

Dipl.-Wi.-Ing. Eberhard Münz
Rot

Identifikation und Diagnose hybrider dynamischer Systeme



Eberhard Münz

Identifikation und Diagnose hybrider dynamischer Systeme

Schriften des
Instituts für Regelungs- und Steuerungssysteme,
Universität Karlsruhe (TH)

Band 01

Identifikation und Diagnose hybrider dynamischer Systeme

von
Eberhard Münz



universitätsverlag karlsruhe

Dissertation, Universität Karlsruhe
Fakultät für Elektrotechnik und Informationstechnik, 2005

Impressum

Universitätsverlag Karlsruhe
c/o Universitätsbibliothek
Straße am Forum 2
D-76131 Karlsruhe
www.uvka.de



Dieses Werk ist unter folgender Creative Commons-Lizenz
lizensiert: <http://creativecommons.org/licenses/by-nc-nd/2.0/de>

Universitätsverlag Karlsruhe 2006
Print on Demand

ISSN 1862-6688
ISBN 3-86644-029-4

Identifikation und Diagnose hybrider dynamischer Systeme

Zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS

von der Fakultät für Elektrotechnik und Informationstechnik
der Universität Fridericiana zu Karlsruhe
genehmigte

DISSERTATION

von

Dipl.-Wi.-Ing. Eberhard Münz
geboren in Heidelberg

Tag der mündlichen Prüfung: 24. Januar 2006

Hauptreferent: Prof. Dr.-Ing. Volker Krebs

Korreferent: Prof. Dr. rer. nat. Karl-Heinz Waldmann

Rot, den 15. Mai 2006

Vorwort

Mit der Arbeit an der Dissertation höre ich jetzt auf. Irgendwann kann man durch zusätzlichen Aufwand nichts mehr verbessern; nicht weil alles perfekt wäre, sondern eher weil die Fehler genauso vertraut erscheinen wie die korrekten Dinge.

Dieser Zeitpunkt ist ganz gut, um über die Zeit, in der ich an der Dissertation gearbeitet habe, noch einmal nachzudenken. Es war eine sehr schöne Zeit, die Zeit am Institut für Regelungs- und Steuerungssysteme der Universität Karlsruhe (TH). Meistens spannend, oft lehrreich. Eine Zeit mit Höhen und Erfolgen. Sicher, es gab auch Tiefen und Misserfolge. Mit einem Vers aus der Bibel ausgedrückt: „Ein jegliches hat seine Zeit, und alles Vorhaben unter dem Himmel hat seine Stunde“.

Für diese Zeit möchte ich mich bei meinen Mitmenschen bedanken. Besonders bei Herrn Prof. Dr.-Ing. Krebs, dem Leiter des Instituts, der mich und meine Arbeit durch seine Anregungen und die auch über das Fachliche hinaus gehende Gesprächsbereitschaft gefördert hat. Ebenso möchte ich mich bei Herrn Prof. Dr. rer. nat. Waldmann bedanken, der das Korreferat für die Arbeit übernommen hat, als ob dies für ihn eine Selbstverständlichkeit wäre. Durch sein Interesse und die hilfreichen Hinweise hat er der Arbeit wertvolle Impulse gegeben.

Ich bin sicher, dass die Arbeit weniger verständlich wäre und mehr Fehler enthalten würde, wenn nicht Ulla, Dirk, Erhard, Markus, Martin und Mathias die Arbeit bzw. Teile davon so sorgfältig gelesen hätten. Frau Bickel hat mich bei der Erstellung der Zeichnungen unterstützt, die wesentlich zum Verständnis der Arbeit beitragen; Matthias stand mir bei Fragen zu L^AT_EX mit Rat und Tat zur Seite. Herzlichen Dank hierfür.

Den stets freundlichen und hilfsbereiten Kolleginnen und Kollegen, die in meiner Zeit am Institut immer für ein angenehmes Umfeld gesorgt haben, gilt mein herzlicher Dank. Durch anregende Diskussionen, die angenehme Zusammenarbeit und Gespräche – auch abseits vom Fachlichen – haben sie dazu beigetragen, dass ich mich am Institut sehr wohl gefühlt habe.

Fachlich wie menschlich haben mich die von mir betreuten Diplomandinnen und Diplomanden, die Masterarbeiter und Studienarbeiter voran gebracht. Sie haben einen großen Beitrag zur Dissertation geleistet.

Nicht wegzudenken sind die Stützen in meinem persönlichen Umfeld: meine Eltern, meine Freundin Ulla, meine Geschwister und Freunde. Auch ihnen danke ich für die gemeinsame Zeit.

Rot, im Mai 2006

Eberhard Münz

Ich kann's bis heute nicht verwinden,
deshalb erzähl' ich's auch nicht gern:
den Stein der Weisen wollt' ich finden
und fand nicht mal des Pudels Kern.

Heinz Erhardt

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problemstellung und Ziele der Arbeit	2
1.2	Aufbau der Arbeit	3
2	Modellbildung und Modellformen für dynamische Systeme	5
2.1	Grundlegende Begriffe	6
2.2	Systemklassifikation	7
2.2.1	Ereignisgetriebene Systeme	10
2.2.2	Zeitgetriebene Systeme	11
2.2.3	Wertdiskrete und wertkontinuierliche Systeme	11
2.2.4	Ereignisdiskrete Systeme	12
2.2.5	Zeitdiskrete und zeitkontinuierliche Systeme	12
2.2.6	Hybride Systeme	13
2.3	Systembeschreibungen in der Automatisierungstechnik	16
2.3.1	Modellformen zur Modellierung ereignisdiskreter Systeme . .	16
2.3.2	Modellformen für zeitdiskrete und zeitkontinuierliche Systeme	20
2.3.3	Gebräuchliche Modellformen für hybride Systeme	21
I	Identifikation hybrider Systeme	29
3	Grundlagen zur Identifikation	31
3.1	Modellbildungsverfahren	31
3.1.1	Theoretische Modellbildung	32
3.1.2	Experimentelle Modellbildung / Identifikation	32
3.2	Methoden und Aufgaben der Identifikation	33
3.2.1	Prädiktion	33
3.2.2	Simulation	34
3.2.3	Systemanalyse	35
3.2.4	Reglersynthese	35
3.2.5	Diagnose	35
3.3	Identifikation linearer Systeme	36
3.3.1	Schätzung statischer linearer Modelle	37

3.3.2	Schätzung dynamischer linearer Modelle	38
3.4	Identifikation nichtlinearer Systeme	41
3.5	Hybride Systeme	43
3.5.1	Zeitdiskrete Darstellung hybrider Systeme	43
3.5.2	Modellform für die Identifikation	44
4	Methoden zur Identifikation stückweise affiner Systeme	47
4.1	Besonderheiten bei der Identifikation stückweise affiner Modelle . . .	48
4.2	Überblick über existierende Verfahren	50
4.2.1	Detektion von Umschaltungen durch Betrachtung des Prädik- tionsfehlers	50
4.2.2	Identifikation stückweise affiner Systeme mittels Clusterung im Parameterraum	50
4.2.3	Identifikation stückweise affiner Systeme mit gemischt-ganz- zahliger linearer und quadratischer Optimierung	51
4.2.4	Weitere Verfahren	52
4.2.5	Diskussion der existierenden Verfahren	52
4.3	Lokale Kennfeldoptimierung	54
4.3.1	Achsenparallele Clusterung im Identifikationsraum	59
4.3.2	Optimierung von Form und Lage der Cluster	60
4.4	Identifikation stückweise affiner Systeme mit globaler Optimierung .	66
4.4.1	Methoden zur globalen Optimierung	74
4.4.2	Aufteilung des Identifikationsraumes mit Hilfe von Graphen .	75
5	Anwendung der hybriden Identifikationsverfahren	79
5.1	Beispiele zur Identifikation durch lokale Kennfeldoptimierung	79
5.1.1	Identifikation eines unstetigen statischen Beispielkennfeldes .	79
5.1.2	Kompressorkennfeld eines Turboladers	83
5.2	Beispiele zur Identifikation durch nichtlineare Optimierung	86
5.2.1	Identifikation eines statischen Beispielkennfeldes	86
5.2.2	Identifikation eines springenden Balls	92
5.2.3	Identifikation eines Zwei-Tank-Systems	95
II	Diagnose hybrider Systeme mit diskreten Sensoren	109
6	Grundlagen zur Diagnose hybrider dynamischer Systeme	111
6.1	Begriffsbildung	112
6.1.1	Fehlerdetektion	115
6.1.2	Fehlerisolation	115
6.1.3	Fehleridentifikation	116
6.2	Klassifikation der Diagnoseverfahren	116

6.2.1	Wissensbasierte Verfahren	116
6.2.2	Signalgestützte Verfahren	117
6.2.3	Modellgestützte Verfahren	117
6.3	Stochastische Zustandsrekonstruktion für hybride Systeme	118
6.3.1	Stochastische Grundlagen	121
6.3.2	Beobachtbarkeit und Rekonstruierbarkeit hybrider Systeme	126
6.3.3	Prinzip der stochastischen Zustandsrekonstruktion bei hybriden Systemen mit diskreten Messungen	128
6.3.4	Beobachterstruktur	134
6.3.5	Systemanforderungen für den Einsatz des hybriden stochastischen Beobachters	135
6.3.6	Informationsgewinn durch den hybriden stochastischen Beobachter	135
7	Verfahren zur Diagnose hybrider Systeme	141
7.1	Diagnose hybrider Systeme mit kontinuierlicher Sensorik	142
7.2	Diagnose hybrider Systeme mit diskreter Sensorik	143
7.2.1	Einsatz eines einzelnen hybriden stochastischen Beobachters	144
7.2.2	Mehrere hybride stochastische Beobachter	146
7.2.3	Einzelner hybrider stochastischer Beobachter mit Parameternachführung	147
7.2.4	Mehrere hybride stochastische Beobachter mit Parameternachführung	151
7.2.5	Mehrere hybride stochastische Beobachter mit Parameternachführung und Gütekriterium	152
8	Anwendung der hybriden Diagnoseverfahren	157
8.1	Aufbau des Zwei-Tank-Systems	157
8.2	Diskrete Sensorik	158
8.3	Fehlerarten	160
8.4	Nichtlineares Simulationsmodell	161
8.5	Batchprozess	162
8.6	Stochastischer Beobachter	164
8.7	Anwendung der Diagnoseverfahren	169
8.7.1	Einsatz eines einzelnen stochastischen Beobachters	170
8.7.2	Mehrere stochastische Beobachter	170
8.7.3	Ein stochastischer Beobachter mit Parameternachführung	172
8.7.4	Mehrere stochastische Beobachter mit Parameternachführung	174
8.7.5	Mehrere stochastische Beobachter mit Parameternachführung und Gütekriterium	180
8.8	Vergleich der Verfahren	182

9 Zusammenfassung	185
Anhang	189
A Grundbegriffe zur Graphentheorie	189
B Delaunay-Netze	193
B.1 Delaunay-Triangulierung und Simplexe	193
B.2 Baryzentrische Koordinaten	194
C Generieren von Fehlern am Zwei-Tank-Laborsystem	197
C.1 Geometrische Beschreibung	197
C.2 Identifikation	199
Literatur	201

Bezeichnungen

Allgemeine Bezeichnungen zur Modellbildung

Σ	Dynamisches System
\mathcal{D}	Definitionsbereich einer Funktion
\mathcal{T}	Menge der Zeitpunkte eines dynamischen Systems
\mathcal{X}	Menge der Zustände eines dynamischen Systems
\mathcal{U}	Menge der Eingangsgrößen eines dynamischen Systems
\mathcal{Y}	Menge der Ausgangsgrößen eines dynamischen Systems
f_x	Zustandsübergangsfunktion eines dynamischen Systems
f_y	Ausgangsfunktion eines dynamischen Systems
\mathbb{R}	Körper der reellen Zahlen
\mathbb{R}^+	Menge der positiven reellen Zahlen einschließlich 0
\mathbb{Z}	Körper der ganzen Zahlen
\mathbb{Z}^+	positive reelle Zahlen einschließlich 0
\mathbb{N}	Menge der natürlichen Zahlen
n	Dimension des Zustandsraums
p	Dimension des Eingangsraums
q	Dimension des Ausgangsraums
$\ \underline{x}\ $	verallgemeinerte Norm des Vektors \underline{x}
$\ \underline{x}\ _1$	1-Norm des Vektors \underline{x} ; $\ \underline{x}\ _1 = \sum_i x_i $
$\ \underline{x}\ _2$	2-Norm des Vektors \underline{x} ; $\ \underline{x}\ _2 = \sqrt{\sum_i x_i^2}$
$\ \underline{M}\ $	Norm der Matrix \underline{M}
$\ \underline{M}\ _2$	2-Norm der Matrix \underline{M} ; $\ \underline{M}\ _2 = \sqrt{\sum_i \sum_j m_{ij}^2}$

Netz-Zustands-Modell

NZM	Netz-Zustands-Modell
x_H	hybrider Zustand des NZM
x_D	diskreter Zustand des NZM

x_{Di}	i -te Komponente des diskreten Zustands
\underline{x}_K	kontinuierlicher Zustand des NZM
x_{Ki}	i -te Komponente des kontinuierlichen Zustands
\underline{u}_D	diskrete Eingangsgrößen des NZM
u_{Di}	i -te Komponente der diskreten Eingangsgrößen
\underline{u}_K	kontinuierliche Eingangsgrößen des NZM
u_{Di}	i -te Komponente der kontinuierlichen Eingangsgrößen
D/K-Interface	Diskret/Kontinuierliches Interface
K/D-Interface	Kontinuierlich/Diskretes Interface
\underline{v}_K	Kopplungsgrößen des D/K-Interface
\underline{v}_D	Kopplungsgrößen des K/D-Interface
IPN	Interpretiertes Petri-Netz
EZM	Erweitertes Zustandsraummodell
$\Gamma(\underline{x}_D)$	Menge der möglichen kontinuierlichen Zustände unter dem diskreten Zustand \underline{x}_D (Neutrale Menge)
Ω_i	Triggermenge
$\Omega_i^*(\underline{x}_D)$	aktive Triggermenge unter dem diskreten Zustand \underline{x}_D
$\partial\Omega_i$	Rand der Triggermenge

Stückweise affine Systeme

\mathcal{M}	stückweise affines System
\mathcal{I}	Definitionsbereich des stückweise affinen Systems
\mathcal{I}_j	Teilmenge des Definitionsbereichs, auf dem eine affine Systembeschreibung gilt
M	Anzahl der Teilsysteme
j	Index zur Nummerierung der Teilmodelle
\underline{A}_j	Dynamikmatrix von Teilsystem j
\underline{B}_j	Eingangsmatrix von Teilsystem j
\underline{c}_j	affiner Anteil von Teilsystem j
\underline{C}_j	Ausgangsmatrix von Teilsystem j

Identifikation

N	Anzahl der Messdaten
i	Index zur Nummerierung der Messdaten
$y(i)$	gemessene Daten
$\underline{\varepsilon}(i)$	Rauschen bei Messdatum i
z_{ij}	Zugehörigkeit von Datenpunkt i zum Teilmodell j

MCS	Multilevel Coordinate Search (globaler Optimierungsalgorithmus) [HN99]
s.t.	subject to (unter den Nebenbedingungen)

Stochastische Grundlagen

X, Y	Zufallsvariablen
$\underline{X}, \underline{Y}$	Zufallsvektoren
$m, E\{X\}$	Erwartungswert bzw. Erwartungswertoperator
$\underline{V}_X, \text{Cov}\{X\}$	Kovarianzmatrix bzw. Kovarianzoperator
$\{\underline{X}(t), t \in \mathcal{T}\}, \underline{X}_t$	vektorieller stochastischer Prozess
$\underline{X}(t)$	vektorielle Zufallsvariable des stochastischen Prozesses $\{\underline{X}(t), t \in \mathcal{T}\}$ zum Zeitpunkt t
$\underline{x}_\omega(t)$	Realisierung bzw. Musterfunktion des stochastischen Prozesses
$\underline{f}_X(t)$	Verteilungsdichte der Zufallsvariablen X
$\{\underline{\beta}(t), t \in [0, \infty)\}, \underline{\beta}_t$	vektorieller Wiener-Prozess
$\underline{\beta}(t)$	vektorielle Zufallsvariable des Wiener-Prozesses zum Zeitpunkt t
$\underline{\beta}_\omega(t)$	Musterfunktion des vektoriellen Wiener-Prozesses
\underline{Q}	Diffusionsmatrix eines vektoriellen Wiener-Prozesses
$\{\underline{\xi}(t), t \in [0, \infty)\}, \underline{\xi}_t$	vektorielles weißes Rauschen
\underline{S}	Steermatrix einer linearen stochastischen Differentialgleichung

Diagnose

HSB	hybrider stochastischer Beobachter
Ξ_K	Vertrauensmenge des HSB
j	Index zur Nummerierung der Systemparameter
P_j	Verstärkung des proportionalen Teils des PI-Algorithmus
I_j	Verstärkung des integrierenden Teils des PI-Algorithmus
N	Anzahl der unterschiedlichen Typen von Messereignissen bei dem betrachteten Prozess
$a \in \{1, \dots, N\}$	Index zur Nummerierung der unterschiedlichen Messereignisse
M	Anzahl der bisher aufgetretenen Messereignisse

$l \in \{1, \dots, M\}$	Index zur Nummerierung der Messereignisse
M_a	Anzahl der bisher aufgetretenen Messereignisse vom Typ a
$l_a \in \{1, \dots, M_a\}$	Index zur Nummerierung der Messereignisse vom Typ a

Kapitel 1

Einleitung

Da die meisten leistungsfähigen und modernen Methoden zur Automatisierung als Grundlage ein Modell des betrachteten Prozesses erfordern, muss für deren Einsatz zunächst dieses Prozessmodell erstellt werden. Die Aufgabe der Modellbildung ist daher zu lösen, bevor die eigentliche Aufgabe, nämlich die gezielte Beeinflussung des Prozesses, angegangen werden kann.

Die dazu aufgebauten Modelle müssen ausreichend genau für die jeweilige Anwendung sein, gleichzeitig muss aber darauf geachtet werden, dass die Komplexität der Modelle möglichst klein bleibt, da ansonsten die Automatisierung und der Reglerentwurf erschwert oder gar praktisch unmöglich werden.

Eine aufwändige theoretische Modellbildung, die von bekannten physikalischen und chemischen Zusammenhängen ausgeht und oft in komplexen Modellen mit partiellen Differentialgleichungen resultiert, führt deshalb nicht unbedingt zu einem Modell, das aus Sicht der Automatisierungstechnik geeignet ist.

Eine Alternative und sinnvolle Ergänzung zur theoretischen Modellbildung stellt die Identifikation dar, mit deren Hilfe Modelle aus Messdaten gewonnen werden können. Damit wird es möglich, sowohl Teile des Gesamtmodells zu erstellen als auch Modelle zu vereinfachen, da viele Verfahren zur Ordnungsreduktion auf Identifikationsverfahren beruhen. Zudem kann der Modellbildungsprozess mit Hilfe der Identifikationsverfahren häufig verkürzt werden, da bei vielen Prozessen das notwendige Wissen in Form von chemischen und physikalischen Zusammenhängen oft nicht in ausreichendem Maße zur Verfügung steht.

Identifikationsverfahren sind mittlerweile für viele Systemklassen entwickelt worden, so dass zahlreiche leistungsfähige Methoden sowohl für lineare als auch für nichtlineare Systeme zur Verfügung stehen. Für hybride und stückweise definierte Systeme besteht bisher allerdings das Problem, dass die methodischen Voraussetzungen für eine verbreitete Anwendung zum Teil noch nicht vorhanden sind. Deshalb soll diese Arbeit einen Beitrag zur Schließung dieser Lücken leisten, indem einsatzfähige Methoden zur Identifikation hybrider Systeme entwickelt werden.

Sind im Modellbildungsprozess geeignete Modelle erstellt worden, so können diese anschließend zur Simulation, zum Reglerentwurf und für weitere automatisierungstechnische Aufgaben eingesetzt werden.

Eine dieser weiteren Aufgaben stellt die Diagnose dar, der bei immer komplexer werdenden technischen Anlagen und Prozessen eine große Bedeutung zukommt; zusammen mit der Komplexität steigt nämlich auch der Kapitaleinsatz. Somit wachsen auch die Kosten für Ausfallzeiten von Anlagen. Zudem kann die Kundenzufriedenheit bei technischen Produkten durch fehlerfreien Betrieb gesteigert werden; falls trotzdem einmal ein Fehler auftritt, soll dieser schnell gefunden und behoben werden können. Um diese Ziele zu erreichen, ist es notwendig, durch die Diagnose die Fehler möglichst früh zu erkennen, zu lokalisieren und abhängig von der Größe des Fehlers geeignete Maßnahmen einzuleiten.

Bei der Diagnose hybrider Systeme sind nun hybride Modelle notwendig, die entweder aufwändig aus der theoretischen Modellbildung hervorgehen können oder – meist einfacher – durch geschickte Kombination von Identifikation und theoretischer Modellbildung gewonnen werden. Daher besitzt die Identifikation auch für die Diagnose einen großen Stellenwert.

1.1 Problemstellung und Ziele der Arbeit

Die Arbeit behandelt die zwei Problemstellungen der Identifikation und der Diagnose hybrider Systeme.

Ein erstes Ziel bei der Identifikation hybrider Systeme ist es, zunächst geeignete hybride Modellformen zu diskutieren und die beste Modellform für die Identifikation auszuwählen. In den bisher bekannten Arbeiten wird diesem Schritt sehr wenig Bedeutung beigemessen, so dass diese Problematik hier ausführlich diskutiert wird.

Ein weiteres Ziel stellt die Darstellung der spezifischen Probleme dar, die die hybride Identifikationsaufgabe aufwirft. Aus den bekannten Verfahren und aus den speziellen Problemen der hybriden Identifikationsaufgabe resultieren Anforderungen, die geeignete Verfahren erfüllen müssen. Entsprechend dieser Forderungen sollen Methoden entwickelt werden, die die gestellten Ansprüche erfüllen. Zusätzlich sollen die entwickelten Verfahren an Beispielen veranschaulicht werden.

Auch bei der Diagnose hybrider Systeme muss zunächst das erforderliche theoretische Fundament gelegt werden. Da die hier betrachteten Diagnosemethoden mit ausschließlich diskreten Messungen arbeiten sollen, besteht das erste Etappenziel darin, den Systemzustand mit Hilfe der diskreten Messungen zu ermitteln. Darauf aufbauend sollen Diagnosealgorithmen entwickelt werden, die die Fehlerdetektion,

Fehlerisolation und Fehleridentifikation leisten. Auch diese Verfahren sollen an einem Anwendungsbeispiel verdeutlicht werden.

Diese auf den ersten Blick voneinander unabhängigen Arbeitspakete erhalten dadurch einen tieferen Zusammenhang, dass für die Diagnose hybrider Systeme zunächst eine Modellbildung erforderlich ist. Die Identifikation stellt mögliche Modelle oder Teilmodelle für die spätere Diagnose bereit. Das Thema des ersten Teils dieser Arbeit ist deshalb die Identifikation hybrider Systeme, bevor im zweiten Teil Möglichkeiten für die Diagnose solcher Systeme präsentiert werden.

1.2 Aufbau der Arbeit

Zunächst werden in Kapitel 2 die gemeinsamen Grundlagen für die folgenden beiden Teile der Arbeit vorgestellt. Dabei werden grundlegende Begriffe der Modellbildung in der Automatisierungstechnik, Modelleigenschaften und die entsprechenden Modellformen diskutiert. Im daran anschließenden Teil I wird die Identifikation hybrider Systeme betrachtet. Danach behandelt Teil II die Diagnose hybrider Systeme mit ausschließlich diskreten Sensoren.

In Teil I werden in Kapitel 3 zuerst einige spezifische Grundlagen zur Identifikation dynamischer Systeme eingeführt und die dort verwendeten Begriffe erklärt. Anschließend wird als eine der wichtigsten Methoden zur Identifikation die Methode der kleinsten Quadrate eingeführt. Einige Identifikationsmethoden für nichtlineare Systeme werden danach kurz erläutert. Daraus ergibt sich für die Identifikation hybrider Systeme eine geeignete Modellform.

Geeignete Methoden zur Identifikation hybrider Systeme werden in Kapitel 4 entwickelt. Dabei wird zunächst eine Übersicht über bestehende Verfahren gegeben, bevor in Abschnitt 4.3 und 4.4 zwei neue Möglichkeiten zur Identifikation hybrider Systeme vorgestellt werden.

Im letzten Kapitel des ersten Teils werden die beiden eingeführten Methoden zur Identifikation hybrider Systeme anhand von Beispielen genauer erläutert.

Teil II der Arbeit beginnt zunächst damit, dass in Kapitel 6 spezifische Grundlagen zur Diagnose dynamischer Systeme vorgestellt werden. Darüber hinaus wird dort das Konzept des hybriden stochastischen Beobachters sowie die dafür benötigten Grundlagen eingeführt. Dieser hybride stochastische Beobachter bildet die Grundlage für die in Kapitel 7 entwickelten Methoden zur Diagnose hybrider dynamischer Systeme mit diskreten Sensoren. Die diskutierten Methoden werden anhand eines Beispiels in Kapitel 8 näher erläutert und dargestellt.

Abschließend wird die Arbeit in Kapitel 9 zusammengefasst.

Kapitel 2

Modellbildung und Modellformen für dynamische Systeme

Bedingt durch die unterschiedlichen Aufgabenstellungen der Automatisierungstechnik wurden in der Vergangenheit zum Teil sehr spezialisierte Methoden für die einzelnen Anwendungsgebiete entwickelt. Eine Gemeinsamkeit bei vielen Aufgaben in der Automatisierungstechnik besteht darin, dass zu deren Lösung ein Modell des zu Grunde liegenden Prozesses erforderlich ist; anschließend sind anhand des Modells Simulationen mit CASE-Werkzeugen, die Analyse von Modelleigenschaften und schließlich die Synthese einer geeigneten Automatisierungsstruktur möglich. Schon durch diese kurze Beschreibung wird deutlich, dass die Aufgabe der *Modellbildung* die nachfolgenden Aufgaben in hohem Maße beeinflusst.

Dabei kann die Modellbildung in sehr unterschiedlicher Weise erfolgen. Beispielsweise kann ein Modell nur als Prozesswissen eines Menschen vorliegen, der die Parameter eines Reglers einstellt; andererseits kann es vorkommen, dass zur Lösung komplexer automatisierungstechnischer Aufgaben ein kompliziertes mathematisches Modell erstellt werden muss. Ganz allgemein gibt ein Modell also einen Teil der Wirklichkeit wieder. Die Genauigkeit dieses Abbildes hängt zum einen von den Kenntnissen über das abzubildende Verhalten und zum anderen von den Problemen ab, die mit dem Modell gelöst werden sollen; so sind z.B. für den klassischen Reglerentwurf im Frequenzbereich häufig relativ grobe Modelle ausreichend, wohingegen die Simulation sehr genaue Modelle erfordert.

Die große Breite der bekannten unterschiedlichen Modellformen macht es notwendig, die Klasse der in dieser Arbeit betrachteten Modellformen einzuschränken und in den Kontext anderer geläufiger Modellformen einzuordnen. Deshalb werden zunächst einige grundlegende Begriffe aus dem Bereich der Modellbildung eingeführt;

anschließend werden in Form einer Klassifikation einige in der Automatisierungstechnik häufig verwendete Modellklassen vorgestellt. Im weiteren Verlauf dieses Kapitels werden die in dieser Arbeit verwendeten Modellformen erläutert.

2.1 Grundlegende Begriffe

Da mit Hilfe der Automatisierungstechnik Vorgänge in der realen Welt beeinflusst werden sollen, ist es notwendig, diese Prozesse selbst und auch den Abbildungsvorgang, mit dem die Prozesse der realen Welt in ein Modell abgebildet werden, begrifflich genauer zu fassen und gegeneinander abzugrenzen. Dazu werden die Begriffe *Prozess*, *Modell*, *Modellbildung* und *System* eingeführt.

Definition 2.1: Prozess

Ein Prozess ist ein Teil der realen Welt, durch den Materie, Energie oder Information umgeformt, transportiert oder gespeichert wird.

Ein solcher Prozess kann technischer Natur sein wie z.B. die Kalzinierung in einem Drehrohrofen, aber auch ökonomische Prozesse sind Teile der realen Welt, die z.B. in der Volkswirtschaftslehre in Modelle gefasst werden.

Unter einem Modell soll im weiteren Verlauf der Arbeit die abstrakte Beschreibung eines Prozesses verstanden werden.

Definition 2.2: Modell

Ein Modell ist die formale Beschreibung eines Prozesses, der ein bestimmtes Abstraktionsniveau zugrunde liegt.

Die Ableitung eines Modelles für einen Prozess wird als *Modellbildung* bezeichnet. Diese Abbildung wird niemals exakt sein; vielmehr stellt diese Abbildung eine Abstraktion des Prozesses dar. Da die Modellbildung nicht Selbstzweck ist, sondern zielgerichtet erfolgt – z.B. zum Lösen automatisierungstechnischer Aufgaben – und stets mit Aufwand verbunden ist, wird der Modellierer den Prozess nur so genau beschreiben, dass die zugrunde liegende Aufgabe gelöst werden kann.

Um eine Modellbildung durchführen zu können, grenzt der Modellierer den Prozess, der durch das Modell beschrieben werden soll, von der Umwelt ab und führt so oft künstliche Grenzen zwischen dem zu betrachtenden Prozess und dessen Umwelt ein. Häufig werden auch Schnittstellen definiert, über die der Prozess Energie, Materie oder Information mit seiner Umwelt austauscht und oft werden weitergehende Wechselwirkungen zwischen Prozess und Umwelt bei der Modellierung vernachlässigt. Diese Abgrenzung führt zum Systembegriff, wie er in DIN 19226 zu finden ist.

Definition 2.3: System im Sinne der DIN

Ein System ist eine eindeutig abgegrenzte Anordnung von aufeinander einwirkenden Systemteilen mit definierten inneren Größen, definierten Ein- und Ausgangsgrößen und einer definierten Art des Zusammenwirkens zwischen diesen.

Davon abweichend findet sich in der Systemtheorie ein etwas anderes Systemverständnis [GRS97].

Definition 2.4: System im Sinne der Systemtheorie

Ein System ist die Abstraktion eines Prozesses oder Gebildes, das mehrere Signale zueinander in Beziehung setzt.

Die beiden Definitionen sind teilweise nicht miteinander vereinbar, da Definition 2.3 keine Aussagen darüber macht, ob ein System Teil der realen oder der abstrakten Welt ist. Im Gegensatz dazu sieht die enger gefasste Definition 2.4 ein System als Teil der abstrakten Welt an; daher ist der Begriff des Systems im Sinne der Systemtheorie mit dem Modellbegriff aus Definition 2.2 stark verwandt.

Bei der Modellbildung werden nun die Eigenschaften, die der zu modellierende reale Prozess aufweist, oft durch die Modellierung akzentuiert, so dass das Modell eines Prozesses Eigenschaften aufweist, die beim Prozess selbst nicht so stark ausgeprägt sind und erst durch die Abstraktionen bei der Modellbildung betont werden. So wird zum Beispiel kaum ein realer Prozess wirklich einer linearen Prozessbeschreibung genügen; im Hinblick auf einige automatisierungstechnische Ziele kann es aber durchaus ausreichend sein, ein lineares Modell des Prozesses zu erstellen.

Da das Modell zentraler Ausgangspunkt für die Aufgaben der Automatisierungstechnik ist und die verwendeten und weiterführenden Methoden meist auch speziell für eine bestimmte Klasse von Modellen entwickelt worden sind und entwickelt werden, erscheint es sinnvoll, die Eigenschaften von verschiedenen Modellklassen genauer zu betrachten. Dabei wird der Systembegriff im weiteren Verlauf der Arbeit meist so verstanden, wie er in Definition 2.4 wiedergegeben ist.

2.2 Systemklassifikation

Um Systeme klassifizieren zu können, muss man sich zunächst klar machen, dass bei der Klassifikation von Systemen eigentlich auf Merkmale zurückgegriffen wird, die ein abstraktes System erst durch eine formale Systembeschreibung, also durch ein Modell des Prozesses, bekommt. Daher wäre es eigentlich gerechtfertigt, von Modellklassifikation und Modellklassen zu sprechen. Dem Sprachgebrauch im Kontext der

Automatisierungstechnik folgend, werden in dieser Arbeit synonym dazu die Begriffe Systemklassifikation und Systemklassen verwendet.

In diesem Abschnitt wird zunächst zwischen statischen und dynamischen Systemen unterschieden und eine formale Definition für dynamische Systeme angegeben; anhand dieser Definition werden weitere Eigenschaften erläutert, wobei besonders auf die hybriden Systeme eingegangen wird, da sie Thema dieser Arbeit sind. Eine Übersicht über die unterschiedlichen hier beschriebenen Eigenschaften und die Kombination der Eigenschaften zu neuen Systemklassen ist in Abbildung 2.1 zu sehen. Dabei sind die Systemklassen als Ellipsen und mögliche Modellformen als Rechtecke dargestellt.

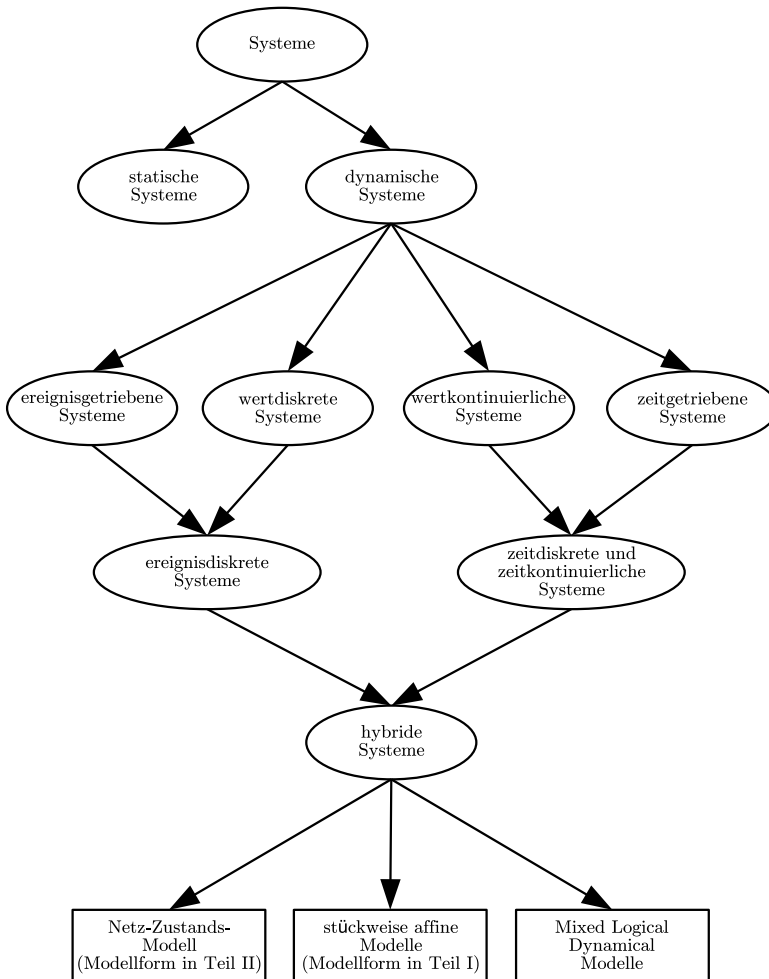


Abbildung 2.1: Übersicht über die Systemklassen

Das Wesen von dynamischen Systemen ist – im Gegensatz zu statischen Systemen – dadurch geprägt, dass die Vorgeschichte des Systems *und* die aktuellen Eingangsgrößen den weiteren Verlauf der Systemgrößen beeinflussen. Demnach ist das Verhalten eines statischen Systems allein durch die momentanen Eingangsgrößen bestimmt. Ein *dynamisches System* ist dagegen ein System, bei dem die momentanen Werte der Ausgangsgrößen nicht nur von den momentanen Werten der Eingangsgrößen, sondern auch vom bisherigen Verlauf der Systemgrößen abhängen.

Bei diesem Verständnis eines dynamischen Systems ist hervorzuheben, dass dynamische Systeme einen Bezug zu einer vom System unabhängigen Variablen – nämlich der Zeit – benötigen. Aufgrund dieser unabhängigen Variablen wird es erst möglich, von den *momentanen* Werten und vom *bisherigen* Verlauf der Systemgrößen zu sprechen. Ein endlicher Automat, der in der Informatik [CL99] normalerweise ohne Zeitbezug eingeführt wird, ist trotzdem ein dynamisches System, da bei diesem eine Reihenfolge der einzelnen Ereignisse aufgrund der kausalen Beziehungen zwischen den einzelnen Ereignissen festgelegt ist. Dadurch entsteht ein Zeitbezug, auch wenn die einzelnen Ereignisse unendlich schnell aufeinander folgen.

Zur Beschreibung von solchen dynamischen Systemen werden entsprechende Formalismen benötigt, die dazu in der Lage sind, die dynamischen Vorgänge geeignet zu beschreiben. In Anlehnung an die Definitionen in [Nen01,Hoh02,Son98] wird deshalb auch hier eine detaillierte Definition für dynamische Systeme angegeben.

Definition 2.5: Dynamisches System

Ein dynamisches System wird durch ein Tupel $\Sigma = [\mathcal{T}, \mathcal{X}, \mathcal{U}, f_x]$ dargestellt. Dabei haben die einzelnen Elemente folgende Eigenschaften:

- \mathcal{T} ist eine Menge von Zeitpunkten aus \mathbb{R}^+ und bildet bezüglich der Addition eine abgeschlossene Gruppe.
- Die nichtleere Menge \mathcal{X} bildet die Zustandsmenge und enthält alle zulässigen Zustände des Systems.
- Die nichtleere Menge \mathcal{U} bildet die Eingangsmenge und enthält alle zulässigen Eingangsgrößen des Systems.
- $f_x : \mathcal{D}_{f_x} \rightarrow \mathcal{X}$ stellt die Zustandsübergangsfunktion dar. Diese Abbildung ist auf einer Menge $\mathcal{D}_{f_x} \subset \{(\tau_1, \tau_2, x, u) \mid \tau_1, \tau_2 \in \mathcal{T}, \tau_1 \leq \tau_2, x \in \mathcal{X}, u \in \mathcal{U}^{[\tau_1, \tau_2]}\}$ definiert und besitzt die folgenden Eigenschaften:
 1. **Nicht-Trivialität:** Für jeden Zustand $x \in \mathcal{X}$ existiert mindestens eine zulässige Steuerung, d.h. es existiert mindestens ein Paar $\tau_1 < \tau_2$ aus \mathcal{T} und eine Steuerung $u \in \mathcal{U}^{[\tau_1, \tau_2]}$, so dass (τ_1, τ_2, x, u) aus dem Definitionsbereich \mathcal{D}_{f_x} der Zustandsübergangsfunktion f_x ist.

2. **Identität:** In jedem Zeitpunkt $\tau_1 \in \mathcal{T}$ stellt die leere Steuerung $\diamond \in \mathcal{U}^{[\tau_1, \tau_1]}$ eine zulässige Steuerung dar und es gilt: $f_x(\tau_1, \tau_1, x, \diamond) = x$.
3. **Kompositionseigenschaft:** Wenn für $\tau_1 < \tau_2 < \tau_3 \in \mathcal{T}$, $u_1 \in \mathcal{U}^{[\tau_1, \tau_2]}$ und $u_2 \in \mathcal{U}^{[\tau_2, \tau_3]}$ gilt: $f_x(\tau_1, \tau_2, x_1, u_1) = x_2$ und $f_x(\tau_2, \tau_3, x_2, u_2) = x_3$, dann gilt für die Steuerung

$$u(t) = \begin{cases} u_1(t) & \text{für } t \in [\tau_1, \tau_2) \\ u_2(t) & \text{für } t \in [\tau_2, \tau_3) \end{cases}$$

$$f_x(\tau_1, \tau_3, x_1, u) = x_3.$$

4. **Auftrenneigenschaft:** Ist eine Steuerung $u \in \mathcal{U}^{[\tau_1, \tau_3]}$ für den Zustand x über $[\tau_1, \tau_3)$ zulässig, so muss auch jede Einschränkung der Steuerung u auf den zeitlichen Gültigkeitsbereich $\mathcal{U}^{[\tau_1, \tau_2]}$ mit $\tau_1 \leq \tau_2 \leq \tau_3$ zulässig sein.

Wird zusätzlich eine Ausgangsmenge \mathcal{Y} und eine Ausgangsabbildung f_y eingeführt, so liegt ein dynamisches System mit Ausgang vor, wobei für $f_y : \mathcal{D}_{f_y} \rightarrow \mathcal{Y}$ und für $\mathcal{D}_{f_y} = \mathcal{T} \times \mathcal{X}$ gilt.

Die starke Verwandtschaft des Systembegriffs und des Modellbegriffs wurde bereits angesprochen. Mit Hilfe der bisher angegebenen Definitionen wird nun deutlich, dass ein dynamisches System nach Definition 2.5 ein Modell – also eine formale Beschreibung – für einen Prozess darstellt, falls solch ein System durch die Modellierung eines Prozesses gewonnen wurde. Andererseits muss ein System nicht unbedingt immer ein Modell sein, denn es kann losgelöst von der realen Welt betrachtet werden. Da allerdings in der Automatisierungstechnik reale Prozesse betrachtet werden, werden die relevanten Systeme immer Modelle für reale Prozesse sein.

Anhand der Eigenschaften, die die einzelnen Mengen und Abbildungen in dynamischen Systemen besitzen, können die dynamischen Systeme, wie in Abbildung 2.1 zu sehen ist, weiter klassifiziert werden.

2.2.1 Ereignisgetriebene Systeme

Bei ereignisgetriebenen Systemen wird ein neuer Systemzustand erst nach dem Auftreten von bestimmten Ereignissen erreicht. Bezogen auf Definition 2.5 heißt das, dass beispielsweise die Eingangsgröße u zu einem bestimmten Zeitpunkt $\tau \in \mathcal{T}$ ein Ereignis für das System darstellt, und somit einen Zustandsübergang auslöst. Falls man den Begriff *Ereignis* etwas weniger abstrakt fassen möchte, so kann man bestimmte Eigenschaften von Zeitfunktionen – dass beispielsweise ein Schwellwert überschritten wurde – als Ereignis definieren. Die Zustandsübergangsfunktion gibt

also, abhängig vom aktuellen Zustand, einen Folgezustand aus, falls die entsprechenden Ereignisse aufgetreten sind.

Definition 2.6: Ereignisgetriebene Systeme

Ein ereignisgetriebenes System ist ein dynamisches System nach Definition 2.5, bei dem der Wert der Zustandsübergangsfunktion nicht von den Zeitpunkten τ_1 und τ_2 abhängt, sondern lediglich davon, ob die Eingangsgrößen u in Abhängigkeit von den Zustandsgrößen x zu den betrachteten Zeitpunkten τ Ereignisse auslösen. Dabei sind die betrachteten Zeitpunkte τ die Teilmenge von \mathcal{T} , die $\tau_1 \leq \tau \leq \tau_2$ erfüllen. Die Zustandsübergangsfunktion eines ereignisgetriebenen Systems lässt sich also als

$$f_x(x, u)$$

darstellen.

2.2.2 Zeitgetriebene Systeme

Bei zeitgetriebenen Systemen wird die Zustandsübergangsfunktion f_x – anders als bei ereignisgetriebenen Systemen – von der Zeit als unabhängige Variable beeinflusst. Das bedeutet, dass das Ergebnis der Abbildung f_x maßgeblich von den beiden Zeitpunkten τ_1 und τ_2 abhängt.

Definition 2.7: Zeitgetriebene Systeme

Ein zeitgetriebenes System ist ein dynamisches System nach Definition 2.5, bei dem der Wert der Zustandsübergangsfunktion maßgeblich von den Zeitpunkten τ_1 und τ_2 beeinflusst wird. Die Zustandsübergangsfunktion f_x lässt sich nicht ohne explizite Abhängigkeiten von τ_1 und τ_2 darstellen.

Weiterhin lässt sich zwischen *zeitvarianten* und *zeitinvarianten* Systemen unterscheiden. Die Zustandsübergangsfunktion zeitvarianter Systeme lässt sich nur mit Hilfe beider Zeitpunkte τ_1 und τ_2 darstellen, wohingegen die Zustandsübergangsfunktion zeitinvarianter Systeme lediglich von der Differenz der beiden Zeitpunkte $\tau_2 - \tau_1$ abhängt.

2.2.3 Wertdiskrete und wertkontinuierliche Systeme

Unter wertdiskreten Systemen sollen in dieser Arbeit Systeme verstanden werden, bei denen die Menge der möglichen Zustände \mathcal{X} , die Menge der möglichen Eingangsgrößen \mathcal{U} und die Menge der möglichen Ausgangsgrößen \mathcal{Y} durch jeweils eine abzählbare Menge dargestellt werden können. Bei wertkontinuierlichen Systemen dagegen

sind die Mengen \mathcal{X} , \mathcal{U} und \mathcal{Y} Teilmengen von reellen Vektorräumen, die zusätzlich topologische Eigenschaften – wie z.B. bestimmte Zusammenhangseigenschaften – aufweisen.

2.2.4 Ereignisdiskrete Systeme

Ein ereignisdiskretes System ist nach [CL99] gegeben, wenn das System sowohl die Eigenschaften eines wertdiskreten Systems als auch die eines ereignisgetriebenen Systems aufweist.

Geht man davon aus, dass ein ereignisdiskretes System ein Modell für einen realen Prozess darstellt, so können bei der Modellierung des Prozesses nur bestimmte Ereignisse von Bedeutung gewesen sein; da die meisten technischen Prozesse zudem durch kontinuierliche Größen charakterisiert werden, muss auch bezüglich dieser Größen bei der Modellierung eine Abstraktion stattgefunden haben. Motiviert sein kann diese Sichtweise beispielsweise dadurch, dass nur diskrete Messgrößen zur Verfügung stehen oder der Prozess nur über diskrete Stellingriffe beeinflusst werden kann. Vor- und Nachteile dieser Darstellung sind u.a. in [Sch04a] diskutiert. Mögliche Modellformen für diese Systemklasse werden in Abschnitt 2.3.1 angegeben.

2.2.5 Zeitdiskrete und zeitkontinuierliche Systeme

Unter zeitdiskreten Systemen versteht man ein zeitgetriebenes System mit einer diskreten Zeitbasis; die Mengen \mathcal{X} , \mathcal{U} und \mathcal{Y} sind jeweils Teilmengen von reellen Vektorräumen. Die Zeitbasis \mathcal{T} eines zeitdiskreten Systems kann also durch eine abzählbare Menge dargestellt werden. Zeitdiskrete Systeme im engeren Sinn benutzen die Zeitbasis $\mathcal{T} = \mathbb{Z}^+$.

Definition 2.8: Zeitdiskretes System

Ein zeitdiskretes System ist ein dynamisches System nach Definition 2.5, bei dem die Zeitbasis \mathcal{T} durch die Menge der ganzen Zahlen \mathbb{Z}^+ dargestellt wird.

Zeitkontinuierliche Systeme im engeren Sinn hingegen sind zeitgetriebene Systeme mit der kontinuierlichen Zeitbasis \mathbb{R}^+ , wobei auch hier die Mengen \mathcal{X} , \mathcal{U} und \mathcal{Y} jeweils Teilmengen von reellen Vektorräumen sind.

Definition 2.9: Zeitkontinuierliches System

Ein zeitkontinuierliches System ist ein dynamisches System nach Definition 2.5, bei dem die Zeitbasis \mathcal{T} durch die Menge der reellen Zahlen \mathbb{R}^+ dargestellt wird.

Da diese Klasse von Systemen schon in den Anfängen der Regelungstechnik in Form von Übertragungsfunktionen, Differentialgleichungen und Differenzgleichungen betrachtet wurde, bestehen hier zahlreiche etablierte Modellformen und Methoden. Eine interessante Zusammenstellung zu den unterschiedlichen Aspekten zeitdiskreter und zeitkontinuierlicher Systeme findet sich in [Hoh02]. Mögliche Modellformen, die auch für diese Arbeit relevant sind, sind in Abschnitt 2.3.2 näher erläutert.

2.2.6 Hybride Systeme

Bei der Kombination von ereignisdiskreten und zeitdiskreten Systemen [BM99,Sch01a] oder bei der Kombination von ereignisdiskreten und zeitkontinuierlichen Systemen [Bra95, Nen01] entstehen Systeme, die in der Literatur als diskret-kontinuierliche oder hybride Systeme bekannt sind; dabei beziehen sich die Begriffe *diskret* und *kontinuierlich* auf die Zustandsmengen der Teilsysteme. Die Zustandsübergangsfunktionen dieser Systeme besitzen also formal betrachtet zum einen Anteile, die sich zeitgetrieben und wertekontinuierlich verändern, andererseits existieren Funktionsanteile, die sich ereignisgetrieben und wertdiskret verändern.

Da hauptsächlich diese Art von Systemen in dieser Arbeit untersucht wird, werden die entsprechenden Modellformen in Abschnitt 2.3.3 ausführlich dargestellt.

Die bei dieser Art von Systemen auftretenden Phänomene sind zuerst in [Bra95] beschrieben worden; es handelt sich dabei um Unstetigkeiten und Umschaltungen, die zumeist aufgrund von Abstraktionen und idealisierten Betrachtungen beim Modellieren eines realen Systems entstehen. So betrachtet man beispielsweise den Aufprallvorgang eines hüpfenden Gummiballs idealisiert, und zwar derart, dass diesem idealisierten Aufprallvorgang keine Zeit beigemessen wird, obwohl für den Auf- und Abbau der Deformation Zeit benötigt wird. Dadurch entsteht ein unstetiger Geschwindigkeitsverlauf.

Solche Abstraktionen werden immer dann zulässig sein, wenn zwei Voraussetzungen erfüllt sind [Nen01]:

- Die Abstraktionen, die bei der Modellierung gemacht werden, tragen dazu bei, dass das Modell einfacher und übersichtlicher wird und damit verschiedene weiterführende automatisierungstechnische Aufgaben erst ermöglicht werden.
- Das automatisierungstechnische Ziel kann trotz der vorgenommenen Modellvereinfachung erreicht werden; d.h. das Modell muss die für das Automatisierungsziel wesentlichen Verhaltensweisen des Prozesses immer noch genau genug wiedergeben.

Diese Voraussetzungen sind auch in der klassischen Regelungstechnik bekannt; dort werden beispielsweise Modelle um Arbeitspunkte linearisiert und auf diese Weise vereinfacht.

Um beispielsweise komplexe nichtlineare Prozesse zu modellieren, deren Gleichungen in geschlossener Form nicht mehr handhabbar wären, können hybride Systeme verwendet werden. Kommt bei diesen Prozessen gleichzeitig eine vollständig ereignisdiskrete Abstraktion nicht in Frage, da ansonsten wesentliche Vorgänge nicht mehr genau genug dargestellt würden, erscheint die Verwendung hybrider Systeme zwingend.

Weiterhin kann man Prozesse als hybride Prozesse betrachten, die durch eine Steuerung beeinflusst werden, bei der einige der Einflussgrößen nur diskrete Werte annehmen können.

Als Merkmal für hybride Systeme im engeren Sinn wird in [Nen01] genannt:

Ein dynamisches System wird als hybrides System bezeichnet, wenn sich der Zustand des Systems auf dem bei der Modellbildung zugrunde gelegten Abstraktionsniveau notwendigerweise aus diskreten und kontinuierlichen Zustandsgrößen zusammensetzt; die zeitliche Entwicklung der Zustandsgrößen wird durch die Wechselwirkung eines ereignisgetriebenen und zeitgetriebenen Modellanteils bestimmt.

Systeme, deren Dynamik in Abhängigkeit von ihrem Zustand umschaltet, können nach dieser Klassifikation als hybrid aufgefasst werden; da aber bei umschaltenden Systemen nicht zwangsläufig eine beiderseitige Wechselwirkung zwischen ereignisgetriebenem Systemanteil und kontinuierlichem Systemanteil auftritt, ist diese Zuordnung nicht zwingend. Allerdings ermöglicht eine hybride Beschreibung oft einen besseren Einblick in das Systemverhalten, so dass es von Vorteil sein kann, eine solche Beschreibung zu wählen.

Die auftretenden Unstetigkeitsphänomene von hybriden Systemen lassen sich in vier grundlegende hybride Phänomene einteilen, wie sie in Abbildung 2.2 gezeigt sind. Dabei wird der Zeitpunkt, zu dem die Unstetigkeit auftritt, mit t_k bezeichnet; x bezeichnet eine Zustandsgröße und der Index K bzw. D zeigt an, dass es sich um eine kontinuierliche bzw. diskrete Größe handelt. Der rechte Grenzwert einer Zustandsgröße zum Zeitpunkt t wird mit $x(t^+)$ bezeichnet.

Diese Phänomene sind mit Ereignissen verbunden und lassen sich unterteilen in Phänomene, die durch interne oder externe Ereignisse ausgelöst werden. *Interne Ereignisse* werden durch die Eigendynamik des hybriden Systems beim Erreichen eines bestimmten Zustandes ausgelöst, während *externe Ereignisse* dadurch ausgelöst werden, dass sich diskrete Eingangsgrößen zu diskreten Zeitpunkten sprunghaft ändern. Da außer dieser Unterscheidung eine weitere Unterscheidung zwischen *Umschaltungen* und *Sprüngen* vorgenommen wird, ergeben sich durch die entsprechenden Kombinationen insgesamt vier hybride Phänomene [Bra95]:

- A1 *Internes Umschalten* bedeutet, dass ein internes Ereignis – also beispielsweise eine Grenzwertüberschreitung – den diskreten Zustand des Systems verändert;

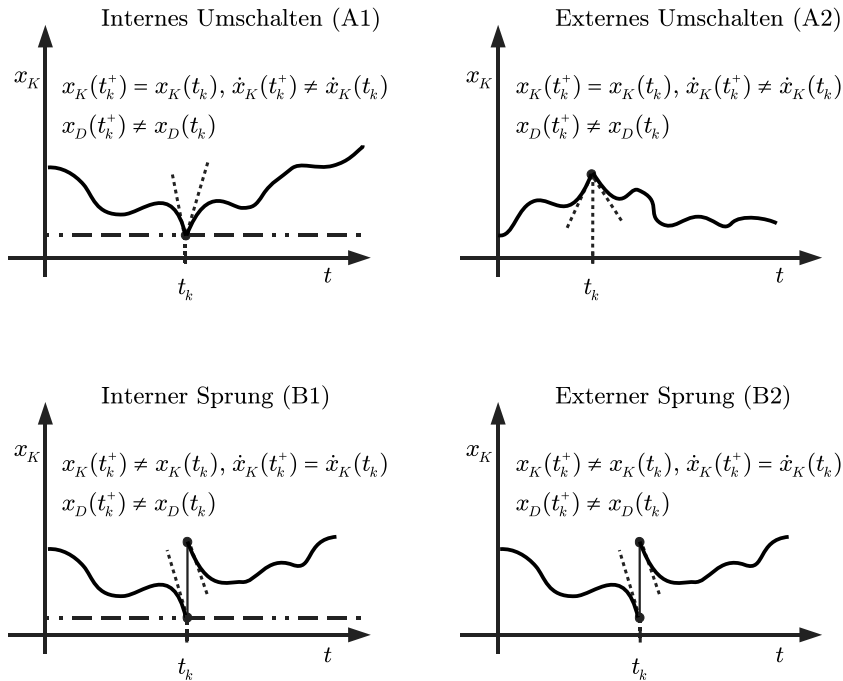


Abbildung 2.2: Charakteristische Verläufe einer kontinuierlichen Zustandsgröße beim Auftreten von Ereignissen in hybriden Systemen

die Veränderung des diskreten Zustandes wiederum schaltet die kontinuierliche Dynamik um.

- A2 Als *externes Umschalten* bezeichnet man den Fall, dass ein externes Ereignis, also die sprunghafte Änderung einer diskreten Eingangsgröße, den diskreten Zustand des Systems verändert; die Veränderung des diskreten Zustandes wiederum wirkt sich als Umschaltung der kontinuierlichen Dynamik aus.
- B1 Ein *interner Sprung* ist dadurch gekennzeichnet, dass ein internes Ereignis den diskreten Zustand des Systems verändert; diese Veränderung des diskreten Zustandes zieht unmittelbar eine sprunghafte Veränderung von kontinuierlichen Zustandsgrößen nach sich. Die Ableitung der springenden Größen verhalten sich dabei stetig. Beispielsweise können Kontaktsituationen in mechanischen Systemen mit Hilfe dieses Phänomens modelliert werden.
- B2 Das entsprechende externe Ereignis wird als *externer Sprung* bezeichnet und tritt beispielsweise auf, wenn in elektrischen Netzen eine Stromquelle zugeschaltet wird.

Um Systeme zu beschreiben, die zu dieser Systemklasse gehören, wurden spezielle

Modellformen entwickelt, von denen einige in Abschnitt 2.3.3 vorgestellt werden.

Nachdem nun verschiedene für die Automatisierungstechnik relevante Systemklassen gegeneinander abgegrenzt wurden, sollen im nächsten Abschnitt mögliche Systembeschreibungen für die einzelnen Klassen vorgestellt werden.

2.3 Systembeschreibungen in der Automatisierungstechnik

Für die Systemklassen, die in der automatisierungstechnischen Praxis häufig vorkommen, gibt es jeweils einige typische Beschreibungsformen. Einige dieser Beschreibungsformen sollen hier ohne Anspruch auf Vollständigkeit kurz vorgestellt werden, damit die dieser Arbeit zugrunde liegenden Beschreibungen eingeordnet werden können.

2.3.1 Modellformen zur Modellierung ereignisdiskreter Systeme

Die ältesten Modelle zur Modellierung ereignisdiskreter Systeme stellen die Automaten dar [AD94, CL99, HMU01]. Diese existieren in sehr unterschiedlichen Ausprägungen, beispielsweise als deterministische oder auch als stochastische Automaten.

Weitere Konzepte, die zur Modellierung von ereignisdiskreten Systemen herangezogen werden können, sind sogenannte *Kellerautomaten* und *Turing-Maschinen* [HMU01]. Diese Konzepte stammen ursprünglich aus der Informatik und wurden dort entwickelt, um Fragen der Komplexität und Berechenbarkeit zu analysieren.

Von den drei genannten Konzepten wird allerdings nur der Automat in nennenswertem Umfang in der Automatisierungstechnik benutzt.

Da der Zustand bei den bisher angesprochenen Konzepten zur Modellierung ereignisdiskreter Systeme in global gültigen Variablen gespeichert ist, sind diese Konzepte zur Modellierung technischer Anlagen, die häufig Nebenläufigkeiten und räumlich verteilte Zustandsindikatoren aufweisen, weniger geeignet. Geeigneter hierfür sind Modellformen, die den verteilten Zustand von nebenläufigen Prozessen auch verteilt speichern. Zur Vereinfachung der Abbildung von solchen Prozessen in ein entsprechendes ereignisdiskretes Modell hat sich die Anwendung von Petri-Netzen als vorteilhaft erwiesen.

Da der Formalismus der Petri-Netze in dieser Arbeit benötigt wird, sich jedoch in der Literatur keine einheitliche Bezeichnungsweise findet, werden im Folgenden

einige wichtige Definitionen und Zusammenhänge angeben.

Definition 2.10: Petri-Netz [Abe90]

Ein Petri-Netz \mathcal{N} wird durch ein 6-Tupel $\mathcal{N} = \{S, \mathcal{T}, \mathcal{K}, c, w, m_0\}$ dargestellt. Dabei gilt:

- S ist eine nicht leere, endliche Stellenmenge,
- \mathcal{T} ist eine nicht leere, endliche Transitionsmenge und
- \mathcal{K} ist eine nicht leere Kantenmenge, die ausschließlich aus Prä- und Postkanten besteht:
 - Jede Präkante verbindet jeweils ein Element aus S mit einem Element aus \mathcal{T} .
 - Jede Postkante verbindet jeweils ein Element aus \mathcal{T} mit einem Element aus S .
- $c : S \rightarrow \mathbb{N}$ beschreibt eine Abbildung, die jeder Stelle eine Kapazität zuordnet,
- $w : \mathcal{K} \rightarrow \mathbb{N}$ stellt eine Abbildung dar, die jeder Kante ein Gewicht zuordnet und
- $m_0 : S \rightarrow \mathbb{N}$ beschreibt die initiale Markenbelegung der Stellen.

Diese Definition findet sich sinngemäß auch in [Bau96, KQ88, CL99] wieder, wobei jedoch – je nach Anwendungskontext – Details unterschiedlich sind. Aus dieser grundlegenden Definition der Petri-Netze wurden für unterschiedliche Anwendungen viele spezielle Netzklassen entwickelt und auch zahlreiche Erweiterungen zu Definition 2.10 hinzugefügt [Bau96]. Dabei ist zu beachten, dass die Komplexität von Petri-Netzen über die von endlichen Automaten hinaus geht, da beispielsweise unendlich viele Marken in einzelnen Stellen nicht von vornherein ausgeschlossen sind. Graphentheoretisch lassen sich Petri-Netze als bipartite Graphen interpretieren, da sie aus zwei unterschiedlichen Arten von Knoten bestehen und Kanten immer nur zwischen zwei Knoten verlaufen, die nicht der gleichen Art angehören. Die Grundbegriffe zur Graphentheorie werden in Anhang A näher erläutert. Zur graphischen Darstellung der Petri-Netze verwendet man häufig Kreise für die Stellen und Balken bzw. Quadrate für die Transitionen. Stellen sind ebenso wie die Kanten mit ihren Kapazitäten bzw. Gewichten beschriftet; unbeschriftete Stellen und Kanten besitzen die Kapazität bzw. das Gewicht 1. Marken werden als schwarze Punkte innerhalb der Stellen dargestellt.

Die Dynamik des Netzes – der so genannte Markenfluss – wird durch Schaltregeln festgelegt. Um diese zu formulieren, werden folgende Begriffe benötigt:

- Mit $\bullet s$ bzw. $\bullet t$ wird der *Vorbereich* einer Stelle, bzw. einer Transition bezeichnet. Der Vorbereich einer Stelle s besteht aus den Transitionen, die über Postkanten mit der Stelle verbunden sind; der Vorbereich einer Transition t besteht aus den Stellen, die über Präkanten mit der Transition verbunden sind.

- Mit $s\bullet$ bzw. $t\bullet$ wird der *Nachbereich* einer Stelle, bzw. einer Transition bezeichnet. Der Nachbereich einer Stelle s besteht aus den Transitionen, die über Präkanten mit der Stelle verbunden sind; der Nachbereich einer Transition t besteht aus den Stellen, die über Postkanten mit der Transition verbunden sind.

Damit kann die Schaltregel definiert werden:

Definition 2.11: Schaltregel für Petri-Netze

Die Schaltregel setzt sich aus der Schaltvoraussetzung und dem eigentlichen Schaltvorgang zusammen:

1. Schaltvoraussetzung: Eine Transition ist schaltfähig, wenn
 - jede Stelle im Vorbereich der Transition eine Anzahl von Marken enthält, die größer oder gleich dem zugehörigen Präkantengewicht ist.
 - jede Stelle im Nachbereich der Transition eine Anzahl von Marken aufnehmen kann, die größer oder gleich dem zugehörigen Postkantengewicht ist, ohne dass die Stelle ihre Kapazität überschreitet.
2. Schaltvorgang: Die Transition schaltet, wenn sie schaltfähig ist (Muss-Schaltregel). Dabei wird jeder Stelle aus dem Vorbereich der Transition die Anzahl an Marken entnommen, die das zugehörige Präkantengewicht angibt. Jeder Stelle des Nachbereichs der Transition wird die Anzahl an Marken hinzugefügt, die das zugehörige Postkantengewicht angibt.

Zwischen zwei gleichzeitig aktivierten Transitionen kann nun ein *Konflikt* bestehen, wenn nach dem Schalten einer Transition die jeweils andere nicht mehr schalten kann. Ebenso können *Kontaktsituationen* auftreten, wenn durch das Schalten einer Transition t_i der Nachbereich einer anderen Transition t_j so beeinflusst wird, dass sie nicht mehr schalten kann.

Die Auflösung von Konflikt- und Kontaktsituationen ist nicht trivial und wird im automatisierungstechnischen Kontext oft durch Zusätze zur Schaltregel aufgelöst. In dieser Arbeit wird davon ausgegangen, dass durch die automatisierungstechnische Interpretation des Netzes solche Situationen mit Hilfe von zusätzlichen Bedingungen aufgelöst werden. Da vorausgesetzt werden kann, dass die hier behandelten Petri-Netze einen Anwendungsbezug besitzen, ist es möglich, den Stellen und Transitionen eine semantische Bedeutung zuzuweisen. Dadurch entstehen *Interpretierte* Petri-Netze, die in der folgenden Definition näher erläutert werden [KQ88, Sch04b].

Definition 2.12: Interpretiertes Petri-Netz

Ein interpretiertes Petri-Netz \mathcal{IN} ist ein 5-Tupel $\mathcal{IN} = (\mathcal{N}, \mathcal{U}_D, \mathcal{Y}_D, \beta, f_{Dy})$ mit:

- dem Petri-Netz \mathcal{N} ,
- der binären Eingangsmenge $\mathcal{U}_D = \{0, 1\}^{p_D}$,
- der diskreten Ausgangsmenge \mathcal{Y}_D ,
- der Abbildung $\beta : \mathcal{T} \times \mathcal{U}_D \rightarrow \{0, 1\}$, die jeder Transition einen booleschen Ausdruck zuordnet, dessen Wahrheitswert von äußeren Größen \mathcal{U}_D abhängt und
- der Abbildung $f_{Dy} : \mathcal{X}_D \rightarrow \mathcal{Y}_D$, die jedem diskreten Zustand des Netzes einen diskreten Ausgangszustand zuordnet.

Die Schaltregel für ein Interpretiertes Petri-Netz muss dementsprechend erweitert werden:

Definition 2.13: Schaltregel bei einem interpretierten Petri-Netz

Eine Transition t_j in einem interpretierten Petri-Netz schaltet entsprechend der Schaltregel aus Definition 2.11, wenn der boolesche Ausdruck, der zur Transition t_j gehört, wahr ist.

Es wird im weiteren Verlauf der Arbeit davon ausgegangen, dass aufgrund der booleschen Ausdrücke keine Konflikte und Kontakte mehr auftreten können.

Da Petri-Netze und auch Automaten ereignisdiskrete Systeme abbilden, ist es angebracht, die beiden Formalismen zu vergleichen:

Während Automaten viele einzelne Zustände besitzen, die immer Zustände des Gesamtsystems sind, besitzen Petri-Netze aufgrund ihrer Markierung einen Zustand, der über alle Stellen des Petri-Netzes verteilt ist. Möchte man nun für ein gegebenes Petri-Netz einen äquivalenten Automaten konstruieren, so ist dies mit Hilfe des Erreichbarkeitsgraphen des Petri-Netzes möglich. Den Erreichbarkeitsgraphen erhält man, indem man ausgehend von der Anfangsmarkierung des Petri-Netzes die erreichbaren Folgemarkierungen ermittelt und diese in einem neuen Graphen darstellt. Besitzt dieser Erreichbarkeitsgraph nur endlich viele Knoten, so stellt er gleichzeitig auch den Graphen für den äquivalenten endlichen Automaten dar. Unter der Voraussetzung, dass das Petri-Netz nur über Stellen mit endlicher Kapazität verfügt, besitzt das Netz einen endlichen Erreichbarkeitsgraphen und ist somit auch durch einen endlichen Automaten darstellbar. Lediglich diese Klasse von Netzen wird im weiteren Verlauf der Arbeit benötigt.

2.3.2 Modellformen für zeitdiskrete und zeitkontinuierliche Systeme

Zeitdiskrete und zeitkontinuierliche Systeme sind Systemklassen, die schon seit den Anfängen der Regelungstechnik behandelt werden. Solange diese Systeme linear und zeitinvariant sind, kann ihr Ein-/Ausgangsverhalten mit Hilfe der klassischen Modelle in Form von zeitkontinuierlichen oder zeitdiskreten Übertragungsfunktionen beschrieben werden. Da diese Modellformen auf lineare und zeitinvariante Systeme beschränkt sind, wurden weitere Modellformen entwickelt, mit denen auch nichtlineare und zeitvariante Systeme dargestellt werden können. Zustandsraummodelle sind dazu in der Lage, das Systemverhalten von nichtlinearen und zeitvarianten Systemen im Zeitbereich mit Hilfe von Differenzgleichungen und Differentialgleichungen zu beschreiben. Darüber hinaus fügt sich der Mehrgrößenfall bei Zustandsraummodellen zwangloser in den Formalismus ein, wohingegen die Beschreibung von Mehrgrößensystemen mit Übertragungsfunktionen einigen Aufwand erfordert.

Da sich diese Arbeit hauptsächlich mit hybriden Systemen beschäftigt, die eine Verallgemeinerung der nichtlinearen Systeme darstellen, werden hier die Zustandsraummodelle eine große Rolle spielen.

