

FORTGESCHRITTENE OPTIMIERUNG UND INVERSE PROBLEME: REGULARISIERUNG INVERSER PROBLEME

Prof. Dr. Bastian von Harrach

Goethe-Universität Frankfurt am Main
Institut für Mathematik

Sommersemester 2021

<http://numerical.solutions>

Inhaltsverzeichnis

1	Einleitung	1
1.1	Beispiel: Risikoabschätzung einer binärer Option	2
1.2	Hintergrund: Ableiten als inverses Problem	3
1.3	Parameteridentifikationsprobleme	3
2	Lineare inverse Probleme	5
2.1	Einige Grundbegriffe	5
2.2	Unbeschränkte lineare Operatoren	9
2.3	Moore-Penrose-Inverse	12
2.4	Kompakte Operatoren	17
2.4.1	Satz von Arzela-Ascoli	17
2.4.2	Kompakte Operatoren	19
2.4.3	Kompaktheit von Integraloperatoren	20
2.4.4	Spektraltheorie kompakter Operatoren	22
3	Regularisierung linearer Probleme	33
3.1	Regularisierung	33
3.1.1	Motivation und Definition	33
3.1.2	Filter-basierte Regularisierung	34
3.1.3	Abgeschnittene SWZ und Tikhonov-Regularisierung	36
3.2	Schwache Konvergenz	38
3.2.1	Sätze von Baire, von der gleichmäßigen Beschränktheit und von Banach-Steinhaus	38

INHALTSVERZEICHNIS

3.2.2	Schwache Konvergenz	40
3.3	Regularisierungsverfahren	42
3.3.1	Parameterwahlstrategien	42
3.3.2	Das Bakushinskii-Veto	45
3.4	Diskrepanzprinzip für das Tikhonov-Verfahren	46

Kapitel 1

Einleitung

Hadamard (1865-1963) nannte ein Problem *wohlgestellt*, falls

1. eine Lösung existiert (*Existenz*),
2. die Lösung eindeutig ist (*Eindeutigkeit*),
3. die Lösung stetig von den Eingangsdaten abhängt (*Stabilität*).

Trifft eine dieser Eigenschaften nicht zu, so spricht man von einem *schlecht gestellten* Problem (engl.: *ill-posed*).

Seien X, Y Hilberträume und $A : X \rightarrow Y$ ein stetig und linear. Dann ist das *direkte Problem*, zu gegebenem $x \in X$ den Vektor $y = Ax$ zu berechnen, offenbar wohlgestellt. Für das dazugehörige *inverse Problem*, eine lineare Gleichung $Ax = y$ zu lösen, bedeuten die Hadamard-Kriterien:

- Existenz: $y \in \mathcal{R}(A)$ bzw. A surjektiv.
- Eindeutigkeit: A injektiv.
- Stabilität: A^{-1} stetig.

In dieser Vorlesung untersuchen wir schlecht gestellte Probleme, bei denen eine oder mehrere dieser drei Eigenschaften nicht erfüllt sind. Wir werden zeigen, in welchem Sinne wir bei Verletzung der ersten beiden Punkte dennoch von einer eindeutigen (und für Anwendungen relevanten) Lösung sprechen können und wie Probleme mit unstetigem A^{-1} dennoch stabil gelöst werden können.

1.1 Ein Beispiel aus der Finanzmathematik: Risikoabschätzung einer binären Option

Wir betrachten eine binäre Option (auch: digitale Option oder Cash-or-Nothing-Option). Der Käufer einer solchen Option erhält zu einem späteren Zeitpunkt (am Verfallszeitpunkt/maturity date) einen festgelegten Betrag (payoff), falls dann der Kurs einer bestimmte Aktie (Basiswert, underlying) über einem gewissen Wert (strike) liegt. Ansonsten erhält er nichts.

Die linke Seite von Abbildung 1.1 zeigt den fairen Preis V einer solchen Option in Abhängigkeit vom Kurs des Underlyings S_0 . Die rote durchgezogene Kurve zeigt dabei analytisch berechnete Werte. (Für solch einfache Beispiele existieren geschlossene Lösungsformeln.) Schwarz gepunktet sind (mit dem sogenannten Monte Carlo-Verfahren) numerisch genäherte Werte eingezeichnet. Beide Kurven stimmen sehr gut überein.

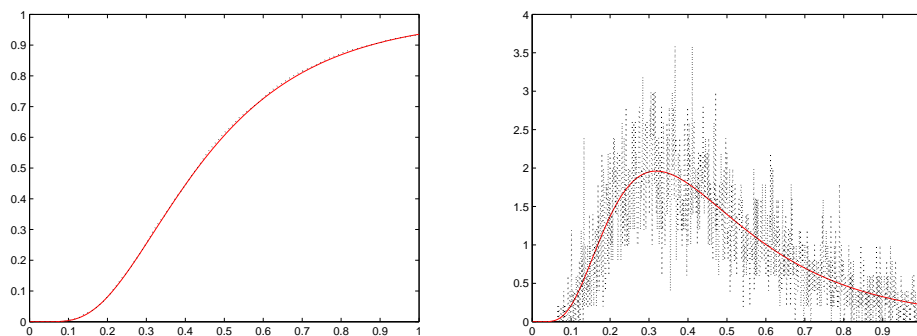


Abbildung 1.1: Analytisch und mit Monte Carlo berechneter Optionspreis (linkes Bild) und daraus jeweils mit finiten Differenzen berechnetes Delta (rechtes Bild).

Zur Risikobewertung einer binären Option ist es wichtig zu wissen, wie stark der Preis V vom Kurs des Underlyings S_0 abhängt, ob etwa schon minimale Kurschwankungen große Wertschwankungen der Option auslösen können. Hierfür relevant ist das sogenannte *Delta*

$$\Delta = \frac{\partial V}{\partial S_0}.$$

Die rechte Seite in Abbildung 1.1 zeigt das mittels finiter Differenzen aus den Kurven auf der linken Seite berechnete Delta.

$$\Delta(S_0) \approx \frac{V(S_0 + h) - V(S_0)}{h}, \quad h = 10^{-3}$$

Während die numerisch berechneten Optionspreise noch sehr gut mit den korrekten (analytisch berechneten) Werten übereinstimmt, sind die daraus numerisch berechneten Ableitungen offenbar völlig unbrauchbar. Der Grund dafür ist, dass durch die Division durch h die in der numerischen Bewertung vorhandenen Fehler um das $1/h = 1000$ -fache verstärkt werden und so das Ergebnis ruinieren.

1.2 Hintergrund: Ableiten als inverses Problem

Der Grund für die beobachtete Instabilität ist, dass die Ableitung einer Funktion nicht stetig von den Funktionswerten abhängt. Ein einfaches Beispiel ist

$$f_n(x) = \frac{1}{n\pi} \sin(n\pi x), \quad f'_n(x) = \cos(n\pi x)$$

Offenbar konvergiert $f_n(x)$ glm. auf $[0, 1]$ gegen die Nullfunktion. Man kann jedoch zeigen, dass $f'_n(x)$ für keinen Punkt $x \in (0, 1]$ konvergiert.

Wir können die Ableitung auch als inverses Problem zu einer linearen Abbildung zwischen Hilberträumen auffassen. Dazu definieren wir

$$A : L^2(0, 1) \rightarrow L^2(0, 1), \quad Af := \int_0^x f(s) ds$$

Man rechnet leicht nach, dass A tatsächlich eine Abbildung von $L^2(0, 1)$ nach $L^2(0, 1)$, sowie linear und stetig ist. Außerdem ist $Af - g = \text{const.}$ genau dann, wenn $f = g'$ (im Sinne der *schwachen* bzw. *distributionellen* Ableitung).

Offenbar ist auch

$$Af'_n = f_n, \quad \|f_n\|_{L^2(0,1)}^2 = \frac{1}{2n^2\pi^2} \quad \text{und} \quad \|f'_n\|_{L^2(0,1)}^2 = \frac{1}{2},$$

d.h. $\|Af'_n\| = \frac{1}{n\pi} \|f'_n\|$. A kann also nicht stetig invertierbar sein. Das Ableiten einer Funktion ist ein *schlecht gestelltes, inverses Problem*.

1.3 Parameteridentifikationsprobleme

Viele Vorgänge in den Natur- und Wirtschaftswissenschaften lassen sich durch partielle Differentialgleichungen beschreiben, deren Lösung die Vorhersage des Verhaltens eines Systems bei vollständiger Kenntnis aller dazu nötigen Parameter ermöglicht. So beschreibt z.B.

$$\begin{aligned} -\nabla \cdot (a(x)\nabla u(x)) &= f(x) & \forall x \in \Omega \\ u(x) &= 0 & \forall x \in \partial\Omega \end{aligned}$$

KAPITEL 1. EINLEITUNG

die stationäre Wärmeverteilung in einem Körper Ω dessen Rand auf Nulltemperatur gehalten wird. $u(x)$ ist dabei die Temperatur im Punkt x . $a(x)$ ist die Wärmeleitfähigkeit und $f(x)$ ist eine von außen angelegte Wärmequelle.

Das dazugehörige inverse Problem ist es, die Wärmeleitfähigkeit des Körpers durch Messungen der Temperatur $u(x)$ zu bestimmen. Wir betrachten ein einfaches eindimensionales Beispiel:

$$-(a(x)u_x(x))_x = f(x) \quad x \in (0, 1)$$

mit $u(0) = 0 = u(1)$. Falls u_x nirgendwo verschwindet, dann erhalten wir

$$a(x) = -\frac{1}{u_x(x)} \left(a(0)u_x(0) + \int_0^x f(s) ds \right)$$

Um a zu berechnen, müssen die Temperaturmessungen $u(x)$ also differenziert werden, was wie im letzten Abschnitt erklärt ein schlecht-gestelltes Problem darstellt. Zusätzliche (nicht-lineare) Instabilitäten können durch die Division durch den möglicherweise kleinen Ausdruck $u_x(x)$ auftreten.

Kapitel 2

Lineare inverse Probleme

Der Aufbau dieses Kapitel folgt an vielen Stellen dem sehr empfehlenswerten Lehrbuch [Rieder].

2.1 Einige Grundbegriffe

Es seien stets X, Y, Z (reelle) Hilberträume mit Skalarprodukten $(\cdot, \cdot)_X$, $(\cdot, \cdot)_Y$ und dadurch induzierten Normen $\|\cdot\|_X$, $\|\cdot\|_Y$.

Definition und Satz 2.1

Sei $A : X \rightarrow Y$ linear. Dann sind äquivalent:

- (a) A ist stetig.
- (b) A ist beschränkt, d.h. es existiert $C > 0$ mit

$$\|Ax\|_Y \leq C \|x\|_X \quad \text{für alle } x \in X.$$

Sei $b : X \times Y \rightarrow \mathbb{R}$ bilinear. Dann sind äquivalent:

- (a) b ist stetig.
- (b) b ist beschränkt, d.h. es existiert $C > 0$ mit

$$|b(x, y)| \leq C \|x\|_X \|y\|_Y \quad \text{für alle } x \in X, y \in Y.$$

Beweis: Wir setzen das als bekannt voraus und verweisen z.B. auf [Alt]. □

KAPITEL 2. LINEARE INVERSE PROBLEME

Definition und Satz 2.2

Den Vektorraum aller stetigen linearen Abbildungen von X nach Y bezeichnen wir mit $\mathcal{L}(X, Y)$. $\mathcal{L}(X, Y)$ ist ein Banachraum mit der Operatornorm

$$\|A\|_{\mathcal{L}(X, Y)} := \sup_{x \neq 0} \frac{\|Ax\|_Y}{\|x\|_X} = \sup_{\|x\|_X=1} \|Ax\|_Y.$$

Wir schreiben auch abkürzend $\mathcal{L}(X) := \mathcal{L}(X, X)$ und $X' := \mathcal{L}(X, \mathbb{R})$. X' heißt auch der Dualraum von X .

Beweis: Auch das setzen wir als bekannt voraus und verweisen auf [Alt]. □

Satz 2.3 (Lax-Milgram)

Es sei X ein Hilbertraum und

$$b : X \times X \rightarrow \mathbb{R},$$

eine stetige, symmetrische, koerzive Bilinearform, d.h.

$$\begin{aligned} b(u, v) &= b(v, u) & \forall u, v \in X & \quad (\text{Symmetrie}), \\ \exists C > 0 : |b(u, v)| &\leq C \|u\| \|v\| & \forall u, v \in X & \quad (\text{Stetigkeit}), \\ \exists \beta > 0 : b(u, u) &\geq \beta \|u\|^2 & \forall u \in X & \quad (\text{Koerzivität}) \end{aligned}$$

und b ist in beiden Komponenten linear. Weiterhin sei $l \in X'$.

Dann existiert genau ein $u \in X$ mit

$$b(u, v) = l(v) \quad \text{für alle } v \in X. \tag{2.1}$$

u hängt stetig und linear von l ab,

$$\|u\|_X \leq \frac{1}{\beta} \|l\|_{X'}.$$

und u ist das eindeutige Minimum von

$$J : X \rightarrow \mathbb{R}, \quad J(v) := \frac{1}{2} b(v, v) - \langle l, v \rangle.$$

Beweis: Übungsaufgabe. □

Bemerkung 2.4

(a) Nimmt man als Bilinearform b das Skalarprodukt $(\cdot, \cdot)_X$ in einem Hilbertraum X , so sind offenbar alle Voraussetzungen von Satz 2.3 erfüllt. Es gibt also zu jeder Linearform $l \in X'$ ein $\xi \in X$, so dass

$$(\xi, x)_X = l(x) \quad \text{für alle } x \in X$$

2.1. EINIGE GRUNDBEGRIFFE

und $\iota : X' \rightarrow X$, $l \mapsto \xi$ ist eine stetige lineare Abbildung.

ι besitzt eine stetige lineare Inverse, denn zu jedem $\xi \in X$ ist durch

$$l : X \rightarrow \mathbb{R}, \quad l(x) := (\xi, x)_X \quad \text{für alle } x \in X$$

eine Linearform definiert mit $\iota(l) = \xi$ und

$$\|l\|_{X'} = \sup_{x \neq 0} \frac{|l(x)|}{\|x\|} = \sup_{x \neq 0} \frac{(\xi, x)_X}{\|x\|} = \|\xi\|_X.$$

X ist daher isometrisch isomorph zu X' . Dies bezeichnet man auch als Riesz-schen Darstellungssatz.

(b) Nach Definition und Satz 2.2 ist X' ein Banachraum. Durch das Skalarprodukt

$$(l_1, l_2)_{X'} := (\iota(l_1), \iota(l_2))_X$$

wird X' sogar zum Hilbertraum.

(c) Sei $A \in \mathcal{L}(X, Y)$. Für jedes $y \in Y$ definiert $l : x \mapsto (Ax, y)_Y$ eine stetige Linearform $l \in X'$ und daher existiert

$$\xi \in X \quad \text{mit} \quad (x, \xi)_X = (Ax, y)_Y.$$

Außerdem hängt in dieser Konstruktion $\xi \in X$ stetig und linear von $l \in X'$ und dieses stetig und linear von $y \in Y$ ab.

Zu jedem $A \in \mathcal{L}(X, Y)$ kann daher der adjungierte Operator $A^* \in \mathcal{L}(Y, X)$ definiert werden durch

$$(x, A^*y)_X := (Ax, y)_Y \quad \forall x \in X, y \in Y.$$

Man zeigt leicht, dass $(A^*)^* = A$.

Satz 2.5

Ist $V \subseteq X$ ein abgeschlossener Unterraum, dann ist V mit dem Skalarprodukt aus X selbst ein Hilbertraum.

Beweis: Offenbar ist $(\cdot, \cdot)_X$ auch ein Skalarprodukt auf V und die Vollständigkeit folgt direkt aus der Abgeschlossenheit und der Vollständigkeit von X . \square

Definition und Satz 2.6

Sei $V \subseteq X$ ein (nicht notwendigerweise abgeschlossener) Unterraum.

KAPITEL 2. LINEARE INVERSE PROBLEME

Wir definieren $P_{\bar{V}} \in \mathcal{L}(X)$ durch $P_{\bar{V}}x := v$ für $x \in X$, wobei

$$v := \arg \min_{\tilde{v} \in \bar{V}} \|x - \tilde{v}\|_X.$$

$P_{\bar{V}}$ heißt orthogonale Projektion auf V .

Außerdem definieren wir das orthogonale Komplement durch

$$V^\perp := \{x \in X : (x, v) = 0\} := \{x \in X : (x, v) = 0 \quad \forall v \in V\}.$$

Dann gilt

(a)

$$v = P_{\bar{V}}x \iff v \in \bar{V} \quad \text{und} \quad v - x \in \bar{V}^\perp,$$

insbesondere ist also $\mathcal{R}(P_{\bar{V}}) = \bar{V}$ und $\mathcal{N}(P_{\bar{V}}) = \bar{V}^\perp$.

(b) $V^\perp = \bar{V}^\perp$.

(c) $X = \bar{V} \oplus V^\perp$.

(d) $(V^\perp)^\perp = \bar{V}$.

