

(Satz von Wiener-Chintschin). Damit gelten insgesamt folgende Beziehungen:

$$\begin{aligned}
 S_{xx}(\omega) &= \int_{-\infty}^{+\infty} R_{xx}(\tau) e^{-j\omega\tau} d\tau \\
 S_{ux}(\omega) &= \int_{-\infty}^{+\infty} R_{ux}(\tau) e^{-j\omega\tau} d\tau \\
 R_{xx}(\tau) &= \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} S_{xx}(\omega) e^{+j\omega\tau} d(j\omega) \\
 R_{ux}(\tau) &= \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} S_{ux}(\omega) e^{+j\omega\tau} d(j\omega).
 \end{aligned} \quad (3.99)$$

Die Beziehungen zwischen der Autokorrelationsfunktion und dem Autoleistungsdichtespektrum in Gl. (3.99) vereinfachen sich für reelle Frequenzen ω , da beide gerade Funktionen sind, zu

$$\begin{aligned}
 S_{xx}(\omega) &= 2 \int_0^{\infty} R_{xx}(\tau) \cos \omega\tau d\tau \\
 R_{xx}(\tau) &= \frac{1}{\pi} \int_0^{\infty} S_{xx}(\omega) \cos \omega\tau d\omega. \\
 \text{Für die mittlere Signalleistung gilt somit entsprechend Gl. (3.67)} \\
 R_{xx}(0) &= \frac{1}{\pi} \int_0^{\infty} S_{xx}(\omega) d\omega = \overline{x^2(t)}.
 \end{aligned} \quad (3.100) \quad (3.101)$$

3.2.4. Beziehungen zwischen Signal- und Systemmodellen

Die Signalbeschreibungen und -modelle sollen in diesem Abschnitt zur Beschreibung des Systemverhaltens im Zeit- und Frequenzbereich verwendet werden. Als Ausgangspunkt werden die grundsätzlichen Beziehungen für kontinuierliche deterministische Eingang- und Ausgangssignale $u(t)$ und $x(t)$ angegeben. Es gilt entsprechend zu Bild 3.24 für den Zusammenhang

$$\begin{aligned}
 &\text{— im Zeitbereich} \\
 x(t) &= \int_{-\infty}^{+\infty} g(\tau) u(t - \tau) d\tau, \\
 &\text{— im Frequenzbereich} \\
 X(j\omega) &= G(j\omega) U(j\omega),
 \end{aligned} \quad (3.102) \quad (3.103)$$

wobei $g(\tau)$ die Gewichtsfunktion und $G(j\omega)$ den komplexen Frequenzgang bedeuten.

In den folgenden Abschnitten soll, aufbauend auf den Korrelationsfunktionen und den Leistungsdichtespektren, ein Zusammenhang zu den genannten Systemmodellen ermittelt werden. Damit ergibt sich die im Bild 3.25 dargestellte Aufgabe.

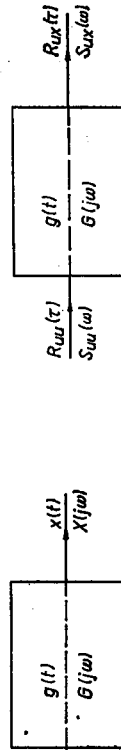


Bild 3.24. Signal- und Systembeschreibungen im deterministischen Fall

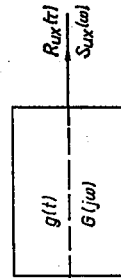


Bild 3.25. Signal- und Systembeschreibungen im stochastischen Fall

3.2.4.1. Beziehungen zwischen den Korrelationsfunktionen $R(\tau)$ und der Gewichtsfunktion $g(t)$

Bei der Ableitung des Zusammenhangs wird von der Berechnungsvorschrift für die Kreuzkorrelationsfunktion entsprechend Gl. (3.49) ausgegangen:

$$R_{ux}(\tau) = \lim_{T_0 \rightarrow \infty} \frac{1}{2T_0} \int_{-T_0}^{+T_0} u(t) x(t + \tau) dt.$$

Mit Gl. (3.52) kann $x(t + \tau)$ ersetzt werden durch

$$x(t + \tau) = \int_{-\infty}^{+\infty} g(\lambda) u(t - \lambda + \tau) d\lambda. \quad (3.104)$$

Damit gilt

$$R_{ux}(\tau) = \lim_{T_0 \rightarrow \infty} \frac{1}{2T_0} \int_{-T_0}^{+T_0} u(t) \int_{-\infty}^{+\infty} g(\lambda) u(t - \lambda + \tau) d\lambda dt. \quad (3.105)$$

Durch Vertauschen der Reihenfolge der Integrationen erhält man aus Gl. (3.105)

$$R_{ux}(\tau) = \int_{-\infty}^{+\infty} g(\lambda) \lim_{T_0 \rightarrow \infty} \frac{1}{2T_0} \int_{-T_0}^{+T_0} u(t) u(t - \lambda + \tau) dt d\lambda$$

und somit die endgültige Beziehung zwischen den Korrelationsfunktionen und der Gewichtsfunktion in der Form

$$R_{ux}(\tau) = \int_{-\infty}^{+\infty} g(\lambda) R_{uu}(\tau - \lambda) d\lambda. \quad (3.106)$$

Diese Beziehung ist eine wesentliche Grundlage für die Ermittlung der Gewichtsfunktion (s. Abschn. 6.2) und stellt eine Analogie zum Faltungsintegral für determinierte Signale von Gl. (3.102) dar.

3.2.4.2. Beziehung zwischen den Leistungsdichtespektren $S(\omega)$ und dem komplexen Frequenzgang $G(j\omega)$

Geht man von dem Zusammenhang zwischen der Kreuzleistungsdichte $S_{ux}(\omega)$ und der Kreuzkorrelationsfunktion $R_{ux}(\tau)$ entsprechend Gl. (3.99)

$$S_{ux}(\omega) = \int_{-\infty}^{+\infty} R_{ux}(\tau) e^{-j\omega\tau} d\tau$$

aus und ersetzt die Kreuzkorrelation durch die Beziehung Gl. (3.106), so gilt

$$S_{ux}(\omega) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(\lambda) R_{uu}(\tau - \lambda) d\lambda e^{-j\omega\tau} d\tau. \quad (3.107)$$

Wird diese Gleichung mit dem Ausdruck

$$e^{-j\omega\lambda} e^{+j\omega\lambda} = 1$$

erweitert, erhält man folgende Beziehung:

$$S_{ux}(\omega) = \int_{-\infty}^{+\infty} g(\lambda) e^{-j\omega\lambda} d\lambda \int_{-\infty}^{+\infty} R_{uu}(\tau - \lambda) e^{-j\omega(\tau - \lambda)} d\tau. \quad (3.108)$$

3.3. Erweiterte Signalmodelle

3.3.1. Grundstrukturen

Die Beschreibung der zeitlichen Verläufe von Signalen durch komplizierte Signalmodelle ist in vielen Gebieten seit Jahrzehnten eine unter der Bezeichnung Zeitreihenanalyse erfolgreich praktizierte Methode [1.50, 3.9, 3.10, 3.13, 3.14, 3.15].

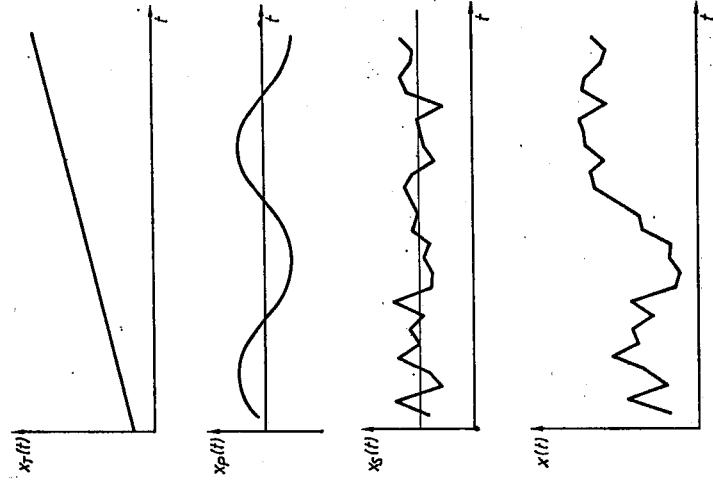


Bild 3.26
Zeitlicher Verlauf eines Signals
und seiner einzelnen Komponenten

Die vielfältig entwickelten Verfahren, besonders auf dem Gebiet der Ökonometrie, finden immer mehr Eingang in technische Bereiche.

In diesem Abschnitt sollen grundlegende Methoden der Analyse von Signalen, die in äquidistanten Abständen erfaßt wurden, vorgestellt werden. Diese zeitlichen Reihenfolgen der Signale werden auch als Zeitreihen bezeichnet. Als Signalmodelle werden mathematische Beschreibungen der deterministischen (determinierte Signalmodelle) bzw. der stochastischen (stochastische Signalmodelle) Signalverläufe angesehen. Die Signalmodelle können z.B. zur Interpretation, zur Filterung und zur Vorhersage der Signalverläufe verwendet werden. Es wird angenommen, daß ein Signal $x(t)$ (Bild 3.26) folgende Komponenten besitzen kann:

$$x(t) = x_T(t) + x_F(t) + x_S(t); \quad (3.112)$$

$x_T(t)$ Signalanteil, der durch einen Polynomansatz beschreibbar ist
 $x_F(t)$ Signalanteil, der durch einen periodischen Ansatz beschreibbar ist
 $x_S(t)$ Signalanteil, der nur durch ein stochastisches Signalmodell beschreibbar ist.

Die Fourier-Transformierte der Gewichtsfunktion ist der komplexe Frequenzgang $G(j\omega)$ und die der Autokorrelationsfunktion entsprechend Gl. (3.99) das Autoleistungsspektrum $S_{xx}(\omega)$. Damit gilt

$$S_{xx}(\omega) = G(j\omega) S_{xx}(\omega). \quad (3.109)$$

Gl. (3.109) ist eine wichtige Bestimmungsgleichung für den komplexen Frequenzgang aus den Leistungsspektren und der Ausgangspunkt für viele weitere Beziehungen für den Entwurf von Filtern und von Steuerungen für stochastisch gestörte Systeme (Tafel 3.5). Eine weitere wichtige Beziehung zur Ermittlung des Amplitudenfrequenzgangs kann aus der Definition des Autoleistungsspektrums abgeleitet werden. Entsprechend Gl. (3.65) gilt

$$S_{xx}(\omega) = \lim_{T_0 \rightarrow \infty} \frac{X_{T_0}(j\omega) X_{T_0}^*(-j\omega)}{2T_0}.$$

Unter Verwendung der bekannten Zusammenhänge von Gl. (3.103) kann diese Beziehung umgeformt werden in

$$S_{xx}(\omega) = G(j\omega) G^*(-j\omega) \lim_{T_0 \rightarrow \infty} \frac{U_{T_0}(j\omega) U_{T_0}^*(-j\omega)}{2T_0}. \quad (3.110)$$

Damit erhält man als zweite Grundbeziehung

$$S_{xx}(\omega) = |G(j\omega)|^2 S_{uu}(\omega). \quad (3.111)$$

Das Leistungsdichtespektrum des Ausgangssignals ist also durch das „Betragsquadrat“ des komplexen Frequenzgangs und das Leistungsdichtespektrum des Eingangssignals bestimmt. Ausgewählte Anwendungen sind wieder in Tafel 3.5 enthalten.

Tafel 3.5. Beziehungen zwischen Leistungsdichtespektren einfacher Strukturen

Struktur	Beziehungen
	$S_{xx}(\omega) = G(j\omega) ^2 S_{u_1 u_1}(\omega) + G(j\omega) ^2 S_{u_2 u_2}(\omega)$ mit $R_{u_1 u_2}(\tau) = 0$ für alle τ
	$S_{xx}(\omega) = G_1(j\omega) ^2 S_{u_1 u_1}(\omega) + G_2(j\omega) ^2 S_{u_2 u_2}(\omega)$ $(\pm) [G_1(j\omega) G_2(-j\omega) S_{u_1 u_2}(\omega) + G_2(j\omega) G_1(-j\omega) S_{u_2 u_1}(\omega)]$
	$S_{x-x}(\omega) = \left \frac{G_1 G_R}{1 + G_1 G_R} \right ^2 S_{u u}(\omega) + \left \frac{1}{1 + G_1 G_R} \right ^2 S_{xx}(\omega)$ mit $R_{u u}(\tau) = 0$ für alle τ $\overline{x^2 u}(\tau) = \frac{1}{\pi} \int_0^\infty S_{x-x}(\omega) d\omega$

Grundsätzlich wird bei allen Modellvorstellungen davon ausgegangen, daß das beobachtete Signal $x(t)$ über sog. „erzeugende Funktionen“ $f(t)$ bei deterministischer Beschreibung und aus einem Rauschsignal $\varepsilon(t)$ über eine Filterung bei der stochastischen Beschreibung entsteht.

Als deterministische Signalmodelle werden hier solche angesehen, die das Signal durch Polynome und/oder periodische Funktionen beschreiben. Damit kann ausgehend vom allgemeinen Ansatz

$$x(t) = f(t) + z(t) \quad (3.113)$$

für die erzeugenden Funktionen $f(t)$ formuliert werden:

– Polynomansatz

$$f(t) = \sum_{i=0}^m a_i t^i,$$

– periodischer Ansatz

$$f(t) = a_0 + \sum_{i=1}^l c_i (\sin \omega_i t + \varphi_i).$$

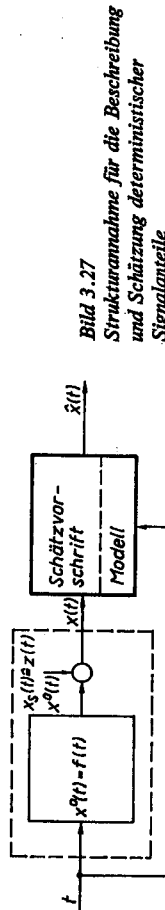


Bild 3.27
Strukturannahme für die Beschreibung
und Schätzung deterministischer
Signalanteile

Im Bild 3.27 ist die angenommene Struktur für die Entstehung eines determinierten Signals $x(t)$ als Funktion der Zeit t dargestellt. Aus dem durch das stochastische Signal $z(t)$ gestörten Signal $x^0(t)$ ist mit Hilfe einer Schätzvorschrift ein Signalmodell durch Minimierung einer Gütefunktion zu ermitteln:

$$Q = f(x(t), \hat{x}(t)).$$

Beim Entwurf stochastischer Signalmodelle kann im einfachsten Fall von der Annahme ausgegangen werden, daß aus einem Signal $\varepsilon(t)$ mit der Eigenschaft des „weißen Rauschens“, d. h.

$$E\{\varepsilon(t)\varepsilon(t+\tau)\} = \begin{cases} \sigma_\varepsilon^2 & \text{für } \tau = 0 \\ 0 & \text{sonst} \end{cases} \quad (3.114)$$

und

$$E\{\varepsilon(t)\} = 0,$$

über ein lineares Filter das korrelierte stochastische Signal $x(t)$ erzeugt wird (Bild 3.28). Als Beschreibungsgleichung wird die lineare stochastische Differentialgleichung in der Form

$$\sum_{j=0}^m c_j x^{(j)}(t) = \sum_{i=1}^l d_i \varepsilon^i(t) \quad (3.115)$$

mit

$$x^{(j)}(t) = d^j x(t)/dt^j, \quad \varepsilon^i(t) = d^i \varepsilon(t)/dt^i$$

allgemein verwendet. Aus dem beobachteten Signal $x(t)$ ist mit Hilfe von Schätzverfahren ein stochastisches Signalmodell so zu entwickeln, daß wiederum eine Gütefunktion Q

ein Minimum wird. Damit wird die Ermittlung der Parameter des linearen Filters, das die beobachteten Signalwerte von $x(t)$ aus dem Rauschsignal $\varepsilon(t)$ erzeugt, zur Aufgabstellung der Theorie linearer stochastischer Prozesse [3.9, 3.10].

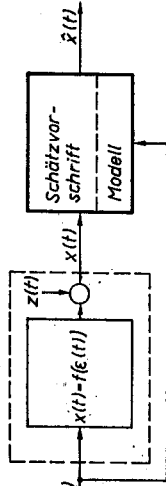


Bild 3.28
Strukturannahme für die Beschreibung
und Schätzung stochastischer Signalanteile

3.3.2. Erweiterte deterministische Signalmodelle

3.3.2.1. Deterministische Signalmodelle mit Polynomansatz

Eine der gebräuchlichsten und universell einsetzbaren analytischen Beschreibungsformen von äquidistant abgetasteten Signalwerten, d. h. der Zeitreihe $x(kT)$; $k = 1, 2, \dots, n$, stellt der folgende Polynomansatz dar:

$$x(kT) = a_0 + \sum_{i=1}^m a_i (kT)^i + z(kT); \quad k = 1, 2, \dots, n. \quad (3.116)$$

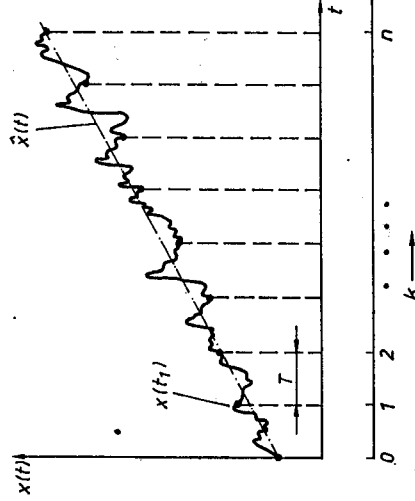


Bild 3.29
Verlauf des gemessenen Signals $x(t)$
und der Schätzung $\hat{x}(t)$

Dabei bedeuten n die Anzahl der beobachteten Meßpunkte und m die Ordnung des Polynomansatzes. Für eine einfache Struktur ist das Signal $x(t)$ und die abgetastete Folge $x(kT)$ im Bild 3.29 dargestellt. Für das Signalmodell kann entsprechend Gl. (3.116) die Beziehung

$$\hat{x}(kT) = \hat{a}_0 + \sum_{i=1}^m \hat{a}_i (kT)^i; \quad k = 1, 2, \dots, n, \quad (3.117)$$

angenommen werden. Zur Vereinfachung der Schreibweise wird bei allen deterministischen Signalmodellen für $x_T(kT) \cong x(kT)$ bzw. $x_T(kT) \cong x(kT)$ gesetzt. In dem Modell sind die Ordnung m und die Modellparameter $\hat{a}_i$ zu schätzen. Die Modellordnung wird häufig nach folgender Strategie ermittelt:

1. Start mit der Ordnung $m = 1$
2. Schätzung der Parameter $\hat{a}_i$

3. Test auf Stationarität des Fehlers $e(kT) = x(kT) - \hat{x}(kT)$ und der Korrelationsfunktion $R_{ee}(\tau)$ auf weißes Rauschen
4. wenn die Tests negativ sind, wird die Modellordnung um 1 erhöht und Schritt 2 realisiert; sonst Abbruch.

Als eine der robustesten Methoden für die Parameterschätzung hat sich die Methode der kleinsten Fehlerquadrate, die *Regression*, auch bei diesem Modelltyp bewährt. Bei der direkten Methode der Regression (s. Abschn. 5.2.2) wird von der Fehlergleichung

$$e(kT) = [x(kT) - \hat{x}(kT)] \quad (3.118)$$

ausgegangen.

Für n Beobachtungen des Signals $x(t)$ zu den Zeitpunkten $k = 1, 2, \dots, n$ erhält man für die Parameter des Modells von Gl. (3.117) folgende Schätzgleichung:

$$\hat{a} = [M^T M]^{-1} M^T x \quad (3.119)$$

mit

$$\begin{aligned} \hat{a}^T &= [\hat{a}_0, \hat{a}_1, \dots, \hat{a}_m] \\ x^T &= [x(1), \dots, x(nT)] \\ M &= \begin{bmatrix} 1 & T & \dots & T^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & nT & \dots & (nT)^m \end{bmatrix} \end{aligned}$$

Als Güterwerte für die Schätzung können berechnet werden

$$\begin{aligned} \text{Reststreuung} \quad s_R^2 &= \frac{1}{n - (m + 2)} \sum_{k=1}^n [x(kT) - \hat{x}(kT)]^2, \quad (3.120) \\ \text{Varianz} \quad \text{var} \{ \hat{a}_i \} &= p_{ii} s_R^2, \\ \text{Kovarianz} \quad \text{cov} \{ \hat{a}_i, \hat{a}_j \} &= p_{ij} s_R^2 \end{aligned}$$

mit der Präzisionsmatrix $P = [M^T M]^{-1}$,

- Konfidenzintervall für die Parameter

$$|a_i - \hat{a}_i| = t_{\alpha} s_R \sqrt{p_{ii}},$$

- Hypothesentest $H_0: a_i = 0$

$$|\hat{a}_i| \geq t_{\alpha} s_R \sqrt{p_{ii}}.$$

Der rechentechnische Aufwand für die Berechnung der Präzisionsmatrix P kann durch die Verwendung der Eulerschen Summenformeln für endliche Folgen wesentlich reduziert werden [3.16]. Ferner muß beachtet werden, daß bei einer hohen Modellordnung und immer größer werdendem Beobachtungsumfang n bei Rechnern mit geringen Rechengenauigkeiten Probleme entstehen. Deshalb ist es häufig günstig, mit einem festen „Beobachtungsfenster“ eine gleitende Schätzung der Modellparameter nach der im

Bild 3.30 angegebenen Strategie vorzunehmen. Diese Vorgehensweise hat die Vorteile einer wesentlichen Reduktion der Rechenzeit – die Präzisionsmatrix wird nur einmal für das gewählte Fenster berechnet – und der numerischen Stabilität. Nachteilig ist die Unsicherheit über den wirklichen Modellansatz, da nur das Beobachtungsfenster zur Verfügung steht. Dadurch sind Vorhersagen über Zeiträume, die die Länge des Beobachtungsfensters überschreiten, in der Regel mit großen Unsicherheiten behaftet. Die Schätzung der Parameter von Polynomansätzen 1. und 2. Ordnung und die trendkorrigierten Signalverläufe werden für ein gewähltes Testbeispiel im Beispiel 3.4 gezeigt.

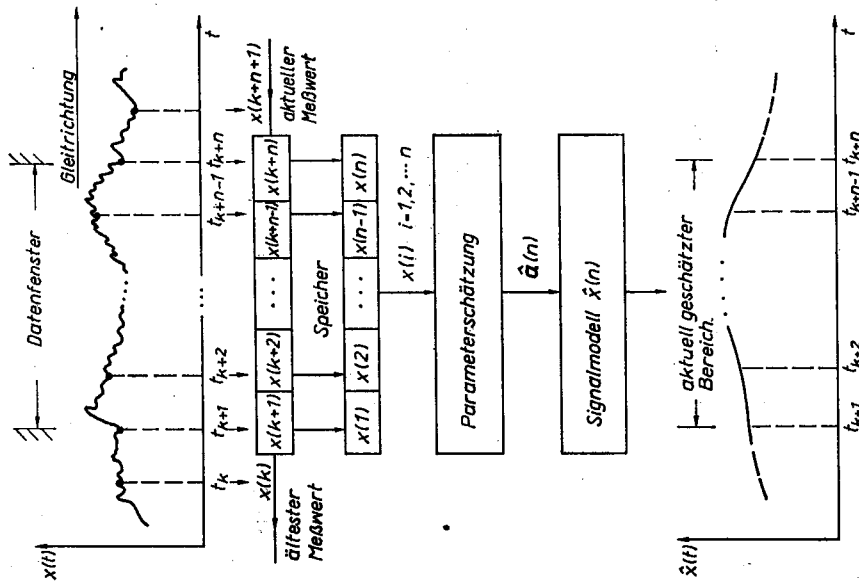
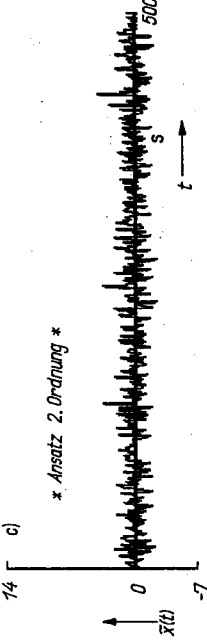
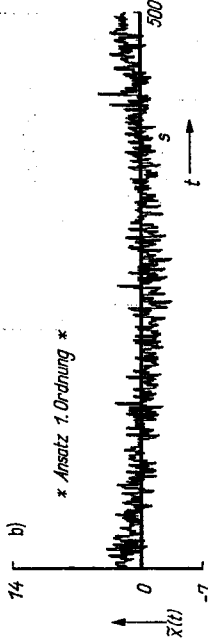


Bild 3.30
Prinzip
der gleitenden Schätzung

Eine weitere Möglichkeit der Echtzeitparameterschätzung ist neben der gleitenden Schätzung die Anwendung von indirekten Schätzverfahren. Durch sie wird es möglich, die Modellparameter aktuell nachzuführen und mathematisch schwer erfassbare zeitliche Verläufe durch einfache zeitvariante Modellansätze anzunähern. Als eine der leistungsstärksten Methoden hat sich die rekursive Regression herausgestellt. Entsprechend den im Abschnitt 5.4.3 durchgeführten allgemeinen Ableitungen dieses Schätzverfahrens gelten für den Modellansatz Gl. (3.117) die Schätzgleichungen der rekursiven Regression:

$$\hat{a}(k+1) = \hat{a}(k) + k(k+1)[x(k+1) - m^T(k+1)\hat{a}(k)] \quad (3.121)$$

Beispiel 3.4	Ermittlung von Signalmodellen mit Polynomansätzen
Gegeben:	Signalgleichung $x(t) = a_0 + a_1 t + a_2 t^2 + z(t),$ $a_0 = -5, \quad a_1 = 1 \cdot 10^{-2} \text{ s}^{-1}, \quad a_2 = 5 \cdot 10^{-5} (\text{s}^{-1})^2,$ $z(t)$ weißes Rauschen, NV (0, 1), Abtastzeit $T = 1 \text{ s}$, Beobachtungen: 500.
Gesucht:	Geschätzte Parameter für die Polynomansätze 1. und 2. Ordnung sowie das jeweils trendkorrigierte Signal $\hat{x}(kT) = x(kT) - \hat{x}(kT)$.
Ergebnis:	Als Schätzwerte aus der Stichprobe ergeben sich - für den Ansatz 1. Ordnung $\hat{a}_0 = -7,06, \quad \hat{a}_1 = 0,035 \text{ s}^{-1},$ - für den Ansatz 2. Ordnung $\hat{a}_0 = -5,05, \quad \hat{a}_1 = 1,0 \cdot 10^{-2} \text{ s}^{-1},$ $\hat{a}_2 = 4,79 \cdot 10^{-5} (\text{s}^{-1})^2.$ Die mit diesen Parametern korrigierten Signalwerte sind im Bild zu Beispiel 3.4 dargestellt.



Die Signalverläufe zeigen, daß bei der Wahl eines Ansatzes 1. Ordnung der Trend nicht eliminiert werden kann. Erst die Erhöhung der Ordnung auf $m = 2$ führt dazu, daß das korrigierte Signal $\hat{x}(kT)$ trendfrei und damit stationär wird.

mit

$$k(k+1) = P(k) m(k+1) [m^T(k+1) P(k) m(k+1) + 1]^{-1}$$

$$m^T(k+1) = [1, (k+1)T, \dots, ((k+1)T)^m]$$

$$P(k) = P(k-1) - k(k) m^T(k) P(k-1). \quad (3.122)$$

Entsprechend den Ausführungen im Abschnitt 5.4.3 ist der in Gl. (3.122) dargestellte Schätzalgorithmus in der Lage, bei geeigneter Wahl der Startmatrix $P(0)$ die Modellparameter anzupassen; er kann sie aber nicht nachführen. Erst durch die Einführung einer Wichtung der Matrix P entsprechend den Beziehungen

$$a) \quad P^*(k) = P(k) + P'; \quad P' = c_1 I, \quad 10^{-2} \leq c_1 \leq 10^2, \quad (3.123)$$

$$b) \quad P^*(k) = \frac{1}{c_2} P(k); \quad 0 < c_2 < 1$$

wird die Nachführung zeitvarianter Modellparameter durch „Vergessen“ zurückliegender Werte möglich. Die Wahl der Start- und Wichtungsparameter für den Schätzalgorithmus der gewichteten rekursiven Regression erfolgt auch für diesen Modelltyp nach den Richtlinien im Abschnitt 5.4.3.

Mit dem geschätzten Signalmodell in Form eines Polynomansatzes können folgende Aufgaben gelöst werden:

- Glättung des Signals $x(t)$,
- Beseitigung tieffrequenter Anteile (Trend) des Signals $x(t)$.

3.3.2.2. Deterministische Signalmodelle mit periodischem Ansatz

Viele Signalverläufe in technischen und nichttechnischen Prozessen enthalten dominierende periodische Komponenten. Obwohl die Beschreibung solcher Signale auch prinzipiell durch die im Abschnitt 3.3.2.1 vorgestellten Signalmodelle mit Polynomansatz erfolgen kann, ist die Verwendung periodischer Ansätze aus Aufwandsgründen (Modellansatz, Rechenaufwand) und häufig wegen der Interpretierbarkeit des Modells angebracht. Die große Bedeutung dieses Modelltyps und die Schwierigkeiten seiner Schätzung hatten zur Folge, daß eine Vielzahl von Verfahren in den letzten Jahrzehnten entwickelt wurde [3.9, 3.10, 3.17, 3.18].

Im Rahmen dieses Buches werden solche Strategien ausgewählt, die einerseits das Grundproblem darstellen und andererseits den heutigen Stand der Rechentchnik berücksichtigen.

Es wird in diesem Fall entsprechend Gl. (3.113) davon ausgegangen, daß das Signal $x(t)$ durch eine determinierte periodische Funktion $f(t)$ und die Störung $z(t)$ erzeugt wird. Mit Hilfe des Additionstheorems

$$\sin(x+y) = \sin x \cos y + \cos x \sin y$$

können die unter der Summe von Gl. (3.113) stehenden Funktionen umgeformt werden in

$$c_1 \sin(\omega_1 t + \varphi_1) = a_1 \cos \omega_1 t + b_1 \sin \omega_1 t$$

$$a_1 = c_1 \sin \varphi_1, \quad b_1 = c_1 \cos \varphi_1 \quad (3.124)$$

$$c_1 = \sqrt{a_1^2 + b_1^2}, \quad \tan \varphi_1 = a_1/b_1.$$

Werden die Beziehungen von Gl. (3.124) in die Gl. (3.113) eingesetzt, erhält man als Signalbeschreibung in den Tastpunkten kT

$$x(kT) = a_0 + \sum_{i=1}^l [a_i \cos \omega_i(kT) + b_i \sin \omega_i(kT)] + z(kT); \quad (3.125)$$

$$k = 1, 2, \dots, n, \quad i = 1, 2, \dots, l,$$

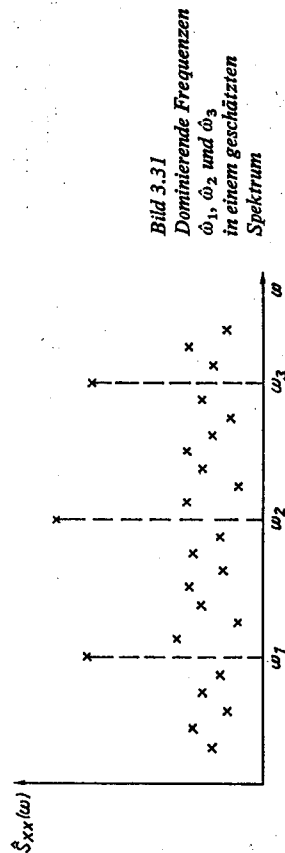
und als Modellgleichung gilt analog

$$\hat{x}(kT) = \hat{a}_0 + \sum_{i=1}^l \hat{a}_i \cos \hat{\omega}_i(kT) + \hat{b}_i \sin \hat{\omega}_i(kT). \quad (3.126)$$

Als Schätzaufgabe für das Signalmodell ist die Ermittlung der Frequenzen $\hat{\omega}_i$ und der Parameter $\hat{a}_0, \hat{a}_i, \hat{b}_i$ auf der Grundlage von n Beobachtungen zu lösen. Da in Gl. (3.126) die Frequenzen $\hat{\omega}_i$ nichtlinear mit der sog. unabhängigen Variablen der Zeit zusammenhängen, ist ihre Ermittlung im Gegensatz zu den Amplitudenparametern $\hat{a}_0, \hat{a}_i, \hat{b}_i$ mittels der Regression i. allg. nicht möglich. Einen Ausweg bildet folgende Strategie:

1. Bestimmung geeigneter Frequenzen $\hat{\omega}_i$,
2. Schätzung der zu den Frequenzen $\hat{\omega}_i$ gehörenden Parameter $\hat{a}_0, \hat{a}_i, \hat{b}_i$,
3. Überprüfung und Anpassung des Modells.

Dabei verschmelzen häufig die Stufen 2 und 3.



Schritt 1. Ermittlung der dominierenden Frequenzen in $x(t)$ mit Methoden der Spektralanalyse.

Ausgangspunkt für die Ermittlung dominierender Frequenzen im Signal $x(t)$ sind die Schätzung des Autoleistungsdichtespektrums $\hat{S}_{xx}(\omega)$ und seine Auswertung. Die Schätzmethoden sind im Abschnitt 3.2.3.4 angegeben. Es werden die Frequenzen $\hat{\omega}_i$ als dominierend angesehen, die sich in ihrem Leistungsanteil wesentlich von den anderen unterscheiden (Bild 3.31). Bei der Anwendung der im Abschnitt 3.2.3.4 vorgestellten Methoden ist zu beachten, daß die vorhandene Datenmenge und die gewählte Abtastzeit die möglichen Frequenzen $\hat{\omega}_i$ beschränken. Außerdem bewirkt die Einführung der Wichtungsfenster bei der Berechnung des Spektrums über die Korrelationsfunktionen zwar eine Glättung des Spektrums, aber damit auch eine Reduzierung der Signifikanz von dominierenden Frequenzen. Somit gilt, daß

- eine Ermittlung sehr eng begrenzter Frequenzbereiche für den Ansatz von Gl. (3.126) nur gesichert ist, wenn eine Variation des Meßwertumfangs die Frequenzen bestätigt,
- die Ermittlung ganz spezieller Frequenzen nur in Ausnahmefällen möglich ist, da $\omega_i = i\omega_0; i = 0, 1, \dots, n/2$ (ω_0 Grundwelle) gilt.

Gegeben:

Signalgleichung

$$x(t) = a_0 + a_1 \cos \frac{2\pi}{125} t + b_1 \sin \frac{2\pi}{125} t + z(t),$$

$$a_0 = 0,15, \quad a_1 = 0,25, \quad b_1 = 0,25,$$

 $z(t)$ weißes Rauschen, NV(0, 1),
Abtastzeit $T = 1$ s,

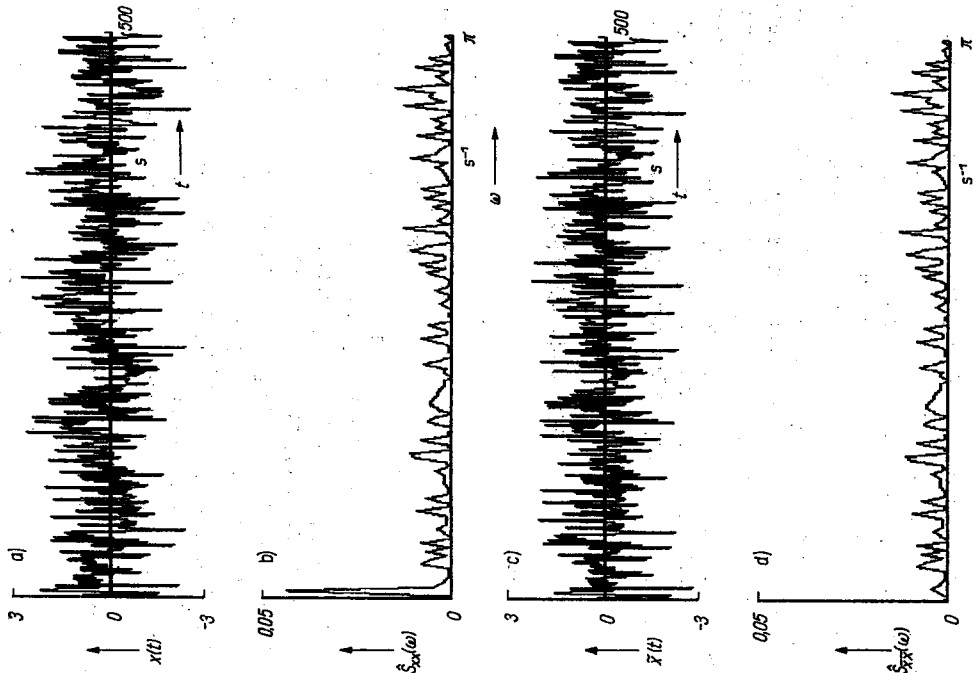
Beobachtungen: 500.

Gesucht:Schätzwerte für die Frequenz $\hat{\omega}_1$ und die Parameter $\hat{a}_1, \hat{b}_1$.**Ergebnis:**

Aus der Spektralanalyse wird im ersten Schritt die Frequenz $\hat{\omega}_1$ zu $\hat{\omega}_1 = 0,04884 \text{ s}^{-1}$ ($\omega_1 = 0,05024 \text{ s}^{-1}$) ermittelt. Die mit dieser Frequenz durchgeführte Schätzung der Parameter a_1 und b_1 entsprechend Gl. (3.128) ergibt das Resultat

$$\hat{a}_0 = 0,199, \quad \hat{a}_1 = 0,18, \quad \hat{b}_1 = 0,40.$$

Eine vorgenommene Signalkorrektur in der Form $\hat{x}(kT) = x(kT) - \hat{x}(kT)$ zeigt sowohl im zeitlichen Verlauf von $x(kT)$ als auch im Leistungsspektrum des korrigierten Signals $\hat{S}_{xx}(\omega)$, daß die dominierende periodische Komponente durch das Signalmodell sehr gut abgebildet wird (Bild zu Beispiel 3.5).



Schritt 2. Schätzung der Parameter $\hat{a}_l$ und $\hat{b}_l$ mit der Methode der Regression bei Verwendung dominierender Frequenzen $\hat{\omega}_l$.

Unter Annahme der Signalbeschreibungen für $x(kT)$ und $\hat{x}(kT)$ in den Gln. (3.125) und (3.126) kann für das Verfahren der Regression für n Beobachtungen der Fehlervektor

$$e(kT) = [x(kT) - \hat{x}(kT)] \quad (3.127)$$

gebildet werden. Als Schätzvorschrift ergibt sich analog den Abschnitten 5.2.2 und 3.2.2.1 für n Beobachtungen ($k = 1, 2, \dots, n$)

$$\hat{s} = [M^T M]^{-1} M^T x \quad (3.128)$$

mit

$$s^T = [\hat{a}_0, \hat{a}_1, \dots, \hat{a}_l, \hat{b}_1, \dots, \hat{b}_l]$$

$$x^T = [x(T), \dots, x(nT)]$$

$$M = \begin{bmatrix} 1 & \cos \hat{\omega}_1 T & \dots & \cos \hat{\omega}_l T & \sin \hat{\omega}_1 T & \dots & \sin \hat{\omega}_l T \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & \cos \hat{\omega}_1(nT) & \dots & \cos \hat{\omega}_l(nT) & \sin \hat{\omega}_1(nT) & \dots & \sin \hat{\omega}_l(nT) \end{bmatrix}.$$

Die Gütwerte der Schätzung werden analog den Beziehungen von Gl. (3.120) berechnet. Als Startparameter für die Anzahl l der Frequenzen sollte $1 \leq l \leq 3$ gewählt werden.

