

Zur Wahl der Startparameter beim rekursiven Prädiktionsfehlerverfahren und zur Suche der Modellordnung nichtlinearer Systemmodelle

H. Reuter

Forschungsbericht Nr. 11/92

Meß-, Steuer- und Regelungstechnik

Übersicht: In diesem Bericht werden zum einen die Erfahrungen des Autors bei der Wahl der Startparameter für ein rekursives Prädiktionsfehlerverfahren am Beispiel der Identifikation eines zeitdiskreten parametrischen Wiener-Modells beschrieben. Hierbei wird auf die auftretenden Konvergenzprobleme eingegangen und es werden Hinweise zu ihrer Vermeidung sowie zur Wahl der notwendigen Anfangsparameter des Schätzalgorithmus gegeben. Zum anderen werden die Ergebnisse der Voruntersuchungen zur Wahl einer optimalen Modellordnung für nichtlineare Systemmodelle am Beispiel der Modellidentifikation eines elektrohydraulischen Translationsantriebes diskutiert.

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

Inhaltsverzeichnis

1	Einleitung	1
2	Das rekursive Prädiktionsfehlerverfahren für ein Wiener-Modell	2
2.1	Beschreibung des Identifikations-Algorithmus	2
2.2	Berechnung des Gradientenvektors für ein Wiener-Modell	3
2.3	Vorgabe der Anfangswerte	6
2.4	Konvergenzverhalten bei Rauscheinfluß	10
3	Bekannte Kriterien zur Wahl der Modellgröße	14
3.1	Vorstellung der Kriterien	14
3.2	Ergebnisse in der praktischen Anwendung	18
4	Zusammenfassung	22
5	Literaturverzeichnis	23

1 Einleitung

Zur Berechnung von Zusammenhängen natürlicher und technischer Prozesse wird oft ein Modell des Ein-/Ausgangsverhaltens benötigt. Dies kann man zum einen durch eine theoretische Modellbildung erhalten, indem man die bekannten physikalischen und andere Zusammenhänge zu einem Modell des Prozesses zusammenfaßt. Sind aber die Zusammenhänge oder der Aufbau des Prozesses zu komplex oder sind nicht alle zur Beschreibung notwendigen Effekte bekannt, kann hiermit oft kein ausreichend genaues Modell gefunden werden. Die andere Möglichkeit ist die der experimentellen Modellbildung, der sogenannten Identifikation, bei der das zu untersuchende System als „black box“ angesehen wird (Schwarz 1991). Hierbei berechnet man aufgrund der bekannten Ein- und Ausgangssignale ein mathematisches Modell, das das Übertragungsverhalten charakterisieren soll.

Die zur Identifikation parametrischer Modelle in einem weiten Bereich verwendeten Least-Squares-Methoden besitzen jedoch den Nachteil, daß hierbei eine Ein-/Ausgangsdarstellung des zu identifizierenden Modells vorliegen muß, die linear in den Parametern ist. Diesen Nachteil besitzt das Prädiktionsfehlerverfahren nicht. Im Gegensatz zum rekursiven Least-Squares-Verfahren ist jedoch beim rekursiven Prädiktionsfehlerverfahren die Wahl der Startparameter sowie der Zeitverlauf des Vergessensfaktors von entscheidender Bedeutung für die Konvergenz. In Abschnitt 2 werden am Beispiel eines zeitdiskreten parametrischen Wiener-Modells die auftretenden Konvergenzprobleme beschrieben und Hinweise zu ihrer Vermeidung gegeben.

Ein anderes Problem bei der Identifikation mit geringen a-priori-Informationen liegt in der Wahl der Modellordnung ohne Kenntnis des realen Prozesses. In Abschnitt 3 werden bekannte Kriterien hierfür beschrieben. Dabei erscheint eine Unterscheidung in Methoden für lineare und für nichtlineare Systemmodelle sinnvoll. Für die Identifikation eines bilinearen Modells in beobachtbarkeitskanonischer Form mit einem rekursiven Least-Squares-Algorithmus werden am Beispiel eines realen elektrohydraulischen Translationsantriebes (Reuter 1992b) die Ergebnisse der algorithmisch dominierten Kriterien diskutiert. Die Zusammenfassung in Abschnitt 4 schließt den Bericht ab.

2 Das rekursive Prädiktionsfehlerverfahren für ein Wiener-Modell

2.1 Beschreibung des Identifikations-Algorithmus

Seit einigen Jahren ist ein Identifikationsverfahren bekannt, das sich auch gut zur Identifikation nichtlinearer Systemmodelle verwenden läßt, da hier die Linearität des Modells in den Parametern, wie sie z.B. beim Least-Squares-Verfahren notwendig ist, nicht gefordert wird. Es handelt sich hierbei um das Prädiktionsfehlerverfahren (PEM: Prediction-Error-Method) (Ljung 1987).

Ausgangspunkt für ein rekursives Prädiktionsfehlerverfahren (RPEM) (Ljung und Söderström 1987) ist die Berechnung des Vorhersage- oder Prädiktionsfehlers

$$e(k) = y(k) - \hat{y}(k) \quad (2.1)$$

zwischen einem gemessenen Ausgangssignal $y(k)$ und einem geschätzten $\hat{y}(k)$. Der weitere Algorithmus wird dann oft wie folgt angegeben:

$$\mathbf{R}(k) = \mathbf{R}(k-1) + \gamma(k) [\boldsymbol{\Psi}(k) \boldsymbol{\Psi}^T(k) - \mathbf{R}(k-1)] \quad (2.2)$$

$$\mathbf{L}(k) = \eta(k) \mathbf{R}^{-1}(k) \boldsymbol{\Psi}(k) \quad (2.3)$$

$$\boldsymbol{\Theta}(k) = \boldsymbol{\Theta}(k-1) + \mathbf{L}(k) e(k) \quad (2.4)$$

Hierbei bezeichnet \mathbf{R} eine Matrix und \mathbf{L} die Suchrichtung für die neu zu schätzenden Parameter des Parametervektors $\boldsymbol{\Theta}$. Der Faktor η bewirkt eine zeitliche Gewichtung der einzelnen Meßwerte und wird im Englischen mit „gain sequence“ bezeichnet, was wohl im Deutschen durch den Begriff Verstärkungsfolge oder Gewichtungsfolge beschrieben werden kann. Der Vektor $\boldsymbol{\Psi}$ enthält die Gradienten des Prädiktors, also des geschätzten Ausgangssignals, nach den zu schätzenden Parametern:

$$\boldsymbol{\Psi}(k) = \frac{\partial \hat{y}(k)}{\partial \boldsymbol{\Theta}} \quad (2.5)$$

Dieser Algorithmus kann jedoch leicht zu numerischer Instabilität führen, da in Gl. (2.3) bei jedem Schritt die Invertierung der $p \times p$ -Matrix \mathbf{R} vorgenommen werden muß, wobei p die Anzahl der zu schätzenden Parameter darstellt. Unter Verwendung von

$$\mathbf{P}(k) = \eta(k) \mathbf{R}^{-1}(k) \quad (2.6)$$

und mit dem Vergessensfaktor

$$\lambda(k) = \frac{\eta(k-1)}{\eta(k)} [1 - \eta(k)] \quad (2.7)$$

ergibt sich folgender Algorithmus:

$$\mathbf{L}(k) = \mathbf{P}(k-1) \boldsymbol{\Psi}(k) \left[\lambda(k) + \boldsymbol{\Psi}^T(k) \mathbf{P}(k-1) \boldsymbol{\Psi}(k) \right]^{-1} \quad (2.8)$$

$$\boldsymbol{\Theta}(k) = \boldsymbol{\Theta}(k-1) + \mathbf{L}(k) \epsilon(k) \quad (2.9)$$

$$\mathbf{P}(k) = \frac{1}{\lambda(k)} \left[\mathbf{P}(k-1) - \mathbf{L}(k) \boldsymbol{\Psi}^T(k) \mathbf{P}(k-1) \right] \quad (2.10)$$

Jedoch auch hier kann es bei der Berechnung von \mathbf{P} wegen der Differenzbildung in Gl. (2.10) zu numerischen Ungenauigkeiten kommen. Ljung und Söderström (1987 : 332) schlagen deshalb eine sogenannte U-D-Faktorisierung nach Biermann ($\mathbf{P} = \mathbf{U} \mathbf{D} \mathbf{U}^T$) vor, wobei \mathbf{U} eine obere Dreiecksmatrix und \mathbf{D} eine Diagonalmatrix darstellt. Die Unterschiede zwischen der Berechnung nach Gl. (2.8) bis Gl. (2.10) und dieser normalisierten Cholesky-Zerlegung (sog. square root free Cholesky decomposition) werden jedoch erst für Werte $p \gg 10$ signifikant.

Um Konvergenz zu erreichen, müssen zudem die geschätzten Parameter ein stabiles Modell beschreiben. Auch hat der Vergessensfaktor beim rekursiven Prädiktionsfehlerverfahren, im Gegensatz zu rekursiven Least-Squares-Verfahren, wo er eine eher untergeordnete Rolle spielt, einen sehr großen Einfluß auf die Konvergenzeigenschaften der Schätzung.

2.2 Berechnung des Gradientenvektors für ein Wiener-Modell

Wendet man dieses Verfahren auf ein Wiener-Modell mit Störfilter (Young 1984) an (Kortmann 1989), so muß hierfür das prädizierte Ausgangssignal $\hat{y}(k)$ und der Gradient $\boldsymbol{\Psi}(k)$ berechnet werden.

