Einführung in die Nachrichtentechnik
Herausgegeben von Alfons Gottwald

Im Zeitalter der Kommunikation ist die ELEKTRISCHE NACHRICHTENTECHNIK eine vielschichtige Wissenschaft: Ihre rasche Entwicklung und Auffächerung zwingt Studenten, Fachleute und Spezialisten immer wieder, sich erneut mit sehr unterschiedlichen physikalischen Erscheinungen, mathematischen Hilfsmitteln, nachrichtentechnischen Theorien und ihren breiten oder sehr speziellen praktischen Anwendungen zu befassen.

EINFÜHRUNG IN DIE NACHRICHTENTECHNIK ist daher eine ebenso vielfältige Aufgabe. Dieser Vielfalt wollen unsere Autoren gerecht werden: Aus ihrer fachlichen und pädagogischen Erfahrung wollen sie in einer REIHE verschiedenartiger Darstellungen verschiedener Schwierigkeitsgrade EINFÜHRUNG IN DIE NACHRICHTENTECHNIK vermitteln.

Estimations-theorie I

Grundlagen und stochastische Konzepte

von
Privatdozent Dr.-Ing. Otmar Loffeld

Mit 52 Bildern

R. Oldenbourg Verlag München Wien 1990

Titel der Habilitationsschrift
„Grundlagen, Konzepte und Anwendungen der Estimationstheorie"

CIP-Titelaufnahme der Deutschen Bibliothek

Loffeld, Otmar:
Estimationstheorie / von Otmar Loffeld. – München ; Wien :
Oldenbourg
(Einführung in die Nachrichtentechnik)

1. Grundlagen und stochastische Konzpete. – 1990
ISBN 3-486-21616-3

Gesamtherstellung: Huber KG, Dießen

ISBN 3-486-21616-3

Inhalt

Estimationstheorie I: Grundlagen und stochastische Konzepte

Estimationstheorie II: Anwendungen – Kalman–Filter

Vorwort

Die vorliegende, zweibändige Darstellung der Grundlagen, Konzepte und Anwendungen der Estimationstheorie entstand während meiner Lehr– und Forschungstätigkeit am Institut für Nachrichtenverarbeitung der Universität–GH–Siegen. Diese Darstellung stellt eine stoffliche Obermenge zu einer zweisemestrigen Wahlvorlesung dar, die ich seit 1986 für alle Elektrotechniker im Hauptstudium an der Universität Siegen halte.

Die Konzepte der Estimationstheorie sind in der ingenieurwissenschaftlichen Ausbildung, speziell in der elektrischen Nachrichtentechnik, noch verhältnismäßig neu und wenig arriviert. Dies stellt ein krasses Mißverhältnis zu ihrer Bedeutung und auch Leistungsfähigkeit dar. Andererseits sind wichtige Grundlagen der Estimationstheorie in Form von Wahrscheinlichkeitstheorie und Stochastik in der Mathematik sehr wohl vorhanden und sehr gut ausgebaut. Dieser Fundus an Grundlagen ist jedoch, selbst für ausgebildete Ingenieure, häufig nur wenig nutzbar, da der durch die Rigorosität der Darstellung bedingte mathematische Formalismus (etwa die Satz–Beweis–Struktur vieler mathematischer Darstellungen) häufig den Blick auf die technisch nutzbare Anwendung verstellt. Eine andere Schwierigkeit der Anwendung mathematischen Grundlagenwissens besteht zum anderen auch häufig darin, daß die 'Ingenieurkunst' aus einer 'Gratwanderung' zwischen nur unter speziellen mathematischen Randbedingungen gültigen Sätzen und ingenieurwissenschaftlichen, rigoros mathematisch allerdings wenig nachvollziehbaren Abstraktionen der realen Welt (wie etwa dem 'weißen Rauschen') besteht.

Die Grundlagen der Estimationstheorie ergeben sich grob aus vier verschiedenen Gebieten. Zwei dieser Gebiete, die Beschreibung von Systemen im Zustandsraum und die Beschreibung von linearen, zeitinvarianten Systemen mit Hilfe von Übertragungsfunktionen im Laplace–, Fourier– und Z–Bereich sind in der ingenieurwissenschaftlichen Ausbildung sehr wohl vertreten. Jedoch werden diese Gebiete häufig wenig zusammenhängend und noch dazu mit verschiedenartiger Zielsetzung zum einen in der Regelungstechnik und zum anderen in der klassischen Nachrichtentechnik vorgestellt. Die beiden anderen Grundlagenlieferanten sind die Wahrscheinlichkeitstheorie und die Stochastik, die häufig immer noch als mathematische Spezialgebiete gelten. Aus dieser Situation heraus ergab sich die Notwendigkeit, die benötigten Grundlagen in einer zusammenhängenden und einheitlichen Form darzustellen und damit gleichzeitig die intermediäre Denkweise der Estimationstheorie klar zu machen. Die Anwendungen dieser Grundlagen in der Estimationstheorie ergeben sich danach umittelbar einleuchtend.

Die Aufteilung dieser inhaltlichen Gesamtheit auf zwei Teilbände mag auf den ersten Blick vielleicht gegen den Anspruch einer zusammenhängenden Darstellung verstoßen, doch erscheint sie nach längerer Betrachtung durchaus (und nicht nur drucktechnisch) sinnvoll: Band I der Darstellung bringt die Grundlagen der Estimationstheorie in Form von Zustandsraumdarstellungen, Wahrscheinlichkeitstheorie und Stochastik und schafft so die Voraussetzungen zum Verständnis der Anwendungen der Estimationstheorie in Form von Kalman–Filtern, die den Inhalt des zweiten Bandes bilden. Damit bilden beide Bände die angestrebte zusammenhängende Darstellung. Um diesen Zusammenhang weiter zu fördern, besitzen beide Teilbände das Gesamt–Inhaltsverzeichnis, zusätzlich wurde eine durchgehende Seitennummerierung gewählt. Auch verfügen beide Bände über das Gesamtsachwortverzeichnis.

Aber auch jeder Teilband bildet allein schon eine abgeschlossene Einheit. Wegen der ausführlichen Darstellung der Grundlagen der modernen Regelungstechnik, der Wahrscheinlichkeitstheorie und der linearen stochastischen Systemtheorie ist Band I für einen breiten Leserkreis interesant, vom Ingenieurstudenten im Hauptdiplom bis zum Ingenieur mit abgeschlossener Ausbildung. Darüberhinaus bietet Band I neben allen wahrscheinlichkeitstheoretischen Grundlagen durchaus schon Anwendungen der Estimationstheorie in Form der Parameterestimation. Band II wendet sich an solche Leser, die schon über die in Band I präsentierten Grundlagen verfügen, mit diesen Grundlagen werden verschiedene Kalman–Filterformulierungen mit verschiedenen Ansätzen abgeleitet und ausführlich diskutiert. Ein Kapitel über die Anwendung von Kalman–Filtern vervollständigt diesen Band.

Abschließend möchte ich mich für die Ermutigung und das positive Interesse von Herrn Prof. Dr.–Ing. R. Schwarte und Herrn Prof. Dr. rer. nat. H. Rühl bedanken. Besonders danke ich meinen Kollegen Dipl.–Ing. I. Aller, Dipl.–Ing. L. Tran Duc, Dr.–Ing. K. Hartmann, sowie den Herren Cand.–Ing. E. Schubert, Cand.–Ing. U. Steinbrecher und Cand.–Ing. F. Klaus, die mir bei den Programmier–, Zeichen– und redaktionellen Arbeiten eine wichtige Hilfe waren.

Ich möchte diese Arbeit meiner Frau Marita widmen. Sie war von der Entstehung dieser Arbeit in vielerlei Hinsicht am unmittelbarsten betroffen, nicht nur durch die Mithilfe bei den Schreibarbeiten. Ich danke ihr für ihre Geduld und Liebe.

Siegen, im Januar 1990 Otmar Loffeld

1 Einführung

1.1 Beschreibung der Problematik

Jeder, der sich mit der Problematik der Meßwertgewinnung und Verarbeitung beschäftigt hat, weiß, daß die Welt voller 'Unschärfe' und 'Ungenauigkeiten' ist. Alles, was wir sehen, beobachten oder messen, verliert bei eingehender Betrachtung seine Genauigkeit, seine eindeutige Bestimmtheit. Um Beispiele für diese Aussage zu finden, braucht man nicht einmal die Quantenmechanik als Paradebeispiel für eine stochastische Interpretation der realen Welt zu bemühen. Jederman findet diese Unschärfe im täglichen Leben, hat sich an sie gewöhnt und sogar Methoden entwickelt, sie zu verringern. Jeder Gegenstand des täglichen Lebens, ein Tisch, ein Zimmer besitzt Maße, die seine Ausdehnung, sein Gewicht, seine Masse usw. beschreiben. Um diese Maße zu bestimmen, macht man Messungen. Obwohl man weiß, daß die Maße, die man durch Messungen zu bestimmen versucht, Eigenschaften der Dinge beschreiben und damit unveränderlich sein sollten, wundert man sich nicht, daß verschiedene Messungen des gleichen Maßes verschiedene und je nach 'Genauigkeit' des Meßvorganges mehr oder weniger stark differierende Ergebnisse erbringen. Auch die Verwendung noch genauerer Meßgeräte ändert an dieser grundsätzlichen Tatsache nichts – die Differenzen zwischen den einzelnen Messungen werden nur kleiner, sie verschwinden aber nicht.

