SYSTEMS & CONTROL ENCYCLOPEDIA

Theory, Technology, Applications

Editor-in-Chief

Madan G Singh

UMIST, MANCHESTER, UK

Volume 4
I-L

SYSTEMS & CONTROL ENCYCLOPEDIA

Theory, Technology, Applications

VOLUME 4
I–L

Editor-in-Chief

Madan G Singh

University of Manchester
Institute of Science and Technology,
Manchester, UK

PERGAMON PRESS

OXFORD · NEW YORK · BEIJING · FRANKFURT
SÃO PAULO · SYDNEY · TOKYO · TORONTO

U.K.	Pergamon Press, Headington Hill Hall, Oxford, OX3 0BW, England
U.S.A.	Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523, U.S.A.
PEOPLE'S REPUBLIC OF CHINA	Pergamon Press, Qianmen Hotel, Beijing, People's Republic of China
FEDERAL REPUBLIC OF GERMANY	Pergamon Press, Hammerweg 6, D-6242 Kronberg, Federal Republic of Germany
BRAZIL	Pergamon Editora, Rua Eça de Queiros, 346, CEP 04011, São Paulo, Brazil
AUSTRALIA	Pergamon Press Australia, P.O. Box 544, Potts Point, N.S.W. 2011, Australia
JAPAN	Pergamon Press, 8th Floor, Matsuoka Central Building, 1–7–1 Nishishinjuku, Shinjuku-ku, Tokyo 160, Japan
CANADA	Pergamon Press Canada, Suite 104, 150 Consumers Road, Willowdale, Ontario M2J 1P9, Canada

---

First edition 1987

Library of Congress Cataloging in Publication Data
Main entry under title:
Systems and control encyclopedia.
Includes bibliographies.
1. System analysis—Dictionaries.
2. Control theory—Dictionaries.
3. Systems engineering—Dictionaries.
I. Singh, Madan G.
QA402.S968 1987 003'.03'21 86-15085

British Library Cataloguing in Publication Data
Systems and control encyclopedia: theory, technology, applications.
1. Control theory—Dictionaries I. Singh, Madan G.
003'.03'21 QA402

ISBN 0–08–028709–3 (set)

Printed in Great Britain by A. Wheaton & Co. Ltd, Exeter

SYSTEMS & CONTROL ENCYCLOPEDIA

Theory, Technology, Applications

EDITORIAL ADVISORY BOARD

HONORARY EDITORIAL ADVISORY BOARD

SUBJECT EDITORS

Applications of Control to Aerospace and Aeronautics

M Pélegrin
Centre d'Etudes et de Recherches de Toulouse
Toulouse, France

Applied Robotics, Control Applications to Manufacturing Industries, Flexible Manufacturing and Ergonomics

P Drazan
University of Wales
Institute of Science and Technology
Cardiff, UK

Building Systems

D J Fisk
Department of the Environment
London, UK

Classical Control, CAD of Control Systems, Multivariable Control

N Munro
University of Manchester
Institute of Science and Technology
Manchester, UK

Database Management, Theory and Applications of Distributed Computing for Control

E Gelenbe
Université Paris-Sud
Paris, France

Distributed Parameter Systems, Discrete Systems

J L Lions
Centre National d'Etudes Spatiales
Paris, France

Environmental Measurements, Measurements in Chemistry

Y Sawaragi
Japan Institute of Systems Research
Kyoto, Japan

Environmental Systems and Agricultural Systems

E Halfon
National Water Research Institute
Burlington, Ontario, Canada

Hierarchical Control, Decentralized Control, Model Simplification for Complex Systems Team Theory, Robustness and Sensitivity

A Titli
Centre National de la Recherche Scientifique
Toulouse, France

History of Systems and Control

S Bennett
University of Sheffield
Sheffield, UK

Management Systems, Educational Systems, Health Care Systems, Social Effects of Automation

A P Sage
George Mason University
Fairfax, Virginia, USA

Measurements in the Biological Area

P A Payne
University of Manchester
Institute of Science and Technology
Manchester, UK

Measurements in the Control Area including Intelligent Instrumentation

M G Mylroi
University of Bradford
Bradford, UK

Modelling and Control of Biological Systems, Biomedical Engineering

D A Linkens
University of Sheffield
Sheffield, UK

Pattern Recognition, Fuzzy Systems

M M Gupta
University of Saskatchewan
Saskatoon, Saskatchewan, Canada

Physical Systems Modelling, Identification, Estimation Theory and Applications, Signal Processing

T Kailath
Stanford University
Stanford, California, USA

L Ljung
Linköping University
Linköping, Sweden

Process Control Applications

A J Niemi

R Ylinen
Technical Research Centre of Finland
Espoo, Finland

Realization Theory for Linear and Nonlinear Systems

C Lobry
Nice, France

M Fliess
Centre National de la Recherche Scientifique
Gif-sur-Yvette, France

Y Rouchaleau
Ecole des Mines de Paris
Valbonne, France

Self-Tuning Regulators and Adaptive Control, Stochastic Systems

D W Clarke
University of Oxford
Oxford, UK

Simulation Techniques, Model Simplification

B P Zeigler
University of Arizona
Tucson, Arizona, USA

Stability Theory, General Systems Theory

G J Klir
State University of New York at Binghamton
Binghamton, New York, USA

Static Optimization, Dynamic Optimization, Optimal Control, Game Theory

A Bensoussan
Institut National de Recherche en Informatique et en Automatique
Le Chesnay, France

Technology of Final Control Elements

J P Elloy
Ecole Nationale Supérieure de Mécanique
Nantes, France

Transportation Systems, Communication Systems

A J Fossard
Ecole Nationale Supérieure de l'Aeronautique et de l'Espace
Toulouse, France

Utility Systems and Energy Systems

A Brameller
University of Manchester
Institute of Science and Technology
Manchester, UK

CONTENTS

I

Identification: Asymptotic Theory

Identification concerns the problem of determining mathematical models from observed data. From a conceptual point of view, an identification method is a mapping from the space of observed input/output data to a space of models. Identification theory concerns the analysis of such mappings. In this article we shall present an analysis of common identification methods. The analysis will be confined to asymptotic properties; that is, we shall only study what happens as the number of observed data approaches infinity.

