

Skript zur Vorlesung

Theoretische Informationstechnik I&II

Prof. Dr. Rudolf Mathar
Rheinisch-Westfälische Technische Hochschule Aachen
Lehrstuhl für Theoretische Informationstechnik
Sommerfeldstraße 24
52056 Aachen

Version vom 15. März 2012



Inhaltsverzeichnis

1	Einleitung	7
1.1	Grobaufbau der Vorlesung	7
2	Stochastische Modellierung	9
2.1	Grundregeln	9
2.2	Bedingte Wahrscheinlichkeit und stochastische Unabhängigkeit	12
2.3	Zufallsvariable und ihre Verteilung	14
2.3.1	Diskrete Zufallsvariable	16
2.3.2	Absolut-stetige Zufallsvariable	17
2.3.3	Erwartungswerte und Momente von Zufallsvariablen	19
2.4	Zufallsvektoren und Transformationen	22
2.4.1	Gemeinsame Verteilung	23
2.4.2	Erwartungswertvektor und Kovarianzmatrix	24
2.4.3	Stochastische Unabhängigkeit	26
2.4.4	Bedingte Verteilungen	27
2.4.5	Transformation von Zufallsvektoren	29
2.5	Mehrwegeausbreitung in Mobilfunkkanälen	33
2.5.1	Rayleigh-Verteilung	36
2.5.2	Rice-Verteilung	38
2.5.3	Lognormal-Verteilung	40
2.6	Die n -dimensionale komplexe Normalverteilung	41
2.7	Stochastische Prozesse	46
2.7.1	Stationäre stochastische Prozesse	49
2.7.2	Lineare Systeme mit stochastischer Eingabe	52
2.7.3	Leistungsdichtespektrum	55
2.7.4	Weißes Rauschen	60
2.7.5	Filterung von weißem Rauschen	64
3	Elemente der Informationstheorie	67
3.1	Grundlagen	67
3.2	Quellenkodierung	72
3.3	Kanalkapazität und Fundamentalsatz	76
3.4	Differentielle Entropie und Transinformation	80

4 Kanäle und ihre Kapazität	89
4.1 Kanäle mit diskreter Eingabe (Discrete input channel)	89
4.2 Reelle Gaußkanäle	92
4.3 Parallele Gaußkanäle	96
4.4 Bandbegrenzte Gaußkanäle	100
4.5 Gaußkanäle mit Feedback	101
4.6 Komplexe MIMO-Kanäle	104
4.6.1 Feste Kanalmatrix H	106
4.6.2 Rayleigh-Fading	109
4.6.3 Obere Schranken für die Kapazität C	111
5 Optimierung und Algorithmen für schwere Probleme	113
5.1 Lineare Programmierung	113
5.2 Branch-and-Bound für ILP	120
5.3 Simulated Annealing	123
5.4 Markov-Ketten und Konvergenz von Simulated Annealing	126
5.5 Konvexe Optimierung	131

Abbildungsverzeichnis

2.1	Wahrscheinlichkeit als Integral Über die Dichte	11
2.2	Mengenzerlegung von Ω	13
2.3	Berechnung von Wahrscheinlichkeiten mit Dichten	19
2.4	2-dimensionale Normalverteilungen	24
2.5	Mehrwegausbreitung	33
2.6	Modell für Mehrwegeausbreitung	33
2.7	Dichten der Rayleigh-Verteilung	38
2.8	Kanten und Winkel im rechtwinkligen Dreieck	39
2.9	Dichten der Rice-Verteilung	40
2.10	Dichten der Rice-Verteilung	41
2.11	Langsames Fading durch verschiedene Positionierung der Mobilstation.	42
2.12	Dichten der Lognormal-Verteilung	42
2.13	Dichten der Lognormal-Verteilung	43
2.14	Verlauf zufälliger Funktionen	47
2.15	Realteil der Bessel-Funktion 1. Art 0'ter Ordnung	52
2.16	Reaktionsverläufe eines Tiefpassfilters	53
2.17	Energiespektrum und Signalenergie	56
2.18	Bandpassfilter	58
2.19	Impulsantwort	58
2.20	Beispiel zum Leistungsdichtespektrum für $f_D = 1$ und $\sigma^2 = 1$.	60
2.21	Gleitendes Mittel und Faltung der Impulsantwort mit sich selbst.	61
2.22	Impulsantwort und ihre Fouriertransformierte	62
2.23	Faltung der Impulsantwort mit sich selbst und die Fouriertransformierte	63
2.24	Autokorrelationsfunktion und Leistungsdichtespektrum von weißem Rauschen	63
2.25	Leistungsdichtespektrum und Autokorrelationsfunktion von farbigem Rauschen	65
3.1	Standardmodell der Informationsübertragung	67
3.2	Bitübertragung mit zufälliger Verfälschung	70
3.3	Entropie, bedingte Entropie und Transinformation	71
3.4	Kodewortübertragung von Wörtern über den diskreten, gedächtnislosen Kanal	78
3.5	Zusammenhänge zwischen Entropie und Transinformation	87

4.1	Generischer Verlauf der Funktion	91
4.2	Kodewörter umgeben von Kugeln	96
4.3	Veranschaulichung von Waterfilling	97
4.4	Abtastung eines bandbegrenzten Signals	100
5.1	Grafische Darstellung eines linearen Programms	114
5.2	Bi-direktionaler Graph	115
5.3	Überdeckung von Verkehrspunkten durch Basisstationen	116
5.4	Konvexe und nicht konvexe Mengen	117
5.5	Extremalpunkte von Mengen	118
5.6	Schnitt von Geraden mit Polytop	119
5.7	Lösung des kombinatorischen Optimierungsproblems als binärer Baum.	120
5.8	Übergangsgraph im Fritchman Kanalmodell	127

1 Einleitung

Eine grobe Einordnung des Gebiets Informationstechnik ist die Umsetzung von Methoden und Verfahren aus

- Nachrichtentechnik (z.B. Übertragungstechnik, Mobilfunk)
- Signalverarbeitung (z.B. Filterung, Transformation, Detektion)
- Informationstheorie (z.B. Codierung, Kapazität, Kryptographie)

in technische Systeme bzw. Komponenten (z.B. Handy, Festplatte, eingebettetes System, Signalprozessor, Speicher).

Die Theoretische Informationstechnik befasst sich mit der Darstellung der theoretischen Grundlagen auf allgemeiner Basis, so dass Anwendungen einheitlich beschrieben werden können. Eine wichtige Rolle spielen dabei, insbesondere in dieser Vorlesung, Methoden, Verfahren und Algorithmen aus

- Stochastik
- Informationstheorie
- Lineare Systeme und Matrixtheorie
- Optimierung

1.1 Grobaufbau der Vorlesung

1. Einleitung

2. Stochastische Modellierung

(Grundregeln, Zufallsvariable, Zufallsvektoren, n-dim. komplexe Normalverteilung, stochastische Prozesse, stationäre stochastische Prozesse, lineare Systeme mit stochastischer Eingabe, Leistungsdichtespektrum, weißes Rauschen)

3. Elemente der Informationstheorie

(Entropie, Transinformation, Kapazität, Fundamentalsatz, differentielle Entropie und Transinformation)

4. Kanäle und ihre Kapazität
(Kapazität von Gaußkanälen und MIMO-Kanälen)
5. Optimierung und Algorithmen für schwere Probleme
(Lineare Programmierung, Branch-and-Bound, Heuristiken für Kanalzuweisung, Optimierung, Planung von Zellnetzen)

2 Stochastische Modellierung

Most students, brought up with a deterministic outlook of physics, find the subject unreliable, vague, difficult. The difficulties persist because of inadequate definition of the first principles, resulting in a constant confusion between assumptions and logical conclusions. Conceptual ambiguities can be removed only if the theory is developed axiomatically. They say that this approach would require measure theory, would reduce the subject to a branch of mathematics, would force the student to doubt his intuition leaving him without convincing alternatives, but I don't think so. I believe that most concepts needed in the applications can be explained with simple mathematics, that probability, like any other theory, should be viewed as a conceptual structure and its conclusions should rely not on intuition but on logic. The various concepts must, of course, be related to the physical world, but such motivating sections should be separated from the deductive part of the theory. Intuition will thus be strengthened, but not at the expense of logical rigor.

Papoulis, Preface to the first edition

2.1 Grundregeln

Die mathematische Beschreibung von Zufallsexperimenten erfolgt auf folgende Weise. Relevante *Ergebnisse* werden zu einer Menge zusammengefasst

Ω : *Ergebnismenge* (oft: $\{0, 1\}$, $\{1, \dots, 6\}$, \mathbb{N} , $\{0, 1\}^{\mathbb{N}}$, \mathbb{R} , \mathbb{R}^n , \mathbb{C}^n)

Relevante *Ereignisse* werden beschrieben durch Teilmengen von Ω

\mathfrak{A} : *Ereignismenge* (bei diskreten Zufallsexperimenten oft $\mathfrak{A} = \mathfrak{P}(\Omega)$)

Der Ausdruck $\mathfrak{P}(\Omega)$ bezeichnet die Menge *aller* Teilmengen von Ω (die Potenzmenge von Ω). Den Ereignissen werden *Wahrscheinlichkeiten* zugeordnet durch eine Funktion $P: \mathfrak{A} \rightarrow [0, 1]$, für die folgendes gelten soll:

1. $P(\Omega) = 1$, $P(\emptyset) = 0$, (*)
2. $P(A^c) = P(\Omega \setminus A) = 1 - P(A) \quad \forall A \in \mathfrak{A}$,
3. $P(A \cup B) = P(A) + P(B) \quad \forall A, B \in \mathfrak{A} \text{ mit } A \cap B = \emptyset$,

4. Für jede Familie $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathfrak{A}$ paarweise *disjunkter* Mengen (d.h. $A_i \cap A_j = \emptyset$ $\forall i, j \in \mathbb{N}$ mit $i \neq j$) gilt

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

Diese Eigenschaft nennt man die σ -Additivität von P .

Aus den Eigenschaften (*) und (**) lassen sich die anderen herleiten.

Seien A, B Ereignisse, also $A, B \in \mathfrak{A}$. Man benutzt folgende Sprechweisen:

- A^c : „A tritt nicht ein“,
- $A \cup B$: „A oder B treten ein“,
- $A \cap B$: „A und B treten ein“.

Beispiel 2.1.1.

- a) Beim Werfen eines Würfels ist die Ergebnismenge $\Omega = \{1, 2, \dots, 6\}$ und die Ereignismenge $\mathfrak{A} = \mathfrak{P}(\Omega) = \{\emptyset, \{1\}, \dots, \{6\}, \{1, 2\}, \dots, \Omega\}$.

- $A = \{1\}, P(A) = \frac{1}{6}$
- $B = \{2, 4, 6\}, P(B) = \frac{3}{6} = \frac{1}{2}$
- $C = \{2, 3, 4, 5\}, P(C) = \frac{4}{6} = \frac{2}{3}$

- b) Die Ergebnismenge sei $\Omega = \mathbb{R}$ und die Ereignismenge \mathfrak{A} sei die kleinste σ -Algebra, die alle Intervalle in \mathbb{R} enthält ($\mathfrak{A} = \mathfrak{P}(\Omega)$ würde in diesem Fall zu Schwierigkeiten bei der Existenz gewisser Wahrscheinlichkeitsverteilungen führen). Das Wahrscheinlichkeitsmaß sei die *Standardnormalverteilung*.

- $A = [a, b], P(A) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$
- $B = [0, \infty), P(B) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{x^2}{2}} dx = \frac{1}{2}$.

Siehe Abbildung 2.1 für eine Veranschaulichung.

Als Basis für ein exaktes Modell benötigt man die folgenden Definitionen.

Definition 2.1.2. (Wahrscheinlichkeitsraum) Sei $\Omega \neq \emptyset$ die Ergebnismenge. $\mathfrak{A} \subseteq \mathfrak{P}(\Omega)$ heißt σ -Algebra (von Ereignissen), wenn

- i. $\Omega \in \mathfrak{A}$,
- ii. $A \in \mathfrak{A} \Rightarrow A^c \in \mathfrak{A}$,
- iii. $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathfrak{A} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$.

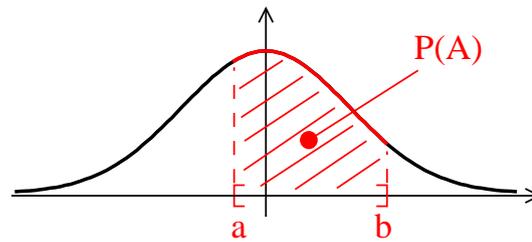


Abbildung 2.1: Wahrscheinlichkeit als Integral über die Dichte der Standardnormalverteilung.

Eine Funktion $P: \mathfrak{A} \rightarrow [0, 1]$ heißt *Wahrscheinlichkeitsverteilung* (*probability distribution*) oder *Wahrscheinlichkeitsmaß* (*probability measure*), wenn gilt

- iv. $P(\Omega) = 1$,
- v. $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ für jede Familie $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathfrak{A}$ paarweise disjunkter Mengen.

$(\Omega, \mathfrak{A}, P)$ wird als *Wahrscheinlichkeitsraum* bezeichnet.

Proposition 2.1.3. (Rechenregeln für Wahrscheinlichkeiten) Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und $A, B, A_n \in \mathfrak{A}$, $n \in \mathbb{N}$. Es gilt

- a) $P(A^c) = 1 - P(A)$.
- b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- c) $A \subseteq B \Rightarrow P(A) \leq P(B)$.
- d) $A \subseteq B \Rightarrow P(B \setminus A) = P(B) - P(A)$.
- e) Falls A_n , $n \in \mathbb{N}$,
 - *aufsteigend*, d.h. $A_n \subseteq A_{n+1} \quad \forall n \in \mathbb{N}$, so gilt

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n),$$

- *absteigend*, d.h. $A_n \supseteq A_{n+1} \quad \forall n \in \mathbb{N}$, so gilt

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

- f) $P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$ für beliebige A_n .

Proposition 2.1.4. (Siebformel und Bonferroni-Ungleichung) Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und $A_1, \dots, A_n \in \mathfrak{A}$. Dann gilt

a) *Siebformel von Poincaré-Sylvester (inclusion-exclusion formula)*

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \dots + (-1)^{n+1} P\left(\bigcap_{i=1}^n A_i\right).$$

b) *Bonferroni-Ungleichung (Bonferroni inequality)*

$$\sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) \leq P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i).$$

2.2 Bedingte Wahrscheinlichkeit und stochastische Unabhängigkeit

Nachfolgend wird die Frage behandelt, wie Wahrscheinlichkeiten festgelegt werden können, wenn das Ergebnis bereits einer bestimmten Teilmenge zugeordnet werden kann.

Beispiel 2.2.1. (Motivation bedingter Wahrscheinlichkeiten) Es wird eine Menge von Datenträgern $\Omega = A \cup B$ betrachtet, die aus Datenträgern der Firma A und der Firma B besteht. I bezeichne die Menge der intakten und D die Menge der defekten Datenträger. Die Wahrscheinlichkeit, dass ein zufällig ausgewählter Datenträger defekt ist, wenn er von der Firma A stammt, kann dann berechnet werden als

$$P(D|A) = \frac{|D \cap A|}{|A|} = \frac{\frac{|D \cap A|}{|\Omega|}}{\frac{|A|}{|\Omega|}} = \frac{P(D \cap A)}{P(A)}.$$

Für eine Menge M bezeichnet $|M|$ die Anzahl der Elemente in M (die *Kardinalität* von M).

Definition 2.2.2. (bedingte Wahrscheinlichkeit) Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und $A, B \in \mathfrak{A}$ mit $P(B) > 0$, dann heißt

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

bedingte Wahrscheinlichkeit (conditional probability) von A unter B (von A , gegeben B).

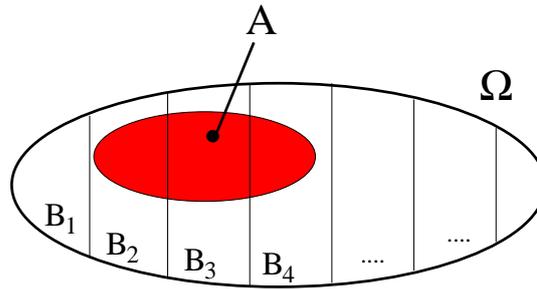


Abbildung 2.2: Mengenzerlegung von Ω

Theorem 2.2.3. (Totale Wahrscheinlichkeit und Bayes-Formel) Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und $\{B_n\}_{n \in \mathbb{N}} \subseteq \mathfrak{A}$ eine *Partition* von Ω , d.h. $\bigcup_{n=1}^{\infty} B_n = \Omega$ und $B_i \cap B_j = \emptyset \forall i, j \in \mathbb{N}$ mit $i \neq j$. Dann gelten

a) *Satz von der totalen Wahrscheinlichkeit (principle of total probability):*

$$P(A) = \sum_{n=1}^{\infty} P(A|B_n) \cdot P(B_n) \quad \text{für alle } A \in \mathfrak{A} \quad (2.1)$$

(Konvention: $* \cdot 0 = 0$)

b) *Bayes-Formel:* Falls $P(A) > 0$, so gilt

$$P(B_n|A) = \frac{P(A|B_n) \cdot P(B_n)}{\sum_{j=1}^{\infty} P(A|B_j) \cdot P(B_j)} \quad \text{für alle } n \in \mathbb{N}. \quad (2.2)$$

Beweis.

a) Die Menge A kann mit Hilfe der Partition $\{B_n\}_{n \in \mathbb{N}}$ in paarweise disjunkte Mengen zerlegt werden, nämlich $A = \bigcup_{n=1}^{\infty} (A \cap B_n)$ und $(A \cap B_i) \cap (A \cap B_j) = \emptyset \forall i, j \in \mathbb{N}$ mit $i \neq j$, siehe Abbildung 2.2. Es folgt

$$\begin{aligned} P(A) &= P\left(\bigcup_{n=1}^{\infty} (A \cap B_n)\right) = \sum_{n=1}^{\infty} P(A \cap B_n) = \sum_{n=1}^{\infty} \frac{P(A \cap B_n)}{P(B_n)} \cdot P(B_n) \\ &= \sum_{n=1}^{\infty} P(A|B_n) \cdot P(B_n), \end{aligned}$$

was mit Hilfe der Konvention $* \cdot 0 = 0$ auch im Fall $P(B_n) = 0$ gilt.

b) Unter Verwendung der Formel aus a) ergibt sich

$$\begin{aligned} P(B_n|A) &= \frac{P(B_n \cap A)}{P(A)} \cdot \frac{P(B_n)}{P(B_n)} = \frac{P(A|B_n) \cdot P(B_n)}{P(A)} \\ &= \frac{P(A|B_n) \cdot P(B_n)}{\sum_{j=1}^{\infty} P(A|B_j) \cdot P(B_j)}. \end{aligned}$$

□

Ein wichtiger Spezialfall ist, wenn die Wahrscheinlichkeit für das Eintreten des Ereignisses A nicht von dem Eintreten des Ereignisses B abhängt, d.h. (für $P(B) > 0$)

$$P(A|B) = P(A) \iff \frac{P(A \cap B)}{P(B)} = P(A) \iff P(A \cap B) = P(A) \cdot P(B).$$

Dieser Zusammenhang wird auf n Ereignisse $A_1, \dots, A_n \in \mathfrak{A}$ bzw. auf Folgen von Ereignissen erweitert und führt zu folgender Definition.

Definition 2.2.4. (stochastische Unabhängigkeit von Ereignissen) Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und $A, B \in \mathfrak{A}$. Dann heißen A und B *stochastisch unabhängig (s.u.) (stochastically independent)*, falls

$$P(A \cap B) = P(A) \cdot P(B).$$

Ereignisse $A_1, \dots, A_n \in \mathfrak{A}$ heißen (*gemeinsam*) *stochastisch unabhängig (jointly stochastically independent)*, falls

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot \dots \cdot P(A_{i_k})$$

für alle $k \leq n$ und Auswahlen $1 \leq i_1 < \dots < i_k \leq n$ gilt.

2.3 Zufallsvariable und ihre Verteilung

Wahrscheinlichkeitsräume $(\Omega, \mathfrak{A}, P)$ dienen allgemein zur Beschreibung von Zufallsexperimenten und zur Modellierung von Zufallseinflüssen.

- Oft interessiert jedoch nicht das gesamte Modell, sondern nur gewisse Teilgrößen. Das ist beispielsweise der Fall bei 5000 Stichproben mit dem Ausgang $0 \stackrel{\wedge}{=} \text{gut}$ oder $1 \stackrel{\wedge}{=} \text{schlecht}$, wobei nur die *Anzahl* der schlechten Stücke relevant ist. Bei dieser Fragestellung wird nicht mehr die ursprüngliche Ergebnismenge $\Omega = \{(x_1, \dots, x_{5000}) \mid x_i \in \{0, 1\}\}$ betrachtet, sondern von Interesse sind nur Ergebnisse aus der Menge $T = \{0, \dots, 5000\}$.

- Die Ergebnisse von Zufallsexperimenten sind oft Zahlen oder Vektoren. Dabei sind arithmetische Operationen oft sehr hilfreich. Ist wie oben $\Omega = \{(x_1, \dots, x_{5000}) \mid x_i \in \{0, 1\}\}$, so lässt sich die Anzahl schlechter Stücke beschreiben durch $\sum_{i=1}^{5000} x_i$. Ein anderes Beispiel bilden etwa gemessene Signalverzögerungszeiten $x_1, \dots, x_n \in \mathbb{R}_+$. Hierbei interessiert insbesondere die mittlere Verzögerungszeit $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

Die Modellierung solcher Fragestellungen erfolgt allgemein mit *Zufallsvariablen*. Zahlreiche weitere Anwendungen werden auftreten.

Definition 2.3.1. (Zufallsvariable) Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum. Eine Abbildung $X: \Omega \rightarrow \mathbb{R}$ heißt *Zufallsvariable* (ZV) oder *Zufallsgröße* (*random variable*, r.v.), wenn

$$X^{-1}((-\infty, x]) = \{\omega \in \Omega \mid X(\omega) \leq x\} \stackrel{(\text{kurz})}{=} \{X \leq x\} \in \mathfrak{A} \quad \forall x \in \mathbb{R}. \quad (2.3)$$

Aus Definition 2.3.1 folgt, dass jede der Mengen

$$\{\omega \in \Omega \mid X(\omega) \leq x\} \stackrel{(\text{kurz})}{=} \{X \leq x\} \subseteq \Omega \quad (2.4)$$

ein Ereignis in \mathfrak{A} ist. Das Ereignis $\{X \leq x\}$ besitzt also eine Wahrscheinlichkeit

$$P(\{\omega \in \Omega \mid X(\omega) \leq x\}) \stackrel{(\text{kurz})}{=} P(X \leq x), \quad x \in \mathbb{R}. \quad (2.5)$$

Dies führt zu folgendem Begriff.

Definition 2.3.2. (Verteilungsfunktion) Die Funktion

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R} \quad (2.6)$$

heißt *Verteilungsfunktion* (*cumulative distribution function*, cdf) der Zufallsvariablen X . $F_X(x) = P(X \leq x)$ besitzt die folgenden Eigenschaften

- F ist monoton steigend (nicht notwendigerweise streng monoton),
- $\lim_{x \rightarrow +\infty} F(x) = 1$ und $\lim_{x \rightarrow -\infty} F(x) = 0$,
- F ist *rechtsseitig stetig*, d.h. $F(x) = \lim_{h \rightarrow 0^+} F(x+h)$.

Umgekehrt lässt sich auch sagen, dass jede Funktion F mit den Eigenschaften i.–iii. die *Verteilungsfunktion* einer Zufallsvariablen ist.

Ein fundamentaler Satz der Wahrscheinlichkeitstheorie besagt, dass die Verteilung jeder Zufallsvariablen X eindeutig durch die zugehörige Verteilungsfunktion beschrieben wird. Die Bezeichnung für die Verteilung von X lautet P^X .

Mathematisches Modell:

<p>Modelliert alle Zufallseinflüsse, wobei das Wissen um die Existenz ausreicht und genauere Kenntnisse oft nicht erforderlich sind.</p>	$(\Omega, \mathfrak{A}, P) \xrightarrow{X} (\mathbb{R}, \mathfrak{B}, P^X)$ $(\mathbb{R}, \mathfrak{B}, P^X)$ ist ebenfalls ein Wahrscheinlichkeitsraum (, in dem die bekannten Regeln gelten), der die interessanten, beobachteten Größen modelliert. P^X ist oft bis auf Parameter bekannt.
--	--

2.3.1 Diskrete Zufallsvariable

Definition 2.3.3. Eine Zufallsvariable X heißt *diskret*, wenn eine höchstens abzählbare Menge $T = \{t_1, t_2, \dots\} \subset \mathbb{R}$ existiert mit

$$P(X^{-1}(T)) = P(\{\omega \in \Omega \mid X(\omega) \in T\}) \stackrel{\text{(kurz)}}{=} P(X \in T) = 1. \quad (2.7)$$

Die Menge T heißt *Träger* der Zufallsvariablen X . Die Funktion $f_X(t_i): T \rightarrow [0, 1]$ mit

$$f_X(t_i) = P(X = t_i), \quad i = 1, 2, \dots \quad (2.8)$$

heißt *Zähldichte* (*probability mass function*).

Beispiel 2.3.4.

- a) *Diskrete Gleichverteilung* (*discrete uniform distribution*) auf $T = \{1, \dots, n\}$
 X heißt diskret gleichverteilt auf $\{1, \dots, n\}$, $n \in \mathbb{N}$, wenn

$$P(X = i) = \frac{1}{n} \quad \forall i = 1, \dots, n.$$

Bezeichnung: $X \sim U(\{1, \dots, n\})$. Beachten Sie die Kurzschreibweise

$$P(X = i) = P(X^{-1}(\{i\})) = P(\{\omega \mid X(\omega) = i\}) .$$

Für alle $A \subseteq T$ gilt

$$P(X \in A) = P\left(X \in \bigcup_{i \in A} \{i\}\right) = \sum_{i \in A} P(X = i) = \frac{|A|}{n} .$$

- b) *Bernoulli-Verteilung*
 X mit Träger $\{0, 1\}$ heißt Bernoulli-verteilt mit Parameter $p \in [0, 1]$, wenn

$$P(X = 1) = p \quad \text{und} \quad P(X = 0) = 1 - p.$$

c) *Binomialverteilung*

X heißt binomialverteilt mit Parameter $p \in [0, 1]$ und Träger $\{0, \dots, n\}$, $n \in \mathbb{N}$, wenn

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Bezeichnung: $X \sim \text{Bin}(n, p)$.

d) *Geometrische Verteilung*

X heißt geometrisch verteilt auf \mathbb{N}_0 ($\mathbb{N}_0 = \mathbb{N} \cup \{0\}$) mit Parameter $p \in (0, 1]$, wenn

$$P(X = k) = (1-p)^k p, \quad k \in \mathbb{N}_0.$$

Bezeichnung: $X \sim \text{Geo}(p)$.

e) *Poisson-Verteilung*

X heißt Poisson-verteilt auf \mathbb{N}_0 mit Parameter $\lambda > 0$, wenn

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0.$$

Bezeichnung: $X \sim \text{Poi}(\lambda)$.

2.3.2 Absolut-stetige Zufallsvariable

Definition 2.3.5. Die Funktion $f: \mathbb{R} \rightarrow \mathbb{R}_+$ sei (uneigentlich Riemann-) integrierbar mit

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

Eine Zufallsvariable X heißt *absolut-stetig*, wenn

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt \quad \forall x \in \mathbb{R}.$$

Die Funktion f heißt (*Verteilungs-*) *Dichte* ((*probability density function, pdf*) von X . Die Zufallsvariable X wird oft als Index an der Dichte notiert, also $f = f_X$.

Mit dem Hauptsatz der Differential- und Integralrechnung folgt, dass $f_X(x) = F'_X(x)$ in allen Stetigkeitsstellen x von f .

Beispiel 2.3.6.

a) X heißt *normalverteilt* (*normally distributed*, *Gaussian distributed*), wenn

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Bezeichnung: $X \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$.

Für die Verteilungsfunktion gilt

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

wobei die Verteilungsfunktion der Standardnormalverteilung (mit $\mu = 0$, $\sigma^2 = 1$) oft tabelliert ist.

b) X heißt *gleich- oder rechteckverteilt* (*uniformly distributed*), wenn

$$\begin{aligned} f_X(x) &= \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{sonst} \end{cases} \\ &= \frac{1}{b-a} \cdot \mathbb{I}_{[a,b]}(x), \quad \text{wobei} \quad \mathbb{I}_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}, \quad A \subseteq \mathbb{R}. \end{aligned}$$

Bezeichnung: $X \sim R(a, b)$, $a < b \in \mathbb{R}$.

c) X heißt *exponentialverteilt* (*exponentially distributed*), wenn

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} = \lambda e^{-\lambda x} \cdot \mathbb{I}_{[0,\infty)}(x).$$

Bezeichnung: $X \sim \text{Exp}(\lambda)$, $\lambda > 0$.

d) X heißt *Gamma-verteilt*, wenn

$$\begin{aligned} f_X(x) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \cdot \mathbb{I}_{[0,\infty)}(x). \\ \Gamma(\alpha) &= \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \text{insbesondere gilt } \Gamma(n) = (n-1)! \quad \forall n \in \mathbb{N}. \end{aligned}$$

Bezeichnung: $X \sim \Gamma(\alpha, \lambda)$, $\alpha, \lambda > 0$.

Als Spezialfälle treten auf

- $\alpha = 1 \rightarrow \text{Exp}(\lambda)$ -Verteilung
- $\alpha = n \in \mathbb{N} \rightarrow \text{Erlang-Verteilung, Erl}(n, \lambda)$.

e) X heißt *Rayleigh-verteilt*, wenn

$$f_X(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \cdot \mathbb{I}_{[0,\infty)}(x).$$

Bezeichnung: $X \sim \text{Ray}(\sigma^2)$, $\sigma^2 > 0$.

f) X heißt *Nakagami- m -verteilt*, wenn

$$f_X(x) = \frac{2}{\Gamma(m)} \left(\frac{m}{\alpha}\right)^m x^{2m-1} e^{-\frac{mx^2}{\alpha}} \cdot \mathbb{I}_{[0,\infty)}(x).$$

Bezeichnung: $X \sim \text{Nak}(m, \alpha)$, $m \in \mathbb{N}$, $\alpha > 0$.

Die Berechnung von Wahrscheinlichkeiten mit Hilfe der Dichte f einer Zufallsvariablen X geschieht wie folgt: Sei $I = (a, b]$ ein Intervall. Dann gilt

$$P(X \in I) = P(a < X \leq b) = \int_a^b f(t) dt.$$

Denn

$$\begin{aligned} P(a < X \leq b) &\stackrel{2.1.3^d)}{=} P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a) = \int_{-\infty}^b f(t) dt - \int_{-\infty}^a f(t) dt = \int_a^b f(t) dt. \end{aligned}$$

Da $F(x) = \int_{-\infty}^x f(t) dt$ sogar stetig ist, gilt obige Aussage auch, wenn

$$I = [a, b] \text{ oder } I = [a, b) \text{ oder } I = (a, b).$$

Allgemein gilt für die Wahrscheinlichkeit beliebiger (meßbarer) Mengen $B \subseteq \mathbb{R}$

$$P(X \in B) = \int_B f(t) dt.$$

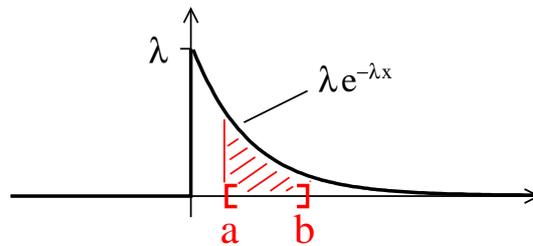


Abbildung 2.3: Veranschaulichung der Berechnung von Wahrscheinlichkeiten mit Dichten (hier: $X \sim \text{Exp}(\lambda)$).

2.3.3 Erwartungswerte und Momente von Zufallsvariablen

Wir beginnen mit einem einfachen motivierenden Beispiel. Die Augenzahl beim symmetrischen Würfelwurf wird modelliert durch die Zufallsvariable $X \sim U(\{1, \dots, 6\})$, d.h. $P(X = k) = \frac{1}{6}$, $k = 1, \dots, 6$.

- Auszahlung = Augenzahl in Euro. Was ist die erwartete oder mittlere Auszahlung?

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5 .$$

- Auszahlung = (Augenzahl)² in Euro = $g(X)$ für die Funktion g mit $g(x) = x^2$.

$$E(X) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \dots + 6^2 \cdot \frac{1}{6} = \frac{91}{6} = 15.1\bar{6} .$$

Definition 2.3.7. (Erwartungswert) Es sei g eine reellwertige Funktion.

- a) X sei eine diskrete Zufallsvariable mit Träger $T = \{x_1, x_2, \dots\}$ und Zähldichte f .

$$E(g(X)) = \sum_{i=1}^{\infty} g(x_i) f(x_i) = \sum_{i=1}^{\infty} g(x_i) P(X = x_i)$$

heißt *Erwartungswert* von $g(X)$, falls $\sum_{i=1}^{\infty} |g(x_i)| f(x_i) < \infty$.

- b) X sei eine absolut-stetige Zufallsvariable mit Dichte f .

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

heißt *Erwartungswert* von $g(X)$, falls $\int_{-\infty}^{\infty} |g(x)| f(x) dx < \infty$.

Mit $g(x) = x$, d.h. $g = \text{id}$, folgt

$$E(X) = \begin{cases} \sum_{i=1}^{\infty} x_i P(X = x_i), & \text{falls } X \text{ diskret} \\ \int_{-\infty}^{\infty} x f(x) dx, & \text{falls } X \text{ absolut-stetig} \end{cases}$$

Eine weitere Methode zur Berechnung des Erwartungswertes ergibt sich wie folgt. Sei $Y = g(X)$ die transformierte Zufallsvariable mit Dichte bzw. Zähldichte f_Y . Dann gilt

$$E(Y) = \begin{cases} \sum_{i=1}^{\infty} y_i f_Y(y_i), & \text{falls } Y \text{ diskret} \\ \int_{-\infty}^{\infty} y f_Y(y) dy, & \text{falls } Y \text{ absolut-stetig} \end{cases}$$

Definition 2.3.8. Seien X, Y Zufallsvariablen, alle auftretenden Erwartungswerte sollen existieren.

- a) $E(X^k)$, $k \in \mathbb{N}_0$, heißt *k-tes Moment* von X ,
 $E((X - E(X))^k)$ heißt *k-tes zentrales Moment* von X .
- b) $\text{Var}(X) = E((X - E(X))^2)$ heißt *Varianz* von X (*variance*).
 $\sqrt{\text{Var}(X)}$ heißt *Standardabweichung* von X (*standard deviation*).
- c) $\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y)))$ heißt *Kovarianz* von X und Y .
 $\text{Corr}(X, Y) = \text{Cov}(X, Y) / \sqrt{\text{Var}(X) \cdot \text{Var}(Y)}$ heißt *Korrelation* von X und Y .

Proposition 2.3.9. (Eigenschaften)

- a) $E(aX + bY) = a E(X) + b E(Y) \quad \forall a, b \in \mathbb{R}$
 (Linearität des Erwartungswertes)
- b) $X \leq Y \implies E(X) \leq E(Y)$
 (Monotonie des Erwartungswertes)
- c) $P(|X| > c) \leq \frac{E(|X|)}{c} \quad \forall c > 0$
 (Markov-Ungleichung)
- d) $\text{Cov}(X, X) = \text{Var}(X), \text{Cov}(X, Y) = E(X \cdot Y) - E(X) E(Y),$
 $\text{Var}(X) = E(X^2) - (E(X))^2$
- e) $\text{Var}(aX + b) = a^2 \text{Var}(X) \geq 0 \quad \forall a, b \in \mathbb{R}$
- f) $\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j)$
- g) $\text{Cov}\left(\sum_{i=1}^n a_i X_i + c, \sum_{j=1}^m b_j Y_j + d\right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(X_i, Y_j)$
- h) $|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \cdot \text{Var}(Y)},$ also $|\text{Corr}(X, Y)| \leq 1$
 (Cauchy-Schwarz Ungleichung)

Die Größen aus Definition 2.3.8 lassen sich folgendermaßen interpretieren:

$E(X)$ mittlerer Wert, Lagemaß

$\text{Var}(X)$ mittlere quadratische Abweichung vom Erwartungswert, Streuungsmaß

$\text{Cov}(X, Y)$ Korrekturterm bei der Berechnung der Varianz von Summen

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$$

$\text{Corr}(X, Y)$ Maß für den linearen Zusammenhang zwischen Zufallsvariablen. Es gilt

$$\exists a, b \in \mathbb{R} : P(Y = aX + b) = 1 \iff |\text{Corr}(X, Y)| = 1.$$

2.4 Zufallsvektoren und Transformationen

Vektoren und Matrizen werden im Folgenden durch Fettdruck gekennzeichnet, \mathbf{A}' bezeichnet die Transponierte einer Matrix oder eines Vektors \mathbf{A} .

Definition 2.4.1. (Zufallsvektor) Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum. Eine Abbildung $\mathbf{X} = (X_1, \dots, X_n)': \Omega \rightarrow \mathbb{R}^n$ heißt *Zufallsvektor*, wenn X_1, \dots, X_n Zufallsvariablen sind.

Ein Zufallsvektor wird aus n Zufallsvariablen auf demselben Wahrscheinlichkeitsraum zusammengesetzt. Das Ziel ist die Beschreibung von gemeinsamen zufälligen Ausgängen, z.B. ein zufälliges Signal $Re^{i\phi}$, wobei R die Amplitude und ϕ die Phase bezeichnet. $\mathbf{X} = (R, \phi)'$ ist dann ein zweidimensionaler Zufallsvektor. Von Interesse ist jetzt die Beschreibung der gemeinsamen Verteilung.

2.4.1 Gemeinsame Verteilung

Definition 2.4.2. Sei $\mathbf{X} = (X_1, \dots, X_n)'$ ein Zufallsvektor. Die Funktion

$$F_{\mathbf{X}}(x_1, \dots, x_n) = P(\{\omega \mid X_1(\omega) \leq x_1, \dots, X_n(\omega) \leq x_n\}) \\ \stackrel{\text{(kurz)}}{=} P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

heißt *gemeinsame Verteilungsfunktion* von $(X_1, \dots, X_n)'$ oder *Verteilungsfunktion* von \mathbf{X} . Bezeichnung: $\mathbf{X} \sim F_{\mathbf{X}}$.

Die gemeinsame Verteilungsfunktion beschreibt die Verteilung des Zufallsvektors \mathbf{X} eindeutig, d.h., die Wahrscheinlichkeiten $P(\mathbf{X} \in A)$ für alle messbaren Ereignisse $A \subseteq \mathbb{R}^n$. Das ist ein tiefliegender Satz der Wahrscheinlichkeitstheorie.

Ein bequemerer Weg um gemeinsame Verteilungen zu beschreiben sind oftmals Dichten.

Definition 2.4.3. Eine (uneigentlich Riemann-) integrierbare Funktion $f_{\mathbf{X}}: \mathbb{R}^n \mapsto \mathbb{R}_+$ heißt *Dichte* von \mathbf{X} (bzw. von $F_{\mathbf{X}}$, der Verteilungsfunktion von \mathbf{X}), wenn

$$F_{\mathbf{X}}(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, \dots, t_n) dt_1 \cdots dt_n \quad \forall x_1, \dots, x_n \in \mathbb{R}.$$

\mathbf{X} (bzw. $F_{\mathbf{X}}$) heißt dann *absolut-stetig* mit Dichte $f_{\mathbf{X}}$. Bezeichnung: $\mathbf{X} \sim f_{\mathbf{X}}$.

Beispiel 2.4.4. (n -dimensionale Normalverteilung, $N_n(\boldsymbol{\mu}, \mathbf{C})$)

Sei $\boldsymbol{\mu} \in \mathbb{R}^n$ und sei $\mathbf{C} \in \mathbb{R}^{n \times n}$ symmetrisch und positiv definit (d.h. $\mathbf{x}'\mathbf{C}\mathbf{x} > 0 \forall \mathbf{x} \neq \mathbf{0}$). Dann lautet die Dichte der n -dimensionalen Normalverteilung

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{C}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}, \quad \mathbf{x} \in \mathbb{R}^n.$$

$|\mathbf{C}| = |\det(\mathbf{C})|$ bezeichnet hierbei den Absolutbetrag der Determinante von \mathbf{C} .

Für den Spezialfall $n = 1$, $C = \sigma^2 > 0$, $\mu \in \mathbb{R}$ gilt

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Es ergibt sich also die Dichte der (eindimensionalen) Normalverteilung wie in Beispiel 2.3.6 a).

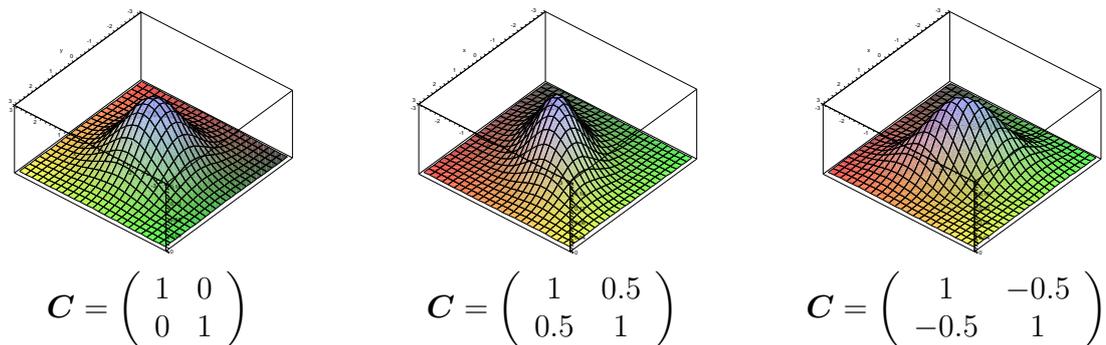


Abbildung 2.4: 2-dimensionale Normalverteilungen mit verschiedenen Kovarianzmatrizen, jeweils mit $\mu = \mathbf{0}$.

In Abbildung 2.4 werden die Dichten von 2-dimensionalen Normalverteilungen mit verschiedenen Kovarianzmatrizen dargestellt. Es ergeben sich Glockenkurven, die symmetrisch oder je nach Struktur der Kovarianzmatrix breit- oder flachgedrückt sind.

Ist $\mathbf{X} = (X_1, \dots, X_n)'$ absolut-stetig mit Dichte $f_{\mathbf{X}}$, so können Wahrscheinlichkeiten von beliebigen (meßbaren) Mengen B des \mathbb{R}^n durch Integration wie folgt berechnet werden

$$P(\mathbf{X} \in B) = \int \cdots \int_B f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

Zum Beispiel erhält man für $B = [a_1, b_1] \times [a_2, b_2]$ speziell

$$P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2) = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f_{\mathbf{X}}(x_1, x_2) dx_1 dx_2$$

oder für $B = \{(x_1, x_2) \mid x_1^2 + x_2^2 \leq 1\}$ (Einheitskreis)

$$P(\mathbf{X} \in B) = P(X_1^2 + X_2^2 \leq 1) = \int \int_{t_1^2 + t_2^2 \leq 1} f_{\mathbf{X}}(t_1, t_2) dt_1 dt_2.$$

2.4.2 Erwartungswertvektor und Kovarianzmatrix

Definition 2.3.7 und Definition 2.3.8 werden nun durch komponentenweise Anwendung auf Zufallsvektoren übertragen.

Definition 2.4.5. Für einen Zufallsvektor $\mathbf{X} = (X_1, \dots, X_n)'$ heißt

$$\text{a) } E(\mathbf{X}) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix} \in \mathbb{R}^n \text{ Erwartungswertvektor von } \mathbf{X},$$

b) die symmetrische Matrix $C(\mathbf{X}) = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq n} \in \mathbb{R}^{n \times n}$ Kovarianzmatrix von \mathbf{X} (Matrix der paarweisen Kovarianzen von X_i und X_j).

Proposition 2.4.6. (Erwartungswert und Kovarianz unter linearen Transformationen)
Sei $\mathbf{X} = (X_1, \dots, X_n)'$ ein Zufallsvektor mit $E(\mathbf{X}) = \boldsymbol{\mu} \in \mathbb{R}^n$, $\text{Cov}(\mathbf{X}) \in \mathbb{R}^{n \times n}$ und $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$ beliebig. Dann gilt

- a) $E(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A}E(\mathbf{X}) + \mathbf{b} \in \mathbb{R}^m$,
 b) $\text{Cov}(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A} \text{Cov}(\mathbf{X}) \mathbf{A}' \in \mathbb{R}^{m \times m}$.

Beweis.

a) $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ bedeutet komponentenweise

$$Y_i = \sum_{k=1}^n a_{ik} X_k + b_i, \quad i = 1, \dots, m.$$

Wegen der Linearität des Erwartungswertes folgt

$$E(Y_i) \stackrel{2.3.9 a)}{=} \sum_{k=1}^n a_{ik} E(X_k) + b_i, \quad \forall i = 1, \dots, m.$$

Also ist

$$E(\mathbf{Y}) = \mathbf{A} E(\mathbf{X}) + \mathbf{b} .$$

b) Mit Proposition 2.3.9 g) gilt

$$\begin{aligned} \text{Cov}(Y_i, Y_j) &= \text{Cov}\left(\sum_{k=1}^n a_{ik} X_k + b_i, \sum_{l=1}^n a_{jl} X_l + b_j\right) \\ &\stackrel{2.3.9 g)}{=} \sum_{k=1}^n \sum_{l=1}^n a_{ik} a_{jl} \text{Cov}(X_k, X_l) \\ &= (\mathbf{A} \text{Cov}(\mathbf{X}) \mathbf{A}')_{i,j} \end{aligned}$$

für alle $1 \leq i, j \leq n$. Insgesamt ist hiermit

$$\text{Cov}(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A} \text{Cov}(\mathbf{X}) \mathbf{A}' .$$

□

Beachten Sie, dass jede Kovarianzmatrix $\mathbf{C} \in \mathbb{R}^{n \times n}$ nicht-negativ definit ist, d.h.

$$\mathbf{a}' \mathbf{C} \mathbf{a} \geq 0 \quad \forall \mathbf{a} \in \mathbb{R}^n .$$

Denn wird in Proposition 2.4.6 b) $\mathbf{A} = \mathbf{a}' \in \mathbb{R}^{1 \times n}$ gewählt, dann gilt mit $\mathbf{a}' = (a_1, \dots, a_n)$

$$0 \leq \text{Var}\left(\sum_{i=1}^n a_i X_i\right) = \text{Var}(\mathbf{a}' \mathbf{X}) = \text{Cov}(\mathbf{a}' \mathbf{X}) = \mathbf{a}' \text{Cov}(\mathbf{X}) \mathbf{a} .$$

2.4.3 Stochastische Unabhängigkeit

Für die Modellierung von stochastischen Effekten ist die stochastische Unabhängigkeit von Zufallsvariablen ein wichtiger Fall.