Der notwendige Rechenaufwand von Gl. (3.128) steigt bei einer Vielzahl von dominierenden Frequenzen und großen Meßwertumfängen enorm an. Eine Reduzierung des Aufwands ist durch die von Anderson [3.10] vorgeschlagenen Vereinfachungen für die Elemente der P -Matrix möglich. Ferner sollte beachtet werden, daß eine zu enge Staffe- lung der dominierenden Frequenzen zu einer schlecht konditionierten P -Matrix aufgrund der eng benachbarten periodischen Funktionen führt. Auf eine weitere Reduzierung des Rechenaufwands bei gleichzeitiger Vereinfachung des Modellansatzes von Gl. (3.126) durch ein Orthogonalsystem der periodischen Funktionen sei an dieser Stelle nur verwiesen [3.10]. Für einen einfachen Test ist im Beispiel 3.5 die in diesem Abschnitt dargestellte Strategie dargestellt.

3.3.3. Erweiterte stochastische Signalmodelle

3.3.3.1. Grundsätze

Neben der Beschreibung von Meßwertfolgen durch deterministische Signalmodelle (s. Abschn. 3.3.2) gewinnt in den letzten Jahren die Beschreibung des Verlaufs und der inneren Zusammenhänge von Signalen auf der Basis der Theorie linearer stochastischer Prozesse durch stochastische Signalmodelle an Bedeutung [2.9, 3.10, 3.11, 3.18, 3.19]. Diese Beschreibungsform ist besonders in der Ökonomie, der Medizin und der Meteorologie verbreitet. Im Bereich der Technik ist ihre Anwendung noch nicht so ausgeprägt.

Ausgangspunkt ist die im Abschnitt 3.3.1 bereits dargelegte Annahme, daß das stochastische Signal $x(t)$ über ein Filter mit den Übertragungseigenschaften $G_F(j\omega)$ aus dem stationären „weißen Rauschen“ $\varepsilon(t)$ erzeugt wird; s. Bild 3.28 und Gl. (3.114). Für den Fall äquidistanter zeitlicher Werte von $\varepsilon(t)$ und $x(t)$ erhält man als allgemeine Beschreibung mit $\varepsilon(kT) \triangleq \varepsilon(k)$ und $x(kT) \triangleq x(k)$ anstelle der stochastischen Differential-

gleichung von Gl. (3.114) die stochastische Differenzengleichung

$$a_0 x(k) + a_1 x(k-1) + \dots + a_r x(k-r) = b_0 \varepsilon(k) + \dots + b_l \varepsilon(k-l). \quad (3.129)$$

Mit Verwendung eines Verschiebeoperators q^{-1} kann Gl. (3.129) durch die Vereinbarung $x(k-n) \triangleq x(k) q^{-n}$ umgeformt werden in

$$A(q^{-1}) x(k) = B(q^{-1}) \varepsilon(k) \quad (3.130)$$

mit den Summenverschiebeoperatoren

$$A(q^{-1}) = [a_0 + a_1 q^{-1} + \dots + a_r q^{-r}]$$

$$B(q^{-1}) = [b_0 + b_1 q^{-1} + \dots + b_l q^{-l}].$$

Die Gl. (3.130) beschreibt einen autoregressiven Prozeß der gleitenden Mittel (autoregressive moving average model oder ARMA-Modell) der Ordnung (r, l) . Aus dieser Beschreibung kann ferner

– das autoregressive Modell (AR-Modell) der Ordnung (r)

$$A(q^{-1}) x(k) = \varepsilon(k), \quad (3.131)$$

– das Modell der gleitenden Mittel (MA-Modell) der Ordnung (l)

$$x(k) = B(q^{-1}) \varepsilon(k) \quad (3.132)$$

abgeleitet werden.

Neben diesen drei Grundtypen werden für die Beschreibung instationärer stochastischer Prozesse Modelle verwendet, die durch eine Differenzbildung der gemessenen benachbarten Signalwerte den instationären Signalanteil (Trend) mit erfassen. Der so abgeleitete autoregressive integrierende Prozeß der gleitenden Mittel der Ordnung (r, d, m) (ARIMA-Modell) wird durch folgende Gleichung beschrieben:

$$A(q^{-1}) \nabla^d x(k) = B(q^{-1}) \varepsilon(k); \quad (3.133)$$

$\nabla^d x(k)$ d -te Differenz der benachbarten Zeitreihenwerte.

Seine Anwendung ist besonders bei kurzen Zeitreihen vorteilhaft, und die Ordnung d der Differenzbildung wird häufig auf ≤ 2 beschränkt.

Ein weiteres stochastisches Signalmodell ergibt sich aus der Verallgemeinerung des ARMA-Modells durch Hinzufügen eines meßbaren Eingangssignals $u(k)$ als Hilfsgröße. Mit den Vereinbarungen für die Systembeschreibung in den Abschnitten 1.4 und 6.3 gilt für diesen autoregressiven Prozeß der gleitenden Mittel mit der Hilfsgröße $u(k)$ der Ordnung (r, m, n) (ARMAX-Modell, autoregressive moving average with exogenous variable) folgende Beziehung:

$$A'(q^{-1}) x(k) = B(q^{-1}) \varepsilon(k) + C(q^{-1}) u(k) \quad (3.133a)$$

mit

$$C(q^{-1}) = [c_0 + c_1 q^{-1} + \dots + c_n q^{-n}].$$

Das ARMAX-Modell, das häufig auch als gesteuertes ARMA-Modell (CARMA-Modell) bezeichnet wird, ist als ein Spezialfall in dem von Aström und Bohlin [1.38] vorgeschlagenen verallgemeinerten Systemmodell enthalten (s. Abschn. 6.3) und soll deshalb an dieser Stelle nicht weiter betrachtet werden.

Im folgenden wird deshalb nur das ARMA-Modell mit seinen Grundtypen, dem AR-Modell und dem MA-Modell, vorgestellt.

3.3.3.2. Autoregressive Modelle (AR-Modell)

Autoregressive Signalmodelle werden vor allem zur Vorhersage des Signalverlaufs (Extrapolation) aus den zurückliegenden Signalwerten verwendet. Aus Gl. (3.130) ergibt sich für $a_0 = 1$

$$x(k) = A'(q^{-1})x(k) + \varepsilon(k) \quad (3.134)$$

mit

$$A'(q^{-1}) = [-a_1 q^{-1}, -\dots, -a_p q^{-p}].$$

Der aktuelle Signalwert $x(k)$ ergibt sich hier aus der bewerteten Signalvergangenheit $A'(q^{-1})x(k)$ und der Zufallsgröße $\varepsilon(k)$. Zur Schätzung der Parameter a_i des AR-Modells wird unter Verwendung der Methode der Regression für n Beobachtungen die Fehlergleichung (s. Abschn. 5.2.2)

$$\hat{\varepsilon}(k) = x(k) - \hat{A}'(q^{-1})x(k) \quad (3.135)$$

mit

$$\hat{\varepsilon}^T(k) = [\hat{\varepsilon}(k), \dots, \hat{\varepsilon}(k+n)]$$

$$x^T(k) = [x(k), \dots, x(k+n)]$$

$$\hat{A}'(q^{-1})x(k) = \begin{bmatrix} -\hat{a}_1 x(k-1) & -\dots & -\hat{a}_p x(k-p) \\ \vdots & & \\ -\hat{a}_1 x(k-1+n) & -\dots & -\hat{a}_p x(k-p+n) \end{bmatrix}$$

benutzt. Damit ergibt sich die Schätzvorschrift für die Parameter zu

$$\hat{a} = [X^T X]^{-1} X^T x \quad (3.136)$$

mit

$$\hat{a}^T = [-\hat{a}_1, \dots, -\hat{a}_p]$$

$$x^T = [x(k), \dots, x(k+n)]$$

$$X = \begin{bmatrix} x(k-1) & \dots & x(k-p) \\ \vdots & & \vdots \\ x(k+n-1) & \dots & x(k+n-p) \end{bmatrix}$$

Die Vorhersage eines Signalwerts um einen Zeittakt T erfolgt dann entsprechend der Beziehung

$$\hat{x}(k+1) = +\hat{a}_1 x(k) + \dots + \hat{a}_p x(k-p). \quad (3.137)$$

Eine Überprüfung des Modells kann mittels der Gütewerte von Gl. (3.120) oder über die Berechnung der Autokorrelationsfunktion $R_{xx}(mT)$, die im Idealfall den Bedingungen von Gl. (3.114) entsprechen sollte, erfolgen.

Anhand der ermittelten Koeffizienten $\hat{a}_i$ kann die Stationarität des Signals sehr einfach dadurch geprüft werden, daß die Lösungen der Gleichung

$$1 + \hat{a}_1 q^{-1} + \dots + \hat{a}_p q^{-p} = 0$$

außerhalb des Einheitskreises liegen müssen [3.2, 3.9]. Natürlich können die Parameter des AR-Modells auch mittels der rekursiven Regression geschätzt werden; s. Gl. (3.121).

Für dieses Modell gilt die rekursive Schätzvorschrift

$$\hat{a}(k+1) = \hat{a}(k) + k(k+1)[x(k+1) - m^T(k+1)\hat{a}(k)] \quad (3.138)$$

mit

$$m^T(k+1) = [x(k), \dots, x(k-r)], \quad k(k+1) \text{ s. Gl. (3.121).}$$

Alle anderen Beziehungen und Startwerte sind wie im Abschnitt 3.3.2.1 zu wählen.

Neben dem Signalmodell von Gl. (3.134) sind die Korrelationsfunktion und das Leistungsdichtespektrum des autoregressiven Prozesses zur Signalanalyse von großer Bedeutung.

Die Korrelationsfunktion $R_{xx}(mT)$ für das stationäre abgetastete Signal $x(k)$ erhält man durch Multiplikation der Gleichung

$$x(k) + a_1 x(k-1) + \dots + a_p x(k-p) = \varepsilon(k) \quad (3.139)$$

mit den Signalwerten $x(k-m)$ und Bildung der Erwartungswerte

$$\begin{aligned} x(k)x(k-m) + a_1 x(k-1)x(k-m) + \dots + a_p x(k-p)x(k-m) \\ = \varepsilon(k)x(k-m), \end{aligned}$$

$$\begin{aligned} E\{x(k)x(k-m)\} + a_1 E\{x(k-1)x(k-m)\} + \dots \\ + a_p E\{x(k-p)x(k-m)\} = E\{\varepsilon(k)x(k-m)\}. \end{aligned} \quad (3.140)$$

In verkürzter Schreibweise $R_{xx}(mT) \triangleq R_{xx}(m)$ und unter Beachtung, daß der Ausdruck

$$E\{\varepsilon(k)x(k-m)\} = 0; \quad m > 0,$$

geht Gl. (3.140) in die Form

$$R_{xx}(m) + a_1 R_{xx}(m-1) + \dots + a_p R_{xx}(m-p) = 0; \quad m > 0 \quad (3.141)$$

über. Analog ergibt sich für $R_{xx}(0)$

$$R_{xx}(0) + a_1 R_{xx}(-1) + \dots + a_p R_{xx}(-p) = \sigma_\varepsilon^2 \quad (3.142)$$

wegen

$$E\{\varepsilon(k)x(k)\} = \sigma_\varepsilon^2.$$

Für die normierte Autokorrelationsfunktion erhält man mit den Gln. (3.141) und (3.142)

$$\begin{aligned} R'_{xx}(m) &= \frac{R_{xx}(m)}{R_{xx}(0)} = -a_1 R'_{xx}(m-1) - \dots - a_p R'_{xx}(m-p); \quad m > 0. \\ & \quad (3.143) \end{aligned}$$

Die Gln. (3.141) und (3.143) werden auch als *Yule-Walker-Gleichungen* bezeichnet. Mit ihrer Hilfe ist ebenfalls eine Schätzung der Parameter a_i möglich [3.9].

Das Leistungsdichtespektrum des autoregressiven Prozesses kann durch Anwendung der Wiener-Chintschinschen Beziehung – s. Abschnitt 3.2.3, Gl. (3.77) – oder über den Zusammenhang zwischen den Leistungsdichten und der Systembeschreibung – s. Abschnitt 3.2.4, Gl. (3.111) – erfolgen. Der letzte Weg soll hier näher erläutert werden, weil er unter den Bedingungen für das Signal $\varepsilon(t)$ und der Kenntnis der Parameter a_i geschlossene Lösungen für das Leistungsspektrum ergibt.

Analog zu Gl. (3.111) gilt für den AR-Prozess

$$S_{xx}(\omega) = |\bar{G}_F(j\omega)|^2 S_\varepsilon(\omega). \quad (3.144)$$

Für den vorliegenden diskontinuierlichen Fall erhält man aus der stochastischen Differenzgleichung (3.134) durch eine Z-Transformation die Pulsübertragungsfunktion des Filters zu [3.2]

$$G_F(z) = \frac{X(z)}{\varepsilon(z)} = \frac{1}{1 + a_1 z^{-1} + \dots + a_p z^{-p}} \quad \text{mit } z = e^{j\omega T}. \quad (3.145)$$

Damit ergibt sich für den Ausdruck $|G_F(j\omega)|^2$ in Gl. (3.144) aus Gl. (3.145) über einige Umformungen die allgemeine Beziehung

$$|G_F(e^{j\omega T})|^2 = 1 / \sum_{i=0}^p \sum_{j=0}^p a_i a_j \cos(i-j) \omega T. \quad (3.146)$$

Das Leistungsdichtespektrum des Signals $\varepsilon(t)$ ergibt sich aus den Annahmen im Abschnitt 3.3.3.1 zu

$$S_{\varepsilon}(\omega) = \sigma_{\varepsilon}^2. \quad (3.147)$$

Das Leistungsdichtespektrum des autoregressiven Prozesses der Ordnung r ist somit gegeben durch

$$S_{xx}(\omega) = \sigma_{\varepsilon}^2 / \sum_{i=0}^r \sum_{j=0}^r a_i a_j \cos(i-j) \omega T. \quad (3.148)$$

Soll das Spektrum aus den gemessenen Signalwerten geschätzt werden, sind in Gl. (3.148) die Koeffizienten $\hat{a}_i$ einzusetzen, und die Streuung des Signals $\varepsilon(t)$ ist als Schätzung aus den Beziehungen

$$\begin{aligned} \hat{\varepsilon}(k) &= x(k) - [\hat{a}_1 x(k-1) + \dots + \hat{a}_r x(k-r)] \\ \sigma_{\varepsilon}^2 &= \frac{1}{n-1} \sum_{k=1}^n (\hat{\varepsilon}(k) - \overline{\hat{\varepsilon}(k)})^2 \end{aligned} \quad (3.149)$$

zu ermitteln.

Im Beispiel 3.6 sind für eine Stichprobe eines autoregressiven Prozesses 1. Ordnung mit positivem Parameter a_1 die Korrelationsfunktion und das Leistungsspektrum dargestellt. Es zeigt ein typisches tieffrequentes Signalverhalten.

Wird in der Gleichung des autoregressiven Prozesses 1. Ordnung der Parameter a_1 negativ gewählt, entsteht ein Signal mit dominierenden hochfrequenten Komponenten. Den typischen Verlauf für diesen Fall zeigt Beispiel 3.7.

3.3.3.3. Modelle der gleitenden Mittel (MA-Modelle)

Das Modell der gleitenden Mittel ist die zweite Grundform des allgemeinen stochastischen Signalmodells von Gl. (3.130). Für dieses Modell wird angenommen, daß sich die Signale $x(k)$ aus den durch die Koeffizienten b_i bewerteten Werten des „weißen Rauschens“ $\varepsilon(k-i)$ nach der Vorschrift

$$x(k) = B(q^{-1}) \varepsilon(k) \quad (3.150)$$

ergeben. Die Gl. (3.150) stellt ein lineares Filter der Zufallsgröße $\varepsilon(k)$ dar; s. Analogie zu digitalen Filtern im Abschnitt 3.4.

Die Schätzung der Parameter des MA-Modells bereitet gegenüber der des AR-Modells größere Schwierigkeiten, weil im MA-Modell die Fehler nichtlineare Funktionen der

Beispiel 3.6 Kernfunktionen eines autoregressiven Prozesses 1. Ordnung ($+1 < a_1 < 0$)

Gegeben:

Signalgleichung

$$x(k) + a_1 x(k-1) = \varepsilon(k),$$

$$a_1 = 0,7,$$

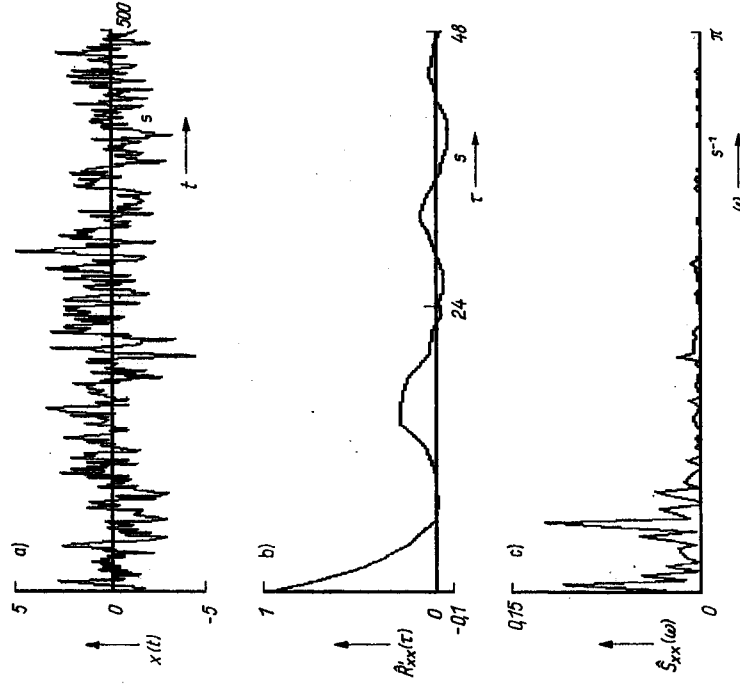
$\varepsilon(k)$ weißes Rauschen, NV (0, 1),
Abtastzeit $T = 1$ s,

Beobachtungen: 500.

Gesucht: Verläufe der Korrelationsfunktion und des Leistungsdichtespektrums.

Ergebnis:

Siehe Bild zu Beispiel 3.6.



Die Verläufe der berechneten Korrelationsfunktion $\hat{R}_{xx}(mT)$ und des Leistungsdichtespektrums $\hat{S}_{xx}(\omega)$ zeigen das typische Verhalten für ein stochastisches Signal mit dominierenden tieffrequenten Signalkomponenten.

Beispiel 3.7 Kernfunktionen eines autoregressiven Prozesses 1. Ordnung ($-1 < a_1 < 0$)

Gegeben:

Signalgleichung

$$x(k) + a_1 x(k-1) = \varepsilon(k),$$

$$a_1 = -0,7,$$

$\varepsilon(k)$ weißes Rauschen, NV (0, 1),

Abtastzeit $T = 1$ s,

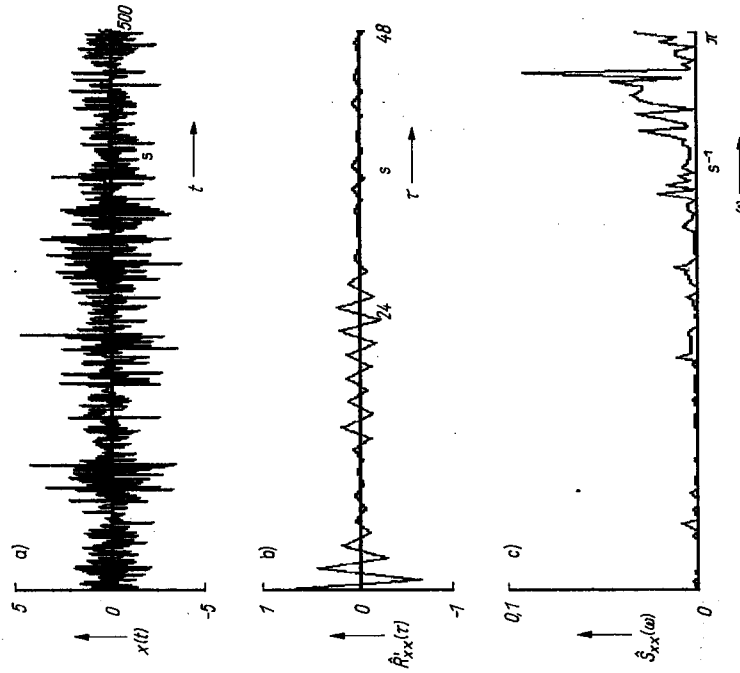
Beobachtungen: 500.

Gesucht:

Verläufe der Korrelationsfunktion und des Leistungsdichtespektrums.

Ergebnis:

Siehe Bild zu Beispiel 3.7.



Im Unterschied zu den im Beispiel 3.6 erhaltenen Ergebnissen ist deutlich zu erkennen, daß bei dem autoregressiven Prozeß 1. Ordnung mit negativem Parameter a_1 die ermittelte Korrelationsfunktion $R_{xx}(mT)$ alternierend das Vorzeichen wechselt und das Leistungsspektrum $S_{xx}(\omega)$ sehr gut die dominierenden hochfrequenten Signalkomponenten zeigt.

Parameter werden. Zur Lösung des Schätzproblems werden dann iterative Verfahren benutzt, auf die hier nur verwiesen werden soll [3.19]. Wesentlich effektiver ist die Ausnutzung des Dualitätsprinzips zwischen MA- und AR-Modell. Liegen die Nullstellen des charakteristischen Polynoms von Gl. (3.150), d.h.

$$1 + b_1 q^{-1} + \dots + b_l q^{-l} = 0, \quad (3.151)$$

außerhalb des Einheitskreises, ist das MA-Modell der Ordnung l in ein AR-Modell der Ordnung r umformbar. Es gilt dann

$$[B(q^{-1})]^{-1} x(k) = \varepsilon(k). \quad (3.152)$$

Damit erhält man als Approximation mit einer hinreichend großen Ordnung r die Beziehung

$$x(k) + a_1 x(k-1) + \dots + a_r x(k-r) = \varepsilon(k),$$

deren Parameter nach den Schätzvorschriften von Gl. (3.136) und Gl. (3.138) ermittelt werden können. Selbst wenn das sich so ergebende Modell mehr Parameter als das MA-Modell besitzt, wird dieser Weg wegen der Eigenschaften der Schätzung und der Interpretierbarkeit häufig bevorzugt.

Die Autokorrelationsfunktion des Signals erhält man wieder durch Multiplikation von Gl. (3.150) mit den Signalwerten $x(k-m)$ und Bilden der Erwartungswerte

$$R_{xx}(m) = E \{ \varepsilon(k) + b_1 \varepsilon(k-1) + \dots + b_l \varepsilon(k-l) \} \\ \times \{ \varepsilon(k-m) + b_1 \varepsilon(k-m-1) + \dots + b_l \varepsilon(k-m-l) \}$$

und somit für $m = 0$

$$R_{xx}(0) = (1 + b_1^2 + \dots + b_l^2) \sigma_\varepsilon^2$$

und für $m > 0$

$$R_{xx}(m) = \begin{cases} (b_m + b_{m+1}b_1 + \dots + b_{l-m}b_l) \sigma_\varepsilon^2 & \text{für } 0 < m \leq l \\ 0 & \text{für } m > l \end{cases} \quad (3.153)$$

Für das Leistungsdichtespektrum erhält man analog zu den Betrachtungen im Abschnitt 3.3.2 die Beziehung

$$S_{xx}(\omega) = \sigma_\varepsilon^2 \sum_{i=0}^l \sum_{j=0}^l b_i b_j \cos(i-j) \omega T. \quad (3.154)$$

3.3.3.4. Modelle eines autoregressiven Prozesses der gleitenden Mittel (ARMA-Modell)

Das Modell des autoregressiven Prozesses der gleitenden Mittel der Ordnung (r, l) vereinigt in sich das AR- und das MA-Modell. Die Bedeutung dieses Modells liegt vor allem darin, daß stationäre Zeitreihen oft durch ein ARMA-Modell mit geringerer Ordnung beschrieben werden können, als dies bei AR- oder MA-Modellen der Fall ist. Entsprechend Gl. (3.129) gilt für diesen Modelltyp mit $a_0 = b_0 = 1$ die Beschreibung

$$x(k) + a_1 x(k-1) + \dots + a_r x(k-r) \\ = \varepsilon(k) + b_1 \varepsilon(k-1) + \dots + b_l \varepsilon(k-l). \quad (3.155)$$

In symbolischer Schreibweise lautet Gl. (3.155)

$$A(q^{-1})x(k) = B(q^{-1})\varepsilon(k) \quad (3.156)$$

mit

$$\begin{aligned} A(q^{-1}) &= [1 + a_1q^{-1} + \dots + a_rq^{-r}] \\ B(q^{-1}) &= [1 + b_1q^{-1} + \dots + b_lq^{-l}]. \end{aligned}$$

Die Schätzung der Parameter der Polynome $A(q^{-1})$ und $B(q^{-1})$ gestaltet sich schwierig, weil die Signalwerte $\varepsilon(k-i)$ nicht meßbar sind. Hinzu kommt, daß die Fehlergleichung nicht linear in den Parametern ist [3.10, 3.14, 3.19].

Eine große Anzahl von Verfahren beruht auf einer iterativen mehrstufigen Lösung oder Linearisierung der Fehlergleichung der aus Gl. (3.156) abgeleiteten Beziehung

$$\varepsilon(k) = [B'(q^{-1})]^{-1} A(q^{-1})x(k) \quad (3.157)$$

mit

$$B'(q^{-1}) = [b_1q^{-1} + \dots + b_lq^{-l}].$$

Weitere Details dieser Lösungswege der Schätzaufgaben, die sehr umfangreich sind, findet man in [3.9, 3.20].

An dieser Stelle soll ein rekursiver Schätzalgorithmus vorgestellt werden, der bei geeigneter Wahl der Ordnungen (r, l) konvergiert und gegenüber den genannten Methoden einfach und robust ist [3.21, 3.22].

Aus Gl. (3.155) ergibt sich durch Umstellung

$$\varepsilon(k) = x(k) - [b_1\varepsilon(k-1) + \dots + b_l\varepsilon(k-l) - a_1x(k-1) - \dots - a_rx(k-r)]. \quad (3.158)$$

In Vektorschreibweise erhält man aus Gl. (3.158)

$$\varepsilon(k) = x(k) - \mathbf{m}^T(k)s \quad (3.159)$$

mit

$$s^T = [b_1 \dots b_l; -a_1 \dots -a_r]$$

$$\mathbf{m}^T(k) = [\varepsilon(k-1) \dots \varepsilon(k-l); x(k-1) \dots x(k-r)].$$

Die Fehlergleichung für die Schätzung kann aus Gl. (3.159) durch Einsetzen der Modellgrößen erstellt werden:

$$\hat{\varepsilon}(k) = x(k) - \hat{\mathbf{m}}^T(k)s \quad (3.160)$$

mit

$$\hat{s}^T = [b_1 \dots b_l; -\hat{a}_1 \dots -\hat{a}_r]$$

$$\hat{\mathbf{m}}^T = [\hat{\varepsilon}(k-1) \dots \hat{\varepsilon}(k-l); x(k-1) \dots x(k-r)].$$

In dieser Fehlergleichung sind neben den Parametern auch die Signalwerte $\hat{\varepsilon}(k-i)$ zu schätzen, da sie nicht unmittelbar gemessen werden können. Es wird deshalb ein Schätzverfahren gesucht, bei dem diese Signalwerte während der Schätzung mit anfallen und das sicher von gewählten Startwerten konvergiert. Aus der Gruppe der rekursiven Schätzverfahren, die diese Forderung erfüllen, hat sich auch in diesem Fall die rekursive Regression bewährt. Unter den für dieses Modell vereinbarten Bezeichnungen gilt entsprechend der Ableitung im Abschnitt 5.4.3 folgende Schätzvorschrift:

$$\hat{\varepsilon}(k+1) = \hat{\varepsilon}(k) + k(k+1)[x(k+1) - \hat{\mathbf{m}}^T(k+1)\hat{s}(k)] \quad (3.161)$$

mit

$$\hat{k}(k+1) = \hat{P}(k)[\mathbf{m}^T(k+1)\hat{P}(k)\hat{\mathbf{m}}(k+1) + 1]^{-1}$$

$$\hat{P}(k) = [\hat{M}(k)^T\hat{M}(k)]^{-1} \text{ und rekursiv}$$

$$\hat{P}(k) = \hat{P}(k-1) - \hat{k}(k)\hat{\mathbf{m}}^T(k)\hat{P}(k-1).$$

Die Schätzung kann u. a. dann abgebrochen werden, wenn für die Korrelationsfunktion $R_{xx}(m)$ gilt

$$R_{xx}(m) = \begin{cases} 0 & \text{für } m \neq 0 \\ \sigma_\varepsilon^2 & \text{für } m = 0 \end{cases} \quad (3.162)$$

Das Problem der Wahl der Startwerte für die Parameter und die Signalwerte $\hat{\varepsilon}(k-i)$ kann verringert werden, wenn durch eine Vorschätzung eines AR-Modells Anhaltspunkte gewonnen werden können. Alle anderen Startwerte und das Konvergenzverhalten entsprechen dem in den Abschnitten 3.3.2.1 und 5.4.3 Gesagten.

Auf die Herleitung der Autokorrelationsfunktion eines ARMA-Prozesses soll im Rahmen dieses Buches verzichtet werden, weil sie sehr aufwendig ist und sehr selten benutzt wird (s. [3.9]). Dagegen kann auf der Basis der Ableitungen für das Leistungsdichtespektrum des AR- und MA-Modells für diesen Prozeß eine geschlossene Lösung in Form von Gl. (3.163) angegeben werden. Es gilt mit Gln. (3.148) und (3.154) für das Leistungsdichtespektrum eines ARMA-Prozesses

$$S_{xx}(\omega) = \sigma_\varepsilon^2 \frac{\sum_{i=0}^l b_i b_j \cos(i-j)\omega T}{\sum_{i=0}^r a_i a_j \cos(i-j)\omega T} \quad (3.163)$$

3.3.3.5. Kombinierte Signalmodelle

In vielen praktischen Fällen haben sich Kombinationen deterministischer und stochastischer Signalmodelle bewährt, weil einerseits die Anzahl der Parameter des Modells wesentlich reduziert werden kann und andererseits ihre Schätzung nicht so viele Probleme mit sich bringt. Aus der Vielzahl der Kombinationsmöglichkeiten wird im Rahmen dieses Buches die wichtige Beschreibung der Signalfolge $x(k)$ durch einen Polynom- oder periodischen Ansatz und eine autoregressive Komponente näher betrachtet. Mit den Beschreibungen der Gln. (3.116), (3.125) und (3.131) ergeben sich damit die kombinierte Polynom-AR-Beschreibung und kombinierte periodische/AR-Beschreibung.

Kombinierte Polynom-/AR-Beschreibung

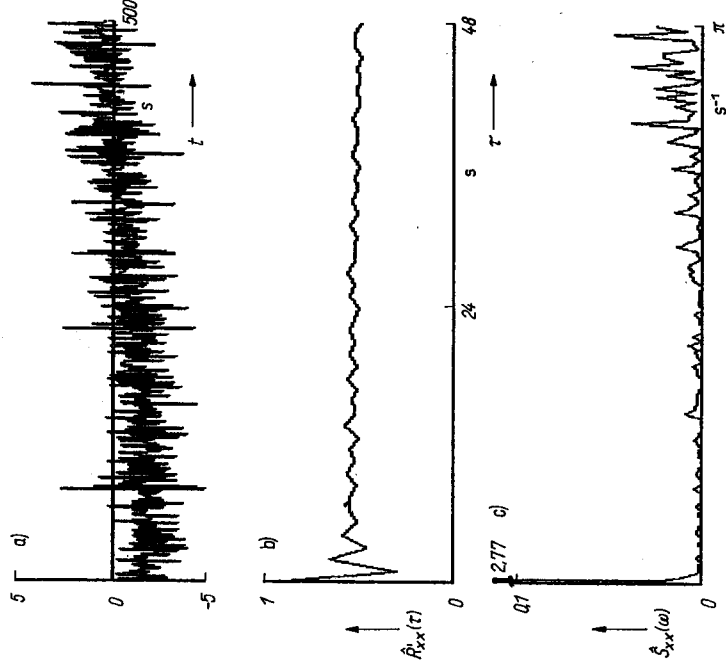
$$x(k) + a_1x(k-1) + \dots + a_rx(k-r) = \bar{d}_0 + d_1(kT) + \dots + d_m(kT)^m + \varepsilon(k) \quad (3.164)$$

Für ein einfaches Modell sind im Beispiel 3.8 der Signalverlauf, die Korrelationsfunktion und das Leistungsdichtespektrum dargestellt.

Zur Schätzung der Parameter a_i und d_i mittels Regression wird Gl. (3.164) zweckmäßigerweise wieder in eine Vektorform umgeschrieben:

$$\varepsilon(k) = x(k) - \mathbf{m}^T(k)s \quad (3.165)$$

Beispiel 3.8	Kombiniertes Signalmodell: autoregressiver/Polynom-Ansatz
Gegeben:	Signalgleichung $x(k) = \varepsilon(k) - a_1 x(k-1) + c_0 + c_1 kT + c_2 (kT)^2,$ $a_1 = -0,5, \quad c_0 = -2, \quad c_1 = 1 \cdot 10^{-3} \text{ s}^{-1},$ $c_2 = 1 \cdot 10^{-3} (\text{s}^{-1})^2,$ $\varepsilon(k)$ weißes Rauschen, NV (0, 1), Abtastzeit $T = 1 \text{ s}$, Beobachtungen: 500.
Gesucht:	Verläufe der Korrelationsfunktion und des Leistungsdichtespektrums.
Ergebnis:	Siehe Bild zu Beispiel 3.8.



Die berechnete Korrelationsfunktion $\hat{R}_{xx}(mT)$ und das Leistungsdichtespektrum $\hat{S}_{xx}(\omega)$ zeigen sehr gut den Anteil der deterministischen und stochastischen Signalkomponente.

mit

$$\mathbf{m}^T = [x(k-1) \dots x(k-r) : 1 \ (kT) \dots (kT)^m]$$

$$\mathbf{s}^T = [-a_1 \dots -a_r : d_0 d_1 \dots d_m].$$

$$\hat{\varepsilon}(k) = \mathbf{x}(k) - \mathbf{M}\hat{\mathbf{s}}, \quad (3.166)$$

$$\hat{\varepsilon}^T(k) = [\hat{\varepsilon}(1), \hat{\varepsilon}(2) \dots \hat{\varepsilon}(n)]$$

$$\mathbf{x}^T(k) = [x(1), x(2) \dots x(n)]$$

$$\mathbf{M} = \begin{bmatrix} x(0) & x(1) & \dots & x(-r+1) & 1 & (1T) & \dots & (1T)^m \\ x(1) & x(0) & \dots & x(-r+2) & 1 & (2T) & \dots & (2T)^m \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x(n-1) & x(n-2) & \dots & x(n-r) & 1 & (nT) & \dots & (nT)^m \end{bmatrix}$$

Wird das Signal $\hat{\varepsilon}(k)$ als Modellfehler betrachtet und seine Beschreibung in Gl. (3.166) verwendet, kann der Parametervektor mit der Methode der Regression (s. Abschn. 5.2.2) nach der Vorschrift

$$\hat{\mathbf{s}} = [\mathbf{M}^T \mathbf{M}]^{-1} \mathbf{M}^T \mathbf{x} \quad (3.167)$$

ermittelt werden. Eine rekursive Schätzung mittels der rekursiven Regression ist ebenfalls möglich (s. Abschn. 5.2.2).

Kombinierte periodische / AR-Beschreibung

$$\begin{aligned} x(k) + a_1 x(k-1) + \dots + a_r x(k-r) \\ = c_0 + \sum_{i=1}^I (c_i \cos \omega_i kT + b_i \sin \omega_i kT) + \varepsilon(k). \end{aligned} \quad (3.168)$$

Im Beispiel 3.9 ist der Verlauf eines Signals, das aus einer autoregressiven Signalkomponente 1. Ordnung, einer periodischen Komponente und einem Gleichanteil besteht, dargestellt.

In Vektorform lautet Gl. (3.168)

$$\varepsilon(k) = \mathbf{x}(k) - \mathbf{m}^T(k) \mathbf{s}, \quad (3.169)$$

$$\mathbf{m}^T(k) = [x(k-1) \dots x(k-r) : 1 : \cos \omega_1 kT \dots \cos \omega_I kT ;$$

$$\sin \omega_1 kT \dots \sin \omega_I kT]$$

$$\mathbf{s}^T = [-a_1 \dots -a_r : c_0 : c_1 \dots c_I : b_1 \dots b_I].$$

Als Gleichungssystem erhält man für n Beobachtungen

$$\hat{\varepsilon}(k) = \mathbf{x}(k) - \mathbf{M}\hat{\mathbf{s}} \quad (3.170)$$

Beispiel
3.9

Kombiniertes Signalmodell: autoregressiv/periodischer Ansatz

Gegeben:

Signalgleichung

$$x(k) = \varepsilon(k) - a_1 x(k-1) + c_0 + c_1 \sin \frac{2\pi}{24} kT,$$

$$a_1 = -0,5, \quad c_0 = -0,2, \quad c_1 = 0,5,$$

 $\varepsilon(k)$ weißes Rauschen, NV(0, 1),
Abtastzeit $T = 1$ s,

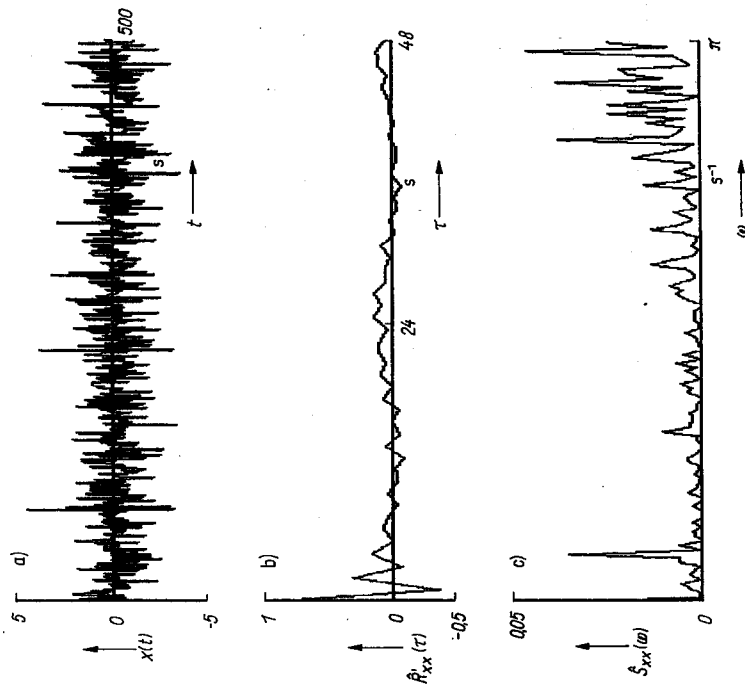
Beobachtungen: 500.

