

# Parameteroptimierung funktionaler Fuzzy-Modelle mit gradientenbasierten Verfahren

Thomas Bernd

Forschungsbericht Nr. 11/96

Meß-, Steuer- und Regelungstechnik

**Übersicht:** Die Modellbildung von dynamischen, nichtlinearen Systemen mit Hilfe funktionaler Fuzzy-Systeme zeigte in verschiedenen Anwendungen gute Ergebnisse. Die Parametrierung eines Fuzzy-Modells läßt sich auf einfache Weise mit dem bisherigen Identifikationsverfahren bestimmen, womit sich Modelle ergeben, die das Ein-/Ausgangsverhalten bereits gut abbilden. Die Modellgüte läßt sich jedoch durch Optimierung der Modellparameter weiter steigern. Der vorliegende Bericht erläutert ausführlich die effizientesten Optimierungsverfahren und untersucht deren Charakteristika anhand von Simulationsstudien. Ein pneumatischer Translationsantrieb dient als Anwendungsbeispiel. Um die Initialisierungsabhängigkeit der bekannten Verfahren zu reduzieren, wird eine Modifikation vorgestellt, die diese Problematik abschwächt.

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

# Inhaltsverzeichnis

<b>Nomenklatur</b>	<b>II</b>
<b>1 Einleitende Übersicht</b>	<b>1</b>
<b>2 Funktionale Fuzzy-Modelle</b>	<b>3</b>
2.1 Analytische Beschreibung funktionaler Fuzzy-Systeme . . . . .	3
2.2 Identifikation von funktionalen Fuzzy-Modellen . . . . .	5
<b>3 Minimierung des mittleren quadratischen Prädiktionsfehlers</b>	<b>8</b>
3.1 Mathematische Grundlagen . . . . .	8
3.2 Levenberg-Verfahren . . . . .	10
3.3 Marquardt-Verfahren . . . . .	11
3.4 Hybridverfahren . . . . .	13
3.5 Dimensionsreduzierende Formulierung der Verfahren . . . . .	15
<b>4 Optimierung von funktionalen Fuzzy-Modellen</b>	<b>17</b>
4.1 Testsystem: Pneumatischer Translationsantrieb . . . . .	17
4.2 Optimierungsergebnisse . . . . .	17
4.3 Initialisierungsabhängigkeit der Modellgüten . . . . .	25
<b>5 Zusammenfassung und Ausblick</b>	<b>31</b>
<b>6 Literatur</b>	<b>33</b>
<b>Anhang</b>	<b>35</b>
<b>A Parameter des identifizierten Modells</b>	<b>35</b>
<b>B Parameterverläufe bei rekursiver Optimierung</b>	<b>37</b>

# Nomenklatur

## Skalare Größen

$a_{i,j}$	Koeffizient der internen Eingangsgröße $y(k-j)$ im $i$ -ten Konklusionspolynom
$b_{i,h}$	Koeffizient der externen Eingangsgröße $u(k-h)$ im $i$ -ten Konklusionspolynom
$c$	Partitionsanzahl
$c_i$	konstanter Term im $i$ -ten Konklusionspolynom
$d$	Abstand
$e$	Fehler
$f$	Konklusionsfunktion
$i_1, i_2$	Steuerstrom
$k$	Abtastschritt
$m_u$	Anzahl der internen Eingangsgrößen $u(k-h)$
$m_y$	Anzahl der externen Eingangsgrößen $y(k-j)$
$n$	Parameteranzahl
$N$	Anzahl der Datensätze
$s$	Initialisierungssamen
$T$	Meßdauer
$T_0$	Abtastzeit
$u$	Eingangsgröße
$V$	mittlerer quadratischer Fehler
$y$	Ausgangsgröße
$\hat{y}$	prädizierte Ausgangsgröße
$\alpha$	reziproker Regularisierungsfaktor
$\gamma$	Dämpfungsfaktor
$\delta$	Regularisierungsfaktor
$\zeta$	Multiplikator
$\theta$	Modellparameter
$\mu$	Zugehörigkeit
$\nu$	Unschärfeparameter
$\varphi$	Winkel
$\varphi_0$	Winkelbegrenzung

## Vektoren und Matrizen

$\mathbf{0}$	Nullvektor
$\mathbf{A}$	Hilfsmatrix
$\mathbf{b}$	Hilfsvektor
$\mathbf{D}$	Formenmatrix
$\mathbf{e}$	Fehlervektor
$\mathbf{H}$	Hessematrix
$\tilde{\mathbf{H}}$	approximierte Hessematrix
$\mathbf{I}$	Einheitsmatrix
$\mathbf{J}$	Jacobimatrix
$\mathbf{x}$	Eingangsvektor
$\mathbf{v}$	Prototyp
$\mathbf{y}$	Ausgangsvektor
$\boldsymbol{\theta}$	Parametervektor
$\boldsymbol{\theta}^*$	Lösung für Parametervektor
$\boldsymbol{\psi}$	Meßvektor
$\boldsymbol{\Psi}$	Meßmatrix

## Operatoren

$\ \cdot\ $	Norm
$\nabla$	Nabla-Operator
$(\cdot)^T$	transponiert
$\angle(\mathbf{a}, \mathbf{b})$	von den Vektoren $\mathbf{a}$ und $\mathbf{b}$ eingeschlossener Winkel
$\rightarrow$	Abbildung

# 1 Einleitende Übersicht

Ein analytisches Modell eines technischen Systems stellt eine wichtige Voraussetzung für einen modellbasierten oder modellgestützten Reglerentwurf dar. Hierbei wird angestrebt, das Ein-/Ausgangsverhalten des Systems mit geringstmöglichem Fehler zu simulieren. Mit Hilfe dieser Modelle lassen sich Regler auslegen und in der Simulation testen, letztlich mit dem Ziel, Kosten, Material- und Zeitaufwand zu minimieren.

Die Modellbildung wird häufig durch nicht genau erfaßbare Nichtlinearitäten erschwert. Speziell in pneumatischen Systemen treten deutliche Reibungseinflüsse auf. Die daraus resultierenden Parameterunsicherheiten lassen bei konventionellen (z. B. theoretischen (Hebisch 1995)) Modellansätzen nur geringe Modellgüten zu. Funktionale Fuzzy-Modelle (Kroll 1995b) dagegen stellen eine analytische Beschreibungsform dar, die nur geringes Vorwissen über das System voraussetzt. Diese lassen sich mit gemessenen Ein-/Ausgangssignalen in einem zweistufigen Verfahren von Kroll (1995b) identifizieren und führten in verschiedenen Anwendungen zu hohen Modellgüten. Allerdings legt die I. Stufe nicht den quadratischen Prädiktionsfehler, anhand dessen die Modellgüte beurteilt wird, sondern ein anderes Gütekriterium als Zielfunktion zugrunde, so daß die Modelle auch nach der II. Stufe nicht optimal parametrisiert sind. Außerdem setzt das Verfahren lineare Parameterabhängigkeiten in den Schlußfolgerungen voraus. Da die Parameterabhängigkeit bez. des Gesamtmodells nichtlinear ist, erfordert die Minimierung des Prädiktionsfehlers ein nichtlineares Optimierungsverfahren.

Engmann (1995) und Kroll (1995b) setzen das Verfahren des steilsten Abstiegs als III. Identifikationsstufe ein, um den Prädiktionsfehler der Einschnittprädiktion zu reduzieren. Dieses Verfahren erweist sich jedoch aufgrund seiner hohen Iterationsanzahl als nicht praktikabel. Ein hybrides Verfahren (Kroll 1995b), das die Verfahren des steilsten Abstiegs und der kleinsten Fehlerquadrate miteinander kombiniert, senkt die erforderliche Iterationszahl bereits deutlich. Durch Substitution des Verfahrens des steilsten Abstiegs durch ein effizienteres ist jedoch eine weitere Verbesserung möglich (Bernd 1996). In (Bernd 1996) werden verschiedene nichtlineare Optimierungsverfahren untersucht, wobei sich die Verfahren von Levenberg und Marquardt (Hoffmann und Hofmann 1971) als wesentlich effizienter als das Verfahren des steilsten Abstiegs herausstellten. Außerdem wird erstmals eine Optimierung der rekursiven Modellauswertung durchgeführt. Gerade bei dieser Prädiktionsart läßt sich sowohl numerisch als auch bei Betrachtung des Zeitreihenverlaufs die Modellgüte beachtlich steigern. Die hohen Dimensionen der für diese Verfahren benötigten Matrizen sowie die Initialisierungsabhängigkeit der erzielbaren Modellgüten stellen jedoch noch Schwachpunkte dar.

Aufgrund der hohen Effizienz der Verfahren von Levenberg und Marquardt sollen diese im vorliegenden Bericht ausführlich erläutert werden. Letzteres umfaßt die mathematischen Grundlagen dieser Verfahren sowie einen Vergleich der theoretischen Überlegungen mit den Ergebnissen in der praktischen Anwendung. Weiterhin wird gezeigt, daß sich

mit einer äquivalenten Darstellung der genannten Verfahren der Nachteil der hohen Matrizen Dimensionen beseitigen läßt. Bezüglich der funktionalen Fuzzy-Modelle besteht die Annahme, daß eine Optimierung der prädiktiven Modellauswertung auch zu einer Verbesserung der rekursiven Modellauswertung führt und umgekehrt. Diese Frage wird in diesem Bericht ebenfalls untersucht und anhand von Parameterstudien geklärt. Wie außerdem gezeigt wird, läßt sich der Nachteil der Initialisierungsabhängigkeit durch eine Modifikation des bisherigen Identifikationsverfahrens abschwächen. Eine systematische Vorgehensweise führt dabei zu Modellen hoher Güte.

Der vorliegende Bericht gliedert sich somit wie folgt: Im 2. Abschnitt folgt eine kurze Darstellung der funktionalen Fuzzy-Modelle und des Identifikationsverfahrens. Eine ausführliche Darstellung der mathematischen Grundlagen der Verfahren von Levenberg und Marquardt gibt der 3. Abschnitt. In diesem Zusammenhang wird die Problematik der hohen Matrizen Dimensionen der beiden Optimierungsverfahren diskutiert und eine äquivalente Darstellung der Verfahren vorgestellt. Ein Vergleich der Verfahren wird im 4. Abschnitt anhand eines Identifikationsdatensatzes für einen pneumatischen Translationsantrieb durchgeführt. Dabei steht die Untersuchung des Konvergenzverhaltens der beiden Verfahren durch ausführliche Parameterstudien und ein Vergleich der prädiktiven und rekursiven Optimierung im Vordergrund. Außerdem wird eine Modifikation des Identifikationsverfahrens erläutert, die den Nachteil der Initialisierungsabhängigkeit der Modellgüten abschwächt. Eine Zusammenfassung und ein Ausblick schließen diesen Bericht ab.

## 2 Funktionale Fuzzy-Modelle

Das Ziel einer Modellbildung ist, den Fehler zwischen einer gemessenen und einer vom Modell prädizierten Ausgangsgröße bei identischem Eingangssignal zu minimieren (vgl. Bild 2.1). In den Abschnitten 2.1 und 2.2 folgt eine Erläuterung funktionaler Fuzzy-Systeme und der Identifikation der darauf basierenden funktionalen Fuzzy-Modelle.

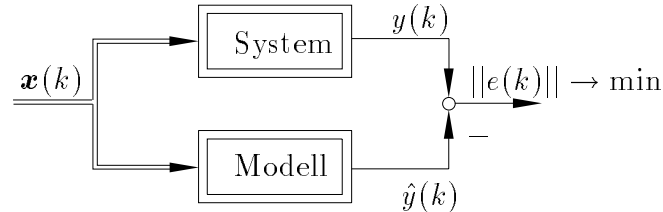


Bild 2.1: Zur Modellbildung

### 2.1 Analytische Beschreibung funktionaler Fuzzy-Systeme

Funktionale MISO- (Multi-Input Single-Output) Fuzzy-Systeme (Takagi und Sugeno 1985) stellen eine mathematische Realisierung einer Abbildung  $\mathbf{x} \rightarrow y$  eines Vektors  $\mathbf{x}$  in ein Skalar  $y$  dar. Das Prinzip der funktionalen Fuzzy-Systeme besteht darin, zur Beschreibung eines Ein-/Ausgangsverhaltens nicht einen einzelnen nichtlinearen (z. B. bilinearen (Schwarz 1991)) Zusammenhang  $y = f(\mathbf{x})$  zu verwenden, sondern den Eingangsgrößenraum in  $c$  Partitionen zu unterteilen und jeder Partition eine (meist lineare) Funktionsgleichung  $y_i = f(\mathbf{x}, \boldsymbol{\theta}_i)$  mit einer speziellen Parametrierung  $\boldsymbol{\theta}_i$ ,  $i = 1, 2, \dots, c$  zuzuordnen (vgl. Bild 2.2). Im Sinne der Fuzzy-Theorie sind die Partitionen unscharf voneinander abgegrenzt, so daß sich zu jedem Eingangsvektor  $\mathbf{x}$  mehrere (bei Verwendung mehrdimensionaler Zugehörigkeitsfunktionen genau  $c$ ) Partialausgangsgrößen erge-

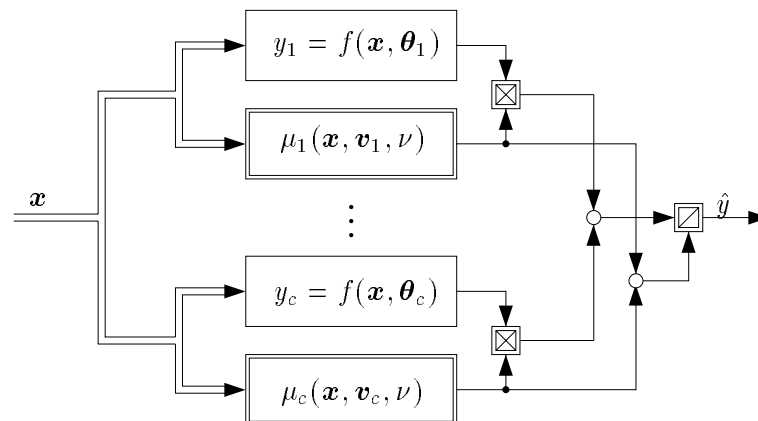


Bild 2.2: Blockschaltbild eines funktionalen Fuzzy-Systems

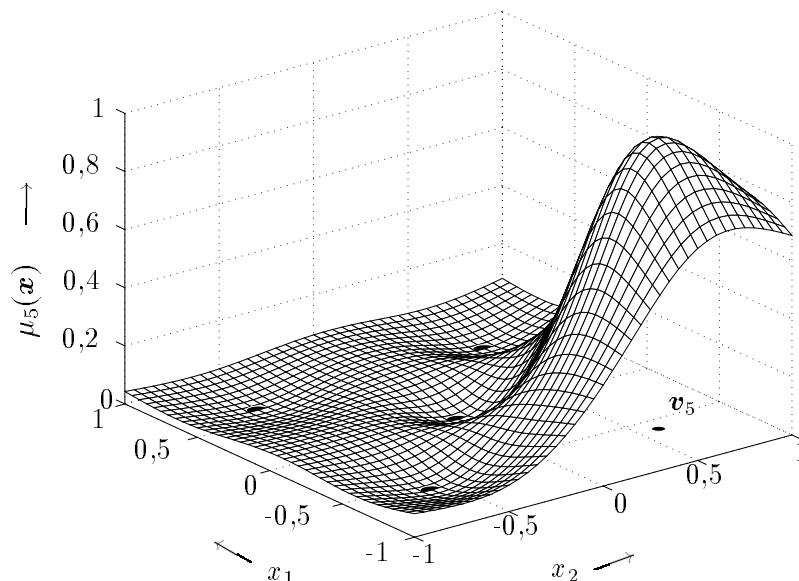
ben. Die Berechnung der resultierenden Ausgangsgröße aus den Partialausgangsgrößen erfolgt mit Hilfe folgender Fuzzy-Regeln:

$$\text{Regel } i: \text{ WENN } (\mathbf{x} \text{ IST } \mathbf{v}_i), \text{ DANN } (y_i = f(\mathbf{x}, \boldsymbol{\theta}_i)) \quad . \quad (2.1)$$

Die Prämisse einer Regel  $i$  bewertet die Zugehörigkeit eines Eingangsvektors zur  $i$ -ten Partition. Zu diesem Zweck ist jeder Partition ein sog. Prototyp  $\mathbf{v}_i$  zugeordnet, mit dessen Hilfe sich ein Abstand  $d_i = \|\mathbf{x} - \mathbf{v}_i\|$  berechnen läßt. Nach Bestimmung aller Abstände ergeben sich dann die Zugehörigkeiten  $\mu_i(\mathbf{x})$  aus den Zugehörigkeitsfunktionen

$$\mu_i(\mathbf{x}) = \begin{cases} \frac{1}{\sum_{j=1}^c \left( \frac{d_j(\mathbf{x})}{d_j(\mathbf{x})} \right)^{\frac{2}{\nu-1}}} & \text{falls } d_j \neq 0 \quad \forall j = 1, \dots, c \\ 0 & \text{falls } d_j = 0 \text{ und } i \neq j \\ 1 & \text{falls } d_j = 0 \text{ und } i = j \end{cases} \quad . \quad (2.2)$$

Zur Veranschaulichung einer Zugehörigkeitsfunktion gemäß Gl. (2.2) soll Bild 2.3 dienen. Es illustriert, daß Datenpunkte in der Nähe des Prototypen  $\mathbf{v}_5$  eine hohe, weiter entfernte eine geringe Zugehörigkeit zu dieser Partition besitzen. Der Unschärfeparameter  $\nu$  beeinflußt dabei die Unschärfe der Partitionierung: Der Wert  $\nu \rightarrow 1$  führt zu einer scharfen, ein größerer Wert dagegen zu einer unscharfen Partitionierung. Wie Kroll (1995b) beweist, bildet  $\nu = 3$  die obere Grenze der für ein gradientenbasiertes Optimierungsverfahren wichtigen Voraussetzung der stetigen Differenzierbarkeit der Zugehörigkeitsfunktionen.



