MATHEMATIK FÜR PHYSIKER I

Prof. Dr. H. Dinges WS 2001/02

Rückblich Die Teile I und II der Veranstaltung befaßten sich mit **komplexen Zahlen** und **Polynomen**. Es gibt dafür kein Skriptum, nur eine Zusammenfassung mit Themenübersicht.

Teil III. Grundideen der Fourier-Analyse

Themenübersicht

Teil III. Grundideen der Fourier-Analyse

11. Vorlesung : Trigonometrische Polynome

Definition

Ein reelles trigonometrisches Polynom vom Grad $\leq N$ ist eine 2π-periodische Funktion der Form

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{N} (a_k \cos kt + b_k \sin kt) \quad \text{für} \quad t \in {}^{\mathbb{R}}/_{2\pi}$$

mit reellen Koeffizienten a_k, b_k. Wenn komplexe Koeffizienten zugelassen sind, dann bekommt man die komplexen trigonometrischen Polynome. Meistens schreibt man die trigonometrischen Polynome lieber mit Hilfe der Funktionen

$$e^{int} = \cos nt + i \sin nt \; ; \quad e^{-int} = \cos nt - i \sin nt$$

Man definiert also:
Ein (komplexes) trigonometrisches Polynom ist eine 2π-periodische Funktion der Form

$$f(t) = \sum_{-N}^{+N} c_n e^{int} \quad \text{mit} \quad c_n \in \mathbb{C}$$

Bemerkungen

1) $\sum c_n e^{int} = c_0 + \sum_{1}^{N} (c_n + c_{-n}) \cos nt + \sum_{1}^{N} i\,(c_n - c_{-n}) \cdot \sin nt$.

2) $c_0 = \frac{a_0}{2}$, $c_n = \frac{1}{2}(a_n - ib_n)$; $\quad c_{-n} = \frac{1}{2}(a_n + ib_n)$ für $n \in \mathbb{N}$.

3) $\sum c_n e^{int}$ ist genau dann reellwertig, wenn $\overline{f(t)} = f(t)$, d.h. $c_n = \bar{c}_{-n}$ für alle $n \in \mathbb{R}$.

Satz :

Sei $\mathfrak{T}^{(N)}$ der Raum der trigonometrischen Polynome vom Grad $\leq N$, $\mathfrak{T}$ der Vektorraum aller trigonometrischen Polynome;

a) $\mathfrak{T}^{(N)}$ ist ein $(2N+1)$-dimensionaler $\mathbb{C}$-Vektorraum.

b) Wenn $f(\cdot) \in \mathfrak{T}^{(N)}$, dann $f'(\cdot) \in \mathfrak{T}^{(N)}$. Die Ableitung $\sum (inc_n)\, e^{int}$ liefert, über die volle Periode 2π integriert, das Integral 0.

c) Das punktweise Produkt trigonometrischer Polynome ist ein trigonometrisches Polynom. Dabei gilt

$$\text{Grad} f_1(t) \leq n_1 \,, \; \text{Grad} f_2(t) \leq n_2 \Longrightarrow \text{Grad}\Big(f_1 \cdot f_2\Big)(t) \leq n_1 + n_2 \quad .$$

d) Mit $f(\cdot)$ ist auch $\bar{f}(\cdot)$ ein trigonometrisches Polynom, ebenso $\Re f(\cdot) = \frac{1}{2}\left(f + \bar{f}\right)(\cdot)$, $\Im f(\cdot) = \frac{1}{2i}\left(f - \bar{f}\right)(\cdot)$.

$$f = \sum c_n e^{int} \Longrightarrow \bar{f} = \sum \bar{c}_{-n} \cdot e^{int} \quad .$$

Satz :

a) Die Koeffizienten von $f(t) = \sum c_n e^{-int}$ ergeben sich durch Integration

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(t) e^{-int} dt \quad .$$

b) Wenn $f(t) = \frac{a_0}{2} + \sum_1^N (a_k \cos kt + b_k \sin kt)$, dann

$$a_k = \frac{1}{\pi} \int_{-\pi}^{+\pi} f(t) \cdot \cos(kt) dt \ , \quad b_k = \frac{1}{\pi} \int_{-\pi}^{+\pi} f(t) \cdot \sin(kt) dt \quad .$$

Der Beweis ergibt sich aus dem

Lemma (Orthogonalitätsrelationen)

$$\frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{ikt} \cdot e^{-ilt} dt = \begin{cases} 1 & \text{falls} \quad k = \ell \\ 0 & \text{falls} \quad k \neq \ell \end{cases}$$

$$\frac{1}{\pi} \int_{-\pi}^{+\pi} \cos(kt) \sin(\ell t) dt = 0 \ , \ \ \frac{1}{\pi} \int_{-\pi}^{+\pi} \cos(kt) \cos(\ell t) kt = \begin{cases} 1 & \text{für} \quad k = \ell \\ 0 & \text{für} \quad k \neq \ell \ . \end{cases}$$

Das **punktweise Produkt** trigonometrischer Polynome wird im Abschnitt „Faltung“ untersucht. Wir erwähnen hier nur

$$\begin{array}{ll} & A(t) = \sum a_k e^{ikt} \ , \ B(t) = \sum b_\ell e^{i\ell t} \ , \ C(t) = A(t) \cdot B(t) = \sum c_n e^{int} \\ \Longrightarrow & c_n = \sum\limits_k a_k \cdot b_{n-k} \ \text{ für alle } \ n \in \mathbb{Z} \ . \end{array}$$

Man nennt die Folge (c_n) das Faltungsprodukt der Folgen (a_k) und (b_k).

Hier interessieren wir uns für das „innere Produkt“ oder **„Skalarprodukt“** zweier trigonometrischer Polynome (beliebigen Grades).

Satz

Man erhält dieselbe komplexe Zahl $\langle f|g \rangle$, wenn man für

$$f(t) = \sum a_k e^{ikt} \quad \text{und} \quad g(t) = \sum b_\ell e^{i\ell t}$$

definiert

$$\langle f|g \rangle = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \bar{f}(t) \cdot g(t) dt$$

oder

$$\langle f|g \rangle = \sum \bar{a}_k \cdot b_k \quad .$$

Beweis mit Hilfe der Orthogonalitätsrelationen.

Satz
Das innere Produkt $\langle\cdot|\cdot\rangle$ hat die Eigenschaften

(i) $\langle f|f\rangle \geq 0 \quad \text{und} \quad \langle f|f\rangle = 0 \Longrightarrow f \equiv 0 \quad .$

(ii) $\langle f_1 + f_2|g\rangle = \langle f_1|g\rangle + \langle f_2|g\rangle$

(iii) $\langle f|\alpha g\rangle = \alpha \cdot \langle f|g\rangle$ für alle $\alpha \in \mathbb{C}$

(iv) $\langle g|f\rangle$ ist konjugiert komplex zu $\langle f|g\rangle$.

Bemerke : Für $\alpha, \beta \in \mathbb{C}$ gilt

$$\begin{aligned} \langle f|\alpha g_1 + \beta g_2\rangle &= \alpha \cdot \langle f|g_1\rangle + \beta\langle f|g_2\rangle \\ \langle \alpha f_1 + \beta f_2|g\rangle &= \bar{\alpha}\langle f_1|g\rangle + \bar{\beta}\langle f_2|g\rangle \end{aligned}$$

Man nennt $\langle\cdot|\cdot\rangle$ eine Sesquilinearform, die im zweiten Argument linear ist.

Satz (Ungleichung von Cauchy-Schwarz-Bunjakovski)

$$|\langle f|g\rangle| \leq \sqrt{\langle f|f\rangle} \cdot \sqrt{\langle g|g\rangle} \quad .$$

Beweis
Es genügt solche Paare f, g zu untersuchen, für welche $\langle f|g\rangle > 0$. Sei f, g ein solches Paar. Der Fall, wo $f(\cdot)$ und $g(\cdot)$ linear abhängig sind, ist trivial.
Wenn f und g linear unabhängig sind, dann gilt für alle reellen α

$$0 < \langle f + \alpha g|f + \alpha g\rangle = \langle f|f\rangle + 2\alpha \cdot \langle f|g\rangle + \alpha^2 \cdot \langle g|g\rangle \quad .$$

Eine quadratische Funktion $a\alpha^2 + 2b\alpha + c$ hat genau dann keine reelle Nullstelle, wenn $b^2 - ac < 0$. Diesen Fall haben wir hier. Also gilt

$$|\langle f|g\rangle|^2 \leq \langle f|f\rangle \cdot \langle g|g\rangle \quad .$$

Hinweis : Die trigonometrischen Polynome bilden das algebraische Herzstück mehrerer analytischer Theorien. Diese Theorien betreffen trigonometrische Reihen

$$\sum_{-\infty}^{+\infty} c_n \cdot e^{int} \quad ,$$

wo die Koeffizientenfolge $(c_n)_{n\in\mathbb{Z}}$ verschiedenen Forderungen zu genügen haben.

A) Bei den sog. **Fourier-Reihen** wird gefordert, dass $(c_n)_n$ **trigonometrische Reihen, quadratisch summable** ist, d.h. $\sum |c_n|^2 < \infty$. Die Theorie der Fourier-Reihen müssen wir zurückstellen, bis wir etwas Integrationstheorie und etwas Theorie der Konvergenz entwickelt haben. Wir werden dann sehen: Die Gesamtheit aller quadratisch summierbaren trigonometrischen Reihen bildet einen **Hilbertraum**.

B) In der Theorie, die wir anschließend skizzieren werden, wird gefordert, dass die $(c_n)_n$ **absolut summabel** sind, d.h. $\sum |c_n| < \infty$. In diesem Fall liefert die (gleichmäßig konvergente) trigonometrische Reihe eine stetige 2π-periodische Funktion

$$f(t) = \sum_{-\infty}^{+\infty} c_n \cdot e^{int} \quad .$$

Die Gesamtheit aller absolut summierbaren trigonometrischen Reihen bildet (mit dem Faltungsprodukt) eine **Banachalgebra**. Bevor wir das beweisen, müssen wir einige Vorbereitungen treffen.

12. Vorlesung : Faltung über $\mathbb{Z}$

Unter einer (komplexen) **Gewichtung** auf $\mathbb{Z}$ verstehen wir eine Familie komplexer Zahlen

$$(a_n)_{n\in\mathbb{Z}} \quad \text{mit} \quad \|a\|_1 = \sum |a_n| \leq \infty \quad .$$

Spezialfälle

a) Wenn nur endlich viele a_n von 0 verschieden sind, sprechen wir von einer **finiten Gewichtung**.

b) Wenn $a_n \geq 0$ und $\sum a_n = 1$, dann sprechen wir von einer **Wahrscheinlichkeitsgewichtung** auf $\mathbb{Z}$.

Definition

Die **charakteristische Funktion** der Gewichtung $(a_n)_{n\in\mathbb{Z}}$ ist die 2π-periodische stetige Funktion

$$A(t) := \sum a_n e^{int} \quad .$$

Bemerke : Die Gewichte a_n ergeben sich aus der charakteristischen Funktion durch Integration

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{+\pi} A(t) \cdot e^{-int} dt \quad .$$

Satz

1) Gewichtungen a und b kann man komplex linear kombinieren. Dabei gilt

$$\begin{array}{rcl} \|a+b\|_1 & \leq & \|a\|_1 + \|b\|_1 \\ \|\alpha \cdot a\|_1 & = & |\alpha| \cdot \|a\|_1 \quad \text{für alle} \quad \alpha \in \mathbb{C} \quad . \end{array}$$

2) Die Gewichtungen a und b kann man auch falten. Dabei gilt

$$\|a * b\|_1 \leq \|a\|_1 \cdot \|b\|_1 \quad .$$

Beweis : Die erste Aussage ist trivial. Betrachten wir das Faltungsprodukt $c = a * b$.

$$c_n = \sum_k a_k \cdot b_{n-k} = \sum_{\left\{(k,\ell):k+\ell=n\right\}} a_k \cdot b_\ell \quad \text{für} \quad n \in \mathbb{R}.$$

Es gilt

$$\begin{aligned} \|a*b\|_1 = \sum_n |c_n| = \sum_n \Big| \sum_k a_k b_{n-k} \Big| \leq \sum_n \sum_{\{k+\ell=n\}} |a_k| \cdot |b_\ell| = \\ = \sum_k |a_k| \cdot \sum_\ell |b_\ell| = \|a\|_1 \cdot \|b\|_1 \quad . \end{aligned}$$

Spezialfall :

Wenn a und b Wahrscheinlichkeitsgewichtungen sind, dann ist auch das Faltungsprodukt $a * b$ eine Wahrscheinlichkeitsgewichtung.

Satz :

Die charakteristische Funktion des **Faltungsprodukts** ist das **punktweise Produkt** der charakteristischen Funktionen

$$\left(\sum_k a_k e^{ikt}\right) \cdot \left(\sum_\ell b_\ell e^{i\ell t}\right) = \sum_n c_n e^{int} \quad \text{wenn} \quad c = a * b \quad .$$

Der Beweis ist trivial.

Notation : Zu jeder Gewichtung a definieren wir die **konjugierte Gewichtung** $b = a^*$ als die Folge mit den Einträgen

$$b_n = \bar{a}_{-n} \quad \text{für alle } n \in \mathbb{Z} \quad .$$

Bemerke : Die charakteristische Funktion der konjugierten Gewichtung gewinnt man durch (punktweise) komplexe Konjugation der charakteristischen Funktion.

$$B(t) := \sum b_n e^{int} = \sum \bar{a}_{-n} e^{int} = \sum \bar{a}_k e^{-ikt} = \overline{\left(\sum a_n e^{int}\right)} = \overline{A(t)} \quad .$$

Notation

Die Stochastiker assoziieren mit einer Wahrscheinlichkeitsgewichtung $(p_n)_n$ $\mathbb{Z}$-wertige Zufallsgrößen T, sodass

$$\mathrm{Ws}(T = n) = p_n \quad .$$

Sie nennen die charakteristische Funktion der Gewichtung die charakteristische Funktion (der Verteilung) von T und notieren

$$A(t) = \mathcal{E}\big(\exp(itT)\big) = \sum_n e^{itn} \cdot \mathrm{Ws}(T = n) \quad ;$$

Bemerke : $\mathcal{E}\Big(\exp\big(it(-T)\big)\Big) = \overline{\mathcal{E}\big(\exp(itT)\big)}$.

Satz : Wenn Z die Summe unabhängiger Zufallsgrößen ist $Z = S + T$, dann ist die Gewichtung zu Z das Faltungsprodukt der Gewichtungen zu S und T

$$\mathrm{Ws}(S+T=n) = \sum_k \mathrm{Ws}(S = k, T = n-k) = \sum_k \mathrm{Ws}(S = k) \cdot \mathrm{Ws}(T = n-k) \ .$$

Beispiel (Geometrische Gewichtung)

1) Ein Zufallsexperiment mit der Erfolgswahrscheinlichkeit p wird unabhängig wiederholt. T bezeichne die Wartezeit bis zum ersten Erfolg. Wir haben

$$\mathrm{Ws}(T = 1) = p\,, \ \ \mathrm{Ws}(T = 2) = (1-p) \cdot p\,, \ldots$$
$$\mathrm{Ws}(T = n) = (1-p)^{n-1} \cdot p \quad .$$

Um Schreibarbeit zu sparen und zu Gunsten der Übersichtlichkeit setzen wir $\alpha = 1 - p$ und betrachten die Wahrscheinlichkeitsgewichtung $a = (a_n)_{n\in\mathbb{Z}}$ mit

$$a_n = \begin{cases} (1-\alpha)\cdot\alpha^{n-1} & \text{für} \quad n = 1, 2, \dots \\ 0 & \text{für} \quad n \le 0 \quad . \end{cases}$$

(„Geometrische Gewichtung mit dem Erwartungswert $\frac{1}{p} = \frac{1}{1-\alpha}$“).

2) Die charakteristische Funktion kann man durch das Aufsummieren einer geometrischen Reihe elementar darstellen:

$$A(t) = \sum a_n e^{int} = (1-\alpha)\cdot\frac{e^{it}}{1-\alpha\cdot e^{it}} \quad .$$

3) Seien S und T unabhängige Wartezeiten zum gleichen Parameter $p = 1-\alpha$. Die charakteristische Funktion von $S+T$ ist $A^2(t)$.

4) Man berechnet leicht die Gewichte von $a * a$:

$$\mathrm{Ws}(S+T=2) = (1-\alpha)^2\ ,\ \mathrm{Ws}(S+T=3) = (1-\alpha)^2\cdot 2\alpha\,,\ \dots$$
$$\mathrm{Ws}(S+T=n+1) = (1-\alpha)^2\cdot n\cdot\alpha^{n-1} \ \text{ für } n = 2, 3, \ \dots$$

Mit der Formel $\sum n\cdot z^{n-1} = \frac{1}{(z-1)^2}$ für $|z| < 1$ kann man die charakteristische Funktion auch direkt aus den Gewichten berechnen

$$\mathcal{E}\exp\big(it(T+S)\big) = (1-\alpha)^2 e^{2it}\cdot\frac{1}{\big(1-\alpha e^{it}\big)^2} = A^2(t) \quad .$$

5) Die Zufallsgröße $T - S$ kann alle Werte $n \in \mathbb{Z}$ mit positiver Wahrscheinlichkeit annehmen. Es gilt

$$\mathrm{Ws}(T-S=0) = \sum_k \mathrm{Ws}(T=k)\cdot\mathrm{Ws}(S=k) =$$
$$= (1-\alpha)^2\cdot\sum_1^\infty \alpha^{k-1}\cdot\alpha^{k-1} = (1-\alpha)^2\cdot\frac{1}{1-\alpha^2}$$
$$\mathrm{Ws}(T-S=n) = \mathrm{Ws}(T-S=-n) =$$
$$= (1-\alpha)^2\cdot\alpha^{|n|}\cdot\frac{1}{1-\alpha^2} \quad \text{für alle } \ n\in\mathbb{Z} \quad .$$

6) Diese „doppeltgeometrische“ Gewichtung kann man auch aus der charakteristischen Funktion ablesen, wenn man beachtet

$$\mathcal{E}\exp\big(it(T-S)\big) = A(t)\cdot\overline{A(t)} = (1-\alpha)^2\cdot\frac{1}{1-\alpha e^{it}}\cdot\frac{1}{1-\alpha e^{-it}} =$$
$$= \frac{(1-\alpha)^2}{1-\alpha^2}\cdot\left(\frac{1}{1-\alpha e^{it}} + \frac{1}{1-\alpha e^{-it}} - 1\right)$$
$$= \frac{(1-\alpha)^2}{1-\alpha^2}\left(\sum_0^\infty \alpha^n\cdot e^{int} + \sum_0^\infty \alpha^n e^{-int} - 1\right)$$

Beispiel (Binomialgewichtung)
Ein Zufallsexperiment mit der Erfolgswahrscheinlichkeit p wird n-mal unabhängig wiederholt. X bezeichne die Anzahl der Erfolge. Man nennt X eine binomialverteilte Zufallsgröße. Die Gewichtung

$$b_k = \begin{cases} \left(\begin{array}{c} n \\ k \end{array} \right) p^k \cdot (1-p)^{n-k} & \text{für } k \in \{0, 1, \ldots, n\} \\ 0 & \text{sonst} \end{cases}$$

nennt man die Binomialgewichtung zum Parameter (n, p).
Die Binomialkoeffizienten

$$\left(\begin{array}{c} n \\ k \end{array} \right) = \frac{1}{k!} n(n-1) \cdot \ldots \cdot (n-k+1)$$

sollten von der Schule her bekannt sein. Man versammelt sie gern im Pascal'schen Dreieck.
Die charakteristische Funktion der Binomialgewichtung ist:

$$B(t) = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) p^k \cdot (1-p)^{n-k} \cdot e^{ikt} = \left(1 - p + p \cdot e^{it}\right)^n \quad .$$

Für $n = 1$ erhält man die charakteristische Funktion der Gewichtung, welche in den Punkt 0 das Gewicht $1 - p$ und in den Punkt 1 das Gewicht p legt („Bernoulli-Gewichtung“ mit Erwartungswert p). $B(t)$ ist die n-te Potenz der charakteristischen Funktion der Bernoulli-Gewichtung; die Anzahl der Erfolge in n Versuchen ist nämlich die Summe von n unabhängigen Bernoulli-Variablen.

13. Vorlesung : Zeitinvariante lineare Filter

In der Nachrichtentechnik beschäftigt man sich mit linearen Filtern (in diskreter oder kontinuierlicher Zeit). Jede Gewichtung $(a_k)_{k\in\mathbb{Z}}$ beschreibt einen zeitdiskreten TLF („timeinvariant linear filter“). Der TLF nimmt eine Folge von Signalen $x(n)$ auf und gibt eine Folge $y(n)$ heraus.

$$y(n) = \sum_k a_k \cdot x(n-k) \quad \text{für alle } n \in \mathbb{Z} \quad .$$

Besonders wichtig sind die „kausalen“ TLF; ihre Gewichte verschwinden für $k < 0$.

$$y(n) = a_0 \cdot x(n) + a_1 \cdot x(n-1) + a_2 \cdot x(n-2) + \ldots \quad .$$

Der Filter verarbeitet hier also die Eingangssignale bis zur Zeit n in linearer Weise zum Ausgangssignal; a_k kann man als den Faktor deuten, mit dem ein Puls vor k Zeitschritten zum Ausgangssignal beiträgt. Die Folge (a_k) heißt die Pulsantwort des TLF.

$$x(n) \longrightarrow \boxed{\text{TLF}_a} \longrightarrow y(n)$$

Wenn das Eingangssignal eine „reine Sinusschwingung“ mit der Kreisfrequenz ω ist,

$$x(n) = C \cdot \exp(i\omega n), \quad n \in \mathbb{Z}$$

dann ist das Ausgangssignal ebenfalls eine reine Sinusschwingung mit der Kreisfrequenz ω. Die komplexe Amplitude wird aber mit einem (von ω abhängenden) Faktor multipliziert

$$y(n) = A(-\omega) \cdot x(n) = \left(\sum_k a_k \cdot e^{-i\omega k} \right) \cdot x(n) \quad ;$$

$$\text{denn} \sum a_k e^{i\omega(n-k)} = e^{i\omega n} \cdot \sum_k a_k e^{-i\omega k} = e^{i\omega n} \cdot A(-\omega) \quad .$$

Die 2π-periodische Funktion

$$\omega \longmapsto A(-\omega)$$

heißt die **Transferfunktion** des Filters; ihr Absolutquadrat $|A(-\omega)|^2$ heißt die Leistungstransferfunktion („power transfer function“).

Lineare Filter TLF_a und TLF_b kann man hintereinanderschalten

$$x(n) \longrightarrow \boxed{\text{TLF}_a} \quad \overset{y(n)}{\longrightarrow} \quad \boxed{\text{TLF}_b} \quad \overset{z(n)}{\longrightarrow}$$

Man erhält den TLF zum Faltungsprodukt $C = a * b$. Die Transferfunktion ist das punktweise Produkt

$$C(-\omega) = A(-\omega) \cdot B(-\omega) \quad \text{für } \omega \in \mathbb{R}/_{2\pi} \quad .$$

Wenn ein Filter die höheren Frequenzen stärker dämpft als die in der Nähe von 0, dann spricht man von einem Tiefpaß. Typische Tiefpässe sind die gleitenden Mittel („moving average“) mit Wahrscheinlichkeitsgewichten.

Beispiel 1

$$y(n) = \frac{1}{3}x(n) + \frac{1}{3}x(n-1) + \frac{1}{3}x(n-2) \quad .$$

Die Transferfunktion ist im Wesentlichen der „Dirichlet-Kern“ (der Ordnung 1), den wir (für alle N) in den Übungen ausführlich studiert haben.