Gesucht:

Verläufe der Korrelationsfunktion und des Leistungsdichtespektrums.

Ergebnis:

Siehe Bild zu Beispiel 3.9.



Die einzelnen Signalkomponenten sind sowohl in der ermittelten Korrelationsfunktion $\hat{R}_{xx}(mT)$ als auch im Leistungsdichtespektrum $\hat{S}_{xx}(\omega)$ zu erkennen. Dabei ist wiederum zu sehen, daß periodische Signalanteile im Leistungsspektrum signifikanter hervortreten als in der Korrelationsfunktion.

mit

$$\hat{\varepsilon}^T(k) = [\hat{\varepsilon}(1) \hat{\varepsilon}(2) \dots \hat{\varepsilon}(n)]$$

$$x^T(k) = [x(1) x(2) \dots x(n)]$$

$$M = \begin{bmatrix} x(0) & \dots & x(-r+1) & 1 & \cos \omega_1 & 1T & \dots & \cos \omega_1 & 1T & \sin \omega_1 & 1T & \dots & \sin \omega_1 & 1T \\ x(1) & \dots & x(-r+2) & 1 & \cos \omega_1 & 2T & \dots & \cos \omega_1 & 2T & \sin \omega_1 & 2T & \dots & \sin \omega_1 & 2T \\ \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ x(n-1) & \dots & x(n-r) & 1 & \cos \omega_1 & nT & \dots & \cos \omega_1 & nT & \sin \omega_1 & nT & \dots & \sin \omega_1 & nT \end{bmatrix}$$

Bei bekannten oder entsprechend der Strategie im Abschnitt 3.3.2.2 ermittelten dominierenden Frequenzen ω_l kann eine Schätzung im Sinne der kleinsten Fehlerquadrate, d. h. mittels Regression (s. Abschn. 5.2.2), nach der Beziehung

$$\hat{s} = [M^T M]^{-1} M^T x \quad (3.171)$$

erfolgen.

Für die hier vorgestellten kombinierten Signalmodelle sollte die Ordnung der deterministischen Teilmodelle klein gehalten werden ($m \leq 2, l \leq 2$) und die Anzahl der Beobachtungen hinreichend groß sein. Die Parameter des deterministischen Teilmodells werden dann relativ unabhängig von der Korrelation und Streuung der Zufallskomponente ermittelt.

Gegenüber den deterministischen Signalmodellen von Abschnitt 3.3.2 besitzen diese kombinierten Modelle dann Vorteile, wenn die stochastische Signalkomponente korreliert ist. Im Vergleich zu den stochastischen Signalmodellen von Abschnitt 3.3.3 bilden sie ferner Signalverläufe mit ausgeprägten deterministischen Anteilen mit wesentlich weniger Parametern sehr gut nach.

3.4. Ausgewählte Methoden der Signalaufbereitung

3.4.1. Notwendigkeit einer Signalaufbereitung

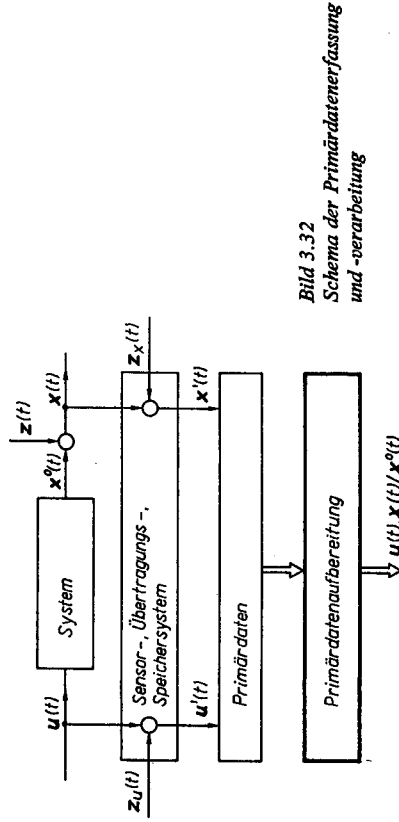
Die mittels Sensoren erfaßten Informationen in Form von zeitlichen Verläufen der Signale stellen bei fehlerfreier Erfassung, Übertragung und Speicherung die genaueste Basis für die weitere Ableitung von Signal- und/oder Systemmodellen dar. Es wird wiederum davon ausgegangen, daß die Information über die Signale in abgetasteter Form als $x(kT)$ vorliegt; dabei wird in diesem Abschnitt vielfach die vereinfachte Schreibweise x verwendet. In den meisten Fällen muß jedoch davon ausgegangen werden, daß in den genannten Stufen (Erfassung, Übertragung und Speicherung) das ursprüngliche Signal verfälscht werden kann. Hinzu kommt, daß die vom Sensor erfaßten Prozeßgrößen/-zustände bereits gestört sein können. Die Verfälschung einer Prozeßgröße kann u. a. ihre Ursache haben im

– Prinzip des Sensors

– Einfluß von „Umweltfaktoren“ und anderen nichtmeßbaren Größen auf den Sensor, die Übertragung und die Speicherung.

Damit ergibt sich das im Bild 3.32 dargestellte allgemeine Schema. Die Eingangs- und Ausgangsgrößen $w(t)$ und $x(t)$ können durch die Störungen z_u und z_x bei der Messung,

der Übertragung und Abspeicherung verfälscht werden. Eine direkte weitere Verwendung der nach diesem Schritt vorliegenden Primärinformationen $u'(t)$ und $x'(t)$ würde eine Abbildung des Verhaltens vom Prozeß einschließlich Sensor-, Übertragungs- und Speicherungssystem, nicht aber des Prozesses allein ermöglichen. Aus diesem Grund ist eine Aufbereitung der Primärdaten bzw. der Primärdaten mittels verschiedener Methoden notwendig. Das Ziel dieser Methoden ist es, die Verfälschung der Prozeßgrößen zu erkennen und sie durch Transformationen, Korrekturen oder Filterung möglichst zu beseitigen. Zu diesen beiden Aufgabenkomplexen sollen in den folgenden Abschnitten ausgewählte Methoden abgeleitet werden, die auf die bereits bekannten Signalmodelle aufbauen und als Ausgangspunkt für weiterführende Betrachtungen angesehen werden können.



3.4.2. Methoden zum Erkennen von Verfälschungen der Prozeßsignale/-zustände

Grundlage für die Methoden zum Erkennen von Verfälschungen der Prozeßsignale/-zustände sind A-priori-Informationen über das „unverfälschte“ Verhalten. Diese Informationen beruhen auf

- theoretischen und/oder experimentellen Analysen des statischen und dynamischen Verhaltens
- Vergleichsmessungen
- Erfahrungen/Intuition der Prozeßspezialisten.

Für die Prüfung des statischen Verhaltens gemessener Signale (auch als Größen bezeichnet) sind deshalb u. a. folgende Aufgaben zu lösen:

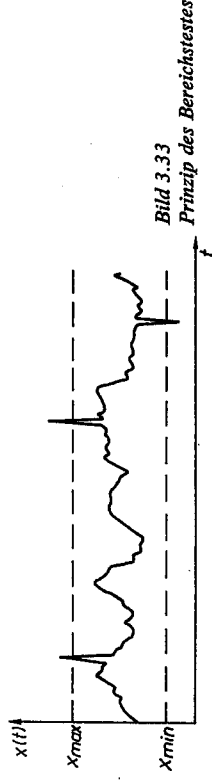
- Ermittlung der analytischen Beziehungen oder der Kennlinien, die bei Signalwandlungen (z. B. AD-Wandlung) auftreten
- Ermittlung von sinnvollen Wertebereichen für die Amplituden der Prozeßsignale
- Ermittlung von analytischen Beziehungen über den Einfluß von Umgebungsbedingungen (z. B. Temperatur) auf das betrachtete Prozeßsignal
- Aufstellen von Bilanzgleichungen (z. B. Masse).

Auf diesem Wissen können die folgenden, einfach zu realisierenden Prüfverfahren für die Prozeßsignale (als allgemeines Signal wird x verwendet) aufgebaut werden:

Bereichstest. Der Bereichstest erfolgt durch einen Vergleich des abgetasteten Signalwerts mit vorgegebenen Grenzwerten für das jeweilige Signal nach der einfachen Vorschrift

$$x_{\min} \leq x(k) \leq x_{\max} \quad (3.172)$$

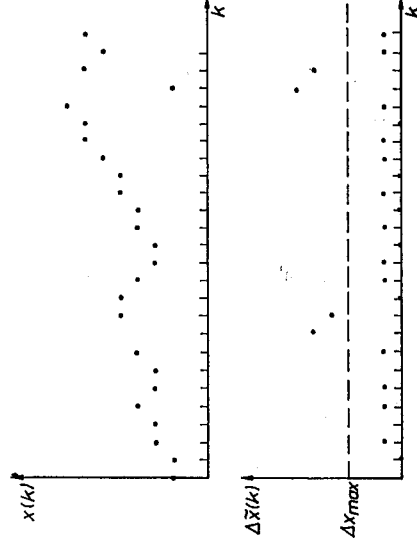
Die Bereichsgrenzen $x_{\min}$ und $x_{\max}$ (Bild 3.33) werden durch theoretische Kenntnisse als „technologisch sinnvolle“ Werte oder durch statistische Auswertungen und Annahmen eines normalverteilten Signals als das Vielfache der Streuung (σ -Bereiche, s. Abschn. 2.2.1) gewonnen.



Differenztest. Der Differenztest beruht auf der Prüfung der Differenz zweier aufeinanderfolgender abgetasteter Signalwerte (Bild 3.34) in der Form

$$|x(k) - x(k-1)| \leq \Delta x_{\max} \quad (3.173)$$

Dieser Test ist in vielen Fällen wirksamer als der Bereichstest, wobei die Festlegung der maximalen Änderung je Tastschritt u. U. Schwierigkeiten bereitet.



Bilanztest. Der Bilanztest beruht auf der Beurteilung mindestens zweier erfaßter Prozeßsignale. Durch Bildung der Bilanzgleichungen (z. B. Masse-, Energie-, Durchflußbilanzen) wird geprüft, ob die Signale bis auf einen vorher festgelegten Fehlerbereich Δx von Null abweichen. Als Prüfgleichung verwendet man die allgemeine Bilanzgleichung

$$\left| \sum_{i=1}^n x_{i, zu}(k) - \sum_{j=1}^m x_{j, ab}(k) - \sum_{l=1}^r x_{l, gesp}(k) \right| \leq \Delta x; \quad (3.174)$$

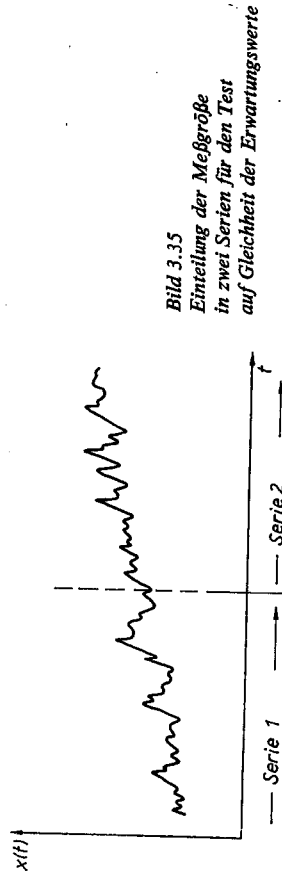
$x_{i, zu}(k)$ i -te zugeführte Größe
 $x_{l, gesp}(k)$ Änderung der l -ten gespeicherten Größe
 $x_{j, ab}(k)$ j -te abgeführte Größe

Wird die Grenze Δx in Gl. (3.174) überschritten, sind Überprüfungen der Meß-, Übertragungs- und Speichersysteme und eine Korrektur der Meßwerte erforderlich. Diese Strategie ist auf vergleichbare Signale beschränkt.

Für die Prüfung des richtigen dynamischen Verhaltens der gemessenen Signale sind u. a. folgende Aufgaben als Voraussetzung zu lösen:

- Ermittlung des normalen und aktuellen dynamischen Verhaltens des Sensors
- Ermittlung der Frequenzlage des Nutz- und des Störsignals.

Die Prüfung des dynamischen Verhaltens erfolgt vorwiegend in Form von Vergleichen des normalen Verhaltens (Standardverhalten) mit dem aktuellen Verhalten. Häufig werden zu diesem Zweck Vergleichsmessungen mit aufgeprägten Signalen vorgenommen. Als eine der wichtigsten Prüfmethoden des dynamischen Verhaltens soll der Test auf tief- und hochfrequente Zeitkomponenten, auch als Trend bezeichnet, vorgestellt werden. Prozeßsignale, die diese Eigenschaft besitzen, werden als instationär bezeichnet, da ihre Verteilung/ihre Momente eine Funktion der Zeit sind. Die Methoden zur Prüfung auf Stationarität beruhen vor allem darauf, daß das beobachtete Signal in einzelne Abschnitte (Serien) unterteilt wird und die Momente der Serien oder daraus abgeleitete Größen durch statistische Testverfahren (s. Abschn. 2.3.3) miteinander verglichen werden (Bild 3.35).



Als Prüfverfahren eignen sich u. a. folgende:

Test auf Gleichheit der Erwartungswerte zweier unabhängiger normalverteilter Grundgesamtheiten (t-Test)

Bei diesem Test soll geprüft werden, ob der Erwartungswert der Serie 1 ($E\{X_1\} = \mu_{x1}$) gleich dem Erwartungswert der Serie 2 ($E\{X_2\} = \mu_{x2}$) ist (Bild 3.35). Dabei wird vorausgesetzt, daß die beiden Serien die gleichen Streuungen besitzen, d. h., $\sigma^2 = \sigma_{x1}^2 = \sigma_{x2}^2$ gilt. Ausgehend von der Hypothese $H_0: \mu_{x1} = \mu_{x2}$ wird für die konkrete Stichprobe der Meßgröße mit den Beobachtungen n_1 für die Serie 1 und n_2 für die Serie 2 die Hypothese verworfen, wenn

$$\left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(n_1 - 1)s_{x1}^2 + (n_2 - 1)s_{x2}^2}} \right| \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \geq t_{\alpha/2, f}; \quad (3.175)$$

$$f = n_1 + n_2 - 2$$

ist [2.3, 2.4]. Damit wird eine Trendermittlung möglich und eine Trendkorrektur erforderlich.

Vergleich der mittleren Streuung einer Serie mit der Streuung der Mittelwerte aller Serien (F-Test)

Die Daten der beobachteten Meßgröße werden entsprechend Bild 3.36 in m Serien mit jeweils n Daten eingeteilt. Aus diesen Serien werden ermittelt:

- die Serienstreuungen s_j^2 ; $j = 1, 2, \dots, m$

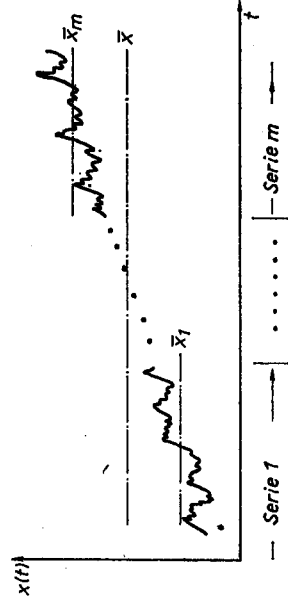
$$s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2,$$

- die mittlere Streuung einer Serie

$$s_x^2 = \sum_{j=1}^m s_j^2 / m,$$

- die Streuung der Mittelwerte

$$s_{\bar{x}}^2 = \frac{1}{m-1} \sum_{j=1}^m (x_j - \bar{x})^2.$$



Als Test wird der F-Test (s. Abschn. 2.3.3) in der Form

$$F = s_{\bar{x}}^2 / s_x^2 \quad (3.176)$$

mit den Freiheitsgraden $f_{s_{\bar{x}}^2} = m-1$ und $f_{s_x^2} = m(n-1)$ verwendet. Das Prozeßsignal x wird mit einer Irrtumswahrscheinlichkeit α als stationär angesehen, wenn

$$F < F_{\alpha, f_{s_{\bar{x}}^2}, f_{s_x^2}} \quad (3.177)$$

Ist. Für den Test ist es zweckmäßig, die Anzahl der Serien $m \geq 4$ und die Anzahl der Beobachtungen in einer Serie $n \geq 10$ zu wählen.

Damit steht eine Reihe von Methoden und Strategien zur Verfügung, um eine Abweichung der Prozeßsignale von ihrem normalen Verhalten zu erkennen. In den folgenden Abschnitten soll auf Methoden näher eingegangen werden, die den Einfluß von erkannten Verfälschungen im statischen oder dynamischen Verhalten der gemessenen Signale reduzieren.

3.4.3. Methoden zur Korrektur gemessener Signale

Nachdem eine Verfälschung der Signale bei der Messung, der Übertragung oder der Speicherung erkannt worden ist, muß diese durch Korrekturmaßnahmen möglichst gut beseitigt werden. Einige weitverbreitete Strategien werden im folgenden angeführt.

3.4.3.1. Eliminierung von Signalwerten

Werden die im Bereichs-, Differenz- oder Bilanztest im Abschnitt 3.4.2 angegebenen Schranken durch die erfaßten Signale verletzt, sind sie als „Ausreißer“ zu eliminieren bzw. ist der letzte als zuverlässig anerkannte Datensatz für diesen Zeitpunkt weiter zu verwenden.

3.4.3.2. Statische Korrektur von Signalen

Die durch das Meßsystem wiedergegebenen Prozeßsignale werden häufig durch dessen statisches Übertragungsverhalten verformt, oder es gehen die Umgebungsbedingungen (Sekundärgrößen) in den erfaßten Wert ein. Der erste Umstand kann bei Kenntnis des statischen Übertragungsverhaltens des Meßsystems in Form von Kennlinien oder analytischen Ausdrücken korrigiert werden. Damit ergibt sich z. B. der korrigierte Meßwert nach der Vorschrift

$$\tilde{x}(k) = a_0 + a_1 x(k) + a_2 x^2(k) + \dots$$

Liegt eine Beeinflussung der erfaßten Prozeßgröße durch Sekundärgrößen vor (z. B. Temperatur), erfolgt eine Korrektur der Daten in der Regel auf der Grundlage von analytischen Ausdrücken der Form

$$\tilde{x}(k) = a_0 + a_1 x(k);$$

$$a_1 = f(\text{Sekundärmeßgrößen}).$$

3.4.3.3. Dynamische Korrektur von Signalen

Unter einer dynamischen Korrektur von Signalwerten soll hier die Korrektur des dynamischen Verhaltens des Meßsystems mit dem Ziel verstanden werden, daß der Endwert der Messung möglichst schnell und genau erreicht wird. Da die Sensorsysteme überwiegend Trägheitsverhalten aufweisen, ist zur Korrektur ein inverses Verhalten, d. h. ein dominierender Vorhalt, notwendig. Dies bereitet technisch einige Probleme. Eine weitere Möglichkeit besteht in der Endwertvorhersage aus beobachteten Anfangsverläufen und bekanntem Sensormodell.

Die Aufgaben der statischen und der dynamischen Korrektur werden in leistungsstarken Meßsystemen durch in die Geräte integrierte Rechnersysteme gelöst. Der Entwurf von Korrekturmaßnahmen kann auch auf der Basis der im Abschnitt 3.4.4 vorgestellten Filter oder durch die Anwendung von Korrekturnetzwerken [3.1] erfolgen.

3.4.4. Filterung von Signalen

Bei der Aufbereitung von Signalen nimmt die Aufgabe der Trennung von Nutz- und Störsignalen durch Filter eine zentrale Stellung ein. Als Filter werden analoge oder digitale Systeme angesehen, die den spektralen Inhalt eines Signals in einem vorgegebenen Frequenzbereich übertragen. Im Rahmen dieses Buches soll auf einige Grundlagen der Trennung von Nutz- und Störsignalen mit digitalen Mitteln näher eingegangen werden, weil sie der Ausgangspunkt für das Studium weiterer leistungsfähiger Verfahren der Filtertheorie sind [3.2, 3.5, 3.23, 3.24].

3.4.4.1. Grundstrategie des Filterentwurfs

Die Strategie des Entwurfs von Filtern für die Trennung von Nutz- und Störsignal geht von der im Bild 3.37 dargestellten Struktur aus. Das Nutzsignal $x^0(t)$ wird durch das Störsignal $z(t)$ verfälscht. Als gemessenes Signal steht $x(t)$ zur Verfügung. Es besteht nun die Aufgabe, ein Filter bzw. eine Vorschrift zum Trennen von Nutz- und Störsignal so zu entwerfen, daß das gefilterte Signal $\tilde{x}(t)$ möglichst dem unverfälschten Signal $x^0(t)$ entspricht. Der Entwurf von optimalen Filtern wurde von Wiener im Sinne des minimalen mittleren quadratischen Fehlers gelöst [3.6, 3.27]. Er soll hier nicht weiter betrachtet werden. Ausgangspunkt für die hier vorgestellte Strategie ist der Filterentwurf auf der Grundlage einer aus der Spektralanalyse oder aus theoretischen Überlegungen vorliegenden A-priori-Information über die Lage der Nutz- und Störsignalspektren.

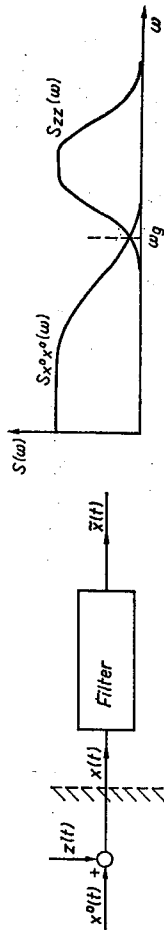


Bild 3.37. Grundstruktur der Filteraufgabe

Bild 3.38. Frequenzlage des Stör- und Nutzsignals

Für die Trennung beider Signalanteile muß immer gelten, daß sich ihre Spektren in der Frequenzlage signifikant unterscheiden (s. z. B. Bild 3.38). Aufgrund der vorliegenden A-priori-Informationen können dann die in Tafel 3.6 vorgestellten vier Filtertypen mit

Tafel 3.6. Zusammenhang zwischen der Lage des Nutz- und Störsignalspektrums und der Filterart

Lage des Leistungsspektrums	Filterart
$x^0(t)$ tieffrequent $z(t)$ hochfrequent	Tiefpaß $ G_F \uparrow$
$x^0(t)$ hochfrequent $z(t)$ tieffrequent	Hochpaß $ G_F \uparrow$
$x^0(t)$ in einem engen Frequenzbereich	Bandpaß $ G_F \uparrow$
$z(t)$ in einem engen Frequenzbereich	Bandsperre $ G_F \uparrow$

ihren idealen Amplitudenfrequenzgängen unterschieden werden. Im Rahmen dieses Buches wird auf die in der Praxis oft verwendeten digitalen Tief- und Hochpaßfilter eingegangen.

3.4.4.2. Digitale Tiefpaßfilter

Bei der Betrachtung der Signalübertragungseigenschaften linearer digitaler Tiefpaßfilter soll von zwei grundsätzlichen Realisierungsstrukturen aus auf den Entwurf von Filtern und auf die Besonderheiten des Übertragungsverhaltens digitaler gegenüber analogen Filtern eingegangen werden.

Als grundlegende Realisierungsstrukturen für lineare digitale Filter existieren nicht-rekursive und rekursive Varianten.

Nichtrekursive Filter

Bei diesem Filtertyp wird der gefilterte Signalwert $\tilde{x}(k)$ aus den ungefilterten Signalwerten $x(k-i)$ nach der Vorschrift

$$\tilde{x}(k) = \sum_{i=-n}^{+n} b_i x(k-i) \quad \text{mit } b_i \geq 0, \quad n \text{ endlich} \quad (3.178)$$

direkt gewonnen. Die Filterstruktur ist schematisch im Bild 3.39 dargestellt.

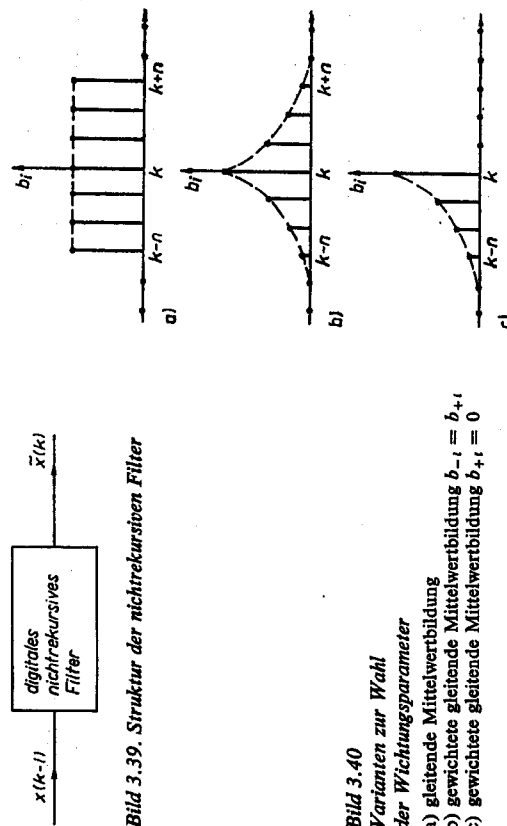


Bild 3.39. Struktur der nichtrekursiven Filter

Bild 3.40

Varianten zur Wahl

der Wichtungparameter

a) gleitende Mittelwertbildung

b) gewichtete gleitende Mittelwertbildung $b_{-i} = b_{+i}$

c) gewichtete gleitende Mittelwertbildung $b_{+i} = 0$

Die Wichtungparameter können nach den folgenden Gesichtspunkten gestaltet werden (Bild 3.40):

$b_i = 1/2n + 1$ für alle i , d. h. gleitende Mittelwertbildung über das Datenfenster $2n + 1$,
 $b_i = b_{-i}$ für alle i , d. h. gewichtete gleitende Mittelwertbildung ohne Phasenverschiebung,

$b_i = 0$ für $i > 0$, d. h. gewichtete gleitende Mittelwertbildung mit Phasenverschiebung.

Häufig wird diese Signalverarbeitung auch als „Glättung“ oder „Mittelwertbildung“ bezeichnet. Die Auswirkung der Gestaltung der Wichtungparameter auf die Güte der Filterung wird in den Beispielen 3.10, 3.11 und 3.12 untersucht.

Für die Darstellung des Übertragungsverhaltens des digitalen nichtrekursiven Filters wird die Antwort auf das mit der Tasterperiode T abgetastete harmonische Eingangssignal der Amplitude 1 betrachtet. Somit gilt für das Eingangssignal des Filters

$$X^*(j\omega) = e^{j\omega k T} \quad (3.179)$$

und für das Ausgangssignal des Filters

$$Y^*(j\omega) = G_F(e^{j\omega T}) e^{j\omega k T}; \quad (3.180)$$

$G_F(e^{j\omega T})$ Filterübertragungsfunktion.

Wird die transformierte Gl. (3.178) in Gl. (3.180) eingesetzt, erhält man die allgemeine Filterübertragungsfunktion zu

$$G_F(e^{j\omega T}) = \sum_{i=-n}^{+n} b_i e^{-j i \omega T}. \quad (3.181)$$

Sind für alle i die Koeffizienten $b_i = b_{-i}$, vereinfacht sich Gl. (3.181) zu der Vorschrift

$$G_F(e^{j\omega T}) = b_0 + 2 \sum_{i=1}^n b_i \cos i \omega T. \quad (3.182)$$

Gl. (3.182) zeigt deutlich, daß der Amplituden- und Phasenfrequenzgang des Filters periodisch mit der Periode $\omega T = 2\pi$ sein wird. Eines der wichtigsten Probleme ist die Ermittlung der Koeffizienten b_i in Abhängigkeit von der Breite n des Datenfensters und der gewählten Grenzfrequenz ω_g , d. h. des Durchlaßbereichs des Filters.

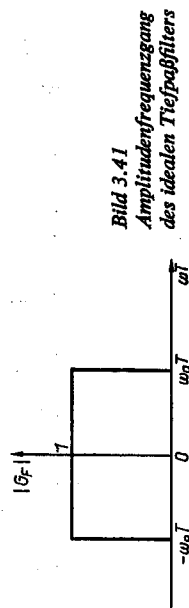


Bild 3.41

Amplitudenfrequenzgang
des idealen Tiefpaßfilters

Angestrebt wird ein idealer Amplitudenfrequenzgang des Filters nach Bild 3.41. Zwischen dem komplexen Frequenzgang $G(e^{j\omega T})$ und den Koeffizienten b_i besteht in Analogie zur Fourier-Reihe (s. Abschn. 3.1.1) der Zusammenhang

$$b_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(e^{j\omega T}) e^{j i \omega T} d\omega \quad (3.183)$$

für beliebige Gewichte b_i und

$$b_i = \frac{1}{\pi} \int_0^{\pi} G(e^{j\omega T}) \cos i \omega T d\omega \quad (3.184)$$

für den Fall, daß für alle i die Beziehung $b_{-i} = b_i$ gilt. Wird in Gl. (3.184) der ideale Frequenzgang entsprechend Bild 3.41 eingesetzt, erhält man

$$b_i = \frac{1}{\pi} \int_0^{\omega_g T} 1 \cdot \cos i \omega T d\omega. \quad (3.185)$$

Als Berechnungsvorschrift für die Gewichte b_i findet man damit

$$b_i = \frac{\omega_s T}{i \omega_s T} \sin i \omega_s T \quad \text{für } i > 0. \quad (3.186)$$

Als endgültige Filtergleichung ergibt sich dann aus Gl. (3.178) durch Einsetzen von Gl. (3.186) der gefilterte Wert im k -ten Zeitpunkt zu

$$\tilde{x}(k) = \frac{\omega_s T}{\pi} \sum_{i=-n}^{+n} \frac{\sin i \omega_s T}{i \omega_s T} x(k-i). \quad (3.187)$$

Die Güte der Anpassung des realisierten Amplitudenfrequenzgangs an den idealen Verlauf im Bild 3.41 hängt ab von dem Produkt $\omega_s T$ und der Länge des Datenfensters n . Je größer n ist, um so besser ist die Anpassung. Der Parameter b_0 ergibt sich aus Gl. (3.186) durch Grenzwertbildung zu

$$b_0 = \frac{\omega_s T}{\pi}.$$

Ein lineares Filter mit der Verstärkung von 1 erhält man durch eine Normierung des Parameters b_0 in der Form

$$b'_0 = 1 - \sum_{i=-n}^{+n} b_i \quad \text{mit } i \neq 0.$$

An zwei einfachen Beispielen soll das Verhalten nichtrekursiver Filter demonstriert werden. Eine Normierung der Filterparameter auf eine Gesamtfilterverstärkung von 1 erfolgt nicht bei den weiteren Betrachtungen. Wenn für alle i die Wichtungparameter $b_i = b_{-i}$ sind, ist die Filterübertragungsfunktion eine reelle Größe, und die Phasenverschiebung des gefilterten Signals $\tilde{x}(k)$ gegenüber dem ungefilterten Signal $x(k)$ ist 0.

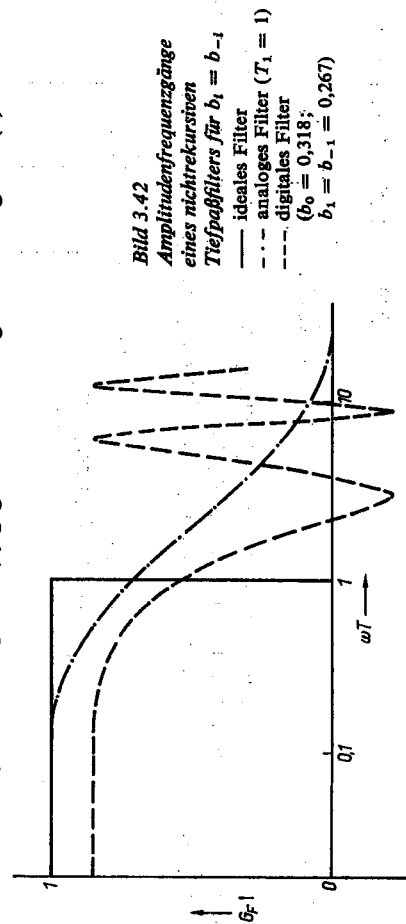


Bild 3.42
Amplitudenfrequenzgänge
eines nichtrekursiven
Tiefpassfilters für $b_1 = b_{-1}$
— ideales Filter
--- analoges Filter ($T_1 = 1$)
- - - digitales Filter
($b_0 = 0,318$;
 $b_1 = b_{-1} = 0,267$)

Für $n = 1$ ergibt sich aus Gl. (3.182) der Amplitudenfrequenzgang zu

$$|G(e^{j\omega T})| = b_0 + 2b_1 \cos \omega T. \quad (3.188)$$

Er ist im Bild 3.42 gemeinsam mit dem idealen Amplitudenfrequenzgang und dem eines analogen Filters 1. Ordnung für die Parameter $\omega_s T = 1$ dargestellt. Die Darstellung zeigt das für Digitalfilter typische periodische Verhalten für den Amplitudenfrequenzgang. Für Werte von $\omega T < \pi$ werden die Leistungsanteile zufriedenstellend gedämpft. Das

Einfluß der Wichtungsarten bei nichtrekursiven Tiefpassfiltern

Beispiel
3.10

Gegeben:

Signalbeschreibung

$$x(t) = a_0 + a_1 \sin \frac{2\pi}{100} t + z(t),$$

$$a_0 = 1, \quad a_1 = 0,5,$$

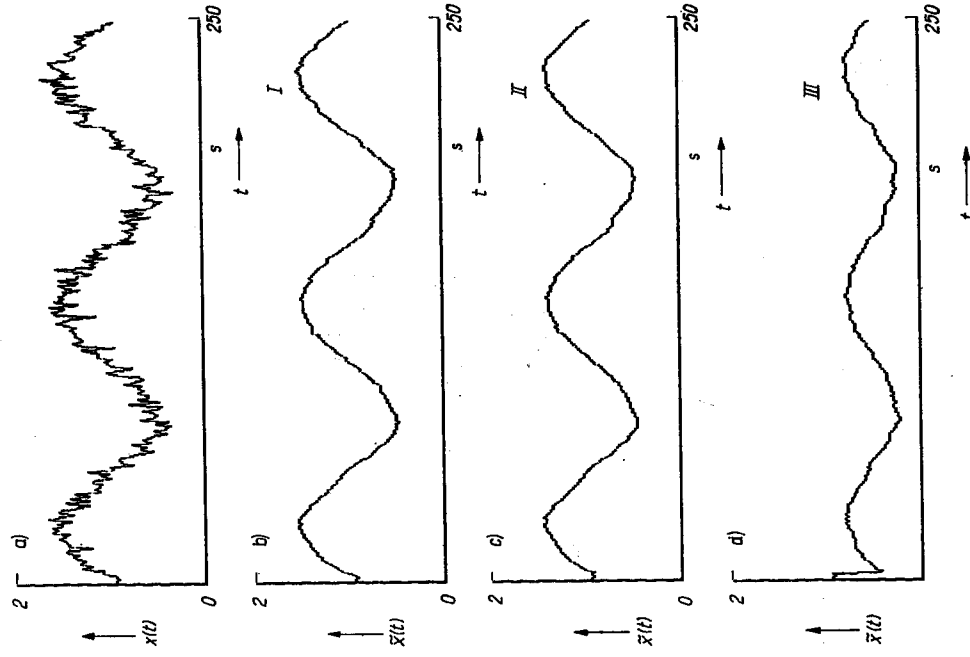
$$z(t) \text{ weißes Rauschen, } 0,1 \{NV(0,1)\},$$

$$\text{Filtergleichung (3.178), } n = 5,$$

$$\omega_s = 0,314 \text{ s}^{-1},$$

$$\text{Abtastzeit } T = 1 \text{ s},$$

$$\text{Beobachtungen: 250.}$$



Beispiel 3.10 Einfluß der Wichtungsarten bei nichtrekursiven Tiefpaßfiltern

Gesucht:

Verläufe des gefilterten Signals $\tilde{x}(k)$ für

- I. alle $b_i = 1/2n + 1 = 0,0909$,
- II. alle $b_{-i} = b_{+i}$ nach Gl. (3.186)
 $b_0 = 0,1, \quad b_1 = 0,098, \quad b_2 = 0,094,$
 $b_3 = 0,086, \quad b_4 = 0,076, \quad b_5 = 0,064,$
- III. alle $b_i = 0$ für $i > 0$ (alle b_i für $i < 0$ s. II).

Ergebnis:

Siehe Bild zu Beispiel 3.10.

- Die dargestellten gefilterten Signalverläufe zeigen, daß
1. die Filterwirkung der Varianten I und II sehr gut die hochfrequente Störkomponente ohne eine Phasenverschiebung des gefilterten gegenüber dem ungefilterten Signal beseitigen,
2. die für die Echtzeitverarbeitung geeignete Variante III ebenfalls die hochfrequente Störung beseitigt, aber neben dem Verstärkungsverlust eine wesentliche Phasennachstellung des gefilterten gegenüber dem ungefilterten Signal hervorruft.

bedeutet, daß die Tasterperiode T bei vorgegebener Grenzfrequenz ω_g entsprechend dem Abtasttheorem von Shannon nach der Vorschrift

$$T \leq \pi/\omega_g$$

zu wählen ist.

Sind alle $b_{+i} = 0; i = 1, 2, \dots, n$, entspricht die nichtrekursive Filtergleichung (3.178) der Faltungssumme (s. Abschn. 6.2). Für den Amplituden- und Phasenfrequenzgang ergeben sich für $n = 1$

$$|G(e^{j\omega T})| = \sqrt{(b_0 + b_1 \cos \omega T)^2 + (b_1 \sin \omega T)^2}$$

$$\varphi(\omega T) = -\arctan \frac{b_1 \sin \omega T}{b_0 + b_1 \cos \omega T} \quad (3.189)$$

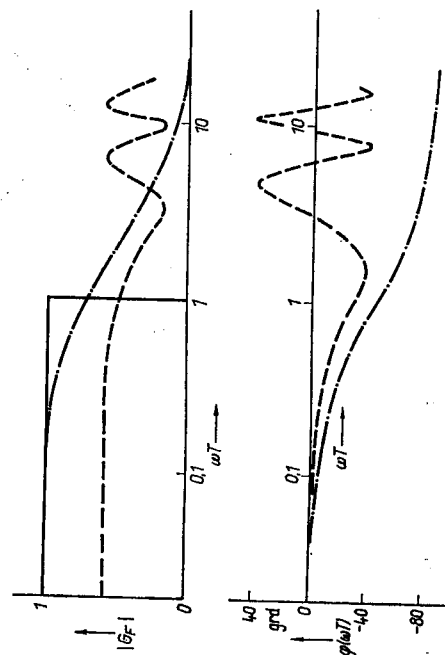


Bild 3.43 Amplituden- und Phasenfrequenzgang eines nichtrekursiven Tiefpaßfilters für $b_{+i} = 0$
 — ideales Filter ($\omega_g T = 1$)
 --- analoges Filter ($T_1 = 1$)
 - · - digitales Filter ($b_0 = 0,318; b_1 = 0,267$)

Beispiel 3.11 Einfluß der Normierung der Verstärkung auf die Wirkung nichtrekursiver Tiefpaßfilter ($b_i = 0$ für $i > 0$)

Gegeben:

Signalbeschreibung

$x(t) = x_1(t) + z(t),$
 $x_1(t) = \begin{cases} 0 & \text{für } t < 100 \text{ s} \\ 10\sigma(t) & \text{für } t \geq 100 \text{ s}, \end{cases}$
 $z(t)$ weißes Rauschen, NV (0, 1),
 Filtergleichung (3.178), $n = 7, \omega_g = 0,628 \text{ s}^{-1},$
 Abtastzeit $T = 1 \text{ s},$
 Beobachtungsbereich: 76 bis 125 s.