Die Erklärungen dieser grundlegenden Tatsache variieren, je nach Standpunkt des Erklärenden, von technisch mathematischen Beschreibungen über philosophische Betrachtungen bis hin zu theologischen Begründungen. Die einzelnen Betrachtungsweisen schließen sich weder gegenseitig aus noch ein, die Unterschiede zwischen ihnen entstehen durch die Verwendung unterschiedlicher Beschreibungsformen der gleichen Problematik. Diese Problematik ist dadurch gekennzeichnet, daß das, was wir sehen, beobachten oder messen, nicht das wirkliche 'Sein' der Dinge, sondern nur eine Interpretation dieses Seins darstellt. Wir interpretieren die Wirklichkeit unter Berücksichtigung der durch unsere Sinne vorgegebenen Erfahrungen, die in Form eines 'Weltbildes' vorliegen.

Das Ziel einer technischen Betrachtungsweise, um die es in dieser Darstellung vorwiegend geht, ist nicht, das 'Sein' der Dinge zu bestimmen oder zu erkennen, sondern ein Modell für diese Dinge zu finden, welches ihr 'sichtbares' Verhalten zufriedenstellend beschreibt. Bezüglich einer derartig bescheidenen Zielsetzung kann dann die reale, aber unbekannte Welt durch die 'Modellwelt' ersetzt werden. Das, was wir sehen oder messen, wird in dieser modellmäßigen Beschreibung als mehr oder weniger stark gestörte

Beobachtung der Modell–Realität interpretiert.

Der Ausdruck 'Beobachtung' mag auf den ersten Blick etwas ungewohnt erscheinen, kennzeichnet aber die Beeinflussung des Meßergebnisses durch die Meßanordnung. Die Projektion eines Quaders in Form seines Schattenbildes ist beispielsweise nur bei einer Beleuchtung mit achsenparallelem Licht, welches senkrecht zur Quaderstirnfläche und zur Bildebene einfällt, ein Rechteck, aus dessen Maßen direkt die Maße der Quaderstirnfläche ermittelt werden können. Bei jeder anderen Beleuchtung wirkt das Schattenbild vergrößert oder verzerrt und ermöglicht kaum noch Rückschlüsse auf das Aussehen des Originalgegenstandes. Der Terminus 'Beobachtung' kennzeichnet damit die systematischen Abhängigkeiten des Meßergebnisses von der Meßanordnung und von den bekannten und damit modellmäßig beschreibbaren Abbildungseigenschaften des Meßsystems.

Der Ausdruck 'Störung' kennzeichnet andererseits die Fehlereinflüsse bei der Messung, die in ihrer Art nicht reproduzierbar sind, sich also von Messung zu Messung unterschiedlich auswirken und damit gerade die stochastische Komponente einer Messung beschreiben.

Die Beseitigung von Störungen, die die in den Messungen enthaltene Information verdecken oder überlagern, bei gleichzeitiger Korrektur der systematischen Fehler, hat eine lange Geschichte, die in ihren Anfangsgründen auf die Arbeiten von C.F. Gauß (1795) zur Bestimmung von Planetentrajektorien /1/ zurückgeht. Weitere, mit den Anfängen dieser Fehlerbeseitigung verknüpfte Namen sind Legendre /2/ und in neuerer Zeit vor allen Dingen Kolmogorov /3,4/, Krein /5,6/, Wiener /7/ und Kalman /8/. Die Liste dieser Namen ist bei weitem nicht vollständig und die Auswahl der erwähnten Autoren recht willkürlich. Eine sehr gute und umfassende Übersicht, in der über 300 Originalveröffentlichungen zitiert und zusammenfassend interpretiert werden, stammt von Kailath /9/. Schon sehr früh hat sich der Begriff 'Estimationstheorie' oder Schätztheorie zur Beschreibung des gesamten Problemkreises der Stör– und Fehlerbeseitigung, sowohl im stochastischen als auch im systematischen Sinn etabliert. Die real unbekannten und nur aus gestörten Messungen beobachtbaren, interessierenden Kenngrößen werden 'geschätzt', wobei dieses 'Schätzen' von einem mathematisch fundierten numerischen Algorithmus durchgeführt wird.

1.2 Zielsetzung und Motivation einer stochastischen Interpretation der Realität

Das zentrale Problem der Estimationstheorie ist die Gewinnung von Information aus gestörten und damit unsicheren Meßwerten. Diese Information muß in den vorliegenden Meßwerten gar nicht direkt enthalten sein, sondern kann unter Umständen nur sehr indirekt beobachtbar sein, wie z.B. die Geschwindigkeit eines fahrenden Autos aus den gestörten Entfernungsmeßwerten, die ein auf der Stoßstange des Autos installierter, berührungslos wirkender Entfernungsmesser liefert. Diese Information kann einerseits benötigt werden, um Regelgrößen für die gezielte Veränderung der betrachteten Vorgänge, Abläufe oder Bewegungen zu gewinnen. Andererseits kann die Gewinnung der Information selbst die Hauptaufgabe, etwa in einer nachrichtentechnischen Problemstellung sein, bei der Übertragungsstörungen und Kanalverzerrungen beseitigt werden sollen. Auch in diesen Fällen muß die gewünschte Information nicht direkt in den vorliegenden Meßwerten enthalten sein, sondern kann auch in codierter Form vorliegen. Ein amplituden– oder frequenzmoduliertes, von Störungen überlagertes Empfangssignal wäre ein Beispiel eines derartigen analogen Übertragungsproblems. Ein solches Problem würde nachrichtentechnisch funktional mit Dekodierung oder Demodulation beschrieben, die gleichzeitig geforderte optimale Störunterdrückung würde auf eine optimale Dekodierung oder optimale Demodulation führen.

Alle diese sogenannten Estimationsprobleme besitzen das gemeinsame Kennzeichen, daß die zu gewinnende Information nur mehr oder weniger direkt durch Sensoren gewonnen werden kann, die ihrerseits wieder Meßfehler verursachen. Diese Meßfehler sind zum einen, wie oben angedeutet, zufälliger oder rauschartiger Natur, zum anderen entstehen systematische Fehler durch das dynamische Verhalten der Sensoren selbst.
Ein anderer Problemkreis ist dadurch gekennzeichnet, daß die interessierenden und aus den vorliegenden Meßwerten zu gewinnenden Kenngrößen nur Modellkenngrößen eines physikalischen Modells der realen Welt darstellen. Solche Modelle stellen aber, wie schon zuvor angedeutet, mathematisch und technisch sinnvolle Abstraktionen der realen Welt dar, die notwendigerweise starke Vereinfachungen aufgrund von Näherungen, unbekannten Parametern oder nicht modellierbaren Effekten enthalten. Weiterhin werden reale Systeme, deren Verhalten durch die Modellkenngrößen beschrieben werden soll, nicht nur von den eigenen bekannten Steuer– oder Eingangsgrößen, sondern auch von unbekannten, nach außen nicht direkt in Erscheinung tretenden Störgrößen (z.B. thermisches Eingangsrauschen eines Übertragungssystems oder nicht modellierte Einflußgrößen aus der Systemumgebung) beeinflußt, so daß das reale Systemverhalten von dem modellierten Systemverhalten in nicht vorhersehbarer Weise abweichen kann.

Als Fazit dieser Überlegungen kann festgehalten werden, daß das Ziel der Estimation die Gewinnung von Information oder Systemkenngrößen aus einer störbehafteten Umgebung ist, die dadurch gekennzeichnet ist, daß sich weder das Störverhalten, noch das Verhalten der unbekannten Systemkenngrößen genau vorhersagen und damit deterministisch beschreiben läßt. Damit wird eine stochastische Interpretation der realen Welt erforderlich und eine Informations– oder Meßdatenverarbeitung, die der stochastischen Natur dieser Problematik Rechnung trägt. Die dazu verwendeten Algorithmen heißen Estimationsalgorithmen. Das Grundproblem der Estimation läßt sich damit zu dem in Bild 1.1 dargestellten Blockschaltbild zusammenfassen, welches eine typische Anwendung eines optimalen Estimationsalgorithmus in Form eines sogenannten Kalman–Filters zeigt.

