

Identifikation funktionaler Fuzzy–Modelle

Andreas Kroll

Forschungsbericht 10/94

Meß-, Steuer- und Regelungstechnik

Übersicht: Dieser Bericht behandelt die Identifikation funktionaler Fuzzy–Modelle. Zuerst erfolgt die Vorstellung einer neuen Prämissenstruktur mit mehrdimensionalen Referenz–Fuzzy–Mengen. Eine solche Menge wird durch einen Schwerpunkt und eine Abstandsnorm festgelegt. Ein neues Verfahren, das ein Modell mit mehrdimensionalen Referenz–Fuzzy–Mengen über Clusterverfahren identifiziert, wird vorgestellt und die Anwendung auf zwei nichtlineare Systeme beschrieben.

Gerhard-Mercator-Universität - GH Duisburg

Meß-, Steuer- und Regelungstechnik

Prof. Dr.-Ing. H. Schwarz

Inhaltsverzeichnis

Nomenklatur	II
1 Einleitende Übersicht	1
2 Verfahren zur Clusterung von Daten	3
2.1 Matrix- und Vektornormen	5
2.2 Fuzzy- c -Means (FCM)	7
2.3 Backer-Algorithmen	9
2.4 Algorithmus von Gustafson und Kessel (Fuzzy Kovarianz-Matrizen)	11
3 Funktionale Fuzzy-Systeme	14
4 Parameteridentifikation des Konklusionspolynoms	20
5 Strukturidentifikation	23
5.1 Orthogonale Partitionierung	23
5.1.1 Wahl des Zugehörigkeitsfunktionstyps	23
5.1.2 Referenz-Fuzzy-Mengen und die Koeffizienten der Konklusionsfunk- tionen	25
5.1.3 Anpassung der Zugehörigkeitsfunktionen	27
5.2 Partitionierung mit mehrdimensionalen Zugehörigkeitsfunktionen	29
6 Beispiele zur Identifikation	31
6.1 Modellierung eines nichtlinearen Testsystems	31
6.2 Modellierung eines Gasofens	39
7 Zusammenfassung und Ausblick	44
8 Literaturverzeichnis	46
Anhang	49
A Verschiebung von Zugehörigkeitsfunktionsflanken	49
B Zur Wahl der Operatoren	52

Nomenklatur

Skalare scharfe Größen:

a_j	: Element von Θ
b_j	: Element von Θ
c_j	: Element von Θ
c	: Clusteranzahl
d_{ik}	: Abstand zwischen i -tem Clusterzentrum ν_i und k -tem Datenelement \mathbf{x}_k gemäß einer vorzugebenden Abstandsnorm
D	: Abstandsnorm (z. B. euklidische)
k	: Zählvariable
m	: Anzahl der eindimensionalen Eingangsgrößen (Dimension des Eingangsdaten- raumes)
n	: Anzahl der Fuzzy-Regeln
n_i	: Anzahl der Daten im i -ten harten Cluster
N	: Anzahl der Datenelemente
p_i	: i -ter Polynomkoeffizient
p_i	: Gütekriterium (mittlerer quadratischer Fehler) bzgl. nicht-normierter Größen
r	: Anzahl der Iterationen beim Gustafson und Kessel-Algorithmus
t	: Zeit
u_i	: i -te Ausgangsgröße
V_r	: Clusterzentrum (r -dimensional, für $r = 0$ ist $V_{r,i} = \mathbf{v}_i$)
w_k	: Gewicht des k -ten Datums
x_i	: i -te Eingangsgröße
y_i	: scharfe Ausgangsgröße der i -ten Regel
y	: scharfe Ausgangsgröße des gesamten Regelsatzes
α_{ji}	: j -te Partialprämissen der i -ten Regel
α	: Parameter der Zugehörigkeitsfunktion
β	: Parameter der Zugehörigkeitsfunktion
γ	: Parameter der Zugehörigkeitsfunktion
δ	: Parameter der Zugehörigkeitsfunktion
ε	: Terminierungsgrenze bei Clusterverfahren
ν	: Unschärfeparameter ($\nu > 1$)
μ_{ik}	: Zugehörigkeit des Punktes / Datums \mathbf{x}_k zum Cluster / zur Partition i
ϕ	: Maximal auftretende Verschiedenheit (z. B. Abstand) zweier Punkte im Datenraum $\phi = \max_{t,k}(\ \mathbf{x}_t - \mathbf{x}_k\)$
τ	: diskrete Totzeit
ω_i	: Gesamtprämissen der i -ten Regel
φ	: zeitliche Erfassungstiefe

Unscharfe Größen:

- A : Fuzzy-Menge
- S_e : Fuzzy-Einermenge
- X : fuzzifizierte Größe x
- \mathbf{X} : fuzzifizierte Größe \mathbf{x}

Vektoren und Matrizen:

- \mathbf{D} : Bewertungsmatrix bei Darstellung einer Norm als quadratische Form
- \mathbf{I} : Einheitsmatrix
- \mathbf{P} : Kovarianzmatrix
- \mathbf{x} : Vektor der Eingangsgrößen, Datum
- \mathbf{U} : Zugehörigkeitsmatrix
- \mathbf{v}_i : Clusterschwerpunkt des i -ten Clusters
- \mathbf{W} : Gewichtsmatrix bei gewichtetem (rekursiven) Least–Squares–Verfahren
- $\boldsymbol{\Theta}$: Parametervektor
- $\boldsymbol{\Psi}$: Meßvektor

Mengen und Räume:

- \mathbb{D} : Definitionsbereich
- \mathbb{D}_N : Definitionsbereich normierter Größen ($\mathbb{D}_N = [-1; 1]$)
- \mathbb{I} : Menge aller c Cluster, $\mathbb{I} = \{1, 2, \dots, c\}$
- \mathbb{I}_k : Menge aller Cluster, bei denen \mathbf{x}_k den Abstand 0 vom Clusterschwerpunkt hat
- $\tilde{\mathbb{I}}_k$: $\tilde{\mathbb{I}}_k = \mathbb{I} \setminus \mathbb{I}_k$
- \mathbb{M}_c : („Harter“) c -Partitionsraum
- \mathbb{M}_x : Menge aller Daten \mathbf{x}_k
- \mathbb{M}_i : Menge der Daten im harten Cluster i
- \mathbb{M}_{fc} : Fuzzy c -Partitionsraum
- \mathbb{N} : Menge der natürlichen Zahlen
- \mathbb{R} : Menge der reellen Zahlen
- \mathbb{R}^m : reeller Vektorraum der Dimension m
- \mathbb{V}_{cN} : Menge aller reellen $c \times n$ -Matrizen

Funktionen:

$\det(\cdot)$: Determinante
$\exp(\cdot)$: Exponentialfunktion
f	: allg. Funktion
g	: allg. Funktion
$\text{sgn}(\cdot)$: Signumfunktion
J	: Kosten- oder Straffunktion
V	: Verlustfunktion
$\mu_A(x)$: Zugehörigkeitsfunktion zur Fuzzy-Menge $A(x)$

Operatoren:

$(\cdot)^{-1}$: Inverse
$(\cdot)^T$: Transponiert
$\max(\cdot)$: Maximum- (MAX-) Operator
$\#(\cdot)$: Kardinalzahl (Anzahl der Elemente)

Sonstige Symbole:

\exists	: es existiert
\forall	: für alle
\setminus	: ohne (Mengenalgebra)
$\hat{(\cdot)}$: Schätzwert für \cdot

Indizes:

$(\cdot)_{a,b,h,l,r,s}$: Laufindizes
$(\cdot)_i$: Regelnummer
$(\cdot)_j$: Nummer der Eingangsgröße
$(\cdot)_k$: Nummer des Datums
$(\cdot)_{mes}$: Gemessene Größe
$(\cdot)_{mod}$: Prädizierte Größe
$(\cdot)^{(l)}$: l -ter Iterationsschritt
$(\cdot)^{(0)}$: Initialisierung

1 Einleitende Übersicht

Mathematische Modelle dienen zur Analyse und Synthese von Regelungssystemen. Diese Systemmodelle erlauben eine Erprobung und Optimierung von Regelungskonzepten in der Simulation. Außerdem können reale Prozesse im allgemeinen nicht auf alle Betriebsfälle und Störungen hin überprüft werden. Bei Systemen hoher Komplexität, falls Nichtlinearitäten vorliegen und / oder die Systemdynamik nicht genau bekannt ist oder nicht oder nur unter zu großem Aufwand analytisch beschreibbar ist, geraten konventionelle Modellbildungsverfahren an ihre Grenzen. Einen Ansatz zur Überwindung dieser Probleme bietet die Fuzzy-Logik (Xu und Lu 1987, Sugeno und Kang 1988, Pedrycz 1991, Shaw und Krüger 1992).

Dieser Bericht befaßt sich mit der Identifikation von funktionalen Fuzzy-Modellen. Funktionale Fuzzy-Modelle unterscheiden sich von den relationalen (Pedrycz 1984, Küpper 1994) dadurch, daß der Folgerungsteil der Fuzzy-Regeln durch eine (scharfe) analytische Funktion gegeben ist und nicht durch Fuzzy-Mengen beschrieben wird. Die Zielsetzung beider Modellierungsarten ist allerdings ähnlich. Es wird im allgemeinen ein zeitdiskretes parametrisches Ein-/Ausgangsgrößenmodell für einen Prozeß bestimmt. Das System wird also als *schwarzer Kasten* (Schwarz 1991) betrachtet. Vor einer eigentlichen Modellidentifikation sind Meßdaten zu gewinnen. Die Festlegung einer geeigneten Abtastzeit sowie geeigneter Testsignale ist allerdings nicht Bestandteil dieses Berichtes. Hier sei z. B. auf Isermann (1988a,b) und Reuter (1993) verwiesen.

Die Identifikation eines Fuzzy-Modells kann in zwei große Aufgaben zerlegt werden: Die Struktur- und die Parameteridentifikation, wobei die Grenzen zwischen der Zuordnung der Teilaufgaben fließend sind. Während die Bestimmung der Eingangsvariablen des Modells immer der Strukturidentifikation und die der Konklusionsparameter der Parameteridentifikation zugeordnet wird, ist die Zuordnung der Identifikation der Ein-/Ausgangsbeziehungen in der Literatur nicht einheitlich dargestellt. In diesem Bericht soll die Gliederung nach Sugeno und Yasukawa (1993) angewendet werden, die sich wie folgt darstellt: Die Strukturidentifikation gliedert sich in die Bestimmung der Eingangsvariablen des Modells (*Teil I*) und die Festlegung der Ein-/Ausgangsbeziehungen (*Teil II*). Im *Teil Ia* werden durch Plausibilitätsüberlegungen über physikalisches Hintergrundwissen potentiell sinnvolle Kandidaten für die Eingangsgrößen bestimmt. Es gibt hierzu keine generelle Lösung, sondern nur heuristisches Vorgehen (vorausgesetzt wird natürlich, daß die zu prädizierende Ausgangsgröße bekannt ist). Im *Teil Ib* werden aus den Eingangskandidaten die konkreten Modelleingangsgrößen bestimmt. Dazu sind nicht-signifikante Kandidaten zu eliminieren und die Signalverzögerungen in allen Eingängen zu bestimmen. Für die Selektion der Eingangsgrößen gibt es systematische Ansätze, die natürlich ein Gütekriterium voraussetzen. Sugeno und Kang (1988) stellen hierzu ein Verfahren über einen Suchbaum mit Elimination der Zweige schlechter Güte beim Übergang von Ebene zu Ebene des Baumes vor. Bei der Identifikation der Ein-/Ausgangsbeziehungen

wird im *Teil IIa* die Anzahl der Fuzzy–Regeln festgelegt. Dieser Teil ist im Zusammenhang mit *Teil IIb*, der Prämissestrukturidentifikation, d. h. der Partitionierung des Eingangsgrößenraumes, zu sehen.

Während sich bei relationalen und funktionalen Fuzzy–Modellen das Problem der Strukturidentifikation in ähnlicher Weise stellt (die Prämisse, d. h. die WENN–Teile der Regeln, sind gleich aufgebaut), unterscheiden sich die bei der Parameteridentifikation zu bestimmenden Parameter wegen der unterschiedlichen Konklusionsformen. Bei den funktionalen Fuzzy–Modellen besteht eine Konklusion aus einer analytischen Funktion. Im allgemeinen wird hier in den in der Literatur (Takagi und Sugeno 1985, Sugeno und Kang 1988, Sugeno und Tanaka 1991, Tan u. a. 1994) wie auch in dem in diesem Bericht vorgestellten neuen Verfahren ein in den Eingangsgrößen lineares Polynom angesetzt. Damit beschränkt sich die Identifikationsaufgabe auf die Bestimmung der Polynomkoeffizienten.

Im vorliegenden Bericht werden die Betrachtungen auf den *Teil II* der Strukturidentifikation und die Parameteridentifikation beschränkt. Nach einer Erläuterung von Clusterverfahren erfolgt eine kurze Beschreibung funktionaler Fuzzy–Systeme. Dabei wird eine neue Prämissestruktur mit mehrdimensionalen Referenz–Fuzzy–Mengen vorgestellt. Anschließend wird das zur Parameteridentifikation verwendete Verfahren der gewichteten kleinsten Fehlerquadrate beschrieben. Die Clusterverfahren werden im 5. Abschnitt zur Strukturidentifikation *Teil II* angewendet. Da viele Veröffentlichungen eindimensionale orthogonale Referenz–Fuzzy–Mengen zur Partitionierung verwenden, werden außerdem einige Betrachtungen zur Wahl dieser Art der Referenz–Fuzzy–Mengen vorgenommen. Danach wird ein neues Identifikationsverfahren vorgestellt, das funktionale Fuzzy–Modelle mit echt–mehrdimensionalen Referenz–Fuzzy–Mengen und linearen Polynomen als Konklusionsfunktionen identifiziert. Eine Anwendung des Verfahrens auf zwei nichtlineare Testsysteme, ein akademisches und ein praktisches Beispiel, erfolgt in Abschnitt 6.

2 Verfahren zur Clusterung von Daten

Bei der Clusterung wird die Anzahl c der Cluster („Haufen“) und eine Abstandsnorm (z. B. die Euklidische) vorgegeben. Dann werden c Clusterschwerpunkte bestimmt und allen N Datenpunkten eine Zugehörigkeit zu den c Clustern zugeordnet, die im Sinne der Abstandsnorm optimal ist. Es handelt sich dabei um ein Optimierungs- (Minimierungs-) Problem bezüglich einer Straf- oder Kostenfunktion J , die von der Datenmenge \mathbb{M}_x und den den Daten zugeordneten Zugehörigkeitsfunktionswerten abhängt. Dazu sind mehrere Cluster-Algorithmen verfügbar, die meistens speziell auf die Detektion bestimmter geometrischer Strukturen im untersuchten Datenraum zugeschnitten sind. Zu beachten ist, daß über die Festlegung der Abstandsnorm $\|\cdot\|$ bei der Optimierung der Kostenfunktion oft die von der Norm implizierte geometrische Clusterform erzwungen wird. Dies kann auch dann der Fall sein, wenn diese in der Datenstruktur evtl. nicht vorhanden ist. Deshalb bedeutet ein minimaler Wert der Kostenfunktion nicht gleichzeitig zwangsläufig eine gute Clusterung, da ja nur auf eine geometrische Struktur hin untersucht wurde.

Der Fuzzy- c -Means-Algorithmus (FCM), auch als Fuzzy-ISODATA bekannt, ist besonders auf hyperellipsoide geometrische Strukturen zugeschnitten. Nach Bezdek (1981) führt er bei Verwendung der euklidischen Abstandsnorm zu Zugehörigkeitsfunktionsoberflächen, die eine Mischung aus c Gauß'schen Wahrscheinlichkeitsdichtefunktionen darstellen (Bezdek 1981:162). Dabei ist c die Anzahl der Cluster. In der Literatur finden sich auch Algorithmen, die auf andere als hyperellipsoide geometrische Strukturen zugeschnitten sind oder sogar lokal unterschiedliche Clusterformen generieren können: Bei den Backer-Algorithmen (Backer 1978) können die Zugehörigkeitsfunktionen auch Maxima nahe den Grenzen des Datenraumes besitzen, was vorteilhaft bei nicht hyperellipsoiden Datenstrukturen ist.

Gustafson und Kessel (Bezdek 1981) stellen einen Algorithmus vor, der lokal unterschiedliche geometrische Strukturen in den Datensätzen erkennen und clustern kann.

Die von Bezdek (1981) entwickelte Familie der Fuzzy- c -Varieties-Algorithmen (FCV) stellt eine Erweiterung des FCM dar. Ein FCV erlaubt die Bestimmung punktförmiger ($r = 0$), linienförmiger ($r = 1$), ebener ($r = 2$) oder hyperplanarer ($r > 2$) Clusterzentren. Im Sonderfall von $r = 0$ stimmt er mit dem FCM überein. Da er allerdings nur den (orthogonalen) Abstand vom Clusterzentrum bestraft, sind z. B. bei $r = 1$ die Clusterzentren Linien unendlicher Länge. Zwei Datenhäufungen nahe der Linie, aber mit großem Abstand zueinander, werden deshalb wie zwei nahe benachbarte bestraft (vgl. Bild 2.1). In diesem Punkt setzt die Familie der Fuzzy-C-Elliptotypen (FCE) an. Die grundlegende Idee entspricht dem FCV für $r = 1$, allerdings werden hier auch die Abstände der Daten entlang der Clusterzentrumlinie bestraft. Über einen Parameter ϑ kann eingestellt werden, wie stark die elliptische in Abgrenzung zur linearen Bestrafung erfolgen soll.



Bild 2.1: Der Fuzzy- c -Varieties-Algorithmus ($r = 1$) bestraft die beiden dargestellten Datenhäufungen nahe der Linie, aber mit unterschiedlicher Streckung in Linienerichtung, gleich.

Diese kurzen Ausführungen sollen illustrieren, daß es den i. allg. optimalen Algorithmus nicht gibt, sondern daß je nach der im Einzelfall vorliegenden Datenstruktur ein darauf zugeschnittener Algorithmus angewendet werden sollte. Oft ist allerdings kein Vorwissen über die geometrische Struktur vorhanden. Bei mehr als zweidimensionalen Datenräumen lassen sich wegen der schlechten Anschaulich- und Visualisierbarkeit oft mit vertretbarem Aufwand keine Aussagen über die Datenstruktur machen. In solchen Fällen kann ein heuristisches Vorgehen zur Ermittlung des geeigneten Algorithmus nötig werden.

Der genauen Beschreibung der Clusterverfahren sei die Definition von Fuzzy- c -Partitionen und harten c -Partitionen vorangestellt:

Definition 2.1 : Fuzzy- c -Partition (Bezdek 1981:26)

Sei $\mathbb{M}_x = \{x_1, \dots, x_k, \dots, x_N\}$ eine endliche (Daten-) Menge mit N Elementen, $c \in \mathbb{N}$ die Anzahl der Partitionen mit $2 \leq c < N$ und \mathbb{V}_{cN} die Menge der reellen $c \times N$ -Matrizen, dann ist der Fuzzy- c -Partitionsraum für \mathbb{M}_x als

$$\mathbb{M}_{fc} = \left\{ U \in \mathbb{V}_{cN} \mid \begin{array}{l} \mu_{ik} \in [0, 1] \quad \forall i, k; \\ \sum_{i=1}^c \mu_{ik} = 1 \quad \forall k, \quad 0 < \sum_{k=1}^N \mu_{ik} < n \quad \forall i \end{array} \right\} \quad (2.1)$$

gegeben.

□

Dabei bezeichnet μ_{ik} den Zugehörigkeitsgrad des Datums x_k zur i -ten Partition. Das i -te Element der j -ten Spalte von \mathbf{U} ($i \in \{1, \dots, c\}, j \in \{1, \dots, N\}$) enthält die Zugehörigkeit des j -ten Datums zur i -ten Partition. Diese Partitionen können, wie in den folgenden Unterabschnitten beschrieben, z. B. durch Clusterung bestimmt werden.

Beispiel: Gegeben seien die in Bild 2.2 gezeigten Zugehörigkeitsfunktionen für *klein*, *mittel* und *groß* sowie die zwei Datenpunkte x_1 und x_2 .

Dann folgt $\mathbf{U} = [\mu_{ij}]$ zu

$$\mathbf{U} = \begin{bmatrix} (x_1) & (x_2) \\ 0 & 0,4 \\ 0,2 & 0,6 \\ 0,8 & 0 \end{bmatrix} \begin{array}{l} (\text{groß}) \\ (\text{mittel}) \\ (\text{klein}) \end{array} . \quad (2.2)$$

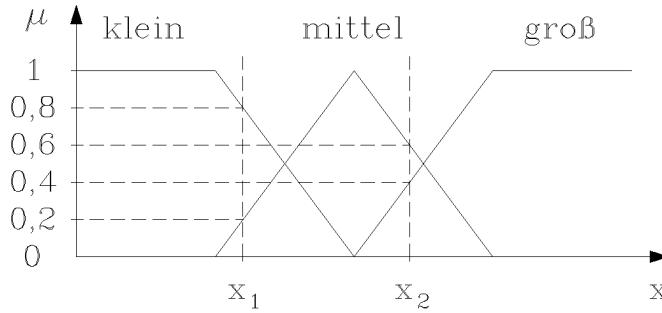


Bild 2.2: Zugehörigkeitsfunktionen für *klein*, *mittel* und *groß* sowie die Datenpunkte x_1 und x_2 .