(e) Für $A \in \mathcal{L}(X, Y)$ ist $\mathcal{N}(A^*) = \mathcal{R}(A)^\perp$.

(f) Für $A \in \mathcal{L}(X, Y)$ ist $\overline{\mathcal{R}(A^*)} = \mathcal{N}(A)^\perp$.

Beweis: Es ist

$$\arg \min_{\tilde{v} \in \bar{V}} \|x - \tilde{v}\|_X = \arg \min_{\tilde{v} \in \bar{V}} \|x - \tilde{v}\|_X^2 = \arg \min_{\tilde{v} \in \bar{V}} \left(\frac{1}{2} (\tilde{v}, \tilde{v})_X - (x, \tilde{v})_X \right)$$

Da \bar{V} ein Hilbertraum ist (Satz 2.5), existiert nach dem Satz von Lax-Milgram (Satz 2.3) genau ein Minimum von $\frac{1}{2} (\tilde{v}, \tilde{v})_X - (x, \tilde{v})_X$, dieses ist die eindeutige Lösung von

$$(v, w) = (x, w) \quad \forall w \in \bar{V} \iff v - x \in \bar{V}^\perp$$

und v hängt stetig und linear von x ab. Damit ist gezeigt, dass $P_{\bar{V}} \in \mathcal{L}(X)$ wohldefiniert ist und (a) gilt. Offenbar ist auch $P_{\bar{V}}^2 = P_{\bar{V}}$, also $P_{\bar{V}}$ tatsächlich eine Projektion.

(b) folgt durch stetige Fortsetzung.

(c) $\bar{V} \cap V^\perp = 0$ ist klar und jedes $x \in X$ lässt sich (stetig) zerlegen in

$$x = P_{\bar{V}}x + (x - P_{\bar{V}}x) \in \bar{V} + V^\perp.$$

(d) klar: $\bar{V} \subseteq (\bar{V}^\perp)^\perp = (V^\perp)^\perp$

Ist umgekehrt $x \in (V^\perp)^\perp$, dann ist nach (a)

$$P_{\bar{V}}x - x \in V^\perp, \quad \text{aber auch} \quad P_{\bar{V}}x - x \in (V^\perp)^\perp,$$

also $(P_{\bar{V}}x - x, P_{\bar{V}}x - x) = 0$ und damit $x = P_{\bar{V}}x \in \bar{V}$.

(e)

$$\begin{aligned} \mathcal{R}(A)^\perp &= \{\eta \in Y : (\eta, y) = 0 \forall y \in \mathcal{R}(A)\} \\ &= \{\eta \in Y : 0 = (\eta, Ax) = (A^*\eta, x) \forall x \in X\} \\ &= \{\eta \in Y : A^*\eta = 0\} = \mathcal{N}(A^*). \end{aligned}$$

(f) Aus $(A^*)^* = A$, (e) und (d) folgt

$$\mathcal{N}(A)^\perp = \mathcal{N}((A^*)^*)^\perp = (\mathcal{R}(A^*)^\perp)^\perp = \overline{\mathcal{R}(A^*)}.$$

2.2 Unbeschränkte lineare Operatoren

Es seien weiterhin X, Y und Z reelle Hilberträume.

Definition 2.7

Sei $V \subseteq X$ ein Unterraum von X . Eine lineare Abbildung $A : V \rightarrow Y$ bezeichnen wir auch als Operator von X nach Y mit Definitionsbereich $\mathcal{D}(A) = V$ und schreiben

$$A : \mathcal{D}(A) \subseteq X \rightarrow Y.$$

Wenn wir die Schreibweise von Definition 2.7 verwenden, wird meistens $\mathcal{D}(A)$ dicht in X liegen und A unbeschränkt sein.

Definition und Satz 2.8

Seien A, B und C (möglicherweise unbeschränkte) lineare Operatoren mit Definitionsbereichen $\mathcal{D}(A), \mathcal{D}(B)$ und $\mathcal{D}(C)$.

(a) Wir schreiben $A \subseteq B$ falls

$$\mathcal{D}(A) \subseteq \mathcal{D}(B) \quad \text{und} \quad Ax = Bx \quad \forall x \in \mathcal{D}(A).$$

Wir schreiben $A = B$ falls zusätzlich $\mathcal{D}(A) = \mathcal{D}(B)$.

KAPITEL 2. LINEARE INVERSE PROBLEME

(b) Wir definieren $A + B$ und AB punktweise auf den natürlichen Definitionsbereichen

$$\mathcal{D}(A + B) = \mathcal{D}(A) \cap \mathcal{D}(B), \quad \mathcal{D}(AB) = \{x \in \mathcal{D}(B) : Bx \in \mathcal{D}(A)\}.$$

Für $0 \neq \alpha \in \mathbb{R}$ definieren wir αA punktweise auf

$$\mathcal{D}(\alpha A) = \mathcal{D}(A).$$

$0A$ ist der (beschränkte und überall definierte) Nulloperator.

(c) Es gelten die folgenden Rechenregeln

$$\begin{aligned} (A + B) + C &= A + (B + C), & (A + B)C &= AC + BC, \\ (AB)C &= A(BC), & A(B + C) &\supseteq AB + AC. \end{aligned}$$

Beweis: Übungsaufgabe □

Definition und Satz 2.9

(a) Die adjungierte Abbildung eines linearen Operators

$$A : \mathcal{D}(A) \subseteq X \rightarrow Y$$

mit dichtem Definitionsbereich $\mathcal{D}(A)$ ist der durch

$$(Ax, y) = (x, A^*y) \quad \forall x \in \mathcal{D}(A), y \in \mathcal{D}(A^*) \quad (2.2)$$

eindeutig definierte lineare Operator $A^* : \mathcal{D}(A^*) \subseteq Y \rightarrow X$ mit Definitionsbereich

$$\mathcal{D}(A^*) := \{y \in Y : x \mapsto (Ax, y) \text{ ist stetig auf } \mathcal{D}(A)\}$$

(b) Sind $A : \mathcal{D}(A) \subseteq Y \rightarrow Z$, $B : \mathcal{D}(B) \subseteq X \rightarrow Y$ und AB lineare Operatoren mit dichten Definitionsbereichen, dann ist

$$(AB)^* \supseteq B^*A^*.$$

Ist $A \in \mathcal{L}(Y, Z)$, dann gilt sogar $(AB)^* = B^*A^*$.

(c) Ein linearer Operator $A : \mathcal{D}(A) \subseteq X \rightarrow X$ heißt symmetrisch, falls

$$(Ax, y) = (x, Ay) \quad \forall x, y \in \mathcal{D}(A).$$

Ist $\mathcal{D}(A)$ dicht in X , so gilt

$$A \text{ symmetrisch} \iff A \subseteq A^*,$$

und A heißt selbstadjungiert, falls $A = A^*$, also zusätzlich $\mathcal{D}(A) = \mathcal{D}(A^*)$ gilt.

2.2. UNBESCHRÄNKTE LINEARE OPERATOREN

Beweis: (a) Offenbar ist $\mathcal{D}(A^*)$ ein Unterraum von Y . Für $y \in \mathcal{D}(A^*)$ kann die Abbildung $x \mapsto (Ax, y)$ zu einer stetigen linearen Abbildung $X \rightarrow \mathbb{R}$ fortgesetzt werden, nach dem Satz von Riesz (Bemerkung 2.4) existiert also genau eine Lösung A^*y der definierenden Gleichung (2.2). Die Linearität von A^* ist klar.

(b) Sei $y \in \mathcal{D}(B^*A^*)$. Es ist zu zeigen, dass

$$y \in \mathcal{D}((AB)^*) \quad \text{und} \quad B^*A^*y = (AB)^*y.$$

Ersteres gilt, falls $x \mapsto (ABx, y)$ stetig ist auf $\mathcal{D}(AB)$.

Um dies zu zeigen, nutzen wir aus, dass für $x \in \mathcal{D}(AB)$ gilt

$$(ABx, y) = (Bx, A^*y) = (x, B^*A^*y),$$

wobei die erste Gleichheit aus $Bx \in \mathcal{D}(A)$ und $y \in \mathcal{D}(A^*)$ und die zweite Gleichheit aus $x \in \mathcal{D}(B)$ und $A^*y \in \mathcal{D}(B^*)$ folgt. Da $x \mapsto (x, B^*A^*y)$ stetig ist, ist also auch

$$x \mapsto (ABx, y) = (x, B^*A^*y) \text{ stetig auf } \mathcal{D}(AB)$$

und damit ist $y \in \mathcal{D}((AB)^*)$ gezeigt. Zusammen mit $x \in \mathcal{D}(AB)$ folgt damit dann aber auch

$$(x, B^*A^*y) = (ABx, y) = (x, (AB)^*y) \quad \forall x \in \mathcal{D}(AB), y \in \mathcal{D}(B^*A^*).$$

Da die linke und rechte Seite stetig in x sind und $\mathcal{D}(AB) \subseteq X$ dicht liegt, folgt daraus, dass

$$(x, B^*A^*y) = (x, (AB)^*y) \quad \forall x \in X, y \in \mathcal{D}(B^*A^*)$$

und somit $B^*A^*y = (AB)^*y$. Damit ist $B^*A^* \subseteq (AB)^*$ gezeigt.

Nun sei A beschränkt. Um $(AB)^* = B^*A^*$ zu beweisen, müssen wir nur noch zeigen, dass $\mathcal{D}((AB)^*) \subseteq \mathcal{D}(B^*A^*)$. Sei also $y \in \mathcal{D}((AB)^*)$. Dann ist für alle $x \in \mathcal{D}(AB)$

$$(x, (AB)^*y) = (ABx, y) = (Bx, A^*y).$$

wobei wir in der zweiten Umformung $A \in \mathcal{L}(X, Y)$ verwendet haben. Es ist also $x \mapsto (Bx, A^*y)$ stetig in $\mathcal{D}(AB) = \mathcal{D}(B)$ und damit $A^*y \in \mathcal{D}(B^*)$. Also ist $y \in \mathcal{D}(B^*A^*)$, womit die Behauptung folgt.

(c) Sei A symmetrisch mit dichtem $\mathcal{D}(A)$. Für alle $y \in \mathcal{D}(A)$ ist

$$x \mapsto (Ax, y) = (x, Ay)$$

stetig, also $y \in \mathcal{D}(A^*)$ und $Ay = A^*y$ für alle $y \in \mathcal{D}(A)$, d.h. $A \subseteq A^*$. Die Rückrichtung ist trivial. \square

Bemerkung 2.10

Achtung! Alle von A gebildeten Ausdrücke hängen automatisch auch von $\mathcal{D}(A)$ ab, insbesondere auch A^ und $\mathcal{D}(A^*)$!*

Es gibt oft mehrere natürliche Möglichkeiten $\mathcal{D}(A)$ zu wählen, der Definitionsbereich muss nicht der „maximal mögliche“ sein. Umgekehrt ist $\mathcal{D}(A^)$ entsprechend Def. 2.9 eindeutig durch A und $\mathcal{D}(A)$ festgelegt und muss ebenfalls nicht mit dem „maximal möglichen“ Definitionsbereich übereinstimmen.*

Beispiel 2.11

Wir definieren die (eindimensionalen) Sobolevräume

$$\begin{aligned} H^1(0,1) &= \{u \in L^2(0,1) : u' \in L^2(0,1)\} \\ H_0^1(0,1) &= \{u \in H^1(0,1) : u(0) = u(1) = 0\} \\ H_\pi^1(0,1) &= \{u \in H^1(0,1) : u(0) = u(1)\} \end{aligned}$$

Auf allen drei Räumen können wir den unbeschränkten Ableitungsoperator definieren

$$\begin{aligned} A_1 : D(A_1) \subset L^2(0,1) &\rightarrow L^2(0,1), & A_1 f &= f', & D(A_1) &= H^1(0,1), \\ A_2 : D(A_2) \subset L^2(0,1) &\rightarrow L^2(0,1), & A_2 f &= f', & D(A_2) &= H_0^1(0,1), \\ A_3 : D(A_3) \subset L^2(0,1) &\rightarrow L^2(0,1), & A_3 f &= f', & D(A_3) &= H_\pi^1(0,1). \end{aligned}$$

Dann kann man zeigen, dass (siehe z.B. [Rudin, Ex. 13.4])

$$A_1^* = -A_2, \quad A_2^* = -A_1, \quad A_3^* = -A_3.$$

2.3 Moore-Penrose-Inverse

In diesem Abschnitt ist immer $A \in \mathcal{L}(X, Y)$ ein linearer beschränkter Operator zwischen Hilberträumen X und Y . Wir interessieren uns für die Lösung der linearen Gleichung

$$y = Ax$$

für möglicherweise nicht injektives und/oder nicht surjektives A .

Satz 2.12

Sei $y \in Y$. Äquivalent sind

(a) $x \in X$ löst $Ax = P_{\overline{\mathcal{R}(A)}} y$

(b) $x \in X$ minimiert das Residuum

$$\|Ax - y\|_Y \leq \|A\xi - y\|_Y \quad \forall \xi \in X.$$

(c) $x \in X$ löst die Normalgleichungen

$$A^*Ax = A^*y. \tag{2.3}$$

Beweis: (a) \iff (b) : Nach Definition und Satz 2.6 gilt

$$Ax = P_{\mathcal{R}(A)}y \iff \|Ax - y\| \leq \|\eta - y\| \quad \forall \eta \in \overline{\mathcal{R}(A)}.$$

und die rechte Aussage ist mit stetiger Fortsetzung äquivalent zu (b).

(a) \iff (c) : Nach Definition und Satz 2.6(a),(e) ist

$$Ax = P_{\mathcal{R}(A)}y \iff Ax - y \in \mathcal{R}(A)^\perp = \mathcal{N}(A^*) \iff A^*Ax = A^*y,$$

also gilt (a) \iff (c). □

Satz 2.13

(a) Zu $y \in Y$ existiert genau dann eine Lösung der Normalgleichungen (2.3), wenn $y \in \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$.

(b) Zu $y \in \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$ existiert genau eine Lösung x^+ der Normalgleichungen mit minimaler Norm,

$$A^*Ax^+ = A^*y \quad \text{und} \quad \|x^+\| \leq \|x\| \quad \forall x \in (A^*A)^{-1}A^*y.$$

Beweis:

(a) Löst x die Normalgleichungen, so ist $Ax - y \in \mathcal{R}(A)^\perp$, also

$$y = Ax + y - Ax \in \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp.$$

Ist umgekehrt $y \in \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$, also $y = Ax + \eta$ mit $\eta \in \mathcal{R}(A)^\perp = \mathcal{N}(A^*)$ dann folgt $A^*y = A^*Ax$.

(b) Die Menge $U := (A^*A)^{-1}A^*y$ ist das Urbild der abgeschlossenen einelementigen Menge $\{A^*y\}$ unter der stetigen Abbildung A^*A . U ist also abgeschlossen und damit vollständig.

Für die eindeutige Existenz eines Elementes $x^+ \in U$ mit minimaler Norm zeigen wir, dass jede minimierende Folge $x_n \in U$, $\lim_{n \rightarrow \infty} \|x_n\| = \inf_{x \in U} \|x\|$ eine

KAPITEL 2. LINEARE INVERSE PROBLEME

Cauchy-Folge ist: U ist offenbar konvex, insbesondere ist für alle $n, m \in \mathbb{N}$ auch $\frac{1}{2}x_n + \frac{1}{2}x_m \in U$ und damit

$$\begin{aligned} \inf_{x \in U} \|x\| &\leq \left\| \frac{1}{2}x_n + \frac{1}{2}x_m \right\| \\ &\leq \frac{1}{2}\|x_n\| + \frac{1}{2}\|x_m\| \rightarrow \inf_{x \in U} \|x\| \quad \text{für } n, m \rightarrow \infty. \end{aligned}$$

Es folgt, dass

$$\begin{aligned} \inf_{x \in U} \|x\|^2 &= \lim_{n, m \rightarrow \infty} \left\| \frac{1}{2}x_n + \frac{1}{2}x_m \right\|^2 \\ &= \lim_{n, m \rightarrow \infty} \left(\frac{1}{4}\|x_n\|^2 + \frac{1}{2}(x_n, x_m) + \frac{1}{4}\|x_m\|^2 \right), \end{aligned}$$

also $\lim_{n, m \rightarrow \infty} (x_n, x_m) = \inf_{x \in U} \|x\|^2$ und damit

$$\lim_{n, m \rightarrow \infty} \|x_n - x_m\|^2 = 0.$$

Da U vollständig ist, existiert $x^+ := \lim_{n \rightarrow \infty} x_n \in U$.

$$\|x^+\| = \lim_{n \rightarrow \infty} \|x_n\| = \inf_{x \in U} \|x\|,$$

also ist x^+ eine Lösung der Normalgleichungen mit minimaler Norm.

Schließlich kann es keine weitere von x^+ verschiedene Lösung \hat{x} der Normalgleichungen mit minimaler Norm geben, da sonst $x^+, \hat{x}, x^+, \hat{x}, \dots$ eine minimierende Folge, jedoch keine Cauchy-Folge wäre. \square

Definition 2.14

Zu $A \in \mathcal{L}(X, Y)$ definieren wir die Abbildung

$$A^+ : \mathcal{D}(A^+) \subset Y \rightarrow X, \quad y \mapsto x^+,$$

mit dem Definitionsbereich $\mathcal{D}(A^+) := \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$, durch die entsprechend Satz 2.13b) eindeutig bestimmte Minimum-Norm-Lösung x^+ der Normalgleichungen.