Beispiel 2

Ein weiterer wichtiger Tiefpass ist der mit geometrisch abfallenden Gewichten

$$y(n) = x(n) + \alpha \cdot x(n-1) + \alpha^2 \cdot x(n-2) + \ldots \quad .$$

Die Transferfunktion ist

$$A(-\omega) = \sum_0^\infty \alpha^k \cdot e^{i\omega k} = \frac{1}{1 - \alpha \cdot e^{i\omega}} \quad .$$

Die Power-Transferfunktion ist

$$|A(-\omega)|^2 = \frac{1}{1 + |\alpha|^2 - 2|\alpha| \cos(\omega - \omega_0)} \ , \ \text{wenn}\, \alpha = |\alpha| \cdot e^{i\omega_0} \quad .$$

(Man erinnere sich, dass wir in der Übung zur Spiegelung am Einheitskreis die Kurven $\left\{ \frac{1}{1+R\cdot e^{it}} \, ; \, t \in (0, 2\pi] \right\}$ studiert haben; $A(-\omega)$ variiert mit ω wie der Abstand des Nullpunkts von einem Kreis in der komplexen Ebene. Wenn $|\alpha|$ nahe bei 1 ist, dann kann man von einer „Resonanz“ bei der Kreisfrequenz ω_0 sprechen).

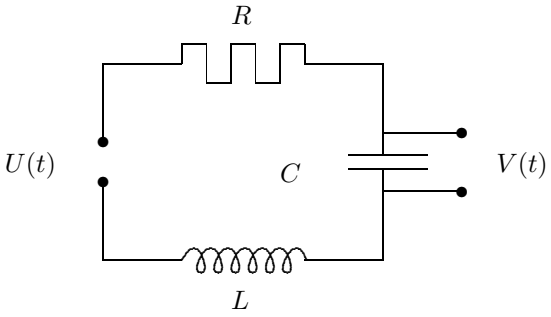
Beispiel 3

Ein TLF, der alles andere als ein Tiefpass ist, ist der „Differenzenoperator“

$$\begin{aligned} y(n) &= x(n) - x(n-1) \\ A(-\omega) &= 1 - e^{-i\omega} = (1 - \cos\omega) + i \sin\omega \\ |A(-\omega)|^2 &= 2(1 - \cos\omega) = 4 \cdot \sin^2 \tfrac{\omega}{2} \quad . \end{aligned}$$

Kontinuierliche Zeit

Ein TLF in kontinuierlicher Zeit ist gegeben durch eine „kontinuierliche Gewichtung“ $\big(a(s) \ : \ s \in \mathbb{R}\big) \ ; \ \int_{-\infty}^{-\infty} \big|a(s)\big| ds < \infty$. Die Technik der gekoppelten Schwingkreise liefert reichhaltige Möglichkeiten, lineare Filter zu realisieren. Die dazugehörigen Gewichtungen werden wir diskutieren, wenn wir über inhomogene lineare Differenzialgleichungen mit konstanten Koeffizienten reden. Hier wenden wir uns sofort der charakteristischen Funktion zu.
Betrachten wir den einfachsten Fall: Für einen einfachen Schwingkreis sei R der Ohm'sche Widerstand, L die Induktivität und C die Kapazität. Wir zwingen dem Schwingkreis die Spannung $U(t)$ auf und greifen die Spannung $V(t)$ am Kondensator ab, die bekanntlich proportional zur Ladung $Q(t)$ ist.



$Q(t)$ ist eine Lösung der inhomogenen Differentialgleichung zweiter Ordnung

$$L \cdot \ddot{Q}(t) + R \cdot \dot{Q}(t) + \frac{1}{C} Q(t) = U(t) \quad .$$

Wenn $U(t) = U_0 e^{i\omega t}$, dann stellt sich nach dem Einschwingvorgang die „Gleichgewichtslösung“ ein.

$$Q(t) = Q_0 \cdot e^{i\omega t} \ , \ \ Q_0 = U_0 \cdot A(-\omega) \quad .$$

Wir gewinnen die Transferfunktion $A(-\omega)$ aus der Differentialgleichung:

$$A(-\omega)\big[L \cdot (-\omega)^2 + R \cdot (i\omega) + \frac{1}{C}\big] = 1 \quad .$$

Die üblichen Bezeichnungen sind

$$\gamma = \frac{R}{L} \quad \text{und} \quad \omega_0^2 = \frac{1}{L \cdot C}$$

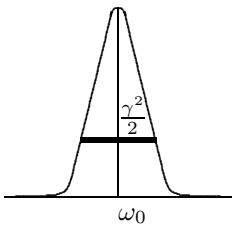
(ω_0 ist die Resonanzfrequenz, wenn die Dämpfung abgeschaltet wird)

$$\begin{aligned} A(-\omega) &= \tfrac{1}{L} \cdot \left(\omega_0^2 - \omega^2 + \gamma(i\omega)\right)^{-1} \\ |A(-\omega)|^2 &= \tfrac{1}{L^2} \cdot \Big(\big(\omega^2 - \omega_0^2\big) + \gamma^2\omega^2\Big)^{-1} \quad . \end{aligned}$$

Eine beliebte Näherungsformel für $\omega \sim \omega_0$ und kleine Dämpfung γ liefert $\big($wegen $\omega_0^2 - \omega^2 = (\omega_0 - \omega)\,(\omega_0 + \omega) \sim 2\omega_0\,(\omega - \omega_0)\big)$.

$$\big|A(-\omega)\big|^2 \approx \frac{1}{4L^2 \cdot \omega_0^2} \cdot \frac{1}{(\omega_0 - \omega)^2 + \frac{\gamma^2}{4}} \quad .$$

Diese Funktion hat ihr Maximum in ω_0. In den Punkten $\omega_0 \pm \frac{1}{2}\gamma^2$ liefert sie die Hälfte des Maximalwerts; man interpretiert daher γ^2 als die „Breite der Resonanzkurve“.



Den Physikstudenten sei hier nachdrücklich empfohlen :

Feynman, *Vorlesungen über Physik*, Band 1, Kap. 23, „Resonanz“.

Allgemeines über integrable Gewichtungen auf $\mathbb{R}$

a) Eine komplexwertige Funktion $a(\cdot)$ mit

$$\int_{-\infty}^{+\infty} \big|a(s)\big| ds < \infty$$

nennen wir eine integrable Gewichtung oder eine komplexwertige Dichte mit der L^1-Norm

$$\|a\|_1 = \int_{-\infty}^{+\infty} \big|a(s)\big| ds \quad .$$

b) Wenn $p(s) \geq 0$ und $\int_{-\infty}^{+\infty} p(s)ds = 1$, dann nennt man $p(\cdot)$ eine Wahrscheinlichkeitsdichte auf $\mathbb{R}$.

c) Von einer integrablen Gewichtung $a(\cdot)$ sagt man, dass sie kompakten Träger hat, wenn ein T existiert, sodass $a(s) = 0$ für $s \notin [-T, +T]$.

d) Der zu $a(\cdot)$ gehörige Filter ist die Operation, welche jeder stetigen beschränkten Funktion $f(\cdot)$ vermöge der Faltungsregel eine Funktion $g = \mathrm{TLF}_a(f)$ zuordnet.

$$g(t) = \int_{-\infty}^{+\infty} a(s) \cdot f(t-s) ds \quad \text{für} \quad t \in \mathbb{R} \quad .$$

e) Wenn $a(\cdot)$ und $b(\cdot)$ integrable Gewichtungen sind, dann definiert man ihr Faltungsprodukt $c = a * b$ als die Gewichtung

$$c(t) = \int a(s) \cdot b(t-s) ds \quad \text{für} \quad t \in \mathbb{R} \quad .$$

Satz : Für integrable Gewichtungen gilt

$$\|a * b\|_1 \leq \|a\|_1 \cdot \|b\|_1 \quad .$$

Wenn $a(\cdot)$ und $b(\cdot)$ Wahrscheinlichkeitsdichten sind, dann ist auch $a * b$ eine Wahrscheinlichkeitsdichte.

Beweis

1) $\int \big|c(t)\big| dt = \int dt \big| \int a(s)b(t-s)ds \big|$

$$\begin{array}{rcl} \leq & \int dt \int \big|a(s)\big|\big|b(t-s)\big| ds & = \quad \int du \int \big|a(s)\big| \cdot \big|b(u)\big| ds \\ = & \int \big|a(s)\big| ds \cdot \int \big|b(u)\big| du & . \end{array}$$

2) Im Falle, wo $a(\cdot)$ und $b(\cdot)$ positiv sind, wird die Ungleichung zu einer Gleichung.

Definition
Die charakteristische Funktion zu $a(\cdot)$ ist die Funktion

$$A(t) = \int e^{its} \cdot a(s)ds \quad \text{für} \quad t \in \mathbb{R} \quad .$$

Satz
Die charakteristische Funktion zum Faltungsprodukt $c = a*b$ ist das punktweise Produkt der charakteristischen Funktionen

$$C(t) = A(t) \cdot B(t) \quad \text{für} \quad t \in \mathbb{R} \quad .$$

Beweis

$$\begin{array}{rcl} C(t) & = & \int e^{its} \cdot c(s)ds \; = \; \int\int e^{it(s-u)} \cdot e^{itu} a(u) \cdot b(s-u)duds \\ & = & \int e^{itu} a(u)du \cdot \int e^{itv} b(v)dv \; = \; A(t) \cdot B(t) \quad . \end{array}$$

13* Regularisierung

Der Leser sollte zuerst die konkreten Beispiele im Abschnitt 14 kennenlernen und dann erst den folgenden Wechsel des Standpunkts mitmachen.
Die Anwendung des Filters TLF_p auf eine (möglicherweise nicht genügend glatte) Funktion $f(\cdot)$ kann man manchmal als eine Glättung von $f(\cdot)$ verstehen, dann nämlich, wenn $p(\cdot)$ glatt ist und auf eine kleine Umgebung des Nullpunkts konzentriert ist. Die Zahl

$$g(t) = \int p(s) \cdot f(t-s)ds$$

ist dann ein gewichtetes Mittel der Funktionswerte in der Nähe von t. Die Mittelung hat zur Folge, dass das Resultat in Abhängigkeit von t ebenso glatt wie der „Regularisierungskern“ $p(\cdot)ds$ ist: Um das zu sehen, schreiben wir

$$g(t) = \int p(t-s) \cdot f(s)ds \quad .$$

Satz : Wenn $p(\cdot)$ stetig differenzierbar ist mit einer integrablen Ableitung, dann gilt für alle beschränkten $f(\cdot)$

$$g'(t) = \int p'(t-s) \cdot f(s)ds \quad \text{ohne Beweis!}$$

Wir studieren nun, wie ein Regularisierer TLF_p auf eine „reine Sinusschwingung“ wirkt. Die Funktion $f(s) = e^{i\omega s}$ wird in eine reine Sinusschwingung mit der Kreisfrequenz ω transformiert, wobei aber die komplexe Amplitude mit einem Faktor multipliziert wird.

$$\int p(s) \cdot e^{i\omega(t-s)}ds = e^{i\omega t} \cdot \int p(s)e^{-i\omega s} = e^{i\omega t} \cdot A(-\omega) \quad .$$

Wenn $|A(-\omega)|^2$ klein ist für große $|\omega|$, dann werden „schnelle Schwankungen“ durch TLF_p stark gedämpft; in diesem Fall verdient der Operator $f \longmapsto \mathrm{TLP}_p$ den Namen „Glätter“.

Beispiele (die eher von theoretischem Interesse sind)

1) Ein beliebter Glätter ist die Faltung mit der Normalverteilung $\mathcal{N}(0, \sigma^2)$. Die Transferfunktion ist

$$A(-\omega) = \exp\left(-\frac{\sigma^2}{2}\omega^2\right)$$

$|A(-\omega)|^2$ ist klein, wenn $|\sigma\omega|$ groß ist.

2) Es gibt lineare Filter, welche die hohen Frequnezen völlig auslöschen. Die Gewichtung sieht etwas merkwürdig aus:

$$p(s) = \frac{S}{\pi}\left(\frac{\sin(sS)}{s \cdot S}\right)^2 \quad \text{für} \quad s \in \mathbb{R} \quad .$$

Die Transferfunktion ist aber sehr einfach:

$$A(-\omega) = \left(1 - \frac{|\omega|}{2S}\right)^+ \quad \text{für} \quad \omega \in \mathbb{R} \quad .$$

Man würde hier nicht von einem Glätter sprechen, weil die Dichte $p(s)$ ausgesprochen schwere Schwänze hat.

3) Es gibt Glätter, die unendlich oft differenzierbar sind und dennoch einen kompakten Träger haben. Man konstruiert solche Glätter z.B. mit Hilfe einer merkwürdigen von Cauchy diskutierten Funktion

$$q(x) = \begin{cases} \exp\left(-\frac{1}{x^2}\right) & \text{für} \quad x > 0 \\ 0 & \text{für} \quad x \leq 0 \end{cases}$$

Die unendlich oft differenzierbare Gewichtung

$$p_\varepsilon(x) = \text{const} \cdot q(-\varepsilon + x) \cdot q(-\varepsilon - x)$$

verschwindet außerhalb $[-\varepsilon, +\varepsilon]$. Durch die Wahl der Konstanten erreicht man, dass $p_\varepsilon(x)$ eine Wahrscheinlichkeitsgewichtung ist.

Hinweis : Wir werden sehen, dass man die Approximation einer 2π-periodischen Funktion durch ihre Fourier-Polynome als eine Glättung auffassen kann. Dabei werden die Dirichlet-Kerne und die Fejér-Kerne eine Rolle spielen.

14. Vorlesung : Charakteristische Funktionen von speziellen Dichten

Wenn man die Eigenschaften von charakteristischen Funktionen

$$A(t) = \int e^{its} a(s) ds$$

in einiger Allgemeinheit studieren will, dann braucht man etwas Maß- und Integrationstheorie („Lebesgue'sches Integral").
Wir wollen hier erst einmal einige elementare Beispiele behandeln. Diese nehmen wir aus der Wahrscheinlichkeitstheorie, weil dort Integrale (neben der „Fläche unter der Kurve") noch andere nützliche Interpretationen haben.
Integrale sind als gewichtete Mittel oder als Erwartungswerte zu deuten.
Als Vorbereitung diskutieren wir

Diskrete gewichtete Mittel

Jedem Element j einer Menge J sei ein komplexer „Funktionswert" h_j zugeordnet. Wenn $|J|$ endlich ist, dann kann man das arithmetische Mittel definieren

$$\bar{h} := \frac{1}{|J|} \sum h_j = \sum p_j \cdot h_j \quad .$$

Hier sind alle p_j gleich $\frac{1}{|J|}$.
Eine Verallgemeinerung des arithmetischen Mittels ist das „mit p gewichtete Mittel". Dabei kann J auch eine unendliche Menge sein.

Definition

Jedem $j \in J$ sei eine Zahl $p_j \geq 0$ zugeordnet, so dass $\sum p_j = 1$. Für jede beschränkte Funktion $h(\cdot)$ auf J heißt

$$\sum p_j h_j = \langle p, h \rangle \quad ,$$

das mit p gewichtete Mittel der Werte von h oder aus das **Integral von h**$(\cdot)$.
(Man sollte sich $h(\cdot)$ als eine beschränkte J-Spalte und p als eine J-Zeile vorstellen.)

Bemerke

Für festgehaltenes p ist $\langle p, \cdot \rangle$ ein lineares Funktional auf dem Vektorraum der beschränkten Funktionen

$$\langle p, \alpha \cdot h_1 + \beta \cdot h_2 \rangle = \alpha \cdot \langle p, h_1 \rangle + \beta \cdot \langle p, h_2 \rangle \quad .$$

Manchmal ist es erforderlich, dieses Funktional auf den VR derjenigen $h(\cdot)$ zu erweitern, für welche gilt $\sum p_j \cdot |h_j| < \infty$. Die Theorie der absolut konvergenten Reihen lehrt, dass es da keine Probleme gibt; man kann wie mit endlichen Summen rechnen.

Die Wahrscheinlichkeitstheoretiker assoziieren gerne zur Gewichtung $(p_j)_{j\in J}$ eine J-wertige Zufallsgröße Z und deuten $\langle p, h\rangle$ als den Erwartungswert der komplexwertigen Zufallsgrößen $h(Z)$.

$$\mathrm{Ws}(Z=j)=p_j \quad \text{für alle} \quad j\in J$$
$$\sum p_j\cdot h_j=\sum h_j\cdot \mathrm{W}s(Z=j)=\mathcal{E}h(Z) \quad .$$

Gewichtungen auf $\mathbb{Z}$

Besondere Bedeutung haben für uns die (Wahrscheinlichkeits)-Gewichtungen auf $\mathbb{Z}$; denn diese kann man falten. Aus dem bereits Gesagten ergibt sich der

Satz

a) Wenn man unabhängige $\mathbb{Z}$-wertige Zufallsgrößen X_1 und X_2 addiert, dann ergibt sich die Gewichtung von X_1+X_2 durch die Faltung der Gewichtungen

$$\mathrm{Ws}\,(X_1+X_2=n)=\sum_k \mathrm{Ws}\,(X_1=k)\cdot\mathrm{Ws}\,(X_2=n-k) \quad \text{für alle} \quad n\in\mathbb{Z} \quad .$$

b) Die charakteristische Funktion einer Gewichtung kann als ein Erwartungswert geschrieben werden

$$\sum e^{itn}\cdot \mathrm{Ws}\,(X=n)=\mathcal{E}\Big(\exp(itX)\Big) \quad .$$

Corollar :

Wenn X_1 und X_2 unabhängig $\mathbb{Z}$-wertig sind, dann gilt für alle $t\in\mathbb{R}$

$$\mathcal{E}\exp\Big(it\,(X_1+X_2)\Big)=\mathcal{E}\exp\,(itX_1)\cdot\mathcal{E}\exp\,(itX_2) \quad .$$

Wahrscheinlichkeitsgewichtungen auf $\mathbb{R}$

Sei $p(x)\geq 0\,,\ \int p(x)=1$. („Wahrscheinlichkeitsgewichtung“).

Sprech- und Bezeichnungsweisen

a) Man sagt von einer reellwertigen Zufallsgröße X, dass sie die Dichte $p(x)$ hat, wenn gilt

$$\mathrm{Ws}\Big(X\in(a,b)\Big)=\int_a^b p(x)dx \quad \text{für alle } a<b \quad .$$

b) Man schreibt auch $\mathrm{Ws}\big(X\in(x,x+dx)\big)=p(x)dx$ für $x\in\mathbb{R}$.

c) Die Stammfunktion

$$F(x) := \int_{-\infty}^{x} p(y)dy = \text{ Ws}(X \leq x)$$

heißt die Verteilungsfunktion der Zufallsgrößen X.

d) Für beschränkte komplexwertige Funktionen $h(\cdot)$ notiert man

$$\int h(x) \cdot p(x)dx = \mathcal{E}h(X) \quad .$$

e) Man notiert $\mathcal{E}\exp(itX) = \int e^{itx}p(x)dx$ (für $t \in \mathbb{R}$) und nennt das Resultat (als Funktion von t betrachtet) die charakteristische Funktion der (Verteilung der) Zufallsgrößen X.

Bemerke : Es hängt von $p(\cdot)$ ab, ob für ein vorgegebenes unbeschränktes $h(\cdot)$ der Erwartungswert $\mathcal{E}h(X)$ wohldefiniert ist. Besonders wichtig ist der Fall $h(X) = X^2$.

Notation

Wenn $\mathcal{E}\,X^2 < \infty$, dann definiert man
$\mathcal{VAR}X = \mathcal{E}X^2 - (\mathcal{E}X)^2 = \mathcal{E}\big((X - \mathcal{E}X)^2\big)$.
Man kann zeigen

Satz

Wenn $\mathcal{E}X^2 < \infty$, dann ist die charakteristische Funktion im Nullpunkt zweimal differenzierbar und es gilt (mit $\mu = \mathcal{E}X$)

$$\mathcal{E}\,\exp(itX) = \exp(it\mu) \cdot \exp\Big(-\frac{1}{2}\sigma^2t^2 + o(t)^2\Big) \quad \text{für } t \to 0 \quad .$$

Corollar : Wenn X_1, X_2 unabhängig sind mit $\mathcal{E}\,X_1^2 < \infty$, $\mathcal{E}\,X_2^2 < \infty$, dann gilt

$$\begin{aligned} \mathcal{E}\,(X_1 + X_2) &= \mathcal{E}X_1 + \mathcal{E}\,X_2 \\ \mathcal{VAR}\,(X_1 + X_2) &= \mathcal{VAR}X_1 + \mathcal{VAR}X_2 \quad . \end{aligned}$$

Beispiel 1 (Gauß'sche Dichten)

Die berühmteste Wahrscheinlichkeitsverteilung ist die Normalverteilung (oder gauß'sche Verteilung) $\mathcal{N}(\mu, \sigma^2)$. Ihre Dichte ist auf den DM 10.- Banknoten abgebildet

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad .$$

Der Spezialfall $\mu = 0$, $\sigma^2 = 1$ liefert die Dichte der Standardnormalverteilung $\mathcal{N}(0, 1)$.