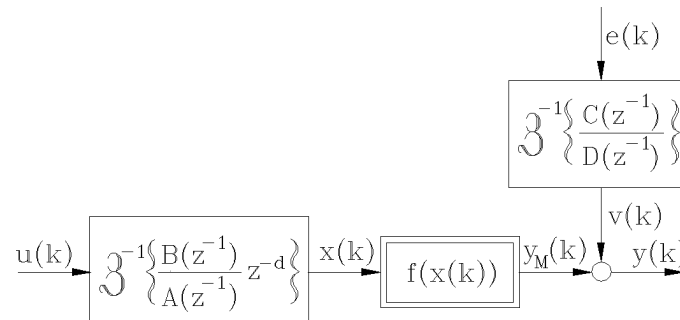


Bild 2.1: Wiener-Modell mit Störfilter

Den dynamischen Teil des Modells beschreibt unter Berücksichtigung einer als bekannt vorausgesetzten Totzeit dT mit der Abtastzeit T die Gleichung

$$X(z^{-1}) = \frac{B(z^{-1})}{A(z^{-1})} z^{-d} U(z^{-1}) \quad (2.11)$$

mit

$$A(z^{-1}) = 1 + a_1 z^{-1} + \dots + a_n z^{-n} \quad (2.12)$$

$$B(z^{-1}) = b_0 + b_1 z^{-1} + \dots + b_n z^{-n} \quad , \quad (2.13)$$

wobei die Modellordnung n als bekannt vorausgesetzt wird und z^{-k} eine zeitliche Verschiebung um k diskrete Schritte darstellt. Mit der Annahme, daß die statische Nichtlinearität durch ein Polynom der bekannten Ordnung q ausreichend approximiert werden kann, ergibt sich das ungestörte Ausgangssignal des Modells zu

$$y_M(k) = \sum_{i=0}^q \gamma_i x^i(k) \quad . \quad (2.14)$$

Zur Nachbildung eines stochastischen Störsignals dient ein Rauschfilter

$$V(z^{-1}) = \frac{C(z^{-1})}{D(z^{-1})} E(z^{-1}) \quad (2.15)$$

mit den beiden Polynomen $C(z^{-1})$ und $D(z^{-1})$ der ebenfalls bekannten Ordnung f ,

$$C(z^{-1}) = 1 + c_1 z^{-1} + \dots + c_n z^{-f} \quad (2.16)$$

$$D(z^{-1}) = 1 + d_1 z^{-1} + \dots + d_n z^{-f} \quad . \quad (2.17)$$

Da ein normalverteiltes weißes Rauschen als Störsignal $\epsilon(k)$ natürlich nicht explizit bekannt ist, für den Prädiktionsfehler aus Gl. (2.1) jedoch auch Normalverteilung angenommen wird, kann dieser Prädiktionsfehler dem Ausgangssignal über das Störfilter auch additiv zugeführt werden (Bild 2.1).

Somit ergibt sich das gesamte Modell zu

$$y(k) = \sum_{i=0}^q \gamma_i \left[\sum_{j=0}^n b_j u(k-j-d) - \sum_{j=1}^n a_j x(k-j) \right]^i + v(k) \quad (2.18)$$

mit

$$v(k) = \sum_{j=1}^f c_j \epsilon(k-j) - \sum_{f=1}^f d_j v(k-j) + \epsilon(k) \quad . \quad (2.19)$$

Der Fehler $\epsilon(k)$ ist aber zum Zeitpunkt k noch nicht bekannt, da er mit diesem Modell erst berechnet werden soll. Da $\epsilon(k)$ jedoch als normalverteilt angenommen wird, kann die beste Schätzung von $v(k)$ durch Weglassen von $\epsilon(k)$ zum Zeitpunkt k erreicht werden (Kortmann 1989).

Mit den zu schätzenden Parametern

$$\Theta = [a_1 \dots a_n, b_0 \dots b_n, \gamma_0 \dots \gamma_q, c_1 \dots c_f, d_1 \dots d_f]^T \quad (2.20)$$

ergibt sich der Gradientenvektor $\Psi(k)$ für diese Modellstruktur zu

$$\Psi(k) = \frac{\partial \hat{y}(k)}{\partial \Theta} = \left[\frac{\partial \hat{y}(k)}{\partial a_1} \quad \dots \quad \frac{\partial \hat{y}(k)}{\partial d_f} \right]^T \quad . \quad (2.21)$$

Die einzelnen Elemente von Ψ werden demnach wie folgt berechnet:

$$\begin{aligned} \frac{\partial \hat{y}(k)}{\partial a_i} &= \sum_{\mu=0}^f d_\mu \left[\sum_{j=1}^q j \gamma_j x^{j-1} \left[-x(k-i-\mu) - \sum_{r=1}^n a_r \frac{\partial x(k-r-\mu)}{\partial a_i} \right] \right] \\ &- \sum_{\nu=1}^f c_\nu \frac{\partial \hat{y}(k-\nu)}{\partial a_i} \quad , \quad i = 1, \dots, n; \quad d_0 = 1 \end{aligned} \quad (2.22)$$

$$\begin{aligned} \frac{\partial \hat{y}(k)}{\partial b_i} &= \sum_{\mu=0}^f d_\mu \left[\sum_{j=1}^q j \gamma_j x^{j-1} \left[u(k-i-\mu-d) - \sum_{r=1}^n a_r \frac{\partial x(k-r-\mu)}{\partial b_i} \right] \right] \\ &- \sum_{\nu=1}^f c_\nu \frac{\partial \hat{y}(k-\nu)}{\partial b_i} \quad , \quad i = 0, \dots, n; \quad d_0 = 1 \end{aligned} \quad (2.23)$$

$$\frac{\partial \hat{y}(k)}{\partial \gamma_0} = 1 - \sum_{\nu=1}^f c_\nu \frac{\partial \hat{y}(k-\nu)}{\partial \gamma_0} \quad (2.24)$$

$$\frac{\partial \hat{y}(k)}{\partial \gamma_i} = \sum_{\mu=0}^f d_\mu x^i(k-\mu) - \sum_{\nu=1}^f c_\nu \frac{\partial \hat{y}(k-\nu)}{\partial \gamma_i} \quad , \quad i = 1, \dots, q; \quad d_0 = 1 \quad (2.25)$$

$$\frac{\partial \hat{y}(k)}{\partial c_i} = \epsilon(k-i) - \sum_{\nu=1}^f c_\nu \frac{\partial \hat{y}(k-\nu)}{\partial c_i} \quad , \quad i = 1, \dots, f \quad (2.26)$$

$$\frac{\partial \hat{y}(k)}{\partial d_i} = -v(k-i) - \sum_{\nu=1}^f c_\nu \frac{\partial \hat{y}(k-\nu)}{\partial d_i} \quad , \quad i = 1, \dots, f \quad (2.27)$$

mit

$$\frac{\partial x(k)}{\partial a_i} = -x(k-i) - \sum_{r=1}^n a_r \frac{\partial x(k-r)}{\partial a_i} \quad , \quad i = 1 \dots n \quad (2.28)$$

$$\frac{\partial x(k)}{\partial b_i} = u(k-i-d) - \sum_{r=1}^n a_r \frac{\partial x(k-r)}{\partial b_i} \quad , \quad i = 0 \dots n \quad (2.29)$$

nach Gl. (2.11).

Wie schon oben angedeutet, ist die Stabilität des geschätzten zeitdiskreten Modells (Schwarz 1979) eine Voraussetzung für eine erwartungstreue Schätzung, also die Konvergenz zu den „richtigen“ Parametern. Dazu müssen die Nullstellen der Polynome $A(z^{-1})$ und $C(z^{-1})$ (Kortmann 1989) innerhalb des Einheitskreises liegen. Dann kann das Modell auch bei beliebiger, aber stetig differenzierbarer, statischer Nichtlinearität nicht mehr instabil werden.

Zur Vermeidung dauerhaft instabiler Modelle ist eine „Identifikationslenkung“ notwendig. Diese wird wie folgt gehandhabt. Mit den Parametern des zuletzt berechneten stabilen Modells wird das neue Ausgangssignal geschätzt. Hiermit wird der RPEM-Algorithmus einmal durchlaufen, um die neuen Parameter zu berechnen. Ist das geschätzte Modell stabil, können die neuen Parameter weiter verwendet werden. Andernfalls müssen alle Größen, die seit dem letzten stabilen Modell berechnet wurden, vergessen werden und der

ganze Durchlauf wird mit dem zuletzt ermittelten stabilen Modell und dem nun folgenden Ein- und Ausgangssignal durchgeführt.

Ein weiterer noch zu beachtender Punkt liegt darin, daß ein linearer Verstärkungsfaktor sowohl im Zählerpolynom $B(z^{-1})$ als auch im Faktor γ_1 enthalten ist. Um die eindeutige Identifizierbarkeit zu ermöglichen, muß eine Normierung durchgeführt, einer dieser Werte vorgegeben werden und auch während des ganzen Rechenlaufes konstant bleiben.

2.3 Vorgabe der Anfangswerte

Zu Beginn dieses Algorithmus werden einige Startwerte benötigt, die, wenn sie nicht von dem zu untersuchenden System bekannt sind, gesetzt werden müssen.

So sind wegen der zeitlichen Rückverschiebung der dynamischen Glieder Signalwerte $u(k)$, $y(k)$, $x(k)$, $e(k)$, $v(k)$ für diskrete Zeiten $k < 0$ notwendig. Hier bietet es sich an, diese Werte zu 0 anzunehmen. Für den Parametervektor setzt man $\Theta(0) = \mathbf{0}$ und für die zur Berechnung des Gradienten Ψ notwendigen Ableitungen $\partial \hat{y}(k \leq 0) / \partial \Theta = \mathbf{0}$ ebenso wie für $\partial x(k \leq 0) / (\partial a_i$ bzw. $\partial b_i) = 0$.

Einen gewichtigen Einfluß auf die Konvergenz besitzen auch die Anfangswerte der Matrix $\mathbf{P}(0)$. Diese sollte für die ersten Versuche als Startwert eine Diagonalmatrix mit $\mathbf{P}(0) = C_1 \mathbf{I}$ sein mit großen Werten C_1 (z.B. $10 \leq C_1 \leq 10000$) oder auch $C_1 \approx 100 / \overline{y^2(k)}$ (Ljung und Söderström 1987). Besitzen Eingangs- und Ausgangssignal signifikant unterschiedliche Amplituden, sollten die Anfangswerte der Diagonalmatrix $\mathbf{P}(0)$ auch unterschiedlich gewählt werden. Wenn in der Anfangsphase der Identifikation geschätzte Parameter sehr stark in einem größeren Wertebereich hin- und herspringen, sollten die zugehörigen Anfangswerte von $\mathbf{P}(0)$ etwas kleiner gewählt werden.

Einen weitaus signifikanteren Einfluß auf das Konvergenzverhalten übt der Vergessensfaktor $\lambda(k)$ aus. Bei einem konstanten Vergessensfaktor ergibt sich die sogenannte Gedächtnis-Zeitkonstante zu

$$T_\lambda \approx \frac{1}{1 - \lambda} \quad . \quad (2.30)$$

Mit $\lambda = 1$ werden alle jemals berechneten Werte gleichgewichtig berücksichtigt. Für $\lambda < 1$ geht der um die Zeit T_λ zurückliegende Vorhersagefehler $e(k - T_\lambda)$ jedoch nur noch mit $e^{-1} \approx 36\%$ in die momentanen Daten mit ein.

Bei zeitvarianten Systemen bietet sich somit ein Vergessensfaktor im Bereich $0.9 \leq \lambda \leq 1$ an. Wegen $T_\lambda(\lambda = 0.9) = 10$ gehen nur noch die Fehler der letzten 10 Tastschritte mit mehr als 36 % in die Schätzung mit ein, so daß auch größere Fehler oder Störungen schon

eine starke Parameteränderung bewirken können.

