

Online-Demodulation stark gestörter winkelmodulierter Signale mit dem Extended Kalman-Filter

Vom Fachbereich Elektrotechnik und Informatik
der Universität-Gesamthochschule Siegen
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften

(Dr.-Ing.)

genehmigte Dissertation

von

Diplom-Ingenieur Dirk Balzer

1. Gutachter: Prof. Dr.-Ing. habil. Otmar Loffeld
2. Gutachter: Prof. Dr.-Ing. Werner Düchting

Tag der mündlichen Prüfung:

14.Mai 1999

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Nachrichtenverarbeitung und Projektbereich 2 des Zentrums für Sensorysysteme (ZESS) der Universität-Gesamthochschule Siegen.

Herrn Prof. Dr.-Ing. habil. O. Loffeld gilt mein herzlicher Dank für die Anregung zu dieser Arbeit und für die Unterstützung bei ihrer Durchführung.

Herrn Prof. Dr.-Ing. W. Düchting möchte ich für die Übernahme des Korreferats als 2. Gutachter danken.

Mein Dank gilt weiterhin Herrn Prof. Dr.-Ing. R. Schwarte für die Übernahme des Vorsitzes der Prüfungskommission.

Ferner gilt mein herzlicher Dank auch der Deutschen Telekom, die mich zu Beginn meiner Promotion freundlicherweise finanziell förderte.

Meinen Kolleginnen und Kollegen Dr.-Ing. Chr. Arndt, Dr.-Ing. A. Hein, Dipl.-Ing. S. Knedlik, Dipl.-Ing. S. Oerder, Dipl.-Ing. D. Justen, Dipl.-Ing. A. Stadermann, Dipl.-Ing. W. Twelsiek und Dipl.-Ing. R. Wurmbach bin ich für fachliche Unterstützung verbunden.

Ein herzliches "Danke schön" geht auch an Dipl.-Ing. S. Gleue, Dipl.-Ing. C. Mähringer, Dipl.-Ing. O. Klauschke, Dipl.-Ing. N. Schweitzer, Dipl.-Ing. E. Hüttenschmidt und Dipl.-Ing. G. Schaffer, die mit ihren Studien- und Diplomarbeiten einen wesentlichen Beitrag zu dieser Arbeit geleistet haben.

Mein ganz besonderer Dank gilt selbstverständlich auch meiner Lebensgefährtin M. Heinrich und meiner Familie, die mir ein stabiles Rückgrat gegeben haben.

Inhaltsverzeichnis:

ABSTRACT	5
SUMMARY	6
1 EINLEITUNG	7
1.1 Kalman-Filter und Nachrichtenübertragung	7
1.2 Online und Offline - Filterung	9
2 THEORETISCHE GRUNDLAGEN	12
2.1 Signaltheoretische Grundlagen	12
2.1.1 Modulationsverfahren.....	12
2.1.1.1 Bandpaßsignale und äquivalente Tiefpaßsysteme.....	12
2.1.1.2 Kontinuierliche Modulationen/ Demodulationen.....	15
2.1.1.2.1 Winkelmodulation.....	16
2.1.1.3 Diskrete Modulationsverfahren.....	18
2.1.2 Demodulationsverfahren.....	18
2.1.2.1 Demodulation mit dem Quadraturempfänger.....	18
2.1.2.2 Demodulation mit PLL.....	22
2.1.3 Systemeigenschaften einer WM-Übertragungsstrecke.....	25
2.1.4 SNR-Definition.....	33
2.1.5 Kanalkapazität	34
2.2 Stochastische Prozesse	36
2.2.1 Gauß-Markov-Prozeß	40
2.2.2 Weißes, gaußverteiltes Rauschen	42
2.2.2.1 Diskretes weißes Rauschen	43
2.2.3 Farbige Rauschprozesse.....	44
2.3 Bandbreitebetrachtungen.....	44
2.3.1 Bandbreitebeziehungen bei Winkelmodulation.....	46
2.3.2 Bandbreiteabschätzung zeitdiskreter Signale	51
2.4 Schätzverfahren.....	53
2.4.1 Schätzwertbestimmung über bedingte Verteilungsdichten	54
2.4.1.1 Informationsfluß im Übertragungskanal	56
2.4.2 Stationäre Filter	58
2.4.2.1 Minimum - Varianz - Methode (Wiener).....	58
2.4.2.2 Maximum Likelihood Estimation	59
2.4.3 Dynamische Filter.....	61
2.4.3.1 Kalman - Filter	62
2.4.3.1.1 Kontinuierliche und diskrete Betrachtungsweise.....	62
2.4.3.2 Nichtlineare Filter (EKF)	64
2.4.3.3 Linearisierung von Zustandsübergangsmatrix und Beobachtungsmatrix	69
2.4.3.3.1 Linearisierung bei nichtlinearer Zustandsübergangsmatrix	70
2.4.3.3.2 Linearisierung bei nichtlinearer Beobachtungsmatrix	71
2.4.3.3.2.1 Betrachtung des Linearisierungsfehlers:	74
2.4.3.4 Filter höherer Ordnung.....	75
2.4.4 Analogien von Extended Kalman-Filter und PLL (Phase Locked Loop):	76
2.5 Informationstheorie	82
2.5.1 Definitionen	82
2.5.2 Informationsgrenze - Cramér-Rao-Grenze	83
2.5.3 Informationsbegriff bei der Demodulation winkelmulierter Signale	86

2.6	Stabilitätsbetrachtung der Kalman-Filter	87
2.6.1	Beobachtbarkeit	87
2.6.2	Steuerbarkeit	89
3	KALMAN-FILTER MODELLIERUNG ZUR DEMODULATION WINKELMODULIRTER SIGNALE	90
3.1	Extended Kalman-Filter zur Phasendemodulation	90
3.1.1	Standardfilter - EKF22 (2 Beobachtungen)	90
3.1.1.1	Beobachtungsmodell	90
3.1.1.2	Zustandsraummodell	92
3.1.1.3	Stabilität	96
3.1.2	Zeitkontinuierliche Modellentwicklung - Singermodell	99
3.1.3	Übergang zur zeitdiskreten Modellbildung	100
3.2	Erweiterung des Beobachtungsvektors	100
3.2.1	Erweiterung des Beobachtungsvektors auf 3 und 4 Beobachtungen	103
3.2.1.1	Zusätzliche Beobachtung: Phase (EKF32pv)	103
3.2.1.2	Zusätzliche Beobachtung: Amplitude	108
3.2.2	Erweiterung des Beobachtungsvektors auf 4 Beobachtungen (EKF44pav)	110
3.3	Adaptive Kalman-Filter	112
3.3.1	Adaptive Anpassung von Parametern im Kalman-Filtermodell	113
3.3.2	Adaptive Anpassung von Parametern durch Maximum Likelihood Schätzung	114
3.4	Smoother	119
3.4.1	Fixed-Intervall Smoothing	120
3.4.1.1	Fraser-Potter Smoother	121
3.4.1.2	Rauch-Tung-Striebel-Smoother oder Meditch-Filter	122
3.4.2	Fixed-Point Smoother	123
3.4.3	Fixed-Lag Smoother	123
3.4.3.1	Fixed-Lag Smoother durch Zustandsraumerweiterung	124
3.4.3.2	Switched-Fixed-Lag Smoother	125
3.4.4	Smoothability:	125
3.4.5	Smoothermodell für die Online-Anwendung	125
3.4.5.1	Fixed-Lag Smoother (Reduced Order Filter: SKF22ro)	127
3.5	Weitere Filteroptimierungen	128
3.5.1	Prefiltering	128
3.5.2	Temporal Update	129
3.5.3	EKF zur Verarbeitung digitaler Modulationsverfahren	129
4	ERGEBNISBETRACHTUNGEN DER SIMULATIONEN UND ONLINE - REALISIERUNG	131
4.1	Testfunktionen und Testsignale	131
4.1.1	Pseudo-Sprachsignal	132
4.1.1.1	Sprachmodell	132
4.1.2	Rauschprozesse	134
4.1.2.1	Diskrete Rauschgeneratoren	134
4.1.2.1.1	Definition unter der Programmiersprache 'C'	134
4.1.2.1.2	Definition unter der Interpretersprache IDL	137
4.1.2.1.3	Weitere Zufallsgeneratoren	137
4.1.2.2	Rauschen der Demodulationsumgebung	137
	-Anpassungstest	138
4.2	Monte-Carlo-Simulation unter IDL und Borland-C	139
4.3	Filtersimulation	140
4.3.1	Simulation unter BorlandC	141
4.3.2	Monte-Carlo-Simulation unter IDL	142
4.4	Auswertung der Online-Filtermodelle	144

4.4.1	Filtertuning	146
4.4.2	Stationäres Verhalten.....	148
4.4.2.1	Startwerte.....	148
4.4.2.2	Schätzfehlervarianz	149
4.4.3	Dynamisches Verhalten	153
4.4.3.1	Stabilität	155
4.4.3.2	Einschwingverhalten	155
4.4.3.3	Modulationsindex.....	157
4.4.4	Verwendung unterschiedlicher Quellsignale.....	157
4.5	Der Meßaufbau.....	159
4.5.1	Hardware.....	159
4.5.1.1	Kalibrierung des Systems.....	161
4.5.2	Software.....	162
4.5.3	Organisation und Design des EKF-Programms.....	162
4.5.4	Optimierung des Kalman-Filters (Zeit=Code)	163
4.5.4.1	Betrachtung der Bandbreite/Laufzeit	164
4.6	Fehlerbetrachtung bei der Online-Demodulation.....	165
4.6.1.1	Soundkarte.....	166
4.6.1.2	Fehler der AD-Wandlerkarte.....	166
4.6.1.3	Quantisiertes Rauschen des Systems.....	166
4.6.1.4	Rundungsfehler	167
4.6.1.5	Modellfehler	167
5	DIE ENTWICKLUNGSUMGEBUNG	168
5.1	Konzept der Übertragungsstrecke	168
5.2	Realisierung der Hardware.....	170
5.3	Software der Entwicklungsumgebung	172
5.3.1	Simulator- Entwicklungsumgebung	172
5.3.1.1	BorlandC-Simulator	172
5.3.1.2	Widget-Simulator	172
5.3.2	Online-Entwicklungsumgebung	173
5.3.2.1	Filtertuning, Filterauswahl	173
5.3.3	Signalgeneration	173
5.3.3.1	Signalerstellung, Modulation und Bandbreitenbetrachtungen unter IDL	174
5.3.3.2	Abspielen der Signale mit Soundkarte aus TXT-File	174
6	ZUSAMMENFASSUNG	175
7	LITERATURVERZEICHNIS.....	176
8	ANHANG	183
8.1	Verzeichnis häufig verwendeter Abkürzungen und Formelzeichen	183
8.2	Sammlung der Filtermodelle.....	186

Abstract

Kalman filters offer a good opportunity for processing heavily disturbed signals. Applied to the demodulation problem of communications an Extended Kalman filter is used as a demodulator of angularly modulated signals. In the range of low signal to noise ratios (0 to 17 dB) a number of several filter models is examined in this work /3, 5, 16, 24, 34, 35, 40, 42, 49, 54, 69, 77/.

The two quadrature components, real and imaginary part of a received signal in the baseband are considered as components of a two dimensional observation vector from which phase, frequency and even amplitude and amplitude derivative are estimated.

Supplementary observations as amplitude and phase, adaptive elements and smoother elements have been tested in an online development environment. It has been built up with a DSP (TMS320 C44) to prove data fusion concepts and information absorption in Kalman filters based on examinations which have been theoretical determined in /5, 42/. It proves that even virtually created additional observations like phase and amplitude determined with the two quadrature components lead to an increase of the absorbed information, which leads to better filter results.

Furtheron stability, accuracy and bandwidth have been observed in the online simulation. They are important criteria for online processing, whereas in case batch processing only the accuracy is the important parameter.

Chapter 3 will be laying the basics in signal theory for the following chapter 4, where the different Kalman filter models developed for the demodulation problem will be described. Chapter 5 will show filtering results and Chapter 6 gives a rough description of the DSP development environment used for online-demodulation /2, 33, 76/. A summary in Chapter 7 concludes this work with results and remarks.

Summary

This work states that Extended Kalman filters can be used for online demodulation of heavily disturbed angularly modulated signals in general. Depending on the used SNR (signal to noise ratio = carrier to noise ratio) and the choice of the modulation signals different Extended Kalman filter models have to be employed. In the investigated range of 0 to 17 dB SNR especially adaptive Kalman filter models show a high accuracy. In this case an external Maximum-Likelihood estimation is preferred allowing a reduction of the filter running time. An additional advantage using external adaptation is the implicit stability of the outcoming adaptive filter algorithm in contrast to parameter estimation by introducing additional state space variables. Modulating signals with low dynamics Kalman filters with precomputed covariances could be applied, where in the stationary case the covariance cycle could be eliminated.

Extended Kalman filter models show especially for negative SNR's the advantages of data fusion, however the filter behaviour is not very stable. The concept of Smoother algorithms is rather unsuitable for online applications, due to the expense in processing to increase filter stability and accuracy. An exception is a heavily reduced smoother model, which retains the filter stability characteristics of smoothers. An employment of this filter is only worth while for negative SNR's, because of high error covariance values.

Using nonlinear filters optimized starting values have to be chosen, because this does not only effect the transient filtering phase, but also the stationary accuracy too. Therefore an optimal transient time can be found by parameter tuning, with the result of an optimized accuracy. Only for further reduction of the transient time the filter error increases.

In parallel to the theoretical investigations a development and evaluation system was elaborated representing a useful development tool not only for online demodulation, but for several Kalman filter applications, like motor management or machine control too.

1 Einleitung

Ein generelles elektrotechnisches Problem stellt die Signalverarbeitung von gestörten Meßsignalen dar. Werden besonders stark gestörte Signale, die einen Störabstand von bis zu 0 dB aufweisen, betrachtet, lassen sich mit einfachen Filtern (Tiefpaßfilterung) keine sinnvollen Filterergebnisse mehr erzielen. Es werden Filter benötigt, die zusätzlich die stochastischen Eigenschaften der Signale und Störungen mit in Betracht ziehen. Mit Kalman-Filtern können diese stochastischen Eigenschaften beispielsweise ausgewertet werden. Beim Kalman-Filter haben wir darüberhinaus den Vorteil, daß sich durch das rekursive Verhalten des Kalman-Filter Algorithmus alle vergangenen, schon bearbeiteten Meßwerte mit in die aktuelle Schätzung einbeziehen lassen.

Weitere Vorteile, die das stochastische Konzept eines Kalman-Filters unterstützen, sind die Tatsachen, daß es in der realen Applikation weder perfekte Meßsensoren noch eine perfekte mathematische Modellierung des gesuchten Systems gibt. Der Kalman-Filter arbeitet genau mit diesen Ungewißheiten, nämlich einer allgemeinen Systembeschreibung, der statistischen Beschreibung der Störgrößen, der Modellgenauigkeit und den Anfangsbedingungen, die wichtig sind für ein schnelles Einschwingen auf minimale Fehler. Bei genauer Kenntnis der gegebenen Gesetzmäßigkeiten und linearen Zusammenhänge erhält man mit einem Kalman-Filter tatsächlich ein Optimalfilter, das heißt, die vorhandene Information, die in den Meßgrößen enthalten ist, wird vollständig ausgewertet und in die Schätzung der gesuchten Regelgröße umgesetzt /38, 39, 48, 49/.

1.1 KALMAN-FILTER UND NACHRICHTENÜBERTRAGUNG

Besondere Aufmerksamkeit gilt in der Nachrichtentechnik der fehlerfreien Übertragung von Daten. Dieser sogenannte Übertragungskanal weist dabei klar definierte Eigenschaften auf. Nicht die gesamte Information kann von A nach B übertragen werden. Das Signal schwächt sich ab und additive Störgrößen beeinflussen die Übertragung. Der beschriebene Nachrichtenfluß kann z.B. als Entropieverteilung im Kanal beschrieben werden.

Durch eine Modulation des Nutzsignals lassen sich Daten wesentlich einfacher übertragen, da die Eigenschaften des Übertragungskanals optimaler ausgenutzt werden. Eine Vielzahl

von Modulationsarten hat sich im Laufe der Jahre entwickelt, die jeweils an die geforderten Kanaleigenschaften angepaßt worden sind.

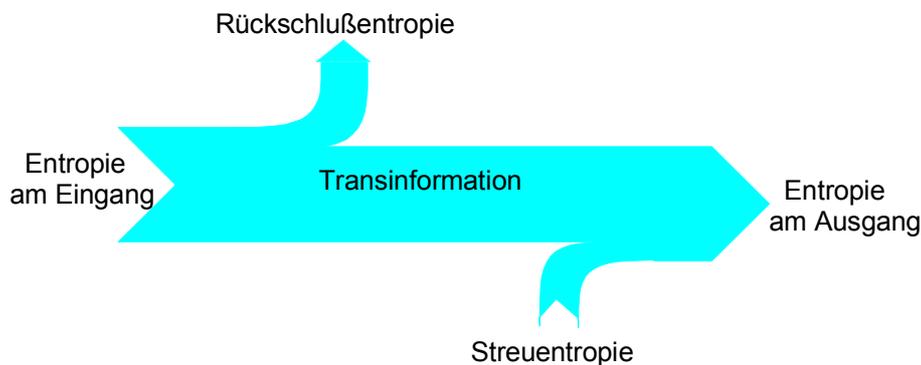


Abb. 1.1: Entropieverteilung am Übertragungskanal

Erwünscht ist es, möglichst viel Information eines Nutzsymbols über einen Nachrichtenkanal zu übertragen, der mit minimaler Bandbreite und Leistung auskommt. Durch eine fehlerfreie Demodulation bei geringen SNR's lassen sich Bandbreite und/oder Leistung eines modulierten Signals absenken. Dies bedeutet auch, daß bei einem ständig anwachsenden Bedarf an Übertragungskapazitäten die Sendeleistungen nicht weiter erhöht werden müssen, um mehr Information zu übertragen. Es begann mit analogen Modulationsarten (AM, PM, FM und diverse Mischformen) und man ist heutzutage bei einer Vielzahl von digitalen Modulationsarten angelangt (PSK, FSK, MSK,...). Ziel dieser Arbeit ist es deshalb, mit Kalman-Filtern Signale von sehr hohem Störanteil zu verarbeiten [18, 27, 29, 43, 45, 60, 62, 75].

Beschränkt man sich auf winkelmodulierte Signale, wird bei der hier betrachteten Quadraturdemodulation das Empfangssignal ins Basisband heruntergemischt und die Quadraturkomponenten $\text{Re}\{r_T(t)\}$, $\text{Im}\{r_T(t)\}$ zur Bestimmung der Phase verwendet, die das Nutzsymbols darstellt. Die Berechnung der Phase aus den beiden Quadraturkomponenten kann einfach über die Arcustangens-Funktion geschehen.

Bei ungestörten Quadraturkomponenten führt dies zu einer korrekten Phase. Bei gestörten Größen führt dies jedoch zu großen Fehlern bei der Berechnung der Phase.

Genau an dieser Stelle setzt das Kalman-Filter an, um die Phase dennoch möglichst genau zu bestimmen. Als Eingangsparameter für das Kalman-Filter werden die beiden Quadraturkomponenten verwendet, um eine Estimation der Phase durchzuführen. Das ausgewählte Kalman-Filter ist in dieser Anwendung ein nichtlineares Filter.

Bei der nichtlinearen Realisierung des Filters werden jedoch -bedingt durch den Algorithmus- Linearisierungen benötigt, die Informationsverlust zur Folge haben. Als lineares Filter eingesetzt, stellt das Kalman-Filter ein Optimalfilter dar. Wie läßt sich die Anwendung in diesem nichtlinearen Fall rechtfertigen? Grundsätzlich werden Linearisierungen vorgenommen, um ein System besser kontrollierbar zu machen. Ferner ist die lineare Systemtheorie bestens bekannt und anwendbar. So ergibt sich hieraus die Folgerung, daß eine Verarbeitung mit einem Kalman-Filter durchaus sinnvoll erscheint, da die Information der beiden Quadraturkomponenten so zusätzlich stochastisch ausgewertet wird. Die Anpassung des Extended Kalman-Filter an die nichtlineare Problemstellung ist ebenfalls Aufgabe dieser Arbeit. Beim Kalman-Filter findet eine sogenannte Datenfusion der Eingangskomponenten statt. Die einzelnen Komponenten führen nicht direkt zu einzelnen Schätzwerten, sondern werden kreuzweise miteinander verknüpft und zu einem einzigen Schätzwert verarbeitet. Dies führt zum einen dazu, daß auch nicht meßbare Größen geschätzt werden können, und zum anderen dazu, daß bei einem multisensoriellen System der Ausfall oder die extreme Störung einer Meßgröße durch andere Meßgrößen kompensiert werden kann.

Von entscheidender Bedeutung ist es, die Kanalkapazität zu optimieren. Entweder wird versucht die Packungsdichte der Datenkanäle auf einem Träger zu maximieren oder es wird versucht die abgestrahlte Leistung zu minimieren. Oder anders formuliert: Bei einem ständig ansteigenden Bedarf an Übertragungskapazitäten wird angestrebt die Sendeleistungen nicht weiter zu erhöhen, um zusätzliche Informationen zu übertragen.

Aus beiden Fällen resultiert die Forderung, immer ungünstigere Signal-Rauschleistungsverhältnisse zuverlässig bearbeiten zu können, wofür Kalman-Filter im allgemeinen prädestiniert sind. Mit einer Vielzahl von Kalman-Filtermodellen wird dies in dieser Arbeit untersucht /3, 5, 6, 7, 8, 16, 24, 34, 35, 40, 42, 49, 54, 69, 77/.

1.2 ONLINE UND OFFLINE - FILTERUNG

Nachdem als erster Schritt das Kalman-Filtermodell zu der Problemstellung entwickelt wurde, folgt die Anpassung des Filters auf den Einsatzort des Filters. Anforderungen im Hinblick auf Sensitivität, Stabilität, Rechnerleistung, Kosten und weiteren Faktoren muß Rechnung getragen werden. Handelt es sich um eine rein statische Anwendung, wird ein vorliegender Datensatz relativ unabhängig von Zeit und Datenmenge bearbeitet. Bei Instabi-

litäten besteht die Möglichkeit zusätzliche Filterungen durchzuführen. Entscheidend ist dann einzig und alleine die optimale Verarbeitung der vorgegebenen Meßdaten.

Bei einem kontinuierlichen Fluß von Daten muß dagegen zusätzlich auf die Prozessierungsdauer und die Stabilität des Filteralgorithmus geachtet werden. Die Prozessierungsdauer geht dabei direkt in die verarbeitbare Bandbreite des Signals ein. Das Kalman-Filter ist ein rechenintensives Filter, das matrizenorientiert arbeitet. Zusätzliche Eingangs- oder / und Ausgangsparameter erhöhen die Anzahl der Rechenoperationen kräftig. Daher wird zur Berechnung ein Digitaler Signal Prozessor (DSP) eingesetzt, der hardwaremäßig bereits Multiplikationen, Divisionen und begrenzt sogar Wurzelfunktionen direkt als einen Befehl ausführen kann. In den verwendeten Entwicklungsumgebungen wurden dabei Prozessoren aus der TMS320 C4x - Familie von Texas Instruments verwendet [2, 33, 76].

Vor dem Einsatz eines Filters in der Online-Anwendung wird jedes Filter auf seine Eigenschaften in Softwaresimulationen getestet. Vorteile einer vorhergehenden Simulation sind die klar definierten Umgebungsparameter und die Möglichkeit erster Nachbesserungen des Filters in der Testphase. Im Online-Einsatz zeigen sich dagegen abweichende Nichtlinearitäten und Störprozesse, die sich erst durch bereits ermitteltes Systemverhalten optimal verarbeiten lassen. Ein erneutes Nachbearbeiten führt schließlich zum endgültigen Filtermodell und zu den gesuchten Schätzwertergebnissen.

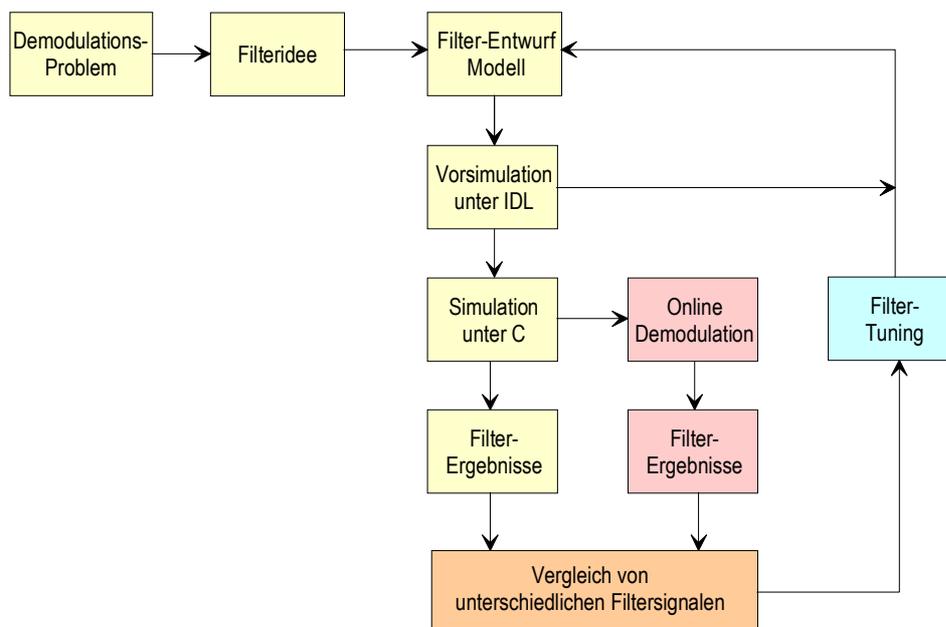


Abb. 1.2: Filterentwurf für Online-Filter

Die Arbeit ist dabei wie folgt aufgebaut: Kapitel 2 beschreibt signaltheoretische Grundlagen, die in Kapitel 3 zur Ableitung der unterschiedlichen Filtermodelle genutzt werden. Im nachfolgenden Kapitel 4 werden Ergebnisse der unterschiedlichen Simulationen aufgezeigt und im letzten Kapitel die verwendete Hardware beschrieben. Diese Arbeit schließt mit einer Zusammenfassung der Ergebnisse und gefundenen Beobachtungen.

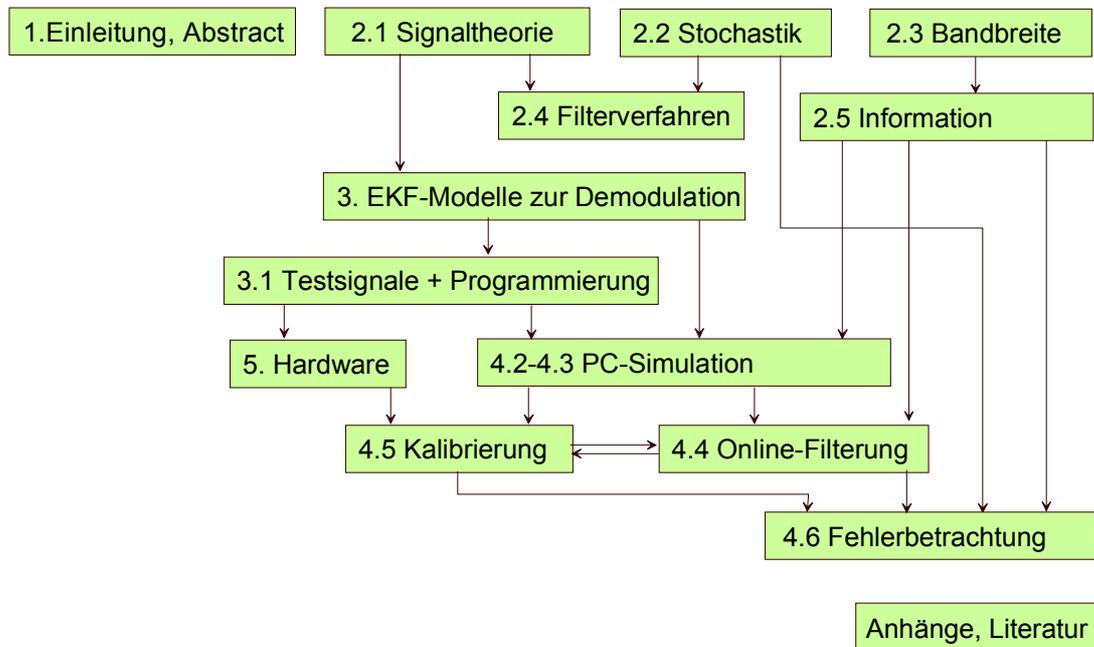


Abb. 1.3: Kapitelübersicht als Struktogramm

2 Theoretische Grundlagen

Zunächst sollen grundlegende Begriffe der Systemtheorie geklärt werden, um mit den gefundenen Beziehungen in den folgenden Abschnitten und Kapiteln leichter arbeiten zu können. Dabei wird sich auf die Grundlagen mit Relevanz aus den Bereichen der Modulationstechnik zeitkontinuierlicher und zeitdiskreter Signale beschränkt. Darüber hinaus werden fundamentale Begriffe der Estimationstheorie inklusive von Rauschprozessen aufgearbeitet. Wichtig ist diese Aufarbeitung auch zur eindeutigen Definition der mathematischen Formelzeichen und technischen Begrifflichkeiten.

2.1 SIGNALTHEORETISCHE GRUNDLAGEN

2.1.1 Modulationsverfahren

Bei der Modulation von Signalen werden diese auf einen Träger aufmoduliert und auf diese Weise über einen Übertragungskanal übertragen. Nach der Übertragung wird das Empfangssignal wieder demoduliert, also von seinem Träger befreit /44, 45, 62, 75/.

2.1.1.1 Bandpaßsignale und äquivalente Tiefpaßsysteme

Bei Übertragungssystemen betrachten wir normalerweise ein bandpaßbegrenztes System. Das Ausgangssignal eines derartigen Systems wird als Bandpaßsignal bezeichnet. Zur vereinfachten Betrachtung werden meistens Bandpaßsignale im äquivalenten Tiefpaßbereich untersucht.

Eine beliebige, reelle mathematische Funktion $s_B(t)$, hier ein Bandpaßsignal, läßt sich aus Sinus- und Cosinus-Schwingungen unterschiedlicher Frequenzen zusammensetzen (Fourier), so daß wir das Bandpaßsignal in gerade und ungerade Anteile aufspalten können:

$$s_B(t) = s_{Bg}(t) + s_{Bu}(t) \tag{2-1}$$

Ferner gilt:

$$\begin{aligned}
 S_B(f) &= \int_{-\infty}^{+\infty} [s_{Bg}(t) + s_{Bu}(t)] \cdot [\cos(2\pi f_0 t) - j \sin(2\pi f_0 t)] dt \\
 &= \operatorname{Re}\{S_B(f)\} + j \cdot \operatorname{Im}\{S_B(f)\}
 \end{aligned} \tag{2-2}$$

Betrachten wir Real- und Imaginärteil von $S_B(f)$ im sog. "äquivalenten Tiefpaßbereich": Dazu verwenden wir ausschließlich die positiven Frequenzanteile von $S_B(f)$, multiplizieren diese mit dem Faktor 2 und verschieben sie anschließend um eine Trägerfrequenz f_0 nach links in Richtung Nullpunkt.

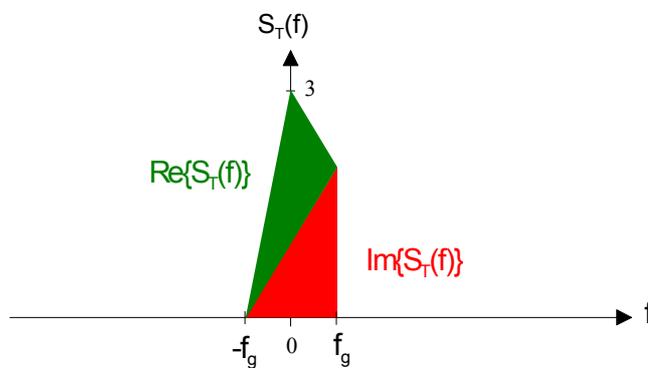


Abb. 2.1: Äquivalenter Tiefpaßbereich

Das erhaltene Spektrum $S_T(f)$ enthält keine Trägerfrequenz f_0 mehr. Die Bandpaßsignal-komponenten mit Hilfe des äquivalenten Tiefpaßbereiches lauten:

$$\begin{aligned}
 \operatorname{Re}\{S_B(f)\} &= \frac{1}{2} \operatorname{Re}\{S_T(f - f_0)\} + \frac{1}{2} \operatorname{Re}\{S_T(-f - f_0)\} \\
 \operatorname{Im}\{S_B(f)\} &= \frac{1}{2} \operatorname{Im}\{S_T(f - f_0)\} - \frac{1}{2} \operatorname{Im}\{S_T(-f - f_0)\}
 \end{aligned} \tag{2-3}$$

Worin liegt der Vorteil der Einführung eines äquivalenten Tiefpaßbereiches? Durch Wegfall der Trägerfrequenz läßt sich das System mathematisch erheblich einfacher beschreiben. Da die Trägerfrequenz keine Nutzinformation enthält, ist durch diese vereinfachte Betrachtungsweise kein Verlust der Signalinformation bei der Betrachtung dieser Signale entstanden. Das Nutzsignal oder Modulationssignal wird vollständig weiter bearbeitet.

Als komplexes Spektrum zusammengefaßt ergibt sich:

$$S_B(f) = \frac{1}{2} S_T(f - f_0) + \frac{1}{2} S_T^*(-f - f_0), \quad (2-4)$$

wobei $S_T^*(f)$ den konjugiert komplexen Anteil darstellt. Durch Rücktransformation in den Zeitbereich erhalten wir dann folgende komplexe Beschreibung:

$$s_B(t) = \frac{1}{2} s_T(t) \cdot e^{j2\pi f_0 t} + \frac{1}{2} [s_T(t) \cdot e^{-j2\pi f_0 t}]^*$$

Für komplexe Zahlen gilt allgemein: $z + z^* = \text{Re}\{2z\}$, was das Signal wieder zu einem reellwertigen Signal werden läßt:

$$s_B(t) = \text{Re}\{s_T(t) \cdot e^{j2\pi f_0 t}\} \quad (2-5)$$

Bislang haben wir das komplexe modulierte Signal $s(t)$ betrachtet, so daß sich insgesamt folgende Beziehung ergibt:

$$\begin{aligned} s(t) &= A \cdot e^{j(2\pi f_0 t + \varphi(t) + \varphi_0)} \\ &= A \cdot e^{j2\pi f_0 t} \cdot e^{j(\varphi(t) + \varphi_0)} \\ &= A \cdot e^{j(\varphi(t) + \varphi_0)} \cdot e^{j2\pi f_0 t} \\ &= s_T(t) \cdot e^{j2\pi f_0 t} \end{aligned}$$

Das Signal $s_T(t)$ wird auch als komplexe Hüllkurve bezeichnet; der Term $e^{j2\pi f_0 t}$ wird analog dazu als komplexer Träger bezeichnet.

$$\begin{aligned} s_B(t) &= s_{T\text{Re}}(t) \cdot \cos(2\pi f_0 t) - s_{T\text{Im}}(t) \cdot \sin(2\pi f_0 t) \\ \text{mit: } s_T(t) &= s_{T\text{Re}}(t) + js_{T\text{Im}}(t) \end{aligned} \quad (2-6)$$

Real- und Imaginärteil der komplexen Hüllkurve werden auch als Quadraturkomponenten¹ bezeichnet.

¹Der Begriff Quadraturkomponente wird auch für den Imaginärteil und Inphasekomponente für den Realteil von $s_T(t)$ verwendet.

- Die Amplitude berechnet sich zu: $A(s_T(t)) = |s_T(t)| = \sqrt{s_{T\text{Re}}^2(t) + s_{T\text{Im}}^2(t)}$
- Die Phase berechnet sich zu: $\varphi(t) = \arctan\left[\frac{s_{T\text{Im}}(t)}{s_{T\text{Re}}(t)}\right]$

Dies läßt sich wieder zur gewohnten komplexen Darstellung zusammenziehen:

$$s_T(t) = |s_T(t)| \cdot e^{j\varphi(t)} = |s_T(t)| \cdot e^{j\varphi(t) + j\varphi_0} \quad |^2 \quad (2-7)$$

2.1.1.2 Kontinuierliche Modulationen/ Demodulationen

Wir unterscheiden zwei grundlegend unterschiedliche Modulationsarten: Die Amplitudenmodulation und die Winkelmodulation. Während bei der Amplitudenmodulation die Amplitude der Trägerschwingung moduliert wird, wird dagegen bei der Winkelmodulation die Momentanphase der Trägerschwingung moduliert. Eine allgemeine, komplexe Beschreibung eines unmodulierten Schwingungssignals $s(t)$ sei als komplexes Signal mit der Trägerfrequenz f_0 gegeben mit:

$$s(t) = a(t) \cdot \exp\{j \cdot (2\pi f_0 t + \varphi(t))\} \quad \text{mit: } \begin{cases} \varphi(t) = \varphi_0 \\ a(t) = a_0 \end{cases} \quad (2-8)$$

Dabei ist φ_0 ein Phasenoffset, der die anfängliche Phasenlage oder auch das Voreilen der Schwingung beschreibt.

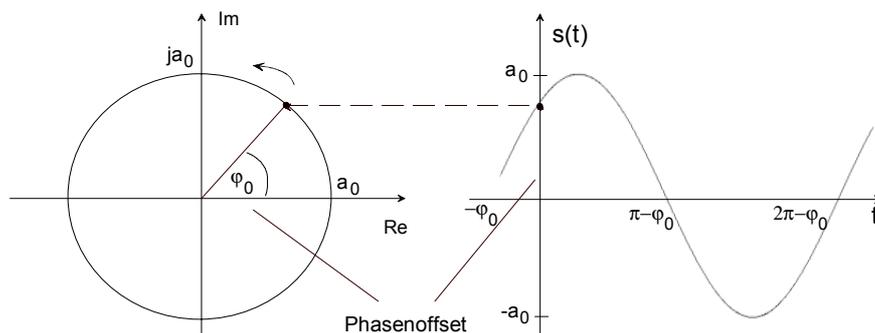


Abb. 2.2: Phasenoffset eines Trägersignals

² φ_0 bezeichnet nur einen Phasenoffset, also eine konstante Größe.

2.1.1.2.1 Winkelmodulation

Bei der Winkelmodulation (WM) bleibt die Amplitude des Sendesignals $s(t)$ konstant, und das Nutzsignal $m(t)$ wird zur Veränderung der Momentanphase verwendet:

$$s(t) = a(t) \cdot \cos(2\pi f_0 t + \varphi(t)) \quad \text{mit:} \quad \begin{cases} \varphi(t) = \text{variable Phase inkl. Phasenoffset } \varphi_0 \\ a(t) = a_0 \end{cases} \quad (2-9)$$

Allgemein gilt außerdem:

$$T = \frac{1}{f_0} \quad \text{Periodendauer der Trägerschwingung}$$

Das zugehörige äquivalente Tiefpaßsignal lautet dann:

$$s_T(t) = a_0 \cdot \exp(j\varphi(t)) \quad (2-10)$$

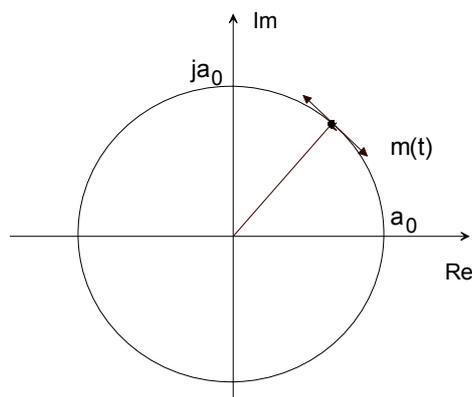


Abb. 2.3: Phasenzeigerdiagramm bei WM

Die verschiedenen Winkelmodulationsarten werden in zwei Gruppen eingeteilt: Die Phasenmodulation und die Frequenzmodulation:

Bei der Phasenmodulation (PM) wird das Nutzsignal $m(t)$ direkt als Phase $\varphi(t)$ verwendet. Der Zusammenhang lautet dann:

$$\varphi(t) = \gamma_{PM} \cdot m(t) + \varphi_0,$$

wobei γ_{PM} als Modulationsindex bezeichnet wird. Er ist definiert durch das Verhältnis von Frequenzhub Δf zu f_M , der Modulationsfrequenz des Nutzsignals $m(t)$:

$$\gamma_{PM} = \frac{\Delta f}{f_M} = \Delta\varphi$$

Er ist somit identisch mit dem (maximalen) Phasenhub $\Delta\varphi$, wenn $|m(t)| \leq 1$. Damit ist der Dynamikbereich bei winkelmodulierten Signalen erheblich größer als bei der AM.

Bei der Frequenzmodulation wird nicht das Nutzsignal $m(t)$ direkt als Phase verwendet, sondern das Kurzzeitintegral über $m(t)$:

$$\varphi(t) = \gamma_{PM} \cdot \int m(t) dt + \varphi_0 \quad (2-11)$$

$$\Rightarrow \dot{\varphi}(t) = \gamma_{PM} \cdot m(t) = f_A(t) \cdot 2\pi$$

Das Signal $m(t)$ findet sich dann in der Ableitung der Phase, der Augenblicksfrequenz $2\pi \cdot f_A$ wieder. Der Zusammenhang zwischen Phasen- und Frequenzmodulation kann wie folgt beschrieben werden:

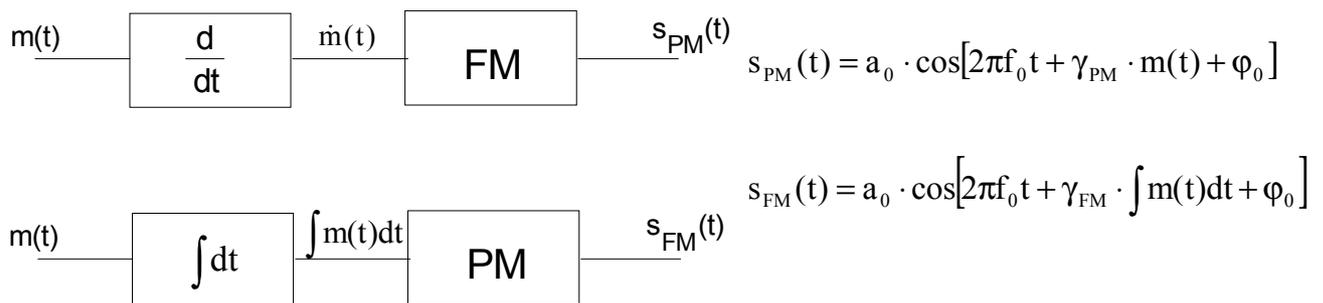


Abb. 2.4 Systemtheoretischer Vergleich von PM und FM

In Verbindung mit einem vorgeschalteten Differenzierglied bzw. einem Integrationsglied läßt sich ein Frequenzmodulator auch als Phasenmodulator verwenden bzw. ein Phasenmodulator auch zur Frequenzmodulation einsetzen.

Bei Kombination von Amplituden- und Winkelmodulation - Quadraturamplitudenmodulation genannt - wird eine Übertragung von 2 unterschiedlichen Signalen $m_1(t), m_2(t)$ mit Hilfe eines Trägers möglich. Ein allgemeines QAM-Signal lautet in komplexer Darstellung:

$$s(t) = a_0 (1 + \mu_{AM} m_1(t)) \cdot \exp j(2\pi f_0 t + \gamma_{WM} m_2(t) + \varphi_0) \quad (2-12)$$

2.1.1.3 Diskrete Modulationsverfahren

Eine andere Möglichkeit Signale zu übertragen, hat sich heutzutage durchgesetzt: Digitale Modulationsverfahren.

Das zeitkontinuierliche Signal wird dazu abgetastet und als Digitalsignal weiterverarbeitet. Entweder werden die zeitlichen Abtastwerte oder die spektralen Komponenten des Signals (Telekommunikation) weiterverwendet. Die digitalen, diskreten Werte können dann durch Modulation übertragen werden. Analog zu den zeitkontinuierlichen Modulationsverfahren ergeben sich die folgenden zeitdiskreten Modulationsverfahren /18, 27, 45, 60/.

2.1.2 Demodulationsverfahren

Der Vorgang zur Rückgewinnung des Nachrichtensignals aus dem modulierten Signal wird als Demodulation, das hierzu notwendige System als Demodulator bezeichnet. Benötigt der Demodulator Frequenz und Phasenwinkel des Trägersignals, so spricht man von kohärenter Demodulation, in anderen Fällen von inkohärenter Demodulation. Die praktische Demodulation kann bei AM - Signalen mit einem Modulationsgrad $\mu_{AM} \leq 1$ auf inkohärente Art durch einfache Betragsbildung mit Tiefpaßfilterung erfolgen. Die realisierte Demodulation winkelmodulierter Signale erfolgt dagegen in mehreren Schritten: Umwandeln, Gleichrichten und Sieben /62, 77/.

Bei der Demodulation mit einem Kalman-Filter werden nach der Transformation in den Tiefpaßbereich alle Demodulationsschritte im Filter zur gleichen Zeit durchgeführt. Der Ansatz zur Berechnung des Filters wird im 4. Abschnitt dieses Kapitels dargestellt.

Durch Verwendung zusätzlicher Zustände zur Amplitudenschätzung lassen sich auch AM-Signale und QAM-Signale mit einem EKF verarbeiten.

2.1.2.1 Demodulation mit dem Quadraturempfänger

Ein modulierte Signal $r(t)$ wird als unverraushtes Signal auf einen Quadraturempfänger gegeben. Durch Multiplikation mit dem Trägersignal und dem um 90° phasenverschobenen Träger ergeben sich die Signale $r_1(t)$ und $r_2(t)$. Multiplikation und Tiefpaßfilterung entsprechen der praktischen Umsetzung einer Frequenzverschiebung:

$$\begin{aligned}
 r(t) &= a_0 \cdot \cos(2\pi f_0 t + \varphi(t)) \\
 &= \frac{a_0}{2} \cdot (\exp[j(2\pi f_0 t + \varphi(t))] + \exp[-j(2\pi f_0 t + \varphi(t))]) \quad |^3 \tag{2-13}
 \end{aligned}$$

Um die zwei Quadraturkomponenten des Empfängers zu berechnen, wird eine Multiplikation mit der Trägerfrequenz und der um 90° phasenverschobenen Trägerfrequenz durchgeführt. Die Multiplikation mit $\cos(2\pi f_0 t)$ und $-\sin(2\pi f_0 t)$ lassen sich zu einer komplexen Größe zusammenfassen, um somit die Berechnung zu vereinfachen. Wir multiplizieren also das Empfangssignal mit $\exp[-j(2\pi f_0 t)] = \cos(2\pi f_0 t) - j\sin(2\pi f_0 t)$. Die komplexe Betrachtungsweise beider Quadraturkomponenten ist zulässig, da die beiden Signale orthogonal werden und somit unabhängig voneinander sind.

$$\begin{aligned}
 \Rightarrow \quad \tilde{r}(t) &= \frac{a_0}{2} \cdot (\exp[j(2\pi f_0 t + \varphi(t))] + \exp[-j(2\pi f_0 t + \varphi(t))]) \cdot \exp[-j(2\pi f_0 t)] \\
 &= \frac{a_0}{2} \cdot (\exp[j(\varphi(t))] + \exp[-j(4\pi f_0 t + \varphi(t))])
 \end{aligned}$$

Anschließend wird dieses komplexe Signal $\tilde{r}(t)$ (stellvertretend für die beiden Teilsignale) so tiefpaßgefiltert, daß die Anteile doppelter Trägerfrequenz verschwinden. Eine weitere Bearbeitung im Frequenzbereich ist sinnvoll, daher transformieren wir $\tilde{r}(t)$ in den Frequenzbereich:

$$\begin{aligned}
 \tilde{R}(f) &= \int_{-\infty}^{+\infty} r(t) \cdot e^{-j2\pi ft} dt \\
 &= \frac{a_0}{2} \int_{-\infty}^{+\infty} e^{-j(4\pi f_0 t + \varphi(t))} \cdot e^{-j2\pi ft} dt + \frac{a_0}{2} \int_{-\infty}^{+\infty} e^{j(\varphi(t))} \cdot e^{-j2\pi ft} dt
 \end{aligned}$$

Der Term $e^{j\varphi(t)}$ wird als Hilfsvariable $s_T(t)$ (äquiv. Tiefpaßsignal) zusammengefaßt:

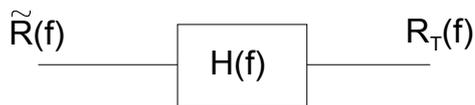
$$\begin{aligned}
 \tilde{R}(f) &= \frac{1}{2} \int_{-\infty}^{+\infty} s_T(t) \cdot e^{-j2\pi ft} e^{-j4\pi f_0 t} dt + \frac{1}{2} \int_{-\infty}^{+\infty} s_T(t) \cdot e^{-j2\pi ft} dt \\
 &= \frac{1}{2} S_T(f + 2f_0) + \frac{1}{2} S_T(f) \tag{2-14}
 \end{aligned}$$

³ $\varphi(t)$ enthält immer auch einen Phasenoffsetwinkel φ_0 .

Das gewählte Tiefpaßfilter muß mindestens die Bandbreite f_{Δ} von $S_T(f)$ besitzen:

$$H(f) = \text{rect} \frac{f}{f_{\Delta}} \bullet - \circ h(t) = f_{\Delta} \cdot \text{si}(\pi \cdot f_{\Delta} \cdot t)$$

Mit diesem Filter erfolgt die Tiefpaßfilterung:



$$R_T(f) = \tilde{R}(f) \cdot H(f) = \frac{1}{2} S_T(f) \tag{2-15}$$

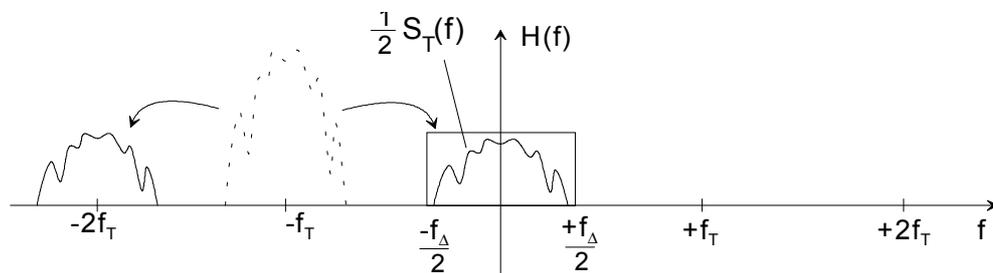


Abb. 2.5 Basisbandfilterung

Die Rücktransformation in den Zeitbereich ergibt:

$$R_T(f) \bullet - \circ r_T(t) = \frac{a_0}{2} \cdot e^{j\varphi(t)} \tag{2-16}$$

$$r_T(t) = \frac{a_0}{2} \cdot e^{j\varphi(t)} = \frac{a_0}{2} \cdot [\cos(\varphi(t)) + j\sin(\varphi(t))] \tag{2-17}$$

Anschließend wird das komplexe Signal wieder in die beiden Teilkomponenten aufgetrennt: Der Realteil von $r_T(t)$ ist das Resultat aus der Multiplikation mit dem $\cos(2\pi f_0 t)$ und der Imaginärteil von $r_T(t)$ repräsentiert das Resultat aus der Multiplikation mit dem phasenverschobenen Trägersignal ($-\sin(2\pi f_0 t)$). Das in den äquivalenten Tiefpaßbereich transformierte Signal ist in seine Quadraturkomponenten aufgespalten worden, so daß wir zusammenfassend die folgenden beiden Signale aus dem Quadraturempfänger erhalten:

$$\frac{r_{T\text{Re}}(t)}{2} = \frac{a_0}{2} \cdot \cos(\varphi(t)) \quad \text{Realteil des Signals } r_T(t), \text{ Inphasekomponente} \tag{2-18}$$

$$\frac{r_{TIm}(t)}{2} = \frac{a_0}{2} \cdot \sin(\varphi(t)) \quad \text{Imaginärteil des Signals } r_T(t), \text{ Quadraturkomponente (2-19)}$$

Um bei einer AM die Hüllkurve zu bestimmen, werden die Quadrate der Quadraturkomponenten gebildet, summiert und die positive Wurzel als gesuchtes Modulationssignal weiterverwendet. Die Amplitude a_0 ist in diesem Fall eine zeitveränderliche Variable $a(t)$.

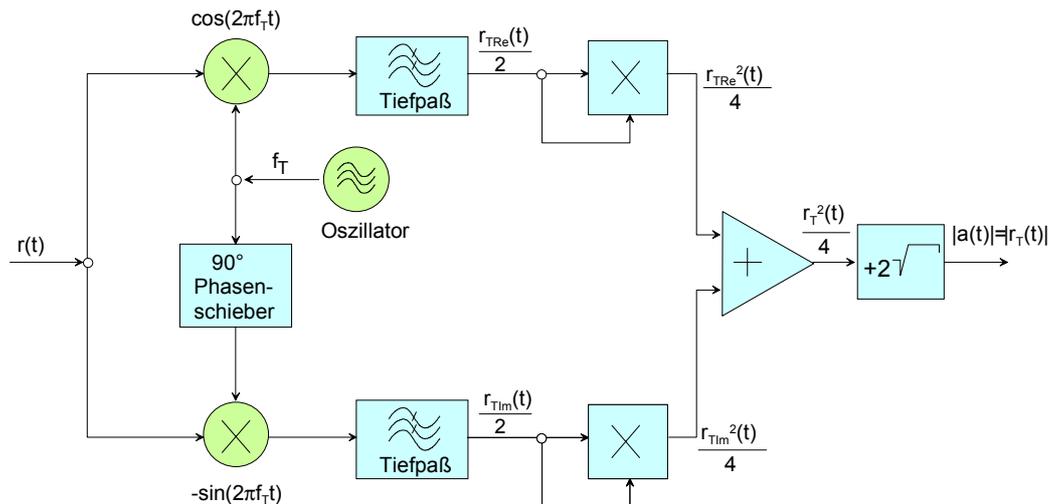


Abb. 2.6: Quadraturempfänger zur Demodulation der Amplitude

Es wird gebildet:

$$\frac{r_{TRe}^2(t)}{4} = \frac{a^2(t)}{4} \cdot \cos^2(\varphi(t))$$

$$\frac{r_{TIm}^2(t)}{4} = \frac{a^2(t)}{4} \cdot \sin^2(\varphi(t))$$

$$\frac{|r_T(t)|}{2} = +\sqrt{\frac{r_{TRe}^2}{4} + \frac{r_{TIm}^2}{4}} = \sqrt{\frac{a^2(t)}{4} \cdot \cos^2(\varphi(t)) + \frac{a^2(t)}{4} \cdot \sin^2(\varphi(t))} = \sqrt{\frac{a^2(t)}{4}} = \frac{|a(t)|}{2}$$

Wir erhalten also zunächst nur die halbe Amplitude! Um die Bildungsvorschrift eines äquivalenten Tiefpaßsignals zu erfüllen, multiplizieren wir das Endergebnis mit dem Faktor 2.

$$|r_T(t)| = |a(t)| \quad (2-20)$$

Der Quadraturempfänger wird nun zur Demodulation des zeitveränderlichen Winkels eingesetzt. Mit den beiden Quadraturkomponenten wird der Arcustangens berechnet, der den Winkel $\varphi(t)$ bestimmt:

$$\varphi(t) = \arctan \left[\frac{\frac{1}{2} r_{T\text{Im}}(t)}{\frac{1}{2} r_{T\text{Re}}(t)} \right] = \arctan \left[\frac{r_{T\text{Im}}(t)}{r_{T\text{Re}}(t)} \right] \quad (2-21)$$

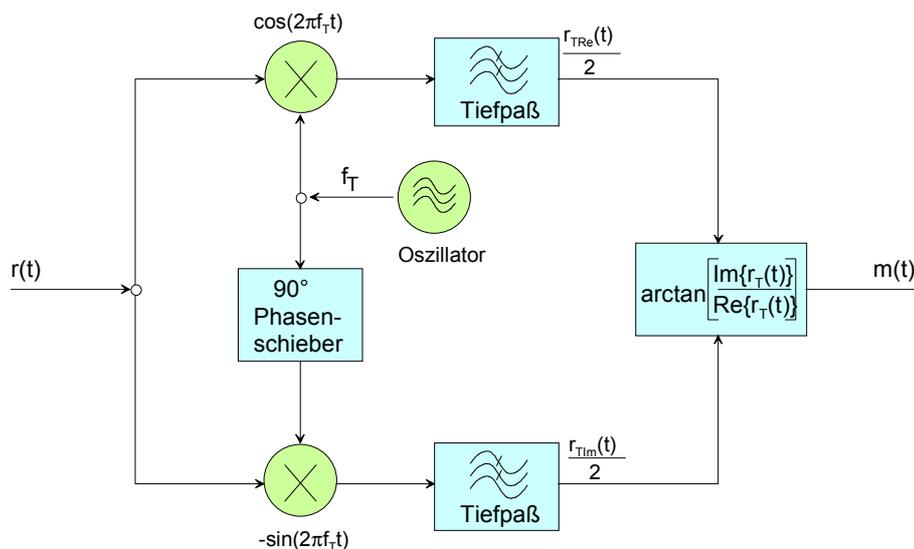


Abb. 2.7: Quadraturempfänger zur Demodulation der Phase

Grundsätzlich wird hierbei vorausgesetzt, daß bei der Modulation der Definitionsbereich der Arcustangens-Funktion nicht überschritten wird, da sonst Mehrdeutigkeiten, d.h. Phasensprünge bei der Demodulation auftreten! Die Quadraturkomponenten müssen einen eindeutigen Quadranten aufspannen (s. 4.5.1.1).

2.1.2.2 Demodulation mit PLL

Ein anderes häufig angewendetes Verfahren der Winkeldemodulation ist die Verwendung von PLL's. Dieser phasengesteuerte Regelkreis wird heutzutage beispielsweise in den meistens FM-Empfängern eingesetzt. Der PLL besteht aus einem Phasendetektor, der das modulierte Signal mit einem unmodulierten, nachlaufenden Träger vergleicht. Der idealerweise als Multiplizierer aufgebaute Phasendetektor wird jedoch oft nur als digitales EX-OR-Gatter aufgebaut. Bei der Realisierung der digitalen Lösung muß das Tastverhältnis des Trägersignals 50% betragen, um keine undefinierten Bereiche zu erhalten, was von Nachteil ist. Ebenso ist es nicht möglich, große Frequenzsprünge sicher zu detektieren /9, 11, 77/.

Je nachdem, ob der im PLL verwendete Phasendetektor linear oder digital arbeitet, wird von einem linearen oder digitalen PLL gesprochen. Diese Mischprodukte sind ein Maß für die Frequenzabweichung des empfangenen Signals zur VCO-Frequenz. Das Resultat wird über einem Tiefpaß zu einer Gleichspannung geglättet, die zur Ansteuerung eines spannungsgesteuerten Oszillators (VCO's) verwendet wird.

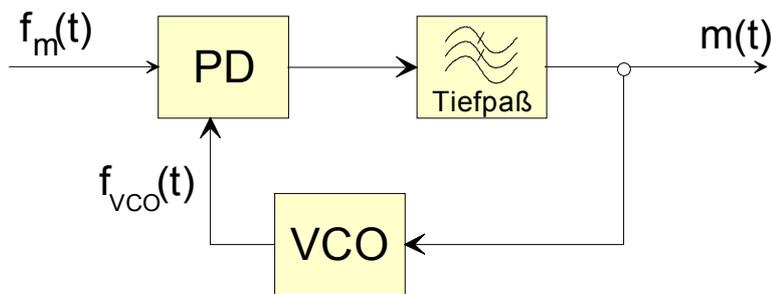


Abb. 2.8: PLL (phase locked loop = Phasenregelkreis)

Der VCO schwingt mit einer Ruhefrequenz $f_{\hat{0}}$, der so konfiguriert werden sollte, daß $f_{\hat{0}}$ direkt der Trägerfrequenz f_0 des Modulationssignals entspricht. In diesem Fall ist die Ansteuerspannung / Regelspannung $m(t)$ Null. Das Empfangssignal, mit der sich um die Trägerfrequenz f_0 ändernden, modulierten Augenblicksfrequenz $f_m(t)$ führt am Phasendetektor (PD) zu Mischprodukten, die nach Glättung zu einer Regelspannung $m(t)$ führen. Eine positive Regelspannung wird bei $f_m > f_{\hat{0}}$ detektiert. Die Ruhefrequenz $f_{\hat{0}}$ wird vom VCO auf $f_{VCO}(t)$ so lange nachgeregelt, bis $f_m(t)$ erreicht ist.

$$f_m(t) = f_{VCO}(t) = f_{\hat{0}} + K \cdot m(t) \quad (2-22)$$

Die Regelspannung bildet auf diese Weise die Frequenzänderungen des Empfangssignals ab, was einer Frequenzdemodulation entspricht. Benötigen wir eine Phasendemodulation, so muß die Regelspannung abschließend integriert werden.

(Der Faktor K ist ein Systemfaktor, der in erster Linie vom VCO abhängig ist, aber auch Verstärkungsfaktoren des PD's enthält.) Ist die Nachregelung erfolgreich, so wird die Regelspannung $m(t)$ Null, der PLL ist eingerastet. Das Verhalten eines PLL's wird in unterschiedliche Arbeitsbereiche eingeteilt, wobei in erster Linie der Fangbereich bei Anwendungen entscheidend ist:

1. Fangbereich: Frequenzbereich, in dem der PLL auf eine vorgegebene Frequenz eines angelegten Signals einrastet.
2. Haltebereich: Frequenzbereich, in dem der PLL in bereits eingerastetem Zustand arbeitet.
3. Ziehbereich (Pull In Range): Frequenzbereich, in dem der PLL auf eine vorgegebene Frequenz eines angelegten Signals möglicherweise noch einrasten kann.
- (4. Pull Out Range: Frequenzbereich, in dem der PLL auch bei sprunghaftigen Frequenzänderungen noch im eingerasteten Zustand verbleibt.)

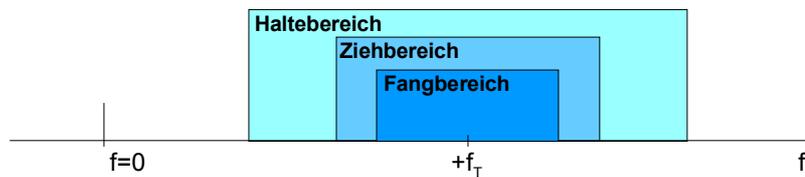


Abb. 2.9: Arbeitsbereiche des PLL

Der PLL wurde hier zunächst aus mehreren Gründen behandelt. Er stellt zum einen einen stabilen und gebräuchlichen Frequenzdemodulator dar und besitzt zum anderen deutliche Parallelen im Vergleich mit dem verwendeten Extended Kalman-Filter. Eine Betrachtung dazu findet sich im Abschnitt 2.4.4. Eine Einspeisung der Phase oder Frequenz als zusätzliche Beobachtung im Kalman-Filter wurde zusätzlich hardwaremäßig vorbereitet.

2.1.3 Systemeigenschaften einer WM-Übertragungsstrecke

Aufgespalten in Funktionsblöcke mit ihren Stoßantworten $h_i(t)$ stellt sich die Übertragung eines Quellsignals $x(t)$ mit WM-Modulation folgendermaßen dar:

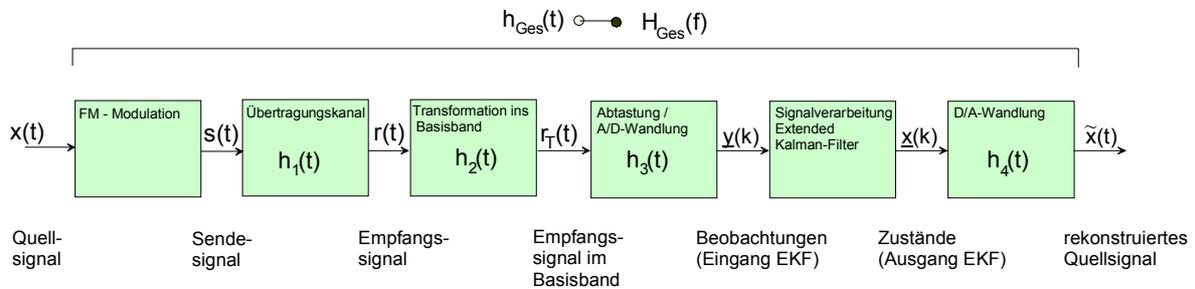


Abb. 2.10: Übertragungsstrecke

Die Abtastung bei der AD-Wandlung und die Rekonstruktion bei der DA-Wandlung wird dabei als so fein quantisiert betrachtet, daß die Quantisierung als linear betrachtet werden darf.

Der 1. Block stellt die bereits beschriebene Winkelmodulation des Quellsignals dar:

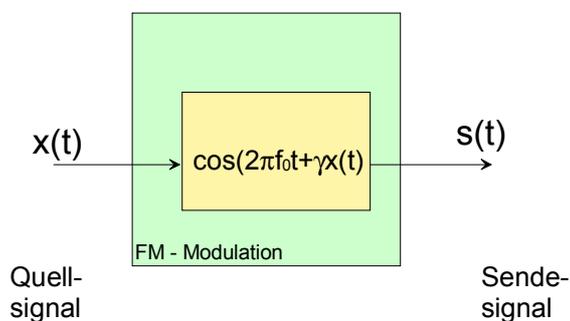


Abb. 2.11: Modulation

Der 2. Block beschreibt die Eigenschaften des WM-Übertragungskanals. Der Zusammenhang zwischen $x(t)$ und $s(t)$ ist dabei absolut nichtlinear, da er mit einer trigonometrischen Funktion beschrieben wird. Das Empfangssignal $r(t)$ eines modulierten Sendesignals hat sich aufgrund von äußeren Signaleinflüssen deutlich verändert. Die Eigenschaften des Empfangssignals sollen im folgenden untersucht und dabei Störungen zusammengefaßt und definiert werden. Mögliche Echos (z.B. durch Reflexionen) sollen dabei nicht als diese betrachtet werden, da sie zusätzliche Parameter mit in die Betrachtung einfließen lassen würden. Es sollen ausschließlich Dämpfung, Verzögerung und die für die Filtermodellierung entscheidende Größe, das additive Übertragungsrauschen das Sendesignal beeinflussen.

Das Rauschsignal wird idealisiert als gaußverteilt und mittelwertfrei angenommen. Besitzt das Übertragungsrauschen nicht diese Eigenschaften, so müssen deterministische Größen zusammengefaßt als zusätzliche Steuergröße U des Kalman-Filters definiert werden und abweichende Verteilungsdichten in der Varianz des Systemrauschens (d.h. als Modellfehler) oder durch ein anderes Filter höherer Ordnung z.B. Gaußfilter berücksichtigt werden.

