

## 2. Wavelets

### 2.1. Die Entwicklungsgeschichte der Wavelet-Transformation

Die geschichtliche Entwicklung der Wavelets ist ein Beispiel dafür, wie Ansätze aus ganz verschiedenen Gebieten von Wissenschaft und Technik zu einer großen übergreifenden Idee führen können. Die vielen verzweigten Wurzeln der Wavelet-Transformation machen es unmöglich, hier einen umfassenden Überblick zu geben. Die folgende Beschreibung der Entwicklung der Wavelets ist im wesentlichen einem Artikel von *Ingrid Daubechies* [1] und anderen einführenden Artikeln [2–4] entnommen. Für ausführlichere Darstellungen sei auf [5] und die Referenzen in [1] verwiesen.

Das erste Wavelet wurde bereits 1910 von *Alfred Haar* erwähnt [6]. Das einfache Haar-Wavelet besitzt zwar einen kompakten Träger (siehe Abschnitt 2.3.4) und ist damit prinzipiell für eine lokalisierende Signalanalyse geeignet. Es ist allerdings nicht stetig, was seine Anwendbarkeit einschränkt. Bereits seit dieser Zeit wurden erste Versuche unternommen, Signale nicht nur mit der Fourier-Analyse, sondern auch mit Hilfe anderer Methoden zu untersuchen. Der Physiker *Paul Levy* beschäftigte sich nach 1930 mit der Brownschen Molekularbewegung und benutzte eine Art Multiskalen-Analyse auf Grundlage der Basisfunktion von *A. Haar*, um Details aus Meßsignalen herauszuarbeiten.

Auch in der weiteren Entwicklung gingen wichtige Impulse oft nicht von Mathematikern, sondern von Wissenschaftlern anderer Fachrichtungen aus, die Lösungen für konkrete Probleme bei der Signalanalyse suchten.

Anfang der 80er Jahren entwickelte *Jean Morlet* eine Alternative zur gefensterten Fourier-Transformation [7], da er als Geophysiker mit Signalen arbeitete, die Merkmale mit ganz verschiedenen Frequenzen und Positionen im Zeitbereich enthielten. Um diese Merkmale mit einer Transformation erfassen zu können, verwendete er als Basis gefensterte Kosinusfunktionen. Durch Stauchen oder Dehnen konnte er sie der zu analysierenden Frequenz anpassen und durch Verschieben der Basisfunktionen

auf der Zeit-Achse das Signal lokal im Bezug auf einen bestimmten Frequenzanteil analysieren. *Morlet* nannte seine Funktionen „*Wavelets of constant shapes*“, die später einfach als „*Wavelets*“ bezeichnet wurden und somit namensgebend waren.

Die Absicht, seine Wavelets auf ein mathematisches Fundament zu stellen, konnte *Morlet* verwirklichen, nachdem er in Kontakt mit *Alex Grossmann* gekommen war. Dieser hatte als Theoretischer Physiker auf dem Gebiet der Quantenmechanik vergleichbare Probleme zu bewältigen wie *Morlet* und erkannte Ähnlichkeiten zwischen Wavelets und dem von ihm verwendeten „Formalismus kohärenter Zustände“. Neben der Entwicklung eines exakten Formalismus [8] untersuchten sie auch verschiedene Anwendungen<sup>1</sup>.

In dieser Zeit hörte der Mathematiker *Yves Meyer* durch einen Zufall von den Arbeiten *Morlets* und *Grossmanns*. Darin bemerkte er starke Übereinstimmungen mit Ansätzen von *A. Calderón* aus den 60er Jahren [10] im Bereich der „Harmonischen Analyse“, einem Zweig der Mathematik, der sich auf Grundlage der Fourier-Analyse entwickelt hatte. *Meyer* beschäftigte sich z.B. mit der Detektion von Singularitäten und anderen lokal auftretenden Artefakten in Signalen und war daher sehr an Wavelets und ihren Lokalisierungseigenschaften interessiert. Innerhalb der Zusammenarbeit mit *Morlet* und *Grossmann* konstruierte er eine orthogonale Wavelet-Basis mit guten Lokalisierungseigenschaften im Zeit- und Frequenzbereich<sup>2</sup>. Unabhängig davon und mit anderen Techniken wurden Wavelets auf Basis von Splinefunktionen entwickelt (*G. Battle* und *P. G. Lemarié* [12, 13], *C. K. Chui* [14]). Die sogenannten Spline-Wavelets finden seitdem vielfältige Anwendungen.

Durch einen weiteren Zufall erkannte *Stéphane Mallat* 1986 Verbindungen zwischen der „Wavelet-Philosophie“ und Algorithmen zur Bildverarbeitung und Bildkomprimierung mittels Multiskalen-Darstellungen (z.B. Schema der *Laplace*-Pyramide [15]). Zusammen mit *Meyer* erarbeitete er die mathematische Formulierung einer Multiskalen-Analyse mit Wavelets [16, 17]. Diese Art von Zerlegung läßt sich mit einem einfachen rekursiven Algorithmus durchführen, der vergleichbar ist mit einer sukzessiven Anwendung von digitalen Filtern.

Damit ergab sich erneut eine Parallele zu einem ganz anderen Forschungsgebiet, nämlich der digitalen Signalverarbeitung im Bereich der Elektro- und Nachrichtentechnik. Dort wurde bereits in den 70er Jahren an Zerlegungen von Signalen in Frequenzbänder (Subband-Kodierung) gearbeitet. Mit Paaren von Hoch- und Tiefpassfiltern (Konjugate Quadratur Filtern (CQF)) konnten 1983 exakte Zerlegungen

---

<sup>1</sup>Später wurde klar, daß zu einem früheren Zeitpunkt *E. Aslaksen* und *J. Klauder* einen vergleichbaren Ansatz in anderem Zusammenhang entwickelt hatten [9].

<sup>2</sup>Auch hier gab es bereits einen Ansatz in dieser Richtung von *J. O. Strömberg*, dessen Bedeutung aber nicht erkannt wurde [11].

und Rekonstruktionen erreicht werden, wie dies später auch mit Wavelets gelang.

Einen weiteren wichtigen Durchbruch stellten die von *Ingrid Daubechies* konstruierten orthogonalen Wavelets mit kompaktem Träger dar [18, 19], die auch in der vorliegenden Arbeit Anwendung finden.

## 2.2. Überblick über Anwendungen von Wavelets

So vielschichtig und verzweigt wie die Geschichte der Wavelet-Transformation sind auch die Gebiete, in denen Wavelets angewendet werden. Aufgrund des adaptiven Charakters der Wavelet-Transformation lassen sich nicht nur bestimmte Signale, sondern ganze Klassen von unterschiedlichen Signalen analysieren. Die Vielzahl von zur Verfügung stehenden Wavelets und Algorithmen erweitert nochmals die Anwendungsmöglichkeiten.

Im folgenden wird ein Überblick über Anwendungen der Wavelet-Transformation gegeben. Zwar steht hierbei die Analytische Chemie im Mittelpunkt, aber auch Arbeiten aus anderen wissenschaftlichen Bereichen werden berücksichtigt, sofern sie interessante Ansätze oder Anwendungen beinhalten. An typischen Beispielen soll damit gezeigt werden, in welchen völlig unterschiedlichen Anwendungsfeldern Wavelets bereits genutzt werden.

### 2.2.1. Anwendungen von Wavelets in der Analytischen Chemie

Ein kürzlich publizierter Übersichtsartikel [20] zeigt die Vielfalt von Anwendungen der Wavelets in der Analytischen Chemie. Die Betrachtung der Anzahl von bisher pro Jahr erschienenen Publikationen macht deutlich, daß Wavelets in diesem Bereich erst in den letzten drei Jahren mit stark steigender Tendenz an Bedeutung gewonnen haben (siehe Abbildung 2.1). Parallel zu dieser Entwicklung wurden besonders im Hinblick auf chemometrische Anwendungen einige einführende Darstellungen und Tutorien zu Wavelets veröffentlicht [2, 4, 21, 22].

Generell stehen zwei Vorgehensweisen offen, um Wavelets für die Verarbeitung chemisch-analytischer Signale einzusetzen:

- **Filterung von Signalen**

Nach der Transformation eines Signals wird dieses im Waveletbereich verändert. Mit der Rücktransformation wird dann ein gefiltertes Signal erhalten, das weiterverarbeitet werden kann.

Die wichtigste derartige Anwendung von Wavelets im Bereich der Analytischen Chemie ist die Glättung von Spektren und Chromatogrammen.

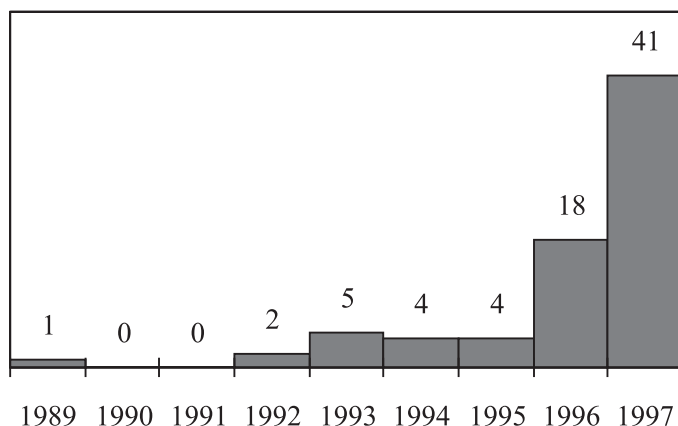


Abbildung 2.1.: Anzahl von Publikationen über Anwendungen von Wavelets in der Chemie zwischen 1989 und 1997 [20].

- **Direkte Signalverarbeitung mit Wavelet-Koeffizienten**

Ein transformiertes Signal wird im Waveletbereich durch die Wavelet-Koeffizienten repräsentiert, die direkt zur weiteren Signalverarbeitung genutzt werden können; vor der Weiterverarbeitung des Signals erfolgt also keine Rücktransformation.

So lassen sich z.B. NIR-Spektren, die zur Identifizierung von Kunststoffen herangezogen werden, nach der Wavelet-Transformation auf der Basis weniger Wavelet-Koeffizienten klassifizieren. Damit ist eine beträchtliche Datenreduktion verbunden, ohne daß wichtige Signalanteile verlorengehen.

Welcher Weg auch gewählt wird, in beiden Fällen werden die speziellen Eigenschaften der Wavelet-Transformation bei der Analyse von Frequenzanteilen genutzt. Im Gegensatz zur Fourier-Transformation können die im Wavelet-Raum aufgetrennt vorliegenden Frequenzanteile weiterhin ihren ursprünglichen Positionen auf der Zeit-Achse (z.B. Retentionszeit, Wellenzahl oder -länge) zugeordnet werden<sup>3</sup>. Diese Lokalisierungseigenschaften sind besonders für die Verarbeitung chemisch-analytischer Signale interessant, bei denen relevante Signalanteile oft stark lokalisiert sind, wie im Falle von Spektren bzw. Chromatogrammen mit ihren Banden bzw. Peaks. Dementsprechend liegt der Schwerpunkt der publizierten chemischen Anwendungen im Bereich der verschiedenen spektrometrischen und chromatographischen Methoden.

---

<sup>3</sup>Die Genauigkeit der Zuordnung hängt aufgrund der Heisenbergschen Unschärferelation von dem verwendeten Zeit-Frequenz-Fenster ab.

### Filterung chemisch-analytischer Signale

Die Filterung chemisch-analytischer Signale hat in vielen Fällen eine Glättung bzw. die Reduzierung von Rauschen zum Ziel. In einigen allgemeiner ausgerichteten Arbeiten wurden verschiedene Verfahren auf der Basis von Wavelets im Bereich der Infrarot-Spektrometrie (IR) [4, 23, 24], der Nahinfrarot-Spektrometrie (NIR) [25] und der Gaschromatographie (GC) [26, 27] untersucht. Dabei konnten z.T. deutlich bessere Ergebnisse als mit klassischen Filterverfahren, wie z.B. Fourier- und Savitzky-Golay-Filtern, erzielt werden. Außerdem erwiesen sich Spline-Wavelets in einem Vergleich mit Fourier-Filtern als vorteilhaft [28].

Generell sind der verwendete Algorithmus und insbesondere die Wahl des Wavelet-Typs entscheidend für die Güte der Glättungen [29]. Ein spezieller Aspekt ist die Bestimmung und Reduzierung von Rauschen, dessen Amplitude von der des zugrundeliegenden Signals abhängt (*heteroscedastic noise*) [30].

V. J. Barclay et al. [23] unterscheiden zwischen einer klassischen Glättung, bei der hochfrequentes Rauschen vermindert wird, und einer amplitudenabhängigen Rauschverminderung, die Signalanteile aller Frequenzen einschließt:

- Bei einer Glättung (*smoothing*) im Wavelet-Raum werden Koeffizienten, die hochfrequente Signalanteile repräsentieren, unabhängig von ihrer Amplitude entfernt.
- Die Rauschunterdrückung (*denoising*) bezieht sich auf den gesamten Frequenzbereich und richtet sich nach den Beträgen der Wavelet-Koeffizienten. Da der Betrag ein Maß für den Informationsgehalt eines Koeffizienten darstellt, können jene mit kleinen Beträgen vernachlässigt werden, ohne daß relevante Signalanteile verlorengehen.

Insbesondere durch *denoising* kann meist ein beträchtlicher Teil der Wavelet-Koeffizienten eliminiert und können Komprimierungsraten von 70 % bis über 90 % erreicht werden. Wie auch in dieser Arbeit gezeigt wird, bleiben dabei meist über 97 % der im Signal enthaltenen Energie (Information) erhalten. Aufgrund der guten Approximationseigenschaften vieler Wavelet-Funktionen ist außerdem sichergestellt, daß die Rekonstruktion eines komprimierten Signals das ursprüngliche Signal sehr gut wiedergibt.

### Spektrometrische Anwendungen von Wavelets

Die Möglichkeiten zur Kodierung von Signalen mit wenigen Wavelet-Koeffizienten können zur Komprimierung von IR- und NIR-Spektren genutzt werden [31–34],

um z.B. effiziente Bibliotheken zu erstellen [35, 36]. Neben der reinen Reduzierung des Speicherbedarfs wird zudem eine raschere Bibliothekssuche erreicht. Die Lokalisierungseigenschaften der Wavelets ermöglichen es außerdem, Peak-Listen für IR-Spektren zu generieren, die u.a. zur Erstellung von Datenbanken herangezogen werden können [37]. Eine spezielle Anwendung stellt die Bestimmung der Bestandteile von Mineralien anhand von Wavelet-gefilterten IR-Spektren dar [38].

Die NIR-Spektrometrie wird vielfach und routinemäßig für quantitative Analysen eingesetzt. Quantitative Kalibrationsmodelle lassen sich auch direkt auf der Basis von Wavelet-Koeffizienten erstellen. Eine Möglichkeit hierfür ist die Verarbeitung der Koeffizienten mit dem klassischen PLS-Verfahren (*Partial Least Squares Regression*) [34, 39, 40]. Dazu werden die durch Filterung reduzierten Koeffizientensätze verwendet, um störende Anteile zu eliminieren und eine Datenreduktion zu erzielen. Im Rahmen der vorliegenden Arbeit wurde ein Verfahren entwickelt, das wenige ausgewählte Wavelet-Koeffizienten durch eine multivariate Regression mit der zu kalibrierenden Eigenschaft verknüpft [41, 42]. Die Selektion einer geeigneten Kombination von Wavelet-Koeffizienten erfolgt dabei mittels eines Genetischen Algorithmus.

Zu dem im Bereich der NIR-Spektrometrie vieldiskutierten Thema des Kalibrationstransfers wurde bisher im Zusammenhang mit Wavelets nur eine Arbeit publiziert. Danach eröffnet die Standardisierung von NIR-Spektren im Waveletbereich Möglichkeiten zur Kompensierung lokal auftretender spektraler Unterschiede [43].

Auch in der UV-VIS-Spektrometrie werden Wavelets zur Filterung [44] und Komprimierung [45] von Spektren eingesetzt. Eine konkrete Anwendung ist die Komprimierung und Klassifizierung von UV-Spektren eines Satelliten zur Beobachtung astronomischer Objekte [46].

Besonders die Astronomie mit ihren vielfältigen spektroskopischen Beobachtungsmethoden ist ein breites Anwendungsgebiet für Wavelet-Methoden, wie eine Übersicht zeigt [47]. So wurden z.B. IR-Spektren einer astronomischen Strahlungsquelle ausgewertet [48] oder die Anzahl der periodisch wechselnden Sonnenflecken analysiert [49].

In Verbindung mit anderen als den bisher genannten spektroskopischen Methoden gibt es heute nur wenige Wavelet-Anwendungen. Dazu zählen die Klassifizierung [50] und Filterung [51] von NMR-Spektren, die Filterung der photoakustischen Spektren von PVC-Materialien [52] und die Ermittlung der Feinstruktur von Röntgenabsorptionsspektren (*extended X-ray absorption fine structure*, EXAFS) [53]. Kürzlich wurden außerdem zwei zusammenhängende Artikel zur Charakterisierung der Profile von Absorptionsbanden in der Graphitrohr-Atomabsorptionsspektrometrie (GF-AAS) veröffentlicht [54, 55].