Wer nicht viel Analysis studiert hat, wird sich vielleicht wundern, dass es gerade der Faktor $\sqrt{2\pi}$ ist, der für die Normierung auf das Gesamtintegral 1

benötigt wird. Er wird vielleicht auch nicht sofort die charakteristische Funktion ausrechnen können. Wir wollen hier das Ergebnis nur angeben:

$$\int_{-\infty}^{+\infty} e^{itx} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx = \exp\left(-\frac{1}{2}t^2\right) \quad \text{für} \quad t \in \mathbb{R} \quad .$$

Man bemerke, dass man das Resultat auf die folgende leicht memorierbare Weise schreiben kann:

$$\frac{1}{\sqrt{2\pi}} \int \exp\Big(-\frac{1}{2}\,(x-it)^2\Big) dx = 1 \quad \text{für alle} \quad t \in \mathbb{R} \quad .$$

Fazit :
Wenn Z standardnormalverteilt ist, dann hat $X = \mu + \sigma \cdot Z$ die Verteilung $\mathcal{N}(\mu, \sigma^2)$ mit der oben angegebenen Dichte und der charakteristischen Funktion

$$\begin{aligned} \mathcal{E}\exp(itX) &= \int e^{itx} f(x) dx \\ &= \mathcal{E}\left(\exp\Big(it(\mu + \sigma Z)\Big)\right) \\ &= e^{it\mu} \cdot \mathcal{E}\exp\Big(i(t\sigma)Z\Big) \\ &= e^{it\mu} \cdot \exp\left(-\tfrac{\sigma^2}{2}t^2\right) . \end{aligned}$$

Der Logarithmus der charakteristischen Funktion ist eine quadratische Funktion

$$\ln\big(\mathcal{E}\,\exp(itX)\big) = it\mu - \frac{1}{2!} \cdot t^2 \cdot \sigma^2 = it(\mathcal{E}X) - \frac{1}{2!}t^2 \cdot \mathcal{VAR}X \quad \text{für alle} \quad t \in \mathbb{R}\,.$$

Beispiel 2 (Die exponentielle Dichte und die Laplace-Dichte)

a) Exponentiell verteilte Zufallsgrößen X treten häufig als Wartezeiten auf (etwa beim radioaktiven Zerfall, λ bezeichnet dort die Zerfallsrate).

$$\mathrm{Ws}\big(X \in (x, x+dx)\big) = \begin{cases} e^{-\lambda x} \cdot \lambda dx & \text{für} \quad x \geq 0 \\ 0 & \text{für} \quad x < 0 \end{cases}$$

Man rechnet leicht nach: $\mathcal{E}\,X = \frac{1}{\lambda},\ \mathcal{VAR}\,X = \frac{1}{\lambda^2}$.

$$\mathcal{E}\,\exp(itX) = \int_0^{\infty} e^{itx} \cdot e^{-\lambda x} \cdot \lambda \cdot dx = \frac{\lambda}{\lambda - it} \text{ für } t \in \mathbb{R} \quad .$$

Zur Erinnerung : Wenn ein Experiment mit der Erfolgswahrscheinlichkeit p unabhängig wiederholt wird, dann gilt für die Wartezeit $\tilde{T}$ bis zum ersten Erfolg.

$$\begin{aligned} \mathcal{E}\,\tilde{T} = \tfrac{1}{p}\,, \quad \mathcal{VAR}\,\tilde{T} &= \tfrac{1-p}{p^2} \\ \mathcal{E}\,\exp(it\tilde{T}) = \quad p \cdot \sum_1^{\infty} (1-p)^{n-1} e^{itn} &= \tfrac{p}{p+(e^{-it}-1)} \quad . \end{aligned}$$

b) Wenn X_1 und X_2 unabhängig exponentiell verteilt sind mit $\mathcal{E}X_1 = \frac{1}{\lambda} = \mathcal{E}X_2$, dann ist $X_1 - X_2$ doppeltexponentiell verteilt (oder „Laplace-verteilt“).

$$\mathrm{Ws}\big(X_1 - X_2 \in (x, x+dx)\big) = \frac{1}{2} e^{-\lambda|x|} \cdot \lambda \cdot dx \quad \text{für alle } x \in \mathbb{R} \quad .$$

$$\mathcal{E}\,\exp\big(it(X_1 - X_2)\big) = \frac{\lambda^2}{|\lambda - it|^2} = \frac{\lambda^2}{\lambda^2 + t^2} \quad \text{für } t \in \mathbb{R} \quad .$$

Bemerke : Auch für $X_1 + X_2$ kann man die Dichte und die charakteristische Funktion leicht explizit ausrechnen

$$\begin{aligned} \mathrm{Ws}\big(X_1 + X_2 \in (x, x+dx)\big) &= \begin{cases} x \cdot e^{-\lambda x} \cdot \lambda^2 dx & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases} \\ \mathcal{E}\,\exp\big(it(X_1 + X_2)\big) &= \left(\tfrac{\lambda}{\lambda - it}\right)^2 \quad \text{für } t \in \mathbb{R} \quad . \end{aligned}$$

Satz : Seien $X_1, X_2, \ldots, X_n$ unabhängig exponentiell verteilt mit demselben Erwartungswert $\mathcal{E}\,X = \frac{1}{\lambda}$, dann gilt

$$\mathrm{Ws}\big(X_1 + \ldots + X_n \in (x, x+dx)\big) = \frac{1}{(n-1)!} \cdot x^{n-1} \cdot e^{-\lambda x} \cdot \lambda^n dx \quad \text{für } x \geq 0$$

$$\mathcal{E}\,\exp\big(it(X_1 + \ldots + X_n)\big) = \left(\frac{\lambda}{\lambda - it}\right)^n \quad \text{für } t \in \mathbb{R}.$$

Beweis durch vollständige Induktion nach n.

Beispiel 3 (Gamma-Dichten)

Es ist Tradition, dass die Erstsemester (des Mathematik- oder Physikstudiums) bei irgendeiner Gelegenheit mit der Gamma-Funktion $\Gamma(\alpha), \alpha > 0$ bekannt gemacht werden.

$$\Gamma(\alpha) := \int_0^\infty x^{\alpha-1} \cdot e^{-x} dx \quad .$$

Durch eine partielle Integration gewinnt man die **Funktionalgleichung**

$$\alpha \cdot \Gamma(\alpha) = \Gamma(\alpha + 1) \quad \text{für alle } \alpha > 0 \quad .$$

Daraus ergibt sich für ganzzahlige Argumente

$$\Gamma(n+1) = n!$$

Man sagt, die Gamma-Funktion sei eine „natürliche“ Fortsetzung der „Fakultätsfunktion“ auf nichtganze Argumente - und man kann das auf vielerlei Weisen begründen.

Uns erscheint $\Gamma(\alpha)$ als der Normierungsfaktor in den sog. Gamma-Dichten.

Definition : Man sagt von einer Zufallsgröße X, sie sei gammaverteilt mit dem Erwartungswert $\alpha (\alpha > 0)$, wenn gilt

$$\mathrm{Ws}\big(X \in (x, x+dx)\big) = \begin{cases} \frac{1}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-x} dx & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases}$$

Satz : Wenn X gammaverteilt ist mit $\mathcal{E}X = \alpha$, dann gilt

$$\mathcal{E}\,\exp(itX) = (1-it)^{-\alpha} \quad .$$

Der Beweis erfordert komplexe Integrationstheorie. Wir wollen durch die folgende Plausibilitätsbetrachtung das Ergebnis leichter memorierbar machen:

$$\int_0^\infty e^{itx} \cdot x^\alpha \cdot e^{-x} \cdot \frac{1}{x} dx = \frac{1}{(1-it)^\alpha} \int_0^\infty (1-it)^\alpha \cdot x^\alpha \cdot e^{-x(1-it)} \cdot \frac{1}{x} dx =$$

$$= \frac{1}{(1-it)^\alpha} \int y^\alpha \cdot e^{-y} \cdot \frac{1}{y} dy \stackrel{!}{=} (1-it)^{-\alpha} \cdot \Gamma(\alpha) \quad .$$

Ergänzung zu Beispiel 3

Das Integral

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

heißt in älteren Büchern das zweite Euler'sche Integral. Es gibt auch ein „erstes Euler'sches Integral“:

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx \quad \text{für } \alpha > 0,\, \beta > 0 \quad .$$

Wir zeigen

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

ausgehend von dem Faktum, dass die Faltung der Gamma-Dichten zu α und β die Gamma-Dichte zu $\alpha+\beta$ ergibt. (Dies folgt aus der angegebenen Formel für die charakteristische Funktion.)

$$\frac{1}{\Gamma(\alpha+\beta)} \cdot y^{\alpha+\beta-1} e^{-y} = \int_0^y u^{\alpha-1} e^{-u} \cdot (y-u)^{\beta-1} e^{-(y-u)} du \cdot \frac{1}{\Gamma(\alpha)} \cdot \frac{1}{\Gamma(\beta)} \text{ für } y > 0 \,.$$

In dieser Formel kann man einiges kürzen. Man erhält :

$$\frac{\Gamma(x)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)} y^{\alpha+\beta-1} = \int_0^y u^{\alpha-1} \cdot (y-u)^{\beta-1} du \quad .$$

Variablentransformation $v = \frac{1}{y}u$ liefert :

$$\frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 v^{\alpha-1} \cdot (1-v)^{\beta-1} dv \quad \text{q.e.d.}$$

Hinweis : Die Wahrscheinlichkeitsdichte

$$p(u) = \begin{cases} \frac{1}{(B\alpha,\beta)} \cdot u^{\alpha-1} \cdot (1-u)^{\beta-1} & \text{für } u \in [0,1] \\ 0 & \text{sonst} \end{cases}$$

heißt die Beta-Dichte zum Parameter α, β. Wenn $X_{\alpha,\beta}$ nach dieser Verteilung verteilt ist, dann gilt $\mathcal{E}\left(X_{\alpha,\beta} = \frac{\alpha}{\alpha+\beta}\right)$.

Beweis

$$\begin{aligned}\int_0^1 x \cdot \tfrac{1}{B(\alpha,\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1} dx &= \tfrac{B(\alpha+1,\beta)}{B(\alpha,\beta)} = \\ &= \tfrac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \cdot \tfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} = \tfrac{\alpha}{\alpha+\beta} \quad .\end{aligned}$$

Beispiel 4

Man sagt von einer Zufallsgröße X, dass sie eine Standard-Cauchy-Dichte hat, wenn gilt

$$\mathrm{Ws}\big(X \in (x, x+dx)\big) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} dx \quad \text{für } x \in \mathbb{R} \quad .$$

Die Stammfunktion ist durch den Arcustangens gegeben.

$$\mathrm{Ws}(X \le x) = \frac{1}{\pi} \cdot \Big[\arctan\, x + \frac{\pi}{2}\Big] = \frac{1}{\pi} \int_{-\infty}^{x} \frac{1}{1+y^2} dy \quad .$$

Die charakteristische Funktion hat eine etwas merkwürdige aber doch ganz elementare Gestalt, nämlich

$$\mathcal{E}\, \exp(itX) = \exp(-|x|) \quad \text{für } t \in \mathbb{R} \quad \text{(ohne Beweis !)}$$

Bemerkungen

1) Die Dichte von $c \cdot X$ nennt man die Cauchy-Dichte zum Parameter c $(c > 0)$.

$$\mathrm{Ws}\big(c \cdot X \in (x, x+dx)\big) = \frac{1}{\pi} \frac{c}{c^2+x^2} dx \quad .$$

Ihre charakteristische Funktion ist

$$\mathcal{E}\big(\exp(it \cdot cX)\big) = \exp(-c \cdot |t|\big) \quad .$$

2) Aus dieser Formel ergibt sich sofort der

Satz

Seien X_1 und X_2 unabhängig standard Cauchy-verteilt und $c_1, c_2 > 0$.
Dann ist $\frac{1}{c_1+c_2}\,(c_1X_1 + c_2X_2)$ standard Cauchy-verteilt.

3) Als Kontrast formulieren wir den

Satz

Seien Z_1 und Z_2 unabhängig standard normalverteilt.
Dann ist

$$\frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}}\,(\sigma_1 Z_1 + \sigma_2 Z_2) \quad \text{standardnormalverteilt.}$$

Der Beweis ergibt sich mit Hilfe der charakteristischen Funktionen

$$\begin{aligned}&\mathcal{E}\, \exp\big(it\,(\sigma_1 Z_1 + \sigma_2 Z_2)\big) = \mathcal{E}\, \exp\,(it\sigma_1 Z_1) \cdot \mathcal{E}\big(\exp\,(it\sigma_2 Z_2)\,\big) \\ = \;& \exp\,\big(-\tfrac{1}{2}\sigma_1^2 t^2\big) \cdot \exp\,\big(-\tfrac{1}{2}\sigma_2^2 \cdot t^2\big) = \exp\big(-\tfrac{1}{2}\,\big(\sigma_1^2+\sigma_2^2\big)\,t^2\big) \quad .\end{aligned}$$

Beispiel 5 (Rechtecks- und Dreiecksdichte)

a) Von einer Zufallsgröße U sagt man, sie sei gleichmäßig (oder uniform) verteilt im Intervall $\left[-\frac{S}{2}, +\frac{S}{2}\right]$, wenn gilt

$$\mathrm{Ws}\big(U \in (u, u+du)\big) = \begin{cases} \frac{1}{S} & \text{für} u \in \left[-\frac{S}{2}, +\frac{S}{2}\right] \\ 0 & \textit{sonst} \end{cases}$$

Man überzeugt sich leicht: $\mathcal{E}U = 0\,,\ \mathcal{VAR}U = \frac{1}{12}S^2$.

$$\mathcal{E}\,\exp(itU) = \frac{2}{S \cdot t}\sin\left(\frac{1}{2}St\right) \quad \text{für } t \in \mathbb{R} \quad .$$

b) Von einer Zufallsgröße V sagt man, sie sei dreiecksverteilt mit der Spannweite $[-S, +S]$, wenn gilt

$$\mathrm{Ws}\big(V \in (v, v+dv)\big) = \frac{1}{S} \cdot \left(1 - \frac{|v|}{S}\right)^2 \quad \text{für } v \in \mathbb{R} \quad .$$

Es gilt offenbar

$$\mathcal{E}V = 0\,,\ \mathcal{VAR} = \frac{1}{6}S^2 \quad .$$

Satz :
Wenn U_1, U_2 unabhängig in $[-S, +S]$ uniform verteilt sind, dann ist $U_1 + U_2$ dreiecksverteilt mit der Spannweite $[-2S, 2S]$.

Beweis : Sei $p(\cdot)$ die Dichte von U_1 und U_2 und $q(\cdot)$ die Dichte von $U_1 + U_2$. Es gilt

$$q(v) = \int p(u) \cdot p(v-u)du \quad .$$

Der Integrand kann nur die Werte $\left(\frac{1}{2S}\right)^2$ und 0 annehmen; er ist ungleich 0, wenn

$$u \in [-S, +S] \cap [-S+v, S+v] \quad \text{denn}$$
$$|v-u| < S \Leftrightarrow -S < v-u < S \Leftrightarrow -S-v < -u < S-v \quad .$$

Die Länge des essentiellen Integrationsintervalls ist also im Falle $v > 0$ gleich $S-(-S+v)$ und im Falle $v < 0$ gleich $S+v+S$, in jedem Falle also $(2S-|v|)^+$. Somit $q(v) = \left(\frac{1}{2S}\right)^2 (2S-|v|)^+ = \frac{1}{2S} \cdot \left(1 - \frac{|v|}{2S}\right)^+$.

Corollar : Wenn V dreiecksverteilt ist mit der Spannweite $[-S, +S]$, dann gilt

$$\mathcal{E}\,\exp(itV) = \Big(\frac{2}{St} \cdot \sin\left(\frac{1}{2}St\right)\Big)^2 \quad .$$

Hinweis : Wenn man auf einem Taschenrechner die Taste „Random Number“ betätigt, dann bekommt man eine sog. Pseudo-Zufallszahl U, die im Einheitsintervall gleichmäßig verteilt ist. Die Dichte von $U - \frac{1}{2}$ ist also die Rechtecksdichte mit der Spannweite $\left[-\frac{1}{2}, +\frac{1}{2}\right]$, $\mathrm{Ws}\big(U \in (u, u+du)\big) = p(u)du$.

a) Wenn man die Taste öfters betätigt, bekommt man $U_1, U_2, \ldots$, die (angeblich) unabhängig sind. $U_1 + U_2 - 1$ hat die Dreiecksdichte mit der Spannweite $[-1, +1]$. $U_1 + U_2 + U_3 - \frac{3}{2}$ hat eine auf $\left[-\frac{3}{2}, +\frac{3}{2}\right]$ konzentrierte Dichte $p * p * p$, die stückweise durch quadratische Funktionen gegeben ist, u.s.w.
Es ist eine interessante Aufgabe MAPLE dazu zu bringen, die Dichten $p * p * p$, $p * p * p * p$, ... zu plotten. Wir behandeln dieses Problem in den Übungen.

b) Praktiker brauchen für Simulation öfters einmal Realisierungen von standardnormalverteilten Zufallsgrößen Z. Für den Hausgebrauch wird empfohlen $Y := U_1 + \ldots + U_{12} - 6$ als approximativ $\mathcal{N}(0,1)$-verteilt gelten zu lassen.

c) Die Koeffizienten der Entwicklung

$$\ln \mathcal{E} \exp(itY) = 6 \cdot \ln \left(\frac{2(1-\cos t)}{t^2}\right) =$$
$$= -\frac{\sigma^2}{2}t^2 + \frac{1}{4!}\kappa_4 \cdot t^4 - \frac{1}{6!}\kappa_6 \cdot t^6 + \ldots$$

heißen die Kumulanten der Verteilung von Y. In unserem Fall haben wir $\sigma^2 = \mathcal{VARZ} = 1$. Die Kumulanten $\kappa_4, \kappa_6, \ldots$ sind leicht zu berechnen; ihre Kleinheit gilt als Indiz für die Ähnlichkeit der Verteilung von Y mit $\mathcal{N}(0, \sigma^2)$.

15. Vorlesung : Die Inversionsformel

Satz : Sei $a(s)$ eine Gewichtung, sodass die charakteristische Funktion $A(t)$ integrabel ist. Dann gilt

$$a(s) = \frac{1}{2\pi} \int e^{-its} A(t) dt \quad \text{für alle } s \in \mathbb{R} \quad .$$

Wir stellen eine grundsätzliche Diskussion dieser Inversionsformel zurück und werfen lieber noch einmal einen Blick auf unsere Beispiele.

Beispiel 1

Wir haben noch keine Methode, um die charakteristische Funktion der gaußischen Dichte herzuleiten; wir kennen aber das Resultat

$$A(t) = \int e^{itx} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx = \exp\left(-\frac{1}{2}t^2\right) \quad .$$

Die Inversionsformel bietet keine Überraschung

$$\frac{1}{2\pi} \int e^{-itx} \cdot A(t) dt = \frac{1}{2\pi} \int e^{-ist} \cdot \exp\left(-\frac{1}{2}t^2\right) dt = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \quad .$$

Hier trägt dieselbe Rechnung wie bei der Berechnung der charakteristischen Funktion.

Beispiel 2

a) Die exponentielle Dichte ist unstetig; die charakteristische Funktion $A(t)$ ist nicht integrabel. Die Inversionsformel gilt in modifizierter Form

$$A(t) = \int_0^\infty e^{itx} e^{-x} dx = \frac{1}{1-it}$$

$$\frac{1}{2\pi} \int_{-T}^{+T} e^{-ixt} \cdot A(t) dt = \frac{1}{2\pi} \int_{-T}^{+T} e^{-ixt} \cdot \frac{1}{1-it} dt \longmapsto \begin{cases} e^{-x} & \text{für} \quad x > 0 \\ 0 & \text{für} \quad x < 0 \end{cases}$$

b) Die doppelexponentielle Dichte hat eine integrable charakteristische Funktion

$$A(t) = \frac{1}{2} \int_{-\infty}^{+\infty} e^{itx} \cdot e^{-|x|} dx = \frac{1}{1+t^2} \quad .$$

Nachdem wir in Beispiel 4 die charakteristische Funktion der Cauchy-Dichte kennen gelernt haben, überrascht uns nicht, was die Inversionsformel hier zu sagen hat.

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ixt} \cdot \frac{1}{1+t^2} dt = \frac{1}{2} \exp(-|x|) \quad \text{für alle } x \quad .$$

(Für eine direkte Herleitung fehlen uns noch die Mittel).

Beispiel 3
Für die Gamma-Dichten mit ganzzahligem Parameter kann man die charakteristische Funktion (durch vollständige Induktion) elementar berechnen. Mit den Mitteln der Integration im Komplexen findet man auch für nicht ganze $\alpha > 0$

$$A(t) = \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{itx} \cdot x^{\alpha-1} e^{-x} dx = \frac{1}{(1-it)^\alpha} \quad .$$

Die Gamma-Dichten sind für $\alpha \leq 1$ im Nullpunkt unstetig: ihre charakteristischen Funktionen sind nicht integrabel. Die (für $\alpha \leq 1$ zu modifizierende) Inversionsformel liefert die bemerkenswerte Identität

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ixt} \cdot \frac{1}{(1-it)^\alpha} a dt = \begin{cases} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} & \text{für} \quad x > 0 \\ 0 & \text{für} \quad x < 0 \end{cases}$$

Beispiel 4 haben wir mit Beispiel 2b) abgehandelt.
Beispiel 5 setzt die Rechtecksdichten und die Dreiecksdichten in Beziehung zu den kontinuierlichen Analoga der Dirichlet- und Fejér-Kerne, die wir in den Übungen ausführlich behandelt haben. In der Übungsaufgabe wird diskutiert: Sei für $S > 0$

$$D^{(S)}(t) := \int_{-S}^{+S} e^{its} ds = \int_0^S 2 \cdot \cos(ts) ds = \frac{2}{t} \cdot \sin(tS)$$

$$F^{(S)}(t) := \frac{1}{S} \int_0^S D^{(v)}(t) dv = \int_{-\infty}^{+\infty} \left(1 - \frac{|v|}{s}\right)^+ e^{itv} dv = \left(\frac{2}{t} \sin\left(\frac{t}{2}\right)\right)^2$$

$$\frac{1}{2\pi} \int_{-T}^{+T} e^{-iut} \cdot \frac{2}{t} \sin t dt \xrightarrow[T\to\infty]{} \begin{cases} \frac{1}{2} & \text{für} \quad u \in (-1,+1) \\ 0 & \text{sonst} \end{cases}$$

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ivt} \cdot \left(\frac{2}{t} \sin\left(\frac{t}{2}\right)\right)^2 dt = (1-|v|)^+ \quad \text{für alle } v \quad .$$

Beispiel 6 (Der harmonische Oszillator)
Wir haben berechnet, wie ein einfacher Schwingkreis eine reine Sinusschwingung modifiziert

$$e^{i\omega t} \longmapsto e^{i\omega t} \cdot A(-\omega) \quad \text{mit}$$

$$A(-\omega) = \frac{1}{LC\left(\omega_0^2 - \omega^2 + \gamma(i\omega)\right)} = \frac{1}{LC} \cdot \frac{1}{\omega_+ + i\omega} \cdot \frac{1}{\omega_- + i\omega} \quad .$$

Die integrable Funktion

$$A(t) = \frac{\omega_+}{\omega_+ - it} \cdot \frac{\omega_-}{\omega_- - it}$$

ist das Produkt zweier charakteristischer Funktionen zu Exponentialgewichtungen (mit den komplexen Parametern ω_+ und ω_-).

Die Pulsantwort $a(\cdot)$ des harmonischen Oszillators ist daher (bis auf einen reellen Faktor) das Faltungsprodukt zweier Exponentialgewichtungen.

$$\begin{aligned} p_{\omega_+}(t) &= e^{-\omega_+ t}\cdot\omega_+ \quad \text{für } t>0\,, \quad =0 \text{ sonst} \\ p_{\omega_-}(t) &= e^{-\omega_- t}\cdot\omega_- \quad \text{für } t>0\,, \quad =0 \text{ sonst}\,. \end{aligned}$$

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\omega t}A(\omega)d\omega = \text{const}\left(p_{\omega_+} * p_{\omega_-}\right)(t) \quad .$$

Das nächste Beispiel erfordert ein gewisses Verständnis für den Begriff der Glättung, den wir in 13* behandelt haben.

Beispiel 7 (Modulierte Sinusschwingungen)
Einen TLF kann man sowohl durch seine Transferfunktion $A(-\omega)$, als auch durch seine Pulsantwortfunktion $a(t)$ charakterisieren.

$$\begin{aligned} A(-\omega) &= \int e^{-i\omega t}a(t)ds \\ a(t) &= \tfrac{1}{2\pi}\int e^{-i\omega t}A(\omega)d\omega \quad . \end{aligned}$$

$A(\tilde{\omega})$ ist der Faktor, mit dem der Filter die (aus der unendlichen Vergangenheit kommende) reine Sinusschwingung $e^{-i\tilde{\omega}t}$ multipliziert. Nun kann man aber in der Realität dem Filter nur Eingaben endlicher Reichweite anbieten. Für die Bestimmung von $A(\tilde{\omega})$ bieten sich schwach modulierte Schwingungen an, wie z.B.

a) $f(t) = f^{(\tilde{\omega},S)}(t) = 1_{[-S,+S]}(t)\cdot e^{-i\tilde{\omega}t}$, oder auch

b) $f(t) = \exp\left(-\frac{t^2}{2S^2}\right)\cdot e^{-i\tilde{\omega}t}$ (für großes S).