Definition 2.4.7. Sei $\mathbf{X} = (X_1, \dots, X_n)'$ ein Zufallsvektor. X_1, \dots, X_n heißen *stochastisch unabhängig (s.u.) (stochastically independent, s.i.)*, wenn

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n) \quad \forall x_1, \dots, x_n \in \mathbb{R}.$$

Definition 2.4.7 ist exakt die Aussage von Definition 2.2.4 für alle Ereignisse der Form

$$A_{x_i} = \{\omega \mid X_i(\omega) \leq x_i\} \stackrel{(\text{kurz})}{=} \{X_i \leq x_i\}, \quad x_1, \dots, x_n \in \mathbb{R}.$$

Denn

$$\begin{aligned} F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) &= P(X_1 \leq x_1, \dots, X_n \leq x_n) \\ &= P\left(\bigcap_{i=1}^n \{X_i \leq x_i\}\right) = P\left(\bigcap_{i=1}^n A_{x_i}\right) = \prod_{i=1}^n P(A_{x_i}) \\ &= \prod_{i=1}^n P(X_i \leq x_i) = \prod_{i=1}^n F_{X_i}(x_i). \end{aligned}$$

Die Auswahl von Indizes in Definition 2.2.4 geschieht durch Setzen von $x_l = \infty$ für Komponenten $l \neq i_1, \dots, i_k$.

Sind Zufallsvariablen stochastisch unabhängig und besitzen dieselbe Verteilung, so heißen sie *stochastisch unabhängig identisch verteilt (stid) (independent identically distributed, i.i.d.)*. Mit Integration und der Definition der Verteilungsfunktion ist leicht zu zeigen:

Proposition 2.4.8. Ist $\mathbf{X} = (X_1, \dots, X_n)'$ absolut-stetig, so gilt:

- a) $f_{\mathbf{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n) \quad \forall (x_1, \dots, x_n)' \in \mathbb{R}^n \Rightarrow X_1, \dots, X_n$ sind s.u.
- b) X_1, \dots, X_n sind s.u. $\Rightarrow \prod_{i=1}^n f_{X_i}(x_i)$ ist eine gemeinsame Dichte von \mathbf{X} .

Beispiel 2.4.9. Seien X_1, \dots, X_n stochastisch unabhängig, $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$. Dann ist

$$\begin{aligned} f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} \left(\prod_{i=1}^n \sigma_i^2\right)^{\frac{1}{2}}} e^{-\sum_{i=1}^n \frac{(x_i - \mu_i)^2}{2\sigma_i^2}} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{C}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu})}, \quad \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

mit $\mathbf{C} = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ eine gemeinsame Dichte von $(X_1, \dots, X_n)'$. Damit folgt, dass stochastisch unabhängige, normalverteilte Zufallsvariablen mit obigem $\boldsymbol{\mu}$ und \mathbf{C} einen $N_n(\boldsymbol{\mu}, \mathbf{C})$ -verteilten Zufallsvektor bilden.

Allgemein gilt die folgende Implikation

$$X_1, \dots, X_n \text{ s.u.} \implies \text{Cov}(\mathbf{X}) = \text{diag}(\sigma_1^2, \dots, \sigma_n^2) \text{ mit } \sigma_i^2 = \text{Var}(X_i).$$

Dies sieht man wie folgt ein. Sind X_i und X_j stochastisch unabhängige Zufallsvariablen, so gilt $E(X_i \cdot X_j) = E(X_i) E(X_j)$, also mit Darstellung 2.3.9 d)

$$\text{Cov}(X_i, X_j) = E(X_i \cdot X_j) - E(X_i) E(X_j) = 0.$$

Die Umkehrung ist im Allgemeinen falsch. Sie ist jedoch richtig bei normalverteilten Zufallsvektoren, d.h.

$$\mathbf{X} \sim N_n(\boldsymbol{\mu}, \text{diag}(\sigma_1^2, \dots, \sigma_n^2)) \implies X_1, \dots, X_n \text{ s.u. und } X_i \sim N(\mu_i, \sigma_i^2).$$

Analog zu Proposition 2.4.8 gilt bei diskreten Zufallsvariablen die folgende Proposition.

Proposition 2.4.10. Seien X_1, \dots, X_n diskrete Zufallsvariablen mit Trägern T_1, \dots, T_n . Dann gilt

X_1, \dots, X_n sind stochastisch unabhängig

$$\iff$$

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i) \quad \forall x_i \in T_i, \quad i = 1, \dots, n.$$

2.4.4 Bedingte Verteilungen

Diskrete Zufallsvektoren

Sei (X, Y) ein diskreter Zufallsvektor mit Träger $T_X \times T_Y$ und gemeinsamer Zähldichte

$$f_{(X,Y)}(x, y) = P(X = x, Y = y), \quad (x, y) \in T_X \times T_Y.$$

Die Definition der *bedingten Zähldichte von X unter Y = y* erfolgt wie in Definition 2.2.2 als elementare bedingte Wahrscheinlichkeit.

$$f_{X|Y}(x|y) = \begin{cases} \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{f_{(X,Y)}(x,y)}{f_Y(y)} & , \text{ falls } f_Y(y) > 0 \\ f_X(x) \text{ (oder bel. andere Zähldichte)} & , \text{ falls } f_Y(y) = 0 \end{cases}$$

$f_{X|Y}(x|y)$ ist eine Zähldichte für jedes feste $y \in T_Y$. Die zugehörige Verteilung heißt *bedingte Verteilung von X unter Y = y*. Es gilt

$$P(X \in A | Y = y) = \sum_{x \in A} f_{X|Y}(x|y), \quad A \subseteq T_X, \quad y \in T_Y.$$

Mit dem Satz von der totalen Wahrscheinlichkeit erhält man:

$$f_X(x) = P(X = x) = \sum_{y \in T_Y} f_{X|Y}(x|y) f_Y(y), \quad x \in T_X.$$

Beispiel 2.4.11. Gegeben sei ein diskreter Zufallsvektor (X, N) , für den folgende Annahmen gelten:

$$P(N = n) = e^{-\lambda} \frac{\lambda^n}{n!}, \quad n = 0, 1, 2, \dots \quad (\sim \text{Poi}(\lambda))$$

$$P(X = k | N = n) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n \quad (\sim \text{Bin}(n, p))$$

Interpretation: Die Zufallsvariable N beschreibt die Anzahl in einem bestimmten Zeitraum angekommener Pakete. Diese werden unabhängig mit Wahrscheinlichkeit p Server 1 und mit Wahrscheinlichkeit $1-p$ Server 2 zugewiesen. Die Zufallsvariable X modelliert die Anzahl der Server 1 zugewiesenen Pakete. Für gegebenes $N = n$ ist X $\text{Bin}(n, p)$ -verteilt. Gesucht ist die Wahrscheinlichkeit, dass Server 1 genau k Pakete zugewiesen werden.

$$\begin{aligned} P(X = k) &= \sum_{n=0}^{\infty} P(X = k | N = n) P(N = n) \\ &= \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!} \\ &= e^{-\lambda} \frac{\lambda^k p^k}{k!} \sum_{n=k}^{\infty} \frac{(\lambda(1-p))^{n-k}}{(n-k)!} \\ &= e^{-\lambda} \frac{(\lambda p)^k}{k!} e^{\lambda(1-p)} \\ &= e^{-\lambda p} \frac{(\lambda p)^k}{k!}. \end{aligned}$$

Die Zufallsvariable X ist also $\text{Poi}(\lambda p)$ -verteilt.

Der Erwartungswert bzgl. der bedingten Verteilung lautet

$$E(g(X) | Y = y) = \sum_{x \in T_X} g(x) f_{X|Y}(x|y), \quad y \in T_Y,$$

der *bedingte Erwartungswert von $g(X)$ unter $Y = y$* (falls existent).

$E(g(X))$ kann aus den bedingten Erwartungswerten wie folgt berechnet werden

$$E(g(X)) = \sum_{y \in T_Y} E(g(X) | Y = y) f_Y(y).$$

Absolut-stetige Zufallsvektoren

Im Fall dass (X, Y) absolut stetig mit der Dichte $f_{X,Y}(x, y)$ ist, erfolgt die Definition der bedingten Dichte analog:

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & , \text{ falls } f_Y(y) > 0 \\ f_X(x) \text{ (oder eine bel. andere Dichte)} & , \text{ falls } f_Y(y) = 0 \end{cases} .$$

$f_{X|Y}(x|y)$ ist eine Dichte für jedes feste $y \in \mathbb{R}$. Die zugehörige Verteilung heißt bedingte Verteilung von X unter $Y = y$.

Es gilt für festes $y \in \mathbb{R}$

$$P(X \in A|Y = y) = \int_A f_{X|Y}(x|y)dx, \quad A \in \mathfrak{B}^1.$$

Analog zum diskreten Fall gelten die folgenden Beziehungen:

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f_{X|Y}(x|y)f_Y(y)dy \\ E(g(X)|Y = y) &= \int_{\mathbb{R}} g(X)f_{X|Y}(x|y)dx \\ E(g(X)) &= \int_{\mathbb{R}} E(g(X)|Y = y)f_Y(y)dy \end{aligned}$$

2.4.5 Transformation von Zufallsvektoren

Betrachtet werden im Folgenden Transformationen $T(\mathbf{X}) \in \mathbb{R}^n$ eines absolut-stetigen Zufallsvektors $\mathbf{X} = (X_1, \dots, X_n)'$ mit Dichte $f_{\mathbf{X}}(x_1, \dots, x_n)$. Vorausgesetzt wird, dass

$$M = \{\mathbf{x} \in \mathbb{R}^n \mid f_{\mathbf{X}}(\mathbf{x}) > 0\} \subseteq \mathbb{R}^n$$

eine offene Menge ist und dass $T: M \rightarrow \mathbb{R}^n$ eine injektive Abbildung ist mit

$$\left| \left(\frac{\partial T_i(x_1, \dots, x_n)}{\partial x_j} \right)_{1 \leq i, j \leq n} \right| > 0, \quad \forall (x_1, \dots, x_n) \in M$$

d.h. der Absolutbetrag der Determinante der Jacobi-Matrix ist positiv.

Theorem 2.4.12. (*Transformationssatz für Dichten*) Unter vorherigen Annahmen besitzt der Zufallsvektor $\mathbf{Y} = T(\mathbf{X})$ eine Dichte

$$\begin{aligned} f_{\mathbf{Y}}(y_1, \dots, y_n) &= \frac{1}{\left| \left(\frac{\partial T_i(\mathbf{x})}{\partial x_j} \right) \Big|_{\mathbf{x}=T^{-1}(y_1, \dots, y_n)} \right|} f_{\mathbf{X}}(T^{-1}(y_1, \dots, y_n)) \\ &= \left| \left(\frac{\partial T_i^{-1}(y_1, \dots, y_n)}{\partial y_j} \right) \right| f_{\mathbf{X}}(T^{-1}(y_1, \dots, y_n)), \quad \forall (y_1, \dots, y_n)' \in T(M). \end{aligned}$$

Im Folgenden wird der obige Transformationssatz für Dichten in verschiedenen Situationen angewendet.

Beispiel 2.4.13. (Simulation, Box-Muller-Verfahren zur Erzeugung von normalverteilten Zufallszahlen aus Standardzufallszahlen) Seien X_1, X_2 i.i.d. gleichverteilt über das offene Intervall $(0, 1)$. Dann sind die Zufallsvariablen

$$Y_1 = \sqrt{-2 \ln X_1} \cos(2\pi X_2), \quad Y_2 = \sqrt{-2 \ln X_1} \sin(2\pi X_2)$$

stochastisch unabhängig und jeweils $N(0, 1)$ -verteilt. Dies sieht man wie folgt ein. Die gemeinsame Dichte von $\mathbf{X} = (X_1, X_2)'$ hat die Form

$$f_{\mathbf{X}}(x_1, x_2) = \mathbb{I}_{(0,1)^2}(x_1, x_2).$$

Mit Theorem 2.4.12 und der Transformation $T: (0, 1) \mapsto \mathbb{R}$ mit

$$T(x_1, x_2) = \left(\sqrt{-2 \ln x_1} \cos(2\pi x_2), \sqrt{-2 \ln x_1} \sin(2\pi x_2) \right) = (y_1, y_2)$$

ist

$$J = \left(\frac{\partial T_i(x_1, x_2)}{\partial x_j} \right)_{1 \leq i, j \leq 2} = \begin{pmatrix} -\frac{\cos(2\pi x_2)}{x_1 \sqrt{-2 \ln x_1}} & -2\pi \sqrt{-2 \ln x_1} \sin(2\pi x_2) \\ -\frac{\sin(2\pi x_2)}{x_1 \sqrt{-2 \ln x_1}} & 2\pi \sqrt{-2 \ln x_1} \cos(2\pi x_2) \end{pmatrix}$$

und damit

$$|\det J| = \left| \frac{-2\pi}{x_1} \cos^2(2\pi x_2) - \frac{2\pi}{x_1} \sin^2(2\pi x_2) \right| = \frac{2\pi}{x_1}.$$

Ferner gilt für die Umkehrabbildung $(x_1, x_2) = T^{-1}(y_1, y_2)$

$$-\ln x_1 \cos^2(2\pi x_2) = \frac{y_1^2}{2} \quad \text{und} \quad -\ln x_1 \sin^2(2\pi x_2) = \frac{y_2^2}{2}.$$

Es folgt also $x_1 = e^{-\frac{y_1^2 + y_2^2}{2}}$. Insgesamt ergibt sich mit Theorem 2.4.12

$$f_{\mathbf{Y}}(y_1, y_2) = \frac{1}{2\pi} e^{-\frac{y_1^2 + y_2^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y_1^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y_2^2}{2}}, \quad y_1, y_2 \in \mathbb{R}.$$

Dies ist das Produkt von zwei $N(0, 1)$ -Dichten, woraus die Behauptung folgt.

Theorem 2.4.14. (Summe von Zufallsvariablen) Sei $\mathbf{X} = (X_1, X_2)'$ ein Zufallsvektor mit der Dichte $f_{\mathbf{X}}(x_1, x_2)$. Dann besitzt $Y = X_1 + X_2$ die Dichte

$$f_Y(y) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(t, y - t) dt.$$

Beweis. Setze in Theorem 2.4.12 $T(x_1, x_2) = (x_1, x_1 + x_2)$, also $T^{-1}(y_1, y_2) = (y_1, y_2 - y_1)$. Für die Jacobi-Determinante gilt dann

$$\left| \left(\frac{\partial T_i(x_1, x_2)}{\partial x_j} \right)_{1 \leq i, j \leq 2} \right| = \left| \det \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \right| = 1$$

und die gemeinsame Dichte von $T(\mathbf{X})$ lautet

$$f_{T(\mathbf{X})}(y_1, y_2) = f_{\mathbf{X}}(y_1, y_2 - y_1) .$$

Die *Randdichte* von $Y = X_1 + X_2$ erhält man durch Integration über die 1. Komponente y_1 zu

$$f_Y(y_2) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(y_1, y_2 - y_1) dy_1 .$$

□

Gilt insbesondere X_1, X_2 s.u., also $f_{\mathbf{X}}(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$, so folgt

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(t) \cdot f_{X_2}(y - t) dt .$$

f_Y heißt *Faltung* (*convolution*) der Dichten f_{X_1} und f_{X_2} .

Als Anwendung von Theorem 2.4.14 erhält man die sogenannte *Faltungsstabilität* von Γ - und Normalverteilungen.

Proposition 2.4.15.

a) X_1, X_2 s.u. , $X_1 \sim \Gamma(\alpha, \lambda)$, $X_2 \sim \Gamma(\beta, \lambda)$, $\alpha, \beta, \lambda > 0$. Dann gilt

$$X_1 + X_2 \sim \Gamma(\alpha + \beta, \lambda) .$$

b) X_1, X_2 s.u. , $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 > 0$. Dann gilt

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) .$$

2.5 Mehrwegeausbreitung in Mobilfunkkanälen

Radiowellenausbreitung unterliegt Effekten wie der Reflexion, Beugung und Dämpfung auf verschiedenen Pfaden zum Empfänger, siehe Abbildung 2.5. Dies führt zu zufälligen Dämpfungen, Phasenverschiebungen und Spektrumausweitungen. Ziel dieses Abschnitts ist es, analytische Modelle für diese Effekte herzuleiten.

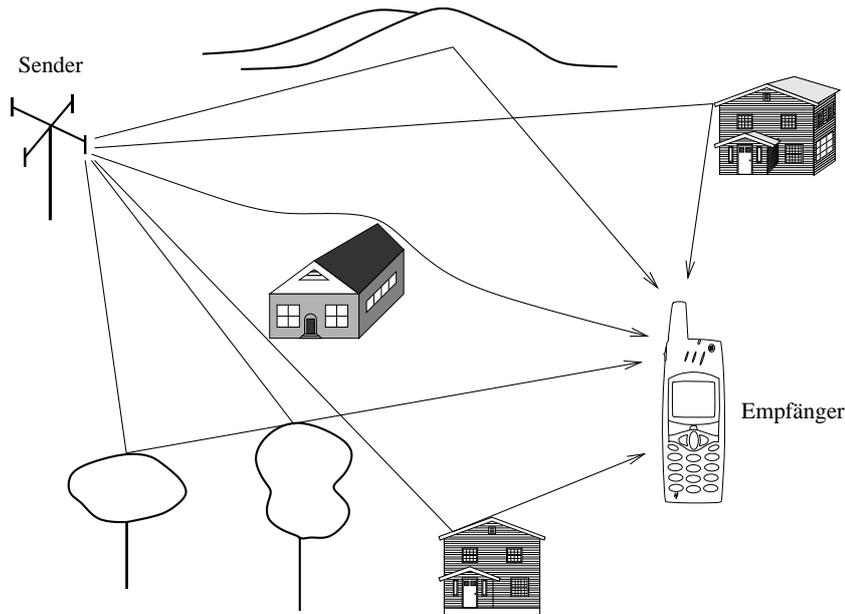


Abbildung 2.5: Mehrwegeausbreitung mit einem festen Sender und einem beweglichen Empfänger (portable receiver).

Unter der Annahme, dass sich der Empfänger mit Geschwindigkeit v bewegt und Signale von n Streupunkten jeweils unter Winkel ϑ_j , $j = 1, \dots, n$, erhält, ergibt sich die Doppler-Verschiebung D_j für Streupunkt j zu

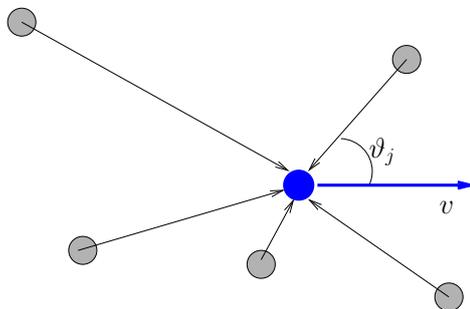


Abbildung 2.6: Modell für Mehrwegeausbreitung

$$D_j = \frac{f}{c} v \cos \vartheta_j,$$

wobei f die Frequenz, c die Lichtgeschwindigkeit und i die komplexe Einheit, $i = \sqrt{-1}$, darstellt, vgl. Abbildung 2.6. Das gesendete Signal lautet für einen unmodulierten Träger $s(t) = e^{i2\pi ft}$. Das empfangene Signal zum Zeitpunkt t über n Streupunkte lautet

$$r(t) = \sum_{j=1}^n A_j e^{i(2\pi f(t + \frac{vt}{c} \cos \vartheta_j) + \Phi_j)},$$

wobei folgende Annahmen gelten:

- $\Phi_j \sim R(0, 2\pi)$ sind zufällige Phasenverschiebungen wegen Reflexion und verschiedener Pfadlängen.
- A_j sind zufällige Pfadverluste (Amplituden) mit
- $E(A_j^2) = \frac{\sigma^2}{n}$, so dass die erwartete Empfangsleistung $\sum_{j=1}^n E(A_j^2) = \sigma^2$ ergibt.
- $A_1, \dots, A_n, \Phi_1, \dots, \Phi_n$ sind gemeinsam stochastisch unabhängig.

Mit

$$c_j = 2\pi f \frac{v}{c} \cos \vartheta_j$$

ergibt sich für das empfangene Signal

$$\begin{aligned} r(t) &= e^{i2\pi ft} \sum_{j=1}^n A_j e^{i(c_j t + \Phi_j)} \\ &= e^{i2\pi ft} \left(\underbrace{\sum_{j=1}^n A_j \cos(c_j t + \Phi_j)}_{=X(t)} + i \underbrace{\sum_{j=1}^n A_j \sin(c_j t + \Phi_j)}_{=Y(t)} \right) \\ &= e^{i2\pi ft} (X(t) + iY(t)). \end{aligned}$$

Für alle c_j, t gilt

- (i) $\cos(c_j t + \Phi_j)$ und $\cos(\Phi_j)$ haben dieselbe Verteilung,
- (ii) $\sin(c_j t + \Phi_j)$ und $\sin(\Phi_j)$ haben dieselbe Verteilung,
- (iii) $E(\cos \Phi_j) = E(\sin \Phi_j) = 0$.

Es folgt

$$\begin{aligned} \mathbb{E}(\sqrt{n}A_j \cos(c_j t + \Phi_j)) &= 0, \\ \mathbb{E}(nA_j^2 \cos^2(c_j t + \Phi_j)) &= \sigma^2 \mathbb{E}(\cos^2 \Phi_j) = \frac{\sigma^2}{2} \text{ und} \\ \text{Var}(\sqrt{n}A_j \cos(c_j t + \Phi_j)) &= \frac{\sigma^2}{2}. \end{aligned}$$

Mit dem zentralen Grenzwertsatz (ZGWS, central limit theorem) gilt weiterhin

$$\begin{aligned} X(t) &= \sum_{j=1}^n A_j \cos(c_j t + \Phi_j) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \sqrt{n}A_j \cos(c_j t + \Phi_j) \\ &\stackrel{\text{as.}}{\approx} \text{N}\left(0, \frac{\sigma^2}{2}\right). \end{aligned}$$

Analog ergibt sich

$$\begin{aligned} Y(t) &= \sum_{j=1}^n A_j \sin(c_j t + \Phi_j) \\ &\stackrel{\text{as.}}{\approx} \text{N}\left(0, \frac{\sigma^2}{2}\right). \end{aligned}$$

Einschub zum zentralen Grenzwertsatz: X_1, X_2, X_3, \dots i.i.d., $\mathbb{E}(X_j) = \mu_j$, $\text{Var}(X_j) = \sigma^2$. Dann gilt

$$\frac{1}{\sqrt{n}\sigma} \sum_{j=1}^n (X_j - \mu_j) \stackrel{\text{as.}}{\approx} \text{N}(0, 1)$$

oder

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j - \mu_j) \stackrel{\text{as.}}{\approx} \text{N}(0, \sigma^2).$$

Hier: $X_j = \sqrt{n}A_j \cos(c_j t + \Phi_j)$, $\mu_j = 0$.

Ferner sind $X(t)$ und $Y(t)$ unkorreliert, denn es gilt

$$\begin{aligned} & \mathbb{E} \left[\left(\sum_{j=1}^n A_j \cos(c_j t + \Phi_j) \right) \cdot \left(\sum_{k=1}^n A_k \sin(c_k t + \Phi_k) \right) \right] \\ &= \sum_{j=1}^n \sum_{k=1}^n \mathbb{E} [A_j A_k \cos(c_j t + \Phi_j) \sin(c_k t + \Phi_k)] \\ &= \sum_{j=1}^n \mathbb{E} \left[A_j^2 \underbrace{\cos(c_j t + \Phi_j) \sin(c_j t + \Phi_j)}_{=\frac{1}{2} \sin(2(c_j t + \Phi_j))} \right] \\ &= \sum_{j=1}^n \frac{\sigma^2}{2n} \mathbb{E}(\sin(2(c_j t + \Phi_j))) = 0. \end{aligned}$$

Als approximatives Modell ergibt sich also (mit ZGWS)

$$r(t) = e^{i2\pi ft}(X(t) + iY(t))$$

mit $X(t), Y(t)$ i.i.d. $\sim N(0, \frac{\sigma^2}{2})$.

Das Signal zur Zeit t (fest) ist also

- zufällig gedämpft durch

$$R = \sqrt{X^2(t) + Y^2(t)} \text{ und}$$

- zufällig phasenverschoben durch

$$\Phi = \angle(X(t) + iY(t)).^1$$

Ziel ist es nun, die gemeinsame Verteilung von (R, Φ) zu bestimmen.

2.5.1 Rayleigh-Verteilung

Im Folgenden sei t fest, schreibe $X = X(t), Y = Y(t)$ und setze $\tau^2 = \frac{\sigma^2}{2}$.

1

$$\angle(x + iy) = \varphi = \begin{cases} \arctan \frac{y}{x} & x > 0, y > 0 \\ \arctan \frac{y}{x} + \pi & x < 0, y > 0 \\ \arctan \frac{y}{x} + 2\pi & x < 0, y < 0 \end{cases}, \text{ wobei } \arctan(\pm\infty) = \pm \frac{\pi}{2}.$$

Das Problem ist also für $X, Y \sim N(0, \tau^2)$ i.i.d. die gemeinsame Verteilung von (R, Φ) zu bestimmen mit

$$R = \sqrt{X^2 + Y^2}, \quad \Phi = \angle(X, Y).$$

Benutze dazu den Transformationssatz für Dichten.

a) Die gemeinsame Dichte von (X, Y) lautet

$$f_{(X,Y)}(x, y) = \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{x^2}{2\tau^2}} \cdot \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{y^2}{2\tau^2}}, \quad (x, y) \in \mathbb{R}^2.$$

b) Mit der Transformation T (in Polarkoordinaten)

$$(r, \varphi) = T(x, y) = \left(\sqrt{x^2 + y^2}, \angle(x, y) \right)$$

und deren Inversen

$$T^{-1}(r, \varphi) = (r \cos \varphi, r \sin \varphi)$$

ergibt sich die Jacobi-Determinante von T^{-1} zu

$$\left| \begin{pmatrix} \frac{\partial T_1^{-1}}{\partial r} & \frac{\partial T_1^{-1}}{\partial \varphi} \\ \frac{\partial T_2^{-1}}{\partial r} & \frac{\partial T_2^{-1}}{\partial \varphi} \end{pmatrix} \right| = \left| \begin{pmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{pmatrix} \right| = |r(\cos^2 \varphi + \sin^2 \varphi)| = |r|.$$

Die Dichte von (R, Φ) ergibt sich gemäß Theorem 2.4.12 zu

$$\begin{aligned} f_{(R,\Phi)}(r, \varphi) &= \frac{1}{2\pi\tau^2} e^{-\frac{-r^2 \cos^2 \varphi + r^2 \sin^2 \varphi}{2\tau^2}} \cdot r \quad r > 0, \quad 0 < \varphi < 2\pi \\ &= \frac{r}{\tau^2} e^{-\frac{r^2}{2\tau^2}} \mathbb{I}_{(0,\infty)}(r) \cdot \frac{1}{2\pi} \mathbb{I}_{(0,2\pi)}(\varphi). \end{aligned}$$

Die gemeinsame Dichte von (R, Φ) ist also das Produkt von zwei Dichten

$$f_R(r) = \frac{r}{\tau^2} e^{-\frac{r^2}{2\tau^2}} \mathbb{I}_{(0,\infty)}(r)$$

und

$$f_\Phi(\varphi) = \frac{1}{2\pi} \mathbb{I}_{(0,2\pi)}(\varphi).$$

Die zugehörigen Zufallsvariablen R und Φ sind damit stochastisch unabhängig. Die Verteilung von R heißt *Rayleigh-Verteilung*, die von Φ ist eine Gleichverteilung auf dem Intervall $[0, 2\pi]$. Abkürzend schreiben wir $R \sim \text{Ray}(\tau^2)$ und $\Phi \sim \text{R}(0, 2\pi)$. Ferner gilt

$$\mathbb{E}(R) = \sqrt{\frac{\pi}{2}} \tau \quad \text{und} \quad \text{Var}(R) = \left(2 - \frac{\pi}{2}\right) \tau^2.$$

Wenn $R \sim \text{Ray}(\tau^2)$, so besitzt die Zufallsvariable $Z = R^2$ eine Dichte

$$f_Z(z) = \frac{1}{2\tau^2} e^{-\frac{z}{2\tau^2}} \mathbb{I}_{(0,\infty)}(z).$$

Also ist $Z \sim \text{Exp}(\frac{1}{2\tau^2})$ bzw. $Z \sim \text{Exp}(\frac{1}{\sigma^2})$. Die Empfangsleistung eines Rayleigh-Fading-Signals ist also exponentialverteilt mit Parameter $\frac{1}{\sigma^2}$, wobei σ^2 die erwartete Empfangsleistung ist.

In Abbildung 2.7 sind Beispiele der Rayleigh-Verteilung für verschiedene Parameter dargestellt.

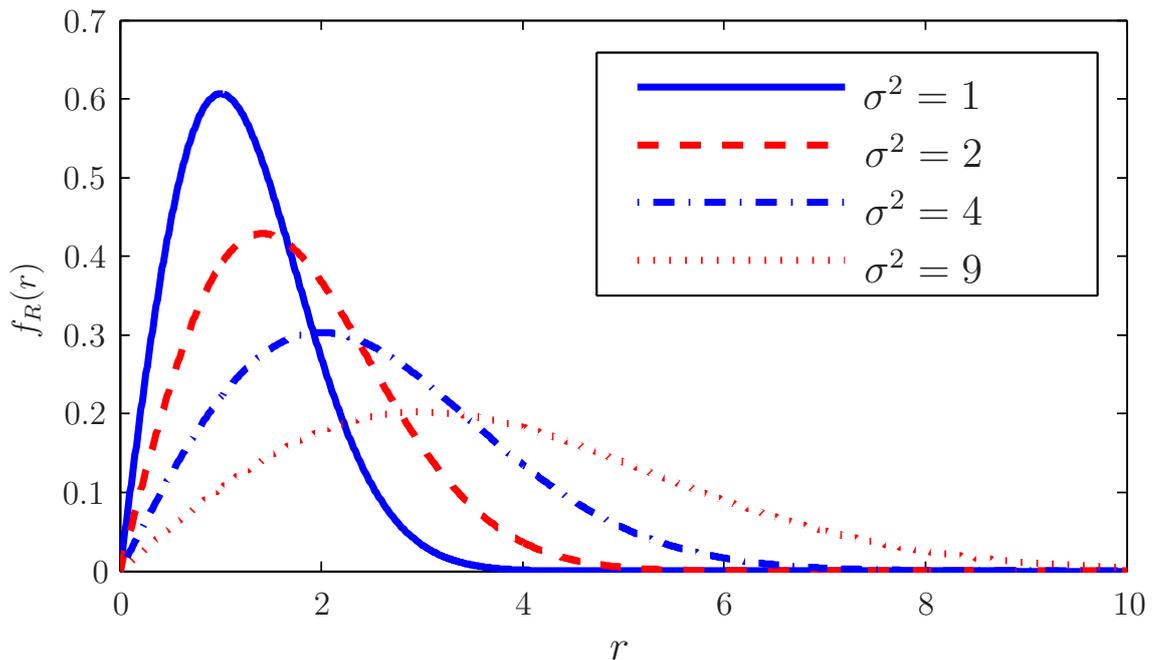


Abbildung 2.7: Dichten der Rayleigh-Verteilung.

Im Rayleigh-Fading-Modell wird keine Einfallrichtung oder Phasenverschiebung besonders ausgezeichnet. Daher gilt $E(X) = E(Y) = 0$. Existiert eine direkte Sichtverbindung (LoS - Line of Sight), gilt i.a. $E(X) = \mu_X$, $E(Y) = \mu_Y$ mit $\mu_X, \mu_Y \neq 0$. Dies führt zur Rice-Verteilung, die im nächsten Abschnitt behandelt wird.

2.5.2 Rice-Verteilung

Die Rice-Verteilung ist ein Fading-Modell für den Fall, dass eine direkte Sichtverbindung zwischen Sender und Empfänger vorliegt.

Seien dazu $X \sim N(\mu_1, \sigma^2)$, $Y \sim N(\mu_2, \sigma^2)$ und X, Y stochastisch unabhängig. Als gemeinsame Dichte ergibt sich durch Produktbildung

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}((x-\mu_1)^2 + (y-\mu_2)^2)}, \quad x, y \in \mathbb{R}.$$

Erfolgt eine Transformation auf Polarkoordinaten (r, φ) wie in Abschnitt 2.5.1, so besitzt $(R, \phi) = T(X, Y)$ für $r \geq 0$, $0 \leq \varphi \leq 2\pi$ die Dichte

$$f_{(R,\phi)}(r, \varphi) = \frac{r}{2\pi\sigma^2} \exp \left\{ -\frac{(r \cos \varphi - \mu_1)^2 + (r \sin \varphi - \mu_2)^2}{2\sigma^2} \right\}.$$

Wird $\mu = \sqrt{\mu_1^2 + \mu_2^2}$ mit $\mu_1 = \mu \cdot \cos \alpha$ und $\mu_2 = \mu \cdot \sin \alpha$ gesetzt (vergleichen Sie dazu auch Abbildung 2.8), ergibt sich die Dichte

$$f_{(R,\phi)}(r, \varphi) = \frac{r}{2\pi\sigma^2} \exp \left\{ -\frac{r^2 + \mu^2}{2\sigma^2} \right\} \exp \left\{ \frac{r\mu}{\sigma^2} (\cos \alpha \cos \varphi + \sin \alpha \sin \varphi) \right\}.$$

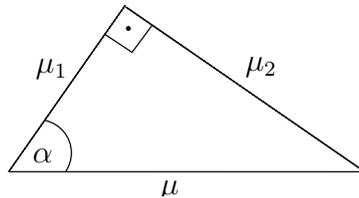


Abbildung 2.8: Kanten und Winkel im rechtwinkligen Dreieck, $\mu = \sqrt{\mu_1^2 + \mu_2^2}$, $\cos \alpha = \frac{\mu_1}{\mu}$, $\sin \alpha = \frac{\mu_2}{\mu}$.

Durch Integration über φ erhält man die Dichte von R zu

$$f_R(r) = \frac{r}{2\pi\sigma^2} \exp \left\{ -\frac{r^2 + \mu^2}{2\sigma^2} \right\} \int_0^{2\pi} \exp \left\{ \frac{r\mu}{\sigma^2} (\cos \alpha \cos \varphi + \sin \alpha \sin \varphi) \right\} d\varphi.$$

Mit $\cos \alpha \cos \varphi + \sin \alpha \sin \varphi = \cos(\alpha - \varphi)$ folgt für das Integral

$$\begin{aligned} \int_0^{2\pi} \exp \left\{ \frac{r\mu}{\sigma^2} \cos(\alpha - \varphi) \right\} d\varphi &= \int_0^{2\pi} \exp \left\{ \frac{r\mu}{\sigma^2} \cos \varphi \right\} d\varphi \\ &= 2 \int_0^{\pi} \exp \left\{ \frac{r\mu}{\sigma^2} \cos \varphi \right\} d\varphi. \end{aligned}$$

Die Funktion

$$I_0(x) = \frac{1}{\pi} \int_0^{\pi} e^{x \cos \vartheta} d\vartheta$$

heißt *modifizierte Bessel-Funktion 1. Art der Ordnung 0*. Allgemein haben die modifizierten Bessel-Funktionen 1. Art der Ordnung k die Gestalt

$$I_k(x) = \frac{1}{\pi} \int_0^{\pi} e^{x \cos \vartheta} \cos(k\vartheta) d\vartheta.$$

Insgesamt lässt sich die Dichte von R damit schreiben als

$$f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2 + \mu^2}{2\sigma^2}} I_0\left(\frac{r\mu}{\sigma^2}\right), \quad r \geq 0.$$

Die zugehörige Verteilung heißt *Rice-Verteilung* mit Parametern $\mu = \sqrt{\mu_1^2 + \mu_2^2}$ und σ^2 . Sie beschreibt zufällige schnelle Fading-Effekte, wenn neben der Mehrwegeausbreitung eine direkte Sichtverbindung zwischen Sender und Empfänger vorliegt. Der Vektor $(\mu_1, \mu_2)'$ korrespondiert hierbei zur Phasenverschiebung durch die dominierende Wellenfront.

Bezeichnung: Rice(μ, σ^2).

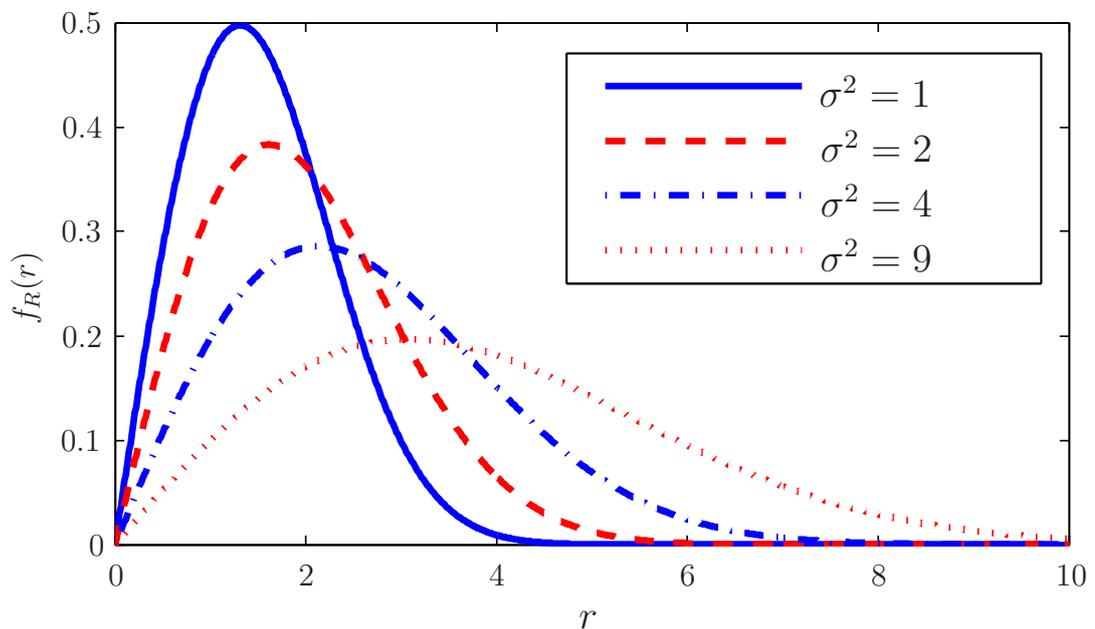


Abbildung 2.9: Dichten der Rice-Verteilung für $\mu = 1$.

2.5.3 Lognormal-Verteilung

Im Folgenden werden Fading-Effekte betrachtet, die durch die Bewegung einer Mobilstation hervorgerufen werden. Die hierdurch verursachten Effekte sind relativ langsam verglichen mit den schnellen Fading-Effekten der vorhergehenden Abschnitte. Eine weitere Ursache für langsames Fading sind sich bewegende Fahrzeuge.

Ergebnisse von Experimenten legen nahe, die langsame Variation des mittleren Pegels durch Multiplikation mit der Zufallsvariablen

$$L = 10^{\frac{G}{10}} \quad \text{mit} \quad G \sim N(\mu_G, \sigma_G^2) \quad (*)$$

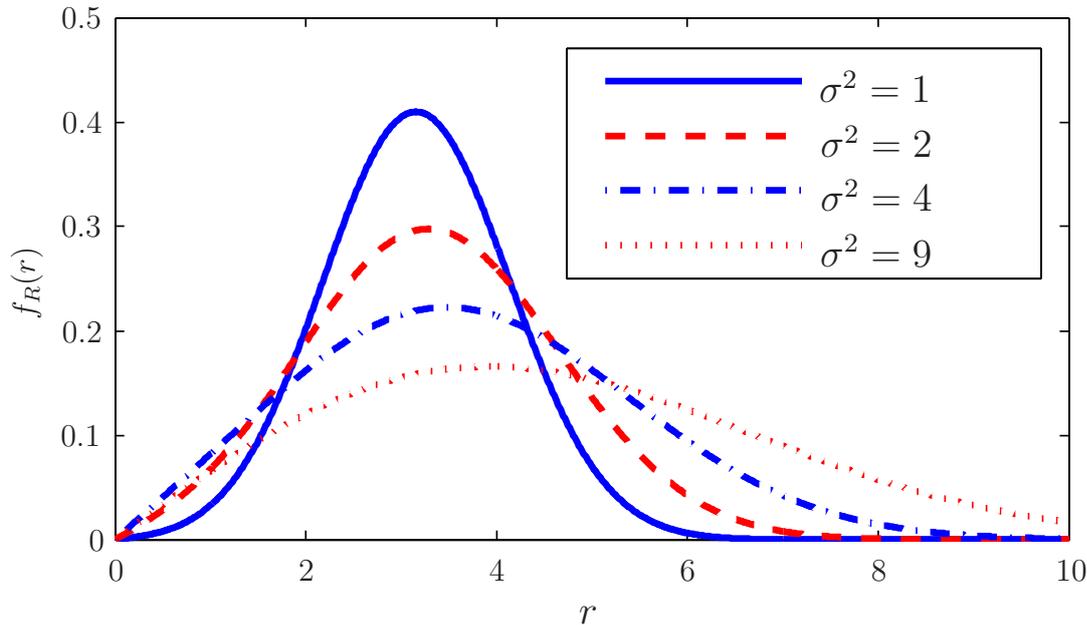


Abbildung 2.10: Dichten der Rice-Verteilung für $\mu = 3$.

zu beschreiben. Typischerweise gilt hierbei für städtische Szenarien $7 \leq \sigma_G \leq 12$. Dies führt auf das Problem, die Verteilung von

$$Y = e^X \quad \text{mit} \quad X \sim N(\mu, \sigma^2)$$

zu bestimmen. Dazu benutzt man Theorem 2.4.12 mit $n = 1$ und

$$T(x) = e^x, \quad T^{-1}(y) = \ln y, \quad T^{-1'}(y) = \frac{1}{y}, \quad y > 0.$$

Es ergibt sich

$$f_Y(y) = \frac{1}{y\sqrt{2\pi\sigma}} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}}, \quad y > 0.$$

Dies ist die Dichte der *Lognormal-Verteilung* mit Parametern μ und σ^2 .
Bezeichnung: $\text{LogN}(\mu, \sigma^2)$.

Eine Zufallsvariable Y ist also lognormal-verteilt, wenn $X = \ln Y$ normalverteilt ist, wie man leicht an (*) ablesen kann.

2.6 Die n -dimensionale komplexe Normalverteilung

Die n -dimensionale Normalverteilung besitzt eine Dichte, wenn die Kovarianzmatrix regulär ist. Wichtig sind aber auch Fälle, in denen sich die Verteilung auf Unterräume konzentriert. Dann existiert keine Dichte, jedenfalls nicht im hier definierten Sinn. Es

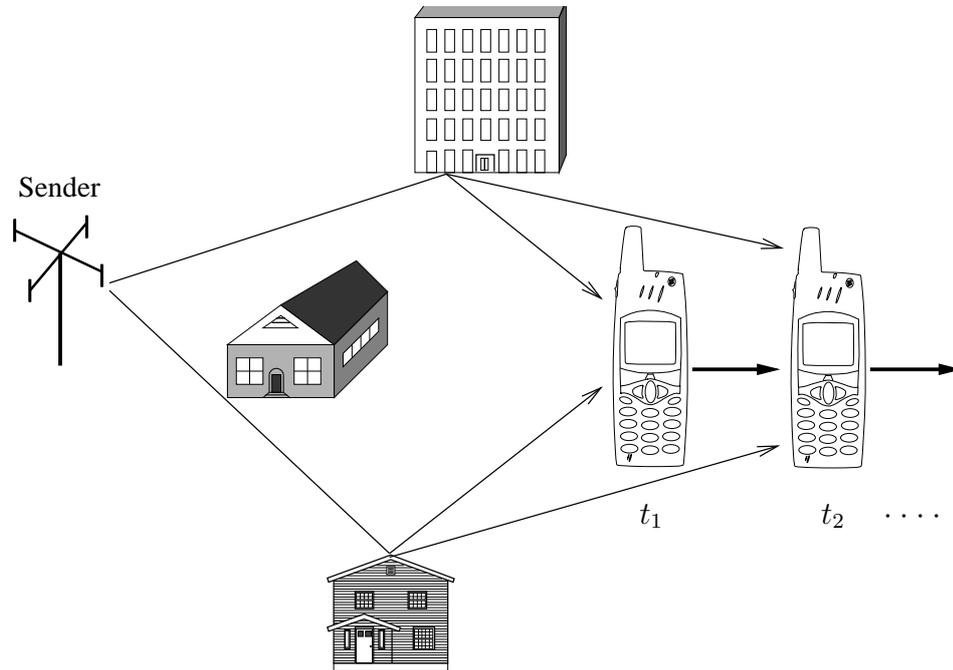


Abbildung 2.11: Langsames Fading durch verschiedene Positionierung der Mobilstation.

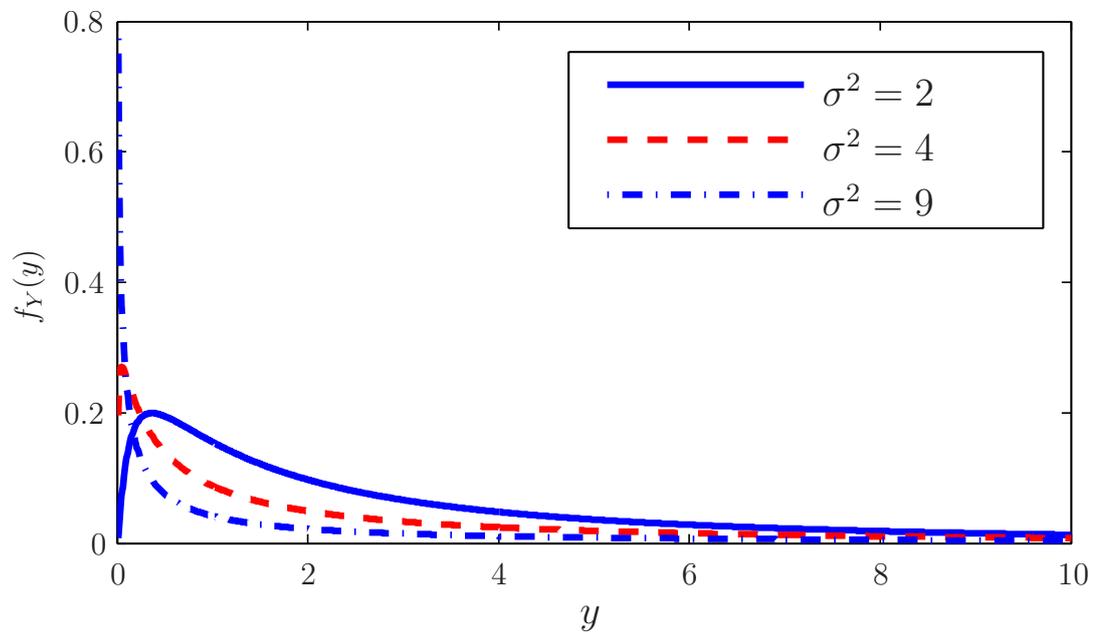


Abbildung 2.12: Dichten der Lognormal-Verteilung für $\mu = 1$.