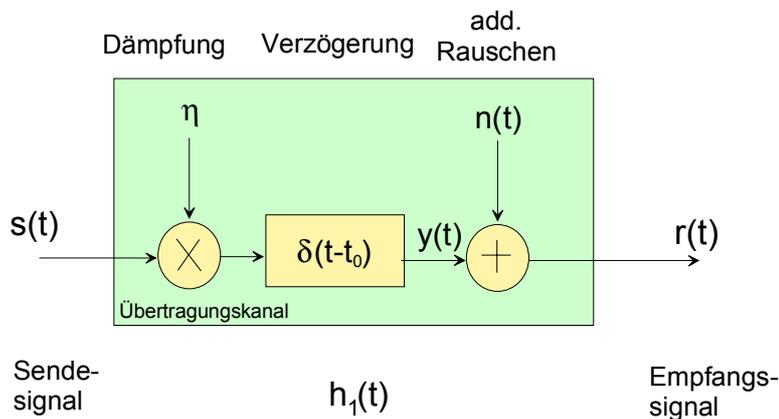


Abb. 2.12: Übertragungskanal

Die systemtheoretische Beschreibung lautet für ein WM-Empfangssignal:

$$r(t) = y(t) + n(t) = \eta \cdot s(t - t_0) + n(t)$$

$$s(t) = a \cdot \cos(2\pi f_0 t + \varphi(t))$$

$$\varphi(t) = \gamma_{\text{WM}} \cdot m(t) + \varphi_0$$

Eigenschaften des Sendesignals $s(t)$ wurden bereits im 1. Abschnitt beschrieben. $m(t)$ stellt das zu übertragende Quellsignal dar. Alle konstanten Terme können zu einem gemeinsamen, neuen Phasenoffset φ_0 zusammengefaßt werden:

Mit: $r(t) = a \cdot \eta \cdot \cos(2\pi f_0(t - t_0) + \gamma_{\text{WM}} \cdot m(t - t_0) + \varphi_{0\text{alt}}) + n(t)$ folgt für den Phasenoffset φ_0 :

$$\varphi_0 = -2\pi f_0 t_0 + \varphi_m + \varphi_{0\text{alt}} \quad (2-23)$$

Der Term $-2\pi f_0 t_0$ ist eine konstante Phasenverschiebung. φ_m ist abhängig vom Verlauf von $m(t)$ und läßt sich nicht direkt bestimmen. Sie entsteht bei der Transformation in den äquivalenten Tiefpaßbereich und wird als Ausbreitungsphase bezeichnet. Die Dämpfung der Amplitude wird ebenfalls verkürzt beschrieben als: $\eta = a \cdot \eta_{\text{alt}}$

Zur Demodulation des Empfangssignals mit einem Quadraturempfänger wird dieses in das Basisband bzw. den äquivalenten Tiefpaßbereich heruntertransformiert. Dies entspricht der Vorverarbeitung des Empfangssignals für die Demodulation mit dem Kalman-Filter.

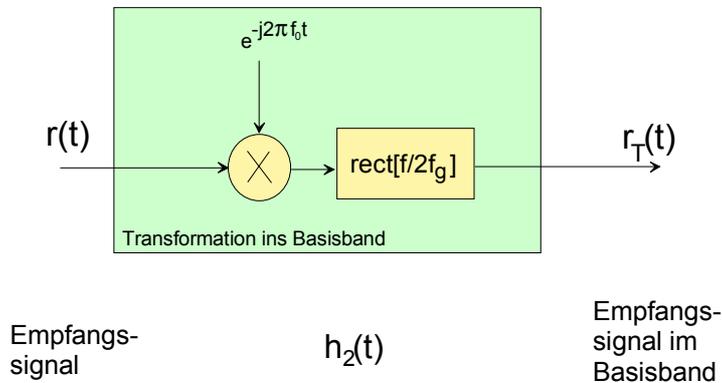


Abb. 2.13: Transformation ins Basisband

Wir betrachten zunächst die Inphasekomponente der Quadraturemodulation und schließen daraus auf die Quadraturkomponente. Das Empfangssignal $r(t)$ wird mit dem phasengleichen Träger multipliziert. Wir erhalten Anteile mit doppelter und ohne Trägerfrequenz:

$$r(t) = \eta \cdot \cos(2\pi f_0 t + \gamma_{\text{WM}} \cdot \varphi(t) + \varphi_0) + n(t) \quad (2-24)$$

$$r(t) \cdot \cos(2\pi f_0 t) = [\eta \cdot \cos(2\pi f_0 t + \gamma_{\text{WM}} \cdot \varphi(t) + \varphi_0) + n(t)] \cdot \cos(2\pi f_0 t)$$

$$r'_T(t) = \frac{\eta}{2} \cdot \cos(4\pi f_0 t + \gamma_{\text{WM}} \cdot \varphi(t) + \varphi_0) + \frac{\eta}{2} \cdot \cos(\gamma_{\text{WM}} \cdot \varphi(t) + \varphi_0) + n(t) \cdot \cos(2\pi f_0 t)$$

Danach erfolgt die Tiefpaßfilterung mit einem idealen Tiefpaß : $H(f) = \text{rect} \frac{f}{2f_g}$

Dessen Stoßantwort $h(t)$ lautet: $h(t) = 2f_g \cdot \text{si}(2\pi f_g t)$

Die Tiefpaßfilterung wird sinnvollerweise im Frequenzbereich durchgeführt.

$$R_{\text{TRe}}(f) = \eta \cdot \frac{1}{2} \mathcal{F}\{\cos(\gamma \cdot \varphi(t) + \varphi_0)\} \cdot \text{rect} \frac{f}{2f_g} +$$

$$\eta \cdot \frac{1}{2} \mathcal{F}\{\cos(4\pi f_0 t + \gamma_{\text{WM}} \cdot \varphi(t) + \varphi_0)\} \cdot \text{rect} \frac{f}{2f_g} + \mathcal{F}\{n(t) \cdot \cos(2\pi f_0 t)\} \cdot \text{rect} \frac{f}{2f_g}$$

Da eine direkte Transformation des letzten Terms aufgrund dessen stochastischer Eigenschaften nicht möglich ist, werden die spektralen Eigenschaften des Terms über die Betrachtung der Erwartungswerte erfolgen:

Das additive Rauschen in den äquivalenten Tiefpaßbereich transformiert lautet hier (komplexe Darstellung):

$$\underline{n}_T(t) = n(t) \cdot e^{j2\pi f_0 t} * 2f_g \text{si}(2\pi f_g t) \quad (2-25)$$

Die Eigenschaften von $\underline{n}_T(t)$ werden durch die beiden Momente des Rauschens beschrieben: Das erste Moment lautet: $E\{\underline{n}_T(t)\} = 0$, da das Signal als mittelwertfrei angenommen wurde. Das 2. Moment $\phi_{\underline{n}_T \underline{n}_T}^L(\tau) = E\{\underline{n}_T(t) \cdot \underline{n}_T(t + \tau)\}$ wird über die Erwartungswerte direkt benachbarter Rauschwerte $\underline{n}_T(t)$ und $\underline{n}_T(t + \tau)$ gebildet und ist im Frequenzbereich als Rauschleistungsdichtespektrum $\phi_{\underline{n}_T \underline{n}_T}^L(f)$ bekannt. Wir bestimmen die Korrelationsfunktionen der beiden Rauschsignale als Erwartungswerte. Die beiden entstehenden Rauschanteile werden dazu zu einem Rauschvektor $\underline{n}_T(t)$ zusammengefaßt:

$$\underline{n}_T(t) = \begin{pmatrix} \underline{n}_{T\text{Re}}(t) \\ \underline{n}_{T\text{Im}}(t) \end{pmatrix} = \begin{pmatrix} n_1(t) \\ n_2(t) \end{pmatrix}$$

Die zu betrachtenden Leistungen und Kreuzleistungen können in der Kovarianzmatrix zusammengefaßt werden:

$$\text{cov}(\underline{n}_T(t), \underline{n}_T(t + \tau)) = \begin{bmatrix} E\{\underline{n}_{T\text{Re}}(t) \cdot \underline{n}_{T\text{Re}}(t + \tau)\} & E\{\underline{n}_{T\text{Re}}(t) \cdot \underline{n}_{T\text{Im}}(t + \tau)\} \\ E\{\underline{n}_{T\text{Im}}(t) \cdot \underline{n}_{T\text{Re}}(t + \tau)\} & E\{\underline{n}_{T\text{Im}}(t) \cdot \underline{n}_{T\text{Im}}(t + \tau)\} \end{bmatrix} \quad (2-26)$$

Das 2. Moment lautet:

$$\begin{aligned} &= E\{\underline{n}_{T\text{Re}}(t) \cdot \underline{n}_{T\text{Re}}(t + \tau)\} \\ &= E\left\{ \left[n(t) \cdot \cos(2\pi f_0 t) * 2f_g \text{si}(2\pi f_g t) \right] \cdot \left[n(t) \cdot \cos(2\pi f_0 (t + \tau)) * 2f_g \text{si}(2\pi f_g (t + \tau)) \right] \right\} \\ &= E \left\{ \int_{-\infty}^{+\infty} n(\xi_1) \cdot \cos(2\pi f_0 \xi_1) \cdot 2f_g \cdot \text{si}[2\pi f_g (t - \xi_1)] d\xi_1 \cdot \int_{-\infty}^{+\infty} n(\xi_2) \cdot \cos(2\pi f_0 \xi_2) \cdot 2f_g \cdot \text{si}[2\pi f_g (t + \tau - \xi_2)] d\xi_2 \right\} \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} N_0 \delta(\xi_2 - \xi_1) \cdot \cos(2\pi f_0 \xi_1) \cdot \cos(2\pi f_0 \xi_2) \cdot 4f_g^2 \cdot \text{si}[2\pi f_g (t - \xi_1)] \cdot \text{si}[2\pi f_g (t + \tau - \xi_2)] d\xi_1 d\xi_2 \end{aligned}$$

Durch Anwendung der Siebeigenschaft fällt die Integration in Richtung ξ_2 heraus:

$$= N_0 \int_{-\infty}^{+\infty} \cos^2(2\pi f_0 \xi_1) \cdot 4f_g^2 \cdot \text{si}[2\pi f_g(t - \xi_1)] \cdot \text{si}[2\pi f_g(t + \tau - \xi_1)] d\xi_1$$

Substitution: $t - \xi_1 = u$ und $-d\xi_1 = du$ mit Änderung der Grenzen und Auflösung von \cos^2 führt zu:

$$\begin{aligned} &= N_0 \int_{-\infty}^{+\infty} \frac{1}{2}(1 + \cos(4\pi f_0(t - u))) \cdot 4f_g^2 \cdot \text{si}[2\pi f_g u] \cdot \text{si}[2\pi f_g(u + \tau)] du \\ &= N_0 \int_{-\infty}^{+\infty} \frac{1}{2} \cdot 4f_g^2 \cdot \text{si}[2\pi f_g u] \cdot \text{si}[2\pi f_g(u + \tau)] du + N_0 \int_{-\infty}^{+\infty} \frac{1}{2} \cos(4\pi f_0(t - u)) \cdot 4f_g^2 \cdot \text{si}[2\pi f_g u] \cdot \text{si}[2\pi f_g(u + \tau)] du \end{aligned}$$

Wir betrachten das 1. Integral als Faltungsintegral und lösen die Cosinus-Fkt. im zweiten Integral weiter auf:

$$= 2N_0 f_g^2 \cdot \text{si}[2\pi f_g \tau] * \text{si}[2\pi f_g \tau] + I_2 \quad (2-27)$$

Durch Betrachtung im Frequenzbereich läßt sich I_2 aus (2-27) weiter vereinfachen:

$$\begin{aligned} &= 2N_0 f_g^2 \cdot \frac{1}{2f_g} \text{si}[2\pi f_g \tau] + 2N_0 f_g^2 \cdot \mathcal{F}^{-1} \left\{ \left[\frac{1}{2} (\delta(f - 2f_0) + \delta(f + 2f_0)) \cdot e^{-j2\pi f t} * \frac{1}{2f_g} \text{rect} \frac{f}{2f_g} \right] \cdot \frac{1}{2f_g} \text{rect} \frac{f}{2f_g} \right\} \\ &= N_0 f_g \cdot \text{si}[2\pi f_g \tau] + 2N_0 f_g^2 \cdot \mathcal{F}^{-1} \left\{ \underbrace{\left[\frac{1}{4f_g} \left(\text{rect} \frac{f - 2f_0}{2f_g} \cdot e^{j4\pi f_0 t} + \text{rect} \frac{f + 2f_0}{2f_g} \cdot e^{-j4\pi f_0 t} \right) \right]}_{=0} \cdot \frac{1}{2f_g} \text{rect} \frac{f}{2f_g} \right\} \\ &= N_0 f_g \cdot \text{si}[2\pi f_g \tau] \end{aligned}$$

Das Resultat für I_2 ist jedoch nur dann Null, wenn $2f_g < 2f_0$, so daß sich schließlich für den Erwartungswert ergibt (mit Resubstitution bei symmetrischer Ergebnisfunktion):

$$E\{n_{\text{TRe}}(t) \cdot n_{\text{TRe}}(t + \tau)\} = N_0 f_g \cdot \text{si}[2\pi f_g \tau] \quad (2-28)$$

Anschließend werden die weiteren Erwartungswerte bestimmt. Bei der Autokorrelation von $n_{\text{TIm}}(t)$ ergibt sich die gleiche Leistung wie bei der ermittelten Autokorrelation von $n_{\text{TRe}}(t)$:

Mit: $E\{n_{\text{TIm}}(t) \cdot n_{\text{TIm}}(t + \tau)\} =$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} N_0 \delta(\xi_2 - \xi_1) \cdot \sin(2\pi f_0 \xi_1) \cdot \sin(2\pi f_0 \xi_2) \cdot 4f_g^2 \cdot \text{si}[2\pi f_g(t - \xi_1)] \cdot \text{si}[2\pi f_g(t + \tau - \xi_2)] d\xi_1 d\xi_2$$

$$= N_0 \int_{-\infty}^{+\infty} \frac{1}{2} (1 - \cos(4\pi f_0(t-u))) \cdot 4f_g^2 \cdot \text{si}[2\pi f_g u] \cdot \text{si}[2\pi f_g(u+\tau)] du$$

$$E\{n_{TIm}(t) \cdot n_{TIm}(t+\tau)\} = N_0 f_g \cdot \text{si}[2\pi f_g \tau] \quad (2-29)$$

Die Kreuzleistungen bzw. Kreuzkovarianzen werden Null, da sich bei Berechnung des Integrals ein $\sin(x)\cos(x)$ -Term ergibt, der zu $\frac{1}{2}\sin(2x)$ als ungerade Funktion zusammengefaßt werden kann. Aufgrund der Unabhängigkeit von Realteilrauschen und Imaginärteilrauschen wurde dies hier auch erwartet.

$$\begin{aligned} E\{n_{TRe}(t) \cdot n_{TIm}(t+\tau)\} &= \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} N_0 \delta(\xi_2 - \xi_1) \cdot \cos(2\pi f_0 \xi_1) \cdot \sin(2\pi f_0 \xi_2) \cdot 4f_g^2 \cdot \text{si}[2\pi f_g(t-\xi_1)] \cdot \text{si}[2\pi f_g(t+\tau-\xi_2)] d\xi_1 d\xi_2 \end{aligned}$$

$$E\{n_{TRe}(t) \cdot n_{TIm}(t+\tau)\} = E\{n_{TIm}(t) \cdot n_{TRe}(t+\tau)\} = 0 \quad (2-30)$$

Aufgrund der Symmetrien berechnet sich der zweite Kreuzterm ebenfalls zu Null. Daraus ergibt sich abschließend die Kovarianzmatrix:

$$\text{cov}(\underline{n}_T(t), \underline{n}_T(t+\tau)) = \begin{bmatrix} f_g N_0 \text{si}(2\pi f_g \tau) & 0 \\ 0 & f_g N_0 \text{si}(2\pi f_g \tau) \end{bmatrix} \quad (2-31)$$

Realteil und Imaginärteil von $n_T(t)$ sind zunächst unkorreliert und somit wegen der Gaußprozeß-Eigenschaft auch voneinander unabhängig. Die Kreuzkorrelation der beiden Anteile ist Null. Es handelt sich folglich um zwei unabhängige Rauschprozesse, die jedoch aufgrund der Tiefpaßfilterung nicht mehr weiß sind. Mit $\varphi_{nn}(\tau) = N_0 \cdot \delta(\tau)$ und mit $\varphi_{n_T n_T}(\tau) = N_0 f_g \cdot \text{si}[2\pi f_g \tau]$ erhalten wir durch Transformation in den Frequenzbereich:

$\phi_{n_T n_T}(f) = N_0 f_g \cdot \frac{1}{2f_g} \text{rect}\left[\frac{f}{2f_g}\right] = \frac{N_0}{2} \text{rect}\left[\frac{f}{2f_g}\right]$	<p>Spektrale Leistungsdichte des Rauschanteils einer der beiden bandbegrenzten Quadraturkomponenten</p> <p style="text-align: right;">(2-32)</p>
--	--

Weiterhin gilt:

$$|H(f)|^2 = \left| \text{rect}\left[\frac{f}{2f_g}\right] \right|^2 = \text{rect}\left[\frac{f}{2f_g}\right]$$

Mit dieser Bandbegrenzung des Tiefpaßfilters $H(f)$ gilt für die nicht bandbegrenzte spektrale Leistungsdichte $\phi_{n_T n_T}(f)$ unter Anwendung der Wiener-Lee-Beziehung und der hypothetisch am Eingang des Tiefpaß-Filters angenommene Rauschleistungsdichte \tilde{N}_0 :

$\phi_{n_T n_T}(f) = \tilde{N}_0 \operatorname{rect}\left[\frac{f}{2f_g}\right] = \frac{N_0}{2} \operatorname{rect}\left[\frac{f}{2f_g}\right]$	<p>Spektr. Leistungsdichte des Rauschanteils einer der beiden unbegrenzten Quadraturkomponenten</p> <p style="text-align: right;">(2-33)</p>
---	--

Durch Vergleich läßt sich der Rückschluß auf die spektrale Leistungsdichte des Rauschsignals $n(t)$ ziehen: Da \tilde{N}_0 als Rauschleistungsdichte der Rauschprozesse der Quadraturkomponenten definiert wurde, ergibt sich somit als Eigenschaft des Rauschsignals $n(t)$:

$$\phi_{nn}(f) = N_0$$

Bei der nachfolgenden Abtastung der Quadraturkomponenten ergeben sich im Zeitbereich folgende Signale:

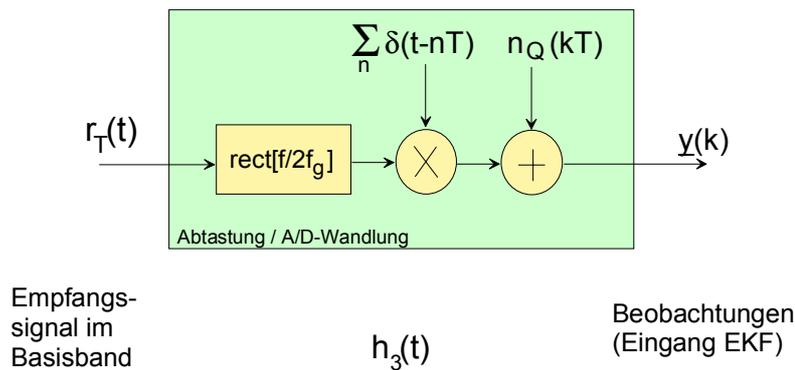


Abb. 2.14: Abtastung

Unter der Annahme einer feinen Quantisierung läßt sich der Quantisierungsfehler durch gleichverteiltes Quantisierungsrauschen $n_Q(kT)$ beschreiben, das in $n_{1,2}(kT)$ enthalten ist:

$$n_{1,2}(kT) = n_T(t) \cdot e^{-j2\pi f_0 t} \cdot \sum_k \delta(t - kT) + n_Q(kT)$$

$$\begin{aligned} \operatorname{Re}\{r_T(kT)\} &= a \cdot \cos(2\pi\gamma m_1(kT - t_0) + \varphi_0 - 2\pi f_0 t_0) + n_1(kT) \\ \operatorname{Im}\{r_T(kT)\} &= a \cdot \sin(2\pi\gamma m_1(kT - t_0) + \varphi_0 - 2\pi f_0 t_0) + n_2(kT) \end{aligned} \tag{2-34}$$

Durch die Abtastung mit Nyquist-rate entsteht aus dem nicht mehr weißen Rauschvektor $\underline{n}_T(t)$ ein zeitdiskreter, weißer Rauschvektor $\underline{n}_T(kT)$. Die Kovarianz lautet nun:

$$\text{cov}(\underline{n}_T(kT), \underline{n}_T(jkT)) = \begin{bmatrix} \frac{1}{2} f_{\Delta} N_0 \delta_{k,j} & 0 \\ 0 & \frac{1}{2} f_{\Delta} N_0 \delta_{k,j} \end{bmatrix} \quad (2-35)$$

$f_{\Delta} = 2f_g$ ist die Bandbreite des Ausgangssignals. Wir erhalten die Delta-Funktion über der Definition des weißen Rauschens als eine Ableitung eines Brown'schen Prozesses. Die Varianz des Brown'schen Prozesses ist linear, da eine Integration von Zufallswerten erfolgt, die eine konstante Varianz (Diffusionskonstante) besitzen. Nach Ableitung ergibt sich eine konstante Varianz, was zu der Delta-Funktion als Autokovarianz des Rauschprozesses führt (s. 2.2.1). Der Autokovarianzkern von $n_T(kT)$ ist selbst im Nullpunkt endlich. Weißes, zeitdiskretes Rauschen besitzt eine endliche Leistung, ist also auch existierbar im Gegensatz zu zeitkontinuierlichem weißem Rauschen.

In den folgenden beiden Funktionsblöcken findet die Demodulation als nichtlineare digitale Signalverarbeitung im Extended Kalman-Filter und die Rekonstruktion des zeitkontinuierlichen Signals statt. Dies läßt sich nicht als Übertragungsfunktion oder Stoßantwort beschreiben. Daher werden die letzten drei Funktionsblöcke von Abb. 2.10 (Abtastung, Verarbeitung, Rekonstruktion) zusammengefaßt betrachtet. Ein derartiges nichtlineares Systemmodell (s. 2.4.3.2) läßt sich allgemein durch sog. Volterra-Reihen darstellen. Eine Aufspaltung in Teilsysteme kann durch das sog. Wienermodell beschrieben werden [69]: Dazu stellt man sich eine Reihenschaltung zweier Übertragungsfunktionen vor. Die erste Übertragungsfunktion ist in diesem Fall linear und ein System mit "Gedächtnis" (Memory System) und das zweite System stellt ein nichtlineares System ohne Gedächtnis (No Memory System) dar, um die nichtlinearen stochastischen Eigenschaften zu beschreiben.

Das hier betrachtete Kalman-Filtermodell kann je nach Betrachtungsweise ebenfalls als Wienermodell (über den Innovationsansatz) interpretiert werden.

Die Übertragungsfunktion kann über die Kanalkapazität (s. 2.1.5) bestimmt werden, die sich über die estimationstheoretische Betrachtung der Information (s. 2.5.1) beschreiben läßt.

Die Rekonstruktion des analogen Quellsignals $x(t)$ läßt sich auch einzeln wieder mit einer Übertragungsfunktion oder Stoßantwort beschreiben:

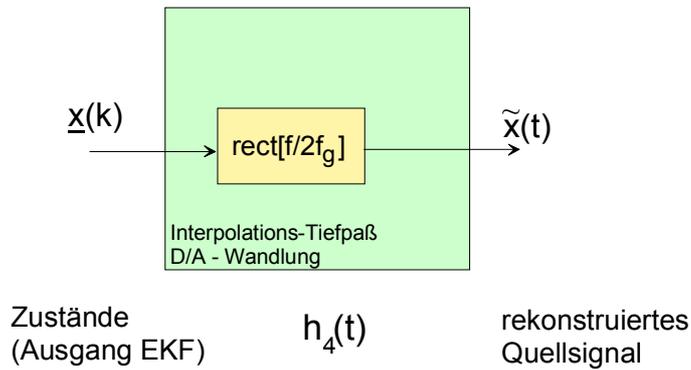


Abb. 2.15: Rekonstruktion

Bei verlustfreier Übertragung des Signals gilt $H_{\text{Ges}}(f) = 1 \cdot \text{rect} \frac{f}{2f_g}$ (s. Abb. 2.10), was jedoch aufgrund der nichtlinearen Eigenschaften der Winkelmodulation/-demodulation nicht möglich sein wird.

2.1.4 SNR-Definition

Das in dieser Arbeit verwendete und angegebene Signal-Rauschleistungsverhältnis SNR wird auch als CNR (Carrier to Noise Ratio) bezeichnet. Es bezeichnet das Verhältnis der Trägerleistung zur Rauschleistung im Bandpaßbereich /3, 44, 45/:

$$\frac{P_S}{P_N} = \frac{\frac{1}{2} a_0^2}{2f_{\Delta} N_0} = \text{SNR} = \text{CNR} \quad (2-36)$$

Durch Transformation in den äquivalenten Tiefpaßbereich ergeben sich am Ausgang halbierte Spannungswerte. Da sich dies jedoch auch auf das Rauschen auswirkt, ändert sich die SNR-Definition bezogen auf beide Quadraturkomponenten nicht:

$$\text{BP-Bereich} \quad \text{SNR}_{\text{reell}} = \frac{\frac{1}{2} a_0^2}{2f_{\Delta} N_0} = \frac{\frac{1}{2} a_0^2 \frac{1}{4}}{\frac{1}{4} 2f_{\Delta} N_0} = \text{SNR}_{\text{komplex}} \quad \text{Äquiv. TP-Bereich}$$

Zusätzlich läßt sich noch ein SNR der beiden einzelnen Quadraturkomponenten betrachten. Es läßt sich dabei zeigen, daß folgendes gilt:

$$\text{CNR} = \text{SNR}_{\text{Im}} + \text{SNR}_{\text{Re}} \quad \text{mit:} \quad \text{SNR}_{\text{Re}} = \frac{E\left\{\frac{a_0^2}{4} \cos^2(\gamma\varphi(t))\right\}}{\frac{1}{2} f_{\Delta} N_0} \quad (2-37)$$

Die beiden SNR's werden im Normalfall unterschiedlich sein, jedoch in der Summe immer konstant sein.

2.1.5 Kanalkapazität

Ein Maß zur Beschreibung des Übertragungskanals ist die Kanalkapazität C . Sie beschreibt die verarbeitbare Informationsmenge des Übertragungskanals: Sie wird ursprünglich als Maximum der Entropie, der maximal übertragenen Information (Transinformation) definiert [3, 44, 45].

$$C = \max_{P(x)} H(x; y) \quad (2-38)$$

Die Transinformation kann folgendermaßen aufgelöst werden:

$$H(x; y) = H(y) - H(y | x) \quad (2-39)$$

Setzt man Shannon's Entropie an, ergibt sich für einen Gaußkanal mit gaußverteilten Größen (s.2.5.1):

$$S = - \int_x \ln(f_x(\xi)) \cdot f_x(\xi) \cdot d\xi = \frac{1}{2} \cdot \ln[(2\pi) \cdot |P|] + \frac{1}{2}$$

Die betrachteten Leistungen sind am Ausgang y die Summe aus Signalleistung und Rauschleistung der Störgröße.

$$\begin{aligned} H(y) = S_y &= - \int_y \ln(f_y(\rho)) \cdot f_y(\rho) \cdot d\rho \\ &= \frac{1}{2} \cdot \ln[(2\pi) \cdot |S + N|] + \frac{1}{2} \end{aligned} \quad (2-40)$$

Die Streuentropie wird durch die gaußverteilte, additiv überlagerte, unabhängige Störgröße beschrieben:

$$H(y | x) = S_{y|x} = - \int_y \ln(f_{y|x}(\rho | \xi)) \cdot f_{y|x}(\rho | \xi) \cdot d\rho$$

$$\begin{aligned}
&= -\int_n \ln(f_n(\eta)) \cdot f_n(\eta) \cdot d\eta \\
&= \frac{1}{2} \cdot \ln[(2\pi) \cdot |N|] + \frac{1}{2}
\end{aligned} \tag{2-41}$$

Damit ergibt sich für die Transformation:

$$\begin{aligned}
H(x; y) &= H(y) - H(y|x) = \frac{1}{2} \cdot \ln[(2\pi) \cdot |S+N|] + \frac{1}{2} - \frac{1}{2} \cdot \ln[(2\pi) \cdot |N|] - \frac{1}{2} \\
&= \frac{1}{2} \cdot \ln[|S+N|] - \frac{1}{2} \cdot \ln[|N|] + \ln(2\pi) - \ln(2\pi) \\
&= \frac{1}{2} \cdot \ln\left[\frac{|S+N|}{|N|}\right] \\
&= \frac{1}{2} \cdot \ln\left[1 + \frac{S}{N}\right] = \frac{1}{2} \cdot \ln[1 + \text{SNR}]
\end{aligned}$$

Somit besteht ein direkter Zusammenhang zum SNR-Verhältnis.

Der Gaußkanal wird aber zusätzlich definiert als idealer Übertragungskanal mit einem idealen Tiefpaß der Grenzfrequenz f_B oder idealem Bandpaß der Bandbreite $f_\Delta = f_B$ und einer gaußförmig verteilten, weißen additiven Störgröße der Rauschleistung N_0 ($N = 2f_B N_0$):

$$H(x; y) = \frac{1}{2} \cdot \ln\left(1 + \frac{S}{2f_B N_0}\right) \tag{2-42}$$

Zusätzlich dazu läßt sich eine zeitbezogene Kanalkapazität C^* definieren. Sie läßt sich aus einer ebenfalls zeitbezogenen Entropie bestimmen:

$$H^*(x; y) = 2f_g \cdot H(x; y) \text{ in Bit/Zeiteinheit}$$

Durch Einsetzen erhält man die zeitbezogene Kanalkapazität mit:

$$H^*(x; y) = C^* = 2f_B \cdot \frac{1}{2} \cdot \ln\left(1 + \frac{S}{2f_B N_0}\right) \tag{2-43}$$

Betrachtet man rein binäre Größen, ergibt sich für die zeitbezogene Kanalkapazität:

$$C^* = f_B \cdot \text{ld} \left(1 + \frac{S}{N} \right) = f_B \cdot \text{ld} \left(1 + \frac{S}{2f_B N_0} \right) \quad \text{in Bit/Zeiteinheit} \quad (2-44)$$

Für einen nicht bandbegrenzten Gaußkanal ($f_B \rightarrow \infty$) erhalten wir für die Kanalkapazität keine unendliche Kanalkapazität, sondern:

$$C_\infty^* = \lim_{f_B \rightarrow \infty} f_B \cdot \text{ld} \left(1 + \frac{S}{N} \right) = \frac{\text{ld}(e)}{2} \cdot \frac{S}{N_0} \cong 0,72 \cdot \frac{S}{N_0} \quad (2-45)$$

Daraus folgt aber zugleich, daß bei einer gewählten Kanalkapazität C_∞^* eine Mindestleistung für S erforderlich wird. Setzt man eine Zeitdauer zur Übertragung eines binären Wertes von $T = \frac{1}{C_\infty^*}$ an, dann läßt sich eine Energie bestimmen, und somit erscheint auch ein erforderliches minimales Signalrauschleistungsverhältnis $\frac{E}{N_0} = \frac{2}{\text{ld}(e)} = 1,39 \hat{=} 1,42\text{dB}$. Diese Grenze wird auch als Shannon-Grenze bezeichnet /45/.

2.2 STOCHASTISCHE PROZESSE

Zum Grundverständnis stochastischer Filter soll hier insbesondere ansatzweise auf Verteilungsfunktionen, Zufallszahlen, Erwartungswerte und Markov-Prozesse eingegangen werden. Diese werden dann im weiteren Verlauf vorausgesetzt.

In mathematischer Betrachtungsweise beschreibt ein stochastischer Prozeß $\underline{x}(\cdot, \cdot)$ ein Ensemble von Musterfunktionen $\underline{x}(\cdot, \omega_j)$, die zeitliche Verläufe der Realisationen von Elementarereignissen ω_j darstellen, die wiederum in einem eigens dafür definierten, fiktiven Ereignisraum Ω stattgefunden haben. Betrachten wir zu einem Zeitpunkt t_i die Musterfunktionen $\underline{x}(\cdot, \omega_j)$, so bezeichnen wir den Vektor, der dadurch entsteht, als Zufallsvektor $\underline{x}(t_i, \cdot)$ oder kurz als $\underline{x}(t_i)$. Durch einen dritten Bezugsraum, die Wahrscheinlichkeit, definieren wir die Auftretenswahrscheinlichkeit der Musterfunktionen, die unter einem definierten Schrankenwert $\underline{\xi}$ liegen, als eine Verteilungsfunktion $F_{\underline{x}(t)}(\cdot)$ an der Stelle $F_{\underline{x}(t)}(\underline{\xi})$. Die Ableitung der Verteilungsfunktion $F_{\underline{x}(t)}(\cdot)$ ergibt die Verteilungsdichtefunktion $f_{\underline{x}(t)}(\cdot)$ /38, 39/.

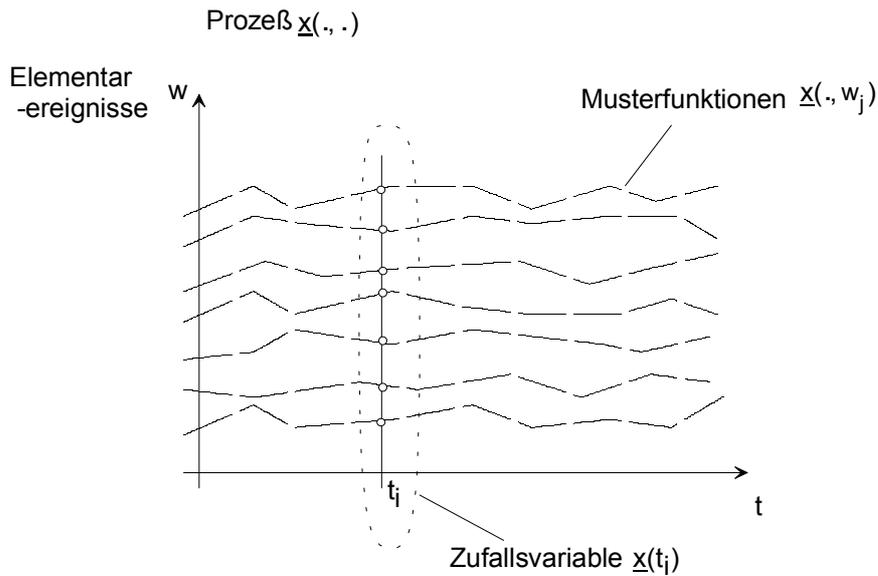


Abb. 2.16: Ereignisraum

Jedem Ereignis als Teilmenge des Ereignisraumes Ω wird dabei praktischerweise eine Zahl aus der Menge der reellen Zahlen zugewiesen. Bei der Zuweisung auf reelle Zahlen sprechen wir von sogenannten Zufallsvariablen $x(\cdot)$, bei der $x(\omega)$ der Abbildung von ω auf die reelle Zahlenachse entspricht. Dies läßt sich für skalare Zufallsvariablen $x(\cdot)$ ebenso wie für vektorielle Zufallsvariablen $\underline{x}(\cdot)$ definieren. Die Zuweisung des Ereignisses wird auch Realisation x_r genannt:

$$x_r = x(\omega) \text{ oder für Vektoren: } \underline{x}_r = \underline{x}(\omega) \quad A_i = \{\omega : \underline{x}(\omega) \in A\}$$

Als ω werden hier alle Elementarereignisse bezeichnet, die mit der Realisation x_r beschrieben werden können. Aus Gründen der Allgemeinheit sollen im folgenden ausschließlich vektorielle Größen behandelt werden.

In \underline{x}_r sind alle Abbildungen von Elementarereignissen zusammengefaßt, die sich im definierten Intervall befinden. In mathematischer Formulierung lauten die Teilmengen A_i im Realisationsraum:

$$A_i = \{\omega : \underline{x}(\omega) \leq \underline{\xi}\} \quad \text{Borel-Feld}$$

In den Teilmengen A_i sind die Elementarereignisse ω zusammengefaßt, die als Realisationen $\underline{x}(\omega)$ unterhalb einer Grenze $\underline{\xi}$, die das halboffene Intervall $I = (0, \underline{\xi}]$ beschreibt,

liegen. Damit läßt sich die Wahrscheinlichkeitsverteilungsfunktion $F_x(\xi)$ oder Verteilungsfunktion genannt beschreiben:

$$F_x(\xi) = P\{\omega: \underline{x}(\omega) \leq \xi\} \quad (2-46)$$

Sie beschreibt die Wahrscheinlichkeit, daß Elementarereignisse im Bereich des vorgegebenen Intervalls mit der Obergrenze ξ liegen. Der Wertebereich einer Verteilungsfunktion liegt immer zwischen 0 und 1. Die Ableitung der Verteilungsfunktion $F_x(\xi)$ wird als Wahrscheinlichkeitsverteilungsdichtefunktion $f_x(\xi)$ oder Verteilungsdichte bezeichnet. Hat die Verteilungsfunktion Unstetigkeiten in ihrem Verlauf, so muß diese in Teilintervalle zur Bildung der Verteilungsdichte zerlegt werden.

Kalman-Filter basieren auf der Betrachtung gaußverteilter Größen:

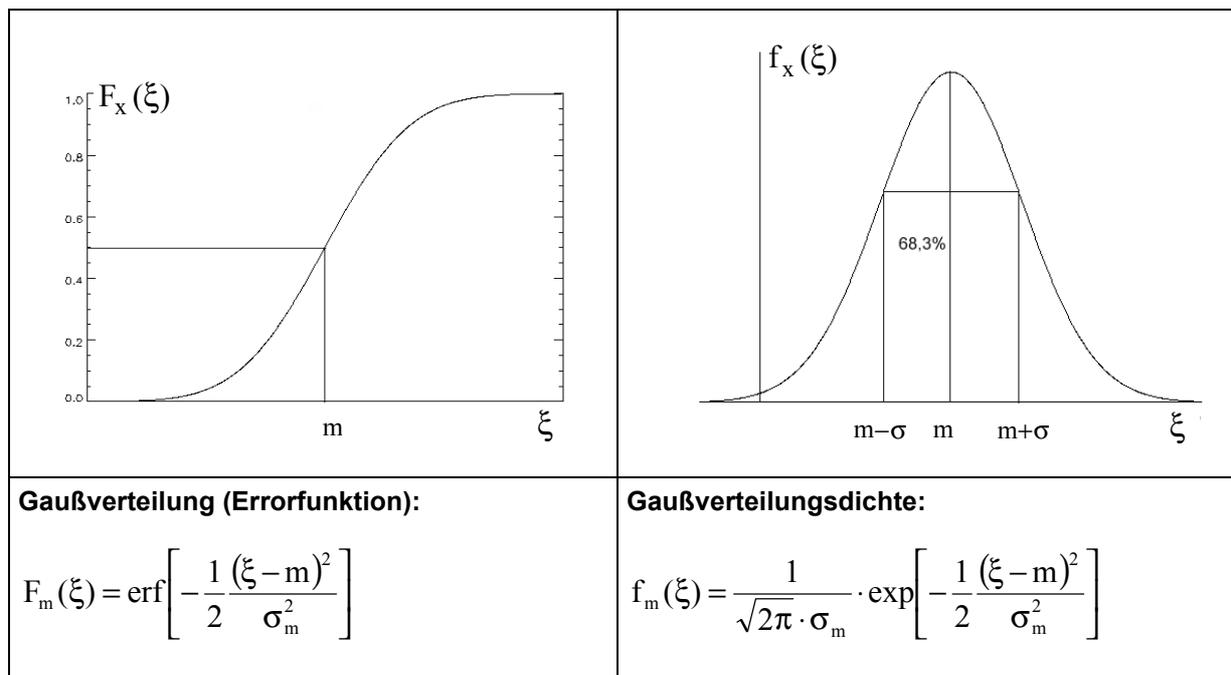


Abb. 2.17: Gaußverteilungsfunktion

Eigenschaften der Zufallsvariablen:

Die Verteilungsfunktion und die Verteilungsdichte einer Zufallsvariablen sind zu deren Beschreibung von elementarer Bedeutung, da sie die einzige Möglichkeit eröffnen, Erwartungswert und Varianz (quadratischer Erwartungswert) einer Zufallsvariablen zu bestimmen. Diese Werte, auch als 1. und 2. Moment bekannt, bieten die einzige Möglichkeit zur Beschreibung eines Rauschsignals:

$$E[\underline{x}] = \int_{-\infty}^{\infty} \underline{\xi} \cdot f_{\underline{x}}(\underline{\xi}) d\underline{\xi} = \underline{m} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} \quad (2-47)$$

Der Erwartungswert stellt gleichzeitig den Mittelwert oder das 1. Moment einer Zufallsvariablen \underline{x} dar.

Zusätzlich läßt sich die quadratische Standardabweichung oder Varianz mit Hilfe der Verteilungsdichte der Zufallsvariablen \underline{x} bestimmen. In diesem Fall wird der Mittelwert \underline{m} mit einbezogen, man spricht vom 2. zentralen Moment (die Varianz wird um den Mittelwert "zentriert") oder auch von der Autokovarianz:

$$E[\underline{x}\underline{x}^T] = \int_{-\infty}^{\infty} (\underline{\xi} - \underline{m}) \cdot (\underline{\xi} - \underline{m})^T \cdot f_{\underline{x}}(\underline{\xi}) d\underline{\xi} = P_{\underline{x}\underline{x}} = \sigma_{\underline{x}}^2 \cdot I \quad (2-48)$$

Wird der Mittelwert nicht mit einbezogen, wird dies als Autokorrelation oder 2. nicht-zentrales Moment von \underline{x} bezeichnet:

$$E[\underline{x}\underline{x}^T] = \int_{-\infty}^{\infty} \underline{\xi}\underline{\xi}^T \cdot f_{\underline{x}}(\underline{\xi}) d\underline{\xi} = \Psi_{\underline{x}\underline{x}} \quad (2-49)$$

Häufig sind die betrachteten Zufallsvariablen erwartungswertfrei, so daß Autokovarianz und Autokorrelation identisch sind.

$$I \cdot \sigma_{\underline{x}}^2 = \Psi_{\underline{x}\underline{x}} = P_{\underline{x}\underline{x}}$$

Eine wichtige Beziehung zwischen dem 2. zentralen und dem 2. nicht zentralen Moment bei rein skalaren Zufallsvariablen lautet:

$$P_{xx} = E[x^2] - (E[x])^2 \quad (2-50)$$

Es können noch weitere höhere Momente bestimmt werden. Sie geben Aufschluß über Symmetrieeigenschaften der Zufallsvariablen und werden jedoch eher selten eingesetzt. Ist z.B. die Verteilungsdichte einer Zufallsvariablen gaußförmig, so wird sie vollständig durch ihr erstes und zweites Moment beschrieben.

Wird eine zeitdiskrete Musterfunktion betrachtet, so erhalten wir mit der Definition auch einen diskreten, stochastischen Prozeß $\underline{x}(\cdot, \cdot)$. Problematisch wird die Definition von zeitkontinuierlichen Prozessen. Um einen zeitkontinuierlichen, stochastischen Prozeß zu erhalten, muß ein "Umweg" über den sogenannten Ereignisraum Ω gewählt werden. Durch die nichtendliche Anzahl von Zwischenwerten einer kontinuierlichen Funktion in einem endlichen Intervall können wir keine direkte Realisation als mathematische Funktion angeben. Der "Umweg" über den Ereignisraum Ω wird gewählt, um eine unendliche Anzahl von Elementarereignissen ω_j , wie sie beispielsweise bei Wärmerauschspannungswerten eines Widerstandes auftreten, in Teilmengen $\underline{x}(\cdot, \omega_j)$ als Ereignisse mit endlichen Grenzen zusammenfassen zu können.

2.2.1 Gauß-Markov-Prozeß

Betrachten wir zunächst einen speziellen Prozeß $x(\cdot, \cdot)$, den sogenannten Brown'sche Prozeß $\beta(\cdot, \cdot)$ oder Wiener-Prozeß. Er wurde vom Botaniker R. Brown (Molekularbewegung in Flüssigkeiten) erstmals Anfang des 19. Jahrhunderts beobachtet und von dem Mathematiker N. Wiener über 100 Jahre später mathematisch beschrieben.⁴ Der Brown'sche Prozeß stellt die Integration eines stochastischen Prozesses $x(\cdot, \cdot)$ dar. Bei der Betrachtung von $\beta(\cdot, \cdot)$ müssen die Inkremente δ mittelwertfrei, gaußverteilt, weiß und statistisch voneinander unabhängig sein:

Der Erwartungswert des Prozesses $\beta(\cdot, \cdot)$ zum Prozeßzeitpunkt t_i , das 1. Moment der Zufallsvariablen $\beta(t_i, \cdot)$ ist gleich Null: $E[\beta(t_i, \cdot)] = 0$. Das zweite Moment, die Varianz des Prozesses ist gekennzeichnet durch einen stetigen, linearen Anstieg: $E[\beta(t_i, \cdot) \cdot \beta(t_i, \cdot)] = q \cdot |t_i - t_0|$. Dabei ist q der sogenannte Diffusionsparameter, der die Varianz zwischen zwei aufeinanderfolgenden Zeitpunkten beschreibt. Die Bestimmung der Kreuzkovarianz zweier Brown'scher Zufallsvariablen führt aufgrund der Erwartungswertfreiheit und der statistischen Unabhängigkeit der Zufallsvariablen zum Wert der Varianz.

⁴Der Mathematiker P. Levy hat den gleichen Prozeß mathematisch anders beschrieben, so daß man auch von Levy-Prozeß oder Wiener-Levy-Prozeß spricht.

Es läßt sich zeigen, daß jeder Prozeß mit unabhängigen Inkrementen ein Markov-Prozeß ist. Zur vollständigen Beschreibung eines stochastischen Prozesses wäre eine vollständige Kenntnis aller Verteilungsdichten zu jedem Prozeßzeitpunkt nötig. Es ist jedoch nicht absolut notwendig, diese vollständige Kenntnis über den Prozeß, d.h. über unendliche viele Verteilungsfunktionen, zu besitzen, da nur die Markov-Eigenschaft gefordert ist. Ist die Verteilungsdichte $F_{\underline{\beta}(t_i)}(\xi)$ einer Zufallsvariablen $\underline{\beta}(t_i, \cdot)$ zum Prozeßzeitpunkt t_i eine bedingte Verteilungsfunktion $F_{\underline{\beta}(t_i)|\underline{\beta}(t_{i-1}),\underline{\beta}(t_{i-2}),\dots,\underline{\beta}(t_k)}(\xi|\underline{b}(t_{i-1}),\underline{b}(t_{i-2}),\dots,\underline{b}(t_k))$, bedingt auf endlich viele, vorhergehende Verteilungsfunktionen, so wird der Prozeß als Markov-Prozeß bezeichnet. Allgemein ist jeder Prozeß mit einem sogenannten "endlichen Gedächtnis" ein Markov-Prozeß. Damit stellt der Brown'sche Prozeß einen speziellen Markov-Prozeß dar /5, 38/.

Sind alle Zufallsvariablen eines stochastischen Prozesses gaußverteilt, wird dieser Prozeß Gauß-Prozeß genannt. Ist er gleichzeitig auch ein Markov-Prozeß, d.h. er besitzt die Markov-Eigenschaft eines endlichen Gedächtnisses, dann wird der Prozeß $x(\cdot, \cdot)$ Gauß-Markov-Prozeß genannt.

Die oben erwähnten Inkremente errechnen sich aus der Differenz zweier aufeinanderfolgender Realisationen: $\delta_k = \delta_k(\omega_n) = \beta(t_k, \omega_n) - \beta(t_{k-1}, \omega_n)$. Die Realisationen $\beta(t_k, \omega_n)$ sind Realisationen einer Musterfunktion $\beta(\cdot, \omega_n)$, die durch das Ereignis ω_n bestimmt werden. Daraus ergibt sich umgekehrt für die Realisationen:

$$\beta(t_k, \omega_n) = \beta(t_{k-1}, \omega_n) + \delta_k(\omega_n)$$

Für den bedingten Erwartungswert $E[\beta(t_k)|\beta(t_{k-1}) = b_{k-1}]$ auf die vorangegangene Zufallsvariable ergibt sich daraus:

$$\begin{aligned} E[\beta(t_k)|\beta(t_{k-1}) = b_{k-1}] &= E[\beta(t_{k-1})|\beta(t_{k-1}) = b_{k-1}] + \underbrace{E[\delta(t_k)|\beta(t_{k-1}) = b_{k-1}]}_{=0} \\ &= E[\beta(t_{k-1})|\beta(t_{k-1}) = b_{k-1}] \\ &= b_{k-1} \end{aligned} \tag{2-51}$$

Verallgemeinert für einen vollständigen Prozeß können wir festhalten: Ist der Erwartungswert gleich der Realisation der vorangegangenen Zufallsvariablen, dann sprechen wir zusätzlich von einem Martingal-Prozeß. Der Brown'sche Prozeß ist also zusätzlich ein Martingal-Prozeß.

2.2.2 Weißes, gaußverteiltes Rauschen

Ein ideales weißes, gaußverteiltes Rauschen kommt weder in der Natur vor, noch kann es mathematisch aus einem Brown'schen Prozeß durch Differentiation hergeleitet werden. Aufgrund seiner Vorteile in der Systemtheorie bei der Beschreibung von mathematisch, idealisierten Systemen wird das weiße, gaußverteilte Rauschen standardmäßig verwendet. Eine praktische Umsetzung von idealem, weißen Rauschen als Software ist ebenfalls nicht realisierbar. Die Betrachtung von idealisiertem, weißen Rauschen stellt eine mathematische zulässige Idealisierung dar, da alle Systeme Bandpaßcharakter haben.

Da jedoch alle betrachteten Systeme bandbegrenzte Systeme sind, stellt ein bandbegrenztetes Rauschen mit einer größeren Bandbreite als der des Systems für das betrachtete System ein ideales, weißes Rauschen dar. Es hängt folglich nur vom Betrachter ab.

Zur Definition: Obwohl mathematisch keine Differenzierbarkeit eines Brown'schen Prozesses existiert, wird sie formal trotzdem berechnet (da vorstellbar). Die Zufallsvariable $w(t, \cdot)$ stellt das weiße, gaußverteilte Rauschen dar:

$$w(t, \cdot) = \frac{d}{dt} \beta(t, \cdot) \quad \text{bzw.} \quad \beta(t, \cdot) = \int_{t_0}^t w(\tau, \cdot) d\tau \quad (2-52)$$

Aus der Definition des Brown'schen Prozesses ergibt sich ein Inkrement als Differenz zweier Integrale:

$$\delta(k) = \beta(t_k, \cdot) - \beta(t_{k-1}, \cdot) = \int_{t_0}^{t_k} w(\tau, \cdot) d\tau - \int_{t_0}^{t_{k-1}} w(\tau, \cdot) d\tau = \int_{t_{k-1}}^{t_k} w(\tau, \cdot) d\tau \quad (2-53)$$

Weißes Rauschen ist folglich mittelwertfrei, d.h. sein Erwartungswert ist gleich Null, da die Inkremente bereits als erwartungswertfrei definiert wurden.

$$E[w(t, \cdot)] = \bar{w}(t, \cdot) = 0 \quad (2-54)$$

Die Kovarianzen (2. Momente) ergeben sich ohne Beweis zu:

$$E[w(t, \cdot) \cdot w(t', \cdot)] = ?$$

$$\Rightarrow \text{Kreuzkovarianz:} \quad E[w(t, \cdot) \cdot w(t', \cdot)] = 0 \quad \text{für } t \neq t'$$

∧ Autokovarianz: $E[w(t, \cdot) \cdot w(t', \cdot)] = q \cdot \delta(t - t')$ für $t = t'$

⇒ Allgemein gilt folglich für die Kovarianz: $E[w(t, \cdot) \cdot w(t', \cdot)] = q \cdot \delta(t)$ (2-55)

Der Beweis läßt sich erneut über die Definition des Brown'schen Prozesses führen. Mit den ersten beiden Momenten ist das weiße Rauschen aufgrund seiner Gaußverteilttheit somit vollständig beschrieben.

Die Gaußverteilttheit gilt als realistische Annahme für die meisten Rauschprozesse. Aufgrund des zentralen Grenzwertsatzes ergibt sich bei Überlagerung einer großen Zahl unterschiedlicher Rauschprozesse mit unterschiedlichen Verteilungsfunktionen eine Gaußverteilung als Gesamtverteilungsfunktion. Warum wird eine Gaußverteilttheit angestrebt? Gaußprozesse lassen sich mathematisch besser behandeln: Eine Gaußverteilungsfunktion ist schon durch ihre ersten zwei Momente (Erwartungswert und Kovarianz) vollständig beschrieben.

2.2.2.1 Diskretes weißes Rauschen

Bei der Verarbeitung von weißem, gaußverteiletem Rauschen muß bei der Simulation, aber auch bei der Online - Verarbeitung auf zeitdiskrete Meßwerte zurückgegriffen werden. D.h. es werden ebenfalls zeitdiskrete Rauschwerte, die Realisationen von unabhängigen Zufallsvariablen darstellen, verarbeitet. Für die Simulation läßt sich weißes Rauschen trotzdem realisieren, da in einer Rechnersimulation zeitdiskrete Rauschwerte verarbeitet werden. Diese können beispielsweise auch durch Abtastung eines farbigen Rauschprozesses erzeugt werden.

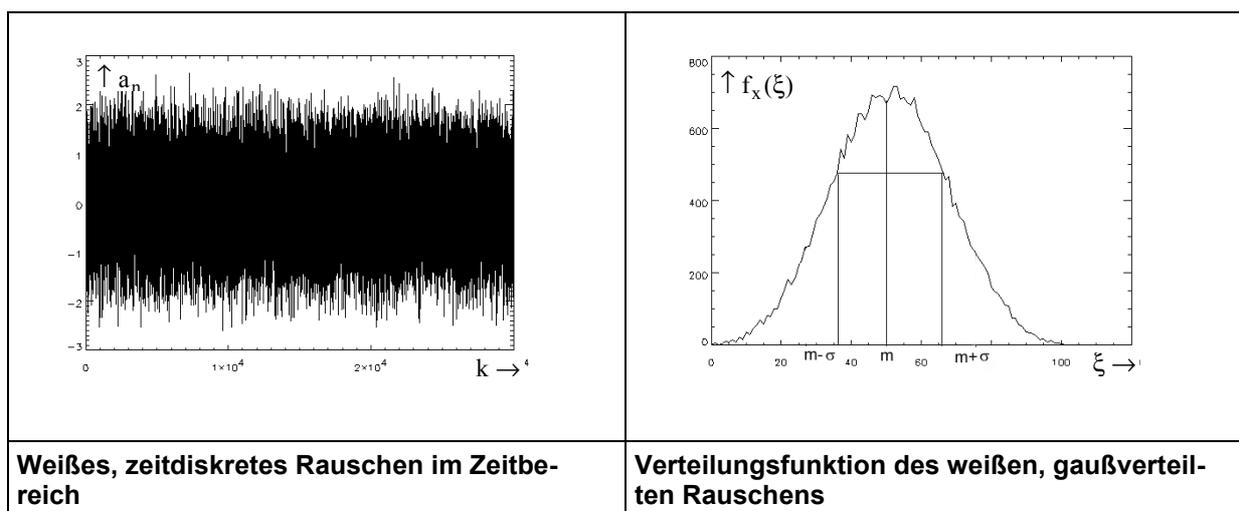


Abb. 2.18: Weißes Rauschen

Auf die Erzeugung von weißen, gaußverteilten Rauschwerten wird bei der Diskussion der Testfunktionen im Abschnitt 5.2.4 unter verschiedenen Programmiersprachen nochmals ausführlich eingegangen.

2.2.3 Farbige Rauschprozesse

Bandbegrenzte Rauschprozesse werden als farbige Rauschprozesse bezeichnet. Alle Frequenzen - von Null bis unendlich - sind bei weißem Rauschen gleichstark vertreten.

Das Leistungsdichtespektrum $\phi_{ww}(f)$, die Fouriertransformierte der Autokorrelationsfunktion $\varphi_{ww}(\tau)$ ist für weißes Rauschen konstant:

$$\phi_{ww}(f) = N_0 \quad \bullet - \circ \quad \varphi_{ww}(\tau) = N_0 \cdot \delta(\tau) = E[w(t, \cdot) \cdot w(t + \tau, \cdot)] \quad (2-56)$$

Die Diffusionskonstante q entspricht der Leistungsdichte N_0 . Die Leistung selbst ist definiert über den Ursprung der Autokorrelationsfunktion $\varphi_{ww}(\tau)$:

$$P = \varphi_{ww}(\tau = 0) = \int_{-\infty}^{+\infty} \phi_{ww}(f) df \quad (2-57)$$

Ein bandbegrenztes Rauschsignal wird auf die Systembandbreite begrenzt, so daß nur im Durchlaßbereich die Frequenzen gleich stark vertreten und im Restbereich gedämpft sind. Die Gaußverteilttheit des farbigen Rauschprozesses wird durch die Bandbegrenzung nicht verändert.

2.3 BANDBREITEBETRACHTUNGEN

Entscheidend für die Wahl eines Modulationsverfahrens ist ebenso dessen Bandbreite B . Im folgenden soll der Bandbreitebedarf von zeitkontinuierlichen und zeitdiskreten Signalen abgeschätzt werden.

Das Spektrum von winkelmodulierten Signalen für ein monofrequentes sinusförmiges Quellsignal ist beispielsweise durch die Entwicklung des Signals in Besselfunktionen bestimmt. Die relative Phasenlage ist dann abwechselnd positiv und negativ. Mit steigender Signalamplitude steigt die Anzahl der Seitenfrequenzen an. Es besteht offensichtlich eine starke

Abhängigkeit der Bandbreite des modulierten Signals und der Bandbreite des Modulationssignals. Deshalb wird zunächst die Bandbreite des Quellsignals bestimmt. Allgemein kann die Bandbreite als Varianz des normierten Leistungsdichtespektrums definiert werden /1, 15, 21, 23, 53/.

Allgemein ist das Problem der Berechnung des Spektrums winkelmodulierter Signale geschlossen nicht möglich. Ein spezielles Verfahren zur Spektralanalyse stellt die Middleton Expansion dar. Bei diesem Verfahren wird das Spektrum in eine Reihe von Teilspektren zerlegt. Diese Teilspektren berechnen sich aus den Faltungsvielfachen des Leistungsdichtespektrums. Die Middleton Expansion zeigt, daß das Spektrum des modulierten Signals dem des Modulationssignals für kleine Signalleistungen sehr ähnelt; für mittlere Leistungen P_x sich die Spektrallinien gaußförmig ausprägen und bei großer Signalleistung P_x die gesamte Form des Spektrums des modulierten Signals gaußförmig ist.

Bei der Betrachtung von unterschiedlichen Nutzsignalen variiert auch deren Bandbreite. Um eine allgemeine Behandlung zu erreichen, bleibt nur die Möglichkeit, die Bandbreite möglichst allgemein abzuschätzen. Dazu wird ein gaußverteiltes Rauschsignal angenommen, was einer Beschreibung eines allgemeinen, beliebigen Nutzsignals entspricht. Es wird dabei zusätzlich angenommen, daß sich das Rauschsignal periodisch wiederholt, so daß eine Fourierreihenanalyse möglich wird. Die Fourierkoeffizienten entsprechen dann den Leistungen der Spektrallinien, die sich bei großem Modulationsindex als normalverteilt herausstellen. Als komplexe Reihenentwicklung betrachtet, läßt sich dies auch folgendermaßen darstellen:

$$s_T(t) = e^{j2\pi\gamma m(t)} = 1 + \frac{j2\pi\gamma}{1!} \cdot m(t) + \frac{[j2\pi\gamma]^2}{2!} \cdot m^2(t) + \frac{[j2\pi\gamma]^3}{3!} \cdot m^3(t) + \dots$$

(2-58)

○ — ●

$$S_T(f) = \delta(f) + 2\pi\gamma \cdot M(f) \cdot e^{j\frac{\pi}{2}} + 2\pi^2\gamma^2 \cdot M(f) * M(f) \cdot e^{j\frac{2\pi}{2}} + \frac{4}{3} \pi^3\gamma^3 \cdot M(f) * M(f) * M(f) \cdot e^{j\frac{3\pi}{2}} + \dots$$

Bei hohem Modulationsindex γ fallen die Faltungsvielfachen von $M(f)$ stärker ins Gewicht, womit sich eine Gaußverteilung ausprägt. Die Bandbreite B läßt sich in diesem Fall als Varianz einer Gaußverteilung beschreiben /53/. Der Verteilungsdichtefunktion entspricht hierbei dem normierten Leistungsdichtespektrum $\tilde{\phi}_{ss}(f)$ des modulierten Signals $s(t)$.

$$B^2 = \int_{-\infty}^{+\infty} (f - f_0)^2 \tilde{\phi}_{ss}(f) df \quad (2-59)$$

2.3.1 Bandbreitebeziehungen bei Winkelmodulation

Generell entspricht der Bandbreitebedarf bei ASK theoretisch der Nyquistrate. BPSK hat einen gleich großen Bandbreitebedarf - der Rauschabstand kann jedoch um 6dB geringer sein als bei ASK. Bei QPSK sinkt der Bandbreitebedarf auf die Hälfte ab. Die Störsicherheit nimmt zwar ab, jedoch ist die Bandbreiteneffizienz erheblich größer. Die MSK benötigt die gleiche Bandbreite wie QPSK. In der Praxis muß grundsätzlich um einen Faktor 1.4 - 2.0 mehr Bandbreite - als theoretisch notwendig - eingesetzt werden.

Die effektive Bandbreite von einem PM-Signal berechnet sich aus dem Produkt aus effektiver Bandbreite von Quellsignal und Wurzel aus der Signalleistung. Dagegen ist die effektive Bandbreite eines FM-Signals direkt proportional zur Wurzel aus der Leistung des abgeleiteten Signals. Die genannten Bandbreitebeziehungen für PM- und FM-Signale werden im folgenden hergeleitet. Dazu wird die Definition der Bandbreite als Varianz des Leistungsdichtespektrums verwendet. Diese Definition benutzt das 2. Moment der normierten Spektralleistungsdichte als ein Maß der Frequenzausbreitung. Die Betrachtung ist bereinigt um die Trägerfrequenz f_0 , d.h. wir betrachten das Bandpaßsignal im äquivalenten Tiefpaßbereich:

Ausgehend von (2-59) gilt:

$$B^2 = \int_{-\infty}^{\infty} f^2 \tilde{\phi}_{yy}(f) df$$

$$\text{mit: } \tilde{\phi}_{yy}(f) = \frac{\phi_{yy}(f)}{\int_{-\infty}^{+\infty} \phi_{yy} df} = \frac{\phi_{yy}(f)}{P_y} \quad (2-60)$$

Im folgenden wird die Formel der Fouriertransformation des Leistungsdichtespektrums $\phi_{yy}(f)$ dazu verwendet, die Bandbreite B genauer zu bestimmen; $\phi_{yy}(\tau)$ stellt die Autokorrelationsfunktion der Zeitfunktion $y(t)$ dar:

$$\begin{aligned}\varphi_{yy}(\tau) &= \int_{-\infty}^{\infty} \phi_{yy}(f) e^{j2\pi f\tau} df \\ \varphi'_{yy}(\tau) &= \int_{-\infty}^{\infty} j2\pi f \cdot \phi_{yy}(f) e^{j2\pi f\tau} df \\ \varphi''_{yy}(\tau) &= \int_{-\infty}^{\infty} (-4\pi^2 f^2) \cdot \phi_{yy}(f) e^{j2\pi f\tau} df\end{aligned}$$

Betrachten wir die 2. Ableitung der Autokorrelationsfunktion im Nullpunkt, dann ergibt sich daraus für die Bandbreite bei normierten Größen:

$$\begin{aligned}\varphi''_{yy}(\tau=0) &= \int_{-\infty}^{+\infty} -4\pi^2 f^2 \cdot \tilde{\phi}_{yy}(f) \cdot P_y df = -4\pi^2 \cdot P_y \cdot B^2 \\ B^2 &= -\frac{\varphi''_{yy}(0)}{4\pi^2 \cdot P_y}\end{aligned}\tag{2-61}$$

Die Autokorrelationsfunktionen von Frequenz- und Zeitbereich müssen auch durch ihre normierten Größen ausgedrückt werden:

$$\begin{aligned}\varphi_{yy}(\tau) &= \tilde{\varphi}_{yy}(\tau) \cdot P_y \\ \Rightarrow B^2 &= -\frac{\tilde{\varphi}''_{yy}(\tau=0)}{4\pi^2}\end{aligned}\tag{2-62}$$

Es wird eine weitere Nebenrechnung notwendig, um die Bandbreite zu bestimmen: Die AKF des Tiefpaßsignals muß gebildet werden.

$$\begin{aligned}\varphi_{yy}(\tau) &= E\{y_T^*(t) \cdot y_T(t+\tau)\} \\ &= E\left\{e^{-j2\pi\gamma m(t)-j\varphi_0+j2\pi\gamma m(t+\tau)+j\varphi_0}\right\} \\ &= E\left\{e^{j2\pi\gamma(m(t+\tau)-m(t))}\right\} \\ &= E\left\{e^{j2\pi\gamma\Delta m(t,t+\tau)}\right\}\end{aligned}$$

Nimmt man voraussetzungsgemäß an, daß $m(t)$ einem Gaußprozeß entstammt, ist $\Delta m(t, t + \tau)$ das Inkrement zwischen zwei Zeitpunkten einer gaußverteilten Zufallsprobe. Die Verteilungsdichte ist gegeben durch:

$$f_{\Delta m}(\xi) = \frac{1}{\sqrt{2\pi} \cdot \sigma_{\Delta m}} \cdot \exp\left[-\frac{1}{2} \frac{\xi^2}{\sigma_{\Delta m}^2}\right]$$

Die charakteristische Funktion für $\Delta m(t, t + \tau)$ lautet:

$$\begin{aligned}
\phi_{\Delta m(\tau)}(u) &= E\{e^{j\Delta m(t,t+\tau)u}\} \\
&= \int_{-\infty}^{\infty} e^{ju\xi} \cdot f_{\Delta m}(\xi) d\xi
\end{aligned} \tag{2-63}$$

Die charakteristische Funktion kann jetzt berechnet werden:

$$\begin{aligned}
\phi_{\Delta m(\tau)}(u) &= \int_{-\infty}^{\infty} \exp[ju\xi] \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma_{\Delta m}} \cdot \exp\left[-\frac{1}{2} \frac{\xi^2}{\sigma_{\Delta m}^2}\right] d\xi \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \cdot \sigma_{\Delta m}} \cdot \exp\left[-\frac{1}{2} \left(\frac{\xi}{\sigma_{\Delta m}} - ju\sigma_{\Delta m}\right)^2 - \frac{1}{2} u^2 \sigma_{\Delta m}^2\right] d\xi \\
&= \exp\left[-\frac{1}{2} u^2 \sigma_{\Delta m}^2\right] \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma_{\Delta m}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} \left(\frac{\xi}{\sigma_{\Delta m}} - ju\sigma_{\Delta m}\right)^2\right] d\xi \\
&= \exp\left[-\frac{1}{2} u^2 \sigma_{\Delta m}^2\right] \cdot 1
\end{aligned}$$

Ein Vergleich des Erwartungswertes der charakteristischen Funktion mit dem gesuchten Erwartungswert von $\varphi_{yy}(\tau)$ ergibt den einzusetzenden Wert für u :

$$\varphi_{yy}(\tau) = \phi_{\Delta m}(u = 2\pi\gamma) = \exp\left[-\frac{1}{2} (2\pi\gamma)^2 \sigma_{\Delta m}^2\right] \tag{2-64}$$

Zunächst wird noch $\sigma_{\Delta m}$ bestimmt, bevor dann die Ergebnisse der Nebenrechnungen in die gesuchte Formel für die Bandbreite des modulierten Signals eingesetzt werden. $\sigma_{\Delta m}$ entspricht der AKF des Differenzsignals des GRP (Gaussian Random Process) $\sigma_{\Delta m} \triangleq \varphi_{\Delta m \Delta m}(0)$.

