

# **Zur Parameterwahl bei Clusterverfahren zur Strukturidentifikation von Fuzzy-Modellen**

Andreas Kroll

Forschungsbericht 18/94

Meß-, Steuer- und Regelungstechnik

**Übersicht:** Fuzzy-Clusterverfahren finden neben Aufgabenstellungen der Mustererkennung zunehmend Anwendung bei der Strukturidentifikation von Fuzzy-Modellen, die in diesem Bericht im Vordegrund steht. Für einen erfolgreichen Einsatz der Verfahren sind einige grundlegende Aspekte zu beachten sowie zahlreiche Parameter der Algorithmen geeignet zu wählen. Diese Problematik wird in diesem Bericht insbesondere im Hinblick auf metrische Clusteralgorithmen mit Zielfunktion behandelt.

Gerhard-Mercator-Universität - GH Duisburg

Meß-, Steuer- und Regelungstechnik

Prof. Dr.-Ing. H. Schwarz

# Inhaltsverzeichnis

<b>Nomenklatur</b>	<b>II</b>
<b>1 Einführung</b>	<b>1</b>
<b>2 Wahl des Datenraumes bei der Clusterung</b>	<b>3</b>
2.1 Alternativen . . . . .	3
2.2 Bewertung der Ansätze . . . . .	5
<b>3 Initialisierung</b>	<b>8</b>
<b>4 Wahl des Unschärfeparameters</b>	<b>15</b>
<b>5 Clustervalidierung</b>	<b>17</b>
<b>6 Orthogonale Partitionierung</b>	<b>21</b>
6.1 Metrische Clusteralgorithmen . . . . .	21
6.2 Überführung in orthogonale Partitionen . . . . .	23
<b>7 Zusammenfassung und Ausblick</b>	<b>26</b>
<b>Anhang</b>	<b>27</b>
<b>A Implementierung von Cluster–Algorithmen</b>	<b>27</b>

# Nomenklatur

## Skalare Größen:

$c$	: Clusteranzahl
$d_{ik}$	: Abstand zwischen $i$ -tem Clusterzentrum $\mathbf{v}_i$ und $k$ -tem Datensatz $\mathbf{x}_k$ gemäß einer vorzugebenden Abstandsnorm
$D$	: Abstandsnorm
$H$	: Partitionsentropie
$H_n$	: normierte Partitionsentropie
$i, k$	: Zählvariable
$m$	: Anzahl der eindimensionalen Eingangsgrößen (Dimension des Eingangsdatenraumes)
$n_i$	: Anzahl der Datensätze im $i$ -ten harten Cluster
$N$	: Anzahl der Datensätze
$pi$	: Gütekriterium (mittlerer quadratischer Fehler) bzgl. nicht-normierter Größen
$x_j$	: $j$ -te Eingangsgröße
$y$	: scharfe Ausgangsgröße des gesamten Regelsatzes
$y_i$	: scharfe Ausgangsgröße der $i$ -ten Regel
$\varepsilon$	: Terminierungsgrenze bei Clusterverfahren
$\phi$	: Maximal auftretende Verschiedenheit (z. B. Abstand) zweier Punkte im Datenraum $\phi = \max_{t,k}(\ \mathbf{x}_t - \mathbf{x}_k\ )$
$\mu_{ik}$	: Zugehörigkeit des Punktes $\mathbf{x}_k$ zum Cluster (zur Partition) $i$
$\nu$	: Unschärfeparameter ( $\nu > 1$ )
$\varrho_i$	: relatives Volumen des $i$ -ten Clusters
$\omega_i$	: Gesamtprämissen der $i$ -ten Regel

## Mengen und Räume:

$A, B$	: Fuzzy-Mengen
$D$	: Definitionsbereich
$\mathbb{D}_N$	: Definitionsbereich normierter Größen ( $\mathbb{D}_N = [-1; 1]$ )
$\mathbb{D}_N^+$	: Definitionsbereich normierter positiver Größen ( $\mathbb{D}_N^+ = [0; 1]$ )
$\mathbb{D}_N^m$	: normierter Definitionsbereich der Dimension $m$ ( $\mathbb{D}_N^m = \mathbb{D}_N \times \dots \times \mathbb{D}_N$ , $m$ -mal)
$\mathbb{I}$	: Menge aller $c$ Cluster, $\mathbb{I} = \{1, 2, \dots, c\}$
$\mathbb{I}_k$	: Menge aller Cluster, bei denen $\mathbf{x}_k$ den Abstand 0 vom Clusterschwerpunkt hat
$\tilde{\mathbb{I}}_k$	: $\tilde{\mathbb{I}}_k = \mathbb{I} \setminus \mathbb{I}_k$
$\mathbb{M}_c$	: („Harter“) $c$ -Partitionsraum
$\mathbb{M}_x$	: Menge aller Datensätze $\mathbf{x}_k$
$\mathbb{M}_i$	: Menge der Datensätze im harten Cluster $i$

---

$\mathbb{M}_{fc}$	: Fuzzy $c$ -Partitionsraum
$\mathbb{M}_x^\mu$	: Menge der um ihre Zugehörigkeiten zu den Clustern erweiterten Datensätze
$\mathbb{R}$	: Menge der reellen Zahlen
$\mathbb{R}^m$	: reeller Vektorraum der Dimension $m$
$\mathbb{W}$	: Wertebereich

### Funktionen und Operatoren:

$d(a, b)$	: Abstand zweier Punkte $a$ und $b$
$f(\cdot), g(\cdot), h(\cdot)$	: allgemeine Funktion
$J(\cdot)$	: Zielfunktion
$\max(\cdot)$	: Maximum
$\min(\cdot)$	: Minimum
$\ \cdot\ $	: Norm
$(\cdot)^T$	: Transponiert

### Vektoren und Matrizen:

$D$	: Bewertungsmatrix bei Darstellung einer Norm als quadratische Form
$x$	: Datensatz
$U$	: Zugehörigkeitsmatrix
$v_i$	: Clusterschwerpunkt des $i$ -ten Clusters, Partitionsmittelpunkt der $i$ -ten Partition

### Sonstige Symbole:

$\forall$	: für alle
$\setminus$	: ohne (Mengenalgebra)
$\hat{(\cdot)}$	: Schätzwert
$\overline{(\cdot)}$	: Mittelwert
$(\cdot)^*$	: Größe bei harter Partitionierung
$\mathcal{A}(\cdot)$	: Kardinalzahl (Anzahl der Elemente)

### Indizes:

$(\cdot)_i$	: Regelnummer
$(\cdot)_j$	: Nummer der Eingangsgröße
$(\cdot)_k$	: Nummer des Datums
$(\cdot)^{(l)}$	: $l$ -ter Iterationsschritt
$(\cdot)_n$	: normierte Größe
$(\cdot)^{(0)}$	: Initialisierung
$(\cdot)_{opt}$	: optimaler Wert

# 1 Einführung

Clusterverfahren wurden besonders für Aufgabenstellungen der Mustererkennung (*pattern recognition*) entwickelt. Sie können geometrische (beispielsweise sphärische, ellipsoide, punkt-, linien- oder bogenförmige) Strukturen in einer durch Datentupel gegebenen Datenmenge identifizieren. Während „harte“ Clusterverfahren nur eine Ja-/Nein-Entscheidung bei der Zuordnung eines Datentupels zu einer Struktur gestatten, ordnen „Fuzzy-Clusterverfahren“ diesem Zugehörigkeiten zwischen Null und Eins zu allen Strukturen zu. Einen Überblick über einige Verfahren geben ?, ?, sowie ?. Besonders die metrischen Verfahren mit Zielfunktion finden auch bei Teilaufgaben der Strukturidentifikation von Fuzzy-Modellen Anwendung.

Fuzzy-Modelle stellen i. allg. zeitdiskrete parametrische Ein-/Ausgangsgrößenmodelle eines Prozesses dar. Das System wird also als *schwarzer Kasten* betrachtet (?). Ein solches Modell besteht aus mehreren Regeln, deren Einflußbereiche durch die sogenannten Fuzzy-Referenzmengen festgelegt werden müssen. Diese können mit Hilfe von Clusterverfahren identifiziert werden (? , ?, ?, ?, ?, ?). Eine Verwendung von mehrdimensionalen Zugehörigkeitsfunktionen, deren Bestimmung durch Clusteralgorithmen erfolgen kann, erlaubt ein Loslösen von den verbreiteten relativ unflexiblen orthogonalen Partitionierungen, die durch die Kombination explizit eindimensionaler Zugehörigkeitsfunktionen entstehen. Das ermöglicht eine deutlich bessere Anpassung der Partitionierung an die Systemcharakteristik, die zudem einfach und automatisiert vorgenommen werden kann. Dadurch lässt sich die Modellparameteranzahl bei gleicher Modellgüte häufig reduzieren.

Bei der Anwendung von Clusteralgorithmen ist eine Vielzahl an Parametern festzulegen, die das Ergebnis der Clusterung stark beeinflussen. Dazu enthält dieser Bericht systematische Untersuchungen, die in den in der Literatur beschriebenen Anwendungen in der Regel fehlen.

Bei der Bestimmung der Fuzzy-Partitionierung des Eingangsgrößenraumes mit Hilfe von Clusterverfahren besitzt die im 2. Abschnitt behandelte Festlegung, ob nur im Eingangsdatenraum oder im Ein-/Ausgangsdatenraum geclustert wird, besondere Bedeutung.

Viele der Clusteralgorithmen, wie beispielsweise der Fuzzy- $c$ -Means-Algorithmus (FCM), konvergieren nur in lokalen Minima oder Sattelpunkten ihrer Zielfunktionen (?), weshalb die Wahl der Initialisierung der Zugehörigkeiten bzw. der Clusterschwerpunkte zu beachten ist (Abschnitt 3).

Einige Zugehörigkeitsfunktionen, wie die bei Fuzzy- $c$ -Means- und Gustafson-und-Kessel-Algorithmus verwendeten, besitzen einen expliziten „Unschärfeparameter“, über den sich die Unschärfe der Fuzzy-Partitionierung festlegen lässt. Seine Wahl hat Einfluss auf die Monotonieeigenschaften des Graphen der Zugehörigkeitsfunktionen sowie auf die Modellgüte (? , ?) und wird in Abschnitt 4 behandelt.

Ein weiterer wichtiger Parameter ist die Anzahl  $c$  der Cluster. Aus dem Bereich der Mustererkennung sind einige Kriterien zur Wahl eines optimalen Wertes für  $c$  bekannt.

---

Diese werden im 5. Abschnitt vorgestellt. Die Aufgabenstellung wird als *Clustervalidierung* bezeichnet, der Bestimmung der tatsächlich vorliegenden Anzahl an geometrischen Strukturen in den Daten.

Für die modellbasierte Synthese von Fuzzy–Reglern wird im allgemeinen von eindimensionalen orthogonalen Fuzzy–Referenzmengen ausgegangen (?, ?, ?, ?). Der 6. Abschnitt behandelt die Frage, wie die mehrdimensionalen (durch Clustering generierten) auf eindimensionale orthogonale Fuzzy–Referenzmengen abgebildet werden können.

Eine Zusammenfassung mit Ausblick schließt den Bericht ab. Im Anhang A wird eine geänderte Terminierungsbedingung für Fuzzy– $c$ –Means– sowie Gustafson–und–Kessel–Algorithmus zur Steigerung der Verarbeitungsgeschwindigkeit vorgeschlagen.

## 2 Wahl des Datenraumes bei der Clusterung

Betrachtet werde die Anwendung der Clusterung zur Fuzzy-Partitionierung des Datenraumes, um bei Fuzzy-Modellen den durch die Prämisse definierten Einflußbereich der Regeln festzulegen. Dieses Vorgehen entspricht der Identifikation von mehrdimensionalen Fuzzy-Referenzmengen für die Eingangsgrößen.

### 2.1 Alternativen

? wendet die Clusterung nur auf den  $m$ -dimensionalen Eingangsdatenraum funktionaler Fuzzy-Modelle an. Das führt zu  $c$  Clustern der Dimension  $m$ , die als mehrdimensionale Fuzzy-Referenzmengen  $A_i$ ,  $1 \leq i \leq c$ , verwendet werden. Dabei definiert die zu jeder Fuzzy-Menge  $A_i$  korrespondierende Zugehörigkeitsfunktion  $\mu_{A_i}$  eine Abbildung

$$\mathbb{D}_1 \times \dots \times \mathbb{D}_m \longrightarrow \mathbb{D}_N^+ = [0, 1] \quad , \quad (2.1)$$

$$\mu_{A_i} = g_1(x_1, \dots, x_m) \quad . \quad (2.2)$$

Die Ausgangsdaten werden bei dynamischen Fuzzy-Modellen insofern berücksichtigt, als daß ein Teil der Eingangsgrößen von verzögert rückgeführten Ausgangsgrößen gebildet wird.