Jedoch auch für zeitinvariante Systeme ist der zeitliche Verlauf des Vergessensfaktors wichtig. So sollte zu Beginn der Identifikationsphase $\lambda(k)$ etwas kleiner sein, da zu diesem Zeitpunkt nur wenige Daten berücksichtigt wurden, der Algorithmus sich noch in einem „Einschwingvorgang“ befindet und deshalb der Einfluß ungesicherter, ungenauer und somit möglicherweise falscher Werte die Modellfindung in eine falsche Richtung lenken. Ljung und Söderström (1987) geben folgenden Zeitverlauf für den Vergessensfaktor an:

$$\lambda(k+1) = \lambda(k) (1 - \Delta\lambda) + \Delta\lambda \quad . \quad (2.31)$$

Günstige Anfangswerte liegen im Bereich $0.9 \leq \lambda(0) \leq 0.99$, wobei für Eingangssignale mit einem Abtastschritt pro Signaltakt der Bereich $0.94 \leq \lambda(0) \leq 0.97$ favorisiert wird. Verwendet man dagegen Eingangssignale mit mehreren Taktzeitpunkten pro Signaltakt, wie z.B. bei einem PRBS (Binäres Pseudo-Rauschsignal), wo das Eingangssignal mindestens w Abtastschritte konstant bleibt, sollte $\lambda(0)$ entsprechend größer gewählt werden, da beispielsweise bei periodischen Signalen die Informationen auf w -mal so viele Werte verteilt sind. Ebenfalls wichtig ist hier die richtige Wahl von $\Delta\lambda$. Dieser Wert bestimmt die Steigung des Zeitverlaufes von $\lambda(k)$ und muß in Abhängigkeit von der Anzahl der zu schätzenden Parameter und von der Modellkomplexität angegeben werden. Ljung und Söderström (1987) halten für ein lineares Modell mit sechs zu schätzenden Parametern $\Delta\lambda = 0.01$ für günstig. Mit zunehmender Parameterzahl und Modellkomplexität muß $\Delta\lambda$ dann kleiner werden, damit sich die zu schätzenden Parameter besser aneinander anpassen können. Wählt man $\Delta\lambda$ zu groß, so werden zu einem früheren Zeitpunkt alle vergangenen Werte berücksichtigt. Damit ist eine Änderung der Parameter auch schwieriger und langwieriger. Es ergibt sich dann entweder eine sehr langsame Konvergenz der Parameter oder es werden fortlaufend instabile Modelle geschätzt. Eine zusätzliche Möglichkeit ist, $\lambda(0)$ für eine gewisse Zeit konstant zu lassen, um die Anlaufphase zu verlängern und um den Einfluß von „Parameterstößen“ in der Anfangsphase der Berechnung schneller abzumindern, da bei der rekursiven Berechnung von $\lambda(k)$ die Steigung von $\lambda(k)$ anfangs am größten ist. Für zunehmende k geht sie dann gegen Null, so daß $\lambda(k)$ sich asymptotisch dem Endwert 1 nähert.

Zur Verifikation des programmierten Verfahrens wurde das Modell von Kortmann (1989) mit dem dynamischen Glied

$$x(k) = \mathcal{Z}^{-1} \left\{ \frac{0.1333 z^{-1} + 0.0667 z^{-2}}{1 - 1.5 z^{-1} + 0.7 z^{-2}} \right\} u(k-d) \quad ; \quad d=0 \quad (2.32)$$

und der nicht invertierbaren statischen Nichtlinearität (Bild 2.2) untersucht:

$$y(k) = x(k) + 3x^2(k) + 1.5x^3(k) \quad . \quad (2.33)$$

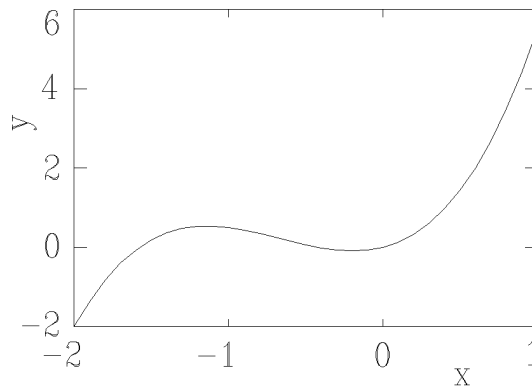


Bild 2.2: Statische Nichtlinearität

Die bei ihm aufgeführten Ergebnisse konnten mit den dort angegebenen Anfangswerten jedoch nicht nachvollzogen werden. Auch bei Variation der Werte von $\mathbf{P}(0)$, $\lambda(0)$, $\Delta\lambda$ und f konnte weder ohne noch mit Rauschen ein konvergierendes Modell gefunden werden, da schon nach einigen Identifikationsschritten das Polynom $C(z^{-1})$ fortlaufend mit Nullstellen außerhalb des Einheitskreises berechnet wurde und somit kein stabiles Modell mehr zu erhalten war.

Nimmt man das Fehlerfilter aus dem Modell heraus oder setzt – was gleichbedeutend damit ist – die Anfangswerte der zur Berechnung der Fehlerfilterparameter c_i und d_i notwendigen Matrix $\mathbf{P}_{jj}(0) = 0$ ($\forall c_i, d_i; i = 1, \dots, f$), ergibt sich auch keine Änderung dieser Parameter. Somit kann bei der günstigen Wahl der anderen Anfangswerte ein stabiles Modell geschätzt werden, wenn die Störung sehr klein bleibt.

Bild 2.3 zeigt das Konvergenzverhalten der geschätzten Parameter dieses Modells (Gl. (2.32) und (2.33)) ohne Berücksichtigung des Fehlerfilters. Eingangssignal ist ein PRBS aus einem fünfstufigen Register mit den Anfangswerten $[1 \ 0 \ 1 \ 0 \ 1]$ (Wernstedt 1989), einem Tastschritt pro Takt und $u = \pm 0.5$. Vorgegeben waren ebenfalls $n = 2$ und $q = 3$. Das konvergierende Modell ergab sich bei Verwendung von $\lambda(0) = 0.95$, $\Delta\lambda = 0.002$ und $\mathbf{P}(0) = 100 \mathbf{I}$ ($\forall a(n), b(n), \gamma(q)$) aus (Gl. (2.11) bis (2.13)). Man erkennt, daß alle geschätzten Parameter erst nach etwas über 1000 Meßwerten in die Nähe der vorgegebenen gelangen. Betrachtet man den Anfangsverlauf der Schätzung, so scheint es, daß das Erreichen der annähernd richtigen Parameterendwerte nur Zufall ist. Daß dieser Verdacht nicht von der Hand zu weisen ist, zeigt folgende Erfahrung. Ändert man bei sonst gleichen Anfangsbedingungen lediglich die Anfangswerte des PRBS – bei einem fünfstufigen Register ergeben sich $2^5 - 1 = 31$ Möglichkeiten – so wird in den meisten Fällen kein stabiles Modell mehr berechnet. Beim Rest der Fälle ergibt sich ein oft völlig anderes Konvergenzverhalten. Als günstig hat sich erwiesen, zu Anfang der Messung mit sich ändernden Eingangswerten zu beginnen.

Bei Verwendung eines nichtrekursiven Fehlerfilters – $D(z^{-1}) = 1 = const.$ – kann ebenfalls

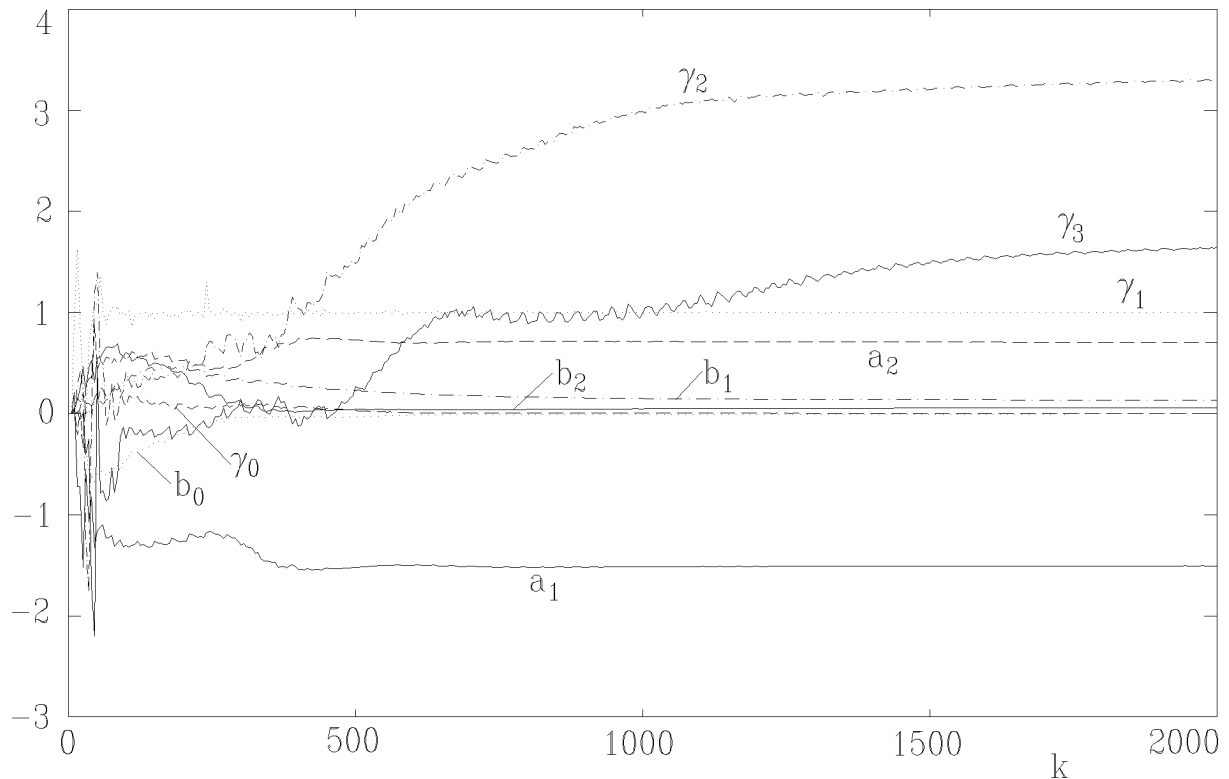


Bild 2.3: Konvergenzverhalten einer Schätzung ohne Störung und Fehlerfilter

ein stabiles Modell berechnet werden, wobei kaum Nullstellen von $C(z^{-1})$ außerhalb des Einheitskreises zu liegen kommen.

