

J. Schultz / V. Krebs

# Methoden der Strukturanalyse zur Identifikation nichtlinearer Prozesse

## Kurzfassung

Zur Identifikation ohne vorherige theoretische Modellbildung hat sich die Modellklasse der Künstlichen Neuronen Netze (KNN) mit externer Dynamik als geeignet erwiesen. Jedoch müssen die zur Beschreibung des Prozesses notwendigen Verzögerungsterme der Ein- und Ausgangsgröße bestimmt werden. Dazu sind beispielsweise Pruningverfahren geeignet, von denen in diesem Beitrag die beiden Methoden Skelettierung und Optimal Brain Damage (OBD) für verrauschte Prozeßausgangsgrößen untersucht werden. Darüber hinaus wird eine neue Methode, die auf lokalen linearen Modellen basiert, für den verrauschten Fall erweitert und ihre Anwendbarkeit demonstriert. Alle Methoden werden am Beispiel eines Rührkesselreaktors präsentiert und miteinander verglichen.

## 1. Einführung

Die Systemanalyse oder Reglersynthese erfordert meist ein Modell der betrachteten Strecke. Dazu ist zunächst eine möglicherweise aufwendige theoretische Modellbildung notwendig. Um diese zu vermeiden, können KNN zum Aufbau von Verhaltensmodellen benutzt werden. Sie haben sich als geeignet erwiesen, nichtlineare dynamische Systeme zu identifizieren [1], [2]. Dies geschieht allein mittels Ein-/Ausgangsdaten des Prozesses. Es wird ein diskretes nichtlineares Black-Box-Modell generiert. Eine mögliche Struktur besteht darin, das KNN mit einer vorgeschalteten Komponente zu versehen, die aus Verzögerungselementen besteht (Abb. 1). Diese Struktur wird als Time-Delay Neural Network (TDNN) bezeichnet. Sie besteht aus einem statischen Neuronen Netz zur Approximation der unbekanntenen Nichtlinearität und einem vorgeschalteten dynamischen Block. Dieser baut einen Vektor aus verzögerten Ein- und

Ausgangsgrößen auf, die dem KNN am Eingang zur Verfügung gestellt werden.

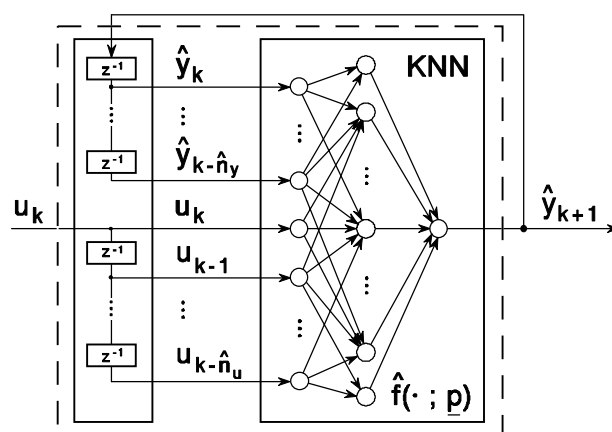


Abb. 1: Dynamisches Modell bestehend aus Künstlichem Neuronen Netz zur Approximation der Nichtlinearität und einem vorgeschalteten Block mit Zeitverzögerungen.

Bei der Verwendung dieser Modellform ist der minimale Satz der zur Beschreibung der Prozeßdynamik notwendigen Verzögerungsterme, die als Eingangsterme der Nichtlinearität eingehen, vor dem Training des KNN zu bestimmen. Verzögerungsterme, die zur Nachbildung der Dynamik des Prozesses nicht erforderlich sind, sollten im Modell nicht verwendet werden, da sie unnötigerweise die Eingangsdimension der Nichtlinearität und damit den Aufwand bei den nachfolgenden Schritten der Identifikation erhöhen. Darüber hinaus hat jeder unnötige Verzögerungsterm der Ausgangsgröße wegen der dynamischen Fortpflanzung des Approximationsfehlers des KNN einen negativen Einfluß auf die Qualität des dynamischen Modellverhaltens [12].

In [3], [4] wurden bereits Untersuchungen zur Bestimmung der relevanten Verzögerungen durchgeführt. Dabei werden als Pruningverfahren die Skelettierung und das OBD herangezogen. Diese Methoden haben das Ziel, unnötige Neuronen bzw. Parameter eines KNN zu be-

stimmen.