### **Chromatographische Anwendungen von Wavelets**

Die allgemeiner ausgerichteten Publikationen zur Glättung von Chromatogrammen in der Gaschromatographie wurden bereits erwähnt [26, 27]. Darüber hinaus wurden in weiteren Arbeiten die Auswirkungen der Filterung mit Wavelets auf die Integration von Peaks, die anschließende Kalibrierung und die Bestimmung von Nachweisgrenzen untersucht [56, 57]. Wavelets eignen sich auch zur Trennung von überlappenden Peaks in Chromatogrammen [58, 59].

Im Bereich der Hochleistungsflüssigchromatographie (HPLC) steht ebenfalls die Glättung von Chromatogrammen im Mittelpunkt [59, 60], um z.B. die Klassifizierung von organischen Verunreinigungen in pharmazeutischen Wirkstoffen zu verbessern [61].

### **Weitere chemisch-analytische Anwendungen**

In einer bereits 1992 publizierten Arbeit wird die chemisch-analytische Anwendung der Wavelet-Transformation in der Fließinjektions-Analyse (FIA) beschrieben [62]. Durch Abtrennung von störendem Rauschen und Basislinienanteilen können Peak-Intensitäten genauer bestimmt und damit z.B. Nachweisgrenzen verbessert werden.

Bei den elektroanalytischen Methoden konzentrieren sich die Aktivitäten auf die Voltammetrie [63–66], wie auch die Übersicht in [20] zeigt, in der außerdem einige Originalarbeiten aus China aufgeführt sind.

Die Möglichkeiten zur Signalverarbeitung mit Wavelets lassen sich in Analogie zur Bildverarbeitung auch für zweidimensionale chemisch-analytische Signale nutzen. *M. Grasserbauer et al.* demonstrieren in mehreren Arbeiten die Entrauschung und Kontrasterhöhung von SIMS-Aufnahmen (Sekundärionenmassenspektrometrie) [67–69]. Bei der SIMS-Technik wird eine Probe lokal mit energiereichen Ionen beschossen, um Teilchen aus der Oberfläche „herauszuschlagen“. Wird eine Probenfläche in dieser Weise abgefahren, erhält man mit der massenspektrometrischen Detektion der Sekundärteilchen eine zweidimensionale Analyse der chemischen Zusammensetzung und kann zudem Tiefenprofile erstellen.

In den Bereich der Materialforschung fällt die Wavelet-basierte Klassifizierung von Mikroskop-Aufnahmen zur Untersuchung von Korrosion [70, 71].

### **2.2.2. Weitere Wavelet-Anwendungen**

Analytische Methoden werden in zunehmendem Maße automatisiert, um z.B. Produktionsprozesse zu überwachen und die Steuerung der Anlagen zu optimieren. Bei

der Auswertung dieser Prozeßdaten geht es u.a. um die Analyse langfristiger Trends und kurzzeitig auftretender Ereignisse im Sinne von Störungen.

Auch hier können Wavelets erfolgreich zur Filterung der Daten eingesetzt werden, um eine Fehlerdiagnose durchzuführen, Trends in Steuerungssystemen zu erkennen [72–77] und Entscheidungen zu treffen [78]. Außerdem ist eine effiziente Komprimierung der in großen Mengen anfallenden Prozeßdaten unter Erhaltung aller wichtigen Details möglich [79].

Nicht nur im Zusammenhang mit Prozeßdaten ist die Untersuchung von sich zeitlich verändernden Signalen von Interesse [80]. So lassen sich bei der Überwachung von Oberflächengewässern periodische (d.h. saisonale oder tägliche wiederkehrende) und kurzzeitige Veränderungen z.B. der Durchflußmenge [81] und der Wasserqualität [82] unterscheiden.

Auch im medizinischen Bereich sind es oft Signalveränderungen, die relevant für die Datenauswertung sind, wie z.B. bei Elektrokardiogrammen (EKG) und Elektroenzephalogrammen (EEG). Bei einem EKG lassen sich die im Rauschen verborgenen Informationen über die Herzklappenausschläge extrahieren und so Herzrhythmusstörungen erkennen [83]. Mit Hilfe der Wavelet-Analyse von EEG-Daten können beispielsweise Wirkungen von Anästhetika besser verstanden und ihre Dosierung für Narkosen optimiert werden [84]. Weiterhin lassen sich Gehirnaktivitäten durch eine Wavelet-basierte Bearbeitung von Kernresonanzaufnahmen genauer analysieren, da kleine Veränderungen besser detektiert und vom Rauschen getrennt werden [85].

Ganz allgemein bieten Wavelets interessante Möglichkeiten für die Bildverarbeitung und im Bereich der Computergraphik [86,87]. Im Mittelpunkt des Interesses steht die Komprimierung von Bildern zur schnelleren digitalen Übertragung und zur effektiven Speicherung großer Datenmengen [88–90].

Die weitverbreiteten Komprimierungsstandards JPEG (Bilddateien) und MPEG (Videodateien) basieren auf der diskreten Kosinus-Transformation (DCT). Werden hierbei hohe Komprimierungsraten verwendet, treten störende blockförmige Artefakte in den dekomprimierten Bildern auf [83]. Wie der Vergleich von JPEG- und Wavelet-Komprimierung in Abbildung 2.2 auf der folgenden Seite zeigt, läßt sich mit Wavelets bei gleicher Komprimierungsrate ein deutlich besseres Ergebnis erzielen. Speziell bei der Übertragung von Bildern via Internet werden daher Wavelet-Algorithmen in zunehmendem Maße in kommerzielle Hard- und Software-Produkte implementiert [91–94].

Bereits seit 1993 werden bei der amerikanischen Bundespolizei (FBI) Fingerabdrücke digital gespeichert, nachdem mit Wavelet-Methoden die wesentlichen Merkmale extrahiert wurden. Damit wird für jeden Abdruck nur noch 5 % des ursprüng-





Originalbild



JPEG-Komprimierung 1:50



Wavelet-Komprimierung 1:50

Abbildung 2.2.: Komprimierung von Bilddaten mit JPEG- und Wavelet-Verfahren [95].

lichen Speicherplatzes benötigt [3,90]. Eine weitere interessante Anwendung sind Klassifizierungen in Sicherheitssystemen, die auf Vermessungen der Iris im menschlichen Auge basieren [96].

Neben der Bildverarbeitung ist auch die Analyse und Komprimierung von Audio-daten, wie Musik und Sprache, ein Anwendungsfeld für Wavelet-Methoden [88,97].

### 2.2.3. Informationsquellen im Internet

Abschließend wird hier noch auf einige Informationsquellen im Internet hingewiesen, die wiederum zahlreiche Verweise enthalten:

- *A. Graps*, die Verfasserin einer Einführung zur Wavelet-Transformation [3], und *O. I. Karsligil* haben informative Internetseiten erstellt, auf denen viele Hinweise und *Links* zu weiteren Seiten, Publikationen und Programmen mit Wavelet-Bezug zu finden sind [98,99].
- Ausgehend von einem einführenden Buch von *S. Mallat* [100] hat *F. Chaplais* eine „Wavelet-Tour“ zusammengestellt [101].
- Der von *W. Sweldens* herausgegebene *Wavelet-Digest* liefert aktuelle Informationen über Konferenzen, Publikationen, sowie Software und bietet ein Diskussionsforum [102].
- Eine Literaturübersicht über Anwendungen von Wavelets in der Chemie, die über den Inhalt von [20] hinausgeht, findet sich auf der Internetseite von *A. K. M. Leung* [103].

## 2.3. Wavelet- und Fourier-Transformation im Vergleich

Einen Zugang zum Verständnis der Eigenschaften der Wavelet-Transformation bietet ein Vergleich mit der Fourier-Transformation [3, 4, 104]. Beides sind lineare Transformationen unter Erhaltung der Signalenergie, d.h. der Informationsgehalt bleibt unverändert und die Umkehrung der Transformationen liefert exakt das Ausgangssignal. Beiden Verfahren ist ferner gemeinsam, daß für ihre Durchführung schnelle Algorithmen entwickelt wurden und daher die Transformationen bei heute zur Verfügung stehender Rechenleistung für viele Anwendungen in Echtzeit erfolgen kann.

Die wichtigste Eigenschaft beider Transformationen ist, daß die Frequenzanteile eines Signals analysiert werden können. Die Wavelet-Transformation ermöglicht aber im Gegensatz zur Fourier-Transformation die Lokalisierung dieser Anteile auch im Zeit- bzw. Ortsbereich. Damit bleibt die Information erhalten, an welcher Stelle im ursprünglichen Signal ein bestimmter und durch Wavelets charakterisierter Frequenzanteil auftritt. So können Signalanteile mit bestimmten Positionen auf der Zeitachse, wie z.B. einzelne Absorptionsbanden in einem Spektrum, gezielt untersucht werden. Der Hauptgrund hierfür liegt in der Art der zur Wavelet-Analyse verwendeten Basisfunktionen, die von ganz unterschiedlicher Natur sind, als die Sinus- und Kosinusfunktionen der Fourier-Analyse [3].

### 2.3.1. Fourier-Reihen und Fourier-Transformation

Nach der von *Joseph Fourier* (1768–1830) im Jahre 1807 veröffentlichten Theorie der Frequenzanalyse, die den Grundstein für die Harmonische Analysis bildet, lassen sich periodische Funktionen als unendliche Reihe von Sinus- und Kosinusfunktionen unterschiedlicher Frequenzen darstellen. Für eine  $2\pi$ -periodische Funktion  $f \in L_2([0, 2\pi])$ <sup>4</sup> gilt [105]:

$$f(t) = a_0 + \sum_{k=1}^{\infty} (a_k \cos kt + b_k \sin kt) . \quad (2.1)$$

Die Koeffizienten  $a_0$ ,  $a_k$  und  $b_k$  lassen sich als Integrale über das Produkt aus der Funktion  $f$  und der jeweiligen Basisfunktion mit der Frequenz  $k \in \mathbb{N}$  berechnen:

$$a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(t) dt, \quad a_k = \frac{1}{\pi} \int_0^{2\pi} f(t) \cos(kt) dt, \quad b_k = \frac{1}{\pi} \int_0^{2\pi} f(t) \sin(kt) dt .$$

---

<sup>4</sup> $L_2([0, 2\pi])$  bezeichnet den Raum der quadratisch integrierbaren Funktionen im Intervall  $[0, 2\pi]$ . Weitere Details sind in Abschnitt 2.4.1 zu finden.

In komplexer Darstellung ergibt sich mit  $e^{ikt} = \cos(kt) + i \sin(kt)$ :

$$f(t) = \sum_{k \in \mathbb{Z}} c_k e^{ikt} . \quad (2.2)$$

Aufgrund der Eigenschaften der komplexen Exponentialfunktion können die Fourier-Koeffizienten  $c_k$  durch Integration berechnet werden:

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-ikt} dt , \quad k \in \mathbb{Z} . \quad (2.3)$$

Ein Koeffizient  $c_k$  charakterisiert in komplexer Darstellung einen Signalanteil der Frequenz  $k$ . Aus dem Real- und Imaginärteil von  $c_k$  lassen sich die Amplitude und die Phase von  $k$  bestimmen.

Zur Darstellung des Frequenzverhaltens von Signalen  $f \in L_2(\mathbb{R})$ <sup>5</sup> wird nicht die Fourier-Reihe, sondern die Fourier-Transformierte bzw. das Fourier-Integral verwendet:

$$f^\wedge(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt , \quad \omega \in \mathbb{R} . \quad (2.4)$$

Durch die Fourier-Transformation wird die übliche Darstellung eines Signals im Zeitbereich, wie z.B. die eines Spektrums auf einer Wellenlängenskala, durch jene im Frequenzbereich, also auf einer Frequenzskala, ersetzt. Dies ist schematisch in den Abbildungen 2.3a und 2.3b dargestellt und wird im weiteren an zwei Beispielen demonstriert.

In Abbildung 2.4a sind die Amplituden der Fourier-Koeffizienten  $c_k$  (Gleichung 2.3) eines Signals gezeigt, in dem sich zwei Sinusfunktionen mit Frequenzen von 50 Hz bzw. 150 Hz und zusätzliches Rauschen überlagern. Die im ursprünglichen Signal kaum erkennbaren Sinusanteile sind nach der Transformation des Signals in den Frequenzbereich sehr gut identifizierbar.

Die Fourier-Transformation ist allerdings nur bedingt geeignet, um z.B. nichtperiodische und insbesondere kurze (transiente) Signalveränderungen zu analysieren. Der Transformierten derartiger Signale ist nicht mehr zu entnehmen, wo in dem analysierten Signal bestimmte Frequenzanteile auftreten.

Dies wird bei der Analyse des zweiten Beispielsignals in Abbildung 2.4b deutlich. Dabei handelt es sich um ein simuliertes Spektrum, in dem zwei Lorentz-Profile als Absorptionsbanden die eigentliche Information darstellen. Zusätzlich wurden eine Basislinie (ein Polynom 2. Grades) und hochfrequentes Zufallsrauschen überlagert [106]. Die logarithmische Auftragung der Amplituden der Fourier-Koeffizienten

---

<sup>5</sup> $L_2(\mathbb{R})$  bezeichnet den Raum der quadratisch integrierbaren Funktionen (*Hilbert-Raum*). Weitere Details sind in Abschnitt 2.4.1 zu finden.

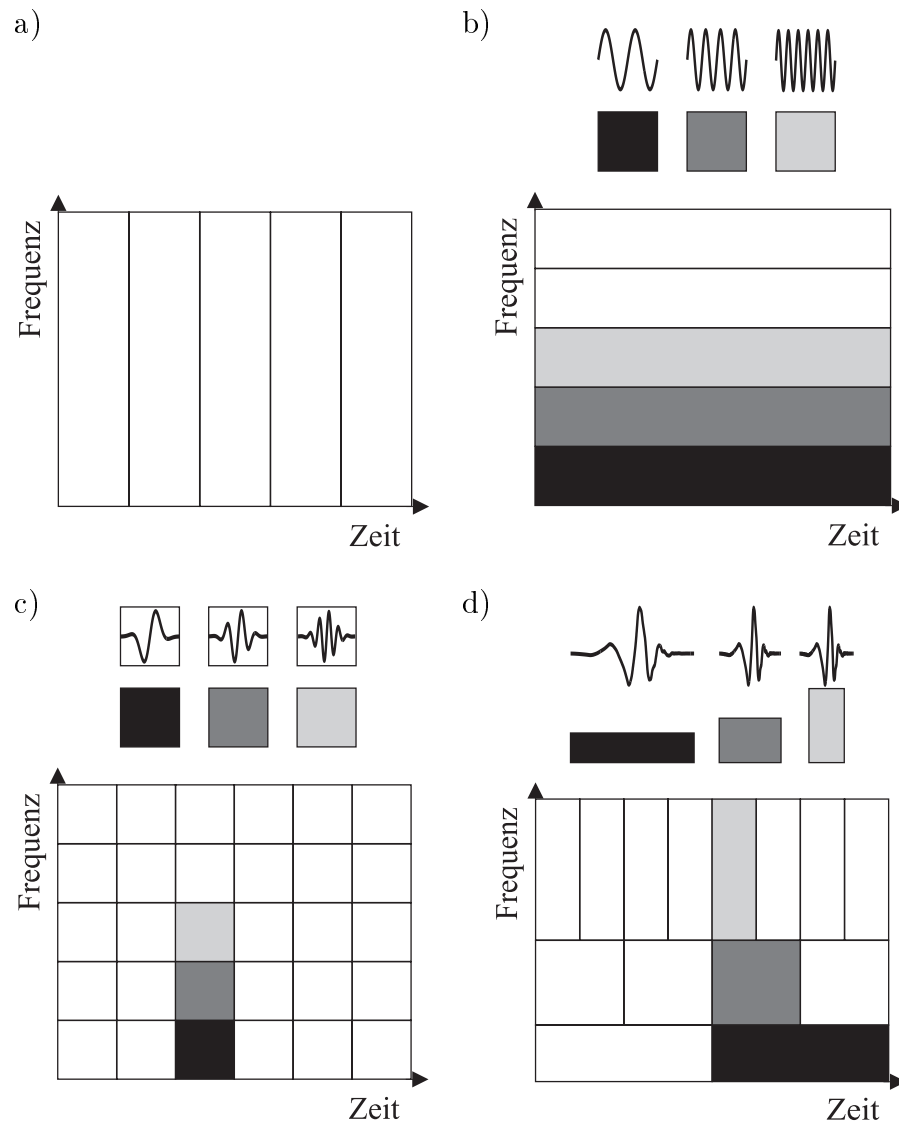


Abbildung 2.3.: Schematische Darstellung der Ausdehnung von Zeit-Frequenz-Fenstern: a) im Zeitbereich, b) nach Fourier-Transformation, c) nach gefensterter Fourier-Transformation und d) nach Wavelet-Transformation [3, 21].

zeigt, daß Frequenzanteile, die den beiden Banden zuzuordnen sind, in der Fourier-Transformierten nicht lokalisierbar sind, da sie über den gesamten Frequenzbereich verteilt sind.