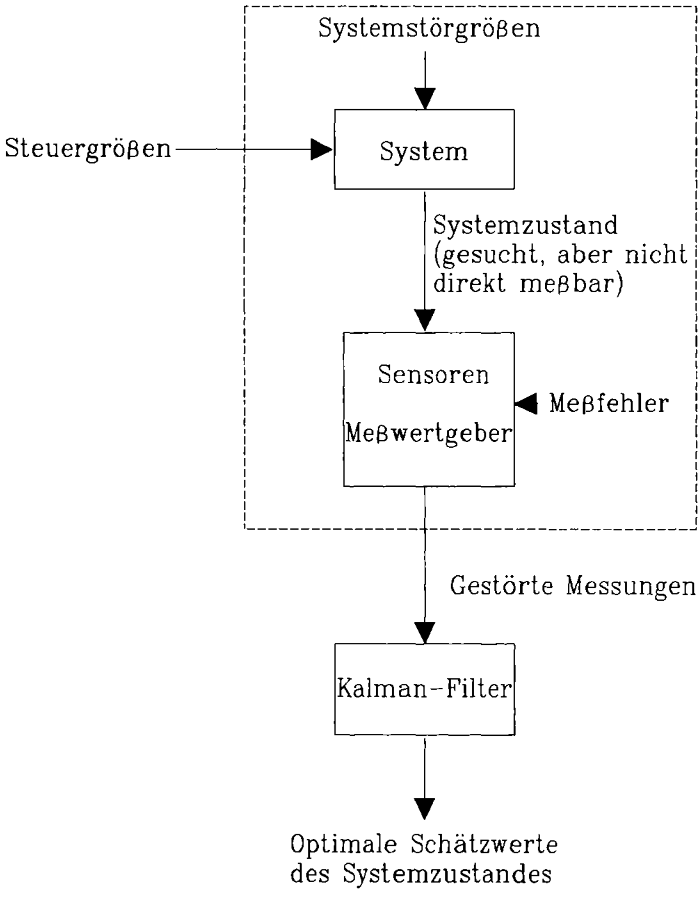

Bild 1.1: Estimationsproblem am Beispiel einer typischen Kalman–Filteranwendung

Unter den zahlreichen Estimationsansätzen wollen wir uns auf lineare, optimale Estimationsalgorithmen beschränken, die im Zustandsraum formuliert werden. Diese Algorithmen werden nach ihrem Entdecker 'Kalman–Filter' genannt und stellen für einen derartig weiten Anwendungsbereich die optimalen Verarbeitungsalgorithmen dar, daß eine Darstellung der Grundlagen der Estimationstheorie, die als Endziel die Ableitung und das Verständnis dieses Algorithmus anstrebt, immer noch hinreichend allgemein ist.

1.3 Optimalität und bedingte Verteilungsdichte, statistisches Konzept eines Filters

Ein statistisches Filter berechnet einen möglichst 'optimalen' Schätzwert einer gewünschten Kenngröße aus einer verrauschten Meßumgebung. Der Ausdruck 'optimal' kennzeichnet dabei die Minimierung eines sinnvollen Fehlerkriteriums, welches je nach Anwendungsfall durchaus unterschiedlich geartet sein kann. Ohne auf die einzelnen unterschiedlichen Fehlerkriterien an dieser Stelle schon eingehen zu wollen, kann man als neutralstes aller möglichen Optimalitätskriterien das Ziel einer optimalen Estimation folgendermaßen formulieren: Das Ziel einer in jeglicher Hinsicht optimalen Estimation wäre die Berechnung der vollständigen bedingten Verteilungsdichtefunktion einer gesuchten Kenngröße, bedingt auf die aktuell vorliegenden Meßwerte. Diese bedingte Verteilungsdichtefunktion enthält alle Information über eine gesuchte Kenngröße, die aus den vorliegenden Meßwerten gewonnen werden kann. Die Information liegt in einer statistischen Formulierung vor; denn die bedingte Verteilungsdichte gibt Auskunft über die Wahrscheinlichkeit eines speziellen Zahlenwertes der gesuchten Kenngröße unter der Voraussetzung, daß die mit dieser Größe zusammenhängenden Meßwerte bestimmte Zahlenwerte angenommen haben. Die Form und das Aussehen dieser bedingten Verteilungsdichtefunktion hängt demzufolge von den Zahlenwerten der gewonnenen Meßwerte ab und gibt Aufschluß über die statistische Zuverlässigkeit, mit der die Aussagen über die Zahlenwerte der gesuchten Kenngröße getroffen werden können:

- Ein ausgeprägtes und schmales Maximum ergibt eine relativ große Wahrscheinlichkeit, mit der die gesuchte Größe die Werte, bei denen die bedingte Verteilungsdichtefunktion maximal wird, annimmt.
- Ein breites, flaches Maximum macht alle Werte, die die gesuchte Größe annimmt, etwa gleichwahrscheinlich und damit Aussagen über diese Werte sehr unsicher.

Basierend auf der bedingten Verteilungsdichtefunktion $f_{x/y_1,y_2,...y_N}$ der gesuchten Größe x, bedingt auf die Meßwerte $y_1...y_N$, kann man verschiedene optimale Schätzwerte definieren:

- Das Maximum der bedingten Verteilungsdichtefunktion (conditional mode), der sogenannte Maximum a posteriori Schätzwert (Map–estimate), beschreibt den Zahlenwert der Größe x, der aufgrund der vorliegenden Meßwerte die höchste Wahrscheinlichkeit besitzt.
- Der bedingte Erwartungswert (conditional mean) beschreibt den Erwartungswert der Größe x aufgrund der vorliegenden Meßwerte und ist das Zentrum der Wahrscheinlichkeits'masse'.
- Der bedingte Medianwert (conditional median) ist ein Schätzwert der Größe x, der dadurch gekennzeichnet ist, daß die Wahrscheinlichkeit, daß der tatsächliche Wert von x größer als dieser Schätzwert ist, genau gleich groß der Wahrscheinlichkeit ist, daß der tatsächliche Wert kleiner als dieser Schätzwert ist.

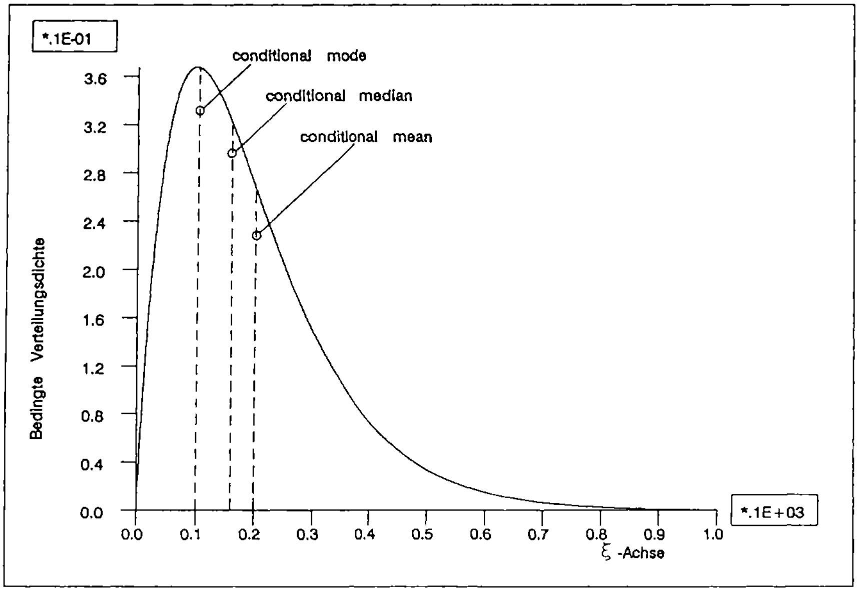

Bild 1.2: Bedingte Verteilungsdichtefunktion einer gesuchten Größe x mit optimalen Schätzwerten

Abbildung 1.2 zeigt eine willkürliche, bedingte Verteilungsdichtefunktion einer gesuchten Größe x, bedingt auf eine Anzahl von Meßwerten $y_1 ... y_N$, mit den eingezeichneten, möglichen optimalen Schätzwerten. Es sind, basierend auf der bedingten Verteilungsdichtefunktion, noch weitere optimale Schätzwertdefinitionen möglich, auf die erst im weiteren Verlauf dieser Darstellung eingegangen werden kann.

Im Falle gaußförmiger, bedingter Verteilungsdichten fallen alle diese optimalen Schätzwerte in einem Punkt zusammen, und der optimale Schätzwert erfüllt gleichzeitig alle drei oben genannten (und auch alle anderen sinnvollen) Optimalitätskriterien.