? werten in der Prämisse jeder Regel jeweils zusätzlich die Prädiktion der zugehörigen Konklusionsfunktion aus, um letztere dann zu gewichten: Bei der Strukturidentifikation wird zuerst eine Clusterung im  $(m + 1)$ -dimensionalen Produktraum aus Eingangsgrößen  $\mathbf{x}$  und Ausgangsgröße  $y$  durchgeführt, die  $(m + 1)$ -dimensionale Fuzzy-Referenzmengen  $A_i$  liefert:

$$\mathbb{D}_1 \times \dots \times \mathbb{D}_m \times \mathbb{W} \longrightarrow \mathbb{D}_N^+ \quad , \quad (2.3)$$

$$\mu_{A_i} = g_2(x_1, \dots, x_m, y) \quad . \quad (2.4)$$

Anschließend werden bei der Parameteridentifikation über das Verfahren der gewichteten kleinsten Fehlerquadrate die  $c$  Konklusionspolynome als lineare Funktionen der  $m$  Eingangsgrößen bestimmt. Die Gewichtung eines Datums bzgl. einer Regel ist durch seine Zugehörigkeit zur entsprechenden Fuzzy-Referenzmenge definiert. Bei der Auswertung des identifizierten Modells liegt zwar der Wert von  $\mathbf{x}$ , nicht aber der von  $y$  vor. Deshalb erfolgen für gegebenes  $\mathbf{x}$  zuerst die Prädiktionen  $\hat{y}_i$  ( $1 \leq i \leq c$ ) durch Auswertung aller  $c$  Konklusionspolynome  $f_i$

$$\mathbb{D}_1 \times \dots \times \mathbb{D}_m \longrightarrow \mathbb{W} \quad , \quad (2.5)$$

$$\hat{y}_i = f_i(x_1, \dots, x_m) \quad , \quad (2.6)$$

in die  $y$  nicht eingeht. Die Gewichtung  $\omega_i$  der  $i$ -ten Regel bei der Zusammenfassung der

Schlußfolgerungswerte  $\hat{y}_i$  aller  $c$  Regeln nach

$$\hat{y} = \frac{\sum_{i=1}^c \omega_i \hat{y}_i}{\sum_{i=1}^c \omega_i} \quad (2.7)$$

wird durch Auswertung des  $i$ -ten Clusters für das Argument  $(\mathbf{x}, \hat{y}_i)$  bestimmt. Es gilt

$$\omega_i = \mu_{A_i}(x_1, \dots, x_m, \hat{y}_i), \quad 1 \leq i \leq c. \quad (2.8)$$

? führen die eigentliche Clusterung *nur* im Ausgangsgrößenraum durch und leiten daraus zugehörige Cluster im Eingangsdatenraum ab. Anschließend erfolgt eine Projektion dieser mehrdimensionalen Eingangsdatencluster auf die Koordinatenachsen des Eingangsdatenraumes. Die somit aus einem mehrdimensionalen Cluster entstehenden  $m$  eindimensionalen Fuzzy-Referenzmengen werden letztlich noch durch trapezoide Zugehörigkeitsfunktionen approximiert und deren genaue Form und Lage optimiert. Zuerst wird dazu vor der Clusterung eine Abbildung des  $(m+1)$ -dimensionalen Ein-/Ausgangsdatenraumes auf den eindimensionalen Ausgangsdatenraum durchgeführt. Dann wird die Clusterung (in  $c$  Cluster) nur im Ausgangsdatenraum  $\mathbb{W}$  durchgeführt

$$\mathbb{W} \longrightarrow \mathbb{D}_N^+ \quad , \quad (2.9)$$

$$\mu_{B_i} = h(y) \quad . \quad (2.10)$$

Dabei sind  $B_i$  die eindimensionalen, durch Clusterung generierten, Fuzzy-Referenzmengen. Durch Erweiterung der  $(m+1)$ -dimensionalen Datensätze um ihre Zugehörigkeiten zu den  $c$  Clustern

$$\mathbb{D}_1 \times \dots \times \mathbb{D}_m \times \mathbb{W} \longrightarrow \mathbb{D}_1 \times \dots \times \mathbb{D}_m \times \mathbb{W} \times \underbrace{\mathbb{D}_N^+ \times \dots \times \mathbb{D}_N^+}_{c-\text{mal}} \quad (2.11)$$

entsteht eine Menge

$$\mathbb{M}_x^\mu = \{(\mathbf{x}_k, y_k, \mu_{B_1}(y_k), \dots, \mu_{B_c}(y_k)) \mid 1 \leq k \leq N\} \quad (2.12)$$

von  $(m+1+c)$ -Tupeln. Anschließend wird jedem der  $c$  eindimensionalen Ausgangsgrößencluster  $B_i$  ein  $m$ -dimensionales Eingangsgrößencluster  $A_i$  zugeordnet. Die Abbildungsvorschrift dazu ist punktweise definiert über

$$\mathbb{W} \times \mathbb{D}_N^+ \longrightarrow \mathbb{D}_1 \times \dots \times \mathbb{D}_m \times \mathbb{D}_N^+ \quad (2.13)$$

$$(y_k, \mu_{B_i}(y_k)) \longrightarrow (\mathbf{x}_k, \mu_{A_i}(\mathbf{x}_k)) \quad \text{mit} \quad (2.14)$$

$$\mu_{A_i}(\mathbf{x}_k) := \mu_{B_i}(y_k) \quad . \quad (2.15)$$

Praktisch können die entstehenden Punkte direkt aus den Elementen von  $\mathbb{M}_x^\mu$  in Gl. (2.12) durch Streichung der irrelevanten Komponenten der  $(m+1+c)$ -Tupel abgeleitet werden: Für das  $i$ -te Eingangsgrößencluster folgt

$$A_i = \{(\mathbf{x}_k, \mu_{B_i}(y_k)) \mid 1 \leq k \leq N\} \quad . \quad (2.16)$$

Wenn sich herausstellt, daß ein Cluster  $A_i$  näherungsweise eine additive Überlagerung aus  $f$  (gaußglockenförmigen) Subclustern darstellt, so wird  $A_i$  in diese  $f$  Subcluster  $A_i^1, \dots, A_i^f$  aufgespalten und letztere werden an Stelle von  $A_i$  verwendet. Falls  $A_i$  nicht konvex ist (obwohl  $B_i$  diese Eigenschaft besitzt), so wird an Stelle des nichtkonvexen exakten Graphen  $\gamma$  (Forster 1984:33) von  $A_i$  dessen konvexe Hüllefläche verwendet.  $\gamma$  führen anschließend noch eine Überführung dieser  $m$ -dimensionalen in eindimensionale trapezoide Zugehörigkeitsfunktionen durch und verwenden diese dann in der üblichen WENN–DANN–Regelstruktur.

## 2.2 Bewertung der Ansätze

Der Ansatz von  $\gamma$  ist auf relationale Fuzzy–Modelle zugeschnitten, für deren Identifikation er auch angewendet wird. Die Idee dabei ist, vereinfacht ausgedrückt, alle Daten mit jeweils etwa gleicher Ausgangsgröße in einer Partition zusammenzufassen. Dies ist bei den relationalen Modellen sinnvoll, da bei jeder Regel ein fester (unscharfer) Partialausgangswert folgt. Der Erfülltheitsgrad der Prämisse führt nur indirekt über eine veränderte Gewichtung der Schlußfolgerung zu einer Verschiebung des Wertes der Ausgangsgröße bei Aggregation und Defuzzifizierung. Bei funktionalen Modellen dagegen wird unabhängig von der Wahl der Fuzzy–Referenzmengen zwischen den Punkten differenziert, die in der Partition einer Regel liegen, da die Schlußfolgerung nicht als Konstante, sondern i. allg. als Polynom in den Eingangsgrößen gegeben ist.  $\gamma$  beschreiben nicht, wie ein Eingangsgrößencluster in Subcluster zu unterteilen ist. Auch bei der eventuell durchzuführenden Umformung nichtkonvexer in konvexe Eingangsgrößencluster besteht Verbesserungspotential, da die Verwendung der konvexen Hüllen nicht zwangsläufig zu optimalen Ergebnissen führt.

Der Ansatz von  $\gamma$  ist an funktionale Modelle angepaßt. Er geht davon aus, daß im Bereich einer Häufung der Systemeingangsgrößen  $\mathbf{x}_k$  der Verlauf der Systemausgangsgröße  $y_k$  sehr nichtlinear, z. B. unstetig, sein kann. Durch die Berücksichtigung von  $y$  bei der Clusterung kann dies durch mehrere Cluster in einem solchen Bereich grundsätzlich besser von dem Modell wiedergegeben werden. Fehler können aber dadurch entstehen, daß die Prädiktion einer linearen Schlußfolgerung zur Auswertung der Referenz–Fuzzy–Mengen mitbenutzt wird. Durch Abweichungen des Modells vom System (Strukturfehler) erfolgt so durch Fehlerfortpflanzung eine Verschlechterung der Modellierungsgüte.

Der Ansatz von  $\gamma$  ist ebenfalls auf funktionale Modelle zugeschnitten. Die Idee bei dieser reinen Eingangsgrößenclusterung besteht darin, bei jeder Regel das lineare Konklusionspolynom besonders gut auf den Bereich anzupassen, in dem die größte Datenkonzentration vorliegt (Bereich um die Clusterschwerpunkte). Dieses Vorgehen kann als Linearisierung des Systemverhaltens jeweils in den Clusterschwerpunkten interpretiert werden. Die Zugehörigkeitsfunktionen bewirken eine Interpolation zwischen den einzelnen Linearisierungen im Bereich zwischen den Clusterschwerpunkten. Die Bereiche, in denen die Sy-

stemzustände vorzugsweise liegen, werden also besonders berücksichtigt. Dabei sollten die Meßdaten repräsentativ sein. Die in den einzelnen Eingangsdatenraumbereichen durchaus unterschiedlichen Ausgangsgrößen können im Gegensatz zum relationalen beim funktionalen Modellansatz besser durch die Konklusionsfunktionen approximiert werden. Da die Prädiktionen  $\hat{y}_i$  nicht zur Auswertung der Zugehörigkeitsfunktionen notwendig sind, besteht hier nicht die erhöhte Fehlerfortpflanzungstendenz wie bei ?. Allerdings geht die Ausgangsgröße nur verzögert als Eingangsgröße in die Clusterung ein. Eine Nachoptimierung der Parameter des Fuzzy-Modells stellt sich bei ? deutlich schwieriger dar: ? wertet die Zugehörigkeitsfunktionen und die Konklusionsfunktionen getrennt aus und verknüpft die beiden Teilergebnisse erst bei der Zusammenfassung aller Regeln gemäß Gl. (2.7). Bei ? geht die Auswertung der Konklusion sowohl in die Zugehörigkeitsberechnung als auch (multiplikativ) in die anschließende Berechnung von  $\hat{y}$  gemäß Gl. (2.7) ein.

Eine nur auf den Eingangsgrößenraum bezogene Clusterung ist nur bei dynamischen Modellen sinnvoll. Bei statischen Fuzzy-Modellen fehlt dagegen die Ausgangsgrößenrückführung und damit die (indirekte) Auswertung der Ausgangsgrößeninformation bei der Clusterung. Abhilfe bietet der Ansatz, die Clusterung im Ein-/Ausgangsdatenraum durchzuführen

$$\mathbb{D}_1 \times \cdots \times \mathbb{D}_m \times \mathbb{W} \longrightarrow \mathbb{D}_1 \times \cdots \times \mathbb{D}_m \times \mathbb{D}_N^+ \quad , \quad (2.17)$$

$$\mu_{\tilde{A}_i} = f(\mathbf{x}, y) \quad , \quad (2.18)$$

die Fuzzy-Referenzmengen  $\tilde{A}_i$  bzw. die Clusterzentren  $\tilde{\mathbf{v}}_i$  dann aber auf den Eingangsdatenraum abzubilden:

$$\mathbb{D}_1 \times \cdots \times \mathbb{D}_m \times \mathbb{W} \times \mathbb{D}_N^+ \longrightarrow \mathbb{D}_1 \times \cdots \times \mathbb{D}_m \times \mathbb{D}_N^+ \quad (2.19)$$

$$(\mathbf{x}, y, \mu_{\tilde{A}_i}(\mathbf{x}, y)) \longrightarrow (\mathbf{x}, \mu_{A_i}(\mathbf{x})) \quad , 1 \leq i \leq c \quad \text{mit} \quad (2.20)$$

$$\mu_{A_i}(\mathbf{x}) := \mu_{\tilde{A}_i}(\mathbf{x}, y) \quad (2.21)$$

bzw.

$$\mathbb{D}_1 \times \cdots \times \mathbb{D}_m \times \mathbb{W} \longrightarrow \mathbb{D}_1 \times \cdots \times \mathbb{D}_m \quad (2.22)$$

$$\tilde{\mathbf{v}}_i(\mathbf{x}, y) \longrightarrow \mathbf{v}_i(\mathbf{x}) \quad , 1 \leq i \leq c \quad \text{mit} \quad (2.23)$$

$$\mathbf{v}_i(\mathbf{x}) := \tilde{\mathbf{v}}_i(\mathbf{x}, y) \quad . \quad (2.24)$$

Falls die durch diese Clusterung generierte Partitionierung jedoch anschließend optimiert wird, bringt dieser Ansatz (außer eventuell besseren Startwerten) bei dynamischen Modellen keinen Vorteil gegenüber der reinen Eingangsgrößenraum-Clusterung. Dies ist offensichtlich, da die resultierenden Referenz-Fuzzy-Mengen genau die gleichen Informationen auswerten.