□

Ein Sonderfall der Fuzzy- c -Partition ist die harte c -Partition:

Definition 2.2 : Harte c -Partition (Bezdek 1981:24)

Sei $\mathbb{M}_x = \{x_1, \dots, x_k, \dots, x_N\}$ eine endliche (Daten-) Menge mit N Elementen, $c \in \mathbb{N}$ die Anzahl der Partitionen mit $2 \leq c < N$ und \mathbb{V}_{cN} die Menge der reellen $c \times N$ -Matrizen, dann ist der harte c -Partitionsraum für \mathbb{M}_x als

$$\mathbb{M}_{fc} = \left\{ U \in \mathbb{V}_{cN} \mid \begin{array}{l} \mu_{ik} \in \{0, 1\} \quad \forall i, k; \\ \sum_{i=1}^c \mu_{ik} = 1 \quad \forall k, \quad 0 < \sum_{k=1}^N \mu_{ik} < n \quad \forall i \end{array} \right\} \quad (2.3)$$

gegeben.

□

2.1 Matrix- und Vektornormen

Die Clusteralgorithmen verwenden Matrix- bzw. Vektornormen für die Auswertung der Terminierungsbedingung ($\|\mathbf{U}^{(l+1)} - \mathbf{U}^{(l)}\|_T < \varepsilon$) bzw. die Bestimmung des Abstandes eines Datenpunktes vom Clusterschwerpunkt ($\|\mathbf{x}_k - \mathbf{v}_i\|_D$). Bei den Matrixnormen erscheinen 1 -, ∞ - und Spektral- (Hilbert-) Norm für den Anwendungszweck als nicht sehr sinnvoll. Geeigneter ist eine Erweiterung der Hölder-Norm (Lancaster und Tismenetsky 1985) auf $r \times s$ -Matrizen $\mathbf{A} = [a_{ij}]$:

$$\|\mathbf{A}\|_p = \left(\sum_{i=1}^r \sum_{j=1}^s |a_{ij}|^p \right)^{\frac{1}{p}}, \quad (2.4)$$

die als Spezialfall für $p = 2$ die euklidische (Frobenius-) Norm beinhaltet, oder eine Erweiterung der Gesamtnorm (Zurmühl und Falk 1986) auf $r \times s$ -Matrizen

$$\|\mathbf{A}\|_G = \sqrt{r \cdot s} \max_{\substack{1 \leq i \leq r \\ 1 \leq j \leq s}} \{|a_{ij}|\}. \quad (2.5)$$

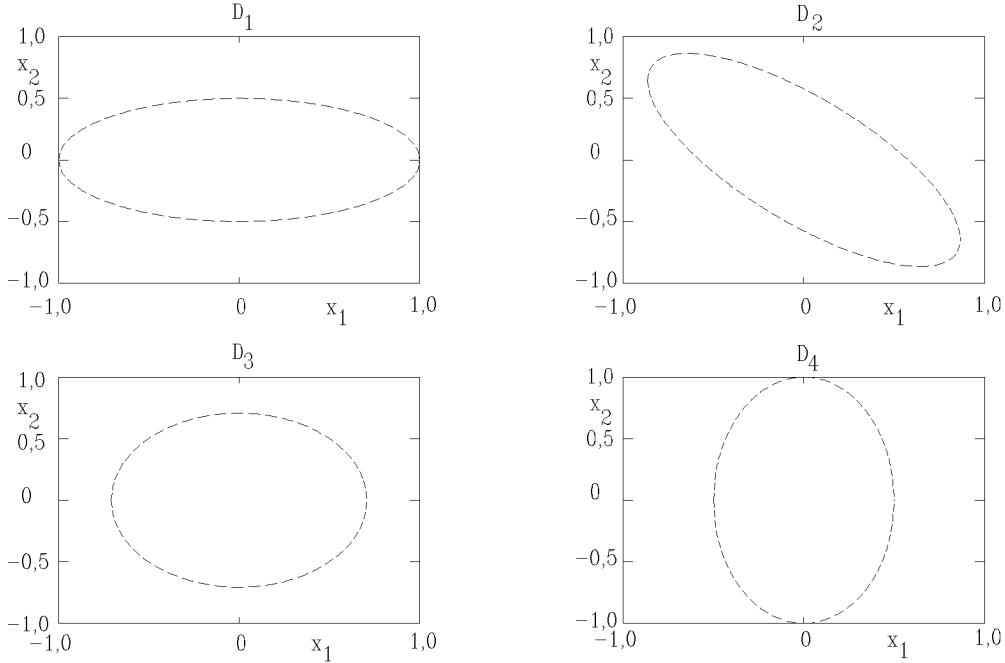


Bild 2.3: Grenzkurve $\|\mathbf{x}\|_{\mathbf{D}_i}^2 = 1$ für $\mathbf{D}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$, $\mathbf{D}_2 = \begin{bmatrix} 3 & \sqrt{5} \\ \sqrt{5} & 3 \end{bmatrix}$, $\mathbf{D}_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, $\mathbf{D}_4 = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$ mit $\det(\mathbf{D}_i) = 4 \quad \forall i$.

Während die Gesamtnorm nur die maximale Abweichung von $\|\mathbf{U}^{(l+1)} - \mathbf{U}^{(l)}\|$ bewertet, werden bei der p -Norm alle Abweichungen bestraft.

Bei Vektornormen sind z. B. 1-, ∞ - und p -Norm üblich. Letztere enthält als Sonderfall ($p = 2$) auch die euklidische (sphärische) Norm (Zurmühl und Falk 1986). Betrachtet werden soll die Darstellung der Norm als quadratische Form

$$\|\mathbf{x}\|_D^2 = \langle \mathbf{x}, \mathbf{x} \rangle_D = \mathbf{x}^T \mathbf{D} \mathbf{x} \quad . \quad (2.6)$$

Je nach Wahl der Matrix \mathbf{D} kann die Grenzkurve gleicher Norm (Equinormalkurve) beeinflußt werden: Für $\mathbf{D} = \mathbf{I}$ folgt die euklidische Norm, die Abweichungen in jede Richtung vom Ursprung gleich bestraft. Damit folgen sphärische Equinormalkurven um den Ursprung, d. h. konzentrische Kreise im zweidimensionalen Fall. Für allgemeine Diagonalmatrizen folgen als Equinormalkurven (Hyper-) Ellipsoide in Richtung einer der Koordinatenachsen, deren Exzentrizität gewählt werden kann. Für vollbesetzte Matrizen kann zusätzlich eine beliebige Lage der Ellipsoide eingestellt werden. Bild 2.3 zeigt exemplarisch die Equinormalkurven $\|\mathbf{x}\|_{\mathbf{D}_i}^2 = d^2 = 1$ für einige \mathbf{D}_i im zweidimensionalen Fall. Hier gilt

$$d^2 = \|\mathbf{x}\|_D^2 = \mathbf{x}^T \mathbf{D} \mathbf{x} = [x_1 \ x_2] \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad . \quad (2.7)$$

Die Equinormalkurven werden durch

$$x_2 = - \frac{a_2 + a_3}{2 a_4} x_1 \pm \sqrt{\frac{d^2}{a_4} - \left[\frac{a_1}{a_4} - \left(\frac{a_2 + a_3}{2 a_4} \right)^2 \right] x_1^2} \quad (2.8)$$

mit $x_1 \in [-x_g; x_g]$ beschrieben. Gilt $4a_1a_4 - (a_2 + a_3)^2 > 0$, so folgt

$$x_g = \sqrt{\frac{d^2}{a_4} \left[\frac{a_1}{a_4} - \left(\frac{a_2 + a_3}{2 a_4} \right)^2 \right]^{-1}}. \quad (2.9)$$

Bei Änderung der absoluten Werte der Elemente von \mathbf{D} unter gleichzeitiger Beibehaltung der relativen Beziehungen, d. h. sei $\mathbf{D}_\sigma = \sigma \mathbf{D}$, kann über den Wert der Determinante $\rho = \det(\mathbf{D}_\sigma) = \sigma \det(\mathbf{D})$ das von der Grenzkurve eingeschlossene Volumen (b. z. w. die Fläche im zweidimensionalen Fall) bei gleicher Kurvenform eingestellt werden (Bild 2.4). Eine Konstantsetzung der Determinante bei gleichzeitiger Variation der Elemente von \mathbf{D} beeinflusst dagegen primär (oder unter gewissen Voraussetzungen sogar ausschließlich) die Lage der Grenzkurve (vgl. Bild 2.3). Im zweidimensionalen Fall berechnet sich die von einer Equinormalkurve $\|x\|_D = d$ umschlossene Fläche zu

$$F = \frac{2 \pi d^2}{\sqrt{4 \det(\mathbf{D}) - (a_2 - a_3)^2}}. \quad (2.10)$$

Für symmetrische Matrizen \mathbf{D} ($a_2 = a_3$) ist F nicht von den einzelnen Werten der a_i , sondern nur von $\det(\mathbf{D})$ abhängig. Eine Beschränkung auf symmetrische Matrizen ist keine wesentliche Einschränkung, da im zweidimensionalen Fall a_2 und a_3 nur als Summe in (2.8) eingehen. Deshalb lässt sich auch für $a_2 = a_3$ eine beliebige Lage und Exzentrizität der Ellipsoide einstellen.

2.2 Fuzzy- c -Means (FCM)

Im folgenden wird der Fuzzy- c -Means Algorithmus beschrieben (Bezdek 1981):

FCM–Algorithmus

1. Clusteranzahl c ($2 \leq c < N$), Abstandsnorm D , geeignete Matrixnorm $\|\cdot\|_T$ für die Terminierungsbedingung, Unschärfeparameter $\nu \in (1, \infty)$ und Anfangswert $\mathbf{U}^{(0)} \in \mathbb{M}_{fc}$ vorgeben. Es sei $l = 0$.
2. Clusterschwerpunkte \mathbf{v}_i berechnen:

$$\mathbf{v}_i^{(l)} = \frac{\sum_{k=1}^N \left(\mu_{ik}^{(l)} \right)^\nu \mathbf{x}_k}{\sum_{k=1}^N \left(\mu_{ik}^{(l)} \right)^\nu} \quad i = 1, \dots, c \quad (2.11)$$

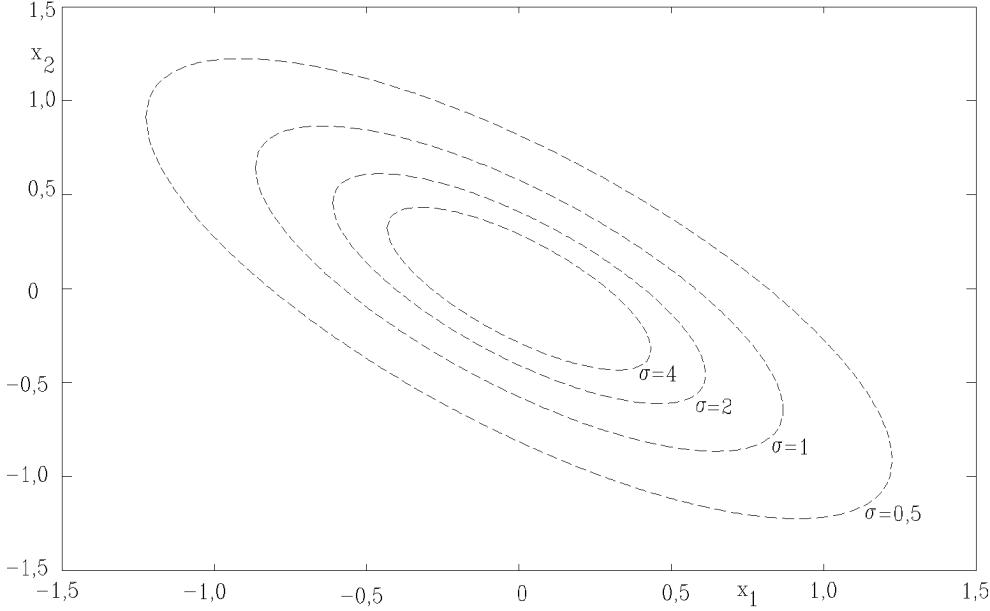


Bild 2.4: Grenzkurven gleicher Norm $\|\mathbf{x}\|_{D_i}^2 = \mathbf{x}^T \sigma \mathbf{D} \mathbf{x} = 1$ mit $\mathbf{D} = \mathbf{D}_2$ aus Bild 2.3.

3. $\mathbf{U}^{(l)}$ zu $\mathbf{U}^{(l+1)}$ aktualisieren :

- (a) Prüfung, ob Singularitäten ($d_{ik} = 0$) vorliegen. Sei $\mathbb{I} = \{1, \dots, c\}$ die Menge aller Cluster und $\mathbb{I}_k^{(l)} = \{i \in \mathbb{I} \mid d_{ik}^{(l)} = \|\mathbf{x}_k - \mathbf{v}_i^{(l)}\|_D = 0\}$ die Menge aller Cluster, bzgl. derer \mathbf{x}_k den Abstand 0 vom Clusterschwerpunkt hat. $\tilde{\mathbb{I}}_k^{(l)} = \mathbb{I} \setminus \mathbb{I}_k^{(l)}$ ist die Menge aller Cluster, bzgl. derer \mathbf{x}_k keine Singularität darstellt.
- (b) Berechne die Zugehörigkeit zu den neuen Clustern

$$\mu_{ik}^{(l+1)} = \begin{cases} \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}^{(l)}}{d_{jk}^{(l)}} \right)^{\frac{2}{\nu-1}}} & \text{für } \tilde{\mathbb{I}}_k^{(l)} = \emptyset \\ \text{(keine Singularitäten)} \\ 0 & \forall i \in \tilde{\mathbb{I}}_k^{(l)} \\ a_{ik}^{(l)} & \forall i \in \mathbb{I}_k^{(l)} \end{cases} \quad \text{für } \mathbb{I}_k^{(l)} \neq \emptyset \quad \text{(Singularitäten liegen vor)} \quad (2.12)$$

mit

$$a_{ik}^{(l)} : \sum_{i \in \mathbb{I}_k^{(l)}} a_{ik}^{(l)} = \sum_{i \in \mathbb{I}_k^{(l)}} \mu_{ik}^{(l)} = 1 \quad \forall i \in \mathbb{I}_k^{(l)} \neq \emptyset \quad . \quad (2.13)$$

Wähle z. B. $a_{ik}^{(l)} = \frac{1}{\#\mathbb{I}_k^{(l)}}$ $\forall i$. Damit ist $\mathbf{U}^{(l+1)} = [\mu_{ik}^{(l+1)}]$.

4. Terminierungsbedingung prüfen:

$$\begin{array}{ll} \text{WENN} & \|\mathbf{U}^{(l+1)} - \mathbf{U}^{(l)}\|_T \leq \varepsilon \\ \text{DANN} & \text{Stop} \end{array} \quad (2.14)$$

SONST $l = l + 1$, zurück zu 2.

□

Diesem Algorithmus liegt als Kostenfunktion

$$J_\nu(\mathbf{U}, \mathbf{v}, \mathbf{D}) = \sum_{k=1}^N \sum_{i=1}^c (\mu_{ik})^\nu \|\mathbf{x}_k - \mathbf{v}_i\|_D^2 = \sum_{k=1}^N \sum_{i=1}^c (\mu_{ik})^\nu (d_{ik})^2 \quad (2.15)$$

mit einer Fuzzy- c -Partition $\mathbf{U} \in \mathbb{M}_{f^c}$ von \mathbb{M}_x und $(d_{ik})^2 = \|\mathbf{x}_k - \mathbf{v}_i\|_D^2$ einer inneren Produktnorm $\|\cdot\|_D$ auf \mathbb{R}^p zugrunde. $\mathbf{v} = [\mathbf{v}_1, \dots, \mathbf{v}_c]$, $\mathbf{v}_i \in \mathbb{R}^m$ ist der Vektor der Clusterschwerpunkte und $\nu \in (1, \infty)$ ein Unschärfeparameter. Die Norm ist zu minimieren. Da d_{ik} quadratisch in J_ν eingeht, handelt es sich beim Cluster-Kriterium um ein quadratisches Fehlerkriterium. Für $\nu \rightarrow 1$ und die euklidische Abstandsnorm geht der FCM in den harten c -Means (HCM, Basis-ISODATA) nach Duda und Hart (Bezdek 1981:55) über. Anderseits wird die Partitionierung immer unschärfer, je größer ν gewählt wird. Bezdek (1981:73) zeigt, daß J_ν mit zunehmendem ν monoton fällt. Der Algorithmus führt zu einem lokalen Minimum von J_ν . Zu beachten ist, daß je nach Datenstruktur ein lokales oder globales Minimum von J_ν keineswegs eine gute Clusterung darstellen muß, da ja nur im Hinblick auf die gewählte Norm optimiert wird.

Bei der Verwendung der euklidischen Norm für $\|\cdot\|_D$ führt der FCM zu sphärischen Clustern. Bezdek (1981:85f) zeigt, daß die Bedingung (2.12) zum Aktualisieren auf $\mu_{ik}^{(l+1)}$ notwendig, aber nicht hinreichend für strikte lokale Minima von J_ν ist (für $\nu > 1$). Theoretisch konvergiert der FCM für unendlich häufiges Iterieren. Für eine endliche Anzahl von Iterationen ist keine Aussage möglich. Erfahrungen zeigen, daß bei Konvergenz nach endlicher Iterationsanzahl im allgemeinen das Ergebnis nahe einem lokalen Minimum liegt (Bezdek 1981).

2.3 Backer-Algorithmen

Die Backer-Algorithmen (Backer 1978) gehen von einer harten Partitionierung des Datenraumes aus. Über eine Prüfung der „Verwandtheit“ (Affinität) jedes Datenpunktes zu jedem harten Cluster werden Zugehörigkeiten der Datenpunkte zu den Clustern abgeleitet. Ein Verwandtheitskriterium, wie z. B. die „mittlere Partitions-Trennbarkeit“ B (*average partitions separability*, Bezdek 1981:157) mit (Backer 1978:106)

$$B(\mathbf{U}, c) = \frac{1}{c-1} \sum_{i=1}^{c-1} \sum_{j=i+1}^c \left[\frac{1}{N} \sum_{k=1}^N (\mu_{ik} - \mu_{jk})^2 \right] \quad (2.16)$$

ist vorzugeben. Jeder harten Partitionierung (durch $(\cdot)^*$ bezeichnet) wird genau eine weiche Partitionierung (wie zuvor aber ohne Stern) zugeordnet. Bezdek (1981) gibt eine Variante der Algorithmen an, die als Verwandtheitskriterium die mittlere Partitions-Trennbarkeit B und die euklidische Abstandsnorm $d_{tk} = \|\mathbf{x}_t - \mathbf{x}_k\|_2$ verwendet:

Backer–Algorithmus

1. Clusteranzahl c ($2 \leq c < N$) und eine harte Partitionierung $\mathbf{U}^{*(0)} = [\mu_{ik}^{*(0)}] \in \mathbb{M}_c$ vorgeben. Berechne $\phi = \max_{t,k} (d_{tk})$. Es sei $l = 0$.

2. Berechne weiche aus harten Clustern:

$$\mu_{ik}^{(l)} = \frac{n_i^{(l)} - \phi \sum_{t=1}^N \mu_{it}^{*(l)} d_{tk}^{(l)}}{N - \phi \sum_{t=1}^N d_{tk}^{(l)}} \quad (2.17)$$

mit \mathbb{M}_i der Menge aller Daten des i -ten harten Clusters und $n_i = \#(\mathbb{M}_i)$.

3. Berechne die Reklassifizierungsmatrix $\mathbf{R} = [r_{hk}]$ zur Neuzuweisung der Daten zu den Partitionen mit

$$r_{ik}^{(l)} = \frac{1}{n_i^{(l)}} \left[\left(\mu_{ik}^{(l)} \right)^2 - \frac{1}{n_i^{(l)} - 1} \sum_{x_t \in \mathbb{M}_{i,-k}^{(l)}} \left(\mu_{it}^{(l)} \right)^2 \right] \quad (2.18)$$

$$r_{jk}^{(l)} = \frac{1}{n_j^{(l)}} \left[\left(\mu_{jk}^{(l)} \right)^2 - \frac{1}{n_j^{(l)} + 1} \sum_{x_t \in \mathbb{M}_{j,+k}^{(l)}} \left(\mu_{jt}^{(l)} \right)^2 \right] \quad \text{für } 1 \leq j \leq c, \ j \neq i \quad (2.19)$$

mit \mathbb{M}_i der Menge aller Daten des i -ten harten Clusters, \mathbb{M}_j der Menge aller Daten des j -ten harten Clusters, $n_i = \#(\mathbb{M}_i)$, $n_j = \#(\mathbb{M}_j)$, $\mathbb{M}_{i,-k} = \mathbb{M}_i \setminus \{\mathbf{x}_k\}$ und $\mathbb{M}_{j,+k} = \mathbb{M}_j \cup \{\mathbf{x}_k\}$.

4. Aktualisiere die harte Partitionierung \mathbf{U}^* (diese Vorschrift führt zum iterativen Optimieren von B):

Für $k = 1, \dots, N$ berechne: $r_{ik}^{(l)} = \max_{1 \leq q \leq c} (r_{qk}^{(l)})$ mit

$$r_{qk}^{(l)} = \begin{cases} r_{ik}^{(l)} & \forall q = i \text{ nach (2.18)} \\ r_{jk}^{(l)} & \forall q \in \{1, \dots, c\} \setminus \{i\} \text{ nach (2.19).} \end{cases} \quad (2.20)$$

Für $t = i$ bleibt die Partitionierung erhalten:

$$\mu_{qk}^{*(l+1)} = \mu_{qk}^{*(l)} \quad \forall q \in \{1, \dots, c\} \quad . \quad (2.21)$$

Für $t \neq i$ wird repartitioniert:

$$\mu_{qk}^{*(l+1)} = \begin{cases} 1 & \forall q = t \\ 0 & \forall q \neq t \end{cases} \quad . \quad (2.22)$$

5. Prüfe Terminierungsbedingung:

WENN	$\mathbf{U}^{*(l+1)} = \mathbf{U}^{*(l)}$
DANN	Stop (keine Änderung der harten Clusterung)
SONST	$l = l + 1$, zurück nach 2.