In [4] wurde darüber hinaus eine leistungsfähige neue Methode angegeben, die auf dem Ansatz von lokalen linearen Modellen basiert. In einem ersten Schritt wird mit Hilfe der Konditionierung einer aus Prozeßdaten aufgebauten Matrix die maximale Verzögerung der Ausgangsgröße festgestellt. Darauf basierend werden die restlichen notwendigen Verzögerungen über den durch eine Least-Squares-Schätzung ermittelten Gradientenvektor bestimmt.

Während den in [3], [4] angestellten Untersuchungen ungestörte Prozeßmeßdaten zugrunde liegen, werden im folgenden diese Methoden für den Fall verrauschter Meßdaten betrachtet. Dazu wird zuvor der in [4] eingeführte Ansatz auf die Berücksichtigung verrauschter Prozeßausgangsgrößen erweitert. Die Untersuchung der Methoden geschieht anhand der Strukturanalyse eines Beispielprozesses.

## 2. Identifikation eines nichtlinearen dynamischen Prozesses

Ein nichtlinearer dynamischer SISO-Prozeß, beschrieben durch

$$y_{k+1} = f\left(y_k, \dots, y_{k-n_y}, u_k, \dots, u_{k-n_u}\right) \quad (1)$$

soll ohne vorherige theoretische Modellbildung identifiziert werden. Es wird vorausgesetzt, daß der Prozeß nicht sprungfähig und im betrachteten Arbeitsgebiet BIBO-stabil ist. Weiterhin wird angenommen, daß die Nichtlinearität  $f(\cdot)$  zeitinvariant, stetig differenzierbar und beschränkt ist. Die maximalen (unbekannten) Zeitverzögerungen, die den Prozeßausgang beeinflussen, werden für die rückgekoppelte Ausgangsgröße mit  $n_y$  und für die Eingangsgröße mit  $n_u$  bezeichnet.

Das Modell (Abb. 1) wird gemäß der nachfolgenden Differenzgleichung

$$\hat{y}_{k+1} = \hat{f}\left(\hat{y}_k, \dots, \hat{y}_{k-\hat{n}_y}, u_k, \dots, u_{k-\hat{n}_u}; \underline{p}\right) \quad (2)$$

mit den maximalen Modellverzögerungen  $\hat{n}_y$ ,  $\hat{n}_u$  angesetzt. Der Vektor  $\underline{p}$  besteht aus den Modellparametern, die z.B. für das Mehrschicht-Perzeptron die Gewichte und Schwellwerte darstellen. Die Parameter in  $\underline{p}$  sind so zu

bestimmen, daß (2) das Ein-/Ausgangsverhalten von (1) möglichst gut nachbildet. Der klassische Ansatz besteht in der Minimierung des Gütekriteriums

$$J = \frac{1}{2} \sum_{k=0}^N (y_{k+1} - \hat{y}_{k+1})^2 \quad (3)$$

$$= \frac{1}{2} \sum_{k=0}^N \left( y_{k+1} - \hat{f}\left(y_k, \dots, y_{k-\hat{n}_y}, u_k, \dots, u_{k-\hat{n}_u}; \underline{p}\right) \right)^2$$

hinsichtlich des Vektors  $\underline{p}$  unter Verwendung einer endlichen Anzahl  $N+1$  von gemessenen Prozeßdaten. Als Ergebnis eines Optimierungsvorganges von (3) hinsichtlich  $\underline{p}$  erhält man ein  $\underline{p}_{\min}$ .

Hierbei sind alle Verzögerungen bis zum Grade  $\hat{n}_y$  bzw.  $\hat{n}_u$  in der Modellstruktur zu berücksichtigen, da ohne theoretische Modellbildung unbekannt ist, welche Zwischenterme unnötig sind. Zur Anwendung der nachfolgend untersuchten Methoden sind die maximalen Verzögerungen mit  $\hat{n}_y > n_y$  und  $\hat{n}_u > n_u$  anzusetzen.

## 3. Pruningmethoden zur Strukturanalyse

Von den Pruningmethoden haben sich zwei Ansätze zur Bestimmung der notwendigen Verzögerungen im Modell (2) - also der notwendigen Eingangsneuronen des KNN - bewährt [3]. Nach einer kurzen Beschreibung dieser Methoden wird auf die Ergebnisse bei der Anwendung auf einen Rührkesselreaktor mit verrauschten Daten eingegangen.