Für zeitdiskrete Systeme bestehen die Zustandsraummodelle aus vektoriellen Differenzgleichungen, bei denen der neue Zustand $\underline{x}(k+1)$ rekursiv aus dem alten Zustand $\underline{x}(k)$ und den Eingangsgrößen $\underline{u}(k)$ berechnet wird:

Definition 2.14: Zeitdiskretes System in Zustandsraumdarstellung

Die Differenzgleichung

$$\underline{x}(k+1) = \underline{f}(\underline{x}(k), \underline{u}(k), k)$$

mit der Ausgangsgleichung

$$\underline{y}(k) = \underline{g}(\underline{x}(k), \underline{u}(k), k)$$

stellt ein zeitdiskretes System in Zustandsraumdarstellung dar. Der Anfangszustand dieses Systems ist gegeben durch $\underline{x}(0) = \underline{x}_0$. Durch die Abbildung \underline{f} wird die Zustandsübergangsfunktion f_x aus Definition 2.5 festgelegt; \underline{g} entspricht der Ausgangsabbildung f_y .

Zur Darstellung zeitkontinuierlicher Systeme können entsprechend vektorielle Differentialgleichungen verwendet werden:

Definition 2.15: Zeitkontinuierliches System in Zustandsraumdarstellung

Die Differentialgleichung

$$\dot{\underline{x}} = \underline{f}(\underline{x}, \underline{u}, t)$$

mit der Ausgangsgleichung

$$\underline{y} = \underline{g}(\underline{x}, \underline{u}, t)$$

stellt ein zeitkontinuierliches System in Zustandsraumdarstellung dar. Der Anfangszustand dieses Systems ist gegeben durch $\underline{x}(0) = \underline{x}_0$. Durch die Abbildung \underline{f} wird wiederum die Zustandsübergangsfunktion f_x aus Definition 2.5 festgelegt; \underline{g} entspricht der Ausgangsabbildung f_y .

Je nach Anwendungskontext werden für die Wertemengen und Bildmengen der Abbildung sowie für die Abbildungen g und f selbst die dafür erforderlichen Eigenschaften gefordert. Da in dieser Arbeit hauptsächlich hybride Systeme betrachtet werden, wird auf die Forderung spezieller Eigenschaften zunächst verzichtet und die Modellformen werden an dieser Stelle lediglich ganz allgemein eingeführt.

2.3.3 Gebräuchliche Modellformen für hybride Systeme

Zur Modellierung hybrider Systeme wurden in den letzten Jahren zahlreiche Möglichkeiten vorgeschlagen. Daher kann hier kein vollständiger Überblick über diese Möglichkeiten gegeben werden; lediglich eine kleine Auswahl von möglichen Modellen wird vorgestellt.

Über die hier vorgestellten Modellformen hinaus finden sich in der Literatur zahlreiche weitere Möglichkeiten wie zum Beispiel hybride Automaten und hybride Programmiersprachen. In [LBR96] findet sich dazu eine Übersicht.

Weite Verbreitung haben hybride Zustandsmodelle erfahren, in denen sich der Systemzustand zum einen aus Zustandsvariablen zusammensetzt, die reelle Werte annehmen, und zum anderen aus Zustandsvariablen, die eine endliche oder abzählbar unendliche Anzahl diskreter Werte annehmen können. Weitere Betrachtungen hierzu sind in [Nen01, Bus01, Mül01] zu finden.

Die nachfolgend vorgestellten Modellformen werden auch in dieser Arbeit verwendet. Das *Netz-Zustands-Modell* findet Verwendung bei der Diagnose hybrider Systeme und stellt die Grundlage des dort verwendeten hybriden stochastischen Beobachters dar. Die *Mixed Logical Dynamical Modelle* finden besondere Verwendung in der zeitdiskreten Betrachtung hybrider Systeme und sind als Übergang zu den stückweise affinen Modellen hilfreich. Schließlich werden die *stückweise affinen Modelle* vorgestellt, die in der Literatur verbreitet als Modellform zur Identifikation hybrider Systeme verwendet werden.

2.3.3.1 Das Netz-Zustands-Modell

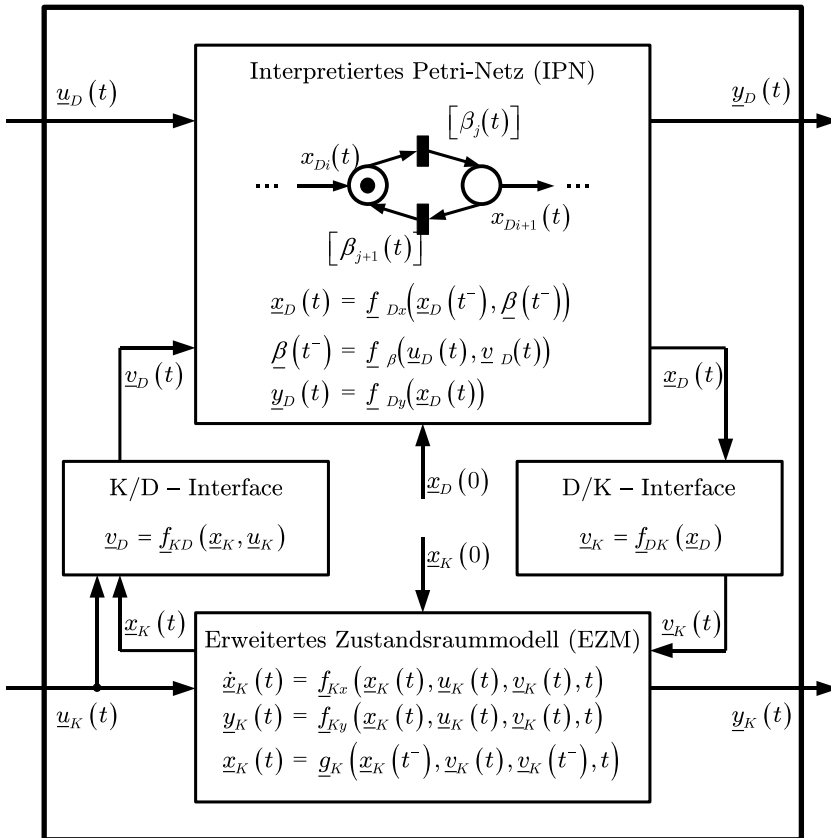


Abbildung 2.3: Struktur des Netz-Zustands-Modells

Das Netz-Zustands-Modell (NZM) wurde in [NK97] vorgestellt. Ziel dieser Modellform ist es, hybride Systeme durch die Verkopplung von ereignisdiskreten und kontinuierlichen Modellformen selbsterklärend darzustellen. Die Struktur des NZM ist in Abbildung 2.3 gezeigt.

Der kontinuierliche Systemanteil wird im NZM durch ein Erweitertes Zustandsraummodell (EZM) repräsentiert, der diskrete Systemanteil durch ein interpretiertes Petri-Netz (IPN). Das IPN erlaubt es – im Gegensatz zu automatenbasierten ereignisdiskreten Modellformen – Nebenläufigkeiten und Synchronisationen übersichtlich darzustellen [Cas93]. Die beiden Modellanteile sind durch geeignete Interface-Komponenten gekoppelt. Diskrete und kontinuierliche Ein- und Ausgangsgrößen stellen die Verbindung zur Außenwelt her.

Der hybride Zustand x_H umfasst diskrete und kontinuierliche Zustandsvariablen und wird somit durch ein 2-Tupel von Vektoren dargestellt: $x_H = (\underline{x}_D, \underline{x}_K)$; er bewegt sich im hybriden Zustandsraum \mathcal{X}_H , der sich entsprechend aus dem Kreuzprodukt von diskretem und kontinuierlichem Zustandsraum zusammensetzt; $\mathcal{X}_H = \mathcal{X}_D \times \mathcal{X}_K$. Die ereignisdiskrete Systemdynamik wird durch ein interpretiertes Petri-Netz [KQ88] nachgebildet, das n_D Stellen besitzt und dessen Markierungsvektor \underline{x}_D den diskreten Modellzustand repräsentiert. Hierbei gibt die i -te Komponente die zum Zeitpunkt t in der Stelle x_{D_i} enthaltene Markenanzahl an. Somit bezeichnet $\underline{x}_D(0)$ den diskreten Anfangszustand. Die möglichen Markierungsvektoren des IPN sind in der endlichen Menge $\mathcal{X}_D = \{\underline{x}_D^{(1)}, \underline{x}_D^{(2)}, \dots, \underline{x}_D^{(|\mathcal{X}_D|)}\}$ zusammengefasst.

Zustandsübergänge des IPN werden durch Ereignisse ausgelöst; diese Ereignisse hängen von der Markierung des IPN und vom Wahrheitsgehalt der booleschen Bedingungen

$$\underline{\beta}(t) = \underline{f}_\beta(\underline{u}_D(t), \underline{v}_D(t))$$

an den Transitionen des IPN ab, wobei die booleschen Bedingungen wiederum Funktionen von den p_D diskreten Eingangsgrößen \underline{u}_D und vom Ausgangsvektor \underline{v}_D des Kontinuierlich/Diskreten Interfaces (K/D-Interface) sind. Der diskrete Eingangsvektor \underline{u}_D ist hierbei stückweise konstant.

Die j -te Transition schaltet unmittelbar, wenn sie bezüglich der Markenbelegung des Netzes aktiviert und der ihr zugeordnete boolesche Schaltausdruck β_j wahr ist (Schaltzwang). Falls mehrere Transitionen nebeneinander schaltfähig sind, schalten diese gleichzeitig. Wird ein Schaltvorgang unmittelbar durch die Änderung von diskreten Eingangsgrößen initiiert, so spricht man von einem externen Ereignis. Führt hingegen die Änderung von internen Verknüpfungsgrößen zu einem Schalten, entspricht dies einem internen Ereignis. Der diskrete Folgezustand

$$\underline{x}_D(t) = \underline{f}_{Dx}(\underline{x}_D(t^-), \underline{\beta}(t^-))$$

hängt dementsprechend vom aktuellen diskreten Zustand und den booleschen Schaltausdrücken ab.

Innerhalb des Diskret/Kontinuierlichen Interfaces (D/K-Interface) wird die ausgangsseitige Interpretation durch die Abbildung \underline{f}_{DK} realisiert, die abhängig vom diskreten Zustand \underline{x}_D den Parametervektor

$$\underline{v}_K(t) = \underline{f}_{DK}(\underline{x}_D(t))$$

festlegt. Die m_K Komponenten von \underline{v}_K ergeben sich aus den Stellenbelegungen \underline{x}_D des IPN durch algebraische oder logische Verknüpfungen und können binäre, aber durchaus auch reelle Werte annehmen.

Da eine Änderung des diskreten Zustandes \underline{x}_D immer nur zu diskreten Zeitpunkten t_k auftritt, kann der Verlauf desselben auch in Form einer Zustandsfolge ($\underline{x}_D(k)$) beschrieben werden, wobei $\underline{x}_D(k)$ den diskreten Zustand nach dem k -ten Schalten angibt. Das Netz befindet sich im Zeitintervall $[t_k, t_{k+1})$ im Zustand $\underline{x}_D(k)$. Sowohl $\underline{x}_D(t)$ als auch $\underline{v}_K(t)$ sind damit stückweise konstant:

$$\underline{x}_D(t) = \underline{x}_D(t_k) = \underline{x}_D(k), \quad \underline{v}_K(t) = \underline{v}_K(t_k), \quad \forall t \in [t_k, t_{k+1}).$$

Das (D/K-Interface) beeinflusst über \underline{v}_K zusammen mit den p_K äußeren kontinuierlichen Eingangsgrößen \underline{u}_K den kontinuierlichen Zustandsvektor \underline{x}_K , der den Anfangswert $\underline{x}_K(0)$ besitzt und der der Dynamik eines Systems von n_K Differentialgleichungen

$$\dot{\underline{x}}_K(t) = \underline{f}_{Kx}(\underline{x}_K(t), \underline{v}_K(t), \underline{u}_K(t))$$

unterworfen ist. Bei bestimmten Änderungen des diskreten Zustandes ist eine Re-initialisierung des kontinuierlichen Zustandsvektors

$$\underline{x}_K(t_k) = \underline{g}(\underline{x}_K(t_k^-), \underline{v}_K(t_k), \underline{v}_K(t_k^-))$$

möglich, wobei $\underline{x}_K(t_k^-)$ und $\underline{v}_K(t_k^-)$ die linksseitigen Grenzwerte zum Zeitpunkt t_k bezeichnen. Nicht jeder Wechsel des diskreten Zustandes muss zu einer Änderung von \underline{v}_K führen und sich damit auf das kontinuierliche Teilsystem auswirken.

Der kontinuierliche Modellanteil wirkt auf den ereignisdiskreten Modellanteil über das K/D Interface ein, das abhängig vom kontinuierlichen Zustand \underline{x}_K und von den kontinuierlichen Eingangsgrößen \underline{u}_K den booleschen Vektor

$$\underline{v}_D(t) = \underline{f}_{KD}(\underline{x}_K(t), \underline{u}_K(t))$$

zur Verfügung stellt. Damit werden Grenzwertüberschreitungen modelliert, die genau dann auftreten, wenn der Vektor, der aus kontinuierlichem Zustand und kontinuierlichen Eingangsgrößen zusammengesetzt ist, in eine definierte Menge Ω_i eintritt. Für die m_D Komponenten von \underline{v}_D ergibt sich:

$$v_{Di}(t) = \begin{cases} 1 & \text{falls } \begin{pmatrix} \underline{x}_K(t) \\ \underline{u}_K(t) \end{pmatrix} \in \Omega_i \\ 0 & \text{sonst} \end{cases} \quad i = 1, \dots, m_D.$$

Abhängig vom aktuellen diskreten Zustand kann eine solche Grenzwertüberschreitung ein internes Ereignis auslösen.

Formal wird das NZM also durch das Tupel

$$\mathcal{H}\{\mathcal{T}, \mathcal{X}_D, \mathcal{X}_K, \mathcal{U}_D, \mathcal{U}_K, \mathcal{Y}_D, \mathcal{Y}_K, \underline{f}_{Dx}, \underline{f}_{\beta}, \underline{f}_{Dy}, \underline{f}_{Kx}, \underline{f}_{Ky}, \underline{g}_K, \underline{f}_{DK}, \underline{f}_{KD}\}$$

beschrieben. Die Zeitbasis des NZMs ist in der bisherigen Darstellung $\mathcal{T} = \mathbb{R}^+$. Hier sind auch Erweiterungen auf Darstellungen möglich, die eine diskrete Zeitbasis betrachten [HSK04a, HSK04b, HM03], was für diese Arbeit bei der Identifikation hybrider Systeme interessant ist. Der diskrete Zustandsraum \mathcal{X}_D wird durch eine Teilmenge von \mathbb{N}^{n_D} dargestellt; der kontinuierliche Zustandsraum \mathcal{X}_K durch eine Teilmenge von \mathbb{R}^{n_K} .

Die Mengen Ω_i , die der Abbildung f_{KD} zugrunde liegen, müssen bestimmte Anforderungen erfüllen und werden im Folgenden *Triggermengen* genannt [Nen01].

Definition 2.16: Triggermenge

Die offene Menge Ω_i heißt *Triggermenge*, wenn ihr Rand $\partial\Omega$ eine $n_K + p_K - 1$ -dimensionale orientierte Untermannigfaltigkeit des $\mathbb{R}^{n_K+p_K}$ ist und somit den Raum $\mathcal{X}_K \times \mathcal{U}_K$ eindeutig in zwei Bereiche unterteilt.

Die Begriffe *orientiert* und *Untermannigfaltigkeit* bleiben hier der Anschauung überlassen; ihre genaue mathematische Bedeutung kann in [Nen01] und in [HSZ96] nachgelesen werden.

Eine Triggermenge Ω_i heißt *aktive Triggermenge* $\Omega_i^*(\underline{x}_D)$ unter dem diskreten Zustand \underline{x}_D , wenn das Eintreten der kontinuierlichen Trajektorie in Ω_i unter dem diskreten Zustand \underline{x}_D einen diskreten Zustandswechsel auslöst.

Die Vereinigung aller aktiven Triggermengen wird mit $\Omega^*(\underline{x}_D) = \bigcup_i \Omega_i^*(\underline{x}_D)$ bezeichnet. Das Komplement von $\Omega^*(\underline{x}_D)$ ist die Menge aller kontinuierlichen Zustände, die keinen diskreten Zustandswechsel auslösen

$$\Gamma(\underline{x}_D) = \{\underline{x}_K \mid \underline{x}_K \notin \Omega^*(\underline{x}_D)\}.$$

und heißt *Neutrale Menge*.

Mit Hilfe des Netz-Zustands-Modells ist es nun möglich, eine große Klasse hybrider Systeme zu modellieren [Nen01]. Vergleicht man diese Darstellungsform mit anderen Beschreibungsformen für hybride Systeme, so stellt man fest, dass diese anderen aus der Literatur bekannten Modellformen ebenbürtig ist. Von besonderem Vorteil ist ihr überschaubarer und modularer Aufbau, wodurch für den Anwender die Übersichtlichkeit gewährleistet ist. Deshalb wird diese Modellform in der vorliegenden Arbeit im Teil II bei der Diagnose hybrider Systeme angewandt.

Zum Abschluss dieses Kapitels werden zwei Modellformen vorgestellt, die aufgrund ihrer diskreten Zeitbasis zur Identifikation geeignet erscheinen. Die *Mixed Logical Dynamical Modelle* waren eine der ersten hybriden Modellformen die eine diskrete Zeitbasis besitzen, was für die Identifikation hybrider Systeme von Vorteil ist. Da

diese Modellform für die Identifikation allerdings zu unübersichtlich und zu komplex ist, werden anschließend die stückweise affinen Modelle eingeführt.

2.3.3.2 Mixed Logical Dynamical Modelle

In [BMM99, BM99] werden Modelle vorgestellt, die einen dynamischen und einen logikbasierten Anteil besitzen. Zusätzlich können auch lineare Beschränkungen in die Beschreibung mit aufgenommen werden. Diese Modellform wird deshalb von den Autoren *Mixed Logical Dynamical Systems* (MLD) genannt. Diese Modellform arbeitet in diskreter Zeit und scheint daher auf den ersten Blick für die Identifikation hybrider Systeme besonders geeignet zu sein.

Allgemein lassen sich diese Systeme in der Form

$$\begin{aligned} \underline{x}(k+1) &= \underline{A} \underline{x}(k) + \underline{B}_1 \underline{u}(k) + \underline{B}_2 \underline{\delta}(k) + \underline{B}_3 \underline{z}(k) \\ \underline{y}(k) &= \underline{C} \underline{x}(k) + \underline{D}_1 \underline{u}(k) + \underline{D}_3 \underline{z}(k) \\ \underline{E}_2 \underline{\delta}(k) + \underline{E}_3 \underline{z}(k) &\geq \underline{E}_1 \underline{u}(k) + \underline{E}_2 \underline{x}(k) \underline{e}_5 \end{aligned}$$

darstellen. Dabei gibt die erste Gleichung die Dynamik des Systems in Abhängigkeit der Zustandsgrößen x_i , den Eingangsgrößen u_i , den logischen Größen δ_i und den Hilfsgrößen z_i an. Die zweite Gleichung beschreibt den Systemausgang und in der dritten Gleichung werden logische und physikalische Randbedingungen in einem Ungleichungssystem verknüpft.

Für die Regelung hybrider Systeme lässt sich aus bekannten Gleichungen das MLD-Systemmodell – teilweise mit Softwareunterstützung – aufstellen. Darauf aufbauend ist es möglich, modellprädiktive Regler zu entwerfen und an verschiedenen Prozessen erfolgreich anzuwenden [TB02, TB03, Buc03].

Allerdings wird hier auch deutlich, dass aufgrund der Vielzahl von Gleichungen und Ungleichungen, die zu den dynamischen Gleichungen hinzu kommt, die Darstellung schnell unübersichtlich und die Anzahl der Parameter sehr groß wird. Zudem sind die Gleichungen nicht mehr einfach interpretierbar, weil die Hilfsgrößen z_i von logischen Größen und von kontinuierlichen Zustandsgrößen abhängen. Da der Anwender für die Identifikation eine sinnvolle Struktur vorgeben muss, bei der anschließend die Parameter zu bestimmen sind, wird es mit dieser Modellform schnell unmöglich, eine Identifikation durchzuführen,

Im nächsten Abschnitt wird die Modellform der stückweise affinen Systeme vorgestellt, die aufgrund ihrer übersichtlichen Struktur besser für die Identifikation von hybriden Systemen geeignet ist.

2.3.3.3 Stückweise affine Systeme

Stückweise affine Systeme und die Regelung solcher Systeme werden unter anderem in [Joh99, Joh03] betrachtet. Weiterhin ist die Analyse der Systemeigenschaften möglich, so dass für die Analyse der Steuerbarkeit und Beobachtbarkeit bereits Methoden angegeben werden konnten [HVS01, CVS04]. Darüber hinaus ist die Modellierungsmächtigkeit dieser Modellform äquivalent zu der der MLD-Systeme, wie in [HDB01] gezeigt wurde. Da sich die stückweise affinen Systeme zudem leicht darstellen lassen, eignen sie sich hervorragend zur Identifikation hybrider Systeme.

Da allerdings keine einheitliche Definitionen in der Literatur zu finden sind, wird hier zunächst eine Definition angegeben.

Definition 2.17: Zeitdiskretes stückweise affines System

Ein stückweise affines System \mathcal{M} ist auf seinem Definitionsbereich $\mathcal{I} = \mathcal{X} \times \mathcal{U}$, dem Kreuzprodukt seines Zustandsraumes und Eingangsraumes, definiert. Dabei sind die Mengen $\mathcal{X} \subset \mathbb{R}^n$ und $\mathcal{U} \subset \mathbb{R}^p$ Teilmengen von euklidischen Vektorräumen.

Die M affinen Teilmodelle

$$\underline{x}(k+1) = \underline{A}_j \underline{x}(k) + \underline{B}_j \underline{u}(k) + \underline{c}_j, \quad j = 1, \dots, M$$

sind jeweils auf einer bestimmten Teilmenge des Definitionsbereichs gültig. Diese Teilmengen \mathcal{I}_j müssen eine vollständige und paarweise disjunkte Überdeckung von \mathcal{I} bilden:

$$\bigcup_{j=1}^M \mathcal{I}_j = \mathcal{I} \quad \text{und} \quad \mathcal{I}_i \cap \mathcal{I}_j = \emptyset \quad \forall i \neq j.$$

Oft werden weitere Eigenschaften für die Teilmengen \mathcal{I}_i gefordert, wie z.B. dass die Gebiete Polytope darstellen müssen. In dieser Arbeit reicht es allerdings aus, wenn die Gebiete \mathcal{I}_j zusammenhängende Teilmengen des Grundbereiches \mathcal{I} sind.

Befindet sich der Systemzustand in einem bestimmten Teilgebiet \mathcal{I}_k des Systems, so entwickelt sich der Zustand nach der dort gültigen Systemdynamik, bis er in ein anderes Gebiet \mathcal{I}_l eintritt.

Da bei zeitdiskreten stückweise affinen Systemen der Zustand und die Eingangsgrößen zwischen den Abtastzeitpunkten stückweise konstant bleiben, muss definiert werden, was passiert, wenn der Folgezustand oder die folgende Eingangsgröße in einem anderen Gültigkeitsbereich liegen als die aktuellen Größen. In dieser Arbeit wird festgelegt, dass ausschließlich der aktuelle Zustand und die aktuelle Eingangsgröße zur Auswahl des Teilmodelles herangezogen wird, mit dem der Folgezustand berechnet wird.

Die Eigenschaft, dass der Systemzustand über der Zeit betrachtet stückweise konstant bleibt, ermöglicht es auch, Sprünge des Zustandes mit stückweise affinen zeitdiskreten Systemen zu modellieren. Zur Darstellung von Sprüngen im Systemzustand ist es notwendig, dass man bei diesen Systemen unstetige rechte Seiten der Differenzgleichungen zulässt [HM03]; dies führt dazu, dass eine kleine Änderung des Zustandes an den Unstetigkeitsstellen nicht mehr zu einer kleinen Änderung im Folgezustand führt; damit wird anschaulich klar, dass der Systemzustand an solchen Unstetigkeitsstellen springen kann.

Dynamikwechsel können hingegen mit stetigen Übergängen von einem Teilmodell in die angrenzenden Teilmodelle modelliert werden, so dass alle hybriden Phänomene mit dieser Modellform darstellbar sind.

Im Gegensatz zu den zeitdiskreten Systemen lassen sich bei stückweise affinen Systemen mit kontinuierlicher Zeitbasis keine Sprünge modellieren, ohne zusätzliche Sprunggleichungen oder andere Formalismen zu benutzen [HM03, HSK04a, HSK04b, SKHK05].

Deswegen sind die stückweise affinen und zeitdiskreten Systeme für die Identifikation hybrider Systeme besonders geeignet.

Zusätzlich könnte eine Ausgangsgleichung für jedes Teilmodell betrachtet werden, was aber in dieser Arbeit nicht notwendig ist und deswegen nicht weiter vertieft wird.

Zusammenfassung

Zum besseren Verständnis der in dieser Arbeit verwendeten Modellformen wurden in diesem Kapitel zunächst einige Grundbegriffe eingeführt, die die Zusammenhänge von realem Prozess und seiner Abstraktion als formales System verdeutlichen. Die Definition eines allgemeinen dynamischen Systems ist anschließend Ausgangspunkt für die Erläuterung verschiedener aus der Literatur bekannter Modellformen und ihrer wesentlichen Eigenschaften, wobei die in dieser Arbeit gebrauchten Modellformen mit anderen Modellformen verglichen und eingeordnet werden. Da das Netz-Zustands-Modell und die stückweise affinen Systeme eine wesentliche Grundlage für diese Arbeit bilden, wurden sie ausführlich vorgestellt.

Teil I

Identifikation hybrider
Systeme

Kapitel 3

Grundlagen zur Identifikation

Wie bereits in Kapitel 2 dargelegt, ist die Modellbildung häufig der erste Schritt zur Lösung eines automatisierungstechnischen Problems. Da die nachfolgenden Schritte stark von der Klasse der gewonnenen Modelle abhängen, ist die Modellbildung auch ein sehr bedeutender Schritt.

In diesem Kapitel sollen einige Begriffe zur experimentellen Modellbildung wiederholt werden. Zur besseren Abgrenzung wird dabei zunächst die theoretische Modellbildung angesprochen, bevor die experimentelle Modellbildung, die einen der Schwerpunkte dieser Arbeit darstellt, erläutert wird.

In Abschnitt 3.2 werden einige grundlegende Techniken zur Identifikation vorgestellt und weiterführende Aufgaben diskutiert, die mit einem identifizierten Modell angegangen werden können.

Die Methode der kleinsten Quadrate wird in Abschnitt 3.3 im Zusammenhang mit der Identifikation linearer Systeme eingeführt, da die in dieser Arbeit entwickelten Methoden als eine Verallgemeinerung dieser Methode verstanden werden können; im darauf folgenden Abschnitt wird ein kurzer Überblick über Methoden zur Identifikation nichtlinearer Systeme gegeben.

Anschließend wird in Abschnitt 3.5 diskutiert, welche der bereits vorgestellten Modellformen besonders für die Identifikation hybrider Systeme geeignet ist.

3.1 Modellbildungsverfahren

Bei der Modellbildung kann zwischen zwei grundsätzlich verschiedenen Herangehensweisen unterschieden werden, nämlich zwischen der theoretischen Modellbildung und der experimentellen Modellbildung oder Identifikation. Die erste Vorgehensweise zeichnet sich dadurch aus, dass zur Modellbildung für ein gegebenes System bekannte Gesetzmäßigkeiten aus den Naturwissenschaften herangezogen werden. Dagegen ist

die zweite Vorgehensweise vor allem dadurch geprägt, dass die Modellbildung durch die Verarbeitung von Messdaten geschieht, die zuvor durch eine geeignete Anregung des zu modellierenden Systems gewonnen wurden.

Gleichwohl existieren fließende Übergänge zwischen den beiden Modellbildungsverfahren; für praktische Probleme werden oft physikalische Überlegungen genutzt, um die Struktur des Modells zu bestimmen; die unbekannt Parameter, die in der so erhaltenen Struktur noch enthalten sind, werden dann häufig mit Hilfe von Messungen bestimmt, also identifiziert.

3.1.1 Theoretische Modellbildung

Bei der *theoretischen Modellbildung* wird mit Hilfe von bekannten Zusammenhängen aus Physik und Chemie versucht, ein Modell aus einem System von Gleichungen herzuleiten. Es ergibt sich ein System gewöhnlicher und/oder partieller Differentialgleichungen, das zu einem theoretischen Modell mit bestimmter Struktur und bestimmten Parametern führt. Oft ist dieses Modell nichtlinearer Natur, von hoher Ordnung und unübersichtlich, so dass es für automatisierungstechnische Anwendungen meist vereinfacht werden muss, z.B. durch Linearisierungen um Arbeitspunkte oder durch Ordnungsreduktion. Zudem müssen einzelne Parameter des Modells – trotz der theoretischen Herleitung – durch Parameteridentifikation bestimmt werden. Da allerdings die genaue Modellstruktur theoretisch abgeleitet wurde und deshalb bekannt ist, können die Zusammenhänge zwischen den einzelnen Modellgrößen interpretiert werden. Deshalb spricht man bei dieser Art von Modellen auch von *White Box Modellen* [Ise92].

3.1.2 Experimentelle Modellbildung / Identifikation

Bei der *experimentellen Modellbildung* bzw. *Identifikation* wird das mathematische Modell aus Messungen der Ein- und Ausgangssignale gewonnen. Dieses experimentelle Modell ermöglicht weniger Einsicht in das Systemverhalten, da lediglich das Ein-/ Ausgangsverhalten betrachtet werden kann. Der genaue funktionale Zusammenhang zwischen dem identifizierten Modell und den physikalischen Daten des Systems bleibt, im Gegensatz zum theoretischen Modell, weitgehend unbekannt.

Bei der experimentellen Modellbildung unterscheidet man *parametrische* und *nicht-parametrische Verfahren*. Parametrische Verfahren erfordern theoretische Vorüberlegungen zur Modellstruktur des zu identifizierenden Systems; anschließend werden durch das Identifikationsverfahren die Parameter bestimmt. Bei dieser Vorgehensweise ist teilweise ein Einblick in innere Zusammenhänge möglich, weshalb man hier von *Grey Box Modellen* spricht.

Nicht-parametrische Verfahren hingegen erfordern keine Vorüberlegung bei ihrer Anwendung; das Ergebnis lässt allerdings auch keinerlei Einblicke in das Systemverhalten zu, weshalb man von *Black Box Modellen* spricht. Meist erhält man eine Übertragungsfunktion in Form einer Wertetabelle oder in Form des Bode- oder Nyquistdiagramms als Ergebnis, in der die Systemparameter bereits implizit enthalten sind. Bei dieser Identifikationstechnik ist offensichtlich, dass die Lösung weiterführender Aufgaben, aufbauend auf dieser Art von Modellen, sehr schwierig ist.

Die Abgrenzung zwischen White Box, Grey Box und Black Box Modellen ist unscharf und es existieren zahlreiche Übergangsformen [Ise92].

3.2 Methoden und Aufgaben der Identifikation

Ganz allgemein lässt sich die Aufgabe der Systemidentifikation als das Problem definieren, eine Relation zwischen gemessenen Datensätzen zu finden, die das beobachtete Verhalten beschreibt. In der Automatisierungstechnik ist man typischerweise daran interessiert, ein Modell zu erstellen, das das Verhalten zwischen den Eingangssignalen eines technischen Prozesses und seinen Ausgangssignalen wiedergibt. Dazu können lineare oder nichtlineare Ansätze verwendet werden. Beiden Ansätzen ist gemein, dass mit dem resultierenden Modell meist nur der erste Schritt einer automatisierungstechnischen Aufgabe gelöst ist. Die Modellierung stellt keinen Selbstzweck dar, sondern dient dazu, die weiteren Schritte der automatisierungstechnischen Aufgabe zu erleichtern. Ein Modell ermöglicht z.B. eine Prädiktion des Systemverhaltens über einen gewissen Zeithorizont, eine Simulation oder auch den Entwurf von Regelungen.

Um ein geeignetes Modell zu erstellen, wird häufig die Minimierung des Prädiktionsfehlers angewandt; dies wird im nächsten Abschnitt näher erläutert. Anschließend werden einige weiterführende Aufgaben wie Simulation, Systemanalyse und Fehlerdetektion erläutert, die auf der Identifikation aufbauen [Nel01].

3.2.1 Prädiktion

Mit Hilfe einer Prädiktion möchte man aus den Messdaten der Ein- und Ausgangssignale vergangener Zeitpunkte und aus den aktuellen Signalen zukünftige Ausgangssignale vorhersagen. Dabei unterscheidet man zwischen der *Ein-Schritt-Prädiktion*, die verbreitet angewandt wird, und der *Mehr-Schritt-Prädiktion*.

Bei der Ein-Schritt-Prädiktion n -ter Ordnung wird ein Modell verwendet, das aus den abgetasteten Eingangssignalen der letzten n Zeitschritte sowie aus den Ausgangssignalen der letzten n Zeitschritte das Ausgangssignal im nächsten Zeitschritt

vorhersagt. Geeignete Modelle für die Prädiktion erhält man, indem der Prädiktionsfehler für eine gegebene Modellstruktur und einen gegebenen Satz von Messdaten minimiert wird. Dabei handelt es sich um eine parametrische Methode zur Systemidentifikation. Als Vorwissen ist hierzu eine geeignete Modellstruktur notwendig; anschließend können die Parameter dieser Modellstruktur so angepasst werden, dass der Ein-Schritt-Prädiktionsfehler möglichst klein wird. Dazu können unterschiedliche Gütemaße zur Anwendung kommen. Üblicherweise wird der quadrierte Ein-Schritt-Prädiktionsfehler minimiert, allerdings können auch andere Normen verwendet werden.

Um nun Prädiktionen über größere Zeithorizonte durchführen zu können, sind zwei Vorgehensweisen möglich:

- Ein Modell, das für die Ein-Schritt-Prädiktion erstellt worden ist, kann für die Mehr-Schritt-Prädiktion mit l Schritten verwendet werden, indem l Ein-Schritt-Prädiktionen durchgeführt werden. Dazu werden die prädizierten Ausgangswerte eines Schrittes im nächsten Schritt verwendet. Diese Vorgehensweise hat den Nachteil, dass sich die Prädiktionsfehler der einzelnen Schritte akkumulieren können; zudem erhält man bei dieser Vorgehensweise ein System mit Rückkopplung, das instabil werden kann.
- Andererseits können Modelle speziell für die Mehr-Schritt-Prädiktion erstellt werden. Hierzu müssen alle Eingangssignale innerhalb des Prädiktionshorizontes mit betrachtet werden, so dass der Datenraum, der bei der Identifikation betrachtet wird, hochdimensional wird und somit meist nicht mehr handhabbar ist. Weiterhin können mit einem solchen Modell immer nur Prädiktionen über den gesamten Zeithorizont durchgeführt werden.

Aufgrund der größeren Flexibilität und der besseren Handhabbarkeit wird zumeist der Ansatz mit Ein-Schritt-Prädiktion verwendet. So wird auch in dieser Arbeit die Minimierung des Ein-Schritt-Prädiktionsfehlers angewandt, wobei der Prädiktionsfehler mit Hilfe der 1-Norm oder 2-Norm des Fehlervektors ermittelt wird.

Da bei der Identifikation abhängig von den darauf aufbauenden Aufgaben jeweils unterschiedliche Ansprüche an die Eigenschaften des identifizierten Systems gestellt werden, wird im Folgenden auf die Anforderungen im Anwendungskontext eingegangen.

3.2.2 Simulation

Verwendet man zur Simulation eines dynamischen Systems ein Modell, das durch die Minimierung eines Ein- oder Mehr-Schritt-Fehlers gewonnen wurde, so ist zu beachten, dass bei der Simulation des Modells die Prädiktion zurückgekoppelt wer-

den muss und somit die normalerweise fehlerbehaftete Prädiktion in die nächste Prädiktion einfließt. Dadurch können sich Fehler fortpflanzen und die Simulation unbrauchbar machen; im schlimmsten Falle kann es aufgrund der Rückkopplung dazu kommen, dass eine Simulation instabil wird.

Die Simulation stellt also hohe Ansprüche an die Genauigkeit des identifizierten Modells und wird daher in der Praxis oft zur Validierung des identifizierten Modells verwendet.

3.2.3 Systemanalyse

Verschiedene Systemeigenschaften lassen sich anhand eines Modells überprüfen, wobei zur Identifikation bereits die Stabilität und die Beobachtbarkeit des Systems voraus gesetzt werden müssen, da ansonsten eine Identifikation nicht möglich ist [Ise92].

Nelles schlägt in [Nel01] vor, durch die Identifikation Einsichten in das tatsächliche Systemverhalten, die Zusammenhänge innerer Systemgrößen und systeminterner Vorgänge zu erhalten, indem man unterschiedliche Modellansätze testet und wiederholt identifiziert, um so auf die inneren Zusammenhänge zu schließen. Dieses Vorgehen, bei dem Modellbildung und naturwissenschaftliche Einsichten sich gegenseitig befruchten, wurde in den Arbeiten [Sch03, HWK⁺04, HRK05] bei der Modellbildung für Feststoff-Brennstoffzellen und für das werkstoffwissenschaftliche Verständnis dieser Systeme erfolgreich angewandt.

3.2.4 Reglersynthese

Da moderne Verfahren zur Reglersynthese meist ein Modell voraussetzen, kann auch hier mit Hilfe eines identifizierten Modells die notwendige Grundlage geschaffen werden. Betrachtet man die Genauigkeit des Modells, die für den Reglerentwurf notwendig ist, so lässt sich feststellen, dass im Vergleich zu Modellen, die zur Simulation verwendet werden, oft ein weniger genaues Modell ausreicht, da bei dieser Anwendung keine Fehlerfortpflanzung stattfindet. Vielmehr werden die Prozessdaten in regelmäßigen Abständen durch Messung neu ermittelt, so dass im wesentlichen ein Modell für eine Ein-Schritt-Prädiktion benötigt wird.

3.2.5 Diagnose

Auch moderne Diagnoseverfahren arbeiten modellbasiert. Die dafür eingesetzten Modelle müssen allerdings eine vergleichsweise hohe Güte aufweisen; ähnlich wie bei der Simulation werden nämlich bei der Diagnose Prozessdaten und Modelldaten über

größere Zeiträume verglichen, so dass sich eine Fehlerfortpflanzung aufgrund ungenauer Modelle negativ auf den Diagnosealgorithmus auswirkt. Bei der Diagnose können Modelle ungenügender Güte schnell zu Fehldiagnosen führen.

Weiterhin arbeiten viele Verfahren zur Diagnose mit Identifikationstechniken, um Fehler detektieren und identifizieren zu können, die sich auf Modellparameter auswirken und diese verfälschen.

Aus dieser kurzen Zusammenstellung wird noch einmal deutlich, welche zentrale Bedeutung dem Modell in der Automatisierungstechnik zukommt. Daher werden systematische Methoden zur Erstellung von Modellen benötigt, mit deren Hilfe es anschließend möglich ist, die zahlreichen weiterführenden Aufgaben anzugehen. Deshalb werden in den nächsten Abschnitten zunächst einige Methoden zur Identifikation linearer und nichtlinearer Systeme vorgestellt. Die entsprechenden Modellformen sind bereits in Kapitel 2 angesprochen worden.

3.3 Identifikation linearer Systeme

Für die Identifikation linearer Systeme existiert ein breiter Kanon an bewährten Methoden, der sowohl parametrische Methoden als auch nicht-parametrische Methoden umfasst. Da in dieser Arbeit vor allen Dingen parametrische Methoden untersucht werden, die zudem noch auf den parametrischen linearen Methoden aufbauen, soll als bekannteste und wichtigste Methode die *Methode der kleinsten Quadrate* betrachtet werden [Ise92, Lju99]. Diese Methode hat aufgrund ihrer brillanten Grundidee, die auf C.F. Gauß [Gau09] zurückgeht, eine große Verbreitung bei der Identifikation parametrischer Modelle erfahren. Zahlreiche Anpassungen und Modifikationen des Grundalgorithmus existieren für die Identifikation dynamischer Systeme, da die Methode der kleinsten Quadrate bei solchen Systemen normalerweise keine konsistente und erwartungstreue Schätzung liefert. Deshalb wurde beispielsweise die *Methode der Hilfsvariablen* [Ise92, Lju99] entwickelt. Andere Erweiterungen der Methode beziehen sich auf die Behandlung zeitvarianter Systeme; auch bei nichtlinearen Systemen wird diese Methode verbreitet eingesetzt.

Die Idee des Verfahrens besteht darin, eine Kurve, die durch eine parametrisch angesetzte Funktion beschrieben wird, so durch einen gemessenen Datensatz zu legen, dass die Summe der Abstandsmaße zwischen der Kurve und den gemessenen Daten minimal wird. Gehen die Parameter der Kurve linear ein und verwendet man als Abstandsmaß die quadrierte 2-Norm des Fehlervektors, so ist eine geschlossene Lösung dieses Problems möglich.

Zunächst wird hier die statische Schätzung mit Hilfe der Methode der kleinsten Quadrate betrachtet.

3.3.1 Schätzung statischer linearer Modelle

Es sind die Datenpaare $\underline{x}(i), y(i) \quad \forall i = 1, \dots, N$ gegeben, für die ein lineares Modell $\underline{y} = \underline{A} \underline{x}$ geschätzt werden soll. Die Datenpaare lassen sich zu Messmatrizen zusammenfassen: $\underline{Y} = (\underline{y}(1), \dots, \underline{y}(N))$ und $\underline{X} = (\underline{x}(1), \dots, \underline{x}(N))$. Nimmt man an, dass die Messungen fehlerfrei sind und auch das lineare Modell das wirkliche Verhalten wiedergibt, so müssen die Messmatrizen die Modellgleichung

$$\underline{Y} = \underline{A} \underline{X}$$

erfüllen.

Da aber bei praktischen Anwendungen fast immer Messfehler vorhanden sind und zudem das angenommene Modell die Realität normalerweise nicht fehlerfrei wiedergibt, kann dieses Gleichungssystem nicht genau erfüllt werden, wenn aufgrund der Anzahl der Messungen mehr Gleichungen als Unbekannte vorhanden sind. In diesem Fall kann man das Gleichungssystem aber immer so lösen, dass die Summe der Fehlerquadrate minimal wird. Dazu kann die Pseudo-Inverse nach Moore und Penrose benutzt werden [Föl94]. Wendet man die Pseudo-Inverse auf die vorherige Gleichung an, so ergibt sich als Schätzung für die Parametermatrix \underline{A}

$$\hat{\underline{A}} = \underline{Y} \underline{X}^+, \quad (3.1)$$

wobei \underline{X}^+ die Pseudo-Inverse bezeichnet. Diese ergibt sich im vorliegenden Fall zu $\underline{X}^+ = \underline{X}^T (\underline{X} \underline{X}^T)^{-1}$, falls die Matrix $\underline{X} \underline{X}^T$ invertierbar ist, wovon in dieser Arbeit stets ausgegangen wird.

Dass diese Lösung $\hat{\underline{A}} = \underline{Y} \underline{X}^T (\underline{X} \underline{X}^T)^{-1}$ tatsächlich die Summe der Fehlerquadrate minimiert lässt sich leicht zeigen, indem man die Matrix \underline{A} zeilenweise schätzt. Dazu betrachtet man die Gleichung

$$\underline{E} = \underline{Y} - \underline{A} \underline{X},$$

wobei in der Matrix \underline{E} die Fehler zwischen Modell und Beobachtung zu finden sind. Betrachtet man nun die k -te Zeile der Fehlermatrix, so kann diese geschrieben werden als

$$\underline{e}_k^T = \underline{y}_k^T - \underline{a}_k^T \underline{X},$$

wobei \underline{a}_k^T die k -te Zeile der Matrix \underline{A} und \underline{y}_k^T die k -te Zeile der Matrix \underline{Y} darstellt.

Will man nun die Summe über die Fehlerquadrate durch geschickte Wahl der Komponenten von \underline{a}_k^T minimieren, so ist das Problem

$$\min \{ \underline{e}_k^T \underline{e}_k \} = \min \left\{ (\underline{y}_k^T - \underline{a}_k^T \underline{X})(\underline{y}_k - \underline{X}^T \underline{a}_k) \right\}$$

zu lösen. Notwendige Bedingung für ein Optimum dieses Problems ist, dass die erste

Ableitung der Zielfunktion null wird. Die Differentiation der Zielfunktion nach dem Parametervektor und Null setzen ergibt

$$2 \underline{a}_k^T \underline{X} \underline{X}^T - 2 \underline{y}_k^T \underline{X}^T = 0.$$

Ist die Matrix $\underline{X} \underline{X}^T$ invertierbar, so erhält man

$$\hat{\underline{a}}_k^T = \underline{y}_k^T \underline{X}^T (\underline{X} \underline{X}^T)^{-1}. \quad (3.2)$$

Aus dieser Gleichung lässt sich zeilenweise Gleichung (3.1) zusammensetzen.

Die hinreichende Bedingung für ein Minimum ist ebenfalls erfüllt, da die Matrix $\underline{X} \underline{X}^T$ positiv definit ist.

Damit eine Schätzung möglich ist, wurde angenommen, dass $\underline{X} \underline{X}^T$ invertierbar ist. Da \underline{X} die Matrix der Eingangsdaten ist, muss der Anwender durch die Auswahl der Messungen sicherstellen, dass diese Bedingung erfüllt wird, indem das zu untersuchende System ausreichend angeregt wird.

Betrachtet man weiter die stochastischen Eigenschaften der Schätzung wie die Erwartungstreue und die Konsistenz, so finden sich in [Lju99] und [Ise92] ausführliche Darstellungen. Dort sind für Konsistenz und Erwartungstreue folgende Voraussetzungen genannt:

- Die Elemente der Fehlervektoren \underline{e}_k^T sind stationär und mittelwertfrei verteilt.
- Die Eingangssignale $\underline{x}(i)$ sind exakt messbar.
- Die Fehler sind nicht mit den Eingangssignalen korreliert.

Unter diesen Voraussetzungen ist die Schätzung erwartungstreu, so dass die Erwartungswerte der geschätzten Parameter mit den tatsächlichen Parametern übereinstimmen. Außerdem ist die Schätzung konsistent; durch eine zunehmende Anzahl von Messwerten wird die Schätzung immer weiter verbessert, da die geschätzten Parameter stochastisch gegen die wahren Parameter konvergieren.

3.3.2 Schätzung dynamischer linearer Modelle

3.3.2.1 Schätzung linearer SISO-Modelle

Dieses Vorgehen zur Parameterschätzung bei statischen Systemen lässt sich durch geschickte Anordnung der Messdaten auch auf dynamische Systeme anwenden [Ise92, Lju99]. Wird hier das Modell als Differenzgleichung mit der Form

$$y(k) - a_1 y(k-1) - \dots - a_{n_a} y(k-n_a) = b_1 u(k-1) + \dots + b_{n_e} u(k-n_e)$$

angenommen, so kann diese Gleichung nach $y(k)$ aufgelöst werden und daher in der Form

$$y(k) = a_1 y(k-1) + \dots + a_{n_a} y(k-n_a) + b_1 u(k-1) + \dots + b_{n_e} u(k-n_e) = \underline{a}^T \underline{y} + \underline{b}^T \underline{u}$$

geschrieben werden. Dabei gilt $\underline{a}^T = (a_1, \dots, a_{n_a})$, $\underline{b}^T = (b_1, \dots, b_{n_e})$, $\underline{y}^T = (y(k-1), \dots, y(k-n_a))$ und $\underline{u}^T = (u(k-1), \dots, u(k-n_e))$. Fasst man nun die Modellparameter zur Parametermatrix $\underline{\Theta} = (\underline{a}^T, \underline{b}^T)$ und die Vektoren \underline{y} und \underline{u} zum Vektor $\underline{m}^T = (\underline{y}^T, \underline{u}^T)$ zusammen, so entsteht die Gleichung

$$y(k) = \underline{\Theta} \underline{m} .$$

Liegt die Messreihe $y(1), \dots, y(N)$ und $u(1), \dots, u(N)$ eines dynamischen Systems vor, so muss zur Schätzung der Modellparameter einer Differenzgleichung der Ordnung $n_a \geq n_e$ die Messmatrix

$$\underline{M} = \begin{pmatrix} y(1) & \cdots & y(N-n_a) \\ \vdots & & \vdots \\ y(n_a) & \cdots & y(N-1) \\ u(1) & \cdots & u(N-n_e) \\ \vdots & & \vdots \\ u(n_e) & \cdots & u(N-1) \end{pmatrix}$$

und der Vektor der Ausgangsgrößen

$$\underline{y}_A = (y(n_a + 1), \dots, y(N))$$

aufgestellt werden. Das Gleichungssystem, das anschließend im Sinne der Methode der kleinsten Quadrate nach der Parametermatrix $\underline{\Theta}$ zu lösen ist, lautet

$$\underline{y}_A = \underline{\Theta} \underline{M}.$$

Eine Schätzung für die Parameter erhält man bei Invertierbarkeit von $\underline{M} \underline{M}^T$ durch

$$\hat{\underline{\Theta}} = \underline{y}_A \underline{M}^T (\underline{M} \underline{M}^T)^{-1}.$$

Dieser Gleichung entspricht im statischen Fall Gleichung (3.2).

3.3.2.2 Schätzung von linearen Zustandsraummodellen

Geht man bei der Identifikation von Mehrgrößensystemen davon aus, dass alle Zustandsgrößen messbar sind, so ist eine Identifikation in Zustandsdarstellung mit

$$\underline{x}(k+1) = \underline{A} \underline{x}(k) + \underline{B} \underline{u}(k)$$

in folgender Weise möglich:

In der Messmatrix

$$\underline{M} = \begin{pmatrix} x_1(1) & \dots & x_n(1) & u_1(1) & \dots & u_p(1) \\ \vdots & & \vdots & \vdots & & \vdots \\ x_1(N-1) & \dots & x_n(N-1) & u_1(N-1) & \dots & u_p(N-1) \end{pmatrix}$$

werden die Zustandsgrößen sowie die Eingangsgrößen gesammelt

Die verzögerten Zustandsgrößen werden in einer zweiten Matrix

$$\underline{Y} = \begin{pmatrix} x_1(2) & \dots & x_n(2) \\ \vdots & & \vdots \\ x_1(N) & \dots & x_n(N) \end{pmatrix}$$

gesammelt.

Die Parametermatrix

$$\underline{\Theta} = \begin{pmatrix} \underline{A}^T \\ \underline{B}^T \end{pmatrix}$$

setzt sich hierbei aus der transponierten Dynamikmatrix und der transponierten Eingangsmatrix zusammen, wobei \underline{A} eine $n \times n$ und \underline{B} eine $n \times p$ Matrix ist.

Verwendet man hier die Schätzgleichung

$$\underline{Y} = \underline{M} \underline{\Theta} ,$$

so erhält man die geschätzte Parametermatrix $\hat{\underline{\Theta}}$ wiederum durch Anwendung der Pseudo-Inversen

$$\hat{\underline{\Theta}} = (\underline{M}^T \underline{M})^{-1} \underline{M}^T \underline{Y} .$$

3.3.2.3 Schätzung von affinen Zustandsraummodellen

Für die Schätzung affiner Systeme können die hier gezeigten Methoden leicht verwendet werden, indem ein Eingang hinzugefügt wird, der über die gesamte Messzeit

als konstant 1 angenommen wird. Damit erhält man für die Zustandsraummodelle eine veränderte Messmatrix

$$\tilde{M} = \begin{pmatrix} x_1(1) & \dots & x_n(1) & u_1(1) & \dots & u_p(1) & 1 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ x_1(N-1) & \dots & x_n(N-1) & u_1(N-1) & \dots & u_p(N-1) & 1 \end{pmatrix}$$

und eine veränderte Parametermatrix

$$\tilde{\Theta} = \begin{pmatrix} \underline{A}^T \\ \underline{B}^T \\ \underline{c}^T \end{pmatrix}.$$

Damit ist die Schätzung affiner Systeme möglich.

Betrachtet man bei der dynamischen Schätzung die Erwartungstreue, so wird deutlich, dass die Voraussetzungen, die bei der statischen Schätzung für die Erwartungstreue notwendig sind, hier nicht mehr erfüllt werden können: Die einzelnen Werte werden bei der dynamischen Schätzung verzögert und wieder verwendet, wodurch die Störungen und das Eingangssignal miteinander korreliert sind.

Um diese Unzulänglichkeiten bei der dynamischen Schätzung mit der Methode der kleinsten Quadrate zu kompensieren, wurde beispielsweise die Methode der Hilfsvariablen entwickelt, mit der es möglich ist, auch für dynamische Schätzungen erwartungstreue Ergebnisse zu erhalten [Lju99, Ise92].

Im Rahmen dieser Arbeit muss auf diese Verfahren nicht weiter eingegangen werden, da hier das Hauptaugenmerk auf der Zuordnung der Daten zu den unterschiedlichen Teilmodellen liegt; die Lösung dieser Aufgabe ist mit Erweiterungen zu der hier gezeigten Methode der kleinsten Quadrate möglich. Anschließend können die Methoden der Hilfsvariablen oder weitere Methoden zur erwartungstreuen Schätzung linearer dynamischer Modelle innerhalb der einzelnen Teilgebiete eingesetzt werden.

3.4 Identifikation nichtlinearer Systeme

Da die nichtlinearen Systeme eine Teilmenge der hybriden Systeme darstellen und bereits Identifikationsmethoden für nichtlineare Systeme vorgeschlagen wurden, erscheint es sinnvoll, diese Methoden zunächst zu betrachten und daraus eventuell durch geschickte Erweiterungen ein Verfahren zur Identifikation hybrider Systeme zu entwickeln.

Die Methoden zur Identifikation nichtlinearer Systeme lassen sich in zwei Gruppen einteilen:

Die erste Gruppe versucht dabei, das nichtlineare Modell durch global gültige Basisfunktionen darzustellen. In der zweiten Gruppe werden lokal gültige Modelle verwendet, zwischen denen umgeschaltet oder interpoliert wird. Einen guten Überblick über unterschiedliche Verfahren der nichtlinearen Identifikation bietet [Nel01].

Zur ersten Gruppe gehören beispielsweise polynomiale Modelle, neuronale Netze aber auch Hammerstein-Modelle. Viele dieser Identifikationsverfahren sind parametrische Verfahren, die sogar linear in den Parametern sind. Bei den Modellen, die linear in ihren Parametern sind, ist die Anwendung der Methode der kleinsten Quadrate möglich, um eine analytische Lösung für die Parameterschätzung zu erhalten. Die Linearität in den Parametern besitzt den weiteren Vorteil, dass das Ergebnis garantiert zu einem globalen Minimum führt. Andere Modelle, die nicht linear in ihren Parametern sind, können nur numerisch im Sinne der Methode der kleinsten Quadrate angepasst werden. Zudem tritt hier häufig das Problem auf, dass das Minimum nicht eindeutig ist und somit ein globales Minimum gesucht werden muss, was sehr aufwändig sein kann.

Zur zweiten Gruppe der nichtlinearen Modelle gehören beispielsweise Fuzzy und Neuro-Fuzzy Modelle, die abschnittsweise definiert sind und deswegen nur lokal gelten.

Zu den abschnittsweise definierten Modellen gehören auch die stückweise affinen Systeme, die in dieser Arbeit vor allem verwendet werden. Diese Modellform wurde bereits in Abschnitt 2.3.3 näher erläutert. Den Geltungsbereichen der einzelnen lokalen Modelle kommt in diesen Modellformen eine entscheidende Bedeutung zu. Um die einzelnen Geltungsbereiche zu beschreiben, werden häufig konvexe Teilgebiete verwendet; besonders verbreitet sind dabei Simplexe, die in einem n -dimensionalen euklidischen Raum durch $n+1$ Punkte festgelegt werden, wie es in Anhang B erläutert ist.

Verwendet man die Aufteilung in Simplexe zusammen mit stückweise affinen Systemen, so kann man immer ein stetiges Modell erzeugen, da ein affines Modell durch die Funktionswerte in den Ecken der Simplexe eindeutig beschrieben werden kann. Setzt man eine feste Aufteilung des zugrunde liegenden Raumes in Simplexe voraus, werden stückweise affine Modelle auch linear in ihren Parametern, wie in [UT96] gezeigt ist.

Möchte man die bisher bekannten nichtlinearen Identifikationsmethoden jedoch auf hybride Systeme anwenden, so stößt man auf mehrere Schwierigkeiten:

- Die bisherigen Methoden versuchen stetige Modelle zu erzeugen; zur Darstellung der hybriden Phänomene muss man allerdings Unstetigkeiten im Modell zulassen [HM03, HSK04a, HSK04b, SKHK05].
- Die Einteilung des Grundbereiches in Gültigkeitsbereiche für die einzelnen affinen Modelle wurde bisher meist heuristisch gelöst.

- Aufgrund der Unstetigkeiten existiert kein Modellansatz, der linear in den Parametern ist. Deshalb kann die Pseudo-Inverse nicht zur Schätzung der Parameter verwendet werden; vielmehr müssen numerische Methoden eingesetzt werden.

Aufgrund dieser Schwierigkeiten werden für die Identifikation hybrider Systeme neue Ansätze benötigt. Dazu wird im nächsten Abschnitt zunächst diskutiert, welche Modellform hierfür geeignet ist. Im nächsten Kapitel werden nach einem Literaturüberblick zwei neue Methoden zur Identifikation hybrider Systeme vorgestellt.

3.5 Hybride Systeme

Da die Identifikation bei der Zuhilfenahme parametrischer Modelle meist mit abgetasteten Größen erfolgt, muss zunächst die Frage geklärt werden, ob sich hybride dynamische Systeme und somit auch die hybriden Phänomene in diskreter Zeit darstellen lassen. Diese Frage wird in Abschnitt 3.5.1 beantwortet. Anschließend wird in Abschnitt 3.5.2 eine geeignete Modellform zur Identifikation hybrider Systeme vorgestellt.

3.5.1 Zeitdiskrete Darstellung hybrider Systeme

Zur zeitdiskreten Darstellung hybrider Systeme wurden erste Untersuchungen in [HM03] vorgenommen; dort wurden die Besonderheiten von hybriden Systemen in diskreter Zeit zunächst phänomenologisch beschrieben. Weitere Untersuchungen zur zeitdiskreten Darstellung hybrider Systeme sind in [Sch04b, HSK04a, HSK04b, SKHK05] zu finden. In diesen Arbeiten wird gezeigt, wie sich die hybriden Phänomene, die in [Bra95] erstmals beschrieben sind, mit Hilfe von Differenzgleichungen darstellen lassen. Zuvor wurden zwar bereits zeitdiskrete Systeme zur Darstellung hybrider Systeme beispielsweise in Form der MLD-Systeme benutzt [BMM99, BM99], allerdings existieren keine formalen Untersuchungen darüber, was beim Übergang von der zeitkontinuierlichen zur zeitdiskreten Darstellung zu beachten ist.

Ergebnis dieser Untersuchungen ist, dass sich abschnittsweise definierte Systemfunktionen bei Differenzgleichungen dazu eignen, die hybriden Phänomene darzustellen. Im Gegensatz zu den Differenzgleichungen, die bisher verbreitet in der Automatisierungstechnik zur Beschreibung nichtlinearer zeitdiskreter Systeme verwendet wurden, muss man bei der Darstellung hybrider Systeme in diskreter Zeit zulassen, dass die Systemfunktionen auch nicht differenzierbare Stellen und sogar Unstetigkeiten aufweisen können. Mit Hilfe der stetigen, aber nicht differenzierbaren Stellen gelingt es, Dynamikwechsel – also Umschaltungen in der Dynamik – darzustellen,

während die Unstetigkeitsstellen für die Darstellung von Sprüngen im Systemzustand benötigt werden.

Aufgrund der Tatsache, dass sich die hybriden Phänomene mit stückweise affinen Systemen darstellen lassen und diese Systemklasse auch schon bei der nichtlinearen Systemidentifikation erfolgreich eingesetzt wurde, wird in dieser Arbeit zur Identifikation hybrider Systeme auf stückweise affine Systeme zurückgegriffen. Im Gegensatz zu den stückweise affinen Systemen, die bei der nichtlinearen Systemidentifikation verbreitet eingesetzt werden [Nel01, Nel99, UT96], sind bei den hier verwendeten Systemen unstetige Übergänge zwischen den verschiedenen Teilmodellen explizit zugelassen. Dies erfordert allerdings komplexere Methoden zur Parameterschätzung.

3.5.2 Modellform für die Identifikation

Die Formalisierung des hybriden Identifikationsproblems wurde bereits in [MK02, MHK03b] vorgenommen; dort wurde auch eine geeignete Modellform vorgeschlagen.

Betrachtet man eine Menge gemessener Daten eines hybriden Systems $\underline{y}(1), \dots, \underline{y}(N)$ und die zugehörigen Anregungssignale $\underline{u}(1), \dots, \underline{u}(N-1)$, so ist es die Aufgabe der Identifikation, ein Modell \mathcal{M} zu finden, das das beobachtete Verhalten beschreibt und mit dem sich das beobachtete Verhalten möglichst gut reproduzieren lässt. Es wird also eine Abbildung gesucht, die den Definitionsbereich \mathcal{I} des Modells auf den Zustandsraum \mathcal{X} abbildet. Der Raum $\mathcal{I} \subset \mathbb{R}^n \times \mathbb{R}^p$ wird fortan auch als *Identifikationsraum* bezeichnet. Wie schon in Abschnitt 2.3.3.3 eingeführt, ist der Raum \mathcal{I} das Kreuzprodukt von Zustands- und Eingangsraum.

Mögliche Identifikationsverfahren müssen nun ein geeignetes zeitdiskretes Modell so parametrieren, dass es die Beobachtungen möglichst gut wiedergibt. Dazu soll in dieser Arbeit ein stückweise affines Modell \mathcal{M} verwendet werden, das aus M affinen Teilmodellen

$$\underline{x}(k+1) = \underline{A}_j \underline{x}(k) + \underline{B}_j \underline{u}(k) + \underline{c}_j, \quad j = 1, \dots, M$$

besteht. Die Messdaten $\underline{y}(k)$ werden dabei über die Messgleichung

$$\underline{y}(k) = \underline{x}(k) + \underline{\varepsilon}(k) \tag{3.3}$$

gewonnen, wobei $\underline{\varepsilon}$ ein bei Messungen unvermeidliches Rauschsignal darstellt. Für die Gültigkeitsbereiche der affinen Teilmodelle werden, wie bereits in Abschnitt 2.3.3 beschrieben, zusammenhängende Teilmengen \mathcal{I}_j von \mathcal{I} vorausgesetzt. Dabei bilden die \mathcal{I}_j eine vollständige und disjunkte Überdeckung von \mathcal{I} :

$$\bigcup_{j=1}^M \mathcal{I}_j = \mathcal{I} \quad \text{und} \quad \mathcal{I}_k \cap \mathcal{I}_j = \emptyset \quad \forall k \neq j.$$

Diese Voraussetzung ist für die Identifikation notwendig, da dadurch die einzelnen Teilgebiete mit Hilfe ihrer Randpunkte beschrieben werden können. Beim identifizierten Modell wird versucht, den Rand dieser Teilgebiete mit endlich vielen Punkten und dazwischen liegenden linearen Mannigfaltigkeiten anzunähern.

Für die Darstellung von Sprüngen im Zustand muss explizit zugelassen werden, dass die affinen zeitdiskreten Modelle an den Grenzen der Teilgebiete unstetig sein dürfen.

Die Verbreitung stückweise affiner Systeme für die Identifikation hybrider Systeme ist auch der Literatur zu entnehmen [FMLM01a, FMLM01b, FM03, FMLM03, Rol03, NTK05]; dort wird oft die Modelklasse der *piecewise affine autoregressive exogenous models* (PWARX) verwendet. Diese Modellform stellt einen Spezialfall der in dieser Arbeit verwendeten Zustandsraumdarstellung dar. Die hier entwickelten Methoden lassen sich ohne Probleme auf die PWARX-Modelle anpassen, da beide Modellformen anschaulich als stückweise definierte lineare Mannigfaltigkeiten in mehrdimensionalen euklidischen Räumen interpretiert werden können.

Zusammenfassung

In Kapitel 3 sind die relevanten Grundlagen und Aspekte für die Identifikation dynamischer Systeme zusammengefasst. Zunächst wurde eine Einordnung der Identifikationsaufgabe in den automatisierungstechnischen Kontext vorgenommen. Die Methode der kleinsten Quadrate wurde vorgestellt und anschließend der Einsatz von nichtlinearen Methoden zur Identifikation hybrider Systeme diskutiert. Als Abschluss dieses Kapitels wurden die hybriden Systeme und ihr Verhalten im Hinblick auf die Identifikation sowie eine hierfür geeignete Modellform erläutert.

Mit der Betrachtung einer geeigneten Modellform für die Identifikation hybrider Systeme sowie den in diesem Kapitel diskutierten Grundlagen kann im folgenden Kapitel die Entwicklung geeigneter Identifikationsalgorithmen für stückweise affine Systeme angegangen werden.

Kapitel 4

Methoden zur Identifikation stückweise affiner Systeme

Die Identifikation hybrider und stückweise affiner Systeme hat in den letzten Jahren beträchtliches Interesse auf sich gezogen. Dies zeigt sich an der wachsenden Zahl von Veröffentlichungen auf diesem Gebiet. Da es allerdings keine einheitliche Definition stückweise affiner Systeme in der Literatur gibt, haben sich auch unterschiedliche Identifikationsverfahren entwickelt. Im vorherigen Kapitel wurde mit Definition 2.17 eine mögliche Definition für stückweise affine Systeme angegeben, auf die die hier vorgestellten Identifikationsmethoden für stückweise affine Systeme aufbauen.

In diesem Kapitel werden nun geeignete Identifikationsalgorithmen entwickelt. Dazu werden zunächst die besonderen Schwierigkeiten dargestellt, die bei der Identifikation stückweise affiner Systeme auftreten.

Anschließend wird ein Überblick über bekannte Verfahren aus der Literatur gegeben, die sich mit der stückweise affinen und hybriden Identifikation beschäftigen. Die Verfahren sowie ihre Vor- und Nachteile werden dabei diskutiert.

Aus dieser Diskussion resultieren einige Forderungen, die ein Identifikationsverfahren für stückweise affine Systeme erfüllen sollte. Die beiden anschließend vorgeschlagenen neuen Verfahren zur Identifikation von Systemen dieser Klasse erfüllen die gestellten Forderungen weitgehend. Diese Verfahren werden in den beiden Abschnitten 4.3 und 4.4 dieses Kapitels vorgestellt.

4.1 Besonderheiten bei der Identifikation stückweise affiner Modelle

Die starke Verkopplung von ereignisgetriebenen und zeitgetriebenen Dynamikanteilen bei hybriden Systemen macht es unmöglich, die einzelnen Teile separat zu betrachten. Daher muss man bei der Identifikation hybrider Systeme stets beide Systemanteile gemeinsam betrachten. In [Hof99, FMLM01a, FM03, Rol03] werden Verfahren vorgeschlagen, die auf unterschiedliche Weise versuchen, mit dieser Schwierigkeit zurechtzukommen.

Da bei der Identifikation hybrider Systeme Umschaltungen in der Dynamik und auch Sprünge der Zustandsgrößen vorkommen können, müssen die entsprechenden Identifikationsverfahren dazu in der Lage sein, die Parameter für Modellformen zu schätzen, die diese hybriden Phänomene [Bra95] wiedergeben können. Als geeignete Modellformen haben sich dazu stückweise affine Systeme in diskreter Zeit erwiesen. In [HM03] wurden dazu erste anschauliche Untersuchungen vorgenommen; weitere Ansätze dazu finden sich in [HSK04a, HSK04b, SKHK05].

Um bei der Identifikation solcher Modelle die Parameter der einzelnen Teilmodelle eines stückweise affinen Systems zu bestimmen, müssen sowohl die Zugehörigkeiten der einzelnen Datenpunkte zu einem affinen Teilmodell als auch die Parameter der einzelnen Teilmodelle bestimmt werden. In Beispiel 4.1 wird ein einfacher Fall der stückweise affinen Identifikation angenommen, der verdeutlicht, dass die Berechnung der Modellparameter nicht unabhängig von den Zugehörigkeitsgebieten geleistet werden kann sondern simultan erfolgen muss, damit ein geeignetes Modell identifiziert werden kann:

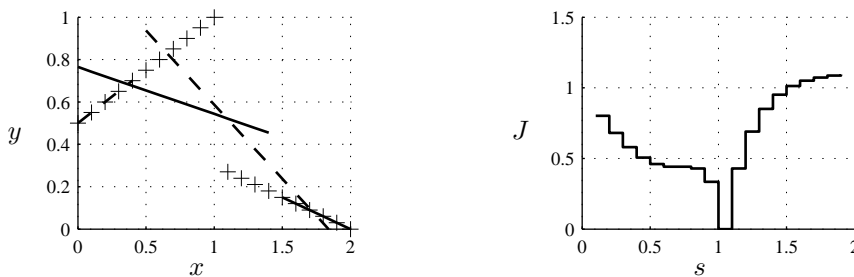


Abbildung 4.1: Identifikation eines stückweise affinen Systems mit zwei Teilsystemen und einer Zustandsgröße

Beispiel 4.1

Im linken Teil von Abbildung 4.1 sind die unverrauschten Messdaten eines zu identifizierenden stückweise affinen Systems, das aus zwei Teilmodellen besteht, als

Kreuze wiedergegeben. Die Aufgabe eines Identifikationsalgorithmus besteht darin, sowohl die Parameter der einzelnen Teilmodelle als auch die Zuordnung der einzelnen Daten zu einem Teilmodell zu ermitteln.