1.4 Optimalität, Kalman–Filter und Wiener–Filter

Ein Kalman–Filter ist ein optimaler, rekursiver Datenverarbeitungsalgorithmus, der von einem Digitalrechner ausgeführt wird. Es ist kein klassisches 'Filter' im Sinne einer 'Blackbox', welche das Frequenzspektrum eines Signales verändert. Seine Eigenschaften sind dadurch gekennzeichnet, daß es für eine breite Klasse von Problemen das optimale Filter ohne jegliche Einschränkung ist und trotzdem eine lineare Struktur besitzt. Für eine noch weitere Klasse von Anwendungen stellt es immerhin noch das optimale lineare Filter dar. Es verwendet alle Meßwerte entsprechend ihrer Genauigkeit, um die gesuchte Größe zu schätzen, und es benötigt dazu gewisse, sogenannte a–priori–Kenntnisse über das dynamische Verhalten des Systems und der Meßwertgeber, über das statistische Verhalten der Meßstörungen und der unbekannten Systemstörgrößen sowie alle verfügbaren Kenntnisse über Anfangswerte. (Die Verwendung von Anfangswerten und Startwerten stellt einen Unterschied speziell zu den Max. Likelihood Algorithmen dar)
Diese benötigten Modell–Vorkenntnisse werden im Zustandsraum mit Hilfe von linearen Systemmodellen, die mit weißem, gaußverteiltem Rauschen (driving noise) und deterministischen Eingangsgrößen 'getrieben' werden, spezifiziert. Die auftretenden Meßstörungen werden ebenfalls durch Zustandsraummodelle, die von weißem, gaußverteiltem Rauschen getrieben werden, modelliert. Zur Kennzeichnung der statistischen Eigenschaften der verwendeten Rauschprozesse reichen dann die ersten beiden Momente, Erwartungswert und Kovarianz, vollkommen aus. Diese Art der Modellierung ist in der Nachrichtentechnik relativ neu, nicht jedoch in der Regelungstechnik, in der Zustandsraumkonzepte im wesentlichen durch R. E. Kalman schon etwa 1960 eingeführt worden sind. Der klassische nachrichtentechnische Ansatz geht von Filterkonzepten aus, die im Frequenzbereich, etwa unter Zuhilfenahme der Fourier– oder Laplace–Transformation, beschrieben werden. In diesem Zusammenhang werden stochastische Prozesse durch ihr Leistungsdichtespektrum beschrieben und Übertragungssysteme durch ihre Übertragungsfunktion. Diese Ansätze führen in Verbindung mit der Minimierung eines quadratischen Fehlerkriteriums auf das bekannte 'Wiener–Filter'. Der Ansatz besitzt jedoch einige

Nachteile, die sich gerade bei praktischen Problemen sehr schwerwiegend auswirken können. Frequenzbereichsmodelle setzen implizit eine Stationarität der betrachteten Probleme voraus. Viele praktische Probleme sind jedoch nicht von stationärer Natur, besitzen demzufolge kein Leistungsdichtespektrum und können mit einem Frequenzbereichsansatz nicht erfaßt werden. Aufgrund des stationären Ansatzes besitzen Wiener–Filter eine zeitinvariante Filterstruktur und erfüllen damit die Optimalitätsanforderungen erst nach Ablauf der Einschwingphase. Während dieser Einschwingphase, deren Dauer ein festes Filterkennzeichen ist, ist das Filter aber nichtoptimal. Kalman–Filter sind dagegen zeitvariant, mit solchen Filtern ist es möglich, gleichzeitig das Einschwingverhalten und die stationäre Filtergenauigkeit zu optimieren. Das Wiener–Filter wird im Frequenzbereich durch seine optimale Übertragungsfunktion beschrieben. Die Realisierung dieser Übertragungsfunktion führt nur in wenigen Fällen auf einfach realisierbare Schaltungen. Kalman–Filter werden dagegen im Zustandsraum durch rekursive Gleichungen beschrieben und sind in ihrer zeitdiskreten Form direkt auf Digitalrechnern zu implementieren. Aufgrund der Beschreibung im Zustandsraum können auch Probleme mit mehreren Eingängen und Ausgängen sehr einfach behandelt werden, eine Tatsache, die auch in der modernen Nachrichtentechnik zunehmend an Bedeutung gewinnt. Wiener–Filter besitzen aus diesen Gründen für praktische Problemstellungen nur noch geringe Bedeutung und werden deshalb in den folgenden Kapiteln dieser Arbeit nicht weiter behandelt. Trotzdem soll ihre geschichtliche Bedeutung, vor allem auch für die Entwicklung der mathematischen Stochastik, nicht unterschlagen werden. Nach Ansicht des Autors stellen die Arbeiten von N. Wiener in Verbindung mit dem Grundlagenwerk von A.N. Kolmogorov die wesentlichen mathematischen Schritte im Übergang von 'Gauß zu Kalman' dar. Der interessierte Leser wird hierzu auf die spezielle mathematische Grundlagenliteratur verwiesen, in der allein das gesamte Lebenswerk von N. Wiener in 4 Sammelbänden /9,10,11,12/ zusammengefaßt ist, von denen Band 3 und /7/ die wesentlichen estimationstheoretischen Arbeiten enthalten.

1.5 Grundlegende Annahmen und ihre physikalische Bedeutung

Wir wollen uns in dieser Darstellung auf lineare Estimationsprobleme in Form linearer Systemmodelle beschränken und argumentieren dazu, wie folgt:

- Lineare Systemmodelle reichen häufig zur Beschreibung eines Problems aus
- In Fällen nichtlinearen Verhaltens kann oft um einen 'Arbeitspunkt' linearisiert werden
- Die lineare Systemtheorie ist einfacher handhabbar als die nichtlineare Systemtheorie

Wir verwenden, wann immer möglich, die ingenieurwissenschaftliche Abstraktion weißer Rauschprozesse. Diese Prozesse sind folgendermaßen charakterisiert: Die Zahlenwerte eines derartigen Rauschprozesses sind 'Realisationen' von unabhängigen Zufallsvariablen.

- Sie sind zeitlich unkorreliert, das heißt, die Kenntnisse über die Vergangenheit eines Rauschprozesses geben keinerlei Auskunft über sein gegenwärtiges und zukünftiges Verhalten.
- Das Leistungsdichtespektrum von weißem Rauschen ist flach, das heißt konstant für alle Frequenzen. Da dies aber eine unendliche Leistung eines solchen Prozesses bedingen würde, muß festgehalten werden:
- Weißes Rauschen existiert in der Realität nicht!

Die Argumentation für die Anwendung eines derartigen Konzeptes ist ingenieurwissenschaftlich:

- Jedes physikalische System hat Bandpaßeigenschaften – es reagiert nur auf Signale mit Frequenzen innerhalb eines bestimmten Bereiches. Frequenzanteile außerhalb dieses Bereiches beeinflussen das Systemverhalten nicht – deshalb kann eine Erweiterung von breitbandigem zu weißem Rauschen das Systemverhalten nicht beeinflussen, aber die benötigte Mathematik wird durch diesen 'Kunstgriff' stark vereinfacht.(Tatsächlich wird sie erst hierdurch handhabbar).
- Zeitkorrelierte Rauschprozesse (farbige Leistungsdichtespektren) werden dann durch 'Formfilter' modelliert, die von weißem Rauschen getrieben werden.

Die ingenieurwissenschaftliche Annäherung eines breitbandigen Rauschprozesses durch einen weißen Rauschprozeß ist in Abbildung 1.3 dargestellt.

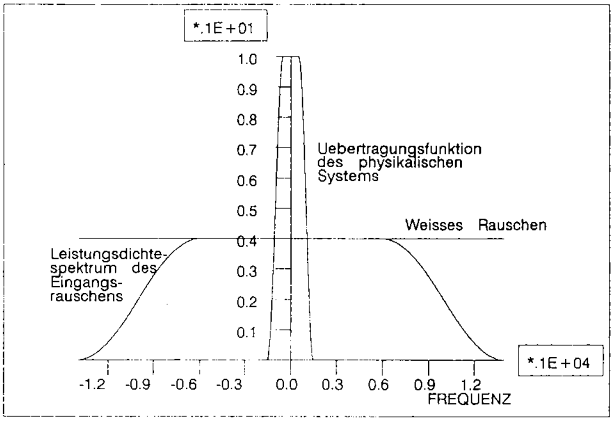

Bild 1.3: Annäherung eines farbigen Rauschprozesses durch einen weißen Rauschprozeß

Als dritte grundlegende Annahme verlangen wir, daß die betrachteten Rauschprozesse nicht nur weiß, sondern Gaußprozesse sind. Dies bedeutet, daß die Wahrscheinlichkeitsdichteverteilung dieser Prozesse zu jedem Zeitpunkt eine gauß'sche Glockenform annimmt. Die Argumentation hierbei ist ingenieurwissenschaftlich:

- Fast alle physikalischen Rauschprozesse werden von einer sehr großen Anzahl verschiedener und unabhängiger Rauschquellen mit infinitesimal kleinen Rauschbeiträgen verursacht – als eine Konsequenz des zentralen Grenzwertsatzes (s. Kapitel 3) tendiert die Verteilungsfunktion einer derartigen Überlagerung von Rauschprozessen gegen eine Gaußverteilung, auch wenn die Anzahl der überlagerten Quellen endlich bleibt.
- Gaußprozesse vereinfachen die Mathematik beträchtlich.
- Sehr oft sind in der Praxis nur die ersten beiden Momente (Erwartungswert (mean) und Kovarianz (covariance)) bekannt. Solange man keine weiteren Kenntnisse höherer Momente besitzt, gibt es keine bessere Annahme als eine Gaußverteilung, die durch die Angabe von Erwartungswert und Kovarianz vollständig beschrieben wird.
- Im Falle von gaußverteilten, bedingten Verteilungsdichten gilt: Ein Kalman–Filter, welches Erwartungswert und Kovarianz der bedingten Verteilungsdichte einer Variablen fortlaufend berechnet, liefert in diesem Falle alle Informationen über die zu schätzende Variable und nicht nur einen Teil aller möglichen Informationen.

