

Karl-Rudolf Koch / Michael Schmidt

**Deterministische
und
stochastische
Signale**

**mit Anwendungen in der
digitalen Bildverarbeitung**

Mit 47 Abbildungen
Dümmlerbuch 7891

FERD.  DÜMMLER^S VERLAG · BONN

Prof. Dr.-Ing. Karl-Rudolf Koch
Dr.-Ing. Michael Schmidt

Institut für Theoretische Geodäsie der
Rheinischen Friedrich-Wilhelms-Universität Bonn

Weitere Dümmlerbücher Geodäsie / Vermessungswesen auf den Seiten 351 ff.

ISBN 3-427-78911-X

Alle Rechte, insbesondere auch die der Übersetzung, des Nachdrucks, des Vortrages, der Verfilmung und Radiosendung sowie jeder Art der fotomechanischen Wiedergabe und der Speicherung in Datenverarbeitungsanlagen, auch auszugsweise, vorbehalten. Das Fotokopieren einzelner Seiten ist nicht gestattet, mit Ausnahme der in §§ 53, 54 UrHG ausdrücklich genannten Sonderfälle.

© 1994 Ferd. Dümmlers Verlag, Kaiserstraße 31-37 (Dümmlerhaus), 53113 Bonn

Printed in Germany by WB-Druck, 87669 Rieden

Vorwort

In dem Maße, wie die digitale Registrierung von Phänomenen unserer Umwelt zunimmt, wächst auch das Bedürfnis nach der Analyse der aufgezeichneten Signale und nach Informationen über Möglichkeiten der Datenauswertung. In den entsprechenden Lehrbüchern wird durchweg entweder die Signalverarbeitung behandelt, also die Analyse deterministischer Signale, oder man beschäftigt sich mit den Zufallsprozessen, den stochastischen Signalen. Die Unterscheidung in deterministische und stochastische Signale ist häufig nicht eindeutig, da die Grenze zwischen beiden Arten von Signalen fließend ist. In diesem Buch werden sowohl die deterministischen als auch die stochastischen Signale behandelt. Dies bietet einmal den Vorteil, daß Verfahren, die zur Analyse deterministischer Signale entwickelt werden, wie Fourier-Reihe, Fourier-Transformation, z-Transformation und Filterung, für die Untersuchung der stochastischen Signale bereits zur Verfügung stehen. Zum anderen wird offensichtlich, wie eng die Methoden zur Analyse deterministischer und stochastischer Signale miteinander verflochten sind. Da beide Arten von Signalen betrachtet werden, muß der Inhalt des Buches eingeschränkt werden, was aber nur bedeutet, daß für spezielle Verfahren auf weitergehende Literatur verwiesen wird. Die Grundlagen zur Analyse digitaler Signale werden in aller Ausführlichkeit behandelt.

Die Untersuchung der deterministischen und der stochastischen Signale beginnt im Eindimensionalen. Es folgt die Verallgemeinerung auf höhere Dimensionen, in denen die zweidimensionalen Signale eingehender erörtert werden. Bekanntlich erzeugt man zweidimensionale Signale in der digitalen Bildverarbeitung, die inzwischen in der Geodäsie und Photogrammetrie so intensiv genutzt wird. Die Beispiele zur Analyse zweidimensionaler deterministischer und stochastischer Signale werden daher aus der digitalen Bildverarbeitung gewählt.

Im Kapitel 2, das den deterministischen Signalen gewidmet ist, werden die oben erwähnten Grundlagen zunächst für den eindimensionalen, dann für den mehrdimensionalen Fall erarbeitet, auf die im Kapitel 3, das die stochastischen Signale behandelt, zurückgegriffen wird. Ausführlich werden im Kapitel 2 die eindimensionalen und zweidimensionalen digitalen Filter erörtert und Methoden zu ihrer numerischen Realisierung diskutiert.

Kapitel 3 behandelt die stochastischen Prozesse und spezielle Zufallsprozesse. Die Schätzung von Momentfunktionen und des Spektrums werden erörtert sowie die Schätzung

von Parametern in speziellen Modellen für die stochastischen Prozesse, um beispielsweise das Problem der optimalen Filterung zu lösen. Kapitel 3 widmet sich intensiv den Signalen mit Markoff-Eigenschaft. Für sie lassen sich die Wahrscheinlichkeitsdichten mit Hilfe der Gibbs-Verteilungen bequem angeben. Probleme der digitalen Bildverarbeitung wie Bildrestaurierung, Kantenextraktion oder Mustererkennung können dann mit diesen Verteilungen gelöst werden.

Allen Studentinnen und Studenten der Universität Bonn sowie Mitarbeiterinnen und Mitarbeitern des Instituts für Theoretische Geodäsie, die durch Korrekturen und Verbesserungsvorschläge am Manuskript zum Erscheinen dieses Buches beigetragen haben, danken wir sehr. Unser besonderer Dank gilt den Herren Dr.-Ing. A. Busch, Dipl.-Ing. J. Klonowski, Dipl.-Ing. M. Köster und MSc. Y.J. Pan für ihre vielen Anregungen und Frau I. Wahl für die Reinschrift des Buches. Schließlich möchten wir noch die gute Zusammenarbeit mit dem Verlag dankend erwähnen.

Bonn, im Oktober 1993

Karl-Rudolf Koch und Michael Schmidt

Inhaltsverzeichnis

1	Einleitung	1
2	Deterministische Signale	3
21	Eindimensionale Signale und lineare Systeme	4
211	Signaldarstellung	4
212	Harmonische Schwingung	5
213	Fourier-Reihe	8
214	Fourier-Transformation	12
215	Abtastung kontinuierlicher Signale	18
216	Lineare Systeme	23
217	Systemfunktion	29
218	Diskrete Fourier-Transformation	33
219	Schnelle Fourier-Transformation	40
22	Eindimensionale digitale lineare Filter	45
221	Allgemeine Definitionen und Eigenschaften	45
222	Allgemeiner Filterentwurf	54
223	Definition und Eigenschaften rekursiver Filter	58
224	Entwurf rekursiver Filter	61
225	Definition und Eigenschaften nichtrekursiver Filter	71
226	Nichtrekursive Elementarfilter	78
227	Entwurf nichtrekursiver Filter	83
228	Implementierung linearer Filter	89
23	Mehrdimensionale Signale und lineare Systeme	92
231	Signaldarstellung, Fourier-Reihe und Fourier-Transformation	92
232	Abtastung kontinuierlicher Signale	94
233	Lineare Systeme	98
234	z-Transformation, Systemfunktion und diskrete Fourier-Transformation	107
235	Zweidimensionale Signale	107
24	Zweidimensionale digitale Filter	112
241	Allgemeine Definitionen und Eigenschaften	112
242	Definition und Eigenschaften zweidimensionaler rekursiver Filter	121

243	Entwurf zweidimensionaler rekursiver Filter	129
244	Definition und Eigenschaften zweidimensionaler nichtrekursiver Filter	134
245	Zweidimensionale nichtrekursive Elementarfilter	143
246	Entwurf zweidimensionaler nichtrekursiver Filter	148
247	Implementierung zweidimensionaler linearer Filter	156
248	Zweidimensionale digitale nichtlineare und verschiebungsvariante Filter	158
3	Stochastische Signale	161
31	Eindimensionale Zufallsprozesse	162
311	Definitionen	162
312	Erwartungswert-, Autokovarianz- und Kreuzkovarianzfunktion	164
313	Stationarität	166
314	Spektrum	169
32	Spezielle Zufallsprozesse	177
321	Weißes Rauschen, Gauß- und Poisson-Prozesse	177
322	Markoff-Prozesse und Markoff-Ketten	179
323	Lineare kontinuierliche Systeme	185
324	Lineare diskrete Systeme und digitale Filter	188
325	Autoregressive Prozesse	191
326	Erweitertes autoregressives Modell	201
327	MA-Prozesse und ARMA-Prozesse	203
33	Schätzungen von Momentfunktionen	209
331	Ergodizität	209
332	Schätzung des Erwartungswertes	215
333	Schätzung der Autokovarianzfunktion	216
34	Schätzungen des Spektrums	220
341	Periodogramm	220
342	Geglättete Schätzung	222
343	Numerische Berechnung der geglätteten Schätzung	229
35	Schätzungen in speziellen Modellen	232
351	Kontinuierliche Optimalfilter	232
352	Diskrete Optimalfilter	235

353	Modell der Prädiktion und Filterung	239
354	Autoregressive Prozesse	241
355	Erweitertes autoregressives Modell	245
36	Zufallsfelder	250
361	Definitionen	250
362	Momentfunktionen, Homogenität und Isotropie	251
363	Spektrum	253
37	Spezielle Zufallsfelder	255
371	Weißes Rauschen und Gauß-Zufallsfelder	255
372	Markoff-Zufallsfelder	256
373	Nachbarschafts-Gibbs-Felder	258
374	Äquivalenz von Markoff-Zufallsfeldern und Nachbarschafts-Gibbs-Feldern	261
375	Markoff-Zufallsfelder mit Gibbs-Verteilungen	263
376	Markoff-Zufallsfelder auf Gittern	270
377	Kausale zweidimensionale autoregressive Prozesse	279
378	Nichtkausale zweidimensionale autoregressive Prozesse	289
379	Anwendungen in der digitalen Bildverarbeitung	299
38	Schätzungen für zweidimensionale Zufallsfelder	316
381	Schätzungen von Momentfunktionen	316
382	Schätzungen des Spektrums	317
383	Parameterschätzungen für Markoff-Zufallsfelder der digitalen Bildverarbeitung	320
	Literatur	331
	Sachverzeichnis	339

1 Einleitung

Natürliche Phänomene, die in Abhängigkeit einer oder mehrerer Veränderlicher mathematisch beschrieben oder meßtechnisch erfaßt werden, bezeichnet man als *Signale*. In der Geodäsie werden beispielsweise Signale bei der elektrooptischen Entfernungsmessung erzeugt, indem ein Trägersignal mit vorgegebener Frequenz generiert wird, das durch ein weiteres Signal mit ebenfalls gegebener Frequenz moduliert wird. In der Photogrammetrie benutzt man zum Beispiel digitale Kameras, um Abbilder der Umwelt zu registrieren. Man erhält Grauwerte von Bildelementen in der Abbildungsebene. Die Grauwerte sind von der Position in der Bildebene abhängig, so daß sie ein zweidimensionales Signal charakterisieren.

Sind Signale vorhersagbar oder determinierbar, spricht man von *deterministischen Signalen*. Dagegen liegen *stochastischen Signalen* Phänomene zugrunde, die sich zufällig ändern und sich nicht exakt beschreiben oder vorhersagen lassen. Das Trägersignal mit dem aufgeprägten Modulationssignal in dem eingangs erwähnten Beispiel der elektrooptischen Streckenmessung läßt sich mathematisch vollständig beschreiben, es stellt ein deterministisches Signal dar. Bei der physikalischen Realisierung des Signals überlagern sich aber elektronische Störanteile, die nicht exakt vorhersagbar sind und daher ein stochastisches Signal bilden.

Im Kapitel 2 werden die deterministischen Signale, im Kapitel 3 die stochastischen Signale behandelt. Deterministische Signale werden durch harmonische Schwingungen dargestellt, deren Summe auf die Fourier-Reihe und das Fourier-Integral führt. Mit der aus dem Fourier-Integral gewonnenen Fourier-Transformation erhält man das Spektrum des Signals, das die komplexen Amplituden des Signals in Abhängigkeit von den Frequenzen angibt. Durch eine Filterung gilt es nun, das Spektrum des Signals derart zu verändern, daß es gewünschte Eigenschaften besitzt. Beispielsweise sollen aus einem Signal Amplituden hoher Frequenzen, die durch Störungen verursacht wurden, herausgefiltert werden. Sowohl eindimensionale als auch mehrdimensionale Signale werden behandelt. Der Entwurf von Filtern bleibt allerdings auf den ein- und zweidimensionalen Fall beschränkt.

Im Gegensatz zu den deterministischen Signalen, bei denen harmonische Schwingungen genügen, sind die stochastischen Signale nur mit größerem Aufwand zu beschreiben. Man benutzt die Wahrscheinlichkeitsdichte des Signals, die aber nicht immer vorliegt, so daß man sich auf die Momentfunktionen, und zwar die Erwartungswertfunktion und die

Autokovarianzfunktion beschränkt. Die Fourier-Transformation der Autokovarianzfunktion führt auf das Spektrum, das den Frequenzinhalt des stochastischen Signals repräsentiert. Bei der Filterung stochastischer Signale wird das Spektrum derart verändert, daß es gewünschte Eigenschaften besitzt. Beim Optimalfilter beispielsweise werden Rauschanteile im Signal unterdrückt.

Für spezielle eindimensionale stochastische Signale und für spezielle mehrdimensionale Zufallsfelder lassen sich die Wahrscheinlichkeitsdichten angeben. Für Zufallsfelder mit Markoff-Eigenschaft folgen die Gibbs-Verteilungen, mit denen Wahrscheinlichkeitsdichten besonders einfach auszudrücken sind. Die Verteilungen sind abhängig von Parametern, die in der Regel unbekannt sind, so daß sie aus einer Realisierung eines Zufallsfeldes zu schätzen sind. Mit diesen Schätzwerten lassen sich dann die Zufallsfelder beschreiben und beispielsweise Texturen, die in digitalen Bildern enthalten sind, erkennen. Auf dieses Beispiel wird ausführlich eingegangen.

Zur besseren Übersicht werden im Text die Beispiele mit dem Zeichen Δ und die Beweise von Sätzen mit dem Symbol \square abgeschlossen.

2 Deterministische Signale

Als deterministisches Signal wird eine Funktion verstanden, die sich mathematisch vollständig beschreiben läßt. Der Wert des Signals ergibt sich aus einem bekannten Bildungsgesetz in Abhängigkeit von einer oder mehreren Variablen, dem sogenannten Index oder Indexvektor. Dementsprechend unterscheidet man zwischen *ein-* und *mehrdimensionalen Signalen*. Besteht das Signal nicht nur aus einem, sondern aus mehreren Werten, liegt ein ein- oder mehrdimensionales *vektorielles Signal* vor. Besondere Bedeutung kommt den periodischen Signalen zu, da sie als Spezialfall auch die nicht periodischen oder aperiodischen Signale enthalten. Ursprünglich liegen die deterministischen Signale in kontinuierlicher oder analoger Form vor. Die numerische Analyse erfordert jedoch diskrete oder digitale Signale, die aus der Digitalisierung der kontinuierlichen Signale gewonnen werden.

Die folgenden Kapitel behandeln zunächst die eindimensionalen Signale und Systeme sowie die eindimensionale Filterung. Anschließend werden die Zusammenhänge auf den mehrdimensionalen Fall verallgemeinert, wobei die zweidimensionalen Signale und Systeme näher betrachtet werden. Zuletzt wird die zweidimensionale Filterung vorgestellt.

21 Eindimensionale Signale und lineare Systeme

211 Signaldarstellung

Ein eindimensionales *kontinuierliches* oder *analoges Signal* wird durch

$$x(t) \quad (211.1)$$

beschrieben, wobei

$$t \text{ mit } t \in \mathbb{R} \quad (211.2)$$

der Index bedeutet, der Element der überabzählbaren Indexmenge $T = \mathbb{R}$ ist. Der Wert des Signals $x(t)$ ist also abhängig vom Index t . Dagegen wird ein eindimensionales *diskretes* oder *digitales Signal* durch

$$x(n) \quad (211.3)$$

dargestellt. In diesem Fall bezeichnet

$$n \text{ mit } n \in \mathbb{Z} = \{0, \pm 1, \pm 2, \dots\} \text{ oder } n \in \mathbb{N}_0 = \{0, 1, 2, \dots\} \quad (211.4)$$

den Index, der Element der abzählbaren Indexmengen $T = \mathbb{Z}$ oder $T = \mathbb{N}_0$ ist. Abbildung 211-1 zeigt ein Beispiel für ein diskretes eindimensionales Signal.

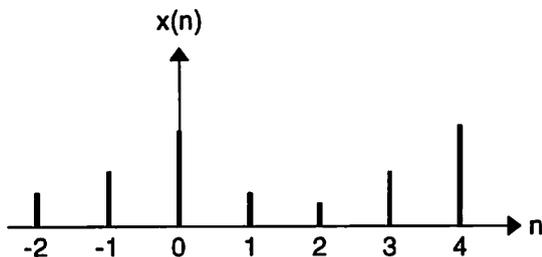


Abb. 211-1: Diskretes Signal $x(n)$

Die Indizes t und n in (211.2) und (211.4) können als Maßeinheiten beispielsweise für die Zeit oder für die Entfernung aufgefaßt werden. Im folgenden werden t und n mit der Zeit identifiziert.

Die *Deltafunktion* $\delta(t)$ ist definiert durch

$$\int_{-\infty}^{\infty} x(t) \delta(t - \tau) dt = x(\tau) \quad (211.5)$$

Sie bewirkt die Ausblendung des Signalwertes $x(\tau)$ aus dem kontinuierlichen Signal $x(t)$. Analog zu (211.5) erhält man den diskreten Signalwert $x(k)$ aus

$$\sum_{n=-\infty}^{\infty} x(n)d(n-k) = x(k) \quad , \quad (211.6)$$

worin $d(n)$ der *diskrete Einheitsimpuls* bedeutet. Er ist definiert durch

$$d(n) = \begin{cases} 1 & \text{für } n = 0 \\ 0 & \text{für } n \neq 0 \end{cases} \quad (211.7)$$

und in Abbildung 211-2 dargestellt.

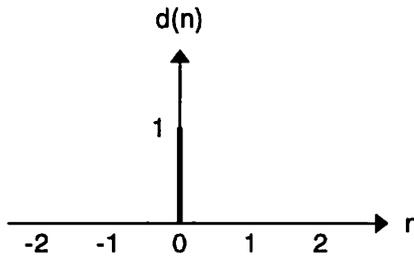


Abb. 211-2: Diskreter Einheitsimpuls

212 Harmonische Schwingung

Kontinuierliche Signale in Abhängigkeit von der Zeit werden häufig durch eine Summe harmonischer Schwingungen beschrieben, so daß auf ihre Darstellung zunächst eingegangen werden soll.

Eine *harmonische Schwingung* $x(t)$ als Funktion der Zeit t ist in reeller Form durch

$$x(t) = C \cos(\omega_0 t - \varphi) \quad (212.1)$$

definiert, worin C die *Amplitude*, ω_0 die *Kreisfrequenz*, $\omega_0 t$ die *Phase* und φ der *Phasenwinkel* bedeuten. Die Kreisfrequenz ω_0 wird mit

$$\omega_0 = 2\pi/T \quad (212.2)$$

durch die *Periode* T der Schwingung dargestellt, für die allgemein

$$x(t+kT) = x(t) \quad (212.3)$$

mit $k \in \mathbb{Z}$ gilt. Aus (212.1) folgt also mit (212.2)

$$C \cos(\omega_0(t+kT) - \varphi) = C \cos(\omega_0 t + 2\pi k - \varphi) = C \cos(\omega_0 t - \varphi) \quad .$$

Wie in Abbildung 212-1 dargestellt, läßt sich die Periode T auch als Zeitdifferenz zwischen dem ersten Maximum der Schwingung auf der positiven t -Achse an der Stelle $t_0 = \varphi/\omega_0$ und dem folgenden Maximum an der Stelle $t_0 + T$ interpretieren. Für die *Frequenz*

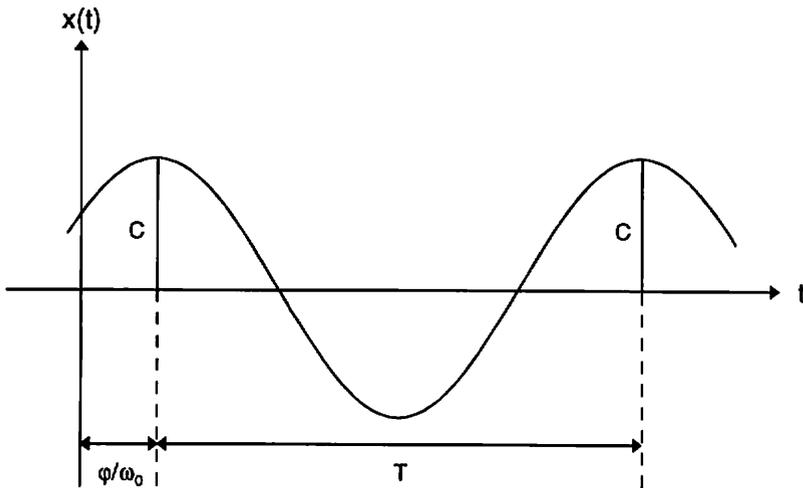


Abb. 212-1: Harmonische Schwingung

f_0 der Schwingung gilt

$$f_0 = 1/T = \omega_0/2\pi . \quad (212.4)$$

Mit dem Additionstheorem der \cos -Funktion ergibt sich aus (212.1)

$$x(t) = C \cos\omega_0 t \cos\varphi + C \sin\omega_0 t \sin\varphi .$$

Führt man die Konstanten A und B durch

$$A = C \cos\varphi$$

$$B = C \sin\varphi \quad (212.5)$$

ein, erhält man die harmonische Schwingung zu

$$x(t) = A \cos\omega_0 t + B \sin\omega_0 t . \quad (212.6)$$

Sind andererseits A und B bekannt, folgt aus (212.5)

$$C = (A^2+B^2)^{1/2}$$

und

$$\varphi = \arctan(B/A) . \quad (212.7)$$

Die beiden Darstellungen (212.1) und (212.6) der harmonischen Schwingung sind also

äquivalent und können durch Umrechnung der Konstanten ineinander überführt werden.

Schließlich soll die harmonische Schwingung (212.1) durch komplexe Größen dargestellt werden. Aus der Eulerschen Formel

$$e^{j\alpha} = \cos\alpha + j\sin\alpha \quad \text{mit} \quad j = (-1)^{1/2}, \quad (212.8)$$

erhält man für $\alpha = 2n\pi$ und $\alpha = (2n+1)\pi$

$$e^{j2n\pi} = 1 \quad \text{und} \quad e^{j(2n+1)\pi} = -1 \quad \text{für} \quad n \in \{0, \pm 1, \pm 2, \dots\}. \quad (212.9)$$

Weiterhin ergeben sich wegen

$$\cos(-\alpha) = \cos\alpha \quad \text{und} \quad \sin(-\alpha) = -\sin\alpha \quad (212.10)$$

aus (212.8) die Zusammenhänge

$$\cos\alpha = \frac{1}{2} (e^{j\alpha} + e^{-j\alpha}) \quad \text{und} \quad \sin\alpha = \frac{1}{2j} (e^{j\alpha} - e^{-j\alpha}). \quad (212.11)$$

Führt man das Ergebnis für $\cos\alpha$ in (212.1) ein, folgt

$$x(t) = \frac{C}{2} e^{-j\varphi} e^{j\omega_0 t} + \frac{C}{2} e^{j\varphi} e^{-j\omega_0 t}. \quad (212.12)$$

Mit den zueinander konjugiert komplexen Koeffizienten D und D^*

$$D = \frac{C}{2} e^{-j\varphi} \quad \text{und} \quad D^* = \frac{C}{2} e^{j\varphi}, \quad (212.13)$$

für die mit (212.5) und (212.8) gilt

$$D = \frac{1}{2} (A - jB) \quad \text{und} \quad D^* = \frac{1}{2} (A + jB), \quad (212.14)$$

läßt sich für (212.12) schreiben

$$x(t) = D e^{j\omega_0 t} + D^* e^{-j\omega_0 t}. \quad (212.15)$$

Damit ist neben (212.1) und (212.6) die dritte Darstellung der harmonischen Schwingung gefunden.

Eine noch kompaktere Form erhält man mit dem komplexen Signal $x_k(t)$

$$x_k(t) = C e^{-j\varphi} e^{j\omega_0 t} = 2D e^{j\omega_0 t}, \quad (212.16)$$

worin $e^{j\omega_0 t}$ als komplexe Schwingung mit der Phase $\omega_0 t$ und $2D$ als komplexe Amplitude interpretierbar sind. Die reelle Zeitfunktion $x(t)$ ergibt sich aus (212.12) zu

$$x(t) = \frac{1}{2} (x_k(t) + x_k^*(t)) = \text{Re}(x_k(t)), \quad (212.17)$$

worin $x_k^*(t)$ wieder das konjugiert komplexe Signal und Re den Realteil der komplexen Funktion bedeuten.

213 Fourier-Reihe

Eine periodische Funktion läßt sich durch eine Summe harmonischer Schwingungen unterschiedlicher Amplituden und Frequenzen annähern. Bei der Wahl der Frequenzen geht man in der Weise vor, daß die sogenannte *Grundfrequenz* festgelegt wird, die sich nach (212.2) aus der Periode der anzunähernden Funktion bestimmt. Die höheren Frequenzen sind ganzzahlige Vielfache der Grundfrequenz. Die Amplituden und Phasen der harmonischen Schwingungen werden durch eine Anpassung an die periodische Funktion nach der Methode der kleinsten Quadrate, also durch eine Parameterschätzung bestimmt.

Es sei $x_p(t)$ der Wert eines periodischen Signals an der Stelle t . Für die Periode T gelte mit (212.2)

$$T = 2\pi/\omega_0, \quad (213.1)$$

wobei ω_0 die Grundfrequenz bezeichnet. Entsprechend (212.6) wird dann $x_p(t)$ dargestellt durch

$$x_p(t) = \frac{A_0}{2} + \sum_{k=1}^K (A_k \cos k\omega_0 t + B_k \sin k\omega_0 t), \quad (213.2)$$

worin k mit $k \leq K$ die Ordnungszahl der sogenannten *Harmonischen* bedeutet.

An N Stellen t_n , die gleichmäßig über die Periode T aus (213.1) verteilt sind, sei die periodische Funktion $x_p(t)$ gegeben. Dabei wird das Intervall zwischen $-\pi/\omega_0 = -T/2$ und $\pi/\omega_0 = T/2$ gewählt, so daß folgt

$$t_n = -\frac{\pi}{\omega_0} + \frac{2\pi}{\omega_0} \frac{n}{N} = -\frac{T}{2} + T \frac{n}{N} \quad \text{für } n \in \{0, 1, \dots, N-1\}. \quad (213.3)$$

Addiert man auf der linken Seite von (213.2) den Fehler e_n der Anpassung, erhält man die Beobachtungsgleichung für die Parameterschätzung

$$x_p(t_n) + e_n = \frac{A_0}{2} + \sum_{k=1}^K (A_k \cos k\omega_0 t_n + B_k \sin k\omega_0 t_n), \quad (213.4)$$

in der die Koeffizienten A_0 , A_k und B_k für $k \in \{1, \dots, K\}$ die unbekannt Parameter bezeichnen. Aus N Beobachtungen sind also $2K+1$ unbekannt Parameter zu schätzen, so daß

$$N \geq 2K+1 \quad (213.5)$$

gelten muß.

Die Koeffizientenmatrix X des Modells für die Parameterschätzung (Koch 1987, S.178) folgt mit

$$X = \begin{bmatrix} 1/2 & \cos\omega_0 t_0 & \sin\omega_0 t_0 & \dots & \cos K\omega_0 t_0 & \sin K\omega_0 t_0 \\ 1/2 & \cos\omega_0 t_1 & \sin\omega_0 t_1 & \dots & \cos K\omega_0 t_1 & \sin K\omega_0 t_1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1/2 & \cos\omega_0 t_{N-1} & \sin\omega_0 t_{N-1} & \dots & \cos K\omega_0 t_{N-1} & \sin K\omega_0 t_{N-1} \end{bmatrix}$$

und hieraus die Normalgleichungsmatrix $X^T X$ (Koch 1987, S.186). Die Sinus- und Cosinus-Funktionen sind zueinander orthogonal wegen

$$\sum_{n=0}^{N-1} \cos k\omega_0 t_n \cos m\omega_0 t_n = \begin{cases} 0 & \text{für } k \neq m \\ N/2 & \text{für } k = m \text{ und } k > 0 \end{cases}$$

sowie der entsprechenden Gleichung für die Sinus-Funktion und

$$\sum_{n=0}^{N-1} \sin k\omega_0 t_n \cos m\omega_0 t_n = 0 .$$

Um die erste Beziehung für $k \neq m$ zu zeigen, wird mit (212.11) umgeformt

$$\sum_{n=0}^{N-1} \cos k\omega_0 t_n \cos m\omega_0 t_n = \frac{1}{4} \sum_{n=0}^{N-1} (e^{jk\omega_0 t_n} + e^{-jk\omega_0 t_n}) (e^{jm\omega_0 t_n} + e^{-jm\omega_0 t_n}) .$$

Nach Multiplikation auf der rechten Seite folgt der erste Summand mit (212.9) und (213.3) zu

$$\frac{1}{4} \sum_{n=0}^{N-1} e^{j(k+m)\omega_0 t_n} = \frac{1}{4} e^{-j(k+m)\pi} \sum_{n=0}^{N-1} e^{2j(k+m)n\pi/N} .$$

Mit der Summe einer geometrischen Reihe

$$\sum_{n=0}^{N-1} (x^k)^n = \frac{1-x^{kN}}{1-x^k} \quad (213.6)$$

erhält man weiter

$$\sum_{n=0}^{N-1} e^{2j(k+m)n\pi/N} = \frac{1-e^{2j(k+m)\pi}}{1-e^{2j(k+m)\pi/N}} ,$$

so daß wegen (212.9) der erste Summand sich zu Null ergibt. Die übrigen Summanden besitzen ebenfalls den Wert Null, und es folgt

$$\sum_{n=0}^{N-1} \cos k \omega_0 t_n \cos m \omega_0 t_n = 0 \quad \text{für } k \neq m .$$

Um die zweite Beziehung zu zeigen, wird wegen $\cos \alpha \cos \beta = \frac{1}{2} [\cos(\alpha - \beta) + \cos(\alpha + \beta)]$ mit (212.9), (212.11), (213.3) und (213.6) gebildet

$$\begin{aligned} \sum_{n=0}^{N-1} \cos k \omega_0 t_n \cos k \omega_0 t_n &= \frac{1}{2} \sum_{n=0}^{N-1} (1 + \cos 2k \omega_0 t_n) \\ &= \frac{N}{2} + \frac{1}{4} \sum_{n=0}^{N-1} (e^{2jk \omega_0 t_n} + e^{-2jk \omega_0 t_n}) \\ &= \frac{N}{2} + \frac{1}{4} \sum_{n=0}^{N-1} (e^{4jk n \pi / N} e^{-2jk \pi} + e^{-4jk n \pi / N} e^{2jk \pi}) \\ &= \frac{N}{2} . \end{aligned}$$

Die restlichen Orthogonalitätsbeziehungen lassen sich entsprechend beweisen. Weiter folgt mit ähnlichen Ableitungen

$$\sum_{n=0}^{N-1} \cos k \omega_0 t_n = 0$$

und die entsprechende Gleichung für die Sinus-Funktion.

Für die Normalgleichungsmatrix $\mathbf{X}'\mathbf{X}$ erhält man daher die Diagonalmatrix

$$\mathbf{X}'\mathbf{X} = (N/2) \text{diag}(1/2, 1, 1, \dots, 1) .$$

Mit ihrer Inversen ergeben sich dann die Schätzwerte der Parameter zu

$$\begin{aligned} A_k &= \frac{2}{N} \sum_{n=0}^{N-1} x_p(t_n) \cos k \omega_0 t_n \quad \text{für } k \in \{0, 1, \dots, K\} \\ B_k &= \frac{2}{N} \sum_{n=0}^{N-1} x_p(t_n) \sin k \omega_0 t_n \quad \text{für } k \in \{1, 2, \dots, K\} . \end{aligned} \quad (213.7)$$

Mit den aus (213.7) bestimmten Koeffizienten A_k und B_k läßt sich dann $x_p(t)$ an einer beliebigen Stelle t mit (213.2) darstellen.

Ist das periodische Signal nicht nur an den diskreten Stellen $x_p(t_n)$ gegeben, sondern liegt ein kontinuierliches Signal $x_p(t)$ vor, dann ist die Periode T im Intervall $-\pi/\omega_0$ und π/ω_0 in sehr viele kleine Zeitabschnitte dt einzuteilen, so daß $T = Ndt$ gilt. Aus (213.1) folgt dann

$$N = \frac{T}{\Delta t} = \frac{2\pi}{\omega_0 \Delta t} .$$

Substituiert man dieses Ergebnis in (213.7), ersetzt die Summationen durch Integrationen und führt in (213.2) für eine möglichst gute Anpassung noch mit $N \rightarrow \infty$ eine unendliche Summe ein, so folgt die *Fourier-Reihe*

$$x_p(t) = \frac{A_0}{2} + \sum_{k=1}^{\infty} (A_k \cos k\omega_0 t + B_k \sin k\omega_0 t) \quad (213.8)$$

mit

$$A_k = \frac{\omega_0}{\pi} \int_{-\pi/\omega_0}^{\pi/\omega_0} x_p(t) \cos(k\omega_0 t) dt \quad \text{für } k \in \{0, 1, \dots, \infty\}$$

$$B_k = \frac{\omega_0}{\pi} \int_{-\pi/\omega_0}^{\pi/\omega_0} x_p(t) \sin(k\omega_0 t) dt \quad \text{für } k \in \{1, 2, \dots, \infty\} . \quad (213.9)$$

Mit den komplexen Koeffizienten D_k läßt sich wegen (212.15) mit der Festlegung $B_0 = 0$ die Fourier-Reihe kompakter schreiben durch

$$x_p(t) = \sum_{k=-\infty}^{\infty} D_k e^{jk\omega_0 t} \quad (213.10)$$

mit

$$D_{-k} = D_k^* , \quad (213.11)$$

wobei D_k^* konjugiert komplex zu D_k ist, so daß aus (212.14) folgt

$$D_k = \frac{1}{2} (A_k - jB_k) \quad \text{und} \quad D_{-k} = \frac{1}{2} (A_k + jB_k) . \quad (213.12)$$

In der Fourier-Reihe (213.10) ist also dem Vielfachen $k\omega_0$ der Grundfrequenz ω_0 die komplexe Amplitude D_k zugeordnet. Für sie folgt nach Substitution von (213.9) mit (212.8)

$$D_k = \frac{\omega_0}{2\pi} \int_{-\pi/\omega_0}^{\pi/\omega_0} x_p(t) e^{-jk\omega_0 t} dt . \quad (213.13)$$

Aus der Darstellung der komplexen Amplitude D_k durch ihren Realteil Re und ihren Imaginärteil Im

$$D_k = \text{Re}(D_k) + j \text{Im}(D_k)$$

$$= |D_k| e^{j\varphi_k} \quad (213.14)$$

mit

$$|D_k| = (\operatorname{Re}(D_k)^2 + \operatorname{Im}(D_k)^2)^{1/2} = (D_k D_k^*)^{1/2} ,$$

$$\varphi_k = \arctan(\operatorname{Im}(D_k) / \operatorname{Re}(D_k))$$

folgt wegen (213.11)

$$|D_k| = |D_{-k}| . \quad (213.15)$$

Die reellen Amplituden $|D_k|$ sind also symmetrisch bezüglich $k = 0$. Ihre Darstellung in Abhängigkeit von den Vielfachen $k\omega_0$ der Grundfrequenz ω_0 bezeichnet man als *diskretes Amplitudenspektrum*. Es veranschaulicht den Frequenzinhalt des Signals, denn die Frequenzen $k\omega_0$, für die die Amplituden $|D_k|$ große Werte annehmen, sind am stärksten in $x_p(t)$ vertreten. Die Abbildung 213-1 zeigt ein Beispiel. Die Darstellung der Phase φ_k in Abhängigkeit von $k\omega_0$ wird *diskretes Phasenspektrum* genannt. Zusammengefaßt heißt D_k auch *komplexes diskretes Amplitudenspektrum* oder *diskretes Spektrum*.

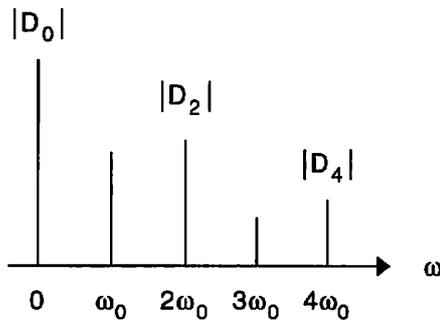


Abb. 213-1: Diskretes Amplitudenspektrum

Ist das Signal $x_p(t)$ eine *gerade Funktion*, gilt also

$$x_p(t) = x_p(-t) , \quad (213.16)$$

dann müssen die Sinus-Terme in der Fourier-Reihe (213.8) verschwinden, so daß nur die Cosinus-Glieder verbleiben. Liegt andererseits eine *ungerade Funktion* mit

$$x_p(t) = -x_p(-t) \quad (213.17)$$

vor, dann enthält die Fourier-Reihe nur die Sinus-Terme.

214 Fourier-Transformation

Wie im vorangegangenen Abschnitt gezeigt, lassen sich periodische Signale als Summe harmonischer Schwingungen darstellen. Liegen aperiodische Signale vor, muß die

Fourier-Reihe zum Fourier-Integral verallgemeinert werden.

Das periodische Signal $x_p(t)$ sei mittels der komplexen Amplituden D_k nach (213.10) und (213.13) dargestellt durch

$$x_p(t) = \sum_{k=-\infty}^{\infty} D_k e^{jk\omega_0 t} \quad \text{mit} \quad D_k = \frac{\omega_0}{2\pi} \int_{-\pi/\omega_0}^{\pi/\omega_0} x_p(t) e^{-jk\omega_0 t} dt . \quad (214.1)$$

Die Periode $T = 2\pi/\omega_0$ aus (213.1) des Signals soll jetzt bis ins Unendliche ausgedehnt werden, so daß die periodischen Wiederholungen des Signals im Unendlichen verschwinden und aus dem periodischen Signal ein aperiodisches entsteht. Dies wird erreicht, wenn ω_0 gegen Null geht. Ersetzt man ω_0 durch $\Delta\omega$ und die Integrationsvariable t durch τ , erhält man anstelle von (214.1)

$$x_p(t) = \sum_{k=-\infty}^{\infty} \left(\frac{\Delta\omega}{2\pi} \int_{-\pi/\Delta\omega}^{\pi/\Delta\omega} x_p(\tau) e^{-jk\Delta\omega\tau} d\tau \right) e^{jk\Delta\omega t} .$$

Man kann nun den Grenzübergang $\Delta\omega$ gegen Null vornehmen, wobei zu beachten ist, daß $k\Delta\omega$ gleich ω wird. Man erhält

$$x_p(t) = \frac{1}{2\pi} \lim_{\Delta\omega \rightarrow 0} \sum_{k=-\infty}^{\infty} \left(\int_{-\pi/\Delta\omega}^{\pi/\Delta\omega} x_p(\tau) e^{-jk\Delta\omega\tau} d\tau \right) e^{jk\Delta\omega t} \Delta\omega .$$

Die Periode wird jetzt unendlich groß, so daß aus dem periodischen Signal $x_p(t)$ das aperiodische Signal $x(t)$ entsteht, folglich

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x(\tau) e^{-j\omega\tau} d\tau \right) e^{j\omega t} d\omega$$

oder

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(j\omega) e^{j\omega t} d\omega \quad \text{mit} \quad X(j\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt . \quad (214.2)$$

Dies sind die beiden *Fourier-Integrale*, die als *Fourier-Transformation* bezeichnet werden. Das erste Integral wird auch *inverse Fourier-Transformation* genannt. Abgekürzt schreibt man die Fourier-Transformation häufig wie folgt

$$x(t) \leftrightarrow X(j\omega) . \quad (214.3)$$

Analog zu (213.14) kann die komplexe Größe $X(j\omega)$ in der Form

$$X(j\omega) = |X(j\omega)| e^{j\varphi(\omega)}$$

dargestellt werden. Die reellen Amplituden $|D_k|$ der Fourier-Reihe (213.10) liefern ein diskretes Amplitudenspektrum. Mit $|X(j\omega)|$ und $\varphi(\omega)$ erhält man nun ein *kontinuierli-*

ches Amplitudenspektrum und ein kontinuierliches Phasenspektrum. Das kontinuierliche Amplitudenspektrum ist für ein reelles Signal $x(t)$ bezüglich $\omega = 0$ symmetrisch. Es folgt nämlich zunächst aus (214.2) wegen (212.13)

$$X^*(j\omega) = X(-j\omega) .$$

Weiter gilt nach (213.14) $|X^*(j\omega)| = |X(j\omega)|$ und somit

$$|X(j\omega)| = |X(-j\omega)| .$$

Abbildung 214-1 zeigt ein Beispiel für ein kontinuierliches Amplitudenspektrum. Zusammengefaßt wird $X(j\omega)$ auch als *komplexes kontinuierliches Amplitudenspektrum* oder als *Spektrum* bezeichnet.

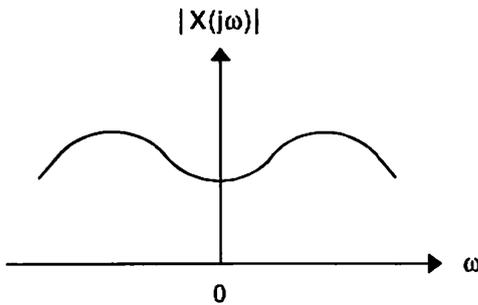


Abb. 214-1: Kontinuierliches Amplitudenspektrum

Bei der Ableitung der Fourier-Transformation waren zeitbegrenzte Signale vorausgesetzt worden, deren Wiederholungen sich ins Unendliche schieben lassen. Das Fourier-Integral wird daher für Signale, die sich bis ins Unendliche erstrecken, nicht konvergieren.

Beispiel 1: Für das konstante Einheitssignal

$$x(t) = 1$$

soll das Spektrum nach (214.2) berechnet werden. Allgemein gilt mit (212.11)

$$\int_{-a}^a e^{-j\omega t} dt = -\frac{1}{j\omega} e^{-j\omega t} \Big|_{-a}^a = \frac{2}{2j\omega} (e^{j\omega a} - e^{-j\omega a}) = \frac{2}{\omega} \sin \omega a . \quad (214.4)$$

Für das Spektrum des konstanten Einheitssignals folgt daher mit

$$X(j\omega) = \int_{-\infty}^{\infty} e^{-j\omega t} dt = \frac{2}{\omega} \sin \infty$$

ein unbestimmter Ausdruck.

▲

Beispiel 2: Die Fourier-Transformation für das Rechtecksignal

$$x(t) = \begin{cases} 1 & \text{für } |t| \leq c \\ 0 & \text{für } |t| > c \end{cases}$$

läßt sich mit (214.4) angeben durch

$$X(j\omega) = \int_{-c}^c e^{-j\omega t} dt = \frac{2}{\omega} \sin \omega c . \quad (214.5)$$

Für das zeitlich begrenzte Rechtecksignal existiert also das Fourier-Integral. Δ

Ist $x(t)$ absolut integrierbar im Sinne von

$$\int_{-\infty}^{\infty} |x(t)| dt < \infty , \quad (214.6)$$

dann existiert das Fourier-Integral $X(j\omega)$ und läßt sich mit (214.2) nach $x(t)$ transformieren (Papoulis 1962, S.9). Es gibt allerdings Funktionen, die nicht absolut integrierbar sind und deren Fourier-Integrale dennoch existieren.

Zerlegt man den Integranden des Fourier-Integrals $X(j\omega)$ in (214.2) mit (212.8) und (212.10) in den Real- und Imaginärteil und spaltet das Integral durch Substitution $t = -\tau$ und anschließender Umbezeichnung auf, erhält man

$$\begin{aligned} X(j\omega) &= \int_{-\infty}^{\infty} x(t) \cos \omega t dt - j \int_{-\infty}^{\infty} x(t) \sin \omega t dt \\ &= \int_0^{\infty} [x(t) + x(-t)] \cos \omega t dt - j \int_{-\infty}^0 [x(t) - x(-t)] \sin \omega t dt . \end{aligned} \quad (214.7)$$

Hieraus erkennt man, daß $X(j\omega)$ reell ist, wenn $x(t)$ gerade ist, wenn also (213.16) gilt. Andererseits ist $X(j\omega)$ imaginär, wenn $x(t)$ ungerade ist, wenn also (213.17) gilt. Der Realteil von $X(j\omega)$ ist gerade und der Imaginärteil ungerade.

Häufig benötigte Fourier-Transformationen befinden sich zum Beispiel in (Stearns 1984, S.48; Papoulis 1977, S.61).

Im folgenden soll noch auf die Faltung eingegangen werden. Mit den Funktionen $x_1(t)$ und $x_2(t)$ wird das Integral definiert

$$x_1(t) * x_2(t) = \int_{-\infty}^{\infty} x_1(\tau) x_2(t - \tau) d\tau , \quad (214.8)$$

das als *Faltung* oder *Konvolution* bezeichnet wird. Das Zeichen $*$ auf der linken Seite

symbolisiert die Faltungsoperation. Sind nun

$$x_1(t) \leftrightarrow X_1(j\omega) \quad \text{und} \quad x_2(t) \leftrightarrow X_2(j\omega) \quad (214.9)$$

zwei Fourier-Transformationspaare, so ist auch

$$x_1(t) * x_2(t) \leftrightarrow X_1(j\omega)X_2(j\omega) \quad (214.10)$$

ein Transformationspaar. Der Faltung (214.8) entspricht also die Multiplikation (214.10) der Fourier-Transformationen. Um diese Beziehung zu beweisen, wird nach (214.2) die Fourier-Transformation $X(j\omega)$ der Faltung gebildet

$$X(j\omega) = \int_{-\infty}^{\infty} x_1(t) * x_2(t) e^{-j\omega t} dt = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1(\tau) x_2(t-\tau) d\tau e^{-j\omega t} dt .$$

Durch die Variablentransformation $u = t - \tau$ mit $du = dt$ und durch Vertauschung der Integrationsfolge erhält man

$$\begin{aligned} X(j\omega) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1(\tau) x_2(u) d\tau e^{-j\omega(u+\tau)} du \\ &= \int_{-\infty}^{\infty} x_1(\tau) e^{-j\omega\tau} \left(\int_{-\infty}^{\infty} x_2(u) e^{-j\omega u} du \right) d\tau \\ &= X_1(j\omega) X_2(j\omega) \end{aligned}$$

und damit die gewünschte Beziehung. Um aber die Integrationsfolgen vertauschen zu können, müssen die Funktionen $x_1(t)$ und $x_2(t)$ quadratisch integrierbar sein (Papoulis 1962, S.27). Es muß also gelten

$$\int_{-\infty}^{\infty} |x_i(t)|^2 dt < \infty \quad \text{für} \quad i \in \{1, 2\} . \quad (214.11)$$

Beispiel 3: Das Dreieckssignal

$$x(t) = \begin{cases} 1 - \frac{|t|}{c} & \text{für} \quad |t| \leq c \\ 0 & \text{für} \quad |t| > c \end{cases} \quad (214.12)$$

läßt sich durch die Faltung zweier Rechtecksignale der Länge $c/2$

$$x_k(t) = \begin{cases} 1 & \text{für} \quad |t| \leq c/2 \\ 0 & \text{für} \quad |t| > c/2 \end{cases} \quad \text{mit} \quad k \in \{1, 2\}$$

erzeugen. Aus (214.8) folgt nämlich

$$\begin{aligned} \frac{1}{c} \int_{-\infty}^{\infty} x_1(\tau) x_2(t-\tau) d\tau &= \frac{1}{c} \int_{-c/2}^{c/2} x_2(t-\tau) d\tau \\ &= \frac{1}{c} \int_{-c/2+t}^{c/2} d\tau = 1 - \frac{t}{c} \quad \text{für } 0 < t < c \end{aligned} \quad (214.13)$$

und ein entsprechendes Ergebnis für $0 > t > c$, so daß das Dreiecksignal erhalten wird. Aus der Fourier-Transformation (214.5) des Rechtecksignals ergibt sich nun mit (214.10) und (214.13) die Fourier-Transformation $X(j\omega)$ des Dreiecksignals (214.12) zu

$$X(j\omega) = \frac{4}{\omega^2 c} \sin^2 \frac{\omega c}{2} . \quad (214.14)$$

Δ

Setzt man $\tau = 0$ in der Definition (211.5) der Deltafunktion $\delta(t)$, folgt

$$\int_{-\infty}^{\infty} \delta(t) x(t) dt = x(0) . \quad (214.15)$$

Hiermit erhält man wegen (212.9)

$$\int_{-\infty}^{\infty} \delta(t) e^{-j\omega t} dt = 1 . \quad (214.16)$$

Dies ist die Fourier-Transformation (214.2) der Deltafunktion, so daß sich

$$\delta(t) \leftrightarrow 1 \quad (214.17)$$

ergibt (Papoulis 1977, S.100). Die Transformation ist in Abbildung 214-2 dargestellt.

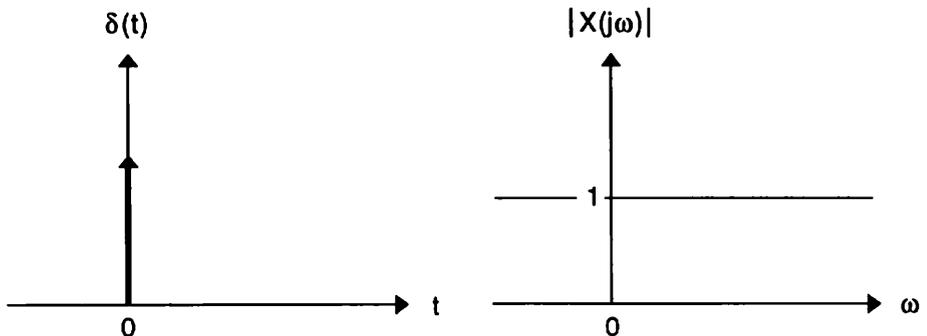


Abb. 214-2: Deltafunktion und ihre Fourier-Transformation

Eine Verallgemeinerung der Fourier-Transformation stellt die Laplace-Transformation dar, um auch die Transformation von Signalen zu erhalten, die mit wachsendem t nicht abnehmen. Existiert das Fourier-Integral aus (214.2)

$$X(j\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt$$

beispielsweise für eine Funktion $x(t)$ nicht, dann läßt sich $x(t)$ mit der abnehmenden Exponentialfunktion $e^{-\alpha|t|}$ für $\alpha > 0$ multiplizieren, so daß für das Produkt als Integrand das Integral existiert. Nimmt man weiter an, daß $x(t) = 0$ für $t < 0$ gilt, erhält man

$$X(\alpha + j\omega) = \int_0^{\infty} x(t)e^{-(\alpha + j\omega)t} dt .$$

Substituiert man schließlich die komplexe Variable s mit $s = \alpha + j\omega$, erhält man die *Laplace-Transformation*

$$X(s) = \int_0^{\infty} x(t)e^{-st} dt . \quad (214.18)$$

215 Abtastung kontinuierlicher Signale

Mit (213.10) wird ein kontinuierliches und periodisches Signal durch ein diskretes Spektrum dargestellt, eine diskrete Funktion also in eine kontinuierliche und periodische Funktion transformiert. Vertauscht man Signal und Spektrum, so bedeutet dies, daß ein beliebiges diskretes Signal in ein kontinuierliches periodisches Spektrum transformiert wird. Ein diskretes Signal $x(k)$ nach (211.3) werde durch *Abtastung* eines kontinuierlichen Signals $x(t)$ an den Stellen $t = k\Delta t$ mit

$$x(k) = x(t = k\Delta t)$$

erhalten, wobei Δt das Abtastintervall bezeichnet. Das komplexe kontinuierliche Amplitudenspektrum $X_p(j\omega)$, das von der Kreisfrequenz ω abhängt, erhält man aus (213.10) mit (213.13), indem t durch ω , ω_0 durch Δt ersetzt und das Vorzeichen der Exponentialfunktionen wie in (213.13) und (213.10) gewählt wird

$$X_p(j\omega) = \sum_{k=-\infty}^{\infty} x(k)e^{-jk\Delta t\omega} \quad (215.1)$$

mit

$$x(k) = \frac{\Delta t}{2\pi} \int_{-\pi/\Delta t}^{\pi/\Delta t} X_p(j\omega)e^{jk\Delta t\omega} d\omega . \quad (215.2)$$

Mit (212.9) gilt

$$e^{-jk\Delta t\omega} = e^{-jk\Delta t(\omega+2n\pi/\Delta t)} \quad \text{für } n \in \{0, \pm 1, \pm 2, \dots\},$$

so daß das Spektrum $X_p(j\omega)$ nach (212.3) periodisch mit der Periode $2\pi/\Delta t$ ist. Man bezeichnet $2\pi/\Delta t$ als *Abtastkreisfrequenz* oder auch als *Abtastfrequenz*.

Um festzustellen, wie das Spektrum $X_p(j\omega)$ des durch Abtastung gewonnenen Signals $x(k)$ sich zu dem Spektrum $X(j\omega)$ des abgetasteten kontinuierlichen Signals $x(t)$ verhält, wird $x(t)$ an den abgetasteten Stellen $t = k\Delta t$ mit $X(j\omega)$ nach (214.2) dargestellt durch

$$x(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(j\omega) e^{jk\Delta t\omega} d\omega. \quad (215.3)$$

Das Integral über die ω -Achse wird in eine unendliche Summe von Integralen über Intervalle der Länge $2\pi/\Delta t$ zerlegt, für die

$$(-\pi+2m\pi)/\Delta t < \omega < (\pi+2m\pi)/\Delta t \quad \text{mit } m \in \{0, \pm 1, \pm 2, \dots\}$$

gelte. Man erhält anstelle von (215.3)

$$x(k) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \int_{(-\pi+2m\pi)/\Delta t}^{(\pi+2m\pi)/\Delta t} X(j\omega) e^{jk\Delta t\omega} d\omega.$$

Mit der Transformation der Variablen $\omega = \Omega + 2m\pi/\Delta t$ folgt weiter

$$x(k) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \int_{-\pi/\Delta t}^{\pi/\Delta t} X(j(\Omega+2m\pi/\Delta t)) e^{jk\Delta t\Omega} e^{2jk m\pi} d\Omega.$$

Setzt man wieder $\Omega = \omega$, ergibt sich schließlich mit (212.9)

$$x(k) = \frac{\Delta t}{2\pi} \int_{-\pi/\Delta t}^{\pi/\Delta t} \frac{1}{\Delta t} \sum_{m=-\infty}^{\infty} X(j(\omega+2m\pi/\Delta t)) e^{jk\Delta t\omega} d\omega.$$

Der Vergleich mit (215.2) liefert das Ergebnis

$$X_p(j\omega) = \frac{1}{\Delta t} \sum_{m=-\infty}^{\infty} X(j(\omega+2m\pi/\Delta t)). \quad (215.4)$$

Aus dieser Beziehung ist ersichtlich, daß nach Berücksichtigung des Faktors $1/\Delta t$, der durch die Abtastung mit $t = k\Delta t$ hervorgerufen wird, das Spektrum $X_p(j\omega)$ des durch Abtastung erhaltenen diskreten Signals sich aus dem Spektrum $X(j\omega)$ des abgetasteten kontinuierlichen Signals ergibt, indem man $X(j\omega)$ um alle positiven und negativen Vielfache der Abtastfrequenz $2\pi/\Delta t$ verschiebt und aufaddiert. Dies ist in Abbildung 215-1

dargestellt.

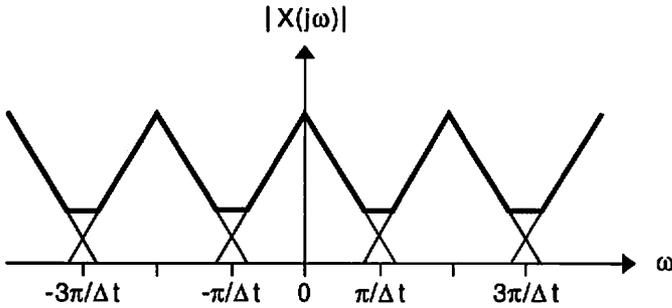


Abb. 215-1: Summe eines um die Abtastfrequenz $2\pi/\Delta t$ verschobenen kontinuierlichen Spektrums

Im allgemeinen werden sich die um $2\pi/\Delta t$ verschobenen kontinuierlichen Spektren überlagern, so daß das Spektrum $X_p(j\omega)$ des diskreten Signals das Spektrum $X(j\omega)$ des kontinuierlichen Signals nur unvollkommen wiedergibt. Man bezeichnet diesen Effekt als *Verzerrung*, im Englischen "aliasing". Dagegen werde nun angenommen, daß das Spektrum des kontinuierlichen Signals bandbegrenzt sei, wobei die Grenze durch die halbe Abtastfrequenz $2\pi/\Delta t$ bestimmt werde, also

$$X(j\omega) = 0 \quad \text{für} \quad |\omega| > \pi/\Delta t . \quad (215.5)$$

Dann folgt aus (215.4)

$$X_p(j\omega) = \frac{1}{\Delta t} X(j\omega) \quad \text{für} \quad |\omega| < \pi/\Delta t . \quad (215.6)$$

Im Intervall $-\pi/\Delta t < \omega < \pi/\Delta t$ reproduziert also das Spektrum $X_p(j\omega)$ des diskreten Signals das Spektrum $X(j\omega)/\Delta t$ des kontinuierlichen Signals. Die Summe (215.4) für diesen Fall ist in Abbildung 215-2 dargestellt.

Damit keine Verzerrung im Spektrum $X_p(j\omega)$ des durch Abtastung erhaltenen Signals auftritt, muß nach (215.6) für die höchste Frequenz ω_m im kontinuierlichen Signal $\omega_m < \pi/\Delta t$ oder

$$\frac{2\pi}{\Delta t} > 2\omega_m \quad (215.7)$$

gelten. Ein bandbegrenzttes Signal ist daher mit einer Frequenz abzutasten, die größer als das Doppelte der höchsten Frequenz ω_m in dem kontinuierlichen Signal ist. Man bezeichnet die Hälfte $\pi/\Delta t$ der Abtastfrequenz als *Nyquist-Frequenz* ω_N , also

$$\omega_N = \pi/\Delta t > \omega_m . \quad (215.8)$$

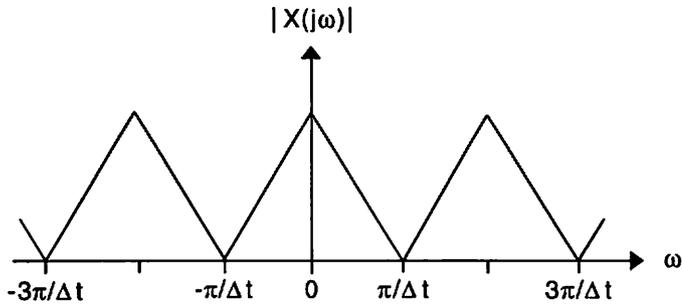


Abb. 215-2: Summe eines um die Abtastfrequenz $2\pi/\Delta t$ verschobenen kontinuierlichen und nach (215.5) bandbegrenzten Spektrums

Liegt nun kein bandbegrenztes Signal vor, enthält $x(t)$ also auch Frequenzanteile im Bereich $|\omega| > \pi/\Delta t$, ist zunächst eine Tiefpaßfilterung durchzuführen, damit für das gefilterte Signal die gewünschte Bandbegrenzung nach (215.5) gilt. Das Tiefpaßfilter wird in (216.17) definiert.

Es soll abschließend die Aufgabe gelöst werden, ein bandbegrenztes kontinuierliches Signal aus seinen Abtastwerten zu rekonstruieren. Die Abtastfrequenz $2\pi/\Delta t$ werde aus der höchsten im Signal vorkommende Frequenz bestimmt. Die Frequenzen ω des Signals liegen also in dem Intervall $-\pi/\Delta t < \omega < \pi/\Delta t$. Das kontinuierliche Signal $x(t)$ folgt dann aus (214.2) mit

$$x(t) = \frac{1}{2\pi} \int_{-\pi/\Delta t}^{\pi/\Delta t} X(j\omega) e^{j\omega t} d\omega \quad (215.9)$$

Das Spektrum $X(j\omega)$ des kontinuierlichen Signals führt mit (215.6) auf das Spektrum $X_p(j\omega)$ des abgetasteten Signals, so daß nach Substitution in (215.9) sich ergibt

$$x(t) = \frac{\Delta t}{2\pi} \int_{-\pi/\Delta t}^{\pi/\Delta t} X_p(j\omega) e^{j\omega t} d\omega$$

und weiter mit $X_p(j\omega)$ aus (215.1)

$$x(t) = \frac{\Delta t}{2\pi} \int_{-\pi/\Delta t}^{\pi/\Delta t} \left(\sum_{k=-\infty}^{\infty} x(k) e^{-jk\Delta t \omega} \right) e^{j\omega t} d\omega .$$

Vertauscht man Integration und Summation folgt

$$x(t) = \sum_{k=-\infty}^{\infty} x(k) f(t - k\Delta t) \quad (215.10)$$

mit der Interpolationsfunktion

$$f(t - k\Delta t) = \frac{\Delta t}{2\pi} \int_{-\pi/\Delta t}^{\pi/\Delta t} e^{j\omega(t - k\Delta t)} d\omega .$$

Wertet man darin das Integral mit (214.4) aus, erhält man

$$f(t - k\Delta t) = \frac{\sin[(\pi/\Delta t)(t - k\Delta t)]}{(\pi/\Delta t)(t - k\Delta t)} . \quad (215.11)$$

Für $t = k\Delta t$ gilt

$$x(t = k\Delta t) = x(k) ,$$

denn

$$\lim_{a \rightarrow 0} \frac{\sin[(\pi/\Delta t)a]}{(\pi/\Delta t)a} = \lim_{a \rightarrow 0} \cos[(\pi/\Delta t)a] = 1 .$$

Die Gleichung (215.10) liefert mit (215.11) unter der Voraussetzung (215.7) das kontinuierliche Signal $x(t)$ aus der *Interpolation* der Abtastwerte $x(k)$. Die Bedingung (215.7) für die Abtastfrequenz stellt zusammen mit der Interpolationsformel (215.10) das *Abtasttheorem* dar.

Durch die Abtastung wird das kontinuierliche Signal $x(t)$ in ein diskretes Signal $x(k)$ überführt. Da die Indizes t und k üblicherweise mit der Zeit identifiziert werden, spricht man von einer *Zeitdiskretisierung*. Für die numerische Verarbeitung sind auch die Signalwerte selbst zu digitalisieren, es ist also eine *Wertdiskretisierung* durchzuführen, worauf hier jedoch nicht eingegangen, sondern auf die Literatur verwiesen wird (Azizi 1983, S.91; Hess 1989, S.13; Hesselmann 1983, S.66).

In Abbildung 215-3 ist das Ablaufschema einer digitalen Signalverarbeitung dargestellt.

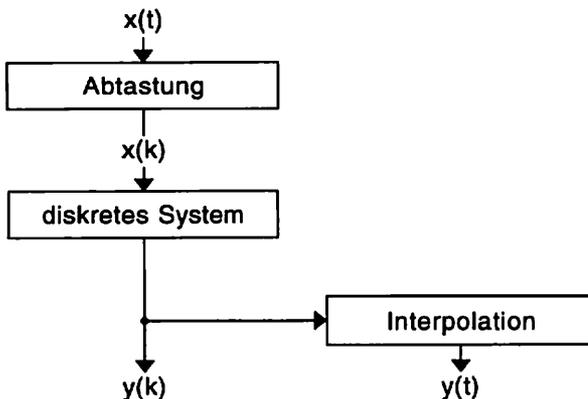


Abb. 215-3: Ablaufschema einer digitalen Signalverarbeitung

Zunächst wird das nach (215.5) bandbegrenzte kontinuierliche Signal $x(t)$ unter Berücksichtigung des Abtasttheorems in das diskrete Signal $x(k)$ überführt. Diesem ersten Teilschritt schließt sich die Transformation durch ein diskretes System beispielsweise unter Modifikation des Frequenzinhaltes in das diskrete Ausgangssignal $y(k)$ an, worauf im folgenden Kapitel eingegangen wird. Will man das diskrete Signal $y(k)$ noch in das kontinuierliche Signal $y(t)$ überführen, besteht der letzte Teilschritt aus dem Interpolationsvorgang nach (215.10).

216 Lineare Systeme

Als *kontinuierliches System* bezeichnet man die Vorschrift, die ein kontinuierliches Eingangssignal $x(t)$ in ein kontinuierliches Ausgangssignal $y(t)$ transformiert, also

$$y(t) = \phi[x(t)] \quad , \quad (216.1)$$

wobei ϕ den Operator bezeichnet, der die Transformation bewirkt. Abbildung 216-1 gibt die schematische Darstellung eines Systems wieder. Die physikalische Realisierung eines Systems bildet beispielsweise ein Stromkreis.



Abb. 216-1: Kontinuierliches System

Führt man Restriktionen für die Transformation (216.1) ein, ergeben sich Klassen von kontinuierlichen Systemen, unter denen die linearen verschiebungsinvarianten Systeme von besonderem Interesse sind. Ein System heißt *linear*, wenn das Superpositionsprinzip gilt. Sind $x_k(t)$ mit $k \in \{1, \dots, K\}$ Eingangssignale, $y_k(t)$ Ausgangssignale und a_k beliebige Konstanten, dann ist das System linear, falls gilt

$$\phi\left[\sum_{k=1}^K a_k x_k(t)\right] = \sum_{k=1}^K a_k \phi[x_k(t)] = \sum_{k=1}^K a_k y_k(t) \quad . \quad (216.2)$$

Ein kontinuierliches System heißt *verschiebungsinvariant*, falls

$$y(t-t_1) = \phi[x(t-t_1)] \quad (216.3)$$

für beliebiges t_1 zutrifft.

Jedes lineare verschiebungsinvariante System läßt sich vollständig durch seine *Impulsantwort* beschreiben, das heißt durch seine Reaktion auf den Impuls, der durch die Deltafunktion $\delta(t)$ gegeben ist. Um das zu zeigen, sei $\delta(t)$ das Eingangssignal eines linearen verschiebungsinvarianten Systems. Die Impulsantwort $h(t)$ folgt dann wegen (216.1) mit

$$h(t) = \phi[\delta(t)] \quad . \quad (216.4)$$

Wegen der Verschiebungsinvarianz (216.3) gilt weiter

$$h(t-t_1) = \phi[\delta(t-t_1)] \quad .$$

Mit der Definition (211.5) der Deltafunktion $\delta(t)$ erhält man, da $\delta(t)$ eine gerade Funktion ist (Papoulis 1977, S.97), für das Eingangssignal $x(t)$

$$x(t) = \int_{-\infty}^{\infty} x(\tau)\delta(t-\tau)d\tau \quad .$$

Die Antwort des Systems auf ein Eingangssignal $x(\tau)\delta(t-\tau)$ ist $x(\tau)\phi[\delta(t-\tau)]$ wegen der Linearität (216.2). Gilt diese Linearität auch für unendlich viele Terme, erhält man

$$y(t) = \phi[x(t)] = \int_{-\infty}^{\infty} x(\tau)\phi[\delta(t-\tau)]d\tau = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau \quad .$$

Der rechte Ausdruck entspricht der Faltung (214.8). Mit der Variablentransformation $u = t-\tau$ und $du = -d\tau$ sowie anschließender Umbenennung mit $u = \tau$ ergibt sich daher schließlich

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau = \int_{-\infty}^{\infty} x(t-\tau)h(\tau)d\tau \quad . \quad (216.5)$$

Hiermit wird offensichtlich, daß ein lineares verschiebungsinvariantes System vollständig durch seine Impulsantwort $h(t)$ beschrieben ist.

Analog den kontinuierlichen Systemen transformieren *diskrete Systeme* diskrete Eingangssignale $x(n)$ in diskrete Ausgangssignale $y(n)$. Fordert man wieder Linearität und Verschiebungsinvarianz entsprechend (216.2) und (216.3), ergibt sich ein diskretes lineares verschiebungsinvariantes System. Ist $h(n)$ nach (216.4) die Impulsantwort, also die Reaktion des Systems auf den mit (211.7) definierten Einheitsimpuls $d(n)$, erhält man anstelle von (216.5)

$$y(n) = x(n) * h(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k) = \sum_{k=-\infty}^{\infty} x(n-k)h(k) \quad . \quad (216.6)$$

Die rechten Ausdrücke in (216.6) bezeichnet man in Analogie zu (214.8) als *diskrete*

Faltung oder *diskrete Konvolution*, so daß wieder die Schreibweise $x(n) * h(n)$ benutzt wird.

Beispiel 1: Es gelte

$$x(n) = 0 \quad \text{für } n < 0 \quad \text{und } n > 3$$

und

$$h(n) = 0 \quad \text{für } n < 0 \quad \text{und } n > 2$$

sowie

$$x(n) \neq 0 \quad \text{und } h(n) \neq 0 \quad \text{für die übrigen Werte von } n ,$$

dann ergibt die diskrete Faltung

$$y(n) = \sum_{k=0}^3 x(k)h(n-k) .$$

Folglich gilt

$$y(n) = 0 \quad \text{für } n < 0 \quad \text{und } n > 5$$

und

$$y(0) = x(0)h(0)$$

$$y(1) = x(0)h(1) + x(1)h(0)$$

$$y(2) = x(0)h(2) + x(1)h(1) + x(2)h(0)$$

$$y(3) = \quad \quad \quad x(1)h(2) + x(2)h(1) + x(3)h(0)$$

$$y(4) = \quad \quad \quad \quad \quad \quad + x(2)h(2) + x(3)h(1)$$

$$y(5) = \quad x(3)h(2)$$

Diskrete lineare verschiebungsinvariante Systeme werden weiter eingeschränkt, indem Stabilität und eventuell Kausalität gefordert werden. Ein *stabiles System* liegt genau dann vor, wenn

$$\sum_{k=-\infty}^{\infty} |h(k)| < \infty \quad (216.7)$$

gilt, so daß ein beschränktes Eingangssignal auch ein beschränktes Ausgangssignal erzeugt. Diese Bedingung muß erfüllt sein, wenn ein System numerisch realisiert werden soll. Ein *kausales System* ist gegeben, wenn das Ausgangssignal an der Stelle $n = n_0$ nur von dem Eingangssignal an den Stellen $n \leq n_0$ und nicht von voraus liegenden Werten

für $n > n_0$ abhängig ist. Wegen (216.6) bedeutet dies

$$h(n) = 0 \quad \text{für } n < 0 \quad . \quad (216.8)$$

Die Bedingung der Kausalität muß erfüllt sein, wenn das System für Echtzeitprobleme eingesetzt wird. Systeme, für die (216.8) nicht zutrifft, werden als *nichtkausal* bezeichnet. Insbesondere sind Systeme mit $h(n) = 0$ für $n > 0$ *antikausal*.

Instruktiv ist die Reaktion eines diskreten linearen verschiebungsinvarianten Systems auf eine harmonische Schwingung als Eingangssignal. Das Ausgangssignal ist nämlich wieder eine harmonische Schwingung, allerdings mit geänderter Amplitude und Phase. Um dies zu zeigen, sei als Eingangssignal entsprechend (212.16) mit $t = n\Delta t$, $x(n) = x(t=n\Delta t)$ und $\Omega = \omega_0\Delta t$ die komplexe Schwingung

$$x(n) = e^{jn\Omega} \quad \text{mit } n \in \{0, \pm 1, \pm 2, \dots\} \quad (216.9)$$

gewählt. Mit (212.9) gilt

$$e^{jn\Omega} = e^{jn(\Omega+2m\pi)} \quad \text{für } m \in \{0, \pm 1, \pm 2, \dots\} \quad ,$$

so daß die Schwingung nach (212.3) mit 2π periodisch ist und $-\pi < \Omega < \pi$ gesetzt wird. Die Frequenz Ω ist daher auf 2π normiert und wird auch als *normierte Frequenz* bezeichnet. Als Ausgangssignal ergibt sich mit (216.6)

$$y(n) = \sum_{k=-\infty}^{\infty} h(k)e^{j(n-k)\Omega} = e^{jn\Omega} \sum_{k=-\infty}^{\infty} h(k)e^{-jk\Omega} \quad .$$

Setzt man

$$H(j\Omega) = \sum_{k=-\infty}^{\infty} h(k)e^{-jk\Omega} \quad , \quad (216.10)$$

folgt mit (213.14)

$$y(n) = H(j\Omega)e^{jn\Omega} = |H(j\Omega)|e^{j\phi(\Omega)}e^{jn\Omega} = |H(j\Omega)|e^{j(\phi(\Omega)/\Omega+n)\Omega} \quad . \quad (216.11)$$

Das Ausgangssignal $y(n)$ ergibt sich also aus der Multiplikation der komplexen Schwingung (216.9) mit dem komplexen Faktor $H(j\Omega)$. Das System bewirkt dabei eine Amplitudenänderung um den reellen Faktor $|H(j\Omega)|$ und eine Phasenverschiebung um den Wert $\phi(\Omega)$. Man bezeichnet die Funktion $H(j\Omega)$, da sie in Abhängigkeit von der Frequenz Ω vorliegt, als *Frequenzantwort* oder *Übertragungsfunktion* des Systems. $H(j\Omega)$ besitzt nach (212.3) die Periode 2π wegen

$$e^{-j\Omega k} = e^{-j(\Omega+2\pi)k} \quad . \quad (216.12)$$

Die Gleichung (216.10) läßt sich auch als Fourier-Reihe der Frequenzantwort $H(j\Omega)$ interpretieren, so daß nach (215.1) und (215.2) die Impulsantwort $h(k)$ wegen

$$h(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(j\Omega) e^{jk\Omega} d\Omega \quad (216.13)$$

als Überlagerung komplexer Schwingungen mit den komplexen Amplituden $H(j\Omega)$ folgt. Das Ausgangssignal $y(n)$ ergibt sich nach (216.6) aus der diskreten Faltung des Eingangssignals $x(n)$ mit der Impulsantwort $h(n)$. Das Spektrum $Y(j\Omega)$ von $y(n)$ erhält man folglich mit (215.1) zu

$$Y(j\Omega) = \sum_{n=-\infty}^{\infty} y(n) e^{-jn\Omega} = \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} x(k) h(n-k) e^{-jn\Omega} . \quad (216.14)$$

Substituiert man $m = n-k$ und $n = m+k$, ergibt sich weiter

$$Y(j\Omega) = \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} x(k) h(m) e^{-j(m+k)\Omega} .$$

Mit

$$X(j\Omega) = \sum_{k=-\infty}^{\infty} x(k) e^{-jk\Omega} \quad \text{und} \quad H(j\Omega) = \sum_{m=-\infty}^{\infty} h(m) e^{-jm\Omega} \quad (216.15)$$

folgt schließlich

$$Y(j\Omega) = H(j\Omega) X(j\Omega) . \quad (216.16)$$

In Analogie zur Fourier-Transformation (214.10) der Faltung kontinuierlicher Signale entspricht (216.16) der Fourier-Transformation der Faltung (216.6) diskreter Signale.

Aus (216.16) folgt das Frequenzverhalten eines linearen verschiebungsinvarianten Systems. Durch Vorgabe der Frequenzantwort $H(j\Omega)$ kann das Spektrum $X(j\Omega)$ des Eingangssignals in einer Weise modifiziert werden, daß ein Spektrum $Y(j\Omega)$ des Ausgangssignals mit gewünschten Eigenschaften im Frequenzbereich erhalten wird. Man spricht dann allgemein von einer *Filterung* des Signals. Wird innerhalb eines Intervalls oder mehrerer Intervalle des Frequenzbereichs die Frequenzantwort vernachlässigbar klein, liegt eine *frequenzselektive Filterung* des Signals vor, worauf im Kapitel 222 näher eingegangen wird.

Beispiel 2: Das ideale Tiefpaßfilter besitzt die Frequenzantwort

$$H(j\Omega) = \begin{cases} 1 & \text{für } |\Omega| \leq \Omega_m \\ 0 & \text{für } \Omega_m < |\Omega| \leq \pi \end{cases} , \quad (216.17)$$

die in Abbildung 216-2 dargestellt ist. Bis zur Grenze Ω_m können alle Frequenzen passieren, alle höheren Frequenzen werden gesperrt. Für die Impulsantwort $h(n)$ folgt mit (214.4) aus (216.13)

$$h(n) = \frac{1}{2\pi} \int_{-\Omega_m}^{\Omega_m} e^{jn\Omega} d\Omega = \frac{1}{n\pi} \operatorname{sinn}\Omega_m \quad (216.18)$$

und damit das gefilterte Signal $y(n)$ aus (216.6). Wegen $h(n) \neq 0$ für $n < 0$ ist das System nach (216.8) nicht kausal. Δ

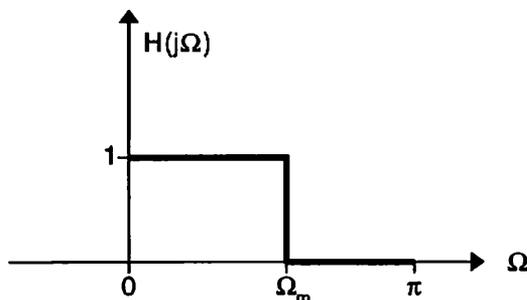


Abb. 216-2: Frequenzantwort des idealen Tiefpassfilters

Für die Filterung ist eine Klasse von linearen verschiebungsinvarianten Systemen von besonderer Bedeutung, deren Eingangssignal $x(n)$ und deren Ausgangssignal $y(n)$ durch eine *Differenzgleichung N-ter Ordnung* mit den konstanten reellen Koeffizienten a_k und b_r verbunden ist,

$$\sum_{k=0}^N a_k y(n-k) = \sum_{r=0}^M b_r x(n-r) . \quad (216.19)$$

Den Differenzgleichungen, die diskrete Systeme definieren, entsprechen die Differentialgleichungen bei stetigen Systemen (Marko 1986, S.166).

Löst man (216.19) für $a_0 \neq 0$ nach $y(n)$ auf

$$y(n) = \sum_{r=0}^M \frac{b_r}{a_0} x(n-r) - \sum_{k=1}^N \frac{a_k}{a_0} y(n-k) , \quad (216.20)$$

erkennt man, daß das Ausgangssignal an der Stelle n sich aus dem Eingangssignal an derselben Stelle, den M vorangegangenen Werten des Eingangssignals und den N vorhergehenden Werten des Ausgangssignals berechnet. Das Ausgangssignal ist das gefilterte Signal.

217 Systemfunktion

So wie die Laplace-Transformation (214.18) zur Verallgemeinerung der Fourier-Transformation (214.2) dient, so kann die *z-Transformation* eines Signals als Verallgemeinerung der Fourier-Reihe (215.1) angesehen werden. Sie ist definiert durch

$$X(z) = \sum_{k=-\infty}^{\infty} x(k)z^{-k} \quad , \quad (217.1)$$

worin z eine komplexe Variable bedeutet.

Stellt man z entsprechend (213.14) durch $z = re^{j\Omega}$ dar und substituiert z in (217.1), folgt

$$X(r, j\Omega) = \sum_{k=-\infty}^{\infty} x(k)r^{-k}e^{-jk\Omega} \quad . \quad (217.2)$$

Mit $r = 1$, also $|z| = 1$, ergibt sich daraus die Fourier-Reihe. Mit (217.2) läßt sich aus der Bedingung (214.6) der absoluten Integrierbarkeit für die Existenz des Fourier-Integrals die Bedingung für die Konvergenz der z -Transformation ableiten, nämlich

$$\sum_{k=-\infty}^{\infty} |x(k)r^{-k}| < \infty \quad . \quad (217.3)$$

Wegen der Multiplikation mit r^{-k} konvergiert für $r > 1$ die z -Transformation von Signalen, deren Fourier-Reihe nicht konvergiert.

Beispiel 1: Die z -Transformation der Folge

$$x(n) = 0 \quad \text{für } n < 0 \quad \text{und} \quad x(n) = a^n \quad \text{für } n \geq 0 \quad ,$$

wobei a reell sei, ergibt sich aus (217.1) mit (213.6) zu

$$X(z) = \sum_{k=0}^{\infty} a^k z^{-k} = \sum_{k=0}^{\infty} (az^{-1})^k = \frac{1}{1-az^{-1}} \quad \text{für } |z| > |a| \quad .$$

Die Transformation $X(z)$ konvergiert also für $|z| > |a|$. Die einzige Singularität, auch *Polstelle* genannt, befindet sich bei $z = a$. Δ

Der diskreten Faltung des Eingangssignals $x(n)$ mit der Impulsantwort $h(n)$ entspricht nach einer z -Transformation die Multiplikation von $X(z)$ und $H(z)$. Mit

$$y(n) = x(n) * h(n) \quad (217.4)$$

aus (216.6) und mit der z -Transformation $Y(z)$ des Ausgangssignals $y(n)$ gilt also, wie

im folgenden gezeigt wird,

$$Y(z) = X(z)H(z) \quad (217.5)$$

Eine entsprechende Beziehung war mit (216.16) für die Fourier-Transformation der diskreten Faltung erhalten worden. Man bezeichnet $H(z)$ als *Systemfunktion* oder auch als *Übertragungsfunktion*. Wird die Systemfunktion auf dem Einheitskreis, das heißt für $|z| = 1$ berechnet, folgt mit $z = e^{j\Omega}$ anstelle der Systemfunktion $H(z)$ die Frequenzantwort $H(j\Omega)$ in (216.10).

Um (217.5) abzuleiten, wird die z -Transformation von (217.4) gebildet. Man erhält mit (216.6) und (217.1)

$$\begin{aligned} Y(z) &= \sum_{k=-\infty}^{\infty} y(k)z^{-k} = \sum_{k=-\infty}^{\infty} \left[\sum_{l=-\infty}^{\infty} x(l)h(k-l) \right] z^{-k} \\ &= \sum_{l=-\infty}^{\infty} x(l) \sum_{k=-\infty}^{\infty} h(k-l)z^{-k} \end{aligned}$$

Mit $m = k-l$ und $k = m+l$ folgt weiter

$$Y(z) = \sum_{l=-\infty}^{\infty} x(l) \left[\sum_{m=-\infty}^{\infty} h(m)z^{-m} \right] z^{-l}$$

und mit (217.1) schließlich (217.5).

Die Differenzengleichung (216.19), die, wie erwähnt, eine Klasse von linearen Systemen charakterisiert, die für die Filterung bedeutsam ist, soll jetzt einer z -Transformation unterworfen werden. Man erhält

$$\sum_{n=-\infty}^{\infty} \left[\sum_{k=0}^N a_k y(n-k) \right] z^{-n} = \sum_{n=-\infty}^{\infty} \left[\sum_{r=0}^M b_r x(n-r) \right] z^{-n}$$

oder

$$\sum_{k=0}^N a_k \sum_{n=-\infty}^{\infty} y(n-k)z^{-n} = \sum_{r=0}^M b_r \sum_{n=-\infty}^{\infty} x(n-r)z^{-n}$$

und mit $n-k = m$, $n = m+k$, $n-r = o$, $n = o+r$

$$\sum_{k=0}^N a_k z^{-k} Y(z) = \sum_{r=0}^M b_r z^{-r} X(z) \quad (217.6)$$

Aus dem Koeffizientenvergleich von (217.6) mit (217.5) folgt die zugehörige Systemfunktion

$$H(z) = \sum_{r=0}^M b_r z^{-r} / \sum_{k=0}^N a_k z^{-k} = Z(z^{-1})/N(z^{-1}) . \quad (217.7)$$

Hiermit ist $H(z)$ als der Quotient der Polynome $Z(z^{-1})$ und $N(z^{-1})$ in z^{-1} dargestellt. Substituiert man im Zählerpolynom $r = M-p$ und $p = M-r$, erhält man wegen

$$Z(z^{-1}) = \sum_{r=0}^M b_r z^{-r} = z^{-M} \sum_{p=0}^M b_{M-p} z^p = z^{-M} Z'(z)$$

mit $Z'(z)$ ein Polynom in z , das auch in der Produktform

$$Z'(z) = b_0 \prod_{p=1}^M (z - z_{0p}) \quad (217.8)$$

dargestellt werden kann. Dabei ergeben sich die *Nullstellen* z_{0p} als Funktionen der Koeffizienten $b_r = b_{M-p}$. Wird auf die gleiche Weise das Nennerpolynom $N(z^{-1})$ aus (217.7) in ein Polynom $N'(z)$ umgewandelt, erhält man die Systemfunktion zu

$$H(z) = \frac{b_0}{a_0} z^{N-M} \frac{\prod_{p=1}^M (z - z_{0p})}{\prod_{q=1}^N (z - z_{\infty q})} . \quad (217.9)$$

Diese Darstellung wird als *Produktform* der Systemfunktion bezeichnet (Azizi 1983, S.135). Sie enthält neben den Nullstellen z_{0p} die *Polstellen* $z_{\infty q}$. Ihre Bedeutung liegt in der Überprüfung der Stabilitätseigenschaft des Systems, wie im Zusammenhang mit (221.18) gezeigt wird.

Wie aus der nach $y(n)$ aufgelösten Form (216.20) der Differenzgleichung (216.19) ersichtlich ist, bedeutet (216.19) eine *rekursive Filterung*, denn das Ausgangssignal wird nicht nur durch Werte des Eingangssignals, sondern auch durch vorangegangene Werte des Ausgangssignals selbst berechnet. Schreibt man (216.20) in eine nichtrekursive Form um, indem das Ausgangssignal $y(n-k)$ auf der rechten Seite von (216.20) durch die Werte der linken Seite von (216.20) für $n-k$ substituiert wird, erhält man

$$y(n) = \sum_{m=0}^{\infty} h(m)x(n-m) , \quad (217.10)$$

worin die Koeffizienten $h(m)$ Funktionen der Koeffizienten a_k und b_r sind. Durch einen Vergleich mit (216.6) läßt sich $h(m)$ als Impulsantwort mit unendlicher Länge interpretieren. Rekursive Filter bezeichnet man daher auch als *Filter mit unendlicher Impulsantwort*, im Englischen "infinite impulse response filter" oder IIR-Filter.

Eine *nichtrekursive Filterung* erhält man mit $a_0 \neq 0$ und $a_k = 0$ für alle $k \neq 0$ in (216.20).

Aus (217.7) folgt dann für die Systemfunktion

$$H(z) = \sum_{k=0}^M (b_k/a_0)z^{-k} \quad (217.11)$$

oder mit

$$h(k) = b_k/a_0 \quad (217.12)$$

die Darstellung

$$H(z) = \sum_{k=0}^M h(k)z^{-k} . \quad (217.13)$$

Aus (216.20) ergibt sich das Ausgangssignal $y(n)$ zu

$$y(n) = \sum_{k=0}^M h(k)x(n-k) . \quad (217.14)$$

Wegen (216.6) ist dies die Faltung des Eingangssignals $x(n)$ mit der Impulsantwort $h(n)$. Aufgrund dieser Beziehung bezeichnet man nichtrekursive Filter auch als *Filter mit endlicher Impulsantwort*, im Englischen "finite impulse response filter" oder FIR-Filter.

Soll für eine vorgegebene Systemfunktion $H(z)$ ein rekursives Filter entworfen werden, sind die Koeffizienten a_k und b_r in der Darstellung (217.7) zu bestimmen, so daß das gefilterte Signal $y(n)$ aus (216.20) berechnet werden kann. Hierauf wird in den Kapiteln 223 und 224 eingegangen. Entsprechend sind beim Entwurf eines nichtrekursiven Filters für eine vorgegebene Systemfunktion $H(z)$ die Koeffizienten b_k/a_0 in (217.11) zu ermitteln, so daß das gefilterte Signal mit (217.14) folgt. Dies wird in den Kapiteln 225 und 227 behandelt.

Anstatt für die nichtrekursive Filterung die Faltung (217.14) zu berechnen, kann es numerisch weniger aufwendig sein, das Eingangssignal $x(n)$ und die Impulsantwort $h(n)$ einer Fourier-Transformation zu unterziehen und anschließend zu multiplizieren, um nach (216.16) das transformierte Ausgangssignal zu erhalten. Die inverse Fourier-Transformation ergibt dann das gefilterte Signal $y(n)$. Auf dieses Verfahren wird im folgenden Kapitel und im Kapitel 228 eingegangen. Die Grundlagen dieser Methode bilden die diskrete und die schnelle Fourier-Transformation.

218 Diskrete Fourier-Transformation

Mit (213.10) wird ein kontinuierliches und periodisches Signal in ein diskretes Spektrum und mit (215.1) ein diskretes Signal in ein kontinuierliches Spektrum transformiert. Um zur diskreten Fourier-Transformation zu gelangen, werden diskrete und periodische Signale und Spektren vorausgesetzt. Man geht von der Darstellung eines periodischen Signals durch eine Fourier-Reihe aus. Aus N gleichmäßig über die Periode T des Signals verteilten Abtastwerten sollen K Harmonische bestimmt werden. Mit $N = 2K+1$ aus (213.5) erhält man daher

$$K = (N-1)/2 . \quad (218.1)$$

Setzt man

$$t = n\Delta t \quad \text{mit } n \in \{0, 1, \dots, N-1\} ,$$

folgt mit (213.1) wegen $T = N\Delta t$ die Grundfrequenz aus

$$\omega_0 = 2\pi/(N\Delta t) . \quad (218.2)$$

Definiert man weiter das diskrete periodische Signal $x_p(n)$ durch

$$x_p(n) = x_p(t=n\Delta t) ,$$

erhält man aus (213.2) und (213.10) in komplexer Darstellung

$$x_p(n) = \frac{1}{N} \sum_{k=-(N-1)/2}^{(N-1)/2} X_p(k) e^{j(2\pi/N)kn} \quad \text{mit } n \in \{0, 1, \dots, N-1\} , \quad (218.3)$$

worin $X_p(k)$ das diskrete Amplitudenspektrum bedeutet. Der Faktor $1/N$ wurde in Übereinstimmung mit der üblichen Notation eingeführt (Oppenheim und Schaffer 1975, S.88; Rabiner und Gold 1975, S.51). Aus (213.7) und (213.12) folgt für das Amplitudenspektrum

$$\begin{aligned} \frac{1}{N} X_p(k) &= \frac{1}{2} (A_k - jB_k) = \frac{1}{N} \sum_{n=0}^{N-1} x_p(n) \left(\cos \frac{2\pi}{N}kn - j \sin \frac{2\pi}{N}kn \right) \\ \frac{1}{N} X_p(-k) &= \frac{1}{2} (A_k + jB_k) = \frac{1}{N} \sum_{n=0}^{N-1} x_p(n) \left(\cos \frac{2\pi}{N}kn + j \sin \frac{2\pi}{N}kn \right) \end{aligned} \quad (218.4)$$

und mit (212.8)

$$X_p(k) = \sum_{n=0}^{N-1} x_p(n) e^{-j(2\pi/N)kn} \quad (218.5)$$

sowie der entsprechenden Beziehung für $X_p(-k)$. Nach Voraussetzung ist das Signal $x_p(n)$ periodisch mit N . Die identische Periode besitzt auch das Amplitudenspektrum $X_p(k)$ wegen

$$e^{-j(2\pi/N)kn} = e^{-j(2\pi/N)(k+mN)n} \quad \text{mit } m \in \{0, \pm 1, \pm 2, \dots\} \quad (218.6)$$

nach (212.9). Spaltet man (218.3) in

$$x_p(n) = \frac{1}{N} \sum_{k=-(N-1)/2}^{-1} X_p(k) e^{j(2\pi/N)kn} + \frac{1}{N} \sum_{k=0}^{(N-1)/2} X_p(k) e^{j(2\pi/N)kn}$$

auf und substituiert in der ersten Summe $k = 1-N$ und $l = k+N$, so erhält man mit $X_p(1) = X_p(1-N)$

$$\begin{aligned} x_p(n) &= \frac{1}{N} \sum_{l=(N+1)/2}^{N-1} X_p(1-N) e^{j(2\pi/N)(1-N)n} + \frac{1}{N} \sum_{k=0}^{(N-1)/2} X_p(k) e^{j(2\pi/N)kn} \\ &= \frac{1}{N} \sum_{l=(N+1)/2}^{N-1} X_p(1) e^{j(2\pi/N)ln} + \frac{1}{N} \sum_{k=0}^{(N-1)/2} X_p(k) e^{j(2\pi/N)kn} . \end{aligned}$$

Faßt man die beiden Summen zusammen, ergibt sich

$$x_p(n) = \frac{1}{N} \sum_{k=0}^{N-1} X_p(k) e^{j(2\pi/N)kn} \quad \text{mit } n \in \{0, 1, \dots, N-1\} . \quad (218.7)$$

Diese Gleichung unterscheidet sich von (218.3) nur durch die Summationsgrenzen. Aufgrund der Periodizität läßt sich also der Beginn der Summation von $k = -(N-1)/2$ auf $k = 0$ verschieben.

Aus dem periodischen Signal $x_p(n)$ soll nun ein Signal $x(n)$ von endlicher Länge dadurch entstehen, daß aus $x_p(n)$ nur die Werte einer Periode übernommen und die übrigen Werte gleich Null gesetzt werden,

$$x(n) = \begin{cases} x_p(n) & \text{für } 0 \leq n \leq N-1 \\ 0 & \text{sonst} . \end{cases} \quad (218.8)$$

Auf die gleiche Weise soll aus dem periodischen Amplitudenspektrum $X_p(k)$ das Spektrum $X(k)$ von endlicher Länge entstehen

$$X(k) = \begin{cases} X_p(k) & \text{für } 0 \leq k \leq N-1 \\ 0 & \text{sonst} . \end{cases} \quad (218.9)$$

Da in (218.5) und (218.7) Werte nur im Intervall von 0 bis $N-1$ auftauchen, erhält man jetzt für das nach (218.8) definierte Signal $x(n)$ von endlicher Länge

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j(2\pi/N)kn} \quad \text{mit } n \in \{0, 1, \dots, N-1\} \quad (218.10)$$

und

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j(2\pi/N)kn} \quad \text{mit } k \in \{0, 1, \dots, N-1\} . \quad (218.11)$$

Diese beiden Beziehungen bezeichnet man im Gegensatz zur Fourier-Transformation (214.2) als *diskrete Fourier-Transformation* und (218.10) auch als *inverse diskrete Fourier-Transformation*. Wie die Ableitung von (218.7) aus (218.3) zeigt, die entsprechend auch für (218.5) gilt, bilden (218.3) und (218.5) mit den Summationsgrenzen von (218.3) ebenfalls eine diskrete Fourier-Transformation. Vielfach wird (218.10) als IDFT und (218.11) als DFT abgekürzt bezeichnet.

Wendet man die z-Transformation (217.1) auf das Signal $x(n)$ an, erhält man wegen (218.8)

$$X(z) = \sum_{n=0}^{N-1} x(n) z^{-n} . \quad (218.12)$$

Die diskrete Fourier-Transformation $X(k)$ und die z-Transformation $X(z)$ sind also durch die Beziehung

$$X(k) = X(z=e^{j(2\pi/N)k}) \quad (218.13)$$

verbunden. Mit $z = e^{j(2\pi/N)k}$ folgt wegen (213.14) $|z| = 1$, so daß die diskrete Fourier-Transformation $X(k)$ aus der z-Transformation $X(z)$ dadurch erhalten wird, daß z auf dem Einheitskreis an den Stellen $\varphi = (2\pi/N)k$ berechnet wird.

Der diskreten Faltung (216.6) entspricht nach der Fourier-Transformation die Multiplikation (216.16) im Frequenzbereich. Für die diskrete Fourier-Transformation sollen jetzt die analogen Beziehungen hergeleitet werden. Das Signal $x_p(n)$ sei wieder periodisch mit der Periode N und ebenfalls die Impulsantwort $h_p(n)$ eines Systems. Bildet man eine der diskreten Faltung (216.6) entsprechende Beziehung mit

$$y_p(n) = \sum_{k=0}^{N-1} x_p(k) h_p(n-k) , \quad (218.14)$$

dann ist $y_p(n)$ ebenfalls mit N periodisch. Dies läßt sich mit (218.7) entsprechend (218.6) zeigen. Die Beziehung (218.14) bezeichnet man als *periodische Faltung*. Sie unterscheidet sich von der Faltung (216.6) dadurch, daß, wenn die Indizes $n-k$ in $h_p(n-k)$ die Periode $0 \leq n-k \leq N-1$ verlassen, die nächste Periode von $h_p(n-k)$ zum Tragen

kommt. Die diskrete Fourier-Transformation der periodischen Faltung (218.14) berechnet sich mit (218.11) zu

$$\begin{aligned}
 Y_p(k) &= \sum_{n=0}^{N-1} \sum_{l=0}^{N-1} x_p(l) h_p(n-l) e^{-j(2\pi/N)kn} \\
 &= \sum_{l=0}^{N-1} x_p(l) \left[\sum_{n=0}^{N-1} h_p(n-l) e^{-j(2\pi/N)k(n-l)} \right] e^{-j(2\pi/N)kl} \\
 &= H_p(k) X_p(k) .
 \end{aligned} \tag{218.15}$$

Der periodischen Faltung (218.14) entspricht also wieder die Multiplikation (218.15).

Geht man mit (218.8) wieder zu den Signalen mit endlicher Länge über und gilt

$$x(n) \neq 0 \quad \text{für} \quad 0 \leq n \leq M-1$$

und

$$h(n) \neq 0 \quad \text{für} \quad 0 \leq n \leq K-1 ,$$

dann folgt im Gegensatz zur periodischen Faltung (218.14) die *lineare Faltung* mit

$$y(n) = \sum_{k=0}^{M-1} x(k) h(n-k) \quad \text{für} \quad n \in \{0, 1, \dots, N-1\} \tag{218.16}$$

und

$$N-1 = M+K-2 . \tag{218.17}$$

Daß $y(n)$ insgesamt N Werte ungleich Null besitzt, ist daraus ersichtlich, daß für k in (218.16) maximal $k = M-1$ und für $n-k$ maximal $n-k = K-1$ gilt, woraus für n der Maximalwert $n = M+K-2$ folgt. Die lineare Faltung (218.16) entspricht der Faltung (217.14), die das Ausgangssignal der nichtrekursiven Filterung liefert.

Bildet man die diskrete Fourier-Transformation von (218.16) nach (218.11), ergibt sich

$$\begin{aligned}
 Y(k) &= \sum_{n=0}^{N-1} y(n) e^{-j(2\pi/N)kn} = \sum_{n=0}^{N-1} \left[\sum_{l=0}^{M-1} x(l) h(n-l) \right] e^{-j(2\pi/N)kn} \\
 &= \sum_{l=0}^{M-1} x(l) \left[\sum_{n=0}^{N-1} h(n-l) e^{-j(2\pi/N)k(n-l)} \right] e^{-j(2\pi/N)kl} \\
 &= H(k) X(k) .
 \end{aligned} \tag{218.18}$$

Der linearen Faltung (218.16) entspricht also das Produkt der diskreten Fourier-Transformationen $H(k)$ und $X(k)$.

Für die Berechnung diskreter Fourier-Transformationen existieren schnelle Algorithmen, wie im folgenden Kapitel gezeigt wird. Es kann daher effizienter sein, das Ausgangssignal nicht über die lineare Faltung (218.16) zu berechnen, sondern das Eingangssignal und die Impulsantwort zu transformieren, die Transformationen wegen (218.18) zu multiplizieren und das Ausgangssignal aus der inversen Transformation zu gewinnen. Diese Art der Berechnung bezeichnet man als *schnelle Faltung*. In Abbildung 218-1 sind die beiden Vorgehensweisen gegenübergestellt.

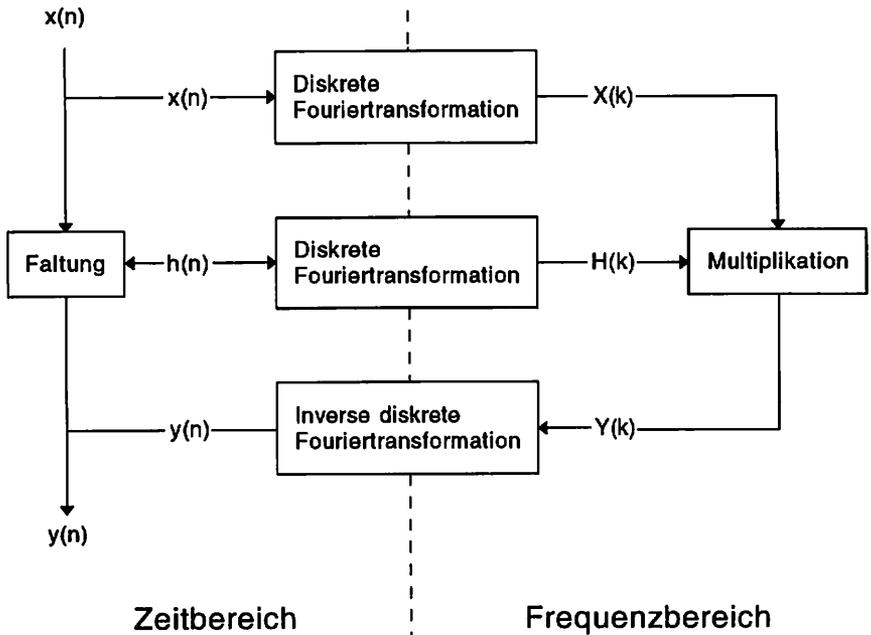


Abb. 218-1: Lineare und schnelle Faltung

Ist bei Anwendung der schnellen Faltung das Eingangssignal im Vergleich zur Impulsantwort sehr lang, kann es für die Fourier-Transformation unvorteilhaft sein, das ganze Eingangssignal zu bearbeiten. Das Eingangssignal $x(n)$ wird dann in Stücke $x_k(n)$ der Länge L zerlegt mit

$$x_k(n) = \begin{cases} x(n) & \text{für } kL \leq n \leq (k+1)L-1 \text{ und } k \in \{0, 1, \dots, P\} \\ 0 & \text{sonst.} \end{cases}$$

Das Eingangssignal $x(n)$ ergibt sich dann als Summe der Stücke $x_k(n)$ aus

$$x(n) = \sum_{k=0}^P x_k(n)$$

und die lineare Faltung als Summe der Faltung der einzelnen Stücke

$$x(n) * h(n) = \sum_{k=0}^P (x_k(n) * h(n)) \quad (218.19)$$

Um die Faltung $x_k(n) * h(n)$ der Stücke zu erhalten, ist gemäß (218.17) die Fourier-Transformation auf der Basis von $L+K-1$ Punkten zu berechnen, falls die Impulsantwort $h(n)$ wieder K Werte umfaßt. Die Stücke des Ausgangssignals, die durch die inverse Transformation gewonnen werden, überlappen sich um $K-1$ Werte. Diese Werte sind nach (218.19) zu addieren und ergeben zusammen mit den sich nicht überlappenden L Werten das Ausgangssignal $y(n)$. Diese Methode des Überlappens und Addierens wird im Englischen mit "overlap-add method" bezeichnet.

Eine weitere Methode, die lineare Faltung für ein sehr langes Eingangssignal nach (218.18) zu berechnen, besteht darin, das Eingangssignal wieder in Stücke der Länge L zu zerlegen, dabei das Eingangssignal aber um $K-1$ Werte zu überlappen, falls K wieder die Länge der Impulsantwort bedeutet. Nach der Fourier-Transformation, der Multiplikation nach (218.18) und der inversen Transformation sind die sich überlappenden Stücke fortzulassen, während die restlichen Stücke aufbewahrt werden und das gewünschte Ausgangssignal ergeben. Im Englischen bezeichnet man dieses Verfahren daher als "overlap-save method" (Oppenheim und Schafer 1975, S.113).

Beispiel 1: Mit (214.5) war die Fourier-Transformation eines kontinuierlichen Rechtecksignals erhalten worden. Entsprechend soll jetzt die diskrete Fourier-Transformation des diskreten Rechtecksignals

$$x(n) = \begin{cases} 1 & \text{für } n \in \{-L, -L+1, \dots, L\} \\ 0 & \text{sonst} \end{cases} \quad (218.20)$$

berechnet werden. Mit

$$\Omega = 2k\pi / (2L+1)$$

erhält man aus (218.11)

$$X(\Omega) = \sum_{n=-L}^L e^{-jn\Omega}$$

und mit (213.6) sowie der Substitution $n = l-L$ und $l = n+L$

$$X(\Omega) = e^{jL\Omega} \sum_{l=0}^{2L} e^{-jl\Omega} = e^{jL\Omega} \frac{1 - e^{-j(2L+1)\Omega}}{1 - e^{-j\Omega}}$$

$$= \frac{e^{jL\Omega} e^{-j(2L+1)\Omega/2} (e^{j(2L+1)\Omega/2} - e^{-j(2L+1)\Omega/2})}{e^{-j\Omega/2} (e^{j\Omega/2} - e^{-j\Omega/2})} .$$

Mit (212.11) folgt hieraus schließlich die diskrete Fourier-Transformation $X(\Omega)$ des diskreten Rechtecksignals (218.20)

$$X(\Omega) = \frac{\sin(2L+1)\Omega/2}{\sin\Omega/2} \quad (218.21)$$

△

Beispiel 2: Die Fourier-Transformation des kontinuierlichen Dreiecksignals ist mit (214.14) gegeben. Die diskrete Fourier-Transformation des diskreten Dreiecksignals

$$x(n) = \begin{cases} 1 - \frac{|n|}{L+1} & \text{für } n \in \{-L, -L+1, \dots, L\} \\ 0 & \text{sonst,} \end{cases} \quad (218.22)$$

worin L eine gerade Zahl bedeutet, berechnet sich mit

$$\Omega = 2k\pi / (2L+1)$$

aus (218.11) zu

$$X(\Omega) = \sum_{n=-L}^L \left(1 - \frac{|n|}{L+1}\right) e^{-jn\Omega} .$$

Mit (213.14) erhält man

$$\begin{aligned} \frac{1}{L+1} \left| \sum_{n=-L/2}^{L/2} e^{-jn\Omega} \right|^2 &= \frac{1}{L+1} \sum_{n_1=-L/2}^{L/2} \sum_{n_2=-L/2}^{L/2} e^{-j(n_1-n_2)\Omega} \\ &= \sum_{k=-L}^L \left(1 - \frac{|k|}{L+1}\right) e^{-jk\Omega} , \end{aligned} \quad (218.23)$$

denn nach Substitution von $k = n_1 - n_2$ gilt $k = \text{const.}$ für die von links unten nach rechts oben verlaufenden Diagonalen des Gitters mit den $(L+1)^2$ Gitterpunkten, über die zu summieren ist, siehe Abbildung 218-2. Löst man die Doppelsumme durch Summation entlang der Diagonalen, erhält man die rechte Seite von (218.23). Aus (218.21) folgt

$$\sum_{n=-L/2}^{L/2} e^{-jn\Omega} = \frac{\sin(L+1)\Omega/2}{\sin\Omega/2}$$

und damit aus (218.23) die diskrete Fourier-Transformation $X(\Omega)$ des diskreten Dreiecksignals (218.22)

$$X(\Omega) = \frac{1}{L+1} \left[\frac{\sin(L+1)\Omega/2}{\sin\Omega/2} \right]^2 . \quad (218.24)$$

Diese Funktion bezeichnet man auch als *Fejer-Kern*.

Δ

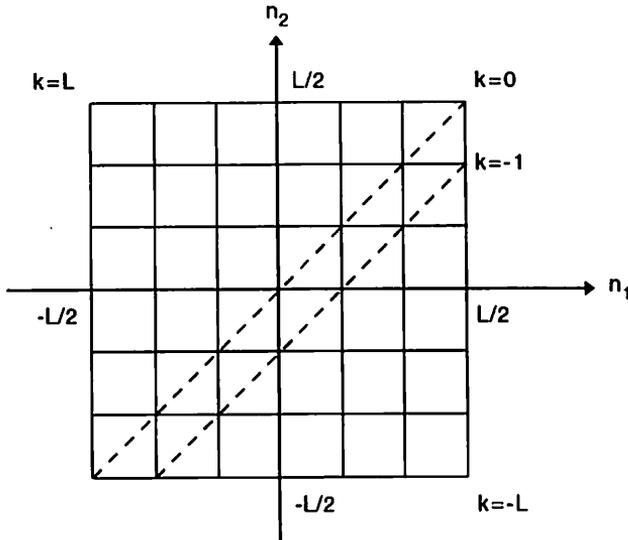


Abb. 218-2: Summation entlang der Diagonalen

219 Schnelle Fourier-Transformation

Die numerische Berechnung der diskreten Fourier-Transformation (218.11) und der inversen diskreten Fourier-Transformation (218.10) soll nun behandelt werden. Mit der Definition

$$W_N = e^{-j(2\pi/N)} \quad (219.1)$$

lautet (218.11)

$$X(k) = \sum_{n=0}^{N-1} x(n) (W_N)^{kn} \quad \text{mit } k \in \{0, 1, \dots, N-1\} \quad (219.2)$$

und (218.10)

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) (W_N)^{-kn} \quad \text{mit } n \in \{0, 1, \dots, N-1\} , \quad (219.3)$$

worin sowohl $x(n)$ als auch $X(k)$ komplex sein können. Die vollständige Auswertung von beispielsweise (219.2) erfordert, da N komplexe Multiplikationen für jeden Wert von k notwendig sind, insgesamt $M_0 = N^2$ komplexe Multiplikationen. Die Tatsache, daß sich die Rechenzeit etwa proportional zur Anzahl der komplexen Multiplikationen verhält (Schüßler 1973, S.195), macht die Notwendigkeit der Reduzierung der Rechenschritte deutlich. Die Entwicklung effizienter Algorithmen basiert auf dem für (219.1) wegen (218.6) geltenden Zusammenhang

$$(W_N)^{kn} = (W_N)^{k(n+N)} = (W_N)^{(k+N)n} \quad , \quad (219.4)$$

der die Anzahl der Rechenoperation deutlich verringert, wie im folgenden für (219.2) gezeigt wird. Entsprechende Überlegungen gelten auch bei der Auswertung von (219.3), da lediglich das Vorzeichen im Exponenten von W_N sich von dem in (219.2) unterscheidet.

Auf Cooley und Tukey (1965) geht ein Algorithmus zur Berechnung der diskreten Fourier-Transformation zurück, der die Entwicklung einer Vielzahl von Verfahren bewirkte, die gemeinsam unter dem Begriff "*schnelle Fourier-Transformation*" oder unter der englischen Bezeichnung "*fast Fourier transform*" einzuordnen sind. Das grundlegende Prinzip liegt in der Aufspaltung einer vollständigen Sequenz von Werten der Länge N in eine Anzahl kürzerer Partialsequenzen. Handelt es sich bei der Sequenz um das Signal $x(n)$, spricht man von *Zeitzerlegung*, bei dem Spektrum $X(k)$ hingegen von *Frequenzzerlegung*. Voraussetzung zur Anwendung dieser Algorithmen ist die Bedingung

$$N = 2^v \quad \text{mit} \quad v \in \{1, 2, \dots\} \quad , \quad (219.5)$$

so daß die Sequenz fortlaufend durch die Division durch zwei in kleinere Partialsequenzen aufgespalten werden kann. In der ersten Stufe der *Zeitzerlegung* wird die Sequenz $x(n)$ mit N Elementen nach

$$x_{i_1}(1) = x(21+i_1) \quad \text{mit} \quad 1 \in \{0, 1, \dots, N/2-1\} \quad \text{und} \quad i_1 \in \{0, 1\} \quad (219.6)$$

in die beiden Partialsequenzen $x_0(n)$ und $x_1(n)$ mit jeweils $N/2$ Elementen aufgespalten. In $x_0(n)$ sind alle Signalwerte mit geradzahligem Argument, in $x_1(n)$ alle Werte mit ungeradzahligem Argument enthalten. Es folgt dann mit (219.1) aus (219.2)

$$\begin{aligned} X(k) &= \sum_{l=0}^{N/2-1} x_0(l) (W_N)^{2lk} + \sum_{l=0}^{N/2-1} x_1(l) (W_N)^{(2l+1)k} \\ &= \sum_{l=0}^{N/2-1} x_0(l) (W_{N/2})^{lk} + (W_N)^k \sum_{l=0}^{N/2-1} x_1(l) (W_{N/2})^{lk} \quad . \end{aligned}$$

Der Faktor $(W_N)^k$ vor der zweiten Summe kann für $k \in \{N/2, \dots, N-1\}$ mit der Substitu-

tion $k = n+N/2$ mit $n \in \{0, \dots, N/2-1\}$ umgeformt werden

$$(\mathbb{W}_N)^{n+N/2} = (\mathbb{W}_N)^n (\mathbb{W}_N)^{N/2} = -(\mathbb{W}_N)^n .$$

Weiterhin gilt wegen (219.4) in den beiden Summen

$$(\mathbb{W}_{N/2})^{1(n+N/2)} = (\mathbb{W}_{N/2})^{1n} .$$

Berücksichtigt man diese Ergebnisse in obiger Gleichung für $X(k)$, indem n wieder durch k ersetzt wird, erhält man

$$X(k+i_1N/2) = X_0(k) + (-1)^{i_1} (\mathbb{W}_N)^k X_1(k) \quad (219.7)$$

mit $k \in \{0, 1, \dots, N/2-1\}$ und $i_1 \in \{0, 1\}$. Darin sind $X_0(k)$ und $X_1(k)$ mit

$$X_{i_1}(k) = \sum_{l=0}^{N/2-1} x_{i_1}(l) (\mathbb{W}_{N/2})^{kl} \quad (219.8)$$

zwei Fourier-Transformationen der Länge $N/2$ und einer Periodizität in k mit der Periode $N/2$. Bei der Berechnung von (219.7) sind $M_1 = 2(N/2)^2 + N/2 = N^2/2 + N/2$ komplexe Multiplikationen auszuführen, da der zweite Summand $N/2$ weitere komplexe Multiplikationen mit $(\mathbb{W}_N)^k$ beinhaltet. Gegenüber $M_0 = N^2$ bedeutet dies aber schon für $N = 4$ eine Verringerung.

In der ersten Stufe wurde also die diskrete Fouriertransformation (219.2) der Länge $N = 2^v$ in zwei diskrete Fouriertransformationen (219.8) der Länge $N/2 = 2^{v-1}$ zerlegt. In der zweiten Stufe werden $X_0(k)$ und $X_1(k)$ analog (219.7) aufgespalten in

$$X_{i_1}(k+i_2N/4) = X_{i_1,0}(k) + (-1)^{i_2} (\mathbb{W}_{N/2})^k X_{i_1,1}(k) \quad (219.9)$$

mit $k \in \{0, 1, \dots, N/4-1\}$ und $i_1, i_2 \in \{0, 1\}$. Darin bedeuten

$$X_{i_1, i_2}(k) = \sum_{l=0}^{N/4-1} x_{i_1, i_2}(l) (\mathbb{W}_{N/4})^{kl} \quad (219.10)$$

vier Fourier-Transformationen der vier Partialsequenzen

$$x_{i_1, i_2}(l) = x_{i_1}(2l+i_2) \quad \text{mit } l \in \{0, 1, \dots, N/4-1\} \quad (219.11)$$

der Länge $N/4 = 2^{v-2}$. Die Anzahl der komplexen Multiplikationen, die zur Berechnung von $X(k)$ notwendig sind, reduziert sich auf $M_2 = 2(2(N/4)^2 + N/4) + N/2 = N^2/4 + 2(N/2)$.

Es läßt sich somit allgemein feststellen, daß in der m -ten Stufe der Zeiterlegung, wobei $m \in \{1, 2, \dots, v\}$ gilt, 2^{m-1} diskrete Fourier-Transformationen

$$X_{i_1 \dots i_{m-1}}(k+i_m N/2^m) = X_{i_1 \dots i_{m-1} 0}(k) + (-1)^{i_m (\mathbb{W}_{N/2^{m-1}})^k} X_{i_1 \dots i_{m-1} 1}(k) \quad (219.12)$$

mit $k \in \{0, 1, \dots, N/2^m - 1\}$ und $i_1, \dots, i_m \in \{0, 1\}$ der Länge 2^{v-m+1} in 2^m diskrete Fourier-Transformationen

$$X_{i_1 \dots i_{m-1} i_m}(k) = \sum_{l=0}^{N/2^m - 1} x_{i_1 \dots i_{m-1} i_m}(l) (\mathbb{W}_{N/2^m})^{kl} \quad (219.13)$$

der 2^m Partialsequenzen

$$x_{i_1 \dots i_{m-1} i_m}(1) = x_{i_1 \dots i_{m-1}}(2l+i_m) \quad \text{mit } l \in \{0, 1, \dots, N/2^m - 1\} \quad (219.14)$$

der Länge 2^{v-m} aufgespalten werden. Dabei verbleiben zur Berechnung von $X(k)$ mit $k \in \{0, 1, \dots, N-1\}$ insgesamt $M_m = (N^2/2^m) + m(N/2)$ komplexe Multiplikationen. In der v -ten Stufe der Zeiterlegung, also für $m = v$ folgt aus (219.13)

$$X_{i_1 \dots i_{v-1} i_v}(k) = x_{i_1 \dots i_{v-1} i_v}(0) \quad (219.15)$$

mit $k = 0$, da $(\mathbb{W}_1)^0 = 1$ gilt. Da mit $(\mathbb{W}_1)^0$ nicht multipliziert zu werden braucht, reduziert sich die Anzahl der komplexen Multiplikationen zur Berechnung von $X(k)$ mit $k \in \{0, 1, \dots, N-1\}$ auf $M_v = v(N/2) = (N/2) \log_2 N$ wegen (219.5). Tatsächlich läßt sich die Anzahl weiter verringern, da selbstverständlich sämtliche Multiplikationen mit $(\mathbb{W}_N)^0$, $(\mathbb{W}_{N/2})^0, \dots$ unterbleiben können.

Zur Erläuterung der *Frequenzerlegung* schreibt man (219.2) mit (219.5) zunächst um in

$$\begin{aligned} X(k) &= \sum_{n=0}^{N/2-1} x(n) (\mathbb{W}_N)^{kn} + \sum_{n=N/2}^{N-1} x(n) (\mathbb{W}_N)^{kn} \\ &= \sum_{n=0}^{N/2-1} [x(n) (\mathbb{W}_N)^{kn} + (\mathbb{W}_N)^{kN/2} x(n+N/2) (\mathbb{W}_N)^{kn}] \\ &= \sum_{n=0}^{N/2-1} [x(n) + (-1)^k x(n+N/2)] (\mathbb{W}_N)^{kn} , \end{aligned} \quad (219.16)$$

da $(\mathbb{W}_N)^{kN/2} = (-1)^k$ wegen (212.9) gilt. Wählt man anstelle von (219.6) in der ersten Stufe der Frequenzerlegung mit $k = 2l+i_1$ für $l \in \{0, 1, \dots, N/2-1\}$ und $i_1 \in \{0, 1\}$ die Aufspaltung

$$X_{i_1}(1) = X(2l+i_1) , \quad (219.17)$$

so folgen aus (219.16) wegen

$$(\mathbb{W}_N)^{(2l+i_1)n} = (\mathbb{W}_N)^{2ln}(\mathbb{W}_N)^{ni_1} = (\mathbb{W}_{N/2})^{ln}(\mathbb{W}_N)^{ni_1}$$

die beiden diskreten Fourier-Transformationen der Länge $N/2$

$$X_{i_1}(1) = \sum_{n=0}^{N/2-1} x_{i_1}(n)(\mathbb{W}_{N/2})^{ln} \quad (219.18)$$

mit

$$x_{i_1}(n) = [x(n) + (-1)^{i_1} x(n+N/2)](\mathbb{W}_N)^{ni_1} \quad (219.19)$$

Zur Berechnung von $X(k)$ für $k \in \{0, 1, \dots, N-1\}$ sind somit $M_1 = 2(N/2)^2 + N/2 = N^2/2 + N/2$ komplexe Multiplikationen auszuführen, wie aus (219.18) und (219.19) deutlich wird.

Allgemein läßt sich also feststellen, daß in der m -ten Stufe der Frequenzerlegung für $m \in \{1, 2, \dots, v\}$ die 2^m diskreten Fourier-Transformationen

$$\begin{aligned} X_{i_1 \dots i_{m-1} i_m}(1) &= X_{i_1 \dots i_{m-1} i_m}(2l+i_m) \\ &= \sum_{n=0}^{N/2^{m-1}-1} x_{i_1 \dots i_{m-1} i_m}(n)(\mathbb{W}_{N/2^m})^{ln} \end{aligned} \quad (219.20)$$

mit $l \in \{0, 1, \dots, N/2^{m-1}-1\}$, $i_1, \dots, i_m \in \{0, 1\}$ und

$$x_{i_1 \dots i_{m-1} i_m}(n) = [x_{i_1 \dots i_{m-1} i_m}(n) + (-1)^{i_m} x_{i_1 \dots i_{m-1} i_m}(n+N/2^m)](\mathbb{W}_{N/2^{m-1}})^{ni_m} \quad (219.21)$$

der Länge 2^{v-m} durch Aufspaltung erhalten werden. Die Anzahl der komplexen Multiplikationen beträgt wie bei der Zeiterlegung $M_m = (N^2/2^m) + m(N/2)$. So gilt dann auch für $m = v$, also in der v -ten Stufe, $M_v = (N/2) \log_2 N$.

Da die schnelle Fourier-Transformation für die numerische Berechnung der diskreten Fourier-Transformation so bedeutsam ist, wird sie in der Literatur ausführlich behandelt. Wegen weiterer Rechenverfahren und Programmimplementierungen sei daher beispielsweise auf (Nussbaumer 1981, S.80; Oppenheim und Schaffer 1975, S.284; Rabiner und Gold 1975, S.356; Stearns 1984, S.105) verwiesen.

22 Eindimensionale digitale lineare Filter

221 Allgemeine Definitionen und Eigenschaften

Der Ausgangspunkt der folgenden Betrachtungen ist die Differenzengleichung (216.19), die mit den Substitutionen $g_k = a_k/a_0$ und $d_r = b_r/a_0$

$$\sum_{k=0}^N g_k y(n-k) = \sum_{r=0}^M d_r x(n-r) \quad (221.1)$$

lautet. Hieraus ergibt sich die Systemfunktion $H(z)$ nach (217.7) als Quotient zweier Polynome in z^{-1}

$$H(z) = \sum_{r=0}^M d_r z^{-r} / \sum_{k=0}^N g_k z^{-k} \quad (221.2)$$

oder nach (217.9) in der Produktform als Funktion der Nullstellen z_{0r} und Polstellen $z_{\infty k}$

$$H(z) = d_0 z^{N-M} \prod_{r=1}^M (z - z_{0r}) / \prod_{k=1}^N (z - z_{\infty k}) . \quad (221.3)$$

Löst man (221.1) nach $y(n)$ auf, erhält man wegen $g_0 = 1$ die Definitionsgleichung eines *kausalen linearen digitalen Filters* mit den *Filterkoeffizienten* d_r und g_k

$$y(n) = \sum_{r=0}^M d_r x(n-r) - \sum_{k=1}^N g_k y(n-k) , \quad (221.4)$$

da, wie im Zusammenhang mit (216.8) erläutert, $y(n)$ an der Stelle $n = n_0$ nur von $x(n)$ an den Stellen $n \leq n_0$ abhängt. Zur Berechnung von (221.4) sind Anfangswerte notwendig. Vielfach wird von $x(n) = y(n) = 0$ für $n < 0$ ausgegangen. Nach (217.10) läßt sich das Ausgangssignal $y(n)$ durch

$$y(n) = \sum_{k=0}^{\infty} h(k)x(n-k) \quad (221.5)$$

darstellen. Für die Impulsantwort gilt die Bedingung (216.8) der Kausalität

$$h(k) = 0 \quad \text{für} \quad k < 0 . \quad (221.6)$$

Die Differenzengleichung (221.1) kann mit $g_N \neq 0$ auch in die Form

$$y(n-N) = \frac{1}{g_N} \left[\sum_{r=0}^M d_r x(n-r) - \sum_{k=0}^{N-1} g_k y(n-k) \right]$$

gebracht werden. Man erhält weiter mit den Substitutionen $m = n - N$ und $n = m + N$

$$y(m) = \frac{1}{g_N} \left[\sum_{r=0}^M d_r x(m+N-r) - \sum_{k=0}^{N-1} g_k y(m+N-k) \right] \quad (221.7)$$

und schließlich mit $p = N - r$, $r = N - p$, $q = N - k$ und $k = N - q$ sowie $\bar{d}_p = d_{N-p}/g_N$ und $\bar{g}_q = g_{N-q}/g_N$

$$y(m) = \sum_{p=N-M}^N \bar{d}_p x(m+p) - \sum_{q=1}^N \bar{g}_q y(m+q) . \quad (221.8)$$

Hiermit liegt ein *nichtkausales lineares digitales Filter* vor, denn zur Berechnung des Ausgangssignals $y(m)$ werden die Eingangs- und Ausgangssignale $x(m+p)$ und $y(m+q)$ benötigt. Wird analog zu der Vorgehensweise, die zu (217.10) führt, das Ausgangssignal $y(m+q)$ auf der rechten Seite von (221.8) durch die linke Seite substituiert, erhält man

$$y(m) = \sum_{k=-\infty}^{M-N} h(k) x(m-k) . \quad (221.9)$$

Für die Impulsantwort gilt also die nichtkausale Bedingung

$$h(k) = 0 \quad \text{für } k > M - N . \quad (221.10)$$

Die Auswertung von (221.8) oder (221.9) erfolgt entlang der m -Achse in negativer Richtung. Man bezeichnet diese nichtkausale Filterung daher als *Filterung nach rückwärts*, während die kausale Vorgehensweise nach (221.4) auch *Filterung nach vorwärts* genannt wird.

Die Systemfunktion (221.2) wurde aus der Differenzgleichung (216.19) beziehungsweise (221.1) hergeleitet, so daß sie sowohl für das kausale als auch das nichtkausale Filter gültig ist. Soll im zweiten Fall $H(z)$ als Funktion der Koeffizienten \bar{d}_p und \bar{g}_q dargestellt werden, substituiert man in (221.2) wie für den Übergang von (221.7) nach (221.8) und erhält

$$H(z) = \sum_{p=N-M}^N \bar{d}_p z^p / \sum_{q=0}^N \bar{g}_q z^q . \quad (221.11)$$

Betrachtet man den Spezialfall $M = N$, so geht (221.3) in

$$H(z) = d_0 \prod_{k=1}^N \frac{z - z_{0k}}{z - z_{\infty k}} \quad (221.12)$$

über. Aus (221.8) ergibt sich mit

$$y(n) = \sum_{r=0}^N \bar{d}_r x(n+r) - \sum_{k=1}^N \bar{g}_k y(n+k) \quad (221.13)$$

die Definitionsgleichung eines *antikausalen linearen digitalen Filters*, da für die Impulsantwort, wie im Zusammenhang mit (216.8) erwähnt, die Bedingung

$$h(k) = 0 \quad \text{für } k > 0 \quad (221.14)$$

gilt.