Die Ausgabe $g(t)$ ist dann (für $|t| \ll S$) nur annähernd gleich $e^{-i\tilde{\omega}t}\cdot A(\tilde{\omega})$. Wir wollen $g(t)$ etwas genauer berechnen.

$$g(t) = \int a(s)f(t-s)ds \quad .$$

Für die charakteristische Funktion gilt also

$$\begin{aligned} G(u) &= \int e^{itu}\cdot g(t)dt = A(u)\cdot F(u)\,, \quad \text{wobei} \\ F(u) &= \int e^{iut}\cdot f(t)dt \quad . \end{aligned}$$

Die Inversionsformel liefert

$$g(t) = \frac{1}{2\pi}\int e^{-iut}A(u)\cdot F(u)du \quad .$$

Im Falle a) haben wir mit

$$\begin{aligned} D^{(S)}(v) &:= \int_{-S}^{+S} e^{ivs}\cdot ds = \tfrac{2\sin(vS)}{v} \quad \text{(siehe Übungen)} \\ F(u) &= \int_{-S}^{+S} e^{iut}\cdot e^{-i\tilde{\omega}t}dt = D^{(S)}(\tilde{\omega}-u) \\ g(t) &= e^{-i\tilde{\omega}t}\cdot\tfrac{1}{2\pi}\int e^{-i(u-\tilde{\omega})t}\cdot A(u)\cdot D^{(S)}(\tilde{\omega}-u)\,du \\ &= e^{-i\tilde{\omega}t}\cdot\int A(\tilde{\omega}-v)\cdot\tfrac{1}{2\pi}e^{ivt}\cdot D^{(S)}(v)dv \quad . \end{aligned}$$

Das Integral ist die Faltung von $A(\cdot)$ mit der komplexen Gewichtung

$$p_t(v)dv = \frac{1}{2\pi} e^{ivt} D^{(S)}(v)dv \quad .$$

Das Gesamtintegral ist 1 für $t = 0$; dennoch handelt es sich nicht um eine Glättung, weil $D^{(S)}(\cdot)$ nicht überall nichtnegativ ist. Für große S ist die Gewichtung auf eine kleine Umgebung des Nullpunkts konzentriert. Man erwartet daher

$$g^{(\tilde{\omega},S)}(t) \approx e^{-i\tilde{\omega}t} \cdot A\,(\tilde{\omega}) \quad \text{für } |t| \text{ klein},$$

jedenfalls dann, wenn $A(\cdot)$ in der Nähe von $\tilde{\omega}$ wenig variiert.
Für größere $|t|$ und $\tilde{\omega}$, in deren Nähe $A(\cdot)$ wenig schwankt, erwarten wir

$$g^{(\tilde{\omega},S)}(t) \approx e^{-i\tilde{\omega}t} \cdot A\,(\tilde{\omega}) \cdot \int \frac{1}{2\pi} e^{ivt} D^{(S)}(v)dv \quad ,$$

wobei nach der Inversionsformel gilt

$$\int \frac{1}{2\pi} e^{ivt} D^{(S)}(v)dv = 1_{[-S,S]}(t) \quad .$$

Im Falle b) wurde die reine Sinusschwingung $e^{-i\tilde{\omega}t}$ mit einer (für großes S weitgestreckten) gaußischen Dichte moduliert. Auch in diesem Fall erwarten wir (für kleine t) als Antwortsignal eine nur wenig modifizierte Sinusschwingung

$$g(t) \approx e^{-i\tilde{\omega}t} \cdot A(\tilde{\omega}) \quad .$$

Wir wollen nun aber genauer berechnen, was herauskommt, wenn $\frac{|t|}{S}$ nicht sehr klein ist.
Die charakteristische Funktion des Eingangssignals ist

$$\begin{aligned} F(u) &= \int e^{iut} e^{-i\tilde{\omega}t} \cdot \exp\left(-\frac{t^2}{2S^2}\right) dt = \\ &= \sqrt{2\pi} \cdot S \cdot \exp\big(-\frac{S^2}{2} \cdot (u-\tilde{\omega})^2\big) \quad . \end{aligned}$$

Die Inversionsformel für $G(u) = A(u) \cdot F(u)$ liefert

$$\begin{aligned} g(t) &= e^{-i\tilde{\omega}t} \cdot \frac{1}{2\pi} \int e^{-i(u-\tilde{\omega})t} \cdot A(u) \cdot \sqrt{2\pi} \cdot S \cdot \exp\big(-\frac{S^2}{2}(u-\tilde{\omega})^2\big) du \\ &= e^{-i\tilde{\omega}t} \cdot \int A\,(\tilde{\omega} - v) \cdot p_t(v)dv \end{aligned}$$

mit

$$p_t(v)dv = \frac{1}{\sqrt{2\pi}} S \cdot \exp\left(-\frac{1}{2}S^2 \cdot v^2\right) \cdot e^{ivt} dv \quad .$$

Für $t = 0$ ist $p_t(v)dv$ eine Wahrscheinlichkeitsdichte. Für $t \neq 0$ ist $p_t(\cdot)$ nicht reellwertig. Für alle t ist $|p_t(v)|dv$ auf die Nähe des Ursprungs konzentriert.
Die Formel für das Gesamtintegral

$$\int p_t(v)dv = \exp\left(-\frac{t^2}{2S^2}\right)$$

zeigt für die $\tilde{\omega}$, in deren Nähe die Transferfunktion wenig gekrümmt ist,

$$g(t) \approx e^{-i\tilde{\omega}t} \cdot A\,(\tilde{\omega}) \cdot \exp\left(-\frac{t^2}{2S^2}\right) \quad .$$

Fazit : Die mit der flachen Normaldichte modulierte Sinusschwingung $e^{-i\tilde{\omega}t}$ wird im Filter approximativ zu der mit derselben Normaldichte modulierten Sinusschwingung $A(\tilde{\omega}) \cdot e^{-i\tilde{\omega}t}$.

16. Vorlesung : Stehende und laufende Wellen

Stellen wir uns eine elastische Saite vor, die an den Stellen $x = 0$ und $x = L$ eingespannt ist. Sie sei mit einer Massendichte $m(x)dx$ belegt und habe in jedem x eine Elastizität $E(x)$, welche angibt, welche Kraft der Dehnung entgegensteht, wobei angenommen wird, dass diese Kraft proportional zur Dehnung ist.
In dem üblichen idealisierenden Modell (für kleine Auslenkungen) hat man nach Newton's Formel

$$m(x) \cdot \frac{\partial^2}{\partial t^2} u(t,x) = E(x) \frac{\partial^2}{\partial x^2} u(t,x) \quad .$$

(Masse × Beschleunigung = rücktreibende Kraft)?
Dass die Dehnung mit der zweiten Ableitung zu tun hat, liegt im Falle der Longitudinalwellen auf der Hand. Bei den transversalen Schwingungen erscheint die rücktreibende Kraft als eine Resultierende der entlang der Saite angreifenden Kräfte.
Uns soll hier nur der Fall interessieren, wo $m(\cdot)$ und $E(\cdot)$ entlang der Saite konstant sind.
Mit $c = \sqrt{\frac{E}{m}}$ haben wir dann die

Gleichung der schwingenden Saite

$$\left(\frac{1}{c^2} \cdot \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} \right) u(t,x) = 0$$
$$u(t,0) = 0 = u(t,L) \quad \text{für alle } t \quad .$$

Die Lösung ist durch die Auslenkung und die Geschwindigkeit zur Zeit eindeutig bestimmt. Seien also $u(0,x)$ und $\dot{u}(0,x)$ gegeben.

Hinweise :

a) Die analoge Gleichung in zwei Raumdimensionen heißt die

Gleichung der schwingenden Membrane

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \right) u(t,x) = 0$$
$$u\big(t,(x,y)\big) = 0 \quad \text{für } (x,y) \text{ am Rand der Membran.}$$

Für die Anfangsbedingungen zur Zeit 0 gilt dasselbe wie für die schwingende Saite.

b) In drei Raumdimensionen wird daraus die

Wellengleichung

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \Delta \right) u(t,x,y,z) = 0$$
$$\text{wobei } \Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad \text{„Laplace Operator“}$$

c) Lassen wir alle Randbedingungen außer Acht, so können wir schon einmal Lösungen eines interessanten Typs hinschreiben, die sogenannten ebenen Wellen.

$$u(t,x,y,z) = \exp\Big(i\big(\omega t - (k_1 x + k_2 y + k_3 z)\big)\Big)$$

mit einer beliebigen „Wellenzahl“ (k_1, k_2, k_3) und der dazu passenden „Kreisfrequenz“

$$\omega = \|k\| \cdot c \quad .$$

Man hat in der Tat

$$\Delta u(t,x,y,z) = \big(-k_1^2 - k_2^2 - k_3^2\big) \cdot u = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} u(t,x,y,z) \quad .$$

Die ebenen Wellen sind ein besonders wichtiger Fall von Lösungen, die man faktorisieren kann.

d) Machen wir nun den allgemeinen „Separationsansatz“ :

$$u(t,x,y,z) = g(t) \cdot h(x,y,z) \quad .$$

Die Wellengleichung fordert für Lösungen dieser Gestalt

$$\frac{1}{c^2} g''(t) \cdot h(x) = g(t) \cdot \Delta h(x)$$

und das bedeutet, dass eine Konstante existiert, sodass

$$\frac{1}{c^2}\frac{g''(t)}{g(t)} = \text{const} = \frac{\Delta h(x)}{h(x)} \quad .$$

Die Lösungen der Gleichung

$$g''(t) = -\omega^2 \cdot g(t)$$

sind natürlich nichts anderes als die Linearkombinationen der Funktionen

$$g(t) = e^{i\omega t} \quad \text{und} \quad g(t) = e^{-i\omega t}$$

oder $g(t) = \cos(\omega t)$ und $g(t) = \sin(\omega t)$.
Sei auf der anderen Seite $h(\cdot)$ eine Lösung der Gleichung

$$\Delta h = -k^2 \cdot h \quad .$$

Dann ist mit $\omega^2 = k^2 \cdot c^2$

$$\exp\,(i\omega t) \cdot h(x)$$

eine Lösung der Wellengleichung.

Besonders einfach ist natürlich der Fall einer einzigen Raumdimension, d.h. der Fall der schwingenden Saite. Bevor wir uns diesem Fall (mit der angegebenen Randbedingung) zuwenden, skizzieren wir kurz den allgemeinen

Hintergrund

Die Theorie der Wellenbewegung ist eine unendliche Geschichte für Mathematiker wie für Physiker. Sie ist auch eine Herausforderung an die Lehre. In Mathematikbüchern wird man schwerlich eine umfassende Definition von Wellenbewegung finden; und in Physikbüchern baut man ohnehin lieber auf konkrete Assoziationen als auf allgemeine Definitionen. Im berühmten Lehrbuch „Physik“ von Gerthsen findet sich (auf Seite 102) die folgende (nach meiner Meinung sehr gelungene) „Definition und Beschreibung“

„Wenn im Innern eines deformierbaren Mediums eine Verschiebung aus der Ruhelage (Deformation) bewirkt oder „erregt“ wird, so bleibt diese nicht auf das Erregungszentrum beschränkt, sondern sie teilt sich den Nachbargebieten mit, die (zeitlich verzögert) ebenfalls deformiert werden. Eine Erregung pflanzt sich nach allen Richtungen mit einer charakteristischen ***Ausbreitungsgeschwindigkeit*** *fort. Wir nennen diesen zeitlich* ***und*** *räumlich veränderlichen Zustand eine* ***Welle“*** *...*

„Flächen im Medium, deren Punkte mit gleicher Phase schwingen, bezeichnen wir als ***Wellenflächen****. Sie umschließen das Erregungszentrum. Ist dieses punktförmig, und ist die Ausbreitungsgeschwindigkeit unabhängig von der Richtung und überall konstant, dann sind die Wellenflächen Kugelflächen (****Kugelwellen****). Liegt das Erregungszentrum im Unendlichen oder mindestens sehr weit entfernt, oder geht die Welle von einer überall mit gleicher Phase schwingenden Ebene aus, dann sind die Wellenflächen Ebenen (****ebene Wellen****),*

Linien, die vom Erregungszentrum ausgehend, die Wellenflächen überall sendkrecht durchsetzen, bezeichnen wir als ***Strahlen****.*

Wenn die Deformation im Erregungszentrum eine harmonische Schwingung der Teilchen um ihre Ruhelage mit der Schwingungsdauer $T = \frac{1}{\nu}$ ist, so setzt sich diese Schwingung durch den ganzen Körper hindurch fort. Benachbarte Teilchen schwingen in der Phase gegeneinander versetzt. In regelmäßigen Abständen folgen aber Teilchen, die in der Schwingungsphase miteinander übereinstimmen. Wir nennen diesen Abstand innerhalb einer Welle die ***Wellenlänge****.*

Während das Erregungszentrum eine volle Schwingung vollführt hat, ist die Erregung bis zu einem Punkt vorgedrungen, in dem die Schwingung nun mit der des Zentrums gleichphasig ist. Sein Abstand ist nach obiger Definition die Wellenlänge. Er ist gleich der Geschwindigkeit, genauer der Phasengeschwindingkeit c der Wellenausbreitung multipliziert mit der Schwingungsdauer T.

$$c = \frac{\lambda}{T} = (2\pi\omega) \cdot \lambda = \nu \cdot \lambda$$

(c ist nicht die Geschwindigkeit eines Körpers, sondern die eines Zustandes.)

Die Gleichung stellt eine Beziehung zwischen den oben definierten Größen c, λ und T dar; sie sagt aber nichts aus über die Abhängigkeit der Phasengeschwindigkeit von der Frequenz“.

Der letzte Satz aus Gerthsens Buch macht deutlich, dass die hier entwickelten Intuitionen wesentlich über die Theorie der Wellengleichung hinausgehen. In der Wellengleichung gibt es nämlich nur eine einzige Konstante c, während bei allgemeineren Wellenphänomenen die „Dispersion“ eine fundamentale Rolle spielt. Die Dispersionsrelation beschreibt, wie die Kreisfrequenz ω von der Wellenzahl (k_1, k_2, k_3) abhängt. Man könnte auch sagen, dass die Dispersionsrelation an-

gibt, wie die Phasengeschwindigkeit von der Wellenzahl abhängt.

$$c = \frac{\Omega\,(k_1, k_2, k_3)}{\|k\|} = \omega \cdot (2\pi\lambda) \quad .$$

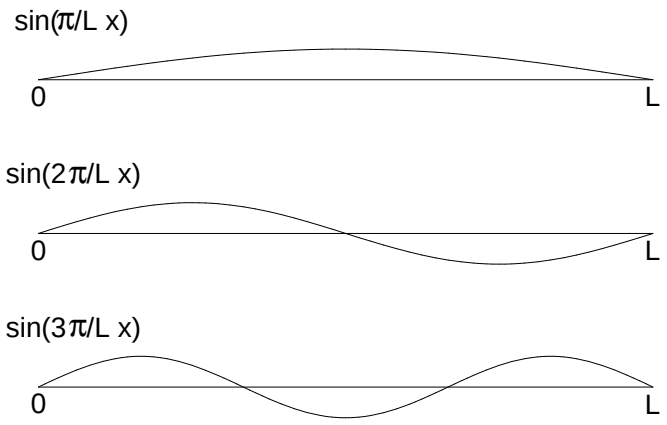
Bemerke : Der Betrag der Wellenzahl hat die Dimension einer reziproken Länge, $\|k\| = \frac{1}{2\pi\lambda}$. In isotropen Medien hängt die Phasengeschwindigkeit nur von der Wellenlänge ab; bei der klassischen Wellengleichung ist sie von der Wellenzahl (k_1, k_2, k_3) gänzlich unabhängig.
Elektromagnetische Wellen breiten sich im materiefreien Raum mit Lichtgeschwindigkeit aus. In einem Medium mit dem Brechungsindex n ist die Phasengeschwindigkeit aber nicht c, sondern $\frac{c}{n}$. Der Brechungsindex ist abhängig von der Frequenz.
Man lese Feynman, *Vorlesungen über Physik*, Band 1, Kap. 31, „Der Ursprung des Brechungsindex".
Der Begriff der Phasengeschwindigkeit ist am Bild der ebenen Wellen entwickelt. Eine mathematische Begründung dafür, dass er für alle Wellenbewegungen fundamental ist, ergibt sich aus der Tatsache, dass man sehr allgemeine Funktionen $u(t, x, y, z)$ aus Funktionen der Form

$$\exp\Big(i\big(\omega t - (k_1 x + k_2 y + k_3 z)\big)\Big)$$

im Sinne einer Integration linear kombinieren kann.
Dieses mathematische Phänomen ist der Gegenstand der Theorie der Fourier-Integrale in höheren Dimensionen.

Eigenschwingungen der Saite

1) Für die in $x = 0$ und $x = L$ eingespannte Saite gibt es eine „Grundschwingung" und eine Folge von Oberschwingungen. Die Wellenzahlen, die hier in Frage kommen, sind die ganzzahligen Vielfachen von $\frac{\pi}{L}$. Die Auslenkungen haben die Form



2) Für $n \in N$ sei $k = \pm\frac{n\pi}{L}$ und $\omega = \pm k \cdot c$. Die Funktion

$$u(t,x) \ := \sin kx \cdot e^{i\omega t}$$

erfüllt die Gleichungen

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}u(t,x) = -k^2 \cdot u(t,x) = \frac{\partial^2}{\partial x^2}u(t,x) \quad .$$

3) Das Produkt $u(t,x)$ kann man auch als Summe schreiben

$$i \cdot u(t,x) = \frac{1}{2}\exp\big(i(\omega t + kx)\big) - \frac{1}{2}\exp\big(i(\omega t - kx)\big) \quad .$$

Die Summanden sind eindimensionale ebene Wellen; man nennt sie laufende Wellen mit den Wellenzahlen $\pm k$ und der Kreisfrequenz ω. Die Phasengeschwindigkeit ist $c = \frac{\omega}{|k|}$.
$u(t,x)$ kann man sich also so zustandekommen denken, dass sich zwei gegeneinander laufende ebene Wellen überlagern, wobei sich die Beiträge in den Positionen $x = 0, \pm L, \pm 2L, \ldots$ zu allen Zeiten auslöschen.

4) Im $\mathbb{C}$-Vektorraum der komplexen Lösungen läßt sich bequemer rechnen als im $\mathbb{R}$-Vektorraum der reellen Lösungen. Die beobachtete Auslenkung kann als der Realteil einer komplexen Lösung verstanden werden. Sie hat die Form

$$v(t,x) = |a| \cdot \sin(kx) \cdot \cos\big(\omega(t - t_0)\big) \quad .$$

Die maximale Amplitude $|a| \cdot \sin(kx)$ wird zu den Zeitpunkten $t_0 + \frac{2\pi}{\omega}m = t_0 + m \cdot T$ $(m \in \mathbb{Z})$ angenommen; $T = \frac{2\pi}{\omega}$ ist die Schwingungsdauer $\nu := \frac{1}{T} = \frac{\omega}{2\pi}$ ist die Frequenz.
Die Geschwindigkeit, mit welcher die Saite durch die Nulllage geht, ist $\pm\omega \cdot |a| \cdot \sin(kx)$. Die kinetische Energie ergibt sich durch Integration des Quadrats.

$$\text{kin.Energie} = \left(\frac{1}{4}M \cdot L\right) \cdot |a|^2 \cdot |\omega|^2 \quad (M = \text{Gesamtmasse})$$

Zu den Zeitpunkten maximaler Auslenkung ist diese Energie in potentielle Energie (Deformationsenergie) umgewandelt.

5) **Die Überlagerung von Oberschwingungen** Die Gesamtheit der Lösungen der Gleichung

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\right)u(t,x) = 0$$
$$u(t,0) = 0 = u(t,L)$$

ist ein Vektorraum. Jede Linearkombination der oben studierten Lösungen ist eine Lösung. Das gilt nicht nur für endliche Linearkombinationen, sondern auch für manche unendliche Überlagerungen

$$u(t,x) = \sum_{-\infty}^{+\infty}(ic_n)\sin\left(\frac{n\pi x}{L}\right) \cdot \ \exp\left(\frac{in\pi ct}{L}\right) \quad .$$

Es handelt sich um eine reelle Lösung, wenn

$$c_{-n} = \bar{c_n} \quad \text{für alle } n \quad .$$

Mit der obigen Rechnung läßt sich das umschreiben

$$\begin{aligned} u(t,x) &= \frac{1}{2}\sum c_n \exp\left(i\frac{n\pi}{L}(x+ct)\right) - \frac{1}{2}\sum c_n \cdot \exp\left(i\frac{n\pi}{L}(-x+ct)\right) \\ &= \frac{1}{2}g(x+ct) - \frac{1}{2}g(-x+ct) \quad . \end{aligned}$$

Bemerke :

a) Die Funktion

$$g(y) = \sum c_n \exp\left(i\frac{n\pi}{L}y\right)$$

ist genau dann reell, wenn $c_{-n} = \bar{c_n}$ für alle n.

b) Zur Zeit 0 sind Auslenkung und Geschwindigkeit

$$\begin{aligned} u(0,x) &= \frac{1}{2}g(x) - \frac{1}{2}g(-x) \\ \frac{1}{c}\frac{\partial}{\partial t}u(0,x) &= \frac{1}{2}g'(x) - \frac{1}{2}g'(-x) \quad . \end{aligned}$$

Wenn die Funktion $g(\cdot)$ ungerade ist, dann verschwindet die Anfangsgeschwindigkeit.
In diesem Falle hat man $u(t,x) = \frac{1}{2}g(x+ct) + \frac{1}{2}g(x-ct)$.

6) Es stellen sich nun zwei Fragen :

I. Wie allgemein ist die durch Überlagerung gewonnene Lösung der Gleichung

$$\begin{aligned} \left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\right) u(t,x) &= 0 \\ u(t,0) = 0 = u(t,L) \quad &\text{für alle } t \quad . \end{aligned}$$

II. Welche Koeffizientenfolgen $(c_n)_n$ kommen in Betracht? Unter welchen Umständen liefert die unendliche Summe tatsächlich eine Lösung?
Diese Fragen haben eine lange und interessante Geschichte.

Historisches zur schwingenden Saite Der berühmte Mathematiker und Enzyklopädist J.B.d'Alembert hat 1747 das folgende Ergebnis publiziert: Wenn man einer elastischen Saite der Länge L, die in den Punkten $x = 0$ und $x = L$ eingespannt ist, in den Punkten x die Auslenkung $g(x)$ gibt und dann losläßt, so wird ihre Auslenkung zu allen Zeiten t gegeben durch

$$u(t,x) = \frac{1}{2}g(x+ct) + \frac{1}{2}g(x-ct) \quad .$$

D'Alembert dachte vermutlich an Funktionen $g(\cdot)$, die durch einen analytischen Ausdruck gegeben sind, welcher die erforderlichen Differentiationen ermöglicht.

L.Euler propagierte zu dieser Zeit schon allgemeinere Auffassungen, was man unter einer Funktion im Sinne der Analysis zu verstehen hat und welche Funktionen $g(\cdot)$ daher in Betracht gezogen werden können. In seinem Lehrbuch von 1755 gibt Euler die folgende Definition:
„Sind nun Größen auf die Art voneinander abhängig, dass keine davon eine Veränderung erfahren kann, ohne zugleich eine Veränderung der anderen zu bewirken, so nennt man diejenige, deren Veränderung man als die Wirkung von der Veränderung der anderen betrachtet, eine Funktion von dieser; eine Benennung, die sich so weit erstreckt, dass sie alle Arten, wie eine Größe durch eine andere bestimmt werden kann, unter sich begreift.“‘
Wenn $g(\cdot)$ eine (in Eulers Sinn beliebige) Funktion ist, die zweimal differenzierbar ist, dann genügt

$$u(t,x) = \frac{1}{2}g(x+ct) + \frac{1}{2}g(x-ct)$$

offenbar der Gleichung $\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\right) u(t,x) = 0$.
D. Bernoulli ging 1753 bei seiner Lösung des Problems der schwingenden Saite von der physikalischen These aus, dass jeder Ton durch Überlagerung von Grund- und Obertönen entsteht. Bernoulli schloß daraus, dass die Auslenkung immer in der Form

$$u(t,x) = \sum_{n=1}^{\infty} c_n \cdot \sin\frac{n\pi x}{L} \cdot \cos\frac{n\pi ct}{L}$$

dargestellt werden kann. Bernoulli meinte, dass man jede mögliche Anfangsauslenkung durch eine Reihe darstellen kann.
Euler, d'Alembert und später auch Lagrange verwarfen die Behauptung Bernoullis. Die völlige Aufklärung dieser Frage mußte bis 1824 vertagt werden, als Fourier die Zweifel an der Gültigkeit der Darstellung einer „beliebigen“ Funktion durch trigonometrische Reihen beseitigte.
Wir können die Frage hier noch nicht sofort klären. Wir müssen vorher einiges über konvergente Folgen und konvergente Reihen lernen.