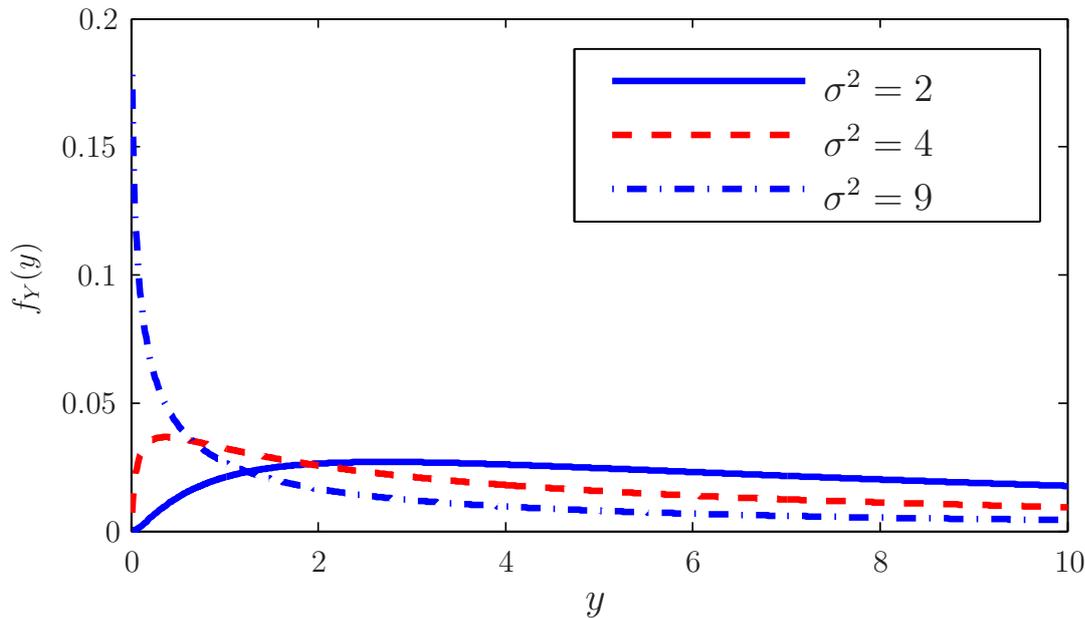


Abbildung 2.13: Dichten der Lognormal-Verteilung für $\mu = 3$.

gilt aber Folgendes.

Sei $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \mathbf{C})$, \mathbf{C} positiv definit, dann gilt

$\mathbf{a}'\mathbf{X} \sim N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\mathbf{C}\mathbf{a})$, falls $\mathbf{a} \neq \mathbf{0}$, oder $P(\mathbf{a}'\mathbf{X} = \mathbf{0}) = 1$, falls $\mathbf{a} = \mathbf{0}$.

Dies führt zu folgender, mit Beispiel 2.4.4 konsistenter Definition.

Definition 2.6.1. Ein reeler Zufallsvektor $\mathbf{X} = (X_1, \dots, X_n)'$ heißt *n-dimensional normalverteilt*, wenn $\mathbf{a}'\mathbf{X}$ für alle $\mathbf{a} \in \mathbb{R}^n$ eindimensional normalverteilt oder einpunktverteilt ist. Dabei heißt eine Zufallsvariable X *einpunktverteilt*, wenn ein $x \in \mathbb{R}$ existiert mit

$$P(X = x) = 1.$$

Es folgt die Erweiterung auf komplexe Normalverteilungen.

Definition 2.6.2. Ein komplexer Zufallsvektor $\mathbf{X} = \mathbf{U} + i\mathbf{V} \in \mathbb{C}^n$ heißt *n-dimensional komplex normalverteilt (complex Gaussian distributed)*, wenn der zusammengesetzte reelle Zufallsvektor

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \in \mathbb{R}^{2n}$$

$2n$ -dimensional normalverteilt ist.

Bezeichnung: $\mathbf{X} \sim \text{CN}$.

Ein wichtiger Spezialfall ist die Verteilung eines komplexen Zufallsvektors, der invariant gegen Drehungen im Komplexen ist. Dies ist auch für allgemeine Verteilungen interessant, nicht nur für normalverteilte.

Definition 2.6.3. Ein komplexer Zufallsvektor \mathbf{X} mit Erwartungswert $E(\mathbf{X}) = \boldsymbol{\mu}$ heißt *zirkulär symmetrisch (circular symmetric) komplex verteilt um $\boldsymbol{\mu}$* , falls für jedes $\theta \in [0, 2\pi)$ gilt

$$(\mathbf{X} - \boldsymbol{\mu})e^{i\theta} \sim (\mathbf{X} - \boldsymbol{\mu}).$$

Proposition 2.6.4. Sei $\mathbf{X} = \mathbf{U} + i\mathbf{V}$ ein n -dimensional komplex normalverteilter Zufallsvektor mit Erwartungswert $E(\mathbf{X}) = \boldsymbol{\mu}$. Der Zufallsvektor \mathbf{X} ist zirkulär symmetrisch komplex verteilt genau dann, wenn eine hermitesche Matrix $\mathbf{Q} \in \mathbb{C}^{n \times n}$ existiert, so dass

$$\text{Cov} \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \text{Re } \mathbf{Q} & -\text{Im } \mathbf{Q} \\ \text{Im } \mathbf{Q} & \text{Re } \mathbf{Q} \end{pmatrix}.$$

Der Zufallsvektor \mathbf{X} heißt dann *zirkulär symmetrisch komplex normalverteilt*.
Bezeichnung: $\mathbf{X} \sim \text{SCN}(\boldsymbol{\mu}, \mathbf{Q})$.

Beachten Sie bei Proposition 2.6.4 die folgenden Punkte:

a)

$$\mathbf{Q} = (q_{ij})_{1 \leq i, j \leq n} = \mathbf{Q}^* \text{ also } q_{ji} = q_{ij}^* \\ \text{d.h.}$$

$$\text{Re}(q_{ji}) = \text{Re}(q_{ij}), \text{Im}(q_{ji}) = -\text{Im}(q_{ij}) \text{ und } \text{Im}(q_{ii}) = 0.$$

$$\begin{pmatrix} \text{Re } \mathbf{Q} & -\text{Im } \mathbf{Q} \\ \text{Im } \mathbf{Q} & \text{Re } \mathbf{Q} \end{pmatrix} \text{ ist also reell und symmetrisch.}$$

b) Falls \mathbf{X} zirkulär symmetrisch komplex normalverteilt ist, gilt

$$E[(\mathbf{X} - E(\mathbf{X}))(\mathbf{X} - E(\mathbf{X}))^*] = \mathbf{Q}.$$

Zirkulär symmetrische komplexe Gaußverteilungen sind eindeutig durch den Erwartungswertvektor $E(\mathbf{X}) = \boldsymbol{\mu}$ und die Kovarianzmatrix \mathbf{Q} bestimmt.

Proposition 2.6.5. Der Zufallsvektor \mathbf{X} sei n -dimensional zirkulär symmetrisch komplex normalverteilt mit Kovarianzmatrix \mathbf{Q} . Die Matrix \mathbf{Q} ist dann nicht-negativ definit (n.n.d.), das heißt für jeden komplexen Vektor $\mathbf{x} \in \mathbb{C}^n$ gilt

$$\mathbf{x}^* \mathbf{Q} \mathbf{x} \in \mathbb{R} \quad \text{und} \quad \mathbf{x}^* \mathbf{Q} \mathbf{x} \geq 0.$$

Beispiel 2.6.6. Sei $X = U + iV$ komplex normalverteilt mit $E(X) = 0$, d.h.

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim N_2(\mathbf{0}, \mathbf{C}), \quad \mathbf{C} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

Sei $X \sim \text{SCN}$, d.h. $\mathbf{C} = \frac{1}{2} \begin{pmatrix} \operatorname{Re} Q & -\operatorname{Im} Q \\ \operatorname{Im} Q & \operatorname{Re} Q \end{pmatrix}$ für ein $Q \in \mathbb{C}$, dann gilt

$$\sigma_1^2 = \frac{1}{2} \operatorname{Re} Q = \sigma_2^2 \quad \text{und} \quad \sigma_{12} = -\frac{1}{2} \operatorname{Im} Q = \frac{1}{2} \operatorname{Im} Q = 0 .$$

Es folgt, dass $Q = \sigma^2$ reell ist und damit gilt

$$U, V \text{ i.i.d. } \sim \text{N} \left(0, \frac{\sigma^2}{2} \right) .$$

Ferner gilt

$$\text{E}(X \cdot X^*) = \text{E}(U^2) + \text{E}(V^2) = \frac{\sigma^2}{2} + \frac{\sigma^2}{2} = \sigma^2 .$$

Im Weiteren gelten für $\mathbf{a} \in \mathbb{C}^n$ und $\mathbf{A} \in \mathbb{C}^{n \times n}$ die Bezeichnungen

$$\hat{\mathbf{a}} = \begin{pmatrix} \operatorname{Re}(\mathbf{a}) \\ \operatorname{Im}(\mathbf{a}) \end{pmatrix} \quad \text{und} \quad \hat{\mathbf{A}} = \begin{pmatrix} \operatorname{Re} \mathbf{A} & -\operatorname{Im} \mathbf{A} \\ \operatorname{Im} \mathbf{A} & \operatorname{Re} \mathbf{A} \end{pmatrix} .$$

Die folgenden Eigenschaften lassen sich leicht zeigen:

- a) $\hat{\mathbf{A}} \hat{\mathbf{x}} = \widehat{\mathbf{A} \mathbf{x}}$,
- b) $\hat{\mathbf{x}}' \hat{\mathbf{y}} = \operatorname{Re}(\mathbf{x}^* \mathbf{y})$,
- c) $\hat{\mathbf{A}}^{-1} = \widehat{\mathbf{A}^{-1}}$,
- d) $\det \hat{\mathbf{A}} = |\det \mathbf{A}|^2$.

Proposition 2.6.7. Sei $\mathbf{X} \sim \text{SCN}(\boldsymbol{\mu}, \mathbf{Q})$, $\boldsymbol{\mu} \in \mathbb{C}^n$, $\mathbf{Q} \in \mathbb{C}^{n \times n}$, \mathbf{Q} regulär, hermitesch, nicht-negativ definit. Dann besitzt \mathbf{X} eine Dichte

$$\begin{aligned} f(\mathbf{x}) &= (\det(\pi \mathbf{Q}))^{-1} \exp\{-(\mathbf{x} - \boldsymbol{\mu})^* \mathbf{Q}^{-1} (\mathbf{x} - \boldsymbol{\mu})\} \\ &= \left(\det(\pi \hat{\mathbf{Q}}) \right)^{-\frac{1}{2}} \exp\left\{ -(\hat{\mathbf{x}} - \hat{\boldsymbol{\mu}})' \hat{\mathbf{Q}}^{-1} (\hat{\mathbf{x}} - \hat{\boldsymbol{\mu}}) \right\} = f(\hat{\mathbf{x}}) . \end{aligned}$$

Beweis. Die Dichte der reellen m -dimensionalen Normalverteilung mit Erwartungswert $\boldsymbol{\nu}$ und Kovarianzmatrix \mathbf{C} lautet (siehe Beispiel 2.4.4)

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{m}{2}} |\mathbf{C}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\nu})' \mathbf{C}^{-1} (\mathbf{y} - \boldsymbol{\nu}) \right\}, \quad \mathbf{y} \in \mathbb{R}^m .$$

Für $\mathbf{X} = \mathbf{U} + i\mathbf{V}$, $\mathbf{x} = \mathbf{u} + i\mathbf{v}$, $\boldsymbol{\mu} = \boldsymbol{\mu}_1 + i\boldsymbol{\mu}_2$, $\mathbf{Q} = \mathbf{R} + i\mathbf{S}$ gilt

$$\begin{aligned}
 & f_{(\mathbf{U}, \mathbf{V})}(\mathbf{u}, \mathbf{v}) \\
 &= \frac{1}{(2\pi)^{\frac{2n}{2}} \left| \frac{1}{2} \begin{pmatrix} \mathbf{R} & -\mathbf{S} \\ \mathbf{S} & \mathbf{R} \end{pmatrix} \right|^{\frac{1}{2}}} e^{-\frac{1}{2}((\mathbf{u}', \mathbf{v}') - (\boldsymbol{\mu}'_1, \boldsymbol{\mu}'_2)) \left[\frac{1}{2} \begin{pmatrix} \mathbf{R} & -\mathbf{S} \\ \mathbf{S} & \mathbf{R} \end{pmatrix} \right]^{-1} \left(\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \right)} \\
 &= \left(\det(\pi \hat{\mathbf{Q}}) \right)^{-\frac{1}{2}} \exp \left\{ -(\hat{\mathbf{x}} - \hat{\boldsymbol{\mu}})' \hat{\mathbf{Q}}^{-1} (\hat{\mathbf{x}} - \hat{\boldsymbol{\mu}}) \right\} \\
 &\stackrel{\text{c),d)}}{=} (|\det(\pi \mathbf{Q})|^2)^{-\frac{1}{2}} \exp \left\{ -(\hat{\mathbf{x}} - \hat{\boldsymbol{\mu}})' \underbrace{\widehat{\mathbf{Q}}^{-1}}_{\substack{\text{a) } \widehat{\mathbf{Q}}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \\ \text{b) } \operatorname{Re}\{(\mathbf{x} - \boldsymbol{\mu})^* \mathbf{Q}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}}} (\hat{\mathbf{x}} - \hat{\boldsymbol{\mu}}) \right\} = f(\hat{\mathbf{x}}) \\
 &= (\det(\pi \mathbf{Q}))^{-1} \exp \left\{ -(\mathbf{x} - \boldsymbol{\mu})^* \mathbf{Q}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\} = f(\mathbf{x}), \\
 &\text{da } \mathbf{Q}^{-1} \text{ hermitesch (das folgt aus } \mathbf{Q} \text{ regulär und hermitesch).}
 \end{aligned}$$

□

Proposition 2.6.8.

- a) Sei $\mathbf{X} \sim \text{SCN}(\boldsymbol{\mu}, \mathbf{Q})$, $\mathbf{A} \in \mathbb{C}^{m \times n} \implies \mathbf{A}\mathbf{X} \sim \text{SCN}(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\mathbf{Q}\mathbf{A}^*)$.
- b) Seien $\mathbf{X} \sim \text{SCN}(\boldsymbol{\mu}_1, \mathbf{Q}_1)$ und $\mathbf{Y} \sim \text{SCN}(\boldsymbol{\mu}_2, \mathbf{Q}_2)$ stochastisch unabhängig $\implies \mathbf{X} + \mathbf{Y} \sim \text{SCN}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \mathbf{Q}_1 + \mathbf{Q}_2)$.

2.7 Stochastische Prozesse

Im Folgenden werden zufällige Folgen oder Kurven von Messwerten behandelt. Das können zum Beispiel zufällige Amplituden oder Phasen als Funktion der Zeit oder der zufällige Kursverlauf einer Aktie sein.

Als Modell dient hierbei ein gegebener *Wahrscheinlichkeitsraum* $(\Omega, \mathfrak{A}, P)$, in dem jedem $\omega \in \Omega$ eine zufällige Funktion $X(t, \omega), t \in T, T \subseteq \mathbb{R}$ zugeordnet wird. Eine grafische Veranschaulichung liefert Abbildung 2.14.

Definition 2.7.1. Sei $T \subseteq \mathbb{R}$. Eine Familie von reell- oder komplexwertigen Zufallsvariablen $\{X(t) | t \in T\}$ auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ heißt *stochastischer Prozess*.

Bei $X(t, \omega) : T \times \Omega \rightarrow \mathbb{R}(\mathbb{C})$ handelt es sich also um eine Zufallsvariable für ein festes $t \in T$ und um eine Funktion von t für ein festes $\omega \in \Omega$.

Für je n Zeitpunkte $t_1, \dots, t_n \in T$ erhält man die gemeinsame Verteilung von $(X(t_1), \dots, X(t_n))'$ durch die gemeinsame Verteilungsfunktion $F_{(X(t_1), \dots, X(t_n))}(x_1, \dots, x_n)$.

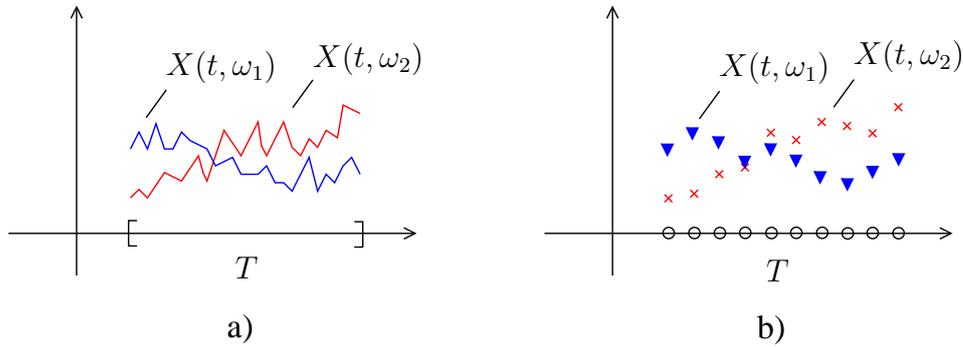


Abbildung 2.14: Verlauf zufälliger Funktionen für kontinuierliches T in a) und diskretes T in b).

Definition 2.7.2. Gegeben sei ein stochastischer Prozess $\{X(t)|t \in T\}$, $t, t_1, t_2 \in T$. Dann heißen

$$\mu_X(t) = E(X(t)) \quad \text{Erwartungswert (-funktion)} \quad ,$$

$$R_{XX}(t_1, t_2) = E(X(t_1)X^*(t_2)) \quad \text{Autokorrelationsfunktion} \quad ,$$

$$C_{XX}(t_1, t_2) = R_{XX}(t_1, t_2) - \mu_X(t_1) \cdot \mu_X^*(t_2) \quad \text{Autokovarianzfunktion} \quad .$$

Beachten Sie, dass für komplexwertige $X(t) = U(t) + iV(t)$ gilt

$$E(X(t)) = E(U(t)) + i E(V(t)) \quad .$$

Außerdem muss berücksichtigt werden, dass in Definition 2.7.2 implizit die Existenz der Erwartungswerte angenommen wird.

Proposition 2.7.3. Es gelten die folgenden Eigenschaften:

a) $R_{XX}(t_1, t_2) = R_{XX}^*(t_2, t_1)$

b) $R_{XX}(t, t) = E(|X(t)|^2) \geq 0$

c) R_{XX} ist eine nicht-negativ definite Funktion, d.h.

$$\forall t_1, \dots, t_n \in T, a_1, \dots, a_n \in \mathbb{C} : \sum_{i,j} a_i a_j^* R_{XX}(t_i, t_j) \geq 0 ,$$

d.h. die Matrix $(R_{XX}(t_i, t_j))_{1 \leq i, j \leq n}$ ist nicht-negativ definit (n.n.d.).

Beweis.

a) $R_{XX}(t_1, t_2) = E(X(t_1)X^*(t_2)) = E[X(t_2)X^*(t_1)]^* = [E(X(t_2)X^*(t_1))]^*$
 $= R_{XX}^*(t_2, t_1)$.

b) $R_{XX}(t, t) = E(X(t)X^*(t)) = E(|X(t)|^2) \geq 0$.

c) Wird $Y = \sum_{i=1}^n a_i X(t_i)$ gesetzt, dann gilt

$$0 \leq E(|Y|^2) = E(YY^*) = E\left(\sum_{i,j} a_i a_j^* X(t_i) X^*(t_j)\right)$$

$$= \sum_{i,j} a_i a_j^* E(X(t_i) X^*(t_j)) = \sum_{i,j} a_i a_j^* R_{XX}(t_i, t_j) .$$

□

Setzt man $R_{XX}(t_1, t_2) = R_1(t_1, t_2) + iR_2(t_1, t_2)$, so folgt aus a)

$$R_1(t_1, t_2) = R_1(t_2, t_1) \quad \text{und} \quad R_2(t_1, t_2) = -R_2(t_2, t_1) ,$$

d.h. der Realteil ist symmetrisch und der Imaginärteil ist *schiefsymmetrisch* (*skew symmetric*).

Beispiel 2.7.4. $X(t) = Ae^{ift}$, $A \sim N(\mu, \sigma^2)$, sei ein komplexes Signal mit zufälliger Amplitude A und fester Frequenz f . Dann ergibt sich der Erwartungswert zu

$$E(X(t)) = e^{ift} E(A) = \mu e^{ift} = \mu_X(t)$$

und die Autokorrelationsfunktion

$$R_{XX}(t_1, t_2) = E(Ae^{ift_1} \cdot Ae^{-ift_2}) = E(A^2) e^{if(t_1-t_2)}$$

$$= ((EA)^2 + \text{Var}(A)) e^{if(t_1-t_2)} = (\mu^2 + \sigma^2) e^{if(t_1-t_2)} .$$

Beispiel 2.7.5. Falls für $Z = \int_{-T}^T X(t)dt$ das Integral existiert oder in geeignetem Sinne definiert ist, handelt es sich bei Z um eine Zufallsvariable

$$Z(\omega) = \int_{-T}^T X(t, \omega)dt .$$

Es gilt dann

$$E(|Z|^2) = \int_{-T}^T \int_{-T}^T E(X(t_1)X^*(t_2)) dt_1 dt_2 = \int_{-T}^T \int_{-T}^T R_{XX}(t_1, t_2) dt_1 dt_2 .$$

Definition 2.7.6. $\{X(t)|t \in T\}$ heißt *Gaußprozess* (*Gaussian (random) process*), wenn für alle $t_1, \dots, t_n \in T$ gilt

$$(X(t_1), \dots, X(t_n))' \text{ ist } n\text{-dimensional normalverteilt .}$$

Für die Autokorrelationsfunktion eines Gaußprozesses $X(t) = U(t) + iV(t)$ mit zirkulär symmetrischen Verteilungen gilt (siehe Beispiel 2.6.6)

$$\begin{aligned} E(X(t_1)X^*(t_2)) - \mu(t_1)\mu^*(t_2) &= E[(U(t_1) + iV(t_1))(U(t_2) - iV(t_2))] - \mu(t_1)\mu^*(t_2) \\ &= \text{Cov}(U(t_1)U(t_2)) + i\text{Cov}(V(t_1)U(t_2)) - i\text{Cov}(U(t_1)V(t_2)) + \text{Cov}(V(t_1)V(t_2)) . \end{aligned}$$

Nach Definition 2.6.2 gilt

$$\text{Cov} \begin{pmatrix} U(t_1) \\ U(t_2) \\ V(t_1) \\ V(t_2) \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \text{Re } \mathbf{Q} & -\text{Im } \mathbf{Q} \\ \text{Im } \mathbf{Q} & \text{Re } \mathbf{Q} \end{pmatrix}$$

für eine komplexe, hermitesche (2×2) -Matrix $\mathbf{Q} = (\sigma_{ij} + i\tau_{ij})_{1 \leq i, j \leq 2}$.
Mit $\sigma(t_1, t_2) = \sigma_{12}$ und $\tau(t_1, t_2) = \tau_{12}$ folgt

$$R_{XX}(t_1, t_2) = \sigma(t_1, t_2) + i\tau(t_1, t_2) + \mu(t_1)\mu^*(t_2) .$$

2.7.1 Stationäre stochastische Prozesse

Definition 2.7.7. Ein *stochastischer Prozess* $\{X(t)|t \in T\}$, $T \subseteq \mathbb{R}$, $s \in T$, heißt

a) *strikt stationär* (*strict-sense stationary*), wenn

$$\forall n \in \mathbb{N}, t_1, \dots, t_n, t_1 + s, \dots, t_n + s \in T : F_{(X(t_1), \dots, X(t_n))} = F_{(X(t_1+s), \dots, X(t_n+s))} ,$$

b) *schwach stationär* (*wide-sense stationary*), wenn

$$\begin{aligned} E(X(t)) &= \mu \quad \forall t \in T , \\ E(X(t_1)X^*(t_2)) &= R_{XX}(t_1 - t_2) \quad \forall t_1, t_2 \in T . \end{aligned}$$

Strikt stationär bedeutet also, dass alle endlich dimensionalen Randverteilungen invariant gegen Zeitverschiebung sind.

Schwach stationär heißt, dass der Erwartungswert konstant ist und die Autokorrelationsfunktion nur von der Differenz $t_1 - t_2$ abhängt. Beachten Sie dabei, dass für die Autokorrelationsfunktion zwar dasselbe Symbol R_{XX} wie in Definition 2.7.2 verwendet wird, es sich jedoch in diesem Fall um eine Funktion nur eines Arguments $t = t_1 - t_2$ handelt.

Beispiel 2.7.8. Für die Beispiele wird die folgende Multiplikationsformel verwendet.

$$\cos(\alpha \pm \beta) = \cos \alpha \cos \beta \mp \sin \alpha \sin \beta \quad (2.9)$$

- a) $X(t) = a \cos(f_0 t + \phi)$ mit $\phi \sim R(0, 2\pi)$ sei ein Signal fester Frequenz und Amplitude mit zufälliger Phase ϕ . Mit der Additionsformel $\cos(\alpha \pm \beta) = \cos \alpha \cos \beta \mp \sin \alpha \sin \beta$ folgt für den Erwartungswert

$$\begin{aligned} \mu_X(t) &= E(X(t)) = a E(\cos(f_0 t + \phi)) \\ &= a E(\cos(f_0 t) \cos \phi - \sin(f_0 t) \sin \phi) \\ &= a \cos(f_0 t) E(\cos \phi) - a \sin(f_0 t) E(\sin \phi) = 0. \end{aligned}$$

Mit der Multiplikationsformel (2.9) gilt für die *Autokorrelationsfunktion*

$$\begin{aligned} R_{XX}(t_1, t_2) &= E(X(t_1)X^*(t_2)) \\ &= E(a^2 \cos(f_0 t_1 + \phi) \cos(f_0 t_2 + \phi)) \\ &= \frac{a^2}{2} E(\cos(f_0(t_1 - t_2)) + \cos(f_0(t_1 + t_2) + 2\phi)) \\ &= \frac{a^2}{2} \cos(f_0(t_1 - t_2)). \end{aligned}$$

- b) Allgemein folgt aus strikter Stationarität stets die schwache, denn

$$\begin{aligned} R_{XX}(t_1, t_2) &= E(X(t_1)X^*(t_2)) = E(X(t_1 + s)X^*(t_2 + s)) \\ &\stackrel{s=t_2-t_1}{=} E(X(t_1 - t_2)X^*(0)) = R_{XX}(t_1 - t_2, 0) \end{aligned}$$

ist eine Funktion, die nur von der Differenz $t_1 - t_2$ abhängt.

Die Umkehrung ist im Allgemeinen falsch. Sie ist jedoch richtig bei Gaußprozessen. Bei Gaußprozessen folgt aus der schwachen Stationarität bereits die strikte.

- c) (Rayleigh-Fading-Prozess)

Ein empfangenes Signal zur Zeit t (s. Kapitel 2.5) sei dargestellt durch

$$r(t) = e^{i2\pi ft} \left(\underbrace{\sum_{j=1}^n A_j \cos(c_j t + \phi_j)}_{X(t)} + i \underbrace{\sum_{j=1}^n A_j \sin(c_j t + \phi_j)}_{Y(t)} \right)$$

mit

$$\begin{aligned} c_j &= 2\pi f \frac{v}{c} \cos \vartheta_j, \\ E(X(t)) &= E(Y(t)) = 0 \quad \forall t, \\ E(X^2(t)) &= E(Y^2(t)) = \frac{\sigma^2}{2} \quad \forall t \text{ und} \\ \text{Cov}(X(t_1), Y(t_2)) &= 0 \quad \forall t_1, t_2. \end{aligned}$$

Man beachte, dass $A_1, \dots, A_n, \phi_1, \dots, \phi_n$ gemeinsam stochastisch unabhängig sind. Bisher wurde die Autokorrelationsfunktion für festes t berechnet, für variables t ergibt sich.

$$\begin{aligned}
 & \mathbb{E}(X(t)X(t + \tau)) \\
 &= \mathbb{E} \left(\sum_{j=1}^n \sum_{k=1}^n A_j A_k \cos(c_j t + \phi_j) \cos(c_k(t + \tau) + \phi_k) \right) \\
 & \stackrel{\phi_1, \dots, \phi_n \text{ s.u.}}{=} \mathbb{E} \left(\sum_{j=1}^n A_j^2 \cos(c_j t + \phi_j) \cos(c_j(t + \tau) + \phi_j) \right) \\
 & \stackrel{(2.9)}{=} \sum_{j=1}^n \left(\mathbb{E}(A_j^2) \frac{1}{2} \mathbb{E}(\cos(c_j \tau) + \cos(2c_j t + c_j \tau + 2\phi_j)) \right) \\
 &= \frac{\sigma^2}{2n} \sum_{j=1}^n \cos(c_j \tau) \\
 &= \frac{\sigma^2}{2n} \sum_{j=1}^n \cos \left(2\pi f \frac{v}{c} \tau \cos \vartheta_j \right) \\
 &= R_{XX}(\tau)
 \end{aligned}$$

$R_{XX}(\tau)$ hängt nur von der Differenz der Zeitpunkte, τ , ab. $X(t)$ ist also schwach stationär.

Seien nun die Einfallswinkel ϑ_j zufällig, d.h. sie werden beschrieben durch Zufallsvariablen $\Theta_j \sim \mathbb{R}(0, 2\pi)$, welche von $\{\phi_k\}$ und $\{A_l\}$ stochastisch unabhängig sind. Dann ergibt sich

$$\begin{aligned}
 R_{XX}(\tau) &= \frac{\sigma^2}{2} \frac{1}{2\pi} \int_0^{2\pi} \cos \left(2\pi f \frac{v}{c} \tau \cos \vartheta \right) d\vartheta \\
 &= \frac{\sigma^2}{2} \frac{1}{\pi} \int_0^{\pi} \cos \left(2\pi f \frac{v}{c} \tau \cos \vartheta \right) d\vartheta \\
 &= \frac{\sigma^2}{2} \operatorname{Re} \left(J_0 \left(2\pi f \frac{v}{c} \tau \right) \right) \\
 &= \frac{\sigma^2}{2} \operatorname{Re} \left(J_0 \left(2\pi \frac{v}{\lambda} \tau \right) \right) \\
 &= \frac{\sigma^2}{2} \operatorname{Re} (J_0(2\pi f_D \tau)), \quad \tau \geq 0,
 \end{aligned}$$

wobei $f_D = \frac{v}{\lambda}$ der maximale Doppler-Shift und

$$J_0(x) = \frac{1}{\pi} \int_0^{\pi} e^{-ix \cos \vartheta} d\vartheta$$

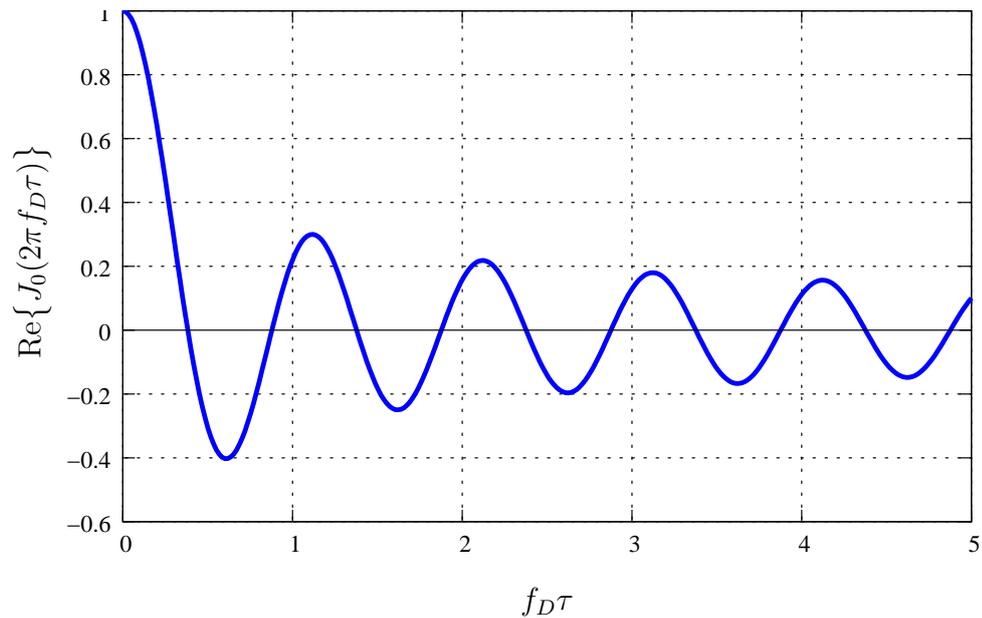


Abbildung 2.15: Realteil der Bessel-Funktion 1. Art 0'ter Ordnung

die Bessel-Funktion 1. Art 0'ter Ordnung ist, siehe Abb. 2.15 für deren Realteil.

Genauso wird

$$R_{YY}(\tau) = \frac{\sigma^2}{2} \operatorname{Re}(J_0(2\pi f_D \tau)), \quad \tau \geq 0$$

hergeleitet.

2.7.2 Lineare Systeme mit stochastischer Eingabe

$x(t)$ bezeichne die Eingabe eines analogen Kanals als eine Funktion der Zeit t , z.B. eine Spannungs- oder Stromkurve. Die Kurve $y(t)$ am Ausgang des Kanals ist in der Regel eine verzerrte, verzögerte und abgeschwächte Version von $x(t)$. Von großer Bedeutung sind hierbei die sogenannten linearen zeitinvarianten Filter. Sie sind nicht nur ein inhärenter, wie oben beschriebener Effekt des Kanals, sondern werden bewusst eingesetzt, um eingegebene Signale zu glätten und bestimmte Frequenzen herauszuschneiden. In diesem Abschnitt wird zunächst ein deterministisches Modell für die lineare zeitinvariante Filterung angegeben, anschließend werden die stochastischen Eigenschaften der Ausgabe eines Filters untersucht, wenn die Eingabe ein stochastischer Prozess ist.

In einem System wird die Eingabe $x(t)$ durch eine Transformation T verändert, die Ausgabe ist $y(t) = T(x(t))$. Der *Tiefpassfilter* aus Abbildung 2.16 ist ein Beispiel eines solchen Systems.

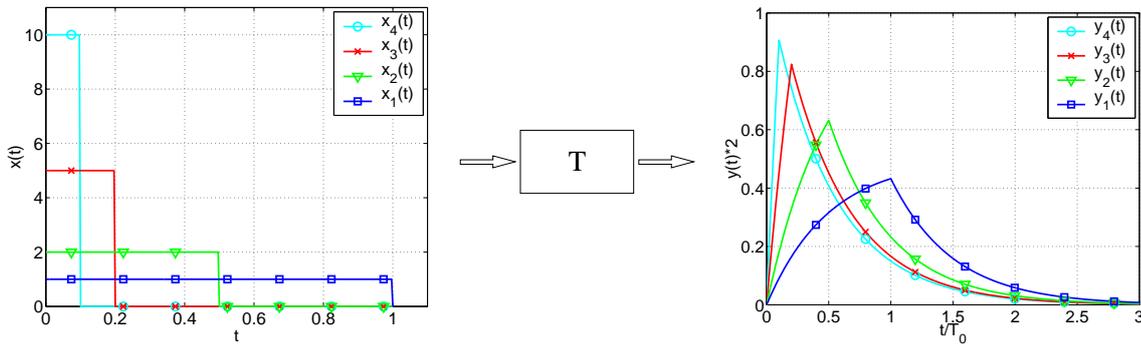


Abbildung 2.16: Reaktionsverläufe eines Tiefpassfilters für *Rechteckimpulse* unterschiedlicher Dauer T_0 mit der Impulsantwort $h(u) = \frac{1}{2}e^{-\frac{u}{2}}$.

Unter der Annahme, dass T linear ist, kann die Ausgabe mit Hilfe einer Funktion h durch die *Faltung* von h und x wie folgt beschrieben werden:

$$y(t) = \int_{-\infty}^{\infty} h(u)x(t-u)du = \int_{-\infty}^{\infty} h(t-u)x(u)du = (h * x)(t) .$$

Systeme dieses Typs heißen *linear zeitinvariant (LTI, linear time-invariant systems* oder *LTI-Filter*).

h heißt *Impulsantwort (impulse response)*. $h(t)$ ergibt sich im Grenzwert als Ausgabe von schmaler werdenden Rechteckimpulsen der Fläche 1.

Oft wird angenommen, dass $h(t) = 0$, falls $t \leq 0$, d.h zur Zeit t erfolgt kein Vorgriff auf zukünftige Werte des Signals. $h(t)$ heißt dann *physikalisch realisierbar*.

Bei der Beschreibung des LTI-Systems ist das Eingangesignal $x(t)$ eine deterministische Funktion der Zeit. Ist die Eingabe $\{X(t)\}$ ein stochastischer Prozess, so ist die Ausgabe $\{Y(t)\}$ eines LTI-Systems ebenfalls ein stochastischer Prozess. Die Frage ist nun, welche Aussagen sich über $\{Y(t)\}$ treffen lassen.

Theorem 2.7.9. Ein schwach stationärer stochastischer Prozess $\{X(t)\}$ sei die Eingabe eines LTI-Systems mit Impulsantwort h , also

$$Y(t) = \int_{-\infty}^{\infty} h(u)X(t-u)du .$$

Dann ist $\{Y(t)\}$ schwach stationär mit

$$\begin{aligned} \mu_Y(t) &= E(Y(t)) = \mu_X \int_{-\infty}^{\infty} h(u)du \quad \text{und} \\ R_{YY}(t) &= \int_{-\infty}^{\infty} h(u) \int_{-\infty}^{\infty} h^*(v)R_{XX}(t-u+v)dvdu . \end{aligned}$$

Beweis. Der Beweis wird nur in seinen wesentlichen Ideen geführt.

$$\begin{aligned}
 \text{(i)} \quad \mu_Y(t) &= \mathbb{E}(Y(t)) = \int_{-\infty}^{\infty} \mathbb{E}(h(u)X(t-u)) du = \int_{-\infty}^{\infty} h(u)\mu_X du . \\
 \text{(ii)} \quad R_{YY}(t_1, t_2) &= \mathbb{E}(Y(t_1)Y^*(t_2)) \\
 &= \mathbb{E} \left[\int_{-\infty}^{\infty} h(u)X(t_1-u) du \int_{-\infty}^{\infty} h^*(v)X^*(t_2-v) dv \right] \\
 &= \int_{-\infty}^{\infty} h(u) \int_{-\infty}^{\infty} h^*(v) \underbrace{\mathbb{E}[X(t_1-u)X^*(t_2-v)]}_{R_{XX}(t_1-t_2-u+v)} dv du .
 \end{aligned}$$

Also ist $R_{YY}(t_1, t_2)$ eine Funktion, die nur von der Differenz $t_1 - t_2 = t$ abhängt. Daraus folgt Stationarität und die obige Darstellung. \square

Beachten Sie, dass im Beweis die folgenden Punkte mathematisch nicht vollständig ausgeführt sind. Einerseits muss der Begriff des Integrals über einen stochastischen Prozess als Zufallsvariable weiter präzisiert werden. Andererseits fehlt noch die Argumentation für die Vertauschung von Erwartungswert und Integration. Beides ist jedoch gesichert, wenn

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(s)h^*(t)R_{XX}(s, t) ds dt$$

existiert.

Die *Kreuzkorrelationsfunktion* (*cross-correlation function*) zwischen zwei stochastischen Prozessen $\{X(t)\}$ und $\{Y(t)\}$ ist definiert als

$$R_{XY}(t_1, t_2) = \mathbb{E}(X(t_1)Y^*(t_2)) .$$

Für die Kreuzkorrelationsfunktion des Eingabe- und Ausgabeprozesses eines LTI-Systems gilt insbesondere

$$\begin{aligned}
 R_{XY}(t_1, t_2) &= \mathbb{E}(X(t_1)Y^*(t_2)) \\
 &= \mathbb{E} \left[X(t_1) \int_{-\infty}^{\infty} h^*(v)X^*(t_2-v) dv \right] \\
 &= \int_{-\infty}^{\infty} h^*(v) \mathbb{E}[X(t_1)X^*(t_2-v)] dv \\
 &= \int_{-\infty}^{\infty} h^*(v)R_{XX}(t_1, t_2-v) dv .
 \end{aligned}$$

Wenn $\{X(t)\}$ schwach stationär ist, hängt $R_{XX}(t_1, t_2)$ nur von der Differenz $t = t_1 - t_2$ ab. Also gilt

$$R_{XY}(t) = \int_{-\infty}^{\infty} h^*(v)R_{XX}(t+v)dv = h^*(-t) * R_{XX}(t) .$$

In Theorem 2.7.9 folgt damit die Darstellung

$$\begin{aligned} R_{YY}(t) &= \int_{-\infty}^{\infty} h(u)R_{XY}(t-u)du = h(t) * R_{XY}(t) \\ &= h(t) * h^*(-t) * R_{XX}(t) , \end{aligned}$$

$R_{YY}(t)$ ergibt sich also als Faltung von $h(t)$, $h^*(-t)$ und $R_{XX}(t)$.

2.7.3 Leistungsdichtespektrum

Es wird zunächst an den Begriff der Fouriertransformierten einer Funktion g erinnert.

Definition 2.7.10. Zwei Funktionen $g(t)$ und $G(f)$ heißen *Paar von Fouriertransformierten (FT-Paar)*, wenn gilt

$$G(f) = \int_{-\infty}^{\infty} g(t)e^{-i2\pi ft} dt \quad \text{und} \quad g(t) = \int_{-\infty}^{\infty} G(f)e^{i2\pi ft} df .$$

Interpretiert man $g(t)$ als deterministisches Signal, so beträgt seine Energie

$$E = \int_{-\infty}^{\infty} |g(t)|^2 dt .$$

Nach dem *Satz von Parseval* gilt die Beziehung

$$\int_{-\infty}^{\infty} |g(t)|^2 dt = \int_{-\infty}^{\infty} |G(f)|^2 df .$$

Betrachtet man die physikalische Interpretation der Ausdrücke, so entspricht $|G(f)|^2$ dem *Energiespektrum* und $|G(f)|^2 \Delta f$ näherungsweise der Signalenergie im Band $[f, f + \Delta f]$. Vergleichen Sie dazu auch Abbildung 2.17.

Im Weiteren wird ein analoges Konzept für schwach stationäre stochastische Prozesse $\{X(t)\}$ entwickelt. $X(t)$ ist für festes t eine Zufallsvariable und gleiches gilt auch für $|X(t)|^2$. $E(|X(t)|^2)$ ist dann die zu erwartenden Momentanleistung zum Zeitpunkt t und es gilt unabhängig von t

$$E(|X(t)|^2) = E(X(t)X^*(t)) = R_{XX}(0) .$$

Sei $R_{XX}(t)$, $S_{XX}(f)$ ein FT-Paar, also $S_{XX}(f)$ die Fouriertransformierte von $R_{XX}(t)$ mit

$$R_{XX}(t) = \int_{-\infty}^{\infty} S_{XX}(f)e^{i2\pi ft} df . \quad (*)$$

Dann folgt daraus

$$E(|X(t)|^2) = R_{XX}(0) = \int_{-\infty}^{\infty} S_{XX}(f)df .$$

Ist $R_{XX}(t)$ die Autokorrelationsfunktion eines schwach stationären stochastischen Prozesses, also nicht-negativ definit, so existiert bei geeignetem Integralbegriff immer eine Funktion $S_{XX}(f)$, für die (*) gilt. Dies besagt der *Satz von Bochner-Khintchin*.

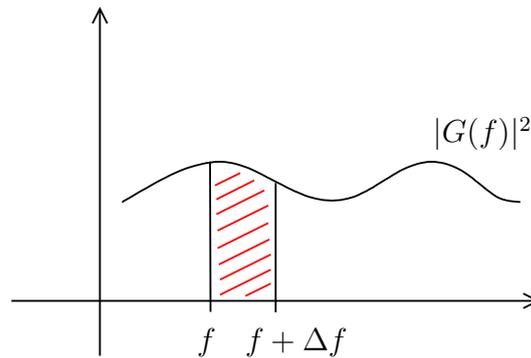


Abbildung 2.17: Energiespektrum und Signalenergie im Band $[f, f + \Delta f]$ des deterministischen Signals $g(t)$.

Definition 2.7.11. Für einen schwach stationären stochastischen Prozess $\{X(t)\}$ mit der Autokorrelationsfunktion $R_{XX}(t)$ heißt die obige Funktion $S_{XX}(f)$ *Leistungsdichtespektrum (power spectral density)*.

Proposition 2.7.12. Für das Leistungsdichtespektrum gelten die folgenden Eigenschaften:

- a) $E(|X(t)|^2) = R_{XX}(0) = \int_{-\infty}^{\infty} S_{XX}(f)df$,
- b) $S_{XX}(f) = S_{XX}(-f)$, falls $R_{XX}(t) \in \mathbb{R}$,
- c) $S_{XX}(f) \geq 0$, insbesondere gilt $S_{XX}(f) \in \mathbb{R}$ für alle $f \in \mathbb{R}$.

Die Frage ist nun, wie sich das Leistungsdichtespektrum bei der LTI-Filterung verhält. Betrachtet wird dazu ein LTI-Filter mit Impulsantwort h , also

$$Y(t) = \int_{-\infty}^{\infty} h(u)X(t - u)du .$$

Das folgende Theorem beschreibt das Leistungsdichtespektrum von $\{Y(t)\}$. $H(f)$ bezeichne die Fouriertransformierte von $h(t)$.