**Bild 2.3:** Beispiel einer Zugehörigkeitsfunktion bei zwei Eingangsgrößen ( $c = 5$ ,  $\nu = 2$ , Euklidische Abstandsnorm)

Nach Auswertung aller Regeln liegen somit  $c$  Partialausgangsgrößen  $y_i(\mathbf{x})$  und  $c$  Zugehörigkeiten  $\mu_i(\mathbf{x})$  vor, woraus sich durch eine gewichtete Mittelwertbildung die resul-



tierende Ausgangsgröße

$$y = \frac{\sum_{i=1}^c \mu_i(\mathbf{x}) y_i(\mathbf{x})}{\sum_{i=1}^c \mu_i(\mathbf{x})} \quad (2.3)$$

ergibt.

## 2.2 Identifikation von funktionalen Fuzzy-Modellen

Die Modellidentifikation mit Hilfe der in Abschnitt 2.1 beschriebenen Fuzzy-Systeme erfordert zunächst die Vorgabe der Partitionenanzahl  $c$ , des Unschärfeparameters  $\nu$  und einer Abstandsnorm. Erläuterungen zur Wahl dieser Parameter gibt Kroll (1995a). Von wesentlicher Bedeutung ist ebenfalls die Wahl der Konklusionsfunktion  $f(\mathbf{x}, \boldsymbol{\theta}_i)$ . Als vorteilhaft erweisen sich lineare Polynome

$$y_i(k) = c_i - \sum_{j=1}^{m_y} a_{ij} y(k-j) + \sum_{h=1}^{m_u} b_{ih} u(k-h) \quad , \quad (2.4)$$

da sich in diesem Fall die Parameter des Konklusionspolynoms  $a_{ij}$ ,  $b_{ih}$  und  $c_i$  mit dem im folgenden beschriebenen Identifikationsverfahren leicht bestimmen lassen. Nach Festlegung der Anzahl der internen Eingangsgrößen  $m_y$  und der externen Eingangsgrößen  $m_u$  verbleibt als Identifikationsaufgabe die Bestimmung der Prototypenkoordinaten und der Parameter des Konklusionspolynoms. Dazu dient folgendes dreistufiges Identifikationsverfahren:

### I. Stufe

In der I. Stufe erfolgt zunächst die Bestimmung der Prototypenkoordinaten. Das Ziel ist hierbei, die Summe aller Abstände  $\|\mathbf{x}(k) - \mathbf{v}_i\|$ , gewichtet mit den jeweiligen Zugehörigkeiten  $\mu_i(k)$ , d. h. die Kostenfunktion (Bezdek 1981)

$$V = \sum_{k=1}^N \sum_{i=1}^c (\mu_i(k))^\nu \|\mathbf{x}(k) - \mathbf{v}_i\|^2 \quad , \quad (2.5)$$

zu minimieren. Die Aufgabe läßt sich mit verschiedenen Fuzzy-Clusteralgorithmen, z. B. der Fuzzy-c-means-Algorithmus (FCM, (Bezdek 1981)) lösen. Das Ergebnis der iterativen Prototypenberechnung ist initialisierungsabhängig.

## II. Stufe

Nach der Durchführung der I. Identifikationsstufe verbleiben lediglich die Konklusionsfunktionsparameter als numerisch unbekannte Größen. Wenn es sich bei den Konklusionsfunktionen um in den Parametern lineare Polynome nach Gl. (2.4) handelt, lassen sich die Konklusionsfunktionsparameter auf direktem Wege mit dem Verfahren der kleinsten Fehlerquadrate (LS-Verfahren) (Isermann 1992)

$$\boldsymbol{\theta} = [\boldsymbol{\Psi}^T \boldsymbol{\Psi}]^{-1} \boldsymbol{\Psi}^T \mathbf{y} \quad \text{mit} \quad \mathbf{y} = [y(1), y(2), \dots, y(N)]^T \quad (2.6)$$

berechnen, wobei die mit den Zugehörigkeiten gewichteten Eingangsgrößen zu den Meßvektoren

$$\begin{aligned} \boldsymbol{\psi}(k) = & [\mu_1, -\mu_1 y(k-1), \dots, -\mu_1 y(k-m_y), \mu_1 u(k-1), \dots, \mu_1 u(k-m_u), \\ & \vdots \\ & \mu_c, -\mu_c y(k-1), \dots, -\mu_c y(k-m_y), \mu_c u(k-1), \dots, \mu_c u(k-m_u)]^T \end{aligned} \quad (2.7)$$

zusammengefaßt werden. Jeder der  $N$  Meßvektoren bildet schließlich eine Zeile der Meßmatrix

$$\boldsymbol{\Psi} = [\boldsymbol{\psi}(1), \boldsymbol{\psi}(2), \dots, \boldsymbol{\psi}(N)]^T \quad . \quad (2.8)$$

Bei Verwendung der Meßvektoren gemäß Gl. (2.7) lautet der Parametervektor nach Anwendung des LS-Verfahrens

$$\boldsymbol{\theta} = [c_1, a_{11}, \dots, a_{1m_y}, b_{11}, \dots, b_{1m_u}, c_2, \dots, b_{cm_u}]^T \quad . \quad (2.9)$$

Die Voraussetzung der Linearität beschränkt die Anwendbarkeit des LS-Verfahrens allerdings auf die prädiktive Modellauswertung der Meßdaten. Bei rekursiver Modellauswertung dagegen stehen die gemessenen Ausgangsgrößen für die Prädiktion und die Notation der Meßmatrix  $\boldsymbol{\Psi}$  nicht zur Verfügung, so daß die Konklusionsfunktionsparameter bei dieser Prädiktionsart nicht direkt mit dem LS-Verfahren berechenbar sind. Die Berechnung der Konklusionsfunktionsparameter bez. der prädiktiven Modellauswertung liefert in einigen Fällen auch bei rekursiver Auswertung gute Ergebnisse. Im allgemeinen wird das Potential des Modellansatzes bei dieser Prädiktionsart durch Berechnung der Konklusionspolynomparameter mit dem LS-Verfahren jedoch nicht ausgeschöpft.

## III. Stufe

Die oben genannte Einschränkung läßt sich in dieser III. Stufe abschwächen und ein weiterer Nachteil beseitigen, der in den unterschiedlichen Zielfunktionen zu sehen ist: In

der II. Stufe erfolgt die Berechnung der Konklusionspolynomparameter bez. des mittleren quadratischen Fehlers

$$V = \frac{1}{N} \sum_{k=1}^N (y(k) - \hat{y}(k))^2 \quad , \quad (2.10)$$

der letztendlich zur Beurteilung der Modellgüte herangezogen wird. Zuvor wurden die Prototypen jedoch bez. einer anderen Zielfunktion (Gl. (2.5)) optimiert. Die sich nach der II. Stufe ergebenden Werte der Modellparameter stellen somit kein lokales Minimum des quadratischen Fehlers dar.

Die Optimierung aller Modellparameter mit einem gradientenbasierten Verfahren, ausgehend von den Parameterwerten nach der II. Stufe, ist somit das Ziel der III. Stufe. Dabei stellt der mittlere quadratische Fehler (s. Gl. (2.10)) die zu minimierende Zielfunktion dar. Die Vorgehensweise zur Optimierung der prädiktiven Auswertung unterscheidet sich dabei nicht von einer rekursiven Optimierung. Deshalb ist im Gegensatz zu dem bisherigen Identifikationsverfahren in dieser III. Stufe auch eine Minimierung des quadratischen Fehlers bei rekursiver Auswertung möglich.

### 3 Minimierung des mittleren quadratischen Prädiktionsfehlers

Die Minimierung des mittleren quadratischen Prädiktionsfehlers stellt ein nichtlineares Optimierungsproblem in den Modellparametern  $\boldsymbol{\theta}$  dar. Aus der Literatur sind zahlreiche nichtlineare Optimierungsverfahren wie das Powell–Verfahren (Foulds 1981), der steilste Abstieg (Entenmann 1976) und das Newton–Raphson–Verfahren (Entenmann 1976) bekannt, mit denen sich die vorliegende Optimierungsaufgabe prinzipiell lösen läßt. Vorteilhafter erscheinen dagegen die speziell auf die Minimierung von quadratischen Fehlern ausgelegten Verfahren von Gauß–Newton, Levenberg und Marquardt (Hoffmann und Hofmann 1971), da sie lediglich die Berechnung der ersten Ableitungen der Zielfunktion erfordern und dennoch quadratisch konvergieren. Die zweiten Ableitungen, deren Berechnung im Fall der funktionalen Fuzzy–Modelle mit unverhältnismäßig hohem Aufwand verbunden wäre, werden dagegen nicht in analytischer Form benötigt, sondern numerisch approximiert. Von den drei zuletzt genannten Verfahren sollen im weiteren die Verfahren von Levenberg und Marquardt zur Anwendung kommen, da diese im Gegensatz zu den Verfahren von Gauß–Newton und des steilsten Abstiegs keine heuristische Bestimmung eines Algorithmusparameters erfordern.

#### 3.1 Mathematische Grundlagen

Die Verfahren von Levenberg und Marquardt basieren auf der in der analytischen Mathematik gebräuchlichen Vorgehensweise zur Bestimmung eines Extremwertes (oder Sattelpunktes). Hierbei sind die ersten partiellen Ableitungen der zu minimierenden Zielfunktion  $V$  zu berechnen und diejenigen Parameterwerte zu bestimmen, die der notwendigen Bedingung  $\nabla V = 0$  genügen. Im  $n$ -dimensionalen Fall ist folglich das Gleichungssystem

$$\nabla V(\boldsymbol{\theta}) = \left[ \frac{\partial V}{\partial \theta_1}, \dots, \frac{\partial V}{\partial \theta_n} \right]^T = \mathbf{0} \quad (3.1)$$

zu lösen. Im weiteren ist mit Hilfe der zweiten partiellen Ableitungen zu prüfen, ob an der Stelle ein Minimum, Maximum oder Sattelpunkt vorliegt.

Eine geschlossene Lösung von Gl. (3.1) ist i. allg., wie auch für funktionale Fuzzy–Modelle, nicht möglich. Eine Taylorreihenentwicklung des Gradienten an der Stelle  $\boldsymbol{\theta}_i$

$$\nabla V(\boldsymbol{\theta}_{i+1}) = \nabla V(\boldsymbol{\theta}_i) + \left. \frac{\partial^2 V(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}_i} (\boldsymbol{\theta}_{i+1} - \boldsymbol{\theta}_i) + \dots \quad , \quad (3.2)$$

und Linearisierung durch Abbruch nach dem zweiten Glied führt dagegen auf die Beziehung

$$\nabla V(\boldsymbol{\theta}_{i+1}) \approx \nabla V(\boldsymbol{\theta}_i) + \left. \frac{\partial^2 V(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}_i} (\boldsymbol{\theta}_{i+1} - \boldsymbol{\theta}_i) = \mathbf{0} \quad , \quad (3.3)$$

die sich zu der Iterationsvorschrift

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - \left[ \frac{\partial^2 V(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_i} \right]^{-1} \nabla V(\boldsymbol{\theta}_i) \quad (3.4)$$

umformen läßt. Die Matrix der zweiten partiellen Ableitungen

$$\mathbf{H}(\boldsymbol{\theta}) := \frac{\partial^2 V(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial^2 V(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_1} & \cdots & \frac{\partial^2 V(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 V(\boldsymbol{\theta})}{\partial \theta_n \partial \theta_1} & \cdots & \frac{\partial^2 V(\boldsymbol{\theta})}{\partial \theta_n \partial \theta_n} \end{bmatrix} \quad (3.5)$$

wird als Hessematrix  $\mathbf{H}$  bezeichnet, mit der sich Gl. (3.4) übersichtlicher schreiben läßt als

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - \mathbf{H}(\boldsymbol{\theta}_i)^{-1} \nabla V(\boldsymbol{\theta}_i) \quad . \quad (3.6)$$

Die iterative Bestimmung des Minimums nach Gl. (3.6) ist in der Literatur (z. B. (Entenmann 1976)) als Newton–Raphson–Verfahren bekannt. Das Verfahren setzt jedoch die explizite Berechnung der zweiten Ableitungen (s. Gl. (3.5)) voraus, so daß der Anwendung von Gl. (3.6) i. allg. ein hoher analytischer Aufwand vorausgeht.

Die Verfahren von Levenberg und Marquardt stellen dagegen eine Modifikation des Newton–Raphson–Verfahrens dar, bei denen sich die Hessematrix numerisch approximieren läßt, wenn es sich bei der zu minimierenden Zielfunktion um einen quadratischen Fehler

$$V = \mathbf{e}^T \mathbf{e} \quad (3.7)$$

handelt, wobei der Vektor  $\mathbf{e} = [e_1, \dots, e_N]^T$  die Fehler zwischen dem  $k$ -ten gemessenen Wert  $y(k)$  und dem  $k$ -ten Schätzwert  $\hat{y}(k)$

$$e_k = y(k) - \hat{y}(k) \quad (3.8)$$

zusammenfaßt. Aus Gl. (3.7) ergibt sich mit der Produktregel für den Gradienten

$$\nabla V = \frac{\partial \mathbf{e}^T}{\partial \boldsymbol{\theta}} \mathbf{e} + \mathbf{e}^T \frac{\partial \mathbf{e}}{\partial \boldsymbol{\theta}} = 2\mathbf{J}^T \mathbf{e} \quad , \quad (3.9)$$

wobei die Matrix der Ableitungen aller  $N$  Komponenten nach allen  $n$  Parametern

$$\mathbf{J} := \frac{\partial \mathbf{e}}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial e_1}{\partial \theta_1} & \cdots & \frac{\partial e_1}{\partial \theta_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial e_N}{\partial \theta_1} & \cdots & \frac{\partial e_N}{\partial \theta_n} \end{bmatrix} \quad (3.10)$$

als Jacobimatrix  $\mathbf{J}$  bezeichnet wird. Die Differentiation von Gl. (3.9) führt auf folgende Darstellung der Hessematrix:

$$\mathbf{H} = 2\mathbf{J}^T \frac{\partial \mathbf{e}}{\partial \boldsymbol{\theta}} + 2 \frac{\partial \mathbf{J}^T}{\partial \boldsymbol{\theta}} \mathbf{e} = 2\mathbf{J}^T \mathbf{J} + 2\nabla \mathbf{J}^T \mathbf{e} \quad . \quad (3.11)$$

In der exakten Formulierung der Hessematrix gemäß Gl. (3.11) ist der zweite Term in der Nähe eines Minimums vernachlässigbar, da die meisten Zielfunktionen an dieser Stelle eine quadratische Form besitzen (Hoffmann und Hofmann 1971). Mit der auf diese Weise approximierten Hessematrix

$$\mathbf{H} \approx 2\mathbf{J}^T \mathbf{J} \quad (3.12)$$

und dem Gradienten in Gl. (3.9) folgt aus Gl. (3.6) die Iterationsvorschrift des Gauß–Newton–Verfahrens:

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - [\mathbf{J}_i^T \mathbf{J}_i]^{-1} \mathbf{J}_i^T \mathbf{e}_i \quad . \quad (3.13)$$

Die Parameterkorrekturen des Gauß–Newton–Verfahrens sind betragsmäßig groß und gewährleisten i. allg. einen schnellen Optimierungsfortschritt. Bei zu großen Abweichungen  $\boldsymbol{\theta}_i - \boldsymbol{\theta}^*$  von der Lösung  $\boldsymbol{\theta}^*$  trifft die vereinfachende Annahme, die zu der Darstellung der Hessematrix in Gl. (3.12) führt, nicht mehr zu, und aus dem vernachlässigten Term kann ein Divergieren des Verfahrens resultieren. Außerdem sind bei der Matrix  $\mathbf{J}^T \mathbf{J}$  im Falle einer überparametrisierung Singularitäten möglich, die das Verfahren scheitern lassen können (Ljung 1987).

