X'	Der <u>Dualraum</u> $\mathcal{L}(X, \mathbb{K})$ des <u>normierten Raums</u> X	30
$(\hat{X}, \hat{d})$	Im Anhang: Die <u>Vervollständigung</u> des <u>metrischen Raums</u> (X, d)	89
Ω	Eine (Rohdaten-)Menge, wenn auf/mit ihr ein <u>Maßraum</u> existiert.....	30
$\bar{z}$	Die <u>konjugiert komplexe Zahl</u> zu $z \in \mathbb{C}$	6

1

Einleitung

Im Maschinellen Lernen existiert eine weit verbreitete und durchaus erfolgreiche Strategie, in welcher sich der Lösung einer gegebenen Aufgabenstellung eher durch Ausprobieren genähert wird, anstelle durch direktes Verständnis der Strukturen/Zusammenhänge: Beispielsweise werden bei der Entwicklung von **Neuronalen Netzen** zumeist lediglich Anzahl/Reihenfolge/Anordnung/Verknüpfung sowie die Aktivierungsfunktionen der einzelnen sog. **Neuronen** – zum Teil sogar zufällig – variiert, um einer Lösung so empirisch näher zu kommen. Dabei werden die Aktivierungsfunktionen oft „nur“ aus einer festen Klasse als geeignet bekannter Funktionen gewählt. Neuronale Netze zu unterschiedlichen Problemstellungen ähneln sich somit in der Praxis weitestgehend und unterscheiden sich in der Regel lediglich in ihrer Variation und Komplexität.

Diese Strategie birgt einige Vorteile: Zum einen lassen sich damit Problemstellungen behandeln, zu denen keine bekannten Gesetzmäßigkeiten vorliegen oder bei denen die Gefahr besteht, dass der zuständige Entwickler die Problemstellung wegen einer (ihm ggf. unbekannten) Voreingenommenheit/Bias oder unkorrekten Annahme verfälscht. Zum anderen lässt sich diese empirische Vorgehensweise automatisieren, d. h. Computer können (Teil-)Lösungen oft alleine finden/trainieren.

Nichtsdestotrotz stoßen auch Computer schnell an ihre Kapazitäts- und Komplexitätsgrenzen, wenn man sie ganz selbstständig und ohne Struktur suchen lässt. Ein Teilbereich des Maschinellen Lernens, in dem eine ausführliche mathematische Theorie dazu existiert, weshalb die oben beschriebene Strategie erfolversprechend sein kann bzw. ist, sind die sog. **Kernmethoden**.

Die Kernmethoden sind in Anlehnung an die durch **Fredholm**, **Hilbert** und **Schmidt** rund um 1900 erforschten **Integralkerne** benannt (vgl. z. B. [25]). Genauso wie letztere handelt es sich bei den Kernen in Kernmethoden um Funktionen mit zwei Parametern, welche je nach Anwendungsfall ebenfalls als Integralkern verwendet werden. Im Rahmen seiner Forschung zu Letzteren bewies **Mercer** [20] im Jahr 1909 einen später nach ihm benannten Satz, welcher von zentraler Bedeutung für die Kernmethoden im Maschinellen Lernen werden sollte. 1935 veröffentlichte **Moore** [21] ein weite-

res wichtiges Puzzleteil (vgl. [27]), wobei seine Veröffentlichung nicht mehr erhältlich zu sein scheint. **Aronszajn** [2] zitierte Moores Ergebnis ein weiteres Mal im Jahr 1950, weshalb dieses in der Literatur den Namen *Satz von Moore* bzw. *Satz von Moore-Aronszajn* erhielt. Zwischen den 1980er und 2010er Jahren entdeckten Forscher und Anwender im Maschinellen Lernen die zwei Sätze für ihre Zwecke (wieder) und entwickelten darauf basierend die Theorie der Kernmethoden.

An dieser Stelle findet diese Arbeit ihren Einstieg: Mit [33] existiert ein gutes, etwas mehr auf Anwender abzielendes Kompendium zu den Kernmethoden im Maschinellen Lernen. Dieses als Basis nehmend, soll die mathematische Grundlage der Kernmethoden-Theorie zusammengestellt und studiert werden. Ziel ist es dabei, die Theorie aus Sicht eines Mathematik-Studierenden im Bachelor oder Informatik-Studierenden mit mathematischer Vorbildung zu erarbeiten, d. h. die in [33] teilweise nur angedeuteten Zusammenhänge mathematisch sauber zu belegen, etwaige Beweislücken zu schließen und kleine Sprünge in der Argumentation zu überbrücken/erläutern.

Beginnend mit einem kleinen Anwendungsbeispiel und der Definition von **Reproducing Kernel Hilbert Spaces (RKHS)** wird der Weg über den **Satz 4.2.1 von Moore(-Aronszajn)** und dessen zentralem Ergebnis für das Maschinelle Lernen – dem sog. „**Kernel-Trick**“ – fortgesetzt. Dies wird zunächst die (theoretische) Existenz von auf Kernmethoden basierenden Lösungen vieler Problemstellungen im Maschinellen Lernen liefern.

Diese theoretische Existenz von Lösungen, welche der Satz von Moore liefert, zwingt den Anwender leider aber oftmals in unendlich-dimensionale Vektorräume und kann den praktischen Nutzen damit in manchen Fällen zunichtemachen. Abhilfe schafft der **Satz 6.5.1 von Mercer**, der eine endlich-dimensionale Approximation des Problems ermöglicht (vgl. [32]) und, wie durch [33] argumentiert, einen geometrischen Rahmen um die zugrunde liegenden Daten spannt.

Der Satz von Mercer bedarf theoretischer Vorergebnisse, damit der Beweis korrekt ausgeführt werden kann: Über ein wenig **Spektraltheorie** lässt sich die **Wurzel positiver Operatoren** finden und der **Einbettungsoperator von $L^\infty(\Omega, \mu)$ nach $L^2(\Omega, \mu)$** stellt sich als **2-summierend** heraus, sofern der Maßraum (Ω, μ) endlich ist. Zudem besitzt der Satz von Mercer eine Voraussetzung, welche seine Anwendung im Maschinellen Lernen erschwert. Aus diesem Grund werden für die Praxis relevante Spezialfälle und Konsequenzen diskutiert. Abschließend wird die Verbindung zwischen dem Satz von Mercer (**Kapitel 6**) und dem Satz von Moore (**Kapitel 4**)/RKHS (**Kapitel 3**) aufgezeigt.

Zum mathematischen Verständnis der Arbeit werden die Kenntnisse eines universitären Kurses der Funktionalanalysis (z. B. [25]) oder entsprechenden Lehrbuchs (z. B. [31]) vorausgesetzt. Einige wichtigen funktionalanalytischen Begriffe (aus [25] zitiert) liegen dazu in **Anhang A** zum Nachschlagen vor. Es wird zudem empfohlen, die Arbeit mittels eines PDF-Betrachters zu lesen, welcher eine Vorschau von bzw. ein Springen zu und zurück von Referenzen/Verlinkungen ermöglicht.

Schließlich, aber nicht zuletzt, möchte sich der Autor herzlich für die sehr solidarische, hilfsbereite, freundliche und überaus fruchtbare Betreuung und Kooperation durch/mit Prof. Dr. Mugnolo, Dr. Lu und Herrn Bifulco bedanken (dessen Titel sich zum Zeitpunkt des Lesens dieser Arbeit bereits dem der zwei Vorbenannten angeglichen haben sollte) sowie bei den korrigierenden Freiwilligen für ihre (wahrscheinlich unerschöpfliche) Suche nach den vielen Kommafehlern. Ein besonderer Dank gebührt auch der Lebensgefährtin des Autors, Tetiana Driabova, da sie die Zeichnungen für die Ergebnis-, Beispiel- und Frageboxen anfertigte und generell Zweiteren in der Zeit der Erstellung dieser Arbeit noch verschrobener aushielt als gewöhnlich.

2

Ein Anwendungsfall – Motivation der Fragestellung

Bevor die in dieser Arbeit zentral betrachteten Objekte – die **Reproducing Kernel Hilbert Spaces (RKHS)** – formal definiert werden, sei an dieser Stelle zunächst ein kleines Anwendungsbeispiel aus dem **Maschinellen Lernen (ML)** zur besseren Einordnung vorgestellt:



Beispiel 2.1.1 (Klassifikationsproblem). Ein Algorithmus ist gesucht, der für ein ihm übergebenes Element entscheiden kann, ob es zu einer bestimmten Gruppe gehört [1] oder nicht [-1]. Man stelle sich dazu beispielsweise *Bilder* (oder *Wörter*) als Elemente vor und als Entscheidung, ob es sich um ein *Katzenbild* (oder *weibliches Substantiv*) handelt oder nicht.

	Katze [1]		keine Katze [-1]	Lampe	feminines Substantiv [1]	blau	kein feminines Substantiv [-1]
---	--------------	---	---------------------	--------------	--------------------------------	-------------	---

Abbildung 2.1: Klassifikationsaufgabe

Wie im vorangegangenen Beispiel zu sehen ist, liegen die Daten in Problemstellungen des Maschinellen Lernens in der Regel nicht in einer bekannten Ordnung vor (z. B. $1 < 2$), es ist keine automatische Metrik oder Norm für die Datenmenge ersichtlich (z. B. wie weit ist das Wort *Lampe* vom Wort *blau* im Kontext von femininen Substantiven „entfernt“?), geschweige denn, dass aufwendigere Operationen wie z. B. **Winkelmessungen** zwischen den Elementen der Datenmenge möglich sind.

Auf der anderen Seite benötigen Algorithmen im Maschinellen Lernen genau diese Messmöglichkeiten und Operationen, um Lösungen generieren bzw. Resultate vorhersagen zu können. Zur Entwicklung von Algorithmen ist es somit sinnvoll, den betrachteten Datenraum durch einen Hilbertraum abzubilden, in dem genau diese benötigten Längen- und Winkelmessungen möglich sind.

Beispiel 2.1.2 (Klassifikationsproblem im Hilbertraum). Gelingt es, die Klassifikationsdaten in einen geeigneten Hilbertraum so abzubilden, dass die beiden Gruppen/Klassen/Datenwolken der zu klassifizierenden Elemente räumlich echt getrennt voneinander liegen, so lässt sich eine (Hyper-)Ebene zwischen den Datenwolken als Entscheidungsgrundlage für den Algorithmus legen.

Der Algorithmus muss dann lediglich messen, ob ein unbekanntes Element auf der linken (Gruppe [1]) oder rechten Seite (Gruppe [-1]) der Ebene liegt und gibt dies als Ergebnis zurück.

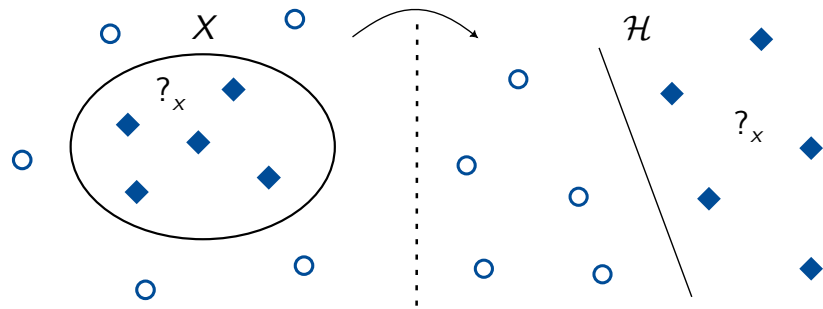


Abbildung 2.2: Transformation der Datenmenge X in geeigneten Hilbertraum $\mathcal{H}$

In Hilberträumen können Algorithmen also basierend auf geometrischen Prinzipien entwickelt werden, was in einer normalen (z. B. ungeordneten) Menge nicht direkt möglich ist. Die Aufgabe, einen maschinellen Lern-Algorithmus zu entwickeln, lässt sich somit in vielen Fällen in die Aufgabenstellung transformieren, einen geeigneten Hilbertraum für die betrachtete Fragestellung zu finden, da in diesem ggf. ein (geometrischer) Algorithmus zur Lösung bereits bekannt ist.

In den folgenden Kapiteln wird hergeleitet werden, dass die **RKHS** für genau diese Herangehensweise einen guten Prototypen darstellen und für die Praxis relevante Eigenschaften besitzen. Zusätzlich zu dem hier bereits Zitierten sind ebenfalls noch weitere Ausführungen in [27, 33] zu finden.

3

Reproducing Kernel Hilbert Spaces (RKHS)

Nach dem einführenden Beispiel in [Kapitel 2](#) soll nun das zentrale Betrachtungsobjekt dieser Arbeit – der **Reproducing Kernel Hilbert Space (RKHS)** – eingeführt und Eigenschaften diskutiert werden. Die Inhalte dieses Kapitels sind [\[27, 33, Chapter 1+2, Chapter I.2\]](#) entnommen, wobei sich bei der Struktur etwas mehr an [\[27\]](#) orientiert wurde. Für einen weiteren Blickwinkel wird [\[33\]](#) empfohlen.

Im Folgenden sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ der reelle oder komplexe Zahlenkörper, X eine beliebige Menge und mit $\mathcal{F}(X, \mathbb{K}) := \{f : X \rightarrow \mathbb{K}\}$ der Vektorraum ([Def. A.2.1](#)) der Funktionen/Abbildungen bezeichnet, die aus dem Definitionsbereich X in den Zahlenkörper $\mathbb{K}$ abbilden (siehe [Bsp. A.2.2](#)).

3.1 Definition und erste Eigenschaften

Definition 3.1.1 (Reproducing Kernel Hilbert Space (RKHS) [\[27, Def. 1.1\]](#)).

Sei $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ ein Hilbertraum ([Def. A.3.5](#)) mit einer Teilmenge $\mathcal{H}$ von $\mathcal{F}(X, \mathbb{K})$. $\mathcal{H}$ wird **Reproducing Kernel Hilbert Space (RKHS)** auf X (über $\mathbb{K}$) genannt, wenn für alle $x \in X$ das lineare Auswertungsfunktional (evaluation functional, [Def. A.4.1](#)) an der Stelle x

$$E_x : \mathcal{H} \rightarrow \mathbb{K}, \quad E_x(f) := f(x) \quad (3.1)$$

beschränkt ist ($E_x \in \mathcal{H}' := \mathcal{L}(\mathcal{H}, \mathbb{K})$, [Def.+Satz A.4.2/Satz A.4.4/Def. A.4.7](#)), d. h. für alle $x \in X$ existiert ein $M_x \geq 0$, sodass:

$$|E_x(f)| \leq M_x \|f\|_{\mathcal{H}} \quad \text{für alle } f \in \mathcal{H}. \quad (3.2)$$

Bemerkung 3.1.2. Die Linearität des Auswertungsfunktionals E_x folgt für jedes $x \in X$ bereits aus der Tatsache, dass $\mathcal{F}(X, \mathbb{K})$ ein Vektorraum ist und muss nicht gesondert geprüft werden. Die Beschränktheit von E_x für jedes $x \in X$ ist somit die eigentlich definierende Eigenschaft eines RKHS.

Beispiel 3.1.3 (Beispiel-RKHS [33, S. 32]).

Man betrachte für $n \in \mathbb{N}$ die Diskretisierung $X := \{x_0, \dots, x_n\}$ des Intervalls $[0, 1]$ gegeben durch $x_i := \frac{i}{n}$ für $i \in \{0, \dots, n\}$ und damit den reellen Vektorraum / die lineare Hülle (Def.+Satz A.2.3)

$$\mathcal{H} := \text{span}\{f_0, \dots, f_n\} \quad \text{mit} \quad f_i : x \mapsto \exp(-|x - x_i|^2) \quad \text{für } x \in X, i \in \{0, \dots, n\}. \quad (3.3)$$

(Zur Veranschaulichung wird auf Abbildung 3.1 verwiesen.) Die Elemente $f \in \mathcal{H}$ besitzen die Form

$$f(x) = \sum_{i=0}^n c_i^{(f)} \exp(-|x - x_i|^2) \quad \text{für } x \in X \text{ und gewisse } c_0^{(f)}, \dots, c_n^{(f)} \in \mathbb{R}. \quad (3.4)$$

Mit dieser Darstellung kann ein Skalarprodukt (Def. A.3.1) auf $\mathcal{H}$ definiert werden (welches tatsächlich wohldefiniert ist – einsehbar mittels identischer Rechnung wie in (4.8)-(4.9)), durch

$$\langle \cdot, \cdot \rangle_{\mathcal{H}} : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}, \quad \langle f, g \rangle_{\mathcal{H}} := \sum_{i=0}^n \sum_{j=0}^n c_i^{(f)} c_j^{(g)} \exp(-|x_i - x_j|^2) \quad \text{für } f, g \in \mathcal{H}, \quad (3.5)$$

denn $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ ist (bi-)linear wegen der Linearität der Summe, symmetrisch wegen der Symmetrie der Doppelsumme sowie von $|\cdot|$ und zudem für alle $f \in \mathcal{H}$ gilt (wegen $t^2 \geq 0$, $\exp(t) > 0$ für alle $t \in \mathbb{R}$):

$$\langle f, f \rangle_{\mathcal{H}} \stackrel{(3.5)}{=} \sum_{i=0}^n \sum_{j=0}^n \left(c_i^{(f)}\right)^2 \exp(-|x_i - x_j|^2) \geq 0 \quad \text{und} \quad \langle f, f \rangle_{\mathcal{H}} = 0 \stackrel{(3.5)}{\Leftrightarrow} f = 0. \quad (3.6)$$

Da zudem $\mathcal{H}$ als lineare Hülle einer endlichen Menge endlich-dimensional und somit gemäß Satz A.2.7 vollständig bezüglich der durch $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ induzierten Norm ist (Bem. A.3.3), folgt, dass $\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}}$ ein Hilbertraum ist. Es verbleibt, die Auswertungsfunktionale zu prüfen – für $j \in \{0, \dots, n\}$, $f \in \mathcal{H}$ gilt:

$$|E_{x_j}(f)| \stackrel{(3.1)}{\stackrel{(3.4)}{=}} \left| \sum_{i=0}^n c_i^{(f)} \exp(-|x_j - x_i|^2) \right| \stackrel{(3.5)}{\stackrel{(3.3)}{=}} |\langle f, f_j \rangle_{\mathcal{H}}| \stackrel{\text{A.3.2}}{\leq} \sqrt{|\langle f_j, f_j \rangle_{\mathcal{H}} \langle f, f \rangle_{\mathcal{H}}|} \stackrel{\text{A.3.3}}{=} \|f_j\|_{\mathcal{H}} \|f\|_{\mathcal{H}}. \quad (3.7)$$

Da f und j beliebig gewählt waren, gilt somit gemäß Def.+Satz A.4.2, dass E_x beschränkt ist für jedes $x \in X$. Ergo ist $\mathcal{H}$ tatsächlich gemäß Def. 3.1.1 ein RKHS auf X (über $\mathbb{R}$).

Diese zugegebenermaßen sehr abstrakte Definition eines RKHS lässt nicht unbedingt sofort erahnen, warum ein solcher für die in Kapitel 2 betrachtete Fragestellung geeignet sein soll. Der Eindruck verbessert sich nach einer Anwendung des Darstellungssatzes von Fréchet-Riesz (Satz A.4.10):

Lemma 3.1.4 (Abbildung von X in RKHS [27, 33, Def. 1.2, 2.2.3]).

Sei $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ ein RKHS auf X (über $\mathbb{K}$). Dann existiert eine Abbildung $\Phi : X \rightarrow \mathcal{H}$ und (mit dieser definiert) eine Funktion $k : X \times X \rightarrow \mathbb{K}$ mit

$$(\Phi(x))(y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}} \quad \text{und} \quad k(x, y) := \overline{\langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}} \quad \text{für alle } x, y \in X. \quad (3.8)$$

Es gelten dann die folgenden Zusammenhänge ($x, y \in X$ und $h \in \mathcal{H}$):

$$k(x, y) = \overline{k(y, x)} \quad \text{und} \quad k(x, y) = \overline{\langle k(\cdot, x), k(\cdot, y) \rangle_{\mathcal{H}}}, \quad (3.9)$$

$$\Phi(x) = k(\cdot, x) \quad \text{und} \quad \|\Phi(x)\|_{\mathcal{H}} = \sqrt{k(x, x)}, \quad (3.10)$$

$$h(x) = \langle h, k(\cdot, x) \rangle_{\mathcal{H}}. \quad (3.11)$$

Beweis. (Die Beweisideen sind [27, 33, Def. 1.2, 2.2.3] entnommen.)

$\mathcal{H}$ ist als RKHS insbesondere ein Hilbertraum und jedes Auswertungsfunktional $E_x : \mathcal{H} \rightarrow \mathbb{K}$ ist per Definition des RKHS für alle $x \in X$ ein beschränktes Funktional, d.h. es gilt gemäß Def. A.4.7: $E_x \in \mathcal{H}'$. Damit sind bereits alle Voraussetzungen des Darstellungssatzes von Fréchet-Riesz erfüllt und dieser liefert für jedes $x \in X$ ein eindeutiges $\varphi_x \in \mathcal{H}$, sodass

$$E_x(h) \stackrel{\text{A.4.10}}{=} \langle h, \varphi_x \rangle_{\mathcal{H}} \quad \text{für alle } h \in \mathcal{H}. \quad (3.12)$$

Die Abbildung $\Phi : X \rightarrow \mathcal{H}$, $x \mapsto \Phi(x) = \varphi_x$ ist somit wohldefiniert und es gilt für alle $x, y \in X$ wie gewünscht:

$$(\Phi(x))(y) = \varphi_x(y) \stackrel{(3.1)}{=} E_y(\varphi_x) \stackrel{(3.12)}{=} \langle \varphi_x, \varphi_y \rangle_{\mathcal{H}} = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}. \quad (3.13)$$

Die in (3.8) definierte Funktion k hat dann die weiteren behaupteten Eigenschaften – für $x \in X$ gilt

$$\Phi(x) \stackrel{(3.13)}{=} \langle \Phi(x), \Phi(\cdot) \rangle_{\mathcal{H}} \stackrel{\text{A.3.1}}{=} \overline{\langle \Phi(\cdot), \Phi(x) \rangle_{\mathcal{H}}} \stackrel{(3.8)}{=} k(\cdot, x), \quad (3.14)$$

womit die linke Seite von (3.10) gezeigt ist und damit ebenfalls (3.11) sowie die rechte Seite von (3.9) folgen – für $x, y \in X$ und $h \in \mathcal{H}$ gelten:

$$h(x) \stackrel{(3.1)}{=} E_x(h) \stackrel{(3.12)}{=} \langle h, \Phi(x) \rangle_{\mathcal{H}} \stackrel{(3.14)}{=} \langle h, k(\cdot, x) \rangle_{\mathcal{H}} \quad \text{und} \quad (3.15)$$

$$k(x, y) \stackrel{(3.8)}{=} \overline{\langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}} \stackrel{(3.14)}{=} \overline{\langle k(\cdot, x), k(\cdot, y) \rangle_{\mathcal{H}}}. \quad (3.16)$$

Schließlich folgen die linke Seite von (3.9) und die rechte Seite von (3.10) mit den Eigenschaften des Skalarproduktes – für alle $x, y \in X$ gelten:

$$k(x, y) \stackrel{(3.8)}{=} \overline{\langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}} \stackrel{\text{A.3.1}}{=} \overline{\langle \Phi(y), \Phi(x) \rangle_{\mathcal{H}}} \stackrel{(3.8)}{=} \overline{k(y, x)} \quad \text{und} \quad (3.17)$$

$$\|\Phi(x)\|_{\mathcal{H}} \stackrel{\text{A.3.3}}{=} \sqrt{|\langle \Phi(x), \Phi(x) \rangle_{\mathcal{H}}|} \stackrel{\text{A.3.1}}{\stackrel{\text{A.8.1}}{=}} \sqrt{\langle \Phi(x), \Phi(x) \rangle_{\mathcal{H}}} \stackrel{(3.8)}{=} \sqrt{k(x, x)}. \quad (3.18)$$

Damit sind alle Behauptungen gezeigt. □

Dieses eher elementare Lemma liefert einen kleinen Schatz an Eigenschaften, die im Folgenden eine wichtige Rolle spielen werden. Zudem stellt es die Verbindung zwischen **RKHS** und dem anfänglich betrachteten Problem im **Maschinellen Lernen** her:

Im Beispiel „Klassifikation im Hilbertraum“ und den darauf folgenden Ausführungen wurde eine Abbildung gesucht, welche aus der Menge der Roh-/Klassifikationsdaten X in einen geeigneten Hilbertraum $\mathcal{H}$ abbildet. Die Abbildung Φ in Lemma 3.1.4 erfüllt genau dies.

Beispiel 3.1.5 (Fortsetzung von Bsp. 3.1.3). Im Bsp. 3.1.3 besteht die Rohdatenmenge X aus den $n + 1$ Punkten $x_0, \dots, x_n \in [0, 1]$. Mit der Abbildung $(\Phi(x_i))(x) := \exp(-|x - x_i|^2) \stackrel{(3.3)}{=} f_i(x) \in \mathcal{H}$ für $x \in X$, $i \in \{0, \dots, n\}$ werden diese in den RKHS $\mathcal{H}$ abgebildet. Die Punkte in der Rohdatenmenge X werden somit auf Funktionen in $\mathcal{H}$ abgebildet, wie Abbildung 3.1 für die Fälle $n = 10$ und $n = 1000$ veranschaulicht.

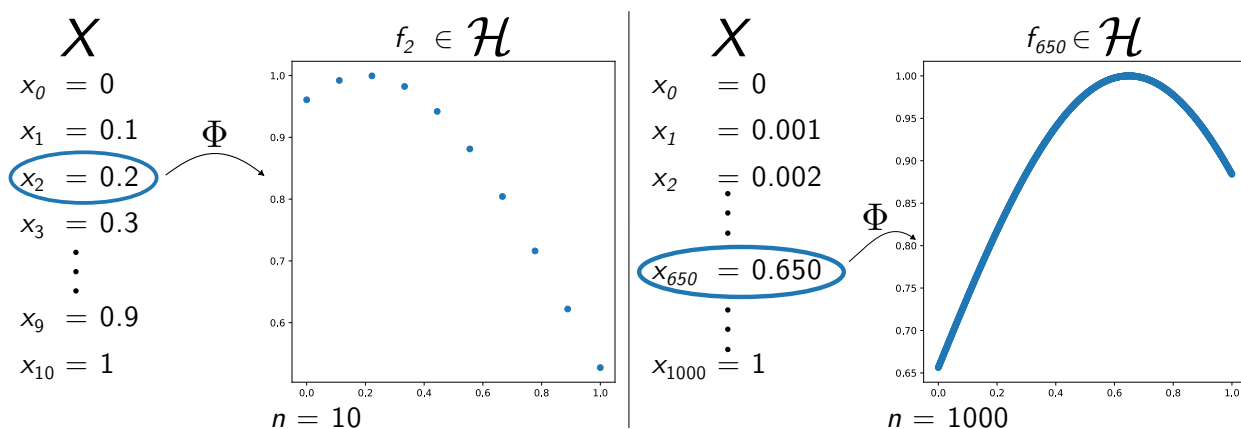


Abbildung 3.1: RKHS $\mathcal{H}$ mit Abbildung $\Phi: X \rightarrow \mathcal{H}$ aus Bsp. 3.1.3 für $n = 10$ bzw. $n = 1000$.

Die mithilfe von Φ definierte Funktion $k: X \times X \rightarrow \mathbb{K}$ besitzt noch praktischere Eigenschaften, weshalb das Augenmerk im Folgenden mehr auf die Funktion k und nicht auf Φ gelegt werden wird:

1. Die Abbildung Φ lässt sich mittels (3.10) jederzeit wieder aus k gewinnen.
2. Mittels (3.8) und (3.10) lassen sich sowohl die **Norm** als auch das **Skalarprodukt** für durch Φ abgebildete Elemente aus X im RKHS $\mathcal{H}$ berechnen. Somit lassen sich geometrische Algorithmen für Aufgabenstellungen im Maschinellen Lernen entwickeln, welche ausschließlich mit der Berechnung von Winkeln oder Abständen auskommen und dessen Berechnungen mithilfe der Funktion k direkt auf der Menge X erfolgen können.
3. Besonders wichtig ist hervorzuheben, dass die Funktion k auf dem Definitionsbereich $X \times X$ wirkt und in einen Zahlenraum abbildet. Über den RKHS $\mathcal{H}$ wird dabei nur „im Hintergrund gelaufen“ und dieser muss bei Kenntnis der Funktion k gar nicht unbedingt selbst bekannt sein. Dieser Fakt wird eine zentrale Rolle in Kapitel 4 und vor allem in Kapitel 5 spielen. Insgesamt wird in diesen beiden Kapiteln der Funktion k noch eine weitere wichtige Rolle zuteil werden.
4. Die Funktion k „rettet“ mittels (3.9) die Symmetrie des Skalarprodukts aus dem RKHS $\mathcal{H}$ in den Rohdatenraum X herüber. Generell weist die Funktion k mit (3.8), (3.9) und (3.10) auch etwas symmetrischere Eigenschaften auf, als Φ es tut. Man bemerke, dass sich im reellen Fall die Beziehung (3.9) zu $k(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}$ für alle $x, y \in X$ ergibt, d. h. die Funktion k stellt genau das Skalarprodukt im reellen RKHS dar. Für den komplexen Fall wurde die Darstellung in der Arbeit in dieser Form gewählt, da sie symmetrischer ist – im Gegensatz zu $k(x, y) = \overline{\langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}} = \langle \Phi(y), \Phi(x) \rangle_{\mathcal{H}}$ für alle $x, y \in X$ gemäß Def. A.3.1.
5. Eine solche Funktion k ließe sich auch für beliebige wohldefinierte Abbildungen $\Phi: X \rightarrow \mathcal{H}$ für einen beliebigen (nicht unbedingt Reproducing Kernel) Hilbertraum definieren. Dann wären zwar nicht alle Eigenschaften aus Lemma 3.1.4 erfüllt, aber Norm und Skalarprodukt ließen sich auch dann berechnen.

Aus diesen Gründen besitzt die Funktion k einen zentralen Namen in Bezug auf RKHS.

3.2 Der Reproduzierende Kern (RK)

Definition 3.2.1 (Der reproduzierende Kern [27, 33, Def. 1.2, Def. 2.9]).

Für ein RKHS $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ auf X (über $\mathbb{K}$) wird die Funktion $k : X \times X \rightarrow \mathbb{K}$, eindeutig definiert mittels (3.8) in Lemma 3.1.4, der **reproduzierende Kern (reproducing kernel / RK) von $\mathcal{H}_k$** genannt. Die Eigenschaft (3.11) nämlich, dass

$$h(x) = \langle h, k(\cdot, x) \rangle_{\mathcal{H}_k} \quad \text{für alle } h \in \mathcal{H}_k, x \in X \quad (3.19)$$

gilt, wird die **reproduzierende Eigenschaft** genannt.

Korollar 3.2.2 (Die reproduzierende Eigenschaft definiert den RK [27, 33, Def. 1.2, Def. 2.9]).

Sei $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ ein RKHS auf X (über $\mathbb{K}$) und sei die Funktion $k : X \times X \rightarrow \mathbb{K}$, sodass für sie die reproduzierende Eigenschaft gilt. Dann ist k der reproduzierende Kern von $\mathcal{H}_k$.

Beweis. Es sei $\tilde{k}$ der reproduzierende Kern von $\mathcal{H}_k$ und es gelten für k obige Voraussetzungen. Betrachtet man im Beweis zu Lemma 3.1.4 wie die Abbildung $\Phi : X \rightarrow \mathcal{H}_k$ gefunden wurde, so war diese mittels (3.12) so definiert, dass

$$h(x) = E_x(h) \stackrel{(3.12)}{=} \langle h, \Phi(x) \rangle_{\mathcal{H}_k} \quad \text{für alle } h \in \mathcal{H}_k, x \in X \quad (3.20)$$

gilt, was sich als dieselbe Gleichung wie die reproduzierende Eigenschaft (3.19) herausstellt. Da k ebenfalls letztere erfüllt, gilt also $\Phi(x) = k(\cdot, x)$ für alle $x \in X$, denn der Darstellungssatz von Fréchet-Riesz besagt, dass zu jedem $x \in X$ das Element $\Phi(x) \in \mathcal{H}_k$ eindeutig bestimmt ist, welches (3.20) erfüllt. Für den reproduzierenden Kern $\tilde{k}$ von $\mathcal{H}_k$ gilt damit für alle $x, y \in X$:

$$\tilde{k}(x, y) \stackrel{(3.8)}{=} \overline{\langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}_k}} \stackrel{A.3.1}{=} \langle k(\cdot, y), k(\cdot, x) \rangle_{\mathcal{H}_k} \stackrel{(3.19)}{=} (k(\cdot, y))(x) = k(x, y). \quad (3.21)$$

Ergo gilt $\tilde{k} = k$, d. h. k ist der reproduzierende Kern von $\mathcal{H}_k$. □

Beispiel 3.2.3 (Fortsetzung von Bsp. 3.1.5). Der **reproduzierende Kern** des RKHS $\mathcal{H}_k$ in Bsp. 3.1.3 besitzt eine kompakte und symmetrische Form: $k : X \times X \rightarrow \mathbb{R}$, $k(x, y) := \exp(-|x - y|^2)$ und ist eine Variante des sog. **Gaußschen Kerns** (vgl. [33, S. 21]).

Ein weiteres interessantes und nützlich werdendes Ergebnis bzgl. des RKHS und seinem zugehörigen RK ist folgendes:

Proposition 3.2.4 (Die reproduzierende Kernfunktionschar liegt dicht im RKHS [27, Prop. 2.1]).

Sei $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ ein RKHS auf X (über $\mathbb{K}$) mit reproduzierendem Kern $k : X \times X \rightarrow \mathbb{K}$. Dann liegt die lineare Hülle der Funktionen $k(\cdot, x)$, $x \in X$ dicht in $\mathcal{H}_k$, d. h. $\overline{\text{span}\{k(\cdot, x) : x \in X\}} = \mathcal{H}_k$ (Def. A.1.7 / Def. A.1.5).

Beweis. (Der Beweis ist [27, Prop. 2.1] entnommen und lediglich um wenige Erläuterungen erweitert.) Es gilt für $A := \text{span}\{k(\cdot, x) : x \in X\}$:

$$h \in A^\perp \quad \Leftrightarrow \quad h(x) \stackrel{(3.19)}{=} \langle h, k(\cdot, x) \rangle_{\mathcal{H}_k} \stackrel{A.3.10}{=} 0 \quad \text{für alle } x \in X \quad \Leftrightarrow \quad h = 0. \quad (3.22)$$

Folglich gilt wegen $A \subseteq \overline{A}$, dass $\overline{A}^\perp \subseteq A^\perp = \{0\}$, und da natürlich $\langle h, 0 \rangle = \langle h, 0 \cdot 0 \rangle \stackrel{\text{A.3.1}}{=} 0 \cdot \langle h, 0 \rangle = 0$ für alle $h \in \mathcal{H}_k \supset \overline{A}$ gilt, folgt, dass $\overline{A}^\perp = \{0\}$.

Nun ist $\overline{A}$ als Abschluss einer linearen Hülle definitionsgemäß ein abgeschlossener Unterraum von $\mathcal{H}_k$ und mittels Satz A.3.11 lässt sich $\mathcal{H}_k$ somit zerlegen in

$$\mathcal{H}_k \stackrel{\text{A.3.11}}{=} \overline{A} \oplus \overline{A}^\perp = \overline{A} \oplus \{0\}. \quad (3.23)$$

Es liegt also jedes Element $0 \neq h \in \mathcal{H}_k$ zwangsläufig in $\overline{A}$ und da A selbst ein Vektorraum ist, gilt auch $0 \in A \subseteq \overline{A}$. Es folgt insgesamt die Behauptung:

$$\mathcal{H}_k = \overline{A} = \overline{\text{span}\{k(\cdot, x) : x \in X\}}. \quad (3.24)$$

□

Korollar 3.2.5. Sei $\mathcal{H}_k$ ein RKHS auf X (über $\mathbb{K}$).

1. Ist X endlich (d. h. $|X| < \infty$), so ist $\mathcal{H}_k$ **endlich-dimensional** und somit isomorph zu $\mathbb{K}^n$ (für ein $n \in \mathbb{N}$).
2. Ist X abzählbar unendlich, so ist $\mathcal{H}_k$ **separabel** (Def. A.1.8). Ist $\mathcal{H}_k$ dann unendlich-dimensional, so ist $\mathcal{H}_k$ isometrisch-isomorph zu

$$\ell^2(\mathbb{N}) := \left\{ (x_n)_{n \in \mathbb{N}} \subset \mathbb{K} : \left(\sum_{k=1}^{\infty} |x_k|^2 \right)^{\frac{1}{2}} \right\}, \quad (3.25)$$

anderenfalls erneut isomorph zu $\mathbb{K}^n$ (für ein $n \in \mathbb{N}$).

Beweis.

1. Ist X endlich, so ist die lineare Hülle $\text{span}\{k(\cdot, x) : x \in X\}$ endlich-dimensional und somit gemäß Satz A.2.7 vollständig. Mit Prop. 3.2.4 folgt, dass $\mathcal{H}_k \stackrel{\text{3.2.4}}{=} \text{span}\{k(\cdot, x) : x \in X\}$ damit endlich-dimensional ist. Der Nachweis der Isomorphie eines beliebigen n -dimensionalen Hilbertraums zu $\mathbb{K}^n$ lässt sich z. B. [25, Theorem 2.2.6+Remark 2.2.7] entnehmen ($n \in \mathbb{N}$).
2. Ist X abzählbar unendlich, so lässt sich zusammen mit der Separabilität von $\mathbb{K}$ und Prop. 3.2.4 zeigen (z. B. mit [25, Prop. 2.5.1]), dass $\mathcal{H}_k$ separabel ist. Der Beweis der isometrischen Isomorphie zu $\ell^2(\mathbb{N})$, falls $\mathcal{H}_k$ unendlich-dimensional ist, lässt sich dann beispielsweise ebenfalls [25, Lemma 6.5.18] entnehmen.

□

Zudem liefert Prop. 3.2.4 noch folgendes instruktives

Korollar 3.2.6 (Eineindeutige Beziehung zwischen RK und RKHS, [27, S. 27]).

Die Beziehung zwischen einem RKHS $\mathcal{H}_k$ und seinem reproduzierenden Kern k ist eineindeutig, d. h. $\mathcal{H}_k$ kann keinen weiteren reproduzierenden Kern $\tilde{k}$ haben, mit $k \neq \tilde{k}$ und es gibt keinen weiteren RKHS $\tilde{\mathcal{H}}_k$, welcher den reproduzierenden Kern k besitzt und nicht isometrisch-isomorph zu $\mathcal{H}_k$ ist.

Beweis. (Die Beweisideen sind [27, S. 27] entnommen.)

Die Hinrichtung folgt bereits aus der Def. 3.2.1 des reproduzierenden Kerns bzw. genauer aus dem Darstellungssatz von Fréchet-Riesz, mit welchem k in Lemma 3.1.4 eindeutig konstruiert wird.

Die Rückrichtung liefert dann Prop. 3.2.4: Seien $\mathcal{H}_k$ und $\tilde{\mathcal{H}}_k$ zwei RKHS auf X über $\mathbb{K}$, die beide den reproduzierenden Kern k haben. Prop. 3.2.4 besagt dann, dass die lineare Hülle $V := \text{span}\{k(\cdot, x) : x \in X\}$ dicht in $\mathcal{H}_k$ und $\tilde{\mathcal{H}}_k$ liegt. Insbesondere gilt also $V \subseteq \mathcal{H}_k \cap \tilde{\mathcal{H}}_k$ und für alle $f, g \in V$ gibt es $n, m \in \mathbb{N}$ sowie zugehörige $x_1, \dots, x_n, y_1, \dots, y_m \in X$ und $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_m \in \mathbb{K}$ mit

$$f = \sum_{i=1}^n \alpha_i k(\cdot, x_i) \quad \text{und} \quad g = \sum_{j=1}^m \beta_j k(\cdot, y_j). \quad (3.26)$$

Damit ergibt sich für das jeweilige Skalarprodukt beider RKHS (mit $*$ $\in \{\mathcal{H}_k, \tilde{\mathcal{H}}_k\}$):

$$\langle f, g \rangle_* \stackrel{(3.26)}{=} \left\langle \sum_{i=1}^n \alpha_i k(\cdot, x_i), \sum_{j=1}^m \beta_j k(\cdot, y_j) \right\rangle_* \stackrel{\text{A.3.1}}{=} \sum_{i=1}^n \sum_{j=1}^m \alpha_i \overline{\beta_j} \langle k(\cdot, x_i), k(\cdot, y_j) \rangle_* \quad (3.27)$$

$$\stackrel{(3.9)}{=} \sum_{i=1}^n \sum_{j=1}^m \alpha_i \overline{\beta_j} \overline{k(x_i, y_j)} \quad \text{für alle } f, g \in V \subseteq \mathcal{H}_k \cap \tilde{\mathcal{H}}_k. \quad (3.28)$$

Ergo sind die beiden Skalarprodukte $\langle \cdot, \cdot \rangle_{\mathcal{H}_k}$ und $\langle \cdot, \cdot \rangle_{\tilde{\mathcal{H}}_k}$ auf $V \times V$ gleich und deren Einschränkung $\langle \cdot, \cdot \rangle_V := \langle \cdot, \cdot \rangle_{\mathcal{H}_k}|_{V \times V} = \langle \cdot, \cdot \rangle_{\tilde{\mathcal{H}}_k}|_{V \times V}$ ist das Skalarprodukt auf V bzw. die dadurch induzierte Norm/Metrik von V ist diejenige bzgl. der $\mathcal{H}_k = \overline{V}$ und $\tilde{\mathcal{H}}_k = \overline{V}$ gemäß Prop. 3.2.4 abgeschlossen/vervollständigt (Def.+Satz A.1.19) sind (man beachte: $\mathcal{H}_k, \tilde{\mathcal{H}}_k$ sind als Hilberträume nicht nur abgeschlossen, sondern sogar vollständig). Gemäß Def.+Satz A.1.19 ist jede Vervollständigung aber bis auf eine isomorphe Isometrie (Def.+Satz A.1.18) identisch, was die Behauptung liefert. $\square$

Kor. 3.2.6 zeigt, dass die Verbindung zwischen einem RKHS und seinem reproduzierenden Kern sehr stark ist. Das folgenden Kapitel 4 wird beide endgültig miteinander verschweißen.



Hauptergebnis des Kapitels:

Jeder Reproducing Kernel Hilbert Space $\mathcal{H}_k$ besitzt einen eindeutigen Reproduzierenden Kern, mit dem sich Norm und Skalarprodukt in $\mathcal{H}_k$ berechnen lassen, ohne dass $\mathcal{H}_k$ dabei explizit bekannt sein muss.

4

Der Satz von Moore(-Aronszajn)

In [Kapitel 3](#) wurde der **Reproducing Kernel Hilbert Space (RKHS)** und sein zugehöriger **reproduzierender Kern (RK)** eingeführt. [Kapitel 4](#), dessen Grundlage erneut [\[27, Chapter 2\]](#) ist, befasst sich weiter mit dem RK, erarbeitet Eigenschaften und schließt mit einem erstaunlichen Resultat, dem **Satz 4.2.1 von Moore(-Aronszajn)**. Der Beginn bildet das folgende Ergebnis für RK.

4.1 Positiv semidefinite Funktionen

Proposition 4.1.1 (Positivitätseigenschaft der RK [\[27, Prop. 2.13\]](#)).

Sei $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ ein $\overline{RKHS}$ auf X (über $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$) mit $\overline{RK} k$. Dann gilt für jedes $n \in \mathbb{N}$ sowie jede (endliche) Folge $\{x_1, \dots, x_n\} \subseteq X$ mit beliebigen $\alpha_1, \dots, \alpha_n \in \mathbb{K}$:

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} k(x_i, x_j) \geq 0. \quad (4.1)$$

Beweis. (Aus [\[27, Prop. 2.13\]](#) entnommen und lediglich um ein paar ausführlichere Schritte ergänzt.) Sei für $n \in \mathbb{N}$ die Folge $\{x_1, \dots, x_n\} \subseteq X$ sowie $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ beliebig, aber fest. Dann folgt sogleich:

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} k(x_i, x_j) \stackrel{(3.9)}{=} \sum_{i,j=1}^n \alpha_i \overline{\alpha_j} \overline{\langle k(\cdot, x_i), k(\cdot, x_j) \rangle_{\mathcal{H}_k}} \stackrel[\text{A.8.1}]{\text{A.3.1}} \overline{\left\langle \sum_{i=1}^n \overline{\alpha_i} k(\cdot, x_i), \sum_{j=1}^n \overline{\alpha_j} k(\cdot, x_j) \right\rangle_{\mathcal{H}_k}} \quad (4.2)$$

$$\stackrel{\text{A.3.3}}{=} \overline{\left\| \sum_{i=1}^n \overline{\alpha_i} k(\cdot, x_i) \right\|_{\mathcal{H}_k}^2} \stackrel{\text{A.2.4}}{\geq} 0. \quad (4.3)$$

□

Diese Eigenschaft der **RK** erinnert dabei sehr an folgenden Begriff aus der Linearen Algebra:

Definition 4.1.2 (Positiv semidefinite Matrix [4, 16, Def. 4.5.5, Def 3.4.16]).

Sei $K := (k_{ij}) \in M_{n,n}(\mathbb{R})$ (bzw. $K \in M_{n,n}(\mathbb{C})$) für $n \in \mathbb{N}$ eine symmetrische (bzw. Hermite'sche) Matrix, d. h. $k_{ij} = \overline{k_{ji}}$ für alle $i, j \in \{1, \dots, n\}$. K heißt **positiv semidefinit**, wenn für alle $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ (bzw. $\alpha \in \mathbb{C}^n$) gilt:

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} k_{ij} \geq 0 \quad \Leftrightarrow \quad \alpha^T K \alpha \geq 0. \quad (4.4)$$

Die Ähnlichkeit der zwei Eigenschaften erklärt, weshalb der Begriff auf Funktionen erweitert ist.

Definition 4.1.3 (Positiv semidefinite Funktion [27, Def. 2.12]).

Sei X eine Menge und $f : X \times X \rightarrow \mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ eine Funktion in zwei Variablen. f heißt **positiv semidefinit**, wenn für alle $x, y \in X$ gilt, dass

$$f(x, y) = \overline{f(y, x)} \quad (4.5)$$

und für alle $n \in \mathbb{N}$, $\{x_1, \dots, x_n\} \subseteq X$ und $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ gilt, dass

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} f(x_i, x_j) \geq 0. \quad (4.6)$$

Bemerkung 4.1.4.

1. In [33, Def. 2.5] wird eine leicht andere, aber äquivalente Definition genutzt: Für eine endliche Folge $\{x_1, \dots, x_n\} \subseteq X$ ($n \in \mathbb{N}$) wird die **Gramsche Matrix** $F := (f(x_i, x_j)) \in M_{n,n}(\mathbb{K})$ definiert. Die Funktion ist genau dann **positiv semidefinit**, wenn jede Gramsche Matrix für f positiv semidefinit ist (d. h. $n \in \mathbb{N}$ und $\{x_1, \dots, x_n\} \subseteq X$ beliebig). Im Falle eines abzählbar unendlichen Grundraums X erlaubt diese Definition die Anschauung von f als (∞, ∞) -Matrix.
2. In der Literatur wird der Begriff **positiv semidefinit** auch unter (leicht) anderen Namen geführt (vgl. [27, 33, S. 17, S. 30]). In der englischsprachigen Literatur (z. B. [27, 31, 33]) ist der Begriff *positive definite* oder nur *positive* geläufig (definit im Gegensatz zu semidefinit heißt dann *strictly positive (definite)*). Im Kontext von RKHS nennt beispielsweise [27] positiv semidefinite Funktionen auch nur noch *kernel functions* und fordert von diesen aber implizit die positive Semidefinitheit.

Es folgt nun sogleich das

Korollar 4.1.5 ([27, Prop. 2.13]). *Jeder reproduzierende Kern ist positiv semidefinit.*

Beweis. Folgt aus Prop. 4.1.1 und Def. 4.1.3 in Verbindung mit (3.9). □

Das erstaunliche und zentrale Ergebnis dieses Kapitels ist nun, dass ebenfalls die Rückrichtung dieser Implikation gilt. Der **Satz 4.2.1 von Moore(-Aronszajn)** wird sich als mithin größte Daseinsberechtigung der RKHS im Kontext des Maschinellen Lernens herausstellen.

4.2 Satz von Moore(-Aronszajn)

Satz 4.2.1 (Satz von Moore(-Aronszajn) [2, 21, 27, Theorem 2.14]).

Sei X eine Menge und $k : X \times X \rightarrow \mathbb{K}$ eine positiv semidefinite Funktion. Dann existiert ein RKHS $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ auf X (über $\mathbb{K}$), sodass k der reproduzierende Kern von $\mathcal{H}_k$ ist.

Beweis. (Die Beweisideen und -struktur wurden [27, Theorem 2.14] entnommen und um etwas ausführlichere Erläuterungen erweitert. Lediglich im Unterpunkt 2 wurde eine abweichende Begründung gewählt. In Unterpunkt 1 wurde eine Beweisidee aus [33, S. 32 ff.] übernommen, die in [27, Theorem 2.14] nur angedeutet ist.)

Aufgrund der etwas aufwendigeren Struktur des Beweises sei zunächst eine Skizze vorgeschoben:

Beweisskizze:

1. Man finde ein geeignetes Skalarprodukt $\langle \cdot, \cdot \rangle_V$, sodass die lineare Hülle $\text{span}\{k(\cdot, x) : x \in X\} := V$ ein euklidischer/unitärer Vektorraum ist (Def. A.3.1):
Aus Prop. 3.2.4 ist bekannt, dass im Fall eines RKHS $\mathcal{H}_k$ die lineare Hülle der zugehörigen Kernfunktionen $\{k(\cdot, x) : x \in X\}$ dicht in $\mathcal{H}_k$ liegen muss. Dies motiviert genau diese als die Wahl für V , dem Prototyp-Vektorraum des zu findenden RKHS.
2. Man finde einen Hilbertraum $(\hat{\mathcal{H}}, \langle \cdot, \cdot \rangle_{\hat{\mathcal{H}}})$ als die Vervollständigung (Def.+Satz A.1.19) von $(V, \langle \cdot, \cdot \rangle_V)$, d. h. man finde eine Isometrie $\Phi : V \rightarrow \hat{\mathcal{H}}$ mit $\overline{\Phi(V)} = \hat{\mathcal{H}}$:
Jeder metrische Raum V (Def. A.1.1) lässt sich zwar zu einem vollständigen Raum $\hat{\mathcal{H}}$ (Def. A.1.17) vervollständigen, jedoch unter der Einschränkung, dass lediglich das Bild einer Isometrie von V in $\hat{\mathcal{H}}$ enthalten sein muss und ggf. nicht mehr V selbst. Prüft man Def. 3.1.1, so muss ein RKHS insbesondere Unterraum von $\mathcal{F}(X, \mathbb{K})$ sein, d. h. ein Raum bestehend aus Funktionen von X nach $\mathbb{K}$. V war als solcher konstruiert, doch bei dessen Vervollständigung $\hat{\mathcal{H}}$ ist demnach $V \subseteq \hat{\mathcal{H}}$ nicht mehr gewährleistet und insbesondere auch nicht $\hat{\mathcal{H}} \subseteq \mathcal{F}(X, \mathbb{K})$. Aus diesem Grunde fährt man fort mit:
3. Man finde einen Hilbertraum $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ mit $\mathcal{H}_k \subseteq \mathcal{F}(X, \mathbb{K})$, sodass $\mathcal{H}_k$ isometrisch-isomorph zu $\hat{\mathcal{H}}$ ist und $V \subseteq \mathcal{H}_k$ gilt:
Mit dem kleinen Umweg über den Hilbertraum $\hat{\mathcal{H}}$ hat man den euklidischen/unitären Vektorraum $(V, \langle \cdot, \cdot \rangle_V)$ nun echt zu einem Hilbertraum $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ vervollständigt und die erste notwendige Bedingung eines RKHS aus Def. 3.1.1 ist erfüllt.
4. Man zeige, dass für alle $x \in X$ das Auswertungsfunktional $E_x : \mathcal{H}_k \rightarrow \mathbb{K}$ an der Stelle x beschränkt ist:
Damit ist vollständig gezeigt, dass es sich bei dem konstruierten Hilbertraum $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ tatsächlich um einen RKHS handelt.
5. Man zeige, dass k der reproduzierende Kern von $\mathcal{H}_k$ ist.

