SCHLECHT GESTELLTE PROBLEME

FRANK NATTERER

Vorlesungsskript

im

Sommersemester 1986

WWU Münster

INHALTSVERZEICHNIS

0. EINLEITUNG

Eine Aufgabe heißt gut gestellt, wenn ihre Lösung durch die Daten wohlbestimmt ist und stetig von den Daten abhängt. Andernfalls heißt die Aufgabe schlecht gestellt.

Es wird ein Überblick über das Auftreten schlecht gestellter Probleme gegeben, sowie eine einheitliche mathematische Theorie und numerische Verfahren vorgestellt.

0.1. Beispiel eines gut gestellten Problems

Wir betrachten folgendes (Dirichletsches) Randwertproblem:

Gesucht ist eine Funktion f auf einem Gebiet Ω mit folgenden Eigenschaften:

i) $\Delta f = 0$ (d.h. f harmonisch)

ii) $f_{|\partial\Omega} = g$ (g ist die "Datenfunktion") .

Sei speziell $\Omega = D_1(0)$ der Kreis um 0 mit Radius 1, wir führen Polarkoordinaten

$$x_1 = r \cos \varphi$$
$$x_2 = r \sin \varphi$$

ein und schreiben den Laplace-Operator entsprechend um:

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \varphi^2} \quad .$$

Wir nehmen nun an, daß sich die gesuchte Funktion f in eine gleichmäßig konvergente Fourierreihe entwickeln läßt:

$$f(r,\varphi) = \sum_{k\in\mathbb{Z}} f_k(r) e^{ik\varphi} \quad .$$

Dann gilt:

ONS

$$\Delta f = \sum_{k\in\mathbb{Z}} \left(f_k'' + \frac{1}{r} f_k' - \frac{k^2}{r^2} f_k\right) e^{ik\varphi} = 0$$

$$\Rightarrow \quad f_k'' + \frac{1}{r} f_k' - \frac{k^2}{r^2} = 0 \qquad \forall\, k \in \mathbb{Z}$$

Wir machen nun den Ansatz:

$$f_k(r) = r^q$$

$$\Rightarrow \quad (q(q-1) + q - k^2) r^{q-2} = 0 \qquad \forall\, k \in \mathbb{Z}$$

$$\Rightarrow \quad q = \pm |k| \quad .$$

Unser Ansatz liefert also zwei linear unabhängige Lösungen, die wir linear kombinieren können:

$$f_k(r) = a_k\, r^{|k|} + b_k\, r^{-|k|} \quad .$$

Wir wollen nur reguläre Lösungen zulassen, so daß sich

$$b_k = 0 \qquad \forall\, k \neq 0$$

ergibt, da sonst f_k eine Singularität im Nullpunkt hätte.

Also:

$$f(r,\varphi) = \sum_{k\in\mathbb{Z}} a_k\, r^{|k|}\, e^{ik\varphi} \quad ,$$

wobei die Koeffizienten a_k aus der Randbedingung:

$$f(1,\varphi) = \sum_{k\in\mathbb{Z}} a_k\, e^{ik\varphi} = g(\cos\varphi, \sin\varphi)$$

zu bestimmen sind:

$$a_k = \frac{1}{2\pi} \int_0^{2\pi} g(\cos\varphi, \sin\varphi) e^{-ik\varphi} \, d\varphi \quad .$$

Oben eingesetzt ergibt das

$$f(r,\varphi) = \frac{1}{2\pi} \sum_{k\in\mathbb{Z}} r^{|k|} e^{ik\varphi} \int_0^{2\pi} g(\cos\psi, \sin\psi) e^{-ik\psi} \, d\psi$$

$$= \int_0^{2\pi} P(r,\varphi,\psi) g(\cos\psi, \sin\psi) \, d\psi$$

mit $$P(r,\varphi,\psi) = \frac{1}{2\pi} \sum_{k\in\mathbb{Z}} r^{|k|} e^{ik(\varphi-\psi)} \quad ,$$

dem sogenannten Poissonschen Integralkern.

Man verifiziert nun leicht:

i) $$P(r,\varphi,\psi) = \frac{1}{2\pi} \frac{1 - r^2}{1 - 2r\cos(\varphi-\psi) + r^2}$$

ii) $P(\cdot, \cdot, \psi)$ ist $\forall\, \psi \in [0,2\pi)$ stetig auf D

iii) $$\int_0^{2\pi} P(r,\varphi,\psi) \, d\psi = 1 \qquad \forall\, r \in [0,1), \ \forall\, \varphi \in [0,2\pi) \quad .$$

Damit bekommen wir die Abschätzung:

$$|f(r,\varphi)| \leq \int_0^{2\pi} |P(r,\varphi,\psi) g(\cos\psi, \sin\psi)| \, d\psi \leq \max_{\psi[0,2\pi)} |g(\cos\psi, \sin\psi)| \cdot 1$$

Ist nun anstelle der exakten Funktion g, nur eine mit einem Fehler behaftete Datenfunktion g_δ, deren Fehler durch

$$\max_{\psi\in[0,2\pi)} |g(\cos\psi, \sin\psi) - g_\delta(\cos\psi, \sin\psi)| \leq \delta$$

abgeschätzt werden kann, bekannt, so müssen wir das Problem

$$\Delta f_\delta = 0$$

$$f_\delta|_{\partial\Omega} = g_\delta$$

lösen.

Wegen: $\Delta(f-f_\delta) = 0$, $(f-f_\delta)|_{\partial\Omega} = g - g_\delta$

gilt demnach folgende Abschätzung für den Fehler der Lösung:

$$|(f-f_\delta)(r,\varphi)| \leq \max_{\psi\in[0,2\pi)} |(g-g_\delta)(\cos\psi,\sin\psi)| \leq \delta \quad .$$

Wir fassen zusammen:

a) das Problem ist eindeutig lösbar (wurde nicht gezeigt)

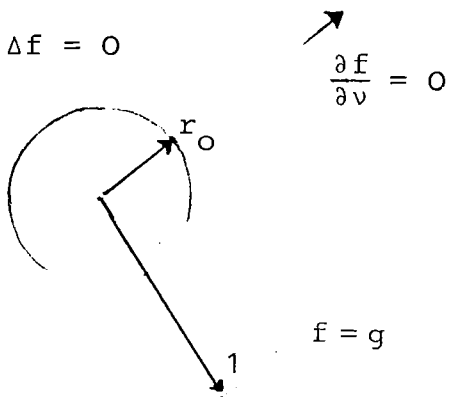
b) es ist "für alle" g lösbar, d.h. g muß keine restriktiven Voraussetzungen erfüllen

c) f hängt stetig von g ab, d.h. "kleine" Fehler in g führen zu "kleinen" Fehlern in f.

0.2 Beispiel eines schlecht gestellten Problems

Wir wandeln das obige Problem nur geringfügig ab:

Sei $\Omega = \{x \in \mathbb{R}^2 \mid r_o < \|x\| < 1, r_o < 1\}$ ein Kreisring.



ν - äußerer Normalenvektor

Wir suchen eine Funktion f mit:

i) $\Delta f = 0$ in Ω

ii) $f(x) = g(x)$, für $\|x\| = 1$

iii) $\frac{\partial f}{\partial \nu} = 0$ auf dem äußeren Rand von Ω.

BEMERKUNG: Probleme dieses Typs tauchen in der Elektrokardiographie und der Seismologie auf.

α) Analytischer Lösungsversuch

Wie oben entwickeln wir f in eine Fourierreihe:

$$f(r,\varphi) = \sum_{k\in\mathbb{Z}} f_k(r) e^{ik\varphi}$$

mit $f_k(r) = a_k r^k + b_k r^{-k}$, wobei nun i.a. $b_k \neq 0$ sein wird.

Die Randbedingungen liefern:

$$f(1,\varphi) = \sum_{k\in\mathbb{Z}} (a_k + b_k) e^{ik\varphi} = g(\cos\varphi, \sin\varphi)$$

sowie

$$0 = \frac{d}{dr}\Big|_{r=1} \left[\sum_{k\in\mathbb{Z}} (a_k r^k + b_k r^{-k}) e^{ik\varphi} \right] = \sum_{k\in\mathbb{Z}} k(a_k - b_k) e^{ik\varphi}$$

$\Rightarrow a_k - b_k = 0 \Rightarrow a_k = b_k$

Damit:

$$a_k + b_k = \frac{1}{2\pi} \int_0^{2\pi} g(\cos\varphi, \sin\varphi) e^{-ik\varphi} d\varphi$$

und mit $\tilde{a}_k = a_k + b_k = 2a_k$:

$$f(r,\varphi) = \sum_k \tilde{a}_k (r^k + r^{-k}) e^{ik\varphi} .$$

Eine Reihe diesen Typs ist i.a. nicht konvergent für $r < 1$.

Wir fassen zusammen:

a) das Problem ist eindeutig lösbar (wurde nicht gezeigt)

b) das Problem ist nur für solche Funktion g lösbar, für die die Fourierkoeffizienten a_k hinreichend schnell fallen, d.h. an g sind starke Regularitätsanforderungen zu stellen.

c) keinerlei stetige Abhängigkeit der Lösung von g:

sei g_δ eine leicht gestörte Version einer hinreichend regulären Funktion g, so wird g_δ diese Regularitätseigenschaften i.a. nicht besitzen, d.h. das Problem für g_δ wird erst gar nicht lösbar sein.

β) <u>Lösungsversuch des diskretisierten Problems - Differenzenverfahren</u>

Wir modifizieren ein wenig:

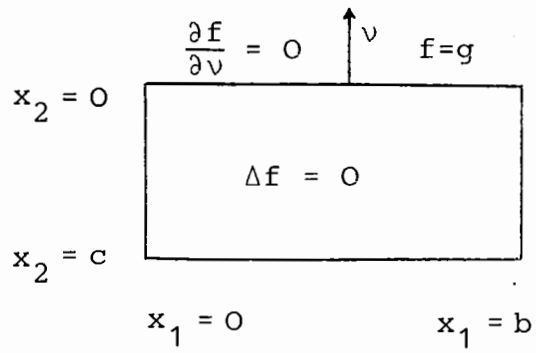
Sei $\Omega = \{x \in \mathbb{R}^2 \mid 0 \leq x_1 \leq b\ ,\ c \leq x_2 \leq 0\}$ ein Rechteck.

Wir suchen eine Funktion f mit:

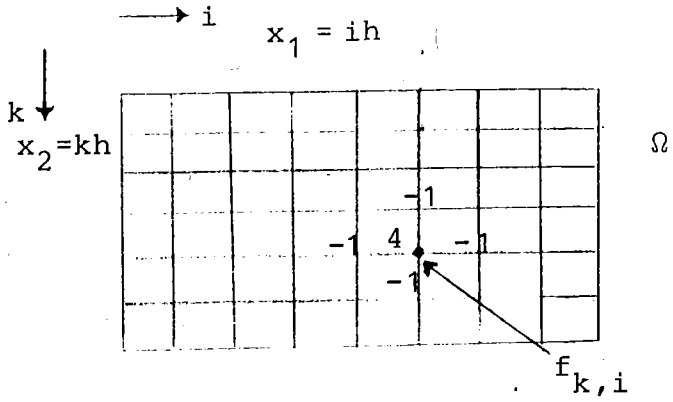
i) $\Delta f = 0$ in Ω

ii) $f(x_1,0) = g(x_1)\ ,\ 0 \leq x_1 \leq b$

iii) $\frac{\partial f}{\partial \nu} = 0$ auf dem oberen Rand von Ω.



Wir wählen eine äquidistante Gitterzerlegung von Ω und berechnen eine diskrete Version des Laplace - Operators:



$$\frac{\partial^2 f}{\partial x_1^2}(ih, kh) \approx \frac{f_{k,i+1} - 2f_{k,i} + f_{k,i-1}}{h^2}$$

$$\frac{\partial^2 f}{\partial x_2^2}(ih, kh) \approx \frac{f_{k+1,i} - 2f_{k,i} + f_{k-1,i}}{h^2}$$

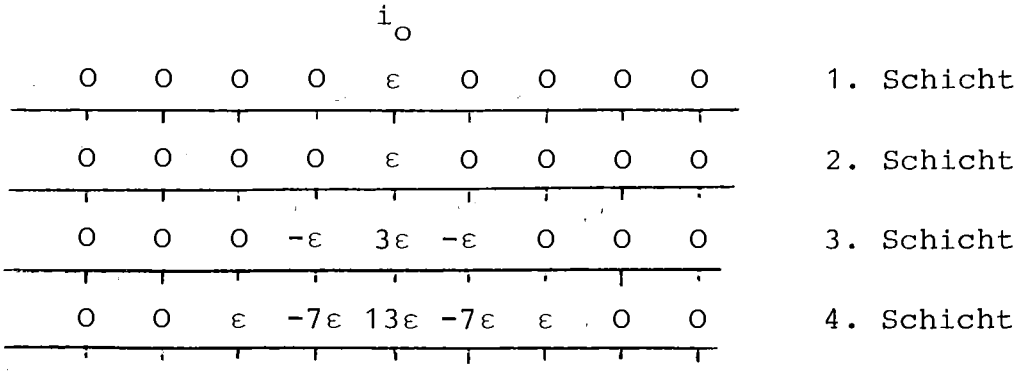
$$\Rightarrow \quad \Delta f(ih,kh) \approx (-4f_{k,i} + f_{k,i+1} + f_{k,i-1} + f_{k+1,i} + f_{k-1,i})/h^2$$

$$\Delta f = 0 \quad \Rightarrow \quad -4f_{k,i} + f_{k,i+1} + f_{k,i-1} + f_{k+1,i} + f_{k-1,i} = 0,$$

$$\frac{\partial f}{\partial \nu} = 0 \quad \Rightarrow \quad f_{0i} = f_{1i} \qquad i = 1, \ldots, K$$

Wegen $f = g$ auf dem oberen Rand kennen wir damit also die $f_{k,i}$ in den beiden obersten Schichten und können dann sehr einfach die darunterliegenden Schichten berechnen.

Das Problem liegt nun bei der Fehlerfortpflanzung. Nehmen wir an, wir hätten g nur in einem Punkt i_o mit einem Fehler ε versehen, die Fortpflanzung dieses Fehlers ergibt sich aus nachfolgendem Schema:



u.s.w.

Das heißt, es ist mit einem katastrophalen Anwachsen des Fehlers zu rechnen.

Abhilfe bei diesem Problem:
Wir brauchen Zustatzinformationen (a-priori-Informationen) um sinnvoll rechnen zu können, z.B. die folgende: es existiert eine Lösung für $\|x\| \geq R$, $r_o < R < 1$, das bedeutet aber:

$$\text{Konvergenz von } \sum_{k\in\mathbb{Z}} a_k (R^k + R^{-k}) e^{ik\varphi} \quad ,$$

also:

$$\text{Konvergenz von } a_k \text{ gegen Null "sehr schnell"} \quad .$$

I ALLGEMEINE THEORIE SCHLECHT GESTELLTER PROBLEME

1. GUT UND SCHLECHT GESTELLTE PROBLEME

Es seien X,Y normierte Räume und $A : X \to Y$ ein linearer Operator, wir definieren damit:

DEFINITION 1.1 (Hadamard): Das Problem

$$Af = g$$

ist gut (korrekt, sachgemäß) gestellt, falls die drei folgenden Bedingungen erfüllt sind:

i) $Af = g$ ist $\forall\, g \in Y$ lösbar

ii) die Lösung ist eindeutig

iii) aus $g_n \xrightarrow{\|\cdot\|_Y} g$ folgt $f_n \xrightarrow{\|\cdot\|_X} f$, wobei $Af_n = g_n$.

Andernfalls heißt $Af = g$ schlecht gestellt.

FOLGERUNG: $Af = g$ ist schlecht gestellt, falls (mindestens) eine der drei folgenden Bedingungen zutrifft:

i) A nicht surjektiv

ii) A nicht injektiv

iii) A^{-1} unstetig, d.h. unbeschränkt.

Praktische Schwierigkeiten bei der Behandlung schlecht gestellter Aufgaben sind folglich

i) Unlösbarkeit

ii) Mehrdeutigkeit

iii) Numerische Schwierigkeiten: mit $Af = g$, $Af_\delta = g_\delta$

wird selbst bei kleinen Datenfehlern $\|g - g_\delta\|_Y$ der Fehler der Näherungslösung $\|f - f_\delta\|_X$ groß sein.

BEISPIELE:

i) Sei $X = Y = C[0,1]$ und $Af(x) = \int_0^x f(y)\,dy$, so ist

$$Af = g \quad , \text{ (d.h. } g' = f \text{)}$$

schlecht gestellt, denn A ist nicht surjektiv (wir müßten uns auf $Y' = Y \cap \{f \mid f(0) = 0\}$ einschränken) und vor allem ist die Umkehrabbildung, die Differentiation keine stetige Abb. von $C[0,1]$ nach $C[0,1]$.

ii) Wir betrachten wieder das Beispiel aus 0.2:

$\Delta u = 0$ auf $\Omega := D_1(0) \setminus D_r(0)$; $0 < r < 1$

$u(r,\varphi) = f(\varphi)$, $u(1,\varphi) = g(\varphi)$ auf dem inneren bzw. äußeren Rand von Ω.

$\frac{\partial u}{\partial \nu} = 0$ auf dem äußeren Rand von Ω.

Wir hatten berechnet:

$$u(r,\varphi) = \sum_{k\in\mathbb{Z}} a_k (r^k + r^{-k}) e^{ik\varphi} = f(\varphi)$$

$$u(1,\varphi) = \sum_{k\in\mathbb{Z}} 2a_k e^{ik\varphi} = g(\varphi) \quad ,$$

also $a_k(r^k + r^{-k}) = \frac{1}{2\pi} \int_0^{2\pi} e^{-ik\psi} f(\psi)\,d\psi$

$$\Rightarrow \quad g(\varphi) = \sum_{k\in\mathbb{Z}} \frac{2}{r^k + r^{-k}} \frac{1}{2\pi} \int_0^{2\pi} e^{-ik\psi} f(\psi)\,d\psi\, e^{ik\varphi}$$

$$= \underbrace{\int_0^{2\pi} K(\psi-\varphi) f(\psi)\,d\psi}_{Af(\varphi)} \; ; \quad \text{mit} \quad K(\chi) = \frac{1}{\pi} \sum_{k\in\mathbb{Z}} \frac{1}{r^k + r^{-k}} e^{-ik\chi} .$$

Aus der Reihenentwicklung für K folgt sofort $K \in C^\infty$, und

wir werden sehen, daß die Integralgleichung

$$Af(\varphi) = g(\varphi)$$

mit einem glatten Kern K stets schlecht gestellt ist (s. Satz 1.1).

iii) Wir betrachten

$$\frac{1}{\pi} \oint_{-\infty}^{\infty} \frac{f(y)}{x-y}\, dy = g(x) \quad ,$$

d.h. die sogenannte Hilbert-Transformation, wobei

$$\oint \frac{f(y)}{x-y}\, dy \quad \text{als} \quad \lim_{\varepsilon \to 0} \int_{|x-y|>\varepsilon} \frac{f(y)}{x-y}\, dy$$

definiert ist (sogenannter Cauchy-Hauptwert).

Man kann zeigen, daß

$$\|f\|_{L_2} = \|g\|_{L_2}$$

gilt und damit das Problem für $X = Y = L_2$ gut gestellt ist.

iv) Sei K ein Operator mit $\|K\| < 1$ und $A = \mathbb{1} - K$, dann ist

$Af = g$ gut gestellt,

denn wir haben:

$$A^{-1} = \sum_{\ell=0}^{\infty} K^{\ell} \quad \text{(Neumannsche Reihe)} \quad \text{und folglich}$$

$$\|A^{-1}\| \leq \sum_{\ell=0}^{\infty} \|K\|^{\ell} = \frac{1}{1-\|K\|} < \infty .$$

v) Sei V ein kompakter Operator, d.h. das Bild kompakter Mengen unter V ist relativ kompakt. Wir betrachten wieder:

$$A = \mathbb{1} - V \quad ,$$

aus dem Fredholmschen Alternativsatz folgt:

A injektiv $\Rightarrow$ A surjektiv und A^{-1} stetig;

also

$$Af = g$$

ist gut gestellt, falls A injektiv ist.

Wir kommen nun zu einem ersten Satz über die Schlecht-Gestelltheit von Integralgleichungen.

SATZ 1.1: Es sei

$$Af(x) = \int_M K(x,y) f(y)\,dy \quad , \quad x \in N \quad ,$$

wobei M, N meßbare Teilmengen endlich-dimensionaler Räume mit inneren Punkten sind.

Sei $\int_N \int_M |K(x,y)|^2 dy\,dx \leq C < \infty$,

dann ist

$$A : L_2(M) \to L_2(N) \quad \text{stetig}$$

und A^{-1} unstetig (falls existent).

BEWEIS: Wir führen den Beweis für den Spezialfall $N = M = [0, 2\pi]$

i) Stetigkeit von A:

$$|Af(x)|^2 = \left| \int_0^{2\pi} K(x,y) f(y)\,dy \right|^2 \leq \left(\int_0^{2\pi} |K(x,y) f(y)|\,dy \right)^2$$

$$\leq \int_0^{2\pi} |K(x,y)|^2 dy \int_0^{2\pi} |f(y)|^2 dy$$

$$\Rightarrow \int_0^{2\pi} |Af(x)|^2\,dx \leq \int_0^{2\pi} \left(\int_0^{2\pi} |K(x,y)|^2 dy \int_0^{2\pi} |f(y)|^2\,dy \right) dx,$$

also $$\|Af\|_{L_2} \leq C \cdot \|f\|_{L_2} \quad .$$

ii) Unstetigkeit von A^{-1}:

Wir werden eine Folge $\{f_k\}_{k\in\mathbb{N}}$ angeben, mit:

$$\|f_k\| = 1 \quad \forall\, k \quad \text{und} \quad \|Af_k\| \to 0 \ , \quad \text{für} \quad k \to \infty \ ,$$

woraus die Unstetigkeit von A^{-1} unmittelbar folgt.

Wir setzen konkret:

$$f_k = (2\pi)^{-1/2}\, e^{ikx} \ , \quad k \in \mathbb{Z} \ .$$

Die f_k stellen ein vollständiges Orthonormalsystem in $L_2([0,2\pi])$ dar, für das die Parsevalsche Gleichung:

$$\|g\|^2 = \sum_{k\in\mathbb{Z}} |(f_k,g)|^2 \qquad (*)$$

gilt, hierbei ist $(\cdot\,,\cdot)$ das gewöhnliche L_2 - Skalarprodukt.

Nun gilt:

$$Af_k(x) = \int_0^{2\pi} K(x,y) f_k(y)\,dy = (K(x,\cdot),f_k)$$

$$\Rightarrow \quad \sum_{k\in\mathbb{Z}} |Af_k|^2(x) = \sum_{k\in\mathbb{Z}} |(K(x,\cdot),f_k)|^2 \overset{(*)}{=} \int_0^{2\pi} |K(x,y)|^2 dy$$

und Integration über x liefert:

$$\sum_{k\in\mathbb{Z}} \|Af_k\|^2 = \int_0^{2\pi}\int_0^{2\pi} |K(x,y)|^2 dxdy = C < \infty$$

Aus der Konvergenz der Reihe folgt insbesondere, daß deren Glieder eine Nullfolge bilden:

$$\|Af_k\| \to 0 \ , \quad k \to \infty \qquad \blacksquare$$

BEMERKUNG: Der Beweis benutzt keinerlei spezielle Eigenschaften der o.a. Orthonormalbasis, wir benötigen lediglich die Existenz einer solchen. Dieses Argument rechtfertigt die Spezialisierung im Beweis.

Wir führen einige Begriffe ein:

Wir nennen:

$\int_M K(x,y)f(y)dy = g(x)$ eine Integralgleichung 1. Art,

$f(x) - \int_M K(x,y)f(y)dy = g(x)$ eine Integralgleichung 2. Art

und

$K(x,y)$ eine Kernfunktion.

Integralgleichungen 1. Art sind typischerweise schlecht gestellt, solche 2. Art typischerweise gut gestellt (falls injektiv).

Ob ein Problem gut oder schlecht gestellt ist, hängt (trivialerweise) von den Normen in X, Y ab.

BEISPIELE:

i) $\int_0^x f(y)dy = g(x)$ ist schlecht gestellt, falls

$X = Y = L_2([0,1])$

α) Wähle nun $Y = \{g \in C^1[0,1],\ g(0) = 0\}$ und

setze: $\|g\|_Y = \left(\int_0^1 |g'(x)|^2 dx \right)^{1/2}$, so gilt

mit $X = L_2([0,1])$:

$$\|f\|_X = \left(\int_0^1 |g'(x)|^2 dx \right)^{1/2} = \|g\|_Y .$$

Das Problem ist nun also gut gestellt.

β) Wählt man in X eine neue Norm:

$$\|f\|_X := \sup_{\substack{v\in C^1[0,1] \\ v(0)=v(1)=0}} \left[\left| \int_0^1 f(x)v(x)\,dx \right| \Big/ \left(\int_0^1 |v'(x)|^2 dx \right)^{1/2} \right],$$

so gilt: $\|f\|_X = \|g\|_{L_2}$

und damit ist das Problem gut gestellt.

ii) Es kann vorkommen, daß Funktionale von f gut bestimmt sind, obwohl die Aufgabe $Af = g$ schlecht gestellt ist.

Wir betrachten nochmals das Beispiel aus 0.2:

Wir wollen jetzt nicht die Funktion f auf dem inneren Kreisring explizit bestimmen, sondern nur deren Mittelwert

$$\int_0^{2\pi} f(\varphi)\,d\varphi = \int_0^{2\pi} \sum_{k\in\mathbb{Z}} a_k (r^k + r^{-k}) e^{ik\varphi}\,d\varphi$$

$$= 2a_0 = \frac{1}{2\pi} \int_0^{2\pi} g(\varphi)\,d\varphi \quad ,$$

das Funktional $\int_0^{2\pi} f(\varphi)\,d\varphi$ hängt also stetig von g ab.

Wir hatten bemerkt, daß es wichtig ist, a-priori-Informationen ausnutzen zu können:

Zusätzlich zur Gleichung $Af = g$ haben wir die Information:

$f \in M$ mit $M \subset X$.

Wie läßt sich nun der Fehler $\|f_1 - f_2\|_X$ abschätzen, falls $f_1, f_2 \in M$ und $\|Af_1 - Af_2\|_Y \leq \delta$ gilt?

Folgende Bezeichnungsweise ist nützlich:

$$\alpha(M,\delta) := \sup \{ \|f_1 - f_2\|_X \mid f_1, f_2 \in M, \|Af_1 - Af_2\| \leq \delta \}$$

Mit dieser Notation gilt der

SATZ 1.2: Sei M eine kompakte Teilmenge von X und $A : X \to Y$ stetig und auf M injektiv, dann gilt:

$$\lim_{\delta \to 0} \alpha(M,\delta) = 0 .$$

BEWEIS: Aufgabe 1. ■

2. VERALLGEMEINERTE LÖSUNGEN

Im folgenden seien X, Y Hilbert-Räume und $A : X \to Y$ ein stetiger linearer Operator.

Für $M \subseteq X$ bezeichnen wir mit

$\overline{M}$ den Abschluß von M und mit

$M^{\perp} = \{x \in X,\ x \perp M\}$ das orthogonale Komplement.

Falls M eine lineare Mannigfaltigkeit in X ist, haben wir $(M^{\perp})^{\perp} = \overline{M}$, für beliebige M können wir X in eine orthogonale Summe

$$X = M^{\perp} \oplus \overline{M}$$

zerlegen. Gilt $M = \overline{M}$, so haben wir also: $X = M^{\perp} \oplus M$, d.h. $\forall\, x \in X$ eine eindeutige Zerlegung $x = x_1 + x_2$, mit $x_1 = Px \in M$ und $x_2 \in M^{\perp}$, hierbei ist P die orthogonale Projektion auf M.

Die Operatornorm ist definiert als:

$$\|A\| = \sup_{\|x\|_X = 1} \|Ax\|_Y \quad .$$

Der Nullraum: $N(A) := \{x \in X,\ Ax = O\}$ ist infolge der Stetigkeit von A abgeschlossen, während der Wertebereich:

$R(A) := \{y \in Y \quad \exists\, x \in X$ mit $y = Ax\}$ i.a. nicht abgeschlossen zu sein braucht.

BEISPIEL dazu: Sei $X = Y = L_2(a,b)$ und

$$Af(x) = \int_a^b K(x,y) f(y)\, dy \quad , \quad \text{mit} \quad K \in C^{\infty} \quad ,$$

so gilt: $R(A) \subseteq C^{\infty}[a,b]$. Eine bzgl. der L_2-Topologie konvergente Folge in $R(A)$ muß jedoch nicht notwendig gegen ein Grenzelement aus $C^{\infty}[a,b]$ konvergieren, d.h. das Grenzelement liegt somit nicht notwendig in $R(A)$.

Der adjungierte Operator $A^* : Y \to X$ wird durch

$$(Af,g)_Y = (f,A^*g)_X \qquad \forall\ f \in X,\ g \in Y$$

charakterisiert. Es gilt: $(A^*)^* = A$.

Wertebereich und Nullraum eines Operators bzw. seines Adjungierten sind folgendermaßen miteinander verknüpft.

SATZ 2.1: Es gilt:

i) $R(A)^{\perp} = N(A^*)$

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

BEWEIS:

i) $y \in R(A)^{\perp} \Leftrightarrow (y,Ax) = 0 \qquad \forall\ x \in X$

$\Leftrightarrow (A^*y,x) = 0 \qquad \forall\ x \in X$

$\Leftrightarrow A^*y = 0$

ii) ersetze in i) A durch A^*:

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

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

FOLGERUNG:

$$X = N(A) \oplus \overline{R(A^*)}$$

$$Y = N(A^*) \oplus \overline{R(A)}$$

DEFINITION 2.1: u heißt kleinste - Quadrate - Lösung von $Af = g$

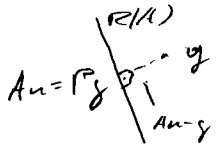
$$\Leftrightarrow \quad \|Au-g\| \leq \|Af-g\| \quad , \quad \forall\, f \in X \quad .$$

SATZ 2.2: Folgende Aussagen sind äquivalent:

i) u ist kleinste - Quadrate - Lösung

ii) $A^*Au = A^*g$ (Normalgleichungen)

iii) $Au = Pg$, wobei P für die orthogonale Projektion auf $\overline{R(A)}$ steht.



BEWEIS:

i) ⇒ ii) Wir schreiben ein beliebiges $f \in X$ in der Form: $f = u + v$

$$\Rightarrow \quad \|Af-g\|^2 = \|Au-g+Av\|^2 = \|Au-g\|^2 + 2\mathrm{Re}(Au-g,Av) + \|Av\|^2$$

$$\geq \|Au-g\|^2$$

Also: $$\|Av\|^2 \geq -2\,\mathrm{Re}\,(Au-g,Av) \qquad \forall\, v \in X$$

BEHAUPTUNG: $(Au-g,Av) = 0 \qquad \forall\, v \in X$

BEWEIS dazu: Wähle $v_o \in X$ beliebig $(\neq 0)$ und $\lambda_n \neq 0$ eine Nullfolge komplexer Zahlen, s.d. $(Au-g, A(\lambda_n v_o))$ rein reell und negativ ist, dann gilt:

$$|\lambda_n|^2 \|Av_o\|^2 \geq 2|\bar{\lambda}_n| \; |(Au-g,Av_o)|$$

Nach Division durch $|\lambda_n|$ folgt die Aussage, da die linke Seite für $n \to \infty$ beliebig klein wird.

Also: $(A^*(Au-g),v) = 0 \quad \forall\, v \in X$

und somit $A^*Au - A^*g = 0$

ii) $\Rightarrow$ iii) Z.Zg. $g - Au \perp$ zu $R(A)$:

$$(g - Au, Av) = (A^*g - A^*Au, v) \overset{ii)}{=} (A^*g - A^*g, v) = 0$$

iii) $\Rightarrow$ i) $\|Af - g\|^2 = \|Au - g + A(f - u)\|^2$

$$\overset{iii)}{=} \|Au - g\|^2 + \|A(f - u)\|^2 \geq \|Au - g\|^2$$

■

FOLGERUNG:

1.) $Af = g$ hat kleinste-Quadrate-Lösung

$\Leftrightarrow$ $g \in R(A) + R(A)^{\perp}$,

denn $g = Au + h$, mit $h \in R(A)^{\perp}$

2.) Die Menge aller $g \in Y$, für die eine kleinste-Quadrate-Lösung existiert, ist dicht in Y, denn:

$$\overline{(R(A) + R(A)^{\perp})} = \overline{R(A)} + R(A)^{\perp} = Y \quad .$$

DEFINITION 2.2: Wir bezeichnen mit

$$A^+g$$

die kleinste-Quadrate-Lösung minimaler Norm, d.h.:

$$u = A^+g \Leftrightarrow \begin{cases} \|Au-g\|_Y \leq \|Af-g\|_Y \quad \forall\, f \in X \\ \|u\|_X \leq \|v\|_X \quad \forall\, v \in X \text{ mit: } \|Au-g\| = \|Av-g\| \end{cases}$$

$u = A^+g$ existiert für $g \in R(A) + R(A)^{\perp}$, denn die Menge M der kleinste-Quadrate-Lösungen ist eine abgeschlossene und konvexe (eventuell leere) Teilmenge von Y, in welcher es folglich ein (eindeutig) bestimmtes Element kleinster Norm gibt. Die Eindeutigkeit rechtfertigt die Bezeichnungsweise von Def. 2.2. Die Abge-

schlossenheit von M sieht man sofort, wenn man bedenkt, daß $M = \{u + v \mid u$ ist kleinste-Quadrate-Lösung, $A^*Av = 0\}$ gilt und A^*A ein stetiger Operator mit folglich abgeschlossenem Nullraum ist.

Bezeichnung: A^+ heißt Moore-Penrose-Inverse.

SATZ 2.3: Es gilt für $g \in R(A) + R(A)^\perp$:

$$f = A^+g \quad \Leftrightarrow \quad \begin{cases} \text{i)} & A^*Af = A^*g \\ \text{ii)} & f \in \overline{R(A^*)} \end{cases}$$

BEWEIS: Nach Satz 2.2 ist i) äquivalent damit, daß f kleinste Quadrat-Lösung ist; nach der Folgerung aus Satz 2.1 hat dieses f genau dann minimale Norm, wenn ii) gilt. ■

DEFINITION 2.3: Der Operator A besitzt eine Singulärwertzerlegung (SVD) genau dann, wenn gilt:

$$Af = \sum_{k=1}^{\infty} \sigma_k (f, f_k)_X \, g_k \quad ,$$

wobei $\{f_k\}$ ein Orthonormalsystem (ONS) in X und $\{g_k\}$ ein ONS in Y ist.

Ferner muß für die Singulärwerte σ_k von A gelten:

$$\sigma_k > 0 \quad \forall\, k \quad \text{(insbesondere also } \sigma_k \in \mathbb{R}) \quad .$$

Aus: $$\|Af\|_Y^2 = \sum_{k=1}^{\infty} |\sigma_k (f, f_k)_X|^2 \leq \sup_k (\sigma_k^2) \sum_{k=1}^{\infty} |(f, f_k)|^2$$

$$\leq \sup_k (\sigma_k^2) \, \|f\|^2$$

folgt sofort: σ_k beschränkt $\Leftrightarrow$ A stetig.

Aus $(Af,g)_Y = \sum_{k=1}^{\infty} \sigma_k (f,f_k)_X (g_k,g)_Y = \left(f , \sum_{k=1}^{\infty} \sigma_k (g,g_k)_Y f_k \right)_X$

folgt: $A^*g = \sum_{k=1}^{\infty} \sigma_k (g,g_k) f_k$.

BEMERKUNG: Insbesondere gelten die Relationen:

$$Af_k = \sigma_k g_k \quad \text{sowie} \quad A^*g_k = \sigma_k f_k \quad \forall\ k \in \mathbb{N} \ .$$

Man rechnet leicht nach, daß

$$A^*Af = \sum_{k=1}^{\infty} \sigma_k^2 (f,f_k) f_k$$

sowie

$$AA^*g = \sum_{k=1}^{\infty} \sigma_k^2 (g,g_k) g_k$$

gilt. Insbesondere also:

$$A^*Af_k = \sigma_k^2 f_k$$

sowie

$$AA^*g_k = \sigma_k^2 g_k \ ,$$

d.h. f_k bzw. g_k sind Eigenfunktionen von A^*A bzw. AA^* zum Eigenwert σ_k^2 .

SATZ 2.4: Sei $g \in R(A) + R(A)^{\perp}$, dann gilt:

$$A^+g = \sum_{k=1}^{\infty} \sigma_k^{-1} (g,g_k) f_k \ .$$

BEWEIS:

i) Zunächst muß die Konvergenz der Reihe gezeigt werden:

Sei $g = Av + h$ mit $h \in R(A)^{\perp}$

$$(g,g_k) = (Av,g_k) + \underbrace{(h,g_k)}_{=0}$$

$$= (v,A^*g_k)$$

$$= \sigma_k (v,f_k)$$

$$\Rightarrow \quad \sum_{k=1}^{\infty} |\sigma_k^{-1}(g,g_k)|^2 = \sum_{k=1}^{\infty} |(v,f_k)|^2 \leq \|v\|^2 < \infty$$

ii) Nach Satz 2.3 ist:

α) $A^*A\left(\sum_{k=1}^{\infty} \sigma_k^{-1}(g,g_k) f_k\right) = A^*g$ und

β) $\sum_{k=1}^{\infty} \sigma_k^{-1}(g,g_k) f_k \in \overline{R(A^*)}$ zu zeigen.