Man erkennt aus (221.3) und (221.12), daß sich die Systemfunktion aus dem Produkt einzelner Teilsystemfunktionen für Teilfilter zusammensetzen läßt. Bezeichnet man die Systemfunktion des ersten Teilfilters mit $H_1(z)$, lautet die z -Transformation $Y_1(z)$ des Ausgangssignals $y_1(n)$ nach (217.5)

$$Y_1(z) = H_1(z)X(z) .$$

Für das zweite Teilfilter mit der Systemfunktion $H_2(z)$ dient $y_1(n)$ als Eingangssignal. Die z -Transformation $Y_2(z)$ des Ausgangssignals $y_2(n)$ ergibt sich dann zu

$$Y_2(z) = H_2(z)Y_1(z) = H_2(z)H_1(z)X(z) .$$

Sind auf diese Weise insgesamt K verschiedene Teilfilter mit $K > 1$ hintereinander geschaltet, erhält man am Ausgang des K -ten Teilfilters die z -Transformation $Y_K(z) = Y(z)$ des Ausgangssignals $y_K(n)$ zu

$$\begin{aligned} Y(z) &= H_K(z)Y_{K-1}(z) = H_K(z)H_{K-1}(z) \dots H_1(z)X(z) \\ &= \left(\prod_{k=1}^K H_k(z) \right) X(z) . \end{aligned} \quad (221.15)$$

Der Vergleich mit (217.5) liefert die Systemfunktion $H(z)$ als Produkt der K Teilsystemfunktionen $H_k(z)$

$$H(z) = \prod_{k=1}^K H_k(z) . \quad (221.16)$$

Analog zu (217.4) entspricht dieser Multiplikation die Faltung der K Teilimpulsantworten $h_k(n)$. Man erhält also die Impulsantwort zu

$$h(n) = h_1(n) * h_2(n) * \dots * h_K(n)$$

und das Ausgangssignal $y(n)$ mit (216.6) zu

$$y(n) = x(n) * h_1(n) * h_2(n) * \dots * h_K(n) ,$$

so daß das Ausgangssignal $x(n) * h_1(n)$ des ersten Teilfilters Eingangssignal des zweiten Teilfilters mit der Impulsantwort $h_2(n)$ wird und so fort. Ein digitales lineares Filter läßt sich also in Teilfilter zerlegen. Sind sie nach (221.16) hintereinander geschaltet, spricht man von der *Serien-* oder *Kaskadenstruktur*, die in Abbildung 221-1 dargestellt ist.

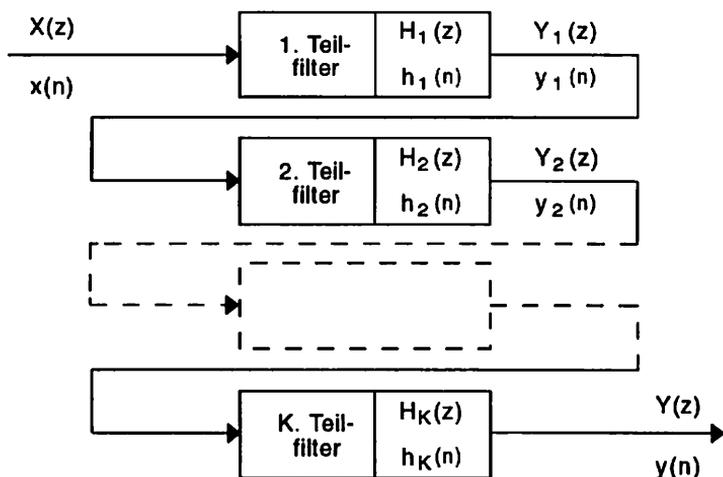


Abb. 221-1: Filter in Kaskadenstruktur

Dient hingegen das Signal $x(n)$ allen K Teilfiltern mit $K > 1$ als Eingangssignal, so lauten die z -Transformationen $Y_k(z)$ des Ausgangssignals $y_k(n)$ mit $k \in \{1, \dots, K\}$

$$Y_k(z) = H_k(z)X(z) .$$

Die z -Transformation $Y(z)$ des Ausgangssignals des gesamten Filterprozesses ergibt sich aus

$$Y(z) = \sum_{k=1}^K Y_k(z) = \left(\sum_{k=1}^K H_k(z) \right) X(z) ,$$

so daß die Systemfunktion $H(z)$ wegen (217.5) zu

$$H(z) = \sum_{k=1}^K H_k(z) \quad (221.17)$$

folgt. Entsprechend erhält man die Impulsantwort $h(n)$ mit

$$h(n) = \sum_{k=1}^K h_k(n)$$

als Summe der K Teilimpulsantworten $h_k(n)$.

Ein digitales lineares Filter läßt sich also auch in Teilfilter zerlegen, die parallel geschaltet sind. Es ist dann in der *Parallelstruktur* realisiert, die in Abbildung 221-2 dargestellt ist.

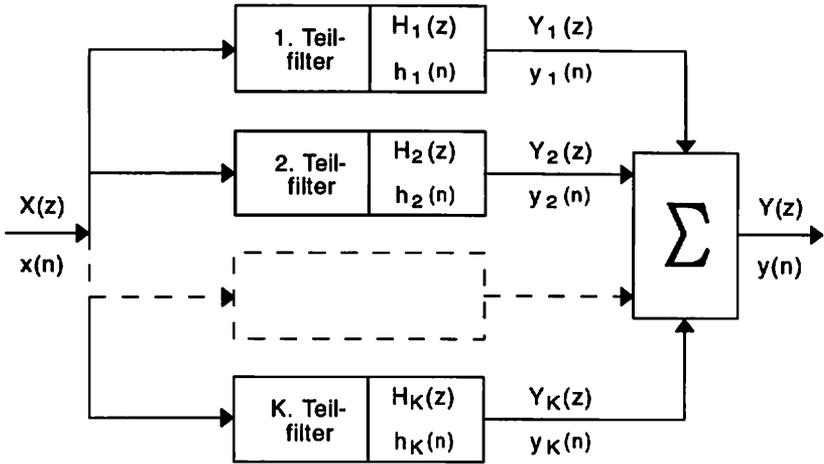


Abb. 221-2: Filter in Parallelstruktur

Mit (221.16) und (221.17) wurden zwei Filterstrukturen vorgestellt, die in den folgenden Kapiteln benötigt werden. Auf weitere Filterstrukturen wird hier nicht eingegangen, da sie hauptsächlich für die Entwicklung effizienter Algorithmen bei der numerischen Realisierung eines Filters von Bedeutung sind (Azizi 1983, S.182; Rabiner und Gold 1975, S.40; Schüßler 1973, S.61).

Es wird nun die *Stabilitätseigenschaft* eines digitalen Filters untersucht, denn wegen (216.7) liefert nur ein stabiles Filter bei einem beschränkten Eingangssignal wieder ein beschränktes Ausgangssignal. Der folgende Satz beinhaltet als Stabilitätskriterium die Polstellen der Systemfunktion (221.3). Es kann daher ohne Einschränkung der Allgemeingültigkeit für die Obergrenzen $M = N$ angenommen werden. Betrachtet wird also das Verhalten der Systemfunktion $H(z)$ aus (221.12) in den Fällen der kausalen Filterung nach (221.6) und der antikausalen Filterung nach (221.14).

Satz: Ein lineares digitales Filter ist genau dann stabil, wenn für die Polstellen $z_{\infty k}$ mit $k \in \{1, \dots, N\}$ der Systemfunktion $H(z)$ aus (221.12) im kausalen Fall (221.6) die Ungleichung $|z_{\infty k}| < 1$ zutrifft und im antikausalen Fall (221.14) die Ungleichung $|z_{\infty k}| > 1$ gilt.

$$(221.18)$$

Beweis: Für die Systemfunktion eines mit (221.6) kausalen Systems erhält man nach (217.1)

$$H(z) = \sum_{n=0}^{\infty} h(n)z^{-n} . \quad (221.19)$$

Die Bedingung für die Konvergenz von $H(z)$ lautet mit $z = re^{j\Omega}$ gemäß (217.3)

$$|H(z)| = \left| \sum_{n=0}^{\infty} h(n)r^{-n}e^{-jn\Omega} \right| \leq \sum_{n=0}^{\infty} |h(n)r^{-n}| < \infty . \quad (221.20)$$

Nimmt man an, daß das System stabil ist, dann ist diese Gleichung wegen (216.7) für $|z| = r \geq 1$ erfüllt. Damit der Nenner in (221.12) nicht verschwindet, muß daher die Ungleichung $|z_{\infty k}| < 1$ zutreffen. Gilt für die Impulsantwort eines Systems hingegen (221.14), lautet die Systemfunktion nach (217.1)

$$H(z) = \sum_{n=-\infty}^0 h(n)z^{-n} . \quad (221.21)$$

Man erhält dann anstelle von (221.20)

$$|H(z)| = \left| \sum_{n=-\infty}^0 h(n)r^{-n}e^{-jn\Omega} \right| \leq \sum_{n=-\infty}^0 |h(n)r^{-n}| < \infty . \quad (221.22)$$

Diese Gleichung ist unter der Annahme der Stabilität für $|z| = r \leq 1$ erfüllt, was bedeutet, daß aus (221.12) die Ungleichung $|z_{\infty k}| > 1$ folgt.

Für (221.12) kann andererseits geschrieben werden

$$H(z) = d_0 \prod_{k=1}^N \frac{1 - z_{0k}/z}{1 - z_{\infty k}/z} = d_0 \prod_{k=1}^N \frac{z_{0k}(1 - z/z_{0k})}{z_{\infty k}(1 - z/z_{\infty k})} . \quad (221.23)$$

Es liege nun ein mit (221.6) kausales System vor und es gelte außerdem $|z_{\infty k}| < 1$ für $k \in \{1, \dots, N\}$. Um Stabilität nachzuweisen, sind die Nennerausdrücke in (221.23) für $z = re^{j\Omega}$ mit $|z| = r \geq 1$ zu untersuchen. Der Nenner $1 - z_{\infty k}/z$ kann mit (213.6) als geometrische Reihe geschrieben werden

$$\frac{1}{1 - z_{\infty k}/z} = \sum_{n=0}^{\infty} (z_{\infty k}/z)^n .$$

Da mit dieser Reihe wegen $|z_{\infty k}|/|z| < 1$ auch $H(z)$ konvergiert, folgt aus (221.20)

$$\sum_{n=0}^{\infty} |h(n)r^{-n}| < \infty$$

und daher mit $r = 1$ die Stabilitätsforderung (216.7). Gilt für die Impulsantwort des Systems die Bedingung (221.14) und ferner $|z_{\infty k}| > 1$ für $k \in \{1, \dots, N\}$, so kann der Anteil $(1 - z/z_{\infty k})$ des Nenners auf der rechten Seite von (221.23) als geometrische Reihe mit $z = re^{j\Omega}$ für $|z| = r \leq 1$ entwickelt werden

$$\frac{1}{1 - z/z_{\infty k}} = \sum_{n=0}^{\infty} (z/z_{\infty k})^n .$$

Mit dieser Reihe konvergiert wiederum $H(z)$, so daß nach (221.22)

$$\sum_{n=-\infty}^0 |h(n)r^{-n}| < \infty$$

folgt. Mit $r = 1$ erhält man wieder die Stabilitätsbedingung (216.7). □

Die Nullstellen z_{0r} für $r \in \{1, \dots, M\}$ und die Polstellen z_{ok} für $k \in \{1, \dots, N\}$ der Systemfunktion (221.2) erfüllen die Bedingungen

$$\sum_{n=0}^M d_n(z_{0r})^{-n} = 0 \quad \text{und} \quad \sum_{n=0}^N g_n(z_{ok})^{-n} = 0 .$$

Ist beispielsweise z_{0r} eine komplexe Zahl, so ist wegen

$$\sum_{n=0}^M d_n(z_{0r})^{-n} = \left(\sum_{n=0}^M d_n(z_{0r})^{-n} \right)^* = \sum_{n=0}^M d_n(z_{0r}^*)^{-n} = 0 \quad (221.24)$$

auch die konjugiert komplexe Zahl z_{0r}^* eine Nullstelle. Es läßt sich also feststellen, daß die Nullstellen einer Systemfunktion entweder reelle Zahlen sind oder, falls sie sich als komplexe Zahlen ergeben, in konjugiert komplexen Paaren auftreten. Diese Aussage gilt in gleicher Weise für die Polstellen.

Soll aus der z -Transformation

$$Y_1(z) = H_1(z)X(z) \quad (221.25)$$

die z -Transformation $X(z)$ zurückgewonnen werden, so wird die Systemfunktion der *inversen Filterung* oder *Dekonvolution*

$$H_2(z) = (H_1(z))^{-1} \quad (221.26)$$

mit $H_1(z) \neq 0$ benötigt. Es folgt dann nämlich aus (221.15) mit (221.25) und (221.26)

$$\begin{aligned} Y(z) &= H_2(z)Y_1(z) = (H_1(z))^{-1}H_1(z)X(z) \\ &= X(z) . \end{aligned} \quad (221.27)$$

Mit $H_1(z) = 0$ verschwindet $Y_1(z)$ und $X(z)$ ist nicht rekonstruierbar. Wie man aus der Produktform (221.3) erkennt, sind die Nullstellen der Systemfunktion $H_1(z)$ nach (221.26) die Polstellen der inversen Systemfunktion $H_2(z)$. Nimmt man beispielsweise an, $H_1(z)$ sei die Systemfunktion eines stabilen kausalen Filters, so gilt nach (221.18) für die Beträge der Polstellen z_{ok} die Bedingung

$$|z_{ok}| < 1 \quad \text{für} \quad k \in \{1, \dots, N\} . \quad (221.28)$$

Das kausale inverse Filter mit der Systemfunktion $H_2(z)$ aus (221.26) ist nur dann stabil,

wenn die Beträge der Nullstellen z_{0r} von $H_1(z)$ die Ungleichung

$$|z_{0r}| < 1 \quad \text{für } r \in \{1, \dots, M\} \quad (221.29)$$

erfüllen. Die kausale inverse Filtergleichung erhält man aus der Differenzgleichung (221.1), wenn man sie mit $d_0 \neq 0$ nach $x(n)$ auflöst. Man bezeichnet ein kausales System als *minimalphasiges Filter*, wenn für seine Systemfunktion die Bedingungen (221.28) und (221.29) gelten (Hess 1989, S.110; Papoulis 1977, S.54).

Der Begriff der Phase eines Filters wird eingeführt, indem man seine Systemfunktion $H(z)$ mit $z = e^{j\Omega}$ als Frequenzantwort in der Form (213.14) schreibt

$$H(j\Omega) = A(\Omega)e^{j\Phi(\Omega)} \quad (221.30)$$

Hierin bedeuten

$$A(\Omega) = |H(j\Omega)| = (H(j\Omega)H^*(j\Omega))^{1/2} = (H(j\Omega)H(-j\Omega))^{1/2} \quad (221.31)$$

die *Amplitude* und

$$\Phi(\Omega) = \arctan(\text{Im}(H(j\Omega))/\text{Re}(H(j\Omega))) \quad (221.32)$$

die *Phase* der Frequenzantwort. Die Darstellung der Amplitude beziehungsweise der Phase in Abhängigkeit von der Frequenz Ω wird als *Amplituden-* beziehungsweise *Phasenspektrum* der Frequenzantwort bezeichnet. Aus (221.31) und (221.32) werden die in der Literatur häufig verwendeten Begriffe der Dämpfung und der Gruppenlaufzeit abgeleitet, siehe beispielsweise (Hess 1989, S.54). Dabei ist die *Dämpfung* $a = a(\Omega)$ allgemein über die Beziehung $a = 20(\log|B_1/B_2|)$ definiert, in der B_1 und B_2 zwei Amplituden in derselben Maßeinheit bedeuten. Sie erhält - obwohl dimensionslos - die Bezeichnung Dezibel (dB), um auf die logarithmische Berechnung hinzuweisen. Mit $B_1 = 1$ und $B_2 = A(\Omega)$ aus (221.31) ergibt sich

$$a(\Omega) = -20(\log A(\Omega)) \quad (221.33)$$

Wie zum Beispiel bei (Buttkus 1991, S.196) erläutert, wird zur Beurteilung des Amplitudenspektrums eines Filters vielfach die Dämpfung herangezogen, da aufgrund der logarithmischen Berechnung Details besser zu erkennen sind. Die *Gruppenlaufzeit* $\tau = \tau(\Omega)$ entspricht der Ableitung der Phase (221.32) nach der Frequenz Ω

$$\tau(\Omega) = d\Phi(\Omega)/d\Omega \quad (221.34)$$

Nach (216.11) bewirkt die Phase $\Phi(\Omega)$ der Frequenzantwort eine Phasenverschiebung des Ausgangssignals gegenüber dem Eingangssignal. Vielfach ist es erwünscht, diese Phasenänderung durch einen linearen Ansatz zu beschreiben. Liegt also die Phase in der Form

$$\Phi(\Omega) = \alpha_1\Omega + \alpha_2 \quad (221.35)$$

vor, wobei α_1 und α_2 reelle Konstanten bedeuten, spricht man von einem *linearphasigen Filter* mit der nach (221.34) konstanten Gruppenlaufzeit

$$\tau(\Omega) = \alpha_1 \quad . \quad (221.36)$$

Setzt man (221.30) mit (221.35) in (216.13) ein, ergibt sich

$$h(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\Omega) e^{j((n+\alpha_1)\Omega+\alpha_2)} d\Omega \quad . \quad (221.37)$$

Ausgehend von dieser Gleichung für die Impulsantwort eines linearphasigen Filters werden im folgenden vier Fälle behandelt, indem für α_1 und α_2 verschiedene Annahmen getroffen werden.

$$\text{a) } \alpha_1 \text{ beliebig und } \alpha_2 = 0 \quad (221.38)$$

Wählt man also $\alpha_2 = 0$ und berücksichtigt, daß $h(n)$ eine reelle Funktion ist, folgt aus (221.37) mit (212.8)

$$h(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\Omega) \cos((n+\alpha_1)\Omega) d\Omega \quad (221.39)$$

und nach Substitution von $m = n+\alpha_1$ und $n = m-\alpha_1$

$$h(m-\alpha_1) = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\Omega) \cos m\Omega d\Omega = h(-m-\alpha_1) \quad . \quad (221.40)$$

Die Impulsantwort ist also wegen (212.10) symmetrisch bezüglich $-\alpha_1$.

$$\text{b) } \alpha_1 = 0 \text{ und } \alpha_2 = 0 \quad (221.41)$$

Mit $\alpha_1 = \alpha_2 = 0$ erhält man aus (221.35)

$$\phi(\Omega) = 0 \quad , \quad (221.42)$$

so daß aus (221.30) die reelle Frequenzantwort

$$H(j\Omega) = A(\Omega) \quad (221.43)$$

folgt. Filter mit einer reellen Frequenzantwort besitzen wegen (221.42) die Phase Null und werden daher als *nullphasige Filter* bezeichnet. Aus (221.40) ergibt sich die Impulsantwort

$$\begin{aligned} h(n) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\Omega) \cos n\Omega d\Omega \\ &= h(-n) \end{aligned} \quad (221.44)$$

nach (213.16) als gerade Funktion.

$$c) \alpha_1 \text{ beliebig und } \alpha_2 = \pi/2 \quad (221.45)$$

Für den Fall erhält man aus (221.37) anstelle von (221.39) wegen $\cos(x+\pi/2) = -\sin x$

$$h(n) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} A(\Omega) \sin((n+\alpha_1)\Omega) d\Omega \quad (221.46)$$

und mit den Substitutionen $m = n+\alpha_1$ und $n = m-\alpha_1$

$$h(m-\alpha_1) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} A(\Omega) \sin m\Omega d\Omega = -h(-m-\alpha_1). \quad (221.47)$$

Die Impulsantwort ist demnach antisymmetrisch bezüglich $-\alpha_1$.

$$d) \alpha_1 = 0 \text{ und } \alpha_2 = \pi/2 \quad (221.48)$$

Mit $\alpha_1 = 0$ und $\alpha_2 = \pi/2$ ergibt sich aus (221.35) die Phase

$$\Phi(\Omega) = \pi/2 \quad (221.49)$$

und aus (221.30) die wegen (212.8) imaginäre Frequenzantwort

$$H(j\Omega) = A(\Omega) e^{j\pi/2} = jA(\Omega) \quad (221.50)$$

Die Impulsantwort erhält man aus (221.46)

$$\begin{aligned} h(n) &= -\frac{1}{2\pi} \int_{-\pi}^{\pi} A(\Omega) \sin n\Omega d\Omega \\ &= -h(-n) \end{aligned} \quad (221.51)$$

Sie ist nach (213.17) eine ungerade Funktion.

Die beschriebenen vier Fälle werden in den Kapiteln 223 und 225 über die rekursiven und nichtrekursiven Filter eingehender untersucht.

222 Allgemeiner Filterentwurf

Beim *Entwurf* digitaler Filter stellt sich die Aufgabe, die von dem System geforderten Eigenschaften zumindest näherungsweise zu realisieren. Es ist somit eine Idealfunktion $f(x)$ durch eine Realfunktion $f_0(x)$ zu approximieren, wobei x eine beliebige Variable bedeutet. Dies kann in Form einer *Entwurfsaufgabe* formuliert werden:

Gegeben sei die Idealfunktion

$$f(x) \quad . \quad (222.1)$$

Gesucht ist dann die Realfunktion

$$f_0(x) \quad (222.2)$$

unter Beachtung einer Modellparametermenge

$$M \quad . \quad (222.3)$$

Die Elemente der Modellparametermenge (222.3) legen das *Toleranzschema* fest, das die Realfunktion (222.2) erfüllen soll.

Das zu realisierende Filter wird nach (221.1) charakterisiert durch die Filterkoeffizienten d_r mit $r \in \{0, \dots, M\}$ und g_k mit $k \in \{1, \dots, N\}$. Faßt man sie mit $M+N+1 = u$ in dem $u \times 1$ Vektor

$$\boldsymbol{\beta} = [d_0, d_1, \dots, d_M, g_1, g_2, \dots, g_N]' \quad (222.4)$$

zusammen, so läßt sich für die Realfunktion (222.2) schreiben

$$f_0(x) = f_0(x, \boldsymbol{\beta}) \quad . \quad (222.5)$$

Alternativ können in (222.4) anstelle der Filterkoeffizienten auch die Null- und Polstellen der Systemfunktion (221.3) gewählt werden. Man erhält dann

$$\boldsymbol{\beta} = [d_0, z_{01}, \dots, z_{0M}, z_{\infty 1}, z_{\infty 2}, \dots, z_{\infty N}]' \quad . \quad (222.6)$$

Die Lösung der Entwurfsaufgabe läßt sich also auf die Bestimmung des Parametervektors $\boldsymbol{\beta}$ aus (222.4) oder (222.6) zurückführen. Hierzu sind eine Reihe Verfahren entwickelt worden, von denen im Kapitel 224 zwei für rekursive Filter und im Kapitel 227 zwei für nichtrekursive Filter vorgestellt werden. Weiterhin lassen sich die Filterkoeffizienten auch auf empirische Weise festlegen, worauf im Kapitel 226 eingegangen wird.

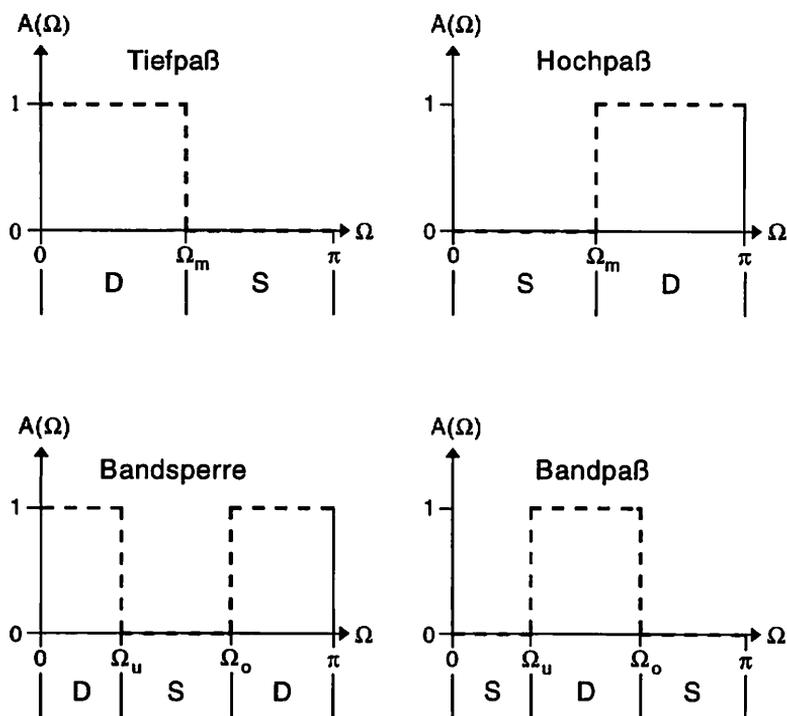
Ziel aller Methoden ist eine möglichst gute Approximation der Idealfunktion durch die Realfunktion. Um dies zu beurteilen, bedient man sich der Approximationsfehlerfunktion

$$e(x) = f_0(x, \boldsymbol{\beta}) - f(x) \quad . \quad (222.7)$$

Bei der Bestimmung der Realfunktion ist die Stabilität des Systems zu beachten. Bedingungen hierfür befinden sich im Satz (221.18).

Die Idealfunktion (222.1) kann allgemein mit beliebigen, das Filter charakterisierenden Funktionen identifiziert werden, so beispielsweise mit der Frequenzantwort, dem Amplitudenspektrum oder der Impulsantwort. Vielfach werden Approximationen idealer frequenzselektiver Filter gewünscht, deren Amplitudenspektren rechteckförmig sind und

entweder Nullphase nach (221.42) oder ein lineares Phasenspektrum nach (221.35) aufweisen. Die wichtigsten Vertreter dieser Idealfilter sind der Tiefpaß, der Hochpaß, der Bandpaß und die Bandsperre. Ihre reellen Amplitudenspektren $A(\Omega)$ sind in Abbildung 222-1 für $0 \leq |\Omega| \leq \pi$ dargestellt. Dabei bedeuten Ω_m , Ω_u und Ω_o vorgegebene Frequenzen.



D = Durchlaßbereich

S = Sperrbereich

Abb. 222-1: Frequenzselektive Filter

Sowohl der Bandpaß als auch die Bandsperre können durch eine Kaskadenschaltung gemäß (221.16) realisiert werden, indem Tief- und Hochpaßfilter hintereinandergeschaltet werden.

Es wird nun anhand von Abbildung 222-2 ein Beispiel zur Formulierung und Lösung einer Entwurfsaufgabe vorgestellt, wobei $f(x) = A(\Omega)$ gesetzt wird.

Beispiel 1: Als Idealfunktion (222.1) sei das reelle Amplitudenspektrum $A(\Omega)$ des

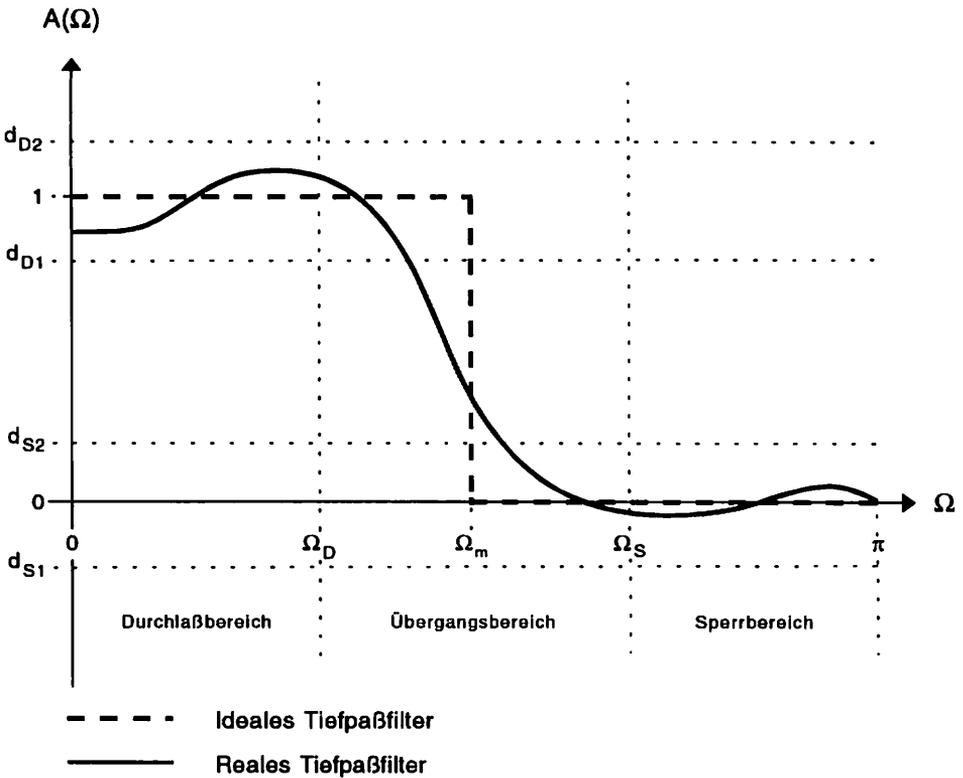


Abb. 222-2: Ideales und reales Tiefpaßfilter

idealen Tiefpaßfilters (216.17) gegeben

$$\text{Durchlaßbereich:} \quad A(\Omega) = 1 \quad \text{für} \quad 0 \leq |\Omega| \leq \Omega_m$$

$$\text{Sperrbereich:} \quad A(\Omega) = 0 \quad \text{für} \quad \Omega_m < |\Omega| \leq \pi \quad . \quad (222.8)$$

Die Lösung der Entwurfsaufgabe liefert als Realfunktion (222.2) das reelle Amplitudenspektrum $A_0(\Omega)$ des realen Tiefpaßfilters, für das nach Abbildung 222-2 gilt

$$\text{Durchlaßbereich:} \quad d_{D1} \leq A_0(\Omega) \leq d_{D2} \quad \text{für} \quad 0 \leq |\Omega| \leq \Omega_D$$

$$\text{Sperrbereich:} \quad d_{S1} \leq A_0(\Omega) \leq d_{S2} \quad \text{für} \quad \Omega_S < |\Omega| \leq \pi$$

$$\text{Übergangsbereich:} \quad d_{S1} \leq A_0(\Omega) \leq d_{D2} \quad \text{für} \quad \Omega_D < |\Omega| \leq \Omega_S \quad . \quad (222.9)$$

Die Modellparametermenge (222.3) setzt sich mit

$$M = \{d_{S1}, d_{S2}, d_{D1}, d_{D2}, \Omega_D, \Omega_S\} \quad (222.10)$$

aus den Toleranzgrößen d_{S1} , d_{S2} , d_{D1} und d_{D2} sowie der Durchlaßgrenzfrequenz Ω_D und der Sperrgrenzfrequenz Ω_S zusammen. ▲

223 Definition und Eigenschaften rekursiver Filter

Durch (221.4) wird ein kausales lineares digitales Filter definiert. Wie bereits im Kapitel 217 erläutert, unterscheidet man die rekursive und die nichtrekursive Filterung. Die erste Summe in (221.4) nennt man daher den nichtrekursiven Anteil und die zweite Summe den rekursiven Anteil, deren Obergrenze den Grad des rekursiven Filters bestimmt. Man bezeichnet also ein Filter als *kausales rekursives lineares digitales Filter N-ten Grades*, falls seine Filtergleichung

$$y(n) = \sum_{r=0}^M d_r x(n-r) - \sum_{k=1}^N g_k y(n-k) \quad (223.1)$$

mit

$$g_N \neq 0 \quad (223.2)$$

lautet. Das Filter ist nach (221.18) genau dann stabil, wenn für die Beträge der N Polstellen $z_{\infty k}$ mit $k \in \{1, \dots, N\}$ der Systemfunktion (221.3) die Bedingung $|z_{\infty k}| < 1$ gilt. Neben dem kausalen Filter wird auch das *antikausale rekursive lineare digitale Filter N-ten Grades* betrachtet. Seine Filtergleichung lautet nach (221.13)

$$y(n) = \sum_{r=0}^N \bar{d}_r x(n+r) - \sum_{k=1}^N \bar{g}_k y(n+k) \quad (223.3)$$

mit

$$\bar{g}_N \neq 0 . \quad (223.4)$$

Dieses Filter ist nach (221.18) genau dann stabil, falls für die Beträge der N Polstellen $|z_{\infty k}| > 1$ zutrifft.

Rekursive Filter sind nach (217.10) durch eine unendlich lange Impulsantwort charakterisiert. Da weiterhin im Falle des kausalen Filters die Bedingung (221.6) und im Fall des antikausalen Filters die Bedingung (221.14) gelten, können die Impulsantworten beider Filtertypen das symmetrische Verhalten in (221.40) und (221.44) sowie das antisymmetrische Verhalten in (221.47) und (221.51) nicht erfüllen. Kausale und antikausale rekursive Filter mit linearer Phase

$$\Phi(\Omega) = \alpha_1 \Omega + \alpha_2 \quad (223.5)$$

sind daher nicht realisierbar.

Es wird nun gezeigt, daß sich ein nullphasiges rekursives Filter konstruieren läßt, indem ein kausales und ein antikausales rekursives Filter nach (221.16) hintereinander geschaltet werden. Die Frequenzantwort eines nullphasigen Filters ist nach (221.43) reell

$$H(j\Omega) = A(\Omega) \quad (223.6)$$

Bezeichnet man die Frequenzantwort des kausalen Filters mit $H_1(j\Omega)$ und führt als weitere Frequenzantwort $H_2(j\Omega)$ ein, so folgt für (223.6) mit (221.16) die Bedingung

$$A(\Omega) = H_1(j\Omega)H_2(j\Omega) \quad (223.7)$$

die nach (221.31) beispielsweise für

$$H_2(j\Omega) = H_1^*(j\Omega) = H_1(-j\Omega) \quad (223.8)$$

erfüllt ist. Die Frequenzantwort $H_1(j\Omega)$ ergibt sich aus (221.2) für $z = e^{j\Omega}$ und $M = N$ zu

$$H_1(j\Omega) = \sum_{r=0}^N d_r e^{-jr\Omega} / \sum_{k=0}^N g_k e^{-jk\Omega} \quad (223.9)$$

und folglich $H_2(j\Omega)$ aus (223.8) zu

$$H_2(j\Omega) = \sum_{s=0}^N d_s e^{js\Omega} / \sum_{l=0}^N g_l e^{jl\Omega} \quad (223.10)$$

Substituiert man in (221.11) $M = N$, $z = e^{j\Omega}$ und

$$\bar{d}_r = d_r \quad \text{und} \quad \bar{g}_k = g_k \quad (223.11)$$

so erhält man (223.10). Mit $H_2(j\Omega)$ ist also die Frequenzantwort eines antikausalen Filters gegeben.

Für die nullphasige Filterung im Zeitbereich wird das Eingangssignal

$$x(n) = y_0(n) \quad (223.12)$$

zunächst vorwärtsgefiltert. Nach (223.1) erhält man das Ausgangssignal

$$y_1(n) = \sum_{r=0}^N d_r y_0(n-r) - \sum_{k=1}^N g_k y_1(n-k) \quad (223.13)$$

das anschließend rückwärtsgefiltert wird. Es ergibt sich das Ausgangssignal

$$y(n) = y_2(n) \quad (223.14)$$

mit (223.3) und (223.11) zu

$$y_2(n) = \sum_{s=0}^N d_s y_1(n+s) - \sum_{l=0}^N g_l y_2(n+l) \quad (223.15)$$

Der Filtervorgang besteht also in der sequentiellen Anwendung der Filtergleichungen

(223.13) und (223.15). Die Reihenfolge ist unerheblich, das Eingangssignal (223.12) muß jedoch vollständig gegeben sein.

Wird die nullphasige Filterung im Frequenzbereich durchgeführt, erhält man mit der z-Transformation $Y_0(z)$ des Eingangssignals (223.12) die z-Transformation $Y_2(z)$ des Ausgangssignals (223.14) aus (221.15) zu

$$\begin{aligned} Y_2(z) &= H_1(z)H_2(z)Y_0(z) \\ &= H(z)Y_0(z) \end{aligned} \quad (223.16)$$

Es soll nun die Systemfunktion $H(z)$ mit $z = e^{j\Omega}$, also die Frequenzantwort (223.6) berechnet werden. Dies ist deswegen sinnvoll, weil sowohl $H_1(j\Omega)$ als auch $H_2(j\Omega)$ Funktionen der Filterkoeffizienten d_r und g_k sind und sich daher ein einfacher Ausdruck für $H(j\Omega)$ finden läßt. Multipliziert man zunächst die Zähler von (223.9) und (223.10), ergibt sich

$$\left(\sum_{r=0}^N d_r e^{-jr\Omega} \right) \left(\sum_{s=0}^N d_s e^{js\Omega} \right) = \sum_{r=0}^N \sum_{s=0}^N d_r d_s e^{-j(r-s)\Omega} .$$

Mit der entsprechenden Beziehung für den Nenner folgt nach Substitution von $r-s = m$ und $k-l = n$

$$H(j\Omega) = \sum_{m=-N}^N a_m e^{-jm\Omega} / \sum_{n=-N}^N b_n e^{-jn\Omega} \quad (223.17)$$

mit den Koeffizienten

$$a_m = \sum_{r=0}^N \sum_{\substack{s=0 \\ r-s=m}}^N d_r d_s = a_{-m} , \quad (223.18)$$

$$b_n = \sum_{k=0}^N \sum_{\substack{l=0 \\ k-l=n}}^N g_k g_l = b_{-n} . \quad (223.19)$$

Die Koeffizienten a_m und b_n sind also symmetrisch bezüglich $m = 0$ und $n = 0$. Berücksichtigt man diese Symmetrien in (223.17), erhält man

$$H(j\Omega) = \left(a_0 + \sum_{k=0}^N a_k (e^{-jk\Omega} + e^{jk\Omega}) \right) / \left(b_0 + \sum_{l=1}^N b_l (e^{-jl\Omega} + e^{jl\Omega}) \right) ,$$

und mit (212.11)

$$H(j\Omega) = \left(a_0 + 2 \sum_{k=0}^N a_k \cos k\Omega \right) / \left(b_0 + 2 \sum_{l=0}^N b_l \cos l\Omega \right) . \quad (223.20)$$

Schließlich ergibt sich mit den Substitutionen

$$\bar{a}_0 = a_0/b_0 \quad \text{und} \quad \bar{b}_0 = 1$$

$$\bar{a}_k = 2a_k/b_0 \quad \text{und} \quad \bar{b}_1 = 2b_1/b_0 \quad \text{für} \quad k, l \in \{1, \dots, N\} \quad (223.21)$$

die kompakte Darstellung

$$H(j\Omega) = \frac{\sum_{k=0}^N \bar{a}_k \cos k\Omega}{\sum_{l=0}^N \bar{b}_l \cos l\Omega} . \quad (223.22)$$

Die Frequenzantwort ist bezüglich $\Omega = 0$ symmetrisch.

224 Entwurf rekursiver Filter

Die im Kapitel 222 allgemein formulierte Entwurfsaufgabe wird nun auf rekursive digitale Filter angewendet. Eine Lösung gewinnt man mit der Wahl einer analytischen Funktion als Realfunktion für das Amplitudenspektrum eines Filters. Die Pol- und Nullstellen der Funktion sind zu ermitteln, um daraus die Filterkoeffizienten zu bestimmen. Dieser Weg wird im folgenden am Beispiel des Butterworthfilters aufgezeigt. Anschließend wird auf ein allgemein anwendbares Verfahren eingegangen, bei dem die Filterkoeffizienten geschätzt werden, indem einem idealen Amplitudenspektrum eine Realfunktion angepaßt wird. Für weitere Verfahren sei auf die Literatur (zum Beispiel Rabiner und Gold 1975, S.210) verwiesen.

Die Frequenzantwort eines nullphasigen rekursiven Filters ist durch (223.6) mit (223.7) und (223.8) definiert und lautet unter Berücksichtigung von (213.14)

$$H(j\Omega) = A(\Omega) = H_1(j\Omega)H_1(-j\Omega) = |H_1(j\Omega)|^2 = (A_1(\Omega))^2 , \quad (224.1)$$

falls $A_1(\Omega)$ die reelle Amplitude von $H_1(j\Omega)$ bedeutet. In dieser Form wird $H(j\Omega)$ auch als *Leistungsverstärkung* bezeichnet (Stearns 1984, S.237). Aus dem Ansatz

$$(A_1(\Omega))^2 = (1 + (C_q(\Omega))^2)^{-1} \quad (224.2)$$

folgt mit

$$C_q(\Omega) = \frac{\tan^q(\Omega/2)}{\tan^q(\Omega_c/2)} \quad (224.3)$$

die Leistungsverstärkung des sogenannten *Butterworthfilters* zu

$$|H_1(j\Omega)|^2 = \left(1 + \frac{\tan^{2q}(\Omega/2)}{\tan^{2q}(\Omega_c/2)}\right)^{-1} \quad (224.4)$$

(Koch 1975; Rabiner und Gold 1975, S.264), worin Ω_c und $q \in \{0, 1, 2, \dots\}$ durch die Modellparameter aus (222.3) bestimmt werden. Mit wachsendem q nimmt (224.4) die Form einer Treppenfunktion an und ist somit zur Approximation eines Tiefpaßfilters geeignet. Soll hingegen ein Hochpaßfilter durch ein Butterworthfilter realisiert werden, sind in (224.4) Ω und Ω_c zu vertauschen.

Mit (224.4) kann die Entwurfsaufgabe aus Kapitel 222 im Fall eines Tiefpaßfilters wie folgt formuliert werden:

Gegeben sei die Idealfunktion $f(x) = H(j\Omega) = A(\Omega)$ des idealen Tiefpaßfilters, für die nach (222.8) gilt

$$\text{Durchlaßbereich: } A(\Omega) = 1 \quad \text{für } 0 \leq |\Omega| \leq \Omega_m$$

$$\text{Sperrbereich: } A(\Omega) = 0 \quad \text{für } \Omega_m < |\Omega| \leq \pi,$$

sowie die Modellparametermenge $M = \{d_{S1}, d_{S2}, d_{D1}, d_{D2}, \Omega_D, \Omega_S\}$, deren Elemente durch

$$d_{S1} = 0, \quad d_{S2} = \delta_S^{-2}, \quad d_{D1} = (1 + \delta_D^2)^{-1}, \quad d_{D2} = 1$$

definiert und in Abbildung 224-1 veranschaulicht sind.

Gesucht ist die Realfunktion $f_0(x) = H_0(j\Omega) = A_0(\Omega) = |H_1(j\Omega)|^2$ gemäß (224.1) eines Butterworthfilters, für die nach (222.9) gelten muß

$$\text{Durchlaßbereich: } d_{D1} \leq A_0(\Omega) \leq d_{D2} \quad \text{für } 0 \leq |\Omega| \leq \Omega_D$$

$$\text{Sperrbereich: } d_{S1} \leq A_0(\Omega) \leq d_{S2} \quad \text{für } \Omega_S < |\Omega| \leq \pi$$

$$\text{Übergangsbereich: } d_{S2} \leq A_0(\Omega) \leq d_{D1} \quad \text{für } \Omega_D < |\Omega| \leq \Omega_S. \quad (224.5)$$

Zunächst ergeben sich nach Abbildung 224-1 aus (224.4) mit $A_0(\Omega) = |H_1(j\Omega)|^2$ die Bedingungengleichungen

$$A_0(\Omega_D) = \left(1 + \frac{\tan^{2q}(\Omega_D/2)}{\tan^{2q}(\Omega_c/2)}\right)^{-1} = d_{D1} = \frac{1}{1 + \delta_D^2},$$

$$A_0(\Omega_S) = \left(1 + \frac{\tan^{2q}(\Omega_S/2)}{\tan^{2q}(\Omega_c/2)}\right)^{-1} = d_{S2} = \frac{1}{\delta_S^2}$$

oder umgeformt

$$\delta_D^2 = \frac{\tan^{2q}(\Omega_D/2)}{\tan^{2q}(\Omega_c/2)} \quad \text{und} \quad \delta_S^2 - 1 = \frac{\tan^{2q}(\Omega_S/2)}{\tan^{2q}(\Omega_c/2)} \quad (224.6)$$

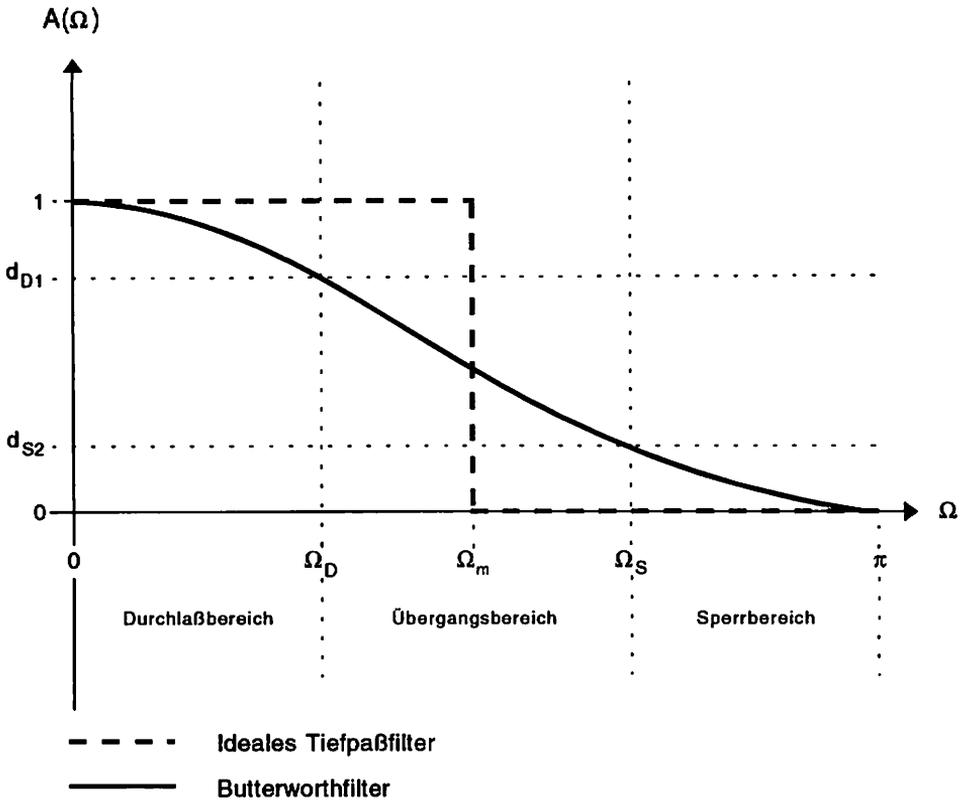


Abb. 224-1: Amplitudenspektrum eines Butterworthfilters

zur Berechnung von Ω_c und q (Hamming 1987, S.238). Wird die erste Gleichung in (224.6) durch die zweite dividiert, folgt

$$\frac{\delta_D}{(\delta_S^2 - 1)^{1/2}} = \left(\frac{\tan(\Omega_D/2)}{\tan(\Omega_S/2)} \right)^q$$

und somit

$$q = \frac{\log(\delta_D / (\delta_S^2 - 1)^{1/2})}{\log(\tan(\Omega_D/2) / \tan(\Omega_S/2))} \quad (224.7)$$

Weiter erhält man aus der ersten Gleichung von (224.6)

$$\Omega_c = 2 \arctan\left(\frac{\tan(\Omega_D/2)}{\delta_D^{-q}}\right) \quad (224.8)$$

Mit der vorgegebenen Modellparametermenge M sind die Parameter Ω_c und q aus (224.7)

und (224.8) zu berechnen. Die Realfunktion $A_0(\Omega)$ geht mit (224.4) und der aus (212.11) folgenden Beziehung

$$\tan \frac{\Omega}{2} = \frac{\sin \Omega/2}{\cos \Omega/2} = -j \frac{e^{j\Omega/2} - e^{-j\Omega/2}}{e^{j\Omega/2} + e^{-j\Omega/2}} = -j \frac{e^{j\Omega} - 1}{e^{j\Omega} + 1} = -j \frac{z-1}{z+1}$$

für $z = e^{j\Omega}$ mit $\rho = \tan(\Omega_c/2)$ über in

$$A_0(\Omega) = \frac{\rho^{2q}}{\rho^{2q} + (-1)^q ((z-1)/(z+1))^{2q}} \quad (224.9)$$

Weiter ergibt sich mit

$$p = (z-1)/(z+1) \quad (224.10)$$

woraus die Ausdrücke $z-pz = 1+p$ und

$$z = (1+p)/(1-p) \quad (224.11)$$

folgen, aus (224.9)

$$A_0(\Omega) = \frac{\rho^{2q}}{\rho^{2q} + (-1)^q p^{2q}} \quad (224.12)$$

Für die Berechnung der $2q$ Polstellen p_m von p in (224.12)

$$p_m = \rho e^{j\beta_m} = \rho(\cos\beta_m + j\sin\beta_m) \quad m \in \{0, 1, \dots, 2q-1\} \quad (224.13)$$

ist zu unterscheiden, ob q gerade oder ungerade ist. Durch Nullsetzen des Nenners in (224.12) folgt im ersten Fall $e^{j2q\beta_m} = -1$ und mit (212.9)

$$\beta_m = (2m+1)\pi/2q \quad (224.14)$$

Ist q hingegen ungerade, gilt $e^{j2q\beta_m} = 1$ und daher

$$\beta_m = m\pi/q \quad (224.15)$$

Im folgenden wird ausschließlich der Fall betrachtet, daß q eine gerade Zahl ist. Dann ergeben sich die Polstellen (224.13) mit (224.14) nach (221.24) als konjugiert komplexe Paare und zwar gilt für p_{k-1} und p_{2q-k} mit $k = \{1, \dots, q\}$ wegen (212.9)

$$\begin{aligned} p_{k-1} &= \rho e^{j(2k-1)\pi/2q} \\ p_{2q-k} &= \rho e^{j(2(2q-k)+1)\pi/2q} = \rho e^{-j(2k-1)\pi/2q} = p_{k-1}^* \end{aligned} \quad (224.16)$$

Weiterhin folgt

$$p_{k-1} = \rho e^{j(2k-1)\pi/2q} = -\rho e^{j(2(k+q)-1)\pi/2q} = -p_{k+q-1} \quad (224.17)$$

Es läßt sich somit feststellen, daß die Polstellen (224.13) mit (224.14) als konjugiert komplexe Paare auftreten und zudem bezüglich der imaginären Koordinatenachse spiegelbildlich auf einem Kreis mit dem Radius ρ angeordnet sind. Sie sind also gleichmäßig auf die vier Quadranten der komplexen Zahlenebene verteilt. Bezeichnet man die Polstellen im ersten Quadranten mit v_k und die im zweiten Quadranten mit w_k , so enthält der dritte Quadrant die konjugiert komplexen Polstellen w_k^* und der vierte Quadrant die konjugiert komplexen Polstellen v_k^* . Es ergeben sich dann unter Berücksichtigung von (224.16) und (224.17) die Zusammenhänge

$$v_k = p_{k-1}$$

$$w_k = p_{q-k}$$

$$w_k^* = p_{q+k-1} = -v_k$$

$$v_k^* = p_{2q-k} = -w_k$$

für $k = \{1, \dots, q/2\}$, die in Abbildung 224-2 verdeutlicht sind.

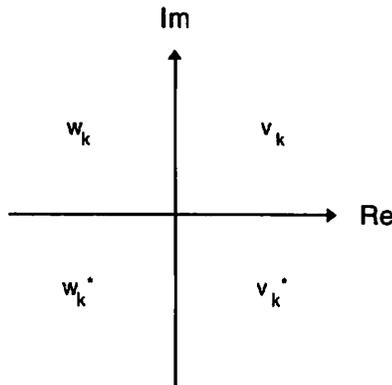


Abb. 224-2: Polstellenverteilung mit $k \in \{1, \dots, q/2\}$

Mit diesen Polstellen kann (224.12) in der Form

$$A_0(\Omega) = \prod_{k=1}^{q/2} \frac{\rho^2}{(p-v_k)(p-v_k^*)} \prod_{k=1}^{q/2} \frac{\rho^2}{(p-w_k)(p-w_k^*)} \quad (224.18)$$

geschrieben werden. Bringt man hierin die aus (224.10) folgende Beziehung

$$p-q_k = \frac{(1-q_k)z - (1+q_k)}{z+1} \quad (224.19)$$

für $q_k \in \{v_k, v_k^*, w_k, w_k^*\}$ ein, ergibt sich schließlich die Realfunktion $A_0(\Omega)$ zu

$$A_0(\Omega) = \prod_{k=1}^{q/2} \frac{e_k(z+1)^2}{(z-t_k)(z-t_k^*)} \prod_{k=1}^{q/2} \frac{c_k(z+1)^2}{(z-s_k)(z-s_k^*)} \quad (224.20)$$

mit

$$c_k = \frac{\rho^2}{(1-v_k)(1-v_k^*)} = \frac{\rho^2}{1+\rho^2-2\rho\cos\beta_{k-1}}, \quad (224.21)$$

$$e_k = \frac{\rho^2}{(1-w_k)(1-w_k^*)} = \frac{\rho^2}{1+\rho^2+2\rho\cos\beta_{k-1}}, \quad (224.22)$$

$$s_k = \frac{1+v_k}{1-v_k} = \frac{1-\rho^2+2j\rho\sin\beta_{k-1}}{1+\rho^2-2\rho\cos\beta_{k-1}}, \quad (224.23)$$

$$s_k^* = \frac{1+v_k^*}{1-v_k^*} = \frac{1-\rho^2-2j\rho\sin\beta_{k-1}}{1+\rho^2-2\rho\cos\beta_{k-1}}, \quad (224.24)$$

$$t_k = \frac{1+w_k}{1-w_k} = \frac{1-v_k^*}{1+v_k^*} = \frac{1-\rho^2+2j\rho\sin\beta_{k-1}}{1+\rho^2+2\rho\cos\beta_{k-1}} = \frac{1}{s_k^*}, \quad (224.25)$$

$$t_k^* = \frac{1+w_k^*}{1-w_k^*} = \frac{1-v_k}{1+v_k} = \frac{1-\rho^2-2j\rho\sin\beta_{k-1}}{1+\rho^2+2\rho\cos\beta_{k-1}} = \frac{1}{s_k}. \quad (224.26)$$

Sie liegt nach (221.16) wegen $A_0(\Omega) = H_0(j\Omega)$ in der Kaskadenform vor

$$A_0(\Omega) = H_1(j\Omega)H_2(j\Omega) = \prod_{k=1}^{q/2} H_{1k}(j\Omega) \prod_{k=1}^{q/2} H_{2k}(j\Omega) = \prod_{k=1}^{q/2} H_{0k}(j\Omega) \quad (224.27)$$

mit den Teilsystemfunktionen

$$H_{1k}(z) = \frac{e_k(z+1)^2}{(z-t_k)(z-t_k^*)} \quad \text{und} \quad H_{2k}(z) = \frac{c_k(z+1)^2}{(z-s_k)(z-s_k^*)} \quad (224.28)$$

sowie

$$H_{0k}(j\Omega) = A_{0k}(\Omega) = H_{1k}(j\Omega) H_{2k}(j\Omega).$$

Erweitert man $H_{2k}(j\Omega)$ mit z^{-2}/z^{-2} , ergibt sich mit $e_k = c_k/(s_k s_k^*)$

$$\begin{aligned} H_{2k}(z) &= \frac{c_k(z+1)^2 z^{-2}}{(z-s_k)(z-s_k^*)z^{-2}} = \frac{c_k(z^{-1}+1)^2}{(1-s_k/z)(1-s_k^*/z)} \\ &= \frac{c_k(z^{-1}+1)^2}{s_k s_k^* (z^{-1}-1/s_k)(z^{-1}-1/s_k^*)} = \frac{e_k(z^{-1}+1)^2}{(z^{-1}-t_k^*)(z^{-1}-t_k)} \\ &= H_{1k}(z^{-1}) \end{aligned} \quad (224.29)$$

oder für $z = e^{j\Omega}$ und $z^{-1} = e^{-j\Omega}$

$$H_{2k}(j\Omega) = H_{1k}(-j\Omega) .$$

Dieses Ergebnis stimmt mit (223.8) überein, so daß $H_{2k}(j\Omega)$ die Frequenzantwort eines antikausalen Systems darstellt.

Die Systemfunktion $H_{1k}(z)$ läßt sich mit $N = 2$ sowie $d_i = d_{ik}$ und $g_i = g_{ik}$ für $i \in \{0, 1, 2\}$ in der Form (223.9) der Systemfunktion eines kausalen rekursiven Filters schreiben

$$H_{1k}(z) = \frac{d_{0k} + d_{1k}z^{-1} + d_{2k}z^{-2}}{g_{0k} + g_{1k}z^{-1} + g_{2k}z^{-2}} . \quad (224.30)$$

Stellt man diesem Ausdruck die aus (224.28) nach Erweiterung um z^{-2}/z^{-2} folgende Beziehung

$$H_{1k}(z) = \frac{e_k(z^2 + 2z + 1)z^{-2}}{(z - t_k)(z - t_k^*)z^{-2}}$$

gegenüber, liefert der Vergleich die Filterkoeffizienten d_{ik} und g_{ik} zu

$$d_{0k} = e_k , \quad d_{1k} = 2e_k , \quad d_{2k} = e_k , \quad g_{0k} = 1 ,$$

$$g_{1k} = -(t_k + t_k^*) = - \frac{2(1 - \rho^2)}{1 + \rho^2 + 2\rho \cos \beta_{k-1}}$$

$$g_{2k} = t_k t_k^* = \frac{(1 - \rho^2)^2 + 4\rho^2 \sin^2 \beta_{k-1}}{(1 + \rho^2 + 2\rho \cos \beta_{k-1})^2} . \quad (224.31)$$

Die Systemfunktion $H_{2k}(z)$ lautet wegen (224.29)

$$H_{2k}(j\Omega) = H_{1k}(z^{-1}) = \frac{d_{0k} + d_{1k}z + d_{2k}z^2}{g_{0k} + g_{1k}z + g_{2k}z^2} . \quad (224.32)$$

Die Polstellen von $H_{1k}(z)$ liegen bei $z = t_k$ und $z = t_k^*$, da dann der Nenner verschwindet. Nach Satz (221.18) ist das kausale System mit der Systemfunktion $H_{1k}(z)$ genau dann stabil, wenn $|t_k| = |t_k^*| < 1$ gilt. Die Polstellen von $H_{2k}(z)$ liegen nach (224.28) bei $z = s_k$ und $z = s_k^*$. Das antikausale System mit der Systemfunktion $H_{2k}(z)$ ist dann stabil, wenn $|s_k| = |s_k^*| > 1$ zutrifft. Bildet man die Beträge von (224.23) und (224.25), erhält man

$$|s_k| = (s_k s_k^*)^{1/2} = \frac{((1 - \rho^2)^2 + 4\rho^2 \sin^2 \beta_{k-1})^{1/2}}{1 + \rho^2 - 2\rho \cos \beta_{k-1}} ,$$

$$|t_k| = (t_k t_k^*)^{1/2} = \frac{((1 - \rho^2)^2 + 4\rho^2 \sin^2 \beta_{k-1})^{1/2}}{1 + \rho^2 + 2\rho \cos \beta_{k-1}} .$$

Für $0 \leq \beta_{k-1} < \pi/2$ und wegen $\rho > 0$ gilt die Ungleichung $|s_k| > |t_k|$, woraus wegen $|s_k| = 1/|t_k|$ schließlich $|t_k| < 1$ und $|s_k| > 1$ folgen. Somit liegen mit $H_{1k}(j\Omega)$ und $H_{2k}(j\Omega)$ Frequenzantworten stabiler Systeme vor.

Es sind nun alle q Teilsystemfunktionen in (224.27) bestimmt. Um die Realfunktion $A_0(\Omega)$ anzugeben, werden die reellen Teilsystemfunktionen $H_{0k}(j\Omega) = A_{0k}(\Omega) = H_{1k}(j\Omega)H_{2k}(j\Omega)$ in (224.27) berechnet. Multipliziert man (224.30) mit (224.32), erhält man nach (223.22) mit $N = 2$

$$H_{0k}(j\Omega) = A_{0k}(\Omega) = \frac{\left(\sum_{i=0}^2 d_{ik}z^{-i}\right)\left(\sum_{i=0}^2 d_{ik}z^i\right)}{\left(\sum_{i=0}^2 g_{ik}z^{-i}\right)\left(\sum_{i=0}^2 g_{ik}z^i\right)} = \frac{\sum_{i=0}^2 \bar{a}_{ik}\cos i\Omega}{\sum_{i=0}^2 \bar{b}_{ik}\cos i\Omega} \quad (224.33)$$

Hierin lauten die aus (223.21) mit (223.18) und (223.19) folgenden Koeffizienten $\bar{a}_i = \bar{a}_{ik}$ und $\bar{b}_i = \bar{b}_{ik}$ für $i \in \{0, 1, 2\}$ mit $b_{0k} = g_{0k}^2 + g_{1k}^2 + g_{2k}^2$

$$\begin{aligned} \bar{a}_{0k} &= (d_{0k}^2 + d_{1k}^2 + d_{2k}^2) / b_{0k} & \bar{b}_{0k} &= 1 \\ \bar{a}_{1k} &= 2(d_{0k}d_{1k} + d_{1k}d_{2k}) / b_{0k} & \bar{b}_{1k} &= 2(g_{0k}g_{1k} + g_{1k}g_{2k}) / b_{0k} \\ \bar{a}_{2k} &= 2(d_{0k}d_{2k}) / b_{0k} & \bar{b}_{2k} &= 2(g_{0k}g_{2k}) / b_{0k} \end{aligned} \quad (224.34)$$

Durch die Hintereinanderschaltung der $q/2$ Teilfilter mit den Teilfrequenzantworten (224.33) für $k \in \{1, \dots, q/2\}$ nach (224.27) erhält man die Frequenzantwort $A_0(\Omega)$ des Butterworthfilters, mit der im Frequenzbereich gefiltert werden kann. Soll die Filterung im Zeitbereich erfolgen, sind die (224.33) zugeordnete Vorwärtsfilterung (223.13) und Rückwärtsfilterung (223.15) mit den Filterkoeffizienten (224.31) anzuwenden. Als Ergebnis erhält man das Ausgangssignal eines Teilfilters. Nach (224.27) sind $q/2$ Teilfilter hintereinanderzuschalten. Das Ausgangssignal des 1. Teilfilters ist also Eingangssignal des folgenden und so fort, wie im Zusammenhang mit (221.16) erläutert.

Im folgenden werden die beschriebenen Zusammenhänge an einem einfachen Beispiel verdeutlicht.

Beispiel 1: Gegeben seien die Idealfunktion $f(x) = H(j\Omega) = A(\Omega)$ mit

$$\text{Durchlaßbereich:} \quad A(\Omega) = 1 \quad \text{für} \quad 0 \leq |\Omega| \leq \Omega_m$$

$$\text{Sperrbereich:} \quad A(\Omega) = 0 \quad \text{für} \quad \Omega_m < |\Omega| \leq \pi$$

und die Modellparametermenge M aus (222.10) mit den Spezifikationen

$$d_{S1} = 0 \quad ; \quad d_{S2} = 0,1975 \quad ; \quad d_{D1} = 0,8 \quad ; \quad d_{D2} = 1 \quad ; \quad \Omega_D = 0,2\pi \quad ; \quad \Omega_S = 0,4\pi \quad .$$

Gesucht ist die Realfunktion $f_0(x) = H_0(j\Omega) = A_0(\Omega)$ gemäß (224.27). Die Berechnung der natürlichen Zahl q aus (224.7) ergibt $q = 2$ und somit folgt aus (224.8) $\Omega_c = 0,274 \pi$. Weiter erhält man $\rho = \tan(\Omega_c/2) = 0,4595$ und aus (224.13) mit (224.14) unter Berücksichtigung der Aufteilung aus Abbildung 224-2 die Polstellen

$$p_0 = 0,3249 + j 0,3249 = v_1$$

$$p_1 = -0,3249 + j 0,3249 = w_1$$

$$p_2 = -0,3249 - j 0,3249 = w_1^*$$

$$p_3 = 0,3249 - j 0,3249 = v_1^* .$$

Aus (224.23) bis (224.26) ergibt sich weiter

$$s_1 = 1,4053 - j 1,1577$$

$$s_1^* = 1,4053 + j 1,1577$$

$$t_1 = 0,4239 - j 0,3492$$

$$t_1^* = 0,4239 + j 0,3492$$

sowie für die Beträge

$$|s_1| = |s_1^*| = 1,8208$$

$$|t_1| = |t_1^*| = 0,5492 .$$

Man erhält daher nach (224.30) und (224.32) mit (224.31) für $e_1 = 0,1135$ die Systemfunktionen

$$H_{11}(z) = \frac{d_{01} + d_{11}z^{-1} + d_{21}z^{-2}}{1 + g_{11}z^{-1} + g_{21}z^{-2}} \quad \text{und} \quad H_{21}(z) = \frac{d_{01} + d_{11}z + d_{21}z^2}{1 + g_{11}z + g_{21}z^2}$$

mit den Filterkoeffizienten

$$d_{01} = 0,1135 ; \quad d_{11} = 0,2270 ; \quad d_{21} = 0,1135$$

$$g_{01} = 1 \quad ; \quad g_{11} = -0,8478 ; \quad g_{21} = 0,3016 . \quad (224.35)$$

Weiter liefert (224.27) mit (224.33) wegen $q = 2$ die gesuchte Realfunktion $A_0(\Omega)$ zu

$$A_0(\Omega) = \frac{\bar{a}_{01} + \bar{a}_{11}\cos\Omega + \bar{a}_{21}\cos 2\Omega}{\bar{b}_{01} + \bar{b}_{11}\cos\Omega + \bar{b}_{21}\cos 2\Omega} .$$

Die darin enthaltenen Koeffizienten \bar{a}_{k1} und \bar{b}_{k1} mit $k \in \{0, 1, 2\}$ lauten mit $b_{01} = 1,8097$

$$\bar{a}_{01} = 4,271 \cdot 10^{-2} ; \quad \bar{a}_{11} = 5,695 \cdot 10^{-2} ; \quad \bar{a}_{21} = 1,424 \cdot 10^{-2}$$

$$\bar{b}_{01} = 1 ; \quad \bar{b}_{11} = -1,2195 ; \quad \bar{b}_{21} = 0,3333 . \quad (224.36)$$

Mit $A_0(\Omega)$ kann im Frequenzbereich gefiltert werden. Die Filtergleichungen im Zeitbereich ergeben sich aus (223.13) und (223.15) mit den Filterkoeffizienten aus (224.35) zu

$$y_1(n) = 0,1135y_0(n) + 0,2270y_0(n-1) + 0,1135y_0(n-2) + 0,8478y_1(n-1) - 0,3016y_1(n-2)$$

$$y_2(n) = 0,1135y_1(n) + 0,2270y_1(n+1) + 0,1135y_1(n+2) + 0,8478y_2(n+1) - 0,3016y_2(n+2).$$

Man berechnet zunächst beispielsweise $y_1(n)$ für alle n in Richtung wachsender Werte für n und anschließend $y_2(n)$ für alle n in der entgegengesetzten Richtung. Wegen $q = 2$ ist nur eine Vorwärts- und Rückwärtsfilterung erforderlich. Δ

Die unbekannt Filterkoeffizienten können auch aus einer Parameterschätzung ermittelt werden. Dieses Verfahren ist allgemein anwendbar und wurde für den Entwurf rekursiver Filter ausführlich bei (Schmidt 1993a,b) behandelt. Wählt man als Idealfunktion wieder das Amplitudenspektrum $A(\Omega)$, erhält man aus (222.7) mit $x = \Omega$ die Fehlerfunktion

$$e(\Omega) = A_0(\Omega, \beta) - A(\Omega) \quad (224.37)$$

Die Realfunktion $A_0(\Omega, \beta)$ sei beispielsweise durch die Frequenzantwort (223.22) des nullphasigen rekursiven Filters gegeben. Anstelle der unbekannt Filterkoeffizienten d_r und g_k nach (222.4) faßt man dann mit $u = 2N+1$ die Koeffizienten \bar{a}_k und \bar{b}_k aus (223.22) im $u \times 1$ Vektor

$$\beta = [\bar{a}_0, \bar{a}_1, \dots, \bar{a}_N, \bar{b}_1, \dots, \bar{b}_N]' \quad (224.38)$$

zusammen und spaltet ihn mit

$$\beta = \beta_0 + \delta\beta \quad (224.39)$$

in den Vektor der Näherungswerte β_0 und den Vektor der unbekannt Zuschläge $\delta\beta$ auf. Man erhält dann für (224.37) nach einer Linearisierung

$$e(\Omega) = A_0(\Omega, \beta_0) - A(\Omega) + \mathbf{x}(\Omega)' \delta\beta \quad (224.40)$$

mit

$$\mathbf{x}(\Omega) = (\partial A_0 / \partial \beta) \Big|_{\beta = \beta_0}.$$

Nach der Diskretisierung von Ω in die n Werte Ω_i mit $i \in \{1, 2, \dots, n\}$ und $n > u$ ergibt sich mit $e_i = e(\Omega_i)$, $y_i = A(\Omega_i) - A_0(\Omega_i, \beta_0)$ und $\mathbf{x}_i = \mathbf{x}(\Omega_i)$ aus (224.40)

$$y_i + e_i = \mathbf{x}_i' \delta\beta \quad (224.41)$$

und mit $\mathbf{e} = (e_i)$, $\mathbf{y} = (y_i)$ und $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]'$ das lineare Modell für die Parame-

terschätzung

$$\mathbf{y} + \mathbf{e} = \mathbf{X} \delta \boldsymbol{\beta} . \quad (224.42)$$

Wird ferner mit der $n \times n$ Matrix \mathbf{P} , die als positiv definit vorausgesetzt wird (Koch 1987, S.179), eine Gewichtung vorgenommen, liefert die beste Schätzung, die Methode der kleinsten Quadrate und die Maximum-Likelihood-Schätzung die Schätzwerte

$$\hat{\delta \boldsymbol{\beta}} = (\mathbf{X}' \mathbf{P} \mathbf{X})^{-1} \mathbf{X}' \mathbf{P} \mathbf{y} \quad (224.43)$$

und schließlich mit

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + \hat{\delta \boldsymbol{\beta}} \quad (224.44)$$

die gesuchte Realfunktion $A_0(\Omega, \hat{\boldsymbol{\beta}})$, mit der im Frequenzbereich zu filtern ist. Anschließend muß überprüft werden, ob die Realfunktion das reelle Amplitudenspektrum eines stabilen Systems repräsentiert. Sollte dies nicht der Fall sein, sind für die Koeffizienten aus (224.38) Restriktionen in die Parameterschätzung einzuführen.

225 Definition und Eigenschaften nichtrekursiver Filter

Wie aus den Erläuterungen zu Beginn des Kapitels 223 folgt, bezeichnet man ein Filter als *kausales nichtrekursives digitales Filter M-ten Grades*, falls seine Filtergleichung

$$y(n) = \sum_{r=0}^M d_r x(n-r) \quad (225.1)$$

mit

$$d_M \neq 0 \quad (225.2)$$

lautet. Dieses nichtrekursive Filter kann mit $g_k = 0$ für $k \in \{1, \dots, N\}$ als Spezialfall des rekursiven Filters mit der Filtergleichung (223.1) interpretiert werden, so daß die im Kapitel 223 zusammengestellten Definitionen und Eigenschaften rekursiver Filter auch für die nichtrekursiven Filter gelten. In Anlehnung an die in der Literatur übliche Darstellungsweise wird die Obergrenze der Summe in (225.1) durch $M = N-1$ substituiert. Die Filtergleichung des nichtrekursiven Filters (N-1)-ten Grades lautet also

$$y(n) = \sum_{k=0}^{N-1} d_k x(n-k) . \quad (225.3)$$

Die Impulsantwort $h(k)$ erhält man aus (217.12) zu

$$h(k) = d_k \quad \text{für } k \in \{0, \dots, N-1\} \quad (225.4)$$

und die Systemfunktion $H(z)$ aus (217.11) zu

$$H(z) = \sum_{k=0}^{N-1} d_k z^{-k} . \quad (225.5)$$

Stellt man sie in der Produktform (221.12) dar

$$H(z) = d_0 \prod_{k=1}^{N-1} \frac{z - z_{0k}}{z - z_{\infty k}} \quad (225.6)$$

und multipliziert Zähler und Nenner aus, folgt aus dem Vergleich mit (225.5) für die Polstellen

$$z_{\infty k} = 0 \quad \text{für } k \in \{1, \dots, N-1\} , \quad (225.7)$$

so daß nach (221.18) ein kausales nichtrekursives Filter stets stabil ist.

In den folgenden Abschnitten a) bis d) werden die nichtrekursiven Filter mit linearer Phase

$$\Phi(\Omega) = \alpha_1 \Omega + \alpha_2 \quad (225.8)$$

nach (221.35) behandelt. Betrachtet man zunächst den Fall

a) α_1 beliebig und $\alpha_2 = 0$

aus (221.38), so muß für die Impulsantwort nach (221.40)

$$h(n - \alpha_1) = h(-n - \alpha_1) \quad (225.9)$$

gelten. Die Impulsantwort ist also symmetrisch bezüglich des Wertes $-\alpha_1$. Da nach (225.4) weiter $h(k) \neq 0$ für $k \in \{0, \dots, N-1\}$ gilt, muß der Symmetriepunkt bei

$$\alpha_1 = -(N-1)/2 \quad (225.10)$$

liegen. Substituiert man in (225.9) $k = n - \alpha_1$ und $n = k + \alpha_1$, folgt mit (225.10)

$$h(k) = h(N-1-k) \quad \text{für } k \in \{0, \dots, N-1\} . \quad (225.11)$$

Bei Berücksichtigung dieser Symmetriebedingung erhält man die Filtergleichung aus (225.3) mit (225.4).

Zur Berechnung der Frequenzantwort ist es sinnvoll, zu unterscheiden, ob N eine gerade oder ungerade Zahl ist.

aa) N gerade Zahl

Für die Frequenzantwort $H(j\Omega)$ gilt nach (221.30) und (225.5) mit (225.8)

$$H(j\Omega) = \sum_{k=0}^{N-1} h(k) e^{-jk\Omega} = A(\Omega) e^{j\alpha_1 \Omega} , \quad (225.12)$$

so daß die Amplitude

$$\begin{aligned}
 A(\Omega) &= \sum_{k=0}^{N-1} h(k)e^{-j(k+\alpha_1)\Omega} \\
 &= \sum_{k=0}^{N/2-1} h(k)e^{-j(k+\alpha_1)\Omega} + \sum_{k=N/2}^{N-1} h(k)e^{-j(k+\alpha_1)\Omega}
 \end{aligned} \quad (225.13)$$

folgt. Wird in der zweiten Summe $k = N-1-n = -2\alpha_1-n$ und $n = N-1-k$ substituiert, erhält man mit (212.11), (225.4) und (225.11)

$$\begin{aligned}
 A(\Omega) &= \sum_{k=0}^{N/2-1} h(k)e^{-j(k+\alpha_1)\Omega} + \sum_{n=0}^{N/2-1} h(N-1-n)e^{j(n+\alpha_1)\Omega} \\
 &= 2 \sum_{k=0}^{N/2-1} d_k \cos(k+\alpha_1)\Omega .
 \end{aligned} \quad (225.14)$$

Mit den Substitutionen

$$\bar{d}_m = 2d_{N/2-m} \quad \text{für } m \in \{1, \dots, N/2\} \quad (225.15)$$

kann (225.14) in die kompaktere Darstellung

$$A(\Omega) = \sum_{k=1}^{N/2} \bar{d}_k \cos(k-1/2)\Omega \quad (225.16)$$

überführt werden. Setzt man (225.16) in (225.12) ein, ergibt sich die Frequenzantwort $H(j\Omega)$ des linearphasigen Filters bei geradzahligem N .

Für $\Omega = \pi$ verschwindet nach (225.16) die Amplitude und damit auch die Frequenzantwort. Nach Abbildung 222-1 sind daher Hochpässe und Bandsperren durch ein linearphasiges Filter mit geradzahligem N nicht realisierbar.

ab) N ungerade Zahl

Anstelle von (225.13) erhält man für die Amplitude $A(\Omega)$ mit (225.10) aus (225.12)

$$\begin{aligned}
 A(\Omega) &= \sum_{k=0}^{N-1} h(k)e^{-j(k+\alpha_1)\Omega} \\
 &= \sum_{k=0}^{(N-3)/2} h(k)e^{-j(k+\alpha_1)\Omega} + h(-\alpha_1) + \sum_{k=(N+1)/2}^{N-1} h(k)e^{-j(k+\alpha_1)\Omega} .
 \end{aligned} \quad (225.17)$$

Substituiert man in der letzten Summe $k = N-1-n = -2\alpha_1-n$ und $n = N-1-k$, ergibt sich analog zu (225.14)

$$\begin{aligned}
 A(\Omega) &= \sum_{k=0}^{(N-3)/2} h(k) e^{-j(k+\alpha_1)\Omega} + h(-\alpha_1) + \sum_{n=0}^{(N-3)/2} h(N-1-n) e^{j(n+\alpha_1)\Omega} \\
 &= 2 \sum_{k=0}^{(N-3)/2} d_k \cos(k+\alpha_1)\Omega + d_{-\alpha_1} .
 \end{aligned} \tag{225.18}$$

Mit den Substitutionen

$$\begin{aligned}
 \tilde{f}_0 &= d_{-\alpha_1} \\
 \tilde{f}_m &= 2d_{-\alpha_1-m} \quad \text{für } m \in \{1, \dots, (N-1)/2\}
 \end{aligned} \tag{225.19}$$

erhält man schließlich aus (225.18)

$$A(\Omega) = \sum_{k=0}^{(N-1)/2} \tilde{f}_k \cos k\Omega . \tag{225.20}$$

Setzt man dieses Ergebnis in (225.12) ein, ergibt sich die Frequenzantwort $H(j\Omega)$ des linearphasigen Filters mit ungeradzahligem N . Mit (225.20) ist die Amplitude $A(\Omega)$ durch eine Summe harmonischer Schwingungen der Form (213.2) bestimmt.

Es wird nun das nullphasige Filter mit

$$b) \alpha_1 = 0 \quad \text{und} \quad \alpha_2 = 0$$

aus (221.41) behandelt, das nach (221.43) die reelle Frequenzantwort

$$H(j\Omega) = A(\Omega) \tag{225.21}$$

besitzt. Wird zunächst Kausalität nach (216.8) gefordert, gilt wegen (221.44) $h(n) \neq 0$ nur für $n = 0$, so daß die Impulsantwort

$$h(n) = c d(n) \tag{225.22}$$

lautet. Hierin bedeuten c eine Konstante und $d(n)$ der Einheitsimpuls (211.7). Setzt man (225.22) in (216.6) ein, ergibt sich

$$y(n) = c x(n) .$$

Das Eingangssignal $x(n)$ und das Ausgangssignal $y(n)$ unterscheiden sich also nur um den konstanten Faktor c . Ein sinnvolles kausales nullphasiges nichtrekursives Filter ist demnach nicht realisierbar, so daß nullphasige nichtrekursive Filter nichtkausal sein müssen.

Mit (225.16) und (225.20) wurden bereits reelle Amplitudenspektren erhalten, die nach (225.21) eine nullphasige Filterung bewirken. Im Fall

ba) N gerade Zahl

ist die Frequenzantwort also mit (225.21) durch (225.16) gegeben. Da die nullphasige Filterung besonders wichtig ist, soll zusätzlich die Filtergleichung hergeleitet werden, wozu zunächst die Impulsantwort benötigt wird, die die Bedingung (221.44) erfüllt.

Nach (225.13) gilt

$$A(\Omega) = \sum_{k=0}^{N-1} h(k) e^{-j(k+\alpha_1)\Omega} \quad (225.23)$$

Substituiert man hierin $k = n - \alpha_1 + 1/2$ und $n = k + \alpha_1 - 1/2$, erhält man mit (225.10)

$$\begin{aligned} A(\Omega) &= \sum_{n=-N/2}^{N/2-1} h(n - \alpha_1 + 1/2) e^{-j(n+1/2)\Omega} \\ &= \sum_{n=-N/2}^{N/2-1} h(n+N/2) e^{-j(n+1/2)\Omega} = \sum_{n=-N/2}^{N/2-1} h_1(n) e^{-j(n+1/2)\Omega} \end{aligned} \quad (225.24)$$

worin

$$h_1(n) = h(n+N/2) \quad \text{für } n \in \{-N/2, \dots, N/2-1\} \quad (225.25)$$

die Impulsantwort des wegen (216.8) nichtkausalen nullphasigen Filters bedeutet. Da N eine gerade Zahl ist, liegt der Symmetriepunkt der Impulsantwort bei $-1/2$. Die Filtergleichung erhält man daher aus (216.6) zu

$$y(n) = \sum_{k=-N/2}^{N/2-1} h_1(k) x(n - (k+1/2)) \quad (225.26)$$

Unter Berücksichtigung von (225.25), (225.4) und (225.15) kann (225.26) umgeschrieben werden in

$$y(n) = \frac{1}{2} \sum_{k=1}^{N/2} \bar{d}_k (x(n+k-1/2) + x(n-k+1/2)) \quad (225.27)$$

Für die Berechnung des gefilterten Signals wird somit wegen $(n+k-1/2)$ und $(n-k+1/2)$ das Eingangssignal zwischen den diskreten Werten benötigt, was als Nachteil eines nichtrekursiven Filters mit Nullphase und geradzahligem N anzusehen ist. Im Fall

bb) N ungerade Zahl

gilt mit (225.21) die Frequenzantwort (225.20). Um die Filtergleichung abzuleiten, wird die Impulsantwort benötigt. Substituiert man in (225.23) $k = n - \alpha_1$ und $n = k + \alpha_1$, erhält man mit (225.10)

$$A(\Omega) = \sum_{n=-(N-1)/2}^{(N-1)/2} h(n-\alpha_1) e^{-jn\Omega} = \sum_{n=-(N-1)/2}^{(N-1)/2} h_1(n) e^{-jn\Omega}, \quad (225.28)$$

worin

$$h_1(n) = h(n+(N-1)/2) \quad \text{für } n \in \{-(N-1)/2, \dots, (N-1)/2\} \quad (225.29)$$

die Impulsantwort des wegen (216.8) nichtkausalen nullphasigen Filters bedeutet. Seine Filtergleichung erhält man aus (216.6) zu

$$y(n) = \sum_{k=-(N-1)/2}^{(N-1)/2} h_1(k) x(n-k) \quad (225.30)$$

und unter Berücksichtigung von (225.29), (225.4) und (225.19) zu

$$y(n) = \frac{1}{2} \sum_{k=0}^{(N-1)/2} \bar{T}_k (x(n+k) + x(n-k)) . \quad (225.31)$$

Die im Zusammenhang mit (225.27) erwähnten Nachteile eines nullphasigen nichtrekursiven Filters mit einem geradzahigen Wert für N treten bei der Wahl eines ungeradzahigen Wertes für N nicht auf. Für nichtrekursive Filter mit Nullphase sind deshalb ungerade N zu bevorzugen.

In dem Fall, daß N eine ungerade Zahl ist, kann die Frequenzantwort (225.21) mit (225.20) auch aus (223.22) abgeleitet werden. Für nichtrekursive Filter verschwinden nämlich die Koeffizienten $\bar{b}_1, \dots, \bar{b}_N$. Ersetzt man außerdem die obere Summationsgrenze N durch $(N-1)/2$ und schreibt $\bar{a}_k = \bar{T}_k$, erhält man

$$H(j\Omega) = \sum_{k=0}^{(N-1)/2} \bar{T}_k \cos k\Omega .$$

Betrachtet wird nun der Fall

c) α_1 beliebig und $\alpha_2 = \pi/2$

aus (221.45). Die Impulsantwort erfüllt wegen (221.47) die Bedingung

$$h(n-\alpha_1) = -h(-n-\alpha_1) . \quad (225.32)$$

Substituiert man hierin $k = n-\alpha_1$ und $n = k+\alpha_1$, ergibt sich mit (225.10)

$$h(k) = -h(N-1-k) \quad \text{für } k \in \{0, \dots, N-1\} . \quad (225.33)$$

Die Filtergleichung erhält man aus (225.3) mit (225.4) unter Berücksichtigung dieser Antisymmetriebedingung. Zur Berechnung der Frequenzantwort wird zunächst wieder der Fall

ca) N gerade Zahl

untersucht. Man erhält für $H(j\Omega)$ mit (212.8) entsprechend (225.12)

$$H(j\Omega) = \sum_{k=0}^{N-1} h(k)e^{-jk\Omega} = A(\Omega)e^{j(\alpha_1\Omega+\pi/2)} = jA(\Omega)e^{j\alpha_1\Omega}, \quad (225.34)$$

mit der reellen Amplitude

$$\begin{aligned} A(\Omega) &= -j \sum_{k=0}^{N-1} h(k)e^{-j(k+\alpha_1)\Omega} \\ &= -j \left[\sum_{k=0}^{N/2-1} h(k)e^{-j(k+\alpha_1)\Omega} + \sum_{k=N/2}^{N-1} h(k)e^{-j(k+\alpha_1)\Omega} \right]. \end{aligned} \quad (225.35)$$

Substituiert man in der zweiten Summe $k = N-1-n = -2\alpha_1-n$ und $n = N-1-k$, folgt unter Berücksichtigung von (212.11), (225.4), (225.32) und (225.15)

$$\begin{aligned} A(\Omega) &= -j \left[\sum_{k=0}^{N/2-1} h(k)e^{-j(k+\alpha_1)\Omega} + \sum_{n=0}^{N/2-1} h(N-1-n)e^{j(n+\alpha_1)\Omega} \right] \\ &= -2 \sum_{k=0}^{N/2-1} d_k \sin(n+\alpha_1)\Omega = - \sum_{k=1}^{N/2} \bar{d}_k \sin(k-1/2)\Omega. \end{aligned} \quad (225.36)$$

Setzt man dieses Ergebnis in (225.34) ein, erhält man die Frequenzantwort $H(j\Omega)$ des linearphasigen Filters. Für

cb) N ungerade Zahl

ergibt sich anstelle von (225.35) analog zu (225.17)

$$A(\Omega) = -j \left[\sum_{k=0}^{(N-3)/2} h(k)e^{-j(k+\alpha_1)\Omega} + h(-\alpha_1) + \sum_{k=(N+1)/2}^{N-1} h(k)e^{-j(k+\alpha_1)\Omega} \right], \quad (225.37)$$

wobei wegen (225.32)

$$h(-\alpha_1) = 0 \quad (225.38)$$

gelten muß. Berücksichtigt man dieses Ergebnis in (225.37) und beachtet weiter (212.11), (225.4), (225.32) und (225.19), erhält man für die reelle Amplitude

$$A(\Omega) = - \sum_{k=1}^{(N-1)/2} \bar{I}_k \sin k\Omega \quad (225.39)$$

und aus (225.34) die Frequenzantwort $H(j\Omega)$. $A(\Omega)$ verschwindet sowohl für $\Omega = 0$ als

auch für $\Omega = \pi$. Nach Abbildung 222-1 sind Systeme mit einem solchen Amplitudenspektrum als Bandpässe einzusetzen.

Behandelt wird abschließend der Fall

$$d) \alpha_1 = 0 \quad \text{und} \quad \alpha_2 = \pi/2$$

aus (221.48), für den nach (221.50) die imaginäre Frequenzantwort

$$H(j\Omega) = jA(\Omega) \quad (225.40)$$

gültig ist. Zur Herleitung der Filtergleichungen kann analog zu (225.23) bis (225.31) vorgefahren werden, wobei in (225.24) die Antisymmetrie (225.33) zu berücksichtigen ist. Ohne hier auf die einzelnen Zwischenschritte einzugehen, erhält man für

da) N gerade Zahl

die Filtergleichung

$$y(n) = \frac{1}{2} \sum_{k=1}^{N/2} \bar{d}_k (x(n+k-1/2) - x(n-k+1/2)) \quad (225.41)$$

und für

db) N ungerade Zahl

die Filtergleichung

$$y(n) = \frac{1}{2} \sum_{k=0}^{(N-1)/2} \bar{f}_k (x(n+k) - x(n-k)) \quad (225.42)$$

Auf den Nachteil von (225.41) im Vergleich zu (225.42) wurde bereits im Zusammenhang mit (225.27) hingewiesen.

226 Nichtrekursive Elementarfilter

Es werden nun einige nichtrekursive Elementarfilter vorgestellt, deren Impulsantworten auf empirische Weise definiert sind. Sie sind einerseits bei manchen eindimensionalen Problemstellungen sinnvoll einsetzbar, andererseits werden aus ihnen, wie im Kapitel 245 gezeigt wird, zweidimensionale Elementarfilter konstruiert, die häufig in der digitalen Bildverarbeitung Verwendung finden.

Man bezeichnet ein kausales nichtrekursives Filter $(N-1)$ -ten Grades als *Rechteckfilter*,

falls für die Filterkoeffizienten d_n gilt

$$d_n = \frac{1}{N} \quad \text{für } n \in \{0, \dots, N-1\} . \quad (226.1)$$

Durch die Filterkoeffizienten ist nach (225.4) die Impulsantwort $h(n)$ unmittelbar bestimmt. Da die Bedingung (225.11) erfüllt wird, ist wegen (225.21) ein nullphasiges Rechteckfilter realisierbar. Seine Filtergleichung ist bei ungeradzahligem N mit (225.31) unter Berücksichtigung von (225.19) durch

$$y(n) = \frac{1}{N} \left[x(n) + \sum_{k=1}^{(N-1)/2} (x(n+k) + x(n-k)) \right] \quad (226.2)$$

gegeben. Die Berechnung von $y(n)$ entspricht also einer Mittelwertbildung. Die reelle Frequenzantwort lautet nach (225.21) mit (225.20)

$$H(j\Omega) = R^{N-1}(\Omega) = \frac{1}{N} + \frac{2}{N} \sum_{k=1}^{(N-1)/2} \cos k\Omega \quad (226.3)$$

und ist in Abbildung 226-1 für $N = 5$ dargestellt. Das Rechteckfilter bewirkt eine Tief-

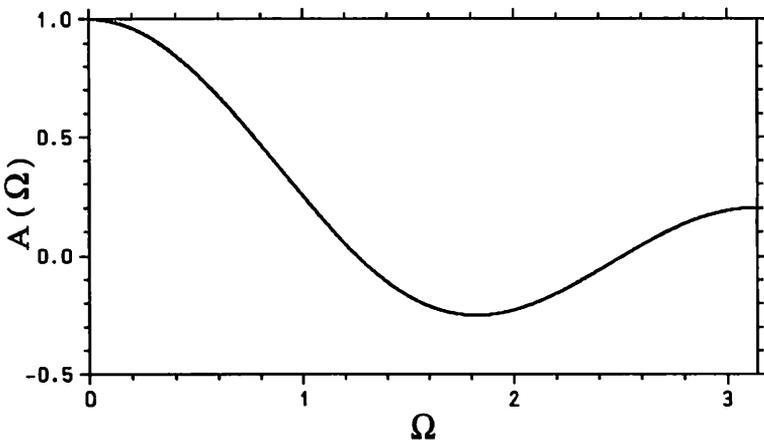


Abb. 226-1: Amplitude des Rechteckfilters

paßfilterung, wobei jedoch gegenüber dem idealen Tiefpaßfilter große Approximationsfehler verbleiben. Diese sind auch bei einer Erhöhung des Filtergrades nicht zu verringern.

Ähnlich verhält sich das *Binomialfilter*, dessen Filterkoeffizienten lauten

$$d_n = 2^{-N+1} \frac{(N-1)!}{(N-1-n)! n!} \quad \text{für } n \in \{0, \dots, N-1\} . \quad (226.4)$$

Aufgrund des symmetrischen Verhaltens der Filterkoeffizienten mit $d_n = d_{N-1-n}$ ist nach (225.11) wieder eine nullphasige Filterung mit der bei ungeradzahligem N aus (225.21) mit (225.20) folgenden Frequenzantwort

$$H(j\Omega) = B^{N-1}(\Omega) = 2^{-N+1} \left(\frac{(2\alpha)!}{\alpha! \alpha!} + \sum_{k=1}^{\alpha} \frac{2(2\alpha)!}{(\alpha+k)! (\alpha-k)!} \cos k\Omega \right) \quad (226.5)$$

möglich, wobei $\alpha = -\alpha_1 = (N-1)/2$ gilt. Abbildung 226-2 zeigt (226.5) für $N = 5$.

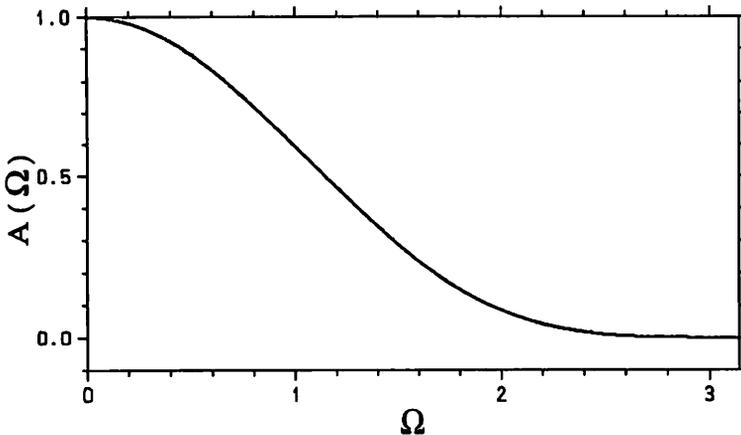


Abb. 226-2: Amplitude des Binomialfilters

Es wird nun ein Filter betrachtet, dessen Filterkoeffizienten lauten

$$d_n = \begin{cases} 1/N & \text{für } n \in \{0, \dots, (N-3)/2, (N+1)/2, \dots, N-1\} \\ 0 & \text{für } n = (N-1)/2, \end{cases} \quad (226.6)$$

wobei N als ungerade Zahl vorausgesetzt wird. Da die Filterkoeffizienten wieder ein symmetrisches Verhalten aufweisen, ist eine nullphasige Filterung möglich. Die reelle Frequenzantwort folgt also aus (225.21) mit (225.20) zu

$$H(j\Omega) = V^{N-1}(\Omega) = \frac{2}{N} \sum_{k=1}^{(N-1)/2} \cos k\Omega. \quad (226.7)$$

Vergleicht man dieses Ergebnis mit (226.3), so unterscheidet sich dieses Filter vom Rechteckfilter nur durch das Verschwinden des ersten Summanden in der Frequenzantwort. Die Filtergleichung ist wieder durch (226.2) gegeben, wobei jedoch der zentrale Eingangssignalwert $x(n)$ unberücksichtigt bleibt

$$y(n) = \frac{1}{N} \left[\sum_{k=1}^{(N-1)/2} x(n+k) + x(n-k) \right] . \quad (226.8)$$

Dieses Filter wird zur Konstruktion eines zweidimensionalen Filters benötigt.

Sowohl Rechteck- als auch Binomialfilter sind Tiefpaßfilter und bewirken eine Glättung des Eingangssignals. Vielfach ist es in der digitalen Bildverarbeitung erwünscht, rasche Änderungen der Grauwerte in Form von Kanten herauszuheben. Als nichtrekursive Elementarfilter finden dann die Ableitungsfiler erster und zweiter Ordnung Verwendung, die auf der Diskretisierung der ersten und zweiten partiellen Ableitung des Eingangssignals basieren. Die Filterkoeffizienten des *Ableitungsfilters erster Ordnung* oder *Gradientenfilters* $(N-1)$ -ten Grades lauten

$$d_n = \begin{cases} (N-1)^{-1} & \text{für } n = 0 \\ -(N-1)^{-1} & \text{für } n = N-1 \\ 0 & \text{sonst .} \end{cases} \quad (226.9)$$

Wegen (225.4) ist nach (225.32) die Impulsantwort antisymmetrisch, so daß $\alpha_2 = \pi/2$ in (225.8) gilt. Die imaginäre Frequenzantwort ergibt sich für $\alpha_1 = 0$ bei geradzahligem N aus (225.40) mit (225.36) und bei ungeradzahligem N aus (225.40) mit (225.39) zu

$$H(j\Omega) = G^{N-1}(j\Omega) = -j \frac{2}{N-1} \sin\left(\frac{N-1}{2}\Omega\right) . \quad (226.10)$$

Werden zwei Gradientenfilter mit den Frequenzantworten $G^{K-1}(j\Omega)$ aus (226.10) hintereinander geschaltet, resultiert nach (221.16) die reelle und nach (221.43) nullphasige Systemfunktion

$$\begin{aligned} H(j\Omega) &= L^{N-1}(\Omega) = G^{K-1}(j\Omega)G^{K-1}(j\Omega) \\ &= - \frac{4}{(K-1)(K-1)} \left[\sin\left(\frac{K-1}{2}\Omega\right)\sin\left(\frac{K-1}{2}\Omega\right) \right] \\ &= - \frac{2}{(K-1)(K-1)} \left[1 - \cos((K-1)\Omega) \right] . \end{aligned} \quad (226.11)$$

Dies ist eine Systemfunktion der Form (225.21) mit (225.20), so daß für den Grad des Filters aus $(N-1)/2 = K-1$ folgt $N-1 = 2K-2$. Die Filterkoeffizienten d_n mit $n \in \{0, \dots, N-1\}$ ergeben sich dann aus (225.19) zu

$$d_n = \begin{cases} (K-1)^{-2} & \text{für } n \in \{0, N-1\} \\ -2(K-1)^{-2} & \text{für } n = K-1 = (N-1)/2 \\ 0 & \text{sonst .} \end{cases} \quad (226.12)$$

Mit (226.11) wird die Frequenzantwort eines verallgemeinerten *Ableitungsfilters zweiter Ordnung* oder *Laplacefilters* beschrieben, wie das folgende Beispiel zeigt.

Beispiel 1: Ein Gradientenfilter ersten Grades, für den also $N = 2$ in (226.10) gilt, ist nach (226.9) durch die Filterkoeffizienten

$$d_0 = 1 \quad , \quad d_1 = -1 \quad (226.13)$$

definiert, so daß nach (226.10) die Frequenzantwort

$$G^1(j\Omega) = -2j\sin(\Omega/2) \quad (226.14)$$

folgt. Die Filtergleichung lautet nach (225.41) mit (225.15)

$$y(n) = x(n+1/2) - x(n-1/2)$$

und stellt damit die diskretisierte erste partielle Ableitung an der Stelle $x(n)$ dar. Weiter ergibt das Produkt (226.11) mit $K = 2$ und $N = 3$ die Frequenzantwort des Laplacefilters zu

$$L^2(\Omega) = G^1(j\Omega)G^1(j\Omega) = -2 + 2\cos\Omega \quad , \quad (226.15)$$

woraus nach (226.12) die Filterkoeffizienten

$$d_0 = 1 \quad , \quad d_1 = -2 \quad , \quad d_2 = 1 \quad (226.16)$$

folgen. Nach (225.31) mit (225.19) erhält man die Filtergleichung

$$y(n) = x(n-1) - 2x(n) + x(n+1) \quad , \quad (226.17)$$

die der diskretisierten zweiten partiellen Ableitung an der Stelle $x(n)$ entspricht, denn es gilt $y(n) = [x(n+1) - x(n)] - [x(n) - x(n-1)]$. Das Laplacefilter mit der Frequenzantwort $L^2(\Omega)$ läßt sich auch über die Parallelstruktur konstruieren. Führt man das Identitätsfilter mit der Frequenzantwort $I(\Omega) = I = 1$ ein, ergibt sich aus (221.17) mit $K = 2$ für $H_1(j\Omega) = 4B^2(\Omega)$ und $H_2(j\Omega) = -4I$ das Ergebnis

$$H(j\Omega) = 4(B^2(\Omega) - I) = L^2(\Omega) \quad , \quad (226.18)$$

wobei $B^2(\Omega) = (1 + \cos\Omega)/2$ aus (226.5) für $N = 3$ folgt. Δ

Werden verschiedene Elementarfilter in Parallel- und/oder Kaskadenstruktur zusammengesetzt, können Filter resultieren, die eine vorgegebene Idealfunktion weitaus besser approximieren als bei ihrer Einzelanwendung. Dennoch werden in der Praxis zumeist Filter mit wesentlich günstigerem Approximationsverhalten benötigt. Gesucht werden also Verfahren, die eine möglichst gute Approximation der Idealfunktion durch die Realfunktion erreichen, worauf im folgenden Kapitel eingegangen wird.

227 Entwurf nichtrekursiver Filter

Zur Approximation einer Idealfunktion durch eine Realfunktion im Fall nichtrekursiver Filter werden in diesem Kapitel zwei Verfahren vorgestellt. Der Beschreibung des allgemein anwendbaren Schätzverfahrens im linearen Modell folgen Erläuterungen zur Fensterfunktionsmethode. Weitere Verfahren entnehme man der Literatur (zum Beispiel Cappellini et al. 1978, S.53).

Das Schätzverfahren im linearen Modell wurde für rekursive Filter am Ende des Kapitels 224 vorgestellt. Für nichtrekursive Filter reduziert sich der unbekannte Vektor β der Filterkoeffizienten aus (222.4) zu

$$\beta = [d_0, d_1, \dots, d_{N-1}]' \quad (227.1)$$

Als Idealfunktion (222.1) wird im folgenden wieder das Amplitudenspektrum $A(\Omega)$ gewählt. Die Realfunktion $A_0(\Omega)$ ist dann je nach Vorgabe der Koeffizienten α_1 und α_2 der linearen Phase (221.35) sowie des Filtergrades $N-1$ durch (225.16), (225.20), (225.36) oder (225.39) anzusetzen. Für ein nullphasiges Filter mit ungeradem N erhält man nach (225.20)

$$A_0(\Omega) = \sum_{k=0}^{(N-1)/2} \bar{f}_k \cos k\Omega \quad (227.2)$$

Faßt man anstelle von (227.1) mit $(N+1)/2 = u$ die u Koeffizienten \bar{f}_k mit $k \in \{0, \dots, (N-1)/2\}$ im $u \times 1$ Vektor β zusammen

$$\beta = [\bar{f}_0, \bar{f}_1, \dots, \bar{f}_{(N-1)/2}]' \quad (227.3)$$

und führt weiterhin den $u \times 1$ Vektor

$$\mathbf{x}(\Omega) = [1, \cos\Omega, \dots, \cos(N-1)\Omega/2]' \quad (227.4)$$

ein, so läßt sich (227.2) in der Form

$$A_0(\Omega, \beta) = \mathbf{x}(\Omega)' \beta \quad (227.5)$$

schreiben. Der Approximationsfehler (222.7) lautet dann

$$e(\Omega) = \mathbf{x}(\Omega)' \beta - A(\Omega) \quad (227.6)$$

Nach der Diskretisierung von Ω in die n Werte Ω_i mit $i \in \{1, \dots, n\}$ und $n > u$ erhält man mit $e_i = e(\Omega_i)$, $\mathbf{x}_i = \mathbf{x}(\Omega_i)$ und $y_i = A(\Omega_i)$ für (227.6)

$$y_i + e_i = \mathbf{x}_i' \beta \quad (227.7)$$

analog zu (224.41). Faßt man die Elemente y_i im $n \times 1$ Vektor y und die Elemente e_i im

$n \times 1$ Vektor \mathbf{e} zusammen und bildet die $n \times u$ Matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]'$, ergibt sich das lineare Modell der Parameterschätzung

$$\mathbf{y} + \mathbf{e} = \mathbf{X}\boldsymbol{\beta} . \quad (227.8)$$

Mit der Gewichtsmatrix \mathbf{P} der y_i folgt die Schätzung (Koch 1987, S.188)

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}\mathbf{y} = [\hat{f}_0, \hat{f}_1, \dots, \hat{f}_{(N-1)/2}]' . \quad (227.9)$$

Der Residuenvektor $\hat{\mathbf{e}}$ ergibt sich zu

$$\hat{\mathbf{e}} = \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y} . \quad (227.10)$$

Für die Realfunktion erhält man aus (227.5) die Schätzung

$$\hat{A}_0(\Omega) = A_0(\Omega, \hat{\boldsymbol{\beta}}) = \mathbf{x}(\Omega)' \hat{\boldsymbol{\beta}} . \quad (227.11)$$

Mit den Schätzwerten $\hat{\boldsymbol{\beta}}$ aus (227.9) ist auch die Filtergleichung (225.31) bestimmt.

Beispiel 1: Gegeben sei die Idealfunktion (222.8) für das Amplitudenspektrum eines Tiefpaßfilters mit $\Omega_m = 0,41\pi$, die in Abbildung 227-1 dargestellt ist. Die Diskretisierung

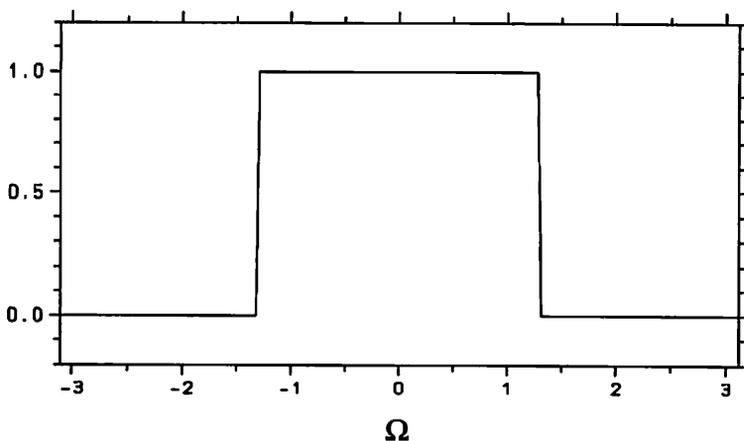


Abb. 227-1: Idealfunktion $A(\Omega)$

erfolgt in Schritten von $\Delta\Omega = 2\pi/n$ mit $n = 200$, so daß sich für die n Komponenten y_i mit $i \in \{1, \dots, n\}$ des Vektors \mathbf{y} ergibt

$$y_i = A(\Omega_i) = 1 \quad \text{für} \quad 0 \leq |\Omega_i| \leq \Omega_m$$

$$y_i = A(\Omega_i) = 0 \quad \text{für} \quad \Omega_m < |\Omega_i| \leq \pi$$

mit $\Omega_i = -\pi + i\Delta\Omega$. Als Realfunktion wird $A_0(\Omega_i) = \mathbf{x}_i' \boldsymbol{\beta}$ aus (227.5) für $N = 99$ mit

(227.3) und (227.4) gewählt, so daß die Anzahl der unbekannt Filterkoeffizienten $u = (N+1)/2 = 50$ lautet. Mit den Elementen

$$x_{ij} = \cos((j-1)\Omega_i) \quad \text{für } j \in \{1, \dots, u\} \quad \text{und } i \in \{1, \dots, n\}$$

der Matrix \mathbf{X} aus (227.8) sowie der Gewichtsmatrix $\mathbf{P} = \mathbf{I}$ ergeben sich die Schätzwerte aus (227.9).

Die Normalgleichungsmatrix aus (227.9) ist eine Diagonalmatrix, denn aufgrund der gewählten Diskretisierung von Ω ist die periodische Idealfunktion $A(\Omega)$ gleichmäßig über den Bereich einer Periode, also über 2π gegeben. Es gilt analog zu (213.3)

$$\Omega_i = -\pi + 2\pi i/n \quad \text{für } i \in \{1, \dots, n\} .$$

Substituiert man in (213.2) $A_0/2 = \bar{f}_0$, $A_k = \bar{f}_k$, $B_k = 0$, $K = (N-1)/2$ und $\omega_0 t_j = \Omega_j$, erhält man mit $x_p(t) = A_0(\Omega)$ den Ansatz (227.2) für die Realfunktion. Die Schätzwerte der unbekannt Filterkoeffizienten können daher aus der ersten Gleichung in (213.7) berechnet werden und lauten

$$\hat{f}_0 = \frac{1}{n} \sum_{i=1}^n y_i \quad \text{und} \quad \hat{f}_k = \frac{2}{n} \sum_{i=1}^n y_i \cos(k\Omega_i) \quad \text{für } k \in \{1, \dots, (N-1)/2\} .$$

Die Abbildungen 227-2 und 227-3 zeigen die geschätzte Realfunktion $\hat{A}_0(\Omega)$ und die Approximationsfehlerfunktion $\hat{e}(\Omega) = \hat{A}_0(\Omega) - A(\Omega)$. Das Überschwingen der Realfunktion im Bereich der Sprungstelle der Idealfunktion wird als Gibbssches Phänomen bezeichnet und ist auch bei einer Erhöhung des Filtergrades nicht zu verringern (Marko 1986, S.10).

▲

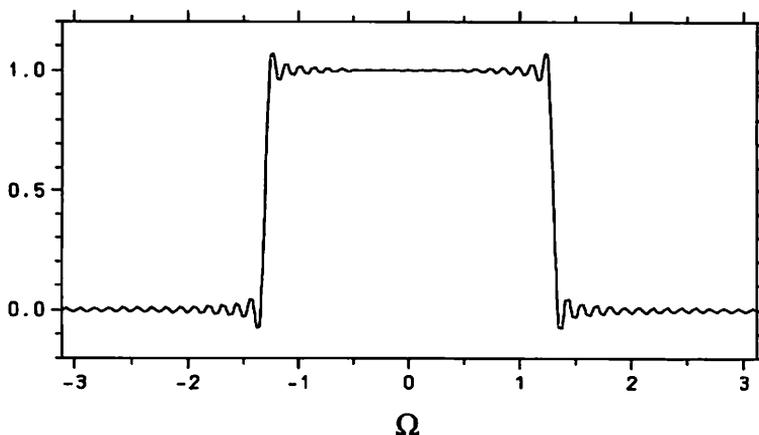


Abb. 227-2: Realfunktion $\hat{A}_0(\Omega)$

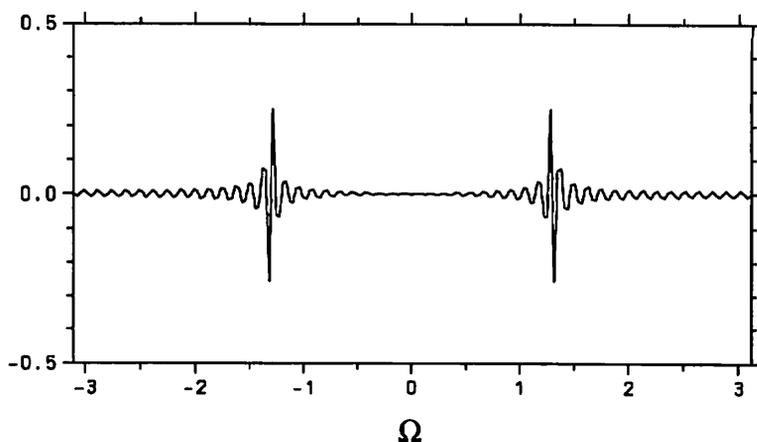


Abb. 227-3: Approximationsfehlerfunktion $\hat{e}(\Omega)$

Die Parameterschätzung (227.9), die der Methode der kleinsten Quadrate entspricht, führt zu einer Glättung der Fehler e_i , was beispielsweise bei digitalen Filtern mit scharf begrenzten Systemfunktionen unerwünscht ist. Sinnvoller ist dann die Anwendung der *Tschebyscheff-Approximation*, bei der

$$\max |e_i| = \min \quad \text{für } i \in \{1, \dots, n\} \quad (227.12)$$

mit e_i aus (227.7) gefordert wird. Zur Lösung von (227.12) kann der *Remez-Algorithmus* verwendet werden (Rabiner und Gold 1975, S.136; Rutishauser 1976, S.155). Andererseits ist auch die Nutzung der Methode der kleinsten Quadrate unter Berücksichtigung von Ungleichungsnebenbedingungen möglich (Fritsch 1987).

Als weiteres Entwurfsverfahren für nichtrekursive Filter wird die *Fensterfunktionsmethode* behandelt. Diese Technik beruht darauf, unendlich lange Impulsantworten, die nach (217.10) auf rekursive Filter führen, derart zu modifizieren, daß sie nach (217.14) als endlich lange Impulsantworten nichtrekursiver Filter dienen können. Auch können auf diese Weise sehr lange Impulsantworten in wesentlich kürzere umgewandelt werden.

Eingeführt wird die diskrete Funktion $w(n)$, die wegen

$$w(n) \begin{cases} \neq 0 & \text{für } n \in \{n_1, \dots, n_2\} \\ = 0 & \text{sonst} \end{cases} \quad (227.13)$$

nur innerhalb eines zusammenhängenden Bereichs von Null verschiedene Werte aufweist und daher als *Fensterfunktion* bezeichnet wird. Ihre Fourier-Transformation lautet nach (216.14)

$$W(j\Omega) = \sum_{n=n_1}^{n_2} w(n)e^{-jn\Omega} \quad (227.14)$$

und wird *Spektralfenster* genannt. Als Fensterfunktion kann beispielsweise das Rechtecksignal (218.20) oder das Dreiecksignal (218.22) gewählt werden.

Bei der Fensterfunktionsmethode wird als Idealfunktion $f(x)$ eine unendlich lange Impulsantwort $h(n)$ gewählt

$$f(x) = h(n) \quad (227.15)$$

mit

$$h(n) \neq 0 \quad \text{für } n \in \{0, \pm 1, \pm 2, \dots\} .$$

Die Realfunktion $f_0(x)$ erhält man mit der Fensterfunktion (227.13) zu

$$f_0(x) = h_0(n) = w(n)h(n) \quad (227.16)$$

mit

$$h_0(n) \begin{cases} \neq 0 & \text{für } n \in \{n_1, \dots, n_2\} \\ = 0 & \text{sonst} . \end{cases}$$

Die Realfunktion weist also nur noch $n_2 - n_1 + 1$ von Null verschiedene Werte auf.

Die Filtergleichung erhält man aus der diskreten Faltung (216.6) zu

$$y_0(n) = \sum_{k=n_1}^{n_2} h_0(k)x(n-k) . \quad (227.17)$$

Soll die Filterung im Frequenzbereich durchgeführt werden, ist nach (216.10) die Frequenzantwort

$$H_0(j\Omega) = \sum_{k=n_1}^{n_2} h_0(k)e^{-jk\Omega} = \sum_{k=n_1}^{n_2} w(k)h(k)e^{-jk\Omega} \quad (227.18)$$

zu berechnen. Mit (216.13) ergibt sich unter Berücksichtigung von (227.14), wobei β eine Integrationsvariable bedeutet

$$\begin{aligned} H_0(j\Omega) &= \frac{1}{2\pi} \sum_{k=n_1}^{n_2} w(k) \int_{-\pi}^{\pi} H(j\beta) e^{jk\beta} d\beta e^{-jk\Omega} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} H(j\beta) \sum_{k=n_1}^{n_2} w(k) e^{-jk(\Omega-\beta)} d\beta \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} H(j\beta) W(j(\Omega-\beta)) d\beta . \end{aligned} \quad (227.19)$$

Vergleicht man dieses Ergebnis mit (214.8), läßt sich schreiben

$$H_0(j\Omega) = \frac{1}{2\pi} H(j\Omega) * W(j\Omega) . \quad (227.20)$$

Bis auf den Faktor $1/2\pi$ ist dies die Faltung der Frequenzantwort $H(j\Omega)$ mit dem Spektralfenster $W(j\Omega)$. Die Frequenzantwort $H_0(j\Omega)$ enthält somit auch Frequenzanteile, die aus dem Spektralfenster und nicht aus $H(j\Omega)$ resultieren. Dieser Effekt wird als "spektrale Verschmierung" bezeichnet, die bei unterschiedlichen Fensterfunktionen verschieden stark ist. Geeignete Fensterfunktionen mit geringen Verfälschungen werden bei (Cappellini et al. 1978, S.69) vorgestellt.

Beispiel 2: Wählt man als Idealfunktion die Impulsantwort (216.18) des idealen Tiefpaßfilters, erhält man aus (227.16) die Realfunktion

$$h_0(n) = w(n) \frac{\sin n\Omega_m}{n\pi} . \quad (227.21)$$

Mit dem Rechteckfenster (218.20)

$$w(n) = \begin{cases} 1 & \text{für } n \in \{n_1 = -L, \dots, n_2 = L\} \\ 0 & \text{sonst} \end{cases} \quad (227.22)$$

folgt

$$h_0(n) = \begin{cases} (\sin n\Omega_m)/n\pi & \text{für } n \in \{-L, \dots, L\} \\ 0 & \text{sonst} . \end{cases} \quad (227.23)$$

Wählt man $L = (N-1)/2$, erhält man aus (227.17) nach (225.30) eine nullphasige Filtergleichung

$$y(n) = \sum_{k=-(N-1)/2}^{(N-1)/2} h_0(k)x(n-k) = \frac{1}{2} \sum_{k=0}^{(N-1)/2} \bar{T}_k(x(n+k)+x(n-k)) \quad (227.24)$$

mit

$$\begin{aligned} \bar{T}_0 &= h_0(0) , \\ \bar{T}_m &= 2h_0(m) \quad \text{für } m \in \{1, \dots, (N-1)/2\} . \end{aligned} \quad (227.25)$$

Das Spektralfenster $W(j\Omega)$ ergibt sich aus (218.21) zu

$$W(j\Omega) = \frac{\sin(N\Omega/2)}{\sin(\Omega/2)} . \quad (227.26)$$

228 Implementierung linearer Filter

Zunächst werden rekursive und nichtrekursive digitale lineare Filter vergleichend gegenübergestellt. Eine generelle Bevorzugung der einen oder anderen Art ist nicht möglich, vielmehr weisen beide Formen Vor- und Nachteile auf, die im Einzelfall abzuwägen sind. Ein Vorteil rekursiver Systeme liegt in der Nutzung analytischer Funktionen gemäß (224.4), die einen einfachen Entwurf mit relativ geringem Rechenaufwand ermöglichen. Dies hat jedoch einen Verlust an Flexibilität zur Folge, da analytische Funktionen nur für die Realisierung von Hoch- und Tiefpaßfiltern existieren. Der Filtergrad rekursiver Systeme kann zumeist wesentlich niedriger angesetzt werden als der nichtrekursiver Systeme, was bereits aus dem Vergleich von (217.10) mit (217.14) deutlich wird. Während nichtrekursive Systeme wegen (225.7) stets die Forderung (216.7) nach Stabilität erfüllen, muß bei rekursiven Systemen diese Eigenschaft nach Satz (221.18) geprüft werden. Zur Berechnung der Polstellen kann der Schur-Cohn-Algorithmus verwendet werden, der beispielsweise bei (Henrici 1974, S.494) beschrieben ist. Ein weiterer Vorteil nichtrekursiver Systeme liegt in der Realisierung linearer Phasengänge gemäß (225.8). Schließlich ist noch die Realisierung von Elementarfiltern sowie die einfache Berechnung der Filterkoeffizienten bei Entwurfsaufgaben über eine Parameterschätzung zu nennen, wie dies am Beispiel 1 des Kapitels 227 vorgeführt wurde. Die Koeffizienten der Frequenzantwort eines rekursiven Filters können nach (224.43) zwar ebenfalls über eine Parameterschätzung bestimmt werden, jedoch sind zur Festlegung geeigneter Näherungswerte für die Linearisierung aufwendige Iterationen notwendig (Schmidt 1993b). Als ein Nachteil nichtrekursiver Filter ist die bei (225.27) und (225.41) auftretende Verschiebung des Eingangssignals zwischen die Berechnungsstellen des Ausgangssignals bei geradzahligem N zu nennen. Als wichtiges Ergebnis ist festzustellen, daß nach (223.6) und (225.21) sowohl rekursive als auch nichtrekursive Systeme mit Nullphase realisierbar sind. Diese Systeme sind nichtkausal, sie können aber verwendet werden, wenn das Eingangssignal $x(n)$ für $n \in \{0, \pm 1, \pm 2, \dots\}$ vollständig gegeben ist.

Die *Implementierung rekursiver Filter* kann nicht über die diskrete Faltung erfolgen, da die Impulsantwort (217.10) unendlich viele von Null verschiedene Werte aufweist. Vielmehr ist direkt von der Differenzgleichung (223.1) oder bei nullphasiger Filterung von (223.13) und (223.15) auszugehen. Zweckmäßigerweise führt man darüberhinaus die Filterung in der Kaskadenform durch, wie das für den Zeit- und Frequenzbereich im Kapitel 224 am Beispiel des Butterworthfilters vorgeführt wurde.

Die *Implementierung nichtrekursiver Filter* kann entweder im Zeitbereich oder im Frequenzbereich erfolgen. Im ersten Fall läßt sich für das kausale Filter das Ausgangssignal

$y(n)$ unmittelbar aus der diskreten Faltung nach (225.3) mit (225.4) gewinnen, wobei N Multiplikationen nötig sind. Diese reduzieren sich bei Ausnutzung der Symmetrie- und Antisymmetriebedingungen der Filterkoeffizienten nach (225.11) oder (225.33). Ist N zum Beispiel eine ungerade Zahl, so verbleiben im Fall der nullphasigen Filterung nach (225.31) nur noch $(N+1)/2$ Multiplikationen.

Wenn eine kurze Rechenzeit gewünscht wird, sollte bei einer langen Impulsantwort geprüft werden, die Filterung im Frequenzbereich vorzunehmen, da dann die schnelle Fourier-Transformation angewendet werden kann, auf die im Kapitel 219 eingegangen wurde. Der Ablauf der Filterung im Frequenzbereich, die man, wie bereits im Kapitel 218 erwähnt, als schnelle Faltung bezeichnet, wird im folgenden beschrieben.

Gegeben sei das Signal

$$x(n) \quad \text{mit} \quad n \in \{0, \dots, M-1\} \quad (228.1)$$

und die Impulsantwort (225.4)

$$h(n) \quad \text{mit} \quad n \in \{0, \dots, N-1\} \quad (228.2)$$

eines kausalen Systems oder die Impulsantwort (225.29)

$$h_1(n) \quad \text{mit} \quad n \in \{-(N-1)/2, \dots, (N-1)/2\} \quad (228.3)$$

eines nichtkausalen Systems bei ungeradzahligem N . Für die Anwendung der schnellen Fourier-Transformation wird zunächst mit (218.17) die ganzzahlige Größe L

$$L = 2^v \geq M+N-1 \quad \text{mit} \quad v \in \{1, 2, \dots\} \quad (228.4)$$

bestimmt, die die Bedingung (219.5) erfüllt. Aus (228.1) und (228.2) werden die Signale

$$\tilde{x}(n) = \begin{cases} x(n) & \text{für } n \in \{0, \dots, M-1\} \\ 0 & \text{für } n \in \{M, \dots, L-1\} \end{cases} \quad (228.5)$$

$$\tilde{h}(n) = \begin{cases} h(n) & \text{für } n \in \{0, \dots, N-1\} \\ 0 & \text{für } n \in \{N, \dots, L-1\} \end{cases} \quad (228.6)$$

konstruiert. Mit der diskreten Fourier-Transformation (219.2) erhält man weiter

$$\tilde{X}(k) = \sum_{n=0}^{L-1} \tilde{x}(n) (W_L)^{kn} \quad (228.7)$$

$$\tilde{H}(k) = \sum_{n=0}^{L-1} \tilde{h}(n) (W_L)^{kn} \quad (228.8)$$

für $k \in \{0, \dots, L-1\}$. Multipliziert man diese Ergebnisse nach (218.18) miteinander, er-

gibt sich

$$Y(k) = \bar{H}(k)\bar{X}(k) \quad \text{für } k \in \{0, \dots, L-1\} . \quad (228.9)$$

Schließlich erhält man über die inverse diskrete Fourier-Transformation (219.3) das gefilterte Ausgangssignal

$$y(n) = \int_{\square} \sum_{k=0}^{L-1} Y(k) (W_L)^{-kn} \quad \text{für } n \in \{0, \dots, L-1\} . \quad (228.10)$$

Ist M in (228.1) eine sehr große Zahl, kann eine Zerlegung wie für (218.19) vorgenommen werden.

Im Fall der nichtkausalen Impulsantwort $h_1(n)$ aus (228.3) bildet man anstelle von (228.6)

$$\bar{h}_1(n) = \begin{cases} h_1(n) & \text{für } n \in \{-(N-1)/2, \dots, (N-1)/2\} \\ 0 & \text{für } n \in \{(N+1)/2, \dots, L-1-(N-1)/2\} . \end{cases} \quad (228.11)$$

Die Fourier-Transformation von $\bar{h}_1(n)$ lautet dann

$$\bar{H}_1(k) = \sum_{n=-(N-1)/2}^{L-1-(N-1)/2} \bar{h}_1(n) (W_L)^{kn}$$

für $k \in \{0, \dots, L-1\}$. Substituiert man $n = m - (N-1)/2$ und $m = n + (N-1)/2$, folgt

$$\bar{H}_1(k) = \sum_{m=0}^{L-1} \bar{h}_1(m - (N-1)/2) (W_L)^{km} (W_L)^{-k(N-1)/2} . \quad (228.12)$$

Mit (218.18) erhält man dann

$$Y_1(k) = \bar{H}_1(k)\bar{X}(k) \quad \text{für } k \in \{0, \dots, L-1\} \quad (228.13)$$

und anstelle von (228.10) das gefilterte Signal

$$y(n) = \int_{\square} \sum_{k=0}^{L-1} Y_1(k) (W_L)^{-kn} \quad \text{für } n \in \{-(N-1)/2, \dots, L-1-(N-1)/2\} . \quad (228.14)$$

23 Mehrdimensionale Signale und lineare Systeme

231 Signaldarstellung, Fourier-Reihe und Fourier-Transformation

Bei vielen Anwendungen werden nicht nur eindimensionale Signale, sondern auch mehrdimensionale Signale erzeugt und beobachtet. Registrieren beispielsweise verschiedene Empfänger Satellitensignale, gewinnt man vierdimensionale Signale. Der Signalindex ist dann, wie bereits zu Beginn des Kapitels 2 erwähnt, ein vierdimensionaler Vektor. Die drei ersten Komponenten werden durch die drei räumlichen Koordinaten der Empfänger und die vierte Komponente durch die Zeitpunkte der Registrierungen gebildet. In der digitalen Bildverarbeitung werden überwiegend zweidimensionale Signale verarbeitet, auf die ausführlich im Kapitel 235 eingegangen wird.