Verbessert werden kann das Konvergenzverhalten noch durch folgende zusätzliche Möglichkeit bei Verwendung des Fehlerfilters. Wird in der Anfangsphase der Identifikation bei vielen hintereinanderfolgenden Meßwerten ein instabiles Modell geschätzt, so ist die Wahrscheinlichkeit, daß hierbei noch ein konvergentes Modell entsteht, relativ klein. Durch ein zusätzliches „Anstoßen“ der Identifikation kann diese in der Großzahl der Fälle also nur verbessert werden. Deshalb wird nach 30 ohne Unterbrechung aufeinanderfolgenden Schritten, bei denen die Berechnung ein instabiles Modell ergab, die Parameter des Fehlerfilters c_i , $d_i = 0$ gesetzt ($i = 1, \dots, f$). Da beim rekursiven Prädiktionsfehlerverfahren Konvergenz erst nach vielen Meßwerten auftritt, bei den hier durchgeführten Rechnungen erst nach frühestens 1000, kann dieses „Anstoßen“ für die ersten 1000 Meßwerte erlaubt werden, sofern genügend Meßwerte zur Verfügung stehen.

Mit einem rekursiven Fehlerfilter und $\mathbf{P}_{jj}(0) \ll 1; (\forall d_i; i = 1, \dots, f)$ – der Einfluß der Parameter d_i auf das Modell nimmt sehr langsam zu – blieb das Modell stabil, solange die Schätzung von d_i keinen im Vergleich zu $C(z^{-1})$ signifikanten Beitrag zum Modell lieferte. Erst als sich die geschätzten Werte d_i deutlicher auswirkten, wurde und blieb das geschätzte Modell instabil.

2.4 Konvergenzverhalten bei Rauscheinfluß

Beim rekursiven Prädiktionsfehlerverfahren (RPDM) treten bezüglich des Konvergenzverhaltens der Identifikation folgende Möglichkeiten auf:

1. Es kann auf die Dauer (d.h. mit den zur Verfügung stehenden Meßdaten) kein stabiles Modell gefunden werden.
2. Divergenz: Es wird zwar ein stabiles Modell berechnet, aber einige Modellparameter steigen exponential während des Untersuchungszeitraumes (oder auch monoton) und laufen auseinander. Dies gilt vor allem für die Parameter der statischen Nichtlinearität, die in diesem Fall nicht invertierbar ist.
3. Keine Konvergenz: Die Werte der geschätzten Parameter springen entweder auch nach langer Identifikationsdauer noch stark hin und her oder sie oszillieren periodisch. Für diesen Fall sind bei periodischen Eingangssignalen zwei Hauptursachen möglich:
 - (a) Das geschätzte Modell ist über mehrere Zeitpunkte instabil.
 - (b) Der Steigungsparameter $\Delta\lambda$ des Vergessensfaktorverlaufes $\lambda(k)$ wurde zu klein gewählt. Eine Erhöhung von $\Delta\lambda$ bedeutet aber nicht zwangsläufig eine bessere Konvergenz der Schätzung.
4. Konvergenz, aber „falsches“ Modell: Beim Prädiktionsfehlerverfahren soll, wie bei vielen anderen Identifikationsverfahren auch, die Verlustfunktion

$$V_V = \frac{1}{2} \sum_{k=1}^N e^2(k | \Theta) \quad (2.34)$$

minimiert werden. Bei der hier auftretenden Anzahl von Modellparametern des Parametervektors Θ ist es einsichtig, daß die Verlustfunktion V_V ein globales Minimum, aber auch mehrere lokale Minima besitzen kann. Bei einem gegen 1 gehenden Vergessensfaktor ($\lambda(k) \approx 1$) ist es der Schätzung trotz Berechnung einer Suchrichtung \mathbf{L} (Gl. (2.8)) oft nicht mehr möglich, den Bereich eines „schlechten“ Minimums zu verlassen und in ein „gutes“ Minimum zu gehen. Ein Modell ist eindeutig „falsch“, wenn das Übertragungsverhalten überhaupt nicht wiedergegeben wird. Ob ein Minimum dagegen „gut“ ist, hängt von den Genauigkeitsanforderungen ab, die an eine Modellschätzung gestellt werden. Deshalb ist der Übergang zum nächsten Punkt auch fließend.

5. Konvergenz gegen „richtiges“ Modell: Die geschätzten Modellparameter müssen nicht zwangsläufig zum globalen Minimum der Verlustfunktion gehören. Das Modell muß aber das Übertragungsverhalten qualitativ wiedergeben und gestellte Anforderungen bezüglich der Modellgenauigkeit erfüllen.

Ein dem Ausgangssignal additiv überlagertes Rauschsignal besitzt natürlich genauso einen erheblichen Einfluß auf das Konvergenzverhalten der Schätzung wie die Anfangswerte des Eingangssignals (siehe Abschnitt 2.3).

Mit zunehmender Stärke des Rauschanteils wird auch die Konvergenz der Schätzung schlechter. Dieser Einfluß soll durch Berücksichtigung des Fehlerfilters mit geeigneter Wahl der Ordnung f vermindert werden. Wie diese Ordnung f zu wählen ist, darüber kann keine genauere Aussage getroffen werden. Sie muß jedoch „ausreichend groß“ sein (Young 1984). Dies sollte in der Regel $f > 2 \dots 3 n$ sein.

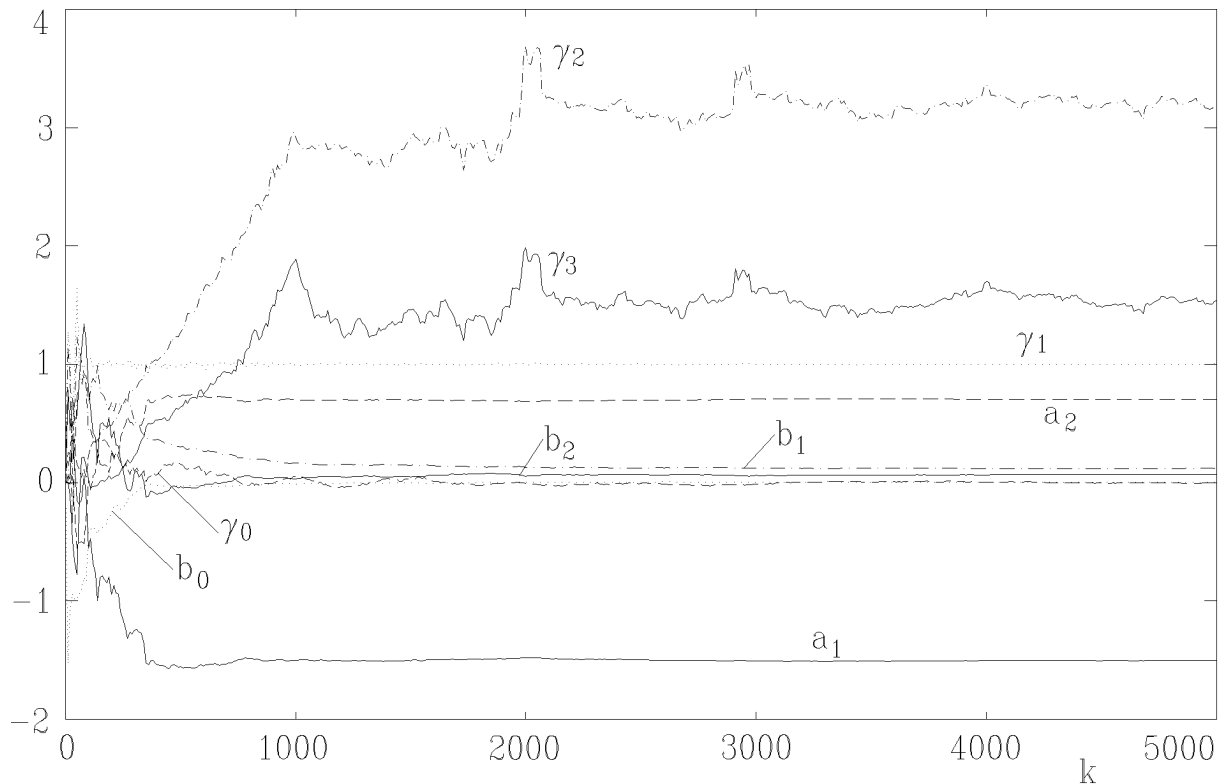


Bild 2.4: Konvergenzverhalten einer Schätzung mit Störung und Fehlerfilter

Der Einfluß des Rauschsignals auf das Konvergenzverhalten kann auch durch Erhöhung des Abtastrates bei gleichbleibender Taktzeit oder umgekehrt erreicht werden. Für das weiter oben aufgeführte Modell wurden deshalb pro Takt des PRBS 5 Abtastschritte gewählt mit $u = \pm 0.5$. Zudem betragen $\lambda(0) = 0.985$, $\Delta\lambda = 0.0008$ und $f = 10$. $D(z^{-1}) = 1$ ergab dann für $\mathbf{P}_{jj}(0) = 0$ ($\forall d_i$) bei $\mathbf{P}_{jj}(0) = 100$ ($\forall a_i, b_i, \gamma_i, c_i$). Als zusätzliche Störung wurde dem Ausgangssignal ein normalverteiltes Rauschsignal r mit Standardabweichung $\sigma(r) = 0.2$ additiv überlagert, so daß sich mit $\sigma(y) = 1.1$ ein Störsignal-Nutzsignalverhältnis $V = \sigma(r)/\sigma(y) = 0.182$ ergab (Reuter 1992a). Wegen der Identifizierbarkeit wurde während der Identifikation $\gamma_1 = 1$ gehalten.

Parameter	a_1	a_2	b_0	b_1	b_2	γ_0	γ_1	γ_2	γ_3
Schätzung	-1.5024	0.702	0.0041	0.1202	0.0686	-0.0041	1.0	3.2022	1.539
Modell	-1.5	0.7	0	0.1333	0.0667	0	1	3	1.5

Tabelle 2.1: Gegenüberstellung der geschätzten und der vorgegebenen Modellparameter

Bei 20 verschiedenen Rauschsignalen ergaben sich nach 5000 Werten in fünf Fällen gute konvergente Schätzungen. Eines dieser fünf Ergebnisse zeigen die nächsten beiden Bilder 2.4 und 2.5 sowie Tabelle 2.1.

