

Dr. Jochen Kerdels, Prof. Dr. Gabriele Peters

Modul 63312

Interaktive Systeme

Lehrveranstaltung

**Interaktive Systeme I: Konzepte und Methoden des
Computersehens**

LESEPROBE

Fakultät für
**Mathematik und
Informatik**

Das Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung und des Nachdrucks bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (Druck, Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung der FernUniversität reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Vorwort

Es freut uns, Sie als Teilnehmerinnen und Teilnehmer des Kurses *01698 – Interaktive Systeme I* begrüßen zu dürfen. Der vorliegende Kurs hat das Ziel, Ihnen die zentralen Konzepte und Methoden des Computersehens zu vermitteln. Das Gebiet des sogenannten *Computersehens* – der visuellen Wahrnehmung der natürlichen Umgebung durch einen Computer – kombiniert eine große Zahl unterschiedlicher und z.T. anspruchsvoller Verfahren und Konzepte u.a. der klassischen Signalverarbeitung, Analysis und linearen Algebra, Statistik und Wahrscheinlichkeitstheorie, sowie des maschinellen Lernens. Diese methodische Vielfalt macht auf der einen Seite den besonderen Reiz des Computersehens aus, zuweilen stellt sie jedoch auf der anderen Seite eine gewisse Einstiegshürde dar. In diesem Sinne verfolgt der Kurs zwei Ziele: Die repräsentative Auswahl der in diesem Kurs vorgestellten Konzepte und Methoden soll Ihnen einen Eindruck der Vielfalt des Gebietes des Computersehens vermitteln und Ihnen gleichzeitig ein solides Verständnis der zentralen konzeptionellen und mathematischen Grundlagen verschaffen. Das in diesem Kurs vermittelte Wissen befähigt Sie somit einerseits, einen großen Teil „alltäglicher“ Aufgaben des Computersehens praktisch bewältigen zu können, und es befähigt Sie andererseits, weiterführender Fachliteratur mit ausreichenden konzeptionellen und mathematischen Fähigkeiten zu begegnen.

Die ersten beiden Kurseinheiten haben ihren Schwerpunkt in der „klassischen“ Signalverarbeitung. Sie stellen das theoretische Fundament dar, auf dem alle weiteren Verfahren implizit oder explizit aufbauen. Kurseinheit 1 definiert zunächst die grundlegenden Eigenschaften der in diesem Kurs behandelten Signale. Anschließend wird in einem kurzen Abschnitt erläutert, wie analoge Signale in digitale Signale überführt werden. Die zweite Hälfte der Kurseinheit konzentriert sich auf eine zentrale Eigenschaft fast aller signalverarbeitender Systeme: Linearität. Darauf aufbauend wird die mathematische Operation der *Faltung* eingeführt und im Detail untersucht. Kurseinheit 2 baut auf das zuvor gelegte Fundament auf und widmet sich in Gänze einer eingehenden und ausführlichen Behandlung der diskreten Fourier-Transformation. Ausgehend von einer fast informellen Beschreibung, die sich auf die Vermittlung der zugrundeliegenden Konzepte und Ideen konzentriert, wird im Anschluss die komplexwertige, mathematische Beschreibung der Fourier-Transformation erarbeitet. Es folgt eine vollständige und ausführliche Beschreibung der sogenannten *Fast Fourier Transform* (FFT). Die Kurseinheit endet mit einer Einordnung der diskreten Fourier-Transformation in die Familie der Fourier-Transformationen und einer Erweiterung der Fourier-Transformation auf den zweidimensionalen Fall.

Im Anschluss daran werden in der dritten Kurseinheit ausgewählte Themen der Bildverarbeitung im Detail vorgestellt. Hierbei werden Ihnen nicht nur die Konzepte der einzelnen Methoden vermittelt, sondern insbesondere auch die jeweils verwendeten mathematischen Beschreibungen. Beginnend mit einer kurzen Einführung in verschiedene Bildformate, Farbmodelle und der realen Organisation der Bilddaten auf einem Computersystem, bilden im Anschluss die Themen *Segmentierung*, *Kantendetektion* und *merkmalbasierte Verfahren* den Schwerpunkt der restlichen Kurseinheit. Die vorgestellten Verfahren sind in diesem Zusammenhang insbesondere unter dem Aspekt ausgewählt worden, dass sie ein solides Grundwissen bzgl. der wichtigsten Konzepte und mathematischen Methoden der digitalen Bildverarbeitung vermitteln.

Die vierte Kurseinheit legt ihren Schwerpunkt auf erweiterte Methoden aus den Bereichen der Statistik, Wahrscheinlichkeitsrechnung und des maschinellen Lernens, die eine Interpretation von Signalen bzw. Daten auf einem semantisch höheren Niveau ermöglichen. Die vorgestellten Verfahren sind hierbei repräsentativ für Methoden, die am Ende einer Signalverarbeitungskette stehen und die wahrgenommenen visuellen Signale schließlich in eine inhaltlich interpretierbare Form überführen.

Hagen im Mai 2016

Jochen Kerdels, Gabriele Peters

Noch ein Hinweis zum Sprachgebrauch. Da wissenschaftliche Texte und Lehrtexte möglichst objektiv formuliert sein sollten, und da es sich bei der Annahme, die Verwendung des einen Geschlechts schließe das andere automatisch ein, um einen Irrtum handelt, ist der vorliegende Text weder in ausschließlich femininer noch in ausschließlich maskuliner Form verfasst. Vielmehr verwenden wir im Wechsel mal die männliche, mal die weibliche Form und orientieren uns im Übrigen an den Empfehlungen zur „Gleichbehandlung im Sprachgebrauch“ der Gesellschaft für Informatik e.V. von 1999 und an den „Richtlinien für einen nicht-sexistischen Sprachgebrauch“ der UNESCO von 1993.

Kurseinheit 1

Grundlagen der Signalverarbeitung

Inhaltsübersicht

1.1	Signale und ihre zugrunde liegenden Prozesse	12
1.1.1	Statistische Beschreibung	14
1.2	Analog-Digital-Wandlung	31
1.3	Lineare Systeme	36
1.3.1	Superpositionsprinzip	40
1.3.2	Dekomposition von Signalen	42
1.3.3	Analyse nicht-linearer Systeme	48
1.4	Faltung diskreter Signale	49
1.4.1	Eigenschaften der Faltung	54
1.4.2	Elementare Impulsantworten	55
1.4.3	Korrelation	59
1.5	Faltung kontinuierlicher Signale	60
1.5.1	Dirac-Impuls	61
1.5.2	Faltung	63
1.6	Selbsttestaufgaben	65
1.7	Lösungen der Selbsttestaufgaben	69

Lernziele:

In dieser Kurseinheit lernen Sie zunächst die grundlegenden Eigenschaften digitaler Signale im Kontext der Signalverarbeitung kennen. Im Anschluss wird erläutert, wie diese Signale durch lineare Systeme verändert und miteinander kombiniert werden können. Den Abschluss der Kurseinheit bildet eine Betrachtung der *Faltung* – einer mathematischen Operation, die die Kombination und Filterung von Signalen ermöglicht und im engen Zusammenhang mit den zuvor beschriebenen linearen Systemen steht. Da die in dieser Kurseinheit vorgestellten Konzepte und Methoden zu den wichtigsten Grundlagen der digitalen Signalverarbeitung zählen, werden diese besonders ausführlich behandelt, um eine solide Grundlage für das weitere Verständnis der folgenden Kurseinheiten zu legen.

Das Arbeits- und Forschungsgebiet der Mensch-Computer-Interaktion (MCI) umfasst ein weites Spektrum an Themengebieten und interessanten Problemstellungen. Es reicht von „klassischen“ Fragestellungen, wie der Gestaltung von grafischen Benutzungsschnittstellen und dem Entwurf direkter, haptischer Schnittstellen, über die Erforschung zukünftiger Schnittstellen, die sich die Erfassung, Verarbeitung und Interpretation von Sprache, Gestik, Mimik und Emotion zu eigen machen, zu sehr grundsätzlichen Fragestellungen, wie etwa der Frage, wie die Semantik komplexer Informationen von einem Computersystem erlernt werden kann.

Interaktionsbegriff

Ein wichtiger und zentraler Aspekt der Mensch-Computer-Interaktion, der als verbindendes Element aller Teilgebiete der MCI aufgefasst werden kann, ist hierbei der Begriff der *Interaktion*, wie er bereits im Kurs 01697 „Einführung in Mensch-Computer-Interaktion“ definiert wurde:

**Definition 1.1** (Interaktion).

Interaktion ist der zielgerichtete^a, wechselseitige Transfer von Information zwischen zwei oder mehr Entitäten.

^aHier ist das *Ziel* im Sinne von Intention / Absicht / Plan gemeint, nicht die Richtung des Informationstransfers. Die Richtung des Informationstransfers ist jedoch meist ebenfalls durch das *Ziel* implizit mit vorgegeben.

Essenzieller Bestandteil dieser Kommunikation zwischen den Entitäten einer Interaktion ist die Verarbeitung der eingehenden Signale auf Seiten des Computers. Sie ist eine zwingende Voraussetzung dafür, dass der Kommunikationsprozess zwischen Mensch und Computer *wechselseitig* stattfinden kann und der Computer adäquat auf die vom Menschen transferierte Information¹ reagieren kann.

¹oder allgemein: auf die Information von anderen an der Interaktion beteiligten Entitäten

In diesem Sinne befasst sich diese Kurseinheit zunächst mit den Grundlagen allgemeiner Signalverarbeitung, ohne sich dabei auf eine konkrete Form der Interaktion zwischen Mensch und Computer zu beschränken. Die folgenden Abschnitte orientieren sich an dem äußerst empfehlenswerten Buch „Digital Signal Processing“ von Steven W. Smith [Smi02]².

²Das Buch ist online unter der Adresse <http://www.dspguide.com> kostenfrei und legal zum Download verfügbar.

1.1 Signale und ihre zugrunde liegenden Prozesse

Lernziele:

In diesem Abschnitt werden Sie die grundlegenden Beziehungen zwischen Signalen und ihren zugrunde liegenden Prozessen kennenlernen. In diesem Zusammenhang werden die wesentlichen Aspekte einer statistischen Beschreibung der Signale und einer darauf basierenden Approximation der statistischen Eigenschaften der zugehörigen Prozesse betrachtet. Besondere Aufmerksamkeit wird hierbei der Erstellung und den Eigenschaften von Histogrammen gewidmet, da diese in vielen Algorithmen und Verfahren der Mensch-Computer-Interaktion Verwendung finden.

Signalbegriff

Im Kontext der Mensch-Computer-Interaktion findet der Begriff des *Signals* auf verschiedene Weise Verwendung. So werden beispielsweise die Nachrichten innerhalb einer grafischen Benutzungsschnittstelle zuweilen als Signale³ bezeichnet. An anderer Stelle, im Rahmen der Erkennung und Interpretation von Gesten und Mimik, werden bestimmte, charakteristische Bewegungen oder spezifische Gesichtsausdrücke mit dem Begriff des *Signals* verknüpft. An wieder anderer Stelle dienen Signale dazu, die Benutzerin über einen Zustandswechsel des Systems zu informieren, zum Beispiel durch die Wiedergabe eines Signaltons. Diese und viele andere Verwendungen des Signalbegriffes im MCI-Zusammenhang finden im Allgemeinen in einem jeweils sehr spezifischen und engen Kontext statt und sind nicht Gegenstand dieser Kurseinheit.

An dieser Stelle soll es vielmehr um Signale im folgenden Sinne gehen:



Definition 1.2 (Signal).

Ein **Signal** ist die Beschreibung einer Variablen in Abhängigkeit einer anderen Variablen.

abhängige &
unabhängige
Variablen

kontinuierliche &
diskrete Signale

Der Temperaturverlauf eines Tages ist ein Beispiel für ein Signal in diesem Sinne. Das Signal beschreibt in diesem Fall die Veränderung der Variablen *Temperatur* in Abhängigkeit der Variablen *Zeit*. In diesem Beispiel ist die Temperatur die sogenannte *abhängige Variable* und die Zeit die sogenannte *unabhängige Variable*. Beide Variablen – sowohl die Temperatur als auch die Zeit – können Werte aus einem kontinuierlichen Wertebereich annehmen. Man spricht dementsprechend von einem *kontinuierlichen Signal*. Für die Verwendung innerhalb eines Computersystems wird ein kontinuierliches Signal üblicherweise mittels eines Analog-Digital-Wandlers *digitalisiert*. Man spricht anschließend von einem *diskreten Signal*. In diesem Zusammenhang bezieht sich der Begriff „diskret“ sowohl auf die unabhängige als auch auf die abhängige Variable. Im Rahmen der Digitalisierung eines Signals wird entlang der

³z.B. beim *Signals&Slots*-Konzept des GUI-Frameworks Qt

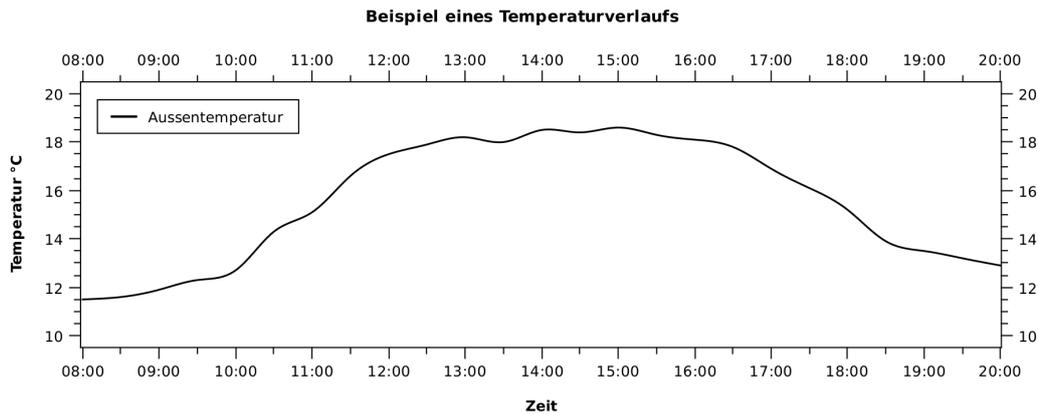


Abbildung 1.1: Darstellung eines Temperaturverlaufs.

unabhängigen Variablen in regelmäßigen⁴ Abständen die abhängige Variable *quantisiert*, d.h. die abhängige Variable wird auf einen Wert innerhalb eines endlichen, ganzzahligen Intervalls abgebildet. Handelt es sich z.B. um einen 10-Bit Analog-Digital-Wandler, so wird die abhängige Variable auf einen Wert aus dem Intervall $\{0..1023\}$ abgebildet. Wie dieser Wert zu interpretieren ist, hängt vom jeweiligen *Messbereich*⁵ des Analog-Digital-Wandlers ab. Ginge der Messbereich im Falle unseres Beispiels von $-51,2^{\circ}\text{C}$ bis $+51,1^{\circ}\text{C}$, so würde die Temperatur in Schritten von $0,1^{\circ}\text{C}$ quantisiert. Die Schrittweite, mit der entlang der unabhängigen Variablen die abhängige Variable quantisiert wird, wird als *Abtastrate* (engl. *sampling rate*) bezeichnet. In seltenen Fällen gibt es auch Mischformen zwischen kontinuierlichen und diskreten Signalen, bei denen jeweils eine der beiden Variablen kontinuierlich und die andere diskret ist. Ein Signal mit einer diskreten unabhängigen Variablen wird hierbei auch *zeitdiskret* genannt und ein Signal mit einer diskreten abhängigen Variablen *wertdiskret*.

Quantisierung

Abtastrate

Abbildung 1.1 illustriert die übliche Darstellungsweise eines Signals in Form eines Graphen. Hierbei repräsentiert die vertikale Achse⁶ die abhängige Variable (hier: die Temperatur) und die horizontale Achse⁷ die unabhängige Variable (hier: die Zeit). Das in Abbildung 1.1 dargestellte Signal ist ein kontinuierliches Signal. Handelt es sich um ein diskretes Signal, so werden die einzelnen Messwerte meist als einzelne Punkte dargestellt, sofern ihre Anzahl relativ gering (< 100) ist. Grundsätzlich sollte man sich hierbei jedoch nicht rein auf die Darstellung verlassen, sondern über die Beschriftung des Graphen bzw. den jeweiligen Kontext überprüfen, ob es sich um ein kontinuierliches oder um ein diskretes Signal handelt.

Zuordnung der Achsen

⁴Es werden üblicherweise regelmäßige Abstände verwandt. Dies ist jedoch nicht zwingend erforderlich.

⁵auch *Eingangssignalebereich* genannt

⁶Die vertikale Achse wird auch als y-Achse, Ordinate oder Amplitude bezeichnet.

⁷Die horizontale Achse wird auch als x-Achse, Abszisse oder Domäne bezeichnet.

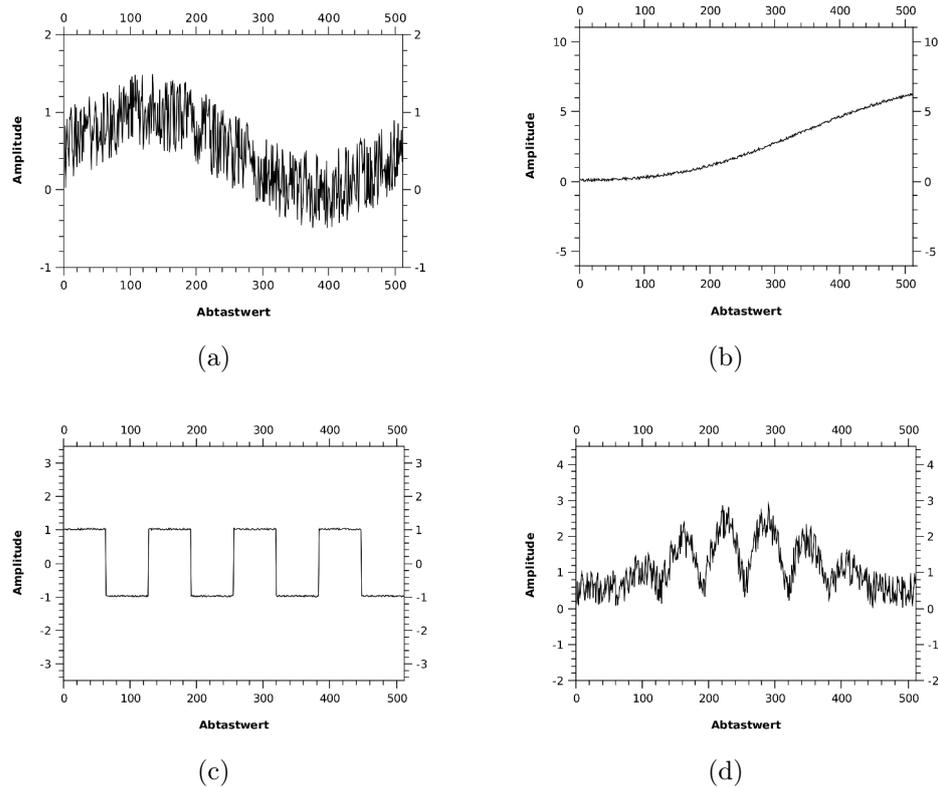


Abbildung 1.2: Vier Beispiele für einfache Signale.

1.1.1 Statistische Beschreibung

Charakterisierung
& Vergleich
von Signalen

Nachdem nun der vorherige Abschnitt die grundlegenden Eigenschaften und den allgemeinen Aufbau von Signalen definiert hat, stellt sich die Frage, wie verschiedene, konkrete Signale charakterisiert und untereinander verglichen werden können. Betrachtet man beispielsweise die vier in Abbildung 1.2 dargestellten Signale, so erscheinen die Unterschiede und Gemeinsamkeiten zwischen den Signalen offensichtlich:

- Signal (a) scheint eine niederfrequente Schwingung mit hohem Rauschanteil⁸ zu sein.
- Signal (b) hat einen deutlich geringeren Rauschanteil und beschreibt einen fast linearen Anstieg der Amplitude.
- Signal (c) ist ein klassisches Rechtecksignal mit einem ebenfalls nur sehr geringem Rauschanteil.
- Signal (d) hingegen ist ein komplexes Signal mit zunächst steigender und anschließend wieder fallender Amplitude und einem mit Signal (a) vergleichbarem Rauschanteil.

⁸Der Begriff *Rauschanteil* wird an dieser Stelle rein informal verwendet, um einen subjektiven Eindruck zu beschreiben.

Einem menschlichen Betrachter fällt es in der Regel leicht, diese charakteristischen Eigenschaften der Signale anhand einer einfachen visuellen Prüfung der Graphen dieser Signale abzuleiten. Entspringen die Signale einer bestimmten Domäne – handelt es sich zum Beispiel um Gehirnstromwellen, die durch ein EEG⁹ erfasst wurden – so sind Expertinnen auf dem jeweiligen Gebiet häufig in der Lage, eine sehr präzise und aussagekräftige Charakterisierung der Signale vorzunehmen. Im Falle der EEG-Signale kann beispielsweise eine erfahrene Expertin rein aufgrund der visuellen Prüfung auf mögliche Krankheiten, z.B. Epilepsie, schließen.

visuelle Prüfung

Auf den ersten Eindruck hin mag die Methode der *visuellen Prüfung* sehr subjektiv, unpräzise und unzuverlässig erscheinen. Es ist jedoch vielfach so, dass in vielen Anwendungsbereichen, wie z.B. auch in der klinischen Analyse von EEG-Daten, eine ausschließlich computergestützte Analyse der Signale noch nicht die Qualität einer manuellen Expertenanalyse erreicht. Ein Blick auf Abbildung 1.3 mag einen Eindruck davon vermitteln, warum die Entwicklung geeigneter, rein computergestützter Analysen so schwierig ist. Die dargestellte Tabelle zeigt die jeweils ersten 40 Abtastwerte der vier Beispielfunktionen (a) bis (d). Der Versuch, nur anhand dieser Tabellendaten¹⁰ eine Charakterisierung der Signale vorzunehmen, erscheint ungleich schwerer im Vergleich zur visuellen Darstellung. Dies liegt zum einen daran, dass durch die Betrachtung einzelner Datenpunkte der größere Kontext verloren geht und zum anderen, dass die textuelle Darstellung der Datenpunkte keinen visuellen, „qualitativen“ Vergleich bzgl. ihrer relativen räumlichen Position erlaubt. So ist beispielsweise bei der tabellarischen Darstellung die Abschätzung des Rauschanteils deutlich schwieriger als bei einer grafischen Darstellung. Um dennoch *nützliche* Aussagen über die Charakteristik eines Signals machen zu können, bedient man sich u. A. Mitteln der Statistik, welche in ihren Grundzügen in den folgenden Unterabschnitten näher beschrieben werden.