Impliziert wird bei den funktionalen Modellansätzen, daß sich ein System bereichsweise nahezu linear verhält und keine Unstetigkeiten aufweist. Ist das aber der Fall, so sind von ihrer Konzeption her relationale Modelle besser geeignet, da diese das Systemverhalten

punktweise mit i. allg. großer Auflösung abbilden und dazwischen interpolieren. Dagegen bilden die funktionalen Modelle eher bereichsweise ab, wobei unscharfe Übergangsbereiche generiert werden.

### 3 Initialisierung

Bei der Initialisierung der Zugehörigkeiten beim FCM– sowie beim Gustafson–und–Kessel–Algorithmus (?) können entweder die Zugehörigkeitsmatrix  $\mathbf{U}^{(0)}$  oder direkt die Clusterschwerpunkte  $\mathbf{v}_i^{(0)} (1 \leq i \leq c)$ , vorgegeben werden. Bei Vorgabe von  $\mathbf{U}^{(0)}$  werden über

$$\mathbf{v}_i^{(l)} = \frac{\sum_{k=1}^N \left(\mu_{ik}^{(l)}\right)^\nu \mathbf{x}_k}{\sum_{k=1}^N \left(\mu_{ik}^{(l)}\right)^\nu} \quad (3.1)$$

(für  $l = 0$ ) die  $\mathbf{v}_i^{(0)}$  berechnet. Es wird die abkürzende Schreibweise  $\mu_{ik} = \mu_i(\mathbf{x}_k)$  verwendet. Dabei gilt

**Satz 3.1** : Gegeben sei die Menge  $\mathbb{M}_x$  aller  $N$  Datensätze. Bei Wahl von  $\mathbf{U}^{(0)} = [\mu_{ik}^{(0)}]$  mit  $\mu_{ik}^{(0)} \in [0, 1] \forall i \in \{1, \dots, c\}, k \in \{1, \dots, N\}$  und  $\sum_{i=1}^c \mu_{ik}^{(0)} = 1 \forall k \in \{1, \dots, N\}$  liegen alle  $c \geq 2$  über Gl. (3.1) bestimmten Clusterschwerpunkte  $\mathbf{v}_i$  auf oder innerhalb der konvexen Hülle um  $\mathbb{M}_x$ .  $\square$

**Beweis:**

a) Sei  $c \geq 2$  und für ein  $i \in \{1, \dots, c\}$  sei  $\mu_{ij}^{(0)} = 1$  sowie  $\mu_{if}^{(0)} = 0 \quad \forall f \in \{1, \dots, N\} \setminus \{j\}$ . Aus Gl. (3.1) folgt

$$\mathbf{v}_i^{(0)} = \frac{\sum_{k=1}^N \left(\mu_{ik}^{(0)}\right)^\nu \mathbf{x}_k}{\sum_{k=1}^N \left(\mu_{ik}^{(0)}\right)^\nu} = \mathbf{x}_j \quad . \quad (3.2)$$

Es kann also mindestens ein Clusterschwerpunkt in einem beliebigen Datensatz  $\mathbf{x}_j$  plaziert werden.

b) Sei  $c \geq 2$  und für ein  $i \in \{1, \dots, c\}$  sei  $\mu_{ij}^{(0)} \neq 0, \mu_{ig}^{(0)} \neq 0, \mu_{if}^{(0)} = 0 \quad \forall f \in \{1, \dots, c\} \setminus \{j; g\}$ . Aus Gl. (3.1) folgt

$$\mathbf{v}_i^{(0)} = \frac{(\mu_{ij}^{(0)})^\nu \mathbf{x}_j + (\mu_{ig}^{(0)})^\nu \mathbf{x}_g}{(\mu_{ij}^{(0)})^\nu + (\mu_{ig}^{(0)})^\nu} = \frac{\mathbf{x}_j + \left(\frac{\mu_{ig}^{(0)}}{\mu_{ij}^{(0)}}\right)^\nu \mathbf{x}_g}{1 + \left(\frac{\mu_{ig}^{(0)}}{\mu_{ij}^{(0)}}\right)^\nu} = \frac{\mathbf{x}_j + \alpha \mathbf{x}_g}{1 + \alpha} \quad . \quad (3.3)$$

Wegen  $\sum_{i=1}^c \mu_{ik}^{(0)} = \mu_{ig}^{(0)} + \mu_{ij}^{(0)} = 1$  und  $\mu_{ig}^{(0)}, \mu_{ij}^{(0)} \in (0, 1)$  folgt

$$\alpha = \left(\frac{1 - \mu_{ij}^{(0)}}{\mu_{ij}^{(0)}}\right)^\nu \in (0, \infty) \quad . \quad (3.4)$$

Einsetzen von

$$\gamma := \frac{\alpha}{1 + \alpha} \in (0, 1) \quad (3.5)$$

liefert

$$\mathbf{v}_i^{(0)} = (1 - \gamma)\mathbf{x}_j + \gamma\mathbf{x}_g = \mathbf{x}_j + \gamma(\mathbf{x}_g - \mathbf{x}_j) \quad . \quad (3.6)$$

Also liegt  $\mathbf{v}_i^{(0)}$  auf der Verbindungsgeraden von  $\mathbf{x}_j$  nach  $\mathbf{x}_g$ . In den Grenzfällen  $\gamma \rightarrow 0$  oder  $\gamma \rightarrow 1$  folgt Fall a).

Aus a) und b) folgt Satz 3.1.  $\square$

Die Vorgabe von  $\mathbf{U}^{(0)}$  ist also ein Spezialfall der Vorgabe der  $\mathbf{v}_i^{(0)}$ . Eine Verletzung der Bedingung

$$\sum_{i=1}^c \mu_{ik}^{(0)} = 1 \quad \forall k \in \{1, N\} \quad (3.7)$$

kann zu Clusterschwerpunkten außerhalb der konvexen Hülle um  $\mathbb{M}_x$  führen. Übliche Strategien (?) zielen auf eine Zufallsbelegung von  $\mathbf{U}^{(0)}$  unter Beachtung von Gl. (3.7) ab, wodurch gemäß Satz 3.1 eine zufällige Plazierung der  $\mathbf{v}_i^{(0)}$  innerhalb oder auf der konvexen Hülle von  $\mathbb{M}_x$  folgt. Auszuschließen ist dabei der Fall, daß bei festen  $i \in \{1, \dots, c\}$

$$\mu_{ik}^{(0)} = a_i = \text{const.} \quad \forall k \in \{1, \dots, N\} \quad (3.8)$$

gilt, weil dann

$$\mathbf{v}_i^{(0)} = \frac{\sum_{k=1}^N (\mu_{ik}^{(0)})^\nu \mathbf{x}_k}{\sum_{k=1}^N (\mu_{ik}^{(0)})^\nu} = \frac{a_i \sum_{k=1}^N \mathbf{x}_k}{N a_i} = \bar{\mathbf{x}}_k \quad \forall i \in \{1, \dots, c\} \quad (3.9)$$

folgt und damit alle  $c$  Clusterschwerpunkte in den Punkt  $\bar{\mathbf{x}}_k$  fallen.

Werden die  $\mathbf{v}_i^{(0)}$  so vorgegeben, daß Gl. (3.7) nicht erfüllt ist, so führt die Zugehörigkeitsberechnung zu den Clusterschwerpunkten bereits nach der ersten Iteration des FCM- oder des Gustafson-und-Kessel-Algorithmus zu deren Einhaltung. Die Zuordnungsvorschrift lautet wie folgt (?):

$$\mu_{ik}^{(l+1)} = \begin{cases} \frac{1}{\sum_{j=1}^c \left( \frac{d_{ik}^{(l)}}{d_{jk}^{(l)}} \right)^{\frac{2}{\nu-1}}} & \text{für } \mathbb{I}_k^{(l)} = \emptyset \\ 0 & \forall i \in \tilde{\mathbb{I}}_k^{(l)} \\ a_{ik}^{(l)} & \forall i \in \mathbb{I}_k^{(l)} \end{cases} \quad \begin{array}{l} \text{(keine Singularitäten)} \\ \text{(Singularitäten liegen vor)} \end{array} \quad (3.10)$$

Dabei sei  $\mathbb{I}_k^{(l)} = \{i \in \mathbb{I} \mid d_{ik}^{(l)} = \|\mathbf{x}_k - \mathbf{v}_i^{(l)}\|_D = 0\}$  die Menge aller Cluster, bzgl. derer  $\mathbf{x}_k$  den Abstand 0 vom Clusterschwerpunkt hat (singulärer Fall).  $\mathbb{I} = \{1, \dots, c\}$  stellt die Menge aller  $c$  Cluster und  $\tilde{\mathbb{I}}_k = \mathbb{I} \setminus \mathbb{I}_k^{(l)}$  die Menge der Cluster dar, bzgl. derer  $\mathbf{x}_k$  keine Singularität darstellt. Außerdem ist

$$\sum_{i \in \mathbb{I}_k^{(l)}} a_{ik}^{(l)} = \sum_{i \in \mathbb{I}_k^{(l)}} \mu_{ik}^{(l)} = 1 \quad \forall i \in \mathbb{I}_k^{(l)} \neq \emptyset \quad . \quad (3.11)$$

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

Es gilt also

**Satz 3.2** : Die Zuordnungsvorschrift beim Fuzzy- $c$ -Means-Algorithmus gemäß Gl. (3.10) und (3.11) führt zu  $\mathbf{U}^{(l)}$ , die Gl. (3.7) erfüllen.  $\square$

**Beweis:** Der Beweis wird von ? geführt.

**Satz 3.3** : Die Zuordnungsvorschrift beim Gustafson-und-Kessel-Algorithmus (entspricht Gln. (3.10) und (3.11) unter Verwendung lokaler Abstandsnormen) führt zu einem  $\mathbf{U}^{(l)}$ , das Gl. (3.7) erfüllt.  $\square$

**Beweis:**

- a) Im Fall von Singularitäten ist dies per definitionem wegen Gl. (3.11) der Fall.
- b) Liegen keine Singularitäten vor, so wird der Beweis formal wie für den Fuzzy- $c$ -Means-Algorithmus geführt (?) mit dem Unterschied, daß für jedes Cluster eine andere Abstandsnorm gilt.  $\square$

Bei dem von ? beschriebenen Backer-Algorithmus werden die Fuzzy-Cluster aus harten Clustern berechnet. Deshalb muß hier zur Initialisierung eine harte Clusterung  $\mathbf{U}^{*(0)}$  vorgegeben werden; Clusterschwerpunkte berechnet der Algorithmus nicht. Es gilt

**Satz 3.4** : Die Abbildungsvorschrift (?)<sup>1</sup>

$$\mu_{ik}^{(l)} = \frac{n_i^{(l)} - \frac{1}{\phi} \sum_{t=1}^N \mu_{it}^{*(l)} d_{tk}^{(l)}}{N - \frac{1}{\phi} \sum_{t=1}^N d_{tk}^{(l)}} \quad (3.12)$$

von einer harten auf eine weiche Clusterung (mit  $\mathbb{M}_i$  der Menge aller Daten des  $i$ -ten harten Clusters und  $n_i = \mathcal{A}(\mathbb{M}_i)$ ) führt zu Zugehörigkeiten, die Gl. (3.7) erfüllen.  $\square$

---

<sup>1</sup>Die von ? aus ? zitierte Formel (2.17) ist bereits in der erstgenannten Quelle nicht korrekt.