Die Leistung P_m wird als quadratische Größe verwendet:

$$\begin{aligned}
\sigma_{\Delta m}^2(0) &= E\{(m(t+\tau) - m(t))^2\} \\
&= E\{(m(t+\tau))^2 + E\{m(\tau)\}^2\} - 2 \cdot E\{(m(t+\tau) \cdot m(\tau))\} \\
&= \varphi_{mm}(0) + \varphi_{mm}(0) - 2 \cdot \varphi_{mm}(\tau) \\
&= 2[P_m - \varphi_{mm}(\tau)] \\
&= 2[P_m(1 - \rho_{mm}(\tau))]
\end{aligned} \tag{2-65}$$

Durch Einsetzen von (2-64) in (2-65) ergibt sich:

$$\begin{aligned}\varphi_{yy}(\tau) &= \exp\left[-\frac{1}{2}(2\pi\gamma)^2 \cdot 2 \cdot (P_m - \varphi_{mm}(\tau))\right] \\ &= \exp\left[4\pi^2\gamma^2 P_m (\tilde{\varphi}_{mm}(\tau) - 1)\right] \quad \text{mit : } \tilde{\varphi}_{mm} = \frac{\varphi_{mm}}{P_m}\end{aligned}$$

Außerdem gilt für P_y : $P_y = \varphi_{yy}''(0) = 1$

Die 2. Ableitung ergibt an der Stelle '0' :

$$\begin{aligned}\varphi_{yy}'(\tau) &= 4\pi^2\gamma^2 P_m \tilde{\varphi}_{mm}'(\tau) \cdot \exp\left[4\pi^2\gamma^2 P_m (\tilde{\varphi}_{mm}(\tau) - 1)\right] \\ \varphi_{yy}''(\tau) &= \left(4\pi^2\gamma^2 P_m \tilde{\varphi}_{mm}'(\tau)\right)^2 \cdot \exp\left[4\pi^2\gamma^2 P_m (\tilde{\varphi}_{mm}(\tau) - 1)\right] + \\ &\quad 4\pi^2\gamma^2 P_m \tilde{\varphi}_{mm}''(\tau) \cdot \exp\left[4\pi^2\gamma^2 P_m (\tilde{\varphi}_{mm}(\tau) - 1)\right] \\ \varphi_{yy}''(0) &= \left(4\pi^2\gamma^2 P_m \tilde{\varphi}_{mm}'(0)\right)^2 \cdot 1 + 4\pi^2\gamma^2 P_m \tilde{\varphi}_{mm}''(0) \\ \varphi_{yy}''(0) &= 4\pi^2\gamma^2 P_m \tilde{\varphi}_{mm}''(0) \quad \text{da : } \tilde{\varphi}_{mm}(0) = 1 ; \tilde{\varphi}_{mm}'(0) = 0\end{aligned}$$

Für B ergibt sich daraus:

$$\begin{aligned}B^2 &= -\frac{\varphi_{yy}''(0)}{4\pi^2 \cdot \varphi_{yy}(0)} = -\frac{1}{4\pi^2} \cdot 4\pi^2\gamma^2 P_m \tilde{\varphi}_{mm}''(0) & (2-66) \\ B^2 &= -\gamma^2 P_m \tilde{\varphi}_{mm}''(0)\end{aligned}$$

Es wurde eine Beziehung zwischen Bandbreite des modulierten Signals mit der AKF des Quellsignals hergeleitet. In den abschließenden Schritten der Herleitung wird die bekannte Beziehung zwischen φ_{yy} und B erneut verwendet.

$$B_y^2 = -\frac{\tilde{\varphi}_{yy}''(0)}{4\pi^2} ; B_m^2 = -\frac{\tilde{\varphi}_{mm}''(0)}{4\pi^2} ; \varphi_{yy}''(0) = 4\pi^2\gamma^2 P_m \tilde{\varphi}_{mm}''(0)$$

$$\Rightarrow \varphi_{yy}''(0) = 4\pi^2\gamma^2 P_m \tilde{\varphi}_{mm}''(0)$$

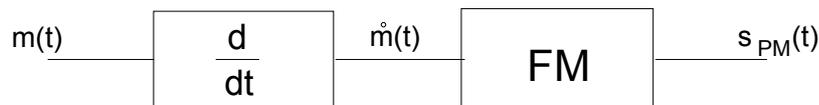
$$\begin{aligned}-4\pi^2 B_y^2 &= 4\pi^2\gamma^2 P_m B_m^2 (-4\pi^2) \\ B_y^2 &= 4\pi^2\gamma^2 P_m B_m^2 & (2-67) \\ B_y &= 2\pi \gamma \sqrt{P_m} B_m\end{aligned}$$

Die effektive Bandbreite eines PM-Signals läßt sich als Produkt aus Bandbreite und Wurzel der Leistung des Quellsignals beschreiben.

Bei der folgenden Untersuchung für FM-Signale ergibt sich ein Unterschied zu PM-Signalen:

Unter Verwendung des Wiener-Lee-Theorems läßt sich aus dem bereits gefundenem Ergebnis für die PM direkt die Bandbreitenbetrachtung für frequenzmodulierte Signale durchführen.

Die Beziehung zwischen FM und PM kann als Übertragungsglied gesehen werden (s. 2.1.1.2.1):



Es gilt allgemein:

$$\phi_{ss}(f) \cdot |H(f)|^2 = \phi_{gg}(f) \quad \text{(Wiener-Lee-Theorem)}^5 \quad (2-68)$$

Ein Differenzierglied im Fourierbereich lautet: $\frac{d}{dt} \circ - \bullet j2\pi f = H(f)$

Mit (2-68) kann die gesuchte Beziehung zwischen PM und FM aufgestellt werden:

$$\begin{aligned} \phi_{FM}(f) &= \phi_{\dot{m}m}(f) = \phi_{mm}(f) \cdot |2\pi f|^2 \\ \phi_{\dot{m}m}(f) &= \phi_{mm}(f) \cdot (2\pi f)^2 \end{aligned} \quad (2-69)$$

Daraus ergibt sich mit dem Wiener-Lee-Theorem auch für den Zeitbereich:

$$\varphi_{\dot{m}m}(\tau) = \frac{d^2}{d\tau^2} \varphi_{mm}(\tau)$$

Das Parseval'sche Gesetz stellt eine nützliche Beziehung zwischen Leistung, Spektrum und der Leistungsdichtefunktion her:

⁵ s(t) ist das Eingangssignal, g(t) das Ausgangssignal und H(f) die Übertragungsfunktion dieses Systems

$$\begin{aligned}
P_{\text{in}} &= \varphi_{\text{min}}(0) = \int_{-\infty}^{\infty} \phi_{\text{min}}(f) df \\
&= \int_{-\infty}^{\infty} 4\pi^2 f^2 \phi_{\text{min}}(f) df \\
&= 4\pi^2 P_m \int_{-\infty}^{\infty} f^2 \tilde{\phi}_{\text{min}}(f) df && \text{in normierte Teile zerlegt !} \\
P_{\text{in}} &= 4\pi^2 P_m B_m^2 && \text{(2-70)}
\end{aligned}$$

Für die Bandbreite in Abhängigkeit von Leistung und Bandbreite des Quellsignals wird die gefundene Beziehung für die Frequenzmodulation eingesetzt:

$$\begin{aligned}
B_y^2 &= 4\pi^2 \gamma^2 \cdot B_m^2 \cdot P_m = \gamma^2 \cdot P_{\text{in}} && \text{(2-71)} \\
B_y &= \gamma \cdot \sqrt{P_{\text{in}}}
\end{aligned}$$

Die effektive Bandbreite des FM-Signals wird durch die Wurzel der Leistung des abgeleiteten Quellsignals gewichtet mit dem Modulationsindex γ , beschrieben.

Somit ist die effektive Bandbreite des FM-Signals unabhängig von der Bandbreite des Quellsignals, was einen großen Unterschied zur Bandbreite des PM-Signals darstellt! Die Spektren ändern sich folglich bei verschiedenen Quellsignalen unterschiedlich bei PM und FM (Das Spektrum PM modulierter Signale ist meist breiter als das Spektrum von FM-Signalen.).

2.3.2 Bandbreiteabschätzung zeitdiskreter Signale

Die Bandbreite eines zeitkontinuierlichen Modulationssignals $y(t) = a(t)e^{j\varphi(t)}$ kann allgemein aus den Mittelwerten der Fourierfrequenz $\bar{\omega}$ und der mittleren Augenblicksfrequenz $\bar{\Omega}$ eines Signals bestimmt werden. Als Ergebnis der Untersuchung beider Mittelwerte erhalten wir nach /72/:

$$\bar{\omega} = \bar{\Omega} = \frac{\int_{-\infty}^{+\infty} \Omega(t) |a(t)|^2 dt}{\int_{-\infty}^{+\infty} |a(t)|^2 dt} \quad \text{(2-72)}$$

$a(t)$ stellt eine zusätzliche zeitabhängige Amplitude des modulierten Signals dar. Für die zeitdiskrete Analyse der Mittelwerte und Varianzen von Fourierfrequenz und

Augenblicksfrequenz ergeben sich jedoch unterschiedliche Gleichungen für $\Omega(n)$ und $\omega(n)$. Gegeben ist das komplexe, zeitdiskrete Signal $y(n) = a(n)e^{j\varphi(n)}$. Die zugehörige diskrete Fouriertransformierte lautet:

$$Y(k) = \frac{1}{N} \sum_{n=0}^{N-1} y(n) e^{-j2\pi nk/N} = \frac{1}{N} \sum_{n=0}^{N-1} a(n) \cdot e^{j\varphi(n)} \cdot e^{-j2\pi nk/N} \quad (2-73)$$

Über die Betrachtung im sogenannten zirkularen Betrachtungsraum lassen sich die Gleichungen für die beiden Mittelwerte ($\bar{\Omega}(n)$ und $\bar{\omega}(n)$) bestimmen:

$$\begin{aligned} \bar{\omega} &= \arg \left[\sum_{n=0}^{N-1} y(n+1) y^*(n) \right] \\ \bar{\Omega}(n) &= \arg \left[\sum_{n=0}^{N-1} y(n) y^*(n-1) \left| \frac{y(n)}{y(n-1)} \right| \right] \end{aligned} \quad (2-74)$$

Ebenso lassen sich die Varianzen V_{ω} und V_{Ω} bestimmen:

$$\begin{aligned} V_{\omega} &= 1 - \left(\sum_{n=0}^{N-1} |y(n)|^2 \right)^{-1} \left| \sum_{n=0}^{N-1} y(n) y^*(n-1) \right| \\ V_{\Omega} &= 1 - \left(\sum_{n=0}^{N-1} |y(n)|^2 \right)^{-1} \left| \sum_{n=0}^{N-1} y(n) y^*(n-1) \left| \frac{y(n)}{y(n-1)} \right| \right| \end{aligned} \quad (2-75)$$

Durch Vergleich mit der statistischen Varianz über den zirkularen Betrachtungsraum läßt sich dann die Bandbreite ermitteln. Die Bandbreite B_{ω} ist über die Gaußverteilungsfunktion der Wahrscheinlichkeitsdichte definiert:

$$B_{\omega} = 2\sqrt{2 \ln 2} \cdot \sigma_{\omega} \quad (2-76)$$

Der Vorfaktor ergibt sich durch die 3dB Grenzfrequenz der Gaußglocke. Durch weiteres Einsetzen folgt abschließend als Näherung die Bandbreite in Abhängigkeit der beiden statistischen Varianzen:

$$B_{\omega/\Omega} \approx 3.33 \sqrt{V_{\omega/\Omega}} \quad (2-77)$$

2.4 SCHÄTZVERFAHREN

Schätzverfahren werden zur Extraktion von Nutzinformation aus gestörten Meßgrößen angewandt. Dabei ergeben sich unterschiedliche Möglichkeiten, einen Schätzwert zu bestimmen, die abhängig von der benötigten Genauigkeit, Art der Störung und der erforderlichen Rechengeschwindigkeit sind. Das einfachste und auch schnellste Verfahren ist eine rein deterministische Betrachtung der Meßwerte eines stationären Problems.

Chronologisch und historisch betrachtet ist der nächste Schritt eine zusätzliche Auswertung der stochastischen Vorgaben in einem ebenfalls stationären Problem. Gauß begann bereits 1795 mit der stochastischen Betrachtung von ungenauen Meßwerten ("Theoria Motus Corporum Coelestium" 1809). Etwa zur gleichen Zeit entwickelte der Mathematiker Legendre ebenso eine erste Form der LMS (Least Mean Squares) Schätzung. Die Schätzwerttheorie wurde erst in den 50er Jahren unseres Jahrhunderts von Wiener und Kolmogorov durch die Maximum-Likelihood Schätzung weiterentwickelt. Seit dieser Zeit nahm die Estimationstheorie eine rasante Entwicklung. Mehrere Schätzverfahren, die von ihrer Herleitung und Idee her unterschiedlich sind, ergeben jedoch in der Realisation der Schätzwertbestimmung im Schätzalgorithmus bei gaußverteilten Größen sogar identische Berechnungsvorschriften. Dazu zählen:

Maximum - Likelihood Estimation, Maximum a posteriori Estimation, Minimum Varianz Estimation (MLS - Minimum Least Squares Methode), Least Mean Squares (LMS) und Weighted Least Squares Estimation.

Handelt es sich um instationäre, dynamische Systeme, so werden rekursive Filter verwendet, um den Informationsgehalt der vergangenen Meßwerte in die Bestimmung des Schätzwertes mit einzubeziehen. Dabei lassen sich für lineare Problemstellungen erneut rein deterministische Regelungen aufbauen (Luenberger Beobachter) oder unter Einbeziehung von sogenannten Kalman-Filtern die stochastischen Kenngrößen mitverwerten. Hierbei wird im folgenden von gaußverteilten, weißen Rauschprozessen ausgegangen.

Zur Lösung von nichtlinearen Problemen werden nichtlineare Filter benötigt. Angewandte Verfahren sind hier eine Weiterentwicklung des linearen Kalman-Filters zum Extended Kalman-Filter (Kalman-Bucy-Filter) oder weitere Kalman-Filter höherer Ordnung /3, 12, 24, 30, 35, 36, 39, 48, 49, 50, 52, 63, 65, 74/.

2.4.1 Schätzwertbestimmung über bedingte Verteilungsdichten

Existieren mehrere Messungen mit bekannten Verteilungsdichten $f_{x|y_1}(\xi|\rho_1), f_{x|y_2}(\xi|\rho_2)$ zur Bestimmung einer gesuchten Größe, dann lassen sich diese Verteilungsdichten zu einer gemeinsamen, bedingten Verteilungsdichte $f_{x|y_1, y_2}(\xi|\rho_1, \rho_2)$ zusammenfassen, um die gesuchte Größe - den Schätzwert \hat{x} zu bestimmen.

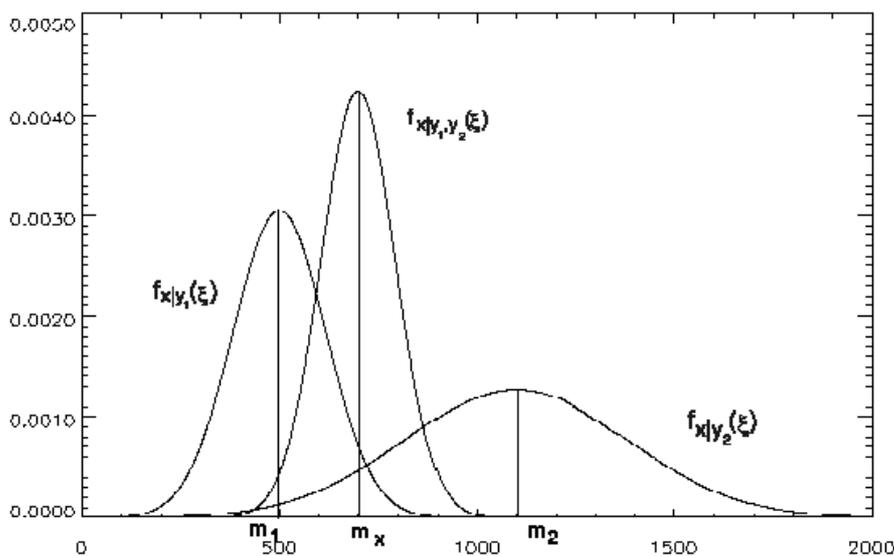


Abb. 2.19: Bedingte Verteilungsfunktionen

Die Basis zum Verständnis der bedingten Verteilungsdichten liefert der Satz von Bayes. Er wurde von Bayes Anfang des 18. Jahrhunderts zunächst als Hypothese aufgestellt. Er betrachtet die Beziehung zweier Verteilungen zueinander:

$$p(x|y) = \frac{p(x) \cdot p(y|x)}{p(y)} = \frac{p(y, x)}{p(y)} \quad (2-78)$$

$$f_{x|y}(\underline{\xi} | \underline{\rho}) = \frac{f_x(\underline{\xi}) \cdot f_{y|x}(\underline{\rho} | \underline{\xi})}{f_y(\underline{\rho})} = \frac{f_{y,x}(\underline{\rho}, \underline{\xi})}{f_y(\underline{\rho})} \quad (2-79)$$

$$f_{x|y}(\underline{\xi} | \underline{\rho}) \cdot f_y(\underline{\rho}) = f_{y,x}(\underline{\rho}, \underline{\xi})$$

Mittelwert und Varianz der gemeinsamen, bedingten Verteilungsdichtefunktion $f_x(\xi)$ bestimmen sich folgendermaßen:

$$m_x = \left(\frac{\sigma_{y_2}^2}{\sigma_{y_1}^2 + \sigma_{y_2}^2} \right) \cdot y_1 + \left(\frac{\sigma_{y_1}^2}{\sigma_{y_1}^2 + \sigma_{y_2}^2} \right) \cdot y_2 = \hat{x} \quad (2-80)$$

$$\sigma_x^2 = \frac{\sigma_{y_1}^2 \sigma_{y_2}^2}{\sigma_{y_1}^2 + \sigma_{y_2}^2} \quad (\text{Varianzen addieren sich als reziproke Werte}) \quad (2-81)$$

Der bestimmte Mittelwert m_x ist zugleich auch der gesuchte optimale Schätzwert \hat{x} . Gleichzeitig ist dieser Schätzwert bei Gaußverteiltetheit auch der Maximum-Likelihood-Schätzwert (Maximum der Gaußglocke) und der Minimum-Varianz-Schätzwert.

Nach der statischen Behandlung von bedingten Verteilungsdichtefunktionen kann man sich - bei der Ermittlung der bedingten Verteilungsdichte und des optimalen Schätzwertes - die Verteilungsdichtefunktionen auch als zeitlich aufeinanderfolgende Verteilungsdichtefunktionen vorstellen. So läßt sich eine bedingte Verteilungsdichtefunktion $f_{x(t_2)|y(t_1),y(t_2)}(\xi | \rho_1, \rho_2)$ zum Zeitpunkt t_2 folgendermaßen bestimmen: Durch Einsetzen und Umformen der obigen Gleichung des Mittelwertes einer bedingten Verteilungsdichte ergibt sich:

$$\hat{x}(t_2) = m_x = y(t_1) + \left(\frac{\sigma_{y_1}^2}{\sigma_{y_1}^2 + \sigma_{y_2}^2} \right) \cdot (y(t_2) - y(t_1))$$

$$\Rightarrow \quad (2-82)$$

$$\hat{x}(t_2) = \hat{x}(t_1) + K(t_2) \cdot (y(t_2) - \hat{x}(t_1)) \quad \text{mit } \hat{x}(t_1) = y(t_1)$$

In Worten bedeutet dies: Der optimale Schätzwert $\hat{x}(t_2)$ läßt sich beschreiben als die Summe aus vorherigem Schätzwert $\hat{x}(t_1)$ und der Differenz der aktuellen Messung zur vorherigen Messung, gewichtet mit einem Korrekturterm K , der sich aus den Varianzen der beiden Messungen zusammensetzt. Da die vorherige Messung bereits durch eine Schätzung optimiert wurde, kann stattdessen der genauere Schätzwert $\hat{x}(t_1)$ eingesetzt werden. Diese Gleichung hat rekursiven Charakter und stellt in der Regelungstechnik die Standardgleichung zur Systembeschreibung dar. Für die Varianzen gilt parallel dazu:

$$\sigma_x^2(t_2) = \sigma_x^2(t_1) - K(t_2) \cdot \sigma_x^2(t_1)$$

Diese Gesetzmäßigkeiten gelten für stationäre und dynamische Systeme. Bei einem dynamischen System wissen wir, daß die Kenntnis über die Genauigkeit einer Schätzung bei bekanntem Startwert mit fortlaufender Schätzung abnimmt.

Stellt man ein allgemeines, zeitkontinuierliches Modell für ein lineares System auf, so lauten die Systemgleichungen:

$$\dot{x}(t) = F(t) \cdot x(t) + B(t) \cdot u(t) + G(t)w(t) \quad \text{Zustandsübergangsgleichung} \quad (2-83)$$

$$y(t) = C \cdot x(t) + v(t) \quad \text{Beobachtungsgleichung} \quad (2-84)$$

oder der beschriebene Regelkreis als Blockdiagramm dargestellt:

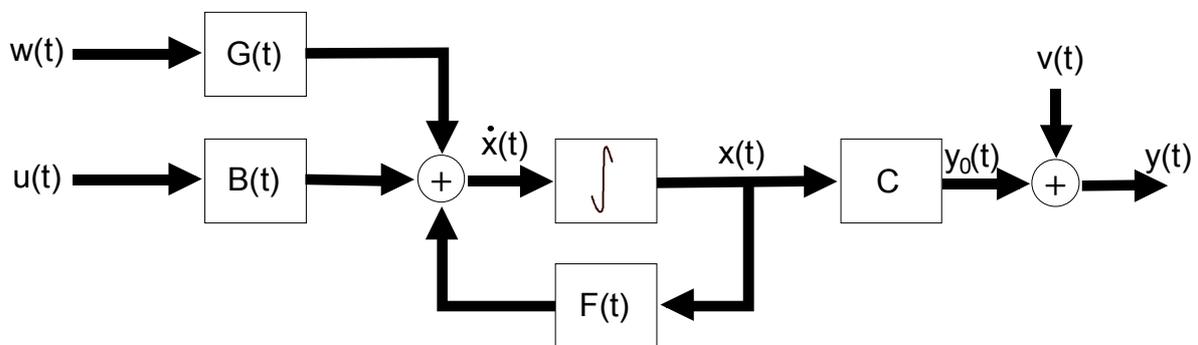


Abb. 2.20: Zeitkontinuierliches Modell

Zunächst soll das stationäre Problem betrachtet werden, um übliche Verfahren zu beschreiben, und im Anschluß daran erfolgt eine Erweiterung auf die dynamische Problemstellung.

2.4.1.1 Informationsfluß im Übertragungskanal

Der Satz von Bayes steht in direktem Zusammenhang zum Informationsfluß im Übertragungskanal.

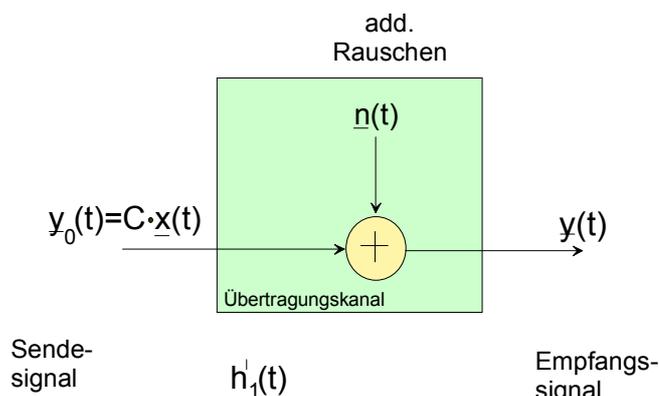


Abb. 2.21: Übertragungskanal

Unter Vereinfachung der Stoßantwort $h_1(t)$ aus Abschnitt 2.1.3 erhält man den für die Modellbildung des Kalman-Filters entscheidenden Teil des Übertragungskanal.⁶

Über den Erwartungswert der logarithmierten Verteilungsdichtefunktionen läßt sich die Entropie bestimmen. Für die Entropien am Übertragungskanal ergibt sich somit basierend auf den Satz von Bayes:

$$\begin{aligned}
 H(\underline{y}, \underline{x}) &= - \int_{\underline{\Xi}} \int_{\underline{\Theta}} \ln f_{\underline{y}, \underline{x}}(\underline{\rho}, \underline{\xi}) \cdot f_{\underline{y}, \underline{x}}(\underline{\rho}, \underline{\xi}) \cdot d\underline{\rho} d\underline{\xi} \\
 &= - \int_{\underline{\Xi}} \int_{\underline{\Theta}} \ln [f_{\underline{x}|\underline{y}}(\underline{\xi} | \underline{\rho}) \cdot f_{\underline{y}}(\underline{\rho})] \cdot f_{\underline{y}, \underline{x}}(\underline{\rho}, \underline{\xi}) \cdot d\underline{\xi} d\underline{\rho} \\
 &= - \int_{\underline{\Xi}} \int_{\underline{\Theta}} [\ln f_{\underline{x}|\underline{y}}(\underline{\xi} | \underline{\rho}) + \ln f_{\underline{y}}(\underline{\rho})] \cdot f_{\underline{y}, \underline{x}}(\underline{\rho}, \underline{\xi}) \cdot d\underline{\xi} d\underline{\rho} \\
 &= - \int_{\underline{\Xi}} \int_{\underline{\Theta}} \ln f_{\underline{x}|\underline{y}}(\underline{\xi} | \underline{\rho}) \cdot f_{\underline{y}, \underline{x}}(\underline{\rho}, \underline{\xi}) \cdot d\underline{\xi} d\underline{\rho} - \int_{\underline{\Xi}} \int_{\underline{\Theta}} \ln f_{\underline{y}}(\underline{\rho}) \cdot f_{\underline{y}, \underline{x}}(\underline{\rho}, \underline{\xi}) \cdot d\underline{\xi} d\underline{\rho} \\
 &= - \int_{\underline{\Xi}} \int_{\underline{\Theta}} \ln f_{\underline{x}|\underline{y}}(\underline{\xi} | \underline{\rho}) \cdot f_{\underline{y}, \underline{x}}(\underline{\rho}, \underline{\xi}) \cdot d\underline{\xi} d\underline{\rho} - \int_{\underline{\Theta}} \ln f_{\underline{y}}(\underline{\rho}) \cdot \int_{\underline{\Xi}} f_{\underline{y}, \underline{x}}(\underline{\rho}, \underline{\xi}) \cdot d\underline{\xi} d\underline{\rho} \\
 &= - \int_{\underline{\Theta}} \left[\int_{\underline{\Xi}} \ln f_{\underline{x}|\underline{y}}(\underline{\xi} | \underline{\rho}) \cdot f_{\underline{y}, \underline{x}}(\underline{\rho}, \underline{\xi}) \cdot d\underline{\xi} \right] d\underline{\rho} - \int_{\underline{\Theta}} \ln f_{\underline{y}}(\underline{\rho}) \cdot f_{\underline{y}}(\underline{\rho}) \cdot d\underline{\rho} \\
 &= + \int_{\underline{\Theta}} H_{\underline{x}|\underline{y}}(\underline{\xi} | \underline{y} = \underline{\rho}) d\underline{\rho} - \underbrace{\int_{\underline{\Theta}} \ln f_{\underline{y}}(\underline{\rho}) \cdot f_{\underline{y}}(\underline{\rho}) \cdot d\underline{\rho}}_{H_{\underline{y}}}
 \end{aligned}$$

$$H_{\underline{y}, \underline{x}} = +H_{\underline{x}|\underline{y}} + H_{\underline{y}} \quad \text{Verbundentropie} \quad (2-85)$$

Analog läßt sich $H_{\underline{x}, \underline{y}}$ berechnen, so daß sich hier folgender Term ergibt:

$$H_{\underline{x}, \underline{y}} = +H_{\underline{y}|\underline{x}} + H_{\underline{x}} \quad (2-86)$$

⁶ Die Störgröße $n(t)$ in Abb. 3.21 entspricht $v(t)$ der Systemmodell bzw. Kalman-Filtermodellbeschreibung

Durch Gleichsetzen der beiden gleichen Verbundentropien ($H_{x,y} = H_{y,x}$) erhält man beispielsweise nach H_y aufgelöst:

$$\Leftrightarrow H(y) = H(x) - H(x|y) + H(y|x) \quad (2-87)$$

\uparrow \uparrow \uparrow \uparrow
 Ausgangs- Eingangs- Rückschluß- Streu-
 entropie entropie entropie entropie

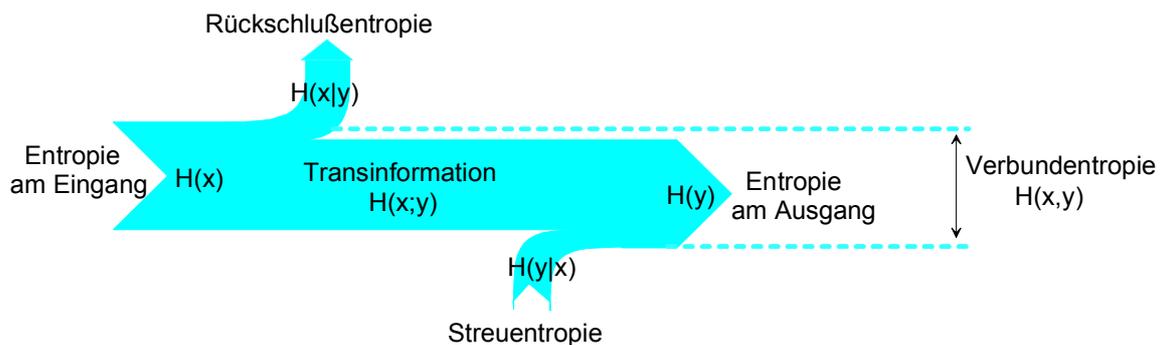


Abb. 2.22: Entropien am Übertragungskanal

Wird der Informationsfluß betrachtet, dann wird die Addition der Rauschgröße als Streuentropie und als Rückschlußentropie interpretiert. Die Streuentropie beschreibt die zusätzliche Störung durch das Rauschen direkt auf den Ausgang; die Rückschlußentropie bezeichnet die Auswirkung des Rauschens in Bezug auf den Eingang.

2.4.2 Stationäre Filter

Bezogen auf das Systemmodell bedeutet die stationäre Problematik, daß die Zustandsübergangsmatrix der Einheitsmatrix entspricht, da keine zeitliche Änderung des gesuchten Zustandes vorliegt.

2.4.2.1 Minimum - Varianz - Methode (Wiener)

MLS (Minimum Least Squares - Verfahren) und LMS (Least Mean Squares - Verfahren) beschreiben das gleiche Verfahren der Minimum Varianz Methode. Die Varianz der Filterschätzwerte muß minimiert werden, was bedeutet, daß das Quadrat der Abweichungen des Schätzwertes für $\underline{x}(t_0)$ vom a priori Schätzwert $\hat{\underline{x}}(t_0)$, die Quadratsumme der Fehler $\underline{w}(t_0)$ und die Quadratsumme der Beobachtungsfehler möglichst klein werden sollen.

Dazu wird die Ableitung der vom Zustand \underline{x} abhängigen Varianz berechnet. Die gefundene Beziehung wird gleich Null gesetzt und kann dann nach \underline{x} (Filterschätzwert) aufgelöst werden. Zur Kontrolle wird zusätzlich die 2. Ableitung berechnet, die Aufschluß über Minimum oder Maximum der gefundenen Extremstelle gibt.

2.4.2.2 Maximum Likelihood Estimation

Bei der ML-Estimation (Maximum Likelihood) wird die Verteilungsdichtefunktion eines Signals dazu verwendet, eine Gleichung für die Filterschätzwerte zu finden. Dazu muß das Maximum der Verteilungsfunktion der Messung, bezogen auf den gesuchten Schätzwert $f_{\underline{y}/\underline{x}}(\underline{\rho}/\underline{\xi})$, bestimmt werden. Die Ableitung der bedingten Verteilungsdichtefunktion wird folglich gleich Null gesetzt. Danach wird der Ausdruck nach \underline{x} aufgelöst. Es besteht keine Abhängigkeit zu vorherigen Meßwerten! Da die Verteilungsdichtefunktion der gesuchten Ausgangsgröße normalerweise unbekannt ist, wird mit Hilfe der bekannten Verteilungsdichte nach der Regel von Bayes umgeformt /38/.

$$f_{\underline{x}/\underline{y}}(\underline{\xi}/\underline{\rho}) = ? \quad \Rightarrow \quad f_{\underline{x}/\underline{y}}(\underline{\xi}/\underline{\rho}) = \frac{f_{\underline{y}/\underline{x}}(\underline{\rho}/\underline{\xi})f_{\underline{x}}(\underline{\xi})}{f_{\underline{y}}(\underline{\rho})} \quad (2-88)$$

Maximum-Bestimmung:

$$\left. \frac{d}{d\underline{\xi}} f_{\underline{x}/\underline{y}}(\underline{\xi}/\underline{\rho}) \right|_{\underline{\xi}=\hat{\underline{x}}_{\text{MAP}}} = \left. \frac{d}{d\underline{\xi}} \left[\frac{f_{\underline{y}/\underline{x}}(\underline{\rho}/\underline{\xi})f_{\underline{x}}(\underline{\xi})}{f_{\underline{y}}(\underline{\rho})} \right] \right|_{\underline{\xi}=\hat{\underline{x}}_{\text{MAP}}} = \underline{0}^T \quad (2-89)$$

($\hat{\underline{x}}_{\text{MAP}}$ stellt den Maximum-a-posteriori-Schätzwert dar, der im folgenden Abschnitt erläutert wird.) Durch Logarithmierung der gaußverteilten, bedingten Verteilungsdichte und weiteres Auflösen erhält man:

$$\left. \frac{d}{d\underline{\xi}} \ln f_{\underline{y}/\underline{x}}(\underline{\rho}/\underline{\xi}) - \frac{1}{2} \cdot \frac{d}{d\underline{\xi}} \left[\underline{\xi} - \underline{m}_{\xi} \right]^T \cdot \underline{P}_{\underline{xx}}^{-1} \cdot \left[\underline{\xi} - \underline{m}_{\xi} \right] \right|_{\underline{\xi}=\hat{\underline{x}}_{\text{MAP}}} = \underline{0}^T \quad (2-90)$$

Um den sogenannten Maximum-Likelihood-Schätzwert zu erhalten, wird der Grenzwert für $\underline{P}_{\underline{xx}}$ gegen unendlich geführt ($\underline{P}_{\underline{xx}}^{-1}$ gegen Null). Dies bedeutet, \underline{x} ist gaußverteilt mit unbekannten Parametern.

Aus den Maximum-Likelihood-Grenzwerten der rechten und der linken Seite ergibt sich:

$$\frac{d}{d\underline{\xi}} \ln f_{\underline{y}/\underline{x}}(\underline{\rho}/\underline{\xi}) \Big|_{\underline{\xi}=\hat{\underline{x}}_{MLE}} = \underline{0}^T \quad (2-91)$$

Die Funktion $\ln f_{\underline{y}/\underline{x}}(\underline{\rho}/\underline{\xi})$ wird als Likelihood-Funktion bezeichnet. Anschließend kann der gesuchte ML-Estimator berechnet werden.

$$f_{\underline{y}/\underline{x}}(\underline{\rho}/\underline{\xi}) = \left[(2\pi)^{1/2} \cdot \det(\mathbf{R})^{1/2} \right]^{-1} \cdot \exp \left\{ -\frac{1}{2} \cdot [\underline{\rho} - \mathbf{C} \cdot \underline{\xi}]^T \cdot \mathbf{R}^{-1} \cdot [\underline{\rho} - \mathbf{C} \cdot \underline{\xi}] \right\}$$

$$\frac{d}{d\underline{\xi}} \ln f_{\underline{y}/\underline{x}}(\underline{\rho}/\underline{\xi}) \Big|_{\underline{\xi}=\hat{\underline{x}}_{MLE}} = \frac{d}{d\underline{\xi}} \ln \left[\sqrt{2\pi \cdot \det(\mathbf{R})}^{-1} \cdot \exp\{\dots\} \right] = \underline{0}^T$$

$$\frac{d}{d\underline{\xi}} \left[-\frac{1}{2} \cdot \left([\underline{\rho} - \mathbf{C} \cdot \underline{\xi}]^T \cdot \mathbf{R}^{-1} \cdot [\underline{\rho} - \mathbf{C} \cdot \underline{\xi}] \right) \right] \Big|_{\underline{\xi}=\hat{\underline{x}}_{MLE}} = \underline{0}^T \quad (2-92)$$

Die Berechnung dieser Ableitung muß elementweise durchgeführt werden. Es ergibt sich jedoch eine ähnliche Schreibweise wie in obiger Nebenrechnung:

$$-\mathbf{C}^T \cdot \mathbf{R}^{-1} \cdot [\underline{\rho} - \mathbf{C} \cdot \underline{\xi}] \Big|_{\underline{\xi}=\hat{\underline{x}}_{MLE}} = \underline{0}^T \quad (2-93)$$

$$\mathbf{C}^T \mathbf{R}^{-1} \underline{\rho} = \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C} \cdot \hat{\underline{x}}_{MLE} \quad (2-94)$$

$$\hat{\underline{x}}_{MLE} = [\mathbf{C}^T \mathbf{R}^{-1} \mathbf{C}]^{-1} \cdot \mathbf{C}^T \mathbf{R}^{-1} \underline{\rho}$$

(\mathbf{C} entspricht dem differentiellen H_{Diff}) $\underline{\xi}$ ist der Maximum-Likelihood-Schätzwert $\hat{\underline{x}}(t)$ und $\underline{\rho}$ ist der dazugehörige Meßwert $\underline{y}(t)$.

Möchte man die ML-Estimation in einem dynamischen, zeitdiskreten System rekursiv einsetzen, so benötigt man die allgemeinen Filtergrößen. Die Maximum-Likelihood-Berechnung wird jedoch weiterhin nur stationär durchgeführt:

$$\dot{\underline{x}}(t) = \mathbf{A} \cdot \underline{x}(t) + \underline{w}$$

$$\underline{y}(t) = \mathbf{C} \cdot \underline{x}(t) + \underline{v}$$

An sich stellt die Maximum Likelihood Estimation keine echte Schätzung dar, sondern es wird nur die inverse Übertragungsfunktion angewendet, um aus dem Meßwert den zu erwartenden Zustandswert zu berechnen. Bei fehlender Störgröße und bekannter Übertragungsfunktion ist das Ergebnis fehlerfrei.

Durch Einsetzen von $\underline{x}(t) = \hat{\underline{x}}(t) \hat{=} \underline{\xi}(t)$ erhält man:

$$\Rightarrow \underline{y}(t) = C \cdot [C^T R^{-1} C]^{-1} \cdot C^T R^{-1} \cdot \underline{y}(t) + \underline{v}(t) \quad |^7$$

Die optimale Umsetzung eines Systemmodells in einer derartigen Prädiktor-Korrektor-Struktur läßt sich auch recht anschaulich am Blockschaltbild darstellen.

Das Modell der Prädiktor-Korrektor-Struktur stellt das Abbild des Systemmodells dar. Besitzt man folglich genaue Kenntnis über das reale System und kann dies mit einem Systemmodell mathematisch beschreiben, so läßt sich auch ein ebenso genaues Filtermodell dazu ableiten.

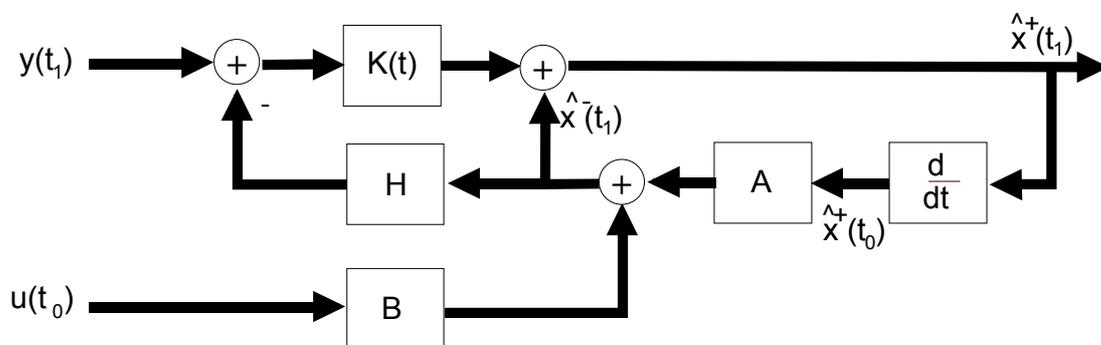


Abb. 2.23: Filtermodell

2.4.3 Dynamische Filter

Der Beobachter (Luenberger) stellt eine einfache Regelung dar. Ohne Verwendung stochastischer Kenngrößen wird die Meßgröße aufgrund des ermittelten Fehlers nachgeregelt. Der Luenberger-Beobachter stellt einen Kalman-Filter ohne Kovarianzenzyklus dar. Während beim im folgenden beschriebenen Kalman-Filter ein Schätzzyklus und ein Varianzenzyklus parallel und verkettet ablaufen, wird beim Luenberger-Beobachter nur der Schätzzyklus realisiert. Ein PLL (s. 3.1.2.2 und 3.4.3) kann man in seiner Funktionsweise als Realisierung eines zeitkontinuierlichen Beobachters verstehen.

⁷ Das Bilden dieser inversen Matrix ist hier grundsätzlich immer möglich, da die Matrix symmetrisch ist.

2.4.3.1 Kalman - Filter

Beim Kalman-Filter werden die stochastischen Kenngrößen zusätzlich im Regelkreis verwendet. Bei einem linearen System und gaußverteilten Störgrößen läßt sich eine optimale Filterung der gesuchten Größen erreichen - man spricht beim Kalman-Filter vom Optimalfilter. Diese Eigenschaften bieten sich auch für nichtlineare Systeme an; durch Linearisierungen mit einem sogenannten Extended Kalman-Filter lassen sich die guten Eigenschaften eines Optimalfilters nutzen.

Um ein Kalman-Filter einsetzen zu können, wird eine deterministische Modellbeschreibung des vorliegenden Systems benötigt. Mit diesem Zustandsraummodell wird dann das Kalman-Filter berechnet. Dabei wird bei jedem Schätzwert anhand der aktuellen Fehlerkovarianzen die aktuelle Gewichtung des Filtermodells neu bestimmt. Das Filter stellt ein optimiertes rekursives Datenfilter dar, das für lineare Problemstellungen entwickelt wurde.

Im Gegensatz zum Wiener-Filter muß beim Kalman-Filter von einer weißen, gaußverteilten Störgröße ausgegangen werden. Die Kenntnis andersartiger Verteiltheit muß im Zustandsmodell mit eingebunden werden, da der Filter nicht auf der Auswertung von Leistungsdichtespektren basiert.

2.4.3.1.1 Kontinuierliche und diskrete Betrachtungsweise

Zuerst wurde das zeitdiskrete Optimalfilter von R.E. Kalman /24/ entwickelt und einige Zeit später das sogenannte Kalman-Bucy-Filter von Bucy /12, 24, 74/ entwickelt. Beide Betrachtungsweisen und deren Herleitungen sind wichtig für die Anwendung und das Verständnis des Filters. Die Praxis bietet fast durchgehend zeitkontinuierliche Prozesse, die jedoch nur in zeitdiskreten Algorithmen verarbeitet werden können. Aus diesem Grunde läßt sich außerdem eine Mischform aus zeitdiskreter und zeitkontinuierlicher Darstellung betrachten. Die sich ergebenden Algorithmen unterscheiden sich nur bei der Bestimmung der Kovarianzen.

Dieser Unterschied wird folgendermaßen begründet: Die Diskretisierung kann als äquidistante, ideale Abtastung beschrieben werden: $t_k = k \cdot T$ (Zur Vereinfachung wird T normalerweise weggelassen.) Die Ableitung $\dot{x}(t_k)$ beschreibt die Steigung zum Zeitpunkt t_k . Wird die Ableitung im diskreten Modell betrachtet, so ergibt sich ein Differenzenquotient,

der jedoch die Steigung zwischen zwei Zeitpunkten t_k und t_{k+1} beschreibt. Durch die anderen mathematischen Gesetzmäßigkeiten einer Differentialgleichung ergibt sich ein Unterschied bei der Bestimmung der Kovarianzen /38/:

$$\begin{aligned} \dot{\underline{x}}(t) &= F \cdot \underline{x}(t) + B \cdot \underline{u}(t) + G \cdot \underline{w}(t) \\ \underline{y}(t) &= C \cdot \underline{x}(t) + \underline{v}(t) \end{aligned} \tag{2-95}$$

Der allgemeine Lösungsansatz für die Differentialgleichung lautet:

$$\underline{x}(t) = \phi(t, t_0) \cdot \underline{x}(t_0) + \int_{t_0}^t \phi(t, \tau) \cdot B \cdot \underline{u}(\tau) d\tau + \int_{t_0}^t \phi(t, \tau) \cdot G \cdot \underline{w}(\tau) d\tau \tag{2-96}$$

Zur grundsätzlichen Betrachtung des Systems gehen wir von einem störungsfreien System aus, so daß das stochastische Signal gleich Null ist und damit nicht weiter betrachtet wird. Im linearen und zeitinvarianten Fall ist eine Lösung im Laplace-Bereich deutlich einfacher. Daraus folgt:

$$\begin{aligned} \underline{x}(k+1) &= F \cdot \underline{x}(k) + B \cdot \underline{u}(k) + G \cdot \underline{w}(k) \\ \underline{y}(k) &= C \cdot \underline{x}(k) + \underline{v}(k) \end{aligned} \tag{2-97}$$

Die Praxis hat jedoch gezeigt, daß nur die Programmierung des zeitdiskreten Modells zu einem stabilen Algorithmus führt, so daß im weiteren nur zeitdiskrete Modelle behandelt werden sollen. Die Änderungen im Kovarianzenzyklus sind mitunter sehr klein, und Störungen in der Symmetrie der Matrizen führen zu großen Ungenauigkeiten und Instabilitäten, so daß wir bei der Anwendung von Kalman-Filtern grundsätzlich nur zeitdiskrete Modelle betrachten.

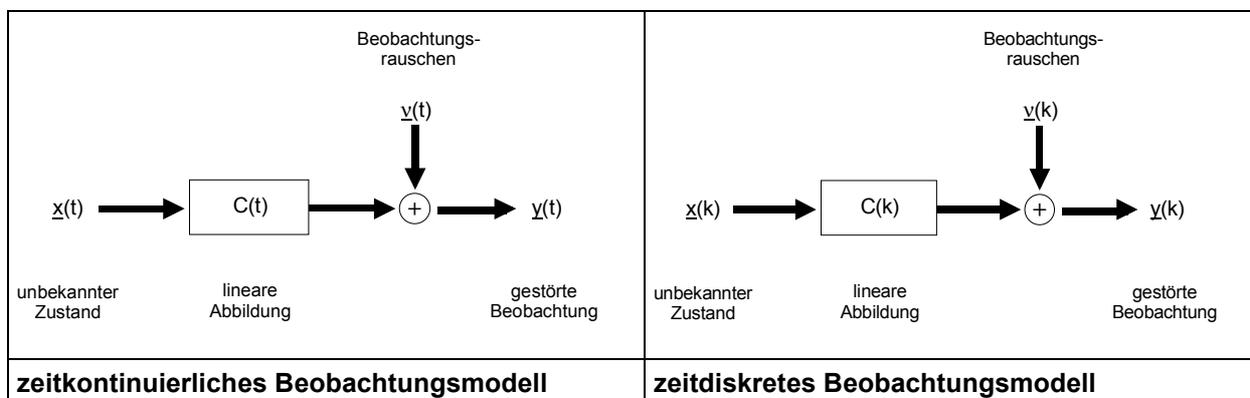


Abb. 2.24: Beobachtungsmodell

Das von R. E. Kalman entwickelte Filter ist ursprünglich ein lineares Filter, da die hier zu untersuchende Problemstellung jedoch nichtlinear ist, wird mit dem sog. erweitertem linearisiertem (extended linearised) Kalman-Filter gearbeitet.

2.4.3.2 Nichtlineare Filter (EKF)

Bei der Betrachtung nichtlinearer Problemstellungen, wie wir sie in dieser Arbeit finden, wird der Einsatz des Extended (linearised) Kalman-Filters notwendig /3, 5, 39, 40, 41, 42, 48/. Es wird jetzt das gesamte Systemmodell für den nichtlinearen, zeitdiskreten Fall betrachtet:

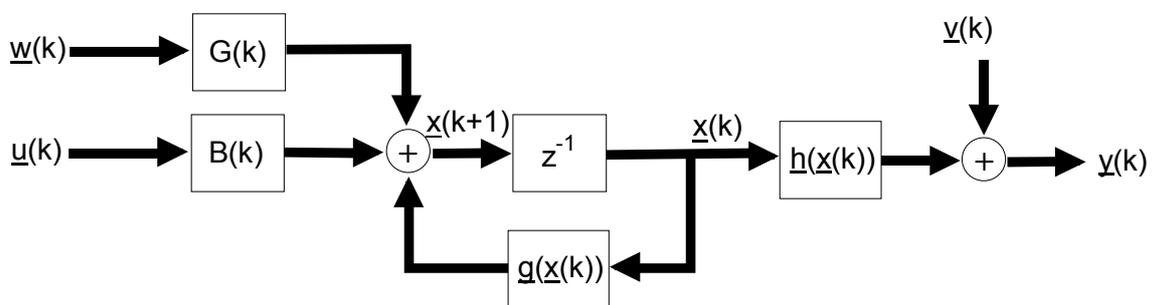


Abb. 2.25: Nichtlineares Systemmodell

Spaltet man das Modell in ein Beobachtungsmodell und ein Zustandsmodell anhand der zwei systembeschreibenden Gleichungen auf, so sieht man im Beobachtungsmodell die nichtlineare Abbildung in der Übertragungsfunktion $\underline{h}(\underline{x}(k))$ und die additive Störung mit der Rauschgröße $\underline{v}(k)$.

$$\underline{y}(k) = \underline{h}(\underline{x}(k)) + \underline{v}(k) \quad \text{nichtlineare Beobachtungsgleichung} \quad (2-98)$$

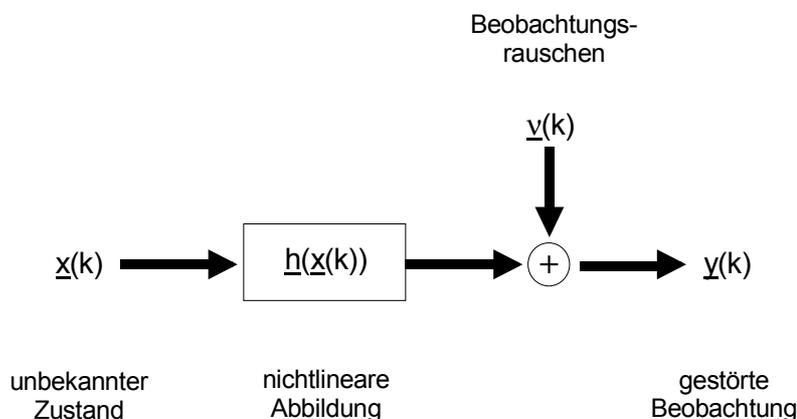


Abb. 2.26: Nichtlineares Beobachtungsmodell

Die Rauschgröße $\underline{v}(k)$ wird als weiß, gaußverteilt und mittelwertfrei angenommen und ist definiert über das 2. Moment vollständig beschrieben:

$$E\{\underline{v}(k)\} = \underline{0} \quad \wedge \quad E\{\underline{v}(k) \cdot \underline{v}(j)^T\} = R \cdot \delta_{k,j} \quad (2-99)$$

Im Zustandsraum, der losgelöst von den Beobachtungsgrößen betrachtet werden kann, wird der "innere Antrieb" des Systems beschrieben. Betrachten wir einen komplexen nichtlinearen Zusammenhang zwischen zwei aufeinanderfolgenden Zuständen, so kann auch die Zustandsübergangsmatrix nichtlinear sein und zu einem nichtlinearen Kalman-Filter führen.

$\underline{U}(k)$ mit B (Gewichtungsmatrix) ist ein bekannter Offset, der unabhängig von der Zustandsgröße addiert wird.

$$\underline{x}(k+1) = A(\underline{x}(k)) + B(k) \cdot \underline{u}(k) + G(k) \cdot \underline{w}(k); \quad \text{mit } G(k) = I \text{ (Einheitsmatrix)}$$

Zustandsgleichung

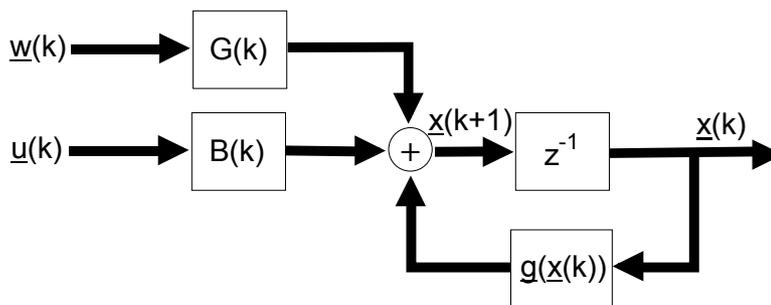


Abb. 2.27: Nichtlineares Zustandsübergangsmodell

Die Rauschgröße $\underline{w}(k)$ sei ebenfalls weiß, gaußverteilt und mittelwertfrei und kann analog zu $\underline{v}(k)$ vollständig durch die Kovarianz beschrieben werden:

$$E\{\underline{w}(k)\} = \underline{0} \quad \wedge \quad E\{\underline{w}(k) \cdot \underline{w}(j)^T\} = Q \cdot \delta_{k,j} \quad (2-100)$$

Aus der ausführlichen Beschreibung des zeitdiskreten Systemmodells wird das Filtermodell entwickelt: Dazu wird grundsätzlich das Zustandsübergangsmodell übernommen. Das bedeutet, die Zustände (gesuchte Filtergrößen) werden durch eine Anregung mit einem inneren Rauschprozeß (Driving Noise) eigenständig erzeugt. Das vorhandene Beobachtungsmodell wird nun, da die Beobachtungen als Eingangsgrößen und nicht als Ausgangsgröße zur Verfügung stehen, "invertiert" und zur Steuerung des inneren Prozesses verwendet.

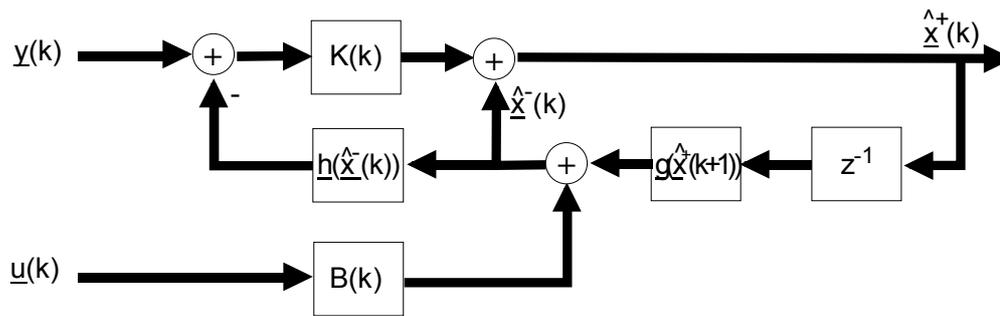


Abb. 2.28: Allgemeines nichtlineares Kalman-Filtermodell

Wohin bilden sich die beiden Rauschprozesse im Kalman-Filtermodell ab? Die beiden Rauschprozesse des Systemmodells werden nur noch im Kovarianzenzyklus durch deren Kovarianzen beschrieben.

Im folgenden soll das Filtermodell und dessen Funktionsweise näher betrachtet und mit Worten beschrieben werden: Das daraus abgeleitete Kalman-Filtermodell rechnet auf zwei Ebenen den kommenden Filter-Schätzwert der Meßreihe aus. Zum einen wird der Schätzwert der Systemvariablen x und zum anderen parallel die dazugehörige Verteilungsdichte betrachtet. Der Estimationsvorgang besteht aus zwei Schritten: Im ersten Schritt - der Prädiktion - wird die sog. Prädiktionsdichte $f_{x(k)|Y(k-1)}(\xi_k | Y_{k-1})$ berechnet. Diese Dichte wird nur aus schon existierender alter Information gewonnen. Sie wird durch die Prädiktionsfehlerkovarianz P^- beschrieben, da wir von gaußverteilten Größen ausgehen. Sie bestimmt die maximal mögliche Zustandsänderung eines nachfolgenden, vorbestimmten Zustands und setzt sich aus der "alten" Filterfehlerkovarianz und der Varianz des erzeugenden inneren Zustandsprozesses, Driving Noise genannt, zusammen. Diese Varianz oder Rauschleistung Q entspricht der Rauschleistung der Zufallsgröße $\underline{w}(k)$ des Systemmodells. Mit ihr werden Modellfehler linearer und nichtlinearer Problemstellungen mit einbezogen. Große Fehler bedingen eine große Varianz von Q .

Im zweiten Schritt wird die Filterdichte bestimmt. Diese enthält jetzt auch die neuen Meßwertinformationen. Beschrieben wird sie ebenfalls durch die ersten beiden Momente. Die Filterfehlerkovarianz P^+ wird durch die Prädiktionsfehlerkovarianz P^- und die angenommene Rauschleistung R der Störgröße $\underline{v}(k)$ der Messung bestimmt und wird als 'observation noise' bezeichnet. Dabei gehen R und P^- ebenfalls in die Bestimmung des Kalman-Gains ein. Dieser entscheidende Steuerparameter ist das Bindeglied vom Kovarianzenzyklus zum Schätzwertzyklus und stellt eine Zusammenfassung mehrerer Parameter dar. Am Kalman-

Gain kann direkt das Filterverhalten beobachtet werden. Er stellt die Gewichtung der Korrekturinformation $\hat{\underline{x}}_k^-$ zum vorhandenen Meßwert \underline{y}_k dar. Das Kalman-Filtermodell besteht folglich aus zwei parallel ablaufenden Schätzungen vom 1. Moment und 2. Moment der vorhandenen Verteilungsdichte, die durch einfließende Beobachtungen gesteuert werden. Die Prädiktion enthält die Dynamik des Systemmodells.

Herleitung: Für die Prädiktion wird eine Beziehung des gesuchten Zustandes $\hat{\underline{x}}_k$ zur vorhergehenden Messung ausgewertet. Diese wird durch die bedingte Verteilungsdichtefunktion $f_{\underline{x}(k)|\underline{Y}(k-1)}(\underline{\xi}_k|\underline{Y}_{k-1})$, der Prädiktionsdichte, beschrieben. Mit der Voraussetzung, daß der modellierende Prozeß $\underline{w}(k)$ und die bereits bekannte, bedingte Verteilungsdichte $f_{\underline{x}(k)|\underline{Y}(k)}(\underline{\xi}_k|\underline{Y}_k)$ gaußverteilt sind, läßt sich beweisen, daß $f_{\underline{x}(k)|\underline{Y}(k-1)}(\underline{\xi}_k|\underline{Y}_{k-1})$ ebenfalls gaußverteilt sein muß:

$$f_{\underline{x}(k)|\underline{Y}(k-1)}(\underline{\xi}_k|\underline{Y}_{k-1}) = \frac{1}{\sqrt{(2\pi)^n \cdot |\mathbf{P}_k^-|}} \cdot e^{-\frac{1}{2}(\underline{\xi}_k - \hat{\underline{x}}_k^-)^T \cdot (\mathbf{P}_k^-)^{-1} \cdot (\underline{\xi}_k - \hat{\underline{x}}_k^-)} \quad (2-101)$$

Die vollständige Beschreibung einer Gaußverteilungsdichte erfolgt über das 1. und 2. Moment, was zu den gesuchten beiden Prädiktionsgleichungen führt:

$$\mathbf{1. Moment:} \left\{ \begin{array}{l} \hat{\underline{x}}_k^- = E\left\{\underline{x}_k \mid \underline{y}_{k-1}\right\} = \int_{-\infty}^{+\infty} \underline{\xi}_k \cdot f_{\underline{x}(k)|\underline{Y}(k-1)}(\underline{\xi}_k|\underline{Y}_{k-1}) d\underline{\xi}_k \end{array} \right. \quad (2-102)$$

$$\mathbf{2. Moment:} \left\{ \begin{array}{l} \mathbf{P}_k^- = E\left\{\left(\underline{x}_k - \hat{\underline{x}}_k^+\right) \cdot \left(\underline{x}_k - \hat{\underline{x}}_k^+\right)^T \mid \underline{Y}(k-1) = \underline{Y}_{k-1}\right\} \\ = \int_{-\infty}^{+\infty} \left(\underline{\xi}_k - \hat{\underline{x}}_k^-\right) \cdot \left(\underline{\xi}_k - \hat{\underline{x}}_k^-\right)^T \cdot f_{\underline{x}(k)|\underline{Y}(k-1)}(\underline{\xi}_k|\underline{Y}_{k-1}) d\underline{\xi}_k \end{array} \right. \quad (2-103)$$

Im Filterschritt wird die gesuchte Filterdichte $f_{\underline{x}(k)|\underline{Y}(k)}(\underline{\xi}_k|\underline{Y}_k)$ bestimmt, die den Zusammenhang zu dem aktuellen Zustand \underline{x}_k und der Beobachtung \underline{y}_k herstellt. Sie wird bestimmt über die Anwendung des Satzes von Bayes:

$$f_{\underline{x}(k),\underline{Y}_k}(\underline{\xi}_k, \underline{Y}_k) = f_{\underline{Y}(k)}(\underline{Y}_k) \cdot f_{\underline{x}(k)|\underline{Y}(k)}(\underline{\xi}_k|\underline{Y}_k) \quad \text{oder:} \quad f_{\underline{x},\underline{Y}} = f_{\underline{Y}} \cdot f_{\underline{x}|\underline{Y}}$$

$$\begin{aligned}
 \Rightarrow f_{\underline{x}(k)|\underline{Y}(k)}(\underline{\xi}_k|\underline{Y}_k) &= f_{\underline{x}(k),\underline{Y}_k}(\underline{\xi}_k,\underline{Y}_k) \cdot \frac{1}{f_{\underline{Y}(k)}(\underline{Y}_k)} \\
 &= \frac{f_{\underline{x}(k),\underline{y}(k),\underline{Y}(k-1)}(\underline{\xi}_k,\underline{y}_k,\underline{Y}_{k-1})}{f_{\underline{y}(k),\underline{Y}(k-1)}(\underline{y}_k,\underline{Y}_{k-1})} \\
 &= \frac{f_{\underline{y}(k)|\underline{x}(k),\underline{Y}(k-1)}(\underline{y}_k|\underline{\xi}_k,\underline{Y}_{k-1}) \cdot f_{\underline{x}(k),\underline{Y}(k-1)}(\underline{\xi}_k,\underline{Y}_{k-1})}{f_{\underline{y}(k),\underline{Y}(k-1)}(\underline{y}_k,\underline{Y}_{k-1}) \cdot f_{\underline{Y}(k-1)}(\underline{Y}_{k-1})} \\
 &= \frac{f_{\underline{y}(k)|\underline{x}(k),\underline{Y}(k-1)}(\underline{y}_k|\underline{\xi}_k,\underline{Y}_{k-1}) \cdot f_{\underline{x}(k)|\underline{Y}(k-1)}(\underline{\xi}_k|\underline{Y}_{k-1}) \cdot f_{\underline{Y}(k-1)}(\underline{Y}_{k-1})}{f_{\underline{y}(k)|\underline{Y}(k-1)}(\underline{y}_k|\underline{Y}_{k-1}) \cdot f_{\underline{Y}(k-1)}(\underline{Y}_{k-1})} \\
 &= \frac{f_{\underline{y}(k)|\underline{x}(k),\underline{Y}(k-1)}(\underline{y}_k|\underline{\xi}_k,\underline{Y}_{k-1}) \cdot f_{\underline{x}(k)|\underline{Y}(k-1)}(\underline{\xi}_k|\underline{Y}_{k-1})}{f_{\underline{y}(k)|\underline{Y}(k-1)}(\underline{y}_k|\underline{Y}_{k-1})}
 \end{aligned}$$

Durch Anwenden des Logarithmus erhält man erneut die Entropien, die den Informationsfluß beschreiben (vgl. 2.4.1.1):

Nach dem Beweis der Gaußverteilttheit der Filterdichte werden ebenso wie bei der Prädiktion die ersten beiden Momente zur Beschreibung der bedingten Verteilungsdichte bestimmt:

$$\mathbf{1. Moment:} \left\{ \hat{\underline{x}}_k^+ = E\left\{\underline{x}_k|\underline{y}_k\right\} = \int_{-\infty}^{+\infty} \underline{\xi}_k \cdot f_{\underline{x}(k)|\underline{Y}(k)}(\underline{\xi}_k|\underline{Y}_k) d\underline{\xi}_k \right. \quad (2-104)$$

$$\mathbf{2. Moment:} \left\{ \begin{aligned} P_k^+ &= E\left\{\left(\underline{x}_k - \hat{\underline{x}}_k^+\right) \cdot \left(\underline{x}_k - \hat{\underline{x}}_k^+\right)^T \middle| \underline{Y}(k) = \underline{Y}_k\right\} \\ &= \int_{-\infty}^{+\infty} \left(\underline{\xi}_k - \hat{\underline{x}}_k^-\right) \cdot \left(\underline{\xi}_k - \hat{\underline{x}}_k^-\right)^T \cdot f_{\underline{x}(k)|\underline{Y}(k)}(\underline{\xi}_k|\underline{Y}_k) d\underline{\xi}_k \end{aligned} \right. \quad (2-105)$$

Diese führen nach zahlreichen Umformungen (Matrizenumformungen) unter Anwendungen unterschiedlicher Lemmata zu den unten aufgeführten Kalman-Filtergleichungen. Um ein nichtlineares Extended Linearised Kalman-Filter zu realisieren, sind teilweise Linearisierungen notwendig. (Im betrachteten Fall ist das Beobachtungsmodell nichtlinear, so daß hier die Beobachtungsmatrix linearisiert werden muß, hingegen die Residuen bleiben nichtlinear. In der Regel wird die Beobachtung $\underline{y}(\underline{x}(k))$ in eine Taylorreihe entwickelt, die bereits nach dem 1. Glied abgebrochen wird.) Zusammenfassend läßt sich das beschriebene, zeitdiskrete nichtlineare Filtermodell durch die folgenden Größen allgemein beschreiben:

Die allgemeinen Größen des zeitdiskreten Extended linearised Kalman-Filters lauten:

Prädiktionsschätzwert:	$\hat{\underline{x}}_{k+1}^- = \underline{g}(\hat{\underline{x}}_k^+) \cdot \hat{\underline{x}}_k^+ + B(k) \cdot \underline{u}(k)$
Prädiktionsfehlerkovarianz:	$P_{k+1}^- = A_k(\hat{\underline{x}}_k^+) \cdot P_k^+ \cdot A_k(\hat{\underline{x}}_k^+)^T + Q$
Residuen:	$\underline{r}_{k+1} = \underline{y}_{k+1} - \underline{h}_{k+1}(\hat{\underline{x}}_{k+1}^-)$
Filterschätzwert:	$\hat{\underline{x}}_{k+1}^+ = \hat{\underline{x}}_{k+1}^- + K_{k+1} \cdot \underline{r}_{k+1}$
Filterfehlerkovarianz:	$P_{k+1}^+ = P_{k+1}^- - K_{k+1} \cdot H_{k+1} \cdot P_{k+1}^-$

Linearisierte Beobachtungsmatrix:	$H_{k+1} = \frac{d}{d\underline{x}_{k+1}} \underline{h}_{k+1} \Big _{\hat{\underline{x}}_{k+1}^-}$
Kalman-Gain-Matrix:	$K_{k+1} = \left[P_{k+1}^- \cdot H_{k+1}^T \right] \cdot \left[H_{k+1} \cdot P_{k+1}^- \cdot H_{k+1}^T + R_{k+1} \right]^{-1}$

2.4.3.3 Linearisierung von Zustandsübergangsmatrix und Beobachtungsmatrix

Die bereits beschriebenen Linearisierungen werden folgendermaßen durchgeführt: Im allgemeinen verwendet man die allgemeine Taylorreihenentwicklung:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^n(x_0) \cdot (x - x_0)^n}{n!} \cong f(x_0) + \frac{df(x_0)}{dx} \cdot (x - x_0) \quad (2-106)$$

Nichtlinearitäten können bei der Zustandsübergangsmatrix aber auch zwischen Beobachtung und Schätzwert auftreten. In beiden Fällen wird mit Hilfe der 1. Näherung einer Taylorreihenentwicklung eine Linearisierung durchgeführt, da sich die Inversen der nichtlinearen Matrizen nicht bilden lassen /41/.