Kontext & qualitativer Vergleich

Arithmetisches Mittel und Standardabweichung

Wie im vorhergehenden Abschnitt motiviert, dienen die nachfolgend beschriebenen, statistischen Maßzahlen und Verfahren der Bestimmung von charakteristischen Eigenschaften diskreter Signale. Anhand dieser Eigenschaften können Signale z.B. miteinander verglichen oder bestimmten Klassen zugeordnet werden.

statistische Maßzahlen

Eine erste, einfache und intuitive Maßzahl ist der durchschnittliche Abtastwert eines Signals. Dieser Wert wird als *arithmetisches Mittel* des Signals bezeichnet und wird aus der Summe über alle Abtastwerte dividiert durch ihre Anzahl berechnet:

⁹Ein Elektroenzephalograph ist ein Gerät zur Messung von Gehirnstromwellen über an der Kopfhaut angebrachte Elektroden.

¹⁰Die Daten in der Tabelle sind natürlich unvollständig, da sie nur die ersten 40 Werte umfassen. Aber auch eine seitenfüllende Darstellung aller 512 Werte hätte vermutlich nur wenig mehr Erhellendes beigetragen.

Abtastwert	(a)	(b)	(c)	(d)
0	0.229	0.0883	1.02	0.0291
1	0.0246	0.102	1.02	0.142
2	0.827	0.162	1.03	0.757
3	0.0282	0.107	1.01	0.147
4	0.547	0.0904	1.03	0.112
5	0.913	0.187	1.02	0.723
6	0.531	0.186	1	0.283
7	0.531	0.0153	1.03	0.46
8	0.302	0.0102	1.04	0.852
9	0.925	0.0924	1.03	0.994
10	0.909	0.0724	1.03	0.875
11	1.06	0.132	1.01	0.739
12	0.936	0.181	1.02	0.601
13	0.532	0.0575	1.01	0.251
14	0.428	0.198	1.02	0.227
15	0.174	0.0941	1.01	0.586
16	0.654	0.0666	1.02	0.483
17	1.1	0.0596	1.01	1
18	0.428	0.194	1.02	0.295
19	1.06	0.0193	1.01	0.747
20	0.291	0.0393	1.02	0.736
21	0.593	0.155	1.01	0.262
22	0.884	0.0715	1.05	0.587
23	0.571	0.214	1.02	0.999
24	0.345	0.0551	1	0.955
25	0.996	0.147	1.03	0.738
26	1.02	0.2	1	0.288
27	0.695	0.0879	1.05	0.675
28	0.368	0.0917	1.03	0.664
29	0.834	0.22	1	0.708
30	0.484	0.112	1.03	0.157
31	1.07	0.155	1.03	0.618
32	0.745	0.0904	1.03	0.875
33	0.976	0.0981	1.02	1
34	0.783	0.0585	1.03	0.106
35	0.952	0.203	1.02	0.57
36	0.787	0.103	1.05	0.134
37	0.912	0.122	1.03	0.734
38	0.96	0.205	1.05	0.633
39	1.2	0.0575	1.02	0.189
⋮	⋮	⋮	⋮	⋮

Abbildung 1.3: Die ersten 40 Abtastwerte der vier Signalbeispiele aus Abbildung 1.2.

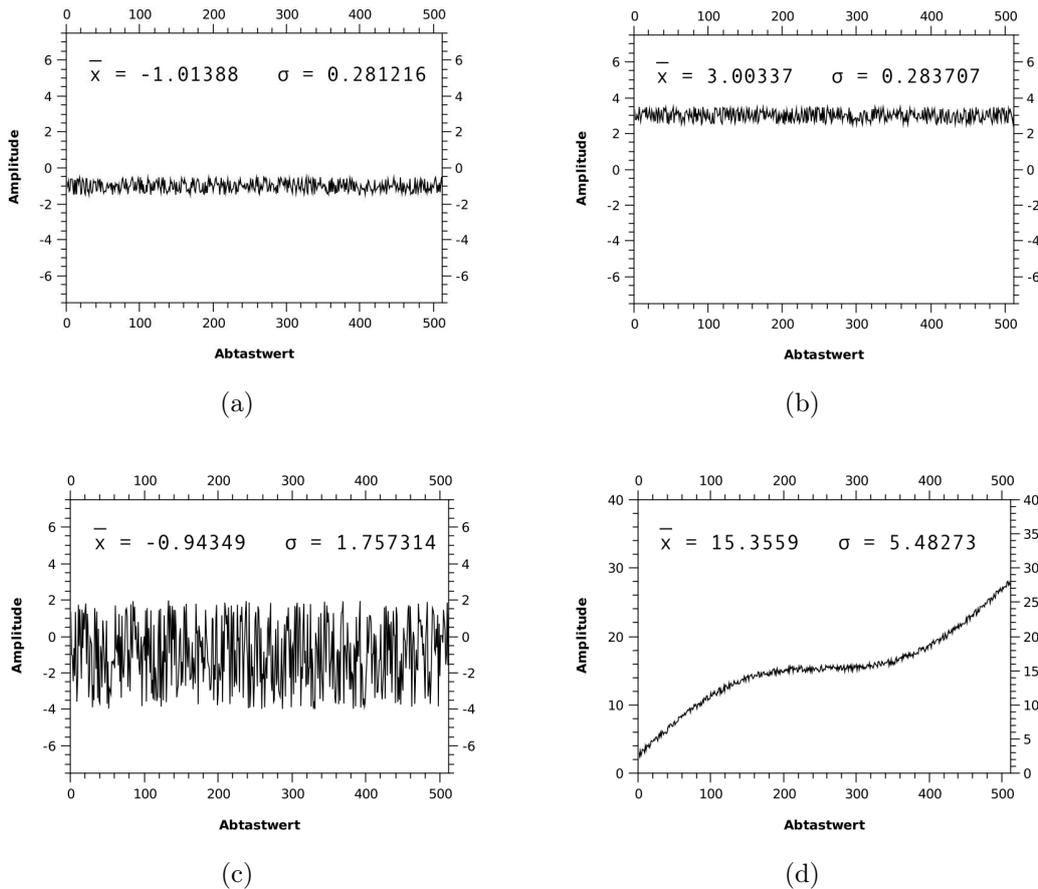


Abbildung 1.4: Arithmetisches Mittel und Standardabweichung von vier Beispielsignalen.

Definition 1.3 (Arithmetisches Mittel).

Gegeben sei das Signal $x = (x_0, x_1, \dots, x_{N-1})$ bestehend aus N Abtastwerten. Dann ist

$$\bar{x} := \frac{1}{N} \sum_{i=0}^{N-1} x_i$$

das **arithmetische Mittel** des Signals x .



Abbildung 1.4 zeigt vier Beispielsignale und ihre jeweiligen arithmetischen Mittel. Der Vergleich von Signal (a) mit Signal (b) illustriert hierbei die wesentliche Aussagekraft des arithmetischen Mittels. Es beschreibt die Verschiebung des Signals entlang der y-Achse (Amplitude). In der Elektrotechnik wird das arithmetische Mittel auch DC-Offset, Gleichanteil, oder Gleichstrom- bzw. Gleichspannungsanteil genannt. Im gegebenen Beispiel ist das Signal (b) gegenüber dem Signal (a) um etwa +4 Einheiten verschoben. Der Vergleich von Signal (a) mit Signal (c) zeigt andererseits eine Schwäche dieser statistischen Maßzahl auf. Beide Signale haben in etwa das gleiche arithmetische Mittel, jedoch variieren die Abtastwerte von Signal (c) in weit größerem Maße

Gleichanteil

als die Werte von Signal (a). Diese Eigenschaft der Signale geht durch die Summenbildung bei der Berechnung des arithmetischen Mittels verloren.

Streuungsmaß

Eine Möglichkeit, die Stärke der Variation mit einer Maßzahl – einem sogenannten *Streuungsmaß* – zu erfassen, ist die *mittlere absolute Abweichung*. Sie wird berechnet, indem für jeden Abtastwert des Signals der absolute Abstand zum arithmetischen Mittel des Signals gebildet, anschließend aufsummiert und schließlich mit der Anzahl der Abtastwerte normiert wird:



Definition 1.4 (Mittlere absolute Abweichung).

Gegeben sei das Signal $x = (x_0, x_1, \dots, x_{N-1})$ bestehend aus N Abtastwerten und es sei \bar{x} das arithmetische Mittel dieses Signals. Dann ist

$$x_{AD^a} := \frac{1}{N} \sum_{i=0}^{N-1} |x_i - \bar{x}|$$

die *mittlere absolute Abweichung* des Signals x .

^aDas Kürzel *AD* ist der englischen Bezeichnung „absolute deviation“ entlehnt.

alternative
Streuungsmaße

Ein Nachteil der mittleren absoluten Abweichung ist die Verwendung der Betragsfunktion. Diese erschwert eine analytische Behandlung, da sie nicht überall differenzierbar ist. Zwei alternative Streuungsmaße, welche diese Problematik umgehen, sind die Varianz und die Standardabweichung. Sie verwenden anstelle der Betragsfunktion das Quadrat der Differenz von Abtastwert und arithmetischem Mittel des Signals:



Definition 1.5 (Varianz).

Gegeben sei das Signal $x = (x_0, x_1, \dots, x_{N-1})$ bestehend aus N Abtastwerten und es sei \bar{x} das arithmetische Mittel dieses Signals. Dann ist

$$\sigma^2 := \frac{1}{N} \sum_{i=0}^{N-1} (x_i - \bar{x})^2$$

die *Varianz* des Signals x .



Definition 1.6 (Standardabweichung).

Gegeben sei das Signal $x = (x_0, x_1, \dots, x_{N-1})$ bestehend aus N Abtastwerten und es sei \bar{x} das arithmetische Mittel dieses Signals. Dann ist

$$\sigma = \sqrt{\sigma^2} := \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (x_i - \bar{x})^2}$$

die *Standardabweichung* des Signals x .

Die Standardabweichung ist somit lediglich die Quadratwurzel der Varianz. Betrachtet man erneut Abbildung 1.4 und vergleicht die Signale (a) bis (c), so charakterisieren arithmetisches Mittel und die Standardabweichung die einzelnen Signale bereits recht aussagekräftig. Handelt es sich bei dem zu beschreibenden Signal jedoch um ein sogenanntes *nicht stationäres* Signal, so verlieren arithmetisches Mittel und Standardabweichung an Aussagekraft. Ein Beispiel für ein derartiges, nicht stationäres Signal zeigt Abbildung 1.4d. Mit den stationären Signalen (a) bis (c) als Modellannahme, würde man sich das Signal (d) nur anhand seiner Werte für arithmetisches Mittel und Standardabweichung als stationäres Signal um einen Amplitudenwert von etwa 15 herum zentriert und mit einem relativ hohen Rauschanteil versehen vorstellen. Diese Vorstellung ist in Anbetracht der grafischen Darstellung von Signal (d) in Abbildung 1.4 offensichtlich inkorrekt. Das Beispiel zeigt, dass im Falle zweifelhafter, meist implizit gemachter Grundannahmen – hier die Annahme es handle sich um ein stationäres Signal – die Verwendung und Interpretation statistischer Maßzahlen zu falschen Schlussfolgerungen über die Charakteristik des untersuchten Signals führen kann. Derartigen Grundannahmen bzw. Modellannahmen sollte grundsätzlich eine hohe Aufmerksamkeit geschenkt werden, da nur in den seltensten Fällen die gemachten Schlussfolgerungen *allgemeine* Gültigkeit besitzen.

nicht stationäre
Signale

implizite
Grundannahmen

In diesem Zusammenhang besteht ein weiterer wichtiger Aspekt in der bewussten Unterscheidung zwischen dem *Signal* und dem diesem Signal *zugrunde liegenden Prozess*. Ein Beispiel hierfür liefert ein klassisches Münzwurfexperiment. Ein Münzwurf ist ein binäres Ereignis. Er generiert entweder *Kopf* oder *Zahl*, wobei die Wahrscheinlichkeit für jedes der beiden möglichen Ergebnisse genau 50% beträgt. Bildet man das Ergebnis *Kopf* auf den Wert 1 ab und das Ergebnis *Zahl* auf den Wert 0, so hat der *Münzwurfprozess* ein arithmetisches Mittel von exakt 0,5. Führt man jedoch ein konkretes Münzwurfexperiment durch, z.B. über 1000 Würfe, und fasst die Ergebnisse der einzelnen Würfe als Signal auf, so hat dieses Signal in der Regel ein leicht von 0,5 abweichendes arithmetisches Mittel. Während die Wahrscheinlichkeiten und das daraus abgeleitete arithmetische Mittel des zugrunde liegenden Münzwurfprozesses konstant sind, variieren die arithmetischen Mittel der durch Münzwurfexperimente zu Stande gekommenen Signale bei jedem Durchlauf. Diese Abweichung vom theoretisch erwarteten Mittelwert wird auch als *zufällige Abweichung* oder *Zufallsfehler* bezeichnet.

Signal vs. zugrunde
liegender Prozess

Zufallsfehler

Abbildung 1.5a illustriert diesen Zufallsfehler und zeigt die Entwicklung des arithmetischen Mittels des Signals eines Münzwurfexperimentes¹¹ (rote, durchgezogene Linie) in Abhängigkeit der Anzahl der Würfe (Abtastwerte). Man erkennt, dass sich das arithmetische Mittel mit zunehmender Zahl der Würfe stabilisiert, jedoch auch nach 1000 Würfeln noch deutlich vom theoretisch erwarteten Mittelwert abweicht. Auf den ersten Blick mag dieser Zufallsfehler nach 1000 Würfeln unintuitiv hoch erscheinen. Bei einem arithmetischen Mittel von 0,48 nach 1000 Würfeln ist es jedoch so, dass gerade einmal 20 *Zahl*-Ergebnisse (abgebildet auf 0) „zuviel“ stattgefunden haben und entsprechend 20 *Kopf*-

¹¹Es handelt sich um ein theoretisches Münzwurfexperiment auf Basis von echten Zufallszahlen.

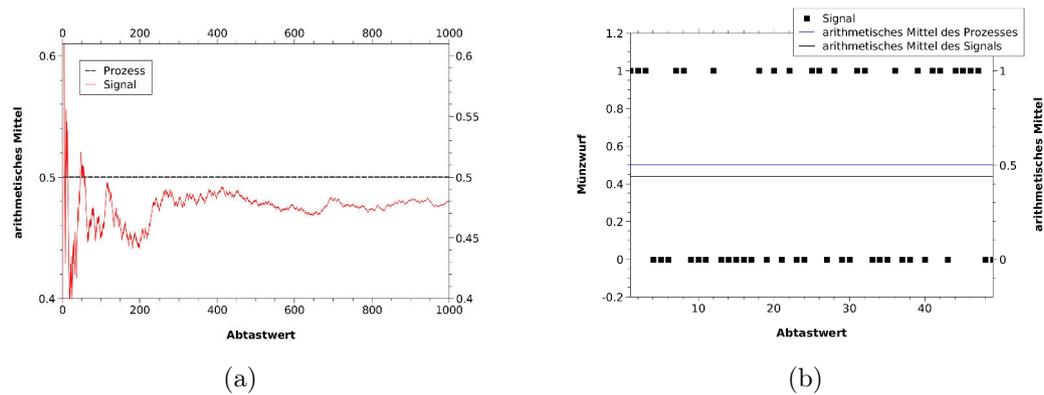


Abbildung 1.5: Erläuterung der geschätzten Varianz. **(a)** Darstellung des arithmetischen Mittels eines Münzwurfexperimentes. Die schwarze, gestrichelte Linie zeigt das konstante arithmetische Mittel von 0,5 des zugrunde liegenden Münzwurfprozesses. Die rote, durchgezogene Linie zeigt das arithmetische Mittel des Signals („Beobachtung“) bei zunehmender Anzahl von Münzwürfen. **(b)** Darstellung eines Münzwurfexperimentes mit 50 Würfeln. Das arithmetische Mittel des Münzwurfexperimentes (0,44) ist als fett gezeichnete, schwarze Linie dargestellt. Das arithmetische Mittel des Münzwurfprozesses (0,5) ist durch eine dünne, blaue Linie angegeben.

Ergebnisse (abgebildet auf 1) „zuwenig“. Aus dieser Perspektive betrachtet erscheinen die Werte durchaus plausibel¹².

Fehlerfort-
pflanzung

Das Auftreten von Zufallsfehlern in statistischen Maßzahlen hat insbesondere dann weitergehende Auswirkungen, wenn die betroffenen Maßzahlen ihrerseits wieder in die Berechnung weiterer Maßzahlen einfließen¹³. So wird beispielsweise das potenziell fehlerbehaftete, arithmetische Mittel für die Berechnung der Varianz verwendet. Welche Auswirkungen hat in diesem Fall die Verwendung eines fehlerbehafteten arithmetischen Mittels auf die Varianz? Um diese Frage zu untersuchen soll erneut das Beispiel des Münzwurfexperimentes verwendet werden. Abbildung 1.5b zeigt das Ergebnis eines solchen Experimentes nach 50 Würfeln. Das aus diesen 50 Werten berechnete arithmetische Mittel (0,44) ist als fett gezeichnete, schwarze Linie dargestellt. Das „ideale“ arithmetische Mittel des zugrunde liegenden Münzwurfprozesses (0,5) ist als dünne, blaue Linie eingezeichnet. Da es sich um horizontale Linien handelt, vereinfacht sich die übliche Geradengleichung von $ax + b$ auf den konstanten Wert b , da bei horizontalen Geraden die Steigung a gleich 0 ist. Im Falle des berechneten arithmetischen Mittels ist $b = 0,44$, im Falle des idealen arithmetischen Mittels des Münzwurfprozesses ist $b = 0,5$. Betrachtet man nun die oben beschriebene Berechnung der Varianz, so lässt sich erkennen, dass man die Varianz auch als durchschnittlichen Abstand der Werte eines Signals von der durch das arithmetische Mittel beschriebenen horizontalen Geraden verstehen kann. Vernachlässigt man für einen Moment den Normierungsfaktor $1/N$, so ergibt

¹²Ein weiteres Experiment mit 10.000 Würfeln (nicht abgebildet) ergab ein arithmetisches Mittel von 0,5015 – also in diesem Fall etwa nur 15 *Kopf*-Ergebnisse „zuviel“ bei 10.000 Ereignissen insgesamt.

¹³Dieses Phänomen wird auch als *Fehlerfortpflanzung* bezeichnet

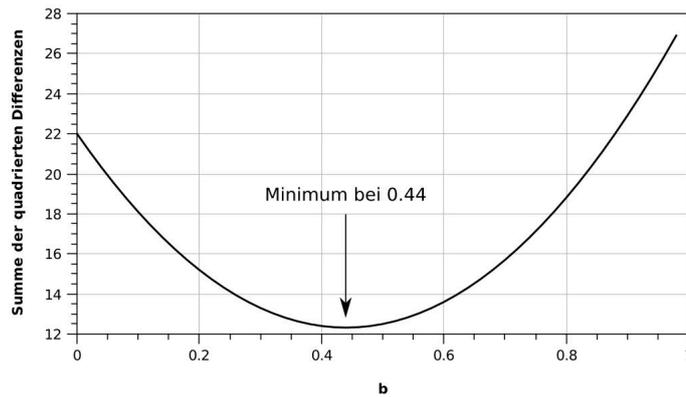


Abbildung 1.6: Minimaler quadratischer Abstand. Die Kurve zeigt die Summe der quadratischen Abstände der einzelnen Werte des Münzwurfexperimentes zu verschiedenen horizontalen Geraden mit Werten von $b = 0$ bis $b = 1$. Die Kurve nimmt ihr Minimum genau bei der Geraden an, die dem arithmetischen Mittel der Werte des Münzwurfexperimentes entspricht ($b = 0.44$).

sich für die Summe der quadratischen Abstände der Werte eines Signals zu einer beliebigen horizontalen Geraden:

$$\sum_{i=0}^{N-1} (x_i - b)^2$$

In Abbildung 1.6 wurde diese Summe nun für verschiedene Werte von b aus dem Intervall $[0, 1]$ und den 50 Werten x_i aus dem Münzwurfexperiment ausgewertet. Es zeigt sich, dass die Summe der quadratischen Abstände – und damit auch die Varianz – ihr Minimum genau beim berechneten arithmetischen Mittel ($b = 0.44$) der Werte x_i hat! Andersherum betrachtet bedeutet dies, dass für alle anderen Werte von b die Summe – und damit die Varianz – einen höheren Wert annimmt.

Zurück zur eingangs gestellten Frage: Welche Auswirkungen hat die Verwendung eines fehlerbehafteten arithmetischen Mittels ($b = 0.44$ anstelle von $b = 0.5$) auf die Varianz? Aus dem in Abbildung 1.6 gezeigten Minimum lässt sich folgern, dass auch, wenn man das „ideale“ arithmetische Mittel des zugrundeliegenden Prozesses nicht kennt, man sich sicher sein kann, dass die Varianz, die man mit dem berechneten arithmetischen Mittel errechnet hat, immer zu klein ist im Vergleich zu der Varianz, die sich mit dem „idealen“ arithmetischen Mittel ergeben würde. Um diesen Effekt insbesondere bei kurzen Signalen zu kompensieren, findet man häufig die folgenden Definitionen von Varianz und Standardabweichung:



Definition 1.7 (Schätzung der Varianz).

Gegeben sei das Signal $x = (x_0, x_1, \dots, x_{N-1})$ bestehend aus N Abtastwerten und \bar{x} sei das arithmetische Mittel dieses Signals. Dann ist

$$\tilde{\sigma}^2 := \frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - \bar{x})^2$$

die **geschätzte Varianz** des dem Signal x zugrunde liegenden Prozesses.



Definition 1.8 (Schätzung der Standardabweichung).

Gegeben sei das Signal $x = (x_0, x_1, \dots, x_{N-1})$ bestehend aus N Abtastwerten und \bar{x} sei das arithmetische Mittel dieses Signals. Dann ist

$$\tilde{\sigma} = \sqrt{\tilde{\sigma}^2} := \sqrt{\frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - \bar{x})^2}$$

die **geschätzte Standardabweichung** des dem Signal x zugrunde liegenden Prozesses.

Normierungsfaktor

Die Definitionen 1.7 und 1.8 unterscheiden sich von den Definitionen 1.5 und 1.6 nur durch einen angepassten Normierungsfaktor ($\frac{1}{N-1}$), der insbesondere dann zum Tragen kommt, wenn das Signal nur wenige Abtastwerte (N) enthält. Die Verwendung dieser angepassten Formeln für Varianz und Standardabweichung sollten immer dann zum Einsatz kommen, wenn es darum geht, Varianz oder Standardabweichung des zugrunde liegenden Prozesses abzuschätzen.