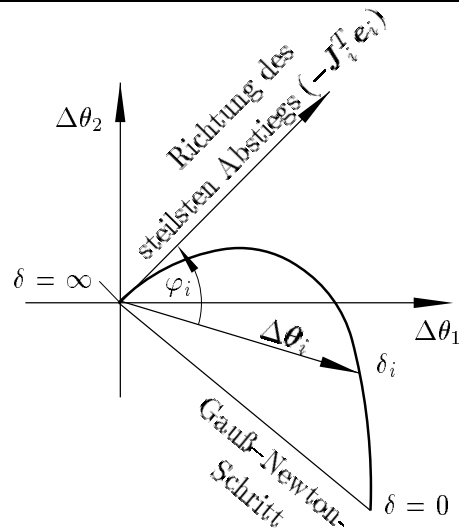
Levenberg und Marquardt beheben diese beiden Schwierigkeiten durch eine Modifikation des Gauß–Newton–Verfahrens, bei der sich Singularitäten durch eine Regularisierung der in Gl. (3.12) approximierten Hessematrix

$$\tilde{\mathbf{H}} = \mathbf{H} + \delta \mathbf{I} \quad , \quad (3.14)$$

d. h. eine Vergrößerung der Hauptdiagonalelemente um einen positiven Skalar  $\delta$ , vermeiden lassen. Bild 3.1 illustriert den Einfluß der Regularisierung auf Betrag und Richtung der Parameterkorrektur: Mit zunehmenden  $\delta$  nähert sich die Suchrichtung zunehmend der Suchrichtung des steilsten Abstiegs (das ist der negative Gradient) und gleichzeitig vermindert sich der Betrag der Parameterkorrektur. Zur Berechnung des Regularisierungsfaktors  $\delta$  verfolgen Levenberg und Marquardt unterschiedliche Strategien, die im folgenden erläutert werden.

## 3.2 Levenberg–Verfahren

Levenberg betrachtet die Bestimmung des Regularisierungsfaktors  $\delta$  wiederum als ein-dimensionales Optimierungsproblem bei konstantem  $\boldsymbol{\theta}_i$  und  $\mathbf{J}_i$ . Die Berechnung von  $\delta$



**Bild 3.1:** Einfluß des Regularisierungsfaktors auf Betrag und Richtung der Parameterkorrektur

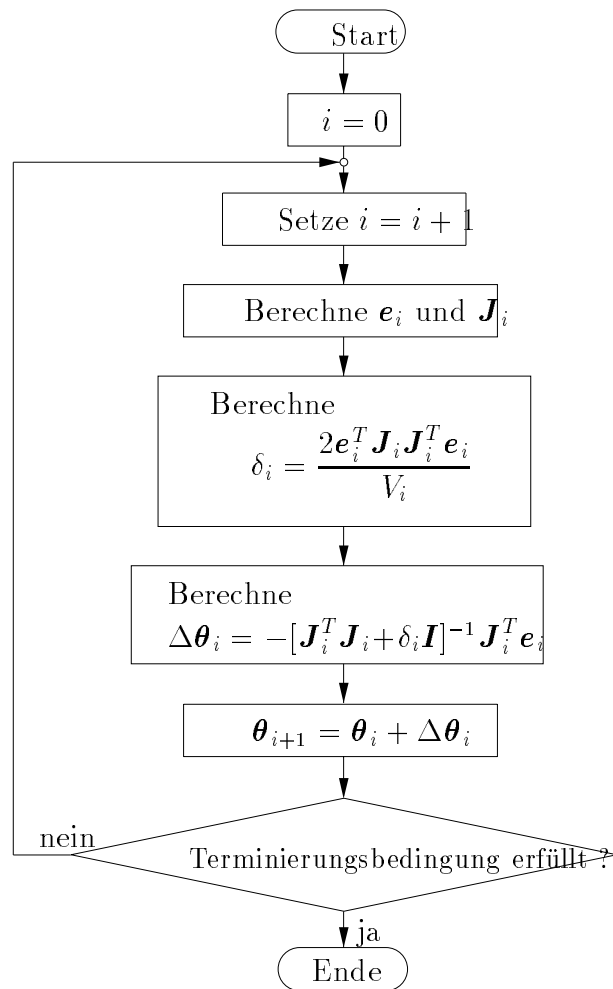
erfolgt in der Weise, daß sich in jedem Iterationsschritt die größtmögliche Funktionswertabnahme ergibt. Da aber  $\delta$  nicht explizit als Funktionsparameter auftritt, schätzt Levenberg den Wert von  $\alpha := 1/\delta$  durch Formulierung einer Tangentengleichung an der Stelle  $\alpha = 0$  und durch Bestimmung des Schnittpunktes mit der Abszisse ab. Die dafür benötigte Tangentensteigung erhält er durch Bildung des Differentialquotienten  $dV_{i-1}/d\alpha$  und den Grenzübergang  $\alpha \rightarrow 0$ . Das Ergebnis dieser Abschätzung, deren genaue mathematische Herleitung in (Hoffmann und Hofmann 1971) zu finden ist, ist die Bestimmungsgleichung für den Regularisierungsfaktor:

$$\delta_i = \alpha_i^{-1} = \frac{2\mathbf{e}_i^T \mathbf{J}_i \mathbf{J}_i^T \mathbf{e}_i}{V_i} \quad . \quad (3.15)$$

Der Fall  $V_i = 0$  bereitet bei der Berechnung des Regularisierungsfaktors nach Gl. (3.15) numerische Probleme. Allerdings ist dies nur möglich, wenn  $\mathbf{e}_i \equiv \mathbf{0}$  gilt, das Modell den Identifikationsdatensatz also exakt prädiziert. Dann ist das Optimierungsziel erreicht und es sind keine weitere Iterationen und Berechnungen des Regularisierungsfaktors erforderlich, so daß sich eine weitere mathematische Untersuchung dieses Falls erübrigt. Unter Berücksichtigung von Gl. (3.15) gibt Bild 3.2 den Algorithmus des Levenberg-Verfahrens in einem übersichtlichen Flußdiagramm an.

### 3.3 Marquardt-Verfahren

Das Marquardt-Verfahren zielt bei der Berechnung von  $\delta$  im Gegensatz zum Levenberg-Verfahren nicht auf eine größtmögliche Funktionswertabnahme, sondern auf eine maximale Parameterkorrektur in jedem Iterationsschritt ab. Bei der Berechnung der Parameterkorrektur soll jedoch beachtet werden, daß in jedem Iterationsschritt der neue Parametervektor zu einer Funktionswertabnahme führt. Diese wichtige Eigenschaft des Marquardt-



**Bild 3.2:** Algorithmus des Levenberg–Verfahrens nach Hoffmann und Hofmann (1971)

Verfahrens wird als Stabilität bezeichnet und in (Entenmann 1976) folgendermaßen definiert:

**Definition 3.1:** Ein Optimierungsverfahren heißt stabil, wenn bei einer Minimierungsaufgabe in jedem Iterationsschritt

$$V(\boldsymbol{\theta}_{i+1}) < V(\boldsymbol{\theta}_i) \quad \text{mit} \quad \boldsymbol{\theta}_{i+1} \neq \boldsymbol{\theta}_i \quad (3.16)$$

gilt, die Funktionswerte mit zunehmender Iterationenfolge also streng monoton abnehmen.  $\square$

Die maximale Parameterkorrektur entspricht dem Gauß–Newton–Schritt mit  $\delta = 0$  (vgl. Bild 3.1). Marquardt beginnt im ersten Iterationsschritt deshalb mit einem sehr kleinen Wert für  $\delta$ , z. B.  $\delta = 10^{-2}$ , und überprüft, ob die Stabilitätsbedingung  $V(\boldsymbol{\theta}_{i+1}) < V(\boldsymbol{\theta}_i)$  erfüllt ist. Im Fall einer Funktionswertabnahme wird der Regularisierungsfaktor verkleinert, um im nächsten Iterationsschritt eine größere Parameterkorrektur zuzulassen (vgl.



Bild 3.3). Bei  $V(\boldsymbol{\theta}_{i+1}) > V(\boldsymbol{\theta}_i)$  wird  $\delta$  dagegen solange mit einem Faktor  $\zeta$  (z. B.  $\zeta = 10$ ) multipliziert, bis die Stabilitätsbedingung erfüllt ist, und erst dann der Iterationsschritt abgeschlossen. Allerdings begrenzt Marquardt den Wert von  $\delta$  in der Weise, daß der Winkel  $\varphi_i = \angle(\Delta\boldsymbol{\theta}_i, -\mathbf{J}_i^T \mathbf{e}_i)$  zwischen der momentanen Korrektur  $\Delta\boldsymbol{\theta}_i$  und dem negativen Gradienten den Wert  $\varphi_0 = 45^\circ$  nicht unterschreitet. Damit soll eine zu starke Annäherung an die Suchrichtung des steilsten Abstiegs vermieden werden, da diese aufgrund der Vernachlässigung der Zielfunktionskrümmung ungünstiger ist und mehr Iterationen erfordert. Ist auch bei diesem  $\varphi$  bzw.  $\delta$  keine Funktionswertabnahme zu verzeichnen, erfolgt eine Dämpfung der Parameterkorrektur durch Multiplikation mit einem Faktor  $\gamma \leq 1$

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i + \gamma \Delta\boldsymbol{\theta}_i \quad , \quad (3.17)$$

der ausgehend von  $\gamma = 1$  solange iterativ halbiert wird, wie Instabilität vorliegt. Darüber hinaus erachtet Marquardt es für zweckmäßig, den Parameterraum in geeigneter Weise zu normieren, mit dem Ziel, langgestreckte Konturen der Zielfunktion möglichst zu Kugeln zu transformieren und günstigere Suchrichtungen zu erhalten (Hoffmann und Hofmann 1971). Zusammenfassend ist der Marquardt-Algorithmus in Bild 3.3 dargestellt.

### 3.4 Hybridverfahren

Eine Besonderheit funktionaler Fuzzy-Modelle bilden die unterschiedlichen Parameter-typen: Während die Modellausgangsgröße in nichtlinearer Weise von den Prototypen  $\mathbf{v}_i$  abhängt, gehen die Konklusionspolynomparameter linear in die Prädiktion ein, wenn es sich um lineare Konklusionspolynome gemäß Gl. (2.4) handelt und die übrigen Parameter konstant gehalten werden. Wie in der II. Stufe des Identifikationsverfahrens lassen sich dann die Konklusionspolynomparameter direkt und global optimal (bezogen auf die übrigen Modellparameter) mit dem LS-Verfahren berechnen. Diese Eigenschaft läßt sich bei der Optimierung durch Verwendung eines zweistufigen Verfahrens nutzen. Ein derartiges Hybridverfahren (Kroll 1995b) führt in einer ersten Stufe zunächst einen Iterationsschritt eines nichtlinearen Optimierungsverfahrens lediglich für die nichtlinearen Parameter aus. Darauf folgt die Berechnung der verbleibenden linearen Parameter mit dem LS-Verfahren (vgl. Bild 3.4).

Engmann (1995) zeigt, daß sich bereits der Optimierungfortschritt des Verfahrens des steilsten Abstiegs durch eine Kombination mit dem LS-Verfahren deutlich (ca. um den Faktor 1000) steigern läßt. Durch Verwendung eines der zuvor beschriebenen Verfahren von Levenberg oder Marquardt läßt sich eine weitere Effizienzsteigerung erzielen (Bernd 1996, Nelles 1995).

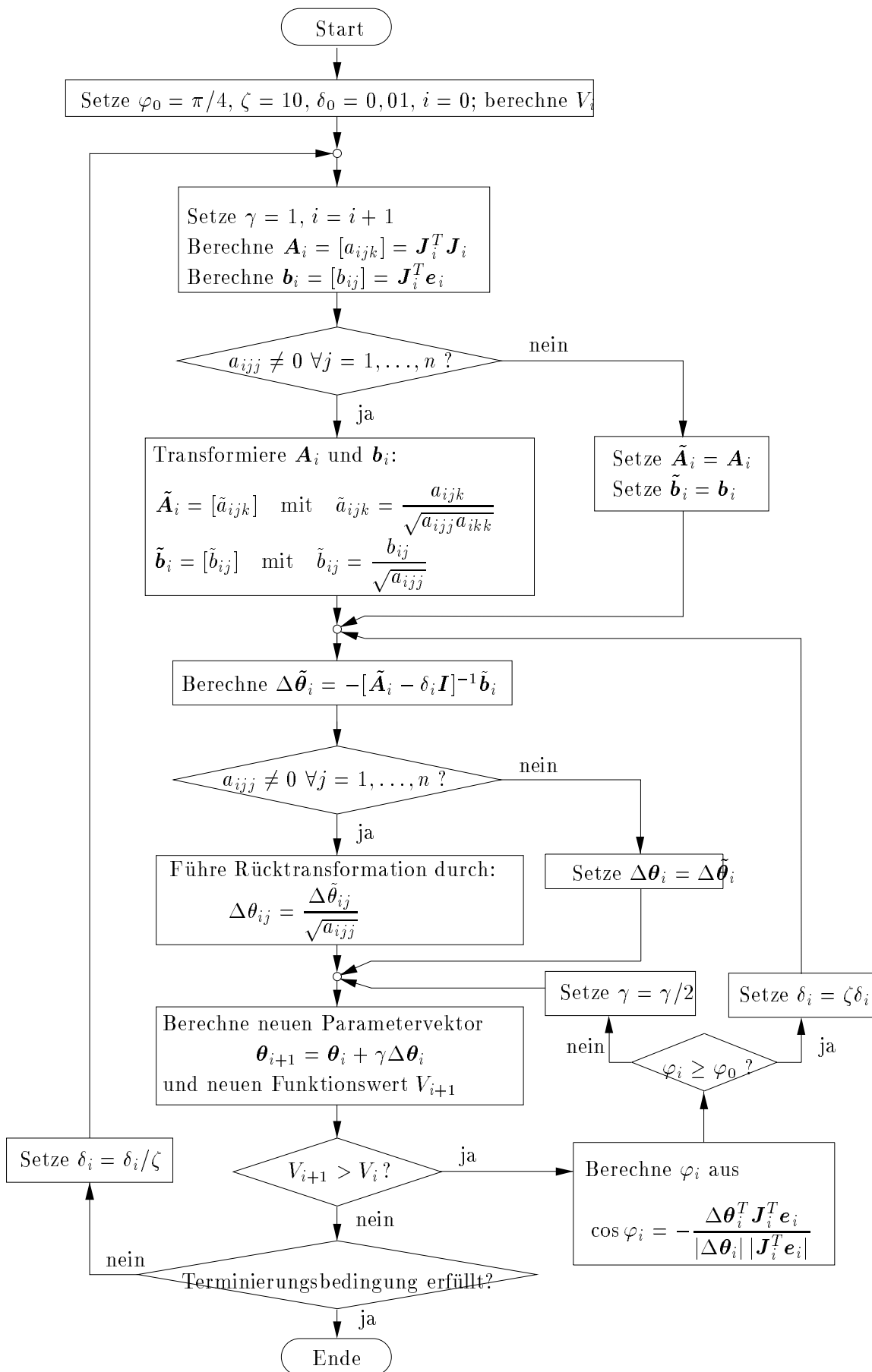


Bild 3.3: Algorithmus des Marquardt-Verfahrens nach Hoffmann und Hofmann (1971)