Einen Ausweg für die Analyse derartiger Signalmerkmale bietet die Wavelet-Transformation, wie später in Abschnitt 2.3.3 und Abbildung 2.7 gezeigt wird.

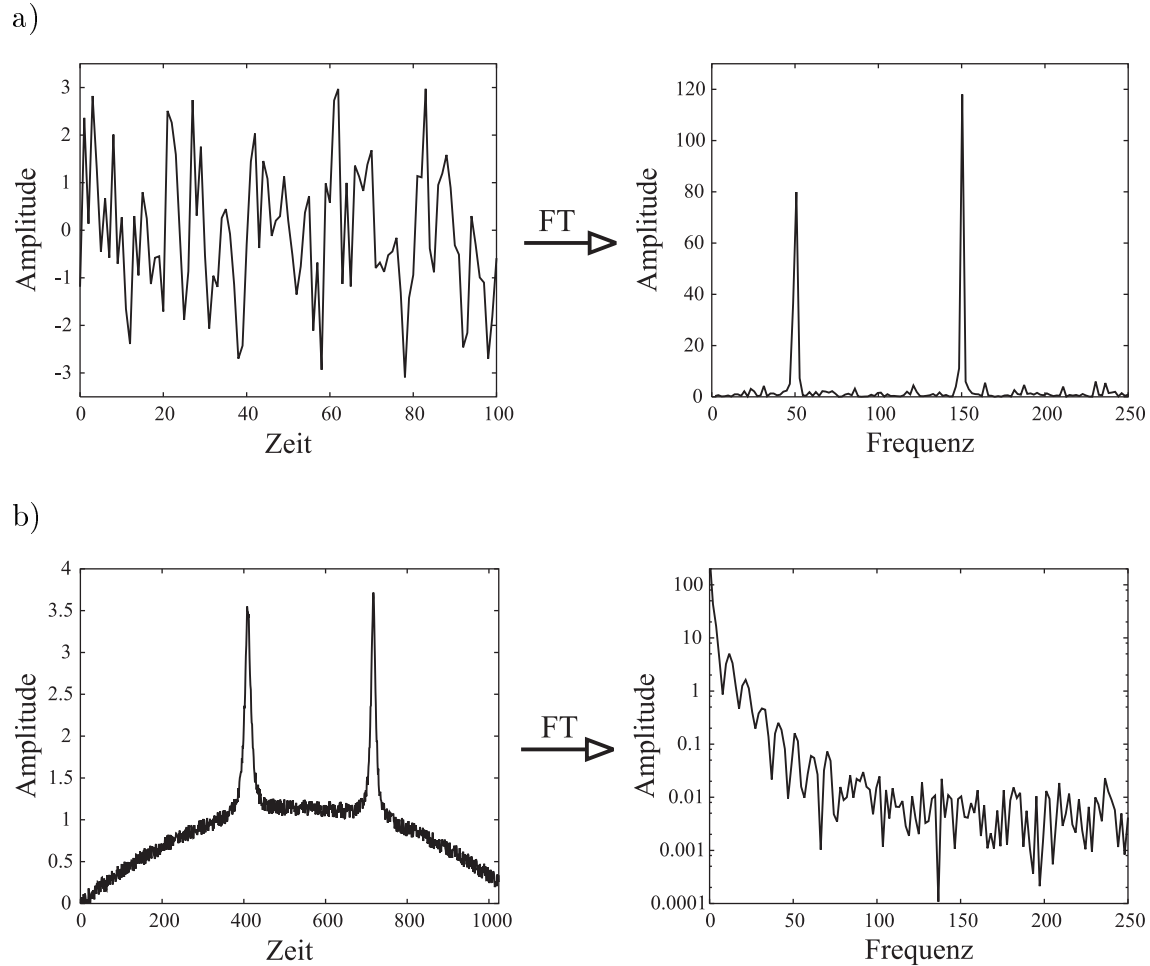


Abbildung 2.4.: Beispiele für die Fourier-Transformation (FT) von Signalen: a) Verrauschte Überlagerung zweier Sinusfunktionen, b) simuliertes Spektrum.

### 2.3.2. GEFENSTERTE FOURIER-TRANSFORMATION

Um die Zeit-Auflösung der Fourier-Transformation zu verbessern, wurde von *D. Gabor* [107] die gefensterete Fourier-Transformation (*Windowed* bzw. *Short Time Fourier Transform*, WFT bzw. STFT, siehe z.B. [108]) entwickelt. Bei dieser wird das Signal innerhalb eines definierten Fensters analysiert. Durch sukzessives Verschieben des Fensters wird eine lokale Frequenzanalyse über das ganze Signal erreicht, deren Auflösung in Zeit- und Frequenzbereich allerdings von der konstant gehaltenen Fensterbreite abhängt (siehe Abbildung 2.3c).

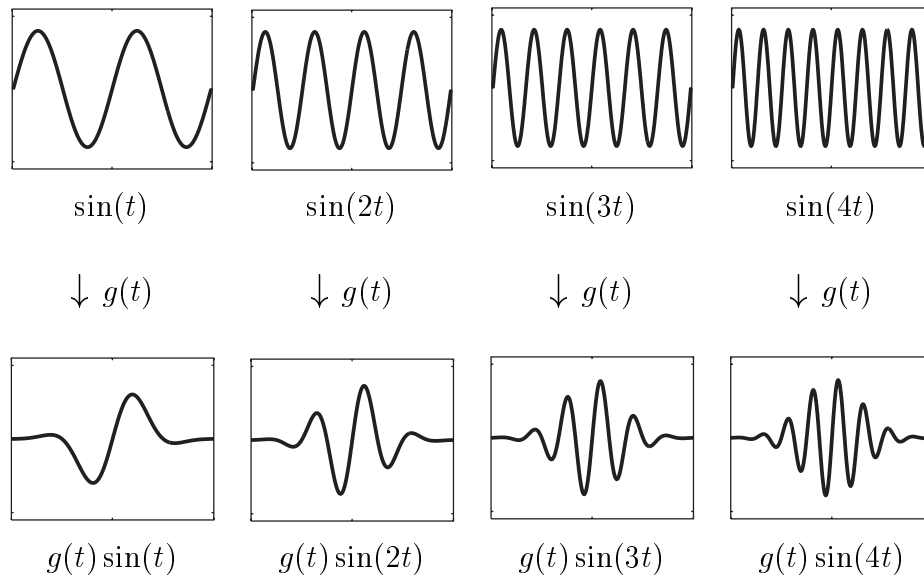


Abbildung 2.5.: Basisfunktionen der gefensterter Fourier-Transformation am Beispiel von Sinusfunktionen multipliziert mit einer Gauß-Funktion  $g(t)$  als Apodisationsfunktion.

Um Artefakte am Rand des Signalausschnitts zu vermeiden, wird das Signal innerhalb des Fensters mit einer Apodisationsfunktion  $g(t)$  — z.B. mit einer Gauß-Funktion  $e^{-x^2}$  — multipliziert, deren Zentrum  $s$  in der Fenstermitte liegt und die zu den Rändern hin rasch abklingt. Ausgehend von Gleichung 2.4 ergibt sich für die gefensterter Fourier-Transformation [4]:

$$T_f(s, \omega) = \int_{-\infty}^{\infty} f(t)g(t \leftrightarrow s)e^{-i\omega t} dt .$$

Die Wirkung der Apodisationsfunktion läßt sich auch in Bezug auf die Basisfunktionen der Fourier-Transformation  $\sin(t)$  und  $\cos(t)$  darstellen. Wie in Abbildung 2.5 am Beispiel von Sinusfunktionen gezeigt wird, ergibt sich mit den lokalisierten Sinus- und Kosinusfunktionen unterschiedlicher Frequenzen eine neue Basis für die Analyse des Signals. Die Lokalisierungseigenschaften sind allerdings abhängig von der Ausdehnung der sich überlappenden Fenster. Aufgrund der Heisenbergschen Unschärferelation  $\Delta t \cdot \Delta \omega \geq C > 0$  ist der Flächeninhalt des Zeit- und Frequenz-Fensters nach unten beschränkt [108]. Der Wert der Konstante  $C$  hängt von der verwendeten Normierung der Fourier-Transformierten ab.

Somit führt die Wahl eines im Zeitbereich schmalen Fensters zu einer relativ guten Auflösung im Zeitbereich und zu einer niedrigeren im Frequenzbereich. Bei der

Wahl eines breiten Zeit-Fensters kehren sich die Verhältnisse um. Wird für die Analyse einzelner Signalanteile eine bestimmte Auflösung gewünscht, so muß das Zeit-Frequenz-Fenster entsprechend angepaßt werden. Die Wavelet-Analyse bietet hier eine Alternative, da sie selbstadaptiv arbeitet, d.h. ein Signal wird immer in verschiedenen Auflösungen für Zeit- und Frequenzbereich analysiert.

### 2.3.3. Wavelet-Transformation

Bei einer Wavelet-Analyse geht man von einem sogenannten Mutter-Wavelet  $\psi$  aus, von dem unterschiedliche Formen und ganze Familien zur Verfügung stehen (siehe Abschnitt 2.3.4). Allen Wavelet-Funktionen ist gemeinsam, daß sie im Zeitbereich außerhalb eines begrenzten Intervalls rasch gegen Null abfallen oder sogar den Wert Null besitzen, wie im Falle der Wavelets mit kompaktem Träger.

Im Gegensatz zur gefensterten Fourier-Transformation ist allerdings die Breite des wirksamen Intervalls oder Fensters variabel. Dies wird durch Stauchen oder Strecken des Wavelets (Faktor 2 bei dyadischer Dilatation) im Zeitbereich erreicht. Damit erhält man Wavelets, die an die verschiedenen zu analysierenden Frequenzanteile angepaßt sind. Wie in Abbildung 2.6 schematisch gezeigt, wird durch die Translation der Wavelets einer Skala  $j$  auf bestimmte Positionen  $k$  die Lokalisierung der Frequenzanteile auf der Zeit-Achse ermöglicht. Als Skala wird in diesem Zusammenhang ein Frequenzband bezeichnet.

Abbildung 2.3d auf der Seite 17 verdeutlicht, wie sich mit der Dilatation der Wavelets im Zeitbereich das Zeit-Frequenz-Fenster verändert. Da die Fläche des Fensters aufgrund der Konstruktion der Wavelets konstant bleiben muß, bewirkt eine Verdoppelung der Auflösung im Zeitbereich eine Halbierung jener im Frequenzbereich. Durch Verschiebung der Fenster entlang der Zeit-Achse wird die gesamte Zeit-Frequenz-Ebene abgedeckt [22].

In Analogie zu den Gleichungen 2.1 und 2.2 läßt sich damit eine Funktion  $f$  auf Basis der Wavelet-Funktionen  $\psi_{j,k}$  darstellen:

$$f(t) = \sum_{j,k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(t) .$$

Die Wavelet-Koeffizienten  $d_{j,k}$  können im Falle einer orthonormalen Wavelet-Basis (siehe Abschnitt 2.4.1) als das Integral über das Produkt aus Wavelet und Funktion berechnet werden:

$$d_{j,k} = \int_{-\infty}^{+\infty} f(t) \overline{\psi_{j,k}(t)} dt . \quad (2.5)$$

Ein Koeffizient kann anschaulich so interpretiert werden, daß sein Betrag  $|d_{j,k}|$  die Amplitude der Frequenz  $2^j$  im Signal angibt. Durch  $k$  wird die Position des Frequenzanteils im Zeitbereich auf der Skala  $j$  bestimmt.



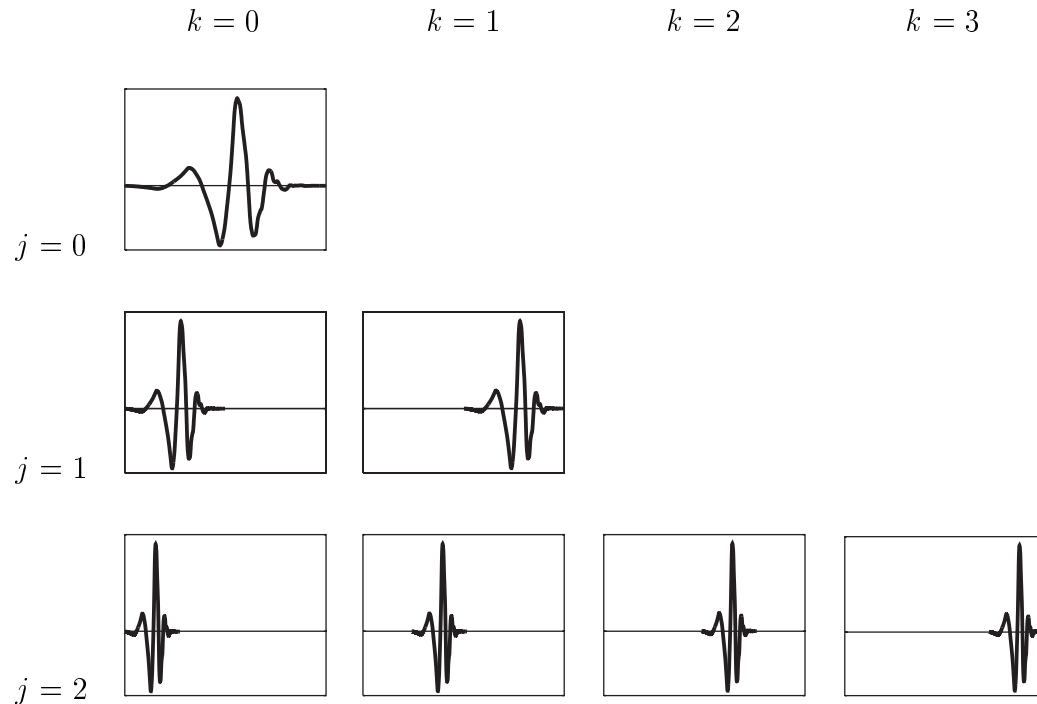


Abbildung 2.6.: Dilatation und Translation eines Wavelets (Typ Daubechies 4 aus [109]).

Abbildung 2.7 zeigt das Ergebnis der Wavelet-Zerlegung des simulierten Spektrums, das bereits in Abbildung 2.4 verwendet wurde. Die durch eine Wavelet-Transformation erhaltenen Koeffizienten  $d_{j,k}$  sind skalenweise in Blockdiagrammen aufgetragen. Die Ordinate gibt den Wert eines Koeffizienten an, während die Abszisse den Index des Koeffizienten über alle Skalen (also nicht Index  $k$ ) wiedergibt.

Ausgehend von der Skala  $j = 0$  werden mit steigendem  $j$  immer höherfrequente Signalanteile durch die Koeffizienten repräsentiert, bis bei  $j = 8$  nur noch hochfrequentes Rauschen erfaßt wird. Dabei verdoppelt sich die Anzahl der Koeffizienten pro Skala und der Bereich auf der Zeit-Achse, den ein einzelner Koeffizient abdeckt, wird bei dyadischer Dilatation jeweils halbiert. Deutlich zu sehen ist, daß in den Skalen  $j = 3$  bis  $j = 6$  Signalanteile der beiden Banden kodiert sind. Dagegen finden sich Basislinienanteile bzw. Rauschen in niedrigeren bzw. höheren Skalen wieder.

Mit der Wavelet-Transformation erhält man die Analyse von Signalen in verschiedenen Auflösungen bezüglich der Frequenzanteile (siehe auch Abbildung 2.10, Seite 35) unter Erhaltung der Lokalisierungsinformation. Eine mathematische Darstellung der Transformation als Multiskalen-Analyse folgt in Abschnitt 2.4.