2.4.3.3.1 Linearisierung bei nichtlinearer Zustandsübergangsmatrix

Mit dem Ansatz der allgemeinen Systemgleichungen bei nichtlinearer Zustandsübergangsmatrix wird die Problematik graphisch aufbereitet und anhand des Kalman-Filterformelsatzes ausgewertet:

$$\begin{aligned}\underline{x}(k+1) &= \underline{f}(\underline{x}(k)) + \mathbf{B} \cdot \underline{u}(k) + \mathbf{G} \cdot \underline{w}(k) \\ \underline{y}(k) &= \mathbf{C} \cdot \underline{x}(k) + \underline{v}(k)\end{aligned}\quad (2-107)$$

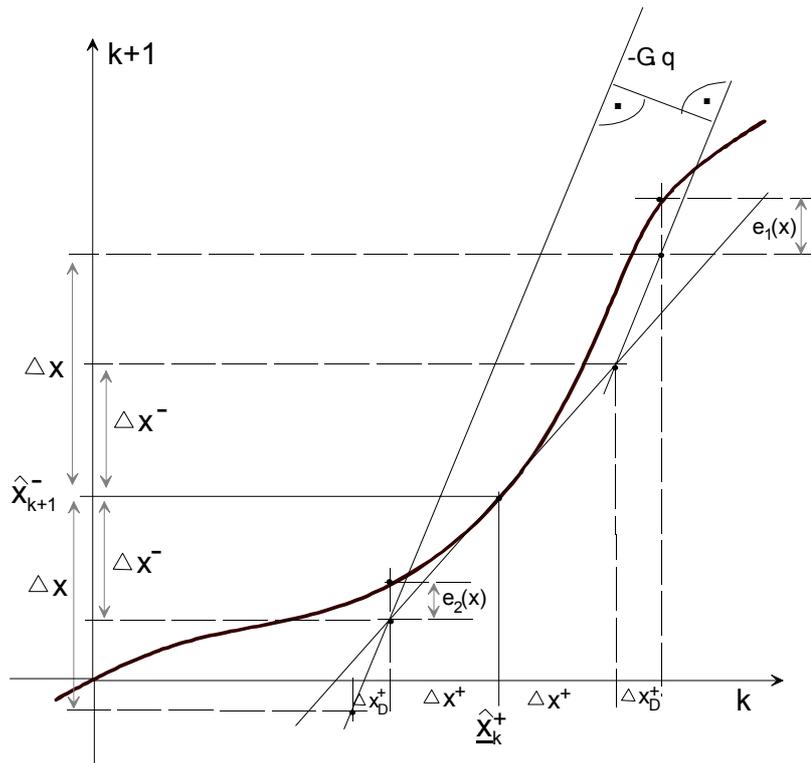


Abb. 2.29: Linearisierung eines nichtlinearen Zustandsübergangs

Die Anwendung einer Linearisierung wirkt sich im Kalman-Filter auch auf die Kovarianzen des Kalman-Filters aus. Prädiktion und Filterung werden als zwei getrennte Zeitpunkte betrachtet, zwischen denen die Linearisierung durchgeführt wird und auf diese Weise Fehler entstehen.

Die bereits definierte Filterfehlerkovarianz beschreibt den entstandenen Fehler zwischen dem korrekten Zustand und dem Filterschätzwert:

$$\mathbf{P}_k^+ = E \left\{ \Delta \underline{x}^+ \cdot (\Delta \underline{x}^+)^T \right\} \quad \text{mit: } \Delta \underline{x}^+ = \underline{x}_k - \hat{\underline{x}}_k^+$$

Bei der Bestimmung des nächsten Filterschätzwertes zum Zeitpunkt $k+1$ wird zuerst der Prädiktionsschritt durchgeführt. Dabei wird aufgrund des gegebenen nichtlinearen Zusammenhangs der Zustandsübergang mit einer Taylorreihe in 1. Näherung linearisiert. Es wird dabei um den Punkt $\hat{\underline{x}}_k^+$ entwickelt:

$$\underline{x}_{k+1} \cong \underline{f}(\underline{x}_k) + \underline{w}_k = \underline{f}(\hat{\underline{x}}_k^+) + \frac{d}{d\underline{x}_k} \underline{f}(\underline{x}_k) \Big|_{\underline{x}_k = \hat{\underline{x}}_k^+} \cdot (\underline{x}_k - \hat{\underline{x}}_k^+) + \dots + \underline{w}(k) \quad (2-108)$$

Die Differenz $\Delta \underline{x}^-$ von \underline{x}_{k+1} und $\hat{\underline{x}}_{k+1}^-$ beschreibt den Linearisierungsfehler, dabei ist $\hat{\underline{x}}_{k+1}^- = \underline{f}(\hat{\underline{x}}_k^+)$:

$$\Delta \underline{x}^- = \underline{x}_{k+1}^- - \hat{\underline{x}}_{k+1}^- = \frac{d}{d\underline{x}_k} \underline{f}(\underline{x}_k) \Big|_{\underline{x}_k = \hat{\underline{x}}_k^+} \cdot (\underline{x}_k - \hat{\underline{x}}_k^+) + \underline{w}_k \quad (2-109)$$

Um die Prädiktionsfehlerkovarianz zu erhalten, wird der Erwartungswert der quadrierten Differenz gebildet. Die Erwartungswerte der Kreuzterme ergeben Null, so daß sich folgendes Resultat ergibt:

$$\begin{aligned} P_k^- &= E \left\{ \Delta \underline{x}^- \cdot (\Delta \underline{x}^-)^T \right\} \\ &= \frac{d}{d\underline{x}_k} \underline{f}(\underline{x}_k) \Big|_{\underline{x}_k = \hat{\underline{x}}_k^+} \cdot E \left\{ \Delta \underline{x}^+ \cdot (\Delta \underline{x}^+)^T \right\} \cdot \frac{d}{d\underline{x}_k} \underline{f}(\underline{x}_k) \Big|_{\underline{x}_k = \hat{\underline{x}}_k^+} + G \cdot Q \cdot G^T \\ &= A'(\hat{\underline{x}}_k^+) \cdot P_k^+ \cdot A'^T(\hat{\underline{x}}_k^+) + G \cdot Q \cdot G^T \end{aligned} \quad (2-110)$$

Der nichtlineare Zustandsübergang wird als Matrix A' bezeichnet, was dem direkten Vergleich mit einem linearen System dient. Im Vergleich zum linearen Zustandsübergang erscheint anstelle von A die 1. Ableitung der Zustandsübergangsmatrix A .

Durch eine starke Änderung im Zustandsverlauf könnte der gesuchte korrekte Zustand im Bereich $\hat{\underline{x}}_{k+1}^- \pm \Delta \underline{x}^-$ nicht mehr zu finden sein, so daß der mögliche Bereich vergrößert wird auf den Bereich $\hat{\underline{x}}_{k+1}^- \pm \Delta \underline{x}$. Die Größe von $\Delta \underline{x}$ wird durch die Steigung der Geraden $G \cdot q$ (Diffusionskonstante q) bestimmt.

2.4.3.3.2 Linearisierung bei nichtlinearer Beobachtungsmatrix

Die Linearisierung bei nichtlinearem Zusammenhang zwischen Beobachtung und Zustand verläuft grundsätzlich analog, jedoch ergeben sich unterschiedliche Auswirkungen auf die

Kovarianzen. (Bei der Demodulation winkelmodulierter Signale ist ein derartiger nichtlinearer Zusammenhang zwischen Beobachtung und Zustand gegeben.) Ausgehend vom Prädiktionsschritt wird aus dem Prädiktionsschätzwert $\hat{\underline{x}}_k^-$ durch Transformation ein prädiktiver errechneter Meßwert \hat{y}_k^- bestimmt. Anschließend wird die Differenz zum realen Meßwert y_k gebildet:

$$\text{res}_k = \Delta y_k = y_k - \hat{y}_k^- \quad (2-111)$$

Dieses Residuum wird in den Zustandsraum zurücktransformiert, mit dem Kalman-Gain gewichtet und zum Prädiktionsschätzwert $\hat{\underline{x}}_k^-$ hinzuaddiert. Das Resultat ist der gesuchte Filterschätzwert $\hat{\underline{x}}_k^+$.

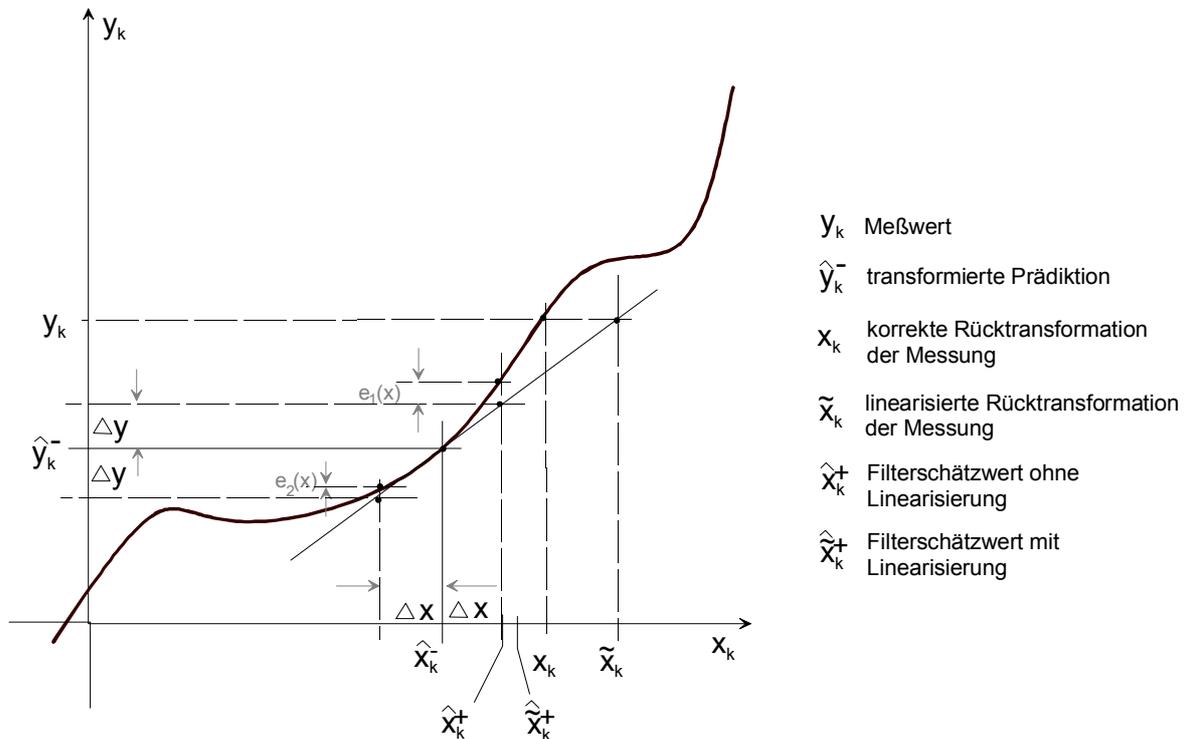


Abb. 2.30: Linearisierung einer nichtlinearen Beobachtung

$\hat{\underline{x}}_k^-$ ist in diesem Fall der Entwicklungspunkt für die Taylorreihenentwicklung der nichtlinearen Beobachtungsgleichung. Zur Betrachtung der Zusammenhänge genügt die Betrachtung des ungestörten Falls mit $\underline{v}_k = \underline{0}$:

Mit der Beobachtungsgleichung $\underline{y}_k = \underline{h}(\underline{x}_k) + \underline{v}_k$ erhalten wir:

$$\underline{h}(\underline{x}_k) = \underline{h}(\hat{\underline{x}}_k^-) + \frac{d}{d\underline{x}} \underline{h}(\underline{x}) \Big|_{\hat{\underline{x}}_k^-} \cdot (\underline{x}_k - \hat{\underline{x}}_k^-) + \dots$$

$$\underline{h}(\underline{x}_k) - \underline{h}(\hat{\underline{x}}_k^-) \cong \frac{d}{d\underline{x}} \underline{h}(\underline{x}) \Big|_{\hat{\underline{x}}_k^-} \cdot (\underline{x}_k - \hat{\underline{x}}_k^-) \quad \text{1.Naherung} \quad (2-112)$$

$$\Leftrightarrow \underline{y}_k - \hat{\underline{y}}_k^- \cong \frac{d}{d\underline{x}} \underline{h}(\underline{x}) \Big|_{\hat{\underline{x}}_k^-} \cdot (\underline{x}_k - \hat{\underline{x}}_k^-)$$

$$\Delta \underline{y}_k \cong H_{\text{diff}} \cdot \Delta \underline{x}_k \quad (2-113)$$

Ebenso lat sich der Zusammenhang zwischen Filterschatzwert und Mewert bestimmen. Als Ansatz wird die Gleichung zur Berechnung des Filterschatzwertes eines linearen Kalman-Filters angesetzt (unter der Bedingung, da keine Storung vorliegt: $R=0$):

$$\hat{\underline{x}}_k^+ = \hat{\underline{x}}_k^- + [P_k^- \cdot C^T] \cdot [C \cdot P_k^- \cdot C^T]^{-1} \cdot \Delta \underline{y}$$

Mit der Bedingung, da der Entwicklungspunkt $\hat{\underline{x}}_k^- = 0$ sein soll, ergibt sich allgemein /5/:

$$\hat{\underline{x}}_k^+ = K \cdot \Delta \underline{y} \quad \text{mit: } K = [P_k^- \cdot C^T] \cdot [C \cdot P_k^- \cdot C^T]^{-1} \quad (2-114)$$

Da $\hat{\underline{x}}_k^+$ hier identisch ist mit $\Delta \underline{x}$ (wg. $\hat{\underline{x}}_k^- = 0$), erkennt man allgemein im Vergleich mit der nichtlinearen Betrachtung, da allgemein - unabhangig vom Entwicklungspunkt - gilt:

$$\Delta \underline{x} = K \cdot \Delta \underline{y} \quad (2-115)$$

Fur die Kovarianzen ergibt sich nun:

$$P_k^- = E\{\Delta \underline{x} \cdot \Delta \underline{x}^T\} \quad \text{Pradiktionsfehlerkovarianz} \quad (2-116)$$

$$P_k^+ = [I - K \cdot H_{\text{diff}}] \cdot P_k^- \quad \text{Filterfehlerkovarianz} \quad (2-117)$$

$$\text{mit } K = E\{\Delta \underline{x} \cdot \Delta \underline{x}^T\} \cdot H_{\text{diff}}^T \cdot [H_{\text{diff}} \cdot E\{\Delta \underline{x} \cdot \Delta \underline{x}^T\} \cdot H_{\text{diff}}^T + R]^{-1}$$

Betrachtet man sich die Komponenten des Kalman-Gain genauer, so lat sich mit $\Delta \underline{y}_k \cong H_{\text{diff}} \cdot \Delta \underline{x}_k$ auch eine Abhangigkeit vom Mewert zum linearisierten transformierten

berechneten Meßwert entdecken :

$$\begin{aligned}
 \mathbf{K} &= \mathbf{E}\{\underline{\Delta x} \cdot \underline{\Delta x}^T\} \cdot \mathbf{H}_{\text{diff}}^T \cdot \left[\mathbf{E}\{(\mathbf{H}_{\text{diff}} \cdot \underline{\Delta x}) \cdot (\mathbf{H}_{\text{diff}} \cdot \underline{\Delta x})^T\} + \mathbf{R} \right]^{-1} \\
 &= \mathbf{E}\{\underline{\Delta x} \cdot \underline{\Delta x}^T\} \cdot \mathbf{H}_{\text{diff}}^T \cdot \left[\mathbf{E}\{\underline{\Delta y} \cdot \underline{\Delta y}^T\} + \mathbf{R} \right]^{-1} \\
 &= \mathbf{P}_k^- \cdot \mathbf{H}_{\text{diff}}^T \cdot \left[\mathbf{P}_y^- + \mathbf{R} \right]^{-1}
 \end{aligned} \tag{2-118}$$

\mathbf{P}_y^- bezeichnet die transformierte Prädiktionsfehlerkovarianz.

2.4.3.3.2.1 Betrachtung des Linearisierungsfehlers:

Grundsätzlich ist nur der gesuchte Schätzwert und dessen Fehler interessant für eine weitere Auswertung. Jedoch entstehen bei der Linearisierung der Beobachtungsmatrix Fehler im Beobachtungsraum. Ausgehend von dem transformierten prädiktiven Meßwert \hat{y}^- aus der Prädiktion \hat{x}^- ergibt sich über die Bestimmung des $\underline{\Delta x}$ die Möglichkeit, ein $\underline{\Delta y}$ zu berechnen. Durch Rücktransformation von $\hat{x}^- + \underline{\Delta x}$ ergibt sich $\hat{y}^- + \underline{\Delta y}$. Dieser Meßwert ist fehlerbehaftet durch die Approximation mit einer Taylorreihenentwicklung. Der entstehende Fehler lautet:

$$\mathbf{e}_1(\underline{x}) = (\hat{y}^- + \underline{\Delta y}) - y \tag{2-119}$$

Zur Berechnung des Fehlers sollen im folgenden skalare Größen verwendet werden. Es werden die bekannten Größen eingesetzt:

$$\mathbf{e}_1(x_k) = h(x_k^- + \Delta x) - h(x_k)$$

Dieser Fehler entspricht exakt den fehlenden Gliedern der Taylorreihenentwicklung, die bei der Approximation vernachlässigt wurden. Die Differenz des linearisierten und des nicht-linearisierten Terms ergibt:

$$\begin{aligned}
 \mathbf{e}_1(x_k) &= h(\hat{x}_k^-) + \mathbf{H}_{\text{Diff}} \cdot \Delta x - \\
 &\left[h(\hat{x}_k^-) + \frac{d}{dx} h(\hat{x}_k^-) \cdot (x_k - \hat{x}_k^-) + \frac{1}{2} \frac{d^2}{d^2 x} h(\hat{x}_k^-) \cdot (x_k - \hat{x}_k^-)^2 + \frac{1}{6} \frac{d^3}{d^3 x} h(\hat{x}_k^-) \cdot (x_k - \hat{x}_k^-)^3 + \dots \right]
 \end{aligned}$$

$$\Rightarrow e_1(x_k) = \frac{1}{2} \frac{d^2}{dx^2} h(\hat{x}_k^-) \cdot (x_k - \hat{x}_k^-)^2 + \frac{1}{6} \frac{d^3}{dx^3} h(\hat{x}_k^-) \cdot (x_k - \hat{x}_k^-)^3 + \dots$$

Diese lassen sich (vgl. /13/ S.279) zu dem Lagrange'schen Restglied zusammenfassen:

$$e_1(x_k) = \frac{d^2}{dx^2} h(\hat{x}_k^- + \vartheta(x_k - \hat{x}_k^-)) \cdot \frac{1}{2!} \cdot (x_k - \hat{x}_k^-)^2 \quad \text{mit } \vartheta \in (0,1)$$

$$e_1(x_k) = \frac{\Delta x_k^2}{2} \cdot \frac{d^2 h(\hat{x}_k^- + \vartheta \cdot \Delta x_k)}{dx^2} \quad \text{(2-120)}$$

Der maximale Fehler ist abhängig von der Wahl der Größe ϑ . Diese Größe hat ihren Ursprung im Mittelwertsatz der Differentialrechnung, so daß Werte im Bereich von $h(\hat{x}_k^-)$ bis $h(\hat{x}_k)$ betrachtet werden. (Mit den Nullstellen, der nach ϑ abgeleiteten Funktion $h'(\hat{x}_k)$ ergeben sich die Extremstellen, so daß sich das Maximum des Fehlers leicht bestimmen läßt!).

2.4.3.4 Filter höherer Ordnung

Werden beim Einsatz von nichtlinearen Kalman-Filtern auch Momente höherer Ordnung mit bedacht, so spricht man von Filtern höherer Ordnung. Die zur Näherung verwendete Taylorreihe wird dann nicht nur bis zum 1. Glied, sondern bis zu weiteren Gliedern berechnet. Bricht man nach dem 2. Glied ab, so nennt man das Filter ein Gaußfilter. Eine optimale Filterung von nicht gaußförmigen Verteilungsdichten wird so möglich.

Eine weitere Möglichkeit zur Verbesserung der Filterschätzwerte wird durch den Einsatz iterativer Kalman-Filter erreicht. Fehler, die durch die Linearisierung mit einer Taylorreihe entstanden sind, werden durch zusätzliche Filterdurchgänge (Prädiktionen) verringert. Schätzwert, Kalman-Gain und Prädiktionsfehler werden mehrfach bis zu einer vordefinierten Varianzgrenze und/oder Maximalanzahl der Durchläufe berechnet. Oft wird auch eine Mischung aus beiden Filtern gewählt, um die Genauigkeit des Schätzwertes zu verbessern.

In diesem Zusammenhang sei auch Stratonovitch erwähnt, der sich mit der Lösung nichtlinearer Gleichungen beschäftigt hat und dessen Lösungen bei Extended Kalman-Filtern ebenfalls eingesetzt werden /24/.

Es ergeben sich Filter, die sehr recht rechenintensiv und beim Einsatz von Online-Filtern eher ungeeignet sind. Die Genauigkeit wird dabei oft auch zu Lasten der Stabilität eines Filters verbessert, so daß in Bezug auf die Anwendung nicht notwendigerweise eine Verbesserung des Filters erreicht wird /5, 48, 50/.

2.4.4 Analogien von Extended Kalman-Filter und PLL (Phase Locked Loop):

Betrachtet man einen PLL - phasengesteuerten Regelkreis genauer, so werden Parallelen eines Extended Kalman-Filter zur Demodulation winkelmodulierter Signale deutlich /3, 9/. Die Zustandsmodellierung wird beim PLL offensichtlich vom VCO übernommen. Wie sieht es jedoch mit den weiteren Größen des Extended Kalman-Filters im Vergleich zum PLL aus? Da es sich beim PLL um eine skalare Betrachtung handelt, wird der Kalman-Filter ebenfalls skalar betrachtet. Ausgehend von einem allgemeinen FM modulierten Sendesignal (z.B. Radio) lautet das Sendesignal:

$$s(t) = \cos(2\pi f_T t + 2\pi\gamma \int_{-\infty}^t m(\tau) d\tau). \quad \text{FM-moduliertes Signal} \quad (2-121)$$

Das Empfangssignal wird dann idealisiert mit einem rein additiven Störsignal $v(t)$ überlagert, welches in seiner diskreten Form als weiß, mittelwertfrei und gaußverteilt definiert werden soll. Somit gilt für die Kovarianz $E\{v(k) \cdot v(l)\} = R \cdot \delta(k,l)$. Nach dem Herunterschieben auf eine Zwischenfrequenz (ZF) der Frequenz f_0 erhält man dann $y(t)$, welches abgetastet als Beobachtung des Kalman-Filters bzw. des PLL's zur Verfügung gestellt wird:

$$\begin{aligned} y(k) &= \cos(2\pi f_0 kT + 2\pi\gamma \cdot x(k)) + v(k) \\ &= h(x(k)) + v(k) \end{aligned} \quad \text{Beobachtungsgleichung} \quad (2-122)$$

Der Zustand $x(k)$ beschreibt die Phase des FM-Signals, die das gesuchte Quellsignal $m(k)$ enthält. Die Ableitung des Zustands $x(k)$ ergibt dabei $m(k)$. Mit der nichtlinearen Beobachtung $h(x(k))$ lautet die Gleichung:

$$y(k) = h(x(k)) + v(k) \quad \text{mit} \quad h(\hat{x}_k^-) = \cos(2\pi f_0 kT + 2\pi\gamma \cdot \hat{x}_k^-) \quad (2-123)$$

Bei der Modellbildung des Kalman-Filters wird eine Linearisierung der Beobachtung über Taylorreihenentwicklung notwendig, welches zum bereits bekannten H_{Diff} führt:

$$H_{\text{Diff}}(\hat{x}_k^-) = \frac{dh(\hat{x}_k^-)}{d\hat{x}_k^-} = -2\pi\gamma \cdot \sin(2\pi f_0 kT + 2\pi\gamma \cdot \hat{x}_k^-) \quad (2-124)$$

Die Filtergleichungen lauten dann:

$$\hat{x}_k^+ = \hat{x}_k^- + K_k \cdot \text{res}_k \quad \text{Filterschätzwert des EKF} \quad (2-125)$$

$$P_k^+ = P_k^- - P_k^- \cdot K_k \cdot H_{\text{Diff}} \quad \text{Filterfehlerkovarianz des EKF} \quad (2-126)$$

mit $\text{res}_k = y_k - h(\hat{x}_k^-)$ (Residuum) und $K_k = P_k^- \cdot H_{\text{Diff}} \cdot [H_{\text{Diff}} \cdot P_k^- \cdot H_{\text{Diff}} + R]^{-1}$ (Kalman-Gain)

Wählt man den Innovationsansatz zur Kalman-Filtermodellbeschreibung, der von der Residuensequenz ausgeht, so läßt sich durch Abwandlung ein Modell erzeugen, welches der Modellbeschreibung eines PLL's am meisten ähnelt (s. /3, 42/).

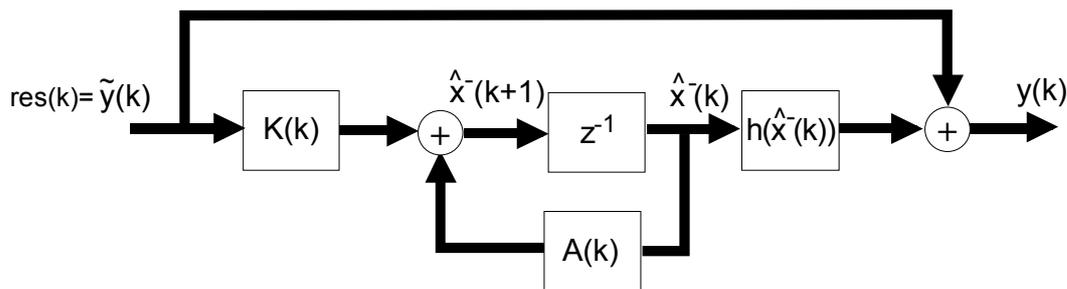


Abb. 2.31: Modell des Innovationsansatzes

Dazu wird die hintere Summenbildung von Abb. 2.31 in Abb. 2.32 nach vorne verlegt:

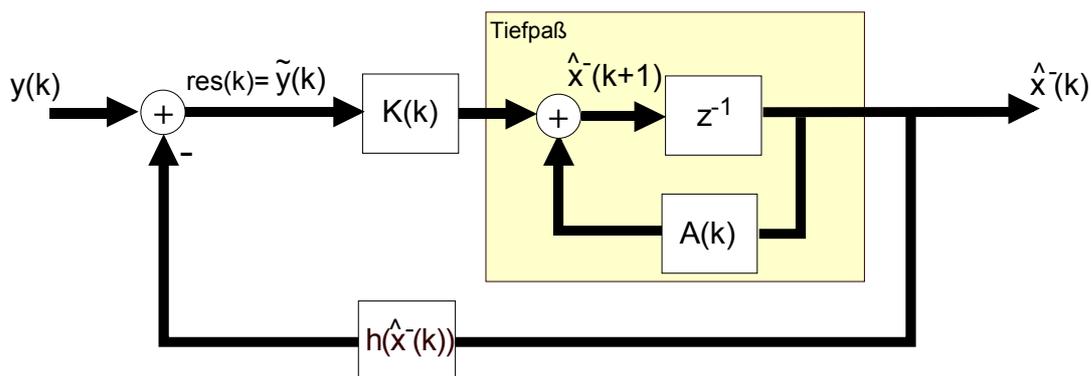


Abb. 2.32: PLL ähnliches Kalman-Filtermodell

Danach wird das Blockschaltbild des PLL's betrachtet:

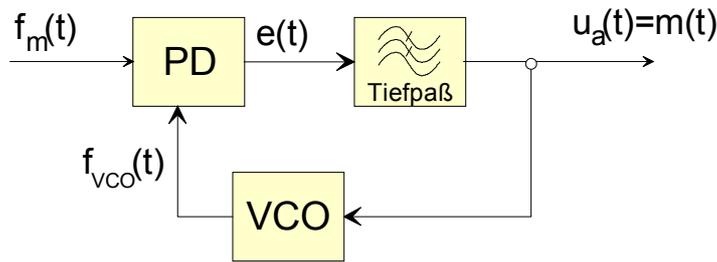


Abb. 2.33: PLL-Blockschaltbild

PLL	EKF zur FM-Demodulation
e(t)	$e(k) = \text{res}(k) \cdot K(k)$
VCO+90°-Phasenverschiebung	$h(\hat{x}^-(k))$
Tiefpaß	Zustandsgenerierung

Tab. 2.1: Analogien zwischen PLL und Extended Kalman-Filter

Die vergleichbare Ausgangsgröße des Phasendetektors (PD) am EKF $e(k)$ wird mit dem bekannten $y(k)$ folgendermaßen bestimmt. Durch Einsetzen von $y(k) = \cos(2\pi f_0 kT + 2\pi\gamma \cdot x(k)) + v(k)$ und Anwendung der Additionstheoreme ergibt sich (Dabei werden die trigonometrischen Umformungen im Hinblick auf eine Differenzbildung von Phase und Schätzphase durchgeführt):

$$\begin{aligned}
 e(k) &= \text{res}(k) \cdot K(k) = [y(k) - h(\hat{x}_k^-)] \cdot K(k) & (2-127) \\
 &= [\cos(2\pi f_0 kT + 2\pi\gamma \cdot x(k)) + v(k) - \cos(2\pi f_0 kT + 2\pi\gamma \cdot \hat{x}_k^-)] \cdot K(k) \\
 &= [\cos(2\pi f_0 kT + 2\pi\gamma \cdot x_k) + v(k) - \cos(2\pi f_0 kT + 2\pi\gamma \cdot \hat{x}_k^-)] \\
 &\quad \cdot \frac{[-2\pi\gamma \cdot \sin(2\pi f_0 kT + 2\pi\gamma \cdot \hat{x}_k^-) \cdot P^{-1}]}{[4\pi^2 \gamma^2 \cdot \sin^2(2\pi f_0 kT + 2\pi\gamma \cdot \hat{x}_k^-) \cdot P^{-1} + R]}
 \end{aligned}$$

Durch Taylorreihenentwicklung für das Kalman-Gain läßt sich der Nenner eliminieren, um so die entstehende Gleichung als Summe von Frequenzvielfachen der Trägerfrequenz f_0 ausdrücken zu können. Dazu wird die Sinus-Schwingung in H_{Diff} durch z substituiert und die Taylorreihenentwicklung nach (2-106) durchgeführt:

$$K(0) = 0 ; K'(0) = -\frac{(2\pi\gamma)^3 \cdot (P^-)^2}{R} ; K''(0) = 0$$

$$K(z) = -C_1 \cdot z + \text{h.o.t.} \quad \text{mit : } C_1 = \frac{(2\pi\gamma)^3 \cdot (P^-)^2}{R} \quad (2-128)$$

$$K(\hat{x}_k^-) = -C_1 \cdot \sin(2\pi f_0 kT + 2\pi\gamma \hat{x}_k^-) + \text{h.o.t.}$$

Terme höherer Ordnung (h.o.t.) zu betrachten ist wenig sinnvoll, da in der nachfolgenden Tiefpaßfilterung alle Terme mit Trägerfrequenz oder Vielfachen von f_0 herausgefiltert werden. Durch Einsetzen der Taylorreihe erhält man:

$$e(k) = \left[\cos(2\pi f_0 kT + 2\pi\gamma \cdot x_k) + v(k) - \cos(2\pi f_0 kT + 2\pi\gamma \cdot \hat{x}_k^-) \right] \cdot \left[-C_1 \cdot \sin(2\pi f_0 kT + 2\pi\gamma \hat{x}_k^-) + \text{h.o.t.} \right] \quad (2-129)$$

Durch Anwendung der Additionstheoreme läßt sich das Ergebnis in eine Summe von Schwingungstermen umformen:

$$e(k) = \left[-\frac{C_1}{2} \cdot \sin(2\pi\gamma \cdot x_k - 2\pi\gamma \cdot \hat{x}_k^-) \right] - \left[\frac{C_1}{2} \cdot \sin(4\pi f_0 kT + 2\pi\gamma \cdot x_k + 2\pi\gamma \cdot \hat{x}_k^-) \right] + \left[\frac{C_1}{2} \cdot \sin(4\pi f_0 kT + 4\pi\gamma \cdot \hat{x}_k^-) \right] - v_k \cdot \frac{C_1}{2} \cdot \sin(2\pi f_0 kT + 2\pi\gamma \cdot \hat{x}_k^-) + \text{h.o.t.}$$

$$= \left[-\frac{(2\pi\gamma)^3 \cdot (P^-)^2}{2R} \cdot \sin(2\pi\gamma \cdot x_k - 2\pi\gamma \cdot \hat{x}_k^-) \right] - \left[\frac{(2\pi\gamma)^3 \cdot (P^-)^2}{2R} \cdot \sin(4\pi f_0 kT + 2\pi\gamma \cdot x_k + 2\pi\gamma \cdot \hat{x}_k^-) \right] + \left[\frac{(2\pi\gamma)^3 \cdot (P^-)^2}{2R} \cdot \sin(4\pi f_0 kT + 4\pi\gamma \cdot \hat{x}_k^-) \right] - v_k \cdot \frac{(2\pi\gamma)^3 \cdot (P^-)^2}{2R} \cdot \sin(2\pi f_0 kT + 2\pi\gamma \cdot \hat{x}_k^-) + \text{h.o.t.}$$

Durch die skalare Betrachtung werden die durchgeführten Umformungen erst ermöglicht. Bei Untersuchung des ungestörten Betriebs ergibt sich für $v(k)=0$. Weiterhin können die Terme höherer Ordnung vernachlässigt werden. Somit ergibt sich für die betrachtete Größe $e(k)$:

$$\begin{aligned}
 e(k) = & \left[-\frac{(2\pi\gamma)^3 \cdot (P^-)^2}{2R} \cdot \sin(2\pi\gamma \cdot x_k - 2\pi\gamma \cdot \hat{x}_k^-) \right] - \\
 & \left[\frac{(2\pi\gamma)^3 \cdot (P^-)^2}{2R} \cdot \sin(4\pi f_0 kT + 2\pi\gamma \cdot x_k + 2\pi\gamma \cdot \hat{x}_k^-) \right] + \\
 & \left[\frac{(2\pi\gamma)^3 \cdot (P^-)^2}{2R} \cdot \sin(4\pi f_0 kT + 4\pi\gamma \cdot \hat{x}_k^-) \right].
 \end{aligned} \tag{2-130}$$

Nach Tiefpaßfilterung erhalten wir:

$$\begin{aligned}
 e(k) = & -\frac{(2\pi\gamma)^3 \cdot (P^-)^2}{2R} \cdot \sin(2\pi\gamma \cdot x_k - 2\pi\gamma \cdot \hat{x}_k^-) \\
 = & -\frac{(2\pi\gamma)^3 \cdot (P^-)^2}{2R} \cdot \sin(2\pi\gamma \cdot \Delta x)
 \end{aligned} \tag{2-131}$$

Unter der Annahme, daß die Differenz Δx von gesuchtem Zustand und Filterschätzwert klein sind, läßt sich $\sin(2\pi\gamma\Delta x)$ durch $2\pi\gamma\Delta x$ nähern:

$$e(k) = -\frac{(2\pi\gamma)^4 \cdot (P^-)^2}{2R} \cdot \Delta x \tag{2-132}$$

Betrachten wir die zu $e(k)$ äquivalente Größe des PLL-Blockschaltbildes (Abb. 2.33) mit den eingezeichneten Größen, so erhalten wir:

$$\begin{aligned}
 e(t) = & f_m(t) \cdot f_{VCO}(t) \\
 = & u_0 \cdot \cos(2\pi f_0 kT + 2\pi\gamma \int_{-\infty}^t m(\tau) d\tau) \cdot \sin(2\pi \hat{f}_0 kT + k \int_{-\infty}^t u_a(\tau) d\tau) \\
 = & \frac{u_0}{2} \cdot \sin\left(2\pi(f_0 - \hat{f}_0)kT + 2\pi\gamma \int_{-\infty}^t m(\tau) d\tau - k \int_{-\infty}^t u_a(\tau) d\tau\right) + \\
 & \frac{u_0}{2} \cdot \sin\left(2\pi(f_0 + \hat{f}_0)kT + 2\pi\gamma \int_{-\infty}^t m(\tau) d\tau + k \int_{-\infty}^t u_a(\tau) d\tau\right)
 \end{aligned} \tag{2-133}$$

Sind f_0 und \hat{f}_0 gleich:

$$e(t) = \frac{u_0}{2} \cdot \sin\left(2\pi\gamma \int_{-\infty}^t m(\tau) d\tau - k \int_{-\infty}^t u_a(\tau) d\tau\right) + \frac{u_0}{2} \cdot \sin\left(4\pi f_0 kT + 2\pi\gamma \int_{-\infty}^t m(\tau) d\tau + k \int_{-\infty}^t u_a(\tau) d\tau\right) \tag{2-134}$$

Vergleicht man $e(k)$ (s. (2-130)) mit $e(t)$ (s. (2-134)) des Extended Kalman-Filters, so wird direkt die Ähnlichkeit der beiden Gleichungen offensichtlich. Bei beiden Gleichungen bilden sich Terme 0-ter und doppelter Frequenz f_0 ; die Ansteuerung eines VCO's mit einer Steuerungsspannung führt zu einer Frequenzmodulation des Steuersignals $u_a(t)$.

Rauschsignale (tiefpaßbegrenzte) werden durch Mischen mit der Zwischenfrequenz auf ein Frequenzband hochgemischt, so daß störende Offsets durch eine Tiefpaßoperation eliminiert werden können. Der nachfolgende Tiefpaß muß folglich auf die Bandbreite des modulierten Signals abgestimmt sein. Terme doppelter Zwischenfrequenz können als redundant betrachtet werden, da ihr Informationsgehalt bereits im Basisband enthalten ist.

Für die zeitkontinuierliche PLL-Größe $e_T(t)$ ergibt sich:

$$e_T(t) = \frac{u_0}{2} \cdot \sin \left(2\pi\gamma \int_{-\infty}^t m(\tau) d\tau - k \int_{-\infty}^t u_a(\tau) d\tau \right) \quad (2-135)$$

Da für kleine Phasenwinkel $\sin \varphi \cong \varphi$, gilt im eingeschwungene, stabilen Zustand des PLL's:

$$e_T(t) \cong \frac{u_0}{2} \cdot \left[2\pi\gamma \int_{-\infty}^t m(\tau) d\tau - k \int_{-\infty}^t u_a(\tau) d\tau \right] \quad (2-136)$$

Dies bedeutet, daß im Phasendetektor tatsächlich die Differenz der beiden Phasen gebildet wurde und die beiden Phasen verglichen wurden. Die Änderung der Phasendifferenz muß nach dem Einschwingen zu Null werden:

$$\frac{d e_T(t)}{dt} \cong \frac{u_0}{2} \cdot [2\pi\gamma \cdot m(t) - k u_a(t)] = 0$$

$$\Rightarrow k u_a(t) = 2\pi\gamma \cdot m(t) \quad (2-137)$$

Hieraus ergibt sich direkt, daß $u_a(t)$ und $m(t)$ proportional sind, was beabsichtigt war. Wird der PLL nicht auf der definierten Zwischenfrequenz des Empfangssignals betrieben, wird $u_a(t)$ einen Spannungsoffset besitzen. Analog läßt sich das Einschwingen des EKF erklären, dessen Filterschätzwert \hat{x}_k^+ ebenfalls direkt die Phase bzw. Frequenz einer modulierten Schwingung abbildet.

2.5 INFORMATIONSTHEORIE

Aus der Definition des Kalman-Filters heraus ergeben sich weitere Fragen nach der Genauigkeit und des verarbeiteten Informationsgehalts. Außerdem stellt sich die gleiche Frage beim Vergleich unterschiedlicher Filteralgorithmen. Was ist Information? Der Informationsbegriff tritt in einer Vielzahl von unterschiedlichen Definitionen im technischen und nicht-technischen Bereich auf. Beschränken wir uns auf die Definition der Information aus dem naturwissenschaftlichen Bereich. Der Begriff der Information hängt eng mit dem Begriff der Entropie (Maß für Unordnung oder Ordnung) zusammen; viele Definitionen verwenden die Entropie deshalb auch als Maß für die Information. Zunächst wollen wir allgemeine Definitionen des Informationsbegriffes ansetzen, um anschließend auf die Betrachtung von Information im Zusammenhang mit der Kalman-Filterung winkelmodulierter Signale einzugehen /5/.

2.5.1 Definitionen

Die Informationsmenge (auch: Informationsgehalt) wird von Hartley (1928) als Logarithmus der Auftrittswahrscheinlichkeit eines Zeichens definiert:

$$I_i = -\ln p(x_i) \quad (2-138)$$

Da $p(x_i)$ nur zwischen 0 und 1 definiert ist, erhalten wir stets eine positive Informationsmenge I . Sie ist eine in der Nachrichtentechnik übliche Definition zur Beschreibung von Nachrichtenkanälen. Damit ist I proportional zur Bandbreite und zur Übertragungszeit.

Werden digitale Signale bearbeitet, so sind die Informationen binär, die Maßeinheit der Information wird in diesem speziellen Fall durch das BIT ausgedrückt. Nur im Fall, daß die Elementmenge bekannt ist, läßt sich eine Maßeinheit für die Information festlegen. Ist die Auftrittswahrscheinlichkeit $p(x_i)$ gleichverteilt, so gilt für N Werte: $p(x_i) = \frac{1}{N}$. Daraus folgt für die Information des Zeichens x_i :

$$\begin{aligned} I_i &= -\ln p(x_i) = -\text{ld} \left(\frac{1}{N} \right) \text{BIT} \\ &= +\text{ld} N = \frac{\ln N}{\ln 2} \text{BIT} \end{aligned} \quad (2-139)$$

$\frac{1}{\ln 2}$ stellt lediglich eine Konstante dar, die den binären Zahlenraum zum Ausdruck bringt. Die Informationsmenge I ist direkt abhängig von der Anzahl der Elemente N .

Treten keine gleichverteilten Elemente auf, so muß die Definition erweitert werden. Shannon (1949) hat die Informationsdefinition dahingehend verallgemeinert:

$$I_{\text{Shannon}} = \sum_{i=1}^N P(x_i) \cdot \text{ld} \frac{1}{P(x_i)} = - \sum_{i=1}^N P(x_i) \cdot \text{ld} P(x_i) \quad (2-140)$$

$$I_{\text{Shannon}} = - \frac{1}{\ln 2} \sum_{i=1}^N P(x_i) \cdot \ln P(x_i) \quad (\text{Schreibweise mit natürlichem Logarithmus})$$

(Bei Gleichverteilungen reduziert sich die Gleichung wieder auf Hartley's Definition.) Shannon's Definition wird auch als Informationsentropie oder Entropie bezeichnet, um den mathematischen Definitionscharakter der Information hervorzuheben.

Shannon's kontinuierliche Definition der Information (Entropie), eine verallgemeinerte Form von Hartley's Definition (s. (2-140)) lautet:

$$S = - \int_x \ln(f_x(\underline{\xi})) \cdot f_x(\underline{\xi}) \cdot d\underline{\xi} \quad \text{Shannon} \quad (2-141)$$

$$G = \int_x \ln \left(\frac{f_x(\underline{\xi})}{f_{I_x}(\underline{\xi})} \right) \cdot f_x(\underline{\xi}) \cdot d\underline{\xi} \quad \text{Kullback} \quad (2-142)$$

2.5.2 Informationsgrenze - Cramér-Rao-Grenze

Mit der Cramér-Rao-Grenze wird die maximal erreichbare Varianzgrenze bei der Estimation definiert. Daß die Cramér-Rao-Bound tatsächlich die minimale Fehlervarianzgrenze darstellt, läßt sich beweisen, indem man eine Funktion $t_y(\underline{\rho})$ definiert, so daß $\underline{\xi} = \underline{x}$ (unbiased Estimator). Wir bilden den bedingten Erwartungswert von $t_y(\underline{\rho})$ bedingt auf \underline{x} :

$$\int_P t_y(\underline{\rho}) \cdot f_{y|x}(\underline{\rho}|\underline{\xi}) \cdot d\underline{\rho} = \underline{\xi} \quad |^* \quad (2-143)$$

* Die Integration erfolgt über P (groß Rho)

Durch Umstellen der Gleichung und unter der Bedingung, daß $\underline{\xi}$ keine Funktion von $\underline{\rho}$ ist und außerdem das Integral über eine Verteilungsdichtefunktion gleich eins ist, dann gilt:

$$I_1 = \int_{\underline{p}} \left(t_{\underline{y}}(\underline{\rho}) - \underline{\xi} \right) \cdot f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \cdot d\underline{\rho} = \underline{0}$$

Wir bilden die Ableitung des Integrals I_1 nach $\underline{\xi}$:

$$\begin{aligned} \frac{\partial}{\partial \underline{\xi}} I_1(\underline{\rho}, \underline{\xi}) &= \frac{\partial}{\partial \underline{\xi}} \int_{\underline{p}} \left(t_{\underline{y}}(\underline{\rho}) - \underline{\xi} \right) \cdot f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \cdot d\underline{\rho} = \underline{0} \\ &= - \int_{\underline{p}} f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \cdot d\underline{\rho} + \int_{\underline{p}} \left(t_{\underline{y}}(\underline{\rho}) - \underline{\xi} \right) \cdot \frac{\partial}{\partial \underline{\xi}} f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \cdot d\underline{\rho} \\ &= - \int_{\underline{p}} f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \cdot d\underline{\rho} + \int_{\underline{p}} \left(t_{\underline{y}}(\underline{\rho}) - \underline{\xi} \right) \cdot \left[\frac{\partial}{\partial \underline{\xi}} \ln f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \right]^T \cdot f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \cdot d\underline{\rho} \\ I &= + \int_{\underline{p}} \left(t_{\underline{y}}(\underline{\rho}) - \underline{\xi} \right) \cdot \left[\frac{\partial}{\partial \underline{\xi}} \ln f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \right]^T \cdot f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \cdot d\underline{\rho} \quad I^8 \end{aligned}$$

Eine weitere Zusammenfassung ist nicht möglich, so daß man die Schwartz'sche Ungleichung zur Interpretation der Erwartungswerte ansetzen kann:

$$\begin{aligned} &\Rightarrow \\ I &\leq + \int_{\underline{p}} \left(t_{\underline{y}}(\underline{\rho}) - \underline{\xi} \right) \cdot \left(t_{\underline{y}}(\underline{\rho}) - \underline{\xi} \right)^T \cdot f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \cdot d\underline{\rho} \cdot \int_{\underline{p}} \left[\frac{\partial}{\partial \underline{\xi}} \ln f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \right] \cdot \left[\frac{\partial}{\partial \underline{\xi}} \ln f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \right]^T \cdot f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \cdot d\underline{\rho} \\ &\Leftrightarrow \\ I \cdot \left[\int_{\underline{p}} \left[\frac{\partial}{\partial \underline{\xi}} \ln f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \right] \cdot \left[\frac{\partial}{\partial \underline{\xi}} \ln f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \right]^T \cdot f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \cdot d\underline{\rho} \right]^{-1} &\leq + \int_{\underline{p}} \left(t_{\underline{y}}(\underline{\rho}) - \underline{\xi} \right) \cdot \left(t_{\underline{y}}(\underline{\rho}) - \underline{\xi} \right)^T \cdot f_{\underline{y}|\underline{x}}(\underline{\rho}|\underline{\xi}) \cdot d\underline{\rho} \end{aligned}$$

Auf der rechten Seite steht die Varianz der gaußverteilten, erwartungswertfreien Schätzgröße.

⁸ I=Einheitsmatrix

$$I \cdot \left[\int_{\underline{p}} \left[\frac{\partial}{\partial \underline{\xi}} \ln f_{y|x}(\underline{p}|\underline{\xi}) \right] \cdot \left[\frac{\partial}{\partial \underline{\xi}} \ln f_{y|x}(\underline{p}|\underline{\xi}) \right]^T \cdot f_{y|x}(\underline{p}|\underline{\xi}) \cdot d\underline{p} \right]^{-1} \leq \text{var}(t_y(\underline{p})) \quad (2-144)$$

Dieser Grenzwert, den die Varianz der Zustandsschätzung nicht unterschreiten kann, ist die Cramér-Rao-Grenze, die sich bei Kenntnis aller stochastischen und deterministischen Kenngrößen ergibt. Sie ist über die inverse Matrix der Fisher'schen Information definiert, die auf der linken Seite der Formel steht.

$$\text{var}(\hat{\underline{x}}) \geq F^{-1}$$

$$[F]^{-1} = \left[\int_{\underline{p}} \left[\frac{\partial}{\partial \underline{\xi}} \ln f_{y|x}(\underline{p}|\underline{\xi}) \right] \left[\frac{\partial}{\partial \underline{\xi}} \ln f_{y|x}(\underline{p}|\underline{\xi}) \right]^T \cdot f_{y|x}(\underline{p}|\underline{\xi}) \cdot d\underline{p} = 0 \right]^{-1} \quad (2-145)$$

Die Fisher'sche Information gibt uns ebenfalls einen Hinweis auf die Kreuzinformation zwischen den Komponenten des Zufallsvektors. Die Cramér-Rao-Grenze ist ganz allgemein die unterste mögliche Varianzgrenze, unabhängig von der Verteiltheit der Störgrößen.

Bei der hier vorkommenden Cramér-Rao-Grenze wird die Phasenfehlerverteilungsdichte untersucht, die sich bei unterschiedlichen SNR's ergibt. Dazu berechnet man die Steigungen an den Stellen der Maxima der Verteilungsdichtefunktionen zu unterschiedlichen SNR's. Aufgrund der ausgewählten Extremstellen müssen die Steigungen alle negativ sein. Die positivierten Werte werden dann als Fisher's Information **F** bezeichnet. Der reziproke Verlauf bezeichnet die Cramér-Rao-Grenze (kleinstmögliche Varianzgrenze). Diese Varianzwerte stellen direkt Varianzen von Gaußverteilungen dar, die man idealisiert zur Beschreibung der Varianzen des Phasenfehlers einsetzen kann (s. 3.2.1.1).

Die Meßinformation steigt im Verlauf der Messungen selber auf einen festen Wert an. Im Gegenzug nimmt die Fehlerinformation ab, die sich durch die Inverse der Fehlerkovarianz P_f bestimmen läßt /5/.

$$C^T \cdot R^{-1} \cdot C = \mathbf{F} \quad \text{Fisher's Information (allgemein)} \quad (2-146)$$

Der Informationsfluß im Filter läßt sich folgendermaßen zusammenfassen:

$$[P_k^+]^{-1} = [P_k^-]^{-1} + C^T \cdot R^{-1} \cdot C \quad (2-147)$$

(Filterinformation=Prädiktionsinformation+Meßinformation)

2.5.3 Informationsbegriff bei der Demodulation winkelmodulierter Signale

Bei der Demodulation winkelmodulierter Signale mit Extended Kalman-Filtern gilt es, Schätzwerte der Phase auf ihren Informationsgehalt hin zu untersuchen. Außerdem wird der Informationsbegriff in einem weiteren Bereich angewendet: Wir sprechen bei dem übertragenen Nutzsignal auch von Transinformation oder übertragener Information.

Im ersten Fall gilt es, die Wahrscheinlichkeit der richtigen Realisierung eines Schätzwertes in Bezug auf die Meßgröße zu betrachten und mit dem Begriff der Information zu belegen. Ist die Wahrscheinlichkeit, eine korrekte Estimation durchgeführt zu haben, sehr hoch, dann ist auch die Information dieses Schätzwertes sehr hoch, d.h. die Fehlervarianz der Verteilungsdichtefunktion muß sehr klein gewesen sein. Es handelt sich um eine statische Betrachtung eines Zeitpunktes. Bei einer starken Störung $n(t)$ des Signals, d.h. bei einer großen Varianz, ist die korrekte Estimation sehr "gewagt". Die Information, die wir zu diesem Schätzwert erhalten, ist recht gering. Oder anders herum gesagt, ist die Rauschleistung (Varianz) gering, so wird der ermittelte Schätzwert recht genau sein, d.h. die Information wird groß sein.

$$r(t) = A \cdot \cos(2\pi f_0 t + \varphi(t)) + n(t) \quad (2-148)$$

Im zweiten Bereich geht es um die Betrachtung der zu übertragenden Nutzinformation von $\varphi(t)$, einer dynamischen Betrachtung. Die Information einer Übertragung wächst mit der steigenden Modulation des Trägers. Bei fehlender Modulation ($\varphi(t) = 0$) des Trägers ist die Information Null und wächst dann bei Modulation ($\varphi(t) = m(t)$) an. Maximal wird sie bei Modulation mit einem möglichst zufälligen, abwechslungsreichen Signal, zu einem Rauschsignal ($m(t) = w(t)$). Bei Modulation mit einem gaußverteilten Rauschsignal läßt sich die Information bestimmen, da wir die Kenntnis der Verteilungsdichtefunktion vollständig über Varianz und Mittelwert besitzen.

Wie wir sehen, ist in den beiden zu betrachtenden Bereichen, wo man den Informationsbegriff verwenden kann, entgegengesetztes Verhalten zu beobachten. Bei steigender Varianz des jeweiligen Rauschprozesses steigt bzw. fällt der Informationsgehalt.

Im nächsten Abschnitt wird die Stabilitätsbetrachtung ebenfalls anhand der Betrachtung des Informationszuwachses durchgeführt. Ebenso finden wir weitere Anwendungen und Betrachtungen der Fisher'schen Information unter 4.4 wieder.

2.6 STABILITÄTSBETRACHTUNG DER KALMAN-FILTER

Bei der Betrachtung der Kalman-Filter muß die Stabilität und Beobachtbarkeit der zugrunde liegenden Zustandsraummodelle grundsätzlich auch gegeben sein. Darauf soll hier jedoch nicht weiter eingegangen werden. Grundsätzlich sind Kalman-Filter asymptotisch stabil, wenn sie stochastisch beobachtbar und stochastisch steuerbar sind. Das bedeutet, daß bei einer ungünstig gewählten Initialisierung des Filters dieses trotzdem auf einen Grenzwert zuläuft, da der Initialisierungswert von Durchlauf zu Durchlauf geringeren Einfluß auf den Schätzwert hat.

Beim Kalman-Filter spielt dagegen die Divergenz eine wichtigere Rolle. Mit ihr wird ein Auseinanderlaufen von Schätzwert und korrektem Wert beschrieben, welches aufgrund von Quantisierungsfehlern, Rundungsfehlern (Rechnerungenauigkeit), einem ungenauen Systemmodell oder einem unkorrekten Driving Noise Parameter entstehen kann /38/.

2.6.1 Beobachtbarkeit

Aussagen über stochastische Beobachtbarkeit des Filters können grundsätzlich über die Untersuchung des Ranges der Matrix M (Gram'sche-Matrix) gewonnen werden. Dies gilt für lineare und nichtlineare Systeme. Die Bedingung lautet also:

$$\text{rang}\{M(k_f, k_0)\} = n \quad \text{Bedingung für Beobachtbarkeit (2-149)}$$

$$\text{Mit : } M(k_f, k_0)^T = \sum_{k=k_0}^{k_f} \phi(k, k_0)^T \cdot C(k)^T \cdot C(k) \cdot \phi(k, k_0) \quad \text{Beobachtbarkeitsmatrix (2-150)}$$

Die Grenze k_f wird begrenzt durch die Anzahl der vorhandenen Zustände. Sie beschreibt die minimale Anzahl der zeitlichen Schritte, die nötig sind, die vollständige Auswirkung der einzelnen Zustandsänderungen auf die Beobachtung/Messung zu beschreiben.

Der Rang dieser Matrix muß also gleich der Matrixgröße sein, so daß die Invertierbarkeit der Matrix ebenfalls gegeben ist.

Bei der Betrachtung von zeitinvarianten Systemen vereinfacht sich $\phi(k, k_0)$ zur konstanten Zustandsübergangsmatrix A und somit auch die Untersuchung der Matrix M . Die Summe der Matrizen läßt sich auch vereinfachen, indem man sich vorstellt, die Multiplikation einer Matrix C_a mit deren Transponierten C_a^T zu betrachten. Der Rang beider muß gleich sein und

dem Rang der Matrix M entsprechen, so daß sich der Rechenaufwand der Matrizenmultiplikationen auf ein Drittel reduziert. Wir untersuchen also nur den Rang der Matrix C_a zur Überprüfung der Beobachtbarkeit (notwendige und hinreichende Bedingung laut /38/).

$$\text{rang}\{M(k_f, k_0)\} = \text{rang}\{C_a(k_f, k_0)\} = \text{rang} \left\{ \begin{bmatrix} C(k_0) \\ C_a(k_1) \cdot A \\ \vdots \\ C(k_f) \cdot A^f \end{bmatrix} \right\} = n \quad (2-151)$$

Neben der hier betrachteten deterministischen Beobachtbarkeit, die stochastische Störgröße wurde durch $R = E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ realisiert, läßt sich auch über die Betrachtung der Fisher'schen Information die Beobachtbarkeit untersuchen. Zwar enthält die Fisher'sche Information, direkt betrachtet, keine Zustandsübergänge. Der kontinuierliche Anstieg der verarbeiteten Information gibt jedoch Aufschluß über die vorhandene Stabilität des Filtermodells. Die Bestimmung der Fisher'schen Information bedarf zwar meistens eines größeren Rechenaufwandes, aber neben der Beobachtbarkeitsaussage erhält man zusätzlich die verarbeitete Informationsmenge.

Diese wird bei der Berechnung des Kalman-Gains bereits bestimmt und kann somit fortlaufend ohne weiteren Aufwand mituntersucht werden (s. 2.5). Die allgemeine Definition lautet:

$$F = C^T \cdot R^{-1} \cdot C$$

Auch bei nichtlinearen Systemen läßt sich die Stabilität derart über die Betrachtung der Information untersuchen:

$$F = H_{\text{Diff}}^T \cdot R^{-1} \cdot H_{\text{Diff}} \quad (2-152)$$

Fügt man bei der Betrachtung der Beobachtbarkeitsmatrix die stochastische Störgröße mit ein, so wird der nahe Zusammenhang der beiden Methoden sichtbar :

$$M(k_f, k_0)^T = \sum_{k=k_0}^{k_f} A(k, k_0)^T \cdot H_{\text{Diff}}(k)^T \cdot R^{-1} \cdot H_{\text{Diff}}(k) \cdot A(k, k_0)$$

Im hier betrachteten Fall sind die beiden Rauschprozesse unkorreliert, so daß sich nur ein Vorfaktor, die reziproke Beobachtungsrauschleistung, ergibt.

2.6.2 Steuerbarkeit

Die hier als Stabilitätsbedingung geforderte stochastische Steuerbarkeit entspricht der stochastischen Erreichbarkeit des Systems für nicht singuläre Zustandsübergangsmatrizen. Diese läßt sich über den Rang der Matrix W_D bestimmen. Es wird hier der zweite Teil der allgemeinen Systemgleichung untersucht:

Läßt sich mit Hilfe des Steuervektors ein beliebiger Ausgangszustand des Systems erreichen?

$$\text{rang}\{W_D(k_f, k_0)\} = n \quad \text{Bedingung für vollständige Erreichbarkeit}$$

$$\text{mit: } W_D(k_f, k_0) = B_A(k_f, k_0) \cdot B_A(k_f, k_0)^T$$

Dies läßt sich vereinfachen zu der Bedingung, daß der Rang der Matrix B_A gleich n sein muß (notwendige und hinreichende Bedingung laut /38/).

$$\text{rang}\{B_A(k_f, k_0)\} = \text{rang}\{\phi(k_f, k_0 + 1) \cdot B(k_0) \cdot \dots \cdot B(k_f - 1)\} = n$$

3 Kalman-Filter Modellierung zur Demodulation winkelmodulierter Signale

Der Einsatz eines Kalman-Filters zur Demodulation winkelmodulierter Signale läßt sich grundsätzlich mit einem nichtlinearen Kalman-Filter, einem sogenannten Extended Linearised Kalman-Filter - EKF - unter Verwendung der Quadraturkomponenten im Basisband realisieren. Dieses Kalman-Filter mit zwei Beobachtungseingängen und zwei Filterschätzwerten am Ausgang, soll im folgenden als EKF22 bezeichnet werden. Es ist der Ausgangspunkt unserer weiteren Filtermodelle und soll somit als Standard und Bezugsgröße definiert werden. Zunächst wird ausführlicher auf die Bestimmung des Beobachtungs- und Systemmodells eingegangen. Im Anschluß daran folgen weitere abweichende Filtermodelle, bei denen - basierend auf den EKF22 - Modifikationen vorgenommen wurden, um die Effektivität und Genauigkeit des Filters zu steigern. Dies soll auch unter dem Blickwinkel einer Online-Verarbeitung geschehen. Dazu zählen unter anderem die Erweiterung der Beobachtungen mit sog. virtuellen Meßwerten, eine adaptive Verbesserung der Systemvarianz und eine Betrachtung von Smootheralgorithmen.

Zur Nomenklatur: Die folgenden Extended Linearised Kalman-Filter werden abgekürzt mit **EKF xyz** . ' x ' bezeichnet die Anzahl der Beobachtungen, ' y ' bezeichnet die Anzahl der Zustände und unter ' z ' werden optional Zusatzkürzel zur weiteren Klassifizierung verwendet. Adaptive und Smoother-Modelle werden mit **AKF xyz** bzw. mit **SKF xyz** abgekürzt. (Eine vollständige Liste der verwendeten Abkürzungen befindet sich im Anhang.)

3.1 EXTENDED KALMAN-FILTER ZUR PHASENDEMULATION

3.1.1 Standardfilter - EKF22 (2 Beobachtungen)

3.1.1.1 Beobachtungsmodell

Grundlage der folgenden Kalman-Filtermodelle ist das nachfolgend beschriebene Beobachtungsmodell. Als Meßwert stehen die Quadraturkomponenten eines winkelmodulierten

Signals im Basisband zur Verfügung. Wie bereits unter 2.1.3 bewiesen, können die abgetasteten Störsignalanteile $v_1(k)$ und $v_2(k)$ des Real- und Imaginärteils im Basisband weiterhin als weiß und gaußverteilt betrachtet werden, sofern ein weißes, gaußverteiltes Rauschsignal als Störgröße am Übertragungskanal angesetzt wurde.

Daraus läßt sich das Beobachtungsmodell bestimmen:

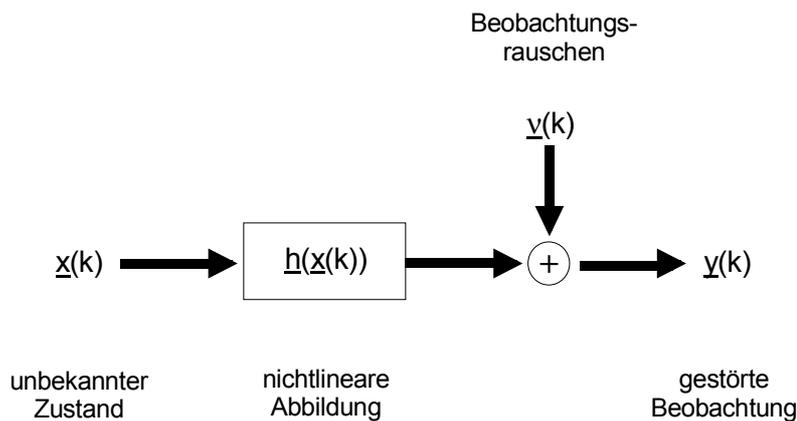


Abb. 3.1: Beobachtungsmodell

Die Winkelmodulation, Amplitudendämpfung η und Zeitverzögerung t_0 des Empfangssignals $r(t)$ wird im Modell als nichtlineare Abbildung und das Nutzsignal durch den Zustand $x(k)$ beschrieben. Die beiden Quadraturkomponenten bilden die gestörten Beobachtungsgrößen $y_1(k)$ und $y_2(k)$:

$$\begin{aligned} y_1(k) &= \text{Re}\{r_T(k)\} = \eta \cdot \cos(2\pi\gamma m_1 kT + \varphi_0) + v_1(k) \\ y_2(k) &= \text{Im}\{r_T(k)\} = \eta \cdot \sin(2\pi\gamma m_1 kT + \varphi_0) + v_2(k) \end{aligned} \quad (3-1)$$

Im Phasenoffset φ_0 sind Zeitverzögerung und anfänglicher Phasenoffset inbegriffen (s. 2.1.3). Da bei der Kalman-Filterung grundsätzlich Kurvenverläufe rekonstruiert werden, ist es ohne vorherige Kenntnis einer korrekten Anfangsphase (bedingt eine Definition von Startwerten) nicht möglich, den korrekten absoluten Phasenwert zu schätzen. Es werden nur die Änderungen betrachtet. Aus diesem Grunde kann der Phasenoffset φ_0 bei der Beschreibung der Beobachtungsgleichung weggelassen werden. Die nichtlineare Beobachtungsgleichung vektoriell zusammengefaßt lautet:

$$\underline{y}(k) = \underline{h}(\underline{x}(k)) + \underline{v}(k) \quad \text{mit :}$$

$$\Rightarrow \underline{y}(k) = \begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix} = \begin{bmatrix} \cos[2\pi\gamma \cdot x_1(k)] \\ \sin[2\pi\gamma \cdot x_1(k)] \end{bmatrix} + \underline{v}(k) \quad (3-2)$$

mit : $E\{\underline{v}(k)\} = \underline{0}$ und $\text{cov}(\underline{v}(k), \underline{v}(j)) = R \cdot \delta_{k,j}$

Ferner ist die Erzeugung des Nutzsignals wichtig für die Erstellung des Systemmodells. Sie wird im Zustandsraummodell beschrieben.