Dies kann geschehen, indem man ein Gütemaß

$$J = \sum_{i=1}^N (y(i) - \hat{y}(i))^2$$

minimiert.

Die Werte \hat{y} werden dabei durch ein stückweise affines Modell

$$\hat{y} = \begin{cases} a_1 x + c_1 & \text{für } x < s \\ a_2 x + c_2 & \text{für } x \geq s \end{cases}$$

berechnet. Die Gültigkeitsbereiche können im vorliegenden Fall durch den Parameter s bestimmt werden, der als Grenze zwischen den zwei Teilmodellen angesehen werden kann.

Wählt man die Grenze zwischen den beiden Teilmodellen zu $s_1 = 0,5$ und führt in den resultierenden Teilbereichen eine Parameterschätzung nach der Methode der kleinsten Quadrate durch, so ergeben sich die gestrichelt dargestellten Teilmodelle. Die mit durchgezogenen Linien eingezeichneten Teilmodelle ergeben sich, wenn man die Grenze zwischen den Teilmodellen zu $s_2 = 1,5$ wählt. Das Identifikationsproblem kann als Optimierungsproblem über alle Modellparameter a_1, a_2, c_1, c_2 und s aufgefasst werden. Es ist stetig differenzierbar in den Parametern der einzelnen Teilmodelle, aber stückweise konstant im Parameter s . Um dies zu zeigen, wurde im rechten Teil von Abbildung 4.1 der Verlauf des Gütemaßes über verschiedene Werte von s aufgetragen; dabei wurden die Parameter der affinen Teilmodelle jeweils optimal im Sinne der Methode der kleinsten Quadrate gewählt.

Das Beispiel zeigt, dass das Identifikationsproblem für stückweise affine Systeme bei naiver Vorgehensweise – also bei der vollständigen Enumeration aller Möglichkeiten – zu einem nicht differenzierbaren Problem führt.

Ein naiver Algorithmus zur Lösung des stückweise affinen Identifikationsproblems, der alle Möglichkeiten ausprobiert, müsste also folgendermaßen vorgehen:

Für jede mögliche Aufteilung der Datenpunkte in eine vorgegebene Anzahl von Teilmodellen werden die Parameter der Teilmodelle z.B. nach der Methode der kleinsten Quadrate bestimmt. Die Aufteilung, die zur kleinsten quadratischen Fehlersumme über alle Datenpunkte führt, wird als beste Aufteilung angenommen und die entsprechenden Parameter der Teilmodelle werden im stückweise affinen Modell

verwendet. Dieser naive Algorithmus besitzt jedoch massive Komplexitätsprobleme, da die Anzahl der möglichen Aufteilungen der Datenpunkte kombinatorisch mit der Anzahl der Datenpunkte und mit der Anzahl der Teilmodelle wächst.

Um diese kombinatorische Explosion des naiven Algorithmus zu vermeiden, wurden in [Hof99, FMLM01a, FMLM01b, FM03, FMLM03, Rol03] Verfahren entwickelt, die zwar meist nicht optimal sind, aber dennoch ein gutes Modell aus den vorhandenen Messdaten ermitteln.

4.2 Überblick über existierende Verfahren

4.2.1 Detektion von Umschaltungen durch Betrachtung des Prädiktionsfehlers

In [Hof99] wird ein Verfahren vorgestellt, das aufgrund der Betrachtung des Prädiktionsfehlers versucht, Datenpunkte zu ermitteln, die als mögliche Grenzpunkte zwischen verschiedenen affinen Dynamiken in Frage kommen. Dies geschieht durch die Anwendung eines Fensters, das entlang einer abgetasteten Trajektorie verschoben wird. Dabei müssen zwei Parameter festgelegt werden. Zum einen ist dies die Fenstergröße und zum anderen der Prädiktionsfehler, ab dem von einer wechselnden Dynamik ausgegangen werden soll. Diese Grenzpunkte werden im weiteren Verlauf des Verfahrens mit Hilfe eines speziell entwickelten Clusterverfahrens zu Gebieten zusammen gefasst, in denen das Umschalten zwischen den verschiedenen linearen Dynamiken stattfindet. Weiterhin werden in [Hof99] die Ermittlung des diskreten Modells und die Ermittlung von Sprunggleichungen beschrieben. Die beiden letzten Schritte sind bei der Identifikation von hybriden Systemen allerdings nicht notwendig, wenn man stückweise affine Modelle in diskreter Zeit als Modellform verwendet, wie in [HM03, HSK04a, SKHK05] bereits gezeigt wurde.

Nachteil dieser Methode ist, dass man zwei Parameter des Algorithmus durch gezieltes Probieren festlegen muss. So müssen die Fenstergröße und der Grenzwert für den Prädiktionsfehler festgelegt werden. Diese Parameter sind vorab schwer zu bestimmen und auch eine heuristische Systemkenntnis hilft bei diesen Parametern wenig, da es Parameter des Algorithmus und keine Parameter des zu ermittelnden Modells sind.

4.2.2 Identifikation stückweise affiner Systeme mittels Clustering im Parameterraum

Ein weiterer interessanter Ansatz ist in den Arbeiten [FMLM01a, FM03] zu finden.

Bei diesem Ansatz wird davon ausgegangen, dass die abgetasteten Daten in einem bestimmten Bereich des Zustandsraums vorhanden sind. Zu jedem Datenpunkt werden bei dieser Methode c Nachbarpunkte gesucht; für jede dieser Gruppen wird jeweils ein affines Modell geschätzt und die erhaltenen Parameter werden als Punkte im Parameterraum dargestellt. Gehören alle $c + 1$ Datenpunkte zum gleichen affinen linearen Modell, so werden die identifizierten Parameter der einzelnen Schätzungen wegen des zu erwartenden Rauscheinflusses zwar nicht identisch sein, aber dennoch auf einen kleinen Bereich im Parameterraum begrenzt sein. Im Parameterraum versucht man dann mit Hilfe eines modifizierten *K-means Algorithmus* die so erhaltenen Punkte zu clustern und auf diese Weise eine Zuordnung der Messdaten zu den affinen Teilmodellen vorzunehmen. Eine anschließende erneute Schätzung der Parameter mit der vorgenommenen Zuordnung verbessert im Sinne der Konsistenz die Güte der Schätzung.

Auch bei dieser Methode muss die Gruppengröße, die zur Schätzung der Parameter verwendet wird, durch den Benutzer gewählt werden. Für den modifizierten *K-means Algorithmus* muss ein weiterer Parameter, nämlich die Anzahl der Teilmodelle, festgelegt werden. Während eine geeignete Gruppengröße wiederum nicht aus a priori Wissen festgelegt werden kann, ist die Anzahl der Teilmodelle leichter aus heuristischem Systemwissen ableitbar.

4.2.3 Identifikation stückweise affiner Systeme mit gemischt-ganzzahliger linearer und quadratischer Optimierung

In [Rol03] wird ein Verfahren vorgestellt, das versucht die Methode der kleinsten Quadrate auf stückweise affine Systeme zu erweitern. Somit wird die Gebietseinteilung bzw. die Zuordnung der Daten zu unterschiedlichen Teilmodellen und die Parameterschätzung für die Teilmodelle simultan aufgrund einer Optimierung vorgenommen.

Dazu werden Hyperebenen, die den Identifikationsraum teilen und affine Modelle in den einzelnen konvexen Polyedern angenommen. Es wird ein Optimierungsproblem angegeben, das es ermöglicht, sowohl die Parameter der teilenden Hyperebenen als auch der einzelnen affinen Modelle zu ermitteln. Für das Optimierungsproblem wird gezeigt, dass es sich als gemischt-ganzzahliges quadratisches Optimierungsproblem bzw. als gemischt-ganzzahliges lineares Optimierungsproblem darstellen lässt und somit mit bekannten Verfahren gelöst werden kann.

Nachteil dieser Methode ist, dass aufgrund der angenommenen Modellstruktur nur stückweise affine Systeme als Modelle zugelassen werden können, die stetige Übergänge zu den Nachbarmodellen aufweisen. Damit ist die Darstellung von Sprüngen im Zustand nicht möglich, wie in Kapitel 2 bereits erläutert wurde.

4.2.4 Weitere Verfahren

In [NTK05] wird vorgeschlagen, eine Likelihood-Funktion für die Identifikation von stückweise affinen Systemen zu verwenden. Das Problem, dass die Likelihood-Funktion viele lokale Minima aufweist, bleibt dort allerdings ungelöst.

Vidal schlägt in [VSS03] vor, eine hybride Entkopplungsbedingung zu benutzen, bei der für jedes einzelne Datum entschieden wird, zu welchem Teilmodell es gehört, unabhängig von der topologischen Lage der Daten im Identifikationsraum. Dieses Verfahren bietet den Vorteil einer analytischen Lösung für Systeme mit bis zu vier Teilmodellen [Vid04]. Allerdings wird bei diesem Verfahren nicht beachtet, dass durch Störungen schnell die eindeutige Lösbarkeit des Gleichungssystems verloren geht. Die Lösung des Gleichungssystems im Sinne der Methode der kleinsten Quadrate führt nicht immer zu sinnvollen Lösungen, da auch bei dieser Problemformulierung lokale Minima nicht vermieden werden können.

4.2.5 Diskussion der existierenden Verfahren

Die bisherigen Verfahren lassen sich in unterschiedliche Gruppen einteilen:

- Die erste Verfahrensgruppe geht davon aus, dass ein System vorliegt, das durch ein stetiges mehrdimensionales Kennfeld angenähert werden kann. Dem entsprechend wird die Systemfunktion durch affine Modelle gebildet, die stetige Übergänge in die Nachbargebiete aufweisen.

Weiter lässt sich diese erste Verfahrensgruppe in Verfahren unterteilen, die die Stützstellen heuristisch verteilen wie z.B. in [UT96] (vergleiche Abschnitt 3.4) und in Verfahren, die die Gebietseinteilung aus einer Optimierung ermitteln [Rol03].

Aufgrund der Tatsache, dass diese Verfahren der ersten Gruppe keine unsteiligen Übergänge zwischen den Teilmodellen zulassen, können keine Sprünge im Zustand dargestellt werden. Deshalb sind diese Verfahren, die zur Identifikation nichtlinearer Systeme entwickelt wurden, zur Identifikation hybrider Systeme nicht geeignet.

- Die Verfahren von [Hof99], [FM03] und die Verfahren von [NTK05] und [Vid04] gehören einer zweiten Gruppe von Verfahren an, die die Darstellung von unsteiligen Übergängen ermöglicht.

[Hof99] und [FM03] gehen wie folgt vor:

- Zunächst werden Gebiete ermittelt, in denen die zeitgetriebene Dynamik jeweils mit einem affinen Modell beschrieben werden kann.

- Anschließend werden für die ermittelten Gebiete lokale Modelle identifiziert.

Oft ist diese Vorgehensweise sehr rechenzeitintensiv, da sehr viele affine Modelle geschätzt werden müssen, um die Stellen für einen Dynamikwechsel zu bestimmen. Weiterhin erschwert diese lokale Sichtweise das Einbringen von Vorwissen wie z.B. das gewünschte Abstraktionsniveau.

Das Verfahren von [Hof99] verwendet zudem einen unnötig komplizierten Modellansatz, da sowohl das ereignisdiskrete System als auch Sprunggleichungen separat identifiziert werden. Wie bereits in Abschnitt 3.5 dargelegt, ist die Darstellung von Zustandssprüngen mit Hilfe der Unstetigkeiten im zeitdiskreten Modell möglich, so dass keine gesonderten Sprunggleichungen identifiziert werden müssen [HM03].

Bei den Verfahren von [FM03] und [Vid04] ist es nicht möglich, zusammenhängende Gültigkeitsbereiche der einzelnen Teilmodelle zu erzwingen.

Die Verfahren von [NTK05] und [Vid04] besitzen das Problem, dass bei verrauschten Messdaten keine eindeutigen Lösungen für die dort formulierten Probleme existieren; daher werden die Verfahren in den zitierten Aufsätzen nur auf sehr einfache Beispiele angewandt.

Diese Diskussion zeigt, dass die Identifikation von stückweise affinen Systemen keineswegs eine triviale Aufgabe ist. Die schwerwiegendsten Nachteile der einzelnen Verfahren sind bereits genannt; zur Entwicklung besserer Verfahren für die Identifikation stückweise affiner Systeme muss deshalb auf folgende Merkmale geachtet werden:

- Die Verfahren sollten möglichst ohne Parameter auskommen, die vom Benutzer eingestellt werden müssen.
- Vorwissen des Anwenders sollte leicht in den Modellansatz integrierbar sein.
- Bei verrauschten Messdaten muss das Verfahren mit möglichen suboptimalen Lösungen zurecht kommen.
- Der Zusammenhang der Teilgebiete, auf denen die einzelnen Teilmodelle gelten, muss vom Anwender erzwingbar sein.

Die in den Abschnitten 4.3 und 4.4 beschriebenen Identifikationsalgorithmen stellen mögliche Lösungen des hybriden Identifikationsproblems dar, die gleichzeitig die gestellten Forderungen erfüllen.

In Abschnitt 4.4.2 wird zudem eine Möglichkeit vorgeschlagen, mit der zusammenhängende Gültigkeitsbereiche der Teilmodelle auch nachträglich erzwungen werden

können. Mit diesem Verfahren können auch die Verfahren von [FM03] und [Vid04] ergänzt werden.

4.3 Identifikation stückweise affiner Systeme durch lokale Kennfeldoptimierung

Da der Anwender eines Identifikationsalgorithmus in den meisten Fällen umfangreiches Vorwissen über das zu identifizierende System besitzt, das den Identifikationsprozess wirkungsvoll unterstützen kann, wurde das Verfahren zur lokalen Kennfeldoptimierung entwickelt, das es ermöglicht, dieses Vorwissen zu nutzen.

Die in diesem Abschnitt entwickelte Methodik zur Identifikation stückweise affiner Systeme besteht aus zwei Schritten:

- Zunächst werden die vorliegenden Messdaten geclustert, so dass in den einzelnen Clustern ein lokales Modell die Daten gut repräsentiert.
- Anschließend wird die Lage und die Form der einzelnen Cluster optimiert.

Das Vorwissen des Anwenders über bestimmte Prozesseigenschaften wird in den Identifikationsprozess eingebracht, bevor die Clusterung vorgenommen wird. Hierbei sind insbesondere Informationen über die Anzahl der Teilmodelle und ihre Form von Bedeutung, da hieraus die ungefähre Anzahl der Cluster bestimmt werden kann. Liegen weitere Informationen über lokale Modelle und deren Lage im Identifikationsraum vor, so kann auch dieses Vorwissen leicht in Form von Randbedingungen für die Clusterung formuliert werden.

Grundlage der lokalen Kennfeldoptimierung ist die Formalisierung des stückweise affinen Identifikationsproblems, wie es in [MK02, MHK03b] angegeben ist. Dort wurde die Methode der kleinsten Quadrate, wie sie für lineare Systeme zur Parameterschätzung bekannt ist, auf den unstetigen, stückweise affinen Fall erweitert.

Das Identifikationsproblem kann mit Hilfe der Modellform aus Abschnitt 3.5.2 als Optimierungsproblem formuliert werden:

Bei der Identifikation eines stückweise affinen Systems soll das Gütemaß

$$J = \sum_{i=1}^N \|\underline{y}(i) - \hat{y}(i)\| + \alpha M \quad (4.1)$$

minimiert werden. Dabei sind die $\underline{y}(i)$ die Messdaten, die durch verrauschte Messungen aus den Zustandsgrößen $\underline{x}(i)$ gewonnen werden; $\hat{y}(i)$ sind die Prädiktionswerte des stückweise affinen Modells und $\|\cdot\|$ bezeichnet eine verallgemeinerte Norm, bei

der die Dreiecksungleichung zu einer Monotonieeigenschaft abgeschwächt werden kann. Bei diesen verallgemeinerten Vektornormen wird in dieser Arbeit anstatt der Dreiecksungleichung gefordert, dass $\|\underline{x}\| > \|\underline{y}\| \Leftrightarrow \|\underline{x}\|_1 > \|\underline{y}\|_1$.

Das Gütemaß muss nun bezüglich der Parameter $\underline{A}_j, \underline{B}_j, \underline{c}_j$, aller möglichen vollständigen und disjunkten Überdeckungen des Identifikationsraumes \mathcal{I} , und bezüglich der Anzahl der Teilmodelle M minimiert werden. Der reelle Parameter $\alpha > 0$ gewichtet dabei die Anzahl der Teilmodelle. Wenn α zunimmt, so nimmt die Anzahl der Teilmodelle im Optimum ab und umgekehrt. Daher kann $1/\alpha$ als Gewichtungsfaktor für den Aufwand interpretiert werden, den der Anwender bei der Identifikation betreiben möchte.

Dieses Optimierungsproblem wirft allerdings zwei große Schwierigkeiten auf:

- Da die Anzahl $M \in \mathbb{N}$ der Teilmodelle ganzzahlig ist und im Gütemaß auftaucht, liegt ein gemischt-ganzzahliges Optimierungsproblem vor.
- Für jede mögliche Anzahl von Teilmodellen müssen alle möglichen vollständigen und disjunkten Überdeckungen von \mathcal{I} betrachtet werden. Diese bilden in reellen Räumen eine überabzählbare unendlich große Menge. Da mit diskreten Messdaten gearbeitet wird, ist zwar nur eine endliche Teilmenge von Interesse, jedoch wächst diese Menge kombinatorisch mit der Anzahl der Messdaten an.

In [MK02] wurde ein Verfahren vorgeschlagen, das die bereits beschriebene zweistufige Vorgehensweise aufgreift. Der dortige Ansatz versucht das beschriebene Optimierungsproblem direkt zu lösen, indem eine bestimmte Anzahl M von Teilmodellen vorgegeben wird und die Form der Teilgebiete auf Simplexe beschränkt wird. Schwierigkeiten mit diesem Verfahren treten vor allem bei großen Datenmengen auf, da die benötigte Rechenzeit immer noch zu stark anwächst. Um diese Schwierigkeiten zu umgehen, wurden die dort verwendeten Cluster- und Optimierungsverfahren weiterentwickelt [Hen02, Hod02] und in [MHK03b] veröffentlicht. Dieses Verfahren wird in diesem Abschnitt vorgestellt.

Dazu wird zunächst anhand des elementaren Beispiels 4.1 aufgezeigt, wie das Komplexitätsproblem der hybriden Identifikation durch das neue Verfahren verringert werden kann.

Bisher wurde an Beispiel 4.1 gezeigt, dass bei naivem Vorgehen ein unstetiges Optimierungsproblem gelöst werden muss. Liegt nun ein Identifikationsraum höherer Dimension als im Beispiel vor, so nimmt bei naiver Vorgehensweise die Anzahl der möglichen Unterteilungen stark zu. Aufgrund der großen Anzahl ist es nicht mehr möglich, alle möglichen Unterteilungen des Identifikationsraums in vernünftiger Zeit auszuprobieren und in der Optimierung zu berücksichtigen. Weiterhin ist es bei der Parameterschätzung mit Hilfe von Optimierungsverfahren von Vorteil, mit stetig differenzierbaren Funktionen zu arbeiten, da für solche Funktionen leistungsfähige

numerische Verfahren zur Verfügung stehen. Dagegen verursacht die Minimierung von nicht differenzierbaren oder unstetigen Funktionen erhebliche Schwierigkeiten.

Um die Probleme, die mit nicht differenzierbaren Gütemaßen verbunden sind, zu umgehen, existieren zwei mögliche Wege:

- Zum einen können die *diskreten Messpunkte durch eine stetig differenzierbare Hyperfläche* ersetzt werden, wie sie beispielsweise durch Spline-Interpolation entsteht; anschließend kann das Integral über die quadratische Abweichung zwischen der Hyperfläche und dem parametrisch vorliegenden Modell minimiert werden. Aufgrund dieser Modifikation ist das Gütemaß stetig differenzierbar, und die Minimierung kann beispielsweise mit Hilfe von Gradientenverfahren vorgenommen werden.
- Zum anderen können die *Unstetigkeiten im parametrischen Modell mit Hilfe von Sigmoid-Funktionen angenähert werden*; auch dadurch erhält man wieder ein stetig differenzierbares Gütemaß.

Im Folgenden wird nun die zweite Möglichkeit näher betrachtet.

4.3.0.1 Approximation einer Sprungfunktion

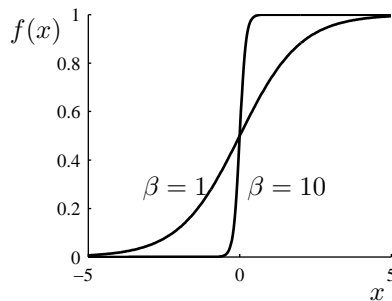


Abbildung 4.2: Sigmoid-Funktion für $\beta = 1$ und $\beta = 10$

Durch den Einsatz stückweise affiner Modelle, entstehen Modellfunktionen, die, wie im Beispiel 4.1 gezeigt ist, unstetig sein können. Verwendet man nun anstatt der unstetigen Modellfunktion eine sehr steile, aber dennoch stetig differenzierbare Approximationsfunktion, so kann der unstetige Übergang zwischen den Teilmodellen in Beispiel 4.1 durch die Minimierung eines stetig differenzierbaren quadratischen Gütemaßes ermittelt werden.

Dazu wird im oben angeführten Beispiel der sprungförmige Übergang von einem affinen Modell zum anderen durch einen stetig differenzierbaren Übergang ersetzt. Die Parameter der derart modifizierten Modellfunktion können anschließend durch

Minimierung eines stetig differenzierbaren quadratischen Gütemaßes ermittelt werden.

Zur Approximation der Sprungfunktion mit einer stetig differenzierbaren Funktion wird dabei die Sigmoid-Funktion verwendet. Die Sigmoid-Funktion

$$f(x) = \frac{1}{1 + e^{-\beta x}}, \quad \beta > 0$$

ist eine geeignete stetig differenzierbare Funktion zur Approximation der Sprungfunktion. Der Definitionsbereich $\mathcal{D}(x)$ ist gegeben durch $x \in \mathbb{R}$, der Wertebereich $\mathcal{W}(f)$ durch $\mathcal{W} = (0, 1) \subseteq \mathbb{R}$. Darüber hinaus ist die Funktion stetig differenzierbar und monoton steigend. Von besonderem Nutzen ist die Tatsache, dass $f(x)$ bei konstantem β sehr schnell gegen 0 bzw. gegen 1 strebt, wenn x von 0 verschieden ist. Die Steilheit des Übergangs kann durch den Parameter β im Exponenten der Exponentialfunktion beeinflusst werden. Es kann gezeigt werden, dass die Sigmoid-Funktion für $\beta \rightarrow \infty$ gegen die Einheitssprungfunktion konvergiert. Diese Eigenschaften sind in Abbildung 4.2 verdeutlicht.

Nimmt man hybride Systeme als Abstraktion nichtlinearer Systeme an, bei denen sehr steile, aber dennoch stetig differenzierbare Übergänge der Systemfunktion durch Sprünge dargestellt werden, so erscheint es vom praktischen Standpunkt aus sinnvoll, diese Art von Funktionen für die Identifikation hybrider oder schaltender Systeme einzusetzen.

4.3.0.2 Erzeugung der Fensterfunktionen

Um affine Modelle ein- und wieder ausblenden zu können, sind zwei Sprungfunktionen notwendig, die eine sogenannte Fensterfunktion bilden. Mit Hilfe der Sigmoid-Funktion und ihrer Eigenschaften kann man auch geeignet approximierte Fensterfunktionen w definieren.

Die Fensterfunktion

$$w(x) = \frac{1}{1 + e^{-\beta x}} \cdot \frac{1}{1 + e^{\beta(x-x_0)}}, \quad x_0 > 0 \quad (4.2)$$

entsteht dadurch, dass man die Sigmoid-Funktion zuerst an der Ordinatenachse spiegelt, anschließend verschiebt und mit der ursprünglichen Funktion multipliziert. In diesem Fall erhält man eine Fensterfunktion die auf dem Intervall $(0, x_0)$ näherungsweise 1 ist. Multipliziert man nun ein affines Modell mit einer solchen Fensterfunktion, so erhält man ein affines Modell, das nur auf einem Teilgebiet gültig ist. Mit Hilfe der Fensterfunktionen und der affinen Modelle kann so eine Parameterschätzung vorgenommen werden, indem man – analog zur Methode der kleinsten Quadrate – ein

Gütemaß minimiert. Das Gütemaß

$$J = \sum_{i=1}^N (y(i) - f_{\text{app}}(x(i)))^2$$

besteht auch hier aus der Summe über die quadratischen Abstände zwischen Messpunkten und Approximationsfunktion. Dabei sind mit $y(i)$ die Messdaten dargestellt und $f_{\text{app}}(x(i))$ ist der Wert, der durch die Approximationsfunktion an der Stelle x_i prädiziert wird. Fortan ersetzt also der Wert der Approximationsfunktion f_{app} die Werte $\hat{y}(i)$, die durch die stückweise affinen Modelle in Gleichung (4.1) prädiziert werden. Die Gewichtung für den Parameter M entfällt hier, da die Anzahl der Teilmodelle als gegeben angenommen wird.

Diese Approximationsfunktion ist ausreichend, um über \mathbb{R} eine Identifikation durchzuführen. Muss die Identifikation allerdings über einem höherdimensionalen Identifikationsraum durchgeführt werden, der mehrere Zustandsgrößen und Eingangsgrößen besitzt, so sind die Fensterfunktionen und die affinen Modelle auf die höherdimensionalen Räume zu erweitern. Man erhält dadurch eine räumlich verallgemeinerte Approximationsfunktion.

4.3.0.3 Verallgemeinerung der Approximationsfunktion

Um höherdimensionale Probleme behandeln zu können, benötigt man eine über dem Identifikationsraum $\mathcal{I} \subset \mathbb{R}^{n+p}$ definierte Approximationsfunktion. Die *vektorielle Approximationsfunktion*

$$\underline{f}_{\text{app}}(\underline{x}, \underline{u}) = \sum_{j=1}^M \underline{g}_j(\underline{x}, \underline{u}) w_j(\underline{x}, \underline{u}) \quad (4.3)$$

mit den vektoriellen Argumenten \underline{x} und \underline{u} ist in Verallgemeinerung der eindimensionalen Approximationsfunktion eine Summe über M Funktionen, wobei M wieder die Anzahl der affinen Teilmodelle darstellt.

Es werden auch hier wieder affine Modelle $\underline{g}_j(\underline{x}, \underline{u})$ verwendet, die mit Hilfe der zugehörigen Fensterfunktionen $w_j(\underline{x}, \underline{u})$ im entsprechenden Teilgebiet eingblendet werden. Jede Fensterfunktion

$$w_j(\underline{x}, \underline{u}) = \prod_m \frac{1}{1 + \exp(-\beta(\underline{\alpha}_{j,m,1}^T \underline{x} + \underline{\alpha}_{j,m,2}^T \underline{u} + \alpha_{j,m,3}))}. \quad (4.4)$$

wird durch Produktbildung über mehrere Sigmoidfunktionen erzeugt, analog zum skalaren Fall, der in Gleichung (4.2) gezeigt ist.

Im Exponenten der einzelnen Sigmoidfunktionen w_j stehen dabei die Gleichungen der Hyperebenen, die das jeweilige Teilgebiet j beranden. Damit ist es möglich, Fensterfunktionen für konvexe, von m Hyperebenen berandete Gebiete zu konstruieren.

Ersetzt man nun in Gleichung (4.3) die jeweiligen Fensterfunktionen w_j gemäß Gleichung (4.4) und die affine Abbildung $\underline{g}_j(\underline{x}, \underline{u})$ durch die ausführliche Darstellung

$$\underline{g}_j(\underline{x}, \underline{u}) = \underline{A}_j \underline{x} + \underline{B}_j \underline{u} + \underline{c}_j ,$$

so ergibt sich ein geschlossener Ausdruck für die Approximationsfunktion

$$\underline{f}_{\text{app}} = \sum_{j=1}^M \left[(\underline{A}_j \underline{x} + \underline{B}_j \underline{u} + \underline{c}_j) \prod_m \frac{1}{1 + \exp(-\beta(\underline{\alpha}_{j,m,1}^T \underline{x} + \underline{\alpha}_{j,m,2}^T \underline{u} + \alpha_{j,m,3}))} \right] .$$

Aufgabe der Identifikation ist es nun, die Parameter $\underline{A}_j, \underline{B}_j, \underline{c}_j, \underline{\alpha}_{j,m,1}, \underline{\alpha}_{j,m,2}$ und $\alpha_{j,m,3}$ dieser Approximationsfunktion zu bestimmen. Dazu wurde, wie eingangs erwähnt, ein zweistufiges Verfahren mit Clusterung und Optimierung entwickelt, das im Folgenden beschrieben wird.

4.3.1 Achsenparallele Clusterung im Identifikationsraum

In der ersten Stufe des Verfahrens wird eine geeignete *Clusterung der Daten* vorgenommen. Dazu wird der Identifikationsraum durch achsenparallele Hyperebenen in Hyperquader unterteilt. Die Anzahl dieser Hyperebenen wird in vielen Fällen durch Vorwissen bekannt sein.

Um bei der Datenclusterung ein stetig differenzierbares Gütemaß zu erhalten, wird die im vorherigen Abschnitt beschriebene Approximationsfunktion verwendet. Es werden ausschließlich Fensterfunktionen eingesetzt, die den Identifikationsraum achsenparallel unterteilen. Dadurch entsteht ein Gitter, das den Identifikationsraum in Hyperquader unterteilt, wodurch sich die Anzahl der Parameter drastisch reduziert.

Die optimale Lage der einzelnen Hyperebenen wird gewonnen, indem man mit Hilfe von numerischen Optimierungsalgorithmen das quadratische Gütemaß

$$J = \sum_{i=1}^N (\underline{y}(i) - \underline{\hat{y}}(i))^2$$

minimiert. Der Schätzwert $\underline{\hat{y}}(i)$ ist durch die Approximationsfunktion $\underline{f}_{\text{app}}(\underline{x}(i))$ gegeben.

Liegt ein System vor, das nur achsenparallel berandete Triggermengen und lokal jeweils eine affine Dynamik aufweist, so findet das vorgeschlagene Clusterverfahren

die entsprechenden Schaltflächen sehr genau. In den einzelnen Teilgebieten können anschließend die lokalen Modelle beispielsweise durch die Methode der kleinsten Quadrate geschätzt werden.

Liegt ein System vor, das nicht ausschließlich achsenparallel berandete Triggermengen aufweist, so muss der Clusterung der nun beschriebene Schritt folgen.

4.3.2 Optimierung von Form und Lage der Cluster

Die durch die Clusterung gewonnene Einteilung des Identifikationsraums in Hyperquader stellt nicht unbedingt die am besten geeignete Unterteilung des Identifikationsraums dar, da beispielsweise nicht alle Triggermengen achsenparallel berandete sein müssen. Daher werden die Eckpunkte der Hyperquader als Startlösung für eine weitere Optimierung verwendet.

Dazu muss zunächst die Einteilung der Hyperquader in Bereiche erfolgen, mit denen eine weitere Optimierung effizient möglich ist. Weiterhin soll es möglich sein, zwischen den einzelnen Gebieten für affine Modelle sowohl stetige als auch unstetige Übergänge in die Nachbargebiete zu erzeugen. Für diesen Zweck sind aus der geometrischen Datenverarbeitung Simplexe als geeignete Grundbereiche bekannt [Ede87]. Ein Simplex im n -dimensionalen Euklidischen Raum wird dabei durch $n + 1$ Punkte, die nicht auf einer Hyperebene liegen, definiert. Um diese Simplexe – ausgehend von dem Raster, das mit Hilfe der achsenparallelen Clusterung erzeugt wurde – zu erhalten, wird die Delaunay-Triangulierung eingesetzt. Weiterhin ist zur Entscheidung, ob ein einzelner Punkt im Inneren eines Simplex liegt, der Formalismus der Baryzentrischen Koordinaten notwendig. Beides ist in Anhang B beschrieben. Damit ist die Grundlage für eine sukzessive Veränderung der Position der Stützstellen gelegt, welche in den nächsten Abschnitten ausführlich beschrieben wird.

Der Optimierungsalgorithmus, der im Folgenden eingesetzt wird, besteht aus einigen Schritten, die so lange wiederholt werden, bis ein Abbruchkriterium erfüllt ist. Die Verschiebung der Stützstellen wird dabei in einem dieser Schritte durchgeführt. In Abbildung 4.3 ist ein Flussdiagramm dieses Algorithmus zu sehen; er besteht aus vier Schritten, denen die Überprüfung der Abbruchbedingung folgt. Diese vier Schritte werden in den folgenden Abschnitten gesondert aufgeführt und diskutiert.

Das Prinzip des Algorithmus ist angelehnt an das Prinzip des *Simulated Annealing* [Pha00]:

Aus der Werkstoffkunde ist bekannt, dass bei der Abkühlung von Schmelzen Kristalle unterschiedlicher Größe entstehen. Dabei sinkt mit der Temperatur die Gesamtenergie der Schmelze. Je nach Kristallisation werden sehr unterschiedliche lokale Energieminima erreicht. Bei schneller Abkühlung erreicht das System üblicherweise

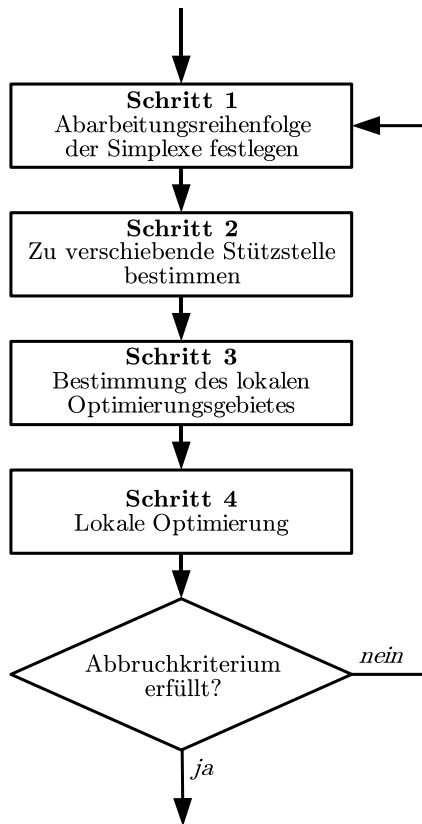


Abbildung 4.3: Flussdiagramm zur iterativen Verbesserung der Lage der Stützstellen nach dem Prinzip des Simulated Annealing

rasch ein lokales Minimum mit relativ hoher Energie; dagegen kann bei genügend langsamer Abkühlung das globale Energieminimum erreicht werden.

In dem entwickelten Algorithmus ist deshalb eine Strategie implementiert, die lokale Kristallisationskeime bildet indem Stützstellen, nachdem sie lokal optimiert wurden, durch die Eigenschaft „blockiert“ unbeweglich gemacht werden. Stellt sich nach der Bewertung heraus, dass diese Kristallisationskeime gut sind, so werden sie beibehalten und umgebende, noch frei bewegliche Stützstellen lagern sich um diese Kristallisationskeime an und werden blockiert. Umgekehrt werden Stützstellen, die bereits blockiert sind, wieder frei gegeben, falls sich in ihrer Umgebung die frei beweglichen Stützstellen weiter bewegen. Die einzelnen Schritte werden im Folgenden ausführlich erläutert.

4.3.2.1 Abbruchkriterium

Das Abbruchkriterium sorgt dafür, dass der Algorithmus abbricht, sobald durch eine Verschiebung der Stützstellen keine Verbesserung des Gesamtfehlers mehr möglich ist. Deshalb werden die Stützstellen entweder mit der Eigenschaft „blockiert“ oder „nicht blockiert“ versehen; blockierte Stützstellen werden durch den Optimierungsalgorithmus nicht verschoben. Solange durch eine lokale Verschiebung einer Stützstelle eine Verbesserung zu erzielen ist und somit eine Veränderung in der Konfiguration der Simplexecken stattfindet, werden die Stützstellen in der unmittelbaren Umgebung der verschobenen Stützstelle für weitere Verschiebungen frei gegeben. Wenn sich dagegen durch die Verschiebung einer Ecke der Fehler nicht mehr verringern lässt, so erhält die Stützstelle den Status blockiert. Somit ist zum einen sichergestellt, dass es keinen Abbruch gibt, solange noch durch eine lokale Verschiebung eine Verbesserung des Gesamtfehlers möglich ist, zum anderen wird dadurch gewährleistet, dass der Algorithmus letztlich terminiert, wie später noch erläutert wird.

4.3.2.2 Schritt 1: Abarbeitungsreihenfolge der Simplexe festlegen

Auch wenn das Gesamtkonzept darauf ausgerichtet ist, einen Gesamtfehler zu minimieren, so bestehen dennoch verschiedene Strategien diesem Ziel näher zu kommen. Da der vorgeschlagene Algorithmus iterativ vorgeht, muss zunächst eine Reihenfolge festgelegt werden, in der die Stützstellen verschoben werden.

Die am nächsten liegende Sortierung der Simplexe geschieht mit Hilfe der Varianz der enthaltenen Daten. Die Varianz ist ein Maß für die Streuung der Datenpunkte innerhalb eines Simplex, bezogen auf die Gleichung der geschätzten Hyperebene. Ist die Varianz eines Simplex im Vergleich zu anderen Simplexen groß, dann ist zu erwarten, dass die Aufteilung der Teilgebiete in diesem Simplex besonders schlecht ist; es könnte beispielsweise ein Sprung im Kennfeld mitten durch den Simplex verlaufen. Dadurch enthält der Simplex Datenpunkte, die zwei verschiedenen affinen Modellen zugeordnet werden müssen. Bei der Optimierung soll genau diese Trennung der Daten erreicht werden.

Für die Festlegung der Reihenfolge wird also immer zuerst der Simplex gewählt, der die größte Varianz aufweist, da möglicherweise der Rand einer Triggermenge innerhalb des Simplex liegt. Durch eine Verschiebung entsprechender Stützstellen wird dann versucht, die Ränder des Simplex möglichst auf den Rand einer Triggermenge zu legen.

4.3.2.3 Schritt 2: Zu verschiebende Stützstelle bestimmen

Nachdem ein Simplex ausgewählt wurde, muss entschieden werden, welche der nicht blockierten Ecken verschoben werden soll, um eine Verkleinerung des Gesamtfehlers zu erzielen.

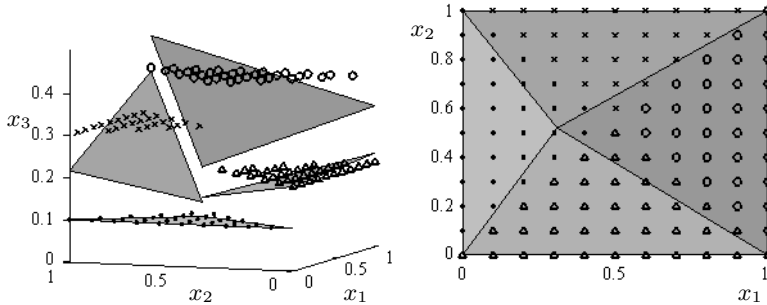


Abbildung 4.4: Suboptimale Aufteilung der Teilgebiete

In Abbildung 4.4 ist eine Beispielkonfiguration gezeigt. Im linken Teil ist eine perspektivische Darstellung gewählt, während im rechten Teil die Zuordnung der Daten zu den Simplexen gezeigt ist. Die Darstellung der Datenpunkte ist dabei so gewählt, dass Daten, die im Originalsystem zum gleichen Teilmodell gehören, die gleiche Darstellung besitzen. Hierbei ist zu erkennen, dass keine optimale Aufteilung des Gebietes vorliegt. Im Beispiel ist außerdem deutlich zu sehen, dass die Verschiebung der mittleren Stützstelle die größte Verbesserung bewirkt, da dadurch die Aufteilung des originalen Kennfeldes und die Aufteilung des identifizierten Modells miteinander in Übereinstimmung gebracht werden können.

Als Ansatzpunkt für einen Formalismus, der systematisch bestimmt, welche Stützstelle verschoben werden soll, dient die Betrachtung des Gradienten des Gütemaßes

$$J = \frac{1}{2} \sum_{i=1}^N (\underline{y}(i) - \hat{y}(i))^2,$$

bezüglich der Koordinaten der Eckpunkte des Simplex. Diese Ableitung kann allerdings nicht direkt ermittelt werden, da die Funktion J nach den Koordinaten der Eckpunkte nicht differenzierbar ist, wie bereits in Beispiel 4.1 gezeigt wurde; $\hat{y}(i)$ muss daher wiederum mit Hilfe der stetig differenzierbaren Approximationsfunktion dargestellt werden.

Die Fensterfunktionen der Approximationsfunktion sind an dieser Stelle nicht mehr auf den achsenparallelen Fall eingeschränkt, sondern stimmen mit den Begrenzungen der Simplexe überein.

4.3.2.4 Schritt 3: Bestimmung des lokalen Optimierungsgebietes

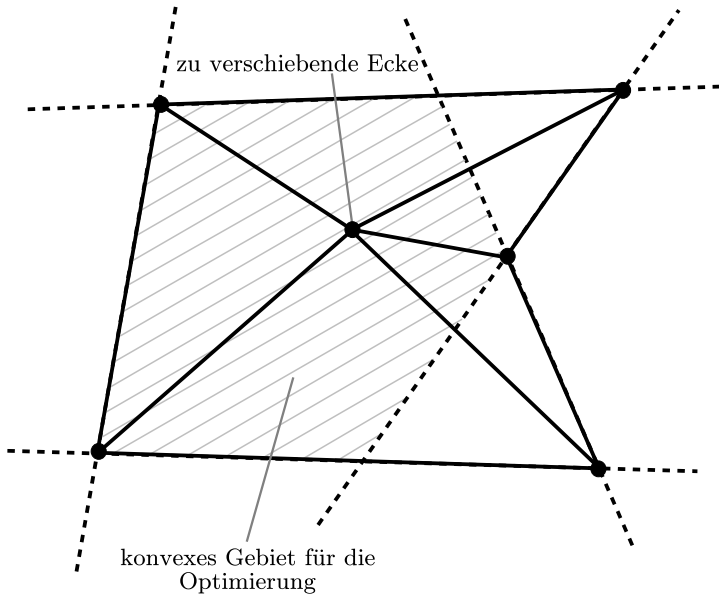


Abbildung 4.5: Auswahl eines konvexen Gebietes für die lokale Optimierung

Nachdem mit Hilfe des Gradienten die zu verschiebende Ecke ausgewählt wurde, wird zunächst ein Gebiet ausgewählt, innerhalb dessen die Ecke verschoben werden darf. Da beim Verschieben einer Ecke alle angrenzenden Simplexe betroffen sind, werden zunächst alle angrenzenden Simplexe zu einem Gebiet vereinigt, wie in Abbildung 4.5 gezeigt ist. Das dadurch entstehende Gebiet ist nicht notwendigerweise konvex. Da aber zur Optimierung nur ein konvexes Teilgebiet dieses Gebietes zugelassen werden kann, damit bei der Verschiebung der Ecke keine Überlappungen der Simplexe entstehen, wird ein konvexes Teilgebiet ausgewählt. Das konvexe Teilgebiet wird durch die Hyperebenen der beteiligten Simplexe eingeschlossen, die die zu verschiebende Ecke nicht beinhalten; diese Hyperebenen sind im zweidimensionalen Fall Geraden und in Abbildung 4.5 durch gestrichelte Linien dargestellt. Das dadurch definierte konvexe Teilgebiet ist durch Schraffur hervorgehoben.

Innerhalb dieses konvexen Teilgebietes wird die Ecke dann so verschoben, dass das quadratische Gütemaß minimal wird; dazu wird wiederum eine sigmoidale Approximationsfunktion benötigt. Diese Optimierung kann – wie bereits die Clusterung – mit Hilfe von numerischen Optimierungsverfahren ausgeführt werden.

4.3.2.5 Schritt 4: Lokale Optimierung

Die Ableitung des lokalen Gütemaßes

$$J_{\text{lokal}} = \frac{1}{2} \sum_{i=1}^N (\underline{y}(i) - \hat{y}(i))^2$$

kann gebildet werden, wenn der Term $\hat{y}(i)$ durch eine Funktion ersetzt wird, die stetig differenzierbar ist. Da durch die sigmoidale Approximationsfunktion eine solche Funktion vorhanden ist, besteht die Möglichkeit, Optimierungsverfahren einzusetzen, die Gradienteninformationen ausnutzen. Dies bringt gegenüber ungerichteten Suchverfahren eine erhebliche Aufwandsverkleinerung mit sich.

Mit einem solchen Verfahren wird das lokale Gütemaß, das mit Hilfe der Approximationsfunktion gebildet wird, auf dem ausgewählten Gebiet minimiert. Anschaulich bedeutet dies, dass bei der lokalen Optimierung neue Koordinaten für die zu verschiebende Stützstelle gesucht werden. Gleichzeitig werden neue Parameter für die affinen Teilmodelle in den angrenzenden Simplexen ermittelt.

Nachdem Schritt 4 ausgeführt ist, wird für jede der beteiligten Stützstellen der Status „blockiert“ und „nicht blockiert“ entsprechend dem Ergebnis der Optimierung neu gesetzt. Anschließend wird die Abbruchbedingung überprüft und gegebenenfalls eine Wiederholung der Schritte 1 bis 4 durchgeführt.

Für das Blockieren und Freigeben der Ecken gelten dabei folgende Regeln:

- Eine Stützstelle, die in Schritt 2 für die Verschiebung ausgewählt wurde und trotzdem bei einer lokalen Optimierung nicht verschoben wird, wird blockiert, da sie sich lokal betrachtet bereits im Optimum befindet.
- Eine Stützstelle, deren Nachbarstützstelle verschoben wurde, wird freigegeben, da sich das lokale Optimierungsgebiet dieser Stützstelle, wie es in Schritt 3 ermittelt wird, verändert hat.

4.3.2.6 Terminiertheit der Optimierung

Da der Wert des Gütemaßes über die gesamte Optimierung hinweg eine monoton fallende Funktion darstellt, kann mit dessen Hilfe gezeigt werden, dass die Optimierung auf jeden Fall terminiert. Dieser Sachverhalt lässt sich durch folgende Plausibilitätsbetrachtung veranschaulichen:

- Genau dann, wenn eine Ecke verschoben wird, nimmt das Gütemaß ab. Nur dabei können Ecken, die zuvor blockiert waren, wieder freigegeben werden.

- Das Gütemaß nimmt nur dann nicht ab, wenn durch eine lokale Veränderung keine Verbesserung erreicht werden kann. Dabei wird die betreffende Ecke blockiert.

Da das Gütemaß bei jedem Verschiebungsschritt abnimmt, kann der Fall einer sich wiederholenden Verschiebungssequenz nicht auftreten. Aufgrund der Beschränktheit des Gütemaßes nach unten muss der Fall, dass keine Verbesserung mehr möglich ist, zwangsläufig auftreten. Dies führt dazu, dass schließlich alle Ecken blockiert sind.

Anwendungsbeispiele zu dieser Methode sind in Abschnitt 5.1 sowie in [Kal03, MHK03b, MHK03a, MHK04] zu finden.

Das hier vorgeschlagene Verfahren erfüllt mehrere der oben genannten Forderungen. So sind nur sehr wenige Parameter vom Benutzer einzustellen, nämlich der Parameter β der Sigmoid-Funktionen und die Anzahl der Teilmodelle bzw. der Stützstellen. Zur Bestimmung der Anzahl der Stützstellen existieren auch systematische Verfahren, die beispielsweise in [NTK05] angesprochen sind. Aufgrund der Anschaulichkeit des Verfahrens ist Vorwissen leicht integrierbar; Teilgebiete des Identifikationsraumes können – falls Vorwissen über diese Gebiete vorliegt – leicht von der Optimierung ausgenommen werden. Das Verfahren ist ohne Schwierigkeiten auch bei verrauschten Messdaten anwendbar. Der Zusammenhang der Teilgebiete ist aufgrund der Struktur des Verfahrens gesichert.

Ein Nachteil des Verfahrens ist, dass relativ große Rechenzeiten in Kauf genommen werden müssen, weshalb beim zweiten vorgeschlagenen Verfahren das Hauptaugenmerk auf der Rechenzeit liegt.

4.4 Identifikation stückweise affiner Systeme mit globaler Optimierung

Im vorherigen Abschnitt wurde zur Identifikation stückweise affiner Systeme eine Methode aufgezeigt, die durch lokale Optimierungsschritte die Identifikation eines stückweise affinen Kennfeldes vornimmt. Dabei wurde die Schwierigkeit, die durch die un stetigen Kennfelder auftauchen, dadurch umgangen, dass die un stetigen Modellfunktionen durch stetig differenzierbare Modellfunktionen mit sehr steilen Übergängen angenähert wurden. Als Ergebnis dieser Vorgehensweise erhält man die Teilmodelle und die simplexförmigen Gültigkeitsbereiche dieser Teilmodelle.

Durch die Anwendung der gemischt-ganzzahligen Optimierung, wie in [Rol03] und in Abschnitt 4.2.3 beschrieben, erhält man ebenfalls Teilmodelle und die entsprechenden Gültigkeitsbereiche dieser Teilmodelle. Allerdings können hier verfahrensbedingt keine Unstetigkeiten zwischen Teilmodellen auftreten, was aber für die Darstellung

aller hybriden Phänomene erforderlich ist. Darüber hinaus wird durch die gemischt-ganzzahlige Programmierung ein Verfahren vorgeschlagen, das sehr aufwändig ist. Entsprechend sind in [Rol03] nur sehr einfache Anwendungsbeispiele zu finden.

Um ein Verfahren zu erhalten, das auch für komplexere Probleme einsetzbar ist, bedarf es also einer neuen Idee, wie sie in ihren Grundzügen bereits in [VSS03] zu finden ist:

Für die Identifikation von stückweise affinen Systemen interessiert nicht unbedingt der Grundbereich, auf dem ein Teilmodell definiert ist, vielmehr ist von Bedeutung, zu welchem Teilmodell einzelne Messdaten gehören. Die Rekonstruktion des Gültigkeitsbereiches für die einzelnen Teilmodelle kann erfolgen, nachdem die Parameter der Teilmodelle und die Zugehörigkeit der einzelnen Messdaten zu einem Teilmodell identifiziert sind.

Ein Ansatz, der diese Vorgehensweise aufgreift und für ideale Daten bzw. wenig Rauschen geeignet ist, ist in [MSK04] zu finden. Dort wird eine Formulierung des Identifikationsproblems vorgestellt, die auf ein stetig differenzierbares Gütemaß ohne Nebenbedingungen führt. Die Weiterentwicklung dieses Verfahrens resultiert in folgender Problemformulierung, die auch in [MK05] verwendet wird:

Anstatt einer einzigen Modellfunktion werden M affine Modelle verwendet, so dass für jeden Datenpunkt der Fehler

$$\underline{y}(i) - \hat{y}_j(i)$$

zu allen affinen Modellen $j = 1, \dots, M$ betrachtet werden muss.

Der Prädiktionwert des affinen Modells j für das i -te Datum wird dabei durch $\hat{y}_j(i)$ repräsentiert. Da ein Datenpunkt immer nur einem Teilmodell angehören kann, muss für jeden Punkt einer der Abstände ausgewählt werden. Dies geschieht mit Hilfe einer binären Zugehörigkeitsvariable z_{ij} ; zusätzlich muss über eine Nebenbedingung sichergestellt werden, dass immer genau ein Modell jedem Datum zugeordnet ist. Daraus ergibt sich für den statischen Fall die Problemformulierung

$$\min \left\{ \sum_{i=1}^N \sum_{j=1}^M \left\| \left(\underline{y}(i) - \hat{y}_j(i) \right) \right\| z_{ij} \right\}$$

s.t.

$$z_{ij} \in \{0, 1\}$$

$$\sum_{j=1}^M z_{ij} = 1 \quad \forall i .$$

Möchte man nun stückweise affine dynamische Modelle schätzen, so wird $\hat{y}_j(i)$ durch

die stückweise affine Modellbeschreibung ersetzt und die so prädierten Werte müssen mit den gemessenen Werten $\underline{y}(i+1)$ verglichen werden:

$$\begin{aligned} & \min \left\{ \sum_{i=1}^{N-1} \sum_{j=1}^M \left\| (\underline{y}(i+1) - \underline{A}_j \underline{y}(i) - \underline{B}_j \underline{u}(i) - \underline{c}_j) \right\| z_{ij} \right\} \\ & \text{s.t.} \\ & z_{ij} \in \{0, 1\} \\ & \sum_{j=1}^M z_{ij} = 1 \quad \forall i. \end{aligned} \tag{4.5}$$

Um dieses Problem zu lösen, können nun gemischt-ganzzahlige Methoden zum Einsatz kommen; damit wäre allerdings keine Reduktion der Komplexität des Problems erreicht. Daher wird zunächst die Relaxation des gemischt-ganzzahligen Optimierungsproblems genauer betrachtet; die Relaxation erhält man, indem man die Ganzzahligkeitsbedingung $z_{ij} \in \{0, 1\}$ fallen lässt und durch die Bedingung $z_{ij} \in [0, 1]$ ersetzt:

$$\begin{aligned} & \min \left\{ \sum_{i=1}^{N-1} \sum_{j=1}^M \left\| (\underline{y}(i+1) - \underline{A}_j \underline{y}(i) - \underline{B}_j \underline{u}(i) - \underline{c}_j) \right\| z_{ij} \right\} \\ & \text{s.t.} \\ & z_{ij} \in [0, 1] \\ & \sum_{j=1}^M z_{ij} = 1 \quad \forall i. \end{aligned} \tag{4.6}$$

Die Lösung eines gemischt-ganzzahligen Problems ist im Allgemeinen verschieden von der Lösung der Relaxation dieses Problems; es lassen sich nur untere Schranken für den optimalen Zielfunktionswert mit Hilfe der Relaxation berechnen. In dem vorliegenden Fall lässt sich aber zeigen, dass aus der optimalen Lösung des kontinuierlichen Problems sehr leicht eine optimale Lösung für das gemischt-ganzzahlige Problem konstruiert werden kann. Unter weiteren sehr schwachen Voraussetzungen ist die Lösung des kontinuierlichen Problems sogar ganzzahlig.

Um dies zu zeigen, wird zunächst die Frage nach der Existenz einer optimalen Lösung für die beiden Optimierungsprobleme beantwortet.

Setzt man voraus, dass die einzelnen Teilmodelle der affinen Systeme ausreichend angeregt sind, dass also genügend Messdaten vorliegen und alle Teilmodelle durch affine Funktionen mit endlichen reellen Parametern darstellbar sind, so existiert nach Satz 1.2.2 in [Alt02] ein globales Minimum für die beiden Optimierungsprobleme.

Die beiden Forderungen verhindern, dass die Zielfunktion gegen ein Infimum strebt, wenn bestimmte Einträge in den Parametermatrizen gegen unendlich streben.

Nun lässt sich die Äquivalenz der beiden Problemformulierungen zeigen.

Satz 4.1

Eine optimale Lösung von Problem (4.5) existiert genau dann, wenn auch eine optimale Lösung für Problem (4.6) existiert.

Existiert eine optimale Lösung, so gilt, dass der optimale Zielfunktionswert von Problem (4.5) mit dem optimalen Zielfunktionswert des zugehörigen relaxierten Problems (4.6) übereinstimmt.

Beweis

Es wird angenommen, dass eine optimale Lösung (*) mit den Parametern \underline{A}_j^* , \underline{B}_j^* , \underline{c}_j^* und \underline{Z}^* von Problem (4.5) existiert. Lösung (*) kann in Problem (4.6) eingesetzt werden und ist für dieses Problem zulässig.

Um zu zeigen, dass die für Problem (4.6) zulässige Lösung (*) auch optimal ist, nimmt man an, dass eine optimale Lösung (***) mit den Parametern \underline{A}_j^{**} , \underline{B}_j^{**} , \underline{c}_j^{**} und \underline{Z}^{**} für das relaxierte Problem existiert, die einen kleineren Zielfunktionswert erzielt als Lösung (*).

Aus Lösung (***) kann man nun eine in den Elementen \underline{Z} ganzzahlige optimale Lösung zu Problem (4.6) konstruieren, indem man für jeden Datenpunkt k die Teilmodelle, die bezüglich der verwendeten Norm minimalen Fehler zum betrachteten Datenpunkt haben, in den Mengen $M_k \subset \{1, \dots, M\}$ sammelt. Dann können bei jeder Menge M_k zwei Fälle auftreten:

1. Die Menge M_k besteht aus genau einem Element l . In diesem Fall ist der zugehörige Parameter z_{kl} bereits 1 und die restlichen Parameter z_{kj} sind 0. Wäre dies nicht der Fall, dann könnte der Zielfunktionswert weiter verringert werden, indem der Parameter z_{kl} um $\varepsilon > 0$ vergrößert wird; eines der anderen $z_{kj} > 0$ müsste entsprechend um ε verringert werden, damit die Lösung zulässig bleibt. Dann wäre Lösung (***) aber nicht optimal für Problem (4.6) und ein Widerspruch zu den Annahmen läge vor. Daher muss $z_{kl} = 1$ und die übrigen $z_{kj} = 0$ sein.
2. Die Menge M_k besteht im zweiten Fall für den Messpunkt k aus mehr als einem Element. Das bedeutet, dass diese Punkte zu mehreren Teilmodellen einen gleich großen Fehler aufweisen. Deshalb ändert sich der Wert der Gütefunktion nicht, wenn die Zugehörigkeiten zwischen diesen nächsten, gleich weit entfernten Modellen verändert werden, solange die Summe dieser Zugehörigkeiten konstant bleibt. In einem Optimum muss $\sum_{l \in M_k} z_{kl} = 1$ gelten.

Andernfalls könnte der Zielfunktionswert weiter verringert werden, indem einer der Parameter z_{kl} mit $l \in M_k$ um $\varepsilon > 0$ vergrößert wird; eines der anderen $z_{kj} > 0$ mit $j \notin M_k$ müsste entsprechend um ε verringert werden, damit die Lösung zulässig bleibt. Dann wäre Lösung (**) aber nicht optimal für Problem (4.6) und ein Widerspruch zu den Annahmen läge vor. Daher kann für ein beliebiges $l \in M_k$ $z_{kl} = 1$ gewählt werden; alle anderen z_{kj} werden 0 gesetzt.

Die so konstruierte Lösung (**) besitzt den gleichen Zielfunktionswert wie die Lösung (**) und ist in \underline{Z} ganzzahlig, also zulässig für Problem (4.5). Gleichzeitig liegt ein Widerspruch zu den Annahmen vor, da der Zielfunktionswert der optimalen Lösung (*) von Problem (4.5) nicht größer sein kann als der Zielfunktionswert der Lösung (**), wenn diese in in Problem (4.5) eingesetzt wird. Daher kann eine Lösung (**), die bei Problem (4.6) besser ist als Lösung (*) nicht existieren. Somit ist Lösung (*) optimal für Problem (4.6).

Umgekehrt existiert eine optimale Lösung für Problem (4.5), wenn eine optimale Lösung für das relaxierte Problem (4.6) existiert, da eine in \underline{Z} ganzzahlige optimale Lösung nach der oben angegebenen Prozedur aus der Lösung des relaxierten Problems (4.6) konstruiert werden kann.

Die Übereinstimmung der Zielfunktionswerte der optimalen Lösungen wurden bereits implizit bei der Konstruktion der ganzzahligen Lösung gezeigt. *q.e.d*

Nach Satz 4.1, reicht es also aus, das relaxierte Problem (4.6) zu lösen, um eine Lösung für das gemischt-ganzzahlige Problem (4.5) zu finden. Somit ist das gemischt-ganzzahlige Problem auf ein kontinuierliches restringiertes Optimierungsproblem zurückgeführt.

Um spezielle Eigenschaften dieses Optimierungsproblems zu untersuchen, wird nun das Optimierungsproblem für den Fall der 1-Norm näher untersucht, da sich mit dieser Norm ein quadratisches Optimierungsproblem ergibt. Dazu wird das ursprüngliche Optimierungsproblem umgeschrieben zu

$$\min \left\{ \sum_{i=1}^{N-1} \sum_{j=1}^M \|(\underline{y}(i+1) - \underline{A}_j \underline{y}(i) - \underline{B}_j \underline{u}(i) - \underline{c}_j)\|_1 z_{ij} \right\}$$

s.t.

$$z_{ij} \in [0, 1] \tag{4.7}$$

$$\sum_{j=1}^M z_{ij} = 1 \quad \forall i .$$

Hierin bezeichnet $\|\cdot\|_1$ die 1-Norm eines Vektors:

$$\|(x_1, \dots, x_n)^T\|_1 = |x_1| + \dots + |x_n|.$$

Dieses spezielle Problem kann nun als quadratisches Optimierungsproblem aufgeschrieben werden, indem man die Beträge mit Hilfe von Ungleichungen eliminiert. Da

$$\min |x|$$

äquivalent ist zu

$$\begin{aligned} \min s \\ \text{s.t.} \\ s \geq x \\ s \geq -x, \end{aligned}$$

kann Problem (4.7) wie folgt geschrieben werden:

$$\begin{aligned} \min \left\{ \sum_{i=1}^{N-1} \sum_{j=1}^M \sum_{k=1}^n s_{ij}^k z_{ij} \right\} \\ \text{s.t.} \\ (s_{ij}^1, \dots, s_{ij}^n)^T \geq \underline{y}(i+1) - \underline{A}_j \underline{y}(i) - \underline{B}_j \underline{u}(i) - \underline{c}_j \\ (s_{ij}^1, \dots, s_{ij}^n)^T \geq -\underline{y}(i+1) + \underline{A}_j \underline{y}(i) + \underline{B}_j \underline{u}(i) + \underline{c}_j \\ z_{ij} \in [0, 1] \\ \sum_{j=1}^M z_{ij} = 1 \quad \forall i. \end{aligned} \tag{4.8}$$

Durch Zusammenfassung der Variablen s_{ij}^k zu einem Vektor $\underline{s}_{ij} = (s_{ij}^1, \dots, s_{ij}^n)^T$,

erhält man die übersichtlichere Schreibweise

$$\min \left\{ \sum_{i=1}^{N-1} \sum_{j=1}^M z_{ij} \cdot \underbrace{(1, \dots, 1)}_n s_{ij} \right\}$$

s. t.

$$\begin{aligned} s_{ij} &\geq \underline{y}(i+1) - \underline{A}_j \underline{y}(i) - \underline{B}_j \underline{u}(i) - \underline{c}_j \\ s_{ij} &\geq -\underline{y}(i+1) + \underline{A}_j \underline{y}(i) + \underline{B}_j \underline{u}(i) + \underline{c}_j \\ z_{ij} &\in [0, 1] \\ \sum_{j=1}^M z_{ij} &= 1 \quad \forall i. \end{aligned}$$

In dieser Darstellung wird eine strukturelle Eigenschaft dieses Problems deutlich: Die Zielfunktion ist indefinit, da keine quadratischen Terme auftauchen, sondern ausschließlich gemischte Terme. Daher besitzt die Hesse-Matrix dieses Problems keine Einträge auf der Hauptdiagonalen. Somit weist das Problem kein eindeutiges Minimum auf. Dies wird auch an der Tatsache sichtbar, dass zu jedem zulässigen Punkt $\underline{x}_0 = (s_{11}^1, \dots, s_{NM}^n, z_{11}, \dots, z_{NM})^T$ von Problem (4.8) weitere $M! - 1$ äquivalente Punkte existieren, die man durch eine Permutation der Nummerierung der Teilmole erhält. Es existieren also mindestens $M!$ globale Minima.

Aufgrund der Indefinitheit der Hesse-Matrix können weitere lokale Minima existieren, in denen Abstiegsverfahren terminieren. Um wirklich das globale Optimum für dieses Problem zu finden, sind weitere Überlegungen notwendig.

Dazu werden die Optimierungsprobleme (4.5) und (4.6) noch einmal transformiert, so dass die Zuordnungsvariablen aus der Problemstellung verschwinden. Wie bereits herausgearbeitet wurde, werden die Zuordnungsvariablen z_{ij} im globalen Minimum immer so gewählt, dass die Daten dem Teilmodell zugeordnet werden bei dem der geringste Fehler resultiert. Dies lässt sich dann schreiben als

$$\min \left\{ \sum_{i=1}^{N-1} \min_j \| \underline{y}(i+1) - \underline{A}_j \underline{y}(i) - \underline{B}_j \underline{u}(i) - \underline{c}_j \| \right\}. \quad (4.9)$$

Satz 4.2

Das Optimierungsproblem (4.5) besitzt ein globales Minimum mit den Parametern \underline{A}_j^* , \underline{B}_j^* und \underline{c}_j^* genau dann, wenn das Optimierungsproblem (4.9) an der gleichen Stelle ein globales Minimum besitzt.

Beweis:

Der Beweis kann durch die gleichen Überlegungen erbracht werden, wie der Beweis zu Satz 4.1. *q.e.d.*

Die bisherigen Darstellungen des stückweise affinen Identifikationsproblems besitzen den gemeinsamen Nachteil, dass es keine konvexen Optimierungsprobleme sind. Falls die Daten nicht verrauscht sind und falls keine Unstetigkeiten im Kennfeld vorhanden sind, lassen sich die Probleme oft gut mit Abstiegsverfahren lösen. Allerdings kann es auch schon bei sehr einfachen Fällen dazu kommen, dass diese Abstiegsverfahren lediglich lokale Minima ermitteln, wie im folgenden Beispiel deutlich wird.

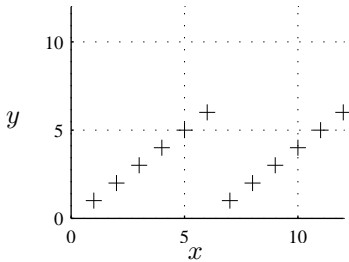


Abbildung 4.6: Messdaten für Beispiel 4.2

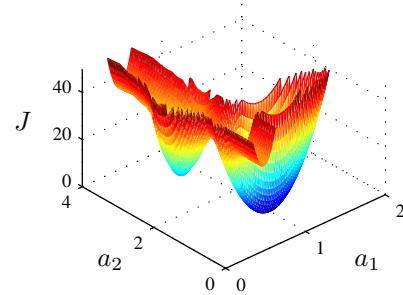


Abbildung 4.7: Gütemaß für Beispiel 4.2; Funktionswerte, die größer als 50 sind, sind nicht dargestellt

Beispiel 4.2

Für die idealen Messdaten aus Abbildung 4.6 soll ein stückweise affines Modell identifiziert werden. Die Anzahl der Teilmodelle beträgt $M = 2$. Damit eine graphische Darstellung des Problems möglich ist, werden die affinen Anteile für die Modelle zu $c_1 = 0$ und $c_2 = -6$ vorgegeben. Dann müssen lediglich die Parameter a_1 und a_2 der beiden Geraden identifiziert werden. Diese Identifikation kann nach Satz 4.2 im dynamischen Fall durch Minimierung von

$$\min \left\{ \sum_{i=1}^{N-1} \min_j \left\| \underline{y}(i+1) - \underline{A}_j \underline{y}(i) - \underline{B}_j \underline{u}(i) - \underline{c}_j \right\| \right\}$$

vorgenommen werden. Da das vorliegende Problem statisch ist und keine Eingänge besitzt, vereinfacht es sich zu

$$\min_{a_1, a_2} \left\{ \sum_{i=1}^{12} \min \left\{ \|y(i) - a_1 x(i) - 0\|, \|y(i) - a_2 x(i) + 6\| \right\} \right\}.$$

Das Gütemaß für dieses Problem ist in Abbildung 4.7 dargestellt, wobei die quadratische Norm benutzt wurde. Offensichtlich gibt es hier zwei lokale Minima; das rechte Minimum stellt gleichzeitig das globale Minimum dar.

Konventionelle Abstiegsverfahren sind daher zur Identifikation stückweise affiner

Modelle nur bedingt geeignet. Es ist notwendig, geeignete Verfahren anzuwenden, wie sie in der Literatur [Han92, Rat92, HN99] bereits bekannt sind.

4.4.1 Methoden zur globalen Optimierung

Für die vorgestellten Formulierungen des Optimierungsproblems wurde gezeigt, dass lokale Minima existieren, die nicht immer eine geeignete Lösung des Optimierungsproblems darstellen. Daher müssen globale Verfahren zum Einsatz kommen, mit denen diese Probleme gelöst werden können.