1. Identification Methods

Here, we adopt the same notation and framework as detailed elsewhere in the Encyclopedia (see *Identification: Time-Domain Methods*). We thus consider a set of models

$$\mathcal{M} = \{\mathcal{M}(\theta) | \theta \in D_{\mathcal{M}}\} \tag{1}$$

consisting of individual models that are characterized as predictors:

$$\mathcal{M}(\theta): \hat{y}(t|\theta) = g_{\mathcal{M}}(\theta; t, z^{t-1}) \tag{2}$$

Here

$$z^t = [y(1), u(1), \ldots, y(t), u(t)] \tag{3}$$

denotes the sequence of inputs (u) and outputs (y) up to time t. The prediction error at time t is denoted by

$$\varepsilon(t, \theta) = y(t) - \hat{y}(t|\theta) \tag{4}$$

An identification method $\mathcal{I}$ is now a mapping from z^N to $\mathcal{M}$:

$$\mathcal{I}: z^N \to \mathcal{M}(\hat{\theta}_N) \tag{5}$$

or, if we prefer to work directly with the parameters, a mapping from z^N to $D_{\mathcal{M}}$:

$$z^N \to \hat{\theta}_N \tag{6}$$

Two basic principles to determine such mappings are discussed elsewhere (see *Identification: Time-Domain Methods*):

$$\hat{\theta}_N = \arg\min_{\theta \in D_{\mathcal{M}}} V_N(\theta, z^N) \tag{7a}$$

$$V_N(\theta, z^N) = \frac{1}{N}\sum_{t=1}^{N} l[t, \theta, \varepsilon(t, \theta)] \tag{7b}$$

and

$$\hat{\theta}_N = \arg_{\theta \in D_{\mathcal{M}}} f_N(\theta, z^N) = 0 \tag{8a}$$

$$f_N(\theta, z^N) = \frac{1}{N}\sum_{t=1}^{N} \zeta(t, \theta)\varepsilon(t, \theta) \tag{8b}$$

In the criterion minimization approach (7), l is a scalar-valued "norm" that measures the "size" of the prediction error. In the correlation approach (8), $\{\zeta(t)\}$ is a sequence of vectors, formed from past data, and possibly also dependent on θ:

$$\zeta(t) = \zeta(t, z^{t-1}, \theta) \tag{9}$$

Further details and examples of identification are given elsewhere (see *Identification: Time-Domain Methods*).

2. Asymptotic Analysis

We now ask the question: what are the properties of the mapping (5), which is implicitly defined by (7) or (8)? There are two ways to approach this question.

(a) Generate data z^N with known characteristics. Apply the mapping (5) (corresponding to a particular identification method) and evaluate the properties of $\mathcal{M}(\hat{\theta}_N)$. Such methods are known as *simulation* studies.

(b) Assume certain properties of z^N and try to calculate what the inherited properties of $\mathcal{M}(\hat{\theta}_N)$ are. This is known as *analysis*.

The mapping (5) is in general fairly complicated and nonlinear. Moreover, the data z^N are often considered to be a realization of a stochastic process. This means that although the mapping (5) itself is deterministic the analysis has to be performed in a probabilistic framework. The parameter $\hat{\theta}_N$ will be a random variable with a distribution that is inherited from the probabilistic properties of z^N in a nontrivial manner. All this implies that in general we can describe only the *asymptotic properties* of $\hat{\theta}_N$ as N tends to infinity. The typical asymptotic aspects are to establish to what $\hat{\theta}_N$ converges as $N \to \infty$ and what the asymptotic distribution of $\hat{\theta}_N$ is.

There are several reasons why such results are of interest. We need them in order to make rational decisions about experiment design and model validation, as well as about the choices of l and ζ in the identification methods.

Here, we shall discuss such asymptotic, analytic results. The analysis is a fairly technical problem, and we cannot provide the theory itself here. Instead, we shall quote the basic results of the analysis, and comment upon the implications for various user choices.

The fact that the analysis only deals with asymptotic results implies certain limitations. With asymptotic results we know the properties of $\hat{\theta}_N$ for "large" N. However, the theory usually does not provide any information about how large N has to be for the results to be applicable. It may be $N = 100$ or $N \approx 10^6$, which

obviously makes a big difference to the user. Therefore, to get some insight into the properties of $\hat{\theta}_N$ for realistic values of N, the analysis must be complemented with simulation studies.

3. The Concept of Identifiability

The term identifiability is often used to express certain properties related to the identification problem. Many different definitions have been given for this concept, and these have been surveyed by Nguyen and Wood (1982). Without going into details about the different possibilities, we may single out the following basic approaches.