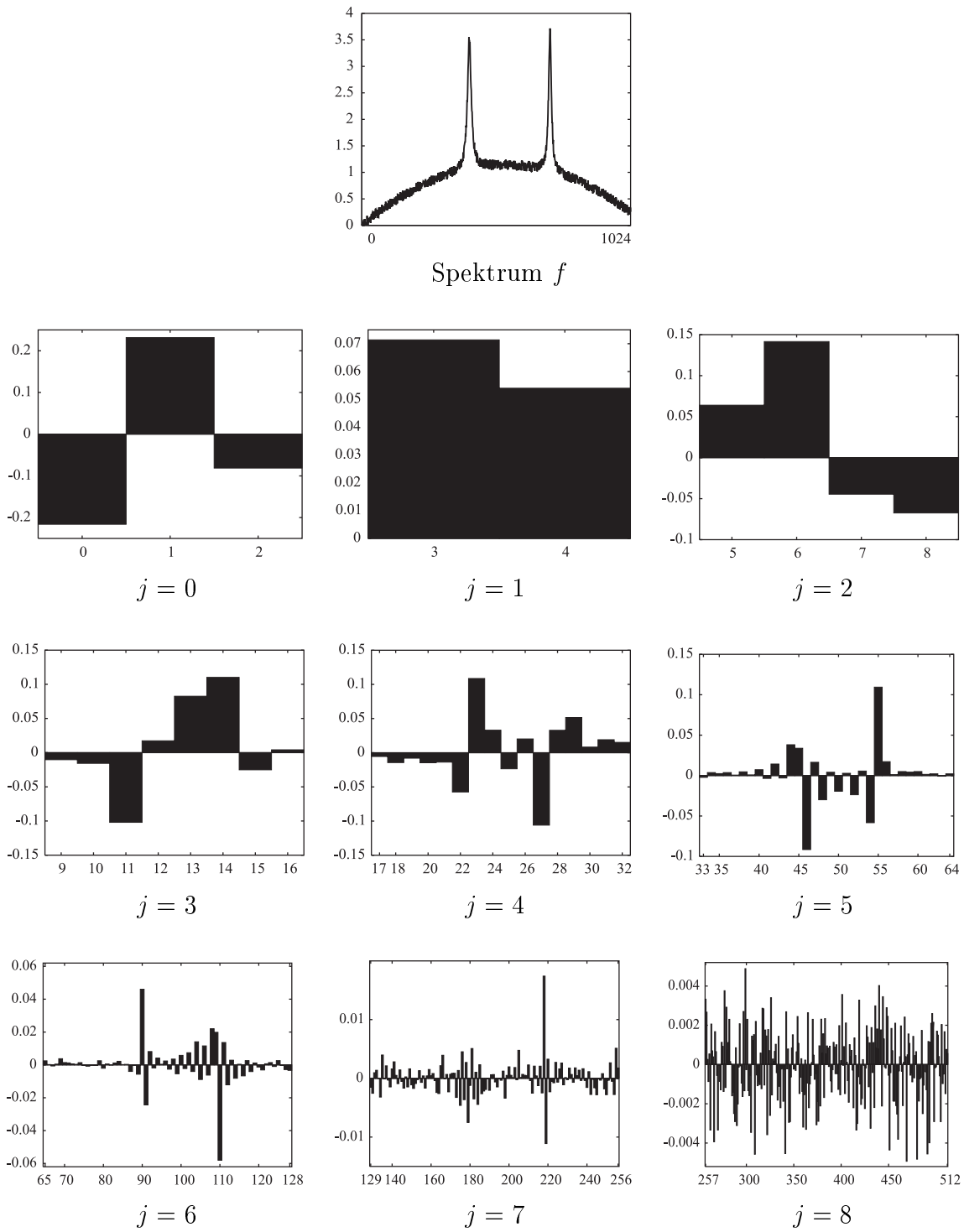


Abbildung 2.7.: Wavelet-Zerlegung eines simulierten Spektrums  $f$  mit dem Sturm-Liouville-Wavelet (s-Wavelet).

### 2.3.4. Arten von Wavelets, Wavelet-Familien und ihre Eigenschaften

Neben der adaptiven Eigenschaft der Wavelet-Transformation bezüglich der Zeit-Frequenzanalyse von Signalen ist die große Zahl an verfügbaren Wavelet-Basisfunktionen mit z.T. sehr verschiedenen Eigenschaften bemerkenswert.

Von den vielen in der Literatur beschriebenen Wavelet-Typen (eine Übersicht gibt [110]) sind ein bedeutender Teil in Wavelet-Toolboxen (z.B. [109]) oder anderer Software implementiert. Für die vorliegende Arbeit wurden die folgenden Wavelets und Wavelet-Familien ausgewählt. Einige typische Beispiele sind in Abbildung 2.8 dargestellt, während die Abbildungen A.1 und A.2 auf den Seiten 143 und 144 im Anhang alle verwendeten Typen zeigen.

- Das **Haar-Wavelet** geht auf *Alfred Haar* zurück [6] und stellt den einfachsten Fall eines Wavelets mit kompaktem Träger dar. Allerdings ist diese Funktion nicht stetig, was bei Rekonstruktionen zu unerwünschten Effekten führen kann.
- **Daubechies-Wavelets** sind von *Ingrid Daubechies* entwickelte orthonormale Wavelets mit kompaktem Träger [18], die im Gegensatz zum Haar-Wavelet mit beliebiger Glattheit konstruiert werden können.
- **Coiflets** wurden ebenfalls von *I. Daubechies* auf Anregung von *R. Coifman* konstruiert [110, 111]. Die zugehörigen speziellen Skalierungsfunktionen (siehe Abschnitt 2.4) mit besseren Approximationseigenschaften unterscheiden diese von den Daubechies-Wavelets.
- **Biorthogonale Wavelets** wurden von *Albert Cohen et al.* [112] und auch von *C. K. Chui* (Spline-Wavelets mit kompaktem Träger [14]) entwickelt, bei deren Konstruktion mehr Freiheitsgrade zur Verfügung stehen. So lassen sich z.B. symmetrische Wavelet-Funktionen erzeugen, die besonders für die Bildverarbeitung nützlich sind.
- Für Signale auf kompakten Intervallen wurden spezielle Wavelets entwickelt (siehe Abschnitt 2.4.7): **Tschebyscheff-Wavelets** von *T. Kilgore* und *J. Prestin* [113] sowie **Sturm-Liouville-Wavelets** von *U. Depczynski* [114], die im weiteren als t- und s-Wavelets bezeichnet werden.

Bereits die in Abbildung 2.8 dargestellten Wavelets lassen erahnen, daß zwischen den verschiedenen Typen sowohl Gemeinsamkeiten als auch deutliche Unterschiede bestehen [109]. Im folgenden werden wichtige Eigenschaften von Wavelets allgemein erläutert, in denen sich die einzelnen Typen z.T. deutlich unterscheiden.

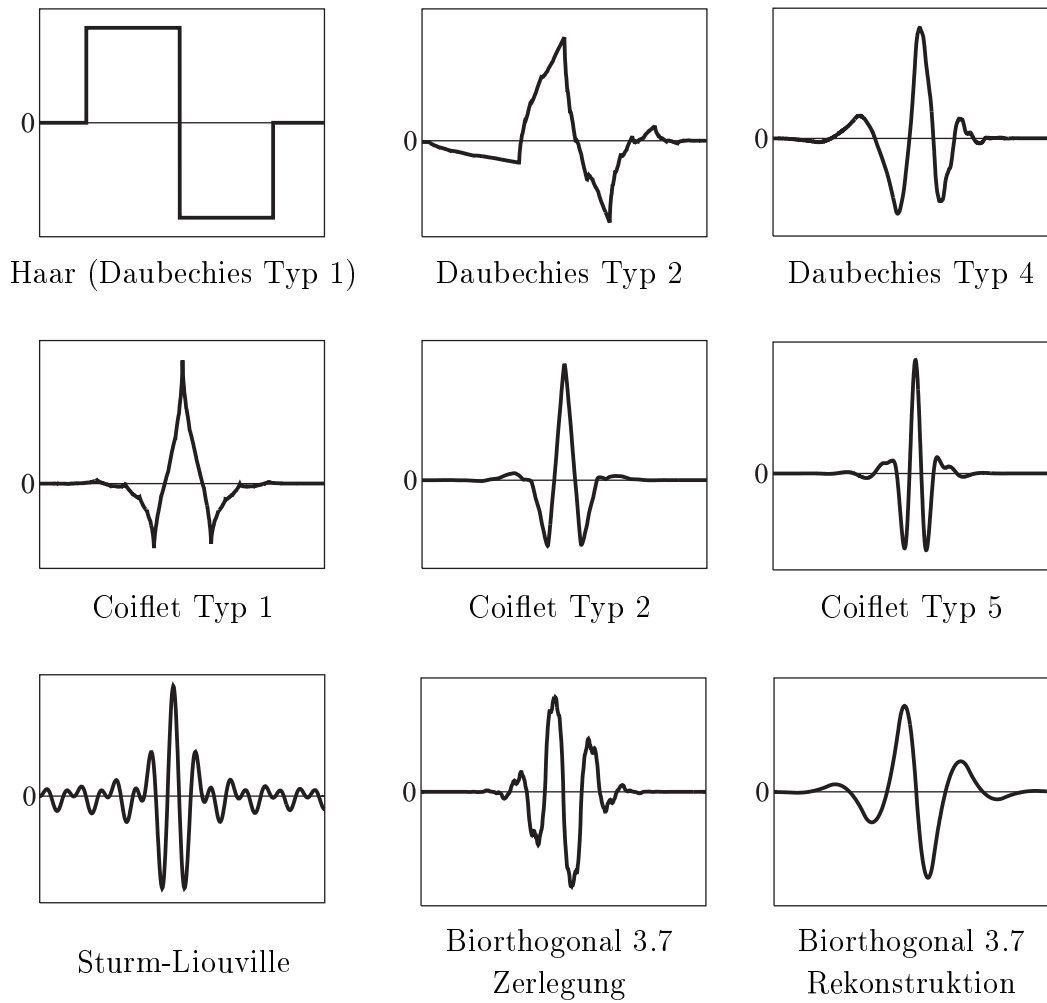


Abbildung 2.8.: Beispiele für Wavelet-Funktionen.

- Verschwindende Momente (*vanishing moments*)

Je nach Konstruktion verfügt ein Wavelet über eine bestimmte Anzahl an „verschwindenden Momenten“. Das  $k$ -te Moment einer Funktion  $f$ , also das Integral über das Produkt dieser Funktion (hier das Wavelet  $\psi$ ) und der  $k$ -ten Potenz der unabhängigen Variablen  $t$ , besitzt dabei den Wert 0:

$$m_k = \int t^k \psi(t) dt = 0 .$$

Beispielsweise gibt die Ordnung  $N$  eines Daubechies-Wavelets an, wieviele verschwindende Momente, also Momente mit dem Wert Null, dieses Wavelet besitzt [83]:

$$\int t^k \psi_N(t) dt = 0 \quad \text{mit} \quad k = 0, \dots, N \Leftrightarrow 1 .$$

Anschaulich gibt diese Zahl an, daß Polynome bis zum Grad  $N \Leftrightarrow 1$  nicht durch das Wavelet erfaßt werden. So machen z.B. drei verschwindende Momente ein Wavelet „blind“ für konstante, lineare und quadratische Signalanteile. Diese gehen bei der Signal-Analyse allerdings nicht verloren, sondern werden, wie später dargestellt wird, bei der Multiskalen-Analyse im Grundraum  $V_0$  (siehe Abschnitt 2.4.6) abgebildet.

Dies führt u.a. zu einer starken Komprimierung des Signals, da die vom Wavelet nicht erfaßbaren Signalanteile mit den wenigen Koeffizienten des Grundraumes kodiert werden. Durch diese kompakte Kodierung der Hauptsignalanteile im Grundraum können andere Anteile, wie z.B. Singularitäten und Unstetigkeiten im Signal, besser detektiert werden, da die entsprechenden Koeffizienten stärker hervortreten.

Konstruktionsbedingt und aufgrund der oszillierenden Form besitzt jedes Wavelet ein verschwindendes Moment mit  $k = 0$ , d.h. das Integral über eine Wavelet-Funktion besitzt den Wert Null [115]:

$$\int_{-\infty}^{+\infty} \psi_{j,k}(t) dt = 0 .$$

- Träger (*support*)

Der Träger eines Wavelets bezeichnet den Bereich, in dem die Wavelet-Funktion das Signal abtastet oder analysiert. Dieser kann kompakt sein, also eine endliche Ausdehnung haben, oder unendlich sein, wobei die Funktion allerdings zu beiden Seiten hin rasch gegen Null abklingt.

Gerade im Bezug auf die kompakten Träger waren die von *I. Daubechies* entwickelten orthogonalen Wavelets ein Durchbruch. Wie beim Haar-Wavelet ist hier der „wirksame Bereich“ der Wavelets auf der Zeitachse definiert. Dies wird aber bei den Daubechies-Wavelets mit beliebig glatten und stetigen Funktionen erreicht, die für viele Anwendungen besser geeignet sind als das „eckige“ Haar-Wavelet. Mit den sich ergebenden endlichen Filterfolgen ließen sich außerdem schnelle Algorithmen entwickeln (siehe Abschnitt 2.4.4).

- Glattheit bzw. Regularität (*smoothness, regularity*)

Die Glattheit der Wavelet-Funktion, d.h. die Anzahl der stetigen Ableitungen, ist besonders für die Rekonstruktion von Signalen nach der Bearbeitung ihrer Wavelet-Transformierten von Bedeutung. So sollte die Regularität des Wavelets im Bereich derjenigen des Signals liegen, um genügend glatte Rekonstruktionen zu erhalten. Wichtig ist dies im Bereich der Bildverarbeitung, wo sich Artefakte besonders störend auswirken.

Einen tabellarischen Überblick über die Eigenschaften der verwendeten Wavelets aus der MATLAB Wavelet Toolbox [109] gibt Abschnitt A.1 im Anhang auf der Seite 139.

## 2.4. Die Wavelet-Transformation als Multiskalen-Analyse

Die Wavelet-Transformation eines Signals läßt sich mit Hilfe der Multiskalen-Analyse beschreiben, bei der das Signal auf verschiedenen Skalen, also mit unterschiedlichen Auflösungen, analysiert wird. Hintergrund hierfür ist, daß Signale meist Merkmale und Strukturen sehr unterschiedlicher Ausdehnung und Frequenz enthalten und es daher für eine Analyse keine einzelne optimale Auflösung gibt.

Algorithmen für Multiskalen-Analysen mit Filtern sind zwar schon seit längerem z.B. in der Bildverarbeitung bekannt, aber erst *S. Mallat* entwickelte einen solchen mit Wavelets [16, 17]. Bei diesem iterativen Verfahren wird das Signal mit Hilfe einer Skalierungsfunktion auf verschiedenen Skalen dargestellt, wobei sich die Auflösung von einer Skala auf die folgende halbiert (dyadische Dilatation). In Richtung niedrigerer Auflösungen werden immer gröbere, in der anderen Richtung detailliertere Darstellungen (Approximationen) des Signals erhalten. Mit Wavelets werden nur die Differenzen zwischen zwei aufeinanderfolgenden Skalen kodiert, also die Details, die bei einer Halbierung der Auflösung verlorengehen [5].

Wie in Abschnitt 2.4.4 gezeigt wird, läßt sich das Zusammenspiel von Wavelet und zugehöriger Skalierungsfunktion bei der Multiskalen-Analyse mit der sukzessiven Anwendung von Hoch- und Tiefpassfiltern in der Signalverarbeitung vergleichen.

In den folgenden Abschnitten wird die Multiskalen-Analyse mathematisch formuliert, nachdem zunächst einige verwendete Begriffe erläutert und Definitionen gegeben wurden. Allgemein sind diese in mathematischen Lehrbüchern zur Analysis [116, 117], aber auch in Darstellungen der Wavelet-Transformation [88, 90] zu finden. Im weiteren werden als Signale  $f$  zunächst Funktionen aus  $L_2(\mathbb{R})$  betrachtet,

um dann die Zerlegung auch auf Signale zu übertragen, deren Definitionsbereich ein endliches Intervall  $[a, b]$  ist ( $f \in L_2([a, b])$ ).

### 2.4.1. Der Hilbert-Raum $L_2(\mathbb{R})$

$L_2(\mathbb{R})$  bezeichnet den *Hilbert*-Raum der quadratisch integrierbaren Funktionen, die ein endliches Intergral<sup>6</sup> besitzen:

$$L_2(\mathbb{R}) := \left\{ f; \int_{-\infty}^{+\infty} |f(t)|^2 dt < +\infty \right\} .$$

Dieser Raum ist mit dem Skalarprodukt

$$\langle f, g \rangle = \int_{-\infty}^{+\infty} f(t) \overline{g(t)} dt$$

ausgestattet. Mit dem Skalarprodukt der beiden Funktionen  $f$  und  $g$  ( $f, g \in L_2(\mathbb{R})$ ) kann der wichtige Begriff der Orthogonalität eingeführt werden. Ergibt das Skalarprodukt  $\langle f, g \rangle$  den Wert Null, so sind die Funktionen per Definition orthogonal, also unkorreliert.

Bei Wavelet-Transformationen mit orthogonalen Basen<sup>7</sup> sind die einzelnen Wavelets orthogonal, was vorteilhaft bei Modifizierungen der Wavelet-Transformierten ist: Die Veränderung eines bestimmten Wavelet-Koeffizienten wirkt sich nicht auf die übrigen aus. Daher ist die Wirkung auf den entsprechenden Signalanteil beschränkt und das Signal bleibt ansonsten unverändert.