Histogramm, Wahrscheinlichkeits- und Dichtefunktion

Histogramme

Einfache statistische Maßzahlen, wie das zuvor beschriebene arithmetische Mittel oder die Varianz, charakterisieren Signale häufig nur unzureichend. Eine sehr verbreitete Methode, Signale genauer zu beschreiben, sind sogenannte *Histogramme*. Das Histogramm eines Signals beschreibt für jede mögliche Ausprägung der Abtastwerte, wie häufig diese im Signal vorkommt. Die *Ausprägung* eines Abtastwertes bezeichnet ein Element des Wertebereiches der Abtastwerte. Handelt es sich beispielsweise um ein Signal, das mit einem 8-Bit Analog-Digital-Wandler erzeugt wurde, so gibt es $2^8 (= 256)$ verschiedene Ausprägungen, die die Abtastwerte des Signals einnehmen können, z.B. aus dem Wertebereich $\{-128, \dots, 127\}$ oder dem Wertebereich $\{0, \dots, 255\}$. Grundsätzlich könnte sich der Wertebereich der Abtastwerte aber auch auf Elemente anderer Art beziehen. Ist das „Signal“ beispielsweise ein Text, so würde der Wertebereich die verschiedenen Buchstaben und Satzzeichen der jeweiligen Sprache umfassen. Das zugehörige Histogramm würde dementsprechend die Häufigkeit der einzelnen Zeichen des Textes beschreiben¹⁴ (s. Abb. 1.8). Unabhängig von der Art der

Häufigkeit der Ausprägungen

¹⁴Derartige Histogramme finden beispielsweise in der Kryptographie Anwendung.

Elemente muss jeder Wertebereich W jedoch endlich und abzählbar sein, um ein Histogramm über W erstellen zu können. Daher kann immer eine Kodierung $\varphi = W \rightarrow \{0, \dots, M - 1\}$ mit $M \in \mathbb{N}$ definiert werden, die die Elemente des jeweiligen Wertebereiches W auf eine Teilmenge der natürlichen Zahlen abbildet. Dementsprechend kann o.B.d.A.¹⁵ ein Histogramm auf folgende Weise definiert werden:

Definition 1.9 (Histogramm).

Gegeben sei das Signal $x = (x_0, x_1, \dots, x_{N-1})$ mit $x_i \in \{0, \dots, M - 1\}$ bestehend aus N Abtastwerten. Dann ist

$$H^x = (H_0^x, H_1^x, \dots, H_{M-1}^x)$$

mit

$$H_j^x = \sum_{i=0}^{N-1} \delta[j - x_i], \quad j = 0, \dots, M - 1$$

und

$$\delta[n] = \begin{cases} 1 & \text{wenn gilt } n = 0, \\ 0 & \text{wenn gilt } n \neq 0, \quad n \in \mathbb{Z}. \end{cases}$$

das **Histogramm** des Signals x .

Die Funktion $\delta[n]$ ist eine Variante des Kronecker-Deltas δ_{ij} , die häufig in der Signalverarbeitung eingesetzt wird und über den ganzen Zahlen definiert ist^a.

^asiehe auch Abb. 1.20 und Def. 1.23

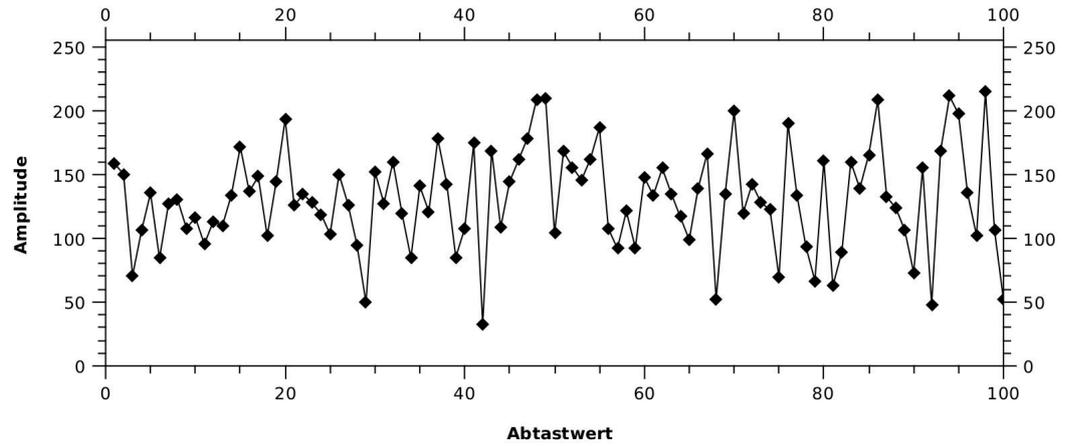


Abbildung 1.7a zeigt 100 Abtastwerte eines Signals mit Werten aus dem Wertebereich $\{0, \dots, 255\}$. Das zu diesem Signal gehörige Histogramm zeigt Abbildung 1.7b. Dem Histogramm kann entnommen werden, dass beispielsweise der Abtastwert 51 genau einmal, der Wert 84 zweimal und der Wert 127 genau dreimal in dem Signal vorkommt. Neben dieser rein quantitativen Information über die einzelnen Ausprägungen der Abtastwerte, vermittelt die „Form“ eines Histogramms einen qualitativen, charakteristischen Eindruck von der Art, wie die Abtastwerte des betrachteten Signals verteilt sind. Diese „Form“ des Histogramms wird auch als *Verteilung* bezeichnet. Da das zugrunde liegende Signal des Histogramms in Abbildung 1.7b nur 100 Abtastwerte umfasst, ist der Anteil zufälliger Abweichungen relativ hoch und die Verteilung der Abtastwerte ist nur schemenhaft zu erkennen. Wie auch beim zuvor beschriebenen arithmetischen Mittel verringert sich der relative Anteil der Zufallsfehler mit einer zunehmenden Anzahl der Abtastwerte eines Signals. Das in Abbildung 1.7c dargestellte Histogramm eines Signals mit 100.000 Abtastwerten illustriert diesen Effekt. Die Verteilung der Abtastwerte tritt wesentlich deutlicher hervor und zeigt, dass in diesem Fall das zugrunde liegende Signal einer sogenannten *Normalverteilung*

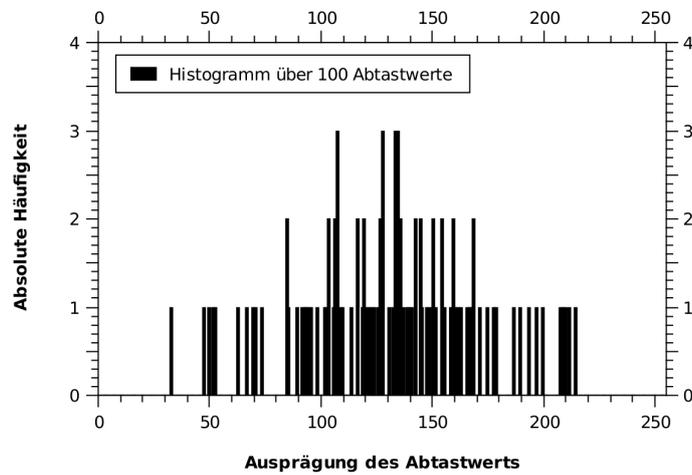
Form eines Histogramms
→ Verteilung

Normalverteilung

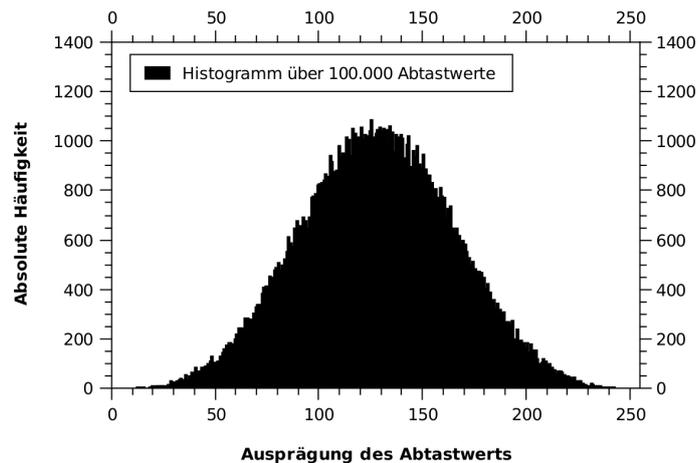
¹⁵o.B.d.A. → ohne Beschränkung der Allgemeinheit



(a)



(b)



(c)

Abbildung 1.7: Signal eines Zufallsprozesses und zugehörige Histogramme. (a) Die ersten 100 Abtastwerte eines Signals mit 100.000 Abtastwerten aus dem Wertebereich $\{0, \dots, 255\}$. (b) Histogramm des in (a) dargestellten Teilsignals. (c) Histogramm des vollständigen Signals mit 100.000 Abtastwerten.

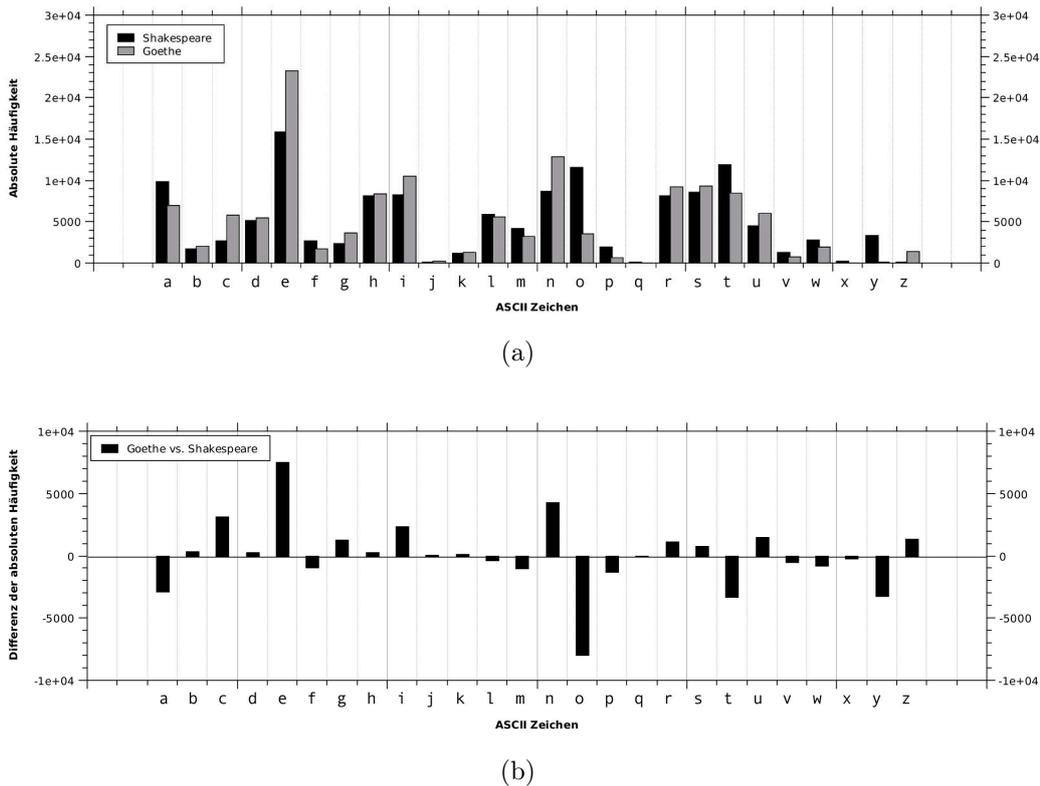


Abbildung 1.8: Histogramme von Texten. **(a)** Histogramme über zwei in ASCII kodierte Texte von Goethe und Shakespeare. **(b)** Differenz der in (a) dargestellten Histogramme.

folgt. Es existieren zahlreiche charakteristische Verteilungen dieser Art¹⁶, die in vielen Fällen Rückschlüsse auf die Prozesse erlauben, die den jeweiligen Signalen zugrunde liegen. So könnte es sich im Falle einer Normalverteilung um einen Zufallsprozess handeln¹⁷.

Neben dem Vergleich mit bekannten Wahrscheinlichkeitsverteilungen und der damit verbundenen Ableitung von Eigenschaften bzgl. der zugrunde liegenden Prozesse, können Histogramme auch für den direkten Vergleich mehrerer Signale verwendet werden. Ein Beispiel hierfür zeigt Abbildung 1.8. Teilbild (a) stellt die Histogramme zweier Texte – *Hamlet* von William Shakespeare und *Faust II* von Johann Wolfgang von Goethe – im direkten Vergleich dar¹⁸. Auf den ersten Blick erscheinen beide Histogramme recht ähnlich, mit jeweils drei dominanten Bereichen a bis i, l bis o und r bis u. Bildet man jedoch die elementweise Differenz der beiden Histogramme, wie im Teilbild (b) dargestellt, so treten die Unterschiede deutlich hervor. Die Unterschiede sind darauf zurückzuführen, dass der Text von Shakespeare in englischer Sprache und der Text von Goethe in deutscher Sprache verfasst ist. Insbesondere die Elemente $\{a, c, e, n, o, t, y\}$ er-

direkter
Signalvergleich

¹⁶Für eine Liste derartiger Verteilungen siehe z.B. den Artikel zur [Wahrscheinlichkeitsverteilung](#) auf Wikipedia.

¹⁷In der Tat wurde das Signal durch die additive Überlagerung von vier Pseudozufallsfolgen generiert.

¹⁸Die Darstellung ist beschränkt auf die Kleinbuchstaben kodiert im ASCII-Standard (American Standard Code for Information Interchange).

scheinen also geeignet, englische und deutsche Texte voneinander unterscheiden zu können. Eine derartige Analyse der Buchstabenhäufigkeit kann beispielsweise in der Kryptographie bei der Entschlüsselung sogenannter monoalphabetischer Chiffren¹⁹ eingesetzt werden.

praktischer Aspekt

Ein praktischer Aspekt von Histogrammen besteht in ihrer Verwendung zur effizienten Berechnung von arithmetischem Mittel und Varianz des zugehörigen Signals, insbesondere dann, wenn das Signal eine hohe Anzahl von Abtastwerten aufweist. Mithilfe des Histogramms lassen sich arithmetisches Mittel \bar{x} und Varianz σ^2 wie folgt berechnen:

effiziente
Berechnung von
arithmetischem
Mittel und Varianz

$$\bar{x} = \frac{1}{N} \sum_{i=0}^{M-1} iH_i$$

$$\sigma^2 = \frac{1}{N} \sum_{i=0}^{M-1} (i - \bar{x})^2 H_i$$

Durch die Verwendung des Histogramms wird vor allem die Berechnung der Varianz beschleunigt, da nur noch M -mal eine Quadrierung durchgeführt werden muss statt N -mal (s. Def. 1.5). Bei einem Signal von beispielsweise 100.000 Abtastwerten aus einem Wertebereich von $\{0, \dots, 255\}$ müssen also ca. 390-mal weniger Quadrierungen durchgeführt werden im Vergleich zu der in Definition 1.5 beschriebenen *klassischen* Methode.

absolute Häufigkeit

Wie bereits bei der Betrachtung von arithmetischem Mittel und Varianz ist die Unterscheidung zwischen Signal und zugrunde liegendem Prozess auch in Bezug auf Histogramme von großer Bedeutung. Ein Histogramm beschreibt immer nur die Verteilung von Abtastwerten eines diskreten, endlichen Signals. Betrachtet man erneut die beiden in Abbildung 1.8a dargestellten Histogramme der Texte von Shakespeare und Goethe, so fällt auf, dass die abhängige Variable als *absolute Häufigkeit* beschrieben wird. Gerade wenn, wie in diesem Beispiel, zwei Histogramme miteinander verglichen werden sollen, stellt diese Art der Darstellung ein Problem dar, da man den Histogrammen nicht ansehen kann, ob sie aus zwei Signalen mit einer gleichen Anzahl von Abtastwerten generiert wurden²⁰. Wäre zum Beispiel der dem Goethe-Histogramm zugrunde liegende Textabschnitt doppelt so lang wie der Textabschnitt, der für das Shakespeare-Histogramm verwendet wurde, so wären die einzelnen Histogrammwerte nicht direkt miteinander vergleichbar. Die Werte des Goethe-Histogramms wären im Schnitt doppelt so hoch wie die des Shakespeare-Histogramms und eine elementweise Differenz wie in Abbildung 1.8b dargestellt, würde zu vollkommen anderen und möglicherweise inkorrekten Schlussfolgerungen führen.

Histogramm-
normalisierung

Um Histogramme unabhängig von der Länge der jeweils zugrunde liegenden Signale vergleichen zu können, müssen die Histogramme vor dem Vergleich normalisiert werden. Hierbei ist die im folgenden Merksatz formulierte Feststellung hilfreich:

¹⁹Ein Beispiel für eine monoalphabetische Chiffre ist eine einfache Permutation der Buchstaben über eine feste Verschiebung ($a \rightarrow b, b \rightarrow c, \dots, z \rightarrow a$)

²⁰In diesem Beispiel wurde dies natürlich sichergestellt.

Merksatz 1.1 (Summe der Histogrammwerte).

Gegeben sei das Signal $x = (x_0, x_1, \dots, x_{N-1})$ mit $x_i \in \{0, \dots, M-1\}$ bestehend aus N Abtastwerten und sei $H^x = (H_0^x, H_1^x, \dots, H_{M-1}^x)$ das Histogramm des Signals. Dann gilt

$$N = \sum_{j=0}^{M-1} H_j^x,$$

d. h. „die Summe der Histogrammwerte H_j^x entspricht der Anzahl der Abtastwerte N .“



Aus dieser Feststellung ergibt sich direkt die Definition eines *normalisierten Histogramms*:

Definition 1.10 (Normalisiertes Histogramm).

Gegeben sei das Signal $x = (x_0, x_1, \dots, x_{N-1})$ mit $x_i \in \{0, \dots, M-1\}$ bestehend aus N Abtastwerten. Dann ist

$$\hat{H}^x = (\hat{H}_0^x, \hat{H}_1^x, \dots, \hat{H}_{M-1}^x)$$

mit

$$\hat{H}_j^x = \frac{1}{N} \sum_{i=0}^{N-1} \delta[j - x_i], \quad j = 0, \dots, M-1$$

und

$$\delta[n] = \begin{cases} 1 & \text{wenn gilt } n = 0, \\ 0 & \text{wenn gilt } n \neq 0, \quad n \in \mathbb{Z}. \end{cases}$$

das **normalisierte Histogramm** des Signals x .



In einem auf diese Weise normalisierten Histogramm nehmen die einzelnen Histogrammeinträge \hat{H}_j^x Werte aus dem reellen Intervall $[0, 1]$ an und die Summe aller Histogrammeinträge ergibt genau 1 unabhängig von der Länge des zugrunde liegenden Signals. Dementsprechend können normalisierte Histogramme direkt miteinander verglichen werden. Darüber hinaus approximiert ein normalisiertes Histogramm \hat{H}^x eines Signals x die *Wahrscheinlichkeitsfunktion* des Prozesses, der dem Signal zugrunde liegt. In diesem Zusammenhang wird der Prozess als diskrete Zufallsvariable modelliert, die genau M mögliche Ausprägungen besitzt. Die Wahrscheinlichkeitsfunktion ordnet jeder dieser Ausprägungen eine Wahrscheinlichkeit zu, mit der die Zufallsvariable die jeweilige Ausprägung annimmt:

Summe der
Histogrammwerte

Approximation
der Wahrscheinlich-
keitsfunktion


Definition 1.11 (Wahrscheinlichkeitsfunktion).

Es sei X eine diskrete Zufallsvariable, die eine abzählbare Menge $C = \{h_0, h_1, \dots\}$ an Ausprägungen besitzt. Jeder dieser Ausprägungen h_j kann eine Wahrscheinlichkeit $P(X=h_j) = p_j \in [0, 1]$ zugeordnet werden, mit der die Zufallsvariable X die Ausprägung h_j annimmt.

Die **Wahrscheinlichkeitsfunktion** ist dann gegeben durch eine Funktion $f : C \rightarrow [0, 1]$ mit

$$f(h) = P(X=h) \quad \text{und} \quad \sum_{h \in C} f(h) = 1.$$

Histogramm vs.
Wahrscheinlich-
keitsfunktion

Das Verhältnis von Wahrscheinlichkeitsfunktion und normalem Histogramm entspricht dem zuvor beschriebenen Verhältnis von arithmetischem Mittel des Münzwurfprozesses und arithmetischem Mittel des konkreten Münzwurfexperimentes. Genau wie das durch ein Münzwurfexperiment bestimmte arithmetische Mittel enthält auch das normalisierte Histogramm Zufallsfehler, die es von der Wahrscheinlichkeitsfunktion des zugrunde liegenden Prozesses abweichen lassen. Die Größe des Zufallsfehlers ist auch in diesem Fall direkt von der Anzahl N der Abtastwerte des Signals abhängig, über das das Histogramm gebildet wurde.

Dichtefunktion

Während die Wahrscheinlichkeitsfunktion für eine *diskrete* Zufallsvariable definiert ist, ist die Wahrscheinlichkeitsdichtefunktion oder auch nur *Dichtefunktion* das Analogon für eine *kontinuierliche* Zufallsvariable:


Definition 1.12 (Dichtefunktion).

Es sei X eine kontinuierliche, reellwertige Zufallsvariable.