Ein *n*-dimensionales kontinuierliches Signal wird durch

$$\mathbf{x}(\mathbf{t}) \quad (231.1)$$

beschrieben, wobei der *n*-dimensionale Indexvektor

$$\mathbf{t} = [t_1, t_2, \dots, t_n]' \quad \text{mit } \mathbf{t} \in \mathbb{R}^n \quad (231.2)$$

Element der überabzählbaren Indexmenge $T = \mathbb{R}^n$ ist. Ein *n*-dimensionales diskretes Signal wird durch

$$\mathbf{x}(\mathbf{n}) \quad (231.3)$$

dargestellt. In diesem Fall ist der *n*-dimensionale Indexvektor

$$\mathbf{n} = [n_1, n_2, \dots, n_n]' \quad \text{mit } \mathbf{n} \in \mathbb{Z}^n \quad \text{oder} \quad \mathbf{n} \in \mathbb{N}_0^n \quad (231.4)$$

Element der abzählbaren Indexmenge $T = \mathbb{Z}^n$ oder der abzählbaren Indexmenge $T = \mathbb{N}_0^n$. Es können auch *n*-dimensionale Signale vorliegen, bei denen für Komponenten t_i des Vektors \mathbf{t} gelte $t_i \in \mathbb{R}$ für Komponenten t_j mit $j \neq i$ jedoch $t_j \in \mathbb{Z}$. In diesem Fall stellt das Signal eine Mischform aus einem kontinuierlichen und einem diskreten Signal dar.

Das kontinuierliche periodische Signal $x_p(\mathbf{t})$ ist als Verallgemeinerung von (212.3) definiert durch

$$x_p(\mathbf{t} + \mathbf{T}\mathbf{k}) = x_p(\mathbf{t}) \quad (231.5)$$

mit $\mathbf{k} \in \mathbb{Z}^n$ und \mathbf{T} der $n \times n$ Periodizitätsmatrix. Sie beinhaltet die periodische Struktur des Signals in bezug auf das gewählte *t*-Koordinatensystem. Ist das Signal beispielsweise in den Richtungen der Achsen des *t*-Koordinatensystems periodisch, liegt also ein Signal

mit rechtwinkliger achsenparalleler Periodizität vor, ergibt sich die Matrix T als Diagonalmatrix

$$T = \text{diag}(T_1, T_2, \dots, T_n) . \quad (231.6)$$

Mit T_i sind die Perioden des Signals bezüglich der t_i -Achsen bezeichnet, aus denen sich die Grundfrequenzen

$$\omega_{0i} = \frac{2\pi}{T_i} \quad \text{für } i \in \{1, \dots, n\} \quad (231.7)$$

berechnen. Im folgenden wird stets von (231.6) ausgegangen. Aus (231.5) folgt dann mit (231.6)

$$x_p(t_1+k_1T_1, t_2+k_2T_2, \dots, t_n+k_nT_n) = x_p(t_1, t_2, \dots, t_n) = x_p(t) . \quad (231.8)$$

Für $k_2 = \dots = k_n = 0$ erhält man

$$x_p(t_1+k_1T_1, t_2, \dots, t_n) = x_p(t) . \quad (231.9)$$

Für feste Werte von t_2, \dots, t_n kann daher das Signal $x_p(t)$ nach (213.10) und (213.13) in eine Fourier-Reihe entwickelt werden

$$x_p(t) = \sum_{k_1=-\infty}^{\infty} D_{k_1}(t_2, \dots, t_n) e^{jk_1\omega_{01}t_1} \quad (231.10)$$

mit

$$D_{k_1}(t_2, \dots, t_n) = \frac{\omega_{01}}{2\pi} \int_{-\pi/\omega_{01}}^{\pi/\omega_{01}} x_p(t) e^{-jk_1\omega_{01}t_1} dt_1 . \quad (231.11)$$

Aus dieser Beziehung folgt nun wegen

$$x_p(t_1, t_2+k_2T_2, t_3, \dots, t_n) = x_p(t) , \quad (231.12)$$

daß $D_{k_1}(t_2, \dots, t_n)$ für feste Werte t_3, \dots, t_n bezüglich t_2 periodisch ist und in eine Fourier-Reihe entwickelt werden kann

$$D_{k_1}(t_2, \dots, t_n) = \sum_{k_2=-\infty}^{\infty} D_{k_1k_2}(t_3, \dots, t_n) e^{jk_2\omega_{02}t_2} \quad (231.13)$$

mit

$$D_{k_1k_2}(t_3, \dots, t_n) = \frac{\omega_{02}}{2\pi} \int_{-\pi/\omega_{02}}^{\pi/\omega_{02}} D_{k_1}(t_2, \dots, t_n) e^{-jk_2\omega_{02}t_2} dt_2 . \quad (231.14)$$

Fährt man in dieser Weise fort, berücksichtigt man also im nächsten Schritt die Periodizität von $D_{k_1k_2}(t_3, \dots, t_n)$ in t_3 , so erhält man schließlich durch fortlaufende Substitution die *n-dimensionale Fourier-Reihe* im Fall der rechtwinkligen Periodizität unter Berücksichtigung von (231.6) und (231.7) zu

$$\begin{aligned}
 x_p(t) &= \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_n=-\infty}^{\infty} D_{k_1 \dots k_n} e^{j(k_1 \omega_{01} t_1 + \dots + k_n \omega_{0n} t_n)} \\
 &= \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_n=-\infty}^{\infty} D(\mathbf{k}) e^{j2\pi \mathbf{k}' \mathbf{T}^{-1} t}
 \end{aligned} \tag{231.15}$$

mit

$$D(\mathbf{k}) = D_{k_1 \dots k_n} = \det \mathbf{T}^{-1} \int_{-\pi/\omega_{0n}}^{\pi/\omega_{0n}} \dots \int_{-\pi/\omega_{01}}^{\pi/\omega_{01}} x_p(t) e^{-j2\pi \mathbf{k}' \mathbf{T}^{-1} t} dt \quad . \tag{231.16}$$

Die n -dimensionale Fourierreihe wird zum n -dimensionalen Fourier-Integral verallgemeinert, indem die Perioden T_i aus (231.6) bis ins Unendliche ausgedehnt werden. Wird die zu (214.2) führende Vorgehensweise auf (231.15) und (231.16) angewandt, erhält man

$$x(t) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} X(j\boldsymbol{\omega}) e^{j\boldsymbol{\omega}' t} d\boldsymbol{\omega} \tag{231.17}$$

mit

$$X(j\boldsymbol{\omega}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x(t) e^{-j\boldsymbol{\omega}' t} dt \quad . \tag{231.18}$$

Die $n \times 1$ Vektoren $\boldsymbol{\omega}$, die die Kreisfrequenzen ω_1 bis ω_n enthalten, spannen den n -dimensionalen Frequenzraum auf. Wie schon in (231.16) sind auch in (231.17) und (231.18) n Integrationen auszuführen. Die Integrale (231.17) und (231.18) bilden die *n-dimensionale Fourier-Transformation*, wobei (231.17) auch als *inverse n-dimensionale Fourier-Transformation* bezeichnet wird. Analog zu (214.3) kann wieder die symbolische Schreibweise

$$x(t) \mapsto X(j\boldsymbol{\omega}) \tag{231.19}$$

für die Fourier-Transformation geschrieben werden. Mit $X(j\boldsymbol{\omega})$ ist das kontinuierliche Amplitudenspektrum des kontinuierlichen aperiodischen Signals $x(t)$ gegeben.

232 Abtastung kontinuierlicher Signale

Im Kapitel 215 wurde das diskrete eindimensionale Signal durch Abtastung des kontinuierlichen eindimensionalen Signals erhalten, wobei $t = k\Delta t$ gesetzt wurde. Verallgemeinert man dieses Vorgehen auf den n -dimensionalen Fall, lassen sich zwischen den Indexvektoren $\mathbf{t} \in \mathbb{R}^n$ und $\mathbf{k} \in \mathbb{Z}^n$ die linearen Beziehungen

$$\mathbf{t} = \mathbf{A}\mathbf{k} \tag{232.1}$$

aufstellen, worin

$$\mathbf{A} = (\mathbf{a}_i) \quad \text{mit } i \in \{1, \dots, n\} \quad (232.2)$$

die $n \times n$ Abtastmatrix bedeutet. Die Abtastung erfolgt in den Richtungen der n linear unabhängigen Spaltenvektoren

$$\mathbf{a}_i = \Delta t_i \mathbf{e}_i \quad (232.3)$$

mit den Abtastintervallen Δt_i . Die Einheitsvektoren \mathbf{e}_i bilden eine Basis im \mathbb{R}^n (Koch 1987, S.11). Handelt es sich hierbei um eine Orthonormalbasis, so spannen sie ein n -dimensionales kartesisches Koordinatensystem auf, das als \mathbf{s} -Koordinatensystem bezeichnet sei. Ein Ortsvektor \mathbf{s} in diesem Koordinatensystem wird mittels der Drehmatrix \mathbf{D} durch

$$\mathbf{t} = \mathbf{D}\mathbf{s} \quad (232.4)$$

in den Ortsvektor \mathbf{t} des \mathbf{t} -Koordinatensystems transformiert. Setzt man (232.1) mit (232.4) gleich und löst nach \mathbf{s} auf, erhält man

$$\mathbf{s} = \mathbf{D}^{-1}\mathbf{A}\mathbf{k} = \mathbf{V}\mathbf{k}$$

mit $\mathbf{V} = \mathbf{D}^{-1}\mathbf{A} = (\mathbf{v}_i)$ und $i \in \{1, \dots, n\}$. Die Spaltenvektoren \mathbf{v}_i erhält man also durch die Transformation der Basisvektoren \mathbf{a}_i in das \mathbf{s} -Koordinatensystem zu

$$\mathbf{v}_i = [0, \dots, 0, \Delta t_i, 0, \dots, 0]^T, \quad (232.5)$$

wobei die i -te Komponente von \mathbf{v}_i das Abtastintervall Δt_i enthält. Mit (231.6) wurde eine rechtwinklige achsenparallele Periodizität vorausgesetzt. In gleicher Weise wird im folgenden eine achsenparallele Abtastung unterstellt, so daß $\mathbf{D} = \mathbf{I}$ und $\mathbf{A} = \mathbf{V}$ gilt. Wegen (232.5) erhält man also die Abtastmatrix \mathbf{A} als Diagonalmatrix

$$\mathbf{A} = \text{diag}(\Delta t_1, \Delta t_2, \dots, \Delta t_n) . \quad (232.6)$$

Ersetzt man unter diesen Voraussetzungen in (231.15) und (231.16) \mathbf{t} durch $\boldsymbol{\omega}$ sowie ω_0 durch Δt_i beziehungsweise in Matrixendarstellung $2\pi\Gamma^{-1}$ durch \mathbf{A} und tauscht jeweils das Vorzeichen der Exponentialfunktion, erhält man analog zu (215.1) und (215.2)

$$X_p(j\boldsymbol{\omega}) = \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_n=-\infty}^{\infty} x(\mathbf{k}) e^{-j\mathbf{k}'\mathbf{A}\boldsymbol{\omega}} \quad (232.7)$$

mit

$$x(\mathbf{k}) = \frac{\det \mathbf{A}}{(2\pi)^n} \int_{-\pi/\Delta t_n}^{\pi/\Delta t_n} \dots \int_{-\pi/\Delta t_1}^{\pi/\Delta t_1} X_p(j\boldsymbol{\omega}) e^{j\mathbf{k}'\mathbf{A}\boldsymbol{\omega}} d\boldsymbol{\omega} . \quad (232.8)$$

Es ist nun zu untersuchen, in welchem Zusammenhang das Spektrum $X_p(j\boldsymbol{\omega})$ des durch Abtastung gewonnenen diskreten Signals $x(\mathbf{k})$ mit dem Spektrum $X(j\boldsymbol{\omega})$ des kontinuier-

lichen Signals $x(t)$ aus (231.18) steht. Zunächst ergibt sich aus (231.17) für $t = \mathbf{A}\mathbf{k}$

$$x(\mathbf{k}) = x(t = \mathbf{A}\mathbf{k}) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} X(j\boldsymbol{\omega}) e^{j\mathbf{k}'\mathbf{A}\boldsymbol{\omega}} d\boldsymbol{\omega} \quad (232.9)$$

mit $\mathbf{A} = \mathbf{A}'$ wegen (232.6). Der n -dimensionale Frequenzraum wird in n -dimensionale rechteckige Volumenelemente mit dem Rauminhalt $(2\pi)^n / \det \mathbf{A}$ zerlegt. Es folgt dann mit

$$(-\pi + 2n_i\pi) / \Delta t_i < \omega_i < (\pi + 2n_i\pi) / \Delta t_i \quad (232.10)$$

für $n_i \in \mathbb{Z}$ und $i \in \{1, \dots, n\}$ aus (232.9)

$$x(\mathbf{k}) = \frac{1}{(2\pi)^n} \sum_{n_1=-\infty}^{\infty} \dots \sum_{n_n=-\infty}^{\infty} \int_{(-\pi+2n_n\pi)/\Delta t_n}^{(\pi+2n_n\pi)/\Delta t_n} \dots \int_{(-\pi+2n_1\pi)/\Delta t_1}^{(\pi+2n_1\pi)/\Delta t_1} X(j\boldsymbol{\omega}) e^{j\mathbf{k}'\mathbf{A}\boldsymbol{\omega}} d\boldsymbol{\omega}.$$

Hieraus erhält man mit der Variablentransformation $\omega_i = \Omega_i + 2n_i\pi / \Delta t_i$ beziehungsweise vektoriell $\boldsymbol{\omega} = \boldsymbol{\Omega} + 2\pi\mathbf{A}^{-1}\mathbf{n}$

$$x(\mathbf{k}) = \frac{1}{(2\pi)^n} \sum_{n_1=-\infty}^{\infty} \dots \sum_{n_n=-\infty}^{\infty} \int_{-\pi/\Delta t_n}^{\pi/\Delta t_n} \dots \int_{-\pi/\Delta t_1}^{\pi/\Delta t_1} X(j(\boldsymbol{\Omega} + 2\pi\mathbf{A}^{-1}\mathbf{n})) e^{j\mathbf{k}'\mathbf{A}\boldsymbol{\Omega}} e^{j2\pi\mathbf{k}'\mathbf{n}} d\boldsymbol{\Omega}$$

und weiter mit der Substitution $\boldsymbol{\Omega} = \boldsymbol{\omega}$ sowie (212.9)

$$x(\mathbf{k}) = \frac{\det \mathbf{A}}{(2\pi)^n} \int_{-\pi/\Delta t_n}^{\pi/\Delta t_n} \dots \int_{-\pi/\Delta t_1}^{\pi/\Delta t_1} \det \mathbf{A}^{-1} \sum_{n_1=-\infty}^{\infty} \dots \sum_{n_n=-\infty}^{\infty} X(j(\boldsymbol{\omega} + 2\pi\mathbf{A}^{-1}\mathbf{n})) e^{j\mathbf{k}'\mathbf{A}\boldsymbol{\omega}} d\boldsymbol{\omega}.$$

Der Vergleich mit (232.8) liefert den Zusammenhang

$$X_p(j\boldsymbol{\omega}) = \det \mathbf{A}^{-1} \sum_{n_1=-\infty}^{\infty} \dots \sum_{n_n=-\infty}^{\infty} X(j(\boldsymbol{\omega} + 2\pi\mathbf{A}^{-1}\mathbf{n})) \quad (232.11)$$

als Verallgemeinerung von (215.4). Definiert man weiter $\mathbf{U} = 2\pi\mathbf{A}^{-1}$, erhält man schließlich

$$X_p(j\boldsymbol{\omega}) = \det \mathbf{A}^{-1} \sum_{n_1=-\infty}^{\infty} \dots \sum_{n_n=-\infty}^{\infty} X(j(\boldsymbol{\omega} + \mathbf{U}\mathbf{n})) \quad (232.12)$$

Diese Gleichung beschreibt den Zusammenhang zwischen dem Spektrum $X(j\boldsymbol{\omega})$ des abgetasteten kontinuierlichen Signals $x(t)$ und dem Spektrum $X_p(j\boldsymbol{\omega})$ des durch Abtastung erhaltenen Signals $x(\mathbf{k})$. Die Periodizität von $X_p(j\boldsymbol{\omega})$ wird durch die Matrix \mathbf{U} bestimmt in Analogie zur Matrix \mathbf{T} aus (231.5). Sie ergibt sich aufgrund von (232.6) als Diagonalmatrix

$$\mathbf{U} = \text{diag}(2\pi/\Delta t_1, 2\pi/\Delta t_2, \dots, 2\pi/\Delta t_n) \quad (232.13)$$

und wird wie \mathbf{T} als Periodizitätsmatrix bezeichnet. Die Diagonalelemente stellen die Ab -

tastkreisfrequenzen oder Abtastfrequenzen dar.

Das kontinuierliche Signal $x(t)$ wird nun als bandbegrenzt angenommen. Dabei seien die Abtastintervalle Δt_i so gewählt, daß gilt

$$X(j\omega) = 0 \quad \text{für} \quad |\omega_i| \geq \pi/\Delta t_i, \quad i \in \{1, \dots, n\}. \quad (232.14)$$

Aufgrund von (232.13) ergibt sich dann aus (232.12)

$$X_p(j\omega) = \det A^{-1} X(j\omega) \quad \text{für} \quad |\omega_i| < \pi/\Delta t_i. \quad (232.15)$$

In diesem Fall reproduziert das Spektrum $X_p(j\omega)$ des diskreten Signals das Spektrum $X(j\omega)$ des kontinuierlichen Signals $x(t)$ bis auf den konstanten Faktor $\det A^{-1}$. Dies bedeutet gleichzeitig, daß bei gegebenem Spektrum $X_p(j\omega)$ das Spektrum $X(j\omega)$ des kontinuierlichen Signals berechnet werden kann

$$X(j\omega) = \det A X_p(j\omega). \quad (232.16)$$

Setzt man zuerst diesen Ausdruck in (231.17) ein und anschließend (232.7), erhält man unter Berücksichtigung der Bandbegrenzung

$$\begin{aligned} x(t) &= \frac{\det A}{(2\pi)^n} \int_{-\pi/\Delta t_n}^{\pi/\Delta t_n} \dots \int_{-\pi/\Delta t_1}^{\pi/\Delta t_1} X_p(j\omega) e^{j\omega' t} d\omega \\ &= \frac{\det A}{(2\pi)^n} \int_{-\pi/\Delta t_n}^{\pi/\Delta t_n} \dots \int_{-\pi/\Delta t_1}^{\pi/\Delta t_1} \left[\sum_{k_1=-\infty}^{\infty} \dots \sum_{k_n=-\infty}^{\infty} x(\mathbf{k}) e^{-j\omega' \mathbf{A}\mathbf{k}} \right] e^{j\omega' t} d\omega \\ &= \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_n=-\infty}^{\infty} x(\mathbf{k}) f(t - \mathbf{A}\mathbf{k}), \end{aligned} \quad (232.17)$$

worin

$$f(t - \mathbf{A}\mathbf{k}) = \frac{\det A}{(2\pi)^n} \int_{-\pi/\Delta t_n}^{\pi/\Delta t_n} \dots \int_{-\pi/\Delta t_1}^{\pi/\Delta t_1} e^{j\omega' (t - \mathbf{A}\mathbf{k})} d\omega \quad (232.18)$$

die Interpolationsfunktion bedeutet. Für sie läßt sich mit (215.11) weiter schreiben

$$\begin{aligned} f(t - \mathbf{A}\mathbf{k}) &= \prod_{i=1}^n \frac{\Delta t_i}{2\pi} \int_{-\pi/\Delta t_i}^{\pi/\Delta t_i} e^{j\omega_i (t_i - \Delta t_i k_i)} d\omega_i \\ &= \prod_{i=1}^n \frac{\sin[(\pi/\Delta t_i)(t_i - \Delta t_i k_i)]}{(\pi/\Delta t_i)(t_i - \Delta t_i k_i)}. \end{aligned} \quad (232.19)$$

Durch (232.17) mit (232.19) kann also aus dem diskreten Signal $x(\mathbf{k})$ das kontinuierliche Signal $x(t)$ wiedergewonnen werden. Dies ist jedoch nur unter den folgenden Voraussetzungen möglich. Bezeichnet man im n -dimensionalen Frequenzraum die höchsten Frequenzen des kontinuierlichen Signals mit ω_{im} für $i \in \{1, \dots, n\}$, so müssen die Abtast-

Frequenzen $2\pi/\Delta t_i$ wegen (232.14) die Bedingungen

$$2\pi/\Delta t_i > 2\omega_{im} \quad (232.20)$$

oder die Abtastintervalle die Ungleichungen $\Delta t_i < \pi/\omega_{im}$ erfüllen. Führt man analog zu (215.8) die *Nyquist-Frequenzen* ω_{iN} ein, ergibt sich

$$\omega_{iN} = \pi/\Delta t_i > \omega_{im} . \quad (232.21)$$

Aufgrund von (232.13) wird die Matrix $U/2$ auch als Nyquist-Frequenzmatrix bezeichnet. Die Bedingungen (232.20) zusammen mit der Interpolationsformel (232.17) bilden das *n-dimensionale Abtasttheorem*.

Das n-dimensionale Abtasttheorem behandelt die Abtastung der Abszissen n-dimensionaler Signale. Für die Isoliniendarstellung insbesondere zweidimensionaler Signale benötigt man die Ordinatenabtastung, deren Regeln bei (Meier und Borkowski 1993) untersucht werden.

233 Lineare Systeme

Ein *kontinuierliches System* überführt ein Eingangssignal $x(t)$ in ein Ausgangssignal $y(t)$

$$y(t) = \phi[x(t)] \quad (233.1)$$

vermittels des Transformationsoperators ϕ . Das System ist *linear*, falls mit den Konstanten a_k für $k \in \{1, \dots, K\}$ das Superpositionsprinzip

$$\phi\left[\sum_{k=1}^K a_k x_k(t)\right] = \sum_{k=1}^K a_k \phi[x_k(t)] = \sum_{k=1}^K a_k y_k(t) \quad (233.2)$$

erfüllt wird. Das System ist außerdem *verschiebungsinvariant*, wenn für beliebige Vektoren t_1

$$y(t-t_1) = \phi[x(t-t_1)] \quad (233.3)$$

zutrifft.

Analog zu (211.5) ist die *n-dimensionale Deltafunktion* $\delta(t)$ aufgrund ihrer Ausblendeigenschaft definiert, so daß sich das Eingangssignal $x(t)$ in der Form

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x(\tau) \delta(t-\tau) d\tau = x(t) \quad (233.4)$$

schreiben läßt. Als *Impulsantwort* $h(t)$ bezeichnet man die Reaktion des Systems auf die Deltafunktion als Eingangssignal

$$h(t) = \phi[\delta(t)] \quad . \quad (233.5)$$

Aus (233.1) folgt dann mit (233.4) und (233.5)

$$y(t) = \phi[x(t)] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x(\tau) \phi[\delta(t-\tau)] d\tau = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x(\tau) h(t-\tau) d\tau \quad . \quad (233.6)$$

Der rechte Ausdruck entspricht der *n-dimensionalen Faltung*. Analog zu (216.5) läßt sich also schreiben

$$\begin{aligned} y(t) &= x(t) * h(t) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x(\tau) h(t-\tau) d\tau = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h(\tau) x(t-\tau) d\tau \quad . \end{aligned} \quad (233.7)$$

Ein *diskretes System* überführt ein diskretes Eingangssignal $x(\mathbf{n})$ in ein diskretes Ausgangssignal $y(\mathbf{n})$. Ist das System wieder linear und verschiebungsinvariant, erhält man anstelle von (233.7) die *diskrete n-dimensionale Faltung*

$$\begin{aligned} y(\mathbf{n}) &= x(\mathbf{n}) * h(\mathbf{n}) \\ &= \sum_{\mathbf{k}_1=-\infty}^{\infty} \dots \sum_{\mathbf{k}_n=-\infty}^{\infty} x(\mathbf{k}) h(\mathbf{n}-\mathbf{k}) = \sum_{\mathbf{k}_1=-\infty}^{\infty} \dots \sum_{\mathbf{k}_n=-\infty}^{\infty} h(\mathbf{k}) x(\mathbf{n}-\mathbf{k}) \quad . \end{aligned} \quad (233.8)$$

Hierin bedeutet

$$h(\mathbf{n}) = \phi[d(\mathbf{n})] \quad (233.9)$$

wieder die Impulsantwort als Reaktion des diskreten Systems auf den *n-dimensionalen Einheitsimpuls*

$$d(\mathbf{n}) = \begin{cases} 1 & \text{für } \mathbf{n} = \mathbf{0} \\ 0 & \text{sonst} \end{cases} \quad (233.10)$$

als Eingangssignal. Das System ist *stabil*, falls die Bedingung

$$\sum_{\mathbf{k}_1=-\infty}^{\infty} \dots \sum_{\mathbf{k}_n=-\infty}^{\infty} |h(\mathbf{k})| < \infty \quad (233.11)$$

gilt. Sie bewirkt, daß ein beschränktes Eingangssignal ein beschränktes Ausgangssignal erzeugt. Vielfach wird die *Kausalität* des Systems durch

$$h(\mathbf{n}) = 0 \quad \text{für alle } \mathbf{n} \notin \mathbb{N}_0^n \quad (233.12)$$

definiert. Diese strenge Form der Kausalität läßt sich allerdings modifizieren, beispielsweise durch

$$h(\mathbf{n}) = 0 \quad \text{für alle } \mathbf{n} \notin \{n_1=0, n_2 \in \mathbb{N}_0, \dots, n_n \in \mathbb{N}_0\} \cup \{n_1 \in \mathbb{N}, n_2 \in \mathbb{Z}, \dots, n_n \in \mathbb{Z}\} . \quad (233.13)$$

In diesem Fall ist das System nur bezüglich n_1 streng kausal. Von Bedeutung ist dieser Ansatz immer dann, wenn n_1 mit der Zeit identifiziert wird.

Analog zu (216.9) werden wieder die normierten Frequenzen $\Omega_i = \omega_i \Delta t_i$ für $i \in \{1, \dots, n\}$ eingeführt. Sie seien im $n \times 1$ Vektor $\mathbf{\Omega}$ zusammengefaßt, für den wegen (232.6) $\mathbf{\Omega} = \mathbf{A}\mathbf{\omega}$ gilt. Aus (232.7) und (232.8) folgt dann mit der Funktionaldeterminante $\det \mathbf{J} = \det \mathbf{A}^{-1}$ (Koch 1987, S.83) die Fourier-Reihe

$$X(j\mathbf{\Omega}) = \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_n=-\infty}^{\infty} x(\mathbf{k}) e^{-j\mathbf{k}'\mathbf{\Omega}} \quad (233.14)$$

mit

$$x(\mathbf{k}) = \frac{1}{(2\pi)^n} \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} X(j\mathbf{\Omega}) e^{j\mathbf{k}'\mathbf{\Omega}} d\mathbf{\Omega} . \quad (233.15)$$

Als Eingangssignal $x(\mathbf{n})$ des Systems diene die komplexe n -dimensionale Schwingung

$$x(\mathbf{n}) = e^{j\mathbf{n}'\mathbf{\Omega}} . \quad (233.16)$$

Das Ausgangssignal $y(\mathbf{n})$ folgt dann aus (233.8) zu

$$y(\mathbf{n}) = \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_n=-\infty}^{\infty} h(\mathbf{k}) e^{j(\mathbf{n}-\mathbf{k})'\mathbf{\Omega}} . \quad (233.17)$$

Definiert man die *Frequenzantwort* des Systems durch

$$H(j\mathbf{\Omega}) = \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_n=-\infty}^{\infty} h(\mathbf{k}) e^{-j\mathbf{k}'\mathbf{\Omega}} , \quad (233.18)$$

ergibt sich aus (233.17) mit (233.16)

$$y(\mathbf{n}) = H(j\mathbf{\Omega}) e^{j\mathbf{n}'\mathbf{\Omega}} . \quad (233.19)$$

Die Darstellung (233.18) entspricht der Fourier-Reihe (233.14). Die Impulsantwort $h(\mathbf{k})$ kann daher nach (233.15) aus der Frequenzantwort berechnet werden

$$h(\mathbf{k}) = \frac{1}{(2\pi)^n} \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} H(j\mathbf{\Omega}) e^{j\mathbf{k}'\mathbf{\Omega}} d\mathbf{\Omega} . \quad (233.20)$$

Bedeutsam für die Filterung digitaler Signale sind die linearen verschiebungsinvarianten Systeme, die durch die Differenzgleichung

$$\sum_{k_1=0}^{K_1} \dots \sum_{k_n=0}^{K_n} a(\mathbf{k}) y(\mathbf{n}-\mathbf{k}) = \sum_{l_1=0}^{L_1} \dots \sum_{l_n=0}^{L_n} b(l) x(\mathbf{n}-l) \quad (233.21)$$

definiert sind, worin $a(\mathbf{k})$ und $b(l)$ reelle konstante Koeffizienten bedeuten. Löst man diese Gleichung für $a(\mathbf{0}) \neq 0$ nach $y(\mathbf{n})$ auf, erhält man

$$y(\mathbf{n}) = \frac{1}{a(\mathbf{0})} \left[\sum_{l_1=0}^{L_1} \dots \sum_{l_n=0}^{L_n} b(l)x(\mathbf{n}-l) - \sum_{\substack{k_1=0 \\ \mathbf{k} \neq \mathbf{0}}}^{K_1} \dots \sum_{k_n=0}^{K_n} a(\mathbf{k})y(\mathbf{n}-\mathbf{k}) \right] . \quad (233.22)$$

Das Ausgangssignal wird also rekursiv aus dem Eingangssignal berechnet.

234 z-Transformation, Systemfunktion und diskrete Fourier-Transformation

Die *n-dimensionale z-Transformation* $X(\mathbf{z})$ eines n -dimensionalen Signals $x(\mathbf{k})$ ist definiert durch

$$X(\mathbf{z}) = \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_n=-\infty}^{\infty} x(\mathbf{k})z_1^{-k_1} \dots z_n^{-k_n} , \quad (234.1)$$

worin \mathbf{z} ein $n \times 1$ Vektor komplexer Variablen z_i mit $i \in \{1, \dots, n\}$ bedeutet. Bildet man die z -Transformation der Faltung (233.8), erhält man

$$\begin{aligned} Y(\mathbf{z}) &= \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_n=-\infty}^{\infty} y(\mathbf{k})z_1^{-k_1} \dots z_n^{-k_n} \\ &= \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_n=-\infty}^{\infty} \left[\sum_{l_1=-\infty}^{\infty} \dots \sum_{l_n=-\infty}^{\infty} x(l)h(\mathbf{k}-l) \right] z_1^{-k_1} \dots z_n^{-k_n} \\ &= \sum_{l_1=-\infty}^{\infty} \dots \sum_{l_n=-\infty}^{\infty} x(l) \left[\sum_{k_1=-\infty}^{\infty} \dots \sum_{k_n=-\infty}^{\infty} h(\mathbf{k}-l) z_1^{-(k_1-l_1)} \dots z_n^{-(k_n-l_n)} \right] \\ &\quad z_1^{-l_1} \dots z_n^{-l_n} \end{aligned}$$

und weiter mit $\mathbf{k}-l = \mathbf{p}$

$$Y(\mathbf{z}) = X(\mathbf{z})H(\mathbf{z}) , \quad (234.2)$$

wobei

$$H(\mathbf{z}) = \sum_{p_1=-\infty}^{\infty} \dots \sum_{p_n=-\infty}^{\infty} h(\mathbf{p})z_1^{-p_1} \dots z_n^{-p_n} \quad (234.3)$$

als *n-dimensionale Systemfunktion* bezeichnet wird. Für $z_i = e^{j\Omega_i}$ geht (234.3) in die Frequenzantwort (233.18) über.

Es wird nun die Systemfunktion für ein System abgeleitet, das durch die Differenzen-

gleichung (233.21) charakterisiert ist. Bildet man dazu ihre z -Transformation, erhält man

$$\begin{aligned} & \sum_{n_1=-\infty}^{\infty} \dots \sum_{n_n=-\infty}^{\infty} \sum_{k_1=0}^{K_1} \dots \sum_{k_n=0}^{K_n} a(\mathbf{k}) y(\mathbf{n}-\mathbf{k}) z_1^{-(n_1-k_1)} \dots z_n^{-(n_n-k_n)} \\ & \qquad \qquad \qquad z_1^{-k_1} \dots z_n^{-k_n} \\ & = \sum_{n_1=-\infty}^{\infty} \dots \sum_{n_n=-\infty}^{\infty} \sum_{l_1=0}^{L_1} \dots \sum_{l_n=0}^{L_n} b(\mathbf{l}) x(\mathbf{n}-\mathbf{l}) z_1^{-(n_1-l_1)} \dots z_n^{-(n_n-l_n)} \\ & \qquad \qquad \qquad z_1^{-l_1} \dots z_n^{-l_n} \end{aligned}$$

und weiter mit (234.1)

$$\left[\sum_{k_1=0}^{K_1} \dots \sum_{k_n=0}^{K_n} a(\mathbf{k}) z_1^{-k_1} \dots z_n^{-k_n} \right] Y(\mathbf{z}) = \left[\sum_{l_1=0}^{L_1} \dots \sum_{l_n=0}^{L_n} b(\mathbf{l}) z_1^{-l_1} \dots z_n^{-l_n} \right] X(\mathbf{z}) \quad (234.4)$$

Vergleicht man dieses Ergebnis mit (234.2), erhält man die Systemfunktion $H(\mathbf{z})$ wegen

$$H(\mathbf{z}) = \left(\sum_{l_1=0}^{L_1} \dots \sum_{l_n=0}^{L_n} b(\mathbf{l}) z_1^{-l_1} \dots z_n^{-l_n} \right) / \left(\sum_{k_1=0}^{K_1} \dots \sum_{k_n=0}^{K_n} a(\mathbf{k}) z_1^{-k_1} \dots z_n^{-k_n} \right) \quad (234.5)$$

als Verhältnis von Polynomen in $z_1^{-1} \dots z_n^{-1}$. Sie soll nun in der Form (234.3) dargestellt werden. Durch fortlaufende Substitution der linken Seite von (233.22) für $y(\mathbf{n}-\mathbf{k})$ auf der rechten Seite von (233.22) erhält man das Ausgangssignal in der Form

$$y(\mathbf{n}) = \sum_{k_1=0}^{\infty} \dots \sum_{k_n=0}^{\infty} h(\mathbf{k}) x(\mathbf{n}-\mathbf{k}) \quad (234.6)$$

Nach (233.8) und (233.12) ist dies eine n -dimensionale Faltung mit streng kausaler Impulsantwort. Anstelle von (234.5) läßt sich daher nach (234.3) auch schreiben

$$H(\mathbf{z}) = \sum_{p_1=0}^{\infty} \dots \sum_{p_n=0}^{\infty} h(\mathbf{p}) z_1^{-p_1} \dots z_n^{-p_n} \quad (234.7)$$

Die Impulsantwort ergibt sich also als Funktion der Koeffizienten $a(\mathbf{k})$ und $b(\mathbf{l})$, wie der Vergleich mit (234.5) zeigt. Aufgrund ihrer unendlichen Länge werden die n -dimensionalen rekursiven Filter analog zu (217.10) als *n -dimensionale Filter mit unendlicher Impulsantwort* bezeichnet.

Die Gleichungen (234.5) bis (234.7) gelten für die rekursive Filterung. Im Falle der nichtrekursiven Filterung lautet (233.22) mit $a(\mathbf{0}) \neq 0$ und $a(\mathbf{k}) = 0$ für $\mathbf{k} \neq \mathbf{0}$

$$y(\mathbf{n}) = \sum_{l_1=0}^{L_1} \dots \sum_{l_n=0}^{L_n} (b(\mathbf{l})/a(\mathbf{0})) x(\mathbf{n}-\mathbf{l}) \quad (234.8)$$

Substituiert man hierin

$$h(\mathbf{1}) = b(\mathbf{1})/a(\mathbf{0}) \quad , \quad (234.9)$$

erhält man nach (233.8) die Faltung des Eingangssignals $x(\mathbf{n})$ mit der Impulsantwort $h(\mathbf{n})$. Nichtreursive Systeme sind also durch eine Impulsantwort endlicher Länge charakterisiert, weswegen sie analog zu (217.14) auch *n-dimensionale Filter mit endlicher Impulsantwort* genannt werden. Ihre Systemfunktion ergibt sich aus (234.3) zu

$$H(z) = \sum_{l_1=0}^{L_1} \dots \sum_{l_n=0}^{L_n} h(\mathbf{1}) z_1^{-l_1} \dots z_n^{-l_n} \quad . \quad (234.10)$$

In Analogie zur diskreten eindimensionalen Fourier-Transformation ergibt sich die *diskrete n-dimensionale Fourier-Transformation* zu

$$\begin{aligned} x(\mathbf{n}) &= \frac{1}{N_1 N_2 \dots N_n} \sum_{k_1=0}^{N_1-1} \dots \sum_{k_n=0}^{N_n-1} X(\mathbf{k}) e^{j2\pi(k_1 n_1/N_1 + \dots + k_n n_n/N_n)} \\ &= \det N^{-1} \sum_{k_1=0}^{N_1-1} \dots \sum_{k_n=0}^{N_n-1} X(\mathbf{k}) e^{j2\pi \mathbf{k}' N^{-1} \mathbf{n}} \end{aligned} \quad (234.11)$$

mit

$$X(\mathbf{k}) = \sum_{n_1=0}^{N_1-1} \dots \sum_{n_n=0}^{N_n-1} x(\mathbf{n}) e^{-j2\pi \mathbf{k}' N^{-1} \mathbf{n}} \quad (234.12)$$

für

$$k_i, n_i \in \{0, \dots, N_i - 1\} \quad \text{und} \quad i \in \{1, \dots, n\} \quad ,$$

wobei (234.11) entsprechend (218.10) wieder als *inverse diskrete n-dimensionale Fourier-Transformation* bezeichnet wird. Die $n \times n$ Diagonalmatrix N ist definiert durch

$$N = \text{diag}(N_1, \dots, N_n) \quad . \quad (234.13)$$

Bestätigen lassen sich (234.11) und (234.12) dadurch, daß sie auf n diskrete eindimensionale Fourier-Transformationen (218.10) und (218.11) zurückgeführt werden können. Anstelle von (234.12) kann geschrieben werden

$$X(\mathbf{k}) = \sum_{n_1=0}^{N_1-1} e^{-j2\pi k_1 n_1/N_1} \left[\sum_{n_2=0}^{N_2-1} \dots \sum_{n_n=0}^{N_n-1} x(\mathbf{n}) e^{-j2\pi(k_2 n_2/N_2 + \dots + k_n n_n/N_n)} \right] \quad . \quad (234.14)$$

Für einen festen Wert n_1 ist der Ausdruck in Klammern eine $(n-1)$ -dimensionale Fourier-Transformation. Führt man also

$$z(n_1, k_2, \dots, k_n) = \sum_{n_2=0}^{N_2-1} \dots \sum_{n_n=0}^{N_n-1} x(\mathbf{n}) e^{-j2\pi(k_2 n_2/N_2 + \dots + k_n n_n/N_n)} \quad (234.15)$$

ein, erhält man aus (234.14) die eindimensionale Fourier-Transformation

$$X(\mathbf{k}) = \sum_{n_1=0}^{N_1-1} z(n_1, k_2, \dots, k_n) e^{-j2\pi k_1 n_1 / N_1} \quad (234.16)$$

Wird nun (234.15) analog zu (234.14) aufgespalten, resultieren eine (n-2)-dimensionale Fourier-Transformation und eine weitere eindimensionale Fourier-Transformation entsprechend (234.15) und (234.16). Führt man in dieser Weise fort, erhält man schließlich anstelle der diskreten n-dimensionalen Fourier-Transformation n eindimensionale diskrete Fourier-Transformationen. Entsprechend läßt sich die inverse diskrete n-dimensionale Fourier-Transformation (234.11) in n eindimensionale inverse diskrete Fourier-Transformationen überführen. Die im Kapitel 219 behandelten schnellen Fourier-Transformationen sind daher auch auf (234.11) und (234.12) anwendbar.

Gemessen am Rechenaufwand kann es günstiger sein, anstelle der Faltung (233.8) die Fourier-Transformationen des Eingangssignals und der Impulsantwort nach (234.12) zu bilden, die Fourier-Transformationen zu multiplizieren und anschließend das Ergebnis der inversen Fourier-Transformation (234.11) zu unterziehen. Zur Herleitung dieser Zusammenhänge für ein nach (233.12) streng kausales System seien das Eingangssignal

$$x(\mathbf{n}) \neq 0 \quad \text{für} \quad 0 \leq n_i \leq M_i - 1 \quad (234.17)$$

und die Impulsantwort

$$h(\mathbf{n}) \neq 0 \quad \text{für} \quad 0 \leq n_i \leq K_i - 1 \quad (234.18)$$

gegeben. Das Ausgangssignal $y(\mathbf{n})$ erhält man aus (233.8) zu

$$y(\mathbf{n}) = \sum_{k_1=0}^{M_1-1} \dots \sum_{k_n=0}^{M_n-1} x(\mathbf{k}) h(\mathbf{n}-\mathbf{k}) \quad (234.19)$$

für

$$0 \leq n_i \leq N_i - 1 \quad \text{mit} \quad N_i - 1 = M_i + K_i - 2 \quad (234.20)$$

Die Gleichung (234.19) wird als *lineare Faltung* bezeichnet. Für k_i gilt in (234.19) maximal $k_i = M_i - 1$ und für $n_i - k_i$ maximal $n_i - k_i = K_i - 1$ nach (234.18). Somit ergibt sich für n_i ein Maximalwert von $n_i = M_i + K_i - 2$.

Bildet man die diskrete Fourier-Transformation von (234.19) nach (234.12), folgt

$$\begin{aligned} Y(\mathbf{k}) &= \sum_{n_1=0}^{N_1-1} \dots \sum_{n_n=0}^{N_n-1} y(\mathbf{n}) e^{-j2\pi \mathbf{k}' \mathbf{N}^{-1} \mathbf{n}} \\ &= \sum_{n_1=0}^{N_1-1} \dots \sum_{n_n=0}^{N_n-1} \left[\sum_{l_1=0}^{M_1-1} \dots \sum_{l_n=0}^{M_n-1} x(\mathbf{l}) h(\mathbf{n}-\mathbf{l}) \right] e^{-j2\pi \mathbf{k}' \mathbf{N}^{-1} \mathbf{n}} \end{aligned}$$

$$\begin{aligned}
&= \sum_{l_1=0}^{M_1-1} \dots \sum_{l_n=0}^{M_n-1} x(l) \left[\sum_{n_1=0}^{N_1-1} \dots \sum_{n_n=0}^{N_n-1} h(n-1) e^{-j2\pi \mathbf{k}' \mathbf{N}^{-1} (n-1)} \right] e^{-j2\pi \mathbf{k}' \mathbf{N}^{-1} \mathbf{l}} \\
&= \mathbf{H}(\mathbf{k}) \mathbf{X}(\mathbf{k}) \quad . \quad (234.21)
\end{aligned}$$

Das Ausgangssignal $y(\mathbf{n})$ ergibt sich dann aus (234.11) zu

$$y(\mathbf{n}) = \det \mathbf{N}^{-1} \sum_{k_1=0}^{N_1-1} \dots \sum_{k_n=0}^{N_n-1} Y(\mathbf{k}) e^{j2\pi \mathbf{k}' \mathbf{N}^{-1} \mathbf{n}} \quad . \quad (234.22)$$

Diese Vorgehensweise, die als *schnelle Faltung* bezeichnet wird, stellt die Alternative zur linearen Faltung (234.19) dar. Im Einzelfall ist zu entscheiden, welche Art der Berechnung vorzuziehen ist. Ist das Eingangssignal im Vergleich zur Impulsantwort sehr lang, kann es analog zur eindimensionalen Darstellung nach (218.19) wieder in Bereiche zerlegt werden. Die Bereiche sind dann unter Berücksichtigung von Überlappungen zu transformieren.

235 Zweidimensionale Signale

Bei vielen Anwendungen, beispielsweise in der digitalen Bildverarbeitung, liegen Signale in einer Ebene vor. Im folgenden werden daher die wichtigsten Ergebnisse aus den Kapiteln 231 bis 234 für den zweidimensionalen Fall, also $n = 2$ zusammengestellt.

Ein *zweidimensionales kontinuierliches Signal* ist nach (231.1) mit $\mathbf{t} = [t_1, t_2]'$ durch

$$\mathbf{x}(\mathbf{t}) = x(t_1, t_2) \quad (235.1)$$

gegeben. Im Fall einer achsenparallelen Abtastung erhält man mit der Abtastmatrix

$$\mathbf{A} = \begin{bmatrix} \Delta t_1 & 0 \\ 0 & \Delta t_2 \end{bmatrix} = \text{diag}(\Delta t_1, \Delta t_2) \quad (235.2)$$

aus (232.6) das *zweidimensionale diskrete Signal* $x(\mathbf{k}) = x(\mathbf{t}=\mathbf{A}\mathbf{k})$ oder mit $\mathbf{k} = [k_1, k_2]'$ = $[m, n]'$

$$\mathbf{x}(\mathbf{k}) = x(m, n) = x(t_1=m\Delta t_1, t_2=n\Delta t_2) \quad (235.3)$$

aus (231.3).

Die *zweidimensionale Fourier-Transformation* und die *inverse zweidimensionale Fourier-Transformation* sind nach (231.17) und (231.18) mit $\boldsymbol{\omega} = [\omega_1, \omega_2]'$ durch

$$\mathbf{x}(t_1, t_2) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X(j\omega_1, j\omega_2) e^{j(\omega_1 t_1 + \omega_2 t_2)} d\omega_1 d\omega_2 \quad (235.4)$$

und

$$X(j\omega_1, j\omega_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t_1, t_2) e^{-j(\omega_1 t_1 + \omega_2 t_2)} dt_1 dt_2 \quad (235.5)$$

gegeben. Das Spektrum des durch Abtastung gewonnenen diskreten Signals $x(k, l)$ aus (235.3) erhält man mit (232.7) unter Berücksichtigung von (235.2) zu

$$X_p(j\omega_1, j\omega_2) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} x(k, l) e^{-j(k\Delta t_1 \omega_1 + l\Delta t_2 \omega_2)} \quad (235.6)$$

Die inverse Transformation lautet nach (232.8)

$$x(k, l) = \frac{\Delta t_1 \Delta t_2}{4\pi^2} \int_{-\pi/\Delta t_2}^{\pi/\Delta t_2} \int_{-\pi/\Delta t_1}^{\pi/\Delta t_1} X_p(j\omega_1, j\omega_2) e^{j(k\Delta t_1 \omega_1 + l\Delta t_2 \omega_2)} d\omega_1 d\omega_2 \quad (235.7)$$

Wird das aus (232.20) folgende zweidimensionale Abtasttheorem erfüllt, reproduziert das Spektrum $X_p(j\omega_1, j\omega_2)$ nach (232.16) das Spektrum $X(j\omega_1, j\omega_2)$ bis auf den Faktor $\det A = \Delta t_1 \Delta t_2$. In diesem Fall kann aus dem diskreten Signal $x(k, l)$ mittels der Interpolationsfunktion (232.19) das kontinuierliche Signal $x(t_1, t_2)$ nach (232.17) wiedergewonnen werden.

Ein *kontinuierliches System* überführt ein Eingangssignal $x(t_1, t_2)$ durch die Transformationsgleichung (233.1)

$$y(t_1, t_2) = \phi[x(t_1, t_2)] \quad (235.8)$$

in das Ausgangssignal $y(t_1, t_2)$. Wird das System als *linear* und *verschiebungsinvariant* angenommen, ist nach (233.7) die Transformation (235.8) durch die *zweidimensionale Faltung*

$$y(t_1, t_2) = x(t_1, t_2) * h(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\tau_1, \tau_2) x(t_1 - \tau_1, t_2 - \tau_2) d\tau_1 d\tau_2 \quad (235.9)$$

gegeben. Liegt ein diskretes Eingangssignal $x(m, n)$ vor, erhält man für ein lineares und verschiebungsinvariantes diskretes System die *diskrete zweidimensionale Faltung* nach (233.8)

$$y(m, n) = x(m, n) * h(m, n) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h(k, l) x(m-k, n-l) \quad (235.10)$$

Die Impulsantwort $h(k, l)$ beschreibt das diskrete System vollständig. Sie ist nach (233.9) die Reaktion des Systems auf den *zweidimensionalen Einheitsimpuls*

$$d(k, l) = \begin{cases} 1 & \text{für } k = l = 0 \\ 0 & \text{sonst} \end{cases} \quad (235.11)$$

als Eingangssignal. Das System ist *stabil*, wenn die Bedingung

$$\sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |h(k, l)| < \infty \quad (235.12)$$

gilt. Die strenge Form der *Kausalität* liegt nach (233.12) im Fall

$$h(k, l) = 0 \quad \text{für } k < 0 \quad \text{und/oder } l < 0 \quad (235.13)$$

vor. Bei verschiedenen Anwendungen wird auch von der Definition

$$h(k, l) = 0 \quad \text{für alle } k \quad \text{und } l < 0 \quad \text{sowie } k < 0 \quad \text{und } l = 0 \quad (235.14)$$

Gebrauch gemacht. Sie entspricht der weniger strengen Form der Kausalität aus (233.13) für $n_1 = 1$ und $n_2 = k$. Die Abbildung 235-1 verdeutlicht den Vorgang der Faltung (235.10) unter Berücksichtigung von (235.13) und (235.14). Das m, n -Koordinatensystem ist derart orientiert, daß die positive m -Achse nach unten weist und die positive n -Achse nach rechts zeigt. Berechnet wird das Ausgangssignal (\times) an der Stelle $m = m_1$ und $n = n_1$ jeweils aus den Eingangssignalwerten (\bullet) des unterlegten Bereichs, links im Fall der strengen Kausalität nach (235.13) und rechts für die weniger strenge Form nach (235.14).

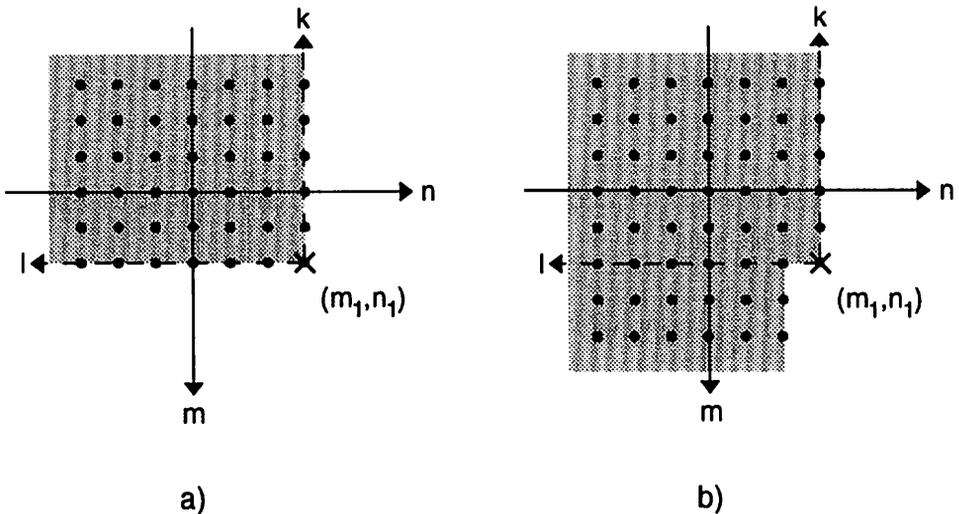


Abb. 235-1: a) streng kausale Faltung nach (235.13), b) kausale Faltung nach (235.14)

Ist die Impulsantwort $h(k, l)$ gegeben und will man die Frequenzantwort $H(j\Omega) = H(j\Omega_1, j\Omega_2)$ mit $\Omega_i = \omega_i \Delta t_i$ berechnen, ist nach (233.18) vorzugehen

$$H(j\Omega_1, j\Omega_2) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h(k, l) e^{-j(k\Omega_1 + l\Omega_2)} \quad (235.15)$$

Liegt andererseits die Frequenzantwort $H(j\Omega_1, j\Omega_2)$ vor, so erhält man aus der zweifachen Integration in (233.20) die Impulsantwort

$$h(\mathbf{k}, \mathbf{l}) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} H(j\Omega_1, j\Omega_2) e^{j(\mathbf{k}\Omega_1 + \mathbf{l}\Omega_2)} d\Omega_1 d\Omega_2 . \quad (235.16)$$

Beispiel 1: Gegeben sei das ideale *zirkulare Tiefpaßfilter* mit der Frequenzantwort

$$H(j\Omega_1, j\Omega_2) = \begin{cases} 1 & \text{für } \Omega_1^2 + \Omega_2^2 = \Omega^2 \leq \Omega_m^2 \\ 0 & \text{für } \Omega_m^2 < \Omega^2 \leq \pi^2 . \end{cases} \quad (235.17)$$

Wegen $H(j\Omega_1, j\Omega_2) = H(j\Omega)$ ist diese Frequenzantwort *rotationssymmetrisch*. Die Impulsantwort berechnet sich aus (235.16) nach Einführung der Polarkoordinaten

$$\Omega = (\Omega_1^2 + \Omega_2^2)^{1/2} \quad \text{und} \quad \varphi = \arctan(\Omega_2/\Omega_1)$$

zu

$$h(m, n) = \frac{1}{4\pi^2} \int_0^{\Omega_m} \int_0^{2\pi} \Omega \exp(j(m\Omega_1 + n\Omega_2)) d\varphi d\Omega . \quad (235.18)$$

Mit den Definitionen $q = (m^2 + n^2)^{1/2}$ und $\tan\theta = n/m$ folgt wegen $\cos\theta = m/q$ und $\cos\varphi = \Omega_1/\Omega$ der Zusammenhang

$$q\Omega \frac{\cos\theta \cos\varphi}{m\Omega_1} = 1 .$$

Erweitert man den Exponent in (235.18) um diesen Ausdruck, erhält man unter Anwendung des Additionstheorems

$$\begin{aligned} h(m, n) &= \frac{1}{4\pi^2} \int_0^{\Omega_m} \int_0^{2\pi} \Omega \exp(jq\Omega \cos\theta \cos\varphi (1 + \frac{n\Omega_2}{m\Omega_1})) d\varphi d\Omega \\ &= \frac{1}{4\pi^2} \int_0^{\Omega_m} \int_0^{2\pi} \Omega \exp(jq\Omega \cos(\varphi - \theta)) d\varphi d\Omega \\ &= \frac{1}{2\pi} \int_0^{\Omega_m} \Omega J_0(q\Omega) d\Omega . \end{aligned}$$

Mit der Variablentransformation $x = q\Omega$ folgt schließlich

$$h(m, n) = \frac{1}{2\pi q^2} \int_0^{\Omega_m} x J_0(x) dx = \frac{\Omega_m}{2\pi q} J_1(q\Omega_m) . \quad (235.19)$$

Hierin bedeuten $J_0(x)$ und $J_1(x)$ die Besselfunktionen erster Art der Ordnung Null und

der Ordnung Eins (Stumpf und Schuler 1973, S.488). Eine graphische Darstellung der Impulsantwort (235.19) ist bei (Dudgeon und Mersereau 1984, S.31) zu finden. Δ

Bildet man die z-Transformation der Faltung (235.10), erhält man aus (234.2) und (234.3) mit $z = [z_1, z_2]$

$$Y(z_1, z_2) = H(z_1, z_2)X(z_1, z_2) \quad , \quad (235.20)$$

wobei

$$X(z_1, z_2) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} x(k, l) z_1^{-k} z_2^{-l} \quad (235.21)$$

$$Y(z_1, z_2) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} y(k, l) z_1^{-k} z_2^{-l} \quad (235.22)$$

die z-Transformationen des Eingangssignals $x(k, l)$ und des Ausgangssignals $y(k, l)$ und

$$H(z_1, z_2) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h(k, l) z_1^{-k} z_2^{-l} \quad (235.23)$$

die zweidimensionale Systemfunktion bedeuten.

Wird die n-dimensionale Differenzengleichung (233.21) für $n = 2$ aufgestellt, erhält man mit $a(k) = a_{k1}$ und $b(l) = b_{pq}$

$$\sum_{k=0}^K \sum_{l=0}^L a_{k1} y(m-k, n-l) = \sum_{p=0}^P \sum_{q=0}^Q b_{pq} x(m-p, n-q) \quad . \quad (235.24)$$

Hieraus ergibt sich mit $a_{00} \neq 0$ die *rekursive Filtergleichung*

$$y(m, n) = \frac{1}{a_{00}} \left[\sum_{p=0}^P \sum_{q=0}^Q b_{pq} x(m-p, n-q) - \sum_{\substack{k=0 \\ (k, l) \neq (0, 0)}}^K \sum_{l=0}^L a_{k1} y(m-k, n-l) \right] \quad , \quad (235.25)$$

die nach (234.6) in die Form

$$y(m, n) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} h(k, l) x(m-k, n-l) \quad (235.26)$$

umgeschrieben werden kann. Die zugehörige Systemfunktion $H(z_1, z_2)$ erhält man aus (235.23) zu

$$H(z_1, z_2) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} h(k, l) z_1^{-k} z_2^{-l} \quad (235.27)$$

oder aus (234.5) zu

$$H(z_1, z_2) = \left(\sum_{p=0}^P \sum_{q=0}^Q b_{pq} z_1^{-p} z_2^{-q} \right) / \left(\sum_{k=0}^K \sum_{l=0}^L a_{kl} z_1^{-k} z_2^{-l} \right) . \quad (235.28)$$

Mit $a_{k1} = 0$ für $(k, l) \neq (0, 0)$ geht (235.25) über in die *nichtrekursive Filtergleichung*

$$y(m, n) = \sum_{p=0}^P \sum_{q=0}^Q (b_{pq}/a_{00}) x(m-p, n-q) \quad (235.29)$$

mit der Impulsantwort nach (234.9)

$$h(p, q) = b_{pq}/a_{00} . \quad (235.30)$$

Ihre Systemfunktion $H(z_1, z_2)$ berechnet sich aus (234.10) zu

$$H(z_1, z_2) = \sum_{p=0}^P \sum_{q=0}^Q h(p, q) z_1^{-p} z_2^{-q} . \quad (235.31)$$

Die diskrete zweidimensionale Fourier-Transformation erhält man aus (234.11) und (234.12) zu

$$x(m, n) = \frac{1}{N_1 N_2} \sum_{k=0}^{N_1-1} \sum_{l=0}^{N_2-1} X(k, l) e^{j2\pi(km/N_1 + ln/N_2)} \quad (235.32)$$

mit

$$X(k, l) = \sum_{m=0}^{N_1-1} \sum_{n=0}^{N_2-1} x(m, n) e^{-j2\pi(km/N_1 + ln/N_2)} \quad (235.33)$$

für

$$k, m \in \{0, \dots, N_1-1\} \quad \text{und} \quad l, n \in \{0, \dots, N_2-1\} .$$

Die Transformation (235.32) wird auch als inverse diskrete zweidimensionale Fourier-Transformation bezeichnet.

Für das Eingangssignal $x(m, n)$ und die Impulsantwort $h(m, n)$ eines nach (235.13) streng kausalen Systems gelte

$$x(m, n) \neq 0 \quad \text{für} \quad 0 \leq m \leq M_1-1 \quad \text{und} \quad 0 \leq n \leq M_2-1 \quad (235.34)$$

$$h(m, n) \neq 0 \quad \text{für} \quad 0 \leq m \leq K_1-1 \quad \text{und} \quad 0 \leq n \leq K_2-1 . \quad (235.35)$$

Das Ausgangssignal $y(m, n)$ folgt aus der linearen Faltung (234.19)

$$y(m, n) = \sum_{k=0}^{M_1-1} \sum_{l=0}^{M_2-1} x(k, l) h(m-k, n-l) \quad (235.36)$$

für $0 \leq m \leq N_1-1$ und $0 \leq n \leq N_2-1$ mit $N_1-1 = M_1+K_1-2$ und $N_2-1 = M_2+K_2-2$.

Als Alternative zur linearen Faltung bietet sich die schnelle Faltung nach (234.21) und (234.22) an. Bildet man die diskrete Fourier-Transformation von (235.36), erhält man nach (234.21)

$$Y(k, l) = H(k, l)X(k, l) \quad (235.37)$$

mit $X(k, l)$ aus (235.33) und nach (234.22) das Ausgangssignal

$$y(m, n) = \frac{1}{N_1 N_2} \sum_{k=0}^{N_1-1} \sum_{l=0}^{N_2-1} Y(k, l) e^{j2\pi(km/N_1 + ln/N_2)} \quad (235.38)$$

Hiermit sind die Grundlagen für die zweidimensionale Filterung gewonnen.

24 Zweidimensionale digitale Filter

241 Allgemeine Definitionen und Eigenschaften

Bei vielen Anwendungen, beispielsweise in der digitalen Bildverarbeitung, sind zweidimensionale Signale zu analysieren. Im folgenden werden daher die zweidimensionalen digitalen linearen Filter behandelt. Die Grundlagen der mehrdimensionalen Filterung sind bei (Dudgeon und Mersereau 1984; Schmidt 1993a) erläutert.

Zur Beschreibung zweidimensionaler linearer digitaler Filter geht man von der Differenzgleichung (235.24) aus, die mit den Substitutionen $g_{kl} = a_{kl}/a_{00}$ und $d_{pq} = b_{pq}/a_{00}$ lautet

$$\sum_{k=0}^K \sum_{l=0}^L g_{kl} y(m-k, n-l) = \sum_{p=0}^P \sum_{q=0}^Q d_{pq} x(m-p, n-q) . \quad (241.1)$$

Entsprechend folgt die Systemfunktion $H(z_1, z_2)$ aus (235.28)

$$H(z_1, z_2) = \left(\sum_{p=0}^P \sum_{q=0}^Q d_{pq} z_1^{-p} z_2^{-q} \right) / \left(\sum_{k=0}^K \sum_{l=0}^L g_{kl} z_1^{-k} z_2^{-l} \right) . \quad (241.2)$$

Löst man (241.1) nach $y(m, n)$ auf, erhält man wegen $g_{00} = 1$

$$y(m, n) = \sum_{p=0}^P \sum_{q=0}^Q d_{pq} x(m-p, n-q) - \sum_{\substack{k=0 \\ (k,l) \neq (0,0)}}^K \sum_{l=0}^L g_{kl} y(m-k, n-l) . \quad (241.3)$$

Nach (235.26) läßt sich für $y(m, n)$ auch schreiben

$$y(m, n) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} h(k, l) x(m-k, n-l) . \quad (241.4)$$

Die Impulsantwort $h(k, l)$ unterliegt also der Bedingung (235.13) der strengen Kausalität

$$h(k, l) = 0 \quad \text{für } k < 0 \quad \text{und/oder} \quad l < 0 , \quad (241.5)$$

weswegen (241.3) die Definitionsgleichung eines kausalen linearen digitalen Filters darstellt. Da (241.5) gleichbedeutend ist mit

$$h(k, l) \neq 0 \quad \text{für } k \geq 0 \quad \text{und} \quad l \geq 0 , \quad (241.6)$$

bezeichnet man (241.3) auch als *Filterung in Richtung des ersten Quadranten* (Cappellini et al. 1978, S.28). Hierfür sind Anfangswerte notwendig, wie Abbildung 241-1 für

$m, n \geq 0$ sowie $P = K$ und $Q = L$ zeigt. Vielfach wird $x(m, n) = y(m, n) = 0$ für $m < 0$ und / oder $n < 0$ gewählt.

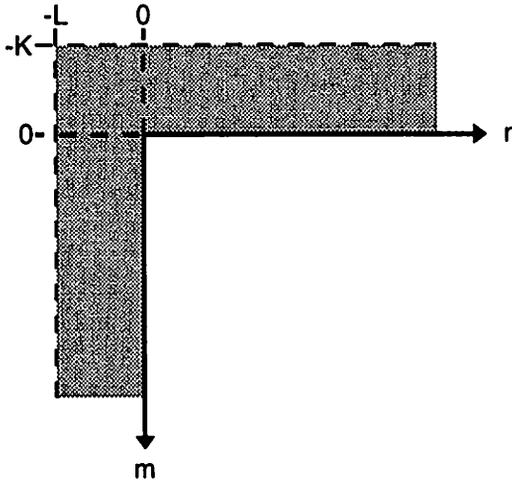


Abb. 241-1: Anfangswerte der kausalen Filterung

Setzt man in (241.1) $P = K$ sowie $Q = L$ und löst nach $y(m-K, n)$ auf, erhält man mit $g_{K0} \neq 0$

$$y(m-K, n) = \frac{1}{g_{K0}} \left[\sum_{p=0}^K \sum_{q=0}^L d_{pq} x(m-p, n-q) - \sum_{\substack{k=0 \\ (k,1) \neq (K,0)}}^K \sum_{l=0}^L g_{kl} y(m-k, n-l) \right].$$

Substituiert man $r = m-K$ und $m = r+K$, ergibt sich weiter

$$y(r, n) = \frac{1}{g_{K0}} \left[\sum_{p=0}^K \sum_{q=0}^L d_{pq} x(r+K-p, n-q) - \sum_{\substack{k=0 \\ (k,1) \neq (K,0)}}^K \sum_{l=0}^L g_{kl} y(r+K-k, n-l) \right]$$

und schließlich mit $u = K-p$, $p = K-u$, $v = K-k$ und $k = K-v$ sowie $d_{K-u, q} / g_{K0} = \bar{d}_{uq}$ und $g_{K-v, l} / g_{K0} = \bar{g}_{vl}$

$$y(r, n) = \sum_{u=0}^K \sum_{q=0}^L \bar{d}_{uq} x(r+u, n-q) - \sum_{\substack{v=0 \\ (v,1) \neq (0,0)}}^K \sum_{l=0}^L \bar{g}_{vl} y(r+v, n-l). \quad (241.7)$$

Dies ist die Filtergleichung eines wegen (235.13) nichtkausalen linearen digitalen Filters. Schreibt man nämlich (241.7) in der Form (235.10), erhält man

$$y(r, n) = \sum_{k=-\infty}^0 \sum_{l=0}^{\infty} h(k, l) x(r-k, n-l) \quad (241.8)$$

und für die Impulsantwort $h(k, l)$ die Bedingung

$$h(k, l) \begin{cases} \neq 0 & \text{für } k \leq 0 \text{ und } l \geq 0 \\ = 0 & \text{sonst .} \end{cases} \quad (241.9)$$

Man bezeichnet (241.7) daher auch als *Filterung in Richtung des zweiten Quadranten*. Die Systemfunktion ist durch (241.2) gegeben. Stellt man sie durch die Koeffizienten \bar{d}_{pq} und \bar{g}_{kl} dar, erhält man

$$H(z_1, z_2) = \left(\sum_{p=0}^K \sum_{q=0}^L \bar{d}_{pq} z_1^p z_2^{-q} \right) / \left(\sum_{k=0}^K \sum_{l=0}^L \bar{g}_{kl} z_1^k z_2^{-l} \right) . \quad (241.10)$$

Analog zu der Vorgehensweise, die auf (241.7) führt, kann man (241.1) auch nach $y(m-K, n-L)$ auflösen

$$y(m-K, n-L) = \frac{1}{\bar{g}_{KL}} \left[\sum_{p=0}^K \sum_{q=0}^L d_{pq} x(m-p, n-q) - \sum_{\substack{k=0 \\ (k,l) \neq (K,L)}}^K \sum_{l=0}^L g_{kl} y(m-k, n-l) \right] .$$

Substituiert man $r = m-K$, $m = r+K$, $s = n-L$ und $n = s+L$ erhält man

$$y(r, s) = \frac{1}{\bar{g}_{KL}} \left[\sum_{p=0}^K \sum_{q=0}^L d_{pq} x(r+K-p, s+L-q) - \sum_{\substack{k=0 \\ (k,l) \neq (K,L)}}^K \sum_{l=0}^L g_{kl} y(r+K-k, s+L-l) \right] .$$

Setzt man weiterhin noch $t = K-p$, $p = K-t$, $u = L-q$, $q = L-u$, $v = K-k$, $k = K-v$, $w = L-l$ und $l = L-w$ und führt die Koeffizienten $d_{K-t, L-u} / \bar{g}_{KL} = \bar{\bar{d}}_{tu}$ sowie $g_{K-v, L-w} / \bar{g}_{KL} = \bar{\bar{g}}_{vw}$ ein, ergibt sich

$$y(r, s) = \sum_{t=0}^K \sum_{u=0}^L \bar{\bar{d}}_{tu} x(r+t, s+u) - \sum_{\substack{v=0 \\ (v,w) \neq (0,0)}}^K \sum_{w=0}^L \bar{\bar{g}}_{vw} y(r+v, s+w) . \quad (241.11)$$

Dies entspricht einer *Filterung in Richtung des dritten Quadranten*, denn für $y(r, s)$ kann auch geschrieben werden

$$y(r, s) = \sum_{k=-\infty}^0 \sum_{l=-\infty}^0 h(k, l) x(r-k, s-l) . \quad (241.12)$$

Für die Impulsantwort $h(k, l)$ gilt also die Bedingung

$$h(k, l) \begin{cases} \neq 0 & \text{für } k \leq 0 \text{ und } l \leq 0 \\ = 0 & \text{sonst .} \end{cases} \quad (241.13)$$

Die Systemfunktion (241.2) lautet mit den Koeffizienten $\bar{\bar{d}}_{pq}$ und $\bar{\bar{g}}_{kl}$

$$H(z_1, z_2) = \left(\sum_{p=0}^K \sum_{q=0}^L \bar{d}_{pq} z_1^p z_2^q \right) / \left(\sum_{k=0}^K \sum_{l=0}^L \bar{g}_{kl} z_1^k z_2^l \right) . \quad (241.14)$$

Löst man schließlich noch (241.1) nach $y(m, n-L)$ auf, substituiert $s = n-L$ und $n = L+s$, erhält man

$$y(m, s) = \frac{1}{g_{0L}} \left[\sum_{p=0}^K \sum_{q=0}^L d_{pq} x(m-p, s+L-q) - \sum_{\substack{k=0 \\ (k,l) \neq (0,L)}}^K \sum_{l=0}^L g_{kl} y(m-k, s+L-l) \right] .$$

Setzt man dann $u = L-q$, $q = L-u$, $v = L-l$ und $l = L-v$, ergibt sich mit den Koeffizienten $\bar{d}_{pu} = d_{p, L-u} / g_{0L}$ und $\bar{g}_{kv} = g_{k, L-v} / g_{0L}$ mit

$$y(m, s) = \sum_{p=0}^K \sum_{u=0}^L \bar{d}_{pu} x(m-p, s+u) - \sum_{\substack{k=0 \\ (k,v) \neq (0,0)}}^K \sum_{v=0}^L \bar{g}_{kv} y(m-k, s+v) \quad (241.15)$$

die *Filterung in Richtung des vierten Quadranten*, denn man kann auch schreiben

$$y(m, s) = \sum_{k=0}^{\infty} \sum_{l=-\infty}^0 h(k, l) x(m-k, s-l) . \quad (241.16)$$

Für die Impulsantwort gilt demnach

$$h(k, l) \begin{cases} \neq 0 & \text{für } k \geq 0 \text{ und } l \leq 0 \\ = 0 & \text{sonst} . \end{cases} \quad (241.17)$$

Außerdem erhält man die Systemfunktion (241.2) mit den Koeffizienten \bar{d}_{pq} und \bar{g}_{kl} zu

$$H(z_1, z_2) = \left(\sum_{p=0}^K \sum_{q=0}^L \bar{d}_{pq} z_1^{-p} z_2^q \right) / \left(\sum_{k=0}^K \sum_{l=0}^L \bar{g}_{kl} z_1^{-k} z_2^l \right) . \quad (241.18)$$

Mit (241.7), (241.11) und (241.15) werden nichtkausale Filtergleichungen erhalten. Während im eindimensionalen Fall die Kausalität für die Echtzeitverarbeitung wichtig ist, spielt sie im zweidimensionalen Fall, beispielsweise in der digitalen Bildverarbeitung, eine untergeordnete Rolle. Ein digitales Bild liegt nämlich im allgemeinen bereits vor der Filterung vollständig abgespeichert vor.

Werden P zweidimensionale Teilfilter mit den Teilsystemfunktionen $H_p(z_1, z_2)$ mit $p \in \{1, \dots, P\}$ und $P > 1$ hintereinandergeschaltet, besitzt das Gesamtfiter mit der Systemfunktion

$$H(z_1, z_2) = \prod_{p=1}^P H_p(z_1, z_2) \quad (241.19)$$

die Kaskadenstruktur analog zu (221.16). Die Impulsantwort $h(m,n)$ ergibt sich analog zu (235.10) und (235.20) aus der Faltung der P Teilimpulsantworten $h_p(m,n)$

$$h(m,n) = h_1(m,n) * h_2(m,n) * \dots * h_p(m,n) . \quad (241.20)$$

Bei Parallelschaltung der P Teilfilter erhält man die Systemfunktion des Gesamtfilters entsprechend (221.17) aus

$$H(z_1, z_2) = \sum_{p=1}^P H_p(z_1, z_2) . \quad (241.21)$$

Es wird nun die Stabilitätseigenschaft eines zweidimensionalen digitalen Filters betrachtet. Der folgende Satz, der auch unter der Bezeichnung *Theorem von Shanks* (Wahl 1980, S.10) bekannt ist, beschränkt sich auf streng kausale Systeme.

Satz: Ein nach (241.5) streng kausales lineares digitales Filter ist genau dann stabil, wenn es kein Wertepaar (z_1, z_2) mit $|z_1| \geq 1$ und gleichzeitig $|z_2| \geq 1$ gibt, für das der Nenner der Systemfunktion (241.2) verschwindet. (241.22)

Beweis: Nach (235.27) und (235.28) gilt für die Systemfunktion (241.2) mit $z_i = r_i e^{j\Omega_i}$ für $i \in \{1, 2\}$ und $|z_i| = r_i$

$$\begin{aligned} H(z_1, z_2) &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} h(k, l) z_1^{-k} z_2^{-l} \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} h(k, l) r_1^{-k} r_2^{-l} e^{-j(k\Omega_1 + l\Omega_2)} . \end{aligned} \quad (241.23)$$

Die Bedingung für die Konvergenz von $H(z_1, z_2)$ lautet analog zu (221.20)

$$|H(z_1, z_2)| \leq \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} |h(k, l) r_1^{-k} r_2^{-l}| < \infty . \quad (241.24)$$

Nimmt man an, daß der Nenner in (241.2) ungleich Null ist, dann ist für $r_1 \geq 1$ und $r_2 \geq 1$ die Bedingung (241.24) erfüllt und liefert für $r_1 = r_2 = 1$ die Stabilitätsforderung (235.12). Setzt man andererseits die Stabilität nach (235.12) voraus, erhält man wegen

$$\left| \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} h(k, l) \right| \leq \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} |h(k, l)| < \infty$$

und mit $r_1 \geq 1$ und $r_2 \geq 1$

$$\left| \sum_{n=0}^{\infty} \sum_{l=0}^{\infty} h(k, l) r_1^{-k} r_2^{-l} \right| \leq \left| \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} h(k, l) \right| < \infty$$

die Konvergenzbedingung (241.24), so daß der Nenner in (241.2) von Null verschieden ist. □

Die Stabilitätsprüfung erweist sich allerdings als wenig praktikabel, denn es ist zum Beispiel für alle komplexen Zahlen $z_2 = r_2 e^{j\Omega_2}$ mit $r_2 \geq 1$ zu untersuchen, ob eine komplexe Zahl $z_1 = r_1 e^{j\Omega_1}$ mit $r_1 \geq 1$ existiert, die den Nenner von (241.2) zu Null werden läßt

$$\sum_{k=0}^K \sum_{l=0}^L g_{kl} r_1^{-k} r_2^{-l} e^{-j(k\Omega_1 + l\Omega_2)} = 0 .$$

Sollte dies der Fall sein, ist das Filter instabil. Neben einer vereinfachten Form des vorgestellten Theorems von Shanks, das von Huang abgeleitet wurde, existieren weitere Stabilitätstheoreme, die beispielsweise bei (Dudgeon und Mersereau 1984, S.191) behandelt werden.

Soll aus der z-Transformation

$$Y_1(z_1, z_2) = H_1(z_1, z_2)X(z_1, z_2) \quad (241.25)$$

die z-Transformation $X(z_1, z_2)$ zurückgewonnen werden, benötigt man analog zu (221.26) die Systemfunktion der *inversen Filterung* oder *Dekonvolution*

$$H_2(z_1, z_2) = (H_1(z_1, z_2))^{-1} , \quad (241.26)$$

denn aus (241.19) folgt

$$H(z_1, z_2) = H_1(z_1, z_2)H_2(z_1, z_2) = 1 . \quad (241.27)$$

Um nach (241.22) zu überprüfen, ob $H_2(z_1, z_2)$ die Systemfunktion eines stabilen und streng kausalen Systems darstellt, ist also der Zähler von $H_1(z_1, z_2)$ auf seine Nullstellen zu untersuchen.

Ein System bezeichnet man als *separierbar*, wenn für seine Systemfunktion $H(z_1, z_2)$ gilt

$$H(z_1, z_2) = H_1(z_1)H_2(z_2) . \quad (241.28)$$

Dann läßt sich aus zwei eindimensionalen Systemfunktionen mittels der Kaskadenstruktur (241.19) eine zweidimensionale Systemfunktion bilden.

Analog zu (221.30) gilt für die komplexe Systemfunktion $H(z_1, z_2)$ mit $z_i = e^{j\Omega_i}$ für $i \in \{1, 2\}$ die Darstellung

$$H(j\Omega_1, j\Omega_2) = A(\Omega_1, \Omega_2)e^{j\Phi(\Omega_1, \Omega_2)} \quad (241.29)$$

mit der *Amplitude*

$$A(\Omega_1, \Omega_2) = |H(j\Omega_1, j\Omega_2)| = (H(j\Omega_1, j\Omega_2)H^*(j\Omega_1, j\Omega_2))^{1/2} \quad (241.30)$$

und der *Phase*

$$\Phi(\Omega_1, \Omega_2) = \arctan(\text{Im}(H(j\Omega_1, j\Omega_2))/\text{Re}(H(j\Omega_1, j\Omega_2))) \quad (241.31)$$

Von besonderer Bedeutung sind digitale Filter mit *linearer Phase*

$$\Phi(\Omega_1, \Omega_2) = \alpha_1\Omega_1 + \alpha_2\Omega_2 + \alpha_3 \quad (241.32)$$

wobei α_1 , α_2 und α_3 beliebige reelle Koeffizienten bedeuten. Setzt man (241.29) mit (241.32) in (235.16) ein, erhält man

$$h(m, n) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} A(\Omega_1, \Omega_2) e^{j((m+\alpha_1)\Omega_1 + (n+\alpha_2)\Omega_2 + \alpha_3)} d\Omega_1 d\Omega_2 \quad (241.33)$$

Wie im Kapitel 221 werden im folgenden für diese Impulsantwort vier verschiedene Fälle behandelt, wobei für die Koeffizienten α_1 , α_2 und α_3 bestimmte Annahmen getroffen werden. Für

$$a) \alpha_1 \text{ beliebig, } \alpha_2 \text{ beliebig und } \alpha_3 = 0 \quad (241.34)$$

erhält man, da $h(m, n)$ eine reelle Funktion ist, mit (212.8) aus (241.33)

$$h(m, n) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} A(\Omega_1, \Omega_2) \cos((m+\alpha_1)\Omega_1 + (n+\alpha_2)\Omega_2) d\Omega_1 d\Omega_2 \quad (241.35)$$

Substituiert man hierin $m+\alpha_1 = k$, $m = k - \alpha_1$, $n+\alpha_2 = l$ und $n = l - \alpha_2$, folgt wegen

$$h(k - \alpha_1, l - \alpha_2) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} A(\Omega_1, \Omega_2) \cos(k\Omega_1 + l\Omega_2) d\Omega_1 d\Omega_2 \quad (241.36)$$

die Symmetriebedingung

$$h(k - \alpha_1, l - \alpha_2) = h(-k - \alpha_1, -l - \alpha_2) \quad (241.37)$$

Der Symmetriepunkt liegt also bei $(-\alpha_1, -\alpha_2)$. Für

$$b) \alpha_1 = \alpha_2 = \alpha_3 = 0 \quad (241.38)$$

erhält man wegen

$$\Phi(\Omega_1, \Omega_2) = 0 \quad (241.39)$$

nach (241.32) aus (241.29) die reelle Frequenzantwort

$$H(j\Omega_1, j\Omega_2) = A(\Omega_1, \Omega_2) \quad (241.40)$$

des *nullphasigen Filters*. Aus (241.35) ergibt sich die Impulsantwort

$$h(k, l) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} A(\Omega_1, \Omega_2) \cos(k\Omega_1 + l\Omega_2) d\Omega_1 d\Omega_2 = h(-k, -l) . \quad (241.41)$$

Im Fall

$$c) \alpha_1 \text{ beliebig, } \alpha_2 \text{ beliebig und } \alpha_3 = \pi/2 \quad (241.42)$$

erhält man aus (241.33) wegen $\cos(x+\pi/2) = -\sin x$

$$h(m, n) = - \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} A(\Omega_1, \Omega_2) \sin((m+\alpha_1)\Omega_1 + (n+\alpha_2)\Omega_2) d\Omega_1 d\Omega_2 \quad (241.43)$$

und mit den Substitutionen $m+\alpha_1 = k$, $m = k-\alpha_1$, $n+\alpha_2 = l$ sowie $n = l-\alpha_2$

$$h(k-\alpha_1, l-\alpha_2) = - \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} A(\Omega_1, \Omega_2) \sin(k\Omega_1 + l\Omega_2) d\Omega_1 d\Omega_2 . \quad (241.44)$$

Die Impulsantwort ist antisymmetrisch bezüglich des Punktes $(-\alpha_1, -\alpha_2)$. Schließlich wird noch der Fall

$$d) \alpha_1 = \alpha_2 = 0 \text{ und } \alpha_3 = \pi/2 \quad (241.45)$$

untersucht. Es ergibt sich mit

$$\Phi(\Omega_1, \Omega_2) = \pi/2 \quad (241.46)$$

aus (241.32) die wegen (212.8) imaginäre Frequenzantwort

$$H(j\Omega_1, j\Omega_2) = jA(\Omega_1, \Omega_2) . \quad (241.47)$$

Die Impulsantwort erhält man aus (241.43) zu

$$h(k, l) = - \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} A(\Omega_1, \Omega_2) \sin(k\Omega_1 + l\Omega_2) d\Omega_1 d\Omega_2 = -h(-k, -l) , \quad (241.48)$$

sie ist antisymmetrisch.

Ein System, das die Symmetriebedingung (241.41)

$$h(k, l) = h(-k, -l) \quad (241.49)$$

erfüllt, wird als *zweiquadrantensymmetrisch* bezeichnet. Die Bedingung führt für die Amplitude $A(\Omega_1, \Omega_2)$ auf den Zusammenhang

$$A(\Omega_1, \Omega_2) = A(-\Omega_1, -\Omega_2) . \quad (241.50)$$

Gilt zusätzlich

$$A(\Omega_1, \Omega_2) = A(-\Omega_1, \Omega_2) , \quad (241.51)$$

erhält man aus (241.41) für die Impulsantwort

$$h(k, l) = h(-k, l) = h(k, -l) = h(-k, -l) . \quad (241.52)$$

In diesem Fall spricht man von einem *vierquadrantensymmetrischen* System. Weiterhin bezeichnet man ein System, das auch noch

$$A(\Omega_1, \Omega_2) = A(\Omega_2, \Omega_1) \quad (241.53)$$

erfüllt, als *oktagonal*.

Mit (235.17) wurde eine rotationssymmetrische Frequenzantwort definiert, deren Amplitude sich wegen

$$A(\Omega_1, \Omega_2) = A(\Omega) \quad \text{mit} \quad \Omega^2 = \Omega_1^2 + \Omega_2^2 \quad (241.54)$$

isotrop, also richtungsunabhängig, verhält. Wie man aus (235.27) für $z_i = e^{j\Omega_i}$ mit $i \in \{1, 2\}$ erkennt, sind für ein diskretes k, l -Gitter rotationssymmetrische Frequenzantworten streng nicht realisierbar, da sich der Exponent $k\Omega_1 + l\Omega_2$ als Linearkombination und nicht als Quadratsumme von Ω_1 und Ω_2 ergibt. Es sind jedoch geeignete Approximationen, beispielsweise durch oktagonale Systeme möglich.

Wie für den eindimensionalen Fall im Kapitel 222 beschrieben, kann auch für den zweidimensionalen Fall eine Entwurfsaufgabe formuliert werden.

Gegeben sei die zweidimensionale Idealfunktion

$$f(x_1, x_2) . \quad (241.55)$$

Gesucht ist dann die zweidimensionale Realfunktion

$$f_0(x_1, x_2) \quad (241.56)$$

unter Berücksichtigung der Modellparametermenge

$$M . \quad (241.57)$$

Im zweidimensionalen Fall wird das zu realisierende Filter (241.1) durch die Filterkoeffizienten d_{pq} mit $p \in \{0, \dots, P\}$ und $q \in \{0, \dots, Q\}$ sowie g_{kl} mit $k \in \{0, \dots, K\}$ und $l \in \{0, \dots, L\}$ für $(k, l) \neq (0, 0)$ wegen $g_{00} = 1$ charakterisiert. Faßt man sie mit $(K+1)(L+1) + (P+1)(Q+1) - 1 = u$ im $u \times 1$ Vektor

$$\beta = [d_{00}, d_{10}, d_{20}, \dots, d_{p0}, d_{01}, \dots, d_{p1}, \dots, d_{0Q}, \dots, d_{pQ}, \\ g_{10}, g_{20}, \dots, g_{K0}, g_{01}, \dots, g_{K1}, \dots, g_{0L}, \dots, g_{KL}]' \quad (241.58)$$

zusammen, lautet die Realfunktion (241.56)

$$f_0(x_1, x_2) = f_0(x_1, x_2, \beta) . \quad (241.59)$$

Die Approximationsfehlerfunktion $e(x_1, x_2)$ berechnet sich aus

$$e(x_1, x_2) = f_0(x_1, x_2, \beta) - f(x_1, x_2) \quad (241.60)$$

In Abbildung 222-1 wurden frequenzselektive eindimensionale Filter veranschaulicht. Entsprechende zweidimensionale Filter sind beispielsweise bei (Gonzales und Woods 1992, S.202) dargestellt.

Die Lösung der Entwurfsaufgabe besteht also in der Bestimmung des Parametervektors β aus (241.58). Es existieren eine Reihe von Verfahren sowohl für rekursive als auch für nichtrekursive Filter. In den Kapiteln 243 und 246 werden einige einfache Methoden vorgestellt.

242 Definition und Eigenschaften zweidimensionaler rekursiver Filter

In Analogie zum eindimensionalen Fall wird in der Definitionsgleichung (241.3) des streng kausalen linearen digitalen Filters die erste Doppelsumme auf der rechten Seite als nichtrekursiver Anteil und die zweite Doppelsumme als rekursiver Anteil bezeichnet. Man nennt also

$$y(m, n) = \sum_{p=0}^P \sum_{q=0}^Q d_{pq} x(m-p, n-q) - \sum_{\substack{k=0 \\ (k,l) \neq (0,0)}}^K \sum_{l=0}^L g_{kl} y(m-k, n-l) \quad (242.1)$$

mit

$$\begin{aligned} g_{k1} &\neq 0 \quad \text{für mindestens ein } l \\ g_{kL} &\neq 0 \quad \text{für mindestens ein } k \end{aligned} \quad (242.2)$$

die Filtergleichung eines zweidimensionalen streng *kausalen rekursiven linearen digitalen Filters vom Grad $\max\{K, L\}$* . Dieses Filter bewirkt, wie bereits im Zusammenhang mit (241.6) erwähnt, eine Filterung in Richtung des ersten Quadranten. Beispiele für Filtergleichungen zweidimensionaler *nichtkausaler rekursiver linearer digitaler Filter vom Grad $\max\{K, L\}$* in den Richtungen des zweiten, dritten und vierten Quadranten sind mit (241.7), (241.11) und (241.15) gegeben.

Nach (241.4), (241.8), (241.12) und (241.16) werden rekursive Filter durch eine unendlich lange Impulsantwort $h(k, l)$ charakterisiert. Da die von Null verschiedenen Werte jeweils nur auf einen Quadranten beschränkt sind, können die Symmetriebedingungen (241.37) und (241.41) sowie die Antisymmetriebedingungen (241.44) und (241.48) von den bislang vorgestellten rekursiven Filtern nicht erfüllt werden. Linearphasige rekursive

Filter in Richtung des i -ten Quadranten mit $i \in \{1, 2, 3, 4\}$ sind demnach nicht realisierbar.

Mit der Kaskadenstruktur (241.19) lassen sich jedoch nullphasige rekursive Filter konstruieren, deren reelle Frequenzantwort nach (241.40) lautet

$$H(j\Omega_1, j\Omega_2) = A(\Omega_1, \Omega_2) \quad (242.3)$$

Bezeichnet man die Frequenzantwort eines Filters in Richtung des ersten Quadranten mit $H_1(j\Omega_1, j\Omega_2)$ folgt mit

$$H_3(j\Omega_1, j\Omega_2) = H_1(-j\Omega_1, -j\Omega_2) \quad (242.4)$$

die Frequenzantwort eines Filters in Richtung des dritten Quadranten, wie ein Vergleich von (241.2) mit (241.14) zeigt. Wegen

$$H_3(j\Omega_1, j\Omega_2) = H_1^*(j\Omega_1, j\Omega_2) \quad (242.5)$$

erhält man nach (241.30) eine reelle Frequenzantwort

$$\begin{aligned} H(j\Omega_1, j\Omega_2) &= H_1(j\Omega_1, j\Omega_2)H_3(j\Omega_1, j\Omega_2) \\ &= A(\Omega_1, \Omega_2) = Z(\Omega_1, \Omega_2)/N(\Omega_1, \Omega_2). \end{aligned} \quad (242.6)$$

Die Frequenzantwort $H_1(j\Omega_1, j\Omega_2)$ des kausalen Filters in Richtung des ersten Quadranten ergibt sich aus (241.2) für $z_i = e^{j\Omega_i}$ mit $i \in \{1, 2\}$ sowie $P = K$ und $Q = L$ zu

$$H_1(j\Omega_1, j\Omega_2) = \left(\sum_{p=0}^K \sum_{q=0}^L d_{pq} e^{-j(p\Omega_1 + q\Omega_2)} \right) / \left(\sum_{k=0}^K \sum_{l=0}^L g_{kl} e^{-j(k\Omega_1 + l\Omega_2)} \right) \quad (242.7)$$

Wegen (242.4) lautet daher $H_3(j\Omega_1, j\Omega_2)$

$$H_3(j\Omega_1, j\Omega_2) = \left(\sum_{p=0}^K \sum_{q=0}^L d_{pq} e^{j(p\Omega_1 + q\Omega_2)} \right) / \left(\sum_{k=0}^K \sum_{l=0}^L g_{kl} e^{j(k\Omega_1 + l\Omega_2)} \right) \quad (242.8)$$

Vergleicht man diesen Ausdruck mit (241.14), erhält man für die Filterkoeffizienten

$$\bar{d}_{pq} = d_{pq} \quad \text{und} \quad \bar{g}_{kl} = g_{kl} \quad (242.9)$$

Die Filtergleichungen des mit (242.6) definierten nullphasigen Systems sind also durch (241.3) für $P = K$ und $Q = L$

$$y_1(m, n) = \sum_{p=0}^K \sum_{q=0}^L d_{pq} y_0(m-p, n-q) - \sum_{\substack{k=0 \\ (k,1) \neq (0,0)}}^K \sum_{l=0}^L g_{kl} y_1(m-k, n-1) \quad (242.10)$$

sowie durch (241.11) unter Berücksichtigung von (242.9) gegeben

$$y_3(m, n) = \sum_{p=0}^K \sum_{q=0}^L d_{pq} y_1(m+p, n+q) - \sum_{\substack{k=0 \\ (k, l) \neq (0, 0)}}^K \sum_{l=0}^L g_{kl} y_3(m+k, n+l) . \quad (242.11)$$

Hierin bedeuten

$$y_0(m, n) = x(m, n) \quad (242.12)$$

das Eingangssignal, $y_1(m, n)$ das Ausgangssignal des Filters in Richtung des ersten Quadranten und

$$y_3(m, n) = y(m, n) \quad (242.13)$$

das Ausgangssignal des Filters in Richtung des dritten Quadranten.

Wird im Frequenzbereich gefiltert, sind nach (242.6) die Frequenzantworten (242.7) und (242.8) miteinander zu multiplizieren. Dies geschieht im folgenden, um für die Frequenzantwort $H(j\Omega_1, j\Omega_2)$ einen einfachen Ausdruck zu erhalten. Betrachtet man zunächst den Zähler $Z(\Omega_1, \Omega_2)$, folgt

$$\begin{aligned} Z(\Omega_1, \Omega_2) &= \left(\sum_{p=0}^K \sum_{q=0}^L d_{pq} e^{-j(p\Omega_1 + q\Omega_2)} \right) \left(\sum_{k=0}^K \sum_{l=0}^L d_{kl} e^{j(k\Omega_1 + l\Omega_2)} \right) \\ &= \sum_{p=0}^K \sum_{q=0}^L \sum_{k=0}^K \sum_{l=0}^L d_{pq} d_{kl} e^{-j((p-k)\Omega_1 + (q-l)\Omega_2)} \end{aligned}$$

und mit den Substitutionen $p-k = m$ sowie $q-l = n$

$$Z(\Omega_1, \Omega_2) = \sum_{m=-K}^K \sum_{n=-L}^L a_{mn} e^{-j(m\Omega_1 + n\Omega_2)} , \quad (242.14)$$

wobei sich die Koeffizienten a_{mn} aus

$$a_{mn} = \sum_{\substack{p=0 \\ p-k=m}}^K \sum_{k=0}^K \sum_{q=0}^L \sum_{l=0}^L d_{pq} d_{kl} = a_{-m, -n} \quad (242.15)$$

berechnen. Berücksichtigt man dieses Symmetrieverhalten in (242.14), erhält man mit (212.11)

$$Z(\Omega_1, \Omega_2) = a_{00} + 2 \sum_{n=1}^L a_{0n} \cos n\Omega_2 + 2 \sum_{m=1}^K \sum_{n=-L}^L a_{mn} \cos(m\Omega_1 + n\Omega_2) . \quad (242.16)$$

Auf gleiche Weise erhält man für den Nenner $N(\Omega_1, \Omega_2)$ in (242.6)

$$N(\Omega_1, \Omega_2) = \left(\sum_{p=0}^K \sum_{q=0}^L g_{pq} e^{-j(p\Omega_1 + q\Omega_2)} \right) \left(\sum_{k=0}^K \sum_{l=0}^L g_{kl} e^{j(k\Omega_1 + l\Omega_2)} \right)$$

$$= b_{00} + 2 \sum_{n=1}^L b_{0n} \cos n\Omega_2 + 2 \sum_{m=1}^K \sum_{n=-L}^L b_{mn} \cos(m\Omega_1+n\Omega_2) \quad (242.17)$$

mit den Koeffizienten

$$b_{mn} = \sum_{\substack{p=0 \\ p-k=m}}^K \sum_{k=0}^K \sum_{q=0}^L \sum_{l=0}^L g_{pq} g_{kl} = b_{-m, -n} \quad (242.18)$$

Zähler und Nenner lassen sich noch kompakter schreiben, was für den Zähler gezeigt wird. Definiert man

$$\bar{a}_{00} = a_{00}$$

$$\bar{a}_{mn} = 2a_{mn} \quad \text{für } m \in \{0, \dots, K\}, n \in \{-L, \dots, L\} \quad \text{und } (m, n) \neq (0, 0) ,$$

erhält man anstelle von (242.16)

$$Z(\Omega_1, \Omega_2) = \sum_{n=0}^L \bar{a}_{0n} \cos n\Omega_2 + \sum_{m=1}^K \sum_{n=-L}^L \bar{a}_{mn} \cos(m\Omega_1+n\Omega_2) \quad (242.19)$$

und eine entsprechende Gleichung für $N(\Omega_1, \Omega_2)$. Die reelle Frequenzantwort ergibt sich dann nach (242.6). Für sie gilt

$$A(\Omega_1, \Omega_2) = A(-\Omega_1, -\Omega_2) , \quad (242.20)$$

sie ist also nach (241.50) zweiquadrantensymmetrisch. Definiert man zusätzlich zu (242.4) die Frequenzantworten

$$H_2(j\Omega_1, j\Omega_2) = H_1(-j\Omega_1, j\Omega_2) \quad (242.21)$$

und

$$H_4(j\Omega_1, j\Omega_2) = H_1(j\Omega_1, -j\Omega_2) , \quad (242.22)$$

so erhält man wegen

$$H_4(j\Omega_1, j\Omega_2) = H_2^*(j\Omega_1, j\Omega_2) \quad (242.23)$$

über die Kaskadenstruktur (241.19)

$$\begin{aligned} H(j\Omega_1, j\Omega_2) &= H_2(j\Omega_1, j\Omega_2)H_4(j\Omega_1, j\Omega_2) \\ &= A(\Omega_1, \Omega_2) = Z(\Omega_1, \Omega_2)/N(\Omega_1, \Omega_2) \end{aligned} \quad (242.24)$$

eine weitere reelle Frequenzantwort und somit eine nullphasige Filterung. Wegen (241.10) ist $H_2(j\Omega_1, j\Omega_2)$ die Frequenzantwort eines Filters in Richtung des zweiten Quadranten. Sie ergibt sich mit (242.21) aus (242.7) zu

$$H_2(j\Omega_1, j\Omega_2) = \left(\sum_{p=0}^K \sum_{q=0}^L d_{pq} e^{j(p\Omega_1 - q\Omega_2)} \right) / \left(\sum_{k=0}^K \sum_{l=0}^L g_{kl} e^{j(k\Omega_1 - l\Omega_2)} \right) . \quad (242.25)$$

Schließlich lautet $H_4(j\Omega_1, j\Omega_2)$, also nach (241.18) die Frequenzantwort eines Filters in Richtung des vierten Quadranten

$$H_4(j\Omega_1, j\Omega_2) = \left(\sum_{p=0}^K \sum_{q=0}^L d_{pq} e^{-j(p\Omega_1 - q\Omega_2)} \right) / \left(\sum_{k=0}^K \sum_{l=0}^L g_{kl} e^{-j(k\Omega_1 - l\Omega_2)} \right) . \quad (242.26)$$

Multipliziert man nach (242.24) die Frequenzantwort (242.25) mit (242.26), erhält man für den Zähler analog zu (242.16)

$$Z(\Omega_1, \Omega_2) = a_{00} + 2 \sum_{n=1}^L a_{0n} \cos n\Omega_2 + 2 \sum_{m=1}^K \sum_{n=-L}^L a_{mn} \cos(m\Omega_1 - n\Omega_2) \quad (242.27)$$

und für den Nenner entsprechend (242.17)

$$N(\Omega_1, \Omega_2) = b_{00} + 2 \sum_{n=1}^L b_{0n} \cos n\Omega_2 + 2 \sum_{m=1}^K \sum_{n=-L}^L b_{mn} \cos(m\Omega_1 - n\Omega_2) . \quad (242.28)$$

Die Frequenzantwort $H(j\Omega_1, j\Omega_2)$ ist also wieder wegen

$$A(\Omega_1, -\Omega_2) = A(-\Omega_1, \Omega_2) \quad (242.29)$$

zweiquadrantensymmetrisch. Die Filtergleichungen sind durch (241.7) und (241.15) gegeben, wenn man

$$\bar{d}_{pq} = \tilde{d}_{pq} = d_{pq} \quad \text{und} \quad \bar{g}_{kl} = \tilde{g}_{kl} = g_{kl} \quad (242.30)$$

substituiert.

Mit (242.6) und (242.24) wurden zwei reelle Frequenzantworten abgeleitet und somit zwei nullphasige zweiquadrantensymmetrische Filter entwickelt. Vielfach ist es erwünscht, rotationssymmetrische Frequenzantworten zu nutzen, die wie im Zusammenhang mit (241.54) erwähnt, streng nicht zu realisieren sind. Eine erste Approximation erhält man über ein vierquadrantensymmetrisches Filter. Ein solches Filter läßt sich mit den Frequenzantworten $H_k(j\Omega_1, j\Omega_2)$ mit $k \in \{1, \dots, 4\}$ aus (242.7), (242.25), (242.8) und (242.26) über

$$H(j\Omega_1, j\Omega_2) = \prod_{k=1}^4 H_k(j\Omega_1, j\Omega_2) = A(\Omega_1, \Omega_2) = Z(\Omega_1, \Omega_2) / N(\Omega_1, \Omega_2) \quad (242.31)$$

ableiten. Berechnet man zunächst wieder den Zähler, ergibt sich

$$Z(\Omega_1, \Omega_2) = \left(\sum_{p_1=0}^K \sum_{q_1=0}^L d_{p_1 q_1} e^{-j(p_1 \Omega_1 + q_1 \Omega_2)} \right) \left(\sum_{p_2=0}^K \sum_{q_2=0}^L d_{p_2 q_2} e^{j(p_2 \Omega_1 - q_2 \Omega_2)} \right)$$

$$\begin{aligned}
 & \cdot \left(\sum_{p_3=0}^K \sum_{q_3=0}^L d_{p_3 q_3} e^{j(p_3 \Omega_1 + q_3 \Omega_2)} \right) \left(\sum_{p_4=0}^K \sum_{q_4=0}^L d_{p_4 q_4} e^{-j(p_4 \Omega_1 - q_4 \Omega_2)} \right) \\
 & = \sum_{p_1=0}^K \dots \sum_{p_4=0}^K \sum_{q_1=0}^L \dots \sum_{q_4=0}^L d_{p_1 q_1} d_{p_2 q_2} d_{p_3 q_3} d_{p_4 q_4} \\
 & \quad e^{-j(p_1 - p_2 - p_3 + p_4) \Omega_1 + (q_1 + q_2 - q_3 - q_4) \Omega_2} .
 \end{aligned}$$

Substituiert man hierin $m = p_1 - p_2 - p_3 + p_4$ und $n = q_1 + q_2 - q_3 - q_4$, erhält man

$$Z(\Omega_1, \Omega_2) = \sum_{m=-2K}^{2K} \sum_{n=-2L}^{2L} a_{mn} e^{-j(m\Omega_1 + n\Omega_2)} \quad (242.32)$$

mit den Koeffizienten

$$\begin{aligned}
 a_{mn} & = \sum_{p_1=0}^K \dots \sum_{p_4=0}^K \sum_{q_1=0}^L \dots \sum_{q_4=0}^L d_{p_1 q_1} d_{p_2 q_2} d_{p_3 q_3} d_{p_4 q_4} \\
 & \quad p_1 - p_2 - p_3 + p_4 = m \quad q_1 + q_2 - q_3 - q_4 = n \\
 & = a_{-m, n} = a_{-m, -n} = a_{m, -n} .
 \end{aligned} \quad (242.33)$$

Berücksichtigt man dieses symmetrische Verhalten in (242.32), ergibt sich mit (212.11)

$$\begin{aligned}
 Z(\Omega_1, \Omega_2) & = a_{00} + 2 \sum_{m=1}^{2K} a_{m0} \cos m \Omega_1 + 2 \sum_{n=1}^{2L} a_{0n} \cos n \Omega_2 \\
 & \quad + 4 \sum_{m=1}^{2K} \sum_{n=1}^{2L} a_{mn} \cos m \Omega_1 \cos n \Omega_2 .
 \end{aligned} \quad (242.34)$$

Wird auf die gleiche Weise der Nenner $N(\Omega_1, \Omega_2)$ in (242.31) berechnet, folgt

$$\begin{aligned}
 N(\Omega_1, \Omega_2) & = b_{00} + 2 \sum_{m=1}^{2K} b_{m0} \cos m \Omega_1 + 2 \sum_{n=1}^{2L} b_{0n} \cos n \Omega_2 \\
 & \quad + 4 \sum_{m=1}^{2K} \sum_{n=1}^{2L} b_{mn} \cos m \Omega_1 \cos n \Omega_2
 \end{aligned} \quad (242.35)$$

mit den Koeffizienten

$$\begin{aligned}
 b_{mn} & = \sum_{p_1=0}^K \dots \sum_{p_4=0}^K \sum_{q_1=0}^L \dots \sum_{q_4=0}^L g_{p_1 q_1} g_{p_2 q_2} g_{p_3 q_3} g_{p_4 q_4} .
 \end{aligned} \quad (242.36)$$

Mit den Definitionen

$$\bar{a}_{00} = a_{00}/b_{00}, \quad \bar{b}_{00} = 1$$

$$\bar{a}_{m0} = 2a_{m0}/b_{00}, \quad \bar{b}_{m0} = 2b_{m0}/b_{00} \quad \text{für } m \in \{1, \dots, 2K\}$$

$$\bar{a}_{0n} = 2a_{0n}/b_{00}, \quad \bar{b}_{0n} = 2b_{0n}/b_{00} \quad \text{für } n \in \{1, \dots, 2L\}$$

$$\bar{a}_{mn} = 4a_{mn}/b_{00}, \quad \bar{b}_{mn} = 4b_{mn}/b_{00} \quad \text{für } m \in \{1, \dots, 2K\} \text{ und } n \in \{1, \dots, 2L\} \quad (242.37)$$

läßt sich schließlich die nullphasige Frequenzantwort $H(j\Omega_1, j\Omega_2)$ aus (242.31) in der kompakten Form

$$H(j\Omega_1, j\Omega_2) = \left(\sum_{p=0}^{2K} \sum_{q=0}^{2L} \bar{a}_{pq} \cos p\Omega_1 \cos q\Omega_2 \right) / \left(\sum_{k=0}^{2K} \sum_{l=0}^{2L} \bar{b}_{kl} \cos k\Omega_1 \cos l\Omega_2 \right) \quad (242.38)$$

schreiben. Sie ist wegen

$$A(\Omega_1, \Omega_2) = A(-\Omega_1, \Omega_2) = A(-\Omega_1, -\Omega_2) = A(\Omega_1, -\Omega_2) \quad (242.39)$$

nach (241.50) und (241.51) vierquadrantensymmetrisch.

Für die Filterung selbst geht man auf die Kaskadenstruktur (241.19) zurück. Bezeichnet man das Eingangssignal des ersten Filters mit der Frequenzantwort $H_1(j\Omega_1, j\Omega_2)$ aus (242.7) nach (242.12) mit $x(m, n) = y_0(m, n)$, ergibt sich das Ausgangssignal mit $y(m, n) = y_1(m, n)$ gemäß (242.10) zu

$$y_1(m, n) = \sum_{p=0}^K \sum_{q=0}^L d_{pq} y_0(m-p, n-q) - \sum_{\substack{k=0 \\ (k, l) \neq (0, 0)}}^K \sum_{l=0}^L g_{kl} y_1(m-k, n-l) .$$

Dieses Ausgangssignal dient als Eingangssignal des zweiten Filters mit der Frequenzantwort $H_2(j\Omega_1, j\Omega_2)$ aus (242.25). Man erhält dann weiter an den Ausgängen der verbleibenden Filter gemäß (241.7), (241.11) und (241.15)

$$y_2(m, n) = \sum_{p=0}^K \sum_{q=0}^L d_{pq} y_1(m+p, n-q) - \sum_{\substack{k=0 \\ (k, l) \neq (0, 0)}}^K \sum_{l=0}^L g_{kl} y_2(m+k, n-l)$$

$$y_3(m, n) = \sum_{p=0}^K \sum_{q=0}^L d_{pq} y_2(m+p, n+q) - \sum_{\substack{k=0 \\ (k, l) \neq (0, 0)}}^K \sum_{l=0}^L g_{kl} y_3(m+k, n+l)$$

$$y_4(m, n) = \sum_{p=0}^K \sum_{q=0}^L d_{pq} y_3(m-p, n+q) - \sum_{\substack{k=0 \\ (k, l) \neq (0, 0)}}^K \sum_{l=0}^L g_{kl} y_4(m-k, n+l) , \quad (242.40)$$

wobei $y(m, n) = y_4(m, n)$ das Ausgangssignal des nullphasigen Gesamtfilters bedeutet, zu dessen Berechnung also vier Durchläufe mit unterschiedlichen Rekursionsrichtungen

notwendig sind.

Es wird nun eine Frequenzantwort $H(j\Omega_1, j\Omega_2)$ betrachtet, die nach (241.28) separierbar ist

$$H(j\Omega_1, j\Omega_2) = H_1(j\Omega_1)H_2(j\Omega_2) \quad (242.41)$$

Setzt man hierin beispielsweise die nullphasige Frequenzantwort $H(j\Omega)$ aus (223.22) einmal für $\Omega = \Omega_1$ und zum anderen für $\Omega = \Omega_2$ ein, erhält man

$$H(j\Omega_1, j\Omega_2) = \frac{\left(\sum_{p=0}^N \bar{a}_p \cos p\Omega_1 \right) \left(\sum_{q=0}^N \bar{a}_q \cos q\Omega_2 \right)}{\left(\sum_{k=0}^N \bar{b}_k \cos k\Omega_1 \right) \left(\sum_{l=0}^N \bar{b}_l \cos l\Omega_2 \right)} = \frac{\sum_{p=0}^N \sum_{q=0}^N \bar{a}_{pq} \cos p\Omega_1 \cos q\Omega_2}{\sum_{k=0}^N \sum_{l=0}^N \bar{b}_{kl} \cos k\Omega_1 \cos l\Omega_2} \quad (242.42)$$

mit den Koeffizienten

$$\bar{a}_{pq} = \bar{a}_p \bar{a}_q \quad \text{und} \quad \bar{b}_{kl} = \bar{b}_k \bar{b}_l \quad (242.43)$$

Dieses Ergebnis ist formal mit (242.38) identisch, so daß also eine vierquadrantensymmetrische Frequenzantwort vorliegt, für die (242.39) gilt. Wegen (241.53) ist sie zudem oktagonale. Die Filterung läuft analog zu (242.40) mit (223.13) und (223.15) in vier Schritten ab

$$y_1(m, n) = \sum_{r=0}^N d_r y_0(m-r, n) - \sum_{k=1}^N g_k y_1(m-k, n)$$

$$y_2(m, n) = \sum_{r=0}^N d_r y_1(m+r, n) - \sum_{k=1}^N g_k y_2(m+k, n)$$

für alle n und

$$y_3(m, n) = \sum_{r=0}^N d_r y_2(m, n-r) - \sum_{k=1}^N g_k y_3(m, n-k)$$

$$y_4(m, n) = \sum_{r=0}^N d_r y_3(m, n+r) - \sum_{k=1}^N g_k y_4(m, n+k) \quad (242.44)$$

für alle m . Im Gegensatz zu (242.40) verlaufen in (242.44) die Rekursionsrichtungen parallel zu den Achsen des m, n -Koordinatensystems.

243 Entwurf zweidimensionaler rekursiver Filter

Es wird nun die am Ende des Kapitels 241 formulierte Entwurfsaufgabe auf rekursive Filter angewendet. Die Aufgabe besteht in der Bestimmung der Filterkoeffizienten d_{pq} und g_{k1} , die nach (241.58) in dem Vektor β zusammengefaßt sind. Eine einfache Lösung erhält man, falls die Frequenzantwort des zu entwickelnden Filters als separierbar vorausgesetzt wird. In diesem Fall wird die zweidimensionale Problemstellung auf zwei eindimensionale zurückgeführt. Demonstriert wird diese Vorgehensweise im folgenden am Beispiel des *zweidimensionalen Butterworthfilters*. Anschließend wird das allgemein anwendbare Verfahren der Parameterschätzung im linearen Modell vorgestellt, wobei als Realfunktion die vierquadrantensymmetrische Frequenzantwort aus (242.38) gewählt wird. Weitere Verfahren werden beispielsweise bei (Wahl 1980, S.20) und (Dudgeon und Mersereau 1984, S.218) behandelt.

Ausgangspunkt der folgenden Betrachtungen ist die separierbare Frequenzantwort $H(j\Omega_1, j\Omega_2)$ aus (242.41). Definiert man

$$H_1(j\Omega_1) = |H_1(j\Omega_1)|^2 \quad (243.1)$$

und

$$H_2(j\Omega_2) = |H_1(j\Omega_2)|^2 \quad (243.2)$$

mit

$$|H_1(j\Omega_i)|^2 = \left(1 + \frac{\tan^{2q}(\Omega_i/2)}{\tan^{2q}(\Omega_c/2)}\right)^{-1} \quad \text{für } i \in \{1, 2\} \quad (243.3)$$

aus (224.4), so folgt aus (242.41) die reelle Frequenzantwort

$$H(j\Omega_1, j\Omega_2) = |H_1(j\Omega_1)|^2 |H_1(j\Omega_2)|^2. \quad (243.4)$$

Mit ihr kann nun die folgende Entwurfsaufgabe formuliert werden, die einen rechteckigen Durchlaßbereich vorsieht.

Gegeben sei die Idealfunktion $f(x_1, x_2) = H(j\Omega_1, j\Omega_2) = A(\Omega_1, \Omega_2)$ des idealen zweidimensionalen Tiefpaßfilters, für die gelten soll

$$\begin{aligned} \text{Durchlaßbereich:} \quad & A(\Omega_1, \Omega_2) = 1 \\ & \text{für } 0 \leq |\Omega_1| \leq \Omega_m \quad \text{und} \quad 0 \leq |\Omega_2| \leq \Omega_m \end{aligned}$$

$$\begin{aligned} \text{Sperrbereich:} \quad & A(\Omega_1, \Omega_2) = 0 \\ & \text{für } \Omega_m \leq |\Omega_1| \leq \pi \quad \text{und/oder} \quad \Omega_m < |\Omega_2| \leq \pi \end{aligned}$$

sowie die im Zusammenhang mit (224.5) eingeführte Modellparametermenge M.

Gesucht ist die Realfunktion $f_0(x_1, x_2) = H_0(j\Omega_1, j\Omega_2) = A_0(\Omega_1, \Omega_2) = |H_1(j\Omega_1)|^2 |H_1(j\Omega_2)|^2$ aus (243.4), für die gelten soll

$$\begin{aligned} \text{Durchlaßbereich:} \quad & d_{D1} \leq A_0(\Omega_1, \Omega_2) \leq d_{D2} \\ & \text{für } 0 \leq |\Omega_1| \leq \Omega_D \quad \text{und} \quad 0 \leq |\Omega_2| \leq \Omega_D \\ \text{Sperrbereich:} \quad & d_{S1} \leq A_0(\Omega_1, \Omega_2) \leq d_{S2} \\ & \text{für } \Omega_S < |\Omega_1| \leq \pi \quad \text{und/oder} \quad \Omega_S < |\Omega_2| \leq \pi \\ \text{Übergangsbereich:} \quad & d_{S2} \leq A_0(\Omega_1, \Omega_2) \leq d_{D1} \\ & \text{für } \Omega_D < |\Omega_1| \leq \Omega_S \quad \text{und/oder} \quad \Omega_D < |\Omega_2| \leq \Omega_S . \end{aligned} \quad (243.5)$$

Mit (224.27) lautet die Realfunktion

$$A_0(\Omega_1, \Omega_2) = A_0(\Omega_1)A_0(\Omega_2) = \prod_{k=1}^{q/2} H_{0k}(j\Omega_1, j\Omega_2) \quad (243.6)$$

mit den Frequenzantworten

$$H_{0k}(j\Omega_1, j\Omega_2) = H_{0k}(j\Omega_1)H_{0k}(j\Omega_2) = A_{0k}(\Omega_1)A_{0k}(\Omega_2) = A_{0k}(\Omega_1, \Omega_2) . \quad (243.7)$$

Wird hierin der Ausdruck (224.33) einmal mit $\Omega = \Omega_1$ und ein weiteres Mal mit $\Omega = \Omega_2$ eingesetzt, folgt nach (242.42)

$$H_{0k}(j\Omega_1, j\Omega_2) = A_{0k}(\Omega_1, \Omega_2) = \frac{\sum_{p=0}^2 \sum_{q=0}^2 \bar{a}_{pqk} \cos p\Omega_1 \cos q\Omega_2}{\sum_{m=0}^2 \sum_{n=0}^2 \bar{b}_{mnk} \cos m\Omega_1 \cos n\Omega_2} \quad (243.8)$$

mit den Koeffizienten aus (242.43)

$$\bar{a}_{pqk} = \bar{a}_{pk}\bar{a}_{qk} \quad \text{und} \quad \bar{b}_{mnk} = \bar{b}_{mk}\bar{b}_{nk} . \quad (243.9)$$

Nach (242.42) ist dies für $N = 2$ die oktagonale Frequenzantwort eines zweidimensionalen rekursiven Filters, dessen Filtergleichungen durch (242.44) gegeben sind. Die Realfunktion des zweidimensionalen Butterworthfilters und seine Filtergleichungen erhält man mit (243.6) durch die Hintereinanderschaltung der $q/2$ Teilfilter.

Beispiel 1: Gegeben sei die in Abbildung 243-1 dargestellte Idealfunktion $f(x_1, x_2) = H(j\Omega_1, j\Omega_2) = A(\Omega_1, \Omega_2)$ mit

$$\begin{aligned} \text{Durchlaßbereich:} \quad & A(\Omega_1, \Omega_2) = 1 \\ & \text{für } 0 \leq |\Omega_1| \leq \Omega_m \quad \text{und} \quad 0 \leq |\Omega_2| \leq \Omega_m \end{aligned}$$

Sperrbereich: $A(\Omega_1, \Omega_2) = 0$
 für $\Omega_m \leq |\Omega_1| \leq \pi$ und/oder $\Omega_m < |\Omega_2| \leq \pi$

und die Modellparametermenge mit den Spezifikationen

$$d_{S1} = 0 ; d_{S2} = 0,1975 ; d_{D1} = 0,8 ; d_{D2} = 1 ; \Omega_D = 0,2\pi ; \Omega_S = 0,4\pi .$$

Gesucht ist die Realfunktion $f_0(x_1, x_2) = H_0(j\Omega_1, j\Omega_2) = A_0(\Omega_1, \Omega_2)$ gemäß (243.4).

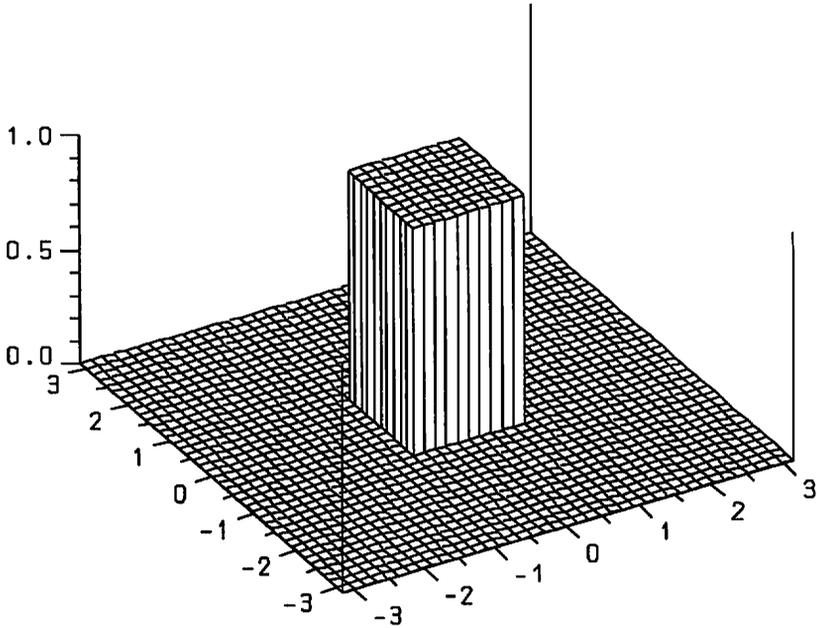


Abb. 243-1: Idealfunktion $A(\Omega_1, \Omega_2)$

Mit den Ergebnissen des Beispiels 1 aus Kapitel 224 erhält man die Koeffizienten (243.9), die in den 3×3 Matrizen \mathbf{A}_1 und \mathbf{B}_1 zusammengefaßt werden. So lautet \mathbf{A}_1

$$\mathbf{A}_1 = \begin{bmatrix} \bar{a}_{001} & \bar{a}_{011} & \bar{a}_{021} \\ \bar{a}_{101} & \bar{a}_{111} & \bar{a}_{121} \\ \bar{a}_{201} & \bar{a}_{211} & \bar{a}_{221} \end{bmatrix} = \begin{bmatrix} 1,824 & 2,432 & 0,608 \\ 2,432 & 3,243 & 0,811 \\ 0,608 & 0,811 & 0,203 \end{bmatrix} \cdot 10^{-3} .$$

Die Matrix \mathbf{B}_1 ergibt sich zu

$$\mathbf{B}_1 = \begin{bmatrix} \bar{b}_{001} & \bar{b}_{011} & \bar{b}_{021} \\ \bar{b}_{101} & \bar{b}_{111} & \bar{b}_{121} \\ \bar{b}_{201} & \bar{b}_{211} & \bar{b}_{221} \end{bmatrix} = \begin{bmatrix} 1 & -1,220 & 0,333 \\ -1,220 & 1,487 & -0,406 \\ 0,333 & -0,406 & 0,111 \end{bmatrix} .$$

Die gesuchte Realfunktion berechnet sich aus (243.8) zu

$$A_0(\Omega_1, \Omega_2) = \frac{\sum_{p=0}^2 \sum_{q=0}^2 \bar{a}_{pq} \cos p \Omega_1 \cos q \Omega_2}{\sum_{k=0}^2 \sum_{l=0}^2 \bar{b}_{kl} \cos k \Omega_1 \cos l \Omega_2}$$

und ist in Abbildung 243-2 veranschaulicht. Weiter ergeben sich die Filtergleichungen nach (242.44) mit $N = 2$ und den Filterkoeffizienten d_{i1} und g_{i1} für $i \in \{0, 1, 2\}$ aus (224.35). Da $q = 2$ in (243.6) gilt, ist jeweils nur eine Vorwärts- und Rückwärtsfilterung entlang der m - und n -Achse erforderlich. Δ

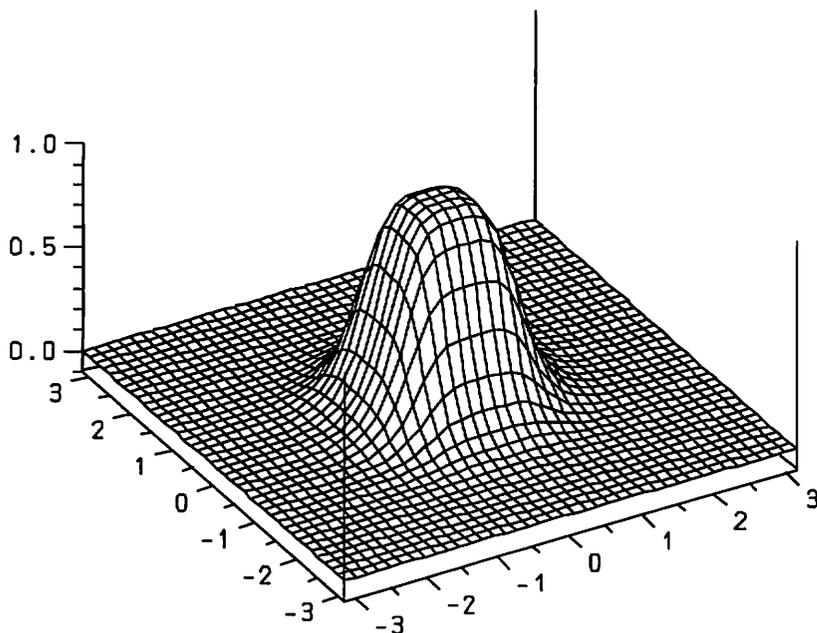


Abb. 243-2: Realfunktion $A_0(\Omega_1, \Omega_2)$

Allgemein lassen sich die unbekannt Filterkoeffizienten, die nach (241.58) im Vektor β zusammengefaßt sind, aus einer Parameterschätzung ermitteln (Schmidt 1993a). Wählt man beispielsweise als Idealfunktion (241.55) das Amplitudenspektrum $A(\Omega_1, \Omega_2)$, erhält man aus (241.60) mit $x_1 = \Omega_1$ und $x_2 = \Omega_2$ die Approximationsfehlerfunktion

$$e(\Omega_1, \Omega_2) = A_0(\Omega_1, \Omega_2, \beta) - A(\Omega_1, \Omega_2) . \quad (243.10)$$

Als Realfunktion $A_0(\Omega_1, \Omega_2, \beta)$ soll die Frequenzantwort (242.38) des vierquadranten-symmetrischen nullphasigen rekursiven Filters dienen. Man faßt also die Koeffizienten \bar{a}_{pq} und \bar{b}_{kl} aus (242.38) mit $2(2K+1)(2L+1) - 1 = u$ im $u \times 1$ Vektor

$$\beta = [\bar{a}_{00}, \bar{a}_{10}, \dots, \bar{a}_{2K,0}, \bar{a}_{01}, \dots, \bar{a}_{2K,1}, \dots, \bar{a}_{0,2L}, \dots, \bar{a}_{2K,2L}, \bar{b}_{10}, \dots, \bar{b}_{2K,0}, \bar{b}_{01}, \dots, \bar{b}_{2K,1}, \dots, \bar{b}_{0,2L}, \dots, \bar{b}_{2K,2L}]' \quad (243.11)$$

zusammen und spaltet ihn analog zu (224.39) mit

$$\beta = \beta_0 + \delta\beta \quad (243.12)$$

in den Vektor der Näherungswerte β_0 und den Vektor der unbekanntenen Zuschläge $\delta\beta$ auf. Nach einer Linearisierung erhält man für (243.10)

$$e(\Omega_1, \Omega_2) = A_0(\Omega_1, \Omega_2, \beta_0) - A(\Omega_1, \Omega_2) + x(\Omega_1, \Omega_2)' \delta\beta \quad (243.13)$$

mit

$$x(\Omega_1, \Omega_2) = (\partial A_0 / \partial \beta) |_{\beta = \beta_0} .$$

Für den Aufbau des linearen Modells werden die Frequenzen Ω_1 und Ω_2 diskretisiert. Bildet man also (243.13) an den Stellen $\Omega_1 = \Omega_{1i}$ mit $i \in \{1, \dots, n_1\}$ und $\Omega_2 = \Omega_{2j}$ mit $j \in \{1, \dots, n_2\}$ erhält man mit $n = n_1 n_2$ die n Gleichungen

$$y_{ij} + e_{ij} = x_{ij}' \delta\beta , \quad (243.14)$$

worin $e_{ij} = e(\Omega_{1i}, \Omega_{2j})$, $y_{ij} = A(\Omega_{1i}, \Omega_{2j}) - A_0(\Omega_{1i}, \Omega_{2j}, \beta_0)$ und $x_{ij} = x(\Omega_{1i}, \Omega_{2j})$ gilt. Mit den $n \times 1$ Vektoren

$$e = \text{vec}E \quad \text{mit} \quad E = (e_{ij}) ,$$

$$y = \text{vec}Y \quad \text{mit} \quad Y = (y_{ij}) \quad (243.15)$$

und der $n \times n$ Matrix

$$X = [x_{11}, x_{21}, \dots, x_{n_1,1}, x_{12}, \dots, x_{n_1,2}, \dots, x_{1,n_2}, \dots, x_{n_1,n_2}]' \quad (243.16)$$

erhält man das lineare Modell

$$y + e = X \delta\beta \quad (243.17)$$

analog zu (224.42). Die Schätzwerte ergeben sich wieder aus (224.43) und (224.44) und liefern die Realfunktion

$$\hat{A}_0(\Omega_1, \Omega_2) = A_0(\Omega_1, \Omega_2, \hat{\beta}) \quad (243.18)$$

mit

$$\hat{\beta} = \beta_0 + \delta\hat{\beta} = \beta_0 + (X'PX)^{-1}X'Py , \quad (243.19)$$

die anschließend einer Stabilitätsprüfung zu unterziehen ist. Stabilitätsbedingungen lassen sich auch mit in die Parameterschätzung einbeziehen, worauf bei (Dudgeon und Mersereau 1984, S.255) hingewiesen wird.