Theorem 2.7.13. $\{X(t)\}$ sei ein schwach stationärer stochastischer Prozess mit Leistungsdichtespektrum $S_{XX}(f)$. Für das Leistungsdichtespektrum des LTI-gefilterten Prozesses $Y(t) = \int_{-\infty}^{\infty} h(u)X(t - u)du$ gilt dann

$$S_{YY}(f) = |H(f)|^2 S_{XX}(f) .$$

Beweis. Der Beweis wird nur in seinen wesentlichen Ideen geführt. Mit Theorem 2.7.9 gilt

$$S_{YY}(f) = \int_{-\infty}^{\infty} \left[\underbrace{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(u)h^*(v)R_{XX}(t-u+v)dudv}_{R_{YY}(t)} \right] e^{-i2\pi ft} dt .$$

Mit der Substitution $s = t - u + v$ ergibt sich

$$\begin{aligned} S_{YY}(f) &= \underbrace{\int_{-\infty}^{\infty} h(u)e^{-i2\pi fu} du}_{H(f)} \underbrace{\int_{-\infty}^{\infty} h^*(v)e^{i2\pi fv} dv}_{H^*(f)} \underbrace{\int_{-\infty}^{\infty} R_{XX}(s)e^{-i2\pi fs} ds}_{S_{XX}(f)} \\ &= |H(f)|^2 S_{XX}(f) . \end{aligned}$$

□

Beispiel 2.7.14. Ein *Bandpassfilter* $H(f)$ der Breite B , zentriert an der Frequenz f_0 (vergleichen Sie Abbildung 2.18), sei gegeben mit

$$H(f) = \begin{cases} 1 & , \quad |f \pm f_0| \leq \frac{B}{2} \\ 0 & , \quad \text{sonst} \end{cases} .$$

Es gilt

$$h(t) = B \operatorname{si}(\pi Bt) \cdot 2 \cos(2\pi f_0 t) , \quad \text{mit } \operatorname{si}(t) = \frac{\sin t}{t} .$$

Sei nun $\{X(t)\}$ ein schwach stationärer stochastischer Prozess und

$$Y(t) = \int_{-\infty}^{\infty} h(u)X(t-u)du ,$$

dann gilt

$$S_{YY}(f) = |H(f)|^2 S_{XX}(f) ,$$

und die erwartete Momentanleistung beträgt

$$\begin{aligned} E(|Y(t)|^2) &= \int_{-\infty}^{\infty} S_{YY}(f)df = \int_{-f_0-\frac{B}{2}}^{-f_0+\frac{B}{2}} S_{XX}(f)df + \int_{f_0-\frac{B}{2}}^{f_0+\frac{B}{2}} S_{XX}(f)df \\ &= 2 \int_{f_0-\frac{B}{2}}^{f_0+\frac{B}{2}} S_{XX}(f)df , \end{aligned}$$

da $S_{XX}(f) = S_{XX}(-f)$. Für kleines B gilt approximativ

$$E(|Y(t)|^2) \approx 2BS_{XX}(f_0) .$$

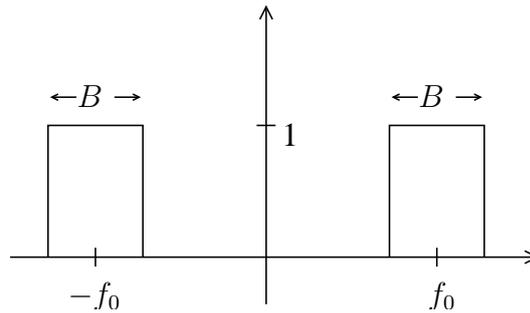


Abbildung 2.18: Bandpassfilter der Bandbreite B , zentriert an der Frequenz f_0 .

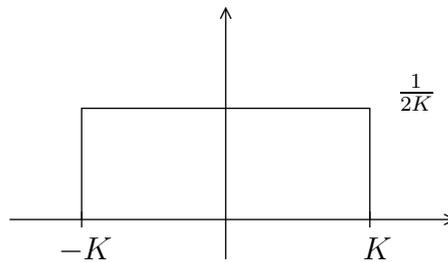


Abbildung 2.19: Impulsantwort $h(t) = \frac{1}{2K}\mathbb{I}_{[-K,K]}(t)$.

Beispiel 2.7.15. a) $\{X(t)\}$ sei ein schwach stationärer stochastischer Prozess mit Autokorrelationsfunktion $R_{XX}(t)$. Die Impulsantwort eines LTI-Systems sei gegeben durch

$$h(t) = \frac{1}{2K}\mathbb{I}_{[-K,K]}(t) ,$$

dann gilt für den LTI-gefilterten Prozess

$$Y(t) = \int_{-\infty}^{\infty} h(u)X(t-u)du = \frac{1}{2K} \int_{t-K}^{t+K} X(u)du .$$

Der auf diese Weise erhaltene Ausgabeprozess wird als *gleitendes Mittel* (*moving average*) bezeichnet. Vergleichen Sie dazu auch die Abbildungen 2.19 und 2.21.

Es gilt nun

$$\begin{aligned} H(f) &= \int_{-\infty}^{\infty} h(t)e^{-i2\pi ft}dt = \frac{1}{2K} \int_{-K}^K e^{-i2\pi ft} dt \\ &= \frac{\sin(2\pi fK)}{2\pi fK} = \text{si}(2\pi fK) , \end{aligned}$$

und mit Theorem 2.7.13 ergibt sich

$$S_{YY}(f) = S_{XX}(f) \frac{\sin^2(2\pi fK)}{(2\pi fK)^2}.$$

Für die Autokorrelationsfunktion des LTI-gefilterten Prozesses gilt mit Theorem 2.7.9

$$R_{YY}(t) = h(t) * h(-t) * R_{XX}(t).$$

Wegen

$$h(t) * h(-t) = \begin{cases} \frac{1}{2K} \left(1 - \frac{|t|}{2K}\right) & , \quad |t| \leq 2K \\ 0 & , \quad \text{sonst} \end{cases}$$

gilt

$$R_{YY}(t) = \int_{-2K}^{2K} \frac{1}{2K} \left(1 - \frac{|u|}{2K}\right) R_{XX}(t-u) du.$$

b) (Rayleigh-Fading Prozess)

Es gilt nach Beispiel 2.7.8

$$R_{XX}(\tau) = E(X(t)X(t+\tau)) = \frac{\sigma^2}{2} \operatorname{Re}(J_0(2\pi f_D \tau)), \quad \tau \geq 0,$$

mit der Bessel-Funktion erster Art 0'ter Ordnung

$$J_0(x) = \frac{1}{\pi} \int_0^\pi e^{-ix \cos \theta} d\theta$$

und dem maximalen Doppler-Shift $f_D = v/\lambda$. Dann gilt für das Leistungsdichtespektrum von $X(t)$

$$S_{XX}(f) = \begin{cases} \frac{\sigma^2}{\pi f_D} \frac{1}{\sqrt{1-(f/f_D)^2}}, & \text{falls } |f| \leq f_D \\ 0, & \text{sonst} \end{cases}.$$

Also gilt $E(X^2(t)) = \int_{-f_D}^{f_D} S_{XX}(f) df$. Das Leistungsdichtespektrum ist für $f_D = 1$ und $\sigma^2 = 1$ in Abbildung 2.20 dargestellt.

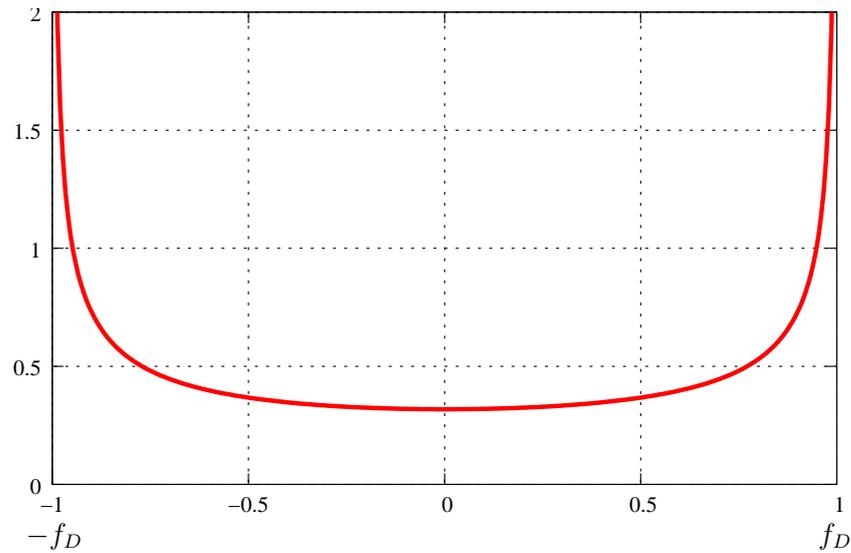


Abbildung 2.20: Beispiel zum Leistungsdichtespektrum für $f_D = 1$ und $\sigma^2 = 1$.

2.7.4 Weißes Rauschen

Zunächst wird folgender Zusammenhang nochmals aufgegriffen. Gegeben sei ein schwach stationärer stochastischer Prozess $\{X(t)\}$ mit Autokorrelationsfunktion $R_{XX}(t)$. Dann ist das Leistungsdichtespektrum (power spectral density) $S_{XX}(f)$ definiert durch die Gleichung

$$R_{XX}(t) = \int_{-\infty}^{\infty} S_{XX}(f) e^{i2\pi ft} df ,$$

und es gilt

$$E(|X(t)|^2) = R_{XX}(0) = \int_{-\infty}^{\infty} S_{XX}(f) df .$$

Der Ausdruck

$$S_{XX}(f) \cdot \Delta f$$

kann näherungsweise als die erwartete Leistung im Band $[f, f + \Delta f]$ interpretiert werden.

Ziel ist es nun, stochastische Prozesse zu beschreiben, deren Leistungsdichtespektrum konstant ist, d.h. alle Frequenzen treten mit der gleichen, positiven Energie auf. Im Zeitbereich entspricht dies der Forderung, dass die Autokorrelationsfunktion für alle $t \neq 0$ den Wert null annehmen soll.

Als Vorüberlegung hierzu dient folgendes Beispiel.

Beispiel 2.7.16. Im Folgenden bezeichnet $\mathfrak{F}(h)$ die Fouriertransformierte der Funktion h . Für

$$h_n(t) = n \cdot \mathbb{I}_{[-\frac{1}{2n}, \frac{1}{2n}]}(t)$$

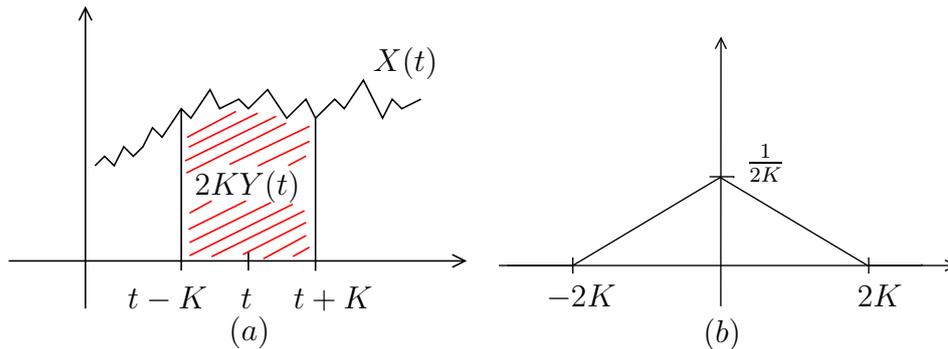


Abbildung 2.21: (a) Gleitendes Mittel, Ausgabeprozess als Fläche unter der Kurve des Eingabeprozesses. (b) $h(t) * h(-t)$, Faltung der Impulsantwort mit sich selbst.

gilt

$$\mathfrak{F}(h_n)(f) = \frac{\sin\left(\frac{\pi f}{n}\right)}{\frac{\pi f}{n}} = \text{si}\left(\frac{\pi f}{n}\right).$$

Dies sieht man wie folgt ein.

$$\begin{aligned} \int_{-\infty}^{\infty} h_n(t) e^{-i2\pi f t} dt &= n \int_{-\frac{1}{2n}}^{\frac{1}{2n}} e^{-i2\pi f t} dt = n \frac{1}{-i2\pi f} e^{-i2\pi f t} \Big|_{-\frac{1}{2n}}^{\frac{1}{2n}} \\ &= n \frac{1}{-i2\pi f} \left(\cos \frac{2\pi f}{2n} - i \sin \frac{2\pi f}{2n} - \cos \frac{2\pi f}{2n} - i \sin \frac{2\pi f}{2n} \right) \\ &= \frac{n}{\pi f} \sin \frac{\pi f}{n} = \text{si}\left(\frac{\pi f}{n}\right). \end{aligned}$$

In Abbildung 2.22 sind h_n und $\mathfrak{F}(h_n)$ grafisch dargestellt.

Für die Faltung der Impulsantwort mit sich selbst gilt

$$g_n(t) = (h_n * h_n)(t) = \begin{cases} n^2 \left(t + \frac{1}{n}\right) & , -\frac{1}{n} \leq t \leq 0 \\ n^2 \left(\frac{1}{n} - t\right) & , 0 \leq t \leq \frac{1}{n} \\ 0 & , \text{sonst} \end{cases}$$

und die entsprechende Fouriertransformierte ergibt sich zu

$$\mathfrak{F}(g_n)(f) = (\mathfrak{F}(h_n)(f))^2 = \frac{\sin^2\left(\frac{\pi f}{n}\right)}{\left(\frac{\pi f}{n}\right)^2} = \text{si}^2\left(\frac{\pi f}{n}\right).$$

Eine grafische Veranschaulichung hierzu liefert Abbildung 2.23.

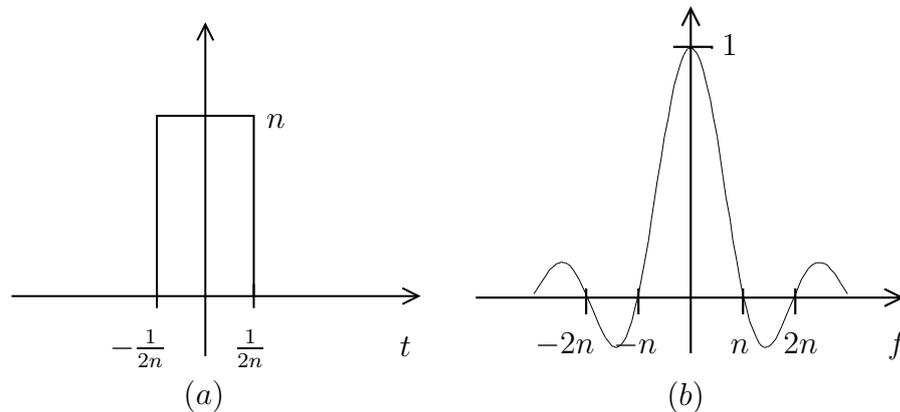


Abbildung 2.22: (a) Impulsantwort $h_n(t) = n \cdot \mathbb{I}_{[-\frac{1}{2n}, \frac{1}{2n}]}(t)$, (b) Fouriertransformierte $\mathfrak{F}(h_n)(f) = \text{si}(\frac{\pi f}{n})$.

Die Grenzwertbetrachtung der mit sich selbst gefalteten Impulsantwort im Zeitbereich liefert

$$g_n(t) \xrightarrow{(n \rightarrow \infty)} \begin{cases} \infty & , t = 0 \\ 0 & , t \neq 0 \end{cases} \quad , \text{ wobei } \int_{-\infty}^{\infty} g_n(t) dt = 1 \quad \text{für alle } n.$$

Der Grenzwert ist keine reellwertige Funktion und wird mit $\delta(t)$ bezeichnet. $\delta(t)$ ist eine verallgemeinerte Funktion bzw. eine Distribution und heißt *Dirac Delta-Funktion* oder *Dirac-Impuls*. Als definierende Eigenschaft dient

$$\int_{-\infty}^{\infty} f(x) \delta(x) dx = f(0) \quad \text{für alle in } x = 0 \text{ stetigen Funktionen } f.$$

Für die Fouriertransformierte $\mathfrak{F}(g_n)$ gilt im Grenzwert

$$\mathfrak{F}(g_n)(f) \xrightarrow{(n \rightarrow \infty)} 1 \quad ,$$

d.h. der Dirac-Impuls besitzt ein konstantes Spektrum mit dem Wert 1.

Im Folgenden werden zu dem vorangegangenen Beispiel analoge Eigenschaften bei stochastischen Prozessen betrachtet. Dabei wird die Autokorrelationsfunktion $R_{WW}(t)$ eines stochastischen Prozesses gesucht, der ein konstantes Spektrum $S_{WW}(f) = \frac{N_0}{2}$ hat.

Definition 2.7.17. Ein schwach stationärer (verallgemeinerter) stochastischer Prozess $\{W(t)\}$ heißt *weißes Rauschen* (*white noise*), wenn

$$S_{WW}(f) = \frac{N_0}{2} \quad \text{für alle } f \in \mathbb{R} .$$

$\{W(t)\}$ heißt *Gaußsches weißes Rauschen* (*white Gaussian noise*), wenn für alle t_1, \dots, t_n der Zufallsvektor $(W(t_1), \dots, W(t_n))'$ gemeinsam (n-dimensional) normalverteilt ist.

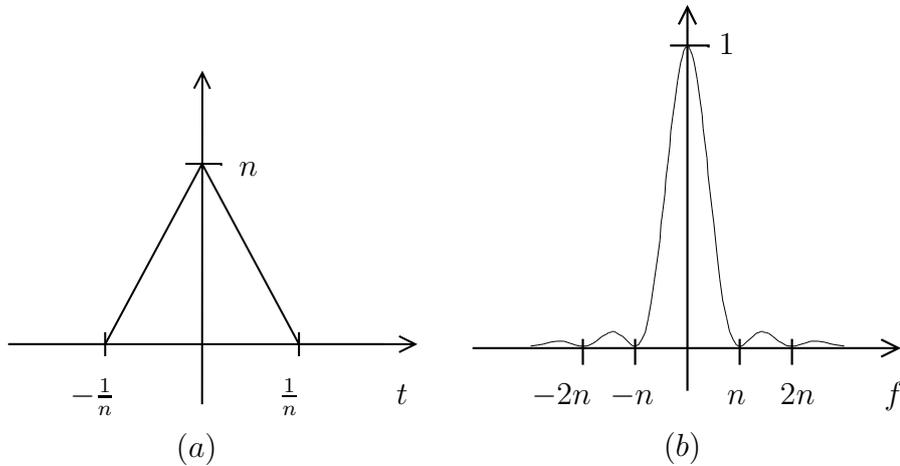


Abbildung 2.23: (a) Faltung der Impulsantwort mit sich selbst $g_n(t) = (h_n(t) * h_n(t))(t)$,
 (b) Fouriertransformierte $\mathfrak{F}(g_n)(f) = \text{si}^2\left(\frac{\pi f}{n}\right)$.

Die Bezeichnung *Weißes Rauschen* wird in Analogie zu dem Begriff des weißen Lichts verwendet, bei dem alle Farben bzw. Lichtfrequenzen vertreten sind. Für weißes Rauschen gilt also

$$R_{WW}(t) = \frac{N_0}{2} \delta(t) \quad \text{und} \quad S_{WW}(f) = \frac{N_0}{2} .$$

Dieser Zusammenhang ist in [Abbildung 2.24](#) grafisch dargestellt.

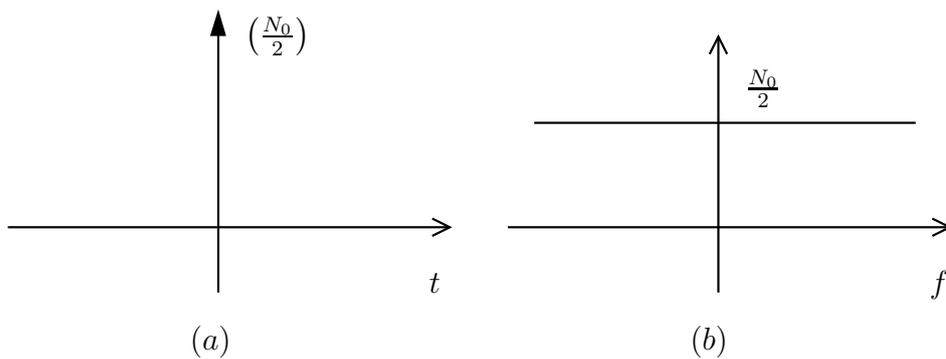


Abbildung 2.24: (a) Autokorrelationsfunktion $R_{WW}(t) = \frac{N_0}{2} \delta(t)$ und (b) Leistungsdichtespektrum $S_{WW}(f) = \frac{N_0}{2}$ von weißem Rauschen.

Weißes Rauschen ist ein universell anwendbares Modell. Es ist jedoch physikalisch

nicht realisierbar, da wegen

$$\int_{-\infty}^{\infty} S_{WW}(f)df = \infty$$

eine unendliche Leistung benötigt würde. Deswegen wird im Folgenden das Rauschen in einem beschränkten Band betrachtet.

2.7.5 Filterung von weißem Rauschen

Gegeben sei ein LTI-System mit Impulsantwort h . Es wird weißes Rauschen $\{W(t)\}$ gefiltert, d.h. die Systemausgabe ist gegeben durch

$$N(t) = \int_{-\infty}^{\infty} h(u)W(t-u)du = (h * W)(t) .$$

Der Ausgabeprozess $\{N(t)\}$ heißt *farbiges Rauschen* (*colored noise*). Mit Theorem 2.7.13 gilt für das Leistungsdichtespektrum

$$S_{NN}(f) = |H(f)|^2 S_{WW}(f) = \frac{N_0}{2} |H(f)|^2 .$$

Beispiel 2.7.18. Die Fouriertransformierte der Impulsantwort des Filters sei

$$H(f) = \mathbb{I}_{[-B,B]}(f) ,$$

wobei B die absolute Bandbreite bezeichnet. Dann gilt

$$S_{NN}(f) = \frac{N_0}{2} \mathbb{I}_{[-B,B]}(f) .$$

Der zugehörige Prozess $\{N(t)\}$ heißt bandbegrenzt weißes Rauschen. Für die Autokorrelationsfunktion gilt

$$\begin{aligned} R_{NN}(t) &= \int_{-B}^B \frac{N_0}{2} e^{i2\pi ft} df = \frac{N_0}{2} \frac{1}{i2\pi t} e^{i2\pi ft} \Big|_{-B}^B \\ &= \frac{N_0}{2} \frac{1}{i2\pi t} (e^{i2\pi Bt} - e^{-i2\pi Bt}) = \frac{N_0}{2} 2B \frac{\sin(2\pi Bt)}{2\pi Bt} \\ &= N_0 B \operatorname{si}(2\pi Bt) . \end{aligned}$$

Beispiel 2.7.19 (*Matched Filter*). Ein deterministisches Signal $s(t)$ wird übertragen und durch additives Rauschen verfälscht. Empfangen wird

$$X(t) = s(t) + W(t) .$$

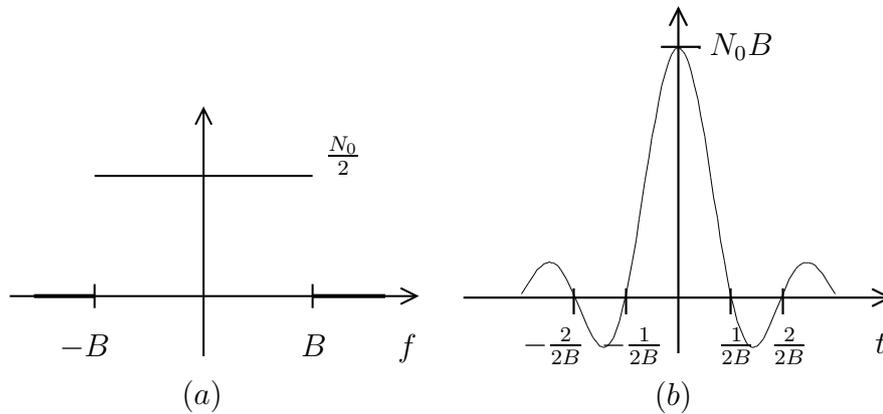


Abbildung 2.25: (a) Leistungsdichtespektrum $S_{NN}(f)$ und (b) Autokorrelationsfunktion $R_{NN}(t)$ von farbigem Rauschen im beschränkten Band $[-B, B]$.

$\{X(t)\}$ wird gefiltert von einem LTI-System mit Impulsantwort $h(t)$. Die Ausgabe des Systems ist dann gegeben durch

$$Y(t) = (h * X)(t) = (h * s)(t) + (h * W)(t) = y_s(t) + N(t) .$$

Das Ziel ist es nun, einen optimalen Filter h zu bestimmen, wobei als Kriterium der *Signal-zu-Rausch Abstand* (*Signal-to-Noise ratio*, SNR) zur Zeit t_0 dient. Dieser wird beschrieben durch das Verhältnis von Signalleistung zur Zeit t_0 zu erwarteter Rauschleistung. Die erwartete Rauschleistung ist dabei aufgrund der Stationarität unabhängig vom Zeitpunkt t . Bezeichnen $S(f)$ und $H(f)$ die Fouriertransformierten von $s(t)$ und $h(t)$, so gilt

$$\text{SNR}_0 = \frac{|y_s(t_0)|^2}{\mathbb{E}(|N(t)|^2)} = \frac{\left| \int_{-\infty}^{\infty} S(f)H(f)e^{i2\pi ft_0} df \right|^2}{\int_{-\infty}^{\infty} \underbrace{|H(f)|^2 S_{WW}(f)}_{=S_{NN}(f)} df} .$$

Unter der Annahme, dass es sich bei $\{W(t)\}$ um weißes Rauschen handelt, gilt für das entsprechende Leistungsdichtespektrum

$$S_{WW}(f) = \frac{N_0}{2}$$

und damit

$$\begin{aligned} \text{SNR}_0 &= \frac{\left| \int_{-\infty}^{\infty} S(f)H(f)e^{i2\pi ft_0} df \right|^2}{\frac{N_0}{2} \int_{-\infty}^{\infty} |H(f)|^2 df} \\ &\stackrel{\text{(Cauchy-Schwarz)}}{\leq} \frac{2}{N_0} \frac{1}{\int_{-\infty}^{\infty} |H(f)|^2 df} \int_{-\infty}^{\infty} |S(f)|^2 \underbrace{|e^{i2\pi ft_0}|^2}_{=1} df \cdot \int_{-\infty}^{\infty} |H(f)|^2 df \\ &= \frac{2}{N_0} \int_{-\infty}^{\infty} |S(f)|^2 df \stackrel{\text{(Parseval)}}{=} \frac{2}{N_0} \int_{-\infty}^{\infty} |s(t)|^2 dt . \end{aligned}$$

Bei der verwendeten *Cauchy-Schwarz Ungleichung* gilt Gleichheit, falls

$$H(f) = S^*(f)e^{-i2\pi ft_0} \quad \text{bzw.} \quad h(t) = s(t_0 - t) \quad (*)$$

Durch die Eigenschaft (*) wird der optimale Empfänger im Sinne von maximalem Signal-zu-Rausch Abstand zur Zeit t_0 (max SNR_0) realisiert, er heißt *matched filter*.

3 Elemente der Informationstheorie

Informationstheorie behandelt nicht den Inhalt oder die Bedeutung von Informationen, sondern Systeme, Modelle und Methoden zur Übermittlung von Informationen bzw. Daten. Das Standardmodell der Informationsübertragung ist in Abbildung 3.1 dargestellt.

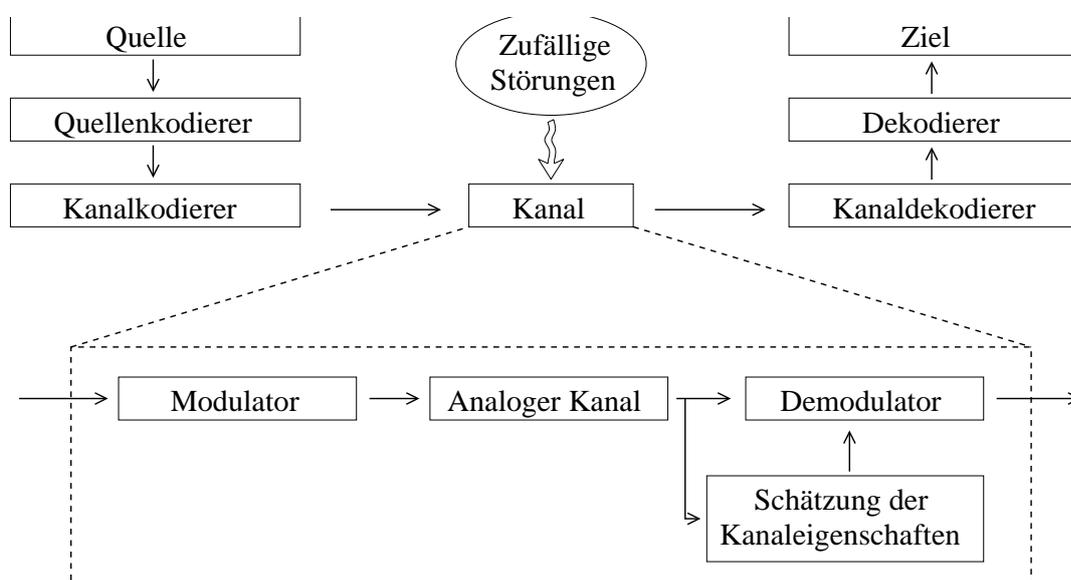


Abbildung 3.1: Standardmodell der Informationsübertragung.

Grundlegende Hilfsmittel stammen aus der stochastischen Modellierung. Begründer der Informationstheorie ist *Claude Elwood Shannon* mit seiner Arbeit *A Mathematical Theory of Communication*, *Bell System Technical Journal*, vol. 27, 1948, pp. 379-423, 623-656.

3.1 Grundlagen

Betrachtet werden zwei Zufallsexperimente mit jeweils vier Ausgängen und den Wahrscheinlichkeiten $(0.9, 0.04, 0.04, 0.02)$ bzw. $(0.25, 0.25, 0.25, 0.25)$. Gesucht ist eine Maßzahl für die Unbestimmtheit des Ausgangs vor der Ausführung oder äquivalent hierzu eine Maßzahl für die Information bzw. den Informationsgewinn nach der Ausführung des Experimentes. Die formale Beschreibung dieser intuitiven Idee lautet wie folgt.

Sei X eine endlich diskrete Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ mit Träger $\mathcal{X} = \{x_1, \dots, x_m\}$ und der Verteilung

$$P(X = x_i) = p_i, \quad i = 1, \dots, m, \quad \text{wobei} \quad p_i \geq 0, \quad \sum_{i=1}^m p_i = 1.$$

Der *Informationsgehalt* des Ereignisses $\{X = x_i\}$ hänge nur von der Wahrscheinlichkeit p_i ab. Der Informationsgehalt wird durch eine Funktion

$$h: [0, 1] \rightarrow \mathbb{R}$$

gemessen, von der man folgende Eigenschaften fordert.

- (i) h ist stetig auf $[0, 1]$,
- (ii) $h(p \cdot q) = h(p) + h(q) \quad \forall p, q \in [0, 1]$,
(Additivität bei stochastischer Unabhängigkeit)
- (iii) $\exists c > 1$ mit $h\left(\frac{1}{c}\right) = 1$.
(Normierung)

Die Eigenschaft (ii) lässt sich wie folgt interpretieren. Sind E_1 und E_2 stochastisch unabhängige Ereignisse mit $P(E_1) = p$ und $P(E_2) = q$, so entspricht der Informationsgehalt von $E_1 \cap E_2$ der Summe der einzelnen Informationsgehalte. Aus den Eigenschaften (i)-(iii) folgt notwendig, dass

$$h(p) = -\log_c p, \quad p \in [0, 1].$$

Definition 3.1.1. Sei X eine endlich diskrete Zufallsvariable und $c > 1$ eine Konstante. Dann heißt

$$I_X(x_j) = -\log_c P(X = x_j) = -\log_c p_j$$

Informationsgehalt des Ereignisses $\{X = x_j\}$ und

$$H(X) = -\sum_{j=1}^m P(X = x_j) \log_c P(X = x_j) = -\sum_{j=1}^m p_j \log_c p_j$$

Entropie von X .

Um nicht-definierte Fälle abzufangen, wird hierbei als Konvention vereinbart, dass $\log 0 = -\infty$ und $0 \cdot \log 0 = 0$. Die Basis des Logarithmus c wird im Folgenden weggelassen, sie gilt als von vornherein vereinbart. Die Entropie ist also der erwartete (mittlere) Informationsgehalt aller Ereignisse.

Es folgt eine entsprechende Definition für endlich diskrete Zufallsvektoren $(X, Y)'$ mit Träger $\mathcal{X} \times \mathcal{Y}$, $\mathcal{X} = \{x_1, \dots, x_m\}$, $\mathcal{Y} = \{y_1, \dots, y_n\}$, und der Verteilung

$$P(X = x_i, Y = y_j) = p_{ij}, \quad \text{wobei} \quad p_{ij} \geq 0, \quad \sum_{i,j} p_{ij} = 1.$$

Definition 3.1.2. $(X, Y)'$ sei ein endlicher diskreter Zufallsvektor. Dann heißt

a) $H(X, Y) = -\sum_{i,j} P(X = x_i, Y = y_j) \log P(X = x_i, Y = y_j)$
 (*gemeinsame*) Entropie von $(X, Y)'$.

b) $H(X|Y) = -\sum_j P(Y = y_j) \sum_i P(X = x_i|Y = y_j) \log P(X = x_i|Y = y_j)$
 $= -\sum_{i,j} P(X = x_i, Y = y_j) \log P(X = x_i|Y = y_j)$
 bedingte Entropie oder Äquivokation (*equivocation*) von X unter Y .

$H(X|Y)$ ist also die mittlere Entropie bezüglich der bedingten Verteilungen.

Proposition 3.1.3. X, Y seien endlich diskrete Zufallsvariablen. Dann gilt

- a) $0 \leq H(X) \leq \log m$
 "=" links $\Leftrightarrow X$ ist einpunktverteilt, d.h. $\exists x_i : P(X = x_i) = 1$
 "=" rechts $\Leftrightarrow X$ ist gleichverteilt, d.h. $P(X = x_i) = \frac{1}{m} \quad \forall i = 1, \dots, m$
- b) $0 \leq H(X|Y) \leq H(X)$
 "=" links $\Leftrightarrow P(X = x_i|Y = y_j) = 1 \quad \forall i, j$ mit $P(X = x_i, Y = y_j) > 0$,
 d.h. X ist total abhängig von Y
 "=" rechts $\Leftrightarrow X, Y$ sind stochastisch unabhängig
- c) $H(X) \leq H(X, Y) \leq H(X) + H(Y)$
 "=" links $\Leftrightarrow Y$ ist total abhängig von X
 "=" rechts $\Leftrightarrow X, Y$ sind stochastisch unabhängig

Definition 3.1.4. X, Y seien endlich diskrete Zufallsvariablen. Dann heißt

$$I(X; Y) = H(X) - H(X|Y)$$

Transinformation oder *Synentropie* (*transinformation, synentropy, mutual information*) von X und Y .

Die Größe $I(X; Y)$ misst, um wieviel die Unbestimmtheit von X im Mittel sinkt, wenn das Ergebnis von Y bekannt ist. Es gilt

$$\begin{aligned} I(X; Y) &= \sum_{i,j} P(X = x_i, Y = y_j) \log \frac{P(X = x_i|Y = y_j)}{P(X = x_i)} \\ &= \sum_{i,j} P(X = x_i, Y = y_j) \log \frac{P(X = x_i, Y = y_j)}{P(X = x_i)P(Y = y_j)}. \end{aligned}$$

Es folgt ein wichtiger Zusammenhang mit der Kullback-Leibler-Distanz. Seien dazu

$$\mathbf{p} = (p_1, \dots, p_k), \quad \mathbf{q} = (q_1, \dots, q_k), \quad p_i, q_i \geq 0, \quad \sum_{i=1}^k p_i = \sum_{i=1}^k q_i = 1$$

diskrete Verteilungen. Dann heißt

$$D(\mathbf{p} \parallel \mathbf{q}) = \sum_{l=1}^k p_l \log \frac{p_l}{q_l}$$

Kullback-Leibler-Distanz zwischen \mathbf{p} und \mathbf{q} . Ersetzt man unter einer geeigneten Abzählung

$$p_l = P(X = x_i, Y = y_j) , \quad q_l = P(X = x_i)P(Y = y_j) ,$$

so folgt, dass $I(X; Y)$ die Kullback-Leibler-Distanz zwischen der gemeinsamen Verteilung $(P(X = x_i, Y = y_j))_{i=1, \dots, m, j=1, \dots, n}$ und der Produktverteilung bei stochastischer Unabhängigkeit $(P(X = x_i)P(Y = y_j))_{i=1, \dots, m, j=1, \dots, n}$ ist.

Beispiel 3.1.5 (Binärer symmetrischer Kanal, BSC (binary symmetric channel)). Bits werden durch einen Kanal übertragen und zufällig verfälscht. Die Bits werden mit Wahrscheinlichkeit ϵ falsch und mit Wahrscheinlichkeit $1 - \epsilon$ richtig übertragen. Sehen Sie hierzu auch Abbildung 3.2. Als Modell der Übertragung dienen die Zufallsvariablen

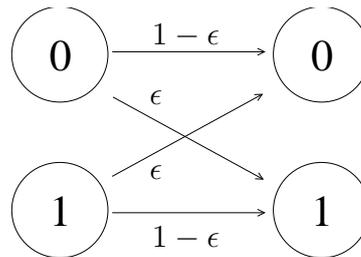


Abbildung 3.2: Bitübertragung mit zufälliger Verfälschung.

$X, Y \in \{0, 1\}$, wobei X die Eingabe und Y die Ausgabe beschreibt. Für die bedingten Wahrscheinlichkeiten gilt dann

$$\begin{aligned} P(Y = 0|X = 0) &= P(Y = 1|X = 1) = 1 - \epsilon , \\ P(Y = 0|X = 1) &= P(Y = 1|X = 0) = \epsilon . \end{aligned}$$

Wir nehmen weiterhin an, dass die Eingabebits gleichverteilt auf $\{0, 1\}$ sind, d.h.

$$P(X = 0) = P(X = 1) = \frac{1}{2} .$$

Hieraus können alle benötigten Verteilungen berechnet werden.

a) Die gemeinsame Verteilung lautet

$$P(X = 0, Y = 0) = P(Y = 0|X = 0) \cdot P(X = 0) = (1 - \epsilon)^{\frac{1}{2}}, \text{ analog dazu} \\ P(X = 1, Y = 1) = \frac{1-\epsilon}{2}, P(X = 0, Y = 1) = P(X = 1, Y = 0) = \frac{\epsilon}{2}.$$

b) Hieraus folgt die Verteilung von Y zu

$$P(Y = 0) = P(Y = 0, X = 0) + P(Y = 0, X = 1) = \frac{1}{2}, P(Y = 1) = \frac{1}{2}.$$

c) Die bedingte Verteilung von X gegeben Y lautet

$$P(X = 0|Y = 0) = \frac{P(X=0, Y=0)}{P(Y=0)} = 1 - \epsilon, \text{ und analog dazu} \\ P(X = 1|Y = 1) = 1 - \epsilon, P(X = 0|Y = 1) = P(X = 1|Y = 0) = \epsilon.$$

Durch Einsetzen in die Definitionen mit Logarithmen zur Basis 2 folgt

$$\begin{aligned} H(X) &= H(Y) = 1 \\ H(X, Y) &= 1 - (1 - \epsilon) \log_2(1 - \epsilon) - \epsilon \log_2 \epsilon \\ H(X|Y) &= -(1 - \epsilon) \log_2(1 - \epsilon) - \epsilon \log_2 \epsilon \\ I(X; Y) &= 1 + (1 - \epsilon) \log_2(1 - \epsilon) + \epsilon \log_2 \epsilon \end{aligned}$$

Diese Größen sind als Funktion von ϵ in Abbildung 3.3 grafisch dargestellt.

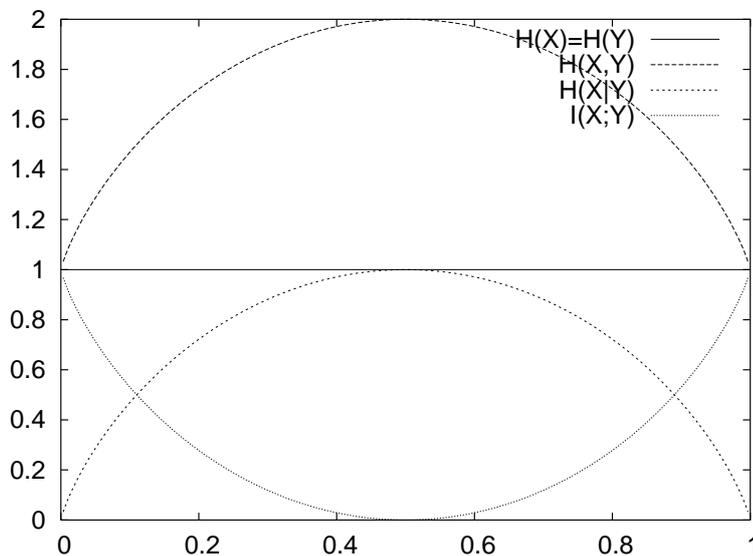


Abbildung 3.3: $H(X) = H(Y)$, $H(X, Y)$, $H(X|Y)$ und $I(X; Y)$.

Proposition 3.1.6. X, Y seien endlich diskrete Zufallsvariablen. Dann gilt

- a) $H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$,
 b) $I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$
 $= H(X) + H(Y) - H(X, Y)$.

Beweis.

a)

$$\begin{aligned} H(X, Y) &= - \sum_{i,j} P(X = x_i, Y = y_j) \\ &\quad \times [\log P(X = x_i, Y = y_j) - \log P(X = x_i) + \log P(X = x_i)] \\ &= - \sum_{i,j} P(X = x_i, Y = y_j) \log P(Y = y_j | X = x_i) - \\ &\quad - \underbrace{\sum_i \sum_j P(X = x_i, Y = y_j) \log P(X = x_i)}_{P(X=x_i)} \\ &\stackrel{D.3.1.2b)}{=} H(Y|X) + H(X) \end{aligned}$$

Die zweite Gleichung wird analog hergeleitet.

- b) $I(X; Y) \stackrel{D.3.1.4}{=} H(X) - H(X|Y) \stackrel{a)}{=} H(X) + H(Y) - H(X, Y)$
 $\stackrel{a)}{=} H(Y) - H(Y|X)$.

□

3.2 Quellenkodierung

Ziel der *Quellenkodierung* ist es, Symbole einer Quelle möglichst dicht zu komprimieren. Dabei sollen die ursprünglichen Symbole eindeutig rekonstruiert werden können. Der folgende Abschnitt behandelt die Frage, was die beste erreichbare Kompression ist.

Betrachtet wird ein *Quellalphabet*

$$\mathcal{X} = \{x_1, \dots, x_m\} ,$$

sowie ein *Kodealphabet*

$$\mathcal{Y} = \{y_1, \dots, y_d\} .$$

Ein Beispiel für ein Quellalphabet ist $\mathcal{X} = \{A, \dots, Z\}$ und für ein Kodealphabet $\mathcal{Y} = \{0, 1\}$.

Buchstaben $x_j \in \mathcal{X}$ werden mit Wörtern über \mathcal{Y} kodiert, beschrieben durch eine *injektive* Abbildung

$$g: \mathcal{X} \rightarrow \bigcup_{l=1}^{\infty} \mathcal{Y}^l: x_j \mapsto g(x_j) = (w_{j_1}, \dots, w_{j_{n_j}}) .$$

$g(x_j)$ heißt *Kodewort* von x_j , n_j heißt *Länge* von $g(x_j)$ und g heißt *Kode* oder *Kodierung*.

Injektivität der Abbildung wird gefordert, damit jedes Kodewort eindeutig zugeordnet werden kann. Wichtig ist, dass auch Konkatenationen von Kodewörtern eindeutig dekodierbar sind.

Definition 3.2.1. Ein Kode g heißt *eindeutig dekodierbar (e.d.)* (*uniquely decodable, uniquely decipherable*), wenn die Abbildung

$$G: \bigcup_{l=1}^{\infty} \mathcal{X}^l \rightarrow \bigcup_{l=1}^{\infty} \mathcal{Y}^l: (x_1, \dots, x_n) \mapsto (g(x_1), \dots, g(x_n))$$

injektiv ist.

Eine wichtige Klasse von eindeutig dekodierbaren Codes bilden die präfixfreien Codes.

Definition 3.2.2. Ein Kode g heißt *präfixfrei (PF)* (*prefix code, instantaneous code*), wenn kein Kodewort Präfix eines Anderen ist. Dabei heißt $\mathbf{a} \in \mathcal{Y}^k$ *Präfix* von $\mathbf{b} \in \mathcal{Y}^l$, $k \leq l$, wenn ein $\mathbf{c} \in \mathcal{Y}^{l-k}$ existiert mit $\mathbf{b} = (\mathbf{a}, \mathbf{c})$.

Proposition 3.2.3. Präfixfreie Codes sind eindeutig dekodierbar.

Diese Aussage sieht man wie folgt ein. Die Dekodierung eines Strings von links beginnend bis zur Identifizierung des ersten Kodewortes ist eindeutig. Dies folgt aus der PF-Eigenschaft. Bei analoger Vorgehensweise gilt gleiches auch sukzessive für den Rest.

Eine Maßzahl für die Güte eines Codes ist die mittlere bzw. erwartete Kodewortlänge. Sei dazu X eine Zufallsvariable mit Träger \mathcal{X} und Verteilung

$$P(X = x_j) = p_j, \quad j = 1, \dots, m .$$

X beschreibt das zufällige Auftreten von Quellbuchstaben. g sei ein Kode mit Kodewortlängen n_1, \dots, n_m . Dann ist

$$\bar{n} = \bar{n}(g) = \sum_{j=1}^m n_j P(X = x_j)$$

die *erwartete Kodewortlänge* von g .

Theorem 3.2.4 (Noiseless Coding Theorem, Shannon 1949).

a) Für alle eindeutig dekodierbaren Codes g gilt

$$\frac{H(X)}{\log d} \leq \bar{n}(g) .$$

b) Es existiert ein präfixfreier Kode g mit

$$\bar{n}(g) \leq \frac{H(X)}{\log d} + 1 .$$

Zum Beweis von Teil b) wird ein Kode konstruiert, der die Ungleichung erfüllt. Teil a) benutzt die Ungleichung von *McMillan (1959)*:

$$\text{Für alle e.d. Codes mit Kodewortlängen } n_1, \dots, n_m \text{ gilt } \sum_{j=1}^m d^{-n_j} \leq 1 .$$

und *Kraft (1949)*:

$$\text{Gilt } \sum_{j=1}^m d^{-n_j} \leq 1 , \text{ so existiert ein PF-Kode mit Kodewortlängen } n_1, \dots, n_m .$$

Mit der Normierung $\log d$ (z.B. $\log 2$ bei binärer Kodierung) ist $H(X)$ eine untere Schranke für $\bar{n}(g)$ für jeden eindeutig dekodierbaren Kode g . Die Schranke wird nicht immer erreicht, asymptotisch jedoch für Blockcodes.