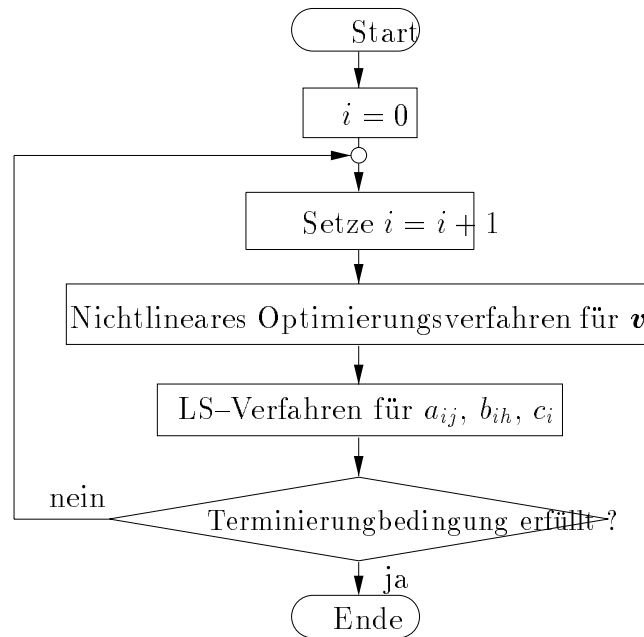


Bild 3.4: Hybridverfahren

### 3.5 Dimensionsreduzierende Formulierung der Verfahren

Die Verfahren von Levenberg und Marquardt benötigen die Jacobimatrix zur Berechnung des Gradienten und zur Approximation der Hessematrix. Diese ist i. allg. von hoher Dimension, woraus sich bei der Umsetzung in einem Rechnerprogramm Speicherprobleme ergeben können. Zur Verdeutlichung dieser Problematik wird der Speicherbedarf für die Jacobimatrix bei der Optimierung eines funktionalen Fuzzy-Modells mit einer mittleren Parameteranzahl angegeben: Als Beispiel dient ein Modell mit  $c = 4$  Prototypen und 4 Eingangsgrößen. Jeder Prototyp besitzt somit 4 Koordinaten und jedes Konklusionspolynom 5 weitere Parameter, so daß insgesamt  $n = 4(4 + 5) = 36$  Parameter zu optimieren sind. Bei einer Identifikation mit  $N = 2000$  Eingangsvektoren besitzt die Jacobimatrix die Dimension  $36 \times 2000$ , wobei jedes Matrixelement (bei doppelter Rechengenauigkeit) 8 Byte belegt. Folglich beträgt der Speicherbedarf der Jacobimatrix hier  $nN \cdot 8\text{Byte} = 576\text{ kByte}$ .

Bei äquivalenten Darstellungen der Verfahren von Levenberg und Marquardt, bei denen die Jacobimatrix selbst nicht benötigt wird, reduziert sich der Speicherbedarf deutlich. Der Gradient läßt sich stattdessen direkt durch

$$\frac{1}{2}\nabla V = \mathbf{J}^T \mathbf{e} = \begin{bmatrix} \frac{\partial e_1}{\partial \theta_1} & \cdots & \frac{\partial e_N}{\partial \theta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial e_1}{\partial \theta_n} & \cdots & \frac{\partial e_N}{\partial \theta_n} \end{bmatrix} \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix} \quad (3.18)$$

$$= \begin{bmatrix} \frac{\partial e_1}{\partial \theta_1} \\ \vdots \\ \frac{\partial e_1}{\partial \theta_n} \end{bmatrix} e_1 + \dots + \begin{bmatrix} \frac{\partial e_N}{\partial \theta_1} \\ \vdots \\ \frac{\partial e_N}{\partial \theta_n} \end{bmatrix} e_N \quad (3.19)$$

$$= \sum_{k=1}^N \nabla e_k e_k \quad (3.20)$$

berechnen und die Hessematrix (vgl. Gl. (3.12)) mit

$$\frac{1}{2} \mathbf{H} \approx \mathbf{J}^T \mathbf{J} = \begin{bmatrix} \frac{\partial e_1}{\partial \theta_1} & \dots & \frac{\partial e_N}{\partial \theta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial e_1}{\partial \theta_n} & \dots & \frac{\partial e_N}{\partial \theta_n} \end{bmatrix} \begin{bmatrix} \frac{\partial e_1}{\partial \theta_1} & \dots & \frac{\partial e_1}{\partial \theta_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial e_N}{\partial \theta_1} & \dots & \frac{\partial e_N}{\partial \theta_n} \end{bmatrix} \quad (3.21)$$

$$= \sum_{k=1}^N \begin{bmatrix} \frac{\partial e_k}{\partial \theta_1} \\ \vdots \\ \frac{\partial e_k}{\partial \theta_n} \end{bmatrix} \begin{bmatrix} \frac{\partial e_k}{\partial \theta_1} & \dots & \frac{\partial e_k}{\partial \theta_n} \end{bmatrix} \quad (3.22)$$

$$= \sum_{k=1}^N \nabla e_k \nabla e_k^T \quad (3.23)$$

approximieren. Bei Verwendung der Gln. (3.20) und (3.23) reduziert sich folglich die Anzahl der zu speichernden Elemente von  $nN$  auf  $N$ , da lediglich der Fehlergradient  $\nabla e_k$  zwischenspeichern ist. In o. g. Beispiel belegen die Hilfsgrößen folglich nur  $N \cdot 8$  Byte = 16 kByte anstatt 576 kByte.

## 4 Optimierung von funktionalen Fuzzy-Modellen

Die in Abschnitt 3 vorgestellten Optimierungsverfahren werden im folgenden bzgl. ihres Konvergenzverhaltens bei der Optimierung eines funktionalen Fuzzy-Modells untersucht. Ein pneumatischer Translationsantrieb stellt dabei das Testsystem dar, das nachfolgend beschrieben wird. Für dieses System erfolgt anschließend die Identifikation und Optimierung eines Modells mit den beschriebenen Verfahren.

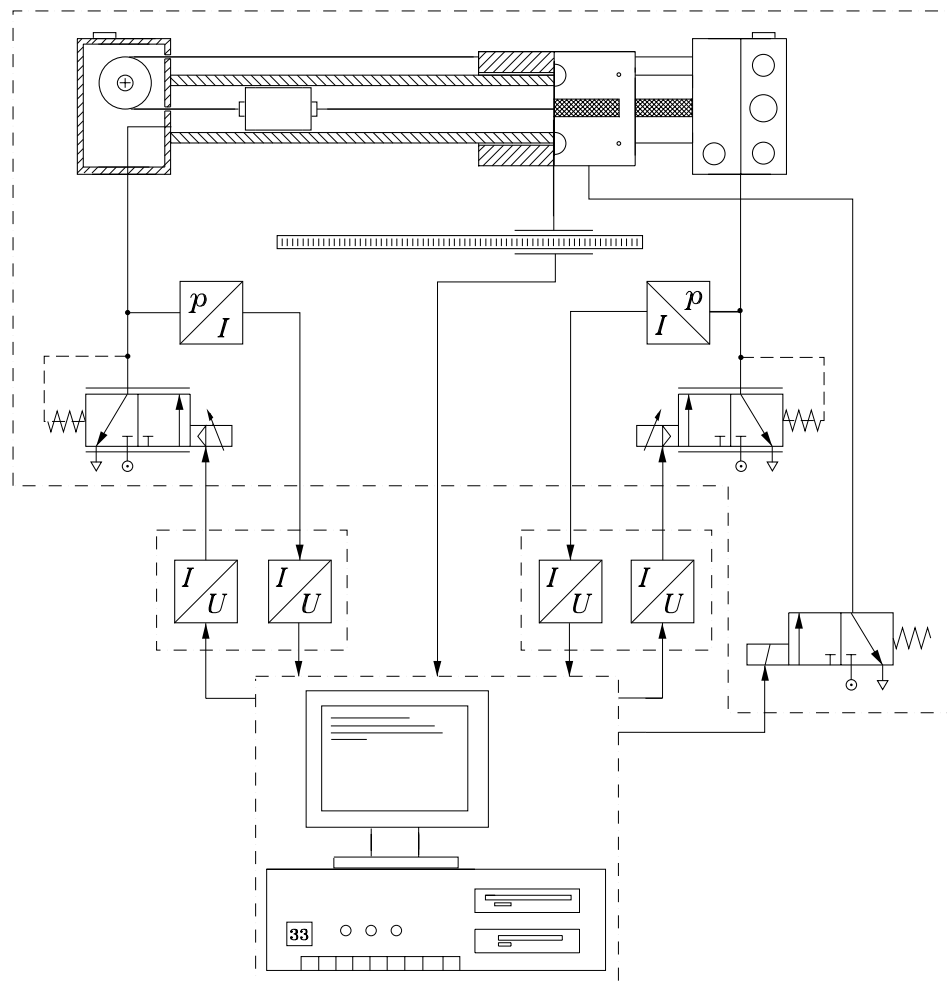
### 4.1 Testsystem: Pneumatischer Translationsantrieb

Gegenstand der folgenden Untersuchungen ist der Prüfstand eines pneumatischen Translationsantriebs. Es handelt sich um einen kolbenstangenlosen Pneumatikzylinder (Bandzylinder LA-85 der Firma Mannesmann Rexroth) mit einem Nennhub von 800 mm. Die Positionierung des Kolbens erfolgt mit Hilfe eines elektrischen Schaltkreises, der zwei Druckservoventile ansteuert. Diese dienen zur Beeinflussung der Drücke in den beiden Zylinderkammern (vgl. Bild 4.1). Als Stellgrößen stehen somit die beiden Steuerströme  $i_1$  und  $i_2$  der Servoventile zur Verfügung. Die inkrementell meßbare Kolbenposition  $y$  stellt die Ausgangsgröße dar. Eine detaillierte Beschreibung des Pneumatikprüfstandes gibt (Hebisch 1995).

Die Modellidentifikation erfolgt mit dem in Bild 4.2a dargestellten Pseudoranschsignal für den Steuerstrom  $i_1$ . Der zweite Steuerstrom wird dabei konstant gehalten. Es handelt sich um eine Folge von Zufallszahlen, deren Werte jeweils über eine Dauer von  $T_i = 0,2$  s konstant gehalten werden. Das mit einer Abtastzeit von  $T_0 = 5$  ms über eine Dauer von  $T = 3$  s aufgezeichnete Ausgangssignal zeigt Bild 4.2b.

### 4.2 Optimierungsergebnisse

Die folgenden Untersuchungen beziehen sich auf ein funktionales Fuzzy-Modell, das mit dem in Bild 4.2 dargestellten und auf die Intervalle  $y \in [-1; 1]$  sowie  $i_1 \in [-1; 1]$  normierten Datensatz (s. Anhang A) identifiziert wurde. Das Modell besitzt  $c = 4$  Regeln, die Eingangsgrößen  $y(k-1)$ ,  $y(k-2)$ ,  $i_1(k-1)$  sowie  $i_1(k-2)$  und benutzt die Mahalanobisnorm sowie  $\nu = 1,6$ . Bei den Konklusionsfunktionen handelt es sich um lineare Polynome gemäß Gl. (2.4) mit  $m_y = m_u = 2$ . Die Berechnung der Prototypen mit dem FCM in der I. Stufe erfordert die Vorgabe von Startwerten, die aus einer MATLAB-Funktion erhalten werden. Diese Funktion liefert bei identischem Argument  $s$ , dem sog. Samen, identische Folgen von Zufallszahlen, so daß reproduzierbare Ergebnisse gewährleistet sind. Hier wird der Wert des Samens auf  $s = 5$  gesetzt. Darauf folgt in der II. Stufe, wie in Abschnitt 2 beschrieben, die Berechnung der Konklusionspolynomparameter mit dem LS-Verfahren. Die nun vorliegenden Parameterwerte dienen als Startwerte für die Optimierung in der III. Stufe.

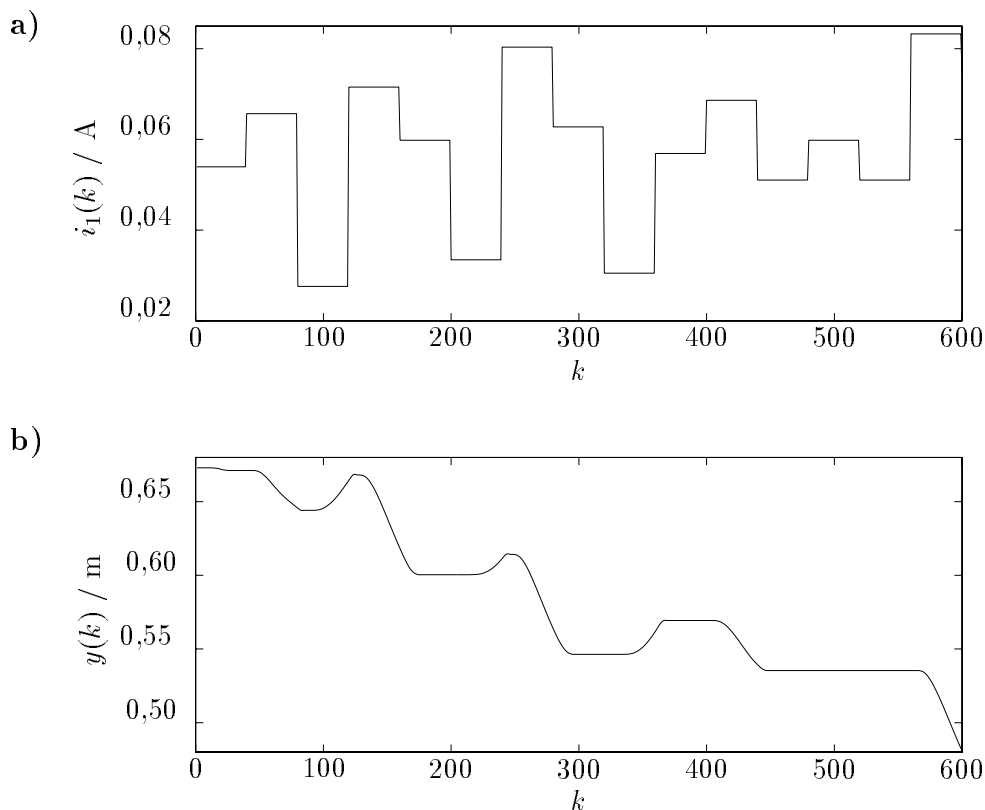


**Bild 4.1:** Vereinfachter Geräteplan des pneumatischen Translationsantriebs

### Konvergenzvergleich der Verfahren von Levenberg und Marquardt

Am Beispiel des zuvor beschriebenen Modells soll nun ein Konvergenzvergleich der Verfahren von Levenberg und Marquardt erfolgen. Weitere gradientenbasierte Verfahren wie jenes von Gauß-Newton und das Verfahren des steilsten Abstiegs werden hierbei nicht mit einbezogen, da sie den Verfahren von Levenberg und Marquardt deutlich unterlegen sind (Bernd 1996). Vielmehr sind die Charakteristika der beiden Verfahren, die sich aus den unterschiedlichen Strategien zur Berechnung des Regularisierungsfaktors ergeben, von Interesse.