Zu α):

$$A^*A\left(\sum_{k=1}^{\infty} \sigma_k^{-1}(g,g_k) f_k\right)$$

$$= \sum_{j=1}^{\infty} \sigma_j^2 \left(\sum_{k=1}^{\infty} \sigma_k^{-1}(g,g_k) f_k, f_j\right) f_j$$

$$= \sum_{j=1}^{\infty} \sigma_j (g,g_j) f_j = A^*g \qquad ..$$

Zu β):

Nach Def. 2.3 gilt: $Au = 0 \Leftrightarrow (u,f_k) = 0 \quad \forall\ k$,

wegen $\left(u, \sum_{k=1}^{\infty} \sigma_k^{-1}(g,g_k) f_k\right) = \sum_{k=1}^{\infty} \sigma_k^{-1} \overline{(g,g_k)} (u,f_k)$,

gilt also $\sum_{k=1}^{\infty} \sigma_k^{-1}(g,g_k) f_k \in N(A)^{\perp} = \overline{R(A^*)}$ $\Big)$. ∎

$= \sum \left(u, \sigma_n^{-1}(g,g_n) f_n\right)$

$= \sum \overline{\sigma_n^{-1}(g,g_n)}\ (u,f_n)$

reell

3. FEHLERABSCHÄTZUNGEN

Wir hatten die Größe

$$\alpha(M,\delta) = \sup_{f_1,f_2\in M} \{\|f_1-f_2\| : \|Af_1-Af_2\| \leq \delta\}$$

eingeführt und in Satz 1.2 gezeigt, daß im Falle kompakter Mengen M gilt: $\alpha(M,\delta) \xrightarrow{\delta\to 0} 0$.

Mit Hilfe dieser Größe wollen wir den Fehler in unserer Standardsituation

$$Af = g \quad , \quad f \in M \quad \text{(a-priori-Information)}$$

g_δ bekannt mit $\|g-g_\delta\| \leq \delta$,

f_δ ist Lösung von $\min_{f\in M} \|Af-g_\delta\|$,

abschätzen, es gilt der

SATZ 3.1: Mit den obigen Bezeichnungen gilt

$$\|f-f_\delta\| \leq \alpha(M,2\delta) \quad .$$

BEWEIS: $\|Af_\delta-g_\delta\| \leq \|Af-g_\delta\| \leq \|g-g_\delta\| \leq \delta$

$$\|Af_\delta-Af\| \leq \|Af_\delta-g_\delta\| + \|g_\delta-g\| \leq 2\delta \quad ,$$

also: $f,f_\delta \in M$, $\|Af_\delta-Af\| \leq 2\delta \overset{\text{Def.}}{\Rightarrow} \|f-f_\delta\| \leq \alpha(M,2\delta)$ ∎

BEMERKUNG:

i) Falls $Af = g$ gut gestellt ist, gilt

$$\alpha(M,\delta) \leq \|A^{-1}\|\delta \quad ,$$

d.h. α ist linear in δ.

ii) Falls $Af = g$ schlecht gestellt ist, wird $\alpha(M,\delta)$ langsamer als δ gegen Null konvergieren;
typisch ist eine Abhängigkeit der Form

$$\alpha(M,\delta) \sim \delta^a \quad , \quad 0 < a < 1 \quad .$$

Es kommen aber auch noch kritische Konvergenzraten, wie z.B.

$$\alpha(M,\delta) \sim (\ell n \frac{1}{\delta})^{-1}$$

vor.

Falls $\alpha(M,\delta) = \delta^a$ gilt, bedeutet das für die Rechengenauigkeit: ist g auf m Dezimalstellen bekannt, also: $\delta \sim \|g\| \, 10^{-m}$, so ist der Fehler in f von der Größenordnung 10^{-am}. Die Lösung f ist also nur auf $a \cdot m$ Dezimalstellen genau zu bestimmen, d.h. der relative Verlust an Dezimalstellen beträgt $1 - a$.

Würde $\alpha(M,\delta) \sim (\ell n \frac{1}{\delta})^{-1}$ gelten, so wäre die Genauigkeit in f proportional zu $\frac{1}{m}$, bei m Dezimalstellen Genauigkeit in g. Wollte man nun f auf nur 10% Genauigkeit bestimmen, so müßte g folglich mit einer Genauigkeit von 10^{-10} bekannt sein.

<u>BEZEICHNUNG:</u> Man bezeichnet ein Problem als

schwach schlecht gestellt, falls $1 - a = \varepsilon$, $0 < \varepsilon << 1$

mäßig " " , falls $a \sim \frac{1}{2}$

sehr " " , falls $a = \varepsilon$, $0 < \varepsilon << 1$, oder bei noch langsamerem Konvergenzverhalten von $\alpha(M,\delta)$, wie z.B. logarithmischem Verhalten.

Wir wollen nun anhand von drei Beispielen zeigen, wie man bei vorgegebener Menge M die Funktion $\alpha(M, \cdot)$ berechnen kann.

i) Sei $Af(x) = \int_0^x f(y)\,dy = g(x)$ und $X = Y = C[0,1]$, versehen mit der Maximum-Norm.

Ferner sei

$$M = \{f \in X : \|f'\| \leq \S\} \quad \text{und} \quad f \in M, \text{ sowie } \|Af\| = \|g\| \leq \delta .$$

Es gilt: $f(x) = \int_y^x f'(t)\,dt + f(y)$

$$\Rightarrow \quad hf(x) = \int_x^{x+h} f(x)\,dy = \int_x^{x+h}\int_y^x f'(t)\,dt\,dy + \int_x^{x+h} f(y)\,dy$$

$$\Rightarrow \quad h|f(x)| \leq \left|\int_x^{x+h}\int_y^x f'(t)\,dt\,dy\right| + |g(x+h) - g(x)|$$

$$\leq \quad h^2\|f'\| \quad + \quad 2\|g\|$$

$$\Rightarrow \quad \|f\| \leq h\rho + \frac{2}{h}\delta$$

Wir versuchen, eine günstige Wahl von h durch Ausbalancieren der Terme auf der rechten Seite zu erzielen:

$$h\rho = \frac{2}{h}\delta \quad \Rightarrow \quad h = \left(\frac{2}{\rho}\delta\right)^{1/2}$$

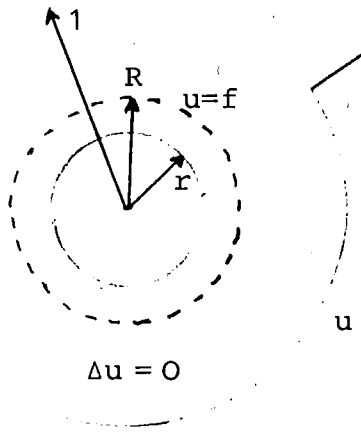
$$\Rightarrow \quad \|f\| \leq 2(2\cdot\rho\delta)^{1/2}$$

Wegen $\alpha(M,\delta) \leq \sup_{\|f\|\leq 2\delta} \{\|f\|, \|Af\| \leq \delta\}$ gilt also:

$$\boxed{\alpha(M,\delta) \leq 4\,\rho^{1/2}\cdot\delta^{1/2}}.$$

D.h. wir haben $a = \frac{1}{2}$.

ii) Wir betrachten erneut das Beispiel O.2:



$\Delta u = O$ auf $= D_1(O) \setminus D_r(O)$

$u = g$ auf dem äußeren Rand von Ω

$\frac{\partial u}{\partial \nu} = O$ "

$u(x) = f(x) \quad \forall\, x$ mit $|x| = R$

Wir wählen: $X = L_2(|x| = R)$, $Y = L_2(|x| = 1)$

Es gilt:

$$f(x) = \sum_{k \in \mathbb{Z}} a_k (R^k + R^{-k}) e^{ik\varphi} \quad ,$$

mit

$$a_k = \frac{1}{4\pi} \int_0^{2\pi} g(\varphi) e^{-ik\varphi} \quad .$$

Sei $M = \{f : \sum_{k \in \mathbb{Z}} |a_k|^2 (r^k + r^{-k})^2 \leq \rho^2 < \infty\}$,

d.h. f existiert noch auf dem Kreis mit Radius r.

Wir nehmen nun an: $f \in M$, $\|g\| = \|Af\| \leq \delta$

Sei nun $p,q \in]1,\infty]$, $\frac{1}{p} + \frac{1}{q} = 1$, dann gilt:

$$\|f\|^2_{L_2(|x|=R)} = \sum_{k \in \mathbb{Z}} |a_k|^2 (R^k + R^{-k})^2$$

$$= \sum_{k \in \mathbb{Z}} \underbrace{|a_k|^{2/p} \frac{(R^k + R^{-k})^2}{(r^k + r^{-k})^{2/q}}}_{=: x_k} \underbrace{(r^k + r^{-k})^{2/q} |a_k|^{2/q}}_{=: y_k}$$

$$\leq \underbrace{\left(\sum_{k\in\mathbb{Z}} |a_k|^2 \, \frac{(R^k+R^{-k})^{2p}}{(r^k+r^{-k})^{2p/q}} \right)^{1/p}}_{=\|x_k\|_p} \underbrace{\left(\sum_{k\in\mathbb{Z}} (r^k+r^{-k})^2 |a_k|^2 \right)^{1/q}}_{=\|y_k\|_q \leq \rho^{2/q},\ \text{da } f\in M}$$

$$\leq \sup_{k\geq 0} \frac{(R^k+R^{-k})^2}{(r^k+r^{-k})^{2/q}} \underbrace{\left(\sum_{k\in\mathbb{Z}} |a_k|^2 \right)^{1/p}}_{=\frac{1}{4}\|g\|^2 \leq \delta^2} \cdot \rho^{2/q} \quad .$$

Es gilt: $$\frac{R^k+R^{-k}}{(r^k+r^{-k})^{2/q}} = \frac{R^{-k}}{r^{-k/q}} \cdot \underbrace{\frac{1+R^{2k}}{(1+r^{2k})^{1/q}}}_{\longrightarrow 1 \text{ mit } k\to\infty}$$

Wir wählen q so, daß

$$\frac{R^{-1}}{r^{-1/q}} = 1$$

gilt, also $R = r^{1/q} \Leftrightarrow q = \frac{\ell n\, r}{\ell n\, R} > 1$

$$\Rightarrow \frac{1}{p} = 1 - \frac{1}{q} = 1 - \frac{\ell n\, R}{\ell n\, r} \quad .$$

Damit haben wir

$$\|f\|_{L_2(|x|=R)} \leq c \left(\frac{\delta}{2}\right)^{1/p} \cdot \delta^{1/q}$$

mit $\frac{1}{p} = 1 - \frac{\ell n\, R}{\ell n\, r} \in]0,1[$.

Somit:

$$\alpha(M,\delta) = \sup_{\|f\|_{L_2(|x|=r)} \leq 2\rho} \{ \|f\|_{L_2(|x|=R)} : \|Af\|_{L_2(|x|=1)} \leq \delta \}$$

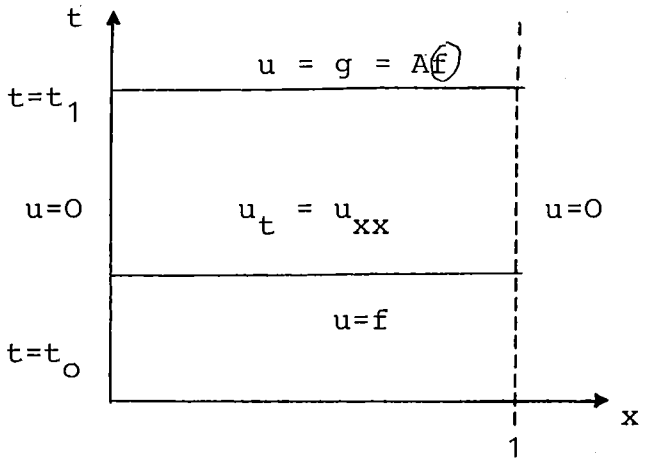
$$\leq c\, 2^{-1/p} \cdot 2^{1/q}\, \delta^{1/p} \cdot \rho^{1/q} \quad .$$

Also: $a = \frac{1}{p} = 1 - \frac{\ell n\, R}{\ell n\, r}$, d.h.

falls: $R - r = \varepsilon << 1$ ist $a \approx 0$, d.h. das Problem sehr schlecht gestellt,

falls $R >> r$ ist das Problem mäßig oder nur schwach schlecht gestellt.

iii) Wir untersuchen folgendes Wärmeleitungsproblem:



Wir nehmen $X = Y = L_2(0,1)$ und

$$\underline{\underline{M}} = \{f \mid f(x) = u(x,t_o),\ u(x,0) = u_o \quad \text{mit}$$

$$\|u_o\|_{L_2(0,1)} \leq \rho\ ,\ u \text{ ist Lösung der}$$

obigen Wärmeleitungsgleichung$\}$.

Wir benötigen folgendes

LEMMA 3.2: Sei $F \in C^2[a,b]$, $F > 0$ und $\frac{d^2}{dt^2} \ln F(t) \geq 0$, d.h. F ist logarithmisch konvex, dann gilt die Abschätzung:

$$F(t) \leq F(a)^{\frac{b-t}{b-a}} \cdot F(b)^{\frac{t-a}{b-a}} .$$

BEWEIS: $\ln F$ konvex:

$$\ln(F(a(1-\lambda) + \lambda b)) \leq (1-\lambda)\ln(F(a)) + \lambda \ln(F(b))$$

$$= \ln(F(a)^{1-\lambda}) + \ln(F(b)^{\lambda})$$

Setze nun: $\lambda = \frac{t-a}{b-a} \Leftrightarrow 1-\lambda = \frac{b-t}{b-a} \Leftrightarrow a(1-\lambda)+\lambda b = t$ und wende die Exponentialfunktion an. ∎

Sei nun $f \in M$, $\|g\| = \|Af\| \leq \delta$,

def.: $F(t) = \int_0^1 u^2(x,t)\,dx$, also: $F(t_o) = \|f\|^2$, $F(t_1) = \|Af\|^2$,

es gilt: $F'(t) = 2\int_0^1 u_t\, u\, dx = 2\int_0^1 u_{xx}\, u\, dx$ ($u_t = u_{xx}$ Wärmeleitungsgl.)

$= -2\int_0^1 u_x^2\, dx$ (Die Randterme fallen wegen $u(0,\cdot) = u(1,\cdot) = 0$ weg.)

$\Rightarrow \quad F''(t) = -4\int_0^1 \underbrace{u_{xt}}_{=u_{xxx}}\, u_x\, dx = 4\int_0^1 u_{xx}^2\, dx$

(Die Randterme verschwinden, da $u(0,t) = u(1,t) \equiv 0$, also auch $u_t = u_{xx} = 0$ für $x = 0,1$.)

$$\frac{d^2}{dt^2} \ln F(t) = \left(\frac{F'}{F}\right)' = \left(\frac{F''}{F} - \frac{F'^2}{F^2}\right) = (F''\cdot F - F'^2)/F^2$$

$$F''F - F'^2 = 4\int_0^1 u_{xx}^2\, dx \int_0^1 u^2\, dx - 4\left(\int_0^1 u_x^2\, dx\right)^2 \geq 0 \ ,$$

denn: $\left(\int_0^1 u_x^2\, dx\right)^2 = \left(-\int_0^1 u_{xx}\, u\, dx\right)^2 \leq \int_0^1 u_{xx}^2\, dx \int_0^1 u^2\, dx$.

Nach Lemma 3.2 gilt mit $a = 0$, $b = t_1$ und $t = t_o$:

$$F(t_o) \leq F(0)^{\frac{t_1-t_o}{t_1}} \cdot F(t_1)^{t_o/t_1}$$

$$\Leftrightarrow \quad \|f\| \leq \|u_o\|^{\frac{t_1-t_o}{t_1}} \cdot \|Af\|^{t_o/t_1}$$

$$\leq \rho^{\frac{t_1-t_o}{t_1}} \cdot \delta^{t_o/t_1} \ .$$

Damit: $\alpha(M,\delta) \leq (2\rho)^{\frac{t_1-t_o}{t_1}}\, \delta^{t_o/t_1}$.

4. REGULARISIERUNG

Wir hatten in dem Fall, daß die Lösung von

$$Af = g$$

nicht existiert oder mehrdeutig ist, den eindeutig bestimmten Ausdruck

$$A^+ g$$

als (verallgemeinerte) Lösung definiert.

Das Unstetigkeitsproblem ist damit aber noch nicht gelöst, denn A^+ ist i.a. unstetig, $A^+ g$ also nicht direkt berechenbar. In diesem Abschnitt wollen wir uns mit konstruktiven Lösungsmöglichkeiten für dieses Problem befassen.

DEFINITION 4.1: Sei $A : X \to Y$ ein linearer stetiger Operator zwischen Hilberträumen. Für alle $\gamma > 0$ sei $R_\gamma : Y \to X$ linear und stetig. R_γ heißt Regularisierung von A, falls gilt

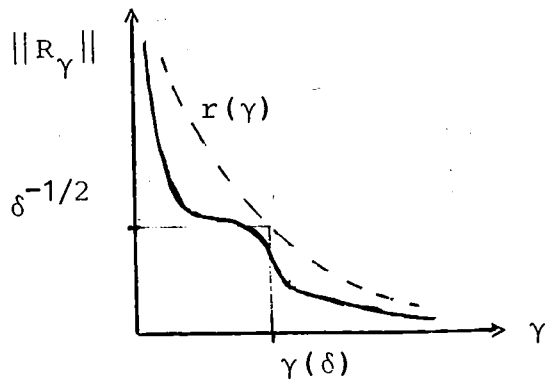
$$R_\gamma g \xrightarrow{\gamma \to 0} A^+ g \qquad \forall\, g \in R(A) + R(A)^\perp \quad ,$$

γ heißt Regularisierungsparameter.

Anwendung dieser Definition auf das Problem:

$$Af = g \;, \quad g_\delta \text{ bekannt mit } \|g - g_\delta\| \leq \delta \quad .$$

Wir wollen $R_\gamma g_\delta$ als Näherungslösung für f_δ nehmen; vorausgesetzt R_γ sei $\forall\, \gamma > 0$ bekannt, haben wir dann das Problem γ als Funktion von δ zu bestimmen. Da A^+ unbeschränkt ist, wird $\|R_\gamma\|$ als Funktion von γ ungefähr wie folgt aussehen:



Angenommen, wir hätten eine Funktion $r(\gamma)$, welche "ungefähr" den Verlauf von $\|R_\gamma\|$ wiedergibt, und die zusätzlichen Eigenschaften:

$$r(\gamma) \geq \|R_\gamma\|$$

$r(\gamma)$ stetig in $]0,\infty[$

$r(\gamma)$ monoton fallend mit $r(\gamma) \xrightarrow{\gamma\to\infty} 0$

aufweist.

Definieren wir dann:

$$\gamma(\delta) = r^{-1}(\delta^{-1/2}) \quad ,$$

so gilt:

i) $\gamma(\delta) \to 0$ für $\delta \to 0$ $\qquad r^{-1}(\infty) = z \Rightarrow r(z) = \infty \Rightarrow z = 0$

ii) $(\|R_{\gamma(\delta)}\| \cdot \delta) \to 0$ für $\delta \to 0$,

denn: $\|R_{\gamma(\delta)}\| \leq r(\gamma(\delta)) = \delta^{-1/2}$.

Falls i) und ii) erfüllt sind, gilt:

$$\|R_{\gamma(\delta)} g_\delta - A^+ g\| \leq \|R_{\gamma(\delta)}(g_\delta - g)\| + \|R_{\gamma(\delta)} g - A^+ g\|$$

$$\leq \underbrace{\|R_{\gamma(\delta)}\| \cdot \delta}_{\xrightarrow[\delta\to 0]{} 0} + \underbrace{\|R_\gamma(\delta) g - A^+ g\|}_{\xrightarrow[\delta\to 0]{} 0} \quad ,$$

nach Def. 4.1

d.h. $R_{\gamma(\delta)} g_\delta$ ist eine Approximation für $A^+ g$.

Praktisches Vorgehen: R_γ, g_δ, δ gegeben, wähle ein "geeignetes" γ und nehme dann $R_\gamma g_\delta$ als Näherung für $A^+ g$.

Die Wahl von γ stößt auf Schwierigkeiten:

γ zu klein: Auswertung von $R_\gamma g_\delta$ ist nicht stabil.

γ zu groß: $R_\gamma g_\delta$ nicht nahe bei $A^+ g$.

Gesucht ist optimales γ zwischen diesen Extrema, Strategien bei der Berechnung eines solchen γ_{opt} sind:

i) Versuch und Irrtum.

ii) Diskrepanz - Prinzip: Bestimme γ so, daß $\|A R_\gamma g_\delta - g_\delta\| \sim \delta$ gilt.

iii) Statistische Techniken.

4.1 Die Regularisierung von Tikhonov (auch: Tychonoff)-Phillips

Wir betrachten das Minimierungsproblem:

$$\min_{f \in X} \left(\|Af - g_\delta\|^2 + \gamma^2 \|f\|^2 \right)$$

und wollen zeigen, daß dessen Lösung f vermöge

$$R_\gamma g_\delta = f$$

eine Regularisierung von A definiert.

Dazu müssen die Eindeutigkeit von f und die in Def. 4.1 geforderten Eigenschaften von R_γ gezeigt werden.

Es gilt: f ist kleinste-Quadrate-Lösung von

$$\begin{pmatrix} A \\ \gamma \mathbb{1} \end{pmatrix} f = \begin{pmatrix} g_\delta \\ 0 \end{pmatrix} \quad ,$$

wobei wir $\begin{pmatrix} A \\ \gamma \mathbb{1} \end{pmatrix}$ als einen Operator vom Hilbertraum X in den Hilbertraum $Y \times X$ auffassen müssen, denn wir haben:

$$\left\| \begin{pmatrix} A \\ \gamma \mathbb{1} \end{pmatrix} f - \begin{pmatrix} g_\delta \\ 0 \end{pmatrix} \right\|^2_{Y \times X} = \|Af - g_\delta\|^2 + \gamma^2 \|f\|^2 \quad .$$

Nach Satz 2.2 ist f charakterisiert durch

$$\begin{pmatrix} A \\ \gamma \mathbb{1} \end{pmatrix}^* \begin{pmatrix} A \\ \gamma \mathbb{1} \end{pmatrix} f = \begin{pmatrix} A \\ \gamma \mathbb{1} \end{pmatrix}^* \begin{pmatrix} g_\delta \\ 0 \end{pmatrix} \quad .$$

Man rechnet sofort nach, daß $\begin{pmatrix} A \\ \gamma \mathbb{1} \end{pmatrix}^* = (A^*, \gamma \mathbb{1})$ gilt; für f gilt somit:

$$(A^*A + \gamma^2 \mathbb{1}) f = A^* g_\delta$$

D.h. f erfüllt die sogenannte regularisierte Normalgleichung. Wir wollen nun zeigen, daß f eindeutig bestimmt ist und stetig von g_δ abhängt, dazu das

LEMMA 4.1: $B : X \to X$ sei linear und stetig, ferner sei B selbstadjungiert $(B = B^*)$ und positiv definit, d.h. $\exists\ \gamma > 0$ s.d.: $(Bf,f) \geq \gamma(f,f)$ $\forall\ f \in X$, dann besitzt B eine auf ganz X definierte stetige Inverse B^{-1}, es gilt $\|B^{-1}\| \leq \frac{1}{\gamma}$.

BEWEIS: B ist injektiv, denn $Bf = 0 \Rightarrow \gamma \|f\|^2 = 0 \Rightarrow f = 0$, also $X = \overline{R(B)} + N(B) = \overline{R(B)}$.

Z.Zg. ist: $R(B)$ ist abgeschlossen:

Sei also $Bf_n \to g$, z.Zg. $g \in R(B)$.

Da $\{Bf_n\}$ konvergiert, ist es insbesondere eine Cauchy-Folge, es gilt:

C.-S.

$$\|Bf_n - Bf_m\| \cdot \|f_n - f_m\| \geq (Bf_n - Bf_m, f_n - f_m) \geq \gamma \|f_n - f_m\|^2$$

$$\Rightarrow \frac{1}{\gamma} \|Bf_n - Bf_m\| \geq \|f_n - f_m\| \quad ,$$

d.h. auch $\{f_n\}$ ist eine Cauchy-Folge, und da X vollständig ist, gilt $f_n \to f$, damit:

$$Bf = B \lim_n f_n = \lim_n Bf_n = g \quad , \quad \text{also} \quad g \in R(B)$$

Ferner gilt:

$$\|Bf\| \geq \gamma \|f\| \quad , \quad \text{setze} \quad h = Bf, \quad \text{so gilt:}$$

$$\gamma \|B^{-1}h\| \leq \|h\| \quad , \quad \text{also:} \quad \|B^{-1}\| \leq \frac{1}{\gamma} \quad .$$

■

Wir wenden nun Lemma 4.1 auf $B = A^*A + \gamma^2 \mathbb{1}$ an, dies ist möglich, da

$$(Bf,f) = (A^*Af + \gamma^2 f, f) = \underbrace{(Af,Af)}_{\geq 0} + \gamma^2 (f,f) \geq \gamma^2 (f,f)$$

gilt.

Wir haben also: $\|B^{-1}\| \leq \frac{1}{\gamma^2}$ und infolge $R_\gamma = B^{-1}A^*$ die Abschätzung:

$$\|R_\gamma\| \leq \frac{1}{\gamma^2} \cdot \|A^*\| \quad .$$

Wir haben bis jetzt also die Wohldefiniertheit und Stetigkeit von $R_\gamma = (A^*A + \gamma^2 \mathbb{1})^{-1}A^*$ gezeigt. Wir müssen nun noch die punktweise Konvergenz von R_γ gegen A^+ zeigen.

SATZ 4.2: Der Operator $A : X \to Y$ besitze eine Singulärwertzer-

legung, dann gilt:

$$R_\gamma g \xrightarrow{\gamma \to 0} A^+ g \quad ,$$

falls $g \in R(A) + R(A)^\perp$.

BEMERKUNG: Die Voraussetzung über die Existenz der Singulärwertzerlegung ist nicht notwendig, vereinfacht jedoch den Beweis.

BEWEIS: Es gilt

$$Af = \sum_{k=1}^{\infty} \sigma_k (f, f_k) g_k \quad , \quad A^* g = \sum_{k=1}^{\infty} \sigma_k (g, g_k) f_k$$

$$A^* A f = \sum_{k=1}^{\infty} \sigma_k^2 (f, f_k) f_k \quad ,$$

wobei $\{f_k\}$ ein ONS in X, $\{g_k\}$ ein ONS in Y ist.

Die regularisierten Normalgleichungen lauten:

$$A^* A f + \gamma^2 f = \sum_{k=1}^{\infty} \sigma_k^2 (f, f_k) f_k + \gamma^2 f = \sum_{k=1}^{\infty} \sigma_k (g_\delta, g_k) f_k = A^* g_\delta$$

$\Rightarrow$ f ist Linearkombination der f_k .

$$\Rightarrow f = \sum_{k=1}^{\infty} (f, f_k) f_k \quad ,$$

Fourierkoeffizienten

durch Einsetzen und anschließenden Koeffizientvergleich folgt:

$$\sigma_k^2 (f, f_k) + \gamma^2 (f, f_k) = \sigma_k (g_\delta, g_k) \quad , \quad \forall\, k$$

$$\Rightarrow (f, f_k) = \frac{\sigma_k}{\sigma_k^2 + \gamma^2} (g_\delta, g_k) = \frac{1}{\sigma_k} \frac{1}{\gamma^2 / \sigma_k^2 + 1} (g_\delta, g_k) \quad , \forall\, k$$

$$\Rightarrow R_\gamma g_\delta = f = \sum_{k=1}^{\infty} \frac{1}{\gamma^2 / \sigma_k^2 + 1} \frac{1}{\sigma_k} (g, g_k) f_k$$

Filter

Es gilt $A^+g = \sum_{k=1}^{\infty} \frac{1}{\sigma_k} (g,g_k) f_k$, also:

$$\| (R_\gamma - A^+) g \|^2 = \sum_{k=1}^{\infty} \underbrace{\left| \left(\frac{1}{\gamma^2/\sigma_k^2 + 1} - 1 \right) \frac{1}{\sigma_k} (g,g_k) \right|^2}_{=: c_k(\gamma)}$$

Wir haben folglich die Situation:

$$F(\gamma) := \sum_{k=1}^{\infty} c_k(\gamma)$$

mit $\lim_{\gamma \to 0} c_k(\gamma) = 0$.

Um hierbei Limes- und Summenbildung vertauschen zu können, müssen wir die gleichmäßige Konvergenz der Funktionenfolge auf $\mathbb{R}^+$ sicherstellen. Diese ist nach dem Weierstraßschen Majorantenkriterium gegeben, falls es (positive) reelle Zahlen C_k gibt, mit

$$\sup_{\gamma \in \mathbb{R}^+} |c_k(\gamma)| < C_k \quad \text{und} \quad \sum_{k=1}^{\infty} C_k < \infty \quad .$$

Wegen: $\left| \frac{1}{\gamma^2/\sigma_k + 1} - 1 \right| = \frac{\gamma^2/\sigma_k^2}{\gamma^2/\sigma_k^2 + 1} < 1 \quad \forall\, \gamma \in \mathbb{R}$,

gilt: $\sup_{\gamma \in \mathbb{R}^+} |c_k(\gamma)| \leq \left| \frac{1}{\sigma_k} (g,g_k) \right|^2 =: C_k$,

mit: $\sum_{k=1}^{\infty} C_k = \|A^+g\|^2 < \infty$.

$$\Rightarrow \lim_{\gamma \to 0} \| (R_\gamma - A^+) g \|^2 = 0$$

■

4.2 Verallgemeinerte Methode von Tikhonov-Phillips

Sei wieder $A : X \to Y$ ein linearer Operator zwischen Hilberträumen, wir betrachten das Minimierungsproblem:

$$\text{Min}(\|Af-g_\delta\|^2 + \gamma^2\|f\|_V^2) \quad ,$$

wobei $\|\cdot\|_V$ eine Norm auf X ist, für die eine Abschätzung der Form

$$\|f\|_X \leq c\|f\|_V$$

gilt, d.h. $\|\cdot\|_V$ ist stärker als $\|\cdot\|_X$.

BEISPIEL: Sei $X = L_2(0,1)$ mit $\|f\|_X = \left(\int_0^1 |f(x)|^2 dx\right)^{1/2}$,

setze $\|f\|_V := \left(\int_0^1 |f(x)|^2 + |f'(x)|^2 dx\right)^{1/2}$.

Man sieht an diesem Beispiel, daß $\|\cdot\|_V$ i.a. nur auf einem Teilraum von X definiert ist und folglich obige Minimierungsaufgabe auf diesen Teilraum einzuschränken ist.

Bezeichnen wir die Lösung dieser Aufgabe mit f, so gibt der folgende Satz Auskunft über den Fehler $f - f_\gamma$:

SATZ 4.3: Es sei $Af = g$, $\|f\|_V \leq \rho$, $\|g-g_\delta\| \leq \delta$, dann gilt:

$$\|f-f_\gamma\| \leq 2L\alpha(\gamma,1)$$

mit $$\alpha(\delta,\rho) = \sup\{\|f\| : \|Af\| \leq \delta, \|f\|_V \leq \rho\} \quad ,$$

sowie $$L = \sqrt{\rho^2+\delta^2/\gamma^2} \quad .$$

BEWEIS:

$$\|Af_\gamma - g_\delta\|^2 + \gamma^2 \|f_\gamma\|_V^2 \leq \|Af - g_\delta\|^2 + \gamma^2 \|f\|_V^2$$

$$\leq \delta^2 + \gamma^2 \rho^2$$

$$\Rightarrow \|Af_\gamma - g_\delta\| \leq \sqrt{\delta^2 + \gamma^2 \cdot \rho^2}\,, \quad \|f_\gamma\|_V \leq \sqrt{\frac{\delta^2 + \gamma^2 \rho^2}{\gamma^2}} = L$$

$$\Rightarrow \|A(f_\gamma - f)\| \leq \|Af_\gamma - g_\delta\| + \|g_\delta - g\|$$

$$\leq \sqrt{\delta^2 + \gamma^2 \rho^2} + \delta$$

$$\leq 2\sqrt{\delta^2 + \gamma^2 \rho^2} = 2\gamma L\,,$$

$$\|f_\gamma - f\|_V \leq \|f_\gamma\|_V + \|f\|_V \leq L + \rho \leq 2L$$

$$\Rightarrow \|f_\gamma - f\| \leq \alpha(2\gamma L, 2L) = 2L\alpha(\gamma, 1)\,,$$

denn die Funktion α erfüllt nach Definition die Homogenitätsbeziehung: $\alpha(t\delta, t\rho) = t\alpha(\delta, \rho) \quad \forall\, t \geq 0$. ■

Anwendung für den Fall $\gamma = \delta/\rho$:

$$\|f_\gamma - f\| \leq 2\sqrt{2}\,\rho\,\alpha\left(\frac{\delta}{\rho}, 1\right) = 2\sqrt{2}\,\alpha(\delta, \rho)$$

Praktisch hat man es, nach erfolgter Diskretisierung, oft mit (n,m)-Matrizen A zu tun; die diskretisierte Version der im obigen Beispiel vorgestellten $\|\cdot\|_V$-Norm wäre dann:

$$\|f\|_V^2 = h\left\{ \sum_{j=0}^{m-1} \left(\frac{f_j - f_{j+1}}{h} \right)^2 + \sum_{j=0}^{m-1} f_j^2 \right\}\,,$$

dies ist eine quadratische Form in f. Es existiert folglich eine symmetrische, streng positiv definite Matrix V mit

$$\|f\|_V^2 = (Vf, f) \quad .$$

Die regularisierten Normalgleichungen hätten somit die Form:

$$(A^*A + \gamma^2 V) f_\gamma = A^* g_\delta \quad .$$

4.3 Die abgeschnittene SVD

Sei $A : X \to Y$ ein linearer, stetiger Operator mit SVD, als Regularisierung von A wählen wir:

$$R_\gamma g = \sum_{\sigma_k \geq \gamma} \frac{1}{\sigma_k} (g, g_k) f_k \quad .$$

Es gilt somit:

$$\|R_\gamma g\|^2 = \sum_{\sigma_k \geq \gamma} \frac{1}{\sigma_k^2} |(g, g_k)|^2 \leq \frac{1}{\gamma^2} \sum_{\sigma_k \geq \gamma} |(g, g_k)|^2 \leq \frac{1}{\gamma^2} \|g\|^2$$

$$\Rightarrow \|R_\gamma\| \leq \frac{1}{\gamma} \quad .$$

$$R_\gamma g_\delta - A^+ g = \sum_{\sigma_k \geq \gamma} \frac{1}{\sigma_k} (g_\delta, g_k) f_k - \sum_{k=1}^{\infty} \frac{1}{\sigma_k} (g, g_k) f_k$$

$$= \sum_{\sigma_k \geq \gamma} \frac{1}{\sigma_k} (g_\delta - g, g_k) f_k - \sum_{\sigma_k < \gamma} \frac{1}{\sigma_k} (g, g_k) f_k$$

$$\|R_\gamma g_\delta - A^+ g\|^2 = \underbrace{\sum_{\sigma_k \geq \gamma} \frac{1}{\sigma_k^2} |(g_\delta - g, g_k)|^2}_{\leq \delta^2 \sum_{\sigma_k \geq \gamma} \frac{1}{\sigma_k^2}} + \underbrace{\sum_{\sigma_k < \gamma} \frac{1}{\sigma_k^2} |(g, g_k)|^2}_{\text{Fehler durch Abschneiden}}$$

Der Beitrag von f_k zu $A^+ g_\delta$ ist:

$$\frac{1}{\sigma_k} (g_\delta, g_k) f_k \quad ,$$

dies ist stabil berechenbar, falls $\sigma_k \sim 1$,

und instabil berechenbar, falls $\sigma_k << 1$.

Bei glatten Funktionen g kann man i.a. hoffen, daß für sehr kleine Werte von σ_k auch das innere Produkt (g, g_k) nahezu gleich Null ist, falls dann

$$\left| \frac{(g, g_k)}{\sigma_k} \right| << 1 \quad ,$$

läßt man $\frac{1}{\sigma_k} (g, g_k) f_k$ in der Aufsummierung weg.

4.4 Digitales Filtern

Wir "definieren" einen Filter F durch die etwas vagen Anforderungen:

$$F(\sigma) = \begin{cases} \sim 1 & , \text{ für große } \sigma \\ \sim 0 & , \text{ für kleine } \sigma \end{cases} ,$$

und wählen als Regularisierung:

$$R_\gamma g = \sum_{k=1} F(\sigma_k) \frac{1}{\sigma_k} (g, g_k) f_k \quad .$$

Nimmt man

$$F(\sigma) = \left(1 + \frac{\gamma^2}{\sigma^2}\right)^{-1} , \quad \gamma \text{ klein},$$

Der Beitrag von f_k zu $A^+ g_\delta$ ist:

$$\frac{1}{\sigma_k} (g_\delta, g_k) f_k \quad ,$$

dies ist stabil berechenbar, falls $\sigma_k \sim 1$,

und instabil berechenbar, falls $\sigma_k << 1$.

Bei glatten Funktionen g kann man i.a. hoffen, daß für sehr kleine Werte von σ_k auch das innere Produkt (g, g_k) nahezu gleich Null ist, falls dann

$$\left| \frac{(g, g_k)}{\sigma_k} \right| << 1 \quad ,$$

läßt man $\frac{1}{\sigma_k} (g, g_k) f_k$ in der Aufsummierung weg.