Betrachtet werden im Folgenden Blockcodes für Blöcke der Länge N . Als Modell dient ein Zufallsvektor $\mathbf{X} = (X_1, \dots, X_N)'$ mit Träger \mathcal{X}^N , der das zufällige Auftreten von Blöcken der Länge N beschreibt. Blockkodierung wird beschrieben durch eine injektive Funktion

$$g_N : \mathcal{X}^N \rightarrow \bigcup_{l=1}^{\infty} \mathcal{Y}^l : (a_1, \dots, a_N) \mapsto g_N(a_1, \dots, a_N) .$$

Diese Situation ist analog zu der vorher Behandelten bis auf den Unterschied, dass jetzt \mathcal{X}^N statt \mathcal{X} das Alphabet ist. Mit Theorem 3.2.4 folgt, dass ein eindeutig dekodierbarer Kode g_N , der sogar präfixfrei gewählt werden kann, existiert mit

$$\frac{H(X_1, \dots, X_N)}{\log d} \leq \bar{n}(g_N) \leq \frac{H(X_1, \dots, X_N)}{\log d} + 1 .$$

Bei Division durch N erhält man

$$\frac{1}{\log d} \frac{H(X_1, \dots, X_N)}{N} \leq \frac{\bar{n}(g_N)}{N} \leq \frac{1}{\log d} \frac{H(X_1, \dots, X_N)}{N} + \frac{1}{N} . \quad (*)$$

$\frac{\bar{n}(g_N)}{N}$ ist die erwartete Kodelänge pro Quellbuchstabe. Sind X_1, \dots, X_N i.i.d., so folgt mit Proposition 3.1.3 c)

$$\frac{1}{\log d} \frac{N H(X_1)}{N} \leq \frac{\bar{n}(g_N)}{N} \leq \frac{1}{\log d} \frac{N H(X_1)}{N} + \frac{1}{N} .$$

Der Grenzübergang $N \rightarrow \infty$ liefert folgende Aussage.

Proposition 3.2.5. $\{X_n\}_{n \in \mathbb{N}}$ sei eine Folge von i.i.d. Zufallsvariablen. Es existiert eine Folge von präfixfreien Codes g_N mit

$$\lim_{N \rightarrow \infty} \frac{\bar{n}(g_N)}{N} = \frac{H(X_1)}{\log d} .$$

Quellen, die durch i.i.d. Zufallsvariablen $\{X_n\}_{n \in \mathbb{N}}$ modelliert werden, heißen *diskrete, gedächtnislose Quellen* (DMS, *discrete memoryless source*).

Eine Folge von Zufallsvariablen $\{X_n\}_{n \in \mathbb{N}}$ heißt (*stark*) *stationär*, wenn

$$P(X_{i_1} = a_1, \dots, X_{i_k} = a_k) = P(X_{i_1+s} = a_1, \dots, X_{i_k+s} = a_k)$$

für alle $1 \leq i_1 < \dots < i_k \in \mathbb{N}$, $s \in \mathbb{N}_0$, $a_1, \dots, a_k \in \mathcal{X}$ gilt.

Im Folgenden werde nun eine stationäre Quelle mit Gedächtnis modelliert. Eine stationäre Folge von Zufallsvariablen $\{X_n\}_{n \in \mathbb{N}}$ heißt *stationäre Quelle* (*stationary source*).

Proposition 3.2.6. Sei $\{X_n\}_{n \in \mathbb{N}}$ stationäre Quelle, dann gilt:

- a) $H(X_N | X_1, \dots, X_{N-1})$ ist monoton fallend in N .
- b) $H(X_N | X_1, \dots, X_{N-1}) \leq \frac{1}{N} H(X_1, \dots, X_N) \quad \forall N \in \mathbb{N}$.
- c) $\frac{1}{N} H(X_1, \dots, X_N)$ ist monoton fallend in N .
- d) $\lim_{N \rightarrow \infty} H(X_N | X_1, \dots, X_{N-1}) = \lim_{N \rightarrow \infty} \frac{1}{N} H(X_1, \dots, X_N) = H_\infty(X)$ (, d.h. beide Grenzwerte existieren).

Definition 3.2.7. Sei $\{X_n\}_{n \in \mathbb{N}}$ stationäre Quelle, dann heißt $H_\infty(X)$ *Entropie* der Quelle.

Insbesondere gilt $H_\infty(X) = H(X_1)$ für $\{X_n\}_{n \in \mathbb{N}}$ i.i.d.

Beispiel 3.2.8 (Englischer Text). In folgender Tabelle sind auszugsweise die Buchstabenhäufigkeiten in der englischen Sprache dargestellt. Diese werden als Wahrscheinlichkeiten interpretiert.

A	B	C	...	E	...	Z
8.04%	1.54%	3.06%	...	12.51%	...	0.09%

Ähnlich können auch Wahrscheinlichkeiten für Digramme festgestellt werden.

AN	TH	\dots	TI
1.81%	3.21%	\dots	1.28%

Die zugehörigen Entropien ergeben sich folgendermaßen.

N	1	2	3	\dots
$\frac{1}{N} H(X_1, \dots, X_N)$	4.14	3.56	3.30	\dots

Aus Experimenten (log zur Basis 2) ergibt sich

$$1 < H_\infty(X) \leq 1.5 (!)$$

Für Blockcodes gilt

$$\frac{1}{\log d} \cdot \frac{H(X_1, \dots, X_N)}{N} \leq \frac{\bar{n}(g^{(N)})}{N} < \frac{1}{\log d} \cdot \frac{H(X_1, \dots, X_N)}{N} + \frac{1}{N},$$

sodass sich folgende Proposition ergibt.

Proposition 3.2.9. Sei $\{X_n\}_{n \in \mathbb{N}}$ stationäre Quelle, dann existiert eine Folge von PF-Blockcodes $g^{(N)}$ mit

$$\lim_{N \rightarrow \infty} \frac{\bar{n}(g^{(N)})}{N} = \frac{H_\infty(X)}{\log d}.$$

Ein englischer Text kann demnach mit genügend langen Blockcodes mit weniger als 1.5 Bit pro Buchstabe kodiert werden.

3.3 Kanalkapazität und Fundamentalsatz

Die Übertragung von Symbolen in einem Kanal wird modelliert durch den Zufallsvektor (X, Y) , wobei die Zufallsvariable X den Träger $\mathcal{X} = \{x_1, \dots, x_m\}$ und Y den Träger $\mathcal{Y} = \{y_1, \dots, y_d\}$ besitzt. X bedeutet die Eingabe und Y die Ausgabe des Kanals. Die Übertragungswahrscheinlichkeiten lauten

$$P(Y = y_j | X = x_i) = p(j|i), \quad i = 1, \dots, m, \quad j = 1, \dots, d.$$

$p(j|i)$ ist die Wahrscheinlichkeit, dass y_j ausgegeben wird, wenn x_i gesendet wurde. Bezeichne $\mathbf{p} = (p_1, \dots, p_m)$, $p_i = P(X = x_i)$, $i = 1, \dots, m$ die Verteilung von X . Die Transinformation des Kanals ist dann nach Definition 3.1.4 gegeben durch

$$\begin{aligned} I(X; Y) &= \sum_{i,j} P(X = x_i, Y = y_j) \log \underbrace{\frac{P(X = x_i | Y = y_j)}{P(X = x_i)}}_{= \frac{P(Y = y_j | X = x_i)}{P(Y = y_j)}} \\ &= \sum_{i,j} p_i p(j|i) \log \frac{p(j|i)}{\sum_l p_l p(j|l)}. \end{aligned}$$

Dies ist ein Maß für die Güte der Informationsübertragung im Kanal. Ziel ist es, $I(X; Y)$ über alle Inputverteilungen (p_1, \dots, p_m) zu maximieren. Dies führt zum Begriff der Kanalkapazität.

Definition 3.3.1.

$$C = \max_{(p_1, \dots, p_m)} I(X; Y)$$

heißt *Kanalkapazität (channel capacity)*.

Beispiel 3.3.2 (Binärer symmetrischer Kanal, BSC). Nach Proposition 3.1.6 gilt

$$I(X; Y) = H(Y) - H(Y|X) ,$$

wobei

$$\begin{aligned} H(Y|X) &= - \sum_i P(X = x_i) \underbrace{\sum_j P(Y = y_j | X = x_i) \log P(Y = y_j | X = x_i)}_{=\epsilon \log \epsilon + (1-\epsilon) \log(1-\epsilon) \text{ (unabhängig von } i \text{ und } j)} \\ &= -\epsilon \log \epsilon - (1 - \epsilon) \log(1 - \epsilon) . \end{aligned}$$

$H(Y|X)$ hängt also nur von ϵ , nicht aber von der Inputverteilung (p_0, p_1) ab. Die Maximierung von $I(X; Y)$ ist also äquivalent zur Maximierung von $H(Y)$ über (p_0, p_1) . $H(Y)$ ist gemäß Proposition 3.1.3 a) maximal, wenn

$$P(Y = 0) = P(Y = 1) = \frac{1}{2} .$$

Dies ist nach Beispiel 3.1.5 der Fall, wenn gilt

$$p_0 = p_1 = \frac{1}{2} .$$

Insgesamt beträgt die Kanalkapazität des BSC damit

$$C = 1 + (1 - \epsilon) \log_2(1 - \epsilon) + \epsilon \log_2 \epsilon ,$$

sie wird angenommen für die Gleichverteilung $p_0 = p_1 = \frac{1}{2}$.

Es werden nicht nur Symbole einzeln übertragen, sondern auch Wörter oder Blöcke der Länge N . Zur Beschreibung ist die folgende direkte Erweiterung des Modells geeignet. Die Eingabe wird durch einen Zufallsvektor $\mathbf{X}_N = (X_1, \dots, X_N)' \in \mathcal{X}^N$ und die Ausgabe durch den Zufallsvektor $\mathbf{Y}_N = (Y_1, \dots, Y_N)' \in \mathcal{Y}^N$ modelliert. In der Regel gilt $\mathcal{X} = \mathcal{Y}$. Insbesondere ist bei der binären Übertragung $\mathcal{X} = \mathcal{Y} = \{0, 1\}$.

Definition 3.3.3. Ein Kanal heißt *diskret und gedächtnislos (DMC, discrete memoryless channel)*, wenn für alle $\mathbf{a} = (a_1, \dots, a_N) \in \mathcal{X}^N$, $\mathbf{b} = (b_1, \dots, b_N) \in \mathcal{Y}^N$ gilt

$$P(\mathbf{Y}_N = \mathbf{b}_N | \mathbf{X}_N = \mathbf{a}_N) = \prod_{i=1}^N P(Y_i = b_i | X_i = a_i) .$$

Eine genauere Beschreibung der Übertragung von Kodewörtern im Kanal ergibt sich wie folgt. Der Eingabekode aus M Kodewörtern der Länge N

$$\mathcal{C}_N = \{\mathbf{c}_1, \dots, \mathbf{c}_M\} \subseteq \mathcal{X}^N$$

wird in einem Kanal mit Übertragungswahrscheinlichkeiten

$$p_N(\mathbf{b}_N | \mathbf{a}_N) = P(\mathbf{Y}_N = \mathbf{b}_N | \mathbf{X}_N = \mathbf{a}_N)$$

übertragen. Nötig ist eine *Dekodierregel*, beschrieben durch eine Funktion

$$h_N: \mathcal{Y}^N \rightarrow \mathcal{C}_N: \mathbf{b}_N \mapsto h_N(\mathbf{b}_N) .$$

Eine grafische Veranschaulichung hierzu liefert Abbildung 3.4.

Zur Bewertung der Übertragungsgüte werden die folgenden Fehlerwahrscheinlichkeiten herangezogen. Die Wahrscheinlichkeit für eine Fehldekodierung, wenn \mathbf{c}_j gesendet wurde, beträgt

$$e_j(\mathcal{C}_N) = P(h(\mathbf{Y}_N) \neq \mathbf{c}_j | \mathbf{X}_N = \mathbf{c}_j) .$$

Hieraus ergibt sich die Wahrscheinlichkeit für einen *Dekodierfehler* als gewichtetes Mittel

$$e(\mathcal{C}_N) = \sum_{j=1}^M e_j(\mathcal{C}_N) P(\mathbf{X}_N = \mathbf{c}_j) .$$

Der schlechteste Fall bei der Fehldekodierung wird durch die maximale Fehlerwahrscheinlichkeit erfasst,

$$\hat{e}(\mathcal{C}_N) = \max_{1 \leq j \leq M} e_j(\mathcal{C}_N) .$$

Im Folgenden werden Kriterien für den Entwurf von Dekodierregeln betrachtet.

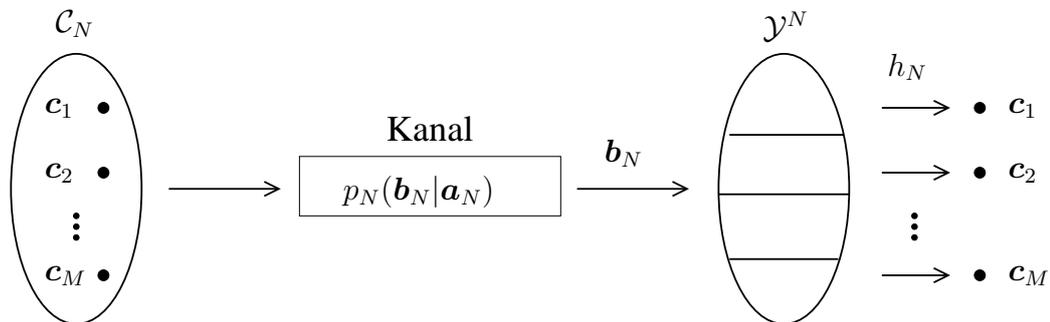


Abbildung 3.4: Kodewortübertragung von Wörtern der Länge N über den diskreten, gedächtnislosen Kanal.

Definition 3.3.4. Eine Dekodierregel $h_N: \mathcal{Y}^N \rightarrow \mathcal{C}$ heißt *ME-Dekodierung* (*minimum error rule, ideal observer*), wenn

$$\mathbf{c}_j = h_N(\mathbf{b}_N) \Rightarrow P(\mathbf{X}_N = \mathbf{c}_j | \mathbf{Y}_N = \mathbf{b}_N) \geq P(\mathbf{X}_N = \mathbf{c}_i | \mathbf{Y}_N = \mathbf{b}_N) \quad \forall i = 1, \dots, M$$

gilt.

Bei ME-Dekodierung wird die Wahrscheinlichkeit maximiert, \mathbf{c}_j gesendet zu haben, wenn \mathbf{b}_N empfangen wird. Der Vorteil liegt darin, dass bei fester Inputverteilung $P(\mathbf{X}_N = \mathbf{c}_j)$, $j = 1, \dots, M$, bei ME-Dekodierung $e(\mathcal{C}_N)$ minimiert wird, wie man zeigen kann. Der Nachteil ist die Abhängigkeit der ME-Dekodierung von der Inputverteilung. Dieser Nachteil entfällt bei ML-Dekodierung.

Definition 3.3.5. Eine Dekodierregel $h_N: \mathcal{Y}^N \rightarrow \mathcal{C}$ heißt *ML-Dekodierung* (*maximum likelihood decoding rule*), wenn

$$\mathbf{c}_j = h_N(\mathbf{b}_N) \Rightarrow P(\mathbf{Y}_N = \mathbf{b}_N | \mathbf{X}_N = \mathbf{c}_j) \geq P(\mathbf{Y}_N = \mathbf{b}_N | \mathbf{X}_N = \mathbf{c}_i) \quad \forall i = 1, \dots, M$$

gilt.

ML-Dekodierungen werden so konstruiert, dass die Wahrscheinlichkeit \mathbf{b}_N zu empfangen, wenn \mathbf{c}_j gesendet wurde, maximiert wird.

Ist die Inputverteilung eine Gleichverteilung, gilt also $P(\mathbf{X} = \mathbf{c}_i) = \frac{1}{M}$ für alle $i = 1, \dots, M$, so ist jede ME-Dekodierung eine ML-Dekodierung und umgekehrt. ME- und ML-Dekodierung sind in diesem Fall äquivalent.

Theorem 3.3.6 (*Shannonscher Fundamentalsatz, Channel Coding Theorem*). Gegeben sei ein diskreter, gedächtnisloser Kanal mit Kanalkapazität C und R mit $0 < R < C$. $M_N \in \mathbb{N}$ sei eine Folge mit

$$\frac{\log M_N}{N} < R .$$

Dann existiert eine Folge von Codes \mathcal{C}_N mit M_N Codewörtern der Länge N und eine Konstante $a > 0$ mit

$$\hat{e}(\mathcal{C}_N) \leq e^{-Na} \xrightarrow{N \rightarrow \infty} 0 .$$

Der Beweis dieses Satzes ist nicht konstruktiv. Man kann nicht ablesen, wie die Codes \mathcal{C}_N zu wählen sind. Als Dekodierregel werden sogenannte Maximum-Likelihood-Dekodierer gewählt.

Beispiel 3.3.7. Gegeben sei ein BSC mit $\epsilon = 0.03$. Aus Beispiel 3.1.5 ergibt sich die Kanalkapazität zu

$$C = 1 + (1 - \epsilon) \log_2(1 - \epsilon) + \epsilon \log_2 \epsilon = 0.8056 .$$

Wähle $R = 0.8$, sodass $R < C$. Nach Theorem 3.3.6 gilt

$$\frac{\log_2 M_N}{N} < R, \text{ also } M_N < 2^{NR}.$$

Wähle also

$$M_N = \lfloor 2^{0.8N} \rfloor.$$

Die folgende Tabelle enthält einige konkrete Zahlen zu dieser Wahl.

N	10	20	30
$ \mathcal{X}^N = 2^N$	1024	1048576	$1.0737 \cdot 10^9$
$M_N = \lfloor 2^{0.8N} \rfloor$	256	65536	$16.777 \cdot 10^6$
Anteil benutzter Kodewörter	25%	6.25%	1.56%

Zur Vereinfachung wird im Folgenden $N \cdot R \in \mathbb{N}$ angenommen. Die Wirkungsweise des Kanals wird jetzt in zeitbezogenen Einheiten beschrieben. Der Kanal übertrage 1 Symbol pro Zeiteinheit. Die Quelle darf dann höchstens R Symbole pro Zeiteinheit produzieren. Dies sieht man wie folgt ein. Bei der Übertragung von Blöcken der Länge N stehen 2^{NR} Kodewörter zur Verfügung. Die Quelle darf höchstens Wörter der Länge $N \cdot R$ erzeugen, damit für jedes Wort der Quelle ein Kodewort zur Verfügung steht. Aus diesen Überlegungen heraus nennt man R die *Quellenrate*.

In Kurzform lässt sich der Fundamentalsatz wie folgt fassen. Ist die Quellenrate kleiner als die Kapazität, so existieren für große Blocklängen Codes beliebig kleiner Fehlerwahrscheinlichkeit.

Als Preis für die kleine Fehlerwahrscheinlichkeit muss man bei den großen Blocklängen aufwendige Kodier- und Dekodieralgorithmen in Kauf nehmen. Ausserdem muss vor der ersten Übertragung zunächst ein Block der Länge N abgewartet werden. Ferner sind konkrete Codes, die exponentiell schnell fallende Fehlerwahrscheinlichkeiten haben, nicht bekannt.

3.4 Differentielle Entropie und Transinformation

Zunächst eine Wiederholung der Definitionen 3.1.1 und 3.1.4. Gegeben seien zwei endlich diskrete Zufallsvariablen X und Y . Dann heißt

$$H(X) = - \sum_{i=1}^m P(X = x_i) \log P(X = x_i)$$

Entropie von X und

$$I(X; Y) = H(X) - H(X|Y) = \sum_{i,j} P(X = x_i, Y = y_j) \log \frac{P(X = x_i|Y = y_j)}{P(X = x_i)}$$

Transinformation von X und Y . Haben die Träger von X und Y dieselbe Mächtigkeit, so ist die *Kullback-Leibler-Distanz* gegeben durch

$$D(X \parallel Y) = \sum_l P(X = x_l) \log \frac{P(X = x_l)}{P(Y = y_l)} .$$

Ziel ist nun die Übertragung dieser Konzepte auf absolut-stetige Verteilungen, also auf Verteilungen, die eine Dichte besitzen (s. Definition 2.4.3).

Definition 3.4.1. X sei eine absolut-stetige Zufallsvariable mit Dichte $f(x)$. Dann heißt

$$H(X) = - \int_{-\infty}^{\infty} f(x) \log f(x) dx$$

differentielle Entropie von X (differential entropy).

Hierbei wird vorausgesetzt, dass das Integral $\int_{-\infty}^{\infty} f(x) \log f(x) dx$ existiert. Dies ist jedoch nicht immer der Fall. Es existieren Dichten, für die obiges Integral nicht definiert ist. Die differentielle Entropie lässt sich wie im diskreten Fall schreiben als

$$H(X) = E(-\log f(X)) .$$

Benutzt wird auch die Schreibweise $H(X) = H(f)$.

Beispiel 3.4.2.

- a) Sei $X \sim R(0, a)$, d.h. X ist gleichverteilt auf dem Intervall $[0, a]$, s. Beispiel 2.3.6 b). Dann ist die differentielle Entropie von X gegeben durch

$$H(X) = - \int_0^a \frac{1}{a} \log \frac{1}{a} dx = -\frac{1}{a} \log \frac{1}{a} \int_0^a dx = \log a .$$

- b) Sei $X \sim N(\mu, \sigma^2)$, d.h. X ist normalverteilt mit $E(X) = \mu$, $\text{Var}(X) = \sigma^2$, s. Beispiel 2.3.6 a). Dann ist die differentielle Entropie von X gegeben durch

$$H(X) = \frac{1}{2} \ln(2\pi e \sigma^2) .$$

Die Übertragung des Konzepts auf mehrdimensionale Dichten und gemeinsame Verteilungen geschieht analog zum bisherigen Vorgehen.

Definition 3.4.3. a) $(X_1, \dots, X_n)'$ sei ein Zufallsvektor mit Dichte $f(x_1, \dots, x_n)$. Dann heißt

$$H(X_1, \dots, X_n) = - \int \cdots \int f(x_1, \dots, x_n) \log f(x_1, \dots, x_n) dx_1 \cdots dx_n$$

gemeinsame differentielle Entropie von $(X_1, \dots, X_n)'$.

b) $(X, Y)'$ sei ein Zufallsvektor mit Dichte $f(x, y)$ und

$$f(x|y) = \begin{cases} \frac{f(x,y)}{f_Y(y)} & , \text{ falls } f_Y(y) > 0 \\ 0 & , \text{ sonst} \end{cases}$$

bezeichne die bedingte Dichte von X unter $Y = y$. Dann heißt

$$H(X|Y) = - \int \int f(x, y) \log f(x|y) dx dy$$

bedingte differentielle Entropie von X unter Y .

Vorausgesetzt wird hierbei die Existenz der auftretenden Integrale.

Beispiel 3.4.4. Sei $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \mathbf{C})$, d.h. n -dimensional normalverteilt gemäß Beispiel 2.4.4 mit Dichte

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{C}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}, \quad \mathbf{x} \in \mathbb{R}^n.$$

Es gilt

$$H(\mathbf{X}) = \frac{1}{2} \ln ((2\pi e)^n |\mathbf{C}|) ,$$

denn

$$\begin{aligned} H(\mathbf{X}) &= - \int f(\mathbf{x}) \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) - \ln \left((2\pi)^{\frac{n}{2}} |\mathbf{C}|^{\frac{1}{2}} \right) \right] d\mathbf{x} \\ &= \frac{1}{2} \mathbb{E} [(\mathbf{X} - \boldsymbol{\mu})' \mathbf{C}^{-1} (\mathbf{X} - \boldsymbol{\mu})] + \frac{1}{2} \ln ((2\pi)^n |\mathbf{C}|) \\ &= \frac{1}{2} \mathbb{E} [\text{tr} (\mathbf{C}^{-1} (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})')] + \frac{1}{2} \ln ((2\pi)^n |\mathbf{C}|) \\ &= \frac{1}{2} \text{tr} \left(\mathbf{C}^{-1} \underbrace{\mathbb{E}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'}_{\substack{\text{(Def. 2.4.5)} \\ \mathbf{C}}} \right) + \frac{1}{2} \ln ((2\pi)^n |\mathbf{C}|) \\ &= \frac{1}{2} \text{tr} \mathbf{I}_n + \frac{1}{2} \ln ((2\pi)^n |\mathbf{C}|) = \frac{1}{2} \ln ((2\pi e)^n |\mathbf{C}|) . \end{aligned}$$

Definition 3.4.5. $(X, Y)'$ sei ein Zufallsvektor mit Dichte $f(x, y)$. Dann heißt

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

(differentielle) Transinformation oder *Synentropie (mutual information)*.

$I(X; Y)$ misst, um wieviel die Unbestimmtheit über das durch X beschriebene Zufallsexperiment fällt, wenn Y bekannt ist. Insofern mißt $I(X; Y)$ den Informationsgehalt von Y über X . Es gilt analog zum endlich diskreten Fall

$$I(X; Y) = \int \int f(x, y) \log \frac{f(x, y)}{f_X(x)f_Y(y)} dx dy = \int \int f(x, y) \log \frac{f(x|y)}{f_X(x)} dx dy ,$$

wobei $f_X(x)$ und $f_Y(y)$ die Dichten von X und Y bezeichnen. Dies sieht man wie folgt ein.

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) = - \int f_X(x) \log f_X(x) dx + \int \int f(x, y) \log f(x|y) dx dy \\ &= - \int \int f(x, y) \log f_X(x) dy dx + \int \int f(x, y) \log f(x|y) dx dy \\ &= \int \int f(x, y) \log \frac{f(x|y)}{f_X(x)} dx dy . \end{aligned}$$

Definition 3.4.6. (Kullback-Leibler-Distanz) $f(\mathbf{x}), g(\mathbf{x})$ seien n -dimensionale Dichten, $\mathbf{x} = (x_1, \dots, x_n)' \in \mathbb{R}^n$. Dann heißt

$$D(f \parallel g) = \int f(\mathbf{x}) \log \frac{f(\mathbf{x})}{g(\mathbf{x})} d\mathbf{x}$$

Kullback-Leibler-Distanz von f und g .

Es gilt

$$I(X; Y) = D(f_{(X,Y)} \parallel f_X \cdot f_Y) .$$

Dies ist die Kullback-Leibler-Distanz zwischen der gemeinsamen Dichte und der Produktdichte im Fall stochastischer Unabhängigkeit.

Proposition 3.4.7. Es gilt

$$D(f \parallel g) \geq 0$$

mit Gleichheit genau dann, wenn $f = g$ (fast sicher).

Beweis.

$$\begin{aligned} -D(f \parallel g) &= - \int f(\mathbf{x}) \log \frac{f(\mathbf{x})}{g(\mathbf{x})} d\mathbf{x} = \int f(\mathbf{x}) \log \frac{g(\mathbf{x})}{f(\mathbf{x})} d\mathbf{x} \\ &= (\log e) \int f(\mathbf{x}) \ln \frac{g(\mathbf{x})}{f(\mathbf{x})} d\mathbf{x} \quad (\ln x \leq x - 1, x \geq 0) \\ &\stackrel{(i)}{\leq} (\log e) \int f(\mathbf{x}) \left(\frac{g(\mathbf{x})}{f(\mathbf{x})} - 1 \right) d\mathbf{x} \\ &= (\log e) \left[\int_{\mathbf{x}: f(\mathbf{x}) > 0} g(\mathbf{x}) d\mathbf{x} - \int f(\mathbf{x}) d\mathbf{x} \right] \\ &\stackrel{(ii)}{\leq} (\log e)(1 - 1) = 0 . \end{aligned}$$

Gleichheit in (i) und (ii) gilt genau unter der angegebenen Bedingung.

□

Aus Proposition 3.4.7 folgt für beliebige n -dimensionale Dichten f, g

$$-\int f(\mathbf{x}) \log f(\mathbf{x}) \leq -\int f(\mathbf{x}) \log g(\mathbf{x})$$

mit Gleichheit falls $f = g$ (fast sicher). Dies folgt aus der Umformung

$$D(f \parallel g) = \int f \log \frac{f}{g} = \int f \log f - \int f \log g \geq 0 .$$

Proposition 3.4.8. $(X, Y)'$ sei ein Zufallsvektor mit Dichte $f(x, y)$.

- a) $I(X; Y) \geq 0$. Gleichheit gilt genau dann, wenn X, Y stochastisch unabhängig sind.
- b) $H(X|Y) \leq H(X)$. Gleichheit gilt genau dann, wenn X, Y stochastisch unabhängig sind.

Beweis.

- a) Wegen Proposition 3.4.7 gilt

$$I(X; Y) = D(f_{(X,Y)} \parallel f_X \cdot f_Y) \geq 0$$

mit Gleichheit genau dann, wenn $f_{(X,Y)} = f_X \cdot f_Y$ (fast sicher), d.h. X, Y stochastisch unabhängig sind. Sehen Sie hierzu auch Proposition 2.4.8.

- b) Es gilt

$$0 \stackrel{a)}{\leq} I(X; Y) = H(X) - H(X|Y)$$

mit Gleichheit genau dann, wenn X, Y stochastisch unabhängig sind.

□

Proposition 3.4.9. $(X, Y)'$ sei ein Zufallsvektor mit Dichte $f(x, y)$.

- a) $H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$ (Kettenregel)
- b) $H(X, Y) \leq H(X) + H(Y)$. Gleichheit gilt genau dann, wenn X, Y stochastisch unabhängig sind.

Beweis.

a) Es gilt

$$\begin{aligned} H(X, Y) &= - \int \int f(x, y) \log f(x, y) \, dx \, dy \\ &= - \int \int f(x, y) \log (f(y|x) f(x)) \, dx \, dy \\ &= - \int \int f(x, y) \log f(x) \, dy \, dx - \int \int f(x, y) \log f(y|x) \, dx \, dy \\ &= H(X) + H(Y|X) . \end{aligned}$$

Der zweite Teil der Gleichung wird analog gezeigt.

b) Es gilt

$$H(X, Y) \stackrel{a)}{=} H(X) + H(Y|X) \leq H(X) + H(Y) .$$

Die Ungleichung folgt aus Proposition 3.4.8 b), wobei Gleichheit genau dann gilt, wenn X, Y stochastisch unabhängig sind.

□

Proposition 3.4.10. \mathbf{X} sei ein absolut-stetiger Zufallsvektor mit Dichte f . $\mathbf{A} \in \mathbb{R}^{n \times n}$ sei regulär und $\mathbf{b} \in \mathbb{R}^n$, dann gilt

$$H(\mathbf{A}\mathbf{X} + \mathbf{b}) = H(\mathbf{X}) + \log(|\mathbf{A}|) .$$

Hierbei bedeutet $|\mathbf{A}| = |\det \mathbf{A}|$, also der Absolutbetrag der Determinante von \mathbf{A} .

Theorem 3.4.11. \mathbf{X} sei ein absolut-stetiger Zufallsvektor mit $E(\mathbf{X}) = \mathbf{0}$, $\text{Cov}(\mathbf{X}) = E(\mathbf{X}\mathbf{X}') = \mathbf{C}$, \mathbf{C} positiv definit. Dann gilt

$$H(\mathbf{X}) \leq \frac{1}{2} \ln ((2\pi e)^n |\mathbf{C}|) ,$$

d.h. $N_n(\boldsymbol{\mu}, \mathbf{C})$ besitzt die größte Entropie unter allen absolut-stetigen Verteilungen mit Kovarianzmatrix \mathbf{C} .

Beweis. Bezeichne

$$\varphi(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{C}|^{\frac{1}{2}}} e^{-\frac{1}{2} \mathbf{x}' \mathbf{C}^{-1} \mathbf{x}} , \quad \mathbf{x} \in \mathbb{R}^n$$

die Dichte der $N_n(\mathbf{0}, \mathbf{C})$ -Verteilung.

Gelte $\mathbf{X} \sim f$, $E(\mathbf{X}) = \mathbf{0}$, $\text{Cov}(\mathbf{X}) = \mathbf{C} = \int \mathbf{x}\mathbf{x}'f(\mathbf{x})d\mathbf{x}$, dann ist

$$\begin{aligned}
 H(\mathbf{X}) &= - \int f(\mathbf{x}) \ln f(\mathbf{x})d\mathbf{x} \\
 &\leq - \int f(\mathbf{x}) \ln \varphi(\mathbf{x})d\mathbf{x} \quad (\text{wegen Proposition 3.4.7}) \\
 &= - \int f(\mathbf{x}) \ln \left[\frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{C}|^{\frac{1}{2}}} e^{-\frac{1}{2}\mathbf{x}'\mathbf{C}^{-1}\mathbf{x}} \right] d\mathbf{x} \\
 &= \ln \left((2\pi)^{\frac{n}{2}}|\mathbf{C}|^{\frac{1}{2}} \right) + \frac{1}{2} \int \mathbf{x}'\mathbf{C}^{-1}\mathbf{x}f(\mathbf{x})d\mathbf{x} \\
 &= \ln \left((2\pi)^{\frac{n}{2}}|\mathbf{C}|^{\frac{1}{2}} \right) + \frac{1}{2} \int \text{tr} \left(\mathbf{C}^{-\frac{1}{2}}\mathbf{x}\mathbf{x}'\mathbf{C}^{-\frac{1}{2}} \right) f(\mathbf{x})d\mathbf{x} \\
 &= \ln \left((2\pi)^{\frac{n}{2}}|\mathbf{C}|^{\frac{1}{2}} \right) + \frac{1}{2} \text{tr} \int \mathbf{C}^{-\frac{1}{2}}\mathbf{x}\mathbf{x}'\mathbf{C}^{-\frac{1}{2}}f(\mathbf{x})d\mathbf{x} \\
 &= \ln \left((2\pi)^{\frac{n}{2}}|\mathbf{C}|^{\frac{1}{2}} \right) + \frac{1}{2} \text{tr} \text{Cov} \left(\mathbf{C}^{-\frac{1}{2}}\mathbf{X} \right) \\
 &= \ln \left((2\pi)^{\frac{n}{2}}|\mathbf{C}|^{\frac{1}{2}} \right) + \underbrace{\frac{1}{2} \text{tr} \mathbf{I}_n}_{=\frac{n}{2}} \quad (\text{Wegen Proposition 2.4.6 b))} \\
 &= \ln \left((2\pi e)^{\frac{n}{2}}|\mathbf{C}|^{\frac{1}{2}} \right) .
 \end{aligned}$$

Damit folgt die Behauptung, da $H(\mathbf{X})$ wegen Proposition 3.4.10 unabhängig vom Erwartungswertvektor ist. □

Im Beweis wurde verwendet, dass

$$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA}) ,$$

falls die Dimension der Matrizen beide Produkte zulässt, und insbesondere

$$\mathbf{x}'\mathbf{C}^{-1}\mathbf{x} = \text{tr}(\mathbf{x}'\mathbf{C}^{-1}\mathbf{x}) = \text{tr} \left(\mathbf{C}^{-\frac{1}{2}}\mathbf{C}^{-\frac{1}{2}}\mathbf{x}\mathbf{x}' \right) = \text{tr} \left(\mathbf{C}^{-\frac{1}{2}}\mathbf{x}\mathbf{x}'\mathbf{C}^{-\frac{1}{2}} \right)$$

gilt. Dabei bezeichnet $\text{tr}(\mathbf{U})$ die *Spur (trace)* einer quadratischen Matrix \mathbf{U} , die als Summe der Diagonaleinträge von \mathbf{U} definiert ist.

Im Folgenden wird noch die bedingte Entropie bei gemischt absolut-stetig/diskreten Zufallsvariablen behandelt. Hierzu sei $(X, Y)'$ ein Zufallsvektor, wobei X eine diskrete Zufallsvariable mit Träger $T = \{x_1, \dots, x_m\}$ ist. Ferner besitze die bedingte Verteilung

von Y , gegeben $X = x_i$ eine bedingte Dichte $f_{Y|X}(y|x_i)$ für alle $i = 1, \dots, m$. Dann gilt

$$H(Y|X = x_i) = - \int_{-\infty}^{\infty} f_{Y|X}(y|x_i) \log f_{Y|X}(y|x_i) dy, \quad i = 1, \dots, m$$

$$H(Y|X) = \sum_{i=1}^m H(Y|X = x_i) P(X = x_i)$$

$$H(X|Y = y) = - \sum_{i=1}^m P(X = x_i|Y = y) \log P(X = x_i|Y = y)$$

$$H(X|Y) = \int_{-\infty}^{\infty} H(X|Y = y) f_Y(y) dy,$$

wobei $P(X = x_i|Y = y) = P(X = x_i) \frac{f_{Y|X}(y|x_i)}{f_Y(y)}$.

In Abbildung 3.5 sind die Zusammenhänge zwischen *Entropie*, bedingter Entropie und *Transinformation* grafisch dargestellt.

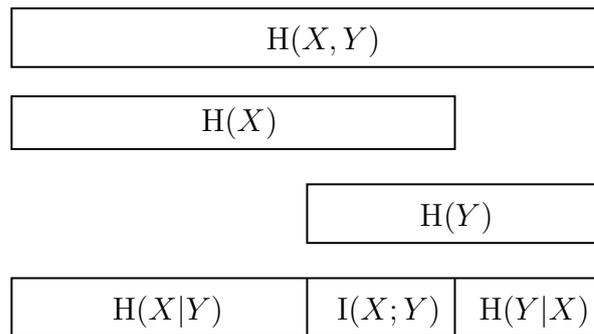


Abbildung 3.5: Zusammenhänge zwischen Entropie und Transinformation.

Beispielsweise liest man die Aussagen von Proposition 3.4.9

$$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y),$$

$$H(X, Y) \leq H(X) + H(Y)$$

und Definition 3.4.5

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X),$$

leicht an vorangegangenem Schaubild ab. Besonders intuitiv wird die Bedeutung der Transinformation als der dargestellte Überlappungsbereich klar.

4 Kanäle und ihre Kapazität

4.1 Kanäle mit diskreter Eingabe (Discrete input channel)

In das Kanalmodell werden diskrete Symbole oder Signale (oft $s_1, \dots, s_m \in \mathbb{C}$) beschrieben durch eine Zufallsvariable X mit Träger $\{x_1, \dots, x_m\}$ und Eingabeverteilung $P(X = x_i) = p_i$, $i = 1, \dots, m$ ($p_i \geq 0, \sum_{i=1}^m p_i = 1$) gegeben. Bei Eingabe von x_i besitzt die Ausgabe eine Dichte

$$f(y|x_i) = f_i(y). \quad (4.1)$$

Beispiel 4.1.1. Discrete input AWGN (additive white Gaussian noise) channel

$$Y = X + Z, \quad Z \sim N(0, \sigma^2)$$

Der Kanal wird durch Input

$$X \in \{x_1, \dots, x_m\} \sim (p_1, \dots, p_m),$$

bedingten Dichten

$$f_i(y) = f(y|x_i) = \varphi(y - x_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-x_i)^2}{2\sigma^2}}, \quad y \in \mathbb{R}$$

auf $\mathbb{C}, \mathbb{R}^n, \mathbb{C}^n$ beschrieben.

Berechnung der Transformation für das allgemeine Modell 4.1 mit $\mathbf{p} = (p_1, \dots, p_m)$.

$$\begin{aligned} I(X; Y) &= I(\mathbf{p}; (f_1, \dots, f_m)) \\ &= H(Y) - H(Y|X) \\ &= H\left(\sum_{i=1}^m p_i f_i\right) - \sum_{i=1}^m p_i H(f_i) \\ &= -\int \left(\sum_{i=1}^m p_i f_i\right) \log \left(\sum_{i=1}^m p_i f_i\right) + \sum_{i=1}^m p_i \int f_i \log f_i \\ &= \sum_{i=1}^m p_i \int f_i \log \frac{f_i}{\sum_{j=1}^m p_j f_j} \\ &= \sum_{i=1}^m p_i D\left(f_i \parallel \sum_{j=1}^m p_j f_j\right) \end{aligned}$$

Die Transformation ist somit gleich der gewichteten Summe von KL-Distanzen/Divergenzen zwischen „Fehlerdichten“.

Zur Bestimmung der Kapazität löse

$$\begin{aligned} C &= \max_{(p_1, \dots, p_m)} I(X; Y) \\ &= \max_{(p_1, \dots, p_m)} I(\mathbf{p}; (f_1, \dots, f_m)) \\ &= \max_{(p_1, \dots, p_m)} \sum_{i=1}^m p_i D \left(f_i \parallel \sum_{j=1}^m p_j f_j \right). \end{aligned}$$

Dies wird jetzt behandelt im Spezialfall des Gauß-Kanals mit binärer Eingabe.

Modell:

- Eingabe $X \in \{-1, 1\}$, $X \sim (p_0, p_1)$
- Ausgabe $Y = \mu X + Z$, $Z \sim N(0, \sigma^2)$, $\mu > 0$

Dann gilt mit Obigem:

$$f_0(y) = \varphi(y + \mu), \quad f_1(y) = \varphi(y - \mu), \quad \varphi(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}},$$

$$\begin{aligned} I(X; Y) &= H(p_0 f_0 + p_1 f_1) - p_0 H(f_0) - p_1 H(f_1) \\ &= H(p_0 \varphi(y + \mu) + p_1 \varphi(y - \mu)) - p_0 H(\varphi(y + \mu)) - p_1 H(\varphi(y - \mu)). \end{aligned}$$

Proposition 4.1.2. Unter dem oben angegebenen Kanalmodell gilt

- a) $I(X; Y) = -p_0 E [\log(p_0 + p_1 e^{-W})] - p_1 E [\log(p_1 + p_0 e^{-W})]$
- b) $C = \max_{\substack{(p_0, p_1) \\ p_0, p_1 \geq 0, p_0 + p_1 = 1}} I(X; Y) = 1 - E [\text{ld}(1 + e^{-W})]$,

wobei W eine Zufallsvariable ist mit $W \sim N\left(\frac{2\mu^2}{\sigma^2}, \frac{4\mu^2}{\sigma^2}\right)$ und ld den Logarithmus dualis zur Basis 2 beschreibt.

Beweis.

a)

$$\begin{aligned} I(X; Y) &= - \int (p_0 \varphi(y + \mu) + p_1 \varphi(y - \mu)) \log (p_0 \varphi(y + \mu) + p_1 \varphi(y - \mu)) \, dy \\ &\quad + p_0 \int \varphi(y + \mu) \log \varphi(y + \mu) \, dy + p_1 \int \varphi(y - \mu) \log \varphi(y - \mu) \, dy \\ &= -p_0 \int \varphi(y) \log \left(p_0 + p_1 e^{-\frac{2\mu(y+\mu)}{\sigma^2}} \right) \, dy - p_1 \int \varphi(y) \log \left(p_1 + p_0 e^{-\frac{2\mu(y+\mu)}{\sigma^2}} \right) \, dy \\ &= -p_0 E [\log(p_0 + p_1 e^{-W})] - p_1 E [\log(p_1 + p_0 e^{-W})] . \end{aligned}$$

- b) Zur Maximierung der Transinformation ist wegen des konstanten Terms $H(Z)$ in a) lediglich

$$H(Y) = H(p_0 f_0 + p_1 f_1), \quad 0 \leq p_0 \leq 1$$

mit $p_1 = (1 - p_0)$ zu betrachten, wobei

$$f_0(y) = \varphi(y + \mu) \text{ und } f_1(y) = \varphi(y - \mu)$$

Die differentielle Entropie ist konkav auf der Menge aller Dichten, d.h. für alle $0 \leq p_0 \leq 1$ gilt mit

$$H(p_0 f_0 + (1 - p_0) f_1) \geq p_0 H(f_0) + (1 - p_0) H(f_1) .$$

Ferner gilt wegen der Symmetrie von φ , dass $f_0(y) = f_1(-y)$, also

$$H(p_0 f_0 + p_1 f_1) = H(p_1 f_0 + p_0 f_1) .$$

Insgesamt ist die Funktion

$$g(p_0) = H(p_0 f_0 + (1 - p_0) f_1)$$

symmetrisch und konkav in der Variablen $p_0 \in [0, 1]$. Generisch wird der Verlauf der Funktion g in Abbildung 4.1 skizziert. Es folgt, dass das Maximum für $p_0 = \frac{1}{2}$

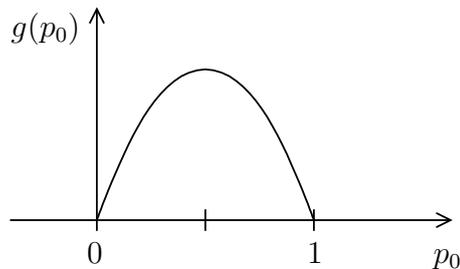


Abbildung 4.1: Skizze des generischen Verlaufs der Funktion

angenommen wird. Einsetzen in a) mit Logarithmen zur Basis 2 liefert

$$C = -2 \cdot \frac{1}{2} \text{E} \left[\log \left(\frac{1}{2} + \frac{1}{2} e^{-W} \right) \right] = 1 - \text{E} [\text{Id}(1 + e^{-W})] .$$

□

4.2 Reelle Gaußkanäle

Betrachtet wird ein physikalischer Kanal mit reeller, deterministischer zeitstetiger Eingabe $x(t)$. Die Ausgabe sei

$$Y(t) = x(t) + W(t) ,$$

wobei $W(t)$ ein zufälliger *Störprozess* ist. Die Eingabe unterliegt einer *Leistungsbeschränkung*, d.h. für die mittlere Leistung im Übertragungsintervall der Länge T gilt

$$\frac{1}{T} \int_0^T x^2(t) dt \leq L .$$

Zur Übertragung von N Symbolen, die reellen Zahlen x_1, \dots, x_N entsprechen, werden N orthonormale Funktionen ϕ_1, \dots, ϕ_N benutzt, d.h.

$$\int_0^T \phi_n(t) \phi_m(t) dt = \delta_{nm} = \begin{cases} 1 & , n = m \\ 0 & , n \neq m \end{cases} .$$

Die Eingabe $x(t)$ ist eine Linearkombination der Funktionen ϕ_n ,

$$x(t) = \sum_{n=1}^N x_n \phi_n(t) , \quad t \in [0, T] .$$

Der Empfänger berechnet nun

$$\begin{aligned} Y_n &= \int_0^T Y(t) \phi_n(t) dt \\ &= \int_0^T \left[\sum_{m=1}^N x_m \phi_m(t) + W(t) \right] \phi_n(t) dt \\ &= x_n + \int_0^T W(t) \phi_n(t) dt . \end{aligned}$$

Falls $W(t)$ Gaußsches weißes Rauschen mit konstantem Leistungsdichtespektrum $\frac{N_0}{2}$ gemäß Definition 2.7.17 ist, gilt

$$\int_0^T W(t) \phi_n(t) dt \sim N \left(0, \frac{N_0}{2} \right) .$$

Als Kanalmodell ergibt sich also

$$Y_n = x_n + Z_n , \quad n = 1, \dots, N ,$$

wobei $Z_n \sim N(0, \frac{N_0}{2})$, mit den Nebenbedingungen

$$\begin{aligned} \frac{1}{T} \int_0^T x^2(t) dt &= \frac{1}{T} \int_0^T \left(\sum_{n=1}^N x_n \phi_n(t) \right)^2 dt \\ &= \frac{1}{T} \sum_{n=1}^N x_n^2 \leq L, \end{aligned}$$

oder äquivalent

$$\overline{x^2} = \frac{1}{N} \sum_{n=1}^N x_n^2 \leq \frac{LT}{N}.$$

Hieraus abgeleitet wird das Modell des *Gaußkanals mit Leistungsbeschränkung L*:

$$Y = X + Z,$$

mit X, Z stochastisch unabhängig, $Z \sim N(0, \sigma^2)$ und $E(X^2) \leq L$. Die Kapazität des Gaußkanals mit Leistungsbeschränkung L ist definiert als

$$C = \max_{E(X^2) \leq L} I(X; Y).$$

Sie lässt sich in geschlossener Form wie folgt angeben.