**Beweis:**

Aus Gründen besserer Übersichtlichkeit wird der Iterationsschritt  $(\cdot)^{(l)}$  in Gl. (3.12) weggelassen. Es gilt

$$\sum_{i=1}^c \mu_{ik} = \sum_{i=1}^c \frac{n_i - \frac{1}{\phi} \sum_{t=1}^N \mu_{it}^* d_{tk}}{N - \frac{1}{\phi} \sum_{t=1}^N d_{tk}} \quad (3.13)$$

mit  $n_i = \mathcal{A}(\mathbb{M}_i)$  und damit  $\sum_{i=1}^c n_i = N$ . Damit folgt

$$\sum_{i=1}^c \mu_{ik} = \frac{N - \frac{1}{\phi} \sum_{i=1}^c \sum_{t=1}^N \mu_{it}^* d_{tk}}{N - \frac{1}{\phi} \sum_{t=1}^N d_{tk}} \quad . \quad (3.14)$$

Jeder Datensatz ist exakt einem Cluster zugeordnet ( $\mu_{it}^* \in \{0; 1\}$ ). Damit ergibt sich

$$\sum_{i=1}^c \sum_{t=1}^N \mu_{it}^* d_{tk} = \sum_{t=1}^N d_{tk} \quad (3.15)$$

und somit

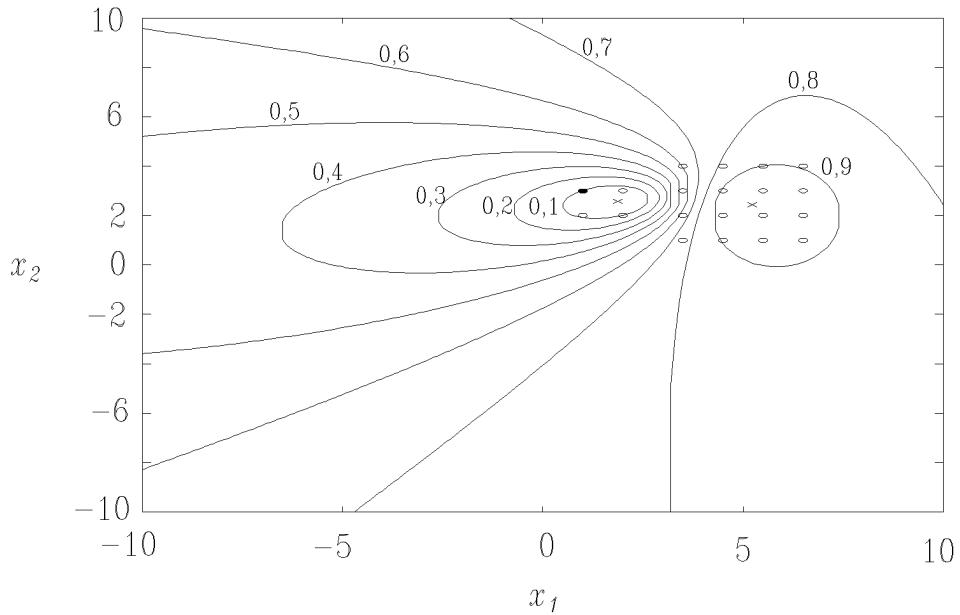
$$\sum_{i=1}^c \mu_{ik} = 1 \quad . \quad (3.16)$$

□

**Praktische Erfahrungen**

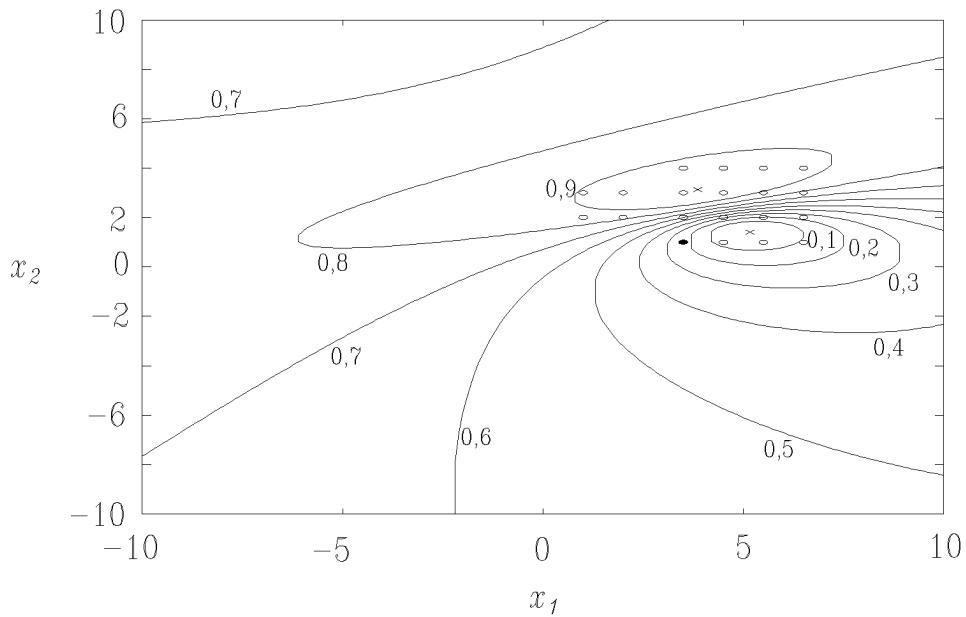
Der FCM–Algorithmus berechnet bei wohlseparierten Clustern (?), also z. B. deutlich voneinander getrennten Datenhäufungen, i. allg. unabhängig von der Initialisierung die gleichen Clusterschwerpunkte. Dies gilt sowohl bei Vorgabe von  $\mathbf{U}^{(0)}$  mit zufällig gewählten Elementen  $\mu_{ij}^{(0)} \in [0 ; 1]$ , wobei  $\sum_{i=1}^c \mu_{ij}^{(0)} = 1 \forall j \in \{1, \dots, N\}$  gilt (siehe z. B. ?), wie auch bei der Vorgabe von zufällig plazierten Clusterschwerpunkten  $\mathbf{v}_i^{(0)}$ . Eine Plazierung der  $\mathbf{v}_i^{(0)}$  außerhalb der konvexen Hülle um  $\mathbb{M}_x$  bewirkt keine Probleme (vergleiche auch Satz 3.2 und 3.3). Bei nichtwohlseparierten Clustern gibt es i. allg. mehrere lokale Minima, zu denen das Verfahren konvergieren kann. Liegen die Datenpunkte beispielsweise auf einer Parabel, so werden die Clusterschwerpunkte auch auf die Parabel gelegt.

Die Ergebnisse einer Clusterung mit dem Gustafson–und–Kessel–Algorithmus sind bereits bei wohlseparierten Datenmustern deutlich von der Initialisierung abhängig (?). Für ein exemplarisch ausgewähltes zweidimensionales Datenmuster, das aus einem kleinen (vier Datenpunkten) und einem großen (16 Datenpunkten) Rechteck besteht, liefert der Algorithmus bei unterschiedlicher Initialisierung (für  $\nu = 2$  sowie  $\varrho_1 = 1$  und  $\varrho_2 = 5$ ) die in

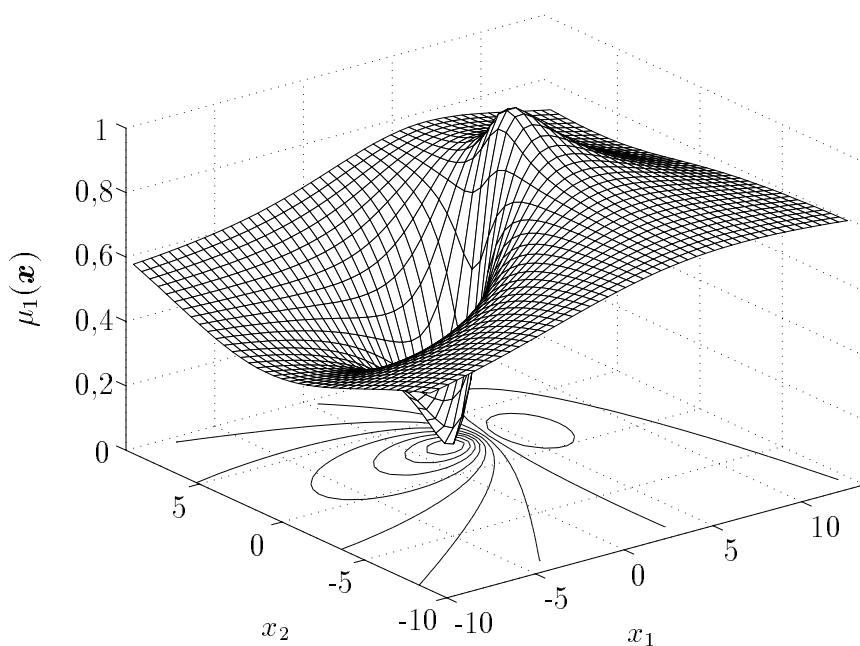


**Bild 3.1:**  $\alpha$ -Schnitte durch  $\mu_1$  (  $\times$  : Clusterschwerpunkt; Initialisierung:  $\circ$  : Datensatz mit  $\mu_1^{(0)} = 0,854$ ,  $\bullet$  : Datensatz mit  $\mu_1^{(0)} = 0,146$  )

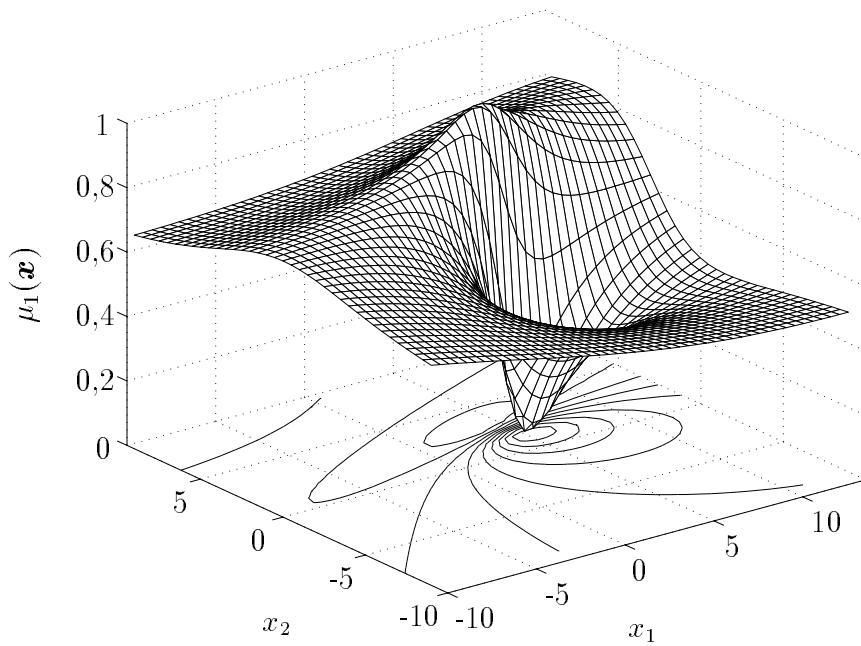
Bild 3.1 und 3.2 dargestellten Clusterungen. Dargestellt werden  $\alpha$ -Schnitte in Schritten zu 0,1 durch die zum ersten Cluster (rechts in Bild 3.1 und oben in Bild 3.2) gehörige Zugehörigkeitsfunktion  $\mu_1$ , die zur räumlichen Anschauung als dreidimensionale Graphiken in den Bildern 3.3 und 3.4 dargestellt sind. Zur Beurteilung, ob die Datenstrukturen erkannt wurden, kann das *Prinzip der maximalen Zugehörigkeit* (?) angewendet werden. Das heißt, jedes Datentupel wird genau dem Cluster zugeordnet, zu dem es die höchste Zugehörigkeit besitzt. Im vorgestellten Beispiel mit  $c = 2$  Clusterzentren muß also eine Zugehörigkeit  $\mu_i(\mathbf{x}_k) > 0,5$  vorliegen, damit  $\mathbf{x}_k$  dem Cluster  $i$  zugeordnet wird. In Bild 3.1 liegen alle Punkte des kleinen Rechtecks innerhalb des  $\alpha$ -Schnittes für  $\mu_1 = 0,5$  (und damit wegen  $\mu_2 = 1 - \mu_1$  alle Punkte des großen Rechteckes innerhalb des  $\alpha$ -Schnittes für  $\mu_2 = 0,5$ ). Die Datenstruktur wurde also erkannt. In Bild 3.2 wurden die beiden Rechtecke offensichtlich nicht erkannt. Wegen der geänderten Initialisierung konvergiert der Algorithmus in ein lokales Minimum der Zielfunktion, dessen zugehörige Clusterung nicht der Vorstellung einer „guten“ Clusterung entspricht. Das Ergebnis der Clusterung hängt außerdem von der Wahl der relativen Clustervolumina  $\varrho_i$  (?) ab. Bei gleicher Initialisierung führen unterschiedliche Kombinationen der  $\varrho_i$  zu unterschiedlichen Clusterungen. Diese Verkopplung der Parameter erschwert die Anwendung des Gustafson–und–Kessel– deutlich gegenüber dem FCM–Algorithmus.