1.6 Ein einfaches Beispiel zur Anwendung eines Estimationsalgorithmus

Wir betrachten ein Problem der berührungslosen Entfernungsmessung mit Hilfe eines Laserradars. Ohne auf das Meßverfahren an dieser Stelle weiter eingehen zu wollen, werden wir annehmen, daß es mit Hilfe eines derartigen, auf der Stoßstange eines fahrenden Autos montierten Meßgerätes möglich ist, mehr oder weniger fehlerbehaftete Entfernungsmessungen zu Hindernissen, auf die sich das Auto zubewegt, zu gewinnen.
Das Ziel einer nachgeschalteten Meßwertverarbeitung ist die Genauigkeitsverbesserung der Meßwerte. Wir nehmen dazu an, daß das Meßgerät zum Zeitpunkt t_1 den Entfernungsmeßwert $y(t_1)=y_1$ liefert. Aufgrund einer implementierten Selbstüberwachung liefert das Meßgerät zusätzlich zum Meßwert y_1 eine Genauigkeitsaussage in Form der Standardabweichung σ_{y1} oder der Varianz σ^2_{y1}, die angibt, wie stark die Zahlenwerte der aktuellen Messung um den wahren, fehlerfreien Meßwert schwanken. Damit besitzen wir gleichzeitig ein Maß für die Unsicherheit des Meßwertes y_1.

Statistisch betrachtet kann der Vorgang des Messens damit mit einer Zufallsvariablen beschrieben werden, deren Zahlenwerte die aktuell vorliegenden Meßwerte sind. Die Schwankung dieser Zahlenwerte um den nominal richtigen Wert kann dann durch die Wahrscheinlichkeitsverteilungsdichte der Meßzufallsvariablen beschrieben werden, für die wir eine Gaußverteilung annehmen, deren Standardabweichung σ_{y1}, bzw. Varianz σ^2_{y1} ein Maß für die Exaktheit der Messung darstellt.

Damit liegt aber auch zum Zeitpunkt t_1 eine Verteilungsdichte der aktuellen Entfernung $x(t_1)$ vor, bedingt auf die Messung $y(t_1)= y_1$.

<u>Was sagt diese bedingte Verteilungsdichte aus?</u>

Diese Verteilungsdichte gibt die Wahrscheinlichkeit an, mit der sich das Fahrzeug zum Zeitpunkt t_1, zu dem die Messung anfällt, an dem Ort $x(t_1) = \xi$ befindet, unter der Bedingung, daß die Messung den Zahlenwert $y(t_1) = y_1$ ergibt. σ_{y1} ist ein direktes Maß der Unsicherheit der Entfernung – je größer, desto ungenauer ist eine auf der Messung beruhende Entfernungsaussage. Diese bedingte Verteilungsdichtefunktion ist in Abbildung 1.4 dargestellt. Basierend auf dieser bedingten Verteilungsdichtefunktion ist ein guter Schätzwert der Entfernung sicherlich der bedingte Erwartungswert:

$$\hat{x}(t_1) = E\{x(t_1)/y(t_1)=y_1\} = y_1$$

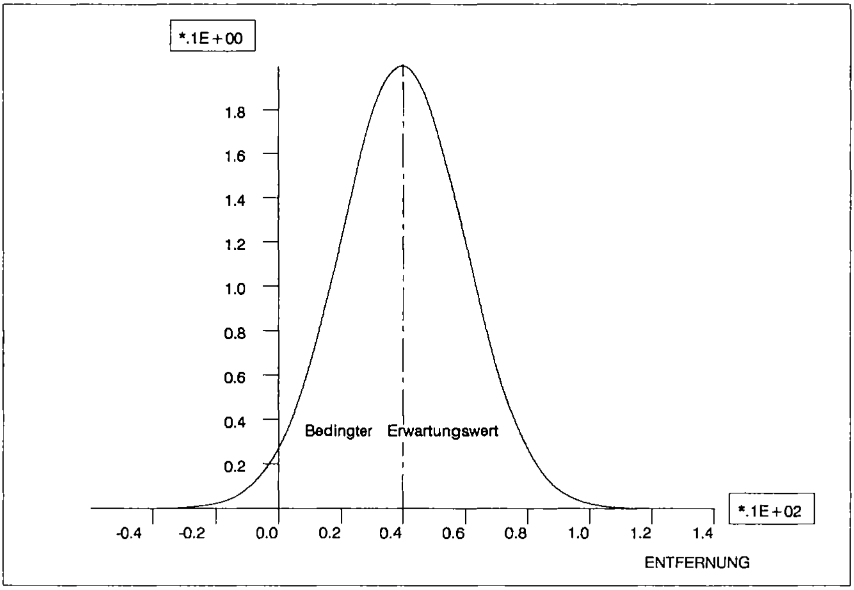

Bild 1.4: Bedingte Verteilungsdichte der Entfernung nach der ersten Messung

Die Varianz dieses Schätzwertes ist:

$$\sigma^2_{\hat{x}}(t_1) = \sigma^2_{y1}$$

Zum Zeitpunkt t_2 fällt nun eine weitere Messung $y(t_2)=y_2$ an, und wir wollen zunächst annehmen, daß die Differenz zwischen den beiden Meßzeitpunkten t_1 und t_2 so gering ist, daß sich die zu bestimmende Entfernung zum Hindernis aufgrund der Eigenbewegung des Fahrzeugs noch nicht nennenswert geändert hat. Die Selbstüberwachung des Meßgerätes kündigt einen relativ exakten Meßwert an, der durch eine Standardabweichung σ_{y2} charakterisiert wird, wobei $\sigma_{y2} \leq \sigma_{y1}$ gelten soll. Das Ergebnis dieser Messung allein wäre eine zweite bedingte Verteilungsdichtefunktion für die aktuelle Entfernung, bedingt auf diese zweite Messung alleine. Diese bedingte Dichte ist in Abbildung 1.5 dargestellt.

Die zweite, genauere Messung soll nun jedoch nicht für sich alleine betrachtet werden, sondern soll dazu verwendet werden, die Genauigkeit der Entfernungsschätzung nach der ersten Messung zu verbessern. Damit suchen wir nach einer optimalen Kombination der beiden Messungen zu einem optimalen Schätzwert der Entfernung, der dann genauer als jede der beiden Einzelmessungen sein müßte.

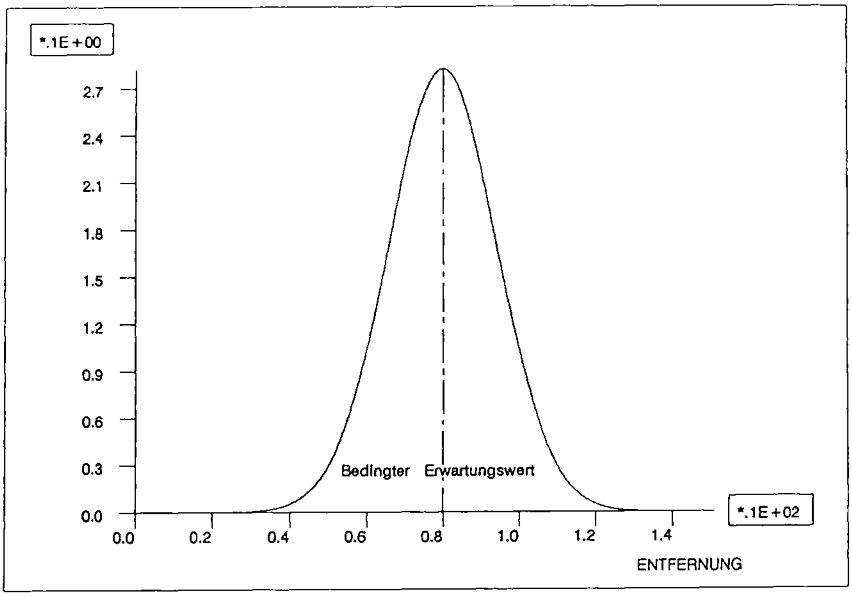

Bild 1.5: Bedingte Verteilungsdichte der Entfernung, bedingt auf die zweite Messung

Bei dieser Suche lassen wir uns von folgender Überlegung leiten:

Eine derartige Kombination muß logischerweise entsprechend der Genauigkeit der jeweiligen Messung durchgeführt werden, das heißt, die zweite, relativ genaue Messung in diesem Beispiel muß stark gewichtet werden, während die erste, relativ ungenaue Messung in einer Kombination nicht überbewertet werden darf.

Es kann mathematisch (s. Kap. 3) gezeigt werden, daß, basierend auf den gemachten Annahmen, auch die bedingte Wahrscheinlichkeitsverteilungsdichtefunktion der Entfernung, bedingt auf beide Messungen, wieder gaußförmig ist und durch die Angabe von bedingtem Erwartungswert μ_x und bedingter Kovarianz $\sigma_x(t_2)^2$ beschrieben werden kann. Diese beiden Kenngrößen werden dann wie folgt berechnet:

$$\mu_x = \frac{\sigma_{y2}^2}{\sigma_{y1}^2 + \sigma_{y2}^2} \cdot y_1 + \frac{\sigma_{y1}^2}{\sigma_{y1}^2 + \sigma_{y2}^2} \cdot y_2$$

und:

$$\frac{1}{\sigma_x^{*2}(t_2)} = \frac{1}{\sigma_{y1}^2} + \frac{1}{\sigma_{y2}^2}$$

Man weist leicht nach, daß die beiden Messungen im umgekehrten Verhältnis zur jeweiligen Meßfehlerkovarianz gewichtet werden und daß die Varianz der Linearkombination von beiden Messungen $\sigma_x^{*2}(t_2)$ kleiner ist als jede der beiden Einzelvarianzen.