Um beispielsweise das Identifikationsproblem in der Darstellung (4.9) zu lösen, ist es notwendig, numerische Verfahren anzuwenden, die der Tatsache Rechnung tragen, dass bei diesem Problem das Minimum nicht eindeutig ist und zusätzlich nicht überall ein Gradient der Funktion angegeben werden kann.

Das in [HN99] beschriebene Verfahren wurde in dieser Arbeit angewandt, um Probleme der Form (4.9) zu lösen. Dazu teilt das Verfahren den Parameterraum in einzelne mehrdimensionale Intervalle auf und berechnet obere und untere Schranken für diese Intervalle. Die weitere Aufteilung der einzelnen Intervalle erfolgt anschließend anhand der bisher berechneten oberen und unteren Schranken.

Da die Problemformulierung zudem mit einer relativ kleinen Anzahl von Variablen auskommt, können mit dieser Vorgehensweise auch Probleme in durchaus praxisrelevanter Größenordnung gelöst werden. Maßgebend für die Anzahl der Parameter ist dabei die Dimension $n + p$ des Identifikationsraumes sowie die Anzahl der Teilmodelle M . Für jedes Teilmodell entstehen $n^2 + np + n$ Parameter, so dass ein Suchraum der Dimension $M \cdot (n^2 + np + n)$ entsteht. Dieses Verfahren garantiert, dass die global optimale Lösung gefunden wird, falls die Anzahl der Unterteilungen, die der Benutzer vorgibt, groß genug gewählt wird.

Andere Verfahren, wie der in [Han92] vorgestellte Algorithmus, zerlegen den Parameterraum selbständig in immer kleinere mehrdimensionale Intervalle und nehmen mit Hilfe der Intervallrechnung in den betreffenden Intervallen eine Abschätzung der Funktionswerte vor. Dadurch wird es mit Hilfe von Branch-and-Bound Algorithmen möglich, verschiedene Intervalle auszuschließen, so dass die betreffenden Intervalle nicht weiter betrachtet werden müssen. Dieser Grundalgorithmus wurde in [Rat92] weiter verbessert.

Um Probleme direkt in der ursprünglichen gemischt-ganzzahligen Form (4.5) zu lösen, können Branch-and-Bound Algorithmen wie MINLP zum Einsatz kommen [Ley98, Ley01]. Diese besitzen zwar den Nachteil, dass viele mögliche Lösungen ausprobiert werden müssen, jedoch ist auch hier garantiert, dass ein globales Optimum für das in dieser Arbeit betrachtete Identifikationsproblem gefunden wird.

In dieser Arbeit wurden die Algorithmen, die in [HN99] und [Ley98, Ley01] vorgestellt sind, anhand der Anwendungsbeispiele erprobt. Diese scheinen am besten geeignet, um die vorliegenden Probleme zu lösen, da sie bereits in einer Form implementiert sind, die eine einfache Verwendung erlaubt. Der Algorithmus von [HN99] steht als MATLAB-Quelltext zur Verfügung; auf den Löser MINLP [Ley98, Ley01] kann über den NEOS-Server [neo05] und das dort zur Verfügung gestellte AMPL-Interface [FGK03] zugegriffen werden. Diese Lösungen sind zudem kostenlos und bereits ausreichend getestet.

Andere Löser, die versprechen das globale Optimierungsproblem zu lösen, wurden ausprobiert, mussten aber wegen folgender Gründe ausgeschlossen werden: Die Löser ACRS (Adaptive Controlled Random Search) und MLOCPSOA sind über den NEOS-Server und das AMPL-Interface verfügbar; allerdings konnte bei diesen Lösern schon mit einfachen Anwendungsbeispielen gezeigt werden, dass sie das globale Optimum nicht finden. Die Algorithmen von [Rat92] liegen nur in XSC-Pascal vor, was eine Anwendung dieser Algorithmen schwierig macht; Quelltexte oder andere verfügbare Software zu [Han92] konnte nicht gefunden werden, so dass dieser Algorithmus nur mit erheblichem Aufwand hätte eingesetzt werden können.

Nachdem die Lösung des globalen Optimierungsproblems mit Hilfe eines geeigneten Algorithmus gefunden wurde, müssen die Daten den Teilmodellen zugeordnet werden. Bei der Lösung des gemischt-ganzzahligen Problems, liegt bereits eine Zuordnung der Daten zum nächsten Teilmodell vor. Die so erhaltene Zuordnung stellt zwar eine mögliche Lösung des Identifikationsproblems dar, da jedoch keine topologischen Informationen in die Optimierung einfließen, ist nicht sichergestellt, dass zusammenhängende Gebiete entstehen. Daher muss die hier erhaltene Zuordnung der Daten zu den Teilmodellen im nächsten Schritt noch so verändert werden, dass zusammenhängende Gültigkeitsbereiche der Teilmodelle zugesichert werden können. Dazu wird im nächsten Abschnitt eine Vorgehensweise vorgeschlagen.

4.4.2 Aufteilung des Identifikationsraumes mit Hilfe von Graphen

Bei der naiven Zuordnung der Messdaten zum nächsten Teilmodell kann es bei vertauschten Daten vorkommen, dass diese Zuordnung zu sehr zersplitterten Gültigkeitsbereichen für die einzelnen Teilmodelle führt. Um zusammenhängende Gebiete zu erhalten, können nun noch Nachbarschaftsinformationen, die in den Messdaten vorhanden sind, ausgenutzt werden.

Dazu ist die Erstellung eines Nachbarschaftsgraphen erforderlich. Dieser Nachbarschaftsgraph ist ein ungerichteter Graph, wie er in Anhang A beschrieben ist. Er gibt an, welche Messdaten im Identifikationsraum benachbart sind. Für zwei benachbar-

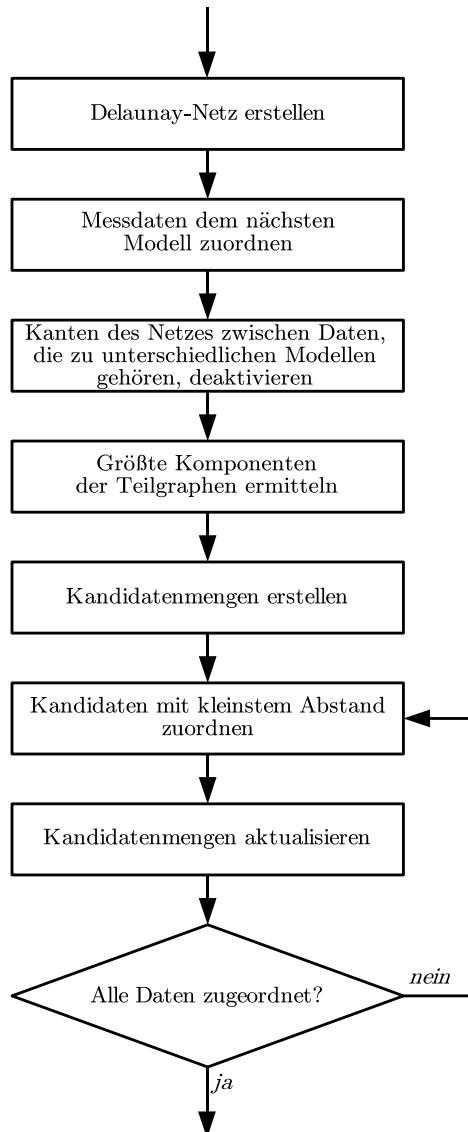


Abbildung 4.8: Flussdiagramm zur Aufteilung des Identifikationsraums mit Hilfe des Nachbarschaftsgraphen

te Datenpunkte wird dies durch eine ungerichtete Kante im Nachbarschaftsgraphen angezeigt. Um solch einen Graphen zu erstellen, gibt es verschiedene Möglichkeiten. Eine davon ist, eine Delaunay-Triangulierung durchzuführen und die Kanten der dabei entstehenden Simplexe als Kanten des Nachbarschaftsgraphen zu verwenden.

Die Delaunay-Triangulierung ist in Anhang B genauer erläutert. Diese Triangulierung besitzt die Eigenschaft, dass kein Datenpunkt innerhalb der Umhyperkugel eines Simplex liegt. Daher ist der so gebildete Graph auch ein sinnvoller Nachbarschaftsgraph.

Dieser Nachbarschaftsgraph muss nun so unterteilt werden, dass zusammenhängende Gebiete entstehen; das bedeutet, dass der Graph in zusammenhängende Teilgraphen unterteilt werden soll. Die Ecken dieser zusammenhängenden Teilgraphen müssen dann komplett zu einem Teilmodell gehören. Außerdem soll das Gütemaß, das bei der Unterteilung des Graphen entsteht, möglichst global minimal sein. Da bei der optimalen Lösung dieser Aufgabe der Aufwand exponentiell mit der Zahl der Ecken im Graphen ansteigt [Sch01b], wird hier nur eine Heuristik zur Lösung dieser Aufgabe abgegeben.

Diese Heuristik wurde in [Phu04] entwickelt und besteht aus mehreren Schritten, die nun nacheinander dargestellt werden sollen:

1. Zunächst muss das Delaunay-Netz für die gegebenen Daten erstellt werden.
2. Anschließend werden alle Messdaten dem nächsten Modell zugeordnet. Dadurch wird der Graph – entsprechend den M Teilmodellen – in M Teile zerlegt, die hier mit Teilgraphen bezeichnet werden. Der Knoten i des Graphen, der das Messdatum i repräsentiert, gehört also zum Teilgraphen j , wenn die Norm des Fehlervektors zwischen Teilmodell j und dem Punkt i minimal ist. Falls die Norm des Fehlervektors für mehrere Modelle minimal wird, muss eines der Modelle ausgesucht werden.
3. Alle Kanten zwischen Messdaten, die zu unterschiedlichen Teilgraphen gehören werden deaktiviert. Dadurch erhält man M Teilgraphen, die nicht miteinander verbunden sind.
4. Die Komponenten der M Teilgraphen werden nun betrachtet. Jeder Teilgraph besitzt eine größte Komponente. Ausgehend von diesen M größten Komponenten werden nun alle Messdaten, die nicht zu einer der größten Komponenten gehören, zu den größten Komponenten zugeordnet.
5. Dazu werden Kandidatenmengen erstellt, indem zunächst alle Kanten, die nicht zu einer größten Komponenten gehören, deaktiviert werden. Alle Knoten, die entsprechend den deaktivierten Kanten zu einer der größten Komponenten benachbart sind, gehören zu den Mengen der Messpunkte, die im nächsten Schritt zugeordnet werden können. So entstehen M Kandidatenmengen.
6. Um aus diesen Kandidatenmengen das Messdatum auszuwählen, das als erstes zugeordnet wird, werden in jeder Kandidatenmenge j die Normen der Fehlervektoren zwischen den Kandidaten und dem in der benachbarten größten Kom-

ponenten j geltenden Teilmodell berechnet. Der Messpunkt, der nach diesem Kriterium den geringsten Wert aufweist, wird dem entsprechenden Teilmodell zugeordnet.

7. Danach müssen die Kandidatenmengen aktualisiert werden.

8. Die Heuristik wird in Schritt 6 fortgesetzt bis alle Messdaten zugeordnet sind.

Dadurch erhält man M zusammenhängende Teilgraphen des Nachbarschaftsgraphen und somit auch zusammenhängende Teilmengen des Identifikationsraumes, auf denen jeweils ein affines Modell gültig ist.

Nachdem so der Zusammenhang der Gültigkeitsbereiche der Teilmodelle sichergestellt wurde, können die einzelnen Teilmodelle noch einmal unabhängig voneinander geschätzt werden.

Dieses Verfahren erfüllt – ebenso wie das vorherige Verfahren – weitgehend die Forderungen, die zu Beginn dieses Kapitels erhoben wurden:

- Die Parameter, die vom Benutzer vorzugeben sind beschränken sich auf die Anzahl der Teilmodelle, zu deren Bestimmung auch hier auf systematische Verfahren zurück gegriffen werden kann [NTK05].
- Aufgrund der Anschaulichkeit des Verfahrens ist auch hier Vorwissen leicht integrierbar.
- Verrauschte Messdaten bereiten keine Schwierigkeiten, da diesem Verfahren ein Optimierungsansatz zugrunde liegt, der eine Verallgemeinerung zur Methode der kleinsten Quadrate darstellt.
- Der Zusammenhang der Teilgebiete kann durch den zweiten heuristischen Schritt gesichert werden.
- Der Hauptnachteil des vorherigen Verfahrens – die große Rechenzeit – ist hier abgemildert, da die eingesetzten Optimierungsalgorithmen sehr leistungsfähig sind und auch bei durchaus praxisrelevanten Problemgrößen eingesetzt werden können.

Anhand einiger Beispiele wird im nächsten Kapitel die Anwendung der Verfahren demonstriert.

Kapitel 5

Anwendung der hybriden Identifikationsverfahren

Die vorgestellten Identifikationsverfahren sollen nun anhand mehrerer Beispiele validiert und verglichen werden. Zunächst werden in Abschnitt 5.1 zwei Beispiele zur Identifikation hybrider Systeme mittels der in Abschnitt 4.3 vorgestellten Methode der lokalen Kennfeldoptimierung vorgestellt.

In Abschnitt 5.2 werden anschließend drei weitere Beispiele vorgestellt, bei denen die Identifikation mit Hilfe der nichtlinearen Optimierungsmethoden aus Abschnitt 4.4 angewandt wird.

Am Ende dieses Kapitels folgt ein Vergleich beider Identifikationsverfahren.

5.1 Beispiele zur Identifikation durch lokale Kennfeldoptimierung

5.1.1 Identifikation eines unstetigen statischen Beispielkennfeldes

In diesem Anwendungsbeispiel wird das Kennfeld aus Abbildung 5.1 betrachtet und eine Identifikation für dieses Kennfeld vorgenommen.

Das betrachtete Kennfeld besitzt zwei Teilmodelle, die in unterschiedlichen Bereichen gelten. Der Bereich 1 (in Abbildung 5.1 dunkel dargestellt) wird durch die Ungleichungen

$$\begin{aligned}x_2 &\geq && 0,2 \\x_1 &\leq && 0,7 \\x_2 &\leq & 0,2 x_1 + & 0,55\end{aligned}$$

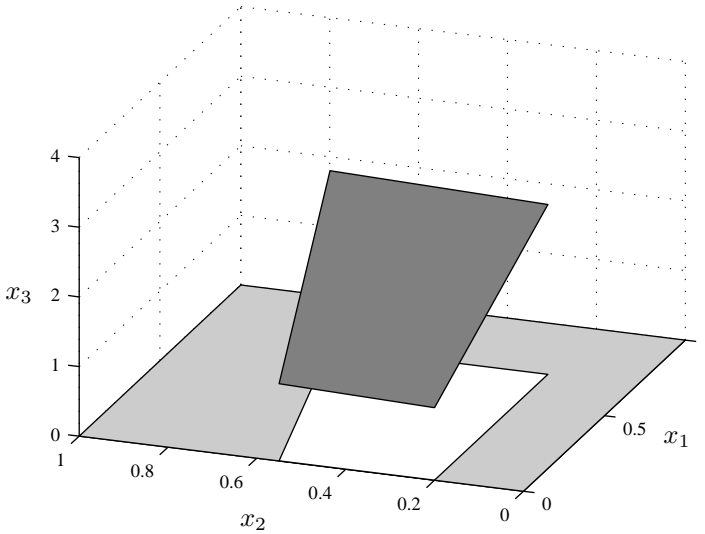


Abbildung 5.1: unstetiges, stückweise affines Beispielkennfeld

beschrieben. Bereich 2 (in Abbildung 5.1 hellgrau dargestellt) ist das Komplement von Bereich 1 auf dem zweidimensionalen Einheitsintervall.

Auf dem Bereich 1 gilt das affine Modell

$$x_3 = 2x_1 + 0,2x_2 + 1.$$

Das Modell auf Bereich 2 wird durch

$$x_3 = 0$$

beschrieben.

Indem man ein regelmäßiges Gitter über den Identifikationsraum legt und an den Gitterpunkten das Kennfeld auswertet, erhält man die Daten für den Identifikationsalgorithmus. Als Abstand der Gitterlinien ist hier 0,03 gewählt, so dass sich Daten ergeben, wie sie in Abbildung 5.2 zu sehen sind.

Die Startkonfiguration der Simplexe wurde vom Benutzer vorgegeben. Dazu wurde zunächst ein Raster aus Stützstellen festgelegt und anschließend mit Hilfe der Delaunay-Triangulierung in Simplexe unterteilt. Danach wurde in jedem Simplex ein affines Modell geschätzt, so dass die Konfiguration entstand, wie sie in Abbildung 5.3 zu sehen ist.

Diese initiale Aufteilung des Grundgebietes führt zu einem großen quadratischen

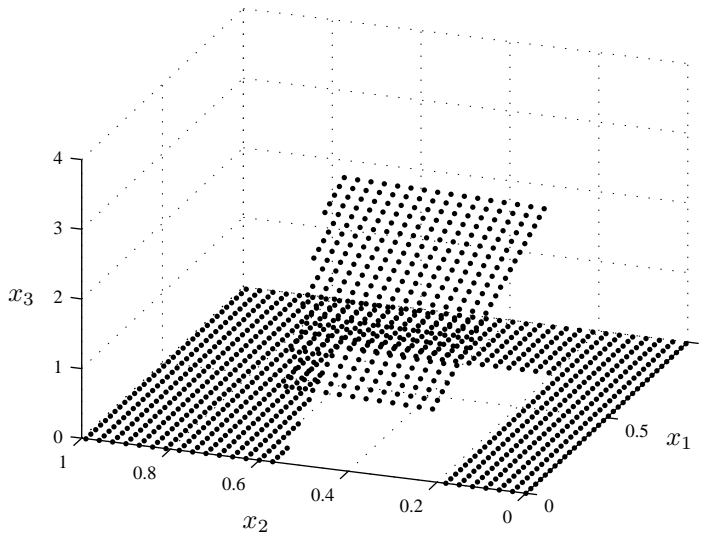


Abbildung 5.2: Datenbasis für die Identifikation des stückweise affinen Beispielkennfeldes

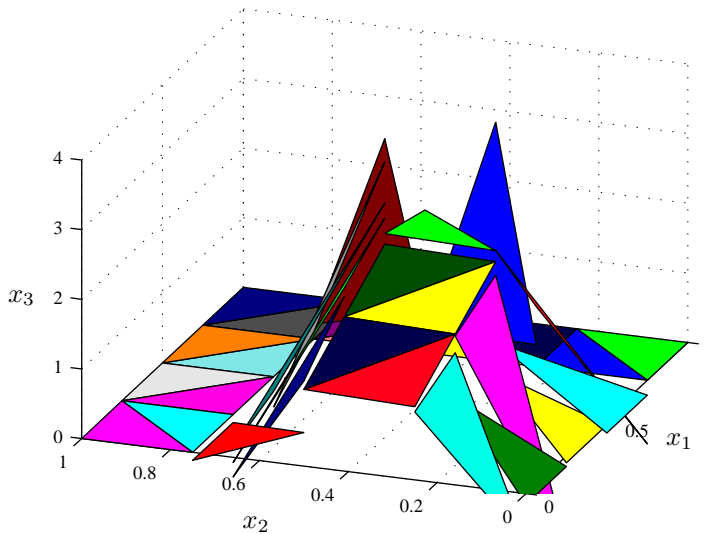


Abbildung 5.3: Ausgangskonfiguration für die Identifikation des Beispielkennfeldes mittels lokaler Kennfeldoptimierung

Fehler, da viele Teilgebiete Mischgebiete sind, die Datenpunkte aus beiden affinen Bereichen des ursprünglichen Modells enthalten.

Nach 89 Durchläufen der Optimierungsschleife wurde der quadratische Gesamtfehler

von einem Startwert von 75,4 auf 0,45 reduziert. Man sieht hier, dass ein Restfehler bleibt, obwohl es sich um ideale Daten handelt. Dieser beruht auf der Tatsache, dass es sich bei dem Verfahren lediglich um eine Folge von lokalen Optimierungsschritten handelt, die nicht notwendigerweise zu einem globalen Optimum führen.

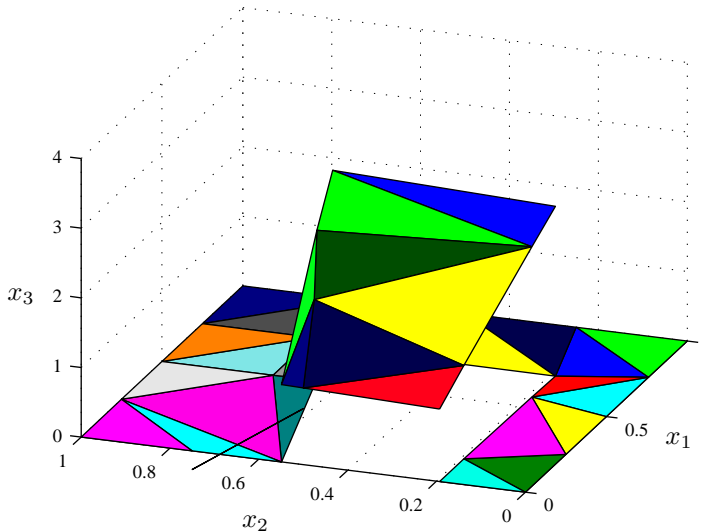


Abbildung 5.4: Ergebnis der lokalen Kennfeldoptimierung für das unstetige Beispielkennfeld

Das Ergebnis ist in Abbildung 5.4 zu sehen. Bei genauer Betrachtung des Ergebnisses wird offensichtlich, in welchem Simplex der Restfehler dieser Aufteilung zu finden ist:

Bei $x_2 = 0,6$ befindet sich ein kleiner und sehr schmaler Simplex, der in der graphischen Darstellung nur als Strich erkennbar ist. In diesem Simplex befinden sich Daten aus beiden Teilmodellen, die den Restfehler verursachen. Trotz dieses Restfehlers ist das Identifikationsergebnis positiv zu bewerten, da die Wahrscheinlichkeit für Anfragepunkte aus dem fehlerhaften Simplex verschwindend gering ist.

Vergleicht man die Anfangskonfiguration in Abbildung 5.3 mit der Endkonfiguration in Abbildung 5.4, so kann man sehen, wie der Algorithmus die einzelnen Stützstellen verschoben hat, da die Färbung der Simplexe während des Algorithmus beibehalten wurde.

5.1.2 Kompressorkennfeld eines Turboladers

In diesem Abschnitt wird das Kennfeld eines Turboladers betrachtet, das den Luftmassenstrom des Kompressors in Abhängigkeit von der Drehzahl des Turboladers und vom Druck im Ladeluftkühler angibt.

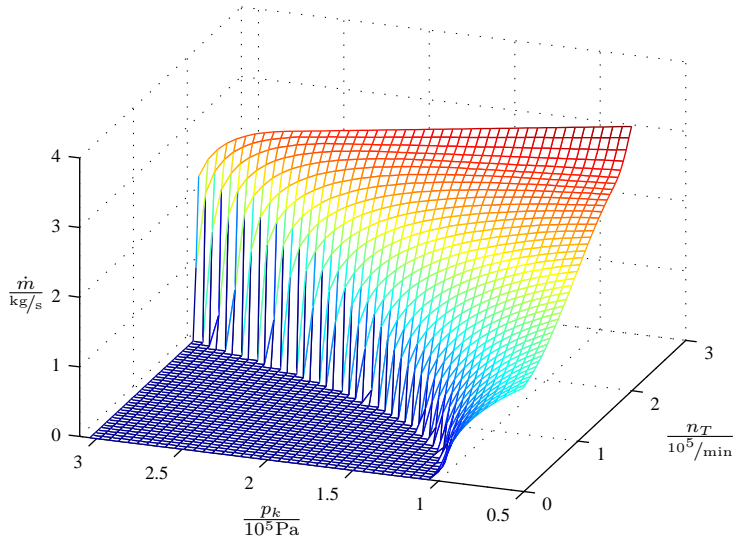


Abbildung 5.5: Kompressorkennfeld des Turboladers

Der Luftmassenstrom, der durch den Kompressor erzeugt wird, weist ein nahezu hybrides Verhalten auf. Ein Kennfeld, wie es für den Common-Rail Vierzylinder Turbodieselmotor OM 611 der DaimlerChrysler AG vermessen wurde, ist in Abbildung 5.5 dargestellt. Das hier gezeigte Kennfeld gilt bei einer Außentemperatur von 25°C und einem Atmosphärendruck von 1013 hPa. Es ist deutlich zu erkennen, dass es einen Bereich für große Drücke und kleine Drehzahlen gibt, in dem der Kompressor keinen Luftmassenstrom fördert, also quasi ausgeschaltet ist. Ein tieferes Verständnis des Turboladers ist an dieser Stelle nicht erforderlich, jedoch können weitere Zusammenhänge in [HP03] nachgelesen werden.

Das Kennfeld soll nun mit Hilfe eines stückweise affinen Systems dargestellt werden. Die Daten, die zur Identifikation des Kennfeldes verwendet wurden, befinden sich an den Schnittpunkten der Gitterlinien, wie sie in Abbildung 5.5 zu sehen sind. Für die Identifikation und die folgenden graphischen Darstellungen wurden die Daten auf das dreidimensionale Einheitsintervall normiert.

Anschließend wurde die in Abschnitt 4.3 beschriebene Methode der lokalen Kennfeldoptimierung angewandt. Als anfängliche Unterteilung des Identifikationsraumes

wurde ein regelmäßiges Gitter verwendet, mit dessen Stützstellen der Identifikationsraum anschließend mit Hilfe der Delaunay-Triangulierung in Simplexe unterteilt wurde. Diese anfängliche Aufteilung des Identifikationsraumes ist in Abbildung 5.6 zu sehen.

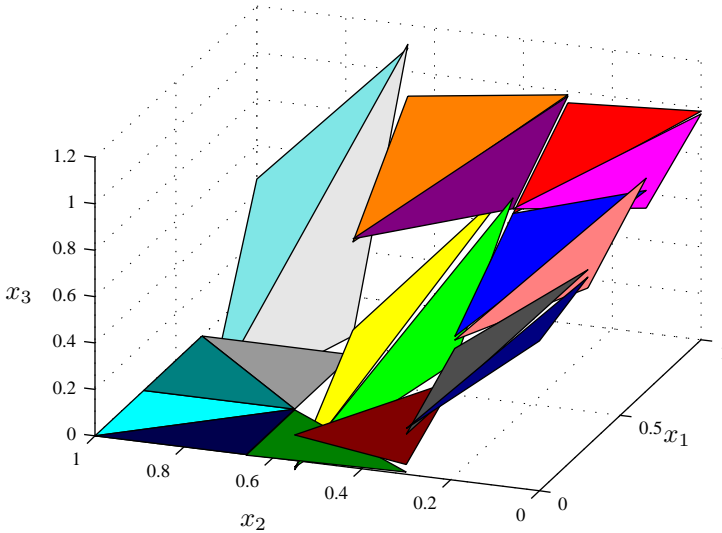


Abbildung 5.6: Ausgangskonfiguration für die Identifikation des Turboladerkennfeldes mittels lokaler Kennfeldoptimierung

Die initiale Aufteilung des Identifikationsraumes führt bei diesem Beispiel zu einem quadratischen Fehler von 5,26. Der Algorithmus hat hier 41 Durchgänge benötigt, bis die Abbruchbedingung erfüllt war. Dabei wurde der quadratische Gesamtfehler auf 0,41 reduziert. Auch hier bleibt ein Restfehler bestehen, was aber in diesem Beispiel auf die Tatsache zurückzuführen ist, dass das ursprüngliche Kennfeld stark nichtlinearer Natur ist und mit Hilfe affiner Teilmodelle nicht fehlerfrei dargestellt werden kann.

In Abbildung 5.7 ist das Ergebnis dieser Optimierung dargestellt. Um die Güte dieser Darstellung besser beurteilen zu können, ist in Abbildung 5.8 das normierte Kennfeld zum Vergleich dargestellt.

Auch bei diesem Anwendungsbeispiel kann man durch Vergleich der Anfangskonfiguration in Abbildung 5.6 mit der Endkonfiguration in Abbildung 5.7 leicht feststellen, in welcher Weise die einzelnen Stützstellen verschoben wurden, da auch in diesem Beispiel die Färbung der einzelnen Simplexe während des Algorithmus beibehalten wurde.

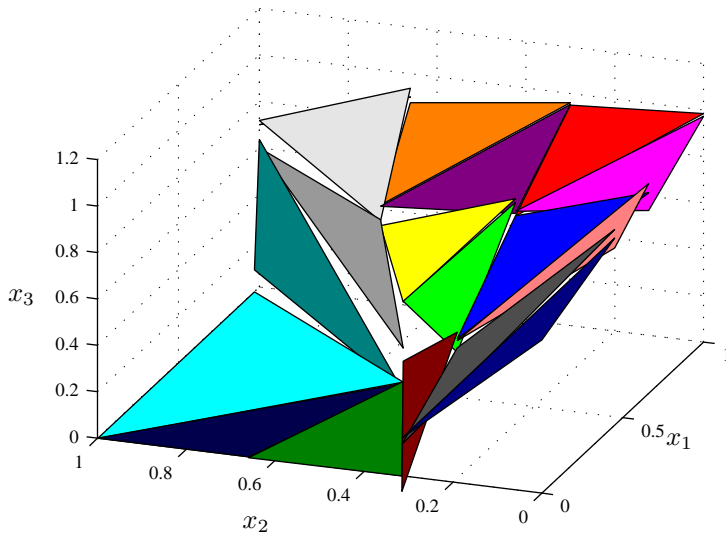


Abbildung 5.7: Ergebnis der lokalen Kennfeldoptimierung für die Identifikation des Turboladerkennfeldes

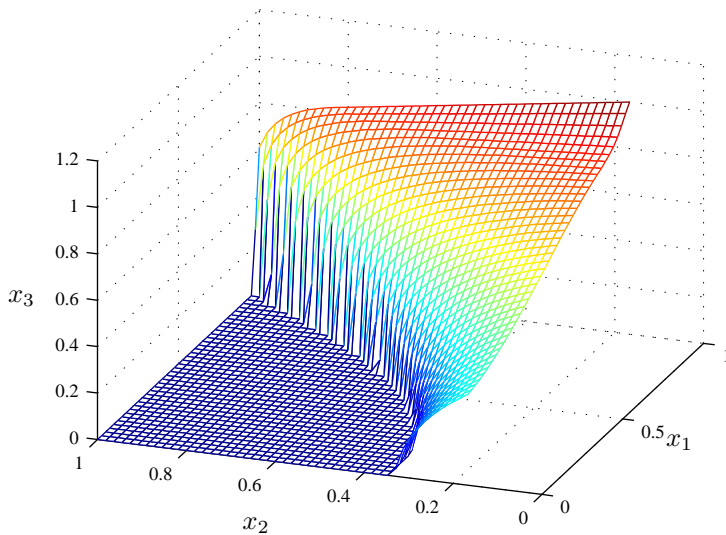


Abbildung 5.8: normiertes Turboladerkennfeld zum Vergleich mit dem Identifikationsergebnis aus Abbildung 5.7

5.2 Beispiele zur Identifikation durch nichtlineare Optimierung

In diesem Abschnitt werden drei Beispiele vorgestellt, die mit Hilfe der Identifikationsmethoden aus Abschnitt 4.4 gelöst werden. Um zunächst die Arbeitsweise der Methode zu veranschaulichen, wird als erstes Anwendungsbeispiel ein stückweise affines, statisches Kennfeld mit zwei Eingangsgrößen und einer Ausgangsgröße betrachtet. Als weiteres Beispiel wird ein springender Ball betrachtet, bei dem ein dynamisches Modell identifiziert und anschließend das identifizierte Modell auch simuliert wird. Zum Schluss wird in diesem Abschnitt ein hybrides Zwei-Tank-System identifiziert und simuliert.

5.2.1 Identifikation eines statischen Beispielkennfeldes

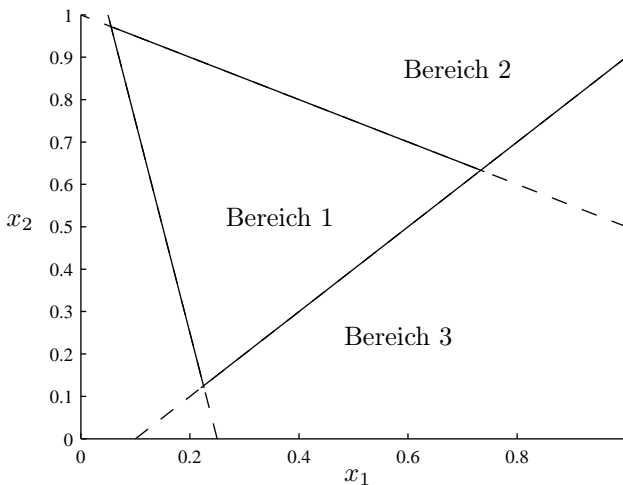


Abbildung 5.9: Aufteilung des Identifikationsraumes bei dem stückweise affinen statischen Beispielkennfeld

In diesem Abschnitt wird anhand eines stückweise affinen und unstetigen Kennfeldes die Arbeitsweise der Identifikation durch globale nichtlineare Optimierung und die anschließende topologische Verbesserung des Ergebnisses verdeutlicht. Das betrachtete Kennfeld besitzt drei Teilmodelle, die in unterschiedlichen Bereichen gelten. Der Bereich 1 wird durch die Ungleichungen

$$\begin{array}{rcl}
 x_1 & \geq & -0,2 \quad x_2 + 0,25 \\
 x_1 & \leq & \quad \quad x_2 + 0,1 \\
 x_1 & \leq & -2 \quad x_2 + 2
 \end{array}$$

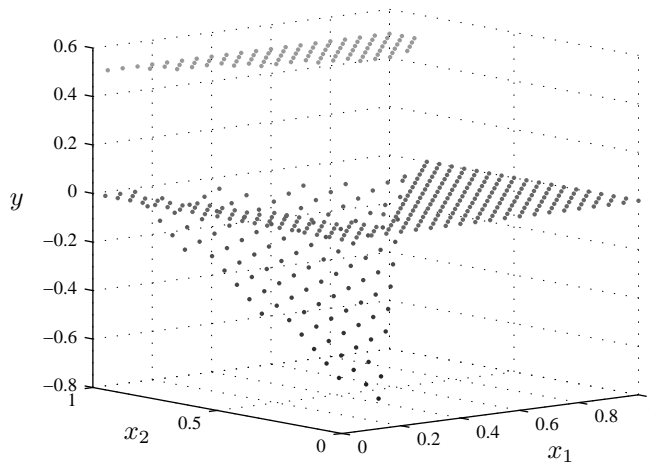


Abbildung 5.10: Darstellung des zu identifizierenden Kennfeldes durch ideale Daten

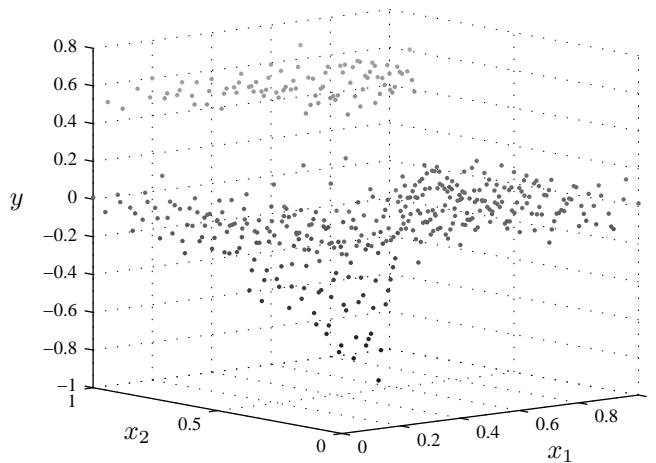


Abbildung 5.11: Datenbasis für die Identifikation des statischen Beispielkennfeldes

beschrieben. Bereich 2 lässt sich durch die Ungleichungen

$$\begin{aligned} x_1 &\geq -0,2 & x_2 &+ 0,25 \\ x_1 &\leq & x_2 &+ 0,1 \\ x_1 &> -2 & x_2 &+ 2 \end{aligned}$$

beschreiben, und Bereich 3 ist das Komplement der Bereiche 1 und 2. Damit ergibt sich die Aufteilung des Grundbereiches, wie sie in Abbildung 5.9 zu sehen ist.

Auf dem Bereich 1 gilt das affine Modell

$$y = 1/2 x_1 + x_2 - 1 = 0.$$

Das Modell auf Bereich 2 wird durch

$$y = 1/2$$

beschrieben und auf Bereich 3 gilt

$$y = 0.$$

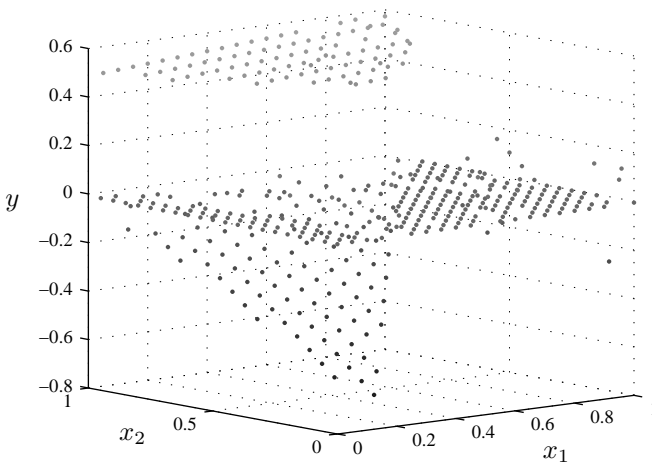


Abbildung 5.12: Lösung des Optimierungsproblems nach der Identifikation mit MCS

Indem das Modell an den Schnittpunkten eines regelmäßigen Gitters mit einem Abstand von 0,05 ausgewertet wird, wird der Identifikationsraum mit Daten gefüllt. Die so erhaltenen idealen Daten sind in Abbildung 5.10 dargestellt. Für die Identifikation werden allerdings nicht diese idealen Daten verwendet. Um die Leistungsfähigkeit des Identifikationsalgorithmus zu zeigen, werden die Daten mit einem normalverteilten Rauschen mit Erwartungswert $m = 0$ und Varianz $\sigma^2 = 0,004$ beaufschlagt. Die resultierenden Daten sind in Abbildung 5.11 zu sehen.

Das Verhältnis von Signal zu Rauschen kann, wie es in [KK98] angegeben ist, berechnet werden. Die Varianz der nicht verrauschten Daten kann bei diesem Beispiel als die Varianz der Funktionswerte berechnet werden. Sie ergibt sich zu $\sigma_{\text{Signal}}^2 = 0,079$. Die Varianz des Rauschens ist bekannt, da das Rauschen mit Hilfe einer Normalverteilung generiert wurde. Das Signal-Rausch-Verhältnis (englisch: Signal-Noise-Ratio

(SNR)) ergibt sich damit zu

$$SNR = 10 \cdot \log_{10} \frac{\sigma_{\text{Signal}}^2}{\sigma_{\text{Rauschen}}^2} = 10 \cdot \log_{10} \frac{0,079}{0,004} = 12,954.$$

Zur Identifikation des Beispiels wird nun das Problem mit Hilfe des MCS-Algorithmus in der Form von Gleichung (4.9) gelöst [HN99]. Dabei wird die quadratische Norm verwendet, so dass das Optimierungsproblem in diesem speziellen Fall die Form

$$\min \left\{ \sum_{i=1}^N \min_j (y(i) - \underline{a}_j^T \underline{x}(i) - c_j)^2 \right\}$$

annimmt, da das Kennfeld statisch ist und das System zwei Eingänge (x_1 und x_2) und einen skalaren Ausgang (y) besitzt.

Nach der Ausführung von MCS mit der angegebenen Gütefunktion erhält man die Lösung, wie sie in Abbildung 5.12 zu sehen ist. Man erkennt, dass es einige Ausreißer gibt, die aufgrund der verrauschten Daten den falschen Modellen zugeordnet worden sind. Betrachtet man die Zugehörigkeit der Datenpunkte zu den verschiedenen Modellen genauer, so ist Abbildung 5.13 hilfreich, in der die Zugehörigkeiten der Daten zu den verschiedenen Modellen mit Hilfe unterschiedlicher Symbole dargestellt sind.

Dieses Optimum, das mit dem MCS-Algorithmus gefunden wurde, minimiert das aufgestellte Gütemaß; da jedoch in dem Gütemaß nicht berücksichtigt wird, dass zusammenhängende Gebiete für die einzelnen Teilmodelle entstehen sollen, kann dies vom Ergebnis auch nicht erwartet werden.

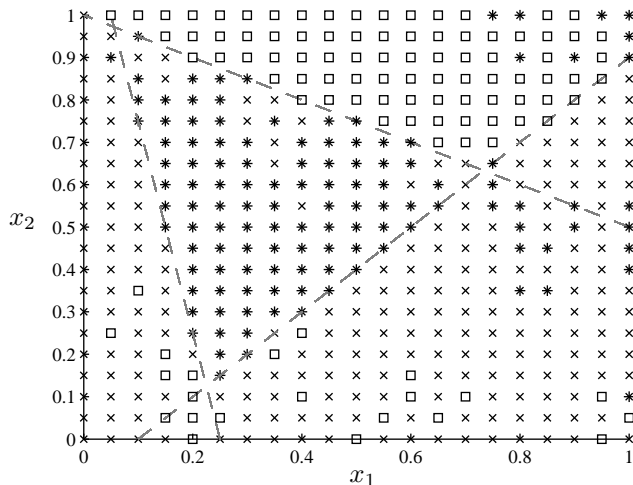


Abbildung 5.13: Zugehörigkeit der Datenpunkte zu den Teilmodellen nach der Optimierung mit MCS

Daher wird nun in einem weiteren Schritt der Zusammenhang der Gültigkeitsbereiche der einzelnen Teilmodelle betrachtet und verbessert. Dazu wird, wie in Abschnitt 4.4.2 bereits beschrieben, ein heuristischer Algorithmus eingesetzt, der den Zusammenhang der Gültigkeitsbereiche der einzelnen Teilmodelle sicherstellt. Zunächst benötigt der Algorithmus Nachbarschaftsinformationen über den vorliegenden Datensatz, der dem Algorithmus in Form eines Delaunay-Graphen zur Verfügung gestellt wird. In diesem Graphen werden nun Subgraphen gebildet, die sich aufgrund der Zugehörigkeit der einzelnen Datenpunkte zu den Teilmodellen ergeben. Anschließend werden in den Subgraphen die größten Komponenten gesucht; das Ergebnis dieses Schrittes ist in Abbildung 5.14 zu sehen.

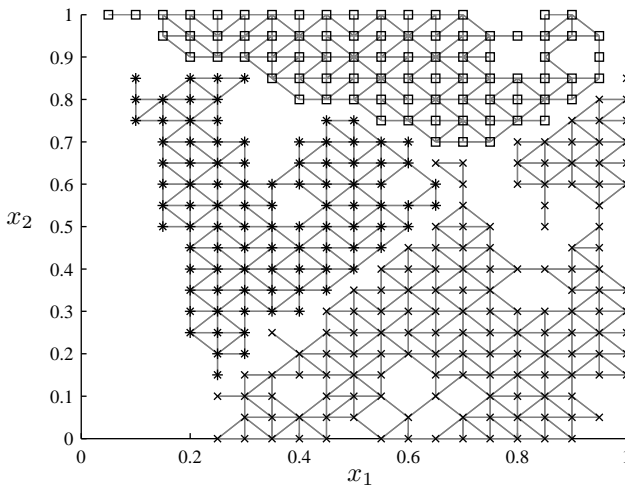


Abbildung 5.14: Die größten Komponenten der Subgraphen

Nachdem die größten Komponenten ermittelt sind, werden zu diesen nach und nach die Datenpunkte hinzugefügt, die zu den Komponenten benachbart sind und zu dem jeweiligen Teilmodell die kleinsten Abstände besitzen; die genaue Vorgehensweise wurde bereits in Abschnitt 4.4.2 beschrieben.

Die Zugehörigkeit der einzelnen Datenpunkte zu den Teilmodellen nach dieser topologischen Optimierung ist in Abbildung 5.15 zu sehen.

Nach dieser Aufteilung können die Modellparameter für die feste Zuordnung noch einmal geschätzt werden. Diese Schätzung mit fester Zuordnung führt zu dem Ergebnis, das in Abbildung 5.16 zu sehen ist.

Dieses Ergebnis kann nun anhand zweier Kriterien beurteilt werden: Zum einen danach, ob die Datenpunkte dem richtigen Modell zugeordnet wurden, zum anderen anhand der Modellparameter, die für die affinen Teilmodelle geschätzt wurden.

Vergleicht man die Aufteilung in die drei Bereiche mit dem Resultat der topologi-

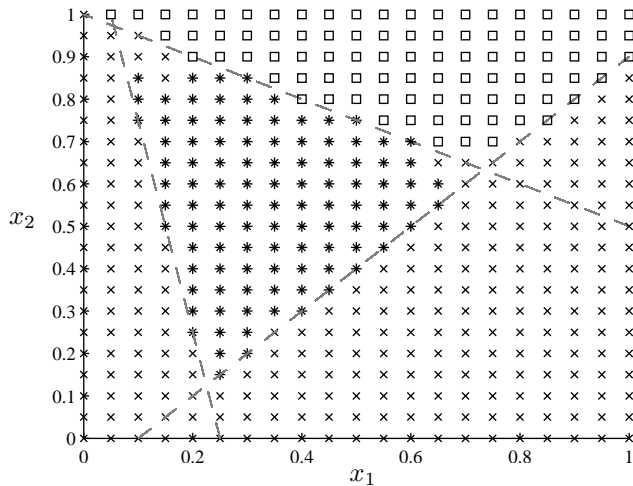


Abbildung 5.15: Zuordnung der Datenpunkte zu den Teilmodellen nach der topologischen Optimierung

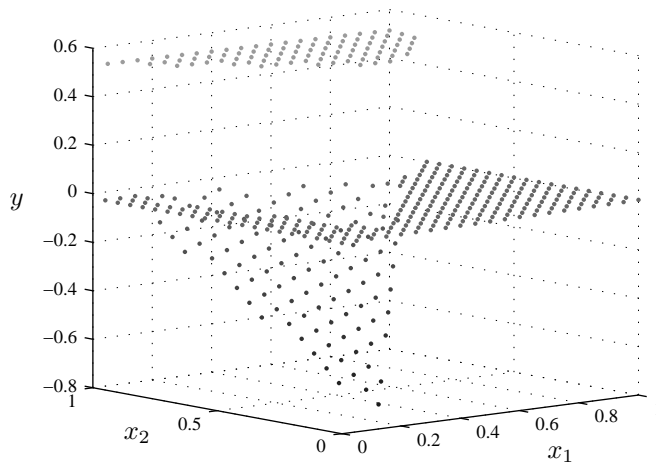


Abbildung 5.16: Darstellung des identifizierten Modells nach topologischer Optimierung und Schätzung der Modellparameter mit fester Zuordnung

schen Optimierung, wie dies in Abbildung 5.15 gezeigt ist, so stellt man fest, dass nur noch wenige Punkte dem falschen Modell zugeordnet sind. Diese falschen Zuordnungen fallen nur wenig ins Gewicht, da die Modelle in dem betreffenden Bereich sehr ähnliche Funktionswerte aufweisen. Dies führt dazu, dass in diesem Bereich die Trennung der beiden Modelle sogar für ideale Daten sehr schwer ist, da sich die Ebenen, die durch die affinen Modelle der Bereiche 1 und 3 aufgespannt werden, hier schneiden.

Unter diesen Gesichtspunkten ist die Zuordnung, die der Identifikationsalgorithmus hier vornimmt, als sehr gut zu bewerten.

Die Parameter, die durch den Algorithmus geschätzt werden, sind ebenfalls sehr nahe an den Parametern des Modells, wie die folgende Tabelle zeigt. Dabei sind in der oberen Zeile die originalen Parameter, in der unteren Zeile die geschätzten Parameter angegeben:

Bereich 1			Bereich 2			Bereich 3		
a_1	a_2	c	a_1	a_2	c	a_1	a_2	c
0,5	1	-1	0	0	0,5	0	0	0
0,5975	1,0196	-1,0488	0,0236	-0,0383	0,5183	-0,0044	0,0248	-0,0051

Insgesamt wird das Modell somit trotz der ungünstigen SNR gut identifiziert, was die Leistungsfähigkeit des Identifikationsverfahrens unterstreicht.

Da das bisher betrachtete Beispiel eine eher abstrakte Anwendung darstellt, sollen in den nächsten beiden Abschnitten Beispiele betrachtet werden, die in ihrer Problemgröße und aufgrund ihrer dynamischen Eigenschaften mit praktischen Problemen vergleichbar sind.

5.2.2 Identifikation eines springenden Balls

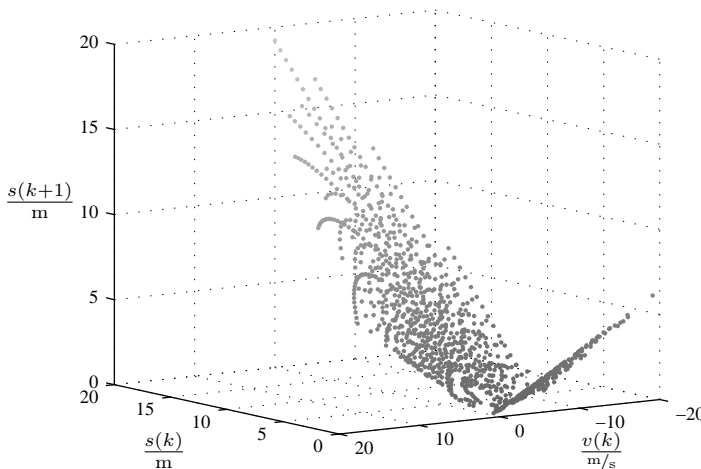


Abbildung 5.17: Abgetastete Daten des springenden Balls: Position des Balls als Folgezustand von Position und Geschwindigkeit

Das Modell des springenden Balls erhält man durch physikalische Modellierung. Der Ball fällt im Schwerfeld der Erde über einer horizontalen Ebene mit einer konstanten Beschleunigung von $9,81 \text{ m/s}^2$, bis er auf die Ebene aufprallt. Der teilelastische Stoß

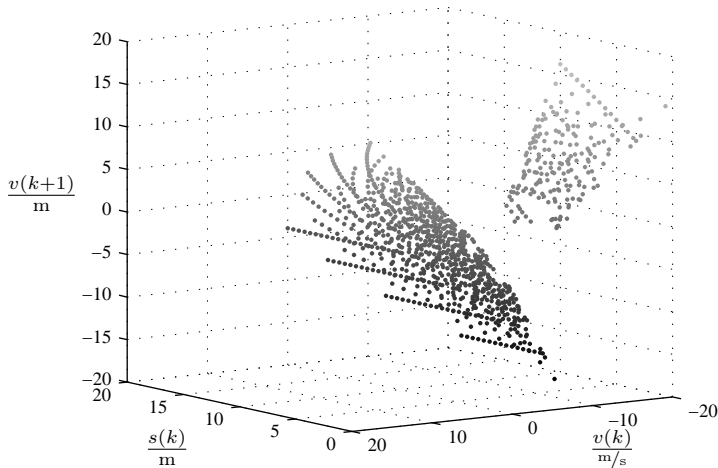


Abbildung 5.18: Abgetastete Daten des springenden Balls: Geschwindigkeit des Balls als Folgezustand von Position und Geschwindigkeit

zwischen der Ebene und dem Ball ist idealisiert so modelliert, dass die Geschwindigkeit des Balls auf 90% verringert wird, sich das Vorzeichen der Geschwindigkeit sprunghaft umkehrt und der Stoß somit momentan abläuft. Daher wird im Modell beim Auftreffen des Balls auf die Ebene die Geschwindigkeit mit -0.9 multipliziert. Die Position des Balls über der horizontalen Ebene ergibt sich aus der Integration der Geschwindigkeit. Die Daten für diesen Versuch wurden gewonnen, indem die Simulation mit verschiedenen Anfangshöhen zwischen 10 m und 20 m gestartet wurde und für jeweils 10 s simuliert wurde; Die Anfangsgeschwindigkeit betrug immer 0 m/s . Abgetastet wurden die Daten mit einer Abtastzeit von $T = 0,4 \text{ s}$. Die resultierenden Daten sind in den Abbildungen 5.17 und 5.18 zu sehen. Dabei sind die Daten mit kleinen Funktionswerten dunkler dargestellt als die Daten mit großen Funktionswerten.

In Abbildung 5.17 ist die Position als Folgezustand von aktueller Position und aktueller Geschwindigkeit dargestellt. Charakteristisch ist der Knick im Kennfeld, der aus der sprunghaftigen Änderung der Geschwindigkeit beim Aufprall des Balls resultiert. Für große Werte der Position liegt die Folgeposition noch vor dem Aufprall des Balls und somit auf der „linken“ Seite des Knicks. Für kleine Werte der Position und betragsmäßig große negative Geschwindigkeiten erfolgt die nächste Abtastung erst nach dem Aufprall des Balls; diese Situationen sind durch den „rechten“ Teil des Kennfelds dargestellt.

Abbildung 5.18 zeigt die Geschwindigkeit als Folgezustand von aktueller Position und aktueller Geschwindigkeit. An der Stelle, an der im vorherigen Kennfeld ein Knick zu sehen ist, befindet sich in diesem Kennfeld ein Sprung, da sich die Geschwindigkeit

des Balls beim Aufprall sprunghaft ändert.

Beim Betrachten der Daten wird sofort klar, dass zur Identifikation des Modells zwei affine Modelle benötigt werden. Zur Bestimmung der Parameter dieser zwei affinen Teilmodelle wurde das Identifikationsproblem verwendet, wie es in Gleichung (4.9) dargestellt ist

$$\min \left\{ \sum_{i=1}^{N-1} \min_j \left\| \underline{y}(i+1) - \underline{A}_j \underline{y}(i) - \underline{B}_j \underline{u}(i) - \underline{c}_j \right\| \right\}.$$

Da das vorliegende Identifikationsproblem keine Eingangsgrößen besitzt, vereinfacht sich das resultierende Optimierungsproblem zu

$$\min \left\{ \sum_{i=1}^{N-1} \min_j \left\| \underline{y}(i+1) - \underline{A}_j \underline{y}(i) - \underline{c}_j \right\| \right\}.$$

Das Modell besitzt zwei Zustandsgrößen; daher sind in jedem affinen Teilmodell 6 Parameter zu bestimmen. Da weiterhin zwei affine Teilmodelle zur Identifikation verwendet werden, muss bei der Identifikation ein 12-dimensionaler Parameterraum durchsucht werden, um die optimale Lösung des angegebenen Problems zu finden. Dieses Problem wurde mit Hilfe des globalen Lösers MCS [HN99] unter MATLAB gelöst. Das resultierende Optimum ist in den Abbildungen 5.19 bis 5.21 dargestellt.

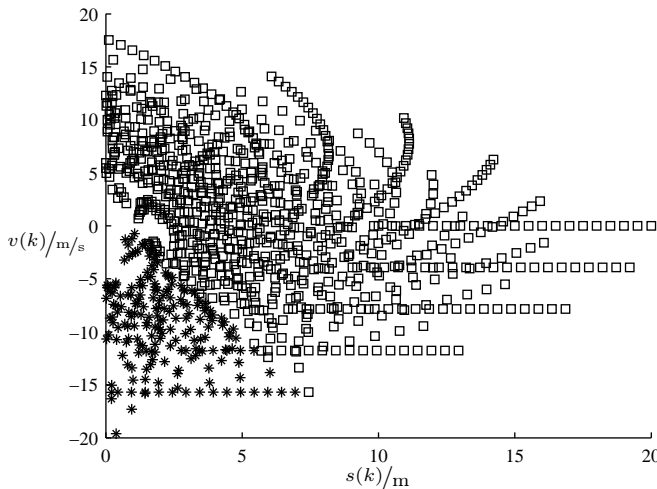


Abbildung 5.19: Zugehörigkeit der Datenpunkte des identifizierten Systems nach der Minimierung mit MCS

In Abbildung 5.19 ist zunächst die Zugehörigkeit der einzelnen Datenpunkte zu den beiden affinen Modellen mit Hilfe zweier unterschiedlicher Symbole gezeigt. Durch die Zugehörigkeit zu den beiden affinen Teilmodellen wird auch der Geltungsbereich

der einzelnen Teilmodelle festgelegt. Da in diesem Beispiel nicht verrauschte Daten betrachtet werden und zusätzlich eine Unstetigkeit im Kennfeld auftritt, ist die Zuordnung der Daten zu den einzelnen Modellen bereits nach der Minimierung des Gütemaßes perfekt, so dass die topologische Optimierung, wie sie in Abschnitt 4.4.2 beschrieben wurde, nicht vorgenommen werden muss.

Das kleinere der beiden Teilmodelle zeigt ein nicht affines Verhalten, wie bereits in [HSK04b, Sch04b] beschrieben wurde. Dies führt dazu, dass dieses Teilmodell nicht exakt durch ein affines Modell repräsentiert werden kann und sich somit auch bei der Simulation des identifizierten Systems ein gewisser Fehler zwischen dem kontinuierlichen Simulationsmodell und dem identifizierten Modell zeigen wird.

Die Abbildungen 5.20 und 5.21 zeigen das identifizierte System, nachdem das Optimierungsproblem mit Hilfe des MCS-Algorithmus gelöst wurde. Beim Vergleich des identifizierten Systems mit den Rohdaten wird deutlich, dass das identifizierte System sehr gut mit den Rohdaten übereinstimmt und deshalb auch ein gutes Simulationsergebnis zu erwarten ist.

In den Abbildungen 5.22 und 5.23 werden die Voraussagen der beiden Modelle für die beiden Zustandsgrößen s und v miteinander verglichen. Hier wird sichtbar, dass der Fehler in der Geschwindigkeit immer dann größer wird, wenn das affine Teilmodell zur Vorhersage des nächsten Zustandes benutzt wird, in dem das nichtlineare Teilmodell approximiert wird. Dies geschieht immer dann, wenn ein Aufprall des Balls zwischen zwei Abtastzeitpunkten liegt. Durch die Integration dieses Fehlers ergibt sich für die Abweichung der Ballposition eine Gerade mit der entsprechenden Steigung.

5.2.3 Identifikation eines Zwei-Tank-Systems

In diesem Abschnitt wird ein hybrides Zwei-Tank-System als Anwendungsbeispiel betrachtet, wie es bereits in [Phu04] beschrieben wurde. Dieses System ist in Abbildung 5.24 dargestellt. Das Zwei-Tank-System besteht aus zwei zylindrischen Tanks, die durch eine Rohrleitung in der Höhe h miteinander verbunden sind. Mit dem Einlassventil V_E kann der Zufluss im Intervall $[0, Q]$ durch die kontinuierliche Eingangsgröße $u_K \in [0, 1]$ eingestellt werden. Die Abflussventile V_1 und V_2 können über einen gemeinsamen diskreten Steuereingang u_D gleichzeitig geöffnet oder geschlossen werden.

Die Systemparameter des Anwendungsbeispiels, ihre physikalische Bedeutung und ihre Werte im Anwendungsbeispiel sind in Tabelle 5.1 zu finden.

Damit durch eine Rechnersimulation Daten gewonnen werden können, muss für das Zwei-Tank-System zunächst ein theoretisches Modell abgeleitet werden. Anschlie-

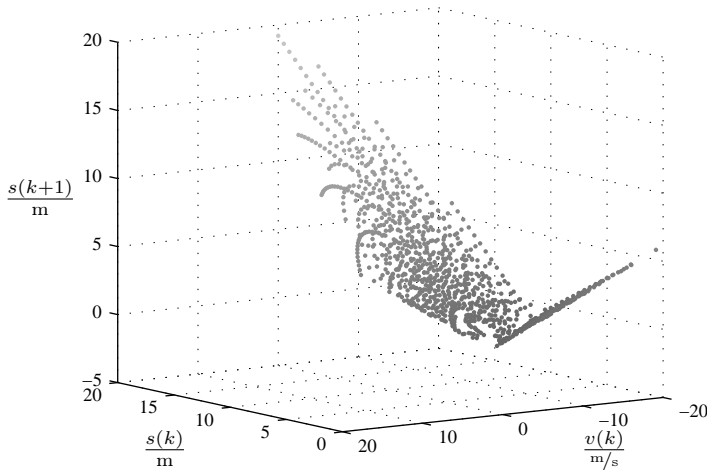


Abbildung 5.20: Kennfeld für die Position des identifizierten Systems nach der Minimierung mit MCS

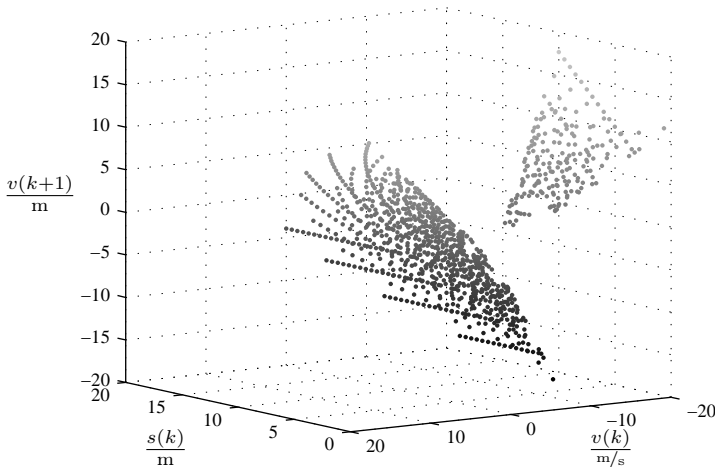


Abbildung 5.21: Kennfeld für die Geschwindigkeit des identifizierten Systems nach der Minimierung mit MCS

ßend soll mit den Daten, die durch Simulationen erzeugt wurden, eine Identifikation des Systemverhaltens vorgenommen werden.

5.2.3.1 Modellbildung des Zwei-Tank-Systems

In der Zustandsraumdarstellung besitzt das System zwei kontinuierliche Zustandsgrößen, nämlich die Füllhöhen der beiden Tanks x_{K1} und x_{K2} , eine kontinuierliche Eingangsgröße – den Zufluss u_K – und die diskrete Eingangsgröße u_D , über die die

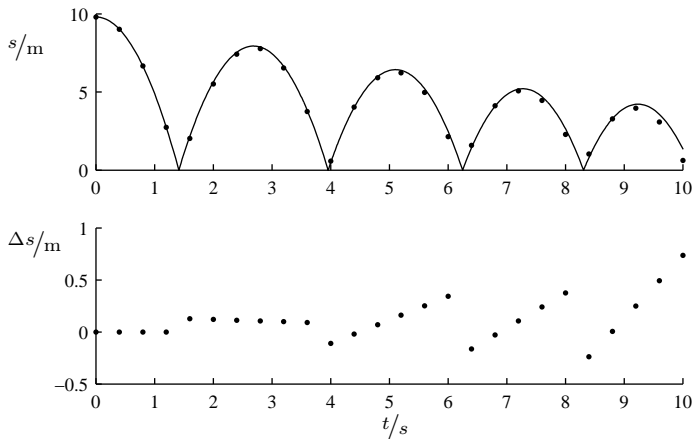


Abbildung 5.22: Oben: Position des Balls für das identifizierte System (\cdot) und das kontinuierliche Simulationsmodell ($-$). Unten: Fehler zwischen identifiziertem System und dem kontinuierlichen Simulationsmodell

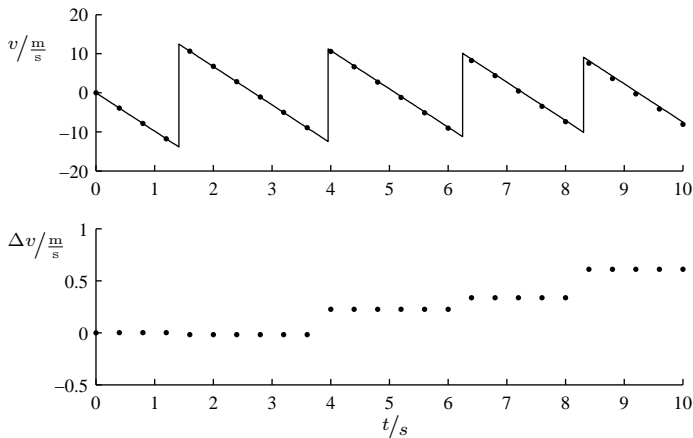


Abbildung 5.23: Oben: Geschwindigkeit des Balls für das identifizierte System (\cdot) und das kontinuierliche Simulationsmodell ($-$). Unten: Fehler zwischen identifiziertem System und dem kontinuierlichen Simulationsmodell

diskreten Auslassventile V_1 und V_2 gleichzeitig beeinflusst werden.

Die hybriden Phänomene, die beim Zwei-Tank-System auftreten, lassen sich auf das Öffnen bzw. Schließen der Auslassventile oder auf das Über- bzw. Unterschreiten der Höhe des Verbindungsrohres durch die Füllhöhe der beiden Tanks zurückführen.

Zur Modellbildung des Systems kann nun mit Hilfe von physikalischen Gleichungen ein Netz-Zustands-Modell erstellt werden, wie es in Abschnitt 2.3.3 vorgestellt wurde. Dazu werden zunächst die Differenzialgleichungen herangezogen, die sich aus den

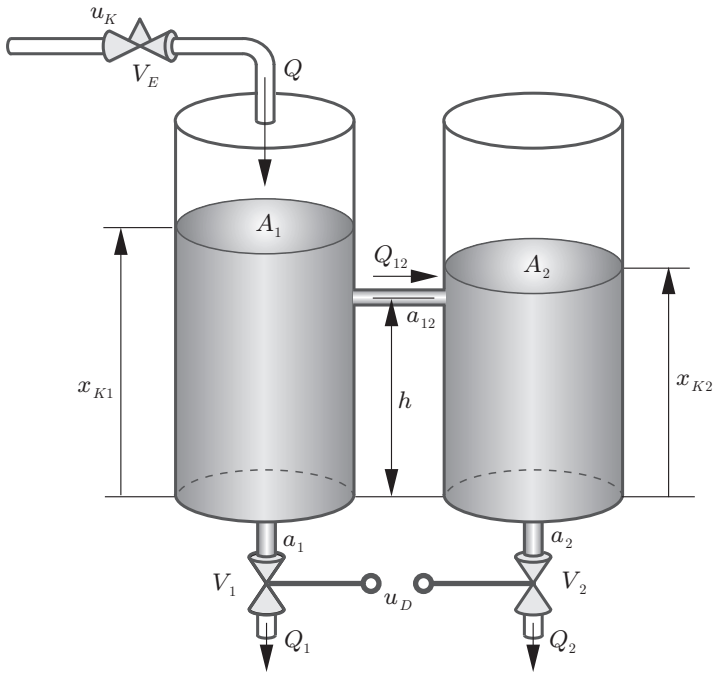


Abbildung 5.24: Zwei-Tank-System

Volumenströmen ableiten lassen

$$\begin{aligned} \frac{dx_{K1}}{dt} &= \frac{Q \cdot u_K - Q_{12} - Q_1}{A_1} \\ \frac{dx_{K2}}{dt} &= \frac{Q_{12} - Q_2}{A_2}. \end{aligned} \quad (5.1)$$

Dabei bezeichnen Q_1 und Q_2 die Volumenströme durch die Auslassventile V_1 und V_2 , Q stellt den Zufluss über das Ventil V_E dar, und Q_{12} gibt den Volumenstrom an, der von Tank 1 nach Tank 2 strömt. Dabei ist das Vorzeichen des Volumenstroms so gewählt, dass ein positiver Volumenstrom vorliegt, wenn Flüssigkeit von Tank 1 nach Tank 2 fließt; im umgekehrten Fall ist das Vorzeichen negativ.

Mit Hilfe des Gesetzes von Toricelli lassen sich bei geöffneten Ventilen V_1 und V_2 für die Volumenströme Q_1 und Q_2 die Gleichungen

$$\begin{aligned} Q_1 &= a_1 \cdot \sqrt{2 \cdot g x_{K1}} \quad \text{und} \\ Q_2 &= a_2 \cdot \sqrt{2 \cdot g x_{K2}} \end{aligned}$$

	Beschreibung	Wert	Einheit
A_1 :	Querschnittsfläche von Tank 1	0,0154	m^2
A_2 :	Querschnittsfläche von Tank 2	0,0154	m^2
a_1 :	Querschnitt des Auslassrohrs von Tank 1	$3,14 \cdot 10^{-4}$	m^2
a_2 :	Querschnitt des Auslassrohrs von Tank 2	$3,14 \cdot 10^{-4}$	m^2
a_{12} :	Querschnitt der Rohrleitung zwischen beiden Tanks	$3,14 \cdot 10^{-4}$	m^2
h_{\max} :	Höhe der Tanks	0,6	m
h :	Höhe der Verbindungsrohrleitung zwischen beiden Tanks	0,3	m
V_E :	kontinuierliches Einlassventil		
V_1 :	diskretes Auslassventil von Tank 1		
V_2 :	diskretes Auslassventil von Tank 2		
Q :	maximaler Durchfluss durch das Einlassventil V_E bei $u_K = 1$	$8 \cdot 10^{-4}$	m^3/s
Q_1 :	Volumenstrom durch das Abflussventil von Tank 1		m^3/s
Q_2 :	Volumenstrom durch das Abflussventil von Tank 2		m^3/s
Q_{12} :	Volumenstrom in der Rohrleitung zwischen beiden Tanks		m^3/s

Tabelle 5.1: Parameter des Zwei-Tank-Systems

ableiten. Falls V_1 und V_2 geschlossen sind, sind die Volumenströme 0.