**Bild 3.2:**  $\alpha$ -Schnitte durch  $\mu_1$  ( $\times$  : Clusterschwerpunkt; Initialisierung:  $\circ$  : Datensatz mit  $\mu_1^{(0)} = 0,854$ ,  $\bullet$  : Datensatz mit  $\mu_1^{(0)} = 0,146$ )



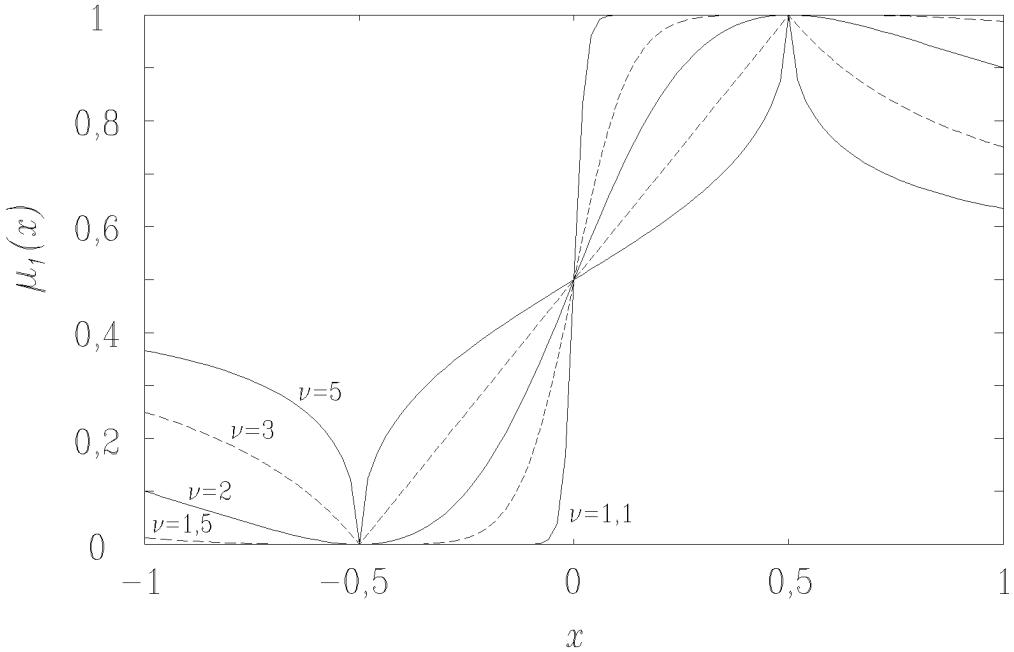
**Bild 3.3:** Zugehörigkeiten  $\mu_1(\mathbf{x})$  zum Bild 3.1, in der  $x_1 - x_2$ -Ebene sind zusätzlich die  $\alpha$ -Schnitte aus Bild 3.1 dargestellt



**Bild 3.4:** Zugehörigkeiten  $\mu_1(\mathbf{x})$  zum Bild 3.2, in der  $x_1 - x_2$ -Ebene sind zusätzlich die  $\alpha$ -Schnitte aus Bild 3.2 dargestellt

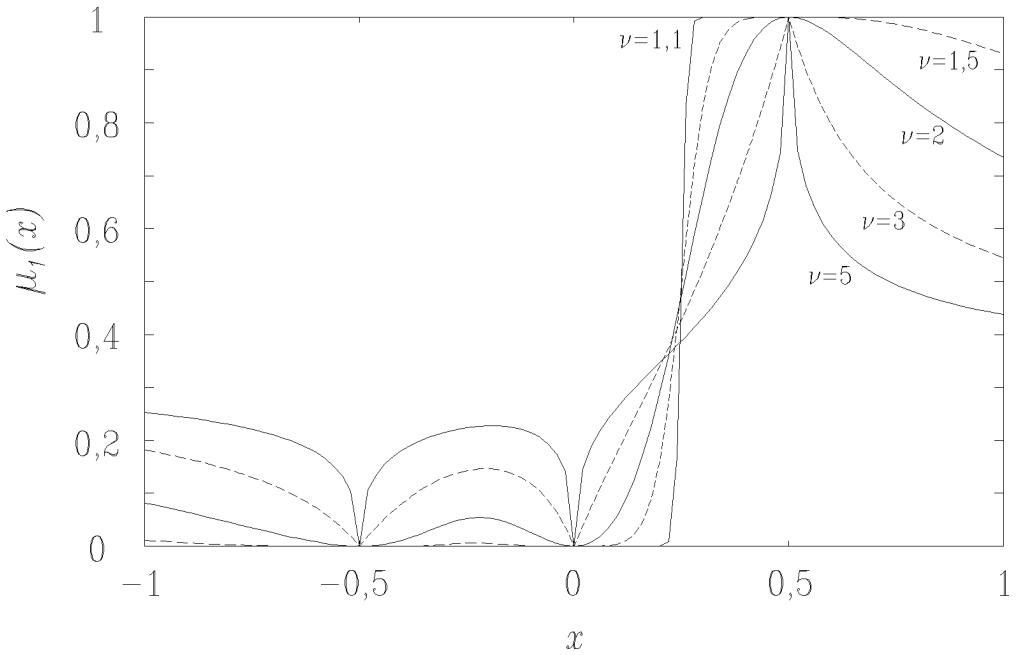
## 4 Wahl des Unschärfeparameters

Beim FCM- und beim Gustafson-und-Kessel-Algorithmus gilt wegen der Zuweisungsvorschrift für die Zugehörigkeiten Gl. (3.10) (unabhängig von der Wahl von  $\nu$ )  $d_{ik} \rightarrow d_{jk}$  für  $\|\mathbf{x}_k\| \rightarrow \infty \quad \forall i, j \in \{1, \dots, c\}$ , da  $\|\mathbf{v}_f\| < \infty \quad \forall f \in \{1, \dots, c\}$ . Damit folgt  $\mu_{ik} \rightarrow 1/c$  für  $\|\mathbf{x}_k\| \rightarrow \infty$ . Für den Verlauf von  $\mu_{ik}$  besitzt der sogenannte Unschärfeparameter  $\nu$  eine zentrale Rolle: Große Werte von  $\nu$  führen zu einer (nicht erwünschten) schnellen Annäherung von  $\mu$  an  $1/c$ . Ein weiterer unerwünschter Effekt liegt darin, daß  $\mu_{ik}$  längs der von  $\mathbf{v}_i$  ausgehenden (beliebigen) Strahlen nicht immer monoton fällt, die Zugehörigkeitsfunktionen sind also nicht konvex. Um jedes andere Clusterzentrum  $\mathbf{v}_j \neq \mathbf{v}_i$  wird eine Mulde in der Topologie von  $\mu_{ik}$  erzwungen. Die Bilder 4.1 und 4.2 illustrieren diese beiden Effekte und verdeutlichen, daß diese für einen kleinen Wert von  $\nu$  (im Beispiel  $\nu \leq 1,5$ ) eine untergeordnete Rolle spielen. Auch in Bezug auf eine hohe Güte eines Fuzzy-Modells (mit

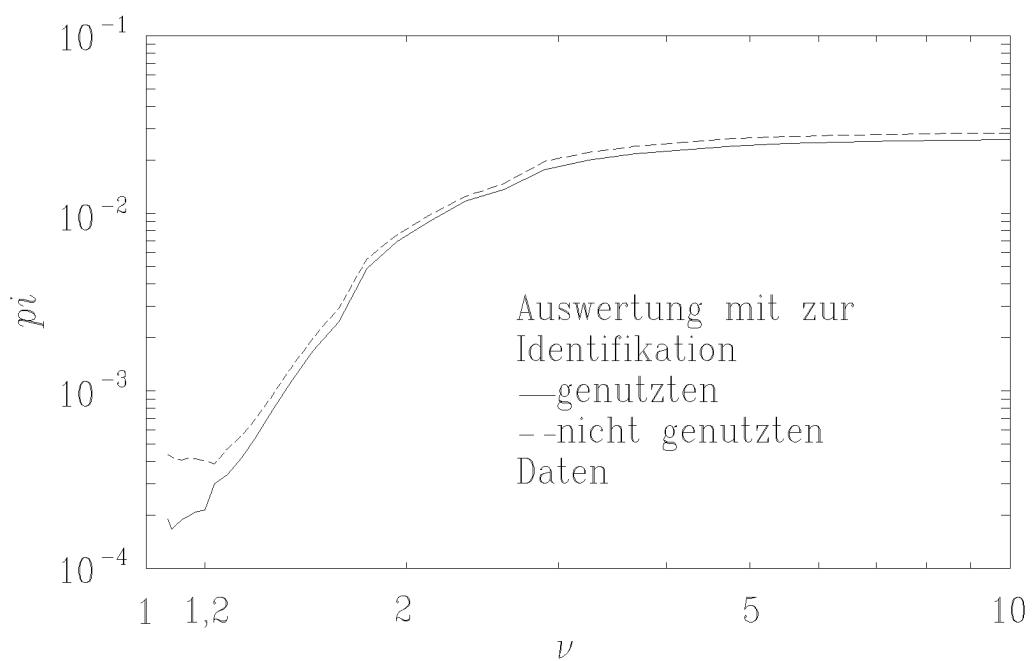


**Bild 4.1:**  $\mu_1(x)$  in Abhängigkeit von  $\nu$  für Clusterschwerpunkte  $\mathbf{v}_1 = 0,5$  und  $\mathbf{v}_2 = -0,5$

durch Clusterung erzeugter Partitionierung) sollte  $\nu$  klein gewählt werden. Eine optimale Anpassung eines Modells an eine Datensatzmenge läßt sich für  $\nu \rightarrow 1$ , also harter Partitionierung, erzielen. Dann wird jedes Datum genau einer Partition zugeordnet. Im Sinne einer robusten Modellidentifikation verschieben sich günstige Werte zu  $\nu \approx 1,2$ . Bild 4.3 zeigt die Abhängigkeit der Güte von  $\nu$  beim ersten nichtlinearen Beispielsystem aus Kroll (1994:31ff). Gegenübergestellt werden eine Modellauswertung mit dem zur Identifikation benutzten Datensatz sowie einem anderen Datensatz (Kreuzvalidierung).



**Bild 4.2:**  $\mu_1(x)$  in Abhängigkeit von  $\nu$  für Clusterschwerpunkte  $\mathbf{v}_1 = 0,5$ ,  $\mathbf{v}_2 = -0,5$  und  $\mathbf{v}_3 = 0$



**Bild 4.3:** Modellgüte  $pi$  für Einschrittprädiktion in Abhängigkeit vom Unschärfeparameter  $\nu$  (Modell mit FCM–Algorithmus generiert,  $c = 10$ )

## 5 Clustervalidierung

Bei der Mustererkennung finden mehrere Kriterien Verwendung, um die im Sinne der betrachteten Datenstruktur richtige Anzahl an Clustern zu bestimmen. Ein naheliegender Ansatz dazu ist die Auswertung der der Clusterung zugrundeliegenden Zielfunktion.

### Clusteranzahl aus Zielfunktionsminimum

FCM- und Gustafson-und-Kessel-Algorithmus minimieren (implizit) die Gesamtsumme der Summen über die gewichteten Abstandsquadrate aller Datenpunkte zum Clusterschwerpunkt innerhalb der Cluster („overall within-group sum-of-squared-error (WGSS)“, ?):

$$J_{1,\nu}(c) = \sum_{k=1}^N \sum_{i=1}^c (\mu_{ik})^\nu \| \mathbf{x}_k - \mathbf{v}_i \|_{D_i}^2 \quad . \quad (5.1)$$

Die Idee liegt dabei in einer Minimierung der gewichteten Streuung der Datensätze in den einzelnen Clustern.