Die **Dichtefunktion** der Zufallsvariablen X ist dann gegeben durch eine Funktion $f : \mathbb{R} \rightarrow [0, 1]$ mit

$$P(a \leq X \leq b) = \int_a^b f(x) dx \quad \text{und} \quad \int_{-\infty}^{+\infty} f(x) dx = 1.$$

Wahrscheinlich-
keitsdichte vs.
Wahrscheinlichkeit

Ein Beispiel für die Dichtefunktion f einer reellwertigen Zufallsvariablen X zeigt Abbildung 1.9. Man beachte die Auszeichnung der abhängigen Variablen als *Wahrscheinlichkeitsdichte* und nicht einfach nur als *Wahrscheinlichkeit*. Auf den ersten Blick könnte man die dargestellte Funktion derart interpretieren, dass z.B. die Zufallsvariable den Wert 105 in etwa 1% aller Fälle²¹ annimmt. Dies ist jedoch **nicht** der Fall. Ganz im Gegenteil: die Wahrscheinlichkeit dafür, dass die Zufallsvariable *exakt* den Wert 105.000000... annimmt ist null. Selbst in der kleinsten Umgebung um den Wert 105 herum gibt es bereits eine unendliche

²¹ $f(105) = 0.01$

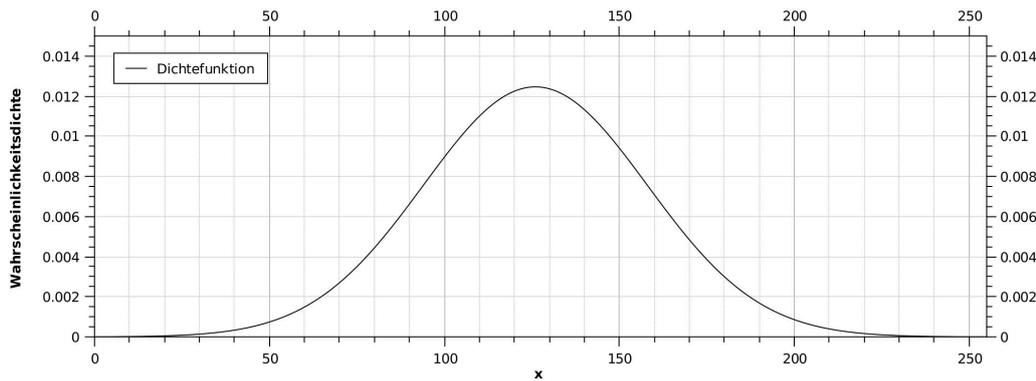


Abbildung 1.9: Beispiel für eine Dichtefunktion. Die hier dargestellte Dichtefunktion ist eine Normalverteilung mit einem arithmetischen Mittel von 127 und einer Standardabweichung von 32.

Anzahl an weiteren Werten, die eben nicht *exakt* den Wert 105 haben. Aus dieser Perspektive ist es einleuchtend, dass es „äußerst unwahrscheinlich“ genau auf den Wert 105 zu treffen. Aus diesem Grund ist es im Falle einer reellwertigen Zufallsvariablen nicht sinnvoll, nach der Wahrscheinlichkeit für das Auftreten eines einzelnen Wertes zu fragen. Vielmehr stellt sich die Frage, wie hoch die Wahrscheinlichkeit ist, dass die Zufallsvariable einen Wert aus einem gegebenen Intervall annimmt. D.h., um die Wahrscheinlichkeit für das Auftreten eines Wertes aus einem Intervall $[a, b]$ zu berechnen, muss die Dichtefunktion f über diesem Intervall integriert werden:

Intervall

$$\int_a^b f(x) dx = P(a \leq X \leq b)$$

Das Integral über dem Intervall $[a, b]$ kann mit

$$\int_a^b f(x) dx = F(b) - F(a)$$

berechnet werden, falls die zur Dichtefunktion f zugehörige *kumulative Verteilungsfunktion* F mit

kumulative Verteilungsfunktion

$$F(c) = P(X \leq c) = \int_{-\infty}^c f(x) dx$$

bekannt ist.

Wie die oben beschriebene Wahrscheinlichkeitsfunktion kann auch die Dichtefunktion einen Prozess beschreiben, der einem Signal zugrunde liegt. Im Gegensatz zu den bisher betrachteten Signalen entstammen die Abtastwerte in diesem Fall jedoch keinem endlichen Wertebereich $\{0, \dots, M - 1\}$, sondern stammen anstelle dessen aus einem kontinuierlichen Intervall $[a, b]$. Um für ein derartiges Signal ein Histogramm erstellen zu können, wird der Wertebereich $[a, b]$ des Signals in M Teilintervalle partitioniert und alle Abtastwerte, deren Werte in ein einzelnes Teilintervall fallen, werden dem entsprechenden

endlicher Wertebereich vs. kontinuierliches Intervall

Binning Histogrammwert zugeordnet. Im englischen wird dieser Vorgang als *Binning*²² bezeichnet:



Definition 1.13 (Intervall-Histogramm).

Gegeben sei das Signal $x = (x_0, x_1, \dots, x_{N-1})$ mit $x_i \in [a, b)$ bestehend aus N Abtastwerten und sei $a = t_0 < t_1 < \dots < t_M = b$ eine Partitionierung des Wertebereichs $[a, b)$ in M Teilintervalle. Dann ist

$$H^x = (H_0^x, H_1^x, \dots, H_{M-1}^x)$$

mit

$$H_j^x = \sum_{i=0}^{N-1} \Theta(x_i - t_j) - \Theta(x_i - t_{j+1}), \quad j = 0, \dots, M-1$$

und

$$\Theta(n) = \begin{cases} 1 & \text{wenn gilt } n \geq 0, \\ 0 & \text{wenn gilt } n < 0, \end{cases} \quad n \in \mathbb{R}.$$

das Histogramm des Signals x über die Teilintervalle $[t_i, t_{i+1})$, $i \in \{0, \dots, M-1\}$.

Die Funktion $\Theta(n)$ wird als Heaviside-Funktion bezeichnet und ist nach dem britischen Mathematiker und Physiker Oliver Heaviside (1850 - 1925) benannt.

Das Histogramm kann analog zur Definition 1.10 normalisiert werden.

Zusammenfassung
in Teilintervalle

Die Verwendung eines Histogramms über Teilintervalle des Wertebereichs kann auch im Falle eines ganzzahligen, endlichen Wertebereichs von Nutzen sein. Ist die Zahl der Abtastwerte im Vergleich zur Größe des Wertebereiches sehr klein, kann es sinnvoll sein, den Wertebereich in Teilintervalle zusammenzufassen. Würde zum Beispiel ein Signal von einem 16-Bit²³ Analog-Digital-Wandler aus 10.000 Abtastwerten bestehen, so wäre das resultierende Histogramm nur spärlich befüllt (ähnlich dem Histogramm in Abbildung 1.7b). Durch die Zusammenfassung der 2^{16} möglichen Ausprägungen auf z.B. 2^8 Intervalle ($0 < 256 < 512 < \dots < 65.536$) würde sich die durchschnittliche Anzahl der Abtastwerte, die einem Histogrammelement zugeordnet werden, um den Faktor 256 erhöhen und somit die Auflösung entlang der vertikalen Achse des Histogramms – zu Lasten der horizontalen Auflösung – deutlich verbessern.

Im vorhergehenden Abschnitt haben Sie die grundlegenden Beziehungen zwischen Signalen und den ihnen zugrunde liegenden Prozessen kennengelernt. Darüber hinaus wurden die wesentlichen Aspekte der statistischen Beschreibung von Signalen und der Approximation statistischer Eigenschaften der

²²dt. *Klasseneinteilung*

²³In diesem Fall könnten die Abtastwerte $2^{16} = 65.536$ verschiedene Werte annehmen.

zugehörigen Prozesse erläutert. Hierbei wurden insbesondere die Erstellung und wesentliche Eigenschaften von Histogrammen betrachtet, da diese in vielen Algorithmen und Verfahren, z.B. dem in Kurseinheit 3 beschriebenen *automatischen Histogrammausgleich*, der Mensch-Computer-Interaktion Verwendung finden.

1.2 Analog-Digital-Wandlung

Lernziele:

Der folgende Abschnitt wird Ihnen die wichtigsten Schritte und Elemente im Kontext der Analog-Digital-Wandlung vorstellen und erläutern. Der zugrunde liegende mathematische Hintergrund wird an dieser Stelle bewußt noch nicht behandelt und auf einen späteren Abschnitt verschoben. Ziel dieses Abschnittes ist die Vermittlung eines prinzipiellen Verständnisses des Prozesses der Analog-Digital-Wandlung.

Wie bereits in den bisherigen Beispielen angedeutet, haben die Signale, die untersucht und weiterverarbeitet werden sollen, häufig ihren Ursprung in der Beobachtung eines natürlichen Prozesses mittels geeigneter Sensoren. Diese wandeln physikalische Größen, z.B. die Lichtintensität oder die Temperatur, in elektrisch messbare Größen um. Im Wesentlichen können hierbei drei Arten der Repräsentation von Messwerten unterschieden werden:

Beobachtung
natürlicher
Prozesse mittels
Sensoren

- die Repräsentation von Messwerten über korrespondierende elektrische Spannungen,
- die Repräsentation von Messwerten über korrespondierende elektrische Ströme und
- die Repräsentation von Messwerten über eine korrespondierende elektrische Widerstandsveränderung.

Üblicherweise werden im Falle der Repräsentation der Messwerte über Ströme oder der Repräsentation über eine Widerstandsveränderung die Ausgaben des jeweiligen Sensors nachträglich in eine Repräsentation über die elektrische Spannung umgewandelt. Dies geschieht unter Zuhilfenahme des *Ohmschen Gesetzes*, welches den Zusammenhang von Spannungsabfall, Strom und Widerstand beschreibt:

Ohmsches Gesetz



Definition 1.14 (Ohmsches Gesetz).

Bezeichne U den Spannungsabfall in Volt, I die Stromstärke in Ampere und R einen ohmschen Widerstand^a.

Dann gilt der als **Ohmsches Gesetz** bezeichnete Zusammenhang:

$$U = R \cdot I$$

^aEin ohmscher Widerstand ist ein idealisierter elektrischer Widerstand, dessen Widerstandswert unabhängig von der jeweiligen Spannung, dem jeweiligen Strom und der jeweiligen Frequenz ist. Die Frequenz bezieht sich hierbei auf elektrische Signale deren Spannung periodisch zwischen zwei Extremwerten wechselt. Die Frequenz beschreibt in diesem Zusammenhang die Anzahl der Perioden pro Sekunde, die die Spannung durchläuft. Sie hat die Einheit *Hertz*.

Repräsentation von
Messwerten mittels
Spannungen

Für die weitere Betrachtung der Analog-Digital-Wandlung reicht es daher aus, nur den Fall der Repräsentation von Messwerten über korrespondierende Spannungen zu berücksichtigen. Als Beispiel soll das in Abbildung 1.10a dargestellte Analogsignal eines fiktiven Sensors dienen. Der Ausgangsspannungsbereich des Sensors soll von 0 Volt bis 10.23 Volt gehen und von einem 10-Bit Analog-Digital-Wandler (ADC) auf einen Wertebereich von 0 bis 1023 abgebildet werden. Wie bereits eingangs erwähnt, wird bei der Analog-Digital-Wandlung eines kontinuierlichen Signals sowohl die unabhängige als auch die abhängige Variable diskretisiert. Die Diskretisierung von unabhängiger und abhängiger Variable verläuft in einem sequenziellen Prozess und findet nicht gleichzeitig statt.

unabhängige
Variable

Abtastung

Im Rahmen der sogenannten *Abtastung* wird zunächst die unabhängige Variable – in diesem Beispiel die Zeit – diskretisiert. Zu diesem Zweck wird das analoge Signal einem sogenannten *Sample-And-Hold*-Schaltkreis zugeführt, der in regelmäßigen Abständen, festgelegt über die *Abtastrate*, den jeweils aktuellen, analogen Signalwert „einfriert“ (s. Abb. 1.10b). Die Veränderungen des analogen Eingangssignals während dieses *Sample-And-Hold*-Schrittes gehen hierbei verloren. Je geringer die Abtastrate ist, desto größer ist dementsprechend der potenzielle²⁴ Diskretisierungsfehler. Das nach der Abtastung entstandene Signal wird auch als *zeitdiskret* bezeichnet. Die einzelnen Werte des Signals sind hierbei weiterhin kontinuierlich.

²⁴Der real auftretende Diskretisierungsfehler ist vom jeweils konkreten Signal abhängig. So würde beispielsweise ein konstantes Signal unabhängig von der Abtastrate keinen Diskretisierungsfehler aufweisen.

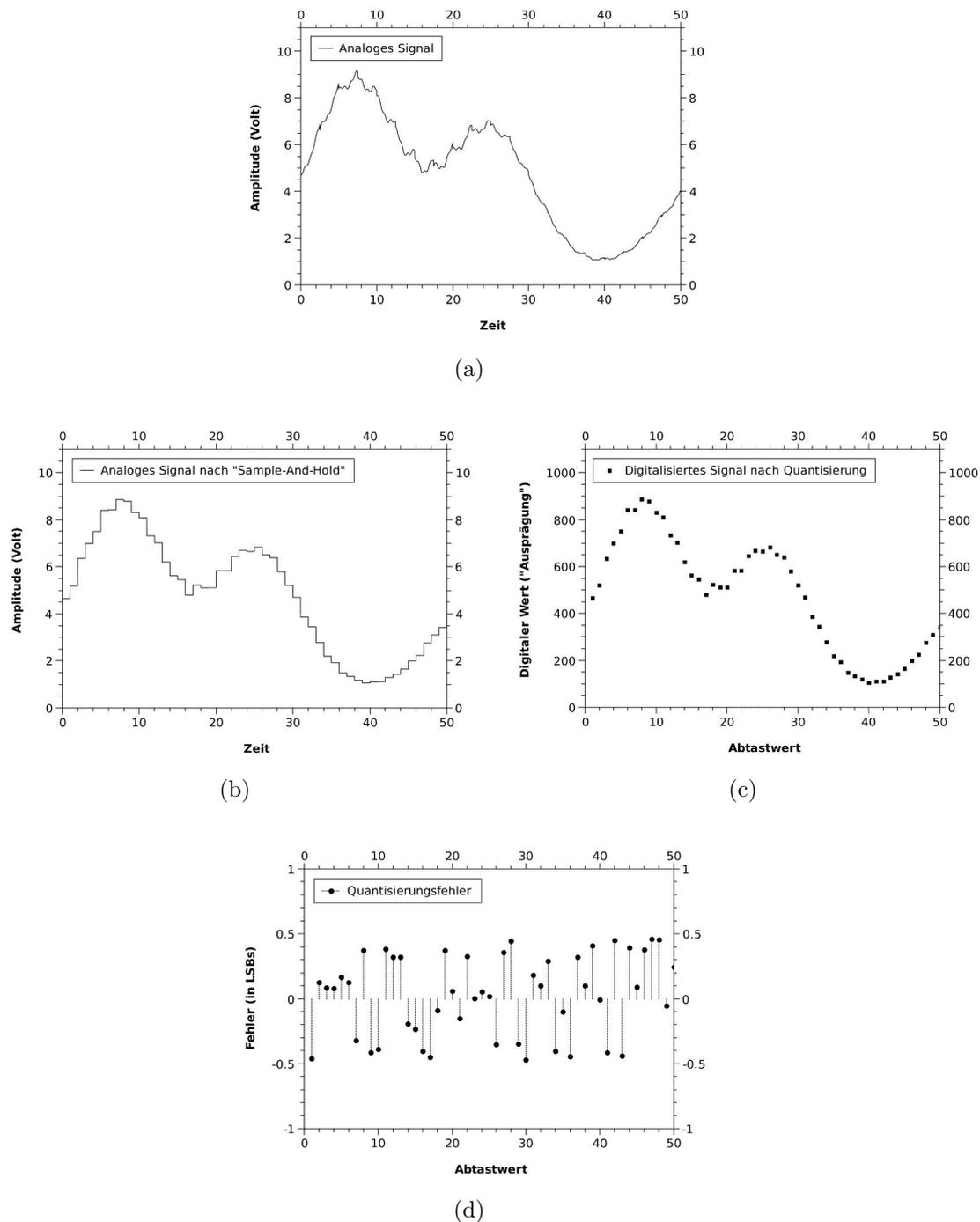


Abbildung 1.10: Schritte der Analog-Digital-Wandlung des Ausgangssignals eines Sensors. **(a)** Das analoge Spannungssignal am Ausgang des Sensors. **(b)** Das weiterhin analoge Spannungssignal am Ausgang des **Sample-And-Hold**-Schaltkreises (Diskretisierung der Zeit \rightarrow Abtastung). **(c)** Das digitale Signal am Ausgang des Analog-Digital-Wandlers (Diskretisierung der Spannungswerte \rightarrow Quantisierung). **(d)** Die Differenz der Signale (b) und (c) entspricht dem Quantisierungsfehler. Die Angabe des Fehlers erfolgt normiert als Anteil bzgl. des niederwertigsten Bits (siehe Text zur näheren Erläuterung).

Quantisierung

abhängige Variable Die Ausgabe des *Sample-And-Hold*-Schaltkreises wird anschließend dem eigentlichen Analog-Digital-Wandler (ADC) zugeführt, dessen Aufgabe in der Diskretisierung der abhängigen Variable – in diesem Beispiel der Spannung – besteht. Aufgrund der vorgeschalteten Abtastung steht dem ADC nun ein jeweils für einen Zeitschritt stabiles, weiterhin analoges Eingangssignal zur Verfügung. Die Umwandlung dieses analogen Eingangssignals auf einen digitalen Wert wird als *Quantisierung* bezeichnet. Die Anzahl der *Quantisierungsstufen*, also die Anzahl der möglichen Ausprägungen des digitalen Wertes, wird üblicherweise über die Bit-Breite des ADC angegeben. In diesem Beispiel beträgt die Bit-Breite des ADC 10-Bit und ermöglicht somit $2^{10} = 1024$ Quantisierungsstufen.

Quantisierungsstufen

Quantisierungsfehler

Abbildung 1.10c zeigt das digitalisierte Signal am Ausgang des ADC. Bildet man die Differenz dieses Signals mit dem Signal am Ausgang des *Sample-And-Hold*-Schaltkreises²⁵ (Abbildung 1.10b), so erhält man die sogenannten *Quantisierungsfehler*, die die Abweichung der digitalen Abtastwerte von ihren korrespondierenden analogen Werten beschreiben. Abbildung 1.10d zeigt die Quantisierungsfehler für das hier verwendete Beispiel. Die Angabe der Fehler erfolgt normiert als Anteil bezüglich des sogenannten „niederwertigsten Bits“ (LSB²⁶). Wie dieser normierte Fehlerwert zu interpretieren ist, wird durch ein einfaches Beispiel klar. Angenommen der analoge Eingangsspannungsbereich eines ADC liegt zwischen 0 Volt und 5.115 Volt und der digitale Wertebereich des Ausgangs ist $\{0, \dots, 1023\}$ gemäß einer Bit-Breite von 10-Bit. Mathematisch²⁷ besteht die Umwandlung eines Eingangswertes auf den entsprechenden Ausgangswert nur aus einer Multiplikation mit einem passend gewählten Konvertierungsfaktor (hier 200). Ein Eingangswert von beispielsweise 2.0105 Volt würde in diesem Fall einen Ausgangswert von 402.1 erfordern. Da der Wertebereich des Ausgangs jedoch ganzzahlig ist, muss der „ideale“ Wert 402.1 auf einen ganzzahligen „realen“ Wert von 402 gerundet werden. Der „fehlende“ Rest von -0.1 („realer“ Wert – „idealer“ Wert) ist der oben beschriebene Quantisierungsfehler bezüglich des niederwertigsten Bits, was jetzt noch näher erläutert wird.

Abweichung digitaler vs. analoger Werte

Betrachtet man die Quantisierungsfehler in Abbildung 1.10d genauer, so fällt auf, dass die Quantisierungsfehler wie ein zufälliges Rauschen mit einer Amplitude von $\pm \frac{1}{2} LSB$ erscheinen. Durch die Quantisierung des analogen Signals (b) wird diesem Signal also ein Rauschen hinzugefügt, dessen Amplitude von der verwendeten Bit-Breite abhängt, also von der Anzahl der für die Quantisierung zur Verfügung stehenden Bits. Im obigen Beispiel (Eingang: 0 – 10.23 Volt, Ausgang: 0..1023) beschreibt das LSB einen Bereich von $10.23/1023 = 0.01$ Volt. Das Rauschen, das durch die Quantisierung entsteht, hat also dementsprechend eine Amplitude von ± 0.005 Volt.

Rauschen mit $\pm \frac{1}{2} LSB$

²⁵unter Zuhilfenahme einer geeigneten Skalierung

²⁶LSB = least significant bit

²⁷In einem konkreten ADC findet die Konvertierung natürlich auf eine andere Weise statt.

Merksatz 1.2 (Rauschen durch Quantisierung).

Die Quantisierung eines analogen Signals fügt dem ursprünglichen Signal ein Rauschen mit Amplitude $\pm \frac{1}{2} \text{LSB}$ hinzu.

„Je geringer die Bit-Breite des eingesetzten Analog-Digital-Wandlers, desto größer ist das im Zuge der Quantisierung hinzugefügte Rauschen.“



Ist die Amplitude des „natürlichen“ Rauschens im analogen Signal bekannt, so kann die Bit-Breite der Quantisierung derart gewählt werden, dass das zusätzliche Rauschen aufgrund der Quantisierung im Verhältnis zum bereits im Signal vorhandenen Rauschen klein ist und hierdurch in diesem „untergeht“.

Da üblicherweise die Bit-Breite der Quantisierung durch die verfügbare Hardware auf wenige Auswahlmöglichkeiten beschränkt ist²⁸, muss ggf. der Eingangsspannungsbereich durch Verschiebung und Skalierung ebenfalls angepasst werden, um eine genauere Quantisierung der Eingangsdaten zu erreichen. Angenommen man habe einen 10-Bit Analog-Digital-Wandler (ADC), der einen Eingangsspannungsbereich von 0 Volt bis 10.23 Volt besitzt und einen Temperatursensor, der den Temperaturbereich von 0°C bis 100°C auf den Spannungsbereich von 0 Volt bis 10 Volt abbildet. In dieser Kombination wird demnach die Temperatur vom ADC in 0.1°C Schritten ausgegeben. Ist man jedoch nur an Messungen aus einem Bereich von z.B. 20°C bis 40°C interessiert, so kann man durch eine geeignete elektronische Schaltung den Ausgangsspannungsbereich des Temperatursensors durch Verschiebung und Skalierung auf -10 Volt bis 40 Volt verändern bevor dieser vom ADC – der weiterhin nur Werte zwischen 0 Volt und 10.23 Volt erfasst – umgewandelt wird. Hierdurch können zwar nur noch Werte von 20°C bis 40°C erfasst werden, dies jedoch nun mit einer Auflösung von 0.02°C, also mit einer fünf mal so großen Genauigkeit.

Anpassung des
Eingangsspannungsbereichs

Erhöhung der
Auflösung

Abtasttheorem

Im Gegensatz zur Frage der Bit-Breite der Quantisierung, ist die Festlegung einer „ausreichenden“²⁹ Abtastrate (auch *Abtastfrequenz* genannt) nicht *allgemeingültig* möglich. Beschränkt man sich jedoch auf die Klasse der kontinuierlichen und bandbeschränkten³⁰ Signale, so kann gezeigt werden, dass genau dann eine fehlerfreie Rekonstruktion des Ausgangssignals möglich ist, wenn die höchste im Signal auftretende Frequenz kleiner ist als die Hälfte der Abtastfrequenz.