244 Definition und Eigenschaften zweidimensionaler nichtrekursiver Filter

Analog zu (225.1) bezeichnet man ein Filter als zweidimensionales streng *kausales nichtrekursives lineares digitales Filter vom Grad $\max\{P, Q\}$* , falls seine Filtergleichung

$$y(m, n) = \sum_{p=0}^P \sum_{q=0}^Q d_{pq} x(m-p, n-q) \quad (244.1)$$

mit

$$\begin{aligned} d_{p0} &\neq 0 \quad \text{für mindestens ein } q \\ d_{pQ} &\neq 0 \quad \text{für mindestens ein } p \end{aligned} \quad (244.2)$$

lautet, die aus (242.1) für $g_{k1} = 0$ folgt. Substituiert man in (244.1) $P = K-1$ und $Q = L-1$, ergibt sich die in der Literatur üblicherweise verwendete Filtergleichung des zweidimensionalen nichtrekursiven Filters vom Grad $\max\{K-1, L-1\}$

$$y(m, n) = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} d_{kl} x(m-k, n-l) \quad (244.3)$$

Die Impulsantwort $h(k, l)$ erhält man aus (235.30) zu

$$h(k, l) = d_{kl} \quad (244.4)$$

und die Systemfunktion $H(z_1, z_2)$ aus (235.31) mit (244.4) zu

$$H(z_1, z_2) = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} d_{kl} z_1^{-k} z_2^{-l} \quad (244.5)$$

Es werden nun die zweidimensionalen nichtrekursiven Filter mit linearer Phase

$$\Phi(\Omega_1, \Omega_2) = \alpha_1 \Omega_1 + \alpha_2 \Omega_2 + \alpha_3 \quad (244.6)$$

nach (241.32) behandelt. Ausgangspunkt der Betrachtungen ist der Fall (241.34)

a) α_1 beliebig, α_2 beliebig und $\alpha_3 = 0$.

Für die Impulsantwort gilt nach (241.37) die Symmetriebedingung

$$h(k-\alpha_1, l-\alpha_2) = h(-k-\alpha_1, -l-\alpha_2) \quad (244.7)$$

so daß aus (244.4) für α_1 und α_2

$$\alpha_1 = -(K-1)/2 \quad \text{und} \quad \alpha_2 = -(L-1)/2 \quad (244.8)$$

als Verallgemeinerung von (225.10) folgt. Substituiert man in (244.7) $m = k-\alpha_1$, $k =$

$m+\alpha_1$, $n = 1-\alpha_2$ und $l = n+\alpha_2$, erhält man mit (244.8)

$$h(m,n) = h(K-1-m, L-1-n) \quad \text{für } m \in \{0, \dots, K-1\} \text{ und } n \in \{0, \dots, L-1\} \quad . \quad (244.9)$$

Die Filtergleichung ergibt sich aus (244.3) mit (244.4) unter Berücksichtigung der Symmetriebedingung (244.9).

Für die Herleitung der Frequenzantwort ist wie bei den eindimensionalen Filtern im Kapitel 225 zu unterscheiden, ob K und L gerade oder ungerade Zahlen sind, wobei insgesamt vier Kombinationen auftreten können. Eindimensionale Filter mit geradzahligem N besitzen gegenüber Filtern mit ungeradzahligem N den Nachteil, daß das zu verarbeitende Eingangssignal zwischen den diskreten Werten benötigt wird. Da es sich bei zweidimensionalen Systemen genauso verhält, wird im folgenden stets von ungeradzahligem Werten für K und L ausgegangen. Der Fall der geradzahligem Werte für K und L wird beispielsweise bei (Cappellini et al. 1978, S.63) behandelt.

Schließlich sind noch die beiden verbleibenden Fälle, also K geradzahlig und L ungeradzahlig sowie K ungeradzahlig und L geradzahlig, zu nennen, die jedoch ohne große Bedeutung sind. Oftmals ist man nämlich, wie schon erwähnt, daran interessiert, Systeme mit genähert rotationssymmetrischen Frequenzantworten einzusetzen, um richtungsunabhängige Filter zu gewinnen. In (244.3) setzt man daher häufig $K = L$ und wählt, wie oben erläutert, $K = L$ ungeradzahlig.

Sind also K und L ungerade Zahlen, erhält man für die Frequenzantwort $H(j\Omega_1, j\Omega_2)$ nach (235.31) und (241.29)

$$H(j\Omega_1, j\Omega_2) = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} h(k, l) e^{-j(k\Omega_1 + l\Omega_2)} = A(\Omega_1, \Omega_2) e^{j(\alpha_1\Omega_1 + \alpha_2\Omega_2)} \quad , \quad (244.10)$$

so daß für die Amplitude $A(\Omega_1, \Omega_2)$ mit (244.8) folgt

$$\begin{aligned} A(\Omega_1, \Omega_2) &= \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} h(k, l) e^{-j((k+\alpha_1)\Omega_1 + (l+\alpha_2)\Omega_2)} \\ &= \sum_{k=0}^{K-1} e^{-j(k+\alpha_1)\Omega_1} \left[\sum_{l=0}^{(L-3)/2} h(k, l) e^{-j(1+\alpha_2)\Omega_2} \right. \\ &\quad \left. + h(k, -\alpha_2) + \sum_{l=(L+1)/2}^{L-1} h(k, l) e^{-j(1+\alpha_2)\Omega_2} \right] \\ &= \sum_{k=0}^{(K-3)/2} e^{-j(k+\alpha_1)\Omega_1} \left[\sum_{l=0}^{(L-3)/2} h(k, l) e^{-j(1+\alpha_2)\Omega_2} + h(k, -\alpha_2) \right] \end{aligned}$$

$$\begin{aligned}
& + \sum_{l=(L+1)/2}^{L-1} h(k, l) e^{-j(1+\alpha_2)\Omega_2} + \left[\sum_{l=0}^{(L-3)/2} h(-\alpha_1, l) e^{-j(1+\alpha_2)\Omega_2} \right. \\
& + h(-\alpha_1, -\alpha_2) + \left. \sum_{l=(L+1)/2}^{L-1} h(-\alpha_1, l) e^{-j(1+\alpha_2)\Omega_2} \right] \\
& + \sum_{k=(K+1)/2}^{K-1} e^{-j(k+\alpha_1)\Omega_1} \left[\sum_{l=0}^{(L-3)/2} h(k, l) e^{-j(1+\alpha_2)\Omega_2} + h(k, -\alpha_2) \right. \\
& + \left. \sum_{l=(L+1)/2}^{L-1} h(k, l) e^{-j(1+\alpha_2)\Omega_2} \right] . \tag{244.11}
\end{aligned}$$

Substituiert man in der Summe vor der letzten eckigen Klammer $k = K-1-m = -2\alpha_1-m$ und $m = K-1-k = -2\alpha_1-k$ sowie in der ersten und der letzten Summe der letzten eckigen Klammer und der letzten Summe in der mittleren eckigen Klammer $l = L-1-n = -2\alpha_2-n$ und $n = L-1-l = -2\alpha_2-1$, erhält man mit (244.9), (212.11) sowie (244.4), wobei die Summen in der mittleren eckigen Klammer die erste Summe der folgenden Gleichung ergeben

$$\begin{aligned}
A(\Omega_1, \Omega_2) & = h(-\alpha_1, -\alpha_2) + 2 \sum_{l=0}^{(L-3)/2} h(-\alpha_1, l) \cos(1+\alpha_2)\Omega_2 \\
& + 2 \sum_{k=0}^{(K-3)/2} \sum_{l=0}^{L-1} h(k, l) \cos((k+\alpha_1)\Omega_1 + (1+\alpha_2)\Omega_2) \\
& = 2 \sum_{l=0}^{(L-3)/2} d_{-\alpha_1, l} \cos(1+\alpha_2)\Omega_2 + d_{-\alpha_1, -\alpha_2} \\
& + 2 \sum_{k=0}^{(K-3)/2} \sum_{l=0}^{L-1} d_{k1} \cos((k+\alpha_1)\Omega_1 + (1+\alpha_2)\Omega_2) . \tag{244.12}
\end{aligned}$$

Mit den Substitutionen

$$\begin{aligned}
\bar{d}_{00} & = d_{-\alpha_1, -\alpha_2} \\
\bar{d}_{0n} & = 2d_{-\alpha_1, -\alpha_2-n} \quad \text{für } n \in \{1, \dots, (L-1)/2\} \\
\bar{d}_{mn} & = 2d_{-\alpha_1-m, -\alpha_2-n} \quad \text{für } m \in \{1, \dots, (K-1)/2\} \\
& \quad n \in \{-(L-1)/2, \dots, (L-1)/2\} \tag{244.13}
\end{aligned}$$

folgt aus (244.12) schließlich

$$A(\Omega_1, \Omega_2) = \sum_{n=0}^{(L-1)/2} \bar{d}_n \cos n\Omega_2 + \sum_{m=1}^{(K-1)/2} \sum_{n=-(L-1)/2}^{n-(L-1)/2} \bar{d}_m \cos(m\Omega_1 + n\Omega_2) \cdot (244.14)$$

Wegen (241.50) ist $A(\Omega_1, \Omega_2)$ die Amplitude eines zweiquadrantenensymmetrischen Systems. Von den ursprünglichen KL Elementen der Impulsantwort $h(k, l)$ werden in (244.12) nur $(KL+1)/2$ benötigt. Sie sind in dem unterlegten Bereich der Abbildung 244-1a) positioniert, in dem das k, l -Koordinatensystem wie in Abbildung 235-1 orientiert ist.

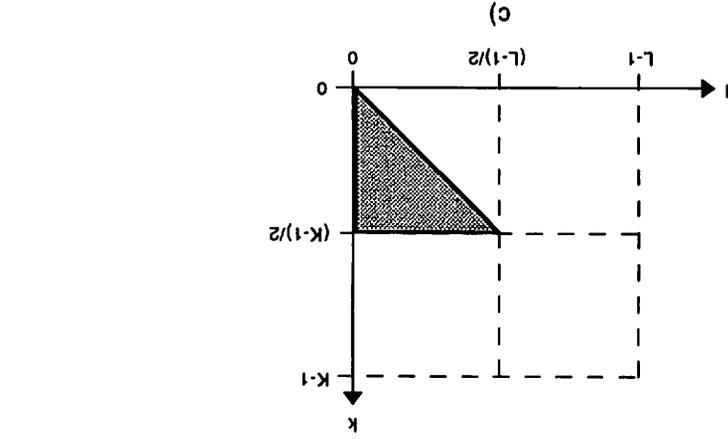
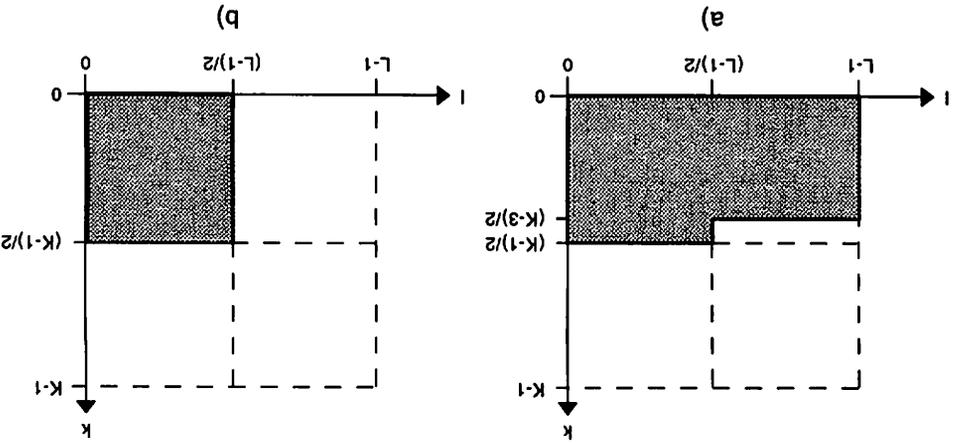


Abb. 244-1: a) zweiquadrantenensymmetrische Impulsantwort
 b) vierquadrantenensymmetrische Impulsantwort
 c) oktagonale Impulsantwort

Setzt man (244.14) in (244.10) ein, ergibt sich die Frequenzantwort des zweidimensionalen nichtrekursiven Filters mit linearer Phase und $\alpha_3 = 0$.

Führt man zusätzlich zu (244.7) die Bedingung

$$h(k-\alpha_1, l-\alpha_2) = h(k-\alpha_1, -l-\alpha_2) \quad (244.15)$$

ein, erhält man mit den Substitutionen $m = k-\alpha_1$, $n = l-\alpha_2$ und $l = n+\alpha_2$ und mit (244.8)

$$h(m, n) = h(m, L-1-n) . \quad (244.16)$$

Die Doppelsumme in der ersten Darstellung von (244.12) läßt sich aufspalten in

$$\begin{aligned} & 2 \sum_{k=0}^{(K-3)/2} \sum_{l=0}^{L-1} h(k, l) \cos((k+\alpha_1)\Omega_1 + (l+\alpha_2)\Omega_2) \\ & = 2 \sum_{k=0}^{(K-3)/2} \left[\sum_{l=0}^{(L-3)/2} h(k, l) \cos((k+\alpha_1)\Omega_1 + (l+\alpha_2)\Omega_2) \right. \\ & \quad \left. + h(k, -\alpha_2) \cos(k+\alpha_1)\Omega_1 + \sum_{l=(L+1)/2}^{L-1} h(k, l) \cos((k+\alpha_1)\Omega_1 + (l+\alpha_2)\Omega_2) \right] . \end{aligned} \quad (244.17)$$

Substituiert man in der letzten Summe $n = l - (L-1) = l + 2\alpha_2$ und $l = n + L - 1 = n - 2\alpha_2$, ergibt sich

$$\begin{aligned} & \sum_{l=(L+1)/2}^{L-1} h(k, l) \cos((k+\alpha_1)\Omega_1 + (l+\alpha_2)\Omega_2) \\ & = \sum_{n=-(L-3)/2}^0 h(k, L-1+n) \cos((k+\alpha_1)\Omega_1 + (n-\alpha_2)\Omega_2) . \end{aligned} \quad (244.18)$$

Substituiert man weiter $l = -n$ und berücksichtigt (244.16), erhält man für die Amplitude $A(\Omega_1, \Omega_2)$ mit (244.17) und (244.18) nach Anwendung der Additionstheoreme für $\cos(x+y)$ und für $\cos(x-y)$ aus (244.12)

$$\begin{aligned} A(\Omega_1, \Omega_2) & = h(-\alpha_1, -\alpha_2) + 2 \sum_{k=0}^{(K-3)/2} h(k, -\alpha_2) \cos(k+\alpha_1)\Omega_1 \\ & \quad + 2 \sum_{l=0}^{(L-3)/2} h(-\alpha_1, l) \cos(l+\alpha_2)\Omega_2 \\ & \quad + 4 \sum_{k=0}^{(K-3)/2} \sum_{l=0}^{(L-3)/2} h(k, l) \cos(k+\alpha_1)\Omega_1 \cos(l+\alpha_2)\Omega_2 \end{aligned}$$

$$\begin{aligned}
&= d_{-\alpha_1, -\alpha_2} + 2 \sum_{k=0}^{(K-3)/2} d_{k, -\alpha_2} \cos(k+\alpha_1)\Omega_1 \\
&\quad + 2 \sum_{l=0}^{(L-3)/2} d_{-\alpha_1, l} \cos(l+\alpha_2)\Omega_2 \\
&\quad + 4 \sum_{k=0}^{(K-3)/2} \sum_{l=0}^{(L-3)/2} d_{k,l} \cos(k+\alpha_1)\Omega_1 \cos(l+\alpha_2)\Omega_2 .
\end{aligned} \tag{244.19}$$

Definiert man weiter

$$\begin{aligned}
\bar{f}_{00} &= d_{-\alpha_1, -\alpha_2} \\
\bar{f}_{m0} &= 2d_{-\alpha_1-m, -\alpha_2} \\
\bar{f}_{0n} &= 2d_{-\alpha_1, -\alpha_2-n} \\
\bar{f}_{mn} &= 4d_{-\alpha_1-m, -\alpha_2-n}
\end{aligned} \tag{244.20}$$

für $m \in \{1, \dots, (K-1)/2\}$ und $n \in \{1, \dots, (L-1)/2\}$, läßt sich (244.19) in der kompakten Form

$$A(\Omega_1, \Omega_2) = \sum_{k=0}^{(K-1)/2} \sum_{l=0}^{(L-1)/2} \bar{f}_{kl} \cos k\Omega_1 \cos l\Omega_2 \tag{244.21}$$

schreiben. Wegen (241.50) und (241.51) ist $A(\Omega_1, \Omega_2)$ die Amplitude eines vierquadrantensymmetrischen Systems. Gegenüber (244.12) hat sich die Anzahl der in (244.19) enthaltenen Elemente der Impulsantwort auf $(K+1)(L+1)/4$ reduziert. Sie sind im Teil b) der Abbildung 244-1 im unterlegten Bereich positioniert.

Setzt man (244.21) in (244.10) ein, erhält man die Frequenzantwort eines zweidimensionalen nichtrekursiven Filters mit linearer Phase und Vierquadrantensymmetrie.

Gilt zusätzlich in (244.21) $K = L$ und

$$\bar{f}_{kl} = \bar{f}_{lk} , \tag{244.22}$$

so ist die Frequenzantwort nach (241.53) oktogonal. Für die Impulsantwort gilt dann

$$h(k, l) = h(l, k) , \tag{244.23}$$

und die Anzahl der unabhängigen Elemente reduziert sich weiter auf $(K+1)(L+3)/8$, wie in Abbildung 244-1c) dargestellt ist.

Betrachtet wird nun der Fall

$$\text{b) } \alpha_1 = \alpha_2 = \alpha_3 = 0 .$$

Wegen (241.39) ergibt sich ein nullphasiges Filter mit der reellen Frequenzantwort

$$H(j\Omega_1, j\Omega_2) = A(\Omega_1, \Omega_2) . \quad (244.24)$$

Wird zunächst ein kausales System vorausgesetzt, erhält man mit der Konstanten c und dem Einheitsimpuls $d(k, l)$ aus (235.11) wegen (241.41) analog zu (225.22) die Impulsantwort

$$h(k, l) = c d(k, l) , \quad (244.25)$$

die nur bewirkt, daß das Ausgangssignal gegenüber dem Eingangssignal um den Faktor c verändert wird. Ein sinnvolles zweidimensionales nichtrekursives Filter ist also nicht-kausal.

Mit (244.14) und (244.22) wurden bereits zwei reelle Amplituden abgeleitet, so daß zwei nullphasige Systeme erhalten werden, wenn man die Ergebnisse in (244.24) einsetzt. Da die nullphasige Filterung besonders wichtig ist, werden im folgenden die Filtergleichungen abgeleitet.

Im Fall der Zweiquadrantensymmetrie ist die Frequenzantwort durch Substitution von (244.14) in (244.24) gegeben. Für die Herleitung der Filtergleichung benötigt man zunächst die Impulsantwort. Nach (244.11) gilt für die Amplitude $A(\Omega_1, \Omega_2)$

$$A(\Omega_1, \Omega_2) = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} h(k, l) e^{-j((k+\alpha_1)\Omega_1 + (l+\alpha_2)\Omega_2)} . \quad (244.26)$$

Substituiert man $k = m - \alpha_1$, $m = k + \alpha_1$, $l = n - \alpha_2$ und $n = l + \alpha_2$, erhält man mit (244.8)

$$\begin{aligned} A(\Omega_1, \Omega_2) &= \sum_{m=-(K-1)/2}^{(K-1)/2} \sum_{n=-(L-1)/2}^{(L-1)/2} h(m - \alpha_1, n - \alpha_2) e^{-j(m\Omega_1 + n\Omega_2)} \\ &= \sum_{m=-(K-1)/2}^{(K-1)/2} \sum_{n=-(L-1)/2}^{(L-1)/2} h_1(m, n) e^{-j(m\Omega_1 + n\Omega_2)} , \end{aligned} \quad (244.27)$$

worin

$$h_1(k, l) = h(k + (K-1)/2, l + (L-1)/2) \quad (244.28)$$

für $k \in \{-(K-1)/2, \dots, (K-1)/2\}$ und $l \in \{-(L-1)/2, \dots, (L-1)/2\}$ die Impulsantwort des zweiquadrantensymmetrischen nullphasigen Filters bedeutet. Um die Impulsantwort $h_1(k, l)$ in Abbildung 244-1a) darzustellen, ist der mit gestrichelten Linien abgegrenzte Bereich der Impulsantwort $h(k, l)$ in k -Richtung um $-(K-1)/2 = \alpha_1$ und in l -Richtung um $-(L-1)/2 = \alpha_2$ zu verschieben.

Die Filtergleichung erhält man aus (235.10) zu

$$y(m, n) = \sum_{k=-(K-1)/2}^{(K-1)/2} \sum_{l=-(L-1)/2}^{(L-1)/2} h_1(k, l) x(m-k, n-l) . \quad (244.29)$$

Berücksichtigt man hierin (244.28), (244.4) und (244.13), ergibt sich schließlich

$$y(m, n) = \frac{1}{2} \sum_{l=0}^{(L-1)/2} \bar{d}_{0l} (x(m, n+1) + x(m, n-1)) \\ + \frac{1}{2} \sum_{k=1}^{(K-1)/2} \sum_{l=-(L-1)/2}^{(L-1)/2} \bar{d}_{kl} (x(m+k, n+1) + x(m-k, n-1)) . \quad (244.30)$$

Im Fall der Vierquadrantensymmetrie ist die Frequenzantwort des nullphasigen Filters mit (244.24) durch (244.21) gegeben. Die Impulsantwort liegt wieder mit (244.28) vor. Für sie gelten wegen (244.7) und (244.15) die Symmetriebedingungen

$$h_1(k, l) = h_1(-k, l) = h_1(-k, -l) = h_1(k, -l) . \quad (244.31)$$

Berücksichtigt man diese Zusammenhänge in (244.29), erhält man mit (244.28), (244.4) und (244.20) die Filtergleichung

$$y(m, n) = \frac{1}{4} \sum_{k=0}^{(K-1)/2} \sum_{l=0}^{(L-1)/2} \bar{f}_{kl} (x(m+k, n-1) + x(m-k, n+1) \\ + x(m+k, n+1) + x(m-k, n-1)) . \quad (244.32)$$

Im Fall eines oktagonalen nullphasigen Filters gilt für die Koeffizienten \bar{f}_{kl} außerdem noch (244.22), so daß sich die Anzahl der Multiplikationen in (244.32) weiter verringert, worauf schon im Zusammenhang mit (244.23) hingewiesen wurde.

Die Frequenzantwort (244.24) mit (244.21) kann auch direkt aus der Frequenzantwort (242.38) des nullphasigen vierquadrantensymmetrischen rekursiven Filters abgeleitet werden. Für nichtrekursive Filter verschwinden die Koeffizienten \bar{b}_{kl} für $(k, l) \neq (0, 0)$. Ersetzt man weiter die Obergrenzen $2K$ und $2L$ durch $(K-1)/2$ und $(L-1)/2$ und schreibt $\bar{a}_{kl} = \bar{f}_{kl}$, ergibt sich

$$H(j\Omega_1, j\Omega_2) = \sum_{k=0}^{(K-1)/2} \sum_{l=0}^{(L-1)/2} \bar{f}_{kl} \cos k\Omega_1 \cos l\Omega_2 . \quad (244.33)$$

Ebenso kann die Frequenzantwort (244.24) mit (244.14) aus (242.19) gebildet werden. Betrachtet wird nun der Fall

c) α_1 beliebig, α_2 beliebig und $\alpha_3 = \pi/2$

aus (241.42). Die Impulsantwort erfüllt wegen (241.44) die Bedingung

$$h(k-\alpha_1, l-\alpha_2) = -h(-k-\alpha_1, -l-\alpha_2) . \quad (244.34)$$

Hieraus folgt für $k = l = 0$

$$h(-\alpha_1, -\alpha_2) = -h(-\alpha_1, -\alpha_2) = 0 \quad (244.35)$$

Substituiert man in (244.34) $m = k - \alpha_1$, $k = m + \alpha_1$, $n = l - \alpha_2$ und $l = n + \alpha_2$ mit α_1 und α_2 aus (244.8), erhält man

$$h(m, n) = -h(K-1-m, L-1-n) \text{ für } m \in \{0, \dots, K-1\} \text{ und } n \in \{0, \dots, L-1\} \quad (244.36)$$

Mit dieser Bedingung ergibt sich die Filtergleichung aus (244.3) mit (244.4).

Die Frequenzantwort $H(j\Omega_1, j\Omega_2)$ erhält man nach (235.31) und (241.29) mit (212.8) zu

$$\begin{aligned} H(j\Omega_1, j\Omega_2) &= \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} h(k, l) e^{-j(k\Omega_1 + l\Omega_2)} \\ &= A(\Omega_1, \Omega_2) e^{j(\alpha_1\Omega_1 + \alpha_2\Omega_2 + \alpha_3)} = jA(\Omega_1, \Omega_2) e^{j(\alpha_1\Omega_1 + \alpha_2\Omega_2)} \quad (244.37) \end{aligned}$$

wobei für die Amplitude $A(\Omega_1, \Omega_2)$ gilt

$$A(\Omega_1, \Omega_2) = -j \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} h(k, l) e^{-j((k+\alpha_1)\Omega_1 + (l+\alpha_2)\Omega_2)} \quad (244.38)$$

Spaltet man die Doppelsumme wie in (244.11) auf und substituiert in der Weise, die von (244.11) auf (244.12) führt, ergibt sich mit (212.11), (244.4) und (244.35)

$$\begin{aligned} A(\Omega_1, \Omega_2) &= -2 \sum_{l=0}^{(L-3)/2} h(-\alpha_1, l) \sin((l+\alpha_2)\Omega_2) \\ &\quad - 2 \sum_{k=0}^{(K-3)/2} \sum_{l=0}^{L-1} h(k, l) \sin((k+\alpha_1)\Omega_1 + (l+\alpha_2)\Omega_2) \\ &= -2 \sum_{l=0}^{(L-3)/2} d_{-\alpha_1, l} \sin((l+\alpha_2)\Omega_2) \\ &\quad - 2 \sum_{k=0}^{(K-3)/2} \sum_{l=0}^{L-1} d_{k1} \sin((k+\alpha_1)\Omega_1 + (l+\alpha_2)\Omega_2) \quad (244.39) \end{aligned}$$

Im Gegensatz zu (244.12) ist dies die Amplitude eines zweiquadrantenantisymmetrischen Systems, denn es gilt

$$A(\Omega_1, \Omega_2) = -A(-\Omega_1, -\Omega_2) \quad (244.40)$$

Setzt man (244.39) in (244.37) ein, erhält man die Frequenzantwort des nichtrekursiven Filters mit linearer Phase und $\alpha_3 = \pi/2$.

Führt man in (244.39) die Substitution

$$\begin{aligned} \bar{d}_{mn} &= 2d_{-\alpha_1-m, -\alpha_2-n} \quad \text{für } m \in \{0, \dots, (K-1)/2\} \\ &\quad \text{und } n \in \{-(L-1)/2, \dots, (L-1)/2\} \end{aligned} \quad (244.41)$$

analog zu (244.13) ein, lautet die Amplitude

$$A(\Omega_1, \Omega_2) = \sum_{n=1}^{(L-1)/2} \bar{d}_{0n} \sin n\Omega_2 + \sum_{m=0}^{(K-1)/2} \sum_{n=-(L-1)/2}^{(L-1)/2} \bar{d}_{mn} \sin(m\Omega_1+n\Omega_2) . \quad (244.42)$$

Stellt man außer der Bedingung (244.34) noch die Forderung

$$h(k-\alpha_1, l-\alpha_2) = -h(k-\alpha_1, -l-\alpha_2) \quad (244.43)$$

auf, läßt sich aus (244.37) die Frequenzantwort eines vierquadrantenantisymmetrischen Systems herleiten. Hierauf wird im folgenden jedoch nicht eingegangen, sondern auf (Cappellini et al. 1978, S.62) verwiesen.

Abschließend wird noch der Fall (241.45)

$$d) \quad \alpha_1 = \alpha_2 = 0 \quad \text{und} \quad \alpha_3 = \pi/2$$

behandelt, für den nach (241.47) die Frequenzantwort

$$H(j\Omega_1, j\Omega_2) = jA(\Omega_1, \Omega_2) \quad (244.44)$$

gültig ist. Setzt man also (244.42) in (244.44) ein, liegt ein zweiquadrantenantisymmetrisches nichtrekursives Filter mit der Phase $\Phi(\Omega_1, \Omega_2) = \pi/2$ aus (241.46) vor. Die Filtergleichung läßt sich analog zu (244.26) bis (244.30) ableiten. Ohne hier die einzelnen Berechnungsschritte anzugeben, erhält man

$$\begin{aligned} y(m, n) &= \frac{1}{2} \sum_{l=1}^{(L-1)/2} \bar{d}_{0l} (x(m, n+l) - x(m, n-l)) \\ &\quad + \frac{1}{2} \sum_{k=0}^{(K-1)/2} \sum_{l=-(L-1)/2}^{(L-1)/2} \bar{d}_{kl} (x(m+k, n+l) - x(m-k, n-l)) . \end{aligned} \quad (244.45)$$

245 Zweidimensionale nichtrekursive Elementarfilter

Wie bereits zu Beginn des Kapitels 226 erwähnt, können zweidimensionale nichtrekursive Elementarfilter aus eindimensionalen Elementarfiltern konstruiert werden. Vielfach wird von separierbaren Systemfunktionen nach (241.28) ausgegangen, so daß sich mit $z_i = e^{j\Omega_i}$ für $i \in \{1, 2\}$ ergibt

$$H(j\Omega_1, j\Omega_2) = H_1(j\Omega_1)H_2(j\Omega_2) . \quad (245.1)$$

Während dieser Ansatz auf der Kaskadenstruktur (241.19) beruht, kann auch die Parallelstruktur (241.21) zur Erzeugung eines zweidimensionalen Elementarfilters nach

$$H(j\Omega_1, j\Omega_2) = H_1(j\Omega_1) + H_2(j\Omega_2) \quad (245.2)$$

genutzt werden. Außerdem werden Kombinationen von (245.1) und (245.2) verwendet, beispielsweise

$$H(j\Omega_1, j\Omega_2) = (H_1(j\Omega_1) + H_2(j\Omega_2))H_3(j\Omega_1) \quad (245.3)$$

Im folgenden werden einige zweidimensionale Elementarfilter vorgestellt. Für weitere Beispiele wird auf (Jähne 1989, S.92; Kraus 1990, S.507) verwiesen.

Die Frequenzantwort eines *zweidimensionalen Rechteckfilters* mit Nullphase erhält man aus der Hintereinanderschaltung zweier eindimensionaler Rechteckfilter, deren Frequenzantworten mit (226.3) vorliegen. Setzt man also für $H_i(j\Omega_i)$ in (245.1) die Frequenzantworten

$$H_1(j\Omega_1) = R^{K-1}(\Omega_1) = \frac{1}{K} + \frac{2}{K} \sum_{k=1}^{(K-1)/2} \cos k\Omega_1 \quad (245.4)$$

$$H_2(j\Omega_2) = R^{L-1}(\Omega_2) = \frac{1}{L} + \frac{2}{L} \sum_{l=1}^{(L-1)/2} \cos l\Omega_2 \quad (245.5)$$

ein, erhält man

$$H(j\Omega_1, j\Omega_2) = R^{K-1, L-1}(\Omega_1, \Omega_2) = \sum_{k=0}^{(K-1)/2} \sum_{l=0}^{(L-1)/2} \bar{T}_{kl} \cos k\Omega_1 \cos l\Omega_2 \quad (245.6)$$

mit den Koeffizienten

$$\bar{T}_{00} = 1/N, \quad \bar{T}_{k0} = 2/N, \quad \bar{T}_{0l} = 2/N, \quad \bar{T}_{kl} = 4/N \quad (245.7)$$

für $k \in \{1, \dots, (K-1)/2\}$ und $l \in \{1, \dots, (L-1)/2\}$, wobei $N = KL$ gilt. Die Frequenzantwort (245.6) ist nach (244.33) die eines vierquadrantensymmetrischen nichtrekursiven Filters mit Nullphase. Die Filtergleichung ist durch (244.32) gegeben. Setzt man darin (245.7) ein, erhält man wegen

$$y(m, n) = \frac{1}{N} \sum_{k=-(K-1)/2}^{(K-1)/2} \sum_{l=-(L-1)/2}^{(L-1)/2} x(m-k, n-l) \quad (245.8)$$

das Ausgangssignal $y(m, n)$ als Mittelwert über die N Eingangswerte $x(m-k, n-l)$. Mit $K = L$ ist das System wegen (244.22) oktagonale.

Beispiel 1: Für $K = L = 3$ erhält man aus (245.6) mit (245.7) wegen $N = 9$ die Frequenzantwort

$$R^{2,2}(\Omega_1, \Omega_2) = \frac{1}{9} (1 + 2(\cos\Omega_1 + \cos\Omega_2) + 4(\cos\Omega_1 \cos\Omega_2)) . \quad (245.9)$$

Die Filtergleichung ergibt sich nach (245.8)

$$y(m, n) = \frac{1}{9} (x(m-1, n-1) + x(m-1, n) + x(m-1, n+1) + x(m, n-1) + x(m, n) + x(m, n+1) + x(m+1, n-1) + x(m+1, n) + x(m+1, n+1)) \quad (245.10)$$

als Mittelwertbildung über den zentralen Eingangssignalwert $x(m, n)$ und den Eingangssignalwerten $x(m-k, n-l)$ der sogenannten *Achter-Nachbarschaft*. Δ

Betrachtet wird nun ein zweidimensionales nichtrekursives Filter, dessen Frequenzantwort aus (245.3) berechnet wird. Setzt man für $H_1(j\Omega_1)$ wieder (245.4) ein und wählt weiter mit der ungeraden Zahl K die Frequenzantwort (226.7)

$$H_2(j\Omega_2) = V^{K-1}(\Omega) = \frac{2}{K} \sum_{k=1}^{(K-1)/2} \cos k\Omega_2 \quad (245.11)$$

sowie

$$H_3(j\Omega_1) = \frac{K}{2K-1} I(\Omega_1) \quad (245.12)$$

mit der Frequenzantwort $I(\Omega_1) = 1$ des Identitätsfilters, ergibt sich

$$H(j\Omega_1, j\Omega_2) = V^{K-1}(\Omega_1, \Omega_2) = \frac{1}{2K-1} [1 + 2 \sum_{k=1}^{(K-1)/2} (\cos k\Omega_1 + \cos k\Omega_2)] . \quad (245.13)$$

Vergleicht man diesen Ausdruck mit (244.21), erhält man die Koeffizienten

$$\bar{T}_{00} = 1/(2K-1), \quad \bar{T}_{0k} = \bar{T}_{k0} = 2\bar{T}_{00}, \quad \bar{T}_{kl} = 0 \quad (245.14)$$

für $k, l \in \{1, \dots, (K-1)/2\}$. Die Frequenzantwort (245.13) ist nach (244.22) die eines oktagonalen Systems, dessen Filtergleichung mit (245.14) aus (244.32) folgt

$$y(m, n) = \frac{1}{2K-1} (x(m, n) + \sum_{k=1}^{(K-1)/2} (x(m+k, n) + x(m-k, n) + x(m, n+k) + x(m, n-k))) . \quad (245.15)$$

Dieses Filter wird zumeist für $K = 3$ verwendet, worauf das folgende Beispiel eingeht.

Beispiel 2: Für $K = 3$ erhält man aus (245.13) die Frequenzantwort

$$V^2(\Omega_1, \Omega_2) = \frac{1}{5} (1 + 2\cos\Omega_1 + 2\cos\Omega_2) . \quad (245.16)$$

Aus der Filtergleichung

$$y(m, n) = \frac{1}{5} (x(m-1, n) + x(m, n) + x(m+1, n) + x(m, n+1) + x(m, n-1)) \quad (245.17)$$

erkennt man, daß sich der Ausgangssignalwert $y(m, n)$ als Mittelwert des zentralen Eingangssignalwertes $x(m, n)$ und den Eingangssignalwerten $x(m-k, n-l)$ aus der sogenannten *Vierer-Nachbarschaft* ergibt. Δ

In den Beispielen 1 und 2 wurden die Achter- und die Vierer-Nachbarschaft eingeführt. Ausgedehntere Nachbarschaften werden im Kapitel 376 definiert.

Ein weiteres, häufig verwendetes nichtrekursives Elementarfilter ist das zweidimensionale Binomialfilter. Man erhält es, indem die Frequenzantworten zweier eindimensionaler Binomialfilter nach (226.5) mit $\alpha = (K-1)/2$ und $\beta = (L-1)/2$ gebildet

$$H_1(j\Omega_1) = B^{K-1}(\Omega_1) = 2^{-K+1} \left(\frac{(2\alpha)!}{\alpha!\alpha!} + \sum_{k=1}^{(K-1)/2} \frac{2(2\alpha)!}{(\alpha+k)!(\alpha-k)!} \cos k\Omega_1 \right) \quad (245.18)$$

$$H_2(j\Omega_2) = B^{L-1}(\Omega_2) = 2^{-L+1} \left(\frac{(2\beta)!}{\beta!\beta!} + \sum_{l=1}^{(L-1)/2} \frac{2(2\beta)!}{(\beta+l)!(\beta-l)!} \cos l\Omega_2 \right) \quad (245.19)$$

und in (245.1) eingesetzt werden. Die Frequenzantwort des zweidimensionalen Binomialfilters lautet dann

$$H(j\Omega_1, j\Omega_2) = B^{K-1, L-1}(\Omega_1, \Omega_2) = \sum_{k=0}^{(K-1)/2} \sum_{l=0}^{(L-1)/2} \bar{F}_{kl} \cos k\Omega_1 \cos l\Omega_2 \quad (245.20)$$

mit den Koeffizienten

$$\bar{F}_{00} = 2^{-(K+L-2)} \frac{(2\alpha)!(2\beta)!}{\alpha!\alpha!\beta!\beta!}, \quad \bar{F}_{01} = 2^{-(K+L-3)} \frac{(2\alpha)!(2\beta)!}{\alpha!\alpha!(\beta+1)!(\beta-1)!},$$

$$\bar{F}_{k0} = 2^{-(K+L-3)} \frac{(2\alpha)!(2\beta)!}{(\alpha+k)!(\alpha-k)!\beta!\beta!},$$

$$\bar{F}_{kl} = 2^{-(K+L-4)} \frac{(2\alpha)!(2\beta)!}{(\alpha+k)!(\alpha-k)!(\beta+l)!(\beta-l)!} \quad (245.21)$$

für $k \in \{1, \dots, (K-1)/2\}$ und $l \in \{1, \dots, (L-1)/2\}$. Die Filtergleichung des für $K = L$ nach (244.22) oktagonalen Binomialfilters erhält man aus (244.32). Dieses Filter ist näherungsweise rotationssymmetrisch, wie graphische Darstellungen seiner Frequenzantworten zeigen (Jähne 1989, S.96).

Das Rechteckfilter und das Binomialfilter sind Tiefpaßfilter und bewirken eine Glättung des Eingangssignals. Möchte man Kanten in einem digitalen Bild lokalisieren, wird zum

Beispiel das *zweidimensionale Laplacefilter* eingesetzt, da es die zweiten Ableitungen ermittelt und daher Krümmungen anzeigt. Das zweidimensionale Laplacefilter entwickelt man wieder aus dem eindimensionalen Laplacefilter, indem die Frequenzantworten

$$H_1(j\Omega_1) = L^2(\Omega_1) = -2 + 2\cos\Omega_1 \quad (245.22)$$

$$H_2(j\Omega_2) = L^2(\Omega_2) = -2 + 2\cos\Omega_2 \quad (245.23)$$

aus (226.15) in (245.2) eingeführt werden. Man erhält als Ergebnis die Frequenzantwort des zweidimensionalen Laplacefilters

$$H_1(j\Omega_1, j\Omega_2) = L^{2,2}(\Omega_1, \Omega_2) = -4 + 2\cos\Omega_1 + 2\cos\Omega_2 \quad (245.24)$$

Der Vergleich mit (244.21) liefert die Koeffizienten

$$\bar{f}_{00} = -4, \bar{f}_{01} = \bar{f}_{10} = 2, \bar{f}_{11} = 0 \quad (245.25)$$

Das zweidimensionale Laplacefilter mit der aus (244.32) folgenden Filtergleichung

$$y(m, n) = x(m+1, n) - 4x(m, n) + x(m-1, n) + x(m, n+1) + x(m, n-1) \quad (245.26)$$

ist wegen (244.22) oktagonal. In (245.26) gehen wieder neben dem zentralen Eingangssignalwert $x(m, n)$ die Eingangssignalwerte aus der Vierer-Nachbarschaft ein.

Neben dem Laplacefilter werden für die Kantenextraktion auch die sogenannten *Sobeloperatoren* eingesetzt, die Gradientenoperatoren darstellen (Jähne 1989, S.107). Sobeloperatoren erhält man zum Beispiel aus (245.3) mit

$$H_1(j\Omega_1) = 2I(\Omega_1) \quad (245.27)$$

$$H_2(j\Omega_2) = 3V^2(\Omega_2) \quad (245.28)$$

$$H_3(j\Omega_1) = -(1/4) G^2(j\Omega_1) \quad (245.29)$$

wobei $I(\Omega_1)$ aus (245.12), $V^2(\Omega_2)$ aus (226.7) und $G^2(j\Omega_1)$ aus (226.10) stammen. Die zweidimensionale Frequenzantwort lautet dann

$$\begin{aligned} H(j\Omega_1, j\Omega_2) &= j\left(\frac{1}{2} \sin\Omega_1 + \frac{1}{2} \sin\Omega_1 \cos\Omega_2\right) \\ &= j\left(\frac{1}{2} \sin\Omega_1 + \frac{1}{4} \sin(\Omega_1 - \Omega_2) + \frac{1}{4} \sin(\Omega_1 + \Omega_2)\right) \end{aligned} \quad (245.30)$$

Vergleicht man dieses Ergebnis mit (244.44) und (244.42), ergeben sich die Koeffizienten

$$\bar{d}_{00} = \bar{d}_{01} = 0, \bar{d}_{1,-1} = 1/4, \bar{d}_{10} = 1/2, \bar{d}_{11} = 1/4 \quad (245.31)$$

Die Filtergleichung erhält man aus (244.45) zu

$$y(m, n) = \frac{1}{8} [x(m+1, n-1) - x(m-1, n+1) + 2x(m+1, n) - 2x(m-1, n)]$$

$$+x(m+1, n+1) - x(m-1, n-1)] \quad (245.32)$$

Ein weiterer Sobeloperator ergibt sich nach Austausch von Ω_1 und Ω_2 in (245.30)

$$H(j\Omega_1, j\Omega_2) = j\left(\frac{1}{2} \sin\Omega_2 - \frac{1}{4} \sin(\Omega_1 - \Omega_2) + \frac{1}{4} \sin(\Omega_1 + \Omega_2)\right) \quad (245.33)$$

In diesem Fall führt der Vergleich mit (244.44) und (244.42) auf die Koeffizienten

$$\bar{d}_{00} = \bar{d}_{10} = 0, \quad \bar{d}_{01} = 1/2, \quad \bar{d}_{1,-1} = -1/4, \quad \bar{d}_{11} = 1/4 \quad (245.34)$$

Folglich lautet die Filtergleichung

$$y(m, n) = \frac{1}{8} [2x(m, n+1) - 2x(m, n-1) - x(m+1, n-1) + x(m-1, n+1) + x(m+1, n+1) - x(m-1, n-1)] \quad (245.35)$$

Die Bedeutung der zweiquadrantensymmetrischen Sobeloperatoren liegt in der Verwendung der Gradientenfilter $G^2(j\Omega_1)$ in (245.30) und $G^2(j\Omega_2)$ in (245.33). Das Filter $G^2(j\Omega_1)$ bildet, wie im Beispiel 1 des Kapitels 226 für $G^1(j\Omega)$ erläutert, die erste Ableitung in Richtung der m -Achse, $G^2(j\Omega_2)$ in Richtung der n -Achse. Die gemeinsame Anwendung von (245.32) und (245.35), worauf im Kapitel 248 noch eingegangen wird, führt also zur Extraktion beliebig orientierter Kanten.

246 Entwurf zweidimensionaler nichtrekursiver Filter

Wie schon erwähnt, sind bei einigen Anwendungen Filter mit rotationssymmetrischen Frequenzantworten erwünscht, da sie sich isotrop verhalten. Man wird daher bestrebt sein, digitale Filter zu entwerfen, die diese Forderung weitgehend erfüllen und ideale Frequenzantworten besser approximieren, als es die Elementarfilter vermögen. Wie im Folgenden erläutert, benötigt man dazu erstens die Idealfunktion eines rotationssymmetrischen Filters und wählt zweitens als Realfunktion die eines vierquadrantensymmetrischen oder oktagonalen Systems.

Zur Lösung der am Ende des Kapitels 241 formulierten Entwurfsaufgabe durch nichtrekursive Filter werden hier die allgemein anwendbare Methode der Parameterschätzung im linearen Modell und die Fensterfunktionsmethode vorgestellt. Weitere Verfahren sind bei (Fiasconaro 1979, S.88; Rabiner und Gold 1975, S.455; Cappellini et al. 1978, S.72) beschrieben.

Das Schätzverfahren im linearen Modell wurde für rekursive Filter am Ende des Kapitels 243 erläutert. Es soll nun auf nichtrekursive Filter angewendet werden, so daß sich der Vektor β der Filterkoeffizienten aus (241.58) zu

$$\beta = [d_{00}, d_{10}, \dots, d_{K-1,0}, d_{01}, \dots, d_{K-1,1}, \dots, d_{0,L-1}, \dots, d_{K-1,L-1}]' \quad (246.1)$$

reduziert. Analog zu der Vorgehensweise im Kapitel 243 wird das Amplitudenspektrum $A(\Omega_1, \Omega_2)$ als Idealfunktion gewählt. Als Realfunktion soll die Amplitude (244.21)

$$A_0(\Omega_1, \Omega_2) = \sum_{k=0}^{(K-1)/2} \sum_{l=0}^{(L-1)/2} \bar{f}_{kl} \cos k\Omega_1 \cos l\Omega_2 \quad (246.2)$$

eines vierquadrantensymmetrischen nullphasigen Filters dienen.

Die Parameterschätzung kann nach (243.10) bis (243.19) ablaufen, wobei in (243.11) $\bar{a}_{kl} = \bar{f}_{kl}$ gilt und $2K$ und $2L$ durch $(K-1)/2$ und $(L-1)/2$ zu ersetzen sind. Bei einer engmaschigen Diskretisierung von Ω_1 und Ω_2 sowie bei einer großen Anzahl unbekannter Filterkoeffizienten wird aber der numerische Aufwand groß, so daß für die Parameterschätzung eine effizientere Methode gewählt wird, die von Fritsch (1982, S.36) entwickelt wurde. Zunächst werden die unbekanntenen Koeffizienten \bar{f}_{kl} aus (246.2) in der $(K+1)/2 \times (L+1)/2$ Matrix

$$B = \begin{bmatrix} \bar{f}_{00} & \bar{f}_{01} & \dots & \bar{f}_{0, (L-1)/2} \\ \bar{f}_{10} & \bar{f}_{11} & \dots & \bar{f}_{1, (L-1)/2} \\ \dots & \dots & \dots & \dots \\ \bar{f}_{(K-1)/2, 0} & \bar{f}_{(K-1)/2, 1} & \dots & \bar{f}_{(K-1)/2, (L-1)/2} \end{bmatrix} \quad (246.3)$$

zusammengefaßt. Führt man weiter mit $(K+1)/2 = u_1$ den $u_1 \times 1$ Vektor

$$\mathbf{x}_1 = \mathbf{x}_1(\Omega_1) = [1, \cos\Omega_1, \dots, \cos((K-1)\Omega_1/2)]' \quad (246.4)$$

und mit $(L+1)/2 = u_2$ den $u_2 \times 1$ Vektor

$$\mathbf{x}_2 = \mathbf{x}_2(\Omega_2) = [1, \cos\Omega_2, \dots, \cos((L-1)\Omega_2/2)]' \quad (246.5)$$

ein, läßt sich für (246.2) schreiben

$$A_0(\Omega_1, \Omega_2) = \mathbf{x}_1' B \mathbf{x}_2 \quad (246.6)$$

Mit dem Kronecker-Produkt \otimes folgt hieraus (Koch 1987, S.48)

$$A_0(\Omega_1, \Omega_2) = \text{vec}(\mathbf{x}_1' B \mathbf{x}_2) = (\mathbf{x}_2' \otimes \mathbf{x}_1') \text{vec} B \quad (246.7)$$

Definiert man $\mathbf{x}' = \mathbf{x}_2' \otimes \mathbf{x}_1'$ und $\beta = \text{vec} B$, stimmt (246.7) mit der Darstellung (227.5) für die Amplitude im eindimensionalen Fall überein. Der Approximationsfehler (241.60) ergibt sich zu

$$e(\Omega_1, \Omega_2) = (\mathbf{x}_2' \otimes \mathbf{x}_1') \text{vec} B - A(\Omega_1, \Omega_2) \quad (246.8)$$

Nach der Diskretisierung von Ω_1 in die n_1 Werte Ω_{1i} mit $i \in \{1, \dots, n_1\}$ und von Ω_2 in

die n_2 Werte Ω_{2j} mit $j \in \{1, \dots, n_2\}$ erhält man mit $e_{ij} = e(\Omega_{1i}, \Omega_{2j})$, $\mathbf{x}_{1i} = \mathbf{x}_1(\Omega_{1i})$, $\mathbf{x}_{2j} = \mathbf{x}_2(\Omega_{2j})$ und $y_{ij} = A(\Omega_{1i}, \Omega_{2j})$ aus (246.8)

$$y_{ij} + e_{ij} = (\mathbf{x}_{2j}' \otimes \mathbf{x}_{1i}') \text{vecB} \quad (246.9)$$

analog zu (243.14). Hieraus ergibt sich weiter mit den $n_1 \times n_2$ Matrizen $\mathbf{E} = (e_{ij})$ und $\mathbf{Y} = (y_{ij})$ sowie der $n_1 \times u_1$ Matrix $\mathbf{X}_1 = [\mathbf{x}_{11}, \mathbf{x}_{12}, \dots, \mathbf{x}_{1n_1}]'$ und der $n_2 \times u_2$ Matrix $\mathbf{X}_2 = [\mathbf{x}_{21}, \mathbf{x}_{22}, \dots, \mathbf{x}_{2n_2}]'$ das lineare Modell der Parameterschätzung

$$\text{vecY} + \text{vecE} = (\mathbf{X}_2 \otimes \mathbf{X}_1) \text{vecB} . \quad (246.10)$$

Führt man noch die $n_1 n_2 \times n_1 n_2$ Gewichtsmatrix $\mathbf{P} = \mathbf{I}$ ein, erhält man nach (227.9) unter Anwendung der Rechenregeln für das Kronecker-Produkt die Schätzung (Fritsch 1982, S.37; Koch 1987, S.22)

$$\begin{aligned} \hat{\text{vecB}} &= ((\mathbf{X}_2' \otimes \mathbf{X}_1') (\mathbf{X}_2 \otimes \mathbf{X}_1))^{-1} (\mathbf{X}_2' \otimes \mathbf{X}_1') \text{vecY} \\ &= (\mathbf{X}_2' \mathbf{X}_2 \otimes \mathbf{X}_1' \mathbf{X}_1)^{-1} (\mathbf{X}_2' \otimes \mathbf{X}_1') \text{vecY} \\ &= [(\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2' \otimes (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1'] \text{vecY} . \end{aligned} \quad (246.11)$$

Für die Matrix $\hat{\mathbf{E}}$ der Residuen ergibt sich aus (246.10)

$$\begin{aligned} \hat{\text{vecE}} &= [(\mathbf{X}_2 \otimes \mathbf{X}_1) [(\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2' \otimes (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1'] - \mathbf{I}] \text{vecY} \\ &= [\mathbf{X}_2 (\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2' \otimes \mathbf{X}_1 (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' - \mathbf{I}] \text{vecY} . \end{aligned} \quad (246.12)$$

Wie in (246.7) können (246.11) und (246.12) weiter umgeschrieben werden in

$$\hat{\mathbf{B}} = (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{Y} \mathbf{X}_2 (\mathbf{X}_2' \mathbf{X}_2)^{-1} \quad (246.13)$$

$$\hat{\mathbf{E}} = \mathbf{X}_1 (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{Y} \mathbf{X}_2 (\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2' - \mathbf{Y} . \quad (246.14)$$

Diese Schätzung der unbekanntem Filterkoeffizienten erweist sich hinsichtlich der benötigten Rechenzeit und des erforderlichen Speicherplatzes als effizienter als die Vorgehensweise aus Kapitel 243, da in (246.13) nur die $u_1 \times u_1$ Matrix $\mathbf{X}_1' \mathbf{X}_1$ und die $u_2 \times u_2$ Matrix $\mathbf{X}_2' \mathbf{X}_2$ zu invertieren sind. Mit (246.13) berechnet sich die Realfunktion (246.6) zu

$$\hat{A}_0(\Omega_1, \Omega_2) = A_0(\Omega_1, \Omega_2, \hat{\boldsymbol{\beta}}) = \mathbf{x}_1' \hat{\mathbf{B}} \mathbf{x}_2 . \quad (246.15)$$

Wird für Ω_1 und Ω_2 die gleiche Diskretisierung vorgenommen, vereinfachen sich die Gleichungen (246.13) und (246.14) wegen $\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}_0$ zu

$$\hat{\mathbf{B}} = (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{Y} \mathbf{X}_0 (\mathbf{X}_0' \mathbf{X}_0)^{-1} , \quad (246.16)$$

$$\hat{\mathbf{B}} = \mathbf{X}_0(\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{Y} \mathbf{X}_0(\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' - \mathbf{Y} . \quad (246.17)$$

Ist die Idealfunktion $A(\Omega_1, \Omega_2)$ rotationssymmetrisch, ergibt sich \mathbf{Y} wegen $\mathbf{Y} = \mathbf{Y}'$ als symmetrische Matrix, so daß aus (246.16) $\hat{\mathbf{B}} = \hat{\mathbf{B}}'$ folgt. Das entwickelte Filter ist dann nach (244.22) oktagonal.

Beispiel 1: Gegeben sei als Idealfunktion das Amplitudenspektrum $A(\Omega_1, \Omega_2)$ eines Tiefpaßfilters. Um eine gute Approximation zu erzielen, wird zwischen der Durchlaßgrenzfrequenz $\Omega_D = 0,14\pi$ und der Sperrgrenzfrequenz $\Omega_S = 0,32\pi$ ein Übergangsbereich analog zu (222.9) eingeführt, in dem die Idealfunktion linear abfällt. Nach der Diskretisierung auf einem $n \times n$ Gitter mit $n = n_1 = n_2$ erhält man für die Elemente $y_{ij} = A(\Omega_{1i}, \Omega_{2j})$ der $n \times n$ Matrix \mathbf{Y} mit $\Omega_{ij}^2 = \Omega_{1i}^2 + \Omega_{2j}^2$

$$\begin{aligned} y_{ij} &= 1 && \text{für } 0 \leq |\Omega_{ij}| \leq \Omega_D \\ y_{ij} &= (\Omega_S - |\Omega_{ij}|) / (\Omega_S - \Omega_D) && \text{für } \Omega_D < |\Omega_{ij}| \leq \Omega_S \\ y_{ij} &= 0 && \text{für } \Omega_S < |\Omega_{ij}| \leq \pi . \end{aligned} \quad (246.18)$$

Als Realfunktion wird $A_0(\Omega_1, \Omega_2)$ aus (246.2) mit $K = L = 15$ gewählt. Die Schätzung der mit $u_1 = u_2 = u = 8$ insgesamt $u^2 = 64$ Koeffizienten in (246.3) folgt wegen $\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}_0$ aus (246.16). Die Anzahl der unabhängigen Koeffizienten reduziert sich wegen $\mathbf{Y} = \mathbf{Y}'$ auf $u(u+1)/2 = 36$. In Abbildung 246-1 bis 246-3 sind die Idealfunktion $A(\Omega_1, \Omega_2)$, die nach (246.15) geschätzte Realfunktion $\hat{A}_0(\Omega_1, \Omega_2)$ sowie die Approximationsfehlerfunktion $\hat{e}(\Omega_1, \Omega_2) = \hat{A}_0(\Omega_1, \Omega_2) - A(\Omega_1, \Omega_2)$ dargestellt. Δ

Die Approximation einer Idealfunktion durch eine Realfunktion nach der Methode der kleinsten Quadrate kann durch die Einführung von Ungleichungen als Nebenbedingungen noch erweitert werden, worauf bei (Fritsch 1982, S.38) eingegangen wird.

Als weiteres Approximationsverfahren wird im folgenden noch die Fensterfunktionsmethode behandelt, bei der sehr lange oder unendlich lange Impulsantworten in kurze oder zumindest endlich lange Impulsantworten umgewandelt werden. Hierzu werden diskrete zweidimensionale Funktionen $w(m, n)$ eingeführt, für die gilt

$$w(m, n) \begin{cases} \neq 0 & \text{für } m \in \{m_1, \dots, m_2\} \text{ und } n \in \{n_1, \dots, n_2\} \\ = 0 & \text{sonst .} \end{cases} \quad (246.19)$$

Sie sind also nur in einem zusammenhängenden zweidimensionalen Bereich von Null verschieden und werden daher als *zweidimensionale Fensterfunktionen* bezeichnet. Ihre Fourier-Transformation, die sich (235.15) entsprechend zu

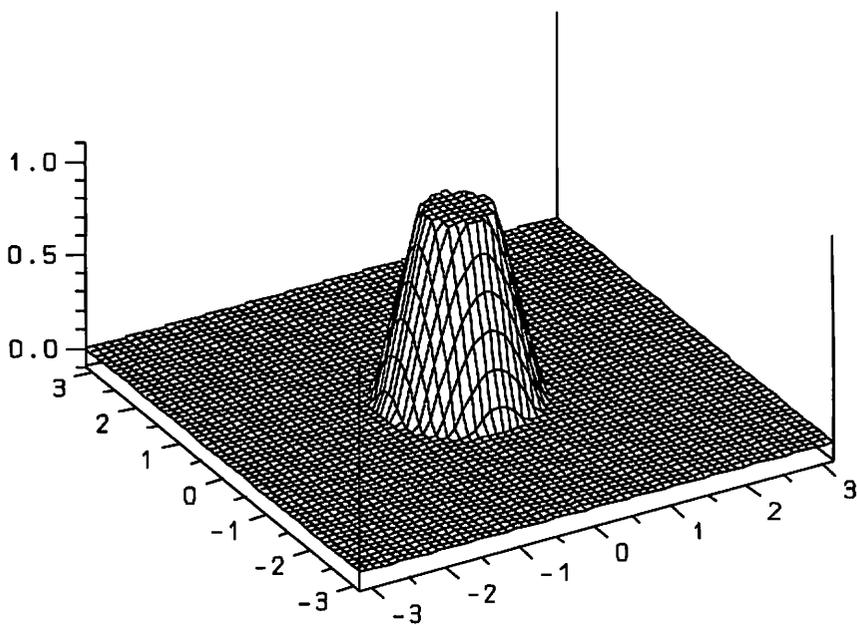


Abb. 246-1: Idealfunktion $A(\Omega_1, \Omega_2)$

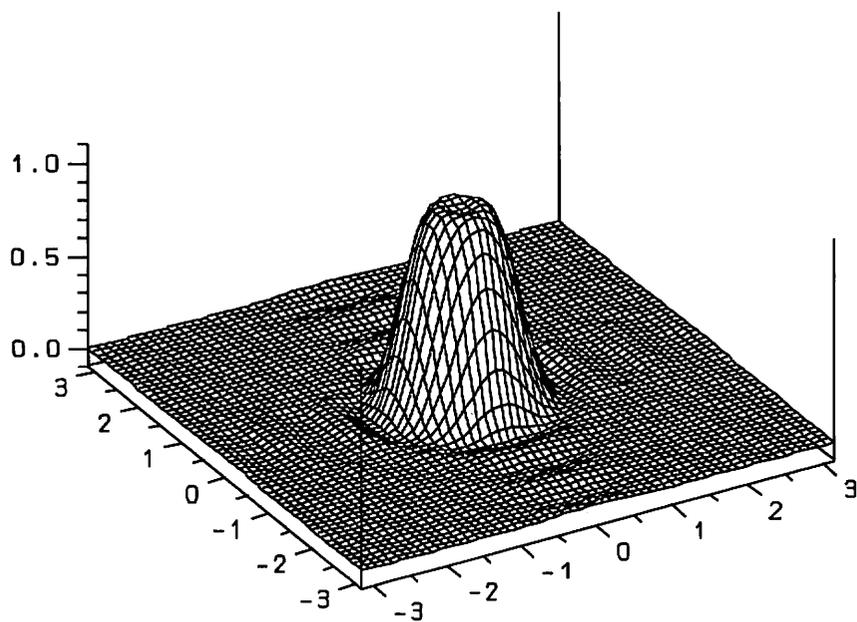


Abb. 246-2: Realfunktion $\hat{A}_0(\Omega_1, \Omega_2)$

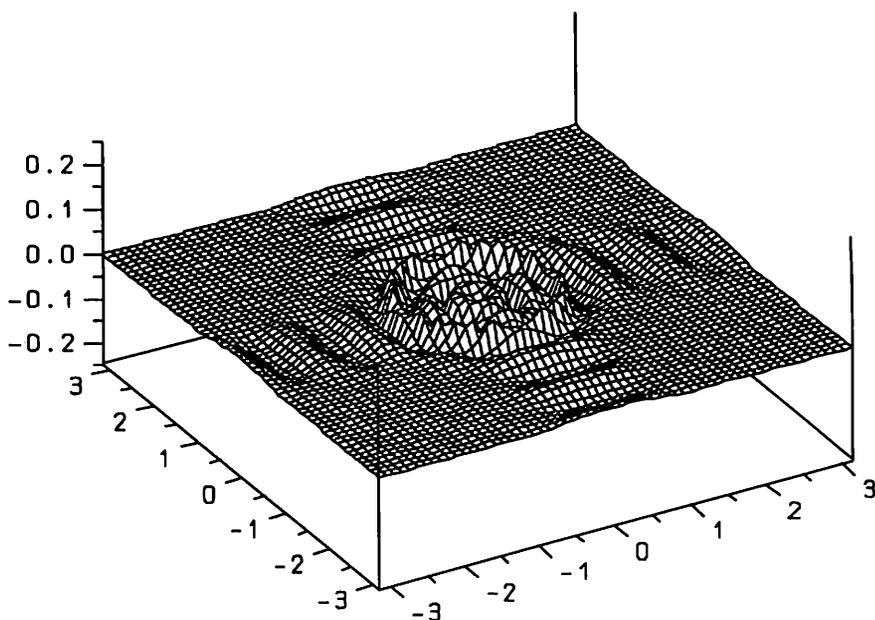


Abb. 246-3: Approximationsfehlerfunktion $\hat{e}(\Omega_1, \Omega_2)$

$$W(j\Omega_1, j\Omega_2) = \sum_{m=m_1}^{m_2} \sum_{n=n_1}^{n_2} w(m, n) e^{-j(m\Omega_1 + n\Omega_2)} \quad (246.20)$$

ergibt, nennt man *zweidimensionales Spektralfenster*.

Als Idealfunktion sei eine unendlich lange Impulsantwort gewählt

$$f(x_1, x_2) = h(m, n) \quad (246.21)$$

mit

$$h(m, n) \neq 0 \quad \text{für } m, n \in \{0, \pm 1, \pm 2, \dots\} .$$

Die Realfunktion erhält man dann mit (246.19) zu

$$h_0(m, n) = w(m, n)h(m, n) \quad (246.22)$$

mit

$$h_0(m, n) \begin{cases} \neq 0 & \text{für } m \in \{m_1, \dots, m_2\} \text{ und } n \in \{n_1, \dots, n_2\} \\ = 0 & \text{sonst} . \end{cases}$$

Mit $h_0(m, n)$ wird also eine endlich lange Impulsantwort gewonnen, die wegen (235.30) auf eine nichtreursive Filterung führt. Die Filtergleichung folgt aus (235.10) zu

$$y(m, n) = \sum_{k=m_1}^{m_2} \sum_{l=n_1}^{n_2} h_0(k, l) x(m-k, n-l) . \quad (246.23)$$

Die zugehörige Frequenzantwort erhält man mit (235.15) und (246.22) zu

$$\begin{aligned} H_0(j\Omega_1, j\Omega_2) &= \sum_{k=m_1}^{m_2} \sum_{l=n_1}^{n_2} h_0(k, l) e^{-j(k\Omega_1 + l\Omega_2)} \\ &= \sum_{k=m_1}^{m_2} \sum_{l=n_1}^{n_2} w(k, l) h(k, l) e^{-j(k\Omega_1 + l\Omega_2)} . \end{aligned} \quad (246.24)$$

Setzt man hierin (235.16) ein, ergibt sich

$$\begin{aligned} H_0(j\Omega_1, j\Omega_2) &= \frac{1}{4\pi^2} \sum_{k=m_1}^{m_2} \sum_{l=n_1}^{n_2} w(k, l) \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} H(j\beta_1, j\beta_2) \\ &\quad e^{j(k\beta_1 + l\beta_2)} d\beta_1 d\beta_2 e^{-j(k\Omega_1 + l\Omega_2)} , \end{aligned}$$

wobei β_1 und β_2 Integrationsvariablen bedeuten. Mit (246.20) erhält man weiter

$$\begin{aligned} H_0(j\Omega_1, j\Omega_2) &= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} H(j\beta_1, j\beta_2) \sum_{k=m_1}^{m_2} \sum_{l=n_1}^{n_2} w(k, l) \\ &\quad e^{-j(k(\Omega_1 - \beta_1) + l(\Omega_2 - \beta_2))} d\beta_1 d\beta_2 \\ &= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} H(j\beta_1, j\beta_2) W(j(\Omega_1 - \beta_1), j(\Omega_2 - \beta_2)) d\beta_1 d\beta_2 . \end{aligned} \quad (246.25)$$

Dieses Ergebnis läßt sich, wie der Vergleich mit (235.9) zeigt, bis auf den Faktor $1/4\pi^2$ als die Faltung der Frequenzantwort $H(j\Omega_1, j\Omega_2)$ mit dem Spektralfenster $W(j\Omega_1, j\Omega_2)$ interpretieren

$$H_0(j\Omega_1, j\Omega_2) = \frac{1}{4\pi^2} H(j\Omega_1, j\Omega_2) * W(j\Omega_1, j\Omega_2) . \quad (246.26)$$

Die Frequenzantwort $H_0(j\Omega_1, j\Omega_2)$ enthält nicht nur Frequenzanteile aus $H(j\Omega_1, j\Omega_2)$, sondern auch aus $W(j\Omega_1, j\Omega_2)$, sie ist also "spektral verschmiert", worauf bereits im Zusammenhang mit (227.20) hingewiesen wurde.

Liegt mit $H(j\Omega_1, j\Omega_2)$ die Frequenzantwort eines nullphasigen Systems vor, gilt nach (244.24) also $H(j\Omega_1, j\Omega_2) = A(\Omega_1, \Omega_2)$, ist es sinnvoll, auch ein nullphasiges Spektralfenster zu wählen. Die Frequenzantwort (246.25) ist dann reell, so daß das resultierende Filter ebenfalls nullphasig ist.

Eine zweidimensionale Fensterfunktion kann beispielsweise als separierbar angenommen

werden

$$w(m, n) = w_1(m)w_2(n) \quad , \quad (246.27)$$

worin $w_1(m)$ und $w_2(n)$ zwei eindimensionale Fensterfunktionen aus (227.13) bedeuten. Eine weitere Möglichkeit besteht in der Definition

$$w(m, n) = w(q) \quad \text{mit} \quad q = (m^2+n^2)^{1/2} \quad , \quad (246.28)$$

wobei $q \in \mathbb{R}$ gilt. In diesem Fall liegt also eine kontinuierliche eindimensionale rotations-symmetrische Fensterfunktion vor. Als $w(q)$ kann das Rechtecksignal des Beispiels 2 oder das Dreiecksignal des Beispiels 3 aus Kapitel 214 verwendet werden. Weitere zwei-dimensionale Fensterfunktionen sind unter anderem bei (Dudgeon und Mersereau 1984, S.118; Rabiner und Gold 1975, S.456) beschrieben.

Beispiel 2: Als Idealfunktion sei die Impulsantwort (235.19) des idealen rotationssymmetrischen Tiefpaßfilters gewählt. Die Realfunktion erhält man dann aus (246.22) zu

$$h_0(m, n) = w(m, n) \frac{\Omega_m J_1(q\Omega_m)}{2\pi q} \quad (246.29)$$

mit $q = (m^2+n^2)^{1/2}$. Führt man das Rechteckfenster

$$w(m, n) = w(q) = \begin{cases} 1 & \text{für } q \leq c \\ 0 & \text{sonst} \end{cases} \quad (246.30)$$

ein, folgt aus (246.29)

$$h_0(m, n) = h_0(q) = \begin{cases} \Omega_m J_1(q\Omega_m) / (2\pi q) & \text{für } q \leq c \\ 0 & \text{sonst} \end{cases} \quad (246.31)$$

Wegen

$$h_0(m, n) = h_0(-m, n) = h_0(-m, -n) = h_0(m, -n) \quad (246.32)$$

ist die Impulsantwort nach (244.31) vierquadrantensymmetrisch und wegen $h_0(m, n) = h_0(n, m)$ nach (244.23) oktagonale. Die Filtergleichung erhält man daher aus (244.32) mit $K = L$ und $(K-1)/2 \geq c$ wegen (246.31). Die Koeffizienten \bar{T}_{kl} berechnen sich zu

$$\bar{T}_{00} = h_0(0)$$

$$\bar{T}_{m0} = 2h_0(q) \begin{cases} \neq 0 & \text{für } q = m \leq c \\ = 0 & \text{sonst} \end{cases}$$

$$\bar{T}_{0n} = 2h_0(q) \begin{cases} \neq 0 & \text{für } q = n \leq c \\ = 0 & \text{sonst} \end{cases}$$

$$\bar{f}_{mn} = 4h_0(q) \begin{cases} \neq 0 & \text{für } q = (m^2+n^2)^{1/2} \leq c \\ = 0 & \text{sonst} \end{cases} \quad (246.33)$$

für $m, n \in \{1, \dots, (K-1)/2\}$.

▲

247 Implementierung zweidimensionaler linearer Filter

Im Gegensatz zu den eindimensionalen rekursiven Filtern sind die zweidimensionalen rekursiven Filter weniger verbreitet. Die Hauptgründe liegen in den aufwendigen Entwurfsverfahren und insbesondere in den Stabilitätsuntersuchungen. Sind indes die Filterkoeffizienten einmal bestimmt, ist die rekursive Filterung im Vergleich zur nichtrekursiven Filterung mit geringerem numerischen Aufwand verbunden, da mit einer endlichen Anzahl von Filterkoeffizienten eine unendliche Impulsantwort erhalten wird, wie der Vergleich von (235.26) mit (235.29) deutlich macht. Die *Implementierung zweidimensionaler rekursiver Filter* kann direkt über die Filtergleichungen, beispielsweise (241.3), (241.7), (241.11) oder (241.15) erfolgen. Im Fall einer nullphasigen Filterung mit Vierquadrantensymmetrie ist nach (242.40) oder (242.44) eine vierfache Rekursion erforderlich. Soll im Frequenzbereich gefiltert werden, können die Frequenzantworten (242.6), (242.24), (242.38) oder (242.42) verwendet werden.

Die *Implementierung zweidimensionaler nichtrekursiver Filter* läßt sich wie im eindimensionalen Fall im Orts- oder im Frequenzbereich vornehmen. Bei der ersten Möglichkeit wird für streng kausale Filter die diskrete Faltung (244.3) angewendet. Soll beispielsweise ein nullphasiges vierquadrantensymmetrisches Filter implementiert werden, so reduzieren sich die KL Multiplikationen in (244.29) aufgrund der Symmetrieeigenschaften (244.31) auf $(K+1)(L+1)/4$ Multiplikationen in (244.32). Eine weitere Verringerung auf $(K+1)(L+3)/8$ Multiplikationen tritt ein, falls eine oktagonale Impulsantwort vorliegt, worauf im Zusammenhang mit (244.23) eingegangen wurde. Für das Beispiel 1 des Kapitels 246 erhält man wegen $K = L = 15$ zunächst 225 Filterkoeffizienten d_{kl} nach (244.3), von denen letztlich aufgrund der Oktogonalität nur 36 Koeffizienten \bar{f}_{kl} zu bestimmen verbleiben.

Soll für die Filterung die Rechenzeit möglichst niedrig gehalten werden, kann es sich lohnen, die schnelle Faltung einzusetzen, also im Frequenzbereich zu filtern. Dies soll im folgenden für die streng kausale Filterung demonstriert werden. Gegeben sei das Signal

$$x(m, n) \quad \text{mit } m \in \{0, \dots, M_1-1\}, n \in \{0, \dots, M_2-1\} \quad (247.1)$$

und die Impulsantwort eines streng kausalen Systems

$$h(k, l) \text{ mit } k \in \{0, \dots, K_1-1\}, l \in \{0, \dots, K_2-1\} . \quad (247.2)$$

Mit den Größen

$$L_1 = 2^v \geq M_1 + K_1 - 1 \quad \text{und} \quad L_2 = 2^\kappa \geq M_2 + K_2 - 1 \quad \text{mit} \quad v, \kappa \in \{1, 2, \dots\} \quad (247.3)$$

konstruiert man sich die dimensionsgleichen Zahlenfelder

$$\tilde{x}(m, n) = \begin{cases} x(m, n) & \text{für } m \in \{0, \dots, M_1-1\}, n \in \{0, \dots, M_2-1\} \\ 0 & \text{für } m \in \{M_1, \dots, L_1-1\}, n \in \{M_2, \dots, L_2-1\} \end{cases} \quad (247.4)$$

$$\tilde{h}(m, n) = \begin{cases} h(m, n) & \text{für } m \in \{0, \dots, K_1-1\}, n \in \{0, \dots, K_2-1\} \\ 0 & \text{für } m \in \{K_1, \dots, L_1-1\}, n \in \{K_2, \dots, L_2-1\} . \end{cases} \quad (247.5)$$

Die diskreten Fourier-Transformationen (235.33) des Signals (247.4) und der Impulsantwort (247.5) lauten mit $W_{L_1} = e^{-j(2\pi/L_1)}$ und $W_{L_2} = e^{-j(2\pi/L_2)}$

$$\tilde{X}(k, l) = \sum_{m=0}^{L_1-1} \sum_{n=0}^{L_2-1} \tilde{x}(m, n) (W_{L_1})^{km} (W_{L_2})^{ln} , \quad (247.6)$$

$$\tilde{H}(k, l) = \sum_{m=0}^{L_1-1} \sum_{n=0}^{L_2-1} \tilde{h}(m, n) (W_{L_1})^{km} (W_{L_2})^{ln} \quad (247.7)$$

für $k \in \{0, \dots, L_1-1\}$ und $l \in \{0, \dots, L_2-1\}$. Sie lassen sich mit den eindimensionalen schnellen Fourier-Transformationen berechnen, wie im Zusammenhang mit (234.12) erläutert. Das gefilterte Ausgangssignal $y(m, n)$ erhält man aus der inversen diskreten Fourier-Transformation (235.32)

$$y(m, n) = \frac{1}{L_1 L_2} \sum_{k=0}^{L_1-1} \sum_{l=0}^{L_2-1} Y(k, l) (W_{L_1})^{-km} (W_{L_2})^{-ln} \quad (247.8)$$

mit $m \in \{0, \dots, L_1-1\}$, $n \in \{0, \dots, L_2-1\}$ und

$$Y(k, l) = \tilde{H}(k, l) \tilde{X}(k, l) . \quad (247.9)$$

Die beschriebenen Zusammenhänge gelten nach (247.2) für eine streng kausale Filterung. Soll indes eine nullphasige Filterung, die nichtkausal ist, vollzogen werden, ist analog (228.11) bis (228.14) zu verfahren.

Die Frage, ob die Filterung im Orts- oder im Frequenzbereich erfolgen soll, wird beispielsweise bei (Mersereau und Dudgeon 1978) angesprochen.

248 Zweidimensionale digitale nichtlineare und verschiebungsvariante Filter

Mit (235.8) wird ein allgemeines zweidimensionales kontinuierliches System

$$y(t_1, t_2) = \phi[x(t_1, t_2)] \quad (248.1)$$

definiert, das ein kontinuierliches Eingangssignal $x(t_1, t_2)$ in ein kontinuierliches Ausgangssignal $y(t_1, t_2)$ überführt. Entsprechend gilt für ein zweidimensionales diskretes System

$$y(m, n) = \phi[x(m, n)] \quad , \quad (248.2)$$

worin $x(m, n)$ das diskrete Eingangssignal und $y(m, n)$ das diskrete Ausgangssignal bedeuten. Bislang wurde das System nach (233.2) als linear und nach (233.3) als verschiebungsinvariant vorausgesetzt, so daß nach (235.10) die Filterung

$$y(m, n) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h(k, l) x(m-k, n-l) \quad (248.3)$$

folgt. Schreibt man hingegen

$$y(m, n) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h(m, n, k, l) x(m-k, n-l) \quad , \quad (248.4)$$

liegt eine *lineare verschiebungsvariante Filterung* vor, da die Impulsantwort $h(m, n, k, l)$ jetzt auch von den Variablen m, n des Ausgangssignals $y(m, n)$ abhängt. Wählt man beispielsweise

$$h(m, n, k, l) = \begin{cases} w(m, n) & \text{für } k=l=0 \\ 0 & \text{sonst} \end{cases} \quad (248.5)$$

erhält man aus (248.4)

$$y(m, n) = w(m, n) x(m, n) \quad . \quad (248.6)$$

Stellt $w(m, n)$ ein zweidimensionales Rechteckfenster dar, für das

$$w(m, n) = \begin{cases} 1 & \text{für } m \in \{m_1, \dots, m_2\} \text{ und } n \in \{n_1, \dots, n_2\} \\ 0 & \text{sonst} \end{cases} \quad (248.7)$$

mit $m_2 - m_1 = M - 1$ und $n_2 - n_1 = N - 1$ gilt, liefert (248.6) die Signalwerte

$$y(m,n) = \begin{cases} x(m,n) & \text{für } m \in \{m_1, \dots, m_2\} \text{ und } n \in \{n_1, \dots, n_2\} \\ 0 & \text{sonst,} \end{cases} \quad (248.8)$$

die als Eingangssignalwerte eines weiteren Filters mit der Transformationsgleichung (248.2)

$$\tilde{y}(k, l) = \phi[y(k, l)] \quad (248.9)$$

dienen können. Bei der *Medianfilterung* werden die MN Werte $y(m,n)$ für $m \in \{m_1, \dots, m_2\}$ und $n \in \{n_1, \dots, n_2\}$ derart in einen $MN \times 1$ Vektor $\mathbf{y} = (y_i)$ geschrieben, daß für die Komponenten y_i gilt

$$y_1 \leq y_2 \leq y_3 \dots \leq y_{MN} . \quad (248.10)$$

Dann bezeichnet

$$y_{\text{med}} = y_{(MN+1)/2} , \quad (248.11)$$

falls MN eine ungerade Zahl ist, oder

$$y_{\text{med}} = \frac{1}{2}(y_{MN/2} + y_{MN/2+1}) , \quad (248.12)$$

falls MN eine gerade Zahl ist, den Median. Das Medianfilter liefert als Ausgangssignalwert $\tilde{y}(k, l)$ an der Stelle $k = k_1$ und $l = l_1$ den Median der Eingangssignalwerte. Werden für die Rechteckfensterweiten M und N ungerade Zahlen vorausgesetzt, bestimmen sich m_1 und n_1 in (248.7) zu $m_1 = k_1 - (M-1)/2$ und $n_1 = l_1 - (N-1)/2$ sowie m_2 und n_2 zu $m_2 = k_1 + (M-1)/2$ und $n_2 = l_1 + (N-1)/2$. Man erhält dann mit (248.11)

$$\tilde{y}(k_1, l_1) = y_{\text{med}} = y_{(MN+1)/2} . \quad (248.13)$$

Das Medianfilter ist aufgrund der Anordnung (248.10) ein *nichtlineares verschiebungsvariantes Filter* und bewirkt eine Glättung des Eingangssignals. Es besitzt gegenüber den linearen Glättungsfilttern, wie den im Kapitel 245 vorgestellten Rechteck- und Binomialfilttern, den Vorteil, daß Ausreißer eliminiert werden (Jähne 1989, S.122; Bates und McDonnell 1989, S.229). Dies zeigt auch das folgende Beispiel.

Beispiel 1: Gegeben seien die Eingangswerte $x(m,n)$ für $m,n \in \{0, \dots, 5\}$ der folgenden Tabelle. Gesucht ist der medianefilterte Ausgangswert $\tilde{y}(3,3)$, wobei für die Fensterweiten $M = N = 3$ gesetzt wird. Mit $m_1 = n_1 = 2$ und $m_2 = n_2 = 4$ ergibt sich der Vektor \mathbf{y} mit den ihren Beträgen nach sortierten Komponenten y_i für $i \in \{1, \dots, 9\}$ des Datenfensters $y(m,n)$ aus (248.8) für $m,n \in \{2,3,4\}$ zu

$$\mathbf{y} = [1, 2, 2, 2, 3, 3, 4, 4, 16]'$$

$x(m,n)$	$n = 0$	1	2	3	4	5
$m = 0$	3	2	4	6	3	4
1	2	2	5	0	2	3
2	3	1	4	2	2	2
3	2	0	3	1	4	4
4	1	0	3	2	16	3
5	2	0	0	3	1	2

Nach (248.13) folgt dann

$$\tilde{y}(3,3) = 3 .$$

Bei dem Wert 16 für $x(4,4)$ handelt es sich offensichtlich um einen Ausreißer. Die Medianfilterung an dieser Stelle lautet

$$\tilde{y}(4,4) = 3 .$$

▲

Ein *nichtlineares verschiebungsinvariantes* Filter wird beispielsweise durch die Filtergleichung

$$y(m,n) = x(m,n)^{1/2} \quad (248.14)$$

definiert. Bezeichnet man das Ausgangssignal $y(m,n)$ aus (245.32) mit $y_1(m,n)$ und das Ausgangssignal $y(m,n)$ aus (245.35) mit $y_2(m,n)$, erhält man mit

$$x(m,n) = y_1(m,n)^2 + y_2(m,n)^2 \quad (248.15)$$

die vollständige Filtergleichung mit den Sobeloperatoren zur Extraktion beliebig orientierter Kanten (Kraus 1990, S.511).

Weitere Einzelheiten und Beispiele für verschiebungsvariante und nichtlineare Filter sind bei (Jähne 1989, S.122; Wahl 1989, S.76; He 1989, S.112) zu finden.

3 Stochastische Signale

Physikalische, geometrische oder chemische Größen können zeitlichen Änderungen unterworfen sein, die zufällig verlaufen und exakt nicht vorhersagbar sind. Als Beispiel sei die Höhe von Grundwasserständen über einen längeren Zeitraum genannt. Zur Beschreibung dieser Phänomene lassen sich daher die im Kapitel 2 behandelten deterministischen Signale nicht verwenden. Man benötigt stochastische Signale, die sich in Abhängigkeit von der Zeit zufällig verhalten. Völlig unregelmäßig verlaufen die zeitlichen Änderungen der Signale im allgemeinen aber nicht, gewisse Intervalle, in denen sich die Werte bewegen, und Erhaltungstendenzen sind erkennbar. Die erwähnten Grundwasserstände werden beispielsweise bei stündlicher Ablesung keine drastischen Sprünge aufweisen. Es bietet sich daher zur Beschreibung eines solchen Phänomens an, eine Zufallsvariable als Funktion der Zeit einzuführen. Man erhält dann als stochastisches Signal einen sogenannten stochastischen Prozeß.

Im folgenden werden zunächst die eindimensionalen Zufallsprozesse behandelt und dann spezielle Zufallsprozesse vorgestellt. Zur Darstellung dieser Prozesse werden Parameter benötigt, die geschätzt werden können. Anschließend folgt die Verallgemeinerung auf Zufallsfelder, wobei die zweidimensionalen Markoff-Zufallsfelder mit ihren Gibbs-Verteilungen im Hinblick auf die digitale Bildverarbeitung besondere Beachtung finden.

31 Eindimensionale Zufallsprozesse

311 Definitionen

Als *stochastischen Prozeß*, *Zufallsprozeß* oder *Zeitreihe* bezeichnet man eine Folge von Zufallsvariablen

$$X(t, s) , \quad (311.1)$$

die im Wahrscheinlichkeitsraum (S, Z, P) eines Experiments mit der Menge S der Elementarereignisse s , der Menge Z der zufälligen Ereignisse und der Wahrscheinlichkeit P definiert sind. Der Index t ist Element der Indexmenge T und wird wie bei den deterministischen Signalen mit der Zeit identifiziert, er kann aber beispielsweise auch die Entfernung repräsentieren. Abgekürzt wird

$$X(t) \quad (311.2)$$

geschrieben, denn ein stochastischer Prozeß läßt sich als Folge vom Index t abhängiger Zufallsvariablen interpretieren.

Die Werte, die ein stochastischer Prozeß für ein bestimmtes Elementarereignis s in Abhängigkeit von t annimmt und für die im folgenden fast ausschließlich reelle Zahlen vorausgesetzt werden, bezeichnet man als eine *Realisierung* des Prozesses. Die Gesamtheit der Realisierungen für sämtliche Elementarereignisse nennt man *Ensemble*. Die Temperaturen in Abhängigkeit von der Zeit in einem Punkt eines Raumes ergeben beispielsweise die Realisierung eines stochastischen Prozesses. Ist das zugrunde liegende Experiment derart definiert, daß die Temperaturen verschiedener Punkte des Raumes betrachtet werden, dann bildet der Temperaturverlauf in allen Punkten das Ensemble des stochastischen Prozesses.

Die Elemente s der Menge S der Elementarereignisse können abzählbar oder überabzählbar sein. Für vorgegebenes t bildet dann $X(t)$ im ersten Fall eine diskrete Zufallsvariable und im zweiten Fall eine stetige Zufallsvariable (Koch 1987, S.96). Ebenso können die Elemente t der Indexmenge T mit $T = \mathbb{N}_0$ oder $T = \mathbb{Z}$ abzählbar oder mit $T = \mathbb{R}$ überabzählbar sein. Für vorgegebenes s liegt dann im ersten Fall der stochastische Prozeß nur an diskreten Zeitpunkten vor, während er im zweiten Fall kontinuierlich gegeben ist. Ein diskreter stochastischer Prozeß mit $T = \mathbb{Z}$ oder $T = \mathbb{N}_0$ wird in Übereinstimmung mit der Notation (211.3) für das diskrete Signal $x(n)$ mit

$$X(n) \quad \text{für} \quad n \in \mathbb{Z} = \{0, \pm 1, \pm 2, \dots\} \quad \text{oder} \quad n \in \mathbb{N}_0 = \{0, 1, 2, \dots\} \quad (311.3)$$

bezeichnet. Im folgenden werden überwiegend abzählbare Mengen S und T angenommen. Die diskreten stochastischen Prozesse kann man sich wieder, wie in den Kapiteln 215 und 232 für die deterministischen Signale erläutert, durch Abtastung der kontinuierlichen stochastischen Prozesse entstanden denken.

Gilt für die Indexmenge $T = \mathbb{I}$ oder $T = \mathbb{R}$, ergibt sich ein *eindimensionaler* stochastischer Prozeß oder Zufallsprozeß. Ist die Folge der Zufallsvariablen nicht von einem skalarwertigen Index t , sondern von einem n -dimensionalen Vektor \mathbf{t} abhängig, gilt also $T = \mathbb{R}^n$, dann folgt ein *n -dimensionaler Zufallsprozeß* oder *n -dimensionales Zufallsfeld*. Liegen keine eindimensionalen Zufallsvariablen, sondern mehrdimensionale Zufallsvariablen vor, ergeben sich *vektorielle Zufallsprozesse* oder *vektorielle Zufallsfelder*. Sie werden zusammen mit den Zufallsfeldern im Abschnitt 36 behandelt. Wegen der höheren Dimension der Zufallsvariablen bezeichnet man vektorielle Zufallsprozesse auch als *multivariate Zufallsprozesse* und vektorielle Zufallsfelder als *multivariate-multidimensionale Zufallsprozesse*. Diese Bezeichnungen werden im folgenden aber nicht verwendet.

Enthält die Indexmenge T abzählbar viele Elemente t_i , ergibt der stochastische Prozeß $X(t_i)$ abzählbar viele Zufallsvariablen. Die gemeinsame Verteilungsfunktion dieser Zufallsvariablen, die gemeinsame Dichte und die Momente können dann wie für mehrdimensionale Zufallsvariablen definiert werden. Sind dagegen die Elemente t der Menge T überabzählbar, führt der stochastische Prozeß $X(t)$ auf Zufallsvariablen unendlicher Dimension. Berücksichtigt man aber in dieser unendlichen Menge jeweils nur endlich viele Elemente $t_i \in T$, beispielsweise $i \in \{0, 1, \dots, n\}$, läßt sich wieder mit der Verteilungsfunktion einer $(n+1)$ -dimensionalen Zufallsvariablen arbeiten, die definiert ist durch

$$F(x_0, \dots, x_n; t_0, \dots, t_n) = P(X(t_0) < x_0, \dots, X(t_n) < x_n) \quad , \quad (311.4)$$

worin F die Verteilungsfunktion, P die Wahrscheinlichkeit des Ereignisses $X(t_0) < x_0, \dots, X(t_n) < x_n$ für beliebige x_0, \dots, x_n bedeuten (Koch 1987, S.102). Diese endlich dimensionale Verteilungsfunktion muß noch gewissen Symmetrie- und Konsistenzbedingungen genügen (Jaglom 1959, S.7; Lamperti 1977, S.250), um als Verteilungsfunktion eines stochastischen Prozesses $X(t)$ zu existieren.

Häufig wird die Verteilungsfunktion eines stochastischen Prozesses nicht bekannt sein, so daß mit ihrer Hilfe die statistischen Eigenschaften eines stochastischen Prozesses nicht zu beschreiben sind. Man muß sich dann mit den Momentfunktionen für das erste und zweite Moment der Zufallsvariablen des Prozesses begnügen.

312 Erwartungswert-, Autokovarianz- und Kreuzkovarianz-funktion

Mit der durch Differentiation der Verteilungsfunktion (311.4) folgenden Wahrscheinlichkeitsdichte, kurz Dichte genannt, und den Randdichten ergibt sich die von t abhängige *Erwartungswertfunktion* $\mu(t)$ des stochastischen Prozesses $X(t)$ als erstes Moment (Koch 1987, S.108) zu

$$\mu(t) = E(X(t)) = \int_{-\infty}^{\infty} xp(x;t)dx \quad , \quad (312.1)$$

worin x die Werte und $p(x;t)$ die Dichte von $X(t)$ bezeichnen. Die Momentfunktion $K(t_1, t_2)$ mit $t_1, t_2 \in T$ des zweiten Momentes des stochastischen Prozesses $X(t)$ folgt mit der Dichte $p(x_1, x_2; t_1, t_2)$ von $X(t_1)$ und $X(t_2)$ zu

$$K(t_1, t_2) = E(X(t_1)X(t_2)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1x_2p(x_1, x_2; t_1, t_2)dx_1dx_2 \quad . \quad (312.2)$$

Sie wird manchmal als *Autokorrelationsfunktion* bezeichnet. Für $t_1 = t_2 = t$ ergibt $K(t, t)$ die durchschnittliche *Leistung* des stochastischen Prozesses $X(t)$

$$E([X(t)]^2) = K(t, t) \quad . \quad (312.3)$$

Diese Bezeichnung folgt aus einer Analogie zur Elektrotechnik. Die elektrische Leistung P ergibt sich aus dem Produkt von Spannung U und Stromstärke I zu $P = UI$. Aufgrund des Ohmschen Gesetzes $I = U/R$, worin R den Widerstand bezeichnet, gilt weiter $P = I^2R = U^2/R$. Repräsentiert nun $X(t)$ die Stromstärke oder die Spannung, so ist (312.3) proportional zu einer durchschnittlichen Leistung.

Die *Autokovarianzfunktion* $C(t_1, t_2)$ ergibt sich mit

$$\begin{aligned} C(t_1, t_2) &= E([X(t_1) - \mu(t_1)][X(t_2) - \mu(t_2)]) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mu(t_1))(x_2 - \mu(t_2))p(x_1, x_2; t_1, t_2)dx_1dx_2 \end{aligned} \quad (312.4)$$

als zweites zentrales Moment. Gilt $t_1 = t_2 = t$, folgt mit $C(t, t)$ die *Varianzfunktion* $\sigma^2(t)$ des Prozesses, also

$$\sigma^2(t) = C(t, t) \quad . \quad (312.5)$$

In Übereinstimmung mit der Bezeichnungsweise in der Statistik wird nicht (312.2), sondern die mit

$$\rho(t_1, t_2) = C(t_1, t_2) / [C(t_1, t_1)C(t_2, t_2)]^{1/2} \quad (312.6)$$

normierte Autokovarianzfunktion $\rho(t_1, t_2)$ häufig *Autokorrelationsfunktion* genannt. Diese Bezeichnung soll im folgenden beibehalten werden. Für vorgegebenes t_1 und t_2 entspricht nämlich die Autokorrelationsfunktion dem Korrelationskoeffizienten in der Statistik. Es gilt daher (Koch 1987, S.112)

$$-1 \leq \rho(t_1, t_2) \leq 1 \quad . \quad (312.7)$$

Für zwei stochastische Prozesse $X(t)$ und $Y(t)$ ist die *Kreuzkovarianzfunktion* $C_{xy}(t_1, t_2)$ definiert durch

$$C_{xy}(t_1, t_2) = E([X(t_1) - \mu_x(t_1)][Y(t_2) - \mu_y(t_2)]) \quad (312.8)$$

mit

$$\mu_x(t) = E(X(t)) \quad \text{und} \quad \mu_y(t) = E(Y(t)) \quad .$$

Gilt $Y(t) = X(t)$ und $t_1 = t_2 = t$, ist die Kreuzkovarianzfunktion mit der Varianzfunktion $\sigma^2(t)$ aus (312.5) identisch.

Die Momentfunktionen eines diskreten stochastischen Prozesses $X(n)$ gemäß (311.3) sind ganz analog definiert. Man braucht lediglich t durch n und t_1, t_2 durch n_1, n_2 zu ersetzen und erhält

$$\mu(n), K(n_1, n_2), C(n_1, n_2), \sigma^2(n), \rho(n_1, n_2), C_{xy}(n_1, n_2) \quad . \quad (312.9)$$

Die Werte der Auto- oder Kreuzkovarianzfunktion für vorgegebenes t_1, t_2 oder n_1, n_2 bezeichnet man als *Autokovarianzen* oder *Kreuzkovarianzen*.

Beispiel 1: Ein stochastischer Prozeß sei durch eine harmonische Schwingung in der Darstellung (212.6) mit

$$X(t) = C(\cos\omega_0 t \cos\varphi + \sin\omega_0 t \sin\varphi) \quad , \quad t \in T$$

gegeben, in der die Amplitude C sowie die Kreisfrequenz ω_0 Konstanten und der Phasenwinkel φ eine Zufallsvariable darstelle, deren Dichte $p(\varphi)$ durch die Gleichverteilung

$$p(\varphi) = \begin{cases} 1/\pi & \text{für } 0 < \varphi < \pi \\ 0 & \text{sonst} \end{cases}$$

gegeben sei. Die Erwartungswertfunktion $\mu(t)$ von $X(t)$ berechnet sich nach (312.1) zu

$$\mu(t) = \int_0^\pi C(\cos\omega_0 t \cos\varphi + \sin\omega_0 t \sin\varphi) \frac{d\varphi}{\pi} = \frac{2C}{\pi} \sin\omega_0 t \quad .$$

Die Momentfunktion $K(t_1, t_2)$ des zweiten Momentes von $X(t)$ folgt nach (312.2) mit

$$K(t_1, t_2) = \int_0^\pi C^2(\cos\omega_0 t_1 \cos\varphi + \sin\omega_0 t_1 \sin\varphi)(\cos\omega_0 t_2 \cos\varphi + \sin\omega_0 t_2 \sin\varphi) \frac{d\varphi}{\pi}$$

$$\begin{aligned}
&= \frac{C^2}{\pi} \int_0^{\pi} (\cos\omega_0 t_1 \cos\omega_0 t_2 \cos^2\varphi + \sin\omega_0 t_1 \sin\omega_0 t_2 \sin^2\varphi \\
&\quad + (\cos\omega_0 t_1 \sin\omega_0 t_2 + \sin\omega_0 t_1 \cos\omega_0 t_2) \sin\varphi \cos\varphi) d\varphi \\
&= \frac{C^2}{2} \cos\omega_0(t_1 - t_2) .
\end{aligned}$$

Während die Erwartungswertfunktion $\mu(t)$ des vorgegebenen Prozesses von t abhängt, ist $K(t_1, t_2)$ lediglich von $t_1 - t_2$ abhängig. Δ

313 Stationarität

Durch die Stationarität lassen sich stochastische Prozesse charakterisieren, die sich in einem stabilen Zustand befinden und sich nicht mit der Zeit entwickeln. Man bezeichnet einen stochastischen Prozeß als *im engeren Sinne stationär*, wenn seine Verteilungsfunktion invariant gegenüber einer Verschiebung des Ursprungs von t ist. Mit (311.4) gilt also für beliebige $t_0, \dots, t_n \in T$ und $t_i + h \in T$ mit $i \in \{0, \dots, n\}$

$$F(x_0, \dots, x_n; t_0 + h, \dots, t_n + h) = F(x_0, \dots, x_n; t_0, \dots, t_n) . \quad (313.1)$$

Identifiziert man wieder mit t die Zeit, dann bedeutet (313.1), daß die statistischen Eigenschaften eines stationären Prozesses nicht mehr von den Zeitpunkten, sondern von Zeitdifferenzen abhängen.

Ein im engeren Sinne stationärer Prozeß $X(t)$ besitzt einen konstanten Erwartungswert μ

$$E(X(t)) = \mu \quad (313.2)$$

und eine nur von der Differenz τ abhängige Momentfunktion

$$K(t_1, t_2) = K(\tau) = E(X(t_1)X(t_1 + \tau)) \quad \text{mit} \quad \tau = t_2 - t_1 , \quad (313.3)$$

denn die aus (313.1) folgenden Dichten sind für alle Zeitpunkte $t \in T$ unverändert, so daß aus (312.1) die Beziehung (313.2) folgt. Weiter gilt für die Dichte $p(x_1, x_2; t_1, t_2)$ in (312.2) wegen (313.1)

$$p(x_1, x_2; t_1, t_2) = p(x_1, x_2; t_1 + h, t_2 + h) = p(x_1, x_2; \tau) \quad \text{mit} \quad \tau = t_2 - t_1 , \quad (313.4)$$

denn die Differenz $t_2 - t_1$ ist invariant gegenüber der Verschiebung des Ursprungs von t_1 und t_2 . Die Substitution von (313.4) in (312.2) ergibt (313.3).

Für einen *im weiteren Sinne stationären* oder *schwach stationären* Prozeß $X(t)$ gilt lediglich, daß seine Erwartungswertfunktion $E(X(t))$ eine Konstante μ ist und daß seine Momentfunktion $K(t_1, t_2)$ von der Differenz τ abhängt

$$E(X(t)) = \mu \quad \text{und} \quad K(t_1, t_2) = K(\tau) \quad \text{mit} \quad \tau = t_2 - t_1 . \quad (313.5)$$

Die Stationarität im weiteren Sinn ist also auf das erste und zweite Moment beschränkt. Ein im engeren Sinne stationärer Prozeß ist auch stationär im weiteren Sinne. Die Umkehrung gilt im allgemeinen nicht. Die Ausnahmen bilden die im Kapitel 321 behandelten Gauß-Prozesse, die die Normalverteilung besitzen, deren Parameter durch das erste und zweite Moment bestimmt sind (Koch 1987, S.135).

Da der Erwartungswert eines stationären Prozesses eine Konstante ist, gilt wegen (312.4), daß auch seine Autokovarianzfunktion nur von τ abhängt

$$C(t_1, t_2) = C(\tau) = E([X(t_1) - \mu][X(t_1 + \tau) - \mu]) \quad \text{mit} \quad \tau = t_2 - t_1 . \quad (313.6)$$

Seine Varianzfunktion (312.5) ist daher die konstante Varianz σ^2

$$\sigma^2(t) = C(0) = \sigma^2 . \quad (313.7)$$

Weiter folgt $C(t_1, t_2) = C(t_2, t_1)$ aus (312.4), so daß sich mit (313.6)

$$C(\tau) = C(-\tau) \quad (313.8)$$

ergibt. Schließlich erhält man aus (312.7)

$$|C(\tau)| \leq C(0) . \quad (313.9)$$

Die Autokovarianzfunktion eines stationären Prozesses ist also wegen (213.16) und (313.8) eine gerade und wegen (313.9) eine beschränkte Funktion.

Beispiel 1: Ein stochastischer Prozeß sei durch eine harmonische Schwingung in der Darstellung (212.6) mit

$$X(t) = A \cos \omega_0 t + B \sin \omega_0 t$$

gegeben, in der A und B voneinander unabhängige Zufallsvariablen darstellen, die gleichverteilt sind mit

$$p(A) = \begin{cases} 1/(2C) & \text{für } -C < A < C \\ 0 & \text{sonst} \end{cases}$$

und für B entsprechend. Die Erwartungswertfunktion $\mu(t)$ von $X(t)$ ergibt sich zu einer Konstanten, denn es gilt mit (312.1)

$$\begin{aligned} \mu(t) &= E(A \cos \omega_0 t) + E(B \sin \omega_0 t) \\ &= \int_{-C}^C (A \cos \omega_0 t \frac{dA}{2C} + B \sin \omega_0 t \frac{dB}{2C}) \\ &= 0 . \end{aligned}$$

Mit diesem Ergebnis folgt für die Autokovarianzfunktion $C(t_1, t_2)$ aus (312.4), da A und B voneinander unabhängig sind und ihre gemeinsame Verteilung aus dem Produkt ihrer Verteilungen sich ergibt (Koch 1987, S.106),

$$\begin{aligned}
 C(t_1, t_2) &= \int_{-C}^C \int_{-C}^C (A \cos \omega_0 t_1 + B \sin \omega_0 t_1) (A \cos \omega_0 t_2 + B \sin \omega_0 t_2) \frac{dA}{2C} \frac{dB}{2C} \\
 &= \frac{1}{4C^2} \int_{-C}^C \int_{-C}^C (A^2 \cos \omega_0 t_1 \cos \omega_0 t_2 + B^2 \sin \omega_0 t_1 \sin \omega_0 t_2 \\
 &\quad + AB (\cos \omega_0 t_1 \sin \omega_0 t_2 + \sin \omega_0 t_1 \cos \omega_0 t_2)) dA dB \\
 &= \frac{1}{4C^2} \int_{-C}^C \left(\frac{2}{3} C^3 \cos \omega_0 t_1 \cos \omega_0 t_2 + 2B^2 C \sin \omega_0 t_1 \sin \omega_0 t_2 \right) dB \\
 &= \frac{C^2}{3} (\cos \omega_0 t_1 \cos \omega_0 t_2 + \sin \omega_0 t_1 \sin \omega_0 t_2) = \frac{C^2}{3} \cos \omega_0 (t_2 - t_1) \\
 &= C(\tau) \quad \text{mit} \quad \tau = t_2 - t_1 .
 \end{aligned}$$

Damit ist der vorgegebene stochastische Prozeß im weiteren Sinne stationär. ▲

Die Kreuzkovarianzfunktion (312.8) der stationären Prozesse $X(t)$ und $Y(t)$ ergibt sich analog zu (313.6) als Funktion nur von τ

$$C_{xy}(t_1, t_2) = C_{xy}(\tau) \quad \text{mit} \quad \tau = t_2 - t_1 . \quad (313.10)$$

Wegen $C_{xy}(t_1, t_2) = C_{yx}(t_2, t_1)$ aus (312.8) gilt außerdem

$$C_{xy}(\tau) = C_{yx}(-\tau) . \quad (313.11)$$

Für einen diskreten Prozeß $X(n)$ gemäß (311.3), der stationär ist, erhält man mit (312.9) anstelle von (313.6) die nur von k abhängige Autokovarianzfunktion

$$\begin{aligned}
 C(n_1, n_2) &= C(k) = E\{[X(n_1) - \mu][X(n_1+k) - \mu]\} \\
 &\quad \text{mit} \quad k = n_2 - n_1 \quad \text{und} \quad k \in \{0, \pm 1, \pm 2, \dots\} .
 \end{aligned} \quad (313.12)$$

Sie besitzt anstelle von (313.8) und (313.9) die Eigenschaften

$$C(k) = C(-k) \quad \text{sowie} \quad |C(k)| \leq C(0) . \quad (313.13)$$

Für die beiden stationären diskreten Prozesse $X(n)$ und $Y(n)$ erhält man die Kreuzkovarianzfunktion zu

$$C_{xy}(n_1, n_2) = C_{xy}(k) \quad \text{mit} \quad k = n_2 - n_1 \quad \text{und} \quad k \in \{0, \pm 1, \pm 2, \dots\} , \quad (313.14)$$

für die nach (313.11) gilt

$$C_{xy}(k) = C_{yx}(-k) . \quad (313.15)$$

Für stationäre Prozesse gibt man häufig Autokovarianzfunktionen vor, oder man paßt eine geschätzte Autokovarianzfunktion - die Schätzung von Autokovarianzen wird im Kapitel 333 behandelt - einer vorgegebenen Funktion an. Im allgemeinen geht eine Autokovarianzfunktion $C(\tau)$ mit wachsendem τ gegen Null. Ihre Abklingeigenschaft wird mit Hilfe der *Halbwertsbreite* τ_0 angegeben, für die

$$C(\tau_0) = \sigma^2/2 \quad \text{mit} \quad C(0) = \sigma^2 \quad (313.16)$$

gilt, wobei σ^2 nach (313.7) die Varianz bedeutet. Zwei häufig verwendete Autokovarianzfunktionen sind in den folgenden beiden Beispielen angegeben, eine umfangreiche Zusammenstellung von Autokovarianzfunktionen befindet sich in (Meier und Keller 1990, S.190).

Beispiel 2: Die Autokovarianzfunktion vom *Exponentialtyp* eines stationären Prozesses ist durch

$$C(\tau) = \sigma^2 e^{-a|\tau|} \quad \text{mit} \quad a > 0 \quad (313.17)$$

gegeben, worin a den Parameter der Autokovarianzfunktion bezeichnet. Ihre Halbwertsbreite τ_0 berechnet sich wegen (313.16) zu

$$\tau_0 = \ln 2/a \quad . \quad (313.18)$$

▲

Beispiel 3: Die Autokovarianzfunktion nach Hirvonen (1962) ist durch

$$C(\tau) = \frac{\sigma^2}{1+(\tau/a)^2} \quad (313.19)$$

gegeben. Die Halbwertsbreite τ_0 ist durch den Parameter a bestimmt

$$\tau_0 = a \quad . \quad (313.20)$$

▲

314 Spektrum

Durch die Fourier-Transformation (214.2) wird das Spektrum eines deterministischen und aperiodischen Signals erhalten. Das Spektrum läßt sich dann mit Hilfe der Frequenzantwort eines linearen Systems nach (216.16) in einer Weise modifizieren, daß der Frequenzinhalt gewünschte Eigenschaften aufweist. Dieser Vorgang wird als *Filterung* bezeichnet.

Den Frequenzinhalt stochastischer Signale möchte man ebenfalls untersuchen und modi-

fizieren. Für die stochastischen Prozesse ergibt sich allerdings die Schwierigkeit, daß deren Fourier-Transformationen im allgemeinen nicht existieren. Dies liegt darin begründet, daß für die Ableitung der Fourier-Transformation (214.2) zeitbegrenzte Signale vorausgesetzt werden, die sich nicht bis ins Unendliche erstrecken. Die Signale müssen nach (214.6) absolut integrierbar sein. Diese Bedingung ist für einen stationären Prozeß, dessen Erwartungswert nach (313.2) eine Konstante ist, nicht erfüllt.

Dagegen ist die Autokovarianzfunktion $C(\tau)$ eines stationären und daher in einem stabilen Zustand sich befindlichen Prozesses im allgemeinen eine zeitbegrenzte Funktion, die aufgrund von (313.9) beschränkt ist. Ihre Fourier-Transformation läßt sich nach (214.2) berechnen, und man bezeichnet sie als *Leistungsspektrum*, *Spektraldichtefunktion* oder kurz *Spektrum* $S(\omega)$ eines stationären Prozesses

$$S(\omega) = \int_{-\infty}^{\infty} C(\tau) e^{-j\omega\tau} d\tau . \quad (314.1)$$

Das Spektrum repräsentiert, wie im folgenden gezeigt wird, den Frequenzinhalt eines stationären Prozesses.

Mit Hilfe der inversen Fourier-Transformation (214.2) kann die Autokovarianzfunktion $C(\tau)$ durch das Spektrum $S(\omega)$ dargestellt werden

$$C(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) e^{j\omega\tau} d\omega . \quad (314.2)$$

Diese Beziehung, deren Herleitung nachfolgend kurz angegeben wird, wird auch als Theorem von Wiener-Khintchine bezeichnet, das zuerst von Wiener (1930) und dann unabhängig von Khintchine (1934) abgeleitet wurde.

Für $\tau = 0$ erhält man mit (313.7) aus (314.2)

$$\sigma^2 = C(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) d\omega . \quad (314.3)$$

Die Varianz σ^2 eines stationären Prozesses ergibt sich also aus der Fläche $S(\omega)/2\pi$ des Spektrums, wobei $S(\omega)d\omega$ den Beitrag des sehr schmalen Frequenzbandes $d\omega$ zu der Varianz enthält. Für stationäre Prozesse $X(t)$ mit $E(X(t)) = 0$ folgt aus (312.3)

$$E([X(t)]^2) = C(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) d\omega , \quad (314.4)$$

so daß die Fläche $S(\omega)/2\pi$ die durchschnittliche Leistung des Prozesses angibt. Daher rührt der Name *Leistungsspektrum*, unter dem aber auch die Fourier-Transformation der Momentfunktion $K(\tau)$ nach (313.3) verstanden wird, so daß (314.4) ohne die Spezialisierung $E(X(t)) = 0$ gilt.

Definiert man das Spektrum als Fourier-Transformation der Autokorrelationsfunktion

$\rho(\tau)$ aus (312.6) und bezeichnet es mit $\bar{S}(\omega)$, dann folgt wegen (313.7) und wegen $\rho(0) = 1$ aus (314.4)

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \bar{S}(\omega) d\omega = 1 \quad \text{mit} \quad \bar{S}(\omega) = S(\omega)/\sigma^2 ,$$

so daß $\bar{S}(\omega)$ die Eigenschaft der Wahrscheinlichkeitsdichte einer Verteilung besitzt. Sie begründet die Bezeichnung Spektraldichtefunktion.

Um die Herleitung des Theorems von Wiener-Khintchine kurz anzudeuten, wird ein im weiteren Sinne stationärer Prozeß $X(t)$ mit dem Erwartungswert $E(X(t)) = 0$ wie ein deterministisches Signal durch eine Fourierreihe dargestellt. Mit (213.2) und (213.10) ergibt sich, falls $k\omega_0 = \omega_k$ gesetzt wird und falls in Übereinstimmung mit der inversen Fourier-Transformation (214.2) der Faktor $1/2\pi$ eingeführt wird,

$$X(t) = \frac{1}{2\pi} \sum_{k=-K}^K D_k e^{j\omega_k t} . \quad (314.5)$$

Hier sei $X(t)$ ein komplexwertiger stochastischer Prozeß, und die Amplituden D_k seien komplexe Zufallsvariablen mit $E(D_k) = 0$, für die Unabhängigkeit, also

$$E(D_k^* D_l) = 0 \quad \text{für} \quad k \neq l \quad (314.6)$$

gelte, wobei D_k^* die konjugiert komplexe Größe bedeutet. Die benutzte Definition des zweiten Moments komplexer Zufallsvariablen stellt eine sinnvolle Erweiterung der entsprechenden Definition für reelle Zufallsvariablen dar.

Mit $K \rightarrow \infty$ erhält man aus (314.5) eine spektrale Darstellung für $X(t)$, falls die Frequenzen $\omega_{-K}, \omega_{-K+1}, \dots, \omega_K$ sich gegenseitig ohne Grenzen annähern. Für den Grenzübergang definiert man den Zuwachs $D(\omega)$ der Amplituden D_k durch ihre Summe mit

$$D(\omega) = \sum_{\omega_k < \omega} D_k , \quad (314.7)$$

wobei wegen (314.6) gilt

$$E([D(\omega_1) - D(\omega_{1-1})]^* [D(\omega_m) - D(\omega_{m-1})]) = 0 \quad \text{für} \quad 1 < m - 1 . \quad (314.8)$$

Den komplexwertigen stochastischen Prozeß $D(\omega)$ bezeichnet man daher als *stochastischen Prozeß mit unabhängigen Zuwächsen*. Die spektrale Darstellung von $X(t)$ folgt nun aus (314.5) mit

$$X(t) = \lim_{K \rightarrow \infty} \left\{ \lim_{(\omega_k - \omega_{k-1}) \rightarrow 0} \frac{1}{2\pi} \sum_{k=-K}^K [D(\omega_k) - D(\omega_{k-1})] e^{j\omega_k t} \right\} . \quad (314.9)$$

Der Grenzwert ist durch das *Stieltjes-Integral* (Fichtenholz 1987, S.81)

$$X(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} dD(\omega) \quad (314.10)$$

bestimmt. Es stellt eine Verallgemeinerung der Fourier-Transformation (214.2) insofern dar, als der komplexwertige Prozeß $X(t)$ als Fourier-Transformation des Differentials $dD(\omega)$ des komplexwertigen Prozesses $D(\omega)$ mit unabhängigen Zuwächsen gefunden wird. Damit hat man die spektrale Darstellung eines stationären stochastischen Prozesses erhalten (Doob 1953, S.527; Jaglom 1959, S.29; Priestley 1981, S.246). Der differentielle Prozeß $dD(\omega)$ läßt sich im allgemeinen nicht angeben, doch wie mit (314.14) gezeigt wird, kann man $E([dD(\omega)]^2)$ mit Hilfe des Spektrums $S(\omega)$ ausdrücken.

Wird der stochastische Prozeß $X(t)$ einer im Kapitel 222 definierten Bandpaßfilterung mit einem sehr schmalen Frequenzband $d\omega = \omega_2 - \omega_1$ unterworfen, folgt für den gefilterten Prozeß $X(d\omega, t)$ aus (314.10)

$$X(d\omega, t) = \frac{1}{2\pi} e^{j\omega_2 t} dD(\omega) . \quad (314.11)$$

Der Beitrag $E([X(d\omega, t)]^2) = E(X(d\omega, t)X^*(d\omega, t))$ des Bandes $d\omega$ zur durchschnittlichen Leistung des gefilterten Prozesses berechnet sich hieraus zu

$$E([X(d\omega, t)]^2) = \frac{1}{4\pi^2} E([dD(\omega)]^2) . \quad (314.12)$$

Führt man als Spektrum $S(\omega)$ des Prozesses die Funktion ein, die die Dichte der Leistung in Abhängigkeit von der Frequenz ω darstellt, dann läßt sich der Beitrag (314.12) des Frequenzbandes $d\omega$ zur Leistung ausdrücken durch

$$E([X(d\omega, t)]^2) = \frac{1}{2\pi} S(\omega)d\omega \quad (314.13)$$

oder

$$E([dD(\omega)]^2) = 2\pi S(\omega)d\omega , \quad (314.14)$$

falls die Konstanten wie in (314.4) gewählt werden.

Mit Hilfe der Spektraldarstellung (314.9) wird jetzt nach (313.6) die Autokovarianzfunktion $C(\tau)$ berechnet. Man erhält mit (314.8) und (314.14)

$$\begin{aligned} C(\tau) &= E(X^*(t)X(t+\tau)) \\ &= \lim_{K \rightarrow \infty} \left\{ \lim_{(\omega_m - \omega_{m-1}) \rightarrow 0} \frac{1}{4\pi^2} \sum_{k=-K}^K \sum_{l=-K}^K e^{j(-\omega_k t + \omega_l(t+\tau))} \right. \\ &\quad \left. E([D(\omega_k) - D(\omega_{k-1})]^* [D(\omega_l) - D(\omega_{l-1})]) \right\} \end{aligned}$$

$$\begin{aligned}
 &= \lim_{K \rightarrow \infty} \left\{ \lim_{(\omega_m - \omega_{m-1}) \rightarrow 0} \frac{1}{2\pi} \sum_{k=-K}^K e^{j\omega_k \tau} S(\omega) d\omega \right\} \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) e^{j\omega \tau} d\omega
 \end{aligned}$$

für $m \in \{k, l\}$ und folglich (314.2), nämlich die Autokovarianzfunktion $C(\tau)$ als inverse Fourier-Transformation des Spektrums $S(\omega)$.

Da die Autokovarianzfunktion wegen (313.8) eine gerade Funktion ist, ist nach (214.7) ihre Fourier-Transformation reell, so daß in (314.1) sofort $S(\omega)$ anstelle von $S(j\omega)$ geschrieben wurde. Weiter folgt aus (214.7), daß $S(\omega)$ nach (213.16) eine gerade Funktion ist

$$S(\omega) = S(-\omega) \quad (314.15)$$

und daß gilt

$$S(\omega) = \int_{-\infty}^{\infty} C(\tau) \cos \omega \tau d\tau \quad \text{mit} \quad C(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) \cos \omega \tau d\omega . \quad (314.16)$$

Die zweite Beziehung in (314.16) läßt sich durch eine (214.7) entsprechende Zerlegung zeigen. Schließlich ist das Spektrum $S(\omega)$ nicht negativ

$$S(\omega) \geq 0 , \quad (314.17)$$

wie mit (341.6) gezeigt wird.

Das Spektrum $S(\omega)$ wurde in (314.1) für stationäre Prozesse eingeführt. Für nicht stationäre Prozesse ist es möglich, für die zeitabhängige Autokovarianzfunktion (312.4) ein zeitabhängiges Spektrum einzuführen, worauf hier aber nicht eingegangen wird.

Das *Kreuzspektrum* ist die Fourier-Transformation der Kreuzkovarianzfunktion $C_{xy}(\tau)$ der stationären Prozesse $X(t)$ und $Y(t)$

$$S_{xy}(j\omega) = \int_{-\infty}^{\infty} C_{xy}(\tau) e^{-j\omega \tau} d\tau . \quad (314.18)$$

Mit (313.11) folgt hieraus

$$S_{xy}(j\omega) = S_{yx}(-j\omega) = S_{yx}^*(j\omega) \quad (314.19)$$

Aus der inversen Fourier-Transformation ergibt sich

$$C_{xy}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{xy}(j\omega) e^{j\omega \tau} d\omega . \quad (314.20)$$

Um das Spektrum eines diskreten stationären Prozesses $X(n)$ zu erhalten, ist zu beachten, daß auch die Autokovarianzfunktion $C(k)$ aus (313.12) nur an diskreten Stellen vor-

liegt. Nimmt man wie im Kapitel 215 an, daß mit

$$C(k) = C(\tau = k\Delta\tau)$$

die diskrete Autokovarianzfunktion $C(k)$ durch Abtastung der kontinuierlichen Autokovarianzfunktion $C(\tau)$ in (313.6) entsteht, wobei $\Delta\tau$ das Abtastintervall bezeichnet, dann ergibt sich das Spektrum $S(\omega)$ von $X(n)$ mit (215.1) anstelle von (314.1) zu

$$S(\omega) = \sum_{k=-\infty}^{\infty} C(k)e^{-jk\Delta\tau\omega} \quad (314.21)$$

Die diskrete Autokovarianzfunktion $C(k)$ folgt durch die inverse Fourier-Transformation (215.2) anstelle von (314.2) mit

$$C(k) = \frac{\Delta\tau}{2\pi} \int_{-\pi/\Delta\tau}^{\pi/\Delta\tau} S(\omega)e^{jk\Delta\tau\omega} d\omega \quad (314.22)$$

Aus den Überlegungen im Zusammenhang mit (215.2) folgt, daß das Spektrum $S(\omega)$ periodisch ist mit der Periode $2\pi/\Delta\tau$. Aus den Frequenzbändern $d\omega$ des Intervalls $-\pi/\Delta\tau \leq \omega \leq \pi/\Delta\tau$ berechnen sich die Beiträge von $S(\omega)d\omega$ zur Varianz σ^2 des diskreten stationären Prozesses

$$\sigma^2 = C(0) = \frac{\Delta\tau}{2\pi} \int_{-\pi/\Delta\tau}^{\pi/\Delta\tau} S(\omega)d\omega \quad (314.23)$$

Das Spektrum $S(\omega)$ ist wieder reell und eine gerade Funktion und (314.16) entsprechend gilt

$$S(\omega) = \sum_{k=-\infty}^{\infty} C(k)\cos k\Delta\tau\omega \quad \text{mit} \quad C(k) = \frac{\Delta\tau}{2\pi} \int_{-\pi/\Delta\tau}^{\pi/\Delta\tau} S(\omega)\cos k\Delta\tau\omega d\omega \quad (314.24)$$

Führt man die Variablentransformation $\Omega = \Delta\tau\omega$ ein, ergibt sich anstelle von (314.21) und (314.22)

$$S(\Omega) = \sum_{k=-\infty}^{\infty} C(k)e^{-jk\Omega} \quad \text{und} \quad C(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\Omega)e^{jk\Omega} d\Omega \quad (314.25)$$

Das Spektrum $S(\Omega)$ ist jetzt mit 2π periodisch, und es wird $-\pi \leq \Omega \leq \pi$ gesetzt. Damit entspricht Ω der für (216.9) eingeführten normierten Frequenz. Substituiert man schließlich noch $2\pi\lambda = \Omega$, ergibt sich mit

$$S(\lambda) = \sum_{k=-\infty}^{\infty} C(k)e^{-2j\pi k\lambda} \quad \text{und} \quad C(k) = \int_{-0,5}^{0,5} S(\lambda)e^{2j\pi k\lambda} d\lambda \quad (314.26)$$

das Spektrum $S(\lambda)$ mit der Periode 1, und es wird $-0,5 \leq \lambda \leq 0,5$ gesetzt.

Anstelle von (314.18) lautet das diskrete Kreuzspektrum

$$S_{xy}(j\Omega) = \sum_{k=-\infty}^{\infty} C_{xy}(k) e^{-jk\Omega} \quad \text{mit} \quad C_{xy}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xy}(j\Omega) e^{jk\Omega} d\Omega \quad (314.27)$$

und

$$S_{xy}(j\Omega) = S_{yx}(-j\Omega) = S_{yx}^*(j\Omega) . \quad (314.28)$$

Beispiel 1: Die Autokovarianzfunktion $C(\tau)$ eines stationären Prozesses sei mit der Konstanten σ^2 durch die Deltafunktion $\delta(\tau)$ mit

$$C(\tau) = \sigma^2 \delta(\tau) \quad (314.29)$$

bestimmt. Mit (214.16) und (314.1) folgt dann das Spektrum $S(\omega)$ zu

$$S(\omega) = \sigma^2 . \quad (314.30)$$

▲

Beispiel 2: Die Autokovarianzfunktion $C(\tau)$ eines stationären Prozesses sei vom Exponentialtyp (313.17)

$$C(\tau) = \sigma^2 e^{-a|\tau|} . \quad (314.31)$$

Zur Berechnung ihrer Fourier-Transformation (314.1) wird das Integral

$$\int_0^{\infty} e^{-a\tau} e^{-j\omega\tau} d\tau = - \frac{1}{a+j\omega} e^{-(a+j\omega)\tau} \Big|_0^{\infty} = \frac{1}{a+j\omega} = \frac{a}{a^2+\omega^2} - j \frac{\omega}{a^2+\omega^2} \quad (314.32)$$

ausgewertet. Da $C(\tau)$ eine gerade Funktion ist, berechnet sich ihre Fourier-Transformation nach (214.7) aus dem Realteil des Integrals (314.32), multipliziert mit dem Faktor 2, so daß für das Spektrum gilt

$$S(\omega) = \sigma^2 \frac{2a}{a^2+\omega^2} . \quad (314.33)$$

▲

Beispiel 3: Die Autokovarianzfunktion $C(\tau)$ eines stationären Prozesses sei durch (313.19) mit

$$C(\tau) = \frac{\sigma^2}{1+(\tau/a)^2}$$

gegeben. Das Spektrum $S(\omega)$ des Prozesses erhält man aus (314.1) zu

$$S(\omega) = \sigma^2 \int_{-\infty}^{\infty} \frac{a^2}{a^2+\tau^2} e^{-j\omega\tau} d\tau . \quad (314.34)$$

Um dieses Integral zu lösen, wird nach (314.2) die inverse Fourier-Transformation des Spektrums (314.33) gebildet, die die Autokovarianzfunktion (314.31) ergibt, folglich

$$\sigma^2 e^{-a|\tau|} = \frac{\sigma^2}{2\pi} \int_{-\infty}^{\infty} \frac{2a}{a^2 + \omega^2} e^{j\omega\tau} d\omega .$$

Durch Transformation der Variablen mit $\omega = -\varepsilon$ und $\tau = \omega$ folgt das Spektrum $S(\omega)$ durch Vergleich mit (314.34) zu

$$S(\omega) = a\pi\sigma^2 e^{-a|\omega|} . \quad (314.35)$$

Δ

Beispiele für Spektren diskreter stochastischer Prozesse befinden sich in den Kapiteln 325, 327, 377 und 378.