(2.23)

Dieser Algorithmus kann im Gegensatz zum Fuzzy- c -Means auch Zugehörigkeitsfunktionen mit Maxima nahe den Grenzen des Datenraumes liefern. Somit kann er bei nicht hyperellipsoiden Datenstrukturen zu besseren Ergebnissen führen als der FCM. Der Algorithmus konvergiert und führt zu einem globalen Maximum von B bezüglich der durch die Zuordnung zu den harten Partitionen möglichen weichen Partitionen (Bezdek 1981:159). Die Konvergenz tritt nach einer endlichen Anzahl von Iterationen ein, wodurch als Terminierungsbedingung ein Vergleich auf Identität aufeinanderfolgender harter Partitionierungen möglich ist. Als Parameter des Algorithmus treten die Anzahl der Cluster c , die Wahl der Abstandsnorm $d = \|\cdot\|$ und des Verwandtheitskriteriums sowie die Festlegung von ϕ auf. So kann die Unschärfe der Partitionierung durch Einführung eines Faktors m' in

$$\phi = m' \max_{k,j} (d_{kj}), \quad (2.24)$$

der bisher zu 1 gesetzt wurde, beeinflußt werden.

Der Reklassifizierung liegt die Idee zu Grunde, durch elementweise Austauschvorgänge zwischen den harten Partitionen die im Sinne eines Gütekriteriums (implizit in (2.18) bzw. (2.19) enthalten) optimale harte Zuordnung der Daten zu den Clustern zu erreichen. Für den Reklassifizierungsvorgang sind pro Datum $c - 1$ und somit insgesamt $N(c - 1)$ Austauschvorgänge während eines Iterationsschrittes zu betrachten. Wie bereits zuvor erwähnt, werden aus der harten Clusterung die unscharfen Zugehörigkeiten abgeleitet.

2.4 Algorithmus von Gustafson und Kessel (Fuzzy Kovarianz-Matrizen)

Der FCM ist wegen der zugrundeliegenden einheitlichen Norm gut für hyperellipsoide geometrische Datenstrukturen geeignet. Zur Clusterung von lokal unterschiedlichen geometrischen Strukturen, z. B. lokal teilweise sphärische und teilweise ellipsoide Strukturen, stellen Gustafson und Kessel (Bezdeck 1981) einen Algorithmus vor, der eine lokale Variation der Abstandsnorm erlaubt. Statt einer global wirkenden Matrix-Norm \mathbf{D} mit $d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\|_D$ gilt also für jedes der c Cluster jeweils eine Norm \mathbf{D}_i mit $d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\|_{D_i}$. Bei dem Algorithmus ist bezüglich $\|\mathbf{x}\|_{D_i}^2 = \mathbf{x}^T \mathbf{D}_i \mathbf{x}$ für jedes \mathbf{D}_i der Wert seiner Determinante $\det(\mathbf{D}_i) = \rho_i > 0$ festzulegen. Die Wahl von ρ_i legt dabei das Volumen des i -ten Clusters fest, die Variation von \mathbf{D}_i bei konstantem ρ_i entspricht

der Suche nach der optimalen Clusterform (siehe auch Abschnitt 2.1). Als Straffunktion findet

$$\hat{J}_\nu(\mathbf{U}, \mathbf{v}, \mathbf{D}_1, \dots, \mathbf{D}_c) = \sum_{k=1}^N \sum_{i=1}^c (\mu_{ik})^\nu \| \mathbf{x}_k - \mathbf{v}_i \|_{D_i}^2 \quad (2.25)$$

mit den lokalen Matrixnormen $\mathbf{D}_1, \dots, \mathbf{D}_c$ Verwendung. Gustafson und Kessel formulieren ihren Algorithmus, der teilweise auf den FCM–Algorithmus zurückgreift, wie folgt:

Fuzzy Kovarianz–Matrizen–Algorithmus

1. Clusteranzahl c ($2 \leq c < N$), geeignete Matrixnorm $\|\cdot\|_T$ für die Terminierungsbedingung, Unschärfe parameter $\nu \in (1, \infty)$, $\rho_i \in [0, \infty)$ ($1 \leq i \leq c$) und Anfangswert $\mathbf{U}^{(0)} \in \mathbb{M}_{fc}$ vorgeben. Es sei $l = 0$.
2. Clusterschwerpunkte \mathbf{v}_i berechnen (wie beim FCM):

$$\mathbf{v}_i^{(l)} = \frac{\sum_{k=1}^N (\mu_{ik}^{(l)})^\nu \mathbf{x}_k}{\sum_{k=1}^N (\mu_{ik}^{(l)})^\nu} \quad i = 1, \dots, c \quad (2.26)$$

3. Die c Matrizen

$$\mathbf{S}_{fi}^{(l)} = \sum_{k=1}^N (\mu_{ik}^{(l)})^\nu (\mathbf{x}_k - \mathbf{v}_i^{(l)}) (\mathbf{x}_k - \mathbf{v}_i^{(l)})^T, \quad (2.27)$$

$\det(\mathbf{S}_{fi})$ und \mathbf{S}_{fi}^{-1} für alle $1 \leq i \leq c$ berechnen.

4. Lokale Matrizen \mathbf{D}_i

$$\mathbf{D}_i = (\rho_i \det(\mathbf{S}_{fi}))^{\frac{1}{m}} \mathbf{S}_{fi}^{-1}, \quad 1 \leq i \leq c \quad (2.28)$$

(mit der Dimension m des Eingangsdatenraums) berechnen.

5. Aktualisiere $\mathbf{U}^{(l)}$ zu $\mathbf{U}^{(l+1)}$ (Vorgehen wie beim FCM, aber mit lokal unterschiedlichen Abstandsmaßen):

$$d_{ik}^{(l)} = \| \mathbf{x}_k - \mathbf{v}_i^{(l)} \|_{D_i} \quad 1 \leq i \leq c, \quad 1 \leq k \leq N \quad (2.29)$$

$$\mu_{ik}^{(l+1)} = \begin{cases} \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}^{(l)}}{d_{jk}^{(l)}} \right)^{\frac{2}{\nu-1}}} & \text{für } \mathbb{I}_k^{(l)} = \emptyset \\ 0 & \forall i \in \tilde{\mathbb{I}}_k^{(l)} \\ a_{ik}^{(l)} & \forall i \in \mathbb{I}_k^{(l)} \end{cases} \quad \begin{array}{l} \text{(keine} \\ \text{Singularitäten)} \\ \text{(Singularitäten} \\ \text{liegen vor)} \end{array} \quad (2.30)$$

mit

$$a_{ik}^{(l)} : \sum_{i \in \mathbb{I}_k^{(l)}} a_{ik}^{(l)} = \sum_{i \in \mathbb{I}_k^{(l)}} \mu_{ik}^{(l)} = 1 \quad \forall \quad i \in \mathbb{I}_k^{(l)} \neq \emptyset \quad . \quad (2.31)$$

$$\text{Wähle z. B. } a_{ik}^{(l)} = \frac{1}{\#(\mathbb{I}_k^{(l)})} \quad \forall i. \text{ Damit ist } \mathbf{U}^{(l+1)} = \left[\mu_{ik}^{(l+1)} \right] \quad .$$

6. Terminierungsbedingung prüfen:

$$\begin{aligned} \text{WENN} \quad & \|\mathbf{U}^{(l+1)} - \mathbf{U}^{(l)}\|_T \leq \varepsilon \\ \text{DANN} \quad & \text{Stop} \\ \text{SONST} \quad & l = l + 1, \text{ zurück zu 2.} \end{aligned} \quad (2.32)$$

3 Funktionale Fuzzy-Systeme

Ein Funktional-Fuzzy-System mit m Eingangsgrößen und einer Ausgangsgröße beruht auf n Regeln der Form

$$R_i : \text{WENN } (X_1 \text{ IST } A_{1,r(i)}^i) \text{ UND } \dots \text{ UND } (X_m \text{ IST } A_{m,r(i)}^i) \text{ DANN } y_i = f_i(x_1, \dots, x_m) \quad (3.1)$$

mit

- x_j : scharfe Eingangsgröße, $j = 1, \dots, m$, $x_j \in \mathbb{D}_{x_j}$
- X_j : unscharfe Eingangsgröße, $j = 1, \dots, m$,
- $A_{j,r(i)}^i$: r -te Referenz-Fuzzy-Menge der j -ten Eingangsgröße der i -ten Regel,
- y_i : scharfe Ausgangsgröße der i -ten Regel und
- $f_i(x_1, \dots, x_m)$: Funktion der m scharfen Eingangsgrößen.

Ein solches System kann sowohl einen Regler als auch ein Modell beschreiben. Die Auswertung der Prämisse der Fuzzy-Regeln erfolgt wie beim Relational-Fuzzy-Regler: Für einen Eingangsgrößenvektor $\mathbf{X} = [X_1, \dots, X_m]^T$ bezeichnet

$$\alpha_{j,i} = X_j \text{ IST } A_{j,r(i)}^i \quad (3.2)$$

den Erfülltheitsgrad der j -ten Partialprämisse der i -ten Regel bzgl. der r -ten Referenz-Fuzzy-Menge. Bei Fuzzifizierung über Fuzzy-Einermengen S_e , deren Zugehörigkeitsfunktionen als

$$\mu_{S_e}(x) = \begin{cases} 1 & \forall x = e \\ 0 & \forall x \neq e \end{cases} \quad (3.3)$$

definiert sind, vereinfacht sich (3.2) zu

$$\alpha_{j,i} = \mu_{A_{j,r(i)}^i}(x_j) \quad . \quad (3.4)$$

Diese Fuzzifizierungsmethodik soll im folgenden vorausgesetzt werden. Bei multiplikativer Verknüpfung der Partialprämissen berechnet sich der Erfülltheitsgrad ω_i der Prämisse der i -ten Regel zu

$$\omega_i = \prod_{j=1}^m \alpha_{j,i} \quad . \quad (3.5)$$

Für die Konklusionspolynome f_i gibt Buckley (1993) für einen Regler mit zwei Eingangsgrößen ($m = 2$) einen allgemeinen Ansatz als

$$y_i = f_i(x_1, x_2) = \sum_{a=0}^{h_a(i)} \sum_{b=0}^{h_b(i)} p_i(a, b) x_1^a x_2^b \quad i = 1, \dots, n \quad (3.6)$$

an, wobei p_i reelle Polynomkoeffizienten bezeichnen. Takagi und Sugeno (1985), Sugeno und Kang (1986, 1988) wie auch Sugeno und Tanaka (1991) berechnen dagegen die (Partial-) Modellausgangsgrößen als in den m Eingangsgrößen lineare Polynome zu

$$f_i(x_1, \dots, x_m) = p_i(0) + \sum_{a=1}^m p_i(a) x_a \quad . \quad (3.7)$$

In die Berechnung der n Größen y_i gehen die Erfülltheitsgrade der Prämisen ω_i nicht ein. Sie werden bei der Aggregation berücksichtigt. Bei der Aggregation über eine gewichtete Mittelwertbildung (Takagi und Sugeno 1985) gilt

$$y = \frac{\sum_{i=1}^n y_i \omega_i}{\sum_{i=1}^n \omega_i} \quad , \quad (3.8)$$

wobei mindestens ein ω_i ($i \in \{1, 2, \dots, n\}$) größer als Null vorausgesetzt wird. Buckley (1993) führt bei der Aggregation keine Normierung der Gewichte ω_i der Größen y_i im Sinne einer Summation zu eins – wie in (3.8) – durch und erhält deshalb

$$y = \sum_{i=1}^n y_i \omega_i \quad . \quad (3.9)$$

Die Gleichungen (3.1), (3.2) und (3.5) beschreiben eine Prämisenstruktur, bei der die Mehrdimensionalität des Eingangsgrößenraumes durch die Verknüpfung explizit eindimensionaler Zugehörigkeitsfunktionen vorgenommen wird. In dem in diesem Bericht in Abschnitt 5.2 vorgestellten neuen Verfahren werden bei der Strukturidentifikation über Clusterverfahren explizit mehrdimensionale Zugehörigkeitsfunktionen betrachtet. Dabei wird in der Prämisse der i -ten Regel abgefragt, wie groß die Affinität (Wesensverwandtschaft) des Datums \mathbf{x}_k zum Partitions- oder Clusterschwerpunkt \mathbf{v}_i unter der Randbedingung ist, daß weitere $c-1$ Partitions- oder Clusterschwerpunkte $\mathbf{v}_1, \dots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_c$ gegeben sind. Es werden also *Punktaffinitäten* betrachtet. Eine solche Regelstruktur läßt sich als

$$R_i : \begin{array}{l} \text{WENN } (\mathbf{x} \text{ IST } \mathbf{v}_i) |_{\mathbf{v}_1 \dots \mathbf{v}_c} \\ \text{DANN } y_i = f_i(\mathbf{x}) \end{array} \quad (3.10)$$

mit

- x_j : scharfe Eingangsgröße, $j = 1, \dots, m$, $x_j \in \mathbb{D}_{x_j}$,
- \mathbf{x} : Vektor der scharfen Eingangsgrößen mit $\mathbf{x} = [x_1, \dots, x_m]^T \in \mathbb{D}_{x_1} \times \dots \times \mathbb{D}_{x_m}$,
- \mathbf{v}_i : den Partitions- oder Clusterschwerpunkten der i -ten Partition bzw. des i -ten Clusters,
- y_i : scharfe Ausgangsgröße der i -ten Regel und
- $f_i(\mathbf{x})$: Funktion der m scharfen Eingangsgrößen

schreiben. Während die Clusterschwerpunkte explizit durch die Clusteralgorithmen berechnet werden können, bleibt noch der Begriff *Partitionsschwerpunkt* zu erläutern. Ausgehend von einer orthogonalen Partitionierung kann statt der explizit eindimensional definierten orthogonalen Referenz-Fuzzy-Mengen auf explizit mehrdimensional definierte Zugehörigkeitsfunktionen übergegangen werden. Dazu sind die Mittelpunkte (z. B. die geometrischen) $\tilde{\mathbf{v}}$ einer jeden (orthogonalen) Partition, die sogenannten Partitionsschwerpunkte, zu ermitteln und mit diesen genau wie mit den Clusterschwerpunkten zu verfahren. Es gilt also $\tilde{\mathbf{v}}_i = \mathbf{v}_i$.

Für die Auswertung der Prämisse in (3.10) wird die folgende, vom Fuzzy- c -Means-Algorithmus übertragene, Vorschrift verwendet:

$$\mu_{ik} = \begin{cases} \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{\nu-1}}} & \text{für } \mathbb{I}_k = \emptyset \\ 0 & \forall i \in \tilde{\mathbb{I}}_k \\ a_{ik} & \forall i \in \mathbb{I}_k \end{cases} \begin{array}{l} \text{für } \mathbb{I}_k = \emptyset \\ \text{(keine} \\ \text{Singularitäten)} \\ \text{für } \mathbb{I}_k \neq \emptyset \\ \text{(Singularitäten} \\ \text{liegen vor)} \end{array} \quad (3.11)$$

Dabei ist

- \mathbf{x} : das k -te Datum,
- d_{ik} : der Abstand zwischen dem Datum \mathbf{x}_k und dem Partitions- oder Clusterschwerpunkt \mathbf{v}_i nach der Abstandsnorm D , $d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\|_D$,
- c : die Anzahl der Partitions- oder Clusterschwerpunkte,
- ν : der Unschärfeparameter mit $\nu \in (1, \infty)$,
- \mathbb{I} : der Menge aller Cluster $\mathbb{I} = \{1, \dots, c\}$,
- \mathbb{I}_k : die Menge aller Cluster, bei denen \mathbf{x}_k den Abstand 0 vom Clusterzentrum besitzt ($d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\|_D = 0$) und
- $\tilde{\mathbb{I}}_k$: die Menge aller Cluster, bezüglich derer \mathbf{x}_k keine Singularität darstellt ($\tilde{\mathbb{I}}_k = \mathbb{I} \setminus \mathbb{I}_k$).

Die Fallunterscheidung

$$\mu_{ik} = a_{ik} \quad \forall i \in \mathbb{I}_k \neq \emptyset \quad (3.12)$$

mit

$$\sum_{i \in \mathbb{I}_k} a_{ik} = \sum_{i \in \mathbb{I}_k} \mu_{ik} = 1 \quad (3.13)$$

besagt, daß, falls ein \mathbf{x}_k in einen mehrfachen Clusterschwerpunkt fällt, die Zugehörigkeiten zu diesen Clustern jeweils kleiner Eins sein müssen. Dafür ist die Randbedingung verantwortlich, daß die Summe der Zugehörigkeiten zu allen Partitionen für alle Punkte des Definitionsräumes gleich Eins sein muß. Sinnvoll ist hierbei z. B. eine Wahl

$$a_{ik} = \frac{1}{\#(\mathbb{I}_k)} \quad \forall i \quad . \quad (3.14)$$

Da der Fall mehrfacher Clusterschwerpunkte in einem Punkt praxisfern ist, kann vereinfachend angenommen werden, daß maximal ein Clusterschwerpunkt in einem Punkt des

Eingangsgrößenraumes liegen kann. Ist $\mathbb{I}_k \neq \emptyset$, dann enthält es genau ein Cluster h , für das $d_{hk} = 0$ gilt. Damit vereinfacht sich (3.11) zu

$$\mu_{ik} = \begin{cases} \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{\nu-1}}} & \text{für } \mathbb{I}_k = \emptyset \\ & \text{(keine} \\ & \text{Singularitäten)} \\ 0 \quad \forall \quad i \in \tilde{\mathbb{I}}_k & \text{für } \mathbb{I}_k \neq \emptyset \\ 1 \quad \forall \quad i = h & \text{(Singularitäten} \\ & \text{liegen vor)} \end{cases} \quad (3.15)$$

Die Gln. (3.11) und (3.15) gewährleisten, daß die Summe der Zugehörigkeiten eines Datums \mathbf{x}_k zu allen c Partitionen exakt gleich Eins ist. Im Fall von Singularitäten ist dies offensichtlich. Für $\mathbb{I} = \emptyset$ gilt:

$$\begin{aligned} \sum_{i=1}^c \mu_{ik} &= \sum_{i=1}^c \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{\nu-1}}} \\ &= \sum_{i=1}^c \frac{1}{(d_{ik})^{\frac{2}{\nu-1}} \sum_{j=1}^c \left(\frac{1}{d_{jk}} \right)^{\frac{2}{\nu-1}}} \\ &= \frac{1}{\sum_{j=1}^c \left(\frac{1}{d_{jk}} \right)^{\frac{2}{\nu-1}}} \sum_{i=1}^c \frac{1}{(d_{ik})^{\frac{2}{\nu-1}}} \\ &= 1 \end{aligned} \quad (3.16)$$

Das heißt, das mit (3.11) und (3.15) vorgestellte Verfahren zur Festlegung der Referenz-Fuzzy-Mengen führt per se zu einem Fuzzy-Informationssystem (Meyer-Gramann und Jüngst 1993). Anstelle der in (3.11) und (3.15) verwendeten global einheitlichen Abstandsnorm kann auch eine lokal variierende Abstandsnorm wie beim Algorithmus von Gustafson und Kessel (Bezdek 1981) eingesetzt werden. Auch hier gilt $\sum_{i=1}^c \mu_{ik} = 1$, wie an (3.16) abzulesen ist.

Für den Fall eines zweidimensionalen normierten Eingangsgrößenraumes, einer für alle Partitionen einheitlichen Euklidischen Norm als Abstandsnorm und drei Clustern mit den Schwerpunkten

$$\mathbf{v}_1 = [-0,5; 0,5]^T, \quad \mathbf{v}_2 = [-0,5; -0,5]^T \quad \text{und} \quad \mathbf{v}_3 = [0,5; 0]^T \quad (3.17)$$

sieht die zur Referenz-Fuzzy-Menge des ersten Clusters korrespondierende echt-mehrdimensionale Zugehörigkeitsfunktion nach (3.15) mit $\nu = 1,13$ wie in Bild 3.1 gezeigt aus. Je kleiner ν gewählt wird, desto schärfer werden die Partitionen. Für $\nu \rightarrow 1$ folgen letztlich

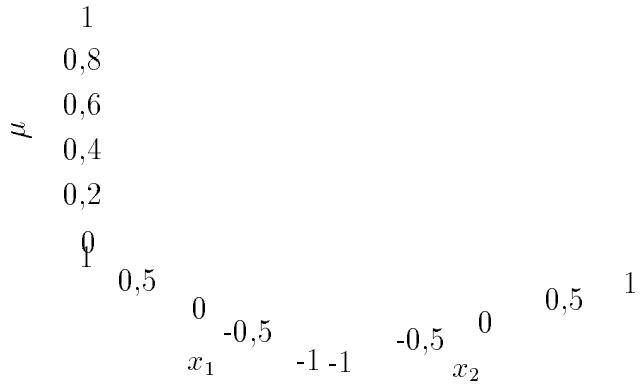


Bild 3.1: Zugehörigkeiten μ_{1k} von $\mathbf{x}_k = [x_1; x_2]^T$ zum ersten Cluster für $\nu = 1, 13$

harte Partitionen ($\mu_{ik} \in \{0; 1\}$). Für größere Werte von ν werden die Zugehörigkeitsfunktionen unschärfer. Bild 3.2 zeigt $\mu_{1k}(\mathbf{x}_k)$ für eine Auswertung von (3.17) mit $\nu = 2$ gemäß (3.15). Die beiden Mulden liegen genau in den beiden Schwerpunkten \mathbf{v}_1 und \mathbf{v}_2 des 2. und 3. Clusters. Hier gilt $\mu_{1k}(\mathbf{v}_2) = \mu_{1k}(\mathbf{v}_3) = 0$.

Mit steigendem Wert von ν wird allerdings der Abstand eines Datenpunktes vom Clusterschwerpunkt stärker bestraft. Dadurch folgen für große Werte von ν Zugehörigkeitsfunktionen, die im Schwerpunkt des korrespondierenden Clusters 1 liefern, in den Schwerpunkten der anderen Cluster 0 und im übrigen Bereich nahezu $\mu = 1/c$. Bild 3.3 illustriert dies für einen bereits relativ großen Wert $\nu = 5$.