(*a*) *Identifiability as parameter consistency.* There is one parameter $\theta_0 \in D_{\mathcal{M}}$ that corresponds to the "true system," usually expressed as

$$\{\varepsilon(t, \theta_0)\} \text{ is a sequence of independent random variables} \quad (10)$$

The parametrization (1) is then said to be identifiable under the chosen identification method and for the particular data set if

$$\hat{\theta}_N \to \theta_0 \quad \text{with probability 1 as} \quad N \to \infty \quad (11)$$

(*b*) *Identifiability as parameter uniqueness.* Identifiability is defined as the situation where the estimate $\hat{\theta}_N$ converges to a *unique* value θ^* as $N \to \infty$, and we are not concerned with the question of whether θ^* is a "true" value. This situation is at hand when the parametrization (1) and the data set z are such that no two different values of $\theta \in D_{\mathcal{M}}$ (except perhaps on a null space) give the same predictions.

(*c*) *Identifiability as system consistency.* Identifiability is defined as the situation where $\mathcal{M}(\hat{\theta}_N)$ converges to a true description of the system, regardless of whether $\hat{\theta}_N$ itself converges to a unique value.

4. Convergence Properties

In Eqns. (7) and (8) the criterion function $V_N(\theta, z^N)$ and the function $f_N(\theta, z^N)$ are defined as sums of random variables $l[t, \theta, \varepsilon(t, \theta)]$ and $\varepsilon(t, \theta)\zeta^{\mathrm{T}}(t, \theta)$, respectively. If these variables were stationary sequences of independent random variables, an elementary law of large numbers would tell us that the sums in question would converge to their expected values. Now, the variables clearly are dependent, but under weak conditions the dependence will decay fairly rapidly with the time difference (i.e., the prediction error now is almost independent of what happened a long time ago). Conditions of this character are known in statistics as "mixing conditions." Under such suitable conditions, the law of large numbers will still apply:

$$V_N(\theta, z^N) \to \bar{V}(\theta)$$

uniformly in θ with probability 1 as

$$N \to \infty \quad (12)$$

$$f_N(\theta, z^N) \to \bar{f}(\theta)$$

uniformly in θ with probability 1 as

$$N \to \infty \quad (13)$$

Here

$$\bar{V}(\theta) = \bar{E}l[t, \theta, \varepsilon(t, \theta)], \quad \bar{f}(\theta) = \bar{E}\zeta(t, \theta)\varepsilon(t, \theta) \quad (14)$$

where the notation $\bar{E}$ is defined as

$$\bar{E}h(t) = \lim_{N \to \infty} \frac{1}{N} \sum_{1}^{N} Eh(t) \quad (15)$$

with an implied assumption that the limit exists. See, for example, Ljung (1978) for technical details.

With (12) it follows that the value $\hat{\theta}_N$ that minimizes $V_N(\theta, z^N)$ will converge with probability 1 to θ^*, the value that minimizes $\bar{V}(\theta)$. If the minimum is not unique we obtain convergence into a set:

$$\hat{\theta}_N \to D_c, \quad \text{with probability 1 as} \quad N \to \infty \quad (16)$$

$$D_c = \{\theta \mid \bar{V}(\theta) = \min_{\theta} \bar{V}(\theta)\}$$

Similarly, the solution $\hat{\theta}_N$ to $f_N(\theta, z^N) = 0$ will, under (13), converge with probability 1 to the solution θ^* of $\bar{f}(\theta) = 0$.

These results are very natural, but also quite general. Notice, for example, that (12) and (13) will hold even without a specific assumption about how the true data are generated. We need thus not assume that the true system belongs to the model set for the quoted convergence results to hold.

5. Asymptotic Distribution

We shall now investigate in what manner the estimate $\hat{\theta}_N$ approaches the limit θ^*. Let us first consider the case (8). We have, by Taylor's expansion,

$$\begin{aligned} 0 &= f_N(\theta_N, z^N) \\ &= f_N(\theta^*, z^N) + f_N'(\xi_N, z^N)^{\mathrm{T}}(\theta_N - \theta^*) \end{aligned} \quad (17)$$

Here f_N' is the gradient of f with respect to θ and ξ_N is a value in a $|\hat{\theta}_N - \theta^*|$ neighborhood of θ^*. Now,

$$\sqrt{N} f_N(\theta^*, z^N) = \frac{1}{\sqrt{N}} \sum_{1}^{N} \zeta(t, \theta^*)\varepsilon(t, \theta^*) \quad (18)$$

is a normalized sum of random variables with zero mean values (recall that $\bar{f}(\theta^*) = \bar{E}\zeta(t, \theta^*)\varepsilon(t, \theta^*) = 0$ by definition). Under suitable, and weak, conditions the central limit theorem will be applicable to Eqn. (18) (see, for example, Ljung and Caines (1979) for some technical details). Then

$$\sqrt{N} f_N(\theta^*, z^N) \in \mathrm{AsN}(0, Q) \quad (19)$$

where

$$Q = \lim_{N \to \infty} EN \cdot f_N(\theta^*, z^N) f_N^{\mathrm{T}}(\theta^*, z^N) \quad (20)$$