3.1.1.2 Zustandsraummodell

Gesucht ist eine Differenzgleichung, die das zustandsgenerierende System beschreibt. Die zeitdiskrete Systemgleichung, die rekursiv formuliert ist, lautet zunächst allgemein:

$$\underline{x}(k+1) = A(k) \cdot \underline{x}(k) + \underline{u}(k) + G(k) \cdot \underline{w}(k); \text{ mit } G(k) = I \text{ (Einheitsmatrix)} \quad (3-3)$$

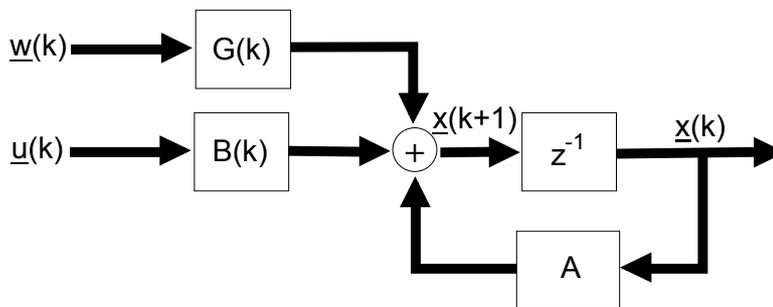


Abb. 3.2: Zustandsmodell

Die gesuchte Zustandsgröße $\underline{x}(k)$ stellt hier die modulierende Phase dar. Sie beschreibt das Nutzsinal. Da der Verlauf der Phase als völlig beliebig angenommen wird, eignet sich ein Markov-Pozeß für die Zustandsgenerierung. Die Änderung des Signalverlaufs wird dabei durch eine gaußverteilte Zufallsgröße bestimmt:

$$\dot{x}(t) = w(t)$$

Zeitdiskret formuliert führt diese Differentialgleichung 1. Ordnung zu einem Gleichungssystem mit zwei Gleichungen:

$$\begin{aligned} x_1(k+1) &= x_1(k) + x_2(k) \\ x_2(k+1) &= x_2(k) + w(k) \end{aligned} \quad (3-4)$$

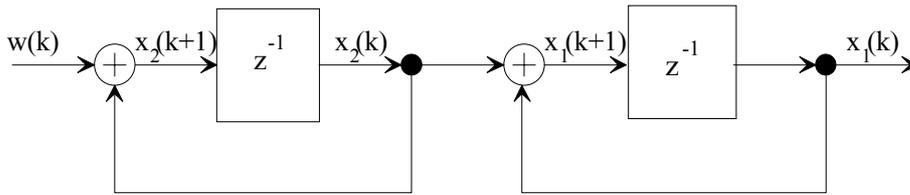


Abb. 3.3: Markov-Modell

Der Zustandsvektor $\underline{x}(k)$ lautet:

$$\underline{x}(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix}$$

Phase

Phasenänderung $\hat{=}$ Frequenz

$x_1(k)$ stellt die Phase und $x_2(k)$ die Phasenänderung oder Momentanfrequenz dar. Für die Phasenänderung $\dot{\varphi}(t)$ gilt:

$$\dot{\varphi}(t) = \omega(t) = 2\pi f_1(t) \tag{3-5}$$

Daraus ergibt sich die Zustandsgleichung:

$$\underline{x}(k+1) = A(k) \cdot \underline{x}(k) + \underline{u}(k) + G(k) \cdot \underline{w}(k)$$

$$\Rightarrow \begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} w_1(k) \\ w_2(k) \end{bmatrix} \tag{3-6}$$

Zustandsgleichung

$A(k)$ stellt die Zustandsübergangsmatrix und $G(k)$ stellt die Gewichtungsmatrix für den Rauschprozeß $w(k)$ dar. Die Größe $\underline{u}(k)$, hier Null, bezeichnet bekannte, überlagerte Größen, die hier zunächst nicht vorhanden sein sollen. Die Zufallsvariable $w_2(k) = w(k)$ ist definiert über deren Verteilungsdichte bzw. deren charakteristischer Werte:

$$E\{\underline{w}(k)\} = \underline{0}$$

$$E\{\underline{w}(k) \cdot \underline{w}(j)^T\} = Q \cdot \delta_{k,j} \tag{3-7}$$

Hiermit ist das gesamte Systemmodell definiert. Eine Zusammenfassung wird im Blockschaltbild des Systemmodells gezeigt:

Die linearisierte Beobachtungsmatrix H_{k+1} lautet:

$$\begin{aligned}
 H_{k+1} &= \frac{d}{d\hat{\underline{x}}_{k+1}} y_{k+1} \bigg|_{\hat{\underline{x}}_{k+1}^-} = \begin{bmatrix} \frac{\partial \cos(2\pi\gamma x_1(k))}{\partial x_1(k)} & \frac{\partial \cos(2\pi\gamma x_1(k))}{\partial x_2(k)} \\ \frac{\partial \sin(2\pi\gamma x_1(k))}{\partial x_1(k)} & \frac{\partial \sin(2\pi\gamma x_1(k))}{\partial x_2(k)} \end{bmatrix} = \\
 &= \begin{bmatrix} -2\pi\gamma \cdot \sin(2\pi\gamma \hat{x}_1^-(k)) & 0 \\ 2\pi\gamma \cdot \cos(2\pi\gamma \hat{x}_1^-(k)) & 0 \end{bmatrix} \tag{3-11}
 \end{aligned}$$

Die linearisierte Beobachtungsmatrix wird ebenso für die Bestimmung der Kalman-Gain-Matrix benötigt:

$$K_{k+1} = [P_{k+1}^- \cdot H_{k+1}^T] \cdot [H_{k+1} \cdot P_{k+1}^- \cdot H_{k+1}^T + R]^{-1} \tag{3-12}$$

Bei der Auswahl der Startwerte wird die direkte Berechnung der Phase verwendet:

$$x_1(0) = \arctan\left(\frac{\text{Im}\{r_T(k)\}}{\text{Re}\{r_T(k)\}}\right) \tag{3-13}$$

Der Startwert für $x_2(0)$ wird zunächst auf Null gesetzt. Die Phasenänderung wird im folgenden Verlauf durch das Driving Noise Q bestimmt. Die Startwerte für die Fehlerkovarianzmatrix $P^+(0)$ werden wie folgt festgelegt. P_{11}^+ definiert den gewünschten Fehler zwischen zwei aufeinanderfolgenden Schätzphasen; da noch keine genaue Kenntnis vorliegt, wird dieser Wert zu Beginn groß gewählt z.B. 0.5...1. Die weiteren Varianzen P_{12}^+ , P_{21}^+ und P_{22}^+ werden als klein angenommen (~ 0.1).

Ein weiteres Augenmerk gilt R und Q . R bezeichnet die Rauschmatrix, welche die Varianzen von Real- und Imaginärteilrauschen bezeichnet. Die Kreuzkovarianzen sind aufgrund der Unabhängigkeit gleich Null (s. 2.1.3).

$$R = \begin{pmatrix} R_{11} & 0 \\ 0 & R_{22} \end{pmatrix} \quad \text{mit: } R_{11} = R_{22} = N_0 f_g \text{si}(2\pi f_g \tau) \text{ (für } \tau = 0) \tag{3-14}$$

⁹ Die Rauschkovarianzen R_{11} und R_{22} werden als Rauschleistung 'Rausch' abgekürzt verwendet.

Die Driving Noise Matrix Q beschreibt die Varianzen des Systemrauschens. Die Wahl dieses Parameters läßt sich nicht direkt auf einen Wert festlegen, da zusätzlich auch Modellfehler mit diesem Parameter beschrieben werden.

$$Q = \begin{pmatrix} Q_{11} & Q_{21} \\ Q_{12} & Q_{22} \end{pmatrix} \quad (3-15)$$

$Q_{12}=Q_{21}=0$. Die Kreuzkovarianzen werden zu Null gesetzt, da keine Auswirkungen von Phase auf Frequenz und umgekehrt existieren, die nicht bereits beschrieben wurden. Im allgemeinen muß der Driving Noise Parameter deutlich kleiner (Größenordnung: 10^{-3}) als das Beobachtungsrauschen R gewählt werden, da der Kurvenverlauf der Zufallsvariablen sich deutlich langsamer verändert, als die überlagerten Rauschwerte. Andernfalls wäre eine Filterung unmöglich. Q darf jedoch auch nicht zu klein gewählt werden, da für $Q=0$ der Filter unabhängig von Meßwerten arbeitet. Die Genauigkeit würde der Filter als hochpräzise annehmen und jeden Meßwert mit beliebig kleiner Störung als zu ungenau behandeln.

Bei der Suche eines geeigneten Startwertes werden zunächst größere Werte für Q gewählt und das Einschwingverhalten des Filters beobachtet. Das Einschwingverhalten läßt sich z.B. bei den Filterkovarianzen gut beobachten. Durch Verkleinern von Q verringern sich die Einschwingzeit und die Höhe der Überschwinger, bis sie bei zu kleinem Q zu Instabilitäten führen. Das Q sollte etwas größer als dieser Grenzwert gewählt werden, um einen stabilen Filter zu erhalten.

Eine Betrachtung der Einheiten ist recht unübersichtlich, führt auch nicht zu relevanten Ergebnissen und wird damit nicht weiter betrachtet. Phase, Frequenz und später auch Amplitude und Amplitudenänderung werden alle gemeinsam in einem Zustandsvektor behandelt, so daß die entstehenden Kreuzterme entsprechende Einheiten erhalten.

3.1.1.3 Stabilität

Bei der Stabilitätsbetrachtung wird nur die stochastische Beobachtbarkeit untersucht, da keine überlagerte deterministische Größe U existiert. Zunächst wird dazu die Matrix C_a und anschließend deren Rang bestimmt.

$$C_a = \begin{bmatrix} C(k_0) \\ C(k_0) \cdot A \end{bmatrix} = \begin{bmatrix} H_{\text{Diff}} \\ H_{\text{Diff}} \cdot A \end{bmatrix} \quad (3-16)$$

$$\begin{aligned}
 \text{rang}\{C_a\} &= \text{rang} \left\{ \begin{array}{c} \left[\begin{array}{cc} -2\pi\gamma\sin(2\pi\gamma x_1) & 0 \\ 2\pi\gamma\cos(2\pi\gamma x_1) & 0 \end{array} \right] \\ \hline \left[\begin{array}{cc} -2\pi\gamma\sin(2\pi\gamma x_1) & 0 \\ 2\pi\gamma\cos(2\pi\gamma x_1) & 0 \end{array} \right] \cdot \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \end{array} \right\} = \\
 &= \text{rang} \left\{ \begin{array}{c} \left[\begin{array}{cc} -2\pi\gamma\sin(2\pi\gamma x_1) & 0 \\ 2\pi\gamma\cos(2\pi\gamma x_1) & 0 \end{array} \right] \\ \hline \left[\begin{array}{cc} -2\pi\gamma\sin(2\pi\gamma x_1) & -2\pi\gamma\sin(2\pi\gamma x_1) \\ 2\pi\gamma\cos(2\pi\gamma x_1) & 2\pi\gamma\cos(2\pi\gamma x_1) \end{array} \right] \end{array} \right\} = 2
 \end{aligned}$$

q.e.d.

Laut Abschnitt 2.6 genügt jedoch auch eine Betrachtung der Positivdefinitheit der Fisher'schen Information.

Die Auswertung des Fisher'schen Informationszuwachses zur Untersuchung der stochastischen Beobachtbarkeit ergibt:

$$\mathbf{F}_{k+1} = (\mathbf{A}^{-1}_{k+1})^T \cdot \mathbf{F}_k \cdot \mathbf{A}^{-1}_{k+1} + \mathbf{H}_{\text{Diff } k+1}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}_{\text{Diff } k+1}$$

$$\mathbf{F}_0 = \mathbf{C}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{C} = \mathbf{H}_{\text{Diff}}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}_{\text{Diff}} =$$

$$= \begin{bmatrix} -2\pi\gamma \cdot \sin(2\pi\gamma x_1(k)) & 2\pi\gamma \cdot \cos(2\pi\gamma x_1(k)) \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\text{Rausch}} & 0 \\ 0 & \frac{1}{\text{Rausch}} \end{bmatrix} \cdot \begin{bmatrix} -2\pi\gamma \cdot \sin(2\pi\gamma x_1(k)) & 0 \\ 2\pi\gamma \cdot \cos(2\pi\gamma x_1(k)) & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -2\pi\gamma \cdot \sin(2\pi\gamma x_1(k)) & 2\pi\gamma \cdot \cos(2\pi\gamma x_1(k)) \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{-2\pi\gamma}{\text{Rausch}} \cdot \sin(2\pi\gamma x_1(k)) & 0 \\ \frac{2\pi\gamma}{\text{Rausch}} \cdot \cos(2\pi\gamma x_1(k)) & 0 \end{bmatrix}$$

$$= \begin{bmatrix} +4\pi^2\gamma^2 & 0 \\ \text{Rausch} & 0 \\ 0 & 0 \end{bmatrix} \tag{3-17}$$

Nur die Phasenkomponente liefert erwartungsgemäß einen Informationsbeitrag. Um das Informationsinkrement zu bestimmen, verwenden wir die etwas allgemeinere Form der linearen Zustandsübergangsmatrix \mathbf{A} , wie sie aus dem Singer-Modell hergeleitet wird (s. 3.1.2: $\beta=0$):

$$\mathbf{A} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \quad \mathbf{A}^{-1} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} +1 & -0 \\ -T & +1 \end{bmatrix}^T = \begin{bmatrix} 1 & -T \\ 0 & 1 \end{bmatrix}$$

$$\begin{aligned}
 \mathbf{F}_1 &= \begin{bmatrix} 1 & 0 \\ -T & 1 \end{bmatrix} \cdot \begin{bmatrix} +4\pi^2\gamma^2 & \\ \text{Rausch} & 0 \\ & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & -T \\ 0 & 1 \end{bmatrix} + H_{\text{Diff}1}^T \cdot R^{-1} \cdot H_{\text{Diff}1} \\
 \Rightarrow &= \frac{+4\pi^2\gamma^2}{\text{Rausch}} \cdot \begin{bmatrix} 1 & 0 \\ -T & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & -T \\ 0 & 1 \end{bmatrix} + H_{\text{Diff}1}^T \cdot R^{-1} \cdot H_{\text{Diff}1} \\
 &= \frac{+4\pi^2\gamma^2}{\text{Rausch}} \cdot \begin{bmatrix} 1 & -T \\ -T & T^2 \end{bmatrix} + H_{\text{Diff}1}^T \cdot R^{-1} \cdot H_{\text{Diff}1}
 \end{aligned}$$

Für das erste Inkrement ergibt sich nun:

$$\mathbf{F}_1 = \frac{+4\pi^2\gamma^2}{\text{Rausch}} \cdot \begin{bmatrix} 2 & -T \\ -T & T^2 \end{bmatrix} \quad \text{und} \quad \mathbf{F}_2 = \frac{+4\pi^2\gamma^2}{\text{Rausch}} \cdot \begin{bmatrix} 3 & -3T \\ -3T & 5T^2 \end{bmatrix}$$

Die Hauptdiagonalelemente sind positivdefinit und somit kann auch von Stabilität ausgegangen werden: Das Ergebnis wurde für eine sinusförmige Modulation betrachtet. Deutlich sichtbar ist der erwartete lineare Anstieg der Information gegen unendlich:

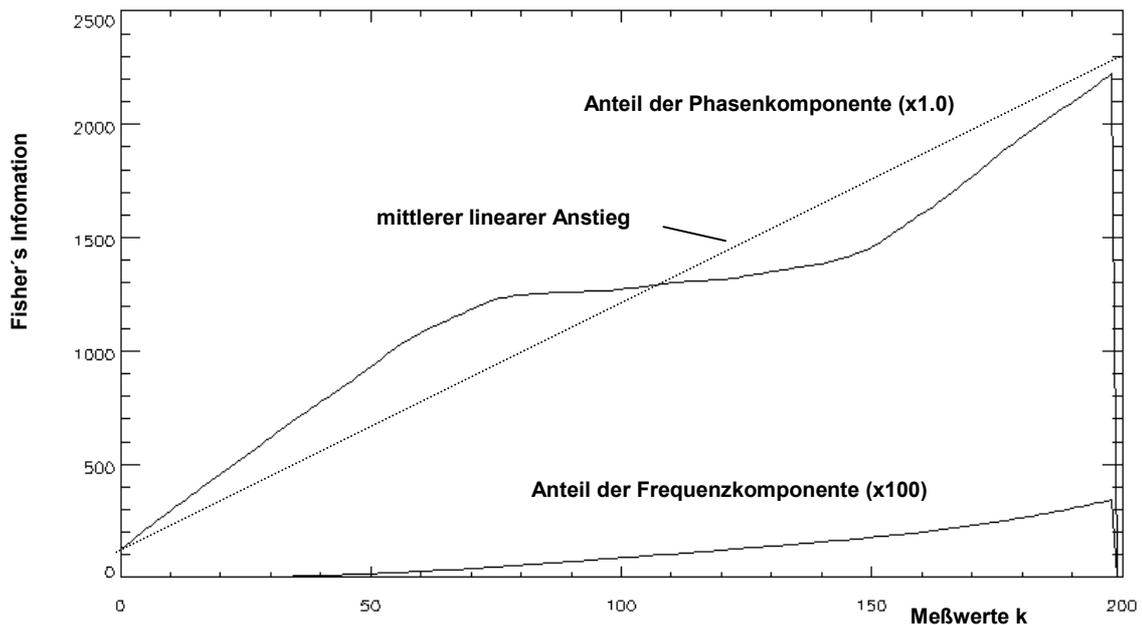


Abb. 3.7: Zeitlicher Informationsverlauf über der Zeit k des Phasen- und Frequenzanteils der Fisher'schen Informationsmatrix (sinusförmiges Quellsignal)

Der Anteil der Frequenzkomponente ist zu jedem Zeitpunkt k deutlich kleiner und beginnt auch erst ab dem 1. Inkrement an zu wachsen.

3.1.2 Zeitkontinuierliche Modellentwicklung - Singermodell

Wird die Zustandsübergangsmatrix aus dem zeitkontinuierlichen Modell entwickelt, so ergibt sich eine veränderte Matrix $A(k)$. Beim sogenannten Singermodell werden, als einfachste Umsetzung einer Differentialgleichung, die Zustandsübergänge über die Ableitung des nächsten Zustandes definiert [24, 30, 35]. In diesem Fall werden Phase und Frequenz betrachtet. (Bei zusätzlicher Betrachtung der Frequenzänderung muß diese als Zustand zusätzlich betrachtet werden. Sie dient hier nur der Transparenz der Modellbildung beim Singermodell, wurde jedoch nicht als Filter realisiert.)

$x_1(t) = \varphi(t)$	Phase
$x_2(t) = \dot{\varphi}(t) \hat{=} f(t) = \dot{x}_1(t)$	Phasenänderung = (Kreis-)Frequenz
$[x_3(t) = \ddot{\varphi}(t) = \dot{f}(t)] = \dot{x}_2(t)$	[Frequenzänderung]

Außerdem muß die Frequenzänderung mit einem Designparameter $-\beta$ ($0 < \beta < 1$) beschrieben werden, da die Frequenz nicht konstant sein soll. Das Minuszeichen gibt an, daß nur negative Werte ein sinnvolles, stabiles System ergeben.

Der Parameter $-\beta$ bestimmt zudem das Einschwingverhalten und somit die Stabilität des Filters. Man nennt einen derartigen Ansatz auch stabilen Modellansatz. Daraus ergibt sich die Zustandsgleichung eines zeitkontinuierlichen Modells für $A(k)$:

$$\dot{\underline{x}}(t) = \begin{bmatrix} 0 & 1 \\ 0 & -\beta \end{bmatrix} \cdot \underline{x}(t) + B \cdot \underline{u}(t) + G \cdot \underline{w}(t) \quad (3-18)$$

$$\left[\text{bzw.: } A(k) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -\beta \end{bmatrix} \right]$$

Bei der Betrachtung eines zeitkontinuierlichen Modells wirkt sich das zeitkontinuierliche Zustandsmodell ebenso auf die Prädiktion der Fehlerkovarianzen aus. Die Matrix P^- lautet nun:

$$P^- = A \cdot P^+ + P^+ \cdot A^T + G \cdot Q \cdot G^T \quad (3-19)$$

Die Filterfehlerkovarianz P^+ des Filterschritts bleibt dagegen identisch zum diskreten Modell, da es sich nur um eine diskrete Übergabe der Werte handelt:

$$P^+ = P^- - K \cdot H_{\text{Diff}} \cdot P^- = (I - K \cdot H_{\text{Diff}}) \cdot P^- \quad (3-20)$$

3.1.3 Übergang zur zeitdiskreten Modellbildung

Beim Übergang der zeitkontinuierlichen Beschreibung eines zeitkontinuierlichen Modells in eine zeitdiskrete Modellbeschreibung wird eine Diskretisierung zu $t=kT$ Abtastzeitpunkten durchgeführt. Dazu wird $A(k)$ über die inverse Laplacetransformierte gebildet:

$$A(k) = \phi(T) = \mathcal{L}^{-1} \left\{ (s \cdot I - F)^{-1} \right\}_{t=T} \quad (3-21)$$

$$\begin{aligned} A(k) &= \mathcal{L}^{-1} \left\{ \begin{bmatrix} s & -1 \\ 0 & s + \beta \end{bmatrix}^{-1} \right\}_{t=T} = \mathcal{L}^{-1} \left\{ \frac{1}{s \cdot (s + \beta)} \begin{bmatrix} s + \beta & 1 \\ 0 & s \end{bmatrix} \right\}_{t=T} \\ &= \mathcal{L}^{-1} \left\{ \begin{bmatrix} \frac{1}{s} & \frac{1}{s(s + \beta)} \\ 0 & \frac{1}{s + \beta} \end{bmatrix} \right\}_{t=T} = \begin{bmatrix} \sigma(t) & \frac{1}{\beta} (1 - e^{-\beta t}) \\ 0 & e^{-\beta t} \end{bmatrix}_{t=T} \end{aligned} \quad (3-22)$$

In erster Näherung der Taylorreihenentwicklung und für kausale Systeme ($t > 0$) folgt für $A(kT)$:

$$A(k) = \begin{bmatrix} 1 & \frac{1}{\beta} (1 - (1 - \beta T)) \\ 0 & 1 - \beta T \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 - \beta T \end{bmatrix} \quad (3-23)$$

Wählt man in der Zustandsübergangsmatrix $\beta = 0$, so erhält man die Zustandsbeschreibung eines Markov-Modells. T beschreibt die Abtastzeitkonstante. Je größer die Komponente gewählt wird, desto stärker wird die Frequenzkomponente in die Filterung mit einbezogen.

Bei den beschriebenen Modellen wurde $T = 1$ und $\beta = 0$ gewählt. Der im Vergleich zum zeitdiskreten EKF22 etwas veränderte Filter wird als EKF22ss (Singer-Modell) bzw. EKF22ts ($T \neq 1$) in der Auswertung beschrieben.

3.2 ERWEITERUNG DES BEOBACHTUNGSVEKTORS

Der Effekt der Datenfusion, das kreuzweise Auswerten der Messungen im Kalman-Filter, läßt sich durch zusätzliche Beobachtungen des Phasenzustands noch wirkungsvoller ausnutzen. Im Gegensatz zur Datenintegration, bei der pro Meßwert jeweils ein Schätzwert unab-

hängig von anderen, zeitgleichen Meßwerten ermittelt wird, erhält man bei der Datenfusion durch Auswertung aller vorhandenen Meßwerte pro Schätzwert eine höhere Absorption der in den Messungen enthaltenen Information.

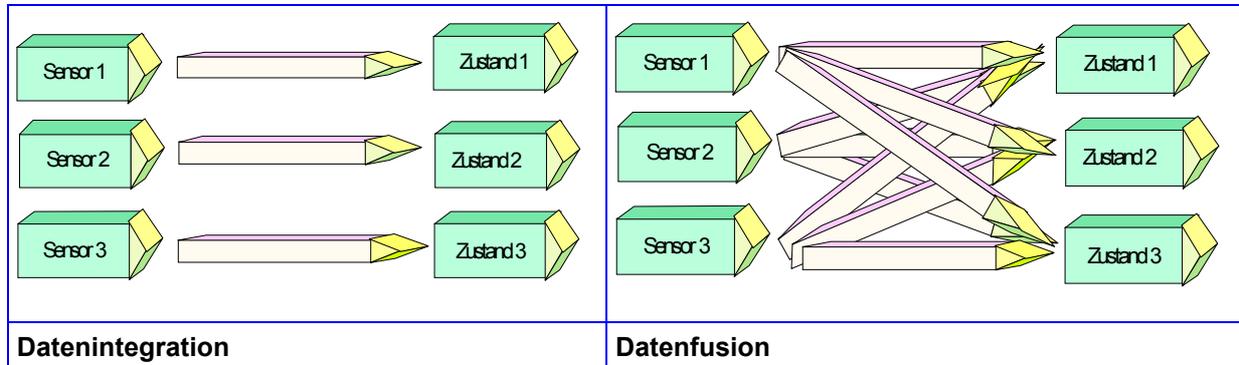


Abb. 3.8: Datenfusion und Datenintegration

Läßt sich die Anzahl der unterschiedlichen Beobachtungen für einen Zustand erhöhen, dann wird auch die verarbeitbare Information über den Zustand größer, was zu einer höheren Genauigkeit führt. Darüberhinaus wird der gesuchte Zustand unabhängiger von einem einzigen Sensor. Bei Ausfall eines Sensors wird die gesuchte Schätzgröße (hier die Phase) nicht vollständig ausfallen, sondern von den weiteren Sensoren zwar weniger akkurat, aber weiterhin geschätzt.

Durch Erweiterung des Beobachtungsvektors um weitere Sensordaten läßt sich somit eine verbesserte Filterschätzung des gesuchten Zustands erlangen. Diese Erweiterung macht insbesondere bei der Bearbeitung nichtlinearer Probleme Sinn, da hier Linearisierungen zu Filtergenauigkeiten führen. Eine zusätzliche Beobachtung des gesuchten Zustands führt so zu einer Minimierung der Linearisierungsfehler. Der Informationsfluß bei unterschiedlich vielen Beobachtungen ist in Abb. 5.9 abgebildet. Linearisierungen wirken wie ein Sieb auf die Information am Eingang: Je größer die Löcher sind, um so mehr Information kann ungehindert zum Schätzwert weiterverarbeitet werden. Dabei ist es sogar möglich, bei der Erweiterung des Beobachtungsvektors sogenannte virtuelle Meßwerte einzusetzen. Dies bedeutet, kein zusätzlicher Sensor produziert zusätzliche Meßdaten, sondern aus den bereits bestehenden Sensordaten wird eine neue Datenreihe erzeugt, die als zusätzliche Beobachtung im EKF für genauere Filterschätzungen sorgt.

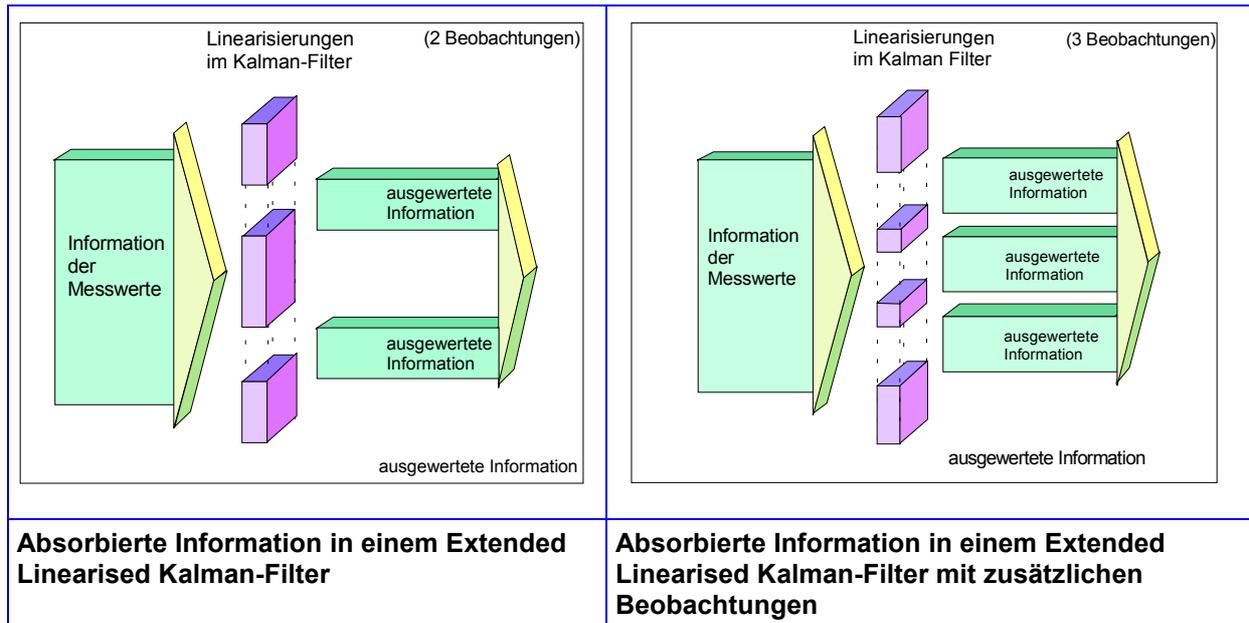


Abb. 3.9: Absorbierte Information

Die genannte Ausfallsicherheit würde sich in diesem Fall sicherlich nicht erhöhen, aber die Linearisierungsfehler des Modells lassen sich auf diese Weise verringern. Die Verwendung virtueller Meßdaten ist jedoch nur dann sinnvoll, wenn zwischen den neuen Meßdaten und den existierenden realen Daten kein linearer Zusammenhang besteht.

Im folgenden sollen zwei mögliche Erweiterungen betrachtet werden: Die Phase und die Amplitude als zusätzliche Meßgrößen. Beide lassen sich sowohl konventionell als auch virtuell aus den vorhandenen Quadraturkomponenten erzeugen.

Der Informationsgewinn unter Verwendung virtueller Beobachtungen wurde in /5, 39/ bereits ausführlich anhand von Simulationen bewiesen. Die inverse Information wächst bei negativen SNR's stark an. Bei 4 Beobachtungen ergibt sich ein deutlich geringerer Anstieg, was auf eine bessere Verarbeitung der vorhandenen Information hinweist. Insbesondere bei negativen SNR's machen sich die Informationsgewinne durch zusätzliche Beobachtungen durch virtuelle Meßwerte bemerkbar.

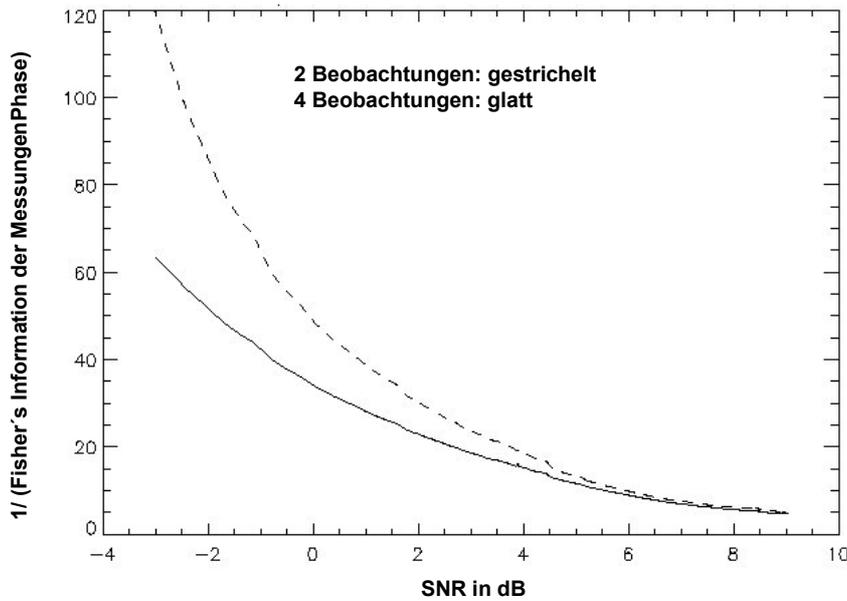


Abb. 3.10: Fisher's Information bei 2 bzw. 4 Beobachtungen

3.2.1 Erweiterung des Beobachtungsvektors auf 3 und 4 Beobachtungen

Zunächst wird der Beobachtungsvektor jeweils um einen zusätzlichen Sensor erweitert. Anschließend werden beide zusätzlichen Sensoren eingesetzt. Jede Erweiterung um zusätzliche Beobachtungen oder Zustände führt zu einem erheblich ansteigenden Rechenaufwand, der die Geschwindigkeit der Filter herabsetzt und somit auch die damit gekoppelte Abtastrate.

3.2.1.1 Zusätzliche Beobachtung: Phase (EKF32pv)

Als dritte Beobachtungsgröße soll beim EKF32pv (\Rightarrow 3 Beobachtungen und 2 Zustände) die virtuelle Meßphase $\tilde{\varphi}(k)$ eingesetzt werden. Diese kann über einen PLL oder Ähnliches direkt aus dem Empfangssignal gewonnen oder virtuell aus den beiden Quadraturkomponenten erzeugt werden. Zur Berechnung ergibt sich im Modell folgende Vorschrift für die 3. Beobachtung:

$$y_3(k) = \arctan\left(\frac{y_2(k)}{y_1(k)}\right) \cong 2\pi\gamma m_1 kT + \varphi_0 + v_3(k) \quad (3-24)$$

Im Definitionsbereich der Arcustangens-Fkt. ergibt sich (besonders genau für kleine Argumente) die Näherung nach Taylorreihenentwicklung erster Ordnung. Durch diese nichtlineare Verknüpfung ergibt sich in der Näherung für den Rauschprozeß $v_3(k)$ ein Prozeß, der nicht

mehr statistisch unabhängig von den Prozessen $v_1(k)$ und $v_2(k)$ ist. Er stellt vielmehr eine komplizierte mathematische Verknüpfung dar. Die Gleichung (3-24) wäre aus rein skalarer Sicht nur erfüllt, wenn $v_1(k)$ und $v_2(k)$ identisch wären, hier gilt jedoch die stochastische Betrachtungsweise. Der Beobachtungsvektor lautet dann:

$$\underline{y}(k) = \underline{h}(\underline{x}(k)) + \underline{v}(k)$$

$$\Rightarrow \underline{y}(k) = \begin{bmatrix} y_1(k) \\ y_2(k) \\ y_3(k) \end{bmatrix} = \begin{bmatrix} \operatorname{Re}\{r_T(k)\} \\ \operatorname{Im}\{r_T(k)\} \\ \tilde{\varphi}(k) = \arctan \frac{\operatorname{Im}\{r_T(k)\}}{\operatorname{Re}\{r_T(k)\}} \end{bmatrix} = \begin{bmatrix} \cos[2\pi\gamma \cdot x_1(k)] \\ \sin[2\pi\gamma \cdot x_1(k)] \\ 2\pi\gamma \cdot x_1(k) \end{bmatrix} + \begin{bmatrix} v_1(k) \\ v_2(k) \\ v_3(k) \end{bmatrix} \quad |^{10} \quad (3-25)$$

$$\text{mit } E\{\underline{v}(k)\} = \underline{0}$$

Für das H_{k+1} ergibt sich nun:

$$H_{k+1} = \frac{d}{d\underline{x}_{k+1}} \underline{y}_{k+1} \Big|_{\hat{\underline{x}}_{k+1}^-} = \begin{bmatrix} \frac{\partial \cos(2\pi\gamma x_1(k))}{\partial x_1(k)} & \frac{\partial \cos(2\pi\gamma x_1(k))}{\partial x_2(k)} \\ \frac{\partial \sin(2\pi\gamma x_1(k))}{\partial x_1(k)} & \frac{\partial \sin(2\pi\gamma x_1(k))}{\partial x_2(k)} \\ \frac{\partial 2\pi\gamma x_1(k)}{\partial x_1(k)} & \frac{\partial 2\pi\gamma x_1(k)}{\partial x_2(k)} \end{bmatrix}$$

$$H_{k+1} = \begin{bmatrix} -2\pi\gamma \cdot \sin(2\pi\gamma x_1(k)) & 0 \\ +2\pi\gamma \cdot \cos(2\pi\gamma x_1(k)) & 0 \\ 2\pi\gamma & 0 \end{bmatrix} \quad (3-26)$$

Filterergebnisse mit realen und virtuellen Daten der zusätzlichen Phase werden im Kapitel 4 diskutiert.

Die Rauschmatrix R muß ebenfalls an das neue Beobachtungsmodell angepaßt werden. Da der Rauschprozeß v_3 nicht unabhängig von v_1, v_2 ist, ergeben sich für die Kreuzkovarianzen σ_{cr} der Rauschmatrix R folgende Terme:

¹⁰ Der Definitionsbereich der virtuellen Phase $\tilde{\varphi}$ ist auf einen Bereich von 2π beschränkt (Modulo 2π)

$$\mathbf{R} = \begin{bmatrix} \sigma_R^2 & 0 & -\sin x_1(k) \cdot \sigma_{\text{cross}}^2 \\ 0 & \sigma_R^2 & \cos x_1(k) \cdot \sigma_{\text{cross}}^2 \\ -\sin x_1(k) \cdot \sigma_{\text{cross}}^2 & \cos x_1(k) \cdot \sigma_{\text{cross}}^2 & \sigma_\varphi^2 \end{bmatrix} \quad (3-27)$$

mit:

$$\sigma_\varphi^2 = \frac{1}{2} \cdot \exp\left\{-\frac{a_0^2}{2 \cdot \sigma_R^2}\right\} \cdot \sum_{n=0}^{\infty} \frac{1}{n!} \cdot \left(\frac{a_0}{\sqrt{2 \cdot \sigma_R^2}}\right)^n \cdot \Gamma\left(\frac{n}{2} + 1\right) \cdot \sum_{k=0}^n \binom{n}{k} \cdot \left\{ 2 \cdot \pi^2 \cdot \frac{\sin[\pi \cdot (2k - n)]}{\pi \cdot (2k - n)} - 4 \cdot \frac{\cos[\pi \cdot (2k - n)]}{(2k - n)^2} + \frac{4}{\pi} \cdot \frac{\sin[\pi \cdot (2k - n)]}{(2k - n)^3} \right\}$$

$$\text{und: } \sigma_{\text{cross}}^2 = \left[\sqrt{2\pi \cdot \sigma_R^2} \cdot \Phi\left(-\frac{a_0}{\sigma_R}\right) - \frac{\sigma_R^2}{a_0} \cdot \left(\exp\left\{-\frac{a_0^2}{2 \cdot \sigma_R^2}\right\} - 1 \right) \right]$$

Diese Terme werden in /5/ ausführlich hergeleitet. Zur Berechnung genügen jedoch auch die folgenden Näherungen, so daß sich dann für die Kreuzkovarianzen und Phasenfehlerkovarianz ergeben:

$$\sigma_{\text{cross}}^2 = a_0 \frac{\sigma_R^2}{a_0^2 + \sigma_R^2} \quad \text{und} \quad \sigma_\varphi^2 = \frac{a_0^2}{2 \cdot \text{SNR}} \cdot 1.25 + 0.013 \quad (3-28)$$

Letzterer Term läßt sich durch einfaches Kurvenfitting der exakten Phasenfehlerkovarianz σ_φ^2 mit der reziproken Fisher's Information der Quadraturkomponenten bestimmen. Mit einem Faktor von 1.25 und einem kleinen Offset erhält man einen recht genauen Wert (<4% rel. Fehler) σ_φ^2 .

Zur Näherung von Gleichung (3-28):

Der reziproke Wert der Fisher'schen Information der Quadraturkomponenten, die Summe der Varianzen der beiden überlagerten, gaußverteilten Rauschprozesse der Quadraturkomponenten im Verhältnis zur Signalleistung läßt sich auch mit Hilfe des SNR's beschreiben:

$$\text{SNR} = \frac{a_0^2}{\sigma_{\text{vr}}^2 + \sigma_{\text{vl}}^2} = \frac{a_0^2}{2 \cdot \sigma_{\text{N}}^2} \quad (3-29)$$

Nach der Varianz umgestellt ergibt sich:

$$\sigma_N^2 = \frac{a_0^2}{2 \cdot \text{SNR}} \quad (3-30)$$

Im folgenden wird von der hier verwendeten normierten Amplitude $a_0=1$ ausgegangen.

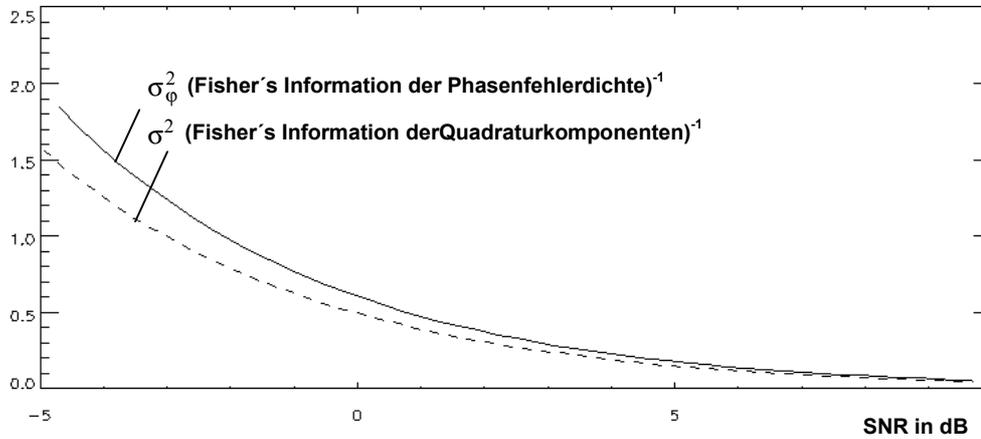


Abb. 3.11: Inverse Fisher'sche Information

Die Phasenfehlerkovarianz σ_ϕ^2 der Quadraturkomponenten wird bestimmt durch das Signal-Rauschleistungsverhältnis und kann im Falle großer SNR's und normierter Amplitude genähert werden /58/:

$$\sigma_N^2 = \frac{1}{2 \cdot \text{SNR}} = \tilde{\sigma}_\phi^2 \cong \sigma_\phi^2 \quad (3-31)$$

Die hier gefundene Näherung gilt jedoch insbesondere für kleine SNR's und weist nur einen relativen Fehler von $\tilde{\sigma}_\phi^2$ zu σ_ϕ^2 von max. 4% auf (-10..+10 dB).

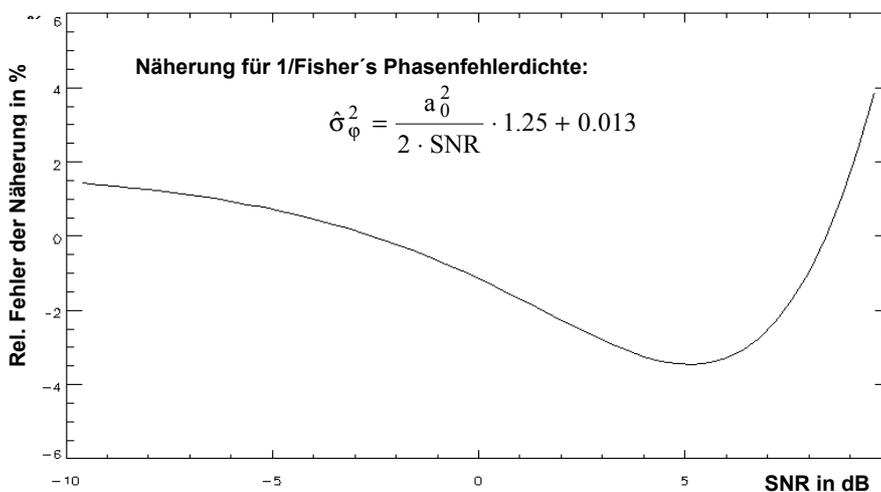


Abb. 3.12: Näherung der Phasenfehlerkovarianz

Der Bezug von σ_ϕ^2 zum SNR läßt sich an einer Zeigerbetrachtung verdeutlichen:

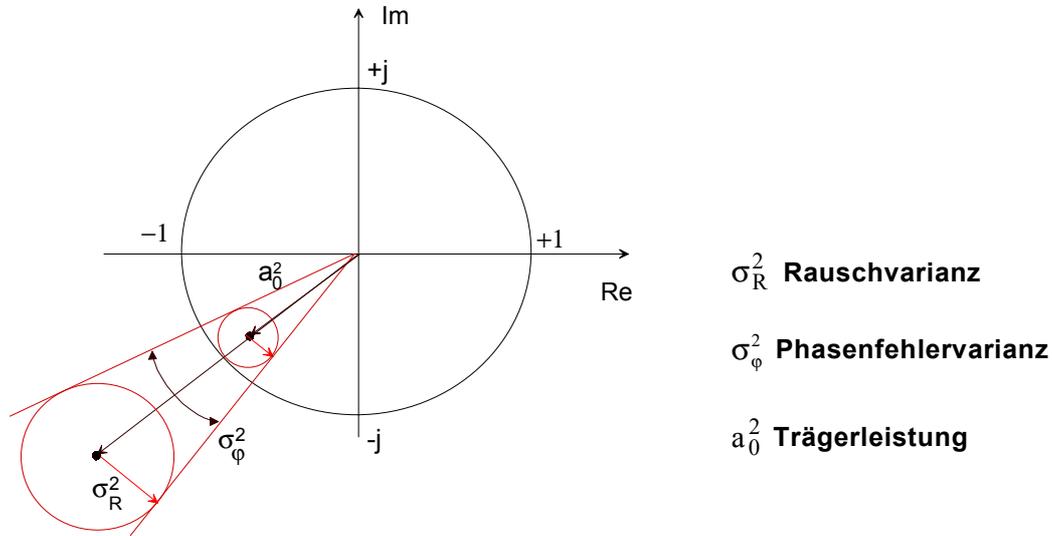


Abb. 3.13: Abhängigkeit von Phasenfehlervarianz zum SNR-Verhältnis

Das Zeigerdiagramm zeigt die Verarbeitung von Quadraturkomponenten im 3. Quadranten. Bei konstantem SNR muß die Rauschleistung auch bei steigender Signalleistung vergrößert werden. Die Verarbeitung muß jedoch im Bereich von $\pi/2$ bleiben, damit eine fehlerfreie Phasenrekonstruktion möglich ist.

Stabilitätsbetrachtung mit C_a (s. 2.6.1):

$$C_a = \begin{bmatrix} C(k_0) \\ C(k_0) \cdot A \\ C(k_0) \cdot A^2 \end{bmatrix} = \begin{bmatrix} H_{\text{Diff}} \\ H_{\text{Diff}} \cdot A \\ H_{\text{Diff}} \cdot A^2 \end{bmatrix} \quad (3-32)$$

$$\text{rang}\{C_a\} = \text{rang} \left\{ \begin{array}{cc} -2\pi\gamma \sin(2\pi\gamma x_1) & 0 \\ 2\pi\gamma \cos(2\pi\gamma x_1) & 0 \\ 2\pi\gamma & 0 \\ \hline -2\pi\gamma \sin(2\pi\gamma x_1) & -2\pi\gamma \sin(2\pi\gamma x_1) \\ 2\pi\gamma \cos(2\pi\gamma x_1) & 2\pi\gamma \cos(2\pi\gamma x_1) \\ 2\pi\gamma & 2\pi\gamma \\ \hline -2\pi\gamma \sin(2\pi\gamma x_1) & 4\pi\gamma \sin(2\pi\gamma x_1) \\ 2\pi\gamma \cos(2\pi\gamma x_1) & 4\pi\gamma \cos(2\pi\gamma x_1) \\ 2\pi\gamma & 4\pi\gamma \end{array} \right\} = 2$$

Die Untersuchung der Fisher'schen Information oder auch die Betrachtung der nichtdeterministischen Beobachtbarkeit $R \neq I$ (Einheitsmatrix) ergibt:

$$\mathbf{F} = H_{\text{Diff}}^T \cdot R^{-1} \cdot H_{\text{Diff}}$$

$$= \begin{bmatrix} -2\pi\gamma \sin(2\pi\gamma x_1) & 2\pi\gamma \cos(2\pi\gamma x_1) & 2\pi\gamma \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \sigma_R^2 & 0 & -\sin(2\pi\gamma x_1)\sigma_{Cr}^2 \\ 0 & \sigma_R^2 & +\cos(2\pi\gamma x_1)\sigma_{Cr}^2 \\ -\sin(2\pi\gamma x_1)\sigma_{Cr}^2 & +\cos(2\pi\gamma x_1)\sigma_{Cr}^2 & \sigma_\varphi^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} -2\pi\gamma \sin(2\pi\gamma x_1) & 0 \\ 2\pi\gamma \cos(2\pi\gamma x_1) & 0 \\ 2\pi\gamma & 0 \end{bmatrix}$$

Für R^{-1} ergibt sich folgender Term:

$$R^{-1} = \frac{1}{\det R} \cdot \begin{bmatrix} \sigma_R^2 \sigma_\varphi^2 - \cos^2(2\pi\gamma x_1)\sigma_{Cr}^4 & -\cos(2\pi\gamma x_1)\sin(2\pi\gamma x_1)\sigma_{Cr}^4 & \sin(2\pi\gamma x_1)\sigma_{Cr}^2 \sigma_R^2 \\ -\cos(2\pi\gamma x_1)\sin(2\pi\gamma x_1)\sigma_{Cr}^4 & \sigma_R^2 \sigma_\varphi^2 - \sin^2(2\pi\gamma x_1)\sigma_{Cr}^4 & -\cos(2\pi\gamma x_1)\sigma_{Cr}^2 \sigma_R^2 \\ \sin(2\pi\gamma x_1)\sigma_{Cr}^2 \sigma_R^2 & -\cos(2\pi\gamma x_1)\sigma_{Cr}^2 \sigma_R^2 & \sigma_R^4 \end{bmatrix}$$

$$\text{mit: } \det R = \sigma_R^2 (\sigma_R^2 \sigma_\varphi^2 - \sin^2(2\pi\gamma x_1)\sigma_{Cr}^4) \quad (3-33)$$

$$\mathbf{F}_{k+1} = (A^{-1}_{k+1})^T \cdot \mathbf{F}_k \cdot A^{-1}_{k+1} + H_{\text{Diff}_{k+1}}^T \cdot R^{-1} \cdot H_{\text{Diff}_{k+1}}$$

Wird anschließend die Fisher'sche Informationsmatrix gebildet, so erhält man erneut einen Term im 1. Element der Matrix, alle weiteren sind Null. Die Berechnung des Fisher'schen Inkrements führt also zum gleichen Resultat, die Berechnung ist hier jedoch erheblich aufwendiger.

3.2.1.2 Zusätzliche Beobachtung: Amplitude

Als dritte Beobachtungsgröße soll hier die Meßamplitude $\tilde{a}(k)$ verwendet werden (\Rightarrow EKF32av, EKF34av). Die Amplitude kann über die Hüllkurvendemodulation direkt aus dem Empfangssignal gewonnen werden (s. 2.1.2) oder erneut virtuell aus den beiden Quadraturkomponenten erzeugt werden. Um bei diesem Filtermodell einen Informationsgewinn zu bekommen, muß eine Zustandsschätzung für die Amplitude durchgeführt werden. Zur Berechnung ergibt sich im Beobachtungsmodell zunächst folgende Vorschrift für die 3. Beobachtung:

$$y_3(k) = \tilde{a}(k) = \sqrt{\text{Im}\{r_T(k)\}^2 + \text{Re}\{r_T(k)\}^2} \quad (3-34)$$

Die Beobachtungsgleichung lautet:

$$\begin{aligned} \underline{y}(k) &= \underline{h}(\underline{x}(k)) + \underline{v}(k) \\ \Rightarrow \underline{y}(k) &= \begin{bmatrix} y_1(k) \\ y_2(k) \\ y_3(k) \end{bmatrix} = \begin{bmatrix} \text{Re}\{r_T(k)\} \\ \text{Im}\{r_T(k)\} \\ \tilde{a}(k) = \arctan \frac{\text{Im}\{r_T(k)\}}{\text{Re}\{r_T(k)\}} \end{bmatrix} \\ &= \begin{bmatrix} a(k) \cdot \cos[2\pi\gamma \cdot x_1(k)] \\ a(k) \cdot \sin[2\pi\gamma \cdot x_1(k)] \\ a(k) \end{bmatrix} + \begin{bmatrix} v_1(k) \\ v_2(k) \\ v_3(k) \end{bmatrix} \end{aligned} \quad (3-35)$$

mit: $E\{\underline{v}(k)\} = \underline{0}$

Für das H_{k+1} bei unverändertem Zustandsmodell ergibt sich nun:

$$H_{k+1} = \frac{d}{d\underline{x}_{k+1}} \underline{y}_{k+1} \Big|_{\hat{\underline{x}}_{k+1}} = \begin{bmatrix} -2\pi\gamma \cdot a(k) \cdot \sin(2\pi\gamma \cdot x_1(k)) & 0 \\ +2\pi\gamma \cdot a(k) \cdot \cos(2\pi\gamma \cdot x_1(k)) & 0 \\ 0 & 0 \end{bmatrix} \quad (3-36)$$

Ohne eine Schätzung der Amplitude kann sich kein Informationsgewinn ergeben. Das Zustandsmodell wird also erweitert. Der zu schätzende Zustandsvektor $\underline{x}(k)$ lautet (s. auch /5/):

$$\underline{x}(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \\ x_4(k) \end{bmatrix} = \begin{bmatrix} m(k) \\ \dot{m}(k) \\ a(k) \\ \dot{a}(k) \end{bmatrix}$$

- Phase**
- Frequenz (Phasenänderung)**
- Amplitude des Empfangssignals**
- Amplitudenänderung des Empfangssignals**

Das Kalman-Filter läßt sich neben der Phasenschätzung also auch zur Schätzung der Amplitude einsetzen. Auch ohne erweiterten Beobachtungsvektor, ist eine Amplitudenschätzung möglich (s. EKF24). Zur Zustandsgenerierung der Amplitude wird ebenfalls von einem beliebig sich verändernden Signal ausgegangen, das mit einem vereinfachten Markov-Prozeß realisiert wird. Die Zustandsübergangsmatrix A lautet dann:

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Die linearisierte Beobachtungsmatrix $H_{\text{Diff}}(k)$ lautet (EKF34pa):

$$H_{\text{Diff}}(k) = \begin{bmatrix} -x_3(k) \cdot 2\pi\gamma \sin(2\pi\gamma x_1(k)) & 0 & \cos(2\pi\gamma x_1(k)) & 0 \\ x_3(k) \cdot 2\pi\gamma \cos(2\pi\gamma x_1(k)) & 0 & \sin(2\pi\gamma x_1(k)) & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (3-37)$$

Grundsätzlich sind Phase (primär gesuchter Zustand) und Amplitude zwar unabhängig (Zylinderkoordinaten), jedoch wird die Amplitude nichtlinear mit den beiden Beobachtungen beschrieben, so daß geringe Schwankungen einer an sich konstant gewählten Amplitude auch Information über die gesuchte Phase enthalten. Filterergebnisse mit realen und virtuellen Daten der zusätzlichen Amplitude werden im Kapitel 4 diskutiert.

Für die angepaßte Rauschmatrix R und die Systemrauschmatrix Q erhält man:

$$R = \begin{bmatrix} \sigma_R^2 & 0 & c_{\text{ar}} \\ 0 & \sigma_R^2 & c_{\text{ai}} \\ c_{\text{ar}} & c_{\text{ai}} & 1 \end{bmatrix} \quad (3-38)$$

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma_{\text{sysnoise}}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.001 \cdot \sigma_{\text{sysnoise}}^2 \end{bmatrix} \quad (3-39)$$

Die Kreuzkovarianzen c_{ar} und c_{ai} von Amplitudenrauschen und Störung der Quadraturkomponenten sind ebenfalls recht aufwendige Terme, die in /5/ berechnet wurden:

$$c_{\text{ar}} = \sqrt{2} \cdot \sigma \cdot a_0 \cdot \cos(2\pi\gamma x_1) \cdot \exp\left\{-\frac{a_0^2}{2 \cdot \sigma^2}\right\} \cdot \left[\sum_{n=0}^{\infty} \frac{1}{(n!)^2} \cdot \Gamma\left(n+1+\frac{1}{2}\right) \cdot \left(n \cdot \left(\frac{a_0^2}{2 \cdot \sigma^2}\right)^{n-1} - \left(\frac{a_0^2}{2 \cdot \sigma^2}\right)^n \right) \right]$$

$$c_{\text{ai}} = \sqrt{2} \cdot \sigma \cdot a_0 \cdot \sin(2\pi\gamma x_1) \cdot \exp\left\{-\frac{a_0^2}{2 \cdot \sigma^2}\right\} \cdot \left[\sum_{n=0}^{\infty} \frac{1}{(n!)^2} \cdot \Gamma\left(n+1+\frac{1}{2}\right) \cdot \left(n \cdot \left(\frac{a_0^2}{2 \cdot \sigma^2}\right)^{n-1} - \left(\frac{a_0^2}{2 \cdot \sigma^2}\right)^n \right) \right]$$

3.2.2 Erweiterung des Beobachtungsvektors auf 4 Beobachtungen (EKF44pav)

Es ist ebenfalls möglich, die beiden zusätzlichen Komponenten Meßphase $\tilde{\varphi}(k)$ und Meßamplitude $\tilde{a}(k)$ im Kalman-Filter mit auszuwerten (\Rightarrow EKF44pav). Der Beobachtungsvektor wird dann auf 4 Komponenten erweitert, was zu erheblich höherem Rechenaufwand führt (s. 4.5.4.1).

Die Beobachtungsgleichung lautet:

$$\underline{y}(k) = \underline{h}(\underline{x}(k)) + \underline{v}(k)$$

$$\Rightarrow \underline{y}(k) = \begin{bmatrix} y_1(k) \\ y_2(k) \\ y_3(k) \\ y_4(k) \end{bmatrix} = \begin{bmatrix} \operatorname{Re}\{r_T(k)\} \\ \operatorname{Im}\{r_T(k)\} \\ \tilde{\varphi}(k) = \arctan \frac{\operatorname{Im}\{r_T(k)\}}{\operatorname{Re}\{r_T(k)\}} \\ \tilde{a}(k) = \sqrt{\operatorname{Re}^2\{r_T(k)\} + \operatorname{Im}^2\{r_T(k)\}} \end{bmatrix} = \begin{bmatrix} x_3(k) \cos[2\pi\gamma \cdot x_1(k)] \\ x_3(k) \sin[2\pi\gamma \cdot x_1(k)] \\ 2\pi\gamma \cdot x_1(k) \\ x_3(k) \end{bmatrix} + \begin{bmatrix} v_1(k) \\ v_2(k) \\ v_3(k) \\ v_4(k) \end{bmatrix}$$

$$\text{mit } E\{\underline{v}(k)\} = \underline{0}$$

(3-40)

Für das $H_{\text{Diff}}(k+1)$ ergibt sich:

$$H_{\text{Diff}}(k+1) = \frac{d}{d\underline{x}_{k+1}} \underline{h}_{k+1} \Big|_{\hat{\underline{x}}_{k+1}^-} = \begin{bmatrix} -2\pi\gamma \cdot \sin(2\pi\gamma x_1(k)) & 0 \cos(2\pi\gamma x_1(k)) & 0 \\ +2\pi\gamma \cdot \cos(2\pi\gamma x_1(k)) & 0 \sin(2\pi\gamma x_1(k)) & 0 \\ 2\pi\gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (3-41)$$

Die Zustandsgleichung bleibt weiterhin unverändert. Filterergebnisse mit realen und virtuellen Daten der zusätzlichen Phase und Amplitude werden im Kapitel 5 diskutiert. (Stabilisierend wirkt sich eine schwache Amplitudenmodulation ($\mu_{\text{AM}} < 0.1$) anstelle einer konstanten Amplitude auf das gesamte Filtermodell aus, wie Simulationen gezeigt haben /5/. Die Amplitudenänderung ist in diesem Fall verschieden von Null, wie auch durch den Markov-Prozeß bei der Zustandsgenerierung angenommen; das reale System wurde so an das entwickelte Modell angepaßt.)

Mit $u(k)=0$ und $G(k)=I$) ergibt für die Systemgleichungen:

$$\underline{x}(k+1) = A(k) \cdot \underline{x}(k) + \underline{u}(k) + G(k) \cdot \underline{w}(k) = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \\ x_4(k) \end{bmatrix} + \begin{bmatrix} 0 \\ w_2(k) \\ 0 \\ w_4(k) \end{bmatrix} \quad \text{Zustands-} \\ \text{gleichung}$$

$$\underline{y}(k) = \underline{h}(\underline{x}(k)) + \underline{v}(k) = \begin{bmatrix} \operatorname{Re}\{r_T(k)\} \\ \operatorname{Im}\{r_T(k)\} \\ \tilde{\varphi}(k) \\ \tilde{a}(k) \end{bmatrix} = \begin{bmatrix} x_3(k) \cos[2\pi\gamma \cdot x_1(k)] \\ x_3(k) \sin[2\pi\gamma \cdot x_1(k)] \\ 2\pi\gamma \cdot x_1(k) \\ x_3(k) \end{bmatrix} + \begin{bmatrix} v_1(k) \\ v_2(k) \\ v_3(k) \\ v_4(k) \end{bmatrix} \quad \text{Beobachtungs-} \\ \text{gleichung}$$

Die Rauschmatrix R und die Systemrauschmatrix Q lauten:

$$R = \begin{bmatrix} \sigma_R^2 & 0 & -\sin x_1(k)\sigma_{\text{cross}}^2 & c_{\text{ar}} \\ 0 & \sigma_R^2 & \cos x_1(k)\sigma_{\text{cross}}^2 & c_{\text{ai}} \\ -\sin x_1(k)\sigma_{\text{cross}}^2 & \cos x_1(k)\sigma_{\text{cross}}^2 & \sigma_\phi^2 & 0 \\ c_{\text{ar}} & c_{\text{ai}} & 0 & 1 \end{bmatrix} \quad (3-42)$$

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma_{\text{sysnoise}}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.001 \cdot \sigma_{\text{sysnoise}}^2 \end{bmatrix} \quad (3-43)$$

Das Amplitudenrauschen σ_a^2 wird zu eins zusammengefaßt. Es existieren keine Kreuzkovarianzen von der Amplitude zur Phase, da Amplitude und Phase in der Gaußnäherung orthogonal sind. Dies ist jedoch nicht ganz korrekt, da über das Bildungsgesetz der virtuellen Amplitude ein direkter Zusammenhang zu den Quadraturkomponenten existiert - eine Berechnung dieser Zusammenhänge ist jedoch noch nicht erfolgt und stellt sich als mathematisch recht kompliziert heraus.

Die Kreuzkovarianzen von Amplitudenrauschen und Störung der Quadraturkomponenten sind bereits im letzten Abschnitt diskutiert worden.

Die Systemrauschvarianz zur Erzeugung der Phasenschätzwerte wird mit $\sigma_{\text{sysnoise}}^2$ bezeichnet. Da die Änderung der Amplitude als gering angenommen wird, wird das Driving Noise Q_{44} erheblich kleiner als Q_{22} gewählt.

3.3 ADAPTIVE KALMAN-FILTER

Eine weitere Möglichkeit, das Standardfilter EKF22 zu verbessern, ist die adaptive Nachregelung der ansonsten konstanten Parameter R oder/und Q. Beide Parameter anzupassen, würde einen enormen Zusatzaufwand bedeuten und ist schon deshalb wenig sinnvoll. Darüberhinaus stehen die beiden Größen in engem Zusammenhang zueinander. In dieser Arbeit wird die Anpassung des wichtigeren Parameters Q behandelt, da unter Q neben der Abhängigkeit zur Leistung des Beobachtungsruschens auch die Modellungenauigkeiten vereint sind. Die adaptive Schätzung von Q, der Varianz des Driving Noise, kann durch eine

Schätzung im Kalman-Filtermodell, aber auch durch eine externe Schätzung nach jedem Kalman-Filterschätzwert geschehen /19, 63, 69, 74/. Grundsätzlich lassen sich adaptive auch mit den erweiterten Kalman-Filtermodellen problemlos kombinieren.

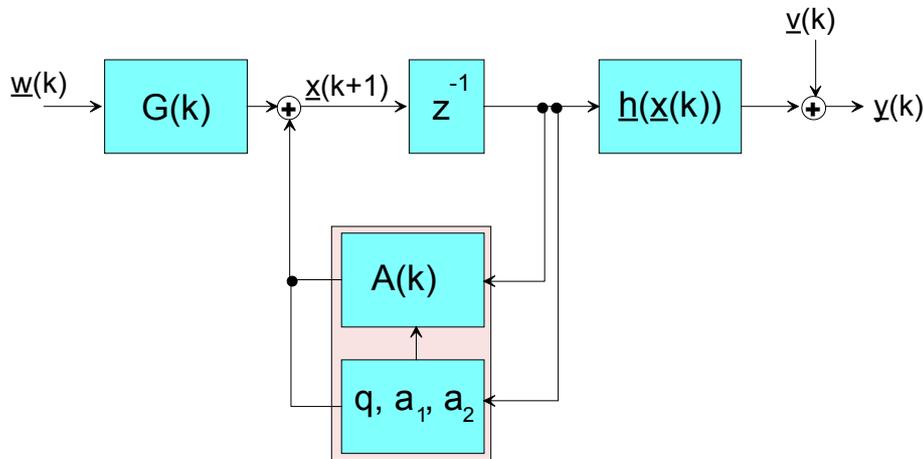


Abb. 3.14: Allgemeines Systemmodell zur adaptiven Schätzung des Parameters Q

3.3.1 Adaptive Anpassung von Parametern im Kalman-Filtermodell

Durch die Erweiterung des Zustandsvektors $\underline{x}(k)$ wird die Generierung eines Parameterverlaufs für Q ermöglicht. Im Modell ließe sich diese Parameterkontrolle als stochastischer Check bezeichnen. Durch Erzeugung eines Markov II - Prozesses mit variablen Parametern - als Zustände realisiert - wird ein dynamischer Parameter Q realisiert. Auf diese Weise wird ein sich langsam verändernder Parameter Q generiert, der abhängig von der Diffusionskonstanten q und den beiden Parametern a_1 und a_2 ist.

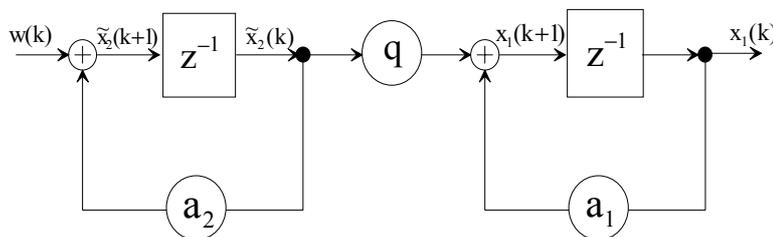


Abb. 3.15: Driving Noise Generierung im Zustandsraum

Mit dem abgebildeten Modell ergeben sich die folgenden Zustandsgleichungen:

$$\begin{aligned} x_1(k+1) &= a_1 \cdot x_1(k) + q \cdot \tilde{x}_2(k) \\ \tilde{x}_2(k+1) &= a_2 \cdot \tilde{x}_2(k) + w(k) \end{aligned} \quad (3-44)$$

Dazu müssen drei zusätzliche Zustände im Zustandsvektor eingefügt werden. Die gesamte vektoriell zusammengefaßte Zustandsgleichung der adaptiven Filterrealisierung lautet nun:

$$\underline{x}(k+1) = A(k) \cdot \underline{x}(k) + \underline{u}(k) + G(k) \cdot \underline{w}(k)$$

$$= \begin{bmatrix} x_3 & x_5 & 0 & 0 & 0 \\ 0 & x_4 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \underline{x}(k) + \begin{bmatrix} 0 \\ w(k) \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{mit: } \underline{x}(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \\ x_4(k) \\ x_5(k) \end{bmatrix} = \begin{bmatrix} m(k) \\ \dot{m}(k) \\ a_1 \\ a_2 \\ q \end{bmatrix}$$

Eine zusammenfassende Darstellung gibt die Abbildung des Filtermodells wieder:

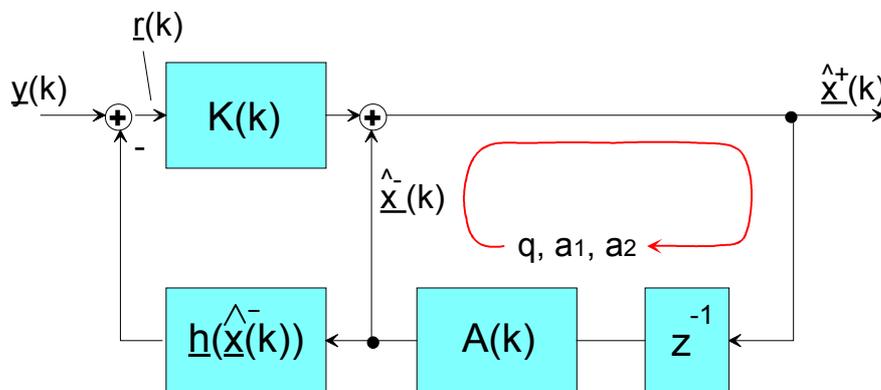


Abb. 3.16: Adaptives Kalman-Filtermodell: Schätzung von q , a_1 , a_2 als additive Zustandsvariablen

Bei geeigneter Wahl des Parameters a_1 kann auf den zweiten Parameter verzichtet werden, und es wird nur ein Markov-Prozeß 2. Ordnung vorgetäuscht. Der Zustandsvektor wurde im Hinblick auf die Online-Realisierbarkeit auf 4 Zustände begrenzt.