Sind diese 5 Unterpunkte nachgewiesen, so sind alle Voraussetzungen von Def. 3.1.1 und Def. 3.2.1 erfüllt und die Behauptung folgt.

Schritt 1: Man finde ein geeignetes Skalarprodukt $\langle \cdot, \cdot \rangle_V$, sodass die lineare Hülle $\text{span}\{k(\cdot, x) : x \in X\} := V$ ein euklidischer/unitärer Vektorraum ist (Def. A.3.1):

Sei die Abbildung $\langle \cdot, \cdot \rangle_V : V \times V \rightarrow \mathbb{K}$ definiert durch

$$\langle v, w \rangle_V := \sum_{i=1}^n \sum_{j=1}^m \alpha_i \overline{\beta_j k(x_i, y_j)}, \quad (4.7)$$

wobei $v, w \in V$ die jeweiligen Darstellungen $v = \sum_{i=1}^n \alpha_i k(\cdot, x_i)$ bzw. $w = \sum_{j=1}^m \beta_j k(\cdot, y_j) \in V$ besitzen, mit $m, n \in \mathbb{N}$, $\alpha_i, \beta_j \in \mathbb{K}$ und $x_i, y_j \in X$ für $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m\}$.

Da die Darstellungen von v und w nicht eindeutig sein müssen, ist als erstes zu prüfen, ob diese Abbildung überhaupt wohldefiniert ist: Seien dazu $v = \sum_{i=1}^{\tilde{n}} \tilde{\alpha}_i k(\cdot, \tilde{x}_i)$ bzw. $w = \sum_{j=1}^{\tilde{m}} \tilde{\beta}_j k(\cdot, \tilde{y}_j) \in V$ zwei weitere Darstellungen von $v, w \in V$, mit $\tilde{m}, \tilde{n} \in \mathbb{N}$, $\tilde{\alpha}_i, \tilde{\beta}_j \in \mathbb{K}$ und $\tilde{x}_i, \tilde{y}_j \in X$ für $i \in \{1, \dots, \tilde{n}\}$, $j \in \{1, \dots, \tilde{m}\}$. Dann gilt:

$$\begin{aligned} \langle v, w \rangle_V &\stackrel{(4.7)}{=} \sum_{i=1}^n \sum_{j=1}^m \alpha_i \overline{\beta_j k(x_i, y_j)} \stackrel{\text{A.8.1}}{=} \sum_{i=1}^n \overline{\alpha_i w(x_i)} = \sum_{i=1}^n \sum_{j=1}^{\tilde{m}} \alpha_i \overline{\tilde{\beta}_j k(x_i, \tilde{y}_j)} \\ &\stackrel{(4.5)}{=} \sum_{i=1}^n \sum_{j=1}^{\tilde{m}} \alpha_i \overline{\tilde{\beta}_j k(\tilde{y}_j, x_i)} = \sum_{j=1}^{\tilde{m}} \overline{\tilde{\beta}_j u(\tilde{y}_j)} = \sum_{i=1}^{\tilde{n}} \sum_{j=1}^{\tilde{m}} \tilde{\alpha}_i \overline{\tilde{\beta}_j k(\tilde{y}_j, \tilde{x}_i)} \stackrel{(4.5)}{=} \sum_{i=1}^{\tilde{n}} \sum_{j=1}^{\tilde{m}} \tilde{\alpha}_i \overline{\tilde{\beta}_j k(\tilde{x}_i, \tilde{y}_j)}. \end{aligned} \quad (4.8)$$

$$(4.9)$$

Somit ist $\langle \cdot, \cdot \rangle_V$ tatsächlich wohldefiniert und es verbleibt zum Abschluss, die Eigenschaften (S1)-(S5) aus Def. A.3.1 eines Skalarproduktes nachzuprüfen.

Seien dazu fortan $v, v', w \in V$ beliebig aber fest mit den zugehörigen Darstellungen als Linearkombinationen $v = \sum_{i=1}^n \alpha_i k(\cdot, x_i)$, $v' = \sum_{i'=1}^{n'} \alpha'_{i'} k(\cdot, x'_{i'})$ und $w = \sum_{j=1}^m \beta_j k(\cdot, y_j)$ mit $n, n', m \in \mathbb{N}$, $\alpha_i, \alpha'_{i'}, \beta_j \in \mathbb{K}$, $x_i, x'_{i'}, y_j \in X$ für $i \in \{1, \dots, n\}$, $i' \in \{1, \dots, n'\}$ und $j \in \{1, \dots, m\}$ sowie $\kappa \in \mathbb{K}$ beliebig aber fest.

(S1)+(S2): Da $V := \text{span}\{k(\cdot, x) : x \in X\}$ als lineare Hülle bzgl. der Addition und der skalaren Multiplikation abgeschlossen ist, gilt, dass $\kappa v + v'$ ebenfalls in V enthalten ist und ergo eine Darstellung der Form $\kappa v + v' = \sum_{i''=1}^{n''} \alpha''_{i''} k(\cdot, x''_{i''})$ besitzt für ein $n'' \in \mathbb{N}$ und gewisse $\alpha''_{i''} \in \mathbb{K}$, $x''_{i''} \in X$ mit $i'' \in \{1, \dots, n''\}$. Es gilt damit:

$$\langle \kappa v + v', w \rangle_V \stackrel{(4.7)}{=} \sum_{i''=1}^{n''} \sum_{j=1}^m \alpha''_{i''} \overline{\beta_j k(x''_{i''}, y_j)} \stackrel{(4.5)}{=} \sum_{i''=1}^{n''} \sum_{j=1}^m \alpha''_{i''} \overline{\beta_j k(y_j, x''_{i''})} = \sum_{j=1}^m \overline{\beta_j (\kappa v + v')(y_j)} \quad (4.10)$$

$$\stackrel{\text{A.2.2}}{=} \sum_{j=1}^m \overline{\beta_j \left(\kappa \sum_{i=1}^n \alpha_i k(y_j, x_i) + \sum_{i'=1}^{n'} \alpha'_{i'} k(y_j, x'_{i'}) \right)} \quad (4.11)$$

$$\stackrel{(4.5)}{=} \kappa \sum_{i=1}^n \sum_{j=1}^m \alpha_i \overline{\beta_j k(x_i, y_j)} + \sum_{i'=1}^{n'} \sum_{j=1}^m \alpha'_{i'} \overline{\beta_j k(x'_{i'}, y_j)} \stackrel{(4.7)}{=} \kappa \langle v, w \rangle_V + \langle v', w \rangle_V. \quad (4.12)$$

(S1)+(S2) folgen nun, wenn man berücksichtigt, dass die Fälle $\kappa = 1 \in \mathbb{K}$ und $v' = 0 \in V$ mit behandelt wurden.

(S3): Es gilt mit der Symmetrie von k :

$$\langle v, w \rangle_V \stackrel{(4.7)}{=} \sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j \overline{k(x_i, y_j)} \stackrel{(4.5)}{=} \sum_{i=1}^n \sum_{j=1}^m \beta_j \overline{\alpha_i k(y_j, x_i)} \stackrel{(4.7)}{=} \overline{\langle w, v \rangle_V}. \quad (4.13)$$

(S4): Es gilt mit der positiven Semidefinitheit von k :

$$\langle v, v \rangle_V \stackrel{(4.7)}{=} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \overline{k(x_i, x_j)} \stackrel{(4.5)}{=} \sum_{j=1}^n \sum_{i=1}^n \overline{\alpha_j} \alpha_i k(x_j, x_i) \stackrel{(4.6)}{\geq} 0. \quad (4.14)$$

(S5): Zuerst mache sich der Leser klar, dass im Beweis der Cauchy-Schwarzschen Ungleichung (siehe [Satz A.3.2](#)) lediglich die Eigenschaften (S1)-(S4) verwendet werden, nicht aber (S5). Das bedeutet, dass mit dem bis hierhin Bewiesenen die Cauchy-Schwarzsche Ungleichung bereits gilt:

$$|\langle v, w \rangle_V|^2 \leq \langle v, v \rangle_V \langle w, w \rangle_V \quad \text{für alle } v, w \in V. \quad (4.15)$$

Damit gilt dann für alle $x \in X$:

$$|v(x)|^2 = \left| \sum_{i=1}^n \alpha_i k(x, x_i) \right|^2 \stackrel{(4.5)}{=} \left| \sum_{i=1}^n \alpha_i \overline{k(x_i, x)} \right|^2 \stackrel{(4.7)}{=} |\langle v, k(\cdot, x) \rangle_V|^2 \stackrel{(4.15)}{\leq} \langle v, v \rangle_V \langle k(\cdot, x), k(\cdot, x) \rangle_V. \quad (4.16)$$

Es folgt somit für alle $x \in X$ aus $\langle v, v \rangle_V = 0$, dass $v(x) = 0$ ist und damit, dass $v = 0$ ist. Umgekehrt gilt offensichtlich auch für $v = 0$

$$\langle v, v \rangle_V \stackrel{(4.7)}{=} \sum_{i=1}^n \sum_{j=1}^n 0 \cdot \overline{0} \cdot \overline{k(x_i, x_j)} = 0, \quad (4.17)$$

(man beachte die gezeigte Wohldefiniertheit von $\langle \cdot, \cdot \rangle_V$) und es ist gezeigt, dass $v = 0 \Leftrightarrow \langle v, v \rangle_V = 0$.

Damit sind alle Bedingungen aus [Def. A.3.1](#) erfüllt, d. h. es wurde bewiesen, dass $\langle \cdot, \cdot \rangle_V$ ein Skalarprodukt auf V definiert und $(V, \langle \cdot, \cdot \rangle_V)$ damit ein euklidischer bzw. unitärer Vektorraum ist.

Schritt 2: Man finde einen Hilbertraum $(\widehat{\mathcal{H}}, \langle \cdot, \cdot \rangle_{\widehat{\mathcal{H}}})$ als die Vervollständigung von $(V, \langle \cdot, \cdot \rangle_V)$, d. h. man finde eine Isometrie $\Phi : V \rightarrow \widehat{\mathcal{H}}$ ([Def.+Satz A.1.18](#)) mit $\Phi(V) = \widehat{\mathcal{H}}$:

Dies ist gerade eine Anwendung von [Satz B.5](#) in Verbindung mit [Def.+Satz A.1.19](#).

Schritt 3: Man finde einen Hilbertraum $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ mit $\mathcal{H}_k \subseteq \mathcal{F}(X, \mathbb{K})$, sodass $\mathcal{H}_k$ isometrisch-isomorph zu $\widehat{\mathcal{H}}$ ist und $V \subseteq \mathcal{H}_k$ gilt:

In Schritt 1 wurde ein als Prototyp eines RKHS geeigneter Funktionenunterraum von $\mathcal{F}(X, \mathbb{K})$ zu einem euklidischen/unitären Vektorraum V befördert, welcher in Schritt 2 vervollständigt wurde. Leider ist über die Vervollständigung $\widehat{\mathcal{H}}$ aber nicht viel bekannt, insbesondere nicht, ob die Elemente darin überhaupt noch in $\mathcal{F}(X, \mathbb{K})$ liegen. Man weiß lediglich, dass eine Isometrie Φ existiert, sodass $\Phi(V) = \widehat{\mathcal{H}}$ gilt. Diese sowie die zu Hilfe eilende reproduzierende Eigenschaft von k sind aber genug, um aus $\widehat{\mathcal{H}}$ einen Hilbertraum $\mathcal{H}_k \subseteq \mathcal{F}(X, \mathbb{K})$ zu (re-)konstruieren:

Kor. B.6 liefert nämlich, dass die Isometrie Φ nicht nur definitionsgemäß den Wert der Norm von Elementen in V für deren Φ -Bilder in $\hat{\mathcal{H}}$ erhält, sondern ebenfalls den Wert des Skalarproduktes: Es gilt für alle $v, w \in V$ und deren Φ -Bilder $\hat{v} := \Phi(v), \hat{w} := \Phi(w) \in \hat{\mathcal{H}}$, dass

$$\langle \hat{v}, \hat{w} \rangle_{\hat{\mathcal{H}}} \stackrel{\text{B.6}}{=} \langle v, w \rangle_V. \quad (4.18)$$

Diese Tatsache ermöglicht nun die Betrachtung des Operators (via Definition: $\hat{k}(x, \cdot) := \Phi(k(\cdot, x)) \in \hat{\mathcal{H}}$ für $x \in X$)

$$L : \hat{\mathcal{H}} \rightarrow \mathcal{F}(X, \mathbb{K}), \quad (L\hat{h})(x) := \langle \hat{h}, \hat{k}(x, \cdot) \rangle_{\hat{\mathcal{H}}} \quad \text{für } \hat{h} \in \hat{\mathcal{H}}, x \in X. \quad (4.19)$$

Für jedes $v = \sum_{i=1}^n \alpha_i k(\cdot, x_i) \in V := \text{span}\{k(\cdot, x) : x \in X\}$ mit $n \in \mathbb{N}$, $\alpha_i \in \mathbb{K}$, $x_i \in X$, $i \in \{1, \dots, n\}$ und $\hat{v} := \Phi(v)$ gilt dann für alle $x \in X$:

$$(L\hat{v})(x) \stackrel{(4.19)}{=} \langle \hat{v}, \hat{k}(x, \cdot) \rangle_{\hat{\mathcal{H}}} \stackrel{(4.18)}{=} \langle v, k(\cdot, x) \rangle_V \stackrel{(4.7)}{=} \sum_{i=1}^n \sum_{j=1}^1 \alpha_i \overline{1k(x_i, x)} \stackrel{4.5}{=} \sum_{i=1}^n \alpha_i k(x, x_i) = v(x). \quad (4.20)$$

Es folgt somit $v \in L(\hat{\mathcal{H}})$ für alle $v \in V$, d. h. $V \subseteq L(\hat{\mathcal{H}})$.

Nun ist L wegen der Linearität des Skalarproduktes $\langle \cdot, \cdot \rangle_{\hat{\mathcal{H}}}$ sowie der punktweisen Addition und skalaren Multiplikation in $\mathcal{F}(X, \mathbb{K})$ ebenfalls linear mittels

$$(L(\alpha\hat{h} + \beta\hat{g}))(x) \stackrel{(4.19)}{=} \langle \alpha\hat{h} + \beta\hat{g}, \hat{k}(\cdot, x) \rangle_{\hat{\mathcal{H}}} \stackrel{\text{A.3.1}}{=} \alpha \langle \hat{h}, \hat{k}(\cdot, x) \rangle_{\hat{\mathcal{H}}} + \beta \langle \hat{g}, \hat{k}(\cdot, x) \rangle_{\hat{\mathcal{H}}} \quad (4.21)$$

$$\stackrel{(4.19)}{=} (\alpha L\hat{h} + \beta L\hat{g})(x) \quad \text{für alle } \hat{h}, \hat{g} \in \hat{\mathcal{H}}, \alpha, \beta \in \mathbb{K}, x \in X. \quad (4.22)$$

Da L somit gemäß **Def. A.4.1** ein wohldefinierter linearer Operator ist, ergibt sich folglich die sinnvolle Wahl des gesuchten RKHS $\mathcal{H}_k := L(\hat{\mathcal{H}}) \subseteq \mathcal{F}(X, \mathbb{K})$ durch das Bild von L , denn mit **Def.+Satz A.4.3** ist $\mathcal{H}_k$ als solches ein Unterraum von $\mathcal{F}(X, \mathbb{K})$ und, wie eben gesehen, gilt $V \subseteq L(\hat{\mathcal{H}}) =: \mathcal{H}_k$.

Nun betrachte man noch den Kern von L : Dazu mache man sich klar, dass wegen der Definition von $V := \text{span}\{k(\cdot, x) : x \in X\}$ und **Kor. B.3** insbesondere $D := \text{span}\{\hat{k}(\cdot, x) : x \in X\}$ dicht in $\hat{\mathcal{H}}$ liegt. Zu jedem $\hat{g} \in \hat{\mathcal{H}}$ gibt es somit eine Folge $(\hat{g}_n)_{n \in \mathbb{N}} \subset D$, sodass $\hat{g}_n \rightarrow \hat{g}$ in $\hat{\mathcal{H}}$. Es gilt dann

$$\hat{h} \in \ker L \stackrel{\text{A.4.3}}{\Rightarrow} L\hat{h} = 0. \quad (4.23)$$

$$\stackrel{(4.19)}{\Rightarrow} \langle \hat{h}, \hat{k}(\cdot, x) \rangle_{\hat{\mathcal{H}}} = 0 \quad \text{für alle } x \in X. \quad (4.24)$$

$$\stackrel{\text{A.2.3}}{\Rightarrow} \langle \hat{h}, d \rangle_{\hat{\mathcal{H}}} = 0 \quad \text{für alle } d \in D. \quad (4.25)$$

$$\stackrel{\text{A.1.7}}{\Rightarrow} \lim_{n \rightarrow \infty} \langle \hat{h}, \hat{g}_n \rangle_{\hat{\mathcal{H}}} = 0 \quad \text{für alle } \hat{g} \in \hat{\mathcal{H}}, (\hat{g}_n)_{n \in \mathbb{N}} \subset D \text{ mit } \hat{g}_n \rightarrow \hat{g}. \quad (4.26)$$

$$\stackrel{\text{A.3.6}}{\Rightarrow} \langle \hat{h}, \hat{g} \rangle_{\hat{\mathcal{H}}} = 0 \quad \text{für alle } \hat{g} \in \hat{\mathcal{H}}. \quad (4.27)$$

$$\stackrel{\text{A.3.1}}{\Rightarrow} \hat{h} = 0. \quad (4.28)$$

Da umgekehrt (z. B. mit **Def.+Satz A.4.3**) trivialerweise gilt, dass $0 \in \ker L$ ist, folgt somit $\ker L = \{0\}$. Ebenfalls gemäß **Def.+Satz A.4.3** ist mit L damit eine Bijektion zwischen $\hat{\mathcal{H}}$ und $\mathcal{H}_k$ gefunden, was insbesondere bedeutet, dass für jedes $h \in \mathcal{H}_k$ ein eindeutiges $\hat{h} \in \hat{\mathcal{H}}$ existiert, sodass $h = L\hat{h}$.

Dies ermöglicht nun die wohldefinierte Definition der Abbildung

$$\langle \cdot, \cdot \rangle_{\mathcal{H}_k} : \mathcal{H}_k \times \mathcal{H}_k \rightarrow \mathbb{K}, \quad \langle h, g \rangle_{\mathcal{H}_k} := \langle \hat{h}, \hat{g} \rangle_{\hat{\mathcal{H}}} \quad \text{für } h, g \in \mathcal{H}_k, \quad (4.29)$$

wobei $\hat{h}, \hat{g}$ die zu $h, g \in \mathcal{H}_k$ eindeutigen Elemente in $\hat{\mathcal{H}}$ sind, sodass $h = L\hat{h}$ und $g = L\hat{g}$ gelten. Da aufgrund von (4.29) für die Abbildung $\langle \cdot, \cdot \rangle_{\mathcal{H}_k}$ die Eigenschaften (S1)-(S5) in der Def. A.3.1 eines Skalarproduktes trivialerweise 1-zu-1 aus den Eigenschaften (S1)-(S5) des Skalarproduktes $\langle \cdot, \cdot \rangle_{\hat{\mathcal{H}}}$ folgen, hat man hiermit ein Skalarprodukt auf $\mathcal{H}_k$ gefunden – in anderen Worten: $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ ist ein euklidischer bzw. unitärer Vektorraum.

Mit dieser Wahl des Skalarproduktes folgt nun insbesondere auch, dass für alle $\hat{h} \in \hat{\mathcal{H}}$ gilt:

$$\|L\hat{h}\|_{\mathcal{H}_k} \stackrel{\text{A.3.3}}{=} \sqrt{\langle L\hat{h}, L\hat{h} \rangle_{\mathcal{H}_k}} \stackrel{(4.29)}{=} \sqrt{\langle \hat{h}, \hat{h} \rangle_{\hat{\mathcal{H}}}} \stackrel{\text{A.3.3}}{=} \|\hat{h}\|_{\hat{\mathcal{H}}}. \quad (4.30)$$

Aus Def.+Satz A.1.18 folgt damit, dass L (sogar) eine bijektive Isometrie zwischen $(\hat{\mathcal{H}}, \langle \cdot, \cdot \rangle_{\hat{\mathcal{H}}})$ und $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ bildet und zudem, dass ersterer ebenso ein Hilbertraum ist, denn letzterer ist als solcher vollständig. Beide Hilberträume sind also isometrisch-isomorph zueinander.

Mit ziemlich großem Aufwand ist endlich mit dem Hilbertraum $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ eine Vervollständigung von V gefunden, in der $V \subseteq \mathcal{H}_k$ echt enthalten ist und für die gilt: $\mathcal{H}_k \subseteq \mathcal{F}(X, \mathbb{K})$. Die ersten beiden Bedingungen in der Def. 3.1.1 eines RKHS sind erfüllt.

Bevor es mit Schritt 4 weiter geht, merke man sich noch folgendes hilfreiche Ergebnis mit den in Schritt 3 eingeführten Bezeichnungen: Sei zu $h \in \mathcal{H}_k$ weiterhin $\hat{h} \in \hat{\mathcal{H}}$ das zugehörige (eindeutige) Element, sodass $h = L\hat{h}$ gilt. Dann folgt für alle $h \in \mathcal{H}_k$ und $x \in X$:

$$h(x) = (L\hat{h})(x) \stackrel{(4.19)}{=} \langle \hat{h}, \hat{k}(\cdot, x) \rangle_{\hat{\mathcal{H}}} \stackrel{(4.29)}{=} \langle h, k(\cdot, x) \rangle_{\mathcal{H}_k}. \quad (4.31)$$

Schritt 4: Man zeige, dass für alle $x \in X$ das Auswertungsfunktional $E_x : \mathcal{H}_k \rightarrow \mathbb{K}$ an der Stelle x beschränkt ist:

Sei $x \in X$. Für das Auswertungsfunktional an der Stelle x gilt dann $(E_x(h) := h(x))$ für $h \in \mathcal{H}_k$:

$$|E_x(h)| \stackrel{(3.1)}{=} |h(x)| \stackrel{(4.31)}{=} |\langle h, k(\cdot, x) \rangle_{\mathcal{H}_k}| \stackrel{\text{A.3.2}}{\leq} \|k(\cdot, x)\|_{\mathcal{H}_k} \|h\|_{\mathcal{H}_k}. \quad (4.32)$$

Da bereits $V \subseteq \mathcal{H}_k$ bewiesen wurde, ist insbesondere $\|k(\cdot, x)\|_{\mathcal{H}_k}$ endlich, d. h. das Auswertungsfunktional E_x ist gemäß Def.+Satz A.4.2 beschränkt und weil $x \in X$ beliebig war, folgt die Behauptung.

Es wurden damit insgesamt alle Bedingungen aus Def. 3.1.1 geprüft und es ist nachgewiesen, dass der in den Schritten 1-3 konstruierte Hilbertraum $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ tatsächlich ein RKHS ist.

Schritt 5: Man zeige, dass k der reproduzierende Kern von $\mathcal{H}_k$ ist:

Aus (4.31) liest man sofort die reproduzierende Eigenschaft von k ab – für $h \in \mathcal{H}_k$ und $x \in X$ gelten:

$$h(x) \stackrel{(4.31)}{=} \langle h, k(\cdot, x) \rangle_{\mathcal{H}_k}. \quad (4.33)$$

Kor. 3.2.2 besagt nun gerade, dass k damit der reproduzierende Kern des RKHS $\mathcal{H}_k$ ist.

Schließlich sind nun alle Behauptungen bewiesen: Es wurde mit $\mathcal{H}_k$ ein RKHS konstruiert, zu dem k der reproduzierende Kern ist.

□

Der auf den ersten Blick vielleicht unscheinbar aussehende **Satz 4.2.1 von Moore** liefert nun folgendes wichtiges Korollar:

Hauptergebnis des Kapitels:



Korollar 4.2.2 (Positiv semidefinite Funktionen und reproduzierende Kerne [2, 21, 27, Theorem 2.14]).

Eine Funktion $k : X \times X \rightarrow \mathbb{K}$ ist genau dann ein reproduzierender Kern, wenn sie positiv semidefinit ist. In diesem Fall besitzt sie einen eindeutigen Reproducing Kernel Hilbert Space (RKHS).

Beweis. Dies besagen gerade Satz 4.2.1, Kor. 4.1.5 und Kor. 3.2.6 gemeinsam. $\square$

5

Der Kernel-Trick

In [Kapitel 3](#) und [Kapitel 4](#) wurden die Begriffe **RKHS** und **reproduzierender Kern** eingeführt und erkannt, dass unter gewissen Positivitätseigenschaften (siehe [Def. 4.1.3](#) / [Kor. 4.2.2](#)) jeder RK seinen eigenen RKHS „rekonstruiert“. Was diese sehr abstrakten Begriffe und Erkenntnisse mit konkreten Aufgabenstellungen im Maschinellen Lernen zu tun haben, soll dieser Abschnitt beleuchten, dessen Grundlage nun [[33](#), Chapter 2] ist.

Es sei kurz an die Motivation der RKHS aus [Kapitel 2](#) erinnert: Im Maschinellen Lernen werden Problemstellungen auf einem Rohdatenraum X betrachtet, wobei lediglich bekannt ist, dass X eine Menge ist. Für die Entwicklung von Algorithmen wünscht man sich hingegen eine Hilbertraumstruktur, um beispielsweise geometrische Lösungsmethoden anwenden zu können. Die RKHS mit ihren jeweiligen RK ermöglichen nun genau dies.



Beispiel 5.1.1 (Reproduzierender Kern im Klassifikationsproblem).

In [Kapitel 2](#) wurde gemutmaßt, dass sich Klassifikationsaufgaben in Hilberträumen ggf. mittels Hyperebenen lösen lassen. In einem reellen RKHS $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ beschreiben z. B. die Mengen $H_{w,\alpha} := \{x \in \mathcal{H}_k : \langle x, w \rangle_{\mathcal{H}_k} = \alpha\}$ mithilfe des Skalarproduktes alle darin existierenden Hyperebenen, wobei $H_{w,\alpha}$ für unterschiedliche $w \in \mathcal{H}_k$, $\alpha \in \mathbb{R}$ jeweils eine Hyperebene beschreibt [[25](#), Example 4.6.8]. Betrachtet man (3.8) in [Lemma 3.1.4](#), so ließe sich ein Klassifikations-Algorithmus f mittels einer geeigneten (punktetrennenden) Hyperebene $H_{w,\alpha}$ nur mithilfe der Kernfunktion k definieren:

$$f : X \rightarrow \{-1, 1\}, \quad f(x) = \begin{cases} 1, & \text{falls } k(x, \tilde{w}) < \alpha \\ -1, & \text{sonst,} \end{cases} \quad (5.1)$$

wobei $\tilde{w} \in X$ dergleichen ist, dass $w = k(\tilde{w}, \cdot) \in \mathcal{H}_k$ (vgl. [Lemma 3.1.4](#)).

Nun ermöglichen insbesondere die Erkenntnisse aus [Kapitel 4](#) in diesem Kontext eine weitere Verbesserung, welche eine der Hauptdaseinsberechtigungen von RKHS in Problemstellungen des Maschinellen Lernens liefert:

Bemerkung 5.1.2 (Kernel-Trick [33, Remark 2.8]).

Ist ein Algorithmus gegeben, welcher bezüglich eines reproduzierenden Kerns k formuliert ist, so lässt sich ein alternativer Algorithmus konstruieren, indem man k durch einen anderen reproduzierenden Kern $\tilde{k}$ ersetzt.

Da k als reproduzierender Kern seinen zugehörigen RKHS $\mathcal{H}_k$ reproduziert, tut dies gemäß [Satz 4.2.1 von Moore](#) auch jeder andere reproduzierende Kern $\tilde{k}$ mit dessen zugehörigem RKHS $\mathcal{H}_{\tilde{k}}$. Ein auf $\mathcal{H}_k$ operierender Algorithmus kann somit durch Tauschen von k und $\tilde{k}$ mittels des **Kernel-Tricks** zu einem anderen/neuen Algorithmus transformiert werden, welcher auf $\mathcal{H}_{\tilde{k}}$ operiert. Anstelle den „richtigen“ Hilbertraum zu suchen, ist eine alternative Lösungsstrategie nun also, den „richtigen“ reproduzierenden Kern k für das konkrete Problem zu finden.

Beispiel 5.1.3 (Kernel-Trick im Klassifikationsproblem [33, S. 32]).

Auf den Algorithmus f aus [Bsp. 5.1.1](#) kann nun der **Kernel-Trick** angewandt werden, da er (nur) von der Kernfunktion k abhängt. Man vergleiche dazu links den RKHS $\mathcal{H}_k$ mit RK $k(x, y) = \langle x, y \rangle$ sowie rechts den RKHS $\mathcal{H}_{\tilde{k}}$ mit RK $\tilde{k}(x, y) = \langle x, y \rangle^2$ für $X = \mathbb{R}^2$.

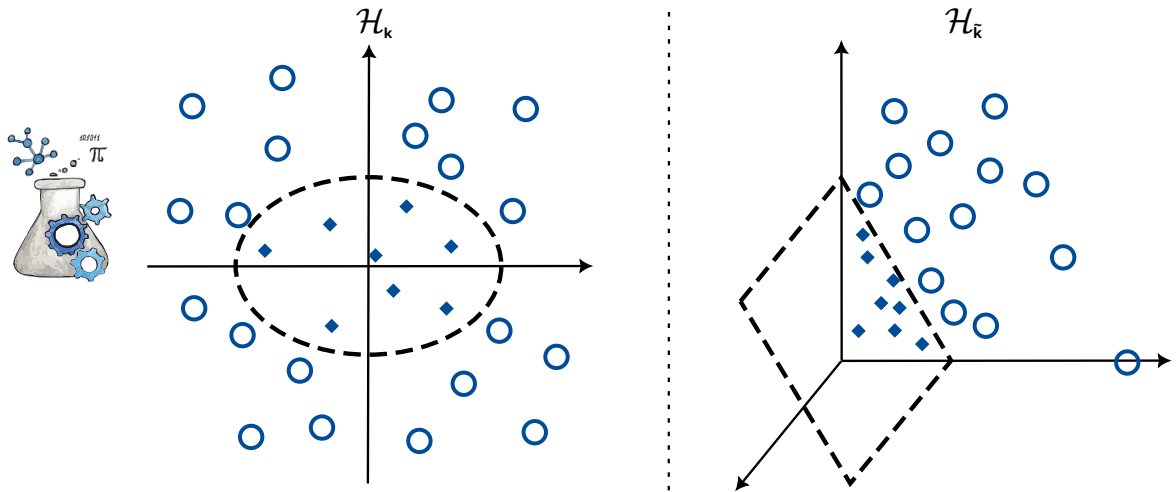


Abbildung 5.1: Der Kernel-Trick

Mithilfe des **Kernel-Tricks** können Entwickler im Maschinellen Lernen somit einen Algorithmus in einem beliebigen, ggf. unbekannten oder für die Problemstellung als ideal angenommenen, theoretischen Hilbertraum erfinden. Sie müssen lediglich darauf achten, dass dieser nur im Sinne eines reproduzierenden Kerns ausgedrückt werden kann, d.h. zum Beispiel lediglich vom Skalarprodukt oder der Norm im Hilbertraum abhängt und nicht etwa von weiteren kern-unabhängigen Eigenschaften. Dann lässt sich ein solcher theoretischer Algorithmus durch den Kernel-Trick vervollständigen,

indem ein geeigneter reproduzierender Kern mittels empirischer Tests dafür gefunden wird (vgl. [33, S. 34 ff.]).

Findet man auf diese Art und Weise einen reproduzierenden Kern, mit welchem der Algorithmus zuverlässig gute Ergebnisse liefert, so hat man die maschinelle Lernaufgabe gelöst, ohne dabei den verwendeten Hilbertraum jemals explizit gekannt zu haben (vgl. [33, Chapter 1]): Weil der reproduzierende Kern k direkt auf der Rohdatenmenge X wirkt, ist die genaue Kenntnis des Hilbertraums in diesem Fall aber eben auch nicht nötig. Es wird lediglich darauf vertraut, dass die Kernfunktion einen geeigneten bzw. den korrekten Hilbertraum „im Hintergrund“ reproduziert, wenn die Ergebnisse des Algorithmus stabil bleiben.

Das in der [Einleitung](#) anfangs erwähnte beliebte empirische Vorgehen im Maschinellen Lernen zur Lösungsfindung erhält somit im Bereich der Kern-Methoden über den [Satz 4.2.1 von Moore](#) und aus diesem folgend, dem [Kernel-Trick](#) seine mathematisch-theoretische Grundlage und Legitimation.

Darüber hinausgehend ermöglicht das beschriebene Vorgehen noch eine Erweiterung: Vertraut man im Falle stabiler Ergebnisse darauf, dass ein für die Problemstellung geeigneter RKHS $\mathcal{H}_k$ gefunden wurde, so kann die Abbildung $\Phi : X \rightarrow \mathcal{H}_k$ via (3.10) aus der Kernfunktion zurückgewonnen und der RKHS darüber erforscht werden. Es ist damit möglich, Metriken, Normen und andere Eigenschaften auf der Rohdatenmenge X zu entdecken, die vielleicht auf den ersten Blick nicht ersichtlich waren. Durch das Finden eines geeigneten RKHS – und bei Vertrauen darauf, dass die stabilen Ergebnisse suggerieren, dass dieser auch „der richtige RKHS“ ist – lassen sich somit „rückwärts“ Hilbertraumeigenschaften auf der Rohdatenmenge X finden. Eine konkrete Anwendung dieses Prinzips ist die sog. **Principle Component Analysis (PCA)** (siehe [33, Chapter 14]).



Hauptergebnis des Kapitels:

Mithilfe des [Kernel-Tricks](#) ermöglichen [RKHS](#) eine alternative Lösungsmethode für Aufgabenstellungen im Maschinellen Lernen: Durch Ausprobieren verschiedener [reproduzierender Kerne](#) wird ein funktionierender Algorithmus gefunden, ohne dass die [Hilbertraumstruktur](#) der Rohdatenmenge X (im Vorhinein) bekannt ist.

6

Der Satz von Mercer

In [Kapitel 4](#) und [Kapitel 5](#) wurden mit dem [Satz 4.2.1 von Moore](#) und dem [Kernel-Trick](#) die Grundlage gelegt, um für eine beliebige Rohdatenmenge X mittels positiver Kernfunktionen Hilberträume zu finden, in welchen sich das maschinelle Lernproblem theoretisch besser behandeln lässt. Auf der Suche nach dem „passenden“ Hilbertraum war es nicht einmal nötig, diesen explizit zu kennen.

In einigen Anwendungen (z. B. in der **Principle Component Analysis (PCA)**, vgl. [\[33, Chapter 14\]](#), weitere Ausführungen im Folgenden) ist es aber nötig, die Struktur des gefundenen Hilbertraums zu kennen/erforschen zu können. Dabei stellt sich heraus, dass für viele Probleme der entstehende RKHS sehr hochdimensional wird (vgl. [Kor. 3.2.5](#)). Als Funktionenunterräume von $\mathcal{F}(X, \mathbb{K})$ sind RKHS oft sogar unendlich-dimensional, was diese unhandlich in der Praxis machen kann.

Um dieses Problem der Unendlich-Dimensionalität in den Griff zu bekommen, eilt der [Satz 6.5.1 von Mercer](#) zu Hilfe, das Hauptergebnis dieses Kapitels. Dessen Beweis bedarf einiger **spektral-theoretischer** Vorbereitungen sowie Vorüberlegungen zu **Wurzeln positiver Operatoren** und zu **2-summierenden Operatoren**. Die Grundlage dieses Kapitels ist [\[1, 7, 10, 18, 33\]](#).

6.1 Etwas Spektraltheorie

In Vorbereitung der Betrachtung von Wurzeln positiver Operatoren, sei zunächst ein Ergebnis aus der Spektraltheorie betrachtet.

Proposition 6.1.1 (Spectral Mapping Theorem [\[34, Prop. 3.1.9\]](#)).

Sei $\mathcal{A}$ eine Banachalgebra mit Einselement ([Def. A.5.2](#)), $a \in \mathcal{A}$ und $p = \sum_{i=0}^n c_i X^i \in \mathbb{C}[X]$ ein Polynom mit $n \in \mathbb{N}_0$, $c_i \in \mathbb{C}$ für $i \in \{1, \dots, n\}$. Dann gilt für das Spektrum ([Def. A.5.4](#)) von $p(a)$:

$$\sigma(p(a)) = p(\sigma(a)). \quad (6.1)$$

Beweis. (Der Beweis ist [34, Prop. 3.1.9] entnommen und lediglich ausformuliert.)

Zunächst sei angemerkt, dass $p(a)$ tatsächlich ein Element von $\mathcal{A}$ ist, weil $\mathcal{A}$ bzgl. der endlichen Summe und Multiplikation abgeschlossen ist. Es sei dann $\lambda \in \mathbb{C}$ beliebig, aber fest und o.B.d.A. angenommen, dass $c_n \neq 0$ und $n > 0$, denn für $p = c_0$ ist $p(a) - \lambda \text{Id} = (c_0 - \lambda) \text{Id}$ genau dann invertierbar (Def.+Satz A.5.3), wenn $\lambda \neq c_0$ gilt, d.h. $\sigma(p(a)) = \{c_0\} = p(\sigma(a))$ gemäß Satz A.5.5. Laut dem Fundamentalsatz der Algebra (Satz A.8.2) gibt es nun $\lambda_1, \dots, \lambda_n \in \mathbb{C}$, sodass

$$p - \lambda = c_n \prod_{i=1}^n (X - \lambda_i) \quad \text{und entsprechend} \quad p(a) - \lambda \text{Id} = c_n \prod_{i=1}^n (a - \lambda_i \text{Id}), \quad (6.2)$$

wobei das Produkt in (6.2) kommutativ ist. Aus Letzterem folgt, dass für jedes $i \in \{1, \dots, n\}$ ebenfalls $a - \lambda_i \text{Id}$ mit $b := p(a) - \lambda \text{Id}$ kommutiert, denn:

$$(a - \lambda_i \text{Id})b \stackrel{(6.2)}{=} (a - \lambda_i \text{Id})c_n \prod_{i=1}^n (a - \lambda_i \text{Id}) = c_n \prod_{i=1}^n (a - \lambda_i \text{Id})(a - \lambda_i \text{Id}) \stackrel{(6.2)}{=} b(a - \lambda_i \text{Id}). \quad (6.3)$$

Man bemerke nun, dass b genau dann invertierbar ist, wenn $a - \lambda_i \text{Id}$ invertierbar ist für alle $i \in \{1, \dots, n\}$: Sind nämlich alle $a - \lambda_i \text{Id}$ invertierbar, so ist mit der Gruppeneigenschaft der invertierbaren Elemente $\mathcal{G}(\mathcal{A})$ (Def.+Satz A.5.3) ebenfalls b invertierbar. Ist umgekehrt b invertierbar, dann gilt für jedes $i \in \{1, \dots, n\}$ mit der Kommutativität in (6.2) und (6.3):

$$(a - \lambda_i \text{Id}) \left(c_n \prod_{\substack{j=1 \\ j \neq i}}^n (a - \lambda_j \text{Id}) b^{-1} \right) \stackrel{(6.2)}{=} c_n \prod_{i=1}^n (a - \lambda_i \text{Id}) b^{-1} \stackrel{(6.2)}{=} b b^{-1} = \text{Id} \quad \text{und} \quad (6.4)$$

$$\left(c_n \prod_{\substack{j=1 \\ j \neq i}}^n (a - \lambda_j \text{Id}) b^{-1} \right) (a - \lambda_i \text{Id}) = c_n \prod_{\substack{j=1 \\ j \neq i}}^n (a - \lambda_j \text{Id}) b^{-1} (a - \lambda_i \text{Id}) b b^{-1} \quad (6.5)$$

$$\stackrel{(6.3)}{=} c_n \prod_{\substack{j=1 \\ j \neq i}}^n (a - \lambda_j \text{Id}) b^{-1} b (a - \lambda_i \text{Id}) b^{-1} \quad (6.6)$$

$$= c_n \prod_{\substack{j=1 \\ j \neq i}}^n (a - \lambda_j \text{Id}) (a - \lambda_i \text{Id}) b^{-1} \stackrel{(6.2)}{=} b b^{-1} = \text{Id}. \quad (6.7)$$

Folglich sind also alle $a - \lambda_i \text{Id}$ invertierbar, sofern $p(a) - \lambda \text{Id}$ es ist und die behauptete Äquivalenz der Invertierbarkeit ist gezeigt. Dies liefert aber gerade mit der Def. A.5.4 des Spektrums die Behauptung:

$$\lambda \in p(\sigma(a)) \Leftrightarrow p(\mu) - \lambda = 0 \text{ für ein } \mu \in \sigma(a) \quad (6.8)$$

$$\stackrel{(6.2)}{\Leftrightarrow} \lambda_i \in \sigma(a) \text{ für ein } i \in \{1, \dots, n\} \Leftrightarrow \lambda \in \sigma(p(a)). \quad (6.9)$$

□

6.2 Die Wurzel eines positiven Operators

Mithilfe dieses spektraltheoretischen Ergebnisses lässt sich nun die **Wurzel eines positiven Operators** finden. Letzterer ist folgendermaßen definiert.

Definition und Satz 6.2.1 (Positiver Operator [33, Theorem 2.10]).

Sei $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ ein Hilbertraum und $T \in \mathcal{L}(\mathcal{H})$ ein beschränkter Operator darauf. T heißt **positiv**, wenn

$$\langle Tx, x \rangle_{\mathcal{H}} \geq 0 \quad \text{für alle } x \in \mathcal{H}. \quad (6.10)$$

Jeder positive Operator besitzt rein nicht-negatives Punktspektrum (Def. A.4.15), d. h. für jeden Eigenwert $\lambda \in \mathbb{C}$ von T gilt: $0 \leq \lambda \leq \|T\|_{\mathcal{L}(\mathcal{H})}$.

Beweis.

Sei $Te = \lambda e$ für einen Eigenwert $\lambda \in \mathbb{C}$ von T mit zugehörigem Eigenvektor (Def. A.4.15) $e \in \mathcal{H}$. Dann gilt:

$$\lambda \|e\|_{\mathcal{H}}^2 \stackrel{\text{A.3.3}}{=} \lambda \langle e, e \rangle_{\mathcal{H}} \stackrel{\text{A.3.1}}{=} \langle \lambda e, e \rangle_{\mathcal{H}} \stackrel{\text{A.4.15}}{=} \langle Te, e \rangle_{\mathcal{H}} \stackrel{(6.10)}{\geq} 0. \quad (6.11)$$

Da e als Eigenvektor gemäß Def. A.4.15 nicht Null sein darf, gilt $\|e\|_{\mathcal{H}} > 0$, wodurch $\lambda \geq 0$ aus (6.11) und in Verbindung mit Def.+Satz A.5.6 die Behauptung folgt. $\square$

Eine besondere Eigenschaft jedes positiven, selbstadjungierten (Def. A.4.14) Operators ist, dass er eine **Wurzel** besitzt.

Proposition 6.2.2 (Wurzel eines positiven, selbstadjungierten Operators [29, 11, Theorem VI.9]).

Sei $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ ein Hilbertraum und $T \in \mathcal{L}(\mathcal{H})$ ein positiver, selbstadjungierter (beschränkter) Operator darauf. Dann existiert ein beschränkter, selbstadjungierter Operator $S \in \mathcal{L}(\mathcal{H})$ mit $S^2 = T$.

Beweis. (Der Beweis und seine zugehörigen -ideen sind [29, 11, Theorem VI.9] entnommen.)

Gemäß Def.+Satz 6.2.1 gilt für das Spektrum von T : $\sigma(T) \subset K := [0, \|T\|_{\mathcal{L}(\mathcal{H})}]$. Da K kompakt ist (Def. A.1.9) und $\sqrt{\cdot} \in C(K, \mathbb{R})$ gilt, folgt mit dem Satz A.2.18 von Stone-Weierstraß, dass es eine Folge $(p_n)_{n \in \mathbb{N}} \subset \mathbb{R}[X]$ von reellen Polynomen gibt, sodass $p_n \rightarrow \sqrt{\cdot}$ in $(C(K, \mathbb{R}), \|\cdot\|_{\infty})$ gilt. Insbesondere ist $(p_n)_{n \in \mathbb{N}}$ damit eine Cauchyfolge (Def.+Satz A.1.11) in $(C(K, \mathbb{R}), \|\cdot\|_{\infty})$.

Mithilfe von $(p_n)_{n \in \mathbb{N}}$ lässt sich die Operatorenfolge $(p_n(T))_{n \in \mathbb{N}} \subset \mathcal{L}(\mathcal{H})$ definieren, welche wohldefiniert ist, da die Banachalgebra $\mathcal{L}(\mathcal{H})$ bzgl. endlicher Summe und Multiplikation abgeschlossen ist. Zudem folgt aus der Selbstadjungiertheit von T , dass für ein beliebiges reelles Polynom $q := \sum_{i=0}^m c_i X^i \in \mathbb{R}[X]$ ebenfalls $q(T)$ selbstadjungiert ist, mit $m \in \mathbb{N}_0$, $c_i \in \mathbb{R}$ für $i \in 0, \dots, m$, denn für $x, y \in \mathcal{H}$ gilt:

$$\langle (q(T))x, y \rangle_{\mathcal{H}} \stackrel{\text{A.3.1}}{=} \sum_{i=0}^m c_i \langle T^i x, y \rangle_{\mathcal{H}} \stackrel{\text{A.4.14}}{=} \sum_{i=0}^m c_i \langle x, T^i y \rangle_{\mathcal{H}} \stackrel{\substack{\text{A.3.1} \\ \text{A.8.1}}}{=} \langle x, (q(T))y \rangle_{\mathcal{H}}. \quad (6.12)$$

Damit sind alle $p_n(T)$, sowie $(p_n - p_m)(T)$ selbstadjungiert und insbesondere normal (Def. A.5.7) für $n, m \in \mathbb{N}$. Es folgt für $\varepsilon > 0$ und zugehörigem $n_{\varepsilon} \in \mathbb{N}$, sodass $\|p_n - p_m\|_{\infty} \leq \varepsilon$ für $n, m \geq n_{\varepsilon}$:

$$\|p_n(T) - p_m(T)\|_{\mathcal{L}(\mathcal{H})} = \|(p_n - p_m)(T)\|_{\mathcal{L}(\mathcal{H})} \stackrel{\text{A.5.8}}{=} r((p_n - p_m)(T)) \quad (6.13)$$

$$\stackrel{\text{A.5.6}}{=} \sup \{|\lambda| : \lambda \in \sigma((p_n - p_m)(T))\} \quad (6.14)$$

$$\stackrel{\text{6.1.1}}{=} \sup \{|(p_n - p_m)(\lambda)| : \lambda \in \sigma(T)\} \quad (6.15)$$

$$\stackrel{\text{6.2.1}}{\leq} \sup_{\lambda \in K} |p_n(\lambda) - p_m(\lambda)| = \|p_n - p_m\|_{\infty} \leq \varepsilon. \quad (6.16)$$

Ergo ist $(p_n(T))_{n \in \mathbb{N}}$ eine Cauchyfolge in $\mathcal{L}(\mathcal{H})$ und da Letzterer gemäß Satz A.4.4 wegen der Vollständigkeit von $\mathcal{H}$ auch vollständig ist, existiert ein $S \in \mathcal{L}(\mathcal{H})$ als der Grenzwert von $(p_n(T))_{n \in \mathbb{N}}$.

S ist erneut selbstadjungiert, denn für $x, y \in \mathcal{H}$ ist Def. A.4.14 erfüllt:

$$\langle Sx, y \rangle_{\mathcal{H}} \stackrel{\text{A.3.6}}{=} \lim_{n \rightarrow \infty} \langle (p_n(T))x, y \rangle_{\mathcal{H}} \stackrel{\text{A.4.14}}{=} \lim_{n \rightarrow \infty} \langle x, (p_n(T))y \rangle_{\mathcal{H}} \stackrel{\text{A.3.6}}{=} \langle x, Sy \rangle_{\mathcal{H}}. \quad (6.17)$$

Schließlich bleibt zu zeigen, dass $S^2 = T$ gilt: Zunächst einmal ist wegen der Algebra-Eigenschaft (Def. A.5.1) von $\mathbb{R}[X]$ ebenfalls $q_n := p_n^2 - X \in \mathbb{R}[X]$ für alle $n \in \mathbb{N}$. Zudem ist wegen $p_n, \sqrt{\cdot} \in C(K, \mathbb{R})$ und der Kompaktheit von K die Norm $\|p_n + \sqrt{\cdot}\|_{\infty} \leq M$ beschränkt mit einem $M > 0$. Mit den obigen Überlegungen folgt damit auf gleiche Weise:

$$\|(p_n(T))^2 - T\|_{\mathcal{L}(\mathcal{H})} = \|q_n(T)\|_{\mathcal{L}(\mathcal{H})} \stackrel{\text{A.5.8}}{=} r(q_n(T)) \stackrel{\text{A.5.6}}{=} \sup \{|\lambda| : \lambda \in \sigma(q_n(T))\} \quad (6.18)$$

$$\stackrel{\text{6.1.1}}{=} \sup \{|q_n(\lambda)| : \lambda \in \sigma(T)\} \stackrel{\text{6.2.1}}{\leq} \sup_{\lambda \in K} |(p_n(\lambda))^2 - \lambda| \quad (6.19)$$

$$\leq \sup_{\lambda \in K} |p_n(\lambda) - \sqrt{\lambda}| |p_n(\lambda) + \sqrt{\lambda}| \leq M \|p_n - \sqrt{\cdot}\|_{\infty} \rightarrow 0 \quad \text{für } n \rightarrow \infty. \quad (6.20)$$

Damit ist auch die letzte Behauptung bewiesen:

$$S^2 = \left(\lim_{n \rightarrow \infty} p_n(T) \right)^2 = \lim_{n \rightarrow \infty} (p_n(T))^2 = T. \quad (6.21)$$

□

Bemerkung 6.2.3. Die Wurzel S in Prop. 6.2.2 ist sogar eindeutig und ebenfalls positiv z. B. gemäß [29, Theorem VI.9]. Da dieser Fakt im Weiteren aber nicht benötigt wird, sei aus Platzgründen auf die Teilbeweise verzichtet.

Schließlich benötigt der Beweis zu Satz 6.5.1 noch eine Aussage zu den Eigenvektoren der soeben gefundenen Wurzel, welche mit den nächsten beiden Korollaren erarbeitet wird.

Korollar 6.2.4 (Eigenräume der Wurzel eines positiven, selbstadjungierten Operators [29, 11, Theorem VI.9]).

Sei $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ ein Hilbertraum, $T \in \mathcal{L}(\mathcal{H})$ ein positiver, selbstadjungierter, (beschränkter) Operator darauf und $S \in \mathcal{L}(\mathcal{H})$ die selbstadjungierte Wurzel gemäß Prop. 6.2.2, d. h. $S^2 = T$. Dann gilt für jeden positiven Eigenwert $\lambda > 0$ von T (d. h. $\lambda \neq 0$, denn T ist positiv):

$$\ker(T - \lambda \text{Id}) = \ker(S - \sqrt{\lambda} \text{Id}) \oplus \ker(S + \sqrt{\lambda} \text{Id}). \quad (6.22)$$

Zudem ist die Summe orthogonal (Def. A.3.10).

Beweis. (Die Beweisideen sind [29, 11, Theorem VI.9] entnommen.)

Zunächst einmal ist die Summe auf der rechten Seite von (6.22) tatsächlich direkt, denn ist $x \in \ker(S - \sqrt{\lambda}\text{Id}) \cap \ker(S + \sqrt{\lambda}\text{Id})$, d. h. gilt $(S - \sqrt{\lambda})x = (S + \sqrt{\lambda})x = 0$, dann liefert die Subtraktion beider Gleichungen $2\sqrt{\lambda}x = 0$. Da aber $\lambda > 0$ angenommen war, folgt $x = 0$ und somit die Direktheit der Summe.