Gesucht:

Verläufe des gefilterten Signals $\tilde{x}(k)$ für

a) Wichtung (unnormierte Verstärkung)

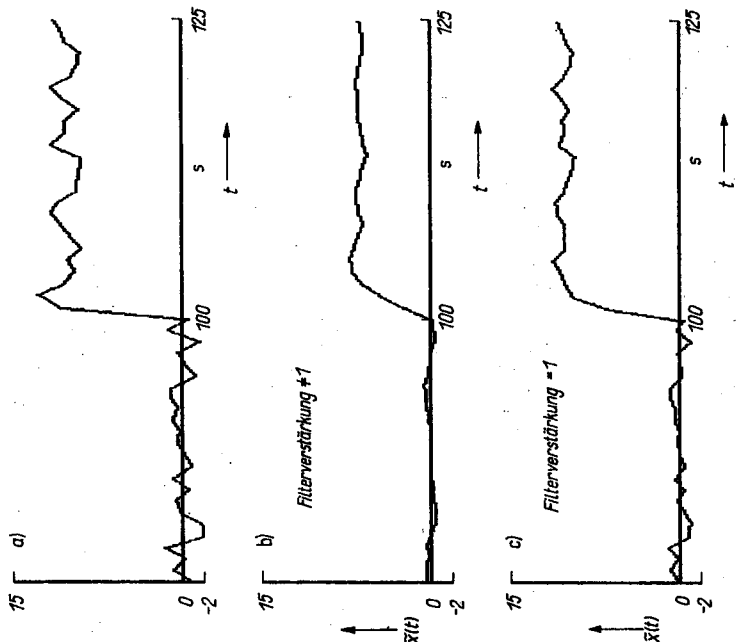
$b_0 = 0,2, \quad b_{-1} = 0,187, \quad b_{-2} = 0,151,$
 $b_{-3} = 0,100, \quad b_{-4} = 0,047, \quad b_{-5} = 9 \cdot 10^{-9},$
 $b_{-6} = -0,031, \quad b_{-7} = 0,043,$

b) Wichtung (normierte Verstärkung)

$b_0 = 0,503, \text{ alle } b_{-i} \text{ wie bei a).}$

Ergebnis:

Siehe Bild zu Beispiel 3.11.



Die Verläufe der gefilterten Signale $\tilde{x}(k)$ zeigen sehr gut das dynamische und statische Verhalten des nichtrekursiven Filters. Gegenüber dem nicht auf eine Verstärkung von 1 normierten Filter erreicht das normierte Filter bei gleichem dynamischem Verhalten den stationären Wert des ungefilterten Signals $x(t)$.

Beispiel 3.12	Wirkung des Filters „15-Punkte-Gleitmittel von Spencer“
---------------	---

Gegeben:

Signalbeschreibung

$$x(t) = a_1 \sin \frac{2\pi}{250} t + z(t),$$

$$a_1 = 4, z(t) \text{ weißes Rauschen, NV } (0, 1),$$

Filtergleichung (3.178), $n = 7$,

Abtastzeit $T = 1$ s,

Beobachtungen: 500.

Gesucht:

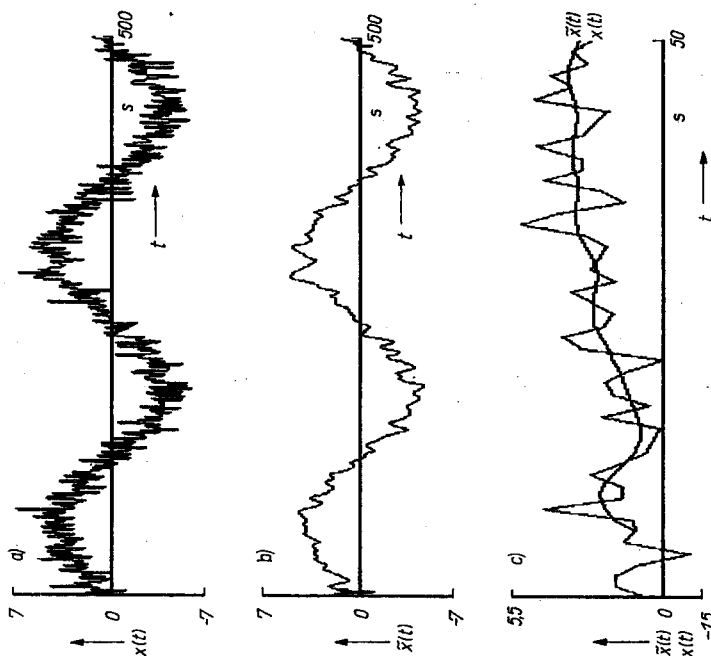
Verläufe des gefilterten Signals $\tilde{x}(k)$

a) über alle Beobachtungen,

b) als Vergleich des ungefilterten und gefilterten Signals für 50 Beobachtungen.

Ergebnis:

Siehe Bild zu Beispiel 3.12.



Die Verläufe der gefilterten Signale entsprechen prinzipiell denen vom Beispiel 3.10. Das Filter beseitigt hochfrequente Signalkomponenten, aber wesentlich besser als die einfachen nichtrekursiven Filter. In b) ist der Anfang der Signalverläufe dargestellt. Das Bild zeigt erneut, daß eine Filterung erst möglich ist, wenn $(2n + 1)$ Signalwerte (d.h. in diesem Fall 15) vorliegen. Zwischen dem gefilterten und dem ungefilterten Signalwert besteht auch hier keine Phasenverschiebung.

Der Amplituden- und der Phasenfrequenzgang für dieses nichtrekursive Filter sowie für das ideale und analoge Filter 1. Ordnung sind für einen typischen Fall im Bild 3.43 dargestellt. Es zeigt deutlich den periodischen Verlauf der Frequenzgänge auch dieses digitalen Filtertyps gegenüber dem analogen Filter. Gleichzeitig wird die Phasennacheilung des gefilterten Signals im Frequenzbereich $\omega T < \pi$ gegenüber der phasenfren Filterung bei symmetrischer Wichtung sehr gut sichtbar.

Das bedeutet, daß Filter mit $b_{+i} = 0$ für alle i zwar zur Echtzeitdatenverarbeitung gut geeignet sind, aber immer eine Phasennacheilung aufweisen. Wird ein Filter für $b_{-i} = b_i$ für alle i verwendet, ist keine Phasenverschiebung des gefilterten Signals vorhanden. Probleme treten dagegen in diesem Fall bei der Filterung der ersten und letzten n Signalwerte der betrachteten Probe sowie im Echtzeiteinsatz auf.

Den Einfluß verschiedener Wichtungsformen auf die Filterwirkung zeigt für einen Testfall Beispiel 3.10. Im Beispiel 3.11 wird der Einfluß der Normierung der Filterverstärkung auf 1 anhand eines Signalabschnitts untersucht.

Da die Filterwirkung häufig nicht ausreicht, wird die nichtrekursive Filterung in zwei oder mehr Stufen wiederholt. In [3.19] sind verschiedene Vorschläge für diese Strategie enthalten. Ein weitverbreitetes Filter ist das sog. „15-Punkte-Gleitmittel von Spencer“, das aus einer Reihenschaltung von vier linearen Filtern entwickelt wurde.

Die Gewichte für das Filter entsprechend Gl. (3.178) sind mit $b_i = b_{-i}$

$$\begin{aligned} b_0 &= 0,2312 & b_{-4} &= b_{+4} = 0,0094 \\ b_{-1} &= b_{+1} = 0,2094 & b_{-5} &= b_{+5} = -0,0156 \\ b_{-2} &= b_{+2} = 0,1437 & b_{-6} &= b_{+6} = -0,00187 \\ b_{-3} &= b_{+3} = 0,0656 & b_{-7} &= b_{+7} = -0,0094. \end{aligned}$$

Im Beispiel 3.12 ist die Wirkung dieses nichtrekursiven Tiefpaßfilters dargestellt.

Rekursive Filter

Der gefilterte Signalwert $\tilde{x}(k)$ wird bei diesem Filter rekursiv aus den vorangegangenen gewichteten gefilterten Signalwerten $x(k-j)$ und den gewichteten ungefilterten Signalwerten $x(k-i)$ nach der allgemeinen Vorschrift

$$\tilde{x}(k) = \sum_{i=0}^n b_i x(k-i) + \sum_{j=1}^m a_j \tilde{x}(k-j) \quad (3.190)$$

berechnet. Das Strukturschema dieses Filters ist im Bild 3.44 dargestellt. Die rekursiven Filter beinhalten als Sonderfall das nichtrekursive Filter von Gl. (3.178).

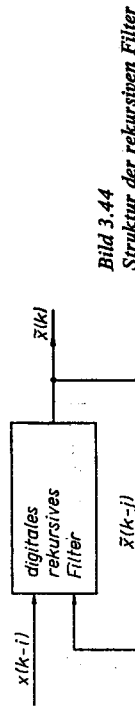


Bild 3.44
Struktur des rekursiven Filter

Im folgenden sollen die wesentlichsten Aspekte dieses Filtertyps am Beispiel eines rekursiven Filters 1. Ordnung betrachtet werden. Als Gleichungen erhält man im Zeitbereich

$$\tilde{x}(k) = b x(k) + a \tilde{x}(k-1) \quad (3.191)$$

und für die Pulsübertragungsfunktion

$$G(z) = \frac{b}{1 - az^{-1}}. \quad (3.192)$$

Zur Darstellung des Amplituden- und des Phasenfrequenzgangs wird wieder von harmonischen abgetasteten Signalen ausgegangen und somit $z = e^{j\omega T}$ gesetzt. Damit ergeben sich nach einigen einfachen Zwischenschritten folgende Beziehungen:

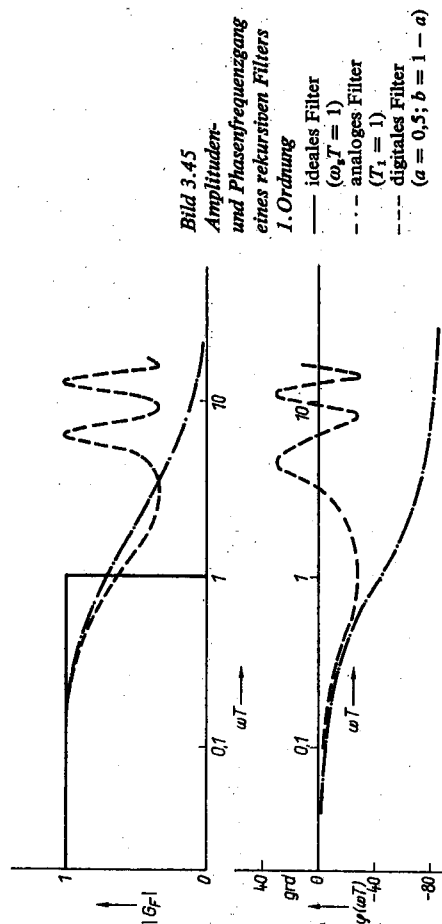
für den Amplitudenfrequenzgang

$$|G(e^{j\omega T})| = b/\sqrt{1 + a^2 - 2a \cos \omega T}, \quad (3.193)$$

für den Phasenfrequenzgang

$$\varphi(\omega T) = -\arctan \frac{a \sin \omega T}{1 - a \cos \omega T}. \quad (3.194)$$

In vielen Fällen wird durch die Wahl des Parameters $b = 1 - a$ das Übertragungsverhalten des Filters so normiert, daß es im stationären Fall die Verstärkung von 1 besitzt (s. auch Bemerkung bei nichtrekursiven Filtern). Im Bild 3.45 sind der Amplituden- und der Phasenfrequenzgang beispielhaft dargestellt im Vergleich zu einem idealen und analogen Tiefpaßfilter. Es zeigt sich wieder das durch die Abtastung bedingte periodische Verhalten (Periode $\omega T = 2\pi$) und das günstige Tiefpaßverhalten im Bereich $\omega T < \pi$. Für höhere Frequenzen ist das Filter nicht geeignet. Mit Zunahme des Wertes des Parameters a ($0 < a < 1$) erlaubt der Bereich für Tiefpaßverhalten nimmt der Tiefpaßcharakter zu, d.h., die Grenzfrequenz ω_g nimmt ab.



Wenn man die Bandbreite des Filters über die Grenzfrequenz ω_g nach der Beziehung

$$|G(e^{j\omega T})| \geq G(\omega T = 0)/\sqrt{2}; \quad 0 \leq \omega T \leq \omega_g T \quad (3.195)$$

definiert, ergibt sich aus den Gln. (3.193) und (Gl. 3.193) mit $b = 1 - a$

$$\frac{1 - a}{\sqrt{1 + a^2 - 2a \cos \omega_g T}} = \frac{1}{\sqrt{2}}.$$

Nach Umformung folgt daraus der Zusammenhang zwischen der Grenzfrequenz ω_g und dem Parameter a zu

$$a = 2 - \cos \omega_g T - \sqrt{(3 - \cos \omega_g T)(1 - \cos \omega_g T)}. \quad (3.196)$$

Beispiel 3.13 Wirkung eines rekursiven Tiefpaßfilters 1. Ordnung

Gegeben:

Signalbeschreibung

$$x(t) = a_1 \sin \frac{2\pi}{250} t + z(t),$$

$$a_1 = 0,5, z(t) \text{ weißes Rauschen, NV } (0, 1),$$

Filtergleichung (3.191),

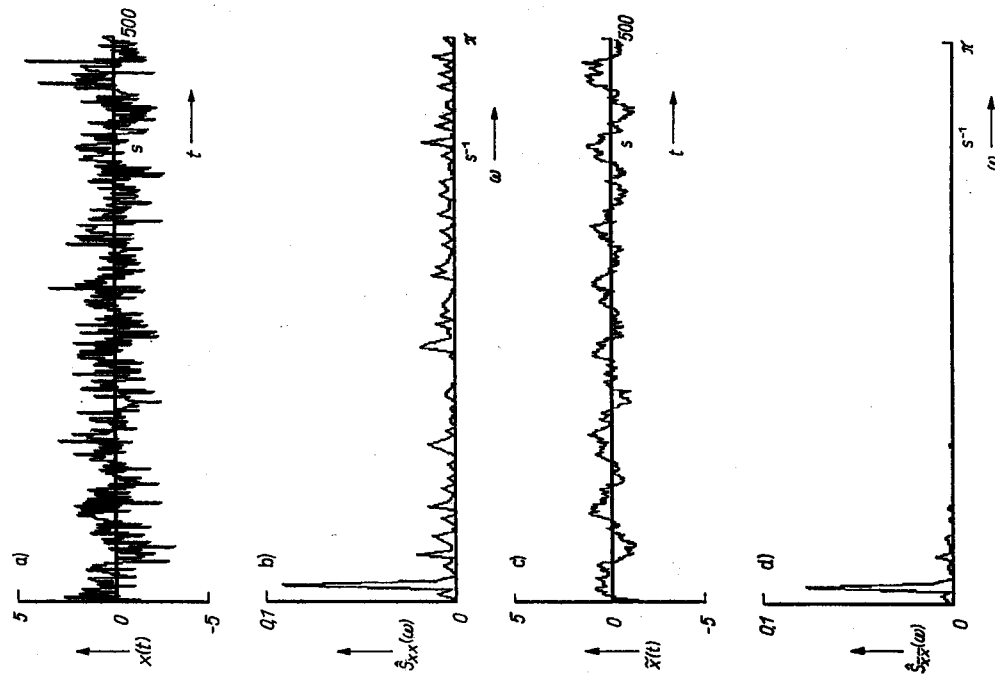
Abtastzeit $T = 1$ s,

Beobachtungen: 500.

Gesucht: Gefiltertes Signal $\hat{x}(k)$ ohne hochfrequente Anteile.

Ergebnis:

Siehe Bild zu Beispiel 3.13.



Die Ergebnisse zeigen, daß das auf einer Spektralanalyse entworfene Filter mit einer gewählten Grenzfrequenz $\omega_g = 0,196 \text{ s}^{-1}$ (daraus folgt der Filterparameter $a = 0,824$) die hochfrequenten Signale stark dämpft. Dies ist aus dem Verlauf des gefilterten Signals $\hat{x}(t)$ und seines Leistungsspektrums $\hat{S}_{\hat{x}\hat{x}}(\omega)$ deutlich zu erkennen. Eine vollständige Beseitigung der hochfrequenten Signale ohne eine starke Verfälschung des Nutzsignals ist durch ein Tiefpaßfilter 1. Ordnung nicht möglich.

Damit können aus a priori vorliegenden Kenntnissen über die Grenzfrequenz ω_z die Tastperiode T und der Filterparameter a nach Gl. (3.196) ermittelt werden. Die Wirkung eines rekursiven Tiefpassfilters 1. Ordnung ist für einen einfachen Testfall im Beispiel 3.13 dargestellt.

Eine besonders für die Echtzeitverarbeitung gut geeignete Filtervorschrift mit der Möglichkeit zur einfachen Gestaltung eines „exponentiellen Vergessens“ kann aus Gl. (3.191) gewonnen werden.

Wenn für den Parameter $b = 1 - a$ eingesetzt wird, so daß die Filterverstärkung im stationären Fall gleich 1 ist, erhält man die Filtervorschrift

$$\tilde{x}(k) = (1 - a)x(k) + a\tilde{x}(k-1) \quad \text{mit} \quad 0 \leq a \leq 1. \quad (3.197)$$

Das generelle Verhalten von Amplituden- und Phasenfrequenzgang entspricht dem Filter von Gl. (3.192). Durch die Größe des Parameters a kann das Filterverhalten sehr gut gewählt werden. Ist der Parameter $a = 0$, erfolgt keine Filterung des Signals, und ist $a = 1$, wird nur der letzte Wert weiter verarbeitet; das gefilterte Signal bleibt konstant. In Abhängigkeit von der Größe a zwischen diesen Eckwerten können die zurückliegenden abgetasteten Signalwerte $x(k-i)$ mit exponentiell abfallenden Gewichten c_i bewertet werden. Wird zum Zeitpunkt $k = 0$ die Filterung mit $\tilde{x}(-1) = 0$ begonnen, so ergibt sich aus Gl. (3.197) für den

Tastpunkt 1 ($k = 0$): $\tilde{x}(0) = 0 + (1 - a)x(0)$

Tastpunkt 2 ($k = 1$): $\tilde{x}(1) = a[(1 - a)x(0)] + (1 - a)x(1)$

Tastpunkt 3 ($k = 2$): $\tilde{x}(2) = a[a(1 - a)x(0)] + (1 - a)x(1) + (1 - a)x(2)$

$$\tilde{x}(2) = a^2(1 - a)x(0) + a(1 - a)x(1) + (1 - a)x(2).$$

Aus diesen Beziehungen ist bereits zu erkennen, daß das Gewicht im jeweiligen Tastpunkt sich aus der allgemeinen Beziehung

$$c_i = (1 - a)a^i; \quad i = 0, 1, \dots, n, \quad (3.198)$$

ergibt. Für $0 < a < 1$ erhält man damit eine typische exponentielle Wichtung (Vergessen), wie sie für einige ausgewählte Parameterwerte a im Bild 3.46 dargestellt ist. Wenn die Parameter c_i und die Signalwerte $x(k-i)$ verwendet werden, ist das gleiche Verhalten wie beim nichtrekursiven Filter der Gl. (3.178) vorhanden.

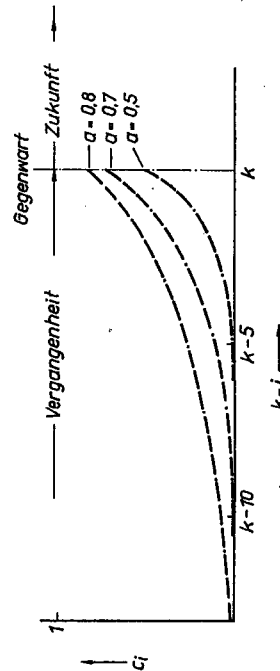


Bild 3.46
Verlauf der Wichtungsparameter c_i

Die Wirkung des in Gl. (3.197) vorgestellten Tiefpassfilters mit exponentiellem Vergessen wird als Funktion der Vergessensrate (Parameter a) im Beispiel 3.14 untersucht. Ist der Abfall des Amplitudenfrequenzgangs im Bereich oberhalb der Grenzfrequenz ω_z nicht steil genug, können mehrere Filter 1. Ordnung in Reihe oder Filter mit komple-

Beispiel 3.14 Wirkung des rekursiven Tiefpassfilters mit exponentiellem Vergessen

Gegeben:

Signalbeschreibung

$$x(t) = a_0 + a_1 \sin \frac{2\pi}{125} t + a_2 z(t),$$

$$a_0 = 1, \quad a_1 = 2, \quad a_2 = 0,5,$$

$z(t)$ weißes Rauschen, NV (0, 1),

Filtergleichung (3.197),

$$a = 0,6; 0,8; 0,95,$$

Abtastzeit $T = 1$ s,

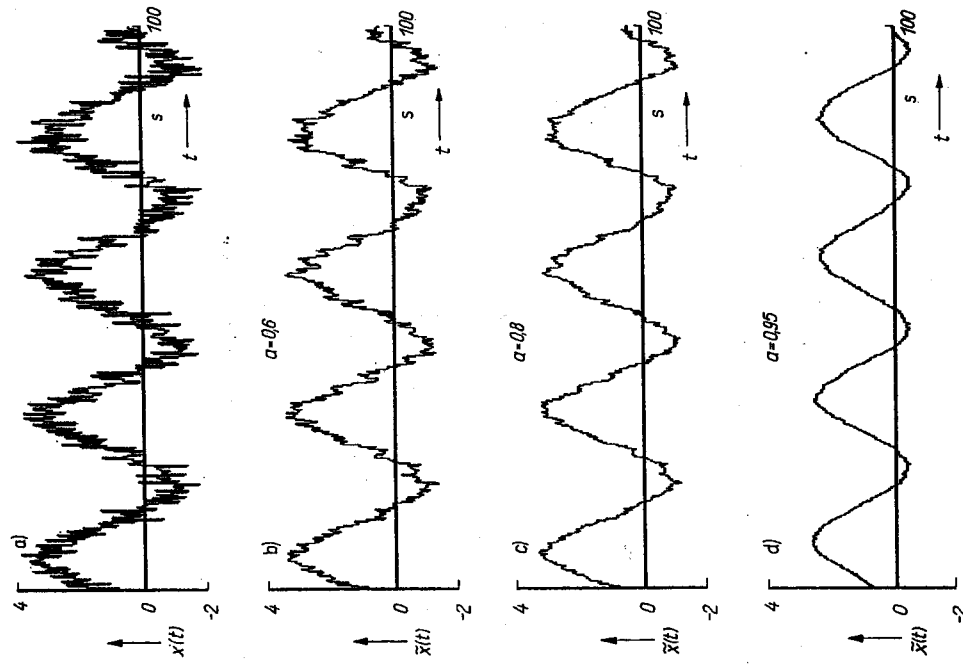
Beobachtungen: 100.

Verläufe der gefilterten Signale für verschiedene Parameter a .

Gesucht:

Ergebnis:

Siehe Bild zu Beispiel 3.14.



Die Verläufe der gefilterten Signale zeigen, daß

1. mit wachsendem Parameter a die Dämpfung der hochfrequenten Signalanteile zunimmt,
2. mit wachsendem Parameter a die Phasennäherung des gefilterten Signals gegenüber dem ungefilterten Signal zunimmt.

nen Polen in Form von Butterworth- oder Tschebyscheff-Filtern das gewünschte Verhalten erzeugen. Der Entwurf der Filter erfolgt dann zweckmäßig mit Hilfe bekannter Verfahren für kontinuierliche Systeme und anschließender Transformation in den zeitdiskreten Bereich [3.22, 3.23].

3.4.4.3. Digitale Hochpaßfilter

In vielen Systemen treten u. a. aufgrund von Alterungserscheinungen oder von Wachstums-/Sterbeprozessen tieffrequente Störungen auf, die den Nutzsignalen überlagert sind. Damit bestehen für die Signalverarbeitung folgende Aufgaben:

- Erkennen der tieffrequenten Störsignale (Analyse des Spektrums zur Festlegung der Grenzfrequenz ω_s oder zum Test auf Instationarität/Trend; s. Abschn. 3.4.2)
- Entwurf eines Hochpaßfilters oder Schätzung der Trendfunktion in Form eines Polynomansatzes (s. Abschn. 3.3.2.1)
- Filterung/Korrektur des gemessenen Signals.

Bei den weiteren Betrachtungen soll wieder die digitale Realisierung von Filtern mit Hochpaßeigenschaften im Vordergrund stehen. Zu generellen Fragen des Entwurfs im kontinuierlichen und diskreten Fall muß auf die weiterführende Literatur verwiesen werden [3.23, 3.24].

Im folgenden werden die beiden Möglichkeiten des Entwurfs eines digitalen Hochpaßfilters 1. Ordnung und der Trendermittlung und -korrektur näher betrachtet.

Hochpaßfilter 1. Ordnung

Das digitale Hochpaßfilter 1. Ordnung hat die Aufgabe, das Nutzspektrum oberhalb einer Grenzfrequenz ω_s ungedämpft zu übertragen und das unterhalb von ω_s liegende Störspektrum möglichst vollständig zu unterdrücken. Der ideale Amplitudenfrequenzgang des Filters ist in Tafel 3.6 dargestellt.

Für das digitale Hochpaßfilter 1. Ordnung, das in diesem Fall ein rekursives Filter ist, gilt analog zu Abschnitt 3.4.4.2 für den normierten Fall ($b_0 = 1 + a$) im Zeitbereich

$$\tilde{x}(k) = (1 + a)x(k) + a\tilde{x}(k-1) \quad (3.199)$$

und für die Pulsübertragungsfunktion

$$G(z) = \frac{(1 + a)}{1 - az^{-1}}. \quad (3.200)$$

Das erstrebte typische Hochpaßverhalten wird für den Parameterbereich $-1 < a < 0$ durch die Gl. (3.199) und (3.200) erzeugt. Für die Betrachtung des Frequenzverhaltens wird wiederum davon ausgegangen, daß als Eingangssignal eine harmonisch abgetastete Schwingung verwendet wird. Damit erhält man aus Gl. (3.200) den Amplitudenfrequenzgang zu

$$|G(e^{j\omega T})| = \frac{(1 + a)}{\sqrt{1 + a^2 - a \cos \omega T}} \quad (3.201)$$

und den Phasenfrequenzgang zu

$$\varphi(\omega T) = -\arctan \frac{a \sin \omega T}{1 - a \cos \omega T}$$

bei einem Parameterbereich für a von $-1 < a < 0$. Für ein Beispiel ist der typische Verlauf des Amplituden- und des Phasenfrequenzgangs im Bild 3.47 dargestellt. Es zeigt wieder das periodische Verhalten von Digitalfiltern, das durch eine geeignete Wahl von ωT ($\omega T < \pi$) beachtet werden muß. Je größer der Parameter a ist, um so besser wird der Signalanteil im Frequenzbereich unterhalb $\omega_s T$ gedämpft und der oberhalb von $\omega_s T$ mit der Verstärkung von 1 übertragen.

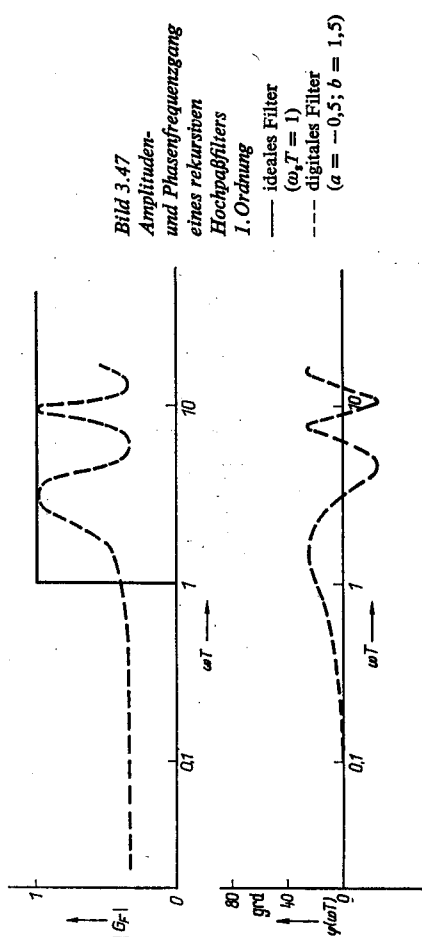


Bild 3.47
Amplituden-
und Phasenfrequenzgang
eines rekursiven
Hochpaßfilters
1. Ordnung

— ideales Filter
($\omega_s T = 1$)
--- digitales Filter
($a = -0,5$; $b = 1,5$)

Für den Zusammenhang zwischen dem Filterparameter a und der Bandbreite ω_s des Filters gilt analog zu den Betrachtungen beim Tiefpaß

$$\frac{1 + a}{\sqrt{1 + a^2 - 2a \cos \omega_s T}} = \frac{1}{\sqrt{2}} \quad \text{mit} \quad -1 < a < 0. \quad (3.202)$$

Aus der Gleichung ergibt sich dann die Beziehung für die Berechnung von a zu

$$a = -(2 + \cos \omega_s T) + \sqrt{(3 + \cos \omega_s T)(1 + \cos \omega_s T)}. \quad (3.203)$$

Für einen Testfall sind im Beispiel 3.15 der Entwurf und die Wirkung eines Hochpaßfilters 1. Ordnung dargestellt.

Reicht die Flankensteilheit des Amplitudenfrequenzgangs nicht aus, so können auch beim Hochpaßfilter mehrere Filter 1. Ordnung in Reihe geschaltet oder Butterworth- oder Tschebyscheff-Filter verwendet werden [3.23].

Trendermittlung und -korrektur

Eine Hochpaßfilterung kann algorithmisch leicht realisiert werden, wenn davon ausgegangen wird, daß ein tieffrequentes Störsignal/Trend in Form von deterministischen Signalmodellen beschrieben werden kann (s. Abschn. 3.3.2). Als zwei typische Formen seien hier der tieffrequente periodische Störeinfluß und ein tieffrequenter Störeinfluß, der mit Hilfe eines Polynomansatzes beschrieben werden kann, betrachtet (Bild 3.48).

Bei der Ermittlung der tieffrequenten Signalanteile und der anschließenden Korrektur ist die Lösung folgender Aufgaben notwendig:

1. Test auf Instationarität (s. Abschn. 3.4.2) bzw. Spektralanalyse zur Ermittlung der Frequenzen der tieffrequenten Störung (s. Abschn. 3.3.2);

Gegeben:

Signalbeschreibung

$$x(t) = a_1 \sin \frac{2\pi}{250} t + z(t),$$

 $a_1 = 0,5$, $z(t)$ weißes Rauschen, NV (0, 1),

Filtergleichung (3.199),

Abtastzeit $T = 1$ s,

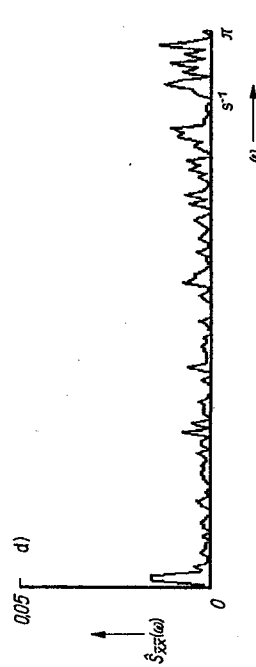
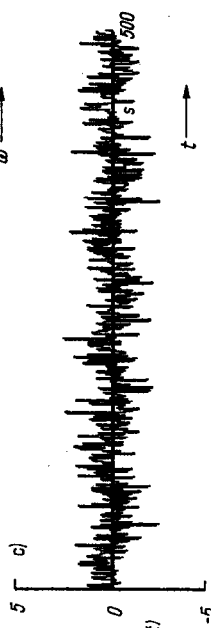
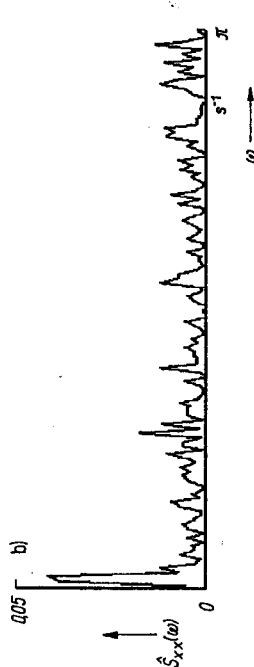
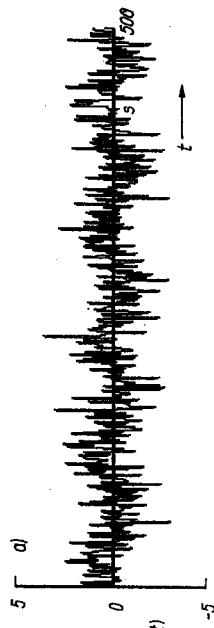
Beobachtungen: 500.

Gesucht:

Gefilterte Signalwerte $\tilde{x}(k)$ ohne tieffrequente Anteile.

Ergebnis:

Siehe Bild zu Beispiel 3.15.



Der Verlauf der gefilterten Signalwerte $\tilde{x}(k)$ und der Vergleich der Leistungsspektren des ungefilterten und des gefilterten Signals $S_{xx}(\omega)$ und $S_{\tilde{x}\tilde{x}}(\omega)$ zeigen, daß die tieffrequenten Signaleanteile stark gedämpft werden. Aus der Analyse des Leistungsdichtespektrums und von Zusatzinformationen kann die Grenzfrequenz zu $\omega_g = 1,31 \text{ s}^{-1}$ gewählt werden. Das ergibt einen Parameter a für das Hochpaßfilter entsprechend Gl. (3.203) zu $a = -0,233$. Eine vollständige Dämpfung aller tieffrequenten Signaleanteile ohne eine starke Verfälschung des Nutzsignals ist mit einem digitalen Hochpaßfilter 1. Ordnung nicht möglich.

2. Ermittlung der Signalmodelle für die tieffrequente Störung $z(t)$ in der diskreten Form

$$\hat{z}(kT) = \hat{a}_0 + \sum_{i=1}^m \hat{a}_i (kT)^i; \quad m \leq 3 \quad (3.204)$$

bzw.

$$\hat{z}(kT) = \hat{a}_0 + \sum_{i=1}^l [\hat{a}_i \cos \hat{\omega}_i (kT) + \hat{b}_i \sin \hat{\omega}_i (kT)];$$

$$l \leq 3, \quad k = 1, 2, \dots, n,$$

entsprechend den im Abschnitt 3.2.2 vorgestellten Verfahren;

3. Korrektur des gemessenen Signals nach der Vorschrift

$$\tilde{x}(kT) = x(kT) - \hat{z}(kT); \quad (3.205)$$

4. Test der Signalwerte $\tilde{x}(kT)$ auf Stationarität. Im negativen Fall muß mit Schritt 2 wieder begonnen und die Ordnung oder die Gesamtstruktur des Modells verändert werden.

Da mit dieser Strategie einzelne tieffrequente periodische Störkomponenten aus dem Signal eliminiert werden können, ist das Verhalten dem einer Bandsperrfilter äquivalent. Diese Strategie ist hinsichtlich ihrer Methodik bevorzugt im Off-line-Betrieb einzusetzen. In den Beispielen 3.4 und 3.5 ist die beschriebene Vorgehensweise für einfache Testfälle dargestellt.

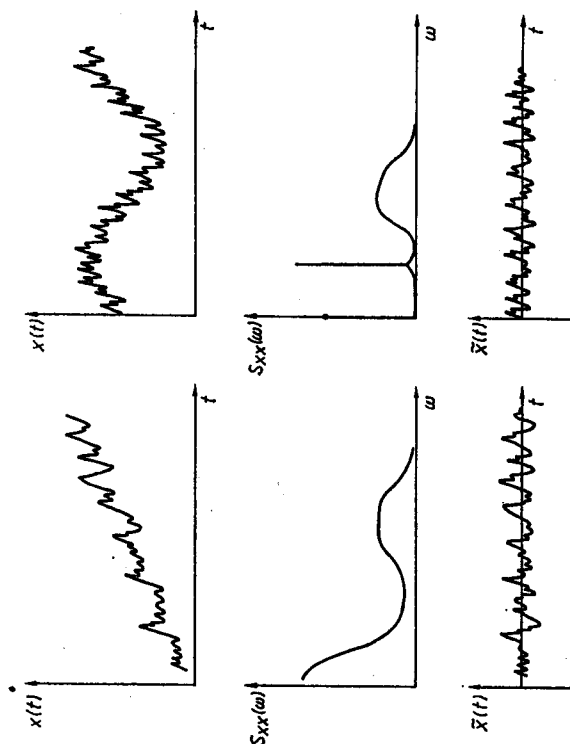


Bild 3.48. Signalverläufe und Leistungsdichtespektren tieffrequent gestörter Signale

3.5. Übungsaufgaben zum Abschnitt 3

Aufgabe 3.1

Berechnen und skizzieren Sie die Autokorrelationsfunktion für das Signal

$$x(t) = a_0 \sin(\omega_0 t + \varphi)$$

mit

$$a_0 = 311 \text{ V} \quad \text{und} \quad \omega_0 = 314 \text{ s}^{-1} \hat{=} f = 50 \text{ Hz!}$$

Berechnen Sie $U = \sqrt{R_{xx}(0)}$!

Aufgabe 3.2

Berechnen Sie die Autokorrelationsfunktion $R_{xx}(mT)$ über eine Periode ($n = 8$) der in Tafel 3.7 angegebenen diskreten Zeitfunktion!

Tafel 3.7
Signalwerte für Aufgabe 3.2

iT	1	2	3	4	5	6	7	8
$x(iT)$	0	1	2	1	0	-1	-2	-1

Aufgabe 3.3

Berechnen Sie die Autokorrelationsfunktion des in Tafel 3.8 angegebenen nichtperiodischen diskreten Signals!

Tafel 3.8
Signalwerte für Aufgabe 3.3

iT	1	2	3	4	5	6	7	8
$x(iT)$	0	1	2	-1	0	-1	-2	-1

Aufgabe 3.4

Berechnen Sie die Autokorrelationsfunktion für das in Tafel 3.9 dargestellte periodische Signal (12 Werte $\hat{=} 1$ Periode) bis zu einer Verschiebung von $m = 14T$

- a) ohne Zentrierung der Werte
b) mit Zentrierung der Werte!

Beziehen Sie die Autokorrelationsfunktion auf den Wert $R_{xx}(0)$! Skizzieren Sie den Verlauf $R_{xx}(mT)$!

Tafel 3.9. Signalwerte für Aufgabe 3.4

iT	1	2	3	4	5	6	7	8	9	10	11	12
$x(iT)$	2,0	2,4	2,5	2,4	2,0	1,5	1,0	0,6	0,5	0,6	1,0	1,5

Aufgabe 3.5

An einem reinen Totzeitsystem werden die in Tafel 3.10 angegebenen Eingangs- und Ausgangssignale gemessen.

Berechnen Sie die Kreuzkorrelationsfunktionen $R_{xx}(mT)$ für $0 \leq m \leq 9$, und ermitteln Sie die Totzeit T_1 !