Mit der bei mehrdimensionalen Zugehörigkeitsfunktionen sinnvollen Vorgabe, daß der Eingangsgrößenraum und das Argument der Zugehörigkeitsfunktionen die gleiche Dimension besitzen (wie dies bei einer Clusterung auch automatisch der Fall ist), stellt der Zugehörigkeitsgrad μ_{ik} des Datums \mathbf{x}_k zur Partition i direkt auch den (Gesamt-) Erfülltheitsgrad der Regelprämissen dar, also gilt

$$\omega_i(\mathbf{x}_k) = \mu_{ik} \quad . \quad (3.18)$$

Dadurch entfällt die explizite Verknüpfung der eindimensionalen Partialprämissen wie bei der oben beschriebenen Regelstruktur nach (3.1).

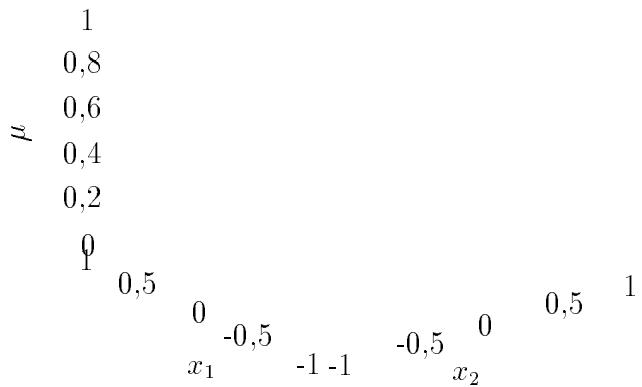


Bild 3.2: Zugehörigkeiten μ_{1k} von $\mathbf{x}_k = [x_1; x_2]^T$ zum ersten Cluster für $\nu = 2$

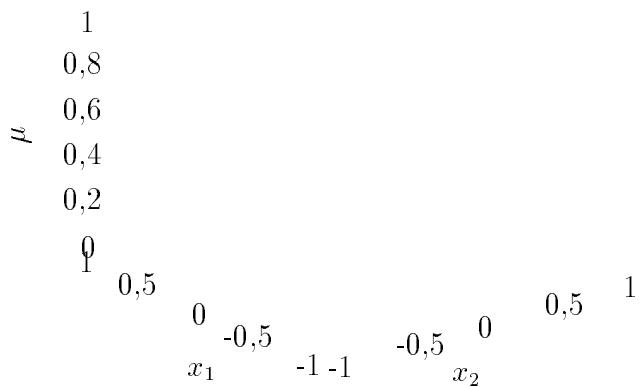


Bild 3.3: Zugehörigkeiten μ_{1k} von $\mathbf{x}_k = [x_1; x_2]^T$ zum ersten Cluster für $\nu = 5$

4 Parameteridentifikation des Konklusionspolynoms

Eine der Aufgaben bei der Identifikation eines funktionalen Fuzzy–Modells besteht darin, die Parameter der n Schlußfolgerungsfunktionen zu bestimmen. Setzt man (wie allgemein üblich: Sugeno und Kang 1988, Sugeno und Tanaka 1991, Takagi und Sugeno 1985, Tan u. a. 1994) als Funktionen in den Eingangsgrößen lineare Polynome an, so sind diese linear in ihren Parametern, so daß Least–Squares–Verfahren (LS–Verfahren) angewendet werden können (Reuter 1992).

Da jede Fuzzy–Regel eine der c Partitionen des Eingangsgrößenraumes beschreibt, sind die zur Identifikation zur Verfügung stehenden Daten den einzelnen Partitionen zuzuordnen. Dadurch entstehen bei Verwendung eindimensionaler orthogonaler Referenz–Fuzzy–Mengen im unscharfen Überlappungsbereich der benachbarten Partitionen Konflikte: Nutzt man ein Datum nur zur Identifikation der Partition, zu der es die größte Zugehörigkeit besitzt, so werden i. allg. nicht alle Daten berücksichtigt, die in den Einzugsbereich ($\omega > 0$) der Partition fallen (d. h. es werden nicht alle *relevanten* Daten ausgewertet). Verwendet man dagegen alle Daten im Einzugsbereich, so führt ein Standard–LS–Verfahren zur Überbetonung der Daten in den Übergangsbereichen zwischen den Partitionen. Bei den im vorherigen Abschnitt vorgestellten echt–mehrdimensionalen Referenz–Fuzzy–Mengen ist ein Standard–LS–Verfahren nicht sinnvoll anwendbar. Abhilfe bietet die Nutzung eines gewichteten LS–Verfahrens (Isermann 1988a), das bereits von Sugeno und Tanaka (1991) eingesetzt wurde. Dabei wird jedes Datum, gewichtet entsprechend seiner Zugehörigkeit zu der zu identifizierenden Partition, bei der Identifikation *jeder* Partition berücksichtigt. Dem LS–Verfahren liegt dann nicht mehr eine Verlustfunktion

$$V = \mathbf{e}^T \mathbf{e} \quad (4.1)$$

mit

$$e(k) = x(k)_{mes} - x(k)_{mod} \quad \text{und} \quad (4.2)$$

$$\mathbf{e} = [e(1), \dots, e(N)]^T \quad (4.3)$$

zugrunde, sondern

$$V = \mathbf{e}^T \mathbf{W} \mathbf{e} \quad (4.4)$$

mit einer Diagonalmatrix \mathbf{W} . Die Diagonalelemente w_{kk} von \mathbf{W} sind dabei die Gewichtungen des k -ten Datums x_k . Für diese Form der Gewichtung gibt Isermann (1988a) das rekursive Verfahren der kleinsten Fehlerquadrate (Weighted Recursive Least Squares Verfahren, WRLS) wie folgt an:

Gesucht seien die Parameter einer linearen Differenzengleichung

$$y(k) = -\hat{a}_1 y(k-1) - \dots - \hat{a}_{\varphi_y} y(k-\varphi_y) + \hat{b}_1 x(k-\tau-1) + \dots + \hat{b}_{\varphi_x} x(k-\tau-\varphi_x) \quad (4.5)$$

mit bekannter diskreter Totzeit τ des Eingangssignals $x(k)$. Die Parameter werden zum Vektor

$$\hat{\boldsymbol{\Theta}} = [\hat{a}_1, \dots, \hat{a}_{\varphi_y}, \hat{b}_1, \dots, \hat{b}_{\varphi_x}]^T \quad (4.6)$$

zusammengefaßt. Dann bestimmt sich der $(k+1)$ -te Schätzwert $\hat{\boldsymbol{\Theta}}(k+1)$ zu

$$\hat{\boldsymbol{\Theta}}(k+1) = \hat{\boldsymbol{\Theta}}(k) + \gamma(k)[y(k+1) - \boldsymbol{\Psi}^T(k+1)\hat{\boldsymbol{\Theta}}(k)] \quad (4.7)$$

mit

$$\gamma(k) = \frac{\boldsymbol{P}_w(k)\boldsymbol{\Psi}(k+1)}{\boldsymbol{\Psi}^T(k+1)\boldsymbol{P}_w(k)\boldsymbol{\Psi}(k+1) + \frac{1}{w(k+1)}} \quad (4.8)$$

und

$$\boldsymbol{P}_w(k+1) = [\mathbf{I} - \gamma(k)\boldsymbol{\Psi}^T(k+1)]\boldsymbol{P}_w(k) \quad . \quad (4.9)$$

Dabei ist

$$\boldsymbol{\Psi}(k) = [-y(k-1), -y(k-2), \dots, -y(k-\varphi_y), x(k-\tau-1), \dots, x(k-\tau-\varphi_x)]^T \quad (4.10)$$

der k -te Datenvektor.

Zu einer i. allg. gewünschten Modellbildung für Systeme mit m Eingangsgrößen gelangt man durch eine Ergänzung von Gleichung (4.5), (4.6) und (4.10) wie folgt:

$$\begin{aligned} y(k) = & -\hat{a}_1 y(k-1) - \dots - \hat{a}_{\varphi_y} y(k-\varphi_y) + \\ & + \hat{b}_1 x_1(k-\tau_{x1}-1) + \dots + \hat{b}_{\varphi_{x1}} x_1(k-\tau_{x1}-\varphi_{x1}) \\ & + \dots + \end{aligned} \quad (4.11)$$

$$+ \hat{c}_1 x_m(k-\tau_{xm}-1) + \dots + \hat{c}_{\varphi_{xm}} x_m(k-\tau_{xm}-\varphi_{xm}) \quad , \quad (4.12)$$

$$\hat{\boldsymbol{\Theta}} = [\hat{a}_1, \dots, \hat{a}_{\varphi_y}, \hat{b}_1, \dots, \hat{b}_{\varphi_{x1}}, \dots, \hat{c}_1, \dots, \hat{c}_{\varphi_{xm}}]^T \quad (4.13)$$

und

$$\begin{aligned} \boldsymbol{\Psi}(k) = & [-y(k-1), -y(k-2), \dots, -y(k-\varphi_y), \\ & x_1(k-\tau_{x1}-1), \dots, x_1(k-\tau_{x1}-\varphi_{x1}), \\ & \dots \\ & x_m(k-\tau_{xm}-1), \dots, x_m(k-\tau_{xm}-\varphi_{xm})]^T \quad . \end{aligned} \quad (4.14)$$

Wie beim ungewichteten Verfahren sind hier die Startwerte für $\hat{\boldsymbol{\Theta}}(0)$ und $\boldsymbol{P}_w(0)$ vorzugeben. Reuter (1992) und Ben Farhat (1994) geben Hinweise zur Wahl der Startwerte.

Das WRLS-Verfahren wird wegen der c Partitionen auch c -mal durchgeführt. Das Gewicht $w_k^i \in [0; 1]$ des k -ten Datums bei der Identifikation der i -ten Partition wird dabei gleich der Zugehörigkeit zu dieser Partition gesetzt. Daten mit der Zugehörigkeit Null sollten zur Erhöhung der Rechengeschwindigkeit und zur Vermeidung numerischer Probleme beim WRLS-Verfahren nicht berücksichtigt werden. Dies ist numerisch exakt, da für $w(k+1) = 0$ aus den obigen Gleichungen $\hat{\Theta}(k+1) = \hat{\Theta}(k)$ folgt.

Während die im folgenden Abschnitt beschriebenen Clusterverfahren bewirken, daß die Summe der Zugehörigkeiten eines Datums zu allen Clustern exakt gleich 1 ist, hängt dies bei Partitionierung über orthogonale Referenz-Fuzzy-Mengen von der Festlegung der (eindimensionalen) Referenz-Fuzzy-Mengen und des Verknüpfungsoperators bzgl. der Partialprämissen ab. Im folgenden seien Referenz-Fuzzy-Mengen angenommen, deren Zugehörigkeitsfunktionen sich im gesamten Definitionsbereich für jede Stelle zu Eins addieren (verbreitetes Vorgehen). Bei einer multiplikativen Verknüpfung der Partialprämissen wie in (3.5) ergibt die Summe der (Gesamt-) Zugehörigkeiten eines Datums zu allen Partitionen exakt 1 (siehe Anhang B). Bei Wahl der Gewichte w_k^i gleich der Gesamtzugehörigkeit $\omega_i(\mathbf{x}_k)$ zur Partition i haben alle Daten bzgl. der gesamten Identifikation das gleiche Gewicht. Kein Datum wird im (kritischen) Übergangsbereich über- oder unterbewertet.

Werden die Partialprämissen dagegen über den Minimum-Operator

$$\omega_i(\mathbf{x}_k) = \min_{1 \leq j \leq m} (\alpha_{ji}) \quad (4.15)$$

verknüpft, so erfolgt eine Überbewertung der Daten im Übergangsbereich (siehe Anhang B), wenn $w_k^i = \omega_i(\mathbf{x}_k)$ gewählt wird.

In diesem Bericht wird von multiplikativer Verknüpfung nach (3.5) und $w_k^i = \omega_i(\mathbf{x}_k)$ ausgegangen. Es besteht aber auch die Möglichkeit, die Datengewichtung gezielt zu beeinflussen. So können die $\omega_i(\mathbf{x}_k)$ z. B. über eine *konzentrierende* oder *kontrastverstärkende* Funktion auf die w_k^i abgebildet werden. Die Beschreibung von Konzentration, Kontrastverstärkung u. a. m. entnehme man z. B. Zadeh (1973).

5 Strukturidentifikation

Bei der Strukturidentifikation von Fuzzy-Modellen (sowohl relationaler als auch funktionaler) stellt sich die Aufgabe einer geeigneten Partitionierung des Eingangsgrößenraumes. Ein einfaches Vorgehen besteht in einer gleichförmigen Partitionierung der Definitionsbezirke der Eingangsgrößen. Eine solche regelmäßige Modellstruktur stimmt im allgemeinen aber nicht mit der Struktur des nichtlinearen Systems überein. Bessere Ergebnisse bei der Modellierung sind durch Anpassung der Zugehörigkeitsfunktionsform und -lage an das System zu erwarten. Hier existieren Ansätze, die eine Anpassung der jeweils bzgl. einer Eingangsgröße definierten Referenz-Fuzzy-Mengen vornehmen und solche, die sich von der strikten Zuordnung der Zugehörigkeitsfunktionen zu einer Dimension (d. h. einer Eingangsgröße) lösen. Zu letzteren gehören z. B. die Clusteranalyse der Testdaten und eine daraus abgeleitete Partitionierung des Eingangsgrößenraums. Der folgende Unterabschnitt geht auf die häufig benutzten orthogonalen eindimensionalen Referenz-Fuzzy-Mengen ein. Anschließend wird ein neues Verfahren mit mehrdimensionalen Zugehörigkeitsfunktionen vorgestellt.

5.1 Orthogonale Partitionierung

Beim modellbasierten Entwurf funktionaler Fuzzy-Regler finden Modelle mit orthogonaler Partitionierung des Eingangsgrößenraumes häufig Verwendung (Sugeno und Kang 1986, Tanaka und Sugeno 1992). Deshalb sollen in diesem Abschnitt einige Betrachtungen zu dieser Form der Prämisse gemacht werden. Wegen der Bedeutung der Wahl der Referenz-Fuzzy-Mengen für eine gute Modellierung werden im folgenden Abschnitt einige typische Fuzzy-Mengentypen vorgestellt und daran anschließend der Zusammenhang zwischen der Wahl der Referenz-Fuzzy-Mengen und der Parameter der Konklusionspolynome betrachtet.

5.1.1 Wahl des Zugehörigkeitsfunktionstyps

Im folgenden soll von auf $\mathbb{D}_N = [-1,1]$ normierten Eingangsgrößen des Modells ausgegangen werden. Durch die Gewichtung der Partialausgangsgrößen y_i der n einzelnen Regeln mit den Erfülltheitsgraden ω_i der entsprechenden Prämisen nach (3.8) oder (3.9) erreicht man eine Partitionierung des Raums der Eingangsgrößen

$$\mathbb{D}_{x_1} \times \dots \times \mathbb{D}_{x_m} = \mathbb{D}_N \times \dots \times \mathbb{D}_N = \mathbb{D}_N^m \quad (5.1)$$

in unscharfe Subräume. Diese Partitionierung hängt von Anzahl, Form und Lage der mit den Referenz-Fuzzy-Mengen korrespondierenden Zugehörigkeitsfunktionen ab. Beispielsweise führen Einer-Fuzzy-Mengen-Zugehörigkeitsfunktionen nach (3.3) dazu, daß die in den Konklusionen definierten Regelgesetze nach (3.6) oder (3.7) nur punktuell gelten. Bei treppenförmigen Zugehörigkeitsfunktionen, d. h.

$$\mu_{A_j^i}(x_j) = \begin{cases} 1 & \forall x_{ug,j}^i \leq x_j \leq x_{og,j}^i \\ 0 & \text{sonst} \end{cases} \quad (5.2)$$

und lückenlosem Angrenzen benachbarter Referenz–Fuzzy–Mengen, d. h. es sei $\mathbf{x} = [x_1, \dots, x_j, \dots, x_m]^T \in \mathbb{D}_N^m$ und es gelte $\exists \mu_{A_{jr}^i}(x_j) = 1, i \in \{1, \dots, n\}$ für alle $x_j \in \mathbb{D}_N$ mit $j \in \{1, \dots, m\}$, dann erhält man Partitionen, in denen jeweils nur genau ein $f_i(x_1, \dots, x_m)$ einen Beitrag zu y liefert, da die ω_i nur Werte $\omega_i \in \{0, 1\}$ annehmen können. Dagegen führen Zugehörigkeitsfunktionen mit „weichen“ Flanken, d. h. Flankenbreiten > 0 , zu unscharfen Rändern der Partitionen. Damit entstehen Übergangsbereiche zwischen benachbarten Referenz–Fuzzy–Mengen, wenn der Eingangsgrößenraum \mathbb{D}_N^m lückenlos durch die Regelbasis abgedeckt wird, d. h. es ist $\mathbf{x} = [x_1, \dots, x_j, \dots, x_m]^T$, dann gilt für jedes $x_j \in \mathbb{D}_N$ mit $j \in \{1, \dots, m\}$: $\exists \mu_{A_j^i}(x_j) > \sigma, i \in \{1, \dots, n\}$. Für die Bedingung der Lückenlosigkeit reicht $\sigma = 0$. Um „kräftige“ Regeln zu erhalten, sollte z. B. $\sigma = 0,5$ gewählt werden (Lee 1990). Zu den Zugehörigkeitsfunktionen mit „weichen“ Flanken zählen z. B. trapezoide nach

$$\mu_T(x) = \begin{cases} 0 & x \leq \alpha \\ \frac{x - \alpha}{\beta - \alpha} & \alpha < x \leq \beta \\ 1 & \beta < x \leq \gamma \\ \frac{\delta - x}{\delta - \gamma} & \gamma < x \leq \delta \\ 0 & x > \delta \end{cases}, \quad (5.3)$$

die für $\beta = \gamma$ in triangulare übergehen, sowie gemischt trigonometrisch–lineare nach

$$\mu_{TL}(x) = \begin{cases} 0 & x \leq \alpha \\ 0,5 + 0,5 \cos \left(\pi \left(\frac{x - \alpha}{\beta - \alpha} + 1 \right) \right) & \alpha < x \leq \beta \\ 1 & \beta < x \leq \gamma \\ 0,5 + 0,5 \cos \left(\pi \frac{x - \gamma}{\delta - \gamma} \right) & \gamma < x \leq \delta \\ 0 & x > \delta \end{cases} \quad (5.4)$$

oder Zugehörigkeitsfunktionen mit S–Flanken (Driankov u. a. 1993) nach

$$\mu_{TS}(x) = \begin{cases} 0 & x \leq \alpha \\ 2 \left(\frac{x - \alpha}{\beta - \alpha} \right)^2 & \alpha < x \leq \frac{\alpha + \beta}{2} \\ 1 - 2 \left(\frac{x - \beta}{\beta - \alpha} \right)^2 & \frac{\alpha + \beta}{2} < x \leq \beta \\ 1 & \beta < x \leq \gamma \\ 1 - 2 \left(\frac{x - \gamma}{\delta - \gamma} \right)^2 & \gamma < x \leq \frac{\gamma + \delta}{2} \\ 2 \left(\frac{x - \delta}{\delta - \gamma} \right)^2 & \frac{\gamma + \delta}{2} < x \leq \delta \\ 0 & x > \delta \end{cases} \quad (5.5)$$

(siehe Bild 5.1). Die Auswahl der passenden Zugehörigkeitsfunktionen erfolgt in Abhängigkeit von der Anwendung. Während $\mu_{TL}(x)$ und $\mu_{TS}(x)$ stetig differenzierbar sind, gilt dies nicht für $\mu_T(x)$. Falls eine Optimierung der Zugehörigkeitsfunktionen über Gradientenverfahren erfolgen soll, ist deren stetige Differenzierbarkeit erforderlich, und damit sind trapezoide Funktionen für diese Anwendung nicht einsetzbar. In diesen Fällen finden auch gaußförmige Zugehörigkeitsfunktionen nach

$$\mu_G(x) = \exp\left(-\left(\frac{x-a}{b}\right)^2\right) \quad (5.6)$$

Anwendung (z.B. Tan u. a. 1994)

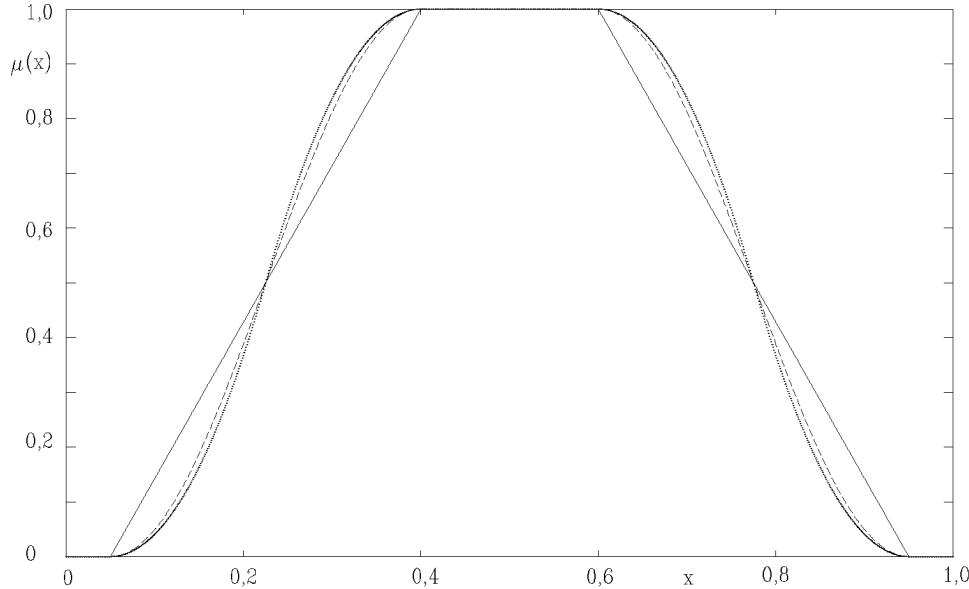


Bild 5.1: Trapezoide (–) und gemischtrigonometrisch-lineare (--) Zugehörigkeitsfunktionen, Zugehörigkeitsfunktion mit S-Flanken (···) für $\alpha = 0,05$; $\beta = 0,4$; $\gamma = 0,6$ und $\delta = 0,95$

In den Übergangsbereichen liefern mehrere Regeln Beiträge zur resultierenden Ausgangsgröße y . Dadurch erhält man einen stetigen Verlauf von y beim Übergang zwischen benachbarten Partitionen.

