

Zur analytischen Beschreibung von Fuzzy-Systemen

Andreas Kroll

Forschungsbericht Nr. 4/95

Meß-, Steuer- und Regelungstechnik

Übersicht: In diesem Bericht wird eine analytische Beschreibung von Fuzzy-Systemen mit mehrdimensionalen Fuzzy-Referenzmengen hergeleitet. Dazu erfolgen Stetigkeits- und Differenzierbarkeitsuntersuchungen für Zugehörigkeitsfunktionen, die aus Fuzzy-Clusterverfahren abgeleitet werden können. Die vorgestellte Beschreibung gestattet es, ungewichtete nichtlineare Verfahren aus der konventionellen Modellidentifikation für die Identifikation solcher Fuzzy-Modelle einzusetzen. Dazu wird ein zweistufiges, in beiden Stufen nichtlineares Identifikationsverfahren angegeben.

Gerhard-Mercator-Universität - GH Duisburg
Meß-, Steuer- und Regelungstechnik
Prof. Dr.-Ing. H. Schwarz

Inhaltsverzeichnis

Nomenklatur	II
1 Einleitende Übersicht	1
2 Eigenschaften mehrdimensionaler Fuzzy-Mengen	4
2.1 c -Partitionierung	4
2.2 Zugehörigkeitsfunktionen vom Fuzzy-Clustertyp	5
2.3 Interpretation der Abhängigkeit vom Unschärfeparameter	13
2.4 Zugehörigkeitsfunktionen vom e -Typ	15
2.5 Abstandsnormen	16
2.6 Diskussion der Zugehörigkeitsfunktionen	17
3 Analytische Beschreibung von Fuzzy-Systemen	19
3.1 Funktionale Fuzzy-Systeme	19
3.2 Relationale Fuzzy-Systeme	23
4 Fuzzy-Systeme und Basisfunktionsnetzwerke	26
4.1 Radialbasisfunktionsnetzwerke	26
4.2 Verwandtschaft von Fuzzy-Systemen und Basisfunktionsnetzwerken	28
5 Nutzung analytischer Modellbeschreibungen bei der Identifikation	31
5.1 Wiener- und Hammerstein-Modelle mit statischer Fuzzy-Nichtlinearität	31
5.2 Identifikation dynamischer Fuzzy-Modelle	35
5.3 Identifikationsverfahren	36
6 Strukturelektion, Güte und Modellvalidierung	40
7 Zusammenfassung und Ausblick	43
Literaturverzeichnis	44

Nomenklatur

Skalare Größen

a_i, b_i	Koeffizienten einer Differenzgleichung
c	Clusteranzahl, entspricht der Anzahl der Regeln
c_i	konstanter Term im i -ten Konklusionspolynom
d_{D_j}	j -tes Element der Formmatrix \mathbf{D} (Diagonalmatrix)
d_{ik}	Abstand zwischen i -tem Clusterzentrum \mathbf{v}_i und k -tem Datensatz \mathbf{x}_k
d_{ijk}	relativierter Abstand, $d_{ijk} = d_{ik}/d_{jk}$
g_i	Gewichtung der i -ten Regel ($g_i \in [0 ; 1]$)
i	Regel- / Partitionsnummer
k	diskretes Zeitargument / Datensatznummer
K_L	Verstärkung des linearen Teilsystems beim Hammerstein-Modell
m	Anzahl der eindimensionalen Eingangsgrößen (Dimension des Eingangsdatenraumes)
n_a	Ordnung des Nennerpolynoms einer rationalen Übertragungsfunktion
n_b	Ordnung des Zählerpolynoms einer rationalen Übertragungsfunktion
N	Anzahl der Datensätze
p	Exponent bei p -Norm
$p_i(a)$	a -ter Polynomkoeffizient der i -ten Regel
r	Ausgangsgröße der statischen Nichtlinearität beim Hammerstein-Modell
s_i	Abzisse des Schwerpunktes der i -ten Fuzzy-Menge
u	externe Eingangsgröße
x_j	j -te Eingangsgröße
y	scharfe Ausgangsgröße des gesamten Regelsatzes
y_i	scharfe Ausgangsgröße der i -ten Regel
α	Gewicht der i -ten Radialbasisfunktion
ε_F	Prädiktionsfehler $\varepsilon_F = y - \hat{y}$
ϕ_i	i -te Fuzzy-Basisfunktion
κ	Potenz des relativierten Abstandes d_{ijk} , $\kappa = 2/(\nu - 1)$
$\mu_{i,0}$	Wert von μ_i^{ga} im i -ten Prototypen
μ_{ik}	Zugehörigkeit des Punktes \mathbf{x}_k zum Cluster (zur Partition) i
μ^{cl}	Zugehörigkeitsfunktion (alle Prototypen ausgewertet)
μ^{ga}	Zugehörigkeitsfunktion (nur ein Prototyp ausgewertet)
ν	Unschärfeparameter ($\nu > 1$)
$\tilde{\sigma}_i$	Parameter der Zugehörigkeitsfunktionen vom e-Typ
σ_j^2	Streuung / Varianz
σ_{jl}^2	Kovarianz
τ	diskrete Ersatzzeit
Ψ	Radialbasisfunktion
ω_i	Gesamtprämisse der i -ten Regel

Vektoren und Matrizen

D	Formmatrix
s	Parametervektor aller Konklusionen bei relationalen Fuzzy-Systemen
U	Zugehörigkeitsmatrix
v_i	Prototyp des i -ten Clusters, Partitionsmittelpunkt der i -ten Partition
x_k	k -ter Datensatz
X_{D1}	Meßvektor zu θ_{D1}
X_{D2}	Meßvektor zu θ_{D2}
Θ_i	Parametervektor des i -ten Konklusionspolynoms bis auf den konst. Term
Θ_{D1}	Parametervektor aller Konklusionspolynome bei dyn. Fuzzy-Systemen (Fall 1)
Θ_{D2}	Parametervektor aller Konklusionspolynome bei dyn. Fuzzy-Systemen (Fall 2)
Θ_F	Parametervektor (Konklusionsfunktionen, lineares Teilsystem) eines Hammerstein-Modells (funktionales Fuzzy-Modell)
Θ_R	Parametervektor (Konklusionen, lineares Teilsystem) eines Hammerstein-Modells (relationales Fuzzy-Modell)
Θ_S	Parametervektor aller Konklusionspolynome bei stat. Fuzzy-Systemen
φ	mit s korrespondierender Meßvektor
φ_F	mit θ_F korrespondierender Meßvektor
φ_R	mit θ_R korrespondierender Meßvektor

Mengen und Räume

U_i	eindimensionale Fuzzy-Referenzmenge bzgl. der Ausgangsgröße der i -ten Regel
Y_i	unscharfe (Partial-) Ausgangsgröße der i -ten Regel
\mathbb{D}_x	Definitionsbereich von x
\mathbb{I}	Menge aller c Cluster, $\mathbb{I} = \{1, 2, \dots, c\}$
\mathbb{I}_k	Menge aller Cluster, bei denen x_k den Abstand 0 vom Prototypen hat
$\tilde{\mathbb{I}}_k$	$\tilde{\mathbb{I}}_k = \mathbb{I} \setminus \mathbb{I}_k$
\mathbb{I}_N	$\mathbb{I}_N = \{1, \dots, N\}$
\mathcal{S}	Sequenz von Werten
\mathbb{V}_{cN}	Menge aller reellen $c \times N$ -Matrizen

Funktionen und Operatoren

$\text{cov}(\cdot)$	Kovarianzmatrix
$\exp(\cdot)$	Exponentialfunktion
$f(\cdot), g(\cdot), h(\cdot)$	allgemeine Funktion
$f_i(\cdot)$	i -te Konklusionsfunktion
$J(\cdot)$	Zielfunktion, Gütefunktion
$\ln(\cdot)$	Logarithmus naturalis
$\ \cdot\ $	Norm
$(\cdot)^{-1}$	Inverse
$(\cdot)^T$	Transponiert
$\overline{(\cdot)}$	Mittelwert
$\mathcal{A}(\cdot)$	Kardinalzahl (Anzahl der Elemente)

Sonstige Symbole

\forall	Allquantor
\setminus	Differenz (Mengenalgebra)
$\hat{(\cdot)}$	Schätzwert

Indizes

i	Regelnummer
j	Nummer der Eingangsgröße
k	Nummer des Datums
opt	optimaler Wert
0	Größe im Entwicklungspunkt
n	normiert

1 Einleitende Übersicht

Durch Fuzzy-Systeme lassen sich Systeme über linguistische WENN-DANN-Regeln beschreiben. Im allgemeinen läßt sich eine solche Beschreibungsform nicht durch einen geschlossenen analytischen Ausdruck ersetzen. Viele der bekannten Verfahren zur Identifikation von Modellen oder zur Optimierung von Systemen wie beispielsweise Gradientenverfahren erfordern aber eine solche Beschreibung und zudem auch noch Eigenschaften wie Stetigkeit oder stetige Differenzierbarkeit. In diesem Bericht werden deshalb verschiedene Klassen von Fuzzy-Systemen vorgestellt, die sich durch die beiden letztgenannten Eigenschaften auszeichnen.

Besondere Bedeutung bei der Erzielung bestimmter mathematischer Merkmale kommt der Wahl der Zugehörigkeitsfunktionen zu, die ausführlich diskutiert wird. Die Wahl hängt von der Anwendung des Fuzzy-Systems ab. Bei der Synthese von Fuzzy-Reglern erfolgt häufig eine Umsetzung von Expertenwissen oder zumindest plausiblen Ursache-Wirkungs-Zusammenhängen in einen Satz von WENN-DANN-Produktionsregeln. In solchen Fällen finden deshalb gewöhnlich Zugehörigkeitsfunktionen mit begrenztem Trägerbereich ihre Anwendung. Diese gestatten eine transparente differenzierte Zuordnung von Stellaktion und verursachender Regel und lassen sich somit einfach nach heuristischen Methoden nachstellen bzw. optimieren. Darüber hinaus gibt es auch Fuzzy-Modellbasierte Verfahren zur Fuzzy-Reglersynthese (Moore und Harris 1992, Vogels 1994, Senger 1995).

Bei der Fuzzy-Modellbildung kann grob zwischen Modellen unterschieden werden, die auf der Auswertung von linguistischem Expertenwissen beruhen sowie solchen, die durch Identifikation eines vorstrukturierten, parametrischen Fuzzy-Modells entstanden sind (Pedrycz 1984, Pedrycz 1985, Takagi und Sugeno 1985, Sugeno und Yasukawa 1993, Kroll 1994a, Küpper 1994). Bezüglich des letzten Ansatzes kann A-priori-Wissen vorteilhaft bei der Vorstrukturierung ausgenutzt werden. Ist kein oder kein geeignetes Expertenwissen verfügbar oder bestehen hohe Ansprüche an die Modellgüte, so ist i. allg. eine Modellidentifikation vorzunehmen.

Zwar lassen sich der Minimum- und der Maximumoperator analytisch beschreiben (Bertram 1992, Behmenburg und Bertram 1993), die mathematisch einfacher handhabbaren Produkt- und Summationsoperatoren führen aber erfahrungsgemäß zu besseren Ergebnissen und sollen deshalb im folgenden behandelt werden. Besonders anzumerken ist, daß der Produktoperator zu glatteren Interpolationen als der Minimumoperator führt (Brown und Harris 1991) und auch bei der Fuzzy-Regelung häufig vorteilhaft ist (Mizumoto 1991). Für die analytische Beschreibbarkeit ist weiterhin die Form der Partitionierung, d. h. die Festlegung der Wirkungsbereiche einzelner Regeln, von Bedeutung. Bei c Regeln, die jeweils in einer von insgesamt c Partitionen gelten, spricht man von einer „ c -Partitionierung“ (Bezdek 1981).

Eine weitere wichtige Eigenschaft besteht in der *geschlossenen, einheitlichen* analytischen Beschreibbarkeit einer jeden Partition. Diese liegt bei den häufig bereichsweise definierten

(beispielsweise triangularen oder trapezoiden) Zugehörigkeitsfunktionen nicht vor. Letztere werden als *Zugehörigkeitsfunktionen mit lokal beschränktem Träger* bezeichnet. Viele dieser Funktionen sind nicht stetig differenzierbar. Zugehörigkeitsfunktionen, die einheitlich geschlossen im gesamten Definitionsbereich definiert sind, heißen *Zugehörigkeitsfunktionen mit unbeschränktem Träger*. Hierzu zählen beispielsweise die Gaußfunktion oder aus Fuzzy-Clustering abgeleitete Zugehörigkeitsfunktionen (Kroll 1994a). Diese erweisen sich als vorteilhaft, wenn z. B. Gradientenverfahren angewendet werden sollen.

Die Verknüpfung von eindimensionalen Gaußfunktionen über das (algebraische) Produkt als Partialprämissenverknüpfungoperator führt zu geschlossen, einheitlich analytisch beschreibbaren Fuzzy-Systemen. Statt einer solchen Verknüpfung mehrerer eindimensionaler Gaußfunktionen kann auch direkt eine mehrdimensionale Gaußfunktion verwendet werden. Bei der i. allg. verwendeten euklidischen Abstandsnorm ergeben sich dann mehrdimensionale, radialsymmetrische Zugehörigkeitsfunktionen. Durch die Verwendung anderer Normen können auch nichtradialsymmetrische Zugehörigkeitsfunktionen realisiert werden. Charakteristisch bleibt allerdings, daß einzig der Abstand zu einem Prototypen (Referenz-Punkt der betrachteten Partition) ausgewertet wird. Eine deutlich unterschiedliche Klasse von Funktionen erhält man, wenn *jede* Funktion nicht nur ihren Referenz-Punkt, sondern *alle* auftretenden Prototypen bewertet. Der Grundgedanke besteht dabei darin, nicht nur den absoluten Abstand vom Referenz-Punkt, sondern die relativen Abstände zu allen Prototypen auszuwerten.

Zu solchen Zugehörigkeitsfunktionen zählen beispielsweise solche, die sich aus Fuzzy-Clusterverfahren ableiten lassen. Sie besitzen gerade bei der Fuzzy-Modellbildung mehrere Vorteile: Mit Hilfe von Fuzzy-Clusterverfahren können die Fuzzy-Referenzmengen direkt identifiziert (und anschließend nachoptimiert) werden (Kroll 1994a, Lohmann 1994, Totz 1995). Die meisten Clusterverfahren führen darüber hinaus zu Zugehörigkeitsfunktionen, die ein Fuzzy-Informationssystem (Meyer-Gramann und Jüngst 1993) bilden. Die Anzahl der Fuzzy-Referenzmengen pro Dimension ist nicht wie bei den eindimensionalen Ansätzen explizit vorzugeben, was unvorteilhaft für eine automatisierte Modellidentifikation ist. Vielmehr werden die mehrdimensionalen Fuzzy-Referenzmengen durch die Clusterverfahren dimensionsübergreifend optimal plaziert, weshalb nur die Gesamtanzahl an Referenzmengen vorzugeben ist. Nicht zuletzt besitzt eine Partitionierung mit derartigen Zugehörigkeitsfunktionen angenehme Inter- und Extrapolationseigenschaften, beispielsweise wenn in einigen Bereichen des Eingangsdatenraumes keine oder keine ausreichende Anzahl an Testdaten liegen. Bei Partitionierungen über Fuzzy-Referenzmengen mit begrenztem Träger führt dies häufig zu Schwierigkeiten.

Diese Vorteile und bereits erfolgreiche Anwendungen bei der Fuzzy-Identifikation (Sin und de Figuero 1993, Kroll 1994a) motivieren zu einer detaillierteren Betrachtung von Fuzzy-Systemen mit solchen Funktionstypen. Im 2. Abschnitt werden einige mathematische Eigenschaften von c -Partitionierungen und von mehrdimensionalen Zugehörigkeitsfunktionen untersucht. Der 3. Abschnitt behandelt die analytische Darstellung funktionaler und relationaler Fuzzy-Systeme. Im 4. Abschnitt werden Vergleiche zwischen

Fuzzy-Systemen und Basisfunktionsnetzwerken, zu denen auch bestimmte Klassen künstlicher Neuronaler Netze gehören, gezogen. Im 5. Abschnitt werden zur Identifikation von einfachen Hammerstein-Modellen mit Fuzzy-Nichtlinearität sowie dynamischen Fuzzy-Modellen geeignete mathematische Beschreibungsformen und ein zweistufiges, in beiden Stufen nichtlineares Identifikationsverfahren vorgestellt. Durch eine Transformation der Meßdaten unter Berücksichtigung der Zugehörigkeiten kann eine Beschreibungsform erreicht werden, die die Anwendung einer Vielzahl von aus der konventionellen Identifikation bekannten Verfahren gestattet. Zusammenfassung und Ausblick schließen den Bericht ab.

2 Eigenschaften mehrdimensionaler Fuzzy-Mengen

Nach der c -Partitionierung werden in diesem Abschnitt Stetigkeit und Differenzierbarkeit von zwei mehrdimensionalen Typen von Zugehörigkeitsfunktionen mit unbeschränktem Träger betrachtet.

2.1 c -Partitionierung

Zu unterscheiden ist zwischen harter

$$\mathbf{U}_c \in \{\mathbf{U}_c = [\mu_{ik}] \in \mathbb{V}_{cN} \mid \mu_{ik} \in \{0 ; 1\} \quad \forall \quad i \in \mathbb{I}, k \in \mathbb{I}_N\} \quad (2.1)$$

und weicher

$$\mathbf{U}_f \in \{\mathbf{U}_f = [\mu_{ik}] \in \mathbb{V}_{cN} \mid \mu_{ik} \in [0 ; 1] \quad \forall \quad i \in \mathbb{I}, k \in \mathbb{I}_N\} \quad (2.2)$$

c -Partitionierung, d. h. der Strukturierung in c Einheiten. Dabei stellt

$$\mathbb{I} := \{1, \dots, c\} \quad (2.3)$$

die Menge aller c Clusternummern und

$$\mathbb{I}_N := \{1, \dots, N\} \quad (2.4)$$

die Menge aller Datensatznummern dar. Es gilt die Abkürzung

$$\mu_{ik} = \mu_i(\mathbf{x}_k) \quad . \quad (2.5)$$

Während bei der harten Partitionierung ein Datensatz \mathbf{x}_k i. allg. genau einer Partitionierung zugeordnet wird und damit

$$\sum_{i=1}^c \mu_{ik} = 1 \quad (2.6)$$

gilt, muß die durch Gl. (2.6) gegebene Bedingung bei der weichen Partitionierung nicht zwangsläufig erfüllt sein. Die meisten durch Fuzzy-Clustering bestimmten Partitionierungen erfüllen Gl. (2.6) und bilden eine Fuzzy-Partitionierung

$$\mathbf{U}_{fb} \in \{\mathbf{U}_{fb} = [\mu_{ik}] \in \mathbb{V}_{cN} \mid \mu_{ik} \in [0 ; 1], \sum_{i=1}^c \mu_{ik} = 1 \quad \forall \quad i \in \mathbb{I}, k \in \mathbb{I}_N\}, \quad (2.7)$$

die dann als Fuzzy-Informationssystem (Meyer-Gramann und Jüngst 1993) bezeichnet wird. Seit kurzer Zeit gibt es aber auch Ausnahmen (Krishnapuram und Keller 1993). Auch bei der Beschreibung der Partitionen jeweils über eine mehrdimensionale Gauß-Funktion wird Gl. (2.6) nicht erfüllt.