Zunächst wird die Optimierung der Prototypen und Konklusionspolynomparameter bzgl. der prädiktiven Modellauswertung betrachtet, wobei die Hybridverfahren aus dem Levenberg- bzw. Marquardt- und dem LS-Verfahren zur Anwendung kommen sollen. Bild 4.3 illustriert die Entwicklung der auf die Startwerte normierten quadratischen Fehler  $V_{\text{präd},i}/V_{\text{präd},0}$ , die sich mit dem Levenberg/LS- und dem Marquardt/LS-Verfahren er-

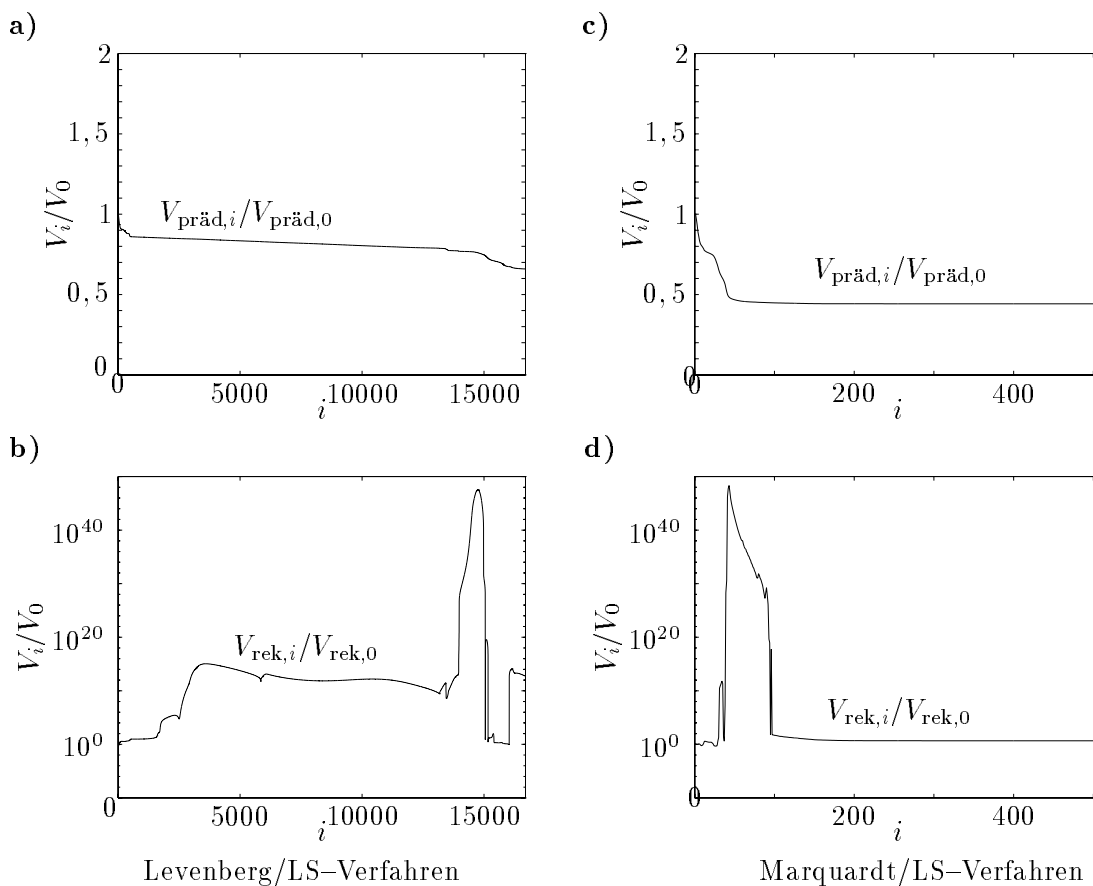


**Bild 4.2:** Eingangssignal a) und gemessenes Ausgangssignal b)

geben. Außerdem ist jeweils der entsprechende Verlauf bei rekursiver Auswertung eingetragen.

In Bild 4.3 fällt zunächst der deutlich schnellere Optimierungsfortschritt und das gute Konvergenzverhalten des Marquardt/LS-Verfahrens im Vergleich zum Levenberg/LS-Verfahren auf. Daraus wird deutlich, daß die von Levenberg angewandte Minimumabschätzung zur Berechnung des Regularisierungsfaktors mit einem großen Fehler behaftet ist und infolge dessen die angestrebte große Funktionswertabnahme nicht erreicht wird. Andererseits führt eine größtmögliche Parameterkorrektur auch zu einer schnellen Funktionswertabnahme, wie der Verlauf des Marquardt/LS-Verfahrens verdeutlicht.

Die Parameterverläufe der beiden Verfahren gestatten eine genauere Analyse. In den Bildern 4.4 bis 4.6 sind die Parameterverläufe, die sich beim Levenberg/LS- und Marquardt/LS-Verfahren ergeben, gegenübergestellt. Dabei sind jeweils gleichartige Parameter aller 4 Regeln zusammengefaßt. Der Vergleich der Parameterverläufe der beiden Verfahren illustriert den in Abschnitt 3 erläuterten Einfluß des Regularisierungsfaktors: Die unterschiedlichen Strategien der beiden Verfahren führen sowohl qualitativ als auch quantitativ zu sehr verschiedenen Parameterverläufen. Insbesondere beim Marquardt/LS-Verfahren sind die angestrebten großen Parameterkorrekturen zu erkennen, die das Verfahren schnell konvergieren lassen. Beim Levenberg/LS-Verfahren wird dagegen deutlich, daß eine Stagnation der Optimierung nicht unbedingt mit geringen Parameteränderungen korreliert,

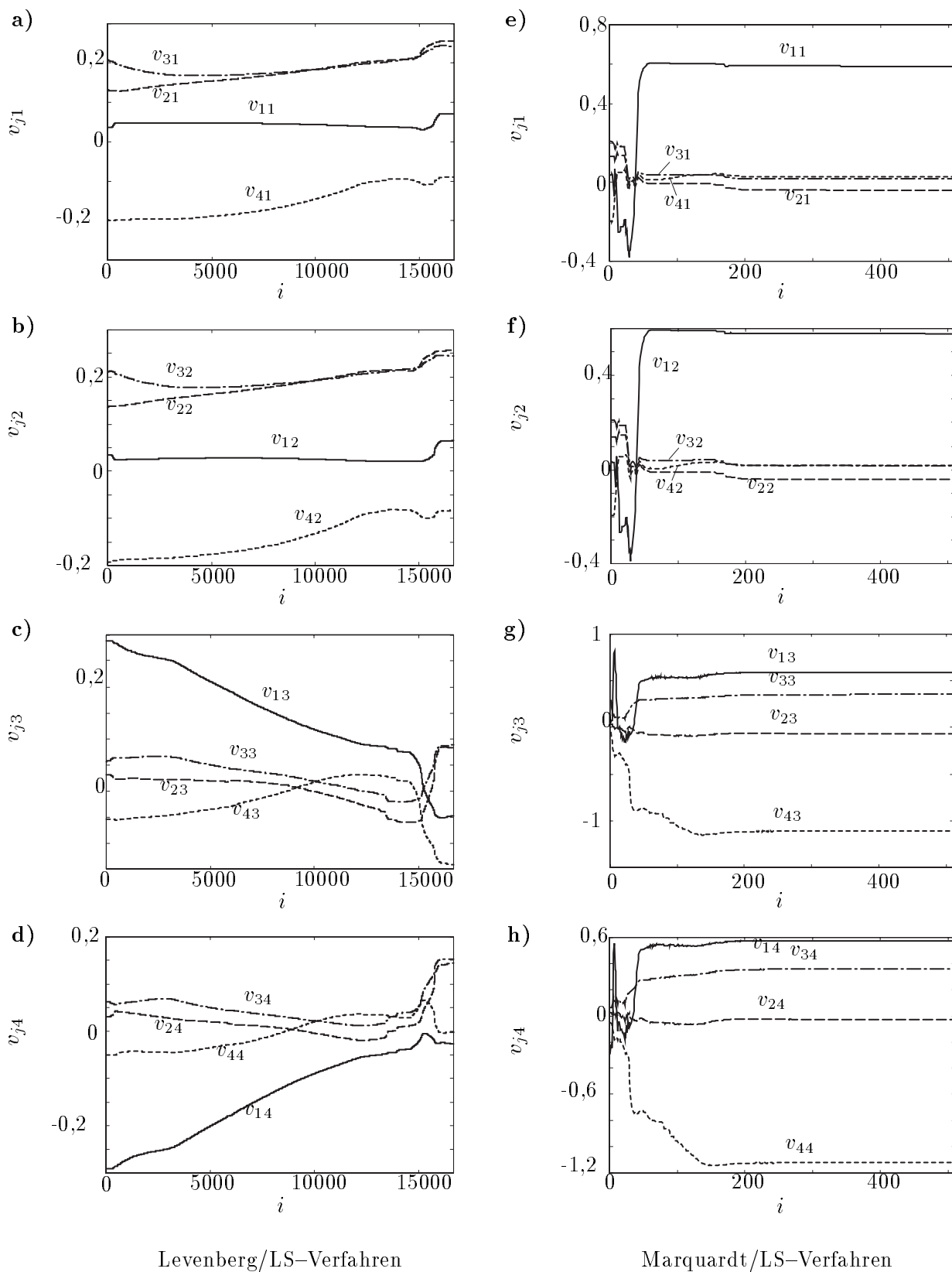


**Bild 4.3:** Optimierung der prädiktiven Modellauswertung

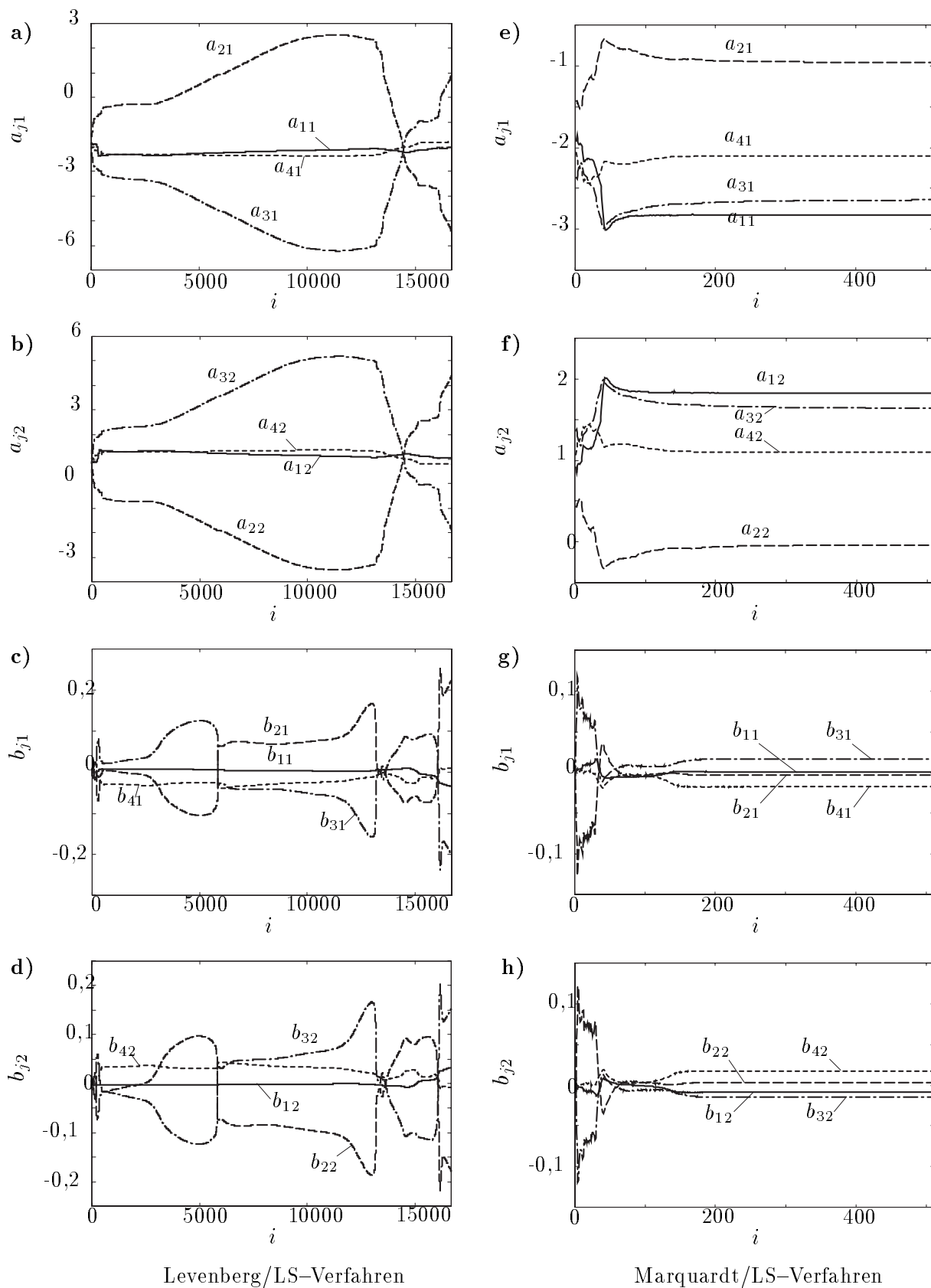
wie die Parameterverläufe (insbesondere der Komponenten  $v_{13}$  und  $v_{14}$ , vgl. Bilder 4.4c und 4.4d) illustrieren. Die Feststellung der Konvergenz sollte sich demnach nicht am Funktionswert, sondern an den Parameterkorrekturen orientieren.

Allerdings ist dabei ein, in mehreren Simulationen der Optimierung verschiedener Fuzzy-Modelle und Verwendung verschiedener Datensätze aufgetretener, Sonderfall zu beachten. Dieser besteht im Konvergieren zweier Prototypen gegen dieselben Werte. Hier tritt dieser Fall bei den Prototypen  $\mathbf{v}_2$  und  $\mathbf{v}_3$  auf. Nach ca. 10000 Iterationen gilt  $v_{21} \approx v_{31}$  (vgl. Bild 4.4a),  $v_{22} \approx v_{32}$  (vgl. Bild 4.4b) und auch die dritten und vierten Komponenten der beiden Prototypen unterscheiden sich nur gering (vgl. Bilder 4.4c und 4.4d).





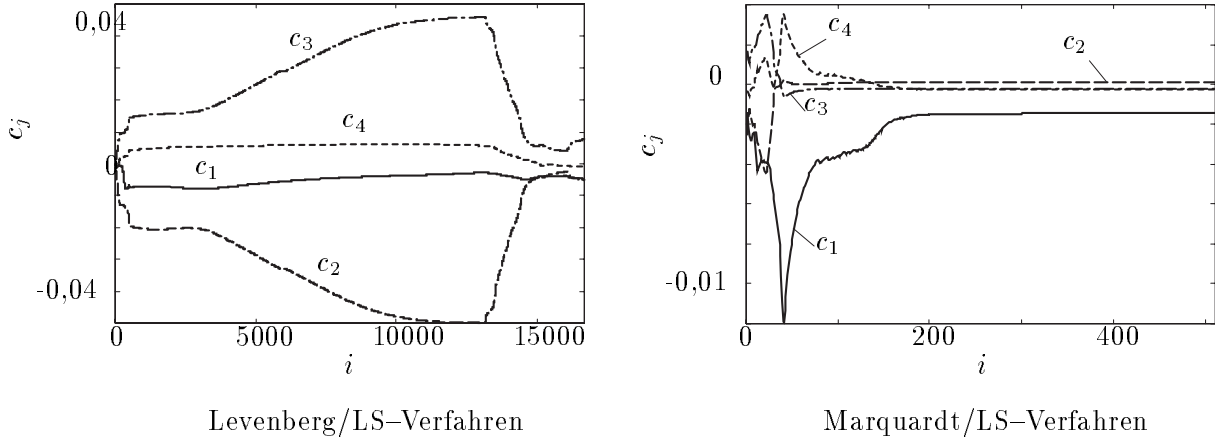
**Bild 4.4:** Parameterverläufe der Prototypenkoordinaten  $v_{jh}$  bei prädiktiver Optimierung (Index  $j$ : Prototypennummer, Index  $h$ : Komponentennummer)



Levenberg/LS-Verfahren

Marquardt/LS-Verfahren

**Bild 4.5:** Parameterverläufe der Konklusionspolynomparameter bei prädiktiver Optimierung (Index  $j$ : Nummer des Konklusionspolynoms)



**Bild 4.6:** Parameterverläufe bei prädiktiver Optimierung

Das Konvergieren der Koordinaten zweier Prototypen gegen dieselben Werte hat zur Folge, daß die überwiegende Zahl der Eingangsvektoren zu beiden Prototypen die gleiche Zugehörigkeit besitzt, mit Ausnahme weniger, sich in unmittelbarer Nähe der Prototypen befindlicher Eingangsvektoren. Wegen  $\mu_2(\mathbf{x}) \approx \mu_3(\mathbf{x})$  gilt hier

$$y \approx \frac{\mu_1(\mathbf{x})y_1(\mathbf{x}) + \mu_2(\mathbf{x})[y_2(\mathbf{x}) + y_3(\mathbf{x})] + \mu_4(\mathbf{x})y_4(\mathbf{x})}{\sum_{i=1}^c \mu_i(\mathbf{x})} \quad (4.1)$$

$$\approx \left[ \sum_{i=1}^4 \mu_i(\mathbf{x}) \right]^{-1} \left[ \mu_1(\mathbf{x})y_1(\mathbf{x}) + \mu_2(\mathbf{x})[c_2 + c_3 - (a_{21} + a_{31})y(k-1) + \right. \quad (4.2) \\ \left. - (a_{22} + a_{32})y(k-2) + (b_{21} + b_{31})i_1(k-1) + \right. \\ \left. + (b_{22} + b_{32})i_1(k-2)] + \mu_4(\mathbf{x})y_4(\mathbf{x}) \right] .$$

Die einzelnen Werte der Konklusionspolynomparameter der Regeln 2 und 3 besitzen somit keinen Einfluß auf die Ausgangsgröße, sondern nur die in Gl. (4.2) auftretenden Parametersummen. Damit erklären sich die symmetrischen Verläufe dieser Parameter und die fortlaufenden Parameterkorrekturen. Eine Konvergenzaussage ist folglich nur nach Überprüfung der Prototypenlage und ggf. der Entwicklung der zugehörigen Summen der Konklusionspolynomparameter möglich.