4.4 Digitales Filtern

Wir "definieren" einen Filter F durch die etwas vagen Anforderungen:

$$F(\sigma) = \begin{cases} \sim 1 & , \text{ für große } \sigma \\ \sim 0 & , \text{ für kleine } \sigma \end{cases} ,$$

und wählen als Regularisierung:

$$R_\gamma g = \sum_{k=1} F(\sigma_k) \frac{1}{\sigma_k} (g, g_k) f_k \quad .$$

Nimmt man

$$F(\sigma) = \left(1 + \frac{\gamma^2}{\sigma^2}\right)^{-1} , \quad \gamma \text{ klein},$$

als Filter, so ist das exakt die Tykhonov-Phillips-Regularisierung (vgl. Beweis von Satz 4.2).

Während

$$F(\sigma) = \begin{cases} 1 & , \ \sigma \geq \gamma \\ 0 & , \ \sigma < \gamma \end{cases}$$

auf die abgeschnittene SVD führt.

4.5 Iterative Methoden

Durch

$$f^{t+1} = B_t f^t + C_t g$$

mit linearen Operatoren B_t, C_t sei ein Iterationsverfahren erklärt.

Wir nehmen an, daß

$$f^t \xrightarrow{t\to\infty} A^+ g \quad , \qquad \forall\, g \in R(A) + R(A)^\perp$$

gilt.

Iteriert man mit der fehlerhaften rechten Seite

$$f^{t+1} = B_t f^t + C_t g_\delta \quad ,$$

so wird

$$f_t \to A^+ g$$

i.a. n i c h t gelten.

Es wird in der Regel jedoch eine sogenannte Semikonvergenz ein-

treten, d.h. das Verfahren wird A^+g nach k Schritten gut approximieren, sich dann jedoch (für $t > k$) wieder verschlechtern.

Als Regularisierung wählt man

$$R_\gamma g = f^k \quad .$$

BEISPIEL: Landweber-Iteration

$$f^{t+1} = (\mathbb{1} - \omega^2 A^*A) f^t + \omega^2 A^*g \ , \quad f^o = 0 \quad .$$

Dieses Verfahren ist konvergent, falls gilt:

$$\omega^2 \cdot \max_k \sigma_k^2 \in]0,2[\quad .$$

BEWEIS dazu:

Sei $$f^t = \sum_{k=1}^{\infty} c_k^t f_k$$

$$\Rightarrow \sum_{k=1}^{\infty} c_k^{t+1} f_k = \sum_{k=1}^{\infty} c_k^t (1 - \omega^2\sigma_k^2) f_k + \omega^2 \sum_{k=1}^{\infty} \sigma_k (g,g_k) f_k$$

$$\Rightarrow c_k^{t+1} = \underbrace{(1 - \omega^2\sigma_k^2)}_{\in\,]-1,1[} c_k^t + \sigma_k (g,g_k)\omega^2 \quad ,$$

d.h. das Verfahren ist für jeden Koeffizienten kontrahierend, folglich gilt:

$$c_k^t \xrightarrow{t\to\infty} c_k$$

mit $$c_k = (1 - \omega^2\sigma_k^2) c_k + \sigma_k (g,g_k)\omega^2$$

$$\Rightarrow \quad c_k = \frac{1}{\sigma_k} (g,g_k) \quad .$$

Wir haben damit die koeffizientenweise Konvergenz des Verfahrens gezeigt. Um die gewünschte Konvergenz

$$f^t \to A^+ g = f$$

zu beweisen, müßten wir noch zeigen, daß wir die Limesbildung $(t\to\infty)$ und die Summenbildung miteinander vertauschen dürfen. Dies geht analog zu dem Vorgehen im Beweis von Satz 4.2, und wir verzichten an dieser Stelle darauf.

Es gilt:

$$f - f^t = \sum_{k=1}^{\infty} (c_k - c_k^t) f_k$$

mit $$(c_k - c_k^{t+1}) = (1 - \omega^2\sigma_k^2)(c_k - c_k^t)$$

$$\Rightarrow \quad c_k - c_k^{t+1} = (1 - \omega^2\sigma_k^2)^t c_k \quad , \quad (\text{wegen } c_k^o = 0) \quad .$$

Also: $$f - f^t = \sum_{k=1}^{\infty} (1 - \omega^2\sigma_k^2)^t \frac{1}{\sigma_k} (g, g_k) f_k$$

$$\Rightarrow \quad f^t = \sum_{k=1}^{\infty} (1 - (1 - \omega^2\sigma_k^2)^t) \frac{1}{\sigma_k} (g, g_k) f_k$$

D.h. es liegt digitales Filtern mit der Filterfunktion

$$F^t(\sigma) = 1 - (1 - \omega^2\sigma^2)^t$$

vor.

Für die großen Singulärwerte ist $\omega \cdot \sigma_k$ nur etwas kleiner als 1, $F(\sigma_k)$ also nahezu gleich 1, für kleine Singulärwerte ist $(1 - \omega^2\sigma_k^2)^t$ nahezu gleich 1 und damit $F(\sigma_k) \sim 0$.

Anhand der oben entwickelten Darstellung für $f - f^t$ sieht man sehr leicht den Grund für die Semikonvergenz:

ist σ_k so, daß gilt $\frac{1}{2} < \omega^2\sigma_k^2 < 1$, so fällt die zugehörige Komponente wie 2^{-t},

ist σ_k so, daß gilt $\omega^2\sigma_k^2 << 1$, so bleibt die zugehörige Komponente praktisch unverändert.

4.6 Regularisierung durch Diskretisierung

Durch Diskretisierung überführt man das ursprüngliche Problem in der Regel in ein System endlich vieler linearer Gleichungen. Die Inverse einer Matrix ist aber (falls existent) stets stetig, so daß wir allein durch Diskretisierung einen Regularisierungseffekt erzielen können.

BEISPIEL:

$$\int_a^b K(x,y)\,f(y)\,dy = g(x) \qquad , \quad c \leq x \leq d$$

Wir diskretisieren das Integral durch eine Quadraturregel:

$$\int_a^b f(y)\,dy \approx h \sum_{j=0}^{m} w_j f(y_j) \qquad , \quad y_j = a + h \cdot j, \quad h = \frac{b-a}{m}$$

mit von h unabhängigen Gewichten w_j.

Die Gewichte können wir z.B. wie folgt wählen:

$w_j = \frac{1}{2}, 1, 1, \ldots, 1, \frac{1}{2}$ - Trapezregel

$w_j = \frac{1}{3}\{1,4,2,4,\ldots,2,4,1\}$ - Simpsonregel (m muß gerade sein)

Die diskretisierte Integralgleichung nimmt damit die folgende Form an:

$$h \sum_{j=0}^{m} K(x_i,y_j) w_j f_j = g(x_i) =: g_i \quad , \quad i = 0,\ldots,m \quad .$$

Die Frage ist nun, ob sich durch Verfeinerung der Diskretisierung (vergrößern von m) immer bessere Näherungen f_j an die exakte Lösung $f(y_j)$ erzielen lassen.

SATZ 4.4: Das Quadraturverfahren ist i.a. nicht konvergent.

BEWEIS: Sei

$$F_m = \begin{pmatrix} f_0 \\ \vdots \\ f_m \end{pmatrix} , \quad G_m = \begin{pmatrix} g_0 \\ \vdots \\ g_m \end{pmatrix} , \quad A_m = (K(x_i,y_j))_{i,j=0,\ldots,m}$$

$$W_m = \operatorname{diag}(w_0,\ldots,w_m) .$$

Das Quadraturverfahren liefert das lineare Gleichungssystem:

$$h\, A_m\, W_m\, F_m = G_m$$

Annahme: A_m sei invertierbar $\forall\, m \in \mathbb{N}$

$$\Rightarrow W_m\, F_m = \frac{1}{h}\, A_m^{-1}\, G_m$$

Die rechte Seite ist von der Wahl der w_j unabhängig, also gilt insbesondere:

$$W_m^{\text{Trapez}} \cdot F_m^{\text{Trapez}} = W_m^{\text{Simpson}}\, F_m^{\text{Simpson}}$$

$$\Rightarrow \qquad F_m^{\text{Trapez}} = \operatorname{diag}\left(\tfrac{2}{3}, \tfrac{4}{3}, \tfrac{2}{3}, \ldots, \tfrac{4}{3}, \tfrac{2}{3}\right) \cdot F_m^{\text{Simpson}}$$

Falls F_m^{Simpson} gegen die richtige Lösung konvergiert, kann dies für F_m^{Trapez} nicht der Fall sein. ∎

Wir können das Quadraturverfahren als Spezialfall einer größeren Klasse von Verfahren ansehen, die wir nun behandeln wollen. Es handelt sich dabei um die sogenannten Projektionsverfahren:

$A : X \to Y$ sei ein linearer Operator zwischen Hilberträumen, zu lösen sei: $Af = g$.

Wir geben uns einen Unterraum:

$$V_n = sp\{v_1,\ldots,v_n\} \subseteq X$$

in X und n lineare Funktionale

$$\psi_1,\ldots,\psi_n$$

auf Y vor.

Von unserer Näherungslösung $f_n \in V_n$ verlangen wir:

$$\psi_i(Af_n - g) = 0 \qquad , \quad i = 1,\ldots,n \quad .$$

Dies sind n Gleichungen mit ebensovielen Unbekannten.

Wir schreiben:

$$f_n = \sum_{j=1}^{n} x_j v_j$$

und setzen ein:

$$\psi_i(A(\sum_{j=1}^{n} x_j v_j) - g) = 0 \quad , \quad i = 1,\ldots,n$$

$$\Rightarrow \sum_{j=1}^{n} (\psi_i A v_j) \cdot x_j - \psi_i g = 0 \quad , \quad i = 1,\ldots,n$$

Das heißt, wir müssen ein lineares Gleichungssystem

$$A^D x = y$$

lösen, mit

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad y = \begin{pmatrix} \psi_1(g) \\ \vdots \\ \psi_n(g) \end{pmatrix}, \quad A^D = (\psi_i A v_j)_{i,j=1,\ldots,n} .$$

BEISPIELE:

i) $Y = C[c,d]$

$\psi_i g = g(x_i)$, $x_i \in [c,d]$, d.h. die Funktionale seien Punktauswertungen.

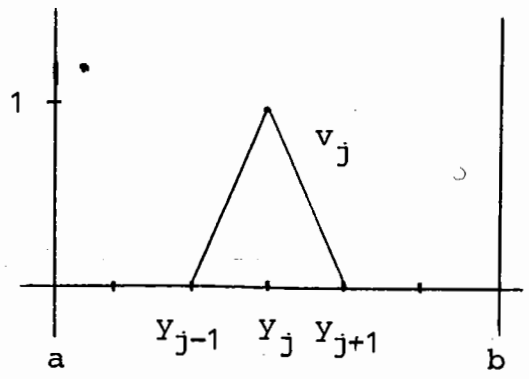
Man nennt dann

$$f_n \in V_n : (Af_n - g)(x_i) = 0 \qquad , \quad i = 1,\ldots,n$$

ein Kollokationsverfahren (nach L. Collatz ∿ 1930).

Sei speziell

$X = L_2(a,b)$ und die v_j Funktionen vom Typ:



V_n ist dann also der Raum der Polygonzüge (lineare Splines) zu den Unterteilungspunkten $y_1,\ldots,y_n$.

Der Integraloperator

$$Af(x) = \int_a^b K(x,y) f(y)\,dy$$

wird diskretisiert durch die Matrix

$$A^D = \Big(\underbrace{\int_a^b K(x_i,y) v_j(y)\,dy}_{\approx K(x_i,y_j)\cdot h,\ \text{mit}\ h = \int_{y_{i-1}}^{y_{i+1}} v_j(y)\,dy} \Big)_{i,j=1,\ldots,n}$$

ii) Momentenverfahren

Y - Hilbertraum

$\psi_i g = (f_i, g)$, zu lösen $(f_i, Af_n - g) = 0$, $i = 1,\ldots,n$

iii) Galerkin-Verfahren

X = Y - Hilberträume

$\psi_i g = (v_i, g)$

und $V_n = sp\{v_1,\ldots,v_n\}$,

also: $f_n \in V_n : (v_i, Af_n - g) = 0$, $i = 1,\ldots,n$,

d.h. bestimme $f_n \in V_n$, s.d. $Af_n - g \perp$ zu V_n.

Wir wollen diese Verfahren einheitlich durch Projektoren beschreiben.

Wir haben Gleichungen vom Typ

$$\psi_i (Af_n - g) = 0 \quad , \quad i = 1,\ldots,n \quad ,$$

$$f_n \in V_n$$

mit $V_n = sp(v_1,\ldots,v_n)$ zu lösen. Wir nehmen dabei an, daß $A^D = (\psi_i A v_j)_{i,j=1,\ldots,n}$ regulär ist, d.h. die Gleichungen $\forall\, g \in Y$ eindeutig lösbar sind.

Vermöge

$$P_n f \longmapsto f_n \quad ,$$

mit f_n ist (eindeutig bestimmte) Lösung von $\psi_i(Af_n - Af) = 0$, $i = 1,\ldots,n$, wird also ein linearer Operator

$$P_n : X \to V_n$$

definiert. Wegen $\psi_i(Af_n - Af_n) = 0 \quad \forall\, i$ gilt trivialerweise

$$P_n^2 f = P_n f \qquad \forall\, f \in X \quad ,$$

d.h. P_n ist ein Projektor.

Ebenso läßt sich ein linearer Operator

$$Q_n : Y \to V_n$$

durch:

$$Q_n g = f_n$$

f_n ist (eindeutig bestimmte) Lösung von $\psi_i(Af_n - g) = 0$, $i = 1,\ldots,n$, definieren.

Wir zeigen

SATZ 4.5: Es gelte

i) $\bigcup_{i=1}^{n} V_n$ ist dicht in X

ii) Das Projektionsverfahren sei durchführbar, d.h. $(\psi_i A v_j)$ sei regulär

iii) A sei nicht injektiv oder A^{-1} sei unstetig, dann gilt:

$$\|Q_n\| \xrightarrow{n\to\infty} \infty \quad .$$

BEWEIS:

$$\|Q_n\| = \sup_{g \neq 0} \frac{\|Q_n g\|}{\|g\|} \geq \sup_{\substack{v \in V_n \\ Av \neq 0}} \frac{\|Q_n Av\|}{\|Av\|} = \sup_{\substack{v \in V_n \\ Av \neq 0}} \frac{\|v\|}{\|Av\|} =: \alpha_n$$

Aufgrund von iii): $\forall\, \varepsilon > 0 \;\; \exists\, f \in X : \|f\| = 1, \|Af\| \leq \varepsilon$

" " i): $\exists$ eine Folge $\{u_n\}_n$ mit $u_n \in V_n \;\forall\, n$

und $u_n \xrightarrow{n\to\infty} f$

insbesondere $\|u_n\| \xrightarrow{n\to\infty} 1$.

Da A stetig ist, gilt

$$Au_n \xrightarrow{n} Af$$

$$\|Au_n\| \to \|Af\| \leq \varepsilon$$

$$\Rightarrow \alpha_n \geq \frac{\|u_n\|}{\|Au_n\|} \xrightarrow{n} \frac{1}{\varepsilon} \quad .$$

und da ε beliebig klein gewählt werden kann:

$$\alpha_n \xrightarrow{n} \infty$$

■

SATZ 4.6: Das Projektionsverfahren sei $\forall\, n$ durchführbar, ferner sei

$$Af = g \quad , \quad \|g - g_\delta\| \leq \delta \; .$$

Sei $f_{n,\delta}$ die Näherung des Projektionsverfahrens für g_δ, d.h.

$$f_{n,\delta} = Q_n g_\delta \quad ,$$

dann gilt die Abschätzung:

$$\|f - f_{n,\delta}\| \leq (1 + \|P_n\|) d(f, V_n) + \|Q_n\| \delta \quad ,$$

mit $d(f, V_n) = \min_{v \in V_n} \|f - v\|$.

BEWEIS: Sei $v \in V_n$ und $f_n = Q_n g$, wegen $P_n|_{V_n} = \mathbb{1}$ gilt

$$v - f_n = P_n(v - f_n) \Rightarrow \|v-f_n\| \leq \|P_n\| \, \|v-f_n\|$$

$$\Rightarrow \|f-f_n\| \leq \|f-v\| + \|v-f_n\| \leq (1 + \|P_n\|)\|v-f_n\| \quad , \forall v \in V_n$$

$$\Rightarrow \|f-f_n\| \leq (1 + \|P_n\|)d(f,V_n)$$

$$\Rightarrow \|f-f_{n,\delta}\| \leq \|f-f_n\| + \underbrace{\|f_n-f_{n,\delta}\|}_{=Q_n(g-g_\delta)} \leq (1 + \|P_n\|)d(f,V_n) + \|Q_n\|\delta$$

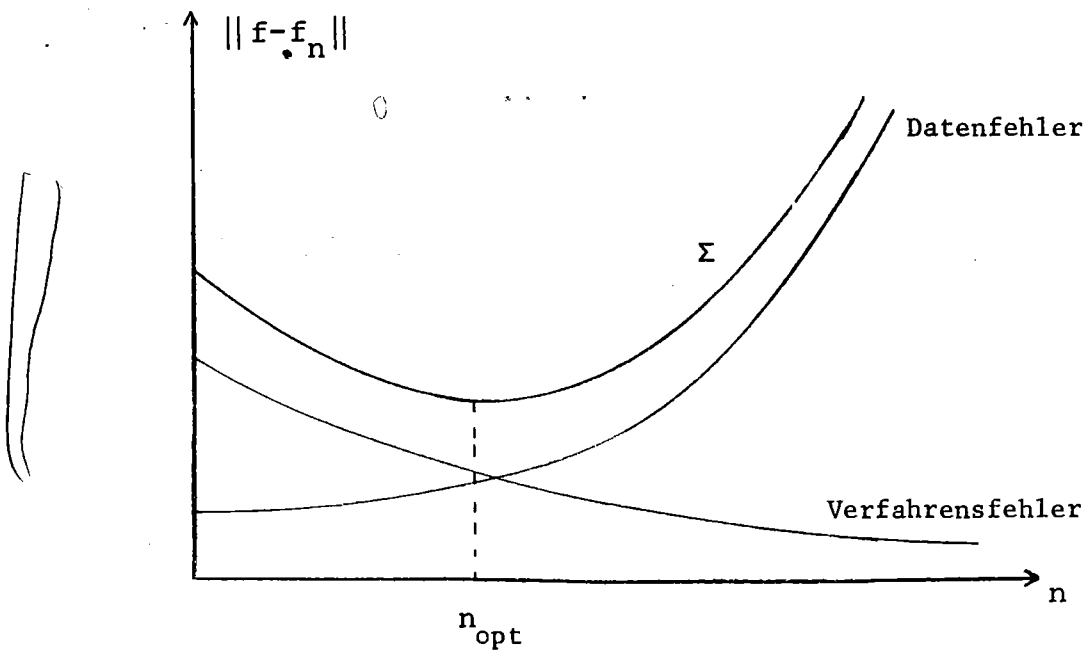
■

Dieses Ergebnis läßt sich wie folgt interpretieren:

$(1 + \|P_n\|)d(f,V_n)$ - Verfahrensfehler

$\|Q_n\|\delta$ - Einfluß des Datenfehlers

Der Verlauf sieht in etwa so aus:



BEISPIELE:

i) $X = C[0,1]$

V_n - Raum der Polygonzüge zur äquidistanten Unterteilung $h = \frac{1}{n}$.

Man kann zeigen (vgl. Prama I):

$$d(f,V_n) \leq \frac{1}{8} n^{-2} \|f''\| \ ,$$

wobei wir $f \in C^2$ als a-priori - Information voraussetzen müssen.

Der Nachweis von

$$\|P_n\| \leq C \qquad \forall\, n$$

ist hingegen i.a. schwierig. Für den Fall

$$(Af)(x) = \int_0^x f(y)\,dy \quad , \quad \psi_i g = (v_i, g) \text{ (also Galerkin-verfahren)}$$

kann gezeigt werden:

$$\|P_n\| \leq C_1 \qquad \forall\, n$$

$$\|Q_2\|_\bullet \leq C_2\, n$$

$$\Rightarrow \|f - f_{n,\delta}\| \leq \frac{1}{8} (1 + C_1)\, n^{-2} \|f''\| + \delta \cdot n$$

Wähle nun $n = [\delta^{2/3}]$, so gilt:

$$\|f - f_{n,\delta}\| \sim O(\delta^{2/3})$$

ii) $X = C[a,b]$, $f \in C^\infty[a,b]$

Um die, gegenüber Beispiel i), bessere a-priori - Information ausnutzen zu können, müssen wir uns nach einem Teilraum V_n mit besseren Approximationseigenschaften umsehen. Denn die

obige Abschätzung für lineare Splines (Polygonzüge) ist (bzgl. der Ordnung) bestmöglich, d.h. auch C^∞ - Funktionen lassen sich durch Polygonzüge der Stützweite $h = \frac{1}{n}$ nur bis zu einer Ordnung von n^{-2} approximieren.

Wählt man nun jedoch

$V_n = sp\{1, x, \ldots, x^n\}$,

so kann man zeigen, daß die Konvergenz

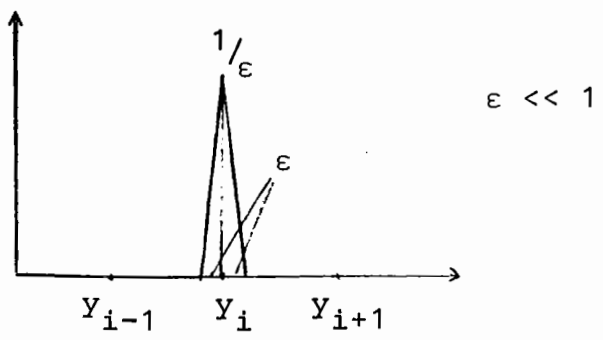
$d(f, V_n) \to 0$

schneller als jede Potenz von n ist.

In diesem Zusammenhang wird auch die negative Aussage von Satz 4.4 klarer:

Wir können das Quadraturverfahren als ein Projektionsverfahren mit $V_n = \{\delta_{y_1}, \ldots, \delta_{y_n}\}$, wobei δ_{y_i} die Deltafunktion zum Punkte y_i ist, interpretieren.

Die Deltafunktionen (eigentlich: Deltadistributionen) δ_{y_i} denke man sich dabei näherungsweise ersetzt durch folgende Funktionen:



Wir können also nicht hoffen, f durch Funktion aus V_n gut zu approximieren.

5. STOCHASTISCHE METHODEN

5.1 Bayes - Schätzung

a) Bedingte Erwartung

Y_1, Y_2 seien vektorwertige Zufallsvariablen (ZV) mit gemeinsamer Verteilung:

$$P(Y_1 \leq \eta_1, Y_2 \leq \eta_2) = \int_{-\infty}^{\eta_1} \int_{-\infty}^{\eta_2} f(y_1, y_2)\, dy_1\, dy_2$$

Die bedingte Wahrscheinlichkeit für $Y_1 \leq \eta_1$ unter der Annahme $\zeta_2 \leq Y_2 \leq \eta_2$ ist dann gegeben durch:

$$P(Y_1 \leq \eta_1 \mid \zeta_2 < Y_2 \leq \eta_2) = \frac{P(Y_1 \leq \eta_1,\ \zeta_2 < Y_2 \leq \eta_2)}{P(\zeta_2 \leq Y_2 \leq \eta_2)}$$

$$= \frac{\int_{-\infty}^{\eta_1} \int_{\zeta_2}^{\eta_2} f(y_1, y_2)\, dy_2 dy_1}{\int_{\mathbb{R}} \int_{\zeta_2}^{\eta_2} f(y_1, y_2)\, dy_2 dy_1} \quad .$$

Durch den Grenzübergang $\zeta_2 \to \eta_2$ erhält man:

$$P(Y_1 \leq \eta_1 \mid Y_2 = \eta_2) = \frac{\int_{-\infty}^{\eta_1} f(y_1, \eta_2)\, dy_1}{\int_{\mathbb{R}} f(y_1, \eta_2)\, dy_1}$$

Damit ergibt sich für die bedingte Verteilung(sdichte):

$$f(\cdot \mid y_2 = \eta_2) = \frac{f(\cdot, \eta_2)}{\int_{\mathbb{R}} f(y_1, \eta_2)\, dy_1}$$

Eine andere Schreibweise ist:

$$f(\cdot \mid y_2) = f(\cdot, y_2) \Big/ \int_{\mathbb{R}} f(y_1, y_2)\, dy_1$$

Der Erwartungswert für Y_1 unter $Y_2 = \eta_2$ ist durch

$$E(Y_1 \mid Y_2 = \eta_2) = \int_{\mathbb{R}} y_1 f(y_1 \mid y_2 = \eta_2)\, dy_1$$

bzw.

$$E(Y_1 \mid Y_2) = \int_{-\infty}^{\infty} y_1 f(y_1 \mid y_2)\, dy_1$$

gegeben.

b) Normalverteilung in $\mathbb{R}^n$

Besitzt Y die Verteilungsfunktion:

$$f(y) = (2\pi)^{-n} \, |A|^{1/2} \, e^{-1/2 (y-\mu)^T A (y-\mu)} \quad ,$$

so sagt man Y ist (μ, A^{-1}) - normalverteilt.

Hierbei ist:

μ - der Mittelwert und

A^{-1} - die Kovarianzmatrix

Wir wollen nun in der Situation

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \quad , \quad Y_1 \in \mathbb{R}^p \quad , \quad Y_2 \in \mathbb{R}^q \quad , \quad p + q = n$$

den Erwartungswert $E(Y_1 \mid Y_2)$ berechnen.

LEMMA 5.1: Sei Y eine n-dimensionale (μ, K^{-1})-normalverteilte ZV.

Sei $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $K = \left(\begin{array}{c|c} K_{11} & K_{12} \\ \hline K_{12}^T & K_{22} \end{array}\right)$.

Dann ist $f(\cdot \mid y_2)$ eine Normalverteilung

mit Mittelwert $\mu_1 + K_{11} K_{22}^{-1}(y_2 - \mu_2)$

und Kovarianz $K_{11} - K_{12} K_{22}^{-1} K_{12}^T$.

BEWEIS: Setze $z_i = (y_i - \mu_i)$, $i = 1,2$, dann gilt:

$$f(y_1 \mid y_2) = \frac{e^{-1/2(z_1^T A_{11} z_1 + z_2^T A_{22} z_2 + 2z_1^T A_{12} z_2)}}{\int_{\mathbb{R}^p} e^{-1/2(z_1^T A_{11} z_1 + A_{22} z_2 + 2z_1^T A_{12} z_2)} dy_1} ,$$

wobei wir $K^{-1} = A = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{pmatrix}$ geschrieben haben.

Durch einfaches Ausmultiplizieren verifiziert man

$$z_1^T A_{11} z_1 + 2z_1^T A_{12} z_2 = (z_1 + A_{11}^{-1} A_{12} z_2)^T A_{11} (z_1 + A_{11}^{-1} A_{12} z_2) - z_2 A_{12}^T A_{11}^{-1} A_{12} z_2 .$$

Setzen wir das ein und kürzen die nur von z_2 (d.h. y_2) abhängigen Terme im Zähler sowie Nenner, erhalten wir mit der Abkürzung $\tilde{\mu}_1 = \mu_1 - A_{11}^{-1} A_{12}(y_2 - \mu_2)$:

$$f(y_1 \mid y_2) = \frac{e^{-1/2(y_1 - \tilde{\mu}_1)^T A_{11} (y_1 - \tilde{\mu}_1)}}{\int_{\mathbb{R}^p} e^{-1/2(y_1 - \tilde{\mu}_1)^T A_{11} (y_1 - \tilde{\mu}_1)} dy_1}$$

Wir haben also eine $(\tilde{\mu}, A_{11}^{-1})$ - Normalverteilung erhalten.

(Substituiert man im Nennerintegral y_1 für $y_1 - \tilde{\mu}_1$ und anschließend y_1 für $A^{1/2} y_1$, bekommt man für das Nennerintegral:

$$\int_{\mathbb{R}^p} e^{-1/2 \| y_1 \|^2} dy_1 \cdot |A^{-1/2}| = (2\pi)^p \cdot |A|^{-1/2} \quad)$$

Aus $K = A^{-1}$ folgt:

$$A_{11}^{-1} = K_{11} - K_{12}\, K_{12}^{-1}\, K_{12}^{T}$$

sowie

$$A_{12} = - A_{11}\, K_{12}\, K_{22}^{-1} ,$$

also: $$A_{11}^{-1}\, A_{12} = - K_{12}\, K_{22}^{-1} ,$$

und damit:

$$\tilde{\mu} = \mu_1 + K_{12}\, K_{22}^{-1}(y_2 - \mu_2) \quad . \qquad \blacksquare$$

LEMMA 5.2: Y sei (μ,K) - normalverteilt, es gelte

$$Z = C\,Y \text{ mit einer regulären Matrix } C .$$

Dann gilt: Z ist $(C\mu, C\,K\,C^{T})$ - normalverteilt.

BEWEIS: Sei f die Verteilungsdichte für Y, so gilt allgemein

$$P(Z \in \Omega) \overset{Z=CY}{=} P(CY \in \Omega) = P(Y \in C^{-1}\Omega)$$

$$= \int_{y \in C^{-1}\Omega} f(y)\,dy = |C^{-1}| \int_{\Omega} f(C^{-1}z)\,dz \quad ,$$

d.h. $|C^{-1}|f(C^{-1}z)$ ist die Verteilungsdichte für Z.

Aus

$$f(y) = (2\pi)^{-n}|K|^{-1/2}\, e^{-1/2(y-\mu)^{T}K^{-1}(y-\mu)}$$

folgt:

$$f(C^{-1}z) = (2\pi)^{-n}|K|^{-1/2}\, e^{-1/2(C^{-1}z-\mu)^{T}K^{-1}(C^{-1}z-\mu)}$$

$$= (2\pi)^{-n}|K|^{-1/2}\, e^{-1/2(z-C\mu)^T C^{-T}K^{-1}C^{-1}(z-C\mu)}$$

Wegen $C^{-T}K^{-1}C^{-1} = (CKC^T)^{-1}$ folgt die Behauptung.

■

c) Bayes - Schätzung

DEFINITION: Y_1, Y_2 seien ZV mit gemeinsamer Verteilung, dann heißt $E(Y_1 \mid Y_2)$ Bayes-Schätzung von Y_1.

d) Anwendung der Bayes-Schätzung

A sei eine q,p - Matrix

f eine p-dimensionale $(\bar{f},F)$ - normalverteilte ZV

n eine q-dimensionale $(0,\sigma\,\mathbb{1})$-normalverteilte ZV.

Wir betrachten das Problem

$$Af = g + n \quad .$$

Mit der ZV n wollen wir das Rauschen (engl.: noise) in den gemessenen Daten statistisch modellieren.
Es ist daher sinnvoll, als Mittelwert 0 und als Kovarianz - matrix $\sigma\mathbb{1}$ (d.h. unkorrelierte Meßfehler) anzunehmen.

Wir führen die ZV

$$Y = \begin{pmatrix} f \\ n \end{pmatrix}$$

ein. Wir machen nun die (etwas kühne) Annahme, daß f und n unkorreliert sind, dann ist

Y normalverteilt mit

Mittelwert $\begin{pmatrix} \bar{f} \\ 0 \end{pmatrix}$ und Kovarianz $\begin{pmatrix} F & 0 \\ 0 & \sigma^2\mathbb{1} \end{pmatrix}$.

Definieren wir nun

$$Z = C\,Y \quad \text{mit} \quad C = \begin{pmatrix} \mathbb{1} & 0 \\ A & \mathbb{1} \end{pmatrix}, \text{ d.h. } \quad Z = \begin{pmatrix} f \\ Af + n \end{pmatrix} ,$$

so ist Z nach Lemma 5.2 normalverteilt mit dem

Mittelwert $\begin{pmatrix} \bar{f} \\ A\bar{f} \end{pmatrix}$ und der

Kovarianz: $\begin{pmatrix} \mathbb{1} & 0 \\ A & \mathbb{1} \end{pmatrix}\begin{pmatrix} F & 0 \\ 0 & \sigma^2\mathbb{1} \end{pmatrix}\begin{pmatrix} \mathbb{1} & A^T \\ 0 & \mathbb{1} \end{pmatrix} = \begin{pmatrix} F & FA^T \\ AF & AFA^T+\sigma^2\mathbb{1} \end{pmatrix}$

Wir wenden nun Lemma 5.1 an, um $f_B = E(f \mid g)$ zu berechnen, es gilt:

$$f_B = \bar{f} + FA^T(AFA^T + \sigma^2\mathbb{1})^{-1}(g - A\bar{f}) \quad .$$

Setzen wir speziell: $F = \mathbb{1}$ und $\bar{f} = 0$, so ist

$$f_B = A^T(AA^T + \sigma^2\mathbb{1})^{-1}$$

$$= (A^TA + \sigma^2\mathbb{1})^{-1}A^Tg \quad .$$

D.h. wir erhalten die Tykhonov-Phillips-Regularisierung. Die Richtigkeit der letzten Gleichung sieht man leicht, wenn man die Identität

$$A^T(AA^T + \sigma^2\mathbb{1}) = (A^TA + \sigma^2\mathbb{1})A^T$$

von links und rechts mit dem Inversen des jeweiligen Klammerausdrucks multipliziert.

Aussagen

5.2 Methode der maximalen Entropie

Wir betrachten folgendes kombinatorische Problem: Gegeben seien

n Teilchen , die in

m Klassen mit n_i, $i = 1,\ldots,m$ Elementen

aufgeteilt werden sollen, d.h. $\sum_{i=1}^{m} n_i = n$.

Wir wollen k - die Anzahl aller möglichen Aufteilungen bestimmen.

Berechnet man die Anzahl aller möglichen Permutationen dieses n-Teilchen - Systems, kommt man zur Identität:

$$k \cdot n_1! \cdot \ldots \cdot n_m! = n! \quad .$$

Also

$$k = \frac{n!}{n_1! \cdot \ldots \cdot n_m!} \quad .$$

Wir stellen uns im folgenden k, k_i, $i = 1,\ldots,m$ als sehr groß vor, so daß wir von der Stirlingschen Formel:

$$n! = \sqrt{2\pi}\ e^{-n} \cdot n^{n+1/2}(1 + O(\tfrac{1}{n}))$$

Gebrauch machen können:

$$k \sim \frac{\sqrt{2\pi}\, e^{-n} \cdot n^n \cdot n^{1/2}}{(\sqrt{2\pi})^m\ e^{-n}\ n_1^{n_1} \cdot \ldots \cdot n_m^{n_m}\ (n_1 \cdot \ldots \cdot n_m)^{1/2}} \quad .$$

$$\Rightarrow\ k^{1/n} \sim (2\pi)^{\frac{1-m}{2n}} \cdot \frac{n^{f_1 + \ldots + f_m}}{n_1^{f_1} \cdot \ldots \cdot n_m^{f_m}} \left(\frac{n}{n_1 \cdot \ldots \cdot n_m} \right)^{\frac{1}{2n}} , \text{ mit } f_i ::= \frac{n_i}{n}$$

$$= (2\pi)^{\frac{1-m}{2n}}\ f_1^{-f_1} \cdot \ldots \cdot f_m^{-f_m} \left(\frac{n}{n_1 \cdot \ldots \cdot n_m} \right)^{\frac{1}{2n}}$$

$$\Rightarrow \frac{1}{n}\log k \sim \frac{1-m}{2n}\log(2\pi) - \sum_{i=1}^{m} f_i \log(f_i) + \frac{1}{2n}\left(\log(n) - \sum_{i=1}^{m}\log(n_i)\right)$$

Für $n >> m$ gilt somit:

$$\frac{1}{n}\log k \sim -\sum f_i \log(f_i) \quad ,$$

wobei $\sum_{i=1}^{m} f_i = 1$.

Man bezeichnet

$$S := \frac{1}{n}\log k$$

als die Entropie des n-Teilchen-Systems.

Das Prinzip der maximalen Entropie lautet:

Bei einem unbekannten System, das durch"Häufigkeiten" $f_1,\ldots,f_m$ $\sum_{i=1}^{m} f_i = 1$, $0 \leq f_i \leq 1$ beschrieben ist, vermutet man, daß

$$S = -\sum_{i=1}^{m} f_i \log f_i$$

maximal ist.

Anwendung dieses Prinzips auf unterbestimmte Gleichungssysteme:

$$Af = g$$

A (n,m) - Matrix mit $n < m$

Suche unter allen Lösungen von $Af = g$ diejenige mit maximaler Entropie:

$$\text{maximiere} \quad -\sum_{i=1}^{m} f_i \log f_i$$

unter den Nebenbedingungen: $Af = g$, $0 \leq f_i \leq 1$, $\sum_{i=1}^{m} f_i = 1$.

II NUMERISCHE LINEARE ALGEBRA SCHLECHT KONDITIONIERTER SYSTEME

Eine Programmbibliothek für die Behandlung schlecht gestellter Probleme müßte Programme für die Berechnung von

- Verallgemeinerten Lösungen
- Regularisierten Lösungen
- SVD
- Quadratischen Optimierungsaufgaben unter Nebenbedingungen
- Max. Entropie-Lösungen

enthalten. Auf die entsprechenden Algorithmen wollen wir in diesem Teil eingehen.