A^+ heißt verallgemeinerte oder Moore-Penrose-Inverse von A .

Satz 2.15

(a) $\mathcal{D}(A^+) = Y$, genau dann wenn $\mathcal{R}(A)$ abgeschlossen ist.

(b) $\mathcal{R}(A^+) = \mathcal{N}(A)^\perp$.

(c) Löst x die Normalgleichungen $A^*Ax = A^*y$, so ist $x^+ = P_{\mathcal{N}(A)^\perp}x$.

(d) A^+ ist linear.

Beweis: (a) Gilt $\mathcal{R}(A) = \overline{\mathcal{R}(A)}$, so ist $Y = \overline{\mathcal{R}(A)} \oplus \mathcal{R}(A)^\perp = \mathcal{D}(A^+)$.

Ist umgekehrt $Y = \overline{\mathcal{R}(A)} \oplus \mathcal{R}(A)^\perp = \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$, dann folgt

$$\overline{\mathcal{R}(A)} = P_{\overline{\mathcal{R}(A)}}(\overline{\mathcal{R}(A)} \oplus \mathcal{R}(A)^\perp) = P_{\overline{\mathcal{R}(A)}}(\mathcal{R}(A) \oplus \mathcal{R}(A)^\perp) = \mathcal{R}(A).$$

(b), „ \subseteq “ Wir zeigen zuerst $\mathcal{R}(A^+) \subseteq \mathcal{N}(A)^\perp$. Sei dazu $x^+ = A^+y \in \mathcal{R}(A^+)$ mit $y \in \mathcal{D}(A^+)$. Angenommen $x^+ \notin \mathcal{N}(A)^\perp$, dann würde ein $x \in \mathcal{N}(A)$ existieren mit $(x^+, x) \neq 0$. Für alle $\lambda \in \mathbb{R}$ löst dann aber auch $x^+ + \lambda x$ die Normalgleichungen, und wegen

$$\|x^+ + \lambda x\|^2 = \|x^+\|^2 + 2\lambda(x^+, x) + \lambda^2 \|x\|^2$$

können wir λ so wählen, dass $\|x^+ + \lambda x\| < \|x^+\|$ gilt (z.B. mit der Wahl $\lambda := -(x^+, x)/\|x\|^2$). Dies widerspricht aber der Minimalnormeigenschaft von x^+ .

(b), „ \supseteq “+(c) Aus $A^*Ax = A^*y$ folgt $A^*A(x - x^+) = 0$ und damit

$$(A(x - x^+), A(x - x^+)) = (A^*A(x - x^+), (x - x^+)) = 0.$$

Es ist also $x - x^+ \in \mathcal{N}(A) = (\mathcal{N}(A)^\perp)^\perp$. Wegen (b), „ \subseteq “ gilt außerdem auch $x^+ \in \mathcal{N}(A)^\perp$, so dass mit Definition und Satz 2.6 folgt, dass $x^+ = P_{\mathcal{N}(A)^\perp}x$. Damit ist (c) gezeigt.

Da jedes $x \in \mathcal{N}(A)^\perp$ selbst Lösung der Normalgleichungen zu $y := Ax$ ist und $x = P_{\mathcal{N}(A)^\perp}x$ erfüllt, folgt damit auch $\mathcal{N}(A)^\perp \subseteq \mathcal{R}(A^+)$.

(d) Sind $y_1, y_2 \in \mathcal{D}(A^+)$ und $\lambda \in \mathbb{R}$, dann löst $A^+y_1 + \lambda A^+y_2$ offenbar die Normalgleichungen zu $y_1 + \lambda y_2$. Mit (c) und (b) folgt

$$A^+(y_1 + \lambda y_2) = P_{\mathcal{N}(A)^\perp}(A^+y_1 + \lambda A^+y_2) = A^+y_1 + \lambda A^+y_2.$$

Satz 2.16

Die Moore-Penrose-Inverse A^+ ist die einzige Abbildung

$$\mathcal{R}(A) \oplus \mathcal{R}(A)^\perp \subseteq Y \rightarrow X,$$

die die folgenden vier Moore-Penrose-Axiome erfüllt

$$\begin{aligned} AA^+A &= A & \text{auf } X, & & A^+A &= P_{\overline{\mathcal{R}(A^*)}} & \text{auf } X, \\ A^+AA^+ &= A^+ & \text{auf } \mathcal{D}(A^+), & & AA^+ &= P_{\overline{\mathcal{R}(A)}}, & \text{auf } \mathcal{D}(A^+). \end{aligned}$$

KAPITEL 2. LINEARE INVERSE PROBLEME

Beweis: Da A^+y die Normalgleichungen löst folgt mit Satz 2.12(a)

$$AA^+y = P_{\overline{\mathcal{R}(A)}}y \quad \forall y \in \mathcal{D}(A^+)$$

und damit auch

$$AA^+Ax = P_{\overline{\mathcal{R}(A)}}Ax = Ax \quad \forall x \in X.$$

Jedes $x \in X$ löst die Normalgleichungen zu $y := Ax \in \mathcal{D}(A^+)$, also folgt aus Satz 2.15c)

$$A^+Ax = x^+ = P_{\overline{\mathcal{R}(A^*)}}x$$

und damit wegen $\mathcal{R}(A^+) = \mathcal{N}(A)^\perp = \overline{\mathcal{R}(A^*)}$ auch

$$A^+AA^+y = P_{\overline{\mathcal{R}(A^*)}}A^+y = A^+y \quad \forall y \in \mathcal{D}(A^+).$$

Erfülle nun umgekehrt eine Abbildung $A^\circ : \mathcal{D}(A^\circ) = \mathcal{D}(A^+) \subset Y \rightarrow X$ die Moore-Penrose-Axiome. Für $y \in \mathcal{D}(A^+)$ und $x = A^\circ y$ ist dann

$$Ax = AA^\circ y = P_{\overline{\mathcal{R}(A)}}y,$$

Nach Satz 2.12 löst $x = A^\circ y$ also die Normalgleichungen. Mit Satz 2.15c) folgt

$$A^+y = P_{\overline{\mathcal{R}(A^*)}}A^\circ y = A^\circ AA^\circ y = A^\circ y.$$

Bemerkung 2.17

(a) Aus

$$\mathcal{N}(A^*A) = \{x : A^*Ax = 0\} \subseteq \{x : (Ax, Ax) = (A^*Ax, x) = 0\} = \mathcal{N}(A)$$

folgt $\mathcal{N}(A^*A) = \mathcal{N}(A)$.

(b) Ist A injektiv, so besitzt A^*A eine Linksinverse $(A^*A)^{-1}$ und es gilt

$$A^+y = (A^*A)^{-1}A^*y \quad \text{für alle } y \in \mathcal{D}(A^+).$$

(c) Endlich-dimensionale Vektorräume sind (bzgl. jeder Norm) vollständig. Daher gilt für lineare beschränkte Operatoren mit $\dim \mathcal{R}(A) < \infty$ stets, dass $\overline{\mathcal{R}(A)} = \mathcal{R}(A)$ und damit $\mathcal{D}(A^+) = Y$. Für Operatoren mit unendlich-dimensionalem Bild gilt dies jedoch im Allgemeinen nicht, dies zeigen wir in den Übungsaufgaben.

2.4 Kompakte Operatoren

Auch in diesem Abschnitt seien X, Y weiterhin stets Hilberträume.

2.4.1 Satz von Arzela-Ascoli

Wir setzen als bekannt voraus, dass eine Teilmenge $\Omega \subseteq X$ *kompakt* heißt, falls jede offene Überdeckung von X eine endliche Teilüberdeckung besitzt, und dass dazu äquivalent ist, dass jede Folge in Ω eine konvergente Teilfolge besitzt. Im endlich-dimensionalen ist eine Menge nach dem Satz von Bolzano-Weierstraß genau dann kompakt wenn sie abgeschlossen und beschränkt ist. Dies gilt jedoch im unendlich-dimensionalen nicht mehr. Um konvergente Teilfolgen in unendlich-dimensionalen Funktionenräumen zu erhalten ist der folgende Satz nützlich.

Satz 2.18 (Arzela-Ascoli)

Sei $\Omega \subset X$ kompakt und $(u_k)_{k \in \mathbb{N}} \subset C(\Omega)$ sei eine gleichmäßig beschränkte und gleichgradig stetige Folge stetiger Funktionen $u_k : \Omega \rightarrow \mathbb{R}$, d.h.

$$\exists C > 0 : \sup_{x \in \Omega} |u_k(x)| \leq C \quad \forall k \in \mathbb{N} \quad (2.4)$$

und

$$\sup_{k \in \mathbb{N}} \sup_{\substack{x, y \in \Omega, \\ \|x - y\| \leq \delta}} |u_k(x) - u_k(y)| \rightarrow 0 \quad \text{für } \delta \rightarrow 0. \quad (2.5)$$

Dann besitzt $(u_k)_{k \in \mathbb{N}}$ eine konvergente Teilfolge in $C(\Omega)$ bezüglich der Supremumsnorm $\|u\|_{C(\Omega)} := \sup_{x \in \Omega} |u(x)|$.

Beweis: (a) Wir zeigen zuerst, dass $(u_k)_{k \in \mathbb{N}}$ *Fast-Cauchy-Folgen* besitzt, d.h. zu jedem $\varepsilon > 0$ existiert eine Teilfolge $(u_{k_i})_{i \in \mathbb{N}}$ mit

$$\limsup_{i, j \rightarrow \infty} \|u_{k_i} - u_{k_j}\|_{C(\Omega)} < \varepsilon. \quad (2.6)$$

Hierzu überdecken wir $[-C, C] \subset \mathbb{R}$ und $\Omega \subset X$ mit Kugeln vom Radius $\delta > 0$ und werden zeigen, dass wir für jedes $\varepsilon > 0$ die Eigenschaft (2.6) durch hinreichend kleine Wahl von $\delta > 0$ sicherstellen können. Seien also $(x_m)_{m=1}^M \subset \Omega$ und $(y_l)_{l=1}^L \subset [-C, C]$, sodass

$$\Omega \subset \bigcup_{m=1}^M B_\delta(x_m) \quad \text{und} \quad [-C, C] \subset \bigcup_{l=1}^L B_\delta(y_l).$$

¹Wir schreiben dies auch abgekürzt als $\sup_{k \in \mathbb{N}} |u_k(x) - u_k(y)| \rightarrow 0$ für $\|x - y\|_X \rightarrow 0$.

KAPITEL 2. LINEARE INVERSE PROBLEME

Wegen (2.4) bildet jedes Element u_k der Funktionenfolge die Punkte $(x_m)_{m=1}^M$ nach $(u_k(x_m))_{m=1}^M \subset [-C, C]$ ab, wir können also jedem u_k (mindestens) eine Abbildung

$$\pi : \{1, \dots, M\} \rightarrow \{1, \dots, L\}$$

mit $|u_k(x_m) - y_{\pi(m)}| < \delta$ zuordnen.

Da es nur endlich viele solcher Abbildung gibt, existiert ein π , dem unendlich viele u_k , also eine Teilfolge $(u_{k_i})_{i \in \mathbb{N}}$, zugeordnet werden. Für jedes $x \in \Omega$ existiert ein $m \in \{1, \dots, M\}$ mit $x \in B_\delta(x_m)$ und es gilt

$$\begin{aligned} |u_{k_i}(x) - u_{k_j}(x)| &\leq |u_{k_i}(x) - u_{k_i}(x_m)| + |u_{k_i}(x_m) - y_{\pi(m)}| \\ &\quad + |y_{\pi(m)} - u_{k_j}(x_m)| + |u_{k_j}(x_m) - u_{k_j}(x)| \\ &\leq 2\delta + 2 \sup_{k \in \mathbb{N}} \sup_{\|x-y\|_X < \delta} |u_k(x) - u_k(y)| \end{aligned}$$

Die rechte Seite unterschreitet wegen (2.5) jedes vorgegebene $\varepsilon > 0$ für hinreichend kleine $\delta > 0$. Für jedes $\varepsilon > 0$ finden wir also eine Teilfolge mit der Eigenschaft (2.6).

- (b) Nun konstruieren wir mit einem sogenannten *Diagonalfolgenargument* aus den Fast-Cauchy-Folgen richtige Cauchy-Folgen. Nach (a) existiert eine Teilfolge² $(u_{1,k})_{k \in \mathbb{N}} \subseteq (u_k)_{k \in \mathbb{N}}$, die (2.6) mit $\delta_1 = 1$ erfüllt. Auf diese Teilfolge können wir wiederum (a) anwenden und finden auf diese Weise eine Teilfolge $(u_{2,k})_{k \in \mathbb{N}} \subseteq (u_{1,k})_{k \in \mathbb{N}}$, die (2.6) mit $\delta_2 = \frac{1}{2}$ und darin wiederum eine, die (2.6) mit $\delta_3 = \frac{1}{3}$ erfüllt, u.s.w.

Wir nehmen nun die Diagonalfolge $(u_{k,k})_{k \in \mathbb{N}}$, d.h. das erste Glied der Teilfolge zu $\delta_1 = 1$, dann das zweite Glied der Teilfolge zu $\delta_2 = \frac{1}{2}$, u.s.w. Diese Diagonalfolge $(u_{k,k})_{k \in \mathbb{N}}$ ist eine Teilfolge der ursprünglichen Folge $(u_k)_{k \in \mathbb{N}}$. Außerdem ist für jedes $l \in \mathbb{N}$ die Diagonalfolge $(u_{k,k})_{k \in \mathbb{N}}$ ab den l -ten Folgenglied Teilfolge von $(u_{l,k})_{k \in \mathbb{N}}$. Es gilt daher

$$\limsup_{j,k \rightarrow \infty} \|u_{j,j} - u_{k,k}\|_{C(K)} \leq \delta_l = \frac{1}{l} \quad \text{für alle } l \in \mathbb{N},$$

und damit $\lim_{j,k \rightarrow \infty} \|u_{j,j} - u_{k,k}\|_{C(K)} = 0$. Die Teilfolge $(u_{k,k})_{k \in \mathbb{N}}$ ist also eine Cauchy-Folge und damit konvergent, da $C(\Omega)$ vollständig ist. \square

²Hier und im Folgenden schreiben wir für zwei Folgen $(v_k)_{k \in \mathbb{N}} \subseteq (u_k)_{k \in \mathbb{N}}$, falls eine streng monoton wachsende Folge natürlicher Zahlen $k_1 < k_2 < \dots$ existiert, so dass $v_j = u_{k_j} \forall j \in \mathbb{N}$.

2.4.2 Kompakte Operatoren

Definition 2.19

$A \in \mathcal{L}(X, Y)$ heißt kompakt, falls das Bild jeder in X beschränkten Folge eine konvergente Teilfolge in Y besitzt bzw. (äquivalent dazu) wenn $\overline{A(B)}$ kompakt ist für alle beschränkten Teilmengen $B \subset X$. Die Menge aller kompakten Operatoren bezeichnen wir mit $\mathcal{K}(X, Y)$.

Satz 2.20

- (a) $\mathcal{K}(X, Y)$ ist ein Vektorraum. Jede Hintereinanderausführung von kompakten Operatoren und stetigen linearen Operatoren ist wiederum kompakt.
- (b) Jeder stetige lineare Operator mit endlich dimensionalem Bild ist kompakt (die sogenannten degenerierte Operatoren).
- (c) $\mathcal{K}(X, Y)$ ist abgeschlossen bezüglich der Operatornorm $\|\cdot\|_{\mathcal{L}(X, Y)}$. Insbesondere ist also wegen (b) jeder Grenzwert degenerierter Operatoren kompakt.

Beweis: (a) ist trivial.

(b) folgt aus dem Satz von Bolzano-Weierstrass.

(c) Sei $(K_j)_{j \in \mathbb{N}} \subset \mathcal{K}(X, Y)$ mit $K_j \rightarrow K \in \mathcal{L}(X, Y)$. Sei $(x_n)_{n \in \mathbb{N}} \subset X$ eine beschränkte Folge. Dann existiert eine Teilfolge $(x_{1,n})_{n \in \mathbb{N}} \subseteq (x_n)_{n \in \mathbb{N}}$, sodass $K_1 x_{1,n}$ konvergiert. Von dieser Teilfolge existiert wiederum eine Teil-Folge $(x_{2,n})_{n \in \mathbb{N}} \subseteq (x_{1,n})_{n \in \mathbb{N}}$, sodass $K_2 x_{2,n}$ konvergiert und davon eine Teil-Folge $(x_{3,n})_{n \in \mathbb{N}} \subseteq (x_{2,n})_{n \in \mathbb{N}}$, sodass $K_3 x_{3,n}$ konvergiert u.s.w.