Bemerkung : Die Absolutquadrate der Amplituden c_n geben Auskunft, wie sich die Gesamtenergie auf die Obertöne verteilt.

$$E_n = \text{const} \cdot n^2 \cdot |c_n|^2 \quad .$$

Da die Gesamtenergie endlich ist, ist

$$g'(y) = \frac{\pi}{L}\sum c_n(in)\ \exp\left(i\frac{n\pi}{L}y\right)$$

durch eine quadratsummierbare Reihe gegeben.
Die Theorie läßt also Auslenkungen $u(t,x)$ zu, die nicht stetig differenzierbar sind.

17. Vorlesung : Weitere Bemerkungen über Faltung und Fourier-Transformation

1) Jeder Gewichtung $a(\cdot)$ auf $\mathbb{R}$ ist eine charakteristische Funktion zugeordnet

$$A(t) = \int e^{ist} \cdot a(s)ds \quad \text{für } t \in \mathbb{R} \quad .$$

Man kann zeigen, dass $A(\cdot)$ stetig ist mit

$$A(t) \longrightarrow 0 \quad \text{für } t \longrightarrow \pm\infty \quad \text{(Lemma von Riemann-Lebesgue)}$$

Man möchte gerne sagen, dass die Gewichtung $a(\cdot)$ durch ihre charakteristische Funktion eindeutig bestimmt ist. Die Aussage ist richtig, wenn man $a(\cdot)$ als eine Äquivalenzklasse von Funktionen auffaßt; integrable Funktionen, die sich nur auf einer Lebesgue-Nullmenge unterscheiden, liefern dieselbe charakteristische Funktion. (Wir mußten z.B. bei den Rechtecksdichten nicht auf die Werte in den Sprungstellen achten.)

2) Die Inversionsformel funktioniert (in direkter Weise) nur dann, wenn die charakteristische Funktion $A(t)$ integrabel ist:

$$a(s) = \frac{1}{2\pi} \int e^{-ist} A(t)dt \quad \text{für } s \in \mathbb{R}$$

ist in diesem Fall stetig. Wenn $A(\cdot)$ nicht integrabel ist, dann ist der Weg von $A(\cdot)$ zur Gewichtung technisch komplizierter; man muß mit Limiten argumentieren. **Konvergenz** wird ein Thema im Teil V dieser Veranstaltung sein.

3) Die Inversionsformel impliziert, dass gewisse stetige Funktionen $a(s)$ als eine (kontinuierliche!) **Überlagerung** von (nichtintegrablen!) Funktionen $e_t(\cdot) = e^{it\cdot}$ dargestellt werden können. Es sind diejenigen stetigen $a(\cdot)$, die eine integrable charakteristische Funktion besitzen. Dies ist eine Regularitätsbedingung, die nicht leicht auf andere Art zu fassen ist.

4) Sei $a(\cdot)$ eine Gewichtung mit integrabler charakteristischer Funktion $A(t)$ und der Eigenschaft, dass auch die Ableitung $a'(\cdot)$ eine integrable charakteristische Funktion $B(t)$ besitzt. Es gilt dann $i \cdot B(t) = t \cdot A(t)$. Dies erweist sich durch partielle Integration

$$B(t) = \int e^{ist} \cdot a'(s)ds = e^{ist} \cdot a(s) \Big|_{-\infty}^{+\infty} - it \int e^{ist} a(s)ds = -i \cdot t \cdot A(t) \quad .$$

Die **Differentiation** $\frac{1}{i}\frac{\partial}{\partial s}$ spiegelt sich also bei den charakteristischen Funktion als eine Multiplikation mit der Variablen t. – Daraus wird ein interessantes Thema, wenn man genauer studiert, welche Funktionen $a(s), A(t), a'(s), B(t)$ hier in Betracht kommen. (Die Physiker erfahren davon, wenn es in der Quantenmechanik um den Impulsoperator geht.)

5) Die Gewichtung $a(\cdot)$ ist durch den Filter

$$\mathrm{TLF}_a(\cdot) \; : \quad f(\cdot) \longmapsto \int a(s) f(\cdot - s) ds$$

eindeutig bestimmt. Dabei muß man den Definitionsbereich von $\mathrm{TLF}_a(\cdot)$ nicht unbedingt besonders groß wählen. (Für alle beschränkten meßbaren $f(\cdot)$ ist das Integral wohldefiniert.) Es genügt $\mathrm{TLF}_a(f)$ für recht spezielle f (bis auf eine Lebesgue-Nullfunktion genau) zu kennen, um die Gewichtung $a(\cdot)$ zu identifizieren. Es genügt z.B. die Familie $\{e_t(\cdot) \; : t \in \mathbb{R}\}$; denn es gilt $\mathrm{TLF}_a\,(e_t) = A(t)$, und die charakteristische Funktion $A(\cdot)$ identifiziert in jedem Fall die Gewichtung $a(\cdot)$, wie wir bereits oben gesagt (aber nicht bewiesen) haben.

6) Aus der charakteristischen Funktion $A(\cdot)$ ergibt sich besonders leicht das „Antwortssignal“ $\mathrm{TLF}_a(b)$ für diejenigen „Eingangssignale“ $b(s)$, welche integrabel sind und noch dazu eine integrable charakteristische Funktion $B(t)$ besitzen

$$B(t) = \int e^{ist} b(s) ds\,, \quad b(s) = \frac{1}{2\pi} \int e^{-ist} B(t) dt \quad .$$

Für diese b gilt nämlich

$$\mathrm{TLF}_a(b) = a * b = \frac{1}{2\pi} \int e^{-ist} \cdot B(t) \cdot A(t) dt \quad \text{für alle}\, s \quad .$$

Viele interessante „Eingangssignale“ b sind von dieser Gestalt. Wir haben im Beispiel (7a) aber auch andere Eingangssignale kennengelert, solche, für welche $B(t) \cdot A(t)$ nicht notwendigerweise integrabel ist. Wir haben damals nicht darauf hingewiesen,dass die Inversionsformel nicht direkt anwendbar ist. Man braucht spezielle Limiten von Integralen, die sog. uneigentlichen Integrale, wenn man die Rechnung exakt machen will.

7) **Hinweis :** Die numerische Berechnung eines Faltungsprodukts $a * b$ ist manchmal recht aufwendig. Für die approximative numerische Berechnung von Fourier-Integralen gibt es aber effiziente Verfahren (Stichwort „Fast Fourier Transform“ FFT). Es wird daher empfohlen, $a * b$ als die inverse Fourier-Transformierte des Produkts $A(t) \cdot B(t)$ zu berechnen.

$$a * b(s) = \frac{1}{2\pi} \int e^{-ist} B(t) \cdot A(t) dt \quad .$$

Diese Empfehlung gilt übrigens auch bei der diskreten Faltung. Betrachten wir die Multiplikation zweier Zahlen, die in Dezimalschreibweise gegeben sind :

$$\begin{array}{rcl} a & = & a_n \cdot 10^n + a_{n-1} \cdot 10^{n-1} + \ldots + a_0 + a_{-1} \cdot 10^{-1} + a_{-2} \cdot 10^{-2} + \ldots \\ b & = & b_m \cdot 10^m + \ldots + b_0 + b_{-1} \cdot 10^{-1} + b_2 \cdot 10^{-2} + \ldots \\ \text{mit} & & a_j, b_j \in \{0, 1, 2, \ldots, 9\} \quad . \end{array}$$

Für das Produkt $a \cdot b$ bekommen wir die Darstellung

$$c = \sum c_n \cdot 10^n \quad \text{mit} \quad c_n = \sum_n a_k \cdot b_{b-k} \quad .$$

Allerdings sind die c_n i.Allg. keine Ziffern $\in \{0, 1, 2, \ldots, 9\}$. Die Dezimaldarstellung des Produkts erfordert noch den „Übertrag".
Man kann also sagen, dass die „schriftliche Multiplikation", die man auf der Schule lernt, mit Faltung zu tun hat. Wir überlassen es den Algorithmikern, aufzuzeigen, warum die schnelle Fourier-Transformation einen effektiveren Weg zur praktischen Durchführung von Multiplikation weist.

8) Wir haben in Beispielen gesehen, wie man gewissen Funktionen $\big($wie $a(s)\big)$ andere Funktionen $\big($wie z.B. $A(-\omega)\big)$ zuordnet, und wie man wieder zurückfindet. Wir bewegten uns in der Nähe der sog. **„Fourier-Transformation"**. Die Notationen sind in der Theorie der Fourier-Transformation leider nicht einheitlich. Insoweit man die Transformationen

$$\mathfrak{F} \; : \; f \longmapsto \hat{f} \; ; \quad \mathfrak{F}^{-1} \; : \hat{f} \longmapsto f$$

durch Integrale darstellen kann, haben sie eine sehr ähnliche Gestalt. Dies hat verschiedene Bemühungen nach Angleichung hervorgerufen. Manche Autoren definieren

$$\begin{array}{rccccl} \text{I)} & \hat{f} & = & \mathfrak{F}f(\cdot) & = & \int f(y)e^{-2\pi i\gamma y}dy \quad \text{(Funktion von } \gamma) \\ & f & = & \mathfrak{F}^{-1}\hat{f}(\cdot) & = & \int \hat{f}(\gamma)e^{2\pi i\gamma y}d\gamma \quad \text{(Funktion von } y) \end{array}$$

(z.B. H.Dym und H.P. McKean: *Fourier Series and Integrals*).
Andere Autoren definieren die Fourier-Transformation folgendermaßen :

$$\begin{array}{rccccl} \text{II)} & \hat{f} & = & \mathfrak{F}f(\cdot) & = & \frac{1}{\sqrt{2\pi}}\int f(y)e^{-i\omega y}dy \quad \text{(Funktion von } \omega) \\ & f & = & \mathfrak{F}^{-1}\hat{f}(\cdot) & = & \frac{1}{\sqrt{2\pi}}\int \hat{f}(\omega)e^{i\omega y}d\omega \quad \text{(Funktion von } y) \quad . \end{array}$$

Die Konvention bei MAPLE ist nochmals anders. Man muß also aufpassen.

Wir haben hier den Terminus Fourier-Transformation ganz vermieden; mit unserer Bezeichnung „charakteristische Funktion" stehen wir in der Tradition der Wahrscheinlichkeitstheoretiker. Die Wahrscheinlichkeitstheoretiker betrachten allerdings die charakteristischen Funktionen von integrablen Gewichtungen $a(s)ds$ als einen Spezialfall der charakteristischen Funktionen von Maßen (oder signierten Maßen).

III) $\quad t \longmapsto \int e^{itx}d\mu(x)$.

Diese Funktionen streben nicht notwendigerweise nach 0 für $t \longrightarrow \pm\infty$. Besonders interessant sind charakteristische Funktionen für positive Maße $\mu(\cdot)$; sie heißen positiv semidefinite Funktionen. (Der entscheidende Satz heißt der Satz von Herglotz-Bochner.)

Wir bemerken : Inversionsformeln, die mit Limiten von Integralen arbeiten, sind außer Mode gekommen. „Uneigentliche" Integrale passen nicht in die moderne Integrationstheorie; sie überleben als Mittel der heuristischen Analogiebetrachtung.

9) (Faltung über $\mathbb{R}/2\pi$).

Seien $f(\cdot)$ und $g(\cdot)$ 2π-periodische Funktionen mit

$$\|f\|_1 \; := \int_{-\pi}^{+\pi} |f(s)|ds < \infty \, , \quad \|g\|_1 = \int_{-\pi}^{+\pi} |g(s)|ds < \infty \quad .$$

Man nennt $f(\cdot)$ und $g(\cdot)$ Gewichtungen über $\mathbb{R}/2\pi$. Man definiert für solche Gewichtung das „Faltungsprodukt über $\mathbb{R}/2\pi$“

$$h(t) = \frac{1}{2\pi} \int f(s) \cdot g(t-s) ds \quad \text{für } \mathbb{R}/2\pi,$$

wo die Integration über ein beliebiges Intervall der Länge 2π zu erstrecken ist. Es ist klar, was unter einer Wahrscheinlichkeitsgewichtung auf $\mathbb{R}/2\pi$ zu verstehen ist.
Interessante komplexe Gewichtungen sind die 2π-periodischen Funktion $e_n(s) = e^{ins}$, $n \in \mathbb{Z}$.
Ihre Faltungsprodukte sind leicht auszurechnen

$$e_n * e_m = \begin{cases} 0 & \text{falls } n \neq m \\ e_n & \text{falls } n = m \end{cases}$$

Weitere interessante komplexe Gewichtungen auf $\mathbb{R}/2\pi$ sind die trigonometrischen Polynome

$$f(\cdot) = \sum_{-N}^{+N} a_n \cdot e_n(\cdot) \quad .$$

Falten mit $e_m(\cdot)$ ergibt

$$f * e_m = a_m \cdot e_m(\cdot) \ , \quad a_m = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(s) \cdot e^{-ims} ds$$

Das Falten trigonometrischer Polynome spiegelt sich in der punktweise Multiplikation der Koeffizientenfolgen

$$\Big(\sum a_n e_n\Big) * \Big(\sum b_m e_m\Big) = \sum (a_n \cdot b_n)\, e_n \quad .$$

10) Zu jeder Gewichtung $f(\cdot)$ auf $\mathbb{R}/2\pi$ kann man die Folge der **Fourier-Koeffizienten** definieren

$$a_n := \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(s) \cdot e^{-ins} ds \quad \text{für } n \in \mathbb{Z} \quad .$$

Das trigonometrische Polynom

$$f_N(s) = \sum_{-N}^{+N} a_n e^{ins}$$

nennt man das **approximierende Fourier-Polynom** der Ordnung N für $f(\cdot)$.
Man betrachtet auch die sogenannten Césaro-Mittel der approximierenden Polynome

$$F_N(s) = \frac{1}{N}\big(f_0(\cdot) + f_1(\cdot) + \ldots + f_{N-1}(\cdot)\big) \quad .$$

Satz :

$f_N(\cdot)$ entsteht aus $f(\cdot)$ durch Faltung mit dem Dirichlet-Kern; $F_N(\cdot)$ entsteht durch Faltung mit dem Fejér-Kern.

Beweis

Mit den Dirichlet- und Fejér-Kernen haben wir zahlreiche MAPLE-Übungen gerechnet.

1) Zur Erinnerung :

$$D_N(s) = \sum_{-N}^{+N} e^{ins}\,, \quad F_N(\cdot) = \frac{1}{N}\big(D_0(\cdot) + \ldots + D_{N-1}(\cdot)\big) \quad .$$

2) $(f * D_N)(t) = \frac{1}{2\pi}\int f(s)\cdot D_N(t-s)ds =$
$= \sum_{-N}^{+N} \frac{1}{2\pi}\int f(s)\cdot e^{in(t-s)}ds = \sum e^{int}\cdot \int \frac{1}{2\pi} f(s)e^{-ins}ds \quad .$

Bemerke : Der Fejér-Kern ist positiv mit Gesamtintegral 2π über eine volle Periode. Die Faltung mit dem Fejér-Kern ist also eine Regularisierung der Gewichtung $f(\cdot)$.

11) Das Lemma von Riemann-Lebesgue garantiert, dass die Folge der Fourier-Koeffizienten eine Nullfolge ist. Man kann i.Allg. nicht erwarten, dass die Folge absolutsummabel ist.

Beispiel (Eulers Sägezahnfunktion)
$E(\cdot)$ sei die 2π-periodische Funktion mit

$$E(s) = \begin{cases} -\pi - s & \text{für } s \in (-\pi, 0) \\ \pi - s & \text{für } s \in (\pi, 0) \end{cases}$$

Die Fourier-Koeffizienten sind

$$a_n = \frac{1}{2\pi}\int E(s)e^{-ins}ds = -\frac{i}{n} \quad \text{für } n \neq 0 \quad .$$

Das approximierende Fourier-Polynom der Ordnung N ist

$$E_N(s) = \sum_{-N}^{+N} a_n e^{ins} = \sum_{1}^{N} \frac{2}{n}\sin(ns) \quad .$$

Die Folge der Fourier-Koeffizienten ist nicht summabel.

Bemerke : $E_N(\cdot)$ ist die Stammfunktion von

$$D_N(s) - 1 = 2\cdot \sum_{1}^{N} \cos(ns) \quad .$$

Nach unserem Satz entsteht

$$\frac{1}{N}\left(E_0(s) + \ldots + E_{N-1}(s)\right)$$

durch Regularisierung von $E(\cdot)$ mit dem Fejér-Kern der Ordnung N. Wie dies aussieht, sahen wir in den Aufgaben 17 und 21.

Ein ähnliches Beispiel ist die (2π-periodische fortgesetzte) Signumfunktion

$$\begin{aligned}\frac{4}{\pi}\cdot\big(E(s)+E(\pi-s)\big) &= \begin{cases} 1 & \text{für } s\in(0,\pi) \\ -1 & \text{für } s\in(-\pi,0)\end{cases} \\ &= \frac{4}{\pi}\left(\sin s + \frac{1}{3}\sin 3s + \frac{1}{5}\sin 5s + \ldots\right) \quad .\end{aligned}$$

Dieses Beispiel eignet sich besonders gut für die Beobachtung des sog. Gibbs'schen Phänomens. Die Konvergenz hat merkwürdige Züge: Im konkreten Fall bemerkt man, dass das Supremum der approximierenden Fourier-Polynome nicht nach 1 konvergiert.

12) Für jede **stetige** 2π-periodische Funktion $f(\cdot)$ strebt $(f * F_N)(\cdot)$ punktweise gegen $f(\cdot)$. Wenn $f(\cdot)$ stückweise stetig ist mit einer Sprungstelle in $\tilde{s}$, dann konvergieren die regularisierten Funktionen $(f * F_N)$ in diesem Punkt gegen den Mittelwert zwischen dem rechtsseitigen und dem linksseitigen Limes. Dies ist eine gute Art der punktweise Konvergenz.
Entsprechendes kann man bei den approximierenden Fourier-Polynomen $(f * D_N)(\cdot)$ nicht erwarten. Der Dirichlet-Kern ist zwar symmetrisch; die Faltung mit $D_N(\cdot)$ ist aber keine Regularisierung, weil $D_N(\cdot)$ nicht positiv ist. In der Tat gibt es stetige 2π-periodische Funktionen, für welche die Folge der Fourier-Polynome nicht in allen Punkten konvergiert. Nach einem berühmten Satz von Carleson (1966) hat man aber immerhin Konvergenz fast überall, und zwar nicht nur für stetige $f(\cdot)$, sondern auch für alle p-integrablen $f(\cdot)$, d.h.

$$\int_{-\pi}^{+\pi} |f(s)|^p ds < \infty \quad \text{für ein } \; p>1 \quad .$$

13) Jeder integrablen komplexen Gewichtung auf $\mathbb{R}/2\pi$ kann man die Folge der Fourier-Koeffizienten $(a_n)_n$ zuordnen. Man kann zeigen, dass diese Folge die Gewichtung eindeutig bestimmt.

$$\sum_{-\infty}^{+\infty} a_n e^{int}$$

heißt die formale Fourier-Reihe zu $f(\cdot)$. Es ist i.Allg. unklar, was diese unendliche Summe bedeuten könnte.

14) Im Falle $\sum |a_n| < \infty$ ist die Sache einfach: Die Reihe ist absolutsummabel und sie liefert die stetige Version der gegebenen Funktion $f(\cdot)$. Um das einzusehen, muss man wissen,dass eine 2π-periodische Funktion, deren Fourier-Koeffizienten allesamt verschwinden, fast überall verschwindet. Wir haben also in diesem Falle

$$f(t) = \sum a_n e^{int} \quad \text{fast überall .}$$

Dies kann man als Inversionsformel verstehen für die Zuordnung

$$f(\cdot) \longrightarrow (a_n)_n \quad .$$

Sie erfaßt aber nur einen Ausschnitt der Theorie der Fourier-Reihen, weil sie nur von denjenigen $f(\cdot)$ handelt, deren Fourier-Koeffizienten absolut summabel sind. Den Fall, wo die Fourier-Koeffizienten lediglich quadratisch summabel sind, behandeln wir später.

Bemerkung : Die Folge der Fourier-Koeffizienten ist jedenfalls dann absolutsummabel, wenn die Gewichtung $f(\cdot)$ absolutstetig ist und die Ableitung quadratisch integrierbar ist. Diese Situation hatten wir bei der schwingenden Saite.

MATHEMATIK FÜR PHYSIKER I

Prof. Dr. H. Dinges WS 2001/02

Teil IV beginnt in der 20ten Vorlesung, gerade noch vor den Weihnachtsfereien. Die Vorlesungen 18 und 19 waren der Wiederholung und der Klausurvorbereitung gewidmet.

Teil IV. Metrik, Norm, Konvexität

Themenübersicht

Teil IV. Metrik, Norm, Konvexität

20. Vorlesung : Dreiecksungleichung und Subadditivität

Definition

Eine abstrakte Punktmenge S wird dadurch zu einem **metrischen Raum**, dass man eine positive Funktion $d(\cdot,\cdot)$ auf $S\times S$ auszeichnet, welche den Forderungen an eine Metrik genügt.

Definition (Metrik)

Eine Funktion $d(\cdot,\cdot)$ heißt eine **Semi-Metrik**, wenn gilt

(i) $d(P,Q)\geq$ für alle P,Q und $d(P,P)=0$ für alle P.

(ii) $d(P,Q)=d(Q,P)$

(iii) $d(P,R)\leq d(P,Q)+d(Q,R)$ („Dreiecksungleichung“)

Die Semi-Metrik ist eine Metrik, wenn gilt

(iv) $d(P,Q)=0\Longrightarrow P=Q$.

Beispiele :

1) Die reelle Achse wird im Schulunterricht von vorneherein als metrischer Raum eingeführt.

2) Die komplexe Ebene wird meistens mit der Metrik $d(z,w) = |w-z| = \sqrt{(w-z)(\bar{w}-\bar{z})}$ ausgestattet. Für manche Zwecke ist aber der Abstand auf der Riemann'schen Zahlkugel vorteilhafter.

3) Der Anschauungsraum wird meistens mit der euklidischen Metrik ausgestattet. Wir werden aber auch andere Metriken schätzen lernen.

4) Man stelle sich ein Straßennetz vor, wo jeder Straße k eine nichtnegative Zahl $e(k)$ zugeordnet ist. In der Fachsprache: S sei die Scheitelmenge eines ungerichteten Graphen; jeder Kante k sei eine Zahl $e(k)\geq 0$ zugeordnet. Nehmen wir an, dass der Graph zusammenhängend ist. Für jeden Weg von P nach Q betrachten wir die Summe der $e(\cdot)$-Werte auf den Kanten. $d(P,Q)$ sei das Infimum. $d(\cdot,\cdot)$ ist dann eine Semi-Metrik auf S. Wenn es keine Straßen gibt, die ohne Kosten zu begehen sind, dann ist $d(\cdot,\cdot)$ eine Metrik.

5) Sei $S=\{0,1\}^n$ die Menge aller Null-Eins-Folgen der Längen. Eine beliebte Metrik ist die Hamming-Distanz. $d(P,Q)$ gibt an, in wievielen Punkten die beiden Null-Eins-Folgen verschieden sind. Man deutet S als die Menge der Ecken eines n-dimensionalen Würfels, dessen Kanten k mit $e(k)=1$ belegt sind. Der Hamming-Abstand ist dann die Metrik im Sinne des Beispiels 4.