1. FEHLERABSCHÄTZUNGEN

Es seien $A, \tilde{A}$ (n,n) - Matrizen, $b, \tilde{b} \in \mathbb{R}^n$, und es gelte:

$Ax = b$, A regulär, sowie

$\tilde{A}\tilde{x} = \tilde{b}$, wobei wir $\tilde{A}$, $\tilde{b}$ als (kleine) Störungen der exakten Größen A bzw. b ansehen wollen.

Wir wollen nun den Fehler von $\tilde{x}$ in der euklidischen Norm abschätzen und erinnern dabei an die Definition von

$$\kappa(A) := \|A\| \ \|A^{-1}\|$$

als der sogenannten Kondition der Matrix A.

Es gilt das

LEMMA 1.1: Sei $\Theta = \kappa(A) \dfrac{\|A - \tilde{A}\|}{\|A\|} < 1$, dann ist auch $\tilde{A}$ invertierbar, und es gilt die Abschätzung

$$\frac{\|x-\tilde{x}\|}{\|x\|} \leq \frac{\kappa(A)}{1-\theta} \left\{ \frac{\|A-\tilde{A}\|}{\|A\|} + \frac{\|b-\tilde{b}\|}{\|b\|} \right\}$$

BEWEIS: Prama 1, Bulirsch - Stoer, Bd. 1, 3. Aufl., p. 153-154. ∎

BEMERKUNG: $\kappa(A)$ ist also (im wesentlichen) der Verstärkungsfaktor für den relativen Fehler in x gegenüber den Ausgangsfehlern in A sowie b.

Wir wollen dieses Lemma auf die Normalgleichungen

$$A^*Ax = A^*b$$

mit einer (m,n) - Matrix A, $m > n = \text{Rang}(A)$ anwenden.

Für beliebige Matrizen A definieren wir

$$\kappa(A) = \|A\| \, \|A^+\| \quad .$$

Die zur euklidischen Vektornorm gehörende Matrizennorm läßt sich leicht mit Hilfe der SVD bestimmen:

$$Ax = \sum_{k=1}^{n} \sigma_k (x,u_k) v_k \quad , \quad \text{mit } \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n, \text{ wobei } \sigma_k^2 \text{ die postiven Eigenwerte von } A^*A \text{ sind.}$$

$$\Rightarrow \quad \|Ax\|^2 = \sum_{k=1}^{n} \sigma_k^2 |(x,u_k)|^2$$

D.h. das $\sup_{\|x\|=1} \|Ax\|$ wird für $x = u_1$ angenommen und beträgt σ_1, wegen

$$A^+b = \sum_{k=1}^{n} \frac{1}{\sigma_k} (b,v_k) u_k$$

sowie

$$A^*Ax = \sum_{k=1}^{n} \sigma_k^2(x,u_k)u_k$$

gilt also insgesamt

$$\|A\| = \sigma_1 \ , \ \|A^+\| = 1/\sigma_n \ , \quad \|A^*A\| = \sigma_1^2$$

und damit

$$\kappa(A^*A) = \frac{\sigma_1^2}{\sigma_n^2} = \kappa^2(A) \ .$$

D.h. bei Verwendung der Normalgleichungen wird die Konditionszahl quadriert. Für schlecht konditionierte Systeme gilt $\kappa(A) >> 1$, also $\kappa(A^*A) >> \kappa(A)$, nach Lemma 1.1 wird die Lösung x also ungeheuer empfindlich gegenüber Störungen in A bzw. b.

Rechnet man mit t Dezimalen Genauigkeit, so muß gelten

$$\kappa^2(A) \cdot 10^{-t} << 1$$

$$\kappa(A) << 10^{t/2} \ .$$

Es stellt sich nun die Frage, ob die Fehler bei überbestimmten Systemen grundsätzlich mit dem Quadrat der Konditionszahl verstärkt werden, oder ob dieses Phänomen nur bei Verwendung der Normalgleichungen auftritt.

<u>SATZ 1.2:</u> $A, \tilde{A}$ seien (m,n)-Matrizen, $b, \tilde{b} \in \mathbb{R}^m$, $\text{Rang}(A) = n \leq m$, x sei verallgemeinerte Lösung von $Ax = b$, und es sei

$$\theta = \kappa(A) \frac{\|A-\tilde{A}\|}{\|A\|} \frac{\sqrt{2}+1}{\sqrt{2}-1} < 1 \ .$$

arbeitet direkt auf der verallg. Lösung

Dann hat auch $\tilde{A}$ den Rang n, und es gilt für die verallgemeinerte Lösung $\tilde{x}$ von $\tilde{A}\tilde{x} = \tilde{b}$:

$$\frac{\|\tilde{x}-x\|}{\|x\|} \leq \frac{\sqrt{2}+1}{\sqrt{2}-1}\ \frac{\kappa(A)}{1-\theta}\left\{\frac{\|\tilde{A}-A\|}{\|A\|}+\frac{\|\tilde{b}-b\|}{\|b\|}\right\}\left\{1+\left(\frac{1}{2}\kappa(A)\ \frac{\|Ax-b\|}{\|Ax\|}\right)^2\right\}^{1/2}$$

BEMERKUNG:

i) Falls $\|Ax-b\|$ $\left(= \min\limits_{y\in\mathbb{R}^n} \|Ay-b\|\right)$ so klein ist, daß $\kappa(A)\|Ax-b\| / \|Ax\| \approx 1$ ist, so entspricht Satz 1.2 dem Lemma 1.1, d.h. auch für überbestimmte Systeme ist eine Fehlerabschätzung mit dem Verstärkungsfaktor $\kappa(A)$ (und nicht $\kappa(A)^2$) möglich.

ii) Ist P die orthogonale Projektion auf $R(A)$, so gilt (Aufgabe 5b) $Ax = AA^+b = P\,b$, man kann also auf der rechten Seite der Ungleichung Ax durch $P\,b$ ersetzen.

BEWEIS: Die Normalgleichungen

$$A^*Ax = A^*b$$

sind eindeutig lösbar. Mit der Definition

$$r = (b - Ax) / 2\sigma_n \quad , \quad \text{wobei } \sigma_n \text{ der kleinste Singulärwert von } A \text{ ist,}$$

sind die Normalgleichungen äquivalent zu dem System:

$$\underbrace{\begin{pmatrix} 2\sigma_n \mathbb{1} & A \\ A^* & 0 \end{pmatrix}}_{=: C} \begin{pmatrix} r \\ x \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix} .$$

Wir berechnen die Eigenwerte der symmetrischen Matrix C:

$$\begin{pmatrix} 2\sigma_n \mathbb{1} & A \\ A^* & 0 \end{pmatrix} \cdot \begin{pmatrix} s \\ y \end{pmatrix} = \lambda \begin{pmatrix} s \\ y \end{pmatrix}$$

$$\Leftrightarrow \quad \left| \begin{array}{l} 2\sigma_n s + Ay = \lambda s \\ A^*s \qquad\quad = \lambda y \end{array} \right.$$

Multiplikation der ersten Zeile mit A^* ergibt folglich:

$$2\sigma_n \lambda y + A^*Ay = \lambda^2 y$$

$$\Leftrightarrow \quad A^*Ay = (\lambda^2 - 2\lambda\sigma_n)y$$

D.h. y ist Eigenvektor von A^*A zum Eigenwert $\lambda^2 - 2\lambda\sigma_n$, die Eigenwerte von A^*A sind aber gerade $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_n^2 > 0$

$$\Rightarrow \quad \lambda_k^2 - 2\lambda_k\sigma_n - \sigma_k^2 = 0 \qquad , \quad k = 1,\ldots,n$$

$$\Leftrightarrow \quad \lambda_{k_{1,2}} = \sigma_n \pm \sqrt{\sigma_k^2 + \sigma_n^2} \qquad , \quad k = 1,\ldots,n \quad .$$

Damit sind $2n$ von Null verschiedene Eigenwerte von C bestimmt, ferner tritt $\lambda = 2\sigma_n (> 0)$ mit der geometrischen Vielfalt $m - n$ als Eigenwert auf:
es gilt $\dim N(A^*) = m - n$ und mit $s \in N(A^*)$ gilt:

$$C\begin{pmatrix} s \\ 0 \end{pmatrix} = 2\sigma_n \begin{pmatrix} s \\ 0 \end{pmatrix} \quad .$$

Der betragsgrößte Eigenwert von C ist $\lambda_{1_1} = \sigma_n + \sqrt{\sigma_n^2 + \sigma_1^2}$,

der betragskleinste Eigenwert von C ist $\lambda_{n_2} = (1 - \sqrt{2})\sigma_n$.

Insbesondere ist C regulär, und es gilt:

$$\|C\| = \lambda_{1_1} = \sigma_n + \sqrt{\sigma_1^2 + \sigma_n^2} > \sigma_1 = \|A\| \quad ,$$

$$\kappa(C) = \frac{\lambda_{1_1}}{|\lambda_{n_2}|} = \frac{\sigma_n + \sqrt{\sigma_1^2+\sigma_n^2}}{(\sqrt{2} - 1)\sigma_n} \leq \frac{\sigma_1(1 + \sqrt{2})}{\sigma_n(\sqrt{2} - 1)} = \kappa(A)\,\frac{\sqrt{2}+1}{\sqrt{2}-1} \quad .$$

Mit $\tilde{C} = \begin{pmatrix} 2\sigma_n \mathbb{1} & \tilde{A} \\ \tilde{A}^* & 0 \end{pmatrix}$, $\tilde{C}\begin{pmatrix} \tilde{r} \\ \tilde{x} \end{pmatrix} = \begin{pmatrix} \tilde{b} \\ 0 \end{pmatrix}$

wenden wir Lemma 1.1 an

$$\tilde{\theta} = \kappa(C)\,\frac{\|C-\tilde{C}\|}{\|C\|}\;;\quad \|C-\tilde{C}\| = \left\|\begin{pmatrix} 0 & A-\tilde{A} \\ A^*-\tilde{A}^* & 0 \end{pmatrix}\right\| = \|A-\tilde{A}\| \quad .$$

(Um den maximalen Eigenwert von $C-\tilde{C}$ zu berechnen, verfahre man wie oben, mit dem Unterschied, daß man für σ_n Null und für A die Matrix $A-\tilde{A}$ einsetzt.)

Somit:

$$\tilde{\theta} = \kappa(C)\,\frac{\|C-\tilde{C}\|}{\|C\|} \le \frac{\sqrt{2}+1}{\sqrt{2}-1}\;\kappa(A)\,\frac{\|A-\tilde{A}\|}{\|A\|} = \theta < 1$$

Nach Lemma 1.1 ist $\tilde{C}$ folglich invertierbar (d.h. Rang $(\tilde{A}) = n$), und es gilt:

$$\frac{\left\|\begin{pmatrix} r \\ x \end{pmatrix} - \begin{pmatrix} \tilde{r} \\ \tilde{x} \end{pmatrix}\right\|}{\left\|\begin{pmatrix} r \\ x \end{pmatrix}\right\|} \le \frac{\sqrt{2}+1}{\sqrt{2}-1}\;\frac{\kappa(A)}{1-\theta}\left\{\frac{\|A-\tilde{A}\|}{\|A\|} + \frac{\|b-\tilde{b}\|}{\|b\|}\right\}$$

2.

$$\Rightarrow \frac{\|x-\tilde{x}\|}{\|x\|} \le \frac{\sqrt{2}+1}{\sqrt{2}-1}\;\frac{\kappa(A)}{1-\theta}\left\{\frac{\|A-\tilde{A}\|}{\|A\|} + \frac{\|b-\tilde{b}\|}{\|b\|}\right\}\frac{\left(\|r\|^2+\|x\|^2\right)^{1/2}}{\|x\|}$$

Mit:

$$\frac{\left(\|r\|^2+\|x\|^2\right)^{1/2}}{\|x\|} = \left\{1+\left(\frac{\|b-Ax\|}{2\sigma_n\|x\|}\right)^2\right\}^{1/2} \le \left\{1+\left(\frac{\|b-Ax\|}{2\sigma_n\|Ax\|/\|A\|}\right)^2\right\}^{1/2}$$

folgt die Behauptung.

■

Das Problem der "Konditionszahlquadrierung" tritt natürlich auch bei der Verwendung der Tykhonov-Phillips-Regularisierung auf, es gilt (s. Beweis zu Satz 4.2):

$$R_\gamma g = \sum_{k=1}^{n} \frac{\sigma_k}{\gamma^2+\sigma_k^2} (g,g_k) f_k \quad ,$$

also

$$\|R_\gamma g\|^2 = \sum_{k=1}^{n} \left| \frac{\sigma_k}{\gamma^2+\sigma_k^2} \right|^2 |(g,g_k)|^2 \leq \sum_{k=1}^{n} \left| \frac{\gamma}{\gamma^2+\gamma^2} \right|^2 |(g,g_k)|^2 \leq \left(\frac{1}{2\gamma}\right)^2 \|g\|$$

(Die Funktion $\varphi_\gamma(x) := \frac{x}{\gamma^2+x^2}$ ist maximal bei $x = \gamma$.)

Folglich:

$$\|R_\gamma\| \sim \frac{1}{2\gamma} \quad ,$$

während die Eigenwerte der Matrix $A^*A + \gamma^2 \mathbb{1}$ gleich

$$\sigma_k^2 + \gamma^2 \quad , \quad k = 1,\ldots,n$$

sind, demgemäß gilt:

$$\kappa(A^*A + \gamma^2 \mathbb{1}) = \frac{\sigma_1^2+\gamma^2}{\sigma_n^2+\gamma^2} \approx \frac{\sigma_1^2}{\gamma^2} \quad ,$$

$\sigma_1 >> \gamma$
$\gamma >> \sigma_n$

d.h. die Kondition hängt quadratisch von γ ab.

Wir wollen nun Fehlerabschätzungen für die Berechnung der Singulärwerte einer (kleinen) Störungen unterworfenen Matrix angeben. Dazu benötigen wir das

LEMMA 1.3: A, $\tilde{A}$ seien symmetrische (n,n) - Matrizen mit den Eigenwerten $\lambda_1 \geq \ldots \geq \lambda_n$ bzw. $\tilde{\lambda}_1 \geq \ldots \geq \tilde{\lambda}_n$, dann gilt:

$$|\lambda_k - \tilde{\lambda}_k| \leq \|A-\tilde{A}\|, \quad k = 1,\ldots,n \quad .$$

BEWEIS: Der Beweis folgt sofort aus dem Ritzschen Maximum - Minimum - Prinzip:

$$\lambda_k = \max_{N,\dim N=k} \min_{\|x\|=1, x\in N} (Ax,x) \quad .$$

Wegen: $(Ax,x) - (\tilde{A}x,x) = ((A-\tilde{A})x,x) \leq \|A-\tilde{A}\| \, \|x\|^2$

gilt $\forall\, x, \|x\| = 1$:

$$(\tilde{A}x,x) - \|A-\tilde{A}\| \leq (Ax,x) \leq (\tilde{A}x,x) + \|A-\tilde{A}\|$$

folglich gilt für jeden k-dimensionalen Teilraum N :

$$\min_{x\in N} (\tilde{A}x,x) - \|A-\tilde{A}\| \leq \min_{x\in N} (Ax,x) \leq \min_{x\in N} (\tilde{A}x,x) + \|A-\tilde{A}\|$$

Maximumbildung über alle k-dimensionalen Teilräume liefert die Behauptung. ■

Damit können wir folgenden Satz beweisen:

SATZ 1.4: A, $\tilde{A}$ seien (m,n) - Matrizen mit den Singulärwerten $\sigma_1 \geq \ldots \geq \sigma_p$, bzw. $\tilde{\sigma}_1 \geq \ldots \geq \tilde{\sigma}_p$ (d.h. Rang A = Rang $\tilde{A}$ = p).

Dann gilt:

$$|\sigma_k - \tilde{\sigma}_k| \leq \|A-\tilde{A}\| \quad .$$

BEWEIS: Im Beweis zu Satz 1.2 hatten wir die Eigenwerte λ der Matrix

$$C = \begin{pmatrix} O & A \\ A^* & O \end{pmatrix}$$

berechnet: $\lambda \in \{\pm \sigma_k \,,\, k = 1,\ldots,p\} \cup \{0\}$.

Definiert man $\tilde{C}$ mit $\tilde{A}$ anstelle von A, so folgt die Behauptung infolge $||C-\tilde{C}|| = ||A-\tilde{A}||$ sofort aus Lemma 1.3.

□

Der Satz 1.4 im Verein mit Lemma 1.3 zeigt, daß es ungünstig sein kann, die Singulärwerte σ_k einer Matrix A als Quadratwurzel der Eigenwerte von A*A zu berechnen. Bei diesem Vorgehen hat man nach Lemma 1.3 die Abschätzung:

$$|\sigma_k^2 - \tilde{\sigma}_k^2| / \sigma_k \leq \frac{1}{\sigma_k} ||A^*A-\tilde{A}^*\tilde{A}||$$

$$\Rightarrow \quad |\sigma_k - \tilde{\sigma}_k| / \sigma_k \leq \frac{1}{\sigma_k(\sigma_k+\tilde{\sigma}_k)} ||A^*A-\tilde{A}^*\tilde{A}||$$

und folglich bei sehr kleinen Singulärwerten σ_k einen großen Genauigkeitsverlust gegenüber Satz 1.4:

Soll der relative Fehler $|\sigma_k-\tilde{\sigma}_k| / \sigma_k$ klein bleiben, so muß gelten

$$||A^*A-\tilde{A}^*\tilde{A}|| << \frac{1}{\sigma_k^2+\sigma_k\tilde{\sigma}_k} \approx \frac{1}{2\sigma_k^2} ,$$

während nach Satz 1.4 die Genauigkeitsanforderung

$$||A-\tilde{A}|| << \frac{1}{\sigma_k}$$

ausreichend ist.

Numerische Tests zeigen, daß die obigen Abschätzungen auch nicht wesentlich verbessert werden können. Rechnet man auf t Dezimalstellen genau, so lassen sich die Singulärwerte σ_k nur für $\sigma_k > 10^{-t/2}$ genau berechnen, falls σ_k explizit als Quadratwurzel eines Eigenwertes von A*A berechnet wird (vgl. Aufgabe 18).

2. ALGORITHMEN FÜR DIE BERECHNUNG VERALLGEMEINERTER UND REGULARISIERTER LÖSUNGEN

Der Abschnitt II-1 lehrte uns, daß wir bei den in Frage stehenden Algorithmen zur numerischen Behandlung schlecht gestellter Probleme die explizite Berechnung von A^*A vermeiden müssen.

2.1 Die Lösung überbestimmter Gleichungssysteme

Wir haben die Aufgabe

$$Ax = b$$

zu lösen, wobei A eine (m,n)-Matrix mit $\text{Rang } A = n \leq m$ ist. Das explizite Lösen der Normalgleichungen kann man durch eine Q-R-Zerlegung der Matrix A umgehen:

Sei

$$A = Q\,R$$

mit einer unitären (m,n)-Matrix Q (d.h. $Q^*Q = \mathbb{1}_n$) und einer oberen Dreiecksmatrix R.

Die Normalgleichungen lauten

$$A^*Ax = A^*b$$

$$\Leftrightarrow \quad R^*Q^*QRx = R^*Q^*b$$

$$\Leftrightarrow \quad R^*Rx = R^*Q^*b \quad .$$

Da A vollen Rang besitzt, ist R (bzw. R^*) invertierbar und die letzte Gleichung ist äquivalent zu

$$Rx = Q^*b \quad .$$

Infolge

$$A^*A = R^*R$$

haben A und R die gleichen Singulärwerte, mithin

$$\kappa(R) = \kappa(A) \quad .$$

Ist die Q-R-Zerlegung erst einmal bewerkstelligt, kann die Lösung x stabil durch Rückwärtseinsetzen gewonnen werden.

Die Zerlegung

$$A = Q\,R$$

läßt sich durch eine Folge von Householdertransformationen erreichen. Wir multiplizieren A sukzessiv von links mit symmetrischen, unitären (m,n) - Matrizen

$$Q_k \cdot Q_{k-1} \cdot \ldots \cdot Q_1 A \quad ,$$

so daß nach Multiplikation mit Q_k $(k = 1,\ldots,n)$ die ersten k Spalten von $Q_k \cdot \ldots \cdot Q_1 A$ obere Dreiecksform haben. Nach n Schritten hat A obere Dreiecksform, und wir setzen:

$$Q := \text{die ersten n Spalten von } Q_1 \cdot Q_2 \cdot \ldots \cdot Q_n$$

Da das Produkt unitärer Matrizen wieder unitär ist, ist auch Q unitär, und wir haben die gewünschte Zerlegung erreicht. Die Matrix Q wird allerdings nicht explizit berechnet, vielmehr wird Q^*b sukzessiv (z.B. durch Anhängen von b als $(n+1)$-te Spalte von A) mitberechnet. Ferner läßt sich durch die spezielle Gestalt der Matrizen Q_k das Matrizenprodukt effizient berechnen. - Wir müßten bei "normaler" Matrizenmultiplikation, nämlich n

von (m,m) - Matrizen mit (m,n) - Matrizen ausführen, was zu einem Rechenaufwand von $O(n^2m^2)$ Rechenoperationen führen würde.

Die Householdermatrizen haben die Gestalt

$$Q_k = \mathbb{1}_m - 2\, u_k\, u_k^T \quad , \quad \|u_k\| = 1 \quad .$$

Es gilt folglich:

$$Q_k^T = Q_k \qquad \text{und}$$

$$Q_k^T Q_k = \mathbb{1}_m - 2\, u_k\, u_k^T - 2\, u_k\, u_k^T + 4\, u_k \overbrace{u_k^T\, u_k}^{=1} u_k^T = \mathbb{1}_m$$

Wir werden nun die u_k , $k = 1,\ldots,n$ so bestimmen müssen, daß

i) eine Multiplikation mit Q_k die ersten $k-1$ Spalten einer oberen Dreiecksmatrix unverändert läßt.

ii) durch die Multiplikation von Q_k die k-te Spalte unterhalb der Diagonalen mit Nullen gefüllt wird.

Folgende Wahl der u_k erfüllt beide Bedingungen:

Sei $A_j = Q_{j-1} \cdot \ldots \cdot Q_1 A$, und $a_{i\ell}^j$ deren Elemente, dann setze:

$$u_k = \frac{v_k}{\|v_k\|} \quad , \text{ wobei}$$

$$v_k = (\underbrace{0,\ldots,0}_{k-1 \text{ Nullen}}, a_{kk}^k, \ldots, a_{m\,k}^k)^T \pm c_k \cdot e_k \quad , \text{ mit}$$

$$c_k = \left(\sum_{j=k}^{m} |a_{jk}^k|^2 \right)^{1/2} \quad , \; e_k \in \mathbb{R}^m \text{ - k-ter Einheitsvektor.}$$

Aus numerischen Gründen wählt man das Vorzeichen von c_k als:

$$\text{sign}\,(a_{kk}^k) \quad .$$

Mit einem so definierten u_k nimmt Q_k die Form

$$Q_k = \left(\begin{array}{c|c} \mathbb{1}_{k-1} & 0 \\ \hline 0 & * \end{array}\right)$$

an. Daraus folgt aber sofort, daß die ersten k-1 Spalten einer oberen Dreiecksmatrix durch Linksmultiplikation mit Q_k nicht verändert werden, d.h. Bedingung i) ist erfüllt.

Die Bedingung ii) wird auch erfüllt:

Sei a die k-te Spalte von A_k, nach Definition von Q_k

$$\begin{aligned} Q_k a &= a - 2(u_k, a)u_k \\ &= a - \frac{2}{\|v_k\|^2}(v_k, a)v_k \\ &= a - \frac{2}{c_k^2 \pm 2a_k c_k + c_k^2}(c_k^2 \pm c_k a_k)v_k \\ &= a - v_k \end{aligned}$$

Da v_k in den Komponenten $k+1,\ldots,m$ mit a übereinstimmt, folgt die Behauptung.

Hierbei sieht man auch, daß für die Multiplikation von Q_k mit A_{k-1} im wesentlichen nur n innere Produkte der Länge m berechnet werden müssen. Man zeigt leicht, daß der Householderalgorithmus für $m = n$

$$\frac{2}{3}n^3 + O(n^2) \qquad \text{R.O.}$$

benötigt, was dem Rechenaufwand für die Berechnung von A^*A ($\sim \frac{1}{2}n^3$) und anschließender Cholesky-Zerlegung von A^*A ($\sim \frac{1}{6}n^3$) entspricht.

Für $m \geq n$ ist der Rechenaufwand für die Q-R-Zerlegung im Prinzip durch

$$n^2(m-n/3) \qquad \text{R.O.}$$

gegeben.

2.2 Die Lösung unterbestimmter Gleichungssysteme

Sei nun Rang $A = m$, $m \leq n$.

Die verallgemeinerte Lösung von

$$Ax = b$$

ist die eindeutige bestimmte Lösung der Normalgleichungen

$$A^*Ax = A^*b \quad .$$

Zerlegt man nun

$$A^* = Q\,R$$

mit einer unitären (n,m) - Matrix Q und einer oberen Dreiecksmatrix R, so löst

$$x = Qz \qquad \text{mit} \qquad R^*z = b$$

die Normalgleichungen:

$$A^*Ax = A^*R^*Q^*Qz = A^*R^*z = A^*b$$

2.3 Die Lösung regularisierter Gleichungssysteme

Wir hatten im Abschnitt I-4.1 die Äquivalenz der drei Aufgaben:

i) $\min_x \|Ax-b\|^2 + \gamma^2\|x\|^2$

ii) löse $(A^*A + \gamma^2 \mathbb{1})x = A^*b$

iii) berechne die verallgemeinerte Lösung von $\begin{pmatrix} A \\ \gamma \mathbb{1} \end{pmatrix} x = \begin{pmatrix} b \\ 0 \end{pmatrix}$

gezeigt.

D.h. wir könnten für die Berechnung der Tykhonov-Phillips-Regularisierung einfach die Q-R-Zerlegung aus 2.2 auf die Matrix $\begin{pmatrix} A \\ \gamma \mathbb{1} \end{pmatrix}$ anwenden. Diese brute - force - Methode hat in der Praxis natürlich zwei häßliche Makel:

i) Durch das "Anhängen" der einfachen Matrix $\gamma \mathbb{1}$ steigt der Rechenaufwand um n^3 R.O. (nämlich von $n^2(m-\frac{n}{3})$ auf $n^2(m+n-\frac{n}{3})$ R.O.)

ii) Der Regularisierungsparameter liegt im vorhinein nicht fest, sondern muß in aller Regel durch "Probieren" bestimmt werden, d.h. die Q-R-Zerlegung von $\begin{pmatrix} A \\ \gamma \mathbb{1} \end{pmatrix}$ ist für viele verschiedene Werte von γ durchzuführen. Dabei ist es natürlich ärgerlich, daß, obwohl die Matrix kaum verändert wird, jeweils der volle Rechenaufwand nötig wird.

Abhilfe in beiden Punkten schafft eine Methode, die auf L. Eldén zurückgeht und dem die beiden folgenden Ideen zugrunde liegen:

i) Bringe A durch geeignete Variablentransformation in eine Bidiagonalgestalt.

ii) Eliminiere die "angehängte" Matrix $\gamma \mathbb{1}$ mit Hilfe von Givens-Rotationen ohne die Bidiagonalgestalt von A zu zerstören.

Wir wollen das Verfahren von Eldén nun genauer beschreiben:

A sei eine (m,n) - Matrix mit $m \geq n = \text{Rang } A$, gegeben sei das Problem:

$$\min_x \|Ax-b\|^2 + \gamma^2 \|x\|^2$$

$$\Leftrightarrow \min_x (Ax-b, Ax-b) + \gamma^2 (x,x)$$

$$\Leftrightarrow \min_x (U(Ax-b), U(Ax-b)) + \gamma^2 (Vx, Vx) \quad , \quad U, V \text{ unitär}$$

$$\Leftrightarrow \min_y \|UAV^*y - Ub\|^2 + \gamma^2 \|y\| \quad , \quad y = Vx$$

Nun sind U, V so zu wählen, daß

$$B = U A V^*$$

bidiagonal wird, d.h.: $B_{ij} = 0 \quad \forall\, i,j$ mit $i < j$ oder $i > j+1$.

Man gewinnt U, V als Produkt von Householdertransformationen:

$$U = U_n \cdot \ldots \cdot U_1 \quad , \quad V^* = V_1, \ldots, V_{n-1} \quad ,$$

$$A_k = U_k \cdot \ldots \cdot U_1 \, A \, V_1 \cdot \ldots \cdot V_k$$

mit (m,m) - Matrizen U_k und (n,n) - Matrizen V_k

$$U_k = \mathbb{1}_m - 2 u_k^T u_k \quad , \quad V_k = \mathbb{1}_n - 2 v_k^T v_k \quad .$$

Hierbei ist u_k ein normierter Vektor aus dem $\mathbb{R}^m$ mit $k-1$ führenden Nullen:

$$u_k = (\underbrace{0,\ldots,0}_{k-1},\tilde{u}_k) \quad \text{und} \quad v_k \in \mathbb{R}^n, \quad \|v_k\| = 1$$

mit $\tilde{v}_k = (\underbrace{0,\ldots,0}_{k},\tilde{v}_k)$, also $\tilde{u}_k \in \mathbb{R}^{m-k+1}$, $\tilde{v}_k \in \mathbb{R}^{n-k}$.

Somit

$$U_k = \left(\begin{array}{c|c} \mathbb{1}_{k-1} & 0 \\ \hline 0 & \mathbb{1}_{m-k+1} - 2\tilde{u}_k\tilde{u}_k^T \end{array}\right), \quad V_k = \left(\begin{array}{c|c} \mathbb{1}_k & 0 \\ \hline 0 & \mathbb{1}_{n-k} - 2\tilde{v}_k\tilde{v}_k^T \end{array}\right)$$

die Vektoren u_k, v_k werden nun so bestimmt, daß A_k die Form

$$A_k = \left(\begin{array}{c|c} B_k & 0 \\ \hline 0 & * \end{array}\right) \} \; k - \text{"Nullzeilen"}$$

$\underbrace{\quad}_{k}$ "Nullspalten"

hat, wobei B_k eine (k+1,k+1) - Bidiagonalmatrix ist. Bedenkt man, daß die Multiplikation der V_k von rechts ein Operieren auf den Zeilen (anstelle der Spalten, wie bei der Linksmultiplikation der U_k) bedeutet, so wird klar, daß die u_k, v_k vollkommen analog zum gewöhnlichen Householderalgorithmus gewählt werden können:

i) Bestimme u_1 so, daß die erste Spalte von U_1A unterhalb der Hauptdiagonalen mit Nullen besetzt ist, also identisch zum Householderalgorithmus.

Dann wähle v_1 so, daß die erste Zeile von $A_1 = U_1AV_1$ in

in den Elementen 3 bis n gleich Null ist. D.h. ist $a = (a_1, \tilde{a}) \in \mathbb{R}^n$ die erste Zeile von U_1A, so setzt man $v_1((0,\tilde{a})^T + \text{sign}(a_2) \cdot \|\tilde{a}\| \cdot e_2)/\alpha$, α-Normierungsfaktor. Da die erste Spalte von V_1 gleich dem ersten Einheitsvektor im $\mathbb{R}^n$ ist, wird die erste Spalte von U_1A durch Rechtsmultiplikation mit V_1 nicht verändert; $A_1 = U_1AV_1$ hat somit die gewünschte Gestalt.

ii) Wähle u_2 wie im gewöhnlichen Householderverfahren; da die erste Zeile von U_2 aus dem ersten Einheitsvektor besteht, werden die Nullen in der ersten Zeile von A_1 nicht zerstört. Wähle nun v_2 so, daß die Elemente 4 bis n in der 2. Zeile von U_2A eliminiert werden. Die Form von V_2 garantiert dabei, daß die beiden ersten Spalten von U_2A_1 unverändert bleiben, also hat A_2 die gewünschte Form.

iii) - n) Wir führen insgesamt $n-1$ solcher Schritte aus und müssen im n-ten Schritt nur noch einmal von links mit U_n multiplizieren, um in der n-ten Spalte unterhalb der Diagonalen Nullen zu erzeugen.

Insgesamt haben wir also einen doppelten Householderalgorithmus durchgeführt und benötigen demnach

$$\frac{4}{3} n^3 + O(n^2) \quad \text{R.O.} \quad (\text{für } m = n).$$

Haben wir einmal die Transformation von A auf Bidiagonalgestalt berechnet, müssen wir für die verschiedenen Regularisierungsparameter γ lediglich das überbestimmte System

$$\begin{pmatrix} B \\ \gamma \mathbb{1} \end{pmatrix} y = \begin{pmatrix} Ub \\ 0 \end{pmatrix}, \qquad B \text{ - bidiagonal}$$

lösen.

Um die zusammengesetzte Matrix auf der linken Seite auf Bidiagonalgestalt zu bringen, wollen wir Givens-Rotationen verwenden.

DEFINITION: Eine Matrix $R(i,j) = (r_{\ell,k})$ heißt Givens-Rotation, wenn gilt:

$$r_{\ell,k} = \begin{cases} c & , \ \ell = k = i \text{ oder } j \\ s & , \ \ell = j, k = i \\ -s & , \ \ell = i, k = j \\ \delta_{\ell k} & , \text{ sonst} \end{cases} ,$$

mit $c^2 + s^2 = 1$, d.h. $c = \cos\alpha$, $s = \sin\alpha$ für ein $\alpha \in [0, 2\pi)$

D.h.: $R(i,j) =$

$$\begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \end{array} \begin{pmatrix} 1 & & & & & & 0 \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & c & \cdots & s & \\ & & & \vdots & \ddots & \vdots & \\ & & & -s & \cdots & c & \\ & & & & & & \ddots \\ 0 & & & & & & & 1 \end{pmatrix} \begin{array}{l} \\ \\ \\ \\ \\ \\ \\ \end{array} \begin{array}{l} \\ \\ \\ -j \\ \\ -i \\ \\ \end{array}$$

$$\begin{array}{cc} | & | \\ j & i \end{array}$$

Es gilt $R^T(i,j)\, R(i,j) = \mathbb{1}$ sowie

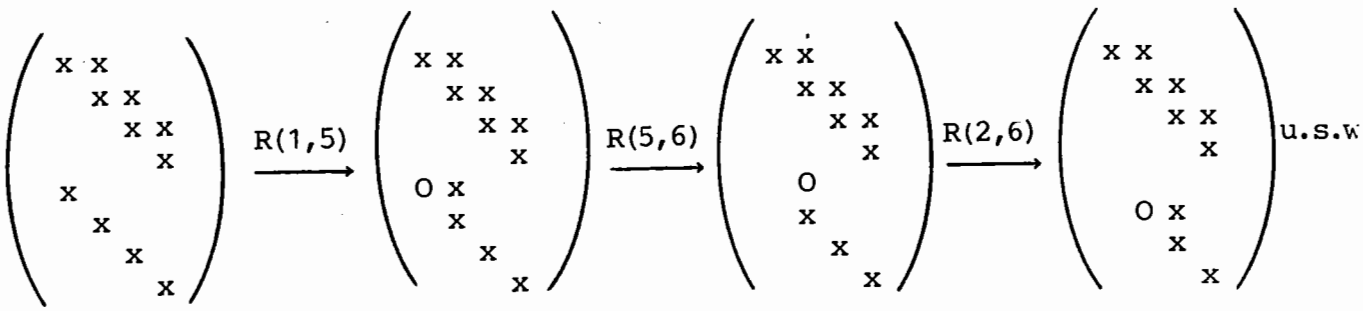
$$(R(i,j) \cdot x)_\ell = \begin{cases} c\,x_j + s\,x_i & , \ \ell = j \\ c\,x_i - s\,x_j & , \ \ell = i \\ x_\ell & , \text{ sonst} \end{cases} ,$$

d.h. Linksmultiplikation von $R(i,j)$ mit einer Matrix verändert diese also nur in den Zeilen i und j; hat ein Vektor jedoch Nullen auf diesen Positionen, so wird er überhaupt nicht verändert. Ferner lassen sich $\forall\, x_i, x_j$ stets c,s $(= \cos\alpha, \sin\alpha)$ so bestimmen, daß

$$x_j \sin\alpha + x_i \cos\alpha = 0 \quad \text{oder} \quad x_i \sin\alpha - x_j \cos\alpha = 0$$

gilt, d.h. eine der beiden Komponenten kann zu Null gemacht werden.

Der Eliminationsprozeß mit Givens-Rotationen sei an einem Beispiel mit $m = n = 4$ verdeutlicht:



D.h. wir multiplizieren $R(m+i,m+i+1) \cdot R(i,m+i)$, $i = 1,2,\ldots,n$ mit jeweils passenden α. Jede Givens-Rotation erfordert $O(1)$, der gesamte Eliminationsprozeß also $O(n)$ R.O.. Die Berechnung von y aus $\begin{pmatrix} B \\ \gamma 1\!\!1 \end{pmatrix} y = \begin{pmatrix} Ub \\ 0 \end{pmatrix}$ erfordert also $O(n)$ R.O., während $x = V^* y = V_1 \cdot \ldots \cdot V_{n-1} y$ in $n^2 + O(n)$ R.O. berechnet werden kann. Insgesamt beträgt der Rechenaufwand für $m = n$:

$\frac{4}{3} n^3 + O(n^2)$ R.O. für die Bidiagonalisierung und

$n^2 + O(n)$ R.O. für die Berechnung des Lösungsvektors bei jedem neuen Wert von γ.

2.4 Die Berechnung der SVD

Wir wollen an dieser Stelle auf einige wesentliche Punkte des Algorithmus von G. Golub und C. Reinsch (s. Num. Math. 14, 403 - 420 (1970)) eingehen.