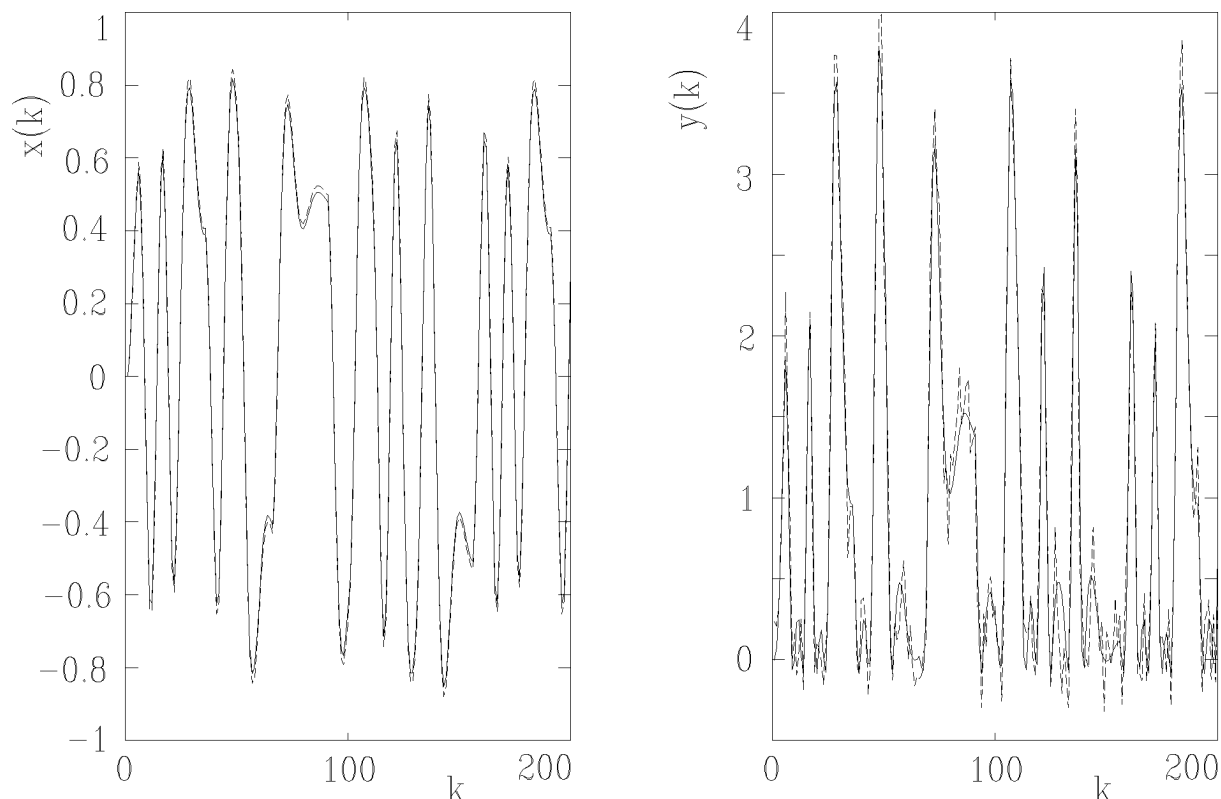


Bild 2.5: links: Zwischensignal $x(k)$ des geschätzten (–) und vorgegebenen Modells (– –)
rechts: Ausgangssignal $y(k)$ des geschätzten Modells (–) und vorgegebenes gestörtes Ausgangssignal (– –)

Bild 2.4 zeigt das Konvergenzverhalten der Schätzung. So ist zu erkennen, daß die Parameter schon nach ca. 1000 Schritten in den „richtigen“ Bereich gekommen sind, die Werte γ_2 und γ_3 aber erst für > 3000 Schritte einen ruhigen Parameterverlauf annehmen. Die Parameter von $A(z^{-1})$ waren schon nach ca. 400 Schritten annähernd konstant, die von $B(z^{-1})$ nach etwa 1000 Schritten. In vielen anderen Fällen war Konvergenz erst viel später zu erkennen (≥ 3000 Schritte). Bild 2.5 zeigt links für 200 Werte eine Gegenüberstellung des Ausgangs des linearen Teilmodells $x(k)$ für das geschätzte Modell (–) und das vorgegebene Modell (– –) sowie rechts für das mit dem beim letzten Schritt geschätzten

Parametersatz erhaltene Ausgangssignal (-) und das vorgegebene gestörte Ausgangssignal (- -). Die anderen vier „gut“ geschätzten Modelle wiesen ein ähnliches Übertragungsverhalten auf.

Zu bemerken ist hierbei noch, daß von den 5000 Schätzungen 18 instabil waren. Bei den anderen vier „guten“ Modellen wurden manchmal über 100 instabile Modelle geschätzt und teilweise die Parameter c_i des Fehlerfilters ein- oder zweimal zu Null gesetzt.

Für die anderen 15 Simulationen konnte in einem Fall kein stabiles Modell gefunden werden. Achtmal divergierten die Parameter γ_2 und γ_3 , viermal konvergierten die Modelle zu falschen Parametern und in 2 Fällen waren die Parameter nach 5000 Meßwerten noch nicht konvergiert, sie lagen aber in der Nähe der tatsächlichen Parameter und gaben qualitativ das Übertragungsverhalten schon recht gut wieder.

Bei 20 Vergleichsrechnungen ohne Berücksichtigung des Fehlerfilters stellten sich sehr ähnliche Ergebnisse bezüglich des Konvergenzverhaltens und der Modellgenauigkeit ein.

3 Bekannte Kriterien zur Wahl der Modellgröße

3.1 Vorstellung der Kriterien

Bei der Anwendung der meisten Identifikationsverfahren wird vorausgesetzt, daß die „Größe“ des Systemmodells bekannt ist. Dies ist jedoch bei praktischen Anwendungen häufig nicht der Fall, da wegen der Komplexität des zu beschreibenden Prozesses eine physikalische Modellbildung oft nicht durchgeführt werden kann und somit keine a-priori-Informationen über die Modellstruktur vorliegen. Die Bestimmung der Modellgröße ist deshalb ein größeres Problem bei der Identifikation.

Das dynamische Verhalten von realen Prozessen, d.h. das der zeitlichen Entwicklung eines Prozesses unter dem Einfluß von Anfangsauslenkungen und/oder einwirkenden Eingangsgrößen, ist in der Umladung von Speichern für Energie und/oder Masse und/oder Information begründet. Die Anzahl dieser Speicher findet sich bei Zustandsraummodellen in der Anzahl der Zustände, also der Modellordnung, wieder (Schwarz 1991). Während bei linearen und bilinearen Modellen mit der Modellordnung die Modellgröße eindeutig festgelegt ist, trifft dies auf nichtlineare Modelle mit anderer Modellstruktur oft nicht zu. Bei Wiener- und Hammerstein-Modellen ist mit der Modellordnung lediglich die Größe des linearen dynamischen Übertragungsgliedes festgelegt. Für die Beschreibung der statischen Nichtlinearität ist zusätzlich die Angabe einer Polynomordnung notwendig. Auch bei den zeitdiskreten Kolmogorov-Gabor-Polynommodellen (Kortmann 1989) wird die Modellgröße durch diese Modellordnung nur sehr unzureichend beschrieben. Hier muß eine Polynomordnung im Sinne einer maximalen Ordnung der multiplikativen Verknüpfung von Ein- und Ausgangsgrößen vorgegeben werden. Aber auch die Verwendung der Anzahl der zu schätzenden Parameter als Definition für die Modellgröße ist zu grob und zu ungenau.

Betrachtet man die allgemeiner bekannten Kriterien zur Auswahl der Modellordnung (Wernstedt 1989, Unbehauen 1988, Isermann 1988), so kann man unterscheiden zwischen algorithmisch dominierten und heuristisch dominierten Verfahren sowie zwischen Verfahren, die nur auf lineare Systemmodelle anwendbar sind und allgemeiner anwendbaren.

- Für lineare Modelle anwendbar ist der Polynomtest. Bei einer Übertragungsfunktion ist die Modellordnung zu groß, wenn Pole durch Nullstellen kompensiert werden. Da solche Pol- und Nullstellen häufig nicht direkt aufeinander liegen, muß der Anwender aufgrund der Pol- und Nullstellenlage visuell und subjektiv entscheiden, ob eine Kompensation vorliegt oder nicht.
- Als leistungsfähig für lineare Modelle hat sich der kombinierte Polynom- und Dominanztest erwiesen (Wernstedt 1989, Unbehauen 1988), bei dem ausgehend von einer

maximalen Ordnung der Übertragungsfunktion Dominanzmaße für die einzelnen Pole herangezogen werden.

- Ein geeignetes neues Kriterium beschreibt Mindel (1992). Ausgehend von einer minimalen Ordnung eines zeitdiskreten Modells zieht er die Pollage in Verbindung mit Eigenschaften der z-Transformation heran.
- Auf allgemeine Modelle anwendbar ist der Signalfehlerstest (Wernstedt 1989), bei dem der Anwender einen optischen Vergleich von gemessenem und identifiziertem Signalverlauf vornimmt und subjektiv entscheidet.
- Ungenau sind auch Determinanztests, die lediglich eine grobe Vorabschätzung der Modellordnung liefern (Unbehauen 1987).

Kortmann (1989) zieht zur Beurteilung der Modellqualität zum einen Anpassungsgütemaße und zum anderen statistische Bewertungskriterien zur Bestimmung der Signifikanz einzelner Parameter im Kolmogorov-Gabor-Polynom heran. Auch die Suche nach der Modellordnung ist ein sehr ähnlich gelagertes Problem, da hier untersucht wird, ob eine höhere Modellordnung ein signifikant besseres Modell ergibt. Nachteil der statistischen Tests ist, daß sie den Gleichungsfehler $e(k) = y(k) - \hat{y}(k)$ mit $y(k)$ als gemessenem und $\hat{y}(k)$ als geschätztem Ausgangssignal als normalverteilte, unkorrelierte und mittelwertfreie Zufallsvariable voraussetzen, was jedoch oft nicht erfüllt ist.

Da ein Modell um so genauer ist, je kleiner der Fehler ist, bietet sich die Verlustfunktion

$$V = \frac{1}{N} \sum_{k=1}^N e^2(k) \quad (3.1)$$

mit N als Anzahl der berücksichtigten Meßwerte an. Günstig hierbei ist eine Modellgröße dann, wenn sich für höhere Modellordnungen keine deutliche Verbesserung, aber für kleinere Modellordnungen eine deutliche Verschlechterung des Verlustfunktionswertes ergibt. Nachteile hierbei sind, daß mit steigender Modellkomplexität bei ungestörten Signalen die Verlustfunktion immer kleiner wird und es keine obere Schranke gibt.

Den Nachteil fehlender Grenzen besitzt das Bestimmtheitsmaß (Sachs 1978)

$$R^2 = \frac{\sum_{k=1}^N (\hat{y}(k) - \bar{y})^2}{\sum_{k=1}^N (y(k) - \bar{y})^2} \quad (3.2)$$

mit \bar{y} als Mittelwert des geschätzten Ausgangssignals $\hat{y}(k)$ nicht mehr. Mit $\hat{y}(k) = y(k)$ ergibt sich $R^2 = 1$ für ein im Sinne eines Fehlers optimales Modell. Die untere Grenze mit $R^2 = 0$ erhält man bei $\hat{y}(k) = \bar{y}$. Dieses Bestimmtheitsmaß ist jedoch nicht definiert für eine konstantes Meßsignal. Ein im Sinne dieses Bestimmtheitsmaßes optimales Modell

ist also unendlich komplex, da $R_{n_p+1}^2 > R_{n_p}^2$ mit n_p als der Anzahl der Modellparameter. Günstig wäre ein Optimum zwischen einem möglichst genauen aber auch möglichst einfachen Modell.