### 3.1 Skelettierung

Bei der Skelettierung [7] wird die Güte des KNN nach Entfernung eines Neurons herangezogen, um zu beurteilen, wie wichtig dieses Neuron ist. Unter Benutzung der quadratischen Summe (3) als Maß für die Netzgüte ergibt sich als Relevanzmaß eines Neurons

$$\rho_i = J_{\text{ohne Neuron } i} - J_{\text{mit Neuron } i} \quad (4)$$

Diese Relevanz wird für jedes Eingangsneuron berechnet. Je geringer  $\rho_i$  ist, desto weniger Auswirkungen hat das  $i$ -te Eingangsneuron. Das Neuron mit dem kleinsten  $\rho_i$  kann entfernt werden, wenn der zugehörige mittlere Approximationsfehler  $\frac{1}{N+1} \cdot J_{\text{ohne Neuron } i}$  unter einer hinreichend kleinen Fehlerschranke liegt.

Der nächste Schritt ist das Training eines KNN mit entferntem Eingangsneuron und damit ohne den zugehörigen Verzögerungsterm. Diese Schritte werden solange wiederholt, bis der mittlere Approximationsfehler des kleinsten  $\rho_i$  die gewählte Fehlerschranke überschritten hat.

### 3.2 Optimal Brain Damage (OBD)

Mit dieser Methode werden diejenigen Netzparameter gefunden, deren Entfernung aus dem Netz den kleinsten Zuwachs im Gütekriterium (3) verursachen. Dazu wird (3) als Taylorreihe um den ermittelten Parametervektor  $\underline{p}_{\min}$ , der (3) minimiert, entwickelt. Der Zuwachs des Gütekriteriums  $\Delta J$  kann als quadratische Funktion in Abhängigkeit von  $\Delta \underline{p} = \underline{p} - \underline{p}_{\min}$  approximiert werden. Mit einigen Vereinfachungen [8] gelangt man zu

$$\Delta J \approx \frac{1}{2} \sum_{i=1}^{n_p} h_{ii} \Delta p_i^2. \quad (5)$$

Bei der Bestimmung von  $\Delta J$  sind nur die von den Eingangsneuronen ausgehenden Parameter zu berücksichtigen, da lediglich der Verbleib dieser Neuronen in der Modellstruktur für die betrachtete Anwendung zur Diskussion steht.

Nach jedem Training wird auch bei dieser Methode das Eingangsneuron mit der geringsten Bedeutung aus der Modellstruktur entfernt, wobei als Maß  $\rho_i$  für die Relevanz des  $i$ -ten Eingangsneurons die Summe der Relevanzmaße aller von diesem Neuron ausgehenden Parameter herangezogen wird. Das Relevanzmaß eines Netzparameters ergibt sich aus (5) zu  $s_i = h_{ii} \Delta p_i^2 / 2$ .

**Beispiel (1. Teil):** Für einen Rührkesselreaktor soll anhand gemessener Daten von Kühltemperatur (=Eingangsgröße  $u$ ) und Reaktortemperatur (=Ausgangsgröße  $y$ ) eine Strukturanalyse durchgeführt werden. Mit der als nicht verfügbar angenommenen zeitkontinuierlichen Systembeschreibung [9] gelangt man mittels Euler-Approximation zu der zeitdiskreten Systemdarstellung

$$y_{k+1} = f(y_k, y_{k-1}, u_k, u_{k-1}). \quad (6)$$

Damit ist  $n_y = 1$  und  $n_u = 1$ . Es wird nachfolgend die normierte Darstellung aus [9] benutzt.

Der betrachtete Identifikationsbereich liegt für  $y$  im Intervall  $[-0,18;0,15]$  und für  $u$  im Intervall  $[-0,19;0,21]$ .