Tafel 3.10
Signalwerte für Aufgabe 3.5

iT	$u(iT)$	$x(iT)$
1	-0,1	-0,2
2	-1,3	0,3
3	0	-0,6
4	-0,5	-0,1
5	-0,2	-1,3
6	-0,3	0
7	0,6	-0,5
8	-1,1	-0,2
9	1,7	-0,3
10	-0,7	0,6
11	0,2	-1,1
12	-0,5	1,7
13	2,4	-0,7
14	1,4	0,2
15	0,2	-0,5
16	-1,7	2,4
17	0,4	1,4

Aufgabe 3.6

Gegeben ist die Korrelationsfunktion

$$R_{xx}(\tau) = \frac{a^2}{2} \cos \omega_0 \tau,$$

wobei $a = 311 \text{ V}$, $\omega_0 = 314 \text{ s}^{-1} \hat{=} f = 50 \text{ Hz}$.

- a) Berechnen und skizzieren Sie das Leistungsdichtespektrum $S_{xx}(\omega)$!
b) Berechnen Sie die Leistung, wenn die Spannung $x(t)$ an einem Widerstand $R = 48,4 \Omega$ abfällt!

Aufgabe 3.7

Ein stochastischer Vorgang habe die Autokorrelationsfunktion $R_{xx}(\tau) = 9 + 25 e^{-5|\tau|}$.

Berechnen Sie

- a) den Mittelwert $\bar{x}$
b) die Standardabweichung σ_x
c) das Autoleistungsspektrum $S_{xx}(\omega)$!

Aufgabe 3.8

Bestimmen Sie mit Hilfe von Näherungsmethoden folgende Parameter des Spektrums $S_{xx}(\omega)$:

a) die Varianz σ_x^2

b) die Grenzfrequenz ω_g

c) den Gleichanteil $S_{xx}(0)$ unter der Voraussetzung, daß das Spektrum durch ein Rechteck der Länge ω_g angenähert wird!

Verwenden Sie die Wertefolge der eindimensionalen Zufallszahl in Tafel 3.11!

Das Spektrum habe in logarithmischer Darstellung einen Abfall von 80 dB/Dekade.

Die Tastzeiten betragen 0,1 s, 1 s, 10 s.

Tafel 3.11. Signalwerte für Aufgabe 3.8

$iT \longrightarrow$	0,00	-1,26	0,12	0,24	-0,50	0,91	-0,21
	-2,25	-0,25	-0,79	0,24	-0,60	-1,19	1,74
	0,28	-0,45	0,41	-0,05	-0,48	-0,11	3,36
	0,76	1,40	0,33	-0,04	-1,09	-0,52	0,37
	2,38	0,43	0,02	-2,07	-1,51	-0,73	
	-0,18	-0,28	0,61	-0,50	-1,71	-0,09	

Aufgabe 3.9

Gegeben sind die in Tafel 3.12 dargestellten vier unmittelbar nacheinander aufgenommenen Serien zu je zehn Beobachtungen einer Zufallsgröße.

a) Bestimmen Sie anhand der Serien 1 und 4

- durch Mittelwertbildung der Serien
- durch Prüfung auf Stationarität mit dem F -Test (bezüglich σ) sowie mit dem t -Test (bezüglich μ), ob ein Trend vorliegt (bei $\alpha = 0,05$)!

b) Schätzen Sie für die Trendfunktion $\hat{x}(t) = at$ den Parameter a ab!

c) Führen Sie eine Trendkorrektur durch, und tragen Sie die korrigierten Serien auf!

d) Zu den korrigierten Meßreihen liegt aus weiteren Messungen eine Schätzung $\sigma = 7$ vor. Es kann angenommen werden, daß die Grundgesamtheit normalverteilt ist.

Bestimmen Sie für $\alpha = 0,3\%$, welche Meßwerte als Fehlmessungen in Frage kommen!

Tafel 3.12. Signalwerte für Aufgabe 3.9

Serie					
1	2		3		4
t/s	$x(t)$	t/s	$x(t)$	t/s	$x(t)$
1	3,3	11	1,1	21	9,0
2	-	12	11,7	22	11,7
3	9,9	13	10,1	23	4,2
4	-13,4	14	1,7	24	4,4
5	-	15	8,6	25	2,9
6	7,1	16	17,0	26	36,3
7	0,4	17	-22,1	27	13,0
8	22,4	18	16,2	28	13,8
9	14,5	19	21,7	29	15,6
10	-1,6	20	9,3	30	14,6
				31	2,0
				32	15,0
				33	13,4
				34	35,6
				35	27,8
				36	12,5
				37	18,6
				38	22,0
				39	10,3
				40	36,6

Aufgabe 3.10

Gegeben sind die in Tafel 3.13 angegebenen Werte einer abgetasteten Zufallsgröße (Abtastzeit $T = 1$ s).

a) Schätzen Sie die Parameter des Trendverlaufs für den Modellansatz $\hat{x}(kT) = a_0 + a_1 kT + a_2 (kT)^2$!

b) Überprüfen Sie die Signifikanz der Parameter für eine Irrtumswahrscheinlichkeit $\alpha = 0,05$!

c) Wie lautet die T -Matrix für den gleichen Modellansatz, jedoch mit einer Abtastzeit $T = 0,5$ s?

Tafel 3.13
Signalwerte für Aufgabe 3.10

t/s	1	2	3	4	5	6
$x(t)$	6,1	11,6	20,4	29,9	41,8	56,2

Aufgabe 3.11

Bestimmen Sie aus den in Tafel 3.14 angegebenen abgetasteten Signalwerten (Abtastzeit $T = 1$ s) die Parameter des Signalmodells

$$\hat{x}(kT) = a_0 + a_1 \sin \omega kT + a_2 \cos \omega kT!$$

Die Frequenz ω wurde aus einer Spektralanalyse zu $\omega = 0,523 \text{ s}^{-1}$ bestimmt.

Tafel 3.14
Signalwerte für Aufgabe 3.11

t/s	$x(t)$
0	0,55
1	0,66
2	1,1
3	1,1
4	0,64
5	-0,02
6	-0,57
7	-1,09
8	-0,99
9	-1,17
10	-0,56
11	-0,1

Aufgabe 3.12

Gegeben sind die in Tafel 3.15 dargestellten Meßwerte eines mit einer Abtastzeit von $T = 1$ s abgetasteten Signals $x(t)$.

Tafel 3.15. Signalwerte für Aufgabe 3.12

t/s	$x(t)$	t/s	$x(t)$	t/s	$x(t)$	t/s	$x(t)$
0	0	9	10,5	18	-1,5	27	-9,5
1	1,24	10	10,35	19	-0,24	28	-10,85
2	5,42	11	8,4	20	-3,42	29	-9,4
3	4,0	12	8,66	21	-6,5	30	-10,66
4	7,93	13	7,16	22	-6,93	31	-5,66
5	7,16	14	4,93	23	-5,66	32	-5,93
6	7,66	15	6,5	24	-9,66	33	-4,5
7	7,4	16	4,42	25	-7,9	34	-4,42
8	11,85	17	1,24	26	-9,85	35	-2,24
				36		36	1,5

Filtern Sie die Meßwertfolge

a) mit dem nichtrekursiven Algorithmus

$$\hat{x}(k) = \frac{\omega_p T}{\pi} \sum_{i=-n}^n \frac{\sin(i\omega_p T)}{i\omega_p T} x(k-i) \quad \text{für } \omega_p = 0,6 \text{ s}^{-1} \quad \text{und } n = 3,$$

b) mit dem rekursiven Algorithmus

$$\hat{x}(k) = \alpha \hat{x}(k-1) + (1-\alpha) x(k) \quad \text{für } \alpha = 0,8!$$

Zeichnen Sie $x(k)$ und $\hat{x}(k)$ für die Varianten a) und b) jeweils in ein Diagramm, und diskutieren Sie die Ergebnisse!

4. Bestimmung des dynamischen Verhaltens ungestörter Systeme

4.1. Systemeigenschaften, Systemmodelle und Testsignale

4.1.1. Voraussetzungen für die Anwendbarkeit deterministischer Methoden

Die in den folgenden Abschnitten vorgestellten Methoden zur Bestimmung des dynamischen Verhaltens gehen davon aus, daß das System durch ein vorwiegend determiniertes Eingangssignal $u(t)$ erregt und das ungestörte Ausgangssignal $x(t)$ gemessen wird. Die Auswertung beider Signale ermöglicht die Ermittlung eines Modells des Systems. Für die deterministischen Methoden zur Modellbestimmung müssen in der Regel folgende Voraussetzungen an das System erfüllt sein, um eine ausreichende Modellgüte zu erzielen:

1. Das System ist ungestört/gering gestört.
2. Das Systemverhalten ist linear.
3. Das Systemverhalten ist zeitinvariant.
4. Es handelt sich um Systeme mit einer Eingangs- und einer Ausgangsgröße.
5. Das Systemverhalten ist durch konzentrierte Parameter beschreibbar.

Unter gering gestört wird ein Niveau der Störung kleiner als 10 % des Ausgangssignals angesehen. Das Ziel der weiteren Ausführungen besteht nun darin, Strategien zur Bestimmung von dynamischen Systemmodellen vorzustellen, die es gestatten, eine günstige Zuordnung von gewünschter Systembeschreibung, möglichem Testsignal und zweckmäßiger Auswertemethode zu treffen.

4.1.2. Verwendete Systemmodelle

Das Ziel der verschiedenen Methoden zur Modellbestimmung des dynamischen Verhaltens besteht in der Ermittlung

- von *nichtparametrischen* (grafisch-numerischen) Systembeschreibungen in Form von Gewichts-, Übergangs- oder Frequenzgangfolgen

$$g(t_i), h(t_i), G(j\omega_i); \quad i = 1, 2, \dots, N$$

oder

- von *parametrischen* Systembeschreibungen in Form von analytischen Ausdrücken für die Gewichtsfunktion $g(t)$, die Übergangsfunktion $h(t)$ oder den Frequenzgang $G(j\omega)$ oder

- der *Systemzustände* durch einen Beobachter, bei gegebener Zustands- und Ausgangsgleichung des Systems.

Die möglichen und praktisch bewährten Wege zur Erstellung der genannten Systemmodelle sind im Bild 4.1 in einer Übersicht dargestellt. Viele der Verfahren zeichnen sich dabei durch ihre einfache geräte technische Realisierung und Auswertemethodik aus. Sie

haben auch nach Einführung der rechnergestützten Methoden (s. Abschnitte 5 und 6) nicht an Bedeutung verloren und dienen den leistungsstärkeren Methoden u. a. zur Festlegung der Tasterperiode T und von Startwerten als vorausgehender Analyseschritt.

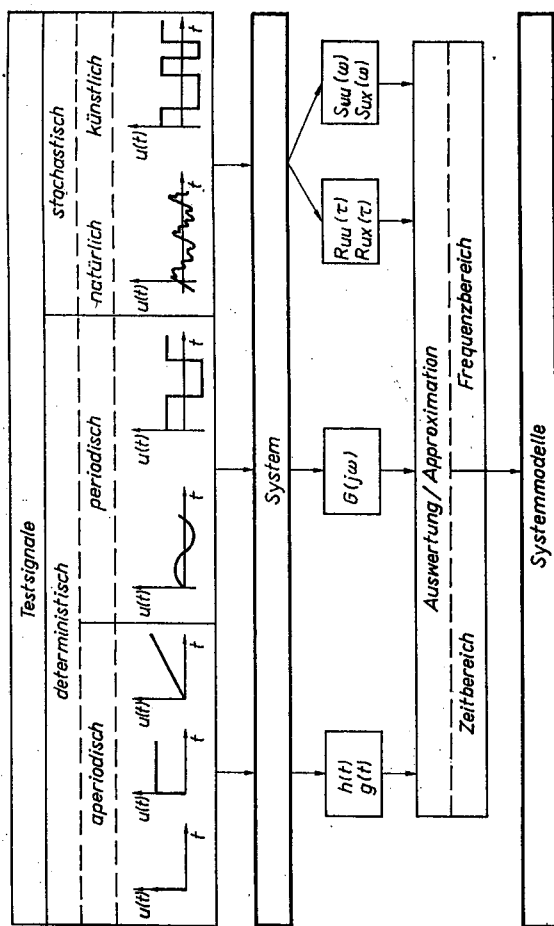


Bild 4.1. Zusammenstellung verschiedener Testsignale, Auswertemethoden und Systemmodelle

4.1.3. Ausgewählte Testsignale

Für die Modellbestimmung können als Testsignale verwendet werden (s. Bild 4.1): *Aperiodische Testsignale*. Zu dieser Gruppe von Testsignalen gehören u. a. die Sprung-, Impuls- und Rampenfunktion. Die Testsignale sind häufig einfach zu realisieren; die notwendige Beobachtungszeit des Systems ist gegenüber anderen Verfahren gering, und die Auswerteverfahren sind relativ einfach (s. Abschn. 4.2).

Periodische Testsignale. Als Signale dieser Gruppe sind sinusförmige oder andere periodische (z. B. Rechtecksignale) Testsignale anzusehen. Ihre Realisierung ist aufwendiger als die der aperiodischen Testsignale. Die Beobachtungszeit des Systems wird aufgrund der notwendigen Einschwingzeit für jede Frequenz und der Messungen für jede Frequenz ω , groß.

Dafür liefern die Auswertemethoden eine hohe Modellgüte, und sie sind relativ unempfindlich gegenüber Störungen (Korrelationsprinzip).

Stochastische und pseudostochastische Testsignale. Stochastische Signale können im normalen Betrieb am Systemeingang auftreten oder künstlich, u. a. in Form von Folgen (PRBS-, Plackett-Burman-, m -Folgen; s. Abschn. 6.4) aufgebracht werden. Die Verwendung der normal im Betrieb auftretenden stochastischen Schwankungen hat den Vorteil, daß das System nicht gestört werden muß. Als entscheidender Nachteil muß aber häufig die zu geringe Erregung des Systems in den das System charakterisierenden Frequenzbereichen angesehen werden. Daraus resultieren sehr lange Beobachtungszeiten. Die Auswertung des gemessenen Eingangs- bzw. Ausgangssignals kann über die Entfaltung (siehe Abschn. 4.2.3) oder die Berechnung der Korrelationsfunktionen $R_{uu}(\tau)$ und $R_{uv}(\tau)$

(s. Abschn. 3.2.2) im Zeitbereich oder durch die Berechnung der Leistungsdichtespektren $S_{uu}(\omega)$ und $S_{xx}(\omega)$ (s. Abschn. 3.2.3) im Frequenzbereich erfolgen. Der Auswerteaufwand steigt gegenüber den ersten beiden Testsignalarten weiter an.

Die allgemeinen Zusammenhänge zwischen den Testsignalen, der auszuwertenden Systemantwort und den möglichen Systembeschreibungen sind im Bild 4.1 dargestellt.

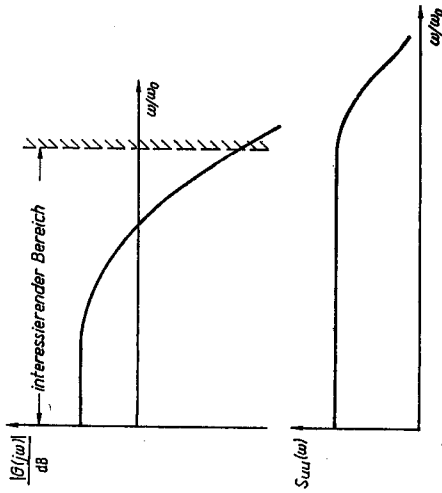


Bild 4.2
Zusammenhang
zwischen dem interessierenden
Frequenzbereich des Systems
und dem Leistungsspektrum
eines idealen Testsignals

Für die Auswahl eines geeigneten Testsignals aus den oben angeführten Grundtypen muß neben dem Aspekt der einfachen Realisierbarkeit und dem Meß- und dem Auswerteaufwand auch der Gesichtspunkt der Modellgüte betrachtet werden. Grundsätzlich ist ein System durch ein Modell nur dann hinreichend gut nachbildbar, wenn das Leistungsdichtespektrum des Testsignals mindestens den Frequenzbereich des Systems anregt, der für die Lösung des vorliegenden Problems notwendig ist (Bild 4.2). Da dieser Bereich häufig nicht bekannt ist, wird ein Testsignal, das die Eigenschaften des „weißen Rauschens“ besitzt, bezüglich der Modellgüte Vorteile aufweisen. Betrachtet man das Amplitudenspektrum der verschiedenen Testsignale (s. Abschn. 3.1, Tafel 3.2), so kann für die Analyse des dynamischen Verhaltens folgende Rangfolge angegeben werden:

1. Dirac-Impuls, periodische Signale mit den Frequenzen ω_i ; $i = 1, 2, \dots$, stochastische Signalfolgen
2. Impulsfunktion
3. Sprungfunktion
4. Rampenfunktion.

Außer bei den Signalen vom Rang 1 nimmt bei allen anderen Testsignalen die Leistung mit der Frequenz ab. Als Kompromiß aufgrund der Anregung des Systems haben sich die Impulsfunktion, die Sprungfunktion, periodische Signale und stochastische Signalfolgen durchgesetzt.

Für die Bestimmung des statischen Verhaltens, d. h. den tieffrequenten Bereich des Systems, ist eindeutig die Sprungantwort (s. Abschn. 4.2.1) zu bevorzugen. Den proportionalen Übertragungsfaktor K_p von Systemen mit Ausgleich [3.1] erhält man aus dem stationären Wert der Sprungantwort $x(\infty)$ (Bild 4.3):

$$h(\infty) = \frac{x(\infty)}{u_0} = \lim_{p \rightarrow 0} p \frac{1}{p} K_p \frac{\prod_{n=1}^m (1 + pT_{Dn})}{\prod_{i=1}^n (1 + pT_i)} = K_p. \quad (4.1)$$

Für Systeme ohne Ausgleich erhält man analog den integralen Übertragungsfaktor K_I aus der Asymptote, in die die Sprungantwort für große t einmündet (Bild 4.3):

$$h(\infty) = \frac{x(\infty)}{u_0} = \lim_{p \rightarrow 0} p^2 \frac{1}{p} \frac{K_I}{p} \frac{\prod_{n=1}^m (1 + pT_{Dn})}{\prod_{i=1}^n (1 + pT_i)} = K_I. \quad (4.2)$$

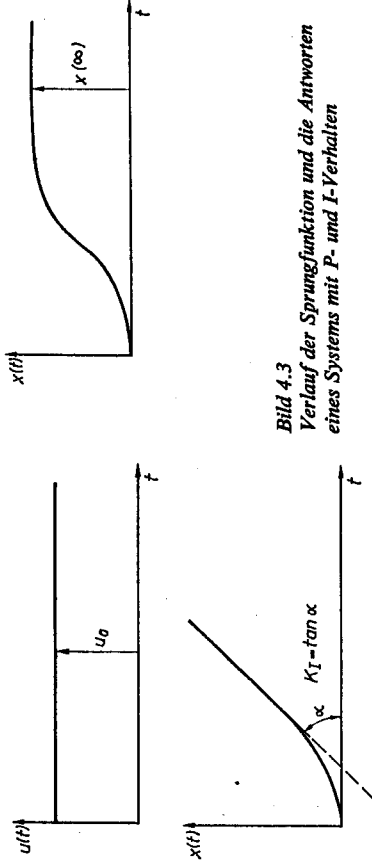


Bild 4.3
Verlauf der Sprungfunktion und die Antworten
eines Systems mit P- und I-Verhalten

Damit wird an dieser Stelle bereits deutlich, daß eine Kombination verschiedener Testsignale in den Varianten

- Sprungfunktion – Impulsfunktion
- Sprungfunktion – periodische Signale
- Sprungfunktion – stochastische Signalfolgen

zur effektiven Ermittlung der den tiefen und hohen Frequenzbereich des Systems beschreibenden Parameter verwendet werden sollte. Weiterführende Betrachtungen sind in [1.37, 1.40, 1.46] enthalten. Eine andere Möglichkeit besteht in der kombinierten Auswertung der Sprungantwort in der offenen und geschlossenen Kette, weil der geschlossene Kreis eine Anhebung der Leistung des Testsignals im höheren Frequenzbereich ermöglicht [1.31].

In den folgenden Abschnitten werden einige typische und in der Praxis bewährte Verfahren für die Bestimmung von dynamischen Systemmodellen im Zeitbereich (Abschnitt 4.2), im Frequenzbereich (Abschn. 4.3) und als Zustandsbeschreibung (Abschnitt 4.4) vorgestellt.

4.2. Ein-/Ausgangs-Modelle im Zeitbereich

4.2.1. Modellbestimmung anhand von Kenngrößen der Sprungantwort bzw. der Übergangsfunktion

Ausgangspunkt aller im weiteren betrachteten Verfahren ist die aufgenommene Sprungantwort $x(t)$ bzw. die normierte Sprungantwort, die Übergangsfunktion $h(t)$. Die Ermittlung der Parameter des Modells aus der Sprungantwort erfolgt bei den hier vorgestellten Verfahren anhand von Kenngrößen des Verlaufs der Systemantwort, z. B. %-Werte

von $h(t)$, Wendepunktkoordinaten, wobei in der Mehrzahl der Methoden ein monotoner Verlauf der Sprungantwort vorausgesetzt wird. Bei nichtmonotonen Sprungantworten müssen zusätzlich Vorhalt- oder Schwingungsglieder in den durch Trägheitsglieder gebildeten allgemeinen Ansatz aufgenommen werden. Selbst bei monoton verlaufenden Systemantworten mit reinem Trägheitsverhalten ist es schwierig, mehr als drei unterschiedliche Zeitkonstanten zu ermitteln. Deshalb wurde von Strejc [4.1] ein Modell mit n gleichen Zeitkonstanten und von Radtke [4.2] ein Modell mit n harmonisch gestaffelten Zeitkonstanten vorgeschlagen. Für viele Aufgaben des Steuerungsentwurfs sind diese Approximationen ausreichend.

Neben der Bestimmung der statischen Übertragungsfaktoren K_0 (im folgenden immer mit K bezeichnet) und K_1 spielt die Ermittlung der Totzeit T_t eines Systems eine große Rolle. Sie kann bei der Auswertung von Sprungantworten noch relativ gut aus der Ermittlung der Zeitdifferenz vom Sprungfunktionsbeginn bis zur Anstiegsänderung $dx(t)/dt \neq 0$ des Ausgangssignals gewonnen werden. Für Systeme höherer Ordnung nimmt die Genauigkeit mit der Zunahme der Ordnung ab. Dann ist eine günstig gewählte Totzeit T_t gegeben, wenn der restliche Teil der Sprungantwort/Übergangsfunktion durch reine Trägheitsglieder gut approximiert werden kann.

Für folgende Systemmodelle werden Methoden zur Struktur- und/oder Parameterbestimmung vorgestellt, die teilweise auf heuristischen Elementen beruhen:

1. $P-T_1$ - und $P-T_1-T_{D1}$ -System,
2. $P-T_n-T_1$ -System
2 unterschiedliche Zeitkonstanten
 n gleiche Zeitkonstanten
3. $I-T_n$ -System,
 n gestaffelte Zeitkonstanten,
4. $G(j\omega)$ aus der Übergangsfunktion $h(t)$.

4.2.1.1. Modellbildung von $P-T_1$ - und $P-T_1-T_{D1}$ -Systemen

Das Ziel dieser Methoden besteht darin, aus der gemessenen Sprungantwort $x(t)$ bzw. Übergangsfunktion $h(t)$ die Parameter der Übertragungsfunktion $G(p)$ zu bestimmen. Als Systemstrukturen werden angenommen

$$G(p) = \frac{K}{(1 + pT_1)} \quad \text{bzw.} \quad G(p) = \frac{K(1 + pT_{D1})}{(1 + pT_1)}. \quad (4.3)$$

Für die Parameterbestimmung des $P-T_1$ -Systems gilt:

Verstärkung K

Aus der Übergangsfunktion

$$x(t) = u_0 K (1 - e^{-t/T_1}) \quad (4.4)$$

folgt mit $t \rightarrow \infty$ für den Parameter K

$$K = \frac{x(\infty)}{u_0}. \quad (4.5)$$

Zeitkonstante T_1

Wegen der 1. Ordnung – Gl. (4.3) – kann der Parameter T_1 entsprechend Bild 4.4 aus folgenden Punkten der Übergangsfunktion bestimmt werden:

- I. $h(T_1) = 0,632$
- II. $h(1,2T_1) = 0,7$
- III. aus Subtangente $s(t_0)$ an der Stelle t_0 .

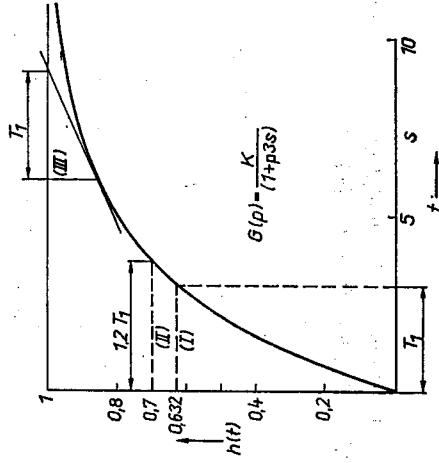


Bild 4.4
Auswertung einer Übergangsfunktion
eines $P-T_1$ -Systems

Die Variante III ergibt sich aus der Konstruktion der Subtangente $s(t_0)$ an einer beliebigen Stelle t_0 der Übergangsfunktion $h(t)$ (s. Bild 4.3). Aus der Gleichung für die Tangente an die Übergangsfunktion im Zeitpunkt $t = t_0$

$$h_T(t) = h(t_0) + \dot{h}(t_0)(t - t_0) \quad (4.6)$$

mit

$$\begin{aligned} h(t_0) &= 1 - e^{-t_0/T_1} \\ \dot{h}(t_0) &= \frac{1}{T_1} e^{-t_0/T_1}, \quad h(t) = \frac{h(t) - h(0)}{h(\infty) - h(0)} \end{aligned}$$

und deren Schnittpunkt t_s mit der Asymptote $h(\infty) = 1$ erhält man

$$h_T(t_s) = 1 = 1 - e^{-t_0/T_1} + \frac{t_s - t_0}{T_1} e^{-t_0/T_1} \quad (4.7)$$

und daraus für die Subtangente

$$s(t_0) = t_s - t_0 = T_1. \quad (4.8)$$

Die Beziehung von Gl. (4.8) gilt auch für unnormierte Übergangsfunktionen von $P-T_1$ - und T_1-T_{D1} -Gliedern.

Für die Parameterbestimmung des $P-T_1-T_{D1}$ -Systems gilt

Verstärkung K : nach Gl. (4.5).

Zeitkonstante T_1 : aus der Subtangente an der Stelle t_0 (Bild 4.5).

Vorhaltzeitkonstante T_{D1} : Aus Gl. (4.3) erhält man für den Anfangswert

$$x(0) = u_0 K \frac{T_{D1}}{T_1}. \quad (4.9)$$

Damit kann die Vorhaltzeitkonstante T_{D1} aus

$$T_{D1} = T_1 h(0) = T_1 \frac{x(0)}{u_0} \quad (4.10)$$

bestimmt werden (Bild 4.5).

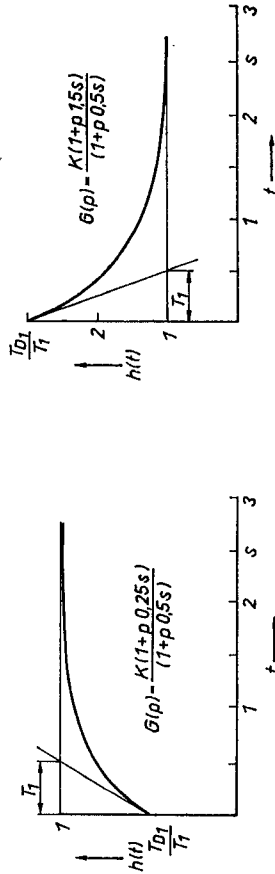


Bild 4.5. Auswertung der Übergangsfunktionen eines $P-T_1-T_{D1}$ -Systems

4.2.1.2. Modellbestimmung von $P-T_1-T_2$ -Systemen

Hat die Übergangsfunktion $h(t)$ einen Wendepunkt, so muß eine Approximation durch eine $P-T_1-T_2$ -Struktur, $n = 2$ vorgenommen werden. Die Bestimmung der Parameter

- Verstärkung K (analog zu Abschn. 4.1.3)
- Totzeit T_1 (Bild 4.6)

erfolgt für alle Verfahren und Modellstrukturen gleich.

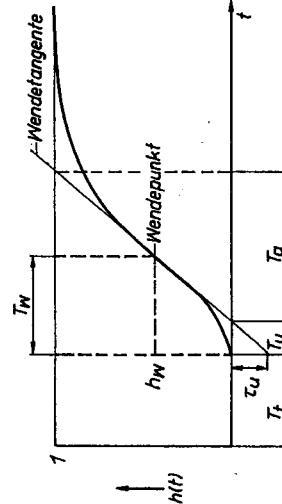


Bild 4.6
Bestimmung der Kenngrößen
der Übergangsfunktion
eines $P-T_1-T_2$ -Systems

Im Rahmen dieses Buches sollen zur Struktur- und Parametermittlung folgende bewährte Modellansätze verwendet werden:

1. Approximation durch zwei unterschiedliche Zeitkonstanten; s. Gl. (4.11)
2. Approximation durch n gleiche Zeitkonstanten; s. Gl. (4.12)
3. Approximation durch n harmonisch gestaffelte Zeitkonstanten; s. Gl. (4.17).

Mit diesen drei Strukturen kann das Verhalten vieler Systeme hinreichend genau beschrieben werden.

Approximation durch zwei unterschiedliche oder n gleiche Zeitkonstanten

Das von Stretjck [4.1] entwickelte Verfahren gestattet die Approximation des Systems durch

$$G(p) = \frac{K e^{pT_1}}{(1 + pT_1)(1 + pT_2)} \quad (4.11) \quad \text{Modell I}$$

$$G(p) = \frac{K e^{-pT_1}}{(1 + pT)^n} \quad (4.12) \quad \text{Modell II}$$

Die Entscheidung über die Wahl der jeweiligen Modellstruktur wird auf der Grundlage der Koordinaten T_W und h_W des Wendepunktes sowie der Werte von τ_u , T_1 und T_2 , die sich aus der Tangente durch den Wendepunkt gemäß Bild 4.6 ergeben, getroffen.

Danach wird entsprechend den Werten in Tafel 4.1 der in Gl. (4.11) oder in Gl. (4.12) angegebene Modelltyp gewählt.

Tafel 4.1
Grenzen der Wendepunktparameter
zur Wahl des Modelltyps

Modelltyp	h_W	τ_u
I	$\leq 0,264$	$\leq 0,104$
II	$> 0,264$	$> 0,104$

Die Parameter T_1 und T_2 des Modells I werden entsprechend Bild 4.7 aus den Kenngrößen $h(t_{0,7})$, $T_{0,7}$, $T_{0,7/4}$ und $h(t_{0,7/4})$ bestimmt.

Nach Untersuchungen von [4.1] ist der Wert $h(t_{0,7}) = 0,7$ (Bild 4.7) nahezu unabhängig vom Verhältnis T_2/T_1 und stark abhängig von der Summenzeitkonstante. Es gilt:

$$T_{0,7} = 1,2(T_1 + T_2) \quad \text{mit einem Fehler } \leq 1,7\% \quad (4.13)$$

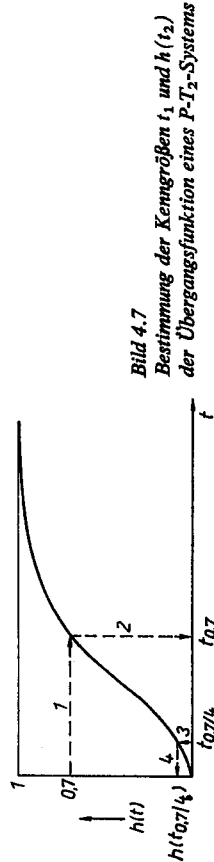


Bild 4.7

Bestimmung der Kenngrößen t_1 und $h(t_2)$
der Übergangsfunktion eines $P-T_2$ -Systems

Auf der anderen Seite ist der Funktionswert $h(t_{0,7/4})$ maximal abhängig vom Verhältnis T_2/T_1 . Mit (4.11) findet man

$$T_1 = \frac{t_{0,7}}{1,2(1 + T_2/T_1)} \quad (4.14)$$

Das Verhältnis T_2/T_1 wird aus Tafel 4.2 entsprechend dem Wert $h(t_{0,7/4})$ ermittelt.

Ist das Zeitkonstantenverhältnis T_1/T_2 groß, wird der Verlauf der Übergangsfunktion $h(t)$ für größere Zeit fast ausschließlich durch die Zeitkonstante T_1 bestimmt, da der

Tafel 4.2
Ermittlung des Verhältnisses T_2/T_1

$h(t_{0,7/4})$	T_2/T_1
0,260	0
0,200	0,1
0,174	0,2
0,150	0,3
0,135	0,4
0,131	0,5
0,126	0,6
0,125	0,7
0,124	0,8
0,123	0,9
0,123	1

durch die Zeitkonstante T_2 beeinflusste Teilvorgang bereits abgeklungen ist. Dies wird sofort bei der Betrachtung der Gleichung für die Übergangsfunktion $h(t)$ deutlich. Es gilt für diesen Modelltyp:

$$h(t) = 1 - h_1(t) + h_2(t) \quad (4.15)$$

mit

$$h_1(t) = \frac{T_1}{T_1 - T_2} e^{-t/T_1}, \quad h_2(t) = \frac{T_2}{T_1 - T_2} e^{-t/T_2}.$$

Für $t > (2 \dots 5) T_2$, d. h. im Endbereich der Übergangsfunktion, ist nur noch der Einfluß von $h_1(t)$, d. h. der Zeitkonstanten T_1 , vorhanden. Damit ergibt sich eine zweite einfache Möglichkeit der Bestimmung beider Zeitkonstanten in folgenden Schritten:

1. Ermittlung von $T_1 + T_2$ entsprechend Gl. (4.13)
2. Ermittlung von T_1 aus der Subtangente für größere Zeiten t entsprechend Abschnitt 4.2.1.1.

Liegen die Parameter τ_u und h_w des Wendepunktes im Wertebereich für das Modell II, so sind die Parameter n und T' aus Gl. (4.12) zu bestimmen. Für den Fall von n gleichen Zeitkonstanten wurde bewiesen, daß der Ordinatenwert τ_u von der Ordnungszahl n abhängt. Damit gilt die Beziehung $n = f(\tau_u)$, die in Tafel 4.3 dargestellt ist.

Tafel 4.3. Kenngrößen der Übergangsfunktion für P - T_n -Systeme (n gleiche Zeitkonstanten)

n	τ_u	h_w	T_w/T'	T_u/T'	T_u/T'
1	0	0	0	0	1
2	0,104	0,264	1	0,282	2,718
3	0,218	0,323	2	0,805	3,695
4	0,319	0,353	3	1,425	4,463
5	0,410	0,371	4	2,100	5,119
6	0,493	0,384	5	2,811	5,699
7	0,570	0,394	6	3,549	6,226
8	0,642	0,401	7	4,307	6,711

Die Zeitkonstante T' wird für ein gewähltes n entsprechend Tafel 4.3 aus den Kenngrößen T_u/T' , T_w/T' oder T_u/T' (s. Bild 4.6) ermittelt. Die nicht für die Bestimmung von n benutzten Kenngrößen können zur Kontrolle bzw. Korrektur verwendet werden. Die in Tafel 4.3 dargestellten Zusammenhänge sind durch analytische Berechnung aus vorgegebenem bekanntem Systemverhalten gewonnen worden [4.1].

Für die Übergangsfunktion $h(t)$ des Systems von Gl. (4.12) erhält man den Ausdruck

$$h(t) = 1 - \left[\sum_{s=0}^{n-1} \frac{(t/T')^s}{s!} \right] e^{-t/T'} \quad (4.16)$$

Aus der Wendetangentenkonstruktion erhält man dann u. a. folgende Beziehungen, die zahlenmäßig für $n = 1 \dots 10$ in Tafel 4.3 dargestellt sind:

$$T_u/T' = \frac{(n-2)!}{(n-1)^{n-2}} e^{n-1},$$

$$T_w/T' = n - 1 - \frac{(n-2)!}{(n-1)^{n-2}} \left[e^{n-1} \sum_{s=0}^{n-1} \frac{(n-1)^s}{s!} \right].$$

Da in praktischen Fällen in der Regel das System ungleiche Zeitkonstanten besitzt, liegen die Werte für die Kenngrößen τ_u und h_w zwischen den in Tafel 4.3 angegebenen. In diesen Fällen ist die Ordnung n zu wählen, die den Werten τ_u bzw. h_w am nächsten ist.

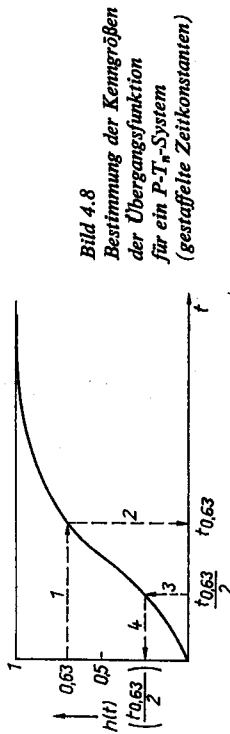


Bild 4.8
Bestimmung der Kenngrößen
der Übergangsfunktion
für ein P - T_n -System
(gestaffelte Zeitkonstanten)

Approximation durch ein Modell mit harmonisch gestaffelten Zeitkonstanten

Kann mit dem Modell von Gl. (4.11) bzw. Gl. (4.12) keine ausreichende Approximation des Systemverhaltens erreicht werden, ist eine stärkere Staffelung der Zeitkonstanten des Systems zu vermuten. Ein Modell, das sich leicht beschreiben läßt und zu einem einfachen Auswertverfahren führt, wurde von Radtke [4.2] entwickelt. Das Modell hat die Form

$$G(p) = \frac{K}{\prod_{j=1}^n \left(1 + p \frac{T'_j}{T_j} \right)} \quad (4.17)$$

Es wurde nachgewiesen, daß bei diesem Modell zwischen der Zeit $t_{0.63}$ (Bild 4.8) und der Summenzeitkonstante T_z ein Zusammenhang der Form

$$t_{0.63} = T_z (1,045 \pm 0,045) \quad (4.18)$$

besteht. Mit

$$T_z = \sum_{j=1}^n T_j = T' \sum_{j=1}^n \frac{1}{j}$$

folgt für die Bestimmung der Zeitkonstante T'

$$T' = \frac{1}{1,045 \sum_{j=1}^n \frac{1}{j}} t_{0.63} = D(n) t_{0.63} \quad (4.19)$$

Für die Ermittlung der Ordnungszahl n erwies sich der Funktionswert $h(t_{0.63}/2)$ der Übergangsfunktion als sehr geeignet (Bild 4.8). In Tafel 4.4 sind die benötigten Kenngrößen für dieses Modell dargestellt.

In Analogie zu dem von Radtke für das Modell mit gestaffelten Zeitkonstanten entwickelten Verfahren wurde von Sponer [4.3] ein Auswertverfahren zur Bestimmung der Parameter eines Modells mit n gleichen Zeitkonstanten Gl. (4.12) entwickelt. Es umgeht die häufig schwierige Bestimmung des Wendepunktes der Übergangsfunktion.