„ $\ker(T - \lambda\text{Id}) \subseteq \ker(S - \sqrt{\lambda}\text{Id}) \oplus \ker(S + \sqrt{\lambda}\text{Id})$ “:

Sei $0 \neq e \in \ker(T - \lambda\text{Id})$, d. h. e ist ein Eigenvektor zum Eigenwert λ von T . Man definiere nun

$$u := (S + \sqrt{\lambda}\text{Id})e \quad \text{und} \quad v := (S - \sqrt{\lambda}\text{Id})e. \quad (6.23)$$

Dann gelten $u \in \ker(S - \sqrt{\lambda}\text{Id})$ und $v \in \ker(S + \sqrt{\lambda}\text{Id})$, denn

$$(S - \sqrt{\lambda}\text{Id})u = (S - \sqrt{\lambda}\text{Id})(S + \sqrt{\lambda}\text{Id})e = (T - \lambda\text{Id})e = 0 \quad \text{und} \quad (6.24)$$

$$(S + \sqrt{\lambda}\text{Id})v = (S + \sqrt{\lambda}\text{Id})(S - \sqrt{\lambda}\text{Id})e = (T - \lambda\text{Id})e = 0. \quad (6.25)$$

Zudem gilt wegen $\lambda > 0$:

$$u - v = (S + \sqrt{\lambda}\text{Id})e - (S - \sqrt{\lambda}\text{Id})e = 2\sqrt{\lambda}e \Rightarrow e = \frac{1}{2\sqrt{\lambda}}(u - v). \quad (6.26)$$

Damit ist $e \in \ker(S - \sqrt{\lambda}\text{Id}) \oplus \ker(S + \sqrt{\lambda}\text{Id})$ gezeigt und da $e \in \ker(T - \lambda\text{Id})$ beliebig gewählt war, folgt $\ker(T - \lambda\text{Id}) \subseteq \ker(S - \sqrt{\lambda}\text{Id}) \oplus \ker(S + \sqrt{\lambda}\text{Id})$.

„ $\ker(S - \sqrt{\lambda}\text{Id}) \oplus \ker(S + \sqrt{\lambda}\text{Id}) \subseteq \ker(T - \lambda\text{Id})$ “:

Sei nun $e = u + v \in \ker(S - \sqrt{\lambda}\text{Id}) \oplus \ker(S + \sqrt{\lambda}\text{Id})$ mit $u \in \ker(S - \sqrt{\lambda}\text{Id})$ und $v \in \ker(S + \sqrt{\lambda}\text{Id})$. Dann gilt $Su = \sqrt{\lambda}u$ sowie $Sv = -\sqrt{\lambda}v$ und damit:

$$Tu = S^2u = S(\sqrt{\lambda}u) = \lambda u \quad \text{und} \quad Tv = S^2v = S(-\sqrt{\lambda}v) = \lambda v. \quad (6.27)$$

Ergo sind $u, v \in \ker(T - \lambda\text{Id})$ und da letzterer gemäß [Def.+Satz A.4.3](#) ein Vektorraum ist, folglich auch $e = u + v \in \ker(T - \lambda\text{Id})$. Da $e \in \ker(S - \sqrt{\lambda}\text{Id}) \oplus \ker(S + \sqrt{\lambda}\text{Id})$ beliebig gewählt war, folgt die behauptete Inklusion.

Es bleibt zu zeigen, dass die Summe in (6.22) orthogonal ist. Seien dazu $x \in \ker(S - \sqrt{\lambda}\text{Id})$ und $y \in \ker(S + \sqrt{\lambda}\text{Id})$. Dann gilt:

$$\sqrt{\lambda}\langle x, y \rangle_{\mathcal{H}} \stackrel{\text{A.3.1}}{=} \langle \sqrt{\lambda}x, y \rangle_{\mathcal{H}} \stackrel{\text{A.4.15}}{=} \langle Sx, y \rangle_{\mathcal{H}} \stackrel{\text{A.4.14}}{=} \langle x, Sy \rangle_{\mathcal{H}} \stackrel{\text{A.4.15}}{=} \langle x, -\sqrt{\lambda}y \rangle_{\mathcal{H}} = -\sqrt{\lambda}\langle x, y \rangle_{\mathcal{H}}. \quad (6.28)$$

Da aber $\lambda > 0$ vorausgesetzt war, folgt $\langle x, y \rangle_{\mathcal{H}} = 0 \Leftrightarrow x \perp y$ und wegen der beliebigen Wahl von x, y die Orthogonalität von $\ker(S - \sqrt{\lambda}\text{Id})$ und $\ker(S + \sqrt{\lambda}\text{Id})$. $\square$

Korollar 6.2.5 (Eigenvektoren der Wurzel eines positiven, selbstadjungierten Operators [29, 11, Theorem VI.9]).

Sei $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ ein Hilbertraum, $T \in \mathcal{L}(\mathcal{H})$ ein positiver, selbstadjungierter, (beschränkter) Operator darauf und $S \in \mathcal{L}(\mathcal{H})$ die selbstadjungierte Wurzel gemäß [Prop. 6.2.2](#), d. h. $S^2 = T$. Dann gilt für jeden positiven Eigenwert $\lambda > 0$ von T (d. h. $\lambda \neq 0$, denn T ist positiv):

Existiert für $n \in \mathbb{N}$ ein Orthonormalsystem ([Def. A.3.12](#)) $e_1, \dots, e_n \subset \mathcal{H}$ von zu λ zugehörigen Eigenvektoren für T , dann existiert ein Orthonormalsystem $\tilde{e}_1, \dots, \tilde{e}_n$ von Eigenvektoren für S , die jeweils entweder den Eigenwerten $\sqrt{\lambda}$ oder $-\sqrt{\lambda}$ zugeordnet sind.

Beweis. (Die Beweisideen sind [29, 11, Theorem VI.9] entnommen.)

Kor. 6.2.4 liefert direkt die Existenz der orthonormalen Eigenvektoren $\tilde{e}_1, \dots, \tilde{e}_n \in \ker(S - \sqrt{\lambda}\text{Id}) \cup \ker(S + \sqrt{\lambda}\text{Id})$. Diese sind dann entweder $\sqrt{\lambda}$ oder $-\sqrt{\lambda}$ zugeordnet. $\square$

Bemerkung 6.2.6. Für die Wurzel eines selbstadjungierten, positiven Operators existiert eine umfangreiche weitere Theorie. So folgt aus **Kor. 6.2.4** in Verbindung mit der unter **Bem. 6.2.3** behaupteten Positivität der Wurzel, dass die Eigenvektoren eines positiven, selbstadjungierten Operators zu Eigenwerten ungleich Null, mit denen seiner Wurzel übereinstimmen. Aus Platzgründen wurde in dieser Arbeit lediglich das minimal Notwendige für die aufgeführten Beweise erarbeitet.

6.3 2-summierende Operatoren

Neben der Wurzel eines positiven Operators benötigt der Beweis des **Satzes 6.5.1 von Mercer** noch den Begriff der **2-summierenden Operatoren**. Die allgemeinere Definition der **p - q -summierenden Operatoren** sei dafür hier gegeben.

Definition 6.3.1 (p - q -summierende Operatoren [18, Def. 1.d.4]).

Seien $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ Banachräume (**Def. A.2.8**), $T \in \mathcal{L}(X, Y)$ ein beschränkter Operator und $1 \leq p \leq q < \infty$. T heißt **p - q -summierend**, wenn ein $C > 0$ existiert, sodass für alle $n \in \mathbb{N}$ sowie $x_1, \dots, x_n \in X$ gilt:

$$\left(\sum_{i=1}^n \|Tx_i\|_Y^p \right)^{\frac{1}{p}} = C \sup_{\|x'\|_{X'} \leq 1} \left(\sum_{i=1}^n |\langle x', x_i \rangle_{X', X}|^q \right)^{\frac{1}{q}}. \quad (6.29)$$

Gilt $p = q$, so nennt man T simpel **p -summierend**.

Bemerkung 6.3.2.

Ihren Ursprung haben die p - q -summierenden Operatoren 1969 in [28, Def. 2.1.1]. Dort werden die sog. **absolutsummierenden Abbildungen** auf allgemeinen **lokalkonvexen Räumen** eingeführt. Die Definition lautet darin, dass eine Abbildung $T : X \rightarrow Y$ **absolutsummierend** ist, wenn sie unbedingt (**Def. A.2.17**) in absolut konvergente Reihen (**Def. A.2.16**) abbildet, d.h. im Fall von normierten Räumen X, Y , dass für $(x_i)_{i \in I} \subset X$ gilt:

$$\sum_{i \in I} x_i \text{ konvergiert unbedingt} \Rightarrow \sum_{i \in I} \|Tx_i\|_Y < \infty \quad (\text{„konvergiert absolut“}). \quad (6.30)$$

Aus dieser Definition stammt auch der dann besser nachvollziehbare Begriff „**absolutsummierend**“. [28, Satz 2.2.1] zeigt weiter, dass im Fall von normierten Räumen die Definition von absolutsummierenden Operatoren äquivalent zu (6.29) ist, für $p = q = 1$ (siehe **Def. C.1.1**). Dieser Fakt und die Ähnlichkeit von (6.29) zu den $\ell^p(\mathbb{N})$ -Normen motivierte in der Folgezeit die Erweiterung der absolutsummierenden Operatoren hin zu den in **Def. 6.3.1** eingeführten p - q -summierenden Operatoren. Die absolutsummierenden sind dann nämlich gerade der Spezialfall der **1-summierenden Operatoren**.

Vorbereitend für den **Satz 6.5.1** von Mercer ist relevant, dass für endliche Maßräume (Ω, μ) (**Def. A.6.4**) der **Einbettungsoperator von $L^\infty(\Omega, \mu)$ nach $L^2(\Omega, \mu)$** 2-summierend ist und daraus resultierend auch jeder Operator, der beschränkt aus dem $L^2(\Omega, \mu)$ in den $L^\infty(\Omega, \mu)$ abbildet. (Bei begrifflichen Unklarheiten wird auf **Abschnitt A.6** sowie **Bsp. A.2.14** und **Bsp. A.3.9** verwiesen.)

Proposition 6.3.3 (Einbettungsoperator von $L^\infty(\Omega, \mu)$ nach $L^2(\Omega, \mu)$ ist 2-summierend für endliche Maßräume [18, Prop. 1.d.5]).

Sei (Ω, μ) ein endlicher Maßraum. Dann ist der Einbettungsoperator $\iota : L^\infty(\Omega, \mu) \rightarrow L^2(\Omega, \mu)$, $\iota f = f$ für $f \in L^\infty(\Omega, \mu)$ 2-summierend und es gilt für alle $n \in \mathbb{N}$ und $f_1, \dots, f_n \in L^\infty(\Omega, \mu)$:

$$\left(\sum_{i=1}^n \|\iota f_i\|_{L^2(\Omega, \mu)}^2 \right)^{\frac{1}{2}} = \sqrt{\mu(\Omega)} \sup_{\|g\|_{(L^\infty(\Omega, \mu))', L^\infty(\Omega, \mu)} \leq 1} \left(\sum_{i=1}^n |\langle g, f_i \rangle_{(L^\infty(\Omega, \mu))', L^\infty(\Omega, \mu)}|^2 \right)^{\frac{1}{2}}. \quad (6.31)$$

Beweis. (Die Beweisideen sind [18, Prop. 1.d.5] entnommen.)

Sei $n \in \mathbb{N}$ sowie zugehörige $f_1, \dots, f_n \in L^\infty(\Omega, \mu)$ beliebig, aber fest. Dann gilt:

$$\sum_{i=1}^n \|\iota f_i\|_{L^2(\Omega, \mu)}^2 \stackrel{\text{A.2.14}}{=} \int_{\Omega} \sum_{i=1}^n |f_i(x)|^2 d\mu(x) \stackrel{\text{A.2.14}}{\leq} \mu(\Omega) \left\| \sum_{i=1}^n |f_i|^2 \right\|_{L^\infty(\Omega, \mu)} := \mu(\Omega) M. \quad (6.32)$$

Um die rechte Seite weiter abschätzen zu können, bediene man sich nun des Satzes A.4.12 von Hahn-Banach: Dazu wähle man zunächst $\varepsilon > 0$ beliebig, aber fest. Aus der Definition von $M := \left\| \sum_{i=1}^n |f_i|^2 \right\|_{L^\infty(\Omega, \mu)}$ und Bsp. A.2.14 folgt, dass es eine Menge $\emptyset \neq \Omega' \subseteq \Omega$ gibt, sodass gilt:

$$M - \varepsilon \leq \sum_{i=1}^n |f_i(x)|^2 \leq M \quad \text{für alle } x \in \Omega'. \quad (6.33)$$

Man wähle nun $x \in \Omega'$ fest und betrachte das auf $V := \text{span}(f_1, \dots, f_n)$ definierte Auswertungsfunktional (Def. A.4.1) an der Stelle x

$$\tilde{g}_x : V \rightarrow \mathbb{K}, \quad \tilde{g}_x(f) = f(x). \quad (6.34)$$

$\tilde{g}_x$ ist linear, denn für $f, h \in V$ und $\alpha, \beta \in \mathbb{N}$ gilt $\tilde{g}_x(\alpha f + \beta h) = \alpha f(x) + \beta h(x) = \alpha \tilde{g}_x(f) + \beta \tilde{g}_x(h)$. Zudem ist es beschränkt, wegen $|\tilde{g}_x(f)| = |f(x)| \leq \|f\|_V$ mit $\|\tilde{g}_x\|_{V'} \leq 1$. Nach dem Hahn-Banach Theorem A.4.12 gibt es somit ein beschränktes lineares Funktional $g_x \in (L^\infty(\Omega, \mu))'$ mit $\|g_x\|_{(L^\infty(\Omega, \mu))'} \leq 1$ und $g_x|_V = \tilde{g}_x$. Mit diesem erhält man nun

$$M - \varepsilon \stackrel{(6.33)}{\leq} \sum_{i=1}^n |f_i(x)|^2 \stackrel{(6.34)}{=} \sum_{i=1}^n |\langle g_x, f_i \rangle_{(L^\infty(\Omega, \mu))', L^\infty(\Omega, \mu)}|^2 \quad (6.35)$$

$$\stackrel{\text{A.4.12}}{\leq} \sup_{\|g\|_{(L^\infty(\Omega, \mu))'} \leq 1} \sum_{i=1}^n |\langle g, f_i \rangle_{(L^\infty(\Omega, \mu))', L^\infty(\Omega, \mu)}|^2. \quad (6.36)$$

Da $\varepsilon > 0$ beliebig gewählt ist, folgt somit

$$M \leq \sup_{\|g\|_{(L^\infty(\Omega, \mu))'} \leq 1} \sum_{i=1}^n |\langle g, f_i \rangle_{(L^\infty(\Omega, \mu))', L^\infty(\Omega, \mu)}|^2 \quad (6.37)$$

und die Behauptung schließlich durch Kombination von (6.32) mit (6.37) und Wurzelziehen. $\square$

Korollar 6.3.4 (2-summierende Operatoren auf endlichen Maßräumen [18, Lemma 1.d.4]).

Sei (Ω, μ) ein endlicher Maßraum $T \in \mathcal{L}(L^2(\Omega, \mu)) \cap \mathcal{L}(L^2(\Omega, \mu), L^\infty(\Omega, \mu))$. Dann ist T 2-summierend und es gilt für alle $n \in \mathbb{N}$ und alle $f_1, \dots, f_n \in L^2(\Omega, \mu)$:

$$\left(\sum_{i=1}^n \|T f_i\|_{L^2(\Omega, \mu)}^2 \right)^{\frac{1}{2}} \leq \sqrt{\mu(\Omega)} \|T\|_{\mathcal{L}(L^2(\Omega, \mu), L^\infty(\Omega, \mu))} \sup_{\|g\|_{L^2(\Omega, \mu)} \leq 1} \left(\sum_{i=1}^n |\langle g, f_i \rangle_{L^2(\Omega, \mu)}|^2 \right)^{\frac{1}{2}}. \quad (6.38)$$

Beweis. (Dieses Korollar ist ein Spezialfall von [18, Lemma 1.d.4].)

Die Behauptung folgt mit dem Einbettungsoperator $\iota : L^\infty(\Omega, \mu) \rightarrow L^2(\Omega, \mu)$, $\iota f = f$ direkt aus Prop. 6.3.3 unter Zuhilfenahme des adjungierten Operators T' von T (Def.+Satz A.4.13):

$$\sum_{i=1}^n \|T f_i\|_{L^2(\Omega, \mu)}^2 = \sum_{i=1}^n \|\iota T f_i\|_{L^2(\Omega, \mu)}^2 \stackrel{6.3.3}{\leq} \mu(\Omega) \sup_{\|g\|_{(L^\infty(\Omega, \mu))'} \leq 1} \sum_{i=1}^n |\langle g, T f_i \rangle_{(L^\infty(\Omega, \mu))', L^\infty(\Omega, \mu)}|^2 \quad (6.39)$$

$$\stackrel{A.4.13}{=} \mu(\Omega) \sup_{\|g\|_{(L^\infty(\Omega, \mu))'} \leq 1} \sum_{i=1}^n \left| \left\langle \frac{\|T\|_{\mathcal{L}(L^2(\Omega, \mu), L^\infty(\Omega, \mu))}}{\|T'\|_{\mathcal{L}((L^\infty(\Omega, \mu))', (L^2(\Omega, \mu))')}} T' g, f_i \right\rangle_{(L^2(\Omega, \mu))', L^2(\Omega, \mu)} \right|^2 \quad (6.40)$$

$$\stackrel{A.3.1}{\leq} \mu(\Omega) \|T\|_{\mathcal{L}(L^2(\Omega, \mu), L^\infty(\Omega, \mu))}^2 \sup_{\|g\|_{(L^2(\Omega, \mu))'} \leq 1} \sum_{i=1}^n |\langle g, f_i \rangle_{(L^2(\Omega, \mu))', L^2(\Omega, \mu)}|^2 \quad (6.41)$$

$$\stackrel{A.4.10}{=} \mu(\Omega) \|T\|_{\mathcal{L}(L^2(\Omega, \mu), L^\infty(\Omega, \mu))}^2 \sup_{\|g\|_{L^2(\Omega, \mu)} \leq 1} \sum_{i=1}^n |\langle g, f_i \rangle_{L^2(\Omega, \mu)}|^2. \quad (6.42)$$

Gemäß Def. 6.3.1 ist T somit wie behauptet 2-summierend. (Im vorletzten Schritt wurde dabei verwendet, dass wegen $T' \in \mathcal{L}((L^\infty(\Omega, \mu))', (L^2(\Omega, \mu))')$ für alle $g \in (L^\infty(\Omega, \mu))'$ mit $\|g\|_{(L^\infty(\Omega, \mu))'} \leq 1$ gilt, dass

$$\left\| \frac{1}{\|T'\|_{\mathcal{L}((L^\infty(\Omega, \mu))', (L^2(\Omega, \mu))')}} T' g \right\|_{(L^2(\Omega, \mu))'} \stackrel{A.4.4}{\leq} \frac{\|T'\|_{\mathcal{L}((L^\infty(\Omega, \mu))', (L^2(\Omega, \mu))')}}{\|T'\|_{\mathcal{L}((L^\infty(\Omega, \mu))', (L^2(\Omega, \mu))')}} \|g\|_{(L^\infty(\Omega, \mu))'} \leq 1 \cdot 1 \quad (6.43)$$

ist und folglich, dass das Supremum in der vorgenommenen Weise weiter abgeschätzt werden darf.) $\square$

Damit sind nun fast alle benötigten Vorresultate für den Satz 6.5.1 von Mercer erarbeitet. Es fehlt noch ein winziger nützlicher Zusammenhang zwischen dem $L^2(\Omega, \mu)$ und dem $L^1(\Omega, \mu)$.

6.4 $L^2(\Omega, \mu)$ liegt dicht in $L^1(\Omega, \mu)$

Proposition 6.4.1 ($L^2(\Omega, \mu)$ liegt dicht in $L^1(\Omega, \mu)$ [14, 8, Theorem 7.8]).

Sei (Ω, μ) ein Maßraum. Dann ist die Menge $L^2(\Omega, \mu)$ dicht in $(L^1(\Omega, \mu), \|\cdot\|_{L^1(\Omega, \mu)})$.

Beweis. (Die Bewesideen sind [14, 8, Theorem 7.8] entnommen.)

Sei $f \in L^1(\Omega, \mu)$ beliebig. Man definiere $(f_n)_{n \in \mathbb{N}} \subset \mathcal{F}(\Omega)$ durch $f_n := f \cdot \mathbf{1}_{\{|f| \leq n\}}$ für $n \in \mathbb{N}$. Es gilt:

$$\int_{\Omega} |f_n(x)|^2 d\mu(x) = \int_{\Omega \cap \{|f| \leq n\}} |f(x)|^2 d\mu(x) \leq n \int_{\Omega \cap \{|f| \leq n\}} |f(x)| d\mu(x) \stackrel{A.2.14}{\leq} n \|f\|_{L^1(\Omega, \mu)} < \infty. \quad (6.44)$$

Folglich ist $(f_n)_{n \in \mathbb{N}} \subset L^2(\Omega, \mu)$ und da definitionsgemäß $|f_n| \leq |f|$, folgt mit der Monotonie des Integrals (Satz A.6.15), dass $(f_n)_{n \in \mathbb{N}} \subset L^1(\Omega, \mu)$ und wegen der Vektorraumstruktur von Letzterem somit auch $(f_n - f)_{n \in \mathbb{N}} \subset L^1(\Omega, \mu)$. Damit ist der Satz A.6.19 von Lebesgue anwendbar und es folgt:

$$\lim_{n \rightarrow \infty} \|f_n - f\|_{L^1(\Omega, \mu)} \stackrel{\text{A.2.14}}{=} \lim_{n \rightarrow \infty} \int_{\Omega} |f_n(x) - f(x)| d\mu(x) \stackrel{\text{A.6.19}}{=} \int_{\Omega} \lim_{n \rightarrow \infty} |f_n(x) - f(x)| d\mu(x) = 0. \quad (6.45)$$

Ergo gilt $f_n \rightarrow f$ in $(L^1(\Omega, \mu), \|\cdot\|_{L^1(\Omega, \mu)})$ und da $f \in L^1(\Omega, \mu)$ beliebig gewählt war, folgt die Behauptung. $\square$

Alle Zutaten für das namensgebende Ergebnis dieses Kapitels sind nun vorbereitet.

6.5 Der Satz von Mercer

Satz 6.5.1 (Satz von Mercer) (auf endlichen Maßräumen) [18, 33, Theorem 3.a.1, Theorem 2.10]). Sei (Ω, μ) ein endlicher Maßraum und $k \in L^\infty(\Omega^2, \mu^2)$ (d. h. es existieren $M > 0$ und Nullmengen (Def. A.6.8) $N, \tilde{N} \subseteq \Omega$, sodass $\sup_{x \in \Omega \setminus N} \sup_{y \in \Omega \setminus \tilde{N}} |k(x, y)| \leq M$ gilt) der Form, dass der Operator

$$T_k : L^2(\Omega, \mu) \rightarrow L^2(\Omega, \mu), \quad (T_k f)(x) := \int_{\Omega} k(x, y) f(y) d\mu(y) \quad \text{für } f \in L^2(\Omega, \mu), x \in \Omega \quad (6.46)$$

selbstadjungiert und positiv ist, d. h. für alle $f \in L^2(\Omega, \mu)$ gilt:

$$\langle T_k f, f \rangle_{L^2(\Omega, \mu)} \stackrel{\text{A.3.9}}{=} \int_{\Omega} \int_{\Omega} k(x, y) f(x) \overline{f(y)} d\mu(y) d\mu(x) \geq 0. \quad (6.47)$$

Dann gelten:

1. Die höchstens abzählbar unendlich vielen, mit Multiplizität gezählten Eigenwerte $(\lambda_i)_{i \in I} \subset \mathbb{C}$ (demnach $I \subseteq \mathbb{N}$) sind allesamt nicht-negativ (reell) und zudem gilt (falls I endlich ist, die Folge mit Null unendlich oft fortgesetzt/aufgefüllt): $(\lambda_i)_{i \in \mathbb{N}} \in \ell^1(\mathbb{N})$, d. h. $\sum_{i \in I} |\lambda_i| < \infty$.
2. Jedes Bild von T_k ist fast überall (Def. A.6.8) beschränkt (d. h. $T_k(L^2(\Omega, \mu)) \subseteq L^\infty(\Omega, \mu)$). Insbesondere ist dies jede Eigenfunktion zu einem Eigenwert von T_k ungleich Null.
3. Sind $(e_i)_{i \in I} \subset L^2(\Omega, \mu)$ die paarweise verschiedenen, $\|\cdot\|_{L^2(\Omega, \mu)}$ -normierten Eigenfunktionen (d. h. $\|e_i\|_{L^2(\Omega, \mu)} = 1$), jeweils zugehörig zu den Eigenwerten $(\lambda_i)_{i \in I}$ von T_k ($I \subseteq \mathbb{N}$), dann konvergiert die (Spektral-)Reihenentwicklung von k fast überall absolut (punktweise):

$$k(x, y) = \sum_{i \in I} \lambda_i e_i(x) \overline{e_i(y)} \quad \text{für } x, y \in \Omega \text{ fast überall.} \quad (6.48)$$

4. Die (Spektral-)Reihenentwicklung von k (6.48) konvergiert genau dann sogar gleichmäßig absolut fast überall (bzw. überall) (Bsp. A.2.9, Def. A.6.8), wenn die folgende Reihe fast überall (bzw. überall) konvergiert:

$$k(x, x) = \sum_{i \in I} \lambda_i |e_i(x)|^2. \quad (6.49)$$

Beweis. (Die Beweisstruktur ist [18, Theorem 3.a.1] und in Punkt 4. die Argumentation [1, 7, 26, 33, Theorem 6.1.1, Theorem III.22, Problem 2.24] entnommen und vereinigt.)

1. Da (Ω, μ) und somit auch (Ω^2, μ^2) ein endlicher Maßraum ist, gilt gemäß [Bsp. A.2.14](#) auch: $k \in L^\infty(\Omega^2, \mu^2) \subseteq L^2(\Omega^2, \mu^2)$. Es sind damit bereits alle Bedingungen von [Def.+Satz A.4.18](#) erfüllt, d. h. T_k ist ein (wohldefinierter) kompakter ([Def.+Satz A.4.6](#)) Hilbert-Schmidt Integraloperator auf $L^2(\Omega, \mu)$ sowie selbstadjungiert und positiv nach Voraussetzung.

Damit lässt sich für T_k der Spektralsatz für kompakte selbstadjungierte Operatoren ([Satz A.4.17](#)) anwenden und dieser liefert den ersten Teil der Behauptung: T_k besitzt höchstens abzählbar unendlich viele Eigenwerte $(\lambda_i)_{i \in I}$ mit jeweils endlicher Multiplizität (also $I \subseteq \mathbb{N}$), welche gemäß [Def.+Satz 6.2.1](#) allesamt nicht-negativ und reell sind, denn T_k ist positiv.

Aufgrund der Abzählbarkeit und Nicht-Negativität der Eigenwerte von T_k sei im Folgenden o. B. d. A. $I = \mathbb{N}$ und $\lambda_i > 0$ für alle $i \in \mathbb{N}$ angenommen, denn die Behauptung $\sum_{i \in I} |\lambda_i| < \infty$ ist für endliches I trivialerweise erfüllt und bei Existenz von $\lambda_i = 0$ auch dann erfüllt, wenn man Letztere aus der Folge $(\lambda_i)_{i \in I}$ entfernt.

Aus der Selbstadjungiertheit und Positivität von T_k lässt sich mithilfe von [Prop. 6.2.2](#) weiter schließen, dass eine Wurzel von T_k existiert, d. h. es gibt einen beschränkten, selbstadjungierten Operator $S \in \mathcal{L}(L^2(\Omega, \mu))$ für den $S^2 = T_k$ gilt.

Für S folgt mit [Kor. 6.2.5](#) – da jeder Eigenwert $\lambda_i > 0$ von T_k wie erwähnt endliche Multiplizität besitzt, d. h. sein Eigenraum endlich-dimensional ist – dass orthonormale Eigenfunktionen $\tilde{e}_{i,1}, \dots, \tilde{e}_{i,n_i} \in L^2(\Omega, \mu)$ der Wurzel S existieren (für $n_i = \dim(\ker(T_k - \lambda_i \text{Id})) \in \mathbb{N}$), die jeweils entweder dem Eigenwert $\sqrt{\lambda_i} > 0$ oder $-\sqrt{\lambda_i} < 0$ zugeordnet sind (für jedes $i \in \mathbb{N}$).

Da dies für jeden Eigenwert $\lambda_i > 0$ von T_k gilt, existiert somit ein Orthonormalsystem $(e_i)_{i \in \mathbb{N}} \subset L^2(\Omega, \mu)$ von Eigenfunktionen für S , wobei e_i jeweils einem Eigenwert ν_i von S zugeordnet ist, sodass $\nu_i^2 = \lambda_i$ gilt für alle $i \in \mathbb{N}$.

Um damit endlich die Aussage zu belegen, muss im Folgenden unter größerem Aufwand noch gezeigt werden, dass $S \in \mathcal{L}(L^2(\Omega, \mu))$ ebenfalls beschränkt von $L^2(\Omega, \mu)$ nach $L^\infty(\Omega, \mu)$ abbildet mit $\|S\|_{\mathcal{L}(L^2(\Omega, \mu), L^\infty(\Omega, \mu))} \leq \sqrt{M} < \infty$. Sei dieses Ergebnis aber zunächst einmal als bereits bewiesen angenommen. Dann folgt mit dem vorbereiteten [Kor. 6.3.4](#), dass S 2-summierend ist und daraus gemeinsam mit allen obigen Überlegungen unter Zuhilfenahme der Besselschen Ungleichung ([Satz A.3.14](#)):

$$\sum_{i=1}^{\infty} |\lambda_i| \stackrel{6.2.5}{=} \sum_{i=1}^{\infty} |\nu_i^2| \stackrel{A.3.12}{=} \sum_{i=1}^{\infty} |\nu_i|^2 \|e_i\|_{L^2(\Omega, \nu)}^2 \stackrel{A.2.4}{=} \sum_{i=1}^{\infty} \|\nu_i e_i\|_{L^2(\Omega, \mu)}^2 \stackrel{A.4.15}{=} \sum_{i=1}^{\infty} \|S e_i\|_{L^2(\Omega, \mu)}^2 \quad (6.50)$$

$$= \lim_{n \rightarrow \infty} \sum_{i=1}^n \|S e_i\|_{L^2(\Omega, \mu)}^2 \stackrel{6.3.4}{\leq} \lim_{n \rightarrow \infty} M \mu(\Omega) \sup_{\|g\|_{L^2(\Omega, \mu)} \leq 1} \sum_{i=1}^n |\langle g, e_i \rangle_{L^2(\Omega, \mu)}|^2 \quad (6.51)$$

$$\stackrel{A.3.14}{\leq} M \mu(\Omega) \sup_{\|g\|_{L^2(\Omega, \mu)} \leq 1} \|g\|_{L^2(\Omega, \mu)}^2 = M \mu(\Omega) < \infty. \quad (6.52)$$

Ergo ist die Behauptung unter 1. korrekt, sobald die oben als bereits bewiesen angenommen Unterbehauptung gezeigt ist, nämlich dass $S \in \mathcal{L}(L^2(\Omega, \mu)) \cap \mathcal{L}(L^2(\Omega, \mu), L^\infty(\Omega, \mu))$ gilt mit $\|S\|_{\mathcal{L}(L^2(\Omega, \mu), L^\infty(\Omega, \mu))} \leq \sqrt{M} < \infty$. Dies werde nun im Folgenden erledigt:

Zunächst gilt für $f \in L^2(\Omega, \mu) \subseteq L^1(\Omega, \mu)$ (vgl. Bsp. A.2.14):

$$\|Sf\|_{L^2(\Omega, \mu)}^2 \stackrel{\text{A.3.3}}{=} \langle Sf, Sf \rangle_{L^2(\Omega, \mu)} \stackrel{\text{A.4.14}}{=} \langle S^2 f, f \rangle_{L^2(\Omega, \mu)} \stackrel{\text{6.2.2}}{=} \langle Tf, f \rangle_{L^2(\Omega, \mu)} \quad (6.53)$$

$$\stackrel{\text{A.3.9}}{\stackrel{\text{(6.46)}}{=}} \int_{\Omega} \int_{\Omega} k(x, y) f(y) d\mu(y) \overline{f(x)} d\mu(x) \stackrel{\text{A.2.14}}{\leq} M \|f\|_{L^1(\Omega, \mu)}^2 \quad (6.54)$$

Folglich ist S ebenfalls als Operator von $(L^2(\Omega, \mu), \|\cdot\|_{L^1(\Omega, \mu)})$ (man beachte die unterschiedliche Norm) nach $L^2(\Omega, \mu)$ beschränkt.

Da gemäß Prop. 6.4.1 die Menge $L^2(\Omega, \mu)$ dicht in $(L^1(\Omega, \mu), \|\cdot\|_{L^1(\Omega, \mu)})$ liegt, existiert mit Satz A.4.5 ein (eindeutiger) beschränkter Operator $\tilde{S} \in \mathcal{L}(L^1(\Omega, \mu), L^2(\Omega, \mu))$, welcher S auf ganz $L^1(\Omega, \mu)$ fortsetzt (Man beachte, dass $L^2(\Omega, \mu)$ vollständig und zudem ein Vektorraum, ergo Unterraum von $L^1(\Omega)$ im Fall des endlichen Maßraums (Ω, μ) ist). Es gilt zudem gemäß Satz A.4.5 noch $\|\tilde{S}\|_{\mathcal{L}(L^1(\Omega, \mu), L^2(\Omega, \mu))} \leq \sqrt{M}$.

Wegen der Beschränktheit von $\tilde{S}$ existiert mit Def.+Satz A.4.13 sein adjungierter Operator $\tilde{S}' \in \mathcal{L}((L^2(\Omega, \mu))', (L^1(\Omega, \mu))')$ mit $\|\tilde{S}'\|_{\mathcal{L}((L^2(\Omega, \mu))', (L^1(\Omega, \mu))')} \leq \sqrt{M}$. $\tilde{S}'$ ist dabei so charakterisiert, dass für alle $x \in L^1(\Omega, \mu)$, $y' \in (L^2(\Omega, \mu))'$ gilt:

$$\langle \tilde{S}' y', x \rangle_{(L^1(\Omega, \mu))', L^1(\Omega, \mu)} = \langle y', \tilde{S} x \rangle_{(L^2(\Omega, \mu))', L^2(\Omega, \mu)}. \quad (6.55)$$

Sei nun kurzzeitig $y' \in (L^2(\Omega, \mu))'$ festgehalten. Dann existieren nach dem Rieszschen Darstellungssatz (Satz A.4.9) zwei Elemente $g \in L^2(\Omega, \mu)$ und $\tilde{g} \in L^\infty(\Omega, \mu)$, sodass für alle $h \in L^2(\Omega, \mu)$, $f \in L^1(\Omega, \mu)$ gilt:

$$y'(h) = \int_{\Omega} g(x) h(x) d\mu(x) \quad \text{und} \quad (\tilde{S}' y')(f) = \int_{\Omega} \tilde{g}(x) f(x) d\mu(x). \quad (6.56)$$

Assoziiert man nun $\tilde{S}' g$ mit $\tilde{g}$, so lässt sich $\tilde{S}'$ in bekannter Art und Weise auch als Operator zwischen $L^2(\Omega, \mu)$ und $L^\infty(\Omega, \mu)$ interpretieren. Als dieser ist er zudem linear und beschränkt mit $\|\tilde{S}'\|_{\mathcal{L}(L^2(\Omega, \mu), L^\infty(\Omega, \mu))} \leq \sqrt{M}$, da der Rieszsche Darstellungssatz die isometrisch-isomorphe Beziehung für $(L^2(\Omega, \mu))' \cong L^2(\Omega, \mu)$ und $(L^1(\Omega, \mu))' \cong L^\infty(\Omega, \mu)$ liefert. Außerdem ergibt sich dann (6.55) zu

$$\int_{\Omega} (\tilde{S}' g)(x) f(x) d\mu(x) \stackrel{\text{(6.56)}}{=} \langle \tilde{S}' y', f \rangle_{(L^1(\Omega, \mu))', L^1(\Omega, \mu)} \stackrel{\text{(6.55)}}{=} \langle y', \tilde{S} f \rangle_{(L^2(\Omega, \mu))', L^2(\Omega, \mu)} \quad (6.57)$$

$$\stackrel{\text{(6.56)}}{=} \int_{\Omega} g(x) (\tilde{S} f)(x) d\mu(x) \quad \text{für alle } f \in L^1(\Omega, \mu). \quad (6.58)$$

$y' \in (L^2(\Omega, \mu))'$ war bisher beliebig, aber festgehalten, d.h. das bisher erarbeitete gilt darum für alle $y' \in (L^2(\Omega, \mu))'$, ergo gelten (6.57)+(6.58) für alle $f \in L^1(\Omega, \mu)$, $g \in L^2(\Omega, \mu)$ wegen der isometrisch-isomorphen Beziehung $(L^2(\Omega, \mu))' \cong L^2(\Omega, \mu)$.

Nun folgt aus der Selbstadjungiertheit von S aber ebenfalls für alle $f, g \in L^2(\Omega, \mu)$ eine ähnliche Bedingung für S und alle $f, g \in L^2(\Omega, \mu)$:

$$\int_{\Omega} \overline{(Sg)(x)} f(x) d\mu(x) \stackrel{\text{A.3.9}}{=} \langle f, Sg \rangle_{L^2(\Omega, \mu)} \stackrel{\text{A.4.14}}{=} \langle Sf, g \rangle_{L^2(\Omega, \mu)} \stackrel{\text{A.3.9}}{=} \int_{\Omega} \overline{g(x)} (Sf)(x) d\mu(x). \quad (6.59)$$

Diese 2 Gleichungen (6.58)+(6.59) lassen sich nun in Kombination mit $L^2(\Omega, \mu) \subseteq L^1(\Omega, \mu)$ im endlichen Maßraumfall gewinnbringend verbinden. Es gilt für alle $f, g \in L^2(\Omega, \mu)$:

$$\langle f, Sg - \tilde{S}'\bar{g} \rangle_{L^2(\Omega, \mu)} \stackrel{\text{A.3.9}}{=} \int_{\Omega} f(x) \overline{(Sg)(x)} f(x) - (\tilde{S}'\bar{g})(x) f(x) d\mu(x) \quad (6.60)$$

$$\stackrel{\text{(6.58)}}{=} \int_{\Omega} \overline{g(x)} (Sf)(x) - \overline{g(x)} (\tilde{S}f)(x) d\mu(x) = 0, \quad (6.61)$$

da $Sf = \tilde{S}f$ für alle $f \in L^2(\Omega, \mu)$, denn $\tilde{S} \in \mathcal{L}(L^1(\Omega, \mu), L^2(\Omega, \mu))$ ist als die stetige Fortsetzung von $S \in \mathcal{L}(L^2(\Omega, \mu))$ definiert. Die Gleichung (6.61) gilt für alle $f, g \in L^2(\Omega, \mu)$, was wegen der positiven Definitheit des Skalarprodukts $Sg = \tilde{S}'\bar{g}$ im $L^2(\Omega, \mu)$ liefert und somit gemäß Bsp. A.6.17, dass $Sg = \tilde{S}'\bar{g}$ fast überall für alle $g \in L^2(\Omega, \mu)$. Damit schließt man schließlich:

$$\|Sg\|_{L^\infty(\Omega, \mu)} \stackrel{\text{A.2.14}}{=} \|\tilde{S}'\bar{g}\|_{L^\infty(\Omega, \mu)} \stackrel{\text{A.2.14}}{=} \|\tilde{S}'g\|_{L^\infty(\Omega, \mu)} \quad (6.62)$$

$$\stackrel{\text{A.4.4}}{\leq} \|\tilde{S}'\|_{\mathcal{L}(L^2(\Omega, \mu), L^\infty(\Omega, \mu))} \|g\|_{L^2(\Omega, \mu)} \stackrel{\text{A.2.14}}{\leq} \sqrt{M} \|g\|_{L^2(\Omega, \mu)}. \quad (6.63)$$

Mit ziemlich großem Aufwand ist damit final gezeigt, dass $S \in \mathcal{L}(L^2(\Omega, \mu), L^\infty(\Omega, \mu))$ gilt mit $\|S\|_{\mathcal{L}(L^2(\Omega, \mu), L^\infty(\Omega, \mu))} \leq \sqrt{M}$ und die Behauptung unter 1. ist vollständig bewiesen.

2. Mit dem endlichen Maßraum Ω und $k(x, y) \leq M$ fast überall in $\Omega \times \Omega$ gilt für alle $f \in L^2(\Omega, \mu)$:

$$\|T_k f\|_{L^\infty(\Omega, \mu)} \stackrel{\text{A.2.14}}{=} \sup_{x \in \Omega \setminus N} \left| \int_{\Omega} k(x, y) f(y) dy \right| \stackrel{\text{A.2.14}}{\leq} M \int_{\Omega} |f(y)| dy \stackrel{\text{A.2.14}}{=} M \|f \cdot 1\|_{L^1(\Omega, \mu)} \quad (6.64)$$

$$\stackrel{\text{A.2.15}}{\leq} M \|f\|_{L^2(\Omega, \mu)} \|1\|_{L^2(\Omega, \mu)} \stackrel{\text{A.2.14}}{=} \sqrt{\mu(\Omega)} M \|f\|_{L^2(\Omega, \mu)} < \infty, \quad (6.65)$$

ergo $T_k(L^2(\Omega, \mu)) \subseteq L^\infty(\Omega, \mu)$. Ist $\lambda \neq 0$ ein Eigenwert von T_k und $e \in L^2(\Omega, \mu)$ die zugehörige Eigenfunktion, so ist ebenfalls wie behauptet $e \in L^\infty(\Omega, \mu)$, denn es gilt:

$$\|e\|_{L^\infty(\Omega, \mu)} \stackrel{\text{A.2.4}}{=} \left| \frac{1}{\lambda} \right| \|\lambda e\|_{L^\infty(\Omega, \mu)} \stackrel{\text{A.4.15}}{=} \left| \frac{1}{\lambda} \right| \|T_k e\|_{L^\infty(\Omega, \mu)} \stackrel{\text{(6.65)}}{<} \infty. \quad (6.66)$$

3. O. B. d. A. sei fortan $I = \mathbb{N}$ angenommen (z. B. indem für endliches I die Folgen $(\lambda_i)_{i \in I}$ und $(e_i)_{i \in I}$ mit 0 unendlich fortgesetzt/aufgefüllt werden) sowie zur besseren Lesbarkeit definiert:

$$E_i : \Omega \times \Omega \rightarrow \mathbb{K}, \quad E_i(x, y) := e_i(x) \overline{e_i(y)} \quad \text{für } x, y \in \Omega, \quad i \in \mathbb{N}. \quad (6.67)$$

Betrachtet man dann die zwei Teilsummenfolgen $(s_n(x, y))_{n \in \mathbb{N}}$ und $(\tilde{s}_n(x, y))_{n \in \mathbb{N}}$ mit $s_n(x, y) := \sum_{i=1}^n |\lambda_i E_i(x, y)|$ sowie $\tilde{s}_n(x, y) := \sum_{i=1}^n \lambda_i E_i(x, y)$ für $x, y \in \Omega$, so gilt für $n, m \in \mathbb{N}$ mit $m \leq n$:

$$\left\{ \begin{array}{l} \|s_n - s_m\|_{L^2(\Omega^2, \mu^2)} \\ \|\tilde{s}_n - \tilde{s}_m\|_{L^2(\Omega^2, \mu^2)} \end{array} \right\} = \left\{ \begin{array}{l} \left\| \sum_{i=m}^n |\lambda_i E_i| \right\|_{L^2(\Omega^2, \mu^2)} \\ \left\| \sum_{i=m}^n \lambda_i E_i \right\|_{L^2(\Omega^2, \mu^2)} \end{array} \right\} \stackrel{\text{A.2.4}}{\leq} \sum_{i=m}^n |\lambda_i| \|E_i\|_{L^2(\Omega^2, \mu^2)} \quad (6.68)$$

$$\stackrel{\text{A.2.14}}{=} \sum_{i=m}^n |\lambda_i| \left(\int_{\Omega} \int_{\Omega} |e_i(x) \overline{e_i(y)}|^2 d\mu(x) d\mu(y) \right)^{\frac{1}{2}} \quad (6.69)$$

$$\stackrel{\text{A.8.1}}{=} \sum_{i=m}^n |\lambda_i| \left(\int_{\Omega} |e_i(x)|^2 d\mu(x) \int_{\Omega} |e_i(y)|^2 d\mu(y) \right)^{\frac{1}{2}} \quad (6.70)$$

$$\stackrel{\text{A.2.14}}{=} \sum_{i=m}^n |\lambda_i| \|e_i\|_{L^2(\Omega, \mu)}^2 = \sum_{i=m}^n |\lambda_i| \cdot 1^2 = \sum_{i=m}^n |\lambda_i| \rightarrow 0 \quad (n \rightarrow \infty). \quad (6.71)$$

Damit sind $(s_n)_{n \in \mathbb{N}}$ und $(\tilde{s}_n)_{n \in \mathbb{N}}$ Cauchyfolgen in $L^2(\Omega^2, \mu^2)$ und wegen dessen Vollständigkeit (siehe Bsp. A.2.14) folgt, dass die beiden Reihen $\sum_{i=1}^{\infty} |\lambda_i E_i|$ und $\sum_{i=1}^{\infty} \lambda_i E_i$ in $L^2(\Omega^2, \mu^2)$ konvergieren. Als Grenzwert (Def. A.1.10) von $(\tilde{s}_n)_{n \in \mathbb{N}}$ in $L^2(\Omega^2, \mu^2)$ ist die nachfolgende Funktion $\hat{k} \in L^2(\Omega^2, \mu^2)$ also wohldefiniert mit

$$\hat{k}(x, y) := \sum_{i=1}^{\infty} \lambda_i E_i(x, y) \stackrel{(6.67)}{=} \sum_{i=1}^{\infty} \lambda_i e_i(x) \overline{e_i(y)} \quad \text{für } x, y \in \Omega \quad (6.72)$$

und die Konvergenz der Reihe ist dabei absolut (punktweise) fast überall gemäß Theorem A.6.18. (Man beachte dabei die absolute Konvergenz fast überall jeder Teilfolge.)

Um schließlich noch die Gleichheit fast überall von k und $\hat{k}$ einzusehen, beachte man, dass unter 1. gezeigt wurde, dass T_k ein selbstadjungierter Hilbert-Schmidt Integraloperator ist und gemäß dessen Def.+Satz A.4.18 damit insbesondere kompakt ist. Für alle $f \in L^2(\Omega, \mu)$ und fast alle $x \in \Omega$ liefert demnach der Spektralsatz für kompakte, selbstadjungierte Operatoren:

$$(T_k f)(x) = \sum_{i=1}^{\infty} \lambda_i \langle f, e_i \rangle_{L^2(\Omega, \mu)} e_i(x). \quad (6.73)$$

Man folgert für beliebige $f \in L^2(\Omega, \mu)$ und fast alle $x \in \Omega$:

$$\langle k(x, \cdot) - \hat{k}(x, \cdot), f \rangle_{L^2(\Omega, \mu)} \stackrel{\text{A.3.1}}{=} \langle k(x, \cdot), f \rangle_{L^2(\Omega, \mu)} - \left\langle \sum_{i=1}^{\infty} \lambda_i e_i(x) \overline{e_i}, f \right\rangle_{L^2(\Omega, \mu)} \quad (6.74)$$

$$\stackrel{\text{A.3.6}}{=} \langle k(x, \cdot), f \rangle_{L^2(\Omega, \mu)} - \sum_{i=1}^{\infty} \lambda_i \langle \overline{e_i}, f \rangle_{L^2(\Omega, \mu)} e_i(x) \quad (6.75)$$

$$\stackrel{\text{A.3.9}}{=} \int_{\Omega} k(x, y) \overline{f(y)} d\mu(y) - \sum_{i=1}^{\infty} \lambda_i \langle \overline{f}, e_i \rangle_{L^2(\Omega, \mu)} e_i(x) \quad (6.76)$$

$$\stackrel{(6.46)}{=} (T_k \overline{f})(x) - (T_k \overline{f})(x) = 0. \quad (6.77)$$

Da $f \in L^2(\Omega, \mu)$ beliebig war, folgt mit der Definitheit des Skalarproduktes: $k(x, \cdot) - \hat{k}(x, \cdot) = 0$ in $L^2(\Omega, \mu)$ für fast alle $x \in \Omega$, d. h. gemäß Bsp. A.6.17: $k(x, y) = \hat{k}(x, y)$ fast überall auf $\Omega \times \Omega$.

4. Man bemerke, dass sich beide Fälle gleichzeitig behandeln lassen: Im Fall $* = L^\infty(\Omega, \mu)$ ziehe man dazu von Ω diejenige Nullmenge ab, sodass danach (6.49) mit $* = \infty$ anstelle von $* = L^\infty(\Omega, \mu)$ auf (diesem o. B. d. A. gleich benanntem) Ω gilt und $|k|$ darauf durch M beschränkt ist. Für die Hinrichtung folgt dann aus (6.49) für $n, m \in \mathbb{N}$, $m \leq n$:

$$\left\{ \begin{array}{l} \|s_n - s_m\|_\infty \\ \|\tilde{s}_n - \tilde{s}_m\|_\infty \end{array} \right\} = \left\{ \begin{array}{l} \left\| \sum_{i=m}^n |\lambda_i E_i| \right\|_\infty \\ \left\| \sum_{i=m}^n \lambda_i E_i \right\|_\infty \end{array} \right\} \stackrel{\text{A.2.14}}{\leq} \sup_{x, y \in \Omega} \sum_{i=m}^n |\lambda_i e_i(x) \overline{e_i(y)}| \quad (6.78)$$

$$\stackrel{\text{A.2.13}}{\leq} \sup_{x, y \in \Omega} \left(\sum_{i=m}^n |\sqrt{\lambda_i} e_i(x)|^2 \right)^{\frac{1}{2}} \left(\sum_{i=m}^n |\sqrt{\lambda_i} e_i(y)|^2 \right)^{\frac{1}{2}} \quad (6.79)$$

$$\stackrel{\text{A.2.14}}{=} \left\| \sum_{i=m}^n \lambda_i |e_i|^2 \right\|_* \stackrel{\text{(6.49)}}{\xrightarrow{\text{A.1.11}}} 0 \quad (n \rightarrow \infty). \quad (6.80)$$

Folglich sind $(s_n)_{n \in \mathbb{N}}$ und $(\tilde{s}_n)_{n \in \mathbb{N}}$ nun sogar Cauchyfolgen in $\mathcal{F}^b(\Omega \times \Omega)$ und wegen auch dessen Vollständigkeit (siehe Bsp. A.2.9) folgt abermals, dass die beiden Reihen $\sum_{i=1}^\infty |\lambda_i E_i|$ und $\sum_{i=1}^\infty \lambda_i E_i$ gleichmäßig konvergieren auf $\Omega \times \Omega$. Dies ist die Hinrichtung der Behauptung.

Wenn umgekehrt (6.73) gleichmäßig konvergiert, dann konvergiert offensichtlich (6.49) und auch die Rückrichtung ist gezeigt. □

Korollar 6.5.2 (Gleichförmig beschränkte Eigenfunktionen [18, Theorem 3.a.1]).

Es gelten die Bedingungen und Benennungen in Satz 6.5.1, ohne dass bekannt sei, ob die Zusatzbedingung unter Punkt 4. erfüllt ist. Seien stattdessen die $\|\cdot\|_{L^2(\Omega, \mu)}$ -normierten Eigenfunktionen $(e_i)_{i \in I}$ gleichförmig in der $\|\cdot\|_{L^\infty(\Omega, \mu)}$ -Norm beschränkt, d. h. es existiere ein $C > 0$ mit

$$\sup_{i \in I} \|e_i\|_{L^\infty(\Omega, \mu)} \leq C < \infty. \quad (6.81)$$

Dann ist die Zusatzbedingung unter Punkt 4. in Satz 6.5.1 ebenfalls mit dieser Voraussetzung erfüllt und ergo konvergiert $k(x, y) = \sum_{i=1}^\infty \lambda_i e_i(x) \overline{e_i(y)}$ erneut absolut gleichmäßig fast überall.

Beweis. (Die Beweisidee ist [18, Theorem 3.a.1] entnommen.)