Abtastfrequenz

Diese als *Abtasttheorem* bekannte Aussage wurde in der ersten Hälfte des 20. Jahrhunderts von mehreren Forschern unabhängig voneinander formuliert. Im Westen wurde das Abtasttheorem 1948 von Claude Shannon formuliert, der sich auf Arbeiten von Harry Nyquist und Edmund Taylor Whittaker stützte. Im

²⁸die meisten AD-Wandler haben 8-, 10-, 12- oder 16-Bit

²⁹Eine *ausreichende* Abtastrate kann dann als gegeben angenommen werden, wenn aus dem digitalisierten Signal das ursprüngliche Signal fehlerfrei rekonstruiert werden kann.

³⁰*bandbeschränkt* bedeutet, dass in einem Signal nur Frequenzen aus einem bestimmten, endlichen Bereich (dem Frequenzband) vorhanden sind.

russischen Sprachraum wurde das Abtasttheorem bereits 1933 von Wladimir Alexandrowitsch Kotelnikow eingeführt.

An dieser Stelle soll vorerst nur der Hinweis auf die Existenz des Abtasttheorems gegeben werden. In der nächsten Kurseinheit, die sich ausführlich mit der Fourier-Transformation und der Repräsentation von Signalen im Frequenzbereich befassen wird, wird die vom Abtasttheorem beschriebene Abhängigkeit von Abtastfrequenz und Bandbreite des Signals deutlich werden.

Der vorhergehende Abschnitt hat Ihnen kurz und knapp die wichtigsten Schritte und Elemente der Analog-Digital-Wandlung nähergebracht. Ihnen sollte der Prozess vom analogen Eingangssignal über die Abtastung und Quantisierung hin zum digitalen Ausgangssignal klar sein, und Sie sollten erläutern können, an welchen Stellen welche Stör- und Fehlerquellen innerhalb dieses Prozesses auftreten können (z.B. Quantisierungsfehler).

1.3 Lineare Systeme

Lernziele:

Das folgende Unterkapitel wird Ihnen die wesentlichen Eigenschaften linearer Systeme und die zentralen Ideen und Techniken der Analyse dieser Systeme erläutern. Das Konzept der *Linearität* ist eine der wichtigsten theoretischen Grundlagen für eine große Zahl der existierenden Methoden zur Signalverarbeitung. Sowohl das anschließende Unterkapitel über die „Diskrete Faltung“ (1.4), als auch die folgende Kurseinheit, die sich intensiv mit der „Fourier-Analyse“ auseinandersetzt, basieren auf der im Folgenden beschriebenen theoretischen Basis.

Die bisherigen Abschnitte dieser Kurseinheit haben sich einzig auf Signale (s. Def. 1.2) und Möglichkeiten ihrer Charakterisierung, z.B. durch Histogramme, konzentriert. In diesem Abschnitt sollen nun *Systeme* betrachtet werden. Es folgt die Definition des Begriffs *System*, wie er im weiteren Verlauf dieses Kurs verwendet wird.



Definition 1.15 (System).

Als **System** kann jedweder Prozess bezeichnet werden, der in Reaktion auf ein Eingangssignal ein Ausgangssignal erzeugt.

Abbildung 1.11 zeigt eine schematische Darstellung eines derartigen Systems. Die in der Abbildung dargestellten eckigen Parameterklammern des Eingangssignals x und des Ausgangssignals y sollen andeuten, dass es sich bei beiden



Abbildung 1.11: Schematische Darstellung eines Systems mit Eingangssignal $x[n]$ und Ausgangssignal $y[n]$.

Signalen um *diskrete* Signale handelt und dementsprechend insgesamt um ein diskretes System. Für die Darstellung von kontinuierlichen Signalen werden im weiteren Verlauf runde Parameterklammern verwendet ($\rightarrow x(t)$ kontinuierlich vs. $x[n]$ diskret). Darüber hinaus werden im Folgenden die Eingangssignale eines Systems mit x und die Ausgangssignale eines Systems mit y bezeichnet.

Die Untersuchung von Systemen ist über ein weites Spektrum praktischer Anwendungsfälle motivierbar, sei es die Entstörung sowohl drahtgebundener als auch drahtloser Kommunikationskanäle (z.B. DSL³¹-Leitungen, Mobilfunk, Satellitennavigation etc.), Teile des Bereichs der Bildverarbeitung (z.B. nachträgliches Hinzufügen von Bildschärfe) oder auch der Equalizer an einer Stereoanlage – bei all diesen Prozessen handelt es sich um Systeme die Eingangssignale auf eine spezifische Art und Weise verändern und entsprechend modifizierte Ausgangssignale erzeugen. Hierbei muss es sich bei dem jeweils untersuchten System nicht zwingend um einen technischen Prozess handeln. Auch die Untersuchung natürlicher Phänomene über das Aussenden und Empfangen eines Testsignals fallen unter die Definition 1.15. Ein Beispiel hierfür sind seismische Untersuchungen im Rahmen der Geologie, bei denen Stoßwellen in den Boden eingebracht werden und die Reflektion dieser Wellen mittels Geophone wieder aufgenommen werden. Das zu untersuchende System ist in diesem Fall der jeweils vorherrschende, geologische Aufbau.

praktische
Anwendungsfälle

natürliche
Phänomene

Trotz dieser Vielfalt an verschiedenen Systemen lässt sich der überwiegende Teil dieser Systeme überraschenderweise einer einzelnen Klasse – den *linearen Systemen* – zuordnen:

³¹Das Kürzel *DSL* steht für den Ausdruck *Digital Subscriber Line*, engl. für *Digitaler Teilnehmeranschluss*


Definition 1.16 (Lineares System).

Ein System S wird als **lineares System** bezeichnet, wenn es die folgenden zwei Bedingungen erfüllt:

- S ist homogen:

$$\begin{array}{l} \text{Aus} \quad x[n] \rightarrow \boxed{S} \rightarrow y[n] \\ \text{folgt} \quad k \cdot x[n] \rightarrow \boxed{S} \rightarrow k \cdot y[n]. \end{array}$$

- S ist additiv:

$$\begin{array}{l} \text{Aus} \quad x_1[n] \rightarrow \boxed{S} \rightarrow y_1[n] \\ \text{und} \quad x_2[n] \rightarrow \boxed{S} \rightarrow y_2[n] \\ \text{folgt} \quad x_1[n] + x_2[n] \rightarrow \boxed{S} \rightarrow y_1[n] + y_2[n]. \end{array}$$

Die Bedingung der Homogenität besagt, dass eine Skalierung der Amplitude des Eingangssignals $x[n]$ mit einem konstanten Wert k zu einer Skalierung der Amplitude des Ausgangssignals $y[n]$ um den gleichen Faktor führt.

Die Bedingung der Additivität besagt, dass es keinen Unterschied macht, ob man zwei Signale $x_1[n]$ und $x_2[n]$ separat durch ein System schickt und die Ausgangssignale anschließend addiert, oder ob man die zwei Signale $x_1[n]$ und $x_2[n]$ zuerst addiert und erst dann die Summe der Signale durch das System verarbeiten lässt. In beiden Fällen erhält man das gleiche Ausgangssignal. Anders ausgedrückt: Die Summe mehrerer Signale durchläuft ein lineares System ohne dass sich die einzelnen Signale dabei beeinflussen.

keine gegenseitige
Beeinflussung

Neben diesen zwei Bedingungen gibt es noch eine dritte Eigenschaft, die bei fast allen linearen Systemen anzutreffen ist. Es handelt sich hierbei um die sogenannte *Verschiebungsinvarianz*:


Definition 1.17 (Verschiebungsinvarianz).

Ein System S heißt **verschiebungsinvariant**, wenn gilt:

$$\begin{array}{l} \text{Aus} \quad x[n] \rightarrow \boxed{S} \rightarrow y[n] \\ \text{folgt} \quad x[n + s] \rightarrow \boxed{S} \rightarrow y[n + s]. \end{array}$$

Die Eigenschaft der Verschiebungsinvarianz³² besagt, dass das System für ein bestimmtes Eingabemuster immer das gleiche Ausgabemuster erzeugt, unabhängig davon, wann das Eingabemuster eingegeben wurde bzw. welche anderen

zeitunabhängig

³²Manchmal wird die Verschiebungsinvarianz auch als *Zeitinvarianz* bezeichnet.

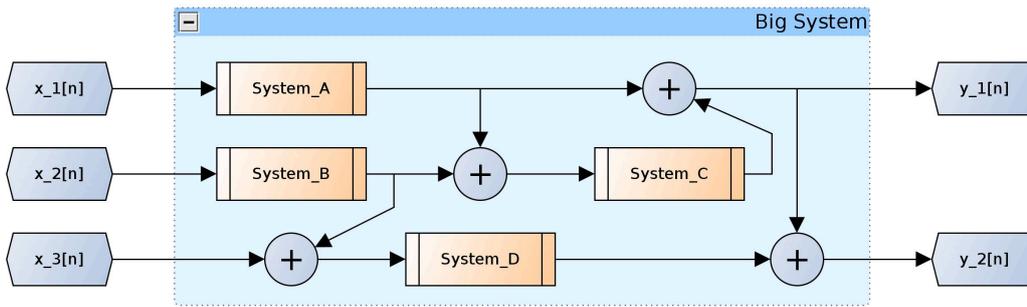


Abbildung 1.12: Schematische Darstellung eines zusammengesetzten, linearen Systems.

Muster zuvor eingegeben wurden. Häufig wird mit dieser Eigenschaft auch ein Idealzustand beschrieben, der von einem realen System nur approximiert wird. So könnte das System z.B. ein Verstärker sein, der die Amplitude des Eingangssignals im Idealfall verdoppeln soll, jedoch aufgrund seiner Bauteile eine leichte Temperaturabhängigkeit aufweist und bei Temperaturen um 0°C einen realen Verstärkungsfaktor von 1.98 und bei Temperaturen um 50°C einen realen Verstärkungsfaktor von 2.03 hat. Bei aufwendig gestalteten Systemen sind derartige Effekte meist bekannt und werden durch entsprechende Ausgleichsmechanismen korrigiert. Darüber hinaus werden die linearen Eigenschaften eines realen Systems meist nur über einem bestimmten *Arbeitsbereich* garantiert. Außerhalb dieses Arbeitsbereiches kann sich das System dann durchaus nicht-linear verhalten. Ein klassisches Beispiel für ein System mit einem eingeschränkten, linearen Arbeitsbereich ist der Transistor.

Arbeitsbereich

Aus der Linearität eines Systems lassen sich noch weitere grundsätzliche Eigenschaften ableiten. So sind lineare Systeme *kommutativ*, d.h. es gilt:

$$\begin{aligned} \text{aus } x[n] &\rightarrow \boxed{\boxed{A} \rightarrow \boxed{B}} \rightarrow y[n] \\ \text{folgt } x[n] &\rightarrow \boxed{\boxed{B} \rightarrow \boxed{A}} \rightarrow y[n] \end{aligned}$$

Kommutativität

Bei einer sogenannten *Kaskade*³³ von linearen Systemen, spielt es demnach keine Rolle, in welcher Reihenfolge ein Eingangssignal die Systeme durchläuft. Trotz dieser Eigenschaft kann es aufgrund der bereits oben erwähnten nicht-idealen Eigenschaften realer Systeme nötig sein, die Reihenfolge der Systeme zu optimieren, um die Einflüsse dieser nicht-idealen Eigenschaften auf das Gesamtsystem zu minimieren.

Kaskade

Über die einfache Hintereinanderschaltung von Systemen hinaus, können Signale innerhalb eines zusammengesetzten Systems auch addiert werden, ohne dass die Eigenschaft der Linearität dabei verloren geht. Ein Beispiel für ein solches System zeigt Abbildung 1.12. Das dargestellte Gesamtsystem hat mehrere Eingänge und besteht intern nur aus linearen Systemen und der Addition

zusammengesetzte Systeme

³³Hintereinanderschaltung

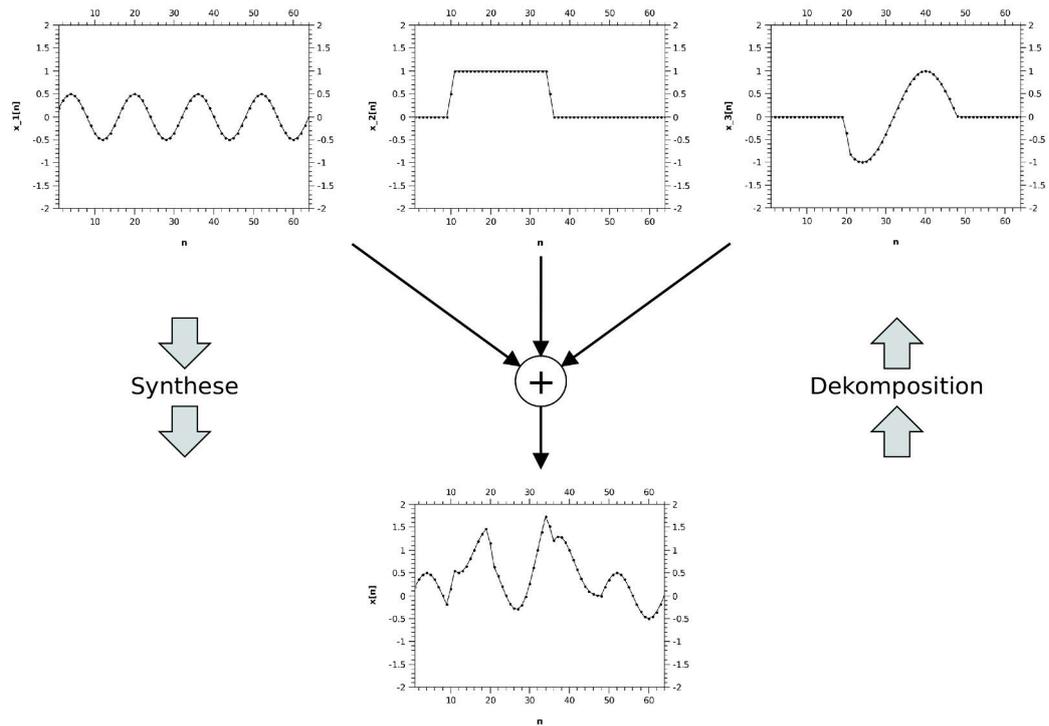


Abbildung 1.13: Schematische Darstellung der Synthese und Dekomposition eines Signals aus drei Ausgangssignalen.

von Signalen. Damit ist das Gesamtsystem ebenfalls linear. Würde anstelle der Addition auch eine Multiplikation von Signalen vorkommen, so würde das Gesamtsystem seine Linearität verlieren. Man beachte, dass diese Aussage für die Multiplikation mit einer Konstanten nicht gilt (s. Bedingung der Homogenität).

1.3.1 Superpositionsprinzip

Kombination
linearer Systeme

Ausgehend von dem in Abbildung 1.12 gezeigten linearen System stellt sich die Frage, wie in einem komplexen linearen System Signale miteinander kombiniert werden können, so dass das System seine linearen Eigenschaften behält. Die Antwort hierfür lässt sich direkt aus den zwei Bedingungen (Homogenität und Additivität) für lineare Systeme ableiten:



Merksatz 1.3 (Kombination von Signalen).

Innerhalb eines linearen Systems können Signale nur durch Skalierung^a und Addition miteinander kombiniert werden.

^amit einem konstanten Wert

Jede andere Art der Kombination von Signalen führt zum Verlust der Linearität. Eine Kombination von Signalen auf diese Weise wird auch als *Synthese* bezeichnet. Der umgekehrte Prozess, die Zerlegung eines Signals in mehrere

additive Komponenten, wird als *Dekomposition* bezeichnet. Abbildung 1.13 stellt diesen Zusammenhang noch einmal schematisch dar.

Synthese &
Dekomposition

Der Prozess der Dekomposition ist im Gegensatz zur Synthese nicht eindeutig und dementsprechend im Allgemeinen schwieriger zu handhaben. Während die Summe (Synthese) einiger Zahlen, z.B. 11, 13 und 17, immer – unabhängig von der Reihenfolge der Operationen – ein Ergebnis ($11+13+17=41$) hervorbringt, hat die Dekomposition einer Zahl in additive Komponenten eine sehr hohe Zahl möglicher Ergebnisse (z.B. $41=40+1$, $41=20+10+11$ etc.).

mehrere
Möglichkeiten der
Dekomposition

Der oben beschriebene Prozess der Synthese ist auch als *Superpositionsprinzip* bekannt und fasst die beiden Bedingungen für lineare Systeme in eine Bedingung zusammen:

Definition 1.18 (Superpositionsprinzip).

Für ein lineares System S gilt das **Superpositionsprinzip**:

$$\begin{array}{l} \text{Aus} \quad \quad \quad x_1[n] \rightarrow \boxed{S} \rightarrow y_1[n] \\ \text{und} \quad \quad \quad x_2[n] \rightarrow \boxed{S} \rightarrow y_2[n] \\ \text{folgt} \quad c_1 \cdot x_1[n] + c_2 \cdot x_2[n] \rightarrow \boxed{S} \rightarrow c_1 \cdot y_1[n] + c_2 \cdot y_2[n]. \end{array}$$

mit c_1, c_2 zwei beliebigen Konstanten.



Zusammen mit dem Prozess der Dekomposition ist das Superpositionsprinzip der zentrale Aspekt für das Verständnis der Analyse von Signalen und Systemen. Angenommen es soll die Veränderung eines komplexen Signals $x[n]$ durch ein System S untersucht werden. Das Superpositionsprinzip erlaubt es nun (s. Abb. 1.14), das Eingangssignal $x[n]$ in eine Menge einfacher Komponenten $\{x_0[n], x_1[n], \dots, x_{M-1}[n]\}$ aufzuteilen (Dekomposition), die einzelnen Komponenten vom System S in Komponenten des Ausgangssignals $\{y_0[n], y_1[n], \dots, y_{M-1}[n]\}$ überführen zu lassen, und schließlich durch Addition der Komponenten $y_i[n]$ das endgültige Ausgangssignal $y[n]$ zu erzeugen (Synthese).

zentraler Aspekt

Das Superpositionsprinzip garantiert in diesem Fall, dass das auf diese Weise erzeugte Ausgangssignal $y[n]$ identisch mit dem Ausgangssignal ist, das bei der direkten Verarbeitung des Eingangssignals $x[n]$ durch das System S entstanden wäre. Dank der Aufteilung des komplexen Eingangssignals in einfache Komponenten reicht es also aus, zu wissen, wie das System diese einfachen Komponenten verändert, um die Veränderung des komplexen Signals durch das System bestimmen und analysieren zu können.

Analyse mittels
Zerlegung in
einfache Signale

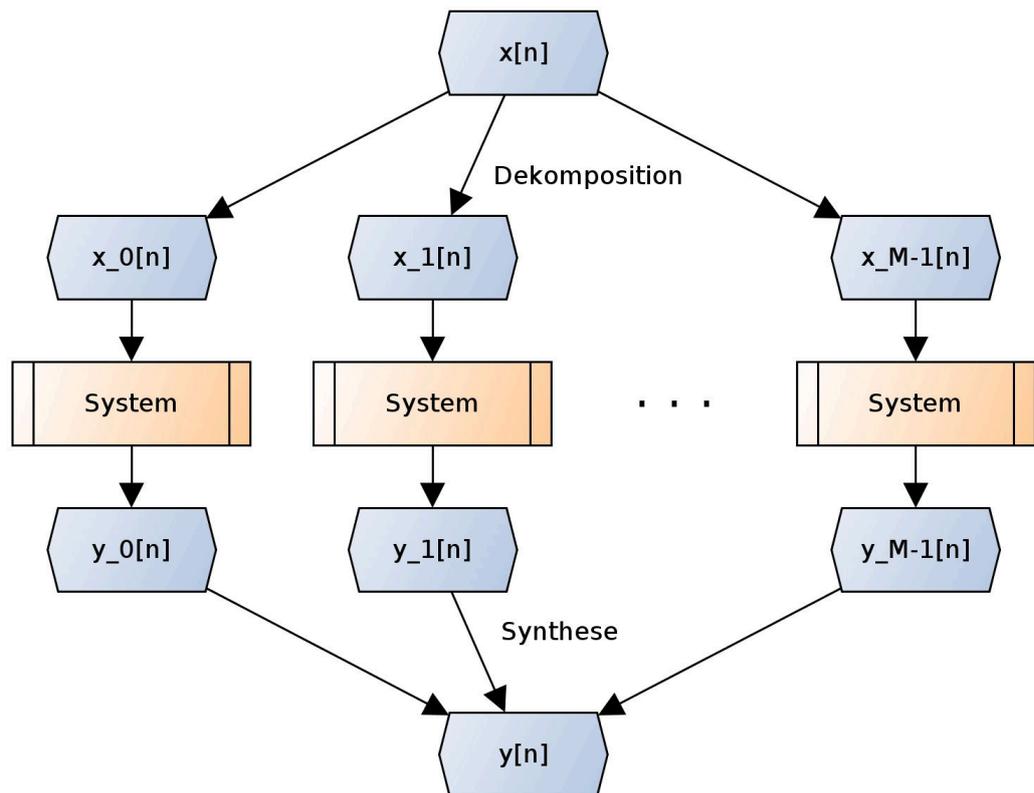


Abbildung 1.14: Superpositionsprinzip. Verwendung des Superpositionsprinzips für die Analyse komplexer Signale und Systeme.

1.3.2 Dekomposition von Signalen

Wie bereits im vorhergehenden Abschnitt erwähnt ist der Prozess der Dekomposition im Gegensatz zur Synthese nicht eindeutig. Für ein gegebenes Signal $x[n]$ existieren eine Vielzahl möglicher Dekompositionen bzw. Zerlegungen. Im Folgenden sollen die wichtigsten und häufigsten Zerlegungen kurz vorgestellt werden.

Vielzahl möglicher
Zerlegungen

Impulszerlegung

Die Impulszerlegung (s. Abb. 1.15) ist eine der einfachsten, aber auch eine der wichtigsten Dekompositionen:



Definition 1.19 (Impulszerlegung).

Es sei $x[n]$ ein Signal der Länge N . Dann ist die **Impulszerlegung** $D_{IMP}(x)$ definiert als:

$$D_{IMP}(x) = \{x_0, x_1, \dots, x_{N-1}\},$$

mit

$$x_i[n] = \delta[n - i] \cdot x[i], \quad i \in \{0, \dots, N - 1\}$$

und $\delta[n]$ dem Kronecker-Delta (s. Def. 1.9).