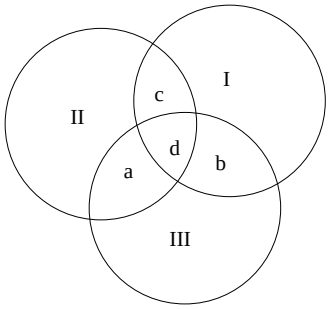
Gewisse Sprechweisen, die für den Anschauungsraum geläufig sind, kann man mit Vorteil auf allgemeine metrische Räume übertragen.

Bemerkungen zur Hamming-Distanz

Jeder Punkt $x^* \in \{0,1\}^n$ hat genau n unmittelbare Nachbarn und $\binom{n}{2}$ Nachbarn im Abstand $= 2$. Man sagt, die abgeschlossene Kugel $B\,(x^*, \leq 1)$ mit dem Mittelpunkt x^* und dem Radius 1 enthält genau $n+1$ Punkte und die Anzahl der Punkte in der Kugel mit dem Radius 2 ist

$$\left|B\,(x^*, \leq 2)\,\right| = 1 + n + \binom{n}{2} \quad .$$

Ein merkwürdiger Raum ist der Raum $S = \{0,1\}^7$. In diesem Raum mit $2^7 = 2^4 \cdot 2^3$ Punkten gibt es $2^4 = 16$ paarweise disjunkte Kugeln mit dem Radius 1. Man benützt dieses Faktum zur Konstruktion eines „fehlerkorrigierenden“ Code. Fehlerentdeckende und fehlerkorrigierende Code dienen dazu, die Übertragung von Bit-Folgen durch gestörte Kanäle sicherer zu machen. Man sendet z.B. statt jedes Quadrupels von Bits ein besonderes 7-Tupel, wo die zusätzlichen drei Bits als Kontrollbits fungieren. Das folgende Diagramm veranschaulicht einen beliebten Code dieser Art (Er heißt Hamming's (7,4)-Code). Der Empfänger ist in der Lage, das gemeinte Quadrupel abcd auch dann zu rekonstruieren, wenn die Übertragung (höchstens) eines der 7 gesendeten Bits abcd I II III verändert hat. Das Diagramm stellt das „Codebuch“ auf übersichtliche Weise dar.



Die Kontrollbits I, II und III werden so gewählt, daß die Summe in jedem Kreis gerade ist. Beispielsweise wird das (a b c d) = (1 0 0 1)-Quadrupel in das 7-Tupel (a b c d, I, II, III) = 1 0 0 1 1 0 0 kodiert.
Wenn nun der Empfänger ein 7-Tupel empfängt, welches der Forderung nicht genügt, dann kann er es durch Abänderung in genau einer Position passend machen.
Zum Beweis müssen wir Fälle unterscheiden: Wenn ein empfangenes 7-Tupel alle drei Gleichungen verletzt, dann ist d zu korrigieren. Wenn genau eine der Gleichungen verletzt ist, dann ist das entsprechende Kontrollbit zu korrigieren. Wenn genau zwei Gleichungen verletzt sind, ist a bzw. b bzw. c zu korrigieren. Das Decodieren ist also genauso einfach wie das Codieren.

Bemerke : Wenn mit allzugroßer Wahrscheinlichkeit mehr als eines der sieben Bits falsch übertragen wird, dann ist Hamming's (7,4)-Code nicht anwendbar.

Satz

Die Summe zweier Semi-Metriken auf S ist eine Semi-Metrik.

Der Beweis ist trivial. Der Satz ist der erste in einer Reihe von Sätzen über die Konstruktion von Semi-Metriken aus Semi-Metriken.

Beispiele

1) Auf dem Raum $\mathbb{R}^2$ aller reellen 2-Spalten ist

$$d_1(x,y) = d_1\big(\begin{pmatrix}x_1\\x_2\end{pmatrix},\begin{pmatrix}y_1\\y_2\end{pmatrix}\big) := |y_1 - x_1| + |y_2 - x_2|$$

eine Metrik. Sie heißt die ℓ_1-Metrik.

2) Wir werden uns auch mit der ℓ_2-Metrik befassen

$$d_2(x,y) = d_2\big(\begin{pmatrix}x_1\\x_2\end{pmatrix},\begin{pmatrix}y_1\\y_2\end{pmatrix}\big) := \sqrt{|y_1 - x_1|^2 + |y_2 - x_2|^2} \quad .$$

Sie wird auch die euklidische Metrik genannt.
Die Art und Weise, wie sie aus den Semi-Metriken

$$|y_1 - x_1| \text{ und } |y_2 - x_2| \qquad (\text{auf dem } \mathbb{R}^2)$$

entsteht, wird uns beschäftigen.

3) Für $p \geq 1$ definiert man die ℓ_p-Metrik

$$d_p\big(\begin{pmatrix}x_1\\x_2\end{pmatrix},\begin{pmatrix}y_1\\y_2\end{pmatrix}\big) = \big(|y_1 - x_1|^p + |y_2 - x_2|^p\big)^{1/p} \quad .$$

Es ist nicht trivial, daß diese Funktion $d_p(\cdot,\cdot)$ die Dreiecksungleichung erfüllt. Wir werden dazu einiges zu sagen haben; Stichworte sind die Hölder'sche Ungleichung und die Minkowskische Ungleichung .

Satz

Wenn $d(\cdot,\cdot)$ eine Metrik ist, dann sind auch

$$\frac{d(\cdot,\cdot)}{1+d(\cdot,\cdot)} \quad \text{und} \quad d(\cdot,\cdot) \wedge 1 \quad \text{Metriken.}$$

Beweis :

Sei $d(P,Q) = s$, $d(Q,R) = t$. Wir wissen $d(P,R) \leq s + t$. Wir müssen zeigen

$$\frac{d(P,R)}{1+d(P,R)} \leq \frac{s}{1+s} + \frac{t}{1+t}.$$

Die Funktion $x \longmapsto \frac{x}{1+x}$ ist isoton.
Es genügt daher zu zeigen

$$\frac{s+t}{1+s+t} \leq \frac{s}{1+s} + \frac{t}{1+t} \qquad \text{für alle } s,t \in \mathbb{R}_+ \quad .$$

Dies ist aber trivial; denn

$$\frac{s}{1+s+t} \leq \frac{s}{1+s}\,;\ \frac{t}{1+s+t} \leq \frac{t}{1+t} \qquad \text{für alle } s,t \in \mathbb{R}_+ \quad .$$

Die Behauptung, dass $d(\cdot,\cdot) \wedge 1$ eine Metrik ist, lassen wir als Übung.

Satz

Sei $d(\cdot,\cdot)$ eine Semi-Metrik und $F(\cdot)$ auf $\mathbb{R}_+$ eine Funktion mit

(i) $F(r) \geq 0$, für alle r , $\quad F(0) = 0$

(ii) $r \leq s \Longrightarrow F(r) \leq F(s)$ (Monotonie)

(iii) $F(r+s) \leq F(r) + F(s)$ (Subadditivität)

Dann ist $F\big(d(\cdot,\cdot)\big)$ eine Semimetrik.

Beweis :

Wir müssen die Dreiecksungleichung nachweisen.
Offenbar folgt aus (ii) und (iii)

$$F\big(d(P,R)\big) \leq F\big(d(P,W) + d(Q,R)\big) \leq F\big(d(P,Q)\big) + F\big(d(Q,R)\big)\,.$$

Beispiele :

Wir haben oben bereits zwei monotone subadditive Funktionen $F(\cdot)$ kennen gelernt, nämlich $F(r) = \frac{r}{1+r}$, $F(r) = \min\{r, 1\}$.

Satz

Seien $d_1(\cdot,\cdot), \ldots, d_n(\cdot,\cdot)$ Semimetriken auf S und $F(\cdot)$ eine monotone subadditive Funktion auf $\mathbb{R}^n_+$. Dann ist

$$e(\cdot,\cdot) = F\big(d_1(\cdot,\cdot), \ldots, d_n(\cdot,\cdot)\big)$$

eine Semimetrik.
Der Beweis ist derselbe.

Beispiel

$$F\left(s_1, \ldots, s_n\right) = \sqrt{s_1^2 + \ldots + s_n^2} \text{ auf } \mathbb{R}^n_+$$

ist offenbar positiv und monoton. $F(\cdot)$ ist außerdem subadditiv; denn

$$F^2(s,t) = \sum \left(s_j + t_j\right)^2 =$$

$$= \sum s_j^2 + \sum t^2 + 2\sum s_j \cdot t_j \leq \sum s_j^2 + \sum t_j^2 + 2 \cdot \sqrt{\sum s_j^2} \cdot \sqrt{\sum t_j^2} =$$

$$= \big(F(s) + F(t)\big)^2 \quad .$$

Wir werden zeigen, dass auch für beliebiges $p \geq 1$ die Funktion

$$F\left(r_1, \ldots, r_n\right) = \big(\sum |r_j|^p\big)^{1/p}$$

die Voraussetzungen erfüllt. Diese Aussage heißt die Minkowskische Ungleichung; wir werden sie im Kapitel über die p-Normen bewiesen.

Definition

Ein Vektorraum V wird dadurch zu einem **normierten Vektorraum**, dass man eine Funktion auf V auszeichnet, welche den Forderungen an eine Norm genügt. (Wir denken an $\mathbb{C}$- oder $\mathbb{R}$-Vektorräume.)

Definition (Seminorm)
Eine Funktion $p(\cdot)$ auf einem Vektorraum V heißt eine **Seminorm**, wenn gilt

(i) $p(0) = 0,\ p(v) \geq 0$ für alle v

(ii) $p(\alpha \cdot v) = |\alpha| \cdot p(v)$ für alle Skalare α

(iii) $p\,(v_1 + v_2) \leq p\,(v_1) + p\,(v_2)$

Man nennt $p(\cdot)$ eine Norm, wenn zusätzlich gilt

(iv) $p(v) = 0 \Longrightarrow v = 0$.

Hinweis : Normen benützt man zur Konstruktion spezieller Metriken auf affinen Räumen. Der Abstand $d(P, Q)$ ist die Norm des Verschiebungsvektors $\overrightarrow{PQ}$. Die Eigenschaften der Norm garantieren, dass $d(\cdot, \cdot)$ die Dreiecksungleichung erfüllt.

21. Vorlesung : Affine Räume, konvexe Funktionen

Die Punkte des Anschauungsraums nannte man früher manchmal Ortsvektoren. Die Bezeichnung ist deswegen unpassend, weil man „Ortsvektoren“ nicht wie wirkliche Vektoren (d.h. Elemente eines Vektorraums) addieren und mit Skalaren multiplizieren kann. Wirkliche Vektoren sind die „Verschiebungsvektoren“ oder „Translationen“, im Schulunterricht manchmal Pfeilklassen genannt. Zu jedem Punktepaar P, Q gibt es genau eine Translation des Anschauungsraums, welche P in Q überführt, man bezeichnet sie mit $\overrightarrow{PQ}$. Translationen kann man hintereinanderschalten (und rückgängig machen). Die Gesamtheit aller Translationen ist eine kommutative Gruppe. Translationen kann man mit einem reellen Faktor strecken. Die Gesamtheit aller Translationen ist ein Vektorraum.
Angeleitet von diesen Vorstellungen über den Anschauungsraum definiert man

Definition

Ein (reeller) affiner Raum ist eine Punktmenge S zusammen mit einem (reellen) Vektorraum V, welcher einfach transitiv auf S wirkt. Wenn der Vektorraum die Dimension n hat, sagt man, S sei ein n-dimensionaler (reeller) affiner Raum.

Zur Erinnerung: Wir hatten es schon einmal mit einer Menge zu tun, auf welcher eine Gruppe (allerdings kein Vektorraum) „transitiv“ (allerdings nicht „einfach transitiv“) wirkt. Wir hatten festgestellt: Die Gruppe G derjenigen Drehungen, die den Einheitswürfel in sich überführen, wirkt transitiv auf die sechspunktige Menge Ω der Flächen; jedes Element von Ω kann nämlich in jedes andere Element durch ein geeignetes $\varphi_g(\cdot)$ übergeführt werden.

$$\forall \omega', \omega'' \in \Omega \; \exists g \in G \; : \varphi_g(\omega') = \omega'' \, .$$

Diese Eigenschaft von $\{\varphi_g(\cdot) \; : g \in G\}$ wird Transitivität genannt.

In unserem Fall gibt es mehrere g, welche ω' in ω'' überführen, insbesondere gibt es mehrere g, welche ω' in sich überführen. Die Gesamtheit dieser g heißt die Fixgruppe von ω' $\big($bzgl. der Transformationsgruppe $\{\varphi_g(\cdot) \; : g \in G\}\big)$.
Im Falle der Translationen eines affinen Raums besteht die Fixgruppe eines beliebigen Punkts nur aus dem Nullvektor. Das macht einen wesentlichen Unterschied zu dem Beispiel der Drehungen des Würfels . Eine Gruppenaktion $\big\{\varphi_g(\cdot) \; : g \in G\big\}$ heißt einfach transitiv, wenn die Fixgruppe eines beliebigen Punkts nur aus der Gruppeneins besteht.
(Im Falle kommutativer Gruppen schreibt man die Verknüpfung meistens additiv und man nennt das neutrale Element das Nullelement.)

Konstruktion affiner Teilräume

Zu jedem Punktepaar $P_0, P_1 \, (P_0 \neq P_1)$ in einem affinen Raum kann man die Verbindungsgerade konstruieren; sie ist die Menge der Punkte von der Gestalt

$$P = P_0 + \lambda \cdot \overrightarrow{P_0 P_1} = P_0 + \lambda \, (P_0 - P_1) = (1 - \lambda) P_0 + \lambda \cdot P_1 \quad (\lambda \in \mathbb{R}) \, .$$

Ist P_2 ein Punkt, der nicht auf dieser Geraden liegt, dann gewinnt man die Ebene durch die Punkte P_0, P_1, P_2 als die Gesamtheit aller Punkte von der Gestalt

$$Q = P_0 + \lambda \cdot \overrightarrow{P_0 P_1} + \mu \cdot \overrightarrow{P_0 P_1} = (1 - \lambda - \mu) P_0 + \lambda \cdot P_1 + \mu \cdot P_2 \, .$$

Diese Ebene ist offenbar ein zweidimensionaler affiner Raum. Der Raum der Verschiebungen ist der Vektorraum, der von $\overrightarrow{P_0P_1}$ und $\overrightarrow{P_0P_2}$ aufgespannt wird. Man kann denselben Teilvektorraum natürlich auch mit anderen Paaren von Vektoren aufspannen, z.B. mit $\overrightarrow{P_1P_0}$ und $\overrightarrow{P_1P_2}$.

Satz

Seien $P_0, P_1, \ldots, P_m$ Punkte in einem affinen Raum und M die Menge aller Punkte der Gestalt

$$Q = \lambda_0 P_0 + \lambda_1 P_1 + \ldots + \lambda_m P_m \text{ mit } \sum \lambda_j = 1\,.$$

Dann ist M der kleinste affine Teilraum, welcher die Punkte P_j enthält. M heißt der von $\{P_j : j = 0, 1, \ldots, m\}$ aufgespannte affine Raum oder auch die **affine Hülle** der Punktmenge $\{P_j : j = 0, \ldots, m\}$. Wenn M die Dimension m hat, sagt man, daß die Punkte P_j sich in allgemeiner Lage befinden; kein P_j liegt in dem von den übrigen Punkten aufgespannten affinen Teilraum.
Der Beweis ist trivial.

Satz : Zu jeder Teilmenge eines affinen Raums gibt es einen kleinsten sie umfassenden affinen Teilraum. (Man nennt ihn die affine Hülle.)

Beweis : Der Durchschnitt aller die gegebene Menge umfassenden affinen Teilräume ist die affine Hülle.

Sprechweise : Ein $(n-1)$-dimensionaler affiner Teilraum eines n-dimensionalen affinen Raums heißt eine Hyperebene.

Hinweis : Mit Hyperebenen und Durchschnitten von Hyperebenen werden wir uns später bei der Theorie der linearen Gleichungssysteme ausführlich befassen.

Bei den bisherigen Konstruktionen war es nicht wichtig, dass der Koeffizientenkörper der Vektorräume der Körper $\mathbb{R}$ ist. Die speziellen Eigenschaften von $\mathbb{R}$ werden erst jetzt wichtig, wenn wir „Verbindungsstrecken“ betrachten.

Definition

Seien P_0, P_1 Punkte in einem reellen affinen Raum. Die Verbindungsstrecke ist die Menge der Punkte von der Form

$$P = (1-\lambda)P_0 + \lambda \cdot P_1 \text{ mit } \lambda \in [0,1]\,.$$

Definition (Konvexe Menge)

Eine Teilmenge K eines affinen Raums heißt eine konvexe Menge, wenn gilt

$$\forall P_0, P_1 \in K\, \forall \lambda \in [0,1] \quad (1-\lambda)P_0 + \lambda P_1 \in K\,.$$

Satz (Konvexe Hülle)

Zu jeder Teilmenge eines affinen Raums gibt es eine kleinste sie umfassende konvexe Menge. (Man nennt sie die konvexe Hülle.)

Beweis : Der Durchschnitt aller die gegebene Menge umfassenden konvexen Mengen ist eine konvexe Obermenge und zwar offenbar die kleinste.

Satz

Seien $P_0, \ldots, P_m$ Punkte in einem affinen Raum und K die Menge aller Punkte der Gestalt

$$P = \lambda_0 P_0 + \lambda_1 P_1 + \ldots + \lambda_m P_m \,, \quad \lambda_j \geq 0 \sum \lambda_j = 1 \,.$$

Dann ist K die konvexe Hülle von $\{P_j \ : j = 0, \ldots, m\}$.

Bemerkung : Wenn die P_j sich in allgemeiner Lage befinden, dann nennt man K ein m-dimensionales Simplex mit den Extremalpunkten P_j. In diesem Falle gibt es zu jedem $P \in K$ genau ein System von Koeffizienten $\lambda_0, \lambda_1, \ldots, \lambda_m, (\lambda_j \geq 0 \sum \lambda_j = 1)$, so dass

$$P = \sum \lambda_j \cdot P_j \,.$$

Man nennt die λ_j die **Schwerpunktskoordinaten** von P.

Definition (Konvexe Funktion)

Eine Funktion $k(\cdot)$ auf einem affinen Raum heißt eine konvexe Funktion, wenn gilt

$$\forall P_0, P_1 \, \forall \lambda \in [0,1] \quad k\big((1-\lambda)P_0 + \lambda P_1\big) \leq (1-\lambda) k\,(P_0) + \lambda \cdot k\,(P_1) \,.$$

Es ist bequem, auch $+\infty$ als möglichen Wert von $k(\cdot)$ zuzulassen. Aus der Definition der Konvexität folgt, dass der Endlichkeitsbereich $\big\{P \ : k(P) < \infty\big\}$ eine konvexe Menge K ist. In älterer Literatur findet man den Begriff einer (endlichwertigen) konvexen Funktion über der konvexen Menge K. Wenn man eine solche Funktion auf den gesamten Raum fortsetzt, indem man sie auf dem Komplement von K gleich $+\infty$ setzt, dann erhält man eine konvexe Funktion in unserem Sinn.

Sprechweisen

Eine Funktion $h(\cdot)$ heißt eine **konkave** Funktion, wenn $-h(\cdot)$ konvex ist.
Eine Funktion $f(\cdot)$, die sowohl konvex als auch konkav ist, heißt eine **affine** Funktion. $f(\cdot)$ ist genau dann eine affine Funktion, wenn gilt

$$\forall P, Q \, \forall \lambda \quad f\big((1-\lambda)P + \lambda Q\big) = (1-\lambda) f(P) + \lambda \cdot f(Q) \,.$$

Hinweis : Es ist etwas unglücklich, wenn man die affinen Funktionen manchmal auch lineare Funktionen nennt. Lineare Funktionen gibt es, streng genommen, nur auf Vektorräumen; es sind diejenigen affinen Funktionen, die im Nullpunkt verschwinden.

Satz : Eine affine Funktion $f(\cdot)$ auf dem affinen Raum S liefert eine lineare Funktion auf dem Vektorraum der Verschiebungen

$$\ell(\overrightarrow{PQ}) = f(Q) - f(P)\,.$$

Durch $\ell(\cdot)$ und den Wert $f(P^*)$ in einem beliebig gewählten Punkt P^* ist $f(\cdot)$ eindeutig bestimmt. Für alle Q gilt nämlich

$$f(Q) = f\left(P^* + \overrightarrow{P^*Q}\right) = f\left(P^*\right) + \ell\left(\overrightarrow{P^*Q}\right)\,.$$

Bemerke :

a) Die Gesamtheit aller affinen Funktionen auf dem 3-dimensionalen Anschauungsraum ist ein vierdimensionaler Vektorraum.

b) Die Menge der konvexen Funktionen (über einem affinen Raum S) ist kein Vektorraum.Man kann konvexe Funktionen addieren und mit positiven Skalaren multiplizieren. Man kann aber nicht subtrahieren. Man sagt, dass die Menge der konvexen Funktionen über S ein konvexer Kegel ist. Dieser Funktionenkegel hat die höchst bemerkenswerte Eigenschaft, dass er gegenüber (punktweiser) Maximumsbildung abgeschlossen ist.

Satz

Zu jeder Funktion $g(\cdot)$ auf S, die mindestens eine konvexe Minorante besitzt, gibt es eine größte konvexe Minorante.

Beweis :

Das punktweise Maximum aller konvexen Minoranten ist konvex. Es ist offenbar die größte konvexe Minorante.

Einen Anschluß an die Schulmathematik liefert der

Satz

Eine reellwertige Funktion über einem offenen Intervall ist genau dann konvex, wenn die rechtsseitige (oder die linksseitige) Ableitung isoton ist.
(Beweis in den Übungen)

Beispiele von konvexen Funktionen über $\mathbb{R}$

1) $k_1(x) = a_0 + a_1 x + a_2 x^2$ mit $a_2 \geq 0$

2) $k_2(x) = \begin{cases} -\ell n x & \text{für} \quad x > 0 \\ +\infty & \text{für} \quad x \leq 0 \end{cases}$

3) $k_3(x) = \begin{cases} x\ell n x & \text{für} \quad x \geq 0 \\ +\infty & \text{für} \quad x < 0 \end{cases}$

4) $k_4(x) = |x| = x^+ + x^-$ $\left(x^+ = \max\{0, x\},\ x^- = \max\{0, -x\}\right)$

5) $k_5(x) = \sum p_j\,|x - x_j|$ mit $p_j \geq 0$, x_j beliebig.

Die Ableitung von $k_5(\cdot)$ ist konstant in jedem Intervall, in welchem kein x_j liegt. Sie springt im Punkt x_j um den Betrag $2p_j$. Man skizziere z.B.

$$k(x) = \frac{1}{2}|x-1| + \frac{1}{2}|x+1| = \max\big\{|x|-1, 0\big\}\,.$$

Hinweis : Das Verhalten von konvexen Funktionen in den Randpunkten des Endlichkeitsbereichs soll uns hier nicht weiter interessieren; wir kommen darauf zurück, wenn wir uns mit unterhalbstetigen konvexen Funktionen befassen.
Wir betrachten lediglich das Beispiel $k_3(\cdot)$.
Die Ableitung existiert in $(0,\infty)$. $k_3'(x) = \ell nx + x\cdot\frac{1}{x}$. Die Konvexität ist gesichert, wenn wir $k_3(0)$ positiv festsetzen. Die Festsetzung $k_3(0) = 0$ macht $k_3(\cdot)$ zu einer unterhalbstetigen Funktion.

Sprechweise

Eine Funktion $F(\cdot)$ auf einem reellen Vektorraum V heißt **positiv homogen**, wenn gilt

$$F(\alpha v) = \alpha \cdot F(v) \text{ für alle } \alpha \geq 0,\, v \in V.$$

Satz

Eine positiv homogene Funktion $F(\cdot)$ ist genau dann konvex, wenn sie subadditiv ist.