Für den Volumenstrom Q_{12} sind vier Fälle zu unterscheiden:

$$Q_{12} = \begin{cases} \sqrt{2 \cdot g \cdot |x_{K1} - x_{K2}|} \cdot \text{sign}(x_{K1} - x_{K2}) & \text{für } x_{K1} > h \wedge x_{K2} > h \\ -\sqrt{2 \cdot g \cdot (x_{K2} - h)} & \text{für } x_{K1} \leq h \wedge x_{K2} > h \\ \sqrt{2 \cdot g \cdot (x_{K1} - h)} & \text{für } x_{K1} > h \wedge x_{K2} \leq h \\ 0 & \text{sonst.} \end{cases}$$

Damit ist die Beschreibung des kontinuierlichen Anteils vollständig.

Der diskrete Anteil setzt sich aus drei Phänomenen zusammen, nämlich aus der Über- bzw. Unterschreitung der Höhe des Verbindungsrohres durch die Füllhöhen und der

Umschaltung der Auslassventile. Diese Phänomene können durch vier Triggermengen

$$\begin{aligned}\Omega_1 &= \{\mathcal{X}_K/x_{K1} > h\} \\ \Omega_2 &= \{\mathcal{X}_K/x_{K1} \leq h\} \\ \Omega_3 &= \{\mathcal{X}_K/x_{K2} > h\} \\ \Omega_4 &= \{\mathcal{X}_K/x_{K2} \leq h\}\end{aligned}$$

nachgebildet werden.

Die Komponenten des Ausgangsvektors \underline{v}_D des Kontinuierlich/Diskreten-Interfaces sind von diesen vier Triggermengen abhängig; v_{Di} ist 1, solange sich der kontinuierliche Zustand innerhalb der entsprechenden Triggermenge befindet. Der Vektor \underline{v}_D und der Eingang für die Abflussventile u_D beeinflussen die diskreten Zustände.

Über das Diskret/Kontinuierliche-Interface wird die für den aktuellen diskreten Zustand gültige kontinuierliche Dynamik ausgewählt. Dazu werden die im NZM in Abbildung 5.25 gezeigten booleschen Gleichungen verwendet.

In Abbildung 5.25 wird das gesamte NZM für das Zwei-Tank-System dargestellt.

Die Implementierung dieses Modells erfolgte in MATLAB und SIMULINK, wobei der diskrete Systemanteil mit Hilfe von STATEFLOW modelliert wurde.

5.2.3.2 Identifikation des Zwei-Tank-Systems

Zur Identifikation des dynamischen Verhaltens des Zwei-Tank-Systems muss festgelegt werden, welche Größen für die Identifikation verwendet werden. Da die beiden Zustandsgrößen x_{K1} und x_{K2} messbar sind, können sie auch zur Identifikation benutzt werden. Ebenso wird die Einflussgröße u_K des kontinuierlichen Einlassventils benötigt. Da das dynamische Verhalten identifiziert werden soll, muss das identifizierte Modell mit Hilfe dieser Größen die Zustände im nächsten Abtastschritt vorhersagen können. Daher wird in diesem Beispiel ein dreidimensionaler Identifikationsraum betrachtet, der von den beiden Zustandsgrößen und der Eingangsgröße aufgespannt wird; über diesem Identifikationsraum muss nun das stückweise affine Modell identifiziert werden.

Für die Identifikation wird hier der Fall betrachtet, dass die diskreten Auslassventile geöffnet und die Messdaten gleichmäßig im gesamten Eingangsraum verteilt sind.

Um den Identifikationsraum gleichmäßig mit Daten zu füllen, wird dieser mit einem achsenparallelen Gitter aufgeteilt. An den Gitterpunkten werden durch Simulation die Folgezustände berechnet. Anschließend werden diese simulierten Daten durch ein gaußsches Rauschen mit Erwartungswert $m = 0$ und einer Varianz von $\sigma^2 = \left(\frac{0.01}{3}\right)^2$ verfälscht. Die Varianz wurde so gewählt, damit die 3σ Grenze der Dichtefunktion

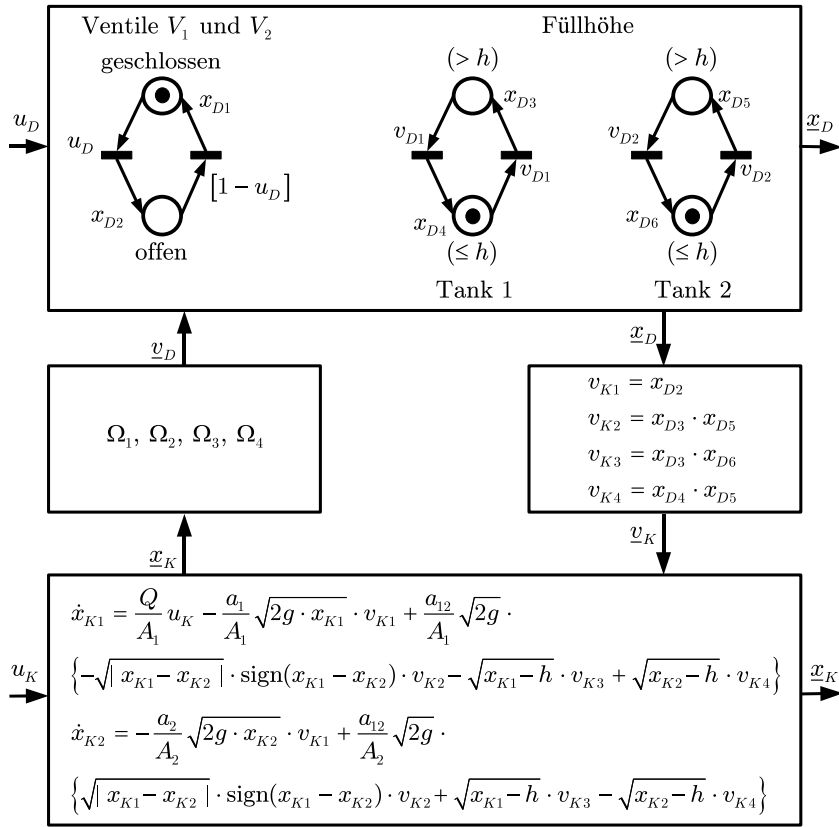


Abbildung 5.25: Netz-Zustand-Modell des Zwei-Tank-Systems

gerade 1 cm entspricht. Mit den so gewonnen Daten wird die Identifikation nun durchgeführt.

Da die Daten vierdimensionale Vektoren sind, wird bei den folgenden Darstellungen der Daten die Eingangsgröße $u_K = 0,6$ gesetzt. In Abbildung 5.26 sind zunächst die Daten ohne Rauschen dargestellt. Die Daten mit Rauschen sind in Abbildung 5.27 zu sehen.

Zur Identifikation des Zwei-Tank-Systems nach der Methode, die in Abschnitt 4.4 beschrieben ist, muss zunächst die Anzahl der Teilmodelle durch Vorkenntnisse über die Charakteristik des Systems bestimmt werden. Beim Zwei-Tank-System ist aufgrund des Aufbaus klar, dass es wegen der Verbindungsleitung zwischen den beiden Tanks mindestens vier Teilmodelle geben muss.

Da eine weitere interne Umschaltung der Dynamik in Abhängigkeit des Füllstandverhältnisses erfolgt, wenn beide Füllstände die Höhe der Verbindungsleitung über-

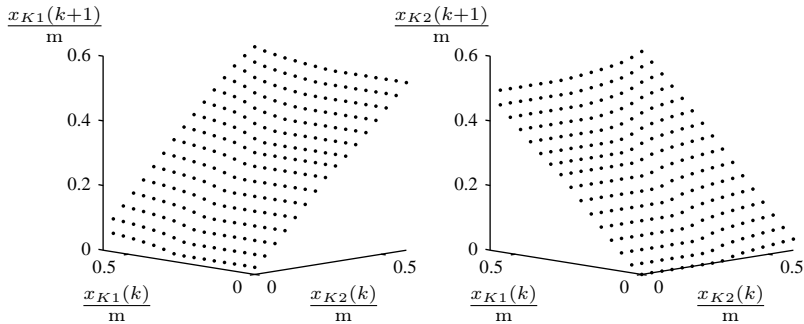
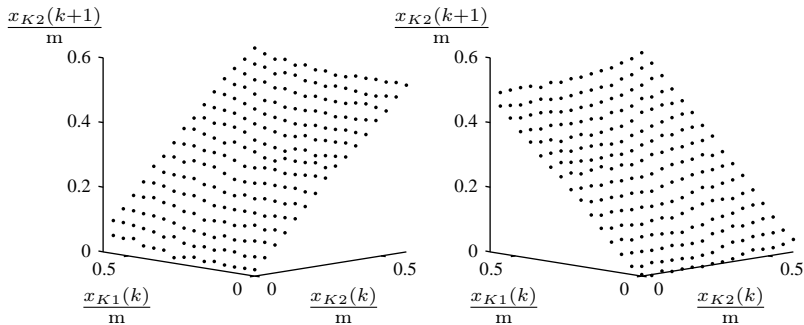


Abbildung 5.26: Ideale Daten des Zwei-Tank-Systems

Abbildung 5.27: Verrauschte Daten des Zwei-Tank-Systems mit $\sigma^2 = \left(\frac{0.01}{3}\right)^2$

schritten haben, werden weitere Teilmodelle bei der Identifikation benötigt. Des Weiteren sind einige Teilmodelle aufgrund der Wurzelfunktion in den Zustandsgleichungen nichtlinear. Diese nichtlinearen Teilmodelle müssen ebenfalls mit mehreren affinen Teilmodellen angenähert werden. Deshalb wird das Systemverhalten in diesem Beispiel mit sechs Teilmodellen identifiziert.

Zur Identifikation muss weiterhin entschieden werden, in welcher Form das Optimierungsproblem gelöst werden soll. Da das System einen dreidimensionalen Identifikationsraum besitzt und als System 2. Ordnung dargestellt werden soll, sind für jedes affine Teilmodell 8 Parameter zu schätzen. Da weiterhin 6 Teilmodelle verwendet werden, sind insgesamt 48 Parameter zu schätzen. Dieser Parameterraum kann von Lösern, wie dem MCS-Löser [HN99], die garantiert ein globales Minimum finden, nicht mehr gehandhabt werden. Deshalb wurde das Optimierungsproblem in der Form von Problem (4.5) verwendet und mit dem Löser MINLP gelöst [Ley98]. Diese Darstellung weist zwar den Nachteil auf, dass die Darstellung gemischt-ganzzahlig ist und wegen der großen Punkteanzahl von $n = 2816$ entsprechend viele Zuordnungsvariablen betrachtet werden müssen, allerdings garantiert diese Vorgehensweise, dass das globale Minimum gefunden wird. Bei der Lösung des nicht konvexen

Problems (4.6) mit Abstiegsverfahren erhält man hingegen meist nur unbrauchbare lokale Optima.

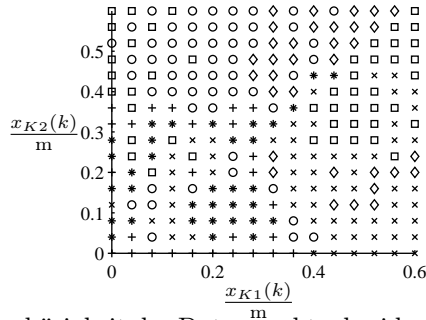


Abbildung 5.28: Zugehörigkeit der Datenpunkte des identifizierten Systems nach der Minimierung mit MINLP

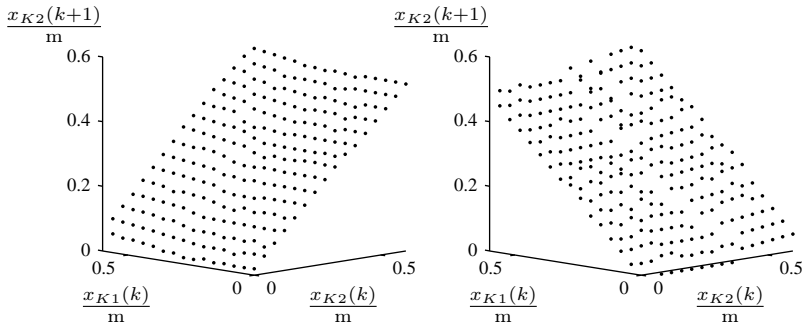


Abbildung 5.29: Kennfelder des identifizierten Systems nach der Minimierung mit MINLP

Die Identifikation des Systems erfolgt nun durch die numerische Minimierung des Optimierungsproblems (4.5). Das Ergebnis dieser Minimierung ist in den Abbildungen 5.28 und 5.29 als Schnitte durch den vierdimensionalen Raum bei $u_K = 0,6$ dargestellt.

Die Zugehörigkeit der Datenpunkte zu den Teilmodellen wird in Abbildung 5.28 durch die unterschiedlichen Symbole gezeigt. Abbildung 5.29 stellt das Folgezustandskennfeld dar. Wie zu erwarten ist, ergeben sich durch die ermittelten Zugehörigkeiten zu den Teilmodellen keine zusammenhängende Gebiete, weshalb noch eine topologische Optimierung der Geltungsbereiche der einzelnen Teilmodelle durchgeführt wird.

5.2.3.3 Topologische Optimierung des Zwei-Tank-Systems

Zunächst werden die größten zusammenhängenden Komponenten der identifizierten Teilmodelle gesucht. Diese größten Komponenten müssen wieder im dreidimensionalen Identifikationsraum betrachtet werden. Für die graphische Darstellung wäre dies aufgrund der Datenmenge sehr unübersichtlich. Deshalb wird auch hier wieder nur ein zweidimensionaler Schnitt dargestellt, bei dem der Zufluss $u_K = 0,6$ gewählt wird. In Abbildung 5.30 sind die Datenpunkte dargestellt, die zu einer der größten Komponenten gehören und gleichzeitig in der Schnittebene $u_K = 0,6$ liegen. Die verschiedenen Symbole, die zur Darstellung der Datenpunkte gewählt wurden, stehen dabei für die sechs verschiedenen größten Komponenten.

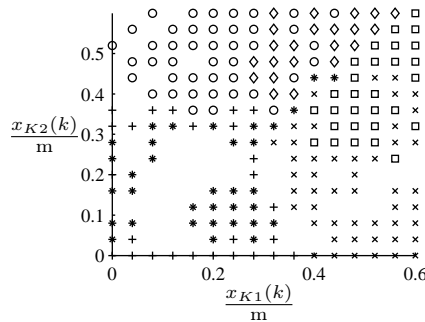


Abbildung 5.30: Größte zusammenhängende Komponenten

Als nächstes werden die Datenpunkte, die nicht zu einer größten Komponente gehören, mit Hilfe des in Abschnitt 4.4.2 vorgestellten Optimierungsalgorithmus zu den größten Komponenten hinzugefügt. Danach besitzt das System eine neue Zuordnung, die in Abbildung 5.31 wiederum mit Hilfe des Schnittes bei $u_K = 0,6$ dargestellt ist. Die Parameter der einzelnen affinen Teilmodelle werden auf Grund-

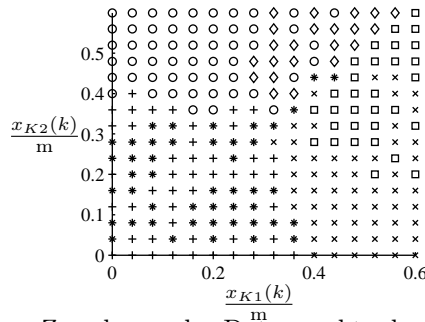


Abbildung 5.31: Neue Zuordnung der Datenpunkte des Systems nach der topologischen Optimierung

lage dieser neuen Zuordnung mit Hilfe der Methode der kleinsten Quadrate noch einmal identifiziert. Das so optimierte System ist in Abbildung 5.32 zu sehen. Mit

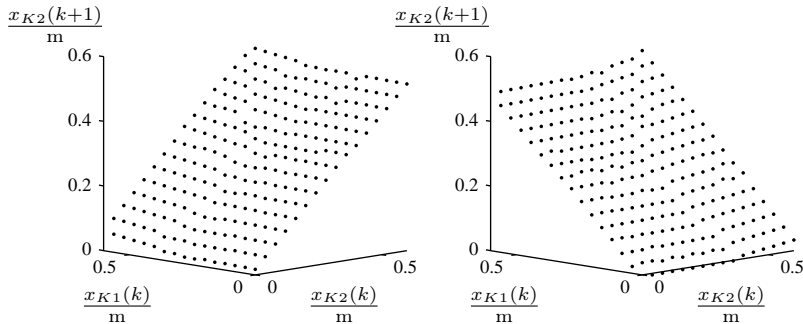


Abbildung 5.32: Kennfelder des identifizierten Systems nach Minimierung und topologischer Optimierung

Hilfe des so erhaltenen Modells, kann die Füllhöhe der beiden Tanks im nächsten Abtastschritt in Abhängigkeit der aktuellen Füllhöhen und der aktuellen Eingangsgrößen vorhergesagt werden. Mit Hilfe der Rückkopplung der vorhergesagten Werte und mit einem festgelegten Eingangsgrößenverlauf kann somit eine Simulation des Systems durchgeführt werden. Dies erfolgt im nächsten Abschnitt.

5.2.3.4 Simulation des Zwei-Tank-Systems

Um das identifizierte System zu simulieren, müssen die Füllhöhen der beiden Tanks im nächsten Abtastschritt – ausgehend von den aktuellen Füllhöhen der beiden Tanks und dem Zufluss – berechnet werden. Hier tritt das Problem auf, dass der Punkt, der die Eingangsgröße im Eingangsraum darstellt, normalerweise nicht genau auf den Datenpunkten liegt, die im Eingangsraum gleichmäßig verteilt sind. Daher muss für einen solchen Anfragepunkt zunächst ein Teilmodell ausgewählt werden. Dies geschieht durch die Annahme, dass der Anfragepunkt zum gleichen Modell gehört wie der Datenpunkt, der dem Anfragepunkt im Eingangsraum am nächsten liegt. In der Simulation wird also zuerst das Datum gesucht, das am nächsten zum Anfragepunkt liegt. Anschließend wird der Folgezustand aus dem aktuellen Zustand und den Eingangsgrößen berechnet. Durch Rückkopplung des Folgezustandes kann eine Simulation auch über größere Zeiträume durchgeführt werden. Da bei den extremen Füllständen ganz voller bzw. ganz leerer Tank das nächste affine Modell bei der Simulation verwendet werden würde, fände eine Extrapolation der affinen Modelle statt, die zu sinnlosen Simulationsergebnissen führen würde. Deshalb muss an den Rändern des Modells zusätzliches Wissen für eine erfolgreiche Simulation eingebracht werden: ist der ermittelte Folgezustand kleiner als 0 m, so wird der Folgezustand auf

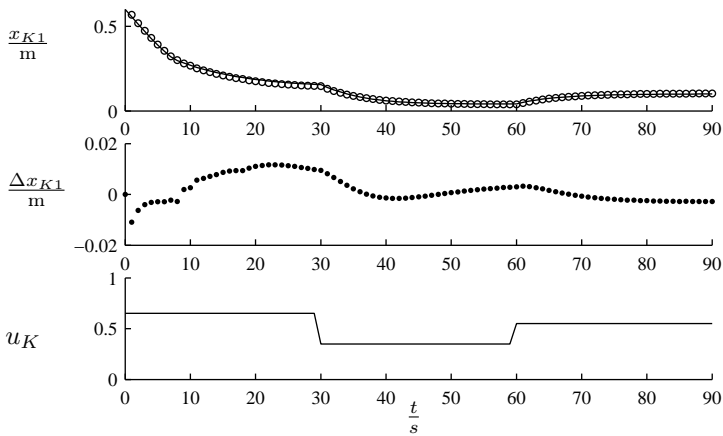


Abbildung 5.33: Ergebnis der Simulation für den Füllstand des ersten Tanks

0 m gesetzt, ist der Folgezustand größer als 0,6 m, so wird der Folgezustand auf 0,6 m gesetzt.

Als Anfangszustand wird nun $x_{K1} = 0,6$ und $x_{K2} = 0,6$ gewählt. Weiterhin wird die Eingangsgröße für 30 Sekunden zu 0,65, für weitere 30 Sekunden zu 0,35 und schließlich für 30 Sekunden zu 0,55 gewählt. Die sich so ergebende 90 sekündige Simulation ist in den Abbildungen 5.33 und 5.34 gezeigt. Im oberen Teil wird jeweils die Füllhöhe des theoretischen Modells und die des identifizierten Modells gezeigt. Im mittleren Teil ist die Differenz zwischen den beiden Simulationen aufgetragen; die Eingangsgröße u_K ist jeweils im unteren Teil dargestellt.

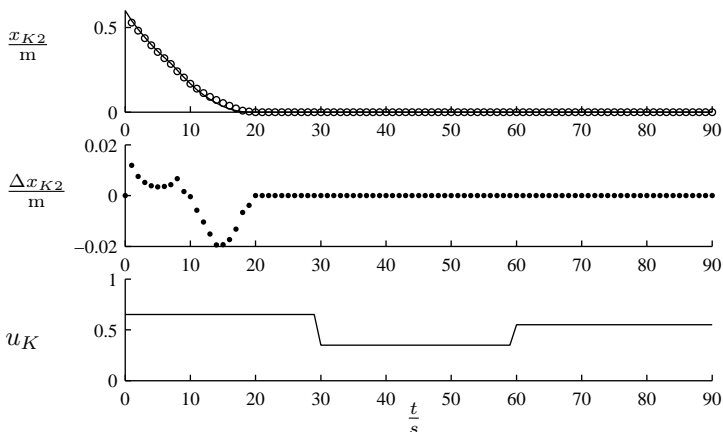


Abbildung 5.34: Ergebnis der Simulation für den Füllstand des zweiten Tanks

Insgesamt ist festzustellen, dass das identifizierte System sehr gut mit dem theo-

retischen System übereinstimmt. Während des Verlaufs der Simulation wurde der Zufluss zweimal geändert. Die erste Änderung erfolgte bei 30 und die zweite Änderung bei 60 Sekunden. Man erkennt, dass der Fehler nach der zweiten Änderung einen nicht glatten Übergang besitzt. Dieses Phänomen zeigt, dass die Dynamik des Systems durch die Änderung des Zuflusses stark beeinflusst wird, so wie es auch zu erwarten ist. Um den Fehler weiter zu verringern, müsste die Anzahl der Teilmodelle erhöht werden.

Die Abweichungen zwischen beiden Simulationen haben zweierlei Ursachen:

Zum einen wurden die Daten durch Rauschen verfälscht und zum anderen kann das System nicht perfekt durch affine Teilgebiete nachgebildet werden, da eine Wurzelkennlinie in der Systembeschreibung des Originalsystems auftritt.

Vergleich der beiden Verfahren

Vergleicht man nun die beiden Verfahren zur Identifikation stückweise affiner Modelle, so lässt sich feststellen, dass das Verfahren der lokalen Kennfeldoptimierung eine sehr anschauliche Vorgehensweise darstellt. Dieses Verfahren ist iterativ und verbessert sukzessive eine Anfangslösung, so dass es auch jederzeit vom Anwender unterbrochen werden kann und trotzdem eine gültige Lösung des Identifikationsproblems vorliegt. Diese anschauliche Vorgehensweise wird allerdings mit einer relativ großen Rechenzeit erkaufte. Zudem geht die Anschauung verloren, wenn der Identifikationsraum mehr als zwei Dimensionen besitzt. Die weiteren Forderungen, die in Abschnitt 4.2.5 gestellt worden sind, werden von diesem Verfahren erfüllt.

Das zweite Verfahren besitzt seine Stärke bei der Identifikation größerer Modelle, da es weniger Rechenzeit in Anspruch nimmt als das Verfahren der lokalen Kennfeldoptimierung. Es ist nicht ganz so anschaulich wie dieses und daher wird auch das Einbringen von Vorwissen etwas erschwert. Ebenso wie das Verfahren zur lokalen Kennfeldoptimierung erfüllt es die Forderungen, die in Abschnitt 4.2.5 an mögliche Identifikationsverfahren gestellt worden sind.

Damit stehen – wie in den Anwendungsbeispielen bereits gezeigt wurde – zwei leistungsfähige Verfahren zur Identifikation stückweise affiner Systeme zur Verfügung.

Teil II

**Diagnose hybrider Systeme
mit diskreten Sensoren**

Kapitel 6

Grundlagen zur Diagnose hybrider dynamischer Systeme

Komplexer werdende industrielle Anlagen und der damit verbundene steigende Kapitaleinsatz haben zur Folge, dass ein fehlerfreier Betrieb über große Zeiträume gewährleistet sein muss. Damit kommt auch der Überwachung dieser Anlagen und der Fehlerdiagnose eine immer größer werdende Bedeutung zu.

Für die Diagnose von dynamischen Systemen wurden bereits zahlreiche Verfahren entwickelt. Dabei hat sich bei fast allen Verfahren eine mehrstufige Vorgehensweise etabliert. Ausgehend vom fehlerfreien Fall wird das Systemverhalten mit dem Modellverhalten verglichen und mögliche Abweichungen werden analysiert. Stellt man eine Abweichung fest, die auf einen Fehler schließen lässt, so ist damit der erste Schritt der Fehlerdiagnose, nämlich die *Fehlerdetektion*, abgeschlossen. Anschließend muss die Frage beantwortet werden, welcher Teil des Prozesses die Abweichung im Verhalten verursacht hat; dieser Schritt wird als *Fehlerisolation* bezeichnet. Danach ist von Interesse, ob der Prozess mit dem vorliegenden Fehler weiter betrieben werden kann. Um dies zu entscheiden, werden häufig Informationen über die Quantität des Fehlers benötigt, weshalb versucht wird, die technischen Parameter des Prozesses – insbesondere der fehlerhaften Teile und Baugruppen – aufgrund der Abweichungen neu zu ermitteln. Dieser Schritt wird hier als *Fehleridentifikation* bezeichnet.

Die Bezeichnung der einzelnen Schritte ist in der Literatur nicht einheitlich; so gliedert beispielsweise Frank [Fra94] den Diagnosevorgang in die Abschnitte Fehlerdetektion, Fehlerisolation und Fehleranalyse. Die hier verwendeten Ausdrücke werden in Abschnitt 6.1 definiert.

In Abschnitt 6.2 werden die verschiedenen Diagnoseverfahren wie wissensbasierte, signalgestützte und modellgestützte Verfahren kurz umrissen und ihre Funktionsweise wird erläutert.

Die zur Zeit in der Literatur bekannten Diagnoseverfahren gehen meistens davon aus, dass viele Größen des Systems kontinuierlich gemessen werden können; dies trifft besonders auch für die Verfahren zu, die für hybride Systeme entwickelt wurden [MKB95, MB97, Mos97]. Aber auch bei Verfahren, die für andere Systemklassen entwickelt wurden, ist diese Voraussetzung weit verbreitet. Stellvertretend seien hier [PFC89, Nyb99, PFC00, PULT00, SFP03] genannt.

Die Aufgabe, die dagegen in dieser Arbeit im Mittelpunkt steht, ist die Entwicklung von Diagnoseverfahren für hybride Systeme mit diskreter Sensorik. Dabei wird auch aufgezeigt, warum bestehende Verfahren nicht ohne weiteres für diesen Anwendungsfall erweitert werden können.

Um diese Diagnoseverfahren entwickeln zu können, wird in einem ersten Arbeitsschritt ein Beobachter entworfen, der aufgrund diskreter Messungen den Zustand eines hybriden Systems rekonstruiert [Gis03]. Dieser Beobachter wird in Abschnitt 6.3 vorgestellt. Darauf aufbauend wird in Kapitel 7 ein beobachtergestütztes Diagnoseverfahren vorgestellt. Dieses Diagnoseverfahren arbeitet mit diskreten Messungen und wurde in [Röm04] entwickelt.

6.1 Begriffsbildung

In diesem Abschnitt soll zunächst die Begriffsbildung für die Diagnose vorgenommen werden. Außerdem werden einige Begriffe erläutert, die für das weitere Verständnis der Arbeit notwendig sind.

Der Begriff *Diagnose* wird dabei wie folgt verstanden [wik05]:

Definition 6.1

Diagnose (griechisch: dia = durch, gnosein = kennen) ist der Vorgang der Erkenntnisgewinnung durch die Zuordnung eines Phänomens oder einer Gruppe von Phänomenen zu einer Kategorie. Auch das Resultat einer solchen Klassifizierung bezeichnet man als Diagnose.

Um eine Verwechslung der beiden unterschiedlichen Begriffe zu vermeiden, wird in dieser Arbeit das Resultat der Klassifizierung *Diagnoseergebnis* genannt.

Einen guten Überblick über die Begriffe, die in der automatisierungstechnischen Diagnose verwendet werden, gibt [Lem95]. Hier finden sich Verweise auf zahlreiche DIN-Normen, die in diesem Zusammenhang von Bedeutung sind.

Bevor die Bedeutung des Begriffs *Diagnose* im weiteren Verlauf genauer erläutert wird, sind zunächst noch einige Definitionen vorzunehmen.

Definition 6.2: Fehler [Lem95]

Als Fehler Δy wird in einem technischen System eine *unzulässige* Abweichung eines Merkmales y_M vom Sollverhalten y_S , also $\Delta y = y_M - y_S > \varepsilon$, bezeichnet.

Ein Merkmal ist im Falle der Diagnose technischer Systeme eine messbare Größe, die von ihrem Sollwert abweicht.

Je nach Art der Abweichung kann ein Fehler in verschiedene Kategorien eingeteilt werden. Es werden hier Begriffe wie *zufälliger Fehler*, *schleichender Fehler*, *grober Fehler* und *mehrfacher Fehler* verwendet, die im Folgenden definiert werden.

Definition 6.3: Zufälliger Fehler

Zufällige Fehler sind Fehler $\Delta y \leq \varepsilon$, die so klein sind, dass der Messwert y_M noch innerhalb eines Vertrauensintervalls um den Sollwert liegt. Darüber hinaus haben die festgestellten Abweichungen den Mittelwert $\overline{\Delta y} = 0$, so dass $\sum_{i=0}^n \Delta y_i/n \rightarrow 0$ für $n \rightarrow \infty$ gilt.

Zufällige Fehler können durch Messrauschen oder andere Rauschprozesse verursacht werden.

Definition 6.4: Schleichender Fehler

Ein Fehler wird als schleichend bezeichnet, wenn die Abweichung $\Delta y < \varepsilon$ so gering ist, dass der Messwert y_M noch innerhalb des Vertrauensintervalls liegt. Weiterhin weisen die Abweichungen jedoch einen Mittelwert $\overline{\Delta y} \neq 0$ auf, so dass $\sum_{i=0}^n \Delta y_i/n \rightarrow \delta$ mit $\delta \neq 0$ für $n \rightarrow \infty$ gilt.

Typische Beispiele für schleichende Fehler sind Verschleiß- und Ermüdungserscheinungen sowie verengte Leitungsquerschnitte, die z.B. durch Kalkablagerungen entstehen, aber auch sehr kleine Leckagen.

Definition 6.5: Grober Fehler

Grobe Fehler sind Fehler Δy , für die $\Delta y > \varepsilon$ gilt, so dass die Messwerte nicht mehr innerhalb eines Vertrauensintervalls liegen.

Plötzliches Versagen oder Ausfall von Bauteilen und Maschinen, Verstopfen einer Rohrleitung durch Fremdkörper sowie größere Leckagen in Rohrleitungen sind Beispiele für grobe Fehler.

Unabhängig von den bisherigen Fehlern wird der mehrfache Fehler wie folgt definiert.

Definition 6.6: Mehrfacher Fehler

Treten mehrere Fehler in zeitlichem Abstand so nacheinander auf, dass der erste Fehler bereits als solcher erkannt werden konnte (Fehlerdetektion) und dessen Diagnose abgeschlossen ist, bevor der nächste Fehler eintritt, wird von mehrfachen Fehlern gesprochen.

Aufgabe eines Diagnosesystems ist es zunächst zu entscheiden, ob ein Fehler aufgetreten ist oder ob sich der Prozess innerhalb der *normalen Spezifikation* befindet. Diese Entscheidung muss dem Bedienpersonal mitgeteilt werden. Die Mitteilung erfolgt häufig in Form von Alarmen.

Definition 6.7: Fehlalarm

Der Begriff *Fehlalarm* entspricht dem so genannten *Fehler erster Art*. Ein Fehler erster Art liegt vor, wenn die so genannte Ausgangshypothese H_0 abgelehnt wird, obwohl sie richtig ist.

Von einem Fehler erster Art spricht man, wenn man einen Fehler annimmt, der in Wirklichkeit nicht vorhanden ist. Die Ausgangshypothese H_0 besagt, dass das System fehlerfrei arbeitet. Diese Hypothese wird jedoch zu Unrecht abgelehnt. Die folgende Definition dient in erster Linie dazu, den Fehler erster und zweiter Art deutlich voneinander abzugrenzen und ergibt sich eigentlich von selbst.

Definition 6.8: nicht erfolgter Alarm

Der Begriff *nicht erfolgter Alarm* entspricht dem so genannten *Fehler zweiter Art*. Ein Fehler zweiter Art liegt vor, wenn die Ausgangshypothese H_0 nicht abgelehnt wurde, obwohl sie falsch ist. Hier wird also nicht erkannt, dass Abweichungen vom Sollwert vorliegen.

Von einem Fehler zweiter Art spricht man, wenn man es versäumt, einen Fehler, der in Wirklichkeit vorhanden ist, zu erkennen. Die Ausgangshypothese H_0 besagt, dass das System fehlerfrei arbeitet. Diese Hypothese wird jedoch nicht abgelehnt, obwohl sie falsch ist. Natürlich ist ein Fehler zweiter Art für ein technisches System weitaus verheerender als ein Fehler erster Art, der lediglich eine entsprechende Reaktion des Bedienpersonals auf eine fälschlich angenommene Abweichung hervorruft. Es kann aber ebenso verheerend sein, wenn ein Fehler zwar als solcher erkannt wird, die Diagnose jedoch eine falsche Ursache ermittelt. Dies führt zum Begriff des Falschalarms.

Definition 6.9: Falschalarm

Ein *Falschalarm* liegt vor, wenn in Wirklichkeit zwar ein Fehler aufgetreten ist, der vom Diagnosesystem isolierte – und möglicherweise auch identifizierte – Fehler jedoch ein anderer als der tatsächliche ist.

Im Rahmen dieser Arbeit soll der Vorgang der Diagnose in die im Folgenden beschriebenen drei Teilschritte zerlegt werden.

6.1.1 Fehlerdetektion

Das Erkennen von Abweichungen ist die Grundlage jeglicher Diagnose. Der Begriff des Fehlers wird nach DIN40041 als *unzulässige Abweichung eines Merkmales vom Sollverhalten* definiert. Dies bedeutet, dass der Vorgang der Fehlerdetektion grundsätzlich einen Soll-Istwert-Vergleich beinhaltet und nach dessen Ergebnis entschieden werden muss, ob ein Fehler vorliegt oder nicht. Da Soll- und Istwerte infolge von Messfehlern, Messrauschen und anderen Störeinflüssen niemals vollständig übereinstimmen, müssen geeignete Verfahren eingesetzt werden, die die festgestellten Abweichungen bewerten und schließlich entscheiden, ob eine festgestellte Abweichung als Fehler einzustufen ist. Die Differenz zwischen der aktuellen, gemessenen oder geschätzten, und der nominellen Referenzgröße wird nach [Fra94] als *Residuum* bezeichnet.

Definition 6.10: Fehlerdetektion

Fehlerdetektion bezeichnet das Erkennen von unzulässigen Abweichungen (Fehlern) zwischen dem realen System und dessen Prozessmodell.

6.1.2 Fehlerisolation

Nachdem im ersten Schritt ein Residuum generiert und das Auftreten eines Fehlers erkannt wurde, folgt nun dessen Auswertung. Ziel dieser Auswertung ist es, den Fehler einzugrenzen, ihn einer Baugruppe, einem Bauteil oder – wie im Rahmen dieser Arbeit – direkt einem Parameter des Prozessmodells zuzuordnen. Gelingt die Zuordnung, ist der dritte Schritt der Diagnose – die Fehleridentifikation – eindeutig. Gelingt keine Fehlerisolation, kann es sich um einen Fehlalarm oder um einen schwer isolierbaren Fehler handeln. Im Falle der schwer isolierbaren Fehler stehen mehrere mögliche Ursachen für das fehlerhafte Verhalten des Systems zur Auswahl, von denen der Algorithmus zur Fehlerisolation keine mehr ausschließen kann. Daher kann die folgende Fehleridentifikation nur für jede in Frage kommenden Ursache die Größe der möglichen Fehler bestimmen.

Definition 6.11: Fehlerisolation

Die Fehlerisolation ordnet einem detektierten Fehler einen physikalischen Parameter als Ursache zu.

6.1.3 Fehleridentifikation

Unter dem Begriff Fehleridentifikation soll im Rahmen dieser Arbeit die Bestimmung der Größenordnung der aufgetretenen Fehler verstanden werden. Nachdem im vorangegangenen Schritt schon eine Zuordnung des Fehlers zu einem bestimmten Parameter des Prozessmodells erfolgt ist, wird hier unter der Fehleridentifikation die Bestimmung der Abweichung des Parameters von seinem Sollwert verstanden. Damit können geeignete Gegenmaßnahmen eingeleitet werden, da der Parameter und dessen Abweichung vom Sollwert bekannt sind. Gegenmaßnahmen können beispielsweise eine autonome Störungsbeseitigung (Umschaltung auf ein redundantes System), eine Fehlermeldung an das Bedienpersonal (Fehler kann im Rahmen der üblichen Wartungsarbeiten behoben werden o.ä.) oder eine Notabschaltung sein.

Definition 6.12: Fehleridentifikation

Fehleridentifikation bedeutet die Feststellung der quantitativen Änderung des als Ursache für den detektierten Fehler isolierten Parameters mit dem Ziel, geeignete Gegenmaßnahmen ergreifen zu können.

Damit sind die wichtigsten Begriffe zur Diagnose dynamischer Systeme eingeführt, so dass im folgenden Abschnitt einzelne Diagnoseverfahren genauer betrachtet werden können.

6.2 Klassifikation der Diagnoseverfahren

Die Klassifizierung von Diagnoseverfahren kann nach [Lem95] anhand verschiedenster Kriterien vorgenommen werden. Mögliche Kriterien sind die Art der Prozessan-kopplung, die Verarbeitungsart von Signalen oder die Art der Repräsentation des verwendeten Prozesswissens. Tabelle 3.2 in [Lem95] liefert eine Klassifikation der Diagnoseverfahren nach Verfahrensprinzip, Verfahrensbasis und Beschreibungsbasis. Eine in der Literatur weit verbreitete Einteilung, die auch von [Fra94, Lem95, BKLS03, Gis03] verwendet wird, ist die Unterteilung in wissensbasierte, signalge-stützte und modellgestützte Verfahren.

6.2.1 Wissensbasierte Verfahren

Lässt sich als Folge der Komplexität des betrachteten Prozesses mit vertretbarem Aufwand kein hinreichend genaues quantitatives Modell des Prozesses erstellen, bieten sich wissensbasierte Verfahren zur Lösung der Diagnoseaufgabe an. Zum Einsatz kommen hier unter anderem Neuronale-Netze, Fuzzy-Logic oder Expertensysteme.

Weiterhin können auch qualitative Prozessmodelle, sogenannte Wissensbeobachter, verwendet werden.

6.2.2 Signalgestützte Verfahren

Zu den ältesten und am weitesten verbreiteten Diagnoseverfahren gehören die signalgestützten Verfahren. Ziel dieser Verfahren ist es, aus den gemessenen Signalen des Prozesses, die möglichst viel Information über die zu diagnostizierenden Fehler enthalten sollen, Merkmale zu extrahieren, die Rückschlüsse auf den aufgetretenen Fehler zulassen. Die Bestimmung von Grenzwerten, Trends, Varianzen und Korrelationen sowie die Spektralanalyse ergeben typische Merkmale.

Da bei den signalgestützten Verfahren kein Modell des realen Systems zum Einsatz kommt, hält sich die Leistungsfähigkeit dieser Verfahren in Grenzen, besonders wenn Fehler auftreten, die sich lediglich auf die Dynamik des Prozesses auswirken.

6.2.3 Modellgestützte Verfahren

Wenn es gelingt, das dynamische Verhalten eines Prozesses vollständig und fehlerfrei mit Hilfe eines Prozessmodells wiederzugeben und alle Größen, die zur Bestimmung aller möglichen Fehler notwendig sind, ohne Störungen zu messen, so ist man durch Vergleich der Messdaten mit den Modelldaten in der Lage, jeden Fehler im Prozess fast sofort zu detektieren, zu isolieren und zu identifizieren [Fra94].

Allerdings gelingt es praktisch nie, dieses Prozessmodell zu erstellen; darüber hinaus ist es in den seltensten Fällen möglich, alle interessierenden Größen zu messen. Weiterhin ist eine exakte, störungsfreie Messung, wie sie hier verlangt wird, in der Realität unmöglich. Dies hat zur Folge, dass die Abweichungen von Messdaten und Modelldaten zwar aufgenommen und verarbeitet werden können, allerdings ist zu beachten, dass sich die gemessenen Abweichungen nicht nur aus Abweichungen aufgrund von Fehlern, sondern darüber hinaus auch aus Modellierungsfehlern und Messfehlern zusammensetzen. Dennoch zählen die modellgestützten Verfahren zu den leistungsfähigsten Diagnosemethoden.

Die modellgestützten Verfahren lassen sich weiter in *beobachtergestützte Verfahren* und in *Parameterschätzverfahren* untergliedern. Der Unterschied zwischen den Verfahren liegt darin, dass bei den beobachtergestützten Verfahren die Differenz der Ausgangsgrößen zwischen Modell und Prozess zur Residuenbildung und somit zur weiteren Diagnose verwendet wird, wohingegen bei den Parameterschätzverfahren versucht wird, die Diagnose mit der Differenz zwischen den Referenzmodellparametern und geschätzten Modellparametern durchzuführen.

Zur zufriedenstellenden Lösung der Diagnoseaufgabe ist es applikationsbedingt oft notwendig, Kombinationen der obigen Verfahren anzuwenden.

Die Verfahren, die in dieser Arbeit vorgestellt bzw. entwickelt werden, sind in die Klasse der modellgestützten Diagnoseverfahren, genauer in die Klasse der beobachtgestützten Verfahren, einzugliedern.

Nachdem nun die Begriffe zur Diagnose dynamischer Systeme vorgestellt wurden und ein Überblick über die unterschiedlichen Verfahren gegeben wurde, wird im nächsten Abschnitt ein grundlegender Baustein für die in Kapitel 7 diskutierte Diagnose hybrider Systeme mittels diskreter Messungen vorgestellt.

6.3 Stochastische Zustandsrekonstruktion für hybride Systeme mit diskreten Messungen

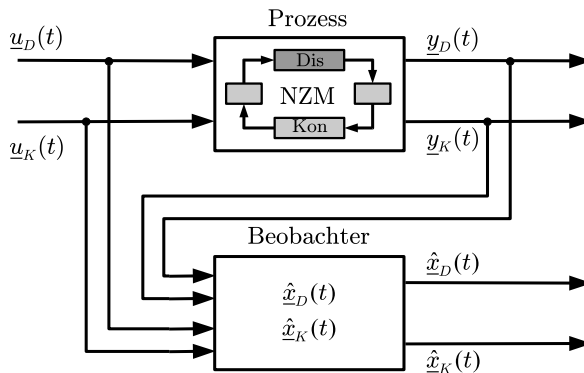


Abbildung 6.1: Allgemeiner hybrider deterministischer Beobachter

Sollen Steuergrößen ermittelt werden, die es erlauben, dass ein Prozess ein bestimmtes Steuer- bzw. Regelungsziel erreicht oder soll überprüft werden, ob ein Prozess korrekt arbeitet oder ob er ein Verhalten zeigt, das auf einen Fehler im Prozess schließen lässt, so sind durch Messung am Prozess gewonnene Informationen auszuwerten.

Sind nicht alle Zustandsgrößen, die zur Erfüllung der Automatisierungsaufgabe nötig sind, bekannt bzw. messbar, so wird häufig versucht, mittels vorhandener Messdaten und eines Prozessmodells die noch fehlenden Zustandsgrößen zu schätzen. Diese Aufgabe wird in der Regelungstechnik als Beobachtung bezeichnet [Sch01a]. Voraussetzung für die erfolgreiche Beobachtung ist die Prozesseigenschaft der *Beobachtbarkeit*.

Im Folgenden wird besonders auf die Beobachtung hybrider Systeme mit messbaren diskreten Zustandsgrößen eingegangen, bei denen die Beobachtungsaufgabe darin besteht, den kontinuierlichen Zustand des Systems zu schätzen. Dieser Systemklasse kommt durch die in der Regel preisgünstigere diskrete Sensorik eine besondere Bedeutung zu [MSK04,Sch01a]. Ein deterministischer Beobachter für diese Prozessklasse wurde bereits in [Sch01a] angegeben und diskutiert.

Zur Einordnung der hier vorliegenden Beobachtungsaufgabe ist in [Sch01a] eine Übersicht über die verschiedenen Konfigurationen der Beobachtungsaufgabe gegeben. Dort werden im wesentlichen drei Fälle und die entsprechenden Strukturen unterschieden:

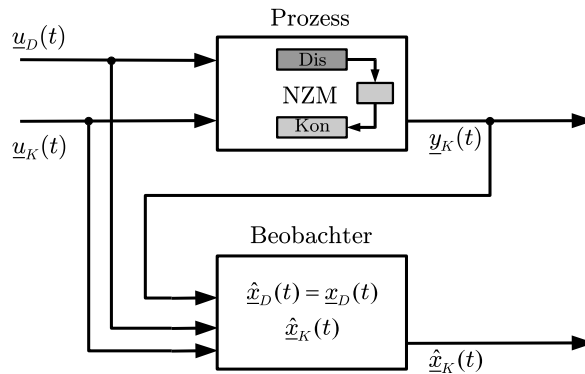


Abbildung 6.2: Beobachter für ein geschaltetes System

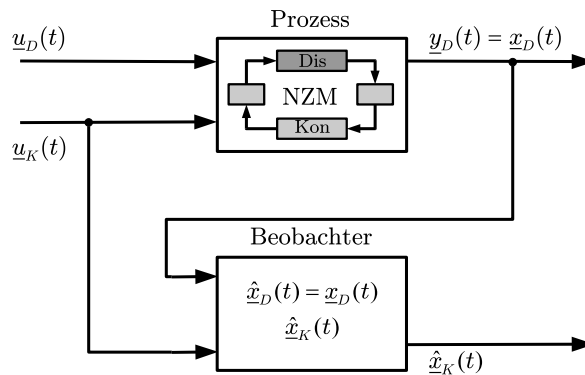


Abbildung 6.3: Hybrider deterministischer Beobachter

1. Zunächst kann der Fall der allgemeinen Beobachtung eines hybriden Systems bei gegebenen kontinuierlichen und diskreten Ausgangsgrößen und den entsprechenden Eingangsgrößen, wie es in Abbildung 6.1 gezeigt ist, betrachtet werden. Dabei werden sowohl der diskrete als auch der kontinuierliche Anteil des Zustandes geschätzt.
2. Als weiterer Fall wird die Beobachtung der kontinuierlichen Zustandsgrößen eines geschalteten Systems angegeben, wie es in Abbildung 6.2 gezeigt ist. Dabei darf der zugrunde liegende Prozess keine interne Wirkung der kontinuierlichen Zustandsgrößen auf den diskreten Systemanteil besitzen, wie im NZM von Abbildung 6.2 angedeutet ist. Bei einem geschalteten System wird zusätzlich davon ausgegangen, dass aufgrund der diskreten Eingangsgrößen die diskreten Zustandsgrößen bekannt sind und deshalb statisch aus den diskreten Eingangsgrößen berechnet werden können.
3. Als letzter Fall wird die Beobachtung der kontinuierlichen Zustandsgrößen betrachtet, wenn die diskreten Zustandsgrößen bekannt sind. Diese Konfiguration ist in Abbildung 6.3 für den deterministischen Fall gezeigt. Diese Beobachtungsaufgabe soll in der vorliegenden Arbeit genauer betrachtet und auf den stochastischen Fall erweitert werden.

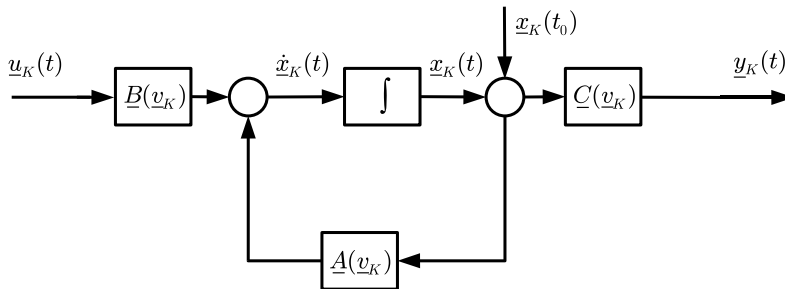


Abbildung 6.4: Lineares Erweitertes Zustandsraummodell

In [Sch01a] bildet ein deterministisches Netz-Zustands-Modell mit stückweise affiner Dynamik, wie es in Abschnitt 2.3.3 vorgestellt wurde, die Grundlage für den dort entwickelten deterministischen hybriden Beobachter. Im Unterschied zu diesem können bei dem hier vorgestellten stochastischen hybriden Beobachter die Einflüsse stochastischer Störgrößen berücksichtigt werden, so dass als Grundlage der Betrachtungen in dieser Arbeit das deterministische, lineare Erweiterte Zustandsraummodell (Abbildung 6.4) zusätzliche stochastische Eingangsgrößen erhält (Abbildung 6.5).

Damit sind alle Größen des Erweiterten Zustandsraummodells als Zufallsvariablen aufzufassen. Beschränkt man sich bei der Betrachtung der stochastischen Prozesse auf die Klasse der Gauß-Prozesse, so genügt zur Beschreibung der Zufallsvariablen

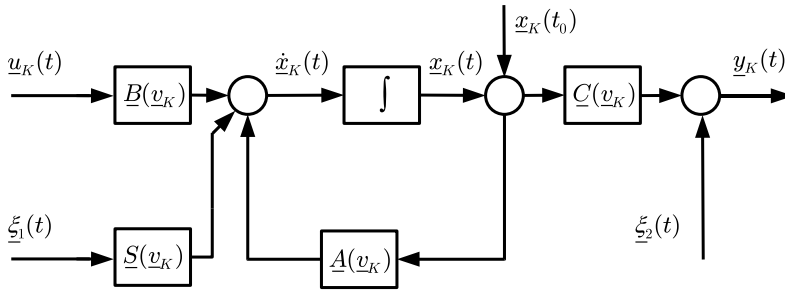


Abbildung 6.5: Lineares, stochastisch Erweitertes Zustandsraummodell

die Kenntnis des ersten und zweiten Momentes. Sind diese bekannt, so ist die Verteilung zu jedem Zeitpunkt eindeutig bestimmt [Lof90]. Um mit diesem Ansatz auch nichtlineare Prozesse beobachten zu können, ist es notwendig, das nichtlineare Verhalten mittels affiner Teilmodelle anzunähern. Auf diese Weise wird in Kapitel 8 der hier vorgestellte Ansatz am Zwei-Tank-System eingesetzt.

Die Aufgabe des stochastischen Beobachters ist demzufolge nicht die Schätzung deterministischer (Zustands-)Größen, sondern die Schätzung von Erwartungswerten und Kovarianzmatrizen. Diese stochastischen Größen beschreiben die Zufallsvariablen des beobachteten Systems.

Der hybride stochastische Beobachter wurde in [MK04, Gis03] eingeführt und beschrieben. An dieser Stelle soll er ausführlich diskutiert werden, da er eine wichtige Grundlage für die Diagnose hybrider Systeme mit diskreten Messungen darstellt. Die dazu verwendeten stochastischen Grundlagen werden in Abschnitt 6.3.1 in aller Kürze dargestellt, da die für stochastische lineare und affine Differentialgleichungen benötigten Zusammenhänge weitgehend als bekannt vorausgesetzt werden können. Für weitergehende Zusammenhänge und eine mathematisch fundierte Darstellung wird auf die Bücher [Arn73] und [Lof90] verwiesen. In Abschnitt 6.3.2 wird auf die Begriffe Beobachtbarkeit und Rekonstruierbarkeit eingegangen, bevor in den Abschnitten 6.3.3 und 6.3.4 die Arbeitsweise und Struktur des Beobachters beschrieben wird.

6.3.1 Stochastische Grundlagen

In Kapitel 2 wurde die deterministische Modellbildung von dynamischen Systemen erläutert. Die aus der deterministischen Betrachtung hervorgehenden zeitgetriebenen, ereignisdiskreten und hybriden Modelle sind leistungsfähige Beschreibungsformen, die allerdings den realen Prozess nicht immer zufriedenstellend abbilden. Dies

liegt daran, dass im realen Prozess das Systemverhalten zusätzlich durch Störungen beeinflusst wird. Deshalb ist es nicht immer möglich, aus bekannten Eingangsgrößen und einem gegebenen Anfangszustand eine deterministische Trajektorie des Systems zu berechnen. Zusätzlich können durch stochastische Einflüsse Ungenauigkeiten und Unsicherheiten bei der Modellierung abgebildet werden.

Falls unbekannte Einflüsse auf das zu betrachtende System wirken, so können die Zustandsverläufe eines realen Prozesses oft als *stochastischer Prozess* dargestellt werden. Es wird unterstellt, dass den Störungen, die die Zustands- und Ausgangsgrößen beeinflussen, gewisse stochastische Eigenschaften zugeordnet werden können [Lof90].

Im Anwendungsfall der hybriden Systeme stellt sich die Frage, welche Variablen durch unbekannte Störungen beeinflusst werden. In dieser Arbeit soll der Fall betrachtet werden, dass das ereignisdiskrete Teilsystem durch ein deterministisches Teilsystem beschrieben werden kann, während das kontinuierliche Teilsystem durch eine Modellform dargestellt wird, in der Störungen mitmodelliert werden.

Für das Netz-Zustands-Modell bedeutet dies, dass das Erweiterte Zustandsraummodell um einen stochastischen Anteil erweitert werden muss, wodurch die ursprünglich deterministischen Zustands- und Ausgangsgrößen zu stochastischen Größen werden. Für einen festen Zeitpunkt t_1 ergeben sich offensichtlich Zufallsvariablen bzw. Zufallsvektoren; diese werden – um sie von den deterministischen Größen abzuheben – durch Großbuchstaben gekennzeichnet. Damit eine Verwechslung eines Zufallsvektors mit einer Matrix ausgeschlossen ist, werden für Zufallsvariablen nur die Buchstaben X und Y verwendet.

Im linearen Fall erhält man somit folgendes Zustandsraummodell:

$$\frac{d\underline{X}}{dt} = \underline{A} \underline{X} + \underline{B} \underline{u} + \underline{S} \underline{\xi}_1, \quad (6.1)$$

$$\underline{Y} = \underline{C} \underline{X} + \underline{D} \underline{u} + \underline{\xi}_2. \quad (6.2)$$

Diese formale Erweiterung der Gleichungen um die Steuermatrix \underline{S} und einen Rauschprozess $\underline{\xi}_1$, der die Störungen modelliert, hat zur Folge, dass die Lösung der Differentialgleichung \underline{X} und damit auch \underline{Y} jeweils durch einen stochastischen Prozesse dargestellt werden muss.

Die Lösung der stochastischen Differentialgleichung ist, wie im deterministischen Fall, von einem Anfangszustand abhängig. Dieser Anfangszustand $\underline{X}(t_0)$ kann bei einer stochastischen Differentialgleichung sowohl ein deterministischer Vektor als auch ein Zufallsvektor mit einer zugehörigen Dichtefunktion sein.

Vergleicht man nun die ursprüngliche deterministische Gleichung mit der stochastischen Differentialgleichung, so muss aus Konsistenzgründen gelten, dass die Lösung \underline{x} der deterministischen Gleichung eine mögliche Realisierung des stochastischen Pro-

zesses \underline{X} sein muss. Der stochastische Prozess besteht anschaulich aus einer Schar von möglichen Trajektorien, weshalb der Zustandsvektor zu einem bestimmten Zeitpunkt nur mit stochastischen Mitteln, also beispielsweise mit Hilfe von Dichtefunktionen, beschrieben werden kann. Im Fall von linearen oder affinen stochastischen Differentialgleichungen kann bei gaußschen Rauschprozessen und einer gaußschen Anfangsverteilung die Dichtefunktion des Zufallsvektors \underline{X} analytisch angegeben werden. Für nichtlineare Differentialgleichungen oder nicht gaußsche Zufallsprozesse ist dies allgemein nicht möglich.

6.3.1.1 Lineare und affine stochastische Differentialgleichungen

Da in dieser Arbeit nur lineare und affine stochastische Differentialgleichungen von Bedeutung sind, soll die Diskussion auf deren spezielle Eigenschaften beschränkt bleiben.

Lineare stochastische Differentialgleichungen

Wie bereits erläutert wurde, entstehen die linearen stochastischen Differentialgleichungen, indem einer linearen deterministischen Zustandsdifferentialgleichung ein weißer, gaußverteilter Rauschprozess additiv hinzugefügt wird

$$d\underline{X}_t = \left(\underline{A} \underline{X}_t + \underline{B} \underline{u} + \underline{S} \underline{\xi}_1 \right) dt . \quad (6.3)$$

Hierbei sind:

- $\underline{\xi}_1(t)$ der Zufallsvariablenvektor des Rauschprozesses zum Zeitpunkt t ,
- \underline{A} die Dynamikmatrix,
- \underline{B} die Eingangsmatrix,
- $\underline{u}(t)$ der deterministische Eingangsvektor und
- \underline{S} die Steuermatrix.

In Abbildung 6.6 sind die Beziehungen noch einmal graphisch zusammengestellt.

Mit der formalen Gleichheit zwischen der Ableitung eines Wiener-Prozesses und dem weißen gaußschen Rauschen $d\underline{\beta}_t = \underline{\xi}_1(t)dt$ [Arn73] kann Gleichung (6.3) umgeschrieben werden zu

$$d\underline{X}_t = \underline{A} \underline{X}_t dt + \underline{B} \underline{u} dt + \underline{S} d\underline{\beta}_t .$$

Diese Differentialgleichung besitzt eine *fast sichere* und eindeutige Lösung [Arn73].

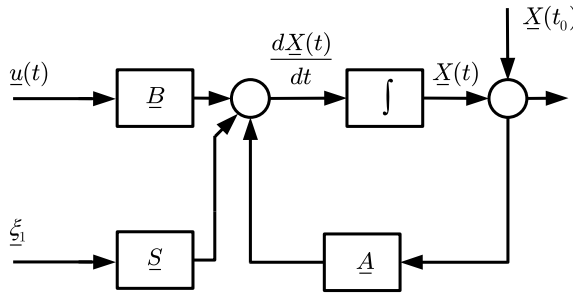


Abbildung 6.6: Darstellung der linearen stochastischen Differentialgleichung aus Gleichung (6.3)

Da die stochastischen Integrale die gleichen Linearitätseigenschaften wie gewöhnliche Riemann-Integrale besitzen, kann zur Bestimmung der Lösung dieser Differentialgleichung zunächst wie bei der deterministischen Zustandsdifferentialgleichung verfahren werden. Damit erhält man

$$\begin{aligned} \underline{X}_t &= \underline{\Phi}(t, t_0) \underline{X}(t_0) \\ &+ \int_{t_0}^t \underline{\Phi}(t, \tau) \underline{B} \underline{u}(\tau) d\tau + \int_{t_0}^t \underline{\Phi}(t, \tau) \underline{S} d\underline{\beta}_\tau \end{aligned} \quad (6.4)$$

mit der Transitionsmatrix $\underline{\Phi}(t, \tau) = e^{\underline{A}(t-\tau)}$.

Der Beweis, dass dieser formale Ansatz eine Lösung der stochastischen Differentialgleichung ist, kann beispielsweise in [Lof90, Arn73] nachgelesen werden.

Die Lösung \underline{X}_t der stochastischen Differentialgleichung enthält einen linear transformierten Anfangszufallsvektor, ein deterministisches Riemann-Integral und ein stochastisches Integral; diese drei Anteile sind additiv verknüpft.

Der *deterministische Anteil* der Lösung $\underline{\Phi}(t, t_0) \underline{X}(t_0) + \int_{t_0}^t \underline{\Phi}(t, \tau) \underline{B} \underline{u}(\tau) d\tau$ entspricht dem erwarteten Trend bzw. der Drift, ausgehend vom erwarteten Anfangsvektor $E\{\underline{X}(t_0)\}$. Die Drift wird beeinflusst vom deterministischen Eingangsvektor und repräsentiert im Falle der linearen stochastischen Differentialgleichungen die Veränderung des Erwartungswertes.

Der *stochastische Anteil* $\int_{t_0}^t \underline{\Phi}(t, \tau) \underline{S} d\underline{\beta}_\tau$ ist ein vektorieller stochastischer Prozess über dem Zeitintervall $[t_0, t]$. Der Wiener-Prozess bzw. die Brownsche Bewegung stellt die unbekanntenen Einflüsse auf das System wie z.B. Störungen, Unsicherheiten und Ungenauigkeiten bei der Modellierung des realen Prozesses dar. Da die Zufallsvariablen des Brownschen Prozesses und der Anfangszufallsvektor $\underline{X}(t_0)$ gaußverteilt und voneinander unabhängig sind, resultiert der Lösungsprozess \underline{X}_t als ein Gauß-Prozess. Die Angabe von Erwartungswert und Kovarianzmatrix genügt deshalb zur

Beschreibung des Prozesses $\underline{X}(t)$. Wegen der Unabhängigkeit der einzelnen Terme lautet der Erwartungswert der Lösung (6.4)

$$\begin{aligned} \underline{m}_{\underline{X}}(t) &= \text{E} \{ \underline{X}_t \} \\ &= \text{E} \{ \underline{\Phi}(t, t_0) \underline{X}(t_0) \} + \text{E} \left\{ \int_{t_0}^t \underline{\Phi}(t, \tau) \underline{B} \underline{u}(\tau) d\tau \right\} \\ &\quad + \text{E} \left\{ \int_{t_0}^t \underline{\Phi}(t, \tau) \underline{S} d\underline{\beta}_{-\tau} \right\}. \end{aligned}$$

Da der Erwartungswert des stochastischen Integrals verschwindet [Arn73], folgt:

$$\underline{m}_{\underline{X}}(t) = \underline{\Phi}(t, t_0) \cdot \text{E} \{ \underline{X}(t_0) \} + \int_{t_0}^t \underline{\Phi}(t, \tau) \underline{B} \underline{u}(\tau) d\tau.$$

Wie man sieht, stimmt der Erwartungswert der Lösung der stochastischen Differentialgleichung mit der Lösung der deterministischen Differentialgleichung überein. Deshalb ist die Differentialgleichung für den Erwartungswert

$$\dot{\underline{m}}_{\underline{X}}(t) = \underline{A} \underline{m}_{\underline{X}}(t) + \underline{B} \underline{u} \quad (6.5)$$

identisch mit der deterministischen Differentialgleichung.

Für die Kovarianzmatrix

$$\underline{V}_{\underline{X}}(t) = \text{E} \left\{ [\underline{X}(t) - \underline{m}_{\underline{X}}(t)] [\underline{X}(t) - \underline{m}_{\underline{X}}(t)]^T \right\}$$

zum Zeitpunkt t folgt die Gleichung

$$\begin{aligned} \underline{V}_{\underline{X}}(t) &= \text{Cov} \{ \underline{\Phi}(t, t_0) \underline{X}(t_0) \} + \text{Cov} \left\{ \int_{t_0}^t \underline{\Phi}(t, \tau) \underline{S} d\underline{\beta}_{-\tau} \right\} \\ &= \underline{\Phi}(t, t_0) \underline{V}_{\underline{X}}(t_0) \underline{\Phi}^T(t, t_0) + \int_{t_0}^t \underline{\Phi}(t, \tau) \underline{S} \underline{Q} \underline{S}^T \underline{\Phi}^T(t, \tau) d\tau, \end{aligned}$$

da der deterministische Ausdruck $\int_{t_0}^t \underline{\Phi}(t, \tau) \underline{B} \underline{u}(\tau) d\tau$ die Kovarianz $\underline{0}$ besitzt. Aus dieser Gleichung erhält man nach Differentiation und einigen Rechenschritten eine Matrizendifferentialgleichung für die Kovarianzmatrix:

$$\dot{\underline{V}}_{\underline{X}}(t) = \underline{A} \underline{V}_{\underline{X}}(t) + \underline{V}_{\underline{X}}(t) \underline{A}^T + \underline{S} \underline{Q} \underline{S}^T. \quad (6.6)$$

Die genaue Herleitung der Gleichung (6.6) kann in [Lof90, Arn73] nachgelesen werden.

Affine stochastische Differentialgleichungen

Wird zur rechten Seite einer linearen stochastischen Differentialgleichung ein konstanter Vektor \underline{c} dazu addiert, so spricht man von einer *affinen stochastischen Differentialgleichung*. Diese besitzt die Form

$$d \underline{X}_t = \underline{A} \underline{X}_t + \underline{B} \underline{u} + \underline{c} + \underline{S} d\beta_t.$$

Die affine stochastische Differentialgleichung kann als eine lineare stochastische Differentialgleichung interpretiert werden, wenn die Eingangsmatrix \underline{B} durch $\tilde{\underline{B}} = [\underline{B}, \underline{c}]$ und der Eingangsvektor $\underline{u}(t)$ durch $\tilde{\underline{u}}(t)$ mit $\tilde{\underline{u}}^T(t) = [\underline{u}^T(t), 1]$ ersetzt werden [Sch01a]. Diese lineare stochastische Differentialgleichung

$$d\underline{X}_t = \underline{A} \underline{X}_t + \tilde{\underline{B}} \tilde{\underline{u}} + \underline{S} d\beta$$

hat die gleiche Lösung und die gleichen Momente wie in Gleichung (6.4) beschrieben. Es ist zu berücksichtigen, dass der neue Eingangsvektor $\tilde{\underline{u}}$ eine nicht beeinflussbare Komponente besitzt.

Damit sind die wichtigsten stochastischen Grundlagen für den hybriden stochastischen Beobachter zusammengestellt. Dieser Beobachter beruht vor allem auf den affinen stochastischen Differentialgleichungen. Durch die affinen Differentialgleichungen steht eine Möglichkeit zur Verfügung, mit der man nichtlineare Systeme annähern kann. Gleichzeitig kann man die Vorteile der linearen bzw. affinen Differentialgleichungen nutzen, so dass lediglich Erwartungswert und Kovarianzmatrix zur vollständigen Beschreibung des Prozesses betrachtet und simuliert werden müssen.

6.3.2 Beobachtbarkeit und Rekonstruierbarkeit hybrider Systeme

Die Begriffe der Beobachtbarkeit und der Rekonstruierbarkeit hybrider Systeme sind in [Sch01a] für den deterministischen Fall angegeben. Unter der *Beobachtbarkeit* des Anfangszustandes versteht man die eindeutige Bestimmung des Anfangszustandes eines hybriden Systems bei vollständiger Kenntnis der Eingangs- und Ausgangsfunktion in endlicher Zeit. Dies bedeutet, dass eine Systemtrajektorie eindeutig vom Endzustand aus bis zu ihrem Anfangszustand bestimmbar ist, wenn die Eingangs- und Ausgangsfunktionen bekannt sind.

Definition 6.13: Beobachtbarer Anfangszustand eines hybriden Systems

Der Anfangszustand \underline{x}_0 eines diskret-kontinuierlichen Systems (hybriden Systems) heißt beobachtbar, wenn er anhand der Eingangsfunktionen und anhand der resultierenden Ausgangsfunktionen in endlicher Zeit eindeutig ermittelt werden kann.

Unter *Rekonstruierbarkeit* versteht man die eindeutige Bestimmung des Zustandes eines hybriden Systems bei vollständiger Kenntnis der Eingangs- und Ausgangsfunktionen in endlicher Zeit. Dies bedeutet, dass eine Systemtrajektorie eindeutig aus dem Anfangszustand und der Eingangs- und Ausgangsfunktion bis zum aktuellen Zustand bestimmbar ist.

Definition 6.14: Rekonstruierbarer Zustand eines hybriden Systems

Der Zustand $\underline{x}(t)$ eines diskret-kontinuierlichen Systems (hybriden Systems) heißt rekonstruierbar zum Zeitpunkt t , wenn er anhand der Eingangsfunktion und der aus ihr resultierenden Ausgangsfunktion ermittelt werden kann.