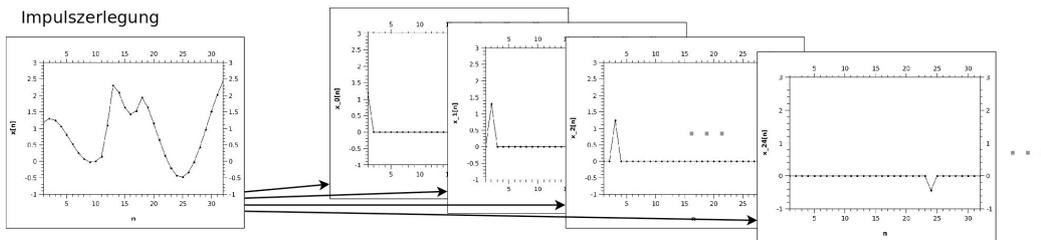


Abbildung 1.15: Impulszerlegung. Schematische Darstellung der Impulszerlegung eines Signals $x[n]$ der Länge N in N Signalkomponenten. Jede Komponente enthält genau einen Wert des Ausgangssignals x . Alle anderen Werte sind null.

Ein Eingangssignal $x[n]$ der Länge N wird durch die Impulszerlegung in N Signalkomponenten $x_i[n]$ aufgeteilt, wobei jede Komponente x_i nur an Position i den Abtastwert des Eingangssignals enthält und an allen anderen Positionen den Wert null hat. Ein Signal, das nur an einer Stelle einen von null verschiedenen Wert besitzt, wird auch als *Impuls* bezeichnet – daher der Name dieser Zerlegung. Durch die isolierte Betrachtung der einzelnen Abtastwerte des Eingangssignals, kann die Reaktion eines Systems auf ein Eingangssignal alleine dadurch untersucht und bestimmt werden, wenn die Reaktion des Systems auf einen Impuls bekannt ist. Diese Vorgehensweise findet sich auch bei der Methode der Faltung wieder, die in Abschnitt 1.4 näher erläutert wird.

Signal-
komponenten
Impuls

Stufenzerlegung

Die Stufenzerlegung (s. Abb. 1.16) ist der Impulszerlegung sehr ähnlich. Anstelle der direkten Abtastwerte des Eingangssignals liegen der Stufenzerlegung jedoch die *Veränderungen* der Abtastwerte des Eingangssignals zugrunde:

Veränderungen der
Abtastwerte

Definition 1.20 (Stufenzerlegung).

Es sei $x[n]$ ein Signal der Länge N . Dann ist die **Stufenzerlegung** $D_{STP}(x)$ definiert als:

$$D_{STP}(x) = \{x_0, x_1, \dots, x_{N-1}\},$$

mit

$$x_i[n] = \begin{cases} x[0] & \text{wenn } i = 0, \\ \Theta(n - i) \cdot (x[i] - x[i - 1]) & \text{wenn } i \in \{1, \dots, N - 1\}. \end{cases}$$

und $\Theta(n)$ der Heaviside-Funktion (s. Def. 1.13).



Ein Eingangssignal $x[n]$ der Länge N wird durch die Stufenzerlegung in N Signalkomponenten $x_i[n]$ aufgeteilt, wobei die ersten $i - 1$ Werte jeder Signalkomponente den Wert null haben und alle folgenden Werte ($i, \dots, N - 1$) den Wert der Veränderung des Eingangssignals von Position $i - 1$ auf Position i

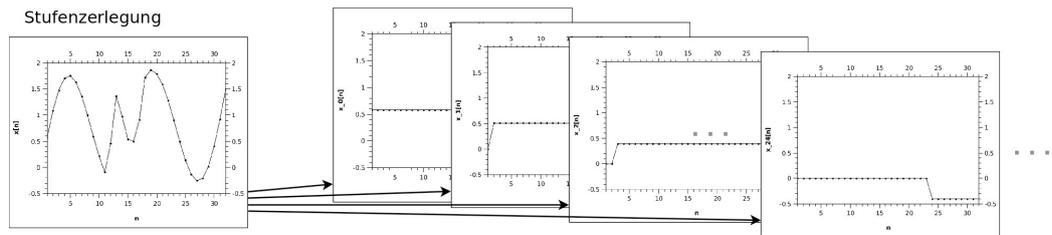


Abbildung 1.16: Stufenzerlegung. Schematische Darstellung der Stufenzerlegung eines Signals $x[n]$.

Reaktion des
Systems

haben ($x[i] - x[i - 1]$). Da der erste Abtastwert des Eingangssignals keinen Vorgänger besitzt und somit keine Differenz gebildet werden kann, wird die Signalkomponente $x_0[n]$ gesondert behandelt und erhält konstant den Wert $x[0]$. Während bei der Impulszerlegung das Wissen um die Reaktion eines Systems auf einen Impuls genutzt wird, steht bei der Stufenzerlegung die Reaktion eines Systems auf eine *Änderung* des Eingangssignals im Fokus.

Zerlegung in geraden und ungeraden Anteil

gerade und
ungerade
Symmetrie

Die Zerlegung eines Eingangssignals in einen geraden und einen ungeraden Anteil (s. Abb. 1.17) erzeugt aus dem Eingangssignal zwei Signalkomponenten. Die Bezeichnungen „gerader Anteil“ bzw. „ungerader Anteil“ könnten vermuten lassen, dass hiermit die Abtastwerte auf geraden und ungeraden Positionen gemeint sind. Dies ist an dieser Stelle jedoch nicht gemeint. Eine derartige Zerlegung existiert zwar ebenfalls, sie wird jedoch erst im nächsten Abschnitt beschrieben. Im Falle der hier beschriebenen Zerlegung sind nicht die Positionen der Abtastwerte, sondern die *Symmetrien* der beiden Signalkomponenten gemeint. Eine der erzeugten Komponenten hat eine sogenannte *gerade Symmetrie* während die andere Komponente eine *ungerade Symmetrie* besitzt.

Spiegelung &
Vorzeichen

Ein Signal $x[n]$ mit Länge N wird als symmetrisch bezeichnet, wenn sich die Abtastwerte des Signals an der Stelle $N/2$ betragsmäßig spiegeln, d.h. $|x[N/2 + k]| = |x[N/2 - k]|$ mit $k \in \{0, \dots, N/2\}$. Sind die Vorzeichen auf beiden Seiten der Spiegelung gleich, d.h. es handelt sich um eine Spiegelung entlang der vertikalen Achse durch $N/2$, wird die Symmetrie als *gerade* bezeichnet. Haben die beiden Seiten der Spiegelung entgegengesetzte Vorzeichen, d.h. es handelt sich um eine Punktspiegelung am Punkt $(N/2, 0)$, so wird die Symmetrie als *ungerade* bezeichnet.

Periodizität

Im Rahmen dieser Symmetriebetrachtung wird das Signal als periodisch angenommen, d.h. an Stelle $x[N]$ beginnt das Signal erneut und hat den Wert $x[0]$, an Stelle $x[N + 1]$ hat es den Wert $x[1]$, usw. Alternativ kann man die Indizes der Signale auch als kongruent modulo N verstehen, wie in folgender Definition der Zerlegung in einen geraden und ungeraden Anteil geschehen:

Zerlegung in geraden/ungeraden Anteil

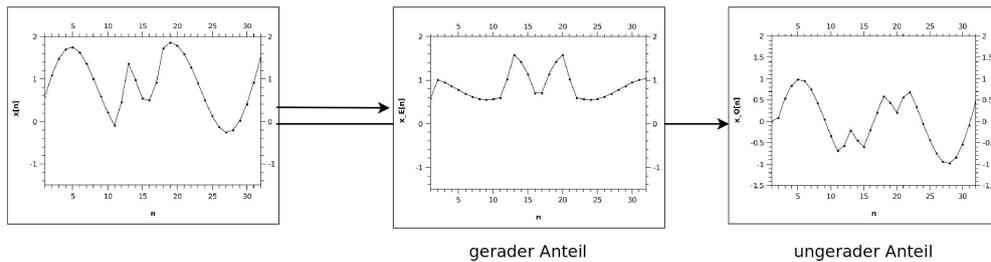


Abbildung 1.17: Zerlegung in geraden und ungeraden Anteil. Schematische Darstellung der Zerlegung eines Signals $x[n]$ in einen geraden und ungeraden Anteil. Man beachte die Sonderbehandlung an Stelle $x[0]$: Für den geraden Anteil gilt $x_E[0] = x[0]$ und für den ungeraden Anteil gilt $x_O[0] = 0$. Diese Sonderbehandlung liegt darin begründet, dass im Kontext der Symmetriebetrachtung das Signal als periodisch angesehen wird.

Definition 1.21 (Zerlegung in geraden und ungeraden Anteil).

Es sei $x[n]$ ein Signal der Länge N . Dann ist die Zerlegung $D_{E/O}(x)$ des Signals x in eine Komponente x_E mit gerader Symmetrie und eine Komponente x_O mit ungerader Symmetrie wie folgt definiert:

$$D_{E/O}(x) = \{x_E, x_O\},$$

mit

$$x_E[n] = \frac{x[n \bmod N] + x[(N - n) \bmod N]}{2},$$

$$x_O[n] = \frac{x[n \bmod N] - x[(N - n) \bmod N]}{2}.$$

Man beachte, dass sich durch die Betrachtung des Eingangssignals als periodisches Signal die Gleichung für $x_E[0]$ zu $x[0]$ vereinfacht und sich die Gleichung für $x_O[0]$ zu 0 vereinfacht. Der Hintergrund, warum es sinnvoll bzw. notwendig sein kann, ein Signal als periodisches Signal aufzufassen, wird im Zuge der später erläuterten *Fourier-Analyse* deutlicher werden (s. Kurseinheit 2).

Warum die in Def. 1.21 gegebene Berechnungsvorschrift immer eine gültige³⁴ Zerlegung eines beliebigen Eingangssignals erzeugt, mag auf den ersten Blick nicht sofort klar sein. Bildet man jedoch die Summe aus $x_E[n]$ und $x_O[n]$ und gruppiert die einzelnen Terme um, so erhält man:

$$x_E[n] + x_O[n] = \frac{x[n]}{2} + \frac{x[n]}{2} + \frac{x[N - n]}{2} - \frac{x[N - n]}{2} = x[n]$$

Die Summe aus $x_E[n]$ und $x_O[n]$ ergibt also immer $x[n]$.³⁵

³⁴d.h. eine Zerlegung, deren Summe dem Eingangssignal entspricht

³⁵ E und O stehen für *even* (gerade) und *odd* (ungerade).



Gültigkeit der Berechnungsvorschrift

Zerlegung in gerade/ungerade Abtastwerte

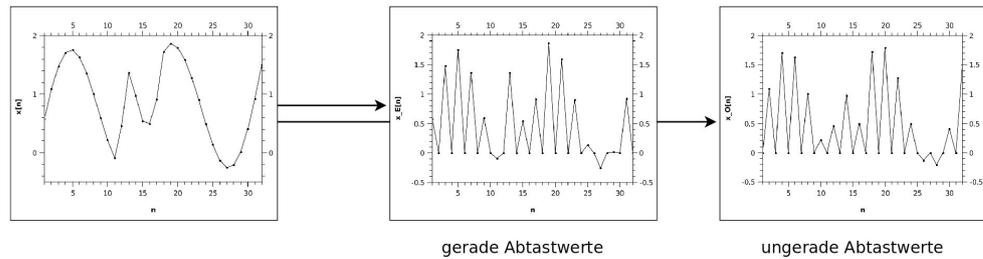


Abbildung 1.18: Zerlegung in gerade und ungerade Abtastwerte. Schematische Darstellung der Zerlegung eines Signals $x[n]$ in seine geraden und ungeraden Abtastwerte. „Gerade“ bzw. „ungerade“ bezieht sich hierbei auf den Index des jeweiligen Abtastwertes, nicht auf seinen Betrag.

Zerlegung in gerade und ungerade Abtastwerte

gerade und ungerade Die Zerlegung des Eingangssignals in gerade und ungerade Abtastwerte (s. Abb. 1.18) erzeugt zwei Signalkomponenten, die jeweils die Abtastwerte an den geraden bzw. ungeraden Positionen des Eingangssignals enthalten:



Definition 1.22 (Zerlegung in gerade und ungerade Abtastwerte).

Es sei $x[n]$ ein Signal der Länge N . Dann ist die Zerlegung $D_{EOS}(x)$ des Signals x in zwei Signalkomponenten x_E und x_O mit jeweils den geraden bzw. ungeraden Abtastwerten des Signals x wie folgt definiert:

$$D_{EOS}(x) = \{x_E, x_O\},$$

mit

$$x_E[n] = x[n] \cdot ((n + 1) \bmod 2),$$

$$x_O[n] = x[n] \cdot (n \bmod 2).$$

Diese möglicherweise etwas seltsam anmutende Dekomposition wird im Zuge der *Fast Fourier Transformation* (FFT) verwendet. Die FFT ist ein divide-and-conquer Algorithmus³⁶, der – im Gegensatz zur *naiven* Implementation – die schnelle Berechnung einer Fourier-Transformation ermöglicht (s. Abschnitt 2.3).

³⁶Divide-and-conquer (engl. für „teile und herrsche“) Algorithmen lösen eine Problemstellung durch die rekursive Aufteilung eines Problems in einfacher zu lösende Teilprobleme. Die Laufzeit liegt hierbei im Allgemeinen in der Größenordnung von $n \log n$.

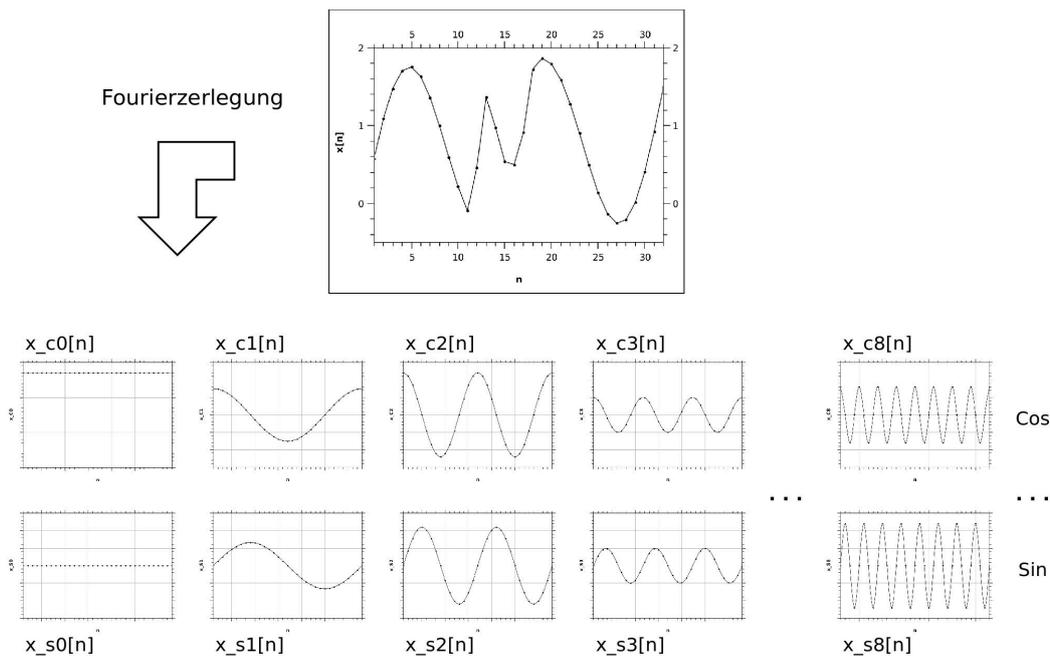


Abbildung 1.19: Fourier-Zerlegung. Schematische Darstellung der Fourier-Zerlegung eines Signals $x[n]$ in $N + 2$ Kosinus- und Sinusanteile.

Fourier-Zerlegung (→ Fourier-Transformation)

Die zuvor behandelten Dekompositionen, abgesehen von der Zerlegung in symmetrische Anteile und der Stufenzerlegung, erzeugen jeweils Signalkomponenten, die als direkte Teilmengen des Eingangssignals verstanden werden können. Hierbei werden die einzelnen Abtastwerte der Eingangssignale „in Gänze“ einer der jeweiligen Signalkomponenten zugeordnet und selber nicht weiter aufgeteilt. Bislang wick nur die Zerlegung des Eingangssignals in zwei symmetrische Komponenten von diesem Schema ab und teilte jeden Abtastwert in zwei Summanden auf. Die Fourier-Zerlegung eines Signals, auch Fourier-Transformation genannt, teilt ebenfalls die einzelnen Abtastwerte in mehrere, potenziell von null verschiedene Summanden auf.

Aufteilung in mehrere Summanden

Für ein Eingangssignal $x[n]$ mit Länge N erzeugt die Fourier-Zerlegung³⁷ $N + 2$ Signalkomponenten (s. Abb. 1.19). Die Signalkomponenten bestehen aus jeweils $N/2 + 1$ Kosinus- und Sinusfunktionen ohne Phasenversatz³⁸ und mit aufsteigender Frequenz. Die niedrigste „Frequenz“ stellen hierbei zwei konstante Signalkomponenten dar. Es sind die beiden Komponenten mit *null* Kosinus- bzw. Sinusschwingungen. Die Kosinusfunktion nimmt dabei den „Gleichanteil“ des Eingangssignals an. Der Gleichanteil entspricht der Gesamtverschiebung des Eingangssignals entlang der abhängigen Variablen³⁹. Die Sinusfunktion

Zerlegung in Kosinus- und Sinusfunktionen

³⁷Die Bezeichnung „die Fourier-Zerlegung“ ist etwas irreführend, da es – wie in Kurs-einheit 2 erläutert wird – mehrere Varianten der Fourier-Transformation gibt. In diesem Fall bezieht sich der Begriff *Fourier-Zerlegung* auf die reellwertige Variante der diskreten Fourier-Transformation.

³⁸Als *Phase* bezeichnet man die Verschiebung einer periodischen Funktion entlang der unabhängigen Variablen (üblicherweise entlang der „x-Achse“).

³⁹üblicherweise entlang der „y-Achse“

nimmt hingegen immer einen konstanten Wert von 0 an. Für alle weiteren Signalkomponenten erhöht sich die Zahl der ganzen Schwingungen der Kosinus- und Sinusfunktionen mit jeder Signalkomponente um eine weitere Schwingung.

Hieraus wird ersichtlich, dass die Frequenzen der Kosinus- und Sinusfunktionen der Signalkomponenten nicht vom *Inhalt* des Eingangssignals x , sondern nur von der Länge N des Signals abhängen. Da der Phasenversatz für alle Signalkomponenten ebenfalls fest vorgegeben ist⁴⁰, bleibt als einzig freier Parameter für jede Signalkomponente die frei wählbare, konstante Amplitude der jeweiligen Kosinus- oder Sinusfunktion. Bei der an dieser Stelle beschriebenen Fourier-Transformation handelt es sich um eine reellwertige, diskrete Variante der Fourier-Transformation. Im späteren Verlauf des Kurses wird darüber hinaus auch eine komplexwertige Variante sowohl in ihrer diskreten als auch ihrer kontinuierlichen Form vorgestellt. Daher soll an dieser Stelle noch keine Definition für die Berechnung einer Fourier-Transformation angegeben werden, da diese ausführlich in der nächsten Kurseinheit beschrieben und erläutert wird.

Die Motivation, sich überhaupt für die Zerlegung eines Signals in sinusoidale Signalkomponenten zu interessieren, hat im Wesentlichen drei Hauptgründe:

- natürliche Signalquellen

 - Viele Signale, z.B. Audiosignale, bestehen aus der Überlagerung einer Vielzahl sinusförmiger Funktionen. Durch die Zerlegung in die entsprechenden Signalkomponenten, kann diese Überlagerung in gewisser Weise „rückgängig“ gemacht werden, und es zeigen sich die Frequenzanteile, aus denen ein solches Signal zusammengesetzt ist. Darüber hinaus ist es auf diese Weise möglich, nachträglich bestimmte Frequenzanteile gezielt zu verändern oder ganz zu entfernen, z.B. um bestimmte Störgeräusche zu entfernen.
- Analyse linearer Systeme

 - Lineare Systeme haben die Eigenschaft, sinusförmige Funktionen nur in ihrer Amplitude und Phase, nicht aber in ihrer Frequenz zu verändern. Somit ist es möglich – ähnlich wie bei der Impulszerlegung – lineare Systeme über ihr spezifisches Verhalten gegenüber einfachen sinusförmigen Signalen zu charakterisieren.
- Fourier-Analyse

 - Die Fourier-Transformation ist die Grundlage für das Gebiet der sogenannten *Fourier-Analyse*. Diese stellt eine Vielzahl z.T. mächtiger mathematischer Werkzeuge zur Verfügung, die von einer Vielzahl moderner Algorithmen für die Signalverarbeitung genutzt werden.

1.3.3 Analyse nicht-linearer Systeme

Nachdem sich die vorhergehenden Abschnitte intensiv mit der Analyse von linearen Systemen beschäftigt haben, stellt sich die Frage, welche Strategien es für die Untersuchung nicht-linearer Systeme gibt. Die Antwort ist einerseits ernüchternd und andererseits betont sie die Wichtigkeit des Konzeptes der Linearität: Der faktisch einzig existierende Ansatz für die Analyse nicht-linearer Systeme besteht darin, sie als zumindest teilweise lineare Systeme anzusehen

⁴⁰Der Phasenversatz ist überall 0.

und diese Subteile getrennt zu untersuchen oder das nicht-lineare System über eine Transformation zu linearisieren. In einem nicht unerheblichen Teil der Fälle wird die Nichtlinearität eines zu untersuchenden Systems schlicht ignoriert.

Der vorhergehende Abschnitt hat Ihnen das wichtige Konzept der *Linearität* im Kontext der Analyse linearer Systeme erläutert. Die beschriebenen Eigenschaften dieser Systeme und die verschiedenen Ansätze, diese Eigenschaften für die Analyse der Systeme zu verwenden, bilden die theoretische Basis für eine Vielzahl von Methoden der Signalverarbeitung. Insbesondere die im Folgenden beschriebene Faltung und die in der nächsten Kurseinheit beschriebene Fourier-Analyse bauen auf den hier vermittelten Grundlagen auf.

Linearität

1.4 Faltung diskreter Signale

Lernziele:

Der folgende Abschnitt dieser Kurseinheit stellt eine zentrale Methode der digitalen Signalverarbeitung vor: *Die diskrete Faltung*. Neben den grundlegenden Eigenschaften dieser mathematischen Operation wird insbesondere der enge Zusammenhang zwischen der Faltungsoperation und den zuvor beschriebenen linearen Systemen dargestellt. Es wird gezeigt, dass mittels der Faltungsoperation und der Impulsantwort eines Systems jedes Eingangssignal in ein entsprechendes Ausgangssignal überführt werden kann. Der Abschnitt endet mit einer kurzen Vorstellung und Abgrenzung einer weiteren mathematischen Operation, die der Faltung mathematisch sehr ähnlich, konzeptionell jedoch eigenständig ist: die Korrelation.

