Identification of continuous-time systems

G.P. Rao and H. Unbehauen

Abstract: System identification is a well-established field. It is concerned with the determination of particular models for systems that are intended for a certain purpose such as control. Although dynamical systems encountered in the physical world are native to the continuous-time domain, system identification has been based largely on discrete-time models for a long time in the past, ignoring certain merits of the native continuous-time models. Continuous-time-model-based system identification techniques were initiated in the middle of the last century, but were overshadowed by the overwhelming developments in discrete-time methods for some time. This was due mainly to the 'go completely digital' trend that was spurred by parallel developments in digital computers. The field of identification has now matured and several of the methods are now incorporated in the continuous time system identification (CONTSID) toolbox for use with Matlab. The paper presents a perspective of these techniques in a unified framework.

1 Introduction

It is essential to know and understand an object before we venture to handle it. An object formally referred to as a system is known through modelling and identification and can be understood by analysis. Modelling and identification techniques help develop knowledge about a system. They are prerequisites to many practices in engineering and technology and are especially important in the field of automatic control. Modelling by itself is a vast area rich in a host of well-established methods, which are based on a variety of principles. Among them, modelling of physical systems on the basis of physical principles is widely practiced. Application of the physical laws occurring in physical system phenomena provides a generic mathematical description, the key parameters in which are to be determined through the process of identification and system parameter estimation. The field of system identification grew both in size and diversity over the last several decades and it was surveyed at different stages. Astrom and Eykhoff [1] surveyed the field in 1971 with focus on discrete-time (DT) models that were predominant at that time and the subject is presented comprehensively in text books [2, 3]. The first significant survey of the field with focus on continuous-time approaches by Young [4] appeared in 1981. Subsequently, Unbehauen and Rao tracked further rapid developments in the field [5–8]. Several books [9–11] have been dedicated to the subject of identification of CT systems. There are books [12–14] dealing with the application of orthogonal functions to identification of CT systems.

IEE Proceedings online no. 20045250

doi:10.1049/ip-cta:20045250

Paper first received 25th November 2004 and in final revised form 30th June 2005

G. P. Rao is a Member of the UNESCO-EOLSS Joint Committee, PO Box 2623, Abu Dhabi, United Arab Emirates

H. Unbehauen is with the Control Engineering Laboratory, Department of Electrical Engineering and Information Science, Ruhr-University, 44780 Bochum, Germany

E-mail: unbehauen@esr.ruhr-uni-bochum.de

The system identification problem, characterised by Zadeh [15] in terms of three entities: a class of models, a class of input signals and a criterion, is depicted in Fig. 1. System identification will be successful and the results useful if the problem is well posed in terms of these entities. The class of models should be appropriate and the set of signals should be illuminative enough to reveal important system characteristics, that have the property of persistent excitation relative to the model class. A classification of the main methods into output error (OE), equation error (EE), prediction error (PE) methods etc. is according to the error minimisation criteria. Continuous-time models call for special (and additional) signal processing considerations as will be outlined in this paper. The general setting of the CT-based identification methods is shown in Fig. 2. As the distinction between CT and DT comes through parametric models, our focus here will be on such models.

At the outset, we briefly discuss the general premises of the two main approaches to identification of continuous-time systems:

(*a*) *The indirect approach* which involves identification of a discrete-time (DT) model in an 'all-digital setting' and transformation into continuous-time (CT) form.
(*b*) *The direct approach* in which the CT model is identified straightaway.

In the present digital age that stretches over nearly half a century into the past, there has been a tremendous surge of digital methods that swept across all fields of science and technology in general and the field of systems and control in particular. Surprisingly, the systems and control community was not swept off its feet from the CT base in the identification of CT systems, although the indirect approach with its 'go completely digital' attitude had dominated for some time. There are a number of advantages in describing real systems by CT models and disadvantages in discretisation of CT systems as follows:

(i) *CT models provide a good insight into the system properties:* Despite great inroads made by digital computers and the enormous spread of DT-based methods, CT will remain as the natural basis of our understanding of the physical world because physical laws (Newton's,

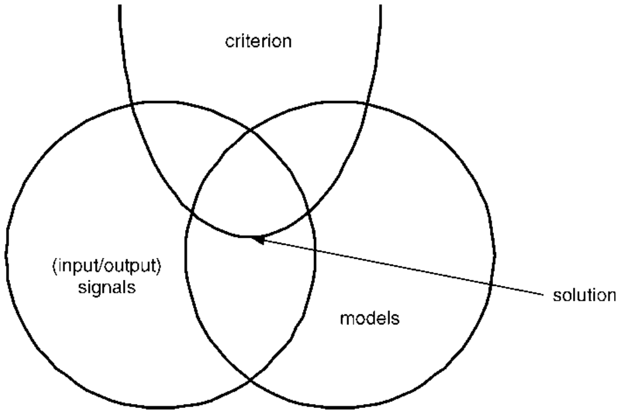


Fig. 1 *System identification problem*

Faraday's etc.) are in CT with no chances of their being rewritten in DT. Also in the human-made field of economics, models carry time-derivative related parameters which are basically in continuous time. The decline of analogue computers and the rise of digital computers spawned a great amount of literature in DT-based modelling, identification and control. Nevertheless, these developments failed to shake the faith of many traditional control engineers in CT approaches. Most of the practical control systems still rely on the ubiquitous, three-term proportional-integral-derivative (PID) controller with its predominant CT lineage. Its design is usually based on traditional CT concepts and implementation follows an 'artificial' process of discretisation. CT models are very helpful in the process of fault diagnosis by virtue of their transparency.

(ii) *CT models preserve partial knowledge:* The process of discretisation itself is associated with some undesirable consequences. In general, a strictly proper CT rational transfer function $G(s)$ with n poles, transformed into $G(z)$ in the z-plane remains rational and possesses generically $n-1$ zeros which cannot be expressed in closed form in terms of the s-plane parameters and the sampling time T_S.

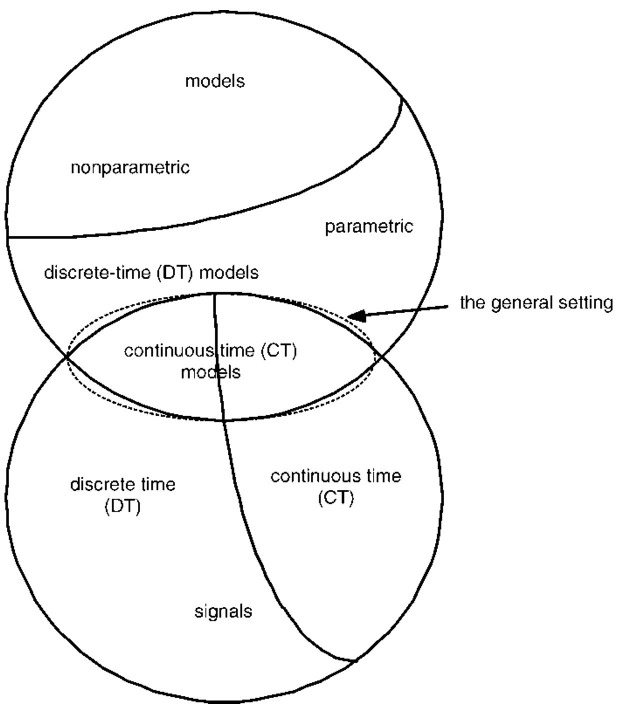


Fig. 2 *General setting for identification of continuous-time systems*

For example, even if $G(s) = K/[(s+a)(s+b)(s+c)]$ has no finite zeros and has just K as the single unknown parameter in the numerator, its DT version takes the form $G_z(z) = (b_1z^{-1} + b_2z^{-2} + b_3z^{-3})/(1 + a_1z^{-1} + a_2z^{-2} + a_3z^{-3})$, in which, in the event of identification, three 'unknown' parameters b_1, b_2 and b_3 are to be estimated. Thus, CT models preserve partial knowledge, if any, of the system parameters and efforts may be limited to the determination of the unknowns only, but such 'knowledge' is inseparably mixed with the unknowns and lost in the process of discretisation, thereby forcing the effort of identification in DT to assume full 'ignorance' of all the parameters.

(iii) *Discretisation may render CT models nonminimum phase:* In general, if $G(s)$ has m zeros and n poles, $n > m$, the corresponding $G_z(z)$ has a total of $n-1$ zeros z_{zi}, of which, as the sampling time T_S goes to zero, m approach to $z_{zi} = 1$ as $\exp(sT_S)$ and the remaining $n-m-1$ zeros approach those of the polynomial

$$B_{n-m}(z) = \sum_{i=1}^{n-m} b_i^{n-m} z^{n-m-i}$$

where

$$b_k^{n-m} = \sum_{l=1}^{k} (-1)^{k-l} l^{n-m} \binom{n-m+1}{k-l},$$
$$k = 1, 2, \ldots, (n-m),$$

and, for $n - m > 2$, lie on or outside the unit circle in the z-plane.

Consider, for example, the CT-model

$$G(s) = \frac{a}{s(s+a)}$$

then it is easily found that the corresponding pulse transfer function involving a zero-order hold is given by

$$G_z(z) = \frac{Az + B}{C(z-1)(z-D)}$$

where $A = e^{-aT} + aT - 1$, $B = 1 - e^{-aT} - aTe^{-aT}$, $C = a$ and $D = e^{-aT}$. The zero of $G_z(z)$, $z_{z1} = -(B/A)$. In the case of $a = 1$, it follows that z_{z1} lies within the unit circle in the z-plane for $T_S > 2$, but moves outside it when $T_S < 2$. For the practically interesting values, in this case of $T_S < 0.5$, the zero z_{z1} moves far outside as shown in Fig. 3. Thus, discretisation may turn a healthy native CT model into a problematic one with non-minimum phase properties. A more detailed discussion of this aspect is available in [11].

(iv) *Discretisation gives rise to undesirable sensitivity problems at high sampling rates:* The z-plane is a transformation of the s-plane through the transcendental relation $z = e^{sT_S}$. The jω-axis of the s-plane maps onto the unit circle in the z-plane with the left and right halves of the s-plane about its jω-axis transformed, respectively, into the inside and outside of the unit circle in the z-plane. In practice, the sampling frequency must be higher than the bandwidth of the system. One 'rule of thumb' is to select T_S such that $\lambda_m T_S \leq 0.5$, where λ_m is the magnitude of the largest eigenvalue of the system. In practice, it is desirable to make the sampling interval much smaller than the value specified by this rule. The result of such a choice is to force all poles to lie in a small lens-shaped region in the z-plane as shown in Fig. 4.

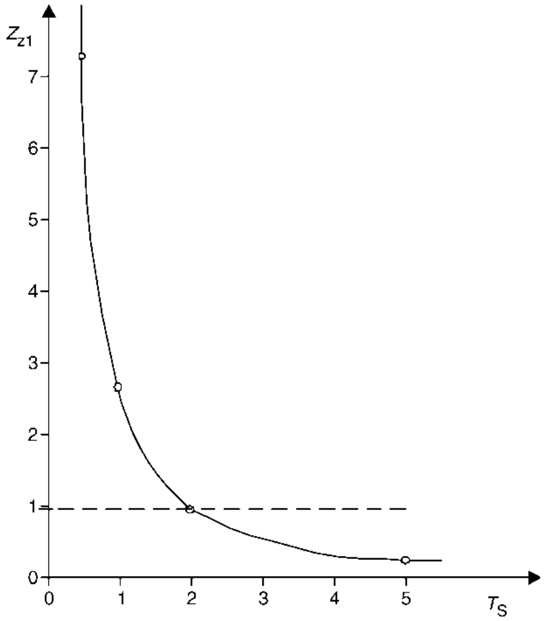


Fig. 3 *Dependence of the zero z_{z1} of $G_z(z)$ on the sampling time*

The topographical features of the z-plane are invariant with the sampling interval T_S. As $T_S \to 0$, the z-plane does not tend to become the s-plane. As $z = e^{sT_S}$, in the limit all the s-plane is mapped onto the 'black hole' (1, j0) in the z-plane. The results of DT methods do not converge in the limit to those corresponding to the original CT model. The return from the conventional discrete-time (DT) model to the original CT model is not possible without assumptions on the information between the sampling intervals.

There are unconventional discrete-time (UDT) methods for discretising of CT models. The differential equation representing the model of a continuous-time system can be discretised to obtain descriptions in the δ-domain which has different variants depending on the choice of approximation.

Table 1 shows some possibilities in which each form of δ is related to the conventional discrete-time operator q^{-1} defined by $x(t_{k-1}) = q^{-1}x(t_k)$, where $T_S = t_k - t_{k-1}$ is the sampling time.