2.2 Zugehörigkeitsfunktionen vom Fuzzy-Clustertyp

Betrachtet werden Zugehörigkeitsfunktionen, die alle Partitionsschwerpunkte der c -Partition bewerten. Sie sind bereits vom Fuzzy- c -Means- (FCM) und Gustafson-und-Kessel-Algorithmus her bekannt und lassen sich wie folgt beschreiben (Bezdek 1981):

$$\tilde{\mu}_{ik}^{cl} = \left\{ \begin{array}{l} \frac{1}{\sum_{j=1}^c \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\|\mathbf{x}_k - \mathbf{v}_j\|} \right)^{\frac{2}{\nu-1}}} \quad \left. \begin{array}{l} \text{für } \mathbb{I}_k = \emptyset \\ \text{(keine Singularitäten)} \end{array} \right\} \\ \\ \left. \begin{array}{l} 0 \quad \forall i \in \tilde{\mathbb{I}}_k \\ a_{ik} \quad \forall i \in \mathbb{I}_k \end{array} \right\} \quad \left. \begin{array}{l} \text{für } \mathbb{I}_k \neq \emptyset \\ \text{(Singularitäten liegen vor)} \end{array} \right\} . \quad (2.8)$$

Dabei sei $\mathbb{I}_k = \{i \in \mathbb{I} \mid d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\|_D = 0\}$ die Menge aller Cluster, bzgl. derer \mathbf{x}_k den Abstand 0 vom Clusterschwerpunkt hat. $\tilde{\mathbb{I}}_k = \mathbb{I} \setminus \mathbb{I}_k$ ist die Menge aller Cluster, bzgl. derer \mathbf{x}_k keine Singularität darstellt. Außerdem gilt

$$\sum_{i \in \tilde{\mathbb{I}}_k} a_{ik} = \sum_{i \in \mathbb{I}_k} \mu_{ik} = 1 \quad \forall i \in \mathbb{I}_k \neq \emptyset , \quad (2.9)$$

was beispielsweise durch eine Wahl $a_{ik} = 1/\mathcal{A}(\mathbb{I}_k) \quad \forall i$ erreicht werden kann.

Die Bedingung nach Gl. (2.9) bedeutet, daß in jedem Punkt maximal ein Clusterschwerpunkt liegen darf. Da dies keine Einschränkung für eine Fuzzy-Partitionierung eines Fuzzy-Modells bedeutet, wird im folgenden stets

$$a_{ik} = 1 \quad (2.10)$$

gesetzt:

$$\mu_{ik}^{cl} = \left\{ \begin{array}{l} \frac{1}{\sum_{j=1}^c \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\|\mathbf{x}_k - \mathbf{v}_j\|} \right)^{\frac{2}{\nu-1}}} \quad \left. \begin{array}{l} \text{für } \mathbb{I}_k = \emptyset \\ \text{(keine Singularitäten)} \end{array} \right\} \\ \\ \left. \begin{array}{l} 0 \quad \forall i \in \tilde{\mathbb{I}}_k \\ 1 \quad \forall i \in \mathbb{I}_k \end{array} \right\} \quad \left. \begin{array}{l} \text{für } \mathbb{I}_k \neq \emptyset \\ \text{(Singularitäten liegen vor)} \end{array} \right\} . \quad (2.11)$$

Das heißt, \mathbb{I}_k ist immer eine leere oder einelementige Menge.

Die mathematischen Eigenschaften von μ^{cl} sind nicht offensichtlich. Deshalb sollen Betrachtungen zu Stetigkeit und Differenzierbarkeit vorgenommen werden. Es gilt

Satz 2.1 : Zugehörigkeitsfunktionen nach Gl. (2.11) sind stetig in \mathbf{x} . □

Beweis: Potentielle Unstetigkeitsstellen sind nur die Clusterzentren $\mathbf{v}_j, j \in \mathbb{I}$. Zuerst wird eine Annäherung an den eigenen Prototypen betrachtet:

$$\lim_{\|\delta\| \rightarrow 0} \mu_i^{cl}(\mathbf{x} = \mathbf{v}_i + \delta) = \lim_{\|\delta\| \rightarrow 0} \frac{1}{1 + \sum_{\substack{j=1 \\ j \neq i}}^c \left(\frac{\|\delta\|}{\|\mathbf{v}_i - \mathbf{v}_j + \delta\|} \right)^{\frac{2}{\nu-1}}} = 1 \quad , \quad i \in \mathbb{I} \quad (2.12)$$

Bei Annäherung an den Prototypen einer anderen Partition gilt:

$$\begin{aligned} \lim_{\|\delta\| \rightarrow 0} \mu_i^{cl}(\mathbf{x} = \mathbf{v}_f + \delta) &= \lim_{\|\delta\| \rightarrow 0} \frac{\|\mathbf{x} - \mathbf{v}_f\|^{\frac{2}{\nu-1}}}{\|\mathbf{x} - \mathbf{v}_i\|^{\frac{2}{\nu-1}} \sum_{j=1}^c \left(\frac{\|\mathbf{x} - \mathbf{v}_j\|}{\|\mathbf{x} - \mathbf{v}_j\|} \right)^{\frac{2}{\nu-1}}} \quad , \quad f \in \mathbb{I} \setminus \{i\} \\ &= \lim_{\|\delta\| \rightarrow 0} \frac{\|\mathbf{x} - \mathbf{v}_f\|^{\frac{2}{\nu-1}}}{\|\mathbf{x} - \mathbf{v}_i\|^{\frac{2}{\nu-1}} \left(1 + \sum_{\substack{j=1 \\ j \neq f}}^c \left(\frac{\|\mathbf{x} - \mathbf{v}_j\|}{\|\mathbf{x} - \mathbf{v}_j\|} \right)^{\frac{2}{\nu-1}} \right)} = 0 \end{aligned} \quad (2.13)$$

Damit folgt der vorstehende Satz. \square

Lemma: Zugehörigkeitsfunktionen nach Gl. (2.11) mit einer p -Norm als Abstandsnorm sind für $1 < \nu < 3$ nach \mathbf{x} stetig differenzierbar. \square

Beweis: Es erfolgt eine Betrachtung des Argumentes der p -Norm in Gl. (2.8) in ihren m einzelnen Koordinaten als

$$\mu_i(\mathbf{x}) = \left[\sum_{j=1}^c \left(\frac{|x_g - v_{g,i}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p}{|x_g - v_{g,j}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,j}|^p} \right)^{\frac{2}{p(\nu-1)}} \right]^{-1} \quad , \quad i \in \mathbb{I} \quad (2.14)$$

mit

$$\mathbf{x} = [x_1, \dots, x_g, \dots, x_m]^T \quad , \quad (2.15)$$

$$\mathbf{v}_j = [v_{1,j}, \dots, v_{g,j}, \dots, v_{m,j}]^T \quad , \quad (2.16)$$

$$\nu > 1 \quad \text{und}$$

$$p \geq 1 .$$

Im folgenden wird die stetige Differenzierbarkeit jeder einzelnen Koordinate von \mathbf{x} betrachtet. Aus deren Nachweis folgt die stetige Differenzierbarkeit von \mathbf{x} . Die einzigen Problemstellen sind die c Prototypen, die im folgendem betrachtet werden sollen.

Sei $x_g > v_{g,i}$ und $x_g > v_{g,j} \quad \forall i \in \mathbb{I}, j \in \mathbb{I} \setminus \{i\}$, dann gilt

$$\frac{\partial \mu_i(\mathbf{x})}{\partial x_g} = \frac{\frac{2}{p(\nu-1)}}{\left[1 + \sum_{\substack{j=1 \\ j \neq i}}^c \left(\frac{|x_g - v_{g,i}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p}{|x_g - v_{g,j}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,j}|^p} \right)^{\frac{2}{p(\nu-1)}} \right]^2} \quad (2.17a)$$

$$\cdot \sum_{\substack{j=1 \\ j \neq i}}^c \left[\left(\frac{|x_g - v_{g,i}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p}{|x_g - v_{g,j}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,j}|^p} \right)^{\frac{2/p+1-\nu}{\nu-1}} \right] \quad (2.17b)$$

$$\cdot \left[\frac{p|x_g - v_{g,i}|^{p-1} \left(|x_g - v_{g,j}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,j}|^p \right)}{\left(|x_g - v_{g,j}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,j}|^p \right)^2} \right] \quad (2.17c)$$

$$- \left[\frac{\left(|x_g - v_{g,i}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p \right) p|x_g - v_{g,j}|^{p-1}}{\left(|x_g - v_{g,j}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,j}|^p \right)^2} \right] \quad (2.17d)$$

Für den Fall $x_g < v_{g,i}$ ist der erste Faktor $p|x_g - v_{g,i}|^{p-1}$ in Teilgl. (2.17c) und (2.17d) durch $-p|x_g - v_{g,i}|^{p-1}$ zu ersetzen. Dies ist für die Grenzwertbetrachtung im folgenden wichtig, da die *beidseitigen* Grenzwerte der Ableitungen zu vergleichen sind. Für den Fall, daß $x_g > v_{g,j} \forall j \in \mathbb{I} \setminus \{i\}$ nicht gilt, ist folgendes zu ändern: Bei allen Summanden in Teilgl. (2.17b) bis (2.17d), bei denen $x_g < v_{g,j}$ gilt, ist der Faktor $p|x_g - v_{g,j}|^{p-1}$ durch $-p|x_g - v_{g,j}|^{p-1}$ zu ersetzen.

Fall 1: Betrachtet werde nun der Fall $x_g \rightarrow v_{g,i}$, dabei gelte:

$$\sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p = 0, \quad (2.18)$$

d. h., es erfolgt eine Annäherung von \mathbf{x} in x_g -Richtung an \mathbf{v}_i . In alle anderen Richtungen hat \mathbf{x} den Abstand 0 von \mathbf{v}_i . Für $x_g \rightarrow v_{g,i}$ geht der Nenner in Gl. (2.17a) gegen -1 , die Nenner in Gl. (2.17b) bis (2.17d) sind positiv. In den Zählern von Gl. (2.17b) bis (2.17d) sind die Terme

$$\left(|x_g - v_{g,i}|^p \right)^{\frac{2/p+1-\nu}{\nu-1}} |x_g - v_{g,i}|^{p-1} = |x_g - v_{g,i}|^{\frac{3-\nu}{\nu-1}} \quad (2.19)$$

und

$$(|x_g - v_{g,i}|^p)^{\frac{2/p+1-\nu}{\nu-1}} |x_g - v_{g,i}|^p = |x_g - v_{g,i}|^{\frac{2}{\nu-1}} \quad (2.20)$$

ausschlaggebend, es gilt

$$x_g \rightarrow v_{g,i} \Rightarrow \frac{\partial \mu_i(\mathbf{x})}{\partial x_g} \rightarrow \begin{cases} 0 & \text{für } 1 < \nu < 3 \\ \begin{matrix} 0 > a > -\infty & x_g > v_{g,i} \\ 0 < a < \infty & x_g < v_{g,i} \end{matrix} & \text{für } \nu = 3 \\ \begin{matrix} -\infty & x_g > v_{g,i} \\ \infty & x_g < v_{g,i} \end{matrix} & \text{für } \nu > 3 \end{cases} \quad (2.21)$$

mit einer Konstanten a .

Fall 2: Betrachtet wird nun der Fall $x_g \rightarrow v_{g,h} \neq v_{g,i}$ und dazu die rechte Seite von Gl. (2.17a) mit $(|x_g - v_{g,h}|/|x_g - v_{g,i}|)^{4/(\nu-1)}$ erweitert. Es gelte

$$\sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,h}|^p = 0 \quad , \quad (2.22)$$

d. h., es erfolgt eine Annäherung von \mathbf{x} in x_g -Richtung an v_h . Für $x_g \rightarrow v_{g,h}$ geht dann der Nenner aus Gl. (2.17a) gegen $-|x_g v_{g,i}|^{4/(\nu-1)} \neq 0$. Im Zähler von Gln. (2.17b) bis (2.17d) ist der Summand für $j = h$ für den Grenzwert entscheidend. Dieser wird durch die Terme (Vorfaktoren ungleich Null werden zur Verbesserung der Übersichtlichkeit weggelassen)

$$\left(\frac{1}{|x_g - v_{g,h}|^p} \right)^{\frac{2/p+1-\nu}{\nu-1}} \frac{|x_g - v_{g,h}|^p}{|x_g - v_{g,h}|^{2p}} |x_g - v_{g,h}|^{\frac{4}{\nu-1}} = |x_g - v_{g,h}|^{\frac{2}{\nu-1}} \quad (2.23)$$

und

$$\left(\frac{1}{|x_g - v_{g,h}|^p} \right)^{\frac{2/p+1-\nu}{\nu-1}} \frac{|x_g - v_{g,h}|^{p-1}}{|x_g - v_{g,h}|^{2p}} |x_g - v_{g,h}|^{\frac{4}{\nu-1}} = |x_g - v_{g,h}|^{\frac{3-\nu}{\nu-1}} \quad (2.24)$$

bestimmt, so daß

$$x_g \rightarrow v_{g,h} \Rightarrow \frac{\partial \mu_i(\mathbf{x})}{\partial x_g} \rightarrow \begin{cases} 0 & \text{für } 1 < \nu < 3 \\ \begin{matrix} 0 < a < \infty & x_g > v_{g,h} \\ 0 > a > -\infty & x_g < v_{g,h} \end{matrix} & \text{für } \nu = 3 \\ \begin{matrix} \infty & x_g > v_{g,h} \\ -\infty & x_g < v_{g,h} \end{matrix} & \text{für } \nu > 3 \end{cases} \quad (2.25)$$

mit einer Konstanten a gilt.

Aus der Gleichheit von rechts- und linksseitigen Grenzwerten folgt der vorstehende Satz. \square

Lemma: Zugehörigkeitsfunktionen nach Gl. (2.11) mit einer p -Norm als Abstandsnorm sind für $1 < \nu < 3$ nach \mathbf{v}_i stetig differenzierbar. \square

Beweis: Ausgehend von Gl. (2.14), (2.15) und (2.16) gilt für $x_g > v_{g,i}$

$$\begin{aligned} \frac{\partial \mu_i(\mathbf{x})}{\partial v_{g,i}} &= \frac{2}{p(\nu-1)} \sum_{\substack{j=1 \\ j \neq i}}^c \left\{ \left(\frac{|x_g - v_{g,i}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p}{|x_g - v_{g,j}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,j}|^p} \right)^{\frac{2/p+1-\nu}{(\nu-1)}} \right. \\ &\quad \cdot \left. \frac{-p|x_g - v_{g,i}|^{p-1}}{|x_g - v_{g,j}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,j}|^p} \right\} \\ &\quad \cdot (-1) \left[1 + \sum_{\substack{j=1 \\ j \neq i}}^c \left(\frac{|x_g - v_{g,i}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p}{|x_g - v_{g,j}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,j}|^p} \right)^{\frac{2}{p(\nu-1)}} \right]^{-2} \end{aligned} \quad (2.26)$$

Für $x_g < v_{g,i}$ ist der Faktor $-p|x_g - v_{g,i}|^{p-1}$ im Zähler durch $p|x_g - v_{g,i}|^{p-1}$ zu ersetzen.

Fall 1: Betrachtet wurde nun der Fall $x_g \rightarrow v_{g,i}$, dabei gelte:

$$\sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p = 0 \quad , \quad (2.27)$$

d. h. es erfolgt eine Annäherung von \mathbf{x} in x_g -Richtung an \mathbf{v}_i . Für $x_g \rightarrow v_{g,i}$ bleibt der Nenner in Gl. (2.26) eine endliche negative Zahl. Im Zähler ist der Term

$$|x_g - v_{g,i}|^{\frac{2+p-\nu p}{\nu-1}} |x_g - v_{g,i}|^{p-1} = |x_g - v_{g,i}|^{\frac{3-\nu}{\nu-1}}$$

ausschlaggebend. Es gilt:

$$x_g \rightarrow v_{g,i} \Rightarrow \frac{\partial \mu_i(\mathbf{x})}{\partial v_{g,i}} \rightarrow \begin{cases} 0 & \text{für } 1 < \nu < 3 \\ \left. \begin{array}{l} 0 < a < \infty \\ 0 > a > -\infty \end{array} \right\} \begin{array}{l} x_g > v_{g,i} \\ x_g < v_{g,i} \end{array} & \text{für } \nu = 3 \\ \left. \begin{array}{l} \infty \\ -\infty \end{array} \right\} \begin{array}{l} x_g > v_{g,i} \\ x_g < v_{g,i} \end{array} & \text{für } \nu > 3 \end{cases} \quad (2.28)$$

Fall 2: Betrachtet wird nun der Fall $x_g \rightarrow v_{g,h} \neq v_{g,i}$ und dazu die rechte Seite von Gl. (2.26) mit $(|x_g - v_{g,h}| |x_g - v_{g,i}|)^{4/(\nu-1)}$ erweitert. Es gelte

$$\sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,h}|^p = 0 \quad , \quad (2.29)$$

d. h., es erfolgt eine Annäherung von \mathbf{x} in x_g -Richtung an \mathbf{v}_h . Für $x_g \rightarrow v_{g,h}$ geht dann der Nenner in Gl. (2.26) gegen eine negative endliche Zahl. Im Zähler ist der Term

$$\left(\frac{1}{|x_g - v_{g,h}|} \right)^{\frac{2+p-\nu}{\nu-1}} \frac{1}{|x_g - v_{g,h}|^p} |x_g - v_{g,h}|^{\frac{4}{\nu-1}} = |x_g - v_{g,h}|^{\frac{2}{\nu-1}} \quad (2.30)$$

des Summanden für $j = h$ ausschlaggebend. Es folgt

$$x_g \rightarrow v_{g,h} \Rightarrow \frac{\partial \mu_i(\mathbf{x})}{\partial v_{g,i}} \rightarrow 0 \quad . \quad (2.31)$$

Aus der Gleichheit von rechts- und linksseitigen Grenzwerten folgt der vorstehende Satz. \square

Lemma: Zugehörigkeitsfunktionen nach Gl. (2.11) mit einer p -Norm als Abstandsnorm sind für $1 < \nu < 3$ nach $\mathbf{v}_i \neq \mathbf{v}_j$ stetig differenzierbar. \square

Beweis: Ausgehend von den Gln. (2.14), (2.15) und (2.16) gilt für $x_g > v_{g,h} \neq v_{g,i}$

$$\frac{\partial \mu_i(\mathbf{x})}{\partial v_{g,h}} = \frac{\frac{-2}{p(\nu-1)} \frac{\left(|x_g - v_{g,i}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p \right)^{\frac{2}{p(\nu-1)}}}{\left(|x_g - v_{g,h}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,h}|^p \right)^{\frac{2/p-1+\nu}{\nu-1}}} p |x_g - v_{g,h}|^{p-1} (-1)}{\left[\sum_{j=1}^c \left(\frac{\left(|x_g - v_{g,i}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p \right)^{\frac{2}{p(\nu-1)}}}{|x_g - v_{g,j}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,j}|^p} \right)^2 \right]} \quad . \quad (2.32)$$

Für $x_g < v_{g,h}$ ist der letzte Faktor (-1) im Zähler durch $(+1)$ zu ersetzen.