Bei linearen Systemen sind Beobachtbarkeit und Rekonstruierbarkeit äquivalent zueinander, da die Systemtrajektorien sowohl in der Richtung positiver Zeit als auch bei einer Umkehrung der Zeitrichtung eindeutig sind. Bei hybriden Systemen bleibt diese Eigenschaft erhalten, wenn man einige zusätzliche Forderungen aufstellt:

Nach [Sch01a] ist die Äquivalenz von Beobachtbarkeit und Rekonstruierbarkeit bei hybriden Systemen mit messbaren diskreten Zustandsgrößen gegeben, wenn alle Sprünge des kontinuierlichen Zustandes bei den diskreten Zustandswechseln, bezüglich Ausgangszustand und Sprungziel, eindeutig sind. Aus dem Zustand direkt vor dem Sprung muss das Sprungziel ermittelbar sein; umgekehrt muss aus dem Sprungziel der Zustand vor dem Sprung ermittelt werden können. Zusätzlich muss gefordert werden, dass die kontinuierlichen Trajektorien unter den einzelnen diskreten Zuständen eindeutig sind und auch bei der Umkehr der Zeitrichtung eindeutig bleiben, wie dies bei linearen Systemen der Fall ist.

Diese Forderungen werden sofort klar, wenn man sich vor Augen führt, dass bei einem deterministischen System der momentane Zustand immer eindeutig aus dem Anfangszustand und dem Verlauf der Eingangsgrößen hervorgeht. Findet nun im Falle messbarer diskreter Zustandsgrößen ein bekanntes Ereignis zu einem ebenfalls bekannten Zeitpunkt statt, so ist es bei eindeutigen kontinuierlichen Trajektorien möglich, den aktuellen Zustand des Systems zu bestimmen (\rightarrow Rekonstruierbarkeit).

Umgekehrt kann unter Umkehr der Zeit – ausgehend von einem kontinuierlichen Zustand – der Zustand unmittelbar nach dem letzten Ereignis bestimmt werden, wenn eindeutige kontinuierliche Trajektorien vorausgesetzt sind und der Ereigniszeitpunkt bekannt ist. Ist der Ausgangspunkt eines jeden möglichen Sprunges eindeutig aus dem Sprungziel zu bestimmen, so kann mittels der Sprungabbildung der Zustand unmittelbar vor dem letzten Ereignis bestimmt werden. Auf diese Weise lässt sich der Zustand des hybriden Systems beliebig weit zurückverfolgen (\rightarrow Beobachtbarkeit).

Für praktische Anwendungen – z.B. für die Regelung und auch für Diagnoseanwendungen – ist es häufig entscheidend, dass die Zustandsrekonstruktion, also die

Ermittlung des aktuellen Systemzustandes, vorgenommen werden kann, wohingegen die Ermittlung des Anfangszustandes eher theoretische Bedeutung besitzt. Der Begriff der Beobachtung wird jedoch häufig verwendet, wenn eigentlich die Zustandsrekonstruktion gemeint ist, da für viele praktisch relevante Systemklassen Beobachtbarkeit und Rekonstruierbarkeit äquivalent zueinander sind. So ist diese Äquivalenz beispielsweise für lineare Systeme gegeben; für den deterministischen Fall der stückweise affinen hybriden Systeme sind die Voraussetzungen für die Äquivalenz von Beobachtbarkeit und Rekonstruierbarkeit in [Sch01a] angegeben.

Die praktisch relevantere Aufgabe der Zustandsrekonstruktion wird in den nächsten Abschnitten für den Fall eines stochastisch beeinflussten kontinuierlichen Anteils erarbeitet.

6.3.3 Prinzip der stochastischen Zustandsrekonstruktion bei hybriden Systemen mit diskreten Messungen

Das Prinzip des hybriden stochastischen Beobachters beruht auf der Tatsache, dass aufgrund der diskreten Messungen auf mögliche kontinuierliche Zustände geschlossen werden kann. Mit jeder diskreten Messung ist es unter bestimmten Voraussetzungen möglich, die Menge der möglichen kontinuierlichen Zustände weiter zu verringern. Ist nach einer gewissen Anzahl von Messungen die Menge der möglichen Zustände klein genug, so kann diese Menge für die praktische Anwendung als Zustand des Systems interpretiert werden.

Um das Problem der stochastischen Zustandsrekonstruktion im mathematischen Sinne korrekt zu lösen, wären im Falle der hybriden Systeme auch bei linearen bzw. affinen Dynamikgleichungen komplizierte Verteilungsfunktionen zu betrachten. Dies hat folgende Gründe:

- Ist keine Information über die Anfangsverteilung gegeben, so wäre es am besten, eine Gleichverteilung über der initialen Neutralen Menge des diskreten Anfangszustandes anzunehmen.
- Zusätzlich müsste man die Begrenzungen der Neutralen Menge in den einzelnen diskreten Zuständen betrachten und beachten, dass der Zustand diese Grenzen nicht überschreitet, solange kein Ereignis festgestellt wurde.

Da unter den geschilderten Umständen die Lösung der stochastischen Differentialgleichung kein Gauß-Prozess mehr ist, werden zwei Vereinfachungen vorgenommen, so dass mit einfachen mathematischen Mitteln eine Zustandsrekonstruktion durchgeführt werden kann. Dazu bietet es sich an, eine anfängliche Gauß-Verteilung anzunehmen und weiterhin die Information über die Begrenzungen der aktuellen Neutralen Menge außer Acht zu lassen. Dann werden die Lösungen der einzelnen

linearen bzw. affinen Differentialgleichungen durch Gauß-Prozesse beschrieben, und damit reicht die Betrachtung von Erwartungswert und Kovarianzmatrix aus.

Zu Beginn der Rekonstruktionsprozedur ist die Menge der möglichen kontinuierlichen Zustände $\mathcal{X}_K(0)$ gerade die Neutrale Menge unter dem initialen diskreten Zustand $\underline{x}_D(0)$:

$$\mathcal{X}_K(0) = \Gamma(\underline{x}_D(0)). \quad (6.7)$$

Diese Menge der initial möglichen Zustände muss zunächst mit einer Anfangsverteilung angenähert werden, wozu ein Vertrauensniveau γ gewählt wird. Dieses Vertrauensniveau beeinflusst die Vertrauensmenge $\Xi_K(t)$, die später bei der Zustandsrekonstruktion berechnet wird. Je höher das Vertrauensniveau gewählt wird, umso sicherer gilt, dass der wirkliche Systemzustand innerhalb der Vertrauensmenge liegt. Die Vertrauensmenge wiederum ist eng mit der Dichte der Normalverteilung

$$f_{\underline{X}}(\underline{x}, t) = \frac{\sqrt{\det(\underline{V}_{\underline{X}}^{-1}(t))}}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2}((\underline{x}-\underline{m}_{\underline{X}})^T \underline{V}_{\underline{X}}^{-1}(t) (\underline{x}-\underline{m}_{\underline{X}}))} \quad (6.8)$$

verknüpft. Die Höhenlinien dieser Dichte, bzw. den Rand einer Vertrauensmenge, im folgenden Vertrauensgrenze genannt, erhält man daraus zu

$$q(\underline{x}) = (\underline{x} - \underline{m}_{\underline{X}})^T \underline{V}_{\underline{X}}^{-1}(t) (\underline{x} - \underline{m}_{\underline{X}}). \quad (6.9)$$

Diese Höhenlinien sind geometrisch gesehen n -dimensionale Hyperellipsoide.

Für die Kovarianzmatrix dieser Verteilung gilt stets:

$$\det(\underline{V}_{\underline{X}}) = \det \begin{pmatrix} \sigma_1^2 & r_{12}\sigma_1\sigma_2 & \dots & r_{1n}\sigma_1\sigma_n \\ r_{12}\sigma_1\sigma_2 & \sigma_2^2 & \dots & r_{2n}\sigma_2\sigma_n \\ \vdots & & \ddots & \vdots \\ r_{1n}\sigma_1\sigma_n & r_{2n}\sigma_1\sigma_n & \dots & \sigma_n^2 \end{pmatrix} \neq 0. \quad (6.10)$$

Der Korrelationskoeffizient $r_{ij} \in [-1, 1]$ ist der Quotient der Kovarianz σ_{ij} und des Produktes der Standardabweichungen $\sigma_1 \cdot \sigma_2 \cdot \dots \cdot \sigma_n$ und ist ein Maß für die Abhängigkeit der X_i untereinander. Die Wahrscheinlichkeit dafür, dass die Zufallsvariable innerhalb der Vertrauensgrenze (6.9) liegt, kann mittels Integration über die Dichtefunktion (6.8) innerhalb des durch die Vertrauensgrenze festgelegten Gebietes berechnet werden

$$P(\sqrt{q(\underline{X})} < R) = \int_{q(\underline{x}) < R^2} f_{\underline{X}} d\underline{x}$$

und ergibt das Vertrauensniveau γ .

Um nun eine Anfangsverteilung $\Xi_K(0)$ zu konstruieren, sind mehrere Bedingungen zu erfüllen:

- Die Wahrscheinlichkeit für die Vertrauensmenge entspricht dem gewählten Vertrauensniveau:

$$P\left(\sqrt{q(\underline{X})} < R\right) = \gamma.$$

- Die Neutrale Menge des diskreten Anfangszustandes ist vollständig überdeckt:

$$\Gamma(x_D(0)) \subseteq \Xi_K(0).$$

- Schließlich soll die Vertrauensmenge bezüglich ihres Hypervolumens so klein wie möglich sein.

Die Prozedur zur Ermittlung des Anfangszustandes wird an folgendem Beispiel veranschaulicht [Gis03]:

Beispiel 6.1: Ermittlung der initialen Vertrauensmenge

Am Beispiel von Abbildung 6.7 wird die Bestimmung der Dichte der Anfangszufallsvariablen gezeigt.

Gegeben ist dazu ein rechteckiger Ausschnitt eines zweidimensionalen euklidischen Zustandsraums. Dieser rechteckige Ausschnitt soll nun als Anfangsmenge mit den Seitenlängen a_i verstanden werden und durch eine Anfangsverteilung approximiert werden. Das vorgegebene Vertrauensniveau soll dabei $\gamma \approx 0,989$ sein; dieses Vertrauensniveau entspricht gerade der 3σ -Grenze [BH75], so dass R in diesem Beispiel zu 3 gewählt werden muss. Im zweidimensionalen Fall ergibt sich aus Glei-

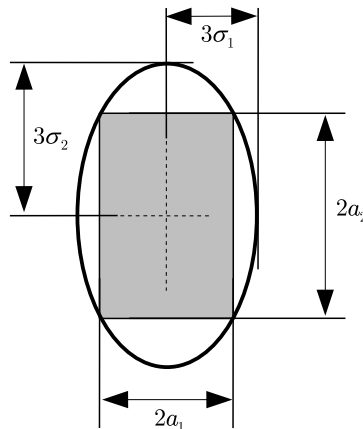


Abbildung 6.7: Approximation einer rechteckigen Anfangsmenge durch eine Gauß-Dichte

chung (6.9) die Ellipsengleichung in der x_1/x_2 Ebene:

$$\frac{(x_1 - m_1)^2}{\sigma_1^2} + \frac{2r_{12}(x_1 - m_1)(x_2 - m_2)}{\sigma_1\sigma_2} + \frac{(x_2 - m_2)^2}{\sigma_2^2} = R^2 = 3^2,$$

wobei die m_i die Koordinaten des Mittelpunktes der Ellipse sind.

Für die Bestimmung der Vertrauensmenge $\Xi_K(0)$ mit dem vorgegebenen Vertrauensniveau ist damit eine Bedingung festgelegt.

Eine zweite Bedingung erhält man aus der Tatsache, dass die Ellipse die Ecken des Rechtecks schneiden soll, da dann die Ellipse das Rechteck vollständig überdeckt. Um einfach rechnen zu können, betrachtet man den Mittelpunkt der Ellipse als Koordinatenursprung und erhält damit:

$$\frac{a_1^2}{\sigma_1^2} + \frac{2r_{12}a_1a_2}{\sigma_1\sigma_2} + \frac{a_2^2}{\sigma_2^2} = 9.$$

Da eine Drehung der Ellipse schon aus anschaulichen Gründen nicht zu einem minimalen Volumen bei gleichzeitiger Überdeckung führen kann, muss $r_{12} = 0$ gewählt werden.

Die Fläche der Ellipse ist nun proportional zum Produkt ihrer Halbmesser, so dass noch folgendes Optimierungsproblem zu lösen ist:

$$\min \sigma_1\sigma_2 \quad \text{s.t.} \quad \frac{a_1^2}{\sigma_1^2} + \frac{a_2^2}{\sigma_2^2} = 9.$$

Löst man dieses Optimierungsproblem, indem man die Lagrangefunktion

$$\sigma_1\sigma_2 + \lambda \left(\frac{a_1^2}{\sigma_1^2} + \frac{a_2^2}{\sigma_2^2} - 9 \right)$$

aufstellt, nach σ_1, σ_2 und λ ableitet und die resultierenden Gleichungen null setzt, so erhält man die Varianzen

$$\begin{aligned} \sigma_1^2 &= \frac{2a_1^2}{9} \quad \text{und} \\ \sigma_2^2 &= \frac{2a_2^2}{9}. \end{aligned}$$

Solange kein Ereignis auftritt, kann die aktuelle Vertrauensmenge für den kontinuierlichen Zustand $\Xi_K(t)$ bestimmt werden, indem der Erwartungswert und die Kovarianzmatrix der Anfangsverteilung mit den Gleichungen (6.5) und (6.6) fortentwickelt werden. Anschließend wird ein Hyperellipsoid mit dem Vertrauensniveau γ um den Erwartungswert $\underline{m}_X(t)$ berechnet. Der Rand des Hyperellipsoids ergibt sich aus der

Gleichung

$$R^2 = (\underline{x} - \underline{m}_{\underline{X}})^T V_{\underline{X}}^{-1}(t) (\underline{x} - \underline{m}_{\underline{X}}).$$

Tritt zum Zeitpunkt t_1 ein internes Ereignis auf, so wird die Vertrauensmenge $\Xi'_K(t_1^+)$ als Hilfsgröße eingeführt, die als Schnittmenge der alten Vertrauensmenge $\Xi_K(t_1)$ und der Triggerrandmenge $\partial\Omega(t_1)$ berechnet wird

$$\Xi'_K(t_1^+) = \Xi_K(t_1) \cap \partial\Omega(t_1); \quad (6.11a)$$

dabei ist $\partial\Omega(t_1)$ der Rand der Triggermenge, die das interne Ereignis verursacht hat. Tritt ein externes Ereignis auf, so kann dieses Ereignis jederzeit eintreten, d.h. das Erreichen einer Triggermenge ist dazu nicht notwendig. Daher ist eine andere Hilfsmenge zu berechnen: Die Elemente dieser Hilfsmenge müssen sowohl Elemente der Neutralen Menge vor dem externen Ereignis als auch der Vertrauensmenge vor dem Ereignis sein:

$$\Xi'_K(t_1^+) = \Xi_K(t_1) \cap \Gamma(x_D(t_1)). \quad (6.11b)$$

Die Hilfsmengen können theoretisch leer sein. Allerdings lassen leere Hilfsmengen den Schluss zu, dass das Prozessverhalten und das Verhalten des Modells stark voneinander abweichen und deshalb ein Fehler vorliegen muss. Hier wird zunächst der fehlerfreie Fall betrachtet, während der Fehlerfall in Kapitel 7 ausführlich behandelt wird.

Wenn ein Ereignis ohne Sprung auftritt, wie es in Abbildung 6.8 gezeigt ist, so erhält man die neue Vertrauensmenge $\Xi_K(t_1^+)$, indem man die als Hilfsgröße berechnete Menge $\Xi'_K(t_1^+)$ mit der neuen Neutralen Menge $\Gamma(x_D(t_1^+))$ schneidet und diese Schnittmenge $\Xi''_K(t_1^+)$ durch die neue Vertrauensmenge genau so approximiert, wie es für die Anfangsverteilung in Beispiel 6.1 beschrieben wurde. Die Menge $\Xi''_K(t_1^+)$ wird also durch ein Hyperellipsoid eingeschlossen, das das Vertrauensniveau γ besitzt. Dadurch erhält man die neue Vertrauensmenge $\Xi_K(t_1^+)$ und die neue Verteilungsdichte $f_{\underline{X}}(\underline{x}, t_1^+)$.

Bei einem Ereignis, das mit einem Zustandssprung verbunden ist, ist das Vorgehen prinzipiell gleich; zu beachten ist, dass die Hilfsmenge $\Xi'_K(t_1^+)$ aber zunächst durch die Sprunggleichung abgebildet werden muss. Das Bild dieser Menge, geschnitten mit der neuen Neutralen Menge $\Gamma(x_D(t_1^+))$, ergibt dann die Menge $\Xi''_K(t_1^+)$, die wieder durch eine neue Vertrauensmenge approximiert wird.

Mit der neuen Vertrauensmenge $\Xi_K(t_1^+)$ kann dann die Vertrauensmenge $\Xi_K(t)$ für $t > t_1$ berechnet werden, indem $\Xi_K(t_1^+)$ nach den Gleichungen (6.5) und (6.6) fortentwickelt wird bis das nächste Ereignis auftritt.

Diese beiden Schritte, nämlich die Fortentwicklung der Vertrauensmenge und der Schnitt der Vertrauensmenge mit einer Triggerrandmenge, werden fortgesetzt bis

die Kovarianzmatrix \underline{V}_X bestimmte Bedingungen erfüllt. Eine mögliche Bedingung ist, dass ihre Elemente klein genug sind. Dann können die Erwartungswerte $E\{\underline{X}_K\}$ als geschätzter Zustand betrachtet werden.

Anschaulich betrachtet kann man sich die Fortentwicklung zwischen zwei Ereigniszeitpunkten im 2-dimensionalen folgendermaßen vorstellen: Wird die initiale Zufallsvariable anhand ihrer Dichte dargestellt, so ergibt sich eine Gauß-Glocke. Das Maximum der Dichte, das gleich dem Erwartungswert ist, bewegt sich bei der Fortentwicklung nun entsprechend der Differentialgleichung (6.5) über die Zustandsebene. Die Ausdehnung der Glocke sowie ihre Ausrichtung werden über die Kovarianzmatrix beeinflusst.

Noch anschaulicher wird es, wenn man die Vertrauensmenge betrachtet. Im 2-dimensionalen Fall wird der Rand der Vertrauensmenge von einer Ellipse gebildet. Die Ellipse erhält man durch den Schnitt der 2-dimensionalen Gauß-Glocke mit einer Ebene in Höhe des gewünschten Vertrauensniveaus.

In Abbildung 6.8 wird die Situation gezeigt, in der ein neuer diskreter Zustand gemessen wird. Dabei wird die Vertrauensmenge $\Xi_K(t_k^+)$ entsprechend der Messinformation verkleinert. In praktischen Anwendungen bietet es sich an, die Varianz dabei nicht auf 0 zu verkleinern, sondern auf einen sehr kleinen Wert zu setzen, so dass die Kovarianzmatrix numerisch invertierbar bleibt. Dies bietet den Vorteil, dass Gleichung (6.9) immer ausgewertet werden kann.

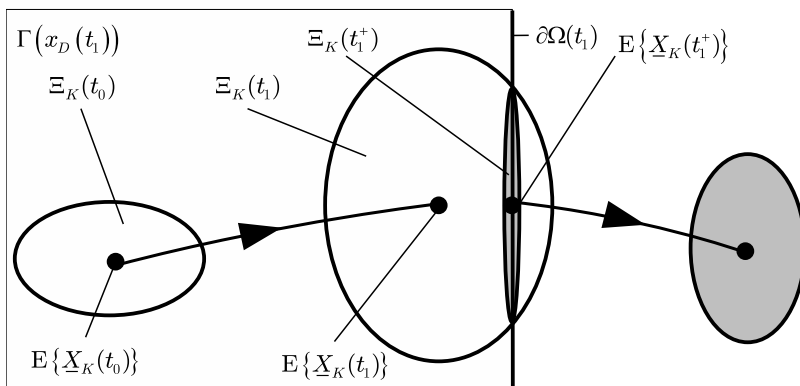


Abbildung 6.8: Verkleinerung der Vertrauensmenge beim Auftreten eines internen Ereignisses

6.3.4 Beobachterstruktur

Der hybride Beobachter für hybride Systeme mit messbaren diskreten Eingangsgrößen hat nach [Sch01a] die Eingangsgrößen $\underline{x}_D(t)$ und $\underline{u}_K(t)$ und liefert einen Schätzwert $\hat{\underline{x}}_K$ für den kontinuierlichen Zustand. Die Struktur dieses Beobachters zeigt Abbildung 6.3.

Erweitert man diesen Beobachter dahingehend, dass die Zustandstrajektorien als stochastische Prozesse aufgefasst werden, gibt der Beobachter eine vektorielle Zufallsvariable aus, die durch ihre Momente $E\{\underline{X}_K(t)\}$ und $\underline{V}_{\underline{X}_K}(t)$ charakterisiert werden kann. Dieser hybride stochastische Beobachter hat die in Abbildung 6.9 gezeigte Struktur. Im hier betrachteten Fall von stückweise affinen Systemen und unabhängiger gaußscher Anfangsverteilung können die Zufallsvariablen durch Angabe des Erwartungswerts und der Kovarianzmatrix eindeutig beschrieben werden.

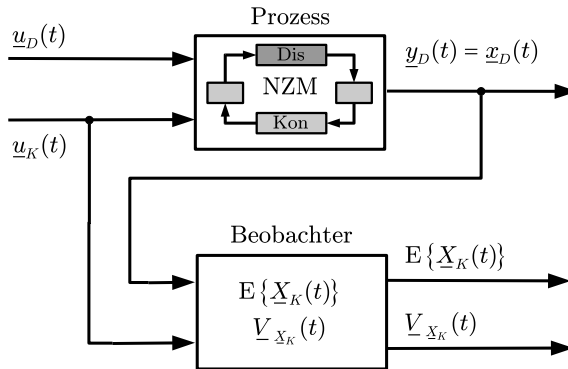


Abbildung 6.9: Hybrider stochastischer Beobachter

Der hybride stochastische Beobachter setzt sich aus zwei Komponenten zusammen:

- Die *Initialisierungskomponente* hat die Aufgabe festzustellen, ob ein internes oder externes Ereignis eingetreten ist. Da ein Ereignis entweder zusätzliche Information über den Zustand bereitstellt oder den Zustand verändert, muss durch die Initialisierungskomponente beim Eintreten von Ereignissen ein neuer Anfangszufallsvektor bereitgestellt werden, der wiederum durch den Erwartungswert und durch die Kovarianzmatrix beschrieben wird.
- Die *Evolutionskomponente* entwickelt den Zustandszufallsvektor, der durch die Initialisierungskomponente bereitgestellt wurde, weiter bis das nächste Ereignis auftritt. Dazu werden Erwartungswert und Kovarianzmatrix entsprechend der kontinuierlichen Eingangsgrößen $\underline{u}_K(t)$, der aktuellen Dynamik und der aktuellen Diffusion fortentwickelt.

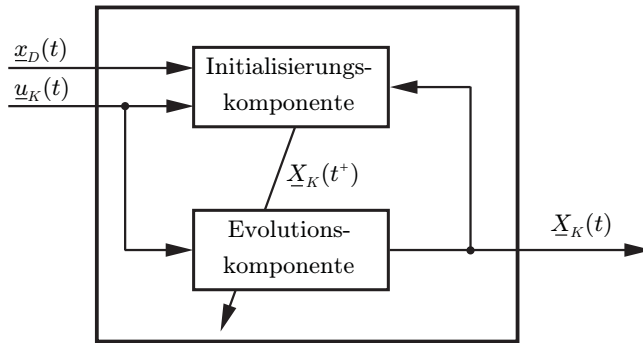


Abbildung 6.10: Struktur des hybriden stochastischen Beobachters

6.3.5 Systemanforderungen für den Einsatz des hybriden stochastischen Beobachters

Nachdem nun anschaulich klar ist, wie der hybride stochastische Beobachter arbeitet, sollen an dieser Stelle die Voraussetzungen für seine Anwendung zusammengestellt werden:

- Der Prozess ist durch ein Netz-Zustands-Modell dargestellt, bei dem das erweiterte Zustandsraummodell durch affine Differentialgleichungen beschrieben wird.
- Die Triggerrandmengen werden durch Hyperebenen beschrieben.
- Zu jeder Triggerrandmenge gibt es einen diskreten Sensor, der die Information bereitstellt, auf welcher Seite der Triggerrandmenge sich der Systemzustand befindet.

Erfüllt ein System diese Anforderungen, so kann ein hybrider stochastischer Beobachter für dieses System entworfen werden. Dass durch diesen Beobachter der Systemzustand auch tatsächlich rekonstruiert werden kann, ist damit allerdings noch nicht sichergestellt. Dazu sind weitere Überlegungen notwendig.

6.3.6 Informationsgewinn durch den hybriden stochastischen Beobachter

Anschaulich wird die Vertrauensmenge des hybriden stochastischen Beobachters bei ausreichend vielen Messereignissen immer wieder durch die jeweilige Messhyperebene geschnitten und dabei orthogonal zur Messhyperebene verkleinert. Damit ist klar,

dass man mindestens n Messereignisse benötigt, um gegenüber der Anfangsverteilung in jeder Dimension n einen Informationsgewinn zu erzielen.

Betrachtet man eine zusätzliche Messung innerhalb des Rekonstruktionsprozesses, so ist klar, dass durch die zusätzliche Messung nur dann neue Information gewonnen wird, wenn der Schnitt von Messhyperebene und Vertrauensmenge die Vertrauensmenge in eine Richtung verkleinert, die zuvor noch nicht verkleinert wurde.

Für die folgende Plausibilitätsbetrachtung werden zwei aufeinander folgende Messungen an zwei verschiedenen diskreten Sensoren betrachtet. Zusätzlich wird vorausgesetzt, dass in dem betrachteten System Beobachtbarkeit und Rekonstruierbarkeit äquivalent zueinander sind [Sch01a]. Die Systemdynamik wird in den einzelnen Gebieten als lineare Systemdynamik angenommen und somit gilt

$$\dot{\underline{x}}_K = \underline{A}(\underline{x}_D) \underline{x}_K + \underline{B}(\underline{x}_D) \underline{u}_K.$$

Verwendet man die Tatsache, dass \underline{A}^0 die Einheitsmatrix ergibt, so kann die Lösung dieser Zustandsdifferentialgleichung abschnittsweise mit Hilfe der Matrix-Exponentialfunktion

$$e^{\underline{A}(\underline{x}_D)t} := \sum_{i=0}^{\infty} \frac{\underline{A}^i(\underline{x}_D)t^i}{i!} = \underline{\Phi}_{\underline{x}_D}(t)$$

in der Form

$$\underline{x}_K(t) = \int_{t_0}^t \underline{\Phi}_{\underline{x}_D}(\tau) \underline{B}(\underline{x}_D) \underline{u}_K(\tau) d\tau + \underline{\Phi}_{\underline{x}_D}(t - t_0) \underline{x}_K(t_0)$$

angegeben werden [Föl94], solange sich der diskrete Zustand nicht ändert. Dabei ist $\underline{x}_K(t_0)$ der Anfangszustand der Differentialgleichung.

Stehen nun lediglich diskrete Messungen zur Verfügung, mit denen festgestellt werden kann, ob der Zustand zu einem bestimmten Zeitpunkt Element einer Hyperebene im Zustandsraum ist, so kann dieses Messereignis zum Zeitpunkt t_1 mit Hilfe der Gleichung

$$\underline{c}_1^T \underline{x}_K(t_1) + \varphi_1 = 0 \tag{6.12}$$

beschrieben werden. Bei dieser ersten Messung wechselt der diskrete Zustand des Systems von \underline{x}_{D1} nach \underline{x}_{D2} . Das nächste Messereignis zum Zeitpunkt t_2 führt wiederum auf eine Messgleichung

$$\underline{c}_2^T \underline{x}_K(t_2) + \varphi_2 = 0, \tag{6.13}$$

wobei der diskrete Zustand von \underline{x}_{D2} nach \underline{x}_{D3} wechselt. Da die Messereignisse aufgrund der Voraussetzungen immer mit den diskreten Zustandswechseln zusammen fallen, gilt zwischen zwei Messereignissen immer eine lineare Dynamik. Aufgrund

dieser bekannten Dynamik des Systems kann man nun $\underline{x}_K(t_2)$ mit Hilfe der Lösung der Zustandsdifferentialgleichung und der bekannten Systemdynamik ausdrücken:

$$\underline{x}_K(t_2) = \int_{t_1}^{t_2} \underline{\Phi}_{x_{D2}}(\tau) \underline{B}(x_{D2}) \underline{u}_K(\tau) d\tau + \underline{\Phi}_{x_{D2}}(t_2 - t_1) \underline{x}_K(t_1).$$

Eingesetzt in Gleichung (6.13) ergibt sich

$$\underline{c}_2^T \int_{t_1}^{t_2} \underline{\Phi}_{x_{D2}}(\tau) \underline{B}(x_{D2}) \underline{u}_K(\tau) d\tau + \underline{c}_2^T \underline{\Phi}_{x_{D2}}(t_2 - t_1) \underline{x}_K(t_1) + \varphi_2 = 0,$$

wobei nun der Anteil, der durch die Steuerung verursacht wurde, mit dem affinen Anteil zu $\tilde{\varphi}_2$ zusammengefasst wird. Man erhält also

$$\underline{c}_2^T \underline{\Phi}_{x_{D2}}(t_2 - t_1) \underline{x}_K(t_1) + \tilde{\varphi}_2 = 0.$$

Durch diese Betrachtung wird offensichtlich, dass ein Informationsgewinn bei der zusätzlichen Messung nur dann vorliegt, wenn der Spaltenvektor $\underline{c}_2^T \underline{\Phi}_{x_{D2}}(t_2 - t_1)$ linear unabhängig von \underline{c}_1^T ist, da das Gleichungssystem aus den Gleichungen (6.12) und (6.13) nur dann einen Freiheitsgrad verliert. Bei zusätzlichen Messungen müssen also immer neue linear unabhängige Zeilen hinzukommen, so dass das Gleichungssystem nach frühestens n Messungen eindeutig nach dem Zustand $\underline{x}_K(t_0)$ aufgelöst werden kann.

Es gilt also zu überprüfen, ob die Matrix

$$\underline{Q}_B = \begin{pmatrix} \underline{c}_1^T \\ \underline{c}_2^T \underline{\Phi}_{x_{D2}}(t_2 - t_1) \\ \underline{c}_3^T \underline{\Phi}_{x_{D3}}(t_3 - t_2) \underline{\Phi}_{x_{D2}}(t_2 - t_1) \\ \vdots \\ \underline{c}_k^T \underline{\Phi}_{x_{Dk}}(t_k - t_{k-1}) \cdot \dots \cdot \underline{\Phi}_{x_{D2}}(t_2 - t_1) \end{pmatrix} \quad (6.14)$$

nach k Messungen den Rang n aufweist. Ist der Rang dieser Matrix n , dann können n linear unabhängige Zeilen ausgewählt und der Systemzustand zum Zeitpunkt t_1 eindeutig berechnet werden.

Ob nun durch verschiedene Messungen überhaupt n linear unabhängige Zeilen gewonnen werden können und somit der Systemzustand beobachtet bzw. rekonstruiert werden kann, hängt – wie man leicht sehen kann – zum einen von den verschiedenen Messhyperebenen ab, die die Zeilenvektoren \underline{c}_i^T festlegen, zum anderen hängt dies von der Systemdynamik ab, die in den einzelnen Teilgebieten die Transitionsmatrizen $\underline{\Phi}_{x_{D_i}}(\cdot)$ bestimmt. Zusätzlich werden die zu betrachtenden Transitionsmatrizen von der Dauer zwischen den Messereignissen beeinflusst.

Aufgrund der Fülle der Abhängigkeiten ist es praktisch unmöglich, ein Kriterium anzugeben, mit dem die Analyse eines hybriden Systems mit stückweise affiner Dynamik auf Beobachtbarkeit möglich ist.

Aus den vorhergehenden Überlegungen wird klar, dass immer wieder Messereignisse auftreten müssen, damit ein hybrides System mit diskreten Messungen beobachtet werden kann. Deshalb wird die Beobachtbarkeit dieser Systemklasse in [Sch01a] sinnvoller Weise so definiert, dass der Anfangszustand nach endlich vielen Messereignissen – und nicht wie bei kontinuierlichen Messungen in endlicher Zeit – ermittelbar sein muss.

Weiterhin sind in [Sch01a] einige notwendige und hinreichende Bedingungen angegeben, die sowohl für einzelne diskrete Trajektorien, als auch für das ganze System die Beobachtbarkeit dieser Systemklasse analysierbar machen sollen. Viele dieser Kriterien gehen jedoch von falschen Voraussetzungen aus und sind somit unbrauchbar. Der Fehler liegt darin begründet, dass eine von (6.14) abweichende Beobachtbarkeitsmatrix \tilde{Q}_B für jeden diskreten Zustandsübergang definiert wird, mit der untersucht wird, ob an einem diskreten Zustandsübergang, der aus dem Zustand \underline{x}_D über eine Messhyperebene mit Normalenvektor \underline{c} stattfindet, zusätzliche Informationen gewonnen werden können. Diese Matrix ist mit Hilfe der Dynamikmatrix $\underline{A}(\underline{x}_D)$ und dem Normalenvektor \underline{c} der Messhyperebene zu

$$\tilde{Q}_B = \begin{pmatrix} \underline{c}^T \\ \underline{c}^T \underline{A}(\underline{x}_D) \\ \vdots \\ \underline{c}^T \underline{A}^{n-1}(\underline{x}_D) \end{pmatrix}$$

definiert. Diese Definition einer Beobachtbarkeitsmatrix ist allerdings nur dann sinnvoll, wenn kontinuierliche Messungen vorliegen. Mit dieser Matrix kann nämlich nicht entschieden werden, ob eine einzige Messung einen Informationsgewinn bringt; man kann damit lediglich entscheiden, ob man mit der über einem Zeitintervall kontinuierlich gemessenen Ausgangsfunktion $\underline{y}(t) = \underline{c}^T \underline{x}_K(t)$ den Zustand bestimmen kann. Die Herleitungen des Kalmanschen Beobachtbarkeitskriteriums in [Föl94] und [Lun97] zeigen dies.

Daher muss man bei den hier angenommenen diskreten Messungen die Transitionsmatrizen zusammen mit den Normalenvektoren der Messhyperebenen betrachten, wie dies in (6.14) gezeigt ist.

Aus diesen Überlegungen wird deutlich, dass der Anwender des hybriden stochastischen Beobachters für zwei Dinge sorgen muss, damit der Beobachter erfolgreich angewendet werden kann:

- Es müssen immer wieder Messereignisse am Prozess stattfinden.

- Eine geeignete Konfiguration von diskreten Sensoren muss am Prozess vorhanden sein.

Zur Validierung eines Beobachters kann der Anwender mit Hilfe von Simulationen überprüfen, ob der Prozess bei den einzelnen diskreten Zustandsübergängen auch immer wieder die benötigten Informationen bereitstellt. Ist dies nicht der Fall, so muss der Anwender durch weitere diskrete Sensoren dafür sorgen, dass die Beobachtbarkeitmatrix (6.14) vollen Rang erhält.

Zusammenfassung zum hybriden stochastischen Beobachter

Zur Zustandsschätzung bei einem hybriden System mit stückweiser affiner Dynamik, bei dem der diskrete Zustand messbar ist, kann ein hybrider stochastischer Beobachter (HSB) zur Bestimmung des kontinuierlichen Zustandes eingesetzt werden. Dieser Beobachter ermittelt einen vektoriellen Erwartungswert und eine Kovarianzmatrix für die nicht messbaren Zustandsgrößen. Sobald die Elemente der Kovarianzmatrix für die Anwendung klein genug sind, kann der Erwartungswert als geschätzter Zustand verwendet werden.

Als Eingänge dienen dem hybriden stochastischen Beobachter der diskrete Zustandsvektor und der kontinuierliche Eingangsvektor. Jede messbare Änderung des diskreten Zustandes wird als Ereignis (vgl. Abbildung 2.2) bezeichnet.

Jedes interne oder externe Ereignis hat eine Änderung des diskreten Zustandes \underline{x}_D des Beobachters zur Folge. Bei einem internen Ereignis wird das Ereignis dadurch ausgelöst, dass der kontinuierliche Zustand den Rand einer aktiven Triggermenge überschreitet. Dadurch wird diskrete Messinformation gewonnen, so dass ein neuer Anfangszufallsvektor bestimmt werden muss. Bei externen Ereignissen wird keine zusätzliche Information gewonnen, jedoch muss der aktuelle Zufallsvektor dem Ereignis entsprechend abgebildet werden. In der Zeit zwischen zwei Ereignissen wird der nach dem letzten Ereignis gültige Anfangszufallsvektor fortentwickelt, wie in Abschnitt 6.3.3 beschrieben.

Der hybride stochastische Beobachter bietet die Möglichkeit, die Zustandsgrößen von hybriden, stückweise affinen Systemen zu beobachten. Nimmt man geringfügige Abweichungen in Kauf, können auch nichtlineare, hybride Systeme nach entsprechender stückweiser affiner Approximation beobachtet werden. Da der hybride stochastische Beobachter die beobachteten kontinuierlichen Zustandsgrößen in Form eines Erwartungswertes und einer Kovarianzmatrix ausgibt, liefert er zusätzliche Information, die es erleichtert, die immer zwischen realem System und Beobachter auftretenden Abweichungen zu beurteilen. Diese Möglichkeit prädestiniert den hybriden stochastischen Beobachter für den Einsatz in der Diagnose hybrider Systeme, wie sie im nächsten Kapitel genauer ausgeführt wird.

Kapitel 7

Verfahren zur Diagnose hybrider Systeme

Die bisher in der Literatur bekannten Verfahren zur Diagnose dynamischer Systeme gehen immer davon aus, dass eine Vielzahl von Messgrößen zur Verfügung steht und diese Messgrößen quasikontinuierlich erfasst werden können. Sensoren für solche Messungen sind in der Regel teuer und stellen selbst ein gewisses Ausfallpotenzial dar, so dass auch die zur Diagnose eingesetzten Sensoren überwacht werden müssen.

Eine Alternative zu teureren kontinuierlichen Sensoren stellen diskrete Sensoren dar, die meist preisgünstiger sind, da sie nur wenige diskrete Messwerte liefern. So liefert beispielsweise ein diskreter Füllstandssensor lediglich die Information, ob der Füllstand größer oder kleiner als seine Einbauhöhe ist. Da mit dieser Art von Sensoren nur wenige Informationen über den Prozess zur Verfügung stehen, muss das dabei eingesetzte Diagnoseverfahren besonders leistungsfähig sein. Es ist anschaulich klar, dass mit dem ausschließlichen Einsatz diskreter Sensoren die Qualität der Diagnoseergebnisse abnimmt. Trotz des reduzierten Informationsgehalts sind solche Ansätze interessant, da die wirtschaftlichen Aspekte auch bei der Diagnose eine entscheidende Rolle spielen.

Um nun das in dieser Arbeit entwickelte Verfahren zur Diagnose hybrider Systeme aufgrund diskreter Messinformationen besser einordnen zu können, wird zunächst in Abschnitt 7.1 das Prinzip eines bekannten Verfahrens beschrieben und anschließend seine Anwendbarkeit auf die hier behandelte Problemstellung diskutiert. Da in der Literatur keine weiteren Verfahren zu dieser Problemstellung gefunden werden konnten, wird in Abschnitt 7.2 ein neues Verfahren vorgestellt, das speziell für den Anwendungsfall hybrider Systeme mit diskreten Sensorinformationen entwickelt wurde.

7.1 Diagnose hybrider Systeme mit kontinuierlicher Sensorik

Zur Diagnose hybrider dynamischer Systeme mit kontinuierlicher Sensorik wurde in [Mos97] ein Verfahren vorgestellt, das in der Lage ist, eine große Klasse dieser Systeme zu diagnostizieren. Grundlage dieses Verfahrens sind Bonddiagramme [Tho74], die um Elemente zur Darstellung des hybriden Systemverhaltens erweitert wurden [MB96, MB95, NBK⁺00]. Nachdem das zu diagnostizierende System als hybrides Bonddiagramm modelliert wurde, wird dieses Diagramm in einen Temporal-Causal-Graph (TCG) transformiert [MKB95, Mos97]; dieser Graph ordnet den Fehlerursachen zukünftige Auswirkungen in Form von Abweichungen in den Messgrößen und in den ersten und zweiten Ableitungen der Messgrößen zu.

Der Teilschritt der *Fehlerisolation* kann somit auf anschauliche Art und Weise bewältigt werden. Zunächst gilt es, nachdem ein Fehler detektiert wurde, durch Rückwärtssuche im TCG alle möglichen Ursachen für diesen Fehler zu finden. Sind alle möglichen Ursachen bestimmt, so wird die Anzahl dieser Kandidaten dadurch reduziert, dass die zukünftigen Auswirkungen jeder möglichen Ursache qualitativ bestimmt werden. Kommen neue Messwerte hinzu, so können mögliche Ursachen, die in ihren prädierten Auswirkungen nicht mit dem gemessenen Prozessverhalten übereinstimmen, nach und nach ausgeschlossen werden.

Der Vergleich von Prädiktion und Messung geschieht, indem zunächst die Abweichung der Messgrößen und die Abweichungen der ersten und zweiten Ableitungen der Messgrößen zu einer *Signatur* zusammengefasst werden. Anschließend müssen die für die verschiedenen Ursachen erzeugten Signaturen mit der Signatur verglichen werden, die durch kontinuierliche Messung gewonnen wurde [MMB99].

Die Fehleridentifikation erfolgt schließlich dadurch, dass der isolierte Parameter des Prozessmodells an die gemessenen Ein- und Ausgangsdaten angepasst wird. Dieser Vorgang wird als *model fitting* bezeichnet und ist in [MBCG99] ausführlich beschrieben.

Um einen Überblick über den gesamten Diagnosevorgang und seine Anwendung zu erhalten, empfiehlt sich die Lektüre von [GBR04]. Eine ausführliche Darstellung dieser Methode für die Diagnose eines Zwei-Tank-Systems ist in [Röm04] zu finden.

In [Röm04] wird versucht, das Verfahren auf hybride Systeme mit ausschließlich diskreten Messungen zu erweitern, was grundsätzlich möglich ist. Die Umsetzung des Verfahrens am dort betrachteten Zwei-Tank-System scheitert jedoch daran, dass zu wenige Messinformationen vorliegen, um erfolgreich zwischen unterschiedlichen Fehlern unterscheiden zu können.

Zwar ist es möglich, nach dem Erkennen von Abweichungen (Fehlerdetektion) zwischen dem realen System und dem linearen, stochastischen Beobachter, eine erfolgreiche Rückwärtssuche im TCG durchzuführen und somit die möglichen Ursachen für die erkannte Abweichung zu bestimmen; allerdings können die durch die ebenfalls erfolgreich mögliche Vorwärtssuche generierten Fehlersignaturen nicht verwendet werden, da mit Hilfe der vorliegenden Messinformationen keine Signaturen für bestimmte Fehlerursachen ausgeschlossen werden können. Die Signaturen beinhalten – wie bereits beschrieben – neben dem Verhalten der Messgrößen auch das Verhalten der ersten und zweiten Ableitung dieser Größen. Die Detektion von Abweichungen in den Ableitungen stellt schon mit kontinuierlicher Sensorik bei äquidistant abgetasteten Systemen ein großes Problem dar. Im Falle von diskreten Sensoren liegt allerdings ein System mit nicht äquidistanten Abtastschritten vor. Damit wird eine sinnvolle Ermittlung von Ableitungen bzw. von Abweichungen der Ableitungen unmöglich, da die einzelnen Messzeitpunkte zeitlich zu weit auseinander liegen, um eine Aussage über die Ableitungen zu ermöglichen.

Diese Möglichkeit der Fehlerdiagnose bei hybriden Systemen mit Hilfe von hybriden Bonddiagrammen und dem TCG kann bei diskreten Messungen lediglich dazu genutzt werden, die Zahl der möglichen Ursachen für einen Fehler zu reduzieren. In weiteren Verfahren müsste die Fehlerisolation der mit diesem Verfahren nicht auszuschließenden Fehler vorgenommen werden. Diese Möglichkeit wäre dazu geeignet, die Fehlerisolation bei Systemen mit einer Vielzahl möglicher Fälle zu erleichtern. Allerdings müssen für die Fehlerisolation zusätzliche Verfahren entwickelt werden. Diese Problematik wird im nächsten Abschnitt weiter erörtert und eine mögliche Lösung wird dort vorgestellt.

7.2 Diagnose hybrider Systeme mit diskreter Sensorik

Dass die Herausforderung, ein System mit ausschließlich diskreten Messungen zu diagnostizieren, groß ist, lässt sich erahnen, da durch die diskreten Sensoren bei den einzelnen Messungen nur verhältnismäßig wenige zusätzlichen Informationen gewonnen werden; darüber hinaus können größere Zeiträume auftreten, in denen überhaupt kein Informationsgewinn stattfindet. Durch die Beschränkung der Information auf wenige Ereigniszeitpunkte, bei denen definierte Schwellwerte über- oder unterschritten werden, gehen alle Informationen über die Ableitungen dieser Größen verloren und können auch mit Differenzenquotienten oder ähnlichen Hilfsmitteln nicht rekonstruiert werden.

Diesen enormen Schwierigkeiten stehen wirtschaftliche Vorteile durch den Einsatz

preisgünstiger Sensoren gegenüber. Dies macht den Einsatz dieser Methodik bei Produkten attraktiv, die in großen Mengen hergestellt werden, aber auch bei Messgrößen, die nur unter großem Aufwand kontinuierlich gemessen werden können.

Daher werden in diesem Abschnitt Ansätze zur Diagnose hybrider Systeme mit ausschließlich diskreter Sensorik dargestellt, die in [Röm04] entwickelt wurden und auf den in Abschnitt 6.3 vorgestellten hybriden stochastischen Beobachter aufbauen.

Die Messgeraden bzw. Messhyperebenen werden zur Erläuterung des Prinzips im weiteren Verlauf des Kapitels als achsenparallel angenommen. Daher können die Messwerte durch Skalare x_i dargestellt werden. Im allgemeinen Fall interessiert der Abstand des Erwartungswertes des Beobachters von der Messhyperebenen. Diese Verallgemeinerung auf beliebige Messgeraden bzw. Messhyperebenen bereitet keine Schwierigkeiten.

7.2.1 Einsatz eines einzelnen hybriden stochastischen Beobachters

Die Möglichkeiten, die ein hybrider stochastischer Beobachter zur Diagnose bietet, sind eher eingeschränkt. Im Grunde kann nur das Auftreten eines Fehlers erkannt werden. Wie dies vor sich geht, zeigen die folgenden Abbildungen. Abbildung 7.1 zeigt einen fehlerfreien Trajektorienverlauf. Der Erwartungswert als Mittelpunkt der Vertrauensellipse wird, wie in Abschnitt 6.3.3 bereits erläutert, nach der herrschenden Dynamik fortentwickelt. Entsprechendes gilt für die Kovarianzmatrix. Die Lage der Vertrauensellipse in der Ebene ändert sich. Tritt nun ein internes Ereignis in Form einer Messung auf, so besitzt die Vertrauensellipse im fehlerfreien Fall mindestens einen Schnittpunkt mit der Messgeraden. Der Erwartungswert $E\{\underline{X}\}$ wird daraufhin an den Messwert x_i angeglichen und die Kovarianzmatrix wird so verändert, dass die Vertrauensellipse orthogonal zur Messgeraden einen minimalen Durchmesser bekommt; für die weitere Rechnung ist es von Vorteil, wenn der minimale Durchmesser der Kovarianzmatrix so gewählt wird, dass die Kovarianzmatrix numerisch invertierbar bleibt. In der Folge werden der neue Erwartungswert und die neue Kovarianzmatrix wieder entsprechend der herrschenden Dynamik entwickelt.

Tritt jedoch ein internes Ereignis in Form einer Messung auf und besitzt die Vertrauensellipse zu diesem Zeitpunkt *noch keinen Schnittpunkt* mit der Messgeraden, wie es Abbildung 7.2 zeigt, so kann davon ausgegangen werden, dass ein Fehler vorliegt. In der Folge muss der Erwartungswert wiederum an den Messwert angeglichen werden; die neue Kovarianzmatrix wird in diesem Falle so gewählt, dass das Stück der Messgeraden, das in der neuen Neutralen Menge liegt, von der neuen Vertrauensellipse eingeschlossen wird.

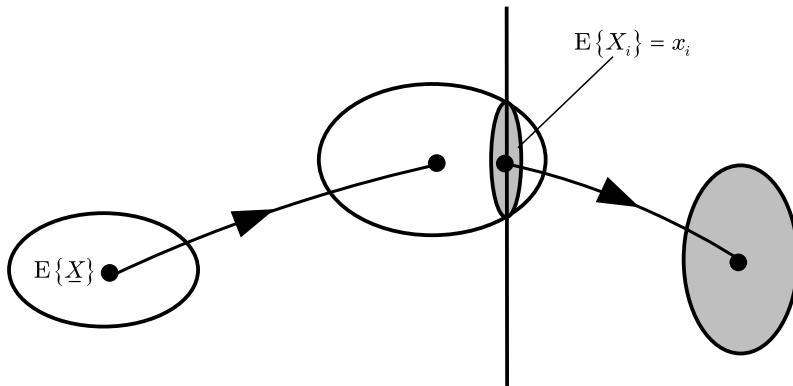


Abbildung 7.1: Fehlerfreier Trajektorienverlauf

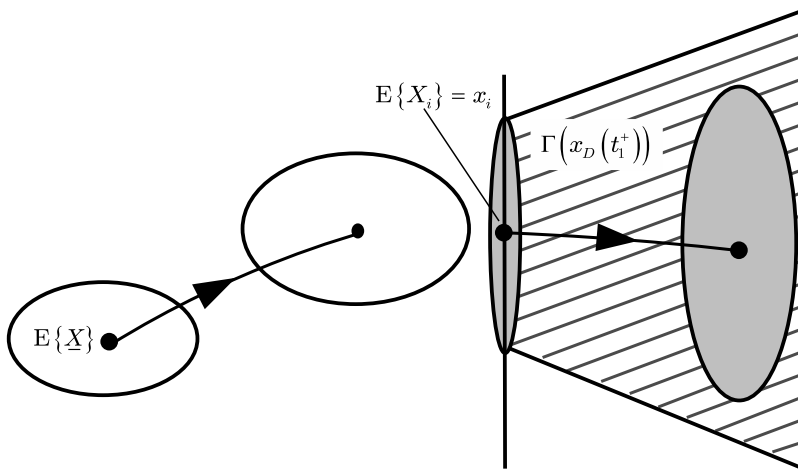


Abbildung 7.2: Fehlerhafte Trajektorie (Fehler A)

Überschreitet der Erwartungswert des Beobachters eine Messgerade, ohne dass jedoch am realen System eine Messung erfolgt, so wird die Vertrauensmenge im Beobachter zunächst weiterentwickelt. Erfolgt danach eine Messung, solange die Schnittmenge zwischen der Vertrauensmenge und der Messgeraden noch nicht leer ist, so liegt ein fehlerfreier Verlauf vor, wie er in Abbildung 7.1 zu sehen ist. Weist dagegen die Vertrauensmenge keinen Schnittpunkt mehr mit der Messgeraden auf, wie es in Abbildung 7.3 dargestellt ist, so liegt wiederum ein Fehler vor.

Es ist darauf hinzuweisen, dass mit der oben dargestellten Methode nur grobe Fehler detektiert werden können. Sollen schleichende Fehler detektiert werden, ist der Beobachter um geeignete Komponenten zu erweitern. Eine solche Erweiterung kann

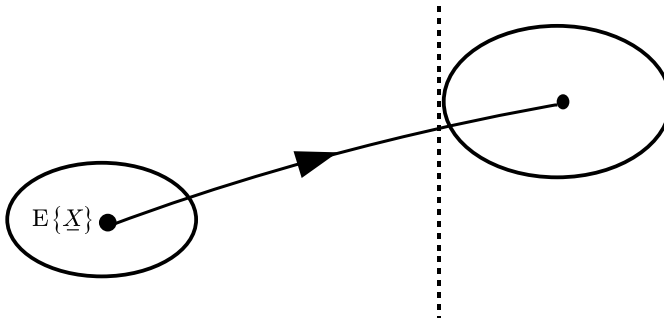


Abbildung 7.3: Fehlerhafte Trajektorie (Fehler B)

beispielsweise eine Trendanalyse der Differenzen $E\{X_i\} - x_i$ sein. Diese doch recht eingeschränkten Möglichkeiten der Diagnose – nämlich lediglich die Detektion grober Fehler – werden im Folgenden durch den Einsatz mehrerer Beobachter erweitert.

7.2.2 Mehrere hybride stochastische Beobachter

Um die Möglichkeiten der Fehlerdiagnose zu erweitern, wird eine Beobachterbank aufgebaut, wie es in [Fra94] beschrieben ist. Dazu werden verschieden parametrisierte Beobachter parallel eingesetzt. Der Beobachter, der dem realen Systemverhalten am nächsten kommt, wird als der gültige Beobachter definiert.

Somit muss also für jeden Fehler, der diagnostiziert werden soll, ein Beobachter implementiert werden, der genau so parametrisiert ist, dass er das fehlerhafte Verhalten wiedergibt. Tritt nun genau dieser Fehler auf, wird die Trajektorie des realen Systems den Vertrauensbereich des mit nominalen Parametern versehenen hybriden stochastischen Beobachters verlassen und in den Vertrauensbereich des Beobachters überwechseln, in dem der entsprechende Fehler modelliert ist. Ein solcher Trajektorienverlauf ist in Abbildung 7.4 angedeutet.

Die Detektion grober Fehler erfolgt hier wie beim Einsatz eines einzelnen Beobachters. Die anschließende Fehlerisolation erfolgt durch Auswahl des Beobachters, dessen Vertrauensbereich einen oder mehrere Schnittpunkte mit der Messhyperebene aufweist; gibt es mehrere solche Beobachter, wird der Beobachter ausgewählt, dessen Erwartungswert den geringsten Abstand $|E\{X_i\} - x_i|$ von der Messhyperebene hat.

Eine Fehleridentifikation könnte bei dieser Vorgehensweise dadurch erfolgen, dass nicht nur für jede Fehlerart, sondern auch für verschiedene Stärken des Fehlers jeweils ein eigener Beobachter implementiert wird.

Damit werden zwei Nachteile dieser Vorgehensweise deutlich:

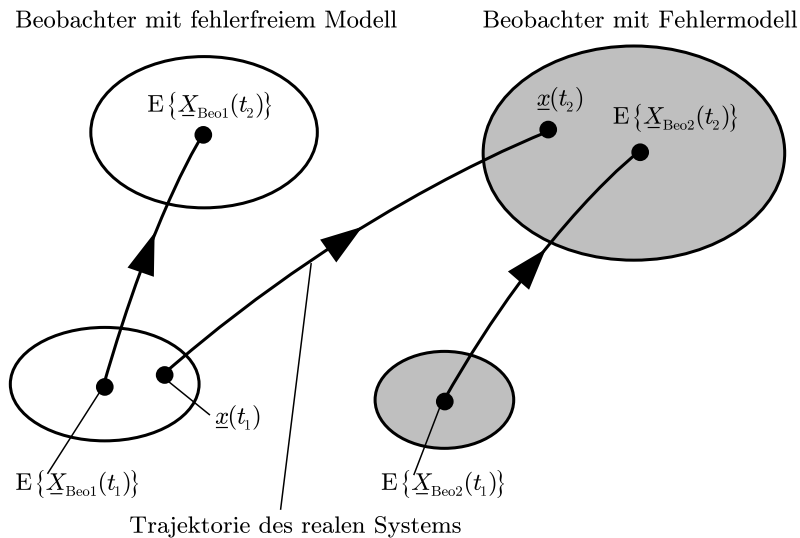


Abbildung 7.4: Diagnose mit mehreren stochastischen Beobachtern

- Eine große Anzahl von Beobachtern wird notwendig, um Fehlerisolation und Fehleridentifikation zu betreiben.
- Bei der Anwendung zeigt sich, dass es mitunter sehr schwierig ist, den richtigen Beobachter auszuwählen. Durch die Korrektur der Erwartungswerte beim Auftreten einer Messung, liegen die Erwartungswerte aller Beobachter sehr nahe beieinander.

Im folgenden Abschnitt wird aufgrund dieser Nachteile eine Möglichkeit vorgeschlagen, die die Anzahl der zur Diagnose nötigen Beobachter verringert und die Fehlerisolation erleichtert.

7.2.3 Einzelner hybrider stochastischer Beobachter mit Parameternachführung

Aufgrund der zuvor genannten Nachteile wurde ein weiteres Verfahren entwickelt, das mit nur einem Beobachter auskommt und es ermöglicht, nicht nur die Art des Fehlers zu bestimmen, sondern auch die Stärke des Fehlers zu identifizieren.

Dieses Verfahren benutzt die Differenz $E\{X_i\} - x_i$ des Beobachters vom realen System, um über PI-Algorithmen [Föl93] die Beobachterparameter nachzuführen und den Beobachter auf diesem Weg wieder mit dem realen System in Übereinstimmung zu bringen. Der Einsatz der PI-Algorithmen ist notwendig, da die Messereignisse zeit-

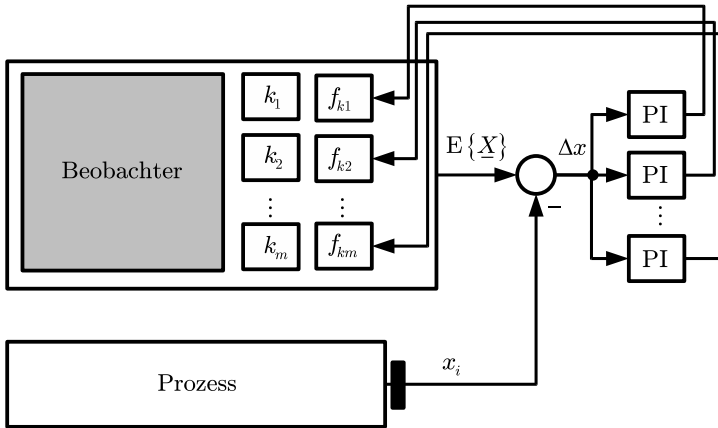


Abbildung 7.5: Diagnose mit einem stochastischen Beobachter und PI-Algorithmus

lich nicht äquidistant auftreten. Daher können keine Parameterschätzverfahren zur Bestimmung der Modellparameter verwendet werden. Zusätzlich besteht die Schwierigkeit, dass die Reihenfolge der Messereignisse variieren kann und sogar einzelne Messereignisse wegfallen können.

Für jeden Parameter k_j des Systems, der einen möglichen Fehler repräsentiert, wird in der Diagnoseeinrichtung ein eigener PI-Algorithmus mit eigenen Verstärkungsfaktoren P_j und I_j verwendet. Dieser PI-Algorithmus sorgt durch die Skalierungsfaktoren f_{kj} für die Anpassung des Parameters k_j .

Der nachgeführte Parameter gibt Auskunft über die Art des Fehlers (Fehlerisolierung, Definition 6.11); die Größenordnung der Parameteränderung lässt Rückschlüsse auf die Größenordnung des Fehlers zu (Fehleridentifikation, Definition 6.12).

Wichtig bei diesem Verfahren ist, dass zur Parameternachführung nicht beliebige Differenzen $E\{X_i\} - x_i$ verwendet werden. Dazu ist die Überlegung hilfreich, dass nicht alle Modellparameter in jedem Gebiet des Zustandsraumes die Systemdynamik beeinflussen. Soll ein Parameter aufgrund einer Messung verändert werden, so muss sichergestellt sein, dass der Parameter im Vorfeld der Messung maßgeblichen Einfluss auf die Systemdynamik hat. Dies wird am Beispiel des Zwei-Tank-Batchprozesses im nächsten Kapitel genauer erläutert.

Die Struktur dieses Ansatzes zeigt Abbildung 7.5; am Prozessausgang ist dabei mit Hilfe einer Transition angedeutet, dass die Messwerte nur zu diskreten Ereigniszeitpunkten zur Verfügung stehen. Die Differenzen zwischen Erwartungswert und Messwert werden den PI-Algorithmen zur Verfügung gestellt, wobei für einen PI-Algorithmus nur die Differenzen bei bestimmten Typen von Messereignissen von

Bedeutung sind. Dabei bildet jeder diskrete Zustandswechsel einen eigenen Typ von Messereignissen.

Um die Arbeitsweise der PI-Algorithmen genauer zu erklären, wird zunächst die Notation vereinbart:

N	bezeichnet die Anzahl der unterschiedlichen Typen von Messereignissen bei dem betrachteten Prozess.
$a \in \{1, \dots, N\}$	ist der Index für die unterschiedlichen Messereignisse.
M	ist die Anzahl aller bisher aufgetretenen Messereignisse.
$l \in \{1, \dots, M\}$	ist der Index zur Nummerierung der Messereignisse.
M_a	ist die Anzahl der bisher aufgetretenen Messereignisse vom Typ a .
$l_a \in \{1, \dots, M_a\}$	ist der Index, der die Messereignisse vom Typ a zählt.

Da hier davon ausgegangen wird, dass die Messhyperebenen im Zustandsraum achsenparallel sind, werden die unterschiedlichen Messereignisse dadurch beschrieben, dass die i -te Komponente des Zustandsvektors einen bestimmten Wert annimmt. Geschieht dies bei einer Messung vom Typ a , so wird diese mit $x_{i,a}$ gekennzeichnet.

Die einzelnen Messereignisse und die zugehörigen Erwartungswerte müssen für den weiteren Verlauf der Arbeit nummeriert werden. Um die i -te Komponente des Erwartungswerts beim l_a -ten auftreten einer Messung vom Typ a zu beschreiben, wird die Notation

$$E\{X_{i,a}^{(l_a,l)}\}$$

eingeführt. Auf diesen Erwartungswert kann auch durch die Zählung aller Ereignisse über den Index l zugegriffen werden. Jeder Erwartungswert zu den Messereignissen kann demnach eindeutig entweder durch den Index l oder durch den Index l_a zusammen mit a beschrieben werden.

Die entsprechende Differenz

$$\Delta x_{i,a}^{(l_a,l)} = E\{X_{i,a}^{(l_a,l)}\} - x_{i,a}$$

zwischen Erwartungswert und Messwert wird auf die gleiche Weise indiziert.

Beispiel 7.1

Es wird angenommen, dass in einem Prozess absolut gezählt bisher M Messereignisse aufgetreten sind. Das letzte aufgetretene Messereignis gehört dabei zum Typ a und stellt eine diskrete Messinformation zur Zustandsgröße i zur Verfügung. Von diesem Typ sind bisher M_a Messereignisse aufgetreten. Auf den Erwartungswert des hybriden stochastischen Beobachters zum Zeitpunkt der letzten Messung kann

nun einerseits über die absolute Zählung der Messereignisse

$$E\{X_{i,\cdot}^{(\cdot,M)}\}$$

zugegriffen werden; andererseits kann der gleiche Wert durch die Zählung der Messereignisse des Typs a

$$E\{X_{i,a}^{(M_a,\cdot)}\}$$

ermittelt werden.

Die einzelnen PI-Algorithmen arbeiten ereignisdiskret und werden durch die Messereignisse des zugehörigen Typs a getriggert. Sie ermitteln Skalierungsfaktoren auf der Grundlage der Differenzen $\Delta x_{i,a}^{(l_a,l)}$. Dazu wird die letzte ermittelte Differenz $\Delta x_{i,a}^{(M_a,M)}$ zunächst mit dem Proportionalfaktor P_j multipliziert. Parallel wird die Summe $\sum_{l_a=1}^{M_a} \Delta x_{i,a}^{(l_a,l)}$ aller bisher beim Messereignis a ermittelten Differenzen mit dem Faktor für den integralen Anteil I_j multipliziert. Je nachdem, wie sich eine Veränderung des Parameters k_j auf das System auswirkt, muss das Vorzeichen von P_j und I_j so gewählt werden, dass eine Gegenkopplung entsteht. Um nun eine Anpassung des Parameters k_j vorzunehmen, wird der Skalierungsfaktor f_{kj}

$$f_{kj} = 1 - P_j \cdot \Delta x_{i,a}^{(M_a,M)} - I_j \cdot \sum_{l_a=1}^{M_a} \Delta x_{i,a}^{(l_a,l)} \quad (7.1)$$

mit Hilfe der Differenzen berechnet, die am Messereignis a ermittelt wurden. Der nachzuführende Parameter k_j wird anschließend mit dem Skalierungsfaktor f_{kj} multipliziert. Damit ergibt sich der neue Parameter zu

$$k_{j_{\text{neu}}} = f_{kj} \cdot k_j. \quad (7.2)$$

Tritt der Fall auf, dass keine Differenzen $\Delta x_{i,a}^{(l_a,l)}$ beim zugehörigen Messereignis a vorliegen, so bleibt der Faktor $f_{kj} = 1$, da der Parameter bereits das Systemverhalten widerspiegelt. Dies ist aus Gleichung (7.1) unmittelbar ersichtlich.

Um die Verstärkungsfaktoren der PI-Algorithmen zu ermitteln, kann die Reaktion der PI-Algorithmen auf einen plötzlich auftretenden Fehler betrachtet werden. Die P_j und I_j müssen anhand dieser „Sprungantworten“ durch gezieltes Probieren eingestellt werden. Eine andere Möglichkeit zur Ermittlung der Verstärkungen ist für die hier gezeigte Anwendung zur Zeit nicht bekannt.

Für das vorgestellte Verfahren genügt der Einsatz eines einzigen Beobachters. Zudem ermöglicht es die Fehlerisolation sowie die Fehleridentifikation. Allerdings zeigt es sich schon bei einer kleinen Anzahl von nachzuführenden Parametern, dass aufgrund der Interaktionen der verschiedenen PI-Algorithmen Instabilitäten auftreten können.

Um diesen Nachteil zu umgehen, wurde ein weiteres Verfahren entwickelt, das lediglich für jede Fehlerart d.h. für jeden Parameter k_j einen eigenen Beobachter verwendet.

7.2.4 Mehrere hybride stochastische Beobachter mit Parameternachführung

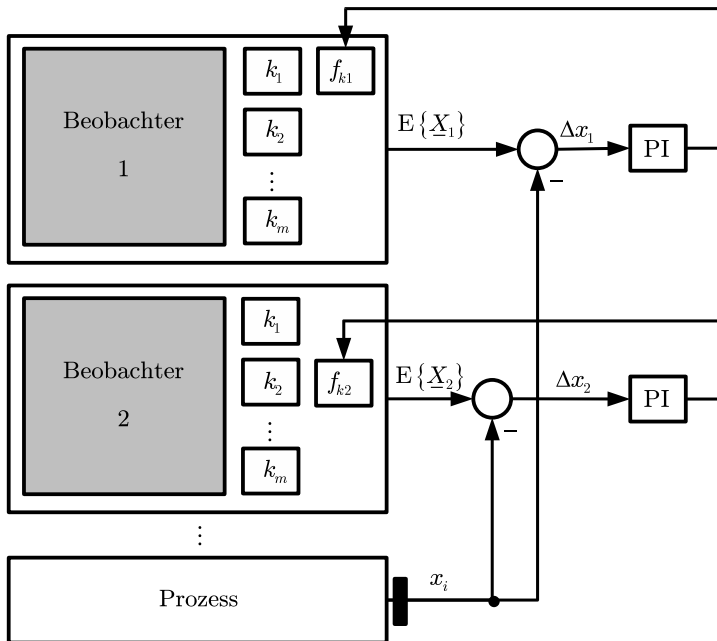


Abbildung 7.6: Diagnose mit mehreren stochastischen Beobachtern und PI-Algorithmus

Im Gegensatz zu obigem Verfahren, bei dem alle PI-Algorithmen auf das gleiche Modell zugreifen und somit mehrere Parameter in einem Modell verändert werden, beeinflusst bei dieser Vorgehensweise jeder PI-Algorithmus genau einen Parameter im zugeordneten Modell. Die Nachführung der Parameter geschieht, wie im vorherigen Abschnitt bereits erwähnt, wenn der Zustand des realen Systems bestimmte Transitionslinien überschreitet und somit Messereignisse auslöst. Auch hier sind die Messungen den PI-Algorithmen so zuzuordnen, dass das reale System in der Vorgeschichte der Messung besonders empfindlich auf den nachzuführenden Parameter reagiert.

Das bedeutet, dass jeder Beobachter unabhängig von allen anderen versucht, die

Differenz $E\{X_{i,a}^{(l_a,l)}\} - x_{i,a}$ für seine bestimmte Messgerade a zu Null auszuregeln. Diese Unabhängigkeit macht das gesamte System zwar stabil, erschwert jedoch die Fehlerisolation, da sich alle Parameter nach dem Auftreten eines Fehlers ändern.