Die bedingte Verteilungsdichte der Entfernung, bedingt auf beide Messungen, ist in Abbildung 1.6 dargestellt.

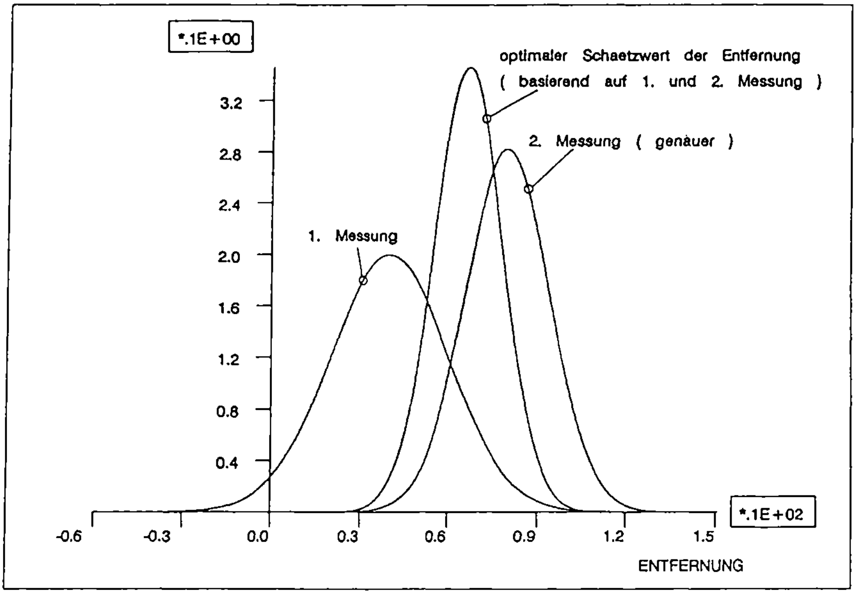

Bild 1.6: Bedingte Verteilungsdichtefunktion der Entfernung, basierend auf beiden Messungen

Hat man nun die bedingte Verteilungsdichtefunktion der Entfernung, bedingt auf beide Messungen, berechnet, bietet sich als bester neuer Schätzwert für die Entfernung der bedingte Erwartungswert an, das heißt:

$$\hat{x}(t_2) = \mu_x$$

Die Varianz dieses Schätzwertes wäre dann: $\sigma_{\hat{x}}^2(t_2)$

Um zu einem rekursiv strukturierten Algorithmus zu gelangen, sollen die Schätzwertgleichungen etwas umgeschrieben werden, und zwar streben wir eine Gleichungsstruktur an, bei der der neue Schätzwert durch Korrektur des vorangegangenen Schätzwertes entsteht (Prädiktor–Korrektor Struktur). Durch Einführung einer Gewichtungsmatrix $K(t_2)$ und durch Ausnutzen der Tatsache, daß $y(t_1) = \hat{x}(t_1)$ gilt, erhalten wir dann aus der vorangegangenen Schätzwertgleichung:

$$\hat{x}(t_2) = y(t_1) + K(t_2) \cdot [y(t_2) - y(t_1)]$$

$$= \hat{x}(t_1) + K(t_2) \cdot [y(t_2) - \hat{x}(t_1)]$$

wobei:

$$K(t_2) = \frac{\sigma_{y1}^2}{\sigma_{y1}^2 + \sigma_{y2}^2} = \frac{\sigma_{\hat{x}}^2(t_1)}{\sigma_{\hat{x}}^2(t_1) + \sigma_{y2}^2} \qquad \text{Kalmangain}$$

und:

$$\sigma_{\hat{x}}^2(t_2) = \sigma_{\hat{x}}^2(t_1) - K(t_2) \cdot \sigma_{\hat{x}}^2(t_1) \qquad \text{Varianz des Schätzwertes}$$

Dies sind die Kalman–Filter Gleichungen für stationäre Probleme.

Interpretation:

Die Kalman–Filter Gleichungen in der Prädiktor–Korrektor–Struktur sind offensichtlich sehr sinnvoll: Jeder neue Schätzwert wird als Linearkombination des letzten Schätzwertes und eines Korrekturtermes berechnet. Dabei stellt der letzte Schätzwert eine Voraussage (Prädiktion) für den neuen Schätzwert dar. Die Differenz zwischen dieser Voraussage und der aktuellen Messung wird mit dem Kalmangain gewichtet und dient dann zur Korrektur der Voraussage. Die Gewichtung der Differenz in Form des Kalmangains $K(t)$ bestimmt sich aus dem Verhältnis von der Voraussagefehlervarianz zur Summe von Voraussagefehlervarianz zuzüglich der Meßstörvarianz. So ist diese Gewichtung $K(t)$ hoch, wenn der Meßwert sehr exakt und die Prädiktion sehr unsicher ist – dies sorgt für eine starke Gewichtung der Korrektur. Andererseits wird die Korrektur sehr gering gewichtet, wenn die Messung selbst sehr unsicher, die Voraussage aber sehr zuverlässig ist.

Zusätzliche Berücksichtigung unbekannter Fahrzeugbewegungen

Wir nehmen nun an, daß das Fahrzeug sich eine ganze Weile weiterbewegt hat, bevor zum Zeitpunkt t_3 eine weitere Messung gemacht werden kann. Die einzigen Kenntnisse, die über die Fahrzeugbewegung vorliegen, sind:

$$\dot{x}(t) = \frac{d}{dt} x(t) = u + w$$

Hierbei stellt u eine mittlere, bekannte, sozusagen nominale Fahrzeuggeschwindigkeit dar. Die Abweichung der tatsächlichen Fahrzeuggeschwindigkeit von dieser konstanten Nominalgeschwindigkeit wird durch eine gaußverteilte Zufallsvariable w mit einem Erwartungswert von Null und einer bekannten Varianz σ_w^2 beschrieben. Dies bedeutet, wir modellieren die Fahrzeuggeschwindigkeit als konstant und nur im Mittel bekannt – die wahre Geschwindigkeit kann von der bekannten Geschwindigkeit um den Zahlenwert der Zufallsvariablen w abweichen.

Die Veränderung der bedingten Verteilungsdichtefunktion der Entfernung von Zeitpunkt t_2 bis zum Zeitpunkt t_3, aufgrund der nur näherungsweise bekannten Eigenbewegung des Fahrzeugs, ist in Bild 1.7 dargestellt.

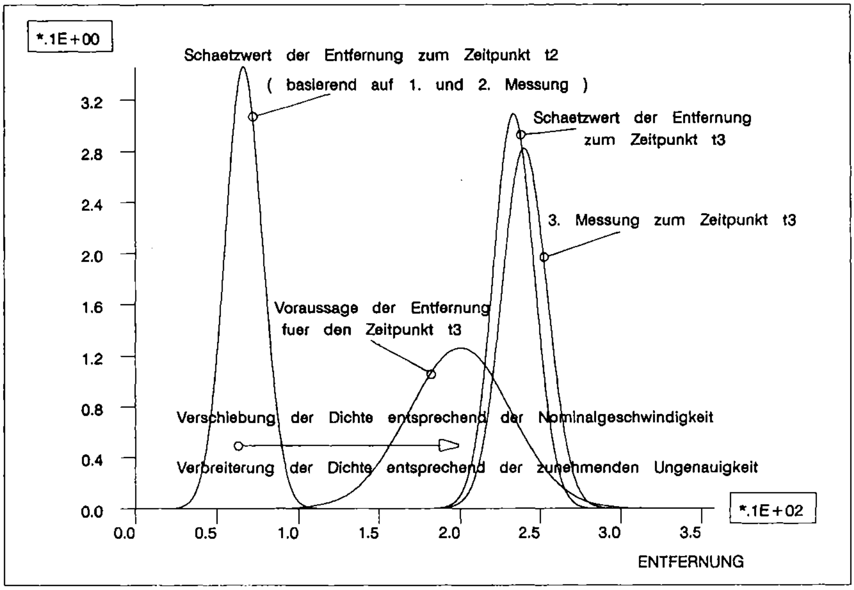

Bild 1.7: Darstellung der verschiedenen bedingten Dichten

Für den Zeitpunkt t_2 wurde die bedingte Verteilungsdichte schon zuvor berechnet. Mit fortschreitender Zeit verschiebt sich die bedingte Dichte mit der nominalen Geschwindigkeit u entlang der ξ–Achse. Gleichzeitig dehnt sich die bedingte Verteilungsdichte symmetrisch zum Erwartungswert aus. Dies entspricht der zunehmenden Unsicherheit aufgrund der unbekannten Realgeschwindigkeit, die sich mit fortschreitender Zeit akkumuliert.