Es wird mittelwertfreies Gaußsches weißes Rauschen  $r_{k+1}$  mit unterschiedlicher Streuung  $\sigma$  additiv auf die Ausgangsgröße aufgeschaltet

$$y_{k+1} = f(y_k, y_{k-1}, u_k, u_{k-1}) + r_{k+1}. \quad (7)$$

Zur Approximation der Nichtlinearität wird ein Mehrschicht-Perzeptron (MLP) mit 2 verdeckten Schichten zu jeweils 6 Neuronen eingesetzt. Als Modellstruktur wird

$$\hat{y}_{k+1} = \hat{f}(\hat{y}_k, \hat{y}_{k-1}, \hat{y}_{k-2}, u_k, u_{k-1}, u_{k-2}; \underline{p}) \quad (8)$$

mit  $\hat{n}_y = 2$  und  $\hat{n}_u = 2$  angesetzt. Zur Approximation werden 9355 verrauschte Prozeßdatentupel  $\left\{ \left[ y_k, y_{k-1}, y_{k-2}, u_k, u_{k-1}, u_{k-2} \right], \left[ y_{k+1} \right] \right\}$  herangezogen. Dabei wird  $\sigma$  zwischen 0,05% und 50% des Bereichs der Ausgangsgröße variiert.

KNN wurden unter Benutzung der verrauschten Prozeßdaten mit den verschiedenen  $\sigma$  trainiert bis keine wesentliche Veränderung von  $J$  mehr stattfand. Anschließend wurden die Skelettierung bzw. die OBD angewandt. In den Tabellen 1 und 2 sind die Relevanzmaße für die Neuronen der einzelnen Verzögerungsterme angegeben, wobei diese jeweils auf den maximalen Wert der Gruppe der Ein- bzw. Ausgangsterme bezogen sind. Dabei muß beachtet werden, daß für  $\sigma \geq 5\%$  das KNN nicht bis zu einem hinreichend kleinen  $J$  minimiert werden konnte.

$\sigma$ [%]	$y_k$	$y_{k-1}$	$y_{k-2}$	$u_k$	$u_{k-1}$	$u_{k-2}$
0,05	1,0	0,12	0,000	0,67	1,0	0,000
0,1	1,0	0,11	0,000	0,68	1,0	0,000
0,5	1,0	0,12	0,000	0,72	1,0	0,000
1,0	1,0	0,12	0,000	0,69	1,0	0,000
5,0	1,0	0,32	0,276	1,0	0,77	0,933
10,0	1,0	0,27	0,001	0,77	1,0	0,104
50,0	1,0	0,21	0,054	0,9	0,72	1,0

Tab. 1: Die bezogenen Relevanzmaße bei Anwendung der Skelettierung auf die trainierten Netze für die unterschiedlichen Streuungen.

Für  $\sigma \leq 1\%$  kann mit beiden Methoden signifikant (mehr als drei Zehnerpotenzen) erkannt werden, daß  $y_{k-2}$  nicht im Modell berücksich-

tigt werden muß. Das gleiche gilt für  $u_{k-2}$ , wobei OBD für  $\sigma = 1, \%$  (lediglich ein Faktor 5) kein brauchbares Ergebnis liefert. Mögliche Ursachen sind in den Vereinfachungen, die zu (5) führen, zu suchen. In diesem Fall müßte ein Netz ohne den Term  $y_{k-2}$  trainiert und für dieses Netz die Analyse erneut durchgeführt werden.

$\sigma$ [%]	$y_k$	$y_{k-1}$	$y_{k-2}$	$u_k$	$u_{k-1}$	$u_{k-2}$
0,05	1,0	0,75	0,000	0,10	1,0	0,000
0,1	1,0	0,59	0,000	0,10	1,0	0,000
0,5	1,0	0,75	0,000	0,10	1,0	0,000
1,0	0,8	1,00	0,001	0,02	1,0	0,004
5,0	1,0	0,34	0,068	0,39	1,0	0,130
10,0	1,0	0,63	0,002	0,60	1,0	0,030
50,0	0,6	1,0	0,207	0,51	1,0	0,160

Tab. 2: Die bezogenen Relevanzmaße bei Anwendung von OBD auf die trainierten Netze für die unterschiedlichen Streuungen.