Fall 1: Betrachtet wird der Fall $x_g \rightarrow v_{g,h} \neq v_{g,i}$. Dazu wird die rechte Seite von Gl. (2.32) mit $(|x_g - v_{g,h}|/|x_g - v_{g,h}|)^{4/(\nu-1)}$ erweitert. Es gelte

$$\sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,h}|^p = 0 \quad , \quad (2.33)$$

d. h., es erfolgt eine Annäherung von \mathbf{x} in x_g -Richtung an \mathbf{v}_h . Für $x_g \rightarrow v_{g,h}$ geht der Nenner in Gl. (2.32) dann gegen eine negative endliche Zahl. Im Zähler ist der Term

$$\left(\frac{1}{|x_g - v_{g,h}|} \right)^{\frac{2-p+\nu}{\nu-1}} |x_g - v_{g,h}|^{p-1} |x_g - v_{g,h}|^{\frac{4}{\nu-1}} = |x_g - v_{g,h}|^{\frac{3-\nu}{\nu-1}} \quad (2.34)$$

ausschlaggebend. Es folgt

$$x_g \rightarrow v_{g,h} \Rightarrow \frac{\partial \mu_i(\mathbf{x})}{\partial v_{g,h}} \rightarrow \begin{cases} 0 & \text{für } 1 < \nu < 3 \\ \left. \begin{array}{l} 0 > a > -\infty \\ 0 < a < \infty \end{array} \right\} \begin{array}{l} x_g > v_{g,h} \\ x_g < v_{g,h} \end{array} & \text{für } \nu = 3 \\ \left. \begin{array}{l} -\infty \\ \infty \end{array} \right\} \begin{array}{l} x_g > v_{g,h} \\ x_g < v_{g,h} \end{array} & \text{für } \nu > 3 \end{cases} . \quad (2.35)$$

Fall 2: Betrachtet wird nun der Fall $x_g \rightarrow v_{g,i} \neq v_{g,h}$. Es gelte

$$\sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p = 0 \quad , \quad (2.36)$$

d. h., es erfolgt eine Annäherung von \mathbf{x} in x_g -Richtung an \mathbf{v}_i . Damit kann direkt

$$x_g \rightarrow v_{g,i} \Rightarrow \frac{\partial \mu_i(\mathbf{x})}{\partial v_{g,i}} \rightarrow 0 \quad (2.37)$$

aus Gl. (2.32) abgelesen werden.

Aus der Gleichheit von rechts- und linksseitigen Grenzwerten folgt der vorstehende Satz. \square

Lemma: Zugehörigkeitsfunktionen nach Gl. (2.11) mit einer p -Norm als Abstandsnorm sind für $\nu > 1$ nach ν stetig differenzierbar. \square

Beweis: Ausgehend von den Gln. (2.14), (2.15) und (2.16) gilt

$$\frac{\partial \mu_i(\mathbf{x})}{\partial \nu} = \frac{\sum_{\substack{j=1 \\ j \neq i}}^c (d_{rel,j})^{\frac{2}{p(\nu-1)}} \frac{-2/p}{(\nu-1)^2} \ln(d_{rel,j})}{\left[1 + \sum_{\substack{j=1 \\ j \neq i}}^c (d_{rel,j})^{\frac{2}{p(\nu-1)}} \right]^2} \quad (2.38)$$

mit

$$d_{rel,j} = \frac{|x_g - v_{g,i}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p}{|x_g - v_{g,j}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,j}|^p} \quad , \quad j \in \mathbb{I} \quad . \quad (2.39)$$

Fall 1: Betrachtet wird der Fall $x_g \rightarrow v_{g,i}$, dabei gelte

$$\sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p = 0 \quad , \quad (2.40)$$

d. h., es erfolgt eine Annäherung von \mathbf{x} in x_g -Richtung an \mathbf{v}_g . Der Fall $x_g \rightarrow v_{g,i}$ ist gleichbedeutend mit $d_{rel,j} \rightarrow 0 \forall j \in \mathbb{I} \setminus \{i\}$.

Im Zähler von Gl. (2.38) liegt der Fall $-\infty/\infty$ vor. Nach de l' Hospital gilt

$$\begin{aligned} \lim_{d_{rel} \rightarrow 0^+} \frac{\ln(d_{rel,j})}{(d_{rel,j})^{\frac{-2/p}{\nu-1}}} &= \lim_{d_{rel,j} \rightarrow 0^+} \frac{\frac{1}{d_{rel,j}}}{\frac{-2/p}{\nu-1} (d_{rel,j})^{\frac{-2/p-\nu+1}{\nu-1}}} \\ &= \frac{\nu-1}{-2/p} \lim_{d_{rel,j} \rightarrow 0^+} (d_{rel,j})^{\frac{2/p}{\nu-1}} = 0 . \end{aligned} \quad (2.41)$$

Es folgt

$$x_g \rightarrow v_{g,i} \Rightarrow \frac{\partial \mu_i(\mathbf{x})}{\partial \nu} \rightarrow 0 . \quad (2.42)$$

Fall 2: Betrachtet wird nun der Fall $x_g \rightarrow v_{g,h} \neq v_{g,i}$, dabei gelte

$$\sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,h}|^p = 0 , \quad (2.43)$$

d. h., es erfolgt eine Annäherung von \mathbf{x} in x_g -Richtung an $\mathbf{v}_h \neq \mathbf{v}_i$. Gl. (2.38) lautet ausführlich

$$\begin{aligned} \frac{\partial \mu_i(\mathbf{x})}{\partial \nu} &= \frac{\frac{-2/p}{(\nu-1)^2}}{\left[1 + \sum_{\substack{j=1 \\ j \neq i}}^c \left(\frac{|x_g - v_{g,i}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p}{|x_g - v_{g,j}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,j}|^p} \right)^{\frac{2}{p(\nu-1)}} \right]^2} \\ &\cdot \sum_{\substack{j=1 \\ j \neq i}}^c \left[\left(\frac{|x_g - v_{g,i}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p}{|x_g - v_{g,j}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,j}|^p} \right)^{\frac{2/p}{\nu-1}} \ln \left(\frac{|x_g - v_{g,i}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p}{|x_g - v_{g,j}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,j}|^p} \right) \right] . \end{aligned} \quad (2.44)$$

Die rechte Seite von Gl. (2.44) wird mit $(|x_g - v_{g,h}|/|x_g - v_{g,h}|)^{4/(\nu-1)}$ erweitert. Für $x_g \rightarrow v_{g,h}$ bleibt der Nenner $\neq 0$. Ausschlaggebend ist der Summand

$$\left(\frac{1}{|x_g - v_{g,h}|} \right)^{\frac{2}{\nu-1}} |x_g - v_{g,h}|^{\frac{4}{\nu-1}} = |x_g - v_{g,h}|^{\frac{2}{\nu-1}} \quad (2.45)$$

für $j = h$ im Zähler, der für $x_g \rightarrow v_{g,h}$ gegen Null geht, sowie der Logarithmus, der gegen Unendlich geht. Nach de l' Hospital gilt (Fall $x_g > v_{g,h}$):

$$\begin{aligned} & \lim_{x_g \rightarrow v_{g,h}} \frac{\ln \left(\frac{|x_g - v_{g,i}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p}{|x_g - v_{g,j}|^p + \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,j}|^p} \right)}{|x_g - v_{g,h}|^{\frac{-2}{\nu-1}}} \\ &= \lim_{x_g \rightarrow v_{g,h}} \frac{\frac{|x_g - v_{g,h}|^p}{|x_g - v_{g,i}|^{p+q_i}} - \frac{p|x_g - v_{g,i}|^{p-1}|x_g - v_{g,h}|^p - (|x_g - v_{g,i}|^{p+q_i})^p |x_g - v_{g,h}|^{p-1}}{|x_g - v_{g,h}|^{2p}}}{\frac{-2}{\nu-1}|x_g - v_{g,h}|^{\frac{-1-\nu}{\nu-1}}} \end{aligned} \quad (2.46)$$

mit $q_i = \sum_{\substack{f=1 \\ f \neq g}}^m |x_f - v_{f,i}|^p$. Wird Gl. (2.46) in einen Minuend und einen Subtrahend zerlegt, so sind die Terme

$$|x_g - v_{g,h}|^p \frac{1}{|x_g - v_{g,h}|^{\frac{-1-\nu}{\nu-1}}} |x_g - v_{g,h}|^{-p} = |x_g - v_{g,h}|^{\frac{\nu+1}{\nu-1}}$$

bzw.

$$|x_g - v_{g,h}|^p \frac{1}{|x_g - v_{g,h}|^{\frac{-1-\nu}{\nu-1}}} |x_g - v_{g,h}|^{-p-1} = |x_g - v_{g,h}|^{\frac{2}{\nu-1}}$$

ausschlaggebend. Beide gehen für $x_g \rightarrow v_{g,h}$ gegen Null. Für den Fall $x_g < v_{g,h}$ folgt das gleiche Ergebnis (es ändern sich nur einige Vorzeichen in Gl. (2.46)). Es folgt

$$x_g \rightarrow v_{g,h} \Rightarrow \frac{\partial \mu_i(\mathbf{x})}{\partial \nu} \rightarrow 0 \quad (2.47)$$

und damit der vorstehende Satz. \square

Satz: Zugehörigkeitsfunktionen nach Gl (2.11) mit einer p -Norm als Abstandsnorm sind für $1 < \nu < 3$ nach der Variablen \mathbf{x} und allen Parametern $(\mathbf{v}_i, \mathbf{v}_j, \nu)$ stetig differenzierbar. \square

Beweis: Folgt direkt aus den vorstehenden Lemmata. \square

2.3 Interpretation der Abhängigkeit vom Unschärfeparameter

Im folgenden werden die charakteristischen Bereiche bezüglich des Unschärfeparameters ν interpretiert. Dann wird Gleichung (2.11) mit

$$d_{jk} = \|\mathbf{x}_k - \mathbf{v}_j\|, \quad d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\|, \quad (2.48)$$

dem durch d_j relativierten Abstand d_i

$$d_{ijk} = d_{ik}/d_{jk} \quad (2.49)$$

von \mathbf{x}_k zu \mathbf{v}_i , und

$$\kappa = \frac{2}{\nu - 1} \quad (2.50)$$

umgeschrieben in

$$\mu_{ik}^{cl} = \left\{ \begin{array}{l} \frac{1}{\sum_{j=1}^c (d_{ijk})^\kappa} \\ 0 \quad \forall i \in \tilde{\mathbb{I}}_k \\ 1 \quad \forall i \in \mathbb{I}_k \end{array} \right\} \begin{array}{l} \text{für } \mathbb{I}_k = \emptyset \\ \text{(keine Singularitäten)} \\ \text{für } \mathbb{I}_k \neq \emptyset \\ \text{(Singularitäten liegen vor)} \end{array} . \quad (2.51)$$

Dabei bewirkt κ eine Progression (für $\kappa > 1$ bzw. $\nu < 3$) oder Degression (für $\kappa < 1$ bzw. $\nu > 3$) der *Abstandsbestrafung* bei der Berechnung der Zugehörigkeit gemäß Gl. (2.23) (Bild 2.1).

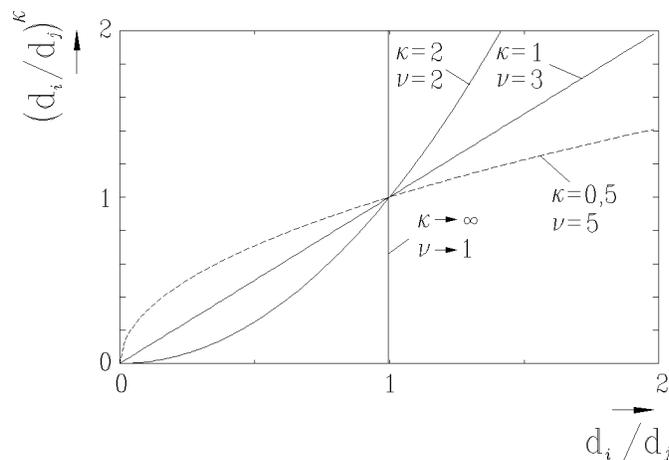


Bild 2.1: Progressive ($\nu < 3$), konstante ($\nu = 3$) oder degressive ($\nu > 3$) Gewichtung durch κ bzw. ν

Durch κ wird aber auch *grundlegend* die Zuordnungsvorschrift zwischen Daten und Partitionen beeinflusst. Betrachtet werde die i -te Partition: $\kappa \rightarrow \infty$ ($\nu \rightarrow 1$) führt zu einer harten Zuordnung der Daten und damit scharfen Zugehörigkeitsfunktion, da, falls ein $j \neq i$ mit $d_{ijk} > 1$ existiert, daraus $\mu_{ik} = 0$ resultiert. Eine Wahl $\infty > \kappa > 1$ ($1 < \nu < 3$) bedeutet, daß kleine Abstände ($d_{ijk} < 1$, d. h., ein Datum ist näher am i -ten als am j -ten Prototypen) unterproportional und große Abstände ($d_{ijk} > 1$) überproportional bestraft werden. Dies führt zu einer Partitionierung unter dem Aspekt, Anhäufungen von Daten durch ein Cluster so zu beschreiben, daß nahe beim Prototypen liegende Daten wenig bestraft werden und somit eine große Zugehörigkeit erhalten. Weit entfernte werden jedoch deutlich ausgegrenzt (üblicher Ansatz bei Fuzzy-Clusterverfahren).

Bei $\kappa < 1$ ($\nu > 3$) werden kleine Abstände ($d_{ijk} < 1$) überproportional bestraft. Das führt dazu, daß Daten selbst nahe dem Schwerpunkt der Partition nur noch relativ gering zugeordnet werden. Die resultierenden Zugehörigkeitsfunktionen fallen sehr schnell um den Partitionsschwerpunkt ab und streben schnell zum asymptotischen Endwert $\mu = 1/c$. Diese Tendenz führt für $\kappa \rightarrow 0$ zu „Punktclustern“, da dann nur noch für singuläre Werte $\mathbf{x}_k = \mathbf{v}_i$ eine Zuordnung mit $\mu_i(\mathbf{x}_k) = 1$ zum i -ten Cluster erfolgt und sonst für alle nicht singularen \mathbf{x}_k $\mu(\mathbf{x}_k) = 1/c$ gilt. Eine solche Partitionierung ist offensichtlich nicht erstrebenswert. Für eine Fuzzy-Partitionierung, bei der Bereiche mit Datenhäufungen zusammengefaßt werden sollen, ist deshalb $1 < \nu < 3$ zu wählen. Die Wahl von ν wird genauer von Kroll (1994b) diskutiert.

2.4 Zugehörigkeitsfunktionen vom e-Typ

Die Eigenschaften von Zugehörigkeitsfunktionen vom e-Typ

$$\mu_i^{ga}(\mathbf{x}) = \mu_{i,0} \exp(-\|\mathbf{x} - \mathbf{v}_i\|^2 / \tilde{\sigma}_i^2) \quad (2.52)$$

hängen von der Wahl der Abstandsnorm $\|\cdot\|$ ab. Für die euklidische Norm $\|\cdot\|_2$ ist μ_i^{ga} offensichtlich sowohl stetig als auch differenzierbar. Sei

$$\Delta \mathbf{x}_i := \mathbf{x} - \mathbf{v}_i \quad (2.53)$$

und

$$\|\Delta \mathbf{x}_i\| = \Delta \mathbf{x}_i^T \mathbf{D} \Delta \mathbf{x}_i \quad (2.54)$$

eine innere Produktnorm mit der Formmatrix (Bronstein und Semendjajew 1991)

$$\mathbf{D} = \begin{bmatrix} d_{D_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & d_{D_m} \end{bmatrix} . \quad (2.55)$$

Dann kann die mehrdimensionale Zugehörigkeitsfunktion durch Gl. (2.52) multiplikativ aus ihren m eindimensionalen Anteilen zusammengesetzt werden:

$$\begin{aligned} \mu_i^{ga}(\mathbf{x}) &= \mu_{i,0} \exp\left(-\sum_{j=1}^m d_{D_j} |x_j - v_j|^2 / \tilde{\sigma}_i^2\right) \\ &= \mu_{i,0} \prod_{j=1}^m \exp(-d_{D_j} |x_j - v_j|^2 / \tilde{\sigma}_i^2) . \end{aligned} \quad (2.56)$$

Dies ist das gleiche Ergebnis wie bei Vorgabe eindimensionaler gaußförmiger Fuzzy-Referenzmengen und dem Produktoperator als Partialprämissenverknüpfungsoperator. Falls eine Skalierung der Meßdaten vor der Verarbeitung erwünscht ist, so kann sie auch durch entsprechende Wahl der d_{D_j} ($1 \leq j \leq m$) ersetzt werden.