5.1.2 Referenz-Fuzzy-Mengen und die Koeffizienten der Konklusionsfunktionen

Für den Fall von Konklusionspolynomen nach (3.7), die linear in den einzelnen Eingangsgrößen sind, soll der Zusammenhang zwischen der Wahl der Referenz-Fuzzy-Mengen und der Polynomkoeffizienten $p_i(j)$, $j \in \{1, \dots, m\}$ bei eindimensionaler orthogonaler Partitionierung untersucht werden. Im Falle treppenförmiger Zugehörigkeitsfunktionen nach (5.2) zeigt sich, daß die Verschiebung einer Referenz-Fuzzy-Mengenflanke unter Beibehaltung des Konklusionspolynoms sich auch durch Beibehaltung der Referenz-Fuzzy-Menge und Änderung des Konklusionspolynoms ausdrücken läßt. Betrachtet wird zuerst der Fall

einer gleichmäßigen Verschiebung zweier benachbarter Referenz–Fuzzy–Mengen (lückenlos aneinandergrenzende Treppenfunktionen). Da überlappungsfreie Referenz–Fuzzy–Mengen vorausgesetzt werden sollen, gilt vor der Verschiebung

$$u(\mathbf{x} \in \mathbb{M}_i) = \frac{u_i \omega_i}{\omega_i} \Big|_{\mathbf{x} \in \mathbb{M}_i} = u_i \Big|_{\mathbf{x} \in \mathbb{M}_i} \quad (5.7)$$

mit

$$\begin{aligned} \mathbb{M}_i &= \{\mathbf{x} \in \mathbb{D}_x^m \mid \mathbf{x} = [x_1, \dots, x_j, \dots, x_m]^T, \\ &\quad x_{ug,j} \leq x_j \leq x_{og,j} \quad \forall j \in \{1, \dots, m\}\} \end{aligned} \quad (5.8)$$

und dem zugehörigen Konklusionspolynom

$$u_i = p_0^i + p_1^i x_1 + \dots + p_m^i x_m \quad . \quad (5.9)$$

Eine lineare Verschiebung zweier benachbarter Referenz–Fuzzy–Mengen um x_c in der Eingangsgröße x_j führt zu einer resultierenden Ausgangsgröße von

$$\tilde{u}(\mathbf{x} \in \tilde{\mathbb{M}}_i) = \frac{u_i \tilde{\omega}_i}{\tilde{\omega}_i} \Big|_{\mathbf{x} \in \tilde{\mathbb{M}}_i} = u_i \Big|_{\mathbf{x} \in \tilde{\mathbb{M}}_i} \quad (5.10)$$

mit

$$\begin{aligned} \tilde{\mathbb{M}}_i &= \{\mathbf{x} \in \mathbb{D}^m \mid \mathbf{x} = [x_1, \dots, x_j, \dots, x_m]^T, \\ &\quad x_{ug,j} + x_c \leq x_j \leq x_{og,j} + x_c \quad \forall j \in \{1, \dots, m\}\} \quad . \end{aligned} \quad (5.11)$$

Dabei ist

$$\begin{aligned} \tilde{u}_i &= p_0^i + p_1^i x_1 + \dots + p_j^i (x_j + x_c) + \dots + p_m^i x_m \\ &= u_i + p_j^i x_c \quad . \end{aligned} \quad (5.12)$$

Dieses Ergebnis lässt sich aber auch durch ein geändertes Konklusionspolynom unter Beibehaltung der ursprünglichen Zugehörigkeitsfunktionen erreichen. Für eine Wahl

$$\hat{u}_i = \hat{p}_0^i + \sum_{j=1}^m \hat{p}_j^i x_j \quad (5.13)$$

mit

$$\hat{p}_0^i = p_0^i + p_j x_c \quad (5.14)$$

gilt

$$\hat{u} = \frac{\hat{u}_i \omega_i}{\omega_i} \Big|_{\mathbf{x} \in Q_i} = \hat{u}_i \Big|_{\mathbf{x} \in \tilde{Q}_i} = \tilde{u} \quad . \quad (5.15)$$

Also unterscheiden sich u_i und \tilde{u}_i nur im statischen Anteil. Damit wurde gezeigt, daß eine Verschiebung zweier benachbarter Referenz–Fuzzy–Mengen einer Eingangsgröße des Fuzzy–Systems durch die Änderung des konstanten Anteils der Konklusion substituiert

werden kann. Das gilt auch für nichtlineare Funktionen f_i , die aber in der betrachteten Koordinate linear sind. Durch wiederholtes Anwenden dieses Verfahrens können auch die Verschiebungen mehrerer Referenz–Fuzzy–Mengen durch die geeignete Wahl des konstanten Anteils im Konklusionspolynom substituiert werden. Bei den an den Grenzen eines Definitions–Intervalls liegenden Referenz–Fuzzy–Mengen ist dabei zu beachten, daß diese nicht verschoben, sondern in Abzissenrichtung gestaucht oder gestreckt werden müssen, damit der Definitionsbereich \mathbb{D}_N am betreffenden Rand weiterhin von der Referenz–Fuzzy–Menge voll abgedeckt wird.

Bei Zugehörigkeitsfunktionen mit „weichen“ Flanken (z. B. trapezoide) führt die Änderung einer oder zweier benachbarter Referenz–Fuzzy–Mengen zu Änderungen der Erfülltheitsgrade ω der betroffenen Regeln, die nicht durch geeignete Änderung der Konklusionsfunktionen, die eine beliebige Form besitzen dürfen, kompensiert werden können. Dies ist bei nichtlinearen Flanken der Zugehörigkeitsfunktion und in den Eingangsgrößen linearen Konklusionspolynomen offensichtlich. Falls nur die Form aber nicht die Lage einer Flanke modifiziert wird, kann bei Zulassung eines Polynoms höherer Ordnung keine generelle Aussage gemacht werden. Hat man z. B. trapezoide Zugehörigkeitsfunktionen, so kann weder mit einem in den Eingangsgrößen linearen Konklusionspolynom noch bei Polynomen höherer Ordnung die nichtlineare Änderung der Zugehörigkeitsfunktion bei Änderung des/der Knickpunkte(s) durch eine Änderung des Konklusionspolynomes substituiert werden. Selbst bei Beschränkung auf den reinen Flankenbereich ($Fl = \{x_i \mid 0 < \mu_1(x_j) < 1 \wedge 0 < \mu_2(x_j) < 1 \wedge 0 < \tilde{\mu}_1(x_j) < 1 \wedge 0 < \tilde{\mu}_2(x_j) < 1\}$ in Bild 5.2) kann gezeigt werden, daß die Substituierbarkeit nicht immer gilt (Anhang A).

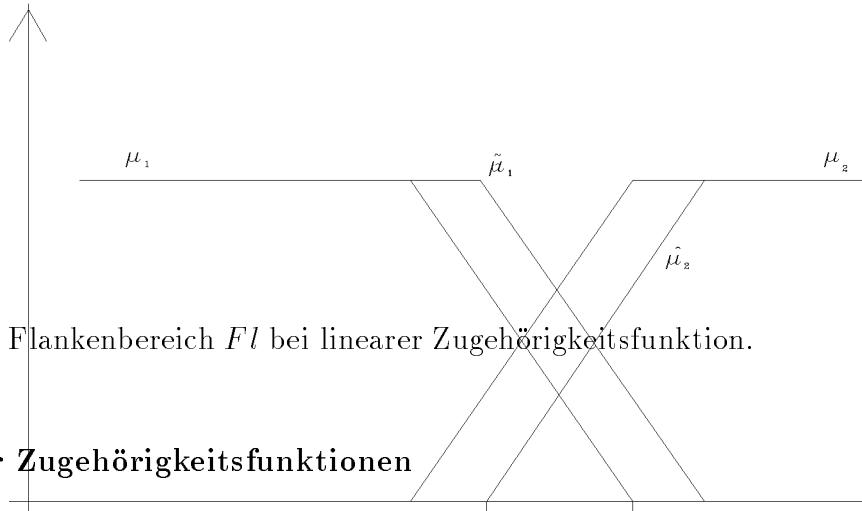


Bild 5.2: Reiner Flankenbereich Fl bei linearer Zugehörigkeitsfunktion.

5.1.3 Anpassung der Zugehörigkeitsfunktionen

Die wahrscheinlich einfachste Form der Partitionierung des Eingangsgrößenraumes besteht in einer gleichförmigen orthogonalen Teilung über Referenz–Fuzzy–Mengen gleicher Form, die nur innerhalb ihres Definitionsbereiches verschoben sind. Dabei wird die Systemanpassung einzig über die Anzahl der Referenz–Fuzzy–Mengen sowie eventuell über deren Form (z. B. Flankenform) vorgenommen. Nachteilig ist hierbei, daß die System– bzw. Datenstruktur nicht direkt berücksichtigt wird. Dadurch werden i. allg. unnötig viele

Regeln, vergleichbar einer unnötig hohen Systemordnung bei konventionellen Modellen, verursacht. Außerdem kann es Probleme geben, wenn zu wenig Daten pro Partition zu deren Identifikation zur Verfügung stehen.

Bei einer heuristischen Partitionierung kann i. allg. die Regelanzahl gesenkt werden. Wegen fehlender Systematik empfiehlt sich diese Vorgehensweise nicht generell. Es existieren aber auch Verfahren, die Referenz–Fuzzy–Mengen systematisch zu modifizieren. Im folgenden sollen für einige dieser Algorithmen kurz die zugrundeliegenden Ideen beschrieben werden: Sugeno und Kang (1988) stellen einen Algorithmus vor, der die Anzahl, Lage und Form trapezoider Zugehörigkeitsfunktionen ändert (Identifikation Teil Ib, II). Optimierungskriterium ist das *Unbiasedness–Kriterium* (Sugeno und Kang 1986). Startpunkt ist ein Modell mit einer Regel, also ein lineares Modell. Bei diesem wird der Defitionsbereich der 1. Eingangsgröße in 2 Teilbereiche aufgeteilt. Dann werden die Prämissenparameter, d.h. die Parameter der trapezoiden Zugehörigkeitsfunktionen identifiziert. Dabei werden die Stützpunkte der Zugehörigkeitsfunktionen so variiert, daß die Differenz zwischen Modellausgang und Meßwerten minimal wird. Anschließend werden die Konklusionsparameter bestimmt. Sugeno und Kang (1988) geben hierzu kein Verfahren an. Hebisch (1992) setzte hierzu das ungewichtete rekursive Verfahren der kleinsten Fehlerquadrate ein. Letztlich wird mit dem Unbiasedness–Kriterium die Güte des Modells bestimmt. Im folgenden Schritt wird statt der 1. die 2. Eingangsgröße und dann alle weiteren entsprechend der oben beschriebenen Form untersucht. Das Modell mit dem geringsten Wert des Unbiasedness–Kriteriums wird weiterverwendet. Ausgehend von diesem Modell werden wieder sukzessive alle Eingangsgrößendefinitionsbereiche einmal mehr unterteilt. Das Modell mit dem geringsten Wert des Gütekriteriums wird ausgewählt. Das Verfahren terminiert, wenn der Wert des Gütekriteriums wieder ansteigt. Die von Hebisch (1992) vorgenommene Anwendung des Algorithmus auf ein technisches hydraulisches System wies allerdings Probleme bei der Konvergenz der Strukturidentifikation auf, deshalb wurde die Prämissenidentifikation heuristisch durchgeführt.

Um die Anzahl der zu identifizierenden Parameter zu reduzieren, beschränken Sugeno und Tanaka (1991) die Anzahl der Freiheitsgrade bei der Festlegung der Zugehörigkeitsfunktion: Die Summe der Zugehörigkeitsgrade zweier benachbarter Zugehörigkeitsfunktionen ist gleich eins. Dadurch reichen zur Festlegung zweier benachbarter Flanken zwei statt vier Parameter aus. Ein ähnliches Verfahren, das orthogonal partitioniert und die beste Partitionierung auswählt (triangulare Zugehörigkeitsfunktionen), stellen Araki u. a. (1991) vor.

Sugeno und Yasukawa (1993) führen die Strukturidentifikation Teil Ib wie Sugeno und Kang (1988) durch, nutzen aber den Fuzzy– c –Means–Clusteralgorithmus für Teil II (die Bezeichnungsweise wurde in Abschnitt 1 eingeführt). Die Cluster werden anschließend mit Hilfe eines Algorithmus (iterativ) auf trapezoide Zugehörigkeitsfunktionen abgebildet. So mit werden Teil IIa und IIb in einem Schritt abgearbeitet.

Tan u. a. (1994) benutzen eine Modellstruktur mit gaußförmigen Referenz–Fuzzy–Mengen und linearen Konklusionspolynomen. Ihr zweistufiger Algorithmus geht von einer Initialisierung der Zugehörigkeitsfunktionen und des Konklusionspolynoms aus. Im ersten Schritt werden mit Hilfe des Verfahrens der kleinsten Fehlerquadrat die Konklusionsparameter bestimmt, wobei die Zugehörigkeitsfunktionen unverändert bleiben. Im 2. Schritt werden die Konklusionspolynomparameter festgehalten und mit Hilfe eines Gradientenverfahrens (Newton-Verfahren) die Zugehörigkeitsfunktionsparameter optimiert. Dabei wird ein quadratisches Fehlerkriterium zu Grunde gelegt. Die beiden Schritte werden wiederholt, bis die gewünschte Modellgenauigkeit erreicht ist oder das Verfahren konvergiert. Zur Anwendbarkeit des Gradientenverfahrens müssen die Partialprämissen multiplikativ miteinander verknüpft werden und die Zugehörigkeitsfunktionen in ihren Parametern stetig differenzierbar sein. Deshalb verwenden Tan u. a. (1994) gaußförmige Zugehörigkeitsfunktionen.

5.2 Partitionierung mit mehrdimensionalen Zugehörigkeitsfunktionen

Das hier vorgestellte neue Verfahren identifiziert Regeln mit einer Struktur nach (3.10). Bei vorgegebener Abstandsnorm sind nur die Partitions- bzw. Clusterschwerpunkte zu bestimmen. Dabei besteht hier die Möglichkeit, von orthogonalen Partitionen auszugehen und deren Mittelpunkte als Partitionsschwerpunkte zu nutzen. Ein sehr systematisches und automatisiertes Vorgehen ergibt sich bei Anwendung der Clusteranalyse zur Bestimmung der Clusterschwerpunkte (als Partitionsschwerpunkte nicht orthogonaler Partitionen). Im 4. Abschnitt wurden dazu bereits mehrere Verfahren vorgestellt. Das Ergebnis einer im Eingangsdatenraum durchgeführten Clusterung sind bei vorgegebener Anzahl der Cluster und einer Abstandsnorm D die Clusterschwerpunkte v_1, \dots, v_c . Somit sind die Prämissen der Fuzzy–Regeln gemäß (3.10) identifiziert. Anschließend erfolgt die in Abschnitt 3 beschriebene Konklusionsparameteridentifikation über das gewichtete rekursive Verfahren der kleinsten Fehlerquadrate. Wie bereits im Abschnitt 4 erläutert wurde, liefern die Clusterverfahren eine im Sinne der jeweils zu Grunde liegenden Abstandsnorm optimierte Partitionierung des Datenraumes. Das bedeutet allerdings nicht, daß die tatsächlich vorliegenden Strukturen oder eine im Sinn der Modellbildung optimale Partitionierung identifiziert wird. Das letztere, aus der Wahl der Gütekriterien resultierende, Problem kann dadurch abgeschwächt werden, daß *nach* der Clusterung eine Nachjustierung der Clusterschwerpunkte durchgeführt wird. Das Gütekriterium hierbei ist das gleiche wie bei der Modellbeurteilung, so daß eine Modellmodifikation in die gewünschte Richtung erfolgt. Im folgenden Abschnitt wurde dazu eine einfache iterative Variations der Clusterschwerpunkte um die von der Clusterung gelieferten guten Anfangswerte durchgeführt. Bei diesem Iterationsverfahren werden die Clusterschwerpunkte in alle Koordinatenrichtungen mit von Iteration zu Iteration abnehmender Schrittweite verschoben.

Verfeinerte Verfahren könnten hier sicher noch Verbesserungen der Güte und eine Steigerung der Rechengeschwindigkeit erzielen. Allerdings sollte das Modell nicht zu fein auf einen speziellen Datensatz optimiert werden, da sonst die Robustheit des Modells abnehmen kann. Dieser Algorithmus geht von einer Situation aus, in der die Eingangsgrößen (z. B. durch physikalische Überlegungen, Expertenwissen, Beschränkung der Anzahl der Modellparameter) und die Signalverzögerungen (z. B. durch Korrelationsanalyse) festgelegt wurden (siehe auch Abschnitt 1). Der folgende Abschnitt beschreibt die Anwendung dieses Verfahrens auf zwei nichtlineare Testsysteme.

6 Beispiele zur Identifikation

In diesem Abschnitt wird die Identifikation von funktionalen Fuzzy-Modellen nach dem zuvor beschriebenen Algorithmus mit mehrdimensionalen Zugehörigkeitsfunktion für zwei nichtlineare Testsysteme beschrieben. Das erste ist ein Testsystem, für das bereits einige Vergleichsergebnisse für relationale Fuzzy-Modelle vorliegen (Küpper 1994a, Suprijadi 1994). Das zweite Beispiel, die von Box und Jenkins (1970) veröffentlichten Meßdaten für einen Gasofen, stellt ein Benchmark-Problem dar. Beim WRLS wurde $\mathbf{P}_w(0) = c\mathbf{I}$ mit $c = 10^4$ und $\tilde{\theta}(0) = \mathbf{0}$ gewählt. Als Abstandskriterium fand bei den Clusterungen die euklidische Norm und als Terminierungsnorm beim FCM die gemäß (2.4) erweiterte Hölder-Norm mit $p = 1$

$$\|\mathbf{U}^{(l+1)} - \mathbf{U}^{(l)}\|_T = \sum_{i=1}^c \sum_{k=1}^N |\mu_{ik}^{(l+1)} - \mu_{ik}^{(l)}| \leq \varepsilon \quad (6.1)$$

Verwendung. Beim Gustafson und Kessel-Algorithmus wurde eine feste Anzahl von 10 Iterationen vorgegeben. Als Gütekriterium wird der mittlere quadratische Fehler verwendet. Ein Vergleich funktionaler Fuzzy-Modelle untereinander oder mit relationalen Fuzzy-Modellen ist schwierig. Für eine näherungsweise Abschätzung soll die Anzahl der Modellparameter als Maßstab angesetzt werden. Die in diesem Bericht beschriebenen echt-mehrdimensionalen Zugehörigkeitsfunktionen sind eindeutig durch ihre Clusterschwerpunkte festgelegt. Dabei ist ein Clusterschwerpunkt ein Vektor, wobei die Komponentenzahl der Dimension des Eingangsdatenraumes entspricht. Deshalb sollte bei jedem Clusterschwerpunkt pro Dimension ein Parameter angesetzt werden.

6.1 Modellierung eines nichtlinearen Testsystems

In diesem Abschnitt wird die Ein-/Ausgangsmodellierung für Testdaten beschrieben, die mit dem nichtlinearen zeitdiskreten System (Küpper 1994a)

$$y(k) = 0,1y(k-1)u(k-2) - 0,2y(k-2) + 0,4u(k-1) \cos(0,8y(k-1)) + 0,3u^2(k-1) \quad (6.2)$$

gewonnen wurden (Anfangswerte $y(0) = y(1) = 0$). Für das Testsignal wurde ein im Intervall $[-1; 1]$ gleichverteiltes Zufallssignal generiert. Dann wird $u(k)$ dadurch generiert, daß jede Zufallszahl für 6 Abtastzeitpunkte in Folge konstant gehalten wird. Bild 6.1 zeigt den Verlauf der Eingangsgröße $u(k)$ und der Ausgangsgröße $\tilde{y}(k+1)$, die nach

$$\tilde{y}(k) = \frac{y(k) - 0,2452}{0,4750} \quad (6.3)$$

aus $y(k)$ berechnet wurde. Diese Transformationsvorschrift dient dazu, den normierten Eingangsgrößenraum besser auszunutzen. Während die Zufallszahlen von $u(k)$ dies bereits gewährleisten, ist das für $y(k)$ nicht der Fall. Die Transformation ist gerade bezüglich der Clusterverfahren sinnvoll. Da absolute Abstandsdifferenzen bewertet werden, könnte es

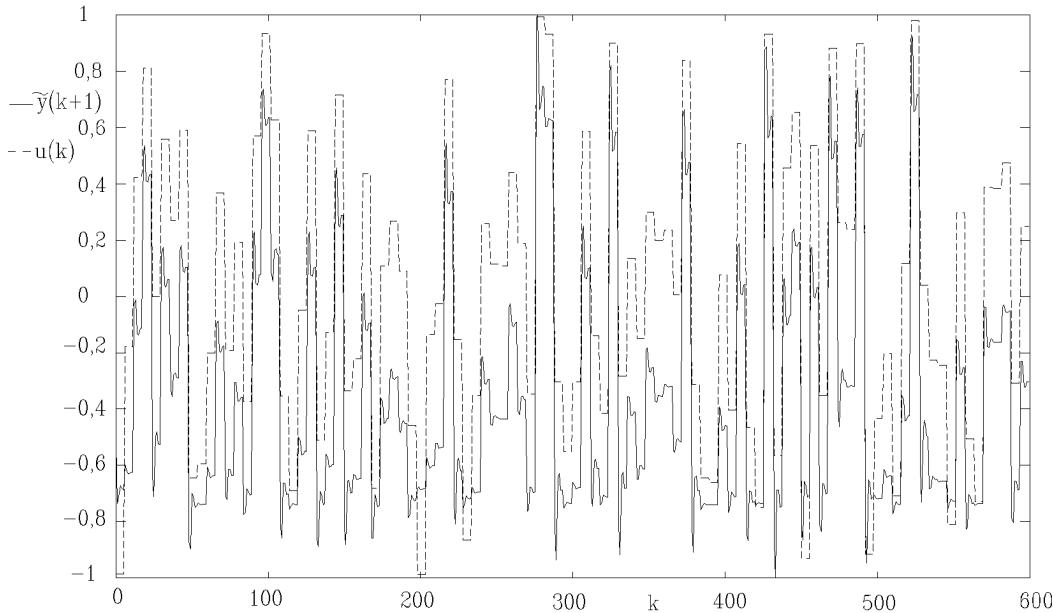


Bild 6.1: Eingangsgröße $u(k)$ und Ausgangsgröße $\tilde{y}(k+1)$ des nichtlinearen Systems

sonst unter Umständen zu einer ungünstigen Über- oder Unterbewertung einzelner Eingangsgrößen kommen.