(Equation (19) means that the random variable $\sqrt{N} f_N$ converges in distribution to the normal distribution with

zero mean and covariance matrix Q.) Under similar conditions as in (13) the matrix $f'_N(\theta, z^N)$ will converge uniformly in θ with probability 1 to some limit $H(\theta)$ as $N \to \infty$. Hence

$$f'_N(\xi_N, z^N) \to H^* \quad \text{with probability 1 as} \quad N \to \infty \tag{21}$$

where

$$H^* = \bar{E}\zeta(t, \theta^*)\psi^T(t, \theta^*) + \bar{E}\frac{d}{d\theta}\zeta(t, \theta)|_{\theta=\theta^*}\varepsilon(t, \theta^*) \tag{22}$$

Here

$$\psi(t, \theta) = -\frac{d}{d\theta}\varepsilon(t, \theta) = \frac{d}{d\theta}\hat{y}(t|\theta) \quad \text{(a dim } \theta \times \text{dim } y \text{ matrix)} \tag{23}$$

Equations (17), (19) and (21) now imply that

$$\sqrt{N}(\hat{\theta}_N - \theta^*) \in \text{AsN}(0, P) \tag{24}$$

where

$$P = [H^*]^{-1}Q[H^*]^{-1} \tag{25}$$

provided, of course, that H^* in (21) is invertible.

In the special case where $\theta^* = \theta_0$ is such that $\{\varepsilon(t, \theta_0)\}$ is a sequence of independent random vectors with zero mean values and covariance matrices Λ_0, Eqns. (20) and (22) simplify to

$$Q = \bar{E}\zeta(t, \theta_0)\Lambda_0\zeta^T(t, \theta_0)$$
$$H^* = \bar{E}\zeta(t, \theta_0)\psi^T(t, \theta_0) \tag{26}$$

Consider now the criterion minimization approach (7) with

$$l(t, \theta, \varepsilon) = \tfrac{1}{2}\varepsilon^T\Lambda_0^{-1}\varepsilon \tag{27}$$

The minimizing value $\hat{\theta}_N$ can also be determined as

$$\hat{\theta}_N = \arg_{\theta \in D_{\mathcal{M}}} \frac{d}{d\theta}V_N(\theta; z^N) = 0$$

where

$$\frac{d}{d\theta}V_N(\theta; z^N) = \frac{1}{N}\sum_{t=1}^{N}\varepsilon(t, \theta)\Lambda_0^{-1}\psi^T(t, \theta) \tag{28}$$

We can therefore apply the previous results with $\zeta(t, \theta_0) = \psi(t, \theta_0)\Lambda_0^{-1}$; so then we obtain that (24) holds with

$$P = [\bar{E}\psi(t, \theta_0)\Lambda_0^{-1}\psi^T(t, \theta_0)]^{-1} \tag{29}$$

This result has a natural interpretation. We see that the asymptotic accuracy of a certain parameter is related to how sensitive the prediction $\hat{y}(t|\theta)$ is with respect to this parameter. Clearly, the more a parameter affects the prediction, the easier it will be to determine its value.

There is another interesting implication of Eqn. (29). Let

$$\bar{V}(\theta) = E\varepsilon^T(t, \theta)\Lambda_0^{-1}\varepsilon(t, \theta) \tag{30}$$

Then we can evaluate how good the model $\mathcal{M}(\hat{\theta}_N)$ is by calculating the value $\bar{V}(\hat{\theta}_N)$. Here $\hat{\theta}_N$ is a random variable, and we may evaluate the expectation of $\bar{V}(\hat{\theta}_N)$ with respect to $\hat{\theta}_N$. This gives, after some straightforward calculations,

$$E\,\bar{V}(\theta_N) \sim E\,\bar{V}(\theta_0) + \frac{\dim \theta}{N} \tag{31}$$

using Eqn. (29) and an assumption that $\{\varepsilon(t, \theta_0)\}$ is a white noise sequence, thus assuming the model set is large enough to include a correct description of the true system. Here " $\sim$ " means asymptotically equal to.

The result (31), which was first derived by Akaike (1973) is remarkable in its generality. It tells us that the expected prediction error variance increases with the number of independent parameters in the model (once the model set is large enough to contain a true value θ_0) irrespective of where the parameters enter the model. Notice, though, that the derivation assumes the matrix in Eqn. (29) to be invertible. It is therefore only the number of "independent" parameters that enters the Eqn. (31).

6. *The Cramér-Rao Bound*

There is a well-known theoretical lower bound for the achievable accuracy of an estimate. This is the Cramér–Rao bound:

$$E(\hat{\theta}_N - \theta_0)(\hat{\theta}_N - \theta_0)^T \geqslant E\left[\frac{\partial}{\partial\theta}\log L(\theta, z^N)\right]^T\left[\frac{\partial}{\partial\theta}\log L(\theta, z^N)\right] \tag{32}$$

for *any* unbiased estimate $\hat{\theta}_N$. Here $L(\theta, z^N)$ is the likelihood function for the estimation problem in question. If we apply this result to an identification problem, where the true prediction errors $\{\varepsilon(t, \theta)\}$ form a sequence of independent, Gaussian random variables with covariance matrices Λ, it is fairly straight-forward to show that the right-hand side of (32) equals P as given by Eqn. (29). The covariance matrix of the asymptotic distribution is thus given by the Cramér–Rao lower bound in this case. The estimator (7) is then said to be asymptotically efficient. See, for example, Goodwin and Payne (1977) for a further discussion.

7. *A Discussion of the Asymptotic Results*

The result (16) can be formulated in the following more suggestive way:

> The estimate converges to the best possible predictor, that is, the best possible approximation of the system, in the model set $\mathcal{M}$ (in the sense of the chosen criterion).