Für jedes feste $j \in \mathbb{N}$ bilden dann fast alle Glieder der Diagonalfolge $(x_{n,n})_{n \in \mathbb{N}}$ eine Teilfolge von $(x_{j,n})_{n \in \mathbb{N}}$, sodass für jedes $j \in \mathbb{N}$ die Folge $(K_j x_{n,n})_{n \in \mathbb{N}} \subset Y$ konvergiert. Damit folgt für alle $j, m, n \in \mathbb{N}$

$$\begin{aligned} & \|Kx_{n,n} - Kx_{m,m}\|_Y \\ & \leq \|Kx_{n,n} - K_j x_{n,n}\|_Y + \|K_j x_{n,n} - K_j x_{m,m}\|_Y + \|K_j x_{m,m} - Kx_{m,m}\|_Y \\ & \leq \|K - K_j\|_{\mathcal{L}(X, Y)} (\|x_{n,n}\|_X + \|x_{m,m}\|_X) + \|K_j x_{n,n} - K_j x_{m,m}\|_Y \\ & \leq \|K - K_j\|_{\mathcal{L}(X, Y)} \sup_{n \in \mathbb{N}} \|x_n\|_X + \|K_j x_{n,n} - K_j x_{m,m}\|_Y. \end{aligned}$$

Für jedes $\varepsilon > 0$ existiert ein $j \in \mathbb{N}$, so dass der erste Summand kleiner als $\varepsilon/2$ ist, und zu diesem j ist für hinreichend große n, m auch der zweite Summand kleiner als $\varepsilon/2$. Es gilt also

$$\lim_{n, m \rightarrow \infty} \|Kx_{n,n} - Kx_{m,m}\|_Y = 0,$$

d.h. $(Kx_{n,n})_{n \in \mathbb{N}}$ ist eine Cauchy-Folge und damit konvergent. □

Satz 2.21

Ist $K \in \mathcal{K}(X, Y)$, so ist $K^* \in \mathcal{K}(Y, X)$.

Beweis: Sei $(y_n)_{n \in \mathbb{N}} \subset Y$ beschränkt. Wir müssen zeigen, dass $(K^* y_n)_{n \in \mathbb{N}}$ eine konvergente Teilfolge besitzt. Dazu definieren wir zuerst

$$l_n : Y \rightarrow \mathbb{R}, \quad l_n(\eta) := (y_n, \eta).$$

Offenbar ist $l_n \in \mathcal{L}(Y, \mathbb{R})$ und

$$|l_n(\eta) - l_n(\eta')| = |(y_n, \eta - \eta')| \leq \|y_n\|_Y \|\eta - \eta'\|_Y \quad \forall \eta, \eta' \in Y,$$

also sind die $(l_n)_{n \in \mathbb{N}}$ als stetige Funktionen auf Y auch gleichgradig stetig.

Sei $B := \{x \in X : \|x\|_X \leq 1\} \subset X$ die Einheitskugel in X . Da K kompakt ist, ist $\Omega := \overline{K(B)} \subset Y$ eine kompakte Teilmenge von Y .

Die Folge der Einschränkungen $(l_n|_{\Omega})_{n \in \mathbb{N}} \subset C(\Omega)$ ist auf Ω gleichmäßig beschränkt und gleichgradig stetig, besitzt also nach dem Satz von Arzela-Ascoli (Satz 2.18) eine in $C(\Omega)$ konvergente Teilfolge $(l_{n_j})_{j \in \mathbb{N}}$.

Für $(l_{n_j})_{j \in \mathbb{N}}$ gilt

$$\lim_{i, j \rightarrow \infty} \sup_{\eta \in \Omega} |l_{n_i}(\eta) - l_{n_j}(\eta)| = 0,$$

und damit insbesondere

$$\begin{aligned} \lim_{i, j \rightarrow \infty} \|K^*(y_{n_i} - y_{n_j})\|_X &= \lim_{i, j \rightarrow \infty} \sup_{x \in B} |(K^*(y_{n_i} - y_{n_j}), x)_X| \\ &= \lim_{i, j \rightarrow \infty} \sup_{\eta \in K(B)} |(y_{n_i} - y_{n_j}, \eta)_Y| = 0. \end{aligned}$$

$(K^*(y_{n_j}))_{j \in \mathbb{N}}$ ist also eine Cauchy-Folge, und damit konvergent. □

2.4.3 Kompaktheit von Integraloperatoren

Satz 2.22

Sei $\Omega \subset \mathbb{R}^n$ ein beschränktes Gebiet. Zu einer Funktion $k \in L^2(\Omega \times \Omega)$ definieren wir den Operator

$$K : L^2(\Omega) \rightarrow L^2(\Omega), \quad (Kf)(x) := \int_{\Omega} k(x, y) f(y) dy.$$

K heißt (Fredholm-)Integraloperator zur (Fredholm-)Kernfunktion k .

Dann ist K kompakt, $\|K\|_{\mathcal{L}(L^2(\Omega))} \leq \|k\|_{L^2(\Omega \times \Omega)}$ und $K^* : L^2(\Omega) \rightarrow L^2(\Omega)$ erfüllt

$$(K^*g)(x) := \int_{\Omega} k(y, x) g(y) dy \quad \text{für alle } g \in L^2(\Omega), x \in \Omega \text{ (f.ü.)}$$

Beweis: Für alle $f \in L^2(\Omega)$ ist

$$\begin{aligned} \|Kf\|_{L^2(\Omega)}^2 &= \int_{\Omega} |(Kf)(x)|^2 dx = \int_{\Omega} \left| \int_{\Omega} k(x,y)f(y) dy \right|^2 dx \\ &\leq \int_{\Omega} \left(\int_{\Omega} |k(x,y)|^2 dy \right) \left(\int_{\Omega} |f(y)|^2 dy \right) dx \\ &= \|k\|_{L^2(\Omega \times \Omega)}^2 \|f\|_{L^2(\Omega)}^2, \end{aligned}$$

also

$$K \in \mathcal{L}(L^2(\Omega)) \quad \text{und} \quad \|K\|_{\mathcal{L}(L^2(\Omega))} \leq \|k\|_{L^2(\Omega \times \Omega)}.$$

Da K offenbar linear von k abhängt, folgt damit auch dass $K \in \mathcal{L}(L^2(\Omega))$ stetig von $k \in L^2(\Omega \times \Omega)$ abhängt. Da $C(\overline{\Omega} \times \overline{\Omega})$ dicht in $L^2(\Omega \times \Omega)$ liegt (siehe z.B. [Forster3, §10, Satz 3]), genügt es nach Satz 2.20 die Kompaktheit von K für $k \in C(\overline{\Omega} \times \overline{\Omega})$ zu zeigen.

Sei also $k \in C(\overline{\Omega} \times \overline{\Omega})$ und $(f_l)_{l \in \mathbb{N}} \subset L^2(\Omega)$ beschränkt. Wir zeigen, dass $(Kf_l)_{l \in \mathbb{N}}$ die Voraussetzungen des Satzes von Arzela-Ascoli erfüllt.

Glm. Beschränktheit:

$$|Kf_l(x)|^2 \leq \int_{\Omega} |k(x,y)|^2 dy \|f_l\|_{L^2(\Omega)}^2 \leq |\Omega| \sup_{y \in \Omega} |k(x,y)|^2 \|f_l\|_{L^2(\Omega)}^2, \quad (2.7)$$

also $\sup_{x \in \Omega, l \in \mathbb{N}} |Kf_l(x)| < \infty$.

Glgd. Stetigkeit:

$$\begin{aligned} |(Kf_l)(x) - (Kf_l)(\xi)|^2 &= \left| \int_{\Omega} (k(x,y) - k(\xi,y)) f_l(y) dy \right|^2 \\ &\leq \|f_l\|_{L^2(\Omega)}^2 \int_{\Omega} |k(x,y) - k(\xi,y)|^2 dy \\ &\leq \|f_l\|_{L^2(\Omega)}^2 |\Omega| \sup_{y \in \Omega} |k(x,y) - k(\xi,y)|^2 \end{aligned}$$

Da k glm. stetig ist und f_l beschränkt ist, folgt

$$\sup_{l \in \mathbb{N}} |(Kf_l)(x) - (Kf_l)(\xi)| \rightarrow 0 \quad \text{für } |x - \xi| \rightarrow 0.$$

Die Folge $(Kf_l)_{l \in \mathbb{N}}$ ist also gleichgradig stetig und insbesondere ist damit auch $Kf_l \in C(\overline{\Omega})$ für alle $l \in \mathbb{N}$.

Nach dem Satz von Arzela-Ascoli besitzt $(Kf_l)_{l \in \mathbb{N}}$ eine in $C(\overline{\Omega})$ konvergente Teilfolge und genauso wie in (2.7) folgt, dass diese auch in $L^2(\Omega)$ konvergiert. K ist also kompakt.

KAPITEL 2. LINEARE INVERSE PROBLEME

Schließlich ist für alle $f, g \in L^2(\Omega)$

$$\begin{aligned}(f, K^*g) &= (Kf, g) = \int_{\Omega} \left(\int_{\Omega} k(x, y) f(y) dy \right) g(x) dx \\ &= \int_{\Omega} f(x) \left(\int_{\Omega} k(y, x) g(y) dy \right) dx,\end{aligned}$$

also $(K^*g)(x) = \int_{\Omega} k(y, x) g(y) dy$. □

Folgerung 2.23

Den Operator aus Abschnitt 1.2

$$A : L^2(0, 1) \rightarrow L^2(0, 1), \quad (Af)(x) := \int_0^x f(y) dy$$

können wir als Integraloperator

$$(Af)(x) = \int_0^1 k(x, y) f(y) dy, \quad \text{mit } k(x, y) := \begin{cases} 1 & \text{für } y < x \\ 0 & \text{sonst} \end{cases}$$

schreiben.

Offenbar ist $k \in L^2((0, 1)^2)$, also ist A nach Satz 2.22 kompakt.

2.4.4 Spektraltheorie kompakter Operatoren

Wir werden in diesem Abschnitt die Eigenwertzerlegung für kompakte selbst-adjungierte Operatoren $K \in \mathcal{K}(X)$ einführen und daraus dann wie im endlich-dimensionalen die Singulärwertzerlegung für allgemeine kompakte Operatoren erhalten. Dabei sei weiterhin X stets ein Hilbertraum.

2.4.4.1 Quadratische Formen und Hilbertraum-ONBs

Satz 2.24

Sei $A \in \mathcal{L}(X)$, $A = A^*$. Dann ist

$$\|A\|_{\mathcal{L}(X)} = \sup_{\|x\|=1} \|Ax\| = \sup_{\|x\|=1} |(Ax, x)|.$$

Beweis: Offenbar ist

$$S := \sup_{\|x\|=1} |(Ax, x)| \leq \sup_{\|x\|=1} \|Ax\| = \|A\|_{\mathcal{L}(X)}.$$

Außerdem gilt für jedes $c > 0$

$$\begin{aligned} 4 \|Ax\|^2 &= \left(A \left(cx + \frac{1}{c} Ax \right), \left(cx + \frac{1}{c} Ax \right) \right) \\ &\quad - \left(A \left(cx - \frac{1}{c} Ax \right), \left(cx - \frac{1}{c} Ax \right) \right) \\ &\leq S \left(\left\| cx + \frac{1}{c} Ax \right\|^2 + \left\| cx - \frac{1}{c} Ax \right\|^2 \right) \\ &= 2S \left(c^2 \|x\|^2 + \frac{1}{c^2} \|Ax\|^2 \right). \end{aligned}$$

Falls $Ax \neq 0$ ist, dann ist auch $x \neq 0$ und mit der Wahl $c^2 := \frac{\|Ax\|}{\|x\|}$ folgt

$$\|Ax\| \leq S \|x\|.$$

Offenbar gilt dies auch falls $Ax = 0$. Es ist also

$$S \geq \sup_{\|x\|=1} \|Ax\| = \|A\|_{\mathcal{L}(X)}$$

und damit die Behauptung gezeigt. □

Satz 2.25

Für $A \in \mathcal{L}(X, Y)$ gilt $\|A^*A\|_{\mathcal{L}(X)} = \|A\|_{\mathcal{L}(X, Y)}^2$.

Beweis: Es gilt

$$\|A\|_{\mathcal{L}(X, Y)}^2 = \sup_{\|x\|=1} \|Ax\|^2 = \sup_{\|x\|=1} (A^*Ax, x) = \|A^*A\|_{\mathcal{L}(X)},$$

wobei wir im letzten Schritt Satz 2.24 verwendet haben. □

Definition 2.26

Sei $A \in \mathcal{L}(X)$, $A = A^*$. A heißt

- (a) positiv, falls $(Ax, x) > 0$ für alle $x \neq 0$.
- (b) positiv semidefinit, falls $(Ax, x) \geq 0$ für alle $x \in X$.
- (c) positiv definit oder koerziv, falls

$$\exists \alpha > 0 : (Ax, x) \geq \alpha \|x\|^2 \quad \text{für alle } x \in X.$$

KAPITEL 2. LINEARE INVERSE PROBLEME

Analog heißt A negativ (neg. semidefinit, neg. definit), falls $-A$ positiv (pos. semidefinit, pos. definit) ist.

Bemerkung 2.27

In der linearen Algebra definiert man als Basis eines Vektorraums eine Teilmenge von Vektoren mit der Eigenschaft, dass sich jedes Element des Vektorraums auf eindeutige Weise als endliche Linearkombination der Elemente dieser Teilmenge schreiben lässt. Dies nennt man auch Hamelbasis.

In einem normierten Raum ist es naheliegend in der Basisdefinition auch unendliche, bezüglich der Norm konvergente Linearkombinationen zuzulassen (sogenannte Schauderbasis). In unendlich-dimensionalen Hilberträumen bezeichnet der Begriff Orthonormalbasis üblicherweise eine Schauderbasis. Wir führen diesen Begriff im Folgenden rigoros ein.

Definition und Satz 2.28

Ein abzählbar unendliches Orthonormalsystem $(e_n)_{n \in \mathbb{N}} \subset X$ heißt Orthonormalbasis (ONB), falls eine der folgenden äquivalenten Bedingungen erfüllt ist:

- (a) Das Erzeugnis $\langle e_1, e_2, \dots \rangle$ (d.h. der Raum aller endlichen Linearkombinationen) liegt dicht in X .
- (b) $x = \sum_{n=1}^{\infty} (x, e_n) e_n$ für alle $x \in X$.
- (c) $(x, \xi) = \sum_{n=1}^{\infty} (x, e_n)(\xi, e_n)$ für alle $x, \xi \in X$.
- (d) $\|x\|^2 = \sum_{n=1}^{\infty} (x, e_n)^2$ für alle $x \in X$.

Dies zeigt, dass wir $x \in X$ mit dem unendlichen langen Vektor seiner Entwicklungskoeffizienten

$$x = \begin{pmatrix} (x, e_1) \\ (x, e_2) \\ \vdots \end{pmatrix} \in l^2(\mathbb{N})$$

identifizieren können und damit wie im endlich-dimensionalen rechnen können.

Hilberträume in denen eine (endliche oder abzählbare unendliche) ONB existiert heißen separabel.

Beweis: Für jedes $x \in X$ und $V_N := \langle e_1, \dots, e_N \rangle$ gilt offenbar

$$\xi := \sum_{n=1}^N (x, e_n) e_n \in V_N, \quad x - \xi \in V_N^\perp,$$

also ist nach Satz 2.6 $P_{V_N}x = \xi$ und damit insbesondere

$$\left\| x - \sum_{n=1}^N (x, e_n) e_n \right\| = \min_{v \in V_N} \|x - v\| = \text{dist}(x, \langle e_1, \dots, e_N \rangle). \quad (2.8)$$

(a) \implies (b): Sei $x \in X$ und

$$x_m \in \langle e_1, e_2, \dots \rangle \quad \text{mit} \quad x_m \rightarrow x.$$

Dann ist $x_m \in V_{N_m}$ für ein $N_m \in \mathbb{N}$ und mit (2.8) folgt für jedes $N \geq N_m$

$$\left\| x - \sum_{n=1}^N (x, e_n) e_n \right\| \leq \|x - x_m\| \rightarrow 0,$$

also $\lim_{N \rightarrow \infty} \sum_{n=1}^N (x, e_n) e_n = x$.

(b) \implies (c): folgt aus der Stetigkeit des Skalarproduktes.

(c) \implies (d): setze $\xi := x$.

(d) \implies (a): Wir wenden (d) auf $\xi := x - \sum_{n=1}^N (x, e_n) e_n$ an und erhalten

$$\left\| x - \sum_{n=1}^N (x, e_n) e_n \right\|^2 = \sum_{n=1}^{\infty} (\xi, e_n)^2 = \sum_{n=N+1}^{\infty} (x, e_n)^2 \rightarrow 0.$$

Dies zeigt (b), und (b) \implies (a) ist trivial. □

2.4.4.2 Spektralsatz für kompakte Operatoren

Satz 2.29

Sei $K \in \mathcal{K}(X)$, $K = K^*$ und K positiv semidefinit. Dann hat K nur abzählbar viele von Null verschiedene Eigenwerte

$$\|K\| = \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots > 0,$$

die sich höchstens im Nullpunkt häufen (d.h. falls unendlich viele Eigenwerte existieren, so ist $\lambda_n \rightarrow 0$).