Im folgenden werden Fuzzy-Modelle vorgestellt, die zur Prädiktion von $y(k)$ die Eingangsgrößen $u(k-1)$, $\tilde{y}(k-1)$ und $\tilde{y}(k-2)$ verwenden.

Vergleichsergebnisse

Suprijadi (1994) identifiziert für das oben beschriebene System selbstlernende Fuzzy-Relational-Modelle. Das beste Modell mit gleichverteilten Referenz-Fuzzy-Mengen (gleichschenklige Dreiecke gleichen Flächeninhalts, benachbarte schneiden sich bei $\mu = 0,5$, drei Referenz-Mengen pro Eingang, zwei für die Konklusion) und 54 Parametern (die Regelgewichte) erreichte eine Güte (mittlerer quadratischer Fehler) von $pi = 7,56 \cdot 10^{-5}$ (600 Iterationen). Bei gleichzeitiger Optimierung der Referenz-Fuzzy-Mengen-Stützpunkte erzielte das beste Modell (92 Parameter) eine Güte von $pi = 1,16 \cdot 10^{-5}$ (600 Iterationen). Dabei wurden bei den Referenz-Fuzzy-Mengen Flankenbreiten (separat) sowie Lage und Höhe der Dreiecksspitze variiert (4 Parameter pro Referenz-Fuzzy-Menge). Küpper (1994) verwendet einen anderen Algorithmus zur Identifikation der gleichen Modellstrukturen als Suprijadi und erreicht dabei für das Modell mit 54 Parametern $pi = 7,89 \cdot 10^{-5}$ (500 Iterationen). Bei gezielter (heuristischer) Änderung der Partitionierung bzgl. $u(k-1)$ (dann 4 Referenz-Fuzzy-Mengen) wird eine Güte von $pi = 1,45 \cdot 10^{-5}$ für ein Modell mit nunmehr 72 Parametern (500 Iterationen) bestimmt.

Eigene Ergebnisse

Bei Anwendung des FCM ($\varepsilon = 100$ in (6.1)) zeigen sich eindeutige Tendenzen bzgl. der Wahl des Unschärfeparameters ν und der Clusteranzahl c . Wie Bild 6.2 zeigt, steigt die Modellgüte mit steigendem c und fallenden ν . Bei diesen Ergebnissen für Einschrittprädiktion wurde

$$pi = \frac{1}{N-1} \sum_{k=2}^{N=600} (y_{mes}(k) - y_{mod}(k))^2 \quad (6.4)$$

betrachtet. Bei den Ergebnissen der rekursiven Modellauswertung in Tabelle 6.1 wurden nur die *vollständig* rekursiv prädizierten Ausgangsgrößen berücksichtigt:

$$pi = \frac{1}{N-3} \sum_{k=4}^{N=600} (y_{mes}(k) - y_{mod}(k))^2 \quad . \quad (6.5)$$

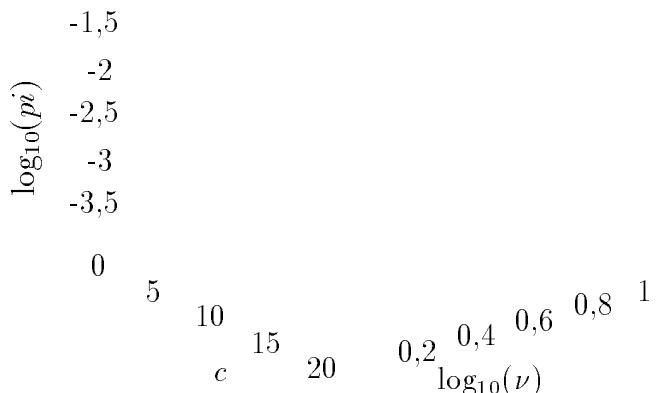


Bild 6.2: Logarithmisch aufgetragene Modellgüte in Abhängigkeit von der Clusteranzahl c und dem Logarithmus des Unschärfeparameters $\log_{10}(\nu)$ bei Anwendung des FCM (Einschrittprädiktion)

Bei Werten von ν nahe bei 1 ergeben sich allerdings numerische Probleme wegen der begrenzten Genauigkeit des verwendeten Computers (80486 CPU). Eine Wahl von $\nu = 1,13$ führt beim FCM zu ausreichenden numerischen Reserven. Für dieses ν erhält man z. B. die in Tabelle 6.1 exemplarisch aufgeführten Ergebnisse (Spalte „FCM“).

c	Anzahl der Parameter	FCM	FCM + Optimierung	Modell- auswertung	
		pi	pi		
7	49	$1,649 \cdot 10^{-4}$	$6,626 \cdot 10^{-5}$	273	Einschritt
8	56	$1,373 \cdot 10^{-4}$	$5,122 \cdot 10^{-5}$	5881	
9	63	$9,283 \cdot 10^{-5}$	$5,652 \cdot 10^{-5}$	341	
10	70	$8,287 \cdot 10^{-5}$	$2,815 \cdot 10^{-5}$	314	
11	77	$8,743 \cdot 10^{-5}$	$1,899 \cdot 10^{-5}$	207	
12	84	$8,065 \cdot 10^{-5}$	$2,266 \cdot 10^{-5}$	353	
13	91	$5,075 \cdot 10^{-5}$	$1,559 \cdot 10^{-5}$	294	
7	49	$1,685 \cdot 10^{-4}$	$6,613 \cdot 10^{-5}$	141	Rekursiv
8	56	$1,422 \cdot 10^{-4}$	$3,535 \cdot 10^{-5}$	133	
9	63	$9,479 \cdot 10^{-5}$	$3,827 \cdot 10^{-5}$	336	
10	70	$8,855 \cdot 10^{-5}$	$1,914 \cdot 10^{-5}$	260	
11	77	$9,440 \cdot 10^{-5}$	$2,422 \cdot 10^{-5}$	224	
12	84	$8,810 \cdot 10^{-5}$	$2,434 \cdot 10^{-5}$	328	
13	91	$6,961 \cdot 10^{-5}$	$2,409 \cdot 10^{-5}$	353	

Tabelle 6.1: Einige Ergebnisse für Identifikation mit FCM für $\nu = 1, 13$ und für Modelle mit nachoptimierten Clusterschwerpunkten

Bild 6.3 zeigt exemplarisch einen Ausschnitt aus Bild 6.1. Dargestellt sind zusätzlich die einschrittprädizierten Werte für $\tilde{y}(k+1)$ bei Clusterung über den FCM mit $\nu = 1, 13$ für $c = 8$ (führt zu 10 Iterationen des FCM) und $c = 13$ (führt zu 25 Iterationen des FCM). Die zugehörigen Gütewerte enthält Tabelle 6.1. Die rekursive Prädiktion wird hier graphisch nicht dargestellt, da die Ergebnisse bei Einschrittprädiktion und rekursiver Prädiktion nur wenig voneinander abweichen (vergleiche numerische Werte in Tabelle 6.1). Dies liegt daran, daß $u(k)$ die dominante Eingangsgröße des Modells ist. Somit nimmt eine Rückführung des Modellausgangs nur Einfluß auf die rezessiven Eingangsgrößen $\tilde{y}(k-1)$ und $\tilde{y}(k-2)$.

Bei 3 Eingangsgrößen ergeben sich 3 Koordinaten der Clusterschwerpunkte und 4 Konkusionspolynomparameter pro Regel. Damit folgt die Gesamtanzahl der Modellparameter zu $7c$. Die oben angegebenen relationalen Fuzzy-Modelle mit 54 Parametern sind etwa um einem Faktor 2 besser als das nicht nachverfeinerte Funktionalmodell mit 56 Parametern. Das Modell von Suprijadi mit 92 Parametern ist etwa um einen Faktor 7 besser als das nicht nachverfeinerte Funktionalmodell mit 91 Parametern. Dies überrascht nicht, da die Referenz-Fuzzy-Modelle auf das Gütemaß hin optimiert werden, nach dem auch die Modellgüte bewertet wird. Bei der Identifikation der Zugehörigkeitsfunktionen über den FCM ist das, wie bereits oben erläutert, nicht der Fall. Diesen Aspekt illustriert die Anwendung des in Abschnitt 5.2 beschriebenen iterativen Optimierungsverfahrens für die

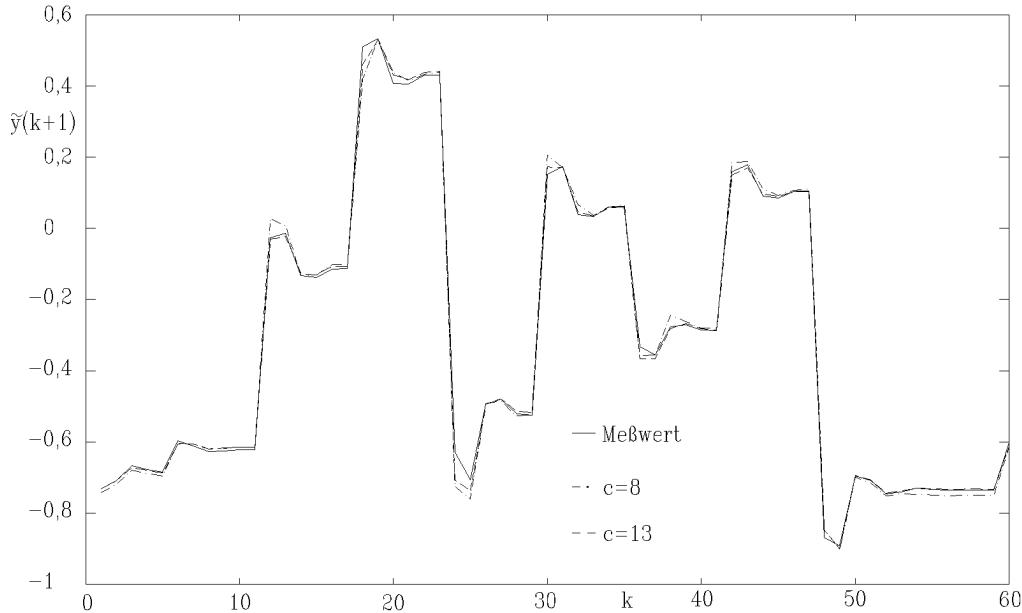


Bild 6.3: Meßwerte und Einschrittprädiktion für über den FCM mit $\nu = 1, 13$ und $c = 8$ bzw. $c = 13$ generierte Modelle

Lage der Clusterschwerpunkte. Durch die Anwendung des Optimierungsverfahrens kann die Güte der Funktional-Modelle deutlich verbessert werden (siehe Tabelle 6.1). Dann ist das funktionale Modell mit 56 Parametern etwa vergleichbar mit den relationalen Modellen mit 54 Parametern und das relationale Modell mit 92 Parametern ist nur noch etwa um einen Faktor 1,5 besser als das funktionale Modell mit 91 Parametern.

Bei der Clusterung werden bei der Initialisierung alle Clusterschwerpunkte nur in der Richtung der ersten Eingangsgröße $u(k-1)$ verteilt. Damit folgt

$$\mathbf{v}_i^{(0)} = \left[-1 + \frac{1 + 2(i-1)}{2c}; 0; 0 \right]^T \quad \forall \quad i = 1, \dots, c \quad . \quad (6.6)$$

Aus den Clusterschwerpunkten lassen sich alle $\mu_{ik}^{(0)}$ nach (3.15) berechnen. Dieser Ansatz wurde auf Grundlage von Wissen über die Struktur des Datensatzes vorgenommen. Vergleiche mit anderen Vorinitialisierungen der Clusterschwerpunkte, bei denen diese nicht auf einer Geraden lagen, zeigten nur geringfügige Abweichungen der Ergebnisse nach den Clusterungen. Dieses Resultat kann so allerdings nicht verallgemeinert werden. Da viele Clusterverfahren in einem lokalen Minimum der Zielfunktion konvergieren, kann die Vorinitialisierung sehr wohl einen Einfluß auf das Ergebnis der Clusterung nehmen. Auch eine Beeinflussung der Konvergenzgeschwindigkeit ist möglich. Für eine Wahl von $\varepsilon = 100$ finden aber beim FCM bei diesem System nur einige wenige (< 10) Iterationen statt, so daß die Konvergenz sehr schnell eintritt.

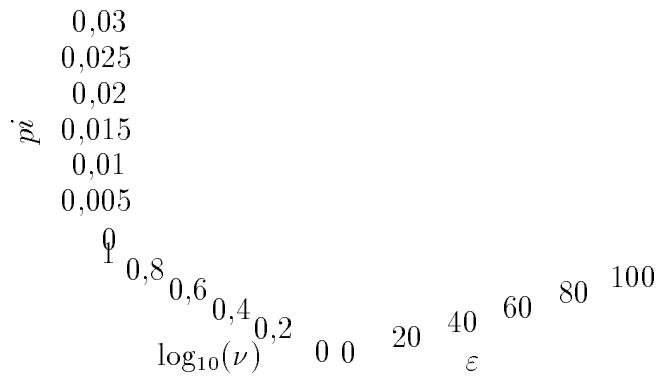


Bild 6.4: Modellgüte in Abhängigkeit vom Logarithmus des Unschärfeparameters $\log_{10}\nu$ und Terminierungsgrenzwert ε für $c = 10$ beim FCM (Einschrittprädiktion)

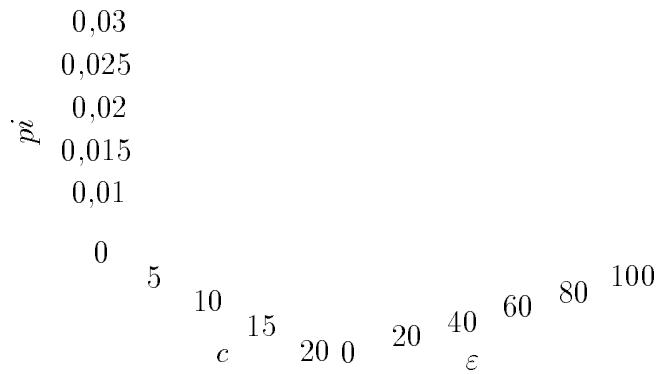


Bild 6.5: Modellgüte in Abhängigkeit von Regelanzahl c und Terminierungsgrenzwert ε für $\nu = 2, 0$ beim FCM (Einschrittprädiktion)

Die Wahl von ε in (6.1) hat keinen wesentlichen Einfluß auf die Modellgüte. Zur Illustration zeigt Bild 6.4 die Güte von Modellen mit $c = 10$ Regeln in Abhängigkeit von $\varepsilon \in \{10, 20, \dots, 100\}$ und $1, 13 \leq \nu \leq 10$. Bild 6.5 zeigt exemplarisch die Abhängigkeit der Modellgüte von der Regelanzahl c und dem Terminierungsgrenzwert ε für festes $\nu = 2,0$. Es bestätigt sich wieder der vernachlässigbare Einfluß der Wahl der Terminierungsgrenze auf die Modellgüte.

Die Anwendung des Algorithmus von Gustafson und Kessel führt zu den gleichen Tendenzen bei der Wirkung der Parameter ν und c , wie Bild 6.6 zeigt. Beim Gustafson und

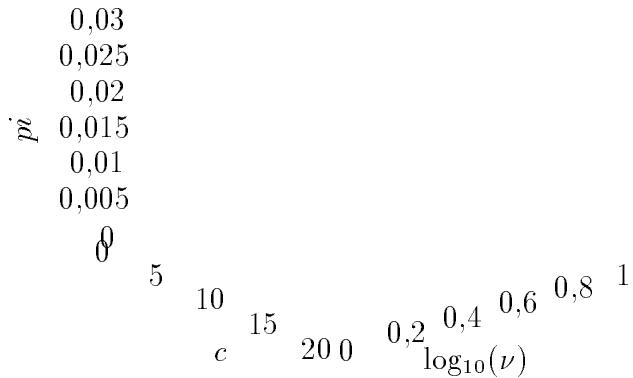


Bild 6.6: Modellgüte in Abhängigkeit von der Clusteranzahl c und dem Logarithmus des Unschärfeparameters $\log_{10}\nu$ bei Anwendung des Gustafson und Kessel–Algorithmus (Einschrittprädiktion)

Kessel–Algorithmus wurde als Terminierungsgrenze ein feste Iterationszahl vorgegeben. Diese hat allerdings genauso wie die Terminierungsgrenze über Vorgabe des ε in (6.1) keinen wesentlichen Einfluß auf die Modellgüte, wie die Bilder 6.7 und 6.8 illustrieren.

Mit dem FCM konnten höhere Güten erreicht werden. Während der FCM bei steigender Clusteranzahl c schnell, beim Unschärfeparameter ν dagegen erst bei sehr kleinen Werten zu höheren Güten führte, ist die Tendenz beim Gustafson und Kessel–Algorithmus vertauscht: Eine Verringerung von ν führt schnell, eine Erhöhung von c dagegen nur langsam zu höheren Güten. Dies kann auf die „bogenförmigen“ Strukturen im ausgewerteten Datensatz zurückgeführt werden. Da sich die lokalen Abstandsnormen hierauf einstellen, ist der Einfluß der Clusteranzahl nicht so dominant. Wichtiger hierbei erscheint, daß die

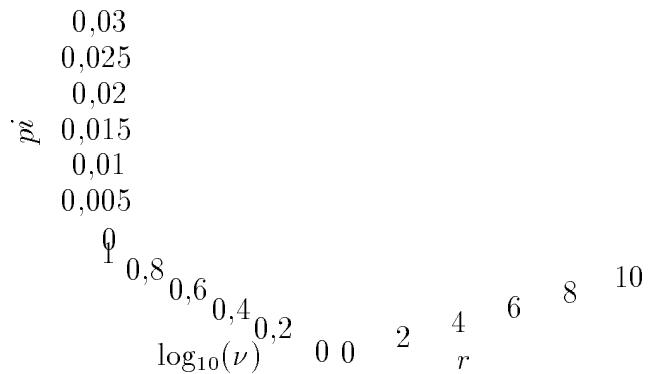


Bild 6.7: Modellgüte in Abhängigkeit von $\log_{10}\nu$ und Anzahl r der Iterationen für $c = 10$ beim Gustafson und Kessel–Algorithmus (Einschrittprädiktion)

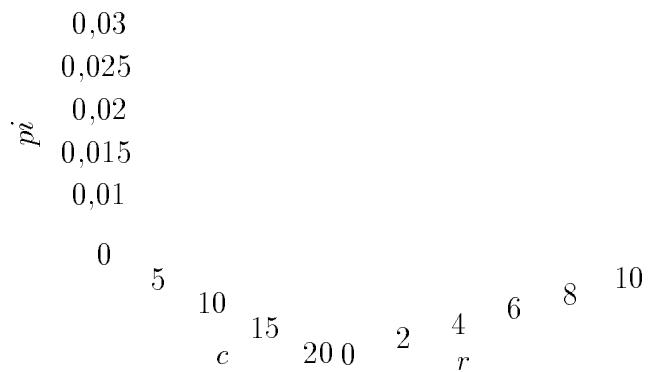


Bild 6.8: Modellgüte in Abhängigkeit von Regelanzahl c und Anzahl r der Iterationen für $\nu = 2, 0$ beim Gustafson und Kessel–Algorithmus (Einschrittprädiktion)

Unschärfe der bereits formangepaßten Cluster richtig gewählt wird.

Anzumerken ist, daß die Clusterung mit global einheitlicher Wahl von $\rho_1 = 1$ durchgeführt wurde. Bei lokal angepaßter Wahl der ρ_i sind wahrscheinlich noch Gütesteigerungen möglich. Allerdings existieren hierzu keine systematischen Vorgehensweisen. Für eine sinnvolle lokale Anpassung der ρ_i müßten A-priori-Informationen über die Cluster vorhanden sein, die im allgemeinen nicht vorliegen.