3.3.2 Adaptive Anpassung von Parametern durch Maximum Likelihood Schätzung

Bei einem Standardmodell des EKF22s wird die Rauschleistungsdichte des Systemrauschens adaptiv mit Hilfe einer "externen" Maximum Likelihood-Schätzung nachgeregelt. Somit wird der entstehende Filteraufwand, der zuvor durch die zusätzlichen (2-3) Zustände erheblich anstieg, vermieden, ohne dabei jedoch auf die adaptiven Eigenschaften zu verzichten. Für das Kalman-Filter wird Q als konstanter Parameter weiterverarbeitet, jedoch wird er anhand der externen Maximum Likelihood-Estimation an die stochastischen Eigenschaften der Messungen angepaßt. Die Kovarianz des Systemrauschens definiert sich nicht mehr allein über

das konstante, gaußverteilte und weiße Markov-Modell, sondern es wird über die Residuen maßgeblich mitbestimmt.

$$\begin{aligned} f_{\underline{x}(k), \underline{Y}(k) | \underline{a}}(\underline{\xi}_k, \underline{\rho}_k | \underline{\alpha}) &= f_{\underline{x}(k) | \underline{Y}(k), \underline{a}} \cdot \frac{f_{\underline{Y}(k), \underline{a}}}{f_{\underline{a}}} = f_{\underline{x}(k) | \underline{Y}(k), \underline{a}} \cdot f_{\underline{Y}(k) | \underline{a}} \\ &= f_{\underline{x}(k) | \underline{Y}(k), \underline{a}} \cdot f_{\underline{y}(k) | \underline{Y}(k-1), \underline{a}} \cdot f_{\underline{Y}(k-1) | \underline{a}} \end{aligned} \quad (3-45)$$

Die beiden Verteilungsdichten sind gaußverteilt:

$$\begin{aligned} f_{\underline{x}(k) | \underline{Y}(k), \underline{a}} &= \frac{1}{\sqrt{2\pi^n \cdot |\mathbf{P}_k^+|}} e^{(\underline{\xi} - \hat{\underline{x}}_k^+)^T \cdot \mathbf{P}_k^{+^{-1}} \cdot (\underline{\xi} - \hat{\underline{x}}_k^+)} \\ f_{\underline{y}(j) | \underline{Y}(j-1), \underline{a}} &= \frac{1}{\sqrt{2\pi^m \cdot |\mathbf{P}_k^-|}} e^{(\underline{\rho} - \underline{h}(\hat{\underline{x}}_k^-))^T \cdot \mathbf{A}_k^{-1} \cdot (\underline{\rho} - \underline{h}(\hat{\underline{x}}_k^-))} \end{aligned} \quad (3-46)$$

Mit: $\mathbf{A}_k = \mathbf{H}_{\text{Diff}} \mathbf{P}_k^- \mathbf{H}_{\text{Diff}}^T + \mathbf{R}_k$

Verwendet man die allgemeine Bedingung des Likelihood-Funktional als Ansatz einer mathematischen Herleitung, so ergibt sich daraus nachfolgende Kausalreihe. Die allgemeine bedingte Verteilungsfunktion (der Filterschätzwerte, bedingt auf alle verfügbaren Beobachtungen) lautet:

$$f_{\underline{x}(k), \underline{Y}(k) | \underline{a}}(\underline{\xi}_k, \underline{\rho}_k | \underline{\alpha}) = f_{\underline{x}(k) | \underline{Y}(k), \underline{a}} \cdot \prod_{j=1}^k f_{\underline{y}(j) | \underline{Y}(j-1), \underline{a}} \quad (3-47)$$

Die passende Likelihood-Funktion $L(\underline{\vartheta}(k), \underline{Y}_k)$ folgt dann durch Logarithmierung der bedingten Verteilungsdichte mit:

$$L(\underline{\vartheta}(k), \underline{Y}_k) = \ln f_{\underline{x}(k) | \underline{Y}_k, \underline{a}} + \sum_{j=1}^k \ln f_{\underline{y}(j) | \underline{Y}_{j-1}, \underline{a}} \quad (3-48)$$

Die Likelihoodfunktion sei - allgemein betrachtet - an der Stelle $\underline{\vartheta}(k) = \underline{\vartheta}^*$ maximal, so daß die die folgende Bedingung der allgemeinen Likelihood-Funktion erfüllt sein muß:

$$\left. \frac{\partial L[\underline{\vartheta}(k), \underline{Y}_k]}{\partial \underline{\vartheta}(k)} \right|_{\underline{\vartheta}(k) = \underline{\vartheta}^*(k)} = \underline{0}^T \quad (3-49)$$

$\underline{\vartheta}(k) = \underline{\vartheta}(\underline{\xi}, \underline{\alpha})$ sei ein Vektor, bestehend aus dem Zustandsvektor $\underline{\xi}$, der Schätzgrößen und einem adaptiven Parameter(vektor) $\underline{\alpha}$, der allgemein die adaptiven Parametergrößen repräsentieren soll: $\underline{\vartheta} = \begin{bmatrix} \underline{\xi} \\ \underline{\alpha} \end{bmatrix}$. Die partielle Ableitungen nach $\underline{\xi}$ und $\underline{\alpha}$, den Elementen des Vektors

$\underline{\vartheta} = \begin{bmatrix} \underline{\xi}^T \\ \underline{\alpha}^T \end{bmatrix}^T$, lauten:

$$\left. \frac{\partial L[\underline{\vartheta}(k), \underline{Y}_k]}{\partial \underline{\vartheta}(k)} \right|_{\underline{\vartheta}(k) = \underline{\vartheta}^*(k)} = \left. \frac{\partial L[\underline{\vartheta}(k), \underline{Y}_k]}{\partial \underline{\xi}(k)} \right|_{\underline{\xi}(k) = \underline{\hat{x}}^*(k)} + \left. \frac{\partial L[\underline{\vartheta}(k), \underline{Y}_k]}{\partial \underline{\alpha}(k)} \right|_{\underline{\alpha}(k) = \underline{a}^*(k)} = \underline{0}^T \quad (3-50)$$

Die Lösung für die Filterzustände führt zu dem bereits bekannten Ergebnis einer Maximum-Likelihood-Schätzung /48/:

$$\left. \frac{\partial L[\underline{\vartheta}(k), \underline{Y}_k]}{\partial \underline{\xi}(k)} \right|_{\underline{\xi}(k) = \underline{\hat{x}}^*(k)} = -[\underline{\xi} - \underline{\hat{x}}] \cdot P_k^{+1} \Big|_{\underline{\hat{x}}^*(k), \underline{a}^*(k)} = \underline{0}^T \quad (3-51)$$

Als Lösungen erhält man die Schätzwerte der Rauschvarianzen Q und R , die als konstant angenommen werden:

$$\begin{aligned} P_k^- &= F \cdot P_k^+ \cdot F^T + G \cdot \hat{Q} \cdot G^T \\ K_k &= P_k^- \cdot H_{\text{Diff}} \cdot [H_{\text{Diff}} \cdot P_k^- \cdot H_{\text{Diff}}^T + \hat{R}]^{-1} \end{aligned} \quad (3-52)$$

Die zweite partielle Ableitung (adaptive Parameter) der Likelihood-Funktion dagegen führt zu folgendem Term:

$$\begin{aligned} \left. \frac{\partial L[\underline{\vartheta}(k), \underline{Y}_k]}{\partial \underline{\alpha}(k)} \right|_{\underline{\alpha}(k) = \underline{a}^*(k)} &= \text{tr} \left\{ P^{+1} \cdot \frac{\partial P^+}{\partial \underline{\alpha}_k} \right\} \Big|_{\underline{\alpha}(k) = \underline{a}^*(k)} - 2 \cdot \sum_{j=1}^k \frac{\partial \underline{\hat{x}}^T}{\partial \underline{\alpha}_k} \cdot H_{\text{Diff}}^T \cdot A_j^{-1} \cdot [\underline{y}_j - \underline{h}(\underline{\hat{x}}_j^-)] \Big|_{\underline{\alpha}(k) = \underline{a}^*(k)} + \\ &\sum_{j=1}^k \text{tr} \left\{ [A_j - A_j^{-1} \cdot [\underline{y}_j - \underline{h}(\underline{\hat{x}}_j^-)] \cdot [\underline{y}_j - \underline{h}(\underline{\hat{x}}_j^-)]^T \cdot A_j^{-1}] \frac{\partial A_j}{\partial \underline{\alpha}_k} \right\} \Big|_{\underline{\alpha}(k) = \underline{a}^*(k)} = \underline{0}^T \end{aligned} \quad (3-53)$$

Dabei wurden die folgenden Zusammenhänge genutzt laut /48/:

$$\frac{\partial}{\partial \alpha_{k0}} \ln |M| = \left[\frac{\partial}{\partial |M|} \ln |M| \right] \cdot \frac{\partial |M|}{\partial \alpha_{k0}} = \frac{1}{|M|} \cdot \frac{\partial |M|}{\partial \alpha_{k0}}$$

$$\begin{aligned}
 &= \text{tr} \left\{ \mathbf{M}^{-1} \cdot \frac{\partial \mathbf{M}}{\partial \alpha_{k0}} \cdot \mathbf{M}^{-1} \right\} \\
 \frac{\partial}{\partial \alpha_{k0}} \mathbf{M}^{-1} &= -\mathbf{M}^{-1} \cdot \frac{\partial \mathbf{M}}{\partial \alpha_{k0}} \cdot \mathbf{M}^{-1} \quad (3-54)
 \end{aligned}$$

Laut /48/ läßt sich der erste Term von (3-53) vernachlässigen, da er kleiner als die Summe des dritten Terms ist. Beim zweiten Term von (3-53) ist die Ableitung vernachlässigbar bei geeigneter Wahl eines α_{k0} . Es verbleibt:

$$\left. \frac{\partial L[\underline{\vartheta}(k), \underline{Y}_k]}{\partial \underline{\alpha}(k)} \right|_{\underline{\alpha}(k) = \underline{\alpha}^*(k)} = \sum_{j=1}^k \text{tr} \left\{ \left[\mathbf{A}_j - \mathbf{A}_j^{-1} \cdot \left[\underline{y}_j - \underline{h}(\hat{\underline{x}}_j^-) \right] \cdot \left[\underline{y}_j - \underline{h}(\hat{\underline{x}}_j^-) \right]^T \cdot \mathbf{A}_j^{-1} \right] \frac{\partial \mathbf{A}_j}{\partial \underline{\alpha}_k} \right\} \Big|_{\underline{\alpha}(k) = \underline{\alpha}^*(k)} = \underline{\mathbf{0}}^T$$

Als 1. Möglichkeit besteht die Schätzung der Beobachtungsrauschleistungsvarianz $\hat{\mathbf{R}} : \mathbf{Q}$ wird dabei als konstant und exakt bekannt angenommen; $\underline{\alpha}_k$ entspricht dabei einem Vektor \underline{r}_j , der aus den zu schätzenden Kovarianzwerten der Matrix R gebildet wird.

$$\text{mit: } \mathbf{H}_{\text{Diff}} \cdot \frac{\partial \mathbf{P}_j^-}{\partial \underline{r}_j} \cdot \mathbf{H}_{\text{Diff}} \approx 0 ; \quad \frac{\partial \mathbf{R}}{\partial \underline{r}_j} \neq 0 \quad \text{und mit } \underline{r} = \begin{bmatrix} \mathbf{R}_{11} \\ \mathbf{R}_{22} \end{bmatrix}$$

$$\underline{\mathbf{0}}^T = \sum_{j=1}^k \left\{ \left[\mathbf{A}_j^{-1} - \mathbf{A}_j^{-1} \cdot \left[\underline{y}_j - \underline{h}(\hat{\underline{x}}_j^-) \right] \cdot \left[\underline{y}_j - \underline{h}(\hat{\underline{x}}_j^-) \right]^T \cdot \mathbf{A}_j^{-1} \right] \right\}$$

Mit der Annahme, daß \mathbf{A}_j^{-1} im betrachteten Bereich ungefähr konstant ist, wird eine Schätzung für die Matrix $\hat{\mathbf{R}}_k$, die Kovarianzmatrix des Meßrauschens, angenommen:

$$\Rightarrow \hat{\mathbf{R}}_k = \frac{1}{N} \sum_{j=k-N+1}^k \text{res}_j \text{res}_j^T - \mathbf{H}_{\text{Diff}} \mathbf{P}_j^- \mathbf{H}_{\text{Diff}}^T \cong \left[\frac{1}{N} \sum_{j=k-N+1}^k \text{res}_j \text{res}_j^T \right] - \mathbf{H}_{\text{Diff}} \mathbf{P}_k^- \mathbf{H}_{\text{Diff}}^T$$

Als 2. Möglichkeit besteht außerdem die Schätzung der Systemrauschleistungsvarianz $\hat{\mathbf{Q}}$:

Ähnlich wie bei der Schätzung von R wird hier R als konstant und exakt bekannt angenommen; $\underline{\alpha}_k$ entspricht dabei einem Vektor \underline{q}_j , der aus den zu schätzenden Kovarianzwerten der Matrix Q gebildet wird.

$$\text{mit: } H_{\text{Diff}} \cdot \frac{\partial P_j^-}{\partial \underline{q}_j} \cdot H_{\text{Diff}}^T = H_{\text{Diff}} \cdot \left[A \cdot \frac{\partial P_{j-1}^+}{\partial \underline{q}_j} \cdot A^T + G \cdot \frac{\partial Q_{j-1}}{\partial \underline{q}_j} \cdot G^T \right] \cdot H_{\text{Diff}}^T \neq 0$$

$$\begin{aligned} 0 &= \sum_{j=1}^k \left[A_j^{-1} - A_j^{-1} \cdot \left[\underline{y}_j - \underline{h}(\hat{\underline{x}}_j^-) \right] \cdot \left[\underline{y}_j - \underline{h}(\hat{\underline{x}}_j^-) \right]^T \cdot A_j^{-1} \right] \\ &= \sum_{j=1}^k \left[A_j^{-1} - \text{res}_j \text{res}_j^T \right] \end{aligned}$$

Es muß weiterhin angenommen werden, daß man sich im eingeschwungenen Zustand (stationären Betrieb) befindet. Dann ist die Änderung des Kalman-Gains K_k vernachlässigbar. Also läßt sich die Gleichung auch mit $K_k = K$ erweitern:

$$\begin{aligned} 0 &= K \cdot \sum_{j=k-N+1}^k \left[A_j^{-1} - \text{res}_j \text{res}_j^T \right] \cdot K^T \\ &= \sum_{j=k-N+1}^k K \cdot \left[A_j^{-1} - \text{res}_j \text{res}_j^T \right] \cdot K^T \end{aligned}$$

$$\text{Mit: } K A_j K^T = P_j^- H_{\text{Diff}} K^T = P_j^- - P_j^+$$

$$= F P_{j-1}^+ F^T + G Q_{j-1} G^T - P_j^+$$

$$\Rightarrow 0 = \sum_{j=k-N+1}^k F P_{j-1}^+ F^T + G Q_{j-1} G^T - P_j^+ - K \text{res}_j \text{res}_j^T K^T \quad (3-55)$$

Mit (3-55) wird eine Schätzung für die Matrix \hat{Q}_k , die Kovarianzmatrix des Systemrauschens definiert:

$$\Rightarrow \hat{Q}_k = \frac{1}{N} \sum_{j=k-N+1}^k G^{-1} \cdot \left(K \text{res}_j \text{res}_j^T K^T + P_j^+ - F P_{j-1}^+ F^T \right) \cdot \left(G^T \right)^{-1} \quad |^{11} \quad (3-56)$$

¹¹Nur wenn die Inverse von G existiert, läßt sich weiter nach Q_{j-1} auflösen. Bei Nichtexistenz von G^{-1} kann die Matrix auch durch seine Pseudo-Inverse ersetzt werden $G^\# = \left[G^T \cdot G \right]^{-1} \cdot G^T$

Über einen Zeitraum von N Meßwerten wird folglich das Driving Noise Q bestimmt und gemittelt. So erhält man eine zusätzliche Stabilität im Hinblick auf Dynamikschwankungen. Da Q ohnehin ein langsam variierender Parameter ist, ergibt sich die Möglichkeit, eine Neuberechnung von Q nur in festgelegten Zeitintervallen durchzuführen.

Die beschriebene Methode kann mit folgendem System- und Kalman-Filtermodell anschaulich beschrieben werden:

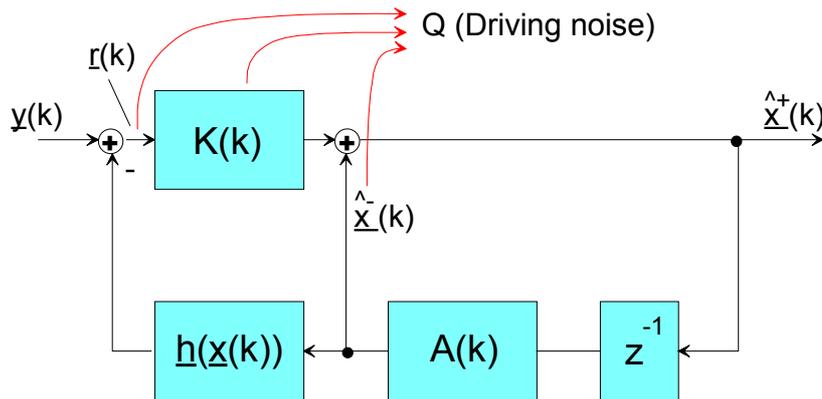


Abb. 3.17: Kalman-Filtermodell

3.4 SMOOTHER

Generell können Schätzwerte in Abhängigkeit von drei unterschiedlichen Ausgangszeitpunkten bestimmt werden. Der Filterschätzwert wird bei der Prädiktion abhängig von vergangenen Meßwerten und beim Filterschritt vom aktuellen Meßwert berechnet. Beim Smoothen werden zusätzlich Meßwerte aus der Zukunft (in bezug auf den Filterzeitpunkt k) zur Estimation mit herangezogen und sich von Vergangenheit und Zukunft mit 2 Filtern dem Filterzeitpunkt genähert. Der gesuchte Schätzwertverlauf wird somit zusätzlich geglättet bzw. an den tatsächlichen Verlauf angenähert. Als Smoothen wird dann meist die Filterung aus der Zukunft zurück (oder auch Backwards-Filtering) bezeichnet /3, 48, 49, 54, 74/. Für den Schätzwert $\hat{x}(k|j)$ gilt:

Prädiktion (Prediction): $k > j$

Filterung (Filtering): $k = j$

Smoothen (Smoothing): $k < j$

Smoothers werden daher generell im Offline-Betrieb eingesetzt. Betrachtet man im Online-Betrieb jedoch bestimmte Intervalle T , so lassen sich auch die Eigenschaften von Smoothern testen. Mit einer zwischengespeicherten Datenmenge von N Meßwerten läßt sich eine Filterverarbeitung mit Smoothern testen. Unterschiedliche Smoother-Verfahren werden beschrieben, auf die Online-Demodulation winkelmodulierter Signale angewendet und getestet. Drei grundsätzliche Typen von Smoothern werden dabei unterschieden: Fixed-Lag Smoother, Fixed-Intervall Smoother und Fixed-Point Smoother. Dabei werden unterschiedliche Strategien verwendet, um den Schätzwert zu optimieren. Bei allen Smoothern werden Rückwärtsfilter (Backward-Filtering) eingesetzt, so daß ein Nachlaufen des Filterschätzwertes gegenüber dem korrekten Verlauf entsteht, die bei der Auswertung berücksichtigt werden muß. Die Smoothertypen werden in den folgenden Unterpunkten beschrieben.

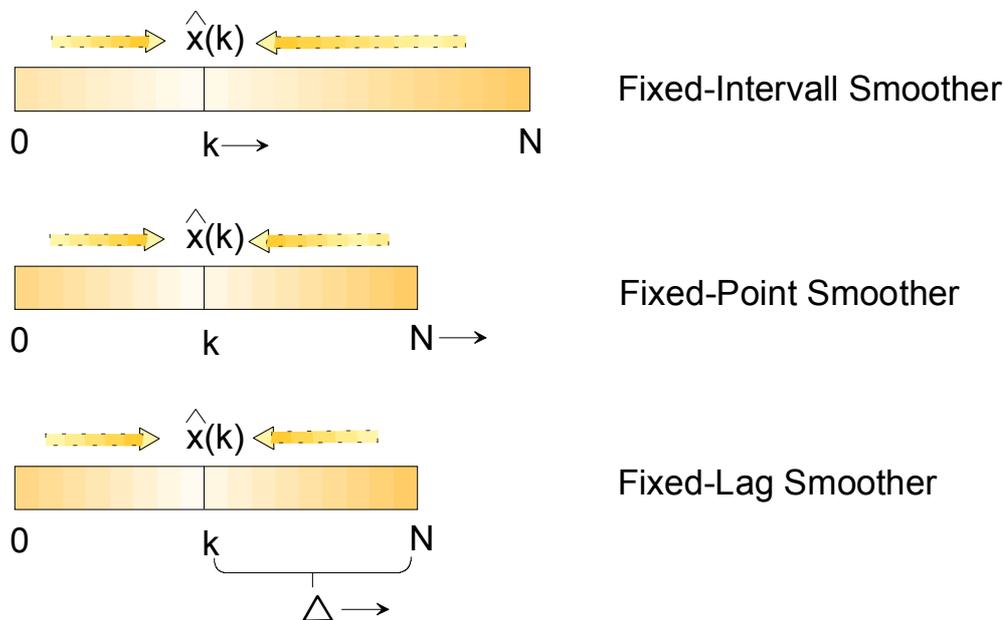


Abb. 3.18: Smoothertypen

3.4.1 Fixed-Intervall Smoothing

Nach einer durchgeführten Messung werden alle Meßwerte zur Berechnung eines einzigen Schätzwertes $\hat{x}(k | j)$ herangezogen.

$$\hat{x}(k | f) = E\{\underline{x}(k) | \underline{Y}(f) = \underline{T}_f\}$$

$$k = 0, 1, 2, \dots, f - 1,$$

alle Meßwerte, $f =$ fester Endzeitpunkt

$$j = f$$

Dieses Schätzverfahren kann nur zur Nachbearbeitung von Meßwerten verwendet werden - nicht zur Online-Verarbeitung! Da immer alle Meßwerte miteinberechnet werden, ist der Rechenaufwand entsprechend groß.

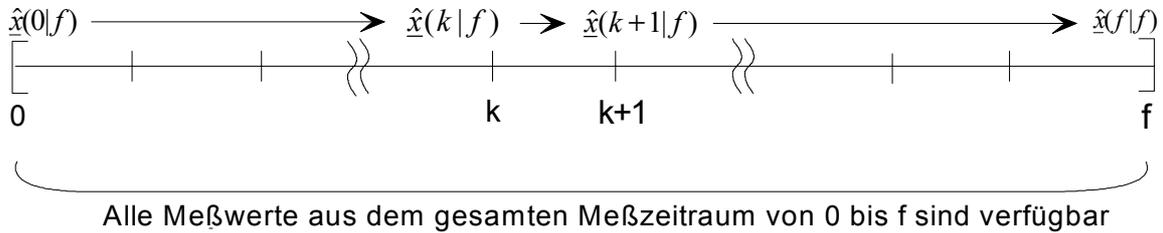


Abb. 3.19: Schätzwertbestimmung beim Fixed-Intervall Smoother

3.4.1.1 Fraser-Potter Smoother

Einem gewöhnlichen vorwärts laufenden Kalman-Filter wird ein rückwärts laufendes Filter angehängt. Dieses verarbeitet die Meßwerte vom Endzeitpunkt $k=f$ an rückwärts bis $k=1$ zum zweiten Mal im sogenannten Rückwärtsfilter. Im dritten Verarbeitungsschritt wird der Schätzwert bestimmt. Dieser kann jedoch auch gleichzeitig zum Rückwärtsfilter berechnet werden /34/.

Vorwärtsfilter:

$$\begin{aligned} P_f^-(k+1) &= AP_f^+(k)A^T + GQG^T \\ \hat{x}_f^-(k+1) &= A\hat{x}_f^+(k) + \underline{u}(k) \end{aligned} \quad (3-57)$$

Rückwärtsfilter:

$$\begin{aligned} P_b^-(k) &= A^{-1}[P_b^+(k+1) + GQG^T](A^{-1})^T \\ \hat{x}_b^-(k) &= A^{-1}\hat{x}_b^+(k+1) - A^{-1}\underline{u}(k) \end{aligned} \quad (3-58)$$

Gesamt-Smootherfilter:

$$\begin{aligned} P_s(k) &= P_b^-(k)(P_f^+(k) + P_b^-(k))^{-1}P_f^+ \\ &= [P_f^+(k)^{-1} + P_b^-(k)^{-1}]^{-1} \\ \hat{x}_s(k) &= [P_f^+(k)^{-1} + P_b^-(k)^{-1}]^{-1} \cdot [P_f^+(k)^{-1}\hat{x}_f^+(k) + P_b^-(k)^{-1}\hat{x}_b^-(k)] \\ &= [I - U]\hat{x}_f^+(k) + [I - V]\hat{x}_b^-(k) \\ &= P_s(k) \cdot [P_f^+(k)^{-1}\hat{x}_f^+(k) + P_b^-(k)^{-1}\hat{x}_b^-(k)] \end{aligned} \quad (3-59)$$

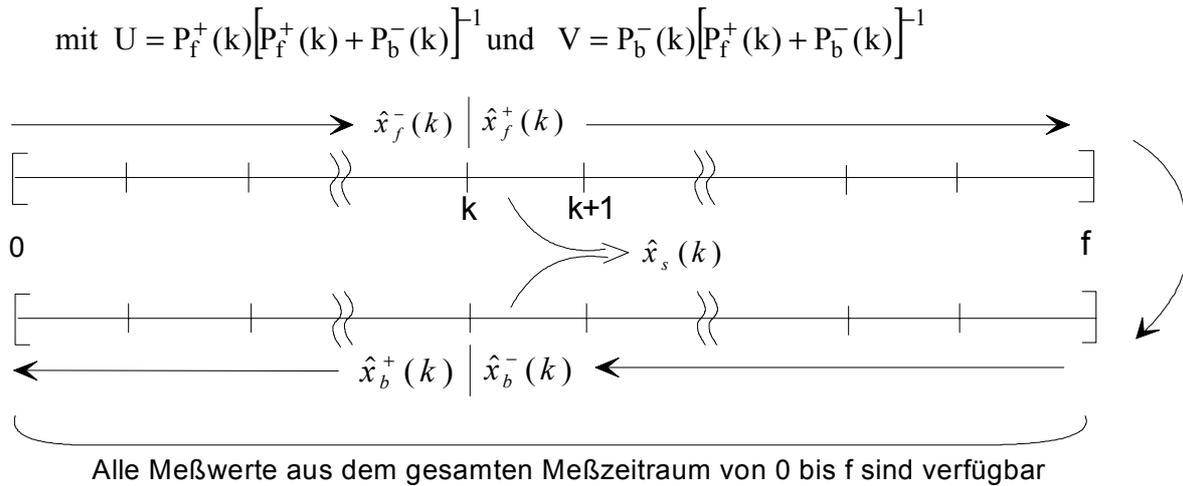


Abb. 3.20: Schätzwertbestimmung beim Fraser-Potter Smoother

Verarbeitungsverlauf des Smootherfilters: Nach Berechnung der Schätzwerte aus Vorwärtsfilter (forward filtering) und Rückwärtsfilter (backward filtering) werden Smootherschätzwert und Smootherskovarianz bestimmt. Je nach Genauigkeit des jeweiligen Filters fließen die Filterschätzwerte unterschiedlich in den Smootherschätzwert ein.

3.4.1.2 Rauch-Tung-Striebel-Smoother oder Meditch-Filter

Durch Weiterentwicklung des Fraser-Potter-Smoother gelangt man zum Rauch-Tung-Striebel-Smoother [52]. Die Filterdurchgänge werden durch fortgeschrittene Herleitung auf einen reduziert, d.h. die Rückwärtsfilterung wird in die Vorwärtsfilterung mit hineingezogen. Er zeichnet sich durch eine programmierfreundliche Formulierung des Rückwärtsfilters aus. Durch Vermeiden von inversen Matrizen kann der Filteralgorithmus des Fixed-Intervall Smoothers stabilisiert und schneller bearbeitet werden. Resultierend daraus wird die Rechenzeit verkürzt, jedoch eine höhere Rechenungenauigkeit zugelassen.

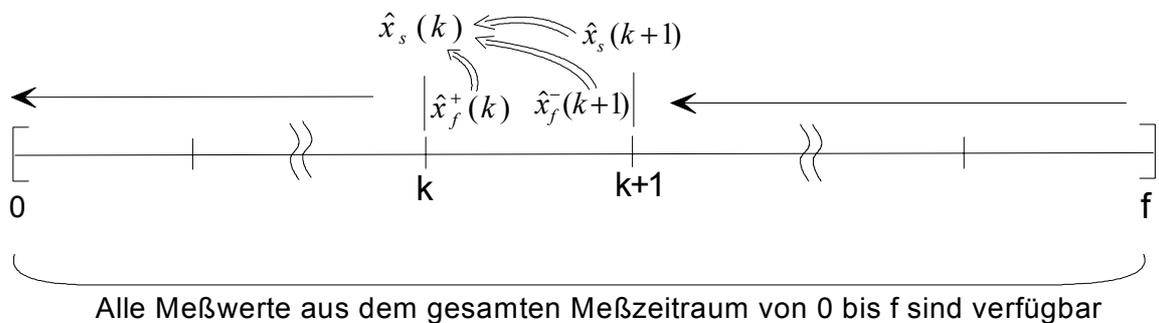


Abb. 3.21: Schätzwertbestimmung beim Rauch-Tung-Striebel Smoother

$$\begin{aligned}
 P_s(k) &= P_f^+(k) - K_{GS}(k) [P_f^+(k) - P_s(k)] K_{GS}^T \\
 \hat{x}_s(k) &= \hat{x}_f^+(k) + K_{GS}(k) \cdot [\hat{x}_s^-(k+1) - \hat{x}_f^-(k)]
 \end{aligned}
 \tag{3-60}$$

mit $K_{GS}(k) = P_f^+(k) \cdot F^T \cdot P_f^-(k+1)^{-1}$

Insgesamt kann jedoch von einem Gewinn bei beiden Smoothern von bis zu 100% ausgegangen werden. Fraser-Ansatz und Meditch-Ansatz sind theoretisch betrachtet äquivalent, d.h. die Filtergleichungen des einen Ansatzes lassen sich in die Gleichungen des anderen durch algebraische Umformungen überführen. Trotz des etwas unterschiedlichen Ansatzes können wir also keine unterschiedlichen Meßergebnisse erwarten.

3.4.2 Fixed-Point Smoother

Ein Fixed-Point Smoother wird eingesetzt, wenn zu einem bestimmten Zeitpunkt k ein genauer Schätzwert $\hat{x}(k|j)$ gesucht wird. Ausgewertet werden dann alle Meßwerte, die bis zu diesem Zeitpunkt gemacht wurden. Ebenso werden alle anfallenden Meßwerte, bezogen auf diesen Zeitpunkt k , mit einberechnet! Der gesuchte Schätzwert wird mit fortschreitender Zeit immer genauer.

Die Realisierung für Online-Applikation macht wenig Sinn, da t immer mitlaufen muß. Beim Fixed-Point Smoother läuft dagegen nur die Gesamtzeit T weiter, um den einen Punkt t genauer zu schätzen.

3.4.3 Fixed-Lag Smoother

Beim Fixed-Lag Smoothing liegt der errechnete Schätzwert immer um N Abtastzeitpunkte zurück (Verzögerung = Fixed Lag). Es werden immer alle vorhandenen Meßwerte zur Bestimmung eines Schätzwertes mit einbezogen.

$$\begin{aligned}
 \hat{x}(k|k+N) &= E\{\underline{x}(k) | \underline{Y}(k+N) = \underline{T}_{k+N}\} \\
 \hat{x}(k|j) &= E\{\underline{x}(k) | \underline{Y}(j) = \underline{T}_j\}
 \end{aligned}
 \tag{3-61}$$

$k = f - N, \quad N < f, \quad N = \text{konstante, natürliche Zahl}$
 $j = k + N$

Bei der Implementierung mit einem Kalman-Filter baut sich der Fixed-Lag zu Beginn bis zur Länge N auf.

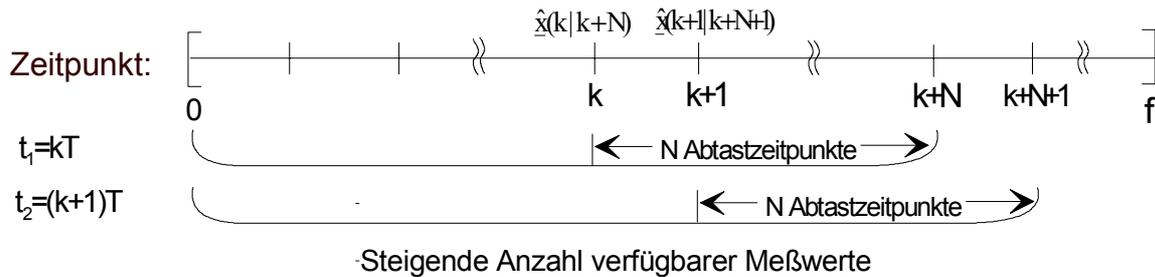


Abb. 3.22: Schätzwertbestimmung beim Fixed-Lag Smoother

Eigenschaften: Bei zeitdiskreter Realisierung entsteht ein finitdimensionales System. Sinnvollerweise wird das System rekursiv entwickelt. Der Algorithmus berechnet jeweils die letzten N Mittelwerte und Kovarianzen. Die Fehlerkovarianz ist ebenso N-fach geglättet/verbessert worden. Die Residuensequenz bleibt beim Smoothermodell identisch mit dem des Kalman-Filtermodells. Die beschriebenen Smoothermodelle eignen sich aufgrund seines enormen Rechenaufwandes nur für Offline-Anwendungen. Um Smoother auch für Online-Anwendungen benutzen können, müssen Modifikationen vorgenommen werden (s. 3.4.5).

3.4.3.1 Fixed-Lag Smoother durch Zustandsraumerweiterung

Eine weitere Realisierung des Fixed-Lag Smoothermodells in einem einzigen Filter kann durch Erweiterung des Zustandsraums erlangt werden. Die Filtergleichungen lauten dann folgendermaßen mit $u(k)=0$, $G(k)=I$ und einem Markov-Prozeß zur Zustandsgenerierung /3, 79/ (auch Delayed States Modell genannt, was die schieberegisterartige Funktionsweise beschreibt):

$$\underline{x}(k+1) = A(k) \cdot \underline{x}(k) + \underline{u}(k) + G(k) \cdot \underline{w}(k)$$

$$= \begin{bmatrix} 1 & -T & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \\ x_4(k) \\ x_5(k) \\ x_6(k) \end{bmatrix} + \begin{bmatrix} 0 \\ w_2(k) \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

**Zustands-
gleichung**

$$\underline{y}(k) = \underline{h}(\underline{x}(k)) + \underline{v}(k) = \begin{bmatrix} \text{Re}\{r_T(k)\} \\ \text{Im}\{r_T(k)\} \end{bmatrix} + \begin{bmatrix} v_1(k) \\ v_2(k) \end{bmatrix}$$

**Beobachtungs-
gleichung**

(3-62)

Im dargestellten Filter sind 2 "Lags" realisiert. Um einen wirklich effektiven Fixed-Lag Smoother zu realisieren, benötigt man mindestens 30 Lags, da diese Anzahl der Filterschritte zum Einschwingen notwendig sind (s. 4.4.3). Die Anzahl der Zustände läge dann bei 60! Die

Filterlaufzeit steigt exponentiell mit den Zuständen an, so daß eine derartige Realisierung für Online-Anwendung keinen Sinn macht.

3.4.3.2 Switched-Fixed-Lag Smoother

Bei auftretenden sprunghaften Meßverläufen können diese mit einem Generalized Likelihood-Ratio-Verfahren bestimmt und an den Kalman-Filter übergeben werden (Adaptiver Kalman-Filter). Wird ein Sprung detektiert, dann wird der Kalman-Filter auf die Default-Werte (Reinitialisierung) zurückgesetzt. Durch eine deutliche Vergrößerung von N , was einen deutlich höheren Rechenaufwand bedeutet, werden jedoch nur geringfügige Verbesserungen erreicht.

3.4.4 Smoothability:

Fraser hat dieses Kriterium eingeführt, um die Notwendigkeit eines Smoothers abschätzen zu können. Sie besagt nichts anderes, als daß mit einem Smoother (einer zusätzlichen Rückwärtsfilterung) eine Verbesserung eintreten muß, um ihn zu rechtfertigen. Ein Unterschied bzw. eine Verbesserung beim Einsatz eines Rückwärtsfilters ergibt sich nur dann, wenn dieser auch durch Prozeßrauschen gesteuert wird. D.h.: Ist das Prozeßrauschen gleich Null, wird sich durch einen Smoother-Filter keine Verbesserung ergeben. Auf eine mathematische Überprüfung soll hier verzichtet werden. Eine praktische Überprüfung der Fisher'schen Information liefert einen kontinuierlichen Informationszuwachs.

3.4.5 Smoothermodell für die Online-Anwendung

Für die Realisierung eines Online-Filters bietet sich folgende Vorgehensweise an:

Lädt man die abgetasteten Meßwerte in einem sogenannten Ringpuffer der Puffergröße von N -Werten, dann stellt N die Maximalgröße \mathbf{f} der zur Verfügung stehenden Daten dar. Wir können somit den Fixed-Intervall Smoother anwenden und $(N-1)$ -Werte im Rückwärtsfilter nutzen. Beim nachfolgenden nächsten Meßwert wird der älteste Meßwert überschrieben und die Rückwärtsfilterung beginnt dann ausgehend vom jetzt aktuellsten Meßwert. Die Vorwärtsfilterung wird dabei fortwährend vom Standardfilter EKF22 durchgeführt. (Zu Beginn einer derartigen Smootherfilterung müssen zunächst N Meßwerte gesammelt werden, bevor eine Rückwärtsfilterung möglich ist.)

Um sinnvolle Startwerte für \hat{x}_b^+ und P_b^+ bei der Rückwärtsfilterung zu erhalten, wird im folgenden Modell bis zum Intervallende (aktuellster Meßwert) vorgefiltert und die dann aktuellen Parameter für \hat{x}_b^+ und P_b^+ eingesetzt. Dadurch läßt sich das Einschwingen des Filters verringern und somit auch die Puffergröße reduzieren. Einfacher gestaltet es sich wie im obigen Fall, immer mit denselben, konstanten Startwerten zu beginnen. Eine weitere Möglichkeit bietet sich durch Heranziehen der Arcustangens-Funktion einen einfachen Schätzwert der Phase zu bestimmen und diese als Startwert einzusetzen. Das Einschwingen des Rückwärtsfilters läßt sich so ebenfalls reduzieren.

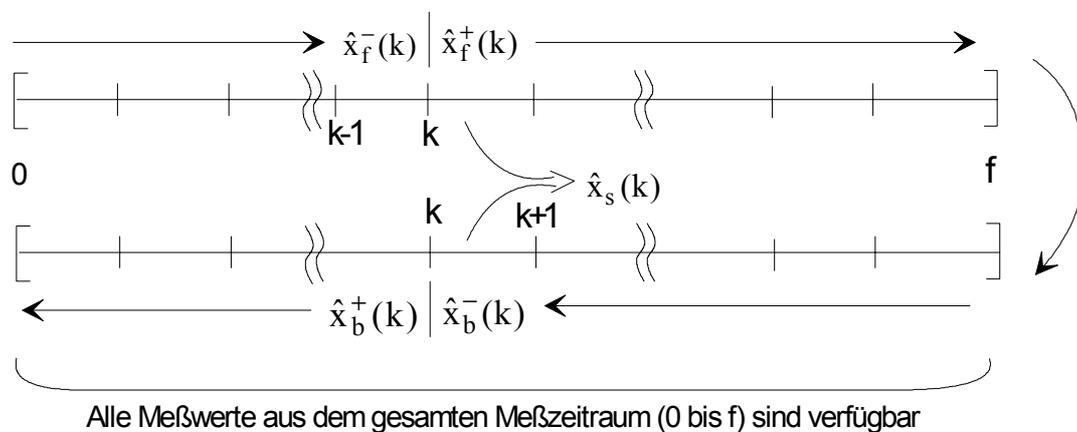


Abb. 3.25: Schätzwertbestimmung $\hat{x}_s(k|k-1, k+1)$ durch Vorfilterung der Startwerte des Rückwärtsfilters

3.4.5.1 Fixed-Lag Smoother (Reduced Order Filter: SKF22ro)

Durch weitere Vereinfachungen eines Filters mit Zustandsraumerweiterung (s. 3.4.3.1) läßt sich das Smootherfilter auf ein stark reduziertes Zustandsmodell mit der Standardbeobachtungsgleichung begrenzen. Die zwei Filterrichtungen werden dann in einem einzigen Filter vereint. Die Rückwärtsfilterung ist jedoch stark reduziert auf vier Schritte, die jedoch unterschiedlich stark gewichtet auf die Vorwärtsfilterung direkt mit eingehen.

Die Filtergleichungen lauten dann folgendermaßen mit $u(k)=0$ und $G(k)=I$:

$$\underline{x}(k+1) = A(k) \cdot \underline{x}(k) + \underline{u}(k) + G(k) \cdot \underline{w}(k) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ a_1 & a_2 & a_3 & a_4 \end{bmatrix} \cdot \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \\ x_4(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ w_4(k) \end{bmatrix} \quad \text{Zustands-} \\ \text{gleichung}$$

$$\underline{y}(k) = \underline{h}(\underline{x}(k)) + \underline{v}(k) = \begin{bmatrix} \operatorname{Re}\{r_T(k)\} \\ \operatorname{Im}\{r_T(k)\} \end{bmatrix} + \begin{bmatrix} v_1(k) \\ v_2(k) \end{bmatrix} \quad \text{Beobachtungs-} \\ \text{gleichung}$$

(3-63)

3.5 WEITERE FILTEROPTIMIERUNGEN

Bei der Entwicklung und dem Test der unterschiedlichen bereits betrachteten Filtermodelle kristallisieren sich die "kritischen" Parameter heraus, die für einen stabilen und präzisen Betrieb notwendig sind. Es zeigt sich jedoch ebenso, daß eine Vielzahl von Parametern sich nach einem Einschwingvorgang kaum noch ändern, so daß sie aus der dauerhaften Berechnung herausgenommen werden können und nur noch zu Beginn der Filterung (\Rightarrow Vorfilterung $\hat{=}$ engl.: Prefiltering) oder temporär, d.h. nach definierten Zeitabschnitten, bestimmt werden müssen. Dazu zählen insbesondere die Filter- und Prädiktionsfehlerkovarianzen P^+ , P^- . Diese Parameter laufen bei linearen Kalman-Filtern gegen deterministische Grenzwerte, die abhängig von R und Q sind. Ebenso ist q ein Parameter, der sich, falls nicht konstant angenommen, nur relativ langsam verändert. Bei Extended Kalman-Filtern kann ein ähnliches Verhalten beobachtet werden, was jedoch in Abhängigkeit von den durchgeführten Linearisierungen steht.

3.5.1 Prefiltering

Bei der Anwendung von Kalman-Filtern läßt sich beobachten, daß die Kovarianzen einem Grenzwert entgegenstreben. Um weitere Berechnungsschritte einzusparen, setzt man die Prädiktionsfehlerkovarianz P^- konstant und reduziert somit den Filteraufwand erheblich. Bei der Realisierung EKF22pr wurde eine Vorfilterung von 50 Filterschritten gewählt, nach denen (bei halbwegs sinnvoller) Parameterwahl der Grenzwert immer erreicht wurde.

Grundsätzlich läßt sich diese Art der Vorfilterung bei allen Filtern anwenden. Eine Verwendung der Vorfilterung bei adaptiven Filtern macht die adaptive Steuerung von Q überflüssig, da dieser Parameter bei der Prädiktion verarbeitet wird. Die nachfolgende Filterung ist dann nicht mehr adaptiv, sondern identisch mit dem EKF22pr.

3.5.2 Temporal Update

Ebenso ist es möglich in konstanten Zeitintervallen die Kovarianz zur Kontrolle neu zu bestimmen und, falls notwendig, die verwendete Kovarianz anzupassen. Ein derartiges Filter wurde hier nicht explizit realisiert. Jedoch kann man sich die Entwicklungsumgebung zur Online-Demodulation entsprechend vorstellen, da hier maximal alle 1000 Filterschritte das Filter erneut gestartet wird. Übernimmt man die Parameter des abgeschlossenen Durchlaufs für den nächsten Durchlauf und paßt die neu ermittelten Parameter an die alten an, so hat man ein derartiges Filter realisiert.

3.5.3 EKF zur Verarbeitung digitaler Modulationsverfahren

Zur Demodulation digitaler Modulationsverfahren haben z.B. Collings und Moore ein Filtermodell entwickelt /16/, das im zylindrischen Koordinatensystem arbeitet. Die Phasenzustände einer PSK werden somit als Phasenzustände des Filtermodells verwendet. Als Beobachtungen werden erneut die Quadraturkomponenten gewählt. Als Zustandsvektor wird hier ebenfalls ein Vektor aus Real- und Imaginärteil verwendet. Die Betrachtung im zylindrischen Koordinatensystem bedeutet aber zugleich eine lineare Filterung, da die Beobachtungen \underline{y} in linearem Zusammenhang zu Real- und Imaginärteil der Phasenzustände stehen. Außerdem wurde mit Hilfe einer Rotationsmatrix die Beschreibung der Phasenänderungen der unterschiedlichen Phasenstufen durchgeführt.

Aufgrund der hohen Anzahl an benötigten Zuständen wurde das Filter nicht implementiert.

Die Definition des Zustandsvektors $\underline{x}(k)$ wird also folgendermaßen aussehen:

$$\underline{x}(k) = \underline{g}(k) = \begin{pmatrix} \underline{g}_k^R \\ \underline{g}_k^I \end{pmatrix} = \begin{pmatrix} \kappa_k \cos(\phi_k) \\ \kappa_k \sin(\phi_k) \end{pmatrix}$$

Verschiebt man die Beobachtungen \underline{y} um einen Winkel θ_k , dann lauten die Quadraturkomponenten:

$$y_1(k) = A_c \kappa_k \cos(\theta_k + \phi_k) = A_c \cos(\theta_k) \cdot \kappa_k \cos(\phi_k) - A_c \sin(\theta_k) \cdot \kappa_k \sin(\phi_k)$$

$$y_2(k) = A_c \kappa_k \sin(\theta_k + \phi_k) = A_c \cos(\theta_k) \cdot \kappa_k \sin(\phi_k) + A_c \sin(\theta_k) \cdot \kappa_k \cos(\phi_k) \quad (3-64)$$

Durch Anwendung der Additionstheoreme und Abspalten des Zustandsvektors ergibt sich für die Beobachtungsgleichung des Systemmodells:

$$\begin{aligned} \underline{y}(k) = \begin{pmatrix} y_{1k} \\ y_{2k} \end{pmatrix} &= C \cdot \underline{g}_k + \underline{w}_k \\ &= \begin{pmatrix} A_c \cos(\theta_k) - A_c \sin(\theta_k) \\ A_c \sin(\theta_k) + A_c \cos(\theta_k) \end{pmatrix} \cdot \begin{pmatrix} g_k^R \\ g_k^I \end{pmatrix} + \begin{pmatrix} w_k^R \\ w_k^I \end{pmatrix} \end{aligned} \quad (3-65)$$

Die Matrix C hat hier die Funktion einer Rotationsmatrix.

4 Ergebnisbetrachtungen der Simulationen und Online - Realisierung

Entscheidend für aussagekräftige Messungen sind die Kenntnisse über den Meßaufbau, d.h. über äußere Einflüsse, Parameter und Randbedingungen, die durch Hardware und Software definiert sind. Bei der Simulation eines Filters sind dabei insbesondere die durch Software gegebenen Randbedingungen entscheidend, bzw. bei praktischen Filterrealisierungen liegt der Schwerpunkt auf Hardware- und Softwareeinflüssen.

Bei der Simulation und dem Online-Test von Filtern im allgemeinen, insbesondere aber auch hier beim Test der Extended Kalman-Filter, haben wir die folgende grundsätzliche Vorgehensweise: Zunächst werden Testfunktionen ausgewählt bzw. Konturverläufe konstruiert. Diese werden im Basisband winkelmuliert und in der Simulation als Beobachtungsgrößen für das Kalman-Filter zur Verfügung gestellt. Bei der Online-Untersuchung der Filter werden die Quadraturkomponenten mit Hilfe einer Soundkarte abgespielt, mit einer AD-Wandlerkarte eingelesen und dem als Software im DSP integrierten Kalman-Filter zur Verfügung gestellt. Da Filterlaufzeit direkt mit der Abtastrate und somit der verarbeitbaren Bandbreite zusammenhängen, muß zuvor eine Kalibrierung und Festlegung der verwendeten Parameter erfolgen. Die Beobachtungsgrößen wurden zuvor mit einem definierten Störsignal (meist gaußverteilt) gemischt. Mit unterschiedlichen Störsignalen ermitteln wir das Verhalten des Filters. Beim Extended Kalman-Filter werden dazu sog. Sensitivitätskurven für das Driving Noise Q ermittelt. Daraus erfolgt eine Optimierung von Startwerten des Driving Noise und der Kovarianzen. Die gefundenen Kovarianzwerte werden durch Monte Carlo Simulation, d.h. durch mehrfache Wiederholung der Simulation gefestigt. Dazu bilden wir den Ensemblemittelwert für den mittleren quadratischen Fehler, bezogen auf die Störleistung über mehrere Simulationen (Monte Carlo). Eine Auswertung der Filterfehlervarianzen und der Stabilität des Filters kann dann durchgeführt werden. Schließlich erfolgt eine vergleichende und grafische Aufarbeitung der gesammelten Meßergebnisse der Filtermodelle in Abhängigkeit von den Parametern S/N-Verhältnis, Modulationsindex und Bandbreite.

4.1 TESTFUNKTIONEN UND TESTSIGNALE

Zum Test der unterschiedlichen Kalman-Filtermodelle werden unterschiedliche Testsignale verwendet. Ein möglichst allgemeines und beliebiges Signal stellt der Signalverlauf eines

Gauß-Markov-Prozesses dar. Ferner wurde ein Sinussignal als Testsignal eingesetzt, das eine stetig sich verändernde Funktion darstellt und somit zu den einfachsten zu bearbeitenden Signalen zählt. Für die Bestimmung der maximal bearbeitbaren Bandbreite, der Laufzeit des Filters und somit der Abtastrate ist die Sinusfunktion ebenfalls hilfreich. Als weitere Signalgrundform wurde ein Rechtecksignal verwendet, das besonders extreme Werteänderungen aufweist und beispielsweise einem Digitalsignal entspricht. Als letztes Testsignal oder Truthsignal wurde ein Pseudo-Sprachsignal verwendet. Im folgenden soll nur die Erzeugung des Pseudo-Sprachsignals erläutert werden.

4.1.1 Pseudo-Sprachsignal

Das Nutzsinal ist im Bereich der Telekommunikation meistens ein Sprachsignal. Für die Simulation der Demodulation winkelmulierter Signale mit Kalman-Filtern wird daher ein Modell der Sprache verwendet, das die charakteristischen Eigenschaften eines Sprachsignals enthält, Pseudo-Sprachsignal genannt. Bei dieser Untersuchung wurde zunächst ein sehr einfaches Modell gewählt, um nur die für das Filter relevanten Signaleigenschaften nachzubilden. Zur Nachbildung der menschlichen Sprechorgane werden sogenannte Vocoder eingesetzt, was jedoch hier nicht erforderlich ist /37, 59, 66/.

Die direkte Modulation des analogen Sprachsignals wird heutzutage durch digitale Modulationsarten überwiegend abgelöst. Nicht das digitalisierte Zeitsignal, sondern die Fourierkoeffizienten eines kurzen Zeitblocks der Sprache werden übertragen. Diese Zeitblöcke werden dazu in 5 bis 20 Filterkoeffizienten eingeteilt. Zur Datenaufnahme werden meist Filterbänke eingesetzt, um die Energie des Sprachsignals in den einzelnen Frequenzbändern zu messen. Dabei werden bis über 35dB SNR angeboten, vielmehr als für exzellente Qualität erforderlich ist.

4.1.1.1 Sprachmodell

Zunächst wird eine akustisch-phonetische Analyse die Sprache einteilen: In der deutschen Sprache gibt es ca. 20 konsonantische Phoneme und 20 kurze/lange Vokale & Diphthonge. Phoneme sind kleinste Bausteine, aus denen die Sprache aufgebaut ist (Sprachlaute).

Durch Koartikulation werden die Phoneme beim Sprechvorgang miteinander verschliffen. Alle 100 ms ändern sich die Spracheigenschaften, so daß Zeitrahmen, die in der Größenordnung von 10-20 ms gewählt werden, zu keinem Informationsverlust führen.

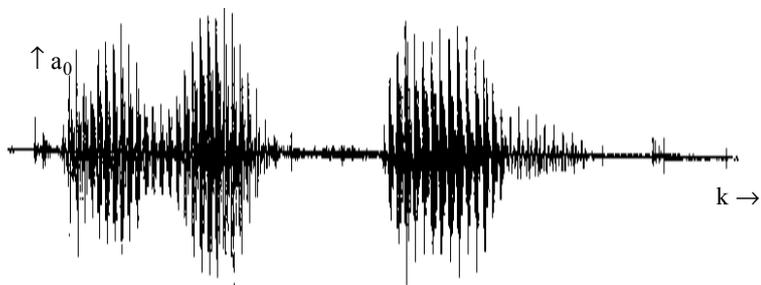


Abb. 4.1: Reale Sprachsignale

Bei der Erzeugung menschlicher Sprache wird der Erzeugungsvorgang in mehrere Bereiche unterteilt. Es folgt ein elektrisches Ersatzschaltbild von Sprache, mit dem eine künstliche Erzeugung eines sprachähnlichen Signals möglich wird. In unserem Fall ist es nicht notwendig bestimmte Buchstaben oder Wörter nachzubilden, sondern es genügt ein Signal mit sprachtypischen Eigenschaften:

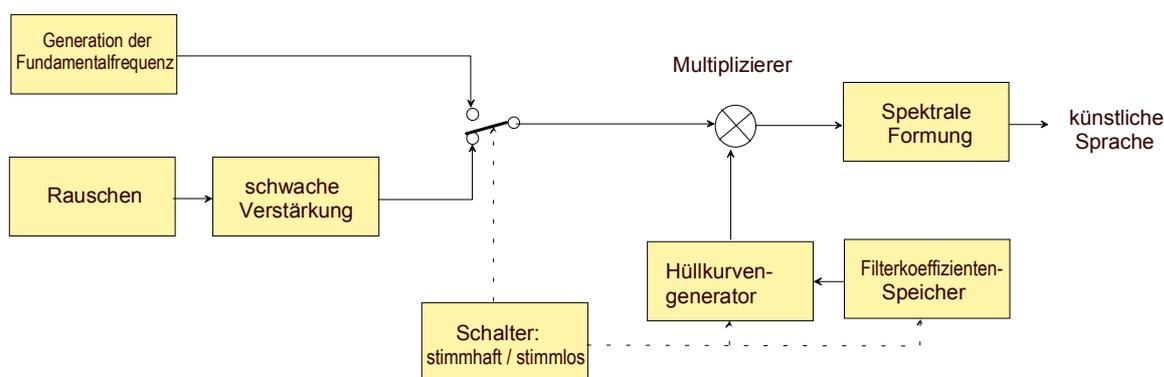


Abb. 4.2: schematisches Modell zur Spracherzeugung

Amplitude und Spektrum sind die beiden charakteristischen Elemente, aus denen sich die Sprache aufbaut. Die Amplitude soll sich kontinuierlich und zufällig ändern, was mit einem Markov-Modell mit oberer und unterer Begrenzung der Werte dargestellt werden kann.

Die Änderung der fundamentalen Frequenz soll sich ähnlich verhalten, so daß dies ebenfalls mit einem Markov-Prozeß mit festem Frequenzbereich definiert werden kann. Der Wechsel von Pulsgenerator (für die fundamentale Frequenz zuständig) und Rauschgenerator (farbiges Rauschen für die Konsonantendarstellung) darf nicht abrupt, sondern nur allmählich geschehen (z.B. Einsatz eines Hamming-Fensters). Der Frequenzbereich der fundamentalen Fre-

quenz, der zur Vokalerzeugung dient, liegt bei ca. 60-150 Hz. (Bei Männern ca. 60-100 Hz und bei Frauen ca. 100-150 Hz). Der Frequenzbereich des farbigen Rauschens liegt etwa im Bereich von 500 Hz-4 kHz /20, 29/. Baugruppen im Telekommunikationsbereich besitzen deshalb eine Bandbreite von 50 Hz - 10kHz.

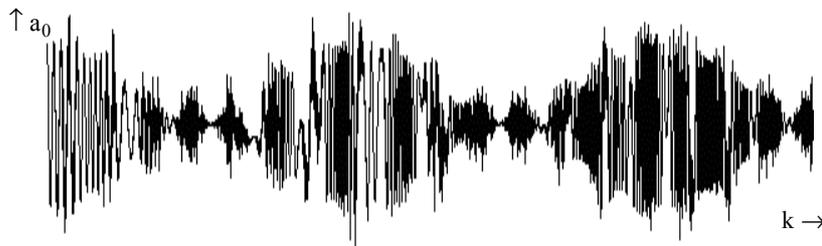


Abb. 4.3: Pseudosprachsignal

4.1.2 Rauschprozesse

Die Definition eines weißen, gaußverteilten Rauschens wurde bereits im Abschnitt 2.2 gegeben. Um jedoch einen derartigen, für das Kalman-Filter notwendigen Rauschprozeß zu erzeugen, gibt es unterschiedliche Lösungsansätze /47, 51, 56, 76/.

4.1.2.1 Diskrete Rauschgeneratoren

Aufgrund der digitalen Verarbeitung wird ein diskretes Rauschen benötigt und dazu die jeweiligen diskreten Rauschgeneratoren eingesetzt. Die Generation eines zufälligen Signals mit einem Random-Generator ist dabei jedoch nicht trivial und soll näher betrachtet werden.

4.1.2.1.1 Definition unter der Programmiersprache 'C'

Die eingesetzten C-Compiler von BorlandC und Texas Instruments stellen beide eine Pseudozufalls-Variable bereit, die aus dem Zahlenvorrat von 0 bis $2^{31} - 2 = 2147483646$ (beim TMS320C40) einen scheinbar zufälligen Wert liefert. Mit der Funktion *srand(seed)* wird dabei ein Zufallsvektor erzeugt, und mit der Funktion *rand()* werden hintereinander Werte dieses Zufallsvektors abgerufen. Nach Auslesen des gesamten Zahlenbereichs (bei 32 Bit eben $2^{31} - 1$) beginnt die Funktion wieder beim ersten Wert.

Dabei existiert jedoch ein vordefinierter Zufallszahlengenerator, bei dem nach einem vordefinierten Muster die Zahlen erzeugt werden, so daß eine Pseudozufallsfolge entsteht. Mit

dem Wert von *seed* wird die zur Generation verwendete Basiszahl ausgewählt. Durch Verknüpfung mit einem Timer läßt sich die zu erwartende Basiszahl zufällig wählen, was die am Ausgang anliegende Zufallszahl noch zufälliger werden läßt. Bei einer derartigen Generation eines Rauschprozesses läßt sich das Rauschen reproduzieren, es ist nicht völlig zufällig! Trotzdem kann es für die angestellten Beobachtungen und Betrachtungen gut eingesetzt werden. Es ist sogar von Vorteil, die Möglichkeit eines reproduzierbaren Rauschsignals zu besitzen, um bestimmte Filterprobleme zu beobachten.

Zunächst liefert die *rand()*-Funktion jedoch gleichverteilte Zufallszahlen, die durch Summenbildung mehrerer gleichverteilter Pseudozufallszahlenfolgen zu einer gaußverteilten Zufallszahlenfolge zusammengefaßt werden müssen. Dies ist möglich, da durch additive Überlagerung beliebig verteilter Signale sich am Ausgang eine Gaußverteilung herausbildet. Die Praxis zeigt, daß wir bereits nach 5 Überlagerungen von Gleichverteilungen eine Gaußverteilung mit einer Genauigkeit von $\sim 2\%$ erhalten. Bei einer betrachteten Länge der Rauschmatrix von 2×1000 Werten ist diese Genauigkeit völlig ausreichend.

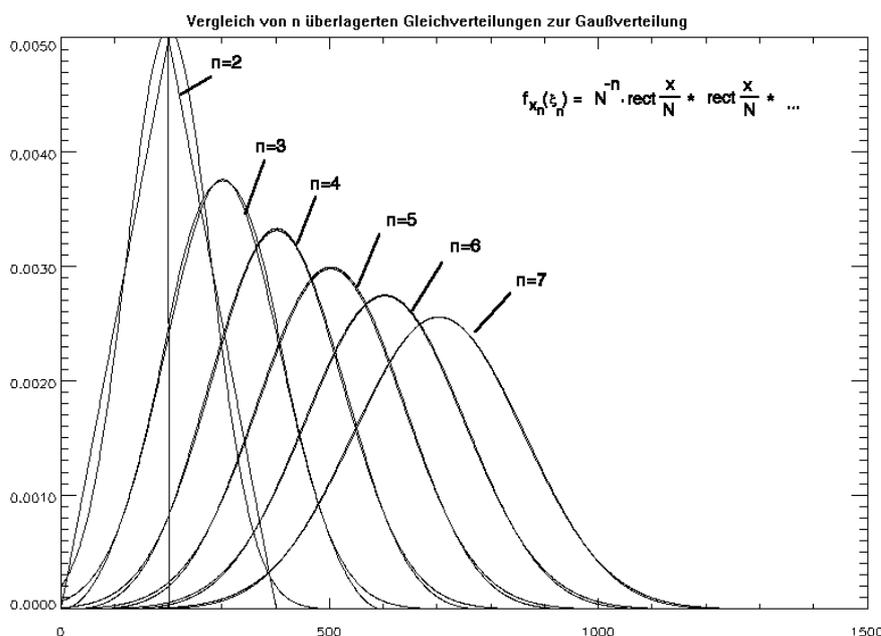


Abb. 4.4:

**Verteilungsfunktion
entstanden durch
Überlagerung
gleichverteilter Si-
gnale im Vergleich
zur Gaußverteilung**

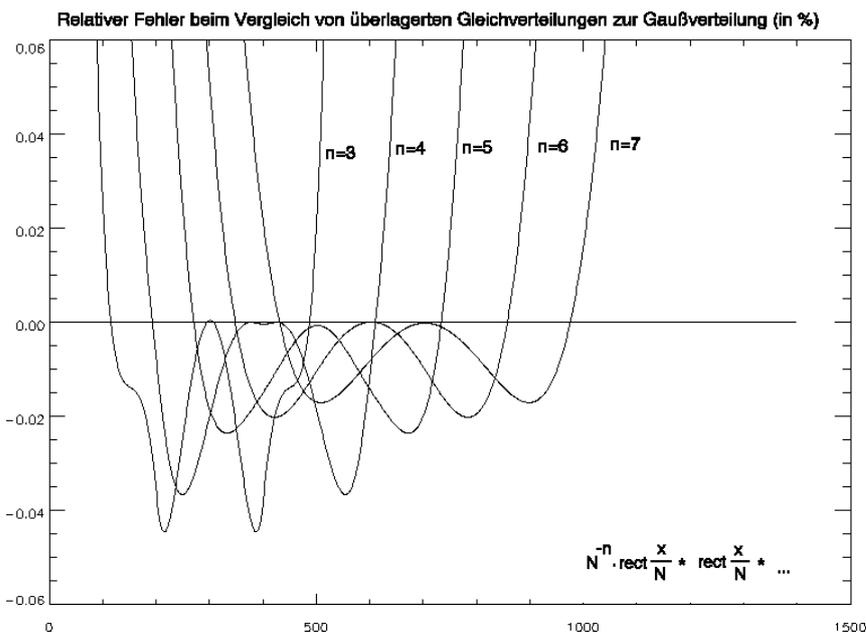


Abb. 4.5:

Rel. Fehler entstanden durch Überlagerung gleichverteilter Signale im Vergleich zur Gaußverteilung

Welche Leistung besitzt das erzeugte Signal? Um ein SNR-Verhältnis angeben zu können, müssen wir die Leistung des Rauschsignals bestimmen. Die Rauschleistung ist definiert über die Varianz des Rauschsignals. Unter der Annahme, daß die diskreten, normierten und somit reellen Werte als quasi zeitkontinuierlich betrachtet werden können, ergibt sich bei einer Gleichverteilung mit der Verteilungsfunktion:

$$f_x(\xi) = \frac{1}{\Delta} \text{rect} \frac{\xi}{\Delta}$$

mit dem Wertebereich: $\Delta = \frac{-1}{2N}$ bis $\frac{+1}{2N}$

(N=Maximalanzahl des Zufallszahlengenerators)

Die Varianz lautet:

$$\begin{aligned} \sigma_x^2 &= \int_{-\infty}^{+\infty} \xi^2 f_x(\xi) d\xi \\ &= \frac{1}{\Delta} \int_{-\infty}^{+\infty} \xi^2 \text{rect} \frac{\xi}{\Delta} d\xi \\ &= \frac{1}{\Delta} \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \xi^2 d\xi = \frac{1}{\Delta} \cdot \frac{1}{3} \xi^3 \Big|_{-\frac{\Delta}{2}}^{+\frac{\Delta}{2}} = \frac{1}{3\Delta} \cdot \left(\frac{\Delta^3}{8} + \frac{\Delta^3}{8} \right) = \frac{\Delta^2}{12} \end{aligned}$$

Durch additive Überlagerung mehrerer Zufallsvariablen falten sich die Verteilungsfunktionen, so daß bei zwei Gleichverteilungen eine Dreieckverteilung entsteht.

Die Leistungen addieren sich, da die beiden Signale unkorreliert sind und sich keine Kreuzkovarianzterme bilden. Die Varianz der Dreieckverteilung $\Lambda\left(\frac{\xi}{\Delta}\right)$ lautet:

$$\sigma_x^2 = \int_{-\infty}^{+\infty} \xi^2 f_x(\xi) d\xi = \frac{2}{12} \Delta^2 = \frac{\Delta^2}{6} \quad \text{mit: } f_x(\xi) = \frac{1}{\Delta^2} \Lambda\left(\frac{\xi}{\Delta}\right)$$

Betrachtung der Gesamtleistung in Abhängigkeit der unkorrelierten Einzelleistungen der Rauschprozesse:

$$\begin{aligned} y &= \sum_i x_i \\ \sigma_x^2 &= E\{y^2\} = E\left\{\sum_i [x_i]^2\right\} = \sum_i \sum_j E\{x_i \cdot x_j\} = \sum_i \sum_j \varphi_{x_i, x_j}(\tau = 0) \\ &= \sum_i \varphi_{x_i, x_i}(\tau = 0) = \sum_i E\{x_i^2\} = \sum_i \frac{1}{12} \Delta^2 \end{aligned}$$

4.1.2.1.2 Definition unter der Interpretersprache IDL

In IDL, da es auf 'C' basiert, existiert eine ähnliche Funktion *randomu()* für das gleichverteilte Rauschen. Eine Funktion für einen gaußverteilten Rauschprozeß wurde mit der Funktion *randomn()* bereits implementiert. Bei beiden Funktionen wird ein Startwert (*seed*) zur Initialisierung des Pseudozufallszahlengenerators gesetzt und gleichzeitig der erste Zufallswert ausgegeben.

4.1.2.1.3 Weitere Zufallsgeneratoren

Anstatt der Generation einer Gaußverteilung über unendlich viele überlagerte Gleichverteilungen zu erzeugen, gibt es weitere Möglichkeiten: Die eine nach Brandt und eine weitere nach Chaos-Regeln wie sie unter /13, 51/ beschrieben wird.

4.1.2.2 Rauschen der Demodulationsumgebung

Der 32 bit DSP liefert bereits nach Überlagerung von fünf gleichverteilten Pseudozufallsfolgen eine annähernd gaußverteilte Verteilungsfunktion. Pro Messung werden je 1000 Zu-

fallswerte für Real- und Imaginärteil benötigt. Danach wird ein neuer Startwert zur Initialisierung des Pseudozufallszahlengenerators gewählt.