Die Wurzel aus dem Skalarprodukt einer Funktion  $f$  mit sich selber wird als  $L_2$ -Norm der Funktion bezeichnet:

$$\|f\| = \sqrt{\langle f, f \rangle} = \left( \int_{-\infty}^{+\infty} f(t) \overline{f(t)} dt \right)^{1/2} .$$

Das Quadrat der Norm ergibt die Energie einer Funktion, die für die Funktionen im Raum  $L_2(\mathbb{R})$  endlich ist:

$$\|f\|^2 = \int_{-\infty}^{+\infty} |f(t)|^2 dt < +\infty . \tag{2.6}$$

$L_2$ -Norm und Energie bleiben bei einer Fourier und Wavelet-Transformation erhalten. Im Falle einer orthonormalen Wavelet-Transformation gilt:

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = \sum_{j,k \in \mathbb{Z}} |d_{j,k}|^2 . \tag{2.7}$$

<sup>6</sup> $L_2$  steht für das *Lebesgue*-Integral (siehe [116]) über quadratisch (dennoch tiefgestellte 2) integrierbare Funktionen.

<sup>7</sup>Bei der Verwendung von normierten orthogonalen Basisfunktionen spricht man von einer orthonormalen oder auch Hilbert-Basis.

Die Kodierung eines Signals durch Wavelet-Koeffizienten  $d_{j,k}$  unter Energieerhaltung sichert eine wichtige Voraussetzung für die Anwendbarkeit der Transformation: Eine Umkehrung ergibt exakt das Ausgangssignal. Ebenso werden lineare Veränderungen des transformierten Signals bei einer Rekonstruktion linear in den Zeitbereich übertragen<sup>8</sup>. Heuristisch betrachtet bleibt bei den Transformationen der Informationsgehalt des Signals konstant.

### 2.4.2. Multiskalen-Analyse von $L_2(\mathbb{R})$

Als Multiskalen-Analyse von  $L_2(\mathbb{R})$  (*Multi Resolution Analysis*, MRA, siehe [83, 88, 90, 106, 118]) wird eine aufsteigende Folge von abgeschlossenen linearen Teilräumen  $\{V_j\}_{j \in \mathbb{Z}}$  aus  $L_2(\mathbb{R})$  bezeichnet, die folgende Bedingungen erfüllen [106]:

1. Schachtelung (*nesting property*):

$$V_j \subset V_{j+1}, \quad j \in \mathbb{Z}.$$

Die einzelnen Teilräume sind so geschachtelt, daß alle Elemente eines Raumes  $V_j$  in  $V_{j+1}$  enthalten sind.

2. Skalierungseigenschaft (*scaling property*):

$$f \in V_j \Leftrightarrow f(2 \cdot) \in V_{j+1}, \quad j \in \mathbb{Z}.$$

Der Raum  $V_{j+1}$  wird aus den mit dem Faktor 2 skalierten Elementen des Raumes  $V_j$  gebildet (dyadische Skalierung)<sup>9</sup>.

3. Translate und Verschiebungsinvarianz (*property of shift-invariance*):

$$f \in V_j \Leftrightarrow f(\cdot \Leftrightarrow 2^{-j}k) \in V_j, \quad j, k \in \mathbb{Z}.$$

Von einer Funktion  $f$  im Raum  $V_j$  sind auch alle dyadischen Translate in diesem Raum enthalten.

4. Vollständigkeit (*property of exhaustion*):

$$\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L_2(\mathbb{R}).$$

Der  $L_2$ -Abschluß der Vereinigung aller Teilräume  $V_j$  ergibt den Raum  $L_2(\mathbb{R})$ . Damit lassen sich alle Funktionen aus  $L_2(\mathbb{R})$  durch Elemente aus einem Raum  $V_j$  ( $j$  genügend groß) beliebig genau approximieren.

<sup>8</sup>Mathematisch kann die Linearität wie folgt formuliert werden:  $\varphi(\alpha x + \beta x) = \alpha \varphi(x) + \beta \varphi(y)$ .

<sup>9</sup>Der Punkt „ $\cdot$ “ steht allgemein für ein Argument der Funktion  $f$ , also z.B. für  $t$  bei  $f(t)$ .



5. Ausschluß von Redundanz (*property of non-redundancy*):

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\} .$$

Der Durchschnitt aller Räume  $V_j$  ist der Raum  $\{0\}$ , der nur die Nullfunktion enthält. Diese ist damit die einzige Funktion, die in allen Teilräumen enthalten ist.

6. *Riesz*-Basis Eigenschaft (*Riesz basis property*):

Es existiert eine Funktion  $\phi \in V_0$ , deren ganzzahlige Translate eine *Riesz*-Basis von  $V_0$  bilden. Dies bedeutet:

- a) Jede Funktion  $f$  im Raum  $V_0$  kann als Linearkombination von Translaten der Skalierungsfunktion  $\phi$  repräsentiert werden

$$f = \sum_{k \in \mathbb{Z}} c_k \phi(\cdot \leftrightarrow k) . \quad (2.8)$$

- b) Es existieren *Riesz*-Schranken  $0 < A \leq B < \infty$ , so daß für alle  $c \in \ell_2(\mathbb{Z})$  gilt:

$$A \cdot \sum_{k \in \mathbb{Z}} |c_k|^2 \leq \left\| \sum_{k \in \mathbb{Z}} c_k \phi(\cdot \leftrightarrow k) \right\|^2 \leq B \cdot \sum_{k \in \mathbb{Z}} |c_k|^2 .$$

Bilden die Translate von  $\phi$  eine orthonormale Basis, so gilt  $A = B = 1$ .

Damit wird ausgedrückt, daß kleine Veränderungen der Koeffizienten nur entsprechend geringe Auswirkungen auf die Funktion haben, so daß sich z.B. Rundungsfehler bei Berechnungen nicht „aufschaukeln“.

Die *Riesz*-Basis Eigenschaft ist von großer praktischer Bedeutung, da so sichergestellt wird, daß die Zerlegung ( $f \rightarrow (c_k)_{k \in \mathbb{Z}}$ ) und Rekonstruktion ( $(c_k)_{k \in \mathbb{Z}} \rightarrow f$ ) eines Signals numerisch stabil ist.

Der Raum  $L_2(\mathbb{R})$  läßt sich demnach durch eine aufsteigende Folge abgeschlossener Teilräume  $V_j \subset L_2(\mathbb{R})$  ( $j \in \mathbb{Z}$ ) darstellen<sup>10</sup>:

$$\{0\} \subset \dots V_{-1} \subset V_0 \subset V_1 \subset \dots V_{j-1} \subset V_j \subset V_{j+1} \dots \subset L_2(\mathbb{R}) .$$

Ein Raum  $V_j$  enthält alle Elemente der vorhergehenden Räume  $V_{j-1}, V_{j-2}, \dots$ . Umgekehrt gehen allerdings bei einem Übergang von  $V_j$  zu  $V_{j-1}$  Elemente verloren.

<sup>10</sup>Bei anderen Darstellungen (z.B. von *I. Daubechies*) erfolgt die Indizierung genau umgekehrt:  $\dots V_j \subset V_{j-1} \dots \subset L_2(\mathbb{R})$ . Beide Indexkonventionen sind äquivalent und bezeichnen ein und dieselbe Zerlegung (siehe [88], S. 196).

Die Konstruktion der Teilräume ist so angelegt, daß mit  $V_{j-1}$  ein gegebenes  $f$  nicht mehr beliebig genau, sondern in unserem Fall nur mit halbiertes Auflösung approximiert werden kann. Alle Elemente des Raumes  $V_j$  können aber durch  $V_{j-1}$  unter Hinzunahme eines geeigneten Wavelet-Raumes  $W_{j-1}$  dargestellt werden:

$$V_j = V_{j-1} \oplus W_{j-1} , \quad (2.9)$$

wobei die Elemente von  $V_{j-1}$  und  $W_{j-1}$  orthogonal sind ( $V_{j-1} \perp W_{j-1}$ ).

Damit wird ausgehend von  $V_j$  die Darstellung des Signals in halber Auflösung mit einer Kodierung der Details im Wavelet-Raum erreicht. Anders betrachtet wird  $V_j$  in einen niederfrequenten ( $V_{j-1}$ ) und einen hochfrequenten Anteil ( $W_{j-1}$ ) zerlegt, wobei nieder- und hochfrequent relativ zu sehen ist, bezogen auf die gegebene Skala  $j$  (Auflösung).

Führt man die Zerlegung des niederfrequenten Anteils  $V_{j-1}$  weiter

$$V_{j-1} = V_{j-2} \oplus W_{j-2} ,$$

so erhält man für  $V_j$ :

$$V_j = \underbrace{V_{j-2} \oplus W_{j-2}}_{=V_{j-1}} \oplus W_{j-1} .$$

Als Ergebnis einer kompletten orthogonalen Multiskalen-Analyse ergibt sich die Wavelet-Zerlegung von  $L_2(\mathbb{R})$  als orthogonale direkte Summe von Wavelet-Räumen:

$$\bigoplus_{j \in \mathbb{Z}} W_j = L_2(\mathbb{R}) .$$

Damit liegt, ähnlich wie nach einer Fourier-Transformation, eine Zerlegung des Signals nach Frequenzen vor, da jeder Wavelet-Raum Signalanteile aus einem bestimmten Frequenzband repräsentiert.

### 2.4.3. Skalierungsfunktion und Wavelet

Bei der Multiskalen-Analyse geht man zunächst von der Skalierungsfunktion  $\phi$  aus, die auch als „Vater-Wavelet“ bezeichnet wird. Letztlich wird das Wavelet erst innerhalb einer solchen Prozedur ausgehend von der gewählten Skalierungsfunktion konstruiert.

Nach den oben genannten Annahmen bildet eine durch dyadische Dilatation und Translation gebildete Menge von Skalierungsfunktionen

$$\{ \phi_{j,k} = 2^{j/2} \phi(2^j \cdot \leftarrow k) , k \in \mathbb{Z} \} ,$$

für jedes  $j \in \mathbb{Z}$  eine *Riesz*-Basis für den Funktionsraum  $V_j$ . So läßt sich ein Signal  $f \in V_j$  nach Gleichung 2.8 wie folgt darstellen:

$$f = \sum_{k \in \mathbb{Z}} c_{j,k} \phi_{j,k} . \quad (2.10)$$

In Analogie dazu bildet die Menge von orthogonalen Wavelets

$$\{\psi_{j,k} = 2^{j/2} \psi(2^j \cdot \leftrightarrow k) , k \in \mathbb{Z}\} ,$$

für jedes  $j \in \mathbb{Z}$  eine *Riesz*-Basis für den Wavelet-Raum  $W_j$  und die Menge  $\{\psi_{j,k} , j, k \in \mathbb{Z}\}$  eine *Riesz*-Basis für  $L_2(\mathbb{R})$ . Damit ergibt sich die Wavelet-Zerlegung einer Funktion  $f \in L_2(\mathbb{R})$ :

$$f = \sum_{j,k \in \mathbb{Z}} d_{j,k} \psi_{j,k} . \quad (2.11)$$

Die Zusammenhänge zwischen Skalierungsfunktion und Wavelet werden durch die Formulierung von Wavelet-Zerlegung und -Rekonstruktion eines Signals deutlich.

### 2.4.4. Die schnelle Wavelet-Transformation

#### Zerlegung eines Signals

Nach Gleichung 2.9 wird bei der Wavelet-Zerlegung ein Funktionsraum  $V_j$  in die beiden Räume  $V_{j-1}$  und  $W_{j-1}$  aufgespalten. Dazu werden die Wavelet-Koeffizienten nicht über die zeitraubende Berechnung des inneren Produktes in Gleichung 2.5 (siehe Seite 20) bestimmt. Statt dessen kommen bei der schnellen Wavelet-Transformation (*Fast Wavelet Transform* (FWT)) zwei Folgen von Filterkoeffizienten  $g, h \in \ell_2(\mathbb{Z})$  zur Anwendung [106]:

$$\underbrace{\phi(2 \cdot \leftrightarrow \ell)}_{V_j} = \underbrace{\sum_{k \in \mathbb{Z}} g_{\ell-2k} \phi(\cdot \leftrightarrow k)}_{V_{j-1}} + \underbrace{\sum_{k \in \mathbb{Z}} h_{\ell-2k} \psi(\cdot \leftrightarrow k)}_{W_{j-1}} , \ell \in \mathbb{Z} .$$

Diese auf *S. Mallat* zurückgehende Art der Multiskalen-Zerlegung ist vergleichbar mit einer sukzessiven Anwendung von Hoch- und Tiefpassfiltern, wie z.B. in der digitalen Signalverarbeitung. In Abbildung 2.9 ist der Algorithmus der Multiskalen-Analyse schematisch dargestellt.

Bei der Berechnung der Wavelet-Koeffizienten für ein Signal  $f \in V_J$  geht man von folgender Darstellung aus [106]:

$$f = \sum_{k \in \mathbb{Z}} c_{J,k} \phi_{J,k} \in V_J ,$$

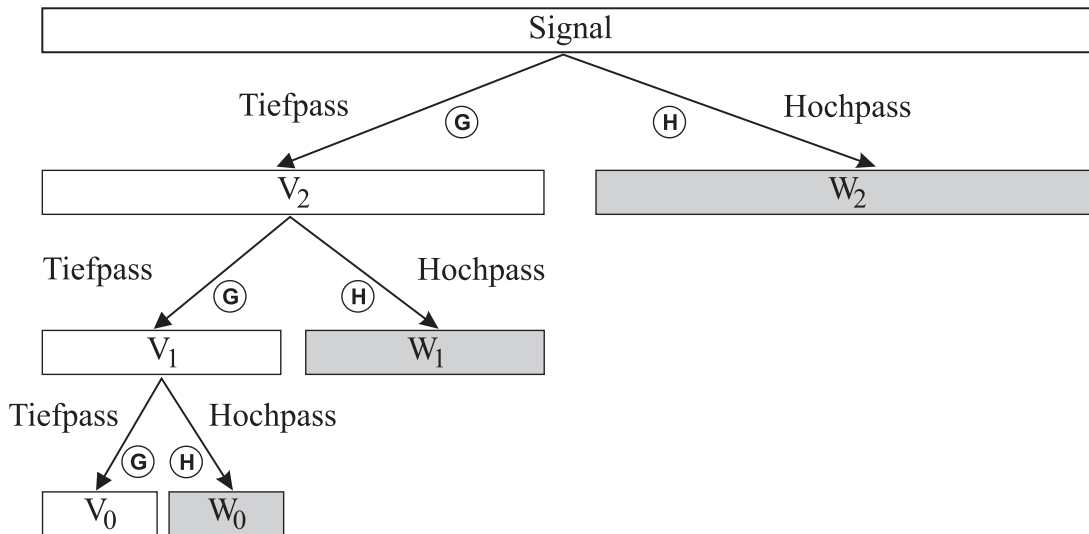


Abbildung 2.9.: Schema der Multiskalen-Analyse nach *S. Mallat* mit Tiefpassfilter  $g$  und Hochpassfilter  $h$  [106].

wobei  $V_J$  den Funktionsraum kennzeichnet, auf den das Signal zu Beginn durch z.B. Projektion oder Approximation abgebildet wird. Mit den Koeffizienten  $c_{J,k}$  werden dann die Wavelet-Koeffizienten  $d_{j,k}$  ( $j < J$ ) rekursiv nach folgenden Gleichungen berechnet:

$$c_{j-1,k} = \sum_{\ell} g_{\ell-2k} c_{j,\ell} ,$$

$$d_{j-1,k} = \sum_{\ell} h_{\ell-2k} c_{j,\ell} .$$

Hierbei handelt es sich um diskrete Faltungen, die sehr schnell berechnet werden können, da es sich bei  $\{h_k\}$  und  $\{g_k\}$  um „kurze“ Folgen handelt. Die Komplexität der Rechenoperationen ist linear  $\mathcal{O}(N)$  im Gegensatz z.B. zur schnellen Fourier-Transformation (FFT) mit der Komplexität  $\mathcal{O}(N \log N)$ .

Eine weitere Vereinfachung besteht darin, daß aus der Filterfolge für eine gegebene orthonormale Skalierungsfunktion die Folge für das Wavelet mit

$$g_k = (\leftrightarrow 1)^k h_{1-k}$$

berechnet werden kann [83]. Damit wird das Wavelet quasi ausgehend von der Skalierungsfunktion konstruiert. Außerdem müssen die Folgen für einen gegebenen Wavelet-Typ nur einmal berechnet werden und liegen dann, wie z.B. in [88], tabelliert vor.