Im Abschnitt 2.3 hatten wir gezeigt, wie man eine Matrix A durch Links- und Rechtsmultiplikation mit unitären Matrizen U,V auf

Bidiagonalgestalt bringen kann:

$$B = U A V^*$$

Wegen $B^*B = V A^* A V^*$ haben B und A die gleichen Singulärwerte, d.h. wir können uns auf die Behandlung bidiagonaler Matrizen beschränken.

Wir wollen die Singulärwerte von B, d.h. die Quadratwurzel aus den Eigenwerten der symmetrischen Tridiagonalmatrix $M = B^*B$ berechnen. Eine sehr effiziente iterative Methode für dieses Problem ist der Q-R-Algorithmus:

$$M_1 = M \quad (= B^*B)$$

(*) $M_k = Q R$, Q unitär, R obere Dreiecksmatrix

$$M_{k+1} = R Q = Q^* M_k Q$$

Um die Konvergenzgeschwindigkeit zu verbessern, verwendet man noch sogenannte Shift-Strategien (Spektralverschiebungen durch Addition von $\pm s\mathbb{1}$ zu M_k bzw. M_{k+1} während der Iteration, mit jeweils geeignetem s).

Unter gewissen Voraussetzungen konvergieren die Diagonalelemente von M_k gegen die Eigenwerte von M, die Bandstruktur von M bleibt dabei für alle M_k erhalten. (s. Prama 1.)

Das Problem liegt wieder in der Berechnung von

$$M = B^*B$$

und der damit einhergehenden Quadrierung der Konditionszahl. Der wesentliche Trick des Algorithmus von Golub - Reinsch ist die

Vermeidung der expliziten Berechnung von B^*B bei obigem Interationsprozeß.

Wegen

$$M_{k+1} = RQ = Q^*Q\,RQ = Q^*M_k\,Q$$

sind M_{k+1} und M_k ähnlich und folglich ist M_k ähnlich zu $M_1 = B^*B$ für alle k.

Betrachte den Iterationsprozeß:

$$B_1 = B$$

(**)

$$B_{k+1} = S^*B_k\,T \qquad \text{, mit unitären Matrizen } S,\ T$$

Es gilt

$$B_{k+1}^*\,B_{k+1} = T^*B_k^*\,S\;\;S^*B_k\,T = T^*B_k^*\,B_k^*T$$

BEMERKUNG: $B_k^*\,B_k$ wird nicht explizit berechnet.

Die Idee besteht nun darin, S und T so zu wählen, daß (**) ein zu (*) äquivalenter Iterationsprozeß ist.

Wir benötigen folgendes

LEMMA 2.1: (Francis) A sei eine quadratische, M eine tridiagonale Matrix, deren Elemente auf der unteren Nebendiagonalen von Null verschieden sind, gilt dann

$$M = Q^*A\,Q$$

mit einer unitären Matrix Q, so ist Q bis auf die Rechtsmultiplikation mit einer Diagonalmatrix mit Elementen ± 1 eindeutig durch A und die erste Spalte von Q bestimmt.

BEWEIS: Sei $M = (m_{ij})$, $Q = (q_1,\ldots,q_n)$, dann lautet die erste Spalte der Gleichung $QM = AQ$:

$$m_{11} q_1 + m_{21} q_2 = A q_1$$

$$\Rightarrow \quad m_{11} = (q_1 , A q_1)$$

$$\Rightarrow \quad m_{21} = \pm \|m_{21} q_2\| = \pm \|Aq_1 - (q_1, A q_1) q_1\|$$

Falls $m_{21} \neq 0$ ist q_2 also bis auf das Vorzeichen durch q_1 und A bestimmt. Sukzessive Anwendung dieses Arguments auf die Spalten 2 bis n der Gleichung $QM = AQ$ liefert die Behauptung.

■

Damit (*) und (**) äquivalente Iterationen darstellen, müssen S, T wie folgt gewählt werden:

i) B_{k+1} bleibt bidiagonal

ii) die erste Spalte von T stimmt bis auf das Vorzeichen mit der ersten Spalte von Q aus (*) überein.

Vorausgesetzt die Nebendiagonalelemente von M sind von Null verschieden, so folgt aus Lemma 1.1 infolge $M_1 = B_1^* B_1$:

$$B_2^* B_2 = T^* M_1 T = D Q^* M_1 Q D = D M_2 D \quad ,$$

wobei D eine Diagonalmatrix mit Elementen ± 1 ist, d.h. es gilt insbesondere $D^{-1} = D$. Somit sind $B_2^* B_2$ und M_2 ähnlich, d.h. besitzen die gleichen Eigenwerte. Da das Produkt von Diagonalmatrizen obigen Typs natürlich wieder vom gleichen Typ ist, kommt man durch wiederholtes Anwenden von Lemma 1.1 zu:

$$B_k^* B_k = D M_k D$$

mit einer entsprechenden Diagonalmatrix D. Die Elemente von M_k sind also (bis auf das Vorzeichen) gleich den entsprechenden Elementen von $B_k^* B_k$. Da M_k gegen eine Diagonalmatrix konvergiert, gilt dies also auch für $B_k^* B_k$ und folglich ebenso für B_k; die Eigenwerte von $B_k^* B_k$ sind gleich denen von M_k, also konvergieren die Diagonalelemente von B_k tatsächlich gegen die Singulärwerte von B.

Die Matrizen S, T mit $B_{k+1} = S^* B_k T$ werden jeweils als ein Produkt von Givens-Rotationen bestimmt:

$$S = S_2 \cdot \ldots \cdot S_n \quad , \quad T = T_2 \cdot \ldots \cdot T_n \quad ,$$

wobei S_i, T_i Givens-Rotationen bzgl. $(i-1,i)$ sind.

Aufgrund der Tridiagonalität von M_k sind nur die beiden oberen Elemente der ersten Spalten von Q (= erste Spalte von $Q_1, Q = Q_1 \cdot \ldots \cdot Q_n$) von Null verschieden, d.h. T_2 kann so gewählt werden, daß die erste Spalte von T_2 (= erste Spalte von T) mit der von Q übereinstimmt. Damit ist die zweite Bedingung an S, T erfüllt, die erste erfüllt man durch folgendes "Chasing - Prozeß":

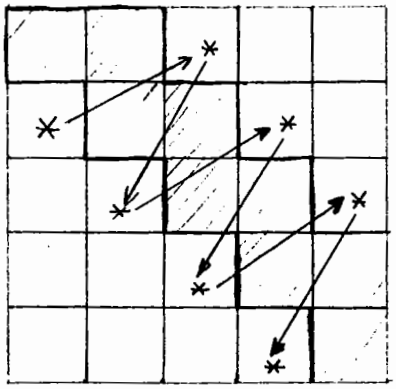
Falls $(B_k T_2)_{21} \neq 0$ wähle S_2 s.d. $(S_2 B_k T_2)_{21} = 0$

Falls $(S_2 B_k T_2)_{13} \neq 0$ wähle T_3 s.d. $(S_2 B_k T_2 T_3)_{13} = 0$

Falls $(S_2 B_k T_2 T_3)_{32} \neq 0$ wähle S_3 s.d. $(S_2 S_2 B_k T_2 T_3)_{32} = 0$

u.s.w.

Für $n = 5$ sei der Rechenablauf veranschaulicht:



Falls die Annahme des Lemmas von Francis nicht erfüllt ist, d.h. ein Element der unteren (und damit auch der oberen) Nebendiagonalen von $B_k^* B_k$ gleich Null ist, zerfällt diese Matrix in zwei Teilblöcke, deren Eigenwerte unabhängig voneinander berechnet werden können:

$$B_k^* B_k = \begin{pmatrix} \square & 0 \\ 0 & \square \end{pmatrix}$$

Für die weiteren Details (Shiftstretegien, Stopkriterien) sei auf die Originalarbeit verwiesen.

3. ITERATIONSVERFAHREN FÜR DIE LÖSUNG UNTER- UND ÜBERBESTIMMTER SYSTEME

3.1 ART (algebraic reconstruction technique, Kaczmarz 1937)

Es sei

$$Ax = b$$

ein überbestimmtes Gleichungssystem, wobei a_j, $j = 1,\ldots,m$ die auf 1 normierten Zeilen von A seien, also ist

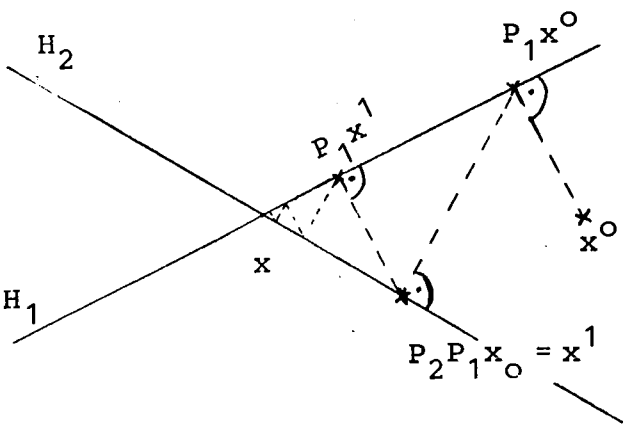
$$(a_j,x) = b_j \quad , \quad j = 1,\ldots,m$$

$$\text{mit} \quad \|a_j\| = 1 \quad , \quad j = 1,\ldots,m$$

zu lösen.

Sei $H_j = \{x \in \mathbb{R}^n \mid (a_j,x) = b_j\}$ und P_j die orthogonale Projektion auf H_j.

Die Idee des ART-Verfahrens ist es, x näherungsweise durch fortlaufende Orthogonalprojektion auf die H_j (d.h. durch Anwendung von P_j) zu bestimmen. Anschaulich:



P_j läßt sich leicht berechnen:

$$P_j x = x - \alpha_j a_j \quad ,$$

α_j wird aus $(a_j , P_j x) = b_j$ bestimmt:

$$(a_j, x) - \alpha_j \underbrace{(a_j , a_j)}_{=1} = b_j$$

$$\Rightarrow \quad \alpha_j = - b_j + (a_j , x)$$

$$\Rightarrow \quad P_j x = x + (b_j - (a_j , x)) a_j$$

$$P := P_m \cdot \ldots \cdot P_1$$

$$x^{k+1} = P x^k = P^k x^o$$

Wir wollen gleich das ART - Verfahren mit Relaxation betrachten:

$$P_j^\omega = (1 - \omega) \mathbb{1} + \omega P_j$$

$$P^\omega := P_m^\omega \cdot \ldots \cdot P_1^\omega$$

$$x^{k+1} = P^\omega x^k$$

Es gilt der

SATZ 3.1: Sei $\omega \in]0,2[$ und $x^o = 0$. Dann konvergiert $x^{t+1} = P^{\omega t} x^o$ gegen die in $R(A^*)$ eindeutig bestimmte Lösung von

$$A^* (D + \omega L)^{-1} (b - Ax) = 0 \quad ,$$

mit $A^* A = \begin{pmatrix} & L^* \\ L & \searrow D \end{pmatrix}$, d.h. $D = \mathbb{1}$, infolge $\|a_j\| = 1$, $\forall j$.

BEMERKUNG:

i) Das ART-Verfahren entspricht dem SOR - Verfahren für das System $AA^*u = b$, $x = A^*u$ (wurde zuerst von Björck - Elfing, 1979 bemerkt).

ii) Der Grenzwert ist (im Gegensatz zu SOR) von ω abhängig.

iii) Der Grenzwert ist von der Reihenfolge der Gleichungen abhängig.

iv) Für $\omega \to 0$ gilt: $x^{\omega} \to A^{+}b$.

BEWEIS: Wir zeigen nur die Äquivalenz zum SOR-Verfahren, die Aussage des Satzes folgt dann aus dem Konvergenzverhalten des SOR - Verfahrens (s. Prama 1); betrachte:

$$(*) \qquad AA^*y = b\,, \quad x = A^*y\,, \quad A = \begin{pmatrix} a_1^* \\ \vdots \\ a_m^* \end{pmatrix}\,, \quad A^* = (a_1,\ldots,a_m)$$

Das Einzelschrittverfahren für $(*)$ lautet:

$$\tilde{y}_i = y_{i-1} + (b_i - a_i^* A^* y_{i-1})e_i$$

$$\Rightarrow A^*y_i = A^*y_{i-1} + (b_i - a_i^* A^* y_{i-1})a_i$$

$$\Leftrightarrow \quad x_i = x_{i-1} + (b_i - a_i^* x_{i-1})a_i$$

Dies ist jedoch ein Iterationsschritt des ART - Verfahrens für $Ax = b$ mit $\omega = 1$, woraus die Äquivalenz des ART - Verfahrens für beliebige $\omega \in]0,2[$ mit dem SOR - Verfahren folgt.

■

3.2 Das CG - Verfahren

Es sei Q eine (symmetrische) positiv definite (n,n) - Matrix, dann sieht man durch Differenzieren, daß die Minimierungsaufgabe:

$$\min_{x \in \mathbb{R}^n} J(x) := \frac{1}{2} (x, Qx) - (b, x) \cdot$$

äquivalent zur Aufgabe

$$Q x = b$$

ist, deren Lösung wir mit x^* bezeichnen wollen. Für großes n kommt eine direkte Invertierung von Q in der Regel nicht in Frage und man ist auf iterative Verfahren angewiesen. Bei der Herleitung von Abstiegsverfahren nutzt man die Äquivalenz zur Minimierungsaufgabe aus:

$x^0, x^1, \ldots$ seien die Iterierten und $d^0, d^1, \ldots$ Richtungen,

wir setzen:

$$x^{t+1} = x^t + \alpha_t d^t$$

und bestimmen α_t derart, daß

$$\frac{1}{2} (x^{t+1}, Qx^{t+1}) - (b, x^{t+1})$$

minimal wird:

$$\frac{1}{2} (x^t + \alpha_t d^t, Q(x^t + \alpha_t d^t)) - (b, x^t + \alpha_t d^t)$$

$$= \frac{1}{2} (x^t, Qx^t) + \alpha_t (d^t, Qx^t) + \frac{1}{2} \alpha_t^2 (d^t, Qd^t) - (b, x^t) - \alpha_t (b, d^t)$$

Dieses quadratische Polynom in α_t hat sein Minimum bei

$$\alpha_t = - \frac{(d^t, Qx^t) - (b, d^t)}{(d^t, Q\, d^t)} = - \frac{(d^t, r^t)}{(d^t, Qd^t)} \quad \text{mit} \quad r^t = Q\, x^t - b \ ,$$

d.h. $r^t = \nabla J(x^t)$.

Je nach der Wahl der Richtungen, unterscheidet man:

i) Verfahren des steilsten Abstiegs: $d^t = - r^t$

ii) Verfahren mit konjugierten Richtungen: $(d^t, Qd^k) = 0, \ t \neq k.$

<u>SATZ 3.2:</u> Für jedes Verfahren mit konjugierten Richtungen gilt: Sei $D_t := sp(d_o, \ldots, d_t)$, dann minimiert x^t das Funktional $J(x)$ in der linearen Mannigfaltigkeit $x^o + D_{t-1}$, und es gilt $r^t \perp D_{t-1}$.

<u>BEWEIS:</u> $x \in x^o + D_{t-1}$ hat die Darstellung

$$x = x^o + \sum_{j<t} c_j d^j \ , \quad x^t = x^o + \sum_{j<t} \alpha_j d^j$$

$$J(x) = \frac{1}{2} (x, Qx) - (b, x) = \frac{1}{2} \underbrace{((x-x^*) , Q(x-x^*))}_{=:E(x)} - \frac{1}{2} (x^*, Q\, x^*) \ ,$$

somit ist Minimierung von J äquivalent zur Minimierung von E, wir betrachten folglich

$$E(x) = (x^o - x^* + \sum_{j<t} c_j d^j, Q(x^o - x^* + \sum_{j<t} c_j d^j))$$

$$= (x^o - x^*, Q(x^o - x^*)) + 2 \sum_{j<t} c_j \underbrace{(x^o - x^*, Qd^j)}_{=(r^o, d^j)} + \sum_{j,k<t} c_j c_k \underbrace{(d^j, Qd^k)}_{=0,\ j \neq k}$$

$$= E(x^o) + 2 \sum_{j<t} c_j (r^o, d^j) + \sum_{j<t} c_j^2 (d^j, Q\, d^j) \ .$$

Das Minimum von E in $x_o + D_{t-1}$ wird genau dann angenommen, wenn

$$\frac{\partial}{\partial c_j} E(x) = 0 \qquad , \quad 1 \leq j < t$$

gilt, also

$$c_j = - \frac{(r^o, d^j)}{(d^j, Qd^j)} \qquad , \quad 1 \leq j < t \quad .$$

Um $c_j = \alpha_j$, d.h. $(r^o - r^j, d^j) = 0$, $j < t$ z.zg., gehen wir zunächst auf die zweite Aussage des Satzes ein:

$r^t \perp d^j$, $j < t$ wird durch Induktion bewiesen:

Induktionsanfang: $t = 0$ ✓

Induktionsannahme: $r^t \perp d^j$, $j < t$ für ein $t \geq 0$

$$r^{t+1} = Q x^{t+1} - b = Q (x^t + \alpha_t d^t) - b = r^t + \alpha_t Q d^t$$

$$\Rightarrow (r^{t+1}, d^j) = (r^t, d^j) + \alpha_j (Qd^t, d^j)$$

$$= \begin{cases} 0 \ , \ j < t & \text{nach Ind.-annahme und Def. der } d^j \\ 0 \ , \ j = t & \text{nach Def. der } \alpha_j \end{cases}$$

Aufgrund der Darstellung von x^j gilt natürlich:

$$r^j = r^o + \sum_{k<j} \alpha_k Q d^k \qquad \text{und demnach für } j < t :$$

$$(r^j - r^o, d^j) = \sum_{k<j} \alpha_k (Q d^k, d^j) = 0 \quad .$$

■

Wir beschreiben nun das Verfahren der konjugierten Gradienten (CG):

$$d^0 = - r^0$$

$$x^{t+1} = x^t + \alpha_t d^t \ , \ \alpha^t = - \frac{(r^t, d^t)}{(Qd^t, d^t)} \ , \ r_t = Q x^t - b$$

$$d^{t+1} = - r^{t+1} + \beta_t d^t \ , \ \beta_t = \frac{(r^{t+1}, Qd^t)}{(Qd^t, d^t)}$$

Mit diesen Definitionen gilt der folgende

SATZ 3.3: CG ist ein Verfahren mit konjugierten Richtungen. Solange $r^t \neq 0$ (d.h. $x^t \neq x^*$) gilt, ist

$$sp(r^0, \ldots, r^t) = sp(d^0, \ldots, d^t) = sp(r^0, Qr^0, \ldots, Q^t r^0) \ .$$

BEWEIS:

i) Z.Zg. $(d^k, Qd^j) = 0 \quad , \quad k \neq j$

Induktionsannahme: Die Aussage sei richtig für $\{d^0, \ldots, d^t\}$.

Sei nun $j \leq t$, betrachte:

$$(d^{t+1}, Qd^j) = - (r^{t+1}, Qd^j) + \beta_t (d^t, Qd^j)$$

$$= 0 \quad ,$$

denn für $j = t$ gilt dies nach Definition von β_t; für $j < t$ verschwindet das zweite innere Produkt nach Induktionsannahme, nun gilt aber (wie gleich bewiesen wird):

$$d^j \in sp(r_o, \ldots, Q^j r^o) \Rightarrow Qd^j \in sp(Qr_o, \ldots, Q^{j+1} r^o) \overset{j<t}{\subset} sp(d^o, \ldots, d^t) \ ,$$

der Induktionsbeweis für den zweiten Teil von Satz 3.2 lehrt

jedoch, daß infolge unserer Induktionsannahme $r^{t+1} \perp d^j$, $j \leq t$ gilt und somit auch das zweite innere Produkt verschwindet.

ii) die Aussage sei richtig für ein $t \geq 0$.

$$r^{t+1} = r^t + \alpha_t Q d^t \overset{I.A.}{\in} sp(r_o,\ldots,Q^{t+1}r_o)$$

$$\Rightarrow \quad d^{t+1} = -r^{t+1} + \beta_t d^t \quad \in \quad sp(r_o,\ldots,Q^{t+1}r_o)$$

also:

$$sp\{r^o,\ldots,r^{t+1}\} \subseteq \{r^o,\ldots,Q^{t+1}r^o\}$$

$$sp\{d^o,\ldots,d^{t+1}\} \subseteq \{r^o,\ldots,Q^{t+1}r^o\}$$

und $\subsetneq$ kann in beiden Fällen nicht sein, da sonst $r^{t+1} \in sp\{d^o,\ldots,d^t\}$ gelten müßte, was für $r^{t+1} \neq 0$ ein Widerspruch zu $r^{t+1} \in sp(d^o,\ldots,d^t)$ ist.

■

SATZ 3.4: Es gilt: x^t hat die Darstellung

$$x^t = P_t(Q)r^o \quad ,$$

mit $P_t \in \mathcal{P}_{t-1}$ = Polynome vom Grad $\leq t-1$ und P_t minimiert das Funktional

$$(x^o - x^*, Q(\mathbb{1} + Q\,P(Q))^2(x^o - x^*))$$

in $\mathcal{P}_{t-1}$.

BEWEIS:

$$E(x^t) \leq E(x) \quad \forall\, x \in x_o + sp\{r^o, Qr^o, \ldots, Q^{t-1}r^o\} \quad ,$$

d.h. $\forall\, x$ mit der Darstellung $x = x_o + P(Q)r^o = x_o + P(Q)(Qx^o - \underbrace{b}_{=Qx^*})$,

wobei $P \in P_{t-1}$, also

$$x - x^* = x^0 - x^* + Q\,P(Q)\,(x^0 - x^*)$$

$$= (\mathbb{1} + Q\,P(Q))\,(x^0 - x^*)$$

$$\Rightarrow \quad E(x) = (x - x^*, Q(x - x^*))$$

$$= ((\mathbb{1} + Q\,P(Q))\,(x^0 - x^*), Q(\mathbb{1} + Q\,P(Q))\,(x^0 - x^*))$$

$$= (x^0 - x^*, Q(\mathbb{1} + Q\,P(Q))^2 (x^0 - x^*))$$

$$\Rightarrow \quad \text{Beh.}$$

Das CG - Verfahren für schlecht gestellte Probleme

Wir wollen das Problem

$$\min_x \|Ax - b\|$$

mit einer (m,n) - Matrix A mit $m \geq n = \text{Rang}(A)$ lösen. Dazu setzen wir $Q = A^*A$ und $c = A^*b$ und minimieren

$$\frac{1}{2}\,(x\,,\,Qx) - (x,c)$$

in x. Die Lösung dieser Minimierungsaufgabe ist bekanntermaßen $x^+ = Q^{-1}\,A^*b = A^+b$.

Mit Hilfe der Singulärwertzerlegung von A hat x^+ die Darstellung:

$$x^+ = \sum_{j=1}^{n} \frac{1}{\sigma_j}\,(b,u_j)v_j \quad , \quad \text{mit} \quad Q\,v_j = \sigma_j^2\,v_j \quad .$$

Wir setzen $\frac{1}{\sigma_j}\,(b,u_j) =: d_j^+$ und $\sigma_j^2 =: \lambda_j$, also

$$x^+ = \sum_{j=1}^{n} d_j^+\,v_j \qquad \text{und} \qquad Qv_j = \lambda_j\,v_j \quad .$$

Wir wollen den Fehler $E(x^t) = (x^t - x^+, Q(x^t - x^+))$ näher untersuchen, nach Satz 3.4 gilt:

$$E(x^t) = (x^o - x^+, (\mathbb{1} + Q P_t(Q))^2 Q(x^o - x^+)) \quad ,$$

wobei P_t dieses Funktional in P_{t-1} minimiert. Wir entwickeln nun x^o nach den v_j : $x^o = \sum_{j=1}^{n} d_j^o v_j$, also $x^o - x^+ = \sum_{j=1}^{n} (d_j^o - d_j^+) v_j$ und somit ergibt sich durch Einsetzen:

$$E(x^t) = \sum_{j=1}^{n} \lambda_j (d_j^o - d_j^+)^2 (1 + \lambda_j P_t(\lambda_j))^2 \quad .$$

Aufgrund der Minimierungseigenschaft von P_t können wir den heuristischen Schluß:

$$\lambda_j (d_j^o - d_j^+)^2 \text{ "groß"} \Rightarrow (1 + \lambda_j P_t(\lambda_j))^2 \text{ "klein"}$$

machen. Nun gilt aber auch:

$$x^t - x^+ = (\mathbb{1} + Q P_t(Q))(x^o - x^+) = \sum_{j=1}^{n} (d_j^o - d_j^+)(1+\lambda_j P_t(\lambda_j)) v_j$$

$$\Rightarrow x^t = (\mathbb{1} + Q P_t(Q)) x^o - Q P_t(Q) x^+$$

$$= \sum_{j=1}^{n} \Big[d_j^o \underbrace{(1 + \lambda_j P_t(\lambda_j))}_{\text{klein}} - d_j^+ \underbrace{\lambda_j P_t(\lambda_j)}_{\sim -1} \Big] v_j$$

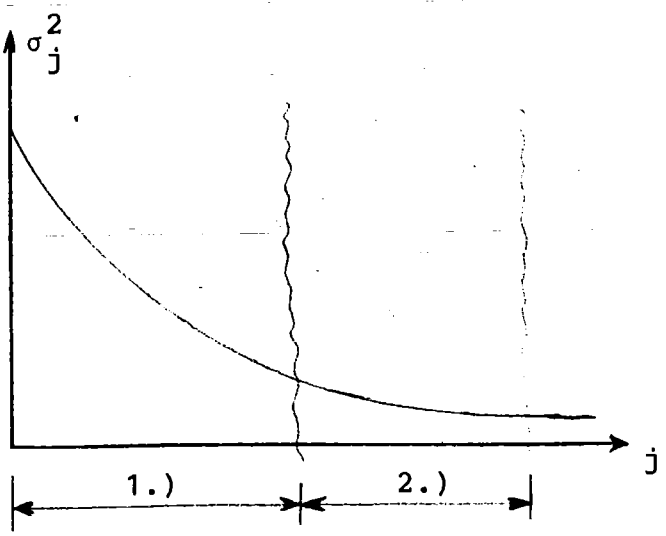
D.h. das CG - Verfahren iteriert "schwerpunktmäßig" auf den Komponenten, in denen $\lambda_j (d_j^o - d_j^+)^2$ groß ist. Die zu großen Eigenwerte bzw. zu großen Startfehlern gehörenden Komponenten werden also bevorzugt iteriert, hingegen arbeitet das Verfahren nicht auf Komponenten mit $d_j^o - d_j^+ \approx 0$.

Diese Beobachtung führt zum

CG - Verfahren mit Neustart:

1) Führe so viele CG - Schritte durch, bis der Fehler in den Unterräumen, in denen $\lambda_j (d_j^o - d_j^+)^2$ groß ist, vernachlässigbar ist.

2) Verwende die so gewonnene Näherungslösung als Startwert für ein neues CG - Verfahren.



Nach dem Neustart iteriert das Verfahren vornehmlich auf den durch "2.)" gekennzeichneten Komponenten.

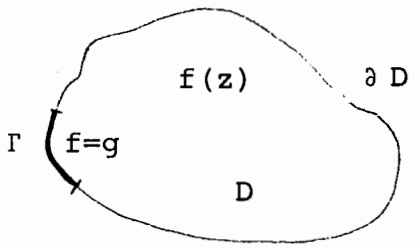
III Spezielle schlecht gestellte Probleme

1. Analytische Fortsetzung

Es sei D ein beschränktes Gebiet und ∂D dessen Rand. Γ sei ein Teilbogen dieses Randes. Wir wollen eine Funktion f vermöge der Vorgaben:

i) f ist analytisch in einem D umfassenden Gebiet

ii) $f_{|\Gamma} = g$



bestimmen.

Der Identitätssatz für holomorphe Funktionen (Fischer, Lieb - Funktionentheorie S. 74) besagt, daß f durch g eindeutig bestimmt ist.

Wir wollen nun die Stabilität dieses Fortsetzungsproblems untersuchen, die Anfänge solcher Stabilitätsbetrachtungen gehen auf Carleman ($\sim$ 1920) zurück.

Wir benötigen im folgenden zwei Hilfsmittel:

i) Das Maximum - Minimum - Prinzip für harmonische Funktionen (auch Randmaximumsatz): Eine auf dem beschränkten Gebiet D harmonische und auf $\overline{D}$ stetige Funktion nimmt ihr Maximum und Minimum auf ∂D an. (u harmonisch $\Leftrightarrow \Delta u = 0$) (s. z.B. Fischer - Lieb, S. 93)

ii) Den Begriff des Harmonischen Maßes:
D sei ein Gebiet, ∂D eine endliche Vereinigung rektifizierbarer Jordanbögen, Γ ein endlicher Teilbogen von ∂D,

dann gibt es genau eine harmonische Funktion $w_{D,\Gamma}(z)$ auf D mit

$$w_{D,\Gamma}(z) = 0 \text{ in den inneren Punkten von } \partial D - \Gamma$$

$$w_{D,\Gamma}(z) = 1 \text{ in den inneren Punkten von } \Gamma .$$

$w_{D,\Gamma}$ heißt das harmonische Maß von Γ bzgl. D.

Aufgrund des Maximumprinzips gilt: $0 \leq w_{D,\Gamma}(z) \leq 1$, $z \in D$.
(s. E. Hille, Analytic Function Theory II, S. 408 ff.)

Es gilt nun der

SATZ 1.1: Sei f analytisch in D, $\Gamma \subset \partial D$, w das harmonische Maß von Γ bzgl. D, und es gelte

$$|f(z)| \leq \delta \quad \text{auf} \quad \Gamma$$

$$|f(z)| \leq \rho \quad \text{auf} \quad \partial D - \Gamma .$$

Dann gilt $\forall\, z \in D$:

$$|f(z)| \leq \delta^{w(z)} \rho^{1-w(z)} .$$

BEWEIS: Sei zunächst $f(z) \neq 0 \quad \forall\, z \in D$ vorausgesetzt.

Betrachte: $h(z) = \ln |f(z)|$, wir behaupten $\Delta h = 0$.

BEWEIS dazu: $\operatorname{Log}(f(z)) = \ln |f(z)| + i \arg(f(z))$,

d.h. h ist Realteil der holomorphen Funktion $\operatorname{Log} \circ f$,

also harmonisch.

Nach Definition gilt:

$$h(z) \leq \begin{cases} \ln \rho , & z \in \partial D - \Gamma \\ \ln \delta , & z \in \Gamma \end{cases}$$

Definiere nun $H(z) = w(z)\ln\delta + (1 - w(z))\ln\rho$, dann gilt trivialerweise $\Delta H = 0$ und

$$H(z) = \begin{cases} \ln\rho & , \quad z \in \partial D - \Gamma \\ \ln\delta & , \quad z \in \Gamma \end{cases}$$

$\Rightarrow$ $h - H \leq 0$ auf ∂D, da $h - H$ harmonisch ist, gilt folglich

$h - H \leq 0$ auf D

$\Rightarrow$ $e^{h(z)} \leq e^{H(z)}$ auf D

$\Rightarrow$ $|f(z)| \leq \exp(w(z) \cdot \ln\delta + (1 - w(z))\ln\rho)$

$= \delta^{w(z)} \cdot \rho^{1-w(z)}$.

Wir müssen noch den Fall einer Nullstelle von f berücksichtigen:

sei $f(z_o) = 0$, also $\lim\limits_{z\to z_o} h(z) = -\infty$, dann existiert für jede Konstante C eine Umgebung U_C von z_o, so daß

$$h(z) < -C \quad \text{auf} \quad \overline{U}_C .$$

Wir wenden nun die obige Argumentation auf das Gebiet $D - \overline{U}_C$ an und erhalten wieder

$$h - H \leq 0 \quad \text{auf} \quad \partial(D - \overline{U_C}) = \partial D \cup \partial U_C \quad ,$$

falls C passend gewählt wurde.

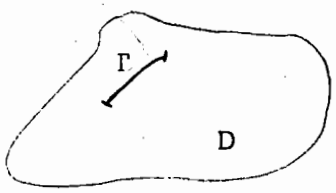
Damit $h \leq H$ und folglich

$$|f(z)| \leq \delta^{w(z)} \rho^{1-w(z)} \quad \text{auf} \quad D - \overline{U_C} \quad .$$

Die Aussage bleibt richtig, falls nun U_C beliebig verkleinert wird, während sie für den Punkt z_o selbst trivial ist, also gilt sie auf ganz D. ∎

2. NUMERISCHE VERFAHREN ZUR ANALYTISCHEN FORTSETZUNG

i) Sei D ein Gebiet, f analytisch in D, stetig auf $\bar{D}$ $f_{|\Gamma} = g$, $\Gamma \subset D$, gesucht ist f in ganz D.

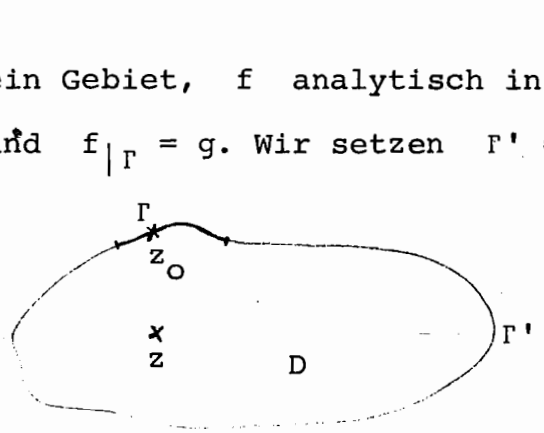


Wir nutzen die Cauchysche Integralformel (CIF) aus

$$g(z) = f(z) = \frac{1}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{\zeta - z}\, d\zeta \qquad \text{für } z \in \Gamma .$$

Dies ist eine Integralgleichung 1. Art für $f_{|\partial D}$, hat man $f_{|\partial D}$ erst einmal bestimmt, kennt man f via CIF auf ganz D. (Die Methode wurde zuerst von J. Douglas ~ 1950 angewendet.)

ii) Sei D ein Gebiet, f analytisch in D, stetig auf $\bar{D}$, $\Gamma \subseteq \partial D$ und $f_{|\Gamma} = g$. Wir setzen $\Gamma' = \partial D - \Gamma$.



Die CIF lautet nun:

$$f(z) = \frac{1}{2\pi i}\left(\int_{\Gamma} \frac{g(\zeta)}{\zeta - z}\, d\zeta + \int_{\Gamma'} \frac{f(\zeta)}{\zeta - z}\, d\zeta \right), \qquad z \in D$$

Wir wählen ein $z_o \in \Gamma$ mit $\inf\limits_{w \in \Gamma'} |z_o - w| > 0$ und führen den Grenzübergang $z \to z_o$ aus:

$$g(z_o) = f(z_o) = \frac{1}{2\pi i} \lim_{z \to z_o} \left(\int_\Gamma \frac{g(\zeta)}{\zeta - z}\, d\zeta \right) + \frac{1}{2\pi i} \int_{\Gamma'} \frac{f(\zeta)}{\zeta - z_o}\, d\zeta$$

Vorausgesetzt wir können den Grenzübergang bilden, ist dies eine Integralgleichung 1. Art für $f_{|\Gamma'}$.

Wir wollen an dem Spezialfall $\Gamma = [-1,1]$, $z_\varepsilon = i\varepsilon$, zeigen, daß es aufwendig sein kann, den Grenzwert zu berechnen.

LEMMA 2.1: Es gilt für $g \in C^1[-1,1]$

$$\lim_{\varepsilon \downarrow 0} \int_{-1}^{1} \frac{g(x)}{x - i\varepsilon}\, dx = \oint_{-1}^{1} \frac{g(x)}{x}\, dx + i\pi g(0) \quad ,$$

wobei $\oint_{-1}^{1} \frac{g(x)}{x}\, dx$ für den Cauchy-Hauptwert, also für

$$\lim_{\delta \downarrow 0} \int_{\delta \le |x| \le 1} \frac{g(x)}{x}\, dx$$

steht.