? erweitern Gl. (5.1) derart, daß außerdem die Streuung der Clusterschwerpunkte maximiert wird:

$$J_{2,\nu}(c) = \sum_{k=1}^N \sum_{i=1}^c (\mu_{ik})^\nu \left( \| \mathbf{x}_k - \mathbf{v}_i \|^2 - \| \mathbf{v}_i - \bar{\mathbf{x}} \|^2 \right) \quad . \quad (5.2)$$

Die Erfahrung zeigt allerdings, daß  $J_{1,\nu}$  (selbst bei wohlseparierten Clustern) i. allg. (streng) monoton fällt. Allerdings ändert sich das Steigungsverhalten bei wohlseparierten Clustern deutlich in Abhängigkeit von  $c$ : Bei steigenden Werten von  $c$  fällt  $J_{1,\nu}$  stark ab, bis die richtige (und eindeutig bestimmte) Clusteranzahl  $c_{opt}$  erreicht wird. Steigt  $c$  weiter an, so nimmt  $J_{1,\nu}$  nur noch langsam ab. Je schlechter die Separierung der Cluster ist, desto weniger deutlich tritt dieser Effekt auf. Dagegen zeigt  $J_{2,\nu}$  bei deutlich separierten kompakten (balligen) Clustern und Werten von  $\nu$  größer als etwa 1,5 in den untersuchten Beispielen ein Minimum bei der wahren Clusteranzahl  $c_{opt}$ , d. h. bei der Anzahl voneinander getrennten geometrischen Strukturen in den Daten. Aber bereits bei beispielsweise stufenförmig (abschnittsweise linear) angeordneten Daten und damit ebenfalls deutlich separierten Clustern tritt kein Minimum bei  $J_{2,\nu}$  auf.

Die von ? angegebene gleiche Gewichtung von Subtrahend und Minuend in Gl. (5.2) beruht auf einer gleichen Gewichtung der Bestrafung der Streuung der Datensätze in den Clustern sowie der Belohnung der Streuung der Clusterzentren bei der Berechnung der „Gesamtkosten“ einer Clusterung. Diese Festsetzung ist willkürlich. Bei Setzung anderer Prioritäten sollte eine entsprechend angepaßte Gewichtung der Teilkosten erfolgen:

$$J_{3,\nu}(c) = \sum_{k=1}^N \sum_{i=1}^c (\mu_{ik})^\nu \left( \| \mathbf{x}_k - \mathbf{v}_i \|^2 - \alpha \| \mathbf{v}_i - \bar{\mathbf{x}} \|^2 \right) \quad . \quad (5.3)$$

Da nur die relativen Werte der Kostenfunktion interessant sind, reicht die Gewichtung eines Terms in Gl. (5.3) aus. Aber auch ein solcher Ansatz führt häufig nicht zu Extrema, da sowohl Subtrahend als auch Minuend oft monoton fallende Funktionen sind. Dazu kommt die Schwierigkeit, daß i. allg. kein Apriori-Wissen über eine geeignete Gewichtung vorliegt.

Bei diesen Ansätzen kann  $J_\nu$ , insbesondere bei schlechter Separierung der Cluster, mehrere lokale Minima bei festem  $c$  besitzen. Dies ist z. B. bei bogenförmig angeordneten Daten der Fall. Bei solchen Datenanordnungen existiert eine „wahre“ Anzahl an Clustern mit punktförmigen Clusterzentren (Prototypen) häufig nicht. Die Erfahrung zeigt aber, daß die zu den lokalen Minima von  $J_{1,\nu}$  gehörenden (durch den FCM-Algorithmus bestimmten) Clusterzentren auf einem solchen Bogen liegen. Die zum (numerischen) globalen Minimum für festes  $c$  gehörende Clusterung muß nicht global optimal für das Fuzzy-Modell sein, da dessen Güte i. allg. nach anderen Gütekriterien bewertet wird. Deshalb können die zu solchen lokalen Minima gehörenden Clusterungen i. allg. als Partitionierungen bei Fuzzy-Modellen eingesetzt werden. Sinnvoll ist dann eine Nachoptimierung der Partitionierung im Sinne des Gütekriteriums bei der Modellbewertung. Zu bemerken bleibt, daß bereits bei der Clusterung ein globales Minimum von  $J_\nu$  nicht gleichbedeutend mit einer guten Clusterung sein muß.

### Partitionsentropie

Ein anderes Kriterium bewertet die *Partitionsentropie*, die wie folgt definiert ist:

**Definition 5.1** : Partitionsentropie (Bezdek 1981:111)

Die Partitionsentropie einer Fuzzy- $c$ -Partition  $\mathbf{U} \in \mathbb{M}_{fc}$  der Datenmenge  $\mathbb{M}_x$  mit  $\mathcal{A}(\mathbb{M}_x) = N$  ist für  $1 \leq c \leq N$  gegeben als

$$H(\mathbf{U}, c) = -\frac{1}{N} \sum_{k=1}^N \sum_{i=1}^c \mu_{ik} \log_a(\mu_{ik}) \quad (5.4)$$

mit der logarithmischen Basis  $a \in (1, \infty)$  und

$$\mu_{ik} \log_a(\mu_{ik}) := 0 \quad \forall \mu_{ik} = 0 \quad . \quad (5.5)$$

□

Mit dieser Definition ist  $H(\mathbf{U}, c)$  ein Maß für die Unschärfe der Partitionierung. Es gilt

**Satz 5.1** (?) : Sei  $\mathbf{U} \in \mathbb{M}_{fc}$  eine Fuzzy- $c$ -Partition von  $N$  Daten. Dann gilt für  $1 \leq c \leq N$  und  $a \in (1, \infty)$

$$0 \leq H(\mathbf{U}, c) \leq \log_a(c) \quad (5.6)$$

$$H(\mathbf{U}, c) = 0 \iff \mathbf{U} \in \mathbb{M}_{c0} \text{ ist hart} \quad (5.7)$$

$$H(\mathbf{U}, c) = \log_a(c) \iff \mathbf{U} = [1/c] \quad . \quad (5.8)$$

□

Das bedeutet: Bei  $\mathbf{U} = [1/c]$  sind die Zugehörigkeitszuweisungen der Daten zu den Clustern maximal unsicher, bei harter  $c$ -Partition maximal sicher. Letzteres entspricht einem Informationsgehalt von Null und damit einem maximalen Informationsgewinn durch die Clusterung. Die Clusteranzahl  $c$  ist also genau so zu wählen, daß  $H(\mathbf{U}, c)$  minimal wird. Eine Strategie läßt sich als zweistufiger Algorithmus formulieren:

### Entropie–Algorithmus (?)

1. Bestimme  $\mathbf{U}^*$  für jedes feste  $c \in \{2, \dots, N-1\}$  so, daß die Partitionsentropie jeweils minimal ist:

$$H(\mathbf{U}_c^*, c) = \min_{\mathbf{U}_c \in \mathbb{M}_{fc}} (H(\mathbf{U}_c, c)) \quad . \quad (5.9)$$

2. Die optimale Partitionsanzahl  $c_{opt}$  führt zur minimalen Partitionsentropie aller  $H(\mathbf{U}_c^*, c)$ , die im 1. Schritt bestimmt wurden:

$$c_{opt} = \{c^* \mid H(\mathbf{U}_{c^*}^*, c^*) < H(\mathbf{U}_c^*, c) \quad \forall c, c^* \in \{2, \dots, N-1\}, c^* \neq c\} \quad . \quad (5.10)$$

□

Über die genaue Wahl von  $a$  in den Gln. (5.4) bis (5.8) wird keine Aussage gemacht, in einem Beispiel wird  $a = e$  verwendet.

Für den Fall monotoner Verläufe von  $H(\mathbf{U}, c)$  für ein gegebenes  $\mathbb{M}_x$  kann  $H(\mathbf{U}, c)$  normiert werden:

### Definition 5.2 Normierte Partitionsentropie (?)

Sei  $\mathbf{U} \in \mathbb{M}_{fc}$  eine Fuzzy- $c$ -Partition von  $\mathbb{M}_x$  und  $\mathcal{A}(\mathbb{M}_x) = N$ . Dann ist die normierte Partitionsentropie von  $\mathbf{U}$  für  $2 \leq c \leq N$  gegeben als

$$\hat{H}(\mathbf{U}, c) = \frac{H(\mathbf{U}, c)}{H_0} = \frac{H(\mathbf{U}, c)}{1 - \frac{c}{N}} \quad . \quad (5.11)$$

□

Der Algorithmus zur Bestimmung einer optimalen Clusteranzahl entspricht dem Entropie–Algorithmus, aber mit  $\hat{H}$  statt  $H$ . Wegen  $H_0 = (1-c/N) \approx 1$  unterscheiden sich normierte und nichtnormierte Partitionsentropie für  $c \ll N$  nur unwesentlich.

Bei den untersuchten Datenstrukturen (Beispiel E20.1 aus ?, Daten, die stufenförmig angeordnet sind, oder solche, die auf einer Parabel liegen, Meßdaten aus einem chemischen Reaktor (?)) konnte mit dem Entropie–Algorithmus sowohl bzgl.  $H$  als auch  $\hat{H}$

bei wohlseparierten Clustern immer  $c_{opt}$  eindeutig bestimmt werden. Dabei unterschieden sich  $H(c_{opt})$  beziehungsweise  $\hat{H}(c_{opt})$  deutlich von den Werten für  $c \neq c_{opt}$ . Bei nichtwohlseparierten Clustern besaß  $H(c)$  Randminima an den Intervallgrenzen des untersuchten Definitionsbereichs von  $c$ .

Die vorgestellten Kriterien zur Clustervalidierung können teilweise zur Bestimmung der Clusteranzahl bei deutlicher Separation der Cluster eingesetzt werden. Mit einer solchen Charakteristik der Ein- und Ausgangsdaten kann bei technischen Systemen i. allg. aber nicht gerechnet werden. In solchen Fällen sind  $J_{1,\nu}$ ,  $J_{2,\nu}$  sowie  $J_{3,\nu}$  monoton. Über Betrachtungen der Partitionsentropie kann auch keine sinnvolle Aussage bzgl. einer (endlichen) optimalen Clusteranzahl gemacht werden. Dann kann die Bestimmung von  $c_{opt}$  über das zur Modellbewertung verwendete Gütekriterium erfolgen: Sollte letzteres ebenfalls monoton von  $c$  abhängen, so existiert  $c_{opt}$  im hier betrachteten Sinne nicht. Es tritt dann nur ein zum Randminimum des Gütekriteriums korrespondierender Wert auf. Eine Festlegung der Partitionsanzahl kann dann durch Vorgabe einer gewünschten Modellgenauigkeit oder Parameteranzahl erfolgen.

## Separationsindizes

? gibt als ein Kriterium zur Überprüfung einer harten Clusterung die *Separationsindizes* an. Diese setzen den minimalen Abstand zwischen Elementen von zwei verschiedenen Clustern, der überhaupt bei allen Daten auftritt, ins Verhältnis zum Maximum der maximalen Durchmesser der  $c$  Cluster (oder deren konvexer Hüllen). Mit steigenden Werten dieser Verhältnisse steigt auch die Separierung der Cluster an. Es gibt feste Schwellen, ab denen eine CS- (compact separated) oder CWS- (compact and well separated) Clusterung vorliegt (Bezdek 1981:146). Bei deutlicher Separierung der Cluster liefert der FCM-Algorithmus gute Approximationen für eine Clusterung, deren Strategie auf einer Maximierung der Separationsindizes beruht. Da bei den betrachteten Aufgabenstellungen der Modellbildung eine deutliche Separierung i. allg. nicht vorliegt und da die Rückführung der unscharfen auf eine scharfe Clusterung einen Verlust an Aussagekraft bewirkt, wird dieser Ansatz hier nicht weiterverfolgt.

## 6 Orthogonale Partitionierung

Bei der funktionalen Fuzzy–Modellbildung nutzen sehr viele Algorithmen (?, ?, ?), bei der funktionalen Fuzzy–Regelung die meisten Verfahren (?, ?, ?), trapezoide Zugehörigkeitsfunktionen. Auch bei relationalen Fuzzy–Modellen überführen beispielsweise ? die über indirekte Clusterung generierten mehrdimensionalen in eindimensionale orthogonale Zugehörigkeitsfunktionen (siehe Abschnitt 2).

Eine Motivation für die orthogonale Partitionierung liegt darin, daß bei linearen Schlußfolgerungspolynomen in den linearen Kernbereichen der Regeln lineare Teilsysteme entstehen, die z. B. bei geschlossenen Regelkreisen mit gleichartigen funktionalen Fuzzy–Reglern eine relativ einfache theoretische Behandlung gestatten. Deshalb soll hier diese Problematik kurz behandelt werden. Dazu ist es zuerst nötig, den mathematischen Hintergrund *metrischer* Clusteralgorithmen zu erläutern.

### 6.1 Metrische Clusteralgorithmen

? beschreibt mehrere *metrische* Clusteralgorithmen. Hier soll der mathematische Hintergrund zu *metrischen* und *normierten* Räumen nachgeliefert werden.