6.2 Modellierung eines Gasofens

In diesem Abschnitt soll die Fuzzy-Modellbildung für einen Gasofen beschrieben werden. Die Ein-/Ausgangsmodellbildung für die von Box und Jenkins (1970) angegebenen 296 Testdatenpaare ist ein Benchmark Problem. Bild 6.9 zeigt den Meßdatenverlauf. Modelliert wird die CO_2 -Emission eines Ofens ($y(t)$ [%]). Das physikalische Stellsignal

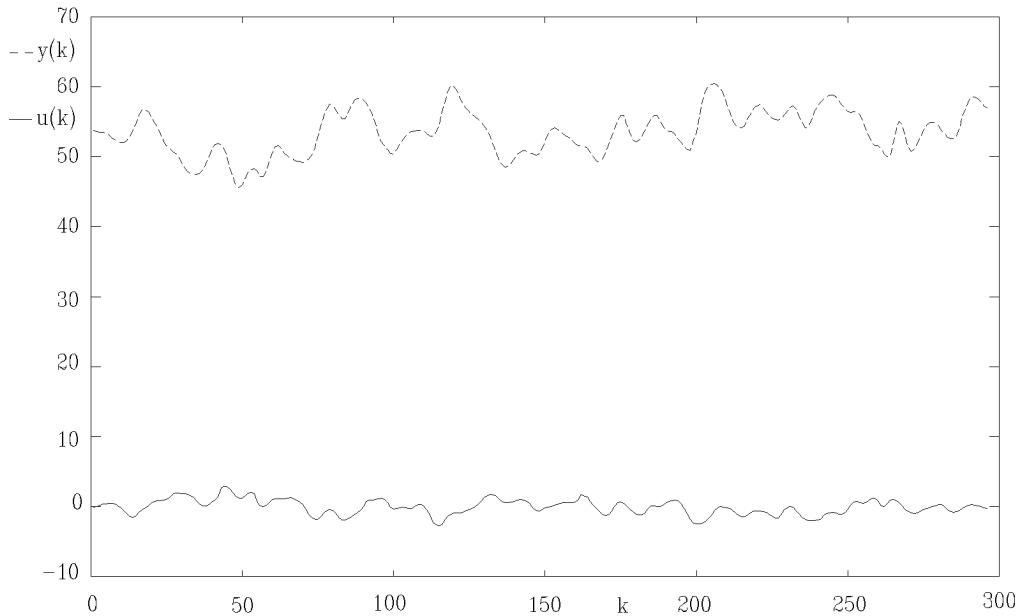


Bild 6.9: Testdaten für den Gasofen nach Box und Jenkins (1970): Eingangssignal $u(k)$ und Ausgangssignal $y(k)$

$u(t)$ steuert einen Methangasvolumenstrom

$$q(t) = (0,6 - 0,04 u(t)) 4,719 \cdot 10^{-4} \frac{\text{m}^3}{\text{s}} . \quad (6.7)$$

Das im Bereich $[-2,716; 2,834]$ liegende Stellsignal führt zu einem Volumenstrom $q(t) \in [2,296 \cdot 10^{-4}; 3,344 \cdot 10^{-4}] \text{ m/s}$. Als Testsignal $u(t)$ wurde ein gefiltertes weißes Rauschen verwendet und die Daten mit $T = 9 \text{ s}$ abgetastet (Küpper 1993). Für die Identifikation wurden Ein- und Ausgangsgröße gemäß

$$\tilde{y}(t) = \frac{y(t) - 53,05}{7,46} \quad \text{und} \quad \tilde{u}(t) = \frac{u(t) - 0,059}{2,78} \quad (6.8)$$

normiert.

Vergleichsergebnisse

Bei Küpper (1993) sowie Sugeno und Yasukawa (1993) finden sich Vergleiche von linearen Modellen mit verschiedenen relationalen und funktionalen Fuzzy-Modellen. Besonders interessant für diesen Bericht ist ein Vergleich mit den beiden funktionalen Fuzzy-Modellen von Sugeno und Tanaka (1991). Deren erstes Modell (M1) prädiziert $y(k)$ mit 2 Regeln aus $\tilde{y}(k-1), \tilde{y}(k-2), \tilde{y}(k-3), \tilde{u}(k), \tilde{u}(k-1)$ und $\tilde{u}(k-2)$ mit einem mittleren quadratischen Fehler von $pi = 0,068$. Hierbei ist zu beachten, daß $\tilde{u}(k)$ zur Berechnung von $y(k)$ benutzt wird, also nicht wirklich ein Schritt in die Zukunft geschaut wird. Das zweite Fuzzy-Modell (M2) prädiziert $y(k)$ mit zwei Regeln aus $\tilde{y}(k-1)$ und $\tilde{u}(k-1)$ mit einer Güte von $pi = 0,359$. Bemerkenswert ist dabei, daß die Prämisse nur $\tilde{y}(k-1)$ bewerten. Die beiden verwendeten halbtrapezoiden Referenz-Fuzzy-Mengen addieren sich an jeder Stelle zu Eins und lassen sich deshalb (zusammen) über zwei Parameter festlegen. Für das erste Modell kommt man damit auf 2 Prämisse und 13 Konklusionsparameter, für M2 entsprechend auf 2 + 4 Parameter. Sugeno und Tanaka (1991) geben nicht an, ob die Modelle zur Einschrittprädiktion oder rekursiv ausgewertet werden. Während Sugeno und Tanaka die ersten 50 Meßdaten bei der Gütebewertung nicht betrachten,

$$pi = \frac{1}{N - 50} \sum_{k=51}^{N=296} (y_{mes}(k) - y_{mod}(k))^2 \quad , \quad (6.9)$$

werden bei den im folgenden beschriebenen eigenen Ergebnissen immer alle Daten berücksichtigt:

$$pi = \frac{1}{N - 1 - \tau} \sum_{k=2+\tau}^{N=296} (y_{mes}(k) - y_{mod}(k))^2 \quad . \quad (6.10)$$

Dabei ist $\tau = \max(\tau_1; \tau_2)$ in $y_{mod}(k) = f(y(k-1-\tau_1), u(k-1-\tau_2))$.

Eigene Ergebnisse

Eine Auswertung des Modells M2 von Sugeno und Tanaka (1991) liefert die etwas bessere Güte von $pi = 0,335$ (6 Parameter). Die leichte Abweichung ist darauf zurückzuführen, daß, wie oben angemerkt, Sugeno und Tanaka (1991) nicht alle Meßdaten zur Güteberechnung verwenden. Bei Verwendung von $\tilde{u}(k-3)$ statt $\tilde{u}(k-1)$ und Zulassung konstanter Anteile in den Konklusionspolynomen, läßt sich die Güte etwa um einen Faktor 2 auf $pi = 0,1475$ steigern (Modell mit 2 Regeln und 8 Parametern). Bei Auswertung des letztgenannten Modells über echt-mehrdimensionale Referenz-Fuzzy-Mengen läßt sich hier die Modellgüte nicht verbessern: Legt man die Partitionsschwerpunkte in die Schwerpunkte der beiden von Sugeno und Tanaka (1991) angegebenen, auf normierte Größen

umgerechneten, (eindimensionalen) Referenz–Fuzzy–Mengen bzgl. $\tilde{y}(k-1)$ (die andere Koordinate wird zu Null gesetzt)

$$v_1 = [47, 80; 0]^T \quad \text{und} \quad v_2 = [56, 13; 0]^T, \quad (6.11)$$

so folgt eine Güte von $pi = 0,1490$. Legt man die Partitionsschwerpunkte in die Mittelpunkte der konstanten Bereiche der beiden von Sugeno und Tanaka (1991) angegebenen, auf normierte Größen umgerechneten, (eindimensionalen) Referenz–Fuzzy–Mengen bzgl. $\tilde{y}(k-1)$ (die andere Koordinate wird zu Null gesetzt)

$$v_1 = [46, 70; 0]^T \quad \text{und} \quad v_2 = [59, 20; 0]^T, \quad (6.12)$$

so folgt eine Güte von $pi = 0,1472$.

Benutzt man die Clusterung (FCM, $\nu = 1,13$) für den gesamten von $\tilde{y}(k-1)$ und $\tilde{u}(k-3)$ aufgespannten Eingangsgrößenraum, so folgen als Clusterschwerpunkte

$$\mathbf{v}_1 = [0, 7840; 50, 77]^T \quad \text{und} \quad \mathbf{v}_2 = [-0.7863; 55, 87]^T \quad (6.13)$$

bei einer Modellgüte von $pi = 0,1495$ (2 Regeln, 10 Parameter). Als Initialisierung wurde $\mathbf{v}_1^{(0)} = [-2/3; 0]^T$ und $\mathbf{v}_2^{(0)} = [2/3; 0]^T$ verwendet. Das letztere Modell wurde allerdings völlig automatisch generiert, ohne jegliche Heuristik bei der Festlegung von Referenz–Fuzzy–Mengen oder Konklusionsfunktionsparametern. Durch eine Erhöhung der Clusteranzahl lässt sich wieder die Modellgüte steigern. Abbildung 6.10 zeigt dies für das Modell mit $\tilde{y}(k-1)$ und $\tilde{u}(k-3)$ als Eingangsgrößen bei Clusterung über den FCM mit $\nu = 1,13$.

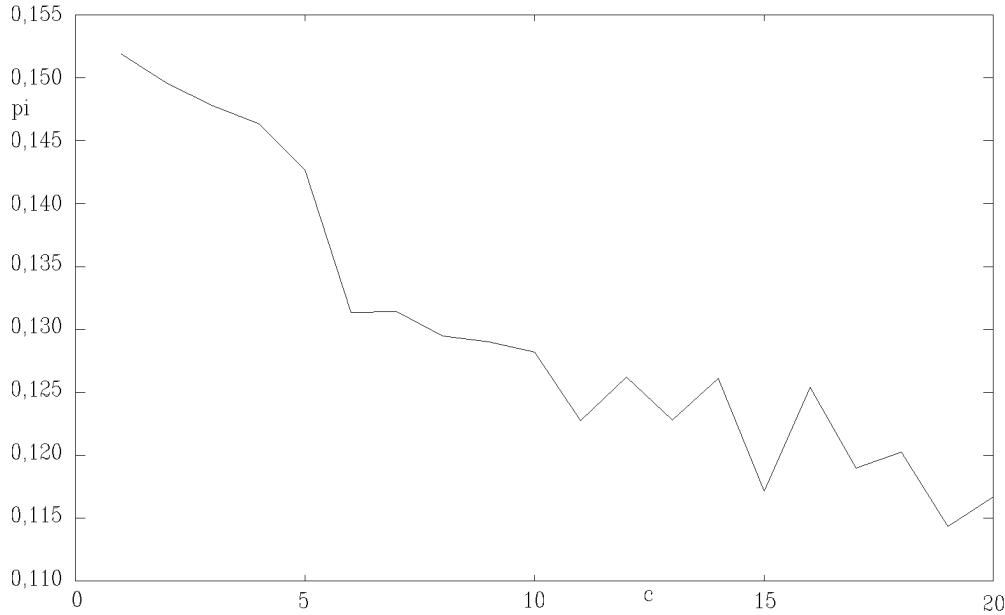


Bild 6.10: Modellgüte in Abhängigkeit von der Clusterzahl c bei Clusterung mit dem FCM mit $\nu = 1,13$

Bei den eigenen Untersuchungen zum Modell M1 wurden, abweichend von Sugeno und Tanaka (1991) bzgl. der Stellgröße $\tilde{u}(k-1)$, $\tilde{u}(k-2)$ und $\tilde{u}(k-3)$ anstelle von $\tilde{u}(k)$, $\tilde{u}(k-1)$ und $\tilde{u}(k-2)$ verwendet. Dies geschah aus der Überlegung, daß das Modell eine (zukünftige) Ausgangsgröße $y(k)$ prädizieren soll. Dann steht aber $\tilde{u}(k)$ nicht als Modelleingangsgröße zur Verfügung. Eine Nachrechnung des Ansatzes von Sugeno und Tanaka mit dieser Änderung brachte eine Güte von $pi = 0,0612$. Werden zwei echt-mehrdimensionale Referenz-Fuzzy-Mengen mit den Partitionsschwerpunkten wie bereits zuvor beschrieben jeweils in der Mitte der konstanten Bereiche der Referenz-Fuzzy-Mengen von Sugeno und Tanaka (1991) (bzgl. normierter Größen) verwendet und anschließend (automatisch) iterativ bzgl. $\tilde{y}(k-1)$ nachverfeinert, so läßt sich eine Güte von $pi = 0,0556$ erreichen.

Bild 6.11 vergleicht Meßdaten und prädizierte Daten für Einschrittprädiktion und rekursive Prädiktion für das Fuzzy-Modell mit 2 Regeln für komplette Eingangsraumclusterung über den FCM ($\nu = 1,13$). Der mittlere quadratische Fehler beträgt dabei für Einschrittprädiktion $pi = 0,150$ und $pi = 2,033$ für rekursive Auswertung des Modells. Wie bereits Bild 6.10 illustriert, führt eine Erhöhung der Anzahl der Partitionen bei ansonsten gleicher Modellstruktur zu keiner wesentlichen Verbesserung. So folgt für $c = 6$ Cluster $pi = 0,131$ für Einschrittprädiktion und $pi = 1,717$ für rekursive Prädiktion.

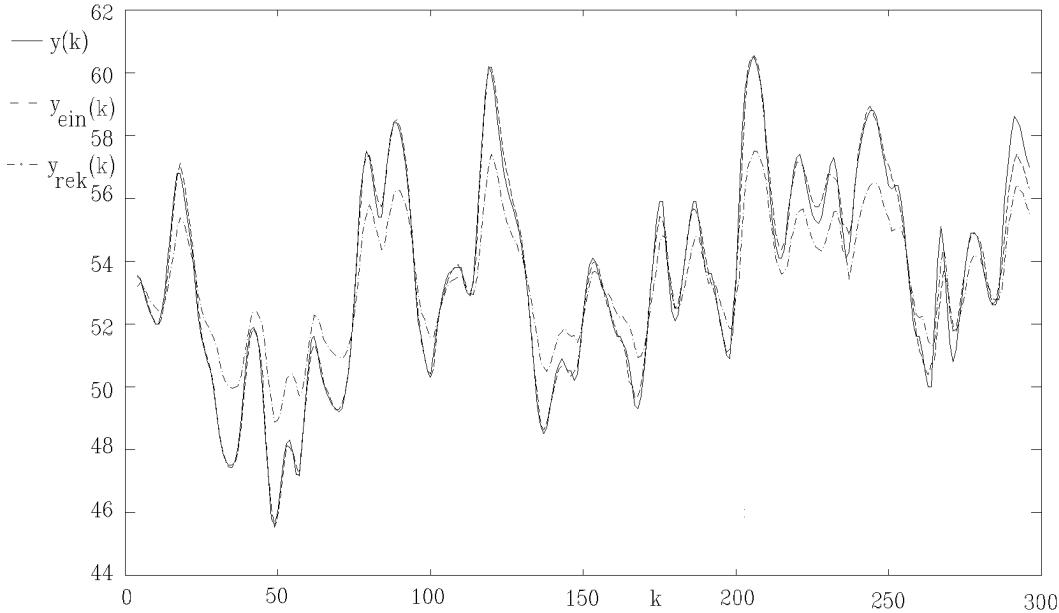


Bild 6.11: Meßdaten $y(k)$, einschrittprädizierte $y_{ein}(k)$ und rekursiv prädizierte Daten y_{rek} nach dem hier vorgestellten Algorithmus für $c = 2$ ($k = 4, \dots, 296$)

Festlegung der Signalverzögerungen

Angesprochen werden soll hier die Problematik bei der Festlegung der Signalverzögerungen. Bei der Festlegung von Totzeiten bezüglich der Stellgrößen eines Systems sollten

immer primär physikalische Überlegungen vorgenommen werden. Sonst besteht die Gefahr, daß bei der Anpassung von Prädiktions- und Meßkurvenverlauf über die numerische Optimierung eines Gütekriteriums durch die Dominanz anderer Effekte (wie z. B. eine i. allg. nicht exakt richtige Modellstruktur) eine Totzeit identifiziert wird, die real nicht vorliegt. Dies kann insbesondere bei einem geschlossenen Regelkreis zu falschen Aussagen führen. Die in Bild 6.12 gezeigte Kreuzkorrelation zwischen $y(k)$ und $u(k)$ favorisiert ein Modell mit $-u(k-5)$ zur Prädiktion von $y(k)$. Da das negative Vorzeichen bzgl. $u(k-5)$

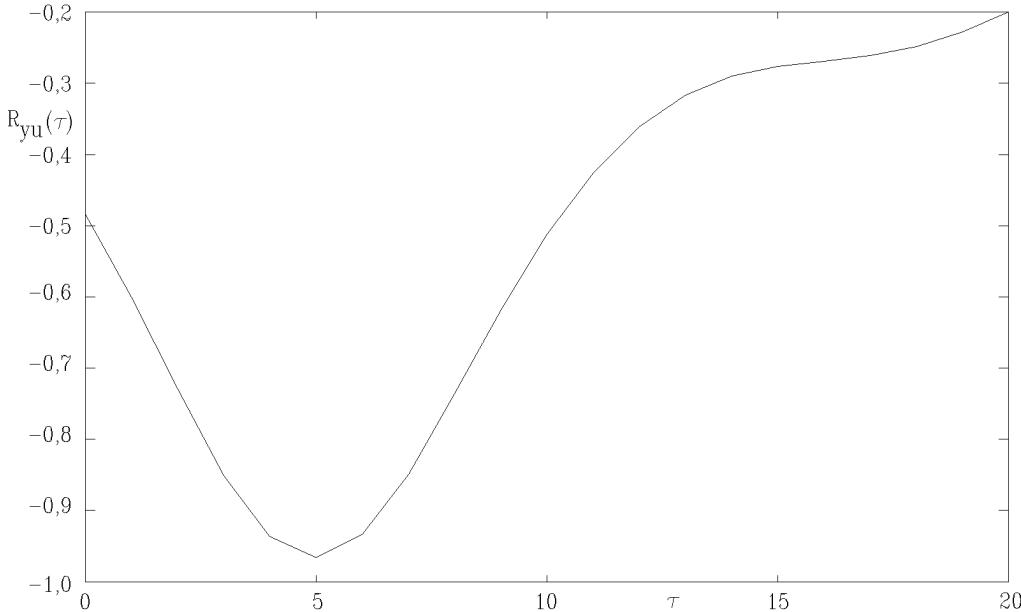


Bild 6.12: Kreuzkorrelation $R_{yu}(\tau)$ der Signale $y(k)$ und $u(k)$ (τ ist die diskrete Zeitverschiebung, $t = \tau T$ mit $T = 9\text{s}$)

automatisch bei der Identifikation der Konklusionsparameter identifiziert wird, können als Modelleingangsgrößen $y(k-1)$ und $u(k-5)$ angesetzt werden. Das gleiche Ergebnis folgt bei der Identifikation eines linearen Modells für $y(k)$ nur mit $\tilde{u}(k-1-\tau)$ als Eingangsgröße und quadratischem Gütekriterium. Dagegen ist für ein Modell mit $\tilde{y}(k-1)$ und $\tilde{u}(k-1-\tau)$ als Eingangsgrößen bezüglich eines quadratischen Gütekriteriums $\tau = 2$ optimal. Diese Änderung der im Sinne der Modellgüte optimalen Signalverzögerung τ_{opt} illustriert, daß τ_{opt} nicht mit der physikalischen Totzeit im realen Prozeß zu verwechseln ist. Diese ist nicht von der Modellstruktur, sondern von physikalischen Zusammenhängen abhängig. Dagegen können die Zeitverzögerungen bezüglich rückgeführter Modellausgangsgrößen im Sinne einer optimalen Modellgüte gewählt werden.

7 Zusammenfassung und Ausblick

Dieser Bericht befaßt sich mit funktionalen Fuzzy–Systemen. Der erste hier vorgestellte neue Aspekt besteht in der Einführung mehrdimensionaler Referenz–Fuzzy–Mengen, die durch ihren Partitionsschwerpunkt und eine Abstandsnorm definiert sind. Dabei wird in der Prämisse abgefragt, wie groß die Affinität eines Datums zum Partitions– oder Clusterschwerpunkt ist, es werden also *Punktaffinitäten* betrachtet. Dadurch ist für jeden Punkt des Eingangsgrößenraumes eine Zugehörigkeit in diese Partition bzw. Fuzzy–Menge und damit zur korrespondierenden Regel definiert. Der zweite in diesem Bericht behandelte Punkt ist die Identifikation von funktionalen Fuzzy–Modellen mit echt–mehrdimensionalen Referenz–Fuzzy–Mengen. Für die Strukturidentifikation wurden Clusterverfahren eingesetzt, für die Konklusionsparameteridentifikation das gewichtete rekursive Verfahren der kleinsten Fehlerquadrate. Hierzu wurden drei Clusterverfahren (FCM, Algorithmen nach Backer sowie Gustafson und Kessel) vorgestellt.

Das vorgestellte Verfahren wurde mit sehr gutem Erfolg auf zwei nichtlineare Testsysteme angewendet. Dabei brachte der FCM die besten Ergebnisse. Der Algorithmus nach Gustafson und Kessel birgt den Nachteil, daß man ohne A–priori–Vorwissen über die Datenstruktur das Potential der lokalen Abstandsnormen nicht richtig nutzen kann. Es konnten eindeutige Vorschriften zur Wahl der Identifikationsalgorithmusparameter gemacht werden, um die Modellgüte zu beeinflussen. Damit ist, im Gegensatz zu noch vielfach verwendeten Trial–and–error–Verfahren bei der Strukturidentifikation, ein sehr systematisches, weitgehend automatisiertes Vorgehen möglich. Dies zeichnet den Algorithmus gegenüber solchen aus, die für eine Partition nur einen lokalen Trägerbereich verwenden. Hierzu zählen die meisten Verfahren mit orthogonalen Partitionen. Dort kann z. B. eine Variation der Referenz–Fuzzy–Mengen zu sprunghaften Güteänderungen führen, die durch Wechsel von Daten zwischen verschiedenen Trägerbereichen hervorgerufen wird. Somit ist es oft nicht möglich, eine Richtlinie für die Wahl der Modellparameter anzugeben. Außerdem kann dort bei ungeeigneter Partitionierung der Fall eintreten, daß einige Partitionen zu wenig Datenpunkte für eine Identifikation oder eventuell gar keine enthalten. Solche Effekte können bei dem hier vorgestellten Verfahren nicht auftreten, da alle Zugehörigkeitsfunktionen Werte größer Null im gesamten Definitionsbereich (ausgenommen die Clusterschwerpunkte) liefern. In den Clusterschwerpunkten liefert nur eine Zugehörigkeitsfunktion $\mu = 1$, alle anderen Null. So werden beim hier vorgestellten Identifikationsverfahren sprunghafte Güteänderungen bei Änderung einzelner Modellparameter vermieden. Da auch die Lage der Daten untereinander vom Identifikationsverfahren berücksichtigt wird, folgt eine (wünschenswerte) globale Anpassung des Modells. Dadurch werden bei Erhöhung der Partitionsanzahl automatisch alle Partitionsschwerpunkte im Sinne des zu Grunde gelegten Gütekriteriums durch die Clusterung optimal gelegt. Es bestehen keine Dimensionsgrenzen für die Verschiebung der Clusterschwerpunkte. Bei anderen Identifikationsverfahren kann i. allg. immer nur eine Dimension feiner unterteilt werden, was im globalen Sinne meistens nicht optimal ist.