The UDT models obtained through the δ-operator converge to their original CT models as the sampling time approaches zero, which is not the case with the usual DT models. This can be seen with reference to Fig. 4. As the δ-operator is related to z as $\delta = (z-1)/T_S$, the origin of the z-plane is shifted to the point (1, j0) and a zooming effect is introduced in direct proportion to the sampling frequency to alleviate the clustering problem. In the limit as T_S approaches zero, the δ-plane approaches the s-plane. This is true also in the case of the more general γ-operator [71, 72]. Discretisation in the delta domain, as an alternative

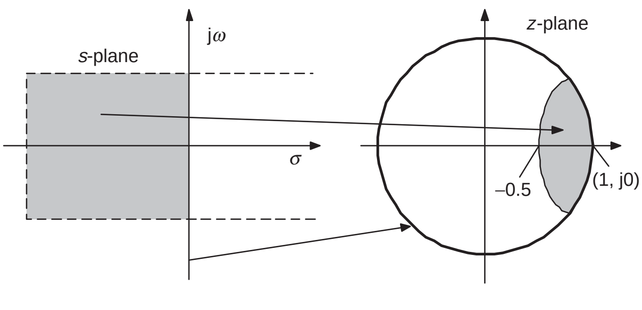


Fig. 4 *Region of normal operation in z-plane*

Table 1: Unconventional DT (UDT) operator δ based on different approximations and its relation to the conventional backward shift operator

	δ	Approximation of the derivative operator
1	$(1-q^{-1})/(q^{-1}T_S)$	forward differences based
2	$(2/T_S)(1+q^{-1})/(1-q^{-1})$	trapezoidal
3	$(1-q^{-1})/T_S$	backward differences based

to the conventional shift operator domain, avoids undue sensitivity at high sampling rates [16] and can be applied to identification, estimation and control [17].

The basic problem lies in the very nature of the relationship between s- and z-domains. The genes of the s- and z-planes are basically different, derivative and shift operations, respectively. Rapid sampling, well above the Nyquist rate, is assumed to ensure a sound return into the CT domain or the s-plane, without concern for the associated sensitivity problems. Practically speaking, the parameters clustered near (1, j0) cause several problems, particularly in finite word length representations. Thus conventional DT models are not in harmony with their descriptions in CT; in the limit of reduced sampling period, they do not converge to the results corresponding to the original CT model. The return from the conventional DT model to the original CT model is not easy. The resurgence of CT-based identification (and control) techniques during the last three decades may be attributed to those factors. Without relying any longer on analogue computers, the present techniques exploit the power of the digital tools. In all these techniques the models remain in their original CT form, or are discretised (unconventionally) to retain the same set of parameters as their CT progenitors.

2 Problem of derivatives in CT model identification

We will consider first linear time-invariant asymptotically stable dynamical systems with input $u(t)$ and output $y(t)$. The input-output description of such a SISO system, in terms of its unknown transfer function $G(s)$, is

$$Y(s) = G(s)U(s) + N(s) \tag{1}$$

A stochastic additive signal $n(t)$ has to be considered together with $y(t)$ to represent reality. In the case of a multi-input multi-output (MIMO) system, we represent the signals as $\boldsymbol{u}(t)$, $\boldsymbol{y}(t)$ and $\boldsymbol{n}(t)$ by vectors of appropriate dimensions and $\boldsymbol{G}(s)$ as the transfer function matrix.

For models which are linear in their parameters, a generic equation of the form:

$$[\textit{transposed vector of measurements}][\textit{parameter vector}] = [\textit{single measurement of output}]$$

is first developed. Using this equation, and measurements at several instants of time, a set of equations is developed and cast in the form:

$$[\textit{matrix of measurements}][\textit{parameter vector}] = [\textit{output measurement vector}]$$

The main difficulty in handling CT models is due to the presence of the derivative operator(s) $d^k/dt^k\{\cdot\}$ associated with the input and output signals. As a simple illustration of this stage, let us consider a first-order transfer function model of the SISO system

$$Y(s)/U(s) = G(s) = b/(1 + as) \tag{2a}$$

which corresponds to the differential equation

$$a\,\mathrm{d}y(t)/\mathrm{d}t + y(t) = bu(t) \tag{2b}$$

By making observations at t_k, $k = 1, 2, 3, \ldots$, a system of equations is generated as follows: $av(k) + y(k) = bu(k)$. In this, $v(k)$ is not available and any attempt at directly realising it either by numerical computation or signal processing through a differentiator will result in accentuation of noise. This difficulty is to be removed by preprocessing the signals in such a way that the undesirable derivative operations are favourably realised. Alternatively, the discretisation of the CT model is to be made in terms of an unconventional discrete time (UDT) operator that is in harmony with its CT counterpart, in the sense that the DT model converges to the original CT version as the sampling interval approaches zero. The various approaches reported in the literature may be classified with reference to the use of the ingredients in the general setting, in which the model class is denoted as CT/UDT for the reasons given in the preceding text:
(i) approaches using DT signals to identify a DT model which is then converted into native CT form,
(ii) approaches using CT signals to directly identify a native CT model, and
(iii) approaches using DT signals giving rise to a UDT model which converges to its native CT.

3 Major approaches to handle signal derivatives in the direct identification of native CT models

Let us consider the model described by (2*b*) in which $v(k) = \mathrm{d}y/\mathrm{d}t$ should be taken into account without performing direct differentiation; some measure of $v(k)$ should be used instead.

3.1 Modulating functions approach

Let the input-output data be available over an interval $[0, t_0]$. Over this interval consider a set of known modulating functions:

$$\begin{aligned}&\{\varphi_n(t)\}, n = 1, 2, \ldots, t \in [0, t_0],\\ &\quad\varphi_n(0) = \varphi_n(t_0) = \mathrm{d}\varphi_n/\mathrm{d}t|_0\\ &\qquad = \mathrm{d}\varphi_n/\mathrm{d}t|_{t_0} = 0, \quad n = 1, 2, \ldots\end{aligned}$$

for which derivatives are known up to an adequate degree. In the case of the first-order model of this example the first derivative will suffice. Multiply the differential equation (2*b*) throughout by $\varphi_n(t)$ and integrate over $[0, t_0]$ to obtain

$$a\int_0^{t_0}\varphi_n(t)\frac{\mathrm{d}y}{\mathrm{d}t}\mathrm{d}t + \int_0^{t_0}\varphi_n(t)y(t)\mathrm{d}t = b\int_0^{t_0}\varphi_n(t)u(t)\mathrm{d}t$$

Integrating the first term by parts and using the terminal conditions,

$$\int_0^{t_0}\varphi_n(t)y(t)\mathrm{d}t - a\int_0^{t_0}\frac{\mathrm{d}\varphi_n}{\mathrm{d}t}y(t)\mathrm{d}t = b\int_0^{t_0}\varphi_n(t)u(t)\mathrm{d}t, \quad n = 1, 2, \ldots$$

The signal-related terms in this equation are computable, albeit offline. The generic *transposed vector of measurement* in this case is given as

$$\left[\int_0^{t_0}\frac{\mathrm{d}\varphi_n}{\mathrm{d}t}y(t)\mathrm{d}t \;\int_0^{t_0}\varphi_n(t)y(t)\mathrm{d}t\right], \quad n = 1, 2, \ldots$$

the 'parameter vector' by $[a\ b]^T$ and the generic 'output measurement' is $\int_0^{t_0}\varphi_n(t)u(t)\mathrm{d}t$, $n = 1, 2, 3, \ldots$.

The modulating function method was first proposed by Shinbrot in 1957 [18]. It is of considerable interest in the identification of nonlinear and time-varying systems and has been used in many applications [19–23]. It may be regarded as the forerunner of many of the techniques that were developed in the subsequent period.

3.2 Poisson moment functional (PMF) approach

The Poisson moment functional (PMF) method [24–47], which is outlined in the following, is one in which the result of computation in the modulating function method becomes a measurement if values of $\varphi_n(t)$ are chosen as those arising out of the impulse response functions of the various stages of a filter chain having identical elements, each having transfer function of the form $1/(s + \lambda)$. The following transformation of a signal $\dot{y}(\tau)$ about $\tau = t$ gives its PMFs as follows:

$$M_i\{\mathrm{d}y/\mathrm{d}t\} \triangleq \int_0^t[(t-\tau)^i/i!]\exp[-\lambda(t-\tau)]\frac{\mathrm{d}y}{\mathrm{d}\tau}\mathrm{d}\tau$$

The PMFs of the derivatives of the process signals $y(t)$ and $u(t)$ can be expressed as linearly weighted sums of the PMFs of these signals themselves. In this case, the transposed generic vector of measurements is given as

$$[M_i\{y(t)\} \quad -\lambda_i M\{y(t)\} - p_i(t)y(0) \quad M_i\{u(t)\}]$$

and the generic 'output measurement' is $M_i\{y(t)\}$. In this vector, $p_i(t)$ is the inverse Laplace transform of $1/(s + \lambda)^{i+1}$.

A set of equations may be developed either by taking PMF transformation at the minimal level of i and varying time t or by PMF transformation at a fixed time t at different levels of i or a combination of both. It is the former strategy that is usually preferred for its simplicity and possibility for online implementation. Higher-order derivative terms of the process signals give rise to their initial values in the measurement vector. As these are unknown, they should be separated and included in the parameter vector as additional unknowns to be estimated together with the usual system parameters. When coupled with a simulation stage, the combined algorithm becomes one of joint state and parameter estimation that is of considerable importance. If λ is chosen to be very large relative to the time constants of the system under identification, and if the PMF transformation is taken about a large time t, the effect of the initial conditions becomes insignificant. Consequently, the terms associated with them can be dropped from the measurement vector. The resulting algorithm estimates only the usual system parameters. However, this estimation would be at the cost of excessive passage of noise through the measurements into the estimates. The book by Saha and Rao [9] is devoted to the PMF method and its many aspects. The basic PMF method was implemented in a microprocessor [34]. Further developments [37–47] include studies on the design of the Poisson filter. The Poisson filter element in its originally proposed form with transfer function $1/(s + \lambda)$ is not normal; its gain at zero frequency depends on λ and the resulting PMFs are termed as ordinary

PMFs (OPMF). With λ in the numerator it is rendered normal and gives rise to normal PMFs (NPMF). A parameter β that is different from λ in the numerator provides an additional degree of freedom in the filter design and gives rise to the so-called generalised PMFs (GPMF). The resulting parameter estimation algorithms are general and flexible in that the Poisson filter may be chosen to meet the needs of the situation. The parameter estimation algorithms are realised in recursive least squares form that is convenient for online applications. Bias-compensation devices have been proposed for the PMF-based least-squares algorithms and they are applied to both open-loop and closed-loop systems. The algorithms have been applied in state space and the related subspace methods that were originally proposed in DT have been extended to CT. The PMF method has been successfully used in practical applications [43, 47].

3.3 Integral equation approach

PMF transformation with $\lambda = 0$ is of particular significance. This leads to the so-called integral equation approach. Repeated integration performed on the original differential equation removes all the derivative terms and paves a way to successful application of parameter estimation algorithms. Diamessis [48–50] was one of the earliest to report on the integration-based approach. The need to realise the integral operation on process signals spurred considerable research and gave rise to a number of useful methods.

The integral operation using a numerical integration formula has been automatically realised by means of a digital filter on sampled process signals [51–61]. Bias compensation features and special problems have also been studied. A typical linear integration filter (a transversal filter) extensively applied is shown in Fig. 5. This is a digital-signal-processing module in which q^{-1} denotes a shift operator and $\{p_i\}$ a set of weights specified by the chosen integration formula.