32 Spezielle Zufallsprozesse

321 Weißes Rauschen, Gauß- und Poisson-Prozesse

Im folgenden werden spezielle und besonders einfach zu beschreibende Zufallsprozesse vorgestellt.

Einen stochastischen Prozeß $X(t)$ bezeichnet man als *weißes Rauschen*, wenn die Zufallsvariable $X(t_1)$ unabhängig ist von $X(t_2)$ für $t_1 \neq t_2$, so daß für die Autokovarianzfunktion

$$C(t_1, t_2) = 0 \quad \text{für} \quad t_1 \neq t_2 \quad (321.1)$$

folgt (Koch 1987, S.111). Stationäres weißes Rauschen wird mit der Konstanten σ_w^2 und der Deltafunktion $\delta(\tau)$ durch die Autokovarianzfunktion

$$C(\tau) = \sigma_w^2 \delta(\tau) \quad \text{mit} \quad \tau = t_2 - t_1 \quad (321.2)$$

charakterisiert. Nach (314.30) ist dann sein Spektrum $S(\omega)$ durch die Konstante

$$S(\omega) = \sigma_w^2 \quad (321.3)$$

gegeben. In Analogie zum weißen Licht sind alle Frequenzen im Spektrum vertreten, daher der Name "weißes Rauschen". Da das Spektrum des stationären weißen Rauschens konstant über sämtliche Frequenzen ω ist, ist nach (314.4) die durchschnittliche Leistung und nach (314.3) die Varianz des weißen Rauschens unendlich groß.

Einen diskreten Prozeß $X(n)$ bezeichnet man als weißes Rauschen, wenn die Zufallsvariable $X(n_1)$ unabhängig von $X(n_2)$ ist für $n_1 \neq n_2$. Mit der Varianz $V(n_1) = \sigma^2(n_1)$ von $X(n_1)$ folgt dann mit dem Einheitsimpuls $d(n)$ aus (211.7) die Autokovarianzfunktion

$$C(n_1, n_2) = \sigma^2(n_1) d(n_2 - n_1) \quad (321.4)$$

Ist der diskrete Prozeß $X(n)$ stationär, ist nach (313.7) die Varianzfunktion $\sigma^2(n_1) = \sigma^2$ eine Konstante, und es ergibt sich mit (313.12)

$$C(k) = \sigma^2 d(k) \quad \text{mit} \quad k = n_2 - n_1 \quad (321.5)$$

Sein Spektrum $S(\omega)$ berechnet sich mit (212.9) nach (314.21)

$$S(\omega) = \sum_{k=-\infty}^{\infty} C(k) e^{-jk\Delta\tau\omega} = \sigma^2 \quad (321.6)$$

zu der Konstanten σ^2 .

Ein stochastischen Prozeß $X(t)$ bezeichnet man als *Gauß-Prozeß* oder *Normalprozeß*, wenn für beliebige Zeitpunkte t mit $t \in \{t_0, t_1, \dots, t_n\}$ die entsprechenden Zufallsvariablen, die im $(n+1) \times 1$ Vektor \mathbf{x} zusammengefaßt seien, also $\mathbf{x} = (X(t_i))$, die multivariate Normalverteilung mit den Parametern $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ besitzen

$$\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad . \quad (321.7)$$

Gauß-Prozesse haben Eigenschaften, die ihre Anwendung erleichtern. Zum einen folgt aus (321.7), daß $E(\mathbf{x}) = \boldsymbol{\mu}$ und $D(\mathbf{x}) = \boldsymbol{\Sigma}$ gelten (Koch 1987, S.135). Man erhält daher aus (312.1) für die Erwartungswertfunktion des Gauß-Prozesses

$$E(X(t_i)) = \mu_i \quad \text{mit} \quad \boldsymbol{\mu} = (\mu_i) \quad (321.8)$$

und mit (312.4) für die Autokovarianzfunktion

$$C(t_i, t_j) = \sigma_{ij} \quad \text{mit} \quad \boldsymbol{\Sigma} = (\sigma_{ij}) \quad . \quad (321.9)$$

Durch das erste und das zweite zentrale Moment ist die Normalverteilung (321.7) eindeutig bestimmt. Gauß-Prozesse, die im weiteren Sinne stationär sind, sind daher auch im engeren Sinne stationär. Außerdem ist eine lineare Funktion normalverteilter Zufallsvariablen wieder normalverteilt (Koch 1987, S.140).

Ein stochastischer Prozeß heißt *Poisson-Prozeß*, falls

$$X(t) = 0 \quad \text{für} \quad t \leq 0$$

und die Poisson-Verteilung mit der Dichte $f(x)$ für $X(t)$ mit $t > 0$ gilt

$$f(x) = e^{-\lambda t} (\lambda t)^x / x! \quad . \quad (321.10)$$

Die Poisson-Verteilung gibt die Wahrscheinlichkeit des Eintreffens von x Punkten in einem Zeitintervall der Länge t an, wobei das Eintreffen von Punkten in sich nicht überlappenden Zeitintervallen voneinander unabhängig sind. Bedeutet x beispielsweise die Anzahl von Fahrzeugen, die in einem Zeitintervall t an einem Ort eintreffen, so ist λ die durchschnittlich zu erwartende Anzahl der ankommenden Fahrzeuge pro Zeiteinheit, in der t definiert ist. Es gilt nämlich mit (312.1) (Koch 1987, S.130)

$$E(X(t)) = \lambda t \quad \text{für} \quad t > 0 \quad . \quad (321.11)$$

Das zweite Moment $E([X(t)]^2)$, das sich auf die gleiche Art wie das erste Moment berechnen läßt (Koch 1987, S.101), ergibt sich zu

$$E([X(t)]^2) = \lambda t (\lambda t + 1) \quad . \quad (321.12)$$

Um die Momentenfunktion $K(t_1, t_2)$ nach (312.2) zu berechnen, wird das Intervall $0 \leq t_1 < t_2$ in die beiden sich nicht überdeckenden Teilintervalle $0 \leq t < t_1$ und $t_1 \leq t < t_2$ zerlegt

$$\begin{aligned} K(t_1, t_2) &= E(X(t_1)X(t_2)) = E(X(t_1)[X(t_2)-X(t_1)+X(t_1)]) \\ &= E(X(t_1)[X(t_2)-X(t_1)]+[X(t_1)]^2) . \end{aligned}$$

Da nach Voraussetzung die Ereignisse in den beiden Teilintervallen voneinander unabhängig sind, ergibt sich mit (321.11) und (321.12)

$$\begin{aligned} K(t_1, t_2) &= E(X(t_1))E(X(t_2)-X(t_1))+E([X(t_1)]^2) \\ &= \lambda t_1 \lambda (t_2-t_1) + \lambda t_1 (\lambda t_1 + 1) = \lambda t_1 (1 + \lambda t_2) \quad \text{für } t_2 > t_1 . \quad (321.13) \end{aligned}$$

Für $t_1 > t_2$ sind t_1 und t_2 in (321.13) zu vertauschen.

Da nach Voraussetzung die Ereignisse in sich nicht überdeckenden Teilintervallen voneinander unabhängig sind, gilt mit den nach der Größe geordneten Elementen t_0, t_1, \dots, t_n der Indexmenge T , daß die durch die folgenden Differenzen gebildeten Zufallsvariablen des Poisson-Prozesses

$$X(t_1)-X(t_0), X(t_2)-X(t_1), \dots, X(t_n)-X(t_{n-1}) \quad (321.14)$$

voneinander unabhängig sind. Man erhält also einen stochastischen Prozeß mit unabhängigen Zuwächsen, der im Zusammenhang mit (314.7) definiert wurde.

Beispiel 1: Personen vor den Schaltern einer Bank oder Programme, die auf die Bearbeitung in einem Rechner warten, bilden eine Warteschlange. Zur Entwicklung von statistischen Modellen für die Warteschlangen muß der stochastische Prozeß der Anzahl der Ankünfte bis zum Zeitpunkt t definiert werden. In der Regel nimmt man für diesen Prozeß einen Poisson-Prozeß an (Heller u.a. 1978, S.179; Mathar und Pfeifer 1990, S.183).

A

322 Markoff-Prozesse und Markoff-Ketten

Es gibt Phänomene, deren Verhalten dadurch charakterisiert wird, daß die zukünftige Entwicklung nur von der Gegenwart und nicht von der Vergangenheit bestimmt ist. Solche Phänomene werden durch Markoff-Prozesse oder Markoff-Ketten beschrieben. Triff beispielsweise in einer Warteschlange zu jeder Zeiteinheit mit einer gewissen Wahrscheinlichkeit ein Kunde ein und wird mit einer weiteren Wahrscheinlichkeit ein Kunde bedient, so ist die Länge der Warteschlange zur Zeiteinheit n nur abhängig von der Länge der Schlange zur vorangegangenen Zeiteinheit $n-1$, nicht aber von der Länge der weiter zurückliegenden Zeiteinheiten.

Einen stochastischen Prozeß $X(t)$ bezeichnet man als *Markoff-Prozeß*, falls für die

Wahrscheinlichkeit des Ereignisses $X(t_n) < x_n$ für beliebiges x_n unter der Bedingung, daß die Zufallsvariablen $X(t_0)$ bis $X(t_{n-1})$ zu den vorangegangenen Zeiten die Werte x_0 bis x_{n-1} annehmen, gilt

$$\begin{aligned} P(X(t_n) < x_n | X(t_0) = x_0, X(t_1) = x_1, \dots, X(t_{n-1}) = x_{n-1}) \\ = P(X(t_n) < x_n | X(t_{n-1}) = x_{n-1}) . \end{aligned} \quad (322.1)$$

Ausgehend von einem Zustand zur Zeit t_{n-1} hängt also die weitere Entwicklung des stochastischen Prozesses zur Zeit t_n nur vom Zustand zur Zeit t_{n-1} ab und nicht von der vorangegangenen Entwicklung.

Durch (322.1) wird ein *Markoff-Prozeß erster Ordnung* definiert. Für einen *Markoff-Prozeß der Ordnung p* gilt

$$\begin{aligned} P(X(t_n) < x_n | X(t_0) = x_0, X(t_1) = x_1, \dots, X(t_{n-1}) = x_{n-1}) \\ = P(X(t_n) < x_n | X(t_{n-1}) = x_{n-1}, \dots, X(t_{n-p}) = x_{n-p}) . \end{aligned} \quad (322.2)$$

Mit den für die Verteilungsfunktion (311.4) getroffenen Voraussetzungen läßt sich die Markoff-Eigenschaft (322.1) durch die bedingte Verteilungsfunktion ausdrücken

$$F(x_n | x_0, x_1, \dots, x_{n-1}) = F(x_n | x_{n-1}) , \quad (322.3)$$

so daß, falls sie nach x_n differenzierbar ist, für die bedingte Dichte des Markoff-Prozesses folgt

$$p(x_n | x_0, x_1, \dots, x_{n-1}) = p(x_n | x_{n-1}) . \quad (322.4)$$

Genau ebenso wie (322.1) charakterisiert diese Beziehung die Markoff-Eigenschaft.

Mit Hilfe der Definition der bedingten Dichte (Koch 1987, S.105)

$$p(x_n | x_0, x_1, \dots, x_{n-1}) = \frac{p(x_0, x_1, \dots, x_n)}{p(x_0, x_1, \dots, x_{n-1})} \quad (322.5)$$

erhält man anstelle von (322.4)

$$p(x_n | x_{n-1}) p(x_0, x_1, \dots, x_{n-1}) = p(x_0, x_1, \dots, x_n) . \quad (322.6)$$

Ansprechend gilt auch

$$p(x_{n-1} | x_{n-2}) p(x_0, x_1, \dots, x_{n-2}) = p(x_0, x_1, \dots, x_{n-1}) .$$

Setzt man dieses Ergebnis rekursiv in (322.6), folgt schließlich die gemeinsame Dichte der Zufallsvariablen $X(t_0)$ bis $X(t_n)$ aus dem Produkt der bedingten Dichten

$$p(x_0, x_1, \dots, x_n) = p(x_n | x_{n-1}) p(x_{n-1} | x_{n-2}) \dots p(x_1 | x_0) p(x_0) . \quad (322.7)$$

Auch für den Markoff-Prozeß der Ordnung p gilt eine entsprechende Beziehung. Man erhält sie mit (325.40) aus (325.47) unter Berücksichtigung der im Zusammenhang mit

(325.24) erläuterten Notation.

Die Randdichte $p(x_n)$ von $X(t_n)$ ergibt sich aufgrund der Definition einer Randdichte (Koch 1987, S.104) und der Definition (322.5) der bedingten Dichte zu

$$\begin{aligned} p(x_n) &= \int_{-\infty}^{\infty} p(x_{n-1}, x_n) dx_{n-1} \\ &= \int_{-\infty}^{\infty} p(x_n | x_{n-1}) p(x_{n-1}) dx_{n-1} . \end{aligned} \quad (322.8)$$

Entsprechend erhält man auch $p(x_{n-1})$ und weiter $p(x_{n-2})$, so daß durch wiederholte Substitution oder unmittelbar aus (322.7)

$$p(x_n) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(x_n | x_{n-1}) p(x_{n-1} | x_{n-2}) \dots p(x_1 | x_0) p(x_0) dx_{n-1} dx_{n-2} \dots dx_0 \quad (322.9)$$

folgt. Sowohl die gemeinsame Dichte (322.7) als auch die Randdichte (322.9) erhält man also aus der Dichte $p(x_0)$ von $X(t_0)$ und den bedingten Dichten $p(x_i | x_{i-1})$.

Ein Markoff-Prozeß besitzt die Markoff-Eigenschaft (322.4) auch in umgekehrter Richtung. Die Wahrscheinlichkeit eines Ereignisses zur Zeit t_n unter der Bedingung, daß alle zukünftigen Ereignisse bis zur Zeit t_{n+p} gegeben sind, hängt nur vom Ereignis zur Zeit t_{n+1} ab

$$p(x_n | x_{n+1}, x_{n+2}, \dots, x_{n+p}) = p(x_n | x_{n+1}) . \quad (322.10)$$

Wendet man nämlich auf der linken Seite von (322.10) die Definition (322.5) der bedingten Dichte an, erhält man mit (322.7)

$$\frac{p(x_n, x_{n+1}, \dots, x_{n+p})}{p(x_{n+1}, \dots, x_{n+p})} = \frac{p(x_{n+1} | x_n) p(x_n)}{p(x_{n+1})} .$$

Aufgrund des Bayes-Theorems (383.2) gilt aber für die rechte Seite

$$\frac{p(x_{n+1} | x_n) p(x_n)}{p(x_{n+1})} = p(x_n | x_{n+1}) , \quad (322.11)$$

so daß (322.10) folgt.

Die Markoff-Eigenschaft (322.4) kann man nicht nur in umgekehrter Richtung nachweisen, sie gilt auch symmetrisch

$$p(x_n | x_0, x_1, \dots, x_{n-1}, x_{n+1}, x_{n+2}, \dots, x_m) = p(x_n | x_{n-1}, x_{n+1}) . \quad (322.12)$$

Aufgrund dieser Eigenschaft lassen sich die Markoff-Prozesse auf Zufallsfelder verallgemeinern, für die Nachbarschaften definiert sind. Hierauf wird im Kapitel 372 eingegangen. Um die Richtigkeit von (322.12) zu zeigen, wird wie beim Beweis von (322.10) vorgegangen. Die linke Seite wird durch (322.5) ausgedrückt und dann (322.7) angewen-

det. Man erhält

$$\frac{p(x_0, x_1, \dots, x_n)}{p(x_0, \dots, x_{n-1}, x_{n+1}, \dots, x_m)} = \frac{p(x_{n+1}|x_n)p(x_n|x_{n-1})}{p(x_{n+1}|x_{n-1})} . \quad (322.13)$$

Weiter ergibt sich zuerst mit (322.7) und dann mit (322.5) für

$$\frac{p(x_{n+1}|x_n)p(x_n|x_{n-1})p(x_{n-1})}{p(x_{n-1})} = \frac{p(x_{n+1}, x_n, x_{n-1})}{p(x_{n-1})} = p(x_{n+1}, x_n|x_{n-1}) .$$

Dieses Ergebnis auf der rechten Seite von (322.13) substituiert, führt schließlich auf (Koch 1990, S.49)

$$\frac{p(x_{n+1}, x_n|x_{n-1})}{p(x_{n+1}|x_{n-1})} = p(x_n|x_{n-1}, x_{n+1})$$

und damit auf (322.12).

In einem Markoff-Prozeß ist die Zufallsvariable $X(t_{n-1})$ unabhängig von der Zufallsvariablen $X(t_{n+1})$, falls der Wert der Zufallsvariablen $X(t_n)$ gegeben ist, denn es gilt

$$p(x_{n+1}, x_{n-1}|x_n) = p(x_{n+1}|x_n)p(x_{n-1}|x_n) . \quad (322.14)$$

Mit der Definition (322.5) der bedingten Dichte und mit (322.7) kann man nämlich die linke Seite von (322.14) umformen

$$p(x_{n+1}, x_{n-1}|x_n) = \frac{p(x_{n-1}, x_n, x_{n+1})}{p(x_n)} = \frac{p(x_{n+1}|x_n)p(x_n|x_{n-1})p(x_{n-1})}{p(x_n)} ,$$

so daß (322.14) mit (322.11) erhalten wird.

Ein stochastischer Prozeß mit unabhängigen Zuwächsen gemäß (321.14), der mit einer Wahrscheinlichkeit von Eins mit $X(t_0) = x_0$ beginnt, ist ein Markoff-Prozeß. Die Zufallsvariable $X(t_n)$ dieses Prozesses läßt sich nämlich mit

$$X(t_n) = x_0 + [X(t_1) - X(t_0)] + [X(t_2) - X(t_1)] + \dots + [X(t_n) - X(t_{n-1})]$$

als Summe unabhängiger Zufallsvariablen $[X(t_i) - X(t_{i-1})]$ darstellen. Die bedingte Verteilung von $X(t_n)$ ist daher nur vom Wert für $X(t_{n-1})$ abhängig, so daß nach (322.4) ein Markoff-Prozeß vorliegt. Ein Poisson-Prozeß ist daher ein Markoff-Prozeß, da der Poisson-Prozeß, wie im Zusammenhang mit (321.14) erwähnt, ein Prozeß mit unabhängigen Zuwächsen darstellt.

Einen diskreten Markoff-Prozeß mit der Indexmenge $T = \mathbb{N}_0$ und einer abzählbaren Menge S von Elementarereignissen bezeichnet man als *Markoff-Kette*. Die Zufallsvariablen $X(n)$ der Markoff-Kette mit $n \in \mathbb{N}_0$ sind also diskrete Zufallsvariable. Anstelle der Markoff-Eigenschaft (322.1) definiert man für die Markoff-Kette

$$P(X(n) = x_n | X(0) = x_0, X(1) = x_1, \dots, X(n-1) = x_{n-1})$$

$$= P(X(n) = x_n | X(n-1) = x_{n-1}) . \quad (322.15)$$

Die bedingte Wahrscheinlichkeit $P(X(n) = x_j | X(n-1) = x_i)$ im Schritt n vom Wert x_i zum Wert x_j bezeichnet man als *Übergangswahrscheinlichkeit*. Abgekürzt soll sie mit

$$p_{ij}(n) = P(X(n) = x_j | X(n-1) = x_i) \quad (322.16)$$

bezeichnet werden. Da bei diskreten Zufallsvariablen die Wahrscheinlichkeit, daß die Zufallsvariable einen bestimmten Wert annimmt, gleich der Dichte ist, bezeichnet $p_{ij}(n)$ auch die Dichte der Zufallsvariablen $X(n)$ mit den Werten x_j unter der Bedingung, daß $X(n-1) = x_i$ gesetzt wird. Es gilt daher (Koch 1987, S.96)

$$\sum_j p_{ij}(n) = 1 , \quad (322.17)$$

wobei über sämtliche Werte x_j zu summieren ist, die $X(n)$ annehmen kann.

Die bedingte Dichte auf der rechten Seite der Markoff-Eigenschaft (322.4) ist für Markoff-Ketten durch die Übergangswahrscheinlichkeit (322.16) bestimmt, also

$$p_{ij}(n) = p(x_n | x_{n-1}) . \quad (322.18)$$

Definiert man weiter die Randdichte $p_i(n)$ der Zufallsvariablen $X(n)$ mit

$$p_i(n) = P(X(n) = x_i) , \quad (322.19)$$

so entspricht dieser Dichte die Dichte $p(x_n)$ in (322.9), also

$$p_i(n) = p(x_n) . \quad (322.20)$$

Aus (322.7) und (322.9) folgt nun, daß für Markoff-Ketten die gemeinsame Dichte der Zufallsvariablen $X(0)$ bis $X(n)$ und die Randdichte der Zufallsvariablen $X(n)$ aus der Anfangsdichte $p_i(0)$ der Kette und den Übergangswahrscheinlichkeiten $p_{ij}(k)$ folgen. Für die gemeinsame Dichte ergibt sich aus (322.7)

$$\begin{aligned} P(X(0) = x_0, X(1) = x_1, \dots, X(n) = x_n) \\ = p_{n-1, n}(n) \dots p_{0, 1}(1) p_0(0) . \end{aligned} \quad (322.21)$$

Zur Berechnung der Randdichte muß anstelle der Integration die Summation treten. Man erhält aus (322.8)

$$\begin{aligned} p_n(n) = P(X(n) = x_n) &= \sum_i P(X(n) = x_n, X(n-1) = x_{(n-1)_i}) \\ &= \sum_i P(X(n) = x_n | X(n-1) = x_{(n-1)_i}) P(X(n-1) = x_{(n-1)_i}) \\ &= \sum_i p_{(n-1)_i, n}(n) p_{(n-1)_i}(n-1) . \end{aligned} \quad (322.22)$$

Führt man mit

$$\mathbf{P}(n) = (p_{ij}(n)) \quad (322.23)$$

die Matrix $\mathbf{P}(n)$ der Übergangswahrscheinlichkeiten $p_{ij}(n)$ ein und mit

$$\mathbf{p}(n) = (p_i(n))$$

den Vektor der Randdichten, erhält man anstelle von (322.22) die Randdichte $p_n(n)$ zu

$$p_n(n) = (\mathbf{p}'(n-1)\mathbf{P}(n))_n$$

oder den gesamten transponierten Vektor $\mathbf{p}'(n)$

$$\mathbf{p}'(n) = \mathbf{p}'(n-1)\mathbf{P}(n) .$$

Entsprechend folgt auch $\mathbf{p}'(n-1)$ und $\mathbf{p}'(n-2)$ und so fort, so daß durch wiederholte Substitution

$$\mathbf{p}'(n) = \mathbf{p}'(0)\mathbf{P}(1)\mathbf{P}(2)\dots\mathbf{P}(n) \quad (322.24)$$

erhalten wird. Dieser Beziehung entspricht (322.9).

Beispiel 1: Im folgenden sollen die Übergangswahrscheinlichkeiten für eine Irrfahrt auf einem eindimensionalen Gitter mit absorbierenden Grenzen angegeben werden. Zu einem Zeitpunkt n mit $n \in \mathbb{N}$ befindet sich ein Gegenstand an der Position k mit $k \in \{0, 1, \dots, s\}$ eines Gitters. Zu jedem Zeitpunkt bewegt er sich um eine Position e_n in positiver Richtung mit $e_n = 1$ oder in negativer Richtung mit $e_n = -1$ auf dem Gitter mit den Wahrscheinlichkeiten

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

Erreicht er die beiden Positionen $k = 0$ oder $k = s$, verläßt er sie nicht mehr.

Es sei $X(n)$ der diskrete stochastische Prozeß, der durch die Position k zur Zeit n bestimmt wird, so daß

$$X(n) = k \quad \text{mit} \quad k \in \{0, 1, \dots, s\}$$

gilt. Weiter erhält man

$$X(n) = \begin{cases} X(n-1)+e_n & \text{für } 1 \leq X(n) \leq s-1 \\ X(n-1) & \text{für } X(n) \in \{0, s\} . \end{cases}$$

Für $X(n)$ ist (322.15) erfüllt, so daß eine Markoff-Kette vorliegt. Die mit (322.23) definierte Matrix der Übergangswahrscheinlichkeiten $p_{ij}(n)$ mit $i, j \in \{0, 1, \dots, s\}$ folgt aus

$$P(n) = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & \dots & 0 & 0 & 0 \\ 0 & 1-p & 0 & p & \dots & 0 & 0 & 0 \\ \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1-p & 0 & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix} .$$

Zum Zeitpunkt m gilt

$$P(m) = P(n) \quad \text{für } m \neq n . \quad (322.25)$$

Eine Markoff-Kette, die diese Eigenschaft erfüllt, bezeichnet man als *homogen*. Δ

Beispiel 2: Bei einer Irrfahrt mit reflektierenden Grenzen verläßt der Gegenstand die Grenze zum Zeitpunkt $k+1$, falls er sie zur Zeit k erreicht hat. Die Matrix $P(n)$ der Übergangswahrscheinlichkeiten ist dann wie folgt aufgebaut

$$P(n) = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 1-p & 0 & p & \dots & 0 & 0 & 0 \\ 0 & 1-p & 0 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1-p & 0 & p \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{bmatrix} .$$

Es gilt wieder wie im 1. Beispiel

$$P(m) = P(n) \quad \text{für } m \neq n ,$$

so daß die Markoff-Kette für die Irrfahrt mit reflektierenden Gittern homogen ist. Δ

323 Lineare kontinuierliche Systeme

Wie für die deterministischen kontinuierlichen Signale soll ein kontinuierliches System einen stochastischen Prozeß $X(t)$ mittels der Transformationsgleichung

$$Y(t) = \phi[X(t)] \quad (323.1)$$

in den kontinuierlichen stochastischen Prozeß $Y(t)$ überführen. Dabei bezeichnen ϕ analog zu (216.1) den Transformationsoperator, $X(t)$ das kontinuierliche Eingangssignal und $Y(t)$ das kontinuierliche Ausgangssignal. Ist das System linear und verschiebungsvariant, so gilt mit (216.5)

$$Y(t) = \int_{-\infty}^{\infty} h(u)X(t-u)du \quad , \quad (323.2)$$

wobei $h(t)$ wieder die Impulsantwort des Systems bedeutet. Bildet man die Erwartungswertfunktion $E(Y(t)) = \mu_y(t)$, folgt aus (323.2) mit $E(X(t)) = \mu_x(t)$

$$\mu_y(t) = \int_{-\infty}^{\infty} E(h(u)X(t-u))du = \int_{-\infty}^{\infty} h(u)\mu_x(t-u)du \quad , \quad (323.3)$$

da $h(t)$ eine deterministische Funktion darstellt, für die $E(h(t)) = h(t)$ gilt. Es soll nun die Momentfunktion (312.2) des zweiten Momentes

$$K_{yy}(t_1, t_2) = E(Y(t_1)Y(t_2)) \quad (323.4)$$

berechnet werden. Zunächst bildet man mit (323.2) die Momentfunktion $K_{xy}(t_1, t_2)$ von $X(t_1)$ und $Y(t_2)$

$$\begin{aligned} K_{xy}(t_1, t_2) = E(X(t_1)Y(t_2)) &= \int_{-\infty}^{\infty} E(X(t_1)h(u)X(t_2-u))du \\ &= \int_{-\infty}^{\infty} E(X(t_1)X(t_2-u))h(u)du \\ &= \int_{-\infty}^{\infty} K_{xx}(t_1, t_2-u)h(u)du \quad . \end{aligned} \quad (323.5)$$

Weiter erhält man für $K_{yy}(t_1, t_2)$ aus (323.4) mit (323.5)

$$\begin{aligned} K_{yy}(t_1, t_2) &= \int_{-\infty}^{\infty} E(X(t_1-u_1)Y(t_2))h(u_1)du_1 \\ &= \int_{-\infty}^{\infty} K_{xy}(t_1-u_1, t_2)h(u_1)du_1 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K_{xx}(t_1-u_1, t_2-u_2)h(u_1)h(u_2)du_1du_2 \quad . \end{aligned} \quad (323.6)$$

Hiermit ist die Momentfunktion des Ausgangssignals in Abhängigkeit von der Momentfunktion des Eingangssignals dargestellt.

Es wird nun vorausgesetzt, daß $X(t)$ ein stationärer Prozeß sei, so daß nach (313.2) und (313.3) $\mu_x(t) = \text{const.}$ und $K_{xx}(t_1, t_2) = K_{xx}(\tau)$ mit $\tau = t_2 - t_1$ gilt. Unter der Voraussetzung, daß $h(u)$ absolut integrierbar ist, was der Stabilitätsbedingung (216.7) diskreter Systeme entspricht, folgt dann aus (323.3)

$$\mu_y(t) = \int_{-\infty}^{\infty} h(u)\mu_x du = \mu_x \int_{-\infty}^{\infty} h(u)du = \mu_y = \text{const.} \quad (323.7)$$

sowie aus (323.6)

$$K_{yy}(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K_{xx}(\tau+u_1-u_2)h(u_1)h(u_2)du_1du_2 = K_{yy}(\tau) . \quad (323.8)$$

Auch das Ausgangssignal $Y(t)$ ist dann wegen (313.5) zumindest im weiteren Sinne stationär.

Bildet man die Differenz $Y_c(t) = Y(t) - \mu_y(t)$, ergibt sich mit (323.2) und (323.3)

$$Y_c(t) = \int_{-\infty}^{\infty} (X(t-u) - \mu_x(t-u))h(u)du = \int_{-\infty}^{\infty} X_c(t-u)h(u)du \quad (323.9)$$

mit $X_c(t) = X(t) - \mu_x(t)$. Die Autokovarianzfunktion $C_{yy}(t_1, t_2)$ lautet dann nach (312.4)

$$C_{yy}(t_1, t_2) = E(Y_c(t_1)Y_c(t_2)) . \quad (323.10)$$

Dieser Ausdruck ist formal mit (323.4) identisch. Ersetzt man daher in (323.6) $K_{yy}(t_1, t_2)$ durch $C_{yy}(t_1, t_2)$ und $K_{xx}(t_1, t_2)$ durch $C_{xx}(t_1, t_2)$, so erhält man mit

$$C_{yy}(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C_{xx}(t_1-u_1, t_2-u_2)h(u_1)h(u_2)du_1du_2 \quad (323.11)$$

den Zusammenhang zwischen den Autokovarianzfunktionen des Eingangs- und des Ausgangssignals. Weiter ergibt sich für einen stationären stochastischen Prozeß $X(t)$ nach (323.8)

$$C_{yy}(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C_{xx}(\tau+u_1-u_2)h(u_1)h(u_2)du_1du_2 . \quad (323.12)$$

Es soll nun das Spektrum $S_{yy}(\omega)$ des stationären Ausgangssignals $Y(t)$ aus dem Spektrum $S_{xx}(\omega)$ des stationären Eingangssignals $X(t)$ berechnet werden. Aus (314.1) folgt zunächst mit (323.12)

$$\begin{aligned} S_{yy}(\omega) &= \int_{-\infty}^{\infty} C_{yy}(\tau)e^{-j\omega\tau}d\tau \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C_{xx}(\tau+u_1-u_2)h(u_1)h(u_2)e^{-j\omega\tau}du_1du_2d\tau \\ &= \left(\int_{-\infty}^{\infty} h(u_1)e^{j\omega u_1}du_1 \right) \left(\int_{-\infty}^{\infty} h(u_2)e^{-j\omega u_2}du_2 \right) \left(\int_{-\infty}^{\infty} C_{xx}(\tau+u_1-u_2) \right. \\ &\quad \left. e^{-j\omega(\tau+u_1-u_2)}d\tau \right) . \end{aligned}$$

Wird im letzten Integral $u = \tau+u_1-u_2$ substituiert, erhält man mit

$$\int_{-\infty}^{\infty} C_{xx}(u)e^{-j\omega u}du = S_{xx}(\omega)$$

wegen (314.1) sowie mit (214.2) und (213.14)

$$S_{yy}(\omega) = S_{xx}(\omega)H(j\omega)H(-j\omega) = S_{xx}(\omega)|H(j\omega)|^2 . \quad (323.13)$$

Darin ist

$$H(j\omega) = \int_{-\infty}^{\infty} h(\tau)e^{-j\omega\tau}d\tau \quad (323.14)$$

die Frequenzantwort des Systems, die für diskrete Systeme mit (216.10) eingeführt wurde.

Zur Interpretation der Transformation (323.13) des Spektrums wird für das Eingangssignal $X(t)$ die spektrale Darstellung (314.10) gewählt

$$X(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} dD_x(\omega) . \quad (323.15)$$

Setzt man diesen Ausdruck in (323.2) ein, folgt mit (323.14)

$$\begin{aligned} Y(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(u)e^{j\omega(t-u)} dD_x(\omega) du \\ &= \int_{-\infty}^{\infty} H(j\omega)e^{j\omega t} dD_x(\omega) = \int_{-\infty}^{\infty} e^{j\omega t} dD_y(\omega) , \end{aligned} \quad (323.16)$$

wobei gilt

$$dD_y(\omega) = H(j\omega)dD_x(\omega) . \quad (323.17)$$

Aus dieser Beziehung erkennt man die Wirkung der Filterung eines stationären Prozesses. Sie entspricht der Filterung (216.16) deterministischer Signale im Frequenzbereich. Durch Vorgabe der Frequenzantwort $H(j\omega)$ wird die spektrale Darstellung $dD_x(\omega)$ des Eingangssignals $X(t)$ in der gewünschten Weise modifiziert, so daß die spektrale Darstellung $dD_y(\omega)$ des Ausgangssignals $Y(t)$ erhalten wird. Da der differentielle Prozeß $dD_x(\omega)$ in der Regel nicht anzugeben ist, wird (323.17) quadriert und anschließend der Erwartungswert gebildet. Man erhält

$$E(|dD_y(\omega)|^2) = E(|H(j\omega)dD_x(\omega)|^2) = |H(j\omega)|^2 E(|dD_x(\omega)|^2)$$

und weiter mit (314.14) die Gleichung (323.13).

Der Filterung (216.16) deterministischer Signale im Frequenzbereich entspricht also die Transformation (323.13) des Spektrums eines stochastischen Signals.

324 Lineare diskrete Systeme und digitale Filter

Im Kapitel 323 wurden ausschließlich kontinuierliche Systeme und Prozesse betrachtet. Liegt nun ein diskreter stochastischer Prozeß $X(n)$ gemäß (311.3) als Eingangssignal

eines diskreten linearen Systems vor, so gilt anstelle von (323.2) wegen (216.6) für das diskrete Ausgangssignal $Y(n)$

$$Y(n) = \sum_{k=-\infty}^{\infty} h(k)X(n-k) . \quad (324.1)$$

Für die Erwartungwertfunktionen $E(X(n)) = \mu_x(n)$ und $E(Y(n)) = \mu_y(n)$ folgt dann

$$\mu_y(n) = \sum_{k=-\infty}^{\infty} h(k)\mu_x(n-k) . \quad (324.2)$$

Die Momentfunktion $K_{yy}(m,n)$ des Ausgangssignals $Y(n)$ ergibt sich (323.6) entsprechend zu

$$K_{yy}(m,n) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} K_{xx}(m-k,n-l)h(k)h(l) . \quad (324.3)$$

Die Autokovarianzfunktion $C_{yy}(m,n)$ erhält man analog zu (323.11) aus (324.3), wenn K_{xx} durch C_{xx} ersetzt wird.

Ist das diskrete System durch ein kausales nichtrekursives digitales Filter N -ten Grades bestimmt, reduziert sich (324.1) nach (217.14) und (225.4) zu

$$Y(n) = \sum_{k=0}^N h(k)X(n-k) = \sum_{k=0}^N d_k X(n-k) . \quad (324.4)$$

Liegt eine kausale rekursive digitale Filterung N -ten Grades vor, folgt aus (324.1) nach (217.10) und (223.1) die Filtergleichung

$$Y(n) = \sum_{k=0}^N d_k X(n-k) - \sum_{k=1}^N g_k Y(n-k) = \sum_{k=0}^{\infty} h(k)X(n-k) . \quad (324.5)$$

Es wird nun vorausgesetzt, daß das diskrete Eingangssignal $X(n)$ stationär sei. Dann gilt in Analogie zu (323.7) für die Erwartungwertfunktion $E(Y(n)) = \mu_y$ aus (324.5)

$$\mu_y = \mu_x \sum_{k=0}^{\infty} h(k) . \quad (324.6)$$

Die Summe auf der rechten Seite läßt sich durch die Frequenzantwort (216.10) des Systems (324.5) für $\Omega = 0$ beziehungsweise durch die Systemfunktion aus (217.5) für $z = 1$ ausdrücken. Man erhält

$$\mu_y = H(z=1)\mu_x . \quad (324.7)$$

Das Ausgangssignal kann nur dann stationär sein, wenn $|H(z=1)| < \infty$ gilt. Es muß also nach (221.20) die Stabilitätsbedingung (216.7)

$$\sum_{n=0}^{\infty} |h(n)| < \infty \quad (324.8)$$

erfüllt sein. Dies ist nach Satz (221.18) gleichbedeutend mit der Forderung, daß die Beträge der Polstellen z_{0k} der Systemfunktion $H(z)$ aus (221.12)

$$|z_{0k}| < 1 \quad \text{für } k \in \{1, \dots, N\} \quad (324.9)$$

erfüllen.

Die Momentfunktion $K_{yy}(m, n)$ ergibt sich entsprechend (323.8) aus (324.3) mit $p = n - m$ zu

$$K_{yy}(m, n) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} K_{xx}(p+k-1)h(k)h(l) = K_{yy}(p) \quad (324.10)$$

Wegen (324.8) ist die Momentfunktion ebenso wie der Erwartungswert (324.6) beschränkt und das Ausgangssignal $Y(n)$ daher zumindest im weiteren Sinne stationär. Die Forderung (324.8) oder die Forderung (324.9) stellen also die *Stationaritätsbedingung* dar.

Wird (324.7) mit der in (221.26) definierten inversen Systemfunktion $(H(z=1))^{-1}$ multipliziert, erhält man

$$\mu_x = (H(z=1))^{-1} \mu_y \quad (324.11)$$

In diesem Fall findet man als Forderung für die Stabilität, daß die Beträge der Nullstellen z_{0k} der Systemfunktion $H(z)$

$$|z_{0k}| < 1 \quad \text{für } k \in \{1, \dots, N\} \quad (324.12)$$

erfüllen, da in (221.12) Zähler und Nenner bei der Invertierung vertauscht werden. Bei (324.12) handelt es sich also um die *Invertierbarkeitsbedingung* eines Filters.

Schreibt man (217.1) entsprechend für die inverse Systemfunktion

$$(H(z))^{-1} = \bar{H}(z) = \sum_{k=0}^{\infty} \bar{h}(k)z^{-k} \quad (324.13)$$

erläutert der zu (324.4) inverse Prozeß

$$X(n) = \sum_{k=0}^{\infty} \bar{h}(k)Y(n-k) = \sum_{k=0}^{\infty} \bar{d}_k Y(n-k) \quad (324.14)$$

mit der Impulsantwort

$$\bar{h}(k) = \bar{d}_k \quad (324.15)$$

Die Autokovarianzfunktion des nach (324.5) gefilterten Signals $Y(n)$ erhält man

(323.12) entsprechend aus (324.10)

$$C_{yy}(p) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} C_{xx}(p+k-l)h(k)h(l) . \quad (324.16)$$

Den Frequenzinhalt des Ausgangssignals liefert die Transformation des Spektrums, die sich mit (314.25) entsprechend (323.13) zu

$$S_{yy}(\Omega) = S_{xx}(\Omega) |H(j\Omega)|^2 \quad (324.17)$$

ergibt. Durch die Wahl der Frequenzantwort

$$H(j\Omega) = \sum_{k=0}^{\infty} h(k)e^{-jk\Omega} \quad (324.18)$$

wird also der Frequenzinhalt des Eingangssignals modifiziert.

Wie bereits im Zusammenhang mit (323.17) erwähnt, entspricht die Transformation (324.17) des Spektrums eines stochastischen Signals der Filterung (216.16) eines deterministischen Signals im Frequenzbereich. Die Transformation (324.16) der Autokovarianzfunktion eines stochastischen Signals ist vergleichbar mit der Filterung (216.6) eines deterministischen Signals im Zeitbereich. Während (324.16) eine zweifache Faltung im Zeitbereich beinhaltet, ist nach (324.17) im Frequenzbereich mit dem Quadrat der Frequenzantwort zu multiplizieren. Diese Transformation gibt daher wegen (221.31) keinen Aufschluß über die durch die Filterung bewirkte Phasenverschiebung des Eingangssignals. Die Phase eines stochastischen Signals ist jedoch auch ohne Interesse, da sie zufälligen Schwankungen unterworfen ist.

325 Autoregressive Prozesse

Mit den autoregressiven Prozessen und den folgenden MA- und ARMA-Prozessen lassen sich diskrete stochastische Prozesse auf einfache Weise beschreiben.

Ein *kausaler autoregressiver Prozeß* der Ordnung p , abgekürzt geschrieben AR(p)-Prozeß, ist definiert durch

$$X(n) = \sum_{k=1}^p a_k X(n-k) + \varepsilon(n) , \quad (325.1)$$

worin a_k konstante reelle Koeffizienten bedeuten. Die Zufallsvariable $X(n)$ berechnet sich demnach aus der gewichteten Summe der p vorangegangenen Werte der Zufallsvariablen, so daß nach (216.8) ein kausaler Prozeß vorliegt. Außerdem wird ein Rauschanteil $\varepsilon(n)$ berücksichtigt. Da die Koeffizienten a_k im allgemeinen nicht bekannt sind

werden im Kapitel 354 Methoden zu ihrer Schätzung vorgestellt.

Die lineare Abhängigkeit der Zufallsvariablen $X(n)$ von den Werten $X(n-k)$ entspricht einer linearen Regression, und da $X(n)$ von den eigenen vorangegangenen Werten abhängt, spricht man von einem autoregressiven Prozeß. Für seine Erwartungswertfunktion erhalte

$$E(X(n)) = \mu_x(n) = 0 \quad (325.2)$$

Wenn diese Bedingung nicht erfüllt ist, sind in (325.1) von den Zufallsvariablen $X(n)$ und $X(n-k)$ lediglich ihre Erwartungswerte zu subtrahieren.

Der Prozeß $\varepsilon(n)$ repräsentiere stationäres weißes Rauschen mit dem Erwartungswert

$$E(\varepsilon(n)) = 0 \quad (325.3)$$

so daß mit der Varianz σ_ε^2 für die Autokovarianzfunktion nach (321.5) gilt

$$C_{\varepsilon\varepsilon}(k) = E(\varepsilon(n)\varepsilon(n-k)) = \sigma_\varepsilon^2 d(k) \quad (325.4)$$

Löst man (325.1) nach $\varepsilon(n)$ auf, folgt mit $a_0 = -1$

$$\varepsilon(n) = - \sum_{k=0}^p a_k X(n-k) \quad (325.5)$$

Dieser Ausdruck entspricht, wie ein Vergleich mit (324.4) zeigt, einer nichtrekursiven Filtergleichung, für die sich aus (217.13) die Systemfunktion

$$H(z) = - \sum_{k=0}^p a_k z^{-k} \quad (325.6)$$

ergibt. Erfüllen die Nullstellen dieser Systemfunktion die Invertierbarkeitsbedingung (324.12), existiert nach (324.14) der zu (325.5) inverse Prozeß, der $X(n)$ als Funktion des Prozesses $\varepsilon(n)$ darstellt

$$X(n) = \sum_{k=0}^{\infty} \bar{a}_k \varepsilon(n-k) \quad (325.7)$$

Nach (324.13) lautet mit (325.6) die zu diesem inversen Prozeß gehörende Systemfunktion

$$(H(z))^{-1} = \bar{H}(z) = \left(- \sum_{k=0}^p a_k z^{-k} \right)^{-1} = \sum_{k=0}^{\infty} \bar{a}_k z^{-k} \quad (325.8)$$

die wegen (324.12) stabil ist. Nach (324.8) gilt dann

$$\sum_{k=0}^{\infty} |\bar{a}_k| < \infty \quad (325.9)$$

so daß $X(n)$ zumindest im weiteren Sinne stationär ist. Stationarität für den autoregressi-

ven Prozeß wird im folgenden vorausgesetzt.

Beispiel 1: Gegeben sei der AR(1)-Prozeß

$$X(n) = a_1 X(n-1) + \varepsilon(n) .$$

Es lautet dann die Systemfunktion $H(z)$ aus (325.6)

$$H(z) = 1 - a_1 z^{-1}$$

mit der Nullstelle

$$z_{01} = a_1 .$$

Die Bedingung für die Invertierbarkeit $|z_{01}| < 1$ nach (324.12) führt somit auf

$$|a_1| < 1 .$$

Dann gilt nach (325.8) für die inverse Systemfunktion

$$(H(z))^{-1} = \bar{H}(z) = \frac{1}{1 - a_1 z^{-1}} = \sum_{k=0}^{\infty} \bar{a}_k z^{-k}$$

und mit (213.6) für ihre Koeffizienten \bar{a}_k

$$\bar{a}_k = a_1^k \quad \text{für } k \in \{0, \dots, \infty\} .$$

Der AR(1)-Prozeß ist zumindest im weiteren Sinne stationär, wenn die Bedingung $|a_1| < 1$ gilt. Δ

Da $X(n)$ sich nach (325.7) durch die Rauschanteile $\varepsilon(n-k)$ für $k \geq 0$ darstellen läßt, folgt für die Kreuzkovarianzfunktion $C_{\varepsilon X}(m)$ von $X(n)$ und $\varepsilon(n)$, wobei $X(n)$, wie erwähnt, als stationär vorausgesetzt wird,

$$C_{\varepsilon X}(m) = E(\varepsilon(n)X(n-m)) = \sum_{k=0}^{\infty} \bar{a}_k E(\varepsilon(n)\varepsilon(n-m-k))$$

und wegen (325.4)

$$C_{\varepsilon X}(m) = E(\varepsilon(n)X(n-m)) = 0 \quad \text{für } m > 0 . \quad (325.10)$$

Multipliziert man (325.1) mit $\varepsilon(n)$ und bildet den Erwartungswert, ergibt sich mit (325.4) und (325.10)

$$C_{\varepsilon X}(0) = E(\varepsilon(n)X(n)) = \sum_{k=1}^p a_k C_{\varepsilon X}(k) + C_{\varepsilon \varepsilon}(0) = \sigma_{\varepsilon}^2 . \quad (325.11)$$

Für die Berechnung der Autokovarianzfunktion $C_{XX}(m)$ eines stationären autoregressiven Prozesses wird (325.1) mit $X(n-m)$ für $m \geq 0$ multipliziert und anschließend der Erwartungswert gebildet. Man erhält dann mit (313.10), (313.11) und (325.2)

$$\begin{aligned}
 C_{xx}(m) &= E(X(n)X(n-m)) = \sum_{k=1}^p a_k E(X(n-k)X(n-m)) + E(\varepsilon(n)X(n-m)) \\
 &= \sum_{k=1}^p a_k C_{xx}(m-k) + E(\varepsilon(n)X(n-m)) \quad (325.12)
 \end{aligned}$$

und mit (325.10) und (325.11) die Autokovarianzfunktion

$$C_{xx}(m) = \sum_{k=1}^p a_k C_{xx}(m-k) + \sigma_\varepsilon^2 d(m) . \quad (325.13)$$

Für $m = 0$ ergibt sich die Varianz $\sigma_x^2 = C_{xx}(0)$ des Prozesses $X(n)$

$$\sigma_x^2 = \sum_{k=1}^p a_k C_{xx}(k) + \sigma_\varepsilon^2 \quad (325.14)$$

und für $m > 0$

$$C_{xx}(m) = \sum_{k=1}^p a_k C_{xx}(m-k) . \quad (325.15)$$

Dividiert man diese Gleichung durch die Varianz σ_x^2 , erhält man nach (312.6) die Autokorrelationsfunktion

$$\rho_x(m) = \sum_{k=1}^p a_k \rho_x(m-k) \quad \text{für } m > 0 , \quad (325.16)$$

die auch als *Yule-Walker Gleichung* bezeichnet wird. Sie ermöglicht die sukzessive Berechnung von $\rho_x(m)$ aus $\rho_x(m-k)$.

Faßt man die Koeffizienten a_k mit $k \in \{1, \dots, p\}$ im $p \times 1$ Vektor β zusammen

$$\beta = [a_1, a_2, \dots, a_p]' \quad (325.17)$$

und bildet weiterhin den $p \times 1$ Vektor

$$c_p = [C_{xx}(1), C_{xx}(2), \dots, C_{xx}(p)]' \quad (325.18)$$

sowie die $p \times p$ Matrix

$$\Sigma_p = \begin{bmatrix} C_{xx}(0) & C_{xx}(1) & \dots & C_{xx}(p-1) \\ C_{xx}(1) & C_{xx}(0) & \dots & C_{xx}(p-2) \\ \dots & \dots & \dots & \dots \\ C_{xx}(p-1) & C_{xx}(p-2) & \dots & C_{xx}(0) \end{bmatrix} , \quad (325.19)$$

so lauten die ersten p Gleichungen (325.15) für die Autokovarianzen $C_{xx}(m)$ mit $m \in \{1, \dots, p\}$ wegen $C_{xx}(k) = C_{xx}(-k)$ in Matrixschreibweise

$$c_p = \Sigma_p \beta . \quad (325.20)$$

Mit $r_p = c_p/\sigma_x^2$ und $R_p = \Sigma_p/\sigma_x^2$ erhält man die p Yule-Walker Gleichungen (325.16) für $m \in \{1, \dots, p\}$ in der Form

$$r_p = R_p \beta . \quad (325.21)$$

Das Spektrum $S_{xx}(\Omega)$ eines stationären AR(p)-Prozesses ergibt sich aus (324.17) angewendet auf (325.7). Mit (325.8) für $z = e^{j\Omega}$ erhält man dann

$$S_{xx}(\Omega) = S_{\varepsilon\varepsilon}(\Omega) |\bar{H}(j\Omega)|^2 = S_{\varepsilon\varepsilon}(\Omega) \left| \left(- \sum_{k=0}^p a_k e^{-jk\Omega} \right)^{-1} \right|^2 ,$$

worin das Spektrum $S_{\varepsilon\varepsilon}(\Omega)$ des weißen Rauschens $\varepsilon(n)$ aus (321.6) folgt, also

$$S_{xx}(\Omega) = \sigma_\varepsilon^2 / \left| \sum_{k=0}^p a_k e^{-jk\Omega} \right|^2 . \quad (325.22)$$

Beispiel 2: Gegeben sei der AR(1)-Prozeß des Beispiels 1

$$X(n) = a_1 X(n-1) + \varepsilon(n) .$$

Aus (325.22) ergibt sich dann mit (212.11) und (213.14)

$$S_{xx}(\Omega) = \sigma_\varepsilon^2 / |1 - a_1 e^{-j\Omega}|^2 = \sigma_\varepsilon^2 / (1 + a_1^2 - 2a_1 \cos \Omega) . \quad \Delta$$

Es soll nun gezeigt werden, daß ein stationärer kausaler AR(p)-Prozeß, dessen Zufallsvariable $X(n)$ und Rauschanteile $\varepsilon(n)$ als normalverteilt vorausgesetzt werden, ein Markoff-Prozeß ist. Dies ist wegen spezieller Anwendungen besonders bei autoregressiven Prozessen in der Ebene von Interesse, worauf im Kapitel 377 eingegangen wird. Für einen Markoff-Prozeß der Ordnung p gilt nach (322.4) unter Berücksichtigung von (322.2) die Markoff-Eigenschaft

$$p(x_n | x_{n-1}, \dots, x_{n-p}, \dots, x_0) = p(x_n | x_{n-1}, \dots, x_{n-p}) . \quad (325.23)$$

Angewandt auf den kausalen AR(p)-Prozeß lautet die Markoff-Eigenschaft

$$p(X(n) | X(n-1), \dots, X(n-p), \dots, X(0)) = p(X(n) | X(n-1), \dots, X(n-p)) , \quad (325.24)$$

in der zur Vereinfachung der Notation wie auch im folgenden die Werte der Zufallsvariablen $X(n)$ nicht mit kleinen Buchstaben, sondern mit $X(n)$ bezeichnet werden. Sind die Zufallsvariablen $\varepsilon(n)$ und $X(n-m)$ normalverteilt, folgt aus (325.10) die Unabhängigkeit von $\varepsilon(n)$ und $X(n-m)$ für $m > 0$ (Koch 1987, S.139). Aus der Definition der Unabhängigkeit (Koch 1987, S.106) ergibt sich

$$p(\varepsilon(n) | X(n-1), \dots, X(0)) = p(\varepsilon(n)) \quad (325.25)$$

und daher mit (325.3)

$$E(\varepsilon(n) | X(n-1), \dots, X(0)) = 0 \quad (325.26)$$

Mit diesem Ergebnis folgt der bedingte Erwartungswert von $X(n)$ aus (325.1) zu

$$E(X(n) | X(n-1), \dots, X(0)) = E(X(n) | X(n-1), \dots, X(n-p)) = \sum_{k=1}^p a_k X(n-k) \quad (325.27)$$

Außerdem erhält man aus (325.1) und (325.4) die bedingte Varianz von $X(n)$ zu

$$V(X(n) | X(n-1), \dots, X(0)) = V(X(n) | X(n-1), \dots, X(n-p)) = \sigma_\varepsilon^2 \quad (325.28)$$

Einen Prozeß, der die beiden letzten Bedingungen erfüllt, bezeichnet man als *Markoff-Prozeß im weiteren Sinne*. Da $X(n)$ als normalverteilt vorausgesetzt wurde, so daß seine Verteilung durch Erwartungswert und Varianz bestimmt ist, erfüllt $X(n)$ auch die Markoff-Eigenschaft (325.24). Ein stationärer AR(p)-Prozeß mit Normalverteilung stellt also einen Markoff-Prozeß dar. Man bezeichnet ihn daher als *Gauß-Markoff-Prozeß*. Umgekehrt läßt sich auch zeigen, daß ein stationärer Gauß-Markoff-Prozeß als AR(p)-Prozeß darstellbar ist (Woods 1972; Chellappa 1985).

Aufgrund der Normalverteilung für $X(n)$ folgt aus (325.27) und (325.28) die bedingte Normalverteilung für $X(n) | X(n-1), \dots, X(n-p)$ mit

$$X(n) | X(n-1), \dots, X(n-p) \sim N\left(\sum_{k=1}^p a_k X(n-k), \sigma_\varepsilon^2\right) \quad (325.29)$$

Im folgenden soll die Markoff-Eigenschaft (325.24) eines AR(p)-Prozesses mit Normalverteilung noch einmal mit Hilfe der Normalverteilungen für $X(n)$ und $\varepsilon(n)$ nachgewiesen werden. Dies ist insofern instruktiv, als auch die bedingte Verteilung (325.29) noch einmal hergeleitet wird. Mit dem $p \times 1$ Vektor

$$\mathbf{x}_p = [X(n-1), X(n-2), \dots, X(n-p)]'$$

kann der AR(p)-Prozeß (325.1) mit (325.17) in die Form

$$X(n) = [\boldsymbol{\beta}', 1] \begin{bmatrix} \mathbf{x}_p \\ \varepsilon(n) \end{bmatrix} \quad (325.30)$$

gebracht werden. Die Matrix $\boldsymbol{\Sigma}_p$ der Autokovarianzen des Vektors \mathbf{x}_p ist durch (325.19) gegeben und die Varianz des Rauschens $\varepsilon(n)$ durch (325.4). Nimmt man für \mathbf{x}_p und $\varepsilon(n)$ jeweils die Normalverteilung an

$$\mathbf{x}_p \sim N(\mathbf{0}, \boldsymbol{\Sigma}_p), \quad \varepsilon(n) \sim N(0, \sigma_\varepsilon^2) \quad (325.31)$$

sind, wie im Zusammenhang mit (325.25) erwähnt, $\varepsilon(n)$ und $X(n-m)$ für $m > 0$ voneinander unabhängig, und für $X(n)$ folgt als lineare Transformation von \mathbf{x}_p und $\varepsilon(n)$ die Nor-

malverteilung (Koch 1987, S.140)

$$X(n) \sim N(0, [\beta', 1] \begin{bmatrix} \Sigma_p & \mathbf{0} \\ \mathbf{0}' & \sigma_\varepsilon^2 \end{bmatrix} \begin{bmatrix} \beta \\ 1 \end{bmatrix}) .$$

Wegen (325.20) erhält man für die Varianz σ_x^2 von $X(n)$

$$\sigma_x^2 = [\beta', \Sigma_p, \sigma_\varepsilon^2] \begin{bmatrix} \beta \\ 1 \end{bmatrix} = \beta' \Sigma_p \beta + \sigma_\varepsilon^2 = \beta' c_p + \sigma_\varepsilon^2 \quad (325.32)$$

in Übereinstimmung mit (325.14).

Weiter werden der $(n-p) \times 1$ Vektor

$$\mathbf{x}_0 = [X(n-p-1), \dots, X(0)]'$$

und der $n \times 1$ Vektor

$$\mathbf{x}_{n-1} = [X(n-1), X(n-2), \dots, X(n-p), X(n-p-1), \dots, X(0)]' = \begin{bmatrix} \mathbf{x}_p \\ \mathbf{x}_0 \end{bmatrix}$$

eingeführt. Zur Berechnung der Dichte auf der rechten Seite von (325.24) wird von dem $(p+1) \times 1$ Vektor

$$\mathbf{x} = \begin{bmatrix} X(n) \\ \mathbf{x}_p \end{bmatrix}$$

ausgegangen, dessen Matrix der Autokovarianzen mit $\Sigma_n = [\sigma_x^2]$ aus (325.32) lautet

$$\Sigma = \begin{bmatrix} \Sigma_n & \Sigma_{np} \\ \Sigma_{pn} & \Sigma_p \end{bmatrix} \quad \text{mit} \quad \Sigma_{np} = \Sigma_{pn}'$$

und

$$\Sigma_{np} = [C_{xx}(1), C_{xx}(2), \dots, C_{xx}(p)] = c_p'$$

aus (325.18) aufgrund der Stationarität. Dann folgt die bedingte Dichte von $X(n) | \mathbf{x}_p$ aus der Normalverteilung (Koch 1987, S.138)

$$X(n) | \mathbf{x}_p \sim N(\Sigma_{np} \Sigma_p^{-1} \mathbf{x}_p, \Sigma_n - \Sigma_{np} \Sigma_p^{-1} \Sigma_{pn}) , \quad (325.33)$$

für die mit (325.14) und (325.20)

$$X(n) | \mathbf{x}_p \sim N(\beta' \mathbf{x}_p, \sigma_\varepsilon^2) ,$$

erhalten wird. Diese bedingte Verteilung ist mit (325.29) identisch. Um (325.24) zu beweisen, ist nun zu zeigen, daß diese Verteilung mit der von $X(n) | \mathbf{x}_{n-1}$ übereinstimmt. Zunächst gilt analog zu (325.33)

$$X(n) | \mathbf{x}_{n-1} \sim N(\Sigma_{n, n-1} \Sigma_{n-1}^{-1} \mathbf{x}_{n-1}, \Sigma_n - \Sigma_{n, n-1} \Sigma_{n-1}^{-1} \Sigma_{n-1, n}) \quad (325.34)$$

mit der Matrix der Autokovarianzen von \mathbf{x}_{n-1}

$$\Sigma_{n-1} = \begin{bmatrix} \Sigma_p & \Sigma_{p0} \\ \Sigma_{0p} & \Sigma_0 \end{bmatrix} \quad \text{mit} \quad \Sigma_{p0} = \Sigma_{0p}' ,$$

deren Inverse lautet (Koch 1987, S.39)

$$\Sigma_{n-1}^{-1} = \begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{F}' & \mathbf{G} \end{bmatrix}$$

mit den Blockmatrizen

$$\mathbf{E} = \Sigma_p^{-1} + \Sigma_p^{-1} \Sigma_{p0}^{-1} \Sigma_{p0} \Sigma_p^{-1}$$

$$\mathbf{F} = -\Sigma_p^{-1} \Sigma_{p0} \mathbf{G}$$

$$\mathbf{G} = (\Sigma_0 - \Sigma_{0p} \Sigma_p^{-1} \Sigma_{p0})^{-1} .$$

Weiterhin folgt aufgrund der Stationarität mit (325.18) und (325.19)

$$\Sigma_{n,n-1} = [C_{xx}(1), C_{xx}(2), \dots, C_{xx}(p), C_{xx}(p+1), \dots, C_{xx}(n)] = [\mathbf{c}_p', \mathbf{c}_0'] .$$

Dabei gilt für den $(n-p) \times 1$ Vektor $\mathbf{c}_0' = [C_{xx}(p+1), \dots, C_{xx}(n)]$ nach (325.15)

$$\mathbf{c}_0 = \Sigma_{0p} \boldsymbol{\beta} ,$$

mit der $(n-p) \times p$ Matrix

$$\Sigma_{0p} = \begin{bmatrix} C_{xx}(p) & C_{xx}(p-1) & \dots & C_{xx}(1) \\ C_{xx}(p+1) & C_{xx}(p) & \dots & C_{xx}(2) \\ \dots & \dots & \dots & \dots \\ C_{xx}(n-1) & C_{xx}(n-2) & \dots & C_{xx}(n-p) \end{bmatrix} .$$

Berücksichtigt man diese Zusammenhänge in (325.34), ergibt sich mit

$$\mathbf{c}_p' \mathbf{E} + \mathbf{c}_0' \mathbf{F}' = \mathbf{c}_p' \mathbf{E} + \boldsymbol{\beta}' \Sigma_{p0} \mathbf{F}' = \mathbf{c}_p' \Sigma_p^{-1} = \boldsymbol{\beta}'$$

$$\mathbf{c}_p' \mathbf{F} + \mathbf{c}_0' \mathbf{G} = \mathbf{c}_p' \mathbf{F} + \boldsymbol{\beta}' \Sigma_{p0} \mathbf{G} = \mathbf{0}$$

die Verteilung

$$X(n) | \mathbf{x}_{n-1} \sim N(\boldsymbol{\beta}' \mathbf{x}_p, \sigma_\epsilon^2) , \quad (325.35)$$

die mit (325.29) übereinstimmt.

Für die im Kapitel 354 zu behandelnde Schätzung der im Vektor $\boldsymbol{\beta}$ zusammengefaßten Koeffizienten a_k des autoregressiven Modells wird die gemeinsame Dichte der $X(n)$ benötigt. Zunächst wird die gemeinsame bedingte Dichte

$$p(X(n+K), \dots, X(n+1), X(n) | X(n-1), \dots, X(n-p)) \quad (325.36)$$

mit $K > 0$ abgeleitet. Allgemein gilt (Koch 1990, S.49)

$$\begin{aligned} p(X(n+K), \dots, X(n+1), X(n) | X(n-1), \dots, X(n-p)) \\ = p(X(n+K), \dots, X(n+1) | X(n), X(n-1), \dots, X(n-p)) \\ p(X(n) | X(n-1), \dots, X(n-p)) . \end{aligned} \quad (325.37)$$

Aufgrund der Markoff-Eigenschaft (325.24) ergibt sich für die erste Dichte auf der rechten Seite dieser Gleichung

$$\begin{aligned} & p(X(n+K), \dots, X(n+1) | X(n), X(n-1), \dots, X(n-p)) \\ & = p(X(n+K), \dots, X(n+1) | X(n), X(n-1), \dots, X(n-p+1)) . \end{aligned} \quad (325.38)$$

Wird nun mit dieser Dichte in der gleichen Weise verfahren wie mit der Dichte (325.36) in (325.37), erhält man für die gemeinsame bedingte Dichte (325.36)

$$\begin{aligned} & p(X(n+K), \dots, X(n+1), X(n) | X(n-1), \dots, X(n-p)) \\ & = p(X(n+K), \dots, X(n+2) | X(n+1), X(n), \dots, X(n-p+2)) \\ & \quad p(X(n+1) | X(n), \dots, X(n-p+1)) p(X(n) | X(n-1), \dots, X(n-p)) . \end{aligned} \quad (325.39)$$

Schließlich ergibt sich mit dieser Vorgehensweise als Ergebnis

$$\begin{aligned} & p(X(n+K), \dots, X(n+1), X(n) | X(n-1), \dots, X(n-p)) \\ & = \prod_{m=0}^K p(X(n+m) | X(n-1+m), \dots, X(n-p+m)) . \end{aligned} \quad (325.40)$$

Führt man die $p \times 1$ Vektoren

$$\mathbf{x}_{p-m} = [X(n-1+m), X(n-2+m), \dots, X(n-p+m)]'$$

für $m \in \{0, \dots, K\}$ ein, erhält man aus (325.29) die Verteilung von $X(n+m) | \mathbf{x}_{p-m}$ zu

$$X(n+m) | \mathbf{x}_{p-m} \sim N(\boldsymbol{\beta}' \mathbf{x}_{p-m}, \sigma_{\epsilon}^2) \quad (325.41)$$

und aus der Produktbildung nach (325.40) die gemeinsame bedingte Dichte von $X(n+K), \dots, X(n+1), X(n) | \mathbf{x}_p$. Die Dichte der Normalverteilung (325.41) lautet (Koch 1987, S.122)

$$\begin{aligned} & p(X(n+m) | X(n-1+m), \dots, X(n-p+m)) \\ & = \left[\frac{1}{2\pi\sigma_{\epsilon}^2} \right]^{1/2} \exp \left[-\frac{1}{2\sigma_{\epsilon}^2} (X(n+m) - \boldsymbol{\beta}' \mathbf{x}_{p-m})^2 \right] . \end{aligned} \quad (325.42)$$

und folglich die gemeinsame Dichte von $X(n+K), \dots, X(n)$ unter der Bedingung, daß $X(n-1), \dots, X(n-p)$ gegeben ist

$$\begin{aligned} & p(X(n+K), \dots, X(n) | X(n-1), \dots, X(n-p)) \\ & = \left[\frac{1}{2\pi\sigma_{\epsilon}^2} \right]^{(K+1)/2} \exp \left[-\frac{1}{2\sigma_{\epsilon}^2} \sum_{m=0}^K (X(n+m) - \boldsymbol{\beta}' \mathbf{x}_{p-m})^2 \right] . \end{aligned} \quad (325.43)$$

Mit dem $(K+1) \times 1$ Vektor

$$\bar{\mathbf{y}} = [X(n+K), X(n+K-1), \dots, X(n)]' \quad (325.44)$$

und der $(K+1) \times p$ Matrix

$$\mathbf{X} = [\mathbf{x}_{p-K}, \mathbf{x}_{p-K+1}, \dots, \mathbf{x}_p]' \quad (325.45)$$

geht (325.43) über in

$$\begin{aligned} & p(X(n+K), \dots, X(n) | X(n-1), \dots, X(n-p)) \\ &= \left[\frac{1}{2\pi\sigma_\varepsilon^2} \right]^{(K+1)/2} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} (\bar{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})' (\bar{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}) \right]. \end{aligned} \quad (325.46)$$

Die gemeinsame Dichte von $X(n+K), \dots, X(n-p)$ berechnet sich nach (322.5) aus

$$\begin{aligned} & p(X(n+K), \dots, X(n+1), X(n), X(n-1), \dots, X(n-p)) \\ &= p(X(n-1), \dots, X(n-p)) p(X(n+K), \dots, X(n) | X(n-1), \dots, X(n-p)). \end{aligned} \quad (325.47)$$

Während der zweite Faktor auf der rechten Seite durch (325.46) gegeben ist, erhält man die Dichte $p(X(n-1), \dots, X(n-p))$ aus (325.31) zu

$$p(X(n-1), \dots, X(n-p)) = \left[\frac{1}{(2\pi)^p \det \boldsymbol{\Sigma}_p} \right]^{1/2} \exp \left[-\frac{1}{2} \mathbf{x}_p' \boldsymbol{\Sigma}_p^{-1} \mathbf{x}_p \right]. \quad (325.48)$$

Setzt man (325.46) und (325.48) in (325.47) ein, erhält man schließlich mit $K+1+p = q$ sowie $\boldsymbol{\Sigma}_p = \sigma_\varepsilon^2 \mathbf{Q}_p$ die gemeinsame Dichte von $X(n+K), \dots, X(n-p)$ zu

$$\begin{aligned} & p(X(n+K), \dots, X(n+1), X(n), X(n-1), \dots, X(n-p)) \\ &= \left[\frac{1}{(2\pi\sigma_\varepsilon^2)^q \det \mathbf{Q}_p} \right]^{1/2} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} (\mathbf{x}_p' \mathbf{Q}_p^{-1} \mathbf{x}_p + (\bar{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})' (\bar{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})) \right]. \end{aligned} \quad (325.49)$$

Mit (325.1) wurde ein kausaler AR(p)-Prozeß definiert. Nichtkausale AR(p)-Prozesse sind im zweidimensionalen Fall von Interesse und werden im Kapitel 378 behandelt. Bei nichtkausalen zweidimensionalen autoregressiven Prozessen läßt sich die gemeinsame bedingte Dichte des Prozesses nicht entsprechend (325.40) ableiten. Im folgenden wird daher die gemeinsame bedingte Dichte (325.43) auf eine Weise hergeleitet, die auch für zweidimensionale autoregressive Prozesse anwendbar ist. Die gemeinsame Dichte der Rauschanteile $\varepsilon(n+K), \varepsilon(n+K-1), \dots, \varepsilon(n)$ folgt wegen (325.3), (325.4) und (325.31) mit (Koch 1987, S.106)

$$\begin{aligned} & p(\varepsilon(n+K), \varepsilon(n+K-1), \dots, \varepsilon(n)) \\ &= \left[\frac{1}{2\pi\sigma_\varepsilon^2} \right]^{(K+1)/2} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{m=0}^K \varepsilon(n+m)^2 \right]. \end{aligned} \quad (325.50)$$

Bildet man (325.5) für $\varepsilon(n+K), \varepsilon(n+K-1), \dots, \varepsilon(n)$, erhält man

$$\varepsilon(n+K) = X(n+K) - a_1X(n+K-1) - a_2X(n+K-2) - \dots - a_pX(n+K-p)$$

$$\varepsilon(n+K-1) = X(n+K-1) - a_1X(n+K-2) - a_2X(n+K-3) - \dots - a_pX(n+K-p-1)$$

...

$$\varepsilon(n) = X(n) - a_1X(n-1) - a_2X(n-2) - \dots - a_pX(n-p) . \quad (325.51)$$

Transformiert man nun die Zufallsvariablen $\varepsilon(n+K), \varepsilon(n+K-1), \dots, \varepsilon(n)$ in der Dichte (325.50) mit (325.51) in die Zufallsvariablen $X(n+K), X(n+K-1), \dots, X(n)$, dann ergibt sich die Jacobische Matrix $(\partial\varepsilon(i)/\partial X(j))$ der Transformation als obere Einheitsdreiecksmatrix. Die Dichte von $X(n+K), X(n+K-1), \dots, X(n)$ unter der Bedingung, daß die Werte der Zufallsvariablen $X(n-1), X(n-2), \dots, X(n-p)$ als Startwerte gegeben sind, folgt daher mit (Koch 1987, S.107)

$$\begin{aligned} p(X(n+K), X(n+K-1), \dots, X(n) | X(n-1), X(n-2), \dots, X(n-p)) \\ = \left[\frac{1}{(2\pi\sigma_\varepsilon^2)} \right]^{(K+1)/2} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{m=0}^K (X(n+m) - \beta' \mathbf{x}_{p-m})^2 \right] . \end{aligned} \quad (325.52)$$

Diese Dichte ist identisch mit (325.43).

326 Erweitertes autoregressives Modell

Es wird nun als Verallgemeinerung des autoregressiven Prozesses das *erweiterte autoregressive Modell* eingeführt, das von dem Ansatz

$$Y(n) = X(n) + b(n) + r(n) \quad (326.1)$$

ausgeht, worin $Y(n)$ das beobachtete Signal, $X(n)$ der AR(p)-Prozeß, $b(n)$ eine Trendfunktion und $r(n)$ das Beobachtungsrauschen bedeuten. Der Trend sei eine beliebige Funktion fester Parameter c_i mit $i \in \{1, \dots, q\}$. In diesem deterministischen Modellansatz werden Anteile des Signals erfaßt, die sich physikalisch erklären und mathematisch beschreiben lassen. Der AR(p)-Prozeß $X(n)$ enthält den zufälligen Anteil, der nicht vom Beobachtungsrauschen $r(n)$ herrührt. Das beobachtete Signal $Y(n)$ wird somit als Summe eines stochastischen Anteils $X(n)$ eines deterministischen Anteils $b(n)$ und des Beobachtungsrauschens $r(n)$ dargestellt. Anwendungen dieses Modells finden sich bei (Schulte 1987, S.70).

Setzt man in (326.1) $n-k$ für n ein, löst nach $X(n-k)$ auf und führt den resultierenden Ausdruck in (325.1) ein, ergibt sich

$$X(n) = \sum_{k=1}^p a_k (Y(n-k) - b(n-k) - r(n-k)) + \varepsilon(n) \quad (326.2)$$

und weiter aus (326.1)

$$Y(n) = \sum_{k=1}^p a_k (Y(n-k) - b(n-k) - r(n-k)) + \varepsilon(n) + b(n) + r(n) . \quad (326.3)$$

Das Beobachtungsrauschen $r(n)$ sei stationäres weißes Rauschen mit der Erwartungswertfunktion

$$E(r(n)) = 0 \quad (326.4)$$

und der Autokovarianzfunktion

$$C_{rr}(k) = \sigma_r^2 d(k) . \quad (326.5)$$

Setzt man neben $r(n)$ auch den Prozeß $X(n)$ wieder als stationär voraus, so ist auch der Prozeß

$$Z(n) = Y(n) - b(n) = X(n) + r(n) \quad (326.6)$$

stationär. Führt man weiterhin

$$\alpha(n) = \varepsilon(n) + r(n) - \sum_{k=1}^p a_k r(n-k) \quad (326.7)$$

ein, erhält man aus (326.3) den Zusammenhang

$$Z(n) = \sum_{k=1}^p a_k Z(n-k) + \alpha(n) , \quad (326.8)$$

der formal mit (325.1) übereinstimmt. Der Prozeß $Z(n)$ ist somit in der Form (325.7) darstellbar. Betrachtet wird nun die Kreuzkovarianzfunktion

$$C_{\alpha z}(m) = E(\alpha(n)Z(n-m)) . \quad (326.9)$$

Setzt man hierin (326.6) für $n-m$ und (326.7) ein, ergibt sich mit $a_0 = -1$

$$\begin{aligned} E((\varepsilon(n) - \sum_{k=0}^p a_k r(n-k))(X(n-m) + r(n-m))) \\ = E(\varepsilon(n)X(n-m)) + E(\varepsilon(n)r(n-m)) - \sum_{k=0}^p a_k E(r(n-k)X(n-m)) \\ - \sum_{k=0}^p a_k E(r(n-k)r(n-m)) . \end{aligned} \quad (326.10)$$

Wird angenommen, daß das Beobachtungsrauschen $r(n)$ sowohl vom Signal $X(n)$ als auch vom Rauschanteil $\varepsilon(n)$ unabhängig ist

$$C_{rx}(k) = E(r(n)X(n-k)) = 0 \quad (326.11)$$

$$C_{r\varepsilon}(k) = E(r(n)\varepsilon(n-k)) = 0 \quad (326.12)$$

für alle k , so folgt aus (326.10) mit (325.11) und (326.5) für $m = 0$

$$C_{\alpha z}(0) = \sigma_{\varepsilon}^2 + \sigma_r^2 \quad (326.13)$$

sowie mit (325.10) für $m \in \{1, \dots, p\}$

$$C_{\alpha z}(m) = -a_m \sigma_r^2 . \quad (326.14)$$

Gilt $m > p$, verschwindet die Kreuzkovarianzfunktion $C_{\alpha z}(m)$. Multipliziert man nun (326.8) mit $Z(n-m)$ für $m \geq 0$ und bildet anschließend den Erwartungswert, erhält man die Autokovarianzfunktion des Prozesses $Z(n)$

$$C_{zz}(m) = E(Z(n)Z(n-m)) = \sum_{k=1}^p a_k C_{zz}(m-k) + C_{\alpha z}(m) . \quad (326.15)$$

Für $m = 0$ ergibt sich die Varianz $C_{zz}(0) = \sigma_z^2$ mit (326.13) zu

$$\sigma_z^2 = \sum_{k=1}^p a_k C_{zz}(k) + \sigma_{\varepsilon}^2 + \sigma_r^2 . \quad (326.16)$$

Außerdem erhält man für $m \in \{1, \dots, p\}$ mit (326.14)

$$C_{zz}(m) = \sum_{k=1}^p a_k C_{zz}(m-k) - a_m \sigma_r^2 . \quad (326.17)$$

Dividiert man (326.17) durch die Varianz (326.16), erhält man analog zu (325.16) die Yule-Walker Gleichung des erweiterten autoregressiven Modells

$$\rho_z(m) = \sum_{k=1}^p a_k \rho_z(m-k) - a_m (\sigma_r^2 / \sigma_z^2) . \quad (326.18)$$

Das Spektrum des erweiterten autoregressiven Modells ist bei (Schulte 1987, S.36) abgeleitet.

Die unbekannt Parameter des erweiterten autoregressiven Modells sind die p autoregressiven Parameter a_k , die q Trendparameter c_i sowie die Varianz σ_{ε}^2 aus (325.4) des autoregressiven Modellanteils und die Varianz σ_r^2 aus (326.5) des Beobachtungsrauschens. Auf die Schätzung dieser unbekannt Parameter wird im Kapitel 355 eingegangen.

327 MA-Prozesse und ARMA-Prozesse

Ersetzt man in (325.1) $X(n-k)$ durch $\varepsilon(n-k)$, die Koeffizienten a_k durch $-b_k$ sowie p durch q , erhält man mit $b_0 = -1$ die Definitionsgleichung eines *MA(q)-Prozesses*

$$X(n) = - \sum_{k=0}^q b_k \varepsilon(n-k) . \quad (327.1)$$

Dabei ist "MA" die Abkürzung für die englische Bezeichnung "moving average", denn $\varepsilon(n)$ wird als gewichteter Durchschnitt der Werte des Prozesses $\varepsilon(n)$ dargestellt, wobei mit wachsendem n die Mittelwertbildung an dem Prozeß $\varepsilon(n)$ entlang bewegt wird. Vergleicht man (327.1) mit (325.7), so zeigt sich, daß der inverse Prozeß eines AR(p)-Prozesses ein MA(∞)-Prozeß ist, dessen Systemfunktion durch (325.8) gegeben ist. Demzufolge erhält man die Systemfunktion des allgemeinen MA(q)-Prozesses (327.1) zu

$$H(z) = - \sum_{k=0}^q b_k z^{-k} . \quad (327.2)$$

Da (325.9) entsprechend

$$\sum_{k=0}^q |b_k| < \infty$$

gilt, ist der MA(q)-Prozeß nach (324.8) zumindest im weiteren Sinne stationär. Die Erwartungswertfunktion $E(X(n))$ berechnet sich nach (324.6) mit $E(\varepsilon(n)) = 0$ gemäß (325.3) zu

$$E(X(n)) = \mu_x = 0 . \quad (327.3)$$

Die Autokovarianzfunktion des MA(q)-Prozesses ergibt sich aus (324.16) mit (325.4) zu

$$C_{xx}(m) = \sigma_\varepsilon^2 \sum_{k=0}^q \sum_{l=0}^q b_k b_l \delta(m+k-l) . \quad (327.4)$$

Aufgrund der Definition (211.7) des Einheitsimpulses kann $C_{xx}(m)$ nur für $m = l-k$ und $|m| \leq q$ von Null verschieden sein. Mit $l = k+m$ und $k = l-m$ erhält man dann aus (327.4)

$$C_{xx}(m) = \sigma_\varepsilon^2 \sum_{k=0}^{q-m} b_k b_{k+m} \quad \text{für } 0 \leq m \leq q . \quad (327.5)$$

Wie oben erwähnt, gilt außerdem

$$C_{xx}(m) = 0 \quad \text{für } m > q . \quad (327.6)$$

Weiter gilt nach (313.13) $C_{xx}(m) = C_{xx}(-m)$ für beliebige Werte von m . Für $m = 0$ folgt aus (327.5) die Varianz σ_x^2 des MA(q)-Prozesses zu

$$\sigma_x^2 = C_{xx}(0) = \sigma_\varepsilon^2 \sum_{k=0}^q b_k^2 . \quad (327.7)$$

Die Autokorrelationsfunktion ergibt sich aus (327.5) und (327.6) durch Division mit (327.7) zu

$$\rho_x(m) = \left(\sum_{k=0}^{q-m} b_k b_{k+m} \right) / \left(\sum_{k=0}^q b_k^2 \right) \quad \text{für } 0 \leq m \leq q$$

$$\rho_x(m) = 0 \quad \text{für } m > q \quad (327.8)$$

und $\rho_x(m) = \rho_x(-m)$ für beliebige Werte von m . Während also die Autokorrelationsfunktion eines MA(q)-Prozesses für $m > q$ verschwindet, gilt dies im Fall eines AR(p)-Prozesses für $m > p$ nach (325.16) nicht.

Das Spektrum $S_{xx}(\Omega)$ des MA(q)-Prozesses berechnet sich aus (324.17) mit (327.2) und (321.6) zu

$$S_{xx}(\Omega) = S_{\varepsilon\varepsilon} |H(j\Omega)|^2 = \sigma_\varepsilon^2 \left| - \sum_{k=0}^q b_k e^{-jk\Omega} \right|^2 . \quad (327.9)$$

Der MA(q)-Prozeß ist invertierbar, falls nach (324.12) für die Nullstellen z_{0k} mit $k \in \{1, \dots, q\}$ der Systemfunktion (327.2) die Bedingung $|z_{0k}| < 1$ gilt. Dann ergibt sich nach (324.14) der zu (327.1) inverse Prozeß

$$\varepsilon(n) = \sum_{k=0}^{\infty} \bar{b}_k X(n-k) , \quad (327.10)$$

der nach (325.5) einem AR(∞)-Prozeß entspricht. Hierin zeigt sich die Dualität von AR(p)-Prozessen und MA(q)-Prozessen. Ist der AR(p)-Prozeß invertierbar, resultiert nach (325.7) ein MA(∞)-Prozeß. Ist hingegen ein MA(q)-Prozeß invertierbar, erhält man nach (327.10) einen AR(∞)-Prozeß.

Beispiel 1: Gegeben sei der MA(1)-Prozeß

$$X(n) = \varepsilon(n) - b_1 \varepsilon(n-1)$$

sowie die Varianz σ_ε^2 . Die Varianz dieses MA(1)-Prozesses berechnet sich aus (327.7) zu

$$\sigma_x^2 = \sigma_\varepsilon^2 (1 + b_1^2) .$$

Die Autokorrelationsfunktion ergibt sich aus (327.8)

$$\rho_x(m) = -b_1 / (1 + b_1^2) \quad \text{für } m = -1 \quad \text{und } m = 1$$

$$\rho_x(m) = 1 \quad \text{für } m = 0$$

$$\rho_x(m) = 0 \quad \text{sonst} .$$

Wegen $b_1 / (1 + b_1^2) = (1/b_1) / (1 + (1/b_1)^2)$ sind die Autokorrelationsfunktionen für MA(1)-Prozesse mit den Parametern b_1 und $1/b_1$ identisch. Zwei unterschiedliche MA(1)-Prozesse können also die gleiche Autokorrelationsfunktion besitzen.

Das Spektrum $S_{xx}(\Omega)$ lautet mit (212.11) nach (327.9)

$$S_{xx}(\Omega) = \sigma_\varepsilon^2 |1 - b_1 e^{-j\Omega}|^2 = \sigma_\varepsilon^2 (1 + b_1^2 - 2b_1 \cos \Omega) .$$

Die Systemfunktion (327.2) ergibt sich zu

$$H(z) = 1 - b_1 z^{-1}$$

und hat die Nullstelle

$$z_{01} = b_1 .$$

Analog der Darstellung im Beispiel 1 des Kapitels 325 erhält man für die Koeffizienten \bar{b}_k des inversen Prozesses

$$\bar{b}_k = b_1^k \quad \text{für } k \in \{0, \dots, \infty\} \quad \text{sowie } |b_1| < 1$$

und aus (327.10) als inversen Prozeß den AR(∞)-Prozeß

$$X(n) = - \sum_{k=1}^{\infty} b_1^k X(n-k) + \varepsilon(n) . \quad \Delta$$

Durch Kombination eines AR(p)- und eines MA(q)-Prozesses erhält man als Verallgemeinerung den ARMA(p, q)-Prozeß. Mit (325.1) und (327.1) ist ein kausaler ARMA(p, q)-Prozeß definiert durch

$$X(n) = \sum_{k=1}^p a_k X(n-k) - \sum_{l=0}^q b_l \varepsilon(n-l) . \quad (327.11)$$

Diese Darstellung entspricht der Filtergleichung (216.20). Mit den Überlegungen, die zu (217.7) führen, erhält man daher als Systemfunktion des ARMA(p, q)-Prozesses mit $a_0 = -1$

$$H(z) = \sum_{l=0}^q b_l z^{-l} / \sum_{k=0}^p a_k z^{-k} . \quad (327.12)$$

Der ARMA(p, q)-Prozeß ist nach (324.9) zumindest im weiteren Sinne stationär, wenn die Polstellen $z_{\infty k}$ für $k \in \{1, \dots, p\}$ der Systemfunktion (327.12) die Bedingung $|z_{\infty k}| < 1$ erfüllen. Weiter ist der ARMA(p, q)-Prozeß nach (324.12) invertierbar, falls die Nullstellen z_{0l} für $l \in \{1, \dots, q\}$ die Bedingung $|z_{0l}| < 1$ erfüllen. Analog zu (217.10) läßt sich $X(n)$ dann als Funktion aller vorangegangener Werte von $\varepsilon(n-k)$ für $k \in \{1, \dots, \infty\}$ darstellen. Setzt man zunächst in (327.11)

$$\bar{\varepsilon}(n) = - \sum_{l=0}^q b_l \varepsilon(n-l) , \quad (327.13)$$

erhält man wegen (325.1) den AR(p)-Prozeß

$$X(n) = \sum_{k=1}^p a_k X(n-k) + \bar{\varepsilon}(n) , \quad (327.14)$$

dessen inverser Prozeß nach (325.7) lautet

$$X(n) = \sum_{k=0}^{\infty} \bar{a}_k \bar{\varepsilon}(n-k) . \quad (327.15)$$

Substituiert man hierin (327.13), erhält man

$$X(n) = - \sum_{k=0}^{\infty} \sum_{l=0}^q \bar{a}_k b_l \varepsilon(n-k-l) . \quad (327.16)$$

Nach (327.1) ist dies ein $MA(\infty)$ -Prozeß. Der zu (327.13) inverse Prozeß lautet nach (327.10)

$$\varepsilon(n) = \sum_{l=0}^{\infty} \bar{b}_l \bar{\varepsilon}(n-l) . \quad (327.17)$$

Setzt man darin $\bar{\varepsilon}(n)$ aus (327.14) ein, ergibt sich wegen

$$\varepsilon(n) = - \sum_{k=0}^p \sum_{l=0}^{\infty} a_k \bar{b}_l X(n-k-l) \quad (327.18)$$

nach (325.5) ein $AR(\infty)$ -Prozeß. Ein $ARMA(p,q)$ -Prozeß läßt sich also, falls die Bedingungen für die Stationarität und die Invertierbarkeit erfüllt sind, nach (327.16) als $MA(\infty)$ -Prozeß oder nach (327.18) als $AR(\infty)$ -Prozeß darstellen.

Multipliziert man (327.11) mit $X(n-m)$ für $m \geq 0$ und bildet den Erwartungswert, folgt die Autokovarianzfunktion eines stationären $ARMA(p,q)$ -Prozesses

$$C_{xx}(m) = E(X(n)X(n-m)) = \sum_{k=1}^p a_k E(X(n-k)X(n-m)) - \sum_{l=0}^q b_l E(\varepsilon(n-l)X(n-m)) . \quad (327.19)$$

Für den Erwartungswert in der letzten Summe gilt mit (325.10) und $n-l = s$ sowie $n = l+s$

$$E(\varepsilon(s)X(s-(m-l))) = C_{\varepsilon x}(m-l) = 0 \quad (327.20)$$

für $m-l > 0$. Somit erhält man für $|m| \leq q$ die Autokovarianzfunktion

$$C_{xx}(m) = \sum_{k=1}^p a_k C_{xx}(m-k) - \sum_{l=0}^q b_l C_{\varepsilon x}(m-l) \quad (327.21)$$

mit $C_{\varepsilon x}(0) = \sigma_{\varepsilon}^2$ aus (325.11) und für $m > q$ wegen (327.20)

$$C_{xx}(m) = \sum_{k=1}^p a_k C_{xx}(m-k) . \quad (327.22)$$

Für $m > q$ ist die Autokovarianzfunktion des $ARMA(p,q)$ -Prozesses mit der Autokovarianzfunktion (325.15) des $AR(p)$ -Prozesses identisch.

Das Spektrum $S_{xx}(\Omega)$ berechnet sich aus (324.17)

$$S_{xx}(\Omega) = S_{\varepsilon\varepsilon}(\Omega) |H(j\Omega)|^2 ,$$

wobei die Frequenzantwort $H(j\Omega)$ durch (327.12) für $z = e^{j\Omega}$ gegeben ist. Man erhält dann mit (321.6)

$$S_{xx}(\Omega) = \sigma_\varepsilon^2 \left| \sum_{l=0}^q b_l e^{-jl\Omega} \right|^2 / \left| \sum_{k=0}^p a_k e^{-jk\Omega} \right|^2 . \quad (327.23)$$

Beispiel 2: Gegeben sei der ARMA(1,1)-Prozeß

$$X(n) = a_1 X(n-1) + \varepsilon(n) - b_1 \varepsilon(n-1) .$$

Für $m = 0$ und $m = 1$ erhält man aus (327.21) mit (325.14) und (327.20)

$$C_{xx}(0) = \sigma_x^2 = a_1 C_{xx}(1) + \sigma_\varepsilon^2 - b_1 C_{\varepsilon\varepsilon}(-1)$$

$$C_{xx}(1) = a_1 \sigma_x^2 - b_1 \sigma_\varepsilon^2 .$$

Für $m > 1$ ergibt sich aus (327.22)

$$C_{xx}(m) = a_1 C_{xx}(m-1) .$$

Die Systemfunktion $H(z)$ aus (327.12) lautet für den ARMA(1,1)-Prozeß

$$H(z) = \frac{-1 + b_1/z}{-1 + a_1/z} = \frac{z - b_1}{z - a_1} .$$

Sie besitzt bei $z = z_{01} = b_1$ die Nullstelle und bei $z = z_{\infty 1} = a_1$ die Polstelle. Der Prozeß ist daher zumindest im weiteren Sinne stationär, falls nach (324.9) $|a_1| < 1$ gilt und invertierbar, wenn nach (324.12) $|b_1| < 1$ zutrifft. Δ

33 Schätzungen von Momentfunktionen

331 Ergodizität

Wie am Ende des Kapitels 311 erwähnt, muß man sich zur Beschreibung eines stochastischen Prozesses häufig auf die Momentfunktionen beschränken. Sind auch sie nicht gegeben, müssen sie geschätzt werden, beispielsweise für die Optimalfilter der Kapitel 352 und 353.

In der Definition (312.1) der Erwartungswertfunktion eines stochastischen Prozesses sowie in den Definitionen (312.2) und (312.4) der Momentfunktionen ist über sämtliche Werte einer Zufallsvariablen des stochastischen Prozesses zu integrieren. Man benötigt also sämtliche Realisierungen des stochastischen Prozesses, das heißt das Ensemble. Für viele stochastische Prozesse liegt aber nur eine Realisierung vor, denn beispielsweise gibt es nur ein einziges Erdschwerefeld oder ein einziges Geomagnetfeld. Man kann sich daher die Frage stellen, ob bereits eine Realisierung eines stochastischen Prozesses alle statistischen Eigenschaften des Prozesses enthält. Ist das der Fall, bezeichnet man den stochastischen Prozeß als *ergodisch*. Voraussetzung für die Ergodizität ist die Stationarität im engeren Sinne, denn nur dann lassen sich die Momentfunktionen eines stochastischen Prozesses aus einer Realisierung ableiten, wenn sie nicht von der Zeit t abhängen.

Der Nachweis der Ergodizität ist in der Regel nicht zu führen, so daß die allgemeine Definition der Ergodizität auf bestimmte statistische Größen, wie den Erwartungswert oder die Autokovarianzfunktion, eingeschränkt wird. Man bezeichnet einen im engeren oder weiteren Sinne stationären Prozeß *ergodisch bezüglich seines Erwartungswertes*, wenn sein Erwartungswert aus einer Realisierung des stochastischen Prozesses zu bestimmen ist. Es soll nun untersucht werden, unter welchen Voraussetzungen die Ergodizität bezüglich des Erwartungswertes vorliegt.

Es sei $X(t)$ ein im engeren oder weiteren Sinne stationärer Prozeß, dann ist nach (313.2) oder (313.5) seine Erwartungswertfunktion

$$E(X(t)) = \mu \quad (331.1)$$

eine Konstante. Aus dem stationären Prozeß im Intervall $(-T, T)$ wird der Mittelwert μ_T

$$\mu_T = \frac{1}{2T} \int_{-T}^T X(t) dt \quad (331.2)$$

gebildet, der eine Zufallsvariable darstellt. Ergodizität bezüglich des Erwartungswertes

ist dann gegeben, wenn mit einer Wahrscheinlichkeit von Eins die Zufallsvariable μ_T den Wert μ für $T \rightarrow \infty$ annimmt, also

$$\lim_{T \rightarrow \infty} \mu_T = \mu . \quad (331.3)$$

Für den Erwartungswert $E(\mu_T)$ von μ_T gilt mit (331.1) und (331.2)

$$E(\mu_T) = \frac{1}{2T} E\left(\int_{-T}^T X(t) dt\right) = \frac{1}{2T} \int_{-T}^T E(X(t)) dt = \mu , \quad (331.4)$$

so daß für die Varianz $V(\mu_T)$ von μ_T mit (312.4) und (313.6) folgt

$$\begin{aligned} V(\mu_T) &= E((\mu_T - \mu)^2) \\ &= E\left(\left[\frac{1}{2T} \int_{-T}^T X(t_1) dt_1 - \mu\right] \left[\frac{1}{2T} \int_{-T}^T X(t_2) dt_2 - \mu\right]\right) \\ &= \frac{1}{4T^2} \int_{-T}^T \int_{-T}^T E([X(t_1) - \mu][X(t_2) - \mu]) dt_1 dt_2 \\ &= \frac{1}{4T^2} \int_{-T}^T \int_{-T}^T C(t_2 - t_1) dt_1 dt_2 . \end{aligned} \quad (331.5)$$

Geht mit $T \rightarrow \infty$ die Varianz $V(\mu_T)$ gegen Null, dann ist mit einer Wahrscheinlichkeit von Eins (331.3) erfüllt. Um die Voraussetzung hierfür zu prüfen, wird das letzte Integral in (331.5) durch die Substitution $t_2 - t_1 = y$ und $t_1 = x$ umgeformt. Für die Funktionaldeterminante $\det J$ der Transformation (Koch 1987, S.83) gilt $\det J = 1$, und das rechteckige Integrationsgebiet wird in ein Parallelogramm transformiert, siehe Abbildung 331-1. An-

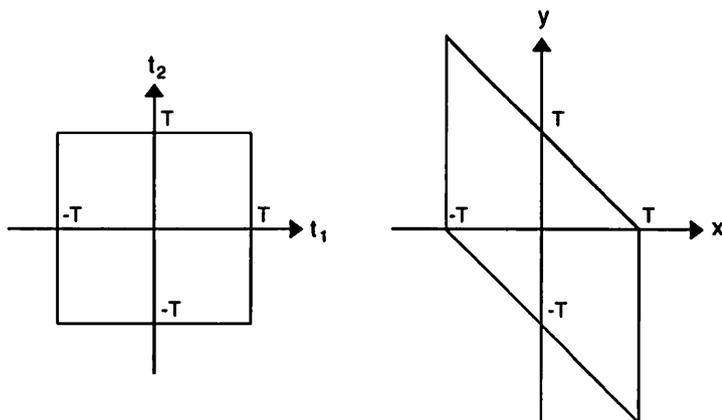


Abb. 331-1: Transformation des Integrationsgebietes

stelle von (331.5) folgt dann

$$\begin{aligned} V(\mu_T) &= \frac{1}{4T^2} \left(\int_{-2T}^0 C(y) \int_{-T-y}^T dx dy + \int_0^{2T} C(y) \int_{-T}^{T-y} dx dy \right) \\ &= \frac{1}{4T^2} \left(\int_{-2T}^0 C(y) (2T+y) dy + \int_0^{2T} C(y) (2T-y) dy \right) . \end{aligned}$$

Mit $y = \tau$ ergibt sich schließlich

$$V(\mu_T) = \frac{1}{2T} \int_{-2T}^{2T} C(\tau) \left(1 - \frac{|\tau|}{2T}\right) d\tau . \quad (331.6)$$

Ist nun die Autokovarianzfunktion $C(\tau)$ nach (214.6) absolut integrierbar

$$\int_{-\infty}^{\infty} |C(\tau)| d\tau < \infty , \quad (331.7)$$

dann gilt

$$\left| \int_{-2T}^{2T} C(\tau) \left(1 - \frac{|\tau|}{2T}\right) d\tau \right| \leq \int_{-2T}^{2T} |C(\tau)| d\tau < \infty ,$$

und es folgt

$$\lim_{T \rightarrow \infty} V(\mu_T) = 0 . \quad (331.8)$$

Damit ist (331.3) erfüllt, so daß der stationäre Prozeß ergodisch bezüglich des Erwartungswertes ist. Die Ungleichung (331.7) gilt, wenn die Autokovarianzfunktion $C(\tau)$ für große Werte von τ gegen Null geht. Um das zu gewährleisten, muß die Realisierung eines stationären Prozesses über einen sehr viel längeren Zeitraum vorliegen, als der Zeitraum beträgt, über den die Autokovarianzfunktion $C(\tau)$ von Null verschiedene Werte annimmt.

Für einen diskreten stationären Prozeß $X(n)$ mit $n \in \{-N, -N+1, \dots, N\}$ ergibt sich anstelle von (331.2) der Mittelwert μ_T aus

$$\mu_T = \frac{1}{2N+1} \sum_{n=-N}^N X(n) \quad (331.9)$$

und die Bedingung für die Ergodizität bezüglich des Erwartungswertes mit $E(X(n)) = \mu$ aus

$$\lim_{N \rightarrow \infty} \mu_T = \mu . \quad (331.10)$$

Der Erwartungswert von μ_T folgt mit

$$E(\mu_T) = \frac{1}{2N+1} E\left(\sum_{n=-N}^N X(n)\right) = \mu \quad (331.11)$$

und seine Varianz mit (313.12) zu

$$\begin{aligned} V(\mu_T) &= E((\mu_T - \mu)^2) \\ &= E\left(\frac{1}{(2N+1)^2} \sum_{n_1=-N}^N [X(n_1) - \mu] \sum_{n_2=-N}^N [X(n_2) - \mu]\right) \\ &= \frac{1}{(2N+1)^2} \sum_{n_1=-N}^N \sum_{n_2=-N}^N C(k) \quad \text{mit } k = n_2 - n_1. \end{aligned} \quad (331.12)$$

Zur Auswertung der Doppelsumme muß über ein Gitter mit $(2N+1)^2$ Gitterpunkten summiert werden. In den von links unten nach rechts oben verlaufenden Diagonalen gilt $k = \text{const.}$, wie in Abbildung 331-2 dargestellt ist und wie schon in Abbildung 218-2 gezeigt wurde. Die Summation über diese Diagonalen ergibt

$$\begin{aligned} V(\mu_T) &= \frac{1}{(2N+1)^2} \sum_{k=-2N}^{2N} C(k) (2N+1 - |k|) \\ &= \frac{1}{2N+1} \sum_{k=-2N}^{2N} C(k) \left(1 - \frac{|k|}{2N+1}\right). \end{aligned} \quad (331.13)$$

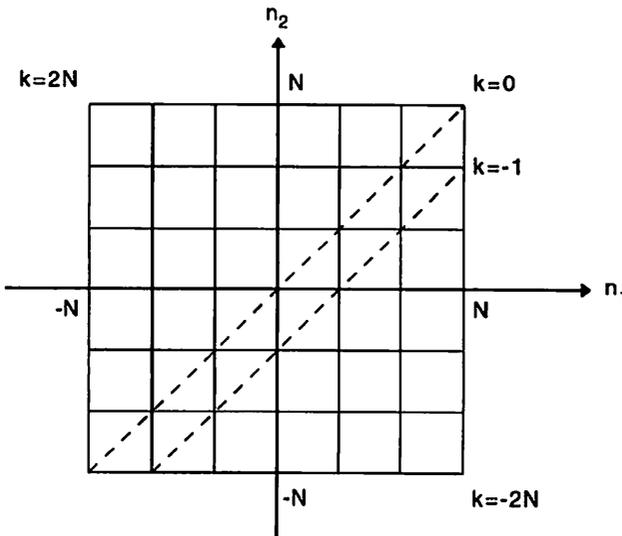


Abb. 331-2: Summation entlang der Diagonalen

Mit $N \rightarrow \infty$ in (331.13) geht die Varianz $V(\mu_T)$ gegen Null, und die Bedingung (331.10) für die Ergodizität ist erfüllt, falls die Autokovarianzfunktion $C(k)$ nach (216.7) absolut summierbar ist

$$\sum_{k=-\infty}^{\infty} |C(k)| < \infty . \quad (331.14)$$

Wie bereits im Zusammenhang von (331.7) erwähnt, bedeutet dies, daß die Autokovarianzfunktion $C(k)$ für große Werte von k gegen Null gehen muß. Die Realisierung des diskreten und stationären Prozesses muß also genügend lang im Vergleich zu dem Bereich sein, in dem die Autokovarianzfunktion $C(k)$ von Null verschiedene Werte annimmt.

Bei der Überprüfung der *Ergodizität* eines im engeren oder weiteren Sinne stationären Prozesses *bezüglich seiner Autokovarianzfunktion* geht man wie beim Erwartungswert vor. Zur Vereinfachung der Ableitungen gelte für den Prozeß $X(t)$

$$E(X(t)) = 0 .$$

Für die Autokovarianzfunktion folgt dann aus (313.6)

$$C(\tau) = E(X(t+\tau)X(t)) . \quad (331.15)$$

Der in dieser Definition zu bildende Mittelwert über das Ensemble wird durch den Mittelwert über die Zeit ersetzt, wobei die Integration auf das Intervall $(-T, T)$ beschränkt bleibt. Der Mittelwert $C_T(\tau)$ folgt dann aus

$$C_T(\tau) = \frac{1}{2T} \int_{-T}^T X(t+\tau)X(t) dt . \quad (331.16)$$

Ergodizität bezüglich der Autokovarianzfunktion ist dann gegeben, wenn mit einer Wahrscheinlichkeit von Eins die Zufallsvariable $C_T(\tau)$ den Wert $C(\tau)$ für $T \rightarrow \infty$ annimmt, folglich

$$\lim_{T \rightarrow \infty} C_T(\tau) = C(\tau) . \quad (331.17)$$

Der Erwartungswert von $C_T(\tau)$ berechnet sich mit (331.15) zu

$$E(C_T(\tau)) = \frac{1}{2T} \int_{-T}^T E(X(t+\tau)X(t)) dt = C(\tau) . \quad (331.18)$$

Seine Varianz folgt daher mit

$$\begin{aligned} V(C_T(\tau)) &= E([C_T(\tau) - C(\tau)]^2) \\ &= \frac{1}{4T^2} \int_{-T}^T \int_{-T}^T E([X(t_1+\tau)X(t_1) - C(\tau)][X(t_2+\tau)X(t_2) - C(\tau)]) dt_1 dt_2 . \end{aligned} \quad (331.19)$$

Da $X(t)$ stationär ist, gilt für den Erwartungswert

$$\begin{aligned} E([X(t_1+\tau)X(t_1)-C(\tau)][X(t_2+\tau)X(t_2)-C(\tau)]) \\ = C(X(t_1+\tau)X(t_1), X(t_2+\tau)X(t_2)) = C(X(t_2-t_1+\tau)X(t_2-t_1)) , \end{aligned} \quad (331.20)$$

denn der Ursprung der t -Achse kann an die Stelle t_1 verschoben werden. Nach Substitution von (331.20) in (331.19) läßt sich das Doppelintegral wie das in (331.5) auswerten.

Man erhält mit (331.6) und $t_2 - t_1 = \omega$

$$V(C_T(\tau)) = \frac{1}{2T} \int_{-2T}^{2T} C(X(\omega+\tau)X(\omega)) \left(1 - \frac{|\omega|}{2T}\right) d\omega . \quad (331.21)$$

Die in $C(X(\omega+\tau)X(\omega))$ auftretenden vierten Momente lassen sich durch zweite Momente ausdrücken, wenn $X(t)$ einen Gauß-Prozeß darstellt. Man erhält (Koch 1987, S.136)

$$\begin{aligned} C(X(\omega+\tau)X(\omega)) &= E([X(t+\omega+\tau)X(t+\omega)-C(\tau)][X(t+\tau)X(t)-C(\tau)]) \\ &= E(X(t+\omega+\tau)X(t+\omega)X(t+\tau)X(t)) + (C(\tau))^2 \\ &\quad - E(X(t+\omega+\tau)X(t+\omega))C(\tau) - E(X(t+\tau)X(t))C(\tau) \\ &= C(\tau)C(\tau) + C(\omega)C(\omega) + C(\omega+\tau)C(\omega-\tau) - (C(\tau))^2 \\ &= (C(\omega))^2 + C(\omega+\tau)C(\omega-\tau) . \end{aligned} \quad (331.22)$$

Nach Substitution dieses Ergebnisses in (331.21) ergibt sich schließlich

$$V(C_T(\tau)) = \frac{1}{2T} \int_{-2T}^{2T} [(C(\omega))^2 + C(\omega+\tau)C(\omega-\tau)] \left(1 - \frac{|\omega|}{2T}\right) d\omega . \quad (331.23)$$

Wenn die Autokovarianzfunktion $C(\tau)$ für große Werte von τ gegen Null geht, dann gilt

$$\lim_{T \rightarrow \infty} V(C_T(\tau)) = 0 , \quad (331.24)$$

und die Bedingung (331.17) für die Ergodizität bezüglich der Autokovarianzfunktion ist erfüllt.

Für einen diskreten stationären Prozeß $X(n)$ mit $n \in \{-N, -N+1, \dots, N\}$ und $E(X(n)) = 0$ erhält man den Mittelwert $C_T(k)$ seiner Autokovarianzfunktion $C(k)$ entsprechend (331.16) aus

$$C_T(k) = \frac{1}{2N+1} \sum_{n=-N}^N X(n+k)X(n) . \quad (331.25)$$

Mit der Definition (313.12) der Autokovarianzfunktion

$$C(k) = E(X(n+k)X(n)) \quad (331.26)$$

ist die Ergodizität bezüglich der Autokovarianzfunktion durch

$$\lim_{N \rightarrow \infty} C_T(k) = C(k) \quad (331.27)$$

gegeben. Der Erwartungswert des Mittels $C_T(k)$ berechnet sich aus

$$E(C_T(k)) = \frac{1}{2N+1} \sum_{n=-N}^N E(X(n+k)X(n)) = C(k) \quad (331.28)$$

und seine Varianz aus

$$\begin{aligned} V(C_T(k)) &= E([C_T(k) - C(k)]^2) \\ &= \frac{1}{(2N+1)^2} \sum_{n_1=-N}^N \sum_{n_2=-N}^N E([X(n_1+k)X(n_1) - C(k)][X(n_2+k)X(n_2) - C(k)]) \\ &= \frac{1}{(2N+1)^2} \sum_{n_1=-N}^N \sum_{n_2=-N}^N C(X(1+k)X(1)) \quad \text{mit } l = n_2 - n_1. \end{aligned} \quad (331.29)$$

Die Summation erfolgt wieder über die Diagonalen wie in (331.13) und die Berechnung der Autokovarianzfunktion unter der Annahme eines Gauß-Prozesses für $X(n)$ wie in (331.22). Man erhält

$$V(C_T(k)) = \frac{1}{2N+1} \sum_{l=-2N}^{2N} ((C(1))^2 + C(1+k)C(1-k)) \left(1 - \frac{|l|}{2N+1}\right). \quad (331.30)$$

Geht die Autokovarianzfunktion $C(k)$ für große Werte von k gegen Null, dann gilt

$$\lim_{N \rightarrow \infty} V(C_T(k)) = 0,$$

und die Bedingung (331.27) für die Ergodizität bezüglich der Autokovarianzfunktion ist erfüllt.

332 Schätzung des Erwartungswertes

Die Methode, die im Kapitel 331 diskutiert wurde, um den Erwartungswert eines stationären Prozesses zu schätzen, bestand darin, die für die Bildung des Erwartungswertes notwendigen Mittelwerte über das Ensemble durch die Mittelwerte über die Zeit zu ersetzen. Ist also $X(t)$ ein im engeren oder weiteren Sinne stationärer Prozeß, erhält man die Schätzung $\hat{\mu}$ seines Erwartungswertes μ nach (331.2) aus

$$\hat{\mu} = \frac{1}{2T} \int_{-T}^T X(t) dt \quad (332.1)$$

und die Varianz $V(\hat{\mu})$ der Schätzung aus (331.6). Die Schätzung ist erwartungstreu wegen (331.4). Ist der stationäre Prozeß ergodisch bezüglich seines Erwartungswertes, dann nimmt nach (331.3) für $T \rightarrow \infty$ die Schätzung $\hat{\mu}$ aus (332.1) den Erwartungswert μ an.

Für den diskreten stationären Prozeß $X(n)$ erhält man nach (331.9) die Schätzung $\hat{\mu}$ des Erwartungswertes μ aus

$$\hat{\mu} = \frac{1}{2N+1} \sum_{n=-N}^N X(n) \quad (332.2)$$

und seine Varianz $V(\hat{\mu})$ aus (331.13). Die Schätzung $\hat{\mu}$ ist wegen (331.11) erwartungstreu. Ist der Prozeß wieder ergodisch bezüglich seines Erwartungswertes, dann geht für $N \rightarrow \infty$ die Schätzung $\hat{\mu}$ gegen den Erwartungswert μ .

333 Schätzung der Autokovarianzfunktion

Ersetzt man wieder bei einem im engeren oder weiteren Sinne stationären Prozeß $X(t)$ mit dem Erwartungswert Null die Mittelwerte über das Ensemble durch die Mittelwerte über die Zeit, erhält man eine Schätzung der Autokovarianzfunktion aus (331.16). Nimmt man an, daß die Realisierung des stochastischen Prozesses $X(t)$ nur im Intervall $(-T, T)$ vorliegt, dann läßt sich die Autokovarianzfunktion $C(\tau)$ in Abhängigkeit von τ nur im Intervall $(-T, T-\tau)$ für $\tau > 0$ oder im Intervall $(-T-\tau, T)$ für $\tau < 0$ schätzen. Die Schätzung $\hat{C}(\tau)$ der Autokovarianzfunktion für $\tau > 0$ berechnet sich daher aus

$$\hat{C}(\tau) = \frac{1}{2T-\tau} \int_{-T}^{T-\tau} X(t+\tau)X(t) dt \quad \text{für } 0 < \tau < 2T. \quad (333.1)$$

Da mit $C(\tau) = C(-\tau)$ aus (313.8) die Autokovarianzfunktion eine gerade Funktion ist, so daß auch $\hat{C}(\tau) = \hat{C}(-\tau)$ gelten muß, werden Schätzungen für $\tau < 0$ im folgenden nicht betrachtet.

Mit (331.15) ergibt sich $\hat{C}(\tau)$ aus (333.1) als erwartungstreu Schätzung von $C(\tau)$

$$E(\hat{C}(\tau)) = C(\tau) \frac{1}{2T-\tau} \int_{-T}^{T-\tau} dt = C(\tau). \quad (333.2)$$

Für die Varianz $V(\hat{C}(\tau))$ der Schätzung $\hat{C}(\tau)$ erhält man mit den entsprechenden Integrationen, die zu (331.6) führen, aus (331.21)

$$V(\hat{C}(\tau)) = \frac{1}{2T-\tau} \int_{-2T+\tau}^{2T-\tau} C(X(\omega+\tau)X(\omega)) \left(1 - \frac{|\omega|}{2T-\tau}\right) d\omega. \quad (333.3)$$

Nimmt man schließlich für $X(t)$ einen Gauß-Prozeß an, ergibt sich mit (331.22)

$$V(\hat{C}(\tau)) = \frac{1}{2T-\tau} \int_{-2T+\tau}^{2T-\tau} [(C(\omega))^2 + C(\omega+\tau)C(\omega-\tau)] \left(1 - \frac{|\omega|}{2T-\tau}\right) d\omega . \quad (333.4)$$

Die Schätzung $\check{C}(\tau)$ der Autokovarianzfunktion $C(\tau)$ mit

$$\check{C}(\tau) = \frac{1}{2T} \int_{-T}^{T-\tau} X(t+\tau)X(t) dt \quad \text{für } 0 < \tau < 2T \quad (333.5)$$

soll ebenfalls betrachtet werden. Sie ist nicht erwartungstreu, denn mit (333.2) folgt

$$E(\check{C}(\tau)) = \left(1 - \frac{\tau}{2T}\right) C(\tau) . \quad (333.6)$$

Für die Varianz $V(\check{C}(\tau))$ der Schätzung $\check{C}(\tau)$ ergibt sich unter der Annahme eines Gauß-Prozesses anstelle von (333.4)

$$V(\check{C}(\tau)) = \frac{1}{2T} \int_{-2T+\tau}^{2T-\tau} [(C(\omega))^2 + C(\omega+\tau)C(\omega-\tau)] \left(1 - \frac{|\omega|+\tau}{2T}\right) d\omega . \quad (333.7)$$

Beim Vergleich der beiden Schätzungen $\hat{C}(\tau)$ aus (333.1) und $\check{C}(\tau)$ aus (333.5) läßt sich zunächst feststellen, daß für beide Schätzungen

$$\lim_{T \rightarrow \infty} \hat{C}(\tau) = C(\tau) \quad \text{und} \quad \lim_{T \rightarrow \infty} \check{C}(\tau) = C(\tau) \quad (333.8)$$

gilt, falls für $X(t)$ die Ergodizität bezüglich der Autokovarianzfunktion erfüllt ist. Geht nämlich die Autokovarianzfunktion $C(\tau)$ für große Werte von τ gegen Null, dann gehen für $T \rightarrow \infty$ auch die Varianzen $V(\hat{C}(\tau))$ aus (333.4) und $V(\check{C}(\tau))$ aus (333.7) gegen Null.

Die erwartungstreue Schätzung $\hat{C}(\tau)$ besitzt den Nachteil, daß wegen $2T - \tau$ im Nenner von (333.3) oder (333.4) ihre Varianz für Werte von τ , die sich $2T$ annähern, sehr groß wird. An den Enden der Funktion $\hat{C}(\tau)$ muß also mit einem sehr unregelmäßigen Verhalten gerechnet werden. Nach (333.7) besitzt die verzerrte Schätzung $\check{C}(\tau)$ diese Eigenschaft nicht. Doch beträgt die Verzerrung, im Englischen "bias", dieser Schätzung nach (333.6)

$$E(\check{C}(\tau)) - C(\tau) = -\frac{\tau}{2T} C(\tau) , \quad (333.9)$$

die für Werte von τ , die in der Nähe von $2T$ liegen, betragsmäßig maximal wird. Da man davon ausgehen kann, daß $C(\tau)$ für große Werte von τ gegen Null geht und auch T entsprechend groß ist, ist der Einfluß der Verzerrung für $\check{C}(\tau)$ weniger gravierend als der der großen Varianz für $\hat{C}(\tau)$. Aus diesen Gründen wird die nicht erwartungstreue Schätzung $\check{C}(\tau)$ der Autokovarianzfunktion $C(\tau)$ gegenüber der erwartungstreuen Schätzung $\hat{C}(\tau)$ bevorzugt.

Die erwartungstreue Schätzung $\hat{C}(n)$ der Autokovarianzfunktion $C(n)$ eines diskreten Prozesses $X(n)$, der im engeren oder weiteren Sinne stationär ist und für den $E(X(n))=0$ angenommen wird, folgt mit (331.25) in Analogie zu (333.1) aus

$$\hat{C}(k) = \frac{1}{2N-k+1} \sum_{n=-N}^{N-k} X(n+k)X(n) \quad \text{für } k \in \{0, 1, \dots, 2N\} , \quad (333.10)$$

denn wegen (331.26) gilt

$$E(\hat{C}(k)) = C(k) . \quad (333.11)$$

Die Varianz $V(\hat{C}(k))$ der Schätzung $\hat{C}(k)$ erhält man, falls $X(n)$ einen Gauß-Prozeß darstellt, (331.30) entsprechend aus

$$V(\hat{C}(k)) = \frac{1}{2N-k+1} \sum_{l=-2N+k}^{2N-k} ((C(l))^2 + C(1+k)C(1-k)) \left(1 - \frac{|l|}{2N-k+1}\right) . \quad (333.12)$$

Die nicht erwartungstreue Schätzung $\check{C}(k)$ der Autokovarianzfunktion $C(k)$, die (333.5) entspricht, berechnet sich aus

$$\check{C}(k) = \frac{1}{2N+1} \sum_{n=-N}^{N-k} X(n+k)X(n) \quad \text{für } k \in \{0, 1, \dots, 2N\} \quad (333.13)$$

mit

$$E(\check{C}(k)) = \left(1 - \frac{k}{2N+1}\right) C(k) \quad (333.14)$$

wegen (331.26). Die Varianz $V(\check{C}(n))$ der Schätzung $\check{C}(n)$ folgt, falls $X(n)$ ein Gauß-Prozeß ist, (331.30) entsprechend aus

$$V(\check{C}(k)) = \frac{1}{2N+1} \sum_{l=-2N+k}^{2N-k} ((C(l))^2 + C(1+k)C(1-k)) \left(1 - \frac{|k+l|}{2N+1}\right) . \quad (333.15)$$

Für den Vergleich der erwartungstreuen Schätzung $\hat{C}(k)$ und der nicht erwartungstreuen Schätzung $\check{C}(k)$ der Autokovarianzfunktion $C(k)$ gilt ganz analog das, was zum Vergleich von $\hat{C}(\tau)$ und $\check{C}(\tau)$ gesagt wurde. Die Schätzung $\check{C}(k)$ ist daher gegenüber der Schätzung $\hat{C}(k)$ zu bevorzugen.

Die Schätzung (333.10) oder die Schätzung (333.13) lassen sich mit Hilfe der schnellen Fourier-Transformation als schnelle Faltung berechnen. Substituiert man nämlich $n = 1-N$ und $l = N+n$, erhält man anstelle von (333.13)

$$\check{C}(k) = \frac{1}{2N+1} \sum_{l=0}^{2N-k} X(1-N+k)X(1-N) .$$

Setzt man weiter $h(-m) = X(m-N)$, ergibt sich

$$h(-k-1) = X(1-N+k) .$$

Mit $x(l) = X(1-N)$ und $-k$ anstelle von k folgt schließlich die Schätzung

$$\tilde{C}(-k) = \frac{1}{2N+1} \sum_{l=0}^{2N+k} x(l)h(k-1) . \quad (333.16)$$

Die Summe auf der rechten Seite entspricht der linearen Faltung (218.16). Sie läßt sich, wie im Zusammenhang mit (218.18) erläutert, als schnelle Faltung berechnen. Da die Autokovarianzfunktion mit $C(k) = C(-k)$ aus (313.13) eine gerade Funktion ist, darf aus (333.16) $\tilde{C}(-k)$ anstelle von $\tilde{C}(k)$ ermittelt werden, da $\tilde{C}(k) = \tilde{C}(-k)$ gelten muß.

Wie bereits im Kapitel 313 erwähnt, kann man die Schätzwerte der Autokovarianzfunktion vorgegebenen Autokovarianzfunktionen anpassen.

Die Schätzung der mit (313.10) definierten Kreuzkovarianzfunktion zweier stationärer Prozesse verläuft ganz analog zur Schätzung der Autokovarianzfunktion.

34 Schätzungen des Spektrums

341 Periodogramm

Das Spektrum $S(\omega)$ eines stationären Prozesses $X(t)$ ergibt sich nach (314.1) als Fourier-Transformation der Autokovarianzfunktion $C(\tau)$ des Prozesses. Um $S(\omega)$ zu schätzen, ist es daher naheliegend, die Schätzung von $C(\tau)$ einer Fourier-Transformation zu unterwerfen. Als geeignete Schätzung für $C(\tau)$ hatte sich $\tilde{C}(\tau)$ aus (333.5) herausgestellt. Mit ihr soll im folgenden gearbeitet werden.

Mit $\tilde{C}(\tau) \neq 0$ für $-2T < \tau < 2T$ aus (333.5) erhält man eine Schätzung $\tilde{S}(\omega)$ von $S(\omega)$ in Analogie zu (314.1) mit der Fourier-Transformation

$$\tilde{S}(\omega) = \int_{-2T}^{2T} \tilde{C}(\tau) e^{-j\omega\tau} d\tau \quad (341.1)$$

oder wegen (314.16), da $\tilde{C}(\tau) = \tilde{C}(-\tau)$ gilt,

$$\tilde{S}(\omega) = \int_{-2T}^{2T} \tilde{C}(\tau) \cos\omega\tau d\tau \quad (341.2)$$

Dieser Schätzung entspricht das Periodogramm für diskrete Prozesse, das in (341.8) eingeführt wird. Der Erwartungswert von $\tilde{S}(\omega)$ berechnet sich mit (333.6) aus

$$E(\tilde{S}(\omega)) = \int_{-2T}^{2T} \left(1 - \frac{|\tau|}{2T}\right) C(\tau) e^{-j\omega\tau} d\tau \quad (341.3)$$

Geht die Autokovarianzfunktion $C(\tau)$ für große Werte von τ gegen Null, dann gilt mit (314.1)

$$\lim_{T \rightarrow \infty} E(\tilde{S}(\omega)) = S(\omega) \quad (341.4)$$

Die Schätzung $\tilde{S}(\omega)$ des Spektrums $S(\omega)$ läßt sich auch direkt aus der Realisierung des stochastischen Prozesses $X(t)$ berechnen, denn es gilt

$$\tilde{S}(\omega) = \frac{1}{2T} \left| \int_{-T}^T X(t) e^{-j\omega t} dt \right|^2 \quad (341.5)$$

Um diese Beziehung zu zeigen, wird mit (213.14) die rechte Seite berechnet. Man erhält

$$\frac{1}{2T} \int_{-T}^T X(t_1) e^{j\omega t_1} dt_1 \int_{-T}^T X(t_2) e^{-j\omega t_2} dt_2 \quad .$$

Substituiert man $t_2 - t_1 = y$, $t_2 = y + t_1$ und $t_1 = x$, ergibt sich mit den gleichen Überlegungen, die zu (331.6) führen,

$$\int_{-2T}^0 \left[\frac{1}{2T} \int_{-T-y}^T X(x)X(y+x) dx \right] e^{-j\omega y} dy + \int_0^{2T} \left[\frac{1}{2T} \int_{-T}^{T-y} X(x)X(y+x) dx \right] e^{-j\omega y} dy .$$

Hieraus folgt dann mit (333.5)

$$\int_{-2T}^{2T} \tilde{C}(y) e^{-j\omega y} dy = \mathfrak{S}(\omega) ,$$

so daß (341.5) bewiesen ist.

Aus (341.5) ist ersichtlich, daß $\mathfrak{S}(\omega) \geq 0$ gilt. Daher folgt wegen (341.4) auch

$$E(\mathfrak{S}(\omega)) \geq 0 \quad \text{und} \quad S(\omega) \geq 0 . \quad (341.6)$$

Für einen diskreten stationären Prozeß $X(n)$ hatte sich die Schätzung $\tilde{C}(k)$ aus (333.13) der Autokovarianzfunktion $C(k)$ als geeignet herausgestellt. Wegen $\tilde{C}(k) \neq 0$ für $k \in \{-2N, -2N+1, \dots, 2N\}$ ergibt sich eine Schätzung $\mathfrak{S}(\Omega)$ des Spektrums $S(\Omega)$ in Analogie zu (314.25) aus

$$\mathfrak{S}(\Omega) = \sum_{k=-2N}^{2N} \tilde{C}(k) e^{-jk\Omega} \quad (341.7)$$

oder wegen (314.24)

$$\mathfrak{S}(\Omega) = \sum_{k=-2N}^{2N} \tilde{C}(k) \cos k\Omega . \quad (341.8)$$

Die Schätzung $\mathfrak{S}(\Omega)$ des Spektrums $S(\Omega)$ bezeichnet man als *Periodogramm*, da mit ihm Periodizitäten im Prozeß $X(n)$ aufzudecken sind. Das Spektrum gibt, wie bereits zu Beginn des Kapitels 314 erläutert, Auskunft über den Frequenzinhalt eines Prozesses.