Ergänzend ist zu erwähnen, daß in dem beschriebenen Fall des Konvergierens zweier Prototypen gegen denselben Wert i. allg. eine geringere Modellgüte zu erwarten ist, da sich die Anzahl der signifikanten Parameter reduziert.

## Bewertung der Verfahren

Eine abschließende Bewertung der beiden Verfahren führt zur Bevorzugung des Marquardt-Verfahrens aufgrund der großen Parameterkorrekturen und der schnellen Konvergenz. Aus praxisorientierter Sicht ist dieses Verfahren auch insofern interessant, als daß

sich bereits mit wenigen Iterationen eine deutliche Reduzierung des mittleren quadratischen Fehlers erzielen läßt, auch ohne das Konvergieren des Verfahrens abzuwarten.

Die großen Parameterkorrekturen des Marquardt-Verfahrens, die bisher als Vorteil bewertet wurden, erweisen sich jedoch in einigen Simulationen als nachteilig, wie sich anhand der Optimierung eines Gasofenmodells feststellen läßt (Bernd 1996). Es zeigt sich, daß das Marquardt-Verfahren häufig zu einer nicht sinnvollen Anordnung der Prototypen führt. Bei einer Optimierung, die zusätzlich die Normparameter einbezieht, verschärft sich die Problematik, da die großen Parameterkorrekturen häufig bereits im ersten Iterationsschritt zu einer nicht positiv definiten Formenmatrix führen. In diesen Fällen ist die Anwendung des Levenberg-Verfahrens aufgrund geringerer Parameterkorrekturen zu empfehlen. Wie die Ergebnisse von Bernd (1996) demonstrieren, ist auch dieses Verfahren effizienter als das Verfahren des steilsten Abstiegs und unterliegt in den meisten Fällen dem Marquardt-Verfahren nicht in der in diesem Abschnitt vorliegenden Deutlichkeit.

### Vergleich der prädiktiven und rekursiven Optimierung

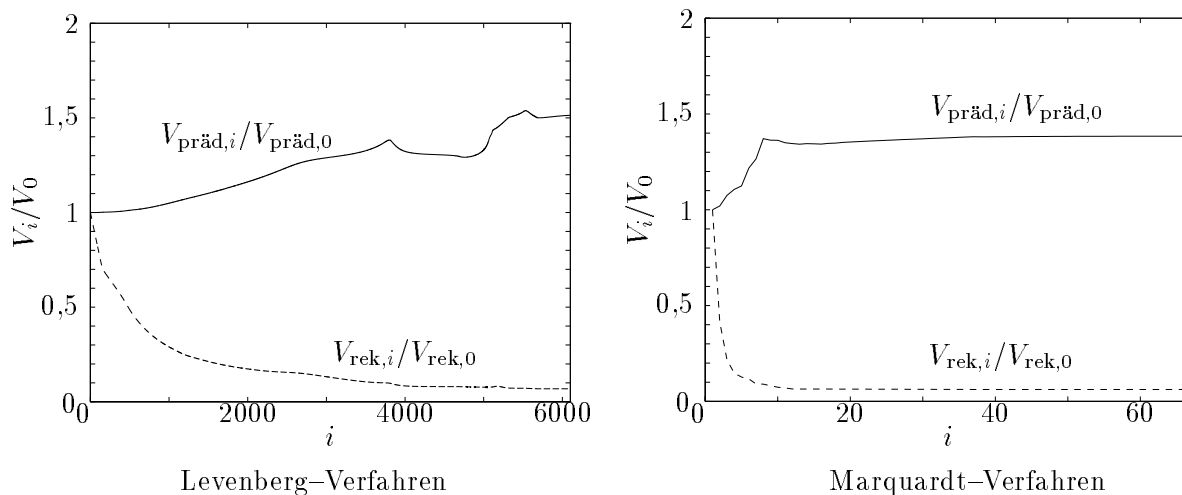
Die bisherige Optimierung bezieht sich auf die prädiktive Modellauswertung. Die Annahme, daß eine prädiktive Optimierung auch zu einer Fehlerreduzierung bei rekursiver Auswertung führt, läßt sich nicht bestätigen, wie Bild 4.3 verdeutlicht. In den Bildern 4.3b und 4.3d sind die Fehler bei rekursiver Auswertung eingezeichnet, die sich während der prädiktiven Optimierung ergeben. Es ist zu erkennen, daß bereits nach wenigen Iterationsschritten der Wert von  $V_{\text{rek}}$  den Bereich sinnvoller Werte verläßt. Im folgenden wird untersucht, ob sich ähnliche Beobachtungen im umgekehrten Fall, d. h. bei Optimierung der rekursiven und gleichzeitiger prädiktiver Auswertung, verzeichnen lassen. Zu diesem Zweck wird das zuvor beschriebene und nach der II. Identifikationsstufe erhaltene Modell mit dem Levenberg- und dem Marquardt-Verfahren bzgl. der rekursiven Auswertung optimiert. Das Simulationsergebnis ist in Bild 4.7 dargestellt.

Aus Bild 4.7 ist ersichtlich, daß sich auch bei rekursiver Optimierung eine Verschlechterung des prädiktiven Modellfehlers ergibt. Allerdings ist hierbei die Zunahme des quadratischen Fehlers  $V_{\text{präd}}$  relativ gering und nur bei numerischer Betrachtungsweise, nicht aber in einer Verschlechterung des Zeitreihenverlaufs erkennbar, wie Bild 4.8 illustriert.

Dieses Ergebnis läßt sich mit einem unterschiedlichen Parametereinfluß bei den beiden Prädiktionsarten begründen: Eine prädiktive Auswertung berechnet die Ausgangsgröße lediglich einen Abtastschritt aus den bekannten Meßwerten voraus und ist somit relativ unempfindlich gegenüber Parameteränderungen. Bei rekursiver Auswertung dagegen wird der Fehler zum Modelleingang zurückgeführt und pflanzt sich auf diese Weise fort. Eine geringe Parameteränderung kann infolgedessen bereits zu einer großen Änderung des Prädiktionsfehlers beitragen. Dies läßt sich wiederum mit Hilfe der Parameterverläufe bestätigen.

Bild 4.9 zeigt exemplarisch die Verläufe von  $v_{j3}$  und  $b_{j1}$  bei Optimierung mit dem Mar-

quardt-Verfahren. Die übrigen Verläufe weisen geringere und überwiegend in der graphischen Darstellung nicht erkennbare Parameteränderungen auf, so daß auf weitere Darstellungen verzichtet wird.

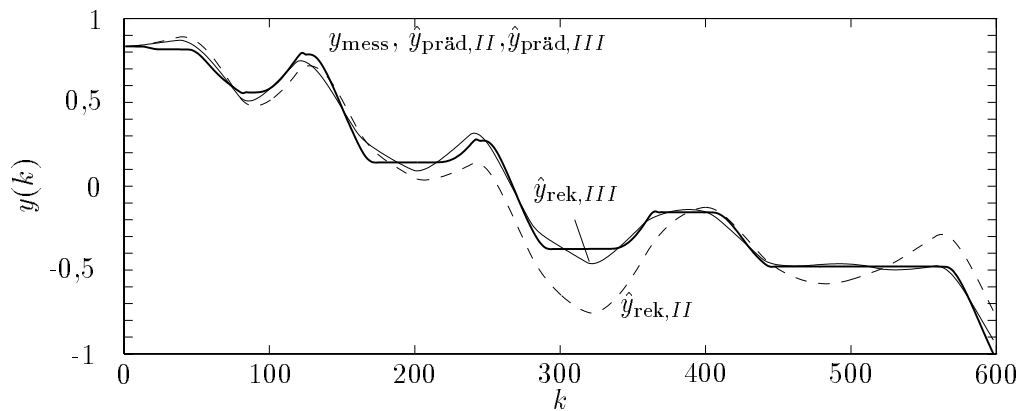


**Bild 4.7:** Optimierung der rekursiven Modellauswertung

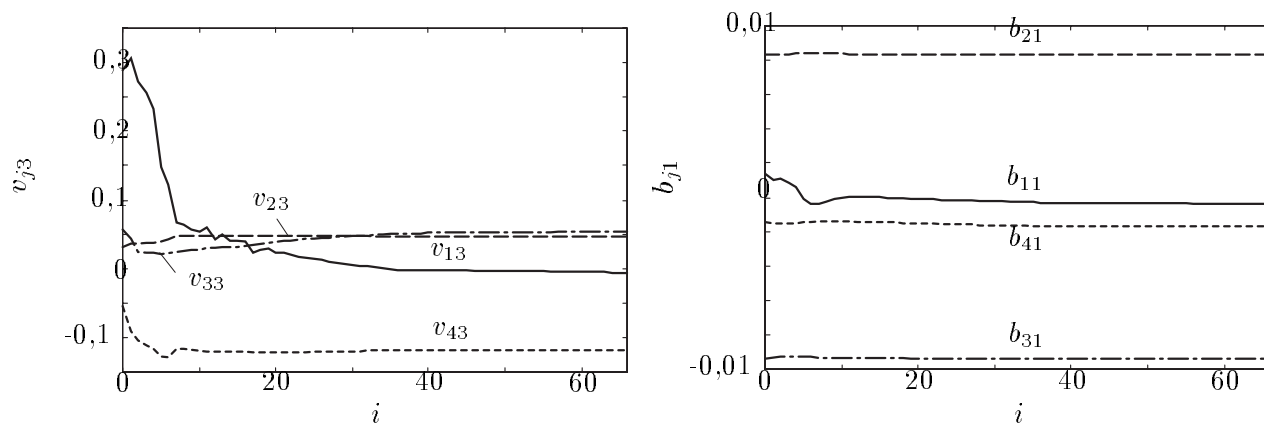
Aus den gegenläufigen Tendenzen der Prädiktionsfehler bei der Optimierung der verschiedenen Prädiktionsarten läßt sich ableiten, daß die zu optimierende Prädiktionsart vorab festzulegen ist. Andererseits ist die Verschlechterung des prädiktiven Fehlers bei rekursiver Optimierung relativ gering im Vergleich zu der deutlichen Verbesserung der rekursiven Auswertung. Eine Verschlechterung der prädiktiven Güte ist zudem nur anhand des numerischen Wertes von  $V_{\text{präd}}$ , nicht aber in der graphischen Darstellung der Zeitreihe erkennbar. Die Verbesserung der rekursiven Auswertung spiegelt sich dagegen sowohl in numerischer als auch graphischer Darstellung wider (vgl. Bild 4.8). Demzufolge ist eine rekursive Optimierung vorzuziehen, wenn nicht explizit eine optimale prädiktive Auswertung angestrebt wird.

### 4.3 Initialisierungsabhängigkeit der Modellgüten

Wie die Ergebnisse von Bernd (1996) zeigen, ist das Ergebnis der Clusterung in der I. Stufe und infolgedessen auch der nachfolgenden Identifikationsstufen stark initialisierungsabhängig. Daraus läßt sich auf eine hohe Multimodalität der Zielfunktion schließen. Dies hat zur Folge, daß verschiedene Initialisierungen zu unterschiedlichen Parameterwerten und Modellgüten führen. Im folgenden wird die bisherige Vorgehensweise und eine Modifikation des Identifikationsverfahrens erläutert, um diesen Nachteil abzuschwächen.



**Bild 4.8:** Gemessene und prädizierte Zeitreihen nach der II. und der III. Stufe bei rekursiver Optimierung; die Zeitreihen bei prädiktiver Auswertung sind in dieser Darstellung vor und nach der Optimierung deckungsgleich mit der Meßreihe



**Bild 4.9:** Parameterverläufe bei rekursiver Optimierung (Marquardt-Verfahren)

### Bisherige Vorgehensweise

Bisher wurde eine mehrfache Durchführung der ersten beiden Identifikationsstufen mit unterschiedlichen Initialisierungen und der Selektion des Modells mit dem geringsten Fehler praktiziert, um das Problem der Initialisierungsabhängigkeit abzuschwächen (Engmann 1995). Diese Vorgehensweise weist jedoch folgende Nachteile auf:

- Die mehrfache Durchführung der ersten beiden Stufen für verschiedene Kombinationen von  $s$  und  $\nu$  ist relativ aufwendig.
- Die Selektion des Modells mit dem geringsten Fehler nach der II. Stufe stellt nicht in jedem Fall einen guten Startwert für eine Optimierung dar.

Während der erstgenannte Nachteil offensichtlich ist, wird die Bedeutung des zweiten Punktes dadurch deutlich, daß Kroll u. a. (1995) mit dem in Bild 4.2 dargestellten Datensatz ein funktionales Fuzzy-Modell mit dem zuvor beschriebenen Modellansatz ( $c = 4$ ,  $m_y = m_u = 2$ , Mahalanobisnorm) für  $\nu = 1,895$  und  $s = 5$  identifizierten, das sich bereits ohne Optimierung durch eine hohe Güte auszeichnet. Dieses Modell weist nach der II. Stufe die sehr geringen quadratischen Fehler  $V_{\text{präd}} = 3,5156 \cdot 10^{-9} \text{ m}^2$  und  $V_{\text{rek}} = 4,5878 \cdot 10^{-5} \text{ m}^2$  auf und übertrifft damit die Güte des in Abschnitt 4.2 vorgestellten Modells nach der II. Stufe deutlich. Die Annahme, daß dieses Modell auch nach einer Optimierung einen geringeren Prädiktionsfehler aufweist, läßt sich jedoch nicht bestätigen. Vielmehr terminiert ein Optimierungsverfahren sofort, da ein lokales Minimum vorliegt, wie auch die Untersuchung der Parametrierung ergibt. Im Gegensatz dazu läßt sich die Güte des in Abschnitt 4.2 beschriebenen Modells ( $c = 4$ ,  $\nu = 1,6$ ,  $s = 5$ ), trotz schlechterer Werte nach der II. Stufe, durch eine Optimierung beachtlich steigern (vgl. Tabelle 4.1). Die mehrfache Durchführung nur der ersten beiden Stufen reicht folglich nicht aus, um ein optimales, initialisierungsbereinigtes Modell zu erhalten. Stattdessen ist jeweils die Optimierung mit einzubeziehen, wodurch sich der Aufwand jedoch deutlich erhöht.

Der beschriebene Nachteil ist besonders dann schwerwiegend, wenn sich bereits mit einem optimierten linearen Modell ( $c = 1$ ) eine höhere Güte erzielen läßt als mit einem optimierten nichtlinearen Modell. Mit dem nichtlinearen Fuzzy-Modellansatz ist nämlich auch stets ein dem linearen äquivalentes Modell bei entsprechender Parametrierung realisierbar. Daraus ergibt sich die Forderung, daß das Optimierungsergebnis aufgrund der höheren Parameteranzahl erst dann zufriedenstellend ist, wenn mindestens die Güte des optimierten linearen Modells erreicht wird.