When the system has no input signal $\{u(t)\}$ (i.e., when the properties of a time series, or signal, $\{y(t)\}$ are modelled), this is a very strong robustness result. It means that we obtain a very meaningful approximate description of the signal.

For a system with an input, the algorithm still makes the best possible out of the situation: when the system is more complex than the model set, it chooses the best approximation available within the model set. This may have some surprising effects on the parameter estimates of the model, which we illustrate in the following simple example.

EXAMPLE 1. Suppose that the system is given by

$$y(t) + a_0 y(t-1) = b_0 u(t-1) + e(t) + c_0 e(t-1) \quad (33)$$

where $\{u(t)\}$ and $\{e(t)\}$ are independent sequences of independent random variables with zero mean values and unit variances. Let the model set be given by

$$\hat{y}(t|\theta) = -ay(t-1) + bu(t-1) \quad (34)$$

It is easy to verify that the values of a and b that give the best predictions when applied to Eqn. (33) are

$$a^* = a_0 - c_0/r_0, \quad b^* = b_0 \quad (35)$$

where

$$r_0 = \mathrm{E}y^2(t) = \frac{r_0^2 + c_0(c_0 - a_0) - a_0 c_0 + 1}{1 - a_0^2}$$

These values give a prediction error variance

$$\bar{V}(a^*, b^*) = 1 + c_0^2 - \frac{c_0^2}{r_0} \quad (36)$$

This variance is smaller than the "true values" a_0 and b_0 inserted into Eqn. (36) would give:

$$\bar{V}(a_0, b_0) = 1 + c_0^2 \quad (37)$$

For example, with $b_0 = a_0 = 0$ and $c_0 = 0.9$, we have $\bar{V}(a^*, b^*) = 1.36$ and $\bar{V}(a_0, b_0) = 1.81$.

When we apply the prediction error method (7) to Eqns. (33) and (34) the estimates $\hat{a}(N)$ and $\hat{b}(N)$ will, according to the general convergence result, converge to the values given by Eqn. (35). The fact that $a^* \neq a_0$ is usually expressed by saying that the estimate is "biased." However, it is clear from Eqns. (36) and (37) that the bias is beneficial for the prediction performance of the model (35). It gives a strictly better predictor for $\hat{a} = \hat{a}^*$ than for $\hat{a} = a_0$.

The example stresses that the algorithm indeed gives us the best possible predictor, and it uses its parameters as vehicles for that. It is, however, important to keep in mind that what is the best approximate description of a system in general depends on the input used.

The ideal situation is of course where the "best possible predictor" is precisely identical to the "true predictor" so that we have obtained a true system description. There are essentially two conditions associated with this. One is that the model set $\mathcal{M}$ should be large enough so that the system $\mathcal{S}$ actually belongs to it. The other is that the experimental condition $\mathcal{X}$ (the input) should be general enough, so that no other model is equivalent to the system under $\mathcal{X}$. This latter condition is illustrated in the following simple example.

EXAMPLE 2. Suppose that the system is given by

$$y(t) + a^0 y(t-1) = b_1^0 u(t-1) + b_2^0 u(t-2) + e(t) \quad (38)$$

and that the input is

$$u(t) \equiv 1 \quad (39)$$

Let the model set be given by

$$\hat{y}(t|\theta) = -ay(t-1) + b_1 u(t-1) + b_2 u(t-2) + e(t) \quad (40)$$

This set is "large enough" to include the true system (38). However, under the input (39) all models $\mathcal{M}(\theta)$ such that

$$b_1 + b_2 = b_1^0 + b_2^0 \quad (41)$$

will give an exact description of the system. All these models will therefore give the best possible predictors, and convergence to the true values $b_1 = b_1^0$, $b_2 = b_2^0$ cannot be guaranteed. The experimental condition (39) is not "general enough."

8. *Properties of Estimates Determined by Numerical Methods*

So far we have discussed the properties of the estimates (7) and (8). In practice the estimates will in most cases have to be determined by numerical search methods that do not necessarily yield the theoretical values defined by Eqns. (7) and (8). The most important aspect, from this point of view, is that the minimization approach (7) will provide estimates that are known only to correspond to *local* minima of Eqn. (7b). It is therefore an important problem for analysis to establish whether there exist nonglobal, local minima of the chosen criterion function. This is in general a difficult problem, depending on the model parametrization, the choice of l and the properties of the data (the system). Clearly for a quadratic criterion (27) and a linearly parametrized model set,

$$\hat{y}(t|\theta) = \theta^{\mathrm{T}} \phi(t, z^t) \quad (42)$$

the criterion function is well behaved. For other choices, only partial results are known. See Åström and Söderström (1974) and Söderström (1975) for some further aspects.

9. *User Choices*

To determine a good identification procedure for a given situation means that the user has to make a number of

choices. He or she has to select experimental conditions (which signals to measure, which inputs to use, sampling rates, etc.), the set of models, and the functions l in Eqn. (7) or ζ in Eqn. (8). The asymptotic results quoted here, like the set D_c, into which the estimates converge (see (16)) and the asymptotic covariance matrix P (see (24) and (25)), are indeed functions of these listed design variables. Rational decisions of the design variables can thus be based on analysis of D_c and P. This analysis may be complex, but certain conclusions can be drawn from general considerations as follows.

(a) The experimental conditions should be such that the predictions become sensitive with respect to interesting parameters (see Eqn. (29)).

(b) The model set should be parsimonious (use few parameters) when describing the system according to Eqn. (31).