Eine *Faltung*, auch *Konvolution* genannt, ist eine mathematische Operation, die zwei Signale miteinander kombiniert. Sie wird durch den Faltungsoperator „ $*$ “ symbolisiert⁴¹. Wie bereits in Abschnitt 1.3 erwähnt, nutzt die Faltung das für lineare Systeme geltende Superpositionsprinzip (s. Def. 1.18), indem die Reaktion eines Systems für jede Signalkomponente separat betrachtet wird und anschließend mittels Addition aus diesen einzelnen Reaktionen ein Ausgabesignal erzeugt wird.

Konvolution

Die Grundlage der Faltungsoperation bilden zwei zentrale Konzepte: Die *Deltafunktion* $\delta[n]$ und die *Impulsantwort* $h[n]$. Die Deltafunktion wurde bereits als „diskrete Variante“ des Kronecker-Deltas in Definition 1.9 eingeführt. Sie liefert nur für den Abtastwert an Stelle null einen Wert von eins zurück und an allen anderen Stellen einen Wert von null (s. Abb. 1.20):

Deltafunktion & Impulsantwort

⁴¹Der Stern „ $*$ “ als Faltungsoperator wird häufig mit dem Punktoperator „ \cdot “ für die Multiplikation verwechselt!

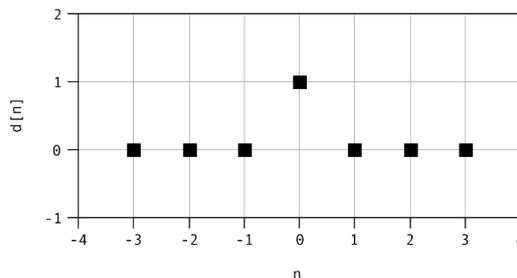


Abbildung 1.20: Deltafunktion. Schematische Darstellung der Deltafunktion $\delta[n]$, die auch als diskrete Variante des Kronecker-Deltas bekannt ist (s. Def. 1.9 und 1.23). Die Deltafunktion hat genau an Stelle null den Funktionswert eins und ansonsten an allen anderen Stellen den Funktionswert null.



Definition 1.23 (Deltafunktion).

Die **Deltafunktion** $\delta[n]$ ist definiert durch

$$\delta[n] = \begin{cases} 1 & \text{wenn gilt } n = 0, \\ 0 & \text{wenn gilt } n \neq 0, \end{cases} \quad n \in \mathbb{Z}.$$

Einheitsimpuls

Ein Signal bzw. eine Funktion, die nur an einer Stelle einen von null verschiedenen Wert zurückliefert, wird auch als *Impuls* bezeichnet, und da die Deltafunktion genau den Wert eins zurückliefert, wird sie auch als *Einheitsimpuls* bezeichnet. Ähnlich wie in der linearen Algebra ein Vektor als Linearkombination von Basisvektoren dargestellt werden kann, ist es mit dem Einheitsimpuls durch Verschiebung und Skalierung möglich, jeden anderen Impuls abzubilden. Angenommen es gäbe einen Impuls $p[n]$ an der Stelle 10 mit einer Amplitude von -1.49 . Dieser Impuls kann durch Verschiebung und Skalierung der Deltafunktion als $p[n] = -1.49 \cdot \delta[n - 10]$ dargestellt werden. Diese Darstellung entspricht exakt der in Definition 1.19 beschriebenen Impulszerlegung eines Signals! Ein Signal kann demnach als eine Summe von skalierten und verschobenen Deltafunktionen aufgefasst werden.

Das zweite Konzept, die *Impulsantwort*, beschreibt die Reaktion eines Systems auf die Eingabe eines Einheitsimpulses:



Definition 1.24 (Impulsantwort).

Es sei S ein lineares, verschiebungsinvariantes System. Dann ist die **Impulsantwort** $h_S[n]$ des Systems gegeben durch

$$\delta[n] \rightarrow \boxed{S} \rightarrow h_S[n]$$

mit $\delta[n]$ der Deltafunktion.

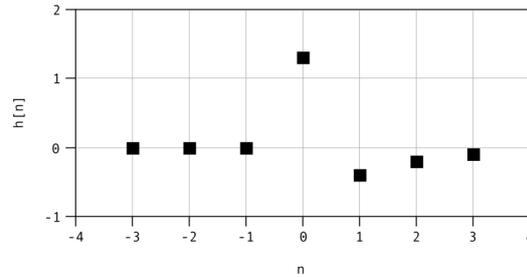


Abbildung 1.21: Impulsantwort. Schematische Darstellung einer möglichen Impulsantwort eines linearen Systems auf die Eingabe einer Deltafunktion $\delta[n]$.

Einen Eindruck davon, wie eine solche Impulsantwort aussehen könnte, liefert die Darstellung in Abbildung 1.21. Es stellt sich nun die Frage, wie sich diese Impulsantwort verändert, wenn ein anderer Impuls als der Einheitsimpuls in das System S eingegeben wird. Die Antwort findet sich in zwei der zuvor beschriebenen Eigenschaften linearer Systeme: der Homogenität (Def.: 1.16) und der Verschiebungsinvarianz (Def.: 1.17). Durch diese Eigenschaften wird garantiert, dass eine Verschiebung des Eingangssignals um den Wert a und eine Skalierung des Eingangssignals um den Faktor b zu einer entsprechenden Verschiebung und Skalierung des Ausgangssignals führt. Die Eingabe eines Impulses $b \cdot \delta[n - a]$ in das System S führt demnach zu einer Impulsantwort $b \cdot h_S[n - a]$. Die Impulsantwort ist also exakt in der gleichen Weise verschoben und skaliert wie die in das System eingegebene Deltafunktion. Das bedeutet, wenn die Systemantwort auf den Einheitsimpuls bekannt ist, sind auch die Antworten auf alle anderen Impulse bekannt.

Impulsantwort auf beliebige Impulse

Folglich kann ein beliebiges Eingangssignal $x[n]$ zunächst mittels der Impulszerlegung (Def.: 1.19) in eine Menge $D_{IMP}(x)$ einzelner Impulse x_i aufgeteilt werden, um im Anschluss für jeden dieser Impulse die Impulsantwort des Systems zu bestimmen. Abschließend kann das Ausgangssignal aus der Menge der Impulsantworten durch Summation synthetisiert werden. Für diesen Prozess – der Bestimmung des Ausgangssignals eines Systems für ein *beliebiges* Eingangssignal – wird nur die Impulsantwort h_S des Systems S bzgl. des Einheitsimpulses benötigt. Die Impulsantwort eines linearen Systems beschreibt somit das Verhalten des Systems *vollständig!*

Dekomposition & Synthese

Die im vorhergehenden Absatz informell beschriebene Vorgehensweise entspricht der Berechnung des Ausgangssignals $y[n]$ eines Systems S mit bekannter Impulsantwort $h_S[n]$ durch eine *Faltung*:

**Definition 1.25** (Faltung).

Gegeben seien das Signal $x[n]$ der Länge N und das Signal $h[n]$ der Länge M . Dann ist die **Faltung von x und h** definiert durch

$$y[i] := \sum_{j=0}^{M-1} h[j] x[i-j], \quad i \in \{0, \dots, M+N-2\}.$$

Das resultierende Signal $y[n]$ hat eine Länge von $N+M-1$. Die Kurzform für die gegebene Berechnungsvorschrift nutzt den Faltungsoperator „*“ und ist definiert als:

$$x[n] * h[n] = y[n].$$

Obwohl die in dieser Definition angegebene Berechnungsvorschrift die Berechnung des Ausgangssignals $y[n]$ auf eine etwas andere Weise vollzieht als die zuvor gegebene, informelle Beschreibung, sind die erzeugten Ausgangssignale für beide Vorgehensweisen identisch. Der Unterschied besteht in der jeweiligen Perspektive auf die Berechnung der einzelnen Werte des Ausgangssignals.

Perspektive des
Eingangssignals

Die informelle Beschreibung betrachtet die Berechnung des Ausgangssignals aus der Perspektive des Eingangssignals. Für jeden Abtastwert des Eingangssignals, also für jeden Impuls der Impulszerlegung, wird berechnet, zu welchen Abtastwerten des Ausgangssignals der Impuls einen Beitrag nach dem Verlassen des Systems leistet. Der folgende Pseudocode führt die Berechnung des Ausgangssignals auf diese Weise durch:

```
memset(y,0,N+M-1);           // Initialisiere das Ausgangssignal
                               // mit Nullen.

for (i = 0; i < N; i++)      // Für jeden Abtastwert des
                               // Eingangssignals ...

    for (j = 0; j < M; j++)  // ... durchlaufe die Impulsantwort
                               // des Systems ...

        y[i+j] += x[i]*h[j]; // ... und addiere das Produkt aus
                               // Eingangssignal und Impulsantwort
                               // auf das Ausgangssignal
```

Eine schematische Darstellung dieses Pseudocodes liefert Abbildung 1.22. Das Eingangssignal $x[n]$ mit Länge 7 ist hierbei in Grüntönen und die Impulsantwort $h[n]$ mit Länge 4 in Blautönen schattiert. Das Ausgangssignal ist mit $y[n]$ bezeichnet und hat eine Länge von $7+4-1=10$. Die dargestellte Bearbeitung der verschachtelten Schleifen erfolgt in der Abbildung zeilenweise von unten nach oben. Das wiederholte Durchlaufen der inneren Schleife mit Index j ist durch die mehrfachen, waagrechten Pfeile angedeutet. Die äußere Schleife mit Index i wird durch den vertikalen Pfeil repräsentiert.

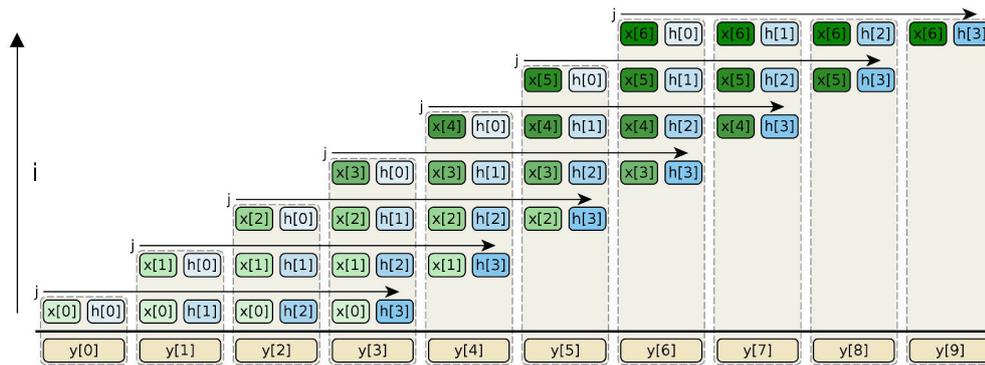


Abbildung 1.22: Faltung aus der Perspektive des Eingangssignals. Für jeden Abtastwert des Eingangssignals $x[n]$ wird bestimmt, wie dieser nach Verlassen des Systems, also nach Skalierung und Verschiebung der Impulsantwort $h[n]$, zu den Abtastwerten des Ausgangssignals $y[n]$ beiträgt.

Im Gegensatz zur informellen Beschreibung betrachtet die in der Definition gegebene Berechnungsvorschrift die Faltung aus der Perspektive des Ausgangssignals. Für jeden Wert des Ausgangssignals wird berechnet, welche Abtastwerte des Eingangssignals zum Ausgangssignal beitragen. Mit dieser Herangehensweise ergibt sich folgender Pseudocode:

Perspektive des
Ausgangssignals

```

for (i = 0; i < N+M-1; i++) // Für jeden Wert des
                           // Ausgangssignals ...

    for (j = 0; j < M; j++) // ... durchlaufe die Elemente
                           // der Impulsantwort ...

        y[i] += x[i-j]*h[j]; // ... und multipliziere das
                              // jeweilige Element der Impuls-
                              // antwort mit dem "passenden"
                              // Abtastwert des Eingangssig-
                              // nals

```

Anhand der schematischen Darstellung dieses Pseudocodes in Abbildung 1.23 ist ersichtlich, dass beide Berechnungsvorschriften – sowohl die der informellen Beschreibung, als auch die der Definition – identische Ausgangssignale erzeugen. Darüber hinaus wird aus der Perspektive des Ausgangssignals ein Problem ersichtlich, das aus der Perspektive des Eingangssignals leicht zu übersehen ist: In den Randbereichen des Ausgangssignals sind Teile der Summe aufgrund „fehlender Teile“ des Eingangssignals unbestimmt. Diese fehlenden Teile sind in Abbildung 1.23 grau unterlegt und mit Fragezeichen beschriftet. Zu den üblichen Vorgehensweisen, diesem Problem zu begegnen, gehört das sogenannte *Padding*, das Auffüllen der fehlenden Werte mit einem konstanten Wert. Dies geschieht auch bei dem oben beschriebenen Verfahren aus Sicht des Eingangssignals, bei dem die fehlenden Abtastwerte implizit mit einem konstanten Wert 0 angenommen werden. Eine weitere Möglichkeit besteht in der Spiegelung des Eingangssignals an den Randbereichen, um die lokale Charakteristik des

Erzeugung
identischer
Ausgangssignale

Padding

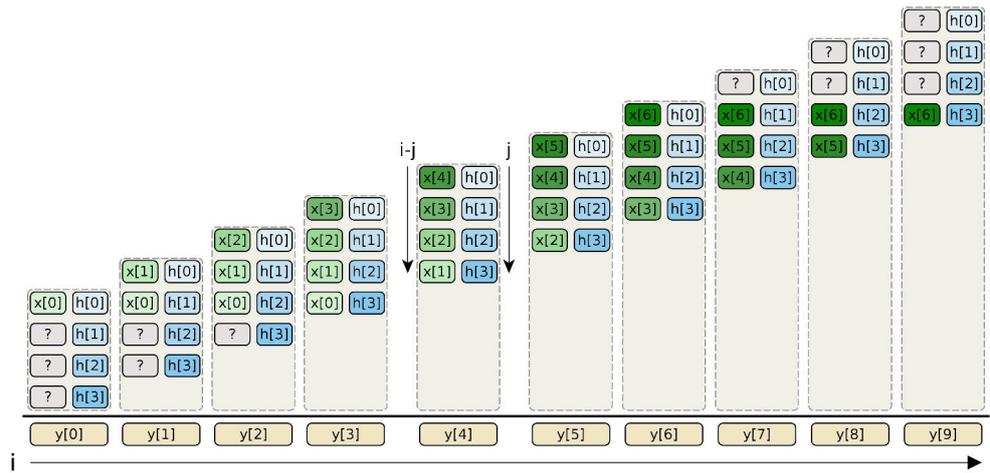


Abbildung 1.23: Faltung aus der Perspektive des Ausgangssignals. Für jeden Abtastwert des Ausgangssignals $y[n]$ wird bestimmt, welche Abtastwerte des Eingangssignals $x[n]$ nach Verlassen des Systems, also nach Filterung mittels der Impulsantwort $h[n]$, zum Ausgangssignal beitragen.

Spiegelung der
Randbereiche

Signals besser anzunähern. Unabhängig davon, auf welche Weise die fehlenden Werte approximiert werden, weisen die Randbereiche des Ausgangssignals einer Faltung immer verfälschte Werte auf. Aus diesem Grund, insbesondere wenn die Länge des Eingangssignals deutlich größer ist als die Länge der Impulsantwort, werden die Randbereiche des Ausgangssignals meist verworfen und nicht für die weitere Verarbeitung verwendet.

1.4.1 Eigenschaften der Faltung

Faltung →
unabhängige
mathematische
Operation

Man beachte, dass innerhalb Definition 1.25 nur allgemein zwei Signale x und h als Operanden der Faltung genannt werden. Keines der Signale wird explizit als *Eingangssignal* oder *Impulsantwort* identifiziert. Dies liegt darin begründet, dass die Faltung eine von der konkreten Anwendung unabhängige mathematische Operation ist, die nach einer festen Berechnungsvorschrift zwei Signale zu einem dritten Signal kombiniert. Sie ist einfach als mathematisches „Werkzeug“ anzusehen, mit dessen Hilfe sich das Ausgangssignal eines linearen Systems mittels der Impulsantwort des Systems berechnen lassen kann. In diesem Anwendungskontext wird die Impulsantwort h_S häufig auch als *Filterkern*, *Filteroperator*, *Faltungskern*, oder *Kernel* (engl.) bezeichnet. Aus der allgemeinen Definition der Faltung geht entsprechend auch hervor, dass es keinen Unterschied macht, welches der beiden Signale (x oder h) in einer konkreten Anwendung die Rolle des „Eingangssignals“ oder die Rolle der „Impulsantwort“ übernimmt.

Verknüpfung
von Faltungs-
operationen

Wie bereits in Abschnitt 1.3 erläutert, können einfache, lineare Systeme zu komplexen Systemen „verschaltet“ werden, die weiterhin alle Eigenschaften eines linearen Systems erfüllen (z.B. Abb. 1.12). Auf die gleiche Weise können auch Faltungsoperationen miteinander verknüpft werden, um z.B. mehrere Impulsantworten miteinander zu kombinieren. Hierbei gelten für die Faltungsoperation die algebraischen Regeln des Kommutativ-, Assoziativ- und Distributivgesetzes:

Merksatz 1.4 (Algebraische Regeln der Faltungsoperation).

Die Faltungsoperation ist

$$\textit{kommutativ:} \quad a[n] * b[n] = b[n] * a[n],$$

$$\textit{assoziativ:} \quad (a[n] * b[n]) * c[n] = a[n] * (b[n] * c[n]),$$

$$\textit{distributiv:} \quad a[n] * b[n] + a[n] * c[n] = a[n] * (b[n] + c[n]).$$



Die Eigenschaften der Kommutativität und Assoziativität können hierbei für die Modellierung kaskadierter Systeme verwendet werden, während die Eigenschaft der Distributivität die Modellierung paralleler Systeme, deren Ausgänge addiert werden, erlaubt.

Modellierung
kaskadierter &
paralleler Systeme

1.4.2 Elementare Impulsantworten

In den bisherigen Abschnitten wurde das Verhältnis von Eingangssignal, Impulsantwort und Ausgangssignal auf eine überwiegend abstrakte Art und Weise beschrieben. Im Folgenden sollen nun einige elementare Impulsantworten vorgestellt werden, um ein besseres Gefühl für die Wirkung dieser Impulsantworten auf das jeweilige Eingangssignal zu vermitteln.

Identität

Die einfachste Impulsantwort besteht aus der *Deltafunktion* $\delta[n]$ selbst. Wird ein Eingangssignal $x[n]$ mit der Deltafunktion gefaltet, so wird das Eingangssignal nicht verändert. Es gilt:

Deltafunktion

$$x[n] * \delta[n] = x[n]$$

Die Deltafunktion ist das *neutrale Element* der Faltung, so wie die Zahl „1“ das neutrale Element für die Multiplikation und die Zahl „0“ das neutrale Element für die Addition ist. In diesem Zusammenhang wird die Deltafunktion auch als *Identität* bzw. *identische Abbildung* der Faltung bezeichnet. Abgesehen von den mathematischen Eigenschaften der Identität wird die Deltafunktion z.B. für die Modellierung verlustfreier Übertragungswege eingesetzt.

Verstärkung

Wird die Deltafunktion mit einem konstanten Faktor k multipliziert, so ergibt sich eine Impulsantwort, die das Eingangssignal um den gleichen Faktor skaliert:

Multiplikation mit
konstantem Faktor

$$x[n] * (k \cdot \delta[n]) = k \cdot x[n]$$

Das auf diese Weise beschriebene System ist ein *Verstärker*, wenn der Faktor größer eins ist und ein *Abschwächer*, wenn der Faktor kleiner eins ist. Bei einem negativen Faktor wird zudem das Eingangssignal an der Achse der unabhängigen Variablen (zumeist die X-Achse) gespiegelt.

Verzögerung

Subtraktion eines konstanten Wertes

Durch die Subtraktion eines konstanten Wertes s auf die unabhängige Variable der Deltafunktion entsteht eine Impulsantwort, die das Eingangssignal um den Wert von s verzögert:

$$x[n] * \delta[n - s] = x[n - s]$$

Ein lineares System, das Signale auf diese Weise verzögert, wird insbesondere für die Modellierung von Signallaufzeiten eingesetzt. Dies gilt sowohl für die Übertragung von Signalen über große Reichweiten, z.B. Funksignale von Raumsonden, als auch für die Übertragung von Signalen in Medien mit geringerer Übertragungsgeschwindigkeit, bei der Verzögerungen schon bei kürzeren Distanzen eine Rolle spielen, z.B. beim Sonar.

Spiegelung des Filterkerns

An dieser Stelle sei noch auf eine „Eigenart“ der Faltung hingewiesen, die am Beispiel der Verzögerung deutlich wird. Warum wird für eine Verzögerung von s der Wert von s subtrahiert und nicht addiert? Sollte dies nicht genau andersherum sein? Ein erneuter Blick auf Abbildung 1.23 kann diese Frage beantworten. Schaut man sich z.B. die Berechnung des Ausgangswertes $y[4]$ an, so ist zu erkennen, dass mit aufsteigendem Index j die Werte des Filterkerns mit den Werten des Eingangssignals $x[n]$ „nach links“, d.h. mit absteigenden Indizes (4,3,2,1), multipliziert werden. Aus der Perspektive des Eingangssignals – also von $x[1]$ „nach rechts bzw. oben“ schauend – erscheint der Filterkern $h[n]$ entlang seiner unabhängigen Variablen gespiegelt ($h[3], h[2], h[1], h[0]$). Allgemein bedeutet dies, dass die Werte eines Filterkerns $h[n]$ mit aufsteigenden Indizes sich auf entsprechend weit *zurückliegende* Werte des Eingangssignals beziehen! Der Wert $h[10]$ eines Filterkerns bezieht sich beispielsweise jeweils auf einen Wert des Eingangssignals, der 10 Abtastwerte bzgl. des aktuellen Ausgangswertes n „in der Vergangenheit“ liegt ($y[n] += x[n - 10] \cdot h[10]$).

Echo

Kombination von Deltafunktionen

Über die Kombination mehrerer ggf. skaliertes und verzögerter Deltafunktionen kann eine Impulsantwort erzeugt werden, die ein oder mehrere Echos dem Eingangssignal hinzufügt. So erzeugt das folgende Beispiel ein Echo im Abstand von 3 Abtastwerten mit einer Amplitude von 50% des Eingangssignals:

$$x[n] * (\delta[n] + 0.5 \cdot \delta[n - 3]) = x[n] + 0.5 \cdot x[n - 3]$$

Das Hinzufügen von Echos findet z.B. bei der Abmischung von Musikstücken Verwendung, um ein natürlicheres Klangbild zu erzeugen.

diskrete Ableitung

Ableitung

Mit einer geeignet geformten Impulsantwort können auch „Analysis-artige“ Operationen durchgeführt werden, z.B. eine angenäherte erste Ableitung des Eingangssignals. Für eine solche Approximation gibt es verschiedene Ansätze.