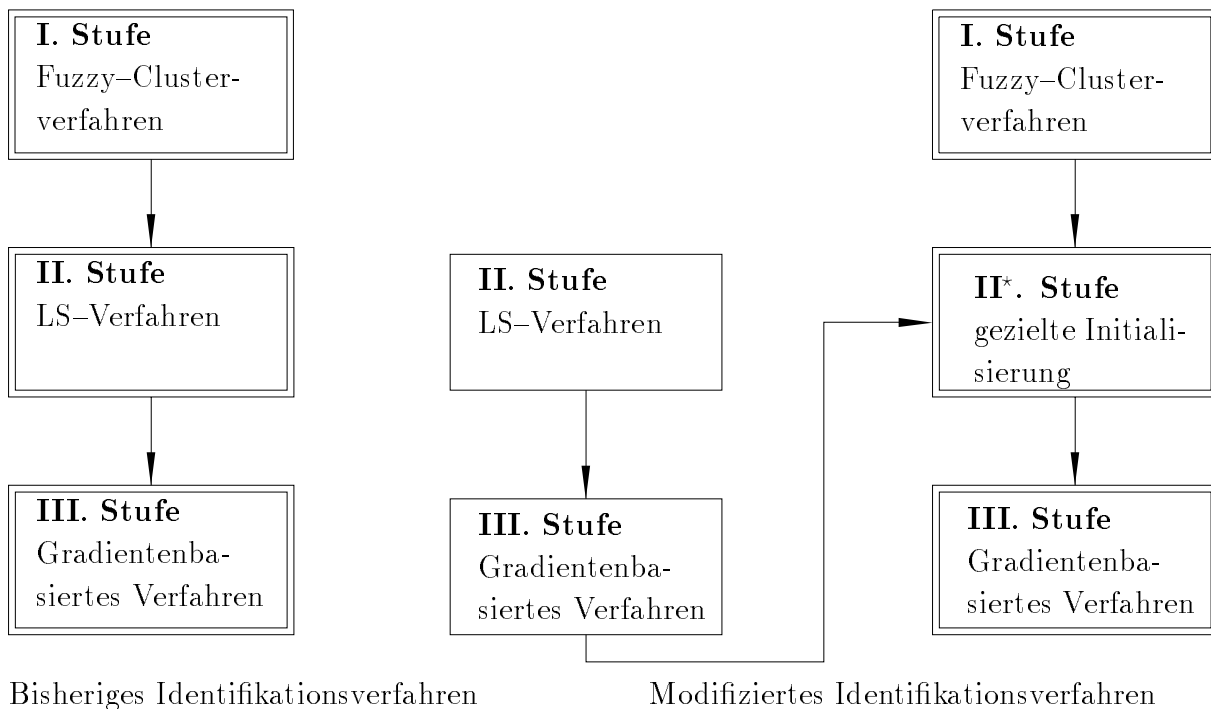
Modell	nach der II. Stufe		nach der III. Stufe	
	$\frac{V_{\text{präd}}}{10^{-9} \text{ m}^2}$	$\frac{V_{\text{rek}}}{10^{-5} \text{ m}^2}$	$\frac{V_{\text{präd}}}{10^{-9} \text{ m}^2}$	$\frac{V_{\text{rek}}}{10^{-5} \text{ m}^2}$
$c = 1$	4,4544	16,8955	30,2056	3,5202
$c = 4$ , $\nu = 1,6$ , $s = 5$	3,9294	21,6955	5,4384	1,3430
$c = 4$ , $\nu = 1,895$ , $s = 5$	3,5156	4,5878	3,5156	4,5878

**Tabelle 4.1:** Vergleich der quadratischen Fehler zweier nichtlinearer funktionaler Fuzzy-Modelle mit einem linearen Modell bei rekursiver Optimierung

### Modifiziertes Verfahren

Auf der Grundlage dieser Forderung läßt sich das Identifikationsverfahren in der Weise modifizieren, daß die Parametrierung des optimierten linearen Modells die Initialisierung für ein nichtlineares Modell darstellt, dessen Güte durch eine nachfolgende Optimierung weiter gesteigert werden soll. Diese Modifikation führt auf folgendes Identifikationsverfahren:

Den Ausgangspunkt stellt das lineare Modell ( $c = 1$ ) dar, das bereits durch Berechnung der Parameterwerte der Konklusionspolynome vollständig bestimmt ist (vgl. Bild 4.10). Diese Berechnung erfolgt (wie bisher) in der II. Stufe mit Hilfe des LS-Verfahrens, aus dem die global optimalen Parameterwerte bzgl. der prädiktiven Modellauswertung hervorgehen. Die erhaltenen Parameterwerte dienen als Startwerte für eine rekursive Optimierung (III. Stufe). Das auf diese Weise erhaltene lineare Modell stellt im weiteren das Referenzmodell dar, dessen Güte mit einem nichtlinearen Modellansatz gesteigert werden soll. Durch identische Initialisierung aller Konklusionspolynomparameter des nichtlinearen Modells mit den zuvor berechneten Parametern des linearen Modells werden exakt die gleichen Zeitreihen prädiziert, unabhängig von der Lage der Prototypen und dem Wert des Unschärfeparameters  $\nu$ . In dem modifizierten Verfahren wird die II. Stufe im folgenden mit „ $\star$ “ gekennzeichnet, um die Initialisierung der Konklusionspolynomparameter von der Berechnung mit dem LS-Verfahren zu unterscheiden. Um sinnvolle Startwerte für die Prototypen zu erhalten, erfolgt deren Berechnung für das nichtlineare Modell wiederum in der I. Stufe mit einem Fuzzy-Clusterverfahren. Daran schließt sich ebenfalls unverändert eine (sinnvollerweise rekursive) Optimierung des Modells an, wobei die Parameter der Partialmodelle (vgl. Bild 4.1) und die Prototypen berechnet werden. Das Ergebnis dieser III. Stufe ist ein Modell höherer oder zumindest gleicher Güte, verglichen mit der Güte des linearen Modells. Die erzielbaren Ergebnisse sollen im folgenden anhand eines Beispiel demonstriert werden.



**Bild 4.10:** Vergleich des bisherigen und des modifizierten Identifikationsverfahrens



## Beispiel

Als Beispiel dient der in Bild 4.2 dargestellte und auf das Intervall  $[-1; 1]$  normierte Datensatz, für den ein funktionales Fuzzy-Modell mit dem modifizierten Identifikationsalgorithmus identifiziert werden soll. Zwecks Vergleichbarkeit mit den bisherigen Ergebnissen wird ein Konklusionspolynom gemäß Gl. (2.4) mit  $m_y = m_u = 2$  gewählt. Nach Anwendung des LS-Verfahrens ergeben sich für das lineare Modell die Fehler

$$V_{\text{präd}} = 4,4544 \cdot 10^{-9} \text{ m}^2 \quad \text{und} \quad V_{\text{rek}} = 1,6896 \cdot 10^{-4} \text{ m}^2 \quad .$$

Durch eine rekursive Optimierung mit dem Marquardt-Verfahren läßt sich der rekursive Prädiktionsfehler auf

$$(V_{\text{präd}} = 3,0206 \cdot 10^{-8} \text{ m}^2) \quad V_{\text{rek}} = 3,5202 \cdot 10^{-5} \text{ m}^2$$

reduzieren. Nun erfolgt der Übergang auf das nichtlineare Fuzzy-Modell: Wiederum wird  $c = 4$ ,  $\nu = 1,6$  sowie  $s = 5$  gewählt und die Prototypen mit Hilfe des FCM berechnet. Nach identischer Initialisierung aller Konklusionspolynomparameter mit den Werten des optimierten linearen Modells resultieren zunächst erneut die Fehler

$$V_{\text{präd}} = 3,0206 \cdot 10^{-8} \text{ m}^2 \quad \text{und} \quad V_{\text{rek}} = 3,5202 \cdot 10^{-5} \text{ m}^2 \quad .$$

Die Optimierung des nichtlinearen Modells wird hier mit dem Levenberg-Verfahren durchgeführt, das als Ergebnis

$$V_{\text{präd}} = 2,8552 \cdot 10^{-8} \text{ m}^2 \quad \text{und} \quad V_{\text{rek}} = 1,6585 \cdot 10^{-5} \text{ m}^2$$

liefert. Das Marquardt-Verfahren führt hier zu einem schlechteren Ergebnis.  $\square$

Wie obiges Beispiel verdeutlicht, läßt sich mit diesem modifizierten Identifikationsverfahren mit geringem Aufwand und systematischer Vorgehensweise ein Modell hoher Güte identifizieren. Die Initialisierung der Prototypen und des Unschärfeparameters beeinflusst zwar weiterhin das Optimierungsergebnis, jedoch in geringerem Maße als zuvor.

Als Nachteil ist dagegen zu nennen, daß auch das auf diese Weise erhaltene Ergebnis wiederum ein lokales Minimum darstellt. Beispielsweise ergibt sich mit dem in Abschnitt 4.3 beschriebenen und optimierten Modell ein geringerer Fehler, wobei die Differenz zu dem in obigen Beispiel beschriebenen Modell gering ist (vgl. Tabelle 4.2). Prinzipiell ist zwar die Bestimmung des globalen Minimums anzustreben, jedoch ist dies nur unter hohem Aufwand, z. B. mit einem stochastischen Verfahren möglich.

Modell	nach der II. Stufe		nach der III. Stufe		optimierte Prädiktionsart, Verfahren
	$\frac{V_{\text{präd}}}{10^{-9} \text{m}^2}$	$\frac{V_{\text{rek}}}{10^{-5} \text{m}^2}$	$\frac{V_{\text{präd}}}{10^{-9} \text{m}^2}$	$\frac{V_{\text{rek}}}{10^{-5} \text{m}^2}$	
$c = 1$	4,4544	16,8955	30,2056	3,5202	rek., Marquardt
$c = 4, \nu = 1, 6, s = 5$	3,9294	21,6955	5,4384	1,3430	rek., Marquardt
$c = 4, \nu = 1, 6, s = 5$	3,9294	21,6955	5,9485	1,4887	rek., Levenberg
$c = 4, \nu = 1, 6, s = 5$	3,9294	21,6955	1,7381	98,1108	präd., Marq./LS
$c = 4, \nu = 1, 6, s = 5$	3,9294	21,6955	2,5890	$1,2 \cdot 10^{14}$	präd., Lev./LS
$c = 4, \nu = 1, 895, s = 5$	3,5156	4,5878	3,5156	4,5878	rek., Marquardt
	nach der II*. Stufe		nach der III. Stufe		
$c = 4, \nu = 1, 6, s = 5$	30,2056	3,5202	28,5518	1,6585	rek., Levenberg

**Tabelle 4.2:** Mittlere quadratische Fehler einiger funktionaler Fuzzy-Modelle

## 5 Zusammenfassung und Ausblick

Der vorliegende Bericht befaßt sich mit zwei effizienten Verfahren zur Optimierung von funktionalen Fuzzy-Modellen. Abschnitt 2 gibt zunächst eine kurze Einführung in die Theorie funktionaler Fuzzy-Systeme und die Identifikation der darauf basierenden funktionalen Fuzzy-Modelle. In diesem Zusammenhang werden die unterschiedlichen Zielfunktionen in den ersten beiden Identifikationsstufen und die Beschränkung des Identifikationsverfahrens auf die Minimierung des prädiktiven Modellfehlers als Motivation für eine Optimierung erläutert. Basierend auf den Ergebnissen von Bernd (1996) folgt eine ausführliche Beschreibung der effizientesten gradientenbasierten Verfahren von Levenberg und Marquardt sowie einer Modifikation des Hybridverfahrens (Kroll 1995b). Das Konvergenzverhalten der Verfahren von Levenberg und Marquardt, die sich nur in den Strategien zur Berechnung eines Algorithmusparameters (Regularisierungsfaktors) unterscheiden, wird anhand eines Modells für einen pneumatischen Translationsantrieb in Abschnitt 4 untersucht. Ausführliche Parameterstudien bestätigen die beim Marquardt-Verfahren in theoretischen Überlegungen angestrebten großen Parameterkorrekturen. Aufgrund der daraus resultierenden schnellen Konvergenz ist dieses Verfahren i. allg. dem Levenberg- und anderen gradientenbasierten Verfahren vorzuziehen. Lediglich, wenn sich die großen Parameterkorrekturen als problematisch erweisen (z. B. bei Optimierung der Normparameter), sollte das Levenberg-Verfahren zur Anwendung kommen.

Ein weiterer Aspekt war ein Vergleich der Optimierung der prädiktiven und der rekursiven Auswertung. Hierbei war eine gegenläufige Tendenz dieser beiden Prädiktionsarten zu erkennen. Die günstigere Alternative besteht in der Optimierung der rekursiven Auswertung, die zu einer beachtlichen Reduzierung des rekursiven und nur geringen Verschlechterung des prädiktiven Fehlers führt.

Schließlich wurde die Problematik der Initialisierungsabhängigkeit des Identifikationsverfahrens diskutiert und eine Modifikation vorgeschlagen. Diese besteht in der Optimierung zunächst eines linearen Modells. Die erhaltenen Parameter dienen als Initialisierung eines nichtlinearen Modells, das wiederum einer Optimierung unterzogen wird. Es wurde gezeigt, daß sich mit dieser systematischen Vorgehensweise Modelle von hoher Güte identifizieren lassen.

Im Rahmen weiterer Untersuchungen ist die Bestimmung des globalen Minimums von Interesse. Hierzu eignen sich stochastische Verfahren, wie z. B. genetische Algorithmen. Wie erste Erfahrungen mit genetischen Algorithmen zeigen (Bernd 1995), ist aufgrund des hohen numerischen Aufwands allerdings keine praktikable Alternative zu dem bisherigen bzw. modifizierten Identifikationsverfahren zu erwarten. Stattdessen sollten die potentiellen Approximationseigenschaften funktionaler Fuzzy-Modelle im Vordergrund stehen.

Wie bereits erwähnt, ist das Ziel der Optimierung (Minimierung des prädiktiven oder rekursiven Fehlers) vorab festzulegen, da die beiden Prädiktionsfehler unterschiedliche Tendenzen aufweisen. Die Untersuchung eines Gütekriteriums, das die quadratischen

Fehler beider Prädiktionsarten miteinander kombiniert (z. B. additiv), kann dagegen zu einer alternativen Zielfunktion führen. Durch eine Optimierung bzgl. dieser Zielfunktion soll eine Reduzierung sowohl des prädiktiven als auch des rekursiven Prädiktionsfehlers erreicht werden.

Ein weiterer zu untersuchender Aspekt ist die Validierung optimierter Modelle. Im Zusammenhang mit einer prädiktiven Optimierung erläutert Kroll (1995b) die Notwendigkeit einer vorzeitigen Terminierung zur Vermeidung einer Überanpassung. Inwiefern dies auch bei Optimierung der rekursiven Auswertung zutrifft, ist noch zu klären.

## 6 Literatur

- Bernd, T.** 1995. *Konstruktive Änderung und Fuzzy-Regelung eines Drehschwingerprüfstandes*. Studienarbeit MSRT. Universität Duisburg.
- Bernd, T.** 1996. *Zur optimalen Fuzzy-Modellierung eines fluidischen Antriebes*. Diplomarbeit MSRT. Universität Duisburg.
- Bezdek, J. C.** 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, London: Plenum Press.
- Engmann, B.** 1995. *Identifikation einer analytischen Klasse von Funktional-Fuzzy-Modellen*. Diplomarbeit MSRT. Universität Duisburg.
- Entenmann, W.** 1976. *Optimierungsverfahren*. Heidelberg: Hüthig.
- Foulds, L.** 1981. *Optimization techniques*. New York: Springer.
- Hebisch, H.** 1995. *Modellierung eines pneumatischen Zylinderantriebes*. Forschungsbericht 7/95 MSRT. Universität Duisburg.
- Hoffmann, U.** und **H. Hofmann.** 1971. *Einführung in die Optimierung*. Weinheim/Bergstr.: Verlag Chemie.
- Isermann, R.** 1992. *Identifikation dynamischer Systeme Bd. I*. Berlin, Heidelberg: Springer.
- Kroll, A.** 1995a. Partition identification of fuzzy models using objective function clustering algorithms. *IEEE Int. Conf. on Systems, Man and Cybernetics*. Vancouver/Kanada, 22.–25.10.1995. 7–12.
- Kroll, A.** 1995b. *Zur analytischen Beschreibung von Fuzzy-Systemen*. Forschungsbericht 4/95 MSRT. Universität Duisburg.
- Kroll, A., H. Reuter** und **H. Hebisch.** 1995. *Zur experimentellen Fuzzy-Modellbildung eines pneumatischen Translationsantriebes*. Forschungsbericht 13/95 MSRT. Universität Duisburg.
- Ljung, L.** 1987. *System identification – theory for the user*. Englewood Cliffs/New Jersey: Prentice-Hall.
- Nelles, O.** 1995. On training radial basis function networks as series-parallel and parallel models for identification of nonlinear dynamic systems. *IEEE International Conference on Systems, Man and Cybernetics*. Vancouver/Kanada, 22.–25.10.95. 4609–4614.