BEWEIS:

$$\int_{-1}^{1} \frac{g(x)}{x - i\varepsilon}\, dx = \int_{-1}^{1} \frac{g(x) - g(0)}{x - i\varepsilon}\, dx + \int_{-1}^{1} \frac{g(0)}{x - i\varepsilon}\, dx$$

Nach dem Mittelwertsatz ist $|g(x) - g(0)| \le |x| \cdot \underbrace{\max_{y \in I} |g'(y)|}_{\le C}$

($I := [\min(0,x), \max(0,x)]$), also

$$\left| \frac{g(x) - g(0)}{x - i\varepsilon} \right| \le \frac{x}{(x^2 + \varepsilon^2)^{1/2}}\, C \le C \quad ,$$

d.h. der Integrand ist unabhängig von ε beschränkt, und folglich

$$\lim_{\varepsilon \downarrow 0} \int_{-1}^{1} \frac{g(x)-g(0)}{x - i\varepsilon}\, dx = \int_{-1}^{1} \frac{g(x)-g(0)}{x}\, dx$$

$$= \int_{0<\delta<|x|\leq 1} \frac{g(x)}{x}\, dx + \int_{-\delta}^{\delta} \frac{g'(0)x+o(|x|)}{x}\, dx \quad (\delta \in]0,1[\text{ beliebig})$$

$$= \int_{0<\delta<|x|<1} \frac{g(x)}{x}\, dx + \int_{-\delta}^{\delta} \frac{o(|x|)}{x}\, dx \ .$$

$$= \not\int \frac{g(x)}{x}\, dx + \underbrace{\lim_{\delta \downarrow 0} \int_{-\delta}^{\delta} \frac{o(|x|)}{x}\, dx}_{=0}$$

Ferner gilt:

$$\int_{-1}^{1} \frac{g(0)}{x-i\varepsilon}\, dx = g(0) \underbrace{\int_{-1}^{1} \frac{x}{x^2+\varepsilon^2}\, dx}_{=0} + g(0)\, i\varepsilon \cdot \int_{-1}^{1} \frac{dx}{x^2+\varepsilon^2}$$

$$\overset{y=\varepsilon x}{=} g(0)\, i\varepsilon^2 \int_{-1/\varepsilon}^{1/\varepsilon} \frac{dy}{\varepsilon^2(1+y^2)} = g(0)\, i \left(\arctan\left(\frac{1}{\varepsilon}\right) - \arctan\left(-\frac{1}{\varepsilon}\right)\right)$$

$$\xrightarrow[\varepsilon \to 0]{} g(0)\, i\, \pi$$

■

ii') Falls $\Gamma = \partial D$:

$$f(z) = \frac{1}{2\pi i} \int_{\Gamma} \frac{g(\zeta)}{\zeta - z}\, d\zeta \quad , \quad z \in D$$

Die Aufgabe ist gut gestellt (das harm. Maß ist $w(z) \equiv 1$).

iii) Aufgabe wie in i), nur sei jetzt $D = D_1(0)$ der Kreis um 0 mit Radius 1.

Man entwickle f in eine Potenzreihe:

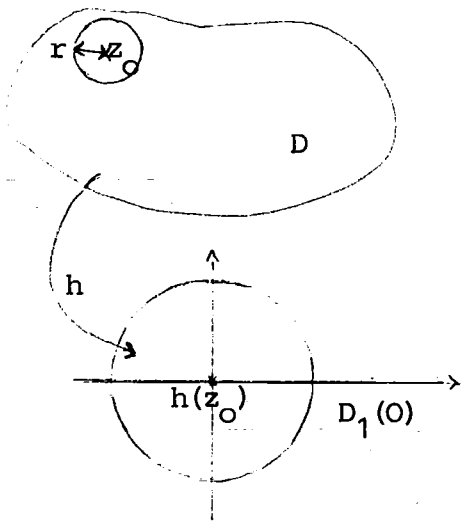
$f(z) = \sum_{k=0}^{\infty} a_k z^k$ und bestimme die Entwicklungskoeffizienten

durch Lösen der Aufgabe:

$$\min \|f-g\|^2_{L_2(\Gamma)} + \gamma^2 \|f\|_{L_2(\partial D)}$$

Diese Methode geht auf K. Miller zurück.

iv)



Die Potenzreihenentwicklung von f im Punkte z_o sei bekannt

$$f(z) = \sum_{k=0}^{\infty} a_k (z-z_o)^k \quad , \quad |z-z_o| < r$$

Falls $D \subsetneq \mathbb{C}$ einfach zusammenhängend ist, können wir es biholomorph auf den Einheitskreis abbilden (Riemannscher Abbildungssatz). Definiere $g(h(z)) = f(z)$ mit $h(D) = D_1(0)$, $h(z_o) = 0$, dann besitzt g eine Potenzreihenentwicklung

$$g(w) = \sum_{k=0}^{\infty} b_k w^k \quad ,$$

deren Koeffizienten b_k aus den a_k, $k \in \mathbb{N}$ berechenbar sind.

v) // Carleman - Funktion (Problem wie in ii))

Eine Funktion $G_\gamma(z,\zeta)$ heißt Carleman - Funktion für Γ und D, falls gilt:

a) $G_\gamma(z,\zeta) = \frac{1}{\zeta - z} + \tilde{G}_\gamma(z,\zeta)$, $\tilde{G}_\gamma$ analytisch in $\zeta \in D$ und stetig auf $\overline{D} \times \overline{D}$

b) $\frac{1}{2\pi} \int_{\Gamma'} |G_\gamma(z,\zeta)||d\zeta| \leq \gamma$

Um damit das Fortsetzungsproblem zu lösen, setzen wir

$$f_\gamma(z) = \frac{1}{2\pi i} \int_{\Gamma} G_\gamma(z,\zeta) f(\zeta) d\zeta \quad ,$$

es gilt aufgrund von a) sowie des Cauchyschen Integralsatzes

$$f(z) = \frac{1}{2\pi i} \int_{\Gamma+\Gamma'} G_\gamma(z,\zeta) f(\zeta) d\zeta$$

und somit

$$|f_\gamma(z) - f(z)| \leq \frac{1}{2\pi} \int_{\Gamma'} |(G_\gamma(z,\zeta) f(\zeta))||d\zeta|$$

$$\leq \gamma \max_{z \in \partial D} |f(z)| \xrightarrow{\gamma \to 0} 0 \quad .$$

Die Carleman - Funktion läßt sich leicht mit Hilfe des harmonischen Maßes angeben:

Sei w das harmonische Maß bzgl. D und Γ, so ist

$$G_\gamma(z,\zeta) = \frac{1}{\zeta - z} \cdot e^{\lambda(z)(\varphi(\zeta) - \varphi(z))}$$

mit: $\operatorname{Re} \varphi = w$, φ auf D analytisch

$$\lambda(z) = \frac{1}{w(z)} \left\{ \ell n(2\pi\gamma) - \ell n \int_{\Gamma'} \left| \frac{d\zeta}{\zeta - z} \right| \right\}$$

eine Carleman - Funktion zu D und Γ.

Die Aufgabe, zu einer harmonischen Funktion w eine holomorphe Funktion φ zu finden, s.d. $\operatorname{Re} \varphi = w$ ist, ist für einfach zusammenhängende Gebiete stets lösbar. Man nimmt $\varphi = w + i\, 2\, \operatorname{Im}(F)$, wobei F eine Stammfunktion der holomorphen Funktion $\frac{\partial w}{\partial z}$ ist.

Wir wollen die Eigenschaften nachrechnen:

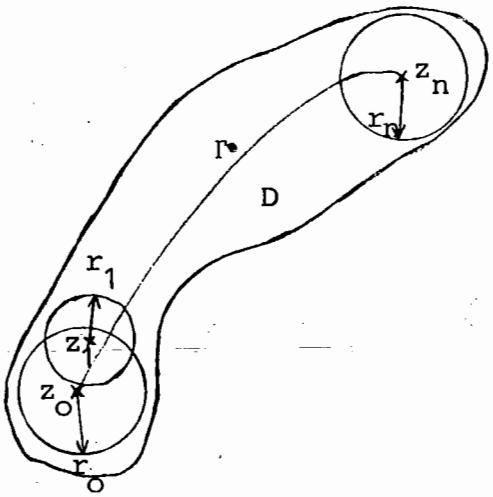
zu a) Die Funktion $\lambda(z)(\varphi(\zeta)-\varphi(z))$ ist trivialerweise analytisch in ζ, für festes z läßt sich diese Funktion in eine Potenzreihe um z entwickeln, in der das konstante Glied fehlt, folglich hat $e^{\lambda(z)(\varphi(\zeta)-\varphi(z))}$ für feste z eine Potenzreihenentwicklung der Form: $1 + \sum_{k=1}^{\infty} b_k(\zeta-z)^k$ und damit G_γ die angegebene Darstellung.

zu b)

$$\int_{\Gamma'} |G_\gamma(z,\zeta)||d\zeta| = \int_{\Gamma'} \frac{1}{|\zeta-z|} e^{\lambda(z)\overbrace{(w(\zeta)-w(z))}^{=0}} |d\zeta|$$

$$= \int_{\Gamma'} \frac{1}{|\zeta-z|} \left(2\pi\gamma \cdot \left(\int_{\Gamma'} \left| \frac{d\zeta}{\zeta-z} \right| \right)^{-1} |d\zeta| = 2\pi\gamma \right.$$

DT vi) Weierstraßsche Methode



Wir entwickeln f in $z_o \in \Gamma$ in eine Potenzreihe und wählen einen Punkt z_1 innerhalb des Konvergenzgebietes, in welchem wir die verschobene Potenzreihenentwicklung berechnen, wir wiederholen solange, bis wir am Punkt z_n angekommen sind.

Sei $f(z) = \sum_{k=0}^{\infty} a_k(z_o)(z-z_o)^k$ für $|z-z_o| < r_o$,

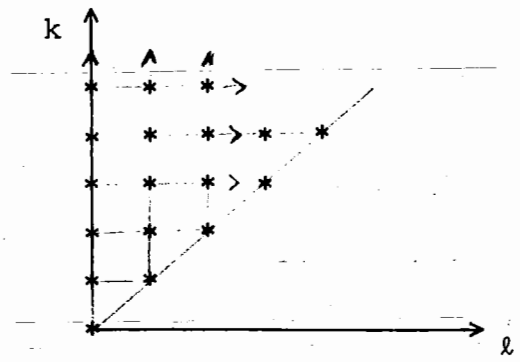
wir berechnen daraus die Potenzreihenentwicklung um z_1 :

$$(z - z_o)^k = (z - z_1 + h_1)^k \quad , \quad \text{mit} \quad h_1 = z_1 - z_o$$

$$= \sum_{\ell=0}^{k} (z - z_1)^\ell \, h_1^{k-\ell} \binom{k}{\ell}$$

$$\Rightarrow \quad f(z) = \sum_{k=0}^{\infty} a_k(z_o) \sum_{\ell=0}^{k} (z - z_1)^\ell \, h_1^{k-\ell} \binom{k}{\ell}$$

Wir dürfen nun die Summationsreihenfolge vertauschen:



$$f(z) = \sum_{\ell=0}^{\infty} (z - z_1)^\ell \underbrace{\sum_{k=\ell}^{\infty} a_k(z_o) h^{k-\ell} \binom{k}{\ell}}_{= a_\ell(z_1)}$$

Mit den Definitionen:

$$a(z_1) = \begin{pmatrix} a_o(z_1) \\ a_1(z_1) \\ \vdots \\ \vdots \end{pmatrix} \quad , \quad (M(h))_{\ell k} = \begin{cases} \binom{k}{\ell} h^{k-\ell} & , \quad \ell \geq k \\ 0 & , \quad \ell < k \end{cases}$$

gilt also:

$$a(z_1) = M(h_1) \cdot a(z_o) \quad .$$

Es gilt das

LEMMA 2.2: Mit den obigen Bezeichnungen besteht die Identität:

$$M(h)M(h') = M(h+h') \quad .$$

BEWEIS: Sei $k \geq \ell$, dann gilt:

$$(M(h)M(h'))_{\ell k} = \sum_{m=\ell}^{k} (M(h))_{\ell m} \cdot (M(h'))_{mk}$$

$$= \sum_{m=\ell}^{k} \binom{m}{\ell}\binom{k}{m} h^{m-\ell} h'^{k-m}$$

$$\overset{m'=m-\ell}{=} \sum_{m'=0}^{k-\ell} \frac{(m'+\ell)!\ k!}{\ell!\ m'!\ (k-m'-\ell)!\ (m'+\ell)!} h^{m'} \cdot h'^{k-m'-\ell}$$

$$= \binom{k}{\ell} \sum_{m'=0}^{k-\ell} \binom{k-\ell}{m'} h^{m'} h'^{(k-\ell)-m'}$$

$$= \binom{k}{\ell} (h+h')^{k-\ell} = (M(h+h'))_{\ell k}$$

■

Wir haben also:

$$a(z_n) = M(h_n) \cdot \ldots \cdot M(h_1) a(z_o) \quad , \quad h_k = z_k - z_{k-1}$$

i.a.
$$\neq M(z_n - z_o) a(z_o)$$ (Matrizenprodukt für ∞-liche Matrizen nicht assoziativ)

Sei nun $a^p(z_i) = \begin{pmatrix} a_o(z_i) \\ \vdots \\ a_p(z_i) \end{pmatrix}$ der vorne abgeschnittene Vektor der Länge p und $M^{p,q}$ die führende p,q-Untermatrix von M, dann gilt:

$$a^p(z_n) = M^{p,p}(h_n) \cdot \ldots \cdot M^{p,p}(h_1) a^p(z_o)$$

$$= M^{p,p}(z_n - z_o) a^p(z_o)$$

Diese Vorgehensweise ist infolge der Instabilität des Problems so nicht brauchbar, sinnvoll ist es hingegen nach jedem Schritt die Anzahl der berechneten Koeffizienten zu verkleinern, d.h. die höheren (schlecht bestimmten) Koeffizienten einfach abzuschneiden, konkret:

Für $\gamma \in]0,1[$ setze $p_k = \lfloor p \gamma^k \rfloor$, (also $p_o = p$) und berechne

$$a_n^p = M^{p_n, p_{n-1}}(h_n) \cdot \ldots \cdot M^{p_1, p}(h_1) a_o^p \quad .$$

Diese Methode stammt von Painlevé (1899) und wurde von Henrici (1966) aktualisiert. Painlevé zeigte folgenden

SATZ 2.3: Es gibt ein $\gamma_o \in]0,1[$, s.d. für $0 < \gamma < \gamma_o$ gilt:

$$\lim_{p \to \infty} \| a_n^p - a^p(z_n) \|_\infty = 0 \quad .$$

BEWEIS: s. P. Henrici: Applied and Computational Complex Analysis, Vol. 1, p. 174 - 178.

3. ABELSCHE INTEGRALGLEICHUNG

Wir wollen in diesem Abschnitt Abelsche Integralgleichungen auf ihre Schlechtgestelltheit hin untersuchen, es handelt sich dabei um Integralgleichungen der Form:

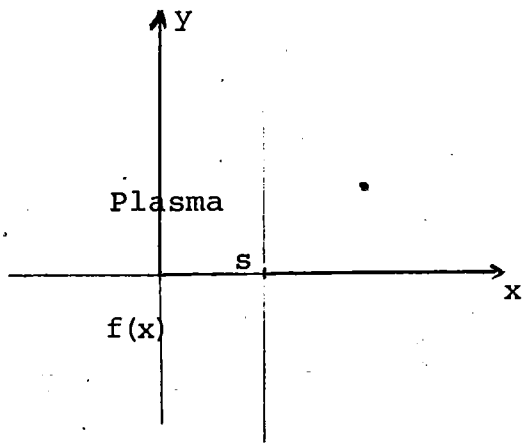
$$\int_0^x \frac{K(x,t)}{|x-t|^{\alpha}} f(t)dt = g(x) \quad , \quad x \geq 0$$

mit $K \in C^1 \times C^1$, $K(x,x) = 1$, $x \geq 0$

und $\alpha \in [0,1[$.

BEISPIELE:

1) Es sei f eine radialsymmetrische Strahlungsdichte, d.h. $f(x,y) = f(\sqrt{x^2+y^2})$, und wir wollen f aus den Daten:



$$g(s) = \int_{-\infty}^{\infty} f(\sqrt{s^2+y^2})\,dy = 2\int_0^{\infty} f(\sqrt{s^2+y^2})\,dy$$

bestimmen.

Mit der Substitution $u = \sqrt{s^2+y^2}$ erhalten wir:

$$g(s) = 2\int_s^{\infty} \frac{uf(u)}{\sqrt{u^2-s^2}}\,du$$

$$= \sqrt{2}\int_s^{\infty} \frac{u^{1/2} f(u)\cdot K(s,u)}{|u-s|^{1/2}}\,du \quad , \quad K(s,u) = \sqrt{2}\,\frac{u^{1/2}}{|u+s|^{1/2}}$$

2) Stereologie (s. L.A. Santalo: Integral Geometrie and Geometric Probability)

Es seien Kugeln verschiedener Radien in einem Körper verteilt, es soll gelten:

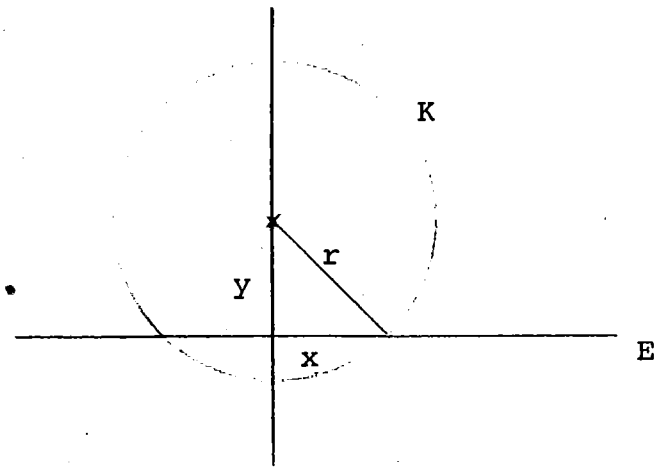
i) Die Verteilung der Radien sei durch $f(r)$ beschrieben.

ii) Die Mittelpunkte der Kugeln seien gleichverteilt.

iii) Die Radien und Mittelpunkte stellen voneinander unabhängige Zufallsvariablen dar.

Gegeben ist die Verteilung $g(x)$ der Kreisradien in einer Schnittebene E durch den Körper, gesucht ist $f(r)$.

Betrachte eine Kugel K mit Radius r:



Der Kreis $K \cap E$ hat einen Radius aus dem Intervall $[x, x+dx]$ genau dann, wenn

der Mittelpunkt von K einen Abstand von E aus dem Intervall $[y, y+dy]$ mit $y = \sqrt{r^2-x^2}$, $dy = \frac{x\,dx}{\sqrt{r^2-x^2}}$ hat.

Wir argumentieren nun wie folgt:

$f(r)dr \quad \hat{=}$ Anzahl der Kugeln in $[r, r+dr]$

$g(x)dx \quad \hat{=}$ Anzahl der Kreise in $[x, x+dx]$

$dy = \frac{x\,dx}{\sqrt{r^2-x^2}} \quad \hat{=}$ Wahrscheinlichkeit dafür, daß eine Kugel mit Radius $\in [y, y+dy]$ als Schnittbild einen Kreis mit Radius $\in [x, x+dx]$ erzeugt.

Also

$$C \cdot \int_x^\infty dy\, f(r)dr = g(x)dx$$

$$\Rightarrow \quad g(x) = c \int_x^\infty \frac{x\, f(x)}{\sqrt{r^2-x^2}}$$

Für den Fall $K \equiv 1$, d.h.

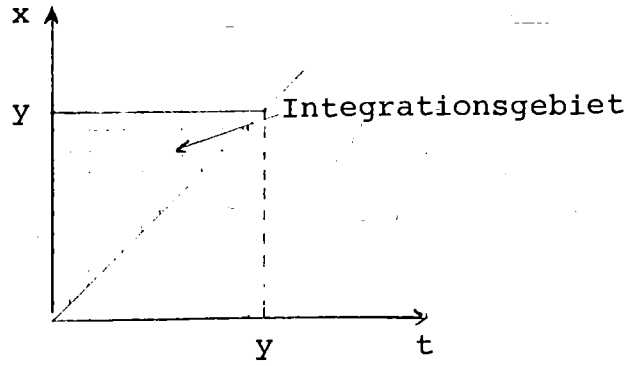
$$g(x) = \int_0^x \frac{f(t)}{(x-t)^\alpha}\, dt$$

kann man leicht eine explizite Inversionsformel angeben (um Triviales auszuschließen sei $\alpha > 0$):

$$g(x) = \int_0^x \frac{f(t)}{(x-t)^\alpha}\, dt$$

$$\Rightarrow \int_0^y (y-x)^{\alpha-1} \int_0^x \frac{f(t)}{(x-t)^\alpha}\, dt\, dx = \int_0^y (y-x)^{\alpha-1}\, g(x)dx$$

Auf der linken Seite dieser Gleichung vertauschen wir nun die Integrationsreihenfolge:



$$\ell.S. = \int_0^y f(t) \int_t^y (y-x)^{\alpha-1}(x-t)^{-\alpha}\, dx\, dt$$

Wir substituieren nun:

$$\xi = \frac{x-t}{y-t} \quad \text{also} \quad x = (1-\xi)t + \xi y \quad , \quad \xi \in [0,1]$$

$$\frac{d\xi}{dx} = \frac{1}{y-t} \quad , \quad x - t = \xi(y-t) \quad , \quad y - x = (1-\xi)(y-t)$$

$$\Rightarrow \quad \ell.S. = \int_0^y f(t) \int_0^1 (1-\xi)^{\alpha-1} (y-t)^{\alpha-1} \cdot \xi^{-\alpha} \cancel{(y-t)^{-\alpha}(y-t)} d\xi\, dt$$

$$= \int_0^y f(t) \underbrace{\int_0^1 (1-\xi)^{\alpha-1} \xi^{-\alpha}\, d\xi}_{= B(1-\alpha,\alpha) \text{ (Betafunktion)}} dt$$

$$= \frac{\Gamma(1-\alpha)\Gamma(\alpha)}{\Gamma(1)} \int_0^y f(t)dt$$

$$= \frac{\pi}{\sin \pi\alpha} \int_0^y f(t)dt$$

$$\Rightarrow \quad f(y) = \frac{\sin \pi\alpha}{\pi} \frac{d}{dy} \int_0^y (y-x)^{\alpha-1} g(x)\, dx \quad .$$

Wir betrachten nun für allgemeines $K \in C^1 \times C^1$, $K(x,x) = 1$

den Fall $\alpha = 0$:

$$(*) \qquad \int_0^x K(x,t)f(t)dt = g(x) \quad ,$$

d.h. die Volterrasche Integralgleichung.

SATZ 3.1: Sei $g \in C^1$, $g(0) = 0$, dann ist (*) eindeutig lösbar, und es gilt für $x \in [0,a]$

$$\|f\|_{L_\infty(0,a)} \leq C(a) \|g'\|_{L_\infty(0,a)} \quad .$$

BEWEIS: (*) ist äquivalent zu

$$f(x) + \underbrace{\int_0^x K_x(x,t)f(t)dt}_{=: Af(x)} = g'(x) \quad .$$

Damit

$$(*) \quad \Leftrightarrow \quad f + Af = g'$$

Da $K \in C^1 \times C^1$ existiert $C_1 > 0$ mit $|K_x(x,t)| < C_1 \quad \forall\, x,t$

$$\Rightarrow \quad |Af(x)| \leq \int_0^a C_1 |f(t)|dt \leq C_1\, a\, \|f\| \quad ,$$

wobei wir den $C[0,a]$ mit der Maximumsnorm

$$\|f\| = \max_{x\in[0,a]} |f(x)|$$

versehen haben.

Also ist

$$\|A\| \leq C_1\, a \quad ,$$

sei nun a so klein, daß gilt $C_1\, a < 1$, so ergibt sich:

$$\| (\mathbb{1} + A)^{-1} \| \leq \frac{1}{1-\|A\|} \quad .$$

(Dies folgt aus der Darstellung von $(\mathbb{1}+A)^{-1}$ als Neumannsche Reihe)

$$\Rightarrow \quad \|f\| \leq \frac{1}{1-\|A\|} \|g'\| \quad .$$

D.h. der Beweis ist für $x \in [0,a]$ mit $a < \frac{1}{c_1}$ erbracht, sei nun $x \in [a,2a]$:

$$f(x) + \int_0^x K_x(x,t) f(t) dt = g'(x)$$

$$\Rightarrow \quad f(x) + \int_a^x K_x(x,t) f(t) dt = g'(x) - \int_0^a K_x(x,t) f(t) dt$$

Wie oben folgt:

$$\|f\|_{L_\infty(a,2a)} \leq \frac{1}{1-\|A\|} \|g'(x) - \int_0^a K_x(x,t) f(t) dt\|_{L_\infty(a,2a)}$$

$$\leq \frac{1}{1-\|A\|} \|g'\|_{L_\infty(a,2a)} + \|A\| \|f\|$$

$$\leq \|g'\|_{L_\infty(0,2a)} \underbrace{\frac{1}{1-\|A\|} \left(1 + \frac{\|A\|}{1-\|A\|} \right)}_{= C(2a)} \quad .$$

Die letzte Argumentation läßt sich beliebig oft wiederholen, d.h. der Satz ist für beliebig große a gezeigt.

■

Wir sind in der Lage, eine Stabilitätsaussage für die Volterrasche Integralgleichungsaufgabe zu machen:

Der obige Satz liefert

$$\|f\| \leq C(a)\|g'\| \quad ,$$

mit $\alpha(\rho,\delta) = \max\{\|f\| : \|f'\| \leq \rho\,,\ \|g\| \leq \delta\}$

gilt dann (vgl. I.3, Numerische Differentiation):

$$\alpha(\rho,\delta) \leq C\,\rho^{1/2}\,\delta^{1/2} \quad .$$

Wir wollen dieses Resultat auf einem anderen Weg herleiten.

SATZ 3.2: Mit den obigen Bezeichnungen gilt

$$\alpha(\rho,\delta) = C\,\rho^{1/2}\,\delta^{1/2} \quad , \quad \text{falls} \quad [0,a]$$

ein hinreichend kleines Intervall ist.

BEWEIS:

$$\int_0^{x+h} K(x+h,t)f(t) - \int_0^{x} K(x,t)f(t)dt = g(x+h) - g(x)$$

$$\Rightarrow \underbrace{\int_x^{x+h} K(x+h,t)f(t)dt}_{=: I_1} - \underbrace{\int_0^{x} [K(x,t) - K(x+h,t)]f(t)dt}_{=: I_2} = g(x+h)-g(x)$$

Wir stellen uns h als klein vor und wollen nun die Ordnung bestimmen, mit welcher die Integrale von h abhängen:

$$I_1 = \int_x^{x+h} K(x+h,t)dt\ f(x) + \underbrace{\int_x^{x+h} K(x+h,t)(f(t)-f(x))dt}_{I_{12}}$$

$$= f(x)\left[\int_x^{x+h} K(x,x)dt + \int_x^{x+h} K(x+h,t)-K(x,x)dt\right] + I_{12}$$

Nun gilt:

$$K(x+h,t) - K(x,x) \overset{|x-t|\leq h}{=\!=\!=} O(h) \quad ,$$

sowie $$f(t) - f(x) \overset{|x-t|\leq h}{\leq} \|f'\| \cdot h \quad ,$$

$$K(x+h,t) \overset{|x-t|<h}{=\!=\!=} 1 + O(h) \quad ,$$

also:

$$I_1 = f(x)(h + O(h^2)) + O(h^2)\|f'\|$$

$$|I_2| \leq \int_0^a |K(x+h,t) - K(x,t)||f(t)|dt$$

$$\leq \quad a\ h\ C_1\|f\|$$

Damit:

$$h(1+O(h))f(x) + O(h^2)\|f'\| - a\ h\ C_1\|f\| = g(x+h) - g(x)$$

$$\Rightarrow \quad f(x) = \frac{g(x+h)-g(x)}{h(1+O(h))} + O(h)\|f'\| + \frac{a\,C_1}{1+O(h)}\|f\| \quad ,$$

sei nun h so klein, daß die Abschätzung $\frac{1}{1+O(h)} \leq 2$ besteht, dann bekommen wir

$$\|f\| \leq \frac{4\|g\|}{h} + C_2\,h\,\|f'\| + 2a\,C_1\|f\| \quad ,$$

sei nun a so klein, daß gilt $2a\,C_1 \leq \frac{1}{2}$, so folgt

$$\|f\| \leq \frac{8\|g\|}{h} + 2\,C_2\,h\,\|f'\|$$

$$\leq 8\,\frac{\delta}{h} + 2\,C_2\,h\,\rho \qquad \left(\text{setze } h = \left(\frac{\delta}{\rho}\right)^{1/2}\right)$$

$$\leq (8 + 2\,C_2)(\delta\,\rho)^{1/2} \quad .$$

■

BEMERKUNG: Die Abschätzung verschlechtert sich, falls

$$K(x,x) = 0 \quad , \quad K_x(x,x) = 1 \qquad \text{gilt.}$$

$$\int_0^x K(x,t)f(t)dt = g(x) \Rightarrow \int_0^x K_x(x,t)f(t)dt = g'(x)$$

$$\Rightarrow \quad f(x) + \int_0^x K_{xx}(x,t)f(t)dt = g''(x)$$

Satz 3.1

$$\Rightarrow \qquad \|f\| \leq C(a)\|g''\|$$

$$\Rightarrow \quad \alpha(\delta,\rho) = C\,\delta^{1/3}\,\rho^{1/3} \quad .$$

Wir wollen nun eine Stabilitätsaussage für die Abelsche Integralgleichungsaufgabe beweisen.

SATZ 3.3: (Vessello, 1984)

Es sei

$$g(x) = Af(x) = \int_0^x (x-t)^{-\alpha}K(x,t)f(t)dt \quad , \quad 0 \leq \alpha < 1, \; x \in [0,a],$$

mit $K \in C^2 \times C^2$, $K(x,x) = 1 \quad \forall\, x$, und $a < \infty$.

Mit der Definition

$$\beta(\delta,\rho) = \max\{\|f\| \,,\, \|Af\| \leq \delta \,,\, \|f'\| + \|f\| \leq \rho\}$$

gilt:

$$\beta(\delta,\rho) \leq C(\alpha,K,a)\,\delta^{\frac{1}{2-\alpha}}\,\rho^{\frac{1-\alpha}{2-\alpha}} \quad .$$

($\|\cdot\|$ bezeichnet hierbei die Maximumnorm)

BEWEIS:

i) Sei zunächst $K \equiv 1$, so daß wir von der Inversionsformel:

$$f = \frac{\sin \pi\alpha}{\pi} g'_\alpha(x) \qquad \text{mit} \qquad g_\alpha(x) = \int_0^x (x-t)^{\alpha-1} g(t)\,dt$$

Gebrauch machen können.

Sei nun $h > 0$, es gilt:

$$\left| \frac{g_\alpha(x+h)-g_\alpha(x)}{h} \right| \le \frac{1}{h} \int_0^{x+h} (x+h-t)^{\alpha-1} g(t)\,dt - \int_0^x (x-t)^{\alpha-1} g(t)\,dt$$

$$\le \frac{1}{h} \left| \int_x^{x+h} (x+h-t)^{\alpha-1} g(t)\,dt \right| + \frac{1}{h} \left| \int_0^x \left\{ (x+h-t)^{\alpha-1} - (x-t)^{\alpha-1} \right\} g(t)\,dt \right|$$

$$\le \frac{\|g\|}{h} \left\{ \underbrace{\int_x^{x+h} (x+h-t)^{\alpha-1}\,dt}_{= \int_0^h \tau^{\alpha-1}\,d\tau = \frac{1}{\alpha} h^\alpha} + \underbrace{\int_0^x (x-t)^{\alpha-1} - (x+h-t)^{\alpha-1}\,dt}_{= -\frac{1}{\alpha}(x-t)^\alpha \Big|_0^x + (x+h-t)^\alpha \Big|_0^x} \right\}$$

$$\le \frac{\|g\|}{\alpha h} \left\{ h^\alpha + x^\alpha + h^\alpha - (x+h)^\alpha \right\} \le \frac{2}{\alpha} h^{\alpha-1} \|g\|$$

Wir entwickeln g_α in eine Taylorreihe um x

$$g_\alpha(x+h) = g_\alpha(x) + h\, g'_\alpha(x) + \frac{h^2}{2} g''_\alpha(\xi) \qquad , \quad 0 < \xi < x+h$$

$$\Rightarrow \quad g'_\alpha(x) = \frac{g_\alpha(x+h)-g_\alpha(x)}{h} - \frac{h}{2} g''_\alpha(\xi)$$

Sei nun $\|g\| \le \delta$, $\|f'\| \le \rho$, so folgt aus obigem:

$$|g'_\alpha(x)| \le \frac{2}{\alpha} h^{\alpha-1} \delta + \frac{h}{2} \|g''_\alpha\|$$

und damit infolge $f^{(i)} = \frac{\sin \pi\alpha}{\pi} g_\alpha^{(i+1)}(x)$, $i = 0,1$:

$$\|f\| \leq \frac{2}{\alpha} h^{\alpha-1} \frac{\pi}{\sin \pi\alpha} \delta + \frac{h}{2} \rho \quad .$$

Wir balancieren die Terme auf der rechten Seite aus, indem wir $h = (\frac{\delta}{\rho})^{\frac{1}{2-\alpha}}$ setzen und erhalten so

$$\|f\| \leq C(\alpha) \delta^{\frac{1}{2-\alpha}} \rho^{\frac{1-\alpha}{2-\alpha}} ,$$

d.h. die Aussage des Satzes für $K \equiv 1$.

ii) Sei nun K beliebig:

$$\int_0^x (x-t)^{-\alpha} K(x,t) f(t) dt = g(x)$$

$$\Rightarrow \int_0^y (y-x)^{\alpha-1} \int_0^x (x-t)^{-\alpha} K(x,t) f(t) dt dx = \int_0^y (y-x)^{\alpha-1} g(x) dy = g_\alpha(y)$$

$$\Rightarrow \int_0^y f(t) \underbrace{\int_t^y (y-x)^{\alpha-1} (x-t)^{-\alpha} K(x,t) dx}_{=: G(y,t)} dt = g_\alpha(y)$$

Wir fahren wie bei der Herleitung der Inversionsformel fort und substituieren $\xi = \frac{x-t}{y-t}$, was zu

$$G(y,t) = \int_0^1 \xi^{\alpha-1} (1-\xi)^{-\alpha} K(\xi(y-t)+t, t) d\xi$$

führt.

Also gilt

$$\lim_{t \to y} G(y,t) = \int_0^1 \xi^{\alpha-1} (1-\xi)^{-\alpha} d\xi = \frac{\pi}{\sin \pi\alpha} \neq 0 .$$

f ist Lösung der Volterraschen Integralgleichung mit g_α als rechter Seite und G als Kernfunktion, da $G(y,y) \neq 0$, können wir Satz 3.1 anwenden:

$$\|f\| \leq C_1(K,a,\alpha) \|g'_\alpha\|$$

Differenzieren wir die Integralgleichung nach y, so kommen wir zu:

$$\frac{\pi}{\sin \pi\alpha} f(y) + \int_0^y G_y(x,t) f(t)dt = g'_\alpha(y)$$

$$\Rightarrow \|g'_\alpha\| \leq C_2(K,a,\alpha) \|f\| \quad .$$

Wir differenzieren ein weiteres Mal:

$$\frac{\pi}{\sin \pi\alpha} f'(y) + G_y(x,y)f(y) + \int_0^y G_{yy}(x,t)f(t)dt = g''_\alpha(y)$$

$$\Rightarrow \|g''_\alpha\| \leq C_3(K,a,\alpha) \; (\|f'\| + \|f\|)$$

Jetzt können wir von der in i) abgeleiteten Abschätzung für g' Gebrauch machen:

Sei $\|g\| \leq \delta$ und $\|f\| + \|f'\| \leq \rho$, so folgt

$$|g'_\alpha(y)| \leq \frac{2}{\alpha} h^{\alpha-1} \delta + \frac{h}{2} C_3(K,a,\alpha)\rho \; ,$$

also mit $h = (\frac{\delta}{\rho})^{1/2-\alpha}$

$$\|g'_\alpha\| \leq C_4(K,a,\alpha) \; \delta^{\frac{1}{2-\alpha}} \rho^{\frac{1-\alpha}{2-\alpha}}$$

$$\Rightarrow \|f\| \leq C_1(K,a,\alpha) \cdot C_4(K,a,\alpha) \; \delta^{\frac{1}{2-\alpha}} \rho^{\frac{1-\alpha}{2-\alpha}} \quad .$$

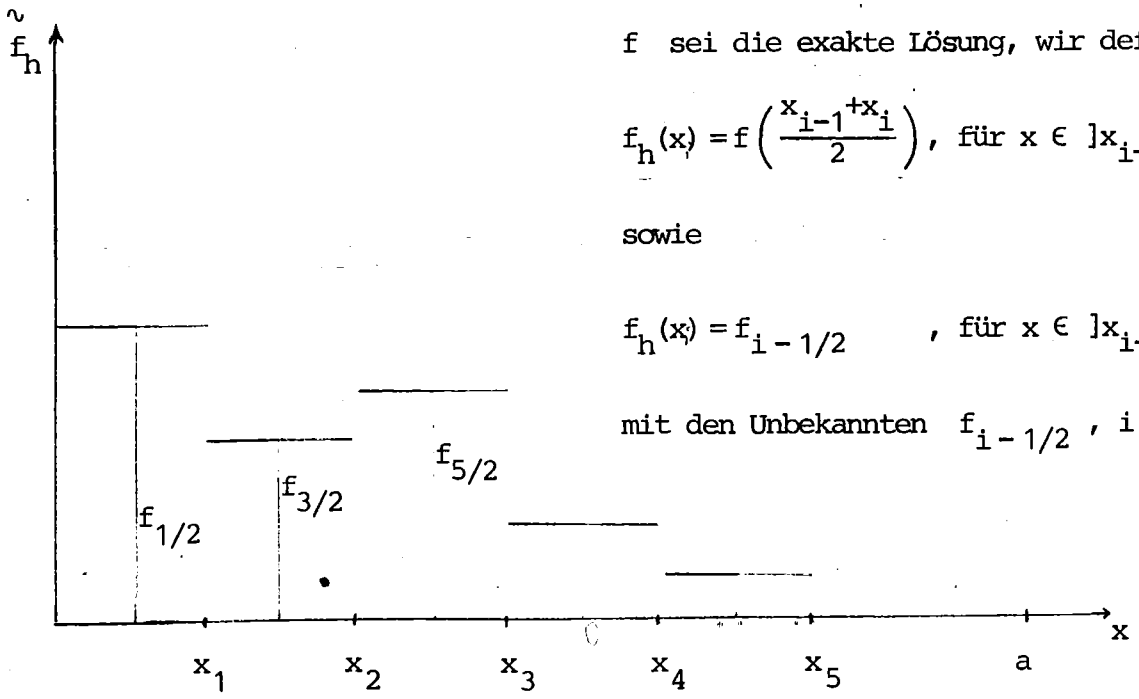
■

Wir wollen eine Stabilitätsaussage für ein numerisches Verfahren zur Behandlung von Volterraschen Integralgleichungen herleiten.