Der Erwartungswert der Schätzung $\mathfrak{S}(\Omega)$ berechnet sich mit (333.14) zu

$$E(\mathfrak{S}(\Omega)) = \sum_{k=-2N}^{2N} \left(1 - \frac{|k|}{2N+1}\right) C(k) e^{-jk\Omega} . \quad (341.9)$$

Geht die Autokovarianzfunktion $C(k)$ für große Werte von k gegen Null, erhält man mit (314.25)

$$\lim_{N \rightarrow \infty} E(\mathfrak{S}(\Omega)) = S(\Omega) . \quad (341.10)$$

Die Schätzung $\mathfrak{S}(\Omega)$ läßt sich direkt aus der Realisierung des diskreten Prozesses $X(n)$ berechnen, denn es gilt

$$\mathfrak{S}(\Omega) = \frac{1}{2N+1} \left| \sum_{k=-N}^N X(k) e^{-jk\Omega} \right|^2 . \quad (341.11)$$

Um dieses Ergebnis zu beweisen, wird die rechte Seite mit (213.14) berechnet. Man erhält

$$\begin{aligned} \frac{1}{2N+1} \left[\sum_{n_1=-N}^N X(n_1) e^{jn_1\Omega} \right] \left[\sum_{n_2=-N}^N X(n_2) e^{-jn_2\Omega} \right] \\ = \frac{1}{2N+1} \sum_{n_1=-N}^N \sum_{n_2=-N}^N X(n_1) X(n_2) e^{-j(n_2-n_1)\Omega} . \end{aligned}$$

Setzt man $k = n_2 - n_1$ und $n_2 = n_1 + k$, so gilt $k = \text{const.}$ entlang der Diagonalen des Gitters mit $(2N+1)^2$ Gitterpunkten, über die zu summieren ist. Summiert man entlang der Diagonalen, erhält man mit (333.13) für $k > 0$ ganz entsprechend wie für (331.13)

$$\sum_{k=-2N}^{2N} \frac{1}{2N+1} \sum_{n_1=-N}^{N-k} X(n_1) X(n_1+k) e^{-jk\Omega} = \sum_{k=-2N}^{2N} \check{C}(k) e^{-jk\Omega} = \mathfrak{S}(\Omega) ,$$

womit (341.11) bewiesen ist.

342 Geglättete Schätzung

Obwohl die Schätzung $\mathfrak{S}(\omega)$ des Spektrums $S(\omega)$ eines stationären Prozesses $X(t)$ und die Schätzung $\mathfrak{S}(\Omega)$ des Spektrums $S(\Omega)$ eines diskreten stationären Prozesses $X(n)$ die Eigenschaft besitzen, daß sie mit wachsender Länge T der Integration und mit wachsender Anzahl N der Summanden gegen $S(\omega)$ und $S(\Omega)$ streben, wie mit (341.4) und (341.10) gezeigt wurde, stellen sie keine geeigneten Schätzungen dar. Die Ursache liegt darin, daß unabhängig von T und unabhängig von N

$$V(\mathfrak{S}(\omega)) \approx [E(\mathfrak{S}(\omega))]^2 \quad \text{und} \quad V(\mathfrak{S}(\Omega)) \approx [E(\mathfrak{S}(\Omega))]^2 \quad (342.1)$$

gilt (Papoulis 1977, S.380; Priestley 1981, S.426). Die Schätzwerte $\mathfrak{S}(\omega)$ und $\mathfrak{S}(\Omega)$ verhalten sich also sehr unruhig, und ihre Varianzen lassen sich durch Vergrößerung von T und N nicht verringern. Dies ist dadurch zu erklären, daß zur Berechnung der Schätzung $\mathfrak{S}(\omega)$ aus (341.1) über sämtliche Schätzungen $\check{C}(\tau)$ zu integrieren ist und daß für $\mathfrak{S}(\Omega)$ aus (341.7) über die Schätzungen $\check{C}(k)$ für sämtliche Werte von k zu summieren ist. Die Varianzen $V(\mathfrak{S}(\omega))$ und $V(\mathfrak{S}(\Omega))$ ergeben sich dann aus der Summe der Varianzen sämtlicher Schätzungen $\check{C}(\tau)$ und $\check{C}(k)$, falls Unabhängigkeit der Schätzungen angenommen wird (Koch 1987, S.118).

Die einfachste Art, die Varianzen $V(\mathfrak{S}(\omega))$ und $V(\mathfrak{S}(\Omega))$ zu reduzieren, besteht darin,

$\tilde{C}(\tau)$ in (341.1) für bestimmte Wertebereiche von τ und $\tilde{C}(k)$ in (341.7) für bestimmte Werte von k zu vernachlässigen. Da bei der Schätzung der Autokovarianzfunktion vorausgesetzt wurde, daß $C(\tau)$ und $C(k)$ für große Werte von τ und k gegen Null gehen, ist es zweckmäßig, die Schätzungen $\tilde{C}(\tau)$ und $\tilde{C}(k)$ an ihren Enden, also für große Werte von τ und k zu vernachlässigen. Noch besser ist es, mit Funktionen $w(\tau)$ und $w(k)$ die Schätzungen $\tilde{C}(\tau)$ und $\tilde{C}(k)$ für große Werte von τ und k geringer zu gewichten als für kleine Werte von τ und k . Man bezeichnet $w(\tau)$ und $w(k)$ als Fensterfunktionen in Übereinstimmung mit der Terminologie des Kapitels 227, wo diese Fensterfunktionen dazu dienten, Impulsantworten von Filtern zu modifizieren.

Mit der Gewichtung von $\tilde{C}(\tau)$ in (341.1) durch die Fensterfunktion $w(\tau)$ erhält man anstelle von $\tilde{S}(\omega)$ die Schätzung $\tilde{S}_w(\omega)$ des Spektrums $S(\omega)$

$$\tilde{S}_w(\omega) = \int_{-2T}^{2T} \tilde{C}(\tau) w(\tau) e^{-j\omega\tau} d\tau . \quad (342.2)$$

Dieser Gewichtung im Zeitbereich entspricht im Frequenzbereich die Faltung, denn durch die inverse Fourier-Transformation (214.2) erhält man anstelle von (341.1) mit der Integrationsvariablen α

$$\tilde{C}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{S}(\alpha) e^{j\alpha\tau} d\alpha .$$

Einsetzen in (342.2) ergibt

$$\tilde{S}_w(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{S}(\alpha) \int_{-2T}^{2T} w(\tau) e^{-j\tau(\omega-\alpha)} d\tau d\alpha .$$

Führt man die Fensterfunktion $w(\tau)$ als gerade und reelle Funktion ein, ist ihre Fourier-Transformation

$$W(\beta) = \int_{-2T}^{2T} w(\tau) e^{-j\tau\beta} d\tau \quad (342.3)$$

nach (214.7) ebenfalls gerade und reell. Sie wird nach (227.14) als Spektralfenster bezeichnet. Die Schätzung $\tilde{S}_w(\omega)$ ergibt sich dann aus der Faltung der Schätzung $\tilde{S}(\alpha)$ mit dem Spektralfenster $W(\omega-\alpha)$, also

$$\tilde{S}_w(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{S}(\alpha) W(\omega-\alpha) d\alpha . \quad (342.4)$$

Die Schätzung $\tilde{S}_w(\omega)$ folgt also aus einer Glättung der Schätzung $\tilde{S}(\alpha)$, wobei als Gewichtsfunktion das Spektralfenster $W(\omega-\alpha)$ dient.

Für sehr große Werte von T ergibt sich der Erwartungswert der geglätteten Schätzung

$\hat{S}_w(\omega)$ mit (341.4) aus (342.4) zu

$$E(\hat{S}_w(\omega)) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\alpha) W(\omega - \alpha) d\alpha . \quad (342.5)$$

Betrachtet man also das Spektrum $S(\alpha)$ durch das Spektralfenster $W(\omega - \alpha)$, erhält man den Erwartungswert der geglätteten Schätzung $\hat{S}_w(\omega)$. Ist das Spektralfenster $W(\omega - \alpha)$ identisch mit der Deltafunktion (211.5)

$$W(\omega - \alpha) = \delta(\omega - \alpha) , \quad (342.6)$$

folgt aus (342.5) für die Verzerrung der Schätzung $\hat{S}_w(\omega)$ der Wert Null

$$E(\hat{S}_w(\omega)) - S(\omega) = 0 . \quad (342.7)$$

In diesem Fall tritt jedoch keine Glättung auf, so daß sich die Varianz der Schätzung aus (342.1) ergibt.

Das Spektralfenster $W(\beta)$ aus (342.3) muß also eine bestimmte Öffnung besitzen, damit die Varianz reduziert wird. Als Folge tritt eine Verzerrung der Schätzung auf. Da die Spektralfenster, wie im folgenden erläutert wird, nicht nur einen Hauptgipfel, sondern auch Nebengipfel besitzen, allerdings von wenigen Prozent der Höhe des Hauptgipfels, werden zusätzliche Verzerrungen durch das Durchsickern des Einflusses von $S(\omega)$ an Frequenzstellen verursacht, die entfernt von der Frequenzstelle liegen, für die die Schätzung erfolgt. Im Englischen wird dieser Effekt "leakage" bezeichnet.

Mit der Öffnung des Spektralfensters vermindert sich die Varianz, die Verzerrung aber steigt an. Gleichzeitig lassen sich daher Varianz und Verzerrung nicht minimieren. Um zu einer *optimalen Fensterfunktion* zu kommen, muß man entweder die Varianz vorgeben und die Verzerrung minimieren oder die Verzerrung vorgeben und die Varianz minimieren (Schlittgen und Streitberg 1991, S.315). Hierauf wird im folgenden nicht eingegangen, es wird lediglich neben dem einfachen Rechteckfenster das häufig verwendete Bartlett-Fenster vorgestellt.

Die einfachste Fensterfunktion $w(\tau)$ erhält man mit dem Rechteckfenster

$$w(\tau) = \begin{cases} 1 & \text{für } |\tau| \leq c \\ 0 & \text{für } |\tau| > c \end{cases} \quad (342.8)$$

und $c < 2T$. Das zugehörige Spektralfenster berechnet sich mit (214.5) aus (342.3) zu

$$W(\beta) = \frac{2}{\beta} \sin\beta c . \quad (342.9)$$

Dieses Spektralfenster besitzt seinen Hauptgipfel bei $\beta = 0$ mit

$$W(\beta=0) = \lim_{\beta \rightarrow 0} \frac{2}{\beta} \sin\beta c = \lim_{\beta \rightarrow 0} 2c \cos\beta c = 2c$$

und weitere Nebengipfel zwischen den Nulldurchgängen bei $\beta = \pm n\pi/c$ mit $n \in \mathbb{N}$. Da das Spektralfenster (342.9) auch negative Werte annehmen kann, können nach (342.4) auch negative Schätzwerte $\hat{S}_w(\omega)$ erhalten werden. Da das Spektrum $S(\omega)$ wegen (314.17) eine nicht negative Funktion ist, führt das Rechteckfenster (342.8) auf keine geeignete Schätzung $\hat{S}_w(\omega)$ des Spektrums.

Das Dreieckfenster

$$w(\tau) = \begin{cases} 1 - \frac{|\tau|}{c} & \text{für } |\tau| \leq c \\ 0 & \text{für } |\tau| > c \end{cases} \quad (342.10)$$

und $c < 2T$ wurde von Bartlett (1955, S.274) für die geglättete Schätzung des Spektrums eingeführt. Das zugehörige Spektralfenster besitzt wegen (214.14) aufgrund von (342.3) die Form

$$W(\beta) = \frac{4}{\beta^2 c} \sin^2 \frac{\beta c}{2} . \quad (342.11)$$

Dieses Fenster besitzt seinen Hauptgipfel bei $\beta = 0$ mit

$$W(\beta=0) = \left(\lim_{\beta \rightarrow 0} \frac{2}{\beta \sqrt{c}} \sin \frac{\beta c}{2} \right)^2 = c \quad (342.12)$$

und Nebengipfel zwischen den Nullstellen bei $\beta = \pm 2n\pi/c$ mit $n \in \mathbb{N}$. Für das Spektralfenster gilt $W(\beta) \geq 0$, so daß mit (342.4) Schätzwerte $\hat{S}_w(\omega) \geq 0$ des Spektrums $S(\omega)$ erhalten werden. Da $S(\omega)$ nach (314.17) eine nicht negative Funktion ist, führt das Spektralfenster (342.11) daher auf eine geeignete Schätzung $\hat{S}_w(\omega)$ des Spektrums.

Weitere häufig verwendete Fensterfunktionen sind die von Tukey und Parzen (Jenkins und Watts 1968, S.244).

Liegt ein diskreter stationärer Prozeß $X(n)$ vor, wird ebenfalls die Schätzung $\hat{C}(n)$ der Autokovarianzfunktion durch eine Fensterfunktion gewichtet, und man erhält anstelle des Periodogramms $\hat{S}(\Omega)$ aus (341.7) die geglättete Schätzung $\hat{S}_w(\Omega)$ des Spektrums $S(\Omega)$ zu

$$\hat{S}_w(\Omega) = \sum_{k=-2N}^{2N} \hat{C}(k) w(k) e^{-jk\Omega} . \quad (342.13)$$

Dieser Gewichtung im Zeitbereich entspricht im Frequenzbereich die Faltung, denn mit der inversen Fourier-Transformation von (341.7) erhält man (314.25) entsprechend

$$\hat{C}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{S}(\lambda) e^{jk\lambda} d\lambda .$$

Durch Einsetzen in (342.13) folgt

$$\mathfrak{S}_w(\Omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathfrak{S}(\lambda) \left[\sum_{k=-2N}^{2N} w(k) e^{-jk(\Omega-\lambda)} \right] d\lambda .$$

Die Fourier-Transformation der Fensterfunktion ergibt nach (227.14) das Spektralfenster

$$W(\gamma) = \sum_{k=-2N}^{2N} w(k) e^{-jk\gamma} , \quad (342.14)$$

so daß die geglättete Schätzung $\mathfrak{S}_w(\Omega)$ als Faltung des Periodogramms $\mathfrak{S}(\lambda)$ mit dem Spektralfenster $W(\Omega-\lambda)$ nach (227.19) folgt

$$\mathfrak{S}_w(\Omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathfrak{S}(\lambda) W(\Omega-\lambda) d\lambda . \quad (342.15)$$

Für große Werte von N folgt der Erwartungswert der geglätteten Schätzung $\mathfrak{S}_w(\Omega)$ wegen (341.10) aus

$$E(\mathfrak{S}_w(\Omega)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\lambda) W(\Omega-\lambda) d\lambda . \quad (342.16)$$

Für das Spektralfenster $W(\gamma)$ gilt also wieder das, was bereits im Zusammenhang mit (342.6) und (342.7) gesagt wurde. Besitzt das Spektralfenster eine schmale Öffnung, ist die Verzerrung der geglätteten Schätzung klein und die Varianz groß, öffnet man das Spektralfenster, wächst die Verzerrung, die Varianz nimmt aber ab.

Die einfachste Fensterfunktion ist wieder das Rechteckfenster mit

$$w(k) = \begin{cases} 1 & \text{für } k \in \{-2L, -2L+1, \dots, 2L\} \\ 0 & \text{sonst} \end{cases} \quad (342.17)$$

und $L < N$. Das zugehörige Spektralfenster erhält man mit (218.21) aus (342.14)

$$W(\gamma) = \frac{\sin(4L+1)\gamma/2}{\sin\gamma/2} . \quad (342.18)$$

Der Hauptgipfel des Spektralfensters liegt bei $\gamma = 0$ mit

$$W(\gamma=0) = \lim_{\gamma \rightarrow 0} (4L+1) \frac{\cos(4L+1)\gamma/2}{\cos\gamma/2} = 4L+1$$

und weitere Nebengipfel zwischen den Nulldurchgängen bei $\gamma = \pm 2n\pi/(4L+1)$ mit $n \in \mathbb{N}$. Da das Spektralfenster auch negative Werte annimmt, können aus (342.15) auch negative Schätzwerte $\mathfrak{S}_w(\Omega)$ berechnet werden, was wegen $S(\Omega) \geq 0$ unerwünscht ist, so daß die Fensterfunktion (342.17) auf keine geeignete Schätzung $\mathfrak{S}_w(\Omega)$ führt.

Das von Bartlett (1955, S.274) eingeführte Dreieckfenster

$$w(k) = \begin{cases} 1 - \frac{|k|}{2L+1} & \text{für } k \in \{-2L, -2L+1, \dots, 2L\} \\ 0 & \text{sonst} \end{cases} \quad (342.19)$$

für $L < N$ führt mit (218.24) und (342.14) auf das Spektralfenster, den sogenannten Fejer-Kern,

$$W(\gamma) = \frac{1}{2L+1} \left[\frac{\sin(2L+1)\gamma/2}{\sin\gamma/2} \right]^2. \quad (342.20)$$

Es besitzt seinen Hauptgipfel bei $\gamma = 0$ mit

$$W(\gamma=0) = \frac{1}{2L+1} \left[\lim_{\gamma \rightarrow 0} (2L+1) \frac{\cos(2L+1)\gamma/2}{\cos\gamma/2} \right]^2 = 2L+1$$

und Nebengipfel zwischen den Nullstellen bei $\gamma = \pm 2n\pi/(2L+1)$ mit $n \in \mathbb{N}$. Der erste und höchste Nebengipfel besitzt etwa 4% der Höhe des Hauptgipfels. Für das Spektralfenster gilt $W(\gamma) \geq 0$, so daß für die Schätzwerte $\mathfrak{S}_w(\Omega) \geq 0$ erhalten wird und damit eine geeignete Schätzung vorliegt.

Weitere häufig benutzte Fensterfunktionen sind die von Tukey und Parzen (Priestley 1981, S.442).

Für die Berechnung der geglätteten Schätzung $\mathfrak{S}_w(\Omega)$ aus (342.13) des Spektrums $S(\Omega)$ eines diskreten stationären Prozesses $X(n)$ wird man die schnelle Fourier-Transformation einsetzen, falls dadurch Rechenzeit gespart werden kann, so daß (342.13) als diskrete Fourier-Transformation zu interpretieren ist. Die Schätzung $\mathfrak{S}_w(\Omega)$ wird sich dann nicht kontinuierlich, sondern an den diskreten Stellen

$$\Omega = \frac{2n\pi}{4N+1} \quad \text{mit } n \in \{-2N, -2N+1, \dots, 2N\} \quad (342.21)$$

ergeben, wie aus dem im Zusammenhang von (218.10) und (218.11) Gesagten folgt.

Bei einer Berechnung der Schätzung $\mathfrak{S}_w(\Omega)$ an diskreten Stellen braucht auch bei einer Anwendung von (342.15) das Integral nicht ausgewertet zu werden. Faßt man nämlich (342.13) und auch (341.7) als diskrete Fourier-Transformation auf, so ergibt die inverse diskrete Fourier-Transformation von (341.7)

$$\check{C}(k) = \frac{1}{4N+1} \sum_{\lambda=-2N}^{2N} \mathfrak{S}(\lambda) e^{jk\lambda} \quad \text{mit } \lambda = \frac{2l\pi}{4N+1}, \quad (342.22)$$

wie aus dem im Zusammenhang von (218.10) und (218.11) Erwähnten folgt. Die Substitution in (342.13) ergibt mit (342.14)

$$\mathfrak{S}_w(\Omega) = \frac{1}{4N+1} \sum_{l=-2N}^{2N} \mathfrak{S}(\lambda) \sum_{k=-2N}^{2N} w(k) e^{-jk(\Omega-\lambda)} = \frac{1}{4N+1} \sum_{l=-2N}^{2N} \mathfrak{S}(\lambda) W(\Omega-\lambda) \quad (342.23)$$

mit

$$\Omega = \frac{2n\pi}{4N+1} \quad \text{und} \quad \lambda = \frac{2l\pi}{4N+1} \quad \text{für} \quad l, n \in \{-2N, -2N+1, \dots, 2N\} .$$

Hieraus folgt die Schätzung $\mathfrak{S}_w(\Omega)$ an diskreten Stellen für Ω .

Abschließend soll noch ein einfaches Verfahren behandelt werden, die geglättete Schätzung $\mathfrak{S}_w(\Omega)$ eines diskreten stationären Prozesses $X(n)$ mit $n \in \{-N, -N+1, \dots, N\}$ an den diskreten Stellen (342.21) zu berechnen. Es besitzt den Vorzug, daß es sich unmittelbar auf Zufallsfelder verallgemeinern läßt. Für die Autokovarianzfunktion gelte

$$C(k) \approx 0 \quad \text{für} \quad k > 2Q \quad \text{mit} \quad Q \ll N . \quad (342.24)$$

Die Folge der Werte für $X(n)$ mit der Länge $2N+1$ werde nun in M Stücke der Länge $4Q+1$ aufgeteilt, also

$$2N+1 = M(4Q+1) . \quad (342.25)$$

Berechnet man in diesen M Stücken der Länge $4Q+1$ jeweils $\check{C}^{(i)}(k)$ des Stückes i nach (333.13) und das Periodogramm $\mathfrak{S}^{(i)}(\Omega)$ des Stückes i nach (341.7), erhält man

$$\mathfrak{S}^{(i)}(\Omega) = \sum_{k=-2Q}^{2Q} \check{C}^{(i)}(k) e^{-jk\Omega} \quad \text{für} \quad i \in \{1, \dots, M\} , \quad (342.26)$$

falls in den Stücken der Werte für $X(n)$ jeweils von $-2Q$ bis $2Q$ gezählt wird. Wegen der Voraussetzung (342.24) kann man annehmen, daß das Periodogramm $\mathfrak{S}^{(i)}(\Omega)$ unabhängig von $\mathfrak{S}^{(j)}(\Omega)$ für $i \neq j$ ist. Die geglättete Schätzung $\mathfrak{S}_w(\Omega)$ des Spektrums ergibt sich daher einfach durch Mittelbildung aus

$$\mathfrak{S}_w(\Omega) = \frac{1}{M} \sum_{i=1}^M \mathfrak{S}^{(i)}(\Omega) \quad \text{für} \quad \Omega = \frac{2n\pi}{4N+1} \quad \text{und} \quad n \in \{-2N, -2N+1, \dots, 2N\} . \quad (342.27)$$

Um die Eigenschaft dieses Schätzers zu untersuchen, wird der Erwartungswert gebildet

$$E(\mathfrak{S}_w(\Omega)) = \frac{1}{M} \sum_{i=1}^M E(\mathfrak{S}^{(i)}(\Omega)) = E(\mathfrak{S}^{(i)}(\Omega)) .$$

Mit (341.9) und (342.26) folgt weiter

$$E(\mathfrak{S}_w(\Omega)) = \sum_{k=-2Q}^{2Q} \left(1 - \frac{|k|}{2Q+1}\right) C(k) e^{-jk\Omega} . \quad (342.28)$$

Im Vergleich hierzu erhält man mit (333.14) den Erwartungswert von $\mathfrak{S}_w(\Omega)$ aus

(342.13), falls für $w(k)$ das Dreieckfenster (342.19) substituiert wird,

$$E(\mathfrak{S}_w(\Omega)) = \sum_{k=-2L}^{2L} \left(1 - \frac{|k|}{2N+1}\right) \left(1 - \frac{|k|}{2Q+1}\right) C(k) e^{-jk\Omega} = \sum_{k=-2Q}^{2Q} \left(1 - \frac{|k|}{2Q+1}\right) C(k) e^{-jk\Omega},$$

denn als Voraussetzung gilt $Q \ll N$ nach (342.24). Die geglättete Schätzung nach (342.27) und die geglättete Schätzung mit Hilfe des Dreieckfensters (342.19) entsprechen sich also, da sie identische Erwartungswerte besitzen.

343 Numerische Berechnung der geglätteten Schätzung

Im folgenden sollen Methoden erläutert werden, die geglättete Schätzung des Spektrums eines stationären Prozesses zu berechnen. Da die numerischen Verfahren besonders für die diskreten Prozesse wichtig sind, bleiben die Diskussionen auf diese Prozesse beschränkt, sie gelten aber analog auch für die kontinuierlichen Prozesse.

Bevor auf die numerische Berechnung eingegangen wird, sollen zwei Verfahren kurz erwähnt werden, die durch Vorverarbeitung des diskreten stationären Prozesses die im Zusammenhang mit (342.16) erwähnte Verzerrung der geglätteten Schätzung verringern.

Bislang wurde nur die Verzerrung diskutiert, die das Spektralfenster $W(\gamma)$ in der geglätteten Schätzung $\mathfrak{S}_w(\Omega)$ aus (342.15) oder (342.23) verursacht. Aber auch das Periodogramm $\mathfrak{S}(\lambda)$ trägt zur Verzerrung bei, da es aus einer endlichen Anzahl von Werten und nicht, wie in (341.10) für (342.16) angenommen wird, aus einer unendlichen Anzahl von Werten berechnet wird. Um diese Verzerrung zu vermindern, multipliziert man die ersten Werte eines diskreten Prozesses $X(n)$ für $n \in \{1, 2, \dots, M\}$ mit einer Kosinusfunktion, im Englischen "cosine taper",

$$\frac{1}{2} \left(1 - \cos \frac{n\pi}{M}\right) \quad \text{für } n \in \{1, 2, \dots, M\} \quad (343.1)$$

und das Ende des Datensatzes mit derselben Funktion, wobei insgesamt etwa 5% bis 20% der Werte des stochastischen Prozesses verändert werden (Schlittgen und Streitberg 1991, S.300).

Die Verzerrung der geglätteten Schätzung kann auch verringert werden, wenn die Maxima eines Spektrums sich glätten lassen, da dann das Durchsickern des Einflusses dieser Maxima auf benachbarte Frequenzstellen vermindert wird, was im Zusammenhang mit (342.7) diskutiert wurde. Weißes Rauschen besitzt nach (321.3) und (321.6) als Spektrum eine Konstante. Vor der Schätzung des Spektrums müßte also ein stationärer Prozeß in weißes Rauschen transformiert werden, dann sollte die Schätzung des Spektrums

erfolgen, das abschließend zurück in das Spektrum des Prozesses zu transformieren wäre. Im Englischen bezeichnet man diese Transformation daher als "prewhitening" (Blackman und Tukey 1958, S.39; Priestley 1981, S.556). Sie läßt sich einfach vornehmen, wenn ein stationärer Prozeß durch den autoregressiven Prozeß (325.1) darzustellen ist. Das Spektrum der Rauschanteile wird dann glatter als das des ursprünglichen Prozesses sein. Man schätzt daher das Spektrum der Rauschanteile und erhält aus (325.22) das Spektrum des ursprünglichen Prozesses.

Um die geglättete Schätzung $\hat{S}_w(\Omega)$ des Spektrums $S(\Omega)$ eines diskreten stationären Prozesses $X(n)$ zu berechnen, können einmal nach (342.27) M Periodogramme berechnet werden, die dann zu mitteln sind, oder die folgenden drei Möglichkeiten sind zu wählen:

1. Die Schätzung $\hat{C}(k)$ der Autokovarianzfunktion $C(k)$ wird aus (333.13) ermittelt. Die Berechnung entspricht der linearen Faltung (333.16), so daß, wie im Kapitel 218 erläutert, die schnelle Faltung angewendet werden kann. Nach Multiplikation von $\hat{C}(k)$ mit einer Fensterfunktion $w(k)$ und diskreter Fourier-Transformation folgt nach (342.13) die geglättete Schätzung $\hat{S}_w(\Omega)$ an den diskreten Stellen (342.21) für Ω . Die diskrete Fourier-Transformation kann mit der schnellen Fourier-Transformation berechnet werden. Gemäß (219.2) ist sie für Summationen definiert, bei denen der Summationsindex mit Null beginnt. In (342.13) ist daher

$$k = n-2N \quad \text{und} \quad n = k+2N \quad (343.2)$$

zu substituieren, und man erhält

$$\hat{S}_w(\Omega) = \sum_{n=0}^{4N} \hat{C}(n-2N)w(n-2N)e^{-j(n-2N)\Omega} \quad (343.3)$$

Bei dieser Art der Berechnung der geglätteten Schätzung $\hat{S}_w(\Omega)$ wird an Rechenoperationen die Faltung, die Multiplikation und die Fourier-Transformation benötigt.

2. Durch inverse Fourier-Transformation der Realisierung des stationären Prozesses $X(n)$ wird nach (341.11) das Periodogramm $\hat{S}(\Omega)$ an den diskreten Stellen (342.21) für Ω berechnet. Durch die diskrete Faltung (342.23) mit einem Spektralfenster $W(\Omega-\lambda)$ ergibt sich $\hat{S}_w(\Omega)$. Für die Fourier-Transformation und für die Faltung lassen sich die schnelle Fourier-Transformation und die schnelle Faltung einsetzen. Dieses Verfahren, die geglättete Schätzung $\hat{S}_w(\Omega)$ zu berechnen, wird häufig angewendet. Es benötigt an Rechenoperationen die Fourier-Transformation, die Multiplikation für die Quadrierung und die Faltung.

3. Wie in Methode 2. wird das Periodogramm $\hat{S}(\Omega)$ berechnet. Durch die inverse diskrete Fourier-Transformation erhält man daraus nach (342.22) die geschätzte Autokovarianzfunktion $\hat{C}(k)$. Wie in der Methode 1. wird sie mit einer Fensterfunktion $w(k)$ multi-

pliziert, so daß $\hat{S}_w(\Omega)$ durch die diskrete Fourier-Transformation (342.13) an den diskreten Stellen (342.21) für Ω folgt. Anstelle der Fourier-Transformationen lassen sich wieder die schnellen Fourier-Transformationen verwenden. Diese Berechnungsmethode für $\hat{S}_w(\Omega)$ benötigt an Rechenoperationen zwei Multiplikationen und zwei Fourier-Transformationen.

Schließlich läßt sich das Spektrum nach (354.25) auch mit Hilfe der geschätzten Parameter der autoregressiven Modelle bestimmen. Man kann zeigen, daß die *Maximum-Entropie-Methode* für die Schätzung des Spektrums mit diesem Verfahren identisch ist (Haykin und Kesler 1979, S.21).

35 Schätzungen in speziellen Modellen

351 Kontinuierliche Optimalfilter

Die Filterung deterministischer Signale war im Kapitel 216 mit Hilfe linearer verschiebungsinvarianter Systeme eingeführt worden. Auf die gleiche Weise wurden im Kapitel 323 die stochastischen Signale gefiltert. Bedeutet wieder $X(t)$ das stochastische Eingangssignal eines linearen verschiebungsinvarianten Systems, dann folgt das stochastische Ausgangssignal $Y(t)$ mit (323.2) zu

$$Y(t) = \int_{-\infty}^{\infty} h(u)X(t-u)du \quad , \quad (351.1)$$

worin $h(u)$ die Impulsantwort bezeichnet.

Eine Realisierung in Form von Messungen soll für das Eingangssignal $X(t)$ vorliegen. Von den Messungen wird angenommen, daß sie zufällige Fehler, das sogenannte *Rauschen*, enthalten. Aus den verrauschten Messungen ist durch Filterung das ideale Signal $Y(t)$ zu rekonstruieren. Das ideale Signal $Y(t)$ soll aber auch vorhergesagt werden, was als *Prädiktion* bezeichnet wird. Gelöst wird dieses Problem dadurch, daß die Impulsantwort $h(u)$ in (351.1) durch eine beste Schätzung (Koch, 1987, S.175) bestimmt wird. Man erhält dann ein *Optimalfilter* oder eine *optimale Prädiktion*.

Um sowohl zu filtern als auch zu präzisieren, wird (351.1) umgeformt zu

$$\hat{Y}(t+\tau) = \int_0^{\infty} h_{\tau}(u)X(t-u)du \quad , \quad (351.2)$$

worin $\hat{Y}(t+\tau)$ die Schätzung des idealen Signals $Y(t+\tau)$ bedeutet. Mit $\tau \leq 0$ wird die Messung $X(t)$ gefiltert, mit $\tau > 0$ wird ein Signal prädiert. Aufgrund der gewählten unteren Grenze im Integral von (351.2) lassen sich Echtzeitprobleme lösen, da Messungen benutzt werden, die in bezug auf den Zeitpunkt t in der Vergangenheit liegen. Vorausgesetzt wird, daß $X(t)$ und $Y(t)$ zumindest im weiteren Sinne stationäre Prozesse darstellen und daß

$$E(X(t)) = 0 \quad \text{und} \quad E(Y(t)) = 0 \quad (351.3)$$

gilt. Der Fehler $e(t+\tau)$ der Schätzung berechnet sich aus

$$e(t+\tau) = \hat{Y}(t+\tau) - Y(t+\tau) \quad . \quad (351.4)$$

Die beste Schätzung bedeutet nun, daß der Erwartungswert des quadratischen Fehlers der Schätzung, der mit $\hat{\sigma}^2$ bezeichnet sei, minimal wird

$$E([Y(t+\tau) - \hat{Y}(t+\tau)]^2) = \hat{\sigma}^2 = \min . \quad (351.5)$$

Substituiert man (351.2), ergibt sich

$$E([Y(t+\tau) - \int_0^{\infty} h_{\tau}(u)X(t-u)du]^2) = \hat{\sigma}^2 = \min . \quad (351.6)$$

Aus $\partial \hat{\sigma}^2 / \partial h_{\tau}(v) = 0$ folgt die sogenannte *Orthogonalitätsrelation*

$$E([Y(t+\tau) - \int_0^{\infty} h_{\tau}(u)X(t-u)du]X(t-v)) = 0 . \quad (351.7)$$

Hieraus erhält man mit (313.6), (313.11) und (351.3) die Kreuzkovarianzfunktion

$$C_{xy}(\tau+v) = \int_0^{\infty} h_{\tau}(u)C_{xx}(v-u)du , \quad (351.8)$$

die als *Wiener-Hopf-Integralgleichung* bezeichnet wird. Sie liefert die Impulsantwort $h_{\tau}(u)$, sofern die Autokovarianzfunktion $C_{xx}(k)$ und die Kreuzkovarianzfunktion $C_{xy}(k)$ gegeben sind. Mit der bekannten Impulsantwort erhält man aus (351.2) die Schätzung $\hat{Y}(t+\tau)$.

Wie bereits zu Beginn dieses Kapitels erwähnt, wird nun berücksichtigt, daß die Messung $X(t)$ das Rauschen $\varepsilon(t)$ enthält,

$$Y(t) = X(t) + \varepsilon(t) \quad \text{mit} \quad E(\varepsilon(t)) = 0 . \quad (351.9)$$

Die Optimalfilterung des Eingangssignals $X(t)$ bedeutet die bestmögliche Reduzierung des Rauschanteils. Das gefilterte, das heißt geschätzte Signal zum Zeitpunkt t ergibt sich für $\tau = 0$ aus (351.2) mit $h_0(u) = h(u)$ zu

$$\hat{Y}(t) = \int_0^{\infty} h(u)X(t-u)du . \quad (351.10)$$

Die Impulsantwort folgt aus (351.8) mit

$$C_{xy}(v) = \int_0^{\infty} h(u)C_{xx}(v-u)du . \quad (351.11)$$

Mit (351.10) und (351.11) ist der *Wiener-Optimalfilter* im Zeitbereich bestimmt (Wiener 1949). Der Integralgleichung (351.11) entspricht wegen (214.10), (314.1) und (314.18) im Frequenzbereich der Zusammenhang

$$S_{xy}(j\omega) = S_{xx}(\omega)H(j\omega) . \quad (351.12)$$

Die Frequenzantwort $H(j\omega)$ des Wiener-Optimalfilters lautet demnach

$$H(j\omega) = S_{xy}(j\omega)/S_{xx}(\omega) . \quad (351.13)$$

Nimmt man an, daß die Prozesse $Y(t)$ und $\varepsilon(t)$ voneinander unabhängig sind, daß also für die Kreuzkovarianzfunktion

$$C_{\varepsilon Y}(v) = 0 \quad (351.14)$$

gilt, dann folgt für die Kreuzkovarianzfunktion $C_{xy}(v)$ und die Autokovarianzfunktion $C_{xx}(v)$ aus (351.9)

$$C_{xy}(v) = E([Y(t) - \varepsilon(t)][Y(t+v)]) = C_{yy}(v) \quad (351.15)$$

$$C_{xx}(v) = E([Y(t) - \varepsilon(t)][Y(t+v) - \varepsilon(t+v)]) = C_{yy}(v) + C_{\varepsilon\varepsilon}(v) . \quad (351.16)$$

Es ergibt sich dann für (351.11)

$$C_{yy}(v) = \int_0^{\infty} h(u)(C_{yy}(v-u) + C_{\varepsilon\varepsilon}(v-u)) du . \quad (351.17)$$

Dieser Faltung im Zeitbereich entspricht im Frequenzbereich anstelle von (351.12) die Gleichung

$$S_{yy}(\omega) = (S_{yy}(\omega) + S_{\varepsilon\varepsilon}(\omega))H(j\omega) \quad (351.18)$$

mit der Frequenzantwort des speziellen Optimalfilters

$$H(j\omega) = \frac{S_{yy}(\omega)}{S_{yy}(\omega) + S_{\varepsilon\varepsilon}(\omega)} . \quad (351.19)$$

Sind die Spektren $S_{yy}(\omega)$ und $S_{\varepsilon\varepsilon}(\omega)$ gegeben, kann die Frequenzantwort berechnet werden. Da sie eine reelle Funktion ist, bewirkt sie nach (221.43) eine nullphasige Filterung. Die mit der inversen Fourier-Transformation (214.2) aus (351.19) berechenbare Impulsantwort

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(j\omega) e^{j\omega t} d\omega \quad (351.20)$$

unterliegt analog zu (221.44) der Bedingung einer geraden Funktion

$$h(t) = h(-t) . \quad (351.21)$$

Der mittlere quadratische Schätzfehler $\hat{\sigma}^2$ berechnet sich mit $\tau = 0$ aus (351.6) aufgrund der Orthogonalitätsrelation (351.7), die für alle v gilt, mit (313.6) und (313.10) zu

$$\begin{aligned} \hat{\sigma}^2 &= E([Y(t) - \hat{Y}(t)]^2) \\ &= E([Y(t) - \int_0^{\infty} h(u)X(t-u)du]Y(t)) \\ &= C_{yy}(0) - \int_0^{\infty} h(u)C_{xy}(u)du . \end{aligned} \quad (351.22)$$

Hieraus folgt durch Substitution von (314.3), (314.20) und (323.14) sowie durch Variablentransformation in (314.20)

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} (S_{yy}(\omega) - \int_0^{\infty} h(u) S_{xy}(-j\omega) e^{-j\omega u} du) d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} (S_{yy}(\omega) - H(j\omega) S_{xy}(-j\omega)) d\omega .\end{aligned}\quad (351.23)$$

Setzt man weiter (351.13) ein, ergibt sich schließlich

$$\hat{\sigma}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} (S_{yy}(\omega) - \frac{S_{xy}(j\omega) S_{xy}(-j\omega)}{S_{xx}(\omega)}) d\omega .\quad (351.24)$$

Mit (351.15) und (351.16) sowie mit $S_{yy}(-j\omega) = S_{yy}(j\omega)$ aus (314.15) vereinfacht sich diese Gleichung zu

$$\hat{\sigma}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S_{yy}(\omega) S_{\varepsilon\varepsilon}(\omega)}{S_{yy}(\omega) + S_{\varepsilon\varepsilon}(\omega)} d\omega .\quad (351.25)$$

Hieraus erkennt man, daß eine perfekte Filterung vorliegt, wenn für die Spektren $S_{yy}(\omega)$ und $S_{\varepsilon\varepsilon}(\omega)$ die Bedingung

$$S_{yy}(\omega) S_{\varepsilon\varepsilon}(\omega) = 0\quad (351.26)$$

zutrifft. Sie ist dann erfüllt, wenn die Spektren $S_{yy}(\omega)$ und $S_{\varepsilon\varepsilon}(\omega)$ sich nicht überlappen. Ist unter dieser Voraussetzung das Spektrum $S_{yy}(\omega)$ mit $S_{yy}(\omega) > 0$ auf den Bereich $\omega_u < |\omega| < \omega_o$ bandbegrenzt und gilt $S_{\varepsilon\varepsilon}(\omega) = 0$ für $\omega_u < |\omega| < \omega_o$, so folgt aus (351.19) für die Frequenzantwort $H(j\omega)$

$$H(j\omega) = \begin{cases} 1 & \text{für } \omega_u < |\omega| < \omega_o \\ 0 & \text{sonst} . \end{cases}\quad (351.27)$$

Nach Abbildung 222-1 liegt dann ein Bandpaß vor.

Bedeutet σ_y^2 die Varianz von $Y(t)$, dann wird der Ausdruck $10(\log(\sigma_y^2/\hat{\sigma}^2))$ *Signal-Rausch-Verhältnis* genannt und mit der Bezeichnung *Dezibel* versehen, worauf schon im Zusammenhang mit (221.33) eingegangen wurde.

352 Diskrete Optimalfilter

Diskrete Messungen $X(n)$ eines Signals sollen dazu benutzt werden, um das ideale diskrete Signal $Y(n)$ aus einer Filterung von $X(n)$ zu gewinnen oder zu präzisieren. Beide Prozesse seien zumindest im weiteren Sinne stationär, und es gelte

$$E(X(n)) = 0 \quad \text{und} \quad E(Y(n)) = 0 . \quad (352.1)$$

Für die Schätzung $\hat{Y}(n+p)$ des Signals $Y(n+p)$ wird in Analogie zu (351.2) der Ansatz

$$\hat{Y}(n+p) = \sum_{k=0}^N h_p(k)X(n-k) \quad \text{mit} \quad p \in \{0, \pm 1, \pm 2, \dots\} \quad (352.2)$$

gewählt. Für $p \leq 0$ wird die Messung $X(n)$ gefiltert, für $p > 0$ wird ein Signal prädiert. Für $p = 0$ liegt wegen

$$\hat{Y}(n) = \sum_{k=0}^N h_0(k)X(n-k) \quad (352.3)$$

eine kausale nichtrekursive Filterung vor, wie man aus (216.8) und (217.14) erkennt. Für die Impulsantwort $h_0(k)$ gilt also

$$h_0(k) = 0 \quad \text{für} \quad k < 0 . \quad (352.4)$$

Substituiert man in (352.2) $m = n+p$ und $n = m-p$ ergibt sich

$$\hat{Y}(m) = \sum_{k=0}^N h_p(k)X(m-(p+k)) .$$

Setzt man weiter $l = p+k$ und $k = l-p$, erhält man

$$\hat{Y}(m) = \sum_{l=p}^{N+p} h_p(l-p)X(m-l) = \sum_{l=p}^{N+p} \bar{h}_p(l)X(m-l) \quad (352.5)$$

mit der Impulsantwort

$$\bar{h}_p(l) = h_p(l-p) . \quad (352.6)$$

Für $p < 0$ liegt also analog zu (225.30) eine nichtkausale nichtrekursive Filterung vor, für $p > 0$ erhält man wie für $p = 0$ eine kausale nichtrekursive Filterung.

Die beste Schätzung des Ausgangssignals bedeutet nach (351.5) im diskreten Fall

$$\begin{aligned} E([Y(n+p) - \hat{Y}(n+p)]^2) \\ = E([Y(n+p) - \sum_{k=0}^N h_p(k)X(n-k)]^2) = \hat{\sigma}^2 = \min . \end{aligned} \quad (352.7)$$

Aus $\partial \hat{\sigma}^2 / \partial h_p(m) = 0$ folgt die diskrete Orthogonalitätsrelation

$$E([Y(n+p) - \sum_{k=0}^N h_p(k)X(n-k)]X(n-m)) = 0 \quad (352.8)$$

für $m \in \{0, \dots, N\}$. Mit (313.12) und (313.14) erhält man hieraus die diskrete Form der

Wiener-Hopf-Integralgleichung

$$C_{xy}(m+p) = \sum_{k=0}^N h_p(k) C_{xx}(m-k) . \quad (352.9)$$

Sind die Kreuzkovarianzen $C_{xy}(m+p)$ und die Autokovarianzen $C_{xx}(m-k)$ gegeben, kann die Impulsantwort $h(k)$ aus (352.9) berechnet werden. Dazu führt man die $(N+1) \times 1$ Vektoren

$$\mathbf{c}_p = [C_{xy}(p), C_{xy}(p+1), \dots, C_{xy}(p+N)]' , \quad (352.10)$$

$$\mathbf{h}_p = [h_p(0), h_p(1), \dots, h_p(N)]' \quad (352.11)$$

sowie die $(N+1) \times (N+1)$ Matrix Σ_{xx} der Autokovarianzen

$$\Sigma_{xx} = \begin{bmatrix} C_{xx}(0) & C_{xx}(1) & \dots & C_{xx}(N) \\ C_{xx}(1) & C_{xx}(0) & \dots & C_{xx}(N-1) \\ \dots & \dots & \dots & \dots \\ C_{xx}(N) & C_{xx}(N-1) & \dots & C_{xx}(0) \end{bmatrix} \quad (352.12)$$

ein und erhält wegen (313.13) das lineare Gleichungssystem

$$\mathbf{c}_p = \Sigma_{xx} \mathbf{h}_p \quad (352.13)$$

mit der Lösung, falls Σ_{xx} als positiv definit vorausgesetzt wird,

$$\mathbf{h}_p = \Sigma_{xx}^{-1} \mathbf{c}_p . \quad (352.14)$$

Die berechnete Impulsantwort $h_p(0), h_p(1), \dots, h_p(N)$ wird in (352.2) eingesetzt und ergibt die Schätzung $\hat{Y}(n+p)$ des Signals $Y(n+p)$. Wie erwähnt, erhält man mit $p \leq 0$ eine Filterung und mit $p > 0$ eine Prädiktion des Signals.

Wie in (351.9) soll die Messung $X(n)$ einen Rauschanteil $\varepsilon(n)$ besitzen

$$Y(n) = X(n) + \varepsilon(n) . \quad (352.15)$$

Unter der Voraussetzung

$$C_{\varepsilon y}(k) = 0 \quad (352.16)$$

gilt dann analog zu (351.15) und (351.16) für die Kreuzkovarianzen $C_{xy}(k)$

$$C_{xy}(k) = C_{yy}(k) \quad (352.17)$$

und für die Matrix Σ_{xx} der Autokovarianzen aus (352.12)

$$\Sigma_{xx} = \Sigma_{yy} + \Sigma_{\varepsilon\varepsilon} . \quad (352.18)$$

Nach Substitution von (352.17) und (352.18) in (352.14) folgt die Impulsantwort und damit die Schätzung $\hat{Y}(n+p)$ des Signals aus (352.2).

Der mittlere quadratische Schätzfehler $\hat{\sigma}^2$ aus (352.7) ergibt sich für $p = 0$ analog zu (351.22)

$$\hat{\sigma}^2 = E[(Y(n) - \hat{Y}(n))^2] = C_{yy}(0) - \sum_{k=0}^N h_0(k) C_{xy}(k) \quad (352.19)$$

und mit (352.10), (352.11) sowie (352.14)

$$\hat{\sigma}^2 = C_{yy}(0) - \mathbf{c}_0' \mathbf{h}_0 = \sigma_y^2 - \mathbf{c}_0' \Sigma_{xx}^{-1} \mathbf{c}_0, \quad (352.20)$$

worin (352.17) und (352.18) gelten, falls die Messung $X(n)$ nach (352.15) den Rauschanteil $\varepsilon(n)$ besitzt und (352.16) vorausgesetzt wird.

Zur Filterung des verrauschten Signals $X(n)$ können anstelle des hier entwickelten Optimalfilters auch Filter mit ähnlichen Eigenschaften eingesetzt werden, beispielsweise die im Kapitel 226 vorgestellten Rechteck- und Binomialfilter.

Es wird nun auf Möglichkeiten eingegangen, die nach Substitution von (352.17) und (352.18) in (352.14) benötigten Autokovarianzen $C_{yy}(k)$ und $C_{\varepsilon\varepsilon}(k)$ zu bestimmen. Die Rauschanteile $\varepsilon(n)$ werden im allgemeinen als voneinander unabhängig angenommen, so daß für $\varepsilon(n)$ stationäres weißes Rauschen mit der Autokovarianzfunktion (321.5) folgt

$$C_{\varepsilon\varepsilon}(k) = \sigma_\varepsilon^2 \delta(k). \quad (352.21)$$

Die Varianz σ_ε^2 ist ein Maß für die Genauigkeit der Messungen $X(n)$ und kann zum Beispiel aus Wiederholungsmessungen geschätzt werden, falls sie nicht bekannt ist.

Sind die Autokovarianzen $C_{yy}(k)$ unbekannt, müssen sie näherungsweise aus den Messungen $X(n)$ nach (333.13) oder (333.16) geschätzt werden. Paßt man die Abklingfunktionen der Form (313.17) oder (313.19) den Schätzwerten an, lassen sich die Autokovarianzen $C_{yy}(k)$ sehr einfach aus einer Abklingfunktion ermitteln.

Eine weitere Möglichkeit, eine Autokovarianzfunktion für $Y(n)$ zu gewinnen, ist durch einen Markoff-Prozeß erster Ordnung gegeben. Dazu wählt man wegen (325.24) einen normalverteilten stationären AR(1)-Prozeß für das ideale Signal $Y(n)$, das durch die Messung $X(n)$ angenähert wird,

$$Y(n) = a_1 Y(n-1) + \varepsilon(n). \quad (352.22)$$

Aufgrund des Beispiels 1 des Kapitels 325 muß $|a_1| < 1$ gelten. Die Autokovarianzfunktion $C_{yy}(k)$ des Signals $Y(n)$ folgt nach (325.15) für $k > 0$ aus

$$C_{yy}(k) - a_1 C_{yy}(k-1) = 0. \quad (352.23)$$

Als Ansatz zur Lösung dieser Differenzgleichung wird

$$C_{yy}(k) = Aa_1^k \quad (352.24)$$

gewählt, wobei A eine beliebige reelle Konstante bedeutet. Setzt man nämlich (352.24) in (352.23) ein, erhält man

$$Aa_1^k - a_1Aa_1^{k-1} = 0 \quad (352.25)$$

Aus (352.24) ergibt sich für $k = 0$

$$C_{yy}(0) = \sigma_y^2 = A \quad (352.26)$$

Somit erhält man

$$C_{yy}(k) = \sigma_y^2 a_1^{|k|} \quad \text{mit } |a_1| < 1 \quad (352.27)$$

Die Varianz σ_y^2 des Prozesses läßt sich nach (333.13) und der Koeffizient a_1 mit den Methoden des Kapitels 354 schätzen.

353 Modell der Prädiktion und Filterung

In (352.15) wird das zu rekonstruierende Signal $Y(n)$ durch die Messung $X(n)$ und das Rauschen $\varepsilon(n)$ ausgedrückt. Benutzt man sämtliche Messungen für die Prädiktion und Filterung, erhält man die Ergebnisse des *Modells der Prädiktion und Filterung* ohne Trendanteil (Koch 1987, S.257; Koch 1990, S.115; Moritz 1980, S.77). Es lautet in der hier verwendeten Notation

$$\begin{aligned} \mathbf{y} &= \mathbf{x} + \boldsymbol{\varepsilon} \quad \text{mit } E(\mathbf{y}) = \mathbf{0}, E(\boldsymbol{\varepsilon}) = \mathbf{0} \quad \text{und} \\ D(\mathbf{y}) &= \sigma^2 \boldsymbol{\Sigma}_{yy}, D(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{\Sigma}_{\varepsilon\varepsilon}, C(\mathbf{y}, \boldsymbol{\varepsilon}) = \mathbf{0} \end{aligned} \quad (353.1)$$

Es seien N Messungen gegeben, so daß man aus (352.2) für $n = N$ erhält

$$\hat{Y}(N+p) = \sum_{k=0}^N h_p(k) X(N-k) \quad (353.2)$$

Nach Einführung des $(N+1) \times 1$ Vektors

$$\mathbf{x} = [X(N), X(N-1), \dots, X(0)]' \quad (353.3)$$

läßt sich für (353.2) mit (352.11) schreiben

$$\hat{Y}(N+p) = \mathbf{h}_p' \mathbf{x} \quad (353.4)$$

Sämtliche N Messungen sollen gefiltert und zusätzlich M Signalwerte prädiiziert werden. Dazu ist $p \in \{-N, -N+1, \dots, -1, 0, 1, \dots, M\}$ in (353.4) zu setzen, und man erhält die Gleichungen

$$\begin{aligned}
\hat{Y}(0) &= \mathbf{h}_{-N}' \mathbf{x} \\
\hat{Y}(1) &= \mathbf{h}_{-N+1}' \mathbf{x} \\
&\dots\dots\dots \\
\hat{Y}(N-1) &= \mathbf{h}_{-1}' \mathbf{x} \\
\hat{Y}(N) &= \mathbf{h}_0' \mathbf{x} \\
\hat{Y}(N+1) &= \mathbf{h}_1' \mathbf{x} \\
&\dots\dots\dots \\
\hat{Y}(N+M) &= \mathbf{h}_M' \mathbf{x} \quad , \quad (353.5)
\end{aligned}$$

die in Matrixform lauten

$$\hat{\mathbf{y}} = \mathbf{H} \mathbf{x} \quad , \quad (353.6)$$

worin

$$\hat{\mathbf{y}} = [\hat{Y}(0), \hat{Y}(1), \dots, \hat{Y}(N-1), \hat{Y}(N), \hat{Y}(N+1), \dots, \hat{Y}(N+M)]' \quad (353.7)$$

ein $(N+M+1) \times 1$ Vektor und \mathbf{H}

$$\mathbf{H} = [\mathbf{h}_{-N}, \mathbf{h}_{-N+1}, \dots, \mathbf{h}_{-1}, \mathbf{h}_0, \mathbf{h}_1, \dots, \mathbf{h}_M]' \quad (353.8)$$

eine zunächst unbekannte $(N+M+1) \times (N+1)$ Matrix bedeuten. Sie läßt sich bestimmen, indem man (352.13) für $p \in \{-N, -N+1, \dots, M\}$ aufstellt und in der Form

$$[\mathbf{c}_{-N}, \dots, \mathbf{c}_0, \dots, \mathbf{c}_M] = \Sigma_{xx} [\mathbf{h}_{-N}, \dots, \mathbf{h}_0, \dots, \mathbf{h}_M] \quad (353.9)$$

zusammenfaßt. Führt man weiter die $(N+M+1) \times (N+1)$ Matrix

$$\Sigma_{xy} = [\mathbf{c}_{-N}, \dots, \mathbf{c}_0, \dots, \mathbf{c}_M]' \quad (353.10)$$

ein, lautet (353.9) unter Berücksichtigung von (353.8)

$$\Sigma_{xy} = \mathbf{H} \Sigma_{xx} \quad . \quad (353.11)$$

Diese Beziehung wird auch als Wiener-Hopf Gleichung bezeichnet (Scharf 1991, S.327).

Mit ihrer Lösung

$$\mathbf{H} = \Sigma_{xy} \Sigma_{xx}^{-1} \quad (353.12)$$

erhält man aus (353.6)

$$\hat{\mathbf{y}} = \Sigma_{xy} \Sigma_{xx}^{-1} \mathbf{x} \quad . \quad (353.13)$$

Da diese Schätzung mit (352.13) gewonnen wird, liegt eine beste Schätzung vor.

Führt man schließlich noch (352.17) und (352.18) beziehungsweise (353.1) in (353.13) ein, folgt

$$\hat{\mathbf{y}} = \Sigma_{yy}^* (\Sigma_{yy} + \Sigma_{\epsilon\epsilon})^{-1} \mathbf{x} \quad (353.14)$$

mit

$$\Sigma_{yy}^* = [\mathbf{c}_{-N}^*, \dots, \mathbf{c}_0^*, \dots, \mathbf{c}_M^*]' \quad .$$

Hierin gilt mit (352.10)

$$\mathbf{c}_p^* = [C_{yy}(p), C_{yy}(p+1), \dots, C_{yy}(p+N)]'$$

und folglich wegen (313.13)

$$\Sigma_{yy}^* = \begin{bmatrix} C_{yy}(N) & C_{yy}(N-1) & \dots & C_{yy}(0) \\ \dots & \dots & \dots & \dots \\ C_{yy}(0) & C_{yy}(1) & \dots & C_{yy}(N) \\ \dots & \dots & \dots & \dots \\ C_{yy}(M) & C_{yy}(M+1) & \dots & C_{yy}(M+N) \end{bmatrix} \quad .$$

Die als Elemente in den Matrizen Σ_{yy}^* , Σ_{yy} und $\Sigma_{\epsilon\epsilon}$ auftretenden Autokovarianzen sind, falls sie nicht bekannt sind, wieder nach (352.21) bis (352.27) zu bestimmen. Die Schätzung $\hat{\mathbf{y}}$ enthält aufgrund der speziellen Wahl der Größe p in (353.4) für $p \leq 0$ einen Filterungsanteil und für $p > 0$ einen Prädiktionsanteil.

An dieser Stelle soll noch das *Kalman-Bucy-Filter* erwähnt werden, das zur Schätzung der Zustandsvariablen eines dynamischen Systems dient. Es läßt sich sehr anschaulich mit der Bayes-Statistik herleiten (Koch 1990, S.92) und wird hier nicht behandelt.

354 Autoregressive Prozesse

Ein autoregressiver Prozeß der Ordnung p oder AR(p)-Prozeß wird nach (325.1) durch eine Linearkombination vorangegangener Werte des Prozesses und durch einen Rauschanteil definiert. Die Koeffizienten a_k für $k \in \{1, \dots, p\}$, die nach (325.17) im $p \times 1$ Vektor $\boldsymbol{\beta}$ zusammengefaßt werden, ergeben sich nach (325.20) aus

$$\boldsymbol{\beta} = \Sigma_p^{-1} \mathbf{c}_p \quad , \quad (354.1)$$

falls die $p \times p$ Matrix Σ_p aus (325.19) positiv definit ist. Unter der Voraussetzung, daß die Autokovarianzen des AR(p)-Prozesses, mit denen Σ_p und \mathbf{c}_p gebildet werden, bekannt sind, soll nun gezeigt werden, daß mit (354.1) eine beste Schätzung der Koeffizienten des autoregressiven Prozesses vorliegt. Stellt man das autoregressive Modell in der Form (325.30) dar, erhält man die beste Schätzung (352.7) entsprechend mit

$$E((X(n) - [\boldsymbol{\beta}', 1] \begin{bmatrix} \mathbf{x}_p(n) \\ \boldsymbol{\epsilon}(n) \end{bmatrix}))^2) = \hat{\sigma}^2 = \min \quad . \quad (354.2)$$

Aus $\hat{\partial}^2/\partial\beta_i = 0$ mit $\beta = (\beta_i)$ und $i \in \{1, \dots, p\}$ folgt die Orthogonalitätsrelation

$$E((X(n) - [\beta', 1] \begin{bmatrix} \mathbf{x}_p \\ \varepsilon(n) \end{bmatrix}))X(n-i)) = 0 \quad (354.3)$$

wegen $\mathbf{x}_p = (X(n-i)) = [X(n-1), X(n-2), \dots, X(n-p)]'$. Weiter erhält man

$$\begin{aligned} E(X(n)X(n-i)) - \beta' E(\mathbf{x}_p X(n-i)) - E(\varepsilon(n)X(n-i)) \\ = C_{xx}(i) - \beta' \mathbf{c}_{i-p} = 0 \end{aligned} \quad (354.4)$$

wegen (325.10) nach Einführung des Vektors

$$\mathbf{c}_{i-p} = [C_{xx}(i-1), C_{xx}(i-2), \dots, C_{xx}(i-p)]' \quad (354.5)$$

Stellt man (354.4) in der Form $C_{xx}(i) = \mathbf{c}_{i-p}' \beta$ für sämtliche Werte von i auf, erhält man mit der Matrix $\Sigma_p = [\mathbf{c}_{1-p}, \mathbf{c}_{2-p}, \dots, \mathbf{c}_0]'$ und dem Vektor \mathbf{c}_p aus (325.18) das Gleichungssystem (325.20). Daher führt (354.1) auf eine beste Schätzung. Sie ist, wie die Ableitung zeigt, unabhängig vom Rauschanteil $\varepsilon(n)$.

Ist die Varianz σ_ε^2 aus (325.4) neben den Koeffizienten a_k für $k \in \{1, \dots, p\}$ unbekannt, faßt man die Gleichungen (325.14) und (325.20)

$$C_{xx}(0) - \mathbf{c}_p' \beta = \sigma_\varepsilon^2$$

$$\mathbf{c}_p - \Sigma_p \beta = 0$$

zusammen zu dem linearen Gleichungssystem

$$\begin{bmatrix} C_{xx}(0) & \mathbf{c}_p' \\ \mathbf{c}_p & \Sigma_p \end{bmatrix} \begin{bmatrix} 1 \\ -\beta \end{bmatrix} = \begin{bmatrix} \sigma_\varepsilon^2 \\ 0 \end{bmatrix} \quad (354.6)$$

Die Lösung kann über den Levinson-Rekursionsalgorithmus erhalten werden (Buttkus 1991, S.238; Jain 1989, S.198).

Die Berechnung der Koeffizienten a_k eines autoregressiven Prozesses nach (354.1) oder (354.6) setzt, wie bereits erwähnt, die Kenntnis der Autokovarianzen $C_{xx}(k)$ voraus. Sollen die Koeffizienten a_k aus der Realisierung des Prozesses $X(n)$ geschätzt werden, bedient man sich zweckmäßig statistischer Schätzverfahren wie der Maximum-Likelihood-Methode.

Die Likelihoodfunktion $L(\beta)$ als Funktion des Vektors β der unbekanntenen Koeffizienten a_k aus (325.17) gewinnt man aus der gemeinsamen Dichte (325.49) der $X(n)$

$$L(\beta) = \left[\frac{1}{(2\pi\sigma_\varepsilon^2)^q \det Q_p} \right]^{1/2} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} (\mathbf{x}_p' Q_p^{-1} \mathbf{x}_p + (\bar{\mathbf{y}} - X\beta)' (\bar{\mathbf{y}} - X\beta)) \right] \quad (354.7)$$

Bei der Maximum-Likelihood-Methode wird β derart geschätzt, daß $L(\beta)$ maximal wird (Koch 1987, S.187). Die Matrix $Q_p = \Sigma_p/\sigma_\varepsilon^2$ der Autokovarianzen ist jedoch nach

(325.13) eine Funktion der unbekannt Parameter β . Die partiellen Ableitungen nach den unbekannt Parametern führen auf nichtlineare Gleichungen, die iterativ zu lösen sind. Anstelle von (354.7) wird daher eine genäherte Likelihoodfunktion $L_0(\beta)$ gewählt. Da im allgemeinen $p \ll K$ gilt, wird sie mit der bedingten Dichte (325.46) gebildet (Bartlett 1955, S.241; Box und Jenkins 1970, S.277). Man erhält

$$L_0(\beta) = \left[\frac{1}{2\pi\sigma_\epsilon^2} \right]^{(K+1)/2} \exp \left[-\frac{1}{2\sigma_\epsilon^2} (\bar{y} - X\beta)' (\bar{y} - X\beta) \right] . \quad (354.8)$$

Nimmt man für die Parameterschätzung N Messungen mit $N = K+1$ an und setzt $n = p$, dann lauten der Vektor \bar{y} und die Matrix X aus (325.44) und (325.45), falls die Reihenfolge der Werte in \bar{y} und der Zeilen in X umgekehrt wird,

$$\bar{y} = [X(p), X(p+1), \dots, X(p+N-1)]' , \quad (354.9)$$

$$X = \begin{bmatrix} X(p-1) & X(p-2) & \dots & X(0) \\ X(p) & X(p-1) & \dots & X(1) \\ \dots & \dots & \dots & \dots \\ X(p+N-2) & X(p+N-3) & \dots & X(N-1) \end{bmatrix} . \quad (354.10)$$

Für die Schätzung der p autoregressiven Parameter a_k werden also $p+N-1$ Werte von $X(n)$ benötigt. Die Elemente der ersten Zeile von X bezeichnet man als Startwerte, da sie im Vektor \bar{y} aus (354.9) nicht enthalten sind.

Die Maximum-Likelihood-Methode liefert mit der genäherten Likelihoodfunktion (354.8) das lineare Normalgleichungssystem (Koch 1987, S.188)

$$X'X \hat{\beta} = X'\bar{y} , \quad (354.11)$$

aus denen, falls $X'X$ positiv definit ist, die Schätzwerte $\hat{\beta}$ von β mit

$$\hat{\beta} = (X'X)^{-1} X'\bar{y} \quad (354.12)$$

folgen. Mit ihnen ergeben sich die Residuen $\hat{\epsilon}$ zu

$$\hat{\epsilon} = X \hat{\beta} - \bar{y} . \quad (354.13)$$

Wegen $\hat{\beta} = [\hat{a}_1, \dots, \hat{a}_p]'$ und $\hat{\epsilon} = [-\hat{\epsilon}(p), \dots, -\hat{\epsilon}(p+N-1)]'$ nach (325.1) erhält man das geschätzte Signal $\hat{X}(n)$ zu

$$\hat{X}(n) = \sum_{k=1}^p \hat{a}_k \hat{X}(n-k) + \hat{\epsilon}(n) \quad (354.14)$$

für $n \in \{p, \dots, p+N-1\}$.

Die Elemente der Normalgleichungsmatrix $X'X = N = (n_{ij})$ berechnen sich aus

$$n_{ij} = \sum_{n=1}^N X(p-1+n-i)X(p-1+n-j) . \quad (354.15)$$

Vergleicht man diesen Ausdruck mit (333.13), läßt sich näherungsweise schreiben

$$n_{ij} \approx N \tilde{C}_{xx}(i-j) . \quad (354.16)$$

Ebenso gilt für die Elemente des Absolutgliedvektors $X' \bar{y} = (1_i)$

$$1_i = N \tilde{C}_{xx}(i) . \quad (354.17)$$

Anstelle von (354.11) erhält man dann das Gleichungssystem

$$\tilde{c}_p = \Sigma_p \beta \quad (354.18)$$

mit dem $p \times 1$ Vektor

$$\tilde{c}_p = [\tilde{C}_{xx}(1), \tilde{C}_{xx}(2), \dots, \tilde{C}_{xx}(p)]' \quad (354.19)$$

und der $p \times p$ Matrix

$$\Sigma_p = \begin{bmatrix} \tilde{C}_{xx}(0) & \tilde{C}_{xx}(1) & \dots & \tilde{C}_{xx}(p-1) \\ \tilde{C}_{xx}(1) & \tilde{C}_{xx}(0) & \dots & \tilde{C}_{xx}(p-2) \\ \dots & \dots & \dots & \dots \\ \tilde{C}_{xx}(p-1) & \tilde{C}_{xx}(p-2) & \dots & \tilde{C}_{xx}(0) \end{bmatrix} , \quad (354.20)$$

das formal mit (325.20) und (354.1) übereinstimmt. Dort ist es jedoch mit den Autokovarianzen $C_{xx}(k)$ formuliert, während (354.18) die Schätzwerte $\tilde{C}_{xx}(k)$ für die Autokovarianzen enthält. Ist Σ_p positiv definit, folgt die Schätzung β der Koeffizienten a_k mit

$$\beta = \Sigma_p^{-1} \tilde{c}_p . \quad (354.21)$$

Wegen der Analogie zu den Yule-Walker-Gleichungen (325.16) bezeichnet man β als *Yule-Walker-Schätzer*. Mit ihm erhält man eine weitere Möglichkeit, die Koeffizienten des autoregressiven Prozesses zu schätzen. Für große Werte von N in (354.15) werden die Schätzwerte aus (354.21) nur geringfügig von denen aus (354.12) differieren. Bei kleinen Werten von N ist (354.12) vorzuziehen, falls dann nicht sogar mit der Likelihoodfunktion (354.7) zu arbeiten ist.

Die Methode der kleinsten Quadrate liefert Schätzungen, die mit denen der Maximum-Likelihood-Methode identisch sind, falls angenommen wird, daß die Beobachtungen im Vektor \bar{y} aus (354.9) voneinander unabhängig sind und gleiche Varianzen besitzen. Als erwartungstreue Schätzung $\hat{\sigma}_\varepsilon^2$ der Varianz σ_ε^2 des Rauschanteils (Koch 1987, S.189) erhält man den sogenannten mittleren quadratischen Prädiktionsfehler

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N-p} (\bar{y} - X\hat{\beta})' (\bar{y} - X\hat{\beta}) = \hat{\sigma}_\varepsilon^2(p) . \quad (354.22)$$

Dieser Fehler wird zur Festlegung des Grades eines autoregressiven Prozesses verwendet, denn neben den Koeffizienten a_k und der Varianz σ_ε^2 ist auch die Ordnung p als unbekannter Parameter des Modells aufzufassen. Er führt auf das Akaike-Kriterium, das nach seiner englischen Bezeichnung "final prediction error" auch FPE-Kriterium genannt wird (Priestley 1981, S.372),

$$\text{FPE}(p) = (1+p/N)\hat{\sigma}_\varepsilon^2(p) \quad (354.23)$$

worin p die Ordnung des jeweiligen Modellansatzes nach (325.1) bedeutet. Mit (354.23) ist der optimale Grad $p = P$ des AR(p)-Prozesses dann gegeben, wenn gilt

$$\text{FPE}(P+1) \geq \text{FPE}(P) \quad (354.24)$$

Ausgehend von einem AR(1)-Prozeß wird also der Grad des Modellansatzes p schrittweise erhöht, bis das FPE-Kriterium sein erstes Minimum aufweist. Ein numerisch effizientes Verfahren zur Bestimmung des optimalen Grades p eines autoregressiven Prozesses entwickelten Barrodale und Erickson (1980) basierend auf der Parameterschätzung (354.12). Zusätzlich zur Schätzung nach (354.12), die für sämtliche aufsteigenden Werte im Beobachtungsvektor \bar{y} formuliert wird, wird die Parameterschätzung auch für absteigende Werte vorgenommen.

Setzt man die Schätzwerte $\hat{\beta}$ aus (354.12) und $\hat{\sigma}_\varepsilon^2$ aus (354.22) in (325.22) ein, erhält man die Schätzung $\hat{S}_{xx}(\Omega)$ des Spektrums $S_{xx}(\Omega)$

$$\hat{S}_{xx}(\Omega) = \hat{\sigma}_\varepsilon^2 / \left| \left(\sum_{k=0}^P \hat{a}_k e^{-jk\Omega} \right) \right|^2 \quad (354.25)$$

Die Schätzung der unbekannt Parameter eines MA(q)-Prozesses, der in (327.1) definiert wurde, oder eines ARMA(p,q)-Prozesses nach (327.11) kann ebenfalls nach der Maximum-Likelihood-Methode erfolgen (Box und Jenkins 1970, S.209; Jenkins und Watts 1968, S.200; Priestley 1981, S.354; Schlittgen und Streitberg 1991, S.181).

355 Erweitertes autoregressives Modell

Für die Parameterschätzung im erweiterten autoregressiven Modell geht man von (326.3) aus. Beim Trend $b(n)$ wird im allgemeinen eine nichtlineare Funktion der unbekannt Parameter c_i mit $i \in \{1, \dots, q\}$ vorliegen, so daß eine Linearisierung

$$b(n) = b_0(n) + \mathbf{d}_0(n)' \delta \mathbf{c} \quad (355.1)$$

mit

$$b_0(n) = b(n, c_{01}, \dots, c_{0q})$$

$$\mathbf{d}_0(n) = [(\partial b(n)/\partial c_1)_0, \dots, (\partial b(n)/\partial c_q)_0]'$$

$$\delta \mathbf{c} = [\delta c_1, \dots, \delta c_q]'$$

notwendig ist. Dabei ist $\delta \mathbf{c}$ wegen $c_i = c_{0i} + \delta c_i$ für $i \in \{1, \dots, q\}$ der $q \times 1$ Vektor der unbekanntenen Zuschläge δc_i zu den Näherungswerten c_{0i} . Mit (355.1) lautet (326.3)

$$\begin{aligned} Y(n) &= \sum_{k=1}^p a_k(Y(n-k) - b_0(n-k)) + b_0(n) + \mathbf{d}_0(n)' \delta \mathbf{c} \\ &- \sum_{k=1}^p a_k \mathbf{d}_0(n-k)' \delta \mathbf{c} - \sum_{k=1}^p a_k r(n-k) + \varepsilon(n) + r(n) . \end{aligned} \quad (355.2)$$

Führt man weiter für die autoregressiven Parameter die Aufspaltung $a_k = a_{0k} + \delta a_k$ mit $k \in \{1, \dots, p\}$ ein, worin a_{0k} wiederum die Näherungswerte und δa_k die Zuschläge bedeuten, so läßt sich bei Vernachlässigung von Produkten kleiner Größe schreiben

$$\begin{aligned} Y(n) &- \sum_{k=1}^p a_{0k}(Y(n-k) - b_0(n-k)) - b_0(n) - \varepsilon(n) \\ &= \sum_{k=1}^p (Y(n-k) - b_0(n-k)) \delta a_k + \mathbf{d}_0(n)' \delta \mathbf{c} - \sum_{k=1}^p a_{0k} \mathbf{d}_0(n-k)' \delta \mathbf{c} \\ &- \sum_{k=1}^p a_{0k} r(n-k) + r(n) . \end{aligned} \quad (355.3)$$

Mit der Substitution

$$\bar{Y}(n) = Y(n) - b_0(n) \quad (355.4)$$

geht (355.3) über in

$$\begin{aligned} \bar{Y}(n) - \sum_{k=1}^p a_{0k} \bar{Y}(n-k) - \varepsilon(n) &= \sum_{k=1}^p \bar{Y}(n-k) \delta a_k + \mathbf{d}_0(n)' \delta \mathbf{c} \\ &- \sum_{k=1}^p a_{0k} \mathbf{d}_0(n-k)' \delta \mathbf{c} - \sum_{k=1}^p a_{0k} r(n-k) + r(n) . \end{aligned} \quad (355.5)$$

Stellt man diese Gleichung für $n \in \{p, \dots, N-1\}$ auf, erhält man mit den Vektoren

$$\bar{\mathbf{y}} = [\bar{Y}(0), \dots, \bar{Y}(N-1)]'$$

$$\mathbf{e} = [\varepsilon(p), \dots, \varepsilon(N-1)]'$$

$$\mathbf{r} = [r(0), \dots, r(N-1)]'$$

und den Matrizen

$$\mathbf{Z} = \begin{bmatrix} a_{0p} & a_{0,p-1} & \dots & a_{01} & -1 & 0 & \dots & 0 & 0 \\ 0 & a_{0p} & \dots & a_{02} & a_{01} & -1 & \dots & 0 & 0 \\ \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & a_{01} & -1 \end{bmatrix}$$

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$$

$$\mathbf{X}_1 = \begin{bmatrix} -\bar{Y}(p-1) & -\bar{Y}(p-2) & \dots & -\bar{Y}(0) \\ -\bar{Y}(p) & -\bar{Y}(p-1) & \dots & -\bar{Y}(1) \\ \dots & \dots & \dots & \dots \\ -\bar{Y}(N-2) & -\bar{Y}(N-3) & \dots & -\bar{Y}(N-p-1) \end{bmatrix}$$

$$\mathbf{X}_2 = \mathbf{Z} \begin{bmatrix} \mathbf{d}_0(0)' \\ \mathbf{d}_0(1)' \\ \dots \\ \mathbf{d}_0(N-1)' \end{bmatrix}$$

das lineare Gleichungssystem

$$\mathbf{X} \boldsymbol{\beta} + \mathbf{Z} \mathbf{r} = \mathbf{Z} \bar{\mathbf{y}} + \mathbf{e} \quad (355.6)$$

mit dem Vektor der unbekannt Parameter

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}, \quad (355.7)$$

der sich zusammensetzt aus

$$\boldsymbol{\beta}_1 = [\delta a_1, \dots, \delta a_p]', \quad (355.8)$$

$$\boldsymbol{\beta}_2 = \delta \mathbf{c} = [\delta c_1, \dots, \delta c_q]'. \quad (355.9)$$

Nach (325.3) und (326.4) gilt für die Erwartungswerte von \mathbf{e} und \mathbf{r}

$$E(\mathbf{e}) = \mathbf{0} \quad \text{und} \quad E(\mathbf{r}) = \mathbf{0}. \quad (355.10)$$

Die Matrizen $D(\mathbf{e})$ und $D(\mathbf{r})$ der Autokovarianzen von \mathbf{e} und \mathbf{r} sind mit (325.4) und (326.5) durch

$$D(\mathbf{e}) = \sigma_e^2 \mathbf{I} \quad \text{und} \quad D(\mathbf{r}) = \sigma_r^2 \mathbf{I} \quad (355.11)$$

gegeben. Man erhält wegen (326.12)

$$D\left(\begin{bmatrix} \mathbf{r} \\ \mathbf{e} \end{bmatrix}\right) = \begin{bmatrix} \sigma_r^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_e^2 \mathbf{I} \end{bmatrix}. \quad (355.12)$$

Mit der Transformation

$$\mathbf{y} = \mathbf{Z} \bar{\mathbf{y}} \quad (355.13)$$

ergibt sich aus (355.6)

$$\mathbf{X} \boldsymbol{\beta} + [\mathbf{Z}, -\mathbf{I}] \begin{bmatrix} \mathbf{r} \\ \mathbf{e} \end{bmatrix} = \mathbf{y}$$

und schließlich mit $\bar{\mathbf{Z}} = [\mathbf{Z}, -\mathbf{I}]$ sowie $\bar{\boldsymbol{\gamma}} = [\mathbf{r}', \mathbf{e}']'$ das gemischte Modell für die Parameterschätzung (Koch 1987, S.244), falls die Elemente der Matrix \mathbf{X} als fest angenommen werden

$$\mathbf{X} \boldsymbol{\beta} + \bar{\mathbf{Z}} \bar{\boldsymbol{\gamma}} = \mathbf{y} \quad \text{mit} \quad D(\bar{\boldsymbol{\gamma}}) = D\left(\begin{bmatrix} \mathbf{r} \\ \mathbf{e} \end{bmatrix}\right) = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} . \quad (355.14)$$

Die Matrix $D(\mathbf{y})$ der Autokovarianzen von \mathbf{y} berechnet sich aus (355.14) zu

$$\begin{aligned} D(\mathbf{y}) &= \bar{\mathbf{Z}} D(\bar{\boldsymbol{\gamma}}) \bar{\mathbf{Z}}' \\ &= [\mathbf{Z}, -\mathbf{I}] \begin{bmatrix} \sigma_r^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_e^2 \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{Z}' \\ -\mathbf{I} \end{bmatrix} \\ &= \sigma_r^2 \mathbf{Z} \mathbf{Z}' + \sigma_e^2 \mathbf{I} = \boldsymbol{\Sigma}_{yy} . \end{aligned} \quad (355.15)$$

Die Normalgleichungen des gemischten Modells folgen aus der Methode der kleinsten Quadrate mit (Koch 1987, S.249)

$$\begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{k} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} . \quad (355.16)$$

Hierin bedeutet \mathbf{k} der Vektor der Lagrangeschen Multiplikatoren. Bevor die Schätzwerte berechnet werden können, sind die unbekanntenen Varianzen σ_r^2 und σ_e^2 in (355.15) zu bestimmen, wozu auf die Varianzkomponentenschätzung (Koch 1987, S.259) zurückgegriffen wird. Setzt man in (355.14)

$$\bar{\mathbf{e}} = -\bar{\mathbf{Z}} \bar{\boldsymbol{\gamma}} , \quad (355.17)$$

erhält man mit (355.15) sowie $\mathbf{V}_r = \mathbf{Z} \mathbf{Z}'$ und $\mathbf{V}_e = \mathbf{I}$ das Gauß-Markoff-Modell

$$\mathbf{X} \boldsymbol{\beta} = \mathbf{y} + \bar{\mathbf{e}} \quad \text{mit} \quad D(\mathbf{y}) = \sigma_r^2 \mathbf{V}_r + \sigma_e^2 \mathbf{V}_e \quad (355.18)$$

mit den unbekanntenen Varianzkomponenten σ_r^2 und σ_e^2 . Es werden zunächst die Varianzkomponenten σ_r^2 und σ_e^2 geschätzt und die Ergebnisse in (355.15) eingesetzt. Man erhält die Schätzung

$$\hat{D}(\mathbf{y}) = \hat{\sigma}_r^2 \mathbf{V}_r + \hat{\sigma}_e^2 \mathbf{V}_e = \hat{\boldsymbol{\Sigma}}_{yy} , \quad (355.19)$$

die in (355.16) anstelle von $\boldsymbol{\Sigma}_{yy}$ eingeführt, die Schätzung für $\boldsymbol{\beta}$

$$\hat{\beta} = (X' \hat{\Sigma}_{yy}^{-1} X)^{-1} X' \hat{\Sigma}_{yy}^{-1} y \quad (355.20)$$

sowie

$$k = \hat{\Sigma}_{yy}^{-1} (y - X \hat{\beta}) \quad (355.21)$$

liefert. Die Lösung (355.20) enthält nach (355.7) die Schätzwerte für die p autoregressiven Parameter und die q Trendparameter. Die Schätzwerte des Zufallsvektors \bar{y} in (355.14) berechnen sich zu (Koch 1987, S.249)

$$\hat{\bar{y}} = \Sigma_{yy} \bar{Z}' k \quad (355.22)$$

oder aufgespalten in \hat{r} und \hat{e}

$$\hat{r} = \hat{\sigma}_r^2 Z' k \quad (355.23)$$

$$\hat{e} = -\hat{\sigma}_e^2 k \quad (355.24)$$

Die gefilterten Beobachtungen $\hat{Y}(n)$ ergeben sich mit (355.23) zu

$$\hat{Y}(n) = Y(n) - \hat{r}(n) \quad (355.25)$$

für $n \in \{0, \dots, N-1\}$. Weiter erhält man mit dem geschätzten Trendanteil

$$\hat{b}(n) = b(n, \hat{\beta}_2) = b(n, \hat{\delta c}_1, \dots, \hat{\delta c}_q) \quad (355.26)$$

das geschätzte Signal

$$\hat{X}(n) = \hat{Y}(n) - \hat{b}(n) \quad (355.27)$$

für $n \in \{0, \dots, N-1\}$, das sich für $n \in \{p, \dots, N-1\}$ auch durch

$$\hat{X}(n) = \sum_{k=1}^p \hat{a}_k \hat{X}(n-k) + \hat{\varepsilon}(n) \quad (355.28)$$

nach (326.2) darstellen läßt.

36 Zufallsfelder

361 Definitionen

Wie bereits im Kapitel 311 erwähnt, bezeichnet man die Folge von Zufallsvariablen

$$X(t, s) \quad (361.1)$$

als *n-dimensionalen Zufallsprozeß* oder *n-dimensionales Zufallsfeld*, falls der Index t der Indexmenge T nicht mehr skalarwertig ist, sondern einen n -dimensionalen Vektor darstellt, beispielsweise

$$t = [t_1, t_2, \dots, t_n]' .$$

Abgekürzt wird wieder

$$X(t) \quad (361.2)$$

geschrieben, da das Zufallsfeld als Folge vom Indexvektor t abhängiger Zufallsvariablen interpretierbar ist.

Die Elemente s der Menge S der Elementarereignisse können wieder abzählbar oder überabzählbar sein. Für einen vorgegebenen Vektor t bildet dann $X(t)$ eine diskrete oder im zweiten Fall eine stetige Zufallsvariable. Ebenso können die Elemente t der Indexmenge T abzählbar oder überabzählbar sein. Im ersten Fall erhält man ein Zufallsfeld an diskreten Punkten oder ein *diskretes Zufallsfeld*, während im zweiten Fall das Zufallsfeld kontinuierlich vorliegt. Als Beispiel eines zweidimensionalen diskreten Zufallsfeldes seien die Grauwerte in den Bildelementen eines digitalen Bildes genannt, wobei die Zufallsvariablen, die die Grauwerte repräsentieren, diskret oder stetig sein können.

Ist das Zufallsfeld nicht skalarwertig, sondern besitzt es mit

$$X(t) = [X_1(t), X_2(t), \dots, X_m(t)]' \quad (361.3)$$

m Komponenten, wird es als *vektorielles Zufallsfeld* bezeichnet. Beispiel eines vektoriellem Zufallsfeldes sind die Geschwindigkeitsvektoren einer Flüssigkeit in Abhängigkeit von der Zeit und von vorgegebenen Punkten eines Raumes. Da die Geschwindigkeitsvektoren drei Komponenten besitzen, liegt für jeden Indexvektor t eine dreidimensionale Zufallsvariable $X(t) = (X_i(t))$ mit $i \in \{1, 2, 3\}$ vor. Der Indexvektor t bezeichnet die Zeit und die Punkte im dreidimensionalen Raum, so daß $t \in \mathbb{R}^4$ gilt.

Mit den Indexvektoren $t \in \{t_0, \dots, t_n\}$ läßt sich für das skalarwertige Zufallsfeld (361.2) wie in (311.4) die Verteilungsfunktion

$$F(x_0, \dots, x_n, t_0, \dots, t_n) = P(X(t_0) < x_0, \dots, X(t_n) < x_n) \quad (361.4)$$

angeben, worin F die Verteilungsfunktion und P die Wahrscheinlichkeit des Ereignisses $X(t_0) < x_0, \dots, X(t_n) < x_n$ für beliebige x_0, \dots, x_n bedeutet. Ersetzt man in (361.4) die eindimensionale Zufallsvariable durch eine mehrdimensionale Zufallsvariable, wird die Verteilungsfunktion für das vektorielle Zufallsfeld (361.3) definiert.

362 Momentfunktionen, Homogenität und Isotropie

Für das Zufallsfeld $X(\mathbf{t})$ aus (361.2) erhält man mit der aus (361.4) folgenden Wahrscheinlichkeitsdichte die Erwartungswertfunktion $\mu(\mathbf{t})$ ganz entsprechend wie in (312.1) zu

$$\mu(\mathbf{t}) = E(X(\mathbf{t})) \quad (362.1)$$

Analog sind die Momentfunktion $K(\mathbf{t}_1, \mathbf{t}_2)$

$$K(\mathbf{t}_1, \mathbf{t}_2) = E(X(\mathbf{t}_1)X(\mathbf{t}_2)) \quad (362.2)$$

und die Autokovarianzfunktion $C(\mathbf{t}_1, \mathbf{t}_2)$

$$C(\mathbf{t}_1, \mathbf{t}_2) = E([X(\mathbf{t}_1) - \mu(\mathbf{t}_1)][X(\mathbf{t}_2) - \mu(\mathbf{t}_2)]) \quad (362.3)$$

definiert.

Die der Stationarität bei stochastischen Prozessen entsprechenden Eigenschaften bezeichnet man bei Zufallsfeldern als *Homogenität*. Die Verteilungsfunktion eines homogenen Zufallsfeldes ist also invariant gegenüber Verschiebungen des Indexvektors \mathbf{t} . Ein homogenes Zufallsfeld besitzt daher analog zu (313.2) einen konstanten Erwartungswert

$$E(X(\mathbf{t})) = \mu \quad (362.4)$$

und (313.3) entsprechend eine nur vom Differenzvektor $\boldsymbol{\tau}$ abhängige Momentfunktion

$$K(\mathbf{t}_1, \mathbf{t}_2) = K(\boldsymbol{\tau}) \quad \text{mit} \quad \boldsymbol{\tau} = \mathbf{t}_2 - \mathbf{t}_1 \quad (362.5)$$

sowie (313.6) entsprechend eine nur von $\boldsymbol{\tau}$ abhängige Autokovarianzfunktion

$$C(\mathbf{t}_1, \mathbf{t}_2) = C(\boldsymbol{\tau}) \quad \text{mit} \quad C(\boldsymbol{\tau}) = C(-\boldsymbol{\tau}) \quad (362.6)$$

Ein Zufallsfeld bezeichnet man als *homogen* und *isotrop*, das heißt richtungsunabhängig, falls die Momentfunktion und damit auch die Autokovarianzfunktion nur vom Betrag von $\boldsymbol{\tau}$ abhängt, also

$$C(\mathbf{t}_1, \mathbf{t}_2) = C(|\boldsymbol{\tau}|) \quad \text{mit} \quad |\boldsymbol{\tau}| = (\boldsymbol{\tau}' \boldsymbol{\tau})^{1/2} \quad (362.7)$$

In der statistischen Beschreibung der Turbulenz wird mit solchen Feldern gearbeitet

(Monin und Yaglom 1975, S.113).

Für ein vektorielles Zufallsfeld $X(\mathbf{t})$ nach (361.3) mit

$$\mathbf{X}(\mathbf{t}) = [X_1(\mathbf{t}), X_2(\mathbf{t}), \dots, X_m(\mathbf{t})]'$$

erhält man die Erwartungswertfunktion $\mu_i(\mathbf{t})$ der i -ten Komponente $X_i(\mathbf{t})$ von $\mathbf{X}(\mathbf{t})$ nach (362.1) zu

$$\mu_i(\mathbf{t}) = E(X_i(\mathbf{t})) \quad \text{für } i \in \{1, \dots, m\} \quad (362.8)$$

sowie die Momentfunktion $K_{ij}(\mathbf{t}_1, \mathbf{t}_2)$ von $X_i(\mathbf{t}_1)$ und $X_j(\mathbf{t}_2)$ nach (362.2) zu

$$K_{ij}(\mathbf{t}_1, \mathbf{t}_2) = E(X_i(\mathbf{t}_1)X_j(\mathbf{t}_2)) \quad \text{für } i, j \in \{1, \dots, m\} \quad (362.9)$$

und die Autokovarianzfunktion $C_{ij}(\mathbf{t}_1, \mathbf{t}_2)$ nach (362.3) zu

$$C_{ij}(\mathbf{t}_1, \mathbf{t}_2) = E([X_i(\mathbf{t}_1) - \mu_i(\mathbf{t}_1)][X_j(\mathbf{t}_2) - \mu_j(\mathbf{t}_2)]) \quad (362.10)$$

Die Erwartungswertfunktion des vektorielles Zufallsfeldes $\mathbf{X}(\mathbf{t})$ wird durch den $m \times 1$ Vektor mit den Komponenten $\mu_i(\mathbf{t})$ aus (362.8) bestimmt und die Momentfunktion des ersten Moments sowie die Autokovarianzfunktion durch die $m \times m$ Matrizen mit den Elementen $K_{ij}(\mathbf{t}_1, \mathbf{t}_2)$ aus (362.9) und $C_{ij}(\mathbf{t}_1, \mathbf{t}_2)$ aus (362.10).

Für ein *homogenes vektorielles Zufallsfeld* erhält man mit (362.5) aus (362.9)

$$K_{ij}(\mathbf{t}_1, \mathbf{t}_2) = K_{ij}(\boldsymbol{\tau}) \quad \text{mit } \boldsymbol{\tau} = \mathbf{t}_2 - \mathbf{t}_1 \quad (362.11)$$

und mit (362.6) aus (362.10)

$$C_{ij}(\mathbf{t}_1, \mathbf{t}_2) = C_{ij}(\boldsymbol{\tau}) \quad (362.12)$$

Liegt ein *homogenes und isotropes vektorielles Zufallsfeld* vor, lassen sich für die Autokovarianzfunktion im Fall der Dimension $m = 3$ vereinfachende Beziehungen angeben, die als *Taylor-Karman-Struktur* bezeichnet werden. Sie wurden für die statistische Beschreibung von Turbulenzen entwickelt (Monin und Yaglom 1975, S.39) und finden beispielsweise in der Erdmessung Anwendung (Meier und Keller 1990, S.95), aber auch bei der Untersuchung geodätischer Netze (Grafarend und Schaffrin 1979).

Beispiel 1: Die Höhen in den Punkten P_i mit den ebenen Koordinaten $\mathbf{t}_i = [x_i, y_i]'$ bilden das Zufallsfeld $X(\mathbf{t}_i)$. Das Zufallsfeld sei homogen und isotrop. Dann ist nach (362.7) die Autokovarianzfunktion lediglich abhängig vom Abstand r der Punkte P_i und P_j mit

$$r = ((x_j - x_i)^2 + (y_j - y_i)^2)^{1/2}.$$

Die Autokovarianzfunktion läßt sich dann beispielsweise nach (313.19) darstellen durch (Koch 1973)

$$C(r) = \frac{\sigma^2}{1+(r/a)^2} . \quad (362.13)$$

363 Spektrum

Die spektrale Darstellung (314.10) eines eindimensionalen stationären Prozesses kann wie folgt auf ein n-dimensionales homogenes Zufallsfeld $X(\mathbf{t})$ erweitert werden

$$X(\mathbf{t}) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{j\boldsymbol{\omega}' \mathbf{t}} dD(\boldsymbol{\omega}) . \quad (363.1)$$

Hierin bedeuten \mathbf{t} der $n \times 1$ Indexvektor, $\boldsymbol{\omega}$ der $n \times 1$ Vektor von Frequenzen und $dD(\boldsymbol{\omega})$ das Differential des komplexwertigen Zufallsfeldes $D(\boldsymbol{\omega})$ mit unabhängigen Zuwächsen. Die Integration umfaßt n Integrale. Entsprechend ergibt sich auch das Theorem von Wiener-Khinchine und damit das Spektrum $S(\boldsymbol{\omega})$ eines n-dimensionalen homogenen Zufallsfeldes als n-dimensionale Fourier-Transformation (231.18) der Autokovarianzfunktion $C(\boldsymbol{\tau})$ aus (362.6), in der $\boldsymbol{\tau}$ ein $n \times 1$ Vektor bedeutet,

$$S(\boldsymbol{\omega}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} C(\boldsymbol{\tau}) e^{-j\boldsymbol{\omega}' \boldsymbol{\tau}} d\boldsymbol{\tau} \quad (363.2)$$

und

$$C(\boldsymbol{\tau}) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} S(\boldsymbol{\omega}) e^{j\boldsymbol{\omega}' \boldsymbol{\tau}} d\boldsymbol{\omega} . \quad (363.3)$$

Das Spektrum eines homogenen vektoriellen Zufallsfeldes $X(\mathbf{t})$ mit m Komponenten wird durch eine $m \times m$ Matrix gebildet. Zwischen den Elementen $S_{ij}(\boldsymbol{\omega})$ dieser Matrix und den Elementen $C_{ij}(\boldsymbol{\tau})$ aus (362.12) der $m \times m$ Matrix der Autokovarianzfunktionen bestehen die Beziehungen

$$S_{ij}(\boldsymbol{\omega}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} C_{ij}(\boldsymbol{\tau}) e^{-j\boldsymbol{\omega}' \boldsymbol{\tau}} d\boldsymbol{\tau} \quad (363.4)$$

und

$$C_{ij}(\boldsymbol{\tau}) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} S_{ij}(\boldsymbol{\omega}) e^{j\boldsymbol{\omega}' \boldsymbol{\tau}} d\boldsymbol{\omega} . \quad (363.5)$$

Für ein homogenes diskretes Zufallsfeld der Dimension n wird wieder angenommen, daß mit (232.1) und (232.6) die diskrete Autokovarianzfunktion $C(k_1, \dots, k_n)$ durch Abtastung aus der kontinuierlichen Autokovarianzfunktion $C(\boldsymbol{\tau})$ erhalten wird

$$C(k_1, \dots, k_n) = C(\tau_1 = k_1 \Delta \tau_1, \dots, \tau_n = k_n \Delta \tau_n) .$$

Anstelle von (363.2) ergibt sich dann mit (232.7) analog zu (314.21) das Spektrum des

homogenen diskreten Zufallsfeldes der Dimension n zu

$$S(\omega_1, \dots, \omega_n) = \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_n=-\infty}^{\infty} C(k_1, \dots, k_n) e^{-j(k_1 \Delta\tau_1 \omega_1 + \dots + k_n \Delta\tau_n \omega_n)} \quad (363.6)$$

und anstelle von (363.3) mit (232.8) analog zu (314.22)

$$C(k_1, \dots, k_n) = \frac{\Delta\tau_1 \dots \Delta\tau_n}{(2\pi)^n} \int_{-\pi/\Delta\tau_n}^{\pi/\Delta\tau_n} \dots \int_{-\pi/\Delta\tau_1}^{\pi/\Delta\tau_1} S(\omega_1, \dots, \omega_n) e^{j(k_1 \Delta\tau_1 \omega_1 + \dots + k_n \Delta\tau_n \omega_n)} d\omega_1 \dots d\omega_n \quad (363.7)$$

Setzt man noch $\Omega_i = \Delta\tau_i \omega_i$, führt man also wie bei (233.14) die normierten Frequenzen Ω_i ein, ergibt sich (314.25) entsprechend

$$S(\Omega_1, \dots, \Omega_n) = \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_n=-\infty}^{\infty} C(k_1, \dots, k_n) e^{-j(k_1 \Omega_1 + \dots + k_n \Omega_n)} \quad (363.8)$$

und

$$C(k_1, \dots, k_n) = \frac{1}{(2\pi)^n} \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} S(\Omega_1, \dots, \Omega_n) e^{j(k_1 \Omega_1 + \dots + k_n \Omega_n)} d\Omega_1 \dots d\Omega_n \quad (363.9)$$

Beispiele für Spektren diskreter Zufallsfelder befinden sich in den Kapiteln 377 und 378.

Das Spektrum eines homogenen diskreten vektoriellen Zufallsfeldes ergibt sich analog zu (363.4) und (363.5).

37 Spezielle Zufallsfelder

371 Weißes Rauschen und Gauß-Zufallsfelder

Von den im Kapitel 321 behandelten speziellen Zufallsprozessen lassen sich das weiße Rauschen und der Gauß-Prozeß bequem auf Zufallsfelder verallgemeinern. Ein Zufallsfeld $X(t)$ bezeichnet man als *weißes Rauschen*, wenn die Zufallsvariablen $X(t_1)$ und $X(t_2)$ für $t_1 \neq t_2$ voneinander unabhängig sind, so daß (321.1) entsprechend für die Autokovarianzfunktion aus (362.3) gilt

$$C(t_1, t_2) = 0 \quad \text{für } t_1 \neq t_2 . \quad (371.1)$$

Ein Zufallsfeld $X(t)$ bezeichnet man als *Gauß-Zufallsfeld*, wenn für beliebige Indexvektoren t mit $t \in \{t_0, \dots, t_n\}$ die entsprechenden Zufallsvariablen, die im $(n+1) \times 1$ Vektor x zusammengefaßt seien, also $x = (X(t_i))$, die multivariate Normalverteilung mit den Parametern μ und Σ besitzen

$$x \sim N(\mu, \Sigma) . \quad (371.2)$$

Entsprechend (321.8) und (321.9) gilt für die Erwartungswertfunktion des Gauß-Zufallsfeldes

$$E(X(t_i)) = \mu_i \quad \text{mit } \mu = (\mu_i) \quad (371.3)$$

und für die Autokovarianzfunktion

$$C(t_i, t_j) = \sigma_{ij} \quad \text{mit } \Sigma = (\sigma_{ij}) . \quad (371.4)$$

Beispiele für Zufallsfelder mit weißem Rauschen und für Gauß-Zufallsfelder befinden sich in den Kapiteln 375 bis 379.

Für ein *vektorielles Zufallsfeld* $X(t)$

$$X(t) = [X_1(t), X_2(t), \dots, X_m(t)]'$$

mit *weißem Rauschen* gilt (371.1) für die Autokovarianzfunktion $C_{ij}(t_1, t_2)$ von $X_i(t)$ und $X_j(t)$ mit $i, j \in \{1, \dots, m\}$, also

$$C_{ij}(t_1, t_2) = 0 \quad \text{für } t_1 \neq t_2 . \quad (371.5)$$

Für ein *vektorielles Gauß-Zufallsfeld* $X(t)$ mit den Indexvektoren $t \in \{t_0, \dots, t_n\}$ besitzen die Zufallsvariablen

$$X_1(t_0), \dots, X_m(t_0), X_1(t_1), \dots, X_m(t_1), \dots, X_1(t_n), \dots, X_m(t_n)$$

eine gemeinsame multivariate Normalverteilung, die (371.2) entsprechend definiert ist.

372 Markoff-Zufallsfelder

Wie bereits im Zusammenhang mit der symmetrischen Markoff-Eigenschaft (322.12) erwähnt, lassen sich die Markoff-Prozesse auf Zufallsfelder verallgemeinern. Voraussetzung hierfür ist, daß ein Nachbarschaftssystem für das Zufallsfeld definiert ist. Dann läßt sich die Markoff-Eigenschaft (322.12) mit Hilfe der Nachbarschaften formulieren. Aufgrund der Markoff-Eigenschaft hängt die bedingte Verteilung einer Zufallsvariablen des Zufallsfeldes nur von den Werten der Zufallsvariablen der Nachbarschaft ab. Dies erleichtert bei vielen Anwendungen, zum Beispiel in der digitalen Bildverarbeitung, ganz wesentlich die Aufstellung von Verteilungen.

Im folgenden sollen die Elemente t_i der Indexmenge T mit $i \in \{0, \dots, n\}$ abzählbar sein, so daß ein diskretes Zufallsfeld vorliegt. Die Indexvektoren t_i lassen sich dann als Knoten eines Graphen interpretieren. Zwei Knoten t_i und t_j mit $i \neq j$ bezeichnet man als *Nachbarn*, falls eine Kante des Graphen sie verbindet. Die Menge der Knoten, die Nachbarn des Knotens t_i sind, bilden die *Nachbarschaft* $\partial\{t_i\}$ des Knotens t_i . Die Menge der Nachbarschaften, die sich mit Hilfe des Graphen entwickeln lassen, bildet das *Nachbarschaftssystem* des Zufallsfeldes.

Bezeichnet man mit $\{t_i\}$ die Menge, die als Element nur den Indexvektor t_i enthält, dann läßt sich die Wahrscheinlichkeit des Ereignisses $X(t_i) < x_i$ unter der Bedingung einführen, daß die Zufallsvariablen $X(t_j)$ an den übrigen Knoten des Graphen die Werte x_j annehmen

$$P(X(t_i) < x_i | X(t_j) = x_j \text{ für alle } t_j \in T \setminus \{t_i\}) . \quad (372.1)$$

Diese Wahrscheinlichkeit wird als *lokale Charakteristik* bezeichnet. Sie entspricht der linken Seite in der symmetrischen Markoff-Eigenschaft (322.12). Man bezeichnet nun das Zufallsfeld $X(t)$ als *Markoff-Zufallsfeld*, falls gilt

$$\begin{aligned} P(X(t_i) < x_i | X(t_j) = x_j \text{ für alle } t_j \in T \setminus \{t_i\}) \\ = P(X(t_i) < x_i | X(t_j) = x_j \text{ für alle } t_j \in \partial\{t_i\}) . \end{aligned} \quad (372.2)$$

Die bedingte Wahrscheinlichkeit des Ereignisses $X(t_i) < x_i$ hängt also nur noch von den Werten der Zufallsvariablen $X(t_j)$ in der Nachbarschaft $\partial\{t_i\}$ von t_i ab.

Wie in (322.3) und (322.4) läßt sich die Markoff-Eigenschaft (372.2) über die bedingte Verteilungsfunktion mit Hilfe der bedingten Dichte eines Markoff-Zufallsfeldes ausdrücken

$$p(x_i | x_0, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = p(x_i | \partial x_i) , \quad (372.3)$$

falls ∂x_i die Werte der Zufallsvariablen $X(t_j)$ in der Nachbarschaft $\partial\{t_i\}$ von t_i bezeichnen. Diese Beziehung gibt wie (372.2) die Markoff-Eigenschaft für Zufallsfelder an. Ist t_i mit $t_i = t_i$ ein Skalar und bezeichnet ∂x_i die Werte x_{i-1} und x_{i+1} der benachbarten Zufallsvariablen $X(t_{i-1})$ und $X(t_{i+1})$, führt (372.3) unmittelbar auf die symmetrische Markoff-Eigenschaft (322.12).

Für ein vektorielles Zufallsfeld mit

$$\mathbf{X}(t) = [X_1(t), X_2(t), \dots, X_m(t)]'$$

nach (361.3) läßt sich (372.1) entsprechend ebenfalls eine lokale Charakteristik definieren, falls x_{ik} einen beliebigen Wert der Komponente $X_k(t_i)$ des vektoriellen Zufallsfeldes bedeutet,

$$\begin{aligned} P(X_1(t_i) < x_{i1}, X_2(t_i) < x_{i2}, \dots, X_m(t_i) < x_{im} | X_1(t_j) = x_{j1}, \\ X_2(t_j) = x_{j2}, \dots, X_m(t_j) = x_{jm} \text{ für alle } t_j \in T \setminus \{t_i\}) . \end{aligned} \quad (372.4)$$

Die lokale Charakteristik bezeichnet die Wahrscheinlichkeit des Ereignisses

$$X_1(t_i) < x_{i1}, X_2(t_i) < x_{i2}, \dots, X_m(t_i) < x_{im}$$

für den Zufallsvektor $X(t_i)$ mit dem Indexvektor t_i unter der Bedingung, daß die Zufallsvektoren $X(t_j)$ an allen übrigen Knoten t_j des Graphen vorgegebene Werte annehmen. Entsprechend (372.2) liegt nun ein *vektorielles Markoff-Zufallsfeld* vor, falls gilt

$$\begin{aligned} P(X_1(t_i) < x_{i1}, \dots, X_m(t_i) < x_{im} | X_1(t_j) = x_{j1}, \dots, X_m(t_j) = x_{jm} \\ \text{für alle } t_j \in T \setminus \{t_i\}) \\ = P(X_1(t_i) < x_{i1}, \dots, X_m(t_i) < x_{im} | X_1(t_j) = x_{j1}, \dots, X_m(t_j) = x_{jm} \\ \text{für alle } t_j \in \partial\{t_i\}) . \end{aligned} \quad (372.5)$$

Die bedingte Wahrscheinlichkeit eines Ereignisses des Zufallsvektors $X(t_i)$ mit dem Indexvektor t_i hängt damit nur noch von den Werten des Zufallsvektors $X(t_j)$ in der Nachbarschaft $\partial\{t_i\}$ von t_i ab. Führt man wieder wie in (322.3) die bedingte Verteilungsfunktion ein, folgt anstelle von (372.5) für die bedingte Dichte des Zufallsvektors $X(t_i)$

$$\begin{aligned} p(x_{i1}, \dots, x_{im} | x_{01}, \dots, x_{0m}, \dots, x_{i-1,1}, \dots, x_{i-1,m}, \\ x_{i+1,1}, \dots, x_{i+1,m}, \dots, x_{n1}, \dots, x_{nm}) = p(x_{i1}, \dots, x_{im} | \partial x_{ik}) , \end{aligned} \quad (372.6)$$

falls ∂x_{ik} die Werte des Zufallsvektors $X(t_j)$ in der Nachbarschaft $\partial\{t_i\}$ von t_i bezeichnet.

373 Nachbarschafts-Gibbs-Felder

Für Zufallsfelder wird jetzt die Gibbs-Verteilung eingeführt. Ihr Ursprung liegt in der statistischen Physik, und sie definiert die Verteilung in Systemen, die sich im Gleichgewichtszustand befinden. Zur Beschreibung eines Ferromagneten benutzte Ising (1925) eine spezielle Gibbs-Verteilung, die in (376.26) angegeben und heute als Ising-Modell bezeichnet wird (Geman und Geman 1984; Kindermann und Snell 1980, S.2). Die Gibbs-Verteilung wird hier als Definition eingeführt, eine physikalische Herleitung findet man in (Landau und Lifschitz 1979, S.75).

Die Werte x_i einer Realisierung des diskreten Zufallsfeldes $X(t_i)$ mit $X(t_i) = x_i$ und $i \in \{0, 1, \dots, n\}$ seien im Vektor \mathbf{x} zusammengefaßt, also $\mathbf{x} = (x_i)$. Die Indexvektoren t_i der Indexmenge T bilden wieder die Knoten eines Graphen. Es sei A mit $A \subset T$ eine Teilmenge der Indexvektoren t_i , die eine gewisse Art der Konfiguration von Knoten im Graphen bilden. Alle möglichen Konfigurationen dieser Art im Graphen, also alle Teilmengen A seien in der Menge D zusammengefaßt. Für die Menge A wird nun die reellwertige Funktion $U_A(\mathbf{x})$, das sogenannte *Potential*, mit den Eigenschaften eingeführt

$$U_A(\mathbf{x}) = U_A(\mathbf{y}) \quad \text{mit} \quad x_i = y_i \quad \text{für alle} \quad t_i \in A, \quad (373.1)$$

wobei $\mathbf{y} = (y_i)$ den Vektor einer Realisierung bedeutet, und

$$U_A(\mathbf{x}) = 0 \quad \text{mit} \quad A = \emptyset. \quad (373.2)$$

Als *Energie* $U(\mathbf{x})$ des Potentials wird die Summe der Potentiale sämtlicher Teilmengen von A in T definiert

$$U(\mathbf{x}) = \sum_{A \in D} U_A(\mathbf{x}). \quad (373.3)$$

Bezeichnet nun $p(\mathbf{x})$ die Wahrscheinlichkeitsdichte des Zufallsfeldes $X(t)$, dann ist die *Gibbs-Verteilung* definiert durch

$$p(\mathbf{x}) = \frac{1}{Z} \exp(-U(\mathbf{x})), \quad (373.4)$$

wobei Z die Normierungskonstante bedeutet. Sie beträgt bei stetigen Zufallsvariablen, falls also die Elemente s der Menge S der Elementarereignisse in (361.1) überabzählbar sind,

$$Z = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp(-U(\mathbf{x})) dx_0 \dots dx_n. \quad (373.5)$$

Bei diskreten Zufallsvariablen oder abzählbaren Elementen s der Menge S wird die Normierungskonstante auch als *Zustandssumme*, im Englischen "partition function", bezeich-

net. Sie berechnet sich zu

$$Z = \sum_{\mathbf{x}} \exp(-U(\mathbf{x})) \quad , \quad (373.6)$$

wobei über sämtliche Werte von \mathbf{x} , also über sämtliche Realisierungen des Zufallsfeldes $X(t)$ zu summieren ist.

Ein Zufallsfeld, das die Gibbs-Verteilung (373.4) besitzt, bezeichnet man als *Gibbs-Zufallsfeld*. Sind die Werte \mathbf{x} des Zufallsfeldes derart beschaffen, daß die Energie $U(\mathbf{x})$ des Zufallsfeldes groß ist, dann sind die Dichtewerte $p(\mathbf{x})$ klein. Umgekehrt bewirkt eine geringe Energie große Dichtewerte.

Anstelle der Darstellung (373.4) für die Dichte der Gibbs-Verteilung findet man auch

$$p(\mathbf{x}) = \frac{1}{Z} \exp[-U(\mathbf{x})/T_0] \quad (373.7)$$

mit der entsprechenden Normierungskonstanten Z . Hierin bedeutet T_0 die Temperatur. Durch sie lassen sich die Maxima der Wahrscheinlichkeitsdichte steuern. Bei hohen Werten von T_0 ergibt sich eine flach verlaufende Dichtefunktion, während kleine Werte von T_0 die Maxima ausprägen. Die Steuerung durch die Temperatur T_0 spielt bei den stochastischen Verfahren zur Parameterschätzung eine Rolle, die im Kapitel 383 behandelt werden.

Bei den durch die Teilmenge A der Knoten t_i gekennzeichneten Konfigurationen sind wieder die von Interesse, die Nachbarschaften beinhalten. Diese Konfigurationen bezeichnet man als *Cliquen*. Es bedeute wieder $\partial\{t_i\}$ die Nachbarschaft des Knotens t_i und $\{t_i\}$ die Menge, die als Element nur den Indexvektor t_i enthält. Eine Teilmenge $C \subset T$ von Knoten der Indexmenge T ist eine Clique, falls

1. $C = \{t_i\}$ oder
2. $t_j \in \partial\{t_k\}$ für alle $t_j, t_k \in C$ und $t_j \neq t_k$ (373.8)

gilt. Eine Clique enthält also entweder ein Element oder, falls sie mehr als ein Element enthält, ist jedes Element der Clique Nachbar jedes anderen Elementes derselben Clique. Um diese Definition zu erläutern, sei auf die Abbildung 373-1 a) verwiesen, in der ein Graph dargestellt ist. Die Cliques, die sich mit dem Knoten des Graphen bilden lassen, der durch den gefüllten Kreis gekennzeichnet ist, sind in Abb. 373-1 b) angegeben.

Es sei nun K die Menge aller Cliques in T . Ein Gibbs-Zufallsfeld bezeichnet man als *Nachbarschafts-Gibbs-Feld*, falls für das Potential der Menge $A \subset T$ gilt

$$U_A(\mathbf{x}) = 0 \quad \text{für} \quad A \notin K \quad . \quad (373.9)$$

Die Dichte $p(\mathbf{x})$ des Nachbarschafts-Gibbs-Feldes ergibt sich dann aus (373.3) und

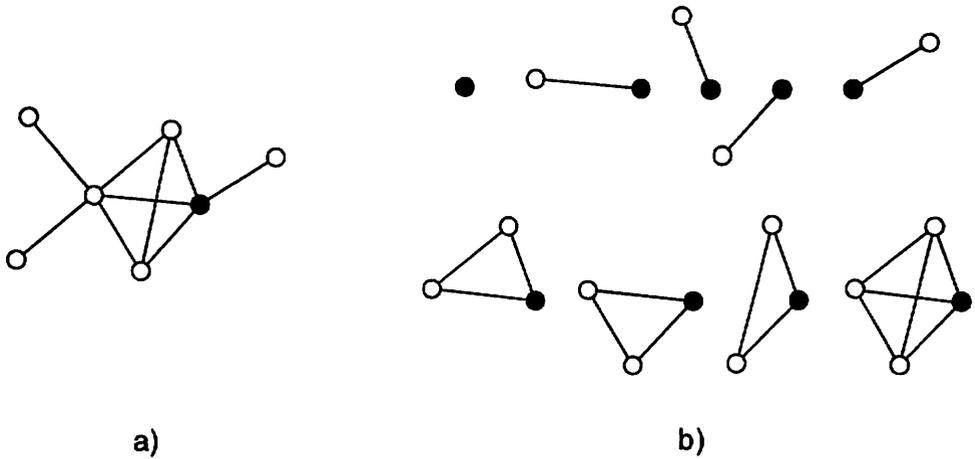


Abb. 373-1: a) Graph und b) Cliques mit dem durch den gefüllten Kreis gekennzeichneten Knoten

(373.4) zu

$$p(\mathbf{x}) = \frac{1}{Z} \exp(-U(\mathbf{x})) \quad (373.10)$$

mit

$$U(\mathbf{x}) = \sum_{C \in \mathcal{K}} U_C(\mathbf{x}) ,$$

worin Z wieder die Normierungskonstante aus (373.5) im Falle stetiger und aus (373.6) im Falle diskreter Zufallsvariablen bedeutet. Die Energie $U(\mathbf{x})$ des Nachbarschafts-Gibbs-Feldes erhält man also aus der Summe über die Potentiale $U_C(\mathbf{x})$ sämtlicher Cliques der Indexmenge T . Mit (373.10) ist die gemeinsame Verteilung eines Nachbarschafts-Gibbs-Feldes gegeben.

Für ein vektorielles Zufallsfeld

$$\mathbf{X}(t) = [X_1(t), \dots, X_m(t)]' \quad \text{und} \quad t \in \{t_0, \dots, t_n\}$$

läßt sich ebenfalls eine Gibbs-Verteilung definieren, so daß ein *vektorielles Gibbs-Zufallsfeld* erhalten wird. Bedeutet x_{ik} wieder einen Wert der Komponente k des Zufallsvektors $\mathbf{X}(t_i)$, also

$$X_k(t_i) = x_{ik} ,$$

und faßt man im Vektor \mathbf{x} die Realisierung des vektoriiellen Zufallsfeldes mit

$$\mathbf{x} = (x_{ik}) \quad \text{für} \quad i \in \{0, 1, \dots, n\}, \quad k \in \{1, 2, \dots, m\} \quad (373.11)$$

zusammen, kann man analog zu (373.1) bis (373.3) die Energie $U(\mathbf{x})$ definieren und erhält die Gibbs-Verteilung

$$p(\mathbf{x}) = \frac{1}{Z} \exp(-U(\mathbf{x})) \quad (373.12)$$

mit

$$Z = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp(-U(\mathbf{x})) dx_{01} \dots dx_{0m} \dots dx_{n1} \dots dx_{nm}$$

für stetige Zufallsvektoren und mit Z aus (373.6) für diskrete Zufallsvektoren, wobei wieder über sämtliche Realisierungen zu summieren ist. Für *vektorielle Nachbarschafts-Gibbs-Felder* erhält man schließlich die Energie $U(\mathbf{x})$ wie in (373.10) aus

$$U(\mathbf{x}) = \sum_{C \in K} U_C(\mathbf{x}) \quad (373.13)$$

374 Äquivalenz von Markoff-Zufallsfeldern und Nachbarschafts-Gibbs-Feldern

Zwar ist die Markoff-Eigenschaft (372.2) oder (372.3) leicht interpretierbar, doch erlaubt sie nicht, einen analytischen Ausdruck für die Dichte eines Markoff-Zufallsfeldes anzugeben. Dagegen ist die Dichte eines Nachbarschafts-Gibbs-Feldes mit (373.10) angebar, die Eigenschaft eines Nachbarschafts-Gibbs-Feldes aber auf den ersten Blick nicht überschaubar. Der jeweilige Mangel dieser beiden Zufallsfelder wird dadurch behoben, daß beide Zufallsfelder äquivalent sind. Markoff-Zufallsfelder besitzen also die bequem angebbaren Gibbs-Verteilungen.

Satz: Ein diskretes Zufallsfeld $X(t_i)$ mit $i \in \{0, 1, \dots, n\}$ ist genau dann ein Markoff-Zufallsfeld, wenn seine Dichte $p(\mathbf{x})$ durch die Gibbs-Verteilung (373.10) des Nachbarschafts-Gibbs-Feldes gegeben ist. (374.1)

Beweis: Es wird gezeigt, daß ein Nachbarschafts-Gibbs-Feld ein Markoff-Zufallsfeld ist. Für die linke Seite der Markoff-Eigenschaft (372.3) erhält man mit der Definition der bedingten Dichte (Koch 1987, S.105)

$$p(x_i | x_0, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = \frac{p(\mathbf{x})}{p(x_0, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)} \cdot$$

Die Randverteilung im Nenner der rechten Seite wird mit Hilfe ihrer Definition (Koch 1987, S.104) durch die gemeinsame Verteilung $p(\mathbf{x})$ ausgedrückt, wobei über die Werte x_j der Zufallsvariablen $X(t_j)$ zu integrieren ist. Man erhält

$$p(x_i | x_0, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = \frac{p(\mathbf{x})}{\int_{-\infty}^{\infty} p(x_0, x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) dx_i}$$

Für die gemeinsame Verteilung $p(\mathbf{x})$ wird jetzt die Gibbs-Verteilung (373.10) des Nachbarschafts-Gibbs-Feldes substituiert

$$p(x_i | x_0, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = \frac{\exp(-\sum_{C \in K} U_C(\mathbf{x}))}{\int_{-\infty}^{\infty} \exp(-\sum_{C \in K} U_C(x_0, x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n)) dx_i} \quad (374.2)$$

Im Nenner auf der rechten Seite lassen sich alle Cliques ausklammern, die den Knoten t_i nicht enthalten, und gegen die Cliques im Zähler kürzen. Die rechte Seite von (374.2) ist daher nur noch von t_i und seinen Nachbarn $\partial\{t_i\}$ abhängig. Man erhält

$$p(x_i | x_0, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = p(x_i | \partial x_i) = \frac{1}{Z_i} \exp(-\sum_{C(\mathbf{t}_i) \in K} U_C(\mathbf{x})), \quad (374.3)$$

worin ∂x_i wieder die Werte der Zufallsvariablen $X(\mathbf{t}_j)$ in der Nachbarschaft $\partial\{t_i\}$ von t_i bedeuten und über sämtliche Cliques $C(\mathbf{t}_i)$ zu summieren ist, die t_i enthalten.

Für die Normierungskonstante Z_i folgt

$$Z_i = \int_{-\infty}^{\infty} \exp(-\sum_{C(\mathbf{t}_i) \in K} U_C(\mathbf{x})) dx_i \quad (374.4)$$

oder im Falle einer diskreten Zufallsvariablen mit den Werten $X(\mathbf{t}_i) = x_{ij}$ sowie $\mathbf{x} = (x_{ij})$ durch Summation über den Index j , wobei j die diskreten Werte für $X(\mathbf{t}_i)$ bezeichnet

$$Z_i = \sum_j \exp(-\sum_{C(\mathbf{t}_i) \in K} U_C(\mathbf{x})) \quad (374.5)$$

Mit (374.3) ist die Markoff-Eigenschaft (372.3) erfüllt, so daß ein Nachbarschafts-Gibbs-Feld ein Markoff-Zufallsfeld ist. Außerdem wird mit (374.3) die bedingte Dichte $p(x_i | \partial x_i)$ der Zufallsvariablen $X(\mathbf{t}_i)$ erhalten.

Der Beweis der Umkehrung des Satzes (374.1) ist sehr viel aufwendiger und wenig instruktiv. Für diskrete Zufallsvariablen ist er bei Grimmet (1973) und Griffeath (1976, S.425) angegeben, siehe auch (Busch 1992, S.13). Der zeitlich erste Beweis, allerdings unter einschränkenden Voraussetzungen, stammt von Averintsev (1970). Bekannt wurde die Äquivalenz von Markoff-Zufallsfeldern und Nachbarschafts-Gibbs-Feldern durch das

unveröffentlichte Theorem von Hammersley und Clifford aus dem Jahre 1971. Der Beweis von Besag (1974) schließt die stetigen Zufallsvariablen ein. \square

Für vektorielle Nachbarschafts-Gibbs-Felder läßt sich ebenfalls zeigen, daß sie Markoff-Zufallsfelder sind. In (374.2) braucht dann lediglich die Randverteilung im Nenner der rechten Seite für sämtliche Komponenten des Zufallsvektors $\mathbf{X}(\mathbf{t}_i)$ berechnet zu werden. Die bedingte Dichte des Zufallsvektors $\mathbf{X}(\mathbf{t}_i)$ ergibt sich entsprechend (374.3) in der Notation der Markoff-Eigenschaft (372.6) des vektoriellen Zufallsfeldes zu

$$p(x_{i1}, \dots, x_{im} | \partial x_{ik}) = \frac{1}{Z_i} \exp\left(- \sum_{C(\mathbf{t}_i) \in K} U_C(\mathbf{x})\right) \quad (374.6)$$

mit

$$Z_i = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\left(- \sum_{C(\mathbf{t}_i) \in K} U_C(\mathbf{x})\right) dx_{i1} \dots dx_{im} \quad (374.7)$$

für stetige Zufallsvariable, wobei über die Werte der Komponenten des Zufallsvektors $\mathbf{X}(\mathbf{t}_i)$ zu integrieren ist. Für diskrete Zufallsvariable ergibt sich die Normierungskonstante durch Summation über sämtliche j Werte der m Komponenten des Zufallsvektors $\mathbf{X}(\mathbf{t}_i)$

$$Z_i = \sum_j \exp\left(- \sum_{C(\mathbf{t}_i) \in K} U_C(\mathbf{x})\right) . \quad (374.8)$$

Durch Verallgemeinerung des Beweises der Umkehrung des Satzes (374.1) auf vektorielle Zufallsfelder ist dann die Äquivalenz von vektoriellen Markoff-Zufallsfeldern und vektoriellen Nachbarschafts-Gibbs-Feldern gezeigt.

Für Markoff-Zufallsfelder ist folglich mit (373.10) die gemeinsame Dichte aller Zufallsvariablen und mit (374.3) die bedingte Dichte einer Zufallsvariablen angebar. Gleiches gilt mit (373.12) und (374.6) für vektorielle Markoff-Zufallsfelder. Beispiele für Markoff-Zufallsfelder mit Gibbs-Verteilungen folgen im nächsten Kapitel.

375 Markoff-Zufallsfelder mit Gibbs-Verteilungen

Im folgenden werden einige für die Anwendung interessante Gibbs-Verteilungen für Markoff-Zufallsfelder diskutiert. Begonnen wird mit einer allgemeinen Darstellung der Energie der Gibbs-Verteilung sowohl für skalarwertige als auch für vektorielle Nachbarschafts-Gibbs-Felder. Durch Einschränkung der Anzahl der Indexvektoren pro Clique wird dann spezialisiert. Als weitere Spezialisierung werden im nächsten Kapitel geeignete Nachbarschaften für Markoff-Zufallsfelder auf Gittern eingeführt.

a) Allgemeine Darstellung der Energie

Ein Markoff-Zufallsfeld $X(\mathbf{t}_i)$ mit $i \in \{0, 1, \dots, n\}$ sei derart definiert, daß der Indexvektor \mathbf{t}_i Nachbar eines jeden Vektors \mathbf{t}_j sei mit $i \neq j$ und $i, j \in \{0, 1, \dots, n\}$. Der Vektor \mathbf{x} mit $\mathbf{x} = (x_i)$ und $X(\mathbf{t}_i) = x_i$ enthalte wieder eine Realisierung des Zufallsfeldes. Die Energie $U(\mathbf{x})$ der Gibbs-Verteilung (373.10) läßt sich dann darstellen durch

$$\begin{aligned}
 U(\mathbf{x}) = & \sum_{i=0}^n x_i U_i(x_i) + \sum_{i=0}^n \sum_{j=i+1}^n x_i x_j U_{i,j}(x_i, x_j) \\
 & + \sum_{i=0}^n \sum_{j=i+1}^n \sum_{k=j+1}^n x_i x_j x_k U_{i,j,k}(x_i, x_j, x_k) + \dots \\
 & + x_0 x_1 \dots x_n U_{0,1,\dots,n}(x_0, x_1, \dots, x_n) .
 \end{aligned} \tag{375.1}$$

Hierin bezeichnet der erste Ausdruck auf der rechten Seite die Summe über die Potentiale der Cliques, die nur einen Indexvektor enthalten, der zweite Ausdruck die Summe über die Cliques mit zwei Indexvektoren und so fort. Die Summationsgrenzen folgen aus der Vorschrift, daß über jede Clique nur einmal zu summieren ist.