Shimojima u. a. (1994) stellen eine Modifikation eindimensionaler Gaußfunktionen zur Erzielung eines linearen Kernbereiches (wie bei trapezoiden Zugehörigkeitsfunktionen) vor (i -te Regel):

$$\mu_j(x_j) = \begin{cases} \exp(-\tilde{\sigma}_{ij}(|x_j - v_{ij}| - c_{ij})) & \forall |x_j - v_{ij}| > c_{ij} \\ 1 & \forall |x_j - v_{ij}| \leq c_{ij} \end{cases} \quad (2.57)$$

mit

$$\mathbf{x}^T = [x_1, \dots, x_j, \dots, x_m] \quad \text{und} \quad (2.58)$$

$$\mathbf{v}_i^T = [v_{i,1}, \dots, v_{i,j}, \dots, v_{i,m}] \quad . \quad (2.59)$$

Eine Erweiterung auf mehrdimensionale Zugehörigkeitsfunktionen und allgemeine p -Normen folgt als

$$\mu_i(\mathbf{x}) = \exp\left(-\sqrt[p]{\sum_{j=1}^m f_{ij}(\mathbf{x})}\right) \quad (2.60)$$

mit

$$f_{ij}(\mathbf{x}) = \begin{cases} \tilde{\sigma}_{ij}(|x_j - v_{ij}|^p - \tilde{c}_{ij}) & \forall |x_j - v_{ij}|^p > \tilde{c}_{ij} \\ 0 & \forall |x_j - v_{ij}|^p \leq \tilde{c}_{ij} \end{cases} \quad . \quad (2.61)$$

Der lineare Kernbereich der i -ten Regel erstreckt sich von $-\tilde{c}_{ij}$ bis $+\tilde{c}_{ij}$ in Richtung der j -ten Eingangsgröße.

2.5 Abstandsnormen

Für eine Wahl $d_{Dj} = 1 \quad \forall j \in \{1, \dots, m\}$ stellt Gl. (2.54) gerade die euklidische Norm und Gl. (2.56) eine mehrdimensionale gaußsche Glockenkurve dar, die konzentrische zirkulare α -Schnitte besitzt. Die α -Schnitte durch Gl. (2.56) bei Wahl unterschiedlicher d_{Dj} in Gl. (2.54) sind dagegen exzentrisch in Richtung der Koordinaten. Bezdek (1981) normiert beispielsweise alle Koordinaten auf ihre jeweilige Streuung:

$$d_{Dj}^{-1} = \sigma_j^2 = \frac{1}{N} \sum_{k=1}^N (x_j(k) - \bar{x}_j)^2 \quad \forall j \in \{1, \dots, m\} \quad . \quad (2.62)$$

Für eine beliebige Orientierung exzentrischer α -Schnitte müssen auch Nebendiagonalelemente von \mathbf{D} besetzt werden (ähnlich wie beim Gustafson-und-Kessel-Algorithmus (Bezdek 1981)). Ein Beispiel hierzu ist die Mahalanobisnorm. Ihre Formmatrix ist die Inverse der Kovarianzmatrix, die für die Meßdaten \mathbf{x} wie folgt definiert ist (Hartung und Elpelt 1992):

$$\mathbf{D} = [\text{cov}(\mathbf{x})]^{-1} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & \\ \sigma_{m1} & \cdots & \cdots & \sigma_m^2 \end{bmatrix}^{-1} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{12} & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & \\ \sigma_{1m} & \cdots & \cdots & \sigma_m^2 \end{bmatrix}^{-1}, \quad (2.63)$$

da die Kovarianzmatrix symmetrisch ist (Hartung und Elpelt 1992). Dabei berechnen sich die Varianzen σ_j^2 gemäß Gl. (2.62) und die Kovarianzen σ_{jl} als (Lipschutz 1980)

$$\sigma_{jl} = \frac{1}{N} \sum_{k=1}^N (x_j(k) - \bar{x}_j)(x_l(k) - \bar{x}_l) \quad . \quad (2.64)$$

Die Kovarianzmatrix und damit auch die Formmatrix der Mahanobisnorm ist positiv definit (Flury und Riedwyl 1983).

Oft wird als Abstandsnorm auch eine allgemeine p -Norm (Kroll 1994a) verwendet. Eine Äquinormalkurve schneidet unabhängig von der Wahl von p die Koordinatenachsen immer im gleichen Punkt. Über die Wahl von p läßt sich der Verlauf zwischen diesen Schnittpunkten einstellen: Beispielsweise führt $p \rightarrow \infty$ zu hexaedrischen, $p = 2$ zu zirkularen und $p = 1$ zu oktaedrischen Äquinormalkurven. Dagegen beeinflußt eine innere Produktnorm die Gewichtung der Anteile der einzelnen Koordinaten. Die Freiheitsgrade beider Normen können zu einer *inneren Produkt- p -Norm* zusammengefügt werden:

$$\|\Delta \mathbf{x}_i\|_{D,p} = \sqrt[p]{(\Delta \mathbf{x}_i^{p/2})^T \mathbf{D} (\Delta \mathbf{x}_i^{p/2})} \quad . \quad (2.65)$$

Dabei ist \mathbf{D} eine positiv definite Matrix. Bei Wahl von \mathbf{D} als Diagonalmatrix nach Gl. (2.55) folgt

$$\|\Delta \mathbf{x}_i\|_{D,p} = \sqrt[p]{\sum_{i=1}^m d_{Di} |\Delta x_i|^p} \quad . \quad (2.66)$$

2.6 Diskussion der Zugehörigkeitsfunktionen

Zugehörigkeitsfunktionen vom μ^{cl} -Typ sind nicht konvex. Für Werte von ν nahe bei Eins und eine Betrachtung eines Bereiches, dessen Abmessungen in der Größenordnung der konvexen Hülle um alle Prototypen v_j liegt, sind die Abweichungen von einer konvexen Form relativ gering (Kroll 1994b). Der nicht konvexe Graph der Zugehörigkeitsfunktionen hängt ursächlich mit der Berücksichtigung aller Prototypen in jeder einzelnen Zugehörigkeitsfunktion zusammen. Dadurch können die μ_i^{cl} nicht unabhängig voneinander eingestellt werden. Ein isoliert lokales Lernen des Modells ist somit nicht möglich. Durch die Interpendenzen zwischen den einzelnen μ_i^{cl} können auch Zugehörigkeitsfunktionen mit nicht ellipsoiden α -Schnitten auftreten.

Zugehörigkeitsfunktionen vom e-Typ sind dagegen konvex und voneinander unabhängig und damit isoliert lokal lernbar, was je nach Anwendung vorteilhaft sein kann (Murray-Smith 1994). Während Zugehörigkeitsfunktionen vom μ^{cl} -Typ im allgemeinen ein Fuzzy-Informationssystem (Meyer-Gramann und Jüngst 1993) bilden, ist dies bei den μ^{ga} -Typen nicht der Fall.

Die explizit mehrdimensionalen Zugehörigkeitsfunktionen bieten den wichtigen Vorteil, daß die Auflösung in den einzelnen Koordinatenrichtungen nicht vorgegeben werden muß. Zum Beispiel durch Clusterverfahren können sie einfach und schnell so plaziert werden,

daß sie optimal (im Sinne des Clusterkriteriums) im Datenraum angeordnet sind. Dies erleichtert eine automatisierte Modellbildung erheblich. Bei Zugehörigkeitsfunktionen mit nicht beschränktem Trägerbereich (z. B. triangulare oder trapezoide) kann bei ungünstiger (meist zu feiner) Partitionierung und unvollständigen Testdaten (ungenügender Systemanregung bei der Aufzeichnung) der Fall eintreten, daß einige Regeln nicht aktiviert werden. Bei der Modellauswertung können diese Regeln durchaus angesprochen werden und liefern dann einen ihrer Initialisierung entsprechenden, und somit nicht immer sinnvollen, Beitrag. Zugehörigkeitsfunktionen mit räumlich nicht beschränktem Trägerbereich wie μ^{cl} und μ^{ga} dagegen führen zu einer im allgemeinen wünschenswerten Interpolation auch in nicht identifizierten Bereichen. Allerdings haben bei ihnen alle Meßdaten Einfluß bei der Identifizierung jeder einzelnen Regel. Ist ein lokales Lernen mit beschränktem Einflußbereich der Meßdaten (also beschränktem Trägerbereich) der Zugehörigkeitsfunktion erforderlich, so sind stückweise definierte Zugehörigkeitsfunktionen wie trapezoide, triangulare oder solche mit S-Flanken zu verwenden. Ist zusätzlich eine $(k-1)$ -fache stetige Differenzierbarkeit gefordert, so können z. B. B-Splines als Zugehörigkeitsfunktionen verwendet werden. Diese sind zudem konvex und lassen sich zu einem Fuzzy-Informationssystem kombinieren (nach Tschiesche 1992).

3 Analytische Beschreibung von Fuzzy-Systemen

In diesem Abschnitt wird eine analytische Beschreibung von Fuzzy-Systemen mit mehrdimensionalen Fuzzy-Referenzmengen hergeleitet. Besondere Beachtung finden dabei die in Abschnitt 2 vorgestellten Zugehörigkeitsfunktionen.

3.1 Funktionale Fuzzy-Systeme

Funktionale Fuzzy-Regeln mit mehrdimensionalen Fuzzy-Referenzmengen lassen sich als

$$\text{WENN } (\mathbf{x}(k) \text{ IST } \mathbf{v}_i) \text{ DANN } y_i(\mathbf{x}(k)) = f_i(\mathbf{x}(k)) \quad (3.1)$$

schreiben mit

- $x_j(k)$: scharfe Eingangsgröße, $j = 1, \dots, m$, $x_j \in \mathbb{D}_{x_j}$,
- $\mathbf{x}(k)$: 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 : Partitions- oder Clusterschwerpunkt der i -ten Partition bzw. des i -ten Clusters,
- $y_i(\mathbf{x}(k))$: scharfe Ausgangsgröße der i -ten Regel und
- $f_i(\mathbf{x}(k))$: Konklusionsfunktion der m scharfen Eingangsgrößen.

Ein Fuzzy-System umfaßt c Regeln, die über

$$y(\mathbf{x}(k)) = \frac{\sum_{i=1}^c y_i(\mathbf{x}(k))\omega_i(k)}{\sum_{i=1}^c \omega_i(k)} \quad (3.2)$$

zusammengefaßt werden. ω_i ist der Erfülltheitsgrad der i -ten Regel, der zu

$$\omega_i(k) = \mu_i(\mathbf{x}(k)) \quad (3.3)$$

gewählt wird. Eine Wahl

$$\omega_i(k) = f(\mu_i(\mathbf{x}(k))) \neq \mu_i(\mathbf{x}(k)), \quad (3.4)$$

bei der Identifikation derartiger Fuzzy-Modelle, um z. B. eine *Konzentration* oder *Kontrastverstärkung* (Zadeh 1973) zu erreichen, bringt keine wesentlichen Vorteile (Lohmann 1994).

Dynamische Fuzzy-Systeme

Jede der c Konklusionen wird aus einem ARX-Modell für das Kleinsignalverhalten um den jeweiligen Entwicklungspunkt angesetzt. Die externen Eingangsgrößen $u(k)$ sind direkte Meßgrößen oder eine analytische Funktion direkter Meßgrößen. Die Konklusionsfunktion soll den Großsignalwert $y_i(k)$ der Ausgangsgröße liefern. Damit läßt sich der ARX-Ansatz für jede Regel als

$$y_i(k) - y_{0,i} = - \sum_{j=1}^n a_{ji}(y(k-j) - y_{0,i}) + \sum_{f=1}^m b_{fi}(u(k-f) - u_{0,i}) \quad (3.5)$$

$$y_i(k) = y_{0,i}(1 + \sum_{j=1}^n a_{ji}) - u_{0,i} \sum_{j=1}^m b_{fi} - \sum_{j=1}^n a_{ji}y(k-j) + \sum_{j=1}^m b_{fi}u(k-f) \quad (3.6a)$$

$$= c_i - \sum_{j=1}^n a_{ji}y(k-j) + \sum_{f=1}^m b_{fi}u(k-f) \quad (3.6b)$$

beschreiben. Der Meßvektor ist damit

$$\mathbf{x}^T(k) = [y(k-1), \dots, y(k-n), u(k-1), \dots, u(k-m)] \quad (3.7)$$

Für die Wahl der Entwicklungspunkte $y_{0,i}$ und $u_{0,i}$ existieren zwei Alternativen:

1. Die Entwicklungspunkte werden über die Prototypen festgelegt oder
2. die Entwicklungspunkte können beliebig plziert werden (Kroll 1994a, Lohmann 1994, Totz 1995).

Im ersten Fall ist damit der konstante Term c_i in Gl. (3.6b) bei vorliegender Partitionierung festgelegt, also kein freier Modellparameter. Tritt eine Eingangsgröße mehrfach verzögert auf (d. h. $n > 1$ und/oder $m > 1$), so ist die Dimension des Eingangsgrößenraumes höher als die Anzahl der Meßgrößen und der Entwicklungspunkt ist nicht eindeutig festgelegt. Dann ist anzugeben, wie die entsprechenden n Komponenten eines jeden Prototypen \mathbf{v}_i bzgl. $y(k)$ bzw. die m Komponenten bzgl. $u(k)$ auf den Entwicklungspunkt $y_{0,i}$ bzw. $u_{0,i}$ abgebildet werden. Eine einfache Vorschrift besteht darin, die mit der am geringsten verzögerten Eingangsgröße $y(k-1)$ bzw. $u(k-1)$ aus $\mathbf{x}(k)$ korrespondierende Komponente eines jeden Prototypen zu verwenden.

Im zweiten Fall ist der konstante Term c_i in Gl. (3.6b) ein freier Modellparameter. Werden die Koeffizienten in Gl. (3.6b) identifiziert, so kann nicht mehr auf die Entwicklungspunkte zurückgeschlossen werden, da das zugehörige Gleichungssystem unterbestimmt ist. Bei einer peripheren Black box-Betrachtung ist dies unerheblich. Es erschwert allerdings die Interpretierbarkeit der Zusammenhänge in der Black box. Andererseits sind durch den zusätzlichen Freiheitsgrad pro Regel deutlich höhere Güten erzielbar.

Bei vorliegender Partitionierung läßt sich Gl. (3.5) im ersten Fall schreiben als

$$y_i(k+1) = y_{0,i} + \Theta_i^T(\mathbf{x}(k) - \mathbf{x}_{0,i}) \quad (3.8)$$

mit

$$\mathbf{x}_{0,i}^T = [y_{0,i}, \dots, y_{0,i}, u_{0,i}, \dots, u_{0,i}] \quad \text{und} \quad (3.9)$$

$$\Theta_i^T = [a_{1,i}, \dots, a_{n_i}, b_{1,i}, \dots, b_{m_i}] \quad . \quad (3.10)$$

Im zweiten Fall folgt

$$y_i(k+1) = [c_i \Theta_i^T] \begin{bmatrix} 1 \\ \mathbf{x}(k) \end{bmatrix} \quad . \quad (3.11)$$

Einsetzen in Gl. (3.2) mit $\omega_i(k) = \mu_i(\mathbf{x}(k))$ ergibt

$$y(k+1) = \frac{\sum_{i=1}^c (y_{0,i} + \Theta_i^T(\mathbf{x}(k) - \mathbf{x}_{0,i})) \mu_i(\mathbf{x}(k))}{\sum_{i=1}^c \mu_i(\mathbf{x}(k))} \quad . \quad (3.12)$$

Substitution von

$$\phi_i(\mathbf{x}(k)) = \phi_i(k) = \frac{\mu_i(\mathbf{x}(k))}{\sum_{i=1}^c \mu_i(\mathbf{x}(k))} \quad , \quad (3.13)$$

der sogenannten Fuzzy-Basisfunktion (Suh und Kim 1994, Wang 1994), liefert

$$y(k+1) = \sum_{i=1}^c y_{0,i} \phi_i(k) + \Theta_i^T(\mathbf{x}(k) - \mathbf{x}_{0,i}) \phi_i(k) \quad (3.14a)$$

$$= \Theta_{D1}^T \mathbf{X}_{D1}(k) \quad (3.14b)$$

mit

$$\Theta_{D1}^T = [y_{0,1}, \dots, y_{0,c}, \Theta_1^T, \dots, \Theta_c^T] \quad \text{und} \quad (3.15)$$

$$\mathbf{X}_{D1}^T(k) = [\phi_1(k), \dots, \phi_c(k), (\mathbf{x}^T(k) - \mathbf{x}_{0,1}^T) \phi_1(k), \dots, (\mathbf{x}^T(k) - \mathbf{x}_{0,c}^T) \phi_c(k)] \quad . \quad (3.16)$$

Im zweiten Fall folgt

$$y_i(k+1) = [c_i \Theta_i^T] \begin{bmatrix} 1 \\ \mathbf{x}(k) \end{bmatrix} \quad , \quad (3.17)$$

$$y_i(k+1) = \frac{\sum_{i=1}^c (c_i + \Theta_i^T \mathbf{x}(k)) \mu_i(\mathbf{x}(k))}{\sum_{i=1}^c \mu_i(\mathbf{x}(k))} \quad (3.18a)$$

$$= \sum_{i=1}^c c_i \phi_i(k) + \Theta_i^T \mathbf{x}(k) \phi_i(k) \quad (3.18b)$$

$$= \Theta_{D_2}^T \mathbf{X}_{D_2}(k) \quad (3.18c)$$

mit

$$\Theta_{D_2}^T = [c_1, \dots, c_c, \Theta_1^T, \dots, \Theta_c^T] \quad \text{und} \quad (3.19)$$

$$\mathbf{X}_{D_2}^T(k) = [\phi_1(k), \dots, \phi_c(k), \mathbf{x}^T(k)\phi_1(k), \dots, \mathbf{x}^T(k)\phi_c(k)] \quad . \quad (3.20)$$

Bei gegebener Partitionierung lassen sich somit z. B. Least-Squares-Verfahren zur Identifikation der Konklusionsfunktionsparameter anwenden. Werden globale Zugehörigkeitsfunktionen verwendet, so ist $y(k+1)$ als einheitlich definierte Funktion der Eingangsgrößen $\mathbf{x}(k)$ darstellbar. Bei Zugehörigkeitsfunktionen, die ein Fuzzy-Informationssystem bilden, vereinfacht sich Gl. (3.13) zu

$$\phi_i(k) = \mu_i(\mathbf{x}(k)) \quad . \quad (3.21)$$

Diese Bedingung ist für μ^{cl} erfüllt und es folgt im ersten Fall

$$y(k+1) = \sum_{i=1}^c [y_{o,i} + \Theta_i^T (\mathbf{x}(k) - \mathbf{x}_{o,i})] \left[\sum_{j=1}^c \frac{\|\mathbf{x}(k) - \mathbf{v}_i\|^\kappa}{\|\mathbf{x}(k) - \mathbf{v}_j\|^\kappa} \right]^{-1} \quad (3.22a)$$

$$= \frac{\sum_{i=1}^c [y_{o,i} + \Theta_i^T (\mathbf{x}(k) - \mathbf{x}_{o,i})] \|\mathbf{x}(k) - \mathbf{v}_i\|^{-\kappa}}{\sum_{j=1}^c \|\mathbf{x}(k) - \mathbf{v}_j\|^{-\kappa}} \quad . \quad (3.22b)$$

Bei μ^{ga} gilt

$$y(k+1) = \frac{\sum_{i=1}^c [y_{i,0} + \Theta_i^T (\mathbf{x}(k) - \mathbf{x}_{o,i})] \mu_{i,0} \exp(-\|\mathbf{x}(k) - \mathbf{v}_i\|^2 / \tilde{\sigma}_i^2)}{\sum_{i=1}^c \mu_{i,0} \exp(-\|\mathbf{x}(k) - \mathbf{v}_i\|^2 / \tilde{\sigma}_i^2)} \quad . \quad (3.23)$$

Fall 2 folgt analog. Die Erweiterung auf mehr als eine externe Eingangsgröße ist trivial und wird deshalb an dieser Stelle aus Übersichtlichkeitsgründen nicht angegeben.