Ausgehend von Gl. (4.18) gilt für gleich große Zeitkonstanten

$$T' = \frac{1}{1,045n} t_{0.63} = D'(n) t_{0.63}.$$

Zur Bestimmung der Ordnung n wird ebenfalls der Funktionswert $h(t_{0.63}/2)$ verwendet, der in Tafel 4.4 als Funktion von n für diesen Modelltyp dargestellt ist.

Tafel 4.4
Kenngrößen
der Übergangsfunktion
für I_1-T_n -Systeme

n	Modell I $h(t_0, \epsilon_{3/2})$	Modell II $h(t_0, \epsilon_{3/2})$	$D'(n)$
1	0,392	0,392	0,957
2	0,291	0,299	0,638
3	0,224	0,242	0,522
4	0,176	0,202	0,459
5	0,139	0,173	0,419
6	0,112	0,149	0,391
7	0,089	0,129	0,368
8	0,072	0,116	0,352
9	0,058	0,103	0,338
10	0,047	0,093	0,326

Weitere sehr umfassend ausgearbeitete Methoden zur Approximation von Übergangsfunktionen sind die *Zeitprozentmethode* von Schwarze [4.4] und die *Momentenmethode* von Simoju [4.5]. Ihre praktische Anwendung ist z. Z. nicht mehr sehr groß.

4.2.1.3. Modellbestimmung von $I_1-T_n-T_f$ -Systemen

Für $I_1-T_n-T_f$ -Systeme sind für die folgenden Modelle:

Modell I

$$G(p) = \frac{K_1}{p(1+pT')^n} \quad (4.20)$$

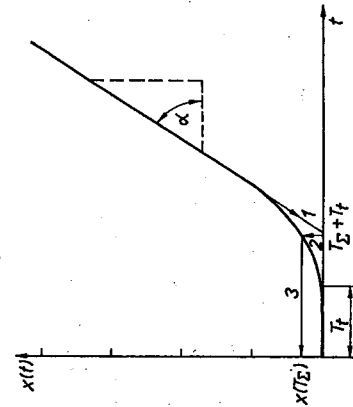
Modell II

$$G(p) = \frac{K_1}{p \prod_{j=1}^n \left(1 + p \frac{T_j'}{T_j}\right)} \quad (4.21)$$

von Schwarze [4.4] und Sponer [4.3] Approximationsregeln aus der Sprungantwort entwickelt worden. Die Totzeit T_f wird für beide Verfahren als erster Parameter wieder abgespalten (Bild 4.9).

Die Approximation des I_1-T_n -Systems durch ein Modell mit n gleichen Zeitkonstanten (Modell I) erfolgt in den Schritten

- Ermittlung von $K_1 = \tan \alpha / u_0$ und $T_2 = nT'$ (Bild 4.9),
- Bestimmung der Ordnung n aus $x(T_2)/(K_1 T_2 u_0)$ (Tafel 4.5),
- Bestimmung der Zeitkonstanten $T' = D_1(n) T_2$; $D_1(n)$ aus Tafel 4.5.

Bild 4.9
Bestimmung der Kenngrößen
der Sprungantwort für ein $I_1-T_n-T_f$ -System

Für die Approximation des Systems mit gestaffelten Zeitkonstanten (Modell II) werden die gleichen Schritte durchlaufen, wobei für die Faktoren $D_1(n)$ und $x(T_2)/(K_1 T_2 u_0)$ entsprechend Tafel 4.5 andere Werte einzusetzen sind.

Tafel 4.5
Kenngrößen
der Sprungantwort
für $I_1-T_n-T_f$ -Systeme

n	Modell I $\frac{x(T_2)}{T_2 K_1 u_0}$	Modell II $\frac{x(T_2)}{T_2 K_1 u_0}$	$D_1(n)$
1	0,368	0,368	1,000
2	0,281	0,271	0,500
3	0,241	0,224	0,333
4	0,218	0,195	0,250
5	0,202	0,174	0,200
6	0,190	0,161	0,167

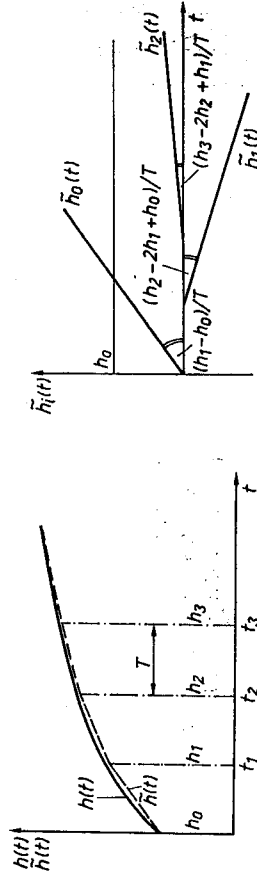
4.2.1.4. Bestimmung diskreter Werte des Frequenzgangs $G(j\omega)$ aus der Übergangsfunktion

Bei dieser von Unbehauen [4.7] vorgeschlagenen Methode wird die im Bild 4.10 dargestellte Übergangsfunktion $h(t)$ durch eine Summe von Geradenstücken $\tilde{h}_i(t)$ angenähert. Damit gilt für die approximierende Übergangsfunktion $\tilde{h}(t)$:

$$\tilde{h}(t) = h_0 + \sum_{i=0}^n \tilde{h}_i(t) \quad (4.22)$$

mit

$$\tilde{h}_i(t) = \begin{cases} 0 & \text{für } t < t_i \\ \frac{b_i}{T} (t - t_i) & \text{für } t \geq t_i \end{cases}$$

Bild 4.10. Prinzip der Zerlegung der Übergangsfunktion $h(t)$

Die Faktoren b_i , die die Differenz der Steigungen zwischen zwei Abschnitten ausdrücken, werden aus den Gleichungen

$$\frac{b_i}{T} = \begin{cases} (h_1 - h_0)/T & \text{für } i = 0 \\ (h_{i+1} - 2h_i + h_{i-1})/T & \text{für } i > 0 \end{cases} \quad (4.23)$$

bestimmt. Für die Übertragungsfunktion $G(p)$ der angenäherten Übergangsfunktion $\tilde{h}(t)$ gilt dann

$$\tilde{G}(p) = h_0 + \frac{1}{pT} \sum_{i=0}^n b_i e^{-p t_i} \quad (4.24)$$

Soll der Frequenzgang $\tilde{G}(j\omega_l)$ für die Frequenz ω_l ermittelt werden, wird $p = j\omega_l$ gesetzt, und man erhält

$$\begin{aligned}\tilde{G}(j\omega_l) &= h_0 - \frac{1}{T\omega_l} \sum_{i=0}^n b_i \sin(iT\omega_l) \\ &- \frac{1}{T\omega_l} \sum_{i=0}^n b_i \cos(iT\omega_l); \quad l = 1, 2, \dots, k.\end{aligned}\quad (4.25)$$

Die Genauigkeit der Approximation steigt mit der Anzahl der Stützstellen. Zufriedenstellende Ergebnisse liefert das Verfahren schon für den Bereich $20 < n < 30$.

Bezüglich der Modellstruktur kann es auch auf Systeme mit konjugiert komplexen Polen (Schwingungsglieder) und mit integriertem Verhalten angewendet werden.

4.2.2. Modellbestimmung anhand von Kenngrößen der Impulsantworten bzw. der Gewichtsfunktionen

Ist die Auftrübung von Sprungfunktionen auf das System nicht zweckmäßig (z. B. bei I-Gliedern, Gefährdung der Lebensfunktion des Systems), wird häufig ein Impuls als Testsignal verwendet. Das Problem dieses Testsignals liegt in der Realisierungsmöglichkeit des „idealen Impulses“, d. h. Amplitudenwert $\Rightarrow \infty$, Impulsbreite $\Rightarrow 0$. Deshalb werden in den folgenden Abschnitten Verfahren angegeben, die sowohl von den idealen als auch von den realen Realisierungsmöglichkeiten des Impulses ausgehen.

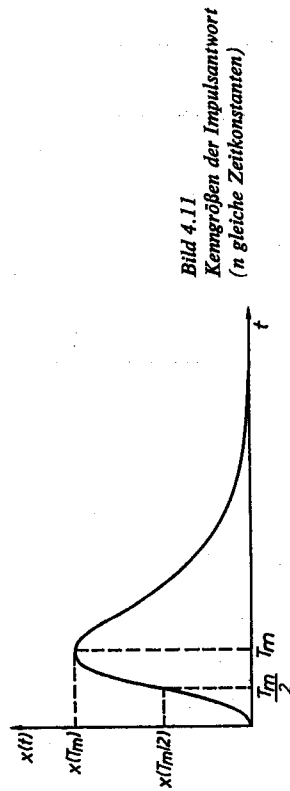


Bild 4.11
Kenngrößen der Impulsantwort
(n gleiche Zeitkonstanten)

4.2.2.1. Auswertung von Impulsantworten bei idealem Eingangssignal (Dirac-Impuls)

Für den Fall, daß das Eingangssignal als Dirac-Impuls angesehen werden kann, wurde von Werner [4.8] für den Modellansatz

$$G(p) = \frac{K}{(1 + pT)^n} \quad (4.26)$$

ein einfaches Verfahren zur Bestimmung der Parameter n und T entwickelt. Als Grenzwert für die Impulsbreite ΔT gilt

$$\Delta T \leq 0,1T'.$$

Aus den Kennwerten der Impulsantwort T_m , $T_m/2$, $x(T_m/2)$ und $x(T_m)$ (Bild 4.11) werden entsprechend Tafel 4.6 die Parameter des Modells in Gl. (4.26) ermittelt. Die einzelnen Schritte sind folgende:

- Aus dem Verhältnis $x(T_m)/(x(T_m/2))$ wird die Ordnung n bestimmt.
- Bei bekannter Ordnung n folgt aus T_m/T' die Zeitkonstante T' .
- Aus $T'x(T_m)/u_0K$ folgt bei bekannter Fläche des Impulses $u_0 = A = 1$ der Verstärkungsfaktor K .

Bei Impulsen mit einer Impulsbreite von $\Delta T > 0,1T'$ werden die Kennwerte vom Verhältnis $\Delta T/T'$ abhängig. Die Auswertung wird dadurch sehr erschwert.

Tafel 4.6
Kennwerte der Impulsantwort
für P-T_n-Systeme
mit n gleichen Zeitkonstanten

$\frac{x(T_m)}{x(T_m/2)}$	n	$\frac{T_m}{T'}$	$\frac{T'x(T_m)}{u_0K}$
1,213	2	1	0,368
1,471	3	2	0,271
1,785	4	3	0,224
2,165	5	4	0,196
2,623	6	5	0,175
3,185	7	6	0,159

4.2.2.2. Auswertung von Impulsantworten unter Beachtung der Impulsbreite des Eingangssignals

Ist die Impulsbreite $\Delta T > 0,1T'$, so ist es notwendig, sie bei der Approximation zu berücksichtigen. Die Methode basiert auf der Approximation durch ein Modell mit n harmonisch gestaffelten Zeitkonstanten; s. Gl. (4.17) [4.2]. Als Kennwerte eignen sich entsprechend Bild 4.12 die Zeiten ΔT , T_m und $T_{00} = t(x_0)$ sowie die Funktionswerte $x_0 = x(\Delta T)$ und $x_m = x(T_m)$. Unter Annahme der Modellstruktur ergeben sich die in Bild 4.13 dargestellten Verläufe für Quotienten aus den Kennwerten. Die Bestimmung der Parameter erfolgt in den Schritten:

- Ermittlung der Kennwerte x_0/x_m und T/T_{00} entsprechend Bild 4.12.
- Ermittlung der Ordnung n aus den Kennwerten x_0/x_m und T/T_{00} im Bild 4.13 a.
- Aus Bild 4.13 b wird mit Hilfe der Größen x_0/x_m und n der Wert für $T'/\Delta T$ entnommen und somit die Zeitkonstante T' ermittelt.
- Der Verstärkungsfaktor K wird bestimmt aus der Beziehung

$$K = \frac{1}{u_0} \frac{x_0}{\tilde{x}_0} \quad \text{mit} \quad \tilde{x}_0 = (1 - e^{-T'/T})^n. \quad (4.27)$$

Eine Überprüfung der Approximation mit verschiedenen ΔT ist in der Praxis sinnvoll.

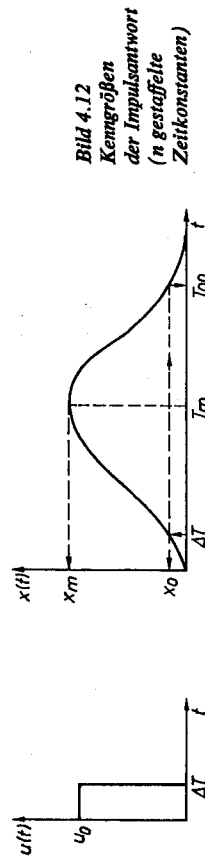


Bild 4.12
Kenngrößen
der Impulsantwort
(n gestaffelte
Zeitkonstanten)

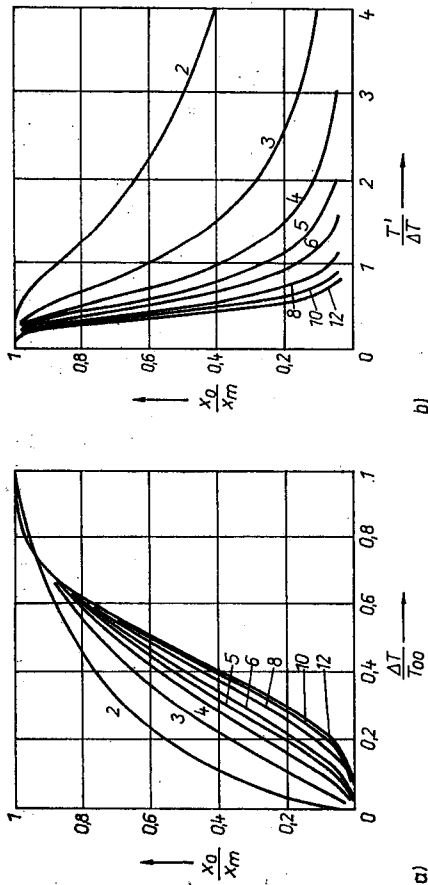


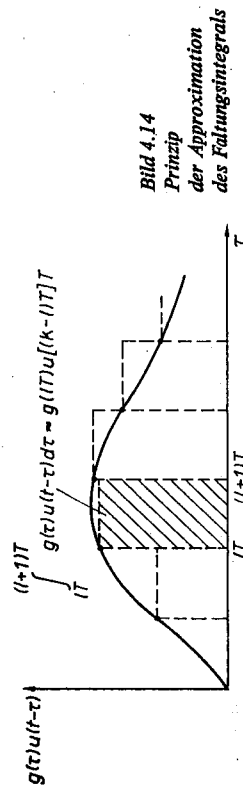
Bild 4.1.3. Verlauf von Quotienten der Kenngrößen aus Bild 4.1.2 zur Approximation von Impulsantworten

a) Diagramm zur Ermittlung der Ordnungszahl n
 b) Diagramm zur Bestimmung der Zeitkonstanten T'

4.2.3. Ermittlung von Stützstellen der Gewichtsfunktion – aus beliebigen Eingangs- und Ausgangssignalen – Entfaltung ohne Ausgleich

Die bisher vorgestellten Verfahren sind auf geometrisch einfache Testsignale, die aktiv auf das System aufgeprägt werden müssen, beschränkt. Hat das aktiv auf das System aufgeprägte oder im normalen Betrieb anliegende Testsignal eine beliebige Form, versagen die bis jetzt vorgestellten Modellbildungsverfahren. Als ein Ausweg bietet sich die Verwendung des Faltungintegrals an, das das Ausgangssignal $x(t)$ über eine Integralbeziehung aus dem Produkt der Gewichtsfunktion $g(t)$ und dem zeitverschobenen Eingangssignal ermittelt. Damit gilt mit $u(t) = 0$ für $t < 0$

$$x(t) = \int_0^t g(\tau) u(t - \tau) d\tau. \quad (4.28)$$



Das Problem besteht in der Auflösung der Integralgleichung nach $g(\tau)$ bzw. nach ihren Stützstellenwerten $g(\tau_i)$ bei bekannten Signalen $x(t)$ und $u(t)$. Im einfachsten Fall kann das Integral mit Hilfe der Rechteckregel aufgelöst werden. Entsprechend Bild 4.14 gilt für Systeme mit einem Gedächtnis von $(r-1)T$, d.h. $g(iT) = 0$ für $i = r$, für das Ausgangssignal

$$x(t) = \sum_{i=0}^{r-1} g(\tau_i) u(t - \tau_i). \quad (4.29)$$

Verwendet man nur die diskreten Werte der Eingangs- und Ausgangssignale $u(k-T) \triangleq u(k)$ und $x(kT) \triangleq x(k)$, so erhält man als Näherung für das Ausgangssignal zum Zeitpunkt $t = kT$ die Faltungssumme

$$x(k) \approx T \sum_{l=0}^{r-1} g(l) u(k-l) \quad (4.30)$$

mit

$$t = kT, \quad \tau = lT, \quad l = 0, 1, \dots, (r-1).$$

In Matrixschreibweise erhält man für j Beobachtungen und die Anfangswerte $x(t) = 0$, $u(t) = 0$ für $t \leq 0$

$$\begin{bmatrix} x(1) \\ x(2) \\ \vdots \\ x(j) \end{bmatrix} = T \begin{bmatrix} u(1) & 0 & \dots & 0 \\ u(2) & u(1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ u(j) & u(j-1) & \dots & u(1) \end{bmatrix} \begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ g(r-1) \end{bmatrix} \quad \begin{matrix} U \\ g \end{matrix}$$

Damit ergibt sich als Systembeschreibung

$$x = TUg. \quad (4.31)$$

Die Auflösung von Gl. (4.31) nach dem Vektor der Stützstellen g wird als *Entfaltung* bezeichnet und ergibt, wenn die Matrix U invertierbar ist,

$$g = \frac{1}{T} U^{-1} x. \quad (4.32)$$

Wird ein bestimmtes Testsignal $u(t)$ immer verwendet, kann die Matrix U^{-1} einmal berechnet und für weitere Untersuchungen abgespeichert werden.

Der Fehler der Approximation hängt wesentlich von der Abtastweite T ab. Sie soll so gewählt werden, daß die Gewichtsfunktion $g(t)$ durch 20 bis 40 Stützstellen beschrieben wird. Nach der Ermittlung der Stützstellen kann dann eine Approximation durch ein parametrisches Modell mit den im Abschnitt 4.2.2 vorgestellten Methoden erfolgen.

Bei dieser Methode ist zu beachten, daß das Testsignal $u(t)$ ein für die Ermittlung der Stützstellenwerte der Gewichtsfunktion $g(iT)$ günstiges Amplitudenspektrum („ideales weißes Rauschen“) besitzt und daß keine Störung $z(t)$ auftreten darf. Ist eine nicht zu vernachlässigende Störung vorhanden, muß eine Schätzung mit der Methode der Entfaltung mit Fehlerausgleich (s. Abschn. 6.2) durchgeführt werden.

4.3. Ein-/Ausgangs-Modelle im Frequenzbereich

Die Auswertung von Frequenzgangmessungen liefert, da alle Frequenzbereiche des Systems im Gegensatz zu den im Abschnitt 4.2 verwendeten Testsignalen mit der gleichen Energie angeregt werden, Modelle mit hoher Genauigkeit. Außerdem sind die Verfahren, die durch ein Korrelationsprinzip den Frequenzgang ermitteln, relativ unempfindlich gegen Störungen. Der Nachteil der Verfahren liegt in den langen Meßzeiten im tief-frequenten Bereich des Systems. Einen Ausweg bilden kombinierte Methoden, z.B. die

Auswertung von Übergangsfunktion (für die den tieffrequenten Bereich beschreibenden Parameter) und Frequenzgang (für die den hochfrequenten Bereich beschreibenden Parameter) [1.27, 1.46].

Außerdem ist zu beachten, daß die Frequenzen ω , des periodischen Testsignals wieder über den „interessierenden Frequenzbereich“ des Systems verteilt sein müssen, um die Struktur und die Parameter eines analytischen Frequenzgangsmodells mit ausreichender Genauigkeit zu ermitteln. Gleichzeitig muß bei allen Meßverfahren bedacht werden, daß der Frequenzgang $G(j\omega)$ der Quotient der komplexen Amplitude von Ausgangs- zu sinusförmigem Eingangssignal im stationären Zustand ist. Das bedeutet, daß bei jeder ausgewählten Frequenz ω , der Einschwingvorgang zur Auswertung abgewartet werden muß, was besonders im tiefen Frequenzbereich zu erheblichen Meßzeiten, Kosten oder Belastungen führen kann (Bild 4.15). Der Ausweg aus diesem Problem wurde oben aufgezeigt. Für die Wahl der Frequenzen des Testsignals hat sich ein konstanter Quotient benachbarter Frequenzen ω_{i+1}/ω_i bewährt, weil dann der interessierende Frequenzbereich im Bode-Diagramm durch Meßwerte mit stets gleichem Abstand beschrieben werden kann.

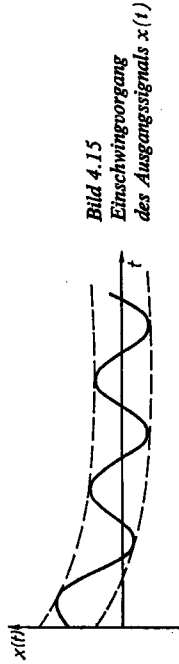


Bild 4.15
Einschwingvorgang
des Ausgangssignals $x(t)$

In den folgenden Abschnitten soll auf einige typische Methoden zur Ermittlung des Frequenzgangs (Abschn. 4.3.1), auf die Approximation des Frequenzgangs durch ein parametrisches Modell in Form eines Zeitkonstantenmodells (Abschn. 4.3.2) und auf die Approximation der Ortskurve durch ein Polynommodell (Abschn. 4.3.3) näher eingegangen werden. Auf eine große Anzahl weiterer Verfahren kann an dieser Stelle nur verwiesen werden [1.37, 1.46].

4.3.1. Methoden zur Aufnahme des Frequenzgangs $G(j\omega)$

Die entwickelten Methoden haben das Ziel, die Komponenten des Frequenzgangs $G(j\omega)$ in einer der Formen

$$G(j\omega) = U(\omega) + jV(\omega) = |G(j\omega)| e^{j\varphi(\omega)} \quad (4.33)$$

zu ermitteln. Werden Phasenminimalsysteme vorausgesetzt, kann die Messung des Frequenzgangs auf die Bestimmung von $|G(j\omega)|$ beschränkt werden, da aus dem Amplitudenfrequenzgang $|G(j\omega)|$ der Phasenfrequenzgang konstruiert werden kann.

4.3.1.1. Synchronaufzeichnungsverfahren

Aus den synchron entsprechend Bild 4.16 registrierten Eingangs- und Ausgangssignalen

$$u(t) = A \sin(\omega t + \varphi_u) \quad \text{und} \quad x(t) = B \sin(\omega t + \varphi_x)$$

kann der Betrag des Frequenzgangs

$$|G(j\omega)| = B/A \quad (4.34)$$

und der Phasenfrequenzgang

$$\varphi(\omega) = \varphi_x - \varphi_u = \omega(t_x - t_u) = \omega t\varphi$$

ermittelt werden. Als günstiger Frequenzbereich für diese Synchronaufzeichnung hat sich $10^{-5} \text{ s}^{-1} < f < 10^{-1} \text{ s}^{-1}$ erwiesen. Problematisch ist die Bestimmung kleiner Phasenverschiebungen.

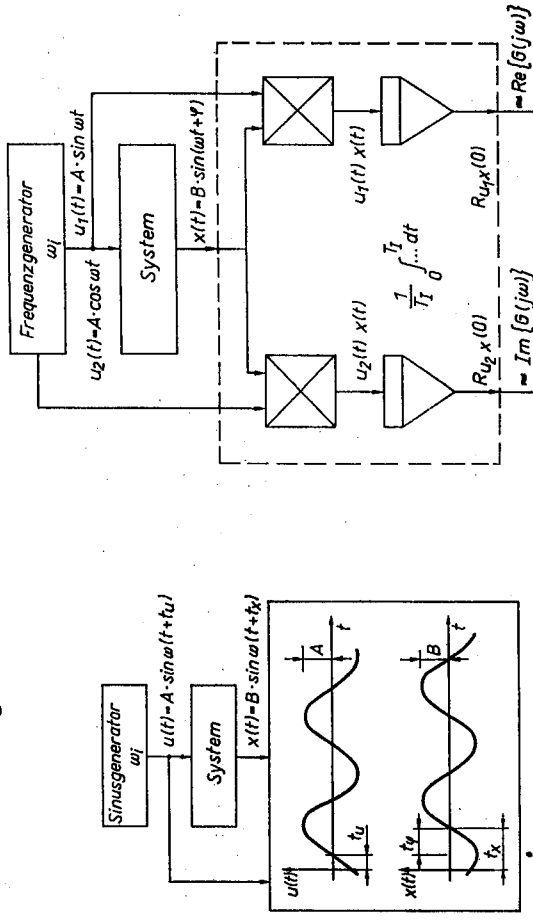


Bild 4.16. Schema des Synchronaufzeichnungsverfahrens

4.3.1.2. Komponentenverfahren

Das Komponentenverfahren hat aufgrund der Möglichkeit der sehr feinen Unterteilung eines großen Frequenzbereichs $10^4 \text{ s}^{-1} < f < 10^2 \text{ s}^{-1}$ und der guten Störunterdrückung durch die Anwendung des Korrelationsprinzips die breiteste Anwendung gefunden. Das Schema dieses Verfahrens ist im Bild 4.17 dargestellt. Der Realteil des Frequenzgangs wird dabei in dem einen Kanal durch die Multiplikation des Systemausgangssignals mit dem Eingangssignal, der Imaginärteil in dem anderen Kanal durch die Multiplikation des Ausgangssignals mit dem um 90° verschobenen Testsignal ermittelt.

Der Realteil ergibt sich aus (s. Abschn. 3.2.2.2)

$$\begin{aligned} R_{u1x}(0) &= \frac{1}{T_1} \int_0^{T_1} AB \sin \omega t \sin(\omega t + \varphi) dt \\ &= \frac{AB}{2T_1} \int_0^{T_1} [\cos \varphi - \cos(2\omega t + \varphi)] dt \end{aligned} \quad (4.35)$$

und für genügend große T_1

$$R_{u1x}(0) = \frac{AB}{2} \cos \varphi = \frac{A^2}{2} |G(j\omega)| \cos \varphi$$

$$R_{u2x}(0) = \frac{A^2}{2} \operatorname{Re}\{G(j\omega)\}. \quad (4.36)$$

Bild 4.17. Schema des Komponentenverfahrens

Für den Imaginärteil erhält man analog

$$\begin{aligned} R_{u_2x}(0) &= \frac{1}{T_1} \int_0^{T_1} AB \cos \omega t \sin (\omega t + \varphi) dt \\ &= \frac{AB}{2T_1} \int_0^{T_1} \sin \omega + \sin (2\omega t + \varphi) dt. \end{aligned} \quad (4.37)$$

Für große T_1 wird

$$R_{u_2x}(0) = \frac{A^2}{2} |G(j\omega)| \sin \varphi = \frac{A^2}{2} \operatorname{Im} \{G(j\omega)\}. \quad (4.38)$$

Ein Störsignal $z(t)$ wird durch das Korrelationsprinzip beseitigt, wenn in der Störung keine periodischen Komponenten der Testsignalfrequenz enthalten sind. Für hinreichend große Mittelungszeiten T_1 gehen die Anteile der Korrelationsfunktion, die zusätzlich in den Gln. (4.36) und (4.37) auftreten,

$$\frac{1}{T_1} \int_0^{T_1} z(t) A \sin \omega t dt \quad \text{bzw.} \quad \frac{1}{T_1} \int_0^{T_1} z(t) A \cos \omega t dt,$$

gegen Null. Bei periodischen Störsignalen wird der Fehler des Frequenzgangs um so größer, je näher Störsignalfrequenz und Testsignalfrequenz zusammenliegen. Weiterführende Betrachtungen zu diesem Problem und zum Einfluß einer endlichen Mittelungszeit auf die Genauigkeit der Frequenzgangbestimmung sind in [1.37, 1.46, 4.9] enthalten.

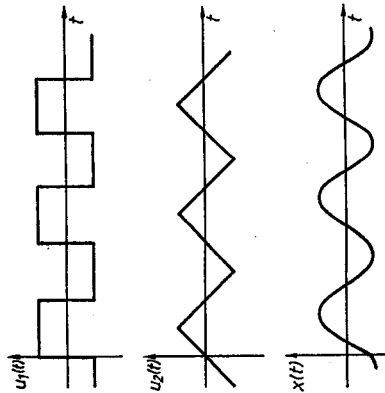


Bild 4.18
Verlauf von rechteck- und dreieckförmigen Testsignalen $u_1(t)$ und $u_2(t)$ und Verlauf der Grundwelle des Ausgangssignals $x(t)$

4.3.1.3. Frequenzgangsmessung mit Rechteck- und Dreieckswellen

Die Ermittlung des Frequenzgangs aus der Analyse des Ausgangssignals bei aufgeprägten Rechteck- oder Dreieckswellen hat die Vorteile der einfachen Realisierbarkeit und teilweise der Erhöhung des Störabstands. So besitzt die Grundwelle des Rechtecksignals einer Amplitude u_0 die Amplitude $u_0 \cdot 4/\pi$ (s. Abschn. 3.1.2), d.h., bei gleicher Aussteuerung wird der Störabstand gegenüber dem sinusförmigen Testsignal um den Faktor 1,27 verbessert. Bei der Verwendung eines Dreieckssignals einer Amplitude u_0 besitzt die Grundwelle die Amplitude $u_0 \cdot 8/\pi^2$, d.h., der Störabstand wird um den Faktor 0,81 verschlechtert.

Für die Rechteckwelle der Amplitude u_0 (Bild 4.18) gilt für die Fourier-Zerlegung entsprechend Abschnitt 3.1.2, Tafel 3.1, die Beziehung

$$u(t) = \frac{4}{\pi} u_0 \left\{ \sin \omega t + \frac{1}{3} \sin 3\omega t + \frac{1}{5} \sin 5\omega t + \dots \right\}. \quad (4.39)$$

Das Ausgangssignal lautet entsprechend

$$x(t) = \frac{4}{\pi} u_0 \left\{ |G(j\omega)| \sin (\omega t + \varphi(\omega)) + \frac{1}{3} |G(j3\omega)| \sin (3\omega t + \varphi(3\omega)) + \dots \right\}. \quad (4.40)$$

Die Ermittlung der Amplitude der Grundwelle im Ausgangssignal erfolgt zweckmäßigerweise durch eine Fourier-Analyse mit Hilfe der im Abschnitt 3.2.3.4 vorgestellten Algorithmen durch einen Digitalrechner. Vereinfachte Verfahren für P-T_s-Systeme sind in [1.40, 4.8] enthalten.

Werden Testsignale in Form von Dreieckswellen verwendet, so lautet die Fourier-Zerlegung (s. Tafel 3.1)

$$u(t) = \frac{8u_0}{\pi^2} \left\{ \sin \omega t - \frac{1}{9} \sin 3\omega t + \frac{1}{25} \sin 5\omega t + \dots \right\}. \quad (4.41)$$

Für das Ausgangssignal gilt damit

$$x(t) = \frac{8u_0}{\pi^2} \left\{ |G(j\omega)| \sin (\omega t + \varphi(\omega)) - \frac{1}{9} |G(j3\omega)| \sin (3\omega t + \varphi(3\omega)) + \dots \right\}. \quad (4.42)$$

Aus Gl. (4.42) ist zu erkennen, daß sich gegenüber der Rechteckwelle der Störabstand verschlechtert und daß die dritte Oberwelle um den Faktor 3 stärker gedämpft wird, was die Ermittlung von $G(j\omega)$ erleichtert.

Neben diesen beiden Testsignalarten wurden sog. *Mehrfrequenzsignale* entworfen, die in einem Meßvorgang die Frequenzgangbestimmung für mehrere Frequenzen gestatten. So enthält z. B. ein in [1.40, 4.10] vorgestelltes binäres Signal neben der Grundwelle ω die Oberwellen mit den Frequenzen $2\omega, 4\omega, 8\omega, 16\omega$ und 32ω . Ihre Anwendung ist jedoch auch beschränkt, weil die Amplituden der Einzelschwingungen des Mehrfrequenzsignals wesentlich geringer sind als bei den bereits genannten Testsignalen und sich damit der Störabstand wesentlich verschlechtert.

Durch die Anwendung der in diesem Abschnitt vorgestellten Methoden erhält man eine diskrete Folge von Frequenzgangswerten. Sie können in der komplexen $G(j\omega)$ -Ebene (Ortskurve) oder als Amplituden- und Phasenfrequenzgang (Bode-Diagramm) [1.2, 3.1] dargestellt werden. Für viele Anwendungen ist es zweckmäßig, das in Form einer Wertefolge bzw. grafisch als Ortskurve oder Frequenzganglinien vorliegende nichtparametrische Modell durch ein parametrisches Modell zu approximieren. Bei der Ermittlung des durch wenige Parameter zu beschreibenden parametrischen Modells muß die Struktur- und Parameterbestimmung gelöst werden. Aus diesem Grunde werden in den folgenden Abschnitten Verfahren zur Approximation des Frequenzgangs durch parametrische Modelle in Form des Zeitkonstantenmodells und des Polynommodells vorgestellt.

4.3.2. Approximation von $G(j\omega)$ durch ein Zeitkonstantenmodell

Aus dem gemessenen Frequenzgang $G(j\omega)$, der oft als Amplituden- und Phasenfrequenzgang aufgetragen wird, sollen die Struktur und die Parameter eines Modells der Form

$$G(p) = K \frac{\prod_{i=1}^n (1 + pT_{Di})}{\prod_{j=1}^m (1 + pT_{Dj})} \quad (4.43)$$

bestimmt werden. Dabei wird vom Gesamtfrequenzgang $G(j\omega)$, vom Amplitudenfrequenzgang $\lg |G(j\omega)|$ oder vom Phasenfrequenzgang $\varphi(j\omega)$ ausgegangen [4.11, 4.12]. Bei Phasenminimalsystemen sind die Verfahren, die von $|G(j\omega)|$ ausgehen, wegen ihrer Einfachheit und Übersichtlichkeit vorteilhaft. Als Phasenminimalsystem werden solche Systeme bezeichnet, die den zu dem vorhandenen Amplitudenfrequenzgang möglichen minimalen negativen Phasenfrequenzgang für alle Frequenzen haben [1.2]. Sie besitzen keine Nullstellen in der rechten p-Halbebene, und für sie existiert eine eindeutige Zuordnung zwischen Amplituden- und Phasenfrequenzgang. Damit fallen für diese Methode alle Systeme mit Allpaß- und Laufzeitverhalten aus der weiteren Betrachtung heraus.

Deswegen kann die Ermittlung eines parametrischen Modells allein auf der Grundlage des Amplitudenfrequenzgangs erfolgen. Das Verfahren beruht auf einer Approximation des in dB über $\lg \omega$ aufgetragenen Amplitudenfrequenzgangs durch Geraden verschiedener Neigungen. Es wird davon ausgegangen, daß der Amplitudenfrequenzgang sich aus einer Summe von Elementarkennlinien mit P-, I-, D-, T_1 - T_{D1} - und T_2^* -Verhalten nach der Vorschrift

$$\lg |G(j\omega)| = \sum_{i=1}^n \lg |G_i(j\omega)| \quad (4.44)$$

ergibt. Die Kennlinien der Elementarsysteme sind in Tafel 4.7 als Geraden dargestellt. Für die Approximation gilt analog zu Gl. (4.44)

$$\lg |\tilde{G}(j\omega)| = \sum_{j=1}^n \lg |\tilde{G}_j(j\omega)|. \quad (4.45)$$

Der Amplitudenfrequenzgang $|G(j\omega)|$ sollte in folgenden Schritten approximiert werden:

1. Darstellung von $|G(j\omega)|$ als Funktion von $\lg \omega/\omega^*$; ω^* wählbare Bezugsfrequenz.
2. Approximation von $|G(j\omega)|$ durch Geraden mit den Neigungen

$$k^* = \frac{d \lg |G(j\omega)|}{d \lg \omega/\omega^*}; \quad k^* = \pm 0, 1, 2, \dots$$

($k^* = -1$ entspricht der Neigung 20 dB/Dekade).

Ändert sich an einer Knickstelle ω , die Neigung um $+$ ($-$) 1, so ist eine reelle Nullstelle (reeller Pol) mit $T_1 = 1/\omega$ anzusetzen. Bei Änderung der Neigung um $+$ ($-$) 2 ist ein konjugiert-komplexes Nullstellen-(Pol-)Paar vorzusehen. Hierbei ist ω_0 mit der Knickfrequenz identisch. An einfachen, weiter voneinander entfernten Knickpunkten ω_i soll die Abweichung des asymptotischen Verlaufs (Geradenapproximation) vom exakten Verlauf, d. h. $\lg |G(j\omega)| - \lg |\tilde{G}(j\omega)|$, $+$ ($-$) 3 dB betragen (Bild 4.19).

Tafel 4.7

Geradenapproximation des Amplitudenfrequenzgangs einfacher Übertragungsfunktionen

Systemverhalten	Approximation
P-Verhalten $G(j\omega) = K$ $k^* = 0$	
I-Verhalten $G(j\omega) = 1/j\omega T_I$ $k^* = -1$	
D-Verhalten $G(j\omega) = j\omega T_D$ $k^* = +1$	
T_1 -Verhalten $G(j\omega) = 1/(1 + j\omega T_1)$ $k^* = -1$	
T_{D1} -Verhalten $G(j\omega) = (1 + j\omega T_{D1})$ $k^* = +1$	

3. Konstruktion des Verlaufs der verfeinerten Approximationskurve unter Anwendung der Korrekturtafeln aus den Bildern 4.20 und 4.21 zur Berücksichtigung der von beobachtbaren Knickstellen herrührenden Abweichungen nach der Vorschrift

$$\lg |\tilde{G}(j\omega)| = \lg |\tilde{G}_j(j\omega)| + \lg |\Delta G(j\omega)| \quad (4.46)$$

mit $\lg |\Delta G(j\omega)|$ als Korrekturwerte entsprechend den Bildern 4.20 und 4.21.

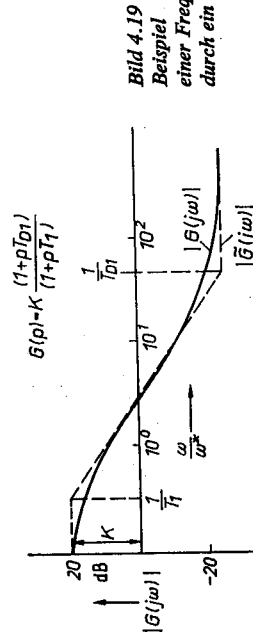


Bild 4.19
Beispiel einer Frequenzgangsapproximation durch ein Zeitkonstantenmodell

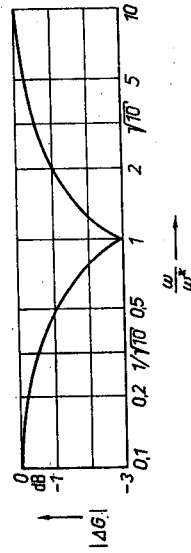


Bild 4.20
Korrekturtafel für die Geradenapproximation von Amplitudenfrequenzgängen für Trägheits- und Vorhaltglieder

4. Vergleich der Amplitudenfrequenzgänge $\lg |G(j\omega)| - \lg |\hat{G}(j\omega)|$ und ggf. eine Korrektur der Knickfrequenzen und Rücksprung zu Punkt 2.
5. Falls die Genauigkeit der Approximation immer noch nicht ausreichend ist, muß mit dem Verfahren von *Linville* in [4.12] eine weitere Verbesserung versucht werden.