Damit startet die Dichte zum Zeitpunkt t_2 mit dem besten Entfernungsschätzwert zu dieser Zeit und bewegt sich dann mit der Nominalgeschwindigkeit u entlang der ξ–Achse fort, entsprechend der modellmäßig bekannten Dynamik des bewegten Fahrzeugs. Die fortschreitende Ausdehnung der bedingten Dichte bei gleichzeitiger Verflachung deutet an, daß die Unsicherheit der Entfernung mit fortschreitender Zeit nach der letzten Messung immer größer wird.

Bevor zum Zeitpunkt t_3 ein neuer Meßwert anfällt, liegen aufgrund der angesammelten Unsicherheit seit der letzten Messung nur noch vage Kenntnisse über die aktuelle Entfernung vor. Diese Vorkenntnisse lassen sich aus dem letzten Schätzwert zum Zeitpunkt t_2 ableiten, zu dem man den Wegbeitrag addiert, der sich aus der Nominalgeschwindigkeit u, multipliziert mit der Zeitdifferenz t_3–t_2, ergibt. Diese Voraussage (Prädiktion) der Entfernung lautet:

$$\hat{x}^{-}(t_3) = \hat{x}(t_2) + u \cdot [t_3 - t_2]$$

Bestmögliche Voraussage des Ortes

Die Varianz dieser Voraussage ist natürlich größer als die letzte Schätzfehlervarianz zum Zeitpunkt t_2, entsprechend der angesammelten Unsicherheit:

$$\sigma_{\hat{x}}(t_3^-)^2 = \sigma_{\hat{x}}(t_2)^2 + \sigma_w^2 \cdot [t_3 - t_2]$$

Prädiktionsvarianz

Nun fällt der dritte Meßwert $y(t_3) = y_3$ mit der Genauigkeit σ_{y3}^2 an. Wiederum liegen also Vorkenntnisse des Aufenthaltsortes, beschrieben durch die Prädiktionsdichte, vor und eine neue, dritte Messung. Die hierin enthaltene Information könnte wiederum durch eine bedingte Verteilungsdichte der Entfernung, bedingt auf die dritte Messung allein, dargestellt werden, analog zu den Überlegungen vor der Verarbeitung des 2. Meßwertes.

Gesucht wird nun wiederum eine optimale Kombination von Vorkenntnissen und neuer Messung. Für diese optimale Kombination von zwei bedingten Verteilungsdichten gilt völlig analog zu den Überlegungen bei der Verarbeitung des 2. Meßwertes:

$$\hat{x}(t_3) = \hat{x}^-(t_3) + K(t_3) \cdot [y_3 - \hat{x}^-(t_3)]$$

Bester Schätzwert, bedingter Erwartungswert, basierend auf den Messungen y_1, y_2 und y_3.

$$K(t_3) = \frac{\sigma_{\hat{x}}(t_3^-)^2}{\sigma_{y3}^2 + \sigma_{\hat{x}}(t_3^-)^2} \qquad \text{Kalmangain}$$

Die Varianz des neuen Schätzwertes ist:

$$\sigma_{\hat{x}}(t_3)^2 = \sigma_{\hat{x}}(t_3^-)^2 \cdot [1 - K(t_3)]$$

Damit ist die bedingte Verteilungsdichte $f_{x(t_3)/y(t_1),y(t_2),y(t_3)}$ mit der Angabe von bedingtem Erwartungswert $\hat{x}(t_3)$ und der bedingten Kovarianz vollständig beschrieben. Nun kann die gesamte Rekursion mit der Berechnung eines neuen Voraussageschätzwertes für den Zeitpunkt t_4, der Berechnung der Prädiktionsvarianz und der daraus folgenden Korrekturgewichtung $K(t_4)$, sowie der anschließenden Verarbeitung des 4. Meßwertes erneut starten. Das Ergebnis dieser Rekursion ist dann der 4. Schätzwert $\hat{x}(t_4)$ usw...

Damit sind die Kalman–Filter Gleichungen für ein eindimensionales Problem in einer zwar heuristischen, aber intuitiv logisch erscheinenden Form hergeleitet worden.

1.7 Ziele und Aufbau der Darstellung

Anhand der vorangegangenen Ausführungen sind die einzelnen Schwerpunkte klar zu erkennen, mit denen sich die folgenden Kapitel dieser Darstellung beschäftigen werden. Kapitel 2 wird sich zunächst der Beschreibung deterministischer Systemmodelle im Zustandsraum widmen. Diese Darstellung stellt die Grundlagen zur Formulierung linearer Systemmodelle mit mehreren Ein– und Ausgängen, die beliebig zeitvariant sein können, bereit. Die Ausführungen dieses Kapitels können und sollen das Studium der entsprechenden Grundlagenliteratur nicht ersetzen, aber wegen der Wichtigkeit des Verständnisses von Zustandsraummodellen für die spätere stochastische Modellbildung und für die Estimationstheorie ist dieses Kapitel jedoch ausführlicher gehalten, als man es von einem reinen Übersichtskapitel erwarten würde.

Kapitel 3 beschäftigt sich mit den wahrscheinlichkeitstheoretischen Grundlagen der Estimation. Das Ziel dieses Kapitels ist die Einführung von Wahrscheinlichkeitskonzepten in einer verständlichen, aber trotzdem (hoffentlich) nicht oberflächlichen Form. Dabei wurde der Tatsache Rechnung getragen, daß nach Kenntnis des Autors gerade zu diesem, für die Estimationstheorie äußerst wichtigen Gebiet kaum deutschsprachige, zugleich grundlagenorientierte und dem Ingenieurverständnis zugängliche Literatur existiert. Ein weiteres Anliegen dieses Kapitels ist die Formulierung statischer Zufallsmodelle zur Beschreibung statistischer Estimationsprobleme. Praktisches Endziel dieses Kapitels ist die Ableitung eines rekursiven, statischen Estimationsalgorithmus. Anhand der Optimalitätsuntersuchung dieses Algorithmus werden die verschiedenen Optimalitätskriterien diskutiert und miteinander verglichen. Daran schließt sich ein eigener Abschnitt an, der dem Konzept orthogonaler Projektionen und ihrer Bedeutung für die Estimation mit minimalem quadratischen Fehler gewidmet ist. Gerade durch das Konzept orthogonaler Projektionen wird dem Ingenieur ein Werkzeug an die Hand gegeben, optimale Estimationsalgorithmen zu formulieren, ohne sich mit den wahrscheinlichkeitstheoretischen Grundlagen beschäftigen zu müssen. Dies ist eine wichtige Vereinfachung der Ingenieurpraxis, stellt aber auch zugleich eine gewisse Gefahr dar, deshalb sollten zumindest einige Grundlagen dieses Konzeptes auch in einer ingenieurmäßigen Darstellung nicht fehlen.

Kapitel 4 schließlich erweitert die wahrscheinlichkeitstheoretischen Grundlagen von Kapitel 3 auf stochastische Prozesse, die zur modellmäßigen Beschreibung linearer, stochastischer Systemmodelle unbedingt benötigt werden. Dabei wird, ähnlich dem Vorgehen in Kapitel 2, zunächst von kontinuierlichen stochastischen Prozessen ausgegangen. Zur Beschreibung dieser Prozesse werden als Basiskonzepte Brown'sche Prozesse, stochastische Integrale und stochastische Differentiale eingeführt. Mit diesen Konzepten können schließlich lineare, stochastische Differentialgleichungen interpretiert und gelöst werden. Analog zum Vorgehen in Kapitel 2 können mit diesen Lösungen dann stochastisch äquivalente, zeitdiskrete stochastische Modelle zu den gegebenen stochastischen Differentialgleichungen abgleitet werden. Bei diesen Betrachtungen werden auch kontinuierliche, weiße Rauschprozesse als ingenieurwissenschaftliche Vereinfachung der kontinuierlichen Stochastik eingeführt und benutzt. Hauptziel dieses Kapitels ist neben der Bereitstellung der Grundlagen stochastischer Prozesse die ingenieurwissenschaftlich motivierte, aber trotzdem mathematisch fundierte Modellbildung linearer, dynamischer

Systeme mit stochastischer Anregung, die zur Beschreibung von Estimationsproblemen unbedingt erforderlich ist.

In Kapitel 5 finden die in den vorangegangenen Kapitel abgeleiteten Grundlagen ihre Anwendung bei der Ableitung und Formulierung des zeitdiskreten, linearen Optimalfilters, des sogenannten Kalman–Filters. Das Kapitel bietet drei unterschiedliche Formulierungen des optimalen Estimationsproblems, die von unterschiedlichen Voraussetzungen ausgehend, jeweils auf den gleichen, optimalen Kalman–Filteralgorithmus führen. Dieser Aufwand erscheint dem Verfasser gerechtfertigt, zumal die unterschiedlichen Herleitungen vollkommen unterschiedliche Aspekte des Kalman–Filters zeigen. Die sich dabei ergebenden, unterschiedlichen Aspekte des Kalman–Filters sollen dazu beitragen, die eigentliche Natur dieses Estimationsalgorithmus und damit einige Grundprinzipien der Estimation besser zu verstehen. Kapitel 5 beschäftigt sich weiterhin mit einigen mathematisch äquivalenten, numerisch jedoch unterschiedlichen Formulierungen des Kalman–Filters, die auch für praktische Probleme von einiger Wichtigkeit sein können. Auch wichtige Fragen wie Filterstabilität, numerisches Verhalten etc. werden in diesem Kapitel kurz andiskutiert. Praktische Probleme wie Filterdivergenz, Divergenztests, Plausibilitätstests von Meßwerten und Ausreißerkontrolle werden behandelt und Lösungsansätze vorgestellt, die in der Praxis von einiger Nützlichkeit sein können. Bei allen Darstellungen dieses Kapitels geht es dem Verfasser allerdings weniger um eine pedantische Realitätsnähe, sondern vielmehr darum, ein zusammenhängendes, tragfähiges Gesamtkonzept zu vermitteln, welches theoretisch abstrakt genug formuliert wird, daß das Gesamtkonzept als solches deutlich wird, auf der anderen Seite aber so verständlich bleibt, daß es, je nach Bedarf, zur Lösung praktischer Probleme herangezogen werden kann.