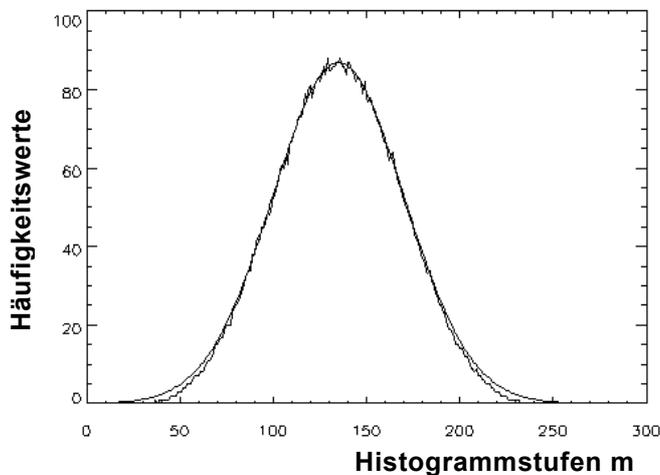


Abb. 4.6: Verteilungsdichtefunktion des Pseudoräuschgenerators für gaußverteilttes weißes Rauschen von $1.000 \times 1.000 = 1.000.000$ Zufallswerten (Aufnahme aus Meßsoftware-Statistikmodul), Vergleich mit Gaußglocke unter IDL

4.1.2.2.1 χ^2 -Anpassungstest

Zur Sicherheit wird ein χ^2 -Anpassungstest durchgeführt. Mit der Hypothese, daß wir eine Gaußverteilung betrachten, wird eine Gaußverteilung in die Kurve hineingefittet (Varianz und Mittelwert sind hier die bestimmenden Parameter). Anhand eines Signifikanzniveaus α und der Freiheitsgrade wird ein kritischer Wert χ_{α}^2 berechnet. Das zu bestimmende χ^2 muß kleiner als der kritische Wert χ_{α}^2 sein, da sonst die Annahme einer Gaußverteilung falsch ist.

Mit der folgenden Formel wird χ_{α}^2 berechnet (s. /13/):
$$\chi_{\alpha}^2 = \frac{1}{2} \left(\sqrt{2m-1} + z_{2\alpha} \right)^2$$

Bei einer Gaußverteilung gibt es einen Freiheitsgrad, so daß sich für die Berechnung von m folgendes ergibt: $m=k-r-1$. Dabei entspricht k der Anzahl der Intervalle im Histogramm und r entspricht Anzahl der Freiheitsgrade: $m=265-1-1=263$;

Ferner ist das Signifikanzniveau $\alpha=0.05$ gegeben. Daraus ergibt sich $z_{2\alpha}$ mit (Tab. 1.1.2.10 /13/) zu:

$$\begin{aligned} 2\phi_0(z_{2\alpha}) &= 1 - 2\alpha \\ \phi_0(z_{2\alpha}) &= \frac{1}{2} - \alpha = \frac{1}{2} - 0.05 = 0.45 \end{aligned} \tag{4-1}$$

$$\Rightarrow \phi_0^{-1}(0.45) = z_{2\alpha} = 1.7$$

Daraus folgt abschließend für χ_α^2 :

$$\Rightarrow \chi_\alpha^2 = \frac{1}{2} \left(\sqrt{2 \cdot 263 - 1} + 1.7 \right)^2 = 605,8$$

χ^2 berechnet sich mit der folgenden Berechnungsvorschrift und ergibt:

$$\chi^2 = \sum_{i=1}^k \frac{(m_i - np_i)^2}{np_i} = 29,68$$

(m_i = Histogrammwerte, n = Gesamtzahl der Messungen, p_i = angenommene Wahrscheinlichkeiten der Gaußverteilung)

(4-2)

Der Vergleich von χ^2 und χ_α^2 zeigt, daß χ^2 kleiner ist und somit die Gaußverteilungsdichte eine wahre Hypothese darstellt! $\chi^2 < \chi_\alpha^2$ q.e.d.

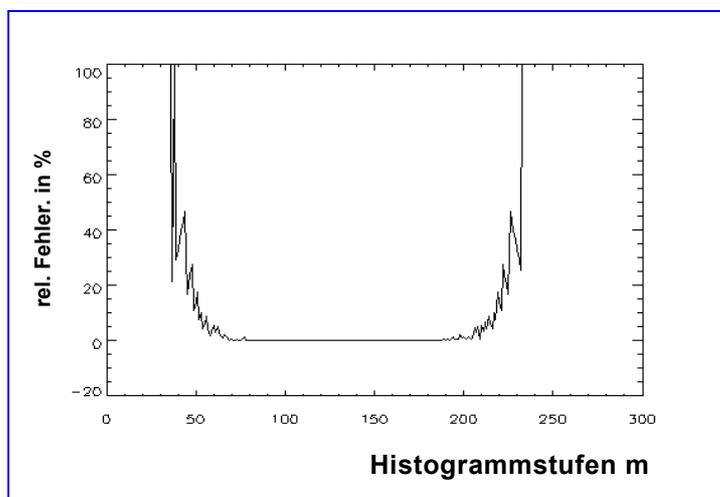


Abb. 4.7: Betrachtung von rel. Fehlern zur Berechnung von χ^2

4.2 MONTE-CARLO-SIMULATION UNTER IDL UND BORLAND-C

Die Durchführung einer Monte-Carlo-Simulation ist die Standard-Methode zur Bestimmung von gesicherten Schätzwertergebnissen. Die zu schätzende Größe muß dabei mit stochastischen Störgrößen behaftet sein. Es folgt zum Test eines Filters eine Vielzahl von Durchläufen/Wiederholungen einer Meßanordnung vollzogen. Diese Arbeitsweise liefert nur bei konstanten, d.h. hier zeitlich unabhängigen Meßbedingungen gültige Aussagen. Bei der

Online-Filterung ist dies jedoch unerheblich, da nur der eingeschwungene Zustand betrachtet werden soll. Wichtig ist jedoch, daß bei der Simulation ausreichend lange Durchläufe verwendet werden. Die Meßergebnisse eines Durchlaufs können wir auch als Musterfunktion $y(\cdot, \omega)$ eines Zufallsprozesses $y(\cdot, \cdot)$ betrachten. (Die Musterfunktionen des Prozesses sind in diesem Fall nicht zeitgleich entstanden, sondern sind zeitlich betrachtet hintereinander aufgetreten. Da jedoch eine zeitliche Unabhängigkeit der Musterfunktionen vorausgesetzt wurde, dürfen wir die Gesetzmäßigkeiten eines Zufallsprozesses anwenden.)

Wieviele Simulationsdurchläufe einer Monte-Carlo-Simulation werden benötigt, um von zuverlässigen Meßergebnissen sprechen zu können? Zunächst werden einige Durchläufe benötigt, um das erste Ergebnis zu bestätigen (bis zu 5000). Ein weiteres Kriterium ist jedoch dann das menschliche Auge bzw. die Bildgröße/-auflösung der Darstellung der Meßkurve. Sie bestimmt in letzter Instanz wie genau das Ergebnis sein muß /64/.

Es besteht grundsätzlich die Möglichkeit, mit einer konstanten Anzahl von Simulationsdurchläufen zu arbeiten, um die Filter auszuwerten oder mit einer angepaßten Anzahl, die zu einer konstanten Welligkeit der betrachteten Meßreihe führt. Im ersten Fall ist das unterschiedliche Filterverhalten direkt sichtbar, im zweiten Fall besteht dagegen diese Möglichkeit nicht. Jedoch lassen sich die verarbeitete Information und die Endzustände besser bestimmen. Ferner ist die Anzahl der benötigten Durchläufe ebenfalls ein Kriterium für das Filterverhalten.

4.3 FILTERSIMULATION

Beim C-Programm wurde die Simulation nahe an der Online-Realisierung programmiert, um eine schnellere Umsetzung auf den DSP zu erreichen. Das Programm erzeugt sich Markov-Prozesse selber, kann aber auch gespeicherte Datenfiles als Musterfunktionen einlesen. Startkovarianzen können geändert werden; Fehlerkovarianzmatrizen, Prädiktionswerte, Übergangsmatrizen und weitere wichtige Parameter können beim Filterdurchlauf direkt betrachtet werden. Die graphischen Tools von BorlandC werden für eine anschließende einfache Auswertung des Filterdurchlaufs benutzt.

4.3.1 Simulation unter BorlandC

In der Simulation wird kein Träger aufmoduliert, so daß die beiden Quadraturkomponenten direkt erzeugt und mit additivem Rauschen gestört werden. Bei der Verteilungsdichte ist trotz des nichtlinearen Zusammenhangs eine gaußförmige Verteilungsdichte der beiden Komponenten zu beobachten.

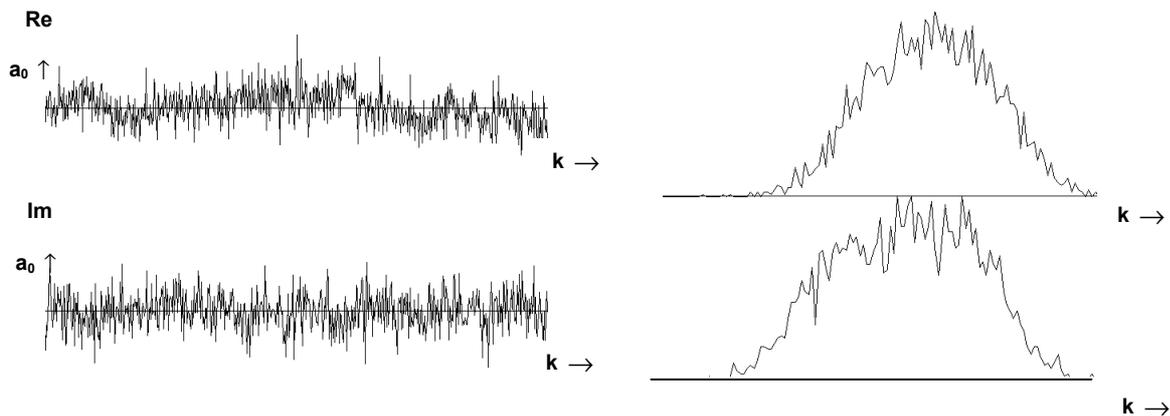


Abb. 4.8: Quadraturkomponenten und deren Verteilungsdichten (Histogramm) des gestörten Empfangssignals

Herkömmliche Demodulatoren werden bei einem SNR von +15 dB betrieben. Die Auswertungen der Online-Demodulation werden den tatsächlichen Gewinn bei Verwendung eines Kalman-Filters aufzeigen. Bei einem SNR-Verhältnis von ca. +4.5 dB ergibt sich beispielsweise:

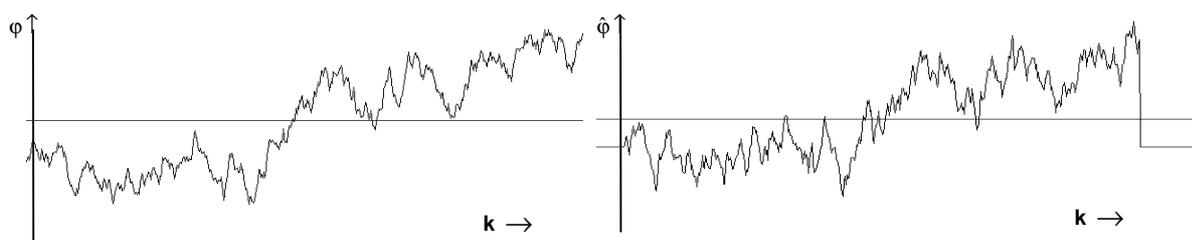


Abb. 4.9: Originalphase (links) und Schätzphase (rechts) bei SNR von ca. +4.5 dB (Quellsignal: Markov1-Signal)

Probleme bei der Online-Demodulation sind die üblichen Demodulationsprobleme wie beispielsweise Kohärenz der Trägerfrequenz, überlagerte Amplitudenmodulationen oder Nichtlinearitäten beim Phasenschieber. Die gezeigte Meßreihe stellt einen Ausschnitt des bereits gezeigten Pseudo-Sprachsignals dar:

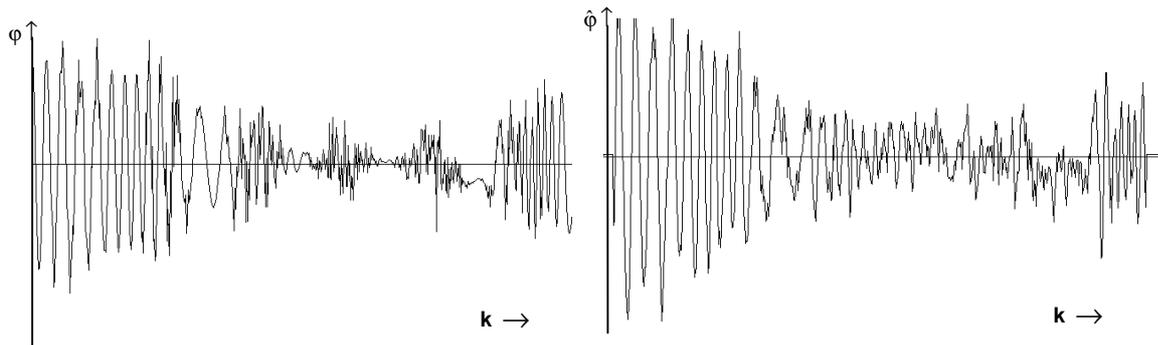


Abb. 4.10: Originalphase (links) und Schätzphase (rechts) bei SNR von ca. +4.5 dB (Quellsignal: Pseudo-Sprachsignal)

Problematisch sind die stark wechselnden Phasenänderungen. Eine adaptive Modellierung bzw. Nachführung des Systemrauschens bietet sich hier als Lösung an. Die Filterergebnisse für die Estimation der Amplitude, die nebenbei anfällt, soll hier nicht weiter betrachtet werden. Sie ist jedoch deutlich schlechter als die Phasenestimation, da aber auch das Systemmodell nicht auf die Amplitudenestimation ausgelegt wurde.

4.3.2 Monte-Carlo-Simulation unter IDL

Mit einer einzigen, zufälligen Musterfunktion läßt sich kein allgemeines Verhalten eines Kalman-Filters beschreiben. Die Qualität der Schätzergebnisse läßt sich durch eine Monte Carlo-Simulation bestimmen. Konvergenzverhalten und Einschwingverhalten können bei der Betrachtung der relativen Varianz (prozentualer Fehler) beobachtet werden.

Als Musterfunktionen wurden hier nur 100 verschiedene Markov-Prozesse eingesetzt, die mit einem Rauschen gestört wurden, so daß ein SNR-Verlauf von -3 dB bis +9 dB beschrieben wurde. (Bei dieser Simulation wurde bei zunehmender Störung des Signals die Fehlervarianz gemessen. Das Filter befand sich somit bei hohem Rauschanteil bereits im eingeschwungenen Zustand, so daß auch bei negativen SNR's Meßergebnisse erlangt werden konnten.)

Eine weitere Meßreihe wurde mit einem Pseudo-Sprachsignal durchgeführt. Das Phasensignal ist für alle 100 Messungen konstant geblieben, nur Amplitude und Rauschen sind bei jeder Musterfunktion neu gewählt worden.

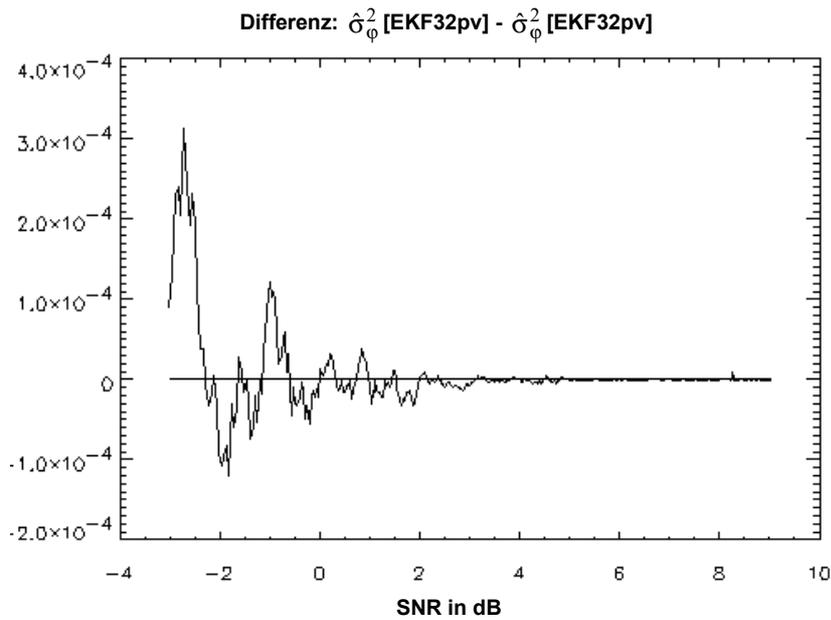


Abb. 4.11:
Differenz von Varianzen
eines Filters mit 2 und
mit 3 Beobachtungs-
variablen bei
Verwendung eines
Markov-Signals

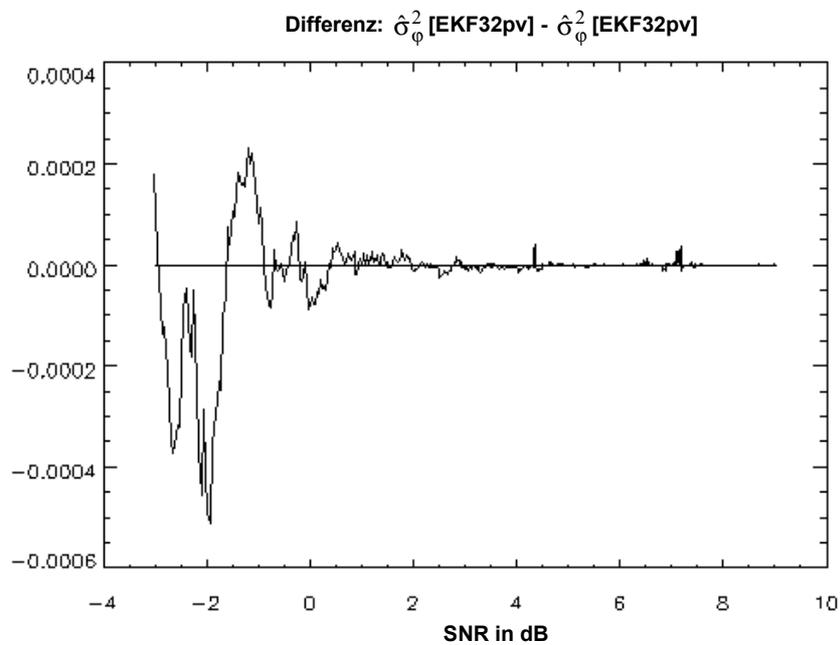


Abb. 4.12:
Differenz von
Varianzen eines Filters
mit 2 und mit 3
Beobachtungs-
variablen bei
Verwendung eines
Pseudo-Sprachsignals

Deutliche Varianzunterschiede sind erst bei sehr kleinen SNR's zu beobachten, so daß der Einsatz von erweiterten Filtermodellen erst in diesem Bereich zu Verbesserungen führt.

4.4 AUSWERTUNG DER ONLINE-FILTERMODELLE

Bei der Vielzahl der beschriebenen Filtermodelle, die auf ihre Online-Eignung hin untersucht wurden, läßt sich nicht jedes Filtersignal mit jedem Filter auswerten. Es wurde ein Auswahl an Vergleichen getroffen, die gemachte Beobachtungen über „typische“ Filterverhaltensweisen repräsentativ darstellen. Die wichtigsten Kriterien der Filterqualität sollen in den folgenden Abschnitten betrachtet werden. Grundsätzlich gilt jedoch, daß abhängig vom SNR unterschiedliche Filter am besten geeignet sind. Weiter ist zu beobachten, daß insbesondere die Wahl der Startwerte einen entscheidenden Einfluß nicht nur auf das Einschwingen, sondern auch auf das stationäre Verhalten der Filter haben.

Wählt man drei wesentliche Filtermerkmale (insbesondere wichtig für die Online-Anwendbarkeit): Schätzfehlergenauigkeit, Bandbreite und Stabilität aus, so lassen sich nach Auswertung der unterschiedlichen Filtermodelle folgende Beobachtungen feststellen:

- Höhere Stabilität wird auf Kosten einer geringeren Genauigkeit bei nur leicht steigender Filterlaufzeit erreicht (z.B. RO-Smoother). Der Einsatz bei negativen SNR's ist hier sinnvoll.
- Höhere Genauigkeit wird durch höhere Laufzeit aber auch wachsender Instabilität erreicht (z.B. Erweiterte Filtermodelle) erreicht. Eine höhere Genauigkeit und höhere Stabilität wird auf Kosten einer stark ansteigenden Filterlaufzeit erreicht (z.B. Smoother).
- Stabilitätsfördernde und/oder genauigkeitsfördernde Maßnahmen führen immer zu einer Verlängerung der Filterlaufzeit. Reduziert man die Filterlaufzeit durch Vorfiltern oder temporales Auffrischen der Filterparameter, so erhält man nur bei Signalen mit geringer Dynamik Verbesserungen in Laufzeit und Genauigkeit.
- Eine Vergrößerung der Kalman-Filterelemente führt zu einer höheren Laufzeit, aber auch (mit Ausnahme der Smootherkonzepte) zu einem instabileren Verhalten der Filter.

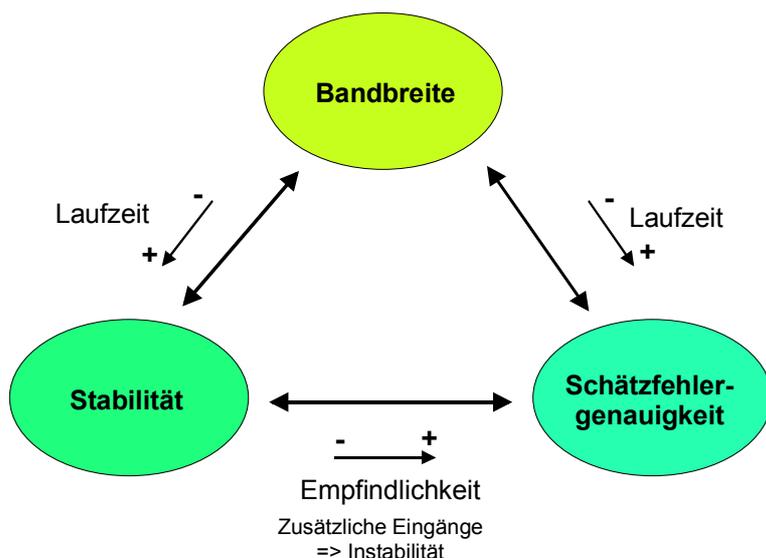


Abb. 4.13: Filtermerkmale

Grundsätzlich läßt sich die Auswertung in stationäres und dynamisches Filterverhalten aufteilen. Bevor eine Messung als gültig akzeptiert wird, muß sich von der Optimalität der Parametereinstellungen überzeugt werden. Erst nach erfolgtem Filtertuning kann der Meßwert bei den folgenden Auswertungen Verwendung finden.

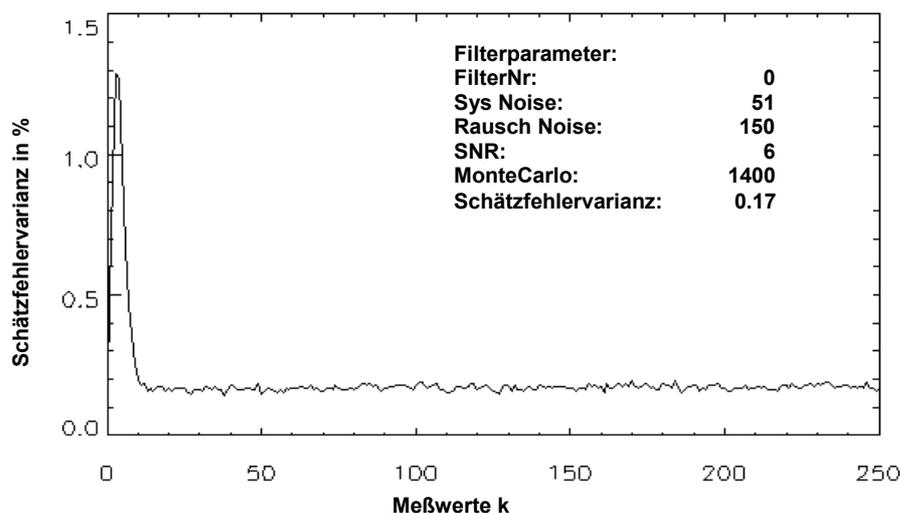


Abb. 4.14: Messung eines Durchlaufs

Bei der Beobachtung der prozentualen Schätzfehlervarianz werden 100 Meßwerte nach Einschwingen betrachtet. Dieses ist nach ca. 50-100 Messungen der Fall, so daß nach 150 Meßwerten mit der Mittelwertbildung begonnen werden kann.

Die nachfolgenden Messungen beziehen sich zunächst auf eine Auswertung von Messungen, bei denen als Quellsignal ein langsam schwingendes Signal verwendet wurde und das Systemmodell einen Markov-Prozeß beschreibt. Dieser kann, da sich beliebige Signale

nachbilden lassen, generell eingesetzt werden. Eine Modellbildung nach dem Singermodell (EKF22ts) führte zu gleichen Meßergebnissen wie beim Standardfilter EKF22.

4.4.1 Filtertuning

Um den optimalen Parametersatz bei der Messung zu verwenden, ist es notwendig, diese durch eine Vielzahl von Testmessungen herauszufinden: Bei einem konstant gewählten SNR (es wurde mit dem kleinsten betrachteten SNR von +16.5 dB begonnen) werden Varianzen für das Beobachtungsrauschen (Variable: Rausch) und das Systemrauschen (Variable: Sysnoise) angesetzt und die sich ergebende prozentuale Filterfehlerschätzkovarianz beobachtet. Durch Variation eines der beiden Parameter ergibt sich bei zunehmender Rauschleistung eine zunehmende Verbesserung des Schätzfehlers und bei weiterem Anstieg schließlich wieder eine langsame Verschlechterung desselben.

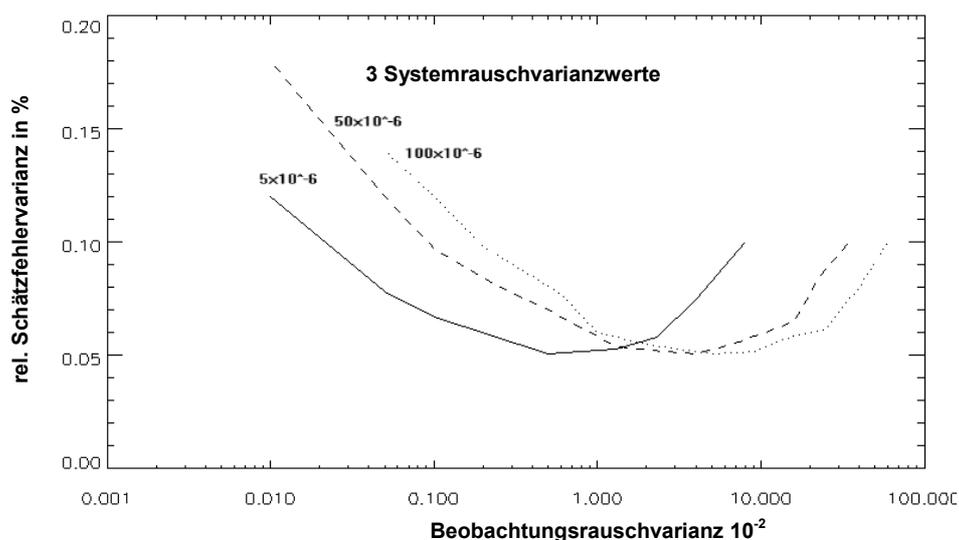


Abb. 4.15: Variation der Rauschvarianz

Bei der Variation der Rauschvarianz und der Wahl dreier unterschiedlicher Systemvarianzen ergeben sich die obigen drei Kurven. Das Minimum aller drei Kurven ist gleich und liegt bei etwa 0.05% beim hier gewählten SNR von +12.5 dB. Jeder Punkt dieser drei Kurvenverläufe besteht aus einer kompletten Messung. Drei Punkte aus der obigen Abbildung wurden im folgenden dargestellt. Die unterste Kurve (Rausch = 0.5) zeigt eine optimale Einstellung bezogen auf den Schätzfehler, die beiden anderen jedoch nicht.

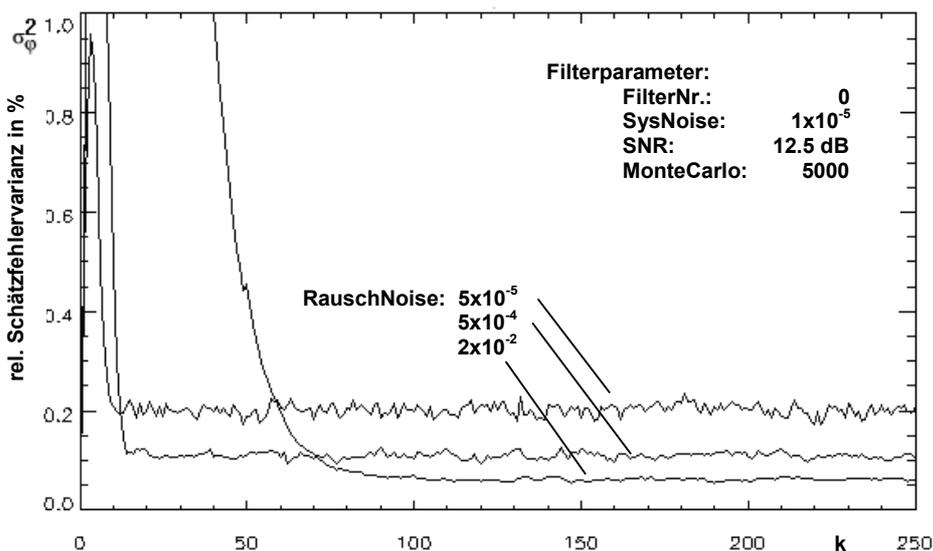


Abb. 4.16: prozentuale Schätzfehlervarianz 3-er Parametereinstellungen

Dies bedeutet, daß bei der Variation der Rauschvarianz bei ungefährender Vorauswahl der Systemvarianz ein korrekter, optimaler Parametersatz gefunden werden kann. Trägt man die gefundenen Parametersätze übereinander auf, so erhält man für jedes Filter ein ungefähr konstantes Verhältnis.

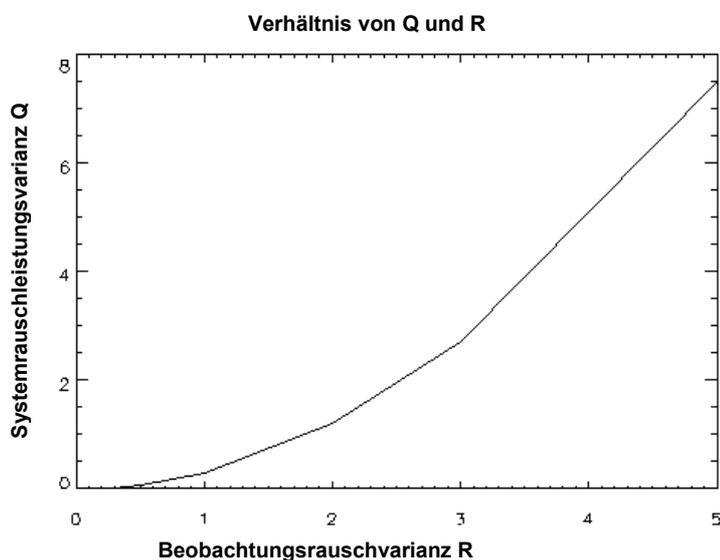


Abb. 4.17: Abhängigkeit von Systemrauschvarianz und Beobachtungsrauschvarianz

Umgekehrt läßt sich auch die Variation der Systemvarianz bei konstanter Rauschvarianz durchführen. Dies bedeutet, daß bei der Suche nach dem optimalen Parametersatz die Variation eines Parameters genügt. Zusätzlich besteht eine Abhängigkeit zum gewählten

SNR, so daß der optimale Parametersatz meist nach wenigen Durchläufen gefunden werden kann.

4.4.2 Stationäres Verhalten

Beim stationären Verhalten wird der eingeschwungene Zustand eines Filters betrachtet. In den meisten Fällen werden Filter im stationären Betrieb eingesetzt. Jedoch ist bei fehlender Stabilität des Filters dessen erneutes Starten notwendig. In diesem Fall ist ein schnelles, kurzes Einschwingen des Filters auf den optimalen stationären Betrieb erwünscht. Jedoch auch im stationären Betrieb besteht eine Abhängigkeit von den Startwerten des Filters.

4.4.2.1 Startwerte

Dazu werden am Beispiel des Standardfilters EKF22s die stationären Filterergebnisse bei unterschiedlichen Startwerten für $\hat{x}_2^+(0)$ (Phasenänderung) und $P_{22}^+(0)$ (Kovarianz der Phasenänderung) begutachtet:

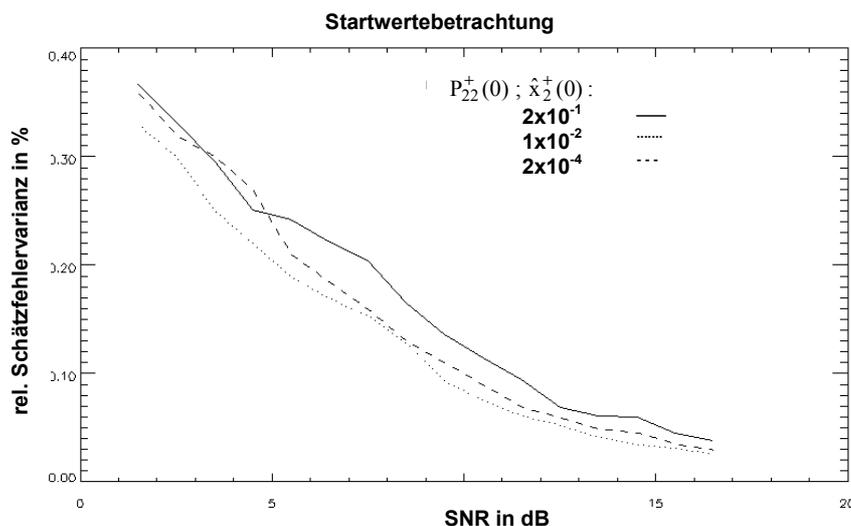


Abb. 4.18: Variation der Startwerte

Bei sehr kleinen SNR's ist die Wahl der Parameter recht schwierig, was die Überlappung der zwei Kurven erklären könnte. Bei der Wahl dieser Filterdurchläufe wurde von einem konstanten Startwert der Phase von $-\frac{3\pi}{4} \approx -2.4$ ausgegangen. Da die Phase bei Online-Filterung jedoch veränderlich ist, läßt sich auf diese Weise nur eine grobe, jedoch stabile Vorgabe des Startwertes vornehmen. Bei dieser Wahl der Startphase und bei klein gewählten

$\hat{x}_2^+(0)$ und $P_{22}^+(0)$ läßt sich das Filter soweit stabilisieren, daß keinerlei Ausreißer mehr auftreten.

Bei kleinen SNR's läßt sich ein genauere Startwert durch die Berechnung der Phase mit Hilfe der Arcustangens-Funktion der gestörten Quadraturkomponenten erreichen. Dies führt jedoch bei kleinen SNR's dazu, daß zufällig extrem schlechte Startwerte entstehen, die zu keinem Einschwingen in der korrekten Phasenlage führen. Diese werden als unerwünschte Ausreißer gezählt. (Jeder Ausreißer wird vom Programm nicht mit in die Anzahl der MC-Simulationen mitgezählt, so daß jeder Ausreißer eine zusätzliche Simulation zur Folge hat. Das Filtertuning wird nur bis zu einer Grenze von 50% Ausreißeranteil durchgeführt, da eine hohes Ausreißerverhalten als negativ für das verwendete Filtermodell gewertet werden kann.)

Auch im stationären Betrieb ist es folglich notwendig, die Filterstartwerte korrekt zu wählen, um einen stabilen Betrieb mit geringer Fehlervarianz zu erreichen. Für jedes Filter ergeben sich andere Startwerte, die empirisch ermittelt werden müssen.

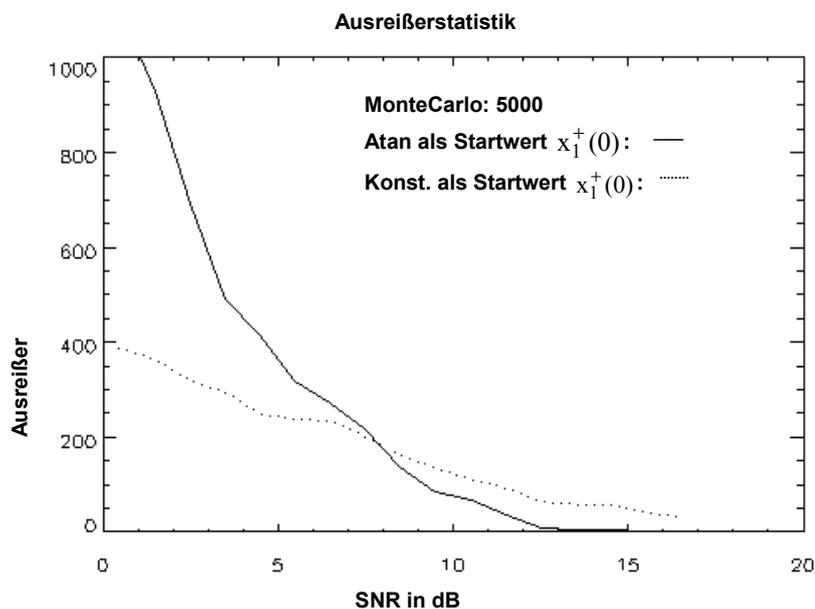


Abb. 4.19: Anzahl der Ausreißer bei 5000 Monte-Carlo-Simulationen

4.4.2.2 Schätzfehlervarianz

Das entscheidende Kriterium für die Filterqualität ist die Filterfehlervarianz. Nachdem durch Filtertuning die optimalen Parametereinstellungen für jedes Filter gewählt wurden, ergibt sich für die unterschiedlichen Filtermodelle folgendes Verhalten. In der folgenden Abbildung

werden die Messungen zu einer Kurve über das SNR aufgetragen. Die Filterfehlervarianzen aller Filter zeigen einen gleichen Kurvenverlauf, deshalb wurden nicht alle Filter mit in die Grafik eingebunden, sondern nur eine Auswahl getroffen.

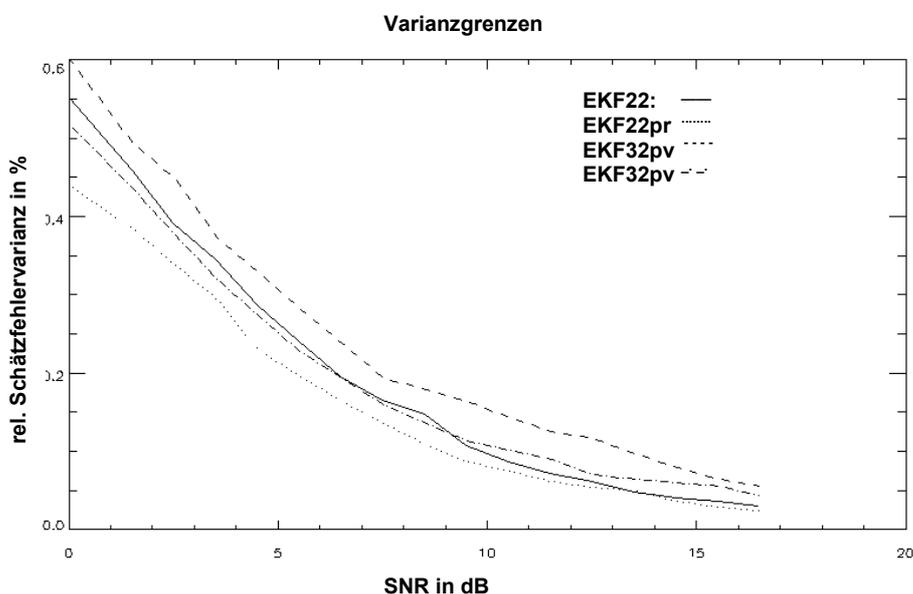


Abb. 4.20: Relative Schätzfehlervarianz unterschiedlicher Filtermodelle

Die Schätzfehlervarianz des Kalman-Filters mit verwendeter Vorfilterung zeigt die besten Filterergebnisse. Es folgt die Gruppe der adaptiven Filter (hier durch AKF22ml) repräsentiert. Die Filter mit additiven virtuellen Beobachtungen ergeben jedoch im Gegensatz zur Theorie schlechtere Varianzwerte. Die eingesetzten Kreuzkovarianzterme und deren Näherungen wurden daraufhin untersucht. Durch Vergrößerung der Kovarianzen wurde die Gewichtung der zusätzlichen Beobachtungen reduziert und eine Verbesserung von $\hat{\sigma}_{\phi}^2$ bis auf Fehlervarianzwerte des Standardfilters beobachtet. Das empirische Verändern der Werte führte zu keiner Verbesserung der dargestellten Meßergebnisse. (Die relativ mäßigen Schätzfehlervarianzen der erweiterten Filtermodelle lassen sich auch auf die gewählte Grenze von 50 % für die Maximalzahl der erlaubten Ausreißer zurückführen.)

Von den Smootherfilter liegen nur Simulationsergebnisse vor, da eine Online-Filterung zu zeitaufwendig ist. Dies wird in der Laufzeitbetrachtung nochmals deutlich gemacht.

Exemplarisch wurden hier für +3.5 dB die unterschiedlichen Filtermodelle in einer anderen Darstellung entzerrt und übersichtlich dargestellt. Dieser SNR-Wert wurde gewählt, da insbesondere bei sehr stark gestörten Signalen die unterschiedlichen Charakteristika der Filtermodelle stärker zum Wirken kommen. Den vorhandenen Informationsgewinn, der durch zu-

sätzliche virtuelle Meßwerte geschaffen wird, schlägt sich insbesondere bei stark gestörten Signalen nieder. Die Rauschleistung ist in diesem Fall bereits fast so groß wie die Signalleistungen der Quadraturkomponenten (s. SNR_{Re} und SNR_{Im} in 2.1.4).

In Farb-/Graustufen unterschieden sind hier die Gruppen, zu denen die unterschiedlichen Filtermodelle zugeordnet sind. Die Filtermodelle mit zusätzlichen virtuellen Beobachtungen zeigen jedoch - anders als erwartet - keine besseren Schätzergebnisse. Deutlich besser verhalten sich dagegen die adaptiven Filtermodelle, bei denen die Systemrauschleistung durch Maximum-Likelihood Schätzung an die aktuellen Verhältnisse angepaßt wird.

Interessanterweise verhält sich das vorgefilterte Kalman-Filter (Prefiltering; EKF22pr) besonders gut. Die in der Einschwingphase gefundenen Parameter werden hier konstant gehalten und die Estimation ohne Kovarianzenzyklus fortgeführt. Weshalb zeigt dieses Filter so gute Eigenschaften, da es sich zudem auch durch eine vorteilhafte schnelle Laufzeit auszeichnet? Nur im Fall von Quellsignalen, die ein konstantes, dynamisches Verhalten aufweisen, läßt sich das Vorfiltern gut einsetzen. Ansonsten führt die Verwendung dieses Filters bei starken dynamischen Änderungen zu Ausreißern.

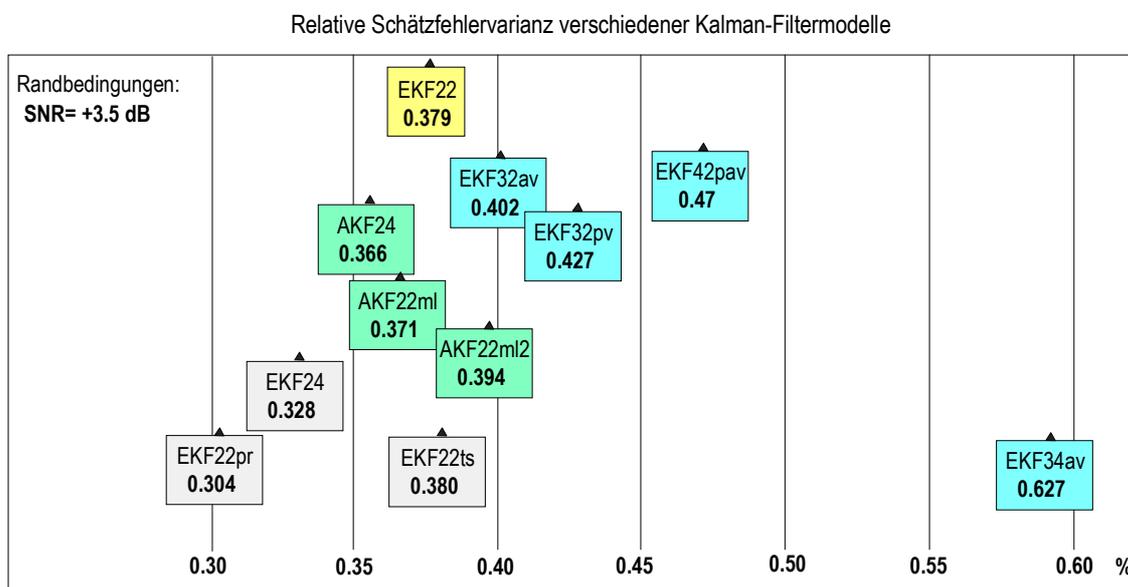


Abb. 4.21: Relative Schätzfehlervarianz verschiedener Kalman-Filtermodelle bei +3.5 dB

Das zusätzlich bessere Filterergebnis läßt sich vermutlich mit der Unabhängigkeit von einer nicht weißen Residuensequenz deuten, die das Kalman-Gain, den wichtigsten Steuerparameter des Filters, ungünstig beeinflussen kann.

Bei +10 dB wurde eine weitere Detailabbildung angefertigt, die zeigt, daß sich grundsätzlich an dem Verhältnis der Filter untereinander kaum etwas verändert hat. Es kann also von einem parallelen Verhalten bei anderen SNR's ausgegangen werden.

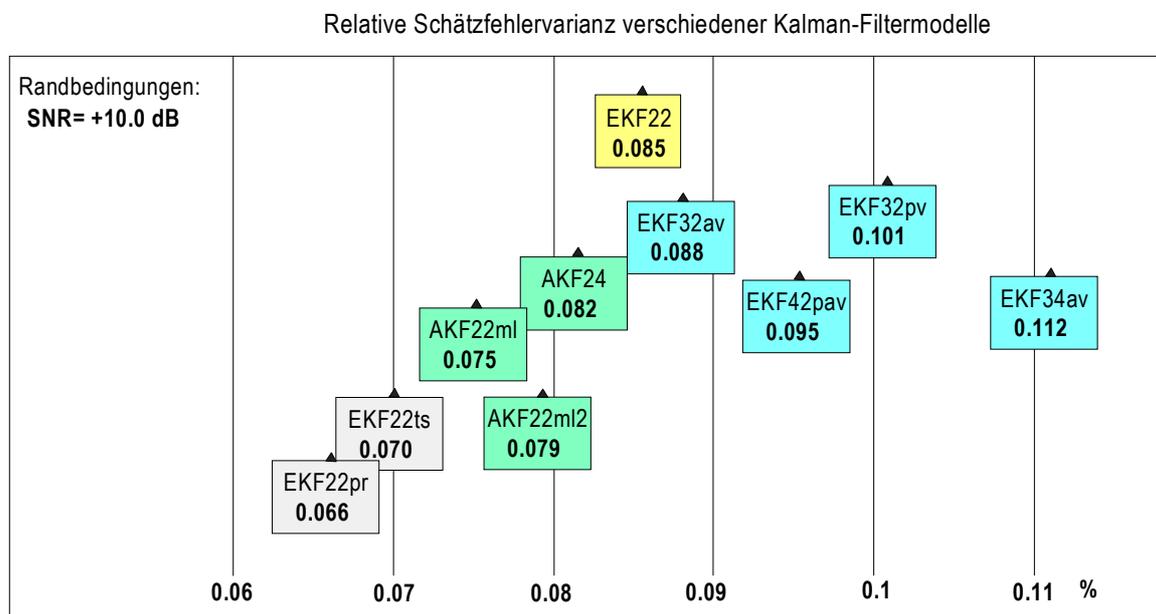


Abb. 4.22: Relative Schätzfehlervarianz verschiedener Kalman-Filtermodelle bei +10 dB

Die nächste Grafik zeigt die Residuensequenz bei einem SNR von +3.5 dB. Deutlich ist hier noch das Signal und ein geringer Offset, d.h. nicht mittelwertfrei, zu erkennen. Beides läßt sich auf das nichtlineare Problem zurückführen. Die Linearisierungen führen zu Fehlern, die deutlich in der Residuensequenz zu beobachten sind und das suboptimale Filtern andeuten. Daher macht es auch Sinn mit zusätzlichen Methoden (adaptiv oder durch additive Elemente) die Linearisierungsfehler zu verringern oder gar auszugleichen.

Bei einem linearen Kalman-Filter wird sich idealerweise eine optimale Filterung des Schätzwertes einstellen und die Residuensequenz wird ein weißes, gaußverteiltes und mittelwertfreies Rauschen darstellen.

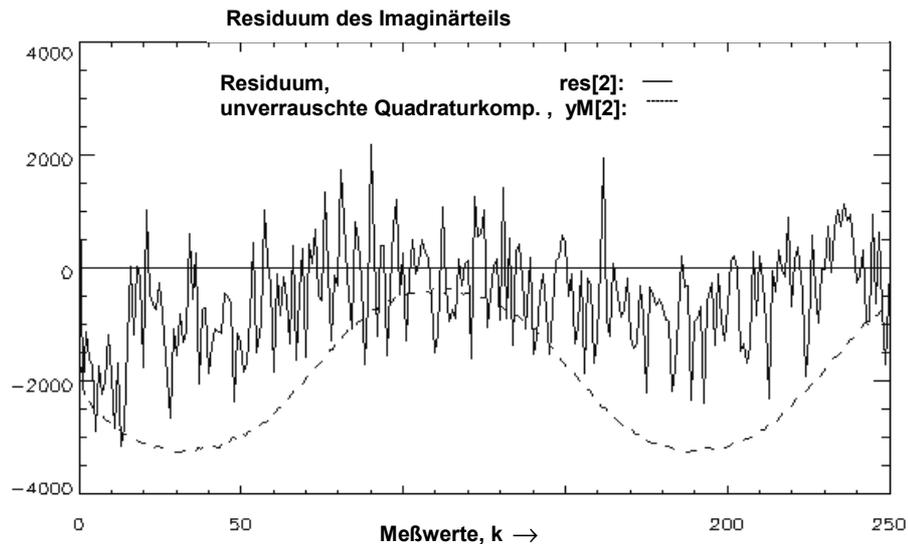


Abb. 4.23: Residuensequenz des Imaginärteils bei 3.5 dB SNR

4.4.3 Dynamisches Verhalten

Interessant für das Online-Filtering ist das dynamische Verhalten der Filter. Dazu zählen in erster Linie das Einschwingverhalten und die Stabilität eines Filters. Die Stabilität eines Filters wird durch die Anzahl der sogenannten Ausreißer bestimmt. Sie werden definiert als alle Monte-Carlo-Simulationen, die zu keinem Einschwingen auf dem bekannten Offset führen: Schwingt der Filter nicht oder auf einem anderen Offset ein, wird er als Ausreißer gewertet.

Interessant ist das dynamische Verhalten bei den adaptiven Filtern. Die folgenden Bilder zeigen die unterschiedliche Anwendung der Maximum-Likelihood-Schätzung der Systemrauschleistung. Zu sehen ist der Verlauf der Prädiktionsfehlerkovarianz (ähnlich mit der Filterfehlerkovarianz) und des adaptierten Parameters Q .

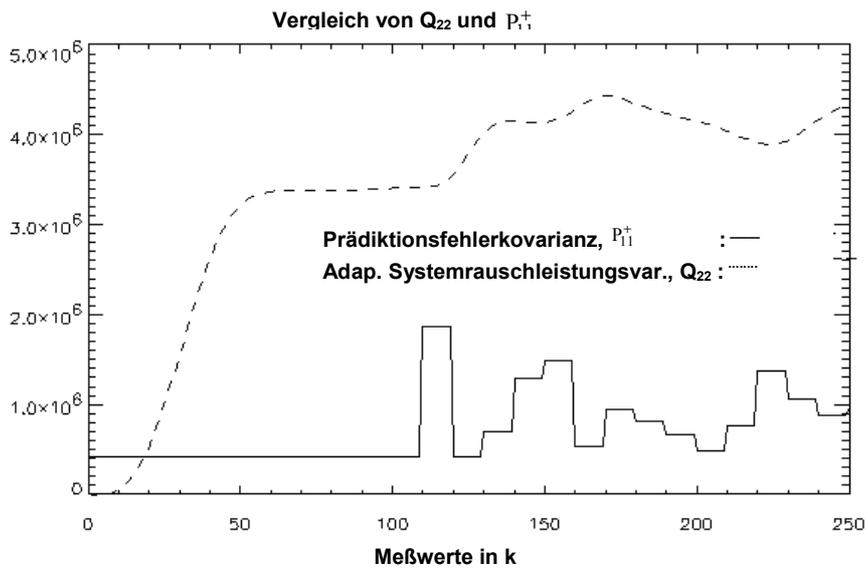


Abb. 4.24: AKF22ml (N=10)

Beim AKF22ml wird alle N-Werte der verwendete Q-Wert aufgefrischt. Zur Berechnung wird die Residuensequenz verwendet und über die vergangenen 10 Werte ausgewertet. Deutlich zu beobachten ist ein Nachlaufen der Prädiktionsfehlervarianzwerte um ca. 10-15 Filterschritte.

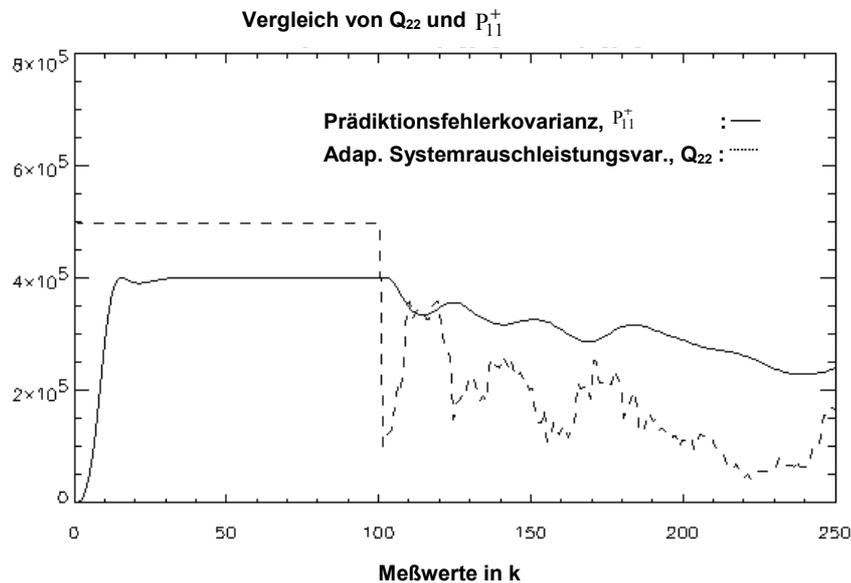


Abb. 4.25: AKF22ml2 (Mittelwert über 10 Werte des Q durch ML-Estimation)

Beim AKF22ml2 wird zu jedem Filterschritt ein neues Q bestimmt. Über einen kleinen Ringpuffer wird ein gleitender Mittelwert über N-Werte über die quadratische Residuensequenz ausgewertet.

4.4.3.1 Stabilität

Ausreißer setzen sich aus den beiden Fällen zusammen: Ein Ausreißer wird gezählt, wenn kein Einschwingen möglich ist, d.h. das Filter keine sinnvollen Schätzwerte liefert. Als Ausreißer werden jedoch auch folgende Filterdurchläufe gewertet, wenn ein Einschwingen in der geforderten Phasenlage (3. Quadrant) nicht möglich war, jedoch das Filter stabil in einer anderen Phasenlage arbeitete. Es soll hier keine Unterscheidung der beiden Fälle durchgeführt werden, obwohl eine Korrektur der Phasenlage, also eines fehlerhaften Offsets, durchaus weggerechnet werden könnte. Grundsätzlich läßt sich durch eine entsprechende Wahl der Startparameter bereits eine hohe Stabilität erreichen. Filtermodelle mit zusätzlichen virtuellen Beobachtungen führen aufgrund der Sensordatenfusion theoretisch zu einer Stabilisierung. Dagegen wirkt jedoch die Komplexität eines Filters. Bei steigender Komplexität wird das Filtermodell auch empfindlicher für Instabilitäten. Wichtet man die Varianzen der zusätzlichen virtuellen Parameter schwächer als theoretisch berechnet, stabilisiert sich das System. Die höhere Informationsverarbeitung der zusätzlichen Beobachtung nimmt jedoch ab.

Beim EKF22pr (Prefiltering) lassen sich bei großen Dynamikschwankungen des Quellsignals Instabilitäten beobachten, insbesondere dann, wenn beim Einschwingen das Quellsignal geringe Dynamik aufwies.

Die adaptiven Filtermodelle erwiesen sich als besonders stabil, da das Residuum (Differenz aus Messung und „berechneter“ Messung) ausgewertet wird.

Betrachtet man die Ausreißerstatistik (s. Abb. 4.19) bei ungünstig gewählten Startwerten, dann läßt sich jedoch das Verhältnis von Ausreißern zur Fehlervarianz betrachten. Die Anzahl der Ausreißer nimmt exponentiell schneller zu als sich $\hat{\sigma}_\varphi^2$ bei sich verschlechterndem SNR vergrößert. Bei verbesserten Startwerten beginnen Ausreißer erst später aufzutreten, jedoch ist die Zunahme dann genauso stark.

4.4.3.2 Einschwingverhalten

Beim Filtertuning lassen sich anhand des dynamischen Verhaltens der Filterfehlervarianz auch Rückschlüsse auf die gewählten Parameter ziehen. Wenn die Systemrauschleistung zu klein gewählt wurde, ist ein längeres Einschwingen des Filters erforderlich. Die

Filterfehlervarianz schwingt stark nach und der gemittelte Varianzwert des als eingeschwingen gewählten Bereichs variiert noch stark.

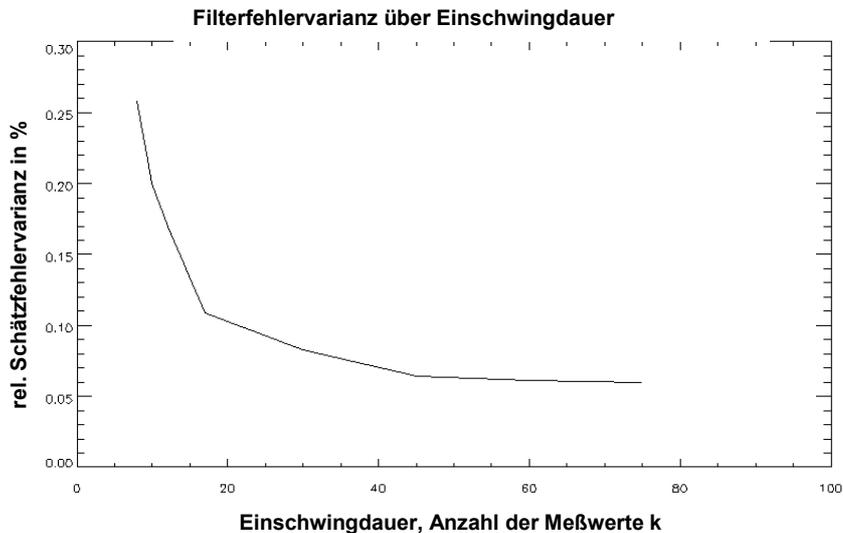


Abb. 4.26: Einschwingverhalten

In Abb. 4.26 wird das Einschwingverhalten bei unterschiedlichen Startwerten des Standardfilters EKF22 getestet. Die Filterfehlervarianz, deren Abhängigkeit zu den Startwerten gegeben wurde, ist bei langem Einschwingen am günstigsten. Dies wird hier, bedingt durch die nicht übereinstimmenden Modelle (System und Beobachtung), auch erwartet. Stimmen beide Modelle überein, so wird sich ein noch schärferer "Knick" der Kurve ergeben. Es gibt dann einen optimalen Parametersatz für Einschwingverhalten und Varianz.

Bei zusätzlichen Beobachtungen wurde beispielsweise ein längeres Einschwingen beobachtet, was sich durch ein komplexeres System erklären läßt. Das Einschwingverhalten läßt sich auch anhand des Prädiktionsfehlerkovarianzwertes P_{11}^+ gut beobachten.

Betrachtet man die Einschwingdauer über den auftretenden Ausreißern, dann läßt sich bis zu einer bestimmten Grenze ein linearer Zusammenhang beobachten. Bei langem Einschwingen ist ein Überschwingen bzw. das Einschwingen auf einem falschen Offset eher zu erwarten, als bei einer geringen Anzahl von Einschwingsschritten. Ein Einschwingen abhängig vom Filtermodell ist jedoch immer erforderlich, so daß sich das Einschwingen beim EKF22 bei 30 Werten einem Minimum nähert.

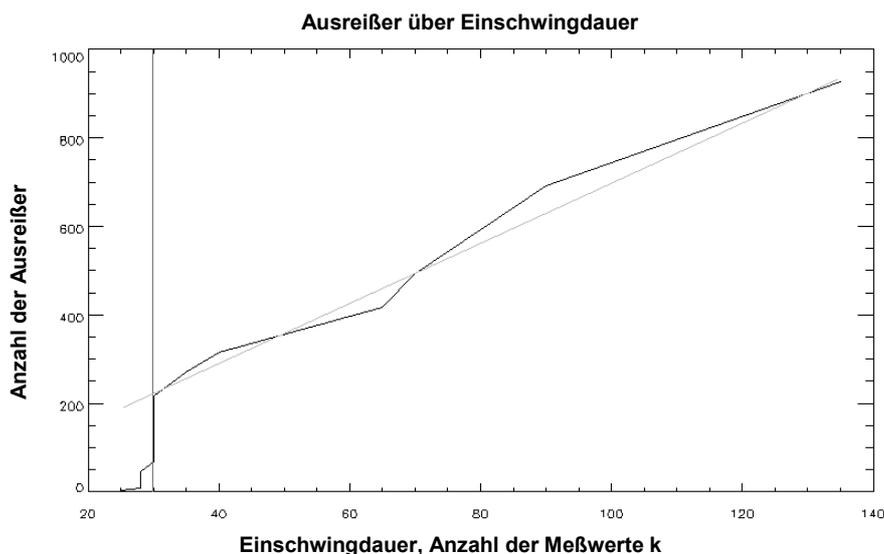


Abb. 4.27: Einschwingdauer über Ausreißer

4.4.3.3 Modulationsindex

Bei allen gemachten Messungen ist der Modulationsindex γ konstant gewählt worden. Es läßt sich zeigen, daß der Modulationsindex unabhängig vom betrachteten SNR bzw. CNR ist. Es wird sich jedoch direkt eine Auswirkung auf die Filterfehlervarianz beobachtet. Bei sehr kleinem γ wird der Träger kaum moduliert, so daß bei konstantem SNR eine hohe Filterfehlervarianz erscheint. Wird γ sehr groß gewählt, können die Quadraturkomponenten den genutzten 3. Quadranten verlassen und bei der Arcustangens-Berechnung der Referenzphase ergeben sich Fehler.

γ wurde als Variable GAMMA bei der Software mehrfach für andere, filterspezifische Parameter verwendet, was jedoch bei den jeweiligen Filterbeschreibungen erklärt wird.

4.4.4 Verwendung unterschiedlicher Quellsignale

Verwendet man anstelle des sinusförmigen Signalverlaufs ein Pseudo-Sprachsignal (s. 4.1.1) oder einen Markov-Prozeß, dann ergibt sich für die Filterfehlerkovarianz folgendes:

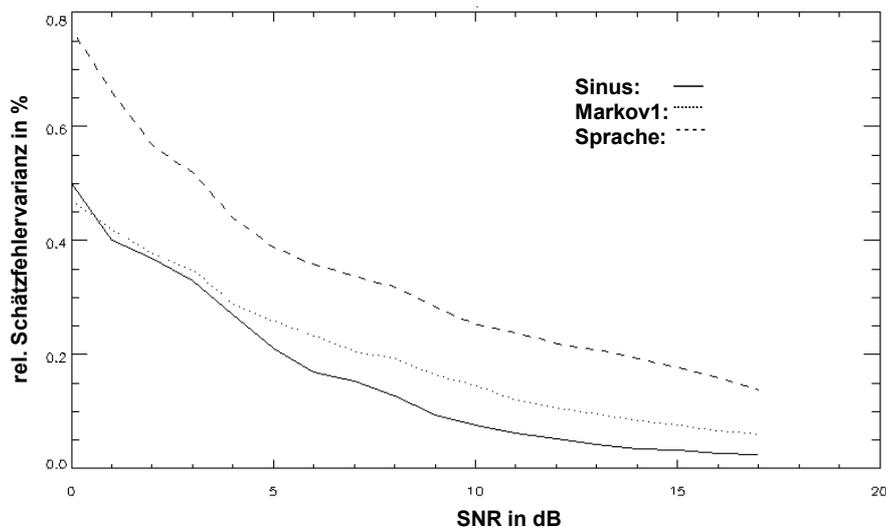


Abb. 4.28: Varianzen unterschiedlicher Quellsignale

Die Bandbreite des sinusförmigen Signals lag dabei bei max. 500 Hz, die des Pseudo-Sprachsignals und des zufälligen Markov-Signalverlaufs bei 4 kHz bzw. 20 kHz.

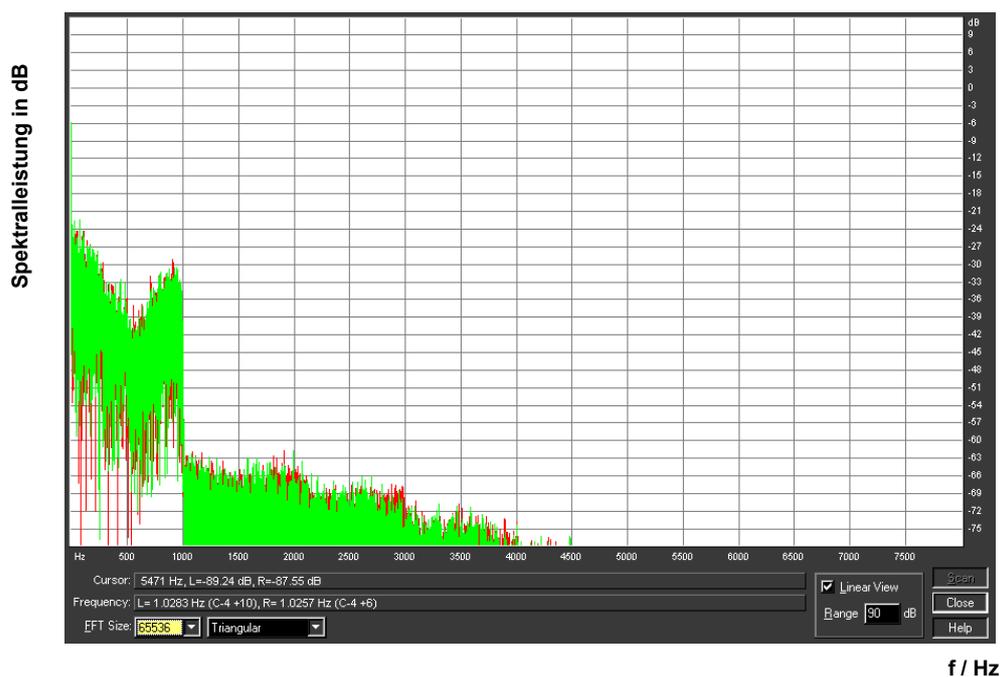


Abb. 4.29: Gemessenes Spektrum des Pseudo-Sprachsignals

Die starke Gewichtung der niedrigen Frequenzen entspricht der Fundamentalfrequenz im Sprachsignal (s. Abb. 4.29). Das Spektrum des Markov1-Signals (s. Abb. 4.30) wurde experimentell durch Betrachtung eines Zeitabschnitts ermittelt. Das Spektrum eines Markov1-Signals läßt sich analytisch nur schwer bestimmen, es ergibt sich jedoch eine Proportionalität mit $\frac{1}{f}$, die auch bei der experimentellen Ermittlung zu beobachten ist.