**Definition 6.1** : Metrik, metrischer Raum (?)

Ist je zwei Elementen  $a, b$  einer beliebigen nichtleeren Menge  $\mathbb{M}$  eine reelle Zahl  $d(a, b)$  so zugeordnet, daß die Axiome

$$\begin{aligned} (M1) \quad d(a, b) &\geq 0 \quad \text{mit } d(a, b) = 0 \Leftrightarrow a = b \quad (\text{Definitheit}) \\ (M2) \quad d(a, b) &= d(b, a) \quad (\text{Symmetrie}) \\ (M3) \quad d(a, b) &\leq d(a, c) + d(c, b) \quad (\text{Dreiecksungleichung}) \end{aligned} \quad (6.1)$$

erfüllt sind, so ist auf  $\mathbb{M}$  die *Metrik*  $d$  eingeführt und  $\mathbb{M}$  ist ein *metrischer Raum*. Die Elemente eines metrischen Raumes heißen Punkte, die Zahl  $d(a, b)$  der Abstand zwischen den Punkten  $a$  und  $b$ .  $\square$

**Beispiel** (?): Wird der Abstand  $d(x, y)$  zweier Punkte  $a, b \in \mathbb{R}^m$  durch

$$d(a, b) := \left( \sum_{i=1}^m (a_i - b_i)^2 \right)^{\frac{1}{2}} \quad (6.2)$$

definiert, so wird die damit auf  $\mathbb{R}^m$  definierte Metrik als *euklidische* bezeichnet. Auch

$$d_\infty(a, b) := \max_{1 \leq i \leq m} (|a_i - b_i|) \quad (6.3)$$

sowie

$$d_1(a, b) := \sum_{i=1}^m |a_i - b_i| \quad (6.4)$$

mit  $a, b \in \mathbb{R}^m$  erklären jeweils eine Metrik auf  $\mathbb{R}^m$ .  $\square$

**Definition 6.2** : Linearer Raum (?)

Eine Menge  $\mathbb{E}$  von Elementen  $a, b$  heißt ein *linearer Raum* bezüglich der reellen Zahlen, wenn in  $\mathbb{E}$  die Operationen Addition und Multiplikation mit reellen Zahlen  $\sigma, \vartheta$  erklärt sind und folgenden Axiomen genügen:

1. Für  $a, b \in \mathbb{E}$  ist die Summe  $a + b \in \mathbb{E}$  eindeutig erklärt, und es gilt:

$$\begin{aligned} (a + b) + c &= a + (b + c) && \text{(Assoziativität)} \\ a + b &= b + a && \text{(Kommutativität)} . \end{aligned} \quad (6.5)$$

Es existiert ein Nullelement  $0 \in \mathbb{E}$  mit der Eigenschaft

$$0 + a = a \quad \forall a \in \mathbb{E} \quad \text{(neutrales Element bzgl. der Addition)} \quad (6.6)$$

und zu jedem  $a \in \mathbb{E}$  existiert ein Element  $-a \in \mathbb{E}$  mit

$$a + (-a) = 0 \quad \text{(inverses Element bzgl. der Addition).} \quad (6.7)$$

2. Zu jedem  $a \in \mathbb{E}$  und jeder reellen Zahl  $\sigma$  gilt  $\sigma \cdot a \in \mathbb{E}$  mit

$$\begin{aligned} \sigma(a + b) &= \sigma a + \sigma b && \text{(Linearität)} \\ (\sigma + \vartheta)a &= \sigma a + \vartheta a && \text{(Distributivität)} \\ (\sigma\vartheta)a &= \sigma(\vartheta a) && \text{(Assoziativität)} \\ 1 \cdot a &= a && \text{(Neutrales Element bzgl. der Multiplikation)} \end{aligned} \quad (6.8)$$

Es folgt  $0 \cdot a = 0 \quad \forall a \in \mathbb{E}$  und  $\sigma \cdot 0 = 0 \quad \forall \sigma \in \mathbb{R}$ . □

**Beispiel** :  $\mathbb{R}$  ist ein linearer Raum (Heuser 1984:118). □

**Definition 6.3** : Normierter Raum (?)

Ein linearer Raum  $\mathbb{E}$  heißt ein *normierter Raum*, wenn jedem  $f \in \mathbb{E}$  eine reelle Zahl  $\|f\|$ , die Norm von  $f$ , so zugeordnet ist, daß die drei Axiome

$$\begin{aligned} (\text{N1}) \quad \|f\| &\geq 0 && \text{mit } \|f\| = 0 \Leftrightarrow f = 0 \\ (\text{N2}) \quad \|\sigma f\| &= |\sigma| \|f\| && \forall \sigma \in \mathbb{R} \\ (\text{N3}) \quad \|f + g\| &\leq \|f\| + \|g\| && g \in \mathbb{E} \end{aligned} \quad (6.9)$$

gelten. □

**Beispiel** : Sei  $p \geq 1$  eine feste reelle Zahl.  $\mathbb{R}^m$  wird durch

$$\|a\|_p := \left( \sum_{i=1}^m |a_i|^p \right)^{\frac{1}{p}} \quad , a = [a_1, \dots, a_m]^T \quad (6.10)$$

ein normierter Raum. Gilt  $p = 2$ , so spricht man vom *euklidischen Raum*. Auch durch die Maximumnorm

$$\|a\|_\infty := \max(|a_1|, \dots, |a_m|) \quad , a = [a_1, \dots, a_m]^T \quad (6.11)$$

wird  $\mathbb{R}^m$  ein normierter Raum. □

Ein normierter Raum  $\mathbb{E}$  ist mit der Metrik

$$d(a, b) := \|a - b\| = \|b - a\| \quad (6.12)$$

ein metrischer Raum. Die Metrik (6.12) besitzt noch die zusätzlichen Eigenschaften

$$\begin{aligned} (E1) \quad d(a + c, b + c) &= d(a, b) & (\text{Translationsinvarianz}) \\ (E2) \quad d(\sigma a, \sigma b) &= |\sigma| d(a, b) & (\text{Stauchungsinvarianz}) \end{aligned} \quad . \quad (6.13)$$

Ist ein metrischer Raum  $\mathbb{E}$  ein linearer Raum und erfüllt die Metrik (E1) und (E2), dann ist  $\mathbb{E}$  mit der Norm

$$\|a\| := d(a, 0) \quad (6.14)$$

ein normierter Raum (?). Anzumerken ist, daß nicht jede Metrik translations- und stauchungsinvariant ist, wie z. B. die „Metrik des französischen Eisenbahnsystems“. Es läßt sich also nicht jeder metrische in einen normierten Raum überführen. Andererseits kann mittels Definition (6.12) jeder normierte Raum in einen metrischen überführt werden.

Der Datenraum  $\mathbb{R}^m$  bei einer Clusterung ist z. B. mit der Norm (6.10) oder (6.11) ein normierter Raum. Mit der Definition des Abstandes über die Norm gemäß (6.12) ist  $\mathbb{R}^m$  auch ein metrischer Raum. Damit ist offensichtlich, daß die von ? beschriebenen Clusteralgorithmen zu den metrischen Clusteralgorithmen gehören.

## 6.2 Überführung in orthogonale Partitionen

? schlägt vor, von zwei  $\alpha$ -Schnitten bei  $\mu = 0,9$  und bei  $\mu = 1,2/c$  (empirische Werte) die maximalen Ausdehnungen in alle Koordinatenrichtungen zu bestimmen und durch die beiden Punkte, die durch Projektion der maximalen Ausdehnung auf die Koordinatenachsen jeweils entstehen, eine Gerade zu legen. Diese definiert dann eine Flanke der eindimensionalen orthogonalen Fuzzy–Referenzmenge, die andere entsteht entsprechend durch Umsetzung der minimalen Ausdehnung. Zu beachten ist dabei, daß der Graph der mehrdimensionalen Zugehörigkeitsfunktionen nicht in alle Richtungen streng monoton ist (Abschnitt 4) und für  $\|x\| \rightarrow \infty$  gegen  $\mu = 1/c$  geht. Durch dieses Vorgehen können nicht abgedeckte Bereiche entstehen, insbesondere in den Randbereichen des Eingangsdatenraumes. Abhilfe bietet eine Ausweitung der äußeren Fuzzy–Referenzmengen in Richtung der Definitionsbereichsgrenzen. Nachteilig ist außerdem, daß i. allg. bei der Clusterung mit dem FCM–Algorithmus die euklidische Abstandsnorm verwendet wird, die zu konzentrischen Kreisen als Equinormalkurven führt. Im Hinblick auf die Umwandlung in orthogonale Partitionen ist es natürlicher, bei der Clusterung eine Metrik mit rechteckigen Equinormalkurven zu wählen. Die Metrik  $d_\infty$  liefert quadratische Equinormalkurven. Nach der Umwandlung sollte eine Optimierung der Fuzzy–Referenzmengen im Hinblick auf das bei der Modellbewertung verwendete Gütekriterium erfolgen.

**Satz 6.1** : Die Metrik  $d_\infty$  ist Translations- und Stauchungs invariant. □

**Beweis:** Translationsinvarianz (E1):

$$\begin{aligned} d_\infty(a+c, b+c) &= \max_{1 \leq i \leq m} (| (a_i + c_i) - (b_i + c_i) |) \\ &= \max_{1 \leq i \leq m} (| a_i - b_i |) = d_\infty(a, b) \end{aligned} \quad (6.15)$$

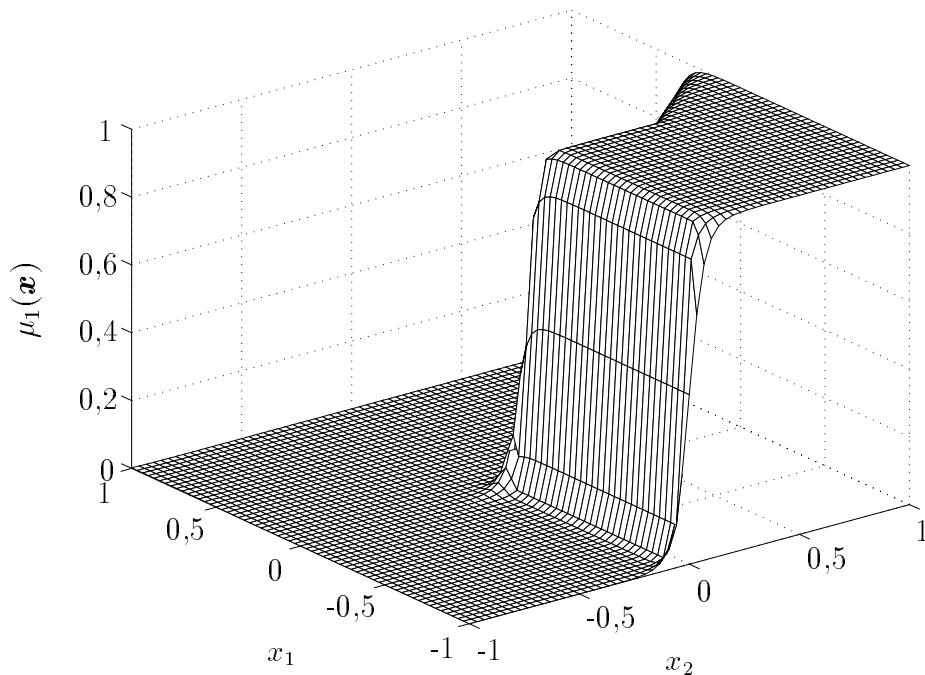
Stauchungsvarianz (E2):

$$\begin{aligned} d_\infty(\sigma a, \sigma b) &= \max_{1 \leq i \leq m} (| \sigma a_i - \sigma b_i |) \\ &= |\sigma| \max_{1 \leq i \leq m} (| a_i - b_i |) \\ &= |\sigma| d_\infty(a, b) \end{aligned} \quad (6.16)$$