Kapitel 6 bringt schließlich ein ausführliches Beispiel für die Anwendung der Kalman–Filtertheorie zur Meßwertverarbeitung bei der Laserentfernungsmessung und vertieft das in dieser Einführung gebrachte, stark vereinfachte Anwendungsbeispiel beträchtlich. Dabei wird mit der stochastischen Modellbildung des Gesamtproblems begonnen, verschiedene Modellvereinfachungen werden vorgeschlagen und in ihrer Wirkungsweise diskutiert. Das auf der exakten Modellbildung beruhende Kalman–Filter wird formuliert und in seiner Wirkungsweise diskutiert. Anschließend wird das auf der vereinfachten Modellbildung beruhende Kalman–Filter vorgestellt und in seiner Wirkungsweise mit dem auf der exakten Modellierung beruhenden Kalman–Filter verglichen. Einige grundsätzliche Überlegungen zum praktischen Filterdesign beschließen dieses Kapitel.

1.8 Literatur zu Kapitel 1

1.) Gauß, C.F., Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientum, Hamburg, 1809 (englische Übersetzung: Dover, New York, 1963)

2.) Legendre, A.M., 'Methode des moindres quarrés, pour trouver le milieu le plus probable entre les résultats de differentes observtions', Mem. Inst. France, S. 149–154, 1810

3.) Kolmogorov, A.N.,'Sur l'interpolation et extrapolation des suites stationaires', C.R. Acad. Sci.,Vol. 208, S.2043, 1939

4.) Kolmogorov, A.N.,'Stationary Sequences in Hilbert Space', in Linear Least–Squares Estimation, ed. Th. Kailath, Dowden, Hutchinson & Ross, Benchmark Book Series/17, Pennsylvania, 1977

5.) Krein, G.M., 'On a generalization of some investigations of G. Szegö, W.M. Smirnov, and A.N. Kolmogorov', Dokl. Akad. Nauk. SSSR, Vol. 46, S.91–94, 1945

6.) Krein, G.M., 'On a problem of extrapolation of A.N. Kolmogorov', Dokl. Akad. Nauk. SSSR, Vol. 46, S.306–309, 1945

7.) Wiener,N., Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications, Technology Press and Wiley, New York, 1949, (Original 1942, Nat. Defense Res. Counsil Rep.)

8.) Kalman,R.E., 'A New Approach to Linear Filtering and Prediction Problems', J. Basic Eng., Vol. 82, S.34–45, Mar. 1960

9.) Kailath, Th., 'A View of Three Decades of Linear Filtering Theory', in Linear Least–Squares Estimation, ed. Th. Kailath, Dowden, Hutchinson & Ross, Benchmark Book Series/17, Pennsylvania, 1977

10.) Wiener,N., Collected Works with Commentaries, ed. P. Masani, Vol. I, MIT Press, Cambridge, 1976

11.) Wiener,N., Collected Works with Commentaries, ed. P. Masani, Vol. II, MIT Press, Cambridge, 1979

12.) Wiener,N., Collected Works with Commentaries, ed. P. Masani, Vol. III, MIT Press, Cambridge, 1981

13.) Wiener,N., Collected Works with Commentaries, ed. P. Masani, Vol. IV, MIT Press, Cambridge, 1985

14.) Maybeck,P.S.,'The Kalman Filter – An Introduction for Potential Users', TM–72–3 Air Force Flight Dynamics Laboratory, Wright Patterson AFB, Ohio, June 1972

2 Deterministische Systemmodelle im Zustandsraum

2.1 Einführung und Zielsetzung

Dieses Kapitel gibt eine kurze Zusammenfassung deterministischer Systemmodelle. Gleichzeitig dienen die Grundlagen dieses Kapitels als Voraussetzung für die stochastischen Systemmodelle, die später betrachtet werden. Systemmodellierung ist eine wesentliche Grundlage für das Kalman–Filterdesign, deshalb ist die Zusammenfassung dieses Kapitels relativ ausführlich.
Zunächst werden zeitkontinuierliche Modelle betrachtet, diese werden zwanglos von der aus der linearen Systemtheorie bekannten Übertragungsfunktion für Systeme mit einem Eingang und einem Ausgang abgeleitet. Dabei wird implizit von kausalen Systemen ausgegangen. Danach wird der Begriff Zustand verallgemeinert und die allgemeine Lösung der linearen Zustandsdifferentialgleichung hergeleitet. In diesem Zusammenhang werden lokale und globale Zustandsübergangsfunktionen eingeführt. Da Estimatoren und Regler typischerweise digital realisiert werden, folgen nach zeitkontinuierlichen Modellen die zeitdiskreten Darstellungen, diese werden aus den zeitkontinuierlichen Modellen abgeleitet. Es werden verschiedene Formen der zeitdiskreten Zustandsraumdarstellung diskutiert und abschließend werden die Begriffe Beobachtbarkeit und Steuerbarkeit behandelt.

Dieses Kapitel befaßt sich ausschließlich mit Zeitbereichsmodellen, da diese eine allgemeinere Darstellung als Frequenzbereichsmodelle ermöglichen.

2.2 Zeitkontinuierliche Systeme

2.2.1 Einführung der Zustandsraumdarstellung

Zur Einführung der Zustandsraumdarstellung werden zunächst lineare, zeitinvariante Systeme (LTI–Systeme) mit einem skalaren Eingang und einem skalaren Ausgang betrachtet.
Die Reaktion eines zeitinvarianten, linearen Systems mit einem Eingang u(t) und einem Ausgang y(t) wird mathematisch durch eine lineare Differentialgleichung n–ter Ordnung mit konstanten Koeffizienten beschrieben, die folgendes Aussehen besitzt:

$$\frac{d^n}{dt^n} y(t) + a_{n-1} \cdot \frac{d^{n-1}}{dt^{n-1}} y(t) + \ldots + a_0 \cdot y(t) = c_p \cdot \frac{d^p}{dt^p} u(t) + \ldots + c_0 \cdot u(t) \tag{2.1.a}$$

Führt man zur Abkürzung die Kurzschreibweise: $\frac{d^n}{dt^n} y(t) = y^{(n)}(t)$ ein, so ergibt sich:

$$y^{(n)}(t) + a_{n-1} \cdot y^{(n-1)}(t) + \dots + a_0 \cdot y(t) = c_p \cdot u^{(p)}(t) + \dots + c_0 \cdot u(t) \tag{2.1.b}$$

Hierbei ist u(t) die Anregungsfunktion und y(t) die Ausgangszeitfunktion des Systems. Die Linearität und Zeitinvarianz kann nun ausgenutzt werden, um das Eingangs – Ausgangsverhalten des Systems im Zeitbereich durch eine Stoßantwort h(t) und im Frequenzbereich durch eine Übertragungsfunktion H(f) zu beschreiben. Zieht man wegen der Beschränkung auf kausale Systeme die besser konvergierende Laplace–Transformation zur Herleitung einer Übertragungsfunktion H(s) vor, so kann man schreiben:

$$Y(s) = H(s) \cdot U(s) \tag{2.2}$$

Die Übertragungsfunktion H(s) findet man aus 2.1. durch Anwenden des Differentiationstheorems und gleichzeitiges Nullsetzen der Anfangsbedingungen. Es ergibt sich dann:

$$s^n \cdot Y(s) + a_{n-1} \cdot s^{n-1} \cdot Y(s) + \dots + a_0 \cdot Y(s) = c_p \cdot s^p \cdot U(s) + \dots + c_0 \cdot U(s) \tag{2.3}$$

Löst man nun nach $H(s) = \frac{Y(s)}{U(s)}$ auf, so erhält man die Übertragungsfunktion:

$$H(s) = \frac{Y(s)}{U(s)} = \frac{c_p s^p + c_{p-1} s^{p-1} + \dots\ c_o}{s^n + a_{n-1} s^{n-1} + \dots + a_0} \tag{2.4}$$

Die Pole von H(s) beschreiben das Verhalten der homogenen Lösung von Gl.2.1 und damit das dynamische Verhalten des Systems, während die Nullstellen von H(s) die Eingangszeitfunktion u(t), bzw. ihre Ableitungen gewichten. Die Ordnung des Nennerpolynoms entspricht dabei der Ordnung der Differentialgleichung.