Wie im Beweis von Satz 6.5.1 sei o. B. d. A. erneut $I = \mathbb{N}$. Da wegen Punkt 1. in Satz 6.5.1 gilt, dass $\sum_{i=1}^\infty |\lambda_i| < \infty$ ist, so folgt sogleich für alle $n, m \in \mathbb{N}$, $m \leq n$:

$$\left\| \sum_{i=m}^n \lambda_i |e_i|^2 \right\|_{L^\infty(\Omega, \mu)} \stackrel{\text{A.2.4}}{\leq} \sum_{i=m}^n |\lambda_i| \|e_i\|_{L^\infty(\Omega, \mu)}^2 \stackrel{\text{(6.81)}}{\leq} C \sum_{i=m}^n |\lambda_i| \stackrel{\text{A.1.11}}{\rightarrow} 0 \quad (n \rightarrow \infty). \quad (6.82)$$

Folglich ist $(\sum_{i=1}^n \lambda_i |e_i|^2)_{n \in \mathbb{N}}$ eine Cauchyfolge in $L^\infty(\Omega, \mu)$ und gemäß Bsp. A.2.14 ist $\sum_{i=1}^n \lambda_i |e_i(x)|^2$ damit gleichmäßig konvergent für fast alle $x \in \Omega$, d. h. die Zusatzbedingung in Punkt 4. gilt. □

Bemerkung 6.5.3.

1. In der originalen Version von [Satz 6.5.1](#) in [\[18, Theorem 3.a.1\]](#) ist die (äquivalente) Zusatzbedingung unter Punkt 4. nicht enthalten, nämlich dass die (Spektral-)Reihenentwicklung [\(6.73\)](#) von k genau dann sogar gleichmäßig absolut (fast überall) konvergiert, wenn $\sum_{i \in I} \lambda_i |e_i|^2$ (fast überall) gleichmäßig konvergiert. Stattdessen behauptet [\[18, Theorem 3.a.1\]](#), dass aus den restlichen Voraussetzungen bereits folge, dass die $L^2(\Omega, \mu)$ -normierten Eigenfunktionen $(e_i)_{i \in I}$ sogar in der $L^\infty(\Omega, \mu)$ -Norm **gleichförmig beschränkt** seien (d. h. $\sup_{i \in I} \|e_i\|_{L^\infty(\Omega, \mu)} < \infty$ gelte). [\[18, Theorem 3.a.1\]](#) behauptet somit, dass mit den Voraussetzungen von [Satz 6.5.1](#) bereits die stärkeren Bedingungen von [Kor. 6.5.2](#) erfüllt seien und verwendet diese implizit in seinem Nachweis der gleichmäßigen Konvergenz der Reihenentwicklung [\(6.73\)](#) von k (mittels Weierstraßschen Majorantenkriteriums). Zum Beispiel [\[5, 6, 19, Remark 6.3.12, Remark 12\]](#) konnten aber zeigen, dass die von [\[18, Theorem 3.a.1\]](#) behauptete gleichförmige Beschränktheit der $L^2(\Omega, \mu)$ -normierten Eigenfunktionen nicht immer unter den Voraussetzungen von [Satz 6.5.1](#) gilt (siehe 3.).

Beim Versuch, diese durch [\[5, 6, 19, Remark 6.3.12, Remark 12\]](#) aufgezeigte Beweislücke in [\[18, Theorem 3.a.1\]](#) zu schließen, ist dieser Arbeit zweierlei gelungen: Zum einen wurde aus Beweisen von Spezialfällen des Satzes von Mercer durch [\[1, 7, Theorem 6.1.1, Theorem III.22\]](#) eine schwächere Bedingung als die durch [\[18, Theorem 3.a.1\]](#) implizit angenommene gleichförmige Beschränktheit der $L^2(\Omega, \mu)$ -normierten Eigenfunktionen entnommen, mit welcher der Beweis dennoch korrekt ausgeführt werden kann: Diese weitere Voraussetzung „ $\sum_{i \in I} \lambda_i |e_i|^2$ konvergiere (fast überall) gleichmäßig“ sei im Folgenden als **Mercer-Punkt-4-Zusatzbedingung** bezeichnet. Zum anderen vermutete [\[23\]](#) korrekt, dass die Zusatzbedingung äquivalent zur gleichmäßigen Konvergenz der Reihenentwicklung [\(6.73\)](#) von k sei – was diese Arbeit bestätigen konnte.

Interessant wird also die weitere Forschung daran sein, unter welchen Umständen die Reihe $\sum_{i \in I} \lambda_i |e_i|^2$ (fast überall) gleichmäßig konvergiert. Aus Punkt 1. und 2. in [Satz 6.5.1](#) entnimmt man, dass die **Hilbert-Schmidt Integraloperator**-Eigenschaft von T_k bereits hilfreiche Teilergebnisse liefert. Vermag es gar zu sein, dass für Hilbert-Schmidt Integraloperatoren diese Bedingung immer gilt und damit, dass der Satz wie durch [\[18, Theorem 3.a.1\]](#) formuliert – ohne die dann nicht mehr benötigte implizite Zusatzbedingung der gleichförmigen Beschränktheit der $L^2(\Omega, \mu)$ -normierten Eigenfunktionen – doch korrekt aufgestellt ist?

**Frage / Ausblick:**

Gilt die **gleichmäßige Konvergenz** im [Satz 6.5.1](#) von Mercer-König ohne die **Mercer-Punkt-4-Zusatzbedingung** unter den restlichen Voraussetzungen des Satzes?

Aus der Mercer-Punkt-4-Zusatzbedingung und der Endlichkeit des Maßraums (Ω, μ) alleine lässt sich beispielsweise für die **Eigenwerte** $(\lambda_i)_{i \in I}$ von T_k und die jeweils zugehörigen $L^2(\Omega, \mu)$ -normierten Eigenfunktionen $(e_i)_{i \in I}$ direkt schließen, dass

$$\sum_{i \in I} |\lambda_i| \stackrel{\substack{\text{A.3.12} \\ \text{6.2.1}}}{=} \sum_{i \in I} \lambda_i \|e_i\|_{L^2(\Omega, \mu)}^2 \stackrel{\substack{\text{A.2.14} \\ \text{A.6.19}}}{=} \int_{\Omega} \sum_{i \in I} \lambda_i |e_i(x)|^2 d\mu(x) \stackrel{\substack{\text{A.2.14} \\ \text{A.6.15}}}{\leq} \mu(\Omega) \left\| \sum_{i \in I} \lambda_i |e_i|^2 \right\|_{L^\infty(\Omega, \mu)} < \infty. \quad (6.83)$$

Aus Satz 6.5.1 bzw. Kor. C.5.3 erkennt man, dass die absolute Summierbarkeit der Eigenwerte aber bereits ohne die Mercer-Punkt-4-Zusatzbedingung gilt, dass sie aber viel aufwendiger nachzuweisen ist. Ist dies ggf. ein Hinweis darauf, dass die folgende Frage bejaht werden kann?



Fragen / Ausblick:

Gibt es eine **einfacher zu prüfende Zusatzbedingung**, welche bei **Hilbert-Schmidt Integraloperatoren** die **Mercer-Punkt-4-Zusatzbedingung** erzwingt und damit äquivalent die **gleichmäßige Konvergenz** in (6.48)?

2. James Mercer bewies 1909 die erste Version des nach ihm benannten Satzes in [20, § 29]. Dabei zeigte er einen Spezialfall der in dieser Arbeit zitierten allgemeineren Variante, denn der Raum $\Omega = [a, b]$ ist darin als **abgeschlossenes Intervall** mit dem **Lebesgue-Maß** ausgestattet und der Integralkern $k \in C([a, b]^2)$ als **stetig** (Def.+Satz A.1.12) angenommen. Mit dem Satz von Dini (Satz A.1.16) gelingt es [20, § 29], die **gleichmäßige Konvergenz** in (6.48) unter diesen spezielleren Bedingungen zu zeigen ohne die Mercer-Punkt-4-Zusatzbedingung vorauszusetzen.
3. Das durch [5, 6, Remark 6.3.12, Remark 12] konstruierte Gegenbeispiel zu [18, Theorem 3.a.1], gelingt ihnen durch die Betrachtung eines **unendlichen, metrischen Graphen mit endlichem Volumen** (einer unendlichen Ansammlung schrumpfender **Dirichlet-Intervalle**, streng genommen eines sog. **Quantengraphen**) auf dem sie den **Laplace-Operator** bzw. die **Wärmeleitungsgleichung** wirken lassen. [5, Remark 6.3.12] zeigt, dass ein solcher einen **endlichen Maßraum** darstellt und der Wärmeleitungsoperator darauf sowohl positiv ist, als auch einen **fast überall beschränkten Integralkern** besitzt, d.h. die Voraussetzungen von [18, Theorem 3.a.1] erfüllt. Besitzen die Kantenlängen des Graphen aber das Infimum Null, so existiert $\sup_{i \in I} \|e_i\|_{L^\infty} = \infty$ gemäß [5, Remark 6.3.12] nicht, was [18, Theorem 3.a.1] widerlegt. Interessanterweise zeigt jedoch ebenfalls [5, Theorem 3.3.1, Remark 3.3.2], dass für den so konstruierten, unendlichen Graphen ein Spezialfall des Satzes von Mercer durch [10, Theorem 2.1.4] gilt, welcher in dieser Arbeit als Satz 6.7.11 vorgestellt wird. Obgleich die Eigenfunktionen anders als durch [18, Theorem 3.a.1] behauptet in der $\|\cdot\|_{L^\infty}$ -Norm keine gemeinsame Schranke besitzen, gilt die **gleichmäßige Konvergenz** von (6.48) in diesem Gegenbeispiel also trotzdem.

Genauso verhält es sich bei dem Gegenbeispiel, welches [19] konstruieren. In diesem handelt es sich um einen **stetigen Integralkern** auf dem **Intervall** $[0, 1]$. Auch dieses Beispiel erfüllt sowohl einen weiteren Spezialfall des Satzes von Mercer durch [1, 7, Theorem 6.1.1, Theorem III.22], welcher in Satz 6.7.3 vorgestellt wird, als auch den unter 2. beschriebenen originalen Satz von Mercer selbst. Folglich besitzen erneut die $\|\cdot\|_{L^2([0,1])}$ -normierten Eigenfunktionen in der $\|\cdot\|_{L^\infty([0,1])}$ -Norm keine gemeinsame Schranke, aber die gleichmäßige Konvergenz von (6.48) gilt trotzdem.

Schließlich stellen [33, Problem 2.24] selbst ein Gegenbeispiel vor, basierend auf der Menge $\mathbb{N}$ der natürlichen Zahlen und einem endlichen Zählmaß. Aufgrund der Einfachheit des Beispiels konvergiert die Darstellung (6.48) des Integralkerns in diesem trivialerweise auch gleichmäßig.

6.6 Konsequenzen für das Maschinelle Lernen

Mit diesem sehr theoretischen und allgemeinen mathematischen Ergebnis im Gepäck kehrt man zum Maschinellen Lernen zurück: Der **Satz 6.5.1 von Mercer** wird nämlich durch [33, S. 37 f.] dazu verwendet – und darin insbesondere die gleichmäßige Konvergenz von (6.48) – um eine endlich-dimensionale Approximation eines mithilfe des Kernel-Tricks gefundenen Hilbertraums zu gewinnen.

Dies geschieht auf folgende Art und Weise: [33, S. 37] vergleichen die Beziehung (3.8) in Bezug auf einen RKHS $\mathcal{H}_k$ und seinem RK k mit dem Ergebnis (6.48) von **Satz 6.5.1**:

$$k(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}_k} \quad \Leftrightarrow \quad k(x, y) = \sum_{i=1}^{\infty} \lambda_i e_i(x) \overline{e_i(y)} \quad \text{für fast alle } x, y \in X = \Omega. \quad (6.84)$$

Um auf der rechten Seite ebenfalls ein Skalarprodukt zu erhalten, interpretieren [33, S. 37] diese als ein solches im $\ell^2(\mathbb{N})$, denn dort hätte es die Form $\langle x, y \rangle_{\ell^2(\mathbb{N})} = \sum_{i=1}^{\infty} x_i \overline{y_i}$. Φ ergibt sich dann als logische Wahl mit $\Phi(x) := (\sqrt{\lambda_i} e_i(x))_{i \in \mathbb{N}}$ für $x \in X$ (was wegen der positiven $(\lambda_i)_{i \in I}$ (6.2.1) wohldefiniert ist) mit resultierendem Hilbertraum $\mathcal{H} := \ell^2(\mathbb{N})$ sowie $k(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}$ für fast alle $x, y \in X = \Omega$ genau so wie im Fall von RKHS.

Anstelle eines oftmals unbekannten Funktionenraums im Falle eines RKHS, erhält man bei obigem Vorgehen mit dem **Satz 6.5.1 von Mercer** somit den ggf. viel handlicheren (separablen) Hilbertraum $\mathcal{H} = \ell^2(\mathbb{N})$. Zudem machen sich [32, 33, S. 28, Prop. 2.11] ebenfalls noch die gleichmäßige Konvergenz in (6.48) zu Nutze: Diese besagt, dass für jedes $\varepsilon > 0$ ein $n \in \mathbb{N}$ existiert, sodass gilt:

$$\left| k(x, y) - \sum_{i=1}^n \lambda_i e_i(x) \overline{e_i(y)} \right| \leq \varepsilon \quad \text{für fast alle } x, y \in X = \Omega. \quad (6.85)$$

In gleicher Weise wie für $\mathcal{H} = \ell^2(\mathbb{N})$ interpretieren [32, 33, S. 28, Prop. 2.11] die endliche Summe in (6.85) erneut als Skalarprodukt in $\mathcal{H}_n = \mathbb{K}^n$ und setzen als Abbildung $\Phi_n : X \rightarrow \mathcal{H}_n$, $\Phi_n(x) = (\sqrt{\lambda_i} e_i(x))_{i=1}^n$ für $x \in X = \Omega$. Somit lässt sich das Skalarprodukt im **endlich-dimensionalen** $\mathcal{H}_n$ bis auf einen Fehler von ε durch k approximieren:

$$\left| k(x, y) - \langle \Phi_n(x), \Phi_n(y) \rangle_{\mathcal{H}_n} \right| \leq \varepsilon \quad \text{für fast alle } x, y \in X = \Omega. \quad (6.86)$$

Vor allem in der **Principle Component Analysis** kann von diesem Ergebnis Gebrauch gemacht werden (vgl. [33, Chapter 14]). Man ist in diesem Anwendungsfall daran interessiert, welche „Dimensionen“ der Rohdatenmenge für ein gegebenes maschinelles Lernproblem von besonderer Relevanz sind. Nun ist die Rohdatenmenge $X = \Omega$ in der Regel aber „nur“ eine gewöhnliche Menge ohne bekannte Vektorraumstruktur (ergo Dimensionen). Aus (6.85) und (6.86) liest man trotzdem ab, dass nur endlich viele Eigenwerte und Eigenfunktionen – ergo Eigenräume in $\Phi(X) \subseteq \mathcal{H} = \ell^2(\mathbb{N})$ dafür verantwortlich sind, den Wert des Skalarprodukts im gefundenen Hilbertraum signifikant zu beeinflussen. Diese Eigenräume scheinen also den für das maschinelle Lernproblem relevanten Informationsgehalt der Rohdatenmenge $X = \Omega$ zu tragen, denn die übrigen beeinflussen das Ergebnis des Algorithmus nicht mehr bedeutend. Hat man den **Kernel-Trick** also auf Funktionen angewandt,

welche die Voraussetzungen von [Satz 6.5.1](#) erfüllen und eine solche gefunden, sodass der Algorithmus stabile Ergebnisse liefert, folgert man dann mittels [\(6.85\)](#) die „relevanten Dimensionen“ in $X = \Omega$.

An dieser Stelle wird auch deutlich, warum [\[33, Theorem 2.10\]](#) die Version des **Satzes von Mercer** in [\[18, Theorem 3.a.1\]](#) präferieren. Dieser stellt an die Rohdatenmenge $X = \Omega$ lediglich die Bedingung, ein endlicher Maßraum zu sein. Das ist zwar bereits eine stärkere Voraussetzung als im [Satz 4.2.1 von Moore](#), welcher mit einer einfachen Menge ohne weitere Zusatzbedingungen auskommt. In der Praxis des Maschinellen Lernens ist es aber möglich mit statistischen Methoden ein Wahrscheinlichkeitsmaß auf der Rohdatenmenge X zu finden und damit die Voraussetzungen von [Satz 6.5.1](#) zu erfüllen.

In der Version des Satzes durch Mercer [\[20, § 29\]](#) selbst aus 1909 ist die Menge der möglichen Rohdatenmengen $X = \Omega = [a, b]$ viel eingeschränkter. Insbesondere muss für die zusätzlich notwendige Stetigkeit des Integralkerns k in Mercers originalem Satz genau die **kanonische Metrik / Norm / Topologie** des $(\mathbb{R}, |\cdot|)$ auf der Rohdatenmenge $X = \Omega = [a, b]$ gelten. Im Maschinellen Lernen ist die Topologie der Rohdatenmenge, aber gerade oftmals gänzlich unbekannt bzw. erst das gewünschte Ergebnis des **Kernel-Tricks**. Der originale Satz von Mercer ist also nur dann hilfreich, wenn sich die Rohdatenmenge zufällig durch ein Intervall $[a, b]$ abbilden lässt.

Leider übersahen [\[33, Theorem 2.10\]](#) anscheinend in ihrer Argumentation aber die fehlende und doch verwendete (in [Bem. 6.5.3](#) diskutierte) Zusatzbedingung im Beweis von Mercer in [\[18, Theorem 3.a.1\]](#), dass die $\|\cdot\|_{L^2(\Omega, \mu)}$ -normierten Eigenfunktionen ebenfalls **gleichförmig** in der $\|\cdot\|_{L^\infty(\Omega, \mu)}$ -Norm beschränkt sein müssen bzw., wie in dieser Arbeit abgeschwächt, die zur gleichmäßigen Konvergenz äquivalente **Mercer-Punkt-4-Zusatzbedingung** gelten muss. Das Resultat von [\[33, S. 37 f.\]](#) ist davon nur in Teilen betroffen: Die gleichmäßige Konvergenz der Reihenentwicklung [\(6.48\)](#) wird von [\[33, Prop. 2.11\]](#) nur für die Approximation des gefundenen Hilbertraums $\mathcal{H} = \ell^2(\mathbb{N})$ durch einen **endlich-dimensionalen** Hilbertraum $\mathcal{H}_n = \mathbb{K}^n$ verwendet. Unberührt bleibt, dass man anstelle eines Funktionenraums/RKHS den ggf. handlicheren Hilbertraum $\mathcal{H} = \ell^2(\mathbb{N})$ erhält. Nichtsdestotrotz muss in Anbetracht von [Bem. 6.5.3](#) in Frage gestellt werden, ob die endlich-dimensionale Approximation des Hilbertraums/Skalarprodukts in der Praxis überhaupt noch immer bzw. in den meisten Anwendungsfällen möglich ist.

Fragen / Ausblick:



- Wie sehr schränkt die **Mercer-Punkt-4-Zusatzbedingung** die Integral-kernfunktionen k in der Praxis des Maschinellen Lernens effektiv ein?
- Unter welchen Bedingungen kommt es im [Satz 6.5.1](#) von Mercer überhaupt zu Fällen, in denen die **Mercer-Punkt-4-Zusatzbedingung** nicht gilt?
- Gibt es geeignete Funktionsklassen k , welche alle Bedingungen in [Satz 6.5.1](#) inkl. der **Mercer-Punkt-4-Zusatzbedingung** erfüllen, mit denen im Maschinellen Lernen der Kernel-Trick angewandt werden kann?

Vermutlich erschwert die **Mercer-Punkt-4-Zusatzbedingung** im [Satz 6.5.1](#) von Mercer-König dessen Anwendbarkeit im Maschinellen Lernen doch um einiges: Für den **Kernel-Trick** möchte der Entwickler/Anwender eigentlich eine große Funktionenklasse experimentell auf seinen Algorithmus

„loslassen“, ohne im Vorhinein viel darüber wissen zu müssen. Die Positivitätseigenschaften (4.6) bzw. (6.47) zu erfüllen, ist bereits eine nicht unerhebliche Einschränkung. Muss zusätzlich noch geprüft werden, dass für jedes k die Reihe $\sum_{i=1}^{\infty} \lambda_i |e_i|^2$ gleichmäßig fast überall konvergiert, so müssen die Eigenwerte und Eigenfunktionen des k zugehörigen Hilbert-Schmidt Integraloperators noch vor Anwendung des Kernel-Tricks bekannt sein. Satz 6.5.1 mag also praktisch nicht hilfreich sein.

6.6.1 Hypothese zum Satz von Mercer im Maschinellen Lernen

Darum sei als Einschub eine Hypothese zur praktischen Anwendung im Maschinellen Lernen erörtert:



Hypothese:

Lässt sich der **Satz 6.5.1 von Mercer** im Maschinellen Lernen „rückwärts“ lesen/anwenden?

Die Ausführungen in [33, Chapter 1, Remark 2.8, Chapter 13] lesen sich so, als ob im Maschinellen Lernen in der Regel zuerst eine geeignete Funktionsklasse gesucht wird, deren Elemente die Voraussetzungen der **Sätze von Moore bzw. Mercer** (Satz 4.2.1 / Satz 6.5.1) erfüllen, um dann z. B. den **Kernel-Trick** darauf anzuwenden. Dazu stellt [33, Chapter 13] eine Vielzahl an Techniken vor, welche von problembezogen bis hin zu allgemein konstruktiv reichen, aber anscheinend eint, dass sie ausgehend von bereits bekannten Beispielfunktionen starten (z. B. **Gaußsche Kerne** in Bsp. 5.1.1).

Diese Arbeit stellt nun die Frage, ob man ebenfalls zu geeigneten Integralkernen gelangt, wenn man vom Ergebnis von **Satz 6.5.1** aus startet, ohne zuvor eine spezielle Beispielfunktion anzunehmen.



Hypothese:

Handelt es sich bei

$$k(x, y) := \sum_{i=1}^{\infty} \lambda_i e_i(x) \overline{e_i(y)} \quad \text{für } x, y \in X = \Omega \quad (6.87)$$

ggf. um eine geeignete **Konstruktionsvorschrift** für (Integral-)Kernfunktionen?

Statt also mit einer geeigneten Funktion k den **Satz 6.5.1** von vorne nach hinten anzuwenden, beginne man mit dessen Ergebnis und konstruiere sich k zuerst damit: Betrachtet man beispielsweise einen Maßraum (Ω, μ) , für den der zugehörige $L^2(\Omega, \mu)$ **separabel** ist, dann existiert gemäß **Def.+Satz A.3.13** (äquivalenterweise) eine abzählbar unendlich große **Orthonormalbasis** $(e_i)_{i \in \mathbb{N}}$ in Letzterem. Dann könnte die Konstruktionsvorschrift lauten, all die $\lambda := (\lambda_i)_{i \in \mathbb{N}} \subseteq c_0(\mathbb{N})$ (**Bsp. A.2.12**) mit $\lambda_i \geq 0$ für alle $i \in \mathbb{N}$ dergleichen zu finden, dass der Ausdruck (6.87) gleichmäßig (fast überall) auf $\Omega \times \Omega$ konvergiert. Ist dies der Fall, so ließe sich die so definierte Funktion k_λ auf identische Weise wie am Anfang von **Abschnitt 6.6** als Skalarprodukt des $\mathcal{H}_\lambda := \ell^2(\mathbb{N})$ interpretieren. Die Abbildung Φ_λ lautet dann erneut $\Phi_\lambda : X \rightarrow \mathcal{H}_\lambda$, $\Phi_\lambda(x) = (\sqrt{\lambda_i} e_i)_{i \in \mathbb{N}}$, sodass am Ende gilt: $k_\lambda(x, y) = \langle \Phi_\lambda(x), \Phi_\lambda(y) \rangle_{\mathcal{H}_\lambda}$. Wegen der konstruktionsbedingten gleichmäßigen Konvergenz in (6.87) ließe sich zudem ebenfalls die endlich-dimensionale Approximation von $\mathcal{H}_\lambda$ wie zu Beginn von **Abschnitt 6.6** erreichen.

Mit dieser Konstruktion muss der Entwickler im Maschinellen Lernen die Positivitätsbedingung (6.47) nicht im Vorhinein prüfen und kann die Mercer-Punkt-4-Zusatzbedingung ignorieren (wobei letztere wegen der gleichmäßigen Konvergenz in (6.87) automatisch erfüllt ist). Zudem ist die Vermutung des Autors, dass sich so **ziemlich alle möglichen Funktionen** k_λ , die auf obige Weise als **Skalarprodukt** interpretierbar sind, durch ebendiese Vorschrift konstruieren lassen: Da im Falle eines separablen $L^2(\Omega, \mu)$ die abgeschlossene lineare Hülle der Orthonormalbasis $(e_i)_{i \in \mathbb{N}}$ definitionsgemäß dicht in Ersterem liegt, sollte die Konstruktionsvorschrift (6.87) den größten Teil der Teilmenge im $L^2(\Omega^2, \mu^2)$ abdecken, deren Elemente k_λ ein Skalarprodukt im $\ell^2(\mathbb{N})$ induzieren.

Je nachdem wie „gutartig“ das Orthonormalsystem $(e_i)_{i \in \mathbb{N}}$ ist, lässt sich ggf. gar eine ganze Teilmenge $L \subseteq c_0(\mathbb{N})$ finden, für deren Elemente $\lambda := (\lambda_i)_{i \in \mathbb{N}} \in L$ der Ausdruck (6.87) immer gleichmäßig konvergiert. Anstelle also den Kernel-Trick auf eine etwaig relativ unbekannte Funktionsklasse anzuwenden, ließe er sich mittels (6.87) dann auf die Teilmenge L des besser bekannten $c_0(\mathbb{N})$ anwenden. Obige Vermutung besagt, dass somit eine viel **umfangreichere Funktionsklasse** $(k_\lambda)_{\lambda \in L}$ mit dem Kernel-Trick getestet werden kann und dass diese zudem **einfacher konstruierbar** ist, als dass zunächst alle Bedingungen der Sätze von Moore bzw. Mercer geprüft werden müssen.

Die Arbeit möchte an dieser Stelle nicht behaupten, Erfinderin obiger (nicht besonders komplizierter) Technik zu sein. In [33, Prop. 4.10] beispielsweise ist die hier beschriebene Idee bereits in Grundzügen zu erahnen. Die explizite Artikulation der Konstruktionsvorschrift (6.87) konnte in der Literatur zwar nicht gefunden werden, aufgrund deren Umfangs ist aber nicht auszuschließen, dass die Idee bereits bekannt ist und aktiv verwendet wird.

Zudem ist unbedingt auszuführen, dass obige Hypothese erst aufkeimte, als [23] die Äquivalenz zwischen der Mercer-Punkt-4-Zusatzbedingung und der gleichmäßigen Konvergenz in (6.73) vermutete: Mit seiner Intuition, die **Rückrichtung** in Punkt 4. von Satz 6.5.1 zu prüfen, entstand bei gleichzeitiger Betrachtung von [33, Problem 2.24] die Idee, Letzteren genauso **rückwärts** zu lesen.

Als interessante Folgefrage ergibt sich, gemeinsam mit Anwendern von Kern-Methoden zu erörtern, inwiefern obige Konstruktionsidee **Relevanz in der Praxis** findet oder finden kann. Es ist durchaus denkbar, dass die mittels (6.87) konstruierte Funktionsklasse zu groß ist, um im Maschinellen Lernen praktisch einsetzbar zu sein. Möglicherweise ist es doch notwendig, erst eine kleinere, als geeignet bekannte Klasse gezielt für das jeweilige Problem auszuwählen, damit das Kernel-Trick-Experiment überhaupt berechenbar bleibt. Auch die Ressourcen eines Computers sind begrenzt und ggf. ist selbst eine Teilmenge des $c_0(\mathbb{N})$ noch zu groß, um darin mittels (6.87) nach einer geeigneten (Integral-)Kernfunktion zu suchen. Vorstellbar wäre z. B., falls sich eine endliche Teilmenge des Orthonormalsystems $(e_i)_{i \in \mathbb{N}}$ finden lässt, welche dicht „genug“ in $L^2(\Omega, \mu)$ liegt, dass ggf. **evolutionäre Algorithmen** (siehe z. B. [9]) diesem Komplexitätsproblem Abhilfe schaffen können. Auch diese Hypothese müsste erst geprüft werden.

Da die Arbeit an dieser Stelle nicht beantworten kann, wie sinnvoll die Konstruktion von Test-Funktionsklassen mittels (6.87) in der Praxis des Maschinellen Lernens tatsächlich ist, seien im Folgenden noch 2 interessante Spezialfälle des Satzes 6.5.1 (aber allgemeiner als die durch Mercer selbst in 1909 aufgestellte Version) vorgestellt. Diese kommen ohne die Mercer-Punkt-4-Zusatzbedingung aus und ermöglichen damit eine einfachere Anwendung des Satzes „von vorne“.

6.7 Spezialfälle des Satzes von Mercer

Für den Beweis des ersten Spezialfalls durch [1, 7, Theorem 6.1.1, Theorem III.22] ist ein Hilfsresultat notwendig, welches in interessanter Weise Kapitel 3 und Kapitel 4 mit diesem Kapitel 6 verbindet:

6.7.1 Stetige Reproduzierende Kerne (RK) auf kompaktem Raum

Lemma 6.7.1 (Stetige reproduzierende Kerne auf Kompaktum [1, 7, Lemma 6.1.3, Lemma III.24]). Sei $\Omega \subseteq \mathbb{R}^d$ kompakt, $k \in C(\Omega \times \Omega)$ stetig und es gelte für alle $f \in L^2(\Omega)$:

$$\int_{\Omega} \int_{\Omega} k(x, y) f(x) \overline{f(y)} \, dx \, dy \geq 0. \quad (6.88)$$

Dann ist k ein reproduzierender Kern (d. h. positiv semidefinit mit Kor. 4.2.2).

Beweis. (Die Beweisideen sind [1, 7, Lemma 6.1.3, Lemma III.24] entnommen.)

Wegen der Kompaktheit von Ω , ist es (mit dem Lebesgue-Maß ausgestattet) ein endlicher Maßraum und zudem mit Prop. A.1.15 das stetige k auf Ω beschränkt. Beides zusammen ergibt, dass $k \in L^2(\Omega^2, \mu^2)$ ist, d. h. gemäß Def.+Satz A.4.18 ist der Operator

$$T_k : L^2(\Omega) \rightarrow L^2(\Omega), \quad (T_k f)(x) := \int_{\Omega} k(x, y) f(y) \, dy \quad \text{für } f \in L^2(\Omega), \, x \in \Omega \quad (6.89)$$

ein wohldefinierter Hilbert-Schmidt Integraloperator. Zudem ist er positiv (Satz 6.5.1) mit (6.88):

$$\langle T_k f, f \rangle_{L^2(\Omega)} \stackrel{\text{A.3.9}}{=} \int_{\Omega} \int_{\Omega} k(x, y) f(x) \overline{f(y)} \, dy \, dx \stackrel{6.88}{\geq} 0. \quad (6.90)$$

Damit sind alle Voraussetzungen von Satz 6.5.1 erfüllt und dieser ergibt $k(x, y) = \sum_{i \in I} \lambda_i e_i(x) \overline{e_i(y)}$ für fast alle $x, y \in \Omega$, wobei $(\lambda_i)_{i \in I}$ positive Eigenwerte von T_k und $(e_i)_{i \in I}$ zugehörige Eigenfunktionen sind. Es folgt nun direkt für fast alle $x, y \in \Omega$ (man beachte $0 < \lambda_i = \bar{\lambda}_i \in \mathbb{R}$ für alle $i \in I$):

$$k(x, y) = \sum_{i \in I} \lambda_i e_i(x) \overline{e_i(y)} \stackrel{\text{(A.8.1)}}{=} \overline{\sum_{i \in I} \lambda_i e_i(y) \overline{e_i(x)}} = \overline{k(y, x)}. \quad (6.91)$$

Wegen der Stetigkeit von k muss somit $k(x, y) = \overline{k(y, x)}$ sogar für alle $x, y \in \Omega$ gelten (da anderenfalls eine offene nicht-leere, ergo Nicht-Nullmenge $M \subseteq \Omega$ existieren würde mit $k(x, y) \neq \overline{k(y, x)}$ für alle $x, y \in M$, was ein Widerspruch zu (6.91) darstellt) und die symmetrische Bedingung (4.5) ist erfüllt.

Um auch die Positivitätsbedingung (4.6) in Def. 4.1.3 nachzuweisen, sei zunächst angemerkt, dass wenn die Funktion k auf ganz $\mathbb{R}^d$ durch Null fortgesetzt wird, (6.88) ebenfalls gilt, wenn man darin Ω durch $\mathbb{R}^d$ ersetzt. Im Folgenden sei mit k diese Fortsetzung bezeichnet, wobei weiterhin nur bekannt sei, dass diese auf $\Omega \times \Omega$ stetig ist.

Nun wähle man $n \in \mathbb{N}$ mit zugehörigen $\{x_1, \dots, x_n\} \subseteq \Omega^\circ$ und $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ beliebig aber fest (und bemerke, dass $x_1, \dots, x_n$ zunächst nur aus dem Inneren (Def. A.1.5) von Ω gewählt sind). Nimmt man dann einen beliebigen Glättungskern $J : \mathbb{R}^d \rightarrow \mathbb{R}$ (Def. A.7.1/Bsp. A.7.2), so folgt, dass die Funktion definiert durch $\tilde{J} : \mathbb{R}^{d+d} \rightarrow \mathbb{R}$, $\tilde{J}(x, y) := J(x)J(y)$ für $x, y \in \mathbb{R}^d$ ebenfalls ein Glättungskern ist:

- (i) Da gemäß Def. A.7.1 $J(x) \geq 0$ für $x \in \mathbb{R}^d$ gilt, ist auch $\tilde{J}(x, y) = J(x)J(y) \geq 0$ für $(x, y) \in \mathbb{R}^{d+d}$.
- (ii) Da gemäß Def. A.7.1 $J(x) = 0$ für $\|x\| > 1$ gilt, ist auch $\tilde{J}(x, y) = J(x)J(y) = 0$ für $\|(x, y)\| := \max\{\|x\|, \|y\|\} > 1$.
- (iii) Es gilt zudem erneut mit Def. A.7.1 zusammen mit dem Satz A.6.20 von Fubini für positive, messbare Funktionen ($\tilde{J}$ ist als Produkt stetiger Funktionen insbesondere stetig und damit gemäß Kor. A.6.12 messbar):

$$\int_{\mathbb{R}^{d+d}} \tilde{J}(x, y) d(x, y) \stackrel{\text{A.6.20}}{=} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} J(x)J(y) dx dy = \int_{\mathbb{R}^d} 1 \cdot J(y) dy = 1. \quad (6.92)$$

Somit sind alle Bedingungen aus Def. A.7.1 erfüllt und $\tilde{J}$ ist wie behauptet ein Glättungskern auf $\mathbb{R}^{d+d}$. Zudem gilt mit der Bezeichnung $\tilde{J}_\varepsilon$ in Def. A.7.1 für alle $x, y \in \mathbb{R}^d$:

$$\tilde{J}_\varepsilon(x, y) \stackrel{(\text{A.44})}{=} \varepsilon^{-(d+d)} \tilde{J}(\varepsilon^{-1}(x, y)) = \varepsilon^{-d} J_\varepsilon(\varepsilon^{-1}x) \varepsilon^{-d} J_\varepsilon(\varepsilon^{-1}y) \stackrel{(\text{A.44})}{=} J_\varepsilon(x) J_\varepsilon(y). \quad (6.93)$$

Betrachtet man nun für $\varepsilon > 0$ die Funktionenschar

$$g_\varepsilon : \Omega \rightarrow \mathbb{K}, \quad g_\varepsilon(x) := \sum_{i=1}^n \alpha_i J_\varepsilon(x_i - x) \quad \text{für } x \in \Omega, \quad (6.94)$$

so ist g_ε für jedes $\varepsilon > 0$ als Linearkombination stetiger Funktionen ebenso stetig und damit insbesondere ein Element von $L^2(\Omega)$, da Ω als kompakt vorausgesetzt ist. Genauso folgert man, dass $\sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} k$ als Linearkombination stetiger Funktionen ebenfalls stetig ist. Es folgt zusammen mit Lemma A.7.3 und dem Satz A.6.21 von Fubini für integrierbare Funktionen (welcher anwendbar ist, da im Integranden von (6.96) eine stetige Funktion steht, welche den kompakten Träger von k teilt, ergo gemäß Prop. A.1.15 einen beschränkten Betrag besitzt, d. h. das Integral von letzterem auf dem kompakten Träger endlich und somit f integrierbar ist):

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} k(x_i, x_j) \stackrel{\text{A.7.3}}{=} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} \left(\lim_{\varepsilon \rightarrow 0} \tilde{J}_\varepsilon * k \right) (x_i, x_j) \quad (6.95)$$

$$\stackrel{\text{A.7.1}}{=} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} \lim_{\varepsilon \rightarrow 0} \int_{\mathbb{R}^{d+d}} \tilde{J}_\varepsilon(x_i - x, x_j - y) k(x, y) d(x, y) \quad (6.96)$$

$$\stackrel{\text{A.6.21}}{\stackrel{(\text{6.93})}{=}} \lim_{\varepsilon \rightarrow 0} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} J_\varepsilon(x_i - x) J_\varepsilon(x_j - y) k(x, y) dx dy \quad (6.97)$$

$$\stackrel{\text{A.7.1}}{=} \lim_{\varepsilon \rightarrow 0} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} k(x, y) \left(\sum_{i=1}^n \alpha_i J_\varepsilon(x_i - x) \right) \left(\sum_{j=1}^n \overline{\alpha_j} J_\varepsilon(x_j - y) \right) dx dy \quad (6.98)$$

$$\stackrel{(\text{6.94})}{=} \lim_{\varepsilon \rightarrow 0} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} k(x, y) g_\varepsilon(x) \overline{g_\varepsilon(y)} dx dy \stackrel{(\text{6.88})}{\geq} \lim_{\varepsilon \rightarrow 0} 0 = 0. \quad (6.99)$$

Damit ist die Positivitätsbedingung (4.6) für das Innere der Menge Ω gezeigt. Dass sie ebenfalls auf der ganzen Menge Ω gilt, folgt nun zusammen mit diesem Vorergebnis aus der Stetigkeit von k . Man

geht in zwei Schritten vor: Zuerst zeigt man, dass das Integral in (6.88) immer ein reelles Ergebnis hat. Auf identische Weise zeigt man dann, dass es ein nicht-negatives Ergebnis hat:

Man nehme für einen Widerspruchsbeweis an, dass ein $n \in \mathbb{N}$ existiert mit $\{x_1, \dots, x_n\} \subset \Omega$ und $\{\alpha_1, \dots, \alpha_n\} \subset \mathbb{K}$, sodass $\sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} k(x_i, x_j) \notin \mathbb{R}$. Dann lässt sich die Funktion $h : \Omega^n \rightarrow \mathbb{K}$ betrachten, definiert durch $h(y_1, \dots, y_n) := \sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} k(y_i, y_j)$ für $(y_1, \dots, y_n) \in \Omega^n$. h ist als Linearkombination von stetigen Funktionen ebenfalls stetig und mit $x := (x_1, \dots, x_n) \in \Omega^n$ ist ein Punkt gefunden, sodass $h(x) \notin \mathbb{R}$ gilt. Da h stetig ist, gibt es gemäß Def.+Satz A.1.12 damit eine offene Kugel U mit Zentrum x , sodass $h(y) \notin \mathbb{R}$ gilt für alle $y \in U \subseteq \Omega$. Als offene Menge in Ω^n enthält U aber ebenfalls einen Schnittpunkt z mit $(\Omega^n)^o$. Für diesen gilt dann sowohl $h(z) \notin \mathbb{R}$, als auch mit dem oben Gezeigten $h(z) = 0 \in \mathbb{R}$ (da $(\Omega^n)^o = (\Omega^o)^n$ gilt). Dies ist ein Widerspruch $\neq$, weshalb die Annahme falsch sein muss und es folgt: Für alle $n \in \mathbb{N}$, $\{x_1, \dots, x_n\} \subset \Omega$ und $\{\alpha_1, \dots, \alpha_n\} \subset \mathbb{K}$ gilt:

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} k(x_i, x_j) \in \mathbb{R}. \quad (6.100)$$

Erneut durch Widerspruchsbeweis mit der Annahme, dass ein $n \in \mathbb{N}$ existiert mit zugehörigen $\{x_1, \dots, x_n\} \subset \Omega$ und $\{\alpha_1, \dots, \alpha_n\} \subset \mathbb{K}$, sodass $\sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} k(x_i, x_j) < 0$ gilt, zeigt man auf absolut identische Weise (indem man in der Argumentation oben „ $\notin \mathbb{R}$ “ durch „ < 0 “ und „ $\in \mathbb{R}$ “ durch „ ≥ 0 “ ersetzt), dass die gewünschte Aussage insgesamt stimmt: Für alle $n \in \mathbb{N}$, $\{x_1, \dots, x_n\} \subset \Omega$ und $\{\alpha_1, \dots, \alpha_n\} \subset \mathbb{K}$ gilt ebenfalls die Positivitätsbedingung (4.6) aus Def. 4.1.3:

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} k(x_i, x_j) \geq 0. \quad (6.101)$$

Def. 4.1.3 ist somit insgesamt erfüllt, d.h. k wie behauptet eine positiv semidefinite Funktion (und gemäß Kor. 4.2.2 ebenfalls ein reproduzierender Kern). $\square$

Bemerkung 6.7.2.

1. Gemäß [7, Lemma III.24] gilt Lemma 6.7.1 im übrigen Wortlaut ebenfalls für ganz allgemeine **kompakte metrische Räume** (X, d) , welche mit einem **regulären Borelmaß** μ ausgestattet sind, sodass für alle nicht-leeren offenen $U \subseteq X$ gilt, dass $\mu(U) > 0$ ist.

Noch interessanter ist, dass abermals gemäß [7, Lemma III.24] (sogar in dem gerade beschriebenen allgemeineren Kontext) die **Rückrichtung** von Lemma 6.7.1 ebenfalls gilt. Da der Beweis sehr technischer Natur ist, sei an dieser Stelle nur eine kurze Beweisskizze gegeben: Man lässt sich durch die Riemannschen Ober- und Untersummen inspirieren und approximiert das Integral unter (6.88) mittels **verallgemeinerter Treppenfunktionen**:

$$\int_{\Omega} \int_{\Omega} k(x, y) f(x) \overline{f(y)} dx dy \approx \sum_{i=1}^n \sum_{j=1}^n k(x_i, x_j) f(x_i) \overline{f(x_j)} c_i \overline{c_j} \quad (6.102)$$

Die rechte Seite der Approximation ist dann gerade von der Form, dass die positive Semidefinitheit von k die Nicht-Negativität der linken Seite liefert. Der Beweis wird technischer, wenn man berücksichtigt, dass sowohl k als auch f komplexwertig sein können.

2. [Lemma 6.7.1](#) und insbesondere die Tatsache, dass die Rückrichtung gemäß [\[7, Lemma III.24\]](#) ebenfalls gilt, haben eine interessante Folge für die Anwendung im Maschinellen Lernen: Ist über den Rohdatenraum $X = \Omega$ bekannt, dass er die Voraussetzungen von [Lemma 6.7.1](#) erfüllt, dann sind die stetigen reproduzierenden Kerne k für die Anwendung im [Satz 4.2.1 von Moore](#) genau diejenigen, die für den [Satz 6.5.1 von Mercer](#) geeignet sind und umgekehrt. Beide Sätze liefern dabei jeweils einen Hilbertraum $\mathcal{H}, \tilde{\mathcal{H}}$, in dem für das Skalarprodukt gilt, dass $\langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}} = k(x, y) = \langle \tilde{\Phi}(x), \tilde{\Phi}(y) \rangle_{\tilde{\mathcal{H}}}$ für alle $x, y \in X = \Omega$ ist. Weder die Hilberträume $\mathcal{H}$ und $\tilde{\mathcal{H}}$, noch die Abbildungen $\Phi : X \rightarrow \mathcal{H}$ und $\tilde{\Phi} : X \rightarrow \tilde{\mathcal{H}}$ müssen dabei übereinstimmen, selbst nicht im isomorphen Sinn. Laut [\[33, S. 39\]](#) tun sie das sogar selten. [Lemma 6.7.1](#) liefert also eine Klasse von Funktionen, mit denen der Kernel-Trick angewendet werden kann, und mit den Sätzen von Moore und Mercer stehen gleich zwei Sätze zur Verfügung, die dem Entwickler im Maschinellen Lernen den geeigneten Hilbertraum für die Problemstellung liefern können.

Dieses in sich alleine bereits interessante Zwischenergebnis liefert nun zusätzlich noch einen nützlichen **Spezialfall** des Satzes [6.5.1](#) von Mercer: Der folgende Satz ist insofern ein Spezialfall, dass der betrachtete Rohdatenraum $X = \Omega$ **kompakt (K)** und der genutzte Integralkern k **stetig (S)** ist (genauso wie im [Lemma 6.7.1](#) angenommen). Für die spätere Diskussion seien dazu die Stellen im Beweis durch **(K)** bzw. **(S)** markiert, in denen diese speziellen Zusatzbedingungen benötigt werden.

6.7.2 Satz von Mercer für stetige Integralkerne auf Kompaktum

Satz 6.7.3 (**Satz von Mercer** für stetige Integralkerne auf Kompaktum [\[1, 7, Theorem 6.1.1, Theorem III.22\]](#)).

Sei $\Omega \subseteq \mathbb{R}^d$ kompakt, $k \in C(\Omega \times \Omega)$ stetig und von der Form, dass der Operator

$$T_k : L^2(\Omega) \rightarrow L^2(\Omega), \quad (T_k f)(x) := \int_{\Omega} k(x, y) f(y) dy \quad \text{für } f \in L^2(\Omega), x \in \Omega \quad (6.103)$$

selbstadjungiert und positiv ist, d. h. für alle $f \in L^2(\Omega)$ gilt:

$$\langle T_k f, f \rangle_{L^2(\Omega)} \stackrel{\substack{\text{A.3.9} \\ (6.103)}}{=} \int_{\Omega} \int_{\Omega} k(x, y) f(x) \overline{f(y)} dy dx \geq 0. \quad (6.104)$$

Dann gelten:

1. Die Bilder von T_k sind stetig, d. h. $T_k(L^2(\Omega)) \subseteq C(\Omega)$. Insbesondere besitzt jede Eigenfunktion zu einem Eigenwert von T_k ungleich Null einen stetigen Repräsentanten, welcher fast überall mit ihr übereinstimmt,
2. Sind $(e_i)_{i \in I} \subset L^2(\Omega)$ die $\|\cdot\|_{L^2(\Omega)}$ -normierten Eigenfunktionen (d. h. $\|e_i\|_{L^2(\Omega)} = 1$) zu den Eigenwerten $(\lambda_i)_{i \in I}$ ungleich Null von T_k (I ist eine abzählbare Indexmenge), dann konvergiert die Reihenentwicklung von k absolut und gleichmäßig (überall) auf Ω :

$$k(x, y) = \sum_{i \in I} \lambda_i e_i(x) \overline{e_i(y)} \quad \text{für } x, y \in \Omega. \quad (6.105)$$

Beweis. (Die Beweisideen sind [1, 7, Theorem 6.1.1, Theorem III.22] entnommen.)

1. Sei $f \in L^2(\Omega)$ und $\varepsilon > 0$ beliebig aber fest (es sei angemerkt, dass T_k gemäß Def.+Satz A.4.18 oder Satz 6.5.1 wohldefiniert ist). Für das stetige k auf der kompakten Menge Ω gilt mit Prop. A.1.15, dass es sogar gleichmäßig stetig (Def. A.1.13) ist (**K+S**). Es existiert somit ein $\delta > 0$, sodass für alle $x, y, z, w \in \Omega$ mit $|x - y| + |z - w| < \delta$ gilt, dass $|k(x, z) - k(y, w)| < (\sqrt{\mu(\Omega)}\|f\|_{L^2(\Omega)})^{-1}\varepsilon$ ist. Damit folgt insbesondere für alle $x, y, z \in \Omega$ mit $|x - y| < \delta$, dass $|k(x, z) - k(y, z)| < (\sqrt{\mu(\Omega)}\|f\|_{L^2(\Omega)})^{-1}\varepsilon$ gilt und hiermit:

$$|(T_k f)(x) - (T_k f)(y)| \stackrel{(6.103)}{\leq} \int_{\Omega} |(k(x, z) - k(y, z))f(z)| dz \quad (6.106)$$

$$\leq \frac{\varepsilon}{\sqrt{\mu(\Omega)}\|f\|_{L^2(\Omega)}} \int_{\Omega} |f(z)| dz \stackrel{A.2.14}{=} \varepsilon \frac{\|f \cdot 1\|_{L^1(\Omega)}}{\sqrt{\mu(\Omega)}\|f\|_{L^2(\Omega)}} \quad (6.107)$$

$$\stackrel{A.2.15}{\leq} \varepsilon \frac{\|f\|_{L^2(\Omega)} \cdot \|1\|_{L^2(\Omega)}}{\sqrt{\mu(\Omega)}\|f\|_{L^2(\Omega)}} \stackrel{A.2.14}{=} \varepsilon \frac{\sqrt{\mu(\Omega)}}{\sqrt{\mu(\Omega)}} = \varepsilon. \quad (6.108)$$

Das bedeutet, $T_k f$ ist gemäß Def. A.1.13 (ebenfalls gleichmäßig) stetig. Mit der beliebigen Wahl von $f \in L^2(\Omega)$ folgt also: $T_k(L^2(\Omega)) \subseteq C(\Omega)$.

Sei $\lambda \neq 0$ nun ein Eigenwert von T_k und $e \in L^2(\Omega)$ eine zugehörige Eigenfunktion. Da $T_k(L^2(\Omega))$ mittels Def.+Satz A.4.3 ein linearer Raum ist, gilt $e = \frac{\lambda}{\lambda}e = \frac{1}{\lambda}T_k e =: \tilde{e} \in T_k(L^2(\Omega)) \subseteq C(\Omega)$. Die letzte Gleichungsfolge gilt dabei (nur) bzgl. der $\|\cdot\|_{L^2(\Omega)}$ -Norm, weshalb e und $\tilde{e}$ gemäß Bsp. A.6.17 fast überall übereinstimmen, und die Behauptung unter 1. ist nachgewiesen.

2. Zunächst mache man sich klar, dass sämtliche Voraussetzungen der allgemeineren Version des Satzes 6.5.1 erfüllt sind - mit Ausnahme der Mercer-Punkt-4-Zusatzbedingung. Wie im Beweis zu Letzterem, sei darum erneut im Folgenden o. B. d. A. $\tilde{I} = \tilde{N}$ angenommen und mit den Erkenntnissen aus 1. sei o. B. d. A. mit $(e_i)_{i \in \mathbb{N}}$ fortan die Folge der **stetigen**, $\|\cdot\|_{L^2(\Omega, \mu)}$ -normierten Eigenfunktionen bezeichnet, zugehörig zu den Eigenwerten $(\lambda_i)_{i \in \mathbb{N}}$.