Proposition 4.2.1. Die Kapazität des Gaußkanals $Y = X + Z$, X, Z stochastisch unabhängig, $Z \sim N(0, \sigma^2)$ mit Leistungsbeschränkung $E(X^2) \leq L$ beträgt

$$C = \frac{1}{2} \ln \left(1 + \frac{L}{\sigma^2} \right).$$

Beweis. Es gilt

$$\begin{aligned} I(X; Y) &\stackrel{D.3.4.5}{=} H(Y) - H(Y|X) = H(Y) - H(X + Z|X) \\ &= H(Y) - H(Z|X) \stackrel{P.3.4.8b)}{=} H(Y) - H(Z). \end{aligned}$$

Aus Beispiel 3.4.2 b) folgt

$$H(Z) = \frac{1}{2} \ln(2\pi e \sigma^2).$$

Es gilt

$$E(Y^2) = E(X + Z)^2 = E(X^2) + 2 \underbrace{(E(X)E(Z))}_{=0} + E(Z^2) \leq L + \sigma^2.$$

Mit Theorem 3.4.11 folgt

$$H(Y) \leq \frac{1}{2} \ln(2\pi e(L + \sigma^2))$$

mit Gleichheit genau dann, wenn Y normalverteilt ist, $E(X^2) = L$ und $E(X) = 0$. Insgesamt ergibt sich

$$\begin{aligned} I(X; Y) &= H(Y) - H(Z) \\ &\leq \frac{1}{2} \ln(2\pi e(L + \sigma^2)) - \frac{1}{2} \ln(2\pi e\sigma^2) \\ &= \frac{1}{2} \ln\left(1 + \frac{L}{\sigma^2}\right). \end{aligned}$$

□

Man beachte, dass $Y = X + Z$ normalverteilt ist, wenn X ebenfalls normalverteilt oder einpunktverteilt ist. Die Kapazität wird also bei normalverteilter Eingabe X mit $E(X^2) = L$ erreicht.

Im diskreten Fall ist die Kapazität eine obere Schranke für die Raten, mit denen fast fehlerfrei übertragen werden kann, siehe Theorem 3.3.6, der Shannonsche Fundamentalsatz. Eine analoge Aussage gilt auch im kontinuierlichen, leistungsbeschränkten Fall. Hierbei werden Symbole $x_1, \dots, x_N \in \mathbb{R}$ übertragen, sie werden zufällig verfälscht und empfangen als

$$Y_i = x_i + Z_i, \quad Z_i \sim N(0, \sigma^2), \quad Z_i \text{ i.i.d.}$$

Die Eingabe ist leistungsbeschränkt, d.h.

$$\frac{1}{N} \sum_{i=1}^N x_i^2 \leq L.$$

Ein Kode \mathcal{C}_N aus M Kodewörtern $\mathbf{x}(k) = (x_1(k), \dots, x_N(k)) \in \mathbb{R}^N$, $k = 1, \dots, M$, der Länge N mit

$$\frac{1}{N} \sum_{i=1}^N x_i^2(k) \leq L \quad \text{für alle } k = 1, \dots, M$$

heißt (M, N) -Kode mit Leistungsbeschränkung L .

Eine Rate $R > 0$ heißt *erreichbar* (*achievable*, *permissible*), wenn eine Folge von $(\lfloor e^{NR} \rfloor, N)$ -Kodes mit Leistungsbeschränkung L existiert, deren maximale Fehlerwahrscheinlichkeit mit $N \rightarrow \infty$ gegen Null konvergiert. Die kleinste obere Schranke aller erreichbaren Raten heißt (reale) Kapazität C_r .

Theorem 4.2.2. Für die reale Kapazität eines Gaußkanals mit Leistungsbeschränkung L gilt

$$C_r = \frac{1}{2} \ln\left(1 + \frac{L}{\sigma^2}\right),$$

also $C_r = C$.

Der Beweis dieses tiefliegenden Zusammenhangs findet sich in *Cover, Thomas, p.242ff.*

Ein intuitives Argument, dass $C_r = \frac{1}{2} \ln \left(1 + \frac{L}{\sigma^2}\right)$ die maximale Rate ist, liefert die folgende Überlegung mit Hilfe von *Kugelpackungen*. Man betrachtet zunächst ein Kodewort $\mathbf{x} = (x_1, \dots, x_N)$ der Länge N mit Leistungsbeschränkung L , d.h.

$$\frac{1}{N} \sum_{i=1}^N x_i^2 \leq L .$$

Empfangen wird

$$Y_i = x_i + Z_i , \quad i = 1, \dots, N ,$$

wobei Z_i i.i.d. und normalverteilt mit Erwartungswert 0 und Varianz σ^2 sind . Es gilt

$$\mathrm{E} (Y_i^2) = x_i^2 + 2x_i \underbrace{\mathrm{E}(Z_i)}_{=0} + \underbrace{\mathrm{E}(Z_i^2)}_{=\sigma^2} = x_i^2 + \sigma^2 ,$$

also

$$\mathrm{E} \left(\frac{1}{N} \sum_{i=1}^N Y_i^2 \right) = \frac{1}{N} \sum_{i=1}^N \mathrm{E} (Y_i^2) = \frac{1}{N} \sum_{i=1}^N (x_i^2 + \sigma^2) \leq L + \sigma^2 ,$$

d.h. im Mittel hat das empfangene Kodewort die Leistungsbeschränkung $L + \sigma^2$. Es folgt

$$\mathrm{E} \left(\sum_{i=1}^N Y_i^2 \right) \leq N(L + \sigma^2) ,$$

d.h. im Mittel liegt das empfangene Kodewort in einer N -dimensionalen Kugel B mit Radius $\sqrt{N(L + \sigma^2)}$. Ihr Volumen beträgt

$$\frac{\pi^{\frac{N}{2}}}{\Gamma \left(\frac{N}{2} + 1 \right)} (N(L + \sigma^2))^{\frac{N}{2}} .$$

Beachten Sie, dass das Volumen einer N -dimensionalen Kugel mit Radius r gegeben ist durch

$$\frac{\pi^{\frac{N}{2}}}{\Gamma \left(\frac{N}{2} + 1 \right)} r^N .$$

Es werden nun M Kodewörter $\mathbf{x}(1), \dots, \mathbf{x}(M)$ gewählt. Für die empfangenen Kodewörter gilt

$$\mathbf{Y}(k) = \mathbf{x}(k) + \mathbf{Z}(k) \sim N_N (\mathbf{x}(k), \sigma^2 \mathbf{I}_N) .$$

Um jedes Kodewort $\mathbf{x}(k)$ wird eine Kugel $B(k)$ mit Radius $\sqrt{N\sigma^2}$ gelegt, wie in Abbildung 4.2 skizziert. Jedes $B(k)$ hat das Volumen

$$\frac{\pi^{\frac{N}{2}}}{\Gamma \left(\frac{N}{2} + 1 \right)} (N\sigma^2)^{\frac{N}{2}} .$$

Die Kodewörter werden so gewählt, dass $B(1), \dots, B(M)$ möglichst disjunkt sind. $\mathbf{x}(k)$ wird dekodiert, falls $\mathbf{Y}(k) \in B(k)$ für $k = 1, \dots, M$. Dann ist die Wahrscheinlichkeit $P(\mathbf{Y}(k) \notin B(k))$ und insgesamt ist die Wahrscheinlichkeit für einen Dekodierfehler klein. Die maximale Anzahl disjunkter Kugeln in B ist

$$\frac{\frac{\pi^{\frac{N}{2}}}{\Gamma(\frac{N}{2}+1)} (N(L + \sigma^2))^{\frac{N}{2}}}{\frac{\pi^{\frac{N}{2}}}{\Gamma(\frac{N}{2}+1)} (N\sigma^2)^{\frac{N}{2}}} = \left(\frac{L + \sigma^2}{\sigma^2} \right)^{\frac{N}{2}} = e^{N \frac{1}{2} \ln(1 + \frac{L}{\sigma^2})},$$

d.h. $R = \frac{1}{2} \ln(1 + \frac{L}{\sigma^2})$ ist die maximal erreichbare Rate.

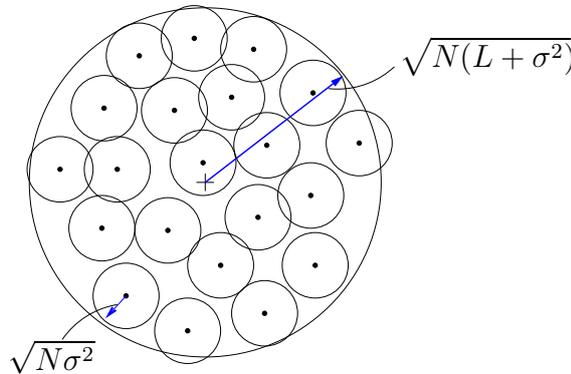


Abbildung 4.2: Kodewörter $\mathbf{x}(k)$ umgeben von Kugeln $B(k)$ mit Radius $\sqrt{N\sigma^2}$.

4.3 Parallele Gaußkanäle

Modelliert wird die parallele Übertragung auf verschiedenen parallelen Kanälen mit unterschiedlichen, auch abhängigen Störungen mit gemeinsamer Leistungsbeschränkung L .

$$\begin{array}{rcccl} X_1 & \longrightarrow & \oplus & \longrightarrow & Y_1 & \mathbf{X} = (X_1, \dots, X_n)' \\ \cdot & & \uparrow & & \cdot & \\ \cdot & & Z_1 & & \cdot & \mathbf{Z} = (Z_1, \dots, Z_n)' \\ \cdot & & & & \cdot & \\ X_n & \longrightarrow & \oplus & \longrightarrow & Y_n & \mathbf{Y} = (Y_1, \dots, Y_n)' = \mathbf{X} + \mathbf{Z} \\ & & \uparrow & & & \\ & & Z_n & & & \end{array}$$

Das genaue Modell lautet wie folgt. Empfangen wird

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z},$$

wobei \mathbf{X} und \mathbf{Z} stochastisch unabhängig sind, $\mathbf{Z} \sim N_n(\mathbf{0}, \Sigma_{\mathbf{Z}})$, $\Sigma_{\mathbf{Z}} > \mathbf{0}$ und die gemeinsame Summen-Leistungsbeschränkung

$$E \left(\sum_{i=1}^n X_i^2 \right) \leq L .$$

gilt. Ziel ist es, die Leistung so auf die verschiedenen Kanäle zu verteilen, dass die Transinformation maximal wird. Bestimmt wird also die Kapazität

$$C = \max_{\sum_{i=1}^n E(X_i^2) \leq L} I(\mathbf{X}; \mathbf{Y}) .$$

Proposition 4.3.1. Für die Kapazität C des oben beschriebenen Kanals gilt

$$C = \frac{1}{2} \sum_{i=1}^n \ln \left(1 + \frac{(\nu - \lambda_i)^+}{\lambda_i} \right) ,$$

wobei $\Sigma_{\mathbf{Z}} = \mathbf{T} \text{diag}(\lambda_1, \dots, \lambda_n) \mathbf{T}'$ eine Spektralzerlegung von $\Sigma_{\mathbf{Z}}$ ist, \mathbf{T} orthogonal und $\lambda_1, \dots, \lambda_n$ die Eigenwerte von $\Sigma_{\mathbf{Z}}$ sind. ν wird so gewählt, dass

$$\sum_{i=1}^n (\nu - \lambda_i)^+ = L .$$

Die Kapazität wird erreicht, falls $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{T} \text{diag}(\nu - \lambda_i)^+ \mathbf{T}')$.

Bei der vorangegangenen Proposition wird die Abkürzung $x^+ = \max\{x, 0\}$ verwendet. x^+ heißt *Positivteil* von $x \in \mathbb{R}$.

Eine Veranschaulichung der Lösung liefert Abbildung 4.3.

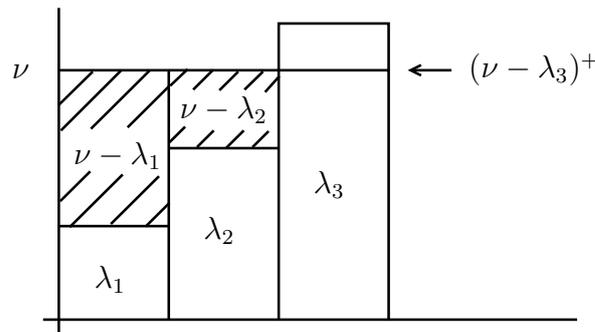


Abbildung 4.3: Veranschaulichung der Lösung, wobei $\sum_{i=1}^n (\nu - \lambda_i)^+ = L$.

Die Lösung wird gefunden durch sogenanntes "water filling", d.h. die Gesamtmenge von L Einheiten "Wasser" wird auf die Eigenwerte "gegossen", sodass sich ein möglichst gleicher Pegel einstellt. Dabei ist auch $(\nu - \lambda_i)^+ = 0$ möglich. Gelte etwa

$$(\nu - \lambda_1)^+ > 0, \dots, (\nu - \lambda_r)^+ > 0 \quad \text{und} \quad (\nu - \lambda_{r+1})^+ = \dots = (\nu - \lambda_n)^+ = 0 .$$

\mathbf{X} besitzt dann keine Dichte (bezüglich des n -dimensionalen Lebesgue-Maßes). Es gilt aber

$$\mathbf{X} \sim \mathbf{T}\mathbf{U} \ , \ \mathbf{U} = (U_1, \dots, U_r, 0, \dots, 0)' \ , \ U_1, \dots, U_r \text{ s.u.} \ , \\ U_i \sim N(0, (\nu - \lambda_i)^+) \ , \ i = 1, \dots, r \ .$$

Die Lösung lässt sich wie folgt interpretieren. Die Kanäle mit wenig Störung, ausgedrückt durch die Eigenwerte λ_i , erhalten beim Senden viel Energie $(\nu - \lambda_i)^+$, die schlechten Kanäle mit $(\nu - \lambda_i)^+ = 0$ werden gar nicht verwendet.

Beweis. Es gilt

$$I(\mathbf{X}; \mathbf{Y}) \stackrel{D.3.4.5}{=} H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}) = H(\mathbf{Y}) - H(\mathbf{X} + \mathbf{Z}|\mathbf{X}) \\ \stackrel{A.8)}{=} H(\mathbf{Y}) - H(\mathbf{Z}|\mathbf{X}) \stackrel{P.3.4.8b)}{=} H(\mathbf{Y}) - H(\mathbf{Z}) \ .$$

Da \mathbf{X} und \mathbf{Z} stochastisch unabhängig sind, gilt

$$\text{Cov}(\mathbf{Y}) = \text{Cov}(\mathbf{X} + \mathbf{Z}) = \text{Cov}(\mathbf{X}) + \text{Cov}(\mathbf{Z}) \ .$$

Also folgt mit Theorem 3.4.11

$$H(\mathbf{Y}) \leq \frac{1}{2} \ln ((2\pi e)^n |\Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Z}}|) \ , \quad (*)$$

wobei Gleichheit gilt, falls $\mathbf{X} \sim N_n(\mathbf{0}, \Sigma_{\mathbf{X}})$. Es bleibt, die obere Schranke in (*) zu maximieren, d.h.

$$\max_{\sum_{i=1}^n E(X_i^2) \leq L} |\Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Z}}| = \max_{\text{tr} \Sigma_{\mathbf{X}} \leq L} |\Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Z}}| \ .$$

Sei $\Sigma_{\mathbf{Z}} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}'$, $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$, \mathbf{T} orthogonal, eine Spektralzerlegung von $\Sigma_{\mathbf{Z}}$. Dann gilt

$$|\Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Z}}| = |\mathbf{T}| \cdot |\mathbf{T}'\Sigma_{\mathbf{X}}\mathbf{T} + \mathbf{\Lambda}| \cdot |\mathbf{T}'| \\ = |\mathbf{T}'\Sigma_{\mathbf{X}}\mathbf{T} + \mathbf{\Lambda}| \\ = |\mathbf{A} + \mathbf{\Lambda}| \ ,$$

wobei $\mathbf{A} = \mathbf{T}'\Sigma_{\mathbf{X}}\mathbf{T}$ mit

$$\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{T}'\Sigma_{\mathbf{X}}\mathbf{T}) = \text{tr}(\Sigma_{\mathbf{X}}\mathbf{T}\mathbf{T}') = \text{tr}(\Sigma_{\mathbf{X}}) \ .$$

Mit der *Hadamard-Ungleichung*, die besagt, dass $|\mathbf{B}| \leq \prod_{i=1}^n b_{ii}$, falls $\mathbf{B} \geq \mathbf{0}$, folgt

$$|\mathbf{A} + \mathbf{\Lambda}| \leq \prod_{i=1}^n (a_{ii} + \lambda_i) \ ,$$

wobei Gleichheit gilt, wenn \mathbf{A} eine Diagonalmatrix ist. Zu lösen bleibt also

$$\max \prod_{i=1}^n (a_{ii} + \lambda_i) ,$$

so dass

$$\sum_{i=1}^n a_{ii} \leq L .$$

Die Lösung lautet

$$a_{ii} = (\nu - \lambda_i)^+ ,$$

wobei ν so gewählt wird, dass $\sum_{i=1}^n (\nu - \lambda_i)^+ = L$ gilt. Also besitzt

$$\max_{\text{tr} \mathbf{\Sigma}_{\mathbf{X}} \leq L} |\mathbf{\Sigma}_{\mathbf{X}} + \mathbf{\Sigma}_{\mathbf{Z}}|$$

eine Lösung

$$\mathbf{\Sigma}_{\mathbf{X}} = \mathbf{T} \text{diag}(a_{ii}) \mathbf{T}' , \quad a_{ii} = (\nu - \lambda_i)^+ .$$

Also gilt

$$\begin{aligned} \max_{\sum_{i=1}^n \mathbb{E}(X_i^2) \leq L} \text{H}(\mathbf{Y}) &= \frac{1}{2} \ln [(2\pi e)^n |\mathbf{T} \text{diag}(a_{ii}) \mathbf{T}' + \mathbf{T} \mathbf{\Lambda} \mathbf{T}'|] \\ &= \frac{1}{2} \ln \left[(2\pi e)^n \prod_{i=1}^n (a_{ii} + \lambda_i) \right] . \end{aligned}$$

Insgesamt ergibt sich

$$\begin{aligned} C &= \max_{\sum_{i=1}^n \mathbb{E}(X_i^2) \leq L} (\text{H}(\mathbf{Y}) - \text{H}(\mathbf{Z})) \\ &= \frac{1}{2} \ln \left[(2\pi e)^n \prod_{i=1}^n (a_{ii} + \lambda_i) \right] - \frac{1}{2} \ln \left[(2\pi e)^n \prod_{i=1}^n \lambda_i \right] \\ &= \frac{1}{2} \sum_{i=1}^n \ln \left(1 + \frac{a_{ii}}{\lambda_i} \right) . \end{aligned}$$

Das Maximum wird angenommen, falls $\mathbf{X} \sim \text{N}_n(\mathbf{0}, \mathbf{T} \text{diag}(a_{ii}) \mathbf{T}')$.

□

Im Spezialfall stochastisch unabhängiger Störungen mit gleicher Kovarianz $\mathbf{\Sigma}_{\mathbf{Z}} = \sigma^2 \mathbf{I}_n$ gilt $\lambda_i = \sigma^2$, $\sum_{i=1}^n (\nu - \sigma^2) = n(\nu - \sigma^2) = L$, also $\nu = \frac{L}{n} + \sigma^2$. Die Kapazität beträgt also

$$C = \frac{n}{2} \ln \left(1 + \frac{L}{\sigma^2} \right) .$$

Wenn $L = nL'$, d.h. für jeden Kanal steht Leistung L' zur Verfügung, so ist die Kapazität das n -fache des einfachen Gaußkanals mit Leistungsbeschränkung L' .

4.4 Bandbegrenzte Gaußkanäle

Theorem 4.4.1 (*Abtasttheorem*, Sampling Theorem, Nyquist 1924). Für das Signal $s(t)$ mit Leistungsbeschränkung

$$\int_{-\infty}^{\infty} |s(t)|^2 dt < \infty$$

gelte

$$s(t) = \int_{-\infty}^{\infty} e^{ift} S(f) df \quad \text{mit } S(f) = 0, \text{ falls } |f| > 2\pi W,$$

d.h. die Spektraldichte von $s(t)$ verschwindet außerhalb von $[-2\pi W, 2\pi W]$. Dann gilt

$$s(t) = \sum_{k=-\infty}^{\infty} s\left(\frac{k}{2W}\right) \frac{\sin(2\pi Wt - k\pi)}{2\pi Wt - k\pi} = \sum_{k=-\infty}^{\infty} s\left(\frac{k}{2W}\right) \text{si}(2\pi Wt - k\pi), \quad t \in \mathbb{R},$$

d.h. $s(t)$ ist durch die Abtastwerte im Abstand $\frac{1}{2W}$ eindeutig bestimmt.

Eine Beweisskizze dieses für die digitale Übertragung grundlegenden Satzes findet sich in *Cover, Thomas, p. 248/249*.

Ist das Signal $s(t)$ bandbegrenzt durch $B = 2\pi W$ (keine Frequenzen größer W), so kann $s(t)$ aus den Abtastwerten $s\left(\frac{k}{2W}\right)$ im Abstand $\frac{1}{2W}$ eindeutig rekonstruiert werden. Eine grafische Veranschaulichung im Intervall $[-\frac{T}{2}, \frac{T}{2}]$ liefert Abbildung 4.4.

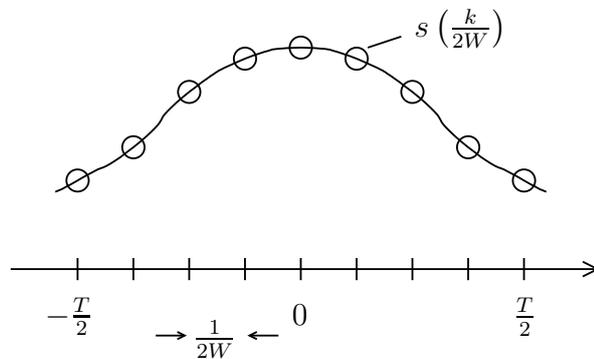


Abbildung 4.4: Abtastung des bandbegrenzten Signals $s(t)$ mit Abstand $\frac{1}{2W}$ im Intervall $[-\frac{T}{2}, \frac{T}{2}]$.

Im Folgenden wird ein stochastisches Modell für die Übertragung in bandbegrenzten Kanälen entwickelt. Es basiert auf dem Abtasttheorem. Betrachtet wird ein zeitstetiges Signal

$$Y(t) = X(t) + N(t).$$

$N(t)$ bezeichnet hierbei farbiges Gaußsches Rauschen (s. Bsp. 2.7.19) mit Leistungsdichtespektrum

$$S_{NN}(f) = \frac{N_0}{2} \mathbb{I}_{[-W,W]}(f)$$

und Autokorrelationsfunktion

$$R_{NN}(t) = \int_{-W}^W \frac{N_0}{2} e^{i2\pi ft} df = N_0 W \frac{\sin(2\pi Wt)}{2\pi Wt} = N_0 W \operatorname{si}(2\pi Wt) .$$

Betrachtet werden $n = 2W$ äquidistante Abtastwerte im Intervall der Länge 1, d.h.

$$Y_i = X_i + Z_i \quad i = 1, \dots, n .$$

Für die Autokorrelationsfunktion gilt

$$R_{NN} \left(\frac{k}{2W} \right) = \int_{-W}^W \frac{N_0}{2} e^{i2\pi f \frac{k}{2W}} df = 0 \quad , \text{ falls } k \neq 0 .$$

Also sind Z_1, \dots, Z_n stochastisch unabhängig. Ferner gilt

$$\operatorname{Var}(Z_k) = R_{NN}(0) = \frac{N_0}{2} 2W = N_0 W = \sigma^2 .$$

Insgesamt folgt, dass $\mathbf{Z} = (Z_1, \dots, Z_n)' \sim N_n(\mathbf{0}, N_0 W \mathbf{I}_n)$. Die Leistungsbeschränkung der X_i folgt wie oben aus der Leistungsbeschränkung des Signals

$$\frac{1}{n} \sum_{i=1}^n \operatorname{E}(X_i^2) \leq L .$$

Mit Proposition 4.3.1 folgt für den Spezialfall stochastisch unabhängiger Z_i

$$C = \frac{n}{2} \ln \left(1 + \frac{L}{\sigma^2} \right) .$$

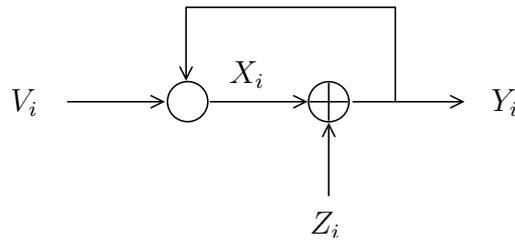
Durch Einsetzen von n und σ^2 ergibt sich

$$C = W \ln \left(1 + \frac{L}{N_0 W} \right) .$$

Dies ist eine der berühmtesten Formeln der Informationstheorie, Shannon 1948.

4.5 Gaußkanäle mit Feedback

Modelliert wird die n -fache Übertragung, bei der die Störung des i -ten Symbols von den Vorgängerstörungen abhängt.



Dabei ist V_i die Benutzereingabe zur Zeit i und

$$X_i = \sum_{j=1}^{i-1} b_{ij} Z_j + V_i$$

die Kanaleingabe zur Zeit i . Mit

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ b_{21} & 0 & 0 & \cdots & 0 \\ b_{31} & b_{32} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & b_{n3} & \cdots & 0 \end{pmatrix}$$

ergibt sich das folgende Modell. Die Kanaleingabe ist

$$\mathbf{X} = \mathbf{B}\mathbf{Z} + \mathbf{V} ,$$

wobei \mathbf{Z} und \mathbf{V} stochastisch unabhängig sind, $\mathbf{Z} \sim N_n(\mathbf{0}, \Sigma_{\mathbf{Z}})$, $\text{Cov}(\mathbf{V}) = \Sigma_{\mathbf{V}}$. Empfangen wird

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z} = (\mathbf{B} + \mathbf{I})\mathbf{Z} + \mathbf{V} ,$$

und es gilt die Leistungsbeschränkung

$$\sum_{i=1}^n \text{E}(X_i^2) \leq L .$$

Proposition 4.5.1. Für die Kapazität C des obigen Kanals gilt

$$C = \max_{\text{tr}(\mathbf{B}\Sigma_{\mathbf{Z}}\mathbf{B}' + \Sigma_{\mathbf{V}}) \leq L} \frac{1}{2} \ln \frac{|(\mathbf{B} + \mathbf{I})\Sigma_{\mathbf{Z}}(\mathbf{B} + \mathbf{I})' + \Sigma_{\mathbf{V}}|}{|\Sigma_{\mathbf{Z}}|} .$$

Beweis. Analog zum Beweis von Proposition 4.3.1 folgt

$$\begin{aligned} I(\mathbf{V}; \mathbf{Y}) &= H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{V}) \\ &= H(\mathbf{Y}) - H((\mathbf{B} + \mathbf{I})\mathbf{Z} + \mathbf{V}|\mathbf{V}) = H(\mathbf{Y}) - H((\mathbf{B} + \mathbf{I})\mathbf{Z}) . \end{aligned}$$

Wie in den Übungen gezeigt wird, gilt

$$H((\mathbf{B} + \mathbf{I})\mathbf{Z}) = H(\mathbf{Z}) + \underbrace{\log(|\mathbf{B} + \mathbf{I}|)}_{=1} = H(\mathbf{Z}) \stackrel{B.3.4.4}{=} \frac{1}{2} \ln((2\pi e)^n |\Sigma_{\mathbf{Z}}|) .$$

Mit Theorem 3.4.11 folgt

$$I(\mathbf{V}; \mathbf{Y}) \leq \frac{1}{2} \ln((2\pi e)^n |(\mathbf{B} + \mathbf{I})\Sigma_{\mathbf{Z}}(\mathbf{B} + \mathbf{I})' + \Sigma_{\mathbf{V}}|) - \frac{1}{2} \ln((2\pi e)^n |\Sigma_{\mathbf{Z}}|) \quad (*)$$

mit Gleichheit genau dann, wenn $\mathbf{Y} = (\mathbf{B} + \mathbf{I})\mathbf{Z} + \mathbf{V}$ normalverteilt ist. Dies gilt, wenn \mathbf{V} normalverteilt ist mit $E(\mathbf{V}) = \mathbf{0}$. Insgesamt folgt

$$C = \max_{\sum_{i=1}^n E(X_i^2) \leq L} I(\mathbf{V}; \mathbf{Y}) = \max_{\text{tr}(\mathbf{B}\Sigma_{\mathbf{Z}}\mathbf{B}' + \Sigma_{\mathbf{V}}) \leq L} I(\mathbf{V}; \mathbf{Y}) ,$$

da

$$\sum_{i=1}^n E(X_i^2) \stackrel{(E(X_i)=0)}{=} \text{tr}(\text{Cov}(\mathbf{X})) = \text{tr}(\text{Cov}(\mathbf{B}\mathbf{Z} + \mathbf{V})) = \text{tr}(\mathbf{B}\Sigma_{\mathbf{Z}}\mathbf{B}' + \Sigma_{\mathbf{V}}) .$$

Die Behauptung folgt durch Anwenden der oberen Schranke (*).

□

Es gilt

- a) $2(\Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Z}}) - \Sigma_{\mathbf{X}+\mathbf{Z}} = \Sigma_{\mathbf{X}-\mathbf{Z}} \geq \mathbf{0}$ (nicht-negativ definit, n.n.d.)
- b) $\mathbf{A} \geq \mathbf{B}$ (d.h. $\mathbf{A} - \mathbf{B}$ n.n.d.) $\implies |\mathbf{A}| \geq |\mathbf{B}|$

Dies wird benutzt, um die Kapazitäten zu vergleichen.

Parallele Gaußkanäle $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$ besitzen gemäß Proposition 4.3.1 die Kapazität

$$C_1 = \max_{\text{tr}(\Sigma_{\mathbf{X}}) \leq L} \frac{1}{2} \ln \frac{|\Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Z}}|}{|\Sigma_{\mathbf{Z}}|} .$$

Der Gaußkanal mit Feedback $\mathbf{Y} = \mathbf{V} + \mathbf{B}\mathbf{Z} + \mathbf{Z} = \mathbf{X} + \mathbf{Z}$ besitzt gemäß Proposition 4.5.1 die Kapazität

$$C_2 = \max_{\text{tr}(\Sigma_{\mathbf{X}}) \leq L} \frac{1}{2} \ln \frac{|\Sigma_{\mathbf{X}+\mathbf{Z}}|}{|\Sigma_{\mathbf{Z}}|} .$$

Es gilt

$$|\Sigma_{\mathbf{X}+\mathbf{Z}}| \leq |2(\Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Z}})| = 2^n |\Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Z}}| ,$$

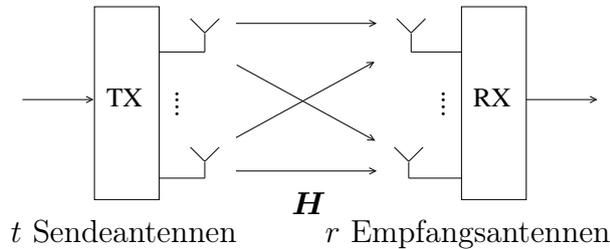
also

$$C_2 \leq \max_{\text{tr}(\Sigma_{\mathbf{X}}) \leq L} \frac{1}{2} \ln \frac{2^n |\Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Z}}|}{|\Sigma_{\mathbf{Z}}|} = C_1 + \frac{n}{2} \ln 2 .$$

Durch Feedback wächst die Kapazität bei n -stelliger Übertragung also um höchstens $\frac{n}{2} \ln 2$.

4.6 Komplexe MIMO-Kanäle

Behandelt werden Mehrantennensysteme mit r Empfangsantennen (*receiver*) und t Sendantennen (*transmitter*). *MIMO* bedeutet *multiple input, multiple output*.



Die Eigenschaften der Übertragungsstrecke von Antenne j zu Antenne i werden durch komplexe Zufallsvariablen h_{ij} mit $i = 1, \dots, r$, $j = 1, \dots, t$ beschrieben, die zu einer Matrix $\mathbf{H} = (h_{ij})_{1 \leq i \leq r, 1 \leq j \leq t}$ zusammengefasst sind.

Das genaue Modell für ein MIMO-System lautet wie folgt. Empfangen wird

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z},$$

wobei der komplexwertige $(t \times 1)$ Zufallsvektor \mathbf{X} die Eingabe ist, und die komplexwertige $(r \times t)$ deterministische oder Zufallsmatrix \mathbf{H} die *Pfadgewinne* (*path gain*) der Übertragungsstrecken h_{ij} , $\mathbf{H} = (h_{ij})$ beschreibt. Der zirkulär symmetrisch komplex normalverteilte $(r \times 1)$ Zufallsvektor \mathbf{Z} mit Erwartungswertvektor $\mathbf{0}$ und Kovarianzmatrix $\sigma^2 \mathbf{I}_r$, d.h. $\mathbf{Z} \sim \text{SCN}(\mathbf{0}, \sigma^2 \mathbf{I}_r)$, beschreibt additive Störungen und der komplexwertige $(r \times 1)$ Zufallsvektor \mathbf{Y} die Ausgabe. Insgesamt ist jede Ausgabe eine Linearkombination der Eingaben mit einem additiven Gaußschen Rauschen.

Es folgt eine kurze Wiederholung der Eigenschaften der komplexen Normalverteilung (vgl. Abschnitt 2.6).

a) $\mathbf{X} = \mathbf{U} + i\mathbf{V} \in \mathbb{C}^n$ heißt komplex normalverteilt, wenn $\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix}$ $2n$ -dimensional normalverteilt ist.

b) \mathbf{X} heißt zirkulär symmetrisch komplex normalverteilt, d.h. $\mathbf{X} \sim \text{SCN}(\boldsymbol{\mu}, \mathbf{Q})$, wenn

$$\text{Cov} \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \text{Re } \mathbf{Q} & -\text{Im } \mathbf{Q} \\ \text{Im } \mathbf{Q} & \text{Re } \mathbf{Q} \end{pmatrix}$$

für eine hermitesche, nicht-negativ definite Matrix $\mathbf{Q} \in \mathbb{C}^{n \times n}$.

c) Sei $\mathbf{X} \sim \text{SCN}(\boldsymbol{\mu}, \mathbf{Q})$, \mathbf{Q} regulär. Dann besitzt \mathbf{X} eine Dichte

$$f(\mathbf{x}) = [\det(\pi \mathbf{Q})]^{-1} \exp \{ -(\mathbf{x} - \boldsymbol{\mu})^* \mathbf{Q}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \}, \mathbf{x} \in \mathbb{C}^n.$$

$$\text{d) } \mathbf{X} \sim \text{SCN}(\boldsymbol{\mu}, \mathbf{Q}) \implies \mathbb{E} \left((\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{X} - \mathbb{E}(\mathbf{X}))^* \right) = \mathbf{Q} .$$

$$\text{e) } \mathbf{X} \sim \text{SCN}(\boldsymbol{\mu}, \mathbf{Q}), \mathbf{A} \in \mathbb{C}^{m \times n} \implies \mathbf{AX} \sim \text{SCN}(\mathbf{A}\boldsymbol{\mu}, \mathbf{AQA}^*) .$$

$$\text{f) } \mathbf{X} \sim \text{SCN}(\boldsymbol{\mu}_1, \mathbf{Q}_1), \mathbf{Y} \sim \text{SCN}(\boldsymbol{\mu}_2, \mathbf{Q}_2), \mathbf{X}, \mathbf{Y} \text{ stochastisch unabhängig} \\ \implies \mathbf{X} + \mathbf{Y} \sim \text{SCN}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \mathbf{Q}_1 + \mathbf{Q}_2) .$$

Insbesondere gilt für das obige Modell mit $\mathbf{Q} = \sigma^2 \mathbf{I}_r$, $\mathbf{Z} = \mathbf{U} + i\mathbf{V}$

$$\text{Cov} \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} = \frac{\sigma^2}{2} \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_r \end{pmatrix} ,$$

d.h. alle Komponenten sind stochastisch unabhängig mit Varianz $\frac{\sigma^2}{2}$.

Benötigt wird im Weiteren die Entropie der zirkulär symmetrischen komplexen Normalverteilung.

Proposition 4.6.1. Sei $\mathbf{X} \sim \text{SCN}(\boldsymbol{\mu}, \mathbf{Q})$, \mathbf{Q} regulär. Dann gilt

$$\text{H}(\mathbf{X}) = \log \det(\pi e \mathbf{Q}) .$$

Beweis. $\boldsymbol{\mu} = \mathbf{0}$ kann gewählt werden. Es gilt

$$\begin{aligned} \text{H}(\mathbf{X}) &= \mathbb{E}[-\log f(\mathbf{X})] \\ &= \log \det(\pi \mathbf{Q}) + \mathbb{E}(\mathbf{X}^* \mathbf{Q}^{-1} \mathbf{X}) \log e \\ &= \log \det(\pi \mathbf{Q}) + \text{tr}(\mathbb{E}(\mathbf{X} \mathbf{X}^*) \mathbf{Q}^{-1}) \log e \\ &= \log \det(\pi \mathbf{Q}) + \underbrace{\text{tr}(\mathbf{Q} \mathbf{Q}^{-1})}_{=n} \log e \\ &= \log(e^n \det(\pi \mathbf{Q})) = \log \det(\pi e \mathbf{Q}) . \end{aligned}$$

□

Man vergleiche dieses Ergebnis mit der Entropie der n -dimensionalen Normalverteilung, $\frac{1}{2} \ln((2\pi e)^n \det \mathbf{C})$, wobei \mathbf{C} die Kovarianzmatrix bezeichnet.

Proposition 4.6.2. \mathbf{X} sei ein absolut-stetiger komplexwertiger Zufallsvektor mit $E(\mathbf{X}) = \mathbf{0}$, $\text{Cov}(\mathbf{X}) = E(\mathbf{X}\mathbf{X}^*) = \mathbf{Q}$. Dann gilt

$$H(\mathbf{X}) \leq \log \det(\pi e \mathbf{Q})$$

mit Gleichheit genau dann, wenn $\mathbf{X} \sim \text{SCN}$ mit $E(\mathbf{X}\mathbf{X}^*) = \mathbf{Q}$.

Der Beweis verläuft analog zu Theorem 3.4.11.

4.6.1 Feste Kanalmatrix H

Es wird angenommen, dass die Pfadgewinne h_{ij} der Übertragungsstrecken fest und bekannt sind. Für die Eingabe gilt die Leistungsbeschränkung L . Das Modell lautet in diesem Spezialfall wie folgt. Empfangen wird

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z} ,$$

wobei \mathbf{X} mit $E(\mathbf{X}^*\mathbf{X}) = \text{tr } E(\mathbf{X}\mathbf{X}^*) \leq L$ die Eingabe beschreibt und

$\mathbf{Z} = \mathbf{U} + i\mathbf{V} \sim \text{SCN}(\mathbf{0}, \sigma^2 \mathbf{I}_r)$, d.h. $\text{Cov} \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} = \frac{\sigma^2}{2} \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_r \end{pmatrix}$, die additiven Störungen. \mathbf{X}, \mathbf{Z} seien stochastisch unabhängig und $\mathbf{H} \in \mathbb{C}^{r \times t}$ fest.

Theorem 4.6.3. Für die Kapazität des obigen Kanals gilt

$$C = \max_{E(\mathbf{X}^*\mathbf{X}) \leq L} I(\mathbf{X}; \mathbf{Y}) = \sum_{i=1}^t \left[\log \left(\frac{\nu \gamma_i}{\sigma^2} \right) \right]^+ ,$$

wobei $\mathbf{H}^*\mathbf{H} = \mathbf{U} \text{diag}(\gamma_1, \dots, \gamma_t) \mathbf{U}^*$, \mathbf{U} unitär, eine Spektralzerlegung ist, und ν so gewählt wird, dass $\sum_{i=1, \gamma_i > 0}^t \left(\nu - \frac{\sigma^2}{\gamma_i} \right)^+ = L$ (*water filling*). Die Kapazität wird angenommen, falls $\mathbf{X} \sim \text{SCN} \left(\mathbf{0}, \mathbf{U} \text{diag} \left(\nu - \frac{\sigma^2}{\gamma_i} \right)^+ \mathbf{U}^* \right)$.

Beweis. Wie im Beweis von Proposition 4.3.1 gilt

$$I(\mathbf{X}; \mathbf{Y}) = H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}) = H(\mathbf{Y}) - H(\mathbf{Z}) .$$

Mit Proposition 4.6.1 gilt

$$H(\mathbf{Z}) = \log \det (\pi e \sigma^2 \mathbf{I}_r) .$$

Es bleibt, $H(\mathbf{Y})$ zu maximieren über alle \mathbf{X} mit $E(\mathbf{X}^*\mathbf{X}) \leq L$. Da

$$E \left(E(\mathbf{X} - E(\mathbf{X}))^* (\mathbf{X} - E(\mathbf{X})) \right) = E(\mathbf{X}^*\mathbf{X}) - (E(\mathbf{X}))^* (E(\mathbf{X})) \leq E(\mathbf{X}^*\mathbf{X}) ,$$

kann man sich auf Zufallsvariablen \mathbf{X} mit $E(\mathbf{X}) = \mathbf{0}$ beschränken.

Für solche \mathbf{X} gilt

$$\mathbf{E}(\mathbf{Y}) = \mathbf{E}(\mathbf{H}\mathbf{X} + \mathbf{Z}) = \mathbf{0}$$

und mit Proposition 4.6.2

$$\mathbf{H}(\mathbf{Y}) \leq \log \det(\pi e \mathbf{R}) ,$$

wobei $\mathbf{R} = \mathbf{E}(\mathbf{Y}\mathbf{Y}^*)$. Gleichheit gilt genau dann, wenn $\mathbf{Y} \sim \text{SCN}(\mathbf{0}, \mathbf{R})$. Dies ist der Fall, wenn $\mathbf{X} \sim \text{SCN}(\mathbf{0}, \mathbf{Q})$. Dann gilt (s.o.)

$$\text{Cov}(\mathbf{Y}) = \sigma^2 \mathbf{I}_r + \mathbf{H}\mathbf{Q}\mathbf{H}^* .$$

Insgesamt folgt

$$\begin{aligned} \mathbf{I}(\mathbf{X}; \mathbf{Y}) &= \mathbf{H}(\mathbf{Y}) - \mathbf{H}(\mathbf{Z}) \\ &\leq \log \det(\pi e (\sigma^2 \mathbf{I}_r + \mathbf{H}\mathbf{Q}\mathbf{H}^*)) - \log \det(\pi e \sigma^2 \mathbf{I}_r) \\ &= \log \det\left(\mathbf{I}_r + \frac{1}{\sigma^2} \mathbf{H}\mathbf{Q}\mathbf{H}^*\right) . \end{aligned}$$

Zur Bestimmung der Kapazität bleibt zu lösen

$$\begin{aligned} &\max_{\text{tr}(\mathbf{Q}) \leq L} \det\left(\mathbf{I}_r + \frac{1}{\sigma^2} \mathbf{H}\mathbf{Q}\mathbf{H}^*\right) \\ &= \max_{\text{tr}(\mathbf{Q}) \leq L} \det\left(\mathbf{I}_t + \frac{1}{\sigma^2} \mathbf{Q}\mathbf{H}^*\mathbf{H}\right) \quad (\text{da } \det(\mathbf{I} + \mathbf{A}\mathbf{B}) = \det(\mathbf{I} + \mathbf{B}\mathbf{A})) \\ &= \max_{\text{tr}(\mathbf{Q}) \leq L} \det\left(\mathbf{I}_t + \frac{1}{\sigma^2} \mathbf{Q}\mathbf{U} \text{diag}(\gamma_i) \mathbf{U}^*\right) \quad (\text{da } \mathbf{H}^*\mathbf{H} \text{ hermitesch}) \\ &= \max_{\text{tr}(\mathbf{Q}) \leq L} \det\left(\mathbf{I}_t + \frac{1}{\sigma^2} \mathbf{\Gamma}^{\frac{1}{2}} \mathbf{U}^* \mathbf{Q} \mathbf{U} \mathbf{\Gamma}^{\frac{1}{2}}\right) \quad (\text{wobei } \mathbf{\Gamma} = \text{diag}(\gamma_i)) \\ &= \max_{\text{tr}(\mathbf{A}) \leq L} \det\left(\mathbf{I}_t + \frac{1}{\sigma^2} \mathbf{\Gamma}^{\frac{1}{2}} \mathbf{A} \mathbf{\Gamma}^{\frac{1}{2}}\right) \quad (\text{da } \text{tr}(\mathbf{U}^* \mathbf{Q} \mathbf{U}) = \text{tr}(\mathbf{Q})) \\ &\leq \max_{\text{tr}(\mathbf{A}) \leq L} \prod_{i=1, \gamma_i > 0}^t \left(1 + \frac{1}{\sigma^2} a_{ii} \gamma_i\right) \quad (\text{Hadamard-Ungleichung}) \end{aligned}$$

Mit dem *water filling*-Prinzip (s. Beweis von Proposition 4.3.1) ergibt sich als Lösung

$$a_{ii} = \begin{cases} \left(\nu - \frac{\sigma^2}{\gamma_i}\right)^+ , & \gamma_i > 0 \\ 0 , & \gamma_i = 0 \end{cases} , \quad \nu \text{ so, dass } \sum_{i=1, \gamma_i > 0}^t \left(\nu - \frac{\sigma^2}{\gamma_i}\right)^+ = L .$$

Insgesamt folgt

$$\begin{aligned}
 C &= \max_{\text{tr}(\mathbf{Q}) \leq L} \log \det \left(\mathbf{I}_r + \frac{1}{\sigma^2} \mathbf{H} \mathbf{Q} \mathbf{H}^* \right) \\
 &= \log \prod_{i=1, \gamma_i > 0}^t \left(1 + \frac{\gamma_i}{\sigma^2} \left(\nu - \frac{\sigma^2}{\gamma_i} \right)^+ \right) \\
 &= \log \prod_{i=1}^t \left(1 + \left(\frac{\gamma_i \nu}{\sigma^2} - 1 \right)^+ \right) = \sum_{i=1}^t \left(\log \frac{\gamma_i \nu}{\sigma^2} \right)^+ .
 \end{aligned}$$

Die Kapazität wird erreicht, wenn $\mathbf{A} = \mathbf{U}^* \mathbf{Q} \mathbf{U} = \text{diag}(a_{ii})$ und $\mathbf{X} \sim \text{SCN} \left(\mathbf{0}, \mathbf{U} \text{diag} \left(\nu - \frac{\sigma^2}{\gamma_i} \right)^+ \mathbf{U}^* \right)$.