### Rekonstruktion eines Signals

Für die Rekonstruktion eines Signals, also den Übergang von  $V_j$  und  $W_j$  nach  $V_{j+1}$ , stehen ebenfalls Filterfolgen ( $p$  und  $q$ ) zur Verfügung. Nach der sogenannten Zwei-Skalen-Relation [106]

$$\phi = \sum_{k \in \mathbb{Z}} p_k \phi(2 \cdot \leftrightarrow k)$$

kann ein  $\phi \in V_j$  durch Linearkombinationen von  $\phi(2 \cdot) \in V_{j+1}$  dargestellt werden, da gilt:  $V_j \subset V_{j+1}$ . Entsprechendes gilt für die Wavelets, die den Raum  $W_j$  repräsentieren:

$$\psi = \sum_{k \in \mathbb{Z}} q_k \phi(2 \cdot \leftrightarrow k) .$$

Die Rekonstruktion läßt sich schließlich als iterativer Algorithmus formulieren [106]:

$$c_{j,k} = \sum_{\ell \in \mathbb{Z}} p_{k-2\ell} c_{j-1,\ell} + q_{k-2\ell} d_{j-1,\ell} .$$

#### 2.4.5. Wavelet-Packets

Zur Vervollständigung der Darstellung wird hier kurz auf die *Wavelet-Packet-Transformation* (WPT) eingegangen, die eine Erweiterung und Generalisierung der Wavelet-Transformation darstellt [88, 109]. Auch können die Konstruktionen einiger Wavelets, wie z.B. der Daubechies-Wavelets, auf *Wavelet-Packets* übertragen werden.

Das Prinzip läßt sich gut anhand der Abbildung 2.9 beschreiben. Die dort schematisch dargestellte Multiskalen-Analyse eines Signals besitzt eine einseitige Baumstruktur. Bei der *Wavelet-Packet-Transformation* werden auch die Wavelet-Räume  $W_j$  weiter zerlegt, so daß sich eine symmetrische zweiseitige Baumstruktur ergibt (siehe z.B. [21, 24, 88, 109, 119]). Damit wird das Signal in mehr Anteile als bei der Wavelet-Transformation zerlegt, die allerdings redundant sind und damit keine Basis bilden. Mit einer derartigen feinen Zerlegung eines Signals ergeben sich weitere Möglichkeiten zur Signalverarbeitung, z.B. eine noch effizientere Kodierung des Signals.

Hierfür muß aber zunächst eine Basis von nicht redundanten Signalanteilen durch die Auswahl entsprechender *Wavelet-Packet*-Räume definiert werden, was einen weiteren Schritt in der Datenverarbeitung nötig macht: die Auswahl der besten Basis (*best bases selection*) [88, 119–121].

Da in der vorliegenden Arbeit keine Signale verarbeitet wurden, für die derartige Zerlegungen erforderlich sind, wurden *Wavelet-Packets* hier nicht verwendet.

### 2.4.6. Analyse von Funktionen auf kompakten Intervallen

Viele Funktionen und die meisten realen Signale, wie z.B. NIR-Spektren, sind nur auf einem kompakten Intervall  $[a, b]$  definiert. Dies führt dazu, daß die Signalzerlegung mittels Multiskalen-Analyse aufgrund der sukzessiven Halbierung der Auflösung auf eine bestimmte Anzahl von Skalen begrenzt ist. Bei Erreichen eines Raumes  $V_0$  ist keine Verkleinerung der Auflösung mehr möglich und somit bricht die Analyse bei  $V_0$ , der größten Approximation des Signals, ab. Die vollständige Zerlegung von  $L_2([a, b])$  in Wavelet-Räume lautet damit:

$$L_2([a, b]) = V_0 \oplus \bigoplus_{j \in \mathbb{N}_0} W_j .$$

Der verbleibende Teilraum  $V_0$  läßt sich mit Hilfe der Skalierungsfunktionen  $\phi_{0,k}$  beschreiben. In  $V_0$  sind jene Signalanteile kodiert, die aufgrund der verschwindenden Momente (siehe Abschnitt 2.3.4) nicht mit der gegebenen Wavelet-Funktion erfaßt werden. Mit den Gleichungen 2.10 und 2.11 erhält man folgende Darstellung für  $f \in L_2([a, b])$ :

$$f = \underbrace{\sum_{k \in J_0} c_{0,k} \phi_{0,k}}_{V_0} + \sum_{j \in \mathbb{N}_0} \sum_{k \in J_j} d_{j,k} \psi_{j,k} . \quad (2.12)$$

Abbildung 2.10 zeigt die Multiskalen-Analyse des simulierten Spektrums  $f$  aus Abbildung 2.7 auf der Seite 22. Klar zu erkennen ist die Verteilung der Signalanteile nach Frequenzen auf die Wavelet-Räume. So enthält der Wavelet-Raum  $W_7$  nur hochfrequentes Rauschen, während die folgenden Räume  $W_6$  bis  $W_3$  u.a. Anteile der beiden Banden enthalten. Parallel zur Analyse niederfrequenter Anteile wird die Approximation in den Funktionsräumen  $V_j$  zum Raum  $V_0$  hin immer größer.

### 2.4.7. Behandlung von Randeffekten

Ein wichtiger Gesichtspunkt bei der Analyse von Signalen auf kompakten Intervallen ist, wie diese am Rand des Intervalls  $[a, b]$  zu behandeln sind, damit keine sogenannten Randeffekte auftreten.

Beim Übergang von Signalen aus  $L_2(\mathbb{R})$  mit unendlicher Ausdehnung zu Signalen auf kompakten Intervallen  $L_2[a, b]$  treten Probleme an den Rändern auf. Außerhalb des Intervalls ist das Signal nicht definiert und daher stehen an den beiden Ränder nicht genügend Daten für die Durchführung einer Transformation zur Verfügung.

Um dieses Problem zu lösen gibt es zwei Möglichkeiten: Das Signal wird über die Intervallgrenzen hinaus erweitert oder die Transformation bzw. die zugehörigen Basisfunktionen werden für die Erfassung der Randbereiche angepaßt.

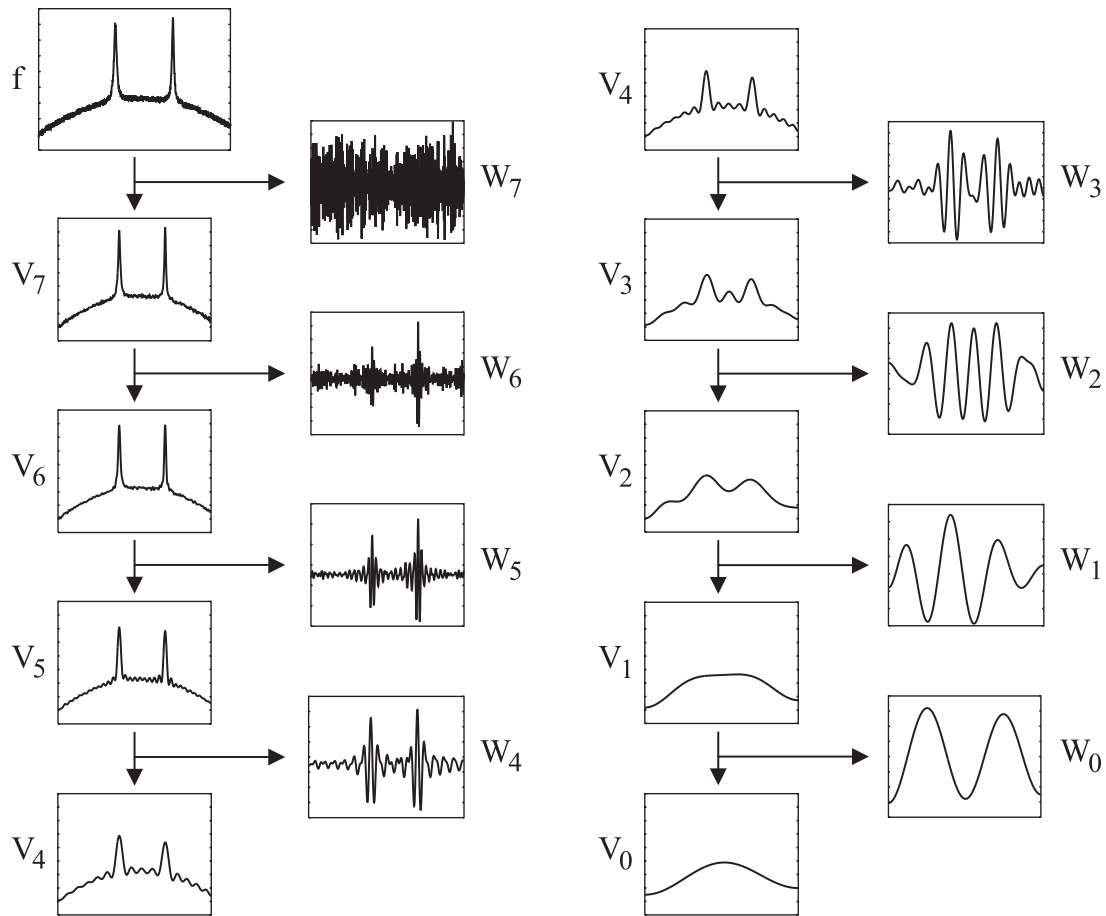


Abbildung 2.10.: Multiskalen-Analyse eines simulierten Spektrums.

Für die Erweiterung eines Signals über das Intervall  $[a, b]$  hinaus werden üblicherweise folgende Techniken angewendet [25, 90, 109, 122]:

1. Fortsetzung mit Nullen (*zero padding*):

$$f(t) := 0 \quad \text{für } t < a \text{ oder } t > b .$$

Hierbei entsteht allerdings im allgemeinen eine Unstetigkeitsstelle beim Übergang vom Signal auf die Erweiterung.

2. Periodisierung:

$$f(t) = f(t + T) \quad ; \quad T = \text{Signallänge}$$

An den beiden Intervallgrenzen wird das ursprüngliche Signal angehängt. Aufgrund verschiedener Randwerte entstehen dabei in der Regel ebenfalls Unstetigkeitsstellen.

3. Polynomiale Erweiterung (*smooth padding*):

Das Signal wird durch Polynome erweitert, womit eine stetige Fortführung erreicht wird.

4. Symmetrisierung (*symmetric extension*):

$$f(t) := \begin{cases} f(2a \leftrightarrow t) & \text{für } t < a , \\ f(2b \leftrightarrow t) & \text{für } t > b . \end{cases}$$

Das Signal wird symmetrisch an den Intervallgrenzen gespiegelt.

5. Punktspiegelung (*reflection*):

$$f(t) := \begin{cases} 2f(a) \leftrightarrow f(2a \leftrightarrow t) & \text{für } t < a , \\ 2f(b) \leftrightarrow f(2b \leftrightarrow t) & \text{für } t > b . \end{cases}$$

Das Signal wird an den Intervallgrenzen punktsymmetrisch gespiegelt.

In den ersten beiden Fällen entstehen Sprünge an den Rändern des erweiterten Signals. Da es mit der Wavelet-Transformation möglich ist, derartige Sprünge zu detektieren, erhält man entsprechend große Rand-Koeffizienten. Allerdings sind diese nicht auf das ursprüngliche Signal zurückzuführen, sondern es handelt sich um Artefakte, die durch die Erweiterung entstehen. Wird die Wavelet-Transformierte z.B. im Sinne eines *denoising* verändert, so wirken sich diese Artefakte stark auf das rekonstruierte Signal aus (siehe Abschnitt 3.2.2).

Bei der Signalerweiterung mittels Symmetrisierung bzw. Punktspiegelung treten am Rand Knickstellen im Signal bzw. der 1. Ableitung auf, die zu Sprüngen in der 1. bzw. 2. Ableitung führen (siehe Abbildung 2.11). Daher kann auch die Verwendung dieser Techniken unerwünschte Randeffekte hervorrufen. Ähnliches gilt in gleicher Weise für eine polynomiale Erweiterung von Signalen (*smooth padding*).

Ein anderer Ansatz derartige Randprobleme zu lösen, ist die Modifizierung der Wavelets, die für die Signal-Analyse am Rand sorgen. Damit verbunden sind komplizierte und rechenaufwendige Anpassungen, deren Komplexität es zunächst verhindert hat, daß derartige Ansätze (z.B. [123, 124]) in der Praxis Verwendung fanden.

Ein praktikabler Ansatz wurde von *T. Kilgore* und *J. Prestin* veröffentlicht, die sogenannte Tschebyscheff-Wavelets speziell für Signale auf kompakten Intervallen konstruierten [113]. Auch die von *U. Depczynski* entwickelten Sturm-Liouville-



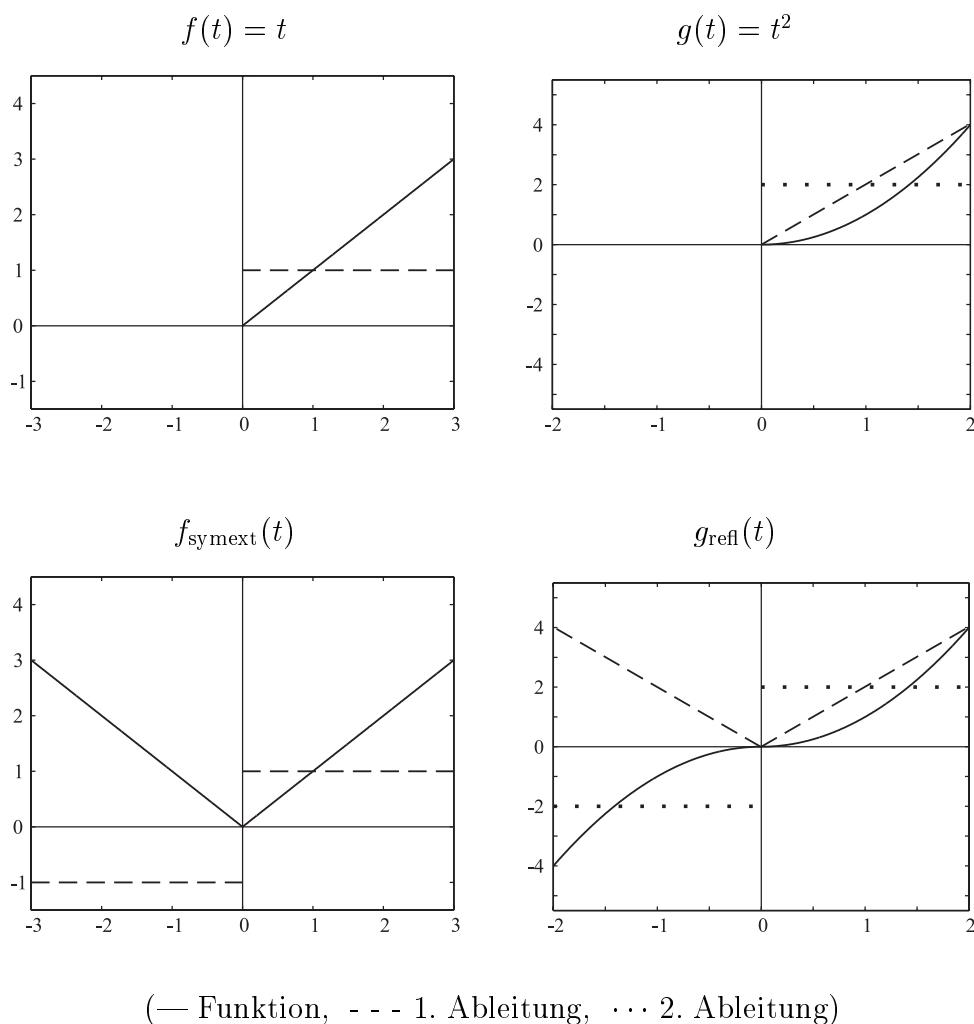


Abbildung 2.11.: Signalerweiterungstechniken zur Minderung von Randeffekten: Symmetrisierung (links) und Punktspiegelung (rechts) [25].

Wavelets<sup>11</sup> [114, 125, 126] sind an die besonderen Randbedingungen angepaßt, womit Signalerweiterungen und die damit verbundenen Effekte entfallen. Außerdem stehen für beide Typen schnelle Zerlegungs- und Rekonstruktionsalgorithmen zur Verfügung, deren Komplexität im Bereich einer schnellen Fourier-Transformation liegen. Die Anwendung von Sturm-Liouville- und Tschebyscheff-Wavelets (s- und t-Wavelets) ist Teil dieser Arbeit.

<sup>11</sup>Die Bezeichnung drückt aus, daß die Konstruktion dieser Wavelets auf Eigenlösungen von Sturm-Liouville-Eigenwertproblemen beruht.

## 2.5. Kriterien und Algorithmen zur Datenverarbeitung mit Wavelet-Koeffizienten

Die Transformation eines Signals  $f$ , das aus  $N$  diskreten Datenpunkten besteht, in den Waveletbereich läßt sich wie folgt formulieren:

$$w = Wf ,$$

mit  $W$  als Matrix für eine diskrete Wavelet-Transformation (DWT) (siehe Gleichung 2.12). Das Ergebnis ist ein Vektor  $w$ , der die Wavelet-Koeffizienten  $d_{j,k}$  enthält. Jeder Koeffizient  $d_{j,k}$  repräsentiert einen bestimmten Signalanteil. Der Index  $j$  kennzeichnet das zugehörige Frequenzband (Skala), während  $k$  die Position des Koeffizienten im Band  $j$  angibt.