(c) The choice of l should be matched to the probabilistic properties of the prediction errors (optimal choice $l(\varepsilon) = -\log f(x)$, where $f(x)$ is the probability density function of $\varepsilon(t, \theta_0)$).

(d) Since the best approximation of the system for the experimental conditions at hand is obtained asymptotically, the chosen conditions should resemble those for which the model is to be used.

A detailed discussion of identification theory and user choices is given in Ljung (1986).

10. Conclusions

Asymptotic identification theory concerns the analysis of asymptotic, statistical properties of models obtained by different identification methods. The key technical tools are (perhaps nonstandard versions of) the law of large numbers and the central limit theorem. Some fairly general results can be derived for the basic identification methods (7) and (8), and such results can form a basis for rational choices of design variables.

See also: Identification: Practical Aspects; Validation of Identified Models; Identification: Frequency-Domain Methods

Bibliography

Akaike H 1973 Information theory and an extension of the maximum likelihood principle. In: Petrov B N, Csáki F (eds.) *Proc. 2nd Int. Symp. Information Theory*. Akadémiai Kaidó, Budapest, pp. 267–81

Åström K J, Söderström T 1974 Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model. *IEEE Trans. Autom. Control* **19,** 769–73

Goodwin G C, Payne R L 1977 *Dynamic System Identification: Experiment Design and Data Analysis*. Academic Press, New York

Ljung L 1978 Convergence analysis of parametric identification methods. *IEEE Trans. Autom. Control* **23,** 770–83

Ljung L 1986 *System Identification Theory for the User*. Prentice Hall, Englewood Cliffs, New Jersey

Ljung L, Caines P E 1979 Asymptotic normality of prediction error estimation for approximate system models. *Stochastics* **3,** 29–46

Nguyen V V, Wood E F 1982 Review and unification of linear identifiability concepts. *SIAM Rev.* **24,** 34–51

Söderström T 1975 On the uniqueness of maximum likelihood identification. *Automatica* **11,** 193–97

L. Ljung

Identification: Basic Problem

System identification is concerned with the problem of building mathematical models of dynamical systems based on input–output measurements. In this article we discuss typical aspects of this problem, and outline basic procedures to handle it. At the same time, this article aims at tying together the different articles in this Encyclopedia that deal with various aspects of the system identification problem. The reader is referred to cross references at the end of this article.

1. Systems and Models

The notion of *systems* plays an important role in modern science; many problems in various fields are solved in a systems-oriented framework. Subjects like control theory, communications theory and operations research tell us how to determine suitable regulators, filters, decision rules, etc. Such theory assumes that a model is available of the system in question. The applicability of the theory is thus critically dependent on the availability of good models.

How does one construct good models of a given system? This question about the interface between the real world and the world of mathematics is crucial. The general answer is that we have to study the system experimentally and make some inference from the observations. In practice there are two main routes. One is to split up the system, figuratively speaking, into subsystems, whose properties are well understood from previous experience. This basically means that we rely upon "laws of Nature" and other well-established relationships that have their roots in earlier empirical work. These subsystems are then joined together mathematically and a model of the whole system is obtained. This route is known as *modelling* (see *Dynamic-Systems Modelling: Basic Principles and Lumped-Parameter Systems*; *Dynamic-Systems Modelling: Distributed-Parameter Models and Discretization*), and does not necessarily involve any experimentation on the actual system. When a model is required of a yet unconstructed system

(such as a projected aircraft) this is the only possible approach.

The other route is based on experimentation. Input and output signals from the system are recorded and are then subjected to data analysis in order to infer a model of the system. This route is known as *identification*. It is often advantageous to try to combine the approaches of modelling and identification in order to maximize the information obtained from identification experiments and to make the data analysis as sensible as possible.

2. Models of Dynamical Systems

Many different types of models of dynamical systems have been developed for various purposes. A brief list of some of the common ones is given below.

2.1 Intuitive or Mental Models

In many cases a model of a system is never formalized. The user works with an intuitive or mental picture of how the system operates, and uses this to solve design problems associated with the system. Such a mental model can be verbalized in a number of different ways, for example, time scales for dominating time constants or frequency ranges for certain resonances, etc.

2.2 Graphic Models

A linear system is fully characterized by its impulse response or by its step response. Plots or tables of this function will thus constitute a model of the system. Such a model can be used, for example, in tuning a PID regulator. An equivalent and more common choice is to describe the frequency-response function $G(i\omega)$ as a function of ω. This function is the Fourier transform of the impulse response and is equal to the transfer function $G(s)$ evaluated on the imaginary axis. The argument of the complex number $G(i\omega)$ describes the phase shift between an input sinusoid of frequency ω and the corresponding output sinusoid, while its absolute value is the output-to-input sinusoid amplitude ratio.

A plot of the frequency-response function, such as the Nyquist or Bode diagram, is a very common model of a linear dynamical system. Several control design techniques have been specifically tailored to such models.

2.3 Analytical Models

For many purposes it is more advantageous to work with analytical, mathematical models, where the relationships between input signals $u(t)$ and output signals $y(t)$ are described by mathematical expressions. Most often, these are basically differential equations, ordinary or partial, simply because most physical phenomena are usually described in that way. A linear model is then

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + a_2 y^{(n-2)}(t) + \cdots + a_n y(t) = b_1 u^{(n-1)}(t) + \cdots + b_n u(t) \quad (1)$$

where (k) as superscript denotes differentiation k times.