Statisches Fuzzy-System

Die Schlußfolgerungen seien als in den m Eingangsgrößen lineare Polynome

$$y_i(k) = y_i(\mathbf{x}(k)) = p_i(0) + \sum_{a=1}^m p_i(a)x_a(k) = p_i(0) + \Theta_i^T \mathbf{x}(k) \quad (3.24)$$

gegeben. Dabei werden direkt Großsignalgrößen miteinander verknüpft, wodurch eine explizite Behandlung der Entwicklungspunkte wie bei den dynamischen Systemen entfällt. Die $x_a(k)$ sind externe Eingangsgrößen $u(k - \tau_u)$, $\tau_u \in \mathbb{N}_0$ oder analytische Funktionen von $u(k - \tau_u)$. Gleichung (3.24) läßt sich entsprechend den dynamischen Systemen als

$$y(k+1) = \sum_{i=1}^c p_i(0)\phi_i(k) + \Theta_i^T \mathbf{x}(k)\phi_i(k) \quad (3.25a)$$

$$= \Theta_S^T \mathbf{X}_S(k) \quad (3.25b)$$

mit

$$\Theta_S^T = [p_1(0), \dots, p_c(0), \Theta_1^T, \dots, \Theta_c^T] \quad \text{und} \quad (3.26)$$

$$\mathbf{X}_S^T(k) = [\phi_1(k), \dots, \phi_c(k), \mathbf{x}^T(k)\phi_1(k), \dots, \mathbf{x}^T(k)\phi_c(k)] \quad (3.27)$$

schreiben. Diese Form entspricht mathematisch (aber nicht inhaltlich) der des zweiten Falles bei den dynamischen Systemen. Zugehörigkeitsfunktion mit unbeschränktem Träger führen auch hier wieder zu einem einheitlich definierten funktionalen Zusammenhang zwischen $\mathbf{x}(k)$ und $y(k)$.

3.2 Relationale Fuzzy-Systeme

Relationale Fuzzy-Regeln mit mehrdimensionalen Fuzzy-Referenzmengen in den Prämissen lassen sich als

$$\text{WENN } (\mathbf{x}(k) \text{ IST } \mathbf{v}_i) \text{ DANN } (Y_i(\mathbf{x}(k)) \text{ IST } U_i)|_{g_i} \quad (3.28)$$

- $\mathbf{x}(k)$: 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 : eindimensionale Fuzzy-Referenzmenge bezüglich der Ausgangsgröße der i -ten Regel,
- $Y_i(\mathbf{x}(k))$: unscharfe (Partial-) Ausgangsgröße der i -ten Regel,
- U_i : eindimensionale Fuzzy-Referenzmenge der i -ten Regel bzgl. der Ausgangsgröße und
- g_i : Gewichtung ($g_i \in [0; 1]$) der i -ten Regel

schreiben. Der Erfülltheitsgrad der Prämisse berechnet sich zu

$$\omega_i(\mathbf{x}(k)) = \mu_i(\mathbf{x}(k)) \quad . \quad (3.29)$$

Als Fuzzy-Referenzmengen der Ausgangsgröße sollen Fuzzy-Mengen, die zu ihren Schwerpunkten s_i symmetrisch sind:

$$\mu_{U_i}(y = s_i - \delta) = \mu_{U_i}(y = s_i + \delta) \quad , \quad (3.30)$$

wie z. B. gleichschenklige Dreiecke, Trapeze oder Fuzzy-Einermengen gleicher Flächen Gewichte, dienen. Eine unterschiedliche Gewichtung der Regeln wird über die Gewichte g_i , die die Erfülltheitsgrade multiplikativ gewichten, erreicht. Die Implikation wurde nach Larsen (Lee 1990)

$$\mu_{U_i}(\mathbf{x})(y) = g_i \omega_i(\mathbf{x}(k)) \mu_{U_i}(y) = g_i \mu_i(\mathbf{x}(k)) \mu_{U_i}(y) \quad , \quad (3.31)$$

die Aggregation der c Regeln des Fuzzy-Systems über den Summationsoperator

$$\mu_Y(\mathbf{x}(k))(y) = \sum_{i=1}^c \mu_{Y_i}(\mathbf{x}(k))(y) = \sum_{i=1}^c g_i \mu_i(\mathbf{x}(k)) \mu_{U_i}(y) \quad (3.32a)$$

durchgeführt. Die Defuzzifizierung erfolgt über die Schwerpunkt-Methode mit:

$$y(\mathbf{x}(k)) = \frac{\sum_{i=1}^c g_i \mu_i(\mathbf{x}(k)) \mu_{U_i}(\mathbf{x}(k))(y)}{\sum_{i=1}^c g_i \mu_i(\mathbf{x}(k))} \quad (3.33a)$$

$$= \frac{\sum_{i=1}^c g_i \mu_i(\mathbf{x}(k)) s_i}{\sum_{i=1}^c g_i \mu_i(\mathbf{x}(k))} \quad . \quad (3.33b)$$

Die Substitution von

$$\phi_i(k) = \frac{g_i \mu_i(\mathbf{x}(k))}{\sum_{i=1}^c g_i \mu_i(\mathbf{x}(k))} \quad (3.34)$$

liefert

$$y(\mathbf{x}(k)) = \sum_{i=1}^c s_i \phi_i(k) \quad . \quad (3.35)$$

Eine Verwendung von μ^{cl} in Gl. (3.33b) ergibt

$$y(\mathbf{x}(k)) = \frac{\sum_{i=1}^c \frac{s_i g_i}{\sum_{j=1}^c \frac{\|\mathbf{x}(k) - \mathbf{v}_j\|^\kappa}}{\sum_{f=1}^c \frac{g_f}{\sum_{j=1}^c \frac{\|\mathbf{x}(k) - \mathbf{v}_j\|^\kappa}}}} \quad (3.36a)$$

$$= \frac{\sum_{i=1}^c s_i g_i \|\mathbf{x}(k) - \mathbf{v}_i\|^{-\kappa}}{\sum_{f=1}^c g_f \|\mathbf{x}(k) - \mathbf{v}_f\|^{-\kappa}} \quad (3.36c)$$

Zu bemerken ist, daß nur für $g_i = 1 \quad \forall i \in \mathbb{I}$ ein Fuzzy-Informationssystem vorliegt. Für μ^{ga} in Gl. (3.33b) folgt entsprechend

$$y(\mathbf{x}(k)) = \frac{\sum_{i=1}^c s_i g_i \mu_{i,0} \exp\left(-\|\mathbf{x}(k) - \mathbf{v}_i\|^2 / \tilde{\sigma}_i^2\right)}{\sum_{i=1}^c g_i \mu_{i,0} \exp\left(-\|\mathbf{x}(k) - \mathbf{v}_i\|^2 / \tilde{\sigma}_i^2\right)} \quad (3.37)$$

Wegen der Redundanz durch $g_i \mu_{i,0}$ kann ohne Beschränkung der Allgemeingültigkeit entweder $g_i \forall i \in \mathbb{I}$ oder $\mu_{i,0} \forall j \in \mathbb{I}$ identisch Eins gesetzt werden. Gleichung (3.35) kann als

$$y(\mathbf{x}(k)) = \mathbf{s}^T \boldsymbol{\varphi}(k) \quad (3.38)$$

mit

$$\mathbf{s}^T = [s_1, \dots, s_c] \quad \text{und} \quad (3.39)$$

$$\boldsymbol{\varphi}^T(k) = [\phi_1(k), \dots, \phi_c(k)] \quad (3.40)$$

dargestellt werden. Statische und dynamische Fuzzy-Systeme unterscheiden sich in der Wahl der Eingangsgrößen, wie bereits bei den funktionalen Systemen in Abschnitt 3.1 beschrieben, sowie in der zeitlichen Beziehung zwischen prädizierten zu ausgewerteten Größen. Bei dynamischen Fuzzy-Systemen gilt

$$y(\mathbf{x}(k)) = y(k+1) \quad , \quad (3.41)$$

bei statischen

$$y(\mathbf{x}(k)) = y(k) \quad . \quad (3.42)$$

4 Fuzzy-Systeme und Basisfunktionsnetzwerke

In diesem Abschnitt wird zuerst kurz die Verwandtschaft zwischen funktionalen und relationalen Fuzzy-Systemen besprochen und anschließend die zu künstlichen Neuronalen Netzen diskutiert. Während sich qualitative Vergleiche zwischen Fuzzy-Systemen und künstlichen Neuronalen Netzen z. B. bei Brown und Harris (1991), Kong und Kosko (1992) wie auch Yamakawa (1993) finden, wird genauer auf die Klasse der Radialbasisfunktionsnetzwerke und deren Verwandte eingegangen.

Die im Abschnitt 3.2 vorgestellten relationalen Fuzzy-Systeme stellen einen Spezialfall der in Abschnitt 3.1 erläuterten funktionalen Fuzzy-Systeme dar. Dabei wurde nur eine bestimmte Systemklasse (Implikationen nach Larsen, Summationsaggregation, symmetrischen und flächengleichen Fuzzy-Referenzmengen für die Konklusion und Schwerpunkt-Defuzzifizierung) betrachtet. Ein Vergleich der Gln. (3.12) bzw. (3.18a) mit (3.33a) zeigt, daß ein funktionales Modell mit konstanter Schlußfolgerung und zusätzlicher Gewichtung der Regeln (vergleichbar $\mu_{i,0}$ in Gl. (3.23)) die gleiche Form hat wie ein relationales. Wie im vorhergehenden Abschnitt erläutert, kann die konstante Gewichtung g_i als Vorfaktor μ_i hinzugerechnet werden, wodurch der Vergleich offensichtlich wird. Wird ein Fuzzy-Modell als gewichtete Superposition von c , in den c Partitionsschwerpunkten identifizierten Teilmodellen aufgefaßt, so entspricht ein relationales Fuzzy-Modell für eine Partition einer Taylorreihenentwicklung bis zum nullten Glied, ein funktionales jedoch einem bis zum ersten Glied um den Entwicklungspunkt. Während beim funktionalen Modell damit eine bessere Modellanpassung über die Konklusion angestrebt wird, soll dies beim relationalen Modell im allgemeinen über eine feinere Partitionierung erreicht werden. Somit wird mit relationalen Fuzzy-Modellen i. allg. eine punktweise, mit funktionalen dagegen eine bereichsweise Abbildung zwischen Ein- und Ausgangsraum beschrieben.

4.1 Radialbasisfunktionsnetzwerke

Eine Übernahme der Idee radialer Basisfunktionen (RBF), die bereits früher zur Approximation von Funktionen eingesetzt wurden (Powell 1987a,b), führt bei künstlichen Neuronalen Netzen zu sogenannten Radialbasisfunktionsnetzwerken, kurz „RBFN“ genannt (Poggio und Girosi 1989, Chen u. a. 1991, Jang und Sun 1993, Suh und Kim 1994). RBFN sind zweischichtige Netzwerke (Bild 4.1), deren Ein-/Ausgangsverhalten sich über

$$f_{RBF}(\mathbf{x}) = \alpha_0 + \sum_{i=1}^c \alpha_i \Psi_i(\|\mathbf{x} - \mathbf{v}_i\|) \quad (4.1)$$

beschreiben läßt. Dabei ist $\|\cdot\|$ eine radialsymmetrische Norm (i. allg. die euklidische), \mathbf{v}_i die Basis und α_i das Gewicht der i -ten Radialbasisfunktion $\Psi : \mathbb{R}^+ \rightarrow \mathbb{R}$. In der Literatur (Jang und Sun 1993) findet sich auch die Normierung

$$f_{RBF,n}(\mathbf{x}) = \frac{\sum_{i=1}^c \alpha_i \Psi_i(\|\mathbf{x} - \mathbf{v}_i\|)}{\sum_{i=1}^c \Psi_i(\|\mathbf{x} - \mathbf{v}_i\|)} \quad (4.2)$$

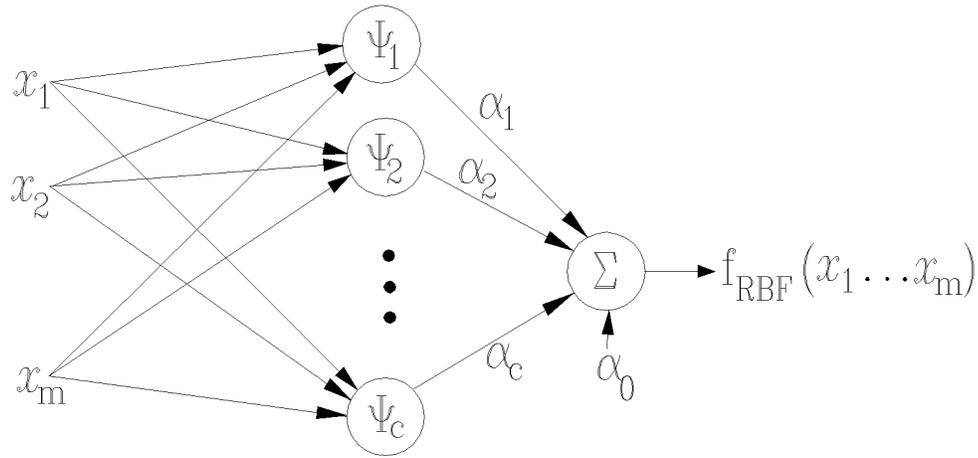


Bild 4.1: Struktur von Radialbasisfunktionsnetzwerken (Chen u. a. 1991)

mit $\alpha_0 \equiv 0$. Ist $\alpha_0 \neq 0$, so liefert ein Ansatz

$$f_{RBF,n}(\mathbf{x}) = \alpha_0 + \frac{\sum_{i=1}^c \alpha_i \Psi_i(\|\mathbf{x} - \mathbf{v}_i\|)}{\sum_{i=1}^c \Psi_i(\|\mathbf{x} - \mathbf{v}_i\|)} \quad (4.3)$$

im Gegensatz zu

$$f_{RBF,n}(\mathbf{x}) = \frac{\alpha_0 + \sum_{i=1}^c \alpha_i \Psi_i(\|\mathbf{x} - \mathbf{v}_i\|)}{1 + \sum_{i=1}^c \Psi_i(\|\mathbf{x} - \mathbf{v}_i\|)} \quad (4.4)$$

einen konstanten Offset, da bei Radialbasisfunktionen $\sum_{i=1}^c \Psi_i(\|\mathbf{x} - \mathbf{v}_i\|) \neq \text{const. } \forall \mathbf{x}$ gilt.

Als Basisfunktion finden z. B. thin-plate-splines, multiquadratische oder die inverse multiquadratische Funktionen und häufig Gaußfunktionen Verwendung (Poggio und Girosi 1989, Chen u. a. 1991, Xu u. a. 1993). Das Zentrum \mathbf{v} einer Basisfunktion soll im folgenden als Prototyp bezeichnet werden. Ein Netzwerk besteht aus Basisfunktionen vom gleichen Typ.

Ein für die Eigenschaften des Netzwerkes wichtiger Aspekt ist die Festlegung von Anzahl und Lage der Prototypen. Anfangs wurden alle N Datenpunkte $\mathbf{x}_k \in \mathbb{M}_x$ als Prototypen und das Netzwerk nur zur Interpolation in den Zwischenräumen genutzt. Bei großen Datenmengen ist dieses Vorgehen nicht praktikabel, weshalb dann nur ein Teil der $\mathbf{x}_k \in \mathbb{M}_x$ als Prototypen verwendet wird. Da eine willkürliche Auswahl zu suboptimalen Ergebnissen führt, nutzen Chen u. a. (1991) das orthogonale Least-Squares-Verfahren (OLS-Verfahren) zur Selektion günstiger $\mathbf{v}_i \in \mathbb{M}_x$. Dennoch erfordert eine hohe Modellgüte eine große Anzahl an Basisfunktionen. Beispielsweise stellen Chen u. a. (1991) drei Anwendungen vor: Bei einer Einschnittprädiktion für eine Zeitreihe ($N=396$) werden 55 RBF

(thin-plate-spline-Funktionen) verwendet. Bei der Identifikation zweier inverser (nicht-linearer) Filter werden einmal bei $N=400$ Datensätzen 12 RBF und das andere Mal bei $N=600$ Datensätzen 70 RBF (Gaußfunktionen) verwendet.