□

Beispiel 4.6.4. Sei $r = t = n$, $\mathbf{H} = \mathbf{I}_n$. Dann liegt keine räumliche Interferenz zwischen den Antennen vor. Die Übertragung findet auf unabhängigen parallelen komplexen Gaußkanälen mit additivem Rauschen \mathbf{Z} statt. Mit Theorem 4.6.3 folgt

$$\mathbf{H}^* \mathbf{H} = \mathbf{I}_n = \text{diag}(1, \dots, 1), \quad \text{also } \gamma_i = 1 \text{ für alle } i = 1, \dots, n .$$

Wähle ν so, dass $\sum_{i=1}^n \left(\nu - \frac{\sigma^2}{1} \right)^+ = L$, also $\nu = \sigma^2 + \frac{L}{n}$. Dann gilt für die Kapazität

$$C = n \log \left(\frac{1}{\sigma^2} \left(\sigma^2 + \frac{L}{n} \right) \right) = n \log \left(1 + \frac{L}{n\sigma^2} \right) = n \log \left(1 + \frac{\rho}{n} \right),$$

wobei $\rho = \frac{L}{\sigma^2}$ hierbei den *SNR* (*signal-to-noise ratio*) bezeichnet. Die Kapazität wird angenommen, falls $\mathbf{X} \sim \text{SCN} \left(\mathbf{0}, \frac{L}{n} \mathbf{I}_n \right)$. Dies ist das Ergebnis von Proposition 4.3.1 mit $\boldsymbol{\Sigma}_{\mathbf{Z}} = \sigma^2 \mathbf{I}_n$, bis auf den Faktor $\frac{1}{2}$, der wegen des Übergangs von reellen zu komplexen Größen entfällt. Im Grenzwert gilt

$$C^* = \lim_{n \rightarrow \infty} n \log \left(1 + \frac{L}{n\sigma^2} \right) = \frac{L}{\sigma^2} \log e = \rho \log e .$$

Beispiel 4.6.5. Sei $\mathbf{H} = \mathbf{1}_{r \times t}$, d.h. eine $r \times t$ -Matrix mit allen Einträgen 1, die maximale räumliche Interferenz repräsentiert. Mit Theorem 4.6.3 gilt

$$\mathbf{H}^* \mathbf{H} = \mathbf{1}_{t \times r} \mathbf{1}_{r \times t} = r \mathbf{1}_{t \times t}$$

mit Eigenwerten $\gamma_1 = r \cdot t$, $\gamma_2 = \dots = \gamma_t = 0$. Der Eigenvektor zu $\gamma_1 = r \cdot t$ ist $\mathbf{u}_1 = \frac{1}{\sqrt{t}} \mathbf{1}_{t \times 1}$. Wähle ν so, dass $\left(\nu - \frac{\sigma^2}{rt} \right)^+ = L$, also $\nu = \frac{\sigma^2}{rt} + L$. Dann gilt für die Kapazität

$$C = \log \frac{1}{\sigma^2} \left(\frac{\sigma^2}{rt} + L \right) rt = \log \left(1 + rt \frac{L}{\sigma^2} \right) = \log(1 + rt\rho)$$

mit SNR $p = \frac{L}{\sigma^2}$. Die Kapazität wird angenommen, falls

$$\mathbf{X} \sim \text{SCN}\left(\mathbf{0}, \frac{1}{\sqrt{t}}\mathbf{1}_{t \times 1} \left(\frac{\sigma^2}{rt} + L - \frac{\sigma^2}{rt}\right) \frac{1}{\sqrt{t}}\mathbf{1}_{1 \times t}\right) = \text{SCN}\left(\mathbf{0}, \frac{L}{t}\mathbf{1}_{t \times t}\right).$$

4.6.2 Rayleigh-Fading

Betrachtet wird das Modell wie oben, aber \mathbf{H} ist jetzt eine Zufallsmatrix. Empfangen wird

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}, \quad (*)$$

wobei \mathbf{X} und \mathbf{Z} stochastisch unabhängig sind, $\mathbf{Z} \sim \text{SCN}(\mathbf{0}, \sigma^2\mathbf{I}_r)$ und $E(\mathbf{X}^*\mathbf{X}) \leq L$. Die Matrix

$$\mathbf{H} = \left(H_{ij}\right) \in \mathbb{C}^{r \times t}$$

besitzt Komponenten H_{ij} , die i.i.d. Zufallsvariable sind mit $\text{Re}(H_{ij}), \text{Im}(H_{ij}) \sim N(0, \frac{1}{2})$ für alle $i = 1, \dots, r$ und $j = 1, \dots, t$. Die Spalten und Zeilen von \mathbf{H} sind damit $\text{SCN}(\mathbf{0}, \mathbf{I}_r)$ – bzw. $\text{SCN}(\mathbf{0}, \mathbf{I}_t)$ –verteilt.

Wir erinnern an die Ergebnisse aus Abschnitt 2.5.1 (Rayleigh-Verteilung). Ist $G = U + iV$ eine komplexwertige Zufallsvariable mit U, V i.i.d. $\sim N(0, \tau^2)$, so gilt für die Darstellung in Polarkoordinaten

$$G = Re^{i\phi},$$

dass R und ϕ stochastisch unabhängig sind. Ferner ist $R \sim \text{Ray}(\tau^2)$ mit Dichte

$$f_R(r) = \frac{r}{\tau^2} e^{-\frac{r^2}{2\tau^2}} \mathbb{I}_{(0, \infty)}(r) \text{ und } \phi \sim R(0, 2\pi) \text{ mit Dichte } f_\phi(\varphi) = \frac{1}{2\pi} \mathbb{I}_{(0, 2\pi)}(\varphi).$$

Durch Multiplikation mit H_{ij} wird also die Phase zufällig gleichverteilt gedreht und die Amplitude zufällig nach einer $\text{Ray}(\frac{1}{2})$ –Verteilung multipliziert. Unter Modell (*) ist das empfangene Signal an jeder Antenne also eine additive Überlagerung der so verfälschten Signale zuzüglich der Störung \mathbf{Z} .

Proposition 4.6.6. Sei $\mathbf{H} = \left(H_{ij}\right) \in \mathbb{C}^{r \times t}$ mit i.i.d. Einträgen $H_{ij} = U_{ij} + iV_{ij}$ und sind U_{ij}, V_{ij} i.i.d. $\sim N(0, \tau^2)$. $\mathbf{U} \in \mathbb{C}^{r \times r}$, $\mathbf{V} \in \mathbb{C}^{t \times t}$ seien unitäre Matrizen. Dann besitzt $\mathbf{U}\mathbf{H}\mathbf{V}^*$ dieselbe Verteilung wie \mathbf{H} .

Beweis. Es genügt zu zeigen, dass $\mathbf{G} = \mathbf{U}\mathbf{H} \sim \mathbf{H}$. Bezeichne dazu $\mathbf{H}_1, \dots, \mathbf{H}_t$ die Spalten von \mathbf{H} . $\mathbf{H}_1, \dots, \mathbf{H}_t$ sind stochastisch unabhängig, $\mathbf{H}_i \sim \text{SCN}(\mathbf{0}, 2\tau^2\mathbf{I}_r)$. \mathbf{G} hat stochastisch unabhängige Spalten $\mathbf{G}_1, \dots, \mathbf{G}_t$ mit

$$\mathbf{G}_i = \mathbf{U}\mathbf{H}_i \stackrel{(P.2.6.5a)}{\sim} \text{SCN}(\mathbf{0}, 2\tau^2\mathbf{U}\mathbf{I}_r\mathbf{U}^*) = \text{SCN}(\mathbf{0}, 2\sigma^2\mathbf{I}_r).$$

□

Im Folgenden wird angenommen, dass der Kanalzustand beim Empfänger bekannt ist (The receiver has perfect *CSI = channel state information*), d.h. der Empfänger kennt die Realisation von \mathbf{H} . Beim Sender ist lediglich die Verteilung von \mathbf{H} bekannt.

Theorem 4.6.7. Unter den Annahmen dieses Abschnitts gilt für die Kapazität des Kanals $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}$ mit $E(\mathbf{X}^*\mathbf{X}) \leq L$

$$C = \max_{E(\mathbf{X}^*\mathbf{X}) \leq L} I(\mathbf{X}; (\mathbf{Y}, \mathbf{H})) = E \left[\log \det \left(\mathbf{I}_r + \frac{L}{t\sigma^2} \mathbf{H}\mathbf{H}^* \right) \right].$$

Die Kapazität wird angenommen, falls $\mathbf{X} \sim \text{SCN}(\mathbf{0}, \frac{L}{t}\mathbf{I}_t)$.

Beweis. Da beim Empfänger perfekte CSI vorliegt, gilt

$$I(\mathbf{X}; (\mathbf{Y}, \mathbf{H})) = \underbrace{I(\mathbf{X}; \mathbf{H})}_{=0, \text{ da } \mathbf{X}, \mathbf{H} \text{ s.u.}} + I(\mathbf{X}; \mathbf{Y}|\mathbf{H}) = I(\mathbf{X}; \mathbf{Y}|\mathbf{H}) = E_{\mathbf{H}} \left(I(\mathbf{X}; \mathbf{Y}|\mathbf{H} = \tilde{\mathbf{H}}) \right).$$

Wie im Beweis von Theorem 4.6.3 gilt

$$I(\mathbf{X}; \mathbf{Y}|\mathbf{H} = \tilde{\mathbf{H}}) \leq \log \det \left(\mathbf{I}_r + \frac{1}{\sigma^2} \tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^* \right) \text{ mit } \mathbf{Q} = E(\mathbf{X}\mathbf{X}^*)$$

mit Gleichheit, falls $\mathbf{X} \sim \text{SCN}(\mathbf{0}, \mathbf{Q})$. Zu lösen bleibt also

$$\max_{\text{tr}(\mathbf{Q}) \leq L, \mathbf{Q} \text{ n.n.d.}} E_{\mathbf{H}} \left[\log \det \left(\mathbf{I}_r + \frac{1}{\sigma^2} \mathbf{H}\mathbf{Q}\mathbf{H}^* \right) \right].$$

Setze $\mathbf{Q} = \mathbf{U}\mathbf{D}\mathbf{U}^*$, \mathbf{U} unitär, \mathbf{D} Diagonalmatrix. Da $\text{tr}(\mathbf{Q}) = \text{tr}(\mathbf{D})$ und $\mathbf{H}\mathbf{U} \sim \mathbf{H}$ wegen P. 4.6.6, bleibt zu lösen

$$\max_{\text{tr}(\mathbf{Q}) \leq L, \mathbf{Q} \text{ n.n.d., diagonal}} E_{\mathbf{H}} \left[\log \det \left(\mathbf{I}_r + \frac{1}{\sigma^2} \mathbf{H}\mathbf{Q}\mathbf{H}^* \right) \right].$$

$\mathbf{\Pi}$ sei eine *Permutationsmatrix*, d.h. in jeder Zeile und Spalte ist genau ein Eintrag 1 und die restlichen Einträge sind 0. $\mathbf{\Pi}$ vertauscht bei Rechtsmultiplikation die Spalten und bei Linksmultiplikation die Zeilen. Es gilt $\mathbf{H}\mathbf{\Pi} \sim \mathbf{H}$, d.h. $\mathbf{H}\mathbf{\Pi}$ und \mathbf{H} besitzen dieselbe Verteilung. Es folgt

$$\begin{aligned} E \left[\log \det \left(\mathbf{I}_r + \frac{1}{\sigma^2} \mathbf{H}\mathbf{Q}\mathbf{H}^* \right) \right] &= E \left[\log \det \left(\mathbf{I}_r + \frac{1}{\sigma^2} \mathbf{H} \underbrace{\mathbf{\Pi}\mathbf{Q}\mathbf{\Pi}'}_{=\mathbf{Q}^{\mathbf{\Pi}}} \mathbf{H}^* \right) \right] \\ &= \frac{1}{t!} \sum_{\mathbf{\Pi}} E \left[\log \det \left(\mathbf{I}_r + \frac{1}{\sigma^2} \mathbf{H}\mathbf{Q}^{\mathbf{\Pi}}\mathbf{H}^* \right) \right] \\ &= E \left[\frac{1}{t!} \sum_{\mathbf{\Pi}} \log \det \left(\mathbf{I}_r + \frac{1}{\sigma^2} \mathbf{H}\mathbf{Q}^{\mathbf{\Pi}}\mathbf{H}^* \right) \right] \\ &\leq E \left[\log \det \left(\mathbf{I}_r + \frac{1}{\sigma^2} \mathbf{H} \left(\frac{1}{t!} \sum_{\mathbf{\Pi}} \mathbf{Q}^{\mathbf{\Pi}} \right) \mathbf{H}^* \right) \right]. \end{aligned}$$

Die letzte Ungleichung gilt, da $\phi(\mathbf{Q}) = \log \det(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^*)$ konkav ist auf der Menge der nicht-negativ definiten Matrizen. Da \mathbf{Q} diagonal ist, gilt $\frac{1}{t} \sum_{\Pi} \mathbf{Q}^{\Pi} = \alpha \mathbf{I}_t$, ein Vielfaches der Einheitsmatrix. Insgesamt folgt

$$\begin{aligned} & \max_{\substack{\text{tr}(\mathbf{Q}) \leq L, \\ \mathbf{Q} \text{ n.n.d., diagonal}}} \mathbb{E}_{\mathbf{H}} \left[\log \det \left(\mathbf{I}_r + \frac{1}{\sigma^2} \mathbf{H}\mathbf{Q}\mathbf{H}^* \right) \right] \\ & \leq \max_{t\alpha \leq L} \mathbb{E}_{\mathbf{H}} \left[\log \det \left(\mathbf{I}_r + \frac{\alpha}{\sigma^2} \mathbf{H}\mathbf{I}_t\mathbf{H}^* \right) \right] \\ & = \mathbb{E}_{\mathbf{H}} \left[\log \det \left(\mathbf{I}_r + \frac{L}{t\sigma^2} \mathbf{H}\mathbf{H}^* \right) \right]. \end{aligned}$$

Das Maximum wird angenommen, falls $\mathbf{Q} = \frac{L}{t} \mathbf{I}_t$ und $\mathbf{X} \sim \text{SCN}(\mathbf{0}, \frac{L}{t} \mathbf{I}_t)$.

□

Mit dem starken Gesetz großer Zahlen gilt im Grenzwert $t \rightarrow \infty$, dass

$$\frac{1}{t} \mathbf{H}\mathbf{H}^* \xrightarrow{(t \rightarrow \infty)} \mathbf{I}_r \quad \text{P-f.s. .}$$

Denn

$$\begin{aligned} i \neq j : \quad & \frac{1}{t} \sum_{l=1}^t H_{il} H_{jl}^* \longrightarrow \mathbb{E}(H_{i1} H_{j1}^*) \stackrel{s.u.}{=} \mathbb{E}(H_{i1}) \mathbb{E}(H_{j1}) = 0 \quad \text{P-f.s. ,} \\ i = j : \quad & \frac{1}{t} \sum_{l=1}^t H_{il} H_{il}^* \longrightarrow \mathbb{E}(|H_{i1}|^2) = \mathbb{E}(U_{i1}^2 + V_{i1}^2) = \frac{1}{2} + \frac{1}{2} = 1 \quad \text{P-f.s. .} \end{aligned}$$

Mit Theorem 4.6.7 folgt im Grenzwert

$$\begin{aligned} C^* &= \lim_{t \rightarrow \infty} \mathbb{E} \left[\log \det \left(\mathbf{I}_r + \frac{L}{t\sigma^2} \mathbf{H}\mathbf{H}^* \right) \right] \\ &= \log \det \left(\left(1 + \frac{L}{\sigma^2} \right) \mathbf{I}_r \right) = \log \left(1 + \frac{L}{\sigma^2} \right)^r = r \log \left(1 + \frac{L}{\sigma^2} \right). \end{aligned}$$

$\log \left(1 + \frac{L}{\sigma^2} \right)$ ist die Kapazität des einfachen komplexen Gaußkanals. Für große t wächst die Kapazität des MIMO-Kanals linear mit der Anzahl der Empfangsantennen.

4.6.3 Obere Schranken für die Kapazität C

Bereits im Beweis von Theorem 4.6.7 wird benutzt, dass $\phi(\mathbf{A}) = \log \det(\mathbf{A})$ eine konkave Funktion auf der Menge der nicht-negativ definiten Matrizen ist. Benutzt wird ferner die *Jensen-Ungleichung* (s. z.B. *Cover, Thomas: Information Theory*), die besagt, dass für

eine konvexe (konkave) Funktion ϕ und integrierbare Zufallsvariablen (Zufallsvektoren, Zufallsmatrizen) \mathbf{X} gilt

$$\mathbb{E}(\phi(\mathbf{X})) \geq (\leq) \phi(\mathbb{E}(\mathbf{X})) .$$

Mit diesen Hilfsmitteln folgt

$$\begin{aligned} C &= \mathbb{E} \left[\log \det \left(\mathbf{I}_r + \frac{L}{t\sigma^2} \mathbf{H}\mathbf{H}^* \right) \right] \leq \log \det \left(\mathbf{I}_r + \frac{L}{t\sigma^2} \underbrace{\mathbb{E}(\mathbf{H}\mathbf{H}^*)}_{=t\mathbf{I}_r} \right) \\ &= \log \det \left(\left(1 + \frac{L}{\sigma^2} \right) \mathbf{I}_r \right) = r \log \left(1 + \frac{L}{\sigma^2} \right) . \end{aligned} \quad (*)$$

Die Schranke wird angenommen, falls $\mathbf{H}\mathbf{H}^* = t\mathbf{I}_r$.

Mit der Identität $\det(\mathbf{I} + \mathbf{A}\mathbf{B}) = \det(\mathbf{I} + \mathbf{B}\mathbf{A})$ (siehe Beweis von Theorem 4.6.3) folgt analog

$$\begin{aligned} C &= \mathbb{E} \left[\log \det \left(\mathbf{I}_t + \frac{L}{t\sigma^2} \mathbf{H}^* \mathbf{H} \right) \right] \leq \log \det \left(\mathbf{I}_t + \frac{L}{t\sigma^2} \underbrace{\mathbb{E}(\mathbf{H}^* \mathbf{H})}_{=r\mathbf{I}_t} \right) \\ &= \log \det \left(\left(1 + \frac{rL}{t\sigma^2} \right) \mathbf{I}_t \right) = t \log \left(1 + \frac{rL}{t\sigma^2} \right) . \end{aligned} \quad (**)$$

Die Schranke wird angenommen, falls $\mathbf{H}^* \mathbf{H} = r\mathbf{I}_t$. Insgesamt folgt aus (*) und (**)

$$C \leq \min \left\{ r \log \left(1 + \frac{L}{\sigma^2} \right), t \log \left(1 + \frac{rL}{t\sigma^2} \right) \right\} .$$

5 Optimierung und Algorithmen für schwere Probleme

In der Informationstechnik treten häufig Optimierungsprobleme auf, beispielsweise optimales Routen in Netzen, kostenminimale Netzauslegung, optimale Kanalzuweisung in GSM, interferenzminimale Konfiguration von Basisstationen, Kodierungs- und Dekodierungsalgorithmen (z.B. kürzeste Wege, Viterbi-Algorithmus). Das Ziel dieses Kapitels ist die Bereitstellung von Methoden zu ihrer Lösung.

Schwere Probleme sind dabei typisch, schwer im Sinne der Komplexitätstheorie, die hierfür den Begriff *NP-schwer* bereitstellt. Der Aufwand zu ihrer Lösung steigt exponentiell mit der Problemgröße, sodass realistische Probleminstanzen nicht in akzeptabler Zeit gelöst werden können. Zu ihrer approximativen Lösung werden *Heuristiken* gebraucht. Die Lösung ist dabei nicht notwendig optimal, sie sollte aber nahe am Optimum liegen. Entsprechende Verfahren sind für den praktischen Einsatz von großem Interesse und werden ebenfalls in diesem Kapitel behandelt.

5.1 Lineare Programmierung

Sei $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$. Dann heißt

$$\begin{aligned} \text{(LP)} \quad & \max \mathbf{c}'\mathbf{x} \\ & \text{s.d. } \mathbf{Ax} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

lineares Programm in kanonischer Form. (LP) wird durch \mathbf{c} , \mathbf{A} , \mathbf{b} konkret bestimmt, d.h. das Tripel $(\mathbf{c}, \mathbf{A}, \mathbf{b})$ repräsentiert eine *Instanz* des linearen Programms. Die Ungleichungen \leq und \geq zwischen den Vektoren sind jeweils komponentenweise zu verstehen.

Eine grafische Repräsentation der Instanz im \mathbb{R}^2 mit

$$\mathbf{c} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \mathbf{A} = \begin{pmatrix} -3 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 3 \\ 6 \\ 4 \end{pmatrix},$$

also

$$\begin{aligned} \max \quad & x_1 + 2x_2 \\ \text{s.d.} \quad & -3x_1 + x_2 \leq 3 \\ & x_1 + x_2 \leq 6 \\ & x_1 \leq 4 \end{aligned}$$

ist in Abbildung 5.1 dargestellt.

Die Gerade $x_1 + 2x_2 = \alpha$ wird hierbei mit wachsendem α nach rechts oben verschoben, bis sie gerade noch den Eckpunkt $(\frac{3}{4}, \frac{21}{4})$ berührt. α hat dann den Wert $\alpha = \frac{45}{4}$, den Zielfunktionswert der optimalen Lösung.

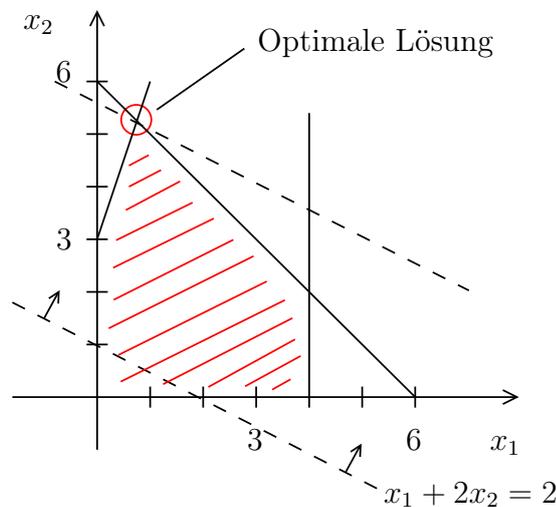


Abbildung 5.1: Grafische Darstellung des obigen linearen Programms mit optimaler Lösung $x_1 = \frac{3}{4}$, $x_2 = \frac{21}{4}$ und zugehörigem Zielfunktionswert $\frac{45}{4}$.

In Abbildung 5.1 erkennt man, dass die Menge der zulässigen Punkte ein *Polytop* bildet, das sich als Durchschnitt von *Halbräumen* ergibt. Jedes \mathbf{x} mit $\mathbf{Ax} \leq \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$, heißt *zulässige Lösung* (*feasible solution*).

Oftmals kommen zu (LP) Ganzzahligkeitsbedingungen hinzu. Die Bezeichnung lautet dann *ILP* (*integer linear program*)

$$\begin{aligned} \text{(ILP)} \quad & \max \mathbf{c}'\mathbf{x} \\ \text{s.d.} \quad & \mathbf{Ax} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}, \mathbf{x} \in \mathcal{S} \subseteq \mathbb{N}_0^n \end{aligned}$$

Beispiel 5.1.1. (Minimum Weighted Vertex Cover)

Gegeben sei ein Graph $G = (V, E)$ mit Knoten $V = \{1, \dots, n\}$ und Kanten $E \subseteq V^{(2)}$. $V^{(2)}$ bedeutet dabei die Menge der zweielementigen Teilmengen. Jedem Knoten seien Kosten c_i zugeordnet. Das Problem lautet nun, eine Auswahl $S \subset V$ von Knoten derart zu

bestimmen, dass jede Kante mit mindestens einem Knoten inzidiert (*Vertex Cover*) und die Kosten minimal sind. Eine Veranschaulichung hierzu liefert Abbildung 5.2, die Kosten sind als reelle Zahlen neben den Knoten notiert. Die Formulierung des Optimierungspro-

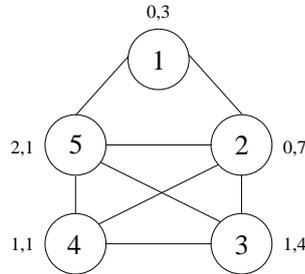


Abbildung 5.2: Zu dem gegebenen Graph $G = (V, E)$ mit $V = \{1, 2, 3, 4, 5\}$, $E = \{\{1, 2\}, \{1, 5\}, \{2, 3\}, \{2, 5\}, \dots, \{4, 5\}\}$ existieren beispielsweise die Vertex Cover $\{1, 2, 3, 4\}$ oder $\{2, 3, 5\}$.

blems als ILP lautet wie folgt. Eingeführt werden die Variablen $\mathbf{x} = (x_1, \dots, x_n)' \in \{0, 1\}^n$ mit der Bedeutung

$$x_i = \begin{cases} 1 & , \text{ Knoten } i \text{ wird in } S \text{ aufgenommen} \\ 0 & , \text{ sonst} \end{cases} .$$

Mit der Notation $\mathbf{c} = (c_1, \dots, c_n)'$ lässt sich das Problem schreiben als

$$\begin{aligned} \min & \quad \mathbf{c}'\mathbf{x} \\ \text{s.d.} & \quad x_i + x_j \geq 1 \quad \forall \{i, j\} \in E, i < j \\ & \quad x_i \in \{0, 1\} \quad \forall i = 1, \dots, n \end{aligned}$$

oder äquivalent als Maximierungsproblem

$$\begin{aligned} \max & \quad -\mathbf{c}'\mathbf{x} \\ \text{s.d.} & \quad -x_i - x_j \leq -1 \quad \forall \{i, j\} \in E, i < j \\ & \quad x_i \in \{0, 1\} \quad \forall i = 1, \dots, n \end{aligned} .$$

Beispiel 5.1.2. (Maximal Covering Location)

Ein Problem aus der Funknetzplanung lautet wie folgt. Aus einer Menge von möglichen Konfigurationen von Basisstationen (Standort, Leistung, Antennenhöhe, Antennencharakteristik, etc.) soll eine beschränkte Anzahl so ausgesucht werden, dass möglichst viel Verkehr überdeckt wird. Eine Veranschaulichung hierzu liefert Abbildung 5.3. Ausgangsdaten sind die Verkehrspunkte $\mathcal{T} = \{1, \dots, T\}$, die Kandidaten für Basisstationen $\mathcal{S} = \{1, \dots, S\}$ und $r_{st} \in \{0, 1\}$ mit der Bedeutung

$$r_{st} = \begin{cases} 1 & , \text{ die Feldstärke von Basisstation } s \text{ ist bei Verkehrspunkt } t \\ & \text{genügend groß (} t \text{ ist von } s \text{ überdeckt)} \\ 0 & , \text{ sonst} \end{cases} .$$

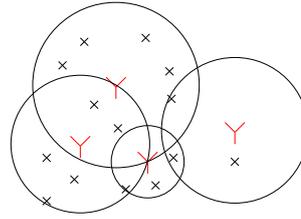


Abbildung 5.3: Skizze der Überdeckung von Verkehrspunkten durch Basisstationen.

Das Problem lautet nun, höchstens s_{max} Kandidaten so auszuwählen, dass möglichst viele Verkehrspunkte überdeckt sind. Die Formulierung als ILP lautet wie folgt. Die Variablen $\mathbf{x} = (x_1, \dots, x_S)' \in \{0, 1\}^S$ haben die Bedeutung

$$x_s = \begin{cases} 1 & , \text{ Kandidat } s \text{ wird ausgewählt} \\ 0 & , \text{ sonst} \end{cases} .$$

Hiermit ergibt sich die Formulierung als Optimierungsproblem wie folgt.

$$\max \sum_{s \in \mathcal{S}, t \in \mathcal{T}} z_{st}$$

$$\text{s.d. } z_{st} \leq r_{st} x_s \quad \forall s \in \mathcal{S}, t \in \mathcal{T} \quad (1)$$

$$\sum_{s \in \mathcal{S}} z_{st} \leq 1 \quad \forall t \in \mathcal{T} \quad (2)$$

$$\sum_{s \in \mathcal{S}} x_s \leq s_{max} \quad (3)$$

$$z_{st}, x_s \in \{0, 1\} \quad \forall s \in \mathcal{S}, t \in \mathcal{T} .$$

z_{st} sind Hilfsvariable. Wegen (1) gilt $z_{st} = 1$ höchstens dann, wenn Basisstation s ausgewählt wird und ihre Feldstärke bei t ausreichend hoch ist. (2) besagt, dass die Überdeckung jedes Verkehrsknotens höchstens einmal gezählt wird. (3) beschränkt die Anzahl der ausgewählten Kandidaten.

Für rechentechnische Zwecke ist es günstiger, lineare Programme in *Standardform* zu betrachten, in der bei den Nebenbedingungen nur Gleichheit auftritt. Es heißt

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.d.} \quad & \mathbf{Ax} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

lineares Programm in Standardform.

Jedes lineare Programm in kanonischer Form lässt sich durch Einführen zusätzlicher *Schlupfvariablen* (*slack variables*) in Standardform überführen. Betrachte

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.d.} \quad & \mathbf{Ax} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned} \quad (1)$$

und mit zusätzlichen Variablen $\mathbf{s} = (s_1, \dots, s_m)' \geq \mathbf{0}$

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.d.} \quad & \begin{pmatrix} \mathbf{A} & \mathbf{I}_m \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{s} \end{pmatrix} = \mathbf{b} \\ & \mathbf{x}, \mathbf{s} \geq \mathbf{0} \end{aligned} \quad (2)$$

Es gelten die folgenden Zusammenhänge. Für jede zulässige Lösung \mathbf{x} von (1) existiert $\mathbf{s} \geq \mathbf{0}$, nämlich $\mathbf{s} = \mathbf{b} - \mathbf{A}\mathbf{x}$, sodass (\mathbf{x}, \mathbf{s}) zulässig ist für (2). Ist umgekehrt (\mathbf{x}, \mathbf{s}) zulässig für (2), so ist \mathbf{x} zulässig für (1). Beide lineare Programme besitzen also dieselben optimalen Werte und Lösungen in \mathbf{x} .

Es folgen Strukturuntersuchungen zum linearen Programm in kanonischer Form

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.d.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

Zunächst sei an folgenden Zusammenhang erinnert. Eine Menge $M \subseteq \mathbb{R}^n$ heißt konvex, wenn $\alpha\mathbf{x} + (1 - \alpha)\mathbf{y} \in M$ für alle $\mathbf{x}, \mathbf{y} \in M$ und alle $0 \leq \alpha \leq 1$ gilt. Eine konvexe Menge enthält also mit je zwei Punkten aus der Menge auch deren Verbindungsgerade. Dieser Zusammenhang ist beispielhaft in Abbildung 5.4 skizziert.

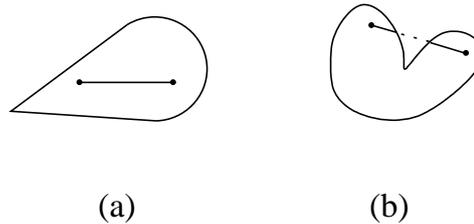


Abbildung 5.4: Die Menge (a) ist konvex, die Menge (b) ist nicht konvex.

Proposition 5.1.3. M_1, \dots, M_n seien konvex. Dann ist $M = \bigcap_{i=1}^n M_i$ ebenfalls konvex.

Beweis. Seien $\mathbf{x}, \mathbf{y} \in \bigcap_{i=1}^n M_i$, $\alpha \in [0, 1]$. Dann gilt $\mathbf{x}, \mathbf{y} \in M_i$ für alle $i = 1, \dots, n$, und wegen der Konvexität der M_i gilt $\alpha\mathbf{x} + (1 - \alpha)\mathbf{y} \in M_i$ für alle $i = 1, \dots, n$. Folglich gilt $\alpha\mathbf{x} + (1 - \alpha)\mathbf{y} \in \bigcap_{i=1}^n M_i$. □

Folgende Sprechweisen und Definitionen sind nützlich zum intuitiven Verständnis.

Definition 5.1.4. Sei $\mathbf{a} \in \mathbb{R}^n$, $b \in \mathbb{R}$ fest. Die Menge

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}'\mathbf{x} = b\}$$

heißt (*affine*) *Hyperebene* und die Menge

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}'\mathbf{x} \leq b\}$$

heißt (*affiner*) *Halbraum*.

Halbräume und auch Hyperebenen sind konvex. Denn für $\mathbf{x}, \mathbf{y} \in H$ und $\alpha \in [0, 1]$ folgt

$$\mathbf{a}'(\alpha \mathbf{x} + (1 - \alpha)\mathbf{y}) = \alpha \underbrace{\mathbf{a}'\mathbf{x}}_{\leq b} + (1 - \alpha) \underbrace{\mathbf{a}'\mathbf{y}}_{\leq b} \leq \alpha b + (1 - \alpha)b = b .$$

Weiterhin ist

$$\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$$

offensichtlich der Durchschnitt von Halbräumen. Sei hierzu

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}'_1 \\ \vdots \\ \mathbf{a}'_m \end{pmatrix}, \quad \mathbf{b} = (b_1, \dots, b_m)' .$$

Dann lässt sich \mathcal{C} schreiben als

$$\mathcal{C} = \{\mathbf{x} \mid \mathbf{a}'_1 \mathbf{x} \leq b_1\} \cap \dots \cap \{\mathbf{x} \mid \mathbf{a}'_m \mathbf{x} \leq b_m\} \\ \cap \{\mathbf{x} \mid x_1 \geq 0\} \cap \dots \cap \{\mathbf{x} \mid x_n \geq 0\} .$$

$\mathcal{C} = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ ist also ein abgeschlossenes konvexes *Polytop*, d.h. der Durchschnitt von Halbräumen. \mathcal{C} enthält auch keine *Gerade*, d.h. keine Menge der Form

$$\{\lambda \mathbf{a} \mid \lambda \in \mathbb{R}\} \quad \text{für ein } \mathbf{a} \neq \mathbf{0} ,$$

da $\mathbf{x} \geq \mathbf{0}$ für alle $\mathbf{x} \in \mathcal{C}$.

Als weiteres Hilfsmittel wird der Begriff *Extremalpunkt* benötigt.

Definition 5.1.5. M sei eine konvexe Menge. $\mathbf{x}_0 \in M$ heißt *Extremalpunkt*, wenn es kein $\mathbf{x}_1 \neq \mathbf{x}_2 \in M$ und kein $\alpha \in (0, 1)$ gibt mit $\mathbf{x}_0 = \alpha \mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2$.

Extremalpunkte \mathbf{x}_0 können nicht als (echte) Konvexkombination von $\mathbf{x}_1 \neq \mathbf{x}_2$ dargestellt werden. Halbräume $\{\mathbf{x} \mid \mathbf{a}'\mathbf{x} \leq b\}$ haben keine Extremalpunkte. Eine grafische Veranschaulichung liefert Abbildung 5.5. Obige Hilfsmittel werden zum Lösen linearer

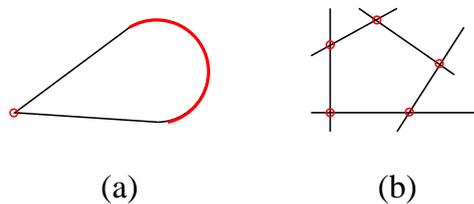


Abbildung 5.5: Extremalpunkte der Mengen (a) und (b) als Punkte bzw. als verdickte Linie dargestellt.

Programme wie folgt verwendet. Der entscheidende Zusammenhang zum Entwurf eines Lösungsalgorithmus für das lineare Programm ist der Folgende.

Theorem 5.1.6. Falls $\mathcal{C} = \{\mathbf{x} \mid \mathbf{Ax} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\} \neq \emptyset$ und die Zielfunktion $\mathbf{c}'\mathbf{x}$ auf \mathcal{C} nach oben beschränkt ist, so gibt es optimale Lösungen von (LP), wobei mindestens eine hiervon *Extremalpunkt* von \mathcal{C} ist.

Der Beweis zu diesem Theorem findet sich beispielsweise in *Roberts & Varberg, p. 141*. Eine grafische Veranschaulichung bietet Abbildung 5.6. Als Konsequenz ergibt sich, dass

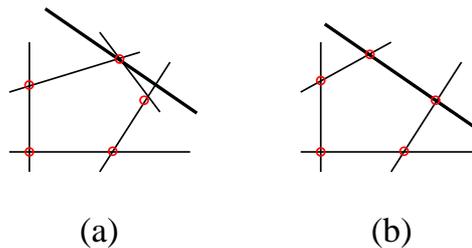


Abbildung 5.6: Schnitt der Geraden mit Polytop (a) liefert eine eindeutige Lösung, die Lösungsmenge in (b) enthält mindestens einen Extremalpunkt.

durch Absuchen der Extremalpunkte mindestens eine optimale Lösung gefunden werden kann. Es gibt höchstens $\binom{m+n}{m}$ Extremalpunkte von \mathcal{C} , Enumeration ist für große Instanzen also hoffnungslos. Für je zwei verschiedene optimale Lösungen $\mathbf{x}_1 \neq \mathbf{x}_2 \in \mathcal{C}$ mit $\mathbf{c}'\mathbf{x}_1 = \mathbf{c}'\mathbf{x}_2 = Z_{opt}$ besteht die gesamte Verbindungsgerade $\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$, $\alpha \in [0, 1]$ aus Optimalstellen. Dies gilt, da für alle $\alpha \in [0, 1]$

$$\mathbf{c}'(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) = \alpha \underbrace{\mathbf{c}'\mathbf{x}_1}_{=Z_{opt}} + (1-\alpha) \underbrace{\mathbf{c}'\mathbf{x}_2}_{=Z_{opt}} = \alpha Z_{opt} + (1-\alpha)Z_{opt} = Z_{opt} .$$

Dantzig hat 1951 einen Algorithmus zur schrittweisen Verbesserung der Zielfunktion auf Extremalpunkten vorgeschlagen, den *Simplex-Algorithmus*. Dieser bewährt sich bei praktischen Problemen hervorragend. Verwendet wird die Darstellung in Standardform $\max \mathbf{c}'\mathbf{x}$, s.d. $\mathbf{Ax} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$. Es gibt Instanzen von (LP), bei denen der Simplex-Algorithmus in m exponentiell viele Extremalpunkte absucht, bis er das Optimum findet. Im *worst case* erfordert der Simplex-Algorithmus also exponentiellen Aufwand.

Eine detaillierte Beschreibung des Simplex-Algorithmus findet sich in *Roberts & Varberg, p. 141-147* oder *Hromkovic, p. 199-209*. Softwarepakete, die den Simplex-Algorithmus implementieren, sind beispielsweise *CPLEX*, *XpressMP* (kommerziell), *GLPK* (*Gnu LP Kit*), *LP-solve* (freie Software).

5.2 Branch-and-Bound für ILP

Gesucht ist

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.d.} \quad & \mathbf{Ax} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}, x_i \in \{0, 1\}, i = 1, \dots, n. \end{aligned}$$

Dies ist ein sogenanntes *kombinatorisches Optimierungsproblem* mit einer Lösungsmenge der Mächtigkeit 2^n . Die Lösungen lassen sich durch einen binären Baum darstellen. Zur Veranschaulichung dient das folgende Beispiel.

Beispiel 5.2.1. Gegeben ist folgende Instanz des ILP

$$\begin{aligned} \max \quad & 77x_1 + 6x_2 + 3x_3 + 6x_4 + 33x_5 + 13x_6 + 110x_7 + 21x_8 + 47x_9 \\ \text{s.d.} \quad & 774x_1 + 76x_2 + 22x_3 + 42x_4 + 21x_5 + 760x_6 + 818x_7 + 62x_8 + 785x_9 \leq 1500 \\ & 67x_1 + 27x_2 + 794x_3 + 53x_4 + 234x_5 + 32x_6 + 797x_7 + 97x_8 + 435x_9 \leq 1500 \\ & x_1, \dots, x_9 \in \{0, 1\} \end{aligned}$$

Die Darstellung der Lösungen der Instanz als binärer Baum liefert Abbildung 5.7.

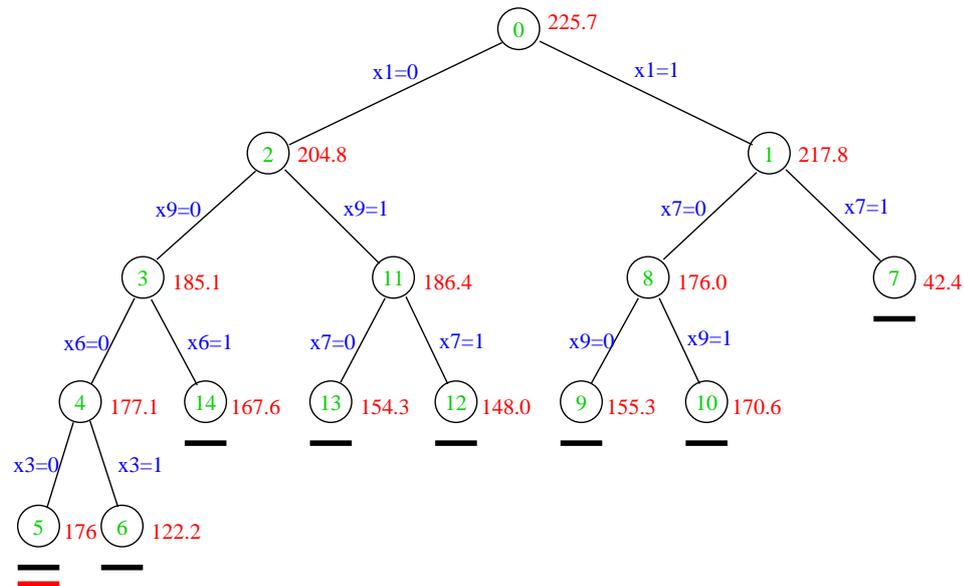


Abbildung 5.7: Lösung des kombinatorischen Optimierungsproblems als binärer Baum.

Die Vorgehensweise ist wie folgt. In jedem Knoten des Baums wird das zugehörige *relaxierte LP* gelöst. Dabei bleiben die zum Knoten führenden Variablen fest und bei allen anderen wird die Ganzzahligkeitsbedingung durch $x_i \in [0, 1]$ ersetzt (*Branching*).

Die Zweige am Knoten, bei denen die Lösung des relaxierten Problems kleiner ist als die beste bisher gefundene ganzzahlige Lösung, werden abgeschnitten. Die Programmierung dieses Verfahrens erfolgt als Programm mit rekursiver Struktur.

Branch-and-Bound mit LP-Relaxierung

Eingabe: 0/1-Programm $Z_{\text{ILP}} = \max\{\mathbf{c}'\mathbf{x} : \mathbf{x} \in \mathcal{C}\}$
mit $\mathcal{C} = \{\mathbf{x} \in \{0, 1\}^n, \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$

Ausgabe: $\mathbf{x}^* = \arg \max\{\mathbf{c}'\mathbf{x} : \mathbf{x} \in \mathcal{C}\}$

Start(){
 Initialisiere $\mathbf{x}^* = \mathbf{0}$, $Z^* = -\infty$;
 BranchAndBoundRelax (Z_{ILP});
}

BranchAndBoundRelax (0/1-Programm $Z_{\text{ILP}} = \max\{\mathbf{c}'\mathbf{x} : \mathbf{x} \in \mathcal{C}\}$){
 Betrachte LP-Relaxierung Z_{LP} von Z_{ILP} ;
 Falls Z_{LP} lösbar und $Z_{\text{LP}} > Z^*$ {
 Falls in der optimalen Lösung \mathbf{x} von Z_{LP} alle $x_i \in \{0, 1\}$ {
 Setze $\mathbf{x}^* = \mathbf{x}$ und $Z^* = Z_{\text{LP}} (= Z_{\text{ILP}})$;
 } sonst {
 Wähle ein $x_i \notin \{0, 1\}$;
 Setze $\mathcal{C}_0 = \mathcal{C} \cap \{\mathbf{x} \in \{0, 1\}^n : x_i = 0\}$
 und $\mathcal{C}_1 = \mathcal{C} \cap \{\mathbf{x} \in \{0, 1\}^n : x_i = 1\}$;
 BranchAndBoundRelax ($Z_{\text{ILP}}^0 = \max\{\mathbf{c}'\mathbf{x} : \mathbf{x} \in \mathcal{C}_0\}$);
 BranchAndBoundRelax ($Z_{\text{ILP}}^1 = \max\{\mathbf{c}'\mathbf{x} : \mathbf{x} \in \mathcal{C}_1\}$);
 }
 }
}

Beispiel 5.2.2. (Kanalzuweisung in GSM)

Bei *GSM* (*Global System for Mobile Communication*) werden Benutzer getrennt, indem verschiedene Kanäle zugewiesen werden (*FDMA*, *Frequency Division Multiple Access*). Zusätzlich werden auf der Zeitachse sich wiederholende Rahmen in jeweils 8 Zeitschlitze (*Slots*) aufgeteilt, die periodisch wiederkehrend benutzt werden (*TDMA*, *Time Division Multiple Access*). Jeder physikalische Kanal trägt also 8 Verbindungen. Kanäle können bei genügendem räumlichen Abstand wiederverwendet werden. Ziel ist es, die Kanäle im GSM-Netz so zuzuweisen, dass keine unzulässige Interferenz auftritt. Als Ausgangsdaten sind vorgegeben die Basisstationen bzw. Zellen $\mathcal{S} = \{1, \dots, S\}$ (ca. 1000 pro Niederlassung), die Kanäle bzw. Frequenzen $\mathcal{F} = \{1, \dots, N\}$ (ca. 60 in den D-Netzen) und der Mindestkanalabstand c_{ij} zwischen Zelle i und j , $i, j = 1, \dots, S$.