Die Struktur dieses Ansatzes ist in Abbildung 7.6 gezeigt. Mit diesem Ansatz ist eine zuverlässige Fehlerdetektion möglich; eine allgemeine Lösung für das Problem der Fehlerisolation und der Fehleridentifikation existiert für dieses Verfahren hingegen nicht. Eine anwendungsbezogene Möglichkeit der Fehlerisolation und Fehleridentifikation könnte durch eine wissensbasierte Auswertung der verschiedenen Verstärkungsfaktoren k_j erreicht werden, die von den PI-Algorithmen im Fehlerfall ermittelt werden. In Abschnitt 8.7.4 wird eine solche anwendungsbezogene Möglichkeit für die Fehlerisolation beim Zwei-Tank-System vorgestellt.

Da die Fehlerisolation im Allgemeinen mit diesem Verfahren nicht möglich ist, können auch keine mehrfachen Fehler nach Definition 6.6 diagnostiziert werden. Man benötigt nämlich eine Möglichkeit, den fehlerhaften Parameter zu bestimmen, muss diesen Parameter anschließend neu identifizieren und im Referenzmodell verwenden, um weitere Fehler zu detektieren.

7.2.5 Mehrere hybride stochastische Beobachter mit Parameternachführung und Gütekriterium

Zur Verbesserung der Fehlerisolation wird nun ein Gütekriterium eingeführt, das bewertet, ob eine vom PI-Algorithmus getätigte Parameternachführung dazu führt, dass das resultierende Modell den Prozess besser beschreibt. Die Struktur dieses Ansatzes zeigt Abbildung 7.7.

Nach dieser Struktur wird zunächst wie beim vorherigen Ansatz mit Hilfe der bereits erläuterten PI-Algorithmen für jeden einzelnen Parameter eine Nachführung vorgenommen. Da alle PI-Algorithmen auf einen Fehler reagieren, wird parallel dazu ein Gütemaß ermittelt, das die Parameternachführung über eine definierte Anzahl zurückliegender Messungen bewertet. Dazu wird der Abstand von Erwartungswert und Messung über die letzten S Messereignisse ermittelt. Dabei kann die Anzahl der verwendeten Messungen beispielsweise absolut angegeben werden. Bei Batchprozessen kann aber auch eine bestimmte Anzahl von Batchdurchläufen angegeben werden.

Haben bis zur Ermittlung des Gütemaßes insgesamt M Messungen stattgefunden, so erhält man das Gütemaß für den Beobachter j , indem man über die Absolutwerte der letzten S Differenzen summiert

$$J_j = \sum_{l=M-S+1}^M |\Delta x_{i,a}^{(l_a,l)}|.$$

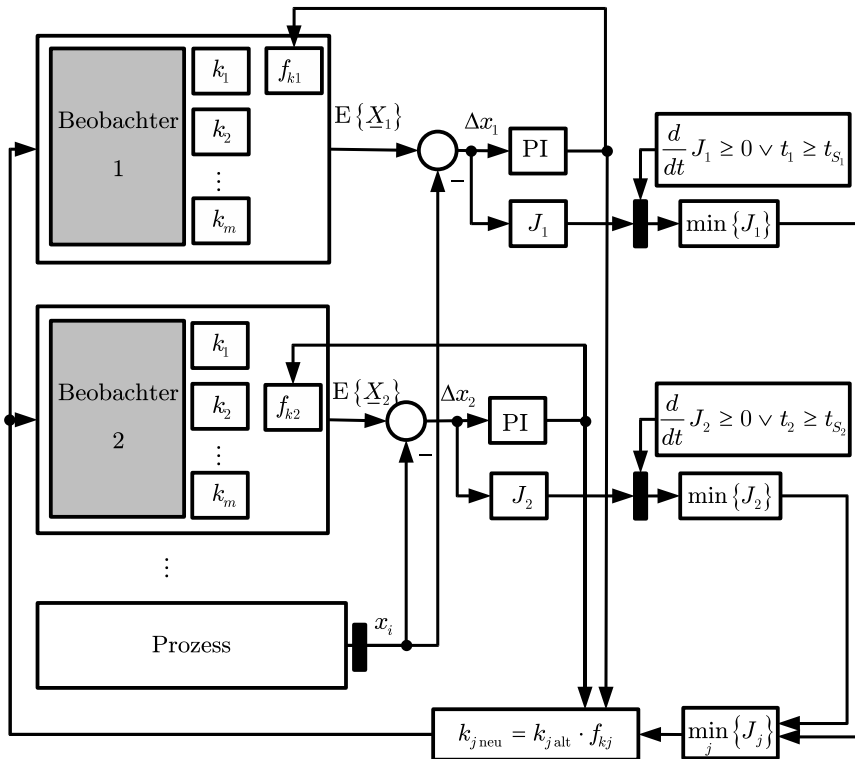


Abbildung 7.7: Diagnose mit mehreren stochastischen Beobachtern, PI-Algorithmus und Gütekriterium

Das Gütemaß ermöglicht es, den Beobachter auszuwählen, der durch seine Parameternachführung die beste Übereinstimmung mit dem realen System beim Überschreiten der unterschiedlichen Messhyperebenen erzielt.

Die Auswahl des Modells, das den aktuellen Systemzustand am besten beschreibt, erfolgt dadurch, dass mit Hilfe des Gütekriteriums nach jedem neuen Messereignis geprüft wird, ob die Parameternachführung durch den PI-Algorithmus erfolgreich war; d.h. es wird versucht, ein Minimum des Gütemaßes zu finden, indem man die Modellparameter verändert.

Die Ermittlung des Minimums kann dadurch geschehen, dass man überprüft, ob der Wert des Gütemaßes nach einer Veränderung abgenommen hat. Die Veränderungen der einzelnen Modellparameter k_j werden so lange zugelassen wie das zugehörige Gütemaß abnimmt. Nimmt ein Gütemaß J_j durch eine Parameteränderung zu, so bedeutet dies, dass ein lokales Minimum überschritten wurde. Der Skalierungsfaktor des PI-Algorithmus, der zum lokal minimalen Gütemaß J_j geführt hat, wird

gespeichert. Damit ist ein lokales Minimum gefunden und die Veränderung des Skalierungsfaktors durch den PI-Algorithmus wird abgebrochen. Zusätzlich wird die Nachführung der Parameter durch ein Zeitlimit begrenzt. Die P- und I-Anteile des PI-Algorithmus müssen bei dieser Art der Minimumbildung so gewählt werden, dass die Skalierungsfaktoren aperiodisch einschwingen. Andernfalls lässt sich nicht der bestmögliche Skalierungsfaktor für den jeweiligen Parameter finden.

Nachdem alle Modellparameter ein Minimum des Gütemaßes erreicht haben oder das Zeitlimit verstrichen ist, kann unter den Gütemaßen J_j das Minimale ausgesucht werden. Dadurch erhält man den Modellparameter, der das aktuelle Verhalten des Systems am besten beschreibt. Dieser Parameter kann anschließend in allen Beobachtern als neuer Nominalwert verwendet werden. Durch die Verwendung der Gütemaße kann die Nachführung eines falschen Parameters verhindert werden, da nur der Parameter weiter verwendet wird, der über den betrachteten Zeithorizont zum kleinsten Gütemaß führt. Die übrigen Parameter werden wieder auf ihre ursprünglichen Werte bzw. deren Skalierungsfaktoren auf 100% gesetzt, da sie das aktuelle Systemverhalten schlechter beschreiben und daher nicht richtig gewesen sein können. Auf diese Art ist es möglich, mehrfache Fehler, wie sie in Definition 6.6 angegeben sind, zu detektieren.

Bezüglich der Gütemaße ist zu beachten, dass bei nichtlinearen Systemen, die affin approximiert wurden, die Gütemaße der Beobachter auch bei fehlerfreiem System einen von Null verschiedenen Wert $J_j \neq 0$ annehmen.

Um die Bewertung der Parameterveränderungen gegenüber Messfehlern robust zu machen, kann die Anzahl der Messereignisse, die im Gütemaß betrachtet werden, erhöht werden.

Dieses Verfahren ermöglicht es, Fehler zu detektieren, zu isolieren und die Quantität der Fehler zu identifizieren, wie in Kapitel 8 am Beispiel des Zwei-Tank-Systems näher ausgeführt wird.

Zusammenfassung

In diesem Kapitel wurden mehrere Verfahren zur Diagnose hybrider Systeme vorgestellt. Im Einzelnen waren dies die Diagnose mittels Bonddiagrammen, die Diagnose mit einem einzelnen hybriden stochastischen Beobachter und die Diagnose mit mehreren Beobachtern. Die Vorgehensweise dieser Verfahren wurde jeweils erläutert und Erweiterungen vorgestellt, die eine bessere Diagnose ermöglichen bzw. bestehende Beschränkungen der jeweiligen Verfahren aufheben.

Die Verfahren, die auf der Grundlage des hybriden stochastischen Beobachters entwickelt und deren theoretische Grundlagen in diesem Kapitel vorgestellt wurden,

werden im folgenden Kapitel auf ein nichtlineares hybrides System mit diskreter Sensorik angewendet. Bei der Anwendung wird deutlich werden, dass das Verfahren aus Abschnitt 7.2.5 mit Parameternachführung und Gütemaß die größte Leistungsfähigkeit besitzt, da mit diesem Verfahren sowohl Fehler detektiert, isoliert als auch identifiziert werden können. Die zuvor vorgestellten Verfahren in den Abschnitten 7.2.1 bis 7.2.4 besitzen zwar einige Schwächen, benötigen jedoch weniger Rechenzeit und sind bei kleinen Diagnoseaufgaben problemlos anwendbar.

Kapitel 8

Anwendung der hybriden Diagnoseverfahren an einem Zwei-Tank-System

Die in Kapitel 7 vorgestellten Diagnoseverfahren für hybride Systeme mit diskreter Sensorik werden in diesem Kapitel wie in [Röm04] auf ein Zwei-Tank-System angewandt. Anhand dieses nichtlinearen hybriden Systems werden die Verfahren getestet und auf ihre Praxistauglichkeit hin überprüft.

8.1 Aufbau des Zwei-Tank-Systems

Das zur Demonstration benutzte Zwei-Tank-System steht als Laboranlage zur Verfügung. Es besteht, wie Abbildung 8.1 zeigt, aus zwei Tanks, die in der Höhe h durch eine Leitung mit dem Querschnitt a_{12} verbunden sind. Die Querschnittsflächen der Tanks werden mit A_1 und A_2 bezeichnet. Die Füllhöhen der Tanks entsprechend mit x_{K1} und x_{K2} .

Darüber hinaus besitzt jeder Tank ein Abflussventil V_1 bzw. V_2 , deren Abflussquerschnitte mit a_1 und a_2 bezeichnet sind. Im Gegensatz zu dem Zwei-Tank-System, das für die Identifikation in Abschnitt 5.2 verwendet wurde, können die Abflussventile hier einzeln angesteuert werden. Eine Pumpe mit Druckleitung zu Tank 1 führt diesem den Volumenstrom Q zu. Die Ventile V_1 und V_2 sind Magnetventile, die entweder vollständig geöffnet oder vollständig geschlossen sind. Der Volumenstrom Q kann über die Pumpenleistung kontinuierlich beeinflusst werden.

Das Zwei-Tank-System stellt ein hybrides System dar, da es sowohl einen ereignisgetriebenen als auch einen zeitgetriebenen Anteil besitzt. Ein Ereignis wäre beispielsweise das Öffnen oder Schließen eines Abflussventils, das als externes Ereignis

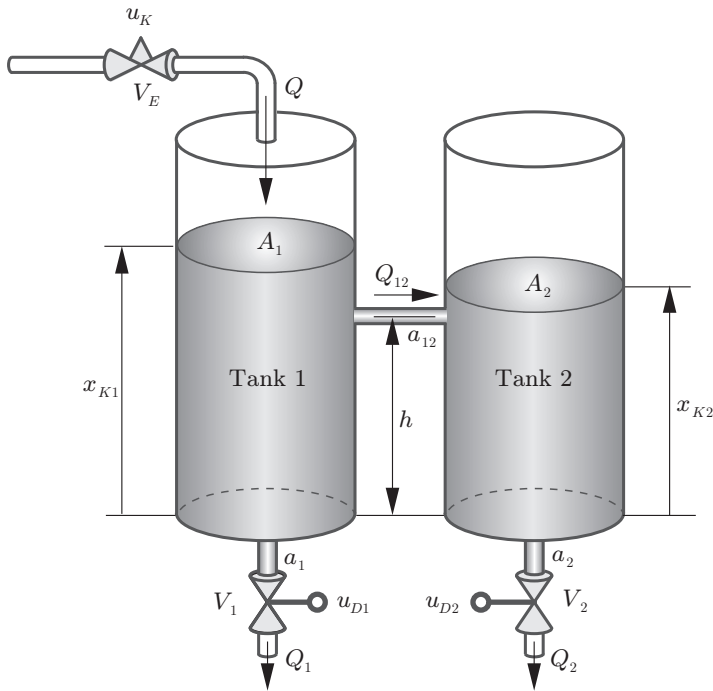


Abbildung 8.1: Aufbau des Zwei-Tank-Systems

bezeichnet wird, da es von außen, z.B. vom Bediener verursacht, auf das System einwirkt. Erreicht die Füllhöhe x_{K1} die Höhe h , beginnt Flüssigkeit von Tank 1 in Tank 2 zu fließen, falls $x_{K2} < x_{K1}$ gilt. Dieses Ereignis wird nicht von außen vorgegeben, sondern ergibt sich aus dem System selbst und wird daher als internes Ereignis bezeichnet. Die verschiedenen hybriden Phänomene wurden bereits in Kapitel 2 genauer erläutert.

Da es sich beim Zwei-Tank-System um ein nichtlineares hybrides System handelt, muss eine geeignete, stückweise affine Approximation erfolgen, damit ein hybrider stochastischer Beobachter für dieses System entworfen werden kann. Darauf wird in Abschnitt 8.5 noch näher eingegangen.

8.2 Diskrete Sensorik

Der Laboraufbau des Zwei-Tank-Systems ist mit kontinuierlichen Sensoren für die Füllstände der Tanks 1 und 2 ausgerüstet. Für diese Anwendung wird jedoch davon ausgegangen, dass lediglich diskrete Sensoren am System eingesetzt werden.

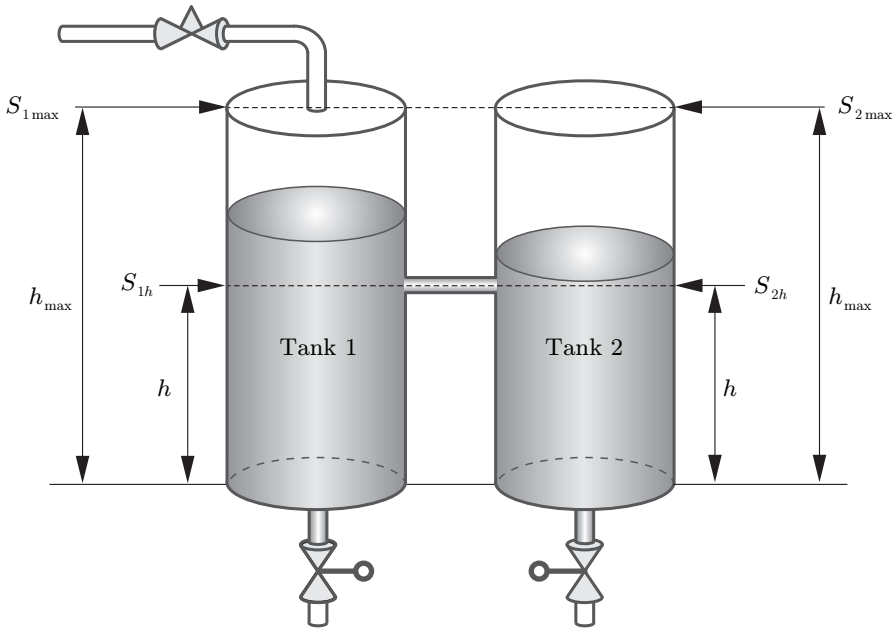


Abbildung 8.2: Diskrete Sensorik des Zwei-Tank-Systems

Da ein diskreter Füllstandssensor lediglich die Information liefert, ob der Füllstand eines Tanks ober- oder unterhalb der Montagehöhe des Sensors liegt, werden die diskreten Sensoren an der Zwei-Tank-Laboranlage dadurch simuliert, dass ein Vergleich der kontinuierlich gemessenen Höhen mit den gewünschten Montagehöhen h_{Montage_i} der diskreten Sensoren durchgeführt wird. Für $x_{Ki} > h_{\text{Montage}_i} + \varepsilon_i$ wird eine 1 ausgegeben, für $x_{Ki} \leq h_{\text{Montage}_i} - \varepsilon_i$ eine 0. Im Bereich $h_{\text{Montage}_i} - \varepsilon_i < x_{Ki} \leq h_{\text{Montage}_i} + \varepsilon_i$ gilt der Wert 0, falls der Bereich von unten betreten wird und der Wert 1, falls der Bereich von oben betreten wird. Diese Hysterese sorgt dafür, dass es beim Schalten des diskreten Sensors nicht zu mehreren schnellen Umschaltungen (Prellen) kommt.

Die diskreten Sensoren sind an beiden Tanks jeweils in der Höhe h und h_{max} angebracht. Die Positionen der Sensoren S_{1h} und S_{2h} fallen mit den Triggerrandmengen des Modells zusammen, so dass bei einer Änderung des Messwertes an einem der beiden Sensoren ein diskretes Ereignis stattfindet. Entsprechend kann jeder mögliche diskrete Zustand, in dem sich das Zwei-Tank-System befinden kann, mit Hilfe der Messwerte der Sensoren S_{1h} und S_{2h} sowie mit Hilfe der Stellung der Abflussventile V_1 und V_2 angegeben werden, so dass das Zwei-Tank-System mit diskreter Sensorik die Voraussetzungen nach Abschnitt 6.3.5 für ein hybrides System mit messbaren diskreten Eingangsgrößen erfüllt. Daher kann der in Abschnitt 6.3 vorgestellte hybride stochastische Beobachter angewandt werden.

Der diskrete Sensor $S_{1_{\max}}$ dient der Steuerung des Zwei-Tank-Batchprozesses. Dieser Batchprozess wird in Abschnitt 8.5 vorgestellt. Der Sensor $S_{2_{\max}}$ wurde eingebaut, um ein Überlaufen von Tank 2 mit Hilfe einer Verriegelungssteuerung zu verhindern. Die verschiedenen Verfahren, die zur Diagnose eingesetzt werden, verfügen ausschließlich über die Information, die die simulierten diskreten Sensoren liefern.

8.3 Fehlerarten

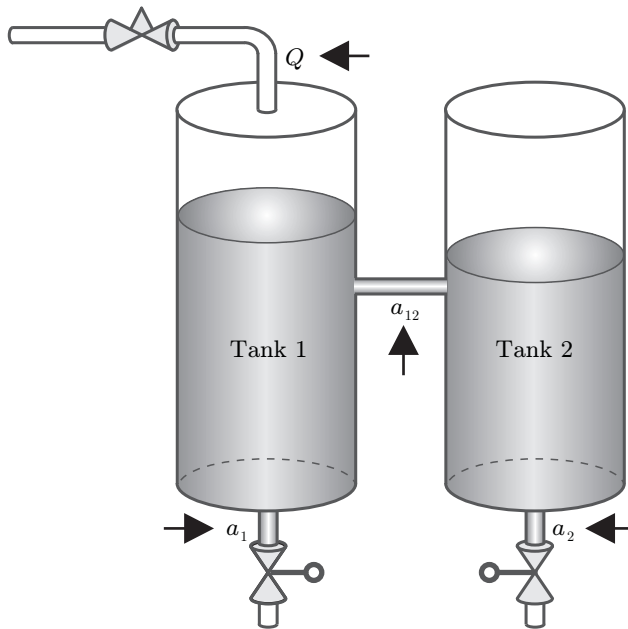


Abbildung 8.3: Fehlerarten des Zwei-Tank-Systems

Um die Aufgabe der Fehlerdiagnose sowohl am nichtlinearen Simulationsmodell als auch am Zwei-Tank-Laborsystem durchführen zu können, beschränken sich die zu diagnostizierenden Fehler auf Änderungen der in Abbildung 8.3 mit Pfeilen gekennzeichneten Parameter. Die Forderung, die Parameteränderungen auch am Laborsystem nachbilden zu können, ergibt folgende mögliche Fehlerarten:

- Der Zufluss kann infolge zu geringer Pumpenleistung, Verschleiß der Pumpe oder einer Verstopfung der Zufussleitung vermindert sein:
 $Q \leq Q_{\text{soll}}$.
- Der Abflussquerschnitt von Tank 1 kann kleiner als angenommen sein z.B. als Folge einer Verstopfung durch Fremdkörper, durch Ablagerungen oder ein

nicht vollständig öffnendes Ventil V_1 :

$$a_1 \leq a_{1_{\text{soll}}}.$$

- Der Abflussquerschnitt von Tank 2 kann vermindert sein z.B. als Folge einer Verstopfung durch Fremdkörper, durch Ablagerungen oder ein nicht vollständig öffnendes Ventil V_2 :

$$a_2 \leq a_{2_{\text{soll}}}.$$

- Die Querschnittsfläche der Verbindungsleitung zwischen Tank 1 und Tank 2 kann aufgrund von Fremdkörpern oder Ablagerungen verkleinert sein:

$$a_{12} \leq a_{12_{\text{soll}}}.$$

Die Fehler werden bei Verwendung des nichtlinearen Simulationsmodells, das in Abschnitt 8.4 erläutert wird, durch Änderung der Modellparameter simuliert. Am Zwei-Tank-Laborsystem kann eine verminderte Pumpenleistung durch Ändern des Parameters für den Zufluss in der Steuersoftware erfolgen. Die übrigen Fehler werden mit Hilfe von Stopfen, wie sie in Anhang C erläutert sind, erzeugt. Diese Stopfen müssen von Hand in das Zwei-Tank-Laborsystem eingesetzt werden.

Die erwähnten Parameter können sich auch zu größeren Werten hin ändern. So bedeutet eine Vergrößerung von $Q > Q_{\text{soll}}$, dass die Pumpe einen größeren Volumenstrom als im Normalbetrieb fördert. Entsprechend bedeutet $a_1 > a_{1_{\text{soll}}}$, dass der Volumenstrom aus Tank 1 größer als im Normalbetrieb ist. Grund dafür kann ein Leck im Tank oder ein durch Korrosion vergrößerter Abflussquerschnitt sein. Gleiches gilt für Tank 2. Eine durch Korrosion vergrößerte Verbindungsleitung zwischen den beiden Tanks würde sich durch $a_{12} > a_{12_{\text{soll}}}$ bemerkbar machen.

8.4 Nichtlineares Simulationsmodell

Um den Beobachter und die notwendigen Algorithmen zur Diagnose am Rechner entwickeln zu können, wurde – als Ersatz für das Laborsystem – ein hybrides Simulationsmodell benötigt.

Das Zwei-Tank-System ist – wie für die Identifikation in Abschnitt 5.2.3 – als Netz-Zustands-Modell modelliert. Der einzige Unterschied zu dem Zwei-Tank-System, das für die Identifikation verwendet wurde, ist, dass die Abflussventile jetzt einzeln angesteuert werden können.

Die Herleitung der nichtlinearen Modellgleichungen wurde bereits in Abschnitt 5.2.3 durchgeführt; die Implementierung erfolgte auch hier in MATLAB/SIMULINK und STATEFLOW.

8.5 Batchprozess

Um die Diagnose am Zwei-Tank-System durchführen zu können, muss das System über einen längeren Zeitraum hinweg betrieben werden. Darüber hinaus muss sich das System so verhalten, dass immer wieder Messereignisse auftreten, da nur so Informationen über das System erlangt werden können. Dabei wird unter einem Messereignis das Schalten eines diskreten Sensors verstanden.

Diese Forderung nach immer wiederkehrenden Messereignissen, die aufgrund der Einschränkung auf diskrete Messungen erhoben werden muss, macht einen entsprechenden Betrieb des Zwei-Tank-Prozesses notwendig. Daher wird der Zwei-Tank-Prozess in diesem Anwendungsbeispiel als Batchprozess betrieben. Aufgrund der sich dadurch ergebenden geschlossenen Trajektorie werden im Prozessverlauf immer wieder Messgeraden geschnitten und somit Messereignisse ausgelöst.

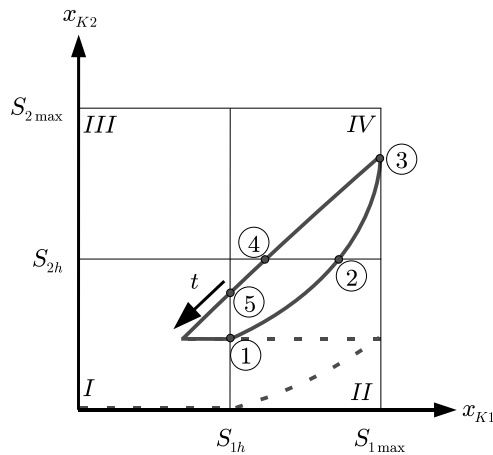


Abbildung 8.4: Zwei-Tank-Batchprozess

Abbildung 8.4 zeigt die Trajektorie des Batchprozesses, der mit Hilfe einer Steuerung, die die Ventile und die Pumpe schaltet, realisiert wurde. Nach dem Anfahrvorgang, der als gestrichelte Linie angedeutet ist, befindet sich das Zwei-Tank-System im Gebiet I der Zustandsebene. Dies bedeutet, dass die Füllhöhen beider Tanks unter der Höhe h der Verbindungsleitung liegen. Zu Beginn des Batchprozesses werden die beiden Abflussventile V_1 und V_2 geschlossen und die Pumpe eingeschaltet, so dass Tank 1 mit dem Volumenstrom $Q = Q_{\text{soil}}$ befüllt wird. Der Füllvorgang bewirkt, dass x_{K1} bis zur Höhe h der Verbindungsleitung steigt. Das Erreichen der Höhe $x_{K1} = h$ ist mit einem messbaren diskreten Ereignis in Punkt 1 verbunden.

Der Batchprozess erreicht nun Gebiet II der Zustandsebene und der Tank 2 beginnt

sich ebenfalls zu füllen. Das messbare interne Ereignis $x_{K2} = h$ bei Punkt 2 zeigt an, dass die Füllhöhe von Tank 2 die Höhe der Verbindungsleitung erreicht hat. Dies entspricht dem Eintritt der Trajektorie in Gebiet *IV*.

Das Erreichen der maximalen Höhe $x_{K1} = h_{1\max}$ bei Punkt 3 wird durch den Sensor $S_{1\max}$ erkannt. Dies bewirkt das Abschalten der Pumpe und das Öffnen der Abflussventile V_1 und V_2 . Es gilt nun $Q = 0$ und beide Tanks beginnen sich zu leeren.

Unterschreitet die Füllhöhe von Tank 2 die Höhe der Verbindungsleitung – gilt also $x_{K2} \leq h$ – wechselt die Trajektorie des Zwei-Tank-Systems bei Punkt 4 wieder von Gebiet *IV* in das Gebiet *II* und in der Folge bei $x_{K1} \leq h$ in Gebiet *I*. Die Gebietswechsel entsprechen wiederum Änderungen des diskreten Zustandes und sind messbare interne Ereignisse.

Sobald die Trajektorie in das Gebiet *I* eindringt, wird ein Timer gestartet. Nach Ablauf der Zeit t werden die Ventile V_1 und V_2 von der Steuerung geschlossen und die Pumpe wird gestartet ($Q = Q_{\text{sol}}$).

Der beschriebene Vorgang wiederholt sich so lange bis das Zwei-Tank-System abgeschaltet wird. Dazu werden beim Zwei-Tank-Laborsystem alle Ventile geschlossen und die Pumpe abgeschaltet, so dass sich das System in einem sicheren Zustand befindet.

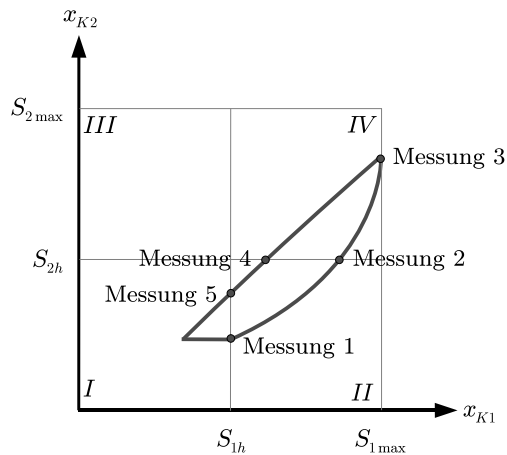


Abbildung 8.5: Diskrete Messungen am Zwei-Tank-Batchprozess

Im Zwei-Tank-Batchprozess finden im Normalbetrieb fünf diskrete Messungen je Zyklus statt. Die Messpunkte sind in Abbildung 8.5 hervorgehoben. Es ist zu beachten, dass es bei einem Fehler im Prozess vorkommen kann, dass sich die Zahl der Messungen je Zyklus ändert. Ein gutes Beispiel hierfür ist eine Verringerung

des Querschnitts der Verbindungsleitung. Dies kann dazu führen, dass die Füllhöhe in Tank 2 nicht mehr über die Höhe h der Verbindungsleitung hinaus ansteigt. In diesem Fall gilt $x_{K2} < h$ für den gesamten Zyklus, so dass die Messungen 2 und 4 entfallen. Aufgrund dieses Phänomens können bei der Identifikation der fehlerhaften Parameter keine Verfahren eingesetzt werden, die auf eine gleichbleibende Reihenfolge der Messereignisse angewiesen sind.

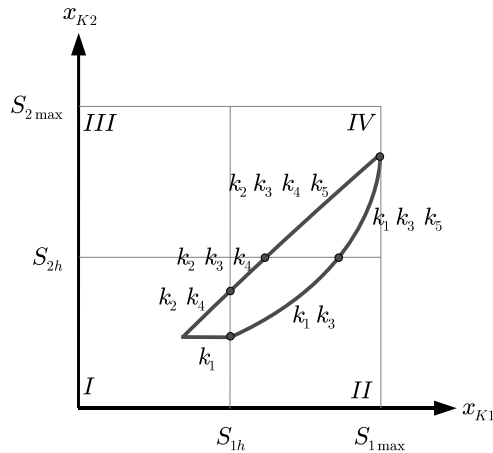


Abbildung 8.6: Wirksame Parameter k_j des Zwei-Tank-Batchprozesses.

Betrachtet man den Batchprozess im Hinblick auf die Einwirkungen der einzelnen Parameter, so erkennt man in Abbildung 8.6, dass sich auf bestimmten Trajektorienabschnitten lediglich bestimmte Parameter k_j auf den Prozess auswirken. Es hängt vom diskreten Zustand des hybriden Zwei-Tank-Systems ab, ob sich ein bestimmter Parameter auf den Prozess auswirkt. Deshalb wird die Bedeutung der Parameter k_j und der Ausgänge v_{K_i} des D/K Interfaces im folgenden Abschnitt erläutert.

8.6 Stochastischer Beobachter

Der stochastische Beobachter ist entsprechend der Ausführungen in Abschnitt 6.3 erstellt. Dazu werden die Geometriedaten des Systems benötigt, die sich in den Parametern k_j widerspiegeln.

Im einzelnen ergeben sich die Parameter k_i zu:

$$\begin{aligned} k_1 &= \frac{Q_{\text{soll}}}{A_1}, \\ k_2 &= \frac{a_1}{A_1} \cdot \sqrt{2g}, \\ k_3 &= \frac{a_{12}}{A_1} \cdot \sqrt{2g}, \\ k_4 &= \frac{a_2}{A_2} \cdot \sqrt{2g}, \\ k_5 &= \frac{a_{12}}{A_2} \cdot \sqrt{2g}, \end{aligned}$$

wobei $g = 9.81 \frac{\text{m}}{\text{s}^2}$ gilt.

Die Flächen A_1 und A_2 sind gleich groß, so dass für die Parameter $k_3 = k_5$ folgt. Deshalb wird in den folgenden Gleichungen nur der Parameter k_3 verwendet.

Die v_{K_i} sind die Ausgänge des D/K-Interface. Sie ergeben sich aus dem diskreten Zustand des Zwei-Tank-Systems und bestimmen die jeweils gültige Dynamik. Im Einzelnen haben die Ausgänge der diskret-kontinuierlichen Schnittstelle, sofern sie den Wert 1 haben, folgende Bedeutung:

- v_{K1} : kontinuierlicher Zustand im Gebiet *I* der Zustandsebene
- v_{K2} : kontinuierlicher Zustand im Gebiet *II* der Zustandsebene
- v_{K3} : kontinuierlicher Zustand im Gebiet *III* der Zustandsebene
- v_{K4} : kontinuierlicher Zustand im Gebiet *IV* der Zustandsebene
- v_{K5} : Abfluss Tank 1 offen
- v_{K6} : Abfluss Tank 2 offen

Zur Einteilung der Gebiete sei auf Abbildung 8.4 verwiesen. Da es sich beim Zwei-Tank-System um ein nichtlineares hybrides System handelt, müssen die nichtlinearen hybriden Differentialgleichungen des Zwei-Tank-Systems zur Anwendung des stochastischen Beobachters zunächst affin approximiert werden. Wie in Abschnitt 5.2 bereits gezeigt ist, ergeben sich aus dem Torricelli-Gesetz die folgenden nichtlinearen Differentialgleichungen:

$$\begin{aligned} \dot{x}_{K1} = & + \underbrace{k_1 Q}_{\text{Zufluss}} - \underbrace{k_2 \sqrt{x_{K1}} v_{K5}}_{\text{Abfluss 1}} - \underbrace{k_3 \sqrt{x_{K1} - h} v_{K2}}_{\text{Gebiet II}} \\ & + \underbrace{k_3 \sqrt{x_{K2} - h} v_{K3}}_{\text{Gebiet III}} - \underbrace{k_3 \sqrt{|x_{K1} - x_{K2}|} \text{sign}(x_{K1} - x_{K2}) v_{K4}}_{\text{Gebiet IV}} \end{aligned} \quad (8.1)$$

$$\begin{aligned}
\dot{x}_{K2} = & - \underbrace{k_4 \sqrt{x_{K2}} v_{K6}}_{\text{Abfluss 3}} + \underbrace{k_3 \sqrt{x_{K2} - h} v_{K2}}_{\text{Gebiet II}} - \underbrace{k_3 \sqrt{x_{K2} - h} v_{K3}}_{\text{Gebiet III}} \\
& + \underbrace{k_3 \sqrt{|x_{K1} - x_{K2}|} \text{sign}(x_{K1} - x_{K2}) v_{K4}}_{\text{Gebiet IV}}. \tag{8.2}
\end{aligned}$$

Die Approximation der Wurzelfunktionen erfolgt abschnittsweise durch die Funktionen w_1, w_2 und w_3 mit

$$\begin{aligned}
w_1 &= s_1 \cdot x_{Ki} && \text{(linear),} \\
w_2 &= s_2 \cdot x_{Ki} + c_2 && \text{(affin),} \\
w_3 &= s_3 \cdot x_{Ki} && \text{(linear).}
\end{aligned}$$

Erst die affine Näherung ermöglicht die Verwendung eines hybriden stochastischen Beobachters. Die Funktion w_1 nähert die Wurzelfunktionen an, die von 0 bis zur Höhe h der Verbindungsleitung aktiv sind. Mit w_2 werden die Terme der Wurzelfunktion approximiert, die von h bis h_{\max} benötigt werden. Für die Terme, die den Volumenstrom zwischen den Tanks beschreiben, wird eine eigene affine Approximation der Wurzelfunktion mit Hilfe der Funktion w_3 vorgenommen.

Der Parameter s_1 wird so ermittelt, dass die Gerade die Wurzelfunktion in $x = 0$ und in $x = h$ schneidet. Die Parameter von w_2 und w_3 werden gewonnen, indem die Wurzelfunktion im Abstand von 0, 1 cm abgetastet und für die resultierenden Punkte eine Approximation mit Hilfe der Methode der kleinsten Quadrate vorgenommen wird. Dadurch erhält man die Parameter

$$\begin{aligned}
s_1 &= 0.18257, \\
s_2 &= 0.077263, \\
s_3 &= 0.24024, \\
c_2 &= 3.2112.
\end{aligned}$$

Das Ergebnis dieser Näherung kann man in Abbildung 8.7 betrachten. Die zwei linearen Funktionen w_1 und w_3 und die affine Funktion w_2 sind darin für $h = 30$ cm und $h_{\max} = 55$ cm veranschaulicht. Damit kann die nichtlineare hybride

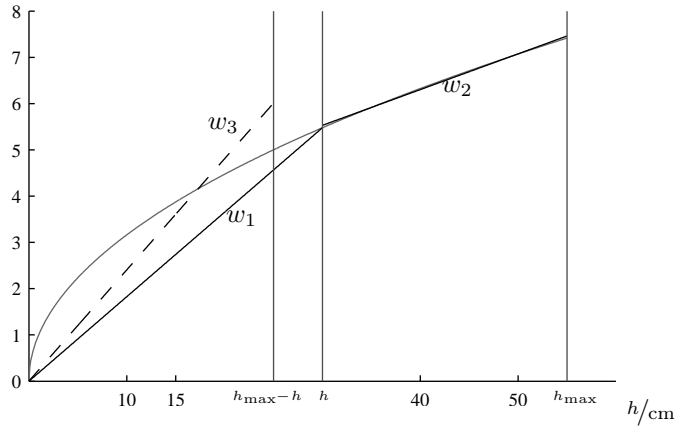


Abbildung 8.7: Affine Approximation der Wurzelfunktion

Differentialgleichung (8.1) in die stückweise affine Differentialgleichung

$$\begin{aligned}
 \dot{x}_{K1} = & + \underbrace{k_1 Q}_{\text{Zufluss}} & (8.3) \\
 & - \underbrace{k_2 s_1 \cdot x_{K1} v_{K1} v_{K5}}_{\text{Abfluss 1, Gebiet I}} - \underbrace{k_2 [s_2 \cdot x_{K1} + c_2] v_{K2} v_{K5}}_{\text{Abfluss 1, Gebiet II}} \\
 & - \underbrace{k_2 s_1 \cdot x_{K1} v_{K3} v_{K5}}_{\text{Abfluss 1, Gebiet III}} - \underbrace{k_2 [s_2 \cdot x_{K1} + c_2] v_{K4} v_{K5}}_{\text{Abfluss 1, Gebiet IV}} \\
 & - \underbrace{k_3 [s_3 \cdot x_{K1} - s_3 \cdot h] v_{K2}}_{\text{Gebiet II}} + \underbrace{k_3 [s_3 \cdot x_{K2} - s_3 \cdot h] v_{K3}}_{\text{Gebiet III}} \\
 & - \underbrace{k_3 [s_3 \cdot x_{K1} - s_3 \cdot x_{K2}] v_{K4}}_{\text{Gebiet IV}}
 \end{aligned}$$

überführt werden. Gleiches gilt für Gleichung (8.2), die nach der affinen Approximation folgende Gestalt annimmt:

$$\begin{aligned}
 \dot{x}_{K2} = & - \underbrace{k_4 s_1 \cdot x_{K2} v_{K1} v_{K6}}_{\text{Abfluss 2, Gebiet I}} - \underbrace{k_4 s_1 \cdot x_{K2} v_{K2} v_{K6}}_{\text{Abfluss 2, Gebiet II}} & (8.4) \\
 & - \underbrace{k_4 [s_2 \cdot x_{K2} + c_2] v_{K3} v_{K6}}_{\text{Abfluss 2, Gebiet III}} - \underbrace{k_4 [s_2 \cdot x_{K2} + c_2] v_{K4} v_{K6}}_{\text{Abfluss 2, Gebiet IV}} \\
 & + \underbrace{k_3 [s_3 \cdot x_{K1} - s_3 \cdot h] v_{K2}}_{\text{Gebiet II}} - \underbrace{k_3 [s_3 \cdot x_{K2} - s_3 \cdot h] v_{K3}}_{\text{Gebiet III}} \\
 & + \underbrace{k_3 [s_3 \cdot x_{K1} - s_3 \cdot x_{K2}] v_{K4}}_{\text{Gebiet IV}} .
 \end{aligned}$$

Um nun die Entwicklung des Erwartungswerts und die Entwicklung der Kovarianzmatrix mit Hilfe der Gleichungen (6.5) und (6.6) berechnen zu können, müssen die obigen Gleichungen in Matrizenform gebracht werden. Zur Entwicklung des Erwartungswerts genügt es, die deterministische Differentialgleichung

$$\dot{\underline{x}} = \underline{A} \underline{x} + \underline{B} \underline{u}$$

zu betrachten. Im einzelnen haben die Vektoren und Matrizen folgendes Aussehen:

$$\dot{\underline{x}} = \begin{pmatrix} \dot{x}_{K1} \\ \dot{x}_{K2} \end{pmatrix},$$

$$\underline{A} = \left(\begin{array}{cc|cc} -k_2 s_1 v_{K1} v_{K5} - k_2 s_1 v_{K3} v_{K5} & & & \\ -k_2 s_2 v_{K2} v_{K5} - k_2 s_2 v_{K4} v_{K5} & & +k_3 s_3 v_{K3} & \\ & -k_3 s_3 v_{K2} & & +k_3 s_3 v_{K4} \\ & -k_3 s_3 v_{K4} & & \\ \hline & & -k_4 s_1 v_{K1} v_{K6} - k_4 s_1 v_{K2} v_{K6} & \\ & +k_3 s_3 v_{K2} & -k_4 s_2 v_{K3} v_{K6} - k_4 s_2 v_{K4} v_{K6} & \\ & +k_3 s_3 v_{K4} & & -k_3 s_3 v_{K3} \\ & & & -k_3 s_3 v_{K4} \end{array} \right),$$

$$\underline{x} = \begin{pmatrix} x_{K1} \\ x_{K2} \end{pmatrix},$$

$$\underline{B} = \left(\begin{array}{c|cc} & -k_2 \frac{c_2}{h} v_{K2} v_{K5} - k_2 \frac{c_2}{h} v_{K4} v_{K5} & \\ k_1 & +k_3 s_3 v_{K2} & \\ & -k_3 s_3 v_{K3} & \\ \hline & -k_4 \frac{c_2}{h} v_{K3} v_{K6} - k_4 \frac{c_2}{h} v_{K4} v_{K6} & \\ 0 & +k_3 s_3 v_{K3} & \\ & -k_3 s_3 v_{K2} & \end{array} \right),$$

$$\underline{u} = \begin{pmatrix} Q \\ h \end{pmatrix}.$$

Dabei ist die zweite Komponente des Eingangs eine Konstante, so dass insgesamt ein affines System vorliegt.

Um den hybriden stochastischen Beobachter zu vervollständigen, müssen die Steuermatrix $\underline{S}(\underline{x}_D)$ und Diffusionsmatrix \underline{Q} festgelegt werden.

Die Steuermatrix $\underline{S}(\underline{x}_D)$, mit der vor allen Dingen Modellungenauigkeiten in jedem diskreten Zustand nachgebildet werden, wurde folgendermaßen gewählt:

$$\underline{S}(\underline{x}_D) = \left(\begin{array}{c|c} b_{SG_1}^1 v_{K1} \\ + b_{SG_1}^2 v_{K2} \\ + b_{SG_1}^3 v_{K3} \\ + b_{SG_1}^4 v_{K4} \\ \hline 0 \\ \hline 0 & b_{SG_2}^1 v_{K1} \\ & + b_{SG_2}^2 v_{K2} \\ & + b_{SG_2}^3 v_{K3} \\ & + b_{SG_2}^4 v_{K4} \end{array} \right) .$$

Die Faktoren $b_{SG_j}^i$ geben an, wie stark die Modellungenauigkeit im Gebiet i der Zustandsebene auf die Zustandsgröße j einwirkt; die oben angegebenen Faktoren $b_{SG_j}^i$ wurden für die Simulation alle zu 1 gewählt, so dass für $\underline{S}(\underline{x}_D)$ in jedem diskreten Zustand die Einheitsmatrix resultiert.

Die Diffusionsmatrix \underline{Q} ist für alle Gebiete konstant und ist die Kovarianzmatrix der Störungen, die auf den Prozess einwirken. Da angenommen wird, dass die beiden Wiener-Prozesse β_i , die auf die beiden Zustandsgrößen wirken, voneinander unabhängig sind, besitzt die Diffusionsmatrix Diagonalgestalt

$$\underline{Q} = \begin{pmatrix} q_{1,1} & 0 \\ 0 & q_{2,2} \end{pmatrix}, \quad (8.5)$$

wobei $q_{1,1}$ und $q_{2,2}$ zu 0,1 gewählt wurden.

Damit sind alle Matrizen angegeben, die benötigt werden, um die Erwartungswerte nach Gleichung (6.5) und die Kovarianzmatrix nach Gleichung (6.6) zu entwickeln.

8.7 Anwendung der Diagnoseverfahren

Auf den folgenden Seiten werden die in Kapitel 7 vorgestellten Diagnoseverfahren am nichtlinearen hybriden Zwei-Tank-System angewendet.

In den Abschnitten 8.7.1 bis 8.7.3 werden die Ergebnisse der Verfahren aus den Abschnitten 7.2.1 bis 7.2.3 vorgestellt. Dabei werden lediglich die jeweiligen Ergebnisse erläutert, da diese Verfahren für das Zwei-Tank-System zwar ausreichend leistungsfähig sind, jedoch nur schwer auf komplexere Systeme übertragen werden können.

Ausführlich vorgestellt wird die Anwendung des Verfahrens aus Abschnitt 7.2.4 in

Abschnitt 8.7.4. Dabei werden zunächst Simulationen mit dem nichtlinearen Simulationsmodell gezeigt. Zusätzlich werden diese Ergebnisse am Laborsystem validiert.

Die Anwendung des Verfahrens aus Abschnitt 7.2.5 am nichtlinearen Simulationsmodell wird in Abschnitt 8.7.5 ausführlich erläutert.

Zum Abschluss werden die Verfahren im Hinblick auf ihre Tauglichkeit für die Diagnose des Zwei-Tank-Systems verglichen.

8.7.1 Einsatz eines einzelnen stochastischen Beobachters

Dieses Verfahren ist in der Simulation und am realen System getestet worden und hat sich als zuverlässig und wenig empfindlich herausgestellt. Mit einer zusätzlich implementierten Trendanalyse können neben groben Fehlern mit Parameteränderungen von etwa 10% auch schleichende Fehler detektiert werden. Hierbei ist jedoch zu bemerken, dass die Simulation der diskreten Sensoren am realen Zwei-Tank-System und die dazu zwingend erforderliche Hysterese die Trendanalyse negativ beeinflusst.

Der Einsatz eines einzelnen stochastischen Beobachters bietet generell nur eine Fehlerdetektion. Für weitere Aufgaben in der Diagnose wie zur Fehlerisolation und Fehleridentifikation kann der einzelne hybride stochastische Beobachter nicht verwendet werden. Im Fehlerfall kann lediglich die Abschaltung der Anlage oder die Verständigung des Bedienpersonals erfolgen. Eine Zuordnung erkannter Abweichungen vom Sollzustand zu entsprechenden Bauteilen, Baugruppen oder Parametern ist aufgrund der Struktur des Verfahrens nicht möglich. Somit müsste im Fehlerfall die gesamte Anlage abgeschaltet und manuell auf Fehler überprüft werden, was bei größeren Anlagen zu einem nicht vertretbaren Aufwand führt.

Ein Vorteil dieser Methode ist, dass schleichende Fehler als solche erkannt werden können. Im Rahmen von üblichen Wartungsintervallen könnte beim Einsatz des einzelnen hybriden stochastischen Beobachters gezielt nach solchen Fehlern gesucht werden. Dies ermöglicht die Behebung von schleichenden Fehlern, bevor es zu Systemausfällen kommt.

8.7.2 Mehrere stochastische Beobachter

Verwendet man mehrere hybride stochastische Beobachter, so können die Fehler, für die die Beobachter parametrisiert sind, sowohl in der Simulation als auch am Zwei-Tank-Laborsystem erkannt werden.

Es ergeben sich jedoch einige Nachteile. Einer dieser Nachteile ist, dass durch die Angleichung der Erwartungswerte aller Beobachter an die Messwerte $E\{X_{Ki}\} = h$

zum Zeitpunkt einer Messung immer wieder Beobachter, die ein völlig falsches Systemverhalten wiedergeben, in die Nähe der Trajektorie des realen Systems gebracht werden. Dieser Umstand wurde bereits in Abschnitt 6.3 erläutert.

Dadurch kann es vorkommen, dass das Verhalten eines Beobachters zum Zeitpunkt einer Messung zufällig mit dem realen System übereinstimmt, obwohl sich die Parameter des Beobachters sehr stark von den Systemparametern unterscheiden. Daher kommt es bei diesem Verfahren immer wieder zu Fehl- und Falschalarmen. Die Anzahl der Fehl- und Falschalarme steigt mit der Gesamtzahl der verwendeten Beobachter natürlich an. Darüber hinaus existiert in der betrachteten Anwendung ein Approximationsfehler, da mit dem hybriden stochastischen Beobachter nur stückweise affine Systeme exakt beschrieben werden können. Diese Tatsache erhöht zusätzlich die Gefahr von Fehllarmen.

Da es sich beim Zwei-Tank-System um ein hybrides System handelt, kommt erschwerend hinzu, dass sich bestimmte Parameter nur in bestimmten Bereichen des Zustandsraumes auswirken, wie in Abschnitt 8.5 erläutert wurde. Daher ist bei den Experimenten auch zu beobachten, dass zwei Beobachter, die sich nur in wenigen Parametern unterscheiden, auf bestimmten Trajektorienabschnitten des Batchprozesses durch ihr Verhalten nicht unterschieden werden können. Tritt nun ein Messereignis auf, so hat man häufig keine Möglichkeit einen der Beobachter auszuschließen. Dieses Problem kann dadurch umgangen werden, dass man für jeden hybriden Zustand angibt, welche Parameter sich gerade auf das System auswirken und nur bestimmte Beobachter für das zuvor beschriebene Auswahlverfahren zulässt. Dies bedeutet jedoch einen zusätzlichen Aufwand an Entscheidungslogik.

Der gravierendste Nachteil dieses Verfahrens ist, dass für jede mögliche Fehlerart und jede mögliche Größenordnung dieser Fehlerart ein eigener Beobachter parametrisiert und parallel zum realen Prozess oder zur nichtlinearen Simulation simuliert werden muss. Möchte man beispielsweise feststellen können, ob der Abfluss eines Tanks zu 10%, 20% oder 30% verstopft ist, so muss man drei Beobachter mit dem jeweils abweichenden Parameter erstellen und betreiben. Dieser Aufwand steigt unverhältnismäßig an, wenn man sich zum Ziel setzt, auch mehrfache Fehler zu detektieren, da dann zusätzlich alle Kombinationen der verschiedenen Fehlerarten und Fehlergrößenordnungen als eigene Beobachter zu implementieren sind.

Eine Fehleridentifikation ist mit diesem Verfahren nur durch unverhältnismäßig hohen Aufwand möglich. Darüber hinaus sind die Aussagen über die Stärke des aufgetretenen Fehlers davon abhängig, wie viele unterschiedliche Beobachter für eine Fehlerart eingesetzt werden und welche Fehlerstärken in diesen Beobachtern modelliert werden. Aufgrund dieser Nachteile ist das Verfahren lediglich für einfache Systeme geeignet.

8.7.3 Ein stochastischer Beobachter mit Parameternachführung

Die Verwendung eines einzigen Beobachters in Verbindung mit mehreren PI-Algorithmen für die relevanten Fehler ermöglicht die Fehlerdetektion, die Fehlerisolation und eine eingeschränkte Fehleridentifikation.

Wie schon in Abschnitt 7.2.1 erläutert, werden die Parameter mittels der Abweichungen Δx_i an das reale System angepasst. In der Vorgeschichte der Messung muss der Prozess möglichst empfindlich auf den nachzuführenden Parameter sein. Betrachtet man noch einmal Abbildung 8.6, so wird die Wahl der Messpunkte zur Parameternachführung, wie sie Abbildung 8.8 zeigt, sofort klar.

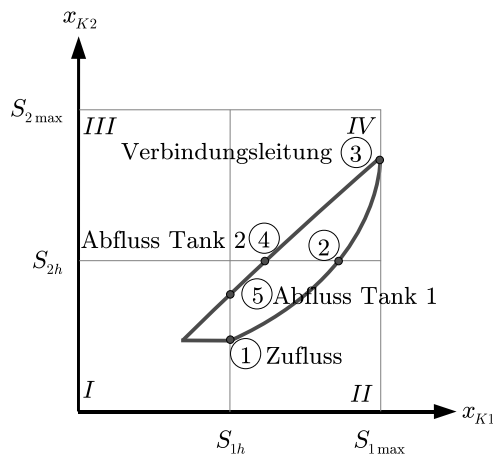


Abbildung 8.8: Zur Parameternachführung verwendete Messpunkte.

Nach dem Anfahren des Batchprozesses befindet sich dieser im Gebiet *I*. Die Abflussventile V_1 und V_2 sind geschlossen und die Pumpe eingeschaltet. Der Prozess wird von nun an bis zur Messung 1 am Sensor S_{1h} lediglich vom Pumpenparameter k_1 beeinflusst. Dies legt nahe, die Differenz Δx_{K1} von Messpunkt 1 zur Nachführung dieses Parameters zu verwenden. Geht man davon aus, dass diese Nachführung annähernd gelingt, so spielen bis zur Messung 3 am Sensor S_{1max} lediglich die Parameter k_3 und k_5 eine Rolle. Diese beiden Parameter sowie ihre Veränderung sind im vorliegenden Fall identisch, da die Querschnitte der beiden Tanks gleich groß sind. Als Ursache ihrer Veränderung kommt im Beispiel nur der Querschnitt der Verbindungsleitung zwischen den beiden Tanks in Frage. Deshalb kann einer der beiden Parameter ausgesucht und für die Nachführung durch einen PI-Algorithmus verwendet werden. Im Beispiel wird der Parameter k_3 verwendet. Nachdem der Messpunkt 3

erreicht ist, werden gleichzeitig die Abflussventile V_1 und V_2 geöffnet und die Pumpe abgestellt, so dass sich bis zur Messung 4 die Parameter k_2 , k_3 , k_4 und k_5 auf den Prozess auswirken, wobei $k_3 = k_5$ gilt. Geht man wiederum davon aus, dass die Parameternachführung von k_3 annähernd gelungen ist, so sind nur noch die Parameter der Tankabflüsse bedeutsam. Der Parameter k_4 wird mit Hilfe der Abweichung, die sich zum Zeitpunkt der Messung 4 ergibt, nachgeführt. Entsprechendes gilt für k_2 zum Zeitpunkt der Messung 5. Messung 2 wird bei diesem Anwendungsbeispiel nicht benutzt.

Leider ist es nicht ohne größere Anstrengungen möglich, dieses Verfahren stabil zu halten. Neben den verwendeten affinen Approximationen haben auch die Verstärkungsfaktoren der PI-Algorithmen Einfluss auf die Stabilität. Bei den affinen Approximationen hat sich gezeigt, dass die oben gewählten Approximationen zum Entwurf eines stabilen Diagnosesystems geeignet sind. Allerdings sind dafür sehr kleine Werte für die Verstärkungen der P- und I-Anteile des PI-Algorithmus notwendig. Die Festlegung der Verstärkungsfaktoren für den PI-Algorithmus ist nur durch gezieltes Probieren möglich, da in diesem Anwendungskontext keine systematischen Verfahren oder Einstellregeln zur Bestimmung dieser Faktoren bekannt sind.

Instabilität bedeutet in diesem Fall, dass sich die unterschiedlichen Parameter des Tank-Systems gegenseitig so beeinflussen, dass sie durch den PI-Algorithmus bis ins Unendliche oder gegen Null verändert werden. Als Folge kann der Beobachter das reale Systemverhalten nicht mehr wiedergeben.

Eine weitere mögliche Ursache für die Instabilität ist der Approximationsfehler, der dazu führen kann, dass der Zustand des nichtlinearen Modells und der Zustand des affin approximierten Modells zum Zeitpunkt der Messung auf unterschiedlichen Seiten der Messhyperebenen liegen. Geht man davon aus, dass der Modellparameter vergrößert werden müsste, wenn man dem nichtlinearen Modell folgt, so wird der Skalierungsfaktor aufgrund des Zustandes des affin approximierten Modells, der auf der anderen Seite der Messhyperebene liegt, verkleinert, also in die falsche Richtung verändert. Zusätzlich werden alle anderen Skalierungsfaktoren im weiteren Verlauf des Beobachtungsprozesses von diesem falsch nachgeführten Parameter beeinflusst. Anschaulich ist klar, dass diese geschlossene Wirkungskette instabil werden kann.

Zur Fehleridentifikation ist noch anzumerken, dass die vom PI-Algorithmus vorgenommenen Parameteränderungen nicht völlig mit den Änderungen der Parameter des realen Systems übereinstimmen. Grund dafür ist wiederum die affine Approximation der Wurzelkennlinie. Dadurch schwingen die Skalierungsfaktoren des Beobachters auch im Normalbetrieb nicht auf 100% ein, sondern weichen um wenige Prozent davon ab. Dies führt dazu, dass die Veränderung eines Parameters um 10% durch den PI-Algorithmus nicht mit einer Änderung des entsprechenden Parameters des realen Systems um 10% gleich gesetzt werden kann. Eine bessere Fehleridentifi-

kation kann über eine Kennlinie erreicht werden, die den Zusammenhang zwischen dem ermittelten Parameterwert und der Stärke des tatsächlich im System vorliegenden Fehlers wiedergibt. Eine solche Kennlinie kann durch die Simulation definierter Fehler und die Betrachtung der Reaktion der PI-Algorithmen erstellt werden. Mit Hilfe der Kennlinie kann die Fehlerstärke des isolierten Fehlers direkt aus den Modellparametern ermittelt werden, die von den PI-Algorithmen verändert wurden.

Gelingt es, dieses Verfahren stabil zu entwerfen, ist es in der Lage, Fehler zu detektieren, isolieren und eingeschränkt zu identifizieren. Die Einschränkung bei der Fehleridentifikation bedeutet, dass die Skalierungsfaktoren, die im Fehlerfall durch die PI-Algorithmen ermittelt werden, aufgrund der Interaktionen nicht verlässlich die Größenordnung der Fehler angeben. Ist die Zahl der Fehlerarten groß, steigt der Aufwand für die Parametrierung der PI-Algorithmen und die Erstellung der Kennlinien für die Fehleridentifikation. Darüber hinaus wird es mit wachsender Zahl parallel arbeitender PI-Algorithmen immer schwieriger, geeignete Werte für die Verstärkungen der PI-Algorithmen zu finden ohne dass das Verfahren instabil wird. Daher ist auch dieses Verfahren nur für einfache Systeme geeignet.

8.7.4 Mehrere stochastische Beobachter mit Parameternachführung

Hatte man beim vorangegangenen Verfahren noch mit Stabilitätsproblemen zu kämpfen, so besitzt das nun diskutierte Verfahren den Vorteil, dass die einzelnen PI-Algorithmen unabhängig voneinander arbeiten. Deshalb ist es einfacher möglich, die Verstärkungsfaktoren der PI-Algorithmen so zu wählen, dass sie stabil arbeiten. Die Anforderungen an die Linearisierung sind nicht so hoch wie zuvor. Darüber hinaus können die PI-Algorithmen so eingestellt werden, dass sie stärker auf die Eingangsgröße Δx reagieren und somit eine schnellere Parameternachführung ermöglichen. Auch ein Überschwingen der Skalierungsfaktoren kann akzeptiert werden, was beim vorangegangenen Verfahren fast zwangsläufig zu Stabilitätsproblemen führte.

Diese Vorteile erkaufte man sich jedoch durch einen erhöhten Rechenaufwand, da für jeden Parameter, der Ursache für das Auftreten eines Fehlers sein kann, ein eigener Beobachter nötig ist.

Die Tatsache, dass die einzelnen Beobachter unabhängig voneinander arbeiten, hat jedoch einen Nachteil. Jeder Beobachter versucht mit seinem Freiheitsgrad zu erreichen, dass die Differenz $E(X_i) - x_i$ zum Zeitpunkt, der für die Nachführung genutzten Messung, möglichst klein ist. Dies führt dazu, dass alle Beobachter auf einen Fehler reagieren.

Um dies zu zeigen, werden mit Hilfe des nichtlinearen Simulationsmodells die einzelnen Fehler simuliert, indem die entsprechenden Parameter des Simulationsmodells

reduziert werden. Zeitlich nacheinander werden die vier verschiedene Fehlerarten in verschiedenen Größenordnungen simuliert:

- Abfluss von Tank 1 ist verstopft,
- Abfluss von Tank 2 ist verstopft,
- Verbindungsleitung zwischen Tank 1 und Tank 2 ist verstopft,
- Zufluss ist verstopft bzw. Pumpe fördert zu wenig.

Dabei wird der Prozess zunächst für $T = 1 \cdot 10^4$ s = 10 000 s im fehlerfreien Zustand betrieben, bevor für $T = 1 \cdot 10^4$ s ein Fehler aktiviert wird. Die Größenordnungen der Fehler betragen dann nacheinander 10%, 20%, 30%, 40% und 50%. Auch diese Fehler werden jeweils für $T = 1 \cdot 10^4$ s aktiviert. Zwischen den einzelnen Fehlern wird immer wieder für $T = 1 \cdot 10^4$ s in den fehlerfreien Zustand gewechselt. Die genaue Abfolge der Fehler, die im Modell simuliert werden, und die Stärke der Fehler ist in Abbildung 8.9 wiedergegeben.

Die entsprechende Reaktion der Beobachter und der PI-Algorithmen zeigt Abbildung 8.10.

Da der Zwei-Tank-Batchprozess im fehlerfreien Betrieb gestartet wird, nehmen die Parameter in den ersten $1 \cdot 10^4$ s ihren Nominalwert an. Hier zeigt sich schon deutlich, dass die Parameter auf leicht von 100% abweichende Werte einschwingen, was auf die schon erwähnten Approximationsfehler und Modellierungsungenauigkeiten zurückzuführen ist.

Abbildung 8.10 zeigt nun, dass immer alle Beobachter auf einen simulierten Fehler reagieren. In der Zeit von $t = 1 \cdot 10^4$ s bis $t = 2 \cdot 10^4$ s wird im Simulationsmodell ein um 10% verstopfter Abfluss in Tank 1 simuliert. Die gestrichelte schwarze Kurve, die den Verlauf des dieser Fehlerart entsprechenden Skalierungsfaktors darstellt, reagiert in dem entsprechenden Zeitfenster wie gewünscht auf den simulierten Fehler, indem sie deutlich absinkt; leider reagiert der Beobachter, der den Parameter für den Zufluss zu Tank 1 nachführt, ebenfalls auf den verstopften Abfluss, was sich deutlich am Verlauf der durchgezogenen schwarzen Kurve ablesen lässt. Diese Kurve nimmt für einen verstopften Abfluss deutlich größere Werte als ihren Nominalwert an. Reduziert sich der Abflussquerschnitt in Tank 1 um mehr als 10%, so nimmt der Skalierungsfaktor für den Zufluss zu Tank 1 über weite Strecken Funktionswerte von über 150% an und kann daher im Graphen teilweise nicht dargestellt werden.

Dieses Verhalten lässt sich leicht erklären: Der Beobachter, der den Parameter für den Zufluss nachführt, berechnet zu dem Zeitpunkt, zu dem die Differenz $E(X_{K1}) - x_{K1}$ gebildet wird, einen geringeren Füllstand in Tank 1 als er im zu überwachenden System vorhanden ist. Diese Abweichung wird zwar von dem verstopften Abfluss in Tank 1 verursacht, allerdings kann der Beobachter für den Zufluss auch nur den

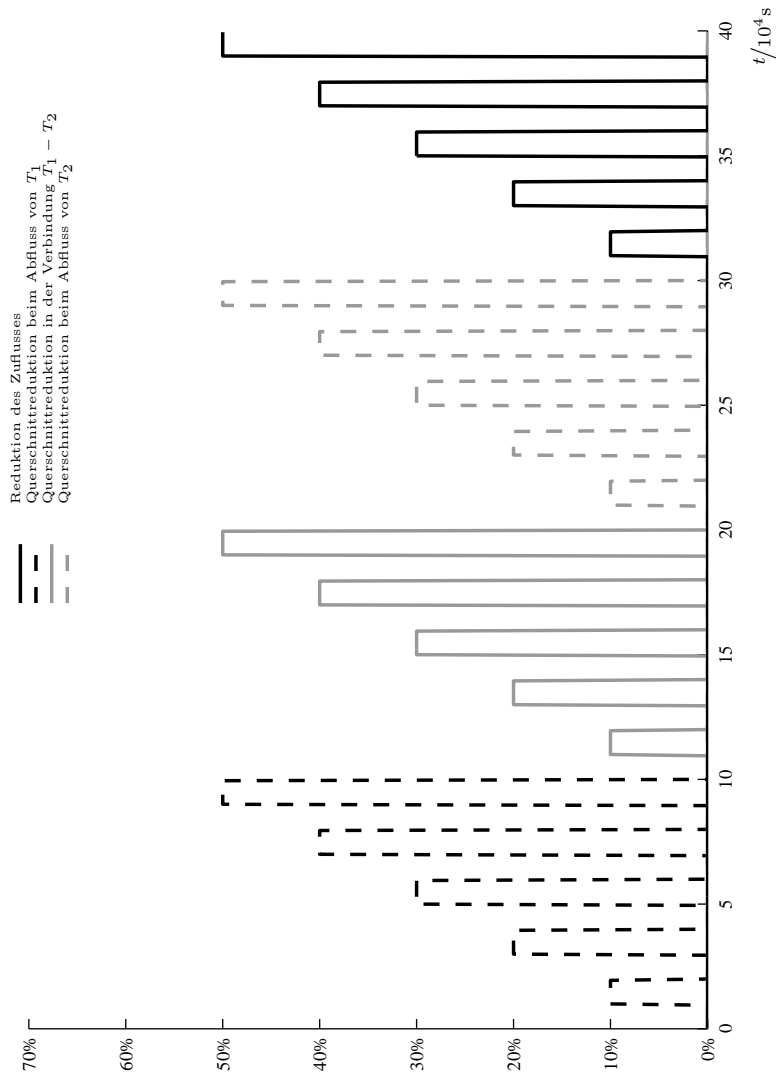


Abbildung 8.9: Abfolge und Stärke der Fehler, die im Modell des Zwei-Tank-Systems simuliert werden.

Parameter für den Zufluss verändern. Deshalb versucht der PI-Algorithmus dieses Beobachters die festgestellte Abweichung durch eine Vergrößerung des Parameters für den Zufluss zu kompensieren.

Auch wenn dieses Verhalten gut interpretieren lässt, erschwert es die Fehlerisolierung, da allein anhand der Skalierungsfaktoren ohne zusätzliches Wissen kein Fehler ausgeschlossen werden kann.

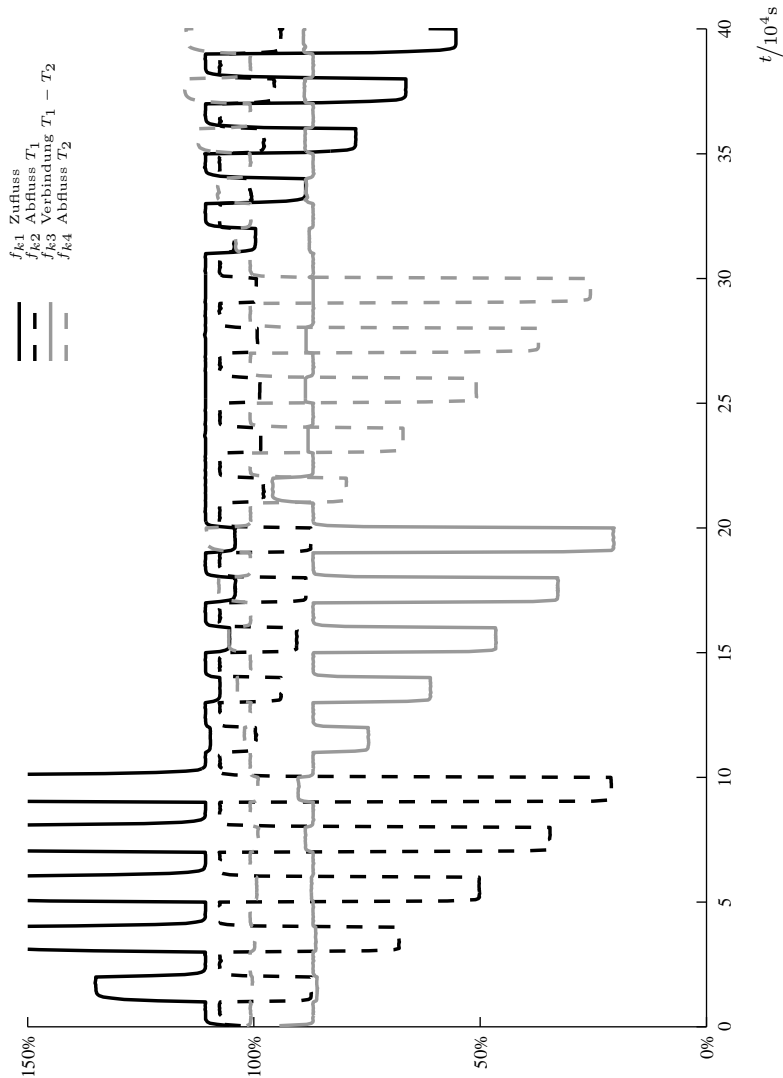


Abbildung 8.10: Simuliertes Verhalten der Skalierungsfaktoren bei den Fehlern, die in Abbildung 8.9 gezeigt sind.