Das Zulassen einer beliebigen Platzierung der Prototypen im Gegensatz zu $\mathbf{v}_i \in \mathbb{M}_x$ kann die Güte der Netzwerke erhöhen. Xu u. a. (1993) schlagen das *rival penalized competitive learning* (RPCL) Verfahren vor, mit dem eine beliebige Platzierung der Prototypen erfolgen kann. Zudem können auch überzählige Prototypen erkannt werden, falls separierte Cluster vorliegen. Dies ist wichtig, da in einem solchen Fall die überzähligen Prototypen zu einem Bias bei den anderen Prototypen führen. Chen u. a. (1992) verwenden ein rekursives (scharfes) Clusterverfahren zur (beliebigen) Platzierung der Prototypen.

Zum Anlernen einer vorgegebenen Netzwerkstruktur können z. B. Gradientenverfahren (Poggio und Girosi 1989) eingesetzt werden (mit dem Problem lokaler Konvergenz). Sollen die Prototypen aus \mathbb{M}_x stammen, so kann das OLS-Verfahren angewendet werden (Chen u. a. 1991). Die Verbindung eines LS-Verfahren zur Identifikation der Gewichte α_i mit einem Clusterverfahren zur beliebigen Platzierung der Prototypen schlagen Chen u. a. (1992) vor.

4.2 Verwandtschaft von Fuzzy-Systemen und Basisfunktionsnetzwerken

Betrachtungen zur Verwandtschaft zwischen RBFN und Fuzzy-Systemen mit eindimensionalen Zugehörigkeitsfunktionen finden sich bei Horikawa u. a. (1992), Suh und Kim (1994) wie auch Jang (1993). Dabei werden auch Synonyme wie *adaptive Netzwerke* (Jang 1993) oder *Fuzzy Neuronale Netzwerke* (Horikawa u. a. 1992) geprägt. Im folgenden wird der Fall der zuvor vorgestellten mehrdimensionalen Zugehörigkeitsfunktionen behandelt. Einige Analogien finden sich bereits bei Jang und Sun (1993).

Das Übertragungsverhalten der in Abschnitt 3.2 diskutierten relationalen Fuzzy-Systeme läßt sich im statischen wie auch dynamischen Fall über

$$f_{FLR,n}(\mathbf{x}) = \frac{\sum_{i=1}^c g_i s_i \mu_i(\mathbf{x})}{\sum_{i=1}^c g_i \mu_i(\mathbf{x})} \quad (4.5)$$

beschreiben, wenn bei der Defuzzifizierung eine Normalisierung auf $\sum_{i=1}^c g_i \mu_i(\mathbf{x})$ vorgenommen wird (siehe Gl. 3.33a). Sonst folgt

$$f_{FLR}(\mathbf{x}) = \sum_{i=1}^c g_i s_i \mu_i(\mathbf{x}) \quad . \quad (4.6)$$

Falls die $\mu_i(\mathbf{x})$ ein Fuzzy-Informationssystem bilden, gilt $f_{FLR,n}(\mathbf{x}) = f_{FLR}(\mathbf{x}) \forall \mathbf{x}$. Für $\alpha_0 \equiv 0$ sind Gln. (4.1) und (4.6) bzw. (4.2) und (4.5) äquivalent, wenn $\psi_i \equiv g_i \mu_i$ und

$s_i \equiv \alpha_i$ gilt. Dabei bilden radialsymmetrische Funktionen $\mu_i(\mathbf{x})$ einen Sonderfall, der beispielsweise für μ_i^{ga} mit der euklidischen Norm vorliegt. Bei anderen Normen wie allgemeinen p -Normen ($p \neq 2$) oder bei anderen Funktionstypen wie μ^{cl} liegt keine Radialsymmetrie vor. Allgemeinere Basisfunktionen, die eine Funktion des *einen* zugehörigen Prototypen sowie \mathbf{x} sind ($\Psi_i = g(\mathbf{x}, \mathbf{v}_i)$), bezeichnen Poggio und Girosi (1989) als Hyperbasisfunktionen (die Radialbasisfunktionen stellen damit einen Sonderfall der Hyperbasisfunktionen dar). Somit läßt sich für Fuzzy-Systeme mit Zugehörigkeitsfunktion vom e-Typ und beliebiger Norm ein äquivalentes Hyperbasisfunktionsnetzwerk angeben. Zugehörigkeitsfunktionen vom μ^{cl} -Typ sind dagegen eine Funktion von *allen* c Prototypen sowie \mathbf{x} ($\mu_i^{cl} = h(\mathbf{x}, \mathbf{v}_1, \dots, \mathbf{v}_c)$). Für sie findet sich bisher kein äquivalentes Netzwerk in der Literatur. Ein entsprechendes Netzwerk soll wegen der Interdependenzen zwischen allen Basisfunktionen ein *Interbasisfunktionsnetzwerk* genannt werden. Da die Rückwirkungsfreiheit auf die anderen $(c-1)$ Interbasisfunktionen bei Modifikation einer einzelnen nicht wie bei Radial- und Hyperbasisfunktionen gegeben ist, liegt eine deutliche qualitative Abgrenzung vor.

Das Übertragungsverhalten aller im Abschnitt 3.1 besprochenen funktionalen Fuzzy-Systeme unterscheidet sich mathematisch deutlich von dem relationaler Systeme. Im Fall dynamischer Fuzzy-Systeme folgt bei gewichteter Mittelwertbildung bei der Bildung der Ausgangsgrößen im ersten Fall (Gl. 3.2)

$$f_{FLF,n}(\mathbf{x}) = \frac{\sum_{i=1}^c (y_{0,i} + \Theta_i^T(\mathbf{x} - \mathbf{x}_{0,i}))\mu_i(\mathbf{x})}{\sum_{i=1}^c \mu_i(\mathbf{x})} \quad . \quad (4.7)$$

Bei einer Ausgangsgrößenermittlung über gewichtete Superposition (Buckley 1993) folgt

$$f_{FLF}(\mathbf{x}) = \sum_{i=1}^c (y_{0,i} + \Theta_i^T(\mathbf{x} - \mathbf{x}_{0,i}))\mu_i(\mathbf{x}) \quad . \quad (4.8)$$

Falls die $\mu_i(\mathbf{x})$ ein Fuzzy-Informationssystem bilden, gilt $f_{FLF,n}(\mathbf{x}) = f_{FLF}(\mathbf{x}) \forall \mathbf{x}$. Statische Fuzzy-Systeme lassen sich beschreiben über

$$f_{FLF,n}^s(\mathbf{x}) = \frac{\sum_{i=1}^c (p_i(0) + \Theta_i^T \mathbf{x})\mu_i(\mathbf{x})}{\sum_{i=1}^c \mu_i(\mathbf{x})} \quad \text{bzw.} \quad (4.9)$$

$$f_{FLF}^s(\mathbf{x}) = \sum_{i=1}^c (p_i(0) + \Theta_i^T \mathbf{x})\mu_i(\mathbf{x}) \quad . \quad (4.10)$$

Selbst wenn in den Gln. (4.7) bis (4.10) die Basisfunktionen μ_i radialsymmetrisch angesetzt werden, so sind die Summanden wegen des von \mathbf{x} abhängigen Gewichtes ($y_{0,i} + \Theta_i^T(\mathbf{x} - \mathbf{x}_{0,i})$) bzw. ($p_i(0) + \Theta_i^T \mathbf{x}$) nicht radialsymmetrisch. Dies gilt auch für den im

Abschnitt 3.1 beschriebenen zweiten Fall. Es läßt sich also kein äquivalentes Radial-, wohl aber ein Hyperbasisfunktionsnetzwerk angeben. Diese Aussage ist auch für den Fall nichtradialsymmetrischer Basisfunktionen, die von einem Prototyp abhängen, gültig. Basisfunktionen, die von mehreren Prototypen abhängen, führen zum oben definierten Interbasisfunktionsnetzwerk.

Geht man von der mathematisch formalistischen auf eine qualitative Betrachtungsweise über, so zeigen sich Parallelen zwischen den vorgestellten Fuzzy-Systemen und den (mehrschichtigen) Netzwerken. Beide bilden den m -dimensionalen Eingangsgrößenraum *nicht-linear* in einen c -dimensionalen Raum ab, in dem die Ausgangsgröße durch eine *lineare* Abbildung auf einen eindimensionalen Raum bestimmt wird. Die nichtlineare Abhängigkeit wird über Basisfunktionen determiniert, deren Wahl damit entscheidend die Güte des Gesamtsystems beeinflußt. Beide Ansätzen besitzen den Vorteil, daß durch mehrdimensionale Basisfunktionen die Dimensionalität des Eingangsgrößenraumes von der Anzahl der Basisfunktionen entkoppelt wird. Damit erfolgt auch eine Entkopplung der Anzahl der Regeln eines Fuzzy-Systems von der Dimensionalität des Eingangsgrößenraumes.

5 Nutzung analytischer Modellbeschreibungen bei der Identifikation

In diesem Abschnitt werden die Identifikation von Hybrid-Modellen, die aus der Serienschaltung eines nichtlinearen statischen Fuzzy-Teilmodells und eines linearen dynamischen Teilmodells aufgebaut sind, sowie auch die Identifikation dynamischer Fuzzy-Modelle behandelt.

Die Aufgabe der Partitionierung besteht dabei darin, den Eingangsgrößenraum in c (unscharfe) Unterräume (Partitionen) aufzuteilen, in denen sich das System hinreichend genau durch das Schlußfolgerungspolynom beschreiben läßt. Die gewichtete Überlagerung der Einzelbeschreibungen liefert das Gesamtmodell. Deshalb unterscheidet sich diese Modellierungsform deutlich von den mathematisch ähnlichen *Fuzzy- c -Regressionsmodellen FCRM* (Hathaway und Bezdek 1993). Diese ermitteln c Regressionen, die jeweils im gesamten Eingangsgrößenraum gelten.

Bei der Identifikation von Fuzzy-Modellen ist die Zugehörigkeit der Datensätze zu den Partitionen bzw. Regeln zu berücksichtigen. Abgesehen vom Verfahren der gewichteten kleinsten Fehlerquadrate bieten konventionelle Identifikationsalgorithmen im allgemeinen keine Möglichkeit, eine Gewichtung der Daten (gemäß ihrer Zugehörigkeit) zu integrieren. In den nächsten beiden Teilabschnitten wird deshalb ein Formalismus angegeben, mit dem durch eine Abbildung der Meßdaten eine Meßmatrix erzeugt wird, in die die Zugehörigkeiten inkorporiert sind. Dadurch können auch andere Identifikationsalgorithmen als das genannte Verfahren der gewichteten kleinsten Fehlerquadrate für die Schätzung der Konklusionsparameter angewendet werden.

Eine Vorabprüfung, wie stark nichtlinear sich ein System verhält und ob eine statische oder dynamische Nichtlinearität vorliegt, ermöglichen Nichtlinearitätstests (Reuter 1995). Als günstig hat sich dabei die parallele Anwendung der Methode der nichtlinearen Auto- und Kreuzkorrelation sowie der Frequenzmethode erwiesen (Reuter 1992, Reuter 1995). Damit kann eine Festlegung der Modellstruktur (Hybrid-Modell bzw. dynamisches Fuzzy-Modell) erfolgen.

Im folgenden werden zuerst analytische Beschreibungen für Hybrid-Modelle sowie für dynamische und statische Fuzzy-Modelle gegeben. Anschließend wird ein zweistufiges Identifikationsverfahren vorgestellt.

5.1 Wiener- und Hammerstein-Modelle mit statischer Fuzzy-Nichtlinearität

Aus der konventionellen Modellbildung (Ljung 1987, Isermann 1988b, Wernstedt 1989) sind einfache Wiener- bzw. Hammerstein-Modelle bekannt, die aus der Serienschaltung eines linearen dynamischen und eines nichtlinearen statischen Teilsystems bzw. umgekehrt

bestehen. Das statische Teilsystem besteht aus einem Polynom in der Eingangsgröße. In einer Verallgemeinerung dieses Ansatzes soll das statische Teilsystem durch ein nichtlineares statisches Fuzzy-Modell ersetzt werden. Damit entsteht ein *einfaches Wiener- bzw. Hammerstein-Modell mit Fuzzy-Nichtlinearität*.

Zur Identifikation von einfachen Wiener- und Hammersteinmodellen können direkte und mehrstufige Strategien angewendet werden. Mehrstufige Strategien beschreibt Wernstedt (1989). Im folgenden wird die direkte Identifikation eines solchen hybriden Modells mit einer (statischen) Fuzzy-Nichtlinearität an Stelle des üblichen Polynomansatzes besprochen. Die wenigsten Einschränkungen erfordert dabei eine Hammerstein-Modellstruktur, bei der zuerst das nichtlineare Teilmodell (Fuzzy-Modell) identifiziert wird.

Hammerstein-Modellstruktur

Betrachtet wird eine Modellstruktur gemäß Bild 5.1.

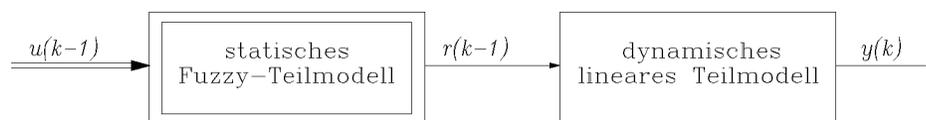


Bild 5.1: Einfaches Hammerstein-Modell mit statischer Fuzzy-Nichtlinearität

Funktionales Fuzzy-Modell

Das lineare Teilmodell läßt sich durch eine rationale Übertragungsfunktion bzw. durch eine Differenzgleichung

$$y(k) = - \sum_{i=1}^{n_a} a_i y(k-i) + \sum_{i=1}^{n_b} b_i r(k-i) \quad (5.1)$$

(ARX-Modell), das Fuzzy-Teilmodell mit Gl. (3.25a) über

$$r(k) = \sum_{j=1}^c \phi_j(k) [p_j(0); \Theta_j^T] \mathbf{u}_S(k) \quad (5.2)$$

mit

$$\mathbf{u}_S^T(k) = [1, u_1(k), \dots, u_m(k)] \quad (5.3)$$

beschreiben. Aus den Gln. (5.1) und (5.2) folgt

$$y(k) = - \sum_{i=1}^{n_a} a_i y(k-i) + \sum_{i=1}^{n_b} b_i \sum_{j=1}^c \phi_j(k-i) [p_j(0); \Theta_j^T] \mathbf{u}_S(k-i) \quad (5.4a)$$

$$= \Theta_F \boldsymbol{\varphi}_F(k) \quad (5.4b)$$

mit

$$\begin{aligned} \Theta_F &= [-a_1, \dots, -a_{n_a}, b_1[p_1(0); \Theta_1^T], \dots, b_1[p_c(0); \Theta_c^T], \dots, \\ &\quad b_{n_b}[p_1(0); \Theta_1^T], \dots, b_{n_b}[p_c(0); \Theta_c^T]] \end{aligned} \quad (5.5)$$

$$= [-a_1, \dots, -a_{n_a}, b_{1,1,0}, \dots, b_{1,1,m}, \dots, b_{1,c,0}, \dots, b_{1,c,m}, \dots, \\ b_{n_b,1,0}, \dots, b_{n_b,1,m}, \dots, b_{n_b,c,0}, \dots, b_{n_b,c,m}] \quad ,$$

$$\begin{aligned} \varphi_F^T(k) &= [y(k-1), \dots, y(k-n_a), \mathbf{u}_S^T(k-1)\phi_1(k-1), \dots, \mathbf{u}_S^T(k-1)\phi_c(k-1), \\ &\quad \dots, \mathbf{u}_S^T(k-n_b)\phi_1(k-n_b), \dots, \mathbf{u}_S^T(k-n_b)\phi_c(k-n_b)] \quad , \end{aligned} \quad (5.6)$$

$$b_{i,j,f} = b_i p_j(f) \quad \text{und} \quad (5.7)$$

$$\Theta_j^T = [p_j(1), \dots, p_j(m)] \quad . \quad (5.8)$$

Die Parameter b_i und $p_j(f)$ können aus den geschätzten Parametern $b_{i,j,f}$ bestimmt werden, wenn die Verstärkung des linearen Teilsystems zu Eins gesetzt wird (Wernstedt 1989):

$$K_L = \frac{\sum_{i=1}^{n_b} b_i}{1 + \sum_{i=1}^{n_a} a_i} \equiv 1 \quad . \quad (5.9)$$

Dann folgt

$$p_j(f) = \frac{\sum_{i=1}^{n_b} b_{i,j,f}}{1 + \sum_{i=1}^{n_a} a_i} \quad \forall f \in \{1, \dots, m\}; j \in \mathbb{I} \quad \text{und} \quad (5.10)$$

$$b_i = \frac{b_{i,j,f}}{p_j(f)} \quad \forall i \in \{1, \dots, n_b\} \quad . \quad (5.11)$$

Relationales Fuzzy-Modell

Das statische relationale Fuzzy-Modell läßt sich durch Gl. (3.35)

$$r(k) = r(\mathbf{u}(k)) = \sum_{i=1}^c s_i \phi_i(k) \quad (5.12)$$

beschreiben, womit

$$\begin{aligned} y(k) &= - \sum_{i=1}^{n_a} a_i y(k-i) + \sum_{j=1}^{n_b} b_j \sum_{i=1}^c s_i \phi_i(k-j) \\ &= \Theta_R \varphi_R(k) \end{aligned} \quad (5.13)$$

mit

$$\begin{aligned}\Theta_R^T &= [-a_1, \dots, -a_{n_a}, b_1 s_1, \dots, b_1 s_c, \dots, b_{n_b} s_1, \dots, b_{n_b} s_c] \\ &= [-a_1, \dots, -a_{n_a}, b_{11}, \dots, b_{1c}, \dots, b_{n_b 1}, \dots, b_{n_b c}] \\ b_{ij} &= b_i s_j \quad \text{und}\end{aligned}\tag{5.14}$$

$$\begin{aligned}\varphi_R(k) &= [y(k-1), \dots, y(k-n_a), \phi_1(k-1), \dots, \phi_c(k-1), \\ &\quad \dots, \phi_1(k-n_b), \dots, \phi_c(k-n_b)]\end{aligned}$$

folgt. Die Parameter b_i und s_i können aus den geschätzten Parametern b_{ij} bestimmt werden, wenn die Verstärkung des linearen Teilsystems zu Eins gesetzt wird (siehe Gl. (5.9)). Dann folgt

$$\begin{aligned}s_j &= \frac{\sum_{i=1}^{n_b} b_{ij}}{1 + \sum_{i=1}^{n_a} a_i} \quad \forall j \in \mathbb{I} \quad \text{und} \\ b_i &= \frac{b_{ij}}{s_j} \quad \forall i \in \{1, \dots, n_b\} \quad .\end{aligned}\tag{5.15}$$

Wiener-Modell

Wie beim Hammerstein-Modell soll ein MISO-Fuzzy-Modell eingesetzt werden (Bild 5.2).