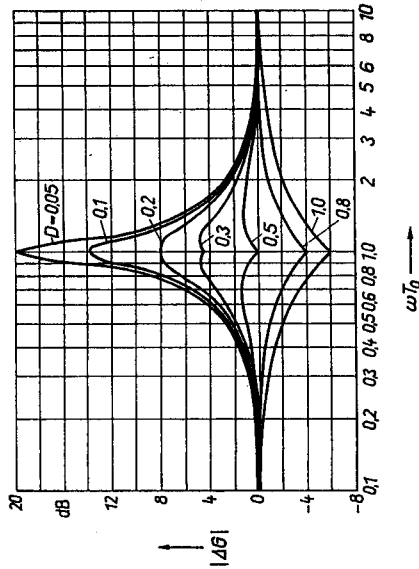


Bild 4.21
Korrekturtafel
für die Geradenapproximation
von Amplitudenfrequenzgängen
für Schwingungsglieder

Das vorgestellte Verfahren zeichnet sich durch seine einfache Handhabung und durch eine für viele Aufgaben ausreichende Güte bezüglich der Struktur- und Parameterermittlung aus. Sind jedoch höhere Genauigkeitsanforderungen zu erfüllen, sollte auf rechnergestützte Methoden zurückgegriffen werden [1.46]. Dabei muß festgestellt werden, daß die Güte dieses Verfahrens wesentlich durch die Erfahrung und das Geschick des jeweiligen Bearbeiters bestimmt wird. Diese Aussage trifft für das im folgenden Abschnitt vorgestellte Verfahren der analytischen Approximation der diskreten Frequenzgangswerte in der $G(j\omega)$ -Ebene, d.h. der Ortskurve, nicht zu.

4.3.3. Approximation von $G(j\omega)$ durch ein Polynommodell

Die Bestimmung der analytischen Beschreibung $G(j\omega)$ eines Systems mit konzentrierten Parametern der Form

$$G(j\omega) = \frac{K}{(j\omega)^m \left[1 + \sum_{i=1}^n a_i (j\omega)^i \right]} = \frac{K}{(j\omega)^m [U(\omega) + jV(\omega)]} \quad (4.47)$$

erfordert

1. die Bestimmung des Grades n und m (d.h. der Struktur),
2. die Bestimmung der Parameter a_i .

Wird das System durch Gl. (4.47) hinreichend genau beschrieben, so können die Strukturparameter n und m aus folgenden Überlegungen gewonnen werden. Trennt man die Beschreibung des Frequenzgangs von Gl. (4.47) in die Teilfrequenzgänge

$$G_1(j\omega) = K/(j\omega)^m; \quad m = 1, 2, \dots, \quad (4.48)$$

$$G_2(j\omega) = 1 / \left(1 + \sum_{i=1}^n a_i (j\omega)^i \right); \quad i = 0, 1, 2, \dots,$$

so ist zu erkennen, daß bei tiefen Frequenzen das Gesamtverhalten durch $G_1(j\omega)$ bestimmt wird:

$$G(j\omega) \approx G_1(j\omega) = K/(j\omega)^m.$$

Im Bild 4.22 ist der Verlauf der Ortskurve für $m = 0$ (P-Glied) bis $m = 3$ (I₃-Glied) dargestellt. Der zweite Teilfrequenzgang $G_2(j\omega)$ hat den Verlauf von T_n -Gliedern. Die Ortskurve durchläuft, von der reellen Achse beginnend, n Quadranten (Bild 4.23). Insgesamt gilt damit für die Strukturparameter

m Anzahl der Quadranten der $G(j\omega)$ -Ebene, die im Uhrzeigersinn zwischen der positiven reellen Achse und dem Quadranten liegt, in dem die Ortskurve mit $\omega = 0$ beginnt (s. Bild 4.22)

n Anzahl der Quadranten, die die Ortskurve berührt (s. Bild 4.23).

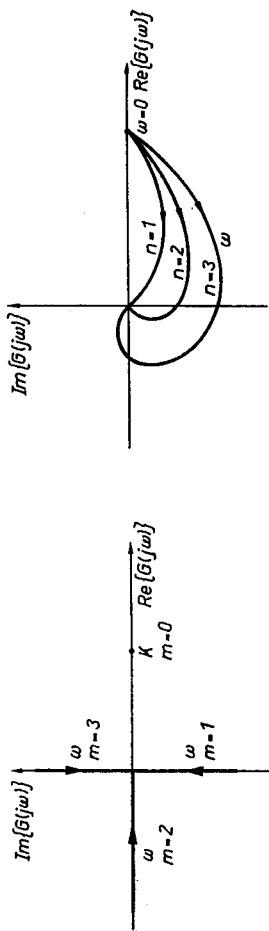


Bild 4.22. Ortskurven des Teilfrequenzgangs

$$G_1(j\omega) = K/(j\omega)^m$$

Bild 4.23. Ortskurven des Teilfrequenzgangs

$$G_2(j\omega) = 1 / \left(1 + \sum_{i=1}^n a_i (j\omega)^i \right)$$

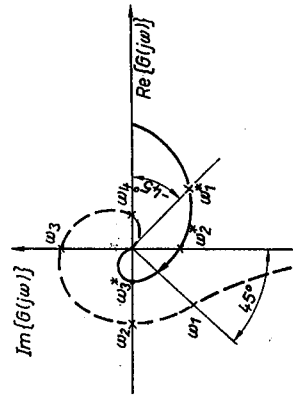


Bild 4.24
Wahl der Interpolationspunkte
bei dem Verfahren von Szveda
 ω_i^* P-Verhalten
 ω_i I-Verhalten

Das System wird dann durch das Modell gut approximiert, wenn keine wesentlichen Nullstellen im System vorhanden sind [3.1]. Nach der Festlegung der Strukturparameter m und n des Modells von Gl. (4.47) können die n Parameter a_i , $i = 0, 1, \dots, n$ ermittelt werden. Die Bestimmungsgleichungen werden dadurch gefunden, daß bei n diskreten Frequenzen ω_i die Amplituden und Phasen des gemessenen Frequenzgangs $G(j\omega)$ mit der Approximation $\hat{G}(j\omega)$ übereinstimmen müssen. Zu relativ einfachen Bestimmungen der Gleichungen gelangt man, wenn man nach *Szveda* [4.13] die Übereinstimmung der Phasen bei n Frequenzen ω_i fordert. Damit sind für die Übereinstimmung zwischen dem Frequenzgang des Modells $\hat{G}(j\omega)$ und dem des Systems $G(j\omega)$ in n Punkten die Bedingungen

$$\phi(\omega_i) - \phi(\omega_i) = 0; \quad i = 1, 2, \dots, n$$

zu erfüllen.

Wählt man als ersten Punkt der Ortskurve die Frequenz ω_1 , bei der erstmals $|U(\omega)| = |V(\omega)|$, und als weitere $(\omega_2 \dots \omega_n)$ die $(n-1)$ Schnittpunkte mit den Koordinatenachsen, wo entweder $U(\omega) = 0$ oder $V(\omega) = 0$ sind, so erhält man ein System von n Gleichungen (Bild 4.24).

Für $n = 7$ erhält man die Bestimmungsgleichungen zu

$$\begin{aligned}
 a_1 &= \frac{1}{\omega_1} \left(1 - \frac{\omega_1^2}{\omega_2^2} \right) \left(1 - \frac{\omega_1^2}{\omega_4^2} \right) \left(1 - \frac{\omega_1^2}{\omega_6^2} \right) \\
 a_2 &= \frac{1}{\omega_2} + \frac{1}{\omega_4} + \frac{1}{\omega_6} \\
 a_3 &= a_1 \left(\frac{1}{\omega_3^2} + \frac{1}{\omega_5^2} + \frac{1}{\omega_7^2} \right) \\
 a_4 &= \frac{1}{\omega_2^2 \omega_4^2} + \frac{1}{\omega_2^2 \omega_6^2} + \frac{1}{\omega_4^2 \omega_6^2} \\
 a_5 &= \frac{1}{\omega_3^2 \omega_5^2} + \frac{1}{\omega_3^2 \omega_7^2} + \frac{1}{\omega_5^2 \omega_7^2} \\
 a_6 &= \frac{1}{\omega_2^2 \omega_4^2 \omega_6^2} \\
 a_7 &= a_1 \frac{1}{\omega_3^2 \omega_5^2 \omega_7^2}.
 \end{aligned} \quad (4.50)$$

Zur Bestimmung der Parameter a_i werden nur die Frequenzen $\omega_1, \dots, \omega_n$ benötigt. Für $n < 7$ sind in den Bestimmungsgleichungen für a_i alle ω_i mit $i > n$ unendlich zu setzen.

Die Ermittlung des statischen Verstärkungsfaktors K erfolgt bei der angenommenen Systemstruktur von Gl. (4.47) entsprechend den Beziehungen

$$\begin{aligned}
 \text{a) } m &= 0 & K &= \lim_{\omega \rightarrow 0} G(j\omega), \\
 \text{b) } m &> 0 & K &= |G(j\omega_1)| \sqrt{2} \omega_1^m \left(1 - \frac{\omega_1^2}{\omega_2^2} \right) \left(1 - \frac{\omega_1^2}{\omega_4^2} \right) \left(1 - \frac{\omega_1^2}{\omega_6^2} \right).
 \end{aligned} \quad (4.51)$$

Ermittlung von

$$\begin{aligned}
 \omega_1 \text{ bei } \varphi(\omega_1) &= -m \frac{\pi}{2} - \frac{\pi}{4} \\
 \omega_i \text{ bei } \varphi(\omega_i) &= -m \frac{\pi}{2} - (i-1) \frac{\pi}{2}; \quad i = 2, 3, \dots, n.
 \end{aligned}$$

und

Die praktische Ermittlung der Frequenzen ω_i (z. B. mit dem Komponentenverfahren) erfolgt durch Erhöhung der Testsignalfrequenz von null bis zu dem Wert, bei dem erstmalig Real- und Imaginärteil gleich sind (ω_1), und dann weiter, bis im Wechsel Real-

und Imaginärteil verschwinden. Die Messung ist zu beenden, wenn kein Schnittpunkt mit einer der Koordinatenachsen mehr festgestellt werden kann.

Für die Approximation des Frequenzgangs durch rationale Übertragungsfunktionen beliebiger Struktur unter Verwendung der Methode der kleinsten mittleren Fehlerquadrate (Regression) sind u. a. von Strobel [1.46] sehr leistungsfähige Verfahren für den Digitalrechner und als Handverfahren entworfen worden. Die Verfahren ermitteln bei vorgegebener Approximationsgüte automatisch die minimale Parameteranzahl.

4.4. Ermittlung des Zustands/Beobachter

4.4.1. Problemstellung

Entsprechend den Darstellungen im Abschnitt 1.4 soll in den weiteren Betrachtungen von folgender Zustandsbeschreibung des Systems ausgegangen werden:

$$\begin{aligned}
 \dot{q}(t) &= Aq(t) + Bu(t) \\
 x(t) &= Cq(t)
 \end{aligned} \quad (4.52)$$

mit den Dimensionen der Vektoren und Matrizen

$$\begin{aligned}
 q(t): [n \times 1], \quad u(t): [l \times 1], \quad x(t): [m \times 1], \\
 A: [n \times n], \quad B: [n \times l], \quad C: [m \times n].
 \end{aligned}$$

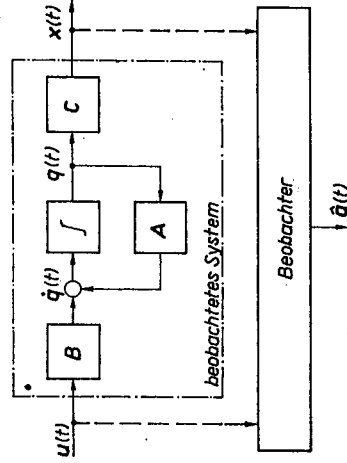


Bild 4.25
Allgemeines Prinzip
der Zustandsbestimmung
durch einen Beobachter

Die Aufgabe besteht nun darin, für das durch Gl. (4.52) gegebene System, von dem die Matrizen A , B und C bekannt sind, eine Vorschrift (Algorithmus) zu entwerfen, die die Rekonstruktion des Zustandsvektors $q(t)$ aus dem gemessenen Eingangssignal oder aus dem gemessenen Eingangs- und Ausgangssignal entsprechend Bild 4.25 gestattet. Für die Ermittlung des Zustands ist ein zweites dynamisches System, der Zustandsbeobachter/Beobachter, zu entwerfen. Er soll so gestaltet sein, daß nach einer endlichen Einschwingzeit der durch den Beobachter ermittelte Zustand $\hat{q}(t)$ mit dem nichtmeßbaren Zustand $q(t)$ möglichst identisch ist.

Die Bestimmung des Zustands gewinnt immer mehr an Bedeutung, weil u. a. folgende Aufgaben gelöst werden können:

1. Gewinnung von mehr und besserer Information über den Prozeß (Bedeutung für die Sensortechnik)

2. möglicher Verzicht auf eine direkte Messung von Prozeßgrößen in den Fällen, in denen keine Meßmöglichkeit besteht, das technische Meßprinzip zu aufwendig gegenüber der rechenstechnischen Lösung und das Beobachtungsprinzip zuverlässiger ist
3. Ergänzung unvollständiger Messungen
4. Ermittlung des Zustands für die Realisierung des Steuerungsgesetzes bei Konzepten der Zustandsrückführung und Optimalsteuerung.

In den folgenden Abschnitten werden für das in Gl. (4.52) dargestellte System die Konzepte der Beobachtung des Zustands auf der Grundlage der gemessenen Eingangssignale $u(t)$ mit einem Parallelmodell des Systems und auf der Grundlage der gemessenen Eingangs- und Ausgangsgrößen mit einem erweiterten Systemmodell (Luenberger-Beobachter) dargelegt. Dabei wird in allen Fällen von der kontinuierlichen Darstellung ausgegangen und dann zur diskontinuierlichen Betrachtung übergegangen. Die trivialen Fälle, daß alle Zustandsgrößen direkt meßbar sind oder aus der Ausgangsgröße direkt durch Bildung von Ableitungen gewonnen werden können, werden hier nicht betrachtet. Außerdem sollen an dieser Stelle nur Beobachter behandelt werden, bei denen die Ordnungen n von System und Beobachter gleich sind (vollständige Beobachter). Ist die Ordnung des Beobachters geringer als die des Systems, wird von reduzierten Beobachtern gesprochen [3.24, 4.14, 4.15].

4.4.2. Zustandsbeobachtung aus den Eingangsgrößen und einem Systemmodell

Die Beobachtung der Zustände erfolgt bei diesem Prinzip durch ein dem System parallelschaltetes Modell unter Verwendung der Eingangssignale $u_i(t)$ (Bild 4.26). Die Zustandsgrößen $\hat{q}_i(t)$ des Modells werden im stationären Fall die Werte $q_i(t)$ des Systems annehmen, wenn gilt

$$\hat{A} = A \quad \text{und} \quad \hat{B} = B.$$

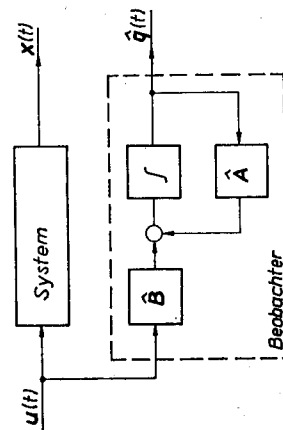


Bild 4.26
Zustandsbeobachtung
mittels Parallelmodell

Damit gilt als allgemeine Beobachtergleichung im kontinuierlichen Fall

$$\dot{\hat{q}}(t) = A\hat{q}(t) + Bu(t) \quad (4.53)$$

mit den Anfangsbedingungen $\hat{q}(t_0) = 0$.

Für diskrete Systeme bzw. diskrete Steuerung kontinuierlicher Systeme erhält man mit der Tasterperiode T und den Vereinbarungen $q(kT) \equiv q(k)$ und $u(kT) \equiv u(k)$ die Beobachtergleichung in der Form

$$\hat{q}(k+1) = A^* \hat{q}(k) + B^* u(k). \quad (4.54)$$

Dabei gelten für die Zusammenhänge zwischen den Matrizen der zeitdiskreten und der kontinuierlichen Zustandsraumdarstellung [1.52, 3.2, 3.24, 4.15]

$$\begin{aligned} A^* &= e^{AT} \\ B^* &= (e^{AT} - I) A^{-1} B \end{aligned} \quad (4.55)$$

$$D^* = D, \quad C^* = C.$$

Zur numerischen Lösung der Gl. (4.55) benutzt man häufig die folgende Reihenentwicklung:

$$A^* = I + AT + A^2 \frac{T^2}{2!} + A^3 \frac{T^3}{3!} + \dots = \sum_{a=0}^{\infty} A^a \frac{T^a}{a!}, \quad (4.56)$$

die in den Ausdruck

$$A^* = I + SA$$

mit

$$S = T \left(I + A \frac{T}{2!} + A^2 \frac{T^2}{3!} + \dots \right) A$$

umgeformt werden kann. Damit erhält man als endgültige Bestimmungsbedingungen

$$\begin{aligned} A^* &= I + SA \\ B^* &= SB. \end{aligned} \quad (4.57)$$

Gegenüber Gl. (4.55) wird die Inversion umgangen; aber durch die notwendige endliche Anzahl von Gliedern der Reihenentwicklung tritt ein Fehler auf. Eine Abschätzung über die erforderliche Zahl der Glieder der Reihenentwicklung m kann durch die Norm des Zuwachsterms

$$\left\| A^m \frac{T^m}{(m+1)!} \right\| \quad (4.58)$$

erfolgen. Die Ermittlung des Zustandsvektors $q(t)$ bzw. $q(k)$ mit dieser Methode hat den Vorteil, daß ein Differenzieren von Signalen vermieden wird. Als Nachteil muß aber angesehen werden, daß das Konvergenzverhalten des Fehlers zwischen System- und Beobachterzustand, d. h.

$$\tilde{q}(t) = q(t) - \hat{q}(t), \quad (4.59)$$

nur von der Dynamik des beobachteten Systems abhängt. In den meisten Systemen ist das Konvergenzverhalten damit durch Trägheitsverhalten gekennzeichnet. Dieser wesentliche Nachteil kann, wie im folgenden Abschnitt gezeigt wird, durch die Erweiterung des in Gl. (4.53) verwendeten Modells mit der Ausgangsgleichung

$$x(t) = Cq(t)$$

beseitigt werden.

4.4.3. Zustandsbeobachtung aus den Eingangs- und Ausgangsgrößen – Luenberger-Beobachter

Den Ausgangspunkt für den von Luenberger [4.16] vorgeschlagenen Zustandsbeobachter bildet die Erweiterung des Systemmodells von Gl. (4.53) um den mit der Verstärkungsmatrix K multiplizierten Ausgangsvektor $x(t)$ zu der Form (Bild 4.27)

$$\dot{\hat{q}}(t) = \hat{A}\hat{q}(t) + \hat{B}u(t) + Kx(t). \quad (4.60)$$

Das Ziel ist es nun, einen Beobachter so zu entwerfen, daß im stationären Fall für den Zustandsfehler

$$\tilde{q}(t) = [q(t) - \hat{q}(t)] = 0 \quad \text{für } t \rightarrow \infty \quad (4.61)$$

gilt. Mit den Systembeschreibungen von Gl. (4.52) und der Modellbeschreibung von Gl. (4.60) werden aus der Differenz beider Gleichungen Bedingungen für die Wahl der Matrizen $\hat{A}$, $\hat{B}$ und K abgeleitet, damit Gl. (4.61) erfüllt ist. Es gilt

$$\dot{\hat{q}}(t) - \dot{q}(t) = Aq(t) + Bu(t) - \hat{A}\hat{q}(t) - \hat{B}u(t) - KCq(t). \quad (4.62)$$

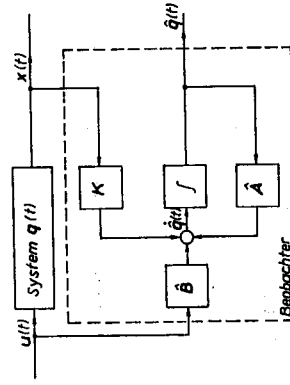


Bild 4.27
Zustandsbeobachtung
aus den gemessenen Eingangs- und Ausgangsgrößen

Durch Erweiterung von Gl. (4.62) um $\pm \dot{A}q(t)$ ergibt sich für

$$\begin{aligned} \dot{\hat{q}}(t) - \dot{q}(t) &= Aq(t) + Bu(t) + \hat{A}[q(t) - \hat{q}(t)] \\ &\quad - \hat{A}q(t) - \hat{B}u(t) - KCq(t) \end{aligned}$$

und durch weitere Umformungen

$$\begin{aligned} \dot{\hat{q}}(t) - \dot{q}(t) &= \hat{A}[q(t) - \hat{q}(t)] + q(t)[A - \hat{A} - KC] \\ &\quad + u(t)[B - \hat{B}]. \end{aligned} \quad (4.63)$$

Aus Gl. (4.63) ist zu erkennen, daß der Fehler und seine Ableitungen für $t \rightarrow \infty$ verschwinden, wenn

$$\begin{aligned} A - \hat{A} - KC &= 0 \\ B - \hat{B} &= 0 \end{aligned}$$

und damit für die Matrizen A , B und K gilt

$$\begin{aligned} \hat{A} &= A - KC \\ \hat{B} &= B; \end{aligned} \quad (4.64)$$

K frei wählbar unter Beachtung notwendiger negativer Eigenwerte von $\hat{A}$.

Damit ergibt sich aus Gl. (4.60) die folgende Gleichung für den Beobachter:

$$\dot{\hat{q}}(t) = [A - KC]\hat{q}(t) + Bu(t) + Kx(t). \quad (4.65)$$

Durch weitere Umformungen kann aus Gl. (4.65) die endgültige Form der Beschreibung des Luenberger-Beobachters gefunden werden. Sie lautet:

$$\dot{\hat{q}}(t) = A\hat{q}(t) + Bu(t) + K[x(t) - \hat{x}(t)]. \quad (4.66)$$

Die Struktur dieser Beobachterform ist im Bild 4.28 dargestellt.

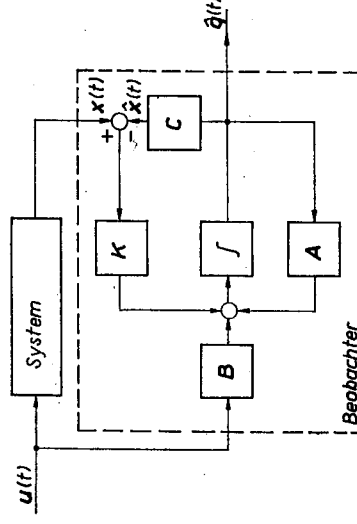


Bild 4.28
Zustandsbeobachtung
mit einem Luenberger-Beobachter

In diskreter Form lautet das Beobachtersgesetz analog zu Abschnitt 4.4.2 und Gl. (4.66):

$$\hat{q}(k+1) = A^*\hat{q}(k) + B^*u(k) + K[x(k) - \hat{x}(k)] \quad (4.67)$$

mit

$$\hat{x}(k) = C^*\hat{q}(k).$$

Damit besteht der Luenberger-Beobachter aus einem Modell des Systems, das um die durch die Verstärkungsmatrix K bewertete Differenz zwischen beobachtetem und vorhergesagtem Ausgangssignalvektor erweitert wird. Die in den Gln. (4.66) und (4.67) dargestellten Beziehungen entsprechen in Grundstruktur und Verhalten den rekursiven Schätzalgorithmen (s. Abschn. 5.4). Durch die Wahl der Elemente der Verstärkungsmatrix K ist bei diesem Beobachtertyp das dynamische Verhalten gut zu beeinflussen. Um die Wirkung der Elemente der Verstärkungsmatrix K auf die Konvergenz der Zustandsbestimmung darzustellen, soll als Ausgangspunkt die Vektordifferentialgleichung des Zustandsfehlers von Gl. (4.63) mit den Annahmen

$$\hat{A} = A - KC \quad \text{und} \quad \hat{B} = B$$

betrachtet werden. Damit gilt

$$[\dot{q}(t) - \dot{\hat{q}}(t)] = \hat{A}[q(t) - \hat{q}(t)]. \quad (4.68)$$

Die Lösung der Vektordifferentialgleichung 1. Ordnung von Gl. (4.68) lautet dann

$$[q(t) - \hat{q}(t)] = e^{\hat{A}t} [q(t_0) - \hat{q}(t_0)]; \quad (4.69)$$

$q(t_0)$ Anfangszustand.

Die Gl. (4.69) zeigt deutlich, daß der Einschwingvorgang des Zustandsfehlers $\tilde{q}(t)$ wesentlich durch die Eigenwerte der Matrix $\hat{A}$ bestimmt wird.

mit

$$\hat{A} = \begin{bmatrix} 0 & \dots & 0 & -\hat{a}_0 \\ 1 & & & \\ 0 & & & \\ \vdots & & & \\ 0 & \dots & 1 & -\hat{a}_{n-1} \end{bmatrix}$$

$$\hat{b}^T = [\hat{b}_0, \hat{b}_1, \dots, \hat{b}_{n-1}]$$

$$\hat{q}^T = [\hat{q}_1(t) \dots \hat{q}_n(t)].$$

Da die Koeffizienten $\hat{a}_i$; $i = 0, 1, \dots, n - 1$, aufgrund der Vorgabe der Eigenwerte für die Matrix $\hat{A}$ durch die Gl. (4.70) bekannt sind, können die Elemente des Vektors $\hat{k}$ aus der Bedingung in Gl. (4.64) einfach ermittelt werden:

$$\hat{A} = A - k\epsilon^T \quad (4.74)$$

und, ausgeschrieben für die einzelnen Elemente:

$$\begin{bmatrix} 0 & \dots & 0 & -\hat{a}_0 \\ 1 & & & \\ 0 & & & \\ \vdots & & & \\ 0 & \dots & 1 & -\hat{a}_{n-1} \end{bmatrix} = \begin{bmatrix} 0 & \dots & 0 & -a_0 - k_1 \\ 1 & & & \\ 0 & & & \\ \vdots & & & \\ 0 & \dots & 1 & -a_{n-1} - k_n \end{bmatrix}$$

Aus dem Vergleich beider Matrizen folgt die Bestimmungsvorschrift für die Elemente des Verstärkungsvektors k zu

$$\hat{a}_0 = a_0 + k_1$$

$$\hat{a}_{n-1} = a_{n-1} + k_{n-1}.$$

Somit gilt allgemein

$$k_i = \hat{a}_{i-1} - a_{i-1}; \quad i = 1, 2, \dots, n. \quad (4.75)$$

Damit sind alle Elemente eines Beobachters festgelegt. Seine Struktur ist im Bild 4.29 dargestellt.

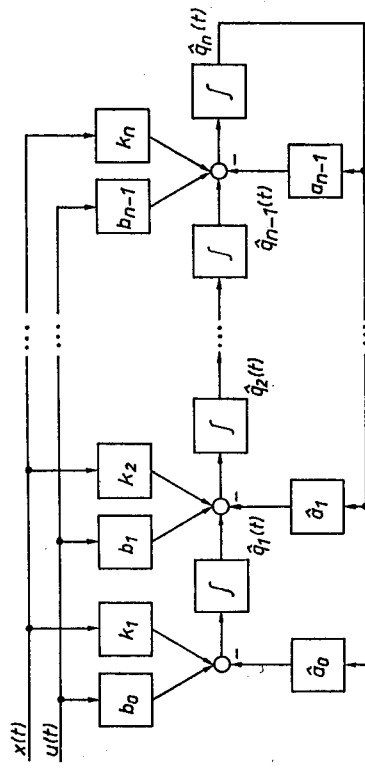


Bild 4.29. Struktur eines Luenberger-Beobachters für 1/1-System

182 4. Bestimmung des dynamischen Verhaltens ungestörter Systeme

Für ein schnelles Abklingen von e^u sind offenbar Eigenwerte von $\hat{A}$ erforderlich, die möglichst weit links auf der imaginären Achse liegen müssen (d.h. großer negativer Realteil). Praktisch wird man aber die Eigenwerte nicht zu groß wählen, weil der Beobachter dann als nahezu differenzierendes System arbeitet. In der Regel werden die Eigenwerte der Matrix $\hat{A}$ so vorgegeben, daß sie links von denen der Matrix A liegen. Damit klingt der Beobachtungsvorgang immer noch schneller ab als die Vorgänge im beobachteten System. Sind die Eigenwerte der Matrix $\hat{A}$ durch diese Überlegungen festgelegt worden, muß man die Verstärkungsmatrix K so bestimmen, daß die Matrix

$$\hat{A} = A - K\epsilon$$

die vorgegebenen Eigenwerte hat. Theoretisch existieren unendlich viele Matrizen K , die die gleichen Eigenwerte aufweisen. Sie besitzen alle das charakteristische Polynom:

$$\det [pI - \hat{A}] = (p - p_1)(p - p_2) \dots (p - p_n) \quad (4.70)$$

bzw.

$$\det [pI - A + KC] = p^n + \hat{a}_{n-1}p^{n-1} \dots + \hat{a}_0. \quad (4.71)$$

Die Existenz einer Lösung für Gl. (4.70) wurde unter der Voraussetzung, daß das System

$$\dot{q}(t) = Aq(t), \quad x(t) = Cq(t) \quad (4.71)$$

beobachtbar ist, in [4.17] nachgewiesen.

Dabei soll unter der Beobachtbarkeit die Eigenschaft des Systems charakterisiert werden, daß aus den gemessenen Ausgangsgrößen $x(t)$ der Zustand $q(t)$ eindeutig ermittelt werden kann.

Die Festlegung der Elemente der Verstärkungsmatrix K ist besonders übersichtlich, wenn die Zustandsgleichungen in die Beobachtungsnormalform [1.2, 4.15, 4.17] übergeführt werden. Für ein System mit einem Ein- und Ausgang gilt dann

$$\dot{q}(t) = Aq(t) + bu(t) \quad (4.72)$$

$$x(t) = c^T q(t)$$

mit

$$A = \begin{bmatrix} 0 & \dots & 0 & -a_0 \\ 1 & & & \\ 0 & & & \\ \vdots & & & \\ 0 & \dots & 1 & a_{n-1} \end{bmatrix}$$

$$b^T = [b_0, b_1, \dots, b_{n-1}], \quad c^T = [0, 0, \dots, 1]$$

$$q^T = [q_1(t), q_2(t), \dots, q_n(t)]$$

$$G(p) = \frac{b_0 + b_1 p + \dots + b_{n-1} p^{n-1}}{a_0 + a_1 p + \dots + p^n}.$$

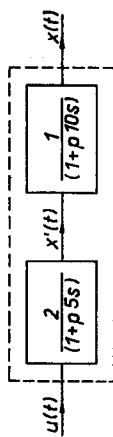
Entsprechend Gl. (4.60) lautet dann die Beobachternormalform:

$$\begin{aligned} \dot{\hat{q}}(t) &= \hat{A}\hat{q}(t) + \hat{b}u(t) + Kx(t) \\ \hat{x}(t) &= \hat{c}^T \hat{q}(t) \end{aligned} \quad (4.73)$$

Untersuchung des Verhaltens des Luenberger-Beobachters (diskontinuierlich)

Beispiel 4.1

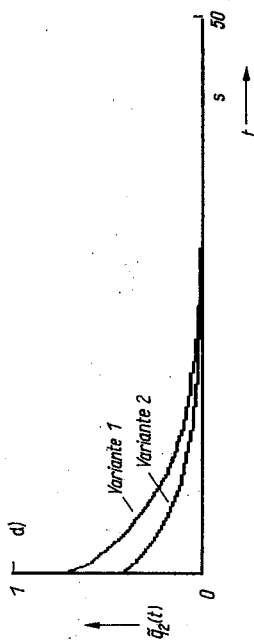
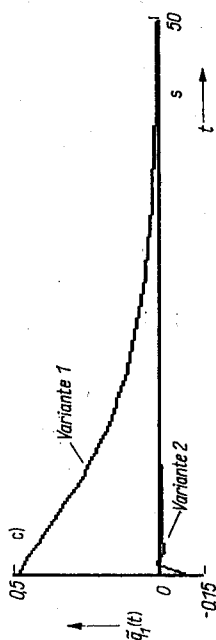
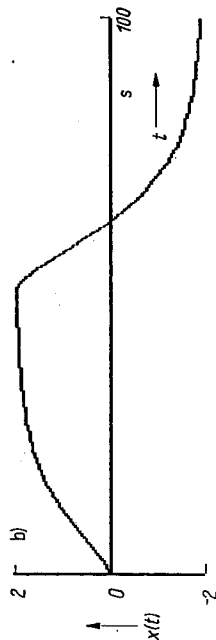
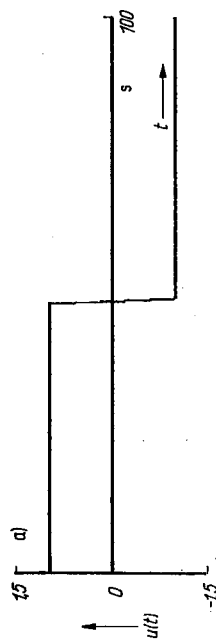
Gegeben: Abgetastete Werte des Eingangs- und Ausgangssignals des Systems (Bild a und Blockschaltbild)



Zustandsgleichungen mit $q_1(t) \triangleq x(t)$, $q_2(t) \triangleq x'(t)$,

$$\begin{bmatrix} \dot{q}_1(t) \\ \dot{q}_2(t) \end{bmatrix} = \begin{bmatrix} -0,1 \text{ s}^{-1} & 0,1 \text{ s}^{-1} \\ 0 & -0,2 \text{ s}^{-1} \end{bmatrix} \begin{bmatrix} q_1(t) \\ q_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0,4 \text{ s}^{-1} \end{bmatrix} u(t),$$

$$x(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} q_1(t) \\ q_2(t) \end{bmatrix},$$



Untersuchung des Verhaltens des Luenberger-Beobachters (diskontinuierlich)

Beispiel 4.1

Abtastzeit $T = 1 \text{ s}$,
Beobachtungszeit: 100 s .

Gesucht: Verlauf der Zustandsfehler $\hat{q}_i(t)$, $i = 1, 2$, in Abhängigkeit von der Lage der Pole des Beobachters z_B zu denen des Systems z_S in der z -Ebene.

Ergebnis: Siehe Bilder b, c, d zu Beispiel 4.1.

1. Verlauf der Eingangs- und Ausgangsgröße des Systems (Bild b),
2. Verlauf der Schätzfehler $\hat{q}_i(t)$ bei gleicher Lage der Pole des Systems und des Beobachters, d.h. $z_S = z_B$ (Bild c),
3. Verlauf des Zustandsfehlers $\hat{q}_i(t)$, wenn die Pole des Beobachters links von denen des Systems in der z -Ebene liegen (Bild d).

Die dargestellten Verläufe der Zustandsfehler $\hat{q}_i(t)$ zeigen, daß bei gleicher Wahl der Lage der Pole von System und Beobachter (d.h. Verstärkungsfaktoren $k_i = 0$) die Konvergenz sehr langsam erfolgt. Je weiter die Lage der Pole des Beobachters von denen des Systems nach links verschoben werden, um so größer werden die Verstärkungsfaktoren k_i , und um so schneller konvergieren die Zustandsfehler zu Null. In dem dargestellten Fall tritt bereits beim Einschwingverhalten des Zustandsfehlers $\hat{q}_1(t)$ ein differenzierendes Verhalten auf. Würden bei dieser gewählten Beobachterstruktur Störungen auf das System wirken, ist eine Zustandsermittlung nicht mehr möglich.

Im diskontinuierlichen Fall kann der Beobachter in analoger Weise entworfen werden. Wenn wieder von in die Beobachternormalform übergeführten Zustandsgleichungen ausgegangen wird, sind die Elemente des Verstärkungsvektors in Gl. (4.67) aus der Beziehung

$$k_i = \hat{a}_{i-1} - a_{i-1}; \quad i = 1, 2, \dots, n$$

zu ermitteln. Die Parameter a_i sind in diesem Fall durch die diskrete Übertragungsfunktion

$$G(z) = \frac{b_0 + b_1 z^{-1} + \dots + b_{n-1} z^{-n}}{a_0 + a_1 z^{-1} + \dots + z^{-n}} \quad (4.76)$$

und die Parameter $\hat{a}_i$ durch die vorgegebenen Eigenwerte z_i in der charakteristischen Gleichung des Beobachters – s. Analogie in Gl. (4.70) und Gl. (4.74) – gegeben [3.24]:

$$\det [zI - A^*] = (z - z_1)(z - z_2) \dots (z - z_n) = \hat{a}_0 + \hat{a}_1 z + \dots + z^n. \quad (4.77)$$

In [3.24] ist eine rechenstechnisch sehr günstige Entwurfsstrategie, die auf der Anwendung spezieller Transformationsmatrizen beruht, enthalten.

Das Verhalten des Luenberger-Beobachters bei der Ermittlung der Zustandsgrößen ist für einen einfachen Testfall im Beispiel 4.1 für den diskontinuierlichen Fall dargestellt. Die Bestimmung der Verstärkungsmatrix K bei Mehrgrößensystemen ist wesentlich unübersichtlicher. Im Rahmen dieses Buches soll auf dieses Problem nicht weiter eingegangen und auf weiterführende Literatur hingewiesen werden [1.52, 4.14, 4.15]. Dies gilt auch für das Problem des Entwurfs von Beobachtern reduzierter Ordnung (wenn einige Zustandsgrößen direkt meßbar sind) und für die Zustandsbeobachtung zeitvarianter Systeme.

Sind die Zustandsgrößen durch bekannte Störgrößen verfälscht, so kann man diesen Einfluß durch den klassischen Weg einer „Störgrößenaufschaltung“ eliminieren. Für die Zustandsgleichung mit Störung $v(t)$ gilt dann

$$\dot{q}(t) = Aq(t) + Bu(t) + Ev(t).$$

(4.78)

Durch Addition eines Terms der Größe

$$Bu^*(t) = -Ev(t)$$

in Gl. (4.78) bzw. von $u^*(t)$ zum Eingangsvektor $u(t)$ kann dieser Störeinfluß kompensiert werden. Treten nichtmeßbare Störungen auf Zustände und Ausgänge auf, versagt das Beobachterprinzip, und die Zustände müssen mit einer Schätzvorschrift (Kalman-Bucy-Filter) ermittelt werden (s. Abschn. 6.5).

4.5. Übungsaufgaben zum Abschnitt 4

Aufgabe 4.1

Gegeben sind die in Tafel 4.8 dargestellten Werte der Sprungantwort eines ungestörten Systems. Bestimmen Sie die Übertragungsfunktion durch Approximation mit zwei unterschiedlichen Zeitkonstanten! Die Sprunghöhe der Testsignale beträgt $u_0 = 1$.