3.4 Orthogonal functions (OF) approach

An interesting way to realise the integrals in the 'integral equation approach' is by representing the process signals in a series of orthogonal functions $\{\theta_i(t),\ i = 1,\ 2, \ldots,\ \infty\}$ over the interval $[0, t_0]$. For the sake of simplicity of illustration, let us consider the first two components of the expansion in the case of all the signals involved in the example:

$$y(t) \simeq y_1\theta_1(t) + y_2\theta_2(t)$$
$$u(t) \simeq u_1\theta_1(t) + u_2\theta_2(t)$$

and insert them in the corresponding integral equation of the system described by (2*a*)

$$ay(t) - ay(0)s(t) + \int_0^t y(\tau)\mathrm{d}\tau = b\int_0^t u(\tau)\mathrm{d}\tau, \quad 0 \le t \le t_0$$

in which $s(t)$ denotes a unit step function at $t = 0$, having

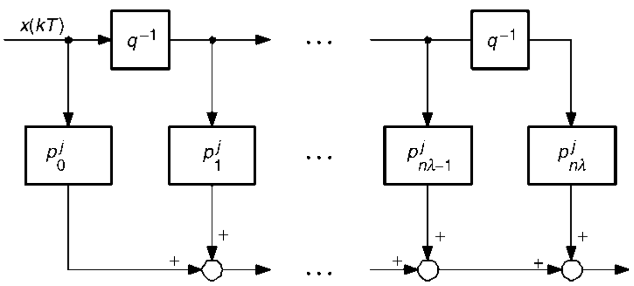


Fig. 5 *Linear integrating filter*

s_1 and s_2 as its spectral components of $\theta_1(t)$ and $\theta_2(t)$. Further let

$$\int_0^t \theta_1(\tau)\mathrm{d}\tau \simeq e_{11}\theta_1(t) + e_{12}\theta_2(t)$$
$$\int_0^t \theta_2(\tau)\mathrm{d}\tau \simeq e_{21}\theta_1(t) + e_{22}\theta_2(t)$$

This integral equation is transformed into algebraic form in which the 'measurement matrix' becomes

$$\begin{bmatrix} y(0)s_1 - y_1 & u_1e_{11} + u_2e_{21} \\ y(0)s_2 - y_2 & u_1e_{12} + u_2e_{22} \end{bmatrix}$$

and the output measurement vector takes the form

$$\begin{bmatrix} y_1e_{11} + y_2e_{21} \\ y_1e_{12} + y_2e_{22} \end{bmatrix}$$

The integral equation approach has been hosted by a wide range of systems of orthogonal functions. These include the systems of piecewise-constant functions [10, 12] such as Walsh, Haar and block pulse functions (BPF) and the systems of continuous functions such as Fourier, Chebyshev, Jacobi, Laguerre, Legendre, Hermite polynomials [13]. The class of general hybrid orthogonal functions (GHOF) proposed by Patra and Rao [14] capture the features of continuity of the continuous systems and of discontinuities of the piecewise-constant systems. The GHOF are capable of efficiently representing a wide range of signals encountered in practice, including those occurring in switched systems. In [14] an extensive list of bibliography on the subject of orthogonal functions is given. The list is mapped onto different fields of applications in systems and control in separate tabular summaries.

Among the systems of orthogonal functions, the BPF are of particular importance by virtue of their simplicity and fundamental nature. Initially for some time, the set of BPF was erroneously held in suspicion of being incomplete, until Rao and Srinivasan [62] definitively established their completeness. This seems to have spurred considerable activity [63–68] in the use of BPF in identification of CT systems. The OF approach was inspired by a paper by Corrington [69] in which Walsh functions have been used in the solution of differential and integral equations. Chen and Hsiao [70] and Rao and Sivakumar [71] independently came up with a method using Walsh functions that is directly addressed to the problem of identification of CT systems. The operational matrix for approximating the integral operation was introduced which reduces models of continuous-time dynamical systems into a computationally convenient form. Rao and Tzafestas [72] surveyed the developments in the use of orthogonal functions during 1975–1985.

4 Unified framework for identification of CT systems

CT model identification is represented in two stages, the first involving a linear dynamic operation (R_{LD}):

4.1 Preparatory stage

The need to generate the time-derivative terms in CT models is eliminated by a class of signal-processing techniques denoted by the operation R_{LD} in this stage. Figure 6 shows the family tree of the various methods denoted by the class R_{LD}. The state variable filter (SVF)

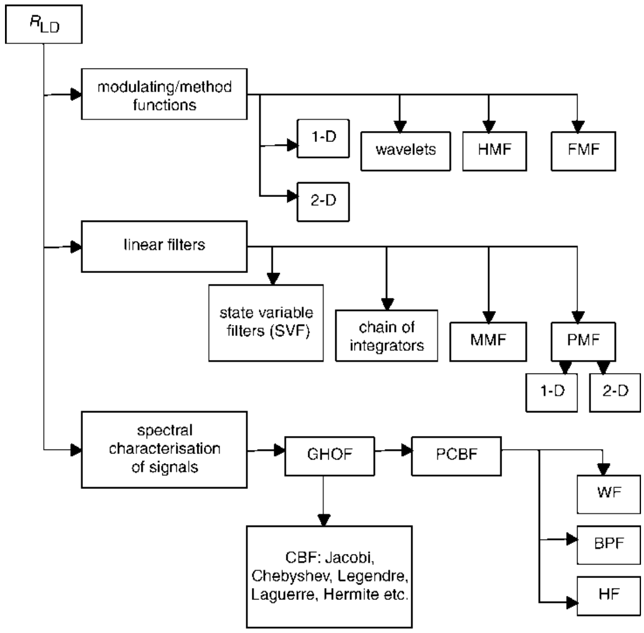


Fig. 6 *Several manifestations of R_{LD}*

MMF = method of multiple filters; PMF = Poisson moment functionals; GHOF = general hybrid orthogonal functions; PCBF = piecewise constant basis functions; CBF = continuous basis functions; WF = Walsh functions; HF = Haar functions; BPF = block pulse functions; HMF = Hartley modulating functions; FMF = Fourier modulating functions

may be regarded also as the forerunner of all the linear filters used in this connection apart from the connections illustrated here among all the variants of the operation R_{LD}. By applying R_{LD} to the input and output data, the system of equations for parameter estimation is generated from the model equation using operational matrices in a matrix-algebraic framework that was developed during the 1970s and 1980s.

The various possibilities for different methods of discretisation of CT models and the relations among them, together with their relationship with signal preprocessing R_{LD}, are given in [73, 74]. From the resulting system of equations, the parameters can be estimated either by an en-bloc computation or by a recursive algorithm. In particular, the BPF method and the PMF method lend themselves to recursive estimation. The δ-operator, as it is known, refers only to the version based on the backward-shift operator. In its more general form it is referred to as the γ-operator. The γ form refers to this case. The case of least-squares parameter estimation of continuous-time ARX models from discrete-time data using the δ-operator has been studied in [75].

4.2 Estimation stage

After the preparatory stage, we now enter the estimation stage. In the estimation stage, standard procedures are applied as in the case of identification of discrete-time systems. The standard methodology used in DT-based methods is applicable here by virtue of the discrete nature of the entities involved here. The discrete entities arise out of the operations in the preparatory stage on continuous-time models and the related signals. Referring to the schemes of Figs. 7 and 8, we will discuss the various approaches in this stage.

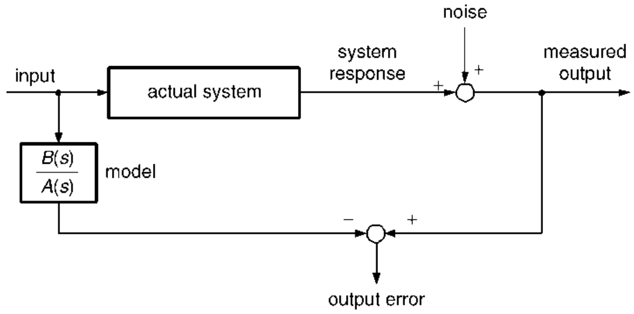


Fig. 7 *Output error scheme*

The model of the actual SISO system is considered in the form:

$$Y(s) = G_M(s)U(s) + N(s) = G_M(s)U(s) + H(s)W(s) \quad (2c)$$

in which the second term on the right-hand side (RHS) of (2*c*) accounts for the combined effects of $n(t) = \mathsf{L}^{-1}\{N(s)\}$, stochastic disturbances, unmodelled dynamics (due to model simplification) and possibly of unknown initial conditions. This term is generally referred to as the noise model and $w(t)$ denotes white noise. The rational model transfer function $G_M(s) = B(s)/A(s)$ contains the parameters to be estimated. The stochastic part of the model is described by the rational transfer function $H(s)$. Table 2 summarises the different choices of $H(s)$ which give rise to different model structures.

Consider the set of measurements sampled at equal intervals of length T_S being represented by the measurement vector:

$$\boldsymbol{y}(N) = [u(k), y(k), k = 1, \ldots, N] \quad (3)$$

Given $\boldsymbol{y}(N)$ and some prior knowledge of the dynamics of the system, the identification problem is to obtain $G_M(s)$ in terms of its parameters which best describe the dynamics of the system in some sense by minimising a chosen norm of the modelling error. Let $G_M(s, \boldsymbol{\theta})$ denote an estimate of the transfer function, where $\boldsymbol{\theta} \in \Re^n$ is the parameter vector. In terms of $G_M(s, \boldsymbol{\theta})$, the input-output description becomes

$$y(t) = \mathsf{L}^{-1}\{G_M(s, \boldsymbol{\theta})U(s) + H(s, \boldsymbol{\theta})W(s)\} \quad (4)$$

The focus of our attention in the present treatment is on the first term in (4). The treatment is also applicable to the second term in (4). A most general parametric form of

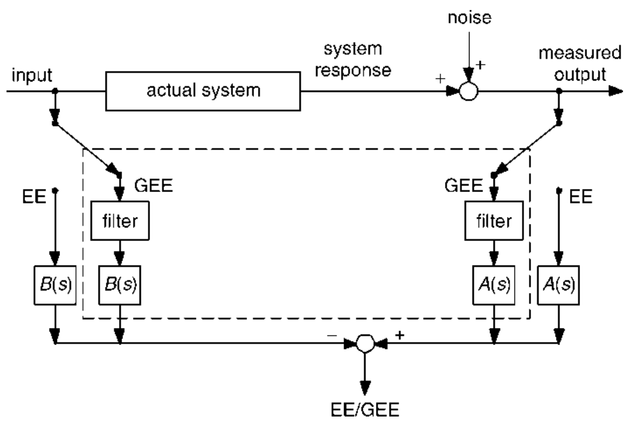


Fig. 8 *Equation error scheme(s)*

Table 2: Different approaches to parameter estimation

Model structure	Name	$H(s)$
I (ARX)	least squares (LS), instrumental variable (IV)	$1/A(s)$
II (ARMAX)	extended matrix model–I	$C(s)/A(s)$
III (ARARX)	extended matrix model–II	$1/[A(s)D(s)]$
IV (ARARMAX)	general	$1/[A(s)D(s)]$

description for (4) is the polynomial black-box model:

$$A(\delta, \boldsymbol{\theta})y(k) = \frac{B(\delta, \boldsymbol{\theta})}{E(\delta, \boldsymbol{\theta})}u(k) + \frac{C(\delta, \boldsymbol{\theta})}{D(\delta, \boldsymbol{\theta})}w(k) \qquad (5)$$

that is the CT counterpart of the DT Box-Jenkins version [10]. In this, δ denotes a DT approximation to the CT differentiation operator d/dt and not the usual backward-shift operator. $u(k)$ and $y(k)$ are the samples of input and output signals, respectively, and $w(k)$ is a sequence of independent and uniformly distributed zero-mean random variables. $A(\delta, \boldsymbol{\theta})$, $B(\delta, \boldsymbol{\theta})$, $C(\delta, \boldsymbol{\theta})$, $D(\delta, \boldsymbol{\theta})$ and $E(\delta, \boldsymbol{\theta})$ are polynomials in δ whose coefficients are arranged to form the parameter vector $\boldsymbol{\theta}$. Specific cases of these polynomials lead to particular models such as autoregressive (AR), moving average (MA), autoregressive and moving average (ARMA) and so on. In particular, to characterise stationary stochastic processes, the following ARMA model is considered:

$$A(\delta, \boldsymbol{\theta})y(k) = C(\delta, \boldsymbol{\theta})w(k) \qquad (6)$$

where $A(\delta, \boldsymbol{\theta}) = \delta^{n_A} + a_1\delta^{n_A-1} + \cdots + a_{n_A}$ and $C(\delta, \boldsymbol{\theta}) = c_0\delta^{n_C} + c_1\delta^{n_C-1} + \cdots + c_{n_C}$, with coefficients which appear in the AR and MA portions, respectively, of the model. In the present context, the terms AR, MA and ARMA refer to the CT context. The ARMA model (6) is commonly used in spectral estimation and time-series analysis. In system identification, where the goal is to characterise the dynamic input-output relation of the underlying process, the following ARX model has to be applied:

$$A(\delta, \boldsymbol{\theta})y(k) = B(\delta, \boldsymbol{\theta})u(k) + w(k) \qquad (7)$$

where $A(\delta, \boldsymbol{\theta}) = \delta^n + a_1\delta^{n-1} + \cdots + a_n$ and $B(\delta, \boldsymbol{\theta}) = b_1\delta^{n-1} + b_2\delta^{n-2} + \cdots + b_n$. Here the MA part is formed from the usually known process input signal. Without resorting to creating a new name, we refer to

$$G(\delta, \boldsymbol{\theta}) = \frac{B(\delta, \boldsymbol{\theta})}{A(\delta, \boldsymbol{\theta})} \qquad (8)$$

as 'deterministic ARMA'. This is nonlinear in the parameters. With this model structure, the model output error (OE) in sampled form is

$$\varepsilon_{OE}(k) = y(k) - \frac{B(\delta, \boldsymbol{\theta})}{A(\delta, \boldsymbol{\theta})}u(k) \qquad (9)$$

A parameter estimation criterion is to minimise, for example, the error function

$$J_{OE}(k) = \sum_{k=1}^{N} \varepsilon_{OE}^2(k) \qquad (10)$$

with respect to $\boldsymbol{\theta}$. As the output error of (9) is nonlinear in parameters, this is a case of nonlinear optimisation. In an attempt to simplify the situation, most of the identification approaches resort to the equation error (EE):

$$\varepsilon_{EE}(k) = A(\delta, \boldsymbol{\theta})y(k) - B(\delta, \boldsymbol{\theta})u(k) \qquad (11a)$$

or the generalised equation error (GEE)

$$\varepsilon_{GEE}(k) = \frac{A(\delta, \boldsymbol{\theta})}{C(\delta, \boldsymbol{\theta})}y(k) - \frac{B(\delta, \boldsymbol{\theta})}{C(\delta, \boldsymbol{\theta})}u(k) \qquad (11b)$$

and as a criterion to minimise

$$J_{EE}(k) = \sum_{k=1}^{N} \varepsilon_{EE}(k)^2 \qquad (12)$$

where $1/E(\delta)$ is a linear-dynamic operator belonging to the class R_{LD}, of adequate order for the removal of the need for direct differentiation of process data [10]. These operators also serve the purpose of prefilters used for removing unimportant frequencies from the process data. As (11) is linear-in-parameters, parameter estimation is simplified to linear direct or recursive least-squares (RLS) estimation. However, EE minimisation has its disadvantages, as will be discussed as follows:

4.2.1 Biased estimation: The parameter estimates will be biased when the EE is not white [76]. Variants of the ordinary least-squares (LS) algorithm such as 'generalised least-squares' and 'instrumental variables' [3] are applied to remove the bias. These and other 'bias compensating least-squares' methods [59, 61] required additional computational effort for bias compenstion. These approaches assume that the measurements are actually generated by an ARMA model. Some methods assume that the measurement noise is Gaussian. However, the performance of some of these may not be satisfactory when there is a significant modelling error, as this component of error may not be Gaussian.

4.2.2 Reducible models (for MIMO systems): Consider a v_i-input v_o-output system described by the transfer function matrix (TFM):

$$\boldsymbol{G}(\delta, \boldsymbol{\theta}) = \begin{bmatrix} G_{11}(\delta, \boldsymbol{\theta}) & G_{12}(\delta, \boldsymbol{\theta}) & \dots & G_{1v_i}(\delta, \boldsymbol{\theta}) \\ G_{21}(\delta, \boldsymbol{\theta}) & G_{22}(\delta, \boldsymbol{\theta}) & \dots & G_{2v_i}(\delta, \boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots \\ G_{v_o1}(\delta, \boldsymbol{\theta}) & G_{v_o2}(\delta, \boldsymbol{\theta}) & \dots & G_{v_ov}(\delta, \boldsymbol{\theta}) \end{bmatrix} \qquad (13)$$

where

$$G_{ij}(\delta, \boldsymbol{\theta}) = \frac{B_{ij}(\delta, \boldsymbol{\theta})}{A_{ij}(\delta, \boldsymbol{\theta})}$$

EE formulation necessitates a canonical form having a least common denominator (CD) of all the elements of the TFM. The CD considerably inflates the unknown parameter vector. To reduce this inflation partially, the TFM is decomposed into multiple-input/single-output (MISO) submodels with several CDs limited only to the rows of the TFM. In this way, a two-stage algorithm was proposed in [77] for DT model identification, and its CT version in [78]. A Gauss-Seidel-type iterative algorithm that does not require a CD was later suggested in [79].

4.2.3 Distribution of estimation errors: Modelling of physical processes is usually associated with a certain amount of undermodelling. This coupled with noise in the measurements, results in biased estimates. Although it is easy to eliminate bias due to measurement noise, the bias resulting from undermodelling can be asymptotically eliminated only by a structured noise model under certain assumptions. It can be distributed over a range of frequencies by careful design of the identification experiment [80], so that such undermodelling is not harmful in the context of the final application of the resulting model. With ARMA modelling, the problem of experiment design for a prescribed distribution of bias over a range of frequencies is not simple and straightforward.

Using Parseval's theorem, the frequency-domain description of the GEE criterion (12), in the limit as $T_S \to 0$, is

$$J_{\mathrm{GEE}}(\omega) = \int_0^\infty \left| \frac{A(j\omega, \boldsymbol{\theta})}{E(j\omega)} U(j\omega) \right|^2 \left| G_0(j\omega) - \frac{B(j\omega, \boldsymbol{\theta})}{A(j\omega, \boldsymbol{\theta})} \right|^2 d\omega \tag{14}$$

where $U(j\omega)$ is the Fourier transform of the input signal and G_0 denotes the true system. The first term on the right-hand side of (14) can be considered as a weighting function that manipulates the second term (bias) over a range of frequencies. With the chosen ARMA model structure, it is clear that this weighting function is a function of the yet unknown $A(\delta, \boldsymbol{\theta})$, which renders online experiment design as impossible. Offline design, however, is shown to be possible by Bapat [37].

5 Models for linear estimation (moving average forms)

$G(\delta, \boldsymbol{\theta})$ is linear in the parameters, if second- and higher-order derivatives of $G(\delta, \boldsymbol{\theta})$, with respect to $\boldsymbol{\theta}$, vanish for all $\boldsymbol{\theta}$, and linearity of a parametrisation is different from the linearity of the model in terms of its input-output behaviour. Even nonlinear models can be linearly parametrised. One situation in which the ARX model of (7) is linearised (with respect to $\boldsymbol{\theta}$) is when its denominator $A(\delta, \boldsymbol{\theta})$ is fixed as some appropriate $A(\delta)$ which leads to the description

$$G(\delta, \boldsymbol{\theta}) = \sum_{i=1}^{n-1} \frac{b_i \delta^{n-i}}{A(\delta)} = \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{b}_d(\delta) \tag{15}$$

in which $\boldsymbol{\theta} = [b_1\ b_2 \cdots b_n]^{\mathrm{T}}$ and

$$\boldsymbol{b}_d(\delta) = \frac{1}{A(\delta)} [\delta^{n-1} \quad \delta^{n-2} \quad \cdots \quad \delta^0]^{\mathrm{T}}$$

A linear-in-parameters model is therefore obtained as

$$y(k) = \frac{B(\delta, \boldsymbol{\theta})}{A(\delta)} u(k) \tag{16}$$

whereby EE = OE and estimation (minimisation) is linear. This leads to an advantageous situation with the following possibilities:

Robust estimation: In the limit as $N \to \infty$, the LS estimate $\hat{\boldsymbol{\theta}}$ in the presence of zero-mean disturbances tends to $\hat{\boldsymbol{\theta}}^*$, where $\hat{\boldsymbol{\theta}}^*$ is the limiting estimate in the absence of disturbances. In particular, if the disturbance term is Gaussian and there is no modelling error, the LS estimate $\hat{\boldsymbol{\theta}}$ is asymptotically normal with mean $\hat{\boldsymbol{\theta}}^*$ and a covariance proportional to the variance of the disturbance. This holds good even for coloured disturbances uncorrelated with the input. This implies that the LS estimation is robust to zero-mean disturbances. Note that the estimates will still be 'biased' due to the inherent undermodelling.

Irreducible model estimation: With MIMO TFM models, because the denominators do not include unknown parameters, the CD formulation does not inflate the parameter vector.

Simplified error distribution problem: The weighting function in (14) now equals

$$\left| \frac{A(j\omega)}{E(j\omega)} U(j\omega) \right|^2$$

The absence of the unknown $\boldsymbol{\theta}$ in this weighting function permits online experiment design for a prescribed bias distribution.

Gray-box modelling: The fixed denominator polynomial $A(\delta)$ in the linear-in-parameters model (15) serves as an additional design variable allowing for effective incorporation of prior knowledge of the process dynamics. By an intelligent choice of this polynomial, even complex systems can be estimated significantly accurately with a smaller number of parameters.

These are the advantages of linear-in-parameters models in system identification. In these models the output is expressed as a linear combination of certain MA components of the input. This leads to the 'generalised moving average model' (GMAM) formulation as

$$y(k) = \sum_{i=1}^{n} \theta_i F_i(\delta) u(k) \tag{17}$$

in which model, the moving-average components of the inputs are formed as the responses of a set of known filters $\{F_i(\delta)\}$ to $u(k)$. These filters form the basis

$$\boldsymbol{b}_{\mathrm{F}}(\delta) = [F_1(\delta) \quad F_2(\delta) \quad \cdots \quad F_n(\delta)]^{\mathrm{T}} \tag{18}$$

of the GMAM structure

$$G(\delta, \boldsymbol{\theta}) = \sum_{i=1}^{n} \theta_i F_i(\delta) \tag{19}$$

With such a parametrisation, the model output error

$$\varepsilon_{\mathrm{OE}}(k) = y(k) - \sum_{i=1}^{n} \theta_i F_i(\delta) u(k)$$

is linear in the set of parameters $\{\theta_i\}$, and, consequently, the minimisation problem of the output error criterion (10) is linear.

Such models evolve very naturally from truncated power-series expansions of the rational transfer function. For example, in the DT case, the system transfer function may be written as

$$G^0(z) = \sum_{i=1}^{\infty} h_i z^{-i} \tag{20}$$

where $\{h_i\}$ is the impulse response sequence. This is approximated as

$$G(z, \boldsymbol{\theta}) = \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{b}_{q1}(z^{-1}) \tag{21}$$

where $\boldsymbol{\theta} = [\theta_1\ \theta_2 \cdots \theta_n]^{\mathrm{T}}$ and $\boldsymbol{b}_{q1}(z^{-1}) = [z^{-1}\, z^{-2} \cdots z^{-n}]^{\mathrm{T}}$. The quality of this approximation depends on the rate of convergence of the impulse-response sequence. The poles of $G^0(z)$ close to the unit circle in the z-domain slow down the rate of convergence. Consequently, a high model order is required for a given tolerance. For these reasons, in rapidly sampled CT systems the rate of convergence of the approximation will be very slow, and in the limit as $T_S \to 0$, the DT poles approach unity and, consequently, the approximation fails to converge.

Furthermore, even in the case of convergent approximations, high model order is required, as the memory of the basis (shift operator) is very short (unity). Hence, model representations having better convergence properties and less sensitivity to sampling rate will be preferable.

In the CT case, the transfer function $G(s)$ may be expanded about $s \to \infty$, as a complex power series in s^{-1} as

$$G^0(s) = \sum_{i=1}^{\infty} h_i (s^{-1})^i \tag{22}$$

leading to the form

$$G(s, \boldsymbol{\theta}) = \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{b}_{s1}(s^{-1}) \tag{23}$$

where $\boldsymbol{\theta} = [h_1\ h_2 \cdots h_n]^{\mathrm{T}}$ and $\boldsymbol{b}_{s1}(s^{-1}) = [s^{-1}\ s^{-2} \cdots s^{-n}]^{\mathrm{T}}$. It is well known that h_i are the CT Markov parameters of $G^0(s)$ which are defined as

$$h_i = \left. \frac{\mathrm{d}^{i-1}}{\mathrm{d}t^{i-1}} g^0(t) \right|_{t=0} \tag{24}$$

where $g^0(t)$ is the impulse response of $G^0(s)$.

Considering a similar expansion of $G^0(s)$ about $s = 0$, we have models parametrised in terms of normalised time moments of the impulse response $g^0(t)$ of $G^0(s)$, i.e.

$$G(s, \boldsymbol{\theta}) = \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{b}_s(s) \tag{25}$$

where $\boldsymbol{\theta} = [m_1\ m_2 \cdots m_n]^{\mathrm{T}}$ and $\boldsymbol{b}_s(s) = [s\ s^2 \cdots s^n]^{\mathrm{T}}$ and

$$m_i = \frac{(-1)^i}{i!} \int_0^{\infty} t^i g^0(t) \mathrm{d}t \tag{26}$$

are the normalised time moments.

Other basis functions are also possible. Well known among these are Laguerre and Kautz filters. Laguerre filters imply a basis,

$$\boldsymbol{b}_{\mathrm{LAG}}(s) = \left[\frac{1}{s+\lambda} \quad \frac{1}{s+\lambda}\left(\frac{s-\lambda}{s+\lambda}\right) \cdots \frac{1}{s+\lambda}\left(\frac{s-\lambda}{s+\lambda}\right)^{n-1} \right]^{\mathrm{T}} \tag{27}$$

with $\lambda > 0$, and Kautz filters imply

$$\boldsymbol{b}_{\mathrm{KAUTZ}}(s) = [\psi_1(s) \quad \psi_2(s) \quad \cdots \quad \psi_n(s)]^{\mathrm{T}} \tag{28}$$

where

$$\psi_{2k-1}(s, b, c) = \frac{s}{s^2 + bs + c} \left[\frac{s^2 - sb + c}{s^2 + sb + c} \right]^{k-1}$$

and

$$\psi_{2k}(s, b, c) = \frac{1}{s^2 + bs + c} \left[\frac{s^2 - bs + c}{s^2 + bs + c} \right]^{k-1}$$

with $b > 0$, $c > 0$ and $k = 1, 2, \ldots$. These bases have been discussed in greater detail in [81]. The role played by the basis in continuous- and discrete-system modelling is discussed in [82].

5.1 Markov parameter models

In CT situations, the practical use of Markov parameters is rare. This is because of the natural but difficult-to-compute form (24) in which Markov parameters are defined for CT systems. The work of Dhawan *et al.* [83] is the first attempt at the use of MP models for SISO CT model identification. The MP model (23) is transformed into an integral equation in which the integrals are realised using block-pulse functions [12], thereby avoiding the derivative route to the realisation of Markov parameters. However, truncation of the MP model as in (23) often leads to poor approximation, due to which the estimation may fail to converge. A simple generalisation of the original MP model to ensure convergent approximations may be found in [83]. Further generalisation of the basis leading to flexible and well-behaved approximations was suggested in [84–87].

Estimation of moving-average models: Consider a v_i-input, v_o-output MIMO system having a transfer matrix $\boldsymbol{G}^0(s)$ and an input-output relationship

$$\boldsymbol{Y}(s) = \boldsymbol{G}^0(s)\boldsymbol{U}(s) + \boldsymbol{N}(s) \tag{29}$$

where $\boldsymbol{Y}(s) \in \Re^{v_o}$, $\boldsymbol{U}(s) \in \Re^{v_i}$ and $\boldsymbol{N}(s) \in \Re^{v_o}$ are the Laplace transformed signal vectors. CT Markov parameters of this system are defined as the coefficients of the power series:

$$\boldsymbol{G}^0(s) = \sum_{l=1}^{\infty} \boldsymbol{H}_l s^{-1} \tag{30}$$

where $\{\boldsymbol{H}_i\}$ is the matrix of Markov parameter sequence (MPS), and denote

$$\boldsymbol{H}_l = \begin{bmatrix} h_{l,11} & h_{l,12} & \ldots & h_{l,1v_i} \\ h_{l,21} & h_{l,22} & \ldots & h_{l,2v_i} \\ \vdots & \vdots & \ddots & \vdots \\ h_{l,v_o1} & h_{l,v_o2} & \ldots & h_{l,v_ov_i} \end{bmatrix}$$

In terms of the MPS,

$$\boldsymbol{y}(t) = \sum_{l=1}^{\infty} \boldsymbol{H}_l \boldsymbol{u}^l(t) + \boldsymbol{n}(t) \tag{31}$$

where $\boldsymbol{u}^l(t)$ is the lth integral of $\boldsymbol{u}(t)$. Assuming absolute convergence of the MPS and thus uniform convergence of partial sums, a truncated MP model is obtained as

$$\boldsymbol{y}(t) = \sum_{l=1}^{n} \boldsymbol{H}_l \boldsymbol{u}^l(t) + \boldsymbol{e}(t) \tag{32}$$

where $\boldsymbol{e}(t)$ includes the truncation (of the MPS) error and the contribution of unknown initial conditions in addition to the usual noise term $\boldsymbol{n}(t)$. This model is valid only when the power-series expansion of $\boldsymbol{G}^0(s)$ is absolutely convergent. Note that, when the system is represented in the sampled domain as $\boldsymbol{G}^0(s^{-1})$, the resulting DT MPS is the impulse-response sequence of the system. For asymptotically stable systems, the DT MPS is absolutely convergent, but, when represented in CT domain, even stable systems may have diverging MP sequences. To ensure absolute convergence and to increase the rate of convergence of the approximation, a more general version of Markov parameters, known as Markov-Poisson parameters, were suggested by Subrahmanyam and Rao [84]. In terms of these parameters, $\boldsymbol{G}^0(s)$ is expanded as

$$\boldsymbol{G}^0(s) = \sum_{l=1}^{\infty} \bar{\boldsymbol{H}}_l \left(\frac{\beta}{s+\lambda}\right)^l \boldsymbol{U}(s) \tag{33}$$

The matrix of Markov-Poisson parameters $\{\bar{\boldsymbol{H}}_l\}$ is related to the matrix of Markov parameters $\{\boldsymbol{\theta}_l\}$ as

$$\boldsymbol{H}_l^* = \frac{1}{\beta^l} \sum_{l=1}^{l} {}^{l-1}C_{l-1} \lambda^{l-1} \boldsymbol{H}_l, \quad l = 1, \ldots; \beta > 1 \tag{34}$$

Thus, the model is

$$\boldsymbol{Y}(s) = \sum_{l=1}^{n} \bar{\boldsymbol{H}}_l \left(\frac{\beta}{s+\lambda}\right)^l \boldsymbol{U}(s) + \boldsymbol{E}(s) \tag{35}$$

implying the basis

$$\boldsymbol{b}_{\mathrm{PF}}(s) = \left[\frac{\beta}{s+\lambda} \quad \left(\frac{\beta}{s+\lambda}\right)^2 \quad \cdots \quad \left(\frac{\beta}{s+\lambda}\right)^n\right]^{\mathrm{T}} \tag{36}$$

The elements of $\boldsymbol{b}_{\mathrm{PF}}(s)$ are the well-known Poisson filters [9] of increasing order, in which λ and β are tunable parameters. This generalisation improves the low-frequency predictive ability of the model. The choice of the filter parameter λ has to be made according to the *a priori* knowledge of the poles of the system. In general, a $\lambda > 0$ is well suited for overdamped systems with poles not very close to the imaginary axis of the s-plane. On the other hand, a $\lambda < 0$ with a large β is appropriate when the poles (complex) of the system are arbitrarily close to the imaginary axis.

Parameter estimation may now be carried out by decomposing the problem into v_o MISO subproblems and considering one subproblem at a time or in parallel. In the sequel, one such MISO problem is considered and the subscript i, that denotes the row index, is dropped mainly for notational simplicity. Further, only n_j parameters are considered for the jth element of the MISO problem. Approximating the derivative operator by δ in the parameter estimation equation we obtain

$$\tilde{y}(k) = \boldsymbol{\varphi}^{\mathrm{T}}(k)\boldsymbol{\theta} \tag{37}$$

where

$$\begin{aligned}
\boldsymbol{\varphi}(k) &= [\boldsymbol{\varphi}_1^{\mathrm{T}}(k) \quad \boldsymbol{\varphi}2^{\mathrm{T}}(k) \quad \cdots \quad \boldsymbol{\varphi}_{v_i}^{\mathrm{T}}(k)]^{\mathrm{T}} \\
\boldsymbol{\varphi}_j^{\mathrm{T}}(k) &= [F_1(\delta)u_j(k) \quad F_2(\delta)u_j(k) \quad \cdots \quad F_{n_j}(\delta)u_j(k)]^{\mathrm{T}}, \\
& \qquad j = 1, \ldots, v_i, \\
\boldsymbol{\theta} &= [\bar{h}_{1,1}, \ldots \bar{h}_{n_1,1} | \cdots | \bar{h}_{1,v_i}, \ldots, \bar{h}_{n_{vi},v_i}]^{\mathrm{T}}
\end{aligned}$$

and

$$F_l(\delta) = \left(\frac{\beta}{\delta+\lambda}\right)^l$$

Next, we define the cost function as

$$\begin{aligned}
J(\boldsymbol{\theta}) &= [\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(0)]^{\mathrm{T}}\boldsymbol{P}(0)^{-1} \\
&\quad \times [\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(0)] + \sum_{k=1}^{N} [y(k) - \boldsymbol{\varphi}^{\mathrm{T}}(k)\boldsymbol{\theta}(k)]^2
\end{aligned} \tag{38}$$

The LS estimate that minimises $J(\boldsymbol{\theta})$ is

$$\begin{aligned}
\hat{\boldsymbol{\theta}}(N) &= \left[\boldsymbol{P}(0)^{-1} + \sum_{k=1}^{N} \boldsymbol{\varphi}(k)\boldsymbol{\varphi}^{\mathrm{T}}(k)\right]^{-1} \\
&\quad \times \left[\boldsymbol{P}(0)^{-1}\hat{\boldsymbol{\theta}}(0) + \sum_{k=1}^{N} \boldsymbol{\varphi}(k)y(k)\right]
\end{aligned} \tag{39}$$

provided the inverse exists. This estimate may be calculated using the conventional recursive least-squares algorithm.

Irreducible ARMA model realisation: Given the estimates of $\bar{\boldsymbol{H}}_l$, $l = 1, \ldots, n$, the first step towards realisation of an irreducible ARMA TFM model is to examine the columns or rows of the Hankel matrix $\boldsymbol{H}(p, q)$ formed from the estimates as

$$\boldsymbol{H}(p,q) = \begin{bmatrix} \bar{\boldsymbol{H}}_1 & \bar{\boldsymbol{H}}_2 & \cdots & \bar{\boldsymbol{H}}_q \\ \bar{\boldsymbol{H}}_2 & \bar{\boldsymbol{H}}_3 & \cdots & \bar{\boldsymbol{H}}_{q+1} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\boldsymbol{H}}_p & \bar{\boldsymbol{H}}_{p+1} & \cdots & \bar{\boldsymbol{H}}_{p+q-1} \end{bmatrix} \tag{40}$$

for predecessor independence [88, 89]. In view of the MISO decomposition,

$$\bar{\boldsymbol{H}}_l = [\bar{h}_{l,1}, \bar{h}_{l,2}, \ldots, \bar{h}_{l,v_i}] \tag{41}$$

Interchanging columns, equation (32) may be written as

$$\boldsymbol{H}(p,q) = [\boldsymbol{H}_1 \quad \boldsymbol{H}_2 \quad \cdots \quad \boldsymbol{H}_{v_i}] \tag{42}$$

where $\boldsymbol{H}_j$, $j = 1, 2, \ldots, v_i$ are the $p \times q$ Hankel matrices of the SISO elements of the MISO submodel. Thus, the problem of structural identification of the MISO model is also decomposed into equivalent problems of finding ranks of Hankel matrices of individual elements over a row. Singular-value decomposition may be used for this purpose.

According to the 'partial realisation theory' [90, 91], given a finite sequence of Markov parameters, it is possible to find a finite-dimensional realisation whose first few Markov parameters are correspondingly equal to the given finite sequence of Markov parameters. Accordingly, given a finite Markov-Poisson parameter sequence, irreducible TFM models can be derived solving the following equations together with (31):

$$h_{l,ij} = b_{l,ij} - \sum_{r=0}^{l-1} h_{r,ij}a_{l-r,ij}, \quad l = 1, 2, \ldots, n_{ij}$$

$$h_{l+n_{ij},ij} = -\sum_{r=1}^{n_{ij}} h_{n_{ij}+l-r,ij}a_{r,ij}, \quad l = 1, 2, \ldots$$

where the ijth element of the TFM is considered to be of the form

$$G_{ij}(s) = \frac{b_{1,ij}s^{n_{ij}-1} + \cdots + b_{n_{ij},ij}}{s^{n_{ij}} + a_{1,ij}s^{n_{ij}-1} + \cdots + a_{n_{ij},ij}}$$

Supposing the system is of this ARMA form, some insight may be given regarding the nature of the MPS:

(i) The MPS is convergent when all the poles of all the elements of the TFM are inside the unit circle centred at the origin of the s-plane. Equivalently, the Markov-Poisson parameter sequence (MPPS) is convergent when all the poles of all the elements of the TFM are inside the circle of radius β centred at $(-\lambda, 0)$ of the pole-zero plot. This circle may be termed as the zone of convergence of the sequence.
(ii) The MPS (or MPPS) is finite if and only if all the poles of all the elements of the TFM lie at the origin of the zone of convergence.

Finitisation of MPS: The usual infinite-length MPS is finite (with length $\max_{i,j}\{n_{ij}\}$ when no common denominator (CD) is assumed, or $\max_j\{n_j\}$ when column-wise CD is assumed) only when the poles of each subsystem of the TFM lie at the origin of the convergence zone. For a known system, all poles can be placed at the centre by state feedback. Then such a modified system will have a finite MPS. In the identification problem, as such state feedback cannot be introduced, because the system itself is unknown, it is possible to introduce the effect of pole-placement on the input-output measurement data by some iterative pole-placement algorithm. For the sake of simplicity, the SISO case is considered in the following.

Consider the state equation of $G^0(s)$ in its controllability canonical form:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}u$$

where

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_n & -a_{n-1} & -a_{n-1} & \cdots & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

The matrix $\boldsymbol{A}$ can be written as

$$\boldsymbol{A} = \boldsymbol{A}_0 - \boldsymbol{b}\boldsymbol{k}^{\mathrm{T}}$$

where

$$\boldsymbol{A}_0 = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{k}^{\mathrm{T}} = [a_n, a_{n-1}, \ldots, a_1]$$

Hence, we have the state equation as

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}_0\boldsymbol{x}(t) + \boldsymbol{b}\bar{u}(t)$$

where $\bar{u}(t) = u(t) - \boldsymbol{k}^{\mathrm{T}}\boldsymbol{x}(t)$ is the filtered input signal. This fictitious system described by the signal pair $\bar{u}(t)$ and $\boldsymbol{x}(t)$ has a finite MPS, as the eigenvalues of $\boldsymbol{A}_0$ are all at zeros. Hence, by transforming the original system into that described by the preceding text, the approximation error due to truncation of the MPS can be made to vanish. This is equivalent to placing the poles of the system at the origin of the convergence zone. Based on this, a time-recursive and iterative algorithm was initially proposed in [84] and extended to MIMO systems [86].

5.2 Time moment models

Like Markov parameters, time moments also play an important role in the field of reduced-order modelling. Despite the wealth of other mathematically sound methods available for reduced-order modelling, the moment matching method is still considered as the simplest and is widely used. In the field of system identification, an approach for multivariable system identification has been proposed in [85].

Estimation of moving-average models: The TFM in (29) may be written in terms of the time moments, which are related to the impulse response as

$$\boldsymbol{G}^0(s) = \int_0^{\infty} \boldsymbol{G}^0(t)e^{-st}\mathrm{d}t = \sum_{l=0}^{\infty} \boldsymbol{M}_l s^l \tag{43}$$

where

$$\boldsymbol{M}_l = \frac{(-1)^l}{l!}\int_0^{\infty} t^l \boldsymbol{G}^0(t)\mathrm{d}t, \quad l = 1, 2, \ldots \tag{44}$$

are the normal time moments of the impulse response, and define

$$\boldsymbol{H}_l = \begin{bmatrix} m_{l,11} & m_{l,12} & \cdots & m_{l,1v_i} \\ m_{l,21} & m_{l,22} & \cdots & m_{l,2v_i} \\ \vdots & \vdots & \ddots & \vdots \\ m_{l,v_o1} & m_{l,v_o2} & \cdots & m_{l,v_ov_i} \end{bmatrix}$$

In terms of the time moment sequence (TMS) $\boldsymbol{M}_l$, the system input-output relation according to (29) becomes

$$\boldsymbol{y}(t) = \sum_{l=0}^{n} \boldsymbol{M}_l \boldsymbol{u}^{(l)}(t) + \boldsymbol{n}(t) \tag{45}$$

where $\boldsymbol{u}^{(l)}(t)$ is the lth derivative of $\boldsymbol{u}(t)$. Assuming absolute convergence of TMS and thus uniform convergence of partial sums, similar to the case of MP modelling (32), the truncated TM model is

$$\boldsymbol{y}(t) = \sum_{l=0}^{n} \boldsymbol{M}_l \boldsymbol{u}^{(l)}(t) + \boldsymbol{e}(t) \tag{46}$$

To validate the use of this model, even for systems with diverging TMS, additional exponential scaling of the series will be necessary to ensure convergence.

To avoid the direct use of derivatives, (46) is operated on both sides by a $(n+1)$th-order Poisson filter operator $\beta^{n+1}/(s+\lambda)^{n+1}$ [9]. Denoting

$$F_{l,n+1}(s) = \beta^{n+1}\frac{s^l}{(s+\lambda)^{n+1}}, \quad l = 0, 1, \ldots, n$$

the time moment (TM) model is

$$F_{0,n+1}(s)\boldsymbol{Y}(s) = \sum_{l=0}^{n} \boldsymbol{M}_l F_{l,n+1}(s)\boldsymbol{U}(s) + \boldsymbol{E}(s) \tag{47}$$

For the ith row of (34) (dropping the subscript i in all relevant symbols), taking into account n_j time moments of the jth MISO subsystem, and letting $n = \max_j\{n_j\}$, the parameter estimation equation in discrete-time is obtained as

$$\tilde{y}^*(k) = \boldsymbol{\varphi}^{\mathrm{T}}(k)\boldsymbol{\theta} \tag{48}$$

where

$$\begin{aligned} \boldsymbol{\varphi}(k) &= [\boldsymbol{\varphi}_1^{\mathrm{T}}(k) \quad \boldsymbol{\varphi}_2^{\mathrm{T}}(k) \quad \cdots \quad \boldsymbol{\varphi}_{v_i}^{\theta}(k)]^{\mathrm{T}}, \qquad (49) \\ \boldsymbol{\varphi}_j^{\mathrm{T}}(k) &= [F_{0,n+1}(\delta)u_j(k) \quad F_{1,n+1}(\delta)u_j(k) \\ &\qquad \cdots \quad F_{n_j,n+1}(\delta)u_j(k)], \\ &\qquad j = 1, \ldots, v_i \end{aligned}$$

and

$$\boldsymbol{\theta} = [m_{0,1}, \ldots, m_{n_1,1}|\cdots|m_{0,v_i}, \ldots, m_{n_{v_i},v_i}]^{\mathrm{T}}$$

Parameter estimation may now be carried out with the usual least-squares algorithm.

Irreducible ARMA model realisation: Given the estimates of $\boldsymbol{M}_l$, $l = 1, \ldots, n$, an irreducible ARMA TFM model can be realised in a manner similar to the case of Markov parameter models. Let

$$\begin{aligned} A_{ij}(s) &= 1 + a_{1,ij}s + \cdots + a_{n_{ij},ij}s^{n_{ij}}, \\ B_{ij}(s) &= b_{0,ij} + b_{1,ij}s + \cdots + b_{n_{ij}-1,ij}s^{n_{ij}-1} \end{aligned}$$

and

$$\boldsymbol{M}_l = \{m_{l,ij};\ i = 1, \ldots, v_o,\ j = 1, \ldots, v_i\}$$

Given the estimates of $\boldsymbol{\theta}_l$, $l = 1, \ldots, n$, the TFM elements can be obtained by solving the following equations:

$$m_{l,ij} = b_{l,ij} - \sum_{r=0}^{l-1} m_{r,ij}a_{l-r,ij}, \quad l = 0, \ldots, n_{ij} - 1$$

and

$$m_{l+n_{ij},ij} = -\sum_{r=1}^{n_{ij}} m_{n_{ij}+1-r,ij} a_{l,ij}, \quad l = 1, 2, \ldots$$

Supposing the system is of this ARMA form, the following remarks are in order:

(i) The TMS is convergent if all the poles of all the elements of the TFM are outside the unit circle centred at the origin of the s-plane. This circle is the zone of the convergence of the sequence.
(ii) The TMS is finite if and only if all the elements of the TFM are denominator free (i.e. have denominator as 1).

For this special case, an iterative algorithm was proposed [85] that finitises the sequence so as to eliminate the truncation error.

Finitisation of TMS: The TMS is finite when all the subsystems of the TFM are denominator free, in which case the length of the TMS is $\max_j\{n_j\}$ and modelling will not involve unmodelled dynamics. This situation can be met by adding fictitious zeros to each subsystem, to cancel their respective denominators. In an identification experiment, this is achievable for ARMA systems as illustrated for the SISO case as follows:

$$Y(s) = \frac{B(s)}{A(s)} U(s)$$

If the denominator $A(s)$ is known, we can write

$$Y(s) = \frac{B(s)}{A(s)} U(s) = \sum_{i=1}^{n-1} b_i s^i \bar{U}(s)$$

where

$$\bar{U}(s) = \frac{1}{A(s)} U(s)$$

Thus the model between $\bar{U}(s)$ and $Y(s)$ has a finite TMS and, therefore, by estimating the denominators and then cancelling them, in an iterative way, it is possible to finitise the TMS, so that the truncation error is removed iteratively. Such an iterative algorithm with detailed analysis was presented in [85].

Choice of parametric form: In reality, modelling error is inevitable and the performance (namely, predictive ability) of the estimated models depends on the choice of model structure and the prior knowledge embedded into the chosen model structures, for a given model order. In the generalised moving average model (GMAM) structure, the following variants are considered for CT system modelling:

(i) Motivated by Markov-Poisson parameter models, a Poisson filter chain:

$$\boldsymbol{b}_{\mathrm{PF}}(s) = \left[\frac{\beta}{s+\lambda} \quad \left(\frac{\beta}{s+\lambda}\right)^2 \quad \cdots \quad \left(\frac{\beta}{s+\lambda}\right)^n \right]^{\mathrm{T}} \tag{50}$$

(ii) Motivated by TM models, with a state-variable filter (SVF),

$$\boldsymbol{b}_{\mathrm{SVF}}(s) = \left[\frac{1}{E(s)} \quad \frac{s}{E(s)} \quad \cdots \quad \frac{s^{n-1}}{E(s)} \right]^{\mathrm{T}} \tag{51}$$

where $1/E(s)$ is an nth order stable filter may be considered to filter the input-output data. A typical choice is an nth order Poisson filter.

The following issues are now studied via numerical examples:

(*a*) *Predictive ability*: The two choices $\boldsymbol{b}_{\mathrm{PF}}(\delta)$ and $\boldsymbol{b}_{\mathrm{SVF}}(\delta)$ are related through a linear nonsingular transformation (for $\lambda \neq 0$), e.g. for $n = 4$ and $\beta = 1$,

$$\boldsymbol{b}_{\mathrm{SVF}}(s) = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -\lambda \\ 0 & 2 & -2\lambda & \lambda^2 \\ 1 & -3\lambda & 3\lambda^2 & \lambda^3 \end{bmatrix} \boldsymbol{b}_{\mathrm{PF}}(s) \tag{52}$$

Hence, for a given model order, models based on these two sets will have the same predictive ability.
(*b*) *The numerical behaviour* of the estimation algorithm is dictated in the LS algorithm by the condition number of the matrix

$$\boldsymbol{R} = \sum_{k=1}^{N} \boldsymbol{\varphi}(k)\boldsymbol{\varphi}^{\mathrm{T}}(k) \tag{53}$$

It has been shown in [84] that use of $\boldsymbol{b}_{\mathrm{PF}}(\delta)$ results in high condition numbers of this matrix (53), as these functions are overlapping and nonorthogonal. On the other hand, the second set $\boldsymbol{b}_{\mathrm{SVF}}(\delta)$ is near-orthogonal [81], which improves the condition number.
(*c*) *Numerical conditioning* may be improved if an intelligently chosen linear transformation of these sets of basis functions is made before parameter estimation commences. When such transformation results in an orthogonal set, the numerical properties of the algorithm will be significantly improved. A popular orthogonal basis is in terms of Laguerre filters:

$$\boldsymbol{b}_{\mathrm{LAG}}(s) = \left[\frac{1}{s+\lambda} \quad \frac{1}{s+\lambda}\left(\frac{s-\lambda}{s+\lambda}\right) \quad \cdots \quad \frac{1}{s+\lambda}\left(\frac{s-\lambda}{s+\lambda}\right)^{n-1} \right]^{\mathrm{T}} \tag{54}$$

The required linear transformations are for ($n = 4$ and $\beta = 1$)

$$\boldsymbol{b}_{\mathrm{LAG}}(s) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & -2\lambda & 0 & 0 \\ 1 & -4\lambda & 4\lambda^2 & 0 \\ 1 & -6\lambda & 12\lambda^2 & -8\lambda^3 \end{bmatrix} \boldsymbol{b}_{\mathrm{PF}}(s) \tag{55}$$

and

$$\boldsymbol{b}_{\mathrm{LAG}}(s) = \begin{bmatrix} 1 & 3\lambda^2 & 3\lambda & 1 \\ -\lambda^3 & -\lambda^2 & \lambda & 1 \\ \lambda^3 & -\lambda^2 & \lambda & 0 \\ 1 & 3\lambda^2 & 3\lambda & 1 \end{bmatrix} \boldsymbol{b}_{\mathrm{SVF}}(s) \tag{56}$$

Several of the time-domain methods of identification are provided in the recently developed continuous-time system identification (CONTSID) Matlab toolbox [92], which has been used to compare all implemented methods in a simulation context and on real data [93].

5.3 Linear filters approach [10]

5.3.1 Basic idea: The earliest use of linear filters in an attempt to avert the derivative measurement problem was in [94]. The method of multiple filters (MMF) using sets of first-order elements in a filter complex [95–98] and the PMF method are two different manifestations of the same mathematical operation to avoid direct time derivatives of process signals in identification of continuous-time models. The PMF formulation, however, made a systematic methodology possible with an extension to distributed parameter systems. This approach is appropriate for the generation of the filtered time derivatives required for the identification of continuous-time models and is known also as state-variable-filter (SVF) method [4]. Similar to the PMF approach in which the input and output signals $u(t)$ and $y(t)$, respectively, of the system to be identified are processed through two identical pre-filter chains, each element of which has transfer function of the form $\beta/(s+\lambda)$, we consider filter banks each consisting of $q = n + m + 1$ lowpass filters described by the transfer functions $1/H_i(s)$, $H_i(s)$ with $i = 1, 2, \ldots, q$ being any Hurwitz polynomials of degree at least equal to n, the order of the system to be identified, and is the degree of its numerator. The coefficients of $H_i(s)$ should be selected such that the dynamics of $1/H_i(s)$ are approximately at least ten times faster than the dynamics of the system to be identified. The linear filter set-up is arranged as in Fig. 9, where the coefficients a_μ and b_ν of the rational system transfer function

$$G(s) = \frac{\sum_{\nu=0}^{m} b_\nu s^\nu}{\sum_{\mu=0}^{n} a_\mu s^\mu} = \frac{B(s)}{A(s)}, \quad a_0 = 1 \tag{57}$$

are to be identified. The filtered input and output signals are given in the frequency domain by

$$U_i^*(s) = H_i^{-1}(s)U(s) \quad \text{and} \quad Y_i^*(s) = H_i^{-1}(s)\frac{B(s)}{A(s)}U(s)$$

and, from these two conditions, it follows directly that

$$U_i^*(s)B(s) = Y_i^*(s)A(s) \tag{58}$$

The inverse Laplace-transformation of (58) provides the ODE

$$\sum_{\nu=0}^{m} b_\nu \frac{d^\nu u_i^*(t)}{dt^\nu} = \sum_{\mu=1}^{n} a_\mu \frac{d^\mu y_i^*(t)}{dt^\mu} + y_i^*(t) \tag{59}$$

If the filtered signals $u_i^*(t)$ and $y_i^*(t)$ as well as their derivatives are known (measurable), then (59) represents an algebraic equation with the unknown system parameters a_μ and b_ν. From the specific structure of the linear filters, the signals $u_i^*(t)$, $y_i^*(t)$ and their derivatives can be obtained directly. For q unknown system parameters, q different filter pairs $1/H_i(s)$ are necessary to obtain the following

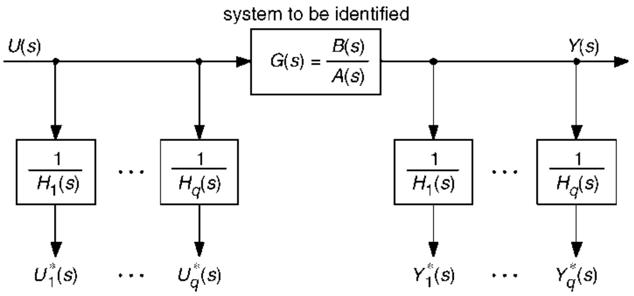


Fig. 9 *Model of the actual system*

equations:

$$b_m \frac{d^m u_1^*}{dt^m} + \cdots + b_0 u_1^* = a_n \frac{d^n y_1^*}{dt^n} + \cdots + a_1 \frac{dy_1^*}{dt} + y_1^*$$

$$\vdots$$

$$b_m \frac{d^m u_q^*}{dt^m} + \cdots + b_0 u_q^* = a_n \frac{d^n y_q^*}{dt^n} + \cdots + a_1 \frac{dy_q^*}{dt} + y_q^*$$

or, in matrix notation,

$$\boldsymbol{y}^*(t) = \boldsymbol{M}(t)\boldsymbol{\theta} \tag{60}$$

with the $(q \times q)$ data (or measurement) matrix

$$\boldsymbol{M}(t) = \left[\begin{array}{ccc|cccc} -\frac{dy_1^*}{dt} & \cdots & -\frac{d^n y_1^*}{dt^n} & u_1^* & \frac{du_1^*}{dt} & \cdots & \frac{d^m u_1^*}{dt} \\ & \vdots & & \vdots & \vdots & & \vdots \\ -\frac{dy_q^*}{dt} & \cdots & -\frac{d^n y_q^*}{dt^n} & u_q^* & \frac{du_q^*}{dt} & \cdots & \frac{d^m u_q^*}{dt} \end{array}\right] \tag{61a}$$

the parameter vector

$$\boldsymbol{\theta} = \left[\begin{array}{cccc|ccc} a_1 & a_2 & \cdots & a_n & b_0 & \cdots & b_m \end{array}\right]^T \tag{61b}$$

and the output data (or measurement) vector of the filtered signals

$$\boldsymbol{y}^*(t) = [y_1^*(t) \quad y_2^*(t) \quad \cdots \quad y_q^*(t)]^T$$

From (60) follows the solution for the parameter vector to be identified as

$$\boldsymbol{\theta} = \boldsymbol{M}^{-1}(t)\,\boldsymbol{y}^*(t) \tag{62}$$

5.3.2 Reduction of the number of filter pairs: Only one filter pair is necessary, if at q different time instants $t_k = kT_S$, with T_S being the sampling period, q pairs of filtered data $\{u^*(t_k), y^*(t_k)\}$ and their derivatives are used for the data matrix

$$\boldsymbol{M}(t_k) = \left[\begin{array}{ccc|ccc} -\left.\frac{dy^*}{dt}\right|_{t=k-q+1} & \cdots & -\left.\frac{d^n y^*}{dt^n}\right|_{t=k-q+1} & u^*|_{t=k-q+1} & \cdots & \left.\frac{d^m u^*}{dt^m}\right|_{t=k-q+1} \\ \vdots & & \vdots & \vdots & & \vdots \\ \left.\frac{dy^*}{dt}\right|_{t=k} & \cdots & -\left.\frac{d^n y^*}{dt^n}\right|_{t=k} & u^*|_{t=k} & \cdots & \left.\frac{d^m u^*}{dt}\right|_{t=k} \end{array}\right] \tag{63a}$$

and the output data vector of filtered signals

$$\boldsymbol{y}^*(t_k) = [y^*(t_{k-q+1}) \quad \cdots \quad y^*(t_k)]^T \tag{63b}$$

Then (60) becomes

$$\boldsymbol{y}^*(t_k) = \boldsymbol{M}(t_k)\boldsymbol{\theta} \tag{64}$$

and the solution for the parameter vector is given by

$$\boldsymbol{\theta} = \boldsymbol{\theta}(t_k) = \boldsymbol{M}^{-1}(t_k)\boldsymbol{y}^*(t_k) \tag{65}$$

i.e. for $t_k \geq qT_S$ the parameter vector $\boldsymbol{\theta}$ can be computed at each sampling time. The approximation will be good for small $T_S < \pi/(5\omega_n)$, where ω_n is the limit frequency of the corresponding signal.

5.3.3 Solution based on LS approach: The case of a stochastically disturbed process output signal is considered. This leads in a statistical sense to a prediction error between the filtered output signal $y^*(t_k)$ and its predicted estimate of the form

$$\varepsilon(t_k) = y^*(t_k) - \boldsymbol{m}^{\mathrm{T}}(t_k)\hat{\boldsymbol{\theta}}(t_{k-1}) \tag{66}$$

where $\boldsymbol{m}^{\mathrm{T}}(t_k)$ is given by the last line vector of (63a). For q different time instants ($k = 1, \ldots, q$), instead of (66) the associated vector form follows

$$\boldsymbol{\varepsilon}(t_k) = \boldsymbol{y}^*(t_k) - \boldsymbol{M}^{\mathrm{T}}(t_k)\hat{\boldsymbol{\theta}}(t_{k-1}) \tag{67}$$

where $\boldsymbol{y}^*(t_k)$, $\boldsymbol{M}^{\mathrm{T}}(t_k)$ and $\hat{\boldsymbol{\theta}}(t_{k-1})$ come from (63b), (63a) and (61b), respectively, and

$$\boldsymbol{\varepsilon}(t_k) = [\varepsilon(t_{k-q+1}) \quad \cdots \quad \varepsilon(t_k)]^{\mathrm{T}} \tag{68}$$

As (67) is linear in the parameter vector $\hat{\boldsymbol{\theta}}(t_k)$, the least-squares (LS) criterion

$$J[\hat{\boldsymbol{\theta}}(t_k)] = \min\left\{\frac{1}{2}\ \boldsymbol{\varepsilon}^{\mathrm{T}}(t_k)\boldsymbol{\varepsilon}(t_k)\right\} \tag{69}$$

provides either the well-known direct solution

$$(\hat{\theta}_{k-1}) = [\boldsymbol{M}^{\mathrm{T}}(t_k)\boldsymbol{M}(t_k)]^{-1}\boldsymbol{M}^{\mathrm{T}}(t_k)\boldsymbol{y}^*(t_k) \tag{70}$$

or the recursive solution

$$\hat{\boldsymbol{\theta}}(t_k) = \hat{\boldsymbol{\theta}}(t_{k-1}) + \boldsymbol{q}(t_k)[y^*(t_k) - \boldsymbol{m}^{\mathrm{T}}(t_k)\hat{\boldsymbol{\theta}}(t_{k-1})] \tag{71a}$$

$$\boldsymbol{q}(t_k) = \boldsymbol{P}(t_{k-1})\boldsymbol{m}(t_k)[1 + \boldsymbol{m}^{\mathrm{T}}(t_k)\boldsymbol{P}(t_{k-1})\boldsymbol{m}(t_k)]^{-1} \tag{71b}$$

$$\boldsymbol{P}(t_k) = \boldsymbol{P}(t_{k-1}) - \boldsymbol{q}(t_k)\boldsymbol{m}^{\mathrm{T}}(t_k)\boldsymbol{P}(t_{k-1}) \tag{71c}$$

where, in addition, initial values for $\hat{\boldsymbol{\theta}}$ and its covariance matrix $\boldsymbol{P}$ have to be selected, which, however, can be done relatively arbitrarily.

It should be mentioned that the principle of the linear-filter approach or state-variable-filter (SVF) method has been applied very successfully in recursive form for the identification of continuous-time models and self-tuning controllers during many years, not only in combination with the LS method but also with other approaches [99–105], such as, for example, stochastic approximation (SA), refined-instrumental-variable (RIV) method, or iterative maximum-likelihood (ML) method. Details of parameter choice, digital realisation and convergence conditions are discussed in [100]. The basic idea of linear filters is also included in those approaches where the filters had been replaced by an algebraic reformulation of transfer function models [105]. Among the methods using linear filters, which are also referred to as 'pre-filters' in linear CT model identification, the SRIVC [102] is optimal in the statistical sense that it iteratively optimises the filter parameters and yields estimates that are consistent and asymptotically efficient (minimum variance). It also provides an estimate of the covariance matrix associated with the parameter estimates, thereby providing standard error bounds on the parameters. An SRIVC algorithm is available in both the CAPTAIN and CONTSID Toolboxes for Matlab, both of which can be downloaded from the web: (http://www.es.lancs.ac.uk/cres/captain/) and (http://www.cran.uhpnancy.fr/contsid/). In Section 10 we will see the results of application of certain methods in the latter Toolbox, in a comparative investigation in which the SRVIC method shows the best in performance.

6 System identification by SRAM approach

A classical technique for identification of linear time-invariant as well as time-varying CT systems is given by the system-reference-adaptive-model (SRAM) approach. The basic idea of this approach will be shown by the following simple example:

6.1 Introductory example

Let us assume the gain factor K of a real system (plant) with pure proportional behaviour to be unknown. In the undisturbed case the identification can be performed easily by arranging an adaptive model in parallel with the real system and applying the same input $u(t)$. The output $y(t)$ of the real system and the model $y_M(t)$ are compared such that the model error

$$e^*(t) = y(t) - y_M(t) = Ku(t) - \hat{K}(t)u(t) \tag{72}$$

is obtained, where $\hat{K}(t)$ is the gain factor of the model which has to be adapted through some criterion such that $\lim_{t\to\infty} \hat{K}(t) = K$. Here K is assumed to be constant, but in general it also can be a time-varying parameter. Selecting the minimisation criterion

$$J[\hat{K}(t)] = \frac{1}{2}e^{*2}(t) = \frac{1}{2}[K - \hat{K}(t)]^2u^2(t) \stackrel{!}{=} \min \tag{73}$$

and applying the gradient method approach [106]

$$\frac{\mathrm{d}\hat{K}}{\mathrm{d}t} = -\alpha\frac{\partial J}{\partial \hat{K}} = \alpha e^*(t,\ \hat{K})u(t) \tag{74}$$

finally yields the adaptation law

$$\hat{K}(t) = \hat{K}(0) + \alpha\int_0^t e^*(\tau, \hat{K})u(\tau)\mathrm{d}\tau \tag{75}$$

Substituting (72) into (74) gives

$$\frac{\mathrm{d}\hat{K}}{\mathrm{d}t} = -\alpha[\hat{K}(t) - K]u^2(t) \tag{76}$$

and introducing the parameter error $\tilde{K}(t) = \hat{K}(t) - K$, with $K = $ constant, leads to

$$\frac{\mathrm{d}\tilde{K}}{\mathrm{d}t} = -\alpha\tilde{K}(t)u^2(t) \tag{77}$$

The solution of the nonlinear DE (77) is given by

$$\tilde{K}(t) = \tilde{K}(0)\mathrm{e}^{-\alpha\int_0^t u^2(\tau)\mathrm{d}\tau} \tag{78a}$$

or

$$\hat{K}(t) = K + [\hat{K}(0) - K]\mathrm{e}^{-\alpha\int_0^t u^2(\tau)\mathrm{d}\tau} \tag{78b}$$

The error $\tilde{K}(t)$ decreases, but convergence of the identified parameter $\hat{K}(t)$ to the real parameter K is only achieved if

$$\lim_{t\to\infty} \tilde{K}(t) = 0 \quad \text{or} \quad \lim_{t\to\infty}\int_0^t u^2(\tau)\mathrm{d}\tau \to \infty \tag{79}$$

To guarantee convergence according to (79) the persistency of excitation condition

$$\int_t^{t+\Delta t} u^2(\tau)\mathrm{d}\tau > 0 \tag{80}$$

must be fulfilled in each time interval $t \leq \tau \leq t + \Delta t$.

The stability of the solution (78*a*) can be investigated by introducing the following Lyapunov function and its time derivative, respectively,

$$V[\tilde{K}(t)] = V[\tilde{K}(0)] - \alpha \int_0^t [e^*(\tau, \tilde{K})]^2 \mathrm{d}\tau \tag{81}$$

$$\frac{\mathrm{d}V(\tilde{K})}{\mathrm{d}t} = -\alpha e^{*2}(\tau, \tilde{K}) \tag{82}$$

with the conditions (i) V positive-definite and (ii) $\dot{V}$ negative-definite. Both of these conditions are fulfilled for

$$(a) \qquad V[\tilde{K}(0)] > \alpha \int_0^t e^{*2}(\tau, \tilde{K})\mathrm{d}\tau \Rightarrow \int_0^t e^{*2}\mathrm{d}\tau < \infty$$
$$\Rightarrow e^* \text{ is bounded}$$

(*b*) $V[\tilde{K}(t) = 0] = 0$,
(*c*) $\dot{V}[\tilde{K}(t)] < 0,\ \tilde{K} \neq 0$,
(*d*) $\dot{V}[\tilde{K}(t) = 0] = 0$

Thus asymptotic stability is also guaranteed under persistent excitation.

Let us now consider the 'case of an additional disturbance' $n(t)$ acting onto the undisturbed system output signal $y_d(t)$. In this case, the error becomes

$$e^*(t) = [y(t) + n(t)] - y_M(t) \tag{83}$$

and from (74) and (83) it follows that

$$\frac{\mathrm{d}\tilde{K}}{\mathrm{d}t} = -\alpha \tilde{K}(t)u^2(t) + \alpha n(t)u(t) \tag{84}$$

If the undisturbed case is asymptotically stable, i.e. $\lim_{t\to\infty} \tilde{K}(t) = 0$, then a bounded disturbance $n(t)$ generates a bounded $\tilde{K}(t)$, because, even for a sufficiently large S/N ratio, $\tilde{K}(t)$ can only converge near to zero. For example let $u(t) = 1$, and, for the moment, $n(t) = 0$. Then a large value of α provides fast convergence for the solution

$$\tilde{K}(t) = \tilde{K}(0)\mathrm{e}^{-\alpha t} \tag{85}$$

as seen from (78*a*). However, for $n(t) \neq 0$, the following nonhomogeneous ODE is obtained from (84):

$$\frac{\mathrm{d}\tilde{K}}{\mathrm{d}t} = -\alpha \tilde{K}(t) + \alpha n(t) \tag{86}$$

It is easy to see from the corresponding solution

$$\tilde{K}(t) = \tilde{K}(0)\mathrm{e}^{-\alpha t} + \alpha \int_0^t \mathrm{e}^{-\alpha(t-\tau)} n(\tau)\mathrm{d}\tau \tag{87}$$

that the smaller α the less effect has $n(t)$. This leads to the contradiction that a large value of α provides fast convergence but low noise immunity, thus a compromise must be made.

Finally, we consider the 'time-varying case' with the gain factor $K = K(t) \neq$ constant. Here, it follows from (77) that

$$\frac{\mathrm{d}\tilde{K}}{\mathrm{d}t} = -\alpha \tilde{K}(t)u^2(t) - \frac{\mathrm{d}K}{\mathrm{d}t} \tag{88}$$

For the special case of $u(t) = 1$ and a drifting gain factor, where $\mathrm{d}K/\mathrm{d}t = \beta =$ constant, it follows from (88) that

$$\frac{\mathrm{d}\tilde{K}}{\mathrm{d}t} = -\alpha \tilde{K}(t) - \beta$$

with the solution

$$\tilde{K}(t) = \left[\tilde{K}(0) + \frac{\beta}{\alpha}\right] \mathrm{e}^{-\alpha t} - \frac{\beta}{\alpha} \tag{89}$$

From (89) it can be seen that better tracking of the time-varying gain factor $K(t)$ can be achieved by a large value of α. After this simple introductory example, let us now consider the general case of parameter estimation by the SRAM approach.

6.2 Parallel SRAM approach via gradient methods

In this Section the slowly time-varying parameters of a linear plant, described by the transfer function $G(s)$ will be identified through adaptation of the parameter vector θ_M of a corresponding parallel model given by the transfer function

$$G_\mathrm{M}(s, \boldsymbol{\theta}_\mathrm{M}) = \frac{Y_\mathrm{M}(s, \boldsymbol{\theta}_\mathrm{M})}{U(s)} = \frac{\sum_{j=0}^m b_{\mathrm{M}j} s^j}{\sum_{i=0}^n a_{\mathrm{M}i} s^i} = \frac{B_\mathrm{M}(s, \boldsymbol{b}_\mathrm{M})}{A_\mathrm{M}(s, \boldsymbol{a}_\mathrm{M})} \tag{90}$$

where

$$\boldsymbol{\theta}_\mathrm{M} = [\, a_{\mathrm{M}0} \ \cdots \ a_{\mathrm{M}n} \ \vdots \ b_{\mathrm{M}0} \ \cdots \ b_{\mathrm{M}m} \,]^\mathrm{T}$$
$$= [\, \boldsymbol{a}_\mathrm{M}^\mathrm{T} \ \vdots \ \boldsymbol{b}_\mathrm{M}^\mathrm{T} \,]^\mathrm{T} \tag{91}$$

usually will be normalised by setting $a_{\mathrm{M}n} = 1$. With the error criterion

$$J(\boldsymbol{\theta}_\mathrm{M}) = \frac{1}{2} e^{*2}(t, \boldsymbol{\theta}_\mathrm{M}) = \frac{1}{2}[y(t) - y_\mathrm{M}(t, \boldsymbol{\theta}_\mathrm{M})]^2 \stackrel{!}{=} \min \tag{92}$$

and application of the gradient method, as defined in (75), follows the identified model parameter vector

$$\boldsymbol{\theta}_\mathrm{M}(t) = \boldsymbol{\theta}_\mathrm{M}(0) + \alpha \int_0^t e^*(\tau, \boldsymbol{\theta}_\mathrm{M}) \frac{\partial}{\partial \boldsymbol{\theta}_\mathrm{M}} y_\mathrm{M}(\tau, \boldsymbol{\theta}_\mathrm{M})\mathrm{d}\tau \tag{93}$$

Defining the vector of sensitivity functions

$$\boldsymbol{v}(t, \boldsymbol{\theta}_\mathrm{M}) = \frac{\partial}{\partial \boldsymbol{\theta}_\mathrm{M}} y_\mathrm{M}(t, \boldsymbol{\theta}_\mathrm{M}) \tag{94}$$

we obtain, for slowly varying parameters in the frequency domain,

$$\frac{\partial}{\partial \boldsymbol{\theta}_\mathrm{M}} Y_\mathrm{M}(s, \boldsymbol{\theta}_\mathrm{M}) = \frac{\partial}{\partial \boldsymbol{\theta}_\mathrm{M}} G_\mathrm{M}(s, \boldsymbol{\theta}_\mathrm{M}) U(s) \tag{95}$$

with the vector of sensitivity filter transfer functions

$$\frac{\partial}{\partial \boldsymbol{\theta}_\mathrm{M}} G_\mathrm{M}(s, \boldsymbol{\theta}_\mathrm{M}) = \left[\frac{\partial G_\mathrm{M}}{\partial \theta_{\mathrm{M}1}} \ \cdots \ \frac{\partial G_\mathrm{M}}{\partial \theta_{\mathrm{M}q}}\right]^\mathrm{T},$$
$$q = n + m + 1 \tag{96}$$

Substituting (90) into (96) finally provides the separated sensitivity transfer functions

$$\frac{\partial}{\partial \boldsymbol{a}_\mathrm{M}} G_\mathrm{M}(s, \boldsymbol{\theta}_\mathrm{M}) = -G_\mathrm{M}(s, \boldsymbol{\theta}_\mathrm{M}) \frac{1}{A_\mathrm{M}(s, \boldsymbol{a}_\mathrm{M})} \boldsymbol{d}_{n-1}(s) \tag{97}$$

and

$$\frac{\partial}{\partial \boldsymbol{b}_{\mathrm{M}}} G_{\mathrm{M}}(s, \boldsymbol{\theta}_{\mathrm{M}}) = \frac{1}{A_{\mathrm{M}}(s, \boldsymbol{a}_{\mathrm{M}})} \boldsymbol{d}_m(s) \tag{98}$$

respectively, where

$$\boldsymbol{d}_\nu = [1 \quad s \quad s^2 \quad \cdots \quad s^\nu]^{\mathrm{T}}, \qquad \nu \in (m, n-1) \text{ and } a_{\mathrm{M}n} = 1 \tag{99}$$

The filters according to (97) and (98) allow us to generate the sensitivity functions needed for the final realisation of the adaptation law (93). The block diagram of this approach is given in Fig. 10.

The following simple example shows the application of this identification method. Let us assume the model transfer function

$$G_{\mathrm{M}}(s, \boldsymbol{\theta}_{\mathrm{M}}) = \frac{b_{\mathrm{M}0} + b_{\mathrm{M}1} s}{a_{\mathrm{M}0} + s} = \frac{B_{\mathrm{M}}(s, \boldsymbol{b}_{\mathrm{M}})}{A_{\mathrm{M}}(s, \boldsymbol{a}_{\mathrm{M}})}$$

where $\boldsymbol{\theta}_{\mathrm{M}} = [a_{\mathrm{M}0} \quad b_{\mathrm{M}0} \quad b_{\mathrm{M}1}]^{\mathrm{T}}$. From (97) and (98), the sensitivity functions are as follows:

$$\frac{\partial G_{\mathrm{M}}}{\partial a_{\mathrm{M}0}} = -G_{\mathrm{M}}(s, \boldsymbol{\theta}_{\mathrm{M}}) \frac{1}{A_{\mathrm{M}}(s, a_{\mathrm{M}0})} = -\frac{b_{\mathrm{M}0} + b_{\mathrm{M}1} s}{(a_{\mathrm{M}0} + s)^2}$$

$$\frac{\partial G_{\mathrm{M}}}{\partial b_{\mathrm{M}0}} = \frac{1}{A_{\mathrm{M}}(s, a_{\mathrm{M}0})} = \frac{1}{a_{\mathrm{M}0} + s}$$

$$\frac{\partial G_{\mathrm{M}}}{\partial b_{\mathrm{M}1}} = \frac{s}{A_{\mathrm{M}}(s, a_{\mathrm{M}0})} = \frac{s}{a_{\mathrm{M}0} + s}$$

Using these sensitivity functions, the adaptation law (93) can be easily realised.

It should be mentioned that the parallel SRAM approach can be simplified considerably by applying either a series (reciprocal) model or a series-parallel model approach [106]. The investigation of stability for the general parallel SRAM approach using the gradient method would be similar to that described in Section 6.1. However, it is easy to understand that the convergence and stability analysis, and especially the selection of an appropriate Lyapunov function, will become more difficult if the structure of the system to be identified is more complicated than in the cases discussed in Sections 6.1 and 6.2. Hence, it seems reasonable to introduce model-based identification approaches which directly rely on stability theory, as discussed in the following Section.

6.3 Realising the SRAM approach by Lyapunov design

The starting point for the derivation of the adaptation law in this method is not a definite cost function, but an error differential equation of the overall system for identification. The adaptation law is to be designed so that the overall system attains a globally asymptotically stable steady state. The application of the direct method of Lyapunov will at first be demonstrated with the aid of a first-order system.

6.3.1 Introductory example [10]: The system to be identified is given by

$$\dot{y}(t) + a y(t) = K u(t) \tag{100}$$

with the unknown and eventually slowly time-varying parameters a and K. A parallel model is selected by

$$\dot{y}_{\mathrm{M}}(t) + a_{\mathrm{M}}(t) y_{\mathrm{M}}(t) = K_{\mathrm{M}}(t) u(t) \tag{101}$$

With the model error $e^*(t) = y(t) - y_{\mathrm{M}}(t)$, the error differential equation may be directly written as

$$\dot{e}^*(t) + a e^*(t) = [K - K_{\mathrm{M}}(t)