Da die punktweise Konvergenz fast überall bereits aus Satz 6.5.1 bekannt ist, prüfe man dann als erstes die punktweise Konvergenz **auf ganz** $\Omega \times \Omega$ von:

$$k(x, y) = \sum_{i=1}^{\infty} \lambda_i e_i(x) \overline{e_i(y)}, \quad x, y \in \Omega. \quad (6.109)$$

Betrachtet man nämlich die Teilsummenfunktion für $x, y \in \Omega$ und $n \in \mathbb{N}$

$$k_n(x, y) := k(x, y) - \sum_{i=1}^n \lambda_i e_i(x) \overline{e_i(y)} \stackrel{(6.109)}{=} \sum_{i>n} \lambda_i e_i(x) \overline{e_i(y)}, \quad (6.110)$$

(wobei das letzte Gleichheitszeichen nur fast überall gilt), dann folgt für $f \in L^2(\Omega)$, $n \in \mathbb{N}$:

$$\int_{\Omega} \int_{\Omega} k_n(x, y) f(x) \overline{f(y)} dx dy \quad (6.111)$$

$$\stackrel{(6.110)}{=} \int_{\Omega} \int_{\Omega} \sum_{i>n} \lambda_i e_i(x) \overline{e_i(y)} f(x) \overline{f(y)} dx dy = \int_{\Omega} \int_{\Omega} \overline{\sum_{i>n} \lambda_i e_i(x) \overline{e_i(y)} f(x) \overline{f(y)}} dx dy \quad (6.112)$$

$$\stackrel{\substack{A.3.9 \\ A.8.1}}{=} \int_{\Omega} \left\langle \sum_{i>n} \overline{\lambda_i e_i} e_i(y), f \right\rangle_{L^2(\Omega)} \overline{f(y)} dy \stackrel{\substack{A.3.6 \\ A.3.1}}{=} \int_{\Omega} \sum_{i>n} \lambda_i \overline{e_i(y)} \langle \overline{e_i}, f \rangle_{L^2(\Omega)} \overline{f(y)} dy \quad (6.113)$$

$$\stackrel{A.3.9}{=} \left\langle \sum_{i>n} \lambda_i \overline{e_i} \langle \overline{e_i}, f \rangle_{L^2(\Omega)}, f \right\rangle_{L^2(\Omega)} \stackrel{\substack{A.3.6 \\ A.3.1}}{=} \sum_{i>n} \lambda_i \langle \overline{e_i}, f \rangle_{L^2(\Omega)} \langle \overline{e_i}, f \rangle_{L^2(\Omega)} \quad (6.114)$$

$$\stackrel{A.8.1}{=} \sum_{i>n} \lambda_i |\langle \overline{e_i}, f \rangle_{L^2(\Omega)}|^2 \stackrel{6.2.1}{\geq} 0. \quad (6.115)$$

Da dies für alle $f \in L^2(\Omega)$ und alle $n \in \mathbb{N}$ nachgewiesen ist und k_n gemäß (6.110) i. V. m. dem unter 1. Gefundenen als endliche Summe von stetigen Funktionen ebenfalls stetig ist, sind somit alle Voraussetzungen von Lemma 6.7.1 erfüllt (**K+S**). Letzteres besagt, dass k_n für jedes $n \in \mathbb{N}$ eine positiv semidefinite Funktion ist und damit insbesondere, dass $k_n(x, x) \geq 0$ für alle $x \in \Omega$ ist (Man wähle dazu in Def. 4.1.3 $n = 1$, $\alpha_1 = 1$ und $x_1 = x$ für jedes $x \in \Omega$). Anders geschrieben heißt das mit (6.110), dass für alle $n \in \mathbb{N}$ und $x \in \Omega$ gilt:

$$\sum_{i=1}^n \lambda_i |e_i(x)|^2 \leq k(x, x) \leq \|k\|_{\infty} \quad \Rightarrow \quad \sum_{i=1}^{\infty} \lambda_i |e_i(x)|^2 \leq \|k\|_{\infty} \quad (6.116)$$

da k als stetige Funktion auf einer kompakten Menge gemäß Prop. A.1.15 und Bsp. A.2.10 durch $\|k\|_{\infty}$ beschränkt ist (**K+S**).

Mit dem Zwischenergebnis (6.116) ist die punktweise Konvergenz von (6.109) auf ganz $\Omega \times \Omega$ nun nicht mehr weit. Man wähle dazu $\varepsilon > 0$ sowie $x \in \Omega$ beliebig aber fest. Es gibt wegen der Konvergenz der Reihe in (6.116) ein $n_x \in \mathbb{N}$, sodass für alle $m, n \geq n_x$ gilt:

$$\sum_{i=m}^n \lambda_i |e_i(x)|^2 \leq \frac{\varepsilon}{\|k\|_{\infty}} \quad (6.117)$$

Daraus schließt man, da die Eigenwerte $(\lambda_i)_{i \in \mathbb{N}}$ gemäß Satz 6.5.1 Punkt 1. allesamt reell und nicht-negativ sind, für die Funktionenfolge $(\tilde{k}_n)_{n \in \mathbb{N}} \in C(\Omega)$, definiert durch folgende jeweils endliche Summe stetiger Funktionen $\tilde{k}_n(y) := \sum_{i=1}^n \lambda_i e_i(x) \overline{e_i(y)}$ für $y \in \Omega$:

$$\|\tilde{k}_n - \tilde{k}_m\|_{\infty} \stackrel{\substack{A.2.9 \\ (6.110)}}{\leq} \sup_{y \in \Omega} \sum_{i=m}^n \lambda_i |e_i(x) \overline{e_i(y)}| \stackrel{\substack{A.2.13 \\ 6.5.1}}{\leq} \sup_{y \in \Omega} \left(\sum_{i=m}^n \lambda_i |e_i(x)|^2 \right) \left(\sum_{i=m}^n \lambda_i |e_i(y)|^2 \right) \quad (6.118)$$

$$\stackrel{(6.117)}{\leq} \frac{\varepsilon}{\|k\|_{\infty}} \|k\|_{\infty} = \varepsilon. \quad (6.119)$$

Ergo ist $(\tilde{k}_n)_{n \in \mathbb{N}}$ eine Cauchyfolge in $(C(\Omega), \|\cdot\|_{\infty})$, was aufgrund dessen Vollständigkeit (siehe Bsp. A.2.10) bedeutet, dass $\lim_{n \rightarrow \infty} \tilde{k}_n = \sum_{i=1}^{\infty} \lambda_i e_i(x) \overline{e_i(\cdot)}$ stetig auf Ω ist. Es gilt aber auch,

dass $k(x, \cdot)$ stetig auf Ω ist und zudem, dass $k(x, \cdot) = \sum_{i=1}^{\infty} \lambda_i e_i(x) \overline{e_i(\cdot)}$ fast überall auf Ω gilt. Aufgrund der Stetigkeit beider Seiten des Gleichheitszeichen, muss die Gleichheit damit überall auf ganz Ω gelten (vgl. identische Argumentation im Beweis zu [Lemma 6.7.1](#)). Da $x \in \Omega$ beliebig gewählt war, folgt damit die punktweise Konvergenz von (6.109) auf ganz $\Omega \times \Omega$.

Um schließlich auch die gleichmäßige Konvergenz von (6.109) einzusehen, seien zwei Beweise gegeben: Ein kürzerer, welcher im Prinzip nur den [Satz A.1.16](#) von Dini anwendet und ein etwas längerer, welcher die nötige Argumentation aus dem Satz von Dini direkt führt und aus dem etwas besser hervorgeht, an welchen Stellen **(K)** bzw. **(S)** genau benötigt werden:

Kurzer Beweis gemäß [1, Theorem 6.1.1]: Man betrachte die Teilsummenfolge $(g_n)_{n \in \mathbb{N}} \subset \mathcal{F}(\Omega)$, definiert durch $g_n := \sum_{i=1}^n \lambda_i |e_i|^2$ für $n \in \mathbb{N}$, wobei jedes g_n mit dem unter 1. Gefundenen als endliche Summe stetiger Funktionen ebenfalls stetig ist (**(K+S)**). Aus [Satz 6.5.1](#) ist zudem bekannt, dass die Eigenwerte $(\lambda_n)_{n \in \mathbb{N}}$ allesamt nicht-negativ und insbesondere reell sind, d. h. $(g_n)_{n \in \mathbb{N}} \subset C(\Omega, \mathbb{R})$. Außerdem folgt aus der oben gezeigten punktweisen Konvergenz von (6.109), dass $(g_n)_{n \in \mathbb{N}}$ gegen $g(x) := \sum_{i=1}^{\infty} \lambda_i |e_i(x)|^2 = \sum_{i=1}^{\infty} \lambda_i e_i(x) \overline{e_i(x)} = k(x, x) \in C(\Omega, \mathbb{R})$ punktweise konvergiert. Wegen der aufsteigenden Summe und der Nicht-Negativität von $(\lambda_i)_{i \in \mathbb{N}}$ folgt darüber hinaus: $g_n(x) \leq g_{n+1}(x)$ für alle $n \in \mathbb{N}$, $x \in \Omega$. Schließlich sind wegen der Kompaktheit **(K)** von Ω damit alle Voraussetzungen aus dem [Satz A.1.16](#) von Dini erfüllt und dieser besagt, dass die Konvergenz von $(g_n)_{n \in \mathbb{N}}$ sogar gleichmäßig ist. Das bedeutet aber gerade, dass die Mercer-Punkt-4-Zusatzbedingung erfüllt ist, nämlich, dass die Reihe $\sum_{i \in I} \lambda |e_i(x)|^2$ auf ganz Ω gleichmäßig konvergiert. Damit folgt aus [Satz 6.5.1](#) 4. die Behauptung.

Ausführlicherer Beweis gemäß [7, Theorem III.22]: Sei $\varepsilon > 0$ beliebig, aber fest. Man definiere damit die Ω -Teilmengen

$$V_n^\varepsilon := \{x \in \Omega : k(x, x) - \sum_{i=1}^n \lambda_i |e_i(x)|^2 < \varepsilon\}, \quad n \in \mathbb{N}. \quad (6.120)$$

Dann gilt für alle $n \in \mathbb{N}$ wegen (6.116), dass $V_n^\varepsilon \subseteq V_{n+1}^\varepsilon$, und wegen 1., dass die V_n^ε offen sind in Ω : Gilt nämlich $x \in V_n^\varepsilon$, d. h. $k(x, x) - \sum_{i=1}^n \lambda_i |e_i(x)|^2 < \varepsilon$, so gibt es wegen der stetigen linken Seite der Abschätzung **(K+S)** eine offene Kugel $U \subseteq \Omega$ mit Zentrum x , sodass $k(y, y) - \sum_{i=1}^n \lambda_i |e_i(y)|^2 < \varepsilon$ für alle $y \in U$ ist (siehe [Def.+Satz A.1.12](#)). Es folgt $U \subseteq V_n^\varepsilon$ und damit, dass jedes V_n^ε offen ist. Zudem gilt mit der punktweisen Konvergenz von (6.109), dass

$$\bigcup_{n \in \mathbb{N}} V_n^\varepsilon = \Omega \quad (6.121)$$

eine offene Überdeckung von Ω bildet. Wegen der Kompaktheit **(K)** von Ω gibt es somit eine endliche Teilüberdeckung (siehe [Def. A.1.9](#)), ergo wegen $V_n^\varepsilon \subseteq V_{n+1}^\varepsilon$ ein n_ε mit $V_{n_\varepsilon}^\varepsilon = \Omega$, welches nicht mehr von $x \in \Omega$ abhängt. Es gilt dann für alle $n \geq n_\varepsilon$:

$$\left\| k - \sum_{i=1}^n \lambda_i |e_i|^2 \right\|_\infty \stackrel{\text{A.2.9}}{\leq} \sup_{x \in \Omega = V_{n_\varepsilon}^\varepsilon} \left\{ k(x, x) - \sum_{i=1}^{n_\varepsilon} \lambda_i |e_i(x)|^2 \right\} \leq \sup_{x \in \Omega} \varepsilon = \varepsilon \quad (6.122)$$

Da $\varepsilon > 0$ beliebig war, besagt dies die gleichmäßige Konvergenz von $k(x, x) = \sum_{i=1}^{\infty} \lambda_i |e_i(x)|^2$ für $x \in \Omega$ gemäß [Bsp. A.2.9](#) und damit erneut, dass die Mercer-Punkt-4-Zusatzbedingung erfüllt ist. Der letzte Teil der Aussage folgt dann abermals aus [Satz 6.5.1](#) Punkt 4. $\square$

Bemerkung 6.7.4. Ebenfalls für [Satz 6.7.3](#) zeigt [\[7, Theorem III.22\]](#), dass er auch für allgemeine komakte, metrische Räume (X, d) gilt, die mit einem regulären Borelmaß ausgestattet sind, wenn für alle nicht-leeren offenen $U \subseteq X$ darin gilt, dass $\mu(U) > 0$ ist. Der Beweis muss dazu nicht wirklich verändert werden.

Mit diesem Spezialfall des Satzes von Mercer gewinnt der Entwickler im Maschinellen Lernen also mit der Stetigkeit eine etwas einfacher zu prüfende Eigenschaft, anstatt wie in [Satz 6.5.1](#) der Mercer-Punkt-4-Zusatzbedingung. Für Letztere mussten bereits vor Anwendung des Kernel-Tricks Kenntnisse über Eigenwerte und Eigenfunktionen des durch den Integralkern k entstehenden Hilbert-Schmidt Integraloperatoren vorliegen. Diese schwierige Vorbedingung entfällt im obigen Fall.

Den dadurch entstehenden Freiheitsgrad tauscht der Entwickler aber dahingehend ein, dass er nun bereits **topologische Eigenschaften** des Rohdatenraums $X = \Omega$ kennen muss. Sowohl für die Stetigkeit der Integralkerne als auch für die Kompaktheit von $X = \Omega$ ist eine Topologie nötig. Es ist instruktiv, sich anhand der Markierungen **(K)** bzw. **(S)** klar zu machen, an welchen Stellen im Beweis genau die Zusatzbedingungen benötigt werden: Ein Großteil wird darauf verwendet, die **punktweise Konvergenz** in [\(6.109\)](#) auf ganz $X = \Omega$ zu erhalten. Im Kontext des Maschinellen Lernens ist diese Eigenschaft ggf. aber gar nicht so relevant wie in der mathematischen Theorie:



Frage / Ausblick:

Wieweit lässt sich [Satz 6.7.3](#) verallgemeinern, wenn die **gleichmäßige Konvergenz** in [\(6.109\)](#) nur auf einer praktisch interessanten Teilmenge gefordert wird?

Trotzdem darf nicht unterschätzt werden, welche zentrale Rolle die punktweise Konvergenz von [\(6.109\)](#) **auf ganz Ω** auch im Nachweis der gleichmäßigen Konvergenz spielt: Sie erlaubt es, dass die Kompaktheit (insbesondere die Abgeschlossenheit) von Ω weiter gilt und nicht durch eine ggf. unbekannte, herauszunehmende (Null-)Menge in Gefahr gerät. Diese ist dann das eigentliche Hauptargument, mit welchem die gleichmäßige Konvergenz nachgewiesen wird.

Nichtsdestotrotz scheint es, dass eine Verallgemeinerung von [Satz 6.7.3](#) für den Kontext des Maschinellen Lernens möglich sein könnte. Zumindest bezüglich der Stetigkeit lässt sich die Aussage von [Satz 6.7.3](#) durch Anwendung des **Satzes A.6.22 von Luzin** leicht gewinnbringend verallgemeinern:

6.7.3 Satz von Mercer für metrischen, regulären Borel-Raum

Korollar 6.7.5 (**Satz von Mercer** für metrischen, regulären Borel-Raum [\[26\]](#)).

Sei $\Omega \subseteq \mathbb{R}^d$ messbar und $k \in L^\infty(\Omega)$, sodass (Ω, μ) ein endlicher Maßraum und der Operator

$$T_k : L^2(\Omega) \rightarrow L^2(\Omega), \quad (T_k f)(x) := \int_{\Omega} k(x, y) f(y) dy \quad \text{für } f \in L^2(\Omega), x \in \Omega \quad (6.123)$$

selbstadjungiert und positiv ist, d. h. für alle $f \in L^2(\Omega)$ gilt

$$\langle T_k f, f \rangle_{L^2(\Omega)} = \int_{\Omega} \int_{\Omega} k(x, y) f(x) \overline{f(y)} dy dx \geq 0. \quad (6.124)$$

Dann gilt: Für jedes $\varepsilon > 0$ existiert eine komakte Menge K mit $\mu(\Omega \setminus K) < \varepsilon$, sodass für die Eigenwerte $(\lambda_i)_{i \in I}$ ungleich Null von T_k (I ist eine abzählbare Indexmenge) und die zugehörigen, paarweise

verschiedenen $\|\cdot\|_{L^2(\Omega, \mu)}$ -normierten Eigenfunktionen $(e_i)_{i \in I} \subset L^2(\Omega, \mu)$ (d. h. $\|e_i\|_{L^2(\Omega, \mu)} = 1$) gilt, dass die (Spektral-)Reihenentwicklung von k absolut und gleichmäßig auf K konvergiert:

$$k(x, y) = \sum_{i \in I} \lambda_i e_i(x) \overline{e_i(y)} \quad \text{für } x, y \in \Omega. \quad (6.125)$$

Beweis. (Die Beweisideen entstanden gemeinsam mit [26].)

Sei $\varepsilon > 0$ beliebig, aber fest. Da $k \in L^\infty(\Omega)$ ist, gibt es eine Teilmenge $\tilde{\Omega} \subseteq \Omega$, auf der k beschränkt ist (und messbar, als Element von $L^\infty(\Omega)$) und es gilt $\mu(\Omega \setminus \tilde{\Omega}) = 0$. Mit dem Satz A.6.22 von Luzin gibt es damit eine kompakte Teilmenge $K \subseteq \tilde{\Omega} \subseteq \Omega$ mit $\mu(\tilde{\Omega} \setminus K) < \varepsilon$, sodass $k|_{K \times K}$ stetig ist. Damit sind für K und $k|_{K \times K}$ alle Voraussetzungen von Satz 6.7.3 erfüllt und es folgt, dass die Reihenentwicklung von k in (6.125) absolut und gleichmäßig auf K konvergiert. Zudem gilt:

$$\mu(\Omega \setminus K) = \mu(((\Omega \cap \tilde{\Omega}) \cup (\Omega \setminus \tilde{\Omega})) \setminus K) = \mu(((\Omega \cap \tilde{\Omega}) \setminus K) \cup ((\Omega \setminus \tilde{\Omega}) \setminus K)) \quad (6.126)$$

$$\stackrel{\text{A.6.6}}{\leq} \mu((\Omega \cap \tilde{\Omega}) \setminus K) + \mu((\Omega \setminus \tilde{\Omega}) \setminus K) \stackrel{\text{A.6.5}}{\leq} \mu(\tilde{\Omega} \setminus K) + \mu(\Omega \setminus \tilde{\Omega}) < \varepsilon + 0 = \varepsilon. \quad (6.127)$$

□

Bemerkung 6.7.6. Gemäß [3, 12, Satz 26.7, Satz VIII.1.18] existieren allgemeinere Versionen des Satzes A.6.22 von Luzin im Kontext von Bem. 6.7.2 und Bem. 6.7.4, d. h. Kor. 6.7.5 lässt sich ebenso auf verallgemeinerte metrische Räume mit regulärem Borelmaß ausweiten, wie Satz 6.7.3 es zulässt. In dieser Arbeit wurde der Spezialfall des Lebesgue-Maßes auf einer Teilmenge des $\mathbb{R}^n$ lediglich aus Platzgründen behandelt und damit die Anforderungen an die Vorkenntnisse des Lesers bei einem universitären Funktionalanalysis-Kurs bleiben.

Für das Maschinelle Lernen sollte diese Variante des Satzes von Mercer aus Kor. 6.7.5 einige Vorteile haben: Die Voraussetzungen sind beinahe wieder identisch zu denen in der allgemeinen Form des Satzes von Mercer in Satz 6.5.1, auf die Prüfung der Mercer-Punkt-4-Zusatzbedingung kann verzichtet werden und trotzdem ist die gleichmäßige Konvergenz von (6.125) auf einer beliebig großen Teilmenge des Rohdatenraums $X = \Omega$ garantiert. Einziger Nachteil dieser Version ist weiterhin, dass eine Topologie auf $X = \Omega$ bekannt sein muss.

Eine Möglichkeit, mit welcher der Entwickler im Maschinellen Lernen zu einer Topologie gelangen könnte, liefert ggf. der Kernel-Trick zusammen mit dem Satz 4.2.1 von Moore: Vielleicht ist es sinnvoll, zuerst mit Letzterem einen (möglicherweise unendlich-dimensionalen) RKHS zu finden und damit eine Topologie „auf der“ Rohdatenmenge, um mit dieser erst Kor. 6.7.5 anzuwenden. Ob dies aber eine effektive Technik darstellt, müsste in der Praxis geprüft werden. Man darf keinesfalls herunterspielen wie einschränkend die Zusatzbedingung einer Topologie in vielen Anwendungsfällen sein kann, denn sie stellt oftmals das gewünschte Ergebnis und eben keine Vorbedingung dar.



Frage / Ausblick:

Liefert der Satz 4.2.1 von Moore eine geeignete Topologie für Kor. 6.7.5?

6.7.4 Satz von Mercer für Heat Kernels

Aus diesem Grund sei noch ein weiterer interessanter Spezialfall des **Satzes 6.5.1 von Mercer** vorgestellt, der keine Topologie auf der Rohdatenmenge erfordert. Um diesen korrekt formulieren zu können, müssen zunächst Begriffe aus den **Partiellen Differentialgleichungen** eingeführt werden:

Definition 6.7.7 (Operatoren-Halbgruppe [15, 24, Def. 2.1, Def. 5.10]).

Sei X ein Banachraum. Eine Familie $(T(t))_{t \geq 0} \subset \mathcal{L}(X)$ von beschränkten linearen Operatoren auf X wird eine **Operatoren-Halbgruppe** genannt, wenn gilt:

$$T(t)T(s) = T(t+s) \quad \text{für alle } t, s \geq 0 \quad \text{und} \quad T(0) = \text{Id}. \quad (6.128)$$

Definition 6.7.8 (Abstraktes Cauchy-Problem (ACP) [24, Def. 5.1]).

Sei X ein Vektorraum, $\psi_0 \in X$ und $A : X \rightarrow X$ ein linearer Operator. Ein **Abstraktes Cauchy-Problem (ACP)** besteht darin, ein (schwach) differenzierbares $\psi : \mathbb{R}_+ \rightarrow X$ zu finden, welches die X -wertige gewöhnliche Differentialgleichung

$$\begin{cases} \frac{d\psi}{dt}(t) = A\psi(t), & t \geq 0 \\ \psi(0) = \psi_0 \end{cases} \quad (6.129)$$

löst.

Ist X ein Banachraum und das abstrakte Cauchy-Problem (ACP) der Form, dass eine eindeutige Lösung $\psi(t)$ ($t \geq 0$) für jeden Anfangswert $\psi_0 \in X$ existiert, so schreibt man (mit einem kleinen Missbrauch der Notation) für die Operator-Familie $(T(t))_{t \geq 0}$, die jedem Anfangswert ψ_0 die eindeutige Lösung des (ACP) zuordnet:

$$e^{tA} := T(t) : X \ni \psi_0 \mapsto \psi(t) \in X, \quad t \geq 0. \quad (6.130)$$

Definition 6.7.9 (Symmetrische Markovsche Halbgruppe, [10, Theorem 1.3.2, Theorem 1.3.3, S. 22]).

Sei $(\Omega, \mathcal{A}, \mu)$ ein endlicher Maßraum mit einem abzählbaren Erzeuger (d. h. es gibt eine abzählbare Mengenfamilie $\mathcal{E}$ mit $\mathcal{A} = \sigma(\mathcal{E})$). Ist $H : L^2(\Omega, \mu) \rightarrow L^2(\Omega, \mu)$ ein selbstadjungierter Operator mit den Eigenschaften:

1. $H \geq 0$, d. h. $\langle Hf, f \rangle_{L^2(\Omega, \mu)} \geq 0$ für alle $f \in L^2(\Omega, \mu)$,
2. e^{tH} ist Positivität-erhaltend für alle $t \geq 0$, d. h. für alle $f \in L^2(\Omega, \mu)$ mit $f \geq 0$, gilt $e^{tH}f \geq 0$ und
3. e^{tH} ist eine Kontraktion in $L^\infty(\Omega)$ für alle $t \geq 0$, d. h. es existiert ein $L < 1$, mit dem für alle $f, g \in L^\infty(\Omega)$ gilt: $\|e^{tH}f - e^{tH}g\|_{L^\infty(\Omega)} \leq L\|f - g\|_{L^\infty(\Omega)}$.

Dann nennt man H eine **Dirichlet-Form** und e^{tH} eine **symmetrische Markovsche Halbgruppe**.

Definition 6.7.10 (Ultrakonstraktive, symmetrische Markovsche Halbgruppe [10, Theorem 2.1.1]).

Eine symmetrische Markovsche Halbgruppe e^{tH} wird **ultrakonstraktiv** genannt, wenn der Operator $e^{tH} : L^2(\Omega, \mu) \rightarrow L^\infty(\Omega, \mu)$ beschränkt ist für alle $t > 0$.

Mit diesen Definitionen ausgestattet lässt sich nun der zweite Spezialfall des **Satzes von Mercer** formulieren, wobei für den Beweis aus Platzgründen auf [10, Theorem 2.1.4] verwiesen wird. Obgleich Letzterer nicht allzu lang ist, erfordert er noch weitere Theorie, welche diese Arbeit sprengen würde.

Satz 6.7.11 (Satz von Mercer für Heat Kernels [10, 23, Theorem 2.1.4]).

Sei e^{tH} eine ultrakonstraktive, symmetrische Markovsche Halbgruppe auf $L^2((\Omega, \mu), \mathbb{R})$, wobei Ω endliches Maß besitze. Dann sind die Eigenwerte von e^{tH} absolutsummierbar für jedes $t > 0$, d. h. ein Element des $\ell^1(\mathbb{N})$. Sind zudem mit $(\lambda_i)_{i \in \mathbb{N}}$ die Eigenwerte von H bezeichnet, in aufsteigender Reihenfolge sortiert und gemäß ihrer Multiplizität gezählt und mit $(e_i)_{i \in \mathbb{N}}$ die jeweils zugehörigen, paarweise verschiedenen $\|\cdot\|_{L^2((\Omega, \mu), \mathbb{R})}$ -normierten Eigenfunktionen, dann gelten:

1. $e_i \in L^\infty((\Omega, \mu), \mathbb{R})$ für alle $i \in \mathbb{N}$ und
2. für den Integralkern $k : \{t\} \times \Omega \times \Omega \rightarrow \mathbb{R}$ von e^{tH} gilt:

$$k(t, x, y) = \sum_{i=1}^{\infty} \exp(-\lambda_i t) e_i(x) e_i(y). \quad (6.131)$$

Die Reihe konvergiert dabei gleichmäßig auf $[\alpha, \infty) \times \Omega \times \Omega$ für alle $\alpha > 0$.

Bemerkung 6.7.12. In seiner Version obigen Satzes setzt [10, Theorem 2.1.4] zwar streng genommen voraus, dass Ω ein **lokalkompakter, zweitabzählbarer Hausdorffraum** und μ ein **Borelmaß** darauf ist, merkt aber bereits selbst [10, S. 12] an, dass allgemeinere Fälle gelten. Gemäß [23] ist es möglich, den Beweis durch [10, Theorem 2.1.4] identisch zu führen, wenn (Ω, μ) als gewöhnlicher endlicher Maßraum vorausgesetzt ist, weshalb Satz 6.7.11 in entsprechender Form angepasst ist.

Diese Version des Satzes von Mercer kommt wie angekündigt zumindest in der Theorie ohne eine Topologie auf der Rohdatenmenge aus. Aber auch dieser Freiheitsgrad kommt nicht kostenfrei: Der Entwickler im Maschinellen Lernen muss nun zunächst ein passendes **abstraktes Cauchy-Problem** finden, dessen **Lösungsoperator** eine ultrakonstraktive, symmetrische Markovsche Halbgruppe ist. Zudem muss dann (mindestens) ein Operator in Letzterer auch noch für die betrachtete Problemstellung geeignet sein. Etwas abstrakter gesagt, ist bei dieser Variante des Satzes von Mercer die **Funktionenklasse** der möglichen Integralkerne viel **eingeschränkter** als in Kor. 6.7.5 oder Satz 6.5.1.

Auf der anderen Seite wurde in **Unterabschnitt 6.6.1** bereits gemutmaß, dass eine allzugroße Integralkernklasse vielleicht gar nicht im Sinne des Entwicklers ist, da das **Kernel-Trick-Experiment** dann ggf. nicht mehr berechenbar ist. Gemäß [23] gewinne man mit Satz 6.7.11 z. B. die Möglichkeit, eine ganze Familie $(e^{tH})_{t \geq 0}$ von Integralkernen für den Kernel-Trick und könne mit dem Parameter t „spielen“. Insbesondere könne die Betrachtung des Grenzfalles $t \rightarrow \infty$ interessant werden. Für weitere Ausführungen wird auf [22] verwiesen. Es ergibt sich folgende



Frage / Ausblick:

In welchen Problemstellungen des Maschinellen Lernens finden beispielsweise **Heat Kernels** Anwendung (d. h. die Lösungsoperatoren der **Wärmeleitungsgleichung**)?

6.8 Interessantes zum Abschluss

Zum Abschluss dieses Kapitels und der gesamten Arbeit seien ein paar interessante Erkenntnisse zusammen gestellt, welche sich aus zwei Aufgaben in [33, Problem 2.24, 2.29] ergeben.

6.8.1 Notwendiges Kriterium für den Satz von Mercer

Korollar 6.8.1 (Notwendiges Kriterium für den Satz von Mercer [33, Problem 2.24]).

Es seien die Bedingungen von Satz 6.5.1 ohne Kenntnis über die Mercer-Punkt-4-Zusatzbedingung erfüllt. Dann gilt für die mit Multiplizität gezählten Eigenwerte $(\lambda_i)_{i \in I}$ und die paarweise verschiedenen, $\|\cdot\|_{L^2(\Omega, \mu)}$ -normierten Eigenfunktionen $(e_i)_{i \in I}$:

$$\sup_{i \in I} \sqrt{\lambda_i} \|e_i\|_{L^\infty(\Omega, \mu)} \leq \sqrt{M} < \infty. \quad (6.132)$$

Beweis. Aus den Voraussetzungen und Punkt 3. von Satz 6.5.1 ist bekannt, dass es drei Nullmengen $N, \tilde{N}, \bar{N}$ gibt, mit denen $\sup_{x \in \Omega \setminus N} \sup_{y \in \Omega \setminus \tilde{N}} |k(x, y)| \leq M < \infty$ gilt und die Reihenentwicklung des Integralkerns k (6.48) auf $\bar{N}$ konvergiert. Sei nun $\hat{N} = N \cup \tilde{N} \cup \bar{N}$ als Vereinigung von Nullmengen auch eine ebensolche (vgl. Kor. A.6.6). Dann folgt mit der Positivität der Eigenwerte (6.2.1):

$$\sup_{i \in I} \lambda_i \|e_i\|_{L^\infty(\Omega, \mu)}^2 \stackrel{\substack{\text{A.2.14} \\ \text{A.2.4}}}{=} \sup_{i \in I} \sup_{x \in \Omega \setminus \hat{N}} \lambda_i |e_i(x)|^2 \stackrel{\substack{6.2.1 \\ \text{A.8.1}}}{\leq} \sup_{i \in I} \sup_{x \in \Omega \setminus \hat{N}} \left| \sum_{j \in I} \lambda_j e_j(x) \overline{e_j(x)} \right| \quad (6.133)$$

$$\stackrel{(6.48)}{=} \sup_{i \in I} \sup_{x \in \Omega \setminus \hat{N}} |k(x, x)| \leq \sup_{i \in I} M = M < \infty. \quad (6.134)$$

Beidseitiges Wurzelziehen liefert die Behauptung. □

Bemerkung 6.8.2. In der Version des Satzes von Mercer in [18, Theorem 3.a.1] wurde behauptet, dass die **gleichförmige Beschränktheit der Eigenfunktionen** aus den übrigen Bedingungen folgen solle, was – wie bereits ausgeführt – durch [5, 6, 19, Remark 6.3.12, Remark 12] widerlegt werden konnte. Kor. 6.8.1 zeigt nun, dass die Beschränktheit zumindest für die Folge $(\sqrt{\lambda_i} \|e_i\|_{L^\infty(\Omega, \mu)})_{i \in I}$ anstelle der Folge $(\|e_i\|_{L^\infty(\Omega, \mu)})_{i \in I}$ gilt. Der Vorfaktor ist gemäß 1. in Satz 6.5.1 dabei „nur“ ein Element des $\ell^2(I) \subseteq \ell^2(\mathbb{N})$.

Zudem könnte Kor. 6.8.1 ggf. ein nützliches **notwendiges Kriterium** für geeignete Integralkerne zur Anwendung im Satz 6.5.1 von Mercer liefern: Obgleich es bestimmt einfacher ist, die fast überall Beschränktheit eines potentiellen Integralkerns zu prüfen, als alle Eigenwerte und Eigenfunktionen vor Anwendung des Satzes von Mercer zu kennen, ist die Berücksichtigung von Kor. 6.8.1 ggf. bei der **Konstruktion** geeigneter Integralkernfunktionen sinnvoll (vgl. Unterabschnitt 6.6.1).

6.8.2 Der Satz von Mercer und RKHS

Die zweite Aufgabe aus [33, Problem 2.29] betrachtet die Verbindung zwischen dem Satz 6.5.1 von Mercer und RKHS. Sie verbindet damit die Hauptbetrachtungsobjekte dieser Arbeit.

In [Unterabschnitt 6.7.1](#) wurde bereits die Äquivalenz von stetigen reproduzierenden Kernen auf einem Kompaktum und ebenso stetigen Integralkernen beschrieben, welche die Voraussetzungen des [Satzes 6.5.1](#) von Mercer darauf erfüllen. Die Stetigkeit wurde dabei insbesondere gebraucht, um den [Satz 6.7.3 von Mercer für stetige Kerne auf einem Kompaktum](#) nachzuweisen.

Für den allgemeinen nicht-stetigen Fall auf einem beliebigen endlichen Maßraum stellt sich heraus, dass [Lemma 6.7.1](#) und damit die Hinrichtung dieser Äquivalenz in leicht abgewandelter Form auch gilt. Mit dem Verlust der Stetigkeit muss dazu aber der endliche Maßraum (ggf.) eingeschränkt werden: Man schneidet die Nullmengen aus dem Maßraum heraus, auf welchen die punktweise Konvergenz der (Spektral-)Reihenentwicklung [\(6.48\)](#) im [Satz 6.5.1](#) von Mercer für endliche Maßräume nicht gilt bzw. auf welchen der Integralkern k nicht beschränkt ist.

Korollar 6.8.3 (Mercer-Kerne sind reproduzierende Kerne [\[33, Problem 2.29, S. 38 ff.\]](#)).

Es seien die Bedingungen von [Satz 6.5.1](#) ohne Kenntnis über die Mercer-Punkt-4-Zusatzbedingung erfüllt, mit dem Zusatz, dass $|k| \leq M < \infty$ und die (Spektral-)Reihenentwicklung [\(6.48\)](#) von k auf ganz $\Omega \times \Omega$ gilt, d. h. es konvergiert (o. B. d. A. seien im Folgenden mit $(\lambda_i)_{i \in I}$ die rein positiven Eigenwerte von T_k bezeichnet, gemäß ihrer Multiplizität gezählt, man vergleiche [Satz 6.5.1 Punkt 1.](#))

$$k(x, y) = \sum_{i \in I} \lambda_i e_i(x) \overline{e_i(y)} \quad \text{für alle } x, y \in \Omega. \quad (6.135)$$

Dann ist k ein reproduzierender Kern.

Beweis. (Die Beweisideen sind [\[33, Problem 2.29\]](#) entnommen.)

Die symmetrische Bedingung [\(4.5\)](#) in [Def. 4.1.3](#) ist für k erfüllt, denn für alle $x, y \in \Omega$ gilt:

$$k(x, y) \stackrel{(6.135)}{=} \sum_{i \in I} \lambda_i e_i(x) \overline{e_i(y)} \stackrel[6.2.1]{A.8.1} \overline{\sum_{i \in I} \lambda_i e_i(x) e_i(y)} \stackrel{(6.135)}{=} \overline{k(y, x)}. \quad (6.136)$$

Zudem ist die Positivitätsbedingung [\(4.6\)](#) in [Def. 4.1.3](#) ebenfalls erfüllt, denn für alle $n \in \mathbb{N}$ mit zugehörigen $x_1, \dots, x_n \in \Omega$ sowie $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ gilt:

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} k(x_i, x_j) \stackrel{(6.135)}{=} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} \sum_{k \in I} \lambda_k e_k(x_i) \overline{e_k(x_j)} \quad (6.137)$$

$$\stackrel[6.2.1]{A.8.1} \sum_{k \in I} \left(\sum_{i=1}^n \sqrt{\lambda_k} \alpha_i e_k(x_i) \right) \overline{\left(\sum_{j=1}^n \sqrt{\lambda_k} \alpha_j e_k(x_j) \right)} \quad (6.138)$$

$$\stackrel{A.8.1}{=} \sum_{k \in I} \left| \sum_{i=1}^n \sqrt{\lambda_k} \alpha_i e_k(x_i) \right|^2 \geq 0. \quad (6.139)$$

Als positiv semidefinite Funktion ist k damit gemäß [Kor. 4.2.2](#) ein reproduzierender Kern. $\square$

Dieses die [Kapitel 3](#), [Kapitel 4](#) und [Kapitel 6](#) schön verbindende [Kor. 6.8.3](#) lässt sich noch weiter konkretisieren. Da k darin als reproduzierender Kern nachgewiesen ist, reproduziert er mit dem [Satz 4.2.1 von Moore](#) einen RKHS $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$, dessen Struktur im Folgenden betrachtet wird.

Korollar 6.8.4 (Struktur des Mercer-RKHS [27, 33, Theorem 11.18, Problem 2.29]).

Es gelten die Bedingungen aus Kor. 6.8.3 und sei $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ der RKHS mit Kern k . Dann gilt:

$$\mathcal{H}_k = \{f \in \mathcal{F}(\Omega) : f = \sum_{i \in I} \alpha_i \sqrt{\lambda_i} e_i \text{ mit } \alpha := (\alpha_i)_{i \in I} \in \ell^2(\mathbb{N})\} \quad \text{und} \quad \langle f, \tilde{f} \rangle_{\mathcal{H}_k} = \langle \alpha, \tilde{\alpha} \rangle_{\ell^2(\mathbb{N})}. \quad (6.140)$$

Beweis. (Die Beweisideen sind [27, Theorem 11.18] entnommen.)

O.B.d.A. sei im Folgenden wie in den Beweisen zuvor $I = \mathbb{N}$ angenommen (z.B. durch Auffüllen aller betroffenen Folgen durch Nullen) und die Eigenwerte $(\lambda_i)_{i \in \mathbb{N}}$ echt positiv. Dann sei zunächst die Teilmenge $\mathcal{H} := \{f \in \mathcal{F}(\Omega) : f = \sum_{i \in I} \alpha_i \sqrt{\lambda_i} e_i \text{ mit } \alpha := (\alpha_i)_{i \in I} \in \ell^2(\mathbb{N})\}$ von $\mathcal{F}(\Omega)$ betrachtet. Diese ist wohldefiniert, denn für alle $f \in \mathcal{H}$ und $x \in \Omega$ gilt:

$$|f(x)| \stackrel{(6.140)}{=} \left| \sum_{i=1}^{\infty} \alpha_i \sqrt{\lambda_i} e_i(x) \right| \leq \sum_{i=1}^{\infty} |\alpha_i \sqrt{\lambda_i} e_i(x)| \stackrel{A.2.11}{=} \left\| \alpha \cdot \left(\sqrt{\lambda_i} e_i(x) \right)_{i \in \mathbb{N}} \right\|_{\ell^1(\mathbb{N})} \quad (6.141)$$

$$\stackrel{A.2.13}{\leq} \|\alpha\|_{\ell^2(\mathbb{N})} \left\| \left(\sqrt{\lambda_i} e_i(x) \right)_{i \in \mathbb{N}} \right\|_{\ell^2(\mathbb{N})} \stackrel{A.2.11}{=} \|\alpha\|_{\ell^2(\mathbb{N})} \left(\sum_{i=1}^{\infty} \lambda_i |e_i(x)|^2 \right)^{\frac{1}{2}} \stackrel{6.2.1}{=} \|\alpha\|_{\ell^2(\mathbb{N})} \quad (6.142)$$

$$\stackrel{(6.135)}{=} \|\alpha\|_{\ell^2(\mathbb{N})} \sqrt{k(x, x)} \leq \|\alpha\|_{\ell^2(\mathbb{N})} \sqrt{M} < \infty. \quad (6.143)$$

Man beachte, dass aus (6.143) insbesondere folgt, dass jedes $f \in \mathcal{H}$ beschränkt ist und wegen der Endlichkeit des Maßraums (Ω, μ) somit ein Element von $L^2(\Omega, \mu)$ ist. Damit folgt mit der Orthogonalität der Eigenfunktionen $(e_i)_{i \in \mathbb{N}}$ in $L^2(\Omega, \mu)$, dass die Darstellung $f = \sum_{i \in I} \alpha_i \sqrt{\lambda_i} e_i \in \mathcal{H}$ durch $\alpha := (\alpha_i)_{i \in \mathbb{N}} \in \ell^2(\mathbb{N})$ eindeutig bestimmt ist, denn angenommen, es gäbe ein weiteres $\alpha^* := (\alpha_i^*)_{i \in \mathbb{N}} \in \ell^2(\mathbb{N})$ mit $f = \sum_{i=1}^{\infty} \alpha_i^* \sqrt{\lambda_i} e_i \in \mathcal{H}$, dann gilt für alle $j \in \mathbb{N}$:

$$0 \stackrel{A.3.1}{=} \langle 0, e_j \rangle_{L^2(\Omega, \mu)} = \langle f - f, e_j \rangle_{L^2(\Omega, \mu)} \stackrel{(6.140)}{=} \left\langle \sum_{i=1}^{\infty} (\alpha_i - \alpha_i^*) \sqrt{\lambda_i} e_i, e_j \right\rangle_{L^2(\Omega, \mu)} \quad (6.144)$$

$$\stackrel{A.3.6}{=} \sum_{i=1}^{\infty} (\alpha_i - \alpha_i^*) \sqrt{\lambda_i} \langle e_i, e_j \rangle_{L^2(\Omega, \mu)} \stackrel{A.3.12}{=} (\alpha_j - \alpha_j^*) \sqrt{\lambda_j}. \quad (6.145)$$

Daraus wiederum folgt, wegen der Positivität der Eigenwerte $(\lambda_i)_{i \in \mathbb{N}}$, dass $\alpha_j = \alpha_j^*$ für alle $j \in \mathbb{N}$ gilt, ergo $\alpha = \alpha^*$ und damit die Eindeutigkeit der Darstellung von f durch α .

Im Folgenden zeige man nun, dass $\mathcal{H}$ tatsächlich wie behauptet ein Hilbertraum ist. $\mathcal{H}$ ist zunächst ein Unterraum von $\mathcal{F}(\Omega)$, denn für alle $f, \tilde{f} \in \mathcal{H}$ und $\eta, \xi \in \mathbb{K}$ gilt:

$$\eta f + \xi \tilde{f} \stackrel{(6.140)}{=} \eta \sum_{i=1}^{\infty} \alpha_i \sqrt{\lambda_i} e_i + \xi \sum_{i=1}^{\infty} \tilde{\alpha}_i \sqrt{\lambda_i} e_i = \sum_{i=1}^{\infty} (\eta \alpha_i + \xi \tilde{\alpha}_i) \sqrt{\lambda_i} e_i = \sum_{i=1}^{\infty} \hat{\alpha}_i \sqrt{\lambda_i} e_i \in \mathcal{H}, \quad (6.146)$$

da $\hat{\alpha} := \eta \alpha + \xi \tilde{\alpha} \in \ell^2(\mathbb{N})$ wegen der Vektorraumeigenschaft von $\ell^2(\mathbb{N})$.

Außerdem definiert die Bilinearform

$$\langle \cdot, \cdot \rangle_{\mathcal{H}} : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{K}, \quad \langle f, \tilde{f} \rangle_{\mathcal{H}} := \langle \alpha, \tilde{\alpha} \rangle_{\ell^2(\mathbb{N})} \quad \text{für } f, \tilde{f} \in \mathcal{H} \quad (6.147)$$

ein Skalarprodukt auf $\mathcal{H}$. Dazu sei zunächst angemerkt, dass wegen der eindeutigen Darstellung von f durch α die Bilinearform wohldefiniert ist. Die Eigenschaften (S1-S5) aus Def. A.3.1 folgen dann umgekehrt, weil $\langle \cdot, \cdot \rangle_{\ell^2(\mathbb{N})}$ selbst ein Skalarprodukt ist. Für $f, \tilde{f}, f^* \in \mathcal{H}$ und $\xi \in \mathbb{K}$ gilt:

$$(S1 + S2) : \quad \langle \xi f + \tilde{f}, f^* \rangle_{\mathcal{H}} \stackrel{(6.147)}{=} \langle \xi \alpha + \tilde{\alpha}, \alpha^* \rangle_{\ell^2(\mathbb{N})} \stackrel{A.3.1}{=} \xi \langle \alpha, \alpha^* \rangle_{\ell^2(\mathbb{N})} + \langle \tilde{\alpha}, \alpha^* \rangle_{\ell^2(\mathbb{N})}, \quad (6.148)$$

$$\stackrel{(6.147)}{=} \xi \langle f, f^* \rangle_{\mathcal{H}} + \langle \tilde{f}, f^* \rangle_{\mathcal{H}}, \quad (6.149)$$

$$(S3) : \quad \langle f, \tilde{f} \rangle_{\mathcal{H}} \stackrel{(6.147)}{=} \langle \alpha, \tilde{\alpha} \rangle_{\ell^2(\mathbb{N})} \stackrel{A.3.1}{=} \overline{\langle \tilde{\alpha}, \alpha \rangle_{\ell^2(\mathbb{N})}} \stackrel{(6.147)}{=} \overline{\langle \tilde{f}, f \rangle_{\mathcal{H}}}, \quad (6.150)$$

$$(S4) : \quad \langle f, f \rangle_{\mathcal{H}} \stackrel{(6.147)}{=} \langle \alpha, \alpha \rangle_{\ell^2(\mathbb{N})} \stackrel{A.3.1}{\geq} 0 \quad \text{und} \quad (6.151)$$

$$(S5) : \quad \langle f, f \rangle_{\mathcal{H}} \stackrel{(6.147)}{=} \langle \alpha, \alpha \rangle_{\ell^2(\mathbb{N})} = 0 \stackrel{A.3.1}{\Leftrightarrow} \alpha = 0 \stackrel{(6.140)}{\Leftrightarrow} f = 0. \quad (6.152)$$

Damit ist $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ gemäß Def. A.3.1 als euklidischer bzw. unitärer Vektorraum nachgewiesen. Insbesondere bemerke man, dass für die Norm aller $f \in \mathcal{H}$ gilt:

$$\|f\|_{\mathcal{H}} \stackrel{A.3.3}{=} \sqrt{\langle f, f \rangle_{\mathcal{H}}} \stackrel{(6.147)}{=} \sqrt{\langle \alpha, \alpha \rangle_{\ell^2(\mathbb{N})}} \stackrel{A.3.3}{=} \|\alpha\|_{\ell^2(\mathbb{N})}. \quad (6.153)$$

Die Vollständigkeit von $\mathcal{H}$ ist damit dann auch leicht einzusehen: Sei dazu $\varepsilon > 0$ und $(f_n)_{n \in \mathbb{N}}$ eine Cauchyfolge in $\mathcal{H}$. Dann existiert gemäß Def.+Satz A.1.11 ein n_ε , sodass für alle $n, m \geq n_\varepsilon$ gilt, dass $\|f_n - f_m\|_{\mathcal{H}} \leq \varepsilon$ ist. Es folgt:

$$\|\alpha_n - \alpha_m\|_{\ell^2(\mathbb{N})} \stackrel{(6.153)}{=} \|f_n - f_m\|_{\mathcal{H}} \leq \varepsilon \quad (6.154)$$

Folglich ist $(\alpha_n)_{n \in \mathbb{N}}$ eine Cauchyfolge in $\ell^2(\mathbb{N})$. Dessen Vollständigkeit (siehe Bsp. A.2.11) liefert einen Grenzwert $\alpha := (\alpha_i)_{i \in \mathbb{N}} \in \ell^2(\mathbb{N})$ und damit ein $f = \sum_{i=1}^{\infty} \alpha_i \sqrt{\lambda_i} e_i \in \mathcal{H}$. Für dieses gilt dann

$$\|f - f_n\|_{\mathcal{H}} \stackrel{(6.153)}{=} \|\alpha - \alpha_n\|_{\ell^2(\mathbb{N})} \rightarrow 0 \quad \text{für } n \rightarrow \infty. \quad (6.155)$$

Also konvergiert die Folge $(f_n)_{n \in \mathbb{N}}$ in $\mathcal{H}$, was diesen vollständig macht gemäß Def. A.1.17, denn $(f_n)_{n \in \mathbb{N}}$ war beliebig gewählt und ergo ist $\mathcal{H}$ wie behauptet als Hilbertraum nachgewiesen.

Um einzusehen, dass $\mathcal{H} \subseteq \mathcal{F}(\Omega)$ sogar ein RKHS ist, muss definitionsgemäß noch das Auswertungsfunktional geprüft werden. Für alle $f \in \mathcal{H}$, $x \in \Omega$ gilt:

$$|E_x(f)| \stackrel{(3.1)}{=} |f(x)| \stackrel{(6.143)}{\leq} \|\alpha\|_{\ell^2(\mathbb{N})} \sqrt{M} \stackrel{(6.147)}{=} \sqrt{M} \|f\|_{\mathcal{H}}. \quad (6.156)$$

Gemäß Def.+Satz A.4.2 ist das Auswertungsfunktional somit für jede Stelle $x \in \Omega$ beschränkt und $\mathcal{H}$ gemäß Def. 3.1.1 ein RKHS.

Es verbleibt den reproduzierenden Kern des RKHS $\mathcal{H}$ zu bestimmen. Dazu bemerke man, dass für alle $x \in \Omega$ die Funktion $k(\cdot, x) = \sum_{i=1}^{\infty} \lambda_i e_i \overline{e_i(x)} = \sum_{i=1}^{\infty} (\sqrt{\lambda_i} e_i(x)) \sqrt{\lambda_i} e_i$ in $\mathcal{H}$ liegt, denn es gilt:

$$\left\| \left(\sqrt{\lambda_i} e_i(x) \right)_{i \in \mathbb{N}} \right\|_{\ell^2(\mathbb{N})} \stackrel{A.2.11}{=} \sum_{i=1}^{\infty} |\sqrt{\lambda_i} e_i(x)|^2 \stackrel{6.2.1}{\stackrel{A.8.1}{=}} \sum_{i=1}^{\infty} \lambda_i e_i(x) \overline{e_i(x)} \stackrel{(6.135)}{=} k(x, x) \leq M < \infty. \quad (6.157)$$

Folglich ist $(\sqrt{\lambda_i} \overline{e_i(x)})_{i \in \mathbb{N}} \in \ell^2(\mathbb{N})$ und $k(\cdot, x) \in \mathcal{H}$ für alle $x \in \Omega$. Es folgt für alle $f \in \mathcal{H}, x \in \Omega$:

$$\langle f, k(\cdot, x) \rangle_{\mathcal{H}} \stackrel{(6.147)}{=} \left\langle \alpha, \left(\sqrt{\lambda_i} \overline{e_i(x)} \right)_{i \in \mathbb{N}} \right\rangle_{\ell^2(\mathbb{N})} \stackrel[\substack{\text{A.2.11} \\ \text{6.2.1}}]{=} \sum_{i=1}^{\infty} \alpha_i \sqrt{\lambda_i} e_i(x) = f(x). \quad (6.158)$$

Mit diesem letzten Ergebnis ist bestätigt, dass k die reproduzierende Eigenschaft (3.19) für den RKHS $\mathcal{H}$ erfüllt. Kor. 3.2.6 besagt aber, dass der durch k reproduzierte RKHS eindeutig ist. Es folgt somit die Behauptung $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k}) = (\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$. $\square$

Korollar 6.8.5 (Orthonormalbasis des Mercer-RKHS [27, Theorem 11.18]).

Seien die Voraussetzungen wie in Kor. 6.8.4. Dann ist $(\sqrt{\lambda_i} e_i)_{i \in I}$ eine Orthonormalbasis von $\mathcal{H}_k$.

Beweis. O.B.d.A. sei wie im Beweis zu Kor. 6.8.4 $I = \mathbb{N}$ angenommen. Wie in Selbigem gezeigt, besitzt jedes $f \in \mathcal{H}_k$ eine eindeutige Darstellung $f = \sum_{i=1}^{\infty} \alpha_i \sqrt{\lambda_i} e_i$ mit einem (ebenfalls eindeutigen) $\alpha := (\alpha_i)_{i \in \mathbb{N}} \in \ell^2(\mathbb{N})$. Es folgt damit für jedes $f \in \mathcal{H}_k$ (zusammen mit der kanonischen $\ell^2(\mathbb{N})$ -Orthonormalbasis/Schauderbasis $s_1 := (1, 0, 0, \dots)$, $s_2 := (0, 1, 0, \dots)$, $\dots$):

$$\sum_{i=1}^{\infty} \langle f, \sqrt{\lambda_i} e_i \rangle_{\mathcal{H}_k} \sqrt{\lambda_i} e_i \stackrel{(6.140)}{=} \sum_{i=1}^{\infty} \langle \alpha, s_i \rangle_{\ell^2(\mathbb{N})} \sqrt{\lambda_i} e_i \stackrel{\text{A.2.11}}{=} \sum_{i=1}^{\infty} \alpha_i \sqrt{\lambda_i} e_i = f. \quad (6.159)$$

Da dies für alle $f \in \mathcal{H}_k$ gilt, folgt die Behauptung mit Def.+Satz A.3.13. $\square$

Korollar 6.8.6 (Mercer-RKHS sind separabel [27, Theorem 11.18]).