Den Nachteil, mit steigender Modellkomplexität immer größer zu werden, besitzen die korrigierten Bestimmtheitsmaße nach Theil (1971)

$$R_T^2 = 1 - (1 - R^2) \frac{N - 1}{N - n_p} \quad (3.3)$$

und Amemiya (1980)

$$R_A^2 = 1 - (1 - R^2) \frac{N + n_p}{N - n_p} \quad (3.4)$$

nicht mehr. Sie enthalten einen „Strafterm“ für eine zunehmende Parameterzahl n des Modells. So kann durch die Aufnahme eines zusätzlichen Parameters, der die Modellgenauigkeit nur geringfügig verbessert, das größere Modell als ungünstiger beurteilt werden als das ursprüngliche. Bei dem korrigierten Bestimmtheitsmaß nach Amemiya wird die Modellkomplexität stärker bestraft. Verwendet man aber eine große Anzahl N von Meßwerten, was bei komplexeren Modellen der Fall ist, so wird hier der Einfluß der Parameteranzahl stark vermindert und unbedeutend.

Der F-Test (Sachs 1978) überprüft für zwei unabhängig gewonnene Zufallsstichproben ihre Varianzen auf Gleichheit oder Homogenität. Die Nullhypothese $\mathbf{H}_0 : \sigma_1^2 = \sigma_2^2$ (beide Varianzen sind gleich) wird abgelehnt, sobald ein aus den Stichprobenvarianzen berechneter Wert $\hat{F} = \sigma_1^2 / \sigma_2^2$ mit einer Irrtumswahrscheinlichkeit α größer ist als ein zugehöriger Tabellenwert. In diesem Fall wird eine Alternativstichprobe akzeptiert. Hierbei wird $\sigma_1^2 > \sigma_2^2$ vorausgesetzt mit dem zugehörigen Tabellenwert $F_{krit}(\alpha, \nu_1, \nu_2)$ der F-Verteilung (ν_1 : Freiheitsgrad des Zählers, ν_2 : Freiheitsgrad des Nenners).

Auf die Bestimmung der Modellgröße angewendet untersucht man, ob sich die Summe des quadratischen Gleichungsfehlers eines Modells M_1 mit n_{p1} Parametern im statistischen Sinne nur unwesentlich von der Summe des quadratischen Gleichungsfehlers eines größeren Modells M_2 mit n_{p2} Parametern unterscheidet. Dabei soll die Modellstruktur von M_1 in der von M_2 enthalten sein. Das hierzu gehörige Hypothesenpaar lautet:

$$\left. \begin{array}{l} \mathbf{H}_0 : \text{Das Modell } M_2 \text{ ist nicht signifikant besser als } M_1 \text{ (Nullhypothese)} \\ \mathbf{H}_1 : \text{Das Modell } M_2 \text{ ist signifikant besser als } M_1 \text{ (Alternativhypothese)} \end{array} \right\} .(3.5)$$

Anschließend wählt man eine Irrtumswahrscheinlichkeit α und berechnet den Wert der Testgröße (Kortmann 1989, Unbehauen 1988)

$$F = \frac{V_1 - V_2}{V_2} \frac{N - n_{p2}}{n_{p2} - n_{p1}} \quad (3.6)$$

V_1 bezeichnet den Wert der Verlustfunktion des Modells M_1 und V_2 den von M_2 . Unter der Annahme, daß der Fehler $\epsilon(k)$ normalverteilt und unkorreliert ist, ergibt sich für V_2 und $V_1 - V_2$ eine voneinander unabhängige χ^2 -Verteilung. Die Testgröße F besitzt somit eine F-Verteilung.

Die Nullhypothese \mathbf{H}_0 wird abgelehnt, wenn

$$F > F_{krit}(\alpha, n_{p2} - n_{p1}, N - n_{p2}) \quad (3.7)$$

verifiziert wird. Die zusätzlichen Parameter bewirken also eine als signifikant erkannte Verbesserung des Übertragungsverhaltens. Die Grenzwerte $F_{krit}(\alpha, n_{p2} - n_{p1}, N - n_{p2})$ können entsprechenden Tabellen entnommen werden (Bronstein 1984, ausführlicher bei Sachs 1978).

Wird die Nullhypothese \mathbf{H}_0 abgelehnt, obwohl sie richtig ist, spricht man von einem Fehler erster Art. Einen Fehler in entgegengesetzter Richtung begeht man, wenn die Hypothese \mathbf{H}_0 angenommen wird, obwohl die Alternativhypothese \mathbf{H}_1 richtig ist. Dies wird als Fehler zweiter Art mit der Auftretenswahrscheinlichkeit β bezeichnet (Fisz 1971). Allgemein gilt (Sachs 1978, Kortmann 1989):

Das Risiko eines Fehlers zweiter Art sinkt bei gegebener Irrtumswahrscheinlichkeit α mit wachsender Anzahl der Meßwerte N . Je größer also N , um so leichter lassen sich Abweichungen von der Nullhypothese erkennen. Und das Risiko eines Fehlers zweiter Art sinkt bei sonst gleichen Bedingungen mit wachsender Irrtumswahrscheinlichkeit α . Die Wahl von α ist also schwierig. Zum möglichst sicheren Nachweis der Alternativhypothese \mathbf{H}_1 sollte α so klein wie möglich gewählt werden. Andererseits sollte α aus statistischen Gründen in Abhängigkeit von N so groß wie möglich werden, denn eine Vergrößerung von α bewirkt eine Verkleinerung von β , also eines Fehlers zweiter Art.

Die Informationskriterien bestehen in ihrer allgemeinen Form

$$I_{krit} = N \ln(V) + f(n_p, N) \quad (3.8)$$

aus zwei Teilen. Der erste Teil bewertet die Approximation des gemessenen Signals durch das geschätzte Ausgangssignal und der zweite die Komplexität des Modells, also die Anzahl der Parameter. Bei steigender Komplexität des Modells sinkt der Wert für den ersten Teil, während der für den zweiten Teil steigt. Somit kennzeichnet das erste Minimum von I_{krit} das beste Modell. Kortmann (1989) verwendet – ausgehend von einem Grundkriterium

$$I_0 = N \ln(V) \quad (3.9)$$

ohne Strafterm – die folgenden vier Informationskriterien:

- FPE ('Final Prediction Error')

$$FPE = N \ln(V) + N \ln\left(\frac{N + n_p}{N - n_p}\right) \quad (3.10)$$

- AIC ('Akaike's Information Criterion')

$$AIC(\Phi_1) = N \ln(V) + \Phi_1 n_p \quad \text{mit } \Phi_1 > 0 \text{ ; oft günstig: } \Phi_1 = 4 \quad (3.11)$$

- LILC ('Khinchin's Law of the Iterated Logarithm Criterion')

$$LILC(\Phi_2) = N \ln(V) + 2n_p \Phi_2 \ln(\ln(N)) \quad \text{mit } \Phi_2 \geq 1 \quad (3.12)$$

- BIC ('Bayesian Information Criterion')

$$BIC = N \ln(V) + n_p \ln(N) \quad (3.13)$$

3.2 Ergebnisse in der praktischen Anwendung

Für ein an einem elektrohydraulischen Translationsantrieb (Köckemann 1988, Dorißen 1989, Schwarz 1991) mittels eines rekursiven Least-Squares-Algorithmus identifiziertes bilineares Modell in beobachtbarkeitskanonischer Form (Reuter 1992b) werden nun die Ergebnisse der Ordnungssuche dargestellt. Dies soll anhand der Verlustfunktion, der Bestimmtheitsmaße, des F-Tests sowie der Informationskriterien geschehen.

Von den 3000 Meßwerten dieser Reihe wurden lediglich die letzten 2500 herangezogen. Während der ersten 500 Meßwerte befanden sich die Modellparameter noch in der Einschwingphase des rekursiven Identifikationsalgorithmus. Einzelne dort auftretende Fehlerspitzen besaßen so einen deutlichen störenden Einfluß auf die Verlustfunktionswerte für die einzelnen Modellordnungen und somit auch auf die Auswahlkriterien. Identifiziert wurden Modelle mit aufsteigender Ordnung von $n = 1 \dots 10$.

Unter Berücksichtigung eines genaueren physikalischen Modells haben Yin (1992) sowie Yin, Ingenbleek und Schwarz (1992) mit Vorgabe schon relativ genauer Anfangsparameter zeitkontinuierliche bilineare Modelle dieser Anlage mit der Modellordnung $n = 4$ bei einem Geschwindigkeitsmodell identifiziert.

Für die Verlustfunktion (Gl. (3.1)) über der Modellordnung ergibt sich der in Bild 3.1 gezeigte Verlauf. In der linken Bildhälfte ist zu erkennen, daß die ermittelten Verlustwerte für die Modelle der Ordnungen $n = 1$ und 2 sich von den übrigen um über eine Größenordnung unterscheiden. Betrachtet man das rechte Teilbild, so bieten sich zwei Ordnungen an. Dies sind die Ordnungen $n = 4$ (für kleinere Ordnungen wird das Modell deutlich schlechter, aber für größere nicht mehr deutlich besser) sowie $n = 7$ (die Verlustfunktion besitzt hier ein lokales Minimum und für größere Ordnungen sind erkennbare Verbesserungen nur minimal).