Die zugehörigen Eigenvektoren v_n können so gewählt werden, dass sie eine ONB von $\mathcal{N}(K)^\perp$ bilden.

Beweis: Für $K = 0$ ist die Aussage trivial. Für $K \neq 0$, existiert wegen Satz 2.24 eine Folge $(x_j)_{j \in \mathbb{N}} \subset X$ mit

$$\|x_j\| = 1, \quad (Kx_j, x_j) \rightarrow \|K\| =: \lambda_1 > 0.$$

Aus

$$\begin{aligned} 0 &\leq \|Kx_j - \lambda_1 x_j\|^2 = \|Kx_j\|^2 - 2\lambda_1(Kx_j, x_j) + \lambda_1^2 \|x_j\|^2 \\ &\leq 2\lambda_1^2 - 2\lambda_1(Kx_j, x_j) \rightarrow 0. \end{aligned}$$

folgt

$$Kx_j - \lambda_1 x_j \rightarrow 0. \quad (2.9)$$

Da K kompakt ist, besitzt Kx_j eine konvergente Teilfolge, o.B.d.A. sei dies bereits Kx_j . Aus (2.9) folgt, dass auch x_j konvergiert. Mit $v_1 := \lim_{j \rightarrow \infty} x_j$ ist dann

$$\|v_1\| = 1, \quad Kv_1 - \lambda_1 v_1 = 0.$$

v_1 ist also Eigenvektor zum Eigenwert λ_1 .

Nun setzen wir $X_1 = \langle v_1 \rangle^\perp$. Für jedes $x \in X_1$ ist

$$(Kx, v_1) = (x, Kv_1) = \lambda_1(x, v_1) = 0,$$

also $K(X_1) \subseteq X_1$. Offenbar ist damit $K_1 := K|_{X_1} \in \mathcal{K}(X_1)$ positiv semidefinit, $K_1 = K_1^*$ und $\|K_1\| \leq \|K\|$. Ist $K_1 = 0$, so ist die Aussage bewiesen, anderenfalls existieren $v_2 \in X_1$ und $\lambda_1 \geq \lambda_2 := \|K_1\| > 0$, sodass $Kv_2 = K_1 v_2 = \lambda_2 v_2$.

Wir fahren so fort und erhalten eine (möglicherweise abbrechende) Folge von Eigenwerten $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots$ mit zugehörigem Orthonormalsystem von Eigenvektoren (v_1, v_2, v_3, \dots) . Die Folge ist genau dann endlich, wenn $K_m = K|_{X_m} = 0$ für ein $m \in \mathbb{N}$ und in diesem Fall ist also

$$X_m = \langle v_1, \dots, v_m \rangle^\perp \subseteq \mathcal{N}(K).$$

Da außerdem für jedes $v \in \mathcal{N}(K)$ gilt, dass

$$(v, v_j) = \frac{1}{\lambda_j}(v, Kv_j) = \frac{1}{\lambda_j}(Kv, v_j) = 0,$$

gilt $\langle v_1, \dots, v_m \rangle^\perp = \mathcal{N}(K)$ und damit $\mathcal{N}(K)^\perp = \langle v_1, \dots, v_m \rangle$ nach Satz 2.6.

Falls die Folge der λ_n unendlich ist, so ist sie wegen Monotonie und Beschränktheit konvergent, es existiert also $\lambda := \lim_{n \rightarrow \infty} \lambda_n \geq 0$. Wäre $\lambda \neq 0$, so folgt aus $\lambda_n \geq \lambda$, dass $\left\| \frac{1}{\lambda_n} v_n \right\| \leq \frac{1}{\lambda}$ für alle $n \in \mathbb{N}$. Da K kompakt ist, hat also $K \left(\frac{1}{\lambda_n} v_n \right) = v_n$ eine konvergente Teilfolge. Dies widerspricht aber der Orthogonalität der v_n , aufgrund derer $\|v_n - v_j\| = \sqrt{2}$ für alle $j, n \in \mathbb{N}$ gilt. Es gilt daher

$$\lim_{n \rightarrow \infty} \lambda_n = 0.$$

Es ist noch zu zeigen, dass $\overline{\langle v_1, v_2, \dots \rangle} = \mathcal{N}(K)^\perp$. Sei dazu $v \in \langle v_1, v_2, \dots \rangle^\perp$. Dann liegt nach Konstruktion $v \in X_n$, also $\|Kv\| = \|K_n v\| \leq \lambda_{n+1} \|v\|$ für alle $n \in \mathbb{N}$ und wegen $\lambda_n \rightarrow 0$ folgt damit $Kv = 0$. Damit ist

$$\langle v_1, v_2, \dots \rangle^\perp \subseteq \mathcal{N}(K).$$

Wie im Fall einer abbrechenden Folge gilt auch $\mathcal{N}(K) \subseteq \langle v_1, v_2, \dots \rangle^\perp$, so dass nach Satz 2.6 $\mathcal{N}(K)^\perp = \overline{\langle v_1, v_2, \dots \rangle}$.

Wie im endlich-dimensionalen gilt, dass Eigenvektoren zu verschiedenen Eigenwerten senkrecht aufeinander stehen, sodass damit auch gezeigt ist, dass keine weiteren (von Null verschiedenen) Eigenwerte existieren können. \square

2.4.4.3 Singulärwertzerlegung kompakter Operatoren

Definition und Satz 2.30

Für jedes $K \in \mathcal{K}(X, Y)$ existiert eine (möglicherweise endliche) Folge

$$\sigma_1 \geq \sigma_2 \geq \dots > 0$$

von sich höchstens im Nullpunkt häufenden Zahlen, sowie Orthonormalfolgen

$$(u_n)_{n \in \mathbb{N}} \subset Y, \quad (v_n)_{n \in \mathbb{N}} \subset X \quad \text{mit} \quad Kv_n = \sigma_n u_n, \quad K^* u_n = \sigma_n v_n.$$

Außerdem ist $\|K\| = \sigma_1$ und $(v_n)_{n \in \mathbb{N}}$ bildet eine ONB von $\mathcal{N}(K)^\perp$.

(u_n, v_n, σ_n) heißt Singulärwertzerlegung.

Beweis: Offenbar ist $K^*K \in \mathcal{K}(X)$ selbstadjungiert und positiv semidefinit. Gemäß Satz 2.29 existiert also eine ONB $(v_n)_{n \in \mathbb{N}}$ von $\mathcal{N}(K^*K)^\perp = \mathcal{N}(K)^\perp$ (siehe Bemerkung 2.17(a)), sowie eine Folge

$$\lambda_1 \geq \lambda_2 \geq \dots > 0,$$

mit $K^*Kv_n = \lambda_n v_n$ und $\lambda_1 = \|K^*K\| = \|K\|^2$ (siehe Satz 2.25).

Mit $\sigma_n := \sqrt{\lambda_n}$ und $u_n := \frac{1}{\sigma_n} Kv_n$ folgt

$$Kv_n = \sigma_n u_n, \quad K^* u_n = \frac{1}{\sigma_n} K^* Kv_n = \sigma_n v_n, \quad \|K\| = \sigma_1$$

und wegen

$$(u_n, u_m) = \frac{1}{\sigma_n \sigma_m} (Kv_n, Kv_m) = \frac{\lambda_n}{\sigma_n \sigma_m} (v_n, v_m) = \delta_{nm}.$$

ist auch $(u_n)_{n \in \mathbb{N}}$ ein Orthonormalsystem. \square

Bemerkung 2.31

Sei $K \in \mathcal{K}(X, Y)$ mit unendlicher Singulärwertzerlegung $(u_n, v_n, \sigma_n)_{n \in \mathbb{N}}$.

(a) Da (v_n) eine ONB von $\mathcal{N}(K)^\perp$ bildet, können wir jedes $x \in X$ gemäß Satz 2.28 schreiben als

$$x = x_0 + \sum_{n=1}^{\infty} (v_n, x) v_n \quad \text{mit } x_0 \in \mathcal{N}(K).$$

Es gilt also

$$Kx = \sum_{n=1}^{\infty} \sigma_n (v_n, x) u_n.$$

(b) Wir definieren $A_N : X \rightarrow Y$ durch

$$A_N x = \sum_{n=1}^N \sigma_n (v_n, x) u_n.$$

Dann ist $\dim \mathcal{R}(A_N) \leq N < \infty$.

Aus Satz 2.28 und Satz 2.30 folgt, dass für alle $x \in X$

$$\begin{aligned} \|Kx - A_N x\|^2 &= \left\| \sum_{n=N+1}^{\infty} \sigma_n (v_n, x) u_n \right\|^2 \leq \sigma_{N+1}^2 \sum_{n=N+1}^{\infty} (v_n, x)^2 \\ &\leq \sigma_{N+1}^2 \|x\|^2 \end{aligned}$$

und damit

$$\|K - A_N\|_{\mathcal{L}(X, Y)} \rightarrow 0, \quad \text{d.h. } A_N \rightarrow K$$

Zusammen mit Satz 2.20 erhalten wir, dass die kompakten Operatoren genau der Abschluss der Operatoren mit endlich-dimensionalen Bild (sog. degenerierte Operatoren) sind.

(c) Wir können

$$Kx = \sum_{n=1}^{\infty} \sigma_n (v_n, x) u_n,$$

auch formal mit unendlich-dimensionalen Matrizen schreiben als

$$K = U \Sigma V^*,$$

mit

$$\begin{aligned}
 U &= (u_1 \ u_2 \ \dots) & V &= (v_1 \ v_2 \ \dots) \\
 V^* &= \begin{pmatrix} v_1^* \\ v_2^* \\ \vdots \end{pmatrix}, & V^*x &= \begin{pmatrix} (v_1, x) \\ (v_2, x) \\ \vdots \end{pmatrix}, \\
 \Sigma &= \begin{pmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots \end{pmatrix}.
 \end{aligned}$$

2.4.4.4 Singulärwertzerlegung und Moore-Penrose-Inverse

Mit Hilfe der Singulärwertzerlegung lässt sich die Moore-Penrose-Inverse eines kompakten Operators explizit angeben.

Satz 2.32

Sei $K \in \mathcal{K}(X, Y)$.

(a) Ist K degeneriert, d.h. $\dim \mathcal{R}(K) < \infty$, so besitzt K eine endliche Singulärwertzerlegung $(u_n, v_n, \sigma_n)_{n=1}^N$ mit $N = \dim \mathcal{R}(K)$.

In dem Fall ist K^+ stetig, $\|K^+\| = \frac{1}{\sigma_N}$, $\mathcal{D}(K^+) = Y$ und es gilt

$$K^+y = \sum_{n=1}^N \sigma_n^{-1}(y, u_n)v_n.$$

(b) Ist K nicht degeneriert, d.h. $\dim \mathcal{R}(K) = \infty$, so besitzt K eine unendliche Singulärwertzerlegung $(u_n, v_n, \sigma_n)_{n=1}^\infty$.

In dem Fall ist K^+ nicht stetig, für jedes $y \in Y$ gilt

$$y \in \mathcal{D}(K^+) = \mathcal{R}(K) \oplus \mathcal{R}(K)^\perp \iff \sum_{n=1}^\infty \frac{(y, u_n)^2}{\sigma_n^2} < \infty, \quad (2.10)$$

(Picard-Kriterium) und für alle $y \in \mathcal{D}(K^+)$ gilt

$$K^+y = \sum_{n=1}^\infty \sigma_n^{-1}(y, u_n)v_n. \quad (2.11)$$

Beweis: Da $\mathcal{R}(K) = \langle u_1, u_2, \dots \rangle$ und die Vektoren u_1, u_2, \dots orthogonal und damit linear unabhängig sind, ist die Anzahl der Singulärwerte und -vektoren gleich $\dim \mathcal{R}(K)$. K ist also genau dann degeneriert, wenn es eine endliche SWZ besitzt.

KAPITEL 2. LINEARE INVERSE PROBLEME

- (a) Es sei $\dim \mathcal{R}(K) = N < \infty$. Dann ist $\mathcal{R}(K) = \overline{\mathcal{R}(K)}$ und daher nach Satz 2.15 $\mathcal{D}(K^+) = Y$. Für $y \in Y$ und $x := K^+y$ gilt

$$x \in \mathcal{N}(K)^\perp \quad \text{und} \quad K^*Kx = K^*y.$$

Wegen $\mathcal{N}(K)^\perp = \langle v_1, \dots, v_N \rangle$ ist deshalb

$$K^+y = x = \sum_{n=1}^N (x, v_n) v_n = \sum_{n=1}^N \frac{1}{\sigma_n} (y, u_n) v_n,$$

wobei wir ausgenutzt haben, dass wegen $K^*Kx = K^*y$ gilt

$$(x, v_n) = \frac{1}{\sigma_n^2} (x, K^*Kv_n) = \frac{1}{\sigma_n^2} (K^*Kx, v_n) = \frac{1}{\sigma_n^2} (K^*y, v_n) = \frac{1}{\sigma_n} (y, u_n).$$

Also ist für alle $y \in Y$

$$\|K^+y\|^2 = \left\| \sum_{n=1}^N \frac{1}{\sigma_n} (y, u_n) v_n \right\|^2 = \sum_{n=1}^N \frac{1}{\sigma_n^2} (y, u_n)^2 \leq \frac{1}{\sigma_N^2} \|y\|^2$$

und daher ist K^+ beschränkt und $\|K^+\| \leq \frac{1}{\sigma_N}$. Mit dieser Summendarstellung folgt auch $K^+u_N = \frac{1}{\sigma_N} v_N$ und daher ist $\|K^+\| = \frac{1}{\sigma_N}$.

- (b) Ist $\dim \mathcal{R}(K) = \infty$, dann folgt mit dem gleichen Argument wie in (a), dass für alle $y \in \mathcal{D}(K^+)$ und $x = K^+y$

$$(x, v_n) = \frac{1}{\sigma_n} (y, u_n).$$

Da $\|x\|^2 = \sum_{n=1}^{\infty} (x, v_n)^2$ folgt damit

$$\sum_{n=1}^{\infty} \frac{(y, u_n)^2}{\sigma_n^2} < \infty \quad \text{und} \quad K^+y = x = \sum_{n=1}^{\infty} (x, v_n) v_n = \sum_{n=1}^{\infty} \frac{1}{\sigma_n} (y, u_n) v_n.$$

Damit ist die Hinrichtung im Picard-Kriterium (2.10) sowie (2.11) gezeigt. Um noch die Rückrichtung im Picard-Kriterium (2.10) zu zeigen sei $Y \in Y$ und $\sum_{n=1}^{\infty} \frac{(y, u_n)^2}{\sigma_n^2}$ konvergiere. Dann bildet

$$\left(\sum_{n=1}^N \sigma_n^{-1} (y, u_n) v_n \right)_{N \in \mathbb{N}}$$

eine Cauchy-Folge, sodass der Grenzwert $x := \sum_{n=1}^{\infty} \sigma_n^{-1} (y, u_n) v_n$ existiert.

Da $K^*y \in \mathcal{N}(K)^\perp$ liegt, ist außerdem

$$K^*y = \sum_{n=1}^{\infty} (K^*y, v_n)v_n = \sum_{n=1}^{\infty} \sigma_n(y, u_n)v_n = K^*Kx,$$

also nach Satz 2.13(a) $y \in D(K^+)$.

Schließlich ist K^+ nicht stetig, da mit den gezeigten Summenformeln folgt, dass $K^+u_n = \frac{1}{\sigma_n}v_n$, und wegen $\sigma_n \rightarrow 0$ daher $\|K^+u_n\|$ beliebig groß wird. \square

Folgerung 2.33

Sei $K \in \mathcal{K}(X, Y)$. K^+ ist genau dann stetig, wenn K degeneriert ist.

Beweis: Dies folgt sofort aus Satz 2.32. \square

Einige weitere wichtige Eigenschaften kompakter Operatoren werden wir in Abschnitt 3.2 behandeln. Wir beenden diesen Abschnitt noch mit einem Beispiel zur Singulärwertzerlegung.

Beispiel 2.34

Wir betrachten wieder den Operator aus Abschnitt 1.2 und Folgerung 2.23

$$A : L^2(0, 1) \rightarrow L^2(0, 1), \quad Af := \int_0^x f(y) dy = \int_0^1 k(x, y)f(y) dy,$$

mit

$$k(x, y) := \begin{cases} 1 & \text{für } y < x \\ 0 & \text{sonst} \end{cases}$$

In den Übungen zeigen wir, dass die Singulärwertzerlegung von A gegeben ist durch

$$\begin{aligned} \sigma_j &= \frac{1}{(j-1/2)\pi}, \\ v_j(x) &= \sqrt{2} \cos((j-1/2)\pi x), \\ u_j(x) &= \sqrt{2} \sin((j-1/2)\pi x). \end{aligned}$$

KAPITEL 2. LINEARE INVERSE PROBLEME

Kapitel 3

Regularisierung linearer Probleme

In diesem Kapitel seien weiterhin X, Y Hilberträume und es sei stets $K \in \mathcal{K}(X, Y)$ nicht-degeneriert mit Singulärwertzerlegung $(u_n, v_n, \sigma_n)_{n \in \mathbb{N}}$.