Gegeben sei die Volterra-Gleichung

$$\int_0^x K(x,t)f(t)dt = g(x) \quad ,$$

die wir durch ein Kollokationsverfahren mit Treppenfunktionen numerisch lösen wollen.



f sei die exakte Lösung, wir definieren:

$$f_h(x) = f\left(\frac{x_{i-1}+x_i}{2}\right), \text{ für } x \in]x_{i-1},x_i]$$

sowie

$$f_h(x) = f_{i-1/2} \quad , \text{ für } x \in]x_{i-1},x_i]$$

mit den Unbekannten $f_{i-1/2}$, $i = 1,2,\ldots$

$$x_i = i\,h \quad , \quad i = 0,1,2,\ldots \quad , h > 0$$

Wir lösen dann die diskretisierte Version

$$\int_0^{x_n} K(x_n,t) \cdot \tilde{f}_h(t)dt = g(x_n) =: g_n \quad , n = 1,2,\ldots$$

$$\Rightarrow \quad \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} K(x_n,t)dt \cdot f_{i-1/2} = g_n$$

mit $a_{n_i} = \int_{x_{i-1}}^{x_i} K(x_n,t)dt$ ist also das obere Dreiecks-

system

$$(*) \qquad \sum_{i=1}^{n} a_{n_i} f_{i-1/2} = g_n \qquad , \quad n = 1,2,3,\ldots$$

zu lösen.

Es gilt der

<u>SATZ 3.4:</u> Sei $K \in C^2 \times C^2$, $K(x,x) = 1$, $f \in C^2$.
Dann gibt es ein $h_o > 0$ und ein $C < \infty$, s.d. für alle $h < h_o$
(*) eindeutig lösbar ist und die Abschätzung

$$|f(x_{i-1/2}) - f_{i-1/2}| \leq C\, h^2 \|f''\|$$

besteht.

<u>BEWEIS:</u>

$$\int_0^{x_n} K(x_n,t) f_h(t)dt = \int_0^{x_n} K(x_n,t) f(t)dt + \int_0^{x_n} K(x_n,t)(f_h - f)(t)dt$$

$$\sum_{i=1}^{n} a_{n_i} f_{i-1/2} = g_n$$

$$\Rightarrow \quad \sum_{i=1}^{n} a_{n_i} (f(x_{i-1/2}) - f_{i-1/2}) = \int_0^{x_n} K(x_n,t)(f_h - f)(t)dt =: p_n$$

Nach Aufgabe 33 gilt:

$$\max_i |f(x_{i-1/2}) - f_{i-1/2}| \leq C \max_i \left| \frac{p_i - p_{i-1}}{h} \right| .$$

$$\left| \frac{p_i - p_{i-1}}{h} \right| = \frac{1}{h} \left| \int_0^{x_i} K(x_i,t)(f_h - f)(t)dt - \int_0^{x_{i-1}} K(x_{i-1},t)(f_h - f)(t)dt \right|$$

$$= \frac{1}{h} \left| \int_{x_{i-1}}^{x_i} K(x_i,t)(f_h-f)(t)dt - \int_0^{x_{i-1}} (K(x_i,t) - K(x_{i-1},t))(f_h-f)(t)dt \right|$$

Aus der Fehlerabschätzung für die Mittelpunktregel für zweimal stetig differenzierbare Integranden folgt, daß wir das erste Integral gegen

$$C_1 \| f'' \| h^3$$

und das zweite infolge $K(x_i,t) - K(x_{i-1},t) = O(h)$ gegen

$$\sum_{j=1}^{i} \tilde{C}_j \| f'' \| h^4$$

abschätzen können.

Da $i \leq \frac{a}{h}$ läßt sich diese Summe gegen

$$C_2 \| f'' \| h^3$$

abschätzen, somit

$$\left| \frac{p_i - p_{i-1}}{h} \right| \leq \frac{1}{h} (C_1 + C_2) \| f'' \| h^3$$

$\Rightarrow$ Beh. ∎

Für Abelsche Integralgleichungen mit $\alpha > 0$ verwendet man die sogenannte Produktintegration:

Man macht den Ansatz

$$f_h(x) = \frac{x-x_i}{h} f_{i-1} + \frac{x_{i+1}-x}{h} f_i \quad \text{für} \quad x \in [x_{i-1}, x_i]$$

und löst dann:

$$\int_0^{x_n} f_h(t) K(x_n,t) (x_n-t)^{\alpha} dt = g(x_n) \qquad , \; n = 1,2,\ldots$$

$$\Rightarrow \sum_{i=1}^{n} \left[\int_{x_{i-1}}^{x_i} (x_i-t)^{-\alpha} K(x_i,t) \, \frac{t-x_i}{h} \, f_{i-1} \, dt \right.$$

$$\left. + \int_{x_{i-1}}^{x_i} (x_i-t)^{-\alpha} K(x_i,t) \, \frac{x_{i+1}-t}{h} \, f_i \, dt \right] .$$

4. INTEGRAL - GEOMETRIE

Wir beschäftigen uns in diesem Abschnitt mit folgendem Problem: Gegeben sei

$M(\xi)$ - Mannigfaltigkeit der Dimension $< m$ im $\mathbb{R}^m$; ξ ist hierbei ein Parameter, welcher den Bereich X durchläuft,

und die Datenfunktion

$$g(\xi) = \int_{M(\xi)} f(x)\, dx \quad ,$$

wobei dx das vom gewöhnlichen Lebesgue-Maß auf $M(\xi)$ induzierte Maß sein soll und f eine auf allen $M(\xi)$, $\xi \in X$ integrierbare reellwertige Funktion ist.

Gesucht ist die Funktion f.

In dieser Allgemeinheit läßt sich kaum etwas über das Problem aussagen, wir beschränken uns deshalb auf Spezialfälle.

Sei $m = 2$ und

$$M(\xi) = \{(x_1,x_2) \in \mathbb{R}^2,\ x_1 = \xi_1 \pm u_{\pm}(x_2,\xi_2)\},\ \xi = (\xi_1,\xi_2) \in \mathbb{R}^1 \times [0,a] \ ,$$

wobei die (nicht von ξ_1 abhängigen!) Funktionen $u_{\pm}$ folgende Voraussetzungen erfüllen sollen:

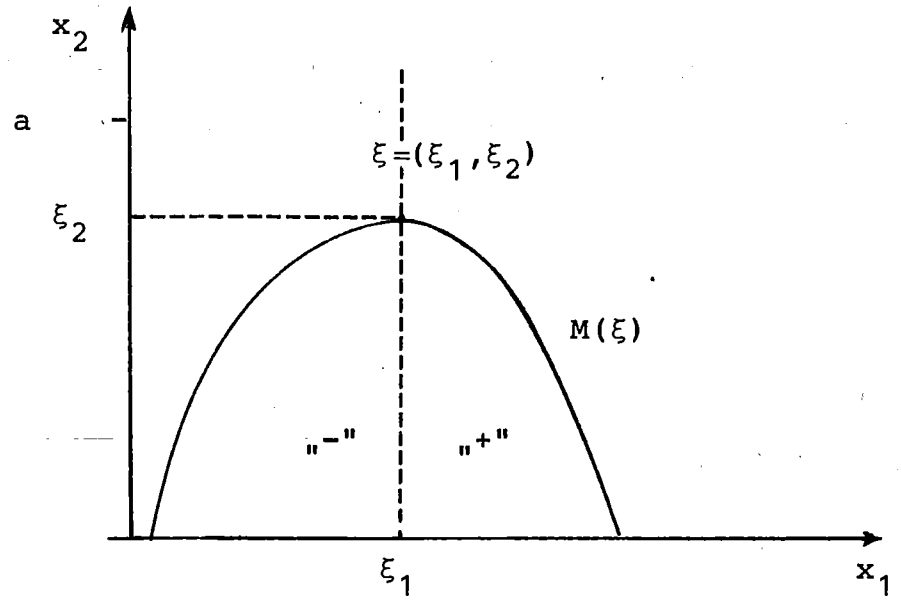
$$\left(\frac{\partial}{\partial x_2} \right)^{\nu} u_{\pm}(x_2,\xi_2) = (\xi_2 - x_2)^{\alpha-\nu} \cdot h_{\pm,\nu}(x_2,\xi_2) \quad \text{für} \quad \nu = 0,1$$

mit

$$\alpha \in]0,1[\qquad \text{und}$$

$$h_{\pm,\nu} \in C^1,\ h_{\pm,\nu}(\xi_2,\xi_2) \neq 0;\ \nu = 0,1,\ \xi_2 \in [0,a] \ .$$

Dazu folgendes Bild:



Aufgrund der angegebenen Voraussetzungen ist

ξ Maximalpunkt der Kurve $M(\xi)$,

$M(\xi)$ invariant gegenüber Verschiebungen in x_1 - Richtung.

Ein Beispiel für eine die Voraussetzungen erfüllende Kurve ist

$$x_1 = \xi_1 \pm \sqrt{\xi_2^2 - x_2^2}$$

der Kreis um $(\xi_1,0)$ mit Radius ξ_2.

Wir beweisen unter den obigen Voraussetzungen den

<u>SATZ 4.1:</u> Sei f stetig mit kompaktem Träger. Dann ist f eindeutig durch

$$g(\xi) \quad , \quad \xi \in \mathbb{R}^1 \times [0,a]$$

bestimmt.

BEWEIS: Das Maß auf der Kurve $M(\xi)$ ist die Bogenlänge, also

$$dx = \left(1 + \left(\frac{dx_1}{dx_2} \right)^2 \right)^{1/2} dx_2 \quad , \text{ somit}$$

$$g(\xi) = \int_{M(\xi)} f(x)\,dx$$

$$= \int_0^{\xi_2} f(\xi_1 - u_-(x_2,\xi_2), x_2) \left(1 + \left(\frac{dx_1}{dx_2} \right)^2 \right)^{1/2} dx_2$$

$$+ \int_0^{\xi_2} f(\xi_1 + u_+(x_2,\xi_2), x_2) \left(1 + \left(\frac{dx_1}{dx_2} \right)^2 \right)^{1/2} dx_2 \quad .$$

Nun ist

$$\frac{dx_1}{dx_2} = \frac{\partial}{\partial x_2} u_\pm(x_2,\xi_2) = (\xi_2 - x_2)^{\alpha-1} h_{\pm,1}(x_2,\xi_2) \quad ,$$

also

$$\left(1 + \left(\frac{dx_1}{dx_2} \right)^2 \right)^{1/2} = (\xi_2 - x_2)^{\alpha-1} \underbrace{\left((\xi_2 - x_2)^{2(1-\alpha)} + h_{\pm,1}^2(x_2,\xi_2) \right)^{1/}}_{=: k_\pm(x_2,\xi_2)}$$

mit $k_\pm \in C^1$ und $k_\pm(\xi_2,\xi_2) > 0$.

$$\Rightarrow g(\xi_1,\xi_2) = \int_0^{\xi_2} (\xi_2 - x_2)^{\alpha-1} \Big\{ f(\xi_1 - u_-(x_2,\xi_2), x_2) k_-(x_2,\xi_2)$$

$$+ f(\xi_1 + u_+(x_2,\xi_2), x_2) k_+(x_2,\xi_2) \Big\} dx_2$$

Wir wollen nun die Fouriertransformation (bzgl. des 1. Arguments) auf diese Gleichung anwenden und erinnern deshalb kurz an deren Definition und wichtigste Eigenschaften.

Sei v eine auf $\mathbb{R}$ definierte integrierbare Funktion, so bezeichnet man mit

$$\hat{v}(\tau) = (2\pi)^{-1/2} \int_{\mathbb{R}} v(t) e^{-it\tau}\, dt$$

deren Fouriertransformierte.

Es gilt

$$v(t) = (2\pi)^{-1/2} \int_{\mathbb{R}} \hat{v}(\tau) e^{it\tau}\, d\tau \qquad \text{(Inversionsformel)}$$

$$\int |v^2(t)|\, dt = \int |\hat{v}(t)|^2\, d\tau \qquad \text{(Plancherel)}$$

ist $v_c(t) = v(t+c) \quad \forall\ t \in \mathbb{R}$, so gilt

$$\widehat{v_c}(\tau) = e^{ic\tau} \hat{v}(\tau) \quad .$$

Mit $\hat{g}(\lambda_1, \xi_2) = (2\pi)^{-1/2} \int_{\mathbb{R}} g(\xi_1, \xi_2) e^{-i\xi_1\lambda_1}\, d\xi_1$ gilt nun nach Vertauschung der Integrationsreihenfolge auf der rechten Seite:

$$\hat{g}(\lambda_1, \xi_2) = \int_0^{\xi_2} (\xi_2 - x_2)^{\alpha-1} \left\{ \widehat{f_{-u_-(x_2,\xi_2)}}(\lambda_1, x_2) \cdot k_-(x_2, \xi_2) + \widehat{f_{u_+(x_2,\xi_2)}}(\lambda_1, x_2) k_+(x_2, \xi_2) \right\} dx_2$$

$$= \int_0^{\xi_2} (\xi_2 - x_2)^{\alpha-1} \hat{f}(\lambda_1, x_2) \underbrace{\left\{ e^{-iu_-(x_2,\xi_2)\lambda_1} k_-(x_2, \xi_2) + e^{iu_+(x_2,\xi_2)\lambda_1} k_+(x_2, \xi_2) \right\}}_{=: K(x_2, \xi_2; \lambda_1)} dx_2$$

Denkt man sich λ_1 als festen Parameter, so ist dies eine Abelsche Integralgleichung für $f(\lambda_1, \cdot)$ mit Kernfunktion $K(\cdot, \cdot; \lambda_1)$, es gilt:

$$u_\pm(\xi_2, \xi_2) = 0 \Rightarrow K(\xi_2, \xi_2; \lambda_1) > 0$$

$K \in C$, K_{ξ_2} ist integrierbar.

Unter diesen (etwas schwächeren) Voraussetzungen bleibt die Aussage von Satz 3.3 gültig, und es gilt somit $\forall\ \lambda_1$

$\hat{f}(\lambda_1, \cdot)$ ist durch $\hat{g}(\lambda_1, \cdot)$ eindeutig bestimmt

und da die Fouriertransformation injektiv ist:

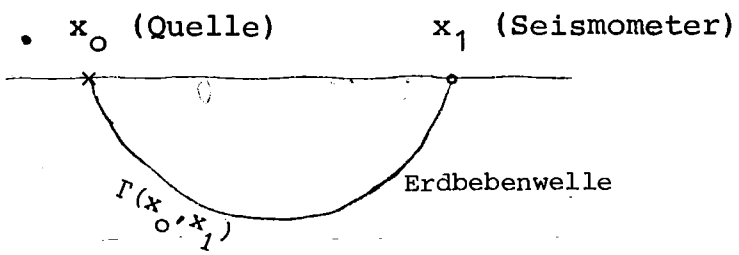
f ist durch g eindeutig bestimmt.

■

Für $\alpha \sim 0$, d.h. $\alpha - 1 \sim -1$ ist die Abelsche Integralgleichung für $\hat{f}(\lambda_1, \cdot)$ „fast gut gestellt"; für $\alpha \sim 1$, d.h. $\alpha - 1 \sim 0$ ist sie sehr schlecht gestellt. Um daraus Aussagen über die Stabilität des Problems: „bestimme f aus g" zu gewinnen, ist jedoch ein genaueres Studium der Abhängigkeit des Kerns $K(\cdot, \cdot, \cdot)$ vom 3. Argument erforderlich.

Beispiele für Anwendungen dieser Theorie:

i) Seismologie



Es wird die Laufzeit(abweichung) der Erdbebenwelle von x_0 nach x_1

$$T(x_0, x_1) = \int_{\Gamma(x_0, x_1)} n \, ds$$

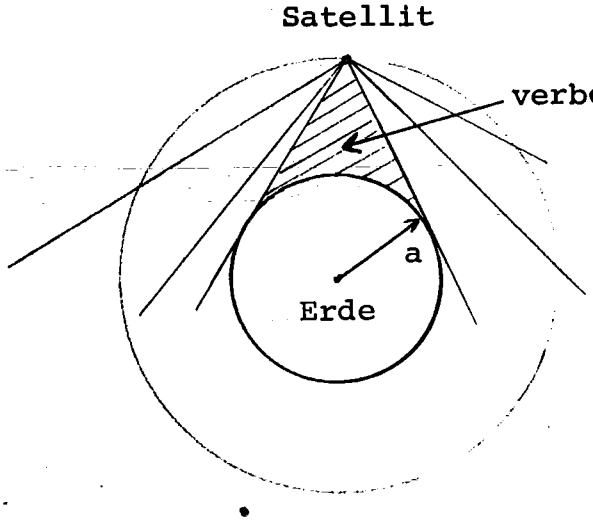
gemessen. Hierbei ist n die Abweichung des Brechungsindex ($\hat{=}$ Kehrwert der lokalen Fortpflanzungsgeschwindigkeit) von einem bekannten Normalwert.

Hat dieser Normalwert eine bestimmte analytische Gestalt, so hat $\Gamma(x_o, x_1)$ die Form eines Halbkreises, d.h. $\alpha = \frac{1}{2}$ in der obigen Theorie.

ii) Radartechnologie

Beim SAR (syntetic aperture radar) mißt man die Integrale der Bodenreflektivität (ground reflectivity) längs eines Halbkreises.

iii) Temperaturprofil der Erdatmosphäre



f ist eine Strahlungsintensität (illuminancy)

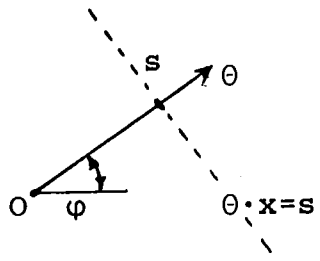
Man mißt:

$g = \int_L f(x)\,dx$, L-Gerade, welche die Erde nicht trifft

Sei Θ ein Einheitsvektor in $\mathbb{R}^2$, so ist

$$\{x \mid x \cdot \Theta = s\}$$

eine Gerade senkrecht zu Θ, welche Θ in s schneidet



Wir definieren

$$g(\Theta,s) = \int_{x\cdot\Theta=s} f(x)\,dx \quad ,$$

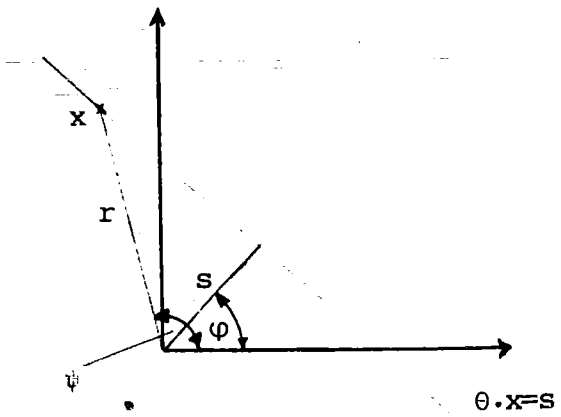
sei nun $\Theta = \begin{pmatrix}\cos\varphi\\ \sin\varphi\end{pmatrix}$; $x = r\,\omega$, $r = |x|$, $\omega = \begin{pmatrix}\cos\psi\\ \sin\psi\end{pmatrix}$,
wir entwickeln g und f in Fourierreihen bzgl. s und φ bzw. r und ψ:

$$g(\Theta,s) = \sum_{\ell\in\mathbb{Z}} g_\ell(s)e^{i\ell\varphi} \quad , \quad f(x) = \sum_{\ell\in\mathbb{Z}} f_\ell(r)e^{i\ell\psi}$$

<u>SATZ 4.2:</u> Sei $f \in C^2$, $f(x) = 0$ für $|x| > R$,
dann ist f in $|x| \geq a$ eindeutig durch $g(\Theta,s)$, $|s| \geq a$ bestimmt.

<u>BEWEIS:</u>

i)



Wir wollen das Integral

$$\int_{\Theta\cdot x=s} f(x)\,dx$$

so darstellen, daß der Winkel ψ zur Integrationsvariablen wird.

$$x = r\cdot\omega,\ x\cdot\Theta = s \Rightarrow r = \frac{s}{\omega\cdot\Theta} \quad , \quad \text{also} \quad x = \frac{s}{\omega\cdot\Theta}\,\omega$$

$$\omega = (\cos\psi\ ,\ \sin\psi)^T \Rightarrow \frac{d\omega}{d\psi} = (-\sin\psi\ ,\ \cos\psi)^T = \omega^\perp \quad (\omega\cdot\omega^\perp = 0)$$

$$\Rightarrow \left|\frac{dx}{d\psi}\right|^2 = \left|\frac{-s\,\omega^\perp\cdot\Theta}{(\omega\cdot\Theta)^2}\,\omega + \frac{s}{\omega\cdot\Theta}\,\omega^\perp\right|^2 = \frac{s^2(\omega^\perp\cdot\Theta)^2}{(\omega\cdot\Theta)^4} + \frac{s^2}{(\omega\cdot\Theta)^2}$$

$$= \frac{s^2}{(\omega\cdot\Theta)^4}\underbrace{\left((\omega^\perp\cdot\Theta)^2 + (\omega\cdot\Theta)^2\right)}_{=\ |\Theta|=1}$$

$$\Rightarrow \left| \frac{dx}{d\psi} \right| = \frac{s}{(\omega\cdot\Theta)^2}$$

$$\Rightarrow \int\limits_{x\cdot\Theta=s} f(x)\,dx = \int\limits_{\Theta\cdot\omega>0} f\left(\frac{s}{\omega\cdot\Theta}\,\omega\right)\frac{s}{(\omega\cdot\Theta)^2}\,d\psi$$

$(\omega\cdot\Theta = \cos\varphi\cos\psi + \sin\varphi\cdot\sin\psi = \cos(\varphi-\psi)\,,\ \Theta\cdot\omega>0 \Leftrightarrow |\varphi-\psi| \leq \frac{\pi}{2}\,)$

ii) Es gilt für beliebige Funktionen h und $\ell \in \mathbb{Z}$

$$\int\limits_0^{2\pi} h(\omega\cdot\Theta)\,e^{i\ell\varphi}\,d\varphi = c_\ell\, e^{i\ell\psi}$$

$$\text{mit}\quad c_\ell = 2\int\limits_{-1}^{1} h(t)\,T_\ell(t)\,(1-t^2)^{1/2}\,dt \quad .$$

Hierbei ist $T_\ell(t) = \cos(\ell \arccos t)$ das Tschebyscheff-polynom 1. Art. (Funk - Hecke - Theorem in der Ebene)

Zum Beweis von ii)

a) sei $\omega = \binom{1}{0}$, d.h. $\psi = 0$

$$\int\limits_0^{2\pi} h(\cos\varphi)\,e^{i\ell\varphi}d\varphi = \int\limits_0^{2\pi} h(\cos\varphi)\,(\cos(\ell\varphi) + \sin(\ell\varphi))\,d\varphi$$

$$= \int\limits_0^{2\pi} h(\cos\varphi)\cos(\ell\varphi)\,d\varphi \qquad (\sin \text{ ist ungerade})$$

$$= 2\int\limits_0^{\pi} h(\cos\varphi)\cos(\ell\varphi)\,d\varphi\,,\ t = \cos\varphi,\ dt = -\sin\varphi\,d\varphi = -(1-t^2)^{\frac{1}{2}}\,d\varphi$$

$$= 2\int\limits_{-1}^{1} h(t)\,T_\ell(t)\,(1-t^2)^{-\frac{1}{2}}\,dt = c_\ell\cdot e^0 = c_\ell$$

b) sei ψ beliebig:

$$\int\limits_0^{2\pi} h(\cos\underbrace{(\varphi-\psi)}_{\varphi'})\,e^{i\ell\varphi}\,d\varphi = \int\limits_0^{2\pi} h(\cos\varphi')\,e^{i\ell(\varphi'+\psi)}\,d\varphi'$$

$$= e^{i\ell\psi} \int_0^{2\pi} h(\cos\varphi)\, e^{i\ell\varphi}\, d\varphi \overset{a)}{=} e^{i\ell\psi}\, c_\ell$$

iii) Wir schreiben nun

$$f(x) = \sum_{\ell\in\mathbb{Z}} f_\ell(r) e^{i\ell\psi}\ ;\ x = r\begin{pmatrix}\cos\psi\\ \sin\psi\end{pmatrix},\quad \psi = \psi(x)$$

$$\text{mit: } f_\ell(r) = \int_0^{2\pi} f\left(r\begin{pmatrix}\cos\psi\\ \sin\psi\end{pmatrix}\right) e^{-i\ell\psi}\, d\psi\ .$$

Da $f \in C^2$, ist $f_\ell = O(\frac{1}{\ell^2})$ und somit die Konvergenz der Fourierreihe absolut und gleichmäßig. Wir werden jetzt eine Beziehung zwischen den Fourierkoeffizienten f_ℓ und g_ℓ, wobei $g(s,\Theta) = \sum_{\ell\in\mathbb{Z}} g_\ell(s) e^{i\ell\varphi}$ gelten soll, herstellen.

Sei $s > 0$ (wegen $g(-s,\omega) = g(s,-\omega)$ ist die Einschränkung unwesentlich):

$$\int_{x\cdot\Theta=s} f_\ell(|x|) e^{i\ell\psi(x)}\, dx \overset{i)}{=} \int_{\omega\cdot\Theta>0} f_\ell\left(\frac{s}{|\omega\cdot\Theta|}\right) \frac{s}{(\omega\cdot\Theta)^2}\, e^{i\ell\psi} d\psi$$

$$= \int_0^{2\pi} h(\omega\cdot\Theta) e^{i\ell\psi}\, d\psi \quad , \text{ mit } \quad h(t) = \begin{cases} 0 & t \le 0 \\ \frac{s}{t^2} f_\ell(\frac{s}{t}) & \text{, sonst} \end{cases}$$

$$\overset{ii)}{=} c_\ell\, e^{i\ell\varphi} \quad , \quad c_\ell = 2\int_0^1 \frac{s}{t^2} f_\ell(\tfrac{s}{t}) T_{|\ell|}(t)(1-t^2)^{-\frac{1}{2}}\, dt$$

Substituiere: $r = \frac{s}{t}$, $dr = -\frac{s}{t^2}\, dt$

$$\Rightarrow\ c_\ell = 2\int_s^\infty f_\ell(r) T_{|\ell|}(\tfrac{s}{r})(1-\tfrac{s^2}{r^2})^{-1/2}\, dr$$

$$\Rightarrow \int_{x\cdot\Theta=s} f_\ell(|x|) e^{i\ell\psi(x)}\, dx = 2\int_s^\infty f_\ell(r)(1-\tfrac{s^2}{r^2})^{-\frac{1}{2}} T_{|\ell|}(\tfrac{s}{r})\, dr\, e^{i\ell\varphi}$$

Insgesamt also:

$$\sum_{\ell\in\mathbb{Z}} g_\ell(s)e^{i\ell\varphi} = g(\Theta,s) = \int_{x\cdot\Theta=s} f(x)\,dx = \sum_{\ell\in\mathbb{Z}} \int_{x\cdot\Theta=s} f_\ell(|x|)e^{i\ell\psi(x)}\,dx$$

$$= \sum_{\ell\in\mathbb{Z}} 2\int_s^\infty f_\ell(r)\left(1-\frac{s^2}{r^2}\right)T_{|\ell|}\left(\frac{s}{r}\right)dr\; e^{i\ell\varphi}$$

$$\Rightarrow \quad g_\ell(s) = 2\int_s^\infty f_\ell(r)\left(1-\frac{s^2}{r^2}\right)T_{|\ell|}\left(\frac{s}{r}\right)dr$$

Diese Integralgleichung für f_ℓ besitzt nach Satz 3.4 eine eindeutige Lösung, d.h. f_ℓ ist $\forall\,\ell$ eindeutig durch g_ℓ und somit auch f eindeutig durch g bestimmt.

■

KOROLLAR 4.3: Zwischen den Fourierkoeffizienten von f und g besteht die Integralgleichung:

$$g_\ell(s) = 2\int_s^\infty f_\ell(r)\left(1-\frac{s^2}{r^2}\right)T_{|\ell|}\left(\frac{s}{r}\right)dr \quad .$$

BEMERKUNG: Falls f rotationssymmetrisch ist, d.h. nur vom Radius und nicht vom Winkel abhängt, gilt $f_\ell(r) \equiv 0$, $\ell \neq 0$. Für $\ell = 0$ läßt sich $f_0(|x|)$ $(= f(x))$ leicht mit Hilfe der Inversionsformel für Abelsche Integralgleichungen aus g_0 berechnen.

Für große ℓ ist $T_\ell|_{[0,1]}$ stark oszillierend.

Wir geben nun an, wie die Integralgleichung in Korollar 4.3 nach f_ℓ aufzulösen ist.

SATZ 4.4: (Cormacksche Inversionsformel, 1964, 1979 Nobelpreis für Medizin)

Es gilt mit den obigen Bezeichnungen:

$$f_\ell(r) = -\frac{1}{\pi} \int_r^\infty (t^2 - r^2)^{-1/2} T_{|\ell|}\left(\frac{t}{r}\right) g'_\ell(t)\,dt$$

BEWEIS: Wir benutzen die Mellin-Transformation in $(0,\infty)$ als Hilfsmittel:

$$Mf(s) := \int_0^\infty f(t) t^{s-1}\,dt \quad , \quad s > 0$$

Eigenschaften der Mellin-Transformation:

i) $$Mf'(s) = \int_0^\infty f'(t) t^{s-1} dt = 0 - \int_0^\infty f(t)(s-1) t^{s-2} dt$$

$$= (1-s) Mf(s-1) \quad , \quad \text{für} \quad s > 1$$

ii) $$M(tf)(s) = \int_0^\infty t f(t) t^{s-1} dt = Mf(s+1) \quad , \quad s > -1$$

iii) $$M(tf')(s) = Mf'(s+1) = -s\,Mf(s) \quad , \quad s > 0$$

Sei $f \overset{\bullet}{*} g\,(u) := \int_0^\infty f(t) g\left(\frac{u}{t}\right) \frac{dt}{t}$, so gilt:

iv) $$M(f * g)(s) = \int_0^\infty \int_0^\infty f(t)\, g\left(\frac{u}{t}\right) \frac{dt}{t}\, u^{s-1} du \quad \text{setze } v = \frac{u}{t},\ du = t\,dv$$

$$= \int_0^\infty \int_0^\infty f(t)\, g(v)\, t^{-1} v^{s-1} t\,dt\,dv$$

$$= Mf(s) \cdot Mg(s) \quad .$$

BEMERKUNG: Bezeichnen wir mit M_+ die multiplikative Gruppe der reellen Zahlen, so läßt sich die Mellin - Transformation als Fouriertransformation auf M_+ auffassen.

SATZ 4.4: (Cormacksche Inversionsformel, 1964, 1979 Nobelpreis für Medizin)

Es gilt mit den obigen Bezeichnungen:

$$f_\ell(r) = -\frac{1}{\pi} \int_r^\infty (t^2 - r^2)^{-1/2} T_{|\ell|}\left(\frac{t}{r}\right) g'_\ell(t)\,dt$$

BEWEIS: Wir benutzen die Mellin-Transformation in $(0,\infty)$ als Hilfsmittel:

$$Mf(s) := \int_0^\infty f(t) t^{s-1}\,dt \quad , \quad s > 0$$

Eigenschaften der Mellin-Transformation:

i) $$Mf'(s) = \int_0^\infty f'(t) t^{s-1} dt = 0 - \int_0^\infty f(t)(s-1) t^{s-2} dt$$

$$= (1-s) Mf(s-1) \quad , \quad \text{für} \quad s > 1$$

ii) $$M(tf)(s) = \int_0^\infty tf(t) t^{s-1} dt = Mf(s+1) \quad , \quad s > -1$$

iii) $$M(tf')(s) = Mf'(s+1) = -s\,Mf(s) \quad , \quad s > 0$$

Sei $f \overset{\bullet}{*} g\,(u) := \int_0^\infty f(t) g\left(\frac{u}{t}\right) \frac{dt}{t}$, so gilt:

iv) $$M(f * g)(s) = \int_0^\infty \int_0^\infty f(t)\, g\left(\frac{u}{t}\right) \frac{dt}{t}\, u^{s-1} du \quad \text{setze } v = \frac{u}{t},\ du = t\,dv$$

$$= \int_0^\infty \int_0^\infty f(t)\, g(v)\, t^{-1} v^{s-1} t\,dt\,dv$$

$$= Mf(s) \cdot Mg(s) \quad .$$

BEMERKUNG: Bezeichnen wir mit M_+ die multiplikative Gruppe der reellen Zahlen, so läßt sich die Mellin - Transformation als Fouriertransformation auf M_+ auffassen.

Das Haarsche Maß auf M_+ ist $d\mu(t) = \frac{dt}{t}$, die Darstellungen von M_+ sind die Abbildungen $s : t \to t^s$, somit: $\hat{f}(s) = \int_0^\infty f(t) t^s \frac{dt}{t}$; iv) entspricht dem Faltungssatz für die Fouriertransformation auf M_+.

Aus dem Buch von Sneddon (The Use of Integral Transforms) entnehmen wir:

f	Mf
$\begin{cases} (1-t^2)^{-\frac{1}{2}} T_\ell(t), & t \le 1 \\ 0, & t > 0 \end{cases}$	$\frac{\pi\, \Gamma(s)\quad 2^{-s}}{\Gamma\left(\frac{\ell+s+1}{2}\right)\Gamma\left(\frac{s+1-\ell}{2}\right)}$, $s > \ell$
$\begin{cases} (1-t)^{-\frac{1}{2}} T_\ell\left(\frac{1}{t}\right), & t \le 1 \\ 0, & t > 1 \end{cases}$	$2^{s-2} \frac{\Gamma\left(\frac{s-\ell}{2}\right)\Gamma\left(\frac{s+\ell}{2}\right)}{\Gamma(s)}$, $s > \ell$

Sei $\ell \in \mathbb{Z}$ beliebig, so können wir die Integralgleichung

$$g_\ell(s) = 2 \int_s^\infty r\, f_\ell(r) T_\ell\left(\frac{s}{r}\right)\left(1 - \frac{s^2}{r^2}\right)^{-1/2} \frac{dr}{r}$$

schreiben als

$$g_\ell(s) = (r \cdot f_\ell * b)(s) \text{ , mit } b(x) = \begin{cases} 2T_{|\ell|}(x)(1-x^2)^{-\frac{1}{2}}, & x < 1 \\ 0, & x \ge 1 \end{cases}$$

$$\overset{\text{iv)}}{\Rightarrow} \quad Mg_\ell = M(r f_\ell) \cdot Mb \quad ,$$

$$\text{mit} \quad Mb(s) = \frac{2\pi\, \Gamma(s)\, 2^{-s}}{\Gamma\left(\frac{|\ell|+1+s}{2}\right)\Gamma\left(\frac{s+1-|\ell|}{2}\right)}$$

$$\Rightarrow \quad \frac{1}{Mb(s-1)} = \frac{\Gamma\left(\frac{s+|\ell|}{2}\right)\ \Gamma\left(\frac{s-|\ell|}{2}\right)}{2\pi\ \Gamma(s-1)2^{-s+1}} = \frac{1}{\pi}\ \frac{\Gamma(s)}{\Gamma(s-1)}\ \frac{\Gamma\left(\frac{s+|\ell|}{2}\right)\Gamma\left(\frac{s-|\ell|}{2}\right)}{\Gamma(s)}\ 2^{s-2}$$

$$= \frac{s-1}{\pi}\ Ma(s) \quad , \quad a(t) = \begin{cases} (1-t)^{-1/2}\ T_{|\ell|}\left(\frac{1}{t}\right) & , t \leq 1 \\ 0 & , t > 1 \end{cases}$$

$$\Rightarrow \quad M(r\, f_\ell)\,(s-1) = \frac{s-1}{\pi}\ Mg_\ell(s-1)Ma(s)$$

$$\overset{iii)}{=} -\frac{1}{\pi}\ M(tg'_\ell)(s-1)\cdot Ma(s)$$

$$\overset{ii)}{\Rightarrow} \quad M(f_\ell)(s) = -\frac{1}{\pi}\ M(g'_\ell)(s)Ma(s) \overset{iv)}{=} -\frac{1}{\pi}\ M(g'_\ell * a)(s)$$

$$\Rightarrow \quad f_\ell(s) = -\frac{1}{\pi}\ g'_\ell * a(r)$$

$$= -\frac{1}{\pi}\ g'_\ell(t)\,a\left(\frac{r}{t}\right)\frac{dt}{t}$$

$$= -\frac{1}{\pi}\ g'_\ell(t)\left(1-\frac{r^2}{t^2}\right)^{-1/2} T_{|\ell|}\left(\frac{t}{r}\right)\ \frac{dt}{t}$$

■

Im Prinzip haben wir mit Satz 4.4 die Lösung unseres Problems erreicht: um f aus den Daten $g(\theta,s)$, $|s| \geq a$, zu berechnen, differenzieren wir diese nach dem 2. Argument und entwickeln die Ableitung in eine Fourierreihe. Deren Koeffizienten werden gegen die o.a. Kernfunktion integriert, um die Fourierkoeffizienten $f_\ell(r)$, für $|r| \geq a$ zu erhalten.