Seien die Voraussetzungen wie in Kor. 6.8.5. Dann ist der RKHS $\mathcal{H}_k$ separabel.

Beweis. Dies ist eine direkte Folge aus Kor. 6.8.5 und Def.+Satz A.3.13. $\square$

Wie Kor. 6.8.4, Kor. 6.8.5 und Kor. 6.8.6 aufzeigen, hat der RKHS sehr gute Eigenschaften, wenn er durch einen sog. **Mercer-Kern** reproduziert wird. Besonders die Orthonormalbasis und die damit einhergehende Separabilität des RKHS sind hilfreich. Als ein solcher Hilbertraum ist der RKHS isometrisch-isomorph zum $\ell^2(\mathbb{N})$ und die Isometrie Ψ lässt sich sogar aus Kor. 6.8.4 direkt ablesen:

$$\Psi : \ell^2(I) \rightarrow \mathcal{H}_k, \quad \Psi(\alpha) := \left\langle \alpha, \left(\sqrt{\lambda_i} \overline{e_i} \right)_{i \in I} \right\rangle_{\ell^2(I)} \quad \text{für } \alpha \in \ell^2(I). \quad (6.160)$$

Es scheint somit, dass sich mit Mercer-Kernen viel einfacher rechnen lässt. Zumindest haben diese eine gute Abrundung dieses Kapitels und der gesamten Arbeit zugelassen. Es sei sich an dieser Stelle herzlich für das aufmerksame Lesen bedankt und für ein Fazit auf die Zusammenfassung bzw. den Abstract verwiesen.

Hauptergebnis des Kapitels:



Mithilfe verschiedener Versionen des **Satzes von Mercer** erhält der Entwickler in Aufgabenstellungen des Maschinellen Lernens mit dem $\mathcal{H} = \ell^2(\mathbb{N})$ einen ggf. handlicheren Hilbertraum als den etwas unbekannten Teilraum von $\mathcal{F}(X)$ im Falle eines **RKHS** / dem **Satz 4.2.1 von Moore**. Unter gewissen Zusatzbedingungen lässt sich $\mathcal{H}$ sogar durch den **endlich-dimensionalen** $\mathbb{K}^n$ approximieren.

Literatur

- [1] Wolfgang Arendt. *Heat Kernels*. 2006. URL: https://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.020/arendt/downloads/internetseminar.pdf (besucht am 03. 07. 2025).
- [2] Nachman Aronszajn. „Theory of Reproducing Kernels“. In: *Transactions of the American Mathematical Society* 68.3 (1950), S. 337–404. ISSN: 0002-9947. DOI: [10.2307/1990404](https://doi.org/10.2307/1990404). URL: <https://www.jstor.org/stable/1990404>.
- [3] Heinz Bauer. *Maß- und Integrationstheorie*. 2. überarb. Aufl. 1992. Reprint 2010. Berlin ; Boston: De Gruyter, 2011. ISBN: 9783110871739. DOI: [10.1515/9783110871739](https://doi.org/10.1515/9783110871739). URL: <https://doi.org/10.1515/9783110871739>.
- [4] Wolfgang Beekmann. *Analysis*. Fernuniversität in Hagen, 2022.
- [5] Patrizio Bifulco. „Contributions to the Diffusion on Graphs and Networks“. Diss. FernUniversität in Hagen, 2025.
- [6] Patrizio Bifulco und Joachim Kerner. „Some spectral comparison results on infinite quantum graphs“. In: *Journal of Mathematical Physics* 65.7 (Juli 2024), S. 073502. ISSN: 0022-2488. DOI: [10.1063/5.0178226](https://doi.org/10.1063/5.0178226). eprint: https://pubs.aip.org/aip/jmp/article-pdf/doi/10.1063/5.0178226/20026684/073502_1_5.0178226.pdf. URL: <https://doi.org/10.1063/5.0178226>.
- [7] Marc Burger. *Functional Analysis I*. 2023. URL: https://metaphor.ethz.ch/x/2023/hs/401-3461-00L/sc/FA1_notes30112.pdf (besucht am 03. 07. 2025).
- [8] F. Conrad, copper.hat und Chen Zhongyuan. *L^q is always dense in L^p ?* 2022. URL: <https://math.stackexchange.com/questions/4396181/> (besucht am 20. 10. 2025).
- [9] David Corne und Michael Lones. *Evolutionary Algorithms*. 2018. URL: <https://arxiv.org/pdf/1805.11014> (besucht am 05. 08. 2025).
- [10] Edward B. Davies. *Heat kernels and spectral theory*. Cambridge tracts in mathematics. Cambridge [u.a.]: Cambridge Univ. Pr, 1989. ISBN: 0521361362. URL: https://digitale-objekte.hbz-nrw.de/storage/2007/11/27/file_158/2237911.pdf.
- [11] Josse van Dobben de Bruyn. *Square root of normal positive operators over real Hilbert spaces*. 2018. URL: [Square%20root%20of%20normal%20positive%20operators%20over%20real%20Hilbert%20spaces](https://arxiv.org/abs/1805.11014) (besucht am 20. 10. 2025).
- [12] Jürgen Elstrodt. *Maß- und Integrationstheorie*. 8. Aufl. 2018. Berlin, Heidelberg: Springer Berlin Heidelberg, Imprint: Springer Spektrum, 2018. ISBN: 9783662579398. DOI: [10.1007/978-3-662-57939-8](https://doi.org/10.1007/978-3-662-57939-8). URL: <https://doi.org/10.1007/978-3-662-57939-8>.

- [13] Peter Haggstrom. *A combinatorial view of the orthogonality of Rademacher functions*. 2016. URL: <https://www.gotohaggstrom.com/A%20combinatorial%20view%20of%20the%20orthogonality%20of%20Rademacher%20functions.pdf> (besucht am 30.05.2025).
- [14] John Kelso Hunter. *Measure theory*. 2011. URL: https://www.math.ucdavis.edu/~hunter/measure_theory/measure_notes.pdf (besucht am 20.10.2025).
- [15] Ansgar Jüngel. *Eine Einführung in die Halbgruppentheorie*. Universität Konstanz, 2001.
- [16] Steffen Kionke. *Lineare Algebra*. Fernuniversität in Hagen, 2023.
- [17] Werner Kirsch. *Maß- und Integrationstheorie*. Fernuniversität in Hagen, 2014.
- [18] Hermann König. *Eigenvalue Distribution of Compact Operators*. booklet. Birkhäuser Basel, 1986. DOI: [10.1007/978-3-0348-6278-3](https://doi.org/10.1007/978-3-0348-6278-3). URL: <https://doi.org/10.1007/978-3-0348-6278-3>.
- [19] Pietro Majer, Delio Mugnolo und Paul Garrett. *Mercer’s Theorem: uniform L^∞ bound on eigenfunctions?* 2011. URL: <https://mathoverflow.net/questions/72573/> (besucht am 25.06.2025).
- [20] James Mercer und Andrew Russell Forsyth. „XVI. Functions of positive and negative type, and their connection the theory of integral equations“. In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 209.441-458 (1909), S. 415–446. DOI: [10.1098/rsta.1909.0016](https://doi.org/10.1098/rsta.1909.0016). eprint: <https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.1909.0016>. URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rsta.1909.0016>.
- [21] Eliakim Hastings Moore. *General Analysis*. Hrsg. von Raymond Walter Barnard. Philadelphia: The American philosophical society, 1935.
- [22] Delio Mugnolo. „Diffusion problems on metric graphs“. In: *CIMPA School 2025 Transport optimal, EDP et optimisation applications en sciences des données*. 2025.
- [23] Delio Mugnolo. *Re: [E-Mails Juli 2025]*. 2025.
- [24] Delio Mugnolo und Joachim Kerner. *Partielle Differentialgleichungen*. Fernuniversität in Hagen, 2024.
- [25] Delio Mugnolo, Joachim Kerner und Hafida Laasri. *Funktionalanalysis*. Fernuniversität in Hagen, 2024.
- [26] Delio Mugnolo und Bingying Lu. *satz von mercer*. [E-Mails Juni/Juli 2025]. 2025.
- [27] Vern I. Paulsen und Mrinal Raghupathi. *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2016.
- [28] Albrecht Pietsch. *Nukleare lokalkonvexe Räume*. Akademie-Verlag Berlin, 1969.
- [29] Michael Reed und Barry Simon. *Methods of Modern Mathematical Physics: Functional analysis*. Methods of Modern Mathematical Physics Bd. 1. Academic Press, 1980. ISBN: 9780125850506. URL: <https://books.google.de/books?id=bvuRuwuFBWwC>.
- [30] Reinhold Remmert und Hans Rosen. *Funktionentheorie*. Fernuniversität in Hagen, 2025.
- [31] Walter Rudin. *Functional Analysis*. McGraw-Hill, 1991. ISBN: 9780070542365.

- [32] Bernhard Schölkopf. *Support Vector Learning*. 1997. URL: https://pure.mpg.de/rest/items/item_1794215_3/component/file_3214422/content (besucht am 28.06.2025).
- [33] Bernhard Schölkopf und Alexander J. Smola. *Learning with Kernels - Support Vector Machines, Regularization, Optimization, and Beyond*. Massachusetts Institute of Technology, 2002.
- [34] Vaikalathur Shankar Sunder. *Functional Analysis: Spectral Theory*. 2000. URL: <https://www.imsc.res.in/~sunder/fa.pdf> (besucht am 19.10.2025).
- [35] Luise Unger. *Mathematische Grundlagen*. Fernuniversität in Hagen, 2023.
- [36] user159517 und user127022. *Uniform convergence in Mercer Theorem for bounded kernels*. 2016. URL: <https://mathoverflow.net/questions/254385/> (besucht am 25.06.2025).

Appendix



Funktionalanalytische Grundlagen

Die Grundlage dieses Anhangs ist [17, 25, 31], aus welchen der überwiegende Teil der Definitionen und Sätze/Theoreme entnommen und ggf. lediglich aus dem Englischen ins Deutsche übersetzt wurden. Die Beweise wurden dabei nicht kopiert, bis auf die Fälle in denen die Beweise selbst in der Argumentation der Arbeit benötigt wurden, können aber den Originalquellen entnommen werden.

A.1 Metrische Räume – Vollständige Räume

Definition A.1.1 (Metrik / Metrischer Raum [25, Def. 1.1.1]).

Eine **Metrik** d auf einer Menge $X \neq \emptyset$ ist eine Abbildung

$$\begin{aligned} d : X \times X &\rightarrow \mathbb{R} \\ (x, y) &\mapsto d(x, y) \end{aligned}$$

sodass für alle x, y, z in X die folgenden Bedingungen erfüllt sind:

$$\text{(M1)} \quad d(x, y) = 0 \Leftrightarrow x = y$$

$$\text{(M2)} \quad d(x, y) = d(y, x)$$

$$\text{(M3)} \quad d(x, z) \leq d(x, y) + d(y, z)$$

Das Paar (X, d) bestehend aus einer Menge $X \neq \emptyset$ und einer Metrik d auf X wird **metrischer Raum** genannt. Man nennt $d(x, y)$ den Abstand zwischen x und y (respektive/bzgl. der Metrik d).

Bemerkung A.1.2 (Umgekehrte Dreiecks-Ungleichung [25, Remark 1.1.2]).

Sei (X, d) ein metrischer Raum. Dann gelten die Ungleichungen.

$$|d(x, z) - d(y, z)| \leq d(x, y) \quad \text{für alle } x, y, z \in X \tag{A.1}$$

und

$$|d(x, y) - d(u, v)| \leq d(x, u) + d(y, v) \quad \text{für alle } x, y, u, v \in X. \quad (\text{A.2})$$

Definition A.1.3 (Offene Kugel in metrischem Raum [25, 1.2]).

Sei (X, d) ein metrischer Raum. Für ein $x \in X$ und ein $r > 0$ wird die Menge

$$B(x, r) := \{y \in M : d(x, y) < r\}$$

die **offene Kugel mit Radius r und Zentrum x** genannt.

Definition A.1.4 (Offene / abgeschlossene Menge [25, Def. 1.2.3, Def 1.2.7]).

Sei (X, d) ein metrischer Raum.

Eine Menge $O \subseteq X$ heißt **offen**, wenn für alle $x \in O$ ein $r > 0$ existiert, sodass $B(x, r) \subseteq O$ gilt.

Eine Menge $F \subseteq X$ heißt **abgeschlossen**, wenn $X \setminus F$ offen ist.

Definition A.1.5 (Inneres/Abschluss/Rand [25, Def. 1.2.9]).

Sei (X, d) ein metrischer Raum und sei $A \subseteq X$.

1. Die Menge $A^\circ := \{x \in X : \exists r > 0 \text{ sodass } B(x, r) \subseteq A\}$ wird das **Innere von A** genannt. Die Elemente von A° heißen **innere Punkte**.
2. Die Menge $\bar{A} := \{x \in X : \forall r > 0 \text{ [gilt] } B(x, r) \cap A \neq \emptyset\}$ heißt der **Abschluss von A** .
3. Die Menge $\partial A := \bar{A} \cap \overline{(X \setminus A)}$ wird der **Rand von A** genannt.

Proposition A.1.6 (Inneres/Abschluss [25, Prop. 1.2.10]).

Sei (X, d) ein metrischer Raum und sei $A \subseteq X$. Dann gilt

1. Das Innere A° ist die größte offene Teilmenge von A .
2. Der Abschluss $\bar{A}$ ist die kleinste abgeschlossene Menge, für die A eine Teilmenge ist.

Definition A.1.7 (Dichte Menge [25, Def. 1.2.9]).

Sei (X, d) ein metrischer Raum. Eine Menge A heißt **dicht in X** , wenn $\bar{A} = X$ gilt.

Definition A.1.8 (Separabler Raum [25, Def. 1.6.1]).

Ein metrischer Raum heißt **separabel**, wenn er eine abzählbare und dichte Teilmenge enthält.

Definition A.1.9 (Kompakter Raum [25, Def. 1.8.1]).

Ein metrischer Raum (X, d) heißt **kompakt**, wenn jede offene Überdeckung von X eine endliche Teilüberdeckung hat, d. h. dass für alle Familien von offenen Mengen $\{O_i : i \in J\}$ mit $X = \bigcup_{i \in J} O_i$ eine endliche Teilmenge $I \subset J$ existiert, sodass $X = \bigcup_{i \in I} O_i$ gilt (J ist dabei beliebige Indexmenge).

Definition A.1.10 (Konvergente Folge [25, Def. 1.3.1]).

Sei (X, d) ein metrischer Raum. Man sagt eine Folge $(x_n)_{n \in \mathbb{N}} \subseteq X$ **konvergiert** zu einem Element $x \in X$ – und man schreibt $\lim_{n \rightarrow \infty} x_n = x$, $x_n \rightarrow x$ für $n \rightarrow \infty$, oder einfach $x_n \rightarrow x$ – wenn für alle $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ existiert, sodass $d(x_n, x) \leq \varepsilon$ für alle $n \geq n_0$ [ist]. Wenn $x_n \rightarrow x$, dann wird x der **Grenzwert** von $(x_n)_{n \in \mathbb{N}}$ genannt.

Definition und Satz A.1.11 (Cauchy-Folge [25, 1.3.2]).

Sei (X, d) ein metrischer Raum. Eine Folge $(x_n)_{n \in \mathbb{N}} \subseteq X$ ist eine **Cauchy-Folge**, wenn für alle $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ existiert, sodass

$$d(x_n, x_m) \leq \varepsilon \quad \text{für alle } n, m \geq n_0.$$

Jede konvergente Folge ist eine Cauchy-Folge.

Definition und Satz A.1.12 (Stetige Abbildung [25, Def. 1.4.1, Theorem 1.4.2, Prop. 1.4.4]).

Seien (X, d_x) und (Y, d_y) zwei metrische Räume. Eine Abbildung $f : X \rightarrow Y$ heißt **stetig im Punkt x_0** , wenn diese beiden äquivalenten Bedingungen erfüllt sind:

1. Für alle $\varepsilon > 0$ existiert ein $\delta_{x_0, \varepsilon} > 0$, sodass

$$d_X(x_0, x) \leq \delta \Rightarrow d_Y(f(x_0), f(x)) \leq \varepsilon.$$

2. f ist **folgenstetig** in x_0 , d. h. für jede Folge $(x_n)_{n \in \mathbb{N}} \subseteq X$, die gegen x_0 konvergiert, gilt $\lim_{n \rightarrow \infty} f(x_n) = f(x_0)$.

Die Funktion f heißt **stetig (auf X)**, wenn eine der drei äquivalenten Bedingungen erfüllt ist:

1. f ist stetig in jedem Punkt $x_0 \in X$.
2. $f^{-1}(O)$ ist offen für jede offene Teilmenge $O \subseteq Y$.
3. $f^{-1}(F)$ ist abgeschlossen für jede abgeschlossene Teilmenge $F \subseteq Y$.

Definition A.1.13 (Gleichmäßig stetige Abbildung [25, Def. 1.4.7]).

Seien $(X, d_X), (Y, d_Y)$ zwei metrische Räume. Dann heißt $f : X \rightarrow Y$ **gleichmäßig stetig**, falls für alle $\varepsilon > 0$ ein $\delta_\varepsilon > 0$ existiert, sodass gilt:

$$d_Y(f(x), f(y)) \leq \varepsilon \quad \text{für alle } x, y \in X, \text{ welche } d(x, y) < \delta_\varepsilon \text{ erfüllen.}$$

Proposition A.1.14 (Metriken sind stetig [25, Example 1.4.3]).

Sei X eine Menge. Dann ist jede Metrik $d : X \times X \rightarrow \mathbb{R}$ eine stetige Funktion von $(X \times X, d)$ nach $(\mathbb{R}, |\cdot|)$.

Proposition A.1.15 (Stetige Funktionen auf kompakten Räumen [25, Prop. 1.8.20]).

Seien $(X, d_X), (Y, d_Y)$ zwei metrische Räume und $f : X \rightarrow Y$ eine stetige Funktion. Dann ist für jede kompakte Teilmenge $K \subset X$ das Bild $f(K)$ kompakt und f ist gleichmäßig stetig von K nach Y .

Satz A.1.16 (Satz von Dini [25, Lemma 2.4.5]).

Sei K ein kompakter metrischer Raum und seien $(f_n)_{n \in \mathbb{N}} \subset C(K, \mathbb{R})$ so, dass

$$f_n(x) \leq f_{n+1}(x), \quad n \in \mathbb{N}, x \in K.$$

Konvergiert die Folge punktweise gegen ein $f \in C(K, \mathbb{R})$, dann ist die Konvergenz gleichmäßig.

Definition A.1.17 (Vollständiger Raum [25, Def. 1.3.8]).

Man nennt einen metrischen Raum (X, d) **vollständig**, wenn jede [darin enthaltene] Cauchy-Folge konvergiert/konvergent ist.

Definition und Satz A.1.18 (Isometrie [25, Prop. 1.8.23]).

Eine Funktion f zwischen zwei metrischen Räumen (X, d_X) und (Y, d_Y) ist eine **Isometrie**, wenn

$$d_Y(f(x), f(y)) = d_X(x, y) \quad \text{für alle } x, y \in X.$$

f ist dann sowohl injektiv als auch stetig. Ist X vollständig, so ist es auch $f(X)$.

Definition und Satz A.1.19 (Vervollständigung [25, Def. 2.8.1, Theorem 2.8.2]).

Ein vollständiger metrischer Raum $(\hat{X}, \hat{d})$ heißt die **Vervollständigung** eines gegebenen metrischen Raumes (X, d) , wenn eine Isometrie $\phi : X \rightarrow \hat{X}$ existiert, sodass $\phi(X)$ dicht in $\hat{X}$ liegt.

Bis auf [isomorphe] Isometrie besitzt jeder metrische Raum (X, d) genau eine Vervollständigung $(\hat{X}, \hat{d})$.

A.2 Normierte Räume – Banachräume

Definition A.2.1 (Vektorraum [4, Def. 1.3.3]).

Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und X eine nicht-leere Menge. Auf X sei eine **Addition** erklärt, d. h. je zwei Elemente $x, y \in X$ sei eindeutig ein Element $x + y$ zugeordnet, mit folgenden Eigenschaften:

(Add₁) Für alle $x, y, z \in X$ gilt $(x + y) + z = x + (y + z)$ (Assoziativität).

(Add₂) Es gibt ein eindeutig bestimmtes Element $0 \in X$, genannt **neutrales Element** oder **Nullvektor**, mit $x + 0 = 0 + x = x$ für alle $x \in X$ (Existenz des neutralen Elements).

(Add₃) Zu jedem Element $x \in X$ gibt es ein eindeutig bestimmtes Element $-x \in X$, das zu x **inverse Element** oder der zu x **entgegengesetzte Vektor**, mit

$$x + (-x) = (-x) + x = 0 \quad (\text{Existenz der inversen Elemente}).$$

(Add₄) Für alle $x, y \in X$ gilt $x + y = y + x$ (Kommutativität).

Auf X sei ferner eine **skalare Multiplikation** erklärt, d. h. für je ein Skalar $\alpha \in \mathbb{K}$ und je einem $x \in X$ sei eindeutig ein Element $\alpha x \in X$ zugeordnet, mit folgenden Eigenschaften:

(Mult₁) Für alle $\alpha, \beta \in \mathbb{K}$ und $x \in X$ gilt $\alpha(\beta x) = (\alpha\beta)x$ (Assoziativität).

(Mult₂) Für alle $x, y \in X$ und $\alpha, \beta \in \mathbb{K}$ gilt
 $\alpha(x + y) = \alpha x + \alpha y, \quad (\alpha + \beta)x = \alpha x + \beta x$ (Distributivität).

(Mult₃) Für alle $x \in X$ gilt $1x = x$.

Dann heißt X (mit dieser Addition und skalaren Multiplikation) ein **Vektorraum** oder **linearer Raum** (über $\mathbb{K}$). Seine Elemente werden als **Vektoren** bezeichnet.

Beispiel A.2.2 (Der Vektorraum $\mathcal{F}(X, \mathbb{K})$ [25, Example 2.1.5]).

Es sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ der reelle oder komplexe Zahlenkörper, X eine beliebige Menge und mit $\mathcal{F}(X, \mathbb{K}) := \{f : X \rightarrow \mathbb{K}\}$ der Raum der Funktionen/Abbildungen bezeichnet, die aus dem Definitionsbereich X in den Zahlenkörper $\mathbb{K}$ abbilden. Dann ist $\mathcal{F}(X, \mathbb{K})$ ein Vektorraum mit der punktweisen Addition und punktweisen skalaren Multiplikation:

$$(f + g)(x) := f(x) + g(x) \quad \forall x \in X, \quad (\text{A.3})$$

$$(\alpha f)(x) := \alpha f(x), \quad \forall x \in X, \alpha \in \mathbb{K}. \quad (\text{A.4})$$

Definition und Satz A.2.3 (Lineare Hülle [35, Def. 6.3.2, Prop. 6.3.4]).

Sei S eine nicht-leere Teilmenge eines Vektorraums V (über $\mathbb{K}$). Die Menge aller Linearkombinationen von endlich vielen Vektoren in S wird die **lineare Hülle von S** , das **Erzeugnis von S** oder der **Aufspann von S** genannt, und mit $\langle S \rangle$ oder $\text{span } S$ bezeichnet. Es gilt:

$$\text{span } S := \left\{ \sum_{i=1}^n \alpha_i v_i : n \in \mathbb{N}, \alpha_i \in \mathbb{K}, v_i \in S \text{ für } i \in \{1, \dots, n\} \right\}.$$

Zudem ist $\text{span } S$ immer ein Unter(vektor)raum von V [und damit bzgl. der Addition und skalaren Multiplikation endlich vieler Elemente abgeschlossen].

Definition A.2.4 (Norm – Normierter Raum [25, Def. 2.1.1]).

Sei X ein Vektorraum über dem Körper $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Eine **Norm** auf X ist eine Abbildung $\|\cdot\| : X \rightarrow [0, +\infty)$, sodass für alle $x, y \in X$ und $\alpha \in \mathbb{K}$ [gelten]:

$$(N1) \quad \|x\| = 0 \Leftrightarrow x = 0$$

$$(N2) \quad \|\alpha x\| = |\alpha| \|x\|$$

$$(N3) \quad \|x + y\| \leq \|x\| + \|y\|$$

Ein **normierter Raum** ist ein Paar $(X, \|\cdot\|)$ bei dem X ein Vektorraum über $\mathbb{K}$ und $\|\cdot\|$ eine Norm auf X ist.

Man verweist auf die Eigenschaften (N1), (N2), (N3) normalerweise, indem man sagt, dass eine Norm positiv definit, homogen ist, und dass sie die Dreiecksungleichung erfüllt. Eine Funktion $\|\cdot\|_s : X \rightarrow [0, +\infty)$, welche (N2), (N3) aber nicht (N1) erfüllt, wird **Seminorm** genannt.

Bemerkung A.2.5 (Metrik auf normiertem Raum [25, Theorem 2.1.4]).

Eine Norm *induziert* auf natürliche Art immer eine Metrik d auf X , indem man

$$d(x, y) := \|x - y\|, \quad x, y \in X$$

setzt. Ein normierter Raum ist somit immer auch ein metrischer Raum mit der so erhaltenen Metrik.

Definition und Satz A.2.6 (Äquivalente Normen [25, Def. 2.2.1, Theorem 2.2.3]).

Zwei Normen $\|\cdot\|_1$ und $\|\cdot\|_2$ auf einem Vektorraum X heißen **äquivalent**, wenn $\alpha, \beta > 0$ existieren, sodass

$$\alpha \|x\|_1 \leq \|x\|_2 \leq \beta \|x\|_1 \quad \text{für alle } x \in X. \quad (A.5)$$

Eine äquivalente Definition lautet: Eine Folge $(x_n)_{n \in \mathbb{N}} \subset X$ konvergiert bzgl. $\|\cdot\|_1$ genau dann, wenn sie bzgl. $\|\cdot\|_2$ konvergiert. In diesem Fall stimmen die Grenzwerte bzgl. $\|\cdot\|_1$ und $\|\cdot\|_2$ überein.

Satz A.2.7 (Endlich-Dimensionale Vektorräume [25, Theorem 2.2.6, Remark 2.2.7]).

Jeder endlich-dimensionale Vektorraum ist ein normierter Raum und alle Normen darauf sind zueinander äquivalent. Zudem folgt aus dem Satz von Bolzano-Weierstraß, dass alle endlich-dimensionalen Vektorräume vollständig bzgl. (einer und somit) aller auf ihnen definierten Normen sind.

Definition A.2.8 (Banachraum [25, Def. 2.6.1]).

Ein vollständiger normierter Raum wird **Banachraum** genannt.

Beispiel A.2.9 ($(\mathcal{F}^b(M), \|\cdot\|_\infty)$ [25, Example 2.1.5, Theorem 2.6.3]).

Sei M eine nicht-leere Menge. Man bezeichnet mit $\mathcal{F}^b(M, \mathbb{K})$ bzw. $\mathcal{F}^b(M)$ die Menge aller beschränkten Funktionen auf M mit Werten in $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. $\mathcal{F}^b(M)$ ist ein Banachraum ausgestattet mit der Norm

$$\|\cdot\|_\infty : \mathcal{F}^b(M) \rightarrow [0, \infty) \quad \|f\|_\infty := \sup_{x \in \Omega} |f(x)|, \quad f \in \mathcal{F}^b(M).$$

Konvergiert eine Funktionenfolge $(f_n)_{n \in \mathbb{N}}$ bzgl. der $\|\cdot\|_\infty$ -Norm, dann nennt man diese auch **gleichmäßig konvergent**.

Beispiel A.2.10 ($(C(\Omega), \|\cdot\|_\infty)$ [25, Example 2.1.5, Example 2.1.7, Prop. 2.6.6]).

Sei (Ω, d) ein kompakter metrischer Raum. Es werde mit $C(\Omega, \mathbb{K})$ bzw. $C(\Omega)$ der Raum der stetigen Funktionen auf Ω mit Werten in $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ bezeichnet. Die Menge $C(\Omega)$ ist ein abgeschlossener Unterraum von $\mathcal{F}^b(\Omega)$, da jede Linearkombination von stetigen Funktionen ebenfalls stetig ist.

Beispiel A.2.11 ($\ell^p(\mathbb{N})$ -Räume [25, Def. 2.1.2, Kor. 2.1.20]).

Sei $1 \leq p \leq \infty$. Dann sind die Folgenräume $(\ell^p(\mathbb{N}), \|\cdot\|_p)$ Banachräume mit $\ell^p(\mathbb{N}) := \{x = (x_n)_{n \in \mathbb{N}} \subset \mathbb{K} : \|x\|_p < \infty\}$ und

$$\|x\|_p := \begin{cases} \left(\sum_{n=1}^{\infty} |x_n|^p \right)^{\frac{1}{p}} & \text{für } 1 \leq p < \infty, \\ \sup_{n \in \mathbb{N}} |x_n| & \text{für } p = \infty. \end{cases} \quad (\text{A.6})$$

Beispiel A.2.12 ($c_0(\mathbb{N})$ [25, Example 2.1.6, Exercise 2.1.23]).

Der normierte Raum $(c_0(\mathbb{N}), \|\cdot\|_\infty)$ ist ein abgeschlossener Unterraum von $(\ell^\infty(\mathbb{N}), \|\cdot\|_\infty)$ mit:

$$c_0(\mathbb{N}) := \{(x_n)_{n \in \mathbb{N}} \in \ell^\infty(\mathbb{N}) : \lim_{n \rightarrow \infty} x_n = 0\}.$$

Satz A.2.13 (Höldersche Ungleichung für Folgen [25, Theorem 2.1.18]).

Die folgenden Aussagen gelten:

(1) Sei $x \in \ell^1(\mathbb{N})$ und $y \in \ell^\infty(\mathbb{N})$. Dann ist $xy \in \ell^1(\mathbb{N})$ und

$$\|xy\|_1 \leq \|x\|_1 \|y\|_\infty.$$

(2) Sei $p, q \in (1, \infty)$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Ist $x \in \ell^p(\mathbb{N})$ und $y \in \ell^q(\mathbb{N})$, dann ist $xy \in \ell^1(\mathbb{N})$ und

$$\|xy\|_1 \leq \|x\|_p \|y\|_q.$$

Beispiel A.2.14 ($L^p(\Omega, \mu)$ -Räume [25, 17, Def. 5.1.2, Theorem 5.1.10, Def. 7.3.1]).

Sei (Ω, μ) ein Maßraum und $1 \leq p \leq \infty$. Dann sind die Funktionsräume $(L^p(\Omega, \mu), \|\cdot\|_{L^p(\Omega, \mu)})$ Banachräume mit $L^p(\Omega, \mu) := \{f : \Omega \rightarrow \mathbb{K} : f \text{ ist messbar und } \|f\|_{L^p(\Omega, \mu)} < \infty\}$ und

$$\|f\|_{L^p(\Omega, \mu)} := \begin{cases} \left(\int_{\Omega} |f(x)|^p d\mu(x) \right)^{\frac{1}{p}} & \text{für } 1 \leq p < \infty, \\ \sup_{x \in \Omega \setminus N} |f(x)| & \text{für } p = \infty \text{ mit einer } \mu\text{-Nullmenge } N, \text{ d. h. } \mu(N) = 0. \end{cases} \quad (\text{A.7})$$

Für den Fall, dass (Ω, μ) ein **endlicher** Maßraum ist (d. h. $\mu(\Omega) < \infty$), dann gilt für $1 \leq p \leq q \leq \infty$, dass $L^q(\Omega, \mu) \subset L^p(\Omega, \mu)$ [25, 17, Exercise 5.2.6, Kor. 7.4.9].

Lemma A.2.15 (Höldersche Ungleichung für Funktionen [25, Lemma 5.1.4]).

Sei $1 \leq p \leq q \leq \infty$ gegeben, so dass $q^{-1} = 1 - p^{-1}$ mit der Konvention, dass $q = \infty$, falls $q^{-1} = 0$. Wenn $f \in L^p(\Omega, \mu)$ und $g \in L^q(\Omega, \mu)$, dann ist $f \cdot g \in L^1(\Omega, \mu)$ und

$$\|fg\|_{L^1(\Omega, \mu)} \leq \|f\|_{L^p(\Omega, \mu)} \|g\|_{L^q(\Omega, \mu)}. \quad (\text{A.8})$$

Definition A.2.16 (Absolut konvergente Reihe [25, Def. 2.6.1]).

Sei $(X, \|\cdot\|_X)$ ein normierter Raum. Für eine Folge $(x_n)_{n \in \mathbb{N}} \subset X$, heißt die Reihe $\sum_{n=1}^{\infty} x_n$ **absolut konvergent**, falls

$$\sum_{n=1}^{\infty} \|x_n\|_X < \infty.$$

Definition A.2.17 (Unbedingt konvergente Reihe [25, Def. 6.5.8]).

Sei $(X, \|\cdot\|_X)$ ein normierter Raum und I eine Indexmenge. Für eine Folge $(x_i)_{i \in I} \subset X$, sagt man dass die Reihe $\sum_{i \in I} x_i$ **unbedingt konvergiert** gegen ein $x \in X$, wenn

- (a) $I_0 := \{i \in I : x_i \neq 0\}$ höchstens abzählbar ist,
- (b) für jede Umordnung $I_0 = \{i_1, i_2, \dots\}$ gilt, dass

$$\sum_{n=1}^{\infty} x_{i_n} = x,$$

bzgl. der Norm $\|\cdot\|_X$. Das bedeutet, dass jede Umordnung dasselbe Ergebnis liefert.

Satz A.2.18 (Satz von Stone-Weierstraß [25, Theorem 2.4.1]).

Sei K eine kompakte Teilmenge des $\mathbb{R}^n$ und A eine Unteralgebra von $C(K, \mathbb{K})$ [d. h. $A \subseteq C(K, \mathbb{K})$ ist abgeschlossen bzgl. Vektorsumme, skalarer Multiplikation und Vektorprodukt], sodass

- A die konstanten Funktionen in $C(K, \mathbb{K})$ enthält,
- wenn $f \in A$ ist, dann ist $\overline{f} \in A$ und
- A punktetrennend in A ist [d. h. für alle $x \neq y$ existiert ein $g \in A$ mit $g(x) \neq g(y)$].

Dann liegt A dicht in $(C(K, \mathbb{K}))$.

A.3 Euklidische/Unitäre Räume – Hilberträume

Definition A.3.1 (Skalarprodukt – euklidischer/unitärer Vektorraum [25, Def. 6.1.1]).

Sei X ein Vektorraum über $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Eine Abbildung

$$\langle \cdot, \cdot \rangle_X : X \times X \rightarrow \mathbb{K}$$

wird **Skalarprodukt** auf X genannt, wenn für alle $x, y, z \in X$ und $\alpha \in \mathbb{K}$ das Folgende gilt:

$$(S1) \quad \langle x + y, z \rangle_X = \langle x, z \rangle_X + \langle y, z \rangle_X,$$

$$(S2) \quad \langle \alpha x, y \rangle_X = \alpha \langle x, y \rangle_X,$$

$$(S3) \quad \langle x, y \rangle_X = \overline{\langle y, x \rangle_X},$$

$$(S4) \quad \langle x, x \rangle_X \geq 0,$$

$$(S5) \quad \langle x, x \rangle_X = 0 \Leftrightarrow x = 0.$$

Wenn $\langle \cdot, \cdot \rangle_X$ ein Skalarprodukt auf X ist, dann wird $(X, \langle \cdot, \cdot \rangle_X)$ euklidischer Vektorraum im Fall von $\mathbb{K} = \mathbb{R}$ und ein unitärer Vektorraum im Fall von $\mathbb{K} = \mathbb{C}$ genannt.

Satz A.3.2 (Cauchy-Schwarzsche Ungleichung [25, Theorem 6.1.5]).

Sei $(X, \langle \cdot, \cdot \rangle)$ ein euklidischer/unitärer Vektorraum. Dann [gilt]

$$|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle$$

für alle $x, y \in X$. Zudem ist die Ungleichung genau dann streng, wenn x, y keine parallelen Vektoren sind, d. h. genau dann, wenn $x + \alpha y \neq 0$ für alle $\alpha \in \mathbb{K}$.

Beweis. ([25, Theorem 6.1.5] entnommen.)

Sei $\alpha \in \mathbb{K}$. Es gilt:

$$0 \leq \langle \alpha x + y, \alpha x + y \rangle = |\alpha|^2 \langle x, x \rangle + \langle y, y \rangle + 2\operatorname{Re}(\alpha \langle x, y \rangle),$$

wobei benutzt wurde, dass $z + \bar{z} = 2a$ für $z = a + ib$ [$\in \mathbb{C}$]. Wählt man $\alpha := -\frac{\overline{\langle x, y \rangle}}{\langle x, x \rangle}$ und berücksichtigt, dass $z \cdot \bar{z} = |z|^2$ [gilt], ergibt sich

$$0 \leq \langle \alpha x + y, \alpha x + y \rangle = \langle y, y \rangle - \frac{|\langle x, y \rangle|^2}{\langle x, x \rangle}$$

und somit die Behauptung. Man beobachte, dass die einzige Ungleichung im Beweis sich genau dann in eine Gleichung verwandelt, wenn $x + \alpha y = 0$, was [noch] zu beweisen war. $\square$

Bemerkung A.3.3 (Norm auf euklidischem/unitärem Vektorraum [25, Theorem 6.1.7]).

Ein Skalarprodukt *induziert* auf natürliche Art immer eine Norm $\|\cdot\|$ auf X indem man

$$\|x\| := \sqrt{\langle x, x \rangle}, \quad x, y \in X$$

setzt. Ein euklidischer/unitärer Vektorraum ist somit immer auch ein normierter Raum mit der so erhaltenen Norm (und mit **Bem. A.2.5** auch ein metrischer Raum). Aus diesem Grund wird ein euklidischer/unitärer Vektorraum über $\mathbb{R}$ oder $\mathbb{C}$ auch **Prä-Hilbertraum** genannt.

Bemerkung A.3.4 (Winkel in euklidischem/unitärem Vektorraum [25, Remark 6.1.6]).

Mit der Cauchy-Schwarzschen Ungleichung ist es möglich einen Winkel zwischen zwei Vektoren $\overrightarrow{xy} := y - x$ und $\overrightarrow{wv} := w - v$ via

$$\theta := \arccos \frac{\langle y - x, w - v \rangle}{\|y - x\| \cdot \|w - v\|}.$$

zu definieren. Man beachte, dass $\theta \in (\frac{\pi}{2}, \pi]$ [ist], wenn $\langle y - x, w - v \rangle < 0$.

Definition A.3.5 (Hilbertraum [25, Def. 6.1.8]).

Ein euklidischer/unitärer Vektorraum $(X, \langle \cdot, \cdot \rangle)$ wird genau dann **Hilbertraum** genannt, wenn er vollständig bzgl. der in [Bem. A.3.3](#) definierten Norm ist, d. h. wenn er ein Banachraum [bzgl. dieser Norm] ist.

Korollar A.3.6 (Stetigkeit des Skalarproduktes [25, Exercise 6.2.5, Solution 6.10.5]).

Sei $(H, \langle \cdot, \cdot \rangle)$ ein Hilbertraum und seien $x_n \rightarrow x, y_n \rightarrow y$ mit $x, y \in H$ $[(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}} \subset H]$. Dann ist das Skalarprodukt stetig und es gilt

$$\langle x, y \rangle = \lim_{n \rightarrow \infty} \langle x_n, y_n \rangle. \quad (\text{A.9})$$

Proposition A.3.7 (Parallelogramm-Gesetz [25, Prop. 6.1.10]).

Sei $(H, \langle \cdot, \cdot \rangle)$ ein Hilbertraum. Dann gilt

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 \quad \text{für alle } x, y \in H. \quad (\text{A.10})$$

Satz A.3.8 (Satz von Fréchet-Jordan-von Neumann [25, Theorem 6.1.11]).

Wenn $(H, \|\cdot\|)$ ein Banachraum über dem Körper $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$ ist, sodass die Norm [die Gleichung] [\(A.10\)](#) erfüllt. Dann ist $(H, \langle \cdot, \cdot \rangle)$ ein Hilbertraum mit

$$\langle x, y \rangle := \frac{1}{4} \left(\|x + y\|^2 - \|x - y\|^2 \right), \quad x, y \in H, \quad (\text{A.11})$$

falls $\mathbb{K} = \mathbb{R}$ bzw.

$$\langle x, y \rangle := \frac{1}{4} \left(\|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2 \right), \quad x, y \in H, \quad (\text{A.12})$$

falls $\mathbb{K} = \mathbb{C}$.

Beispiel A.3.9 ($L^2(\Omega, \mu), \langle \cdot, \cdot \rangle_{L^2(\Omega, \mu)}$ [25, Example 6.1.12]).

Von den $L^p(\Omega, \mu)$ -Räumen aus [Bsp. A.2.14](#) ist ausschließlich $(L^2(\Omega, \mu), \langle \cdot, \cdot \rangle_{L^2(\Omega, \mu)})$ ein Hilbertraum mit

$$\langle f, g \rangle_{L^2(\Omega, \mu)} := \int_{\Omega} f(x) \overline{g(x)} \, d\mu(x) \quad \text{für } f, g \in L^2(\Omega, \mu). \quad (\text{A.13})$$

Definition A.3.10 (Orthogonalität [25, Def. 6.3.1]).

Sei $(H, \langle \cdot, \cdot \rangle)$ ein Hilbertraum und $A \subseteq H$ eine Teilmenge. Man nennt $x \in H$

- 1) **orthogonal** zu y [$y \in H$], wenn $\langle x, y \rangle = 0$: Dann schreibt man $x \perp y$;
- 2) **orthogonal** zu A , wenn $\langle x, y \rangle = 0, \forall y \in A$: Dann schreibt man $x \perp A$.

Die Menge

$$A^\perp := \{x \in H : \langle x, z \rangle = 0 \text{ für alle } z \in A\}$$

wird das **orthogonale Komplement** von A genannt. Es folgt damit, dass $x \perp A$ genau dann gilt, wenn $x \in A^\perp$.

Satz A.3.11 ($H = A \oplus A^\perp$ [25, Theorem 6.3.7]).

Sei $(H, \langle \cdot, \cdot \rangle)$ ein Hilbertraum und $A \subseteq H$ ein abgeschlossener Unterraum (d.h. eine abgeschlossene Teilmenge, die ein Vektorraum mit der Addition und Multiplikation von H ist). Dann gilt

$$H = A \oplus A^\perp.$$

Definition A.3.12 (Orthonormalsystem [25, 6.5.1]).

Sei $(H, \langle \cdot, \cdot \rangle_H)$ ein Hilbertraum. Eine nicht-leere Teilmenge S von H heißt

- (i) **orthogonal**, wenn $x \perp y$ für alle $x, y \in S$ mit $x \neq y$,
- (ii) **orthonormal**, wenn S orthogonal ist und $\|x\|_H = 1$ für alle $x \in S$ gilt,
- (iii) ein **vollständiges Orthonormalsystem** wenn S orthonormal und maximal ist, im folgenden Sinn: Ist $\tilde{S}$ ein anderes Orthonormalsystem von H mit $S \subset \tilde{S}$, dann gilt $S = \tilde{S}$.

Definition und Satz A.3.13 (Orthonormalbasis [25, S. 6.5.12, 6.4.14]).

Ein Orthonormalsystem S in einem Hilbertraum H heißt **Orthonormalbasis**, wenn gilt, dass

$$x = \sum_{e \in S} \langle x, e \rangle e \quad \text{für alle } x \in H.$$

Ein Hilbertraum H ist genau dann separabel, wenn er eine höchstens abzählbar unendlich große Orthonormalbasis besitzt.

Satz A.3.14 (Besselsche Ungleichung [25, Theorem 6.5.5]).

Sei $\{x_n : n \in \mathbb{N}\}$ ein Orthonormalsystem in einem Hilbertraum $(H, \langle \cdot, \cdot \rangle)$. Dann gilt für alle $x \in H$:

$$\sum_{n=1}^{\infty} |\langle x, x_n \rangle|^2 \leq \|x\|^2. \quad (\text{A.14})$$

A.4 Operatoren – Funktionale

Definition A.4.1 ((anti-/lineare) Operatoren / Funktionale [25, Def. 3.1.1]).

Seien X, Y zwei Vektorräume über dem Körper $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Eine Abbildung T von X nach Y wird traditionellerweise **Operator** genannt – oder aber **Funktional** im Falle von $Y = \mathbb{K}$.

Ein Operator $T : X \rightarrow Y$ wird **linear** genannt, falls gilt:

(L1) $T(x + y) = T(x) + T(y)$ für alle $x, y \in X$ und

(L2) $T(\alpha x) = \alpha T(x)$ für alle $x \in X$ und alle $\alpha \in \mathbb{K}$.

Er wird **anti-linear** genannt, falls er (L1) erfüllt und gilt:

(L2') $T(\alpha x) = \bar{\alpha} T(x)$ für alle $x \in X$ und alle $\alpha \in \mathbb{K}$.

Anstelle von $T(x)$ ist es in der Literatur üblich, nur Tx zu schreiben für $x \in X$.

Definition und Satz A.4.2 (beschränkte anti-/lineare Operatoren [25, Def. 3.1.3, Prop. 3.1.5]).

Seien $(X, \|\cdot\|_X)$ und $(Y, \|\cdot\|_Y)$ zwei normierte Räume und $T : X \rightarrow Y$ linear oder anti-linear. Dann heißt T **beschränkt** (respektive/bzgl. $\|\cdot\|_X$ und $\|\cdot\|_Y$), wenn folgende äquivalente Bedingungen erfüllt sind:

- a) Es existiert ein $M > 0$, sodass $\|Tx\|_Y \leq M\|x\|_X$ für alle $x \in X$.
- b) T ist Lipschitz-stetig.
- b) T ist stetig.
- b) T ist stetig in 0.

Definition und Satz A.4.3 (Kern und Bild eines linearen Operators [25, Def. 3.1.2]).

Seien $(X, \|\cdot\|_X)$ und $(Y, \|\cdot\|_Y)$ zwei normierte Räume (über $\mathbb{K} \subseteq \{\mathbb{R}, \mathbb{C}\}$) und $T : X \rightarrow Y$ linear oder anti-linear. Dann sind der **Kern** und das **Bild** von T definiert durch

$$\ker T := T^{-1}(\{0\}) = \{x \in X : Tx = 0\}$$

und

$$\text{ran } T := T(X) = \{Tx \in Y : x \in X\}.$$

$\ker T$ und $\text{ran } T$ sind Unterräume von X bzw. Y und T ist genau dann injektiv, wenn $\ker T = \{0\}$ gilt. Ist T zudem beschränkt, so ist $\ker T$ zudem abgeschlossen.

Satz A.4.4 (Normierter Raum $(\mathcal{L}(X, Y), \|\cdot\|_{\mathcal{L}(X, Y)})$ [25, Prop. 3.1.7, Prop. 3.1.15]).

Es seien $(X, \|\cdot\|_X)$ und $(Y, \|\cdot\|_Y)$ zwei normierte Räume und

$$\mathcal{L}(X, Y) := \{T : X \rightarrow Y \mid T \text{ ist linear und beschränkt}\}$$

die Menge der beschränkten, linearen Operatoren. Dann ist $(\mathcal{L}(X, Y), \|\cdot\|_{\mathcal{L}(X, Y)})$ ebenfalls ein normierter Raum mit der Norm

$$\|T\|_{\mathcal{L}(X, Y)} := \inf\{M > 0 : \|Tx\|_Y \leq M\|x\|_X\} \quad (\text{A.15})$$

$$= \sup_{\|x\|_X=1} \|Tx\|_Y \quad (\text{A.16})$$

$$= \sup_{\|x\|_X \leq 1} \|Tx\|_Y \quad (\text{A.17})$$

$$= \sup_{x \neq 0} \frac{\|Tx\|_Y}{\|x\|_X} \quad \text{für alle } T \in \mathcal{L}(X, Y). \quad (\text{A.18})$$

Ist Y ein Banachraum, so ist dies auch $\mathcal{L}(X, Y)$.

Satz A.4.5 (Stetige Fortsetzung linearer Operatoren [25, Theorem 3.1.11]).

Seien X, Y zwei normierte Räume, sodass Y vollständig ist. Sei D ein dichter Unterraum von X . Dann existiert für jedes $T \in \mathcal{L}(D, Y)$ ein eindeutiges $\tilde{T} \in \mathcal{L}(X, Y)$, welches T fortsetzt, d. h.

$$Tx = \tilde{T}x \quad \text{für jedes } x \in D.$$

Zudem gilt $\|T\| = \|\tilde{T}\|$.

Definition und Satz A.4.6 (Kompakte Operatoren [25, Def. 7.5.1, Theorem 7.5.2]).

Seien X und Y zwei normierte Räume und $T : X \rightarrow Y$ ein linearer Operator. Man nennt T **kompakt**, wenn eine/alle der drei äquivalenten Bedingungen zutreffen:

1. $\overline{T(A)} \subset Y$ ist kompakt immer dann, wenn $A \subset X$ beschränkt ist.
2. $\overline{T(\overline{B(0, 1)})}$ ist kompakt.
3. $(Tx_n)_{n \in \mathbb{N}}$ enthält eine konvergente Teilfolge immer dann, wenn $(x_n)_{n \in \mathbb{N}} \subset X$ eine beschränkte Folge ist.

Definition A.4.7 (Dualraum [25, 4]).

Sei $(X, \|\cdot\|_X)$ ein normierter Raum über $\mathbb{K} \in \{\mathbb{R}, \mathbb{K}\}$. Dann heißt der Raum der beschränkten, linearen Funktionale $\mathcal{L}(X, \mathbb{K})$ der **Dualraum** von X und wird mit X' bezeichnet.

Für ein Element $x' \in X'$ und $x \in X$ ist für die Funktionalauswertung die Schreibweise üblich:

$$\langle x', x \rangle_{X', X} := x'(x). \quad (\text{A.19})$$

Korollar A.4.8 (Norm via Dualraum [25, Kor. 4.2.10]).

Ist X ein normierter Raum dann gilt für alle $x \in X$

$$\|x\| = \sup\{|\langle x', x \rangle| : x' \in X, \|x'\| \leq 1\}. \quad (\text{A.20})$$

Satz A.4.9 (Darstellungssatz von Riesz [25, S. 5.1.27]).

Sei $1 \leq p < \infty$ und $L \in L^p(\Omega, \mu)'$ gegeben. Dann existiert ein eindeutiges $f \in L^q(\Omega)$ mit $\frac{1}{p} + \frac{1}{q} = 1$, mit der Konvention $\frac{1}{0} = \infty$

$$L(g) = \int_{\Omega} fg \, d\mu(x), \quad \forall g \in L^p(\Omega, \mu). \quad (\text{A.21})$$

Zudem gilt $\|f\|_{L^q(\Omega, \mu)} = \|L\|_{L^p(\Omega, \mu)'}$. Insbesondere ist die Abbildung

$$\Phi_q : L^q(\Omega, \mu) \rightarrow L^p(\Omega, \mu)', \quad f \mapsto L_f$$

ein isometrischer Isomorphismus.

Satz A.4.10 (Darstellungssatz von Fréchet-Riesz [25, S. 6.3.9]).

Sei $(H, \langle \cdot, \cdot \rangle)$ ein Hilbertraum. Dann existiert für jedes $f \in H' := \mathcal{L}(H, \mathbb{K})$ ein eindeutiges $y_f \in H$, sodass

$$f(x) := \langle x, y_f \rangle \quad \text{für alle } x \in H. \quad (\text{A.22})$$

Korollar A.4.11 (H und H' sind isometrisch-isomorph [25, Cor. 6.3.10]).

Sei $(H, \langle \cdot, \cdot \rangle)$ ein Hilbertraum über dem Körper $\mathbb{K}$ und $(H', \|\cdot\|_{H'})$ sein Dualraum. Dann ist die Abbildung $\Phi : H \rightarrow H'$ definiert durch

$$[\Phi(y)](\cdot) := \langle \cdot, y \rangle, \quad y \in H, \quad (\text{A.23})$$

isometrisch, bijektiv und zusätzlich linear (wenn $\mathbb{K} = \mathbb{R}$) oder anti-linear (wenn $\mathbb{K} = \mathbb{C}$).

Theorem A.4.12 (Hahn-Banach: Fortsetzung von beschränkten linearen Funktionalen [25, Theorem 4.2.6]).