3.1 Regularisierung

3.1.1 Motivation und Definition

Wir betrachten wieder das Problem $Kx = y$ zu lösen, wobei wir annehmen, dass wir y nur bis auf einen kleinen Messfehler δ kennen, d.h. wir kennen y^δ mit $\|y^\delta - y\| < \delta$.

Ist K nicht degeneriert, so ist die Moore-Penrose-Inverse K^+ nicht überall definiert und nicht stetig, d.h. im Allgemeinen ist $y^\delta \notin \mathcal{D}(K^+)$ und selbst im Falle $y^\delta, y \in \mathcal{D}(K^+)$, so kann (für beliebig kleine $\delta > 0$) die Abweichung $K^+y^\delta - K^+y$ beliebig groß sein.

$$y^\delta \rightarrow y \quad \text{impliziert nicht} \quad K^+y^\delta \rightarrow K^+y.$$

Die naive Anwendung der Moore-Penrose-Inverse scheitert also selbst für beliebig kleine Messfehler!

In diesem Kapitel verfolgen wir die Idee, K^+ durch stetige Approximation R_α zu ersetzen. Der Parameter α steuert dabei wie stetig R_α ist bzw. wie gut R_α die gewünschte Inverse K^+ approximiert. Unser Ziel ist es, mit einer geeigneten *Parameterwahlstrategie* $\alpha = \alpha(\delta, y^\delta)$ zu erreichen, dass

$$R_{\alpha(\delta, y^\delta)}y^\delta \rightarrow x = K^+y \quad \text{für } \delta \rightarrow 0.$$

Definition 3.1

Eine Familie $(R_\alpha)_{\alpha>0}$ von linearen Operatoren $R_\alpha : Y \rightarrow X$ heißt Regularisierung von K^+ für $\alpha \rightarrow 0$, falls

- (a) $R_\alpha \in \mathcal{L}(Y, X)$ für alle $\alpha > 0$ (Stabilität)
- (b) $R_\alpha y \rightarrow K^+ y$ für alle $y \in \mathcal{D}(K^+)$ (pktw. Konvergenz).

Analog verwenden wir auch Regularisierungen $(R_k)_{k \in \mathbb{N}}$ und ersetzen in diesem Fall $\alpha \rightarrow 0$ durch $k \rightarrow \infty$.

3.1.2 Filter-basierte Regularisierung

Eine Reihe wichtiger Regularisierungsmethoden lässt sich einheitlich durch Einführung von Spektralfiltern behandeln. Dahinter steckt die Idee, dass die Schlechtgestelltheit des inversen Problems durch die gegen Null konvergierenden Singulärwerte des kompakten Operators K verursacht wird und wir durch eine dies vermeidende Abänderung der Singulärwerte stetige Approximationen an K^+ erhalten können. Wir führen die nötigen Eigenschaften der Abänderung zunächst abstrakt ein.

Definition 3.2

Sei $(F_\alpha)_{\alpha>0}$ eine Familie beschränkter Funktionen $F_\alpha : (0, \|K\|^2] \rightarrow \mathbb{R}$. $(F_\alpha)_{\alpha>0}$ heißt Filter, falls

$$\lim_{\alpha \rightarrow 0} F_\alpha(\lambda) = 1/\lambda \quad \forall \lambda \in (0, \|K\|^2].$$

Ist zusätzlich $\lambda F_\alpha(\lambda)$ glm. beschränkt in α , d.h.

$$\exists C > 0 : \lambda |F_\alpha(\lambda)| \leq C \quad \forall \alpha > 0, \lambda \in (0, \|K\|^2],$$

so heißt $(F_\alpha)_{\alpha>0}$ regularisierender Filter.

Definition und Satz 3.3

Sei $(F_\alpha)_{\alpha>0}$ ein Filter. Wir definieren damit

$$R_\alpha : Y \rightarrow X, \quad R_\alpha y := \sum_{n=1}^{\infty} F_\alpha(\sigma_n^2) \sigma_n(u_n, y) v_n$$

und schreiben auch formal $R_\alpha = F_\alpha(K^* K) K^*$.¹

¹Die Frage, für welche Funktionen dies mehr als nur formale Bedeutung hat ist Gegenstand des sog. *Funktionalkalküls* und wird uns in dieser Vorlesung nicht beschäftigen.

Dann ist $R_\alpha \in \mathcal{L}(Y, X)$ und

$$\begin{aligned}\|R_\alpha\|_{\mathcal{L}(Y, X)} &\leq \sup_{n \in \mathbb{N}} |F_\alpha(\sigma_n^2)| \|K^*\|_{\mathcal{L}(Y, X)}, \\ \|KR_\alpha\|_{\mathcal{L}(Y)} &\leq \sup_{n \in \mathbb{N}} |F_\alpha(\sigma_n^2)| \sigma_n^2\end{aligned}$$

Ist $(F_\alpha)_{\alpha>0}$ ein regularisierender Filter, so gilt

$$\lim_{\alpha \rightarrow 0} R_\alpha y = K^+ y \quad \text{für alle } y \in \mathcal{D}(K^+),$$

$(R_\alpha)_{\alpha>0}$ ist also eine Regularisierung für K^+ .

Beweis: Für jedes α ist F_α beschränkt und damit

$$\begin{aligned}\|R_\alpha y\|^2 &\leq \sup_{n \in \mathbb{N}} |F_\alpha(\sigma_n^2)|^2 \left\| \sum_{n=1}^{\infty} \sigma_n(u_n, y) v_n \right\|^2 \\ &\leq \sup_{n \in \mathbb{N}} |F_\alpha(\sigma_n^2)|^2 \|K^* y\|^2 \leq \sup_{n \in \mathbb{N}} |F_\alpha(\sigma_n^2)|^2 \|K^*\|^2 \|y\|^2.\end{aligned}$$

Es ist also $R_\alpha \in \mathcal{L}(Y, X)$ und $\|R_\alpha\|_{\mathcal{L}(Y, X)} \leq \sup_{n \in \mathbb{N}} |F_\alpha(\sigma_n^2)| \|K^*\|$. Außerdem ist

$$\begin{aligned}\|KR_\alpha y\|^2 &\leq \sup_{n \in \mathbb{N}} |F_\alpha(\sigma_n^2) \sigma_n^2|^2 \left\| \sum_{n=1}^{\infty} (u_n, y) u_n \right\|^2 \\ &\leq \sup_{n \in \mathbb{N}} |F_\alpha(\sigma_n^2) \sigma_n^2|^2 \|y\|^2,\end{aligned}$$

also $\|KR_\alpha\|_{\mathcal{L}(Y)} \leq \sup_{n \in \mathbb{N}} |F_\alpha(\sigma_n^2)| \sigma_n^2$.

Nun sei $y \in D(K^+)$. Mit Satz 2.32 ist $\frac{1}{\sigma_n}(u_n, y) = (v_n, K^+ y)$ und

$$\begin{aligned}\|(K^+ - R_\alpha)y\|^2 &= \left\| \sum_{n=1}^{\infty} (\sigma_n^{-1} - F_\alpha(\sigma_n^2) \sigma_n) (u_n, y) v_n \right\|^2 \\ &= \sum_{n=1}^{\infty} (\sigma_n^{-1} - F_\alpha(\sigma_n^2) \sigma_n)^2 (u_n, y)^2 \\ &= \sum_{n=1}^N (1 - F_\alpha(\sigma_n^2) \sigma_n^2)^2 (v_n, K^+ y)^2 + \sum_{n=N+1}^{\infty} (1 - F_\alpha(\sigma_n^2) \sigma_n^2)^2 (v_n, K^+ y)^2 \\ &\leq \sum_{n=1}^N (1 - F_\alpha(\sigma_n^2) \sigma_n^2)^2 (v_n, K^+ y)^2 + (1 + C)^2 \sum_{n=N+1}^{\infty} (v_n, K^+ y)^2,\end{aligned}$$

wobei $C > 0$ die Konstante aus der Definition regularisierender Filter ist. Der zweite Summand wird unabhängig von α beliebig klein für $N \rightarrow \infty$ und der erste Summand wird für festes N und $\alpha \rightarrow 0$ beliebig klein. Es folgt

$$R_\alpha y \rightarrow K^+ y,$$

so dass die Behauptung gezeigt ist. \square

3.1.3 Abgeschnittene SWZ und Tikhonov-Regularisierung

Beispiel 3.4

(a) Durch

$$F_\alpha(\lambda) := \begin{cases} \lambda^{-1} & : \lambda \geq \alpha \\ 0 & : \lambda < \alpha \end{cases}$$

ist offenbar ein regularisierender Filter definiert. Er führt zur abgeschnittenen Singulärwertzerlegung

$$R_\alpha y = \sum_{\sigma_n \geq \sqrt{\alpha}} \sigma_n^{-1}(u_n, y) v_n.$$

(b) Durch

$$F_\alpha(\lambda) := \frac{1}{\lambda + \alpha}$$

ist ebenfalls offenbar ein regularisierender Filter definiert. Dieser führt zur Tikhonov-Regularisierung

$$R_\alpha y = \sum_{n=1}^{\infty} \frac{\sigma_n}{\sigma_n^2 + \alpha} (u_n, y) v_n.$$

Bemerkung 3.5 (Motivation der Tikhonov-Regularisierung)

Ist K nicht degeneriert, dann ist K^+ nicht stetig. Es existiert also keine Konstante $C > 0$ mit

$$\|x\|_X \leq C \|Kx\|_Y \quad \forall x \in X.$$

Möglicherweise divergiert also $\|K^+ y^\delta\|_X \rightarrow \infty$ für $\delta \rightarrow 0$, d.h. für immer kleinere Messfehler konvergieren unsere naiven Lösungen i.A. nicht nur nicht, sondern können sogar beliebig groß werden!

Ein naheliegender praktischer Ansatz ist daher das Problem nur so gut es geht zu lösen, und gleichzeitig zu verhindern, dass die Norm der Lösung zu groß wird, d.h. zu y suchen wir den Minimierer $x_\alpha = R_\alpha y$ von

$$\|y - Kx\|_Y^2 + \alpha \|x\|_X^2 \rightarrow \min! \tag{3.1}$$

Der Regularisierungsparameter $\alpha > 0$ steuert dabei, ob die Lösung auf Kosten Ihrer Regularität (hier: ihrer $\|\cdot\|_X$ -Norm) besser zu den Daten passen soll ($\alpha \rightarrow 0$) oder umgekehrt ($\alpha \rightarrow \infty$).

Die Minimierungsaufgabe lässt sich auch schreiben als

$$\left\| \begin{pmatrix} K \\ \sqrt{\alpha}I \end{pmatrix} x - \begin{pmatrix} y \\ 0 \end{pmatrix} \right\|_{Y \times X}^2 = \left\| \begin{pmatrix} Kx - y \\ \sqrt{\alpha}x \end{pmatrix} \right\|_{Y \times X}^2 \rightarrow \min!$$

Gemäß Satz 2.12 löst der Minimierer x_α die Normalgleichungen

$$\begin{pmatrix} K^* & \sqrt{\alpha}I \end{pmatrix} \begin{pmatrix} K \\ \sqrt{\alpha}I \end{pmatrix} x_\alpha = \begin{pmatrix} K^* & \sqrt{\alpha}I \end{pmatrix} \begin{pmatrix} y \\ 0 \end{pmatrix},$$

also

$$(K^*K + \alpha I)x_\alpha = K^*y. \quad (3.2)$$

Wir verwenden die Singulärwertzerlegung von K und erhalten

$$x_\alpha = \sum_{n=1}^{\infty} (x_\alpha, v_n)v_n + x_0, \quad x_0 \in \mathcal{N}(K)$$

und

$$\mathcal{N}(K)^\perp \ni K^*y = \sum_{n=1}^{\infty} (K^*y, v_n)v_n = \sum_{n=1}^{\infty} \sigma_n(u_n, y)v_n.$$

x_α löst also

$$\sum_{n=1}^{\infty} (\sigma_n^2 + \alpha)(x_\alpha, v_n)v_n + \alpha x_0 = (K^*K + \alpha I)x_\alpha = K^*y = \sum_{n=1}^{\infty} \sigma_n(u_n, y)v_n$$

und durch Koeffizientenvergleich folgt

$$x_\alpha = \sum_{n=1}^{\infty} (x_\alpha, v_n)v_n = \sum_{n=1}^{\infty} \frac{\sigma_n}{\sigma_n^2 + \alpha} (u_n, y)v_n.$$

Für die Tikhonov-Regularisierung haben wir also die zwei äquivalenten Darstellungen

$$x_\alpha = \sum_{n=1}^{\infty} \frac{\sigma_n}{\sigma_n^2 + \alpha} (u_n, y)v_n \quad \text{und} \quad x_\alpha = (K^*K + \alpha I)^{-1}K^*y.$$

Die erste Darstellung besitzt den Vorteil, dass wir damit Aussagen für allgemeine Filter-basierte Verfahren auf die Tikhonov-Regularisierung übertragen können. Zur Implementierung verwendet man aber üblicherweise die zweite Darstellung, da diese keine Singulärwertzerlegung benötigt.

Bemerkung 3.6

Für die abgeschnittene Singulärwertzerlegung wissen wir schon aus Satz 2.32, dass

$$\|R_\alpha y\| \rightarrow \infty \quad \text{falls } y \notin \mathcal{D}(K^+).$$

Tatsächlich kann es keine Regularisierung geben für die $\|R_\alpha\|$ beschränkt bleibt. Um dies zu zeigen benötigen wir jedoch noch einige Hilfsmittel aus der Funktionalanalysis.

3.2 Schwache Konvergenz

3.2.1 Sätze von Baire, von der gleichmäßigen Beschränktheit und von Banach-Steinhaus

Satz 3.7

(a) Ist $X = \bigcup_{k \in \mathbb{N}} V_k$ mit abgeschlossenen Mengen V_k , so besitzt mindestens ein V_k nicht-leeres Inneres, d.h.

$$\exists k \in \mathbb{N}, x \in X, \varepsilon > 0 : B_\varepsilon(x) \subset V_k.$$

(Baire'scher Kategoriensatz)

(b) Ist $(A_n)_{n \in \mathbb{N}} \subset \mathcal{L}(X, Y)$ punktweise beschränkt, d.h.

$$\sup_{n \in \mathbb{N}} \|A_n x\| < \infty \quad \forall x \in X,$$

so ist A_n beschränkt, also $\sup_{n \in \mathbb{N}} \|A_n\|_{\mathcal{L}(X, Y)} < \infty$.

(Satz von der gleichmäßigen Beschränktheit)

Beweis: (a) Angenommen kein V_k enthält eine Kugel $B_\varepsilon(x)$. Zu $\varepsilon_0 := 1$ und $x_0 := 0$ ist dann $B_{\varepsilon_0}(x_0) \setminus V_1$ nichtleer und offen. Es existiert also $\varepsilon_1 < 1$, $x_1 \in B_{\varepsilon_0}(x_0)$ mit $B_{\varepsilon_1}(x_1) \subset B_{\varepsilon_0}(x_0) \setminus V_1$.

$B_{\varepsilon_1}(x_1) \setminus V_2$ ist wiederum nichtleer und offen, enthält also eine abgeschlossene Kugel $\overline{B_{\varepsilon_2}(x_2)}$ mit $0 < \varepsilon_2 < 1/2$, $x_2 \in B_{\varepsilon_1}(x_1)$. Wir fahren so fort und wählen dabei immer $\varepsilon_n < 1/n$. Dies ergibt eine Cauchy-Folge $(x_n)_{n \in \mathbb{N}}$, deren Grenzwert $x = \lim_{n \rightarrow \infty} x_n$ in jedem $\overline{B_{\varepsilon_k}(x_k)}$ und damit in keinem V_k liegt. Dies widerspricht $X = \bigcup_{k \in \mathbb{N}} V_k$.

(b) Setze

$$V_k := \{x \in X : \|A_n x\| \leq k \forall n \in \mathbb{N}\}.$$

Die V_k sind offensichtlich abgeschlossen und aus der punktweise Beschränktheit folgt $X = \bigcup_{k \in \mathbb{N}} V_k$. Nach (a) existiert ein K und eine Kugel $B_\varepsilon(x_0)$ mit $B_\varepsilon(x_0) \subset V_K$, also

$$\|A_n(x+x_0)\| \leq K \quad \forall x \in X, \|x\| < \varepsilon, n \in \mathbb{N}.$$

Aus der Linearität von A_n folgt

$$\|A_n x\| \leq \|A_n(x+x_0)\| + \|A_n x_0\| \leq 2K \quad \forall x \in X, \|x\| < \varepsilon, n \in \mathbb{N},$$

und damit

$$\|A_n\| \leq \frac{2K}{\varepsilon} \quad \forall n \in \mathbb{N},$$

womit die Behauptung gezeigt ist. \square

Satz 3.8 (Banach-Steinhaus)

Sei $(A_n)_{n \in \mathbb{N}} \subset \mathcal{L}(X, Y)$.