Üblicherweise ist $c_{ij} \in \{0, 1, 2\}$. Dabei bedeutet

- $c_{ij} = 0$: derselbe Kanal kann in Zelle i und j verwendet werden ,
- $c_{ij} = 1$: benachbarte Kanäle können in Zelle i und j verwendet werden ,
- $c_{ij} = 2$: der Kanalabstand muss mindestens 2 betragen .

r_i bezeichnet die Kanalanforderung in Zelle i , $i = 1, \dots, S$. Üblicherweise ist $r_i \in \{1, 2, 3, 4\}$, je nach Verkehrslast der Zelle. $r_i = 0$ bedeutet keinen Kanal in Zelle i .

Die Mengen $V_i \subset \mathcal{F}$, $G_i \subset \mathcal{F}$ enthalten verbotene bzw. gebotene Kanäle in Zelle i . Diese Informationen sind wichtig, beispielsweise bei angrenzenden Netzen oder um Erfahrungswerte über die Güte von Kanälen berücksichtigen zu können. Als Variable treten im Folgenden die Einträge der Kanalmatrix

$$\mathbf{X} = \begin{pmatrix} x_{ik} \end{pmatrix} \quad 1 \leq i \leq S \quad , \quad x_{ik} \in \{0, 1\} \\ 1 \leq k \leq N$$

auf. Es bedeutet

$$x_{ik} = \begin{cases} 1 , & \text{Kanal } k \text{ wird in Zelle } i \text{ zugewiesen} \\ 0 , & \text{sonst} \end{cases} .$$

Das Optimierungsproblem (*channel assignment problem*) als ILP lautet wie folgt.

$$\begin{aligned} \text{(CAP)} \quad & \max \quad \sum_{i=1}^S \sum_{k=1}^N x_{ik} \\ \text{s.d.} \quad & x_{ik} + x_{jl} \leq 1 \quad (i, k) \neq (j, l) \text{ mit } |k - l| < c_{ij} \quad (1) \\ & x_{ik} + x_{i, k+1} \leq 1 \quad i = 1, \dots, S, \quad k = 1, \dots, N - 1 \quad (2) \\ & \sum_{k=1}^N x_{ik} \leq r_i \quad i = 1, \dots, S \quad (3) \\ & x_{ik} = 0 \quad k \in V_i, \quad i = 1, \dots, S \quad (4) \\ & x_{ik} = 1 \quad k \in G_i, \quad i = 1, \dots, S \quad (5) \\ & x_{ik} \in \{0, 1\} \quad i = 1, \dots, S, \quad k = 1, \dots, N \quad (6) \end{aligned}$$

Die Zielfunktion maximiert die Anzahl zugewiesener Kanäle. (1) stellt sicher, dass Kanal k und l in Zelle i und j nicht gleichzeitig zugewiesen werden, wenn der Mindestabstand unterschritten wird. (2) impliziert, dass in Zelle i nicht zwei benachbarte Kanäle gleichzeitig zugewiesen werden. (3) beschränkt die Anzahl der in Zelle i zugewiesenen Kanäle auf höchstens r_i Stück, aufgrund der Zielfunktion werden also möglichst viele unterhalb der Beschränkung r_i zugewiesen. Die Nebenbedingungen (4) und (5) schreiben die verbotenen und gebotenen Kanäle in Zelle i vor und (6) stellt die Ganzzahligkeit sicher. Würde man in (3) Gleichheit fordern, so entstände das Problem, überhaupt eine zulässige Lösung zu finden. Alle zulässigen Lösungen sind dann optimal. Möglicherweise existiert jedoch überhaupt keine zulässige Lösung. Bei Verwendung von " \leq " in (3) ist man mit Lösungen zufrieden, bei denen nicht alle Kanalanforderungen vollständig erfüllt

sind. Vorteil ist auch, dass der Kanalplan $\mathbf{X} = \mathbf{0}_{S \times N}$ eine zulässige Lösung zum Starten des Simplex-Algorithmus ist.

Das Problem CAP ist *NP-schwer*. Realistische Problemgrößen können nicht mit Branch-and-Bound-Verfahren in vertretbarer Zeit gelöst werden. Nötig sind *Heuristiken*. Besonders bewährt hat sich das folgende *hybride* Verfahren, welches exakte mit zufallsgesteuerter Optimierung verknüpft.

CAP-Heuristik

Eingabe: Zulässiger Kanalplan $\mathbf{X} = (x_{ik})_{\substack{1 \leq i \leq S \\ 1 \leq k \leq N}}$, eventuell auch $\mathbf{X} = \mathbf{0}$

```

Start(){
  Initialisiere  $\mathbf{X}^* = \mathbf{X}$ ;
  Solange CAP ungelöst und Zeitschranke nicht erreicht {
    Wähle zufällig eine Zelle  $i_0 \in \{1, \dots, S\}$ ;
    Wähle eine Umgebung von Zellen  $U(i_0) = \{i_1, \dots, i_m\}$  um  $i_0$  so,
    dass das ILP mit Variablen  $x_{ik}$ ,  $i \in U(i_0)$ ,  $k = 1, \dots, N$  mit
    Branch-and-Bound noch lösbar ist;
    Halte  $x_{ik}$ ,  $i \notin U(i_0)$ ,  $k = 1, \dots, N$  fest. Benutze hierzu  $G_i$ ;
    Löse ILP CAP mit Variablen  $x_{ik}$ ,  $i \in U(i_0)$ ,  $k = 1, \dots, N$ ;
  }
}

```

5.3 Simulated Annealing

Besonders wichtig in der Optimierung sind Heuristiken für schwere kombinatorische Optimierungsprobleme. Eine erfolgreiche Methode ist *Simulated Annealing (SA)*, die dem Abkühlen von Metallschmelzen nachempfunden ist. Gegeben ist dabei die Instanz eines kombinatorischen Optimierungsproblems (\mathcal{S}, f) , wobei \mathcal{S} eine endliche Menge (Lösungsraum) und $f : \mathcal{S} \rightarrow \mathbb{R}$ eine Kostenfunktion ist. Das Problem lautet

$$\min_{i \in \mathcal{S}} f(i),$$

d.h. eine Lösung i^* zu finden mit $f(i^*) \leq f(i)$ für alle $i \in \mathcal{S}$.

Beispiel 5.3.1.

a) *Kanalzuweisung*

$$\mathcal{S} = \left\{ \mathbf{X} = (x_{ik}) \mid x_{ik} \text{ erfüllen die Nebenbedingungen aus (CAP)} \right\}$$

$$f : \mathcal{S} \rightarrow \mathbb{R} : \mathbf{X} \mapsto - \sum_{i,k} x_{ik}$$

- b) *Travelling Salesman Problem (TSP)* oder Handlungsreisendenproblem, Rundreiseproblem. Gegeben sind n Städte mit paarweisen Distanzen d_{ij} , $i, j = 1, \dots, n$. Gesucht ist die Rundreise durch die n Städte mit kürzester Gesamtdistanz.

Der Lösungsraum ist

$$\mathcal{S} = \left\{ \pi = \left(\pi(1), \dots, \pi(n-1), \pi(n) \right) \mid \pi \text{ ist Permutation auf } \{1, \dots, n\}, \pi(n) = n \right\},$$

d.h. alle Rundreisen mit Start in Stadt $n = \pi(0)$ und aufeinander folgenden Besuchen der Städte $\pi(1), \pi(2), \dots, \pi(n-1)$, ehe zu $\pi(n) = n$ zurückgekehrt wird. Die Kostenfunktion ist

$$f(\pi) = \sum_{i=0}^{n-1} d_{\pi(i), \pi(i+1)}.$$

Der Wert der Kostenfunktion $f(\pi)$ ist also die Gesamtentfernung der Rundreise π . Es gilt $|\mathcal{S}| = (n-1)!$.

Typisch für obige Problemstellungen ist ein Lösungsraum, der zu groß ist, um vollständig abgesucht werden zu können. Basis für ein Verfahren, das gute approximative Lösungen liefert, ist folgende Analogie aus der Physik. Beim langsamen Abkühlen von flüssigem Metall wird ein Zustand minimaler Energie angenommen, d.h. eine fast reine Gitterstruktur. Das Simulationsmodell hierfür von *Metropolis, Rosenbluth, Teller (1953)* lautet wie folgt. \mathcal{S} seien Zustände (mögliche Gitterstrukturen), $i \in \mathcal{S}$ besitze die Energie $f(i)$, T ist die Temperatur und k_B die *Boltzmann-Konstante*. Die Zustände werden sukzessive durch folgenden Algorithmus erzeugt.

Metropolis-Algorithmus

Eingabe: Zustand $i \in \mathcal{S}$

Start(){

Erzeuge zufällig einen Zustand $j \in \mathcal{S}$ in der Nachbarschaft von i ;

Falls $f(j) \leq f(i)$ {

Akzeptiere j als neuen Zustand;

} sonst {

Akzeptiere j als neuen Zustand mit Wahrscheinlichkeit $e^{-\frac{f(i)-f(j)}{Tk_B}}$,
ansonsten bleibe bei i ;

}

}

Die Zufallsvariable $X^{(T)}$ beschreibe das System im thermischen Gleichgewicht bei Temperatur T . Gemäß der statistischen Physik besitzt $X^{(T)}$ eine *Boltzmann-Verteilung*

$$P(X^{(T)} = i) = \frac{1}{Z(T)} e^{-\frac{f(i)}{Tk_B}} \quad \text{mit} \quad Z(T) = \sum_{i \in \mathcal{S}} e^{-\frac{f(i)}{Tk_B}}.$$

Später wird sich folgender Zusammenhang herausstellen. Bezeichne $X_n^{(T)}$ den Zustand zur Zeit n im Metropolis-Algorithmus bei Temperatur T , so gilt, wenn die Nachbarschaften geeignet gewählt werden,

$$\lim_{n \rightarrow \infty} P(X_n^{(T)} = i) = P(X^{(T)} = i) \quad \text{für alle } i \in \mathcal{S}.$$

Die Boltzmann-Verteilung ist also die Grenzverteilung im Metropolis-Algorithmus.

Untersucht wird nun, was passiert, wenn im Metropolis-Algorithmus für jedes T das Gleichgewicht erreicht ist und die Temperatur erhöht oder gesenkt wird.

Proposition 5.3.2. Bezeichne $\mathcal{S}_{min} = \{i \in \mathcal{S} \mid f(i) \leq f(j) \text{ für alle } j \in \mathcal{S}\}$ die Zustände minimaler Energie. Dann gilt

$$\begin{aligned} \text{a) } \quad & \lim_{T \rightarrow 0} \frac{1}{Z(T)} e^{-\frac{f(i)}{Tk_B}} = \begin{cases} \frac{1}{|\mathcal{S}_{min}|} & , \quad i \in \mathcal{S}_{min} \\ 0 & , \quad \text{sonst} \end{cases} \\ \text{b) } \quad & \lim_{T \rightarrow \infty} \frac{1}{Z(T)} e^{-\frac{f(i)}{Tk_B}} = \frac{1}{|\mathcal{S}|} \quad , \quad i \in \mathcal{S}. \end{aligned}$$

Wenn sich das System jeweils im thermischen Gleichgewicht befindet und die Temperatur auf 0 abgesenkt wird, erreicht es einen auf \mathcal{S}_{min} gleichverteilten Zustand minimaler Energie.

Diese Prinzipien werden bei *Simulated Annealing* (SA) angewendet (annealing: abkühlen, ausglühen, härten). Andere Bezeichnungen für SA sind *Monte Carlo Annealing*, *Probabilistic Hill Climbing*, *Statistical Cooling*.

Wir setzen im Folgenden voraus, dass für jede Lösung $i \in \mathcal{S}$ eine Menge von Nachbarn \mathcal{S}_i gegeben ist. Die \mathcal{S}_i bilden eine sogenannte Nachbarschaftsstruktur.

Beispiel 5.3.3.

- a) Gegeben sei der Kanalplan \mathbf{X} . Als Nachbarn können solche Kanalpläne gewählt werden, die zulässig sind und sich von \mathbf{X} an höchstens einer Stelle unterscheiden.
- b) TSP. Gegeben sei die Rundreise π . Nachbarn sind solche Rundreisen, bei denen zwei Städte ausgetauscht werden und die Reise dazwischen in umgekehrter Reihenfolge stattfindet (Lin-2-Austausch).

Simulated Annealing

Eingabe: Anfangszustand $i \in \mathcal{S}$, Anfangstemperatur c

```

Start() {
  Erzeuge zufällig Zustand  $j \in \mathcal{S}_i$ ;
  Falls  $f(j) \leq f(i)$  {
     $i = j$ ;
  } sonst {
    Falls  $e^{\frac{f(i)-f(j)}{c}} > \text{random}[0, 1]$  {
       $i = j$ ;
    }
  }
  Berechne neue Temperatur  $c$ ;
}

```

Simulated Annealing ist also der Metropolis-Algorithmus mit Abkühlung der Temperatur. Üblicherweise hält man die Temperatur c für eine vorgegebene Anzahl von Iterationen konstant und kühlt danach jeweils geometrisch ab, also mit $c_{neu} = \alpha \cdot c_{alt}$, $\alpha < 1$, α nahe bei 1.

5.4 Markov-Ketten und Konvergenz von Simulated Annealing

Simulated Annealing repräsentiert einen Typ von Algorithmus, der in ähnlicher Form in der Bildverarbeitung (*Gibbs Sampler*) und beim Trainieren neuronaler Netze (*Boltzmann-Maschinen*) auftritt. Zum Verständnis der Wirkungsweise sind Konvergenzaussagen nützlich. Sie basieren auf der Theorie der Markov-Ketten.

Definition 5.4.1. Eine Folge von Zufallsvariablen $\{X_n\}_{n \in \mathbb{N}_0}$, alle mit demselben höchstens abzählbaren Träger $\mathcal{S} = \{s_1, s_2, \dots\}$, heißt *Markov-Kette (MK)*, wenn

$$P(X_n = i_n \mid X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_n = i_n \mid X_{n-1} = i_{n-1})$$

für alle $n \in \mathbb{N}$, $i_0, \dots, i_{n-1} \in \mathcal{S}$ mit $P(X_{n-1} = i_{n-1}, \dots, X_0 = i_0) > 0$ gilt.

$P(X_n = j \mid X_{n-1} = i)$ heißt *Übergangswahrscheinlichkeit (transition probability)* im n -ten Schritt von i nach j .

Eine Markov-Kette heißt *homogen (homogeneous)*, wenn

$$P(X_n = j \mid X_{n-1} = i) = P(X_m = j \mid X_{m-1} = i)$$

für alle $i, j \in \mathcal{S}$, $n, m \in \mathbb{N}$ mit $P(X_{n-1} = i) > 0$ und $P(X_{m-1} = i) > 0$.

Markov-Ketten beschreiben das Verhalten eines stochastischen Systems mit Zustandsraum \mathcal{S} zu diskreten Zeitpunkten $n \in \mathbb{N}_0$. Die Wahrscheinlichkeit für eine Zustandsänderung hängt nur ab vom aktuellen Zustand, nicht von der Vergangenheit.

Im Folgenden werden homogene Markov-Ketten mit endlichem Zustandsraum $\mathcal{S} = \{0, 1, \dots, r\}$ oder $\mathcal{S} = \{1, \dots, r\}$ betrachtet. Viele der weiteren Aussagen lassen sich auch auf abzählbare Zustandsräume übertragen.

Beispiel 5.4.2. (Fritchman Kanalmodell)

Ein Kanal befindet sich im Zustand *gut* ($= 1$, Bitfehlerrate 10^{-6}) oder *schlecht* ($= 0$, Bitfehlerrate 10^{-3}). Mit Wahrscheinlichkeit p wechselt der Zustand von 1 nach 0 und mit Wahrscheinlichkeit q von 0 nach 1. Das Modell lautet wie folgt. Der Zustandsraum ist $\mathcal{S} = \{0, 1\}$. Der Zustand zum Zeitpunkt n wird beschrieben durch die Zufallsvariable X_n , $n \in \mathbb{N}_0$. Die Übergangswahrscheinlichkeiten sind

$$\begin{aligned} P(X_n = 0 \mid X_{n-1} = 1) &= p & , & \quad P(X_n = 1 \mid X_{n-1} = 0) = q \\ P(X_n = 1 \mid X_{n-1} = 1) &= 1 - p & , & \quad P(X_n = 0 \mid X_{n-1} = 0) = 1 - q . \end{aligned}$$

Eine grafische Veranschaulichung der Übergangswahrscheinlichkeiten liefert Abbildung 5.8.

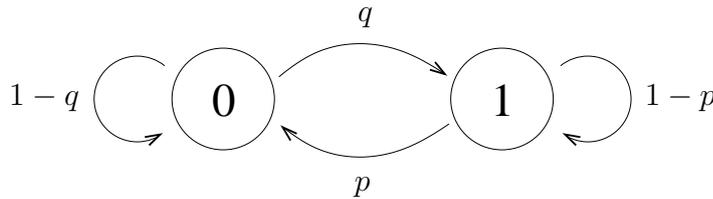


Abbildung 5.8: Übergangsgraph im Fritchman Kanalmodell.

Es folgen einige Bezeichnungen und Konventionen. Die Übergangswahrscheinlichkeiten

$$P(X_n = j \mid X_{n-1} = i) = p_{ij} \quad , \quad i, j \in \mathcal{S}$$

lassen sich darstellen als *Übergangsmatrix*

$$\mathbf{\Pi} = \left(p_{ij} \right)_{i,j \in \mathcal{S}} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ \vdots & \vdots & & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix} .$$

Übergangsmatrizen sind *stochastische Matrizen*, d.h.

$$p_{ij} \geq 0 \quad \text{für alle } i, j \in \mathcal{S} \quad \text{und} \quad \sum_{j \in \mathcal{S}} p_{ij} = 1 \quad \text{für alle } i \in \mathcal{S} .$$

Die Verteilung von X_n , dem Zustand zum Zeitpunkt n , wird beschrieben durch einen stochastischen Vektor

$$\mathbf{p}(n) = (p_1(n), \dots, p_r(n)) = (p_i(n))_{i \in \mathcal{S}} = (P(X_n = 1), \dots, P(X_n = r)) .$$

Die *Anfangsverteilung* einer Markov-Kette ist

$$\mathbf{p}(0) = (p_1(0), \dots, p_r(0)) = (P(X_0 = 1), \dots, P(X_0 = r)) .$$

Proposition 5.4.3 (Berechnung von Verteilungen einer Markov-Kette).

- a) Die Verteilung zum Zeitpunkt n ist eindeutig bestimmt durch die Anfangsverteilung und die Übergangswahrscheinlichkeiten, d.h.

$$\mathbf{p}(n) = \mathbf{p}(n-1)\mathbf{\Pi} = \mathbf{p}(n-2)\mathbf{\Pi}^2 = \dots = \mathbf{p}(0)\mathbf{\Pi}^n .$$

- b) Übergangswahrscheinlichkeiten höherer Stufen lauten

$$p_{ij}^{(n)} = P(X_n = j \mid X_0 = i) = (\mathbf{\Pi}^n)_{i,j} \quad \text{für alle } i, j \in \mathcal{S} .$$

- c) Es gelten die *Chapman-Kolmogorov-Gleichungen*

$$p_{ij}^{(m+n)} = \sum_{l \in \mathcal{S}} p_{il}^{(m)} p_{lj}^{(n)} .$$

Festzustellen ist nun, ob das durch eine Markov-Kette beschriebene System ein *statisches Gleichgewicht* besitzt. Die konkrete Frage ist, ob die Verteilung von X_n mit $n \rightarrow \infty$ gegen eine feste Verteilung konvergiert, d.h.

$$\lim_{n \rightarrow \infty} \mathbf{p}(n) = \mathbf{p}^* \quad \text{mit} \quad \mathbf{p}^* \mathbf{\Pi} = \mathbf{p}^*$$

existiert. Das muss nicht immer der Fall sein.

Beispiel 5.4.4. (Fritchman Kanalmodell mit $p = q = 1$)

Für $p = q = 1$ gilt $\mathbf{\Pi} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\mathbf{\Pi}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, also

$$\mathbf{\Pi}^n = \begin{cases} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, & n \text{ ungerade} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & n \text{ gerade} \end{cases} .$$

Gelte $\mathbf{p}(0) = (1, 0)$, also $P(X_0 = 0) = 1$, $P(X_0 = 1) = 0$, d.h. das System startet im Zustand 0 = *schlecht*. Dann ist

$$\mathbf{p}(n) = (1, 0)\mathbf{\Pi}^n = \begin{cases} (0, 1), & n \text{ ungerade} \\ (1, 0), & n \text{ gerade} \end{cases}$$

eine alternierende Folge, die nicht konvergiert.

Definition 5.4.5. Eine Markov-Kette bzw. Übergangsmatrix heißt

a) *irreduzibel*, wenn für alle $i, j \in \mathcal{S}$ Zustände $i_1, \dots, i_n \in \mathcal{S}$ existieren, sodass

$$p_{i_1 i_1} \cdot p_{i_1 i_2} \cdots p_{i_{n-1} i_n} \cdot p_{i_n j} > 0 ,$$

d.h. jeder Zustand von jedem anderen in endlich vielen Schritten mit positiver Wahrscheinlichkeit erreicht werden kann.

b) *aperiodisch*, wenn sie irreduzibel ist und für alle $i \in \mathcal{S}$ gilt

$$d(i) := \text{ggT} \left\{ n \in \mathbb{N} \mid p_{ii}^{(n)} > 0 \right\} = 1 .$$

Bei Irreduzibilität ist $d(i)$ derselbe Wert für alle $i \in \mathcal{S}$.

Proposition 5.4.6. Ist $\{X_n\}_{n \in \mathbb{N}}$ eine irreduzible, aperiodische Markov-Kette mit endlichem Zustandsraum \mathcal{S} , so existiert eine eindeutige *stationäre Verteilung* \mathbf{p}^* mit $\mathbf{p}^* \mathbf{\Pi} = \mathbf{p}^*$. Ferner gilt

$$\lim_{n \rightarrow \infty} \mathbf{\Pi}^n = \begin{pmatrix} p_1^* & \cdots & p_r^* \\ \vdots & & \vdots \\ p_1^* & \cdots & p_r^* \end{pmatrix} , \quad \mathbf{p}^* = (p_1^*, \dots, p_r^*) \text{ die eindeutige stationäre Verteilung}$$

und $\lim_{n \rightarrow \infty} \mathbf{p}(n) = \mathbf{p}^*$ unabhängig von der Anfangsverteilung.

Simulated Annealing wird nun durch Markov-Ketten modelliert. Im Algorithmus SA bezeichne $X_n^{(c)}$ den Zustand zum Zeitpunkt $n \in \mathbb{N}_0$ bei fester Temperatur c und $p_{ij}^{(c)}$ die Übergangswahrscheinlichkeiten. Dann gilt

$$p_{ij}^{(c)} = \begin{cases} g_{ij} a_{ij}(c) & , \quad i \neq j \\ 1 - \sum_{l \neq i} g_{il} a_{il}(c) & , \quad i = j \end{cases} ,$$

wobei $a_{ij}(c) = \exp \left\{ -\frac{(f(j)-f(i))^+}{c} \right\}$, $i, j \in \mathcal{S}$ und g_{ij} die Generierungswahrscheinlichkeit eines neuen Zustands $j \in \mathcal{S}_i$ bezeichnet. Wählt man nun die Generierungswahrscheinlichkeiten und damit auch die Nachbarschaften so, dass die Matrix $\mathbf{G} = (g_{ij})_{i,j \in \mathcal{S}}$ symmetrisch und irreduzibel ist, so konvergiert die Verteilung von $X_n^{(c)}$ gegen eine Boltzmann-Verteilung.

Proposition 5.4.7. Gelte $\mathbf{G} = (g_{ij})_{i,j \in \mathcal{S}}$ ist symmetrisch, d.h. $g_{ij} = g_{ji}$ für alle $i, j \in \mathcal{S}$, und irreduzibel, so folgt für alle $c > 0$ fest

$$\lim_{n \rightarrow \infty} P(X_n^{(c)} = i) = \frac{1}{Z(c)} e^{-\frac{f(i)}{c}} , \quad i \in \mathcal{S} ,$$

mit der normierenden Konstanten $Z(c) = \sum_{i \in \mathcal{S}} e^{-\frac{f(i)}{c}}$, unabhängig von der Startverteilung.

Das Verhalten der Boltzmann-Verteilung bei Abkühlen der Temperatur c wurde bereits in Proposition 5.3.2 betrachtet. Als Konsequenz ergibt sich

$$\lim_{c \rightarrow 0} \lim_{n \rightarrow \infty} P(X_n^{(c)} = i) = \begin{cases} \frac{1}{|\mathcal{S}_{min}|} & , \quad i \in \mathcal{S}_{min} \\ 0 & , \quad \text{sonst} \quad . \end{cases}$$

Simulated Annealing konvergiert also in Verteilung gegen einen optimalen Zustand. In der Praxis ist bei der Umsetzung des doppelten Limes Vorsicht geboten, da man bei einem endlichen Wert abbrechen muss. Bei genügend langsamen Abkühlen und geeigneter Wahl der Startparameter und Nachbarschaftsstruktur endet der Algorithmus jedoch in der Regel bei Funktionswerten, die sehr nahe am Optimum liegen.

5.5 Konvexe Optimierung

Zur Erinnerung: Eine Menge M heißt *konvex* (convex), wenn

$$\alpha \mathbf{x} + (1 - \alpha) \mathbf{y} \in M \quad (*)$$

für alle $\mathbf{x}, \mathbf{y} \in M, \alpha \in [0, 1]$.

Bemerkung: $\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}$ muss definiert sein. In der Regel ist $M \subseteq V$, V ein linearer Vektorraum, z.B. $V = \mathbb{R}$, $V = \mathbb{R}^n$, $V = \mathbb{C}^{m \times n}$, auch $V = C[a, b]$.

Interpretation von (*): Mit je zwei Punkten liegt auch die Verbindungsgerade in M .

Definition 5.5.1. M sei konvex. Eine Funktion $f : M \rightarrow \mathbb{R}$ heißt konvex (konkav), wenn

$$f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) \underset{(\geq)}{\leq} \alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y})$$

für alle $\mathbf{x}, \mathbf{y} \in M, \alpha \in [0, 1]$.

Interpretation: Die Verbindungsgerade liegt stets oberhalb (unterhalb) der Funktion.

Proposition 5.5.2. $f : M \rightarrow \mathbb{R}$ ist konvex (konkav) genau dann, wenn

$$f\left(\sum_{i=1}^n \alpha_i \mathbf{x}_i\right) \underset{(\geq)}{\leq} \sum_{i=1}^n \alpha_i f(\mathbf{x}_i)$$

für alle $\mathbf{x}_1, \dots, \mathbf{x}_n \in M, \alpha_i \geq 0$ mit $\sum_{i=1}^n \alpha_i = 1$.

Beispiel 5.5.3. Bei der Bestimmung der Kapazität von MIMO-Kanälen mit fester Kanalmatrix $\mathbf{H} \in \mathbb{C}^{r \times t}$ tritt folgende Funktion auf (hier vereinfachend nur für reelle Matrizen):

$$f(\mathbf{Q}) = \log \det(\mathbf{I} + \mathbf{H} \mathbf{Q} \mathbf{H}'), \quad \mathbf{Q} \in \mathbb{R}^{t \times t}$$

(i)

$\mathbb{R}_+^{t \times t}$, die Menge aller nicht-negativ definiten Matrizen ist konvex. Denn:

$$\mathbf{x}'(\alpha \mathbf{Q}_1 + (1 - \alpha) \mathbf{Q}_2) \mathbf{x} = \underbrace{\alpha \mathbf{x}' \mathbf{Q}_1 \mathbf{x}}_{\geq 0} + \underbrace{(1 - \alpha) \mathbf{x}' \mathbf{Q}_2 \mathbf{x}}_{\geq 0} \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^t$$

für alle $\mathbf{Q}_1, \mathbf{Q}_2 \in \mathbb{R}_+^{t \times t}, \alpha \in [0, 1]$.

(ii)

$f(\mathbf{A}) = \log \det \mathbf{A}$, $\mathbf{A} \in \mathbb{R}_+^{t \times t}$, ist konkav. Benutze zum Beweis die Ungleichung von Ky Fan:

$$\det(\alpha \mathbf{A} + (1 - \alpha) \mathbf{B}) \geq (\det \mathbf{A})^\alpha (\det \mathbf{B})^{1-\alpha}$$

für alle $\mathbf{A}, \mathbf{B} \geq 0, \alpha \in [0, 1]$. Also

$$\log \det(\alpha \mathbf{A} + (1 - \alpha) \mathbf{B}) \geq \alpha \log \det \mathbf{A} + (1 - \alpha) \log \det \mathbf{B}.$$

(iii)

$f(\mathbf{Q}) = \log \det(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}')$, $\mathbf{Q} \in \mathbb{R}_+^{t \times t}$, ist konkav. Denn:

$$\begin{aligned} & \log \det(\mathbf{I} + \mathbf{H}(\alpha \mathbf{Q}_1 + (1 - \alpha) \mathbf{Q}_2) \mathbf{H}') \\ &= \log \det(\alpha \underbrace{(\mathbf{I} + \mathbf{H}\mathbf{Q}_1 \mathbf{H}')}_{\geq 0} + (1 - \alpha) \underbrace{(\mathbf{I} + \mathbf{H}\mathbf{Q}_2 \mathbf{H}')}_{\geq 0}) \\ &\stackrel{(ii)}{\geq} \alpha \log \det(\mathbf{I} + \mathbf{H}\mathbf{Q}_1 \mathbf{H}') + (1 - \alpha) \log \det(\mathbf{I} + \mathbf{H}\mathbf{Q}_2 \mathbf{H}') \end{aligned}$$

für alle $\mathbf{Q}_1, \mathbf{Q}_2 \in \mathbb{R}_+^{t \times t}, \alpha \in [0, 1]$.

Die Aussagen gelten analog für komplexe hermitesche Matrizen. Zur Bestimmung der Kapazität löse also

$$\max_{\mathbf{Q} \geq 0, \text{tr } \mathbf{Q} \leq L} \log \det(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^*).$$

Es handelt sich um eine konkave Zielfunktion, die auf einer konvexen Menge maximiert wird, denn $\{\mathbf{Q} \geq 0 \mid \text{tr } \mathbf{Q} \leq L\}$ ist als Durchschnitt konvexer Mengen ebenfalls konvex.

□

Beispiel 5.5.4. Diskreter Kanal mit Eingabe X , Ausgabe Y , Kanalmatrix

$$\mathbf{W} = (w_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}, \quad w_{ij} = P(Y = j \mid X = i)$$

und Inputverteilung

$$\mathbf{p} = (p_1, \dots, p_m), \quad p_i = P(X = i).$$

\mathbf{W} ist aus \mathcal{S} , der Menge der stochastischen $(m \times n)$ -Matrizen, diese ist offensichtlich konvex. \mathbf{p} ist aus \mathcal{D} , der Menge der stochastischen Vektoren der Dimension m , diese ist offensichtlich auch konvex. Mit der Schreibweise

$$\mathbf{W} = \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_m \end{pmatrix}$$

gilt für die Transinformation:

$$\begin{aligned}
 I(X; Y) &= H(Y) - H(Y|X) \\
 &= H(\mathbf{p}\mathbf{W}) - \sum_{i=1}^m p_i H(\mathbf{w}_i) \\
 &= - \sum_{j=1}^n \left(\sum_{i=1}^m p_i w_{ij} \right) \log \left(\sum_{i=1}^m p_i w_{ij} \right) + \sum_{i=1}^m p_i \sum_{j=1}^n w_{ij} \log w_{ij} \\
 &= \sum_{i=1}^m p_i \sum_{j=1}^n w_{ij} \log \frac{w_{ij}}{(\mathbf{p}\mathbf{W})_j} \\
 &= \sum_{i=1}^m p_i D(\mathbf{w}_i \parallel \mathbf{p}\mathbf{W}) \\
 &= I(\mathbf{p}; \mathbf{W})
 \end{aligned}$$

Der Beweis der zweiten Aussage ist recht trickreich.

Es gilt: $I(\mathbf{p}; \mathbf{W})$ ist eine konkave Funktion in $\mathbf{p} \in \mathcal{D}$ und konvex in $\mathbf{W} \in \mathcal{S}$. Zur Bestimmung der Kapazität löse bei festem \mathbf{W}

$$\max_{\mathbf{p} \in \mathcal{D}} I(\mathbf{p}; \mathbf{W}) = \max_{p_i \geq 0, \sum_{i=1}^n p_i = 1} I(\mathbf{p}; \mathbf{W}).$$

□

Typische Probleme sind also: Maximiere (minimiere) eine konkave (konvexe) Funktion unter Ungleichungs- und Gleichungsnebenbedingungen. Allgemeines Lösungsverfahren: Karush-Kuhn-Tucher-Bedingungen (KKT-Bedingungen).

$$\begin{aligned}
 \text{(CP)} \quad & \min && f(\mathbf{x}) \\
 & \text{s.d.} && g_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, s \\
 & && h_j(\mathbf{x}) = 0, \quad j = 1, \dots, m \\
 & && \mathbf{x} \in \mathbb{R}^n \\
 & \text{wobei} && f(\mathbf{x}), g_i(\mathbf{x}), i = 1, \dots, s, \text{ konvex} \\
 & && h_j(\mathbf{x}), j = 1, \dots, m, \text{ affin}
 \end{aligned}$$

Definiere die *Lagrange-Funktion*:

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^s \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^m \mu_j h_j(\mathbf{x})$$

KKT-Bedingungen:

$$\begin{aligned}
 g_i(\mathbf{x}) &\leq 0, \quad i = 1, \dots, s \\
 h_j(\mathbf{x}) &= 0, \quad j = 1, \dots, m \\
 \boldsymbol{\lambda} &\geq 0 \\
 \lambda_i g_i(\mathbf{x}) &= 0, \quad i = 1, \dots, s \\
 \nabla f(\mathbf{x}) + \sum_{i=1}^s \lambda_i \nabla g_i(\mathbf{x}) + \sum_{j=1}^m \mu_j \nabla h_j(\mathbf{x}) &= 0
 \end{aligned}$$

Theorem 5.5.5. Wenn $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ die KKT-Bedingungen erfüllt, so löst \mathbf{x}^* (CP). Wenn \mathbf{x}^* (CP) löst und es existiert ein zulässiger innerer Punkt mit $g_i(\mathbf{x}) < 0$ für alle i , dann existieren $\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*$, so dass $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ die KKT-Bedingungen erfüllt.

Beispiel 5.5.6. (Berechnung der Kapazität für den diskreten Kanal aus Beispiel 5.5.4)
Benutze Darstellung

$$I(\mathbf{p}; \mathbf{W}) = H(\mathbf{pW}) - \sum_{i=1}^m p_i H(\mathbf{w}_i).$$

Nach längerer Rechnung folgt

$$\frac{\partial}{\partial p_i} I(\mathbf{p}; \mathbf{W}) = D(\mathbf{w}_i \parallel \mathbf{pW}) - 1.$$

Löse also:

$$\begin{aligned}
 \max \quad & I(\mathbf{p}; \mathbf{W}) \\
 \text{s.d.} \quad & p_i \geq 0, \quad i = 1, \dots, m \\
 & \sum_{j=1}^m p_j = 1
 \end{aligned}$$

Lagrange-Funktion:

$$L(\mathbf{p}, \boldsymbol{\lambda}, \mu) = I(\mathbf{p}; \mathbf{W}) + \sum_{i=1}^m \lambda_i p_i + \mu \left(\sum_{j=1}^m p_j - 1 \right)$$

KKT-Bedingungen:

$$\begin{aligned}
 p_i &\geq 0, \quad \sum_{j=1}^m p_j = 1 \\
 \lambda_i &\geq 0 \\
 \lambda_i p_i &= 0 \\
 D(\mathbf{w}_i \parallel \mathbf{pW}) + \lambda_i + \nu &= 0, \quad \text{wobei } \nu = \mu - 2
 \end{aligned}$$

Mit Theorem 5.5.5 folgt:

$$\mathbf{p}^* \text{ ist optimal} \Leftrightarrow D(\mathbf{w}_i \parallel \mathbf{p}^* \mathbf{W}) = \xi \quad \forall i \text{ mit } p_i > 0.$$

Hieraus kann man für den Binary Asymmetric Channel (BAC), der durch die Übertragungswahrscheinlichkeiten

$$\begin{aligned} P(Y = 0|X = 0) &= 1 - \varepsilon, & P(Y = 1|X = 0) &= \varepsilon, \\ P(Y = 1|X = 1) &= 1 - \delta, & P(Y = 0|X = 1) &= \delta \end{aligned}$$

charakterisiert ist, schließen

$$p_0^* = \frac{1}{1+b}, \quad p_1^* = \frac{b}{1+b}$$

mit

$$b = \frac{a\varepsilon - (1 - \varepsilon)}{\delta - a(1 - \delta)}, \quad a = \exp\left(\frac{h(\delta) - h(\varepsilon)}{1 - \varepsilon - \delta}\right), \quad h(\varepsilon) = H(\varepsilon, 1 - \varepsilon).$$

Index

- σ -Algebra, 10
- (M,N) -Code mit Leistungsbeschränkung, 94
- Übertragungswahrscheinlichkeit, 76

- Abtasttheorem, 100
- Aquivokation, 69
- Autokorrelationsfunktion, 47, 50
- Autokovarianzfunktion, 47

- Bandpassfilter, 57
- Bayes-Formel, 13
- Bessel-Funktion
 - modifizierte 1. Art, 39
- Boltzmann
 - Konstante, 124
 - Maschine, 126
 - Verteilung, 124, 129
- Bonferroni-Ungleichung, 12
- Box-Muller-Verfahren, 30
- Branch-and-Bound, 120

- Cauchy-Schwarz Ungleichung, 22, 66
- Channel Assignment Problem, 122
- Chapman-Kolmogorov-Gleichung, 128

- Dekodierfehler, 78
- Dekodierregel, 78
 - ME-Dekodierung, 79
 - ML-Dekodierung, 79
- Determinante
 - Absolutbetrag der, 85
 - Jacobi, 31
- Dichte, 17
 - bedingte, 82
 - von Zufallsvektoren, 23
- Dirac-Impuls, 62
- disjunkt, 10

- Energiespektrum, 55
- Entropie, 68, 80, 87
 - bedingte, 69, 87
 - bedingte differentielle, 82
 - differentielle, 81, 91
 - gemeinsame, 69
 - gemeinsame differentielle, 81
 - Kettenregel für, 84
 - linearer Transformationen, 85
 - maximale differentielle, 85
 - mittlere, 69
- Ereignis, 9
- Ereignismenge, 9
- Ergebnis, 9
- Ergebnismenge, 9
- Erwartungswert
 - funktion, 47
 - vektor, 24
 - diskreter Zufallsvariable, 20
 - linearer Transformationen, 25
 - Linearität des, 22
 - Monotonie des, 22
 - stetiger Zufallsvariable, 20
- Extremalpunkt, 118, 119

- Fading, 40
 - Rayleigh, 109
- Faltung, 31
- farbiges Rauschen, 64
- FDMA, 121

- Fouriertransformierte, 55
- Fritchman Kanalmodell, 127, 128
- Funktion
 - nicht-negativ definite, 47
 - orthonormale, 92
- Gaußkanal
 - bandbegrenzter, 100
 - Kapazität des, 93
 - mit Feedback, 101
 - mit Leistungsbeschränkung, 93
 - Kapazität eines, 94
 - paralleler, 96
 - reeller, 92
 - unabhängiger paralleler komplexer, 108
- Gaußprozess, 49
 - Stationarität bei, 50
- Gerade, 118
- Gibbs-Sampler, 126
- gleitendes Mittel, 58
- GSM, 121

- Hadamard-Ungleichung, 98
- Halbraum, 114, 117
- Heuristik, 113, 123
- hybrides Verfahren, 123
- Hyperebene, 117

- Impulsantwort, 53
 - eines LTI-Systems, 58
 - Faltung der, 61
 - physikalisch realisierbare, 53
- Informationsgehalt, 68
- Instanz, 113
 - als binärer Baum, 120
- Jensen-Ungleichung, 111

- Kanal
 - kapazität, 77
 - Übertragung in bandbegrenztem, 100
 - binärer symmetrischer, 70, 77, 79
 - diskreter und gedächtnisloser, 77
 - komplexer MIMO, 104
 - Transinformation eines, 76
- Kanalmatrix
 - feste, 106
 - zufällige, 109
- Kanaluweisung, 121, 123
- Kapazität
 - im diskreten Fall, 94
 - obere Schranke für die, 111
- Kardinalität, 12
- Kode, 73
 - alphabet, 72
 - länge, 73
 - wort, 73
 - eindeutig dekodierbarer, 73, 74
 - erwartete Länge des, 73
 - präfixfreier, 73, 74
- Kodierung, 73
- konvex, 117
- Korrelation, 21
- Kovarianz, 21
 - linearer Transformationen, 25
- Kovarianzmatrix, 25
- Kreuzkorrelationsfunktion, 54
- Kugelpackung, 95
- Kullback-Leibler-Distanz, 70, 81, 83

- Lösung
 - optimale, 114, 119
 - zulässige, 114, 117
- Leistungsbeschränkung, 92
 - mittlere, 95
- Leistungsdichtespektrum, 56
 - des LTI-gefilterten Prozesses, 56
 - konstantes, 60
- Lineares Programm
 - ganzzahliges, 114
 - in kanonischer Form, 113, 117
 - in Standardform, 116
 - relaxiertes, 120
- Markov-Kette, 126

- Anfangsverteilung einer, 128
- aperiodische, 129
- homogene, 126
- irreduzible, 129
- Verteilung einer, 128
- Markov-Ungleichung, 22
- Matched Filter, 64
- Matrix
 - nicht-negativ definite, 47
 - Spektralzerlegung einer, 97, 98
 - stochastische, 127
 - transponierte, 22
- Maximal Covering Location, 115
- Metropolis-Algorithmus, 124
- Mobilfunkkanal
 - Mehrwegeausbreitung im, 33
- Moment
 - k-tes, 21
 - k-tes zentrales, 21
- Momentanleistung
 - erwartete, 57
- Normalverteilung, 18, 81
 - n -dimensionale, 23, 41, 82
 - Faltungstabilität der, 31
- NP-schwer, 113, 123
- Optimierungsproblem
 - als ILP, 115
 - kombinatorisches, 120
- Partition, 13
- Permutationsmatrix, 110
- Pfadgewinn, 104
- Polytop, 114
 - abgeschlossenes konvexes, 118
- Positivteil, 97
- Quellalphabet, 72
- Quelle
 - diskrete gedächtnislose, 75
- Quellenkodierung, 72
- Quellenrate, 80
- Randdichte, 31
- Rate
 - erreichbare, 94
 - maximal erreichbare, 96
- Rauschen
 - farbiges Gaußsches, 101
 - Gaußsches weißes, 92
- Rechteckimpuls, 53
- rechtsseitig stetig, 15
- Satz von Bochner-Khintchin, 55
- Satz von Parseval, 55
- schiefsymmetrisch, 48
- Schlupfvariable, 116
- Shannonscher Fundamentalsatz, 79
- Siebformel, 12
- Signal
 - energie, 55
 - bandbegrenztes, 100
 - deterministisches, 55
 - empfangenes, 34
 - gesendetes, 34
 - komplexes, 48
 - leistungsbeschränktes, 100
- signal-to-noise ratio, 108
- Signal-zu-Rausch Abstand, 65
- Simplex-Algorithmus, 119
- Simulated Annealing, 123, 126, 129
- Störprozess, 92
- Standardabweichung, 21
- Standardnormalverteilung, 10
- stationäre Verteilung, 129
- statistisches Gleichgewicht, 128
- stochastisch unabhängig, 14, 26
 - gemeinsam, 14
 - identisch verteilt, 26
- stochastischer Prozess, 46, 49
 - schwach stationärer, 49
 - strikt stationärer, 49
- Synentropie, 69
 - differentielle, 82
- System

- lineares zeitinvariantes, 53
- TDMA, 121
- Tiefpassfilter, 52
- Träger, 16
- Transformationssatz für Dichten, 29
- Transinformation, 69, 81, 87
 - differentielle, 82
 - eines Kanals, 76
- Travelling Salesman Problem, 124, 125

- Übergangsgraph, 127
- Übergangsmatrix, 127
- Übergangswahrscheinlichkeit, 126
- Übertragung
 - von Symbolen, 76
- Varianz, 21
- Verteilung
 - Bernoulli, 16
 - binomial, 17
 - diskret gleich, 16
 - exponential, 18
 - Gamma, 18
 - Faltungsstabilität der, 31
 - geometrisch, 17
 - gleich, 18, 81
 - Lognormal, 40
 - Dichte der, 41
 - Nakagami-m, 19
 - normal, 18
 - Poisson, 17
 - Rayleigh, 18, 36, 109
 - Rechteck, 18, 81
 - Rice, 38, 40
- Verteilungsdichte, 17
- Verteilungsfunktion, 15
 - gemeinsame, 23
 - von Zufallsvektoren, 23
- Vertex Cover, 114

- Wahrscheinlichkeit, 9
 - bedingte, 12

- Berechnung von, 19, 24
 - totale, 13
- Wahrscheinlichkeitsmaß, 11
- Wahrscheinlichkeitsraum, 11, 14, 46
- Wahrscheinlichkeitsverteilung, 11
- water filling, 97, 106, 107
- weißes Rauschen, 62
 - bandbegrenztes, 64
 - Filterung von, 64
 - Gaußsches, 62

- Zähldichte, 16
- Zufallsgröße, 15
- Zufallsvariable, 15
 - absolut-stetige, 17
 - diskrete, 16
 - einpunktverteilte, 43
 - Summe von, 30
- Zufallsvektor, 22
 - absolut-stetiger, 23
 - komplex normalverteilter, 43
 - zirkulär symmetrisch komplex normalverteilter, 44, 104
 - Entropie des, 105
 - zirkulär symmetrisch komplex verteilter, 44