Sei X ein normierter Raum und V ein Unterraum von X . Dann existiert für jedes beschränkte lineare Funktional $u' : V \rightarrow \mathbb{K}$ ein beschränktes lineares Funktional $x' : X \rightarrow \mathbb{K}$, sodass

$$x'|_V = u' \quad \text{und} \quad \|x'\|_{X'} = \|u'\|_{V'}.$$

Definition und Satz A.4.13 (Adjungierter Operator [25, Def. 4.7.1, Theorem 4.7.4]).

Seien X, Y zwei normierte Räume und sei $T \in \mathcal{L}(X, Y)$. Der **adjungierte Operator** $T' : Y' \rightarrow X'$ ist definiert durch

$$\langle T'y', x \rangle_{X', X} := \langle y', Tx \rangle_{Y', Y}, \quad x \in X, y' \in Y'. \quad (\text{A.24})$$

T' ist ebenfalls linear und beschränkt (d. h. $T' \in \mathcal{L}(Y', X')$) und es gilt

$$\|T'\|_{\mathcal{L}(Y', X')} = \|T\|_{\mathcal{L}(X, Y)}.$$

Definition A.4.14 (Selbstadjungierter Operator [25, Def. 6.7.6]).

Sei $(H, \langle \cdot, \cdot \rangle_H)$ ein Hilbertraum. Ein linearer Operator $T \in \mathcal{L}(H)$ heißt **selbstadjungiert** oder **symmetrisch**, wenn

$$\langle Tx, y \rangle_H = \langle x, Ty \rangle_H \quad \text{für alle } x, y \in H. \quad (\text{A.25})$$

Definition A.4.15 (Punktspektrum eines Operators [25, Def. 7.3.1]).

Sei $A \in \mathcal{L}(X)$ ein beschränkter Operator auf einem Banachraum X (auf $\mathbb{K}$). Die Menge

$$\sigma_p(A) := \{\lambda \in \mathbb{K} : \ker(\lambda \text{Id} - A) \neq \{0\}\}$$

heißt das **Punktspektrum** von A .

Die Elemente von $\sigma_p(A)$ heißen **Eigenwerte** von A analog zum Fall von Matrizen, denn für alle $\lambda \in \sigma_p(A)$ existiert ein Element $x \in X$ ungleich Null, sodass

$$Ax = \lambda x. \quad (\text{A.26})$$

Folglich wird solch ein x **Eigenvektor** zum Eigenwert λ genannt. Ist X ein Funktionenraum, so heißt x auch **Eigenfunktion**. Der zugehörige **Eigenraum** ist die lineare Hülle aller Eigenvektoren zum Eigenwert λ .

Satz A.4.16 (Beschränktes Punktspektrum [25, Theorem 7.1.3]).

Sei $A \in \mathcal{L}(X)$ ein beschränkter linearer Operator auf einem Banachraum X . Dann gilt für jeden Eigenwert $\lambda \in \sigma_p(A)$, dass $|\lambda| \leq \|A\|_{\mathcal{L}(X)}$.

Satz A.4.17 (Spektralsatz für kompakte selbstadjungierte Operatoren [25, Theorem 7.7.5]).

Sei $(H, \langle \cdot, \cdot \rangle_H)$ ein Hilbertraum und T ein kompakter selbstadjungierter Operator darauf. Dann existieren

- eine endliche oder abzählbar unendliche Teilmenge $K \subset H$
- ein Orthonormalsystem $\{e_k : k \in K\}$ und
- eine Folge $(\lambda_k)_{k \in K}$ von reellen Zahlen

sodass,

1. $(\lambda_k)_{k \in K}$ hat genau einen Häufungswert bei 0 immer dann, wenn K unendlich ist.
2. die Folge $(\lambda_k)_{k \in K}$ besteht aus den Eigenwerten von T ungleich Null und $(e_k)_{k \in K}$ sind die zugehörigen Eigenvektoren.
3. höchstens endlich viele Eigenwerte sind gleich (d. h. für alle $\lambda \in \mathbb{R}$ gibt es ein $r \in \mathbb{N}_0$, sodass $\lambda = \lambda_{k_1} = \dots = \lambda_{k_r}$ und alle restlichen Eigenwerte unterscheiden sich von λ) und man sagt, dass alle Eigenwerte von endlicher Multiplizität sind.
4. H kann zerlegt werden mittels

$$H = \ker(T) \oplus \overline{\text{span}\{e_1, e_2, \dots\}}; \quad (\text{A.27})$$

5. Die Reihenentwicklung

$$Tx = \sum_{k \in K} \lambda_k \langle x, e_k \rangle_H e_k \quad (\text{A.28})$$

gilt für alle $x \in H$.

Folglich, wenn insbesondere $\ker(T) = \{0\}$ ist, dann existiert eine Orthonormalbasis von H , die nur aus Eigenvektoren von T besteht.

Definition und Satz A.4.18 (Hilbert-Schmidt Integraloperatoren [25, Def. 7.5.9, Prop. 7.5.10, Theorem 7.5.11]).

Sei (Ω, μ) ein Maßraum und der Operator T definiert durch

$$Tf := \int_{\Omega} k(\cdot, y)f(y) \, d\mu(y), \quad f : \Omega \rightarrow \mathbb{K} \quad (\text{A.29})$$

wobei k eine gegebene Funktion von $\Omega \times \Omega$ nach $\mathbb{K}$ ist, der sog. **(Integral-)Kern** von T . Wenn $k \in L^2(\Omega^2, \mu^2)$ ist, d. h.

$$\int_{\Omega} \int_{\Omega} |k(x, y)|^2 \, d\mu(x) \, d\mu(y) = M < \infty \quad (\text{A.30})$$

gilt, dann wird T **Hilbert-Schmidt Integraloperator** genannt.

Betrachtet man den Hilbert-Schmidt Integraloperator T als Operator auf $L^2(\Omega, \mu)$, dann ist T beschränkt und kompakt.

A.5 Spektraltheorie

Definition A.5.1 (Algebra [25, Def. 7.1.1]).

Sei $\mathcal{A}$ ein Vektorraum über $\mathbb{K}$. Existiert eine zusätzliche Verknüpfung

$$\mathcal{A} \times \mathcal{A} \ni (a, b) \mapsto ab \in \mathcal{A},$$

die:

$$\textbf{(A1)} \quad (xy)z = x(yz),$$

$$\textbf{(A2)} \quad x(y + z) = xy + yz, \quad (x + y)z = xz + yz,$$

$$\textbf{(A3)} \quad \alpha\beta(xy) = (\alpha x)(\beta y)$$

für alle $\alpha, \beta \in \mathbb{K}$ und $x, y, z \in \mathcal{A}$ erfüllt, dann wird $\mathcal{A}$ eine **Algebra** genannt. Die obige Verknüpfung wird normalerweise als die Multiplikation über $\mathcal{A}$ bezeichnet und die Eigenschaften (A1) und (A2) können entsprechend als Assoziativ- bzw. Distributivgesetz angesehen werden.

Man sagt, die Algebra besitzt ein **Einselement**, wenn sie ein neutrales Element bzgl. der Multiplikation enthält – normalerweise mit Id oder e bezeichnet – sodass $\text{Id } x = x \text{Id} = x$ für alle $x \in \mathcal{A}$. Ist die Multiplikation kommutativ, d. h. $xy = yx$ für alle $x, y \in \mathcal{A}$, so heißt $\mathcal{A}$ **kommutative Algebra**.

Definition A.5.2 (Banachalgebra [25, Def. 7.1.2]).

Eine **Banachalgebra** ist ein Paar $(\mathcal{A}, \|\cdot\|)$, für das $\mathcal{A}$ eine Algebra über $\mathbb{K}$ ist und $\|\cdot\|$ eine Norm bzgl. welcher $\mathcal{A}$ vollständig ist, wobei die Norm zudem submultiplikativ ist, d. h.

$$\|xy\| \leq \|x\|\|y\| \quad \text{für alle } x, y \in \mathcal{A}. \quad (\text{A.31})$$

Definition und Satz A.5.3 (Invertierbare Elemente [25, Def. 7.1.5]).

Sei $\mathcal{A}$ eine Banachalgebra mit Einselement. Dann heißt ein Element $a \in \mathcal{A}$ **invertierbar**, wenn es ein $b \in \mathcal{A}$ gibt mit

$$ab = ba = \text{Id}.$$

In diesem Fall heißt b das **Inverse** von a , ist eindeutig und wird mit a^{-1} bezeichnet. Die Menge

$$\mathcal{G}(\mathcal{A}) := \{a \in \mathcal{A} : a \text{ ist invertierbar}\}$$

bildet eine Gruppe bzgl. der Multiplikation über $\mathcal{A}$.

Definition A.5.4 (Resolventenmenge und Spektrum [25, Def. 7.1.8]).

Sei $\mathcal{A}$ eine Banachalgebra mit Einselement. Für $a \in \mathcal{A}$ ist die **Resolventenmenge** definiert durch

$$\rho(a) := \{\lambda \in \mathbb{K} : \lambda \text{Id} - a \in \mathcal{G}(\mathcal{A})\}.$$

Ihr Komplement

$$\sigma(a) := \mathbb{C} \setminus \rho(a)$$

heißt das **Spektrum** von a . Für alle $\lambda \in \rho(a)$ bezeichnet man mit

$$R(\lambda, a) := (\lambda \text{Id} - A)^{-1}$$

die Resolvente von a bei λ .

Satz A.5.5 (Nicht-leeres Spektrum [25, Theorem 7.1.13]).

Sei $\mathcal{A}$ eine Banachalgebra mit Einselement und $a \in \mathcal{A}$. Dann gilt: $\sigma(a) \neq \emptyset$.

Definition und Satz A.5.6 (Spektralradius [25, Def. 7.1.18, Remark 7.1.19]).

Sei $\mathcal{A}$ eine Banachalgebra mit Einselement. Der **Spektralradius** von $a \in \mathcal{A}$ ist die nicht-negative reelle Zahl

$$r(a) := \sup\{|\lambda| : \lambda \in \sigma(a)\}. \quad (\text{A.32})$$

Es gilt:

$$0 \leq r(a) \leq \|a\|. \quad (\text{A.33})$$

Definition A.5.7 (Involutive Banach- / C^* -Algebra [25, Def. 7.1.16]).

Eine **involutive Banachalgebra** ist eine komplexe Banachalgebra mit einer zusätzlichen Verknüpfung $a \mapsto a^* \in \mathcal{A}$, **Involution** genannt, sodass

$$a^{**} = a, \quad (ab)^* = b^*a^* \quad \text{und} \quad (\alpha a + b)^* = \bar{\alpha}a^* + b^*, \quad (\text{A.34})$$

für $\alpha \in \mathbb{C}$ und $a, b \in \mathcal{A}$. Ein Element $a \in \mathcal{A}$ heißt **selbstadjungiert**, wenn $a = a^*$ und **normal**, wenn $a^*a = aa^*$.

Gilt zusätzlich

$$\|a^*a\| = \|a\|^2 \quad \text{für alle } a \in \mathcal{A}, \quad (\text{A.35})$$

Dann nennt man man $\mathcal{A}$ eine **C^* -Algebra** („C-Stern-Algebra“ ausgesprochen).

Satz A.5.8 (Gelfands Formel [25, Theorem 7.1.23]).

Sei $\mathcal{A}$ eine Banachalgebra mit Einselement. Dann gelten die folgende Behauptungen.

1. Der Spektralradius von $a \in \mathcal{A}$ lässt sich mit Gelfands Formel berechnen:

$$r(a) = \lim_{n \rightarrow \infty} \|a^n\|^{1/n}. \quad (\text{A.36})$$

Man beachte, dass dies insbesondere impliziert, dass der Grenzwert auf der rechten Seite immer existiert.

2. Ist $\mathcal{A}$ eine C^* -Algebra mit Einselement und a normal, dann gilt

$$r(a) = \|a\|, \quad (\text{A.37})$$

und folglich

$$\|a\| = r(a^*a)^{1/2} \quad \text{für jedes } a \in \mathcal{A}. \quad (\text{A.38})$$

A.6 Maß- und Integrationstheorie

Definition A.6.1 (σ -Algebra [17, Def. 4.2.4, Prop. 4.2.5]).

Ein Mengensystem $\mathcal{A}$ über einer (nichtleeren) Menge Ω heißt **σ -Algebra**, wenn gilt:

1. $\Omega \in \mathcal{A}$
2. Ist $A \in \mathcal{A}$, dann auch $\Omega \setminus A \in \mathcal{A}$
3. Ist $\{A_i\}_{i \in \mathbb{N}} \subset \mathcal{A}$, dann gilt $A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Definition und Satz A.6.2 (Borelsche σ -Algebra [17, Def. 4.2.18, Satz 4.2.22]).

Sei $\Omega = \mathbb{R}^d$ für ein $d \in \mathbb{N}$. Es existiert eine kleinste σ -Algebra $\mathcal{B}(\mathbb{R}^d)$, die alle offenen Mengen in Ω enthält. $\mathcal{B}(\mathbb{R}^d)$ heißt die **Borel- σ -Algebra** und die Mengen darin heißen **Borelmengen** oder **Borelmessbare Mengen**.

Definition A.6.3 (Maß [17, Def. 4.3.1, Def. 3.2.18]).

Ein **Maß** ist eine Mengenfunktion $\mu : \mathcal{A} \rightarrow [0, \infty]$ auf einer σ -Algebra $\mathcal{A}$ für die gilt:

1. $\mu(\emptyset) = 0$,
2. $\mu(A) \geq 0$ für alle $A \in \mathcal{A}$ und
3. μ ist σ -additiv, d. h. für jede Folge $\{A_n\}_{n \in \mathbb{N}}$ disjunkter Mengen aus $\mathcal{A}$, für die $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ ist, gilt:

$$\mu(A) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n). \quad (\text{A.39})$$

Definition A.6.4 (Messraum / Maßraum [17, Def. 4.3.5]).

Ist $\mathcal{A}$ eine σ -Algebra mit Grundmenge Ω , dann heißt das Paar $(\Omega, \mathcal{A})$ ein **Messraum**.

Ist $(\Omega, \mathcal{A})$ ein Messraum und μ ein Maß auf $\mathcal{A}$, dann heißt das Tripel $(\Omega, \mathcal{A}, \mu)$ ein **Maßraum**.

Ein Maßraum $(\Omega, \mathcal{A}, \mu)$ heißt **endlich**, wenn $\mu(\Omega) < \infty$ gilt und **σ -endlich**, wenn $(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}$ existiert mit $\mu(A_n) < \infty$ und $A_n \subseteq A_{n+1}$ für alle $n \in \mathbb{N}$ sowie $\Omega = \bigcup_{n \in \mathbb{N}} A_n$.

Korollar A.6.5 (Monotonie von Maßen [17, Kor. 3.6.6]).

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Sind $A, B \in \mathcal{A}$ mit $B \subset A$, dann folgt:

$$\mu(B) \leq \mu(A) \quad (\text{Monotonie}). \quad (\text{A.40})$$

Korollar A.6.6 (σ -Subadditivität von Maßen [17, Kor. 3.6.6]).

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Sind $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{A}$ (nicht notwendigerweise disjunkt), dann gilt:

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n) \quad (\sigma\text{-Subadditivität}). \quad (\text{A.41})$$

Proposition A.6.7 (Stetigkeit von Maßen [17, Prop. 4.3.10]).

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum.

1. Ist $A_n \in \mathcal{A}, A_n \nearrow A$ (d. h. $A_n \subseteq A_{n+1}$ für alle $n \in \mathbb{N}$ und $A = \bigcup_{n=1}^{\infty} A_n$) mit $A \in \mathcal{A}$, dann gilt

$$\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A).$$

2. Ist $A_n \in \mathcal{A}, \mu(A_1) < \infty$ und $A_n \searrow A$ (d. h. $A_{n+1} \subseteq A_n$ für alle $n \in \mathbb{N}$ und $A = \bigcap_{n=1}^{\infty} A_n$) mit $A \in \mathcal{A}$, dann gilt

$$\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A).$$

Definition A.6.8 (Nullmenge und fast überall [17, 25, Def. 6.2.1, Def. A.1.7]).

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum.

1. Eine Menge $N \in \mathcal{A}$ heißt **μ -Nullmenge** oder einfach **Nullmenge**, wenn $\mu(N) = 0$ gilt.
2. Man sagt eine Eigenschaft ist **fast überall** erfüllt auf einer Menge $A \in \mathcal{A}$, falls es eine Nullmenge $B \in \mathcal{A}$ mit $B \subseteq A$ gibt, sodass die Eigenschaft in $A \setminus B$ erfüllt ist.

Definition A.6.9 (Messbare Abbildungen [17, Def. 5.2.1]).

Sind $(\Omega_1, \mathcal{A}_1)$ und $(\Omega_2, \mathcal{A}_2)$ zwei Messräume, dann heißt eine Abbildung

$$f : \Omega_1 \rightarrow \Omega_2$$

$(\mathcal{A}_1 - \mathcal{A}_2)$ -messbar, wenn für jedes $A_2 \in \mathcal{A}_2$

$$A_1 = f^{-1}(A_2) \in \mathcal{A}_1$$

gilt.

Bemerkung A.6.10 (Messbare Funktionen [25, Def. A.1.8]).

Eine messbare Abbildung $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}^d))$ die aus einem Messraum $(\Omega, \mathcal{A})$ in die Borelsche σ -Algebra abbildet, nennt man auch **messbare Funktion**.

Satz A.6.11 (Komposition messbarer Abbildungen [17, Satz 5.2.6]).

Seien $(\Omega_1, \mathcal{A}_1)$, $(\Omega_2, \mathcal{A}_2)$ und $(\Omega_3, \mathcal{A}_3)$ Messräume, $f : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$ sei $(\mathcal{A}_1 - \mathcal{A}_2)$ -messbar und $g : (\Omega_2, \mathcal{A}_2) \rightarrow (\Omega_3, \mathcal{A}_3)$ sei $(\mathcal{A}_2 - \mathcal{A}_3)$ -messbar. Dann ist

$$g \circ f : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_3, \mathcal{A}_3)$$

$(\mathcal{A}_1 - \mathcal{A}_3)$ -messbar.

Korollar A.6.12 (Stetigkeit und Messbarkeit [17, Satz 5.2.9]).

Ist $f : \mathbb{R}^d \rightarrow \mathbb{R}^m$ stetig, dann ist f messbar.

Satz A.6.13 (Rechenregeln für messbare Funktionen [17, Satz 5.2.12]).

Ist $(\Omega, \mathcal{A})$ ein Messraum und sind die Funktionen $f_1, f_2 : \Omega \rightarrow \mathbb{R}$ messbar, $\alpha, \beta \in \mathbb{R}$, dann sind auch die folgenden Funktionen messbar:

$$\begin{aligned} &\alpha f_1 + \beta f_2, \\ &f_1 f_2, \\ &\max\{f_1, f_2\} \quad \text{und} \\ &\min\{f_1, f_2\}. \end{aligned}$$

Definition A.6.14 (Integrierbare Funktion [17, Def. 5.5.3]).

Ist $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, dann heißt eine messbare Funktion $f : (\Omega, \mathcal{A}, \mu) \rightarrow \mathbb{R} \cup \{\infty, -\infty\}$ **μ -integrierbar**, wenn gilt:

$$\int_{\Omega} |f(x)| \, d\mu(x) < \infty.$$

Satz A.6.15 (Monotonie des Integrals [17, 25, Satz 5.4.20, Theorem A.1.15]).

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und f, g zwei messbare, nicht-negative Funktionen mit $f(x) \leq g(x)$ für alle $x \in \Omega$. Dann gilt:

$$\int_{\Omega} f(x) \, d\mu(x) \leq \int_{\Omega} g(x) \, d\mu(x).$$

Lemma A.6.16 (Integral und fast überall [25, Lemma 5.1.3]).

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und f eine messbare nicht-negative Funktion mit $\int_{\Omega} f(x) \, d\mu(x) = 0$. Dann gilt $f(x) = 0$ fast überall für $x \in \Omega$.

Beispiel A.6.17 (Gleichheit in $L^p(\Omega, \mu)$ -Räumen). Es seien $p \in [1, \infty]$ und $f, g \in L^p(\Omega, \mu)$, so dass $\|f - g\|_{L^p(\Omega, \mu)} = 0$ gilt. Dann besagt Lemma A.6.16, dass $f(x) = g(x)$ für fast alle $x \in \Omega$ gilt.

Theorem A.6.18 (Konvergenz in $L^p(\Omega, \mu)$ -Räumen [25, Theorem 5.1.11]).

Sei (Ω, μ) ein Maßraum und $1 \leq p \leq \infty$. Es konvergiere $(f_n)_{n \in \mathbb{N}} \subset L^p(\Omega, \mu)$ in $L^p(\Omega, \mu)$ gegen eine Funktion $f \in L^p(\Omega, \mu)$. Dann existiert eine Teilfolge $(f_{n_k})_{k \in \mathbb{N}}$, die punktweise fast überall konvergiert.

Satz A.6.19 (Satz von Lebesgue über majorisierte Konvergenz [17, Satz 6.1.1]).

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und f_n eine Folge messbarer Funktionen auf diesem. Es gelte

$$f(x) = \lim_{n \rightarrow \infty} f_n(x) \quad \text{für alle } x \in \Omega.$$

Wenn es eine integrierbare Funktion g gibt, sodass $|f_n(x)| \leq g(x)$ für alle $n \in \mathbb{N}$ und $x \in \Omega$ gilt, dann folgt:

1. f ist integrierbar,
2. $\int_{\Omega} f(x) \, d\mu(x) = \int_{\Omega} \lim_{n \rightarrow \infty} f_n(x) \, d\mu(x) = \lim_{n \rightarrow \infty} \int_{\Omega} f_n(x) \, d\mu(x)$ und
3. $\lim_{n \rightarrow \infty} \int_{\Omega} |f(x) - f_n(x)| \, d\mu(x) = 0.$

Satz A.6.20 (Satz von Fubini für nicht-negative, messbare Funktionen (Spezialfall $\mathbb{R}^n$) [17, Satz 6.4.19]).

$\Omega_1 \subset \mathbb{R}^n$ und $\Omega_2 \subset \mathbb{R}^m$ seien zwei σ -endliche Lebesgue-Maßräume. Ist $f : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ nicht-negativ und messbar, dann gilt

$$\int_{\Omega_1 \times \Omega_2} f(x, y) \, d(x, y) = \int_{\Omega_1} \int_{\Omega_2} f(x, y) \, dy \, dx \quad (\text{A.42})$$

$$= \int_{\Omega_2} \int_{\Omega_1} f(x, y) \, dx \, dy. \quad (\text{A.43})$$

Satz A.6.21 (Satz von Fubini für integrierbare Funktionen (Spezialfall $\mathbb{R}^n$) [17, 25, Satz 6.4.19, Theorem A.1.22]).

$\Omega_1 \subset \mathbb{R}^n$ und $\Omega_2 \subset \mathbb{R}^m$ seien zwei σ -endliche Lebesgue-Maßräume und $f : \Omega_1 \times \Omega_2 \rightarrow \bar{\mathbb{R}}$ integrierbar. Dann gelten

- (a) $f(\cdot, y) \in L^1(\Omega_1)$ für fast alle y in Ω_2 ;
- (b) $f(x, \cdot) \in L^1(\Omega_2)$ für fast alle x in Ω_1 ;
- (c) $\int_{\Omega_1} f(x, y) \, dx \in L^1(\Omega_2)$;
- (d) $\int_{\Omega_2} f(x, y) \, dy \in L^1(\Omega_1)$;
- (e) Man hat

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f(x, y) \, d(x, y) &= \int_{\Omega_2} \left(\int_{\Omega_1} f(x, y) \, dx \right) dy \\ &= \int_{\Omega_1} \left(\int_{\Omega_2} f(x, y) \, dy \right) dx. \end{aligned}$$

Satz A.6.22 (Satz von Luzin [25, Theorem 5.1.14]).

Sei $\Omega \subset \mathbb{R}^n$ ein endlicher Maßraum. Sei $f : \Omega \rightarrow \mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ eine messbare beschränkte Funktion. Dann existiert für jedes $\varepsilon > 0$ eine kompakte Teilmenge $K \subset \Omega$, sodass $\mu(\Omega \setminus K) < \varepsilon$ und die Einschränkung von f auf K stetig ist.

A.7 Glättungskerne / Mollifier

Definition A.7.1 (Glättungskern / Mollifier / Regularisierte Funktion [25, Def. 5.1.20]).

Ein **Glättungskern** bzw. **Mollifier** ist eine Funktion $J \in C_0^\infty(\mathbb{R}^n)$, sodass

- (i) $J(x) \geq 0$,
- (ii) $J(x) = 0$ für $\|x\| > 1$,
- (iii) $\int_{\mathbb{R}^n} J(x) dx = 1$.

Man definiert zudem,

$$J_\varepsilon(x) := \varepsilon^{-n} J(\varepsilon^{-1}x), \quad (\text{A.44})$$

was impliziert, dass $J_\varepsilon(x) = 0$ für $\|x\| \geq \varepsilon$ und $\int_{\mathbb{R}^n} J_\varepsilon(x) dx = 1$. Die Funktion $J_\varepsilon * f : \mathbb{R}^n \rightarrow \mathbb{K}$, definiert durch

$$(J_\varepsilon * f)(x) := \int_{\mathbb{R}^n} J_\varepsilon(x - y) f(y) dy, \quad x \in \mathbb{R}^n \quad (\text{A.45})$$

wird dann die **regularisierte Funktion** von f genannt.

Beispiel A.7.2 (Beispiel-Glättungskern [25, Def. 5.1.20]).

Ein kanonisches Beispiel für einen Glättungskern / Mollifier ist

$$J : x \mapsto \begin{cases} k \exp\left(-\frac{1}{1-\|x\|^2}\right) & \text{wenn } \|x\| < 1 \\ 0 & \text{sonst,} \end{cases}$$

wobei k so gewählt ist, dass (iii) in Def. A.7.1 erfüllt ist.

Lemma A.7.3 (Konvergenz von regularisierten Funktionen [25, Lemma 5.1.21]).

Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ eine Funktion mit $f(x) = 0$ für $x \in \mathbb{R}^n \setminus \Omega$ (dabei sei $\Omega \subset \mathbb{R}^n$ offen). Ist $f \in C(\Omega)$ und $\Omega_0 \subset \Omega$ kompakt, dann konvergiert $J_\varepsilon * f \rightarrow f$ gleichmäßig in Ω_0 für $\varepsilon \rightarrow 0$.

A.8 Eine Winzigkeit Komplexes sowie Algebra

Definition und Satz A.8.1 (Konjugiert komplexe Zahl [30, S. 1.1.5, 1.2.1]).

Jede Zahl $z \in \mathbb{C}$ besitzt definitionsgemäß eine Darstellung $z = x + iy$ mit eindeutigen $x, y \in \mathbb{R}$. Die Zahl $\bar{z} := x - iy$ wird die **konjugiert komplexe Zahl** zu z genannt. Es gelten für alle $w, z \in \mathbb{C}$ und $r \in \mathbb{R}$:

- 1. $|\bar{z}| = |z|$,
- 2. $z\bar{z} = |z|^2$,
- 3. $\overline{z + w} = \bar{z} + \bar{w}$,
- 4. $\overline{z\bar{w}} = \bar{z}w$ und
- 5. $r = \bar{r}$.

Satz A.8.2 (Fundamentalsatz der Algebra [16, Satz 1.5.25, Kor. 1.5.26]).

Jedes normierte, nicht-konstante Polynom in $\mathbb{C}[X]$ zerfällt in Linearfaktoren.

B

Vervollständigungen von Vektorräumen

Im Beweis zum [Satz 4.2.1](#) von Moore wird an einer Stelle ein euklidischer bzw. unitärer Vektorraum zu einem Hilbertraum vervollständigt.

In der bisherigen universitären Laufbahn des Autors wurde diesbezüglich lediglich die (bis auf Isometrie eindeutige) Vervollständigung eines **metrischen Raumes** in [Def.+Satz A.1.19](#) betrachtet. Die Vervollständigung eines **normierten/euklidischen/unitären Vektorraums** erhält jedoch die Vektorraum-/Norm- und Skalarprodukteigenschaften des vervollständigten Raumes, sodass in diesen Fällen [Def.+Satz A.1.19](#) erweitert werden kann.

Um die Argumentation im Beweis zum Satz von Moore vollständig in dieser Arbeit zu haben, seien die relevanten Ergebnisse ausgeführt, deren Erarbeitung sich für den Autor auch als insgesamt sehr lehrreiche Übungsaufgabe herausgestellt hat. Dieses Kapitel stellt damit lediglich eine genauere Betrachtung von [\[25, Theorem 2.8.2\]](#) dar.

Satz B.1 (Vervollständigung eines Vektorraums [\[25, Theorem 2.8.2\]](#)).

Bis auf Isometrie besitzt jeder metrische Vektorraum (X, d) (über $\mathbb{K}$) genau eine Vervollständigung $(\hat{X}, \hat{d})$, wobei $\hat{X}$ selbst ein Vektorraum über $\mathbb{K}$ ist.

Beweis. [Def.+Satz A.1.19](#) liefert die bis auf Isometrie eindeutige Vervollständigung $(\hat{X}, \hat{d})$ von (X, d) , d. h. es existiert eine Isometrie $\Phi : X \rightarrow \hat{X}$, sodass $\overline{\Phi(X)} = \hat{X}$ (bzgl. $\hat{d}$). Es muss somit lediglich noch nachgewiesen werden, dass $\hat{X}$ ein Vektorraum ist.

Für die nachfolgende Argumentation seien dazu $\hat{x}, \hat{y}, \hat{z} \in \hat{X}$ und $\alpha, \beta \in \mathbb{K}$ fortan beliebig, aber fest. Da $\Phi(X)$ dicht in $\hat{X}$ liegt, existieren Folgen $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}, (z_n)_{n \in \mathbb{N}} \subset X$ mit $\Phi(x_n) \rightarrow \hat{x}$, $\Phi(y_n) \rightarrow \hat{y}$ und $\Phi(z_n) \rightarrow \hat{z}$ in $(\hat{X}, \hat{d})$.

Man definiere nun eine Addition und skalare Multiplikation auf $\hat{X}$ via:

$$\hat{x} + \hat{y} := \lim_{n \rightarrow \infty} \Phi(x_n + y_n) \quad \text{und} \quad \alpha \hat{x} := \lim_{n \rightarrow \infty} \Phi(\alpha x_n). \quad (\text{B.1})$$

Dabei beachte man, dass aufgrund der Stetigkeit von Φ (siehe Def.+Satz A.1.18), die beiden Verknüpfungen wohldefiniert sind, d. h. die Grenzwerte in $\Phi(X) = \hat{X}$ existieren und dass die Ausdrücke nicht von der Wahl der Folgen $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}$ abhängen, mit denen $\Phi(x_n) \rightarrow \hat{x}$ bzw. $\Phi(y_n) \rightarrow \hat{y}$.

Insbesondere wird Φ mit dieser Definition der beiden Verknüpfungen in (B.1) zu einer linearen Abbildung zwischen X und der Menge $(\Phi(X), +, \cdot)$:

$$\Phi(\alpha x + \beta y) \stackrel{(B.1)}{=} \alpha \Phi(x) + \beta \Phi(y) \quad \text{für alle } x, y \in X, \alpha, \beta \in \mathbb{K} \quad (B.2)$$

Es lassen sich damit die Bedingungen (Add₁)-(Add₄) und (Mult₁)-(Mult₃) aus der Def. A.2.1 eines Vektorraums schnell nachprüfen:

(Add₁): Es gilt:

$$(\hat{x} + \hat{y}) + \hat{z} \stackrel{(B.1)}{=} \lim_{n \rightarrow \infty} \Phi(x_n + y_n) + \lim_{n \rightarrow \infty} \Phi(z_n) \stackrel{(B.2)}{=} \lim_{n \rightarrow \infty} \Phi((x_n + y_n) + z_n) \quad (B.3)$$

$$= \lim_{n \rightarrow \infty} \Phi(x_n + (y_n + z_n)) \stackrel{(B.2)}{=} \lim_{n \rightarrow \infty} \Phi(x_n) + \lim_{n \rightarrow \infty} \Phi(y_n + z_n) \stackrel{(B.1)}{=} \hat{x} + (\hat{y} + \hat{z}). \quad (B.4)$$

(Add₂): Für das Element $0_{\hat{X}} := \Phi(0_X) \in \hat{X}$ (wobei $0_X \in X$ das Nullelement von X bezeichne) gelten:

$$\hat{x} + 0_{\hat{X}} \stackrel{(B.1)}{=} \lim_{n \rightarrow \infty} \Phi(x_n + 0_X) = \lim_{n \rightarrow \infty} \Phi(x_n) \stackrel{(B.1)}{=} \hat{x} \quad \text{und} \quad (B.5)$$

$$0_{\hat{X}} + \hat{x} \stackrel{(B.1)}{=} \lim_{n \rightarrow \infty} \Phi(0_X + x_n) = \lim_{n \rightarrow \infty} \Phi(x_n) \stackrel{(B.1)}{=} \hat{x} \quad (B.6)$$

Somit ist $0_{\hat{X}}$ das neutrale Element von $\hat{X}$.

(Add₃): Für das Element $-\hat{x} := \lim_{n \rightarrow \infty} \Phi(-x_n) \in \hat{X}$ gelten:

$$\hat{x} + (-\hat{x}) \stackrel{(B.1)}{=} \lim_{n \rightarrow \infty} \Phi(x_n + (-x_n)) = \Phi(0_X) \stackrel{(B.1)}{=} 0_{\hat{X}} \quad \text{und} \quad (B.7)$$

$$(-\hat{x}) + \hat{x} \stackrel{(B.1)}{=} \lim_{n \rightarrow \infty} \Phi((-x_n) + x_n) = \Phi(0_X) \stackrel{(B.1)}{=} 0_{\hat{X}} \quad (B.8)$$

Somit ist $-\hat{x}$ das zu $\hat{x}$ inverse Element.

(Add₄): Es gilt:

$$\hat{x} + \hat{y} \stackrel{(B.1)}{=} \lim_{n \rightarrow \infty} \Phi(x_n + y_n) = \lim_{n \rightarrow \infty} \Phi(y_n + x_n) \stackrel{(B.1)}{=} \hat{y} + \hat{x} \quad (B.9)$$

und somit auch die Kommutativität der Addition.

(Mult₁): Es gilt:

$$\alpha(\beta \hat{x}) \stackrel{(B.1)}{=} \alpha \lim_{n \rightarrow \infty} \Phi(\beta x_n) \stackrel{(B.2)}{=} \lim_{n \rightarrow \infty} \Phi(\alpha(\beta x_n)) = \lim_{n \rightarrow \infty} \Phi((\alpha\beta)x_n) \stackrel{(B.1)}{=} (\alpha\beta)\hat{x}. \quad (B.10)$$

(Mult₂): Es gelten:

$$\alpha(\hat{x} + \hat{y}) \stackrel{(B.1)}{=} \alpha \lim_{n \rightarrow \infty} \Phi(x_n + y_n) \stackrel{(B.2)}{=} \lim_{n \rightarrow \infty} \Phi(\alpha(x_n + y_n)) = \lim_{n \rightarrow \infty} \Phi(\alpha x_n + \alpha y_n) \quad (B.11)$$

$$\stackrel{(B.2)}{=} \lim_{n \rightarrow \infty} \Phi(\alpha x_n) + \lim_{n \rightarrow \infty} \Phi(\alpha y_n) \stackrel{(B.1)}{=} \alpha \hat{x} + \alpha \hat{y} \quad \text{sowie} \quad (B.12)$$

$$(\alpha + \beta) \hat{x} \stackrel{(B.1)}{=} \lim_{n \rightarrow \infty} \Phi((\alpha + \beta)x_n) = \lim_{n \rightarrow \infty} \Phi(\alpha x_n + \beta x_n) \stackrel{(B.1)}{=} \alpha \hat{x} + \beta \hat{x} \quad (B.13)$$

(Mult₃): Es gilt:

$$1\hat{x} \stackrel{(B.1)}{=} 1 \cdot \lim_{n \rightarrow \infty} \Phi(x_n) \stackrel{(B.2)}{=} \lim_{n \rightarrow \infty} \Phi(1x_n) = \lim_{n \rightarrow \infty} \Phi(x_n) \stackrel{(B.1)}{=} \hat{x}. \quad (B.14)$$

Somit sind alle Bedingungen aus [Def. A.2.1](#) erfüllt und $(\hat{X}, +, \cdot)$ ist als Vektorraum nachgewiesen, welcher bzgl. der Metrik $\hat{d}$ abgeschlossen ist. $\square$

Korollar B.2 (Vektorraumvervollständigung – lineare Isometrie [[25](#), Theorem 2.8.2]).

Sei $(\hat{X}, \langle \cdot, \cdot \rangle_{\hat{X}})$ eine Vervollständigung eines Vektorraums $(X, \langle \cdot, \cdot \rangle_X)$ gemäß [Satz B.1](#) und Φ die zugehörige Isometrie, sodass $\overline{\Phi(X)} = \hat{X}$. Dann ist Φ eine lineare Abbildung zwischen X und $\hat{X}$.

Beweis. Diese Aussage entspricht gerade [\(B.2\)](#) im Beweis von [Satz B.4](#). $\square$

Korollar B.3 (Vervollständigung einer linearen Hülle [[25](#), Theorem 2.8.2]).

Sei $(\hat{X}, \langle \cdot, \cdot \rangle_{\hat{X}})$ eine Vervollständigung gemäß [Satz B.1](#) von $(X, \langle \cdot, \cdot \rangle_X)$, wobei $X = \text{span } D$ die lineare Hülle einer Menge D sei und Φ die zugehörige Isometrie, sodass $\overline{\Phi(X)} = \hat{X}$. Dann liegt die lineare Hülle von $\Phi(D)$ dicht in $\hat{X}$, d. h. $\overline{\text{span } \Phi(D)} = \hat{X}$.

Beweis. „ $\Phi(X) \subseteq \text{span } \Phi(D)$ “: Sei $y \in \Phi(X)$ beliebig aber fest. Dann existiert ein $x \in X$ mit $\Phi(x) = y$. Da $X = \text{span } D$ ist, existieren dann $n \in \mathbb{N}, \alpha_i \in \mathbb{K}, d_i \in D$ mit $i \in \{1, \dots, n\}$, sodass $x = \sum_{i=1}^n \alpha_i d_i$. Weil Φ mit [Kor. B.2](#) linear ist, folgt insgesamt:

$$y = \Phi(x) = \Phi\left(\sum_{i=1}^n \alpha_i d_i\right) \stackrel{B.2}{=} \sum_{i=1}^n \alpha_i \Phi(d_i) \in \text{span } \Phi(D). \quad (B.15)$$

„ $\text{span } \Phi(D) \subseteq \Phi(X)$ “: Sei $y \in \text{span } \Phi(D)$ beliebig aber fest. Dann existieren $n \in \mathbb{N}, \alpha_i \in \mathbb{K}, d_i \in D$ mit $i \in \{1, \dots, n\}$ sodass $y = \sum_{i=1}^n \alpha_i \Phi(d_i)$. Erneut folgt mit der Linearität von Φ (siehe [Kor. B.2](#)):

$$y = \sum_{i=1}^n \alpha_i \Phi(d_i) \stackrel{B.2}{=} \Phi\left(\sum_{i=1}^n \alpha_i d_i\right) \in \Phi(\text{span } D) = \Phi(X). \quad (B.16)$$

$\square$

Satz B.4 (Vervollständigung eines normierten Raums [[25](#), Theorem 2.8.2]).

Bis auf Isometrie besitzt jeder normierte Raum $(X, \|\cdot\|_X)$ (über $\mathbb{K}$) genau eine Vervollständigung $(\hat{X}, \|\cdot\|_{\hat{X}})$, wobei $\hat{X}$ ein Banachraum (über $\mathbb{K}$) ist.

Beweis. Als normierter Raum ist X insbesondere ein Vektorraum (vgl. [Def. A.2.4](#)), wodurch [Satz B.1](#) anwendbar ist. Dieser liefert eine bis auf Isometrie eindeutige Vervollständigung $(\hat{X}, \hat{d})$ von $(X, \|\cdot\|_X)$,

d. h. es existiert eine Isometrie $\Phi : X \rightarrow \hat{X}$, sodass $\overline{\Phi(X)} = \hat{X}$ (bzgl. $\hat{d}$). Insbesondere besagt [Satz B.1](#), dass $\hat{X}$ selbst ein Vektorraum ist und für die Isometrie Φ gilt (man beachte [Kor. B.2](#)):

$$\hat{d}(\Phi(x), \Phi(y)) \stackrel{\text{B.1}}{=} d(x, y) \stackrel{\text{A.2.5}}{=} \|x - y\|_X \quad \text{für alle } x, y \in X. \quad (\text{B.17})$$

$$\Phi(\alpha x + \beta y) \stackrel{\text{B.2}}{=} \alpha \Phi(x) + \beta \Phi(y) \quad \text{für alle } x, y \in X, \alpha, \beta \in \mathbb{K}. \quad (\text{B.18})$$

Es ist somit nur noch zu zeigen, dass eine Norm $\|\cdot\|_{\hat{X}}$ existiert, mit der $(\hat{X}, \|\cdot\|_{\hat{X}})$ ein Banachraum ist.

Wie in der Beweisführung zu [Satz B.1](#) seien für die nachfolgende Argumentation dazu $\hat{x}, \hat{y} \in \hat{X}$ und $\alpha, \beta \in \mathbb{K}$ fortan beliebig aber fest. Da $\Phi(X)$ dicht in $\hat{X}$ liegt, existieren Folgen $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}} \subset X$ mit $\Phi(x_n) \rightarrow \hat{x}$ und $\Phi(y_n) \rightarrow \hat{y}$ in $(\hat{X}, \hat{d})$.

Man definiere damit die Abbildung ($0_{\hat{X}}$ bezeichne dabei das neutrale Element im Vektorraum $\hat{X}$).

$$\|\cdot\|_{\hat{X}} : \hat{X} \rightarrow [0, +\infty), \quad \|\hat{x}\|_{\hat{X}} := \hat{d}(\hat{x}, 0_{\hat{X}}) \quad \text{für } \hat{x} \in \hat{X}. \quad (\text{B.19})$$

Die Abbildung ist wohldefiniert, da $\hat{d}$ nach $[0, +\infty)$ abbildet und die drei Norm-Bedingungen aus [Def. A.2.4](#) lassen sich folgendermaßen prüfen.:

(N₁): Es folgt direkt aus der Eigenschaft der Metrik (M1) (siehe [Def. A.1.1](#)), dass

$$\|\hat{x}\|_{\hat{X}} \stackrel{(\text{B.19})}{=} \hat{d}(\hat{x}, 0_{\hat{X}}) = 0 \stackrel{\text{A.1.1}}{\Leftrightarrow} \hat{x} = 0. \quad (\text{B.20})$$

(N₂): Es gilt mit der Normeigenschaft (N2) von $\|\cdot\|_X$ in [Def. A.2.4](#) und der Isometrie Φ , dass

$$\|\alpha \hat{x}\|_{\hat{X}} \stackrel{(\text{B.19})}{=} \hat{d}(\alpha \hat{x}, 0_{\hat{X}}) \stackrel{(\text{B.18})}{=} \hat{d}(\lim_{n \rightarrow \infty} \Phi(\alpha x_n), \Phi(0_X)) \stackrel{\text{A.1.14}}{=} \lim_{n \rightarrow \infty} \hat{d}(\Phi(\alpha x_n), \Phi(0_X)) \quad (\text{B.21})$$

$$\stackrel{(\text{B.17})}{=} \lim_{n \rightarrow \infty} \|\alpha x_n - 0_X\|_X \stackrel{\text{A.2.4}}{=} \lim_{n \rightarrow \infty} |\alpha| \|x_n - 0_X\|_X \stackrel{(\text{B.17})}{=} |\alpha| \lim_{n \rightarrow \infty} \hat{d}(\Phi(x_n), \Phi(0_X)) \quad (\text{B.22})$$

$$\stackrel{\text{A.1.14}}{=} |\alpha| \hat{d}(\lim_{n \rightarrow \infty} \Phi(x_n), \Phi(0_X)) = |\alpha| \hat{d}(\hat{x}, 0_{\hat{X}}) \stackrel{(\text{B.19})}{=} |\alpha| \|\hat{x}\|_{\hat{X}}. \quad (\text{B.23})$$

(N₃): Ebenso wie unter (N₂) folgt mit der Normeigenschaft (N3) von $\|\cdot\|_X$ in [Def. A.2.4](#) und der Isometrie Φ , dass

$$\|\hat{x} + \hat{y}\|_{\hat{X}} \stackrel{(\text{B.19})}{=} \hat{d}(\lim_{n \rightarrow \infty} \Phi(x_n + y_n), \Phi(0_X)) \stackrel{\text{A.1.14}}{=} \lim_{n \rightarrow \infty} \hat{d}(\Phi(x_n + y_n), \Phi(0_X)) \quad (\text{B.24})$$

$$\stackrel{(\text{B.17})}{=} \lim_{n \rightarrow \infty} \|x_n + y_n - 0_X\|_X \stackrel{\text{A.2.4}}{=} \lim_{n \rightarrow \infty} (\|x_n - 0_X\|_X + \|y_n - 0_X\|_X) \quad (\text{B.25})$$

$$\stackrel{(\text{B.17})}{=} \lim_{n \rightarrow \infty} (\hat{d}(\Phi(x_n), \Phi(0_X)) + \hat{d}(\Phi(y_n), \Phi(0_X))) \quad (\text{B.26})$$

$$\stackrel{\text{A.1.14}}{=} \hat{d}(\lim_{n \rightarrow \infty} \Phi(x_n), \Phi(0_X)) + \hat{d}(\lim_{n \rightarrow \infty} \Phi(y_n), \Phi(0_X)) = \hat{d}(\hat{x}, 0_{\hat{X}}) + \hat{d}(\hat{y}, 0_{\hat{X}}) \quad (\text{B.27})$$

$$\stackrel{(\text{B.19})}{=} \|\hat{x}\|_{\hat{X}} + \|\hat{y}\|_{\hat{X}}. \quad (\text{B.28})$$

Sämtliche Bedingungen aus [Def. A.2.4](#) für eine Norm sind demnach erfüllt und $(\hat{X}, \|\cdot\|_{\hat{X}})$ ist also ein normierter Raum, welcher bzgl. der Metrik $\hat{d}$ abgeschlossen ist.

Da aber auch mit identischer Argumentation wie zum Nachweis von (N₂) und (N₃) gilt, dass

$$\|\hat{x} - \hat{y}\|_{\hat{X}} \stackrel{(B.19)}{=} \hat{d}\left(\lim_{n \rightarrow \infty} \Phi(x_n - y_n), \Phi(0_X)\right) \stackrel{A.1.14}{=} \lim_{n \rightarrow \infty} \hat{d}(\Phi(x_n - y_n), \Phi(0_X)) \quad (B.29)$$

$$\stackrel{(B.17)}{=} \lim_{n \rightarrow \infty} \|x_n - y_n - 0_X\|_X = \lim_{n \rightarrow \infty} \|x_n - y_n\|_X \stackrel{(B.17)}{=} \lim_{n \rightarrow \infty} \hat{d}(\Phi(x_n), \Phi(y_n)) \quad (B.30)$$

$$\stackrel{A.1.14}{=} \hat{d}\left(\lim_{n \rightarrow \infty} \Phi(x_n), \lim_{n \rightarrow \infty} \Phi(0_X)\right) \stackrel{(B.19)}{=} \hat{d}(\hat{x}, \hat{y}), \quad (B.31)$$

folgt, dass $\hat{d}$ gerade die von $\|\cdot\|_{\hat{X}}$ erzeugte Metrik ist. Ergo sind die beiden Räume $(\hat{X}, \|\cdot\|_{\hat{X}})$ und $(\hat{X}, \hat{d})$ topologisch äquivalent, was wegen der Abgeschlossenheit bzgl. $\hat{d}$ wiederum bedeutet, dass $(\hat{X}, \|\cdot\|_{\hat{X}})$ ein Banachraum ist. $\square$

Satz B.5 (Vervollständigung eines euklidischen/unitären Vektorraums [25, Theorem 2.8.2]).

Bis auf Isometrie besitzt jeder euklidische bzw. unitäre Vektorraum $(X, \langle \cdot, \cdot \rangle_X)$ (über $\mathbb{K}$) genau eine Vervollständigung $(\hat{X}, \langle \cdot, \cdot \rangle_{\hat{X}})$, wobei $\hat{X}$ ein Hilbertraum (über $\mathbb{K}$) ist.

Beweis. Als euklidischer/unitärer Vektorraum ist X insbesondere ein normierter Raum $(X, \|\cdot\|_X)$ mit der Norm

$$\|\cdot\|_X : X \rightarrow [0, +\infty), \quad \|x\|_X := \sqrt{\langle x, x \rangle}, \quad \text{für } x, y \in X \quad (B.32)$$

(vgl. Bem. A.3.3). Somit ist Satz B.4 anwendbar und dieser liefert eine bis auf Isometrie eindeutige Vervollständigung $(\hat{X}, \|\cdot\|_{\hat{X}})$ von $(X, \langle \cdot, \cdot \rangle_X)$, d. h. es existiert eine Isometrie $\Phi : X \rightarrow \hat{X}$, sodass $\overline{\Phi(X)} = \hat{X}$ (bzgl. $\|\cdot\|_{\hat{X}}$). Insbesondere besagt Satz B.4, dass $\hat{X}$ selbst ein Banachraum bzgl. $\|\cdot\|_{\hat{X}}$ ist und für die Isometrie Φ gilt (man beachte (B.18)):

$$\|\Phi(x)\|_{\hat{X}} \stackrel{B.4}{=} \|x\|_X \quad \text{für alle } x \in X. \quad (B.33)$$

$$\Phi(\alpha x + \beta y) \stackrel{(B.18)}{=} \alpha \Phi(x) + \beta \Phi(y) \quad \text{für alle } x, y \in X, \alpha, \beta \in \mathbb{K} \quad (B.34)$$

Damit ist nur noch zu zeigen, dass die Norm $\|\cdot\|_{\hat{X}}$ das Parallelogrammgesetz (Prop. A.3.7) erfüllt. Die Behauptung folgt dann mit dem Satz von Fréchet-Jordan-von-Neumann (Satz A.3.8).

Wie in der Beweisführung zu Satz B.4 seien für die nachfolgende Argumentation dazu $\hat{x}, \hat{y} \in \hat{X}$ und $\alpha, \beta \in \mathbb{K}$ fortan beliebig aber fest. Da $\Phi(X)$ dicht in $\hat{X}$ liegt, existieren Folgen $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}} \subset X$ mit $\Phi(x_n) \rightarrow \hat{x}$ und $\Phi(y_n) \rightarrow \hat{y}$ in $(\hat{X}, \|\cdot\|_{\hat{X}})$.

Es gilt dann unter Verwendung von (B.34):

$$\|\hat{x} + \hat{y}\|_{\hat{X}}^2 - \|\hat{x} - \hat{y}\|_{\hat{X}}^2 \stackrel{A.1.14}{=} \lim_{n \rightarrow \infty} \left(\|\Phi(x_n + y_n)\|_{\hat{X}}^2 - \|\Phi(x_n - y_n)\|_{\hat{X}}^2 \right) \quad (B.35)$$

$$\stackrel{(B.33)}{=} \lim_{n \rightarrow \infty} \left(\|x_n + y_n\|_X^2 - \|x_n - y_n\|_X^2 \right) \stackrel{A.3.7}{=} \lim_{n \rightarrow \infty} \left(2\|x_n\|_X^2 + 2\|y_n\|_X^2 \right) \quad (B.36)$$

$$\stackrel{(B.33)}{=} \lim_{n \rightarrow \infty} \left(2\|\Phi(x_n)\|_{\hat{X}}^2 + 2\|\Phi(y_n)\|_{\hat{X}}^2 \right) \stackrel{A.1.14}{=} 2\|\hat{x}\|_{\hat{X}}^2 + 2\|\hat{y}\|_{\hat{X}}^2 \quad (B.37)$$

Somit ist das Parallelogrammgesetz (Prop. A.3.7) erfüllt und mit dem Satz von Fréchet-Jordan-von-Neumann (Satz A.3.8) folgt die Behauptung. $\square$

Korollar B.6 (Vervollständigung und Skalarprodukt [25, Theorem 2.8.2]).

Sei $(\hat{X}, \langle \cdot, \cdot \rangle_{\hat{X}})$ eine Vervollständigung eines euklidischen bzw. unitären Vektorraums $(X, \langle \cdot, \cdot \rangle_X)$ gemäß Satz B.5 und Φ die zugehörige Isometrie, sodass $\Phi(X) = \hat{X}$. Dann gilt für alle $x, y \in X$ mit zugehörigen $\hat{x} := \Phi(x), \hat{y} := \Phi(y) \in \hat{X}$, dass

$$\langle \hat{x}, \hat{y} \rangle_{\hat{X}} = \langle x, y \rangle_X \quad (\text{B.38})$$

In anderen Worten: Die Isometrie behält den Wert des Skalarproduktes zwischen Elementen aus X in dessen Vervollständigung $\hat{X}$ bei.