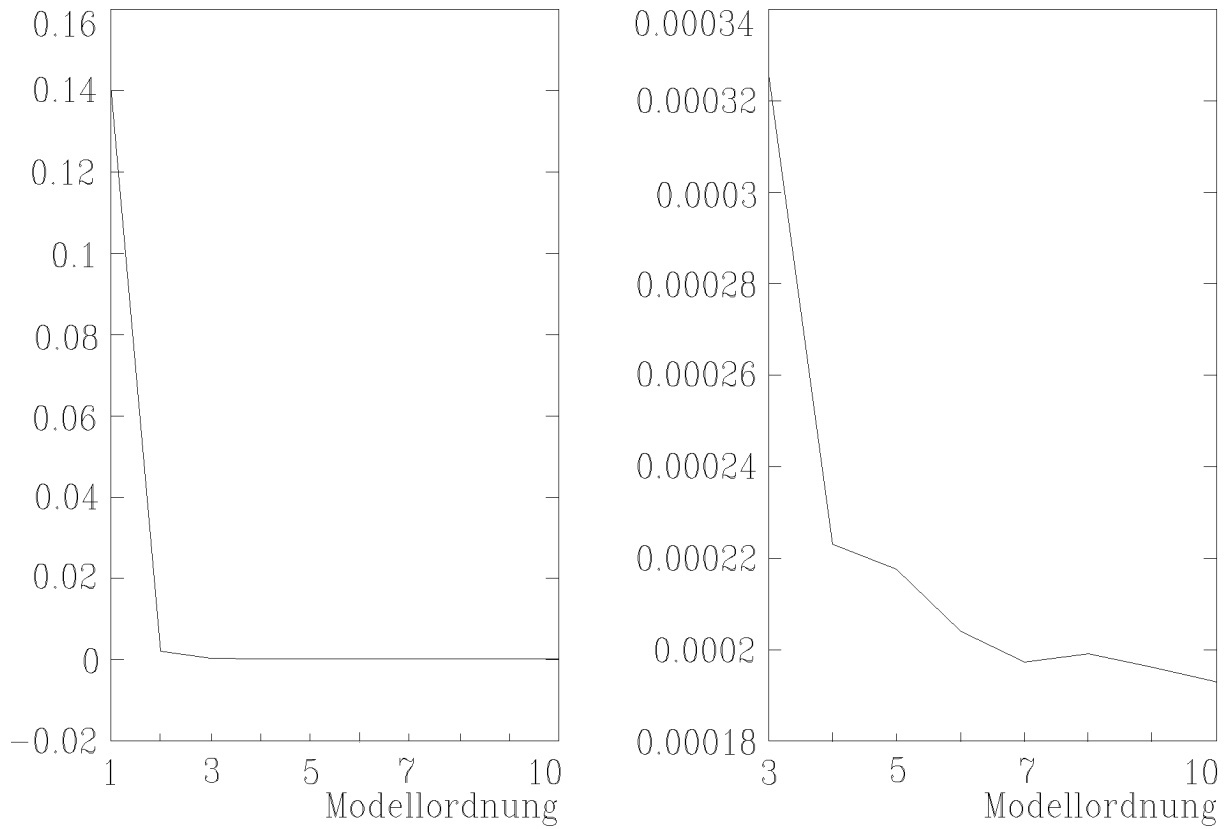


Bild 3.1: Werte der Verlustfunktion V über der Modellordnung

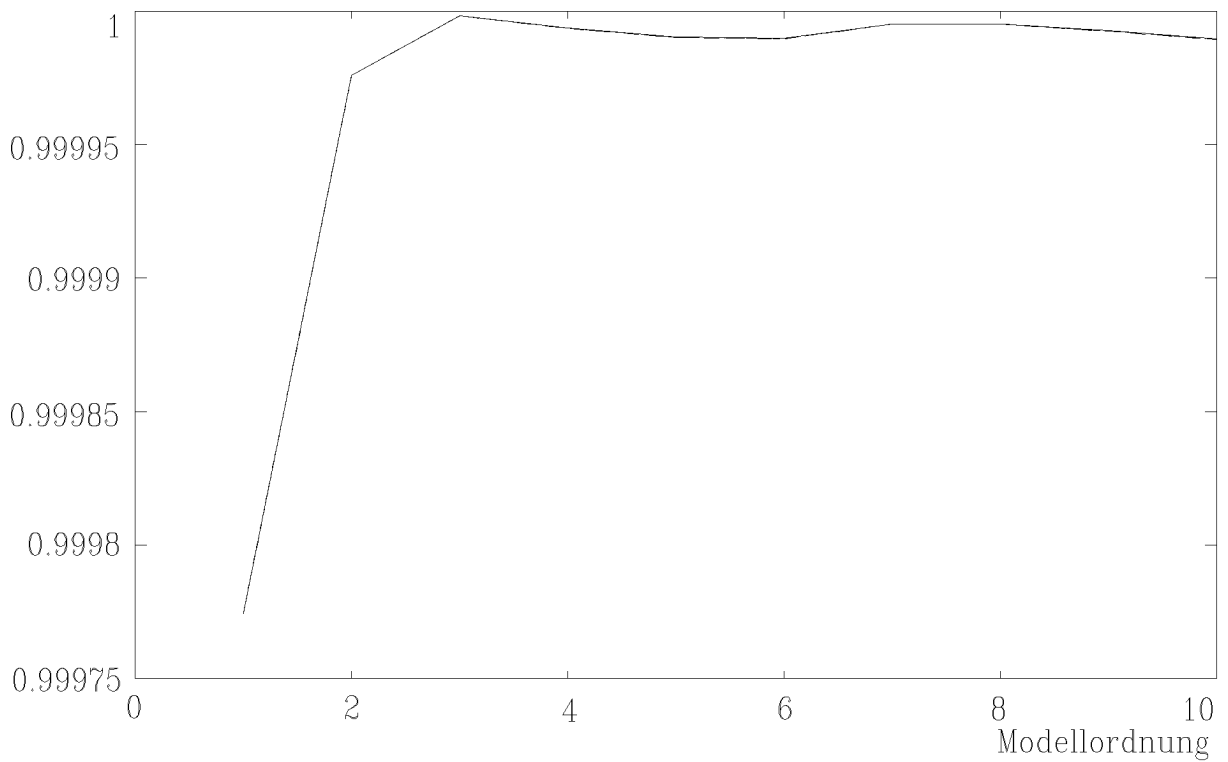


Bild 3.2: Werte der Bestimmtheitsmaße R^2 , R_T^2 und R_A^2

Wie in Bild 3.2 zu sehen ist, gibt es für das ursprüngliche Bestimmtheitsmaß R^2 und die korrigierten R_T^2 und R_A^2 keine erkennbaren Unterschiede. So wird wegen der großen Anzahl notwendiger Meßwerte bei im Vergleich dazu sehr niedriger Anzahl von Parametern der Einfluß der „Strafterme“ insignifikant klein. Auch wird hierbei eine günstige Modellordnung $n = 3$ angedeutet, was im Widerspruch zur Verlustfunktion sowie den anderen Ergebnissen steht.

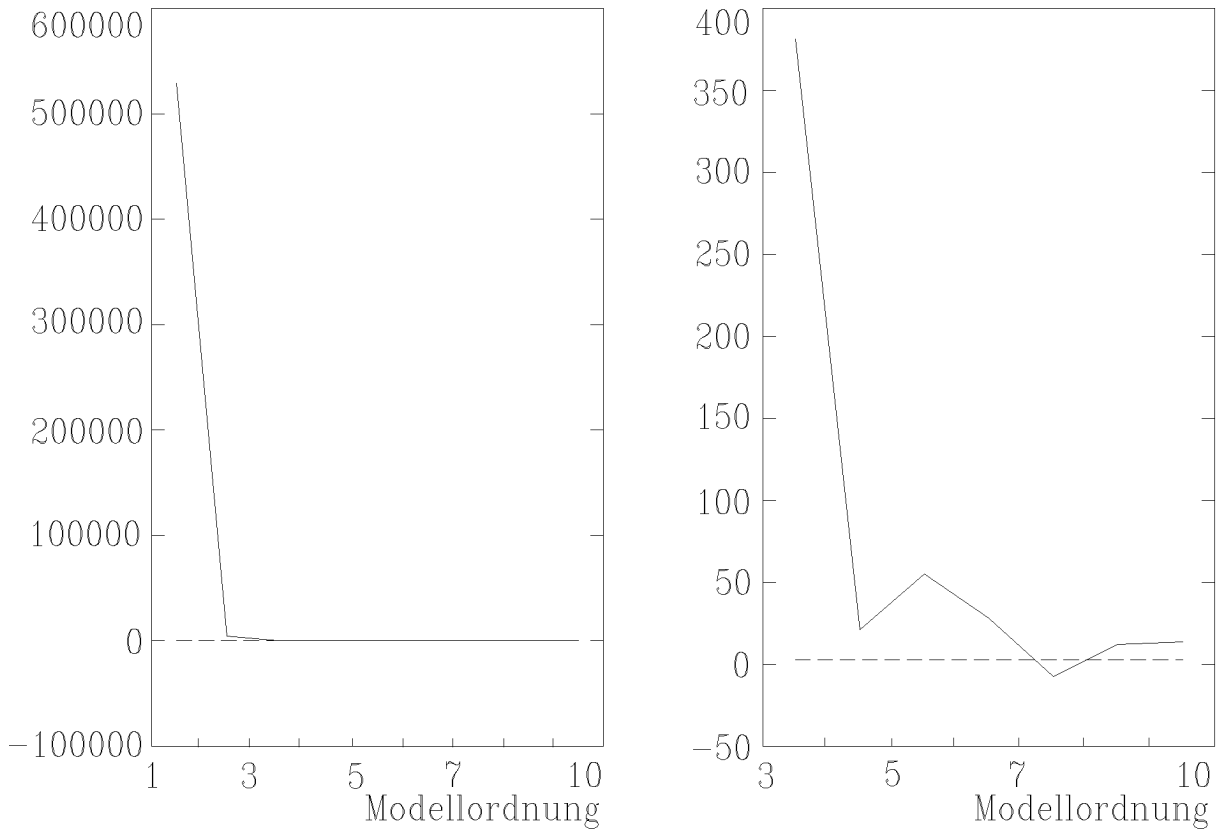


Bild 3.3: Werte der Ordnungssuche mittels F-Test

Bild 3.3 veranschaulicht die Ergebnisse des F-Tests. Das linke Teilbild verdeutlicht wieder die ungünstigen Werte für die Ordnungen $n = 1$ und 2 . Hierbei sind die Werte nicht direkt über der Modellordnung aufgetragen. Die Erhöhung der Modellordnung von $n = 1$ auf $n = 2$ (aufgetragen bei $n = 1.5$) ergibt eine sehr deutliche Verbesserung des Übertragungsverhaltens. Das rechte Teilbild zeigt für die Verwendung der Anzahl n_p der Modellparameter zur Bestimmung des Freiheitsgrades des Signifikanzgrenzwertes F_{krit} (- -) eine optimale Modellordnung $n = 7$. Für eine Erhöhung der Modellordnung von $n = 7$ auf $n = 8$ ist $F < F_{krit}$ (das Modell wird nicht signifikant besser), es wird wegen $F < 0$ sogar schlechter. Ein anderes Minimum wäre $n = 4$, jedoch liegt hierbei $F > F_{krit}$. Berücksichtigt man jedoch nicht die Anzahl n_p der Parameter zur Bestimmung von F_{krit} , sondern direkt die Modellordnung n (bei dem hier identifizierten bilinearen Modell in kanonischer Form verursacht eine Erhöhung der Modellordnung um $\Delta n = 1$ eine Erhöhung der Mo-

dellparameter um $\Delta n_p = 3$), so läge F_{krit} hierfür deutlich höher und $n = 4$ wäre die so bestimmte optimale Modellordnung.