#### 4. Strukturanalyse mittels lokalem linearem Modell

In [3], [4] wurde eine Methode vorgestellt, mit der die relevanten Verzögerungsterme bestimmt werden können, ohne hierzu überdimensionierte KNN trainieren zu müssen. Dazu wird ein Ansatz über ein lineares Modell gemacht, das lokal um einen Punkt  $\underline{x}_p = [y_k, \dots, y_{k-\hat{n}_y}, u_k, \dots, u_{k-\hat{n}_u}]_p$  im Eingangsraum betrachtet wird. Bei der Methode wird zunächst die maximale Verzögerung  $n_y$  ermittelt. Im darauffolgenden Schritt wird mittels der Analyse des geschätzten Gradientenvektors an der Stelle  $\underline{x}_p$  auf weitere unnötige Verzögerungswerte geschlossen.

Um die maximale Verzögerung  $n_y$  zu bestimmen, wird die Konditionierung einer aus Meßdaten aufgebauten Matrix

$$\underline{A}_\kappa = \left[ \Delta y_{-k-\kappa}, \dots, \Delta y_{-k-\hat{n}_y}, \Delta u_{k-\kappa}, \dots, \Delta u_{k-\hat{n}_u} \right] \quad (9)$$

herangezogen.

Die Konditionszahl  $c_\kappa = \text{cond}\{\underline{A}_\kappa\}$  einer Matrix  $\underline{A}_\kappa$  ist das Verhältnis vom betragsmäßig größten zum betragsmäßig kleinsten singulären Wert dieser Matrix. Die singulären Werte sind die positiv genommenen Quadratwurzeln der Ei-

genwerte der Matrix. Eine Matrix ist für Konditionszahlen nahe bei Eins gut konditioniert und schlecht konditioniert, wenn die Konditionszahl sehr viel größer als Eins ist.

In die Vektoren  $\Delta y_{-k-v}^T = [\Delta y_{k-v}^1, \dots, \Delta y_{k-v}^N]$  und  $\Delta u_{k-v}^T = [\Delta u_{k-v}^1, \dots, \Delta u_{k-v}^N]$  geht die Differenz der Elemente der N Meßwerte zu den jeweiligen Elementen von  $\underline{x}_p$  ein. Die Meßdaten müssen aus der näheren Umgebung von  $\underline{x}_p$  gewählt werden, damit der auftretende Linearisierungsfehler klein gehalten wird.

Die Matrix  $\underline{A}_\kappa$  wird schrittweise für  $\kappa = \{\hat{n}_y, \hat{n}_y - 1, \dots, 1\}$  aufgebaut. Die Veränderung der Konditionszahl gibt einen Hinweis auf  $n_y$ . Für  $\hat{n}_y - n_y \leq \kappa \leq \hat{n}_y$  ist  $\underline{A}_\kappa$  gut konditioniert. Ab  $\kappa < \hat{n}_y - n_y$  folgt, daß  $\underline{A}_\kappa$  sich zu einer schlecht konditionierten Matrix ändert, da ab dem Hinzufügen von  $\Delta y_{-k+1+n_y-\hat{n}_y}$  für  $\kappa = \hat{n}_y - n_y - 1$  sich ein Rangabfall in  $\underline{A}_\kappa$  ergibt. Der Wechsel von einer gut zu einer schlecht konditionierten Matrix  $\underline{A}_\kappa$  zeigt an, daß ein Verzögerungswert der Ausgangsgröße mehr als notwendig berücksichtigt wird [4].

Der Rangabfall in  $\underline{A}_\kappa$  wird bei einem großen Linearisierungsfehler und bei Vorhandensein von Rauschen in den Meßdaten schlechter ausgeprägt sein. Deshalb muß im verrauschten Fall bei der Berechnung der Konditionszahlen das Rauschen mit einfließen.

Bei der Methode [4] zur Bestimmung der Verzögerungen mit dem lokalen Modell wird dann auf der gut konditionierten Matrix  $\underline{A}_{\kappa^*}$  aufbauend der Gradientenvektor geschätzt, um mit dessen Hilfe die unnötigen Verzögerungen für die Eingangsgröße und der verbliebenen Verzögerungsterme der Ausgangsgröße zu ermitteln. Auch bei der Schätzung des Gradientenvektors wird das Rauschen zu berücksichtigen sein. Im nachfolgenden wird auf diesen Problempunkt nicht genauer eingegangen, sondern auf [11] verwiesen.