Liegt ein vektorielles Markoff-Zufallsfeld mit dem Vektor \mathbf{x} der Realisierung aus (373.11) vor, kann man die Energie $U(\mathbf{x})$ der Gibbs-Verteilung (373.12) darstellen durch

$$\begin{aligned}
 U(\mathbf{x}) = & \sum_{i=0}^n [x_{i1} U_i(x_{i1}) + x_{i2} U_i(x_{i2}) + \dots + x_{im} U_i(x_{im}) + x_{i1} x_{i2} U_i(x_{i1}, x_{i2}) + \dots \\
 & + x_{i1} x_{im} U_i(x_{i1}, x_{im}) + \dots + x_{i1} x_{i2} \dots x_{im} U_i(x_{i1}, x_{i2}, \dots, x_{im})] + \dots,
 \end{aligned} \tag{375.2}$$

worin die Summen über die Potentiale der Cliques, die mehr als einen Indexvektor enthalten, wie in (375.1) zu bilden sind.

Die Ausdrücke (375.1) und (375.2) geben allgemeine Darstellungen für die Energie $U(\mathbf{x})$ der Gibbs-Verteilungen (373.10) und (373.12) an. Durch Spezialisierung der Nachbarschaften und durch Spezifizierung der Potentiale können daraus verschiedene Verteilungen gewonnen werden.

b) Auto-Modell

Beschränkt man sich auf Cliques mit ein und zwei Indexvektoren, also auf Einer- und Zweier-Cliques, läßt sich die Energie $U(\mathbf{x})$ aus (375.1) mit den Konstanten $U_{i,j}(x_i, x_j) = \beta_{ij}$ vereinfachen zu

$$U(\mathbf{x}) = \sum_{i=0}^n x_i U_i(x_i) + \sum_{i=0}^n \sum_{j=i+1}^n \beta_{ij} x_i x_j . \tag{375.3}$$

Ein Markoff-Zufallsfeld mit dieser Energie einer Gibbs-Verteilung bezeichnet man als *Auto-Modell* (Besag 1974). Dieses Modell ist noch so allgemein, daß aus ihm eine Reihe von Spezialfällen zu entwickeln ist. Hierauf wird im folgenden eingegangen.

c) Gauß-Zufallsfeld als Spezialfall des Auto-Modells

Die bedingte Dichte (374.3) der Zufallsvariablen $X(\mathbf{t}_i)$ eines Markoff-Zufallsfeldes sei durch die Dichte einer Normalverteilung gegeben, und zwar durch

$$p(x_i | \partial x_i) \propto \exp\left\{-\frac{1}{2\sigma^2} \left[x_i - \mu_i - \sum_{\substack{j=0 \\ i \neq j}}^n \beta_{ij} (x_j - \mu_j) \right]^2 \right\}, \quad (375.4)$$

wobei \propto das Proportionalitätszeichen und μ_i sowie μ_j Konstanten bedeuten. Quadriert man die eckige Klammer im Exponenten, ergibt sich der dritte Term zu einer Konstanten, die in der Normierungskonstanten der Verteilung absorbiert werden kann, so daß folgt

$$p(x_i | \partial x_i) \propto \exp\left\{-\frac{1}{2\sigma^2} \left[(x_i - \mu_i)^2 - 2(x_i - \mu_i) \sum_{\substack{j=0 \\ i \neq j}}^n \beta_{ij} (x_j - \mu_j) \right] \right\}. \quad (375.5)$$

Der erste Term im Exponenten repräsentiert das Potential der Einer-Clique mit dem Indexvektor \mathbf{t}_i , der zweite Term die Potentiale der Zweier-Cliquen mit \mathbf{t}_i und \mathbf{t}_j für $i \neq j$.

Um aus der bedingten Dichte (375.5) die gemeinsame Dichte $p(\mathbf{x})$ zu erhalten, ist nach (373.10) über sämtliche Einer- und Zweier-Cliquen zu summieren. Der Summation in (375.1) entsprechend erhält man

$$p(\mathbf{x}) \propto \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=0}^n \left[(x_i - \mu_i)^2 - 2(x_i - \mu_i) \sum_{j=i+1}^n \beta_{ij} (x_j - \mu_j) \right] \right\}. \quad (375.6)$$

Ein Vergleich mit (375.3) zeigt, daß diese Verteilung einen Spezialfall des Auto-Modells darstellt. Da die Potentiale der Zweier-Cliquen unabhängig von der Reihenfolge der Indizes sind, muß

$$\beta_{ij} = \beta_{ji}$$

gelten, so daß mit $-\beta_{ii} = 1$ der Exponent sich umformen läßt

$$\begin{aligned} \sum_{i=0}^n \left[(x_i - \mu_i)^2 - 2(x_i - \mu_i) \sum_{j=i+1}^n \beta_{ij} (x_j - \mu_j) \right] &= \sum_{i=0}^n \sum_{j=0}^n (x_i - \mu_i) (-\beta_{ij}) (x_j - \mu_j) \\ &= (\mathbf{x} - \boldsymbol{\mu})' \mathbf{B} (\mathbf{x} - \boldsymbol{\mu}) \quad \text{mit } \boldsymbol{\mu} = (\mu_i) \quad \text{und } \mathbf{B} = (-\beta_{ij}). \end{aligned}$$

Anstelle von (375.6) ergibt sich dann

$$p(\mathbf{x}) \propto \exp\left\{-\frac{1}{2\sigma^2} (\mathbf{x}-\boldsymbol{\mu})' \mathbf{B}(\mathbf{x}-\boldsymbol{\mu})\right\} . \quad (375.7)$$

Setzt man voraus, daß \mathbf{B} positiv definit ist, folgt also für \mathbf{x} die Normalverteilung

$$\mathbf{x} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{B}^{-1}) . \quad (375.8)$$

Die bedingte Normalverteilung (375.4) für die Zufallsvariable $X(t_i)$ impliziert somit für das Zufallsfeld mit den Zufallsvariablen $X(t_i)$ mit $i \in \{0, 1, \dots, n\}$ das Gauß-Zufallsfeld mit der Verteilung (375.8). Weiter ist die Normalverteilung als Spezialfall des Auto-Modells eine Gibbs-Verteilung und damit das Gauß-Zufallsfeld auch ein Markoff-Zufallsfeld, also ein *Gauß-Markoff-Zufallsfeld*.

Das Ergebnis (375.8) ist lediglich die Bestätigung einer Beziehung für die bedingte Verteilung normalverteilter Zufallsvariablen. Denn gilt die Normalverteilung (375.8), erhält man mit

$$\mathbf{B} = \begin{bmatrix} 1 & -\beta_{01} & -\beta_{02} & \dots & -\beta_{0n} \\ -\beta_{10} & 1 & -\beta_{12} & \dots & -\beta_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ -\beta_{n0} & -\beta_{n1} & -\beta_{n2} & \dots & 1 \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{b}_{01}' \\ \mathbf{b}_{10} & \mathbf{B}_{11} \end{bmatrix}$$

und

$$\mathbf{x}_1 = [x_1, x_2, \dots, x_n]' , \quad \boldsymbol{\mu}_1 = [\mu_1, \mu_2, \dots, \mu_n]'$$

die bedingte Verteilung von x_0 zu

$$x_0 | \mathbf{x}_1 \sim N(\mu_0 + \mathbf{b}_{01}' (\mathbf{x}_1 - \boldsymbol{\mu}_1), \sigma^2)$$

oder

$$x_0 | \mathbf{x}_1 \sim N\left(\mu_0 - \sum_{j=1}^n \beta_{0j} (x_j - \mu_j), \sigma^2\right) .$$

Der Erwartungswert von x_0 und seine Varianz werden nämlich mit der ersten Zeile der Inversen von $\sigma^2 \mathbf{B}^{-1}$ gebildet, wie sich mit Hilfe des Beweises der bedingten Normalverteilung zeigen läßt (Koch 1987, S.138). Dies trifft nicht nur für die erste Zeile, sondern auch für die übrigen Zeilen zu, so daß die Verteilung (375.4) erhalten wird.

Der Erwartungswert der bedingten Normalverteilung (375.4) für $X(t_i)$ beträgt

$$E(X(t_i) | \partial x_i) = \mu_i - \sum_{\substack{j=0 \\ i \neq j}}^n \beta_{ij} (x_j - \mu_j) . \quad (375.9)$$

Aus dieser Darstellung wird die enge Verbindung der durch (375.4) und (375.8) charak-

terisierten Gauß-Zufallsfelder mit den im Kapitel 325 behandelten autoregressiven Prozessen deutlich.

d) Autologistisches Modell

Auf eine besonders einfache Verteilung führt das Auto-Modell (375.3), falls $X(t_i)$ binäre Zufallsvariable darstellen, die nur die Werte Null und Eins annehmen können. Setzt man zunächst $U(x_i) = \alpha_i$, ergibt sich die Energie

$$U(\mathbf{x}) = \sum_{i=0}^n x_i \alpha_i + \sum_{i=0}^n \sum_{j=i+1}^n \beta_{ij} x_i x_j . \quad (375.10)$$

Wird nur über die Cliques summiert, die den Indexvektor t_i enthalten, folgt aus (374.3) für die bedingte Dichte

$$p(x_i=0 | \partial x_i) = \frac{1}{Z_i}$$

und

$$p(x_i=1 | \partial x_i) = \frac{1}{Z_i} \exp\left\{-\left(\alpha_i + \sum_{\substack{j=0 \\ i \neq j}}^n \beta_{ij} x_j\right)\right\} .$$

Die Zustandssumme Z_i berechnet sich nach (374.5) aus der Summe der Exponentialfunktionen dieser beiden bedingten Dichten, und man erhält

$$p(x_i | \partial x_i) = \exp\left\{-x_i \left(\alpha_i + \sum_{\substack{j=0 \\ i \neq j}}^n \beta_{ij} x_j\right)\right\} / \left\{1 + \exp\left[-\left(\alpha_i + \sum_{\substack{j=0 \\ i \neq j}}^n \beta_{ij} x_j\right)\right]\right\} . \quad (375.11)$$

Ein Markoff-Zufallsfeld mit dieser bedingten Verteilung und der aus (375.10) nach (373.10) folgenden gemeinsamen Verteilung bezeichnet man als *autologistisches Modell* (Besag 1974).

e) Autobinomialmodell

Gegeben sei ein diskretes Markoff-Zufallsfeld $X(t)$ mit der Realisierung $X(t_i) = x_i$, wobei $x_i \in \{0, 1, \dots, l\}$ und $i \in \{0, 1, \dots, n\}$ gelte. Die bedingte Dichte (374.3) für $X(t_i)$ sei durch die Dichte der Binomialverteilung (Koch 1987, S.98) gegeben, in der die Anzahl der Versuche durch den Wert l gegeben ist und in der die Wahrscheinlichkeit $p(\partial x_i)$ des Erfolges von den Werten ∂x_i der Zufallsvariablen in der Nachbarschaft $\partial\{t_i\}$ des Indexvektors t_i abhängt. Man erhält

$$p(x_i | \partial x_i) = \binom{l}{x_i} p(\partial x_i)^{x_i} (1-p(\partial x_i))^{l-x_i} . \quad (375.12)$$

Die Wahrscheinlichkeit $p(\partial x_i)$ des Erfolges wird entsprechend der Parameterisierung des Auto-Modells mit $U_i(x_i) = \alpha_i$ in (375.3) dargestellt durch

$$p(\partial x_i) = \exp Q / (1 + \exp Q) \quad (375.13)$$

mit

$$Q = \alpha_i + \sum_{\substack{j=0 \\ i \neq j}}^n \beta_{ij} x_j .$$

Hiermit erhält man

$$p(\partial x_i)^{x_i} (1 - p(\partial x_i))^{1 - x_i} = (\exp Q)^{x_i} (1 + \exp Q)^{-1}$$

und nach Substitution in (375.12) und nach Vernachlässigung der Konstanten

$$p(x_i | \partial x_i) \propto \binom{1}{x_i} [\exp(\alpha_i + \sum_{\substack{j=0 \\ i \neq j}}^n \beta_{ij} x_j)]^{x_i} . \quad (375.14)$$

Für die bedingte Dichte (374.3) ergibt sich

$$\ln p(x_i | \partial x_i) \propto - \sum_{C(\mathbf{t}_i) \in K} U_C(\mathbf{x}) ,$$

so daß aus dem Vergleich mit (375.14) folgt

$$- \sum_{C(\mathbf{t}_i) \in K} U_C(\mathbf{x}) = x_i \alpha_i + x_i \sum_{\substack{j=0 \\ i \neq j}}^n \beta_{ij} x_j + \ln \binom{1}{x_i} .$$

Durch Summation über sämtliche Cliques erhält man wie in (375.1) die Energie $U(\mathbf{x})$

$$-U(\mathbf{x}) = \sum_{i=0}^n x_i \alpha_i + \sum_{i=0}^n \sum_{j=i+1}^n \beta_{ij} x_i x_j + \sum_{i=0}^n \ln \binom{1}{x_i} \quad (375.15)$$

und hiermit nach (373.10) die Gibbs-Verteilung. Ein Markoff-Zufallsfeld mit der bedingten Dichte (375.12) und der Energie (375.15) bezeichnet man als *autobinomiales Modell* (Besag 1974). Wie ein Vergleich dieser Energie mit (375.3) zeigt, handelt es sich um einen Spezialfall des Auto-Modells.

Als bedingte Verteilung eines diskreten Zufallsfeldes läßt sich auch die Poisson-Verteilung (321.10) einführen, deren Parameter wie Q in (375.13) dargestellt werden. Man erhält dann das *Auto-Poisson-Modell* (Besag 1974; Cressie 1991, S.427).

f) Vektorielltes Gauß-Zufallsfeld

Das Gauß-Zufallsfeld (375.8), das als Spezialfall des Auto-Modells (375.3) abgeleitet wurde, läßt sich auch für jede Komponente eines vektoriellen Zufallsfeldes einführen, so

daß die Verteilung eines speziellen vektoriellen Gauß-Zufallsfeldes erhalten wird.

Für das vektorielle Zufallsfeld

$$\mathbf{X}(t) = [X_1(t), \dots, X_m(t)]' \quad \text{und} \quad t \in \{t_0, \dots, t_n\}$$

mit der Realisierung

$$X_k(t_i) = x_{ik} \quad \text{für} \quad i \in \{0, 1, \dots, n\}, \quad k \in \{1, 2, \dots, m\}$$

sei die bedingte Dichte von $X_k(t_i)$ durch die Normalverteilung (375.4) gegeben. In der Notation von (372.6) folgt dann, falls ∂x_{ik} die Werte von $X_k(t_j)$ in der Nachbarschaft $\partial\{t_i\}$ von t_i bezeichnen

$$p(x_{ik} | \partial x_{ik}) \propto \exp\left\{-\frac{1}{2\sigma_k^2} \left[x_{ik} - \mu_{ik} - \sum_{\substack{l=0 \\ i \neq l}}^n \beta_{i,l,k}(x_{lk} - \mu_{lk})\right]^2\right\}. \quad (375.16)$$

Weiter gelte für die bedingte Dichte der Komponenten $X_k(t_i)$ mit $k \in \{1, 2, \dots, m\}$ des vektoriellen Zufallsfeldes

$$\begin{aligned} p(x_{i1}, \dots, x_{im} | \partial x_{ik}) \propto \exp\left\{-\frac{1}{2\sigma_1^2} \left[x_{i1} - \mu_{i1} - \sum_{\substack{l=0 \\ i \neq l}}^n \beta_{i,l,1}(x_{l1} - \mu_{l1})\right]^2 - \dots \right. \\ \left. - \frac{1}{2\sigma_m^2} \left[x_{im} - \mu_{im} - \sum_{\substack{l=0 \\ i \neq l}}^n \beta_{i,l,m}(x_{lm} - \mu_{lm})\right]^2\right\}. \quad (375.17) \end{aligned}$$

Die Summen der Exponentialfunktion sind Konstanten, die in der Normierungskonstanten der Verteilung absorbiert werden können. Man erhält daher anstelle von (375.17)

$$\begin{aligned} p(x_{i1}, \dots, x_{im} | \partial x_{ik}) \propto \\ \exp\left\{-\frac{1}{2\sigma_1^2} \left[(x_{i1} - \mu_{i1})^2 - 2(x_{i1} - \mu_{i1}) \sum_{\substack{l=0 \\ i \neq l}}^n \beta_{i,l,1}(x_{l1} - \mu_{l1})\right] - \dots \right. \\ \left. - \frac{1}{2\sigma_m^2} \left[(x_{im} - \mu_{im})^2 - 2(x_{im} - \mu_{im}) \sum_{\substack{l=0 \\ i \neq l}}^n \beta_{i,l,m}(x_{lm} - \mu_{lm})\right]\right\}. \quad (375.18) \end{aligned}$$

Hieraus ist ersichtlich, daß die Verteilung sich aus Potentialen von Einer-Cliquen mit dem Indexvektor t_i und Potentialen von Zweier-Cliquen mit t_i und t_j für $i \neq j$ zusammensetzt. Faßt man die Werte x_{ik} der Zufallsvariablen $X_k(t_i)$ in den Vektoren \mathbf{x}_k wie folgt zusammen

$$\mathbf{x}_k = (x_{ik}) \quad \text{für} \quad i \in \{0, 1, \dots, n\}, \quad k \in \{1, 2, \dots, m\}$$

und führt man mit

$$\mathbf{x} = [\mathbf{x}_1', \mathbf{x}_2', \dots, \mathbf{x}_m']'$$

den Vektor \mathbf{x} einer Realisierung des vektoriellen Zufallsfeldes $X(\mathbf{t})$ ein, dann ergibt sich die gemeinsame Verteilung $p(\mathbf{x})$ von $X(\mathbf{t})$ nach (373.13) aus der Summation über sämtliche Einer- und Zweier-Cliquen. Mit

$$-\beta_{ii,k} = 1 \quad \text{und} \quad \beta_{ij,k} = \beta_{ji,k}$$

erhält man mit denselben Umformungen, die nach (375.7) führen,

$$p(\mathbf{x}) \propto \exp\left\{-\frac{1}{2\sigma_1^2} (\mathbf{x}_1 - \boldsymbol{\mu}_1)' \mathbf{B}_1 (\mathbf{x}_1 - \boldsymbol{\mu}_1) - \dots - \frac{1}{2\sigma_m^2} (\mathbf{x}_m - \boldsymbol{\mu}_m)' \mathbf{B}_m (\mathbf{x}_m - \boldsymbol{\mu}_m)\right\} \quad (375.19)$$

mit

$$\boldsymbol{\mu}_k = (\mu_{ik}) \quad \text{und} \quad \mathbf{B}_k = (-\beta_{ij,k}) \quad \text{für} \quad k \in \{1, 2, \dots, m\} .$$

Die Gibbs-Verteilung (375.19) eines vektoriellen Zufallsfeldes ist ein Spezialfall der Gibbs-Verteilung, die aus der allgemeinen Darstellung der Energie (375.2) für ein vektorielles Zufallsfeld folgt. Die Verteilung (375.19) bedeutet, daß sich jede Komponente $X_k(\mathbf{t}_i)$ mit $k \in \{1, 2, \dots, m\}$ des vektoriellen Zufallsfeldes normalverteilt ergibt, wobei $X_k(\mathbf{t}_i)$ unabhängig von $X_l(\mathbf{t}_i)$ für $k \neq l$ ist.

Beispiele, in denen die hier entwickelten Gibbs-Verteilungen eingesetzt werden, sind im Kapitel 379 gegeben.

376 Markoff-Zufallsfelder auf Gittern

Für Anwendungen werden häufig Markoff-Zufallsfelder auf regelmäßigen Gittern definiert. In der digitalen Bildverarbeitung repräsentieren beispielsweise die Grauwerte der Bildelemente, die man sich in Gittern angeordnet denken kann, ein Markoff-Zufallsfeld. Die Zugehörigkeit von Bildelementen zu bestimmten Texturen stellt ebenfalls ein Markoff-Zufallsfeld auf einem Gitter dar. Analysiert man die Grauwerte von Bildern, die in verschiedenen Spektralbereichen des Lichts aufgenommen werden, liegt ein vektorielles Markoff-Zufallsfeld vor, dessen Anzahl der Komponenten aus der Anzahl der Spektralbereiche folgt.

Für ein Gitter lassen sich bequem Nachbarschaften definieren, die beliebig ausgedehnt sein können. Die Schnittpunkte eines Gitters seien die Knoten eines Graphen. Diese Knoten werden durch Kanten in Form eines Gitters verbunden, so daß man den in Abbildung 376-1 a) dargestellten Graphen erhält, der die Gestalt eines regelmäßigen Gitters besitzt. Greift man einen Knoten heraus, wobei Knoten am Rande des Gitters zunächst

nicht betrachtet werden, und stellt die Nachbarn dieses Knotens zusammen, ergibt sich die in Abbildung 376-1 b) dargestellte Nachbarschaft von vier Punkten, die sogenannte *Vierer-Nachbarschaft*. Für dieses Nachbarschaftssystem lassen sich nach (373.8) außer der Clique mit dem zentralen Knoten die beiden in Abbildung 376-1 c) dargestellten Cliquentypen bilden. Es sind jeweils zwei nebeneinanderliegende Gitterpunkte in horizontaler und vertikaler Richtung.

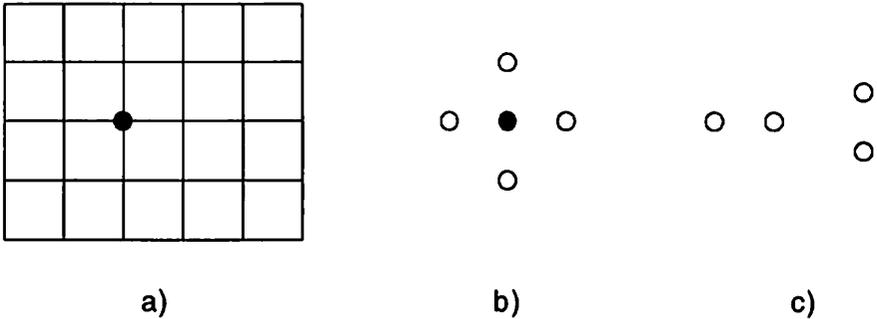


Abb. 376-1: a) Graph des Gitters, b) Vierer-Nachbarschaft, c) Cliquentypen

In Abbildung 376-2 a) ist ein weiterer Graph eines regelmäßigen Gitters dargestellt. In ihm sind die Knoten zusätzlich durch diagonal verlaufende Kanten verbunden. Dieser Graph führt auf die in Abbildung 376-2 b) dargestellte *Achter-Nachbarschaft* mit den neun in Abbildung 376-2 c) aufgeführten Cliquentypen ohne die Clique mit dem zentralen Knoten. Die *Zwölfer-Nachbarschaft* folgt, wenn zusätzlich durch Kanten in horizontaler und vertikaler Richtung jeder zweite Knoten verbunden wird und die *Zwanziger-Nachbarschaft* mit zusätzlichen Verbindungen in schrägen Richtungen, siehe auch Abbildung 376-3. Entsprechend lassen sich die Nachbarschaftssysteme immer weiter ausdehnen.

Ein Markoff-Zufallsfeld mit einer Vierer-Nachbarschaft bezeichnet man auch als *Markoff-Zufallsfeld erster Ordnung* und das Markoff-Zufallsfeld mit der Achter-Nachbarschaft als *Markoff-Zufallsfeld zweiter Ordnung* und so fort in Analogie zur Definition (322.2) eines Markoff-Prozesses der Ordnung p . Markoff-Zufallsfelder bis zur fünften Ordnung, also bis zur Vierundzwanziger-Nachbarschaft sind in Abbildung 376-3 dargestellt. Mit den Zahlen, die die Ordnungen angeben, werden die Knoten bezeichnet, die hinzuzufügen sind, um die Nachbarschaft der jeweiligen Ordnung aufzubauen.

Eine Nachbarschaft läßt sich auch dadurch einführen, daß sämtliche Gitterpunkte eines im Gitter definierten Rechtecks Nachbarn des zentralen Punktes des Rechtecks sind.

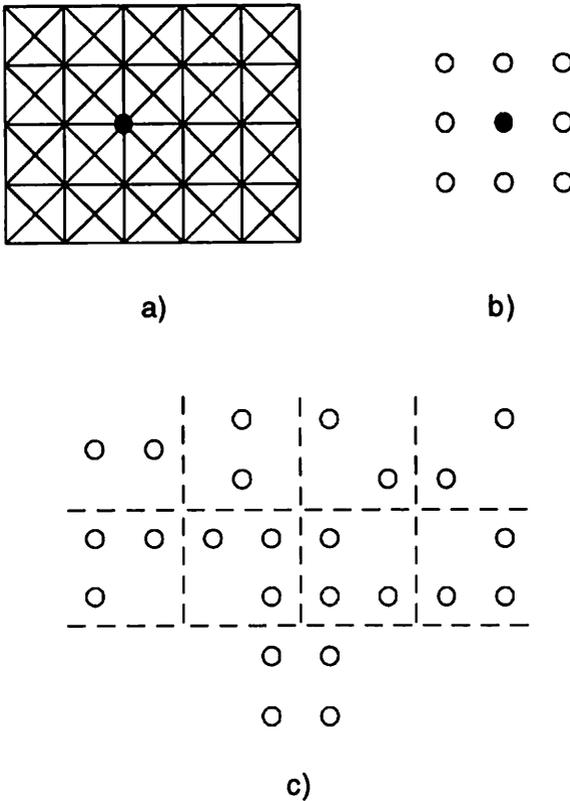


Abb. 376-2: a) Graph des Gitters, b) Achter-Nachbarschaft, c) Cliquentypen

Es sei

$$X(m,n) = X(t_1) \quad \text{mit} \quad t_1 = [m,n]'$$

$$\text{und} \quad m \in \{0, 1, \dots, M\}, \quad n \in \{0, 1, \dots, N\}, \quad l \in \{0, 1, \dots, (M+1)(N+1)-1\} \quad (376.1)$$

ein Markoff-Zufallsfeld für ein Gitter, wobei der Index m in vertikaler Richtung nach unten anwächst, während der Index n in horizontaler Richtung nach rechts zunimmt. Im Gitterpunkt (m,n) wird durch $X(m,n) = x_{mn}$ der Wert x_{mn} für $X(m,n)$ definiert, der im Vektor x der Realisierung des Zufallsfeldes mit $x = (x_{mn})$ gesammelt wird. Bei endlich ausgedehnten Gittern muß über die Werte des Zufallsfeldes jenseits des Randes des Gitters verfügt werden. Die einfachste Annahme besteht in einem sogenannten *freien Rand*. Bei ihm werden die Werte des Zufallsfeldes außerhalb des Gitters zu Null gesetzt. Eine weitere Möglichkeit, den Rand zu behandeln, besteht darin, ein *toroidales Gitter* anzunehmen, bei dem man sich die Werte in den Gitterpunkten am rechten Rand mit den Werten des linken Randes fortgesetzt denkt. Ebenso werden die Werte am unteren Rand

5	4	3	4	5
○	○	○	○	○
4	2	1	2	4
○	○	○	○	○
3	1		1	3
○	○	●	○	○
4	2	1	2	4
○	○	○	○	○
5	4	3	4	5
○	○	○	○	○

Abb. 376-3: Aufbau der Markoff-Zufallsfelder bis zur fünften Ordnung

um die Werte am oberen Rand ergänzt. Entsprechendes gilt für den linken und den oberen Rand. Ein toroidales Gitter entsteht also dadurch, daß man zunächst aus dem ebenen Gitter einen Zylinder formt und dann die offenen Enden zu einem Torus zusammenführt. In Abbildung 376-4 ist ein toroidales Gitter mit neun Gitterpunkten (x) und seine Fortsetzung durch die Randpunkte (o) dargestellt.

(0,0)		(0,2)	(0,0)		(0,2)	(0,0)
○	○	○	○	○	○	○
	○	○	○	○	○	○
(2,0)		(2,2)	(2,0)		(2,2)	(2,0)
○	○	○	○	○	○	○
(0,0)		(0,2)	(0,0)		(0,2)	(0,0)
○	○	○	x	x	x	○
	○	○	x	x	x	○
(2,0)		(2,2)	(2,0)		(2,2)	(2,0)
○	○	○	x	x	x	○
(0,0)		(0,2)	(0,0)		(0,2)	(0,0)
○	○	○	○	○	○	○

Abb. 376-4: Toroidales Gitter mit neun Gitterpunkten (x) und seine Fortsetzung durch die Randpunkte (o)

Im folgenden werden Beispiele für Energien von Gibbs-Verteilungen und für die zugehörigen gemeinsamen und bedingten Verteilungen von Markoff-Zufallsfeldern auf Gittern gegeben.

Die Energie $U(\mathbf{x})$ der Gibbs-Verteilung (373.10) eines Markoff-Zufallsfeldes mit einer Vierer-Nachbarschaft oder eines Markoff-Zufallsfeldes erster Ordnung, in dem nur Einer- und Zweier-Cliquen auftreten, erhält man in der (375.1) entsprechenden Darstellung, falls über sämtliche Gitterpunkte summiert wird, mit

$$U(\mathbf{x}) = \sum_{m=0}^M \sum_{n=0}^N x_{mn} [U_{mn}(x_{mn}) + x_{m+1,n} U_{mn}(x_{mn}, x_{m+1,n}) + x_{m,n+1} U_{mn}(x_{mn}, x_{m,n+1})] \quad (376.2)$$

Da die Summation über die Gitterpunkte erfolgt, dürfen von den Potentialen der Cliquen eines Gitterpunktes nur die aufgeführt werden, deren Cliquen nicht noch einmal bei einem folgenden Gitterpunkt auftauchen.

Die zugehörige bedingte Dichte einer Zufallsvariablen $X(m,n)$ des Markoff-Zufallsfeldes mit einer Vierer-Nachbarschaft ergibt sich nach (374.3) proportional zu

$$p(x_{mn} | \partial x_{mn}) \propto \exp \{ -x_{mn} [U_{mn}(x_{mn}) + x_{m+1,n} U_{mn}(x_{mn}, x_{m+1,n}) + x_{m,n+1} U_{mn}(x_{mn}, x_{m,n+1}) + x_{m-1,n} U_{m-1,n}(x_{mn}, x_{m-1,n}) + x_{m,n-1} U_{m,n-1}(x_{mn}, x_{m,n-1})] \} \quad (376.3)$$

Für diese bedingte Dichte ist über sämtliche Cliquen zu summieren, die den Indexvektor t_1 enthalten, während in der aus (376.2) folgenden gemeinsamen Dichte über jede Clique des Gitters zu summieren ist.

Für ein Markoff-Zufallsfeld mit einer Achter-Nachbarschaft oder für ein Markoff-Zufallsfeld zweiter Ordnung lautet die Energie $U(\mathbf{x})$, falls das Auto-Modell (375.3) mit $U_i(x_i) = \alpha_i$ angenommen wird,

$$U(\mathbf{x}) = \sum_{m=0}^M \sum_{n=0}^N x_{mn} [\alpha_{mn} + x_{m+1,n} \beta_{m+1,n} + x_{m,n+1} \beta_{m,n+1} + x_{m+1,n-1} \beta_{m+1,n-1} + x_{m+1,n+1} \beta_{m+1,n+1}] \quad (376.4)$$

Die bedingte Dichte $p(x_{mn} | \partial x_{mn})$ folgt hieraus entsprechend wie in (376.3) durch Summation über die Cliquen der acht Nachbarn.

Sind die Konstanten α_{mn} und β_{mn} in (376.4) über das gesamte Gitter identisch, läßt sich die Energie $U(\mathbf{x})$ des Auto-Modells darstellen durch

$$U(\mathbf{x}) = \sum_{m=0}^M \sum_{n=0}^N x_{mn} [\alpha + x_{m+1,n} \beta_{1,0} + x_{m,n+1} \beta_{0,1} + x_{m+1,n-1} \beta_{1,-1} + x_{m+1,n+1} \beta_{1,1}] \quad (376.5)$$

beschränkt, und es gelte

$$\beta_{mnkl} = \beta_{k-m, l-n} ,$$

so daß die Koeffizienten in der Nachbarschaft jedes Gitterpunktes identisch sind. Anstelle der Symmetriebedingung (376.9) folgt dann die Bedingung

$$\beta_{k-m, l-n} = \beta_{-(k-m), -(l-n)} \quad (376.13)$$

und anstatt (376.7) die bedingte Dichte

$$p(x_{mn} | \partial x_{mn}) \propto \exp\left\{-\frac{1}{2\sigma^2}[(x_{mn}-\mu_{mn})^2 - 2(x_{mn}-\mu_{mn}) \sum_{k=m-p}^{m+p} \sum_{l=n-q}^{n+q} \beta_{k-m, l-n}(x_{kl}-\mu_{kl})]\right\} .$$

(376.14)

Um die Dichte $p(\mathbf{x})$ der zugehörigen gemeinsamen Verteilung anzugeben, muß wieder wie in (376.8) über sämtliche Einer- und Zweier-Cliquen summiert werden, wobei aber jede Clique nur einmal auftauchen darf. Es ist daher jeweils nur über die Hälfte der Gitterpunkte in dem durch (376.12) definierten Rechteck zu summieren. Man erhält

$$p(\mathbf{x}) \propto \exp\left\{-\frac{1}{2\sigma^2} \sum_{m=0}^M \sum_{n=0}^N [(x_{mn}-\mu_{mn})^2 - 2(x_{mn}-\mu_{mn}) \left(\sum_{l=n+1}^{n+q} \beta_{0, l-n}(x_{ml}-\mu_{ml}) + \sum_{k=m+1}^{m+p} \sum_{l=n-q}^{n+q} \beta_{k-m, l-n}(x_{kl}-\mu_{kl}) \right)]\right\} .$$

(376.15)

Mit (376.13) und $-\beta_{00} = 1$ läßt sich der Exponent umformen

$$\begin{aligned} & \sum_{m=0}^M \sum_{n=0}^N [(x_{mn}-\mu_{mn})^2 - 2(x_{mn}-\mu_{mn}) \left(\sum_{l=n+1}^{n+q} \beta_{0, l-n}(x_{ml}-\mu_{ml}) + \sum_{k=m+1}^{m+p} \sum_{l=n-q}^{n+q} \beta_{k-m, l-n}(x_{kl}-\mu_{kl}) \right)] \\ &= \sum_{m=0}^M \sum_{n=0}^N \sum_{k=m-p}^{m+p} \sum_{l=n-q}^{n+q} (x_{mn}-\mu_{mn})(-\beta_{k-m, l-n})(x_{kl}-\mu_{kl}) , \end{aligned} \quad (376.16)$$

so daß mit $\mathbf{B} = (-\beta_{k-m, l-n})$ die Dichte $p(\mathbf{x})$ der Normalverteilung folgt

$$p(\mathbf{x}) \propto \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu})' \mathbf{B}(\mathbf{x}-\boldsymbol{\mu})\right\} . \quad (376.17)$$

Der Aufbau der Matrix \mathbf{B} hängt von der Größe des mit (376.12) definierten Rechtecks der Nachbarpunkte und von der Annahme über den Rand des Gitters ab. Für ein toroidales Gitter folgt \mathbf{B} als zirkuläre Blockmatrix, wie im Zusammenhang mit (378.42) näher erläutert wird. Mit (378.51) wird außerdem gezeigt, daß die Normalverteilung (376.17) identisch ist mit der Verteilung eines nichtkausalen zweidimensionalen autoregressiven

Prozesses mit Markoff-Eigenschaft.

Im Gegensatz zu der als Rechteck definierten Nachbarschaft (376.12) sind die Nachbarschaften der Markoff-Zufallsfelder unregelmäßig begrenzt, siehe Abbildung 376-3. Um auch für diese Zufallsfelder eine kompakte Darstellung der Verteilungen zu gewinnen, wird mit

$$\Omega = \{r = (m, n), 0 \leq m \leq M, 0 \leq n \leq N\} \quad (376.18)$$

die Menge der Gitterpunkte eingeführt und für jeden Gitterpunkt $r \in \Omega$ die Nachbarschaft N_p des Markoff-Zufallsfeldes der Ordnung p definiert durch

$$N_p = \{s_1, s_2, \dots, s_l\} \quad (376.19)$$

mit

$$\begin{aligned} N_1 &= \{s_1, s_2\} \\ &= \{(1, 0), (0, 1)\} \end{aligned}$$

$$\begin{aligned} N_2 &= \{s_1, s_2, s_3, s_4\} \\ &= \{(1, 0), (0, 1), (1, -1), (1, 1)\} \end{aligned}$$

$$\begin{aligned} N_3 &= \{s_1, s_2, s_3, s_4, s_5, s_6\} \\ &= \{(1, 0), (0, 1), (1, -1), (1, 1), (2, 0), (0, 2)\} \end{aligned}$$

$$\begin{aligned} N_4 &= \{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}\}, \\ & \quad s_7 = (1, -2), \quad s_8 = (1, 2), \quad s_9 = (2, -1), \quad s_{10} = (2, 1) \end{aligned}$$

und so fort. Die Anzahl l der Nachbarschaftspunkte s_i hängt von der Ordnung p ab. Für den Gitterpunkt $r \in \Omega$ ergibt sich die Nachbarschaft im Markoff-Zufallsfeld der Ordnung p aus $r+s_i$ und $r-s_i$ mit $s_i \in N_p$ und $i \in \{1, \dots, l\}$, wobei beispielsweise $r+s_3 = (m+1, n-1)$ gilt.

Mit den Nachbarschaftspunkten $s \in N_p$ folgt für die Energie $U(\mathbf{x})$ des Auto-Modells eines Markoff-Zufallsfeldes der Ordnung p anstelle von (376.5)

$$U(\mathbf{x}) = \sum_{r \in \Omega} x_r \left[\alpha + \sum_{s \in N_p} \beta_s x_{r+s} \right] \quad (376.20)$$

und für die bedingte Dichte $p(x_r | \partial x_r)$ anstelle von (376.6)

$$p(x_r | \partial x_r) \propto \exp \left\{ -x_r \left[\alpha + \sum_{s \in N_p} \beta_s (x_{r+s} + x_{r-s}) \right] \right\} . \quad (376.21)$$

Die bedingte Dichte (376.14) der bedingten Normalverteilung, die als Spezialfall des Auto-Modells abgeleitet wurde, ergibt sich für den Gitterpunkt $r \in \Omega$ mit den Nachbar-

schaftspunkten $s \in N_p$ des Markoff-Zufallsfeldes der Ordnung p zu

$$p(x_r | \partial x_r) \propto \exp\left\{-\frac{1}{2\sigma^2}[(x_r - \mu_r)^2 - 2(x_r - \mu_r) \sum_{s \in N_p} \beta_s (x_{r+s} - \mu_{r+s} + x_{r-s} - \mu_{r-s})]\right\} \quad (376.22)$$

mit

$$\beta_s = \beta_{-s} \quad (376.23)$$

aus (376.13). Die Dichte $p(\mathbf{x})$ der gemeinsamen Verteilung folgt aus (376.15) mit

$$p(\mathbf{x}) \propto \exp\left\{-\frac{1}{2\sigma^2} \sum_{r \in \Omega} [(x_r - \mu_r)^2 - 2(x_r - \mu_r) \sum_{s \in N_p} \beta_s (x_{r+s} - \mu_{r+s})]\right\} . \quad (376.24)$$

Durch die Umformung (376.16) erhält man mit $\mathbf{B} = (-\beta_s)$ die Dichte der Normalverteilung wie in (376.17) zu

$$p(\mathbf{x}) \propto \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{x} - \boldsymbol{\mu})' \mathbf{B}(\mathbf{x} - \boldsymbol{\mu})\right\} . \quad (376.25)$$

Als weiteres Beispiel der Gibbs-Verteilung eines Markoff-Zufallsfeldes auf einem Gitter sei das bereits zu Beginn des Kapitels 373 erwähnte *Ising-Modell* aufgeführt (Geman und Geman 1984; Ising 1925; Kindermann und Snell 1980, S.2). Das Modell wurde ursprünglich für Punkte auf einer Linie formuliert und wird hier für die Gitterpunkte $r \in \Omega$ mit den Nachbarschaftspunkten $s \in N_p$ angegeben. Es stellt einen Spezialfall des autologischen Modells dar. Seine Energie $U(\mathbf{x})$ ergibt sich anstelle von (375.10) in der Darstellung (376.20) aus

$$U(\mathbf{x}) = \sum_{r \in \Omega} x_r \left[\alpha + \sum_{s \in N_p} \beta_s (x_{r+s}) \right] , \quad (376.26)$$

worin α und β_s die Parameter des Modells bezeichnen. Die bedingte Verteilung des Ising-Modells folgt dann mit (375.11) und (376.3) zu

$$p(x_r | \partial x_r) = \exp(x_r T) / (1 + \exp T) \quad (376.27)$$

mit

$$T = -\left[\alpha + \sum_{s \in N_p} \beta_s (x_{r+s} + x_{r-s}) \right] .$$

Abschließend sei das vektorielle Zufallsfeld $\mathbf{X}(t)$ nach (361.3) auf einem Gitter definiert. Für die k -te Komponente $X_k(t_1)$ gelte dann (376.1) entsprechend

$$X_k(m, n) = X_k(t_1) \quad \text{mit} \quad t_1 = [m, n]' \quad \text{und} \\ m \in \{0, 1, \dots, M\}, n \in \{0, 1, \dots, N\}, l \in \{0, 1, \dots, (M+1)(N+1)-1\}, k \in \{1, 2, \dots, K\} . \quad (376.28)$$

Durch

$$X_k(m, n) = x_{mnk} \quad (376.29)$$

werde der Wert x_{mnk} der k -ten Komponente bezeichnet. Nimmt man an, daß die Komponenten voneinander unabhängig sind, braucht in (376.3) an die vorhandenen Indizes nur der Index k hinzugefügt und in (376.2) über den Index k summiert zu werden, um die Gibbs-Verteilung eines vektoriiellen Markoff-Zufallsfeldes mit einer Vierer-Nachbarschaft zu erhalten. Diese Summation wurde in (375.19) gezeigt. Entsprechend ist bei den Markoff-Zufallsfeldern höherer Ordnung zu verfahren.

Beispiele für Anwendungen von Markoff-Zufallsfeldern auf Gittern befinden sich im Kapitel 379.

377 Kausale zweidimensionale autoregressive Prozesse

So wie ein eindimensionaler stochastischer Prozeß durch einen eindimensionalen autoregressiven Prozeß, so läßt sich ein zweidimensionales Zufallsfeld durch einen zweidimensionalen autoregressiven Prozeß auf einfache Weise beschreiben. Im folgenden werden kausale autoregressive Prozesse betrachtet, die also die Bedingung (235.13) der Kausalität erfüllen. Im nächsten Kapitel folgen dann nichtkausale autoregressive Prozesse. Da ebene Zufallsfelder in manchen Anwendungen die Markoff-Eigenschaft besitzen, wird gezeigt, daß kausale autoregressive Prozesse in der Ebene unter gewissen Voraussetzungen Markoff-Zufallsfelder darstellen.

Ein *kausaler zweidimensionaler autoregressiver Prozeß* der Ordnung $\max\{p, q\}$, abgekürzt geschrieben AR(p,q)-Prozeß, ist definiert analog zu (325.1) durch

$$X(m, n) = \sum_{\substack{k=0 \\ (k, l) \neq (0, 0)}}^p \sum_{l=0}^q a_{kl} X(m-k, n-l) + \varepsilon(m, n) \quad (377.1)$$

worin a_{kl} konstante Koeffizienten bedeuten. Der Prozeß $\varepsilon(m, n)$ wird als stationäres weißes Rauschen angenommen, für den der Erwartungswert

$$E(\varepsilon(m, n)) = 0 \quad (377.2)$$

und die Autokovarianzfunktion

$$C_{\varepsilon\varepsilon}(k, l) = E(\varepsilon(m, n)\varepsilon(m-k, n-l)) = \sigma_\varepsilon^2 d(k, l) \quad (377.3)$$

gelten. Für den Erwartungswert von $X(m, n)$ wird

$$E(X(m, n)) = \mu_x(m, n) = \mu_x = 0 \quad (377.4)$$

gesetzt. Löst man (377.1) nach $\varepsilon(m, n)$ auf, ergibt sich mit $a_{00} = -1$

$$\varepsilon(m, n) = - \sum_{k=0}^p \sum_{l=0}^q a_{kl} X(m-k, n-l) , \quad (377.5)$$

was nach (235.29) einer nichtrekursiven Filterung entspricht. Die Systemfunktion lautet dann nach (235.31)

$$H(z_1, z_2) = - \sum_{k=0}^p \sum_{l=0}^q a_{kl} z_1^{-k} z_2^{-l} . \quad (377.6)$$

Wie in (325.7) wird der zu (377.5) inverse Prozeß gebildet, um $X(m, n)$ in Abhängigkeit von $\varepsilon(m, n)$ darzustellen

$$X(m, n) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \bar{a}_{kl} \varepsilon(m-k, n-l) . \quad (377.7)$$

Die zugehörige Systemfunktion gewinnt man, indem (377.6) entsprechend (325.8) invertiert wird. Der AR(p,q)-Prozeß ist somit stabil, wenn es nach Satz (241.22) kein Wertepaar (z_1, z_2) mit $z_1 = r_1 \exp(j\Omega_1)$ und $z_2 = r_2 \exp(j\Omega_2)$ sowie $r_1 \geq 1$ und gleichzeitig $r_2 \geq 1$ gibt, für das $H(z_1, z_2) = 0$ gilt. Ist das der Fall, folgt nach (235.12), da mit (235.26) wegen (377.7) $h(k, l) = \bar{a}_{kl}$ gilt,

$$\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} |\bar{a}_{kl}| < \infty \quad (377.8)$$

analog zu (325.9).

Die Momentfunktion $K_{xx}(m_1, n_1, m_2, n_2)$ des autoregressiven Prozesses $X(m, n)$ berechnet sich nach (362.2) zu

$$K_{xx}(m_1, n_1, m_2, n_2) = E(X(m_1, n_1)X(m_2, n_2)) .$$

Setzt man hierin (377.7) ein, ergibt sich mit $p_1 = m_2 - m_1$ und $p_2 = n_2 - n_1$

$$\begin{aligned} K_{xx}(m_1, n_1, m_2, n_2) &= \sum_{k_1=0}^{\infty} \sum_{l_1=0}^{\infty} \sum_{k_2=0}^{\infty} \sum_{l_2=0}^{\infty} \bar{a}_{k_1 l_1} \bar{a}_{k_2 l_2} E(\varepsilon(m_1 - k_1, n_1 - l_1) \varepsilon(m_2 - k_2, n_2 - l_2)) \\ &= \sum_{k_1=0}^{\infty} \sum_{l_1=0}^{\infty} \sum_{k_2=0}^{\infty} \sum_{l_2=0}^{\infty} K_{\varepsilon\varepsilon}(p_1 + k_1 - k_2, p_2 + l_1 - l_2) \bar{a}_{k_1 l_1} \bar{a}_{k_2 l_2} \\ &= K_{xx}(p_1, p_2) . \end{aligned} \quad (377.9)$$

Gilt (377.8), ist die Momentfunktion beschränkt. Wegen (377.4) ist daher der autoregressive Prozeß $X(m, n)$ aus (377.1) zumindest im weiteren Sinne homogen. Die Forderung (377.8) stellt daher die *Homogenitätsbedingung* dar, die im folgenden als erfüllt angesehen

hen wird.

Ersetzt man in (377.7) m durch $m-k_1$ und n durch $n-k_2$ mit $k_1 \geq 0$ und $k_2 \geq 0$ sowie $k_1+k_2 > 0$, multipliziert die Gleichung mit $\varepsilon(m,n)$ und bildet anschließend den Erwartungswert, folgt

$$\begin{aligned} & E(\varepsilon(m,n)X(m-k_1, n-k_2)) \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \bar{a}_{kl} E(\varepsilon(m,n)\varepsilon(m-(k_1+k), n-(k_2+l))) . \end{aligned}$$

Hieraus folgt wegen (377.3)

$$C_{\varepsilon x}(k_1, k_2) = E(\varepsilon(m,n)X(m-k_1, n-k_2)) = 0 . \quad (377.10)$$

Multipliziert man (377.1) mit $\varepsilon(m,n)$ und bildet erneut den Erwartungswert, gilt

$$\begin{aligned} & E(\varepsilon(m,n)X(m,n)) \\ &= \sum_{k=0}^p \sum_{\substack{l=0 \\ (k,l) \neq (0,0)}}^q a_{kl} E(\varepsilon(m,n)X(m-k, n-l)) + E(\varepsilon(m,n)\varepsilon(m,n)) \end{aligned}$$

und mit (377.3) und (377.10)

$$C_{\varepsilon x}(0,0) = \sigma_{\varepsilon}^2 . \quad (377.11)$$

Faßt man (377.10) und (377.11) zusammen, erhält man

$$C_{\varepsilon x}(k,l) = E(\varepsilon(m,n)X(m-k, n-l)) = \sigma_{\varepsilon}^2 d(k,l) \quad (377.12)$$

für alle $k, l \geq 0$.

Wegen (377.2) und (377.4) gilt $K_{xx}(p_1, p_2) = C_{xx}(p_1, p_2)$ und $K_{\varepsilon\varepsilon}(p_1, p_2) = C_{\varepsilon\varepsilon}(p_1, p_2)$, so daß (377.9) mit der Autokovarianzfunktion identisch ist. Um die Koeffizienten \bar{a}_{kl} zu eliminieren, wird die Autokovarianzfunktion unmittelbar aus (377.1) berechnet. Multipliziert man diese Gleichung mit $X(m-p_1, n-p_2)$ für $p_1, p_2 \geq 0$ und bildet anschließend den Erwartungswert, ergibt sich mit (377.12) und $C_{xx}(p_1, p_2) = C_{xx}(-p_1, -p_2)$ aus (362.6)

$$\begin{aligned} C_{xx}(p_1, p_2) &= E(X(m,n)X(m-p_1, n-p_2)) \\ &= \sum_{k=0}^p \sum_{\substack{l=0 \\ (k,l) \neq (0,0)}}^q a_{kl} E(X(m-k, n-l)X(m-p_1, n-p_2)) + E(\varepsilon(m,n)X(m-p_1, n-p_2)) \\ &= \sum_{k=0}^p \sum_{\substack{l=0 \\ (k,l) \neq (0,0)}}^q a_{kl} C_{xx}(p_1-k, p_2-l) + \sigma_{\varepsilon}^2 d(p_1, p_2) . \end{aligned} \quad (377.13)$$

Für $p_1 = p_2 = 0$ folgt die Varianz des Prozesses $X(m,n)$

$$\sigma_x^2 = C_{xx}(0,0) = \sum_{\substack{k=0 \\ (k,1) \neq (0,0)}}^p \sum_{l=0}^q a_{kl} C_{xx}(k,l) + \sigma_\varepsilon^2 \quad (377.14)$$

und für $p_1 \geq 0$ und $p_2 \geq 0$ mit $p_1 + p_2 > 0$

$$C_{xx}(p_1, p_2) = \sum_{\substack{k=0 \\ (k,1) \neq (0,0)}}^p \sum_{l=0}^q a_{kl} C_{xx}(p_1 - k, p_2 - l) \quad (377.15)$$

Dividiert man (377.15) durch die Varianz (377.14), erhält man die Autokorrelationsfunktion

$$\rho_x(p_1, p_2) = \sum_{\substack{k=0 \\ (k,1) \neq (0,0)}}^p \sum_{l=0}^q a_{kl} \rho_x(p_1 - k, p_2 - l) \quad (377.16)$$

die analog zu (325.16) als *zweidimensionale Yule-Walker Gleichung* bezeichnet wird.

Die Autokovarianzfunktion (377.15) kann entsprechend (325.20) in Matrizenform geschrieben werden, wozu die Koeffizienten a_{kl} in den Vektoren

$$\mathbf{a}_0 = [a_{10}, a_{20}, \dots, a_{p0}]'$$

$$\mathbf{a}_j = [a_{0j}, a_{1j}, \dots, a_{pj}]' \quad \text{für } j \in \{1, \dots, q\}$$

zusammengefaßt werden. Zusätzlich führt man für die Autokovarianzen $C_{xx}(m, n) = \gamma_{m, n}$ die Vektoren

$$\boldsymbol{\gamma}_0 = [\gamma_{1,0}, \gamma_{2,0}, \dots, \gamma_{p,0}]'$$

$$\boldsymbol{\gamma}_j = [\gamma_{0,j}, \gamma_{1,j}, \dots, \gamma_{p,j}]' \quad \text{für } j \in \{1, \dots, q\} \quad (377.17)$$

und die Matrizen

$$\boldsymbol{\Sigma}_{00} = \begin{bmatrix} \gamma_{0,0} & \gamma_{-1,0} & \dots & \gamma_{1-p,0} \\ \gamma_{1,0} & \gamma_{0,0} & \dots & \gamma_{2-p,0} \\ \dots & \dots & \dots & \dots \\ \gamma_{p-1,0} & \gamma_{p-2,0} & \dots & \gamma_{0,0} \end{bmatrix}$$

$$\boldsymbol{\Sigma}_{0j} = \begin{bmatrix} \gamma_{1,-j} & \gamma_{0,-j} & \dots & \gamma_{1-p,-j} \\ \gamma_{2,-j} & \gamma_{1,-j} & \dots & \gamma_{2-p,-j} \\ \dots & \dots & \dots & \dots \\ \gamma_{p,-j} & \gamma_{p-1,-j} & \dots & \gamma_{0,-j} \end{bmatrix} = \boldsymbol{\Sigma}_{j0}'$$

$$\Sigma_{ij} = \begin{bmatrix} \gamma_{0, i-j} & \gamma_{-1, i-j} & \cdots & \gamma_{-p, i-j} \\ \gamma_{1, i-j} & \gamma_{0, i-j} & \cdots & \gamma_{1-p, i-j} \\ \dots & \dots & \dots & \dots \\ \gamma_{p, i-j} & \gamma_{p-1, i-j} & \cdots & \gamma_{0, i-j} \end{bmatrix} = \Sigma_{ji} \quad (377.18)$$

ein und erhält für die Autokovarianzen $C_{xx}(p_1, p_2) = \gamma_{p_1, p_2}$ für $p_1 \in \{0, \dots, p\}$ und $p_2 \in \{0, \dots, q\}$ mit der Einschränkung $p_1 + p_2 > 0$ das lineare Gleichungssystem

$$\begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \dots \\ \gamma_q \end{bmatrix} = \begin{bmatrix} \Sigma_{00} & \Sigma_{01} & \cdots & \Sigma_{0q} \\ \Sigma_{10} & \Sigma_{11} & \cdots & \Sigma_{1q} \\ \dots & \dots & \dots & \dots \\ \Sigma_{q0} & \Sigma_{q1} & \cdots & \Sigma_{qq} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_q \end{bmatrix} \quad (377.19)$$

oder zusammengefaßt

$$c_{pq} = \Sigma_{pq} \beta \quad (377.20)$$

worin c_{pq} und β mit $r = (p+1)q+p$ zwei $r \times 1$ Vektoren und Σ_{pq} eine symmetrische $r \times r$ Matrix bedeuten. Mit $r_{pq} = c_{pq} / \sigma_x^2$ und $R_{pq} = \Sigma_{pq} / \sigma_x^2$ erhält man die zweidimensionalen Yule-Walker Gleichungen.

Das Spektrum $S_{xx}(\Omega_1, \Omega_2)$ des autoregressiven Prozesses $X(m, n)$ ergibt sich aus (363.8) für $n=2$ zu

$$S_{xx}(\Omega_1, \Omega_2) = \sum_{p_1=-\infty}^{\infty} \sum_{p_2=-\infty}^{\infty} C_{xx}(p_1, p_2) e^{-j(p_1\Omega_1 + p_2\Omega_2)} \quad .$$

Ersetzt man hierin $K_{xx}(p_1, p_2) = C_{xx}(p_1, p_2)$ aus (377.9), folgt

$$\begin{aligned} S_{xx}(\Omega_1, \Omega_2) &= \sum_{p_1=-\infty}^{\infty} \sum_{p_2=-\infty}^{\infty} \sum_{k_1=0}^{\infty} \sum_{l_1=0}^{\infty} \sum_{k_2=0}^{\infty} \sum_{l_2=0}^{\infty} C_{\varepsilon\varepsilon}(p_1+k_1-k_2, p_2+l_1-l_2) \\ &\quad \bar{a}_{k_1 l_1} \bar{a}_{k_2 l_2} e^{-j(p_1\Omega_1 + p_2\Omega_2)} \\ &= \left(\sum_{k_1=0}^{\infty} \sum_{l_1=0}^{\infty} \bar{a}_{k_1 l_1} e^{j(k_1\Omega_1 + l_1\Omega_2)} \right) \left(\sum_{k_2=0}^{\infty} \sum_{l_2=0}^{\infty} \bar{a}_{k_2 l_2} e^{-j(k_2\Omega_1 + l_2\Omega_2)} \right) \\ &\quad \sum_{p_1=-\infty}^{\infty} \sum_{p_2=-\infty}^{\infty} C_{\varepsilon\varepsilon}(m_1, m_2) e^{-(m_1\Omega_1 + m_2\Omega_2)} \quad , \end{aligned} \quad (377.21)$$

wobei $m_1 = p_1 + k_1 - k_2$ und $m_2 = p_2 + l_1 - l_2$ substituiert wurde. Der zweite Faktor in (377.21) ist analog zu (325.8) die inverse Systemfunktion von (377.6) für $z_i = \exp(j\Omega_i)$ mit $i \in \{1, 2\}$. Wegen (213.14) und (363.8) ergibt sich somit

$$S_{xx}(\Omega_1, \Omega_2) = S_{\varepsilon\varepsilon}(\Omega_1, \Omega_2) | (H(j\Omega_1, j\Omega_2))^{-1} |^2 \quad (377.22)$$

und mit (377.3) sowie (377.6)

$$S_{xx}(\Omega_1, \Omega_2) = \sigma_\varepsilon^2 / \left| \left(\sum_{k=0}^p a_k e^{j(k\Omega_1 + l\Omega_2)} \right) \right|^2 . \quad (377.23)$$

Bei manchen Anwendungen kann vorausgesetzt werden, daß die Zufallsfelder, mit denen gearbeitet wird, die Markoff-Eigenschaft besitzen. Beispiele hierzu befinden sich im Kapitel 379. Im folgenden soll daher bewiesen werden, daß ein stationärer kausaler AR(p,q)-Prozeß, dessen Zufallsvariablen $X(m,n)$ und dessen Rauschanteile $\varepsilon(m,n)$ als normalverteilt angenommen werden, ein kausales Markoff-Zufallsfeld darstellt. Weitergehende Untersuchungen zu Markoff-Eigenschaften autoregressiver Prozesse findet man bei Kashyap (1984).

Markoff-Zufallsfelder werden im Kapitel 372 aufgrund der symmetrischen Markoff-Eigenschaft (322.12) definiert. Wegen der Kausalität wird diese Definition jetzt abgewandelt, indem aufgrund der einseitigen Markoff-Eigenschaft (322.4) ein kausales zweidimensionales Markoff-Zufallsfeld der Ordnung (p,q) durch die Bedingung eingeführt wird

$$\begin{aligned} & p(X(m,n) | X(m-1,n), X(m-2,n), \dots, X(0,n), X(m,n-1), X(m-1,n-1), \\ & \quad \dots, X(0,n-1), \dots, X(0,0)) \\ & = p(X(m,n) | X(m-1,n), X(m-2,n), \dots, X(m-p,n), X(m,n-1), \\ & \quad X(m-1,n-1), \dots, X(m-p,n-1), \dots, X(m-p,n-q)) , \end{aligned} \quad (377.24)$$

worin wie auch im folgenden die Werte der Zufallsvariablen $X(m,n)$ ebenfalls mit $X(m,n)$ bezeichnet werden. Der Beweis wird analog zu (325.25) bis (325.28) durchgeführt. Sind die Zufallsvariablen $\varepsilon(m,n)$ und $X(m-k,n-1)$ normalverteilt, folgt aus (377.12) für $k \geq 0$ und $l \geq 0$ sowie $k+l > 0$ die Unabhängigkeit. Somit erhält man aus der Definition der Unabhängigkeit für die bedingte Dichte

$$\begin{aligned} & p(\varepsilon(m,n) | X(m-1,n), X(m-2,n), \dots, X(0,n), X(m,n-1), X(m-1,n-1), \\ & \quad \dots, X(0,n-1), \dots, X(0,0)) = p(\varepsilon(m,n)) \end{aligned} \quad (377.25)$$

und daher wegen (377.2)

$$\begin{aligned} & E(\varepsilon(m,n) | X(m-1,n), X(m-2,n), \dots, X(0,n), X(m,n-1), X(m-1,n-1), \\ & \quad \dots, X(0,n-1), \dots, X(0,0)) = E(\varepsilon(m,n)) = 0 . \end{aligned} \quad (377.26)$$

Bildet man den Erwartungswert von $X(m,n)$ unter der Bedingung, daß die Werte für $X(m-1,n), \dots, X(0,0)$ gegeben seien, folgt aus (377.1) unter Berücksichtigung von (377.26)

$$E(X(m,n) | X(m-1,n), X(m-2,n), \dots, X(0,n), X(m,n-1), X(m-1,n-1),$$

$$\begin{aligned}
& \dots, X(0, n-1), \dots, X(0, 0)) \\
& = E(X(m, n) | X(m-1, n), X(m-2, n), \dots, X(m-p, n), X(m, n-1), \\
& \quad X(m-1, n-1), \dots, X(m-p, n-1), \dots, X(m-p, n-q)) \\
& = \sum_{\substack{k=0 \\ (k, l) \neq (0, 0)}}^p \sum_{l=0}^q a_{kl} X(m-k, n-l) .
\end{aligned} \tag{377.27}$$

Außerdem erhält man für die bedingte Varianz von $X(m, n)$ aus (377.1) mit (377.3)

$$\begin{aligned}
& V(X(m, n) | X(m-1, n), X(m-2, n), \dots, X(0, n), X(m, n-1), X(m-1, n-1), \\
& \quad \dots, X(0, n-1), \dots, X(0, 0)) \\
& = V(X(m, n) | X(m-1, n), X(m-2, n), \dots, X(m-p, n), X(m, n-1), \\
& \quad X(m-1, n-1), \dots, X(m-p, n-1), \dots, X(m-p, n-q)) \\
& = V(\varepsilon(m, n)) = \sigma_\varepsilon^2 .
\end{aligned} \tag{377.28}$$

Da $X(m, n)$ normalverteilt ist, gilt wegen (377.27) und (377.28) auch die Markoff-Eigenschaft (377.24). Ein kausaler AR(p,q)-Prozeß, der als Gauß-Prozeß vorausgesetzt wird, stellt somit ein kausales Markoff-Zufallsfeld der Ordnung (p,q) dar, also ein kausales Gauß-Markoff-Zufallsfeld. Der Nachweis der Markoff-Eigenschaft (377.24) über die bedingte Verteilung analog zu (325.30) bis (325.35) ist im zweidimensionalen Fall sehr aufwendig und wird daher hier nicht geführt. Eine Herleitung ist bei (Pan 1994) zu finden. Dort wird allerdings von der Definition (235.14) der Kausalität ausgegangen.

Da $X(m, n)$ normalverteilt ist, folgt aus (377.27) und (377.28) die bedingte Normalverteilung für $X(m, n) | X(m-1, n), X(m-2, n), \dots, X(m-p, n-q)$ mit

$$\begin{aligned}
& X(m, n) | X(m-1, n), X(m-2, n), \dots, X(m-p, n-q) \\
& \sim N\left(\sum_{\substack{k=0 \\ (k, l) \neq (0, 0)}}^p \sum_{l=0}^q a_{kl} X(m-k, n-l), \sigma_\varepsilon^2\right) .
\end{aligned} \tag{377.29}$$

Diese Verteilung soll noch einmal über die Normalverteilungen für $X(m, n)$ und $\varepsilon(m, n)$ abgeleitet werden. Dazu werden die Vektoren

$$\begin{aligned}
& \mathbf{x}_{p0} = [X(m-1, n), X(m-2, n), \dots, X(m-p, n)]' , \\
& \mathbf{x}_{pj} = [X(m, n-j), X(m-1, n-j), \dots, X(m-p, n-j)]' \quad \text{für } j \in \{1, \dots, q\}
\end{aligned} \tag{377.30}$$

mit den Matrizen der Autokovarianzen $D(\mathbf{x}_{p0}) = \Sigma_{00}$, $C(\mathbf{x}_{p0}, \mathbf{x}_{pj}) = \Sigma_{0j}$ und $C(\mathbf{x}_{pi}, \mathbf{x}_{pj}) = \Sigma_{ij}$ aus (377.18) eingeführt. Die Matrix der Autokovarianzen des $r \times 1$ Vektors

$$\mathbf{x}_p = [\mathbf{x}_{p0}', \mathbf{x}_{p1}', \dots, \mathbf{x}_{pq}']' \tag{377.31}$$

ist somit durch Σ_{pq} aus (377.20) gegeben. Mit dem Vektor \mathbf{x}_p und dem Vektor $\boldsymbol{\beta}$ aus (377.20) kann die Definitionsgleichung (377.1) in der Form

$$X(m, n) = [\boldsymbol{\beta}', 1] \begin{bmatrix} \mathbf{x}_p \\ \varepsilon(m, n) \end{bmatrix} \quad (377.32)$$

geschrieben werden. Die Varianz von $\varepsilon(m, n)$ folgt aus (377.3). Nimmt man für \mathbf{x}_p und $\varepsilon(m, n)$, wie bereits erwähnt, Normalverteilungen an

$$\mathbf{x}_p \sim N(\mathbf{0}, \Sigma_{pq}) \quad \text{und} \quad \varepsilon(m, n) \sim N(0, \sigma_\varepsilon^2), \quad (377.33)$$

gilt für $X(m, n)$ als lineare Transformation von \mathbf{x}_p und $\varepsilon(m, n)$ mit (377.10) die Verteilung (Koch 1987, S.140)

$$X(m, n) \sim N(0, [\boldsymbol{\beta}', 1] \begin{bmatrix} \Sigma_{pq} & \mathbf{0} \\ \mathbf{0}' & \sigma_\varepsilon^2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ 1 \end{bmatrix}). \quad (377.34)$$

In Übereinstimmung mit (377.14) ergibt sich wegen (377.20) also die Varianz

$$\sigma_x^2 = \boldsymbol{\beta}' \Sigma_{pq} \boldsymbol{\beta} + \sigma_\varepsilon^2 = \boldsymbol{\beta}' \mathbf{c}_{pq} + \sigma_\varepsilon^2. \quad (377.35)$$

Die Matrix Σ der Autokovarianzen des Vektors $\mathbf{x} = [X(m, n), \mathbf{x}_p']'$ erhält man mit $\Sigma_{mn} = [\sigma_x^2]$ unter Berücksichtigung von (377.17) und (377.20) zu

$$\Sigma = \begin{bmatrix} \Sigma_{mn} & \mathbf{c}_{pq}' \\ \mathbf{c}_{pq} & \Sigma_{pq} \end{bmatrix}. \quad (377.36)$$

Die bedingte Dichte von $X(m, n) | \mathbf{x}_p$ lautet dann (Koch 1987, S.138)

$$X(m, n) | \mathbf{x}_p \sim N(\mathbf{c}_{pq}' \Sigma_{pq}^{-1} \mathbf{x}_p, \Sigma_{mn} - \mathbf{c}_{pq}' \Sigma_{pq}^{-1} \mathbf{c}_{pq}) \quad (377.37)$$

und mit (377.20) und (377.35)

$$X(m, n) | \mathbf{x}_p \sim N(\boldsymbol{\beta}' \mathbf{x}_p, \sigma_\varepsilon^2)$$

in Übereinstimmung mit (377.29).

Analog zu (325.50) bis (325.52) wird nun die gemeinsame bedingte Dichte

$$\begin{aligned} & p(X(m+K, n+L), X(m+K-1, n+L), \dots, X(m, n+L), X(m+K, n+L-1), \dots, X(m, n) | \\ & X(m-1, n+L), \dots, X(m-p, n+L), X(m-1, n+L-1), \dots, X(m-p, n+L-1), \\ & \dots, X(m-1, n), \dots, X(m-p, n), X(m+K, n-1), X(m+K-1, n-1), \\ & \dots, X(m-p, n-1), \dots, X(m-p, n-q)) \end{aligned} \quad (377.38)$$

abgeleitet, wobei $X(m-1, n+L), \dots, X(m-p, n-q)$ die Start- oder Anfangswerte bedeuten. In (377.33) wurde der Rauschanteil $\varepsilon(m, n)$ als normalverteilt vorausgesetzt. Daher erhält man für die gemeinsame Dichte von $\varepsilon(m+K, n+L), \varepsilon(m+K-1, n+L), \dots, \varepsilon(m, n)$ wegen (377.2) und (377.3) (Koch 1987, S.106)

$$p(\varepsilon(m+K, n+L), \varepsilon(m+K-1, n+L), \dots, \varepsilon(m, n+L), \varepsilon(m+K, n+L-1), \dots, \varepsilon(m, n)) \\ = \left[\frac{1}{2\pi\sigma_\varepsilon^2} \right]^{(K+1)(L+1)/2} \exp \left[- \frac{1}{2\sigma_\varepsilon^2} \sum_{k=0}^K \sum_{l=0}^L \varepsilon(m+k, n+1)^2 \right]. \quad (377.39)$$

Ersetzt man in (377.5) m durch $m+k$ und n durch $n+1$, erhält man

$$\varepsilon(m+k, n+1) = - \sum_{i=0}^p \sum_{j=0}^q a_{ij} X(m+k-i, n+1-j) \\ = X(m+k, n+1) - \sum_{\substack{i=0 \\ (i,j) \neq (0,0)}}^p \sum_{j=0}^q a_{ij} X(m+k-i, n+1-j). \quad (377.40)$$

Es werden nun die Vektoren

$$\mathbf{x}_{k10} = [X(m+k-1, n+1), X(m+k-2, n+1), \dots, X(m+k-p, n+1)]' \\ \mathbf{x}_{k1j} = [X(m+k, n+1-j), X(m+k-1, n+1-j), \dots, X(m+k-p, n+1-j)]' \quad (377.41)$$

für $j \in \{1, \dots, q\}$ eingeführt. Bildet man weiter mit $r = (p+1)q+p$ den $r \times 1$ Vektor

$$\mathbf{x}_{k1} = [\mathbf{x}_{k10}', \mathbf{x}_{k11}', \dots, \mathbf{x}_{k1q}']', \quad (377.42)$$

so erhält man für (377.40) mit dem Vektor β aus (377.20)

$$\varepsilon(m+k, n+1) = X(m+k, n+1) - \beta' \mathbf{x}_{k1}. \quad (377.43)$$

Transformiert man nun die Zufallsvariablen $\varepsilon(m+K, n+L), \varepsilon(m+K-1, n+L), \dots, \varepsilon(m, n)$ in der Dichte (377.39) in die Zufallsvariablen $X(m+K, n+L), X(m+K-1, n+L), \dots, X(m, n)$, so folgt die Jacobische Matrix mit den partiellen Ableitungen $(\partial \varepsilon(i, j) / \partial X(k, l))$ aus (377.43). Sie ergibt sich aufgrund der Kausalität als obere Einheitsdreiecksmatrix. Mit Beispiel 1 wird dies im folgenden demonstriert. Die Dichte (377.38), die abgekürzt $p(X(m+K, n+L), \dots, X(m, n) | X(m-1, n+L), \dots, X(m-p, n-q))$ geschrieben wird, lautet daher mit $w = (K+1)(L+1)$

$$p(X(m+K, n+L), \dots, X(m, n) | X(m-1, n+L), \dots, X(m-p, n-q)) \\ = \left[\frac{1}{2\pi\sigma_\varepsilon^2} \right]^{w/2} \exp \left[- \frac{1}{2\sigma_\varepsilon^2} \sum_{k=0}^K \sum_{l=0}^L (X(m+k, n+1) - \beta' \mathbf{x}_{k1})^2 \right]. \quad (377.44)$$

Hieraus folgt schließlich mit der Normalverteilung der Startwerte $X(m-1, n+L), \dots, X(m-p, n-q)$ entsprechend (325.47) die gemeinsame Dichte für $X(m+K, n+L), \dots, X(m-q, n-q)$.

Im Kapitel 354 wird erläutert, daß im eindimensionalen Fall die bedingte Dichte (325.46) anstelle der strengen Likelihoodfunktion als genäherte Likelihoodfunktion zur Parameterschätzung herangezogen wird. Dies bedeutet für den zweidimensionalen Fall,

daß die bedingte Dichte (377.44) als genäherte Likelihoodfunktion $L_0(\beta)$ gewählt wird. Sie läßt sich kompakter schreiben, wenn man den $w \times 1$ Vektor

$$\bar{y} = [X(m+K, n+L), X(m+K-1, n+L), \dots, X(m, n)]' \quad (377.45)$$

und mit den Vektoren x_{k1} aus (377.42) die $w \times r$ Matrix

$$X = [x_{KL}, x_{K-1,L}, \dots, x_{0L}, \dots, x_{00}]' \quad (377.46)$$

einführt. Man erhält dann

$$L_0(\beta) = \left[\frac{1}{2\pi\sigma_e^2} \right]^{w/2} \exp \left[- \frac{1}{2\sigma_e^2} (\bar{y} - X\beta)' (\bar{y} - X\beta) \right] \quad (377.47)$$

analog zur eindimensionalen Darstellung (354.8). Die Schätzung des Vektors β der unbekannt Parameter ergibt sich entsprechend (354.12).

Beispiel 1: Für $K=L=1$ und $p=q=2$ erhält man mit (377.43) die vier Gleichungen

$$\begin{aligned} \varepsilon(m+1, n+1) &= X(m+1, n+1) - a_{10}X(m, n+1) - a_{20}X(m-1, n+1) - a_{01}X(m+1, n) \\ &\quad - a_{11}X(m, n) - a_{21}X(m-1, n) - a_{02}X(m+1, n-1) - a_{12}X(m, n-1) - a_{22}X(m-1, n-1) \\ \varepsilon(m, n+1) &= X(m, n+1) - a_{10}X(m-1, n+1) - a_{20}X(m-2, n+1) - a_{01}X(m, n) \\ &\quad - a_{11}X(m-1, n) - a_{21}X(m-2, n) - a_{02}X(m, n-1) - a_{12}X(m-1, n-1) - a_{22}X(m-2, n-1) \\ \varepsilon(m+1, n) &= X(m+1, n) - a_{10}X(m, n) - a_{20}X(m-1, n) - a_{01}X(m+1, n-1) \\ &\quad - a_{11}X(m, n-1) - a_{21}X(m-1, n-1) - a_{02}X(m+1, n-2) - a_{12}X(m, n-2) - a_{22}X(m-1, n-2) \\ \varepsilon(m, n) &= X(m, n) - a_{10}X(m-1, n) - a_{20}X(m-2, n) - a_{01}X(m, n-1) - a_{11}X(m-1, n-1) \\ &\quad - a_{21}X(m-2, n-1) - a_{02}X(m, n-2) - a_{12}X(m-1, n-2) - a_{22}X(m-2, n-2) . \end{aligned} \quad (377.48)$$

Die Abbildung 377-1 zeigt die zu berechnenden Werte (\times) an den Positionen (m, n) bis $(m+1, n+1)$ aus den Startwerten (\circ). Auf die Vorgabe dieser Anfangswerte wurde schon

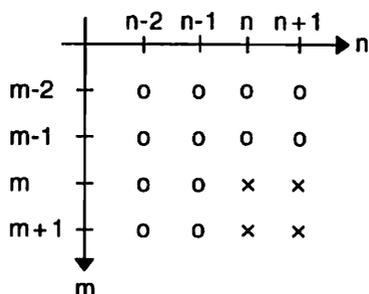


Abb. 377-1: Anordnung der Startwerte

im Zusammenhang mit Abbildung 241-1 hingewiesen.

Faßt man mit $w = (K+1)(L+1) = 4$ die Rauschanteile $\varepsilon(m+1, n+1), \dots, \varepsilon(m, n)$ im 4×1 Vektor \mathbf{e} zusammen

$$\mathbf{e} = [\varepsilon(m+1, n+1), \varepsilon(m, n+1), \varepsilon(m+1, n), \varepsilon(m, n)]'$$

und ebenso die Zufallsvariablen $X(m+1, n+1), \dots, X(m, n)$ im 4×1 Vektor $\bar{\mathbf{y}}$

$$\bar{\mathbf{y}} = [X(m+1, n+1), X(m, n+1), X(m+1, n), X(m, n)]'$$

erhält man aus (377.48) die Jacobische Matrix $\mathbf{B} = (b_{ij}) = (\partial e_i / \partial \bar{y}_j)$ als obere Einheitsdreiecksmatrix

$$\mathbf{B} = \begin{bmatrix} 1 & -a_{10} & -a_{01} & -a_{11} \\ 0 & 1 & 0 & -a_{01} \\ 0 & 0 & 1 & -a_{10} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

wobei $\det \mathbf{B} = 1$ gilt. Mit $r = (p+1)q+p = 8$ lautet der 8×1 Vektor $\boldsymbol{\beta}$ aus (377.20)

$$\boldsymbol{\beta} = [a_{10}, a_{20}, a_{01}, a_{11}, a_{21}, a_{02}, a_{12}, a_{22}]'$$

und die 4×8 Matrix \mathbf{X} aus (377.46)

$$\mathbf{X} = [\mathbf{x}_{11}, \mathbf{x}_{01}, \mathbf{x}_{10}, \mathbf{x}_{00}]'$$

$$= \begin{bmatrix} X(m, n+1) & X(m-1, n+1) & X(m, n) & X(m-1, n) \\ X(m-1, n+1) & X(m-2, n+1) & X(m-1, n) & X(m-2, n) \\ X(m+1, n) & X(m, n) & X(m+1, n-1) & X(m, n-1) \\ X(m, n) & X(m-1, n) & X(m, n-1) & X(m-1, n-1) \\ X(m-1, n) & X(m-2, n) & X(m-1, n-1) & X(m-2, n-1) \\ X(m+1, n-1) & X(m, n-1) & X(m+1, n-2) & X(m, n-2) \\ X(m, n-1) & X(m-1, n-1) & X(m, n-2) & X(m-1, n-2) \\ X(m-1, n-1) & X(m-2, n-1) & X(m-1, n-2) & X(m-2, n-2) \end{bmatrix}.$$

Die Likelihoodfunktion $L_0(\boldsymbol{\beta})$ ergibt sich hiermit aus (377.47).

△

378 Nichtkausale zweidimensionale autoregressive Prozesse

Im zweidimensionalen Fall sollen auch *nichtkausale autoregressive Prozesse* der Ordnung $\max\{p, q\}$, abgekürzt $nAR(p, q)$ -Prozesse geschrieben, betrachtet werden. Bei einem zweidimensionalen autoregressiven Prozeß wird nämlich die Zufallsvariable $X(m, n)$ unter Einbeziehung eines Rauschanteils $\varepsilon(m, n)$ durch die Zufallsvariablen in der

Nachbarschaft von $X(m, n)$ ausgedrückt. Bei den kausalen zweidimensionalen autoregressiven Prozessen ist die Nachbarschaft auf die Zufallsvariablen $X(i, j)$ mit $i \leq m, j \leq n$ und $(i, j) \neq (m, n)$ beschränkt, während die Nachbarschaft bei den nichtkausalen zweidimensionalen autoregressiven Prozessen durch eine in bezug auf $X(m, n)$ symmetrisch liegende Punktanordnung gegeben ist. Zur Vereinfachung wird hier als Nachbarschaft das bereits mit (376.12) eingeführte Rechteck gewählt. Punktanordnungen wie in den Nachbarschaften (376.19), die der Definition eines Markoff-Zufallsfeldes der Ordnung p dienen, sind ebenfalls möglich.

Wie bei den kausalen zweidimensionalen autoregressiven Prozessen wird auch hier bei den nichtkausalen Prozessen die Markoff-Eigenschaft geprüft. Wie im folgenden gezeigt wird, besitzt ein nichtkausaler autoregressiver Prozeß die Markoff-Eigenschaft, falls eine spezielle Autokovarianzfunktion für den Rauschanteil gewählt wird. Ein solcher Prozeß wird auch als *konditionales Markoff-Modell* (CM-Modell) bezeichnet. Er besitzt, wie außerdem gezeigt wird, eine Verteilung, die mit der Normalverteilung als Spezialfall des Auto-Modells eines Nachbarschafts-Gibbs-Feldes identisch ist. Ein nichtkausaler autoregressiver Prozeß, dessen Autokovarianzfunktion durch weißes Rauschen bestimmt ist, besitzt die Markoff-Eigenschaft nicht. Dieser Prozeß wird hier nicht behandelt, untersucht wird er beispielsweise bei (Kashyap und Chellappa 1983).

Ein nAR(p,q)-Prozeß wird also definiert durch

$$X(m, n) = \sum_{\substack{k=-p \\ (k, l) \neq (0, 0)}}^p \sum_{l=-q}^q a_{kl} X(m-k, n-l) + \varepsilon(m, n) \quad (378.1)$$

worin a_{kl} wieder konstante Koeffizienten bedeuten. Der Rauschanteil $\varepsilon(m, n)$ wird mit

$$E(\varepsilon(m, n)) = 0 \quad (378.2)$$

und der Autokovarianzfunktion

$$C_{\varepsilon\varepsilon}(k, l) = E(\varepsilon(m, n)\varepsilon(m-k, n-l)) = \begin{cases} -a_{kl}\sigma_\varepsilon^2 & \text{für } -p \leq k \leq p, -q \leq l \leq q \\ 0 & \text{sonst} \end{cases} \quad (378.3)$$

als stationärer Prozeß vorausgesetzt, wobei $a_{00} = -1$ gilt. Im Gegensatz zu (377.3) wird also kein stationäres weißes Rauschen angenommen. Für den Erwartungswert von $X(m, n)$ gelte

$$E(X(m, n)) = \mu_x(m, n) = \mu_x = 0 \quad (378.4)$$

Löst man (378.1) nach $\varepsilon(m, n)$ auf, erhält man

$$\varepsilon(m, n) = - \sum_{k=-p}^p \sum_{l=-q}^q a_{kl} X(m-k, n-l) \quad (378.5)$$

was nach (244.29) einer nichtkausalen nichtrekursiven Filterung entspricht. Die Systemfunktion lautet nach (244.27) mit (244.24) und $z_i = \exp(j\Omega_i)$ für $i \in \{1, 2\}$

$$H(z_1, z_2) = - \sum_{k=-p}^p \sum_{l=-q}^q a_{kl} z_1^{-k} z_2^{-l} . \quad (378.6)$$

Der zu (378.5) inverse Prozeß ist durch

$$X(m, n) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \bar{a}_{kl} \varepsilon(m-k, n-l) \quad (378.7)$$

gegeben und besitzt die zu (378.6) inverse Systemfunktion $(H(z_1, z_2))^{-1}$. Dieses System ist nur dann stabil, wenn (235.12) zutrifft. Für die Impulsantwort $h(k, l)$ gilt nach (235.10) $h(k, l) = \bar{a}_{kl}$. Die Stabilitätsforderung lautet daher

$$\sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |\bar{a}_{kl}| < \infty \quad (378.8)$$

analog zu (377.8). Die Momentfunktion $K_{xx}(m_1, n_1, m_2, n_2)$ des nAR(p,q)-Prozesses ist durch (377.9) gegeben, wenn dort die Summation jeweils bei $-\infty$ beginnt. Die Momentfunktion ist mit (378.8) ebenso wie der Erwartungswert beschränkt, so daß das Ausgangssignal $X(m, n)$ aus (378.1) zumindest im weiteren Sinne stationär ist. Die Forderung (378.8) stellt somit wieder die *Homogenitätsbedingung* dar, die im folgenden als erfüllt angenommen wird.

Nach (362.6) folgt für die Autokovarianzen der Rauschanteile

$$C_{\varepsilon\varepsilon}(k, l) = C_{\varepsilon\varepsilon}(-k, -l)$$

und daher aus (378.3) für die Koeffizienten a_{kl} die Symmetriebedingung

$$a_{kl} = a_{-k, -l} . \quad (378.9)$$

Ersetzt man in (378.5) m durch $m-k_1$ und n durch $n-k_2$ für alle k_1 und k_2 , multipliziert anschließend mit $\varepsilon(m, n)$ und bildet den Erwartungswert, ergibt sich

$$E(\varepsilon(m, n)\varepsilon(m-k_1, n-k_2)) = - \sum_{k=-p}^p \sum_{l=-q}^q a_{kl} E(\varepsilon(m, n)X(m-(k_1+k), n-(k_2+l)))$$

oder mit (378.3) für $-p \leq k_1 \leq p$ und $-q \leq k_2 \leq q$

$$a_{k_1 k_2} \sigma_{\varepsilon}^2 = \sum_{k=-p}^p \sum_{l=-q}^q a_{kl} C_{\varepsilon x}(k_1+k, k_2+l) . \quad (378.10)$$

Erfüllt wird diese Gleichung durch die Kreuzkovarianzfunktion

$$C_{\varepsilon x}(k, l) = E(\varepsilon(m, n)X(m-k, n-l)) = \sigma_{\varepsilon}^2 d(k, l) . \quad (378.11)$$

Setzt man nämlich (378.11) in (378.10) ein, folgt für die rechte Seite mit (378.9)

$$\sum_{k=-p}^p \sum_{l=-q}^q a_{k1} \sigma_{\epsilon}^2 d(k_1+k, k_2+1) = a_{-k_1, -k_2} \sigma_{\epsilon}^2 = a_{k_1 k_2} \sigma_{\epsilon}^2 . \quad (378.12)$$

Durch die Autokovarianzfunktion (378.3) wird somit eine (377.12) entsprechende Kreuzkovarianzfunktion (378.11) erhalten. Für $k = l = 0$ ergibt sich

$$C_{\epsilon X}(0, 0) = \sigma_{\epsilon}^2 . \quad (378.13)$$

Multipliziert man (378.1) mit $X(m-p_1, n-p_2)$ für alle p_1, p_2 und bildet anschließend den Erwartungswert, folgt mit (378.11) die Autokovarianzfunktion

$$\begin{aligned} C_{XX}(p_1, p_2) &= E(X(m, n)X(m-p_1, n-p_2)) \\ &= \sum_{\substack{k=-p \\ (k, l) \neq (0, 0)}}^p \sum_{l=-q}^q a_{k1} E(X(m-k, n-l)X(m-p_1, n-p_2)) + E(\epsilon(m, n)X(m-p_1, n-p_2)) \\ &= \sum_{\substack{k=-p \\ (k, l) \neq (0, 0)}}^p \sum_{l=-q}^q a_{k1} C_{XX}(p_1-k, p_2-l) + \sigma_{\epsilon}^2 d(p_1, p_2) . \end{aligned} \quad (378.14)$$

Für $p_1 = p_2 = 0$ erhält man die Varianz

$$\sigma_X^2 = C_{XX}(0, 0) = \sum_{\substack{k=-p \\ (k, l) \neq (0, 0)}}^p \sum_{l=-q}^q a_{k1} C_{XX}(-k, -l) + \sigma_{\epsilon}^2 \quad (378.15)$$

und für $(p_1, p_2) \neq (0, 0)$

$$C_{XX}(p_1, p_2) = \sum_{\substack{k=-p \\ (k, l) \neq (0, 0)}}^p \sum_{l=-q}^q a_{k1} C_{XX}(p_1-k, p_2-l) . \quad (378.16)$$

Die Autokorrelationsfunktion $\rho_X(p_1, p_2)$, die zweidimensionale Yule-Walker-Gleichung, ergibt sich aus der Division von (378.16) durch (378.15).

Die Autokovarianzfunktion (378.16) soll wieder in Matrizendarstellung geschrieben werden. Dazu werden die Koeffizienten a_{k1} in den Vektoren

$$\begin{aligned} \mathbf{a}_0 &= [a_{-p, 0}, a_{-p+1, 0}, \dots, a_{-1, 0}, a_{1, 0}, \dots, a_{p, 0}]' \\ \mathbf{a}_j &= [a_{-p, j}, a_{-p+1, j}, \dots, a_{-1, j}, a_{0, j}, a_{1, j}, \dots, a_{p, j}]' \end{aligned} \quad (378.17)$$

für $j \in \{-q, \dots, -1, 1, \dots, q\}$ zusammengefaßt. Ebenso wird mit den Autokovarianzen $C_{XX}(m, n) = \gamma_{m, n}$ verfahren

$$\begin{aligned} \boldsymbol{\gamma}_0 &= [\gamma_{-p, 0}, \gamma_{-p+1, 0}, \dots, \gamma_{-1, 0}, \gamma_{1, 0}, \dots, \gamma_{p, 0}]' \\ \boldsymbol{\gamma}_j &= [\gamma_{-p, j}, \gamma_{-p+1, j}, \dots, \gamma_{-1, j}, \gamma_{0, j}, \gamma_{1, j}, \dots, \gamma_{p, j}]' . \end{aligned} \quad (378.18)$$

Führt man weiterhin die Matrizen

$$\Sigma_{00} = \begin{bmatrix} \gamma_{0,0} & \gamma_{-1,0} & \cdots & \gamma_{-p+1,0} & \gamma_{-p-1,0} & \cdots & \gamma_{-2p,0} \\ \gamma_{1,0} & \gamma_{0,0} & \cdots & \gamma_{-p+2,0} & \gamma_{-p,0} & \cdots & \gamma_{-2p+1,0} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_{p-1,0} & \gamma_{p-2,0} & \cdots & \gamma_{0,0} & \gamma_{-2,0} & \cdots & \gamma_{-p-1,0} \\ \gamma_{p+1,0} & \gamma_{p,0} & \cdots & \gamma_{2,0} & \gamma_{0,0} & \cdots & \gamma_{-p+1,0} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_{2p,0} & \gamma_{2p-1,0} & \cdots & \gamma_{p+1,0} & \gamma_{p-1,0} & \cdots & \gamma_{0,0} \end{bmatrix}$$

$$\Sigma_{0j} = \begin{bmatrix} \gamma_{0,-j} & \gamma_{-1,-j} & \cdots & \gamma_{-2p,-j} \\ \gamma_{1,-j} & \gamma_{0,-j} & \cdots & \gamma_{-2p+1,-j} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{p-1,-j} & \gamma_{p-2,-j} & \cdots & \gamma_{-p-1,-j} \\ \gamma_{p+1,-j} & \gamma_{p,-j} & \cdots & \gamma_{-p+1,-j} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{2p,-j} & \gamma_{2p-1,-j} & \cdots & \gamma_{0,-j} \end{bmatrix} = \Sigma_{j0}'$$

$$\Sigma_{ij} = \begin{bmatrix} \gamma_{0,i-j} & \gamma_{-1,i-j} & \cdots & \gamma_{-2p,i-j} \\ \gamma_{1,i-j} & \gamma_{0,i-j} & \cdots & \gamma_{-2p+1,i-j} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{2p,i-j} & \gamma_{2p-1,i-j} & \cdots & \gamma_{0,i-j} \end{bmatrix} = \Sigma_{ji}' \quad (378.19)$$

ein, so erhält man für die Autokovarianzen $C_{xx}(p_1, p_2)$ für $p_1 \in \{-p, -p+1, \dots, p\}$ und $p_2 \in \{-q, -q+1, \dots, q\}$ aus (378.16) das lineare Gleichungssystem

$$\begin{bmatrix} \gamma_{-q} \\ \cdots \\ \gamma_{-1} \\ \gamma_0 \\ \gamma_1 \\ \cdots \\ \gamma_q \end{bmatrix} = \begin{bmatrix} \Sigma_{-q,-q} & \cdots & \Sigma_{-q,-1} & \Sigma_{-q,0} & \Sigma_{-q,1} & \cdots & \Sigma_{-q,q} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \Sigma_{-1,-q} & \cdots & \Sigma_{-1,-1} & \Sigma_{-1,0} & \Sigma_{-1,1} & \cdots & \Sigma_{-1,q} \\ \Sigma_{0,-q} & \cdots & \Sigma_{0,-1} & \Sigma_{00} & \Sigma_{01} & \cdots & \Sigma_{0q} \\ \Sigma_{1,-q} & \cdots & \Sigma_{1,-1} & \Sigma_{10} & \Sigma_{11} & \cdots & \Sigma_{1q} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \Sigma_{q,-q} & \cdots & \Sigma_{q,-1} & \Sigma_{q0} & \Sigma_{q1} & \cdots & \Sigma_{qq} \end{bmatrix} \begin{bmatrix} a_{-q} \\ \cdots \\ a_{-1} \\ a_0 \\ a_1 \\ \cdots \\ a_q \end{bmatrix}, \quad (378.20)$$

das analog (377.20) in der Form

$$c_{pq} = \Sigma_{pq} \beta \quad (378.21)$$

vereinfacht dargestellt wird. Die Vektoren c_{pq} und β enthalten jeweils $s = (2p+1)(2q+1) - 1$ Komponenten. Σ_{pq} ist eine symmetrische $s \times s$ Matrix. Aus (378.21) lassen sich wieder mit $r_{pq} = c_{pq}/\sigma_x^2$ und $R_{pq} = \Sigma_{pq}/\sigma_x^2$ die zweidimensionalen Yule-Walker Gleichungen aufstellen.

Das Spektrum $S_{\varepsilon\varepsilon}(\Omega_1, \Omega_2)$ des Rauschanteils berechnet sich aus (363.8) mit (378.3) zu

$$\begin{aligned}
 S_{\varepsilon\varepsilon}(\Omega_1, \Omega_2) &= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} C_{\varepsilon\varepsilon}(k, l) e^{-j(k\Omega_1 + l\Omega_2)} \\
 &= - \sum_{k=-p}^p \sum_{l=-q}^q a_{kl} \sigma_{\varepsilon}^2 e^{-j(k\Omega_1 + l\Omega_2)} .
 \end{aligned}$$

Mit $z_i = \exp(j\Omega_i)$ für $i \in \{1, 2\}$ erhält man aus der Systemfunktion (378.6) des Prozesses (378.5) die Frequenzantwort $H(j\Omega_1, j\Omega_2)$, mit der

$$S_{\varepsilon\varepsilon}(\Omega_1, \Omega_2) = \sigma_{\varepsilon}^2 H(j\Omega_1, j\Omega_2) \quad (378.22)$$

folgt. Bezeichnet man die Impulsantwort des Prozesses (378.5) mit $h_1(k, l)$, gilt wegen (378.9) nach (235.10)

$$h_1(k, l) = h_1(-k, -l) = -a_{kl} .$$

Die Frequenzantwort $H(j\Omega_1, j\Omega_2)$ ist somit nach (241.40) reell und nach (241.50) zwei-quadrantensymmetrisch

$$H(j\Omega_1, j\Omega_2) = H(-j\Omega_1, -j\Omega_2) = A(\Omega_1, \Omega_2) \quad (378.23)$$

und lautet nach (244.14)

$$A(\Omega_1, \Omega_2) = 1 - 2 \sum_{l=1}^q a_{0l} \cos(l\Omega_2) - 2 \sum_{k=1}^p \sum_{l=-q}^q a_{kl} \cos(k\Omega_1 + l\Omega_2) . \quad (378.24)$$

Das Spektrum $S_{xx}(\Omega_1, \Omega_2)$ des nAR(p,q)-Prozesses berechnet sich aus (377.22) mit (378.22) und (378.23) zu

$$\begin{aligned}
 S_{xx}(\Omega_1, \Omega_2) &= S_{\varepsilon\varepsilon}(\Omega_1, \Omega_2) (H(j\Omega_1, j\Omega_2))^{-1} (H(-j\Omega_1, -j\Omega_2))^{-1} \\
 &= \sigma_{\varepsilon}^2 / A(\Omega_1, \Omega_2) .
 \end{aligned} \quad (378.25)$$

Wie bereits im Kapitel 377 erwähnt, kann bei verschiedenen Anwendungen vorausgesetzt werden, daß die Zufallsfelder die Markoff-Eigenschaft aufweisen. Im folgenden soll daher bewiesen werden, daß der nAR(p,q)-Prozeß (378.1), dessen Zufallsvariable $X(m, n)$ und dessen Rauschanteil $\varepsilon(m, n)$ als normalverteilt angenommen werden, ein Markoff-Zufallsfeld im Sinne von (372.2) oder (372.3) ist. Es besitzt also die symmetrische Markoff-Eigenschaft im Gegensatz zu dem im Kapitel 377 eingeführten kausalen zweidimensionalen Markoff-Zufallsfeld, das auf der einseitigen Markoff-Eigenschaft basiert. Der Beweis der Umkehrung, daß ein homogenes Gauß-Markoff-Zufallsfeld als nAR(p,q)-Prozeß darstellbar ist, befindet sich bei Woods (1972) und Chellappa (1985). Weitergehende Untersuchungen zu Markoff-Eigenschaften zweidimensionaler autoregressiver Prozesse nahm Kashyap (1984) vor.

Betrachtet wird das Eingangssignal $X(m, n)$ für $m \in \{0, \dots, M\}$ und $n \in \{0, \dots, N\}$, wobei $M > p$ und $N > q$ vorausgesetzt sei. Mit der hier verwendeten Notation lautet dann (372.3),

falls der Wert der Zufallsvariablen $X(m,n)$ ebenfalls mit $X(m,n)$ bezeichnet wird,

$$p(X(m,n) | X(M,N), X(M-1,N), \dots, X(0,N), X(M,N-1), X(M-1,N-1), \dots, X(0,N-1), \dots, X(m+1,n), X(m-1,n), \dots, X(0,0)) = p(X(m,n) | \partial X(m,n)) \quad (378.26)$$

wobei $\partial X(m,n)$ die Werte der Zufallsvariablen $X(m-k,n-1)$ für $k \in \{-p, \dots, p\}$ und $l \in \{-q, \dots, q\}$ außer $(k,l) = (0,0)$ in der Nachbarschaft von $X(m,n)$ bezeichnet. Der Beweis wird analog zu (377.25) bis (377.28) geführt. Sind die Zufallsvariablen $\varepsilon(m,n)$ und $X(m-k,n-1)$ für beliebige k,l außer $(k,l) = (0,0)$ normalverteilt, folgt aus (378.11) die Unabhängigkeit. Es gilt daher für die bedingte Dichte

$$p(\varepsilon(m,n) | \partial X(m,n)) = p(\varepsilon(m,n)) \quad (378.27)$$

Für den Erwartungswert erhält man somit wegen (378.2)

$$E(\varepsilon(m,n) | \partial X(m,n)) = E(\varepsilon(m,n)) = 0 \quad (378.28)$$

Bildet man den Erwartungswert der Zufallsvariablen $X(m,n)$ unter der Bedingung, daß die Werte für $X(M,N), \dots, X(0,0)$ gegeben seien, folgt aus (378.1) mit (378.28)

$$\begin{aligned} E(X(m,n) | X(M,N), X(M-1,N), \dots, X(0,N), X(M,N-1), X(M-1,N-1), \dots, X(0,N-1), \dots, X(m+1,n), X(m-1,n), \dots, X(0,0)) &= E(X(m,n) | \partial X(m,n)) \\ &= \sum_{\substack{k=-p \\ (k,l) \neq (0,0)}}^p \sum_{l=-q}^q a_{kl} X(m-k,n-1) \end{aligned} \quad (378.29)$$

Außerdem folgt für die bedingte Varianz von $X(m,n)$ aus (378.1) mit (378.3)

$$\begin{aligned} V(X(m,n) | X(M,N), X(M-1,N), \dots, X(0,N), X(M,N-1), X(M-1,N-1), \dots, X(0,N-1), \dots, X(m+1,n), X(m-1,n), \dots, X(0,0)) &= V(X(m,n) | \partial X(m,n)) \\ &= V(\varepsilon(m,n)) = \sigma_\varepsilon^2 \end{aligned} \quad (378.30)$$

Da $X(m,n)$ normalverteilt ist, gilt wegen (378.29) und (378.30) auch die Markoff-Eigenschaft (378.26). Der nAR(p,q)-Prozeß stellt also ein Markoff-Zufallsfeld im Sinne von (372.3) dar. Der Nachweis der Markoff-Eigenschaft (378.26) über die bedingte Dichte analog zu (325.30) bis (325.35) im eindimensionalen Fall ist bei (Pan 1994) zu finden.

Aufgrund der Normalverteilung für $X(m,n)$ läßt sich aus (378.29) und (378.30) die bedingte Normalverteilung für $X(m,n) | \partial X(m,n)$ angeben. Man erhält

$$X(m,n) | \partial X(m,n) \sim N\left(\sum_{\substack{k=-p \\ (k,l) \neq (0,0)}}^p \sum_{l=-q}^q a_{kl} X(m-k,n-1), \sigma_\varepsilon^2\right) \quad (378.31)$$

Multipliziert man den quadratischen Ausdruck im Exponenten der Dichte dieser Vertei-

lung aus und absorbiert die Konstante in der Normierungskonstanten, erkennt man, daß die Dichte identisch ist mit der bedingten Dichte (376.14). Substituiert man nämlich in (376.14) $k-m = p_1$ und $l-n = p_2$, ergibt sich mit $\sigma^2 = \sigma_\varepsilon^2$, $\beta_{p_1 p_2} = a_{p_1 p_2}$ und $\mu_{mn} = \mu_{k1} = 0$ in der hier verwendeten Notation

$$p(X(m,n) | \partial X(m,n)) \propto \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} [(X(m,n))^2 - 2 X(m,n) \sum_{\substack{p_1=-p \\ (p_1, p_2) \neq (0,0)}}^p \sum_{\substack{p_2=-q}}^q a_{p_1 p_2} X(m+p_1, n+p_2)]\right\} \quad (378.32)$$

und weiter nach Berücksichtigung von (378.9) sowie $p_1 = -k$ und $p_2 = -l$

$$p(X(m,n) | \partial X(m,n)) \propto \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} [(X(m,n))^2 - 2 X(m,n) \sum_{\substack{k=-p \\ (k, l) \neq (0,0)}}^p \sum_{l=-q}^q a_{k1} X(m-k, n-l)]\right\} \quad (378.33)$$

in Übereinstimmung mit (378.31). Das Gauß-Zufallsfeld als Spezialfall des Auto-Modells besitzt also dieselbe bedingte Verteilung wie der nichtkausale autoregressive Prozeß mit Markoff-Eigenschaft.

Analog zu den Ableitungen (377.30) bis (377.37) läßt sich die bedingte Verteilung (378.31) auch aus den Normalverteilungen für $X(m,n)$ und $\varepsilon(m,n)$ herleiten, was aber hier nicht vorgeführt wird.

Um die gemeinsame Verteilung der $X(m,n)$ für $0 \leq m \leq M$ und $0 \leq n \leq N$ abzuleiten, wird zunächst das toroidale Gitter zugrunde gelegt, siehe Abbildung 376-4. Mit der in (376.18) definierten Menge Ω der Gitterpunkte gilt für ein Zufallsfeld $Y(m,n)$ auf einem toroidalen Gitter

$$Y(m-k, n-l) = \begin{cases} X(m-k, n-l) & \text{für } (m-k, n-l) \in \Omega \\ X((m-k) \bmod (M+1), (n-l) \bmod (N+1)) & \text{für } (m-k, n-l) \notin \Omega \end{cases} \quad (378.34)$$

An dieser Stelle sei noch die mod-Funktion erklärt. Trifft beispielsweise für $m-k$ eine der Ungleichungen $m-k < 0$ oder $m-k > M$ zu, so bedeutet $(m-k) \bmod (M+1) = r$, daß sovielen Vielfache von $(M+1)$ zu dem Wert $m-k$ zu addieren oder zu subtrahieren sind, bis für r die Bedingung $0 \leq r \leq M$ gilt.

Mit (378.34) erhält man anstelle von (378.1) den nAR(p,q)-Prozeß

$$X(m,n) = Y(m,n) = \sum_{\substack{k=-p \\ (k, l) \neq (0,0)}}^p \sum_{l=-q}^q a_{k1} Y(m-k, n-l) + \varepsilon(m,n) \quad (378.35)$$

und anstelle von (378.5)

$$\varepsilon(m, n) = - \sum_{k=-p}^p \sum_{l=-q}^q a_{kl} Y(m-k, n-l) . \quad (378.36)$$

Führt man zunächst die Vektoren

$$\mathbf{x}_j = [X(0, j), X(1, j), \dots, X(M, j)]' \quad (378.37)$$

und

$$\boldsymbol{\varepsilon}_j = [\varepsilon(0, j), \varepsilon(1, j), \dots, \varepsilon(M, j)]' \quad (378.38)$$

für $j \in \{0, \dots, N\}$ ein und definiert weiter

$$\mathbf{x} = [\mathbf{x}_0', \mathbf{x}_1', \dots, \mathbf{x}_N']' \quad (378.39)$$

sowie

$$\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_0', \boldsymbol{\varepsilon}_1', \dots, \boldsymbol{\varepsilon}_N']' , \quad (378.40)$$

erhält man aus (378.36) unter Beachtung von (378.34)

$$\boldsymbol{\varepsilon} = \mathbf{B}\mathbf{x} \quad (378.41)$$

mit der Koeffizientenmatrix

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_0 & \mathbf{B}_1 & \dots & \mathbf{B}_q & \mathbf{0} & \dots & \mathbf{0} & \mathbf{B}_{-q} & \dots & \mathbf{B}_{-1} \\ \mathbf{B}_{-1} & \mathbf{B}_0 & \dots & \mathbf{B}_{q-1} & \mathbf{B}_q & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{B}_{-2} \\ \dots & \dots \\ \mathbf{B}_1 & \mathbf{B}_2 & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{B}_{-q} & \mathbf{B}_{1-q} & \dots & \mathbf{B}_0 \end{bmatrix} \quad (378.42)$$

und den Untermatrizen \mathbf{B}_k für $k \in \{-q, \dots, q\}$, die unter Beachtung der Symmetriebedingung (378.9) aufgestellt sind

$$\mathbf{B}_k = - \begin{bmatrix} a_{0k} & a_{1k} & \dots & a_{pk} & 0 & \dots & 0 & a_{-p,k} & \dots & a_{-1,k} \\ a_{-1,k} & a_{0k} & \dots & a_{p-1,k} & a_{pk} & \dots & 0 & 0 & \dots & a_{-2,k} \\ \dots & \dots \\ a_{1k} & a_{2k} & \dots & 0 & 0 & \dots & a_{-p,k} & a_{1-p,k} & \dots & a_{0k} \end{bmatrix} = \mathbf{B}_{-k}' . \quad (378.43)$$

Die wegen (378.43) symmetrische Matrix \mathbf{B} ist eine zirkuläre Blockmatrix (Davis 1979, S.66). Ebenso sind die Untermatrizen \mathbf{B}_k zirkuläre Matrizen.

Da vorausgesetzt wurde, daß der zu (378.5) inverse Prozeß (378.7) existiert, ist \mathbf{B} regulär, so daß (378.41) nach \mathbf{x} aufgelöst werden kann

$$\mathbf{x} = \mathbf{B}^{-1}\boldsymbol{\varepsilon} . \quad (378.44)$$

Die Matrix $\boldsymbol{\Sigma}_x$ der Autokovarianzen der Komponenten $X(m, n)$ von \mathbf{x} ergibt sich daher

aus

$$\Sigma_{\mathbf{x}} = \mathbf{B}^{-1} \Sigma_{\varepsilon} \mathbf{B}^{-1} , \quad (378.45)$$

worin Σ_{ε} die Matrix der Autokovarianzen $C_{\varepsilon\varepsilon}(k, l)$ der Rauschanteile $\varepsilon(m, n)$ bedeutet. Anstelle von (378.3) erhält man auf einem toroidalen Gitter für die Autokovarianzen

$$E(\varepsilon(m, n) \varepsilon(m-k, n-l)) = C_{\varepsilon\varepsilon}(k_1, k_2) = \begin{cases} -a_{k_1 k_2} \sigma_{\varepsilon}^2 & \text{für } -p \leq k_1 \leq p, \quad -q \leq k_2 \leq q \\ 0 & \text{sonst,} \end{cases} \quad (378.46)$$

wobei für k_1 und k_2 die Rechenvorschriften

$$k_1 = k + \alpha_1(M+1) \quad \text{und} \quad k_2 = l + \alpha_2(N+1) \quad (378.47)$$

mit $\alpha_1, \alpha_2 \in \{-1, 0, 1\}$ lauten. Die Werte für α_i mit $i \in \{1, 2\}$ sind so zu wählen, daß $|k_i| = \min$ gilt. Die Matrix Σ_{ε} ergibt sich dann zu

$$\Sigma_{\varepsilon} = \sigma_{\varepsilon}^2 \mathbf{B} . \quad (378.48)$$

Weiter folgt aus (378.45)

$$\Sigma_{\mathbf{x}} = \sigma_{\varepsilon}^2 \mathbf{B}^{-1} . \quad (378.49)$$

Da die Komponenten $X(m, n)$ des Vektors \mathbf{x} als normalverteilt vorausgesetzt werden, erhält man

$$\mathbf{x} \sim N(\mathbf{0}, \sigma_{\varepsilon}^2 \mathbf{B}^{-1}) ,$$

so daß die gemeinsame Dichte der $X(m, n)$ folgt mit

$$p(\mathbf{x}) \propto \exp\left\{-\frac{1}{2\sigma_{\varepsilon}^2} \mathbf{x}' \mathbf{B} \mathbf{x}\right\} . \quad (378.50)$$

Bei einem Gitter mit offenem Rand enthält die Matrix \mathbf{B} aus (378.42) Nullmatrizen am rechten oberen und linken unteren Rand und die Matrix \mathbf{B}_k aus (378.43) Nullelemente im rechten oberen und linken unteren Teil. Die Matrizen besitzen dann Bandstruktur. Die übrigen von Null verschiedenen Elemente stimmen überein, so daß auch für Gitter mit offenen Rändern mit entsprechenden Matrizen \mathbf{B} die Beziehungen (378.41), (378.48) und (378.50) gültig sind.

Schreibt man (376.16) wie für (378.33) um, erhält man mit (378.34), (378.36) und (378.41)

$$\begin{aligned} \sum_{m=0}^M \sum_{n=0}^N \sum_{k=-p}^p \sum_{l=-q}^q Y(m, n) (-a_{k_1}) Y(m-k, n-l) &= \sum_{m=0}^M \sum_{n=0}^N X(m, n) \varepsilon(m, n) \\ &= \mathbf{x}' \boldsymbol{\varepsilon} = \mathbf{x}' \mathbf{B} \mathbf{x} . \end{aligned} \quad (378.51)$$

Die Dichten (376.17) und (378.50) sind also identisch. Der nichtkausale autoregressive Prozeß mit Markoff-Eigenschaft besitzt also dieselbe gemeinsame Verteilung wie das Gauß-Zufallsfeld als Spezialfall des Auto-Modells.

Wendet man zur Schätzung der unbekanntenen Koeffizienten a_{k1} des autoregressiven Prozesses die Maximum-Likelihood-Methode an und bestimmt die Likelihoodfunktion durch die Dichte (378.50), ergeben sich wegen $(\det \mathbf{B})^{1/2}$ in der Normierungskonstanten von (378.50) nichtlineare Gleichungen für die Schätzwerte der a_{k1} , die hier nicht untersucht werden. Lineare Gleichungssysteme erhält man mit einer genäherten Likelihoodfunktion, die aus der bedingten Normalverteilung (378.31) folgt, wie im Zusammenhang mit (383.24) erläutert wird. Auch für den kausalen autoregressiven Prozeß war mit (377.47) eine genäherte Likelihoodfunktion für die Maximum-Likelihood-Methode betrachtet worden, die auf ein lineares Gleichungssystem für die unbekanntenen Parameter führt.

379 Anwendungen in der digitalen Bildverarbeitung

Für die Analyse digitaler Bilder stellen die auf Gittern definierten Zufallsfelder geeignete mathematische Hilfsmittel dar, um die in den Bildern enthaltenen Phänomene zu modellieren. Die Messungen der Grauwerte der Bildelemente eines digitalen Bildes lassen sich beispielsweise als ein Zufallsfeld auf einem Gitter einführen. Auch die unbekanntenen Grauwerte, die zur Rekonstruktion eines digitalen Bildes aus den Messungen eingeführt werden, repräsentieren ein Zufallsfeld. Die Zugehörigkeit von Bildelementen zu Texturen ist ein weiteres Beispiel für ein Zufallsfeld auf einem Gitter. Liegen die Messungen der Grauwerte eines digitalen Bildes nicht nur in einem Spektralband des Lichtes vor, sondern in mehreren Spektralbändern, zum Beispiel im Spektralband des roten, grünen und blauen Lichts, können diese Messungen durch ein vektorielles Zufallsfeld mit drei Komponenten dargestellt werden, wobei das Zufallsfeld wieder für ein Gitter definiert ist.

Die Grauwerte der Bildelemente eines digitalen Bildes besitzen die Eigenschaft, daß sie im wesentlichen durch die Grauwerte in ihrer unmittelbaren Nachbarschaft beeinflusst werden. Die Wahrscheinlichkeit eines Grauwertes unter der Bedingung, daß sämtliche Grauwerte des Bildes gegeben sind, hängt also nur von den Grauwerten der Bildelemente der Nachbarschaft ab. Auch die Messung des Grauwertes eines Bildelementes oder die Zugehörigkeit eines Bildelementes zu einer Textur werden durch die Verhältnisse in den Bildelementen der Nachbarschaft beeinflusst. Im allgemeinen besitzen Zufallsfelder der digitalen Bildverarbeitung die Markoff-Eigenschaft, sie stellen Markoff-Zufallsfelder dar. Ihre Verteilungen lassen sich somit durch Gibbs-Verteilungen ausdrücken, da aufgrund

von (374.1) Markoff-Zufallsfelder und Nachbarschafts-Gibbs-Felder äquivalent sind. Eine weitere Möglichkeit, die Verteilungen von Zufallsfeldern mit Markoff-Eigenschaft auszudrücken, besteht durch die kausalen und nichtkausalen autoregressiven Prozesse, deren Markoff-Eigenschaft in den Kapiteln 377 und 378 nachgewiesen wird.

Um ein digitales Bild zu restaurieren, um also das bei der Aufnahme verursachte Rauschen zu beseitigen oder um Texturen und Muster in einem Bild zu erkennen, müssen unbekannte Parameter geschätzt werden. Man geht in der Regel so vor, daß die Verteilung der Beobachtungen in Abhängigkeit von den unbekanntem Parametern aufgestellt wird. Sie ergibt die Likelihoodfunktion, die mit $p(\mathbf{y}|\boldsymbol{\beta})$ bezeichnet sei, worin \mathbf{y} den Vektor der Beobachtungen und $\boldsymbol{\beta}$ den Vektor der unbekanntem Parameter bedeuten. Als Schätzwerte für die unbekanntem Parameter $\boldsymbol{\beta}$ lassen sich die Werte bestimmen, die für einen gegebenen Vektor \mathbf{y} von Beobachtungen maximale Werte der Likelihoodfunktion liefert. Diese Maximum-Likelihood-Methode, die bereits im Kapitel 354 angewendet wurde, wird zusammen mit weiteren Schätzverfahren noch einmal im Kapitel 383 behandelt.

In der digitalen Bildverarbeitung liegt häufig Vorinformation über die unbekanntem Parameter vor. Bei der Bildrestaurierung ist beispielsweise zu beachten, daß ein Bild, abgesehen von Unstetigkeitsstellen in Form von Kanten, glatt sein sollte, wenn man sich die dreidimensionale Darstellung seiner Grauwerte durch eine Fläche verbunden denkt. Bei der Mustererkennung weiß man, daß Texturen nicht nur aus wenigen Bildelementen bestehen, sondern sich über größere Flächen erstrecken. Aufgrund der Markoff-Eigenschaften der Zufallsfelder läßt sich diese Vorinformation durch Verteilungen ausdrücken, so daß die Bayes-Statistik angewendet werden kann. In diesem Zusammenhang wird die Verteilung der Vorinformation als Priori-Verteilung bezeichnet, die durch Multiplikation mit der Likelihoodfunktion auf die Posteriori-Verteilung führt, die zur Parameterschätzung genutzt wird. Hierauf wird im Kapitel 383 eingegangen.

Aufgaben der digitalen Bildverarbeitung, die häufig gelöst werden müssen, sind die Bildrestaurierung, die Kantenextraktion und die Texturerkennung. Für diese drei Probleme sollen im folgenden die Likelihoodfunktionen und die Priori-Verteilungen angegeben werden.

a) Bildrestaurierung

Bei der Bildrestaurierung liegen als Messungen die Grauwerte der Bildelemente einer digitalen Aufnahme vor. Sie ist durch Rauschen verfälscht, so daß aus den Messungen die Grauwerte des Bildes zu rekonstruieren sind, die sich bei einem Aufnahmevorgang ohne Rauschen ergeben würden. Das entsprechend (376.1) definierte Zufallsfeld

$$Y(m,n) = Y(t_1) \quad \text{mit} \quad t_1 = [m,n]'$$

$$\text{und} \quad m \in \{0, 1, \dots, M\}, \quad n \in \{0, 1, \dots, N\}, \quad l \in \{0, 1, \dots, (M+1)(N+1)-1\} \quad (379.1)$$

repräsentiere den Meßprozeß der Grauwerte. Die Messung y_{mn} des Grauwertes eines Bildelementes (m,n) ergibt mit $Y(m,n) = y_{mn}$ einen Wert der Zufallsvariablen $Y(m,n)$ und wird mit $\mathbf{y} = (y_{mn})$ im Vektor \mathbf{y} der Messungen aufgeführt. Mit $B(m,n)$ sei entsprechend (379.1) das Zufallsfeld der unbekanntes Grauwerte des unverrauschten Bildes definiert, das als Markoff-Zufallsfeld angenommen wird. Eine Realisierung von $B(m,n)$ sei mit

$$B(m,n) = \beta_{mn} \quad \text{und} \quad \boldsymbol{\beta} = (\beta_{mn}) \quad (379.2)$$

im Vektor $\boldsymbol{\beta}$ der unbekanntes Parameter enthalten.

Sowohl das Zufallsfeld $Y(m,n)$ als auch $B(m,n)$ werden zunächst als skalarwertige Zufallsfelder eingeführt. Später wird angenommen, daß Messungen von Grauwerten in mehreren Frequenzbändern vorliegen, so daß die Verallgemeinerung auf vektorielle Zufallsfelder erfolgt.

Mit der nichtlinearen Funktion $f(\boldsymbol{\beta})$ mit $f(\boldsymbol{\beta}) = (f_{mn}(\boldsymbol{\beta}))$ sei der Abbildungsvorgang bezeichnet. Er wird häufig durch die lineare Punktspreizungsfunktion beschrieben (Bates und McDonnell 1989, S.7), so daß sich die Messungen y_{mn} als zweidimensionale Faltung (235.10) der unbekanntes Grauwerte β_{mn} mit der Punktspreizungsfunktion ergeben, die in (235.10) der Impulsantwort entspricht. Bei einer Punktspreizungsfunktion von endlicher Länge läßt sich die Faltung auch als Matrizenprodukt darstellen, und man erhält $\mathbf{f}(\boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$. Hiermit folgt für die Beziehung zwischen den Beobachtungen \mathbf{y} und den unbekanntes Parametern $\boldsymbol{\beta}$ das lineare Modell in der Bayes-Darstellung (Koch 1990, S.157)

$$E(\mathbf{y}|\boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta} \quad \text{und} \quad D(\mathbf{y}) = \boldsymbol{\Sigma} \quad , \quad (379.3)$$

in der $\boldsymbol{\Sigma}$ die Matrix der Autokovarianzen des Beobachtungsvektors \mathbf{y} bezeichnet.

Repräsentiert das Zufallsfeld $Y(m,n)$ der Messungen ein mit (371.2) definiertes Gauß-Zufallsfeld, ergibt sich aus (379.3) unmittelbar die Likelihoodfunktion, also die Dichte $p(\mathbf{y}|\boldsymbol{\beta})$ der Beobachtungen als Funktion der unbekanntes Parameter

$$p(\mathbf{y}|\boldsymbol{\beta}) \propto \exp\left[-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right] \quad . \quad (379.4)$$

Wie bereits im Zusammenhang mit (375.7) gezeigt wurde, liegt hiermit die Dichte einer Gibbs-Verteilung vor. Das Zufallsfeld $Y(m,n)$ der Messungen ist daher auch ein Markoff-Zufallsfeld.

Im folgenden wird zur Vereinfachung

$$\mathbf{X} = \mathbf{I} \quad \text{und} \quad \boldsymbol{\Sigma} = \text{diag}(\sigma_{00}^2, \sigma_{10}^2, \dots, \sigma_{M0}^2, \sigma_{01}^2, \dots, \sigma_{MN}^2) \quad (379.5)$$

gesetzt, so daß mit $r = (m, n)$ und $r \in \Omega$, wobei Ω die in (376.18) definierte Menge der Gitterpunkte bedeutet, anstelle von (379.4) die Likelihoodfunktion folgt

$$p(\mathbf{y}|\boldsymbol{\beta}) \propto \exp\left[-\sum_{r \in \Omega} (y_r - \beta_r)^2 / 2\sigma_r^2\right] . \quad (379.6)$$

Die bedingte Dichte $p(\mathbf{y}_r | \partial \mathbf{y}_r, \boldsymbol{\beta})$ der Messung y_r , die der bedingten Dichte (374.3) entspricht, ergibt sich dann zu

$$p(\mathbf{y}_r | \partial \mathbf{y}_r, \boldsymbol{\beta}) \propto \exp\left[-(y_r - \beta_r)^2 / 2\sigma_r^2\right] . \quad (379.7)$$

Wie bereits zu Beginn dieses Kapitels erwähnt, liegt bei der Bildrestaurierung für die unbekannt Parameter β_{mn} aus (379.2) die Vorinformation vor, daß das restaurierte Bild, von Kanten abgesehen, glatt sein sollte. Diese Vorinformation läßt sich bequem einbringen, da das Zufallsfeld $B(m, n)$ für die unbekannt Grauwerte nach Voraussetzung ein Markoff-Zufallsfeld und wegen (374.1) auch ein Nachbarschafts-Gibbs-Feld ist. In der bedingten Dichte (374.3) der Gibbs-Verteilung lassen sich daher die Potentiale $U_C(\boldsymbol{\beta})$ der Cliques derart einführen, daß für raue Bilder hohe Energien und damit geringe Dichtewerte folgen. Für die durch (376.18) definierte Nachbarschaft N_p mit $s \in N_p$ des Markoff-Zufallsfeldes der Ordnung p wird dies durch die folgende bedingte Dichte im Gitterpunkt $r = (m, n)$ mit $r \in \Omega$ erreicht

$$p(\beta_r | \partial \beta_r) \propto \exp\left\{-c_\beta \sum_{s \in N_p} [(\beta_r - \beta_{r+s})^2 + (\beta_r - \beta_{r-s})^2]\right\} . \quad (379.8)$$

Je größer die Grauwertdifferenzen in der Nachbarschaft für β_r ausfallen, desto kleiner sind die Dichtewerte der bedingten Verteilung für β_r und desto weniger wahrscheinlich ist die Rekonstruktion eines rauhen Bildes. Der konstante Faktor c_β ermöglicht die Gewichtung der Vorinformation. Die gemeinsame Dichte $p(\boldsymbol{\beta})$, das heißt die Priori-Dichte für $\boldsymbol{\beta}$, folgt wie in (376.2) durch Summation über sämtliche Cliques des Gitters mit

$$p(\boldsymbol{\beta}) \propto \exp\left[-c_\beta \sum_{r \in \Omega} \sum_{s \in N_p} (\beta_r - \beta_{r+s})^2\right] . \quad (379.9)$$

Anstelle der hier vorgestellten Dichten (379.8) und (379.9) lassen sich die Potentiale der Cliques auch derart definieren, daß bei umgekehrten Vorzeichen die Quadrate der Grauwertdifferenzen im Nenner stehen (Geman et al. 1987). Der Vorteil von (379.8) und (379.9) liegt darin begründet, daß es sich um Normalverteilungen handelt (Busch und Koch 1990), da sie Spezialfälle von (376.22) und (376.24) bilden.

Der Maximalwert der Dichte (379.8) ergibt sich für identische Grauwerte in der Nachbarschaft N_p . Die Dichte für die Vorinformation läßt sich auch derart gewinnen, daß Maximalwerte der Dichte für Grauwerte folgen, die auf einer Geraden liegen (Busch 1992, S.34).

Die Priori-Dichte (379.9) für die unbekanntes Grauwerte bewirkt eine Glättung des restaurierten Bildes. Befindet sich aber eine Unstetigkeitsstelle in Form einer Kante im Bild, ist die Glättung zwischen benachbarten Bildelementen unerwünscht, falls sie durch eine Kante getrennt werden. Um die Glättung zu verhindern, muß ein weiteres Zufallsfeld definiert werden, das die Kanten in einem Bild beschreibt. Am anschaulichsten ist es, wenn man annimmt, daß Kantenelemente zwischen den Bildelementen angeordnet sind, wie in Abbildung 379-1 dargestellt ist. Die Kantenelemente werden durch diskrete Zufallsvariable repräsentiert, die den Wert Null annehmen, wenn kein Kantenelement vorhanden ist, oder den Wert Eins, falls ein Kantenelement vorliegt.

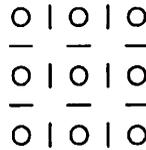


Abb. 379-1: Bild- und Kantenelemente

Es sei $G(\mathbf{t})$ das Zufallsfeld, das die unbekanntes Kantenelemente bildet, mit der Realisierung $\gamma_{mn;k l}$

$$G(\mathbf{t}) = \gamma_{mn;k l} \quad \text{und} \quad \boldsymbol{\gamma} = (\gamma_{mn;k l}) , \quad (379.10)$$

wobei die Indizes das Kantenelement bezeichnen, das zwischen den benachbarten Bildelementen (m, n) und (k, l) liegt. Im Vektor $\boldsymbol{\gamma}$ sei eine Realisierung von $G(\mathbf{t})$ zusammengefaßt. $G(\mathbf{t})$ wird als Markoff-Zufallsfeld angenommen.