Tafel 4.8
Funktionswerte der Sprungantwort
zu Aufgabe 4.1

t/s	x(t)	t/s	x(t)
2	0	24	3,97
4	0,16	26	4,15
6	0,54	28	4,30
8	1,02	30	4,43
10	1,52	32	4,54
12	2,00	34	4,62
14	2,44	36	4,70
16	2,84	38	4,76
18	3,19	40	4,81
20	3,49	42	4,85
22	3,75	44	4,88
		46	4,91
		48	4,93
		50	4,95
		52	4,97
		54	4,98
		56	4,99
		58	5,00
		60	5,00
		62	5,00
		64	5,00
		66	5,00

Aufgabe 4.2

Gegeben sind die in Tafel 4.9 dargestellten Meßwerte der Sprungantwort eines gestörten Systems. Die Sprunghöhe des Testsignals beträgt $u_0 = 1$. Ermitteln Sie die Übertragungsfunktion für die Struktur

a)
$$G(p) = \frac{K}{(1 + pT)^n}$$

b)
$$G(p) = \frac{K}{\prod_{i=1}^n \left(1 + p \frac{T_i}{T_i'}\right)}$$

Überprüfen Sie die erhaltenen Modelle an den Stellen $t_1 = 20$ s und $t_2 = 40$ s!

Tafel 4.9
Funktionswerte der Sprungantwort
zu Aufgabe 4.2

t/s	x(t)	t/s	x(t)
0	0	30	3,93
2	-0,43	32	4,04
4	0,16	34	4,30
6	0,44	36	4,35
8	1,02	38	4,55
10	0,82	40	4,58
12	1,27	42	4,64
14	1,43	44	4,29
16	1,85	46	4,66
18	2,57	48	4,70
20	2,71	50	4,95
22	2,87	52	4,76
24	3,23	54	4,66
26	3,29	56	4,60
28	3,75	58	4,58

Aufgabe 4.3

Gegeben ist in Tafel 4.10 der Kurswinkelverlauf eines Schiffes als Reaktion auf eine Änderung des Steuerruders um 10° . Bestimmen Sie Struktur und Parameter dieses Systems
a) mit dem Verfahren nach Schwarze;
b) mit dem Verfahren nach Sponer!

Tafel 4.10
Funktionswerte der Sprungantwort zu Aufgabe 4.3

t/s	x(t)	t/s	x(t)
0	0	12	19,9
2	1,0	14	24,4
4	3,2	16	28,8
6	6,7	18	33,3
8	11,0	20	37,7
10	15,5		

Aufgabe 4.4

Ermitteln Sie aus den in Tafel 4.11 gegebenen Stützwerten der Impulsantwort die Parameter der Übertragungsfunktion

$$G(p) = \frac{K}{(1 + pT)^n}$$

Überprüfen Sie die Güte des Modells an den Stellen
 $t = 3$ s, 6 s, 9 s, 12 s!

Tafel 4.11
Funktionswerte der Impulsantwort zu Aufgabe 4.4

t/s	x(t)	t/s	x(t)
0	0	10	0,11
1	0,12	11	0,08
2	0,30	12	0,06
3	0,385	13	0,05
4	0,405	14	0,03
5	0,375	15	0,02
6	0,315	16	0,015
7	0,25	17	0,010
8	0,195	18	0,007
9	0,15	19	0,005

Aufgabe 4.5

Aus den Meßwerten der Impulsantwort eines gestörten Systems (s. Tafel 4.12) ist unter Beachtung der Parameter des realen Eingangsimpulses ($\Delta T = 5 \text{ s}$, $u_0 = 4$) ein Modell mit gestaffelten Zeitkonstanten zu ermitteln. Vergleichen Sie das Ergebnis mit den wahren Parametern des Systems $K = 5$, $T_1 = 10 \text{ s}$, $T_2 = 5 \text{ s}$!

Tafel 4.12
Funktionswerte der Impulsantwort
zu Aufgabe 4.5

t/s	x(t)	t/s	x(t)	t/s	x(t)
0	-0,25	22	2,65	44	0,24
2	0,63	24	2,45	46	0,27
4	2,02	26	1,73	48	0,55
6	4,06	28	1,69	50	-0,27
8	4,54	30	1,19	52	0,23
10	4,59	32	1,02	54	-0,09
12	4,98	34	1,09	56	-0,05
14	4,28	36	1,04	58	0,12
16	3,87	38	0,32	60	-0,09
18	3,80	40	0,56	62	0,26
20	2,88	42	0,51	64	-0,25

Aufgabe 4.6

Ermitteln Sie durch eine Geradenapproximation des Amplitudenfrequenzgangs die Übertragungsfunktion eines Phasenminimalsystems aus den in Tafel 4.13 dargestellten Meßwerten!

Tafel 4.13
Funktionswerte des Amplitudenfrequenzgangs zu Aufgabe 4.6

ω/s^{-1}	G /dB	ω/s^{-1}	G /dB
0,1	40,0	4,0	-0,5
0,14	37,0	5,6	-5,0
0,2	34,0	8	-9,0
0,28	31,0	10	-12,0
0,4	27,0	14	-17,0
0,56	24,0	20	-22,5
0,8	20,0	28	-28,0
1,0	17,5	40	-34,0
1,4	13,5	56	-39,5
2,0	8,5	80	-45,5
2,8	4,0	100	-49,5

Aufgabe 4.7

Bei einer Frequenzgangmessung an einem System, das aus einer Reihenschaltung von I- und/oder P-Tn-Gliedern bestehen kann, wurden folgende Werte ermittelt:

$$\begin{aligned} \omega_0 &= 0 && \text{bei } \varphi(\omega_0) = 0 \\ \omega_1 &= 5,97 \cdot 10^{-2} \text{ s}^{-1} && \text{bei } \varphi(\omega_1) = -45^\circ \\ \omega_2 &= 15,1 \cdot 10^{-2} \text{ s}^{-1} && \text{bei } \varphi(\omega_2) = -90^\circ \\ \omega_3 &= 69 \cdot 10^{-2} \text{ s}^{-1} && \text{bei } \varphi(\omega_3) = -180^\circ \\ \omega_4 &\rightarrow \infty && \text{bei } \varphi(\omega_4) = -270^\circ. \end{aligned}$$

Bei der Frequenz ω_0 betrug das Verhältnis $x(t)/u(t) = 2,7$. Bestimmen Sie die Beschreibung des Frequenzgangs nach dem Verfahren von Szewda!

5. Bestimmung des statischen Verhaltens von gestörten Systemen

5.1. Ausgewählte Grundlagen der Schätztheorie

5.1.1. Aufgabe der Schätzmethoden

In den Abschnitten 5 und 6 soll das statische und das dynamische Verhalten von gestörten Systemen ermittelt werden. Entsprechend den Vereinbarungen im Abschnitt 1.4 besteht die Aufgabe der Schätzmethoden darin, ein Modell an das System auf der Grundlage der gemessenen Eingangs- und Ausgangsgrößen und von A-priori-Informationen so anzupassen, daß eine gewählte Zielfunktion Q minimal oder maximal wird. Das Grundschema der Schätzaufgabe ist im Bild 5.1 dargestellt. Durch die auftretende Störung werden die Struktur $\hat{f}$ und die Parameter $\hat{s}$ von Ein-/Ausgangs-Modellen der Form

$$\hat{x} = \hat{f}(m, \hat{s}) \quad (5.1)$$

oder die Zustandsgrößen $\hat{q}$ bei Zustandsmodellen der Form

$$\dot{\hat{q}}(t) = \hat{A}\hat{q}(t) + \hat{B}u(t) \quad (5.2)$$

selbst Zufallsgrößen. Damit ist eine Beschreibung mit den Methoden der mathematischen Statistik (s. Abschn. 2) notwendig. Beschränken wir die Aufgabe auf die wertmäßige Ermittlung von r Parametern (Parameterschätzung) für ein System mit l Eingängen und einem Ausgang, so ist der Parametervektor aus n Beobachtungen zu ermitteln. Es gilt

$$\hat{s} = \hat{s} \{m(1T), \dots, m(nT); x(1T), \dots, x(nT)\}. \quad (5.3)$$

Die Güte der Schätzung hängt damit u. a. von der Anzahl der Beobachtungen, den vorhandenen A-priori-Informationen und der gewählten Zielfunktion ab.

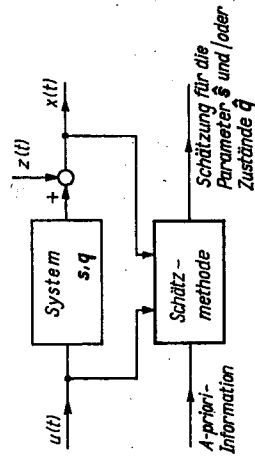


Bild 5.1
Grundschema der Schätzaufgabe

Es werden deshalb in den folgenden Abschnitten die Grundideen wesentlicher Schätzmethoden, Kriterien zur Beurteilung der Schätzung, Möglichkeiten zur Fehlerbildung zwischen System- und Modellsignalen und die Grundstrategien zur Minimierung der Zielfunktion vorgestellt. Alle Ausführungen werden für Modelle des statischen Systemverhaltens durchgeführt. Eine Erweiterung auf Modelle für das dynamische Verhalten erfolgt im Abschnitt 6.

5.1.2. Grundideen wesentlicher Schätzmethoden

Den Ausgangspunkt für die Herleitung der wesentlichen Schätzmethoden bilden in diesem Buch der Umfang an A-priori-Wissen über Wahrscheinlichkeitsverteilungen der Störungen, der Parameter und der Ausgangsgrößen sowie über die Kosten/den Verlust, der durch den Schätzfehler entsteht. Damit stehen als A-priori-Informationen maximal zur Verfügung:

1. die Kosten- oder Verlustfunktion $C(\xi, s)$,
2. die bedingte Wahrscheinlichkeitsdichtefunktion $p(x/s)$,
3. die Wahrscheinlichkeitsdichtefunktion $p(x)$,
4. die Wahrscheinlichkeitsdichtefunktion $p(z)$ und die Kovarianzmatrix $Z = E\{zz^T\}$ der Störung.

Als wesentliche Schätzmethoden, die die genannten A-priori-Informationen unterschiedlich verwenden, sind

- die Bayes-Schätzung,
- die Maximum-Likelihood-Schätzung
- die Markov-Schätzung
- die Regression

entwickelt worden. Es wird im folgenden gezeigt, daß der Umfang der verwendeten A-priori-Information von der Bayes-Schätzung zur Regression abnimmt und daß alle Hauptmethoden unter bestimmten Annahmen ineinander übergeführt werden können.

5.1.2.1. Bayes-Schätzung

Bei der *Bayes-Schätzung* wird davon ausgegangen, daß der bedingte Erwartungswert der Kostenfunktion $C(\xi, s)$, das Risiko für die Wahl von ξ unter Berücksichtigung der Beobachtungen x , minimal werden soll. Es gilt damit

$$E_{x/s} \{C(\xi, s)\} = \int_n C(\xi, s) p(x/s) d^n x \quad (5.4)$$

mit

$$\int_n \dots d^n x, \quad n\text{-faches Integral,}$$

$$d^n x = dx_1, dx_2, \dots, dx_n.$$

Unter Berücksichtigung der Wahrscheinlichkeit für die Werte der Systemparameter $p(s)$ erhält man aus Gl. (5.4) das durchschnittliche Risiko für die Schätzsituation zu

$$R(\xi) = E_s \{E_{x/s} \{C(\xi, s)\}\} \quad (5.5)$$

und somit

$$R(\xi) = \int_r \int_n C(\xi, s) p(x/s) p(s) d^n x d^r s.$$

Durch Anwendung der Bayes-Regel, die der Methode den Namen gab, kann mit den Beziehungen

$$p(x/s)p(s) = p(x, s) = p(s/x)p(x) \quad (5.6)$$

Gl. (5.5) umgeschrieben werden in

$$R(\xi) = p(x) \int_r C(\xi, s) p(s/x) d^r s \int_n d^n x. \quad (5.7)$$

Beachtet man nun, daß $p(x) \geq 0$, kann Gl. (5.7) minimiert werden, indem das zweite Integral bei den beobachteten Werten $x = c$ minimiert wird. Damit gilt als Bestimmungsvorschrift

$$\min_{\xi} \int_r C(\xi, s) p(s/c) d^r s \quad (5.8)$$

und als notwendige Bedingung

$$\frac{\partial}{\partial \xi} \int_r C(\xi, s) p(s/c) d^r s \Big|_{\xi=s} = 0. \quad (5.9)$$

Für einen Parameter s und eine quadratische Kostenfunktion $C(\xi, s) = c(\xi - s)^2$ sind im Bild 5.2 die Werte des Integrals von Gl. (5.8) dargestellt. Es ist zu sehen, daß das durchschnittliche Risiko an der Stelle $\xi = s$ minimal wird. Damit benötigt die Bayes-Schätzung folgende A-priori-Informationen

- die Kosten-/Risikofunktion $C(\xi, s)$
- die bedingte Wahrscheinlichkeitsdichtefunktion $p(x/s)$
- die Wahrscheinlichkeitsdichtefunktion $p(s)$.

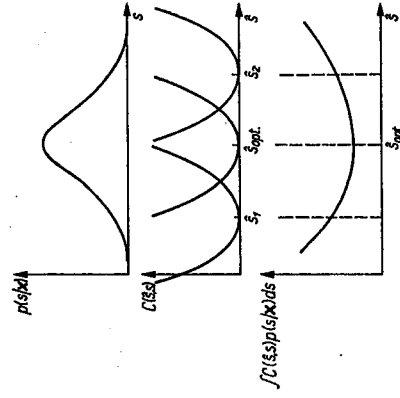


Bild 5.2
Parametermittlung mit dem Bayes-Verfahren

5.1.2.2. Maximum-Likelihood-Schätzung

Wird die Kenntnis der Kostenfunktion $C(\xi, s)$ fallengelassen, so ist grundsätzlich der Parametervektor s so zu wählen, daß die bedingte Wahrscheinlichkeit $p(s/x)$ ein Maximum annimmt. Aus Gl. (5.6) ergibt sich dann

$$\max_s p(s/x) = p(x/s)p(s)/p(x). \quad (5.10)$$

Wird nun ferner angenommen, daß die Verteilung des Parametervektors $p(s)$ nicht bekannt ist und durch eine Gleichverteilung $p(s) = \text{const}$ ersetzt wird, geht die Aufgabe von Gl. (5.10) in die Vorschrift

$$\max_s p(s|x) = \frac{\text{const}}{p(x)} \max_s p(x, s) \quad (5.11)$$

über. Die Wahrscheinlichkeitsdichtefunktion $p(x, s)$ wird nach Einsetzen der Beobachtungen für x in der Form

$$x(1T) = c_1, \dots, x(nT) = c_n \quad \text{bzw.} \quad x = c$$

und des zu schätzenden Parametervektors s als *Likelihood-Funktion* L bezeichnet:

$$L(c_1, \dots, c_n, s) = L(c, s). \quad (5.12)$$

Für das Auffinden des Maximums dieser Funktion, d. h. der Bestimmung des Parametervektors $\hat{s} = s$, müssen folgende Bedingungen erfüllt sein:

$$\left. \frac{\partial}{\partial s} L(c, s) \right|_{s=\hat{s}} = 0 \quad (5.13)$$

oder als logarithmische Funktion

$$\left. \frac{\partial}{\partial s} \ln L(c, s) \right|_{s=\hat{s}} = 0.$$

Damit benötigt die *Maximum-Likelihood-Schätzung* nur die Wahrscheinlichkeitsdichtefunktion $p(x, s)$ als A-priori-Information.

5.1.2.3. Markov-Schätzung – ein Spezialfall der verallgemeinerten Regression

Werden von den möglichen A-priori-Informationen nur noch die Wahrscheinlichkeitsfunktion $p(z)$ und die Kovarianzmatrix $Z = E\{zz^T\}$ benutzt, kann eine Schätzvorschrift abgeleitet werden, die folgende Eigenschaften besitzt:

- Die Schätzung ist erwartungstreu, d. h. $E\{\hat{s}\} = s$.
- Die Schätzung der Parameter erfolgt mit minimaler Varianz.

Wenn angenommen wird, daß der Störvektor eine n -fache Normalverteilung ist, gilt analog zu Abschnitt 2.2.2

$$p(z) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|Z|^{1/2}} \exp - \frac{1}{2} (z^T Z^{-1} z) \quad (5.14)$$

mit

$$E\{z\} = 0, \quad E\{zz^T\} = Z.$$

Wird der Logarithmus der Wahrscheinlichkeitsdichtefunktion für $z = x - M\hat{s}$ von Gl. (5.14) gebildet, erhält man

$$\begin{aligned} \ln p(x - M\hat{s}) &= -\frac{1}{2} \ln \{(2\pi)^n |Z|\} \\ &\quad - \frac{1}{2} \{(x - M\hat{s})^T Z^{-1} (x - M\hat{s})\}. \end{aligned} \quad (5.15)$$

M ist eine allgemeine Meßwertmatrix, die für die speziellen Probleme jeweils nachfolgend präzisiert wird (s. auch Abschn. 1.4).

Das Maximum aus Gl. (5.15) erhält man entsprechend Gl. (5.13) zu

$$\frac{\partial}{\partial \hat{s}} (x - M\hat{s})^T Z^{-1} (x - M\hat{s}) = 0 \quad (5.16)$$

und mit dem Fehlervektor e zwischen System- und Modellausgang $e = (x - M\hat{s})$ zu

$$\frac{\partial}{\partial \hat{s}} e^T Z^{-1} e = 0.$$

Wenn die Inverse der Matrix $M^T Z^{-1} M$ existiert, erhält man die Vorschrift für die *Markov-Schätzung* aus Gl. (5.16) zu

$$\hat{s} = [M^T Z^{-1} M]^{-1} M^T Z^{-1} x. \quad (5.17)$$

Damit ist auch gezeigt, daß die Markov-Schätzung die Eigenschaften der Maximum-Likelihood-Schätzung besitzt, wenn die Störung normal verteilt und ihre Kovarianzmatrix Z bekannt ist.

Setzt man für die Kovarianzmatrix Z eine allgemeine Wichtungsmatrix W ein, erhält man aus den Gln. (5.16) und (5.17) die Schätzvorschrift der verallgemeinerten Regression zu

$$\hat{s} = [M^T W^{-1} M]^{-1} M^T W^{-1} x. \quad (5.18)$$

5.1.2.4. Methode der kleinsten Fehlerquadrate – Regression

Wird davon ausgegangen, daß die Störung z als „weißes Rauschen“ (s. Abschn. 3) angesehen werden kann, ergibt sich die Kovarianzmatrix Z zu

$$Z = cI. \quad (5.19)$$

Damit geht die Schätzvorschrift der Markov-Schätzung Gl. (5.17) in die der *Regression* über:

$$\hat{s} = [M^T M]^{-1} M^T x. \quad (5.20)$$

Die Methode der Regression besitzt unter den oben genannten Bedingungen die Eigenschaften der Maximum-Likelihood-Schätzung. Ist die Störung nicht normalverteilt und korreliert, wird zwar der mittlere Ausgangsfehler minimal, aber die Varianz der Parameter ist größer als bei der Markov-Schätzung (s. Abschn. 5.2.3).

Damit wurde gezeigt, daß die vier grundlegenden Verfahren der Schätztheorie über die A-priori-Informationen verknüpft werden können. Von der Bayes-Schätzung bis zur

Tafel 5.1. Zusammenstellung der grundsätzlichen Schätzmethoden

Methode	A-priori-Informationen	Zielfunktion	Ziel
Bayes-Schätzung	$C(\hat{s}, s)$ $p(x/s)$ $p(s)$	$\int C(\hat{s}, s) p(s) d^*s$	Minimum
Maximum-Likelihood-Schätzung	$p(x, s)$	$L(c, \hat{s})$	Maximum
Markov-Schätzung	Z	$e^T Z^{-1} e$	Minimum
Regression	–	$e^T e$	Minimum

Regression nimmt der Umfang an notwendiger A-priori-Information ab. In Tafel 5.1 sind für die vier Grundsätzverfahren das verwendete Zielkriterium, die A-priori-Informationen und das Ziel der Schätzung in einer Übersicht dargestellt. Da die Ermittlung der Kostenfunktion $C(\hat{s}, s)$ und der mehrdimensionalen Wahrscheinlichkeitsfunktionen $p(x/s)$ und $p(s)$ selten möglich ist, wird im Rahmen dieses Buches auf die weitere Betrachtung der Bayes- und der Maximum-Likelihood-Schätzung verzichtet. Weiterführende Darstellungen dieser Schätzverfahren sind in [1.35, 1.36, 1.38, 1.47] enthalten.

5.1.3. Kriterien zur Beurteilung der Schätzung

Durch die auftretende Störung $z(t)$ werden die zu schätzenden Parameter s und/oder Zustandsgrößen $\hat{q}$ selbst Zufallsgrößen und unterliegen damit den Gesetzen der Statistik. In diesem Abschnitt sollen die wesentlichsten Kriterien zur Beurteilung der Schätzung bzw. des Schätzverlaufs vorgestellt werden. Wie bereits im Abschnitt 2 gezeigt, kann das Verhalten von Zufallsgrößen statisch durch ihre Verteilung oder durch die entsprechenden Momente beschrieben werden.

5.1.3.1. Verwendung von Verteilungen

Die meiste Information über die Güte der Schätzung (des Schätzverlaufs) ist in der Verteilung der geschätzten Parameter

$$p(\hat{s}, T_0) \quad \text{bzw.} \quad p(\hat{s}, n);$$

T_0 Beobachtungsintervall
 n Anzahl der Beobachtungen

enthalten. Zur Beurteilung der Güte der Schätzung wird die Lage der Verteilung, besonders ihr Schwerpunkt $E\{\hat{s}\}$, betrachtet.

Wenn der Erwartungswert der Schätzung gleich dem richtigen Parametervektor ist, spricht man von einer *unverschobenen Schätzung*. Dann gilt

$$E\{\hat{s}\} = s. \quad (5.21)$$

Ist Gl. (5.21) nicht erfüllt, liegt eine *verschobene Schätzung* vor. Für den eindimensionalen Fall sind diese beiden Fälle im Bild 5.3 dargestellt.

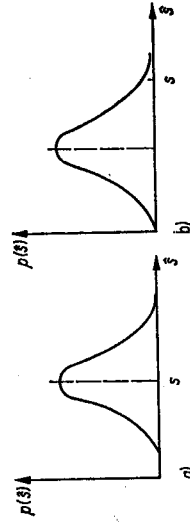


Bild 5.3
Wahrscheinlichkeitsverteilungen für einen geschätzten Parameter
a) unverschobene Schätzung
b) verschobene Schätzung

Da die Schätzung in der Regel auch eine Funktion der Beobachtungszeit T_0 /Beobachtungsanzahl n ist, muß das Konvergenzverhalten der Verteilungen der geschätzten Parameter untersucht werden. Eine Schätzung wird asymptotisch unverschoben bezeichnet, wenn unabhängig von der Beobachtungszeit/Beobachtungsanzahl Gl. (5.21) gilt. Für den eindimensionalen Fall ist dieses Verhalten im Bild 5.4a dargestellt. Nähert sich der Schwerpunkt der Verteilung der geschätzten Parameter mit größer werdender Beobachtungszeit/Beobachtungsanzahl immer mehr dem wahren Parametervektor, wird der Schätzverlauf als asymptotisch verschoben bezeichnet. Er ist im Bild 5.4b dargestellt.

Das Problem der Beurteilung der Güte der Schätzung anhand der Verteilungen der geschätzten Parameter liegt in ihrer Bestimmung und Darstellung. Einfacher und bei Annahme einer Normalverteilung der Parameterschwankungen ausreichend ist die Beschreibung der Schätzung durch die Momente.

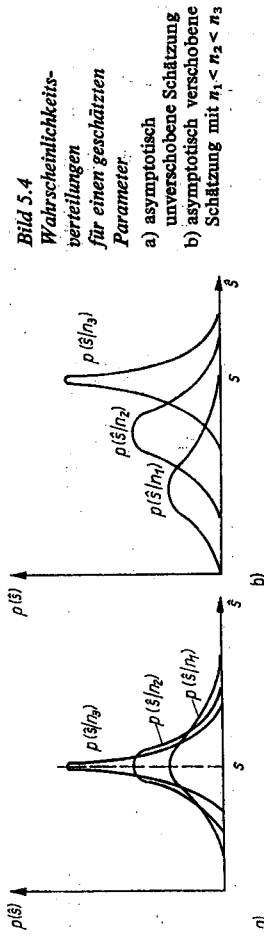


Bild 5.4
Wahrscheinlichkeitsverteilungen für einen geschätzten Parameter
a) asymptotisch unverschobene Schätzung
b) asymptotisch verschobene Schätzung mit $n_1 < n_2 < n_3$

5.1.3.2. Verwendung von Momenten

Zur Beurteilung der Schätzgüte werden das erste normale Moment (d. h. der Erwartungswert $E\{\dots\}$) und das zweite zentrale Moment (d. h. die var $\{\dots\}$) verwendet.

Eine Schätzung wird als *erwartungstreu* bezeichnet, wenn für jede Beobachtungsgröße (T_0, n) gilt

$$E\{\hat{s}\} = s.$$

Konvergiert der Parametervektor $\hat{s}$ mit Zunahme der Beobachtungsgröße (T_0, n) gegen den wahren Parametervektor, wird die Schätzung als *asymptotisch erwartungstreu* bezeichnet. Es gilt damit

$$\lim_{T_0, n \rightarrow \infty} E\{\hat{s}\} = s. \quad (5.22)$$

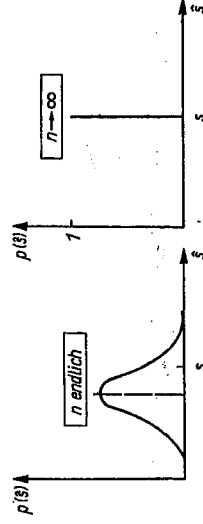


Bild 5.5
Verhalten der Wahrscheinlichkeitsverteilung des geschätzten Parameters bei einer konsistenten Schätzung

Die bis jetzt betrachteten Kriterien enthalten keine Aussage über die Varianz der Schätzungen. Neben einer guten Übereinstimmung des mittleren Modellparametervektors mit dem Systemparametervektor soll die Varianz um den Erwartungswert der Schätzung minimal sein. Diese Bedingung erfüllt eine *konsistente Schätzung*. Für sie gilt

$$P \left(\lim_{T_0, n \rightarrow \infty} |\hat{s} - s| = 0 \right) = 1 \quad (5.23)$$

mit den hinreichenden Bedingungen (Bild 5.5)

$$\lim_{T_0, n \rightarrow \infty} E\{\hat{s}\} = s$$

$$\lim_{T_0, n \rightarrow \infty} \sigma^2\{\hat{s}\} = 0.$$

Soll beurteilt werden, ob durch die Hinzunahme von neuen Daten die Güte der Schätzung weiter verbessert wird, muß ein Vergleich der erzielten Kovarianzen für den Zeitpunkt k und $k+1$ erfolgen. Eine Schätzung wird als *effektiv* (wirksam) bezeichnet, wenn die folgende Gleichung gilt:

$$\text{cov} \{ \hat{s}_{k+1} \} \leq \text{cov} \{ \hat{s}_k \}. \quad (5.24)$$

Mit den auf der Verteilung und den Momenten beruhenden Kriterien kann die Güte der Schätzung und ihr Verlauf hinreichend beschrieben werden.

Die vorgestellten Kriterien zur Beurteilung der Güte der Schätzung sind vom Standpunkt des Entwurfs und der Realisierung von Steuerungsaufgaben nur bedingt zu verwenden. Die Probleme bestehen darin, daß

1. ein System nicht unendlich lange beobachtet werden kann, bevor eine Entscheidung getroffen wird,
2. die Modelle zur Diagnose, Steuerung und Vorhersage überwiegend stark aggregierte Approximationen der Realität sind und deshalb keine „erwartungstreue Abbildung“ möglich ist,
3. die notwendige Modellgüte wesentlich durch das Steuerkonzept festgelegt wird (z. B. benötigten adaptive Steuerkonzepte keine Modelle einer erwartungstreuen Schätzung).

5.1.4. Möglichkeiten zur Fehlerbildung zwischen System- und Modellsignalen

Eine der wesentlichen Grundlagen für die Schätzung der Parameter eines Modells ist die Festlegung des Fehlersignals $e(k)$ zwischen System- und Modellsignalen. Folgende Möglichkeiten existieren:

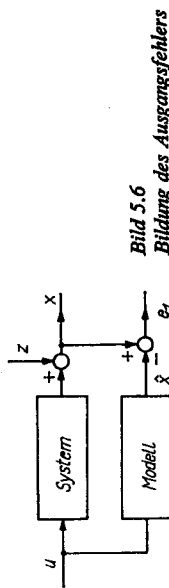


Bild 5.6

Bildung des Ausgangsfehlers

Ausgangsfehler (Bild 5.6). Der Ausgangsfehler $e_1(k)$ zum Zeitpunkt kT wird durch die Gleichung

$$e_1(k) = x(k) - \hat{x}(k) \quad (5.25)$$

festgelegt. Er wird angewendet bei der Schätzung von Parametern statischer Modelle (Abschnitte 5.2 und 5.3), bei der Bestimmung der Stützstellen der Volterra-Reihe (Abschn. 6.2) und bei der Zustandsschätzung (Abschn. 6.5). Bei dynamischen parametrischen Modellen (z. B. der Differenzengleichung) wird der Fehler $e_1(k)$ nichtlinear von den zu schätzenden Parametern abhängig. Wird als diskrete Übertragungsfunktion des Modells

$$\hat{G}(z^{-1}) = \hat{B}(z^{-1})/\hat{A}(z^{-1}) \quad (5.26)$$

mit

$$\hat{B}(z^{-1}) = [\hat{b}_0 + \hat{b}_1 z^{-1} + \dots + \hat{b}_m z^{-m}]$$

$$\hat{A}(z^{-1}) = [\hat{a}_0 + \hat{a}_1 z^{-1} + \dots + \hat{a}_n z^{-n}]$$

angenommen, erhält man für den Ausgangsfehler im Bildbereich

$$E_1(z) = X(z) - \frac{\hat{B}(z^{-1})}{\hat{A}(z^{-1})} U(z).$$

Diese nichtlineare Abhängigkeit erschwert die Schätzung der Parameter dieses Modelltyps. Neben einer iterativen Lösung der Schätz Aufgabe (s. Abschn. 5.3) wird der Weg der Bildung eines verallgemeinerten Fehlers beschrieben.

Eingangsfehler (Bild 5.7). Der Eingangsfehler $e_2(k)$ zum k -ten Zeitpunkt entsteht dadurch, daß das über ein inverses Systemmodell gebildete Eingangssignal $\hat{u}(k)$ vom realen Eingangssignal $u(k)$ entsprechend der Gleichung

$$e_2(k) = u(k) - \hat{u}(k) \quad (5.27)$$

subtrahiert wird. Dieser Fehler wird selten verwendet, weil das inverse Systemmodell in der Regel ein stark differenzierendes Verhalten aufweisen muß (das bedeutet Störungsverstärkung) und außerdem auch bei dynamischen parametrischen Modellen der Fehler nichtlinear mit den Parametern zusammenhängt.

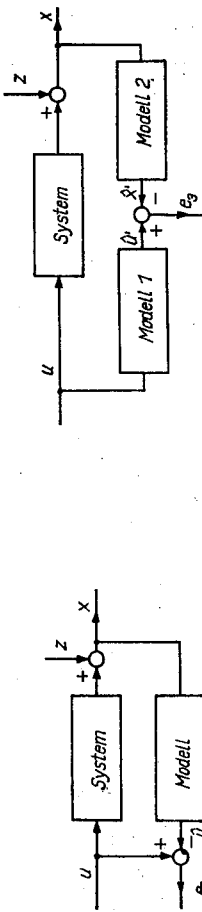


Bild 5.7. Bildung des Eingangsfehlers

Bild 5.8. Bildung des verallgemeinerten Fehlers

Verallgemeinerte Fehler (Bild 5.8). Aus der Bildungsvorschrift des Ausgangs- und Eingangsfehlers kann der verallgemeinerte Fehler abgeleitet werden, wenn das Systemmodell M in die Teilmodelle M_1 und M_2 aufgetrennt werden kann. Der verallgemeinerte Fehler $e_3(k)$ ergibt sich dann aus

$$e_3(k) = \hat{u}'(k) - \hat{x}'(k) \quad (5.28)$$

mit

$$\hat{u}'(k) = f(M_1, u(k)), \quad \hat{x}'(k) = f(M_2, x(k)).$$

Diese Fehlerart ist besonders bei der Schätzung von dynamischen parametrischen Modellen (Differenzengleichung) vorteilhaft, weil die Fehlergleichung linear von den zu schätzenden Parametern abhängt. Es gilt im Bildbereich für Gl. (5.28)

$$E_3(z) = \hat{B}(z^{-1}) U(z) - \hat{A}(z^{-1}) X(z). \quad (5.29)$$

Auf weitere Probleme der Anwendung dieser auch als Gleichungsfehler bezeichneten Fehlerart wird im Abschnitt 6.3 eingegangen.

Auf die Indizierung des Fehlers wird in den folgenden Abschnitten verzichtet, da überwiegend der Ausgangsfehler verwendet wird.

5.1.5. Grundstrategien zur Minimierung der Zielfunktion

Das Problem der Schätzung der Parameter auf der Grundlage der beobachteten Eingangs- und Ausgangsgrößen, der A-priori-Information und der gewählten Zielfunktion ist eine Optimierungsaufgabe. Es sind mit Hilfe eines Schätzverfahrens die Parameter des Modells so zu bestimmen, daß ein minimales Risiko oder eine maximale Wahrscheinlichkeit oder ein minimaler mittlerer quadratischer Fehler erzielt wird. Damit sind prinzipiell alle Methoden der mathematischen Optimierung anwendbar, die in [5.6 bis 5.19] ausführlich dargestellt sind. Im Rahmen dieses Buches wird nur auf Optimierungsmethoden eingegangen, die für die Lösung der Schätz Aufgabe verwendet werden.

Generell vereinfacht sich die Optimierungsaufgabe durch die quadratische Form der Zielfunktion beim Verfahren der Regression und der verallgemeinerten Regression. Die optimale Lösung ist bei linearen Fehlergleichungen ein analytischer Ausdruck in Form der Schätzvorschrift. Die Aufgabe besteht dann in der Anwendung einer günstigen numerischen Lösungsvariante für die Schätzvorschrift (s. Abschn. 5.1.5.1). Besteht zwischen dem Fehler und den zu schätzenden Parametern ein nichtlinearer Zusammenhang, müssen die optimalen Parameter durch eine iterative Lösung mittels Suchverfahren bestimmt werden. Dies ist auch erforderlich, wenn die Lösung der Zielfunktion unter der Beachtung von Gleichungs- oder Ungleichungsbeschränkungen ermittelt werden muß (s. Abschnitt 5.1.5.2).

Die bis jetzt genannten Wege zur Ermittlung der Parameter gehen davon aus, daß aus einem Datenblock mit n Beobachtungen die Lösung direkt oder durch eine Iteration gefunden wird. In vielen Fällen stehen aber laufend neue Informationen zur Verfügung, die zur weiteren Verbesserung des Modells benutzt werden können. Die Schätzverfahren, die in der Lage sind, die Zielfunktion durch Verarbeitung neuer Informationen über die Eingangs- und Ausgangsgrößen zu minimieren, werden als rekursive Verfahren bezeichnet (s. Abschn. 5.1.5.3).

5.1.5.1. Direkte Lösung der Schätz Aufgabe

Die direkte Lösung der Schätz Aufgabe erfolgt anhand der Gleichung

$$\frac{\partial Q}{\partial \beta} = 0 \quad (5.30)$$

in einem Schritt aufgrund von n Beobachtungen. Das allgemeine Schema dieser Verfahren ist im Bild 5.9 dargestellt. Für die Methode der Regression ergibt sich aus Gl. (5.30) so die direkte analytische Lösung zu (s. Abschn. 5.2.2)

$$\beta = [M^T M]^{-1} M^T x. \quad (5.31)$$

Als numerische Lösungsverfahren der Gl. (5.31) haben sich u. a. die in Tafel 5.2 genannten bewährt.

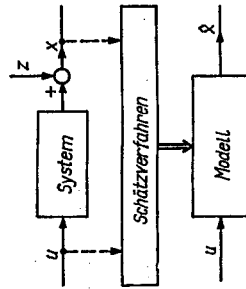


Bild 5.9
Struktur der direkten und der iterativen Schätzmethoden

Tafel 5.2. Zusammenstellung von Lösungen der direkten Schätzmethoden

Numerisches Verfahren	Lösungsgleichung für β	Name des Verfahrens
Matrizeninversion – direkt – Zerlegung in Dreiecksmatrizen	$\beta = [M^T M]^{-1} M^T x$	– Gauß – Gauß-Banachiewicz [5.1, 5.2, 5.3] – Cholesky-Zerlegung
Auflösung linearer Gleichungssysteme	$M^T M \beta = M^T x$	– Gauß-Jordan [5.1, 5.2] – Cholesky-Zerlegung [5.4, 5.5, 5.6]
Transformationen – orthonormal – orthogonal	$M^T M = I$ $M^T M = \begin{bmatrix} m_1^* m_1^* & \dots & 0 \\ 0 & \dots & m_n^* m_n^* \end{bmatrix}$ m_i^* i-te Spalte der orthogonalisierten Meßwertmatrix M^*	– Householder [5.7]

5.1.5.2. Iterative Lösung der Schätz Aufgabe

Eine iterative Lösung der Schätz Aufgabe auf der Grundlage der n Beobachtungen wird erforderlich, wenn die zu schätzenden Parameter nichtlinear in die Fehlergleichung eingehen oder wenn Beschränkungen in der Zielfunktion berücksichtigt werden müssen.

Tafel 5.3. Zusammenstellung von iterativen Schätzverfahren

Verfahrensgruppe	Lösungsgleichung für β	Name des Verfahrens/Literatur
Gradientenfreie Verfahren	$\beta^{i+1} = \beta^i + \Delta \beta^{i+1}$ $\Delta \beta^{i+1} = \alpha(i) \Delta \beta$ $\alpha(i) = \begin{cases} > 0 \\ = 0 \\ < 0 \end{cases}$ für $Q(i) = \begin{cases} < \\ = \\ > \end{cases} Q(i-1)$ $\Delta \beta$ vorgegebener Vektor	Powell [5.8, 5.9] Rosenbrock [5.10] Fibonacci [5.11]
Gradientenverfahren	$\Delta \beta^{i+1} = -\gamma(i+1) R(i+1) \nabla Q(i)$ $\gamma(i)$ Schrittweitenvektor $R(i)$ Richtungsmatrix $\nabla Q(i)$ Gradient zum i-ten Schritt	steilster Abstieg [5.9, 5.11] Newton [5.9, 5.12] Newton-Raphson [5.9, 5.11] Marquardt-Levenberg [5.13, 5.14, 5.15] Powell [5.9, 5.11]
Zufallssuchverfahren	$\Delta \beta^{i+1} = \beta(i+1) R(i+1) \Delta \beta$ mit $\beta(i+1) = \begin{cases} 0 & \text{für } Q(i+1) \geq Q(i) \\ 1 & \text{für } Q(i+1) < Q(i) \end{cases}$ $\Delta \beta$ Zufallsvektor	Brooks [5.16] Rastrigin [5.17] White [5.18] Schwefel [5.19]