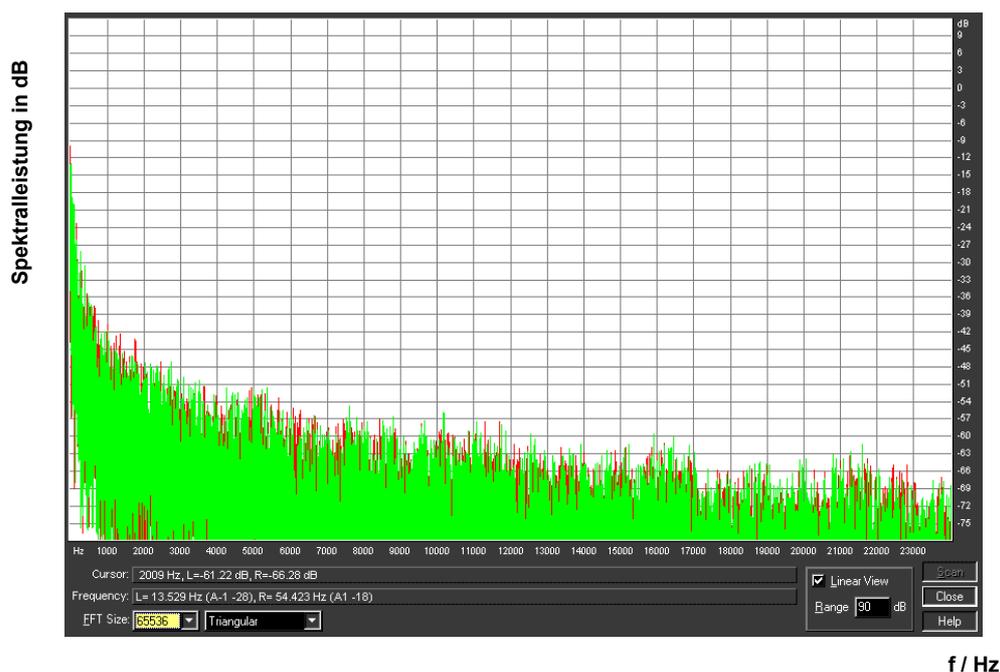


Abb. 4.30: Gemessenes Spektrum des Markov1-Signals

Aufgrund des passenden Systemmodells (ebenfalls nach Markov) ergibt sich bei einem Markov-Signal trotz wesentlich höherer Bandbreite einen recht guten Filterfehler-varianzverlauf insbesondere bei niedrigen SNR's. Unter Verwendung eines Pseudo-Sprachsignals erhalten wir einen relativ schlechten Verlauf, der insbesondere wegen der hohen Dynamikschwankungen zu erklären ist. Außerdem paßt das Systemmodell nur bedingt auf ein Pseudo-Sprachsignal, weil dessen Erzeugung deutlich unterschiedlich ist (s. 4.1.1).

4.5 DER MEßAUFBAU

4.5.1 Hardware

Die gemessenen Randbedingungen von Hardware und Software, die Einfluß auf die Meßstrecke haben, sollen im folgenden näher betrachtet werden. Die ebenfalls als komplette Hardware aufgebaute Übertragungsstrecke soll dabei nicht betrachtet werden, da sich eine exakte Messung der Eingangskomponenten und deren Fehlerbetrachtung als zu aufwendig erwies. Es wird die Übertragungsstrecke mit Signalgeneration über Software und Soundkarte beschrieben:

Aufgrund einer negativen Potentialdifferenz zwischen Soundkarte und AD-Wandlerkarte ergibt sich ein Offset. Die Phasendemodulation wird aus diesem Grunde im 3. Quadranten

durchgeführt. (Bei Betrieb mit einem Signalgenerator ergibt sich kein negativer Spannungsoffset!):

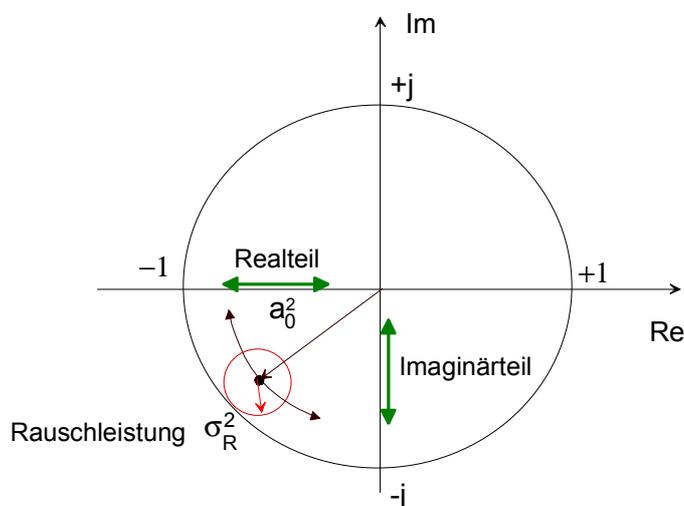


Abb. 4.31: Quadraturkomponenten bei Betrieb im 3. Quadranten

Ein direkter Vergleich des Quellsignals, das zur Erzeugung der Quadraturkomponenten eingesetzt wurde, mit der Schätzphase ist nicht möglich, da die Soundkarte nur einen AC-Ausgang besitzt. Ein möglicher, längere Zeit anhaltender gleichspannungsähnlicher Pegel würde von der Soundkarte zu Null weggeregelt. Aus diesem Grund wird die Bezugsphase direkt aus den Meßwerten (Arcustangens-Funktion) bestimmt, bevor diese durch einen künstlichen, additiven Rauschpegel (wie zu Beginn der Untersuchung angenommen) gestört werden. Aufgrund der beiden negativen Quadraturkomponenten ergibt sich für die Phase ein Bereich zwischen $-\frac{\pi}{2}$ und $-\pi$. (Die Diagrammansicht des Programms liefert diesen Wert in Volt, was als Ausgabe über einen DA-Wandler zu interpretieren ist.) Daraus läßt sich weiter ein Modulationsindex von $\gamma = \frac{\pi}{4}$ mit einem Phasenoffset von $-\frac{3\pi}{4}$ bestimmen.

$$\begin{aligned}
 s_{TRe}(t) &= a_0 \cos\left(-\frac{3\pi}{4} + \frac{\pi}{4}\varphi(t)\right) \\
 s_{TIm}(t) &= a_0 \sin\left(-\frac{3\pi}{4} + \frac{\pi}{4}\varphi(t)\right)
 \end{aligned}$$

Verwendete Quadraturkomponenten

Einen Vorteil bietet dieser negative Spannungsoffset noch: Die Störung mit konstanter Rauschleistung führt somit zu einem schwach schwankendem momentanen SNR.

Nulldurchgänge eines Signals führen ansonsten zu einem negativ unendlichen SNR. Den Offset wegzurechnen ist also nicht nötig.

4.5.1.1 Kalibrierung des Systems

Die von der Soundkarte abgespielten Quadraturkomponenten werden so angesteuert, daß sich eine Amplitude a_0 von ca. 3V ergibt (Aufgrund der Phasenlage erscheint ein Pegel von max. -3.5 V). Die Veränderung des Pegels entspricht einer Veränderung des Modulationsindex. Die Pegel sollten während der Messungen nicht mehr verändert werden, da die Filterergebnisse auch abhängig vom Modulationsindex sind. Ebenso wird bei den Online-Kalman-Filtermessungen der Modulationsindex nicht mehr verändert, so daß eine feste Definition des Systems existiert. Die Rauschleistung wird über das eingestellte SNR und die gemessene Signalleistung definiert.

Bei abgeschaltetem Rauschen wird danach das System auf die minimale Schätzfehlervarianz mit Hilfe der Quadraturkomponenten ausgeregelt. Es erscheint eine relative Schätzfehlervarianz von ca. 0.0035% zwischen berechneter Referenzphase und Kalman-Filterphase des EKF22. Um unterschiedliche Meßreihen vergleichbar zu machen, sollte diese Kalibrierung grundsätzlich durchgeführt werden und vor einer Meßreihe überprüft werden! (Es treten Schwankungen beim Ausgangspegel der Soundkarte auf.)

Grundsätzlich dürfen jedoch keine positiven Spannungspegel auftreten, da sich sonst Fehler in der Filterauswertung einschleichen. Eine Aussteuerung bis in den Randbereich des 3. Quadranten sollte daher vermieden werden. Fehler entstehen durch die Bestimmung der korrekten Bezugsphase aus den unverrauschten Quadraturkomponenten. Eine falsche Phasenberechnung der Bezugsgröße wird dann als Fehler dem Kalman-Filter angelastet, welches jedoch unabhängiger von den Quadranten arbeitet, da er sich an der Änderung des gesuchten Signalverlaufs orientiert. Die Arcustangens-Funktion engt somit bei dieser Anwendung den Arbeitsbereich ein. Grundsätzlich lassen sich jedoch Messungen in einem Pegelbereich von $\pm 5V$ mit 4 AD-Wandlern auswerten. (Der Potentialabgleich kann über ein einfaches Poti ausgeregelt werden.)

Zur Laufzeitmessung wird eine konstante, deterministische Schwingung moduliert und anhand von Schwingungen und Abtastwerten die Abtastrate und somit die Laufzeit eines Filterdurchlaufs ermittelt.

4.5.2 Software

Eine komplette Beschreibung der Software soll hier entfallen, da die Programmierung und der Aufbau der verwendeten Entwicklungsumgebung zwar Hintergrund aber nicht Thema der Arbeit sind. Jedoch sollen einige Fakten zur Optimierung und Struktur der Filterprogramme genannt werden, um den Unterschied zwischen reiner simulierter Demodulation und Online-Demodulation zu verdeutlichen. Daraus ergeben sich auch Erkenntnisse über die Filterlaufzeit bzw. die Bandbreiten, die sich bei den unterschiedlichen Filtermodellen ergeben.

4.5.3 Organisation und Design des EKF-Programms

Bei der Programmierung eines Extended Kalman-Filters müssen Design und Organisation des Programmes wohlüberlegt sein. Um eine möglichst schnelle Verarbeitung zu gewährleisten, muß insbesondere die Struktur des Filters mit möglichst wenigen Schleifen aufgebaut werden. Ferner lassen sich bei der Berechnung der Matrizen Nullelemente vorher erkennen und eliminieren.

Da eine Vielzahl von unterschiedlichen Filtermodellen in der erstellten Entwicklungsumgebung integriert wurden, hat sich die folgende Vorgehensweise bewährt. Das DSP-Hauptprogramm wählt nach PC-Vorgabe ein Filter aus, danach wird die dazugehörige Initialisierung des Filters durchgeführt, die Meßwerte des AD-Wandlers geladen und dann das gewählte Filter gestartet. Filter- und Initialisierungsroutinen sind immer in einem eigenen C-File abgespeichert. Die Variablen der Filter sind entsprechend den Kalman-Filterbezeichnungen benannt und sind in allen Filtern identisch. Es wurde eine globale Definition der Variablen durchgeführt, um feste Speicheradressen im DSP anzulegen. Die Startwerte werden als Prädiktionswerte aufgefaßt, so daß damit der erste Beobachtungsübergang definiert wird. Von dort wird der erste Filterschätzwert aus den Startwerten in einer Unteroutine berechnet, die die Kalman-Filtergleichungen enthält. Anschließend erfolgt die Prädiktion für die Estimation des nächsten Schätzwertes. Nach dem Rücksprung ins Hauptprogramm wird die AD-Wandlerkarte zur nächsten Wandlung gestartet. Durch die direkte Verarbeitung jedes Meßwerts läßt sich direkt die Filtergeschwindigkeit/-laufzeit ermitteln.

4.5.4 Optimierung des Kalman-Filters (Zeit=Code)

An erster Stelle steht die Zeitoptimierung durch Entfernung von Leerberechnungen. So reduziert sich die Laufzeit eines EKF schon um ca. 20%. Umfangreichere Filtermodelle erschließen noch größere Einsparmöglichkeiten.

Wichtiges Kriterium für die Programmierung des Kalman-Filters sind die Symmetrien verschiedener Matrizen. Durch die begrenzte Genauigkeit der Floating-Point-Berechnung kann es bei der Berechnung der unterschiedlichen Matrizenelemente zu Unsymmetrien kommen, die schnell zu Stabilitätskonflikten im Betrieb des Filters führen. Wichtig ist es deshalb, schon im Vorfeld die Matrizenelemente zu kennen und wichtige Symmetrien im Verlauf des Filters zu überprüfen. Eine weitere Möglichkeit besteht darin, die Filtergleichungen noch vor deren Programmierung so zu gestalten, daß zwangsweise Symmetrien beibehalten werden! Bei sensiblen Filtermodellen werden deshalb Fehlerkovarianz- und Prädiktionsfehlerkovarianz auf deren Symmetrie hin untersucht. Ebenso kritisch ist die Bildung der inversen Matrix zur Bestimmung des Kalman-Gains. Extrem kleine Werte bei der Division und die damit verbundenen Rundungsfehler lassen sich vermeiden, indem man zuvor mit einem Faktor multipliziert und nach erfolgter Division den Faktor wieder eliminiert.

Programmtechnisch kann die Berechnung des Störsignals vor die Filterung geschoben werden, so daß bei der Filterung nur noch aus einem angelegten Rauschvektor die Zufallswerte ausgelesen werden. Dies führt zu einer zusätzlichen Beschleunigung um ca. 7% (Bei EKF22 von 14 kHz auf 15 kHz Abtastrate).

Eine weitere Optimierungsmöglichkeit liegt im Bereich der Assemblerprogrammierung. Die C-Programme, deren Code auch als Assemblerfile vom Compiler zur Verfügung gestellt werden kann, eröffnen die Möglichkeit, den Assemblercode weiter zu optimieren. Da jedoch die eingesetzten C-Compiler bereits hochoptimierten, ausführbaren Code liefern, können nur noch geringe Optimierungen der Laufzeit erreicht werden (Größenordnung von 5%), was jedoch den Aufwand nicht rechtfertigen würde. Die Lesbarkeit der Programme ist bei Assemblercode erheblich schlechter als beim C-Programm. Die Optimierungen werden z.B. auch durch das sogenannte Pipelining (das Ineinanderschachteln von Befehlen, das aufgrund der DSP-Struktur möglich ist) erlangt.

4.5.4.1 Betrachtung der Bandbreite/Laufzeit

Bei zusätzlichen Beobachtungen oder Zuständen geht die Rechendauer zwangsläufig in die Höhe. Bei zusätzlichen Beobachtungen vergrößert sich insbesondere der Aufwand erheblich, da sich dies direkt auf die Matrizengröße der zu invertierenden Matrix bei der Bestimmung des Kalman-Gains auswirkt. Bei Matrizen größerer Ordnung muß - um Stabilität zu gewährleisten - zusätzlich der Rang der Matrix überprüft werden.

Ohne Optimierungen erhöht sich der Rechenaufwand von drei auf vier Beobachtungen um über 120 % . Die Berechnung der inversen Matrix benötigt exponentiell mehr Rechenoperationen.

Es folgt eine Auflistung der untersuchten Filter nach Rechenaufwand und gemessener Abtastrate:

Filter	Rechenoperationen	Abtastrate (Messung) in kHz
EKF22 (Extended Kalman-Filter)	ca. 350	ca. 14,0
EKF22s, EKF22ss (zeitoptimierte Modelle, Singermodell)	ca. 260	ca. 18,5
EKF22pr (EKF22 + Vorfilterung)	ca. 210	ca. 19,5
EKF32pv (EKF22 + virtueller Phase)	ca. 680	ca. 10
EKF34av (EKF22 + virt. Amplitude)	ca. 1050	ca. 8,0
EKF42pav (EKF22 + virt. Ampl. & Phase)	ca. 720	ca. 9,5
EKF24 (EKF22 + Amplitudenschätz.)	ca. 730	ca. 10,5
AKF24 (Adaptiver EKF22 durch Zustandserweiterung)	ca. 720	ca. 11,0
AKF22ml (Adaptiver EKF22 mit Max.-Likelihood-Schätzung)	ca. 300	ca. 15,5
AKF22ml2 (Adaptiver EKF22 mit Max.-	ca. 300 .. 400	ca. 14 .. 15,5

Filter	Rechenoperationen	Abtastrate (Messung) in kHz
Likelihood-Schätzung ²⁾		
SKF22fl (N=10) ¹² (Fixed-Lag-Smoother)	ca. 470+ 260 x N = 3070	ca. 3,0
SKF22fp (N=1..30) (Fraser-Potter-Smoother)	ca. 470+260 x N	ca. 1,7-8,0
SKR22ro: (Smoother-Filter mit red. Ordnung)	ca. 570	ca. 12,0
AVG (Kein EKF nur gleit. Mittelwert)	22	ca. 21,0

Tab. 4.1: Einsatz von DSP's (TMS320 C4x / 40 MHz)

Möchte man keinen Kalman-Filter verwenden, muß die Filterwahl "AVG" getroffen werden. In dieser Funktion ist ein gleitender Mittelwert (Moving Average) definiert. Mit dem Gamma-Parameter läßt sich die Breite wählen. Bei Gamma = 1 läuft das Filter quasi ohne Filterung. Der verbleibende restliche Programmcode führt zu einer Abtastrate von 21 kHz mit etwa 120 Rechenoperationen.

Die Leistungsfähigkeit des DSP stellt den Engpaß bei der Erhöhung der Bandbreite dar. Die realisierbaren Bandbreiten lassen sich durch schnellere DSP's optimieren. Die Variante mit 60 MHz (30 MOps) bringt gegenüber der 40 MHz-Variante (20 MOps) einen Zugewinn von 50%, so daß mit dem schnelleren baugleichen Modul direkt Abtastraten von 20-30 kHz erreicht werden. Ein weiteres Modul, bestückt mit dem DSP TMS320C6xx (Floating Point), besitzt eine Taktrate von 200 MHz, so daß ebenfalls deutlich höhere Abtastraten und somit auch Bandbreiten verarbeitet werden können.

4.6 FEHLERBETRACHTUNG BEI DER ONLINE-DEMODULATION

Die Hardwarerealisierung stellt einen hohen Aufwand dar. Zusätzlich birgt sie eine Vielzahl von Fehlerquellen, die mathematisch alle nur schwer erfaßt werden können. Insbesondere stellt sich die Kohärenz des Trägersignals bei der Mischerstufe des Quadraturdemodulators sowie deren exakte 90°-Phasendrehung als kritisch heraus. Diese sind nur bei der genauen

¹² N = Anzahl der Rückwärtsfilterschritte = Größe des Ringpuffers

Grenzfrequenz eines Tiefpasses 2. Ordnung gegeben. Liegt die Grenzfrequenz des Tiefpasses anders als die verwendete Trägerfrequenz, ergeben sich Phasenfehler. Aus diesen Gründen soll im folgenden nur auf die Fehler der Softwarevariante (s. 5.1) eingegangen werden:

4.6.1.1 Soundkarte

Durch Verwendung einer kostengünstigen Soundkarte als Signalquelle ist jedoch keine optimale und originale Übertragung eines winkelmodulierten Signals möglich. Trotz der AC-Entkopplung des Ausgangs wurde die Soundkarte jedoch verwendet, da sich die Referenzphase auch aus den bekannten Quadraturkomponenten direkt zusammensetzen läßt. Durch diese Methode entfallen Fehler, die durch die Übertragungsstrecke entstehen, da eventuelle Fehler auch in der Referenzphase verarbeitet werden. Somit führen Quantisierungsfehler und quantisiertes Rauschen nicht zu einem Fehler in der Auswertung der Filterschätzphasen.

4.6.1.2 Fehler der AD-Wandlerkarte

Fehler, die bei der Abtastung mit der AD-Wandlerkarte entstehen, werden als Quantisierungsfehler bezeichnet und sind aus der Theorie allgemein als gleichverteilt bekannt. Eine Messung ließe sich mit einer Rampenfunktion durchführen. Da diese Fehlerquelle jedoch keine Beeinflussung der Messungen zur Folge hat, wird sie nicht weiter betrachtet.

4.6.1.3 Quantisiertes Rauschen des Systems

Gemessen wurde das quantisierte Rauschen, das das Restrauschen der Hardware beschreibt. Es liegt jedoch deutlich unter einem störenden Pegel. Da anschließend eine Überlagerung mit recht hohen Störgrößen durchgeführt wird, ist eine weitere Betrachtung ebenfalls nicht erforderlich.

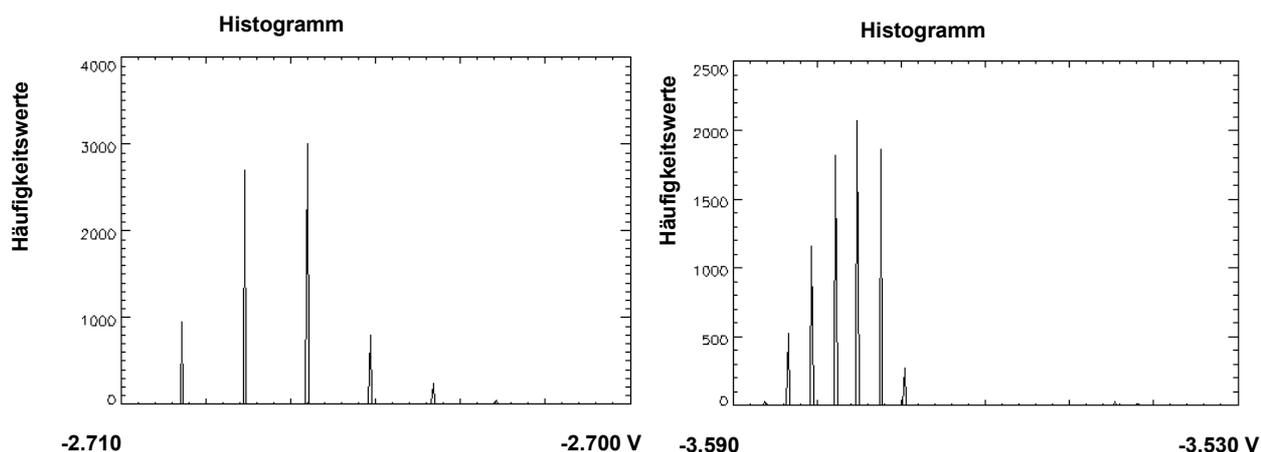


Abb. 4.32: Histogramme von Realteil und Imaginärteil des quantisierten Rauschens

4.6.1.4 Rundungsfehler

Fehlerquellen, die nach der Abtastung entstehen, sind beispielsweise Rundungsfehler des DSP's. Bei einer 32 Bit Architektur umfaßt der Wertebereich einer Fließkommavariablen $5.8774817 \cdot 10^{-39}$ bis $5.8774817 \cdot 10^{+39}$. Werte, die außerhalb dieses Wertebereichs liegen, werden gerundet. Tatsächlich treten Werte im Bereich des Kovarianzenzyklusses auf, die am Rand dieses Zahlenbereichs liegen können, so daß Rundungsfehler auftreten können. Aus diesem Grunde werden Kovarianzzyklen an manchen Stellen auf Positivdefinitheit überprüft und durch Zwangssymmetrierung die Symmetrie einer solchen Matrix wieder hergestellt.

Bei der Übergabe an die PC-Delphi-Umgebung gibt es nur eine Übergabe als long integer. Dies bedeutet, daß Variablen mit einem bestimmten Faktor versehen werden und nur der auftretende Wertebereich der Variablen übertragen wird. Hier sind Rundungsfehler ebenfalls möglich.

4.6.1.5 Modellfehler

Weitere Fehlerquellen, die jedoch in der Filtermodellierung zu suchen sind, sind die Nichtlinearitäten jedes Filtermodells. Hinzu kommen weitere systematische Fehler, die bei hypothetischen Annahme von gaußverteilten Verteilungsdichten der Meßwerte und Schätzwerte gemacht werden. Die Modellfehler lassen sich bei der Kalibrierung der Meßreihe nur in ihrer Summe feststellen. Bei unverrauschten Quadraturkomponenten läßt sich immer noch ein kleiner Fehler feststellen. Diese Filterfehlervarianz liegt beim EKF22 im Bereich von 0.002 - 0.005 %, abhängig von der Parameterwahl (Gamma, Amplitude).

5 Die Entwicklungsumgebung

Entscheidend für einen effektiven Test der bereits beschriebenen Online-Kalman-Filtermodelle ist die Verwendung einer leistungsfähigen Hardware und Software. Aufgrund der aufwendigen, rechenintensiven Matrizenoperationen des Kalman-Filters fiel die Wahl auf ein DSP-gestütztes System. Um bestehende Kenntnisse der TMS320-Familie (Texas Instruments) zu nutzen, wurde der Floating Point DSP TMS320C40 eingesetzt, der zu Beginn dieser Arbeit neu auf den Markt kam. Im Laufe von mehreren Hardwaresystemen und Softwareversionen wurde eine unter WIN95-lauffähige Entwicklungsumgebung geschaffen, über die Download, Online-Parametertuning, Filterauswahl mit direkter Datenübertragung über den PCI-Bus und grafische Darstellungen der Filterergebnisse möglich sind.

5.1 KONZEPT DER ÜBERTRAGUNGSSTRECKE

Bei der Entwicklung einer Simulationsumgebung zum Test von nichtlinearen Kalman-Filtern am Beispiel der Winkelmodulation, wurde ein vollständiger Übertragungskanal aufgebaut. Problematisch war es, Modulatoren zu finden, die gleichzeitig die Trägerfrequenz mit exakter Phasenlage liefern können. Aus diesem Grunde wurden 2 Oszillatoren aufgebaut, die bei gleicher Trägerfrequenz ($f_0=1$ MHz) arbeiten. Einer der beiden wurde nicht moduliert und zusätzlich mit einem Quarz stabilisiert. Der andere wurde mit einem Nutzsignal moduliert. Damit die Mittenfrequenz des modulierten Oszillators nicht schwankt, abhängig von der Art und Amplitude des Nutzsignals, wurde eine zusätzliche kapazitive Kopplung an dem stabilisierten Oszillator verwendet. Bei fehlender Ansteuerung ist so gewährleistet, daß die beiden Oszillatoren absolut phasengleich arbeiten. Ein Nachteil dieser recht primitiven Realisierung ist die geringe Aussteuerbarkeit des Modulators. Bei zu hoher Aussteuerung beginnt die stabilisierte Frequenz zu schwanken, was trotz einer Quarzstabilisierung nicht verhindert werden kann. Die hardwaremäßige Realisierung hat zusätzliche Nachteile: Durch den gesamten Aufbau ergeben sich zusätzliche Störgrößen und Nichtlinearitäten, deren genaue Kenntnis fehlt (s. Abb. 5.1 die Größen η und $n'(t)$). $n'(t)$ muß hier als klein gegenüber der definierten, generierten Störgröße angesehen werden.

Das modulierte Signal sowie das Trägersignal werden am Quadraturempfänger eingespeist, der die beiden Quadraturkomponenten am Ausgang liefert. Beide Ausgangssignale werden mit einem 12 - Bit A/D-Wandler in zeitdiskrete, quantisierte Signale umgeformt, die den unterschiedlichen Kalman-Filtermodellen als Beobachtungsgrößen dienen. Der Kalman-Filter-Algorithmus ist auf einem DSP-Board realisiert worden, der die Schätzwerte als Daten ausgibt. Diese können dann wahlweise zum PC übertragen werden oder per D/A-Wandler als analoges, zeitkontinuierliches Signal direkt wieder ausgegeben werden.

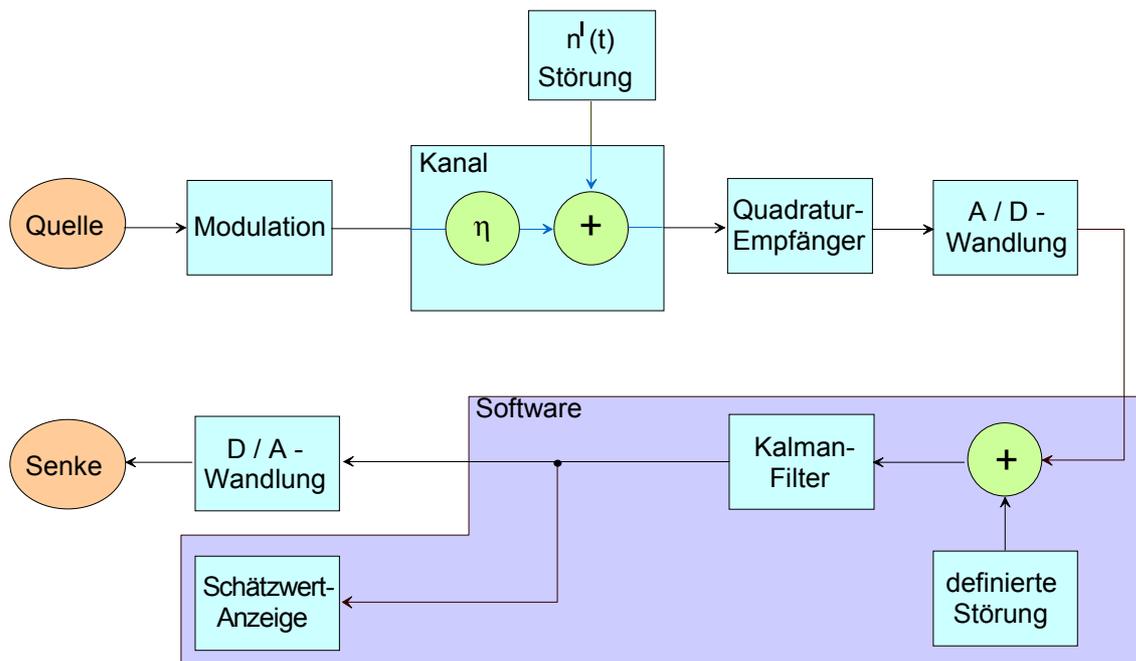


Abb. 5.1: Hardwarevariante der Übertragungsstrecke

Aus diesem Grunde wurde ein weiteres Konzept realisiert, mit dem auch die zuvor beschriebenen Online-Simulationen durchgeführt wurden. Die Modulation eines selbstdefinierten Nutzsymbols wurde als Software (IDL) realisiert. Bei einer derartigen Realisierung können die unbekanntes Störgrößen und Nichtlinearitäten umgangen werden. Bis hin zum Heruntermischen in den Tiefpaßbereich wird eine durchweg digitale Lösung realisiert. Damit ergibt sich so die Möglichkeit, Bandbreite des Nutzsymbols und Bandbreitebedarf von unterschiedlichen Modulationsarten unkompliziert und schnell zu bestimmen. Die errechneten Quadraturkomponenten werden als Daten-File gespeichert und können dann jederzeit von einer handelsüblichen Soundkarte "abgespielt", d.h. D/A-gewandelt werden. Ein weiterer Vorteil ist die 100%-ige Reproduzierbarkeit eines Nutzsymbols.

Der verbleibende Aufbau mit A/D-Wandlung und DSP-gesteuerter Filterverarbeitung ist bei beiden Varianten identisch. An der Schnittstelle der A/D-Wandlung lassen sich grundsätzlich auch andere, hochwertigere Modulationssignale einspeisen.

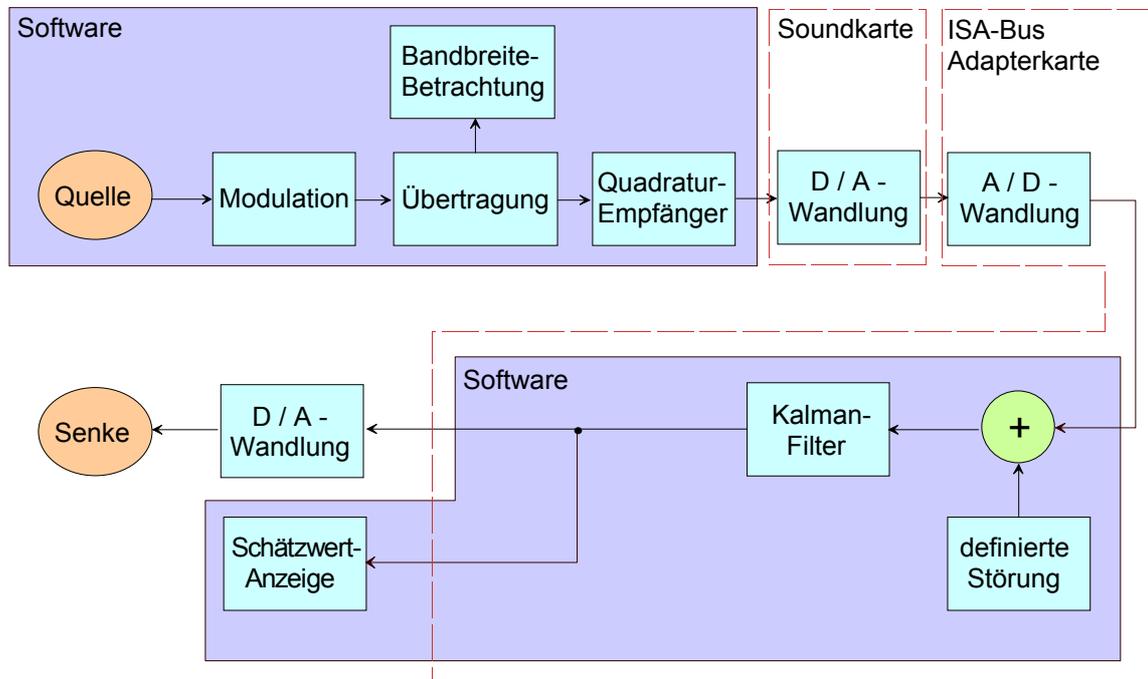


Abb. 5.2: Softwarevariante der Übertragungsstrecke

Trotz der zeitweise nicht zeitkontinuierlichen, schrittweisen Bearbeitung eines Nutzsignals handelt es sich um eine Online-Verarbeitung des Signals.

5.2 REALISIERUNG DER HARDWARE

Die technischen Einzelheiten der Hardware-Module werden in einem gesondertem, erweiterten Anhang beschrieben, da sie nicht direkter Bestandteil der wissenschaftlichen Untersuchung sind. Es sollte jedoch erwähnt werden, daß es sich um einen extrem kostengünstigen Aufbau der Entwicklungsumgebung handelt. Neben dem DSP-Modul (ca. 1600,- DM) entstanden kaum weitere Kosten, so daß inkl. Soundkarte, AD-Wandlern, SARAM die Kosten kaum die 1800,- DM überstiegen haben. Eine Aufzählung gibt Einblick in die bestehenden und verwendeten Komponenten:

- FM-Sender (Trägerfrequenz 1 MHz oder 1.25 MHz)
- Funktionsgenerator (Sinus, Rechteck, Dreieck)
- Empfänger - Modul I (Analog-Multiplizierer MLT04, 1 MHz)

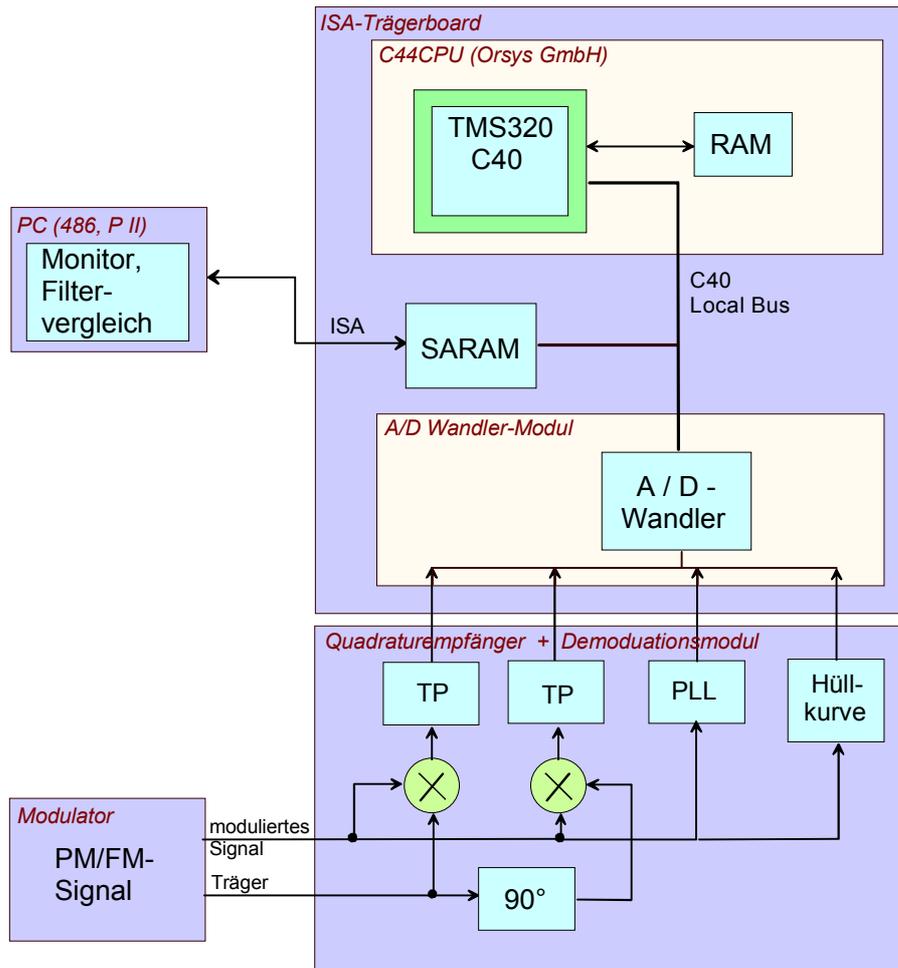


Abb. 5.3: Blockschaltbild der ISA-Bus Adapterplatine

- Empfänger - Modul II (MIQY-1.25D von Mini Circuits; Schottky-Dioden-Netzwerk bei 1.25 MHz)
- 90° - Phasenschieber (1 MHz)
- Empfänger - Modul I (Analog-Multiplizierer MLT04, 1 MHz)
- Frequenzdemodulation mit PLL (NE564 von Philips Semiconductors)
- Hüllkurvendemodulator
- AD/DA-Wandlerkarte I/II (MAX120/122 (Maxim); 12 Bit Auflösung bei 500 ksps)
- Cyclops C40-Karte (Ariel, TMX320C40)
- DA-Wandler (AD768 von Analog Device)

- Embedded DSP-Modul DSPC44 (Orsys, TMS320C44, 40MHz)
- ISA-Bus-Adapterplatine als Trägerplatine für AD-Wandlerkarte und Embedded DSP-Modul (Dual-Ported RAM IDT 70825)

5.3 SOFTWARE DER ENTWICKLUNGSUMGEBUNG

Die verwendete Software soll hier nur knapp beschrieben werden, da dies sonst den Rahmen dieser Arbeit übersteigen würde. Grundsätzlich wurden jedoch 3 völlig voneinander unabhängige Testumgebungen geschaffen, mit denen Kalman-Filter zur Demodulation winkelmodulierter Daten eingesetzt werden können. Unter IDL und BorlandC wurden einfache Simulationsprogramme entwickelt, die einen schnellen Test eines Filters zulassen. Eine Vorauswahl und eine grundsätzliche Machbarkeit der Filter konnte somit schon im Vorfeld geprüft werden. Die Hauptentwicklungsumgebung, die zum endgültigen Test und der Auswertung der Filtermodelle verwendet wurde, ist als Online-Entwicklungsumgebung konzipiert worden und enthält als Hauptblöcke die Prozessierung unter C auf dem DSP-Board und die Weiterverarbeitung der Daten auf dem PC mit Hilfe einer Delphi-Oberfläche.

5.3.1 Simulator- Entwicklungsumgebung

Ausgehend von einem IDL-Programm zur Simulation wurde eine Umsetzung in die Programmiersprache C und eine Weiterentwicklung des IDL-Programms als Widget-Applikation durchgeführt.

5.3.1.1 BorlandC-Simulator

Zum Test der Filter auf dem PC als C-Code wurde dieser Filtersimulator entwickelt. Somit wird die anschließende Umsetzung für die DSP-Programme vereinfacht, da ein Test der DSP-Routinen unter der anfänglichen DSP-Entwicklungsumgebung noch nicht möglich war, so daß Änderungen im Code erheblichen Zeitaufwand bedeuteten.

5.3.1.2 Widget-Simulator

Parallel dazu wurde eine Widget-Simulation entwickelt, die die grafische Umsetzung der Ergebnisse und Zwischenvariablen anschaulich darstellt. Ein anfängliches Filtertuning

konnte so während der Simulation durchgeführt werden. Die Optionen und Möglichkeiten des Widget-Simulators dienen dann als Vorlage des Delphi-Programms, das die Auswertung der "echten" Online-Filterergebnisse letztendlich durchführt. Der Widget-Simulator kann als Vorlage für allgemeine Kalman-Filterentwicklungen weitergenutzt werden.

5.3.2 Online-Entwicklungsumgebung

Die abschließende Entwicklungsumgebung zum Test der unterschiedlichen Kalman-Filtermodelle besteht aus der DSP-seitigen Programmierung der Kalman-Filter auf der im erweiterten Anhang beschriebenen Hardware. Der Programmcode umfaßt zur Zeit etwa 168 kB der vorhandenen 256 kB, so daß für weitere Filtermodelle genügend Platz ist.

Die Organisation der Filtermodelle ist unter Abschnitt 4.5.3 beschrieben. Über das SARAM können Daten bidirektional ausgetauscht werden, so daß nicht nur Filterschätzwerte und Parameter des DSP's übertragen werden, sondern zusätzlich ein Filtertuning des Kalman-Filters auf dem DSP vom PC aus möglich ist.

5.3.2.1 Filtertuning, Filterauswahl

In einem 2x4 KWorte (16 Bit) großen Feld werden die Parameterdaten der beiden Programme ständig ausgetauscht. Eine aktuelle Anzeige der Parameter und Filterergebnisse ist somit gewährleistet. Dabei wird immer eines- der beiden Felder abwechselnd zum Lesen bzw. zum Schreiben genutzt, um einen störungsfreien Datenaustausch zwischen PC und DSP zu gewährleisten und die Übertragung zu beschleunigen.

Ein Zusatzmenü unter Delphi erlaubt es, bitweise den Datentransfer mit Hilfe einer Vielzahl von erstellten Systemroutinen zu kontrollieren. Sie erlauben jederzeit Funktionalitätstests der Entwicklungsumgebung.

5.3.3 Signalgeneration

Das eingespeiste Signal wird unter IDL generiert und basisbandmoduliert. Diese Quadraturkomponenten werden anschließend als File gespeichert und können dann von einer Soundkarte abgespielt werden.

5.3.3.1 Signalerstellung, Modulation und Bandbreitenbetrachtungen unter IDL

Zur Signalerzeugung wurde ein IDL-Programm entwickelt, das zusätzlich Bandbreitenuntersuchungen bei unterschiedlichen Modulationsverfahren durchführt. Als Quellsignale wurden die beschriebenen Testsignale eingesetzt (s. 4.1). Neben der Bandbreitenuntersuchung lassen sich die entstehenden Quadraturkomponenten als TXT-Dateien abspeichern.

5.3.3.2 Abspielen der Signale mit Soundkarte aus TXT-File

Zum Abspielen der TXT-Dateien auf einer handelsüblichen Soundkarte wurde das Programm COOL.EXE (Shareware) verwendet. Die Vor- und Nachteile bei Verwendung einer Soundkarte wurden unter 4.5.1.1 beschrieben.

6 Zusammenfassung

Die Arbeit hat gezeigt, daß Extended Kalman-Filter zur Online-Demodulation winkelmodulierter Signale grundsätzlich geeignet sind. Abhängig vom betrachteten SNR (CNR) und der Wahl der Quellsignale müssen unterschiedliche Extended Kalman-Filtermodelle gewählt werden. Im betrachteten Bereich von 0 bis 17 dB zeigen insbesondere adaptive Kalman-Filtermodelle eine hohe Genauigkeit. Dabei ist die externe Parameter-Anpassung durch Maximum-Likelihood Schätzung zu bevorzugen, da Filterlaufzeit eingespart wird und das Filter nicht durch zusätzliche Komponenten destabilisiert wird. Besitzen die Signale nur geringe Dynamik, dann lassen sich vorgefilterte Kalman-Filter einsetzen, bei denen im stationären Betrieb auf den Kovarianzenzyklus verzichtet wird. Erweiterte Filtermodelle zeigen insbesondere bei negativen SNR's die Vorteile der Datenfusion, ihr Verhalten ist jedoch grundsätzlich instabiler. Die Gruppe der Smoothermodelle ist für Online-Anwendungen eher ungeeignet, da extrem hohe Filterlaufzeiten benötigt werden, um Stabilität und Filtergenauigkeit steigern zu können. Eine Ausnahme stellt die Filterrealisierung mit reduzierter Ordnung dar, bei dem die Stabilitätseigenschaften der Smoother ausgenutzt werden, jedoch ist der ermittelte Fehler der Schätzphase recht hoch, so daß ein Einsatz des Filters nur bei $\text{SNR}'s < 0$ sinnvoll erscheint.

Bei nichtlinearen Filtern muß auf die Wahl von korrekten Startwerten geachtet werden, da sich diese nicht nur auf die Filterstabilität beim Einschwingen auswirken, sondern auch noch auf die Filtergenauigkeit des eingeschwungenen Filters. Es läßt sich beim Filtertuning eine optimale Einschwingzeit ermitteln, bei der die Filtergenauigkeit noch optimal bleibt und erst bei Verkürzen der Einschwingdauer, der Filterfehler ansteigt.

Parallel zu den theoretischen Untersuchungen wurde eine Entwicklungsumgebung aufgebaut, welche ein brauchbares Entwicklungstool nicht nur für die Online-Demodulation darstellt, sondern sich vielseitig auch für andere Anwendungen beispielsweise im Motorenmanagement (Automobilindustrie) oder Maschinenregelungen einsetzen läßt.

7 Literaturverzeichnis

1. Abramson, N.: *Bandwidth and Spectra of Phase-and-Frequency-Modulated Waves* (IEEE Trans. on Signal Processing 5.June, 1993)
2. Analog Devices: Datenbücher für ADSP 210xx (1995-1996)
3. Anderson, B.D.O.; Moore J.B.: *Optimal filtering* (New Jersey:Prentice Hall, ISBN 0-13-638122-7,1979)
4. Anderson, J. B.; Aulin, T.; Sundberg. C.-E.: *Digital Phase Modulation* (Plenum Press London and New York, ISBN 0-306-42195-X, 1986)
5. Arndt, C.: *Informationsgewinnung und -verarbeitung in nichtlinearen dynamischen Systemen* (Dissertation, Universität-GH Siegen, ISBN 3-8265-2304-0, 1997)
6. Balzer, D.; Loffeld, O.; Arndt, C.: *Demodulation of Phase Modulated Signals using an Expanded Observation Model in Kalman Filters* (PIERS'96, Progress in Electromagnetics Research, Symposium, 8-12 Juli 96, Innsbruck, Austria, Proc. 397ff.)
7. Balzer, D.; Loffeld, O.; Arndt, C.: *C40 online demodulation of phase modulated signals using different kalman filter models* (LASER'97/EUROPTO, 16-17. Juni 97 ,München, Germany Proc. of SPIE: Sensors, Sensor Systems and Sensor Data Processing 175ff.)
8. Balzer, D.; Loffeld, O.; Arndt, C.: *Improving nonlinear Kalman Filters with different adaptive elements*, (IASTED'98, 11-13. Februar 1998, Las Palmas, Gran Canaria/Spanien, Proc. of IASTED International Conference 238ff.)
9. Best, R.: *Theorie und Anwendungen des Phase-locked Loops* (VDE-Verlag, ISBN 3-8007-1980-0, 1993)
10. Boashash, B.; O'Shea, P.; Arnold, M. J.: *Algorithms for instantaneous frequency estimation: A comparative study* (Advanced signal processing algorithms SPIE Vol. 1348, Seiten 126-148, 1990)
11. Böhmer, E.: *Elemente der angewandten Elektronik* (Vieweg-Verlag Braunschweig/Wiesbaden, ISBN 3-528-04090-4, 1998)

12. Brammer, K.: *Kalman-Bucy Filter, Anwendungen der Kalman-Filtertechnik* (Oldenbourg, ISBN 3-486-22779-3, 1994)
13. Brandt, S.: *Datenanalyse* (Mannheim [u.a.] : BI-Wiss.-Verl., ISBN 3-411-03200-6, 1992)
14. Bronstein, I.; Semendjajew, K.; Musiol, G.; Mühlig, H.: *Taschenbuch der Mathematik* (Verlag Harri Deutsch, ISBN 3-8171-2004-4, 1999)
15. Campbell C.: *A bandwidth reduction system für speech transmission* (San Francisco, Thesis Massachusetts Institut of Technology, 1959)
16. Collings, I. B., Moore, J. B.: *Adaptive HMM filters for signals in noisy fading channels* (IEEE 0-7803-1775-0/94, Band III p.305-308, 1994)
17. Cook C . E., Bernfeld, M.: *Radar Signals* (Academic press New York, San Francisco, London, 1967)
18. David K., Benkner T.: *Digitale Mobilfunksysteme* (Teubner-Verlag Stuttgart; ISBN 3-519-06181-3, 1996)
19. Engell, S.: *Entwurf nichtlinearer Regelungen* (Oldenbourg Verlag, München, Wien, ISBN 3-486-23065-4, 1995)
20. Fellbaum, K. : *Elektronische Sprachverarbeitung* (Franzis-Verlag, ISBN 3-7723-6532-9, 1991)
21. Fink, L. M.: *Relation between the Spectrum and Instantaneous frequency of a signal* (Academy Peredachi Informatcii, Vol. 2, No. 4, pp. 26-38, 1966)
22. Fliege, N.: *Systemtheorie* (B.G.Teubner Stuttgart, ISBN 3-519-06140-6, 1991)
23. Gabor, D.: *Theory of communication Parts I-III* (British Thomson Houston Co., Ltd., Research Lab., 24.09.1945)
24. Gelb, A.: *Applied Optimal Estimation* (The Analytic Science Corporation/ M.I.T., Reading, Cambridge, Massachussets, London, 1978)

25. Gleue, S.: *Implementierung einer graphischen Simulationsoberfläche zur Demodulation winkelmodulierter Signale mit dem Extended Kalman Filter unter 'IDL'* (Diplomarbeit PB2-ZESS, Uni-GH-Siegen, 1997)
26. Griffiths, L. J.: *Rapid Measurements of Digital Instantaneous Frequency* (IEEE Trans. on Signal Processing Vol. Assp-23, No. 2, April 1975)
27. Hatzold, P.: *Die neue Modulation der D- und E-Netze* (Funkschau 17/1994, S. 54-58)
28. Hüttenschmidt, E.: *PC-Adapterkarte zur Anbindung eines TMS320C44 Moduls über ein SARAM* (Diplomarbeit PB2-ZESS, Uni-GH-Siegen, 1997)
29. ITU-T: *Artificial Voices; Telephone transmission quality, objective measuring apparatus* (ITU-T Recommendation P.50 von 03/93)
30. Jazwinski A.H.: *Stochastic process and filtering* (Acad. Press New York, ISBN 0-12-381550-9, 1979)
31. Kailath, T. Frost, P.: *Innovations approach to least squares estimation: Part II Linear smoothing additive white noise* (IEEE Trans Aut. Control A. 646-655, 1968)
32. Kernigham, B. W., Ritchie, D. M.: *Programmieren in C* (Carl Hanser Verlag München Wien, ISBN 3-446-13878-1, 1983)
33. Kurz, K., Wagener, W.: *Signalprozessoren-Praxis* (Franzis-Verlag, ISBN 3-7723-6502-7, 1991)
34. Klaus, F.: *Entwicklung, Adaption und Analyse von "Fixed Interval Kalman-Smoothern" für die Doppler-Centroidestimation beim Synthetic Aperture Radar* (Diplomarbeit am INV/Uni-GH-Siegen, 1995)
35. Krebs, V.: *Nichtlineare Filterung* (Oldenbourg Verlag München Wien, ISBN 3-486-24751-4, 1980)
36. Landgraf C.: *Stochastische lineare Systeme, Systembeschreibung und optimale Filterung* (Landgraf Berlin : Techn. Univ., 1979)
37. Laface, P., De Mori, R.: *Speech Recognition and Understanding* (Springer Verlag, ISBN 0-387-54032-6, 1992)

-
38. Loffeld, O.: *Estimationstheorie I* (Oldenbourg Verlag München Wien, ISBN 3-486-21616-3, 1990)
 39. Loffeld, O.: *Estimationstheorie II* (Oldenbourg Verlag München Wien, ISBN 3-486-21627-9, 1990)
 40. Loffeld, O.: *Demodulation of noisy Phase or Frequency modulated Signals with Kalman filters* (Proc. IGARSS, 1991)
 41. Loffeld, O., Sack, A.: *Kalman Filters for the Demodulation of Angularly Modulated Signals* (Proc. ICSP'1990, pp. 631-634, Beijing, China 1990)
 42. Loffeld, O.: *Demodulation of noisy phase or frequency modulated signals with Kalman filters* (Proc. ICASSP94, IV pp. 177-190, Adelaide Australia, 1994)
 43. Loffeld, O.: *Nachrichtenverarbeitung* (Vorlesungsunterlagen 1997, Universität-GH Siegen)
 44. Loffeld, O., Balzer, D., Veitenhansl, S., Weiß, M. (1998,1999): *Allgemeine Nachrichtentechnik* (Vorlesungsunterlagen 1999, Universität-GH Siegen)
 45. Lüke, H. D.: *Signalübertragung* (Springer Verlag, ISBN 3-540-58753-5, 1995)
 46. Mähringer, C.: *Online-I&Q-Demodulation von FM/PM-Signalen mit dem DSP TMX320C40 - Entwicklung einer A/D-Wandlerkarte und Kalman-Filter-Implementierung in 'C'* (Diplomarbeit PB2-ZESS, Uni-GH-Siegen, 1996)
 47. Mardia, K. V.: *The Art of Statistical Science* (John Wiley & Sons, ISBN 0-471-93110-1, 1992)
 48. Maybeck, P. S.: *Stochastic models, estimation and control* (Vol. I/II, Academic Press, New York, 1979 bzw. 1980)
 49. Maybeck, P.S.: *Combined Estimation of States and Parameters for Online Applications*, (Rep. T-557. Dissertation, MIT, Cambridge, Massachussets, 1972)
 50. McGarty, T.P.: *Stochastic Systems and State Estimation* (A Wiley-Interscience publication, ISBN 0-471-58400-2, 1974)

-
51. McWilliams, J.F. und Sloane, N.J.A.: *PseudoRandom Sequences and Arrays* (Proceedings of IEEE Vol.64 No.12, Dec.1976)
 52. Meditch, J.S.: *Stochastic optimal linear estimation and control* (Mc GrawHill, New York, 1969)
 53. Middleton, D.: *Stastical communication theory* (Peninsula Publishing, Los Altos, California, 1987)
 54. Moore, J.B.: *Discrete-Time Fixed Lag Smoothing Algorithms* (Automatica, Vol. 9. pp. 163-173, Pergamon Press, Great Britain 1973)
 55. Neufang, O., Rühl, H.: *Elektronik Wörterbuch* (Zimmermann-Neufang, Ulmen, ISBN 3-922410-15-4, 3. Auflage, 1988)
 56. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., *Numerical Recipes in C: The art of scientific computing* (Cambridge University Press, 2. Edition, ISBN 0-521-43108-5, 1994)
 57. Oppenheim, A. V., Schafer, R. W.: *Zeitdiskrete Signalverarbeitung* (Prentice Hall, Englewood Cliffs, New Jersey 07632, ISBN 3-486-22948-6, 1995)
 58. Olk, J.: *Untersuchung von Laufzeitentfernungsmeßsystemen unter besonderer Berücksichtigung des Referenzproblems* (Dissertation, Universität-GH Siegen, ISBN 3-8265-2940-5, 1997)
 59. Picone, J.: *Signal modeling techniques in speech recognition* (Proc. of IEEE Vol.81, No.9, Sept. 1993)
 60. Poguntke: *Vorlesung: Nachrichtenübermittlung* (Universität-GH Siegen, 1990)
 61. Posthoff, C., Woschni, E.-G.: *Funktionaltransformationen der Informationstechnik* (Akademie-Verlag Berlin, 1984)
 62. Prokott, E.: *Modulation und Demodulation* (Elitera, Berlin, ISBN 3-87087-106-7, 1978)
 63. Rauch H.E.: *Solutions to the linear Kalman filtering problem* (IEEE Transact. Aut. Control AC8, 371, 1963)

-
64. Rao, C. R.: *Was ist Zufall?* (Prentice Hall, ISBN 3-8272-9508-4, 1995)
 65. Reilly, A., Boashash, B.: *Advanced Signal Processing Algorithms* (San Diego 07/92; SPIE Vol.1770, 1992)
 66. Ruske, G.: *Automatische Spracherkennung* (Oldenbourg Verlag, ISBN 3-486-22794-7, 1994)
 67. Sadeghi, A.-R.: *Nichtlineare optimale Estimation* (Diplomarbeit Uni-GH-Siegen, 1993)
 68. Schaffer, G.: *IDL-Simulation, Generation und Bandbreiteuntersuchung unterschiedlicher modulierter WAV-Files zum Test von Online-Kalman Filtern* (Diplomarbeit PB2-ZESS, Uni-GH-Siegen, 1998)
 69. Schetzen, M.: *The Volterra and Wiener theories of nonlinear systems* (John Wiley & Sons, ISBN 0-471-04455-5, 1980)
 70. Schumacher, W.: *Untersuchung adaptiver Kalman-Filter zur Lösung des Radarzielverfolgungsproblems* (Dissertation, TU Berlin, Fachber. 10, Verfahrenstechnik, 1979)
 71. Schweitzer, N.; Klauschke, O.: *Entwicklung eines Demodulationssystems winkelmulierter Signale unter Anbindung des DSP TMS320 C44 zur Implementierung von Online-Kalman-Filtern* (Diplomarbeit, Uni-GH-Siegen, 1996)
 72. Sciabassi, R. J. Sun, M.: *Discrete-Time Instantaneous Frequency and Its Computation* (IEEE Transaction on Signal Processing Vol. 41. No. 5, May 1993)
 73. Sommer, D.: *Simultane Zustands- und Parameterestimation für Remote Sensing Anwendungen mit Extended Linearized Kalman-Filtern* (Diplomarbeit INV/Uni-GH-Siegen, 1995)
 74. Sorenson, H. W.: *Kalman Filtering: Theory and Application* (IEEE Press, ISBN 0-87942-191-6, 1985)
 75. Stadler, E.: *Modulationsverfahren* (Vogel-Buchverlag Würzburg, ISBN 3-8023-1486-7, 1993)
 76. Texas Instruments: *Datenbücher für TMS320 C3x/C4x + Compiler, Linker, Debugger und Simulator* (1991-1997)

77. Tietze, U., Schenk, Ch.: Halbleiterschaltungstechnik (Axel-Springer-Verlag, ISBN 3-540-56184-6, 1993)
78. White, L. B., Ross, J. D.: *Adaptive demodulation of phase modulated signals using HMMs* (IEEE Trans. on Signal Processing, IV,164-167, 1993)
79. Zepernick, H.J.: *Untersuchung und Entwicklung von Kalman-Smoothern zur Kantenestimation in Konturverarbeitungsfiltern* (Diplomarbeit INV/Uni-GH-Siegen, 1987)

Internetadressen mit Kalman-Filter-Anwendungen:

http://members.aol.com/kfcas/index.html	Anwendung von Kalman-Filtern in GPS, Navigation und INS
http://www.fkf.net/FKFware.html	Fast Kalman-Filtering Algorithmen
http://tspi.elan.af.mil/kalman-java.html	Java-Applet: Interaktiver Kalman-Filter
http://cs.unc.edu/~welch/kalmanLinks.html	Links auf Kalman-Filter Web-Seiten

8 Anhang

8.1 VERZEICHNIS HÄUFIG VERWENDETER ABKÜRZUNGEN UND FORMELZEICHEN

A(k)	Zustandsübergangsmatrix F (zeitdiskret)
a_0	Amplitude
AKF	Autokorrelationsfunktion
AM	Amplitudenmodulation
ASK	Amplitude Shift Keying (Amplitudenumtastung)
\tilde{a}	Virtuelle Amplitude
A(f)	Amplitudengang, Amplitudenspektrum
B	Bandbreite
BPSK	Binäres PSK
β	Bandbreitendehnfaktor
$\beta(t, \cdot)$	Brownscher Prozeß
c(k), c(t)	Lin. Beobachtungsmatrix (zeitdiskret, zeitkontinuierlich)
C(k), C(t)	Diff. Beobachtungsmatrix (zeitdiskret, zeitkontinuierlich)
CNR	Carrier to Noise Ratio (Verhältnis von Trägerleistung zur Rauschleistung)
γ	Modulationsindex
$\Delta\varphi$	Phasenhub
Δf	Frequenzhub
EKF_{xyz}	Extended Kalman-Filter; 'x' bezeichnet die Anzahl der Beobachtungen, 'y' bezeichnet die Anzahl der Zustände und 'z' wird für optionale weitere Klassifizierungen eingesetzt
erf(x)	Fehlerfunktion, Errorfunction
$E\{x\}$	Erwartungswert x
F	Fisher's Information
F(t)	Zustandsübergangsmatrix F (zeitkontinuierlich)
FM	Frequenzmodulation
$f_m(t)$	Momentanfrequenz
FSK	Frequency Shift Keying (Frequenzumtastung)
$F_{\underline{x}/\underline{y}}(\underline{\xi}/\underline{\rho})$	Bedingte Verteilungsfunktion
$f_{\underline{x}/\underline{y}}(\underline{\xi}/\underline{\rho})$	Bedingte Verteilungsdichtefunktion

f_{Bit}	Bitfolgefrequenz
f_g	Grenzfrequenz
f_{Δ}	Bandbreite des Nutzsignals
f_0	Trägerfrequenz
G	Gewichtungsmatrix für Rauschprozeß
GMSK	Gaussian Minimum Shift Keying
GRP	Gauß'scher Random (Zufalls-) Prozeß
G_{α}	α-Divergenz nach Shannon
$\varphi_{mm}(\tau)$	Autokorrelationsfunktion von $m(t)$
$\tilde{\varphi}_{mm}(\tau)$	Normierte Autokorrelationsfunktion von $m(t)$
$\varphi_{xx}^L(\tau)$	Autokorrelationsfunktion (Leistungssignal)
$\phi(f)$	Charakteristische Funktion
$\phi_{mm}(f)$	Leistungsdichte von $m(t)$
$\tilde{\phi}_{mm}(f)$	Normierte Leistungsdichte von $m(t)$
H(f)	Übertragungsfunktion
H(x(k)), H(x(t))	partielle Ableitung der Beobachtungsmatrix (zeitdiskret, zeitkontinuierlich)
h(x(k)), h(x(t))	nichtlineare Beobachtungsmatrix (zeitdiskret, zeitkontinuierlich)
I	Informationsgehalt, Informationsmenge
I	Einheitsmatrix
K	Kalman-Gain (- Matrix)
MSK	Minimum Shift Keying
$m(t)$	Nutzsignal
η	Dämpfungsgrad
$n(t)$	Rauschsignal
N	Rauschleistung
N_0	Rauschleistungsdichte
P(x_i)	Auftrittswahrscheinlichkeit eines Zeichens
PLL	Phase Locked Loop (Phasengesteuerter Regelkreis)
PM	Phasenmodulation
PM	Phasenmodulation
P_s	Signalleistung
PSK	Phase Shift Keying (Phasenumtastung)
$P_{\underline{xx}}^-(k)$	Prädiktionsfehlerkovarianz von \underline{x}

$P_{xx}^+(k)$	Filterfehlerkovarianz von \underline{x}
$Q(k)$	Driving Noise Matrix des Kalman-Filters, Varianz des Prozeßrauschens
QPSK	Quaternäres PSK
$R(k)$	Rauschmatrix des Kalman-Filters, Varianz des Meßrauschens
$r_T(t)$	Empfangssignal im äquivalenten Tiefpaßbereich
$r(t), y(t)$	Empfangssignal
SNR	Signal to Noise Ratio (Signalrauschleistungsverhältnis)
$s_T(t)$	Sendesignal im äquivalenten Tiefpaßbereich
$s(t)$	Sendesignal
$s_B(t)$	Bandpaßsignals mit $s_{Bg}(t)$ als gerade und $s_{Bu}(t)$ als ungerade Anteile
S_α	α -Information nach Shannon
σ_x^2	Varianz von x (2.Moment)
μ	Modulationsgrad
$\underline{v}(t), \underline{v}(k)$	Rauschvektor des Meßrauschens
v_1, v_2	unabhängige Zustandsvariablen von Rauschprozesse
VCO	Voltage Controlled Oscillator (Spannungsgesteuerter Oszillator)
$\underline{w}(t), \underline{w}(k)$	Rauschvektor des Driving Noise
WM	Winkelmodulation
ω_0	Kreisfrequenz der Trägerschwingung
ω_j	Elementarereignis j
$\underline{x}(\cdot, \cdot)$	Stochastischer Prozeß
$\underline{x}(\cdot, \omega_j)$	Musterfunktion j eines stochastischen Prozesses
$\hat{\underline{x}}^-$	Prädiktionsschätzwert
$\hat{\underline{x}}^+, \hat{\underline{x}}$	Filterschätzwertvektor
$\underline{x}(t_i, \cdot), \underline{x}(t_i)$	Zufallsvariable eines stochastischen Prozesses
ξ	Dummyvariable für x
\bar{x}	Mittelwert von x (1.Moment)
ρ	Dummyvariable für y

Tab. 8.1 Verzeichnis häufig verwendeter Formelzeichen und Abkürzungen

8.2 SAMMLUNG DER FILTERMODELLE

Mit den folgenden Namen werden die Filter in Software (DSP und Delphi-Umgebung) bezeichnet. Wegen der kürzeren Schreibweise sind diese Filternamen auch häufig im Text zu finden.

Filterbezeichnung	Filtergruppe	Zusatzbemerkung
EKF22	Extended Kalman-Filter	Standardmodell
EKF22s	Standard EKF	geschwindigkeitsoptimierter EKF22
EKF22ts	Standard EKF	zeitkontinuierliches Modell, optimiert
EKF22ss	Standard EKF	Singermodell anstatt Markov, optimiert
EKF22pr	Standard EKF	vorgefilterter EKF22
EKF32pv	Standard EKF	additive Phase virtuell
EKF32av	Standard EKF	additive Amplitude virtuell
EKF42pav	Standard EKF	additive Phase + Amplitude virtuell
EKF24	EKF Amplitudenestimation	4 Zustände: (Amplitude und Amplitudenänderung)
EKF34av	EKF Amplitudenestimation	4 Zustände: (Amplitude und Amplitudenänderung) + virtuelle Amplitude
AKF24	Adaptiver Extended Kalman-Filter	Q wird geschätzt mit 2 Parametern
AKF25	Adaptiver EKF	Q wird geschätzt mit 3 Parametern
AKF22ml	Adaptiver EKF	Q wird per Maximum-Likelihood-Schätzung ermittelt
AKF22ml2	Adaptiver EKF	wie AKF22ml, aber zusätzlich Variationsmöglichkeiten der ML-Schätzung
SKF22fl	Smoother	Fixed Lag / Fixed Intervall Smoother
SKF22fp	Smoother	Fraser-Potter Smoother
SKF22ro	Smoother	Reduced Order

Tab. 8.2 Kalman-Filtermodelle