Beweis :

1) $F(\cdot)$ sei subadditiv und positiv homogen.
Für $v, w \in V$ und $\lambda \in [0,1]$ gilt dann

$$F\big((1-\lambda)v + \lambda w\big) \leq F\big((1-\lambda)v\big) + F(\lambda w) = (1-\lambda)F(v) + \lambda \cdot F(w)\,.$$

2) $F(\cdot)$ sei konvex und positiv homogen.
Für $v, w \in V$ gilt dann

$$F(v+w) = 2 \cdot F\left(\frac{v+w}{2}\right) \leq 2 \cdot \left(\frac{1}{2}F(v) + \frac{1}{2}F(w)\right) = F(v) + F(w)\,.$$

Corollar : Die Seminormen sind spezielle positiv homogene konvexe Funktionen.

Satz : K sei eine konvexe Menge in einem Vektorraum V; der Nullpunkt sei innerer Punkt. Es existiert dann genau eine positiv homogene konvexe Funktion, welche im Inneren von K kleiner als 1 ist, und im Äusseren ≥ 1 („Distanzfunktion“). Wenn K symmetrisch zum Nullpunkt ist ($x \in K \Longrightarrow -x \in K$), dann ist die Distanzfunktion eine Seminorm.

22. Vorlesung : Die p-Normen, Dualität

Auf dem Vektorraum X der finiten komplexen Folgen $x = (x_j)_{j\in\mathbb{Z}}$ definiert man (für jedes $p \geq 1$)

$$\|x\|_p := \left(\sum |x_j|^p\right)^{1/p}$$

Wir werden sehen, dass $\|\cdot\|_p$ eine Norm auf X ist.
Die Ungleichung

$$\|x+y\|_p \leq \|x\|_p + \|y\|_p \text{ für alle } x, y \in X$$

heißt die Minkowskische Ungleichung.

Hinweis : Es gibt eine kontinuierliche Variante, die man in der Integrationstheorie studiert.
Sei X ein $\mathbb{C}$-Vektorraum von komplexwertigen Funktionen auf einem endlichen Maßraum, sodass

$$\int \|f\|^p < \infty \text{ für alle } f \in X\,.$$

Man definiert dann

$$\|f\|_p = \left(\int |f|^p\right)^{1/p}$$

und man beweist

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p\,.$$

(Minkowskische Ungleichung).
Der Schlüssel zum Beweis der Minkowskischen Ungleichung ist die Hölder'sche Ungleichung, die wir nun bewiesen wollen.
Wir holen weiter aus.
Das geometrische Mittel positiver Zahlen ist immer kleiner als das arithmetische Mittel.

$$\sqrt{a\cdot b} < \frac{1}{2}(a+b), \text{ wenn } a \neq b\,.$$

Dies sieht man sofort, wenn man die Ungleichung quadriert.

$$a\cdot b < \frac{1}{4}\left(a^2+2ab+b^2\right) \;;\; \text{denn } 0 < \frac{1}{4}\left(a^2-2ab+b^2\right)\,.$$

Als Verallgemeinerung beweisen wir

Lemma
Seien $p, q \geq 1$ mit $\frac{1}{p}+\frac{1}{q}=1$. Für alle $a, b > 0$ gilt dann

$$a^{1/p}\cdot b^{1/q} \leq \frac{1}{p}a + \frac{1}{q}b\,.$$

Beweis : Wir logarithmieren und haben dann zu zeigen

$$\frac{1}{p}\ell na + \frac{1}{q}\ell nb \leq \ell n\left(\frac{1}{p}a + \frac{1}{q}b\right) .$$

Da $\ell n(\cdot)$ konkav ist, haben wir

$$\forall a, b > 0\, \forall \lambda \in [0,1] \quad \ell n\big((1-\lambda)a + \lambda b\big) \geq (1-\lambda)\ell na + \lambda \cdot \ell nb\,,$$

was zu beweisen war.

Bemerke : Der Logarithmus ist strikt konkav. Das bedeutet

$$a^{1/p} \cdot b^{1/q} < \frac{1}{p}a + \frac{1}{q}b \text{ wenn } a \neq b\,.$$

Satz (Höldersche Ungleichung)
Seien $p, q \geq 1$ mit $\frac{1}{p} + \frac{1}{q} = 1$, und seien x, y finite Folgen. Sei

$$\|x\|_p = \big(\sum |x_j|^p\big)^{1/p},\ \|y\|_q = \big(\sum |y_j|^q\big)^{1/q}$$

Es gilt

$$\mid \sum x_j y_j \mid \leq \|x\|_p \cdot \|y\|_q\,.$$

Beweis : Es genügt offenbar, den Fall zu untersuchen, wo alle x_j und y_j strikt positiv sind.

a) In diesem Falle betrachten wir

$$a_j = \frac{(x_j)^p}{\sum (x_j)^p}\,, \quad b_j = \frac{(y_j)^q}{\sum (y_j)^q}\,.$$

Beachte $\sum a_j = 1 = \sum b_j$.
Nach dem Lemma haben wir für alle j

$$\frac{1}{p}a_j + \frac{1}{q}b_j \geq a_j^{1/p} \cdot b_j^{1/q} = \frac{x_j}{\|x\|_p} \cdot \frac{y_j}{\|y\|_q}\,.$$

Summation über j liefert die Behauptung.

$$1 \geq \frac{1}{\|x\|_p \cdot \|y\|_q} \cdot \sum |x_j y_j|\ .$$

b) Gleichheit haben wir nur im Falle, wo

$$a_j = b_j \text{ für alle } j\,, \quad \text{d.h. } |x_j|^p = \text{ const } \cdot |y_j|^q \text{ für alle } j\,.$$

Dies formulieren wir als

Satz : Zu jeder komplexen Folge x gibt es ein y, sodass

$$\sum x_j y_j = \|x\|_p \cdot \|x\|_q \,.$$

Im Spezialfall $p = 2 = q$ schreibt man die Resultate gerne folgendermaßen :

Satz : (Cauchy-Schwarz'sche Ungleichung)

$$\left|\sum \overline{y_j} \cdot x_j\right| \leq \sqrt{\sum |y_j|^2} \cdot \sqrt{\sum |x_j|^2}$$

mit Gleichheit, falls

$$y_j = \text{ const} \cdot x_j \text{ für alle } j \,.$$

Man sagt: Die duale Norm zur p-Norm (auf X) ist die q-Norm. Diese Sprechweise sollte verständlich werden, wenn wir etwas weiter ausholen.

Konstruktion (Duale Norm)
Sei $\|\cdot\|$ eine Norm auf dem $\mathbb{C}$-Vektorraum X aller J-Spalten (J sei hier eine endliche Indexmenge). Die Linearformen auf X beschreibt man durch J-Zeilen $\vartheta = (\vartheta_j)_{j\in J}$; man notiert

$$\vartheta x = \sum \vartheta_j \cdot x_j \,.$$

Man definiert nun die zu $\|\cdot\|$ duale Norm $\|\cdot\|^*$

$$\|\vartheta\|^* := \sup\{|\vartheta x| \,:\, \|x\| \leq 1\} \,.$$

Es handelt sich wirklich um eine Norm; denn

$$\|\alpha \cdot \vartheta\|^* = |\alpha| \cdot \|\vartheta\|^* \text{ und}$$
$$\left\|\vartheta^{(1)} + \vartheta^{(2)}\right\|^* \leq \left\|\vartheta^{(1)}\right\|^* + \left\|\vartheta^{(2)}\right\|^* \,, \text{ wegen}$$
$$\left|\left(\vartheta^{(1)} + \vartheta^{(2)}\right) x\right| \leq \left|\vartheta^{(1)} x\right| + \left|\vartheta^{(2)} x\right| \leq \left\|\vartheta^{(1)}\right\|^* + \left\|\vartheta^{(2)}\right\|^* \text{ für alle } x \text{ mit } \|x\| \leq 1 \,.$$

Theorem
Wenn $\|\cdot\|^*$ die duale Norm zu $\|\cdot\|$ ist, dann ist $\|\cdot\|$ die duale Norm zu $\|\cdot\|^*$.

Beweis : Die Linearformen auf dem Raum X^* der J-Zeilen kann man mit den J-Spalten identifizieren. Für jedes $x \in X$ ist zu zeigen

$$\sup\left\{|\vartheta x| \,:\, \|\vartheta\|^* \leq 1\right\} = \|x\|$$

oder, äquivalent dazu:

(i) Für alle $x \in X,\ \vartheta \in X^*$ gilt

$$|\vartheta x| \leq \|\vartheta\|^* \cdot \|x\| \,.$$

(ii) Zu jedem $x \in X$ und jedem $\varepsilon > 0$ existiert ϑ_ε, so dass

$$\|\vartheta_\varepsilon\| \leq 1 \text{ und } |\vartheta_\varepsilon x| \geq (1-\varepsilon)\|x\| .$$

Die Ungleichung (i) folgt aus der Definition der dualen Norm $\|\cdot\|^*$. Die Existenzaussage (ii) ist nicht schwer zu beweisen. Es existiert sogar ein ϑ mit $\vartheta x = \|x\|$. Das ist auch so im unendlichdimensionalen Vektorraum X der finiten Folgen. In diesem Fall zeigt die Hölder'sche Ungleichung, dass die p-Norm und die q-Norm zueinander dual sind, wenn $\frac{1}{p} + \frac{1}{q} = 1$. Die Minkowskische Ungleichung ergibt sich aus dem allgemeinen Sachverhalt, dass

$$\vartheta \longmapsto \sup\left\{\vartheta x \ : \|x\| \leq 1\right\} =: \|\vartheta\|^*$$

für jede Norm $\|\cdot\|$ eine Norm ist.
Wer einen „direkten" Beweis der Minkowskischen Ungleichung studieren will, ziehe irgendein Lehrbuch Analysis I zu Rate.

Hinweis.
Sei $k(\cdot)$ eine beliebige konvexe Funktion auf einem reellen affinen Raum X. Man definiert dann dazu eine Funktion $k^*(\cdot)$ auf dem Dualraum

$$k^*(\vartheta) \ := \sup\left\{\vartheta x - k(x) \ : x \in X\right\} .$$

$k^*(\cdot)$ heißt eine Legendre-Transformierte von $k(\cdot)$.

Bemerke :

1) Die affine Funktion $x \longmapsto \vartheta x - k^*(\vartheta)$ nennt man die Subtangente zur Richtung ϑ. $k^*(\vartheta)$ ist die kleinste Zahl c, für welche gilt $\forall\, x\ \vartheta x - c \leq k(x)$.

2) Wenn man die Legendre-Transformation auf $k^*(\cdot)$ anwendet, erhält man für alle $x \in X$

$$k^{**}(x) = \sup\left\{\vartheta x - k^*(\vartheta) \ : \vartheta \in X^*\right\} .$$

Ein berühmtes Theorem, welches wir hier noch nicht beweisen können (weil wir den Begriff der Unterhalbstetigkeit noch nicht eingeführt haben), besagt

Theorem
Wenn $k(\cdot)$ eine unterhalbstetige konvexe Funktion ist, dann gilt $k^{**}(x) = k(x)$ für alle x.
Geometrisch gesprochen: Das punktweise Supremum aller Subtangenten ist die größte unterhalbstetige konvexe Minorante.

23. Vorlesung : Prä-Hilberträume

Einen normierten Vektorraum $(V, \|\cdot\|)$ nennt man auch einen Prä-Banachraum. Man nennt ihn einen Prä-Hilbertraum, wenn die Norm die folgende Gleichung erfüllt

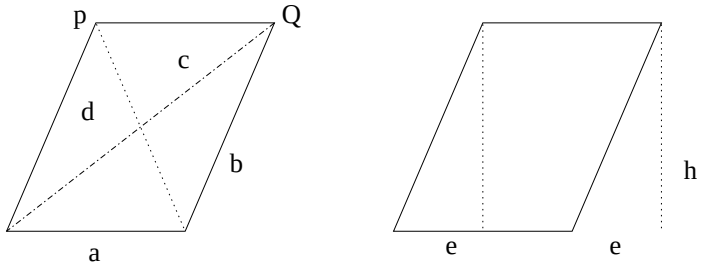
$$\|v+w\|^2 + \|v-w\|^2 = 2\cdot\|v\|^2 + 2\|w\|^2 \quad \text{(Parallelogrammgleichung)}$$

Hinweis : Endlichdimensionale normierte Vektorräume sind vollständig in dem Sinne, den wir später diskutieren werden. Banachräume entstehen durch Vervollständigung aus Prä-Banachräumen. Hilberträume entstehen durch Vervollständigung aus Prä-Hilberträumen.

Satz (1. Beispiel)
Die euklidische Norm im Anschauungsraum erfüllt die Parallelogrammgleichung.

1. Beweis
Der erste Beweis baut elementargeometrisch auf dem Satz von Pythagoras auf. Betrachten wir ein Parallelogramm mit den Seitenlängen a, b. Die Längen der Diagonalen seien c und d. Wir zeigen

$$c^2 + d^2 = 2a^2 + 2b^2$$



Wir fällen das Lot von P und Q auf die Grundlinie und haben nach dem Satz von Pythagoras

$$c^2 = (a+e)^2 + h^2\,, \quad d^2 = (a-e)^2 + h^2\,, \quad b^2 = e^2 + h^2\,.$$

Addition der beiden ersten Gleichungen ergibt

$$c^2 + d^2 = 2a^2 + 2e^2 + 2h^2 = 2a^2 + 2b^2\,.$$

2. Beweis
Wir deuten die Strecken als komplexe Zahlen. Zu zeigen ist

$$|z+w|^2 + |z-w|^2 = 2|z|^2 + 2|w|^2\,.$$

Das ergibt sich sofort, wenn man die komplex konjugierten Zahlen ins Spiel bringt

$$(z+w)(\bar{z}+\bar{w}) + (z-w)(\bar{z}-\bar{w}) = z\bar{z} + w\bar{w} + z\bar{z} + w\bar{w}\,.$$

Die gemischten Produkte fallen weg.

Die folgende ähnliche Rechnung werden wir später brauchen.

Hilfssatz
Sind z, w komplexe Zahlen, so gilt

$$|z+w|^2 - |z-w|^2 = 2z\bar{w} + 2\bar{z}w$$
$$|z+w|^2 - |z-w|^2 + i|z-iw|^2 - i|z+iw|^2 = 4\bar{z}w\,.$$

Der Beweis ist eine simple Rechnung.

Ein zweites Beispiel für den Begriff des Prä-Hilbertraums ist der Vektorraum der trigonometrischen Polynome mit der 2-Norm.

Satz (2. Beispiel)
Betrachte trigonometrische Polynome

$$f(t) = \sum c_n e^{int}\,,\ g(t) = \sum d_n e^{int}$$

mit den Normen

$$\begin{aligned} \|f\| &= \left(\tfrac{1}{2\pi}\int |f(t)|^2 dt\right)^{1/2} = \left(\sum |c_n|^2\right)^{1/2} \text{ bzw.} \\ \|g\| &= \left(\tfrac{1}{2\pi}\int |g(t)|^2 dt\right)^{1/2} = \left(\sum |d_n|^2\right)^{1/2}. \end{aligned}$$

Es gilt

$$\|f+g\|^2 + \|f-g\|^2 = 2\|f\|^2 + 2\|g\|^2$$
$$\|f+g\|^2 - \|f-g\|^2 + i\cdot\|f-ig\|^2 - i\cdot\|f+ig\|^2 =$$
$$= 4\cdot\frac{1}{2\pi}\int \left(\bar{f}\cdot g\right)(t)dt = 4\cdot\sum \bar{c}_n d_n\ .$$

Der Beweis ist eine einfache Rechnung auf der Grundlage der Feststellung. dass die komplexen Zahlen $f(t), g(t)$ (für alle t) und die komplexen Zahlen c_n, d_n (für alle n) die entsprechenden Gleichungen erfüllen (siehe Hilfssatz).

Satz (3. Beispiel)
Sei J eine endliche Indexmenge und V der $\mathbb{C}$-Vektorraum der komplexen J-Spalten. Für $v \in V$ sei v^* die J-Zeile mit den konjugierten Einträgen.
H sei eine $J \times J$-Matrix mit den Eigenschaften

$$H = H^* \quad \text{und} \quad v^* H v > 0 \text{ für alle } v \neq 0\,.$$

(Eine solche Matrix H nennt man eine positiv definite hermitische Matrix).
Die Funktion

$$p(v) = \sqrt{v^* H v} \text{ für } v \in V$$

ist dann eine Norm auf V, welche die Parallelogrammgleichung erfüllt.
$\big(V, p(\cdot)\big)$ ist ein Hilbertraum.

Beweis :

1) Die Parallelogrammgleichung ist (mit Hilfe der Distributiv- und Assoziativgesetze der Matrizenrechnung) leicht nachzurechnen.

2) Man rechnet auch leicht nach

$$p^2(v+w) - p^2(v-w) + ip^2(v-iw) - ip^2(v+iw) = 4v^*Hw\,.$$

3) Die Subadditivität von $p(\cdot)$ ergibt sich aus der Schwarz'schen Ungleichung

$$|v^*Hw| \le p(v)\cdot p(w)\,,$$

die wir unten beweisen. Daraus ergibt sich in der Tat

$$\begin{aligned} p^2(v+w) &= (v+w)^*H(v+w) \\ &= p^2(v) + p^2(w) + w^*Hv + v^*Hw \\ &\le p^2(v) + p^2(w) + 2\cdot p(v)\cdot p(w) = \big(p(v)+p(w)\big)^2\,. \end{aligned}$$

Hilfssatz
Sei S eine $J\times J$-Matrix, so dass

$$v^*Sv \ge 0 \ \text{ für alle } J-\text{Spalten } v\,.$$

Es gilt dann für alle v, w

$$|v^*Sw| + |w^*Sv| \le 2\cdot\sqrt{v^*Sv}\cdot\sqrt{w^*Sw}\,.$$

Beweis :
Es genügt den Fall zu betrachten, wo v^*Sw und w^*Sv reell (und ≥ 0) sind. Die rechte Seite ändert sich nämlich nicht, wenn man v durch $v\cdot e^{i\varphi}$ und w durch $w\cdot e^{i\psi}$ ersetzt. Wir haben für alle reellen λ mit $2b = v^*Sw + w^*Sv$

$$0 \le (v^* + \lambda w^*)\,S(v+\lambda w) = v^*Sv + \lambda 2b + \lambda^2\cdot w^*Sw$$

Aus der Schulmathematik ist bekannt, dass für eine reelle quadratische Funktion $a + 2b\lambda + c\lambda^2$, die keine negativen Werte hat, gilt $b^2 \le ac$. Also gilt hier $b \le \sqrt{a\cdot c}$ q.e.d.
Wenn man ähnliche Überlegungen in unendlichdimensionalen Vektorräumen durchführen will, dann kann man nicht mit Matrizen arbeiten. Man braucht den Begriff der Sesquilinearform und speziell den Begriff der positiv definiten hermitischen Form. Die Algebra zu diesen Begriffen entwickeln wir in der nächsten Vorlesung. Dort werden wir auch den folgenden wichtigen Satz bewiesen.

Theorem (Jordan und v. Neumann)
Sei $(V, \|\cdot\|)$ ein Prä-Hilbertraum. Für $v, w \in V$ definiert man das innere Produkt

$$s(v,w) = \langle v|w\rangle \ := \frac{1}{4}\|v+w\|^2 - \frac{1}{4}\|v-w\|^2 + \frac{i}{4}\|v-iw\|^2 - \frac{i}{4}\|v+iw\|^2\,.$$

Es gilt dann

(i) $\langle u|v+w\rangle = \langle u|v\rangle + \langle u|w\rangle$

(ii) $\langle v|\alpha w\rangle = \alpha\langle v|w\rangle$ für alle $\alpha \in \mathbb{C}$

(iii) $\langle v|w\rangle = \overline{\langle w|v\rangle}$

(iv) $\langle v|v\rangle = \|v\|^2 > 0$ für alle $v \neq 0$

(Ohne Beweis.)

Man nennt $\langle\cdot|\cdot\rangle$ das **innere Produkt** zur Prä-Hilbertraum-Norm $\big(V, \|\cdot\|\big)$.
Das innere Produkt ist eine Spezialität der (Prä)-Hilberträume, welche eine ungeheuere Tragweite hat. Man benützt es z.B. um orthonormierte Koordinaten einzuführen.

Definition

Eine Familie von Vektoren $(e_n)_{n\in I}$ heißt ein Orthonormalsystem (kurz ein ONS), wenn gilt

$$\langle e_n|e_m\rangle = \begin{cases} 1 & \text{falls} \quad n = m \\ 0 & \text{falls} \quad n \neq m \end{cases}$$

Satz (Orthonormalisierungsverfahren von E. Schmidt)

Bezeichne $\langle\cdot|\cdot\rangle$ das innere Produkt in einem Prä-Hilbertraum V.
$v_1, v_2, v_3, \ldots$ seien linear unabhängige Vektoren. Es existiert dann ein ONS $e_1, e_2, e_3, \ldots$, sodass $\{e_1, \ldots, e_m\}$ denselben Teilraum aufspannt wie $\{v_1, \ldots, v_m\}$ (für jedes m).

Beweis

Es gibt ein Vielfaches von v_1, welches die Länge 1 hat. e_1 sei ein solches. Wir konstruieren die e_m rekursiv. $\{e_1, \ldots, e_m\}$ seien bereits konstruiert. v_{m+1} ist nicht Linearkombination der v_j mit $j \leq m$ und daher auch nicht Linearkombination der e_j mit $j \leq m$.

Sei $b_j = \langle e_j|v_{m+1}\rangle$ für $j = 1, \ldots, m$ und $\tilde{e}_{m+1} = v_{m+1} - \sum^{m} b_j e_j$.
Es gilt $\langle e_k|\tilde{e}_{m+1}\rangle = \langle e_k|v_{m+1}\rangle - \sum_j b_j \langle e_k|e_j\rangle = 0$ für $k = 1, \ldots, m$.

Man wähle nun e_{m+1} als ein Vielfaches von $\tilde{e}_{m+1}$ mit der Norm 1.

Bemerke

Sei $e_1, e_2, \ldots$ ein ONS in Prä-Hilbertraum $\big(V, \langle\cdot|\cdot\rangle\big)$. Für die Vektoren

$$v = \sum c_k e_k\,, \quad w = \sum d_k e_k$$

gilt dann

$$\langle v|w\rangle = \sum \bar{c}_k \cdot d_k\,.$$

Insbesondere gilt

$$\langle v|v\rangle = \|v\|^2 = \sum |c_k|^2\,.$$

Hinweis :
Manche Autoren definieren: Ein Vektorraum V wird zu einem Prä-Hilbertraum, wenn man eine Funktion $\langle\cdot|\cdot\rangle$ auf $V \times V$ spezifiziert, welche die Eigenschaften (i) ... (iv) besitzt. Man beweist wie oben die Schwarz?sche Ungleichung

$$|\langle v|w\rangle| \leq \|v\| \cdot \|w\|$$

indem man bemerkt, dass $\langle v+\lambda w|v+\lambda w\rangle$ eine nichtnegative quadratische Funktion von λ ist. Daraus ergibt sich wie im dritten Beispiel die Abschätzung

$$\|v + w\| \leq \|v\| + \|w\| \,.$$

So beweist man, dass $\|\cdot\|$ eine Norm ist, welche die Parallelogrammgleichung erfüllt. Die Norm wird hier also als eine (aus dem inneren Produkt) abgeleitete Struktur aufgefaßt.