Die hier vorgestellten Untersuchungen können in mehrerer Hinsicht fortgeführt werden: Wegen der gleichen Prämissenstruktur von funktionalen und relationalen Fuzzy-Systemen können die vorgestellten mehrdimensionalen Referenz-Fuzzy-Mengen, die damit verbundenen Zugehörigkeitsbestimmung über Ähnlichkeitsvergleich sowie der Algorithmus zur Strukturidentifikation ohne Änderungen direkt auch auf relationale Fuzzy-Systeme angewendet werden. Die Effektivität des bei nicht hyperellipsoiden Datenstrukturen günstigen Backer-Algorithmus sollte geprüft werden. Des Weiteren ist zu untersuchen, nach welchen Kriterien der Clusteralgorithmus systematisch bei wenig Vorwissen über die betrachtete Datenstruktur ausgewählt werden kann. Der vorgestellte einfache Algorithmus zur Feineinstellung der mit Clusteranalyse bestimmten Partitionsschwerpunkte sollte z. B. durch Anwendung eines Gradientenverfahrens im Hinblick auf eine Steigerung von Modellgüte und Rechengeschwindigkeit verbessert werden. Auch wäre eine Prüfung der Robustheit des hier vorgestellten Identifikationsverfahrens bei verrauschten Meßwerten sinnvoll.

8 Literaturverzeichnis

- Araki, S., H. Nomura, I. Hayashi und N. Wakami.** 1991. A self-generating method of fuzzy inference rules. *Proc. of the Conf. on fuzzy engineering towards human friendly systems, Yokohama/Japan, 13-15. Nov. 1991.* 1047-1058.
- Backer, E.** 1978. *Cluster analysis by optimal decomposition of induced fuzzy sets.* Dissertation Universität Delft/Niederlande. Delft: Delft Univ. Press.
- Ben Farhat, H.** 1994. *Selbsteinstellende Funktional-Fuzzy-Regelung.* Studienarbeit MSRT. Universität Duisburg.
- Bezdek, J. C.** 1981. *Pattern recognition with fuzzy objective function algorithms.* New York, London: Plenum Press.
- Box, G. und G. Jenkins.** 1970. *Time series analysis.* Oakland: Holden-Day.
- Buckley, J.J.** 1993. Sugeno type controllers are universal controllers. *Fuzzy Sets and Systems 53.* 299-303.
- Driankov, D., H. Hellendoorn und M. Reinfrank.** 1993. *An Introduction to Fuzzy Control.* Berlin u. a.: Springer.
- Hebisch, H.** 1992. *Identifikation eines Fuzzy-Modells am Beispiel einer hydraulischen Regelstrecke.* Diplomarbeit MSRT. Universität Duisburg.
- Isermann, R.** 1988a. *Identifikation dynamischer Systeme.* Bd. I. Berlin u.a.: Springer.
- Isermann, R.** 1988b. *Identifikation dynamischer Systeme.* Bd. II. Berlin u.a.: Springer.
- Küpper, K.** 1993. *Fuzzy-Modellierung eines Gasofens.* Forschungsbericht Nr. 7/93 MSRT. Universität Duisburg.
- Küpper, K.** 1994a. Self-learning fuzzy models using stochastic approximation. *Proc. of the 3rd Conf. on Control Applications, Glasgow/Scotland, 24-26. Aug. 1994.*
- Küpper, K.** 1994b. *Modellierung mittels eines selbstlernenden Fuzzy-Systems.* Forschungsbericht Nr. 8/94 MSRT. Universität Duisburg.
- Lancaster, P. und M. Tismenetsky.** 1985. *The theory of matrices.* Orlando u. a.: Academic Press.
- Lee, C. C.** 1990. Fuzzy logic in Control Systems: Fuzzy Logic Controller - Part I. *IEEE Transactions on Systems, Man and Cybernetics 20.* 404-418.
- Meyer-Gramann, K.D. und E.-W. Jüngst.** 1993. Fuzzy Control – schnell und kostengünstig implementiert mit Standard-Hardware. *Automatisierungstechnik 41.* 166-172.

- Pedrycz, W.** 1984. An identification algorithm in fuzzy relational systems. *Fuzzy Sets and Systems* 13. 153-167.
- Pedrycz, W.** 1991. Fuzzy modelling: Fundamentals, construction and evaluation. *Fuzzy Sets and Systems* 41. 1-15.
- Reuter, H.** 1992. *Zur Identifikation bilinearer Systemmodelle mit einem robusten Least-Squares-Verfahren*. Forschungsbericht Nr. 9/92 MSRT. Universität Duisburg.
- Reuter, H.** 1993. *Ein neues Verfahren zur Abschätzung einer günstigen Abtastzeit mittels Entropieanalyse*. Forschungsbericht Nr. 9/93 MSRT. Universität Duisburg.
- Schwarz, H.** 1991. *Nichtlineare Regelungssysteme – Systemtheoretische Grundlagen*. München, Wien: Oldenbourg.
- Shaw, I.S. und J.J. Krüger.** 1992. New fuzzy learning model with recursive estimation for dynamic systems. *Fuzzy Sets and Systems* 42. 217-229.
- Sugeno, M. und G. T. Kang.** 1986. Fuzzy Modelling and Control of Multilayer Incinerator. *Fuzzy Sets and Systems* 18. 329–346.
- Sugeno, M. und G. T. Kang.** 1988. Structure Identification of Fuzzy Model. *Fuzzy Sets and Systems* 28. 15–33.
- Sugeno, M. und K. Tanaka.** 1991. Successive identification of a fuzzy model and its applications to prediction of a complex system. *Fuzzy Sets and Systems* 42. 315–334.
- Sugeno, M. und T. Yasukawa.** 1993. A Fuzzy-Logic-Based Approach to Qualitative Modeling. *IEEE Transactions on Fuzzy Systems* 1. 7-31.
- Suprijadi, E. B.** 1994. *Identifikation von Fuzzy-Modellen mittels stochastischer Approximation*. Studienarbeit MSRT. Universität Duisburg.
- Takagi T. und M. Sugeno.** 1985. Fuzzy Identification of Systems and its Applications to Modeling and Control. *IEEE Transactions on Systems, Man, and Cybernetics SMC-15*. 116–132.
- Tan, E., G. Mourot, D. Maquin und J. Rugot.** 1994. Identification of fuzzy models. *Proc. of the International Workshop on Fuzzy-Technologies in Automation and Intelligent Systems, 7.-8.4.1994, Duisburg*. 190-199.
- Tanaka, K. und M. Sugeno.** 1992. Stability analysis and design of fuzzy control systems. *Fuzzy Sets and Systems* 45. 135–156.
- Xu, C.W. und Y.-Z. Lu.** 1987. Fuzzy model identification and self learning for dynamic systems. *IEEE Transactions on Systems, Man, and Cybernetics SMC-17*. 683-689.

- Zadeh, L.A.** 1973. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics SMC-3*. 28-44.
- Zurmühl, R. und S. Falk.** 1986. *Matrizen und ihre Anwendungen*. Band II. Berlin u. a.: Springer.

A Verschiebung von Zugehörigkeitsfunktionsflanken

In diesem Abschnitt wird gezeigt, daß die Verschiebung einer oder zweier benachbarter linearer Referenz–Fuzzy–Mengenflanken selbst im Flankenbereich i. allg. nicht durch eine Modifikation des Konklusionspolynoms der beiden an der Flanke in Verschiebungsrichtung aneinander grenzenden Regeln substituiert werden kann. Betrachtet wird dazu der Flankenbereich Fl zweier Regeln, in den die Änderung der Flanken fällt. Des Weiteren beschränkt wird der Bereich auf das Gebiet, in dem nur diese beiden Regeln feuern (in den Randbereichen von Fl feuern im allgemeinen vier Regeln gleichzeitig). Verschoben werden sollen eine oder zwei benachbarte Flanken in x_j –Richtung. Deshalb kann die Betrachtung auf die Abhängigkeit des Erfülltheitsgrades der Prämisse ω alleine auf x_j beschränkt werden. Es gilt

$$\begin{aligned}\omega_1 &= \omega_{T,1} * \alpha_1(x_j) \quad \text{und} \\ \omega_2 &= \omega_{T,2} * \alpha_2(x_j)\end{aligned}\tag{A.1}$$

mit $\omega_{T,1} \neq f(x_j)$ und $\omega_{T,2} \neq f(x_j)$ sowie

$$\begin{aligned}0 &< \omega_{T,1} \leq 1 \quad \text{und} \\ 0 &< \omega_{T,2} \leq 1 \quad ,\end{aligned}\tag{A.2}$$

da die betrachteten Regeln in Fl feuern. $\omega_{T,1}$ ($\omega_{T,2}$) stellt dabei das Produkt der Erfülltheitsgrade aller m Teilprämissen α_i bis auf α_1 (α_2) dar. Eine Änderung der beiden Flanken führt zu

$$\widetilde{\alpha_1}(x_j) = \alpha_1(x_j) + \Delta\alpha_1(x_j)\tag{A.3}$$

anstelle von $\alpha_1(x_j)$ und

$$\widetilde{\alpha_2}(x_j) = \alpha_2(x_j) + \Delta\alpha_2(x_j)\tag{A.4}$$

anstelle von $\alpha_2(x_j)$. Damit folgt als resultierende Stellgröße

$$\tilde{u} = \frac{\omega_{T,1}(\alpha_1 + \Delta\alpha_1)u_1 + \omega_{T,2}(\alpha_2 + \Delta\alpha_2)u_2}{\omega_{T,1}(\alpha_1 + \Delta\alpha_1) + \omega_{T,2}(\alpha_2 + \Delta\alpha_2)} \quad .\tag{A.5}$$

Zu prüfen ist nun, ob man \tilde{u} auch durch Beibehaltung der Referenz–Fuzzy–Mengen ($\Delta\alpha_1 = \Delta\alpha_2 = 0$) aber Änderung der Konklusionen erhalten kann:

$$\hat{u} = \frac{\omega_{T,1}\alpha_1\hat{u}_1 + \omega_{T,2}\alpha_2\hat{u}_2}{\omega_{T,1}\alpha_1 + \omega_{T,2}\alpha_2} \stackrel{?}{=} \tilde{u} \quad .\tag{A.6}$$

Wegen (A.2) kann man

$$g := \frac{\omega_{T,1}}{\omega_{T,2}}\tag{A.7}$$

abkürzen und erhält durch Gleichsetzen von (A.5) und (A.6)

$$g^2\alpha_1(\alpha_1 + \Delta\alpha_1)(u_1 - \hat{u}_1) + g\alpha_2(\alpha_1 + \Delta\alpha_1)(u_1 - \hat{u}_2) + g\alpha_1(\alpha_2 + \Delta\alpha_2)(u_2 - \hat{u}_1) + \alpha_2(\alpha_2 + \Delta\alpha_2)(u_2 - \hat{u}_2) \stackrel{!}{=} 0 \quad (\text{A.8})$$

Beschreibt man die Flanken vor der Verschiebung durch

$$\alpha_1 = m_1 x + n_1 \quad \text{und} \quad \alpha_2 = m_2 x + n_2 \quad , \quad (\text{A.9})$$

so erhält man durch die Verschiebung

$$\begin{aligned} \Delta\alpha_1 &= \hat{\alpha}_1 - \alpha_1 = \tilde{n}_1 - n_1 =: c_1 \quad \text{und} \\ \Delta\alpha_2 &= \hat{\alpha}_2 - \alpha_2 = \tilde{n}_2 - n_2 =: c_2 \quad . \end{aligned} \quad (\text{A.10})$$

Bei den Konklusionspolynomen können alle nicht von x_j abhängigen Terme mit dem konstanten Anteil zusammengefaßt werden:

$$u_1 = p_1 + p_3 x \quad u_2 = p_2 + p_4 x \quad . \quad (\text{A.11})$$

Die modifizierten Polynome werden durch „:“ beschrieben:

$$\hat{u}_1 = \hat{p}_1 + \hat{p}_3 x \quad \hat{u}_2 = \hat{p}_2 + \hat{p}_4 x \quad . \quad (\text{A.12})$$

Die Ersetzbarkeit der Referenz–Fuzzy–Mengen–Änderung durch andere Konklusionen muß für alle $x_j \in Fl$ gelten. Durch Einsetzen von (A.9) – (A.12) in (A.8) erhält man mit der Bedingung, daß die jeweiligen Koeffizienten der Potenzen von x identisch Null sein müssen, ein Gleichungs–System 4. Ordnung. Hat das System den vollen Rang, so gibt es eine eindeutige Lösung und die Substituierbarkeit von Referenz–Fuzzy–Mengen durch Konklusionspolynomänderung ist gegeben. Nach Anwendung des gaußschen Eliminations–algorithmus enthält die Dreieckdiagonalmatrix die folgenden 4 Elemente:

$$d_1 = -g^2 m_1^2 - g m_1 m_2 \quad , \quad (\text{A.13})$$

$$d_2 = \frac{g^2 n_1 m_2 m_1 - n_2 m_1^2 g - m_2 n_2 g m_1 + m_2^2 n_1}{g m_1} \quad , \quad (\text{A.14})$$

$$d_3 = d_1 \quad \text{und} \quad (\text{A.15})$$

$$\begin{aligned} d_4 &= -(g m_1 c_2^2 n_1 n_2 m_2 - 3g^3 m_1^2 c_1 n_1 n_2^2 + 3g^3 m_1 n_1^2 c_1 m_2 n_2 + g^3 m_1 m_2 c_1^2 n_1 n_2 \\ &\quad + 2g^3 m_1 n_1^3 m_2 n_2 - 2g^3 m_1^2 n_1^2 n_2^2 - 2g^2 m_1^2 n_2^2 n_1 c_2 - 2g^2 m_1^2 n_2^3 n_1 \\ &\quad + 3g^2 m_1 c_1 n_1 n_2^2 m_2 + 3g^2 m_1 n_1^2 c_2 n_2 m_2 + 4g^2 m_1 n_1^2 n_2^2 m_2 \\ &\quad + 2g^2 m_1 c_1 n_1 c_2 n_2 m_2 - g^2 n_1^3 c_2 m_2^2 - 2g^2 c_1 n_1^2 n_2 m_2^2 - 2g^2 n_1^3 n_2 m_2^2 \\ &\quad - g^2 c_1 n_1^2 c_2 m_2^2 - g c_2^2 n_1^2 m_2^2 - 2g n_2^2 n_1^2 m_2^2 - 3g n_2 n_1^2 c_2 m_2^2 - m_1 c_2^2 n_1 n_2 m_2 \end{aligned}$$

$$\begin{aligned}
& -2m_1n_2^3n_1m_2 - 3m_1n_2^2n_1c_2m_2 + 3n_2n_1^2c_2m_2^2 + c_2^2n_1^2m_2^2 + 2n_2^2n_1^2m_2^2 \\
& -g^3m_1^2c_1^2n_2^2 - g^2m_1^2c_1n_2^3 - g^2m_1^2n_2^2c_2c_1 + 2gm_1n_2^3n_1m_2 \\
& +3gm_1n_2^2n_1c_2m_2 - n_2c_1^2g^2m_1n_1m_2 - 3n_2^2m_1n_1m_2gc_1 \\
& -2n_2c_2m_1n_1m_2gc_1 + 2n_1^2n_2^2g^2m_1^2 + 2n_1^3n_2gm_2^2 + gn_1^3c_2m_2^2 \\
& -2n_1^3m_2g^2n_2m_1 + 2n_2c_1gm_2^2n_1^2 + gc_1c_2m_2^2n_1^2 - 3n_1^2m_2g^2c_1n_2m_1 \\
& -4n_1^2n_2^2m_2m_1g - 3n_1^2c_2m_2n_2m_1g + 3n_1n_2^2c_1g^2m_1^2 + 2n_1n_2^2c_2m_1^2g \\
& +2n_1n_2^3m_1^2g + g^2m_1^2c_1^2n_2^2 + gm_1^2c_1n_2^3 + gm_1^2n_2^2c_2c_1) \\
& / [m_1(g^2n_1m_2m_1 - n_2m_1^2g - m_2n_2g m_1 + m_2^2n_1)] =: Z_4/N_4
\end{aligned} \tag{A.16}$$

Bei den betrachteten benachbarten Referenz–Fuzzy–Mengen ist $g = 1$ und (A.13) für alle $m_1 \neq m_2$ erfüllt. (Dies entspricht bereits einer starken Einschränkung, da häufig benachbarte Referenz–Zugehörigkeitsfunktionen so definiert werden, daß sich im Überlappungsbereich die Zugehörigkeitsgrade zu 1 addieren.) Außerdem gilt: $m_1 \neq 0$, $m_2 \neq 0$, ($n_1 \neq 0 \vee n_2 \neq 0$). Damit ist

$$d_2 = \frac{m_1 m_2 (n_1 - n_2) + n_1 m_2^2 - n_2 m_1^2}{m_2} . \tag{A.17}$$

Weil für die benachbarten Flanken $\operatorname{sgn}(m_1) = -\operatorname{sgn}(m_2)$ und $\operatorname{sgn}(n_1) = -\operatorname{sgn}(n_2)$ oder $[(n_1 = 0 \wedge n_2 = 0) \vee (n_1 \neq 0) \wedge (n_2 \neq 0)]$ gilt, sowie $m_1 \neq 0$ und $m_2 \neq 0$ ist, gilt $d_2 \neq 0$. Der Zähler von d_2 tritt zusammen mit einem Faktor m_1 als Nenner von d_4 auf, welcher somit auch $\neq 0$ ist. Beim Zähler zu (A.16) dagegen gibt es Parameterkombinationen, für die das Zählerpolynom den Wert 0 annimmt: Aus $g = 1$ und $n_2 = 0$ folgt beispielsweise $Z_4 = 0$ oder aus $g = 1$ und $n_1 = 0$ folgt $Z_4 = -m_1^2 c_1 n_2^2 (c_1 + c_2 + n_2)$, wobei Kombinationen $c_1 + c_2 + n_2 = 0$ möglich sind.

B Zur Wahl der Operatoren

Im folgenden wird der Einfluß des Partialprämissen–Verknüpfungsoperators auf das Gewicht eines Datums bei der Regelauswertung oder bei der Identifikation erläutert. Exemplarisch wird von einer 2×2 –Partitionierung eines zweidimensionalen Eingangsgrößenraumes ausgegangen (Bild B.1).

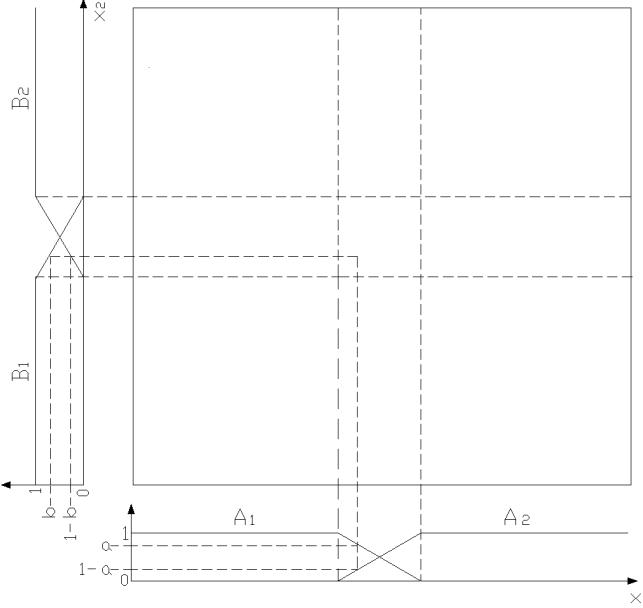


Bild B.1: 2×2 –Partitionierung

Vorausgesetzt sei

$$\sum_i \mu_{A_i}(x_1) = 1 \quad \forall x_1 \in \mathbb{D}_{x_1} \quad \text{und} \quad (B.1)$$

$$\sum_j \mu_{B_j}(x_2) = 1 \quad \forall x_2 \in \mathbb{D}_{x_2} \quad . \quad (B.2)$$

Zur Abkürzung sei

$$a = \mu_{A_1}(x_1) \in [0, 1] \quad \text{und} \quad b = \mu_{B_1}(x_2) \in [0, 1] \quad (B.3)$$

eingeführt. Für den Produkt–Operator folgt die Summe der (Gesamt–) Zugehörigkeiten eines Datums zu

$$\sum_{i=1}^4 \omega_i = a b + (1 - a)b + (1 - a)(1 - b) + a(1 - b) = 1 \quad . \quad (B.4)$$

Für den Minimum–Operator folgt:

$$\sum_{i=1}^4 \omega_i = \min(a, b) + \min(1 - a, b) + \min(1 - a, 1 - b) + \min(a, 1 - b) \quad . \quad (B.5)$$

Bei der nötigen Fallunterscheidung werde der Fall $a > b$ und $1 - a < b$ betrachtet. Dafür ergibt sich

$$\sum_{i=1}^4 \omega_i = 3 - 2a > 1 \quad (\text{B.6})$$

im Übergangsbereich. Ähnliche Ergebnisse folgen für die anderen Fälle.