It is also common to work with models in discrete time. The counterpart of (1) then is

$$y(t_k) + a_1 y(t_{k-1}) + \cdots + a_n y(t_{k-n}) = b_1 u(t_{k-1}) + \cdots + b_n u(t_{k-n}) \quad (2)$$

Here the t_k are the sampling instants. Formulas for how to transform from (1) to (2) when the input is constant between the sampling instants are given, for example, in Kwakernaak and Sivan (1972).

Often the effects of random disturbances on the system and on the measurements are included in the model. A typical example is a stochastic, state-space model

$$x(t_{k+1}) = Fx(t_k) + Gu(t_k) + w(t_k) \quad (3)$$
$$y(t_k) = Hx(t_k) + e(t_k)$$

where w and e are white-noise sequences with certain specified covariance properties.

Analytical models are most often given in the time domain, but this is not an inherent feature. We could have a frequency-domain, analytical model like

$$G(i\omega) = \frac{b_1(i\omega)^{n-1} + b_2(i\omega)^{n-2} + \cdots + b_n}{(i\omega)^n + a_1(i\omega)^{n-1} + \cdots + a_n} \quad (4)$$

which is equivalent to (1).

A complete list of possible analytical models would be too long to compile here. It may be useful, though, to list a number of adjectives that are often used to characterize models.

Static/dynamic. A static model is one where the inputs and outputs are related by algebraic ("instantaneous") expressions. A dynamic model relates these quantities by difference and/or differential equations, which will make the current input affect future outputs also.

Discrete-time/continuous-time. In a discrete-time (or sampled-data) model the relationship between the input and output is expressed in terms of the sampled sequences $u(t)$ and $y(t)$ for $t = 1, 2, \ldots$, while a continuous-time model uses the time-continuous input and output functions. Notice that this issue as such has nothing to do with how the data is collected. We could very well adjust a continuous-time model to discrete-time measured data.

Linear/nonlinear. If the mapping from input to output is linear (or affine), the model is called linear, otherwise it is nonlinear.

Deterministic/stochastic. In a deterministic model the output can be exactly calculated, given the input. If the model also contains random elements, like processes that make the exact calculations of $y(t)$ impossible given the input u^t, then it is called stochastic.

Input–output form/state-space form. If the model contains no auxiliary variables (except stochastic disturbances) in addition to the inputs and the outputs, it is called an input–output model. If it is written as a

system of first-order difference or differential equations, we talk about state-space forms.

Lumped/distributed parameter. A model that is based on a finite number of ordinary differential or difference equations is called lumped. If it uses partial differential equations or an infinite number of difference or differential equations, we call it a distributed parameter model.

Time-invariant/time-varying. A model whose properties do not depend on time is time-invariant, otherwise it is time-varying.

The issue of how to deal with the multitude of different models is further discussed in *Identification: Time-Domain Methods*.

3. Approaches to Identification

The identification procedure can in general terms be described as follows:

(a) Collect input–output data from the process.

(b) Settle for a set of candidate models.

(c) Pick one particular member of the model set as the best representative, guided by the information in the data.

Let us give a conceptual discussion of each of these three steps. More details are given in the next section.

3.1 The Data

The data is sometimes recorded during a specifically designed identification experiment. The objective then is to get as good information about the system as possible. Some methods to determine models require special input sequences, and thus specific experiments. In other cases data from normal operation of the system have to be used.

Nowadays, data is almost always recorded by sampling in discrete time using a digital computer (perhaps after intermediate storage in a data recorder). We shall from here assume that this is the case and we shall denote the data set recorded over N samples by z^N.

3.2 The Set of Models

A set of candidate models is obtained by specifying among which collection of models we are going to look for a suitable one. For graphic models this will typically be the set of all (reasonably smooth) curves corresponding to the set of all linear models in the cases described in the previous section. For analytical models, suitable sets are usually obtained by letting certain parameters in the model descriptions range over a given space (such as the a_i and b_i of (2) or certain entries of the matrices F, G and H in (3)). We shall generally denote a set of models by $\mathcal{M}$.

A model set whose parameters are basically viewed as vehicles for adjusting the fit to data, and do not reflect physical considerations in the system, is called a black-box model (set). Model sets with adjustable parameters with physical interpretations may, accordingly, be called gray boxes.

3.3 Picking a Particular Model in $\mathcal{M}$ Guided by Data

This is "the identification method." There are obviously a vast number of ways to select models. To give some idea of the basic principles, it is useful to distinguish between methods for graphic models and for sets of analytical models.

(*a*) *Graphic time-domain models.* Graphic time-domain models are basically impulse or step responses. Techniques used to determine these are called transient analyses. One simply applies an (approximate) impulse or a step as input and records the corresponding output. If the signal-to-noise ratio is good for the measurements, valuable information about static gain and dominating time constants can be obtained in this manner. There are certain techniques for approximating given step responses with analytical, low-order models, based on the tangent with the largest slope; see Rake (1980) and *Identification: Transient- and Frequency-Response Methods*. The impulse response can also be obtained as the cross-correlation function between the output and the input, when the input sequence is white noise; see for example, *Identification: Correlation Methods* and Godfrey (1980). Apparatus that performs such correlation analysis is commercially available.