Der einfachste Ansatz bildet die Differenz zwischen dem aktuellen Abtastwert und dem vorhergehenden Abtastwert:

$$\begin{aligned} y[n] &= x[n] - x[n-1] \\ &= x[n] * (\delta[n] - \delta[n-1]) \end{aligned}$$

Ein weiterer, häufig eingesetzter Ansatz approximiert die erste Ableitung symmetrisch um einen Abtastwert herum:

$$\begin{aligned} y[n] &= (x[n+1] - x[n-1]) / 2 \\ &= x[n] * (0.5 \cdot \delta[n+1] - 0.5 \cdot \delta[n-1]) \end{aligned}$$

Insbesondere in der Bildverarbeitung wird diese Approximation der ersten Ableitung eines Signals häufig für die Detektion von Kanten im Bild oder für die Beschreibung der Struktur eines lokalen Bildbereiches verwendet.

Bildverarbeitung

diskrete Stammfunktion

Auf ähnliche Weise wie die erste Ableitung des Eingangssignals kann auch die Stammfunktion des Eingangssignals approximiert werden:

Approximation der Stammfunktion

$$\begin{aligned} y[n] &= x[n] + y[n-1] \\ &= x[n] * \Theta(n) \end{aligned}$$

mit $\Theta(n)$ der Heaviside-Funktion (s. Def. 1.13).

Aufgrund der rekursiven Definition der approximierten Stammfunktion wird es notwendig, dass die entsprechende Impulsantwort bis ins positiv Unendliche einen Wert von eins annimmt. Dies kann durch die endliche Kombination von einfachen Deltafunktionen nicht erreicht werden. Aus diesem Grund wird hier die bereits aus Definition 1.13 bekannte Heaviside-Funktion als Impulsantwort verwendet.

Heaviside-Funktion

Tiefpassfilter

Wie zuvor erwähnt werden in Abhängigkeit des jeweiligen Kontextes Impulsantworten oft auch als *Filterkerne* oder einfach nur als *Filter* bezeichnet. Zu den meist verwendeten Filtern gehören hierbei die sogenannten *Tiefpassfilter*. Ein Tiefpassfilter ist eine Impulsantwort, die bei der Faltung mit einem Eingangssignal, nur die tiefen Frequenzanteile im Eingangssignal erhält und die hohen Frequenzanteile „herausfiltert“. Es gibt viele verschiedene Varianten von Tiefpassfiltern. Sie alle bestehen typischerweise aus einer Anzahl an benachbarten, (überwiegend) positiven Werten. Die einfachste Form eines Tiefpassfilters besteht aus einem Rechteckimpuls, der z.B. über zwei Heaviside-Funktionen beschrieben werden kann (s. Abb. 1.24):

Filterkerne

$$y[n] = x[n] * \frac{1}{11} (\Theta(n) - \Theta(n-11))$$

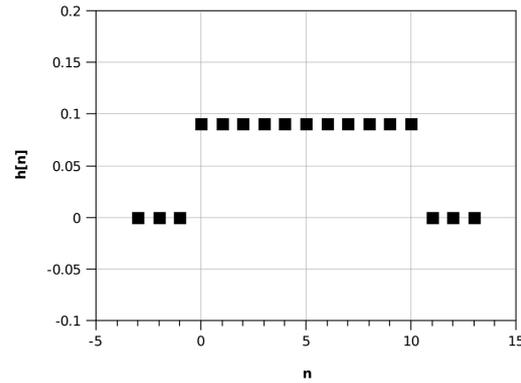


Abbildung 1.24: Ein einfacher, rechteckförmiger Tiefpassfilter.

Der Faktor von $\frac{1}{11}$ sorgt in diesem Fall dafür, dass die Fläche des Rechteckimpulses eins wird und somit nur eine Tiefpassfilterung und nicht etwa auch eine Verstärkung eintritt. Welche Frequenzen der Filter noch passieren lässt und welche Frequenzen herausgefiltert werden, hängt von der Breite des jeweiligen Filters ab. Je breiter der Filter ist, desto tiefer sind die Frequenzen, die herausgefiltert werden.

Hochpassfilter

Konstruktion aus
Tiefpassfilter

Ein Hochpassfilter ist das logische Gegenstück zu einem Tiefpassfilter. Es werden tiefe Frequenzen herausgefiltert und hohe Frequenzen bleiben erhalten. Wie auch bei den Tiefpassfiltern gibt es eine große Vielzahl verschiedener Varianten, wie ein Hochpassfilter aufgebaut sein kann. In der Tat ist es so, dass aus jedem Tiefpassfilter ein Hochpassfilter „konstruiert“ werden kann. Auch in diesem Fall spielt das *Superpositionsprinzip* (vgl. Def.: 1.18) eine wichtige Rolle. Wie bereits oben beschrieben, ist die Deltafunktion die Identität bzgl. der Faltung. Alle Werte im Eingangssignal werden bei Faltung mit der Deltafunktion ohne Veränderung in das Ausgangssignal übernommen. Ein Tiefpassfilter hingegen übernimmt nur die tiefen Frequenzanteile in das Ausgangssignal. Dank des Superpositionsprinzips können diese beiden Vorgänge additiv miteinander verknüpft werden, so dass eine Impulsantwort, die aus einer Deltafunktion besteht von der ein Tiefpassfilter subtrahiert wurde, nur die hohen Frequenzanteile eines Eingangssignals in das Ausgangssignal übernimmt.

Für das vorherige Beispiel eines einfachen, rechteckigen Tiefpassfilters sieht der entsprechende Hochpassfilter folgendermaßen aus (s. Abb. 1.25):

$$y[n] = x[n] * \left(\delta[n - 5] - \frac{1}{11} (\Theta(n) - \Theta(n - 10)) \right)$$

An der Verschiebung der Deltafunktion ist zu erkennen, dass diese üblicherweise symmetrisch im Zentrum des Tiefpassfilters positioniert wird.

Die Konstruktion eines Hochpassfilters aus einer Deltafunktion und einem Tiefpassfilter ist ein Beispiel dafür, wie die mathematischen Eigenschaften

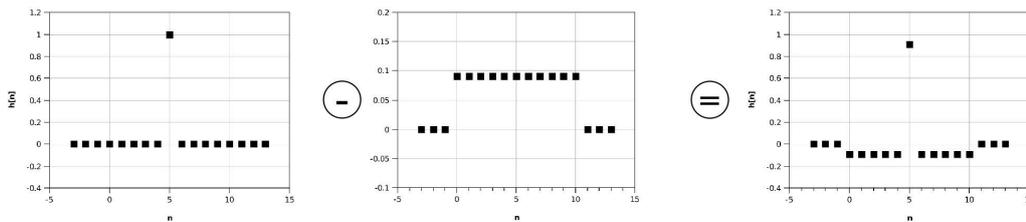


Abbildung 1.25: Ein einfacher Hochpassfilter, der über die Subtraktion eines rechteckigen Tiefpassfilters von einer über dem Tiefpassfilter zentrierten Deltafunktion konstruiert wurde.

linearer Systeme genutzt werden können, um Filterkerne gezielt aus mehreren Komponenten zu konstruieren.

1.4.3 Korrelation

Zum Abschluss dieses Abschnitts soll noch auf ein bekanntes Verfahren der Signalverarbeitung hingewiesen werden, das sich zwar konzeptionell von der zuvor beschriebenen Faltung unterscheidet, mathematisch jedoch fast auf identische Weise berechnet wird. Dementsprechend häufig kommt es diesbezüglich zu Verwechslungen und Missverständnissen.

Korrelation \neq
Konvolution

Mathematisch ist die Korrelation eine Operation, die genau wie die Konvolution zwei Signale miteinander kombiniert. Formal ist die Korrelation wie folgt definiert:

Definition 1.26 (Korrelation).

Es sei $x[n]$ ein Signal der Länge N und $t[n]$ ein Signal der Länge M . Dann ist die **Korrelation von x und t** gegeben durch

$$y[i] = \sum_{j=0}^{M-1} t[j] x[i + j], \quad i \in \{0, \dots, M + N - 2\}.$$

Das resultierende Signal $y[n]$ hat eine Länge von $N + M - 1$ und wird als Kreuzkorrelation bezeichnet. Sind x und t identisch, wird das resultierende Signal auch Autokorrelation genannt.



Auf den ersten Blick könnte der Unterschied zur Faltung (s. Def. 1.25) fast übersehen werden. Innerhalb des Summanden hat sich nur die Subtraktion der Indizes in eine Addition verwandelt. Diese Änderung führt dazu, dass die oben im Abschnitt „Verzögerung“ (1.4.2) beschriebene, gespiegelte Verwendung des Faltungskerns im Kontext der Korrelation nicht auftaucht und das dem Faltungskern entsprechende Signal t „richtig herum“ verwendet wird.

Faltungskern nicht
gespiegelt

Konzeptionell bestimmt die Korrelation für jede Position n eines Eingangssignals $x[n]$ wie *ähnlich* ein Mustersignal $t[n]$ dem Eingangssignal an dieser

Detektion von
bekannten
Signalmustern

Position ist. Anwendung findet diese Form der Signalverarbeitung z.B. dann, wenn die Position eines bekannten Signalmusters in einem Eingangssignal detektiert werden soll. Ein Beispiel hierfür ist z.B. die Detektion von Unterwasserobjekten mittels Sonar. Bei modernen Geräten wird hierzu nicht nur ein einfacher Puls ausgesandt, sondern ein sogenanntes „Chirp“-Signal – ein Signal mit einer ansteigenden Frequenz. Dieses Signal wird von entfernten Objekten zurückreflektiert und von einer Empfangseinheit wieder aufgenommen. Das aufgenommene Signal ist in der Regel stark verrauscht und enthält eines oder mehrere Echos des zuvor ausgesandten „Chirp“-Signals. Mittels einer Kreuzkorrelation lassen sich nun diese Echos identifizieren und über die Verschiebung die Laufzeit der Signale und somit die Distanz der jeweiligen Objekte ermitteln.

Da sich mathematisch die Korrelation von einer Faltung nur durch einen nicht gespiegelten Faltungskern unterscheidet, kann durch eine entsprechende Vorverarbeitung, d.h. eine Spiegelung des Faltungskerns, eine Korrelation mit den Methoden einer Faltung berechnet werden.

Der vorhergehende Abschnitt stellte Ihnen die Operation der (diskreten) Faltung im Kontext der Signalverarbeitung vor. Neben den mathematischen Grundlagen und Eigenschaften der Faltung wurde auch die enge Beziehung zwischen den zuvor vorgestellten linearen Systemen und der Faltungsoperation erläutert. Anhand wichtiger, elementarer Impulsantworten wurde das Verständnis für diesen Zusammenhang weiter vertieft. Abschließend wurde die Operation der Korrelation vorgestellt, die mathematisch der Faltung sehr ähnlich, konzeptionell jedoch eigenständig ist.

1.5 Faltung kontinuierlicher Signale

Lernziele:

Der letzte Abschnitt dieser Kurseinheit wirft einen kurzen Blick darauf, wie die bisher für diskrete Signale beschriebenen Methoden und Konzepte für kontinuierliche Signale formuliert werden können. Im Fokus steht dabei der sogenannte *Dirac-Impuls*, der das kontinuierliche Pendant der zuvor beschriebenen diskreten Deltafunktion ist.

lückenlose
mathematische
Beschreibung

Die Beschreibung der in dieser Kurseinheit vorgestellten Methoden und Konzepte hat sich bislang auf *diskrete* Signale konzentriert. Zum Abschluss dieser Kurseinheit soll nun auch ein Blick auf *kontinuierliche* Signale geworfen werden. Auch wenn kontinuierliche Signale nicht in einem Computer verarbeitet werden können, bieten sie die Möglichkeit, die zugrunde liegenden Prozesse und die korrespondierenden analogen Signale mathematisch zu modellieren. Auf diese Weise wird eine lückenlose mathematische Beschreibung ausgehend vom zugrunde liegenden Prozess über die Analog-Digital-Wandlung hin zum diskreten Signal im Computer möglich.

1.5.1 Dirac-Impuls

Analog zum Vorgehen bei diskreten Signalen können auch kontinuierliche Signale als Folgen von Impulsen aufgefasst werden. Im Gegensatz zu diskreten Signalen besteht jedoch das Problem, dass ein kontinuierliches Signal selbst auf einem endlichen Intervall, z.B. im Bereich von 0 bis 1, unendlich viele Werte besitzt. Im Umkehrschluss bedeutet dies, dass der im diskreten Fall verwendete Einheitsimpuls im kontinuierlichen Fall *unendlich schmal* sein muss. Darüber hinaus sollte es – analog zum diskreten Fall – möglich sein, dass ein kontinuierliches Signal als eine Kombination von skalierten und verschobenen Einheitsimpulsen aufgefasst werden kann. Ein Impuls, der diesen beiden Forderungen genügt, ist der sogenannte Dirac-Impuls:

unendlich viele
Werte → unendlich
schmaler Impuls

Definition 1.27 (Dirac-Impuls).

Der **Dirac-Impuls** $\delta(t)$ ist über die folgenden Eigenschaften definiert:

$$\delta(t) = \begin{cases} \infty & , \text{ falls } t = 0, \\ 0 & , \text{ falls } t \neq 0, \end{cases} \quad t \in \mathbb{R}$$

mit

$$\int_{-\infty}^{+\infty} \delta(t) dt = 1.$$



Der Dirac-Impuls⁴² nimmt demnach nur genau an Stelle 0 den Wert *unendlich* an und hat ansonsten an allen anderen Stellen den Wert null. Die Fläche des Impulses an Stelle 0 beträgt genau 1. Es mag auf den ersten Blick unklar sein, warum der Dirac-Impuls ein geeignetes kontinuierliches Pendant zur diskreten Deltafunktion darstellen soll. Zur Erinnerung: Im Falle der diskreten Deltafunktion $\delta[n]$ wurde diese mit einem Signal $x[n]$ multipliziert und so der Wert des Signals x an Stelle 0 isoliert:

Dirac-Impuls →
Pendant zur
diskreten
Deltafunktion

$$x[n] \cdot \delta[n] = \begin{cases} x[0] & \text{wenn gilt } n = 0, \\ 0 & \text{wenn gilt } n \neq 0, \end{cases} \quad n \in \mathbb{Z}.$$

Würde man dieses Vorgehen 1-zu-1 auf den Dirac-Impuls übertragen, so würde der Wert eines kontinuierlichen Signals $x(t)$ an Stelle 0 mit dem Wert $\delta(0) = \infty$ (unendlich) multipliziert. Der ursprüngliche Wert des Signals x an Stelle 0 würde so verloren gehen. Dieses Vorgehen erscheint also nicht den gewünschten Zweck zu erfüllen.

Es zeigt sich, dass mit Hilfe des Dirac-Impulses die einzelnen Werte eines kontinuierlichen Signals auf eine etwas andere Weise isoliert werden:

$$\int_{-\infty}^{+\infty} x(t) \cdot \delta(t) dt = x(0).$$

⁴²Manchmal wird der Dirac-Impuls analog zum diskreten Fall auch als Deltafunktion bezeichnet.

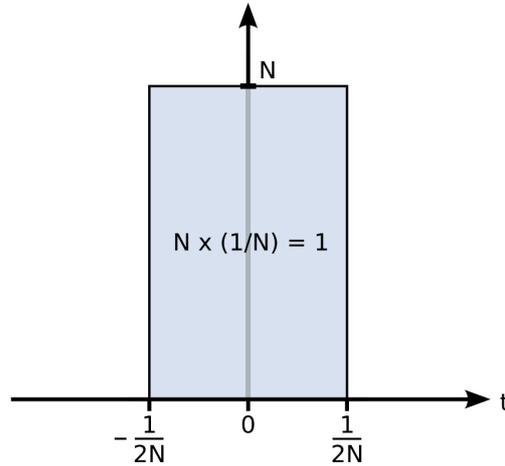


Abbildung 1.26: Impuls mit Einheitsfläche.

Ausblend-
eigenschaft

Da die Fläche des Dirac-Impulses genau eins beträgt und der Dirac-Impuls nur genau an Stelle 0 einen Wert ungleich null hat, wird diese Fläche ausschließlich durch den Wert des Signals $x(t)$ an Stelle 0 skaliert. Mit diesem „Umweg“ über das Integral des Produktes von Signal und Dirac-Impuls kann schließlich der Wert des Signals an Stelle 0 isoliert werden. Dementsprechend wird die Wirkung dieses Integrals mitunter auch als *Ausblendeigenschaft* des Dirac-Impulses bezeichnet.

Es bleibt die Frage, ob es überhaupt möglich ist, einen derartigen Impuls (mathematisch) konstruieren zu können. Immerhin beschreibt die oben genannte Definition des Dirac-Impulses einen Impuls, der nicht nur unendlich schmal und unendlich hoch ist, sondern gleichzeitig auch eine Fläche von genau 1 haben soll. Intuitiv könnte man gar vollständig anzweifeln, ob ein derartiger Impuls überhaupt eine Fläche habe!

Konstruktion eines
Dirac-Impulses

Eine Möglichkeit, einen derartigen Impuls zu konstruieren, zeigt Abbildung 1.26. Sie zeigt einen rechteckigen Impuls, dessen Höhe und Breite über einen Parameter n gesteuert werden:

$$\delta_n(t) = \begin{cases} n & \text{wenn gilt } -\frac{1}{2n} \leq t \leq \frac{1}{2n}, \\ 0 & \text{sonst.} \end{cases}$$

Die Fläche dieses rechteckigen Impulses ist unabhängig vom konkreten Wert des Parameters n immer eins. Für $n \rightarrow \infty$ („ n gegen unendlich“) wird der Impuls unendlich schmal, unendlich hoch und hat weiterhin eine Fläche von genau 1. Im Grenzwert zeigt dieser Impuls demnach das gleiche Verhalten wie ein Dirac-Impuls:

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} x(t) \cdot \delta_n(t) dt = x(0).$$

Abtastung

Ein weiterer Aspekt des Dirac-Impulses besteht darin, dass mit seiner Hilfe die Verbindung zwischen einem kontinuierlichen Signal und seinem abgetasteten, diskreten Gegenpart beschrieben werden kann:

Merksatz 1.5 (Abtastung eines kontinuierlichen Signals).

Wird ein kontinuierliches Signal $x(t)$ in regelmäßigen Abständen Δt abgetastet, so kann das entstehende diskrete Signal $x[n]$ mittels der Ausblendeigenschaft des Dirac-Impulses beschrieben werden:

$$x[n] = \int_{-\infty}^{+\infty} x(t) \cdot \delta(t - n\Delta t) dt, \quad n \in \mathbb{Z}.$$



1.5.2 Faltung

In gleicher Weise wie über die diskrete Deltafunktion $\delta[n]$ durch Verschiebung und Skalierung jeder andere Impuls abgebildet werden kann, kann dies im kontinuierlichen Fall über den Dirac-Impuls $\delta(t)$ geschehen. Damit wird es analog zum diskreten Fall möglich, ein kontinuierliches Signal als Kombination von Impulsen aufzufassen. In diesem Zusammenhang können die für diskrete Signale gemachten Überlegungen aufgegriffen und auf kontinuierliche Signale übertragen werden. So kann auch im kontinuierlichen Fall ein lineares System S über seine Impulsantwort $h_S(t)$ eindeutig beschrieben werden:

Signal \rightarrow
Kombination von
Impulsen

$$\delta(t) \rightarrow \boxed{S} \rightarrow h_S(t)$$

Das Ausgangssignals $y(t)$ eines linearen Systems S kann dementsprechend über die Faltung des Eingangssignals $x(t)$ mit der Impulsantwort $h_S(t)$ des Systems bestimmt werden. Im kontinuierlichen Fall ist die Faltung dabei wie folgt definiert:

Definition 1.28 (Faltung kontinuierlicher Signale).

Seien die kontinuierlichen Signale $x(t)$ und $h(t)$ gegeben. Dann ist die **Faltung** $x(t) * h(t)$ der beiden Signale definiert durch:

$$y(t) = \int_{-\infty}^{+\infty} h(j) \cdot x(t - j) dj$$



Diese Definition ist damit das kontinuierliche Pendant zur Definition 1.25.

Analog können mittels des Dirac-Impulses auch die in Abschnitt 1.4.2 beschriebenen elementaren Impulsantworten nachvollzogen werden:

elementaren
Impulsantworten

Identität: $x(t) * \delta(t) = x(t)$

Verstärkung: $x(t) * (k \cdot \delta(t)) = k \cdot x(t)$

Verzögerung: $x(t) * \delta(t + s) = x(t + s)$

Echo (Beispiel): $x(t) * (\delta(t) + 0.5 \cdot \delta(t + 3)) = x(t) + 0.5 \cdot x(t + 3)$

usw.

Es ist zu erkennen, wie die zuvor für den diskreten Fall beschriebenen Methoden und Konzepte im Wesentlichen 1-zu-1 für den kontinuierlichen Fall übernommen werden können. Dies ist insofern bemerkenswert, da der zuvor beschriebene Dirac-Impuls eben *keine* 1-zu-1-Übertragung der diskreten Deltafunktion für die Verwendung mit kontinuierlichen Signalen ist.

Der letzte Abschnitt dieser Kurseinheit vermittelte Ihnen einen Eindruck davon, wie die in dieser Kurseinheit vornehmlich für diskrete Signale beschriebenen Methoden und Konzepte für kontinuierliche Signale formuliert werden können. Der zentrale Aspekt war diesbezüglich die Dirac-Funktion, die das kontinuierliche Gegenstück zur diskreten Deltafunktion ist. Insbesondere die *Ausblendeigenschaft* des Dirac-Impulses, mit der einzelne Werte eines kontinuierlichen Signals isoliert werden können, ist in diesem Zusammenhang von großer Bedeutung. Sie bildet die Grundlage dafür, dass auch im kontinuierlichen Fall lineare Systeme über ihre Impulsantwort vollständig beschrieben werden können, da sich jedes Signal als eine Kombination von verschobenen und skalierten Dirac-Impulsen darstellen lässt. Darüber hinaus ermöglicht die Ausblendeigenschaft die mathematische Beschreibung der Analog-Digital-Wandlung – also der Beziehung zwischen einem kontinuierlichen Signal und seinem diskreten Gegenstück.

Literaturverzeichnis

- [Smi02] SMITH, Steven: *Digital Signal Processing: A Practical Guide for Engineers and Scientists*. Book and CD ROM. Newnes, 2002. – ISBN 075067444X