Zur Berücksichtigung des Rauschens bei der Ermittlung der Konditionszahlen wird davon

ausgegangen, daß sich das gemessene  $y_{k+1}$ , durch

$$y_{k+1} = f\left(y_k, \dots, y_{k-n_y}, u_k, \dots, u_{k-n_u}\right) + r_{k+1} \quad (10)$$

mit Gaußschem weißen Rauschen  $r_{k+1}$  ergibt. Es kann durch

$$y_{k+1} = y_{l,k+1} + r_{k+1} \quad (11)$$

mit einer unverrauschten Prozeßausgangsgröße  $y_{l,k+1}$  und  $r_{k+1}$  ausgedrückt werden. Damit ist (9) durch  $\underline{A}_k = \underline{L}_k + \underline{R}_k$  darstellbar. Die Matrizen  $\underline{L}_k$  und  $\underline{R}_k$  sind analog (9) aufgebaut, wobei in  $\underline{R}_k$  lediglich die durch das Rauschen verursachten Anteile von (11) wiederzufinden sind. Wird nun die Korrelationsmatrix  $\underline{\Phi}_{AA,k} = \frac{1}{N} \cdot \underline{A}_k^T \underline{A}_k$  berechnet, so gilt wegen der statistischen Unabhängigkeit zwischen  $y_{l,k+1}$  und  $r_{k+1}$ :

$$\underline{\Phi}_{AA,k} = \underline{\Phi}_{LL,k} + \underline{\Phi}_{RR,k}. \quad (12)$$

Die Elemente von  $\underline{\Phi}_{RR,k}$  sind Autokorrelationsfunktionen des Rauschprozesses und werden als bekannt vorausgesetzt.

Die Aufgabe besteht nun darin, aus der Konditionszahl von  $\underline{\Phi}_{AA,k}$ , die aus den gemessenen Prozeßdaten ermittelt wird, und der Matrix  $\underline{\Phi}_{RR,k}$ , auf die Konditionszahl von  $\underline{\Phi}_{LL,k}$  zu schließen. Sie erlaubt dann, wie im nachfolgenden erläutert, eine Aussage hinsichtlich der notwendigen Eingangsterme des Prozesses.

Dazu hilft ein Satz aus [10]: *Wenn  $r_1 \leq \dots \leq r_n, s_1 \leq \dots \leq s_n$  und  $t_1 \leq \dots \leq t_n$  die Eigenwerte der  $n$ -dimensionalen hermiteschen Matrizen  $\underline{R}$ ,  $\underline{S}$  und  $\underline{T} = \underline{R} + \underline{S}$  sind, dann gilt der Zusammenhang  $r_j + s_1 \leq t_j \leq r_j + s_n$ ,  $j = 1, \dots, n$ .*

Werden die Eigenwerte von  $\underline{\Phi}_{AA,k}$  mit  $\alpha_{k,\min}^2, \dots, \alpha_{k,\max}^2$ , von  $\underline{\Phi}_{LL,k}$  mit  $\lambda_{k,\min}^2, \dots, \lambda_{k,\max}^2$  und von  $\underline{\Phi}_{RR,k}$  mit  $\beta_{k,\min}^2, \dots, \beta_{k,\max}^2$  bezeichnet, so folgt unter Berücksichtigung, daß die Eingangsgröße nicht verrauscht ist ( $\beta_{k,\min}^2 = 0$ ) mit obigem Satz durch Einsetzen und Umformung:

$$\alpha_{k,\max}^2 - \beta_{k,\max}^2 \leq \lambda_{k,\max}^2 \leq \alpha_{k,\min}^2 \quad (13)$$

$$\alpha_{k,\min}^2 - \beta_{k,\max}^2 \leq \lambda_{k,\min}^2 \leq \alpha_{k,\min}^2. \quad (14)$$

Der Wert  $\beta_{k,\max}^2$  wird durch Analyse des Rauschens ermittelt.

Mit (13), (14) und  $\lambda_{k,\min}^2 \geq 0, \lambda_{k,\max}^2 \geq 0$  ergeben sich unter Berücksichtigung des Rauschens die folgenden Abschätzungen der Konditionszahlen nach oben:

I. für  $\alpha_{k,\min}^2 - \beta_{k,\max}^2 > 0$ :

$$\text{cond}\{\underline{\Phi}_{LL,k}\} = \frac{\lambda_{k,\max}}{\lambda_{k,\min}} \leq \frac{\alpha_{k,\max}}{\sqrt{\alpha_{k,\min}^2 - \beta_{k,\max}^2}} = c_o$$