(A_n) konvergiert genau dann punktweise gegen ein $A : X \rightarrow Y$, wenn die folgenden beiden Eigenschaften erfüllt sind:

(a) $\exists C > 0 : \|A_n\| \leq C \quad \forall n \in \mathbb{N}.$

(b) $(A_n x)_{n \in \mathbb{N}}$ konvergiert für alle x in einer dichten Teilmenge $V \subseteq X$.

Außerdem ist dann $A \in \mathcal{L}(X, Y)$.

Beweis: „ \Leftarrow “: Sei $x \in X$. Wir zeigen, dass $(A_n x)_{n \in \mathbb{N}}$ eine Cauchy-Folge ist. Für alle $v \in V$ gilt

$$\begin{aligned} \|A_n x - A_m x\| &\leq \|A_n x - A_n v\| + \|A_n v - A_m v\| + \|A_m v - A_m x\| \\ &\leq 2C \|x - v\| + \|A_n v - A_m v\|. \end{aligned}$$

Der erste Summand wird unabhängig von n, m beliebig klein für geeignetes v , der zweite Summand konvergiert für festes v und $n, m \rightarrow \infty$ gegen 0. $(A_n x)_{n \in \mathbb{N}}$ ist also eine Cauchy-Folge und damit konvergent.

Wir können daher $Ax := \lim_{n \rightarrow \infty} A_n x$ für alle $x \in X$ definieren. Die Linearität von A ist klar. Die Stetigkeit folgt aus

$$\|Ax\| = \lim_{n \rightarrow \infty} \|A_n x\| \leq C \|x\|.$$

„ \Rightarrow “: Konvergiert A_n punktweise auf ganz X , so ist es insbesondere punktweise beschränkt, sodass (a) aus Satz 3.7(b) folgt. (b) ist trivial. \square

3.2.2 Schwache Konvergenz

Definition 3.9

Eine Folge $(x_n)_{n \in \mathbb{N}} \subset X$ konvergiert schwach gegen $x \in X$, falls

$$(x_n, \xi) \rightarrow (x, \xi) \quad \text{für alle } \xi \in X.$$

Wir schreiben auch $x_n \rightharpoonup x$.

Offenbar ist der schwache Grenzwert eindeutig und die schwache Grenzwertbildung vertauscht mit den Vektorraumoperationen.

Bemerkung 3.10

Der Begriff Konvergenz ist in der Mathematik festgelegt. Ihn einfach umzudefinieren birgt die Gefahr vieler Missverständnisse. So ist etwa die Definition

$$(a_k)_{k \in \mathbb{N}} \subset \mathbb{R} \quad \text{konvergiert gegen } a \in \mathbb{R}, \text{ falls } a_1 = a.$$

ziemlicher Unsinn und besitzt sicher nicht die bekannten und intuitiv erwarteten Eigenschaften eines Grenzwertbegriffs.

Von Konvergenz sollte man deshalb nur sprechen, wenn es eine Norm, Metrik oder zumindest Topologie gibt, die diese Konvergenz induziert. Wir werden die zu Definition 3.9 gehörige Topologie hier weder explizit angeben noch verwenden. Es ist aber wichtig festzustellen, dass es sie gibt.

Satz 3.11

(a) Es gibt eine Topologie auf X , die die Konvergenz aus Definition 3.9 induziert und die X zu einem topologischen Vektorraum macht, d.h. in dieser Topologie sind die Vektorraumoperationen stetig und einelementige Mengen sind abgeschlossen.

(b) Für $(x_n)_{n \in \mathbb{N}} \subset X$ folgt aus $x_n \rightarrow x$ auch $x_n \rightharpoonup x$.

(c) Schwach konvergente Folgen $(x_n)_{n \in \mathbb{N}} \subset X$ sind beschränkt.

(d) Ist $A \in \mathcal{L}(X, Y)$ und $(x_n)_{n \in \mathbb{N}} \subset X$, $x_n \rightharpoonup x \in X$, so gilt $Ax_n \rightharpoonup Ax \in Y$.

(e) Ist $K \in \mathcal{K}(X, Y)$ und $(x_n)_{n \in \mathbb{N}} \subset X$, $x_n \rightharpoonup x \in X$, so gilt $Kx_n \rightarrow Kx$.

(f) Gilt für eine schwach konvergente Folge

$$(x_n)_{n \in \mathbb{N}} \subset X, \quad x_n \rightharpoonup x,$$

zusätzlich $\|x_n\| \leq \|x\|$ für alle $n \in \mathbb{N}$, so gilt sogar $x_n \rightarrow x$.

Beweis: (a) Hierzu verweisen wir auf [Rudin, Chp. 3].

(b) ist trivial.

(c) Für die Abbildungen

$$A_n \in \mathcal{L}(X, \mathbb{R}), \quad A_n \xi := (x_n, \xi)$$

gilt offenbar $\|A_n\|_{\mathcal{L}(X, \mathbb{R})} = \|x_n\|_X$. (Die Zuordnung $A_n \leftrightarrow x_n$ ist gerade der Riesz-Isomorphismus aus Bemerkung 2.4.) Schwache Konvergenz der Folge $(x_n)_{n \in \mathbb{N}}$ ist äquivalent zur punktweisen Konvergenz der $(A_n)_{n \in \mathbb{N}}$, sodass die Beschränktheit aus Satz 3.8 folgt.

(d) Für alle $y \in Y$ gilt

$$(Ax_n, y) = (x_n, A^*y) \rightarrow (x, A^*y) = (Ax, y),$$

also $Ax_n \rightharpoonup Ax$.

(e) Nach (d) konvergiert $Kx_n \rightharpoonup Kx$. Da nach (c) die Folge $(x_n)_{n \in \mathbb{N}}$ beschränkt ist, konvergiert eine Teilfolge von jeder Teilfolge von $(Kx_n)_{n \in \mathbb{N}}$. Aus (b) und der Eindeutigkeit des schwachen Limes folgt, dass eine Teilfolge jeder Teilfolge gegen Kx konvergiert. Außerhalb jeder Umgebung von Kx können also nur endlich viele Kx_n liegen, sodass $y = \lim_{n \rightarrow \infty} Kx_n$ folgt.

(f) Es ist

$$\limsup_{n \rightarrow \infty} \|x - x_n\|^2 = \limsup_{n \rightarrow \infty} \left(\|x\|^2 - 2(x, x_n) + \|x_n\|^2 \right) \leq 0.$$

Satz 3.12

In einem Hilbertraum X besitzt jede beschränkte Folge eine schwach konvergente Teilfolge.

Beweis: Sei $(x_k)_{k \in \mathbb{N}} \subset X$ beschränkt. Nach Satz 2.5 ist $V := \overline{\langle x_1, x_2, \dots \rangle}$ ein Hilbertraum.

Für jedes $k \in \mathbb{N}$ ist $(x_1, x_k), (x_2, x_k), \dots$ eine beschränkte Folge in \mathbb{R} . Nach dem Satz von Bolzano-Weierstrass existiert also eine Teilfolge $(x_{1,n})_{n \in \mathbb{N}}$ von $(x_k)_{k \in \mathbb{N}}$, sodass $(x_{1,n}, x_1)$ konvergiert und davon wiederum eine Teil-Teilfolge $(x_{2,n})_{n \in \mathbb{N}}$, sodass $(x_{2,n}, x_2)$ konvergiert, u.s.w.

Für die Diagonalfolge $(v_n)_{n \in \mathbb{N}} := (x_{n,n})_{n \in \mathbb{N}}$ gilt also, dass (v_n, x_k) für jedes k konvergiert. Wir definieren

$$A_n \in \mathcal{L}(V, \mathbb{R}) = V', \quad A_n : x \mapsto (v_n, x),$$

KAPITEL 3. REGULARISIERUNG LINEARER PROBLEME

dann folgt $\|A_n\|_{V'} = \|v_n\|_V \leq \sup_{k \in \mathbb{N}} \|x_k\|_X < \infty$ und aufgrund der Linearität konvergiert $A_n x$ für alle $x \in \langle x_1, x_2, \dots \rangle$.

Aus Satz 3.8 folgt, dass ein $A \in \mathcal{L}(V, \mathbb{R})$ existiert mit

$$Ax = \lim_{n \rightarrow \infty} A_n x \quad \forall x \in V$$

und nach dem Satz von Riesz aus Bemerkung 2.4 existiert $v \in V$ mit $(v, x) = Ax$ für alle $x \in V$.

Schließlich ist für alle $x \in X$, $P_V x \in V$ und $x - P_V x \in V^\perp$, also

$$\begin{aligned} (v_n, x) &= (v_n, P_V x) + (v_n, x - P_V x) = (v_n, P_V x) = A_n P_V x \\ &\rightarrow A P_V x = (v, P_V x) = (v, P_V x) + (v, x - P_V x) = (v, x). \end{aligned}$$

Die Teilfolge $(v_n)_{n \in \mathbb{N}}$ von $(x_k)_{k \in \mathbb{N}}$ konvergiert also schwach gegen $v \in V \subseteq X$. \square

Folgerung 3.13

Ist $(x_n)_{n \in \mathbb{N}} \subset X$ beschränkt, so besitzt $(Kx_n)_{n \in \mathbb{N}}$ eine konvergente Teilfolge, deren Grenzwert in $\mathcal{R}(K)$ liegt.

Beweis: Nach Satz 3.12 existiert eine Teilfolge $(x_{n_k})_{k \in \mathbb{N}}$ mit $x_{n_k} \rightharpoonup x$ und aus Satz 3.11(e) folgt $\lim_{k \rightarrow \infty} Kx_{n_k} = Kx$ \square

3.3 Regularisierungsverfahren

Es seien weiterhin X, Y Hilberträume und $K \in \mathcal{K}(X, Y)$ sei nicht-degeneriert mit Singulärwertzerlegung $(u_n, v_n, \sigma_n)_{n \in \mathbb{N}}$.

3.3.1 Parameterwahlstrategien

Satz 3.14

Ist $(R_\alpha)_{\alpha > 0}$ eine Regularisierung von K^+ , dann gilt

(a) $\|R_\alpha\|_{\mathcal{L}(Y, X)} \rightarrow \infty$ für $\alpha \rightarrow 0$.

(b) Ist zudem $\|KR_\alpha\|_{\mathcal{L}(Y)}$ glm. beschränkt in α , dann gilt für alle $y \in \mathcal{D}(K^+)$

$$\|R_\alpha y\|_X \rightarrow \infty \quad \text{für } \alpha \rightarrow 0.$$

Nach Satz 3.3 gilt (b) für regularisierende Filter.

3.3. REGULARISIERUNGSVERFAHREN

Beweis: (a) Angenommen es existiert eine (glm. in α) beschränkte Teilfolge von R_α . Dann wäre K^+ stetig nach dem Satz 3.8 (Banach-Steinhaus). Dann wäre aber nach Korollar 2.33 K degeneriert.

(b) Angenommen es existiert eine beschränkte Teilfolge $\|R_\alpha y\|_X$. Nach Korollar 3.13 existiert dann eine Teilfolge von $KR_\alpha y$, die gegen ein $\eta \in \mathcal{R}(K)$ konvergiert.

Außerdem konvergiert KR_α punktweise auf $\mathcal{D}(K^+)$ gegen $KK^+ = P_{\overline{\mathcal{R}(K)}}$. Aus der Beschränktheit von KR_α folgt mit Satz 3.8 (Banach-Steinhaus), auch $KR_\alpha y \rightarrow P_{\overline{\mathcal{R}(K)}} y$. Es gilt also $P_{\overline{\mathcal{R}(K)}} y = \eta \in \mathcal{R}(K)$ und damit

$$y \in \mathcal{R}(K) + \mathcal{R}(K)^\perp = \mathcal{D}(K^+),$$

so dass die Behauptung durch Kontraposition gezeigt ist. □

Bemerkung 3.15

Für die Abweichung von $x_\alpha^\delta := R_\alpha y^\delta$ vom gesuchten $K^+ y$ gilt

$$\begin{aligned} \|x_\alpha^\delta - K^+ y\| &\leq \|x_\alpha^\delta - R_\alpha y\| + \|R_\alpha y - K^+ y\| \\ &= \|R_\alpha(y^\delta - y)\| + \|R_\alpha y - K^+ y\| \end{aligned}$$

Der erste Summand $R_\alpha(y^\delta - y)$ beschreibt den fortgepflanzten Datenfehler. Gemäß Satz 3.14 konvergiert er (für $y^\delta \notin \mathcal{D}(K^+)$) für $\alpha \rightarrow 0$ gegen unendlich. Der zweite Summand beschreibt den durch die Regularisierung eingeführten Verfahrensfehler. Er konvergiert für $\alpha \rightarrow 0$ gegen Null.

Das Parameterwahlproblem besteht darin, α (in Abhängigkeit von δ und y^δ) so zu wählen, dass der Gesamtfehler $\|x_\alpha^\delta - K^+ y\|$ möglichst klein wird und für $\delta \rightarrow 0$ gegen Null konvergiert.

Definition 3.16

Eine Funktion

$$\alpha : \mathbb{R}^+ \times Y \rightarrow \mathbb{R}^+, \quad (\delta, y^\delta) \rightarrow \alpha(\delta, y^\delta)$$

heißt Parameterwahlstrategie.²

Eine Kombination von Regularisierung und Parameterwahlstrategie heißt Regularisierungsverfahren, falls für jedes $y \in \mathcal{D}(K^+)$

$$R_{\alpha(y^\delta, \delta)} y^\delta \rightarrow K^+ y \quad (\delta \rightarrow 0)$$

²Analog heißt für Regularisierungen $(R_k)_{k \in \mathbb{N}}$ eine Funktion $k(\delta, y^\delta)$ Parameterwahlstrategie oder auch Stoppregel.

für alle $(y^\delta)_{\delta>0} \subseteq Y$ mit $\|y^\delta - y\| \leq \delta$.

Hängt α nur von δ ab, so sprechen wir von einer a priori Parameterwahl sonst von einer a posteriori Parameterwahl.

Satz 3.17

Sei $(R_\alpha)_{\alpha>0}$ eine Regularisierung. Dann erzeugt jede Parameterwahlstrategie $\alpha = \alpha(\delta)$ mit

$$\alpha(\delta) \rightarrow 0, \quad \|R_\alpha\| \delta \rightarrow 0 \quad (\text{für } \delta \rightarrow 0)$$

ein Regularisierungsverfahren.

Beweis: Wie in Bemerkung 3.15 gilt für $y \in \mathcal{D}(K^+)$

$$\|x_\alpha^\delta - K^+y\| \leq \underbrace{\|R_\alpha\| \|y^\delta - y\|}_{\leq \|R_\alpha\| \delta \rightarrow 0} + \underbrace{\|R_\alpha y - K^+y\|}_{\rightarrow 0 \text{ für } \alpha \rightarrow 0},$$

womit die Behauptung folgt. □

Für allgemeine Filter haben wir $\|R_\alpha\|$ schon in Satz 3.3 abgeschätzt, für regularisierende Filter lässt sich dies noch verschärfen:

Satz 3.18

Für regularisierende Filter $(F_\alpha)_{\alpha>0}$ ist

$$\|R_\alpha\| \leq C \sup_{n \in \mathbb{N}} \sqrt{|F_\alpha(\sigma_n^2)|} \leq C \sup_{\lambda \in (0, \|K\|^2]} \sqrt{|F_\alpha(\lambda)|}$$

mit $C = \sup_{\alpha>0, \lambda \in (0, \|K\|^2]} \sqrt{\lambda |F_\alpha(\lambda)|}$.

Beweis: Es ist

$$\begin{aligned} \|R_\alpha y\|^2 &= \left(R_\alpha y, \sum_{n=1}^{\infty} F_\alpha(\sigma_n^2) \sigma_n(u_n, y) v_n \right) = \sum_{n=1}^{\infty} F_\alpha(\sigma_n^2) \sigma_n(u_n, y) (R_\alpha y, v_n) \\ &= \sum_{n=1}^{\infty} F_\alpha(\sigma_n^2) (u_n, y) (KR_\alpha y, u_n) \leq \sup_{n \in \mathbb{N}} |F_\alpha(\sigma_n^2)| \|KR_\alpha\| \|y\|^2. \end{aligned}$$

und aus Satz 3.3 folgt $\|KR_\alpha\| \leq C^2$. □

Beispiel 3.19

(a) Für die abgeschnittene Singulärwertzerlegung

$$F_\alpha(\lambda) := \begin{cases} \lambda^{-1} & : \lambda \geq \alpha \\ 0 & : \lambda < \alpha \end{cases}$$

3.3. REGULARISIERUNGSVERFAHREN

ist offenbar $C = 1$ und damit $\|R_\alpha\| \leq \frac{1}{\sqrt{\alpha}}$. Also erzeugt jede Parameterwahlstrategie $\alpha(\delta)$ mit

$$\alpha(\delta) \rightarrow 0, \quad \frac{\delta}{\sqrt{\alpha(\delta)}} \rightarrow 0$$

ein Regularisierungsverfahren.