Bei der nächsten Fehlerart, die von $t = 10 \cdot 10^4$ s bis $t = 20 \cdot 10^4$ s simuliert wird, handelt es sich um die Verkleinerung des Querschnitts der Verbindungsleitung. Auch hier wird der entsprechende Skalierungsfaktor korrekt nachgeführt, wie an der grauen durchgezogenen Linie in Abbildung 8.10 zu sehen ist. Die restlichen Beobachter reagieren zwar nicht mit der gleichen Stärke auf diesen Fehler, jedoch erschwert auch deren Reaktion die Fehlerisolierung. Ausschließlich anhand der Skalierungsfaktoren

kann kein Fehler ausgeschlossen werden, da alle Faktoren auf den Fehler reagieren.

Auch bei den restlichen beiden Fehlern, die simuliert werden, ergibt sich ein ähnliches Bild: Der Skalierungsfaktor, der den wirklichen Fehler nachbildet, reagiert am stärksten auf den simulierten Fehler; jedoch gibt es auch andere Skalierungsfaktoren, die nennenswert von ihrem Nominalparameter abweichen, so dass für die Fehlerisolierung auch hier zusätzliches Wissen notwendig ist.

8.7.4.1 Validierung am Laborsystem

Das simulierte Verhalten stimmt grundsätzlich mit dem Verhalten des Zwei-Tank-Laborsystems überein. Die Abbildungen 8.11 bis 8.14 zeigen die durch Messung am realen System gewonnenen Daten. Dabei wurde für 5000 s jeweils mit einem Stopfen der Querschnitt des Abflusses von Tank 1 verstopft. Die Größenordnungen der Verstopfungen von Tank 1 sind nicht mit denen in der Simulation identisch, da sich die verwendeten Stopfen zur Querschnittsverkleinerung anders verhalten; so verursacht der Stopfen A eine Reduktion des wirksamen Querschnitts um ca. 28%. Die Durchmesser und die Geometrie der verwendeten Stopfen sowie ihre Wirkung auf das Laborsystem sind in Anhang C zusammengestellt.

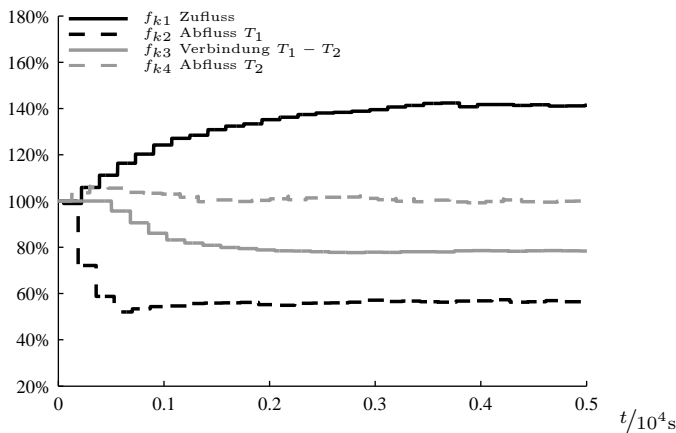


Abbildung 8.11: Messergebnis bei verstopftem Abfluss von Tank 1 am Zwei-Tank-Laborsystem (Stopfen A)

In Abbildung 8.11 ist zu sehen, wie die Skalierungsfaktoren der einzelnen Beobachter auf den mit Stopfen A verkleinerten Abflussquerschnitt reagieren. Wie bereits erläutert wurde, reagiert nicht nur der Beobachter für den tatsächlich im System vorliegenden Fehler auf den verkleinerten Abfluss, sondern auch die anderen Beobachter.

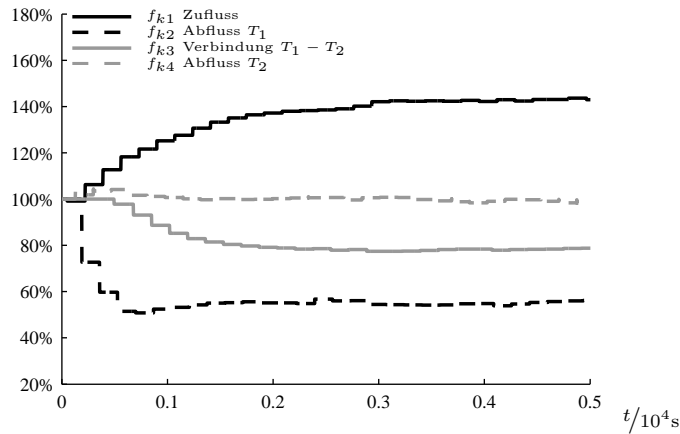


Abbildung 8.12: Messergebnis bei verstopftem Abfluss von Tank 1 am Zwei-Tank-Laborsystem (Stopfen B)

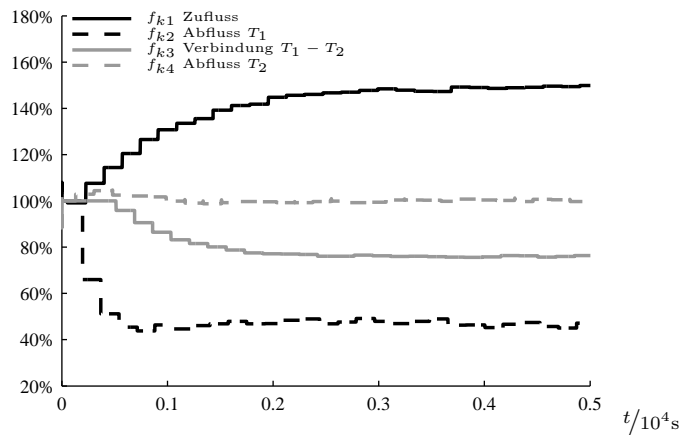


Abbildung 8.13: Messergebnis bei verstopftem Abfluss von Tank 1 am Zwei-Tank-Laborsystem (Stopfen C)

Dieses Verhalten ist qualitativ bei allen Versuchen mit den verschiedenen Stopfen gleich. Die unterschiedlich starke Abnahme des Skalierungsfaktors f_{k2} ist auf die unterschiedlich großen Querschnittsflächen der Bohrungen in den Stopfen zurückzuführen. Da sich die Querschnittsflächen der Bohrungen von Stopfen A und B nur sehr wenig unterscheiden, ist das Messergebnis dieser Versuche auch sehr ähnlich.

Zu dem Verfahren ist anzumerken, dass die Fehlerdetektion, wie an den Beispielen gezeigt, problemlos funktioniert, da die PI-Algorithmen auf Abweichungen der berechneten von den gemessenen Schaltzeitpunkten genau so wie in der Simulation reagieren.

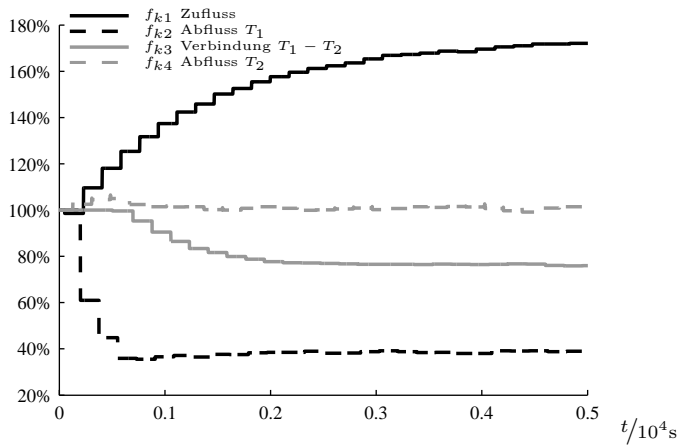


Abbildung 8.14: Messergebnis bei verstopfem Abfluss von Tank 1 am Zwei-Tank-Laborsystem (Stopfen D)

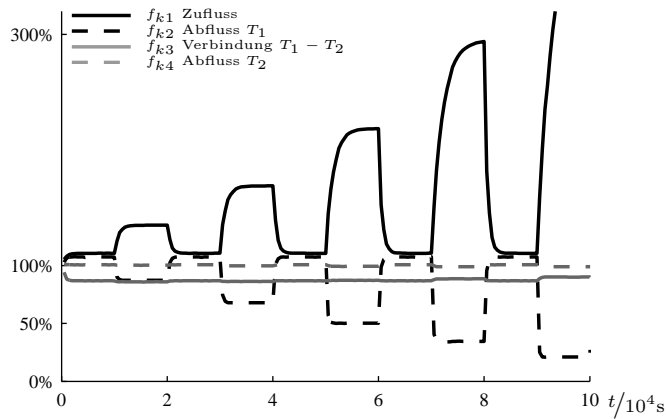
Die Fehlerisolierung ist im Falle des Zwei-Tank-Systems mit diesem Verfahren ebenfalls möglich, wenn man sich darauf beschränkt, dass im Falle des Zwei-Tank-Systems nur Parameterverkleinerungen, also Verstopfungen, als Fehler auftreten können. Dadurch können Modellparameter, die durch die PI-Algorithmen vergrößert werden, von vorne herein als Fehlerursache ausgeschlossen werden. Anhand von Abbildung 8.10 lässt sich nachvollziehen, dass genau der Skalierungsfaktor am stärksten auf den Fehler reagiert, dessen Parameter im Simulationsmodell verändert wurde, um einen Fehler zu simulieren. Mit diesem Kriterium kann für das Zwei-Tank-System eine Fehlerisolierung durchgeführt werden.

Für eine genaue Fehleridentifikation ist anschließend noch die im vorangegangenen Abschnitt erwähnte Kennlinie sinnvoll, mit der die durch die PI-Algorithmen ermittelten Parameter auf die tatsächliche Stärke der Fehler abgebildet werden. Somit ermöglicht dieses Verfahren für das Zwei-Tank-System eine vollständige Diagnose mit Fehlerdetektion, Fehlerisolierung und Fehleridentifikation.

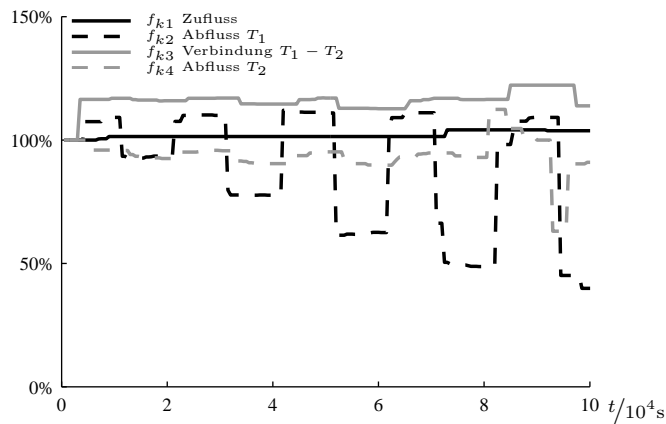
8.7.5 Mehrere stochastische Beobachter mit Parameternachführung und Gütekriterium

Der Versuch, den Nachteil der eingeschränkten Fehlerisolierung, wie er bei der Diagnose mit mehreren stochastischen Beobachtern mit PI-Algorithmus auftritt, mittels eines Gütekriteriums zu beheben, gelingt im Falle des Zwei-Tank-Batchprozesses vollständig.

Dazu werden in Abbildung 8.15 das Verfahren ohne Gütemaß und das Verfahren mit



(a) ohne Gütekriterium



(b) mit Gütekriterium

Abbildung 8.15: Vergleich der Parameter k_i bei der Simulation eines verstopften Abflusses von Tank 1 ohne (a) und mit (b) Verwendung des Gütekriteriums

Gütemaß einander gegenübergestellt. In Abbildung 8.15 (a) sind die ersten 100 000 s aus Abbildung 8.10 noch einmal detaillierter und in einem veränderten Maßstab abgebildet. Hier wird noch einmal die Reaktion des Skalierungsfaktors für den Zufluss auf den Fehler im Abfluss des ersten Tanks deutlich. Weiterhin ist es bei diesem Ansatz nicht ohne weiteres möglich, zu entscheiden, welcher Skalierungsfaktor den Fehler des realen Systems abbildet. Anders verhalten sich die Skalierungsfaktoren bei der Anwendung des Gütemaßes, wie es in Abbildung 8.15 (b) zu sehen ist. Bei dieser Simulation wurde die gleiche Fehlersequenz verwendet, wie sie in Abbildung 8.9 bereits dargestellt ist.

Vor dem Auftreten des ersten Fehlers bei $t = 1 \cdot 10^4$ s ist zu sehen, wie die einzel-

nen Skalierungsfaktoren ihren Wert für den fehlerfreien Fall annehmen; dieser Wert weicht auch hier aufgrund der affinen Approximation von 100% ab. Hierbei wird zur Entscheidung, welcher Parameter verändert werden sollte, das Gütemaß angewandt. Nach dem Auftreten des Fehlers wird als erstes der Parameter für den Abfluss von Tank 1 deutlich verringert, wohingegen die anderen Parameter später und nur geringfügig verändert werden. Durch diese Vorgehensweise wird eine Kombination von Skalierungsfaktoren ermittelt, die das aktuelle Systemverhalten sehr gut beschreibt.

Die Verläufe in Abbildung 8.15 (b) sind viel kantiger als die in (a) gezeigten Verläufe. Dies ist darauf zurückzuführen, dass in (a) jeweils der aktuelle Wert der PI-Algorithmien aufgezeichnet ist, während in (b) gezeigt ist, welche Skalierungsfaktoren nach der Überprüfung durch das Gütemaß in allen Beobachtern als neue Nominalparameter verwendet werden; der Einschwingvorgang der PI-Algorithmien ist also in (b) nicht zu sehen, da erst nach dem Einschwingen der PI-Algorithmien eine Bewertung durch das Gütemaß vorgenommen wird. Mit Hilfe der aktuell gültigen Verstärkungsfaktoren und einer Kennlinie für die Fehleridentifikation ist die Fehlerisolation und die Fehleridentifikation beim Zwei-Tank-System möglich.

Für die Einstellung der maximalen Einschwingzeiten der Skalierungsfaktoren oder der minimalen Änderung der Skalierungsfaktoren, die zum Abbruch des Einschwingvorgangs führen, können keine allgemeinen Einstellregeln angegeben werden. Hier wurden die maximale Einschwingzeit und die minimale Änderung durch gezieltes Probieren eingestellt. Ein Nachteil dieser Vorgehensweise mit Gütemaß zur Unterdrückung der Nachführung falscher Parameter ist die größere benötigte Rechenleistung, da zusätzlich die Gütekriterien berechnet werden müssen. Dem gegenüber steht die Möglichkeit, die Fehlerisolation und die Fehleridentifikation systematisch durchführen zu können.

8.8 Vergleich der Verfahren

In Tabelle 8.8 werden einige der auf den Zwei-Tank-Batchprozess angewandten Diagnoseverfahren anhand ausgewählter Stichpunkte verglichen. Dabei ist das Verfahren aus Abschnitt 8.7.3 nicht in der Tabelle zu finden, da es nur sehr schwierig so entworfen werden kann, dass es stabil arbeitet. Von den aufgeführten Verfahren ist das Verfahren mit mehreren Beobachtern, PI-Algorithmien und Gütemaß das leistungsfähigste, da damit systematisch ein vollständiges Diagnosesystem für die betrachtete Systemklasse erstellt werden kann. Die übrigen Verfahren bieten nicht die Möglichkeit einer systematischen Vorgehensweise, um ein Diagnosesystem zu erstellen, da entweder die Fehlerisolation oder die Fehleridentifikation nicht oder nur eingeschränkt möglich ist. Trotzdem haben die Verfahren ihre Berechtigung, da sie bei weniger komplexen Systemen durchaus eingesetzt werden können und einfacher zu implementieren sind.

Tabelle 8.1: Vergleich der Diagnoseverfahren

Verfahren	einzelner HSB Abschnitt 8.7.1	mehrere HSB Abschnitt 8.7.2	mehrere HSB mit PI Abschnitt 8.7.2	mehrere HSB mit und Gütekriterium Abschnitt 8.7.5
Fehler- detektion	möglich	möglich	möglich	möglich
grobe	möglich	möglich	möglich	möglich
schleichende	mit Trendanalyse	mit Trendanalyse	möglich	möglich
mehrfache	nicht möglich	eingeschränkt mit zu- sätzlichen Beobachtern	möglich	möglich
Fehler- isolation	nicht möglich	mit einem Beobachter je Fehlerart	möglich	möglich
Fehler- identifikation	nicht möglich	eingeschränkt mit meh- reren Beobachtern je Fehlerart und verschie- dener Fehlerausprägung	mit Kennlinie für die Fehleridentifikation	mit Kennlinie für die Fehleridentifikation
Vorteile	einfache Implementierung, geringe Rechenleistung	für wenige Fehlerarten einfache Implementie- rung	robustes Verfahren	guter Kompromiss zwi- schen Rechenleistung und Diagnosemöglich- keiten
Nachteile	sehr eingeschränkte Diag- nosemöglichkeiten	häufige Fehl- und Falsch- alarme, sehr hohe Anforderun- gen an Rechenleistung, große Anzahl von Beob- achtern nötig	hoher Parametrierungs- aufwand für die PI-Algo- rithmen, keine Fehlerisolation und Fehleridentifikation	hoher Parametrierungs- aufwand für die PI-Al- gorithmen und die maxi- malen Einschwingzeiten

Zusammenfassung

Die Aufgabe der Diagnose nichtlinearer und hybrider Systeme mit diskreter Sensorik ist von herkömmlichen Diagnoseverfahren, die auf kontinuierliche Sensorik angewiesen sind, nicht zu lösen. Es steht zu wenig Messinformation zur Verfügung, um die in den meisten Verfahren notwendigen Vergleiche zwischen den Modellinformationen und den gemessenen Informationen vornehmen zu können.

Der hybride stochastische Beobachter bietet die Möglichkeit, diese Aufgabe zu bewältigen. Allerdings ist es aufgrund der fehlenden Messinformation notwendig, den Prozesszustand mit Hilfe von Modellen und Prozesswissen zu rekonstruieren. Hierzu müssen mehrere hybride stochastische Beobachter parallel eingesetzt und um zusätzliche Funktionen erweitert werden. Die Verwendung von PI-Algorithmen ermöglicht es, die durch die Fehler beeinflussten Parameter des Prozessmodells an den veränderten Prozess anzupassen und somit eine Fehlerdetektion vorzunehmen. Die Fehlerisolation kann anschließend erfolgen, indem die Parameterveränderungen, die durch die PI-Algorithmen vorgenommen werden, mit Hilfe von Gütemaßen bewertet werden. Die Fehleridentifikation erfolgt schließlich, indem die nachgeführten Parameter mit Hilfe einer Kennlinie ausgewertet werden.

Dabei bleibt der Rechenaufwand relativ klein, da nur für jede Fehlerart ein eigener Beobachter notwendig ist.

Kapitel 9

Zusammenfassung

Ziel der vorliegenden Arbeit war es, Methoden zur *Identifikation* und *Diagnose hybrider Systeme* zu entwickeln. Die betrachtete Kombination automatisierungstechnischer Aufgaben erscheint zunächst willkürlich; jedoch ist die Diagnose stark von einer zuverlässigen Modellbildung abhängig. Die erstellten Modelle müssen ausreichend genau für die Diagnose sein, so dass Prädiktionen über größere Zeiträume möglich sind; gleichzeitig muss darauf geachtet werden, dass die Komplexität der Modelle möglichst klein bleibt, da ansonsten die Diagnose erschwert oder sogar praktisch unmöglich wird.

Deshalb wurden in dieser Arbeit – nach der Einleitung und einigen systemtheoretischen Grundlagen in Kapitel 2 – zunächst zwei Verfahren vorgestellt, die zur Identifikation hybrider Systeme eingesetzt werden können. Da die Integration von a priori Wissen in diesen Verfahren möglich ist, können diese Verfahren auch zur Unterstützung einer Modellbildung herangezogen werden, die teilweise theoretisch und teilweise experimentell vorgeht.

Für die *Identifikation hybrider Systeme* wurden in Kapitel 3 zunächst einige Grundlagen eingeführt und das Prinzip der Methode der kleinsten Quadrate erläutert. Zur Auswahl einer geeigneten Modellform für die Identifikation wurden die hybriden Systeme und ihr Verhalten im Hinblick auf die Identifikation untersucht. Mit der Betrachtung einer geeigneten Modellform für die Identifikation hybrider Systeme sowie den in diesem Zusammenhang diskutierten Grundlagen konnten anschließend geeignete Identifikationsalgorithmen für stückweise affine Systeme entwickelt werden.

In Kapitel 4 wurden zunächst die besonderen Probleme der hybriden Identifikationsaufgabe erläutert. Anschließend wurden zwei Verfahren zur Identifikation hybrider Systeme vorgestellt. Das erste der beiden Verfahren – die lokale Kennfeldoptimierung – besitzt eine sehr anschauliche Vorgehensweise, bei der iterativ eine Anfangslösung verbessert wird. Dies geschieht dadurch, dass die Grenzen der Gültigkeitsbereiche der einzelnen Teilmodelle mit Hilfe von Gradienteninformationen sukzessive verscho-

ben werden. Diese Gradienteninformation steht aufgrund der Approximation von Sprungfunktionen durch stetig differenzierbare Sigmoidfunktionen zur Verfügung.

Das zweite Verfahren besitzt seine Stärke bei der Identifikation größerer Modelle, da es weniger Rechenzeit in Anspruch nimmt als das Verfahren der lokalen Kennfeldoptimierung. Zudem ist bei dieser Vorgehensweise garantiert, dass ein globales Optimum für das Identifikationsproblem gefunden wird. Es ist nicht ganz so anschaulich wie das der lokalen Kennfeldoptimierung, wodurch das Einbringen von Vorwissen etwas erschwert ist. Für das hybride Identifikationsproblem konnte gezeigt werden, dass mehrere äquivalente Formulierungen existieren, zwischen denen – je nach Anwendung und Leistungsfähigkeit der zur Verfügung stehenden Optimierungsalgorithmen – ausgewählt werden kann. Weiterhin ist das Verfahren robust gegenüber Messrauschen. Der Zusammenhang der Geltungsbereiche der einzelnen Teilmodelle wird durch einen heuristischen Algorithmus sichergestellt.

Anhand der Anwendungsbeispiele in Kapitel 5 wurde deutlich, dass mit den beiden Methoden zwei leistungsfähige Verfahren zur Identifikation von hybriden Systemen zur Verfügung stehen. Diese Beispiele runden den ersten Teil der Arbeit ab.

Auch für die *Diagnose hybrider Systeme* musste zunächst das erforderliche theoretische Fundament in Kapitel 6 gelegt werden. Da die hier betrachteten Diagnosemethoden mit ausschließlich diskreten Messungen arbeiten sollten, bestand das erste Etappenziel im zweiten Teil der Arbeit in der Ermittlung des Systemzustandes aufgrund der diskreten Messungen. Dazu wurde ein hybrider stochastischer Beobachter vorgestellt.

Darauf aufbauend wurden in Kapitel 7 Diagnosealgorithmen entwickelt, um über die Möglichkeit der Fehlerdetektion hinaus auch eine Fehlerisolation und eine Fehleridentifikation auf der Grundlage des hybriden stochastischen Beobachters vorzunehmen. Die prinzipielle Funktionsweise der Fehlerisolation und der Fehleridentifikation wurde in diesem Kapitel erläutert.

Um das Ziel der vollständigen Fehleridentifikation mit möglichst wenigen Beobachtern zu erreichen, kamen schließlich mehrere parallel arbeitende Beobachter zum Einsatz, die jeweils die Möglichkeit besitzen, einen Parameter über einen PI-Algorithmus an das Verhalten des Prozesses anzupassen. Um zu entscheiden, welcher Beobachter das Prozessverhalten am besten beschreibt, wurde ein Gütekriterium eingeführt, das anhand der diskreten Messungen beurteilt, wie gut jeder Beobachter mit dem realen Prozess übereinstimmt. Mit Hilfe dieses Gütekriteriums gelang es, die Veränderung der Parameter, die den aktuellen Fehler des Systems nicht nachbilden können, zu unterdrücken. Dadurch wurde eine zuverlässige Fehlerisolation ermöglicht.

Anhand des nichtlinearen SIMULINK-Modells des Zwei-Tank-Systems wurden in Kapitel 8 die Ansätze validiert. An der Laboranlage des Zwei-Tank-Systems wurden zu-

sätzlich einige Messungen vorgenommen, um zu zeigen, dass das simulierte und das reale Verhalten miteinander übereinstimmen. Um diese Tests zu ermöglichen, wurden Stopfen gefertigt, die die Simulation von Fehlern an der Zwei-Tank-Laboranlage ermöglichen.

Da bei diesem Diagnoseansatz nur mit diskreten Messungen gearbeitet wurde, konnten keine aus der Literatur bekannten Methoden zur Parameteridentifikation zum Einsatz kommen. Dies ist auf die Tatsache zurückzuführen, dass die Information über den realen Prozess lediglich als diskrete Messung mit zeitlich nicht äquidistanten Abtastschritten vorliegt. Um so höher ist zu bewerten, dass mit dem vorgeschlagenen Verfahren die vollständige Diagnose des Zwei-Tank-Systems mit Fehlerdetektion, Fehlerisolation und Fehleridentifikation gelang.

Mit den Methoden, die in dieser Arbeit entwickelt wurden, stehen für den Anwender neue leistungsfähige Werkzeuge bereit, die bei der Modellbildung und bei der Diagnose von hybriden dynamischen Prozessen eingesetzt werden können.

Anhang A

Grundbegriffe zur Graphentheorie

Bei der Einführung der Petri-Netze in Kapitel 2 und zur Aufteilung des Identifikationsraumes in zusammenhängende Gebiete in Kapitel 4 werden einige Grundlagen der Graphentheorie benötigt. Deshalb sollen die dort verwendeten Begriffe an dieser Stelle eingeführt und erläutert werden. Die Begriffe und Definitionen in diesem Anhang sind [NM02] und [Bau96] entnommen.

Zunächst wird definiert, was unter einem Graphen verstanden werden soll:

Definition A.1: Graph

Ein Graph G besteht aus einer Menge E und einer nicht leeren Menge V mit $V \cap E = \emptyset$. Auf E ist eine Abbildung h definiert, die jedem $e \in E$ genau ein Paar von Elementen $v \in V$ zuordnet. Ist das jedem e zugewiesene Paar ungeordnet, so heißt der Graph G *ungerichteter Graph*; ist das zugewiesene Paar geordnet, so spricht man von einem *gerichteten Graphen*. Die Elemente von V bezeichnet man als *Knoten*; die Elemente von E werden bei einem ungerichteten Graphen *Kanten* und in einem gerichteten Graphen *gerichtete Kanten* genannt.

Die Bezeichnung „ungerichtet“ wird meist weggelassen. Nur gerichtete Graphen werden explizit gekennzeichnet.

Bei einem ungerichteten Graphen bezeichnet man die Kanten mit $e = (v, w)$, wobei v und w Elemente der Menge V sind. Diese Knoten bezeichnet man auch als Endknoten von e . Bei einem gerichteten Graphen bezeichnet man die gerichteten Kanten mit $e = [v, w]$. Hier wird v als Startknoten und w als Endknoten von e bezeichnet.

In einem ungerichteten Graphen G heißt ein Knoten v *Nachbar* zu w , wenn eine Kante (v, w) in G existiert. Die Menge der Nachbarn des Knotens v wird mit $N(v)$ bezeichnet.

Ein Knoten v in einem gerichteten Graphen G heißt *Nachfolger* von w , wenn es in G eine gerichtete Kante $[w, v]$ gibt. Umgekehrt heißt w *Vorgänger* von v , falls die gerichtete Kante $[w, v]$ existiert. Die Menge der Nachfolger von v bezeichnet man mit $\mathcal{N}(v)$, die Menge der Vorgänger mit $\mathcal{V}(v)$. In gerichteten Graphen heißen die Vorgänger und die Nachfolger eines Knotens v Nachbarn von v .

Um nun den Begriff des *zusammenhängenden Graphen* bzw. der *zusammenhängenden Komponenten* zu definieren, werden zunächst die Begriffe *Kantenfolge*, *Kantenzug*, *Kette* und *Kreis* benötigt.

Definition A.2: Kantenfolge

Sei $G = (V, E)$ ein ungerichteter Graph. Eine Menge $\hat{E} = \{e_1, \dots, e_k\}$, deren Elemente auch Elemente der Kantenmenge E sind, heißt *Kantenfolge* von G , wenn eine Menge $\hat{V} = \{v_0, \dots, v_k\}$ mit Knoten aus V existiert, so dass $e_i = (v_{i-1}, v_i) \forall i \in \{1, \dots, k\}$ gilt.

Ist $v_0 = v_k$, so heißt die Kantenfolge *geschlossen*, andernfalls *offen*. v_0 heißt *Anfangsknoten* und v_k heißt *Endknoten* der Kantenfolge.

Definition A.3: Kantenzug, Kette, Kreis

Eine Kantenfolge mit lauter verschiedenen Kanten heißt *Kantenzug*.

Eine offene Kantenfolge, in der jeder Knoten nur einmal vorkommt, heißt *Kette*.

Ein *Kreis* ist eine geschlossene Kantenfolge, in der jeder Knoten, außer dem Anfangs- und Endknoten, nur einmal vorkommt.

Definition A.4: Verbundene Knoten, zusammenhängender Graph

Zwei Knoten v und w eines Graphen heißen *verbunden*, wenn es eine Kantenfolge mit den Anfangs- und Endknoten v und w gibt.

Ein Graph $G = (V, E)$ heißt *zusammenhängend*, wenn alle Paare aus V verbunden sind.

Ein nicht zusammenhängender Graph kann in *Komponenten* unterteilt werden.

Definition A.5: Komponenten

Die zusammenhängenden Teilgraphen eines nicht zusammenhängenden Graphen G werden *Komponenten* des Graphen genannt.

Je nach Anwendungskontext können den Kanten (gerichtet oder ungerichtet) und den Knoten des Graphen Gewichte zugeordnet werden.

Im Zusammenhang mit Petri-Netzen taucht der Begriff der *bipartiten Graphen* auf:

Definition A.6: Bipartiter Graph

Ein gerichteter Graph G heißt bipartit, wenn die Knotenmenge V so in zwei Teilmengen V_1 und V_2 zerlegt werden kann, dass jede gerichtete Kante einen Knoten aus V_1 mit einem Knoten aus V_2 verbindet. Dabei ist der Schnitt der beiden Teilmengen leer.

In Petri-Netzen entsprechen die beiden Knotenmengen V_1 und V_2 den Stellen und den Transitionen.

Damit sind die in dieser Arbeit benötigten Grundlagen aus der Graphentheorie eingeführt.

Anhang B

Delaunay-Netze

B.1 Delaunay-Triangulierung und Simplexe

Die Delaunay-Triangulierung stellt eine weit verbreitete Methode zur Beschreibung von unregelmäßigen Oberflächen dar. Diese aus der geometrischen Datenverarbeitung stammende Methode zur Beschreibung von dreidimensionalen Gebilden lässt sich aber auch ohne Probleme auf höherdimensionale Räume übertragen [Ede87].

Die Delaunay-Triangulierung teilt eine Menge gegebener Punkte im n -dimensionalen euklidischen Raum in bestimmter Weise in Simplexe auf, wobei ein Simplex ein Polygon ist, das aus $n + 1$ Ecken besteht. Bei einer gegebenen Menge von Punkten wird durch die Delaunay-Triangulierung jeder Punkt mit seinen natürlichen Nachbarn verbunden. Im zweidimensionalen Fall entspricht ein Simplex einem Dreieck.

Eng verwandt mit der Delaunay-Triangulierung ist das Voronoi-Diagramm, das ebenfalls aus der geometrischen Datenverarbeitung stammt. Das Voronoi-Diagramm einer Menge S , die aus N Punkten besteht, ist eine Zerlegung des Raumes in N Regionen. Jede Voronoi-Region $VR(\underline{p}, S)$ eines Punktes \underline{p} aus S enthält genau die Punkte, die näher zu \underline{p} als zu irgendeinem anderen Punkt in S sind. Wenn als Distanz von zwei Punkten die Länge der geraden Verbindung der Punkte genommen wird, dann resultiert das Voronoi-Diagramm unter der euklidischen Metrik. Die Gewinnung des Voronoi-Diagramms kann anschaulich als ein „Aufblähen“ der Punkte \underline{p} verstanden werden, bis jeder dieser „Luftballons“ an die Grenzen aller anderen Voronoi-Regionen stößt.

Die Verbindung zwischen Voronoi-Diagramm und Delaunay-Triangulierung ist im zweidimensionalen Fall sehr anschaulich und kommt über den Umkreis der Simplexe zustande. Der Mittelpunkt jedes Umkreises liegt dabei immer auf der Ecke eines Voronoi-Polygons. Dadurch ist sicher gestellt, dass der Umkreis jedes Simplex keine weiteren Punkte aus der Punktmenge S enthält. Dieses Kriterium bleibt auch für höherdimensionale Delaunay-Netze erhalten. Es gilt, dass ein gültiges Delaunay-

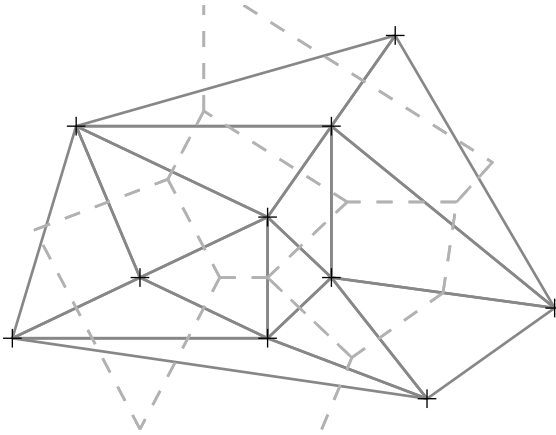


Abbildung B.1: Eine Delaunay-Triangulierung (durchgezogene Linien) mit dem korrespondierenden Voronoi-Diagramm (gestrichelte Linien) einer Punktmenge (+)

Netz aus nicht überlappenden Simplexen besteht und dass kein Punkt innerhalb der Umhyperkugel eines Simplex liegt. Diese Zusammenhänge sind in Abbildung B.1 veranschaulicht.

Eine für diese Arbeit sehr wichtige Eigenschaft der Delaunay-Triangulierung ist, dass sie keine lang gezogenen und schmalen Simplexe liefert. Bei der Identifikation ist dies von Bedeutung, da schmale Bereiche mit höherer Wahrscheinlichkeit keine bzw. nur wenige Datenpunkte enthalten, was dazu führen kann, dass innerhalb eines Simplex kein Modell geschätzt werden kann. Außerdem ist bei solchen lang gezogenen Simplexen die Wahrscheinlichkeit, dass eine Triggerrandmenge den Simplex schneidet höher als bei kompakt geformten Simplexen.

Um bei der Identifikation die Datenpunkte den einzelnen Gebieten zuzuordnen, ist ein Formalismus nötig, der die Entscheidung trifft, ob sich ein Datenpunkt innerhalb oder außerhalb eines Simplex befindet. Hierzu bieten sich die baryzentrischen Koordinaten an.

B.2 Baryzentrische Koordinaten

Baryzentrische Koordinaten eignen sich sehr gut dazu, Punkte im Inneren eines Simplex zu parametrisieren [HL92]. Dies erfolgt über eine lineare Koordinatentransformation, die das Innere eines Simplex auf einen Halbraum abbildet [BS91, HL92].

Definition B.1: Baryzentrische Koordinaten

Unter den Baryzentrischen Koordinaten eines Punktes $\underline{x} \in \mathbb{R}^n$ in Relation zu einem Simplex T_k versteht man die $(n + 1)$ Gewichte b_i , mit denen die Ortsvektoren der Ecken $\underline{p}_1, \dots, \underline{p}_{n+1}$ von T_k versehen werden müssen, damit ihr Schwerpunkt mit dem Punkt \underline{x} zusammenfällt. Dazu muss das Gleichungssystem

$$\sum_{i=1}^{n+1} b_i \cdot \underline{p}_i = \underline{x}$$

gelöst werden. Dieses Gleichungssystem besitzt $n + 1$ Unbekannte und n Gleichungen. Um nun eine eindeutige Lösung zu erhalten, muss zusätzlich

$$\sum_{i=1}^{n+1} b_i = 1$$

gelten [UT96].

Als Folge dieser Definition ergibt sich eine Eigenschaft, die in dieser Arbeit häufig verwendet wird [UT96].

Satz B.1: Zugehörigkeit zum Innenraum eines Simplex

Ein Punkt $\underline{x} \in \mathbb{R}^n$ befindet sich genau dann innerhalb der konvexen Hülle des Simplex T_k , wenn sämtliche baryzentrischen Koordinaten b_i positiv sind: $b_i > 0, i \in (1, \dots, n + 1)$.

Der Punkt \underline{x} liegt genau dann außerhalb der konvexen Hülle von T_k , wenn mindestens eine baryzentrische Koordinate b_i negativ ist.

Zum besseren Verständnis wird an folgendem Beispiel der zweidimensionale Fall vorgestellt:

Beispiel B.1

Es sei $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ ein Punkt in der Ebene \mathbb{R}^2 und $\begin{pmatrix} p_{11} \\ p_{21} \end{pmatrix}, \begin{pmatrix} p_{12} \\ p_{22} \end{pmatrix}, \begin{pmatrix} p_{13} \\ p_{23} \end{pmatrix}$ die Ecken eines Dreiecks. Die Baryzentrischen Koordinaten $\underline{x}_{\text{bar}}$ können dann aus dem Gleichungssystem

$$\begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ 1 & 1 & 1 \end{pmatrix} \underline{x}_{\text{bar}} = \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}$$

berechnet werden.

Anhang C

Generieren von Fehlern am Zwei-Tank-Laborsystem

Um die in der Simulation gewonnen Erkenntnisse und die entwickelten Diagnoseverfahren am realen Zwei-Tank-Laborsystem verifizieren zu können, muss eine Möglichkeit geschaffen werden, am Zwei-Tank-Laborsystem reproduzierbare Fehler zu simulieren. Fehler, die eine verringerte Förderleistung der Pumpe, die Tank 1 befüllt, betreffen, lassen sich einfach durch eine entsprechende Einstellung in der Steuerungssoftware für die Pumpe erzeugen. Schwieriger ist es, die Querschnitte des Durchflusses zwischen den beiden Tanks und die der Abflüsse reproduzierbar zu beeinflussen.

Zu diesem Zweck wurden insgesamt sechs verschiedene Stopfen gefertigt und ihr Verhalten am System identifiziert. Es handelt sich um zwei einschraubbare Stopfen für eine Verengung des Durchflussquerschnittes und vier Stopfen, die in die Abflussöffnung eines jeden Tanks gesteckt werden können, um diese zu verkleinern.

C.1 Geometrische Beschreibung

Der Boden der beiden Tanks des Zwei-Tank-Laborsystems ist wie in Abbildung C.1 beschaffen. Senkrecht im Tankboden befindet sich eine Bohrung mit 8 mm Durchmesser. Im rechten Winkel zu dieser befindet sich eine weitere Bohrung mit 10 mm Durchmesser, die über ein Leitungsstück mit dem Abflussventil verbunden ist.

Die vier Stopfen für die Tankabflüsse haben die Gestalt wie in Abbildung C.2 gezeigt und besitzen unterschiedliche Innendurchmesser d . Die Stopfen sind konisch gefertigt und mit einem Kunststoff-Schrumpfschlauch überzogen, um einen sicheren Halt in den Abflussbohrungen des Zwei-Tank-Systems zu gewährleisten. Sie bestehen aus V2A-Stahl und sind somit nicht rostend. Die Innendurchmesser betragen

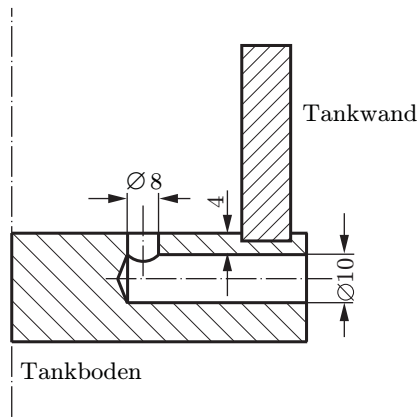


Abbildung C.1: Geometrie des Tankabflusses

im einzelnen 4 mm, 4,5 mm, 5 mm und 5,5 mm.

Zur Verringerung des wirksamen Querschnitts des Durchflusses zwischen den beiden Tanks wurden zwei einschraubbare Kunststoffteile, die einen Innendurchmesser von 10 mm besitzen, durch Einschieben von Kunststoffringen auf die Durchmesser 8 mm und 7 mm reduziert. Diese Kunststoffteile können auf der Innenseite von Tank 1 vor die Durchflussleitung geschraubt werden.

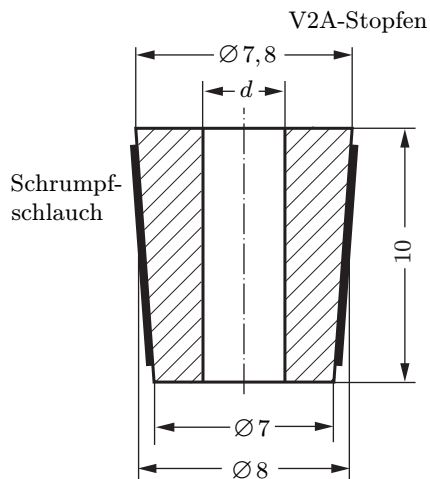


Abbildung C.2: Stopfen für Abflüsse von Tank 1 und 2

C.2 Identifikation

Um die wirksamen Abflussquerschnitte der Tanks zu bestimmen, wird Tank 1 des Zwei-Tank-Systems bis zur Höhe $x_{K1} = 50$ cm befüllt. Danach wird der Abfluss geöffnet und die Zeit gemessen bis der Füllstand in Tank 1 die Höhe $x_{K1} = 20$ cm erreicht hat.

Entsprechend wird bei der Bestimmung der wirksamen Querschnitte des Durchflusses zwischen den beiden Tanks vorgegangen. Tank 1 wird bis zur Höhe $x_{K1} = 55$ cm befüllt, Tank 2 bleibt leer. Nach dem Öffnen des Durchflusses wird die Zeit gemessen bis der Füllstand in Tank 1 die Höhe $x_{K1} = 35$ cm erreicht hat.

Die Messungen werden für jeden Stopfen mehrfach vorgenommen, um Ungenauigkeiten bei der manuellen Zeitnahme durch Mittelung zu eliminieren. Die Bestimmung des wirksamen Abfluss- bzw. Durchflussquerschnittes erfolgt letztendlich unter Verwendung des nichtlinearen Simulationsmodells des Zwei-Tank-Systems. In diesem werden die jeweiligen Parameter so eingestellt, dass eine Simulation des Identifizierungsvorgangs genau so lange dauert wie zuvor am Zwei-Tank-Laborsystem gemessen. Das Ergebnis dieser Simulation führt jeweils auf einen Wert für k_2 bzw. k_4 . Aus diesen Werten lässt sich der wirksame Durchmesser berechnen. Für den Abflussdurchmesser gilt mit $g = 981 \frac{\text{cm}}{\text{s}^2}$ und $d_{\text{Tank}_1} = 14$ cm die Gleichung

$$d_{\text{Abfluss}_1} = d_{\text{Tank}_1} \cdot \sqrt{\frac{k_2}{\sqrt{2} \cdot g}}.$$

Die Ergebnisse für die Abflüsse zeigt Tabelle C.1.

Tabelle C.1: Wirksame Abflussquerschnitte

Durchmesser	berechneter Querschnitt	Zeit
8,0 mm (ohne Stopfen)	50,27 mm ²	78,4 s
5,5 mm (Stopfen A)	23,76 mm ²	108,4 s
5,0 mm (Stopfen B)	19,63 mm ²	111,1 s
4,5 mm (Stopfen C)	15,90 mm ²	122,5 s
4,0 mm (Stopfen D)	12,56 mm ²	140,1 s
Parameter k_2	wirksamer Querschnitt	wirksamer Durchmesser

Tabelle C.1: Wirksame Abflussquerschnitte (Fortsetzung)

$0,066265\sqrt{\text{m}}/\text{s}$	23,03 mm ²	5,41 mm
$0,047942\sqrt{\text{m}}/\text{s}$	16,66 mm ²	4,61 mm
$0,046776\sqrt{\text{m}}/\text{s}$	16,26 mm ²	4,55 mm
$0,042445\sqrt{\text{m}}/\text{s}$	14,75 mm ²	4,33 mm
$0,037097\sqrt{\text{m}}/\text{s}$	12,89 mm ²	4,05 mm

Für den Durchmesser und die beiden Einschraubteile an der Verbindungsleitung gilt entsprechend

$$d_{\text{Durchfluss}} = d_{\text{Tank}_1} \cdot \sqrt{\frac{k_3}{\sqrt{2} \cdot g}}$$

Die Ergebnisse für die Durchflüsse sind in Tabelle C.2 für die originale Verbindungsleitung mit Durchmesser 10 mm und für die beiden Einschraubteile mit Durchmesser 7 mm und 8 mm dargestellt.

Tabelle C.2: Wirksame Durchflussquerschnitte

Durchmesser	berechneter Querschnitt	Zeit
10,0 mm	78,54 mm ²	71,8 s
8,0 mm	50,26 mm ²	77,7 s
7,0 mm	38,48 mm ²	85,0 s
Parameter k_3	wirksamer Querschnitt	wirksamer Durchmesser
$0,076965\sqrt{\text{m}}/\text{s}$	26,75 mm ²	5,84 mm
$0,071140\sqrt{\text{m}}/\text{s}$	24,72 mm ²	5,61 mm
$0,065061\sqrt{\text{m}}/\text{s}$	22,61 mm ²	5,37 mm

Literatur

- [Abe90] ABEL, D.: *Petri-Netze für Ingenieure*. Berlin : Springer, 1990
- [AD94] ALUR, R. ; DILL, D.: A Theory of Timed Automata. In: *Theoretical Computer Science* 126 (1994), S. 183–235
- [Alt02] ALT, W.: *Nichtlineare Optimierung: Eine Einführung in die Theorie, Verfahren und Anwendungen*. Braunschweig, Wiesbaden : Vieweg, 2002
- [Arn73] ARNOLD, L.: *Stochastische Differentialgleichungen: Theorie und Anwendung*. München : Oldenbourg Verlag, 1973
- [Bau96] BAUMGARTEN, B.: *Petri-Netze*. 2. Heidelberg : Spektrum Akademischer Verlag, 1996
- [BH75] BRYSON, A. E. ; HO, Y.-C.: *Applied Optimal Control*. Washington D.C., USA : Hemisphere Publishing Cooperation, 1975
- [BKLS03] BLANKE, M. ; KINNAERT, M. ; LUNZE, J. ; STAROSWIECKI, M.: *Diagnosis and Fault-Tolerant Control*. Berlin : Springer, 2003
- [BM99] BEMPORAD, A. ; MORARI, M.: Control of Systems Integrating Logic, Dynamics, and Constraints. In: *Automatica* 35 (1999), Nr. 3, S. 407–427
- [BMM99] BEMPORAD, A. ; MIGNONE, D. ; MORARI, M.: Moving Horizon Estimation for Hybrid Systems and Fault Detection. In: *Proceedings of the American Control Conference*. San Diego, California, USA, 1999
- [Bra95] BRANICKY, M.S.: *Studies in Hybrid Systems: Modeling, Analysis, and Control*, MIT, Diss., 1995
- [BS91] BRONSTEIN, I. N. ; SEMENDJAJEW, K. A. ; GROSCHE, G. (Hrsg.) ; ZIEGLER, V. (Hrsg.) ; ZIEGLER, D. (Hrsg.): *Taschenbuch der Mathematik*. 25. Frankfurt am Main : Verlag Harri Deutsch, 1991
- [Buc03] BUCHHOLZ, M.: *Modellprädiktive Regelung eines hybriden Systems am Beispiel eines Zwei-Tanks*, Universität Karlsruhe (TH), Institut für Regelungs- und Steuerungssysteme, Studienarbeit 175, 2003

- [Bus01] BUSS, M.: *Methoden zur Regelung hybrider dynamischer Systeme*. München : Technische Universität München, 2001
- [Cas93] CASSANDRAS, C.: *Discrete Event Systems: Modeling and Performance Analysis*. Boston : Aksen Associates and Irwin, 1993
- [CL99] CASSANDRAS, C. ; LAFORTUNE, S.: *Introduction to Discrete Event Systems*. Boston : Kluwer Academic Publishers, 1999
- [CVS04] COLLINS, P. ; VAN SCHUPPEN, J.: Observability of Piecewise-Affine Hybrid Systems. In: ALUR, R. (Hrsg.) ; PAPPAS, G. (Hrsg.): *Hybrid Systems: Computation and Control, Lecture Notes in Computer Science* Bd. 2993. Berlin : Springer, 2004, S. 265–279
- [Ede87] EDELSBRUNNER, H.: *Algorithms in Computational Geometry*. EATCS Monographs in Theoretical Computer Science 19, Springer, 1987
- [FGK03] FOURER, R. ; GAY, D. ; KERNIGHAN, B.W.: *AMPL : A Modeling Language for Mathematical Programming*. Pacific Grove, CA, USA : Thomson/Brooks/Cole, 2003
- [FM03] FERRARI-TRECATE, G. ; MUSELLI, M.: Single-Linkage Clustering for Optimal Classification in Piecewise Affine Regression. In: *Proceedings of Analysis and Design of Hybrid Systems (ADHS 03)*. Saint Malo, France, 2003
- [FMLM01a] FERRARI-TRECATE, G. ; MUSELLI, M. ; LIBERATI, D. ; MORARI, M.: A Clustering Technique for the Identification of Piecewise Affine Systems. In: DI BENEDETTO, M. (Hrsg.) ; SANGIOVANNI-VINCENTELLI, A. (Hrsg.): *Hybrid Systems: Computation and Control, Lecture Notes in Computer Science* Bd. 2034. Springer, 2001, S. 218–231
- [FMLM01b] FERRARI-TRECATE, G. ; MUSELLI, M. ; LIBERATI, D. ; MORARI, M.: Identification of Piecewise Affine and Hybrid Systems. In: *Proceedings of the American Control Conference* Bd. 5. Arlington, USA, 2001, S. 3521–3526
- [FMLM03] FERRARI-TRECATE, G. ; MUSELLI, M. ; LIBERATI, D. ; MORARI, M.: A Clustering Technique for the Identification of Piecewise Affine and Hybrid Systems. In: *Automatica* 39 (2003), S. 205–217
- [Föl93] FÖLLINGER, O.: *Lineare Abtastsysteme*. 5. München : Oldenbourg Verlag, 1993
- [Föl94] FÖLLINGER, O.: *Regelungstechnik*. 8. Heidelberg : Hüthig-Verlag, 1994
- [Fra94] FRANK, P. M.: Diagnoseverfahren in der Automatisierungstechnik. In: *Automatisierungstechnik* 42 (1994), Nr. 2, S. 47–64

- [Gau09] GAUSS, C. F.: *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*. Göttingen : Königliche Gesellschaft der Wissenschaften, 1809
- [GBR04] GUPTA, S. ; BISWAS, G. ; RAMIREZ, J. W.: An Improved Algorithm for Hybrid Diagnosis of Complex Systems. In: *Proceedings of 15th International Workshop on Principles of Diagnosis*. Carcassonne, Frankreich, 2004
- [Gis03] GISELBRECHT, D.: *Entwicklung von Methoden zur Diagnose hybrider dynamischer Systeme*, Universität Karlsruhe (TH), Institut für Regelungs- und Steuerungssysteme, Diplomarbeit 756, 2003
- [GRS97] GIROD, B. ; RABENSTEIN, R. ; STENGER, A.: *Einführung in die Systemtheorie*. Stuttgart : Teubner Verlag, 1997
- [Han92] HANSEN, E.: *Global Optimization Using Interval Analysis*. NewYork : Marcel Dekker Inc., 1992
- [HDB01] HEEMELS, W. P. M. H. ; DE SCHUTTER, B. ; BEMPORAD, A.: Equivalence of Hybrid Dynamical Models. In: *Automatica* 37 (2001), Juli, Nr. 7, S. 1085–1091
- [Hen02] HENNING, J.: *Entwicklung geeigneter Verfahren zur Kennfeldoptimierung stückweise affiner Modelle*, Universität Karlsruhe (TH), Institut für Regelungs- und Steuerungssysteme, Diplomarbeit 744, 2002
- [HL92] HOSCHEK, J. ; LASSER, D.: *Grundlagen der Geometrischen Datenverarbeitung*. Stuttgart : Teubner-Verlag, 1992
- [HM03] HODRUS, T. E. ; MÜNZ, E.: Hybride Phänomene in Zeitdiskreter Darstellung. In: *at-Automatisierungstechnik* 51 (2003), Dezember, Nr. 12, S. 574–582
- [HMU01] HOPCROFT, J. E. ; MOTWANI, R. ; ULLMANN, J. D.: *Introduction to Automata Theory, Languages, and Computation*. 2. Boston, München : Addison-Wesley, 2001
- [HN99] HUYER, W. ; NEUMAIER, A.: Global Optimization by Multilevel Coordinate Search. In: *Journal of Global Optimization* 14 (1999), S. 331–355
- [Hod02] HODRUS, T. E.: *Entwicklung geeigneter Clusterverfahren zur Identifikation stückweise affiner Modelle am Beispiel eines turbogeladenen Dieselmotors*, Universität Karlsruhe (TH), Institut für Regelungs- und Steuerungssysteme, Diplomarbeit 743, 2002
- [Hof99] HOFFMANN, I.: *Identifikation hybrider dynamischer Systeme*. Shaker Verlag, 1999

- [Hoh02] HOHMANN, S.: *Entwurf nichtlinearer Abtastregelungen auf der Basis einer polynomialen Zeitdiskretisierung*. Düsseldorf : VDI-Verlag, 2002
- [HP03] HIERETH, H. ; PRENNINGER, P.: *Aufladung der Verbrennungskraftmaschine*. Wien, New York : Springer, 2003
- [HRK05] HASCHKA, M. ; RÜGER, B. M. ; KREBS, V.: Identification of the Electrical Behavior of a Solid Oxide Fuel Cell in the Time-Domain. In: *Proceedings of Fractional Derivatives and their Applications*, 2005
- [HSK04a] HODRUS, T. E. ; SCHWARZ, M. ; KREBS, V.: A Time Discretization Method for a Class of Hybrid Systems. In: ALLGÖWER, Frank (Hrsg.): *Proceedings of IFAC Workshop on Nonlinear Control Systems*. Stuttgart, 2004
- [HSK04b] HODRUS, T. E. ; SCHWARZ, M. ; KREBS, V.: A Time Discretization Method For a Class of Hybrid Systems With Jumps. In: *Proceedings of IAR Annual Meeting*. Karlsruhe, 2004
- [HSZ96] HACKBUSCH, W. ; SCHWARZ, H. R. ; ZEIDLER, E. ; ZEIDLER, E. (Hrsg.): *Teubner-Taschenbuch der Mathematik*. Stuttgart : B.G. Teubner, 1996
- [HVS01] HABETS, L. ; VAN SCHUPPEN, J.: A Controllability Result for Piecewise-Linear Hybrid Systems. In: *Proceedings of European Control Conference*. Porto, Portugal, 2001
- [HWK⁺04] HASCHKA, M. ; WEICKERT, T. ; KREBS, V. ; SCHÄFER, S. ; IVERS-TIFFÉE, E.: Identification of a Nonlinear Model for the Electrical Behavior of a Solid Oxide Fuel Cell. In: *Proceedings of France Deutschland Fuel Cell Conference*. Belfort, Frankreich, 2004
- [Ise92] ISERMANN, R.: *Identifikation dynamischer Systeme*. 2. Berlin : Springer, 1992
- [Joh99] JOHANSSON, M.: *Piecewise Linear Control Systems*, Department of Automatic Control, Lund Institute of Technology, Diss., 1999
- [Joh03] JOHANSSON, M.: *Piecewise Linear Control Systems : A Computational Approach*. Berlin, Heidelberg : Springer, 2003
- [Kal03] KALEFA, M. B.: *Implementierung und Erprobung eines Verfahrens zur Identifikation hybrider Systeme*, Universität Karlsruhe (TH), Institut für Regelungs- und Steuerungssysteme, Masterarbeit 757, 2003
- [KK98] KAMMEYER, K. D. ; KROSCHEL, K.: *Digitale Signalverarbeitung*. Stuttgart : Teubner Verlag, 1998

- [KQ88] KÖNIG, R. ; QUÄCK, L.: *Petri-Netze in der Steuerungs- und Digitaltechnik*. München : Oldenbourg Verlag, 1988
- [LBR96] LABINAZ, G. ; BAYOUMI, M. ; RUDIE, K.: Modeling and Control of Hybrid Systems. In: *Proceedings of 13th IFAC World Congress*. San Francisco, USA, 1996
- [Lem95] LEMMER, K.: *Diagnose diskret modellierter Systeme mit Petri-Netzen*. Düsseldorf : VDI Verlag, 1995
- [Ley98] LEYFFER, S.: *User Manual for MINLP BB*. University of Dundee, UK, 1998
- [Ley01] LEYFFER, S.: Integrating SQP and Branch-and-Bound for Mixed Integer Nonlinear Programming. In: *Computational Optimization and Applications* 18 (2001), Nr. 3, S. 295 – 309
- [Lju99] LJUNG, L.: *System Identification: Theory for the User*. 2. Upper Saddle River, NJ : Prentice Hall, 1999
- [Lof90] LOFFELD, O.: *Estimationstheorie*. München : Oldenbourg Verlag, 1990
- [Lun97] LUNZE, J.: *Regelungstechnik*. Berlin : Springer, 1997
- [MB95] MOSTERMAN, Pieter J. ; BISWAS, Gautam: Behavior Generation Using Model Switching - a Hybrid Bond Graph Modeling Technique. In: F.E.CELLIER (Hrsg.) ; J.J.GRANDA (Hrsg.): *Proceedings of International Conference on Bond Graph Modeling and Simulation* Bd. 27, 1995, S. 177–182
- [MB96] MOSTERMAN, Pieter J. ; BISWAS, Gautam: A Formal Hybrid Modeling Scheme for Handling Discontinuities in Physical System Models. In: *Proceedings of AAAI-96* Bd. 2. Portland, Oregon, 1996, S. 985–990
- [MB97] MOSTERMAN, Pieter J. ; BISWAS, Gautam: Model-Based Diagnosis of Dynamic Systems. In: *Journées du LIPN, Univ. of Paris-Nord* (1997)
- [MBCG99] MCILRAITH, S. ; BISWAS, G. ; CLANCY, D. ; GUPTA, V. *Towards Diagnosing Hybrid Systems*. März 1999
- [MHK03a] MÜNZ, E. ; HODRUS, T. E. ; KREBS, V.: A Gradient Based Method for the Identification of Hybrid Systems. In: *Proceedings of 4th MATH-MOD*. Wien, 2003
- [MHK03b] MÜNZ, E. ; HODRUS, T. E. ; KREBS, V.: Top-Down Identification of Hybrid Characteristic Maps. In: *Proceedings of ADHS*. Saint Malo, 2003

- [MHK04] MÜNZ, E. ; HODRUS, T. E. ; KREBS, V.: Gradient Based Identification of Hybrid Systems. In: *Mathematical and Computer Modelling of Dynamical Systems* 10 (2004), März, Nr. 1, S. 25–40
- [MK02] MÜNZ, E. ; KREBS, V.: Identification of Hybrid Systems Using a Priori Knowledge. In: *Proceedings of the IFAC World Congress*. Barcelona, Spain, 2002
- [MK04] MÜNZ, Eberhard ; KREBS, Volker: Stochastic State Reconstruction in Piecewise Affine Hybrid Systems Based on Discrete Measurements. In: *Proceedings of NOLCOS 2004*. Stuttgart, 2004
- [MK05] MÜNZ, E. ; KREBS, V.: Continuous Optimization Approaches to the Identification of Piecewise Affine Systems. In: *Proceedings of IFAC World Congress*. Prag, 2005
- [MKB95] MOSTERMAN, Pieter J. ; KAPADIA, Ravi ; BISWAS, Gautam: Using Bond Graphs for Diagnosis of Dynamic Physical Systems. In: *Proceedings of Sixth International Workshop on Principles of Diagnosis*, 1995, S. 81–85
- [MMB99] MANDERS, E. J. ; MOSTERMAN, P. J. ; BISWAS, G.: Signal to Symbol Transformation Techniques for Robust Diagnosis in TRANSCEND. In: *Knowledge Systems Laboratory*, 1999
- [Mos97] MOSTERMAN, Pieter J.: *Hybrid Dynamic Systems: A Hybrid Bond Graph Modeling Paradigm and its Application in Diagnosis*. Nashville, Tennessee, Vanderbilt University, Diss., 1997
- [MSK04] MÜNZ, E. ; SCHULLERUS, G. ; KREBS, V.: On the Use of Hybrid System Identification Methods for Max-Plus-Linear System Identification. In: *Proceedings of IFAC Workshop on Discrete Event Systems*. Reims, Frankreich, 2004
- [Mül01] MÜLLER, C.: *Analyse und Synthese Diskreter Steuerungen Hybrider Systeme mit Petri-Netz-Zustandsraummodellen*. Düsseldorf : VDI-Verlag, 2001
- [NBK⁺00] NARASIMHAN, S. ; BISWAS, G. ; KARSAI, G. ; PASTERNAK, T. ; ZHAO, Feng: Building Observers to Address Fault Isolation and Control Problems in Hybrid Dynamic Systems. In: *Systems, Man, and Cybernetics* 4 (2000), Oktober, S. 2393–2398
- [Nel99] NELLES, O.: *Nonlinear System Identification with Local Linear Neuro-Fuzzy Models*. Aachen : Shaker Verlag, 1999

- [Nel01] NELLES, O.: *Nonlinear System Identification*. Berlin, Heidelberg : Springer, 2001
- [Nen01] NENNINGER, G.: *Modellbildung und Analyse hybrider dynamischer Systeme als Grundlage für den Entwurf hybrider Steuerungen*. Düsseldorf : VDI-Verlag, 2001
- [neo05] *NEOS Server for Optimization*. Dezember 2005. – <http://www-neos.mcs.anl.gov/>
- [NK97] NENNINGER, G. ; KREBS, V.: Modeling and Analysis of Hybrid Systems: A New Approach Integrating Petri Nets and Differential Equations. In: *Proceedings of the IEEE 5th International Workshop on Parallel and Distributed Real-Time Systems*. Genf, Schweiz, 1997, S. 234–238
- [NM02] NEUMANN, K. ; MORLOCK, M.: *Operations Research*. 2. München, Wien : Hanser, 2002
- [NTK05] NAKADA, H. ; TABAKA, K. ; KATAYAMA, T.: Identification of Piecewise Affine Systems Based on Statistical Clustering Technique. In: *Automatica* 41 (2005), Nr. 5, S. 905–913
- [Nyb99] NYBERG, M.: *Model Based Fault Diagnosis*. Linköping, Schweden, Linköping University, Diss., 1999
- [PFC89] PATTON, R. ; FRANK, P.-M. ; CLARK, R.: *Fault Diagnosis in Dynamic Systems*. Prentice Hall, 1989
- [PFC00] PATTON, R. ; FRANK, P.-M. ; CLARK, R.: *Issues of Fault Diagnosis for Dynamic Systems*. Springer, 2000
- [Pha00] PHAM, D. T.: *Intelligent Optimisation Techniques*. London, Berlin, Heidelberg : Springer, 2000
- [Phu04] PHUNG, A. S.: *Identifikation hybrider Systeme durch nichtlineare Optimierung*, Universität Karlsruhe (TH), Institut für Regelungs- und Steuerungssysteme, Masterarbeit 770, 2004
- [PULT00] PATTON, R. J. ; UPPA, F. J. ; LOPEZ-TORIBIO, C. J.: Soft Computing Approaches to Fault Diagnosis for Dynamic Systems: A Survey. In: *Proceedings of the 4th IFAC Symposium on Fault Detection Supervision and Safety for Technical Processes* Bd. 1. Budapest, Juni 2000, S. 298–311
- [Rat92] RATZ, D.: *Automatische Ergebnisverifikation bei globalen Optimierungsproblemen*, Universität Karlsruhe (TH), Diss., 1992

- [Rol03] ROLL, J.: *Local and Piecewise Affine Approaches to System Identification*, Linköping University, Diss., April 2003
- [Röm04] RÖMPERT, M.: *Beobachtergestützte Diagnose hybrider Systeme*, Universität Karlsruhe (TH), Institut für Regelungs- und Steuerungssysteme, Diplomarbeit 771, 2004
- [Sch01a] SCHNABEL, M.: *Diskret-Kontinuierliche Dynamische Systeme: Optimale Steuerung und Beobachtung*. Düsseldorf, Germany : VDI-Verlag, 2001
- [Sch01b] SCHRÖDER, M.: *Gebiete Optimal Aufteilen*, Universität Karlsruhe (TH), Diss., 2001
- [Sch03] SCHMID, K.: *Qualitative dynamische Modelle für das Verhalten von Hochtemperatur-Brennstoffzellen*. Düsseldorf : VDI-Verlag, 2003
- [Sch04a] SCHULLERUS, G.: *Identifikation Zeitbewerteter Ereignisdiskreter Systeme*. Düsseldorf : VDI-Verlag, 2004
- [Sch04b] SCHWARZ, M.: *Zeitdiskretisierung Stückweise Affiner Kontinuierlicher Systeme*, Universität Karlsruhe (TH), Institut für Regelungs- und Steuerungssysteme, Diplomarbeit, 2004
- [SFP03] SIMANI, S. ; FANTUZZI, C. ; PATTON, R. J.: *Model-Based Fault Diagnosis in Dynamic Systems Using Identification Techniques*. London, Berlin, Heidelberg : Springer, 2003
- [SKHK05] SCHWARZ, M. ; KIENCKE, U. ; HODRUS, T. E. ; KREBS, V.: Time Discretization of Piecewise Affine Systems With Sliding Modes. In: *Proceedings of IFAC World Congress*. Prag, 2005
- [Son98] SONTAG, E. D.: *Mathematical Control Theory*. New York, Heidelberg : Springer, 1998
- [TB02] TORRISI, F. D. ; BEMPORAD, A.: *HYDSEL 2.0.5 - User Manual*. ETH Zürich, Institut für Automation, 2002
- [TB03] TORRISI, F. D. ; BEMPORAD, A.: *HYDSEL - a Tool for Generating Computational Hybrid Models for Analysis and Synthesis Problems / ETH Zürich, Institut für Automation. 2003 (AUT 02-03). – Forschungsbericht*
- [Tho74] THOMA, Jean U.: *Grundlagen und Anwendungen der Bonddiagramme*. Essen : Verlag W. Girardet, 1974

-
- [UT96] ULLRICH, T. ; TOLLE, H.: Delaunay Networks for Modelling of Nonlinear Processes. In: *Proceedings of the IASTED Conference on Modelling and Simulation*. Pittsburgh, USA, 1996
- [Vid04] VIDAL, R.: Identification of PWARX Hybrid Models With Unknown and Possibly Different Orders. In: *Proceedings of IEEE American Conference on Control*. Boston, MA, USA, 2004
- [VSS03] VIDAL, R. ; SOATTO, S. ; SASTRY, S.: An Algebraic Geometric Approach to the Identification of Linear Hybrid Systems. In: *Proceedings of IEEE Conference on Decision and Control*. Maui, Hawaii USA, 2003
- [wik05] *Wikipedia: Die Freie Enzyklopädie.* Dezember 2005. – <http://de.wikipedia.org>

Lebenslauf

Persönliche Daten:

Name: Eberhard Münz
Geburtsdatum: 3. März 1976
Staatsangehörigkeit: deutsch

Schulbildung:

09/1982 - 08/1986 Besuch der Grundschule in Rot
09/1986 - 06/1995 Besuch des Gymnasiums „St. Paulusheim“ in Bruchsal

Studium:

10/1995 - 06/2000 Studium des Wirtschaftsingenieurwesens an der Universität Karlsruhe (TH)
01/2000 - 06/2000 Diplomarbeit am Institut für Regelungs- und Steuerungssysteme der Universität Karlsruhe (TH). Thema: „Modifizierte Optimale Regelung zeitvarianter nichtlinearer Mehrgrößensysteme“.

Berufliche Tätigkeit:

09/2000 - 08/2005 Wissenschaftlicher Angestellter bzw. Mitarbeiter am Institut für Regelungs- und Steuerungssysteme der Universität Karlsruhe (TH)

Schriften des Instituts für
Regelungs- und Steuerungssysteme,
Universität Karlsruhe (TH), Band 01



ISSN 1862-6688
ISBN 3-86644-029-4

www.uvka.de