Für die weitere Datenverarbeitung stehen nun zwei Vorgehensweisen offen [24]:

1. Nach der Veränderung des Signals im Waveletbereich ( $w \rightarrow \hat{w}$ ) wird durch die inverse Transformation das entsprechend veränderte Signal  $\hat{f}$  im Zeitbereich erhalten:

$$\hat{f} = W^{-1}\hat{w} . \quad (2.13)$$

Beispielsweise läßt sich die Filterung von Frequenzanteilen im Sinne einer Glättung (*smoothing* und *denoising*, siehe Abschnitt 3.2) auf diese Art einfacher durchführen als im Zeitbereich.

2. Die Wavelet-Koeffizienten können auch direkt zur weiteren Signalverarbeitung genutzt werden; es erfolgt also keine Rücktransformation. Für die entsprechenden Anwendungen werden relevante Koeffizienten herausgefiltert, während die übrigen verworfen werden ( $w \rightarrow \hat{w}$ ). Im allgemeinen ist damit eine z.T. beträchtliche Komprimierung der Ausgangsdaten verbunden, ohne daß wichtige Informationsanteile verlorengehen.

In beiden Fällen sind Algorithmen und Kriterien für die Auswahl relevanter Koeffizienten nötig, von denen einige hier vorgestellt werden.

### 2.5.1. Koeffizientenauswahl nach Skalen

Die Analyse der Frequenzanteile von Signalen unter Erhaltung der Zeitinformation ist eine wichtige Eigenschaft der Wavelet-Transformation. Wie im Abschnitt 2.4 gezeigt, führt die Multiskalen-Zerlegung dazu, daß Signalkomponenten bestimmter Frequenzen durch die Koeffizienten  $d_{j,k}$  bestimmter Skalen  $j$  repräsentiert werden.

Durch Eliminierung von Wavelet-Skalen können Signalanteile des zugehörigen Frequenzbereiches komplett entfernt werden.

Werden z.B. höhere Skalen ausgeschlossen, wird nach der Rekonstruktion ein geglättetes Spektrum erhalten (siehe Abschnitt 3.2). Ohne eine Rücktransformation erhält man eine deutliche Datenreduktion, da sich mit dem Skalen-Index die Anzahl der Wavelet-Koeffizienten in etwa verdoppelt. Die Datenreduktion bei Eliminierung von  $n$  höheren Skalen läßt sich mit  $1 : 2^n$  abschätzen.

Dieses Vorgehen ist aber nur erfolgversprechend, wenn für ein gegebenes Signal feststeht, daß die ausgeschlossenen Skalen keine relevanten Informationsanteile enthalten. So können bei vielen Arten von NIR-Spektren je nach Wavelet-Typ bis zu drei der höchsten Skalen weggelassen werden, ohne die spektrale Struktur bis auf einen oft sogar gewünschten Glättungseffekt zu beeinflussen. Im Gegensatz dazu würden z.B. Raman-Spektren aufgrund der niedrigen Peak-Halbwertsbreiten stärker verändert und eventuell unbrauchbar. Dagegen könnten in diesem Falle niedrigere Frequenzbänder weggelassen werden, was aber zu keiner deutlichen Datenreduktion führen würde, da jene Bänder nur relativ wenige Koeffizienten enthalten (siehe z.B. Abbildung 2.7, Seite 22).

### 2.5.2. Koeffizientenauswahl durch thresholding

Eine andere Möglichkeit, als Koeffizienten nach Skalen zu selektieren, ist die Auswahl nach dem Betrag ihrer Amplitude. Da die Wavelet-Transformation linear ist und die Gesamtenergie erhalten bleibt, gibt die Amplitude eines Koeffizienten  $|d_{j,k}|$  den von ihm repräsentierten Anteil an der Signalenergie wieder. Die Energie des Signals wird meist in relativ wenigen großen Koeffizienten konzentriert. Dagegen wird z.B. weißes Rauschen über die gesamte Wavelet-Transformierte verteilt und führt zu relativ kleinen Koeffizienten, da es sich unter allen orthogonalen Transformationen invariant verhält [5].

Mit *thresholding*, d.h. der Festlegung eines Schwellenwertes, werden Koeffizienten mit größeren Amplituden und somit größerem Informationsgehalt von anderen, die Minoritätsanteile darstellen, abgetrennt. Dafür wird ein Schwellen- oder *threshold*-Wert *thr* bestimmt, der angibt, ab welchem Amplitudenbetrag Koeffizienten erhalten bleiben sollen. Die Selektion der Signalanteile erfolgt hierbei unabhängig von ihrer Frequenz.

Da *thresholding* meist im Sinne von *denoising* (siehe Abschnitt 3.2) zur Eliminierung von Rauschen eingesetzt wird, orientiert sich die Bestimmung des Parameters *thr* in diesen Fällen am Rauschanteil im Signal. Dabei wird ein gemessenes Signal  $f$  als die Summe aus dem zu bestimmenden Signal  $\hat{f}$  und weißem Rauschen  $\varepsilon$  be-

trachtet [127]:

$$f = \hat{f} + \varepsilon .$$

Ziel ist es nun, für  $\hat{f}$  nach Gleichung 2.13 einen modifizierten Koeffizientensatz  $\hat{w}$  zu bestimmen.

Für Signale, die weißes Rauschen enthalten, hat *D. L. Donoho* den sogenannten *universal threshold* definiert [128–130]. Dabei wird für  $f$  zunächst auf der ersten Stufe der Wavelet-Zerlegung ( $W_j$ , höchste Skala) die Standardabweichung  $\sigma$  des Rauschens bestimmt. Der *threshold* wird nun nach

$$thr = \sqrt{2 \log(N)} \cdot \sigma$$

berechnet, wobei  $N$  die Anzahl der Datenpunkte ist. Im Ergebnis bleiben also unabhängig von der Frequenz alle Koeffizienten erhalten, deren Amplituden betragsmäßig über denen von Koeffizienten liegen, die weißes Rauschen repräsentieren.

Der Ansatz hat einige bemerkenswerte Eigenschaften und Vorteile gegenüber anderen. Obwohl er auf eine sehr umfangreiche Klasse von Signalen anwendbar ist, muß als Parameter nur  $\sigma$  festgelegt werden, was aber auch automatisierbar ist. Im Gegensatz zu herkömmlichen Rauschminderungsverfahren werden insbesondere keine Informationen über die Glattheit (Struktur) von  $f$  benötigt, um eine optimale und rauschfreie Approximation  $\hat{f}$  mit identischer Glattheit zu erhalten [5, 130]. Allerdings kann es zu leichtem *underfitting* kommen, d.h. die optimale Approximation des zugrundeliegenden Signals wird nicht ganz erreicht [127].

Für die Bestimmung eines *thresholds*  $thr$  gibt es noch andere Ansätze, auf die hier nur verwiesen wird, da sie im weiteren nicht angewendet werden.

- Im Vergleich zu anderen Glättungsmethoden werden in [131] verschiedene Kriterien zur Festlegung von  $thr$  vorgestellt und getestet.
- *G. P. Nason* [127] stellt die Bestimmung von  $thr$  mittels *cross-validation* vor, wobei der mittlere quadratische Fehler für die Approximation von  $f$  durch  $\hat{f}$  minimiert wird.
- In [23] wird als *threshold* ein Prozentsatz von der Amplitude des größten Wavelet-Koeffizienten festgelegt. Diese signalabhängige Festlegung kann ebenfalls durch Minimierung des Approximationsfehlers optimiert werden.

Unabhängig von der Methode zur Bestimmung von  $thr$  kann *thresholding* auf zwei Arten angewendet werden. Beim restriktiven *hard-thresholding* werden Koeffizienten auf 0 gesetzt, deren Betrag den *threshold*-Wert unterschreiten, während die übrigen

unverändert bleiben:

$$d_{j,k} = \begin{cases} 0 & , \text{ falls } |d_{j,k}| < thr , \\ d_{j,k} & , \text{ falls } |d_{j,k}| \geq thr . \end{cases}$$

Dagegen werden beim *soft-thresholding* — auch *wavelet shrinkage* genannt — vom Betrag der Koeffizienten, die nicht 0 gesetzt werden, der *threshold*-Wert subtrahiert:

$$d_{j,k} = \begin{cases} 0 & , \text{ falls } |d_{j,k}| < thr , \\ \text{sign}(d_{j,k})(|d_{j,k}| \ominus thr) & , \text{ falls } |d_{j,k}| \geq thr , \end{cases}$$

wobei sich der Wert von  $\text{sign}(d_{j,k})$  nach dem Vorzeichen des Koeffizienten  $d_{j,k}$  richtet:  $\text{sign}(d_{j,k}) = 1$  für  $d_{j,k} > 0$  und  $\text{sign}(d_{j,k}) = -1$  für  $d_{j,k} < 0$ .

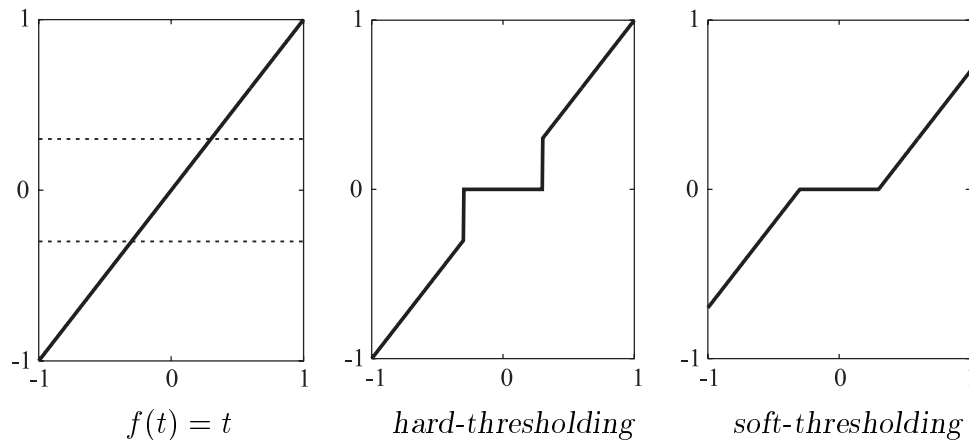


Abbildung 2.12.: *Hard- und soft-thresholding* der Funktion  $f(t) = t$ , *threshold*-Wert 0,3 [109].

Abbildung 2.12 zeigt am Beispiel einer einfachen Funktion die unterschiedlichen Wirkungsweisen von *hard-* und *soft-thresholding*.

Übertragen auf *thresholding* von Wavelet-Koeffizienten bedeutet dies, daß beim *soft-thresholding* die Stufen zwischen den erhalten gebliebenen und den auf Null gesetzten Koeffizienten verkleinert werden. Dadurch werden im allgemeinen glattere Rekonstruktionen als nach *hard-thresholding* erhalten, wobei allerdings auch Details verloren gehen können. Wie stark sich beide Methoden im Ergebnis unterscheiden hängt auch von der Glattheit des verwendeten Wavelets ab. Generell ist keiner der beiden Methoden vorab der Vorzug zu geben, sondern die Wahl ist in Abhängigkeit vom Signal und der Anwendung zu treffen.

### 2.5.3. Koeffizientenauswahl nach dem Energie-Kriterium

Bei diesem Kriterium werden ebenfalls die Amplituden der Wavelet-Koeffizienten herangezogen, um relevante von nicht relevanten Signalanteilen zu trennen. Zunächst werden die Koeffizienten skalenunabhängig nach ihrem Betrag, also nach abnehmendem Anteil an der Energie des Signals (Gleichung 2.7, Seite 27), sortiert. In den modifizierten Koeffizientenvektor  $\hat{w}$  werden nun solange Koeffizienten nach abnehmendem Betrag übernommen, bis ein festgelegter Anteil an der Gesamtenergie erreicht ist. Als einziger signalabhängiger Parameter muß hier also der prozentuale Anteil an der Signalenergie, der erhalten bleiben soll, vorgegeben werden.

### 2.5.4. Koeffizientenauswahl mittels Genetischer Algorithmen

Bei der quantitativen spektrometrischen Analyse auf der Basis von Wavelet-Koeffizienten muß diejenige Kombination von Koeffizienten herausgefunden werden, die zum besten Kalibrationsmodell führt. Da es aufgrund der „kombinatorischen Explosion“ nicht möglich ist, alle Kombinationsmöglichkeiten zu berechnen, gilt es ein abkürzendes Verfahren anzuwenden. Zu diesem Zweck kommen in der vorliegenden Arbeit Genetische Algorithmen zum Einsatz.

## Genetische Algorithmen in der Analytischen Chemie

Genetische Algorithmen (GA) werden hauptsächlich zur Lösung kombinatorischer Optimierungsprobleme eingesetzt, die effizient und in einem vertretbaren Zeitraum gelöst werden müssen. Da auch im Bereich der Analytischen Chemie und der Chemometrie derartige Problemstellungen auftreten, werden Genetische Algorithmen hier verstärkt eingesetzt. Ein Anwendungsschwerpunkt in der Chemometrie ist die Auswahl von Variablen bei der Erstellung multivariater Kalibrationsmodelle in der quantitativen NIR-Spektrometrie:

- Selektion von Wellenlängen bzw. Wellenzahlen oder auch zusammenhängenden Spektralbereichen in Verbindung mit MLR (*Multiple Linear Regression*) und PLS (*Partial Least Squares Regression*) [132–141];
- Bestimmung der optimalen Faktorkombination für die *Principal Component Regression* (PCR) durchgeführt von unserer [142–144] und einer anderen [145] Arbeitsgruppe;
- Auswahl von Fourier-Koeffizienten [146];

- Auswahl von Wavelet-Koeffizienten, durchgeführt im Rahmen der vorliegenden Arbeit [41, 42].

Neben den Anwendungsbeispielen in diesem Bereich wurden auch einführende Artikel für Chemiker und Chemometriker veröffentlicht [147–149]. Einen guten Einstieg in die Thematik bietet das „*Handbook of Genetic Algorithms*“ von *L. Davis* [150]. Auf letzterem baut neben der folgende Kurzbeschreibung auch der Genetische Algorithmus auf, der im Rahmen dieser Arbeit und des DFG-Projektes „NIR/Wavelets“ zur Selektion von Wavelet-Koeffizienten für die quantitative NIR-Spektrometrie entwickelt wurde.

### Das Funktionsprinzip und der Aufbau Genetischer Algorithmen

Das Vorbild für die Entwicklung Genetischer Algorithmen ist der natürliche Kreislauf der Evolution. Nach der von *Charles Darwin* 1859 veröffentlichten Evolutionstheorie beruht die Entstehung und Veränderung der Arten auf einem Kreislauf aus Reproduktion, Mutation und Selektion.

In einem „Kampf ums Dasein“ (*struggle for life*) überleben nur die am besten an die Umwelt angepaßten Individuen einer Art (*survival of the fittest*) und pflanzen sich rascher als die übrigen fort<sup>12</sup>. Diese natürliche Auslese besteht auch zwischen Arten und führt daher zur Ausbildung von neuen veränderten Arten, die andere ökologische Nischen besetzen [152].

Innerhalb von Genetischen Algorithmen werden zur Lösung eines Optimierungsproblems die gleichen Mechanismen angewendet, wie sie im Kreislauf der Evolution wirksam sind (siehe Tabelle 2.1). Als Individuen werden in einem GA potentielle Lösungen des Problems betrachtet, die kodiert als Chromosomen vorliegen und von Generation zu Generation verändert werden können.

Beim Ablauf eines Genetischen Algorithmus (siehe Abbildung 2.13 auf der Seite 45) wird die erste Generation einer Population von Chromosomen zufällig generiert (Initialisierung). Diese Startpopulation enthält in der Regel nicht die gesuchte optimale Lösung, sondern neben einigen besser geeigneten auch viele völlig ungeeignete Lösungen.

---

<sup>12</sup>Das Ergebnis einer solchen Entwicklung entdeckte *Darwin* auf den Galápagosinseln, wo er eine Gemeinschaft verschiedener Finkenarten vorfand, deren Ähnlichkeit er auf eine gemeinsame Ahnenform zurückführte. Durch die Spezialisierung auf bestimmte Nahrung und die Ausbildung ganz unterschiedlicher Schnabelformen haben sich einzelne Arten entwickelt. So konnten die Nahrungsquellen des Biotops optimal genutzt werden und jede Art sicherte damit ihr Überleben [151].

Tabelle 2.1.: Grundzüge der Evolutions-Theorie nach *Charles Darwin* und ihre Umsetzung in Genetischen Algorithmen.

	<b>Natürlicher Kreislauf der Evolution</b>	<b>Genetische Algorithmen</b>
<b>Population von Individuen</b>	Chromosomen tragen die Erbanlagen	Chromosomen sind potentielle Lösungen
<b>Selektionskriterium</b>	Anpassung an die Umwelt ( <i>Survival of the fittest</i> )	Fitness- oder Zielfunktion
<b>Rekombination</b>	Fortpflanzung der „starken“ Individuen	Rekombination von Chromosomen mit hoher Fitness
<b>Mutation</b>	Zufällige Veränderung des Erbgutes	Zufällige Veränderung von Chromosomen
<b>Gendrift</b>	Zu-/Abwanderung von Individuen	Zufällige Generierung neuer Chromosomen

Die Güte aller generierten Lösungen wird mit Hilfe einer Fitnessfunktion (Zielfunktion) bewertet und eine Rangfolge aufgestellt. Die Wahl der Fitnessfunktion ist der entscheidende Schritt für die Anpassung eines Genetischen Algorithmus an ein zu lösendes Problem, da sie das Selektionskriterium für die Optimierung darstellt. So wird z.B. bei der Erstellung von Kalibrationsmodellen in der quantitativen Spektrometrie der Analysenfehler als Optimierungskriterium herangezogen.