II. für  $\alpha_{k,\min}^2 - \beta_{k,\max}^2 \leq 0$ :

$$\text{cond}\{\underline{\Phi}_{LL,k}\} \leq \infty$$

und nach unten:

III. für  $\alpha_{k,\max}^2 - \beta_{k,\max}^2 > 0, \alpha_{k,\min} > 0$ :

$$\text{cond}\{\underline{\Phi}_{LL,k}\} = \frac{\lambda_{k,\max}}{\lambda_{k,\min}} \geq \frac{\sqrt{\alpha_{k,\max}^2 - \beta_{k,\max}^2}}{\alpha_{k,\min}} = c_u$$

IV. für  $\alpha_{k,\max}^2 - \beta_{k,\max}^2 \geq 0, \alpha_{k,\min} > 0$ :

$$\text{cond}\{\underline{\Phi}_{LL,k}\} \geq 0.$$

Aus II. erkennt man, daß eine Aussage über einen Konditionssprung nicht mehr möglich ist, wenn die mittlere Rauschleistung  $\beta_{k,\max}^2 = \overline{n_k^2}$  größer als  $\alpha_{k,\min}^2$  ist.

Die Abschätzung  $c_o$  wird herangezogen, um mit größerer Sicherheit eine Aussage über einen Sprung in der Konditionszahl zu machen.

**Beispiel (2. Teil):** Die Eingangsterme für das Modell des Rührkesselreaktors sollen ermittelt werden. Es sind jeweils 400 Datentupel mit einem Radius von  $1,0 \cdot 10^{-2}$  um ein  $\underline{x}_p$  ausgewählt. Mit  $\sigma = \{100; 10; 1\} \cdot 10^{-5}$  wird das Rauschen verändert. Die Ergebnisse sind in den Tabellen 3 bis 5 aufgelistet. Dabei bezeichnet

$c_A = \text{cond}\{\underline{\Phi}_{AA,k}\}$  und  $c_L = \text{cond}\{\underline{\Phi}_{LL,k}\}$ .

Man erkennt, daß mit  $c_A$  lediglich für  $\sigma = 1,0 \cdot 10^{-5}$  ein signifikanter Sprung von einer gut zu einer schlecht konditionierten Matrix detektiert werden würde. In Tab. 3 und 4 ist mit dem korrigierten Wert  $c_o$  nach I. ein

Sprung von  $\kappa = 1$  zu  $\kappa = 0$  zu detektieren. Damit ist  $\kappa^* = 1$  und  $n_y = \hat{n}_y - \kappa^* = 2 - 1 = 1$ .

Bei einem Rauschsignalpegel von  $\sigma = 1,0 \cdot 10^{-3}$  (entspricht  $\sigma \approx 0,3\%$  im 1. Teil des Beispiels) kann keine korrekte Detektion mehr erfolgen, wie aus Tabelle 5 zu entnehmen ist.

	$\kappa = 2$	$\kappa = 1$	$\kappa = 0$	$\kappa = -1$
$c_A$	1,79	7,323	716,48	753,0
$c_O$	1,79	7,324	2.540,0	$\infty$
$c_L$	1,79	7,323	2.455,1	8050,35
$c_U$	1,79	7,232	716,48	753,0

Tab. 3: Ergebnisse für  $\sigma = 1,0 \cdot 10^{-5}$ .

	$\kappa = 2$	$\kappa = 1$	$\kappa = 0$	$\kappa = -1$
$c_A$	1,59	61,91	88,65	89,87
$c_O$	1,60	63,09	$6,0 \cdot 10^{12}$	$\infty$
$c_L$	1,59	62,09	1.668,88	6.419,30
$c_U$	1,5	61,88	88,59	89,79

Tab. 4: Ergebnisse für  $\sigma = 1,0 \cdot 10^{-4}$ .

	$\kappa = 2$	$\kappa = 1$	$\kappa = 0$	$\kappa = -1$
$c_A$	1,65	5,29	8,69	8,73
$c_O$	1,81	$\infty$	$\infty$	$\infty$
$c_L$	1,54	6,83	2.168,39	7.583,9
$c_U$	1,61	5,01	8,03	7,86

Tab. 5: Ergebnisse für  $\sigma = 1,0 \cdot 10^{-3}$ .