Mit der Wahl $\alpha(\delta) := \delta$ folgt also

$$R_{\alpha(\delta)}y^\delta = \sum_{\sigma_k \geq \sqrt{\delta}} \sigma_k^{-1}(u_k, y^\delta)v_k \rightarrow K^+y \quad \text{für } \delta \rightarrow 0.$$

(b) Für die Tikhonov-Regularisierung

$$F_\alpha(\lambda) := \frac{1}{\lambda + \alpha}$$

gilt offenbar ebenfalls $C = 1$ und $\|R_\alpha\| \leq \frac{1}{\sqrt{\alpha}}$, sodass jede Parameterwahl mit

$$\alpha(\delta) \rightarrow 0 \quad \text{und} \quad \frac{\delta^2}{\alpha(\delta)} \rightarrow 0,$$

zu einem Regularisierungsverfahren führt.

Mit der Wahl $\alpha(\delta) := \delta$ folgt also

$$R_{\alpha(\delta)}y^\delta = (K^*K + \delta I)^{-1}K^*y^\delta \rightarrow K^+y \quad \text{für } \delta \rightarrow 0.$$

3.3.2 Das Bakushinskii-Veto

Satz 3.20 (Bakushinskii-Veto)

Sei $(R_\alpha)_{\alpha>0}$ eine Regularisierung. Es existiert keine Parameterwahl $\alpha = \alpha(y^\delta)$, sodass für alle $y \in \mathcal{D}(K^+)$

$$R_{\alpha(y^\delta)}y^\delta \rightarrow K^+y \quad \text{für } \delta \rightarrow 0. \quad (3.3)$$

Beweis: Angenommen, es gibt eine solche Parameterwahl. Dann definieren wir uns damit die (möglicherweise nicht-lineare) Abbildung

$$R: Y \rightarrow X, \quad R(\eta) := R_{\alpha(\eta)}\eta.$$

Für $y \in \mathcal{D}(K^+)$ folgt mit $y^\delta := y$ aus (3.3), dass $Ry = K^+y$, d.h. $R|_{\mathcal{D}(K^+)} = K^+$. Außerdem folgt aus (3.3) für jede Folge $y^\delta \rightarrow y$, dass

$$R(y^\delta) = R_{\alpha(y^\delta)}y^\delta \rightarrow K^+y = R(y),$$

d.h. R (und damit K^+) ist stetig auf $\mathcal{D}(K^+)$. Dies widerspricht Korollar 2.33. \square

Bemerkung 3.21

- (a) Satz 3.20 zeigt dass die Kenntnis des Datenfehlers notwendig ist für eine konvergente Parameterwahlstrategie. In der Praxis ist dieser jedoch nicht immer bestimmbar und so haben sich neben den mathematisch rigoros gerechtfertigten auch heuristische Parameterwahlstrategie etabliert. Ein Beispiel hierfür ist das sogenannte L-Kurven-Kriterium, in dem $\|x_\alpha^\delta\|$ als Funktion von $\|y^\delta - Kx_\alpha^\delta\|$ (doppeltlogarithmisch) geplottet wird und der zum linken „Eckpunkt“ gehörige Parameter α verwendet wird (vgl. das in der Vorlesung gemalte Bild).
- (b) In der Praxis werden Messungen oft wiederholt durchgeführt um aus der Streuung den Datenfehler schätzen zu können und ihn durch Mittelung zu reduzieren. Mit diesem Ansatz lässt sich unter geeigneten Annahmen tatsächlich das Bakushinskii-Veto umgehen und (in einem stochastischen Sinne) Konvergenz beweisen [HJP20].

3.4 Diskrepanzprinzip für das Tikhonov-Verfahren

In diesem Abschnitt besitze K stets dichtes Bild, d.h. $\overline{\mathcal{R}(K)} = Y$, und wir betrachten nur die Tikhonov-Regularisierung

$$R_\alpha = (K^*K + \alpha I)^{-1}K^*.$$

Sei $y^\delta \in Y$, $\delta > 0$ mit $\|y^\delta - y\| \leq \delta$. Der Parameter $\alpha > 0$ steuert (wie in Bemerkung 3.5 beschrieben), wie gut $x_\alpha^\delta := R_\alpha y^\delta$ zu den gemessenen Daten passt. Um so kleiner α gewählt wird, umso kleiner wird das Residuum

$$\|Kx_\alpha^\delta - y^\delta\|.$$

Dabei erscheint es sinnlos, hier eine höhere Genauigkeit als δ zu fordern. Die PWS

$$\text{„Wähle } \alpha(y^\delta, \delta) \text{ sodass } \|Kx_{\alpha(y^\delta, \delta)}^\delta - y^\delta\| = \delta\text{“}$$

heißt *Diskrepanzprinzip von Morozov*.

Die Existenz eines solches α kann nur dann garantiert werden, wenn K dichtes Bild besitzt. Im allgemeinen Fall $\mathcal{D}(K^+) = \mathcal{R}(K) + \mathcal{R}(K)^\perp$ mit $\mathcal{R}(K)^\perp \neq \{0\}$ könnte nämlich selbst für exakte Daten und exakte Inversion das Residuum

$$K(K^+y) - y \stackrel{\text{Satz 2.16}}{=} P_{\overline{\mathcal{R}(K)}}y - y \in \mathcal{R}(K)^\perp$$

3.4. DISKREPANZPRINZIP FÜR DAS TIKHONOV-VERFAHREN

beliebig groß werden.

Falls K dichtes Bild besitzt (was wir in diesem Abschnitt stets voraussetzen), so existiert jedoch immer ein Parameter $\alpha(y^\delta, \delta)$, der das Diskrepanzprinzip erfüllt, und mit dieser Parameterwahl ergibt sich ein Regularisierungsverfahren, also $R_{\alpha(y^\delta, \delta)} y^\delta \rightarrow K^+ y$.

Um dies zu zeigen, beweisen wir zunächst die in Bemerkung 3.5 gemachten Aussagen.

Lemma 3.22

Sei $\alpha > 0$ und $y^\delta \in Y$.

$K^*K + \alpha I \in \mathcal{L}(X)$ ist stetig invertierbar und

$$x_\alpha^\delta := (K^*K + \alpha I)^{-1} K^* y^\delta = R_\alpha y^\delta = \sum_{n \in \mathbb{N}} \frac{\sigma_n}{\sigma_n^2 + \alpha} (u_n, y^\delta) v_n$$

ist der eindeutige Minimierer von

$$\|y^\delta - Kx\|_Y^2 + \alpha \|x\|_X^2 \rightarrow \min!$$

Beweis: Wegen

$$(K^*K + \alpha I)x = z \iff ((K^*K + \alpha I)x, \xi)_X = (z, \xi)_X \quad \forall \xi \in X.$$

folgt die stetige Invertierbarkeit von $K^*K + \alpha I$ aus dem Satz von Lax-Milgram (Satz 2.3). Die Singulärwertdarstellung von $(K^*K + \alpha I)^{-1}$ folgt sofort aus der von K .

Um die Minimiereigenschaft zu zeigen, sei $x \in X$ mit $x \neq x_\alpha^\delta$. Dann gilt

$$\begin{aligned} & \|y^\delta - Kx\|_Y^2 + \alpha \|x\|_X^2 - \|y^\delta - Kx_\alpha^\delta\|_Y^2 + \alpha \|x_\alpha^\delta\|_X^2 \\ &= -2(y^\delta, Kx)_Y + (x, (K^*K + \alpha I)x)_X + 2(y^\delta, Kx_\alpha^\delta)_Y \\ &\quad - (x_\alpha^\delta, (K^*K + \alpha I)x_\alpha^\delta)_X \\ &= \left((x - x_\alpha^\delta), (K^*K + \alpha I)(x - x_\alpha^\delta) \right)_X \\ &\quad + 2 \left(x, (K^*K + \alpha I)x_\alpha^\delta \right)_X - 2(y^\delta, Kx)_Y \\ &\quad - 2 \left(x_\alpha^\delta, (K^*K + \alpha I)x_\alpha^\delta \right)_X + 2(y^\delta, Kx_\alpha^\delta)_Y \\ &= \|K(x - x_\alpha^\delta)\|_Y^2 + \alpha \|x - x_\alpha^\delta\|_X^2 > 0. \end{aligned}$$

Lemma 3.23

Sei $y^\delta \in Y$.

(a) Die Funktion

$$\alpha \mapsto \left\| Kx_\alpha^\delta - y^\delta \right\|, \quad \text{mit } x_\alpha^\delta := R_\alpha y^\delta,$$

ist stetig und streng monoton steigend.

(b) Es gilt

$$\lim_{\alpha \rightarrow 0} \left\| Kx_\alpha^\delta - y^\delta \right\| = 0 \quad \text{und} \quad \lim_{\alpha \rightarrow \infty} \left\| Kx_\alpha^\delta - y^\delta \right\| = \left\| y^\delta \right\|.$$

Beweis: Das Erzeugnis $\langle u_1, u_2, \dots \rangle$ liegt offensichtlich dicht in $\mathcal{R}(K)$ und damit auch in $\overline{\mathcal{R}(K)} = Y$. Nach Definition und Satz 2.28 ist daher

$$y^\delta = \sum_{n \in \mathbb{N}} (u_n, y^\delta) u_n$$

und damit

$$Kx_\alpha^\delta - y^\delta = \sum_{n \in \mathbb{N}} \left(\frac{\sigma_n^2}{\sigma_n^2 + \alpha} - 1 \right) (u_n, y^\delta) u_n = \sum_{n \in \mathbb{N}} \frac{-\alpha}{\sigma_n^2 + \alpha} (u_n, y^\delta) u_n. \quad (3.4)$$

(a) Aus (3.4) folgt dass

$$\left\| Kx_\alpha^\delta - y^\delta \right\|^2 = \sum_{n \in \mathbb{N}} \left(\frac{1}{\frac{\sigma_n^2}{\alpha} + 1} \right)^2 (u_n, y^\delta)^2$$

und dieses ist offenbar streng monoton steigend in α .

Außerdem ist für $\alpha_1, \alpha_2 > 0$

$$\begin{aligned} \left\| x_{\alpha_1}^\delta - x_{\alpha_2}^\delta \right\|^2 &= \left\| \sum_{n=1}^{\infty} \left(\frac{\sigma_n}{\sigma_n^2 + \alpha_1} - \frac{\sigma_n}{\sigma_n^2 + \alpha_2} \right) (u_n, y^\delta) v_n \right\|^2 \\ &\leq \sup_{n \in \mathbb{N}} \left(\frac{\sigma_n}{\sigma_n^2 + \alpha_1} - \frac{\sigma_n}{\sigma_n^2 + \alpha_2} \right)^2 \left\| y^\delta \right\|^2 \\ &\leq (\alpha_2 - \alpha_1)^2 \frac{\sigma_1^2}{\alpha_1^2 \alpha_2^2} \left\| y^\delta \right\|^2. \end{aligned}$$

x_α^δ und damit insbesondere auch $\left\| Kx_\alpha^\delta - y^\delta \right\|$ hängt also stetig von $\alpha > 0$ ab.

3.4. DISKREPANZPRINZIP FÜR DAS TIKHONOV-VERFAHREN

(b) Wir verwenden wieder (3.4). Für $\alpha \rightarrow 0$ wird in

$$\begin{aligned} \|Kx_\alpha^\delta - y^\delta\|^2 &= \sum_{n=1}^N \left(\frac{\alpha}{\sigma_n^2 + \alpha} \right)^2 (u_n, y^\delta)^2 + \sum_{n=N+1}^{\infty} \left(\frac{\alpha}{\sigma_n^2 + \alpha} \right)^2 (u_n, y^\delta)^2 \\ &\leq \sum_{n=1}^N \left(\frac{\alpha}{\sigma_n^2} \right)^2 (u_n, y^\delta)^2 + \sum_{n=N+1}^{\infty} (u_n, y^\delta)^2 \\ &\leq \left(\frac{\alpha}{\sigma_N^2} \right)^2 \|y^\delta\|^2 + \sum_{n=N+1}^{\infty} (u_n, y^\delta)^2 \end{aligned}$$

der zweite Summand (unabhängig von $\alpha > 0$) für hinreichend große $N \in \mathbb{N}$ beliebig klein und der erste Summand wird für festes N und hinreichend kleines $\alpha > 0$ beliebig klein. Damit folgt

$$\lim_{\alpha \rightarrow 0} \|Kx_\alpha^\delta - y^\delta\| = 0.$$

Für $\alpha \rightarrow \infty$ gilt

$$\|Kx_\alpha^\delta\|^2 = \sum_{n \in \mathbb{N}} \left(\frac{\sigma_n^2}{\sigma_n^2 + \alpha} \right)^2 (u_n, y^\delta)^2 \leq \left(\frac{\sigma_1^2}{\alpha} \right)^2 \|y^\delta\|^2 \rightarrow 0,$$

und damit $\lim_{\alpha \rightarrow \infty} \|Kx_\alpha^\delta - y^\delta\| = \|y^\delta\|$. □

Satz 3.24

(a) Zu jedem $y^\delta \in Y$ und $0 < \delta \leq \|y^\delta\|$ existiert genau ein $\alpha(y^\delta, \delta)$ mit

$$\|Kx_{\alpha(y^\delta, \delta)}^\delta - y^\delta\| = \delta.$$

(b) Die Tikhonov-Regularisierung bildet mit dieser Parameterwahlstrategie³ ein Regularisierungsverfahren, d.h. für jedes $y \in \mathcal{D}(K^+)$ gilt

$$R_{\alpha(y^\delta, \delta)} y^\delta \rightarrow K^+ y$$

für jede Folge $(y^\delta)_{\delta > 0} \subseteq Y$ mit $\|y^\delta - y\| \leq \delta$.

³Strenggenommen ist diese Parameterwahlstrategie nur für $y \neq 0$ und hinreichend kleine $\delta > 0$ wohldefiniert. Eine überall definiertes Regularisierungsverfahren erhält man, indem für $\delta > \|y^\delta\|$, $\alpha := \infty$, also $x_{\alpha(y^\delta, \delta)}^\delta := 0$ gesetzt wird.

Beweis: (a) folgt aus Lemma 3.23

Zum Beweis von (b) beachte zunächst, dass für jede Folge $\delta \rightarrow 0$ aufgrund unserer Parameterwahl und der Minimalitätseigenschaft aus Lemma 3.22 gilt, dass

$$\begin{aligned} \delta^2 + \alpha(y^\delta, \delta) \left\| x_{\alpha(y^\delta, \delta)}^\delta \right\|^2 &= \left\| y^\delta - Kx_{\alpha(y^\delta, \delta)}^\delta \right\|^2 + \alpha(y^\delta, \delta) \left\| x_{\alpha(y^\delta, \delta)}^\delta \right\|^2 \\ &\leq \left\| y^\delta - KK^+y \right\|^2 + \alpha(y^\delta, \delta) \|K^+y\|^2 \\ &\leq \delta^2 + \alpha(y^\delta, \delta) \|K^+y\|^2. \end{aligned}$$

Dabei haben wir im letzten Schritt ausgenutzt, dass $KK^+y = P_{\overline{\mathcal{R}(K)}}y = y$ wegen $\overline{\mathcal{R}(K)} = Y$ gilt (vgl. Satz 2.16). Für alle $\delta > 0$ (und zugehörige y^δ mit $\|y^\delta - y\| \leq \delta$) ist also

$$\left\| x_{\alpha(y^\delta, \delta)}^\delta \right\| \leq \|K^+y\|. \quad (3.5)$$

Nach Satz 3.12 besitzt jede Teilfolge von $x_{\alpha(y^\delta, \delta)}^\delta$ eine schwach konvergente Teilfolge. Sei \hat{x} ein dazugehöriger schwacher Grenzwert. Aus dem Diskrepanzprinzip und Satz 3.11(e) folgt $\|K\hat{x} - y\| = 0$. Da $\|K^+y\|$ die Lösung mit minimaler Norm ist (Definition 2.14) ist $\|K^+y\| \leq \|\hat{x}\|$ und mit (3.5) folgt aus Satz 3.11(f), dass \hat{x} sogar starker Grenzwert der betrachteten Teilfolge ist. Aus (3.5) folgt dann aber $\|\hat{x}\| \leq \|K^+y\|$ und somit $\hat{x} = K^+y$. Insgesamt ist also gezeigt, dass jede Teilfolge von $x_{\alpha(y^\delta, \delta)}^\delta$ eine gegen K^+y konvergente Teilfolge besitzt. Es gilt also

$$x_{\alpha(y^\delta, \delta)}^\delta \rightarrow K^+y.$$

Literaturverzeichnis

- [Alt] H. W. Alt. *Lineare Funktionalanalysis*. 3. Auflage. Springer, Berlin, 1999.
- [Forster3] O. Forster: *Analysis 3. Integralrechnung im \mathbb{R}^n mit Anwendungen*. 3. Auflage, Vieweg, Braunschweig, 1996.
- [HJP20] B. Harrach, T. Jahn, R. Potthast: Beyond the Bakushinkii veto: regularising linear inverse problems without knowing the noise distribution, *Numer. Math.* **145**(3), 581–603, 2020.
- [Rieder] A. Rieder. *Keine Probleme mit Inversen Problemen*. Vieweg, Wiesbaden, 2003.
- [Rudin] W. Rudin: *Functional analysis. Second edition*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, 1991.