Beweis. Im Beweis von Satz B.5 wurde zunächst via Satz B.4 ein Banachraum $(\hat{X}, \|\cdot\|_{\hat{X}})$ als Vervollständigung von $(X, \|\cdot\|_X)$ gefunden. Danach wurde die Existenz des Skalarproduktes $\langle \cdot, \cdot \rangle_{\hat{X}}$ mithilfe des Satzes von Fréchet-Jordan-von Neumann nachgewiesen. Letzterer gibt das Skalarprodukt sogar explizit an und es gilt im Fall von $\mathbb{K} = \mathbb{R}$ mit den oben gewählten Bezeichnungen:

$$\langle \hat{x}, \hat{y} \rangle_{\hat{X}} \stackrel{\text{A.3.8}}{=} \frac{1}{4} \left(\|\hat{x} + \hat{y}\|_{\hat{X}}^2 - \|\hat{x} - \hat{y}\|_{\hat{X}}^2 \right) \stackrel{(\text{B.34})}{=} \frac{1}{4} \left(\|\Phi(x+y)\|_{\hat{X}}^2 - \|\Phi(x-y)\|_{\hat{X}}^2 \right) \quad (\text{B.39})$$

$$\stackrel{(\text{B.33})}{=} \frac{1}{4} \left(\|x+y\|_X^2 - \|x-y\|_X^2 \right) \stackrel{\text{A.3.3}}{=} \frac{1}{4} (\langle x+y, x+y \rangle_X - \langle x-y, x-y \rangle_X) \quad (\text{B.40})$$

$$\stackrel{\text{A.3.1}}{=} \frac{1}{4} (\langle x, x \rangle_X + 2\langle x, y \rangle_X + \langle y, y \rangle_X - \langle x, x \rangle_X + 2\langle x, y \rangle_X - \langle y, y \rangle_X) \stackrel{\text{A.3.8}}{=} \langle x, y \rangle_X. \quad (\text{B.41})$$

Diese Rechnung lässt sich identisch ebenso für den Fall $\mathbb{K} = \mathbb{C}$ durchführen, indem man (A.12) anstelle von (A.11) verwendet:

$$\langle \hat{x}, \hat{y} \rangle_{\hat{X}} \stackrel{\text{A.3.8}}{=} \frac{1}{4} \left(\|\hat{x} + \hat{y}\|_{\hat{X}}^2 - \|\hat{x} - \hat{y}\|_{\hat{X}}^2 + i\|\hat{x} + i\hat{y}\|_{\hat{X}}^2 - i\|\hat{x} - i\hat{y}\|_{\hat{X}}^2 \right) \quad (\text{B.42})$$

$$\stackrel{(\text{B.34})}{=} \frac{1}{4} \left(\|\Phi(x+y)\|_{\hat{X}}^2 - \|\Phi(x-y)\|_{\hat{X}}^2 + i\|\Phi(x+iy)\|_{\hat{X}}^2 - i\|\Phi(x-iy)\|_{\hat{X}}^2 \right) \quad (\text{B.43})$$

$$\stackrel{(\text{B.33})}{=} \frac{1}{4} \left(\|x+y\|_X^2 - \|x-y\|_X^2 + i\|x+iy\|_X^2 - i\|x-iy\|_X^2 \right) \quad (\text{B.44})$$

$$\stackrel{\text{A.3.3}}{=} \frac{1}{4} (\langle x+y, x+y \rangle_X - \langle x-y, x-y \rangle_X + i\langle x+iy, x+iy \rangle_X - i\langle x-iy, x-iy \rangle_X) \quad (\text{B.45})$$

$$\stackrel{\text{A.3.1}}{=} \frac{1}{4} (\langle x, x \rangle_X + \langle y, y \rangle_X - \langle x, x \rangle_X - \langle y, y \rangle_X + 4\text{Re}\langle x, y \rangle_X \quad (\text{B.46})$$

$$+ i\langle x, x \rangle_X + i\langle y, y \rangle_X - i\langle x, x \rangle_X - i\langle y, y \rangle_X + 4\text{Im}\langle x, y \rangle_X) \stackrel{\text{A.3.8}}{=} \langle x, y \rangle_X. \quad (\text{B.47})$$

□



Absolutsummierende Hilbert-Schmidt Operatoren

Bei der Recherche zu einem korrekten Beweis für den **Satz 6.5.1 von Mercer** in der Arbeit, sind Erkenntnisse zu **absolutsummierenden** und **Hilbert-Schmidt (Integral-)Operatoren** erarbeitet worden, welche aber nicht gewinnbringend eingesetzt werden konnten. Der Vollständigkeit halber seien diese hier aber trotzdem aufgeführt. Die Grundlage dieses Anhangs sind [13, 28].

C.1 Absolutsummierende Operatoren

Definition C.1.1 (Absolutsummierende Operatoren [28, Def. 2.1.1, Satz 2.2.1]).

Seien X, Y zwei normierte Räume. Ein Operator $T : X \rightarrow Y$ heißt **(absolut)summierend**, wenn ein $c > 0$ existiert, sodass für alle $n \in \mathbb{N}$ und jede endliche Folge $(x_k)_{k=1}^n \subset X$ gilt, dass (vgl. Def. A.4.7)

$$\sum_{k=1}^n \|Tx_k\|_Y \leq c \sup_{\|x'\|_{X'} \leq 1} \sum_{k=1}^n |\langle x', x_k \rangle_{X', X}|. \quad (\text{C.1})$$

Es stellt sich heraus, dass jeder Hilbert-Schmidt Integraloperator ein absolutsummierender Operator ist. Auf dem Weg dorthin werden noch drei eher theoretische, aber miteinander zusammenhängende analytische Ergebnisse benötigt: Die sog. **Rademacherschen Funktionen**, die **Chintschinsche Ungleichung** und der **Einbettungsoperator von $\ell^1(\mathbb{N})$ nach $\ell^2(\mathbb{N})$** .

C.2 Die Rademacherschen Funktionen

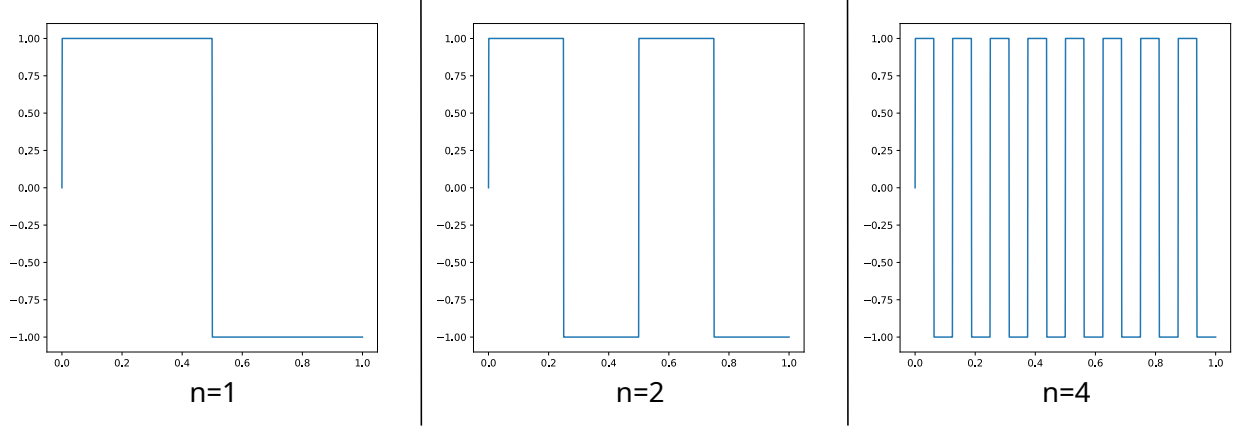


Abbildung C.1: Rademachersche Funktionen r_1, r_2 und r_4

Definition und Satz C.2.1 (Rademachersche Funktionen [13, 28, 2.4.1]).

Als die **Rademacherschen Funktionen** werden für $n \in \mathbb{N}$ bezeichnet (siehe [Abbildung C.1](#)):

$$r_n : [0, 1] \rightarrow \mathbb{R}, \quad r_n(t) := \operatorname{sgn}(\sin(2^n \pi t)) = \begin{cases} 1, & \text{für } \frac{k}{2^n} \leq t < \frac{k+1}{2^n} \text{ mit } k = 0, 2, \dots, 2^n - 2 \\ -1 & \text{für } \frac{k}{2^n} \leq t < \frac{k+1}{2^n} \text{ mit } k = 1, 3, \dots, 2^n - 1. \end{cases} \quad (\text{C.2})$$

Für sie gilt eine Orthonormalitätsbedingung der folgenden Form: Sei $m \in \mathbb{N}_0$ und $n_1, \dots, n_m \in \mathbb{N}$ mit $n_1 \geq \dots \geq n_m$ (was wegen der Kommutativität des Produktes im folgenden Integralausdruck o. B. d. A. angenommen werden darf). Dann gilt:

$$\int_0^1 \prod_{i=1}^m r_{n_i}(t) dt = \begin{cases} 1, & \text{falls } m \text{ gerade und } n_i = n_{i+1} \text{ für } i = 1, 3, 5, \dots, m-1 \\ 0, & \text{sonst.} \end{cases} \quad (\text{C.3})$$

Beweis. (Die zentrale Beweisidee wurde [13] entnommen.)

Betrachtet man die periodische Fortsetzung der Funktion r_1 , nämlich $r : [0, \infty) \rightarrow \mathbb{R}$ mit

$$r(t) := \begin{cases} 1, & 0 \leq t < \frac{1}{2} \\ -1, & \frac{1}{2} \leq t < 1, \end{cases} \quad r(t+1) := r(t) \quad \text{für alle } t \in [1, \infty) \quad (\text{C.4})$$

so erkennt man für $t \in [0, 1]$ und alle $n \in \mathbb{N}$:

$$r(2^{n-1}t) = \begin{cases} 1, & \text{falls } k \leq 2^{n-1}t < k + \frac{1}{2} \text{ für } k \in \mathbb{N}_0 \\ -1, & \text{falls } k + \frac{1}{2} \leq 2^{n-1}t < k + 1 \text{ für } k \in \mathbb{N}_0 \end{cases} \quad (\text{C.5})$$

$$= \begin{cases} 1, & \text{für } \frac{2k}{2^n} \leq t < \frac{2k+1}{2^n} \text{ mit } k \in \mathbb{N}_0 \\ -1 & \text{für } \frac{2k+1}{2^n} \leq t < \frac{2k+2}{2^n} \text{ mit } k \in \mathbb{N}_0 \end{cases} \quad (\text{C.6})$$

$$= r_n(t). \quad (\text{C.7})$$

Mithilfe dieser Darstellung $r_n = r(2^{n-1} \cdot)$ der Rademacherschen Funktionen für alle $n \in \mathbb{N}$ lässt sich nun mittels vollständiger Induktion über $m \in \mathbb{N}_0$ die Behauptung zeigen:

Induktionsanfang:

$m = 0$: Es gilt

$$\int_0^1 \prod_{i=1}^0 r_{n_i}(t) dt = \int_0^1 1 dt = 1. \quad \checkmark \quad (\text{C.8})$$

$m = 1$: Es gilt

$$\int_0^1 \prod_{i=1}^1 r_{n_i}(t) dt = \int_0^1 r(2^{n_1-1}t) dt = \frac{1}{2^{n_1-1}} \int_0^{2^{n_1-1}} r(u) du \quad [\text{Substitution}] \quad (\text{C.9})$$

$$= \frac{2^{n_1-1}}{2^{n_1-1}} \int_0^1 r(u) du = \int_0^{\frac{1}{2}} 1 du - \int_{\frac{1}{2}}^1 1 du = 0. \quad \checkmark \quad [\text{Periode von } r \text{ und } 2^{n_1-1} \geq 1] \quad (\text{C.10})$$

Induktionsannahme: Sei $m > 0$ fest und es gelte für $\tilde{m} \in \{m, m-1\}$

$$\int_0^1 \prod_{i=1}^{\tilde{m}} r_{n_i}(t) dt = \begin{cases} 1, & \text{falls } \tilde{m} \text{ gerade und } n_i = n_{i+1} \text{ für } i = 1, 3, 5, \dots, \tilde{m}-1 \\ 0, & \text{sonst.} \end{cases} \quad (\text{C.11})$$

Induktionsschritt: $m \rightsquigarrow m+1$:

Fall 1: $n_m = n_{m+1}$:

$$\int_0^1 \prod_{i=1}^{m+1} r_{n_i}(t) dt = \int_0^1 \prod_{i=1}^m r_{n_i}(t) \cdot 1 dt = \begin{cases} 1, & \text{falls } m+1 \text{ gerade und } n_i = n_{i+1} \text{ für } i = 1, 3, 5, \dots, m \\ 0, & \text{sonst.} \end{cases} \quad (\text{C.12})$$

[Unter Verwendung der Induktionsannahme in (C.12)]

Fall 2: $n_m \neq n_{m+1} \Rightarrow n_m > n_{m+1}$:

$$\int_0^1 \prod_{i=1}^{m+1} r_{n_i}(t) dt = \int_0^1 \prod_{i=1}^m r_{n_i}(t) \cdot r_{n_{m+1}}(t) dt \quad (\text{C.13})$$

$$= \int_0^1 \prod_{i=1}^m r(2^{n_i-1}t) \cdot r(2^{n_{m+1}-1}t) dt \quad (\text{C.14})$$

$$= \frac{1}{2^{n_{m+1}-1}} \int_0^{2^{n_{m+1}-1}} \prod_{i=1}^m r(2^{n_i-n_{m+1}}u) \cdot r(u) du \quad [\text{Substitution}] \quad (\text{C.15})$$

$$(\text{C.16})$$

$$= \frac{2^{n_{m+1}-1}}{2^{n_{m+1}-1}} \int_0^1 \prod_{i=1}^m r(2^{n_i-n_{m+1}}u) \cdot r(u) du \quad [\text{Periode von } r \text{ und } 2^{n_{m+1}-1} \geq 1 \text{ und } 2^{n_i-n_{m+1}} \geq 1] \quad (\text{C.17})$$

$$= \int_0^{\frac{1}{2}} \prod_{i=1}^m r(2^{n_i-n_{m+1}}u) \, du - \int_{\frac{1}{2}}^1 \prod_{i=1}^m r(2^{n_i-n_{m+1}}u) \, du \quad (\text{C.18})$$

$$= \frac{1}{2} \left(\int_0^1 \prod_{i=1}^m r(2^{n_i-n_{m+1}-1}u) \, du - \int_1^2 \prod_{i=1}^m r(2^{n_i-n_{m+1}-1}u) \, du \right) \quad [\text{Substitution}] \quad (\text{C.19})$$

$$= \frac{1}{2} \int_0^1 \prod_{i=1}^m r(2^{n_i-n_{m+1}-1}u) \, du - \prod_{i=1}^m r(2^{n_i-n_{m+1}-1}u) \, du \quad [\text{Periode von } r \text{ und } 2^{n_i-n_{m+1}-1} \geq 1] \quad (\text{C.20})$$

$$= 0. \quad (\text{C.21})$$

Mit dem Prinzip der vollständigen Induktion folgt die Behauptung. □

Korollar C.2.2 (Rademachersches Orthonormalsystem [13, 28, 2.4.1]).

Die Rademacherschen Funktionen bilden ein Orthonormalsystem des $(L^2([0,1]), \langle \cdot, \cdot \rangle_{L^2([0,1])})$.

Beweis. Es gilt für $n, m \in \mathbb{N}$:

$$\langle r_n, r_m \rangle_{L^2([0,1])} \stackrel{\text{A.3.9}}{=} \int_0^1 r_n(t) r_m(t) \, dt \stackrel{\text{C.2.1}}{=} \begin{cases} 1, & \text{für } n = m \\ 0, & \text{für } n \neq m. \end{cases} \quad (\text{C.22})$$

□

Mithilfe der **Rademacherschen Funktionen** lässt sich nun die nützliche **Chintschinsche Ungleichung** aufstellen und beweisen.

C.3 Die Chintschinsche Ungleichung

Proposition C.3.1 (Chintschinsche Ungleichung [28, 2.4.1]).

Es seien $n \in \mathbb{N}$ und $x_1, \dots, x_n \in \mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ reelle oder komplexe Zahlen. Dann gilt mit den Rademacherschen Funktionen $(r_i)_{i \in \mathbb{N}}$:

$$\left(\sum_{i=1}^n |x_i|^2 \right) \leq \sqrt{3} \int_0^1 \left| \sum_{i=1}^n x_i r_i(t) \right| \, dt. \quad (\text{C.23})$$

Beweis. (Der Beweis ist [28, 2.4.1] entnommen und mit ausführlicheren Schritten ergänzt.)
Man setze

$$\alpha := \int_0^1 \left| \sum_{i=1}^n x_i r_i(t) \right|^2 dt = \int_0^1 \sum_{i=1}^n x_i r_i(t) \overline{\sum_{j=1}^n x_j r_j(t)} dt \stackrel{\text{A.3.9}}{=} \left\langle \sum_{i=1}^n x_i r_i, \sum_{j=1}^n x_j r_j \right\rangle_{L^2([0,1])} \quad (\text{C.24})$$

$$\stackrel{\text{A.3.1}}{=} \sum_{i=1}^n \sum_{j=1}^n x_i \bar{x}_j \langle r_i, r_j \rangle_{L^2([0,1])} \stackrel{\text{C.2.2}}{=} \sum_{i=1}^n x_i \bar{x}_i = \sum_{i=1}^n |x_i|^2. \quad (\text{C.25})$$

Mit denselben Methoden erhält man (mit den reellen Rademacherschen Funktionen):

$$\int_0^1 \left| \sum_{i=1}^n x_i r_i(t) \right|^4 dt = \int_0^1 \left(\sum_{i=1}^n x_i r_i(t) \overline{\sum_{j=1}^n x_j r_j(t)} \right) \cdot \left(\sum_{k=1}^n x_k r_k(t) \overline{\sum_{l=1}^n x_l r_l(t)} \right) dt \quad (\text{C.26})$$

$$\stackrel{\text{A.3.9}}{=} \left| \left\langle \sum_{i=1}^n x_i r_i \overline{\sum_{j=1}^n x_j r_j}, \sum_{k=1}^n x_k r_k \overline{\sum_{l=1}^n x_l r_l} \right\rangle_{L^2([0,1])} \right| \stackrel{\text{A.3.1}}{=} \left| \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n x_i \bar{x}_j \bar{x}_k x_l \langle r_i r_j, r_k r_l \rangle_{L^2([0,1])} \right| \quad (\text{C.27})$$

$$\stackrel{\text{A.3.9}}{=} \left| \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n x_i \bar{x}_j \bar{x}_k x_l \int_0^1 r_i r_j r_k r_l dt \right| \stackrel{(\text{C.3})}{=} \left| 2 \sum_{i=1}^n \sum_{j=1}^n x_i \bar{x}_i \bar{x}_j x_j + \sum_{i=1}^n \sum_{j=1}^n x_i \bar{x}_j \bar{x}_j x_i \right| \quad (\text{C.28})$$

$$\leq 2 \sum_{i=1}^n \sum_{j=1}^n |x_i|^2 |x_j|^2 + \sum_{i=1}^n \sum_{j=1}^n |x_i|^2 |\bar{x}_j|^2 = 3 \left(\sum_{i=1}^n |x_i|^2 \right)^2, \stackrel{(\text{C.25})}{=} 3\alpha^2 \quad (\text{C.29})$$

Mithilfe der Hölderschen Ungleichung (Lemma A.2.15) mit $p = \frac{3}{2}$ und $q = 3$ folgt nun:

$$\alpha^3 \stackrel{(\text{C.25})}{=} \left(\int_0^1 \left| \sum_{i=1}^n x_i r_i(t) \right|^2 dt \right)^3 = \left(\int_0^1 \left| \sum_{i=1}^n x_i r_i(t) \right|^{\frac{2}{3}} \left| \sum_{i=1}^n x_i r_i(t) \right|^{\frac{4}{3}} dt \right)^3 \quad (\text{C.30})$$

$$\stackrel{\text{A.2.14}}{=} \left\| \left| \sum_{i=1}^n x_i r_i \right|^{\frac{2}{3}} \left(\left| \sum_{i=1}^n x_i r_i \right|^4 \right)^{\frac{1}{3}} \right\|_{L^1(\Omega, \mu)}^3 \stackrel{\text{A.2.15}}{\leq} \left\| \left| \sum_{i=1}^n x_i r_i \right|^{\frac{2}{3}} \right\|_{L^{\frac{3}{2}}(\Omega, \mu)}^3 \left\| \left(\left| \sum_{i=1}^n x_i r_i \right|^4 \right)^{\frac{1}{3}} \right\|_{L^3(\Omega, \mu)}^3 \quad (\text{C.31})$$

$$\stackrel{\text{A.2.14}}{=} \left(\int_0^1 \left| \sum_{i=1}^n x_i r_i(t) \right|^{\frac{2}{3} \cdot \frac{3}{2}} dt \right)^{\frac{2}{3} \cdot 3} \left(\int_0^1 \left| \sum_{i=1}^n x_i r_i(t) \right|^{4 \cdot \frac{1}{3} \cdot 3} dt \right)^{\frac{1}{3} \cdot 3} \stackrel{(\text{C.29})}{\leq} \left(\int_0^1 \left| \sum_{i=1}^n x_i r_i(t) \right|^2 dt \right)^2 \cdot 3\alpha^2. \quad (\text{C.32})$$

Insgesamt folgt damit die Behauptung:

$$\left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} \stackrel{(\text{C.25})}{=} \sqrt{\alpha} \stackrel{(\text{C.32})}{\leq} \sqrt{3} \int_0^1 \left| \sum_{i=1}^n x_i r_i(t) \right| dt. \quad (\text{C.33})$$

□

Bemerkung C.3.2. Die vollständige **Chintschinsche Ungleichung** enthält noch eine weitere Abschätzung nach oben, welche die rechte Seite erneut mit der linken und einem konstanten Vorfaktor abschätzt. Sie spielt insbesondere in der Wahrscheinlichkeitsrechnung/Statistik eine wichtige Rolle. Seit der hier zitierten Version in [28, 2.4.1] aus 1969 wurden zudem Verbesserungen des Faktors $\sqrt{3}$ gezeigt. Für die Argumentation dieser Arbeit genügt aber die Version von [28, 2.4.1].

Die **Chintschinsche Ungleichung** gemeinsam mit den **Rademacherschen Funktionen** liefern nun das letzte Puzzleteil, um **Hilbert-Schmidt Integraloperatoren** als **absolutsummierend** nachzuweisen. (Für die Definition der $\ell^p(\mathbb{N})$ -Räume für $p \in [1, \infty]$, betrachte man [Bsp. A.2.11.](#))

C.4 Der Einbettungsoperator von $\ell^1(\mathbb{N})$ nach $\ell^2(\mathbb{N})$

Proposition C.4.1 (Einbettungsoperator von $\ell^1(\mathbb{N})$ nach $\ell^2(\mathbb{N})$ [28, Satz 2.4.2]).

Der Einbettungsoperator $\iota : \ell^1(\mathbb{N}) \rightarrow \ell^2(\mathbb{N})$, $\iota(x) = x$ für alle $x \in \ell^1(\mathbb{N})$ ist absolutsummierend.

Beweis. (Der Beweis ist [28, Satz 2.4.2] entnommen und ausformuliert.)

Zuerst einmal sei angemerkt, dass ι wegen $\ell^2(\mathbb{N}) \subset \ell^1(\mathbb{N})$ und $\iota(\alpha x + \beta y) = \alpha x + \beta y = \alpha \iota(x) + \beta \iota(y)$ für beliebige $x, y \in X := \ell^1(\mathbb{N})$, $\alpha, \beta \in \mathbb{K}$ ein wohldefinierter linearer Operator ist.

Sei nun $m \in \mathbb{N}$ beliebig aber fest. Man betrachte zunächst für $t \in [0, 1]$ die Funktionale

$$x'_t : X \rightarrow \mathbb{K}, \quad x'_t(x) := \sum_{i=1}^m x_i r_i(t) \quad \text{für alle } x = (x_i)_{i \in \mathbb{N}} \in X, \quad (\text{C.34})$$

wobei $(r_i)_{i \in \mathbb{N}}$ die Rademacherschen Funktionen darstellen. Dann ist für jedes $t \in [0, 1]$ das Funktional x'_t linear, da für beliebige $x = (x_i)_{i \in \mathbb{N}}, y = (y_i)_{i \in \mathbb{N}} \in X$, $\alpha, \beta \in \mathbb{K}$ gilt:

$$x'_t(\alpha x + \beta y) \stackrel{(\text{C.34})}{=} \sum_{i=1}^m (\alpha x + \beta y)_i r_i(t) = \alpha \sum_{i=1}^m x_i r_i(t) + \beta \sum_{i=1}^m y_i r_i(t) \stackrel{(\text{C.34})}{=} \alpha x'_t(x) + \beta x'_t(y). \quad (\text{C.35})$$

Zudem ist für jedes $t \in [0, 1]$ das lineare Funktional beschränkt wegen $\sup_{i \in \mathbb{N}} \sup_{t \in [0, 1]} |r_i(t)| = 1$ und

$$|x'_t(x)| \stackrel{(\text{C.34})}{=} \left| \sum_{i=1}^m x_i r_i(t) \right| \leq \sum_{i=1}^m |x_i r_i(t)| \leq \sum_{i=1}^m |x_i| \leq \sum_{i=1}^{\infty} |x_i| \stackrel{\text{A.2.11}}{=} \|x\|_X \quad \text{für alle } x = (x_i)_{i \in \mathbb{N}} \in X. \quad (\text{C.36})$$

Somit gilt für alle $(x'_t)_{t \in [0, 1]}$: $\|x'_t\|_{X'} \leq 1$.

Mit diesem Vorergebnis ist die Behauptung nicht mehr weit. Seien dazu nun zusätzlich $n \in \mathbb{N}$ sowie $x_1 = (x_{1,i})_{i \in \mathbb{N}}, \dots, x_n = (x_{n,i})_{i \in \mathbb{N}} \in X$ beliebig, fest. Es folgt mit der Chintschinschen Ungleichung:

$$\sum_{k=1}^n \left(\sum_{i=1}^m |x_{k,i}|^2 \right)^{\frac{1}{2}} \stackrel{\text{C.3.1}}{\leq} \sum_{k=1}^n \sqrt{3} \int_0^1 \left| \sum_{i=1}^m x_{k,i} r_i(t) \right| dt \stackrel{(\text{C.34})}{=} \sqrt{3} \int_0^1 \sum_{k=1}^n |\langle x'_t, x_k \rangle_{X', X}| dt \quad (\text{C.37})$$

$$\leq \sqrt{3} \int_0^1 \sup_{\|x'\|_{X'} \leq 1} \sum_{k=1}^n |\langle x', x_k \rangle_{X', X}| dt = \sqrt{3} \sup_{\|x'\|_{X'} \leq 1} \sum_{k=1}^n |\langle x', x_k \rangle_{X', X}| \quad (\text{C.38})$$

Da $m \in \mathbb{N}$ beliebig gewählt war, folgt damit die Behauptung – ι ist absolutsummierend:

$$\sum_{k=1}^n \|\iota(x_k)\|_2 \stackrel{\text{A.2.11}}{=} \lim_{m \rightarrow \infty} \sum_{k=1}^n \left(\sum_{i=1}^m |x_{k,i}|^2 \right)^{\frac{1}{2}} \stackrel{\text{(C.38)}}{\leq} \sqrt{3} \sup_{\|x'\|_{X'} \leq 1} \sum_{k=1}^n |\langle x', x_k \rangle_{X', X}|. \quad (\text{C.39})$$

□

Korollar C.4.2 (Einbettungsoperator von $\ell^1(\mathbb{N})$ nach $\ell^2(\mathbb{N})$ ist beschränkt [28, Satz 2.4.2]).
Der Einbettungsoperator ι aus Prop. C.4.1 ist beschränkt, d. h. $\iota \in \mathcal{L}(\ell^1(\mathbb{N}), \ell^2(\mathbb{N}))$.

Beweis. Für $x \in \ell^1(\mathbb{N})$ gilt mit Kor. A.4.8 und, da ι absolutsummierend ist:

$$\|\iota(x)\|_2 \stackrel{\text{(C.39)}}{\leq} \sqrt{3} \sup_{\|x'\|_{(\ell^1(\mathbb{N}))'} \leq 1} |\langle x', x \rangle_{(\ell^1(\mathbb{N}))', \ell^1(\mathbb{N})}| \stackrel{\text{A.4.8}}{=} \sqrt{3} \|x\|_1. \quad (\text{C.40})$$

□

Damit sind nun – mit einiger Vorarbeit – endlich alle Zutaten vorbereitet, um den seit einigen Seiten angekündigten Zusammenhang zwischen **Hilbert-Schmidt Integraloperatoren** und **absolutsummierenden Operatoren** zu zeigen.

C.5 Absolutsummierende Hilbert-Schmidt (Integral-)Operatoren

Auf dem Hilbertraum $H := L^2(\Omega, \mu)$ werden im Folgenden Hilbert-Schmidt Integraloperatoren betrachtet, wobei (Ω, μ) ein Maßraum ist.

Satz C.5.1 (Selbstadjungierte Hilbert-Schmidt Integraloperatoren sind absolutsummierend [28, Theorem 2.5.5]).

Sei (Ω, μ) ein Maßraum und $T_k : H \rightarrow H$ ein selbstadjungierter Hilbert-Schmidt Integraloperator darauf, d. h. es existiere ein $k : \Omega \times \Omega \rightarrow \mathbb{K}$ und $M > 0$ mit

$$(T_k f)(x) = \int_{\Omega} k(x, y) f(y) d\mu(y) \stackrel{\text{A.3.9}}{=} \begin{cases} \langle k(x, \cdot), \bar{f} \rangle_H \\ \langle f, \overline{k(x, \cdot)} \rangle_H \end{cases} \quad \text{für alle } f \in H, x \in \Omega, \quad (\text{C.41})$$

$$\langle T_k f, g \rangle_H = \langle f, T_k g \rangle_H \quad \text{für alle } f, g \in H \quad \text{und} \quad (\text{C.42})$$

$$\int_{\Omega} \int_{\Omega} |k(x, y)|^2 d\mu(y) d\mu(x) = M < \infty. \quad (\text{C.43})$$

Dann ist T_k absolutsummierend.

Beweis. (Der Beweis ist [28, Theorem 2.5.5] entnommen und etwas ausführlicher formuliert.)

Als Hilbert-Schmidt Integraloperator ist T_k gemäß Def.+Satz A.4.18 auch kompakt und, da er voraussetzungsgemäß selbstadjungiert ist, lässt sich der Spektralsatz für kompakte, selbstadjungierte Operatoren anwenden. Es folgt, dass T_k die folgende Darstellung besitzt:

$$T_k f = \sum_{i \in I} \lambda_i \langle f, e_i \rangle_H e_i \quad \text{für alle } f \in H, \quad (\text{C.44})$$

wobei I eine höchstens abzählbar unendliche Indexmenge ist, $(\lambda_i)_{i \in I} \subset \mathbb{C}$ die mit Multiplizität gezählten Eigenwerte von T_k und $(e_i)_{i \in I} \in H$ die jeweils zugehörigen, paarweise verschiedenen orthonormalen Eigenfunktionen sind (d. h. $\|e_i\|_H = 1$ für alle $i \in I$).

Sei aus diesem Grund fortan o. B. d. A. $I = \mathbb{N}$ abzählbar unendlich, indem $(\lambda_i)_{i \in I}$ und $(e_i)_{i \in I}$ mit Folgengliedern konstant Null unendlich oft fortgesetzt seien, falls I endlich sein sollte.

Als Erstes zeige man nun mit der Besselschen Ungleichung, dass für T_k als Hilbert-Schmidt Integraloperator auf einem Hilbertraum die Summe seiner Eigenwertquadrate endlich ist:

$$\sum_{n=1}^{\infty} |\lambda_n|^2 \stackrel{\text{A.3.12}}{=} \sum_{n=1}^{\infty} |\lambda_n|^2 \|e_n\|_H^2 \stackrel{\text{A.2.4}}{=} \sum_{n=1}^{\infty} \|\lambda_n e_n\|_H^2 \stackrel{\text{A.4.15}}{=} \sum_{n=1}^{\infty} \|T_k e_n\|_H^2 \quad (\text{C.45})$$

$$\stackrel{\text{A.2.14}}{=} \sum_{n=1}^{\infty} \int_{\Omega} |(T_k e_n)(x)|^2 d\mu(x) \stackrel{(\text{C.41})}{\stackrel{\text{A.3.1}}{=}} \sum_{n=1}^{\infty} \int_{\Omega} |\overline{\langle k(x, \cdot), e_n \rangle_H}|^2 d\mu(x) \quad (\text{C.46})$$

$$\stackrel{\text{A.6.19}}{\stackrel{\text{A.8.1}}{=}} \int_{\Omega} \sum_{n=1}^{\infty} |\overline{\langle k(x, \cdot), e_n \rangle_H}|^2 d\mu(x) \stackrel{\text{A.3.14}}{\stackrel{\text{A.6.15}}{\leq}} \int_{\Omega} \|\overline{k(x, \cdot)}\|_H^2 d\mu(x) \quad (\text{C.47})$$

$$\stackrel{\text{A.2.14}}{=} \int_{\Omega} \int_{\Omega} |\overline{k(x, y)}|^2 d\mu(y) d\mu(x) \stackrel{\text{A.8.1}}{=} \int_{\Omega} \int_{\Omega} |k(x, y)|^2 d\mu(y) d\mu(x) \stackrel{(\text{C.43})}{=} M < \infty. \quad (\text{C.48})$$

(Im Schritt von (C.46) nach (C.47) durften Integration und Reihe dabei mit dem Satz A.6.19 von Lebesgue getauscht werden, da $\sum_{n=1}^m |\overline{\langle k(x, \cdot), e_n \rangle_H}|^2$ für alle $m \in \mathbb{N}$ nicht-negativ sowie als Komposition und elementare Kombination messbarer Funktionen (Def. A.6.9) gemäß Satz A.6.11+Satz A.6.13 ebenfalls messbar ist, und mit $\|\overline{k(x, \cdot)}\|_H^2$ eine integrierbare Majorante (Def. A.6.14) existiert.)

Man definiere dann die drei Operatoren

$$R : D_R := H \rightarrow \ell^1(\mathbb{N}), \quad Rf := (\lambda_n \langle f, e_n \rangle_H)_{n \in \mathbb{N}}, \quad (\text{C.49})$$

$$\iota : D_{\iota} := R(D_R) \subseteq \ell^1(\mathbb{N}) \rightarrow \ell^2(\mathbb{N}), \quad \iota(x) := x \quad \text{und} \quad (\text{C.50})$$

$$L : D_L := \iota(D_{\iota}) \subseteq \ell^2(\mathbb{N}) \rightarrow H, \quad Ly := \sum_{n=1}^{\infty} y_n e_n \quad (\text{C.51})$$

für $f \in D_R$, $x \in D_{\iota}$ und $y = (y_n)_{n \in \mathbb{N}} \in D_L$. In Kor. C.4.2 wurde bereits nachgewiesen, dass $\iota \in \mathcal{L}(D_{\iota}, \ell^2(\mathbb{N}))$ gilt (d. h. insbesondere wohldefiniert ist) und in Prop. C.4.1, dass ι absolutsummierend ist. Zudem ist die Reihe in (C.51) konvergent in H , da $D_L \stackrel{(\text{C.50})}{=} \iota(R(H))$ als Teilmenge von $\ell^2(\mathbb{N})$ gerade dergleichen gewählt ist, dass mit der Konvergenz von (C.44) die Konvergenz in (C.51) sichergestellt ist. Die (restliche) Wohldefiniertheit sowie die Beschränktheit von R bzw. L folgen dann gemäß Def.+Satz A.4.2 aus (wobei das Vorergebnis der endlichen Eigenwertquadrate notwendig wird):

$$\|Rf\|_1 \stackrel{(C.49)}{=} \|(\lambda_n)_{n \in \mathbb{N}} \cdot (\langle f, e_n \rangle_H)_{n \in \mathbb{N}}\|_1 \stackrel{A.2.13}{\leq} \|(\lambda_n)_{n \in \mathbb{N}}\|_2 \cdot \|(\langle f, e_n \rangle_H)_{n \in \mathbb{N}}\|_2 \quad (C.52)$$

$$\stackrel{A.2.11}{=} \left(\left(\sum_{n=1}^{\infty} |\lambda_n|^2 \right) \left(\sum_{n=1}^{\infty} |\langle f, e_n \rangle_H|^2 \right) \right)^{\frac{1}{2}} \stackrel{(C.48)}{\leq} \sqrt{M} \|f\|_H \quad \text{für alle } f \in D_R = H \text{ und} \quad (C.53)$$

$$\|Ly\|_H \stackrel{A.3.3}{=} \left(\left\langle \sum_{n=1}^{\infty} y_n e_n, \sum_{m=1}^{\infty} y_m e_m \right\rangle_H \right)^{\frac{1}{2}} \stackrel{A.3.1}{=} \left(\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} y_n \bar{y}_m \langle e_n, e_m \rangle_H \right)^{\frac{1}{2}} \stackrel{A.3.6}{=} \quad (C.54)$$

$$\stackrel{A.3.12}{\leq} \left(\sum_{n=1}^{\infty} y_n \bar{y}_n \right)^{\frac{1}{2}} \stackrel{A.8.1}{=} \left(\sum_{n=1}^{\infty} |y_n|^2 \right)^{\frac{1}{2}} \stackrel{A.2.11}{=} \|y\|_2 \quad \text{für alle } y \in D_L \subseteq \ell^2(\mathbb{N}). \quad (C.55)$$

R und L sind zudem linear, da für alle $f, g \in D_R$, $y := (y_n)_{n \in \mathbb{N}}$, $z := (z_n)_{n \in \mathbb{N}} \in D_L$, $\alpha, \beta \in \mathbb{K}$ gilt:

$$R(\alpha f + \beta g) \stackrel{(C.49)}{=} (\lambda_n \langle \alpha f + \beta g, e_n \rangle_H)_{n \in \mathbb{N}} \stackrel{A.3.1}{=} \alpha (\lambda_n \langle f, e_n \rangle_H)_{n \in \mathbb{N}} + \beta (\lambda_n \langle g, e_n \rangle_H)_{n \in \mathbb{N}} \quad (C.56)$$

$$\stackrel{(C.49)}{=} \alpha Rf + \beta Rg \quad \text{und} \quad (C.57)$$

$$L(\alpha y + \beta z) \stackrel{(C.51)}{=} \sum_{n=1}^{\infty} (\alpha y_n + \beta z_n) e_n = \alpha \sum_{n=1}^{\infty} y_n e_n + \beta \sum_{n=1}^{\infty} z_n e_n \stackrel{(C.51)}{=} \alpha Ly + \beta Lz. \quad (C.58)$$

(In (C.58) durfte dabei die Linearität der Reihe angenommen werden, da die beiden auseinander gezogenen Einzelreihen, wie bereits erörtert, konvergieren, d. h. deren Grenzwerte existieren.)

Somit sind ebenfalls $R \in \mathcal{L}(D_R = H, \ell^1(\mathbb{N}))$ und $L \in \mathcal{L}(D_L, H)$ nachgewiesen. Schließlich ergibt die Definition von R, ι und L offensichtlich gerade die Faktorisierung $T_k = L\iota R$, denn für alle $f \in H$ gilt:

$$L\iota Rf \stackrel{(C.49)}{=} (L\iota)((\lambda_n \langle f, e_n \rangle_H)_{n \in \mathbb{N}}) \stackrel{(C.50)}{=} L((\lambda_n \langle f, e_n \rangle_H)_{n \in \mathbb{N}}) \stackrel{(C.51)}{=} \sum_{n=1}^{\infty} \lambda_n \langle f, e_n \rangle_H e_n \stackrel{(C.44)}{=} T_k f. \quad (C.59)$$

Mit dieser relativ aufwendigen Vorarbeit folgt nun endlich die Behauptung, wobei die absolutsummierende Eigenschaft von ι und der adjungierte Operator von R den Hauptbeitrag leisten. Für beliebige $n \in \mathbb{N}$ und $f_1, \dots, f_n \in H$ gilt:

$$\sum_{i=1}^n \|T_k f_i\|_H \stackrel{(C.59)}{=} \sum_{i=1}^n \|L(\iota(Rf_i))\|_H \stackrel{(C.55)}{\leq} \sum_{i=1}^n 1 \cdot \|\iota(Rf_i)\|_2 \quad (C.60)$$

$$\stackrel{(C.39)}{\leq} \sqrt{3} \sup_{\|x'\|_{(\ell^1(\mathbb{N}))'} \leq 1} \sum_{i=1}^n \left| \langle x', Rf_i \rangle_{(\ell^1(\mathbb{N}))', \ell^1(\mathbb{N})} \right| \stackrel{A.4.13}{=} \sqrt{3} \sup_{\|x'\|_{(\ell^1(\mathbb{N}))'} \leq 1} \sum_{i=1}^n |\langle R'x', f_i \rangle_{H', H}| \quad (C.61)$$

$$\stackrel{A.4.13}{=} \sqrt{3} \sup_{\|x'\|_{(\ell^1(\mathbb{N}))'} \leq 1} \sum_{i=1}^n \left| \left\langle \frac{\|R\|_{\mathcal{L}(H, \ell^1(\mathbb{N}))}}{\|R'\|_{\mathcal{L}((\ell^1(\mathbb{N}))', H')}} R'x', f_i \right\rangle_{H', H} \right| \quad (C.62)$$

$$\stackrel{A.3.1}{\leq} \sqrt{3} \|R\|_{\mathcal{L}(H, \ell^1(\mathbb{N}))} \sup_{\|x'\|_{H'} \leq 1} \sum_{i=1}^n |\langle x', f_i \rangle_{H', H}| \stackrel{(C.53)}{\leq} \sqrt{3M} \sup_{\|x'\|_{H'} \leq 1} \sum_{i=1}^n |\langle x', f_i \rangle_{H', H}|. \quad (C.63)$$

Gemäß Def. C.1.1 ist T_k somit wie behauptet absolutsummierend. (Im letzten Schritt wurde dabei verwendet, dass wegen $R' \in \mathcal{L}((\ell^1(\mathbb{N}))', H')$ für alle $x' \in (\ell^1(\mathbb{N}))'$ mit $\|x'\|_{(\ell^1(\mathbb{N}))'} \leq 1$ gilt, dass

$$\left\| \frac{1}{\|R'\|_{\mathcal{L}((\ell^1(\mathbb{N}))', H')}} R' x' \right\|_{H'} \stackrel{\text{A.4.4}}{\leq} \frac{\|R'\|_{\mathcal{L}((\ell^1(\mathbb{N}))', H')}}{\|R'\|_{\mathcal{L}((\ell^1(\mathbb{N}))', H')}} \|x'\|_{(\ell^1(\mathbb{N}))'} \leq 1 \cdot 1 = 1 \quad (\text{C.64})$$

ist und folglich, dass das Supremum in der vorgenommenen Weise weiter abgeschätzt werden darf.) $\square$

Bemerkung C.5.2. [28, Theorem 2.5.5] zeigt eine Verallgemeinerung von Satz C.5.1, denn er beweist ihn einerseits für die Erweiterung der sog. **Hilbert-Schmidt-Operatoren** zwischen allgemeinen Hilberträumen, zu denen die Hilbert-Schmidt Integraloperatoren auf $H = L^2(\Omega, \mu)$ ein Spezialfall sind. Des Weiteren zeigt er ebenfalls die Rückrichtung, d. h. die Hilbert-Schmidt Operatoren sind gerade alle zwischen Hilberträumen existierende **absolutsummierenden Operatoren** und umgekehrt.

Korollar C.5.3 (Eigenwertsumme von selbstadjungierten Hilbert-Schmidt Integraloperatoren [28, Theorem 2.5.5]).

Sei (Ω, μ) ein Maßraum und $T_k : H \rightarrow H$ ein selbstadjungierter Hilbert-Schmidt Integraloperator darauf. Dann gilt für dessen Eigenwerte $(\lambda_i)_{i \in I}$

$$\sum_{i \in I} |\lambda_i|^2 < \infty, \quad (\text{C.65})$$

wobei die Indexmenge I abzählbar ist. Ist $I \cong \mathbb{N}$ unendlich, so gilt also $(\lambda_i)_{i \in \mathbb{N}} \in \ell^2(\mathbb{N})$.

Beweis. (Der Beweis ist ein Teil des Beweises zu [28, Theorem 2.5.5].)

Die Voraussetzungen sind identisch zu denen von Prop. C.4.1. In dessen Beweis wurde die Aussage bereits mittels der Gleichungen (C.45)-(C.48) nachgewiesen. $\square$

Index

Abstraktes Cauchy-Problem, **54**, **55**

Algebra, **28**, **81**

Banach-, **25–27**, **81**

Algorithmus, **3**, **4**, **8**, **21–23**, **41**, **42**

evolutionärer, **44**

Aronszajn, **2**

Dirichlet

-Form, **54**

-Intervall, **40**

Eigen

-raum, **28–30**, **34**, **80**

-vektor/-funktion, **28–30**, **33**, **34**, **36**,
38–43, **45**, **48**, **49**, **52**, **53**, **55**, **56**, **58**,
80, **102**

-wert, **27–30**, **33**, **34**, **36**, **39–41**, **43**, **45**,
48–52, **55–58**, **80**, **102**, **104**

Folge

Cauchy-, **27**, **28**, **37**, **38**, **50**, **59**, **69**

Funktion(en)

integrierbare, **46**, **85**, **102**

messbare, **46**, **85**, **102**

positiv semidefinite, **14**, **15**, **17**, **20**, **25**, **45**,
47, **50**, **57**

Rademachersche, **96**, **98–100**

stetige, **27**, **40**, **42**, **45–53**, **57**, **69**

Funktional, **31**, **77**, **100**

Auswertungs-, **5**, **6**, **7**, **15**, **19**, **31**, **59**

Graph

metrischer, **40**

Quanten-, **40**

Hülle

lineare, **6**, **9–11**, **15**, **16**, **44**, **71**

Halbgruppe

Operatoren-, **54**

symmetrische Markovsche, **54**, **55**

ultrakonstraktive, **54**, **55**

Hyperebene, **4**, **21**

Isometrie, **15**, **17**, **18**, **60**, **70**

Isomorphie, **10**, **48**, **70**

isometrische, **10**, **11**, **15**, **17**, **19**, **35**, **60**, **70**

Kern

Gaußsche, **9**, **43**

Glättungs-, **45**, **46**, **87**

Integral-, **1**, **33**, **40**, **42–44**, **48**, **52**, **55–57**,
81

Reproduzierender, **6**, **9**, **11**, **13–15**, **19–23**,
25, **41**, **45**, **47**, **48**, **57–59**

Kernel-Trick, **2**, **22**, **23**, **41–44**, **48**, **53**, **55**

Klassifikationsproblem/-aufgabe, **3**, **4**, **7**, **21**,
22

Konvergenz

absolute, **30**, **33**, **37–39**, **48**, **53**, **73**

gleichmäßige, **33**, **38–44**, **48**, **51–53**, **55**, **72**

unbedingte, **30**, **73**

Maß

Borel-, **47**, **52**, **53**, **55**

Lebesgue-, **45**, **53**

Wahrscheinlichkeits-, **42**

Matrix, **14**

Gramsche, **14**

positiv semidefinite, **14**

Menge

abgeschlossene, 11, 44, 52, **68**
 Abschluss, 10, **68**
 dichte, 9, 11, 15, 18, 32, 35, 44, **68**
 Innere, 46, 47, **68**
 kompakte, 27, 28, 45, 46, 48–53, 57, **68**
 messbare, 52, **83**
 Null-, 33, 38, 45, 52, 56, 57, **84**
 offene, 45, 47, 51, 52, **68**
 orthogonales Komplement, 9, 10, **76**
 separable, 10, 41, 43, 44, 60, **68**
 vollständige, 10, 11, 15, 19, 28, 35, 37, 38,
 50, 59, **69**

Mercer, 1, **40**

Mercer-Punkt-4-Zusatzbedingung, **33**, 38–40,
 42, 44, 49, 51–53, 56, 57

Metrik, 3, 8, 23, 42, **67**

induzierte, 11, **71**

Moore, **1**

Norm, 3, 6, 8, 11, 18, 22, 23, 28, 30, 35, 40, 42,
 59, **71**

induzierte, 11, **74**

Operator, 18, 27, 35, 45, 54, 55, **77**, 95, 100,
 102

p - q -summierender, 2, 25, **30**, 31, 32, 34

absolutsummierender, 30, **95**, 100–104

adjungierter, 32, 35, **79**, 103, 104

beschränkter, 5, 7, 19, 27–32, 34, 35, 54,
 59, **77**, 100–103

Einbettungs-, 2, 30, **31**, 32, 95, 100, 101

Hilbert-Schmidt (Integral-), 33, 34, 37, 39,
 43, 45, 48, 52, **81**, 95, 100–102, 104

kompakter, 37, **78**, 102

Laplace-, **40**

normaler, 27, **82**

positiver, 2, 25, **27**, 28–30, **33**, 34, 40, 45,
 48, 52, 54

selbstadjungierter, 27–30, 34, 35, 37, 48,
 52, 54, **79**, 82, 101, 102, 104

Orthonormal

-basis, 43, 44, 60, **76**

-system, 29, 30, 34, 44, 58, **76**, 98, 102

Polynom, **25**, 26–28

Principle Component Analysis (PCA), 23, 25,
41

Raum

$C(\Omega)$ -, 48–50, **72**

$L^p(\Omega, \mu)$ -, 30–40, 42–44, 46, 48–50, 52–55,
 58, **72**, 98, 101

$\ell^p(\mathbb{N})$ -, 10, 30, **33**, 41, 42, 56, 58–60, **72**,
 100–102, 104

Banach-, 30, 54, **71**

Dual-, **78**

euklidischer/unitärer Vektor-, 15–17, 19,
 59, **74**

Hilbert-, 3–8, 10, 11, 15, 17, 19, 21–23, 25,
 27–29, 41, 42, 48, 58–60, **75**, 101, 102,
 104

lokalkompakter, **55**

lokalkonvexer, **30**

Maß-, 2, 30–36, 40, 42, 43, 45, 52, 54, 55,
 57, 58, **84**, 101, 104

metrischer, 15, 47, 52, 53, **67**

normierter, 30, **71**, 95

Unter-, 10, 15, 17, 18, 25, 35, **70**

Vektor-, 2, 5, 6, 10, 29, 33, 35, 41, 49, 54,
 58, **70**

zweitabzählbarer Hausdorffraum, **55**

Reproducing Kernel Hilbert Space (RKHS),
 2–4, **5**, 6–11, 13–15, 17–23, 25, 41, 42,
 53, 56–60

Reproduzierende Eigenschaft, **9**, 17, 19, 60

Rohdatenmenge, **5**, 7, 23, 25, 41, 42, 48, 53–55

Satz

Dini, 40, 51, **69**

Fréchet-Riesz (Darstellungs-), 6, 7, 9, 11,
79

Fubini, 46, **86**

Fundamental- der Algebra, 26, **87**

Hahn-Banach, 31, **79**

Lebesgue, 33, **86**, 102

Luzin, 52, 53, **86**

Mercer, 2, 25, 30, 32, **33**, 39–44, 48,
 52–58, 60

- Moore-(Aronszajn), 2, 14, **15**, 20, 25,
42–44, 48, 53, 57, 60
- Riesz (Darstellungs-), 35, **79**
- Spektral- für kompakte selbstadjungierte
Operatoren, 34, 37, **80**, 102
- Stone-Weierstraß, 27, **73**
- Skalarprodukt, 6–8, 11, 15–19, 22, 36, 38,
41–44, 48, 59, **74**
- Spektrum, 25–27, **82**
- Ungleichung
 - Besselsche, 34, **76**, 102
 - Cauchy-Schwarzsche, 17, **74**
 - Chintschinsche, 95, **98**, 100
 - Höldersche, **72**, 73, 99
- Vervollständigung, 11, 15, 17, 19, **70**
- Wärmeleitungsgleichung, **40**, 55
- Winkel, 3, 8, **75**
- Wurzel, 2, 25, **27**, 28–30, 34, 35