## 5. Vergleich der Methoden

Für  $\sigma \leq 1,0\%$ , d.h. praktisch auftretende Störungen von ca.  $\pm 4^\circ (\triangleq 3\sigma)$  im betrachteten Reaktortemperaturbereich von  $[-20^\circ; 100^\circ]$ , führt die Skelettierung zu sehr guten Ergebnissen, wohingegen OBD bereits ab  $\sigma = 1,0\%$  schlechtere Resultate liefert.

Für die Methode mit lokalem linearen Modell müssen die statistischen Eigenschaften des Rauschens berücksichtigt werden, um eine sichere Aussage über die größte Ausgangsverzögerung machen zu können. Für einen Rührkesselreaktor hat die Analyse ab einem Wert  $\sigma \approx 0,3\%$  keine brauchbaren Ergebnisse mehr erzielt.

Im Gegensatz zum unverrauschten Fall [3], zeigen die Pruningmethoden hier einen leichten Vorteil. Die Ursache ist darin zu suchen, daß diese auf einem trainierten KNN aufbauenden Methoden einen ausgleichenden Effekt besitzen, da Daten aus dem gesamten Arbeitsbereich für das Netztraining benutzt werden. Demge-

genüber hat die direkt datengestützte Methode lokalen Charakter.

## Zusammenfassung

Methoden zur Bestimmung der notwendigen Verzögerungsterme zur Modellierung des nichtlinearen Prozesses wurden für den Fall verrauschter Prozeßausgangsgrößen untersucht und erweitert. Am Beispiel eines Rührkesselreaktors wurde demonstriert, daß die vorgestellten Methoden zur Eingangstermauswahl auch für den verrauschten Fall erfolgreich eingesetzt werden können. Es erscheint nicht erforderlich, eine bestimmte Methode zu favorisieren.

## Literaturhinweise

- [1] Hunt, K.J., Sbarboro, D.: "Neural Networks for Control Systems - A Survey", Automatica, 28(6), pp. 1083-1112, 1992.
- [2] Narendra, K. S., Parthasarathy, K.: "Identification and Control of Dynamical Systems Using Neural Networks", IEEE Trans. on Neural Networks, 1(1), pp. 4-27, March 1990.
- [3] Schultz, J., Hillenbrand, S.: "Determining the structure of nonlinear models", ECC '97, July 1-4, 1997, Brussels, Belgium.
- [4] Schultz, J.: "Eine Strategie zur Gewinnung Neuronaler Modelle für nichtlineare dynamische Systeme", VDI-GMA-Kongreß '96 Mess- und Automatisierungstechnik, 10.-11.9.1996, Baden-Baden, VDI-Verlag, S. 733-742.
- [5] Press, W. H.: "Numerical Recipes in C", Cambridge University Press, Second Edition, 1992.
- [6] Reed, R.: "Pruning Algorithms-A survey", IEEE Trans. on Neural Networks, Vol. 4, No. 5, Sept. 1993.
- [7] Mozer, M., Smolensky, P.: "Skeletonization: A Technique for trimming the fat from a network via relevance assessment.", Advances in Neural Information Processing Systems (1), 1988, pp. 107-115.
- [8] Cun, Y., Denker, J.: "Optimal Brain Damage", Advances in Neural Information Processing Systems (2), 1989, pp. 598-605.
- [9] Föllinger, O.: "Nichtlineare Regelungssysteme I", R. Oldenbourg, München, 7. Auflage, 1993.
- [10] Stewart, G. W.: "Introduction to matrix computations", Academic Press, 1973.
- [11] Isermann, R.: "Identifikation dynamischer Systeme Band I/II", Springer-Verlag, 1988.
- [12] Schultz, J., Krebs, V.: "Improvement of nonlinear dynamic models containing Neural Networks based on domain partitioning", ECC '95, Roma, Italy, Sept. 5-8, 1995, S. 1191-1196.

## Autorenangaben:

Dipl.-Ing. Jörg Schultz,  
 Prof. Dr.-Ing. Volker Krebs,  
 Institut für Regelungs- und Steuerungssysteme,  
 Universität Karlsruhe,  
 Kaiserstr. 12,  
 76131 Karlsruhe  
 Tel.: 0721/608-3179  
 Fax: 0721/608-2707  
 email: schultz@irs.etec.uni-karlsruhe.de