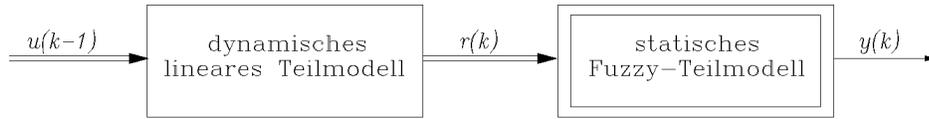


Bild 5.2: Einfaches Wiener-Modell mit statischem Fuzzy-Teilmodell

Das Fuzzy-Modell läßt sich wie zuvor in Gl. (5.2) durch

$$y(k) = \sum_{j=1}^c \phi_j(k) [p_j(0); \Theta_j^T] \mathbf{r}_S(k)\tag{5.16}$$

mit

$$\mathbf{r}_S^T(k) = [1; \mathbf{r}^T(k)]\tag{5.17}$$

darstellen. Da im allgemeinen die Parameter des linearen Teilsystems unbekannt sind, kann durch Clusterung keine Initialpartitionierung bestimmt werden. Da diese damit willkürlich gewählt werden muß, sind die Startwerte schlechter, was sich negativ auf die Konvergenzeigenschaften der Optimierungsverfahren auswirkt.

Regressionsverfahren können hier nicht eingesetzt werden: Bei Entkoppelbarkeit des linearen Systems (Kriterien und Transformationsvorschriften finden sich bei Schwarz (1967) und Schwarz (1991)) gilt

$$r_j(k) = - \sum_{i=1}^{n_a} a_{i,j} r_j(k-i) + \sum_{i=1}^{n_b} b_{i,j} u_j(k-i) \quad \forall j = 1, \dots, m \quad . \quad (5.18)$$

Der aus der Kombination der Gln. (5.16) und (5.18) entstehende Ausdruck läßt sich nicht als Skalarprodukt eines Parameter- und eines Meßvektors darstellen, weil die Ausgangsgröße des linearen Teilmodells die Zwischen- und nicht wie beim Hammerstein-Modell die Ausgangsgröße ist. Dies gilt auch für den Fall eines linearen SISO-Systems. Wegen dieser Problematik wird das einfache Wiener-Modell mit statischer Fuzzy-Nichtlinearität nicht weiterverfolgt.

5.2 Identifikation dynamischer Fuzzy-Modelle

Betrachtet werden dynamische Fuzzy-Modelle (Bild 5.3). Im Fall von funktionalen Fuzzy-



Bild 5.3: Dynamisches Fuzzy-Modell

Modellen gilt Gl. (3.18b) (das diskrete Zeitargument wird um eine Periode verkleinert)

$$y(k) = y(\mathbf{x}(k-1)) = \sum_{i=1}^c \phi_i(\mathbf{x}(k-1)) [c_i; \Theta_i^T] \begin{bmatrix} 1 \\ \mathbf{x}(k-1) \end{bmatrix} \quad , \quad (5.19)$$

im Fall von relationalen Fuzzy-Modellen (siehe Gl. (3.35))

$$y(k) = y(\mathbf{x}(k-1)) = \sum_{i=1}^c s_i \phi_i(\mathbf{x}(k)) \quad . \quad (5.20)$$

Bei zeitdiskreten linearen Ein-/Ausgangsmodellen mit PT_{n_a} -Verhalten sind ein- bis n_a -fach verzögerte Ausgangsgrößenrückführungen notwendig:

$$y(k) = - \sum_{i=1}^{n_a} a_i y(k-i) + \sum_{j=1}^{n_b} b_j u(k-j) \quad . \quad (5.21)$$

Die im Abschnitt 4.1 vorgestellten Modelle sind deshalb durch eine nichtlineare Verstärkung und eine einheitliche PT_{n_a} -Dynamik charakterisiert.

Bei dynamischen (funktionalen oder relationalen) Fuzzy-Modellen mit ein- bis n_a -fach verzögerter Ausgangsgrößenrückführung hat jede Partition eine unabhängig einstellbare PT_{n_a} -Charakteristik. Das Gesamtmodell ist deshalb durch nichtlineare Verstärkung und

Dynamik gekennzeichnet. Somit stellen die Hammerstein- und Wiener-Modelle einen einfachen Sonderfall der dynamischen Fuzzy-Modelle dar. Da sie aber eine deutliche Reduktion der Modellparameteranzahl bewirken, stellen sie eine interessante Alternative dar, falls die lineare Dynamik die Dynamik des zu modellierenden Systems ausreichend approximiert.

5.3 Identifikationsverfahren

Wenn die Partitionierung (d. h. die μ_i) vorgegeben und nicht zu identifizieren ist, so sind die Gln. (5.4), (5.13), (5.19) und (5.20) linear in den Parametern. Damit lassen sich die Parameter des linearen Teilmodells und der Konklusionsfunktionen z. B. über Regression, das Hilfsvariablenverfahren, das Verfahren der verallgemeinerten Regression oder durch die Maximum-Likelihood-Schätzung ermitteln (Wernstedt 1989). Sollen auch die die Partitionierung bestimmenden Parameter identifiziert werden, so kann z. B. das folgende zweistufige Verfahren angewendet werden.

1. Stufe

Durch nichtüberwachtes Lernen wird über Clusterverfahren wie Fuzzy- c -Means- oder Gustafson- und Kessel-Algorithmus (Bezdek 1981) eine Fuzzy-Partitionierung über Zugehörigkeitsfunktionen vom μ^{cl} -Typ vorgenommen. Liegen keine A-priori-Erkenntnisse über eine geeignete Partitionsform vor, so ist es vorteilhaft, die Mahalanobis-Abstandsnorm zu verwenden (Totz 1995). Diese legt die Clusterform nach statistischen Eigenschaften der Meßdaten fest. Die Festlegung von Randbedingungen und Parametern bei der Partitionierung und bei der Anwendung von Clusterverfahren wird von Kroll (1994b) diskutiert. Für Zugehörigkeitsfunktionen vom μ^{ga} -Typ können über Clusterung generierte Prototypen in erster Näherung direkt übernommen werden. Der die Flanken bestimmende Parameter $\tilde{\sigma}_i$ kann aus μ_i^{cl} abgeschätzt werden, und die $\mu_{i,0}$ werden zu Eins initialisiert. Durch überwachtes Lernen mit dem LS- oder RLS-Verfahren erfolgt die Bestimmung der Konklusionsparameter.

2. Stufe

1. Alternative

Die in der 1. Stufe bestimmten Parameter werden durch überwachtes Lernen mit einem Gradientenverfahren optimiert. Da die Zugehörigkeitsfunktionen stetig differenzierbar sind, kann z. B. bei einer quadratischen Kostenfunktion

$$J = \frac{1}{2} \sum_{k=1}^N (y(k) - \hat{y}(k))^2 \quad (5.22)$$

der Gradient analytisch (geschlossen) berechnet werden.

2. Alternative

In jedem Iterationsschritt werden zuerst durch ein Gradientenverfahren die Parameter der Partitionierung (das Schätzproblem ist nichtlinear in den Parametern) und anschließend über das LS-Verfahren die Parameter der Konklusionsfunktionen und die des linearen Teilsystems (dieses Schätzproblem ist linear in den Parametern) beim Hammerstein-Modell identifiziert.

Die 1. Stufe läßt sich in der beschriebenen Form auch bei nicht stetig differenzierbaren Zugehörigkeitsfunktionen durchführen (Kroll 1994a). Dann sind allerdings in der 2. Stufe Fallunterscheidungen durchzuführen oder es ist ein numerisches Gradientenverfahren anzuwenden. Die Idee dieser stufenweisen Identifikation liegt darin, mit Verfahren zu starten, die wenig Vorwissen erfordern, aber auch nur einen begrenzten Wissensgewinn hervorbringen. In den folgenden Stufen (hier eine) liegt dann größeres Vorwissen vor, und geeignete Verfahren können einen dementsprechend größeren Wissensgewinn liefern (Reuter 1995).

Bereits in der 1. Stufe kann ein Modell hoher Güte bei geringer Rechenzeit erzeugt werden (Lohmann 1994, Kroll 1994a, Totz 1995). Damit kann die Forderung nach guten Startwerten, die für die Konvergenz der nichtlinearen Optimierungsalgorithmen in der 2. Stufe notwendig sind (Reuter 1993), i. allg. erfüllt werden. Deshalb reicht für strukturelle Untersuchungen, die Identifikationen erfordern, i. allg. bereits die Ausführung der 1. Stufe aus. Dann besitzen die Parameter aber wegen des LS-Verfahrens in der Regel keine Bias-Freiheit (Isermann 1988a, Reuter 1995). Außerdem sind nur die Parameter des Konklusionspolynoms optimal im Sinne der kleinsten Fehlerquadrate bei gegebener Partitionierung, nicht aber die Partitionierung selbst. Deshalb sollte in der 2. Stufe eine Nachoptimierung aller Parameter erfolgen, die aber nur einmal durchzuführen ist. Dadurch bleibt der gesamte Rechenaufwand, selbst bei iterativen Verfahren in der 2. Stufe, vertretbar.

Während zeitinvariante Systeme sowohl on-line- als auch off-line-Verarbeitung (z. B. über Stapelverarbeitung („Batch“-Betrieb)) gestatten, sind bei zeitvarianten Systemen on-line-fähige Verfahren anzuwenden. Dies kann auch im 1. Schritt z. B. durch eine einmalige Partitionierung über Clusterverfahren und anschließend datenweise Nachkorrektur der Parameter der Konklusionsfunktionen über rekursive Verfahren erfolgen. In der 2. Stufe kann ebenfalls durch Übergang auf rekursive Gradientenverfahren wie das Verfahren der stochastischen Approximation (Isermann 1988b, Küpper 1994) on-line-Fähigkeit erreicht werden.

Eine Vorschrift zur Bestimmung der $\tilde{\sigma}_i$ kann darin liegen, daß sich benachbarte Basisfunktionen maximal bei einem Funktionswert $a \in (0; 1)$ schneiden. Dies kann durch folgendes Verfahren erreicht werden:

Initialisierungsalgorithmus für $\tilde{\sigma}_i$

Ist \mathbf{v}_j der nächste Prototyp zu \mathbf{v}_i im Sinne der Metrik $\|\cdot\|$ und wurde $\tilde{\sigma}_j$ noch nicht bestimmt, so gilt für die Wahl von $\tilde{\sigma}_i = \tilde{\sigma}_j$

$$\mu_i(\mathbf{v}_i + 0,5(\mathbf{v}_j - \mathbf{v}_i)) = \mu_{i,0} \exp(\|0,5(\mathbf{v}_j - \mathbf{v}_i)\|^2 / \tilde{\sigma}_i^2) \stackrel{!}{=} a \quad (5.23)$$

und damit

$$\tilde{\sigma}_i = \frac{0,5\|\mathbf{v}_j - \mathbf{v}_i\|}{\sqrt{\ln(a) - \ln(\mu_{i,0})}} \quad (5.24)$$

Bei der euklidischen Metrik folgt damit:

$$\tilde{\sigma}_i = \frac{1}{2} \sqrt{\frac{\sum_{f=1}^m |\mathbf{v}_i(f) - \mathbf{v}_j(f)|^2}{\ln(a) - \ln(\mu_{i,0})}} \quad (5.25)$$

Wurde $\tilde{\sigma}_j$ bereits bestimmt, so gilt für die Wahl von $\tilde{\sigma}_i$:

Bestimme das $\gamma \in (0 ; 1)$ mit

$$\mu_j(\mathbf{v}_j + \gamma(\mathbf{v}_i - \mathbf{v}_j)) = \mu_{j,0} \exp(\|\gamma(\mathbf{v}_i - \mathbf{v}_j)\|^2 / \tilde{\sigma}_j^2) \stackrel{!}{=} a \quad (5.26)$$

zu

$$\gamma = \frac{\tilde{\sigma}_j \sqrt{\ln(a) - \ln(\mu_{j,0})}}{\|\mathbf{v}_i - \mathbf{v}_j\|} \quad (5.27)$$

Wähle $\tilde{\sigma}_i$ so, daß

$$\mu_i(\mathbf{v}_j + \gamma(\mathbf{v}_i - \mathbf{v}_j)) = \mu_{i,0} \exp(\|(1 - \gamma)(\mathbf{v}_i - \mathbf{v}_j)\|^2 / \tilde{\sigma}_i^2) \stackrel{!}{=} a \quad (5.28)$$

ist:

$$\tilde{\sigma}_i = \frac{|1 - \gamma| \|\mathbf{v}_i - \mathbf{v}_j\|}{\sqrt{\ln(a) - \ln(\mu_{i,0})}} \quad (5.29)$$

Das Clusterverfahren sollte entsprechend der vorliegenden Datenstrukturen gewählt werden. Liegt keine Information über die geometrischen Formen vor, besteht eine einfache Lösung im Rückgriff auf den unproblematischen Fuzzy- c -Means-Algorithmus. Bei Verwendung der Mahalanobis-Norm (Bezdek 1981) können in einem automatisierten Vorgehen gute Ergebnisse erzielt werden (Totz 1995), da die Metrik statistische Eigenschaften der Meßdaten berücksichtigt. Mit einer p -Norm lassen sich beispielsweise hexaedrische ($p \rightarrow \infty$), oktaedrische ($p = 1$) oder sphärische ($p = 2$) Clusterformen sowie Übergangsformen zwischen ihnen realisieren. Zur Clusterung können auch Genetische Algorithmen eingesetzt werden. Falls sie allerdings nur herkömmliche Clusterverfahren ersetzen, so führen sie zu keiner Verbesserung der Ergebnisse (Schulte 1994), verursachen aber deutlich größeren Implementierungsaufwand und längere Rechenzeiten. Deshalb werden solche Verfahren an dieser Stelle nicht weiter vertieft.

6 Strukturelektion, Güte und Modellvalidierung

Einfache Gütekriterien sind beispielsweise

$$J_p = \frac{1}{N} \sum_{k=1}^N |y(k) - \hat{y}(k)|^p, \quad (6.1)$$

das für $p = 2$ den mittleren quadratischen Fehler angibt, oder

$$J_p^* = \sqrt[p]{\frac{1}{N} \sum_{k=1}^N |y(k) - \hat{y}(k)|^p}. \quad (6.2)$$

Dabei muß noch zwischen einschrittprädierten und rekursiv prädierten Werten \hat{y} unterschieden werden. Die Anwendung von J_p oder J_p^* zur Strukturelektion ist problematisch, da sie im allgemeinen eine monotone Erhöhung der Modellparameteranzahl hervorruft. Wird Gl. (6.1) oder (6.2) als Zielfunktion einer Optimierung eingesetzt, kann eine Überanpassung (*overfit*) des Modells an das Testsignal erfolgen. Dies verursacht z. B. bei verrauschten Testsignalen einen Verlust an im Modell abgebildeter Systeminformation.

Strukturelektion

Deshalb werden zur Strukturelektion bei der konventionellen Modellidentifikation Kriterien wie der *final prediction error FPE* (Akaike 1969) oder das *information theoretic criterion ITC* (Ljung 1987) verwendet. Reuter (1995) stellt einige weitere Kriterien vor. Bei der Fuzzy-Modellierung werden häufig *Kreuzvalidierungskriterien* zur Strukturelektion eingesetzt. Während FPE und ITC eine Modellbewertung anhand der zur Identifikation verwendeten Daten vornehmen, greifen diese zusätzlich auf unbenutzte Daten zurück. Zu den letzteren zählen z. B. das *Unbiasedness criterion UC* (Sugeno und Kang 1988).

$$J_{UC} = \sqrt{\sum_{k=1}^{N_A} (\hat{y}_k^{AB} - \hat{y}_k^{AA})^2 + \sum_{k=1}^{N_B} (\hat{y}_k^{BA} - \hat{y}_k^{BB})^2} \quad (6.3)$$

und darauf aufbauend (Horikawa u. a. 1992)

$$J_C = \sqrt{\sum_{k=1}^{N_A} (y_k^A - \hat{y}_k^{AA})^2 + \sum_{k=1}^{N_B} (y_k^B - \hat{y}_k^{BB})^2} + J_{UC}, \quad (6.4)$$

das *Regularity Criterion RC* (Sugeno und Yasukawa 1993)

$$J_{RC} = \frac{1}{2} \left[\frac{1}{N_A} \sum_{k=1}^{N_A} (\hat{y}_k^{AB} - y_k^A)^2 + \frac{1}{N_B} \sum_{k=1}^{N_B} (\hat{y}_k^{BA} - y_k^B)^2 \right] \quad (6.5)$$

oder der *basisdatenunabhängige mittlere quadratische Fehler* (Angenendt 1993)

$$J_{BD} = \frac{1}{f^2 d} \sum_{j=1}^f \sum_{i=1}^f \sum_{k=1}^d (\hat{y}_k^{ij} - y_k^i)^2. \quad (6.6)$$

Bei J_{UC} , J_C und J_{RC} werden die N Meßdatensätze in zwei Sequenzen \mathcal{S}_A und \mathcal{S}_B aufgeteilt mit $\mathcal{A}(A) + \mathcal{A}(B) = N_A + N_B = N$. Das mit \mathcal{S}_A identifizierte Modell liefert bei Auswertung mit \mathcal{S}_A die Werte \hat{y}^{AA} , bei Auswertung mit \mathcal{S}_B die Werte \hat{y}^{BA} . y^A sind die Ausgangsgrößenmeßdaten der Sequenz \mathcal{S}_A . Entsprechendes gilt für das mit \mathcal{S}_B identifizierte Modell. Bei J_{BD} werden f Modelle wie folgt identifiziert:

Sei beim Vorliegen von N Meßdatensätzen d das größte Vielfache von f , so daß $fd \leq N$ gilt. Die ersten $\hat{N} := fd$ der gegebenen Meßdatensätze werden in f Sequenzen $S_1 := \{1, \dots, d\}$, $S_2 := \{d + 1, \dots, 2d\}$ bis $S_f := \{\hat{N} - d + 1, \dots, \hat{N}\}$ zerlegt. Für jede Sequenz S_i wird jeweils ein Modell identifiziert. Sei \hat{y}_k^{ij} die Prädiktion des j -ten Modells (das mit S_j identifiziert wurde) bei Auswertung mit dem k -ten Datensatz der Sequenz S_i und y_k^i die gemessene Ausgangsgröße. Im Fall einer Aufteilung in 2 Sequenzen \mathcal{S}_A und \mathcal{S}_B (wie bei J_{UC} , J_C und J_{RC}) folgt $f = 2$ und $d = \hat{N}/2 =: N_A = N_B$ sowie

$$J_{BD} = \frac{1}{2\hat{N}} \left[\sum_{k=1}^{N_A} (\hat{y}_k^{AB} - y_k^A)^2 + (\hat{y}_k^{AA} - y_k^A)^2 + \sum_{k=1}^{N_B} (\hat{y}_k^{BA} - y_k^B)^2 + \sum_{k=1}^{N_B} (\hat{y}_k^{BB} - y_k^B)^2 \right] \quad (6.7)$$

mit der oben angegebenen Nomenklatur. Eine einfache und sinnvolle Erweiterung liegt in der Zulassung verschieden langer Sequenzen \mathcal{S}_A und \mathcal{S}_B .