Hinweis : P.A. Dirac und viele Physiker nach ihm benützen die Bezeichnung $|v\rangle$ für die Zustände eines quantenmechanischen Systems und sie nennen $|v\rangle$ einen ket-Vektor. Die Linearformen $\langle u|\cdot\rangle$ nennen sie die bra-Vektoren. Die Zahl $\langle w|v\rangle$, das innere Produkt von w und v ist die bracket („Klammer“). Der Ausdruck $|v\rangle\langle u|\cdot\rangle$ wird als eine Projektion auf den von v aufgespannten Vektorraum verstanden.

Diese Notationn findet sich z.B. auch in Feynman's Vorlesungen, Band III, Kapitel 20.

Didaktische Anmerkungen

1) In der Schulgeometrie geht man davon aus, dass der Begriff des rechten Winkels (Orthogonalität) und die Längenmessung intuitiv gegeben sind. Wenn man den Satz von Pythagoras zum sog. Cosinussatz verallgemeinern will, dann stößt man auf den Begriff des inneren Produkts. Um dieses einzuführen, stützt man sich üblicherweise auf kartesische Koordinaten. Es gilt als intuitiv klar, dass es kartesische Koordinaten gibt.

2) Wir haben hier gesehen, dass es auch anders geht. Man braucht nur die Parallelogrammgleichung und die Idee der Polarisierung. Im reellen Fall reduziert sich die Polarisierungskonstruktion auf

$$\langle v|w\rangle = \frac{1}{4}\|v + w\|^2 - \frac{1}{4}\|v - w\|^2 = \frac{1}{2}\|v + w\|^2 - \frac{1}{2}\|v\|^2 - \frac{1}{2}\|w\|^2$$

und das ist im Wesentlichen der Cosinussatz.

3) Das didaktische Problem mit dem Skalarprodukt besteht darin, dass es die Unterscheidung zwischen Vektoren und „Covektoren“ (=Linearformen) überflüssig zu machen scheint. Die Vermischung behindert dann das Verständnis für den Begriff der Dualität. Das Verständnis für die Struktur allgemeiner Vektorräume wird daher leider manchmal durch allzuflüssige Erfahrungen mit dem euklidischen Anschauungsraum erschwert.

Satz :
Das innere Produkt in einem Prä-Hilbertraum V

$$\langle\cdot|\cdot\rangle = s(\cdot,\cdot)$$

liefert eine antilineare isometrische Abbildung in den Dualraum

$$\varphi(\cdot) \ : V \to V^* \,,\ \varphi(v) = s(v,\cdot) = \langle v|\cdot\rangle \,.$$

Beweis :

$$\begin{array}{lrcl} 1) & \varphi(v+w) & = & \varphi(v)+\varphi(w)\,,\ \varphi(\alpha,v) = \bar{\alpha}\cdot\varphi(v) \\ 2) & \|\varphi(v)\|^* & = & \sup\left\{|\langle\varphi(v),w\rangle| \ : \|w\| \le 1\right\} \\ & & & \sup\left\{|\langle v|w\rangle| \ : \|w\| \le 1\right\} = \|v\|^2 \,. \end{array}$$

Beispiel
V sei der Vektorraum der J-Spalten. Der Dualraum V^* kann mit dem Vektorraum der J-Zeilen identifiziert werden; $\vartheta(v) = \langle\vartheta, v\rangle$ ist das „Matrizenprodukt“ $\vartheta v = \sum \vartheta_j v_j$. H sei eine positiv definite hermitische $J \times J$-Matrix. $s(\cdot,\cdot)$ sei das von H erzeugte Skalarprodukt

$$\|v\|^2 = v^* H v \quad s(v,w) = v^* \cdot H \cdot w \,.$$

Offenbar haben wir

$$\varphi(v) = s(v,\cdot) = v^* \cdot H \,.$$

Die duale Norm auf der Menge aller Zeilen ϑ ist

$$\|\vartheta\|^* = \sqrt{\vartheta^* \cdot H \cdot \vartheta} \,.$$

24. Vorlesung : Bilinearformen und Polarisierung

Wir machen hier einen Ausflug in die Algebra.

Definition
Sei V ein $\mathbb{C}$-Vektorraum.
Eine komplexwertige Funktion $t(\cdot,\cdot)$ auf $V \times V$ heißt eine **Bilinearform**, wenn gilt

(i) $t(u, v+w) = t(u,v) + t(u,w)$

(ii) $t(u+v, w) = t(u,w) + t(v,w)$

(iii) $t(v, \alpha w) = \alpha \cdot t(v,w)$ für alle $\alpha \in \mathbb{C}$

(iv) $t(\alpha v, w) = \alpha \cdot t(v,w)$ für alle $\alpha \in \mathbb{C}$.

Sie heißt eine **Sesquilinearform**, wenn anstelle von (iv) gilt

(iv*) $t(\alpha v, w) = \bar{\alpha} \cdot t(v,w)$

Bemerkung
Wenn ein $t(\cdot,\cdot)$ die Bedingung (i) erfüllt, dann sagt man, $t(\cdot,\cdot)$ sei additiv im zweiten Argument. Aus (i) folgert man leicht: Für jedes feste Paar u, v gilt

$$t(u, r \cdot v) = r \cdot t(u,v) \text{ für alle rationalen } r = \frac{m}{n}\,.$$

Es bleibt hier noch offen, ob $f(x) := t(u, x \cdot v)$ eine lineare Funktion auf $\mathbb{R}$ ist. Wir haben zunächst nur die Additivität $f(x+y) = f(x) + f(y)$.

Hinweis: Die Grundlagentheoretiker haben sich mit der Frage beschäftigt, ob es auch noch andere additive $f(\cdot)$ auf $\mathbb{R}$ geben könnte. Es ist erwiesen, daß man nur mit ganz verrückten „Konstruktionen“ zu solchen Funktionen gelangen und dass diese notwendigerweise überall unstetig sind; sie sind noch nicht einmal „borelmeßbar“. Für eine biadditive Funktion $t(\cdot,\cdot)$ $\big($d.h. ein $t(\cdot,\cdot)$ mit den Eigenschaften (i) und (ii)$\big)$, für welche man solche Verrücktheiten ausschließen kann (und das wollen wir hier tun), reduzieren sich die Forderungen (iii) und (iv) bzw (iv*) zu

(iii) $t(v, iw) = i \cdot t(v,w)$

(iv) $t(iv, w) = i \cdot t(v,w)$

bzw.

(iv$*$) $t(iv, w) = t(iv, w)$

Sprechweisen

a) Eine Bilinearform mit
$$t(v,w) = t(w,v)$$
heißt eine **symmetrische** Bilinearform.

b) Eine Bilinearform mit
$$t(v,w) = -t(w,v)$$
heißt eine **alternierende 2-Form** oder auch eine schiefsymmetrische Bilinearform.

c) Eine Sesquilinearform mit
$$t(v,w) = \overline{t(w,v)}$$
heißt eine **hermitische Form**.

Bemerkungen

1) Jede Bilinearform läßt sich in eindeutiger Weise als Summe einer symmetrischen und einer schiefsymmetrischen Bilinearform schreiben:
$$\begin{array}{rcl} t(\cdot,\cdot) & = & b(\cdot,\cdot) + a(\cdot,\cdot) \\ b(v,w) & = & \frac{1}{2}t(v,w) + \frac{1}{2}t(w,v) \\ a(v,w) & = & \frac{1}{2}t(v,w) + \frac{1}{2}t(w,v) \end{array}$$

2) Wenn $t(\cdot,\cdot)$ eine Sesquilinearform ist, dann ist
$$h(v,w) = \frac{1}{2}t(v,w) + \frac{1}{2}\overline{t(w,v)}$$
eine hermitische Form.

Satz

Wenn $b(\cdot,\cdot)$ eine symmetrische Bilinearform ist, dann heißt $q(v) = b(v,v)$ die von $b(\cdot,\cdot)$ erzeugte quadratische Form. Es gilt

(i) $q(\alpha v) = \alpha^2 \cdot q(v)$ für alle α

(ii) $q(v+w) + q(v-w) = 2q(v) + 2q(w)$

(iii) $b(v,w) = \frac{1}{4}q(v+w) - \frac{1}{4}q(v-w)$

Satz (Polarisierungsidentität)

Sei $h(\cdot,\cdot)$ eine hermitische Form und
$$q(v) = h(v,v)\,.$$
Für die reellwertige Funktion $q(\cdot)$ gilt dann

(i) $q(\alpha v) = |\alpha|^2 \cdot q(v)$

(ii) $q(v+w) + q(v-w) = 2q(v) + 2q(w)$

(iii) $h(v,w) = \frac{1}{4}q(v+w) - \frac{1}{4}q(v-w) + \frac{i}{4}q(v-iw) - \frac{i}{4}q(v+iw)$

Hinweis : Es ist ein weithin üblicher Mißbrauch der Sprache, wenn man die einer Sesquilinearform zugeordnete Funktion $q(v) := s(v,v)$ die erzeugte quadratische Form nennt.

Beispiel

J sei eine endliche Indexmenge. V sei der $\mathbb{C}$-Vektorraum aller J-Spalten.
Für $v \in V$ bezeichnet v^t die J-Zeile mit den Einträgen von v; V^* bezeichnet die J-Zeile mit den konjugierten Einträgen.

1) Jede komplexe $J \times J$-Matrix T liefert eine Bilinearform

$$t(v,w) = v^t T w$$

und eine Sesquilinearform

$$s(v,w) = v^* T w\,.$$

Jede Bilinearform und jede Sesquilinearform entsteht aus einer wohlbestimmten Matrix T.

2) T^t bezeichne die transponierte Matrix

$$B = \frac{1}{2}(T + T^t) \text{ und } A = \frac{1}{2}(T - T^t)$$

liefern den symmetrischen und den alternierenden Anteil der Bilinearform $v^T T w$. Für die erzeugte quadratische Form gilt

$$q(v) = v^T T v = v^T \cdot B \cdot v\,.$$

3) T^* bezeichne die hermitisch konjugierte Matrix .

$$H = \frac{1}{2}(T + T^*) \text{ ist hermitisch.}$$

Für die „erzeugte quadratische Form“ gilt

$$q(v) = v^* H v = \Re(v^* T v)\,.$$

Hinweis : Quadratische Matrizen sollte man nicht immer als Darstellungen von Endomorphismen oder (im nichtsingulären Fall) als Darstellungen von Koordinatenwechseln auffassen. Manchmal dienen sie auch der Darstellung von Bilinearformen oder von Sesquilinearformen.

Anhang

Wir zeigen nun, wie man zu einer quadratischen Funktion eine Bilinearform bzw. eine Sesquilinearform gewinnt.

Hifssatz (Polarisierungskonstruktion)
Sei q eine komplexwertige Funktion mit

(i) $q\big((1+i)v\big) = 2i \cdot q(v)$

(ii) $q(v+w) + q(v-w) = 2q(v) + 2q(w)$

Betrachte auf $V \times V$

$$b(v,w) = \frac{1}{4}q(v,w) - \frac{1}{4}q(v-w)\,.$$

Es gilt dann

(i) $b(u, v+w) = b(u,v) + b(u,w)$

(ii) $b(v, iw) = ib(v,w)$

(iii) $b(w,v) = b(v,w)$

(iv) $b(v,v) = q(v)$

Beweis

1) Aus der Parallelogrammgleichung erhalten wir

$$q(0) = 0, q(-w) = q(w)\,,\ q(v+v) = 4 \cdot q(v)\,,\ b(v,v) = q(v)\,.$$

Aus der Eigenschaft (i) und $(1+i)^2 = 2i$ ergibt sich

$$\frac{1}{4}q\Big((1+i)^2 v\Big) = \frac{1}{2}i \cdot q\Big((1+i)v\Big) \text{ und } q(iv) = \frac{1}{4}q(2i \cdot v) = -q(v)\,.$$

2) Wir zeigen

$$b(u,v) + b(u,w) = 2b\left(u, \frac{v+w}{2}\right)\,.$$

Setze $m = u + \frac{v+w}{2}$, $d = \frac{v-w}{2}$. Das ergibt

$$\begin{array}{rcl} u+v = m+d, & & u+w = m-d \\ q(u+v)+q(u+w) & = & 2q(m) + 2q(d) = 2q\left(u + \frac{v+w}{2}\right) + 2q(d) \\ q(u-v)+q(u-w) & = & 2q\left(u - \frac{v+w}{2}\right) + 2q(-d)\,. \end{array}$$

Die Differenz ist

$$\begin{array}{rcl} 4b(u,v) + 4b(u,w) & = & 2q\left(u + \frac{v+w}{2}\right) - 2q\left(u - \frac{v+w}{2}\right) \\ & = & 2 \cdot 4 \cdot b\left(u, \frac{v+w}{2}\right)\,. \end{array}$$

3) Die Symmetrie (iii) ist trivial.

4) Der Beweis von $b(v, iw) = ib(v,w)$ benötigt eine Rechnung:

$$\begin{array}{rcl} 4 \cdot b(v, iw) & = & q(v+iw) - q(v-iw) \\ & = & -q\big((-i)(v+iw)\big) + q\big(i(v-iw)\big) \\ & = & q(w+iv) - q(w-iv) = 4b(iv, w) \\ b\Big(v, (1+i)w\Big) & = & b(v,w) + b(v,iw) = b\big((1+i)v, w\big)\,. \end{array}$$

Wegen $\frac{1+i}{1-i} = i$ und $(1+i)(1-i) = 2$ haben wir weiter

$$\begin{aligned} b(v,iw) &= \tfrac{1}{4}q(v+iw) - \tfrac{1}{4}q(v-iw) \\ &= \tfrac{1}{4}q\Big((1+i)\left[\tfrac{v}{1+i} + \tfrac{w}{1-i}\right]\Big) - \tfrac{1}{4}q\Big((1+i)\left[\tfrac{v}{1+i} - \tfrac{w}{1-i}\right]\Big) \\ &= 2ib\left(\tfrac{v}{1+i}, \tfrac{w}{1-i}\right) = 2ib\left(v, \tfrac{w}{2}\right) = ib(v,w)\,. \end{aligned}$$

Es folgt der

Satz

Eine symmetrische Bilinearform $b(\cdot,\cdot)$ ist durch die erzeugte quadratische Funktion $q(v) = b(v,v)$ eindeutig bestimmt. Zu jedem $q(\cdot)$ mit

$$\begin{aligned} q(\alpha v) &= \alpha^2 \cdot q(v) \text{ für alle } \alpha \in \mathbb{C} \\ q(v+w) + q(v-w) &= 2q(v) + 2q(w) \end{aligned}$$

gibt es genau eine symmetrische Bilinearform, welche $q(\cdot)$ erzeugt.

Einen entsprechenden Satz gibt es für hermitische Formen. (Aus ihm ergibt sich dann sofort der Satz von Jordan und v. Neumann.)

Satz

Eine hermitische Form $h(\cdot,\cdot)$ ist durch die erzeugte „quadratische“ Funktion $q(v) = h(v,v)$ eindeutig bestimmt. Zu jedem $q(\cdot)$ mit

$$\begin{aligned} q(\alpha, v) &= |\alpha|^2 \cdot q(v) \\ q(v+w) + q(v-w) &= 2q(v) + 2q(w) \end{aligned}$$

gibt es genau eine hermitische Form, welche $q(\cdot)$ erzeugt.

Der Beweis egibt sich aus dem analogen

Hifssatz (Polarisierungskonstruktion im hermitischen Fall)
Sei $q(\cdot)$ eine reellwertige Funktion auf V mit

i) $q(iv) = q(v)$

ii) $q(v+w) + q(v-w) = 2q(v) + 2q(w)$

Betrachte auf $V \times V$

$$s(v,w) = \frac{1}{4}q(v+w) - \frac{1}{4}q(v-w) + \frac{i}{4}q(v-iw) - \frac{i}{4}q(v+iw)\,.$$

Es gilt dann

i) $s(u, v+w) = s(u,v) + s(u,v)$

ii) $s(v, iw) = is(v,w)$

iii) $s(w,v) = \overline{s(v,w)}$

iv) $s(v,v) = q(v)$

Beweis :

1) Wegen $\frac{1-i}{1+i} = -i$ haben wir

$$\begin{aligned} q(v) + q(-iv) &= q\left(\frac{v}{1+i} + \frac{iv}{1+i}\right) + q\left(\frac{v}{1+i} - \frac{iv}{1+i}\right) \\ &= 2 \cdot q\left(\frac{v}{1+i}\right) + 2 \cdot q\left(\frac{iv}{1+i}\right) = 4 \cdot q\left(\frac{v}{1+i}\right) . \end{aligned}$$

Also $q(v) = 2 \cdot q\left(\frac{v}{1+i}\right) = 2 \cdot q\left(\frac{v}{1-i}\right)$

$$4s(v,v) = q(v+v) - q(0) + i \cdot q(v - iv) - iq(v + iv) = 4q(v) .$$

2)

$$\begin{aligned} 4\overline{s(w,v)} &= q(w+v) - iq(w - iv) - iq(w - iv) + iq(w + iv) \\ &= q(v+w) - q(v-w) - iq(iw + v) + iq(iw - v) \\ &= 4s(v,w) . \end{aligned}$$

3)

$$\begin{aligned} 4s(v,iw) &= q(v + iw) - q(v - iw) + iq(v + w) - iq(v - w) \\ &= 4i \cdot s(v,w) . \end{aligned}$$

4) Wie oben zeigen wir

$$\begin{aligned} &4s(u,v) + 4s(u,w) = \\ &= \Big(q(u+v) + q(u+w)\Big) - \Big(q(u-v) + q(u-w)\Big) \\ &\quad + i\Big(q(u - iv) + q(u - iw)\Big) - i\Big(q(u+v) + q(u - iw)\Big) \\ &= 2 \cdot q\left(u + \frac{v+w}{2}\right) - 2q\left(u - \frac{v+w}{2}\right) + 2iq\left(u - i\frac{v+w}{2}\right) - 2i \cdot q\left(u + i\frac{v+w}{2}\right) \\ &= 2 \cdot 4 \cdot s\left(u, \frac{v+w}{2}\right) . \end{aligned}$$

Damit ist der Hilfssatz bewiesen.

Der Minkowski-Raum

Das Raum-Zeit-Kontinuum der speziellen Relativitätstheorie ist ein vierdimensionaler reeller affiner Raum. Auf dem $\mathbb{R}$-Vektorraum der Verschiebungen hat man eine ausgezeichnete quadratische Form, die positive und negative Werte annehmen kann. Ein Vektor v heißt zeitartig, wenn $q(v) > 0$ und raumartig, wenn $q(v) < 0$; v heißt ein Vektor auf dem Lichtkegel, wenn $q(v) = 0$.

Es gibt dreidimensionale Unterräume $\tilde{V}$, die nur aus raumartigen Vektoren (und dem Nullvektor) bestehen. In einem solchen Teilvektorraum liefert $-q(\cdot)$ eine euklidische Norm. Ausgehend von irgendwelchen drei linear unabhängigen Vektoren kann man durch das Schmidt?sche Orthogonalisierungsverfahren ein ONS konstruieren. Dazu gibt es einen (bis auf das Vorzeichen eindeutig bestimmten) weiteren Vektor v_0 mit

$$q(v_0) = 1 \quad \text{und} \quad \frac{1}{4}q(v_0 + \tilde{v}) - \frac{1}{4}q(v_0 - \tilde{v}) = 0 \quad \text{für alle } \tilde{v} \in \tilde{V} .$$

Man erhält so eine Basis v_0, v_1, v_2, v_3 mit

$$q\left(\alpha_0 v_0 + \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3\right) = \alpha_0^2 - \left(\alpha_1^2 + \alpha_2^2 + \alpha_3^2\right) .$$

Ein Quadrupel (v_0, v_1, v_2, v_3) mit dieser Eigenschaft nennt man ein **Inertialsystem**.

Definition
Eine lineare Abbildung φ heißt eine **Lorentztransformation**, wenn gilt

$$q\big(\varphi(v)\big) = q(v) \text{ für alle } v\,.$$

Satz:

a) Wenn (v_0, v_1, v_2, v_3) ein Inertialsystem ist, dann auch

$$(w_0, w_1, w_2, w_3) = \left(\varphi\left(v_0\right), \varphi\left(v_1\right), \varphi\left(v_2\right), \varphi\left(v_3\right)\right).$$

b) Die Gesamtheit aller Lorentztransformationen ist eine Gruppe linearer Abbildungen.

Der zweidimensionale Minkowski-Raum

Wir betrachten im Vektorraum der 2-Spalten die quadratische Funktion

$$q\left(\begin{pmatrix} t \\ x \end{pmatrix}\right) = t^2 - x^2\,.$$

Wir interessieren uns für diejenigen 2×2-Matrizen A, für welche gilt

$$q\left(A\begin{pmatrix} t \\ x \end{pmatrix}\right) = q\left(\begin{pmatrix} t \\ x \end{pmatrix}\right).$$

Wir nennen sie (für den Augenblick) die zweidimensionalen Lorentzmatrizen. Wenn man sich einen Überblick über die Gruppe dieser A verschaffen will, dann ist es nützlich, die sog. hyperbolischen Funktionen einzuführen :

$$\begin{array}{rcll} \cosh\alpha & = & \frac{1}{2}\left(e^{\alpha} + e^{-\alpha}\right) & \text{(Cosinushyperbolicus)} \\ \sinh\alpha & = & \frac{1}{2}\left(e^{\alpha} - e^{-\alpha}\right) & \text{(Sinushyperbolicus)}\,. \end{array}$$

Lemma (Additionstheoreme)

$$\begin{gathered} (\cosh\alpha)^2 - (\sinh\alpha)^2 = 1 \\ \cosh(\alpha+\beta) = \cosh\alpha \cdot \cosh\beta + \sinh\alpha \cdot \sinh\beta \\ \sinh(\alpha+\beta) = \sinh\alpha \cdot \cosh\beta + \cosh\alpha \cdot \sinh\beta\,. \end{gathered}$$

Satz

a) Für jedes $\alpha \in \mathbb{R}$ ist

$$A_\alpha := \begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix}$$

eine Lorentzmatrix.

b) Für alle α, β gilt

$$A_\alpha \cdot A_\beta = A_{\alpha+\beta}\,.$$

Beweis

1) Wir schreiben $q(\cdot)$ als ein Matrizenprodukt

$$\begin{aligned} q\begin{pmatrix} t \\ x \end{pmatrix} &= (t,x)\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} t \\ x \end{pmatrix} \\ q\big(A\begin{pmatrix} t \\ x \end{pmatrix}\big) &= (t,x)A^T \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} A\begin{pmatrix} t \\ x \end{pmatrix} \end{aligned}$$

A ist genau dann Lorentzmatrix, wenn gilt

$$A^T \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \cdot A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\,.$$

Denn eine symmetrische quadratische Form auf $\mathbb{R}^2$ bestimmt in eindeutiger Weise die symmetrische Matrix, welche sie erzeugt.

2) Es gilt

$$\begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix} =$$
$$\begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix}\begin{pmatrix} \cosh\alpha & \sinh\alpha \\ -\sinh\alpha & -\cosh\alpha \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \text{ q.e.d.}$$

3) Die Aussage b) ergibt sich aus den Additionstheoremen.

4) Mit

$$A = \begin{pmatrix} c & s \\ s & c \end{pmatrix}$$

sind auch $-A$ und die folgenden Matrizen Lorentzmatrizen

$$\begin{pmatrix} c & s \\ -s & -c \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} A\,,\ \begin{pmatrix} c & -s \\ s & -c \end{pmatrix} = A \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\,.$$

Unsere A_α sind diejenigen Lorentzmatrizen mit der Determinante $= 1$, welche in der linken oberen Ecke positiv sind. Man nennt sie die orthochronen Lorentzmatrizen.

Die entsprechenden Begriffe für den vierdimensionalen Minkowski-Raum studiert man in der speziellen Relativitätstheorie.

Index