Rolf Isermann

Identifikation dynamischer Systeme

Band II

Parameterschätzmethoden,
Kennwertermittlung und Modellabgleich,
Zeitvariante, nichtlineare und
Mehrgrößen-Systeme, Anwendungen

Springer-Verlag

Rolf Isermann

Identifikation dynamischer Systeme

Band II:
Parameterschätzmethoden,
Kennwertermittlung und Modellabgleich,
Zeitvariante, nichtlineare und
Mehrgrößen-Systeme, Anwendungen

Mit 83 Abbildungen

Springer-Verlag Berlin Heidelberg NewYork
London Paris Tokyo 1988

1. Dynamical systems.

Professor Dr.-Ing. Rolf Isermann
Institut für Regelungstechnik
Fachgebiet Regelsystemtechnik
TH Darmstadt
Schloßgraben 1
6100 Darmstadt

ISBN 3-540-18694-8 Springer-Verlag Berlin Heidelberg NewYork
ISBN 0-387-18694-8 Springer-Verlag NewYork Heidelberg Berlin

CIP-Kurztitelaufnahme der Deutschen Bibliothek
Isermann, Rolf:
Identifikation dynamischer Systeme / Rolf Isermann.
Berlin ; Heidelberg ; NewYork ; London ; Paris ; Tokyo : Springer
Bd. II. Parameterschätzmethoden, Kennwertermittlung und Modellablgleich,
Zeitvariante, nichtlineare und Mehrgrößen-Systeme, Anwendungen
1988
ISBN 3-540-18694-8 (Berlin . . .)
ISBN 0-387-18694-8 (NewYork . . .)

Druck: Mercedes-Druck, Berlin; Bindearbeiten: Lüderitz & Bauer, Berlin
2160/3020-543210

Vorwort

Für viele Aufgabenstellungen beim Entwurf, beim Betrieb und bei der Automatisierung technischer Systeme werden in zunehmendem Maße genaue mathematische Modelle für das dynamische Verhalten benötigt. Auch im Bereich der Naturwissenschaften, besonders Physik, Chemie, Biologie und Medizin und in den Wirtschaftswissenschaften hat das Interesse an dynamischen Modellen stark zugenommen. Das grundsätzliche dynamische Verhalten kann dabei auf dem Wege einer theoretischen Modellbildung ermittelt werden, wenn die das System beschreibenden Gesetzmäßigkeiten in analytischer Form vorliegen. Wenn man diese Gesetze jedoch nicht oder nur teilweise kennt, oder wenn einige wesentliche Parameter nicht genau bekannt sind, dann muß man eine experimentelle Modellbildung, Identifikation genannt, durchführen. Hierbei verwendet man gemessene Signale und ermittelt das zeitliche Verhalten innerhalb einer gewählten Klasse von mathematischen Modellen.

Die Systemidentifikation (oder Prozeßidentifikation) ist eine noch relativ junge Disziplin, die sich vor allem in Rahmen der Regelungstechnik seit etwa 1960 entwickelt hat. Sie verwendet Grundlagen und Methoden der Systemtheorie, Signaltheorie, Regelungstheorie und Schätztheorie, und wurde wesentlich geprägt durch die moderne Meßtechnik und digitale Rechentechnik.

In zwei Bänden werden die bekanntesten Methoden der Identifikation dynamischer Systeme behandelt. Dabei wird sowohl auf die Theorie als auch Anwendung eingegangen. Das Werk ist eine Fortsetzung der vom Verfasser im Jahr 1971 im Bibliographischen Institut und im Jahr 1974 im Springer-Verlag erschienenen Bändchen. Der Umfang ist jedoch durch die weitere Entwicklung des Gebietes erheblich angestiegen, so daß die Aufteilung in zwei Bände zweckmäßig war.

Die Behandlung von grundlegenden Methoden der Identifikation dynamischer Systeme erfolgt in Band I. In Kapitel 1 wird zunächst das prinzipielle Vorgehen bei der Identifikation beschrieben. Die einzelnen Methoden werden nach typischen Merkmalen geordnet und es wird eine Übersicht der verschiedenen Anwendungsmöglichkeiten gegeben.

Dann folgt im Kapitel 2 eine kurze Zusammenstellung der mathematischen Modelle linearer dynamischer Systeme für zeitkontinuierliche und zeitdiskrete Signale. Die weiteren Kapitel sind in Teilen zusammengefaßt.

Im Teil A wird zunächst die Identifikation mit nichtparametrischen Modellen für zeitkontinuierliche Signale betrachtet. Dabei wird die Fourieranalyse mit nichtperiodischen Testsignalen, die Frequenzgangmessung mit periodischen Testsignalen und die Korrelationsanalyse mit stochastischen Signalen beschrieben. Dann erfolgt im Teil B die Identifikation mit nichtparametrischen Modellen, aber zeitdiskreten Signalen in Form der Korrelationsanalyse.

Der Teil C widmet sich der Identifikation mit parametrischen Modellen für zeitdiskrete Signale. Der Fall zeitdiskreter Signale wird hier zuerst besprochen, da die zugehörigen Methoden einfacher zu behandeln und weiter entwickelt sind als für zeitkontinuierliche Signale. Es wird zunächst die Parameterschätzung für statische Systeme und dann für dynamische Systeme beschrieben. Die Methode der kleinsten Quadrate in der ursprünglichen, nichtrekursiven Form wird abgeleitet. Dann werden die zugehörigen rekursiven Parameterschätzgleichungen angegeben. Es folgen die Methoden der gewichteten kleinsten Quadrate, mehrere Modifikationen der Methode der kleinsten Quadrate, die Methode der Hilfsvariablen und die stochastische Approximation.

Im Anhang werden verschiedene Grundlagen, Grundbegriffe und Ableitungen zusammengefaßt, die den Stoff einiger Kapitel ergänzen.

Der Band II setzt den Teil C mit einer vertiefenden Behandlung der Parameterschätzmethoden fort. Zunächst werden die Maximum-Likelihood-Methode und die Bayes-Methode beschrieben, die von einer statistischen Betrachtungsweise ausgehen. Dann folgt eine Parameterschätzmethode mit nichtparametrischem Zwischenmodell. In besonderen Kapiteln wird auf die rekursiven Parameterschätzmethoden und damit verbunden, auf die Parameterschätzung zeitvarianter Prozesse eingegangen. Weitere Kapitel über numerisch verbesserte Schätzmethoden, ein Vergleich verschiedener Parameterschätzmethoden, die Parameterschätzung im geschlossenen Regelkreis und verschiedene Probleme (Wahl der Abtastzeit, Ermittlung der Modellordnung, integrale Prozesse, usw.) schließen den Teil C ab.

Zur Identifikation mit parametrischen Modellen, aber zeitkontinuierlichen Signalen in Teil D werden zunächst verschiedene Verfahren zur Parameterbestimmung aus Übergangsfunktionen, die sog. Kennwertermittlung, beschrieben. Dann folgen die Parametereinstellmethoden mit Modellabgleich, die im Zusammenhang mit der Analogrechentechnik entstanden sind, Parameterschätzmethoden für Differentialgleichungen und für gemessene Frequenzgänge.

Der Teil E ist der Identifikation von Mehrgrößensystemen gewidmet. Es werden zunächst die verschiedenen Modellstrukturen und dann geeignete Identifikationsmethoden mittels Korrelation und Parameterschätzung betrachtet.

Einige Möglichkeiten zur Identifikation nichtlinearer Systeme werden in Teil F beschrieben. Hierbei steht die Parameterschätzung von dynamischen Systemen mit stetig und nichtstetig differenzierbaren Nichtlinearitäten im Vordergrund.

Schließlich wird im Teil G auf die praktische Durchführung der Identifikation eingegangen. Es werden zunächst einige Angaben zu praktischen Aspekten, wie besondere Geräte, die Elimination besonderer Störsignale, die Verifikation der erhaltenen Modelle und die Identifikation mit Digitalrechnern gemacht. Dann erfolgen Anwendungsbeispiele für mehrere technische Prozesse. Diese Beispiele zeigen exemplarisch, daß die meisten der behandelten Identifikationsmethoden in verschiedenen Einsatzfällen auch praktisch erprobt wurden.

Das Werk richtet sich an Studenten, Ingenieure in der Forschung und Praxis und an Wissenschaftler aus dem Bereich der Naturwissenschaften, die an einer Einführung und vertieften Behandlung der Identifikation dynamischer Systeme interessiert sind. Dabei werden lediglich Grundkenntnisse der Behandlung linearer, dynamischer Systeme vorausgesetzt. Der erste Band entspricht weitgehend einer Vorlesung (2 Stunden Vorlesung, 1 Stunde Übung) an der Technischen Hochschule Darmstadt ab dem sechsten Semester. Dabei wird der Stoff in der Reihenfolge der Kapitel 1, A1, A2, 2, 3, 4, 5, A3, 6, 7, 8, 9, 10 behandelt.

Viele der Methoden, Untersuchungen und Ergebnisse wurden in zahlreichen Studien- und Diplomarbeiten seit 1966 und in besonderen Forschungsarbeiten seit 1972 erarbeitet. Hierzu möchte ich sowohl den damaligen Studenten als auch den Institutionen zur Forschungsförderung, besonders der Deutschen Forschungsgemeinschaft (DFG) und dem Bundesministerium für Forschung und Technologie (BMFT) sehr danken.

Der Verfasser dankt ganz besonders seinen Mitarbeitern, die in mehrjähriger Zusammenarbeit an der Untersuchung und Entwicklung von Identifikationsmethoden, der Erstellung von Programmpaketen, Simulationen auf Digitalrechnern, Anwendungen mit Prozeßrechnern und Mikrorechnern und schließlich durch das Korrekturlesen wesentlich am Entstehen dieses Buches beteiligt waren. Hierbei danke ich besonders den Herren Dr.-Ing. U. Baur, Dr.-Ing. W. Bamberger, Dr.-Ing. S. Bergmann, Dr.-Ing. P. Blessing, Dr.-Ing. W. Goedecke, Dr.-Ing. H. Hensel, Dipl.-Ing. R. Kofahl, Dr.-Ing. H. Kurz, Dr.-Ing. K.-H. Lachmann, Dr.-Ing. W. Mann, Dipl.-Ing. K.H. Peter, Dr.-Ing. R. Schumann und Dr.-Ing. F. Radke. Mein Dank gilt ferner dem Springer-Verlag für die Herausgabe des Buches. Schließlich möchte ich mich noch sehr bei Frau M. Widulle für die sorgfältige Gestaltung des gesamten Textes mit der Schreibmaschine bedanken.

Darmstadt, April 1987 Rolf Isermann

Inhaltsverzeichnis

12 Maximum-Likelihood-Methode

In den vorausgegangenen Kapiteln wurden Parameterschätzmethoden behandelt, bei denen keine besonderen Annahmen über die Verteilungsdichte des Störsignals oder Fehlersignals gemacht werden mußten. Die Annahme von Modellen, deren Fehlersignal linear in den Parametern ist, erlaubte dann bei der nichtrekursiven Methode der kleinsten Quadrate eine *direkte Verarbeitung* der Daten (in einem Zug), siehe Abschnitt 1.3, was einen rechentechnischen Vorzug bedeutet. Die möglichen Strukturen der Modelle waren jedoch eingeschränkt.

Die in diesem Kapitel beschriebene Maximum-Likelihood-Methode unterscheidet sich prinzipiell von den bisher betrachteten Methoden. Sie geht von einer statistischen Betrachtungsweise aus, bei der eine Funktion der beobachteten Signale und unbekannten Parameter, die Likelihood-Funktion, gebildet wird. Für die Verteilungsdichte des Fehlersignals müssen allerdings bestimmte Annahmen gemacht werden; das Fehlersignal braucht jedoch nicht mehr linear in allen Parametern zu sein. Einfache Verhältnisse ergeben sich allerdings nur bei Annahme einer vormalverteilten Fehlersignals. Die Maximum-Likelihood-Methode ist ein relativ allgemeines Parameterschätzverfahren. Sie erlaubt aufgrund der allgemeinen Schätztheorie Angaben über die asymptotische Güte der Konvergenz. Unter bestimmten Bedingungen ist sie asymptotisch effizient, d.h. es gibt keine anderen erwartungstreuen Schätzverfahren mit kleinerer Varianz.

Über die Entwicklung der Maximum-Likelihood-Methode berichtet Deutsch (1965). Das Prinzip der Maximum-Likelihood-Schätzung geht demnach auf Gauss (1809) zurück. R.A. Fisher (1921) hat sie jedoch als allgemeines Schätzverfahren eingeführt. Seitdem gehört sie zu den grundlegenden statistischen Schätzmethoden.

Zur Parameterschätzung dynamischer Prozesse wurde die Maximum-Likelihood-Methode zuerst von Åström, Bohlin (1966) auf Differenzengleichungen mit korreliertem Ausgangssignal (ARMAX-Modell) angewendet. Kashyap (1970) verwendete sie für Zustandsmodelle mit korrelierten Eingangsstörungen aber nichtkorrelierten Ausgangsstörungen und Mehra (1973) für Zustandsmodelle mit nichtkorreliertem Eingangssignal und nichtkorrelierten Ausgangsstörungen bzw. Mehra (1973) für korrelierte Ausgangsstörungen.

Die Ableitung der nichtrekursiven Maximum-Likelihood-Methode soll im folgenden in Anlehnung an Åström, Bohlin (1966) und Åström (1980) erfolgen. Dann schließt sich die Beschreibung einer rekursiven Form an, die durch Vereinfachungen aus der nichtrekursiven ML-Methode hervorgeht. Schließlich wird noch eine unterste Schranke für die Kovarianzen der Parameterschätzwerte behandelt, die Cramér-Rao-Ungleichung.

12.1 Nichtrekursive Maximum-Likelihood-Methode (ML)

Es wird ein Modell des dynamischen Prozesses in der ARMAX-Form

$$\left.\begin{aligned} \hat{A}(z^{-1})y(z) - \hat{B}(z^{-1})u(z) &= \hat{D}(z^{-1})e(z) && (12.1\text{-}1)\\ \hat{A}(z^{-1}) &= 1 + a_1 z^{-1} + \ldots + a_m z^{-m} \\ \hat{B}(z^{-1}) &= b_1 z^{-1} + b_2 z^{-2} + \ldots + b_m z^{-m} && (12.1\text{-}2)\\ \hat{D}(z^{-1}) &= 1 + d_1 z^{-1} + \ldots + d_m z^{-m} \end{aligned}\right\}$$

angenommen, wobei e ein statistisch unabhängiges normalverteiltes Signal $(0,\sigma_e)$ ist und alle Wurzeln von D(z) im Inneren des Einheitskreises liegen.

Zum Vergleich sei an das Modell der Methode der kleinsten Quadrate erinnert, Gl. (8.1-9),

$$\hat{A}(z^{-1})y(z) - \hat{B}(z^{-1})u(z) = \varepsilon(z). \tag{12.1-3}$$

Der Gleichungsfehler ε mußte nach Satz 8.2 nichtkorreliert sein, damit eine biasfreie Parameterschätzung ermöglicht wird. Aus dem Vergleich von Gl. (12.1-1) und (12.1-3) folgt

$$\varepsilon(z) = \hat{D}(z^{-1})e(z). \tag{12.1-4}$$

Für das Modell Gl. (12.1-1) ist $\varepsilon(z)$ zu einem Prozeß mit gleitendem Mittel (moving average process) erweitert worden. Der Gleichungsfehler ε ist für diesen Fall somit ein korreliertes Signal, das durch das Filter $1/\hat{D}(z^{-1})$ in ein nichtkorreliertes Fehlersignal umgeformt wird, Bild 12.1.

Mit diesen Annahmen muß der Prozeß also die Struktur

$$y(z) = \frac{B(z^{-1})}{A(z^{-1})}u(z) + \frac{D(z^{-1})}{A(z^{-1})}v(z) \tag{12.1-5}$$

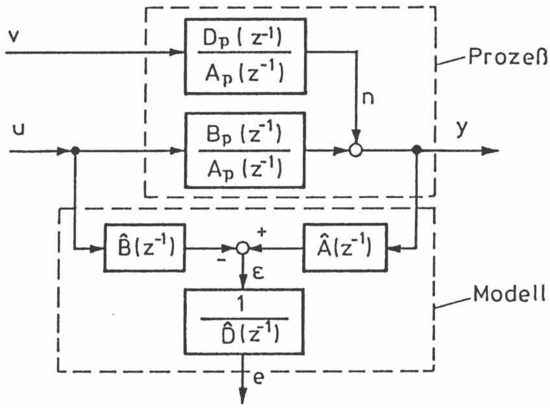

Bild 12.1 Struktur von Prozeß und Modell für die Maximum-Likelihood-Methode

haben (ARMAX-Modell), wenn v ein nichtkorreliertes Störsignal ist, und v=e gesetzt wird. Diese Struktur ergibt sich auch bei der Zustandsdarstellung siehe Abschnitt 18.2.

Zur Ableitung der Maximum-Likelihood-Methode muß das beobachtete Ausgangssignal y(k) eine bestimmte Verteilungsdichte aufweisen. Da sich nur bei Annahme eines normalverteilten Ausgangssignales überschaubare Gleichungen ergeben, sei *Normalverteilung* der y(k) angenommen.

Die bedingte Verteilungsdichte der beobachteten Signalwerte $\{y(k)\}$ für gegebene Eingangssignalwerte $\{u(k)\}$ und für gegebene Parameter $\underline{\theta}=[a_1,\dots,a_m;\ b_1,\dots,b_m;\ d_1,\dots d_m,\sigma_e]$ werde mit

$$p[\{y(k)\}|\{u(k)\},\underline{\theta}] = p[\underline{y}|\underline{u},\underline{\theta}]$$

bezeichnet und sei bekannt, Bild 12.2.

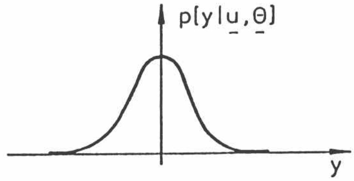

Bild 12.2 Bedingte Verteilungsdichte des beobachteten Signals y(k)

In diese Gleichung für die Verteilungsdichte werden die gemessenen Werte $y_P(k)$ und $u_P(k)$ eingesetzt. Dann erhält man die *Likelihood-*

Funktion

$$p[\underline{y}_P | \underline{u}_P, \underline{\theta}],$$

die man in Abhängigkeit von den unbekannten Parametern θ_i betrachtet, Bild 12.3.

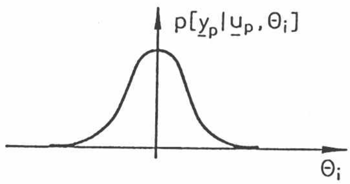

Bild 12.3 Likelihood-Funktion für einen einzigen Parameter θ_i

Da die Parameter θ_i Konstanten sind und keine stochastischen Variablen, ist die Likelihood-Funktion keine Verteilungsdichte der Parameter. Der Methode des Maximum-Likelihood liegt nun der Gedanke zugrunde, daß die besten Werte der unbekannten Parameter θ_i diejenigen sind, die dem beobachteten Ergebnis die größte Wahrscheinlichkeit verleihen. Das sind offensichtlich diejenigen Parameterwerte, die die Likelihood-Funktion maximieren. In bezug auf mehrere Parameter gilt dann als Ausgangsgleichung

$$\frac{\partial}{\partial \underline{\theta}} p[\underline{y}_P | \underline{u}_P, \underline{\theta}] = \underline{0}. \qquad (12.1\text{-}6)$$

Da die $\underline{y}_P = \{y_P(k)\}$ nicht statistisch unabhängig voneinander sind, läßt sich die bisher betrachtete Verteilungsdichte nicht unmittelbar angeben. Zur Bildung der Likelihood-Funktion wird deshalb die Verteilungsdichte

$$p[\underline{e} | \underline{u}_P, \underline{\theta}]$$

des Fehlersignals e(k) verwendet, das ebenfalls wie y(k) normalverteilt ist, falls $A(z^{-1})/D(z^{-1})$ ein *lineares Filter* ist. Es wird als *statistisch unabhängiges Signal* angenommen. Deshalb gilt für seine bedingte Verteilungsdichte bei N gemessenen Signalen

$$p[\underline{e}|\underline{u},\underline{\theta}] = p[e(1)|\underline{u},\underline{\theta}] \cdot p[e(2)|\underline{u},\underline{\theta}] \cdot \ldots \cdot p[e(N)|\underline{u},\underline{\theta}]$$

$$= \prod_{k=1}^{N} p[e(k)|\underline{u},\underline{\theta}]. \qquad (12.1\text{-}7)$$

Diese Funktion muß entsprechend Gl.(12.1-6) abgeleitet werden. Da die Ableitung eines aus vielen Faktoren bestehenden Produktes jedoch unangenehm zu handhaben ist, bildet man den Logarithmus der Likelihood-Funktion

$$L = \ln p[\underline{e}|\underline{u},\underline{\theta}] = \sum_{k=1}^{N} \ln p[e(k)|\underline{u},\underline{\theta}]. \tag{12.1-8}$$

Dadurch wird die Lage der Maxima der Likelihood-Funktion bezüglich der Parameter $\underline{\theta}$ nicht verändert. Die Parameterschätzung erfolgt schließlich durch Lösen der *Maximum-Likelihood-Gleichung*

$$\frac{\partial}{\partial \underline{\theta}} L = \frac{\partial}{\partial \underline{\theta}} \sum_{k=1}^{N} \ln p[e(k)|\underline{u},\underline{\theta}] = \underline{0}. \tag{12.1-9}$$

Diese Gleichung gilt noch für beliebige differenzierbare Verteilungsdichten. Das Einführen einer *Normalverteilung* erleichtert jedoch die folgende Rechnung wesentlich. Mit $E\{e(k)\}=0$ gilt dann für einen einzigen Signalwert zum Zeitpunkt k

$$p[e(k)|\underline{u},\underline{\theta}] = \frac{1}{\sigma_e\sqrt{2\pi}} \exp[-\tfrac{1}{2} e^2(k)/\sigma_e^2], \tag{12.1-10}$$

und für N Signalwerte nach Gl.(12.1-7)

$$p[\underline{e}|\underline{u},\underline{\theta}] = \prod_{k=1}^{N} \frac{1}{\sigma_e\sqrt{2\pi}} \exp[-\tfrac{1}{2} e^2(k)/\sigma_e^2] \tag{12.1-11}$$

wenn die Standardabweichung σ_e aller Fehlersignale e(k) gleich ist. Der Logarithmus der Likelihood-Funktion wird dann nach Gl.(12.1-8)

$$\begin{aligned} L(\underline{\theta}) &= \ln\left[\left[\frac{1}{\sigma_e\sqrt{2\pi}}\right]^N \cdot \prod_{k=1}^{N} \exp[-\tfrac{1}{2} e^2(k)/\sigma_e^2] \right] \\ &= -\frac{1}{2\sigma_e^2} \sum_{k=1}^{N} e^2(k) - N \ln \sigma_e - \frac{N}{2} \ln 2\pi. \end{aligned} \tag{12.1-12}$$

Diese Gleichung muß nun bezüglich der unbekannten Parameter a_i, b_i, d_i und σ_e maximiert werden. Bezeichnet man die in Gl.(12.1-12) auftretende Verlustfunktion mit

$$V(\underline{\theta}) = \frac{1}{2} \sum_{k=1}^{N} e^2(k) \tag{12.1-13}$$

dann muß gelten

$$\frac{\partial}{\partial a_i} V(\underline{\theta}) = 0; \quad \frac{\partial}{\partial b_i} V(\underline{\theta}) = 0; \quad \frac{\partial}{\partial d_i} V(\underline{\theta}) = 0 \tag{12.1-14}$$

$$\frac{\partial L}{\partial \sigma_e} = 2\,\sigma_e^{-3}\, V(\underline{\theta}) - N\,\sigma_e^{-1} = 0. \tag{12.1-15}$$

Aus der letzten Gleichung folgt direkt als Schätzwert für die Varianz