- Schwarz, H.** 1991. *Nichtlineare Regelungssysteme — Systemtheoretische Grundlagen*. München, Wien: Oldenbourg.
- Takagi, T.** und **M. Sugeno.** 1985. Fuzzy identifikation of systems and its applications to modeling and control.. *IEEE Transactions on Systems, Man and Cybernetics SMC-15*. 116–132.

## A Parameter des identifizierten Modells

Aufgrund der guten Approximationseigenschaften (vgl. Bild 4.1) des in Abschnitt 4.2 vorgestellten Modells bei rekursiver Auswertung werden im folgenden die genauen Parameterwerte angegeben, die als Grundlage für weitergehende Untersuchung dienen sollen.

Die Identifikation erfolgt mit dem in Bild 4.2 dargestellten Meßwertverlauf und mit dem durch

$$y = \frac{\tilde{y} - \bar{y}}{|\tilde{y} - \bar{y}|_{\max}} \quad (\text{A.1})$$

sowie

$$i_1 = \frac{\tilde{i}_1 - \bar{i}_1}{|\tilde{i}_1 - \bar{i}_1|_{\max}} \quad (\text{A.2})$$

normierten Datensatz, wobei  $\bar{y} = 0,5856 \text{ m}$ ,  $|\tilde{y} - \bar{y}|_{\max} = 0,1048 \text{ m}$ ,  $\bar{i}_i = 0,0571 \text{ A}$  und  $|\tilde{i}_i - \bar{i}_i|_{\max} = 0,0295 \text{ A}$  betragen. Ausgehend von  $\nu = 1,6$  und  $s = 5$  lauten die Fuzzy-Regeln nach rekursiver Optimierung mit dem Marquardt-Verfahren:

Regel 1:

$$\begin{aligned} & \text{WENN} \begin{bmatrix} y(k-1) \\ y(k-2) \\ i_1(k-1) \\ i_1(k-2) \end{bmatrix} \text{ IST} \begin{bmatrix} 4,167760 \cdot 10^{-2} \\ 2,939014 \cdot 10^{-2} \\ -5,594833 \cdot 10^{-3} \\ 6,759849 \cdot 10^{-3} \end{bmatrix}, \quad (\text{A.3}) \\ & \text{DANN } y_1 = -2,889624 \cdot 10^{-3} + 1,919269y(k-1) - 9,206054 \cdot 10^{-1}y(k-2) \\ & \quad - 3,695335 \cdot 10^{-4}i_1(k-1) + 5,641984 \cdot 10^{-4}i_1(k-2) \end{aligned}$$

Regel 2:

$$\begin{aligned} & \text{WENN} \begin{bmatrix} y(k-1) \\ y(k-2) \\ i_1(k-1) \\ i_1(k-2) \end{bmatrix} \text{ IST} \begin{bmatrix} 1,313575 \cdot 10^{-1} \\ 1,357422 \cdot 10^{-1} \\ 4,675474 \cdot 10^{-2} \\ 2,104797 \cdot 10^{-2} \end{bmatrix}, \quad (\text{A.4}) \\ & \text{DANN } y_2 = -2,244746 \cdot 10^{-3} + 1,429872y(k-1) - 4,313245 \cdot 10^{-1}y(k-2) \\ & \quad + 8,339424 \cdot 10^{-3}i_1(k-1) - 1,218730 \cdot 10^{-2}i_1(k-2) \end{aligned}$$

Regel 3:

$$\begin{aligned} & \text{WENN} \begin{bmatrix} y(k-1) \\ y(k-2) \\ i_1(k-1) \\ i_1(k-2) \end{bmatrix} \text{ IST} \begin{bmatrix} 2,070494 \cdot 10^{-1} \\ 2,108600 \cdot 10^{-1} \\ 5,357028 \cdot 10^{-2} \\ 5,705156 \cdot 10^{-2} \end{bmatrix}, \quad (\text{A.5}) \end{aligned}$$

$$\begin{aligned} \text{DANN } y_3 = & 1,759499 \cdot 10^{-3} + 2,365624y(k-1) - 1,364526y(k-2) \\ & - 9,405826 \cdot 10^{-3}i_1(k-1) + 9,155348 \cdot 10^{-3}i_1(k-2) \end{aligned}$$

Regel 4:

$$\text{WENN } \begin{bmatrix} y(k-1) \\ y(k-2) \\ i_1(k-1) \\ i_1(k-2) \end{bmatrix} \text{ IST } \begin{bmatrix} -1,949657 \cdot 10^{-1} \\ -1,971022 \cdot 10^{-1} \\ -1,184905 \cdot 10^{-1} \\ -5,310534 \cdot 10^{-3} \end{bmatrix}, \quad (\text{A.6})$$

$$\begin{aligned} \text{DANN } y_4 = & -2,414692 \cdot 10^{-4} + 2,061418y(k-1) - 1,062521y(k-2) \\ & - 1,677289 \cdot 10^{-3}i_1(k-1) + 2,119155 \cdot 10^{-3}i_1(k-2) \end{aligned}$$

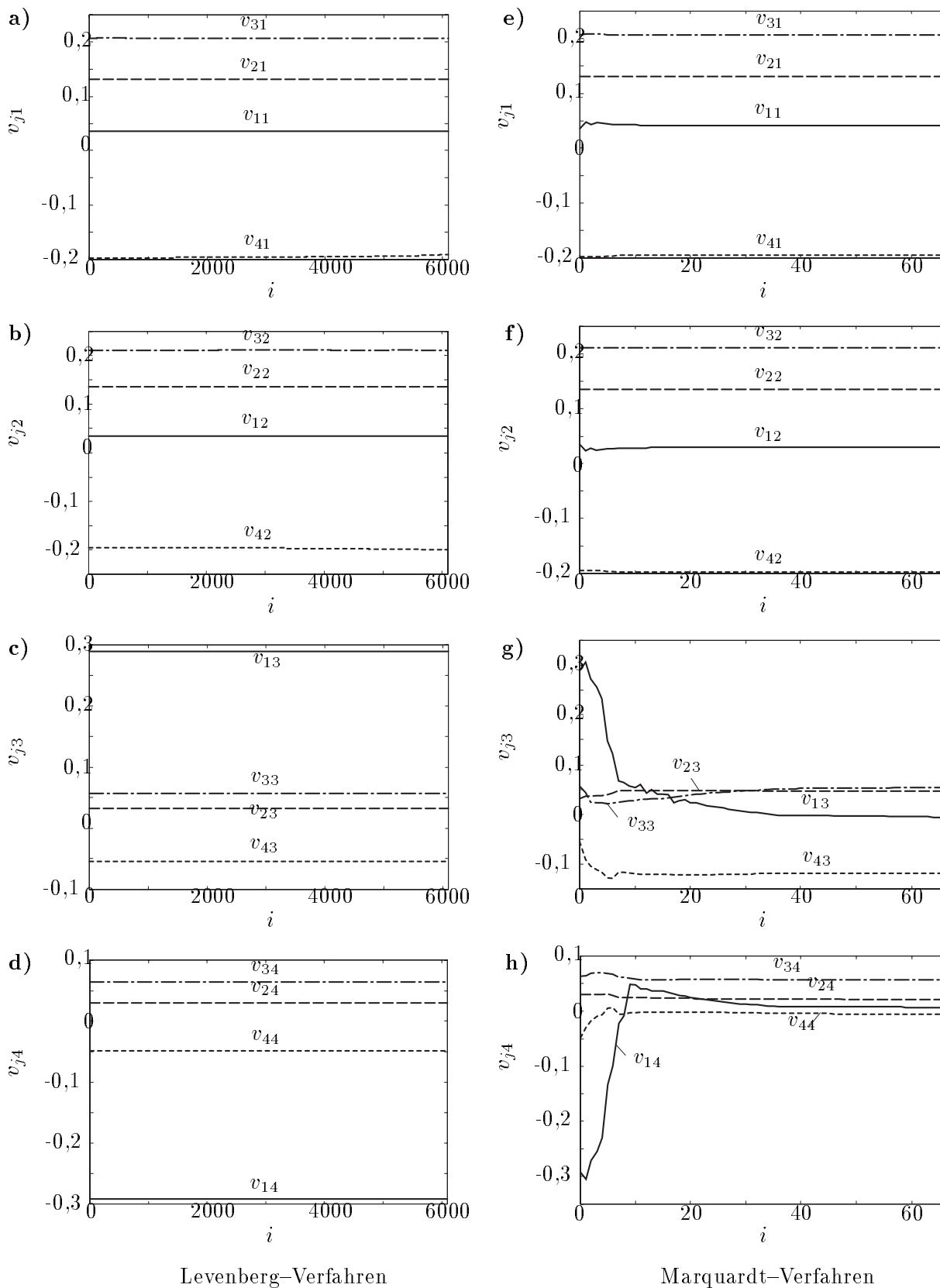
Als Abstandsnorm wurde die Mahalanobisnorm zugrundegelegt, für die sich folgende Formmatrix ergibt:

$$\mathbf{D} = \begin{bmatrix} 2,964460 \cdot 10^4 & -2,963617 \cdot 10^4 & -7,430495 \cdot 10 & 3,460139 \cdot 10^2 \\ -2,963617 \cdot 10^4 & 2,963201 \cdot 10^4 & 7,419174 \cdot 10 & -3,454137 \cdot 10^2 \\ -7,430495 \cdot 10 & 7,419174 \cdot 10 & 5,189392 \cdot 10 & -5,099541 \cdot 10 \\ 3,460139 \cdot 10^2 & -3,454137 \cdot 10^2 & -5,099541 \cdot 10 & 5,602823 \cdot 10 \end{bmatrix} \quad (\text{A.7})$$

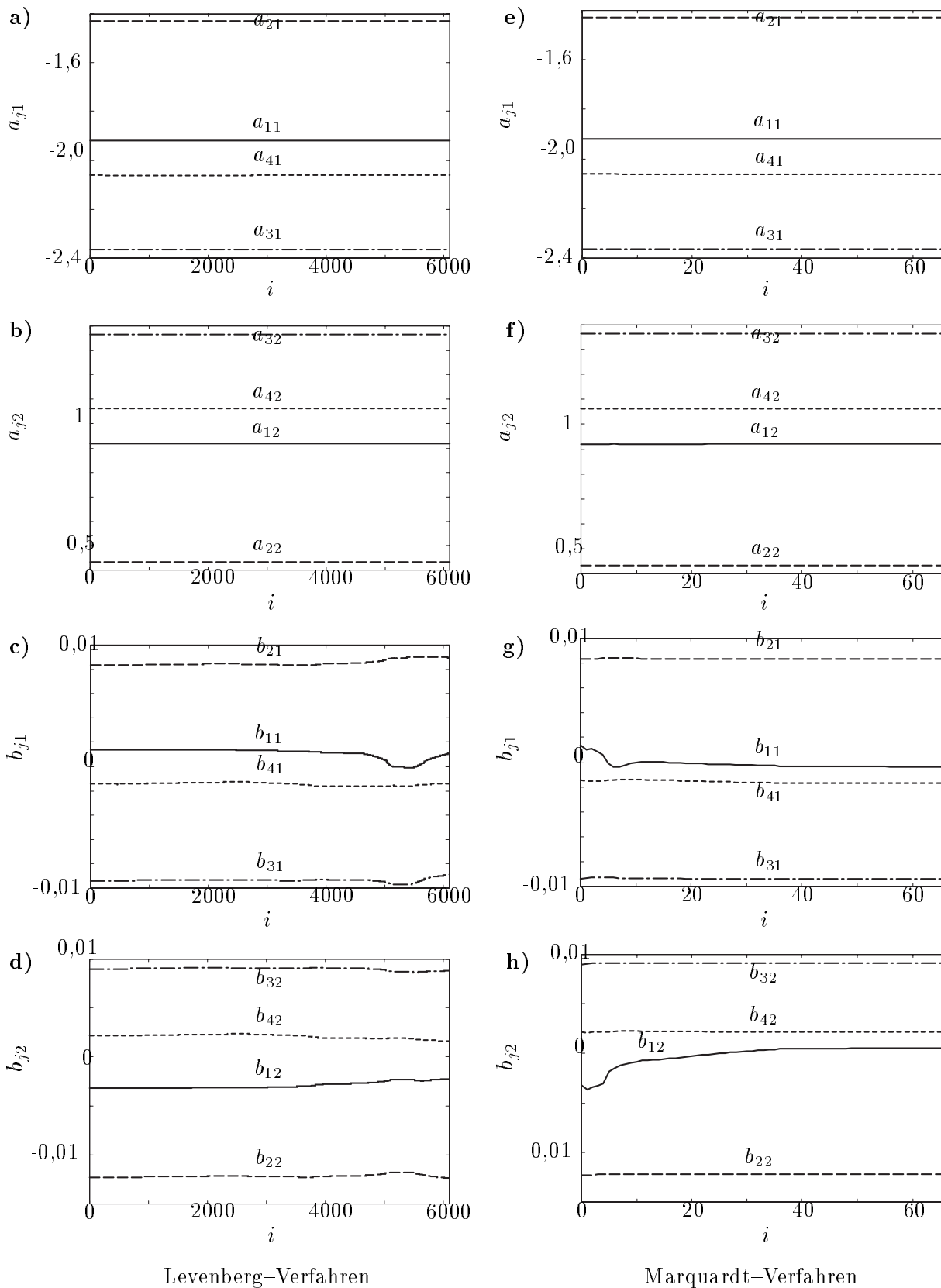


## **B Parameterverläufe bei rekursiver Optimierung**

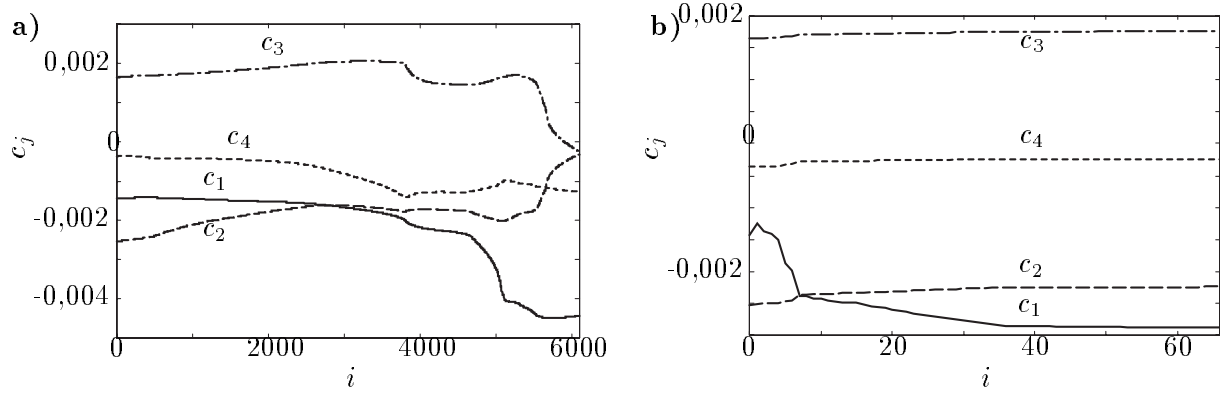
In Abschnitt 4.2 wurde der Einfluß einer rekursiven Optimierung auf die Parameterverläufe diskutiert. Zur Illustration dienten lediglich zwei Parameterverläufe mit den größten Parameteränderungen. Diese treten beim Marquardt–Verfahren auf. Um einen detaillierteren Vergleich zwischen dem Levenberg- und dem Marquardt–Verfahren bei rekursiver Optimierung sowie eine genaue Gegenüberstellung mit der prädiktiven Optimierung zu ermöglichen, werden hier alle Parameterverläufe vorgestellt.



**Bild B.1:** Parameterverläufe der Prototypenkoordinaten  $v_{jh}$  bei rekursiver Optimierung (Index  $j$ : Prototypennummer, Index  $h$ : Komponentennummer)



**Bild B.2:** Parameterverläufe der Konklusionspolynomparameter bei rekursiver Optimierung (Index  $j$ : Nummer des Konklusionspolynoms)



Levenberg-Verfahren

Marquardt-Verfahren

**Bild B.3:** Parameterverläufe bei rekursiver Optimierung