Im Laufe der weiteren Optimierung werden ungeeignete Lösungen verworfen und in der nächsten Generation durch neu gebildete Chromosomen ersetzt. Die Bildung neuer Chromosomen beruht auf den gleichen Mechanismen, wie sie in der Natur wirksam sind: die Rekombination und Mutation von potentiell gut geeigneten Chromosomen, sowie die Einbringung von neuen Chromosomen. In einem Genetischen Algorithmus werden diese Veränderungen mit bestimmten Techniken erreicht und durch Operatoren gesteuert, die im Anhang unter Abschnitt A.3 auf der Seite 145 vorgestellt werden.



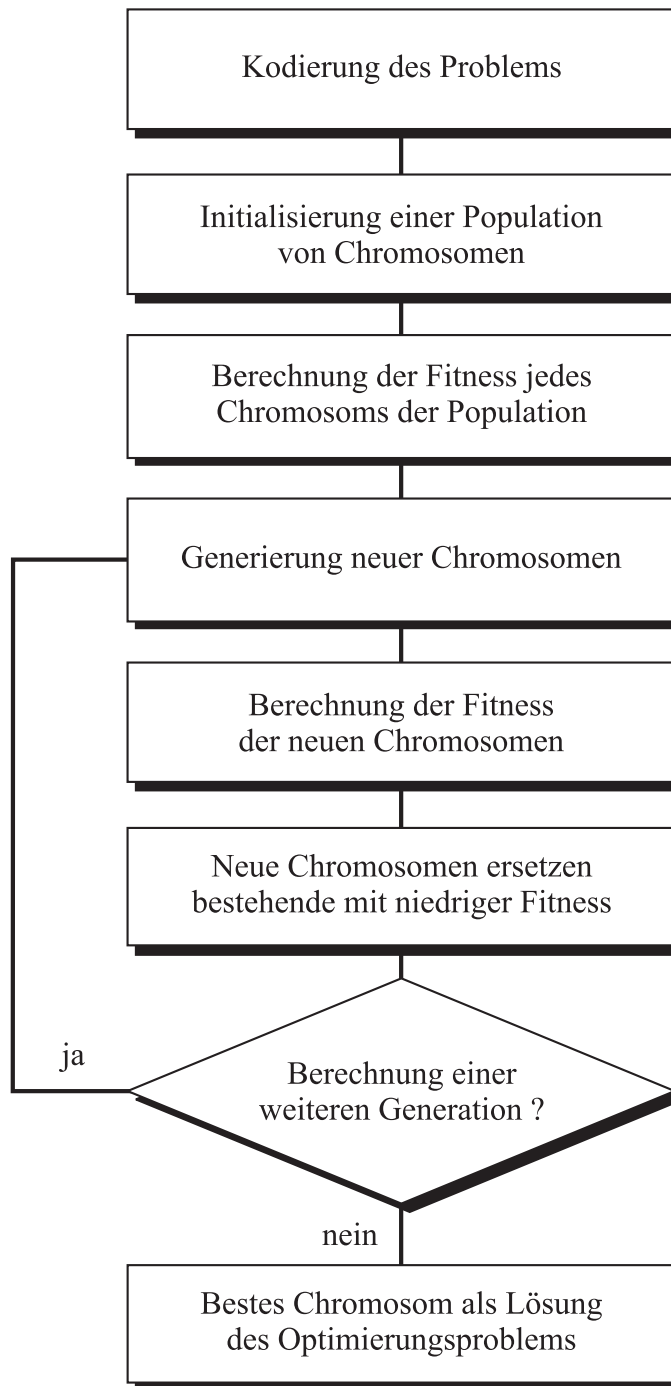


Abbildung 2.13.: Ablaufschema eines Genetischen Algorithmus [143].

Besonders durch die Einbeziehung des bereits vorhandenen und als geeignet bewerteten „genetischen Materials“ in die fortlaufende Bildung von neuen Chromosomen konvergiert die Population sehr schnell in Richtung der optimalen Lösung bezüglich der verwendeten Fitnessfunktion. In der Ausnutzung von Information, die bereits in der Population enthalten ist, liegt ein großer Vorteil genetischer Algorithmen gegenüber zufälligen Suchverfahren (*Random Search* [153]). Damit läßt sich ein kombinatorisches Optimierungsproblem rasch lösen, ohne alle möglichen Kombinationen durchzurechnen, was außerdem ab einer größeren Zahl von Variablen kaum mehr praktikabel ist.

Auch ist die Gefahr, mit Genetischen Algorithmen nur ein lokales Optimum zu erreichen, deutlich geringer als bei anderen Optimierungsmethoden. Betrachtet man alle möglichen Lösungen als „Lösungsraum“, so wird dieser mit einem GA gleichzeitig an verschiedenen Stellen untersucht. Dagegen gehen andere Verfahren (Simplex-Verfahren [154–156], *Stochastic Hill Climbing* [153], *Tabu-Search*-Verfahren [157, 158]) von einem Punkt in diesem Raum aus und bleiben bei der schrittweise Optimierung leichter in einem lokalen Optimum „hängen“. Bei Gradienten-Verfahren kommt hinzu, daß die Fitnessfunktion stetig und differenzierbar sein muß [159]. All diese Verfahren sind nur zur Lösung einfacherer Problemstellung geeignet. Damit sind Problemstellungen gemeint, deren Lösungsräume nur wenige lokale Optima aufweisen.

Ein Verfahren, daß trotz schrittweiser Optimierung in der Lage ist ein globales Optimum zu erreichen, ist das *Simulated Annealing* (SA) [134, 160, 161]. Dieses vereint Merkmale des *Random Search* sowie des *Hill Climbing* und ist in vielen Fällen von vergleichbarer Leistungsfähigkeit wie ein GA [153, 155, 156].

Insbesondere bei komplexen Optimierungsproblemen besitzen Genetische Algorithmen jedoch aufgrund der Abdeckung des gesamten Lösungsraumes und der Verwertung von bereits vorhandener Information deutliche Vorteile. Bei entsprechender Auslegung zeigen diese eine hohe Konvergenz und Robustheit bei akzeptablen Rechenzeiten [162].

Die Konvergenz eines GA in Richtung der optimalen Lösung und die Reproduzierbarkeit der Resultate hängt im wesentlichen von der Populationsgröße  $p$  und der Anzahl der zu berechnenden Generationen  $n_g$  ab. Diese beiden Parameter werden in Abhängigkeit von der Komplexität des kombinatorischen Problems festgelegt.

Neben der Kodierung eines Optimierungsproblems ist die Definition des Optimierungskriteriums von grundlegender Bedeutung. Mit dieser Funktion muß die Güte potentieller Lösungen eindeutig bewertet werden können. Außerdem ist die Fitnessfunktion so auszulegen, daß sie genau einen Wert für die Bewertung eines Chromosoms liefert. Ein Genetischer Algorithmus konvergiert in Richtung eines Ma-

ximums bzw. Minimums dieser Fitnessfunktion.

Die Entwicklung einer Fitnessfunktion zur Optimierung von quantitativen Kalibrationsmodellen im Bereich der NIR-Spektrometrie war ein wichtiger Teil der vorliegenden Arbeit. Nähere Einzelheiten, auch zu Besonderheiten des verwendeten Genetischen Algorithmus, sind in Abschnitt 3.4.1 ab Seite 78 zu finden.

### 2.5.5. Klassifizierung mit Hilfe Neuronaler Netze: Lernende Vektorquantisierung (LVQ)

Im Rahmen der vorliegenden Arbeit wurden neben quantitativen auch qualitative Fragestellungen untersucht. Für die Klassifizierung bzw. Identifizierung verschiedener Materialien wurden Wavelet-transformierte NIR-Spektren verwendet. Im Gegensatz zu der im letzten Abschnitt beschriebenen Problematik der Auswahl von Wavelet-Koeffizienten mittels Genetischer Algorithmen war hierbei ein Verfahren anzuwenden, mit dem qualitative Modelle erstellt werden konnten. Für diesen Zweck wurde ein Verfahren auf Basis eines Neuronalen Netzes ausgewählt, dessen Struktur und Arbeitsweise im folgenden beschrieben wird.

Das von *T. Kohonen et al.* entwickelte Verfahren der Lernenden Vektorquantisierung (*learning vector quantization*, LVQ) [163] ist ein überwachtes Lernverfahren, d.h. die Klassenzuordnung der eingegebenen Objekte muß zur Steuerung des Lernprozesses bekannt sein. Im Falle der in dieser Arbeit verwendeten Datensätze bilden NIR-Spektren oder deren Wavelet-Transformierten die Eingabevektoren, die Klassen zugeordnet werden.

Der zugrundeliegende Algorithmus entspricht einem sehr einfachen Neuronalen Netz mit einer Schicht von Neuronen, den sogenannten Kohonen-Neuronen. Wie Abbildung 2.14 zeigt, ist die Eingabeschicht über Gewichtungs- oder Referenzvektoren (*Codebook*-Vektoren) direkt mit diesen Neuronen verbunden. Das Ziel des Lernprozesses ist eine Anpassung der Gewichtungsvektoren, so daß ähnliche Eingabevektoren den gleichen oder benachbarten Neuronen zugeordnet werden. Dies entspricht einer Abbildung (*mapping*) der  $n$ -dimensionalen Eingabevektoren auf die Neuronenschicht [164]. Die folgende Beschreibung des Algorithmus ist an der Darstellung von *A. Zell* orientiert [165].

Zu Beginn des Trainingprozesses werden in dem durch die Eingabevektoren  $X = (x_1, \dots, x_n)$  aufgespannten  $n$ -dimensionalen Eingaberaum Gewichtungsvektoren so plaziert, daß diese ihn möglichst gut abdecken. Um eine gute Konvergenz des iterativen Algorithmus zu gewährleisten, werden auch Duplikate von Eingabevektoren als Gewichtungsvektoren verwendet. Für jede Klasse von Eingabevektoren gibt es mehrere Gewichtungsvektoren.

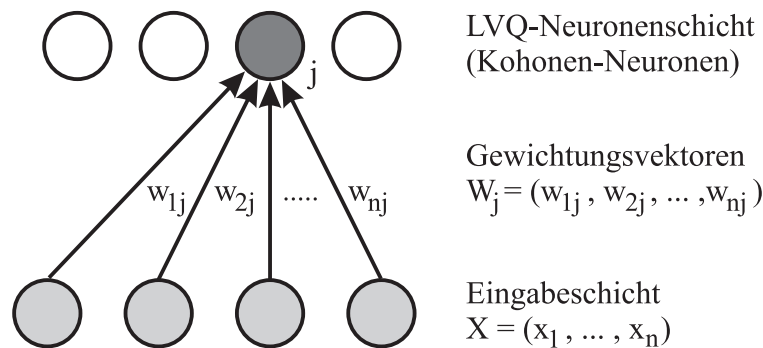


Abbildung 2.14.: Lernende Vektorquantisierung (LVQ): Schema der Netzwerkstruktur [165].

Im weiteren Verlauf wird nun ein neuer oder bereits trainierter Eingabevektor  $X$  mit allen Gewichtungsvektoren  $W_1, \dots, W_m$  verglichen. Das Kohonen-Neuron  $c$ , dessen Gewichtungsvektor  $W_c$  dem Vektor  $X$  am ähnlichsten ist, gibt die Klasse von  $X$  an. Diese Ähnlichkeit wird mathematisch durch die  $\ell_2$ -Norm bestimmt:

$$\|X \Leftrightarrow W_c\| = \min_j (\|X \Leftrightarrow W_j\|) .$$

Kern des Lernprozesses ist, daß bei jedem Iterationsschritt  $t$  eine weitere Anpassung der Gewichtungsvektoren an die Verteilung der Eingabevektoren im Raum erfolgt. Der Gewichtungsvektor  $W_c$  des Neurons  $c$  wird dem Eingabevektor  $X$  noch ähnlicher gemacht, wenn  $c$  der gleichen Klasse wie  $X$  angehört. Im umgekehrten Fall wird der Vektor  $W_c$  so verändert, daß er dem Vektor  $X$  unähnlicher wird. Dies geschieht durch Addition bzw. Subtraktion eines variablen Bruchteils  $\alpha$  des Differenzvektors  $X(t) \Leftrightarrow W_c(t)$ :

$$\begin{aligned} W_c(t+1) &= W_c(t) + \alpha(t)[X(t) \Leftrightarrow W_c(t)], & \text{falls Klasse}(W_c) = \text{Klasse}(X) , \\ W_c(t+1) &= W_c(t) \Leftrightarrow \alpha(t)[X(t) \Leftrightarrow W_c(t)], & \text{falls Klasse}(W_c) \neq \text{Klasse}(X) . \end{aligned}$$

Alle übrigen Neuronen  $W_j$  bleiben unverändert:

$$W_j(t+1) = W_j(t), \quad \text{falls } j \neq c .$$

Der Faktor  $\alpha(t)$  ( $0 < \alpha(t) < 1$ ) bestimmt die Geschwindigkeit des Lernprozesses (Lernrate), die konstant gehalten oder z.B. linear mit der Zahl der Iterationen verändert werden kann. Bei dem hier eingesetzten OLVQ1-Algorithmus (*optimized-learning-rate* LVQ1 [163]) werden für jedes Neuron  $c$  zur Erhöhung der Konvergenz individuelle Lernraten  $\alpha_c(t)$  verwendet.

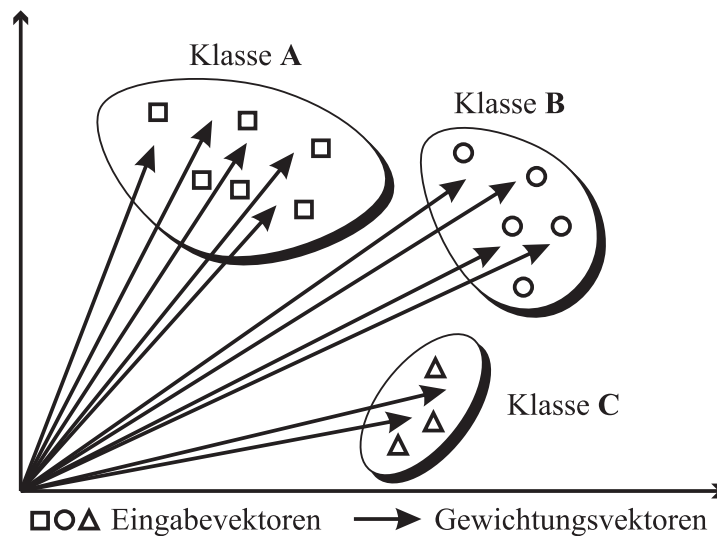


Abbildung 2.15.: Lernende Vektorquantisierung (LVQ): Klassifizierung mittels Gewichtungsvektoren (*Codebook*-Vektoren) [165].

Nach Abschluß des Lernprozesses geben die Gewichtungsvektoren die Topologie der Eingabedaten wieder, indem sie Cluster umschreiben, denen die einzelnen Klassen zugeordnet werden können (siehe Abbildung 2.15). Bei der Verwendung eines trainierten LVQ-Modells zur Klassifizierung eines unbekanntem Eingabevektors  $Y$  wird dieser der Klasse des Neurons  $W_c$  zugeordnet, mit dem der ähnlichste, also der nächste benachbarte Gewichtungsvektor verbunden ist.

Ein Vorteil dieser Klassifizierungsmethode ist, daß die einzelnen im Raum definierten Cluster keiner vorgegebenen Geometrie entsprechen müssen. Durch die Positionierung der Gewichtungsvektoren werden die Grenzen der Cluster unabhängig von einer vorab festzulegenden Form definiert. Die Gefahr einer zu starken Anpassung der Cluster an die Eingabedaten (*overfitting*, siehe Abschnitt 3.4.1.2 auf Seite 82) wird dadurch gemildert, daß zu klassifizierende Objekte, die außerhalb der definierten Cluster liegen, dem nächstgelegenen Cluster zugeordnet werden.

Bei anderen Verfahren, die z.B. auf einer Hauptkomponentenanalyse (*Principal Component Analysis*, PCA) basieren, werden Cluster meist durch ihren Schwerpunkt und einen Radius, der die zugehörigen Objekte einschließt, definiert. Auch bei der Verwendung von mehreren Radien für verschiedene Raumrichtungen entstehen je nach Lage und Homogenität der Objekte größere Cluster, die sich dadurch auch eher überlappen können. Dies erschwert die fehlerfreie Klassifizierung unbekannter Objekte.