Bei der Optimierung oder beim Vergleich verschiedener Modelle interessieren nur relative Werte der Güte. Damit sind die Radizierung bei J_{UC} und der Vorfaktor 0,5 bei J_{RC} irrelevant. Bei J_{UC} erfolgt keine Normierung auf die Datenanzahl, was Vergleiche bei Auswertung verschieden langer Sequenzen erschwert. Außerdem gehen die gemessenen Ausgangsgrößen y_k^A und y_k^B nicht in J_{UB} ein. Weil nur Prädiktionen bewertet werden, handelt es sich um ein *sekundäres Kriterium*. Bei J_{RC} haben für $N_A \neq N_B$ die Daten aus \mathcal{S}_B jeweils ein anderes Gewicht als die aus \mathcal{S}_A . Da im allgemeinen kein A-priori-Wissen vorliegt, auf Grund dessen die Daten einer Sequenz stärker/schwächer gewichtet werden sollten als die der anderen, sollten alle Daten gleich gewichtet werden. Desweiteren bewertet J_{RC} in keiner Weise die Prädiktion mit den zur Identifikation eingesetzten Daten. Sowohl J_C als auch J_{BD} liegt die Idee zugrunde, die Anteile bzgl. Kreuzvalidierung und Systemidentifikation separat und (willkürlich) gleichgewichtet zu berücksichtigen. Außerdem besitzt jedes Datum das gleiche Gewicht. Die Normierung bei J_{BD} gewährleistet bei verschieden großen Datenmengen eine einfache Vergleichbarkeit.

Der Nachteil der Kreuzvalidierungskriterien liegt in der deutlich erhöhten Anzahl an erforderlichen Daten. Damit sie eine signifikante Aussage liefern, müssen bei J_{UC} , J_C und J_{RC} zwei vollständige Identifikationsdatensätze vorliegen, bei J_{BD} sogar f Sätze. Liegen nicht genug Daten vor, so sollte mit reduzierten Kriterien gearbeitet werden, die einen Datensatz nur zur Auswertung verwenden. Dieser kann deshalb im Umfang reduziert werden (Totz 1995).

Modellvalidierung

Neben den bereits angesprochenen Kreuzvalidierungskriterien können zur Modellvalidierung auch Korrelationsverfahren eingesetzt werden. In der Literatur (Ljung 1987) wird angenommen, daß die Sequenz der Abweichungen

$$\varepsilon_F(k) = y(k) - \hat{y}(k) \quad , \quad k = 1, \dots, N \quad (6.8)$$

zwischen Messung und Prädiktion (Bild 6.1) bei gelungener Identifikation ein weißes Rau-

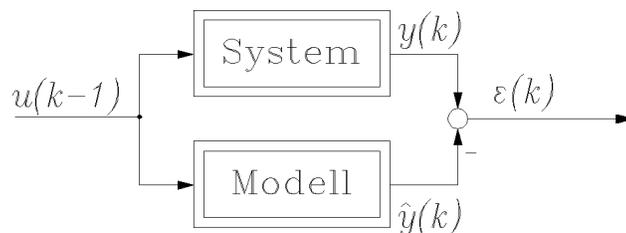


Bild 6.1: System und identifiziertes Modell

schen ist. Das heißt, die Autokorrelationsfunktion $\phi_{\varepsilon\varepsilon}(\tau)$ liefert für $\tau \neq 0$ den Wert Null. Außerdem sollte $\varepsilon_F(k)$ nicht mit alten Werten der Stellgröße $u(k - \tau)$, $\tau > 0$, korreliert sein, da sonst nicht alle Wirkungen von $u(k)$ identifiziert wurden. Dagegen zeigt eine Korrelation von $\varepsilon_F(k)$ und $u(k - \tau)$, $\tau < 0$, an, daß die Ausgangsgröße verzögert als Eingangsgröße verwendet wird (Ljung 1987). Dies wird bei dynamischen Fuzzy-Systemen strukturell zur Erzielung der dynamischen Eigenschaften vorgegeben und führt deshalb zu keiner Negativbewertung.

7 Zusammenfassung und Ausblick

Der vorliegende Bericht behandelt die analytische Beschreibung von Fuzzy-Systemen, deren Prämissen mehrdimensionale Zugehörigkeitsfunktionen mit unbeschränktem Träger verwenden. Diesbezüglich wurden insbesondere Zugehörigkeitsfunktionen vom e-Typ (μ^{ga}) und aus Fuzzy-Clusteralgorithmen abgeleitete Zugehörigkeitsfunktionen (μ^{cl}) betrachtet. Für Fuzzy-Systeme mit solchen stetigen und stetig differenzierbaren Zugehörigkeitsfunktionen wurde eine *einheitliche analytische* Beschreibung angegeben. Die Verwandtschaft und teilweise sogar Äquivalenz zu speziellen künstlichen Neuronalen Netzen wie Radial- oder Hyperbasisfunktionsnetzwerken wurde besprochen. Als analoges Netzwerk zu Fuzzy-Systemen mit Zugehörigkeitsfunktionen vom μ^{cl} -Typ wurde ein Interbasisfunktionsnetzwerk eingeführt.

Bei der Identifikation von Fuzzy-Modellen ist die Zugehörigkeit der Datensätze zu den Partitionen bzw. Regeln zu beachten. Dies erschwerte bisher die Anwendbarkeit konventioneller Identifikationsverfahren. Eine Ausnahme bildet das Verfahren der gewichteten kleinsten Fehlerquadrate, bei dem die Zugehörigkeit eines Datums direkt als sein Gewicht benutzt werden kann.

Aus der analytischen Darstellung der Fuzzy-Systeme erfolgte die Ableitung einer Transformation der Meßwerte, deren Ergebnis ungewichtet zur Identifikation verwendbar ist. Dies erlaubt die Anwendung einer Vielzahl bekannter Identifikationsverfahren. Ein zweistufiges Identifikationsverfahren für dynamische und statische Fuzzy-Modelle sowie für Hammerstein-Modelle mit statischer Fuzzy-Nichtlinearität wurde angegeben: Ein Fuzzy-Clusterverfahren ermittelt in der 1. Stufe eine Partitionierung und anschließend erfolgt eine Identifikation der Parameter der Konklusionsfunktionen durch das ungewichtete Least-Squares-Verfahren. Ein Gradientenverfahren optimiert in der 2. Stufe die Parameter. Für die Modellvalidierung wurden einige Kreuzvalidierungskriterien vorgestellt und diskutiert.

Zur Zeit erfolgt eine Erprobung des vorgestellten Algorithmus. Schwerpunktmäßig werden funktionale Fuzzy-Modelle mit Zugehörigkeitsfunktionen vom μ^{cl} -Typ untersucht. Im Gegensatz zu Zugehörigkeitsfunktionen vom e-Typ kann der Graph von μ^{cl} sehr verschiedene Formen annehmen. Damit ist zu erwarten, daß deutlich weniger Regeln und damit Modellparameter für die gleiche Modellgüte erforderlich sind. Weiterhin sind noch Erfahrungen mit der Wahl und Parametrierung des Gradientenverfahrens in der 2. Identifikationsstufe zu sammeln.

Ein ganz anderes Anwendungsfeld bietet die Fuzzy-Regelung. Durch die auch auf Fuzzy-Regler anwendbare analytische Beschreibung vereinfachen sich Regleroptimierung sowie -adaptation. Dieser Bereich ist Gegenstand zukünftiger Untersuchungen.

Literaturverzeichnis

- Akaike, H.** 1969. *Fitting autoregressive models for prediction*. Forschungsbericht 21. Institut für Statistik und Mathematik, Tohei-Suri-Kenkyusho, Tokyo.
- Angenendt, G.** 1993. *Identifikation eines selbsteinstellenden Fuzzy-Modells*. Diplomarbeit MSRT. Universität Duisburg.
- Behmenburg, C.** und **T. Bertram**. 1993. Informatik aktuell: Fuzzy-Logic, Theorie und Praxis. *Analytische Beschreibung eines Fuzzy-Systems und ihre Anwendung*, hg. von B. Reusch. 111–120. Berlin, Heidelberg, New York: Springer.
- Bertram, T.** 1992. *Herleitung einer analytischen Darstellung für ein Fuzzy-System*. Forschungsbericht 13/92 MSRT. Universität Duisburg.
- Bezdek, J. C.** 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, London: Plenum Press.
- Bronstein, I. N.** und **K. A. Semendjajew**. 1991. *Taschenbuch der Mathematik, Ergänzende Kapitel*. Thun, Frankfurt/Main: Harri Deutsch.
- Brown, M.** und **C. J. Harris**. 1991. A nonlinear adaptive controller: A comparison between fuzzy logic control and neurocontrol. *IMA Journal of Mathematical Control & Information* 8. 239–265.
- Buckley, J. J.** 1993. Sugeno type controllers are universal controllers. *Fuzzy Sets and Systems* 53. 299–303.
- Chen, S., S. A. Billings** und **P. M. Grant**. 1992. Recursive hybrid algorithm for non-linear system identification using radial basis function networks. *Int. Journal of Control* 55. 1051–1070.
- Chen, S., C. F. N. Cowan** und **P. M. Grant**. 1991. Orthogonal least squares algorithm for radial basis function networks. *IEEE Transactions on Neural Networks* 2. 302–309.
- Flury, B.** und **H. Riedwyl**. 1983. *Angewandte multivariate Statistik*. Stuttgart, New York: Gustav Fischer.
- Hartung, J.** und **B. Elpelt**. 1992. *Multivariate Statistik: Lehrbuch der angewandten Statistik*. München: Oldenbourg.
- Hathaway, R. J.** und **J. C. Bezdek**. 1993. Switching regression models and fuzzy clustering. *IEEE Transactions on Fuzzy Systems* 1. 195–204.

- Horikawa, S. I., T. Furuhashi** und **Y. Uchikawa**. 1992. On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm. *IEEE Transactions on Neural Networks* 3. 801–806.
- Isermann, R.** 1988a. *Identifikation dynamischer Systeme I*. Berlin, Heidelberg: Springer.
- Isermann, R.** 1988b. *Identifikation dynamischer Systeme II*. Berlin, Heidelberg: Springer.
- Jang, J.-S. R.** 1993. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics* 23. 665–685.
- Jang, J.-S. R.** und **C.-T. Sun**. 1993. Functional equivalence between radial basis function networks and fuzzy inference systems. *IEEE Transactions on Neural Networks* 4. 156–158.
- Kong, S. G.** und **B. Kosko**. 1992. Adaptive fuzzy systems for backing up a truck-and-trailer. *IEEE Transactions on Neural Networks* 3. 211–223.
- Krishnapuram, R.** und **J. Keller**. 1993. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy-Systems* 1. 98–110.
- Kroll, A.** 1994a. *Identifikation funktionaler Fuzzy-Modelle*. Forschungsbericht 10/94 MSRT. Universität Duisburg.
- Kroll, A.** 1994b. *Zur Parameterwahl bei Clusterverfahren zur Strukturidentifikation von Fuzzy-Modellen*. Forschungsbericht 18/94 MSRT. Universität Duisburg.
- Küpper, K.** 1994. Self-learning fuzzy models using stochastic approximation. *3rd Conference on Control Applications*. Glasgow, Scotland. 1723–1728.
- Lee, C. C.** 1990. Fuzzy logic in control systems: Fuzzy logic controller, part II. *IEEE Transactions on Systems, Man, and Cybernetics* 20. 419–435.
- Lipschutz, S.** 1980. *Wahrscheinlichkeitsrechnung*. Düsseldorf u. a.: McGraw-Hill.
- Ljung, L.** 1987. *System Identification – Theory for the User*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Lohmann, G.** 1994. *Identifikation von Funktional-Fuzzy-Modellen zur systematischen Synthese von Funktional-Fuzzy-Reglern*. Studienarbeit MSRT. Universität Duisburg.
- Meyer-Gramann, K. D.** und **E. W. Jüngst**. 1993. Fuzzy control — schnell und kostengünstig implementiert mit Standard-Hardware. *Automatisierungstechnik* 41. 166–172.

- Mizumoto, M.** 1991. Min-max-gravity versus product-sum-gravity method for fuzzy controls. *4th IFSA World Congress*. Brüssel, Belgien. 127–130.
- Moore, C. G.** und **C. J. Harris.** 1992. Indirect adaptive fuzzy control. *Int. Journal of Control* 56. 441–468.
- Murray-Smith, R.** 1994. Local model networks and local learning. *Fuzzy Duisburg '94, Int. Workshop on Fuzzy-Technologies in Automation and Intelligent Systems*. Universität Duisburg.
- Pedrycz, W.** 1984. An identification algorithm in fuzzy relational systems. *Fuzzy Sets and Systems* 13. 153–167.
- Pedrycz, W.** 1985. Design of fuzzy control algorithms with the aid of fuzzy models. *Industrial Applications of Fuzzy Control*, hg. von M. Sugeno. 153–173. North-Holland: Elsevier Science Publishers B.V.
- Poggio, T.** und **F. Girosi.** 1989. Regularization algorithms for learning that are equivalent to multilayer networks. *Science* 24. 978–982.
- Powell, M. J. D.** 1987a. Radial basis function approximation to polynomials. *Proc. of the 12th Biennial Conference on Numerical Analysis*. Dundee. 223–241.
- Powell, M. J. D.** 1987b. Radial basis functions for multivariable interpolation: A review. *Algorithms for Approximation*, hg. von J. C. Mason und M. G. Cox. 143–167. Oxford: Clarendon Press.
- Reuter, H.** 1992. *Verfahren zur Untersuchung der Nichtlinearität dynamischer Systeme*. Forschungsbericht 03/92 MSRT. Universität Duisburg.
- Reuter, H.** 1993. *Zur Identifikation bilinearer Modelle in kanonischer Form*. Forschungsbericht 08/93 MSRT. Universität Duisburg.
- Reuter, H.** 1995. *Zur Identifikation nichtlinearer Systemmodelle mit wenig A-priori-Informationen*. Dissertation. Universität Duisburg. VDI Fortschrittsberichte Reihe 8. Düsseldorf: VDI-Verlag.
- Schulte, C. M.** 1994. Genetic Algorithms for prototype based fuzzy clustering. *Int. European Congress on Intelligent Techniques and Soft Computing EUFIT*. Aachen. 913–921.
- Schwarz, H.** 1967. *Mehrfachregelungen I — Grundlagen einer Systemtheorie*. Berlin, Heidelberg, New York: Springer.
- Schwarz, H.** 1991. *Nichtlineare Regelungssysteme — Systemtheoretische Grundlagen*. München, Wien: Oldenbourg.

- Senger, M.** 1995. *Fuzzy-Regelung einer Roheisenentschwefelungsanlage*. Studienarbeit MSRT. Universität Duisburg.
- Shimojima, K., T. Fukada** und **N. Hasegawa.** 1994. Fuzzy reasoning with unsupervised/supervised learning — adaptive rule and membership function by genetic algorithm. *Proceedings of the 1st Asian Control Conference, Tokyo*. Department of Mechano-Informatics and Systems, Nagoya University. 133–136.
- Sin, S.-K.** und **R. J. P. de Figuero.** 1993. Fuzzy System design through fuzzy clustering and optimal predefuzzification. *2nd IEEE Conference on Fuzzy Systems*. San Francisco, USA. 190–195.
- Sugeno, M.** und **G. T. Kang.** 1988. Structure identification of fuzzy models. *Fuzzy Sets and Systems* 28. 15–33.
- Sugeno, M.** und **T. Yasukawa.** 1993. A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems* 1. 7–31.
- Suh, I. H.** und **T. W. Kim.** 1994. Fuzzy membership function based neural networks with applications to the visual servoing of robot manipulators. *IEEE Transactions on fuzzy systems* 2. 203–220.
- Takagi, T.** und **M. Sugeno.** 1985. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics* 15. 116–132.
- Totz, O.** 1995. *Verfahren zur Identifikation mehrdimensionaler Referenz-Fuzzy-Mengen bei Funktional-Fuzzy-Modellen*. Studienarbeit MSRT. Universität Duisburg.
- Tschiesche, T.** 1992. *Direkte Regelung von Industrierobotern*. Dissertation. Universität Duisburg. VDI Fortschrittsberichte Reihe 8 Nr. 283. Düsseldorf: VDI-Verlag.
- Vogels, R.** 1994. *Fuzzy-Reglersynthese auf der Basis von Fuzzy-Modellen*. Studienarbeit MSRT. Universität Duisburg.
- Wang, L.-X.** 1994. *Adaptive Fuzzy Systems and Control — Design and Stability Analysis*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Wernstedt, J.** 1989. *Experimentelle Prozeßanalyse*. Berlin: VEB-Verlag Technik.
- Xu, L., A. Krzyzak** und **E. Oja.** 1993. Rival Penalized Competitive Learning for Clustering Analysis, RBF Net, and Curve Detection. *IEEE Transactions on Neural Networks* 4. 636–649.
- Yamakawa, T.** 1993. A fuzzy inference engine in nonlinear analog mode and its application to a fuzzy logic control. *IEEE Transactions on Neural Networks* 4. 496–522.

- Zadeh, L.** 1973. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics* 3. 28–44.