Bei der Bildrestaurierung kann man davon ausgehen, daß die Likelihoodfunktion durch das Zufallsfeld der unbekanntes Kantenelemente nicht beeinflußt wird. Mit (379.6) erhält man also

$$p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) = p(\mathbf{y}|\boldsymbol{\beta}) \quad (379.11)$$

und mit (379.7)

$$p(y_r|\partial y_r, \boldsymbol{\beta}, \boldsymbol{\gamma}) = p(y_r|\partial y_r, \boldsymbol{\beta}) . \quad (379.12)$$

Eine Priori-Dichte für den Vektor $\boldsymbol{\beta}$ der unbekanntes Grauwerte und für den Vektor $\boldsymbol{\gamma}$ der unbekanntes Kantenelemente muß aber eingeführt werden. Mit der Definition (322.5) der bedingten Dichte $p(\boldsymbol{\beta}|\boldsymbol{\gamma})$ von $\boldsymbol{\beta}$ unter der Bedingung, daß $\boldsymbol{\gamma}$ gegeben ist, ergibt sich die Priori-Dichte $p(\boldsymbol{\beta}, \boldsymbol{\gamma})$ von $\boldsymbol{\beta}$ und $\boldsymbol{\gamma}$ zu

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}) = p(\boldsymbol{\beta}|\boldsymbol{\gamma})p(\boldsymbol{\gamma}) . \quad (379.13)$$

Die bedingte Dichte $p(\boldsymbol{\beta}|\boldsymbol{\gamma})$ erhält man durch einfache Modifikation von (379.8), wobei

zu bedenken ist, daß über ein Kantenelement nicht geglättet werden darf, falls es vorhanden ist, wenn also $\gamma_{r;p} = 1$ mit $r = (m, n)$ und $p = (k, l)$ gilt,

$$p(\beta_r | \partial\beta_r, \gamma_{r;p}) \propto \exp\{-c_{\beta\gamma} \sum_{s \in N_p} [(1 - \gamma_{r;r+s})(\beta_r - \beta_{r+s})^2 + (1 - \gamma_{r;r-s})(\beta_r - \beta_{r-s})^2]\} . \quad (379.14)$$

Hierin bedeutet $c_{\beta\gamma}$ wieder eine Konstante für die Gewichtung. Die gemeinsame bedingte Dichte $p(\beta | \gamma)$ folgt analog zu (379.9) aus

$$p(\beta | \gamma) \propto \exp[-c_{\beta\gamma} \sum_{r \in \Omega} \sum_{s \in N_p} (1 - \gamma_{r;r+s})(\beta_r - \beta_{r+s})^2] . \quad (379.15)$$

Um die Priori-Dichte $p(\gamma)$ der Kantenelemente anzugeben, wird für das Markoff-Zufallsfeld $G(t)$ der Kantenelemente, das wegen (374.1) auch ein Nachbarschafts-Gibbs-Feld ist, das Nachbarschaftssystem eingeführt, das die drei in Abbildung 379-2 a), b) und c) dargestellten Cliquentypen enthält, die von Null verschiedene Potentiale erzeugen

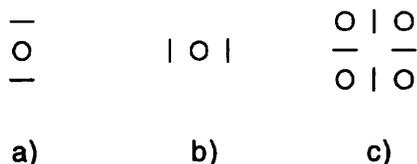


Abb. 379-2: Drei Cliquentypen für Kantenelemente

(Busch und Koch 1990; Geman und Geman 1984). Jeweils zwei der abgebildeten Cliquentypen werden kombiniert, so daß horizontale und vertikale Kantenelemente die in Abbildung 379-3 a) und b) dargestellten Nachbarn besitzen.

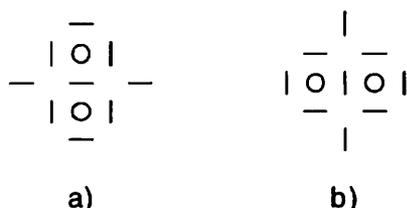


Abb. 379-3: Nachbarn a) eines horizontalen und b) eines vertikalen Kantenelementes

Die bedingte Dichte eines Kantenelementes $\gamma_{r;p}$ erhält man nach (374.3) zu

$$p(\gamma_{r;p} | \partial\gamma_{r;p}) \propto \exp(-c_\gamma \sum_{C(t_0)} U_C(\gamma)) , \quad (379.16)$$

worin c_γ eine Konstante und t_0 den Indexvektor für das Kantenelement $\gamma_{r;p}$ bezeichnet.

Durch Summation über sämtliche Cliques C folgt dann nach (373.10) die Priori-Dichte $p(\gamma)$ der Kantenelemente

$$p(\gamma) \propto \exp(-c_\gamma \sum_C U_C(\gamma)) . \quad (379.17)$$

Um die Potentiale $U_C(\gamma)$ der Cliquentypen anzugeben, muß man berücksichtigen, daß der Cliquentyp der Abbildung 379-2 c) in den in Abbildung 379-4 angegebenen Zuständen und Rotationen dieser Zustände auftreten kann. Die Potentiale können aufgrund theoretischer Ansätze vergeben (Silverman et al. 1990) oder mit Hilfe von Objekten bestimmt werden, die in den auszuwertenden Bildern zu erwarten sind (Nadabar und Jain 1992). Aber auch rein empirisch lassen sich die Potentiale festlegen, wie es in Tabelle 379-1 für die verschiedenen Zustände der Cliquentypen der Abbildung 379-2 geschehen ist (Busch und Koch 1990). Ausgedehntere Nachbarschaften als die in Abbildung 379-3 angegebenen sind bei (Busch 1992, S.29) zu finden.

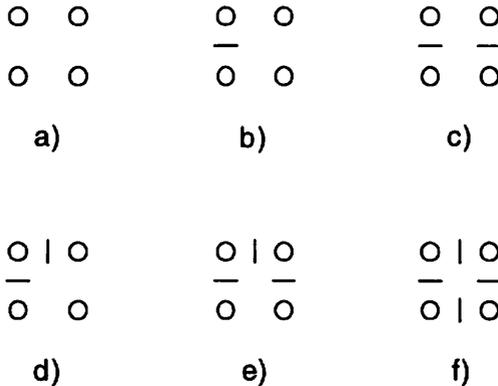


Abb. 379-4: Lage der Kantenelemente a) keine Kante, b) Ende, c) Fortsetzung, d) Richtungswechsel, e) drei Kanten, f) vier Kanten

Lage der Kantenelemente	Potentiale $U_C(\gamma)$
keine Kante	0
Ende	20
Fortsetzung	2
Richtungswechsel	2
drei Kanten	10
vier Kanten	15
parallele Kanten	25

Tabelle 379-1: Potentiale für Cliquentypen

Bislang wurden die Messungen der Grauwerte in nur einem Frequenzband berücksichtigt. Liegen Messungen in mehreren Frequenzbändern vor, sind sowohl für die Messungen als auch für die unbekanntes Grauwerte vektorielle Zufallsfelder zu definieren, deren Anzahl der Komponenten aus der Anzahl der Frequenzbänder folgt, in denen gemessen wurde. Es sei nach (376.28) und (376.29) $Y_k(m,n)$ die k -te Komponente des vektoriellen Zufallsfeldes der Messungen mit dem Wert y_{mnk} , folglich

$$Y_k(m,n) = y_{mnk} \quad \text{mit} \quad \mathbf{y} = (y_{mnk})$$

für $m \in \{0, 1, \dots, M\}$, $n \in \{0, 1, \dots, N\}$, $k \in \{1, 2, \dots, K\}$. (379.18)

Entsprechend gelte für die k -te Komponente des vektoriellen Zufallsfeldes der unbekanntes Grauwerte

$$B_k(m,n) = \beta_{mnk} \quad \text{mit} \quad \boldsymbol{\beta} = (\beta_{mnk}) . \quad (379.19)$$

Nimmt man an, daß die Messungen in den einzelnen Frequenzbändern voneinander unabhängig sind, braucht mit $r = (m,n)$ in der Likelihoodfunktion (379.6) nur der Index k hinzugefügt zu werden, über den dann zu summieren ist. Dies wurde bereits im Zusammenhang mit (376.29) festgestellt. Man erhält

$$p(\mathbf{y}|\boldsymbol{\beta}) \propto \exp\left[-\sum_{r \in \Omega} \sum_{k=1}^K (y_{rk} - \beta_{rk})^2 / 2\sigma_{rk}^2\right] . \quad (379.20)$$

Die bedingte Dichte $p(\boldsymbol{\beta}|\boldsymbol{\gamma})$ ergibt sich entsprechend aus (379.15), da vorausgesetzt werden kann, daß die Grauwerte, die den einzelnen Frequenzbändern zuzuordnen sind, voneinander unabhängig sind. Die Kantenelemente können in der Regel für die Bilder sämtlicher Frequenzbänder als identisch angesehen werden, da die abgebildeten Objekte identisch sind. Man erhält

$$p(\boldsymbol{\beta}|\boldsymbol{\gamma}) \propto \exp\left[-c_{\beta\gamma} \sum_{r \in \Omega} \sum_{s \in N_p} \sum_{k=1}^K (1 - \gamma_{r;r+s}) (\beta_{rk} - \beta_{r+s,k})^2\right] . \quad (379.21)$$

b) Kantenextraktion

Besteht die Aufgabe bei der Analyse digitaler Bilder lediglich darin, Kanten zu extrahieren, muß man berücksichtigen, daß die gemessenen Grauwerte Informationen über die Kanten enthalten. Die Likelihoodfunktion ist daher als Funktion der unbekanntes Kantenelemente zu formulieren, somit $p(\mathbf{y}|\boldsymbol{\gamma})$. Man erreicht dies auf einfache Weise, indem die gemessenen Grauwerte durch ein Polynom ausgedrückt werden, aus dem durch Bilden der ersten und zweiten Ableitung die Steigung und Krümmung ermittelt werden, die Hinweise auf Kanten geben (Haralick 1984). Im Kapitel 245 werden daher zur Kantenextraktion Ableitungsfiler angegeben. Paßt man den gemessenen Grauwerten y_r rechts

und links eines vertikalen Kantenlementes oder über und unter einem horizontalen Kantenlement ein eindimensionales Polynom an, ergibt sich der gemessene Grauwert $y(x)$ in Abhängigkeit von der Abszisse x

$$y(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3, \quad (379.22)$$

falls man sich auf vier Polynomkoeffizienten beschränkt. Die Ableitungen folgen mit

$$dy(x)/dx = \alpha_1 + 2\alpha_2 x + 3\alpha_3 x^2$$

und

$$d^2y(x)/dx^2 = 2\alpha_2 + 6\alpha_3 x.$$

Eine Kante ist beim Nulldurchgang der zweiten Ableitung zu erwarten, falls die erste Ableitung von Null verschieden ist. Man erhält $d^2y(x)/dx^2 = 0$ an der Stelle

$$x_0 = -\alpha_2 / (3\alpha_3).$$

Ein Nulldurchgang im Bereich des Kantenlementes liegt vor, falls $-0,5 \leq x_0 \leq 0,5$ gilt, wobei x in bezug auf das Kantenlement in Einheiten von Bildelementen gezählt wird.

Das Kantenlement $\gamma_{r;p}$ wird jetzt als Einer-Clique interpretiert, dem das Potential

$$U_C(\gamma) = \begin{cases} -c_p \gamma_{r;p} (dy(x)/dx|_{x=x_0})^2 & \text{für } -0,5 \leq x_0 \leq 0,5 \\ 0 & \text{sonst} \end{cases} \quad (379.23)$$

zugewiesen wird, wobei c_p eine Konstante bedeutet (Busch 1992, S.38). Das Kantenlement $\gamma_{r;p}$ erhält also, falls es den Wert $\gamma_{r;p} = 1$ annimmt, ein Potential, das umso kleiner ist, je größer die aus der Polynomannäherung berechnete Steigung ist. Die Likelihoodfunktion $p(\mathbf{y}|\boldsymbol{\gamma})$ folgt dann als Gibbs-Verteilung nach (373.10) durch Summation über sämtliche Einer-Cliquen, also über sämtliche Kantenlemente

$$p(\mathbf{y}|\boldsymbol{\gamma}) \propto \exp(-\sum_C U_C(\boldsymbol{\gamma})) \quad (379.24)$$

Bei Messungen in mehreren Frequenzbändern muß das vektorielle Zufallsfeld (379.18) eingeführt werden.

Stellt man die Kantenlemente als Grauwertstufen dar, ist die entsprechende Likelihoodfunktion bei (Chou und Brown 1990) zu finden.

c) Texturerkennung

Bei der Texturerkennung sind in einem digitalen Bild mit Hilfe der gemessenen Grauwerte Muster oder Texturen zu lokalisieren, um das Bild den Texturen entsprechend zu segmentieren. Neben dem in (379.1) definierten Zufallsfeld $Y(m,n)$ des Meßprozesses

mit der Realisierung y_{mn} ist dann noch das Zufallsfeld einzuführen, das die Zugehörigkeit eines Bildelementes zu einer Textur beschreibt. Entsprechend (376.1) wird definiert

$$E(m,n) = E(t_1) \quad \text{mit} \quad t_1 = [m,n]'$$

$$\text{und} \quad m \in \{0, 1, \dots, M\}, n \in \{0, 1, \dots, N\}, l \in \{0, 1, \dots, (M+1)(N+1)-1\}, \quad (379.25)$$

so daß $E(m,n)$ das Zufallsfeld der Zugehörigkeit zu einer Textur bezeichnet. Von $E(m,n)$ wird angenommen, daß es ein Markoff-Zufallsfeld repräsentiert. Ein Wert von $E(m,n)$ sei mit ε_{mn} bezeichnet, und die Werte für alle Bildelemente werden im Vektor ε zusammengefaßt, also

$$E(m,n) = \varepsilon_{mn} \quad \text{mit} \quad \varepsilon = (\varepsilon_{mn}) . \quad (379.26)$$

Das Zufallsfeld der Zugehörigkeit zu einer Textur wird durch diskrete Zufallsvariable $E(m,n)$ gebildet, denn bezeichnet T die Anzahl der Texturen eines Bildes, so soll

$$\varepsilon_{mn} = \varepsilon \quad \text{mit} \quad \varepsilon \in \{1, 2, \dots, T\} \quad (379.27)$$

gelten. Die Komponenten ε_{mn} des Vektors ε stellen die unbekannt Parameter der Texturerkennung dar.

Der Beobachtungsvektor y der Grauwerte mit der Realisierung y_{mn} des Zufallsfeldes $Y(m,n)$ der Messungen sei wie bei der Bildrestaurierung normalverteilt, so daß $Y(m,n)$ ein Markoff-Zufallsfeld repräsentiert. Bei der Texturerkennung ist zu beachten, daß der Erwartungswert der Beobachtung y_{mn} des Grauwertes eines Bildelementes und ihre Varianz von der Textur abhängen, zu der die Beobachtung gehört. Der Erwartungswert der Beobachtung wird durch die Beobachtungen in der Nachbarschaft des Bildelementes mittels unbekannter, für die jeweilige Textur gültiger Parameter dargestellt und die Varianz für jede Textur als konstant angenommen. Als Likelihoodfunktion eignet sich daher die bedingte Normalverteilung (376.22), die als Spezialfall des Auto-Modells abgeleitet wurde. Mit der in (376.18) definierten Menge Ω der Gitterpunkte und der in (376.19) eingeführten Menge N_p der Nachbarschaftspunkte für das Markoff-Zufallsfeld der Ordnung p ergibt sich in der üblichen Darstellung (375.4) die bedingte Normalverteilung der Messung y_r im Gitterpunkt $r = (m,n) \in \Omega$ unter der Bedingung, daß die Messungen ∂y_r in den Nachbarpunkten fest vorgegeben sind und daß mit $\varepsilon_r = \varepsilon$ im Bildelement $r \in \Omega$ die Textur ε vorliegt,

$$p(y_r | \partial y_r, \varepsilon_r = \varepsilon) \propto \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} [y_r - \mu_\varepsilon - \sum_{s \in N_p} \beta_{s,\varepsilon} (y_{r+s} - \mu_\varepsilon + y_{r-s} - \mu_\varepsilon)]^2\right\} , \quad (379.28)$$

worin σ_ε^2 die unbekannt Varianz der Messung y_r beim Vorliegen der Textur ε , μ_ε das unbekannt gemeinsame Mittel der Beobachtungen y_r für die Textur ε und $\beta_{s,\varepsilon}$ die unbekannt Parameter der Textur ε angeben. Um die gemeinsame Verteilung der Messungen y_r anzugeben, muß jeweils über die Einer- und Zweier-Cliquen der Bildelemente

summiert werden, die zu einer Textur gehören. Nimmt man an, daß die Menge der Bildpunkte, die zur Textur ε gehören, bekannt ist und bezeichnet man diese Menge mit Ω_ε , kann die gemeinsame Verteilung der Beobachtungen \mathbf{y}_ε für die Bildelemente, die zur Menge Ω_ε gehören, in der Form (376.25) ausgedrückt werden. Die gemeinsame Verteilung des gesamten Beobachtungsvektors \mathbf{y} bei gegebenem Vektor ε der Texturzugehörigkeit ergibt sich dann als Summe über die T Texturen ε zu

$$p(\mathbf{y}|\boldsymbol{\varepsilon}) \propto \exp\left\{-\sum_{\varepsilon=1}^T \frac{1}{2\sigma_\varepsilon^2}(\mathbf{y}_\varepsilon - \mathbf{1}\mu_\varepsilon)' \mathbf{B}_\varepsilon(\mathbf{y}_\varepsilon - \mathbf{1}\mu_\varepsilon)\right\} \quad (379.29)$$

mit $\mathbf{1} = [1, 1, \dots, 1]'$. Die Matrix \mathbf{B}_ε enthält die unbekannt Parameter $\beta_{s,\varepsilon}$.

In (379.28) wird der Erwartungswert der Beobachtung des Grauwerts eines Bildelementes für jede Textur durch die Beobachtungen in der Nachbarschaft des Bildelementes unter Einschaltung unbekannter Parameter ausgedrückt. Zur Darstellung des Zufallsfeldes $Y(m,n)$ der Messungen eignen sich daher auch zweidimensionale autoregressive Prozesse, bei denen die Zufallsvariablen des Prozesses in einem Punkt durch die Zufallsvariablen der Nachbarschaft multipliziert mit unbekannt Parametern erhalten werden. Wünscht man, daß der autoregressive Prozeß die Markoff-Eigenschaft besitzt, kann der in Kapitel 377 behandelte kausale autoregressive Prozeß gewählt werden, dessen Nachbarschaft des Bildelementes (m,n) sich jedoch auf die Bildelemente (i,j) mit $i \leq m$ und $j \leq n$ sowie $(i,j) \neq (m,n)$ beschränkt. Günstiger ist die Nachbarschaft eines nicht-kausalen autoregressiven Prozesses gestaltet, die ein Bildelement symmetrisch einschließt. Wählt man einen nichtkausalen zweidimensionalen autoregressiven Prozeß mit Markoff-Eigenschaft zur Darstellung des Zufallsfeldes $Y(m,n)$ der Messungen, ergibt sich (379.28) als bedingte Dichte einer Messung und (379.29) als Dichte sämtlicher Messungen. Dies folgt aus (378.33) und (378.51). Das Gauß-Zufallsfeld als Spezialfall des Auto-Modells und das Zufallsfeld des nichtkausalen autoregressiven Prozesses mit Markoff-Eigenschaft besitzen also identische bedingte und gemeinsame Verteilungen, so daß (379.28) und (379.29) für beide Zufallsfelder gültig sind. Wegen dieser günstigen Eigenschaften brauchen Verteilungen für autoregressive Prozesse im folgenden nicht mehr betrachtet zu werden.

Anwendungen der Normalverteilungen (379.28) und (379.29), die als Spezialfälle des Auto-Modells interpretiert werden, für die Texturerkennung und die Segmentierung digitaler Bilder befinden sich in (Cohen und Cooper 1987; Manjunath und Chellappa 1991; Manjunath et al. 1990). Anwendungen von (379.29), abgeleitet als Verteilung des nicht-kausalen autoregressiven Prozesses mit Markoff-Eigenschaft sind bei (Chellappa und Chatterjee 1985; Cohen und Fan 1992; Cohen et al. 1991 a,b) gegeben. Der kausale autoregressive Prozeß wird für die Bildsegmentierung bei (Bouman und Liu 1991) ange-

wendet, während bei (Mao und Jain 1992) der nichtkausale autoregressive Prozeß mit weißem Rauschen benutzt wird.

Beispiel 1: Es sollen die Beobachtungen der Grauwerte einer Textur generiert werden, die für $T = 1$ die gemeinsame Verteilung (379.29) besitzen. Für die Ordnung p des Markoff-Zufallsfeldes wird $p = 7$ gewählt, so daß sich aus (376.19) die Nachbarschaft

$$N_7 = \{s_1, s_2, \dots, s_{22}\}$$

ergibt. Folgende Parameter $\beta_{s, \varepsilon}$ in (379.28) bestimmen die Textur ε , wobei der Index ε fortgelassen und der Index s wie in (376.19) ausgedrückt wird,

$$\beta_{1,0} = \beta_{0,1} = 0,09 \quad , \quad \beta_{1,-1} = \beta_{1,1} = 0,08 \quad ,$$

$$\beta_{2,0} = \beta_{0,2} = \beta_{1,-2} = \beta_{1,2} = \beta_{2,-1} = \beta_{2,1} = \beta_{2,-2} = \beta_{2,2} = 0,06 \quad ,$$

$$\beta_{3,0} = \beta_{0,3} = -0,07 \quad ,$$

$$\beta_{1,-3} = \beta_{1,3} = \beta_{2,-3} = \beta_{2,3} = \beta_{3,-2} = \beta_{3,2} = \beta_{3,-1} = \beta_{3,1} = -0,06 \quad .$$

Mit diesen Parametern und $\mu_\varepsilon = 0$ sowie $\sigma_\varepsilon^2 = 1$ wird nun in jedem Bildelement eines Bildes mit 128×128 Elementen eine Zufallszahl mit der Normalverteilung (379.28) erzeugt, wobei als Startwerte für sämtliche Bildelemente der Wert Null angenommen wird und die Reihenfolge, in der die Bildelemente durchlaufen werden, zufällig ist, was mittels Gleichverteilung erreicht wird. Mit den so gewonnenen Zufallszahlen wird erneut für jedes Bildelement ein Zufallswert mit der Normalverteilung (379.28) erzeugt, wobei die Reihenfolge, in der die Bildelemente durchlaufen werden, wieder zufällig ist. Ein solcher Durchlauf bildet einen Iterationsschritt des sogenannten "Gibbs sampler" (Geman und Geman 1984), auf den im Kapitel 383 ausführlicher eingegangen wird. Iteriert man genügend häufig, besitzen die generierten Beobachtungen der Grauwerte die aus der bedingten Dichte (379.28) folgende gemeinsame Verteilung (379.29). Hier werden insgesamt vierzigmal in sämtlichen Bildelementen normalverteilte Zufallszahlen erzeugt, wobei darauf geachtet wurde, daß die Parameter $\beta_{s, \varepsilon}$ so gewählt waren, daß die Quadratsumme der Erwartungswerte der generierten Zufallszahlen nicht unbeschränkt anwächst, sondern sich auf einem gewissen Niveau einpendelt. Das Ergebnis der Texturgenerierung zeigt die Abbildung 379-5, in der die generierten Beobachtungen durch Grauwerte dargestellt sind. Deutlich zeigen sich Ringstrukturen, die aufgrund der gewählten Parameter zu erwarten sind. ▲

Möglicherweise können die Abbilder einer Textur in verschiedenen Maßstäben und unterschiedlichen Orientierungen in einem digitalen Bild erscheinen. Greift man eines der Abbilder heraus, in dem die Grauwerte durch die Indizes m und n des Gitters in (376.1)

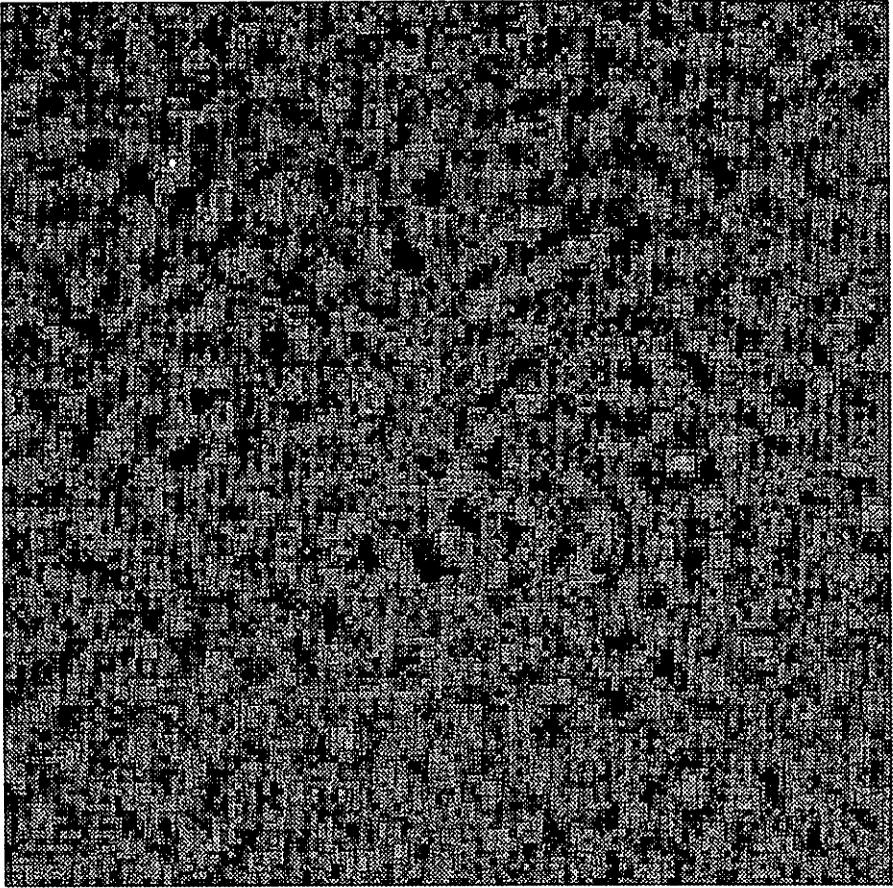


Abb. 379-5: Texturgenerierung

gekennzeichnet sind, und definiert man für diese Abbildung die Parameter $\beta_{s, \epsilon}$ der Textur, dann müssen die Indizes m und n der Grauwerte in den Maßstab und die Orientierung der übrigen Abbildungen transformiert werden. Mit der Drehmatrix \mathbf{R} (Koch 1987, S.50) und der Matrix \mathbf{M} für den Maßstab

$$\mathbf{R} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \quad \text{und} \quad \mathbf{M} = f\mathbf{I} \quad , \quad (379.30)$$

worin θ den Drehwinkel im Gegenuhrzeigersinn und f den Maßstabsfaktor bedeuten, ergeben sich die ganzzahligen Indizes k und l im transformierten System aus

$$\begin{bmatrix} k \\ l \end{bmatrix} = \mathbf{M} \mathbf{R} \begin{bmatrix} m \\ n \end{bmatrix} \quad . \quad (379.31)$$

In (379.28) ist dann mit $r = (m, n)$ und $p = (k, l)$

$$y_r = y_p \quad (379.32)$$

zu setzen, um die bedingte Verteilung der Messungen y der Grauwerte einer Textur zu erhalten, deren Abbilder in unterschiedlichen Maßstäben und Orientierungen auftreten. Rotationen und Maßstabsänderungen für nichtkausale autoregressive Prozesse mit Markoff-Eigenschaft behandeln Cohen et al. (1991 b).

Eine weitere Möglichkeit, die Verteilung der gemessenen Grauwerte y_r in Abhängigkeit von der Zugehörigkeit ε_r zu der Textur ε anzugeben, besteht mit dem autobinominalen Modell. Bezeichnet l die Anzahl der Grauwerte, die ein Bildelement $r \in \Omega$ annehmen kann, ergibt sich aus (375.12) die bedingte Dichte der Messung y_r des Grauwertes im Bildelement r unter der Bedingung, daß die Messungen ∂y_r in den Nachbarpunkten fest vorgegeben sind und daß mit $\varepsilon_r = \varepsilon$ die Textur ε vorliegt,

$$p(y_r | \partial y_r, \varepsilon_r = \varepsilon) = \binom{l}{y_r} p(\partial y_r)^{y_r} (1 - p(\partial y_r))^{l - y_r} \quad (379.33)$$

Die Wahrscheinlichkeit $p(\partial y_r)$ ist abhängig von den Grauwerten der Nachbarschaft und von der Textur, zu der die Grauwerte gehören. Man erhält anstelle von (375.13) entsprechend (376.21)

$$p(\partial y_r) = \exp Q / (1 + \exp Q) \quad (379.34)$$

mit

$$Q = \alpha_\varepsilon + \sum_{s \in N_p} \beta_{s, \varepsilon} (y_{r+s} + y_{r-s}) ,$$

worin α_ε und $\beta_{s, \varepsilon}$ die unbekanntenen Parameter der Textur ε bedeuten. Die Größe und Häufigkeit der Texturen steuern den Parameter α_ε . Der Parameter $\beta_{1, 0, \varepsilon}$ wird nach (376.19) durch die Grauwerte über und unter dem Bildelement r kontrolliert, während $\beta_{0, 1, \varepsilon}$ von den Grauwerten rechts und links von r beeinflusst wird. Für die übrigen Parameter $\beta_{s, \varepsilon}$ gilt Entsprechendes.

Die Likelihoodfunktion, die gemeinsame Verteilung des Beobachtungsvektors y bei gegebenem Vektor ε der Texturzugehörigkeit, erhält man durch Summation jeweils über die Cliques, die einer Textur angehören, und anschließender Summation über sämtliche Texturen. Durch Modifikation von (375.15) entsprechend (376.19) und (379.34) erhält man

$$p(y | \varepsilon) \propto \exp \left\{ - \sum_{\varepsilon=1}^T \left[\sum_{r \in \Omega_\varepsilon} y_r (\alpha_\varepsilon + \sum_{s \in N_p} \beta_{s, \varepsilon} y_{r+s}) + \sum_{r \in \Omega_\varepsilon} \ln \binom{l}{y_r} \right] \right\} \quad (379.35)$$

wobei die Auswirkungen unregelmäßig geformter Gebiete für die Mengen Ω_ε auf die

Summation nicht berücksichtigt werden, so daß darauf zu achten ist, daß über sämtliche Zweier-Cliquen summiert wird. Die Autobinomial-Verteilung für die Texturmodellierung ist bei (Cross und Jain 1983) angegeben. Eine Erweiterung zur Beseitigung der Abhängigkeit vom Histogramm des digitalen Bildes befindet sich bei (Acuna 1992).

Wie bereits zu Beginn des Kapitels erwähnt, beruht die Vorinformation bei der Texturerkennung darauf, daß Texturen nicht nur wenige Bildelemente umfassen, sondern sich über größere Bereiche des digitalen Bildes erstrecken. Diese Information läßt sich als Priori-Verteilung für den Vektor ε der unbekannt Parameter für die Texturzugehörigkeit einbringen, die aufgrund der Markoff-Eigenschaft des Zufallsfeldes (379.25) der Texturzugehörigkeit als Gibbs-Verteilung definiert wird. Als bedingte Dichte für ε_r mit $r \in \Omega$ wird das Auto-Modell gewählt, und zwar in einer Form, die etwas allgemeiner ist als (376.21),

$$p(\varepsilon_r | \partial \varepsilon_r) \propto \exp\{-[\alpha_\varepsilon + \sum_{s \in N_p} \beta_s (I(\varepsilon_r, \varepsilon_{r+s}) + I(\varepsilon_r, \varepsilon_{r-s}))]\} \quad (379.36)$$

mit der Indikatorfunktion

$$I(\varepsilon_r, \varepsilon_q) = \begin{cases} 1 & \text{falls } \varepsilon_r \neq \varepsilon_q \\ 0 & \text{sonst .} \end{cases}$$

Durch die Indikatorfunktion wird festgestellt, ob die Bildelemente in den Zweier-Cliquen der Nachbarschaft N_p des Bildelementes r zu unterschiedlichen Texturen gehören. Die Parameter β_s steuern mit $\beta_s > 0$ die Zugehörigkeit zu derselben Textur, denn mit großen Werten von β_s werden Cliquen aus unterschiedlichen Texturen hohe Energien und damit kleine Dichtewerte zugeordnet. Über die Parameter β_s lassen sich auch die Richtungen der Zugehörigkeit zu Texturen beeinflussen. Für eine mit (376.19) definierte Achter-Nachbarschaft sind beispielsweise kleine Werte für $\beta_{1,0}$ und $\beta_{0,1}$ zu wählen, wenn die Texturen in Blöcken mit überwiegend vertikalen und horizontalen Rändern angeordnet sind. Die Parameter α_ε mit $\varepsilon \in \{1, 2, \dots, T\}$ kontrollieren die Anzahl der Bildelemente, die den einzelnen Texturen zuzuordnen sind. Mit $\alpha_\varepsilon > 0$ bewirken große Werte für α_ε , daß wenige Bildelemente zur Textur ε gehören werden.

Die gemeinsame Verteilung der ε_r , also die Priori-Dichte für den Vektor ε , folgt mit (373.10) entsprechend (376.20) aus

$$p(\varepsilon) \propto \exp\{- \sum_{r \in \Omega} [\alpha_\varepsilon + \sum_{s \in N_p} \beta_s I(\varepsilon_r, \varepsilon_{r+s})]\} . \quad (379.37)$$

Die bedingte Dichte (379.36) für $\alpha_\varepsilon = 0$ und für eine Achter-Nachbarschaft wurde bei (Koch 1991) dazu benutzt, Texturen zu generieren. Dieselbe Verteilung fand mit $\beta_{1,0} = \beta_{0,1}$ und $\beta_{1,-1} = \beta_{1,1}$ bei (Bouman und Liu 1991) als Priori-Verteilung für die Texturer-

kennung Verwendung. Ähnliche Ansätze befinden sich bei (Geman und Geman 1984; Mester 1989, S.43; Therrien et al. 1986) und mit Berücksichtigung von Dreier-Cliquen bei (Derin und Cole 1986; Derin und Elliott 1987; Lakshmanan und Derin 1989). Die Dichte (379.37) für eine Achter-Nachbarschaft mit einer Indikatorfunktion, die die Identität der Texturen zählt, wird bei (Zhao et al. 1992) für Grauwerte formuliert und zur Bildrestaurierung angewendet.

Beispiel 2: Ein Zufallsfeld der Zugehörigkeit zu Texturen, das die Dichte (379.37) besitzt, wird durch den bereits im Beispiel 1 benutzten "Gibbs sampler" generiert. Mit $p=2$ in (376.19) wird die Achter-Nachbarschaft

$$N_2 = \{s_1, s_2, s_3, s_4\}$$

gewählt. Es seien vier Texturen gegeben, so daß $T = 4$ in (379.27) gilt. Die Anordnung dieser vier Texturen sei durch die Parameter

$$\begin{aligned} \alpha_1 &= 0,00, & \alpha_2 &= 0,01, & \alpha_3 &= 0,03, & \alpha_4 &= 0,04, \\ \beta_{1,0} &= 0,4, & \beta_{0,1} &= 0,2, & \beta_{1,-1} &= 1,4, & \beta_{1,1} &= 1,2 \end{aligned}$$

in (379.36) bestimmt. Generiert wird für ein Bild mit 128×128 Bildelementen.

Zunächst ordnet man zufällig mittels Gleichverteilung jedem Bildelement eine der vier Texturen zu. Anschließend wird für jedes Bildelement eine nach (379.36) verteilte Zufallszahl erzeugt, wobei die Reihenfolge, in der die Bildelemente durchlaufen werden, wieder zufällig ist, was durch eine Gleichverteilung erreicht wird. Die nach (379.36) verteilten Zufallszahlen werden durch die inverse Transformation aus einer Gleichverteilung generiert (Rubinstein 1981, S.39; Ripley 1987, S.71). Insgesamt werden sechzigmal die Zufallszahlen in den 128×128 Bildelementen erzeugt, dann hat sich der Exponent in (379.36), der nach (373.4) die Energie der Gibbs-Verteilung angibt, auf einem gewissen Niveau stabilisiert, und die generierten Zugehörigkeiten zu den vier Texturen besitzen die Dichte (379.37).

Das Ergebnis der Generierung zeigt Abbildung 379-6, in der die vier Texturen in abgestuften Grautönen dargestellt sind, wobei die Textur 1 weiß und die Textur 4 schwarz ist. Deutlich ist zu erkennen, daß die einzelnen Texturen sich in größeren Bereichen angeordnet haben. Wegen $\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4$ überwiegt der Anteil der weiß gefärbten Textur 1 im Bild, während die schwarz gefärbte Textur 4 am schwächsten vertreten ist. Mit $\beta_{0,1} = 0,2$ im Vergleich zu $\beta_{1,-1} = 1,4$ und $\beta_{1,1} = 1,2$ herrscht die Anordnung der Texturen in horizontaler Richtung vor. Wegen $\beta_{1,0} = 0,4$ ist die senkrechte Ausrichtung der Texturen ebenfalls ausgeprägt. Da beide Parameter gleichzeitig wirken, treten auch eine Reihe von diagonal ausgerichteten Grenzen zwischen den Texturen auf. ▲

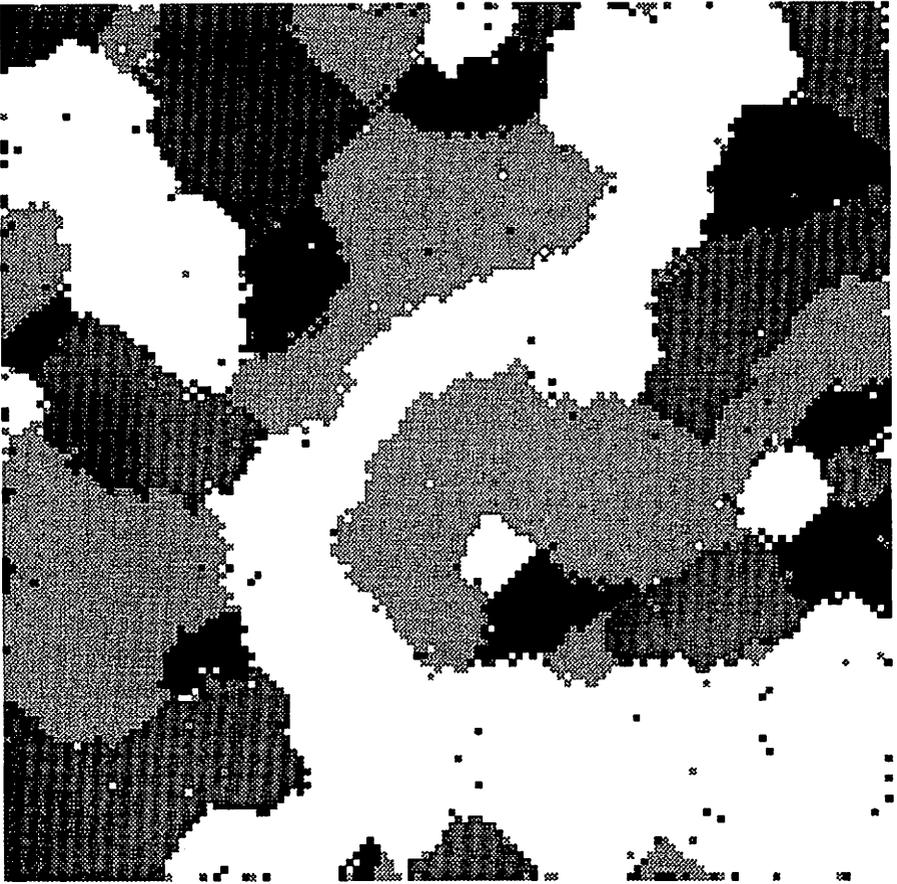


Abb. 379-6: Generierung einer Anordnung von vier Texturen

38 Schätzungen für zweidimensionale Zufallsfelder

381 Schätzungen von Momentfunktionen

Im folgenden sollen Schätzungen der Momentfunktionen und der Spektren nur von diskreten Zufallsfeldern betrachtet werden, da bei praktischen Anwendungen in der Regel diese Zufallsfelder vorliegen. Die Schätzungen für die zweidimensionalen Zufallsfelder werden durch Verallgemeinerung der für eindimensionale Prozesse abgeleiteten Ergebnisse gewonnen.

Das zweidimensionale diskrete Zufallsfeld $X(m,n)$ sei auf einem regelmäßigen Gitter mit

$$X(m,n) \quad \text{für} \quad m \in \{-M, -M+1, \dots, M\} \quad \text{und} \quad n \in \{-N, -N+1, \dots, N\} \quad (381.1)$$

gegeben, worin im Vergleich zu (376.1) eine Nullpunktverschiebung vorgenommen wird, um die Analogie zur Zählung der diskreten Prozesse zu gewinnen. Es wird ein homogenes Zufallsfeld $X(m,n)$ vorausgesetzt, das bezüglich seines nach (362.4) konstanten Erwartungswertes ergodisch sei. Der zur Bildung des Erwartungswertes notwendige Mittelwert über das Ensemble läßt sich dann durch das Mittel über eine Realisierung von $X(m,n)$ ersetzen. Die Schätzung $\hat{\mu}$ des Erwartungswertes μ folgt daher (332.2) entsprechend wie bei eindimensionalen Prozessen mit

$$\hat{\mu} = \frac{1}{(2M+1)(2N+1)} \sum_{m=-M}^M \sum_{n=-N}^N X(m,n) \quad (381.2)$$

Die Autokovarianzfunktion eines homogenen Zufallsfeldes ist nach (362.5) nur von einem Vektor abhängig, der im Falle von ebenen Zufallsfeldern zwei Komponenten besitzt. Die Autokovarianzfunktion eines diskreten zweidimensionalen Zufallsfeldes wird daher mit $C(k, l)$ bezeichnet. Setzt man wieder Ergodizität bezüglich der Autokovarianzfunktion voraus, läßt sich die Schätzung von $C(k, l)$ aus einer Realisierung von $X(m,n)$ bestimmen. In Analogie zur nicht erwartungstreuen Schätzung (333.13) für eindimensionale Prozesse folgt die nicht erwartungstreue Schätzung $\check{C}(k, l)$ der Autokovarianzfunktion $C(k, l)$ von $X(m,n)$ mit $E(X(m,n)) = 0$ aus

$$\check{C}(k, l) = \frac{1}{(2M+1)(2N+1)} \sum_{m=-M}^{M-k} \sum_{n=-N}^{N-l} X(m+k, n+l)X(m, n) = \frac{1}{(2M+1)} \sum_{m=-M}^{M-k} \check{C}_{m+k}(l) \quad (381.3)$$

für $k \in \{0, 1, \dots, 2M\}$, $l \in \{0, 1, \dots, 2N\}$ mit

$$\tilde{C}_{m+k}(1) = \frac{1}{(2N+1)} \sum_{n=-N}^{N-1} X(m+k, n+1)X(m, n) . \quad (381.4)$$

Für feste Werte m und k entspricht (381.4) der nicht erwartungstreuen Schätzung (333.13), so daß $\tilde{C}_{m+k}(1) = \tilde{C}_{m+k}(-1)$ gilt und weiter mit (362.6)

$$\tilde{C}(k, 1) = \tilde{C}(k, -1) = \tilde{C}(-k, 1) = \tilde{C}(-k, -1) \quad (381.5)$$

folgt.

Ist das zweidimensionale diskrete Zufallsfeld $X(m, n)$ nicht nur homogen, sondern auch isotrop, gilt nach (362.7) für die Autokovarianzfunktion

$$C(k, l) = C(p) \quad \text{mit} \quad p = (k^2 + l^2)^{1/2} . \quad (381.6)$$

Wird für die Schätzung der Autokovarianzfunktion wieder Ergodizität vorausgesetzt, ist über alle Wertepaare $X(m, n)$ und $X(q, r)$ zu summieren, die den Abstand p besitzen. Da das Zufallsfeld auf einem Gitter definiert ist, kann diese Bedingung nicht exakt, sondern nur näherungsweise erfüllt werden.

382 Schätzungen des Spektrums

Mit der Fourier-Transformation der nicht erwartungstreuen Schätzung $\tilde{C}(k, l)$ aus (381.3) bis (381.5) der Autokovarianzfunktion erhält man mit (363.8) entsprechend (341.7) das zweidimensionale Periodogramm $\tilde{S}(\Omega_1, \Omega_2)$ zu

$$\tilde{S}(\Omega_1, \Omega_2) = \sum_{k=-2M}^{2M} \sum_{l=-2N}^{2N} \tilde{C}(k, l) e^{-j(k\Omega_1 + l\Omega_2)} \quad (382.1)$$

mit

$$\Omega_1 = \frac{2m\pi}{4M+1} \quad \text{für} \quad m \in \{-2M, -2M+1, \dots, 2M\}$$

und

$$\Omega_2 = \frac{2n\pi}{4N+1} \quad \text{für} \quad n \in \{-2N, -2N+1, \dots, 2N\}$$

aus (342.23).

Für ein homogenes und isotropes zweidimensionales Zufallsfeld erhält man mit der Schätzung von $C(p)$ unter Berücksichtigung von (381.6) ein eindimensionales Periodogramm aus (341.7), so daß sich dieser Fall nicht von den Schätzungen für die eindimensionalen Prozesse unterscheidet.

Das Periodogramm $\tilde{S}(\Omega_1, \Omega_2)$ aus (382.1) stellt keine geeignete Schätzung des Spektrums

$S(\Omega_1, \Omega_2)$ eines homogenen zweidimensionalen diskreten Zufallsfeldes $X(m, n)$ dar. Der Grund liegt in der großen Varianz für $\mathfrak{S}(\Omega_1, \Omega_2)$, die dadurch verursacht wird, daß in (382.1) über sämtliche Schätzungen $\check{C}(k, l)$ der Autokovarianzfunktion zu addieren ist, so daß sich auch die Varianzen der Schätzungen addieren, falls Unabhängigkeit der Schätzungen vorausgesetzt wird (Koch 1987, S.118). Wie im Kapitel 342 für eindimensionale Prozesse ausführlich diskutiert wird, ist das Periodogramm zu glätten, um seine Varianz zu reduzieren.

Eine einfache Methode, eine geglättete Schätzung zu erhalten, besteht darin, das Gitter der Größe $(2M+1)(2N+1)$, auf dem nach (381.1) das homogene Zufallsfeld $X(m, n)$ definiert ist, in kleinere Rechtecke zu zerlegen, in denen die Periodogramme geschätzt werden. Sind diese Schätzungen voneinander unabhängig, ergibt sich wie in (342.27) für einen eindimensionalen Prozeß die geglättete Schätzung aus dem Mittel sämtlicher Periodogramme. Dieses Verfahren setzt voraus, daß für die Autokovarianzfunktion $C(k, l)$ entsprechend (342.24) gilt

$$C(k, l) \approx 0 \quad \text{für } k > 2Q \quad \text{und/oder } l > 2R \quad \text{mit } Q \ll M, R \ll N. \quad (382.2)$$

Das Gitter der Größe $(2M+1)(2N+1)$ wird in P Rechtecke der Größe $(4Q+1)(4R+1)$ aufgeteilt, also

$$(2M+1)(2N+1) = P(4Q+1)(4R+1). \quad (382.3)$$

In diesen P Rechtecken wird nun jeweils nach (382.1) das Periodogramm $\mathfrak{S}^{(i)}(\Omega_1, \Omega_2)$ des Rechteckes i mit $\check{C}^{(i)}(k, l)$ aus (381.3) bis (381.5) berechnet

$$\mathfrak{S}^{(i)}(\Omega_1, \Omega_2) = \sum_{k=-2Q}^{2Q} \sum_{l=-2R}^{2R} \check{C}^{(i)}(k, l) e^{-j(k\Omega_1 + l\Omega_2)} \quad \text{für } i \in \{1, 2, \dots, P\}, \quad (382.4)$$

falls in jedem Rechteck die Werte von $X(m, n)$ jeweils von $-2Q$ bis $2Q$ und von $-2R$ bis $2R$ gezählt werden. Die geglättete Schätzung $\mathfrak{S}_w(\Omega_1, \Omega_2)$ des Spektrums ergibt sich dann durch Mittelung über die P Periodogramme

$$\mathfrak{S}_w(\Omega_1, \Omega_2) = \frac{1}{P} \sum_{i=1}^P \mathfrak{S}^{(i)}(\Omega_1, \Omega_2) \quad (382.5)$$

mit

$$\Omega_1 = \frac{2m\pi}{4M+1} \quad \text{für } m \in \{-2M, -2M+1, \dots, 2M\}$$

und

$$\Omega_2 = \frac{2n\pi}{4N+1} \quad \text{für } n \in \{-2N, -2N+1, \dots, 2N\}.$$

Dieser geglätteten Schätzung entspricht wie bei den eindimensionalen Prozessen eine

Gewichtung der Schätzung $\check{C}(k, l)$ der Autokovarianzfunktion durch die zweidimensionale Fensterfunktion $w(k, l)$, so wie nach (342.13) die geglättete Schätzung $\check{S}_w(\Omega_1, \Omega_2)$ des Spektrums $S(\Omega_1, \Omega_2)$ erhalten wird,

$$\check{S}_w(\Omega_1, \Omega_2) = \sum_{k=-2M}^{2M} \sum_{l=-2N}^{2N} \check{C}(k, l) w(k, l) e^{-j(k\Omega_1 + l\Omega_2)} \quad (382.6)$$

Um hieraus eine (382.5) entsprechende Schätzung zu erhalten, ist aufgrund von (342.28) die Fensterfunktion $w(k, l)$ aus dem Dreieckfenster (342.19) zu entwickeln. Setzt man nach (246.27)

$$w(k, l) = w(k)w(l) \quad \text{für } k \in \{-2K, -2K+1, \dots, 2K\}, \quad l \in \{-2L, -2L+1, \dots, 2L\}, \quad (382.7)$$

wobei $w(k)$ und $w(l)$ eindimensionale Fensterfunktionen bedeuten, ergibt sich eine rechteckige Fensterfunktion. Wünscht man eine rotationssymmetrische Fensterfunktion, kann man nach (246.28)

$$w(k, l) = w((k^2 + l^2)^{1/2}) \quad (382.8)$$

setzen, wobei $w((k^2 + l^2)^{1/2})$ wieder eine eindimensionale Fensterfunktion bezeichnet.

Anstatt die geglättete Schätzung $\check{S}_w(\Omega_1, \Omega_2)$ nach (382.6) durch die Gewichtung der Schätzung $\check{C}(k, l)$ der Autokovarianzfunktion zu berechnen, kann $\check{S}_w(\Omega_1, \Omega_2)$ auch entsprechend (342.23) durch die diskrete Faltung des Periodogramms mit dem zweidimensionalen Spektralfenster erhalten werden. Es ergibt sich

$$\check{S}_w(\Omega_1, \Omega_2) = \frac{1}{(4M+1)(4N+1)} \sum_{m=-2M}^{2M} \sum_{n=-2N}^{2N} \check{S}(\lambda_1, \lambda_2) W(\Omega_1 - \lambda_1, \Omega_2 - \lambda_2) \quad (382.9)$$

mit dem Spektralfenster

$$W(\Omega_1 - \lambda_1, \Omega_2 - \lambda_2) = \sum_{k=-2M}^{2M} \sum_{l=-2N}^{2N} w(k, l) e^{-j(k(\Omega_1 - \lambda_1) + l(\Omega_2 - \lambda_2))}$$

aus (246.20) und mit

$$\Omega_1 = \frac{2p\pi}{4M+1}, \quad \lambda_1 = \frac{2m\pi}{4M+1} \quad \text{für } p, m \in \{-2M, -2M+1, \dots, 2M\}$$

$$\Omega_2 = \frac{2q\pi}{4N+1}, \quad \lambda_2 = \frac{2n\pi}{4N+1} \quad \text{für } q, n \in \{-2N, -2N+1, \dots, 2N\}.$$

383 Parameterschätzungen für Markoff-Zufallsfelder der digitalen Bildverarbeitung

Die Methode der kleinsten Quadrate, die Maximum-Likelihood-Methode und die beste Schätzung zur Bestimmung unbekannter Parameter wurde in den vorangegangenen Kapiteln bereits mehrfach angewendet. Für die Parameterschätzung bei Markoff-Zufallsfeldern in der digitalen Bildverarbeitung benötigt man außerdem die Bayes-Schätzung, da bei der Bildanalyse Vorinformationen vorhanden sind, die für die Schätzung berücksichtigt werden sollten, worauf bereits im Kapitel 379 hingewiesen wurde. Die Bayes-Schätzung wird daher im folgenden kurz vorgestellt.

Die Likelihoodfunktion, das heißt die Wahrscheinlichkeitsdichte der Beobachtungen als Funktion der unbekannt Parameter, sei mit $p(\mathbf{y}|\boldsymbol{\beta})$ bezeichnet, wobei \mathbf{y} den Vektor der Beobachtungen und $\boldsymbol{\beta}$ den Vektor der unbekannt Parameter bedeutet. Die Priori-Dichte des Vektors $\boldsymbol{\beta}$ der unbekannt Parameter sei $p(\boldsymbol{\beta})$, dann ergibt sich die Posteriori-Dichte $p(\boldsymbol{\beta}|\mathbf{y})$ für $\boldsymbol{\beta}$ unter der Bedingung, daß der Beobachtungsvektor \mathbf{y} vorliegt, aus dem *Bayes-Theorem* zu (Koch 1990, S.4)

$$p(\boldsymbol{\beta}|\mathbf{y}) \propto p(\boldsymbol{\beta})p(\mathbf{y}|\boldsymbol{\beta}) . \quad (383.1)$$

Führt man in das Bayes-Theorem die Normierungskonstante c ein, erhält man

$$p(\boldsymbol{\beta}|\mathbf{y}) = \frac{1}{c} p(\boldsymbol{\beta})p(\mathbf{y}|\boldsymbol{\beta}) \quad (383.2)$$

mit

$$c = p(\mathbf{y}) = \int p(\boldsymbol{\beta})p(\mathbf{y}|\boldsymbol{\beta})d\boldsymbol{\beta} ,$$

wobei über den Parameterraum, den $\boldsymbol{\beta}$ aufspannt, zu integrieren ist. Mit Hilfe der Posteriori-Dichte $p(\boldsymbol{\beta}|\mathbf{y})$ lassen sich für die unbekannt Parameter $\boldsymbol{\beta}$ Konfidenzregionen oder Hypothesentests durchführen. Führt man die quadratische Verlustfunktion ein, erhält man als *Bayes-Schätzung* $\hat{\boldsymbol{\beta}}$ von $\boldsymbol{\beta}$ den Erwartungswert (Koch 1990, S.34)

$$\hat{\boldsymbol{\beta}} = \int \boldsymbol{\beta} p(\boldsymbol{\beta}|\mathbf{y})d\boldsymbol{\beta} , \quad (383.3)$$

wobei wieder über den Parameterraum von $\boldsymbol{\beta}$ zu integrieren ist. Für die numerische Lösung eignet sich die Monte-Carlo-Integration (Koch 1990, S.54).

Man kann aber auch den Vektor $\boldsymbol{\beta}$ der unbekannt Parameter durch $\hat{\boldsymbol{\beta}}$ derart schätzen, daß

$$\hat{\boldsymbol{\beta}} = \sup_{\boldsymbol{\beta}} p(\boldsymbol{\beta}|\mathbf{y}) \quad (383.4)$$

gilt. Dies ist die *Maximum-Posteriori-Schätzung*, auch *MAP-Schätzung* genannt (Koch 1990, S.36). Diese Schätzung eignet sich für die Klassifizierung, bei der die unbekannt Parameter mit den in (379.25) definierten Zugehörigkeiten zu Texturen identifiziert werden. Weist man nämlich Fehlklassifikationen identische Verluste und korrekten Klassifikationen keine Verluste zu, ergibt sich als Klassifizierungsregel die MAP-Schätzung der unbekannt Zugehörigkeit zu einer Textur (Koch 1990, S.136).

Ersetzt man im Bayes-Theorem (383.1) die Priori-Dichte $p(\boldsymbol{\beta})$ durch eine konstante, uneigentliche Dichte (Koch 1990, S.9), ergibt sich anstelle der MAP-Schätzung (383.4) die *Maximum-Likelihood-Methode* (Koch 1987, S.188)

$$\hat{\boldsymbol{\beta}} = \sup_{\boldsymbol{\beta}} p(\mathbf{y}|\boldsymbol{\beta}) , \quad (383.5)$$

mit der die Schätzwerte $\hat{\boldsymbol{\beta}}$ derart bestimmt werden, daß sie für einen vorliegenden Vektor \mathbf{y} von Beobachtungen maximale Werte für die Likelihoodfunktion $p(\mathbf{y}|\boldsymbol{\beta})$ liefern. Die Maximum-Likelihood-Methode führt in einem linearen Modell bei normalverteilten Beobachtungen auf Schätzwerte, die mit denen der Methode der kleinsten Quadrate und der besten Schätzung identisch sind (Koch 1987, S.188).

Beispiel 1 (Bildrestaurierung): Als Beispiel für eine Posteriori-Verteilung in der digitalen Bildanalyse soll zunächst die einfachste angegeben werden, nämlich die für die Bildrestaurierung ohne Kantenelemente. Mit der Likelihoodfunktion (379.6) und der Priori-Dichte (379.9) erhält man mit dem Bayes-Theorem (383.1) die Posteriori-Dichte $p(\boldsymbol{\beta}|\mathbf{y})$ der Verteilung für den Vektor $\boldsymbol{\beta}$ der unbekannt Grauwerte der Bildelemente zu

$$p(\boldsymbol{\beta}|\mathbf{y}) \propto \exp\left\{-\sum_{r \in \Omega} \sum_{s \in N_p} [c_{\beta}(\beta_r - \beta_{r+s})^2] + (y_r - \beta_r)^2 / \sigma_r^2\right\} . \quad (383.6)$$

Die zugehörige bedingte Posteriori-Dichte $p(\beta_r | \partial\beta_r, \mathbf{y})$ für den Grauwert β_r des Bildelementes r ergibt sich mit (379.8) zu

$$p(\beta_r | \partial\beta_r, \mathbf{y}) \propto \exp\left\{-\sum_{s \in N_p} c_{\beta}[(\beta_r - \beta_{r+s})^2 + (\beta_r - \beta_{r-s})^2] - (y_r - \beta_r)^2 / 2\sigma_r^2\right\} . \quad (383.7)$$

Wie erwähnt, messen die ersten Terme in (383.6) oder (383.7) die Rauigkeit des Bildes in der Nachbarschaft des Bildelementes $r \in \Omega$, während der letzte Term die Abweichung der Beobachtungen y_r von den unbekannt Grauwerten β_r im Bildelement r angibt. Zur Schätzung des Vektors $\boldsymbol{\beta}$ der unbekannt Grauwerte ist nun die Posteriori-Dichte $p(\boldsymbol{\beta}|\mathbf{y})$ nach (383.4) zu maximieren oder mit $p(\boldsymbol{\beta}|\mathbf{y})$ ist nach (383.3) der Erwartungswert zu bilden. ▲

Die Normierungskonstante einer Posteriori-Dichte der digitalen Bildverarbeitung, beispielsweise (383.6), ergibt sich nach (383.2) als Integral über sämtliche Werte des unbe-

kannten Zufallsvektors β oder, falls die unbekannt Parameter β als diskrete Zufallsvariable aufgefaßt werden, durch Summation über sämtliche Werte. Bei Grauwerten sind dies üblicherweise die Werte von 0 bis 255, so daß $256^{(64 \times 64)}$ Summanden für ein Gitter mit 64×64 Gitterpunkten vorliegen. Analytische Verfahren zur Parameterschätzung nach (383.3) oder (383.4) werden daher nur in Sonderfällen anwendbar und direkte Suchverfahren zur Bestimmung des Maximums (383.4) kaum durchführbar sein.

Wie bereits im Kapitel 379 erläutert, besitzen die Zufallsfelder der digitalen Bildverarbeitung die Markoff-Eigenschaft, so daß die bedingte Posteriori-Dichte der Zufallsvariablen eines Bildelementes, beispielsweise (383.7), nur von der Umgebung dieses Bildelementes abhängen. Die Parameter lassen sich daher mit der sogenannten *Codierungsmethode* (im Englischen "coding methods") (Besag 1974, 1986; Chellappa, 1985; Cressie 1991, S.463) schätzen. Hierbei werden die Punkte eines Gitters derart codiert, daß die bedingten Verteilungen für die codierten Punkte voneinander unabhängig werden. Bei einem Markoff-Zufallsfeld erster Ordnung bedeutet dies, daß die bedingte Verteilung nur in jedem zweiten Gitterpunkt aufgestellt werden darf, die in Abbildung 383-1 mit einem

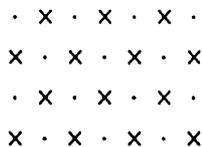


Abb. 383-1: Codierungsschema für ein Markoff-Zufallsfeld erster Ordnung

Kreuz versehen sind. Die Posteriori-Dichte $p(\beta|\mathbf{y})$ ergibt sich dann aus dem Produkt der bedingten Dichten $p(\beta_p|\partial\beta_p, \mathbf{y})$ der K Parameter β_p der codierten Gitterpunkte p zu

$$p(\beta|\mathbf{y}) = \prod_K p(\beta_p|\partial\beta_p, \mathbf{y}) \quad (383.8)$$

Die Posteriori-Dichte $p(\beta|\mathbf{y})$ kann nun dazu benutzt werden, nach (383.3) oder (383.4) die unbekannt Parameter β zu schätzen. Der Vorteil dieses Verfahrens besteht darin, daß die zu (383.8) gehörende Normierungskonstante eine wesentlich einfachere Form besitzt als beispielsweise die zu (383.6) gehörende Normierungskonstante. Der Nachteil besteht darin, daß Schätzungen einmal aus den mit einem Kreuz in Abbildung 383-1 codierten Gitterpunkten folgen, zum anderen aus den mit einem Punkt versehenen Bildelementen, die dann zu mitteln sind.

Diesen Nachteil vermeidet man beim *Pseudolikelihoodverfahren* (Besag 1975; Cressie 1991, S.461), bei dem näherungsweise angenommen wird, daß die gemeinsame Posteriori-Dichte $p(\beta|\mathbf{y})$ sich aus dem Produkt der bedingten Dichten sämtlicher Gitterpunkte

$r \in \Omega$ ergibt

$$p(\boldsymbol{\beta}|\mathbf{y}) = \prod_{r \in \Omega} p(\beta_r | \partial\beta_r, \mathbf{y}) . \quad (383.9)$$

Mit dieser Posteriori-Dichte $p(\boldsymbol{\beta}|\mathbf{y})$ erfolgt dann die Parameterschätzung nach (383.3) oder (383.4).

Für die Anwendung des Pseudolikelihoodverfahrens eignen sich iterative Verfahren, sogenannte *deterministische Verfahren*, die dadurch zu begründen sind, daß wegen (383.9) die Schätzung des Parameters β_r mit Hilfe der bedingten Dichten $p(\beta_r | \partial\beta_r, \mathbf{y})$ erfolgen kann, falls die Werte der Parameter in der Umgebung des Gitterpunktes r bekannt sind. Die unbekanntenen Werte werden deshalb durch ihre Schätzwerte ersetzt, so daß man beispielsweise für die bedingte Dichte (383.7) erhält

$$p(\beta_r | \partial\beta_r, \mathbf{y}) \propto \exp\left\{ - \sum_{s \in N_p} c_{\beta} [(\beta_r - \hat{\beta}_{r+s})^2 + (\beta_r - \hat{\beta}_{r-s})^2] - (y_r - \beta_r)^2 / 2\sigma_r^2 \right\} , \quad (383.10)$$

wobei die Schätzwerte mit einem Dach versehen sind. Die Schätzung für β_r ergibt sich dann aus (383.3) zu

$$\hat{\beta}_r = \int \beta_r p(\beta_r | \partial\beta_r, \mathbf{y}) d\beta_r \quad (383.11)$$

oder aus (383.4) zu

$$\bar{\beta}_r = \sup_{\beta_r} p(\beta_r | \partial\beta_r, \mathbf{y}) . \quad (383.12)$$

Ausgehend von Näherungswerten für die Schätzwerte wird iterativ vorgegangen. In jedem Iterationsschritt berechnet man die unbekanntenen Parameter sämtlicher Bildelemente nach (383.11) oder (383.12), wobei die Schätzwerte unmittelbar die Schätzwerte der vorangegangenen Iteration ersetzen. Die Reihenfolge, unter der die Bildelemente aufgesucht werden, sollte zufällig sein, damit systematische Effekte vermieden werden. Es wird solange iteriert, bis die Schätzwerte sich nicht mehr ändern. Die iterative Parameterschätzung nach (383.11) bezeichnet man auch als ICA-Verfahren (im Englischen "iterated conditional averages") (Johnson et al. 1991) und die iterative Parameterschätzung nach (383.12) als ICM-Verfahren (im Englischen "iterated conditional modes") (Besag 1986). Besonders günstig ist es, mit einer Normalverteilung wie in (383.7) zu arbeiten, da dann der Erwartungswert (383.11) bequem analytisch anzugeben ist (Koch 1990, S.161; Busch und Koch 1990) und da außerdem $\hat{\beta}_r = \bar{\beta}_r$ gilt.

Die iterative Parameterschätzung nach (383.11) oder (383.12) läßt sich auch dadurch interpretieren, daß sie mit Hilfe der Randdichte $p(\beta_r | \mathbf{y})$ der Posteriori-Dichte $p(\boldsymbol{\beta}|\mathbf{y})$ erfolgt. Diese Randverteilung verschafft man sich näherungsweise dadurch, daß für die unbekanntenen Parameter in der Umgebung des Gitterpunktes r die Schätzwerte eingesetzt

werden (Koch 1990, S.58), wie beispielhaft in (383.10) angegeben.

Die deterministischen Verfahren zur Parameterschätzung besitzen den Vorteil, daß sie kurze Rechenzeiten benötigen. Sie erlauben allerdings, da sie auf dem Pseudolikelihood-Verfahren (383.9) basieren, nur eine genäherte Schätzung der Parameter. Diesen Nachteil vermeiden die *stochastischen Verfahren*. Zu ihnen zählt die *Methode der simulierten Abkühlung* (im Englischen "simulated annealing"), die eine MAP-Schätzung nach (383.4) vornimmt, wobei jedoch der Rechenaufwand sehr viel höher als bei den ICA- oder ICM-Verfahren ist. Die Methode der simulierten Abkühlung wurde von Kirkpatrick et al. (1983) aus dem Metropolis-Algorithmus (Metropolis et al. 1953) entwickelt, eine Übersicht geben Aarts und van Laarhoven (1987). Dieses Verfahren simuliert einen Prozeß für die Vergütung von Metallen, bei dem das Metall zunächst erhitzt und dann langsam auf einen energiearmen, harten Zustand abgekühlt wird. In der Darstellung (373.7) der Gibbs-Verteilung $p(\boldsymbol{\beta}|\mathbf{y})$ wird, beginnend mit einer hohen Temperatur T_0 , die Temperatur in jedem Iterationsschritt verringert. Mit hohen Werten von T_0 erhält man eine flach verlaufende Dichtefunktion, während kleine Werte von T_0 die Maxima ausprägen. Mit Näherungswerten für die unbekannt Parameter β_r der Bildelemente r beginnen die Iterationen. In jeder Iteration wird dann für jedes Bildelement ein Zufallswert aus einer Gleichverteilung für den Parameter β_r erzeugt und der neue Wert für β_r akzeptiert, falls die sich ergebende Energiedifferenz ΔU in der Gibbs-Verteilung negativ wird. Falls sie positiv ist, wird der neue Wert mit einer Wahrscheinlichkeit von $\exp(-\Delta U/T_0)$ angenommen. Wenn also $u < \exp(-\Delta U/T_0)$ gilt, wobei u den Wert einer im Intervall $[0, 1]$ gleichverteilten Zufallszahl bedeutet, wird der neue Wert für β_r akzeptiert.

In der von Geman und Geman (1984) vorgeschlagenen Variante des Verfahrens der simulierten Abkühlung wird in jedem Iterationsschritt für jedes Bildelement eine Zufallszahl für β_r aus der zu $p(\boldsymbol{\beta}|\mathbf{y})$ gehörenden bedingten Dichte $p(\beta_r|\partial\beta_r, \mathbf{y})$ generiert, die beispielsweise durch (383.10) gegeben ist. Außerdem wird für jeden Iterationsschritt die Temperatur abgesenkt. Der generierte Wert ersetzt den Wert der vorangegangenen Iteration, und es wird so lange iteriert, bis die Schätzwerte konvergieren und durch die Absenkung der Temperatur gegen den Wert Null eine minimale Energie und damit ein maximaler Dichtewert $p(\boldsymbol{\beta}|\mathbf{y})$ erhalten wird. Bezeichnet t den jeweiligen Iterationsschritt, so wird die Absenkung der Temperatur $T_0(t)$ gesteuert durch

$$T_0(t) = c_t / \ln(1+t) , \quad (383.13)$$

worin c_t eine Konstante bedeutet. Entspricht die bedingte Dichte $p(\beta_r|\partial\beta_r, \mathbf{y})$ wie in (383.10) einer Normalverteilung, existieren zu ihrer Generierung effiziente Verfahren (Marsaglia und Bray 1964). Für die Generierung von Zufallszahlen anderer Verteilungen existieren weitere Methoden (Rubinstein 1981, S.38). Für diskrete Verteilungen bietet

sich das Verfahren der inversen Transformation an (Ripley 1987, S.71).

Die Variante von Geman und Geman (1984) der simulierten Abkühlung beruht auf einer Verbindung des sogenannten "Gibbs sampler" mit dem Abkühlungsprozeß. Beim "Gibbs sampler" wird bei konstanter Temperatur T_0 in jedem Iterationsschritt für jedes Bildelement ein Zufallswert für β_r mit der bedingten Dichte $p(\beta_r | \partial\beta_r, \mathbf{y})$ generiert. Iteriert man genügend häufig, erhält man Zufallszahlen mit der gemeinsamen Dichte $p(\boldsymbol{\beta} | \mathbf{y})$. Die einzelnen Zufallszahlen besitzen die Randdichten für β_r (Gelfand et al. 1990), so daß mit dem "Gibbs sampler" nicht nur die MAP-Schätzung, sondern auch die Bayes-Schätzung (383.4) berechnet werden kann.

Die beschriebenen Methoden zur Parameterschätzung eignen sich nicht nur für die in Beispiel 1 behandelte Schätzung von Grauwerten β_r , auch die in (379.10) definierten Kantenelemente lassen sich sowohl zusammen mit den Grauwerten bei einer Bildrestaurierung als auch allein für die Kantenextraktion berechnen. Ebenfalls kann die Zugehörigkeit (379.26) zu Texturen geschätzt werden. Es braucht nur für das jeweilige Problem mit dem Bayes-Theorem (383.1) die Posteriori-Verteilung aus der Priori-Verteilung und der Likelihoodfunktion aufgestellt zu werden.

Beispiel 2 (Texturerkennung): Abschließend soll noch das Beispiel der Bildsegmentierung aufgrund von Texturen behandelt werden. Außer der in (379.26) definierten Zugehörigkeit ε_r eines Bildelementes r zu einer Textur treten noch zusätzliche unbekannte Parameter auf, die den einzelnen Texturen zuzuordnen sind. Um das Problem der Texturerkennung möglichst allgemein zu formulieren, sollen nicht nur unbekannte Parameter in der Likelihoodfunktion, sondern auch unbekannte Parameter in der Priori-Dichte angenommen werden. Die bedingte Posteriori-Dichte $p(\varepsilon_r | \partial\varepsilon_r, \varepsilon_r = \varepsilon, \mathbf{y})$ für die Zugehörigkeit ε_r des Bildelementes $r \in \Omega$ zur Textur ε folgt aus dem Bayes-Theorem (383.1) mit der Likelihoodfunktion (379.28) und der Priori-Dichte (379.36). Sie ist abhängig von den unbekanntem Parametern α_ε und β_s sowie von μ_ε , σ_ε^2 und $\beta_{s,\varepsilon}$, so daß die Dichte erhalten wird

$$p(\varepsilon_r, \alpha_\varepsilon, \beta_s, \mu_\varepsilon, \sigma_\varepsilon^2, \beta_{s,\varepsilon} | \partial\varepsilon_r, \varepsilon_r = \varepsilon, \mathbf{y}) \propto \exp\left\{-\left[\alpha_\varepsilon + \sum_{s \in N_p} \beta_s (I(\varepsilon_r, \varepsilon_{r+s}) + I(\varepsilon_r, \varepsilon_{r-s}))\right]\right. \\ \left. - \frac{1}{2\sigma_\varepsilon^2} [y_r - \mu_\varepsilon - \sum_{s \in N_p} \beta_{s,\varepsilon} (y_{r+s} - \mu_\varepsilon + y_{r-s} - \mu_\varepsilon)]^2\right\}. \quad (383.14)$$

Die Parameter α_ε und β_s steuern, wie im Zusammenhang mit (379.36) erwähnt, die Zugehörigkeit der Bildelemente zu den einzelnen Texturen und die Anordnung der Texturen, während μ_ε und $\beta_{s,\varepsilon}$ die Parameter bezeichnen, die die einzelnen Texturen beschreiben, wie bei (379.28) bereits erwähnt.

Durch Bilden der Randverteilung für die unbekannte Zugehörigkeit ε_r lassen sich die übrigen unbekannt Parameter in (383.14) eliminieren. Näherungsweise verschafft man sich die Randverteilung dadurch, daß die unbekannt Parameter durch ihre Schätzwerte ersetzt werden (Koch 1990, S.58), wie bereits bei der Begründung der iterativen Verfahren nach (383.11) oder (383.12) erwähnt wurde. Versieht man die Schätzwerte mit einem Dach, ergibt sich anstelle von (383.14) die bedingte Posteriori-Dichte für ε_r

$$p(\varepsilon_r | \partial\varepsilon_r, \varepsilon_r = \varepsilon, \mathbf{y}) \propto \exp\{-[\hat{\alpha}_\varepsilon + \sum_{s \in N_p} \hat{\beta}_s (I(\varepsilon_r, \varepsilon_{r+s}) + I(\varepsilon_r, \varepsilon_{r-s}))] - \frac{1}{2\hat{\sigma}_\varepsilon^2} [y_r - \hat{\mu}_\varepsilon - \sum_{s \in N_p} \hat{\beta}_{s, \varepsilon} (y_{r+s} - \hat{\mu}_\varepsilon + y_{r-s} - \hat{\mu}_\varepsilon)]^2\} . \quad (383.15)$$

Bei Anwendung des Pseudolikelihoodverfahrens (383.9) ergibt sich die gemeinsame Posteriori-Dichte $p(\boldsymbol{\varepsilon} | \mathbf{y})$ aus dem Produkt der bedingten Dichten (383.15) für die einzelnen Bildelemente. Die Zugehörigkeit ε_r des Bildelementes r zur Textur ε läßt sich dann mit dem deterministischen Verfahren nach (383.11) oder (383.12) schätzen. Als stochastisches Verfahren kann die Methode der simulierten Abkühlung angewendet werden.

Um die Schätzwerte $\hat{\alpha}_\varepsilon$, $\hat{\beta}_s$, $\hat{\mu}_\varepsilon$, $\hat{\sigma}_\varepsilon^2$ und $\hat{\beta}_{s, \varepsilon}$ in (383.15) zu berechnen, bedarf es eines segmentierten Bildes. Nun hat aber die Anwendung von (383.14) das Ziel, ein segmentiertes Bild zu erhalten. Das Problem ist also nur iterativ zu lösen, indem ausgehend von einer genäherten Segmentierung die Parameter geschätzt und mit den Schätzwerten das Bild erneut segmentiert wird. Unter der Voraussetzung, daß eine Segmentierung gegeben ist, wird also jetzt die Berechnung der Schätzwerte behandelt.

Ist die Zugehörigkeit der Bildelemente zu den Texturen bekannt, erhält man aus (379.36) die bedingte Dichte der unbekannt Parameter α_ε und β_s zu

$$p(\alpha_\varepsilon, \beta_s | \partial\varepsilon_r) \propto \exp\{-[\alpha_\varepsilon + \sum_{s \in N_p} \beta_s (I(\varepsilon_r, \varepsilon_{r+s}) + I(\varepsilon_r, \varepsilon_{r-s}))]\} . \quad (383.16)$$

Diese Dichte ist für jeden Gitterpunkt $r \in \Omega$ definiert. Bei Anwendung des Pseudolikelihoodverfahrens (383.9) ergibt sich die gemeinsame Dichte $p(\alpha_\varepsilon, \beta_s | \boldsymbol{\varepsilon})$ bei gegebenem Vektor $\boldsymbol{\varepsilon}$ der Texturzugehörigkeit aus dem Produkt der bedingten Dichten (383.16). Da die unbekannt Parameter α_ε und β_s nicht für die einzelnen Bildelemente definiert sind, sondern α_ε für jede Textur und β_s für sämtliche Texturen, läßt sich (383.11) oder (383.12) nicht anwenden, sondern die Schätzung muß nach (383.3) oder (383.4) erfolgen. Bei der Berechnung des Schätzwertes eines Parameters kann man aber wie für (383.11) und (383.12) annehmen, daß die Schätzwerte der übrigen Parameter gegeben sind und dann wieder iterativ vorgehen. Besonders einfach ist es, das Maximum der Dichte für diskrete Werte für α_ε und β_s zu suchen.

Zum Exponenten in (383.16) läßt sich eine Konstante addieren, ohne daß die Dichte sich ändert, da die Konstante in der Normierungskonstanten absorbiert werden kann. Die unbekannt Parameter α_ε mit $\varepsilon \in \{1, 2, \dots, T\}$ aus (379.27) werden daher nur dann eindeutig geschätzt, falls ein Parameter, beispielsweise α_1 , vorgegeben wird.

Bei gegebener Zugehörigkeit der Bildelemente zu den Texturen folgt die bedingte Dichte der unbekannt Parameter μ_ε , σ_ε^2 und $\beta_{s,\varepsilon}$ aus (379.28) mit

$$p(\mu_\varepsilon, \sigma_\varepsilon^2, \beta_{s,\varepsilon} | \partial y_r, \varepsilon_r = \varepsilon) \propto \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} [y_r - \mu_\varepsilon - \sum_{s \in N_p} \beta_{s,\varepsilon} (y_{r+s} - \mu_\varepsilon + y_{r-s} - \mu_\varepsilon)]^2\right\}. \quad (383.17)$$

Diese Dichte ist für jeden Gitterpunkt $r \in \Omega_\varepsilon$ definiert, der zur Menge Ω_ε der Gitterpunkte mit der Textur ε gehört. Wendet man zur Parameterschätzung das Pseudolikelihoodverfahren (383.9) an, folgt die gemeinsame Dichte $p(\mu_\varepsilon, \sigma_\varepsilon^2, \beta_{s,\varepsilon} | \varepsilon_r = \varepsilon, \mathbf{y})$ der Parameter $\mu_\varepsilon, \sigma_\varepsilon^2, \beta_{s,\varepsilon}$ für alle Bildelemente $r \in \Omega_\varepsilon$ aus dem Produkt der Dichten (383.17), so daß man erhält

$$p(\mu_\varepsilon, \sigma_\varepsilon^2, \beta_{s,\varepsilon} | \varepsilon_r = \varepsilon, \mathbf{y}) \propto \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} \sum_{r \in \Omega_\varepsilon} [y_r - \mu_\varepsilon - \sum_{s \in N_p} \beta_{s,\varepsilon} (y_{r+s} - \mu_\varepsilon + y_{r-s} - \mu_\varepsilon)]^2\right\}. \quad (383.18)$$

Die Maximum-Likelihood-Methode (383.5) zur Schätzung der unbekannt Parameter in dieser Dichte führt auf die Methode der kleinsten Quadrate, denn die Quadratsumme

$$\sum_{r \in \Omega_\varepsilon} [y_r - \mu_\varepsilon - \sum_{s \in N_p} \beta_{s,\varepsilon} (y_{r+s} - \mu_\varepsilon + y_{r-s} - \mu_\varepsilon)]^2 \quad (383.19)$$

ist zu minimieren. Die Schätzung $\hat{\mu}_\varepsilon$ des Parameters μ_ε ergibt sich vorweg durch Mittelbildung aus

$$\hat{\mu}_\varepsilon = \frac{1}{n_\varepsilon} \sum_{r \in \Omega_\varepsilon} y_r, \quad (383.20)$$

worin n_ε die Anzahl der Bildelemente in der Menge Ω_ε bedeutet. Für die Methode der kleinsten Quadrate werden die Zeilenvektoren eingeführt

$$\mathbf{x}_r' = [y_{r+s_1} + y_{r-s_1} - 2\hat{\mu}_\varepsilon, y_{r+s_2} + y_{r-s_2} - 2\hat{\mu}_\varepsilon, \dots, y_{r+s_u} + y_{r-s_u} - 2\hat{\mu}_\varepsilon], \quad (383.21)$$

die für jedes Bildelement $r \in \Omega_\varepsilon$ der Textur ε definiert sind, und der Vektor β_ε der unbekannt Parameter für die Textur ε mit

$$\beta_\varepsilon = [\beta_{s_1,\varepsilon}, \beta_{s_2,\varepsilon}, \dots, \beta_{s_u,\varepsilon}]', \quad (383.22)$$

worin der Index u durch die Größe der Nachbarschaft (376.19), das heißt, von der gewählten Ordnung des Markoff-Zufallsfeldes bestimmt wird. Mit $\mathbf{y} = (y_r - \hat{\mu}_\epsilon)$ ergeben sich die Beobachtungsgleichungen der Methode der kleinsten Quadrate zu

$$\mathbf{X}\boldsymbol{\beta}_\epsilon = E(\mathbf{y}) \quad (383.23)$$

mit

$$\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_r, \dots, \mathbf{x}_{n_\epsilon}]'$$

Die Lösungen der Normalgleichungen (Koch 1987, S.186) führen dann auf die Schätzwerte $\hat{\boldsymbol{\beta}}_\epsilon$ der unbekannt Parameter $\boldsymbol{\beta}_\epsilon$

$$\hat{\boldsymbol{\beta}}_\epsilon = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \left[\sum_{r \in \Omega_\epsilon} \mathbf{x}_r \mathbf{x}_r' \right]^{-1} \sum_{r \in \Omega_\epsilon} \mathbf{x}_r (y_r - \hat{\mu}_\epsilon) \quad (383.24)$$

Die Schätzung $\hat{\sigma}_\epsilon^2$ der Varianz σ_ϵ^2 berechnet sich zu

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n_\epsilon - u} \sum_{r \in \Omega_\epsilon} (\mathbf{x}_r' \hat{\boldsymbol{\beta}}_\epsilon - y_r + \hat{\mu}_\epsilon)^2 \quad (383.25)$$

Wie bereits im Zusammenhang mit (379.28) erläutert, läßt sich die bedingte Dichte (383.17) und die daraus abgeleitete gemeinsame Dichte (383.18) über das Auto-Modell eines Nachbarschafts-Gibbs-Feldes oder über das Zufallsfeld eines nichtkausalen autoregressiven Prozesses mit Markoff-Eigenschaft herleiten. Für beide Zufallsfelder gelten daher die Parameterschätzungen (383.24) und (383.25). Die Eigenschaft dieser Schätzungen bezüglich Konsistenz wurden im autoregressiven Modell von Kashyap und Chellappa (1983) untersucht. Benutzt man für die Maximum-Likelihood-Schätzung der Texturparameter nicht die Dichte (383.18), die aus dem Pseudolikelihoodverfahren folgt, sondern die Dichte (379.29) der gemeinsamen Verteilung, ergeben sich nichtlineare Gleichungssysteme für die unbekannt Parameter, die für die numerische Rechnung weniger geeignet sind. Dies wurde bereits im Zusammenhang mit (378.50) erläutert.

Sind die Parameter μ_ϵ , σ_ϵ^2 und $\beta_{s,\epsilon}$ in (383.18) bekannt, kann diese Dichte auch dazu benutzt werden, die in (379.30) eingeführten Parameter für die Drehung und die Maßstabsänderung von Texturen zu schätzen. Man wendet (383.3) oder (383.4) an, wobei, wie im Zusammenhang mit (383.16) erwähnt, iterativ vorgegangen werden kann. Besonders einfach ist es wieder, das Maximum der Verteilung für diskrete Werte der unbekannt Parameter zu suchen.

Wie erwähnt, sind ausgehend von einer genäherten Segmentierung die Parameter α_ϵ und β_s für die Zugehörigkeit zu den Texturen mit Hilfe von (383.16) und die Texturparameter μ_ϵ , σ_ϵ^2 und $\beta_{s,\epsilon}$ nach (383.20), (383.24) und (383.25) zu schätzen. Mit den Schätzwerten ist dann nach (383.15) die Segmentierung des Bildes zu korrigieren. Aufgrund

der korrigierten Segmentierung sind die Parameter erneut zu schätzen, und es ist iterativ bis zur Konvergenz der Schätzung und der Segmentierung fortzufahren. Anwendungen dieser iterativen Schätzverfahren finden sich bei (Manjunath und Chellappa 1991; Won und Derin 1992). Liegen Trainingsgebiete für die Texturen vor, sollte der Iterationsprozeß mit den Schätzwerten der Texturparameter für die Trainingsgebiete starten.

Eine genäherte Segmentierung eines Bildes kann beispielsweise auf der Grundlage der bereits behandelten Kantenextraktion erfolgen. Eine weitere Möglichkeit besteht darin, das Bild systematisch zu unterteilen und aufgrund dieser Segmentierung die Texturparameter nach (383.20), (383.24) und (383.25) zu berechnen. Um zu entscheiden, ob diese Parameter identische Texturen repräsentieren, berechnet man zweckmäßigerweise Abstandsmaße (Chatterjee 1993; Chellappa und Chatterjee 1985; Koch 1990, S.138; Manjunath und Chellappa 1991). Mit dem Vektor θ_ϵ für die Textur ϵ

$$\theta_\epsilon = [\hat{\mu}_\epsilon, \hat{\beta}_\epsilon', \hat{\sigma}_\epsilon^2]' \quad (383.26)$$

erhält man als Abstand $d(i, j)$ zwischen den Texturen der Segmentierung i und j

$$d(i, j) = (\theta_i - \theta_j)' (\theta_i - \theta_j) / [(\theta_i' \theta_i) + (\theta_j' \theta_j)] \quad (383.27)$$

Aufgrund der Abstände faßt man die Segmentierungen, die zu einer Textur gehören, zusammen und bestimmt die Anzahl der Texturen in einem digitalen Bild. Diese Anzahl läßt sich auch mit der zugehörigen Verteilung ableiten. Man erhält das Akaike-Kriterium (Sakamoto et al. 1986), das beispielsweise bei (Won und Derin 1992) angewendet und modifiziert wird.

Um den Rechenaufwand bei den iterativen Segmentierungsverfahren in Grenzen zu halten, werden häufig aus einem digitalen Bild weitere Bilder mit geringerer Auflösung erzeugt und diese Bilder segmentiert. Die Segmentierung wird anschließend durch Verfeinerung der Auflösung schrittweise verbessert (Bouman und Liu 1991; Cohen und Cooper 1987). Unterschiedliche Auflösungen findet man auch bei (Geman et al. 1990). Δ

Beispiel 3: Im Beispiel 1 des Kapitels 379 wurde mit dem "Gibbs sampler" eine Textur aus der bedingten Normalverteilung (379.28) generiert, so daß die gemeinsame Verteilung (379.29) erhalten wird. Das generierte Bild 379-5 soll nun dazu dienen, die Parameter β_ϵ , μ_ϵ und σ_ϵ^2 der Generierung zu schätzen, wobei angenommen wird, daß die gemeinsame Verteilung aus dem Produkt der bedingten Verteilungen folgt. Dann läßt sich das Pseudolikelihoodverfahren (383.9) anwenden, und die Schätzwerte für die Parameter ergeben sich aus (383.24), (383.20) und (383.25). Man erhält

$$\hat{\beta}_\epsilon = [0,084, 0,110, 0,062, 0,084, 0,043, 0,064, 0,070, 0,049, \\ 0,064, 0,059, 0,067, 0,061, -0,060, -0,081, -0,056, -0,055,$$

-0,069, -0,062, -0,056, -0,061, -0,070, -0,052]'

$$\hat{\mu}_\varepsilon = 0,004, \quad \hat{\sigma}_\varepsilon^2 = 1,002.$$

Die für die Generierung der Textur benutzten Parameter werden also durch die Parameterschätzung recht gut reproduziert. ▲

Beispiel 4: Eine Anordnung von vier Texturen wurde im Beispiel 2 des Kapitels 379 mit dem "Gibbs sampler" aus der bedingten Dichte (379.36) generiert. Das generierte Bild 379-6 soll jetzt dazu benutzt werden, die für die Generierung verwendeten Parameter α_ε und β_s mit dem Pseudolikelihoodverfahren zu schätzen. Für die Dichte dieser Parameter wird daher angenommen, daß sie aus dem Produkt der bedingten Dichten (383.16) folgt. Wie im Zusammenhang mit (383.16) beschrieben, wird das Maximum der Verteilung für α_ε und β_s unter der Annahme gesucht, daß jeweils nur ein Parameter unbekannt ist und die übrigen Parameter gegeben sind und daß in jeweils einer Iteration sukzessiv sämtliche Parameter geschätzt werden. Sucht man das Maximum für diskrete Werte mit einem Fortschritt von $\pm 0,01$ für α_ε und $\pm 0,05$ für β_s und gibt man, wie für (383.16) erläutert, $\alpha_1 = 0,00$ vor, erhält man nach drei Iterationen die Schätzwerte

$$\begin{aligned} \hat{\alpha}_2 &= 0,01, & \hat{\alpha}_3 &= 0,01, & \hat{\alpha}_4 &= 0,01, \\ \hat{\beta}_{1,0} &= 0,4, & \hat{\beta}_{0,1} &= 0,3, & \hat{\beta}_{1,-1} &= 1,4, & \hat{\beta}_{1,1} &= 1,2. \end{aligned}$$

Die Ergebnisse zeigen, daß die Parameter β_s sich gut und die Parameter α_ε weniger gut durch die Schätzung reproduzieren lassen. ▲

Literatur

- Aarts, E.H.L. und P.J.M. van Laarhoven (1987) Simulated annealing: a pedestrian review of the theory and some applications. In: Devijver, P.A. und J. Kittler (Hrsg.) Pattern Recognition Theory and Applications, Springer, Berlin, 179-192
- Acuna, C.O. (1992) Texture modeling using Gibbs distributions. Comput Vision Graphics Image Process: Graphical Models Image Process, 54: 210-222
- Averintsev, M.B. (1970) One method of describing random fields with a discrete argument. Problems of Information Transmission 6: 169-175 (Translated from Problemy Peredachi Informatsii, 6: 100-108)
- Azizi, S.A. (1983) Entwurf und Realisierung digitaler Filter. Oldenbourg, München
- Barrodale, I. und R.E. Erickson (1980) Algorithms for least-squares linear prediction and maximum entropy spectral analysis, Part I: Theory and Part II: FORTRAN program. Geophysics, 45: 420-446
- Bartlett, M.S. (1955) An Introduction to Stochastic Processes. University Press, Cambridge
- Bates, R.H.T. und M.J. McDonnell (1989) Image Restoration and Reconstruction. Clarendon, Oxford
- Besag, J.E. (1974) Spatial interaction and the statistical analysis of lattice systems. J Royal Statist Society, B 36: 192-236
- Besag, J.E. (1975) Statistical analysis of non-lattice data. The Statistician, 24: 179-195
- Besag, J.E. (1986) On the statistical analysis of dirty pictures. J Royal Statist Society, B 48: 259-302
- Blackman, R.B. und J.W. Tuckey (1958) The Measurement of Power Spectra. Dover, New York
- Bouman, C. und B. Liu (1991) Multiple resolution segmentation of textured images. IEEE Trans Pattern Anal Machine Intell, 13: 99-113
- Box, G.E.P. und G.M. Jenkins (1970) Time Series Analysis. Holden-Day, San Francisco
- Busch, A. (1992) Bayes-Statistik und Markoff-Felder für die Restaurierung digitaler Bilder. Deutsche Geodätische Kommission, Reihe C, 396, München
- Busch, A. und K.R. Koch (1990) Reconstruction of digital images using Bayesian estimates. Z Photogrammetrie und Fernerkundung, 58: 173-181
- Buttkus, B. (1991) Spektralanalyse und Filtertheorie in der angewandten Geophysik. Springer, Berlin
- Cappellini, V., A.G. Constantinides und P. Emiliani (1978) Digital Filters and Their Applications. Academic Press, London

- Chatterjee, S. (1993) Classification of natural textures using Gaussian Markov random field models. In: Chellappa, R. und A. Jain (Hrsg.) *Markov Random Fields*, Academic Press, Boston, 159-177
- Chellappa, R. (1985) Two-dimensional discrete Gaussian Markov random field models for image processing. In: Kanal, L.N. und A. Rosenfeld (Hrsg.) *Progress in Pattern Recognition 2*, North-Holland, Amsterdam, 79-112
- Chellappa, R. und S. Chatterjee (1985) Classification of textures using Gaussian Markov random fields. *IEEE Trans Acoust Speech Signal Process*, ASSP-33: 959-963
- Chou, P.B. und C.M. Brown (1990) The theory and practice of Bayesian image labeling. *Int J Computer Vision*, 4: 185-210
- Cohen, F.S. und D.B. Cooper (1987) Simple parallel hierarchical and relaxation algorithms for segmenting noncausal Markovian random fields. *IEEE Trans Pattern Anal Machine Intell*, PAMI-9: 195-219
- Cohen, F.S. und Z. Fan (1992) Maximum likelihood unsupervised textured image segmentation. *Comput Vision Graphics Image Process: Graphical Models Image Process*, 54: 239-251
- Cohen, F.S., Z. Fan und S. Attali (1991 a) Automated inspection of textile fabrics using textural models. *IEEE Trans Pattern Anal Machine Intell*, 13: 803-808
- Cohen, F.S., Z. Fan und M.A. Patel (1991 b) Classification of rotated and scaled textured images using Gaussian Markov random field models. *IEEE Trans Pattern Anal Machine Intell*, 13: 192-202
- Cooley, J.W. und J.W. Tukey (1965) An algorithm for the machine calculation of complex Fourier series. *Math Computation*, 19: 297-301
- Cressie, N.A.C. (1991) *Statistics for Spatial Data*. Wiley, New York
- Cross, G.R. und A.K. Jain (1983) Markov random field texture models. *IEEE Trans Pattern Anal Machine Intell*, PAMI-5: 25-39
- Davis, P.J. (1979) *Circulant Matrices*. Wiley, New York
- Derin, H. und W.S. Cole (1986) Segmentation of textured images using Gibbs random fields. *Comput Vision Graphics Image Process*, 35: 72-98
- Derin, H. und H. Elliott (1987) Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Trans Pattern Anal Machine Intell*, PAMI-9: 39-55
- Doob, J.L. (1953) *Stochastic Processes*. Wiley, New York
- Dudgeon, D.E. und R.M. Mersereau (1984) *Multidimensional Digital Signal Processing*. Prentice-Hall, Englewood Cliffs
- Fiasconaro, J.G. (1979) Two-dimensional nonrecursive filters. In: Huang, T.S. (Hrsg.) *Picture Processing and Digital Filtering*, Springer, Berlin, 69-129
- Fichtenholz, G.M. (1987) *Differential- und Integralrechnung III*. Deutscher Verlag der

Wissenschaften, Berlin

- Fritsch, D. (1982) Entwurf digitaler zweidimensionaler nichtrekursiver Filter. Deutsche Geodätische Kommission, Reihe C, 275, München
- Fritsch, D. (1987) On algorithms solving the l_{∞} -approximation in geometric modelling. Proceed ISPRS Intercommission Conf Fast Processing of Photogrammetric Data, Interlaken, Switzerland, Institut für Geodäsie und Photogrammetrie, ETH Zürich, 142-155
- Gelfand, A.E., S.E. Hills, A. Racine-Poon und A.F.M. Smith (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. JASA, 85, 412: 972-985
- Geman, D., S. Geman und C. Graffigne (1987) Locating texture and object boundaries. In: Devijver, P.A. und J. Kittler (Hrsg.) Pattern Recognition Theory and Applications, Springer, Berlin, 165-177
- Geman, D., S. Geman, C. Graffigne und P. Dong (1990) Boundary detection by constrained optimization. IEEE Trans Pattern Anal Machine Intell, 12: 609-628
- Geman, S. und D. Geman (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans Pattern Anal Machine Intell, PAMI-6: 721-741
- Gonzales, R.C. und R.E. Woods (1992) Digital Image Processing. Addison-Wesley, Reading
- Grafarend, E. und B. Schaffrin (1979) Kriterion-Matrizen I - zweidimensionale homogene und isotrope geodätische Netze. Z Vermessungswesen, 104: 133-149
- Griffeath, D. (1976) Introduction to random fields. In: Kemeny, J.G., J.L. Snell und A.M. Knapp (Hrsg.) Denumerable Markov Chains, Springer, Berlin, 425-458
- Grimmet, G.R. (1973) A theorem about random fields. Bull London Math Soc, 5: 81-84
- Hamming, R.W. (1987) Digitale Filter. VCH Verlagsgesellschaft, Weinheim
- Haralick, R.M. (1984) Digital step edges from zero crossing of second directional derivatives. IEEE Trans Pattern Anal Machine Intell, PAMI-6: 58-68
- Haykin, S. und S. Kesler (1979) Prediction-error filtering and maximum-entropy spectral estimation. In: Haykin, S. (Hrsg.) Nonlinear Methods of Spectral Analysis, Springer, Berlin, 9-72
- He, Guangping (1989) Bildverbesserung mittels digitaler Filterung. Veröff Institut für Photogrammetrie und Fernerkundung, 31, Wien
- Heller, W.D., H. Lindenberg, M. Nuske und K.H. Schriever (1978) Stochastische Systeme. Walter de Gruyter, Berlin
- Henrici, P. (1974) Applied and Computational Complex Analysis, Vol. 1. Wiley, New York

- Hess, W. (1989) *Digitale Filter*. Teubner, Stuttgart
- Hessmann, N. (1983) *Digitale Signalverarbeitung*. Vogel, Würzburg
- Hirvonen, R.A. (1962) On the statistical analysis of gravity anomalies. *Annales Academiae Scientiarum Fennicae, Series A, III, 58*, Helsinki
- Ising, E. (1925) Beitrag zur Theorie des Ferromagnetismus. *Z Physik, 31*: 253-258
- Jähne, B. (1989) *Digitale Bildverarbeitung*. Springer Verlag, Berlin
- Jaglom, A.M. (1959) *Einführung in die Theorie der stationären Zufallsfunktionen*. Akademie-Verlag, Berlin
- Jain, A.K. (1989) *Fundamentals of Digital Image Processing*. Prentice-Hall, Englewood Cliffs
- Jenkins, G.M. und D.G. Watts (1968) *Spectral Analysis*. Holden-Day, San Francisco
- Johnson, V.E., W.H. Wong, X. Hu und C. Chen (1991) Image restoration using Gibbs priors: boundary modeling, treatment of blurring, and selection of hyperparameter. *IEEE Trans Pattern Anal Machine Intell, 13*: 413-425
- Kashyap, R.L. (1984) Characterization and estimation of two-dimensional ARMA models. *IEEE Trans Inform Theory, IT-30*: 736-745
- Kashyap, R.L. und R. Chellappa (1983) Estimation and choice of neighbors in spatial-interaction models of images. *IEEE Trans Inform Theory, IT-29*: 60-72
- Khintchine, A. (1934) Korrelationstheorie der stationären stochastischen Prozesse. *Math Annalen 109*: 604-615
- Kindermann, R. und J.L. Snell (1980) *Markov Random Fields and their Applications*. American Mathematical Society, Providence
- Kirkpatrick, S., C.D. Gelatt und M.P. Vecchi (1983) Optimization by simulated annealing. *Science 220*: 671-680
- Koch, K.R. (1973) Höheninterpolation mittels gleitender Schrägebene und Prädiktion. *Vermessung, Photogrammetrie, Kulturtechnik, Mitteilungsblatt, 71*: 229-232
- Koch, K.R. (1975) Rekursive numerische Filter. *Z Vermessungswesen, 100*: 281-292
- Koch, K.R. (1987) *Parameterschätzung und Hypothesentests in linearen Modellen*. Dümmler, Bonn
- Koch, K.R. (1990) *Bayesian Inference with Geodetic Applications*. Springer, Berlin
- Koch, K.R. (1991) Gibbs-Verteilungen in der digitalen Bildverarbeitung. *Z Vermessungswesen, 116*: 560-569
- Kraus, K. (1990) *Fernerkundung Band 2, Auswertung photographischer und digitaler Bilder*. Dümmler, Bonn
- Lakshmanan, S. und H. Derin (1989) Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing. *IEEE Trans Pattern Anal Machine Intell, 11*: 799-813

- Lamperti, J. (1977) Stochastic Processes. Springer, Berlin
- Landau, L.D. und E.M. Lifschitz (1979) Lehrbuch der theoretischen Physik, Bd. V, Statistische Physik, Teil 1. Akademie-Verlag, Berlin
- Manjunath, B.S. und R. Chellappa (1991) Unsupervised texture segmentation using Markov random field models. IEEE Trans Pattern Anal Machine Intell, 13: 478-482
- Manjunath, B.S., T. Simchony und R. Chellappa (1990) Stochastic and deterministic networks for texture segmentation. IEEE Trans Acoust Speech Signal Process, 38: 1039-1049
- Mao, J. und A.K. Jain (1992) Texture classification and segmentation using multiresolution simultaneous autoregressive models. Pattern Recognition, 25: 173-188
- Marko, H. (1986) Methoden der Systemtheorie. Springer, Berlin
- Marsaglia, G. und T.A. Bray (1964) A convenient method for generating normal variables. SIAM Review, 6: 260-264
- Mathar, R. und D. Pfeifer (1990) Stochastik für Informatiker. Teubner, Stuttgart
- Meier, S. und A. Borkowski (1993) Ordinatenabtastung diskreter Signale. Z Vermessungswesen, 118: 11-20
- Meier, S. und W. Keller (1990) Geostatistik. Springer, Wien
- Mersereau, R.M. und D.E. Dudgeon (1978) Two-dimensional digital filtering. In: Mitra, S.K. und M.P. Ekstrom (Hrsg.) Two-Dimensional Digital Signal Processing, Dowden, Hutchinson & Ross, Stroudsburg, 24-37
- Mester, R. (1989) Regionenorientierte Bildsegmentierung unter Verwendung stochastischer Bildmodelle. VDI Fortschrittberichte, Reihe 10, 106, VDI Verlag, Düsseldorf
- Metropolis, N., A.W. Rosenbluth, M.L. Rosenbluth, A.H. Teller und E. Teller (1953) Equation of state calculations by fast computing machines. J Chem Phys, 21: 1087-1092
- Monin, A.S. und A.M. Yaglom (1975) Statistical Fluid Mechanics, Vol. 2. MIT Press, Cambridge
- Moritz, H. (1980) Advanced Physical Geodesy. Wichmann, Karlsruhe
- Nadabar, S.G. und A.K. Jain (1992) Parameter estimation in MRF line process models. Proc 1992 IEEE Comp Soc Conf Comp Vision Pattern Recognition, 528-533, IEEE Computer Society Press, Los Alamitos
- Nussbaumer, H.J. (1981) Fast Fourier Transform and Convolution Algorithms. Springer, Berlin
- Oppenheim, A.V. und R.W. Schaffer (1975) Digital Signal Processing. Prentice-Hall, Englewood Cliffs
- Pan, Y.J. (1994) Autoregressive Modelle für die Analyse digitaler Bilder. Institut für Theoretische Geodäsie der Universität Bonn, in Vorbereitung

- Papoulis, A. (1962) *The Fourier Integral and its Application*. McGraw-Hill, New York
- Papoulis, A. (1977) *Signal Analysis*. McGraw-Hill, Auckland
- Priestley, M.B. (1981) *Spectral Analysis and Time Series*. Academic Press, London
- Rabiner, L.R. und B. Gold (1975) *Theory and Application of Digital Signal Processing*. Prentice-Hall, Englewood Cliffs
- Read, R.R. (1979) Two-dimensional recursive filtering. In: Huang, T.S. (Hrsg.) *Picture Processing and Digital Filtering*, Springer, Berlin, 131-176
- Ripley, B.D. (1987) *Stochastic Simulation*. Wiley, New York
- Rubinstein, R.Y. (1981) *Simulation and the Monte Carlo Method*. Wiley, New York
- Rutishauser, H. (1976) *Vorlesungen über numerische Mathematik, Band 1*. Birkhäuser, Basel
- Sakamoto, Y., M. Ishiguro und G. Kitagawa (1986) *Akaike Information Criterion Statistics*. Reidel, Dordrecht
- Scharf, L.L. (1991) *Statistical Signal Processing*. Addison-Wesley, Reading
- Schlittgen, R. und B.H.J. Streitberg (1991) *Zeitreihenanalyse*. Oldenbourg, München
- Schmidt, M. (1993a) Grundlagen rekursiver und nichtrekursiver Filterung, Teil I: Q-dimensionale Filterung. *Z Vermessungswesen* 118: 74-82
- Schmidt, M. (1993b) Grundlagen rekursiver und nichtrekursiver Filterung, Teil II: Eindimensionale Filterung. *Z Vermessungswesen* 118: 116-121
- Schüßler, H.W. (1973) *Digitale Systeme zur Signalverarbeitung*. Springer, Berlin
- Schulte, S. (1987) Modellierung von Beobachtungsreihen durch ein erweitertes Autoregressives Modell. Deutsche Geodätische Kommission, Reihe C, 327, München
- Silverman, B.W., C. Jennison, J. Stander und T.C. Brown (1990) The specification of edge penalties for regular and irregular pixel images. *IEEE Trans Pattern Anal Machine Intell*, 12: 1017-1024
- Stearns, S.D. (1984) *Digitale Verarbeitung analoger Signale*. Oldenbourg, München
- Stumpf, H. und W. Schuler (1981) *Elektrodynamik*. Vieweg, Braunschweig
- Therrien, C.W., T.F. Quatieri und D.E. Dudgeon (1986) Statistical model-based algorithms for image analysis. *Proc IEEE*, 74: 532-551
- Wahl, F.M. (1980) Der Entwurf zweidimensionaler rekursiver Filter und ihre Anwendung in der Bildverarbeitung. *Nachrichtentechnische Berichte*, 2, Lehrstuhl für Nachrichtentechnik der Technischen Universität München
- Wahl, F.M. (1989) *Digitale Bildsignalverarbeitung*. Springer Verlag, Berlin
- Wiener, N. (1930) Generalized harmonic analysis. *Acta Math* 55: 117-258
- Wiener, N. (1949) *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications*. Wiley, New York
- Won, C.S. und H. Derin (1992) *Unsupervised segmentation of noisy and textured images*

- using Markov random fields. *Comput Vision Graphics Image Process: Graphical Models Image Process*, 54: 308-328
- Woods, J.W. (1972) Two-dimensional discrete Markovian fields. *IEEE Trans Inform Theory*, IT-18: 232-240
- Zhao, Y., X. Zhuang, L. Atlas und L. Anderson (1992) Parameter estimation and restoration of noisy images using Gibbs distributions in hidden Markov models. *Comput Vision Graphics Image Process: Graphical Models Image Process*, 54: 187-197

Sachverzeichnis

- Ableitungsfilter, 81
- Abstandsmaß, 329
- Abtastfrequenz, 19, 97
- Abtastintervall, 18, 95
- Abtastkreisfrequenz, 19, 97
- Abtastmatrix, 95, 105
- Abtasttheorem, 22, 98, 106
 - , n-dimensionales, 98
 - , zweidimensionales, 106
- Abtastung, 18, 94, 105
- Achter-Nachbarschaft, 145, 271, 274, 314
- Akaike-Kriterium, 245, 329
- Amplitude, 5, 7, 118, 165
 - , komplexe, 7, 13
- Amplitudenspektrum, 12, 13, 14, 18, 33, 52, 61, 83, 94, 149
 - , diskretes, 12
 - , kontinuierliches, 13, 14, 18, 94
- Approximationsfehler, 55, 70, 83, 132, 149
- ARMA-Prozeß, 206, 245
- Ausgangssignal, 23, 29, 37, 47, 59, 74, 89, 98, 123, 144, 185, 189, 232, 236, 291
- Autokorrelationsfunktion, 164, 165, 194, 204, 282
- Autokovarianz, 165, 194, 237, 247, 282
- Autokovarianzfunktion, 164, 167, 170, 177, 187, 189, 192, 202, 204, 211, 216, 220, 223, 233, 241, 251, 279, 290, 316
- Auto-Modell, 265, 275, 290, 308, 328
 - -Poisson-Modell, 268

- Bandpaß, 56, 172, 235
- Bandsperre, 56
- Bartlett-Fenster, 224
- Bayes-Schätzung, 320
 - -Statistik, 300
 - -Theorem, 181, 320, 325
- Beobachtungsgleichung, 8, 328
- Bildelement, 299, 313, 321

- Bildrestaurierung, 300, 321
- Bildsegmentierung, 309, 325
- Bildverarbeitung, 92, 112, 299, 320
- Binomialfilter, 79, 146
 - , eindimensionales, 79, 146
 - , zweidimensionales, 146
- Binomialverteilung, 267
- Butterworthfilter, 61, 129
 - , zweidimensionales, 129
- Clique, 259, 262, 264, 271, 302
- Codierungsmethode ("coding method"), 322
- Dämpfung, 52
- Dekonvolution, 51, 117
- Deltafunktion, 4, 17, 24, 98, 177, 224
 - , n-dimensionale, 98
- Dezibel, 52
- Dichte, 164, 180, 197, 256, 265, 274, 295, 302, 322
 - , bedingte, 180, 197, 256, 261, 274, 284, 295, 302, 322
 - , gemeinsame, 198, 298, 326
- Differenzgleichung, 28, 45, 100, 109, 112
- Dreiecksignal, 16, 39, 155
 - , diskretes, 39
 - , kontinuierliches, 16, 155
- Durchlaßbereich, 57, 62, 129
- Durchsickern ("leakage"), 224, 229
- Einer-Clique, 264, 275, 308
- Eingangssignal, 23, 29, 37, 47, 59, 74, 89, 98, 123, 145, 185, 188, 232, 294
- Einheitsimpuls, 5, 24, 99, 106, 177, 204
 - , n-dimensionaler, 99
 - , zweidimensionaler, 106
- Einheitssignal, 14
- Elementarereignis, 162, 250
- Elementarfilter, 78, 143
- Energie, 258, 264, 274, 302, 324
- Ensemble, 162, 209, 216

Entwurfsaufgabe, 54, 61, 120, 129, 148

Ergodizität, 209, 213, 216, 217, 316

- bezüglich der Autokovarianzfunktion, 213, 217, 316

- bezüglich des Erwartungswertes, 209, 216, 316

Erwartungswertfunktion, 164, 178, 186, 189, 192, 202, 204, 209

Eulersche Formel, 7

Exponentialtyp, 169, 175

Faltung, 15, 24, 29, 35, 36, 87, 90, 99, 102, 105, 156, 191, 218, 230, 301, 319

-, diskrete, 24, 29, 87, 99, 319

-, lineare, 36, 104, 219

-, n-dimensionale, 99, 102

-, periodische, 35

-, schnelle, 37, 90, 105, 156, 218, 230

-, zweidimensionale, 106, 301

Fejer-Kern, 40, 227

Fensterfunktion, 86, 151, 223, 230, 319

-, eindimensionale, 86, 319

-, zweidimensionale, 151, 319

Fensterfunktionsmethode, 86, 151

Filter, 27, 31, 45, 53, 61, 71, 79, 83, 89, 109, 121, 134, 143, 156, 189, 234, 280

-, antikausales, 47, 58

-, digitales, 45, 71, 112, 134, 158, 189

-, eindimensionales, 45

-, frequenzselektives, 27, 55

-, inverses, 51, 117

-, kausales, 45, 58, 71, 112, 121, 134, 189, 236

-, lineares, 45, 58, 112, 134, 156, 158

-, linearphasiges, 53, 72, 118, 121, 134

-, minimalphasiges, 52

-, n-dimensionales, 102

-, nichtkausales, 46, 113, 121, 140, 236

-, nichtlineares, 159, 160

-, nichtrekursives, 31, 71, 78, 83, 89, 109, 134, 143, 148, 156, 189, 236, 280

-, nullphasiges, 53, 59, 61, 79, 83, 118, 122, 234

-, rekursives, 31, 58, 61, 89, 109, 129, 156

-, stabiles, 49, 72, 116

-, verschiebungsvariantes, 158

- , zweidimensionales, 112, 121, 129, 134, 143, 148, 156
- mit endlicher Impulsantwort, 32, 103
- mit unendlicher Impulsantwort, 31, 102

Filterkoeffizient, 45, 55, 60, 61, 79, 83, 120, 132

Filterung, 27, 46, 59, 112, 114, 115, 122, 125, 188, 239

- in Richtung des dritten Quadranten, 114, 123
- in Richtung des ersten Quadranten, 112, 122
- in Richtung des vierten Quadranten, 115, 125
- in Richtung des zweiten Quadranten, 114, 124
- nach rückwärts, 46, 59
- nach vorwärts, 46, 59

Fourier-Integral, 13, 15, 94

- -Reihe, 11, 27, 29, 93, 100
- -Transformation, 13, 27, 29, 33, 41, 90, 103, 105, 110, 170, 218, 220, 223
- -, diskrete, 33, 41, 90, 103, 110
- -, inverse, 13, 33, 94, 103, 105, 110, 223
- -, n-dimensionale, 94, 103
- -, schnelle, 41, 90, 218, 227, 230
- -, zweidimensionale, 110

FPE-Kriterium ("final prediction error"), 245

Frequenz, 6, 8, 20, 26, 52, 56, 100

- , normierte, 26, 100

Frequenzantwort, 26, 53, 59, 61, 100, 108, 120, 122, 129, 188, 233

- , oktagonale, 128, 130, 139
- , reelle, 53, 59, 61, 122, 129, 294
- , rotationssymmetrische, 108, 120, 148
- , separierbare, 117, 128, 129
- , vierquadrantensymmetrische, 125, 129, 139, 144
- , zweiquadrantensymmetrische, 124, 140, 294

Frequenzerlegung, 43

Gauß-Markoff-Prozeß, 196

- -Markoff-Zufallsfeld, 266, 285, 294
- -Prozeß, 178, 217, 285
- -Zufallsfeld, 255, 266, 269, 275, 301
- -, vektorielles, 255, 269

geometrische Reihe, 9, 50

gerade Funktion, 12, 15, 54, 167, 173, 216

Gibbs-sampler", 310, 314, 325

Gibbs-Verteilung, 258, 261, 274, 299, 302, 313

- -Zufallsfeld, 259

- -, vektoriell, 260

Gibbssches Phänomen, 85

Gitter, 39, 212, 270, 276, 296, 299, 316, 322

-, toroidales, 272, 276, 296

Gradientenfilter, 81, 148

Graph, 256, 258, 270

Grauwert, 299, 306, 312, 321

Grundfrequenz, 8, 33, 93

Gruppenlaufzeit, 52

Halbwertsbreite, 169

Harmonische, 8, 33

harmonische Schwingung, 5, 8, 74, 165, 167

Hochpaßfilter, 56, 62

Homogenität, 251, 280, 291

Homogenitätsbedingung, 280, 291

ICA-Verfahren ("iterated conditional averages"), 323

ICM-Verfahren ("iterated conditional modes"), 323

Idealfunktion, 55, 83, 120

Imaginärteil, 11, 15

Implementierung, 89, 156

Impulsantwort, 24, 31, 45, 71, 87, 90, 98, 134, 156, 158, 186, 232, 291

Index, 4, 162, 250

Indexmenge, 4, 92, 162, 250

Indexvektor, 92, 94, 163, 250, 264

Interpolation, 22, 97

Invertierbarkeitsbedingung, 190, 192, 207

Irrfahrt, 184

Ising-Modell, 258, 278

Isotropie, 251

Kalman-Bucy-Filter, 241

Kantenelement, 303, 307,

Kantenextraktion, 147, 306, 325, 329

Kaskadenstruktur, 47, 66, 82, 116, 122, 144

Kausalität, 25, 99, 107

komplexe Größe, 7, 11, 64

konjugiert komplexe Größe, 7, 11, 64

Konvolution, 15, 24

Kosinusfunktion ("cosine taper"), 229

Kreisfrequenz, 5, 18, 94

Kreuzkovarianz, 165, 237

Kreuzkovarianzfunktion, 165, 168, 193, 202, 233, 291

Kreuzspektrum, 173, 174

Laplacefilter, 82, 147

- , eindimensionales, 82, 147

- , zweidimensionales, 147

Laplace-Transformation, 18

Leistung, 164, 170, 172

Leistungsspektrum, 170

Leistungsverstärkung, 61

Levinson-Rekursionsalgorithmus, 242

Likelihoodfunktion, 242, 287, 299, 300, 321

lokale Charakteristik, 256

MA-Prozeß, 203, 245

Markoff-Eigenschaft, 180, 195, 256, 261, 277, 284, 294, 299

- -Kette, 182, 184

- -, homogene, 185

- -Modell, 290

- -, konditionales (CM-Modell), 290

- -Prozeß, 179, 196

- -Zufallsfeld, 256, 261, 264, 270, 299

- -, vektorielles, 257, 263, 264, 270

- -Zufallsfeld der Ordnung p , 271, 277, 302

Maximum-Entropie-Methode, 231

- -Likelihood-Methode, 71, 242, 299, 300, 321

- -Posteriori-Schätzung (MAP-Schätzung), 321, 324

Medianfilter, 159

Methode der kleinsten Quadrate, 8, 71, 244, 248, 320

- der simulierten Abkühlung ("simulated annealing"), 324

- Metropolis-Algorithmus**, 324
Modell, 70, 84, 133, 148, 201, 239, 245, 248, 267, 301, 312
 -, autobinomiales, 267, 312
 -, autologistisches, 267
 -, erweitertes autoregressives, 201, 245
 -, gemischtes, 248
 -, lineares, 70, 84, 133, 148, 301
 - - für die Parameterschätzung, 70, 84, 133, 148
 - der Prädiktion und Filterung, 239
Modellparameter, 55, 62, 120, 130
Moment, 164, 167, 178, 186
 -, erstes, 164, 167, 178
 -, zweites, 164, 167, 178, 186
Momentfunktion, 164, 166, 178, 186, 189, 209, 251, 280, 291, 316

Nachbarschaft, 256, 262, 271, 290
Nachbarschafts-Gibbs-Feld, 259, 261, 290, 300, 328
Nachbarschaftssystem, 256, 304
Normalgleichungsmatrix, 9, 243, 328
Normalprozeß, 178
Normalverteilung, 178, 196, 265, 275, 284, 295, 308, 324
 -, bedingte, 196, 266, 285, 295, 308
Nullstelle, 31, 45, 61, 190, 192, 206
Nyquist-Frequenz, 20, 98
 - -Frequenzmatrix, 98

Optimalfilter, 232, 235, 238
 -, diskretes, 235
 -, kontinuierliches, 232
Orthogonalitätsrelation, 233, 236, 242

Parallelstruktur, 48, 82, 144
Parameterschätzung, 8, 70, 84, 132, 148, 245, 300, 320
Periode, 5, 13, 26, 33, 93
Periodogramm, 221, 228, 317
Periodizitätsmatrix, 92
Phase, 5, 52, 72, 118, 191

- Phasenspektrum, 12, 14, 52
 - , diskretes, 12
 - , kontinuierliches, 14
- Phasenverschiebung, 26, 52, 191
- Phasenwinkel, 5, 165
- Poisson-Prozeß, 178
 - -Verteilung, 178
- Polstelle, 29, 31, 45, 61, 206
- Posteriori-Dichte, 300, 320, 322, 325
- Potential, 258, 264, 302
- Prädiktion, 232, 236, 239
- Priori-Dichte, 300, 302, 304, 320, 325
- Produktform der Systemfunktion, 31, 45, 72
- Prozeß, 162, 170, 178, 191, 201, 209, 220, 232, 276, 280, 289, 297, 309
 - , autoregressiver, 191, 201, 204, 238, 241, 276, 279, 289, 309
 - , diskreter, 162, 165, 168, 173, 177, 182, 188, 191, 211, 227
 - , im engeren Sinne stationärer, 166, 178, 209, 216
 - , im weiteren Sinne stationärer, 166, 178, 187, 192, 204, 209, 216, 232
 - , inverser, 190, 192, 204, 280, 297
 - , kausaler, 191, 279, 309
 - , nichtkausaler, 276, 279, 309
 - , schwach stationärer, 166
 - , stationärer, 166, 170, 177, 186, 192, 202, 209, 216, 220, 222, 279
 - , stochastischer, 162, 171, 185
 - , zweidimensionaler, 276, 279
 - mit unabhängigen Zuwächsen, 171, 179
- Pseudolikelihoodverfahren, 322, 326, 330
- Rand, 272
 - , freier, 272
- Randdichte, 164, 181, 323
- Rauschanteil, 191, 202, 237, 241, 284, 289
- Rauschen, 177, 192, 202, 229, 232, 238, 239
 - , weißes, 177, 192, 202, 229, 238, 255, 279
- Realfunktion, 55, 61, 83, 120
- Realisierung, 162, 209, 216, 220, 232, 258, 264, 301
- Realteil, 8, 11, 15
- Rechteckfenster, 88, 155, 158, 224

Rechteckfilter, 79, 144

- , eindimensionales, 79, 144
- , zweidimensionales, 144

Rechtecksignal, 15, 38, 155

- , diskretes, 38
- , kontinuierliches, 15, 155

Remez-Algorithmus, 86

Schätzung, 71, 84, 209, 215, 216, 220, 222, 229, 232, 236, 240, 241, 300, 316, 318, 320

- , beste, 71, 232, 236, 240, 241, 320
- , geglättete, 222, 229, 318

Schur-Cohn-Algorithmus, 89

Schwingung, 5, 26, 100

- , harmonische, 5, 26
- , komplexe, 7, 26, 100
- , n-dimensionale, 100

Serienstruktur, 47

Signal, 1, 3, 4, 7, 12, 13, 18, 20, 33, 92, 94, 105, 161, 169, 188, 190, 201, 232, 239, 249

- , analoges, 4
- , aperiodisches, 12, 94, 169
- , bandbegrenztes, 20
- , deterministisches, 1, 3, 169, 188, 232
- , digitales, 4
- , diskretes, 4, 18, 33, 92, 105
- , eindimensionales, 3, 4, 92, 94
- , komplexes, 7
- , kontinuierliches, 4, 18, 92, 105
- , mehrdimensionales, 3, 92
- , n-dimensionales, 92
- , periodisches, 8, 13, 18, 33, 92
- , stochastisches, 1, 161, 188, 232
- , vektorielles, 3
- , zweidimensionales, 105

Signal-Rausch-Verhältnis, 235

Sobeloperator, 147, 160

Spektraldichtefunktion, 170

spektrale Darstellung, 171, 188, 253

- Verschmierung, 88, 154

- Spektralfenster, 87, 153, 223, 229, 319
 - , zweidimensionales, 153
- Spektrum, 12, 14, 18, 33, 170, 187, 191, 195, 205, 220, 229, 234, 253, 283, 294, 317
 - , diskretes, 12, 18, 33
 - , kontinuierliches, 14, 18, 33
- Sperrbereich, 57, 62, 129
- Stabilität, 25, 99, 107, 116, 186, 190, 291
- Stabilitätsbedingung, 51, 133, 186, 189
- Stationarität, 166, 193, 207, 251
- Stationaritätsbedingung, 190, 207
- Stieltjes-Integral, 171
- Superpositionsprinzip, 23, 98
- System, 23, 26, 49, 67, 98, 106, 116, 120, 137, 148, 158, 185, 189
 - , antikausales, 26, 49, 67
 - , diskretes, 24, 189
 - , kausales, 25, 49, 67, 99, 107, 116
 - , kontinuierliches, 23, 98, 106, 158, 185
 - , lineares, 23, 98, 106, 185, 189
 - , nichtkausales, 26
 - , oktagonales, 120, 144, 148
 - , separierbares, 117
 - , stabiles, 25, 50, 99, 107, 291
 - , verschiebungsinvariantes, 23, 98, 106, 185
 - , vierquadrantensymmetrisches, 120, 139, 148
 - , zweiquadrantensymmetrisches, 119, 137
- Systemfunktion, 30, 45, 67, 71, 101, 112, 189, 192, 204, 283, 291
- Taylor-Karman-Struktur, 252
- Teilfilter, 47, 68, 115, 130
- Teilimpulsantwort, 47, 116
- Teilsystemfunktion, 47, 66, 115
- Texturerkennung, 307, 325
- Theorem von Hammersley und Clifford, 263
 - - Shanks, 116
 - - Wiener-Khintchine, 170, 253
- Tiefpaßfilter, 21, 27, 56, 62, 79, 88, 108, 129, 146, 151, 155
 - , zirkulares, 108, 155
- Toleranzschema, 55

Trend, 201, 239, 245

Tschebyscheff-Approximation, 86

Übergangsbereich, 57, 62, 130

Übergangswahrscheinlichkeit, 183

Übertragungsfunktion, 26, 30

ungerade Funktion, 12, 15, 54

Varianz, 167, 177, 194, 204, 210, 216, 222, 281, 318

Varianzfunktion, 164, 167, 177

Varianzkomponente, 248

Verteilungsfunktion, 163, 180, 251, 256

Verzerrung ("aliasing"), 20

Verzerrung ("bias"), 217, 224, 229

Vierer-Nachbarschaft, 146, 271, 274

Wahrscheinlichkeitsdichte, 164, 251

Wertdiskretisierung, 22

Wiener-Hopf-Gleichung, 240

- - -Integralgleichung, 233, 237

- -Optimalfilter, 233

Yule-Walker-Gleichung, 194, 203, 282, 292

- -, zweidimensionale, 282, 292

- - -Schätzer, 244

Zeitdiskretisierung, 22

Zeitreihe, 162

Zeitzerlegung, 41

zirkulare Blockmatrix, 276, 297

- Matrix, 297

z-Transformation, 29, 35, 101, 109

- , n-dimensionale, 101

Zufallsfeld, 163, 250, 255, 261, 263, 264, 270, 278, 299, 316

- , diskretes, 250, 261, 267, 316

- , homogenes, 251, 316

- , isotropes, 251, 317

- , n-dimensionales, 250

-, vektorielles, 163, 250, 252, 255, 263, 269, 278, 299

Zufallsprozeß, 162, 250

-, eindimensionaler, 163

-, multivariater, 163

-, n-dimensionaler, 163, 250

-, vektorieller, 163

Zufallsvariable, 162, 177, 180, 191, 209, 250, 255, 261, 284, 289

-, diskrete, 162, 182, 262, 303, 322

-, mehrdimensionale, 163, 251

-, stetige, 162

Zustandssumme ("partition function"), 258

Zwanziger-Nachbarschaft, 271

Zweier-Clique, 264, 275, 308

Zwölfer-Nachbarschaft, 271