Die beiden folgenden Lemmata werden jedoch zeigen, daß die Cormacksche Formel für das praktische Rechnen wertlos ist: wir müssen die betragsmäßig sehr große Funktion $T_{|\ell|}\left(\frac{t}{r}\right)$ numerisch gegen eine stark oszillierende Datenfunktion $g'_\ell(t)$ integrieren.

LEMMA 4.5: Für $x \geq 1$ gilt die Abschätzung

$$T_m(x) \geq \frac{1}{2}\left(x + \sqrt{x^2-1}\right)^m \quad .$$

BEWEIS: Für $|x| < 1$ gilt: $T_m(x) = \cos(m \arccos x)$, auf der linken Seite steht jedoch ein Polynom vom Grad m in x, welches $\forall\, x \in \mathbb{R}$ definiert ist. Wir versuchen daher, die rechte Seite holomorph für $x \geq 1$ fortzusetzen:

$\cosh(z) = \cos(iz) \Rightarrow -i \arccos(\cosh(z)) = z$

$\Rightarrow i \operatorname{arcosh} z = \arccos z \Rightarrow \cos(m \arccos(z)) = \cos(i m \operatorname{arcosh}(z))$

$= \cosh(m \operatorname{arcosh}(z))$, also für $x > 1$

$$T_m(x) = \cosh(m \operatorname{arcosh}(x))$$

(wegen $\cos(m \arccos(1)) = \cosh(m \operatorname{arcosh}(1)) = 1$ haben wir den richtigen Zweig des komplexen arcosh gewählt).

Sei $t = \cosh x = \frac{1}{2}(e^x + e^{-x})$, so folgt

$$t + (t^2-1)^{1/2} = \frac{1}{2}(e^x + e^{-x}) + \left(\frac{1}{4}(e^{2x} + 2 + e^{-2x}) - 1\right)^{1/2}$$

$$= \frac{1}{2}(e^x + e^{-x}) + \frac{1}{2}(e^x - e^{-x}) = e^x$$

$$\Rightarrow \quad \operatorname{arcosh}(t) = \ln\left(t + (t^2-1)^{1/2}\right) \quad , t \geq 1$$

$$\Rightarrow \quad T_m(x) = \frac{1}{2}(e^{mt} + e^{-mt}) \quad , t = \ln\left(x + (x^2-1)^{\frac{1}{2}}\right), x \geq 1$$

$$\geq \frac{1}{2}\exp\left(m \ln\left(x + (x^2-1)^{\frac{1}{2}}\right)\right)$$

$$= \frac{1}{2}\left(x + (x^2-1)^{\frac{1}{2}}\right)^m$$

■

LEMMA 4.6: Für $m = |\ell| - 1, |\ell| - 3, \ldots$ gilt:

$$\int_0^\infty g_\ell'(s) s^m \, ds = 0 \quad .$$

BEMERKUNG: Hieraus folgt, daß g_ℓ' für große $|\ell|$ stark oszillierend sein muß.

BEWEIS:

i) $$\int_{-\infty}^{\infty} g(\theta,s) s^m ds = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\underbrace{s\theta + t\theta}_{=x}) \, dt \, \underbrace{s^m}_{=x\cdot\theta} \, ds$$

$$= \int_{\mathbb{R}^2} f(x) (x \cdot \theta)^m \, dx$$

$$=: P_m(\varphi)$$

Aus der Definition folgt, daß $P_m(\varphi)$ ein Polynom vom Grad m in $\theta = \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix}$ ist. Also hat P_m die Darstellung:

$$P_m(\varphi) = \sum_{|\ell| \le m} c_\ell \, e^{i\ell\varphi}$$

ii) $$\int_{\mathbb{R}} g_\ell'(s) s^m ds = -m \int_{\mathbb{R}} g_\ell(s) s^{m-1} ds$$ (die Randterme verschwinden, da g kompakten Träger hat)

$$= -\frac{m}{2\pi} \int_{\mathbb{R}} \int_0^{2\pi} g(\theta,s) e^{i\ell\varphi} \, d\varphi \, s^{m-1} \, ds$$

$$= -\frac{m}{2\pi} \int_0^{2\pi} e^{i\ell\varphi} \int_{\mathbb{R}} g(\theta,s) s^{m-1} \, ds \, d\varphi$$

$$= -\frac{m}{2\pi} \int_0^{2\pi} e^{i\ell\varphi} \cdot P_{m-1}(\varphi) \, d\varphi$$

$$\overset{i)}{=} 0 \quad , \quad \text{falls} \quad |\ell| \ge m$$

iii) $g(\theta,s) = g(-\theta,-s) \Rightarrow g'(\theta,s) = -g'(-\theta,-s)$

$$\Rightarrow \sum_{\ell\in\mathbb{Z}} g'_\ell(s)e^{i\ell\varphi} = -\sum_{\ell\in\mathbb{Z}} g'_\ell(-s)e^{i\ell(\varphi+\pi)}$$

$$= -\sum_{\ell\in\mathbb{Z}} g'_\ell(-s)(-1)^{|\ell|}\, e^{i\ell\varphi}$$

$$\Rightarrow g'_\ell(s) = (-1)^{|\ell|+1}\, g'_\ell(-s)$$

iv) $\int_{\mathbb{R}} g'_\ell(s)s^m\,ds = 0$ falls $m \le |\ell|$ nach ii)

Sei nun $m < |\ell|$ und $m + |\ell|$ ungerade, d.h. $m = |\ell| - 1, |\ell| - 3, \ldots,$ dann ist $g'_\ell(s)s^m$ eine gerade Funktion, denn

$$g'_\ell(-s)(-s)^m \overset{\text{iii)}}{=} (-1)^{m+|\ell|+1}\, g_\ell(s)s^m = g_\ell(s)s^m \quad .$$

Also gilt:

$$0 = \int_{\mathbb{R}} g'_\ell(s)s^m ds = \frac{1}{2}\int_0^\infty g'_\ell(s)s^m ds \quad .$$

■

Wir geben nun eine stabil auswertbare Inversionsformel für die Fourierkoeffizienten $f_\ell(r)$ an, die dafür jedoch den Nachteil besitzt, daß für ihre Auswertung für $|r| \ge a$ die Funktionen $g'_\ell(s)$ auf ganz $\mathbb{R}^+$ (also nicht nur auf $[a,\infty]$) bekannt sein müssen.

SATZ 4.7: Es gilt:

$$f_\ell(r) = -\frac{1}{\pi r}\left\{\int_r^\infty \left(\frac{s^2}{r^2}-1\right)^{-\frac{1}{2}}\left(\frac{s}{r}+\sqrt{\frac{s^2}{r^2}-1}\right)^{-|\ell|} g'_\ell(s)\,ds \right.$$

$$\left. - \int_0^r U_{|\ell|-1}\left(\frac{s}{r}\right) g'_\ell(s)\,ds \right\} .$$

($U_m(x) = \frac{\sin(m+1)t}{\sin t}$, $t = \arccos x$, $|x| \leq 1$ ist das Tschebyscheff-Polynom 2. Art.)

BEWEIS: Sei $Q_{|\ell|}$ irgendein Polynom vom Grad $< |\ell|$ mit der Partät $|\ell| - 1$, d.h.

$$Q_{|\ell|}(-s) = (-1)^{|\ell|-1} Q_{|\ell|}(s) \quad \forall\, s \quad ,$$

dann gilt nach Satz 4.4 u. Lemma 4.5

$$f_\ell(r) = -\frac{1}{\pi r}\int_r^\infty \left(\frac{s^2}{r^2}-1\right)^{-1/2} T_{|\ell|}\left(\frac{s}{r}\right) g'_\ell(s)\,ds$$

$$+ \frac{1}{\pi r}\underbrace{\int_0^\infty Q_{|\ell|}\left(\frac{s}{r}\right) g'_\ell(s)\,ds}_{= 0 \text{ (Lemma 4.6)}}$$

$$= -\frac{1}{\pi r}\int_r^\infty g'_\ell(s)\left[\left(\frac{s^2}{r^2}-1\right)^{-\frac{1}{2}} T_{|\ell|}\left(\frac{s}{r}\right) - Q_{|\ell|}\left(\frac{s}{r}\right)\right] ds$$

$$+ \frac{1}{\pi r}\int_0^r Q_{|\ell|}\left(\frac{s}{r}\right) g'_\ell(s)\,ds$$

Das Polynom $Q_{|\ell|}$ wird nun so gewählt, daß der Ausdruck $(x^2-1)^{-1/2} T_{|\ell|}(x) - Q_{|\ell|}(x)$ für $|x| > 1$ nicht zu groß wird. $(x = \frac{s}{r})$

$$T_{|\ell|}(x) = \cosh(|\ell|\, t) \quad , \; x = \cosh t$$

$$\Rightarrow \quad (x^2-1)^{-1/2} T_{|\ell|}(x) = \left(\cosh^2(t)-1\right)^{-\frac{1}{2}} \cosh(|\ell| t) = \frac{\cosh(|\ell|\, t)}{\sinh(t)}$$

Es läßt sich nun wie in Lemma 4.5 zeigen, daß

$\frac{\sinh(mt)}{\sinh(t)}$, $t = \operatorname{arcosh}(x)$, $x > 1$ die Fortsetzung von

$U_{m-1}(x) = \frac{\sin(m\,t)}{\sin(t)}$, $t = \arccos(x)$, $|x| < 1$ ist.

Man überzeugt sich leicht, daß U_{m-1} die gewünschte Parität hat, ferner gilt $\lim_{x\to\infty} (\cosh(x) - \sinh(x)) = \lim_{x\to\infty} e^{-x} = 0$, so daß mit $Q_{|\ell|}(x) = U_{|\ell|-1}(x)$ der o.a. Ausdruck für große x klein wird.

$$(x^2-1)^{-\frac{1}{2}} T_\ell(x) - \frac{\sinh(|\ell|t)}{\sinh(t)} = \frac{\cosh(|\ell|t) - \sinh(|\ell|t)}{\sinh(t)} \quad , x = \cosh t$$

$$= \frac{e^{-|\ell|t}}{(\cosh^2(t)-1)^{1/2}} \; ; \quad t = \operatorname{arcosh} x = \ell n(x + (x^2-1)^{1/2})$$

$$= (x^2 - 1)^{-1/2} (x + (x^2 - 1)^{1/2})^{-|\ell|} \qquad \Rightarrow \quad \text{Beh.}$$

■

§ 5 LAPLACE - TRANSFORMATION

Es sei f eine auf $(0,\infty)$ definierte Funktion, mit

$$\mathcal{L}f(s) = \int_0^\infty e^{-st} f(t)dt \qquad , s \in \mathbb{C}$$

bezeichnen wir die Laplace-Transformierte von f.

BEMERKUNG: Existiert das Integral für ein $s_o \in \mathbb{R}$, so existiert es $\forall\, s \in \mathbb{C}$ mit $\mathrm{Re}(s) > s_o$.

Die Laplace-Transformation besitzt folgende Eigenschaften:

i) $$\mathcal{L}f'(s) = s\,\mathcal{L}f(s) - f(0)$$

Daraus folgt sofort durch Induktion

$$\mathcal{L}f^{(m)}(s) = s^m \mathcal{L}f(s) - s^{m-1}f(0) - s^{m-2}f'(0) - \ldots - f^{(m-1)}(0) \,.$$

ii) Wird

$$f * g(t) = \int_0^t f(u)g(t-u)du$$

definiert, so gilt:

$$\mathcal{L}(f * g) = \mathcal{L}f \cdot \mathcal{L}g \,.$$

Die Beweise hierfür sind denkbar einfach:

zu i):

$$\mathcal{L}f'(s) = \int_0^\infty e^{-st} f'(t)dt = e^{-st}f(t)\Big|_{t=0}^{t=\infty} + s\int_0^\infty e^{-st}f(t)dt$$

$$= f(0) + s\mathcal{L}f(s) \,, \quad (\lim e^{-st} f(t) = 0, \text{ falls } \mathcal{L}f \text{ existiert}).$$

zu ii):

$$\mathcal{L}(f*g)(s) = \int_0^\infty e^{-st} \int_0^\infty f(u)g(t-u)\,du\,dt$$

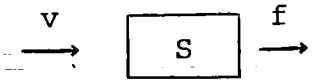
$$= \int_0^\infty f(u) \int_u^\infty e^{-st} g(t-u)\,dt\,du \quad (\text{setze } v = t-u)$$

$$= \int_0^\infty f(u)\, e^{-su} \int_0^\infty e^{-sv} g(v)\,dv$$

$$= \mathcal{L}f(s)\,\mathcal{L}g(s) \quad .$$

Anwendungen:

a) Systemtheorie



Ein lineares System S antworte auf ein bekanntes Eingangssignal v mit dem Ausgangssignal f. Dann läßt sich der Zusammenhang zwischen v und f oft durch eine lineare Differentialgleichung mit konstanten Koeffizienten beschreiben:

$$v = a_m f^{(m)} + a_{m-1} f^{(m-1)} + \ldots + a_o f \quad ,$$

$$f^{(i)}(0) = 0 \quad , \quad i = 0,\ldots,m-1$$

Wenden wir darauf die Laplace-Transformation an, so ergibt sich:

$$\mathcal{L}v(s) = (a_m s^m + a_{m-1} s^{m-1} + \ldots + a_o)\mathcal{L}f(s)$$

$$\Rightarrow \mathcal{L}f = \frac{1}{a_m s^m + \ldots + a_o}\,\mathcal{L}v$$

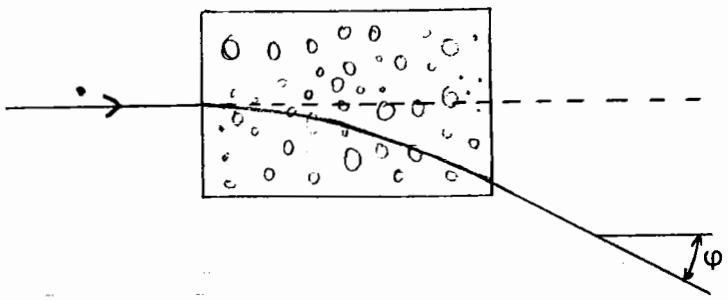
D.h. die Laplace-Transformierte von f ergibt sich einfach durch Polynomdivision aus derjenigen von v.

Für viele Funktionen sind die Laplace-Transformierten berechnet und in Tabellenwerken herausgegeben worden (s. G. Doetsch: Handbuch der Laplace-Transformation).

b) Physik:

Wir haben eine Flüssigkeit, in der kugelförmige Teilchen verschiedener Größen gelöst seien, $f(r)$ sei die Verteilungsdichtefunktion für die Radien. Schicken wir nun einen Laserstrahl durch die Flüssigkeit, so wird dieser mit dem Winkel φ abgelenkt. Aus physikalischen Überlegungen ergibt sich:

$$\varphi = \mathcal{L}f$$



Wir geben nun die Inversionsformel für die Laplace-Transformation an.

SATZ 5.1: Für ein $s_o \in \mathbb{R}$ gelte $\int_0^\infty e^{-s_o t} |f(t)| dt < \infty$.

Dann gilt $\forall\, s \in \mathbb{R},\ s > s_o$

$$f(t) = \frac{1}{2\pi i} \int_{s-i\infty}^{s+i\infty} e^{yt} \mathcal{L}f(y)\, dy \quad .$$

BEWEIS: Der Integrationsweg ist die um s ($> s_o$) verschobene imaginäre Achse. Für $Re(s) > s_o$ ist $\mathcal{L}f(s) = \int_0^\infty e^{-st} f(t)dt$ wohldefiniert und, wie Differentiation unter dem Integral zeigt, sogar holomorph, insbesondere ist $\mathcal{L}f$ auf dem angegebenen Integrationsweg sinnvoll erklärt.

Als Hilfsmittel benötigen wir die Inversionsformel für die Fouriertransformation:

$$\hat{f}(\tau) = (2\pi)^{-1/2} \int_{\mathbb{R}} e^{-i\tau t} f(t)dt$$

$$f(t) = (2\pi)^{-1/2} \int_{\mathbb{R}} e^{i\tau t} \hat{f}(\tau)d\tau \quad .$$

Sei

$$f_s(t) := \begin{cases} e^{-st} f(t) & , \ t \geq 0 \\ 0 & , \ \text{sonst} \end{cases} ,$$

dann gilt nach Definition:

$$\widehat{f_s}(\tau) = (2\pi)^{-1/2} \int_0^\infty e^{-i\tau t} e^{-st} f(t)dt$$

$$= (2\pi)^{-1/2} \mathcal{L}f(s + i\tau)$$

$$\Rightarrow \quad f_s(t) = (2\pi)^{-1/2} \int_{\mathbb{R}} e^{i\tau t} (2\pi)^{-1/2} \mathcal{L}f\underbrace{(s + i\tau)}_{=y;\ i\tau = y-s} d\tau$$

$$= \frac{1}{2\pi i} \int_{s-i\infty}^{s+i\infty} e^{t(y-s)} \mathcal{L}f(y)dy$$

Also

$$e^{-st} f(t) = e^{-st} \frac{1}{2\pi i} \int_{s-i\infty}^{s+i\infty} e^{ty} \mathcal{L}f(y)\,dy$$

$\Rightarrow$ Beh. ∎

In Anwendungen ist $\mathcal{L}f$ in der Regel nur auf $\mathbb{R}$ (oft nur für wenige reelle Argumente) gegeben, um die Inversionsformel anzuwenden, müssen wir folglich $\mathcal{L}f$ analytisch auf $\mathbb{C}$ fortsetzen.

Wir behandeln (dies ist keine wesentliche Einschränkung) den Fall $s_o = 0$, d.h. $\int_0^\infty |f(t)|\,dt < \infty$,

gegeben $\mathcal{L}f_{|[0,\infty[}$

gesucht $\mathcal{L}f(i\,w)$, $w \in \mathbb{R}$.

Wir führen die Pollaczek - Polynome P_n ein.
Diese werden definiert durch die Bedingung:

$$\int_{\mathbb{R}} w(y) P_n(y) P_m(y)\,dy = \delta_{n\,m} \quad , \forall\ n,m \in \mathbb{N}$$

mit $w(y) = |\Gamma(\frac{1}{2} - i\,y)|^2$ und $P_j \in \mathcal{P}_j$ $\forall\ j$.

D.h. die Pollaczek - Polynome sind die Orthogonalpolynome bzgl. der Gewichtsfunktion w und des Intervalls $\mathbb{R}$. Die Pollaczek - Funktionen ψ_n sind definiert durch

$$\psi_n(y) = \frac{1}{\sqrt{\pi}} \Gamma(\frac{1}{2} - i\,y) P_n(y) \quad ,$$

sie sind nach Definition analytisch für $Im(y) < \frac{1}{2}$ und bilden ein vollständiges ONS in $L_2(\mathbb{R})$. Folglich können wir $\mathcal{L}f$

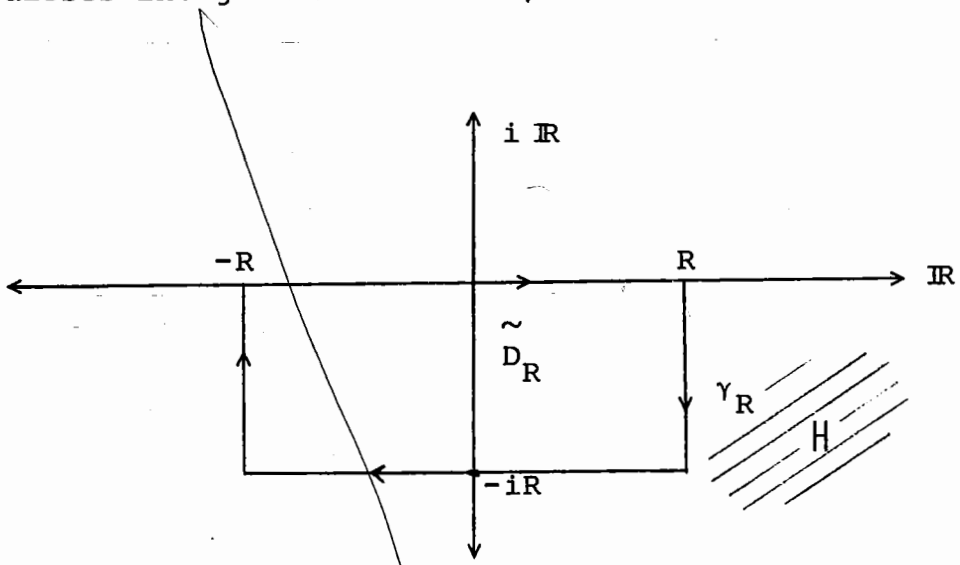
nach den ψ_n entwickeln

$$\mathcal{L}f(iy) = \sum_{m=0}^{\infty} c_n \psi_n$$

mit $$c_n = \int_{\mathbb{R}} \mathcal{L}f(iy)\,\overline{\psi_n(y)}\,dy$$

$$= \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} \mathcal{L}f(iy)\,\overline{\Gamma(\tfrac{1}{2} - iy)}\,P_n(y)\,dy$$

Um dieses Integral auszuwerten, wenden wir den Residuensatz an:



$$\int_{-R}^{R} F(y)\,dy + \int_{\gamma_R} F(z)\,dz = 2\pi i \sum_{z \in \tilde{D}_R} \operatorname{res}_z F(z)$$

mit $$F(y) = \mathcal{L}f(iy)\,\overline{\Gamma(\tfrac{1}{2} - iy)}\,P_n(y)$$

Wir wollen nun natürlich den Grenzübergang $R \to \infty$ bilden und zeigen, daß $\lim\limits_{R\to\infty} \int_{\gamma R} F(z)\,dz = 0$ gilt. Dies vorausgesetzt, folgt dann:

$$c_n = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} F(y)\,dy = 2\sqrt{\pi}\,i \sum_{z \in H} \operatorname{res}_z F(z) \quad .$$

Wir müssen also die Residuen von $\mathcal{L}f(i\,y)\,\overline{\Gamma(\frac{1}{2} - i\,y)}P_n(y)$ in der unteren Halbebene berechnen:

$\mathcal{L}f$ ist holomorph in der rechten Halbebene ($Re(z) > 0$), also ist $\mathcal{L}f(i\,y)$ holomorph in der unteren Halbebene; P_n ist als Polynom eine ganze Funktion, also bleiben nur die Pole der Funktion $\Gamma(\frac{1}{2} - i\,y)$ in der unteren Halbebene.

Wendet man die Funktionalgleichung der Γ-Funktion $\Gamma(z+1) = z\Gamma(z)$ auf $\Gamma(z+n)$ n-mal an, folgt $\Gamma(z+n) = (z+n-1)\cdot\ldots\cdot(z-1)z\Gamma(z)$,

$$\Rightarrow \quad \Gamma(z) = \prod_{j=0}^{m} \frac{1}{z+j}\,\Gamma(z+m+1) \qquad , \quad m = 0,1,2,\ldots$$

$$\Rightarrow \quad \lim_{z\to -m}(z+m)\Gamma(z) = \prod_{j=0}^{m-1} \frac{1}{j-m}\,\Gamma(1) = \frac{(-1)^m}{m!}$$

Die Γ-Funktion hat also an den Stellen $-m$, $m \in \mathbb{N}$ einfache Pole mit dem Residuum $\frac{(-1)^m}{m!}$ und dies sind die einzigen Singularitäten von Γ.

$\Rightarrow$ $\Gamma(\frac{1}{2} - i\,y)$ hat Pole 1. Ordnung mit Residuum $\frac{(-1)^m}{m!}$ in den Punkten $y = -i(m + \frac{1}{2})$, $m = 0,1,2,\ldots$ (also $y \in H$)

$$\Rightarrow \quad c_n = 2\sqrt{\pi}\,i \sum_{m=0}^{\infty} \mathcal{L}f(m+\tfrac{1}{2})P_n(-i\,(m+\tfrac{1}{2}))\,\frac{(-1)^m}{m!} \quad .$$

Da $P_n(-i(m + \frac{1}{2})) \sim m^n$ gilt, ist diese Formel für große n unstabil.

Um $\lim\limits_{R\to\infty} \int\limits_{\gamma R} F(z)dz = 0$ zu zeigen, muß das Wachstumsverhalten der Funktion F genauer untersucht werden. Sei $R = K \in \mathbb{N}$, dann gilt auf den senkrechten Integrationswegen

(*) $\qquad z = \pm K - i t \qquad t \in [0,K]$

und auf dem waagerechten

(**) $\qquad z = - i K + t \qquad t \in [-K,K]$.

Für (*) gilt $\frac{1}{2} - i z = \frac{1}{2} - t \pm i K$ und

für (**) : $\frac{1}{2} - i z = \frac{1}{2} - K - i t$.

Für diese Werte müssen wir die Γ-Funktion abschätzen:

Wegen

$$|\Gamma(x+iy)| \sim (2\pi)^{\frac{1}{2}} e^{-\frac{1}{2}\pi|y|} |y|^{\frac{1}{2}-x} , \quad |y| >> 1$$

gilt für (*):

$$|\Gamma(\tfrac{1}{2} - t \pm i K)| \sim (2\pi)^{\frac{1}{2}} e^{-\frac{1}{2}\pi K} |K|^{-t}$$

$$\leq c\, e^{-\frac{1}{2}\pi K}$$ (s. Abramowitz/Stegun S. 257)

Um $|\Gamma(\frac{1}{2} - K - i t)|$ abzuschätzen, benutzen wir die Funktionalgleichung

$$|\Gamma(\tfrac{1}{2} - K - i t) = \left| \prod_{j=0}^{K} \frac{1}{\frac{1}{2} - K + j + i t} \right| |\Gamma(\tfrac{1}{2} + i t)|$$

$$= \prod_{j=0}^{K} \frac{1}{((\frac{1}{2} - K + j)^2 + t^2)^{1/2}} |\Gamma(\tfrac{1}{2} + i t)|$$

$$\leq 2 \prod_{j=0}^{k-1} \frac{2}{2(K-j)-1} |\Gamma(\tfrac{1}{2} + i t)| = 2^{K+1} \prod_{\ell=1}^{K-1} \frac{1}{2\ell+1} |\Gamma(\tfrac{1}{2} + i t)|$$

$$= 2^{K+1}\, 2^{K-1}\, \frac{\Gamma(K)}{\Gamma(2K)} |\Gamma(\tfrac{1}{2} + i t)|$$

$$= 2^{2K}\, \frac{\Gamma(K)}{\Gamma(2K)} \left(\frac{\pi}{\cosh \pi t} \right)^{1/2}$$ (s. Aramowitz/Stegun S. 256)

$$\sim 2^{2K} \frac{e^{-K} K^{K-1/2}}{e^{-2K} (2K)^{2K-1/2}} \left(\frac{\pi}{\cosh \pi t} \right)^{1/2} \quad \text{(Stirlingsche Formel)}$$

$$= \sqrt{2} \frac{e^K}{K^K} \left(\frac{\pi}{\cosh \pi t} \right) .$$

Insgesamt also $|\Gamma(\frac{1}{2} - i y)| = O(e^{-\alpha K})$, $\alpha > 0$, da P_n nur polynomial wächst und $|\mathcal{L}f(i y)|$ sich auf der unteren Halbebene durch $|\mathcal{L}f(0)|$ abschätzen läßt, folgt:

$$\lim_{K \to \infty} \int_{\gamma_K} F(z)dz = 0 .$$

Wir hatten gesehen, daß die analytische Fortsetzung uns nicht viel hilft, da die Formel für die Entwicklungskoeffizienten c_n unstabil ist.

Wir wollen daher versuchen, die Invertierung von $\mathcal{L}$ mit Hilfe einer Singulärwertzerlegung zu berechnen.

Wir nehmen im folgenden an, daß uns die a-priori Information

$$\text{supp } (f) \subset [1,\gamma]$$

gegeben ist. Also gilt:

$$\mathcal{L}f(s) = \int_1^\gamma e^{-st} f(t)dt ,$$

wir fassen $\mathcal{L}$ daher als Operator von $L_2(1,\gamma)$ in $L_2(0,\infty)$ auf. Also gilt für $\mathcal{L}^* : L_2(0,\infty) \to L_2(1,\gamma)$

$$\mathcal{L}^* g(t) = \int_0^\infty e^{-st} g(s)ds \qquad (t \in [1,\gamma])$$

Und folglich für $\mathcal{L}^*\mathcal{L}$: $L_2(1,\infty) \to L_2(1,\gamma)$

$$\mathcal{L}^*\mathcal{L}f(t) = \int_0^\infty e^{-st} \int_1^\gamma e^{-su} f(u)\,du\,ds$$

$$= \int_1^\gamma f(u) \int_0^\infty e^{-s(t+u)}\,ds\,du$$

$$= \int_1^\gamma f(u)\ \frac{1}{t+u}\ du$$

Diskretisierung dieses Operators führt auf eine Matrix vom Typ der Hilbertmatrix $(H)_{ij} = \frac{1}{i+j+1}$.

Für $\gamma = 5$ seien einige Singulärwerte angegeben:

k	σ_k
1	0,8751
2	0,1935
3	0,0327
4	0,0074
5	0,0014

Bei den praktischen Anwendungen der Laplace-Transformation gilt typischerweise:

i) Nur sehr wenige Singulärwerte sind deutlich von Null verschieden.

ii) $\mathcal{L}f(t)$ ist nur für relativ wenige Werte von t verfügbar.

Wir haben es bei der Laplace-Transformation mit einem Integraloperator vom Typ

$$Af(x) = \int_a^b K(x,y)f(y)dy \quad , \quad K \in C^\infty$$

zu tun. Hierbei ist

$$g(x) = Af(x)$$

bekannt für $x = x_i$, $i = 1,2,\ldots,n$ (n relativ klein).

Wir betrachten die diskrete Version von A:

$$\mathcal{A}: L_2(a,b) \to \mathbb{R}^n \quad , \text{ mit}$$

$$(\mathcal{A}f)_i = Af(x_i) = \int_a^b K(x_i,y)f(y)dy \quad .$$

Mit $g = \begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix}$, $g_i = g(x_i)$

wollen wir die Minimum-Norm-Lösung f_M von

$$\mathcal{A}f = g$$

berechnen, d.h. wir lösen

$$\mathcal{A}\mathcal{A}^* v = g$$

und setzen $f_M = \mathcal{A}^* g$.

Nach Aufgabe 7 gilt:

$$\mathcal{A}^*: \mathbb{R}^n \to L_2(a,b) \quad , \mathcal{A}^* \begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix} = \sum_{i=1}^n K(x_i,y) g_i$$

somit

$$\mathcal{A}\mathcal{A}^*: \mathbb{R}^n \to \mathbb{R}^n, \qquad (\mathcal{A}\mathcal{A}^* g)_k = \int_a^b K(x_k,y) \sum_{i=1}^n K(x_i,y) g_i$$

$$= \sum_{i=1}^n \underbrace{\left(\int_a^b K(x_k,y)K(x_i,y)dy \right)}_{= (\mathcal{A}\mathcal{A}^*)_{ki}} g_i$$

Seien nun v_k die Eigenvektoren von $\mathcal{M}\mathcal{M}^*$ zu den Eigenwerten σ_k $(k = 1,\ldots,n)$, so ist

$$f_M = \sum_{k=1}^{n} \frac{1}{\sigma_k} (g, v_k) u_k \quad \text{mit} \quad u_k(x) = \frac{1}{\sigma_k} \mathcal{M}^* v_k(x) \; .$$

Fallen die Singulärwerte sehr schnell ab, so wird man die Summation vorzeitig abbrechen:

$$f_{M,p} = \sum_{k=1}^{p} \frac{1}{\sigma_k} (g, v_k) u_k \qquad , \quad p \leq n \; .$$

Wir wollen den Fehler $f_{M,p} - f$, wobei f die exakte Lösung sein soll (also $\mathcal{M} f = g$), berechnen

$$f_{M,p} = \sum_{k=1}^{p} \frac{1}{\sigma_k} (\mathcal{M} f, v_k)_{\mathbb{R}^n} u_k = \sum_{k=1}^{p} \frac{1}{\sigma_k} (f, \mathcal{M}^* v_k)_{L_2} u_k$$

$$= \sum_{k=1}^{p} (f, u_k)_{L_2} u_k$$

$$\Rightarrow \quad f_{M,p}(x) = \sum_{k=1}^{p} \int_a^b f(y) u_k(y)\, dy \; u_k(x)$$

$$= \int_a^b f(y) \delta_p(x,y)\, dy$$

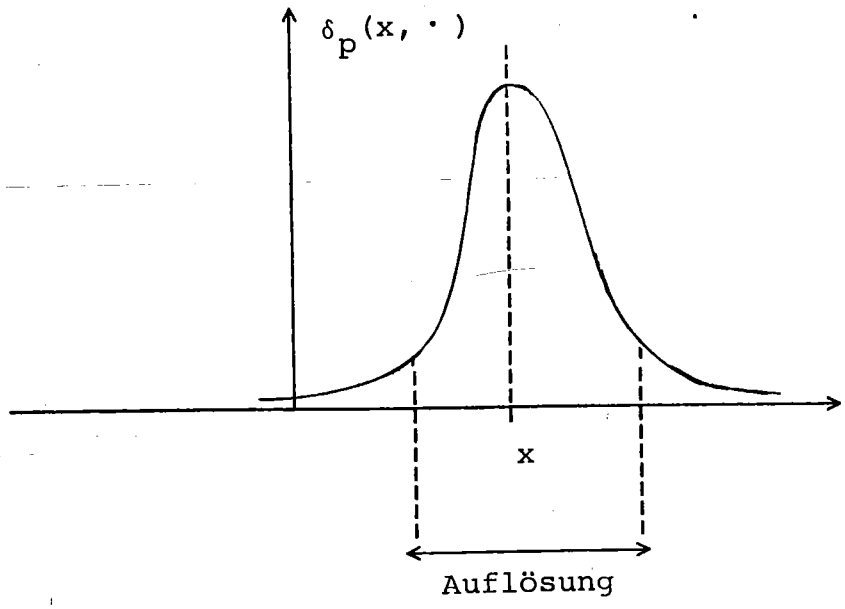
$$\text{mit } \delta_p(x,y) = \sum_{k=1}^{p} u_k(x) u_k(y)$$

Die Näherungslösung $f_{M,p}$ ergibt sich also durch Integration der exakten Lösung gegen eine Kernfunktion $\delta_p(x,y)$. Der Idealfall wäre

$$\delta_p(x,y) = \delta(x - y) \qquad \text{(Diracdistribution)}$$

d.h. $f_{M,p}(x) = f(x) \qquad \forall\, x \; .$

Indem wir die Funktion $\delta_p(x, \cdot)$ für verschiedene Werte von x numerisch berechnen, können wir uns einen Überblick über die Qualität der Näherungslösung $f_{M,p}(x)$ verschaffen. $\delta_p(x, \cdot)$ wird nicht vollständig auf x konzentriert sein, sondern eine gewisse endliche "Auflösung" besitzen.



Liegen zwei "Details" der Funktion f (z.B. zwei nah beieinanderliegende Funktionsspitzen) innerhalb der lokalen Auflösung, so werden diese in $f_{M,p}$ nicht mehr aufgelöst, z.B. würden die zwei Funktionsspitzen in f zu einer einzigen in $f_{M,p}$ "zusammengeschmiert" erscheinen.

BEISPIEL:

$$(\mathcal{M}f)_i = \int_{x_i-d/2}^{x_i+d/2} f(y)dy = g_i \quad , \quad i = 1,\ldots,n$$

$$x_{i+1} - x_i = h$$

Anders formuliert:

$$g_i = (\mathcal{H}f)_i = \int_{x_o}^{x_n} K(x_i,y)f(y)dy \quad \text{mit} \quad K(x_i,y) = \begin{cases} 1 , & |x_i - y| < \frac{d}{2} \\ 0 , & \text{sonst} \end{cases}$$

Also: $\mathcal{H}^* g(y) = \sum_{i=1}^{n} K(x_i,y) g_i$

und falls $d \leq h$:

$$\mathcal{H}^* = d \, \mathbb{1}$$

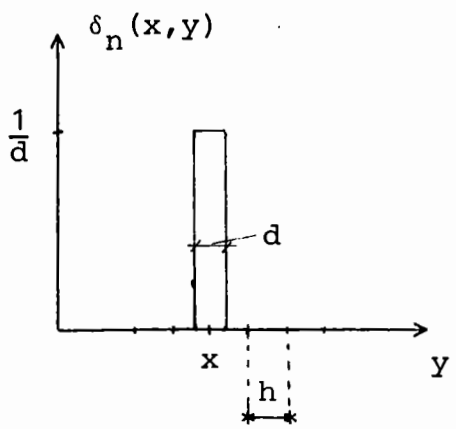
d.h. $\sigma_k = \sqrt{d}$, $u_k(y) = d^{-1/2} K(x_k,y)$, $k = 1,\ldots,n$.

Somit: $f_M(y) = \frac{1}{d} \sum_{k=1}^{n} K(x_k,y) g_k$,

d.h. f_M ist eine Treppenfunktion mit Funktionswerten

$\frac{g_k}{d} = \frac{1}{d} \int_{x_i - d/2}^{x_i + d/2} f(y)dy$, d.h. den lokalen Mittelwerten von f.

Ferner: $\delta_n(x,y) = \frac{1}{d} \sum_{k=1}^{n} K(x_k,x)K(x_k,y) = \begin{cases} \frac{1}{d} , & |x - y| < d \\ 0 , & \text{sonst} \end{cases}$



Für ansteigende Werte von $d > h$ wird der Kern $\delta_n(x,y)$ immer mehr zerfließen, d.h. die Auflösung immer schlechter werden.