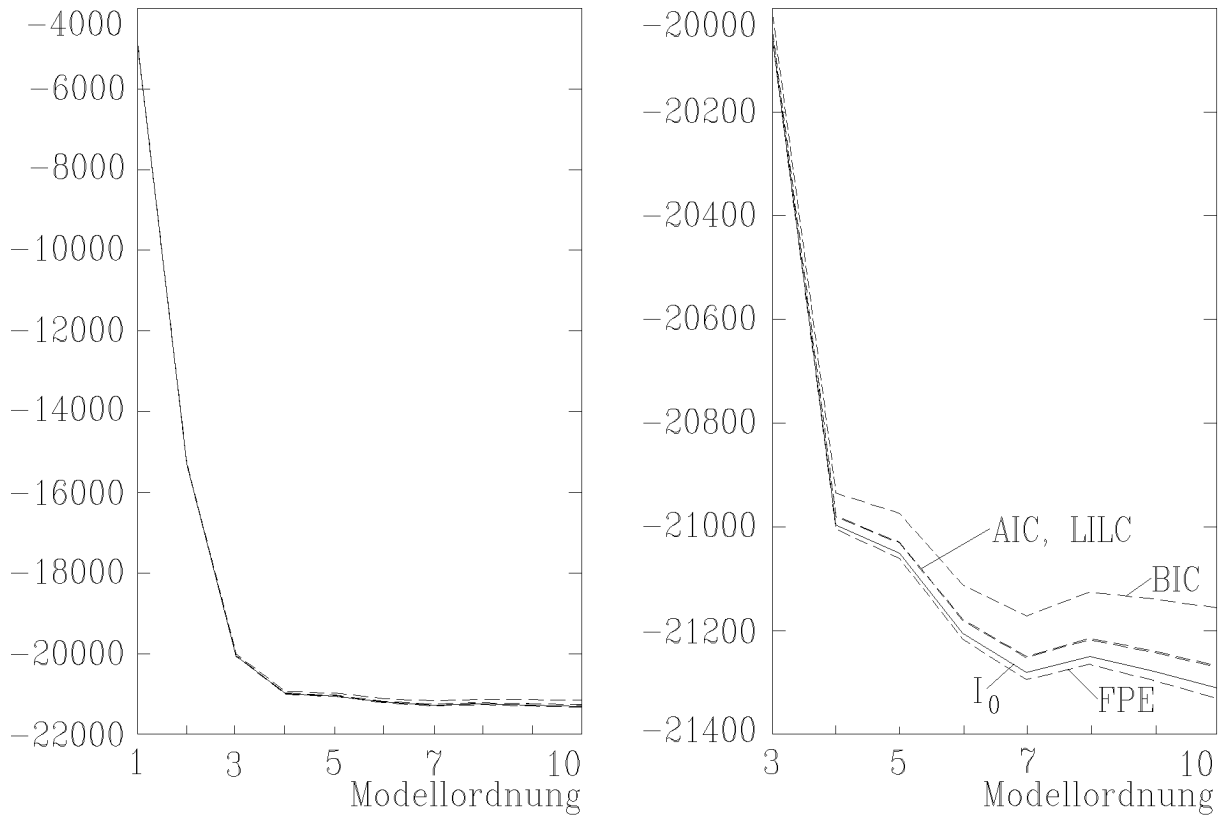


Bild 3.4: Werte der Ordnungssuche für die Informationskriterien

Die Ergebnisse der Informationskriterien zeigt Bild 3.4. Das linke Teilbild verdeutlicht, daß auch hier die Strafterme nur einen sehr geringen Einfluß besitzen. Auf dem rechten Teilbild erkennt man, daß sowohl I_0 (—) als auch die vier Informationskriterien (- -) ihr erstes lokales Minimum bei $n = 7$ besitzen. Der Einfluß der Terme für die Modellgröße ist hier somit vernachlässigbar klein. Bei der Verwendung von Meßwerten mit größeren Beträgen und/oder beispielsweise eines einfachen LS-Verfahrens zur Identifikation (und kein RLS-Verfahren) ergeben sich größere Fehlerwerte und somit auch größere Werte für I_0 (man bekäme bei diesem Anwendungsbeispiel eine optimale Modellordnung $n = 4$). Somit erhalten die Terme $f(n_p, N)$ (Gl. (3.8)) eine ganz andere Gewichtung bezüglich der Wahl der Modellordnung. Bei diesen Kriterien ist also eine starke Abhängigkeit vom Betrag der Verlustfunktion V zu erkennen.

4 Zusammenfassung

In diesem Bericht sind die Erkenntnisse des Autors bei der Anwendung des rekursiven Prädiktionsfehlerverfahrens am Beispiel eines Wiener-Modells sowie über die Verwendung bekannter Kriterien zur Auswahl einer optimalen Modellgröße bei der Identifikation nichtlinearer Systemmodelle zusammengefaßt.

Das Prädiktionsfehlerverfahren eignet sich gerade auch für die Identifikation nichtlinearer Systemmodelle, da hierbei, im Gegensatz z.B. zu den Least-Squares-Verfahren, die Linearität des Modells in den Parametern nicht gefordert wird.

Bezüglich der Konvergenz müssen jedoch gewisse Anforderungen an Signale und Anfangswerte gestellt werden. So führen beliebige Eingangssignalfolgen und dem Ausgangssignal additiv überlagerte Rauschanteile nicht zwangsläufig zu konvergierenden Modellparametern. Zudem muß eine hohe Anzahl von Meßsignalen zur Verfügung stehen. Für den rekursiven Identifikationsalgorithmus ist neben der Wahl der Diagonalmatrix $\mathbf{P}(0)$ der Zeitverlauf des Vergessensfaktors $\lambda(k)$ von sehr wichtiger Bedeutung, für deren Wahl hier Hinweise gegeben werden konnten, da sie einen starken Einfluß auf die Konvergenz der Schätzung besitzen. Für verrauschte Meßsignale ließ sich jedoch kein erkennbarer Unterschied zwischen der Identifikation mit und ohne Berücksichtigung eines nichtrekursiven Fehlerfilters feststellen.

Bei den Kriterien zur Wahl einer optimalen Modellordnung haben sich am Beispiel eines elektrohydraulischen Translationsantriebes deutliche Schwächen herausgestellt. So wird bei der Anwendung der meisten Kriterien an realen Prozessen eine gewisse Erfahrung benötigt, die Nicht-Experten in der Regel nicht besitzen.

Die Bestimmtheitsmaße geben keine Hilfe bei komplexen Modellen mit einer größeren Parameterzahl, welche häufig vorkommen, da bei der damit verbundenen großen Anzahl von notwendigen Meßwerten der Einfluß der Parameteranzahl de facto herausfällt. Auch die Ergebnisse des F-Tests sind stark von der Anzahl N der Meßwerte beeinflußt. Zudem wird hier eine Normalverteilung der Fehler vorausgesetzt, die häufig nicht gegeben ist. Schon bei geringen Abweichungen der Fehler von der Normalverteilung ist aber auch die Zuverlässigkeit dieser Tests nicht mehr gegeben. Bei den Informationskriterien zeigt sich hier eine stärkere Abhängigkeit der Ergebnisse (günstige Modellordnung) von der absoluten Größe der Verlustfunktion.

An den Kriterien zur Wahl einer optimalen Modellgröße müssen noch stärkere Verbesserungen vorgenommen werden, um den Einfluß der Anzahl N der Meßwerte sowie den der absoluten Werte der Verlustfunktion zu eliminieren.

5 Literaturverzeichnis

- Amemiya, T.** 1980. Selection of regressors. *International Economic Review* 21. 331-354.
- Bronstein, N. und K. A. Semendjajew.** 1984. *Taschenbuch der Mathematik*. Thun, Frankfurt/Main: Verlag Harri Deutsch.
- Dorißen, H. T.** 1990. *Zur Minimalrealisierung und Identifikation bilinearer Systeme durch Markovparameter*. Diss. Universität –GH– Duisburg. Fortschr.-Ber. VDI Reihe 8 Nr. 221. Düsseldorf: VDI-Verlag.
- Fisz, M.** 1971. *Wahrscheinlichkeitsrechnung und mathematische Statistik*. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Isermann, R.** 1988. *Identifikation dynamischer Systeme I,II*. Berlin, Heidelberg: Springer-Verlag.
- Köckemann, A.** 1988. *Zur adaptiven Regelung elektro-hydraulischer Antriebe*. Diss. Universität –GH– Duisburg. Fortschr.-Ber. VDI Reihe 8 Nr. 174. Düsseldorf: VDI-Verlag.
- Kortmann, M.** 1989. *Die Identifikation nichtlinearer Ein- und Mehrgrößensysteme auf der Basis nichtlinearer Modellansätze*. Fortschr.-Ber. VDI Reihe 8 Nr. 177. Düsseldorf: VDI-Verlag.
- Ljung, L.** 1987. *SYSTEM IDENTIFICATION: Theory for the User*. Englewood Cliffs: The MIT Press.
- Ljung, L. und T. Söderström.** 1987. *Theory and practice of recursive identification*. Cambridge/Massachusetts, London: Prentice Hall, Inc.
- Mindel, K.** 1992. *Integration von analytischem und heuristischem Wissen in ein System zur experimentellen Modellbildung*. Vom Fachbereich Maschinenbau zur Veröffentlichung freigegebene Dissertation. Universität –GH– Duisburg.
- Reuter, H.** 1992a. *Verfahren zur Untersuchung der Nichtlinearität dynamischer Systeme*. Forschungsbericht MSRT Nr. 03/92. Universität – GH – Duisburg.
- Reuter, H.** 1992b. *Zur Identifikation bilinearer Systemmodelle mit einem robusten Least-Squares-Verfahren*. Forschungsbericht MSRT Nr. 09/92. Universität –GH– Duisburg.
- Sachs, L.** 1978. *Angewandte Statistik*. 5. Auflage. Berlin, Heidelberg, New York: Springer-Verlag.
- Schwarz, H.** 1979. *Zeitdiskrete Regelungssysteme*. Braunschweig: Vieweg.

- Schwarz, H.** 1991. *Nichtlineare Regelungssysteme*. München: Oldenbourg.
- Theil, J.** 1971. *Principles of econometrics*. New York: John Wiley & Sons Inc.
- Unbehauen, H.** 1988. *Regelungstechnik III*. Braunschweig, Wiesbaden: Vieweg.
- Wernstedt, J.** 1989. *Experimentelle Prozeßanalyse*. Berlin: VEB Verlag Technik.
- Yin, X.** 1992. Bilinear Modelling and State-feedback Control of an electro-hydraulic Drive. *IFAC/IMACS/IEEE/IUTAM Workshop on Motion Control for Intelligent Automation*. Perugia/Italien, Oktober 1992.
- Yin, X., R. Ingenbleek und H. Schwarz.** 1992. Observer design for translatory hydraulic drives via parameter identification of a observer canonical form. *4th International Symposium on Systems Analysis and Simulation*. Berlin, August 1992.
- Young, P.** 1984. *Recursive Estimation and Time-Series Analysis*. Berlin, Heidelberg, New-York, Tokyo: Springer-Verlag.