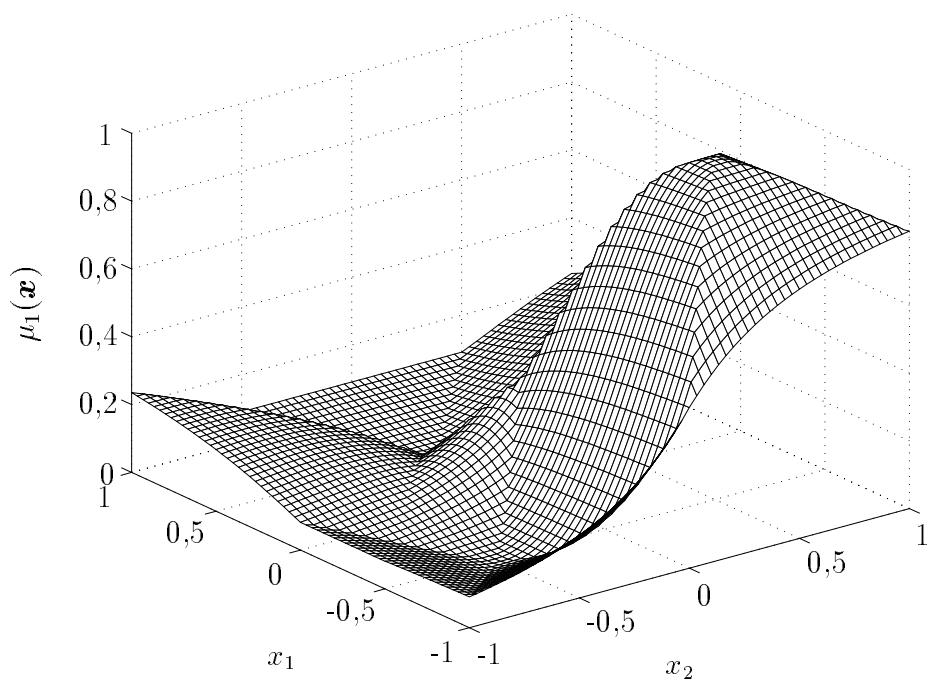
□

Die Translations- und Stauchungsvarianz sind wichtige Eigenschaften, da sie Voraussetzung für die Zulässigkeit von Verschiebung und Stauchung bei einer Normierung der Definitionsbereiche sind (z. B. Normierung von Meßdaten auf den Einheitsraum  $\mathbb{D}_N^m$ ). Die Bilder 6.1 und 6.2 illustrieren die Wirkung von  $d_\infty$  an dem bereits von ? Kroll (1994:17ff) für die euklidische Abstandsnorm gegebenen Beispiel mit den Clusterschwerpunkten

$$\mathbf{v}_1 = [-0,5; 0,5]^T, \quad \mathbf{v}_2 = [-0,5; 0,5]^T \text{ und} \quad \mathbf{v}_3 = [0,5; 0]^T \quad . \quad (6.17)$$



**Bild 6.1:** Zugehörigkeiten  $\mu_1(\mathbf{x})$  zum ersten Cluster für  $\nu=1, 13$  und Auswertung mit  $d_\infty$



**Bild 6.2:** Zugehörigkeiten  $\mu_1(\mathbf{x})$  zum ersten Cluster für  $\nu = 2$  und Auswertung mit  $d_\infty$

## 7 Zusammenfassung und Ausblick

Bei der Anwendung von metrischen Clusteralgorithmen sind verschiedene Parameter geeignet zu wählen, um eine Datenstrukturanalyse erfolgreich durchzuführen. Im Hinblick auf den Einsatz dieser Algorithmen zur Strukturidentifikation bei der Fuzzy–Modellbildung ist der Datenraum, in dem die Analyse durchgeführt wird, geeignet zu wählen. Verschiedene Möglichkeiten hierzu wurden vorgestellt und diskutiert.

Fuzzy– $c$ –Means– sowie Gustafson–und–Kessel–Algorithmus konvergieren nur lokal (?); deshalb ist besonderes Augenmerk auf die Initialisierung zu richten. Gerade der Gustafson–und–Kessel–Algorithmus, bei dem auch noch die Wahl der relativen Clustervolumina (?) das Ergebnis der Clusterung beeinflußt, ist sehr empfindlich gegenüber der Anfangswertwahl (?). Im Gegensatz zur üblichen aufwendigen Initialisierung der Zugehörigkeitsmatrix können auch die Clusterschwerpunkte initialisiert werden. Deren Bewertung bei den Terminierungsbedingungen führt zu einer Vereinfachung und somit einer Steigerung der Verarbeitungsgeschwindigkeit.

Zur Bestimmung der optimalen Anzahl der Partitionen, der sog. Clustervalidierung, wurden verschiedene Kriterien aus dem Bereich der Mustererkennung vorgestellt. Diese besitzen bei nicht wohlseparierten Datenstrukturen, die bei der Modellidentifikation für technische Systeme vorwiegend auftreten, aber i. allg. nur Randmaxima im Definitionsbereich. Hier sollte eine Wahl der Partitionsanzahl nach dem Gütekriterium für das Modell erfolgen oder, falls dessen Abhängigkeit von der Partitionsanzahl monoton ist, z. B. nach der Vorgabe einer gewünschten Parameteranzahl oder Güte des Modells.

Es wurde ein Verfahren vorgeschlagen, um mehrdimensionale auf eindimensionale orthogonale Referenz–Fuzzy–Mengen abzubilden. Dabei sollte die Clusterung bereits mit einer Metrik durchgeführt werden, die der orthogonalen Partitionierung verwandter ist als die euklidische, wie z. B. die Maximum–Norm. In diesem Zusammenhang wurde gezeigt, daß viele der eingesetzten Metriken *translations–* und *stauchungsinvariant* sind. Diese beiden Eigenschaften sind Voraussetzungen für die Zulässigkeit der üblichen Normierung der Meßdaten vor der unscharfen Verarbeitung. Allerdings ist gerade wegen der Translationsinvarianz eine Bereinigung der Meßdaten von Offsets bei der Normierung für die Auswertung der (über Abstände zu Referenzpunkten definierten) Zugehörigkeitsfunktionen überflüssig.

Durch Ersetzen der üblichen euklidischen Abstandsnorm durch eine allgemeine  $p$ –Norm ist bei angepaßter Wahl von  $p$  eine bessere Anpassung der Cluster an die real vorliegenden Datenstrukturen zu erwarten. Hierzu sind weitere Untersuchungen notwendig. Häufig erfolgt eine Normierung jeder Komponente der Meßdaten auf den jeweils maximal auftretenden Betrag dieser Komponente in allen den Meßdaten. Das führt zu einer willkürlichen Festlegung der Gewichtung der Einzelbewertungen der einzelnen Koordinaten bei der Berechnung der Gesamtkosten der Clusterung, die über die Clusteralgorithmen minimiert werden. Deshalb kann eine als quadratische Form definierte Norm Verbesserungen bringen, wenn die Komponenten der zugehörigen Matrix geeignet optimiert werden.

## Literaturverzeichnis

**Backer, E.** 1978. *Clusteranalysis by optimal decomposition of induced fuzzy sets*. Dissertation. Universität Delft/Niederlande.

**Baumann, Y., A. Dorofeyuk und D. Filev.** 1990. Fuzzy identification of nonlinear dynamical systems. *Conference on Fuzzy Logic and Neural Networks*. Iizuka/Japan. 895–898.

**Bezdek, J. C.** 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, London: Plenum Press.

**Bezdek, J. C., R. Ehrlich und W. Full.** 1984. FCM: The fuzzy- $c$ -means clustering algorithm. *Computers & Geosciences* 10. 191–203.

**Bronstein, I. N. und K. A. Semendjajew.** 1991. *Taschenbuch der Mathematik, Ergänzende Kapitel*. Thun, Frankfurt/Main: Harri Deutsch.

**Dunn, J. C.** 1974. Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics* 4. 95–104.

**Engmann, B.** 1994. *Datenstrukturanalyse mit Clusterverfahren*. Seminarvortrag MSRT. Universität Duisburg.

**Forster, O.** 1984. *Analysis 2: Differentialrechnung im  $\mathbb{R}^n$ , gewöhnliche Differentialgleichungen*. Braunschweig: Vieweg.

**Hathaway, R. J. und J. C. Bezdek.** 1988. Recent convergence results for the fuzzy- $c$ -means clustering algorithms. *Journal of Classification* 5. 237–247.

**Heuser, H.** 1983. *Lehrbuch der Analysis. Teil 2*. Stuttgart: Teubner.

**Heuser, H.** 1984. *Lehrbuch der Analysis. Teil 1*. Stuttgart: Teubner.

**Kroll, A.** 1994. *Identifikation funktionaler Fuzzy-Modelle*. Forschungsbericht 10/94 MSRT. Universität Duisburg.

**Lohmann, G.** 1994. *Identifikation von Funktional-Fuzzy-Modellen zur systematischen Synthese von Funktional-Fuzzy-Reglern*. Studienarbeit MSRT. Universität Duisburg.

**Moore, C. G. und C. J. Harris.** 1992. Indirect adaptive fuzzy control. *Int. J. of Control* 56. 441–468.

**Mourot, G., F. Kratz und J. Ragot.** 1993. Fuzzy clustering for pattern recognition diagnosis of technical system: an overview. *1st European Congress on Fuzzy and Intelligent Technologies EUFIT '93*. Aachen. 1211–1217.

**Pedrycz, W.** 1984. An identification algorithm in fuzzy relational systems. *Fuzzy Sets and Systems* 13. 153–167.

**Pedrycz, W.** 1985. An experimental study on fuzzy control using a model car. *Industrial Applications of Fuzzy Control*, hg. von M. Sugeno. 153–173. North–Holland: Elsevier Science Publishers B.V.

**Schwarz, H.** 1991. *Nichtlineare Regelungssysteme — Systemtheoretische Grundlagen*. München, Wien: Oldenbourg.

**Sin, S.-K. und R. J. P. de Figueiro.** 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.** 1985. Fuzzy modelling and control of multilayer incinerator. *Fuzzy Sets and Systems* 18. 329–360.

**Sugeno, M. und G. T. Kang.** 1988. Structure identification of fuzzy models. *Fuzzy Sets and Systems* 28. 15–33.

**Sugeno, M. und K. Murakami.** 1985. An experimental study on fuzzy control using a model car. *Industrial Applications of Fuzzy Control*, hg. von M. Sugeno. 125–138. North–Holland: Elsevier Science Publishers B.V.

**Sugeno, M. und K. Tanaka.** 1991. Successive identification of a fuzzy model and its applications to prediction of a complex system. *Fuzzy Sets and Systems* 42. 315–334.

**Sugeno, M. und T. Yasukawa.** 1993. A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems* 1. 7–31.

**Takagi, T. und M. Sugeno.** 1985. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics SMC*-15. 116–132.

**Tanaka, K. und M. Sugeno.** 1992. Stability analysis and design of fuzzy control systems. *Fuzzy Sets and Systems* 45. 135–156.

## A Implementierung von Cluster–Algorithmen

### Terminierungsbedingung

? vergleicht bei der Terminierungsbedingung die Änderungen der in der Matrix  $\mathbf{U}$  abgelegten Zugehörigkeiten aller  $N$  Meßwerte zu jeweils allen  $c$  Clustern:

$$\|\mathbf{U}^{(l+1)} - \mathbf{U}^{(l)}\|_T \leq \varepsilon_u \quad . \quad (\text{A.1})$$

? wendet hierzu die erweiterte Hölder–Norm

$$\|\mathbf{U}^{(l+1)} - \mathbf{U}^{(l)}\|_T = \left( \sum_{i=1}^c \sum_{k=1}^N \left| \mu_{ik}^{(l+1)} - \mu_{ik}^{(l)} \right|^p \right)^{\frac{1}{p}} \leq \varepsilon_u \quad (\text{A.2})$$

mit  $p = 1$  an. Dazu sind  $Nc$  Summationen vor dem Vergleich auszuführen. Da sich bei festem Unschärfeparameter  $\nu$  bei FCM- und Gustafson–und–Kessel–Algorithmus die Zugehörigkeiten eines beliebigen Punktes im Datenraum zu allen Clustern eindeutig durch Gl. (3.1) aus den  $c$  Clusterschwerpunkten  $\mathbf{v}_i$  berechnen lassen, kann zur Steigerung der Rechengeschwindigkeit auch eine mit wesentlich weniger Operationen verbundene Auswertung der Verschiebung der Clusterschwerpunkte erfolgen:

$$\|[\mathbf{v}_1^{(l+1)} - \mathbf{v}_1^{(l)}, \dots, \mathbf{v}_c^{(l+1)} - \mathbf{v}_c^{(l)}]\|_T \leq \varepsilon_v \quad . \quad (\text{A.3})$$

Eine Anwendung der erweiterten  $p$ –Hölder–Norm (?) liefert

$$\|[\mathbf{v}_1^{(l+1)} - \mathbf{v}_1^{(l)}, \dots, \mathbf{v}_c^{(l+1)} - \mathbf{v}_c^{(l)}]\|_T = \left( \sum_{i=1}^c \sum_{j=1}^m \left| \mathbf{v}_i(j)^{(l+1)} - \mathbf{v}_i(j)^{(l)} \right|^p \right)^{\frac{1}{p}} \leq \varepsilon_v \quad (\text{A.4})$$

mit der Dimension  $m$  des Datenraumes. Bei  $p = 1$  sind statt  $Nc$  nur noch  $m c$  Summationen nötig. Wegen  $N \gg m$  folgt i. allg. eine deutliche Reduzierung des Rechenaufwandes ohne Verlust an Information.

Bei dem von ? angegebenen Backer–Algorithmus werden keine Clusterschwerpunkte bestimmt, sondern die unscharfen Zugehörigkeitsgrade aus iterativ verbesserten harten Clustern abgeleitet. Deshalb kann die oben vorgeschlagene Modifikation der Terminierungsbedingung dort nicht angewendet werden.