(*b*) *Graphic frequency-domain models.* A direct way of determining the value of the frequency-response function of a system at a given frequency is to use a sinusoidal input of that frequency, let the transient die out, and record the phase shift and amplitude change of the output sinusoid. Formally, in discrete time (T = sampling interval) we have

$$u(t) = u_0 \sin(\omega k T) \quad \text{for } kT \leq t < (k+1)T \tag{5}$$

$$y(kT) = y_0 \sin(\omega k T + \phi) + \text{transient} \tag{6}$$

The frequency-response function $H(e^{i\omega})$ is then determined from

$$\arg H(e^{i\omega T}) = \phi, \qquad |H(e^{i\omega T})| = y_0/u_0 \tag{7}$$

The experiment is repeated for a number of different frequencies ω in the range of interest, and a table or graph of $H(e^{i\omega T})$ can be constructed. This technique is known as frequency analysis.

If the measurements are noise-corrupted so that it is difficult to determine y_0 and ϕ directly, it is useful to multiply $y(kT)$ by $\sin(kT\omega)$ and by $\cos(kT\omega)$, respectively, and sum over a number of observations. The phase shift and amplitude gain can then be determined more accurately. The technique is known as frequency analysis by the correlation methods, and equipment for this is commercially available.

With more sophisticated data analysis one may, so to speak, apply all frequencies at the same time and sort

them out afterwards by Fourier techniques. This gives a frequency-response function

$$H(e^{i\omega}) = Y(e^{i\omega})/U(e^{i\omega}) \qquad (8)$$

where Y and U are the discrete Fourier transforms (DFT) of the output and input sequences, respectively. When noise affects the system, (8) is usually a bad estimate ("the periodogram estimate"), since no noise reduction is obtained. Instead various smoothing filters are applied to (8): weighted averages over certain frequency windows are formed. Such techniques, known as spectral analysis, are further described in, for example, Jenkins and Watts (1969) and *Identification: Frequency-Domain Methods*.

(*c*) *Analytical models.* Most modern identification methods deal with the estimation of analytical models, usually described in the time domain. Detailed descriptions of such methods are given in *Identification: Maximum Likelihood Method*; *Identification: Time-Domain Methods*; *Identification: Recursive Methods*; *Identification: Least Squares Method*; *Identification: Asymptotic Theory*; *Identification: Instrumental Variable Techniques*; see also Goodwin and Payne (1977), Åström (1980) and Åström and Eykhoff (1971). The basic idea behind such methods is the following one. Let each of the candidate models "guess" (predict) the next output $y(t)$ based on the information in z^{t-1}. Pick that model which produces the best ("smallest") sequence of errors between guesses and actually recorded outputs. These identification methods are thus characterized by a criterion of fit between a model and the recorded sequence of data.

4. The Identification Procedure

The identification procedure is, in principle, described by the three items listed above: data, model set and identification criterion. In practice, the procedure is characterized by a number of choices which we now list.

Experiment design. The problem here is how to design the identification experiment so that it becomes suitably informative. The choice of inputs, sampling rates, presampling filters, feedback configurations, signals to be measured, etc. are further discussed in Goodwin and Payne (1977) and *Identification: Experiment Design*. Practical limitations and how they may affect the identification result are described in the articles *Identification: Practical Aspects*.

Choice of model set. To select the set of candidate models is without doubt the most important and, at the same time, the most difficult choice. It is here that *a priori* knowledge and engineering intuition and insight has to be combined with formal properties of models and identification methods to facilitate a good result from the identification exercise. Some aspects are discussed in *Identification: Model Structure Determination* and *Validation of Identified Models*. The importance of making use of available insights is further illustrated in the application example below.

Choice of criterion of fit. How to evaluate the quality of a particular model from data is a crucial issue, which is further discussed in *Identification: Time-Domain Methods*.

Calculation of the best model. With given data and a fixed model set and a chosen criterion of fit, the "best" model is implicitly defined. It remains "only" to calculate it, which may involve extensive computations. Good numerical algorithms are required in order to allow for reliable and inexpensive calculations. Some computational aspects are included in *Identification: Time-Domain Methods* and *Identification: Least Squares Method*; see also Gupta and Mehra (1974).

In several applications, the models are required on line, as the system operators and more data becomes available. The reason could be that the models are to be used for some on-line decision, like control (adaptive control), filter tuning (e.g., adaptive noise cancellation) or monitoring (fault detection). This implies certain restrictions on how to calculate the estimates. Such methods are called recursive identification methods (or on-line or real-time identification) and are discussed, for example, in Ljung and Söderström (1983) and *Identification: Recursive Methods*.

Model validation. Once the best model available in the model set has been determined it remains to test whether it is "good enough," that is, whether it is valid for its purpose. This is the problem of model validation, which is further described in *Validation of Identified Models* and *Identification: Model Structure Determination*.

The whole identification procedure is typically an iterative one, in which earlier made choices have to be revised after the model validation step and portions of the procedure repeated. This is illustrated in Fig. 1.

5. The Identification Tool

System identification has become an important tool for solving a number of modelling problems in engineering. Some aspects of the applicability of this tool to real-life problems are discussed in *Identification: Practical Aspects*. Here, we shall briefly comment upon what this tool should look like in the hand of the user. The identification procedure is, as pointed out in Fig. 1, typically an iterative one, where insights and judgements of the user are mingled with formal calculations, extensive data handling and complex algorithms. To make the tool an efficient one, it is therefore more or less necessary to package the software in an interactive environment, with man-machine communication via graphical displays. Several such packages have been developed. A well-known one, IDPAC, developed at the Lund Institute of Technology, Sweden, is described by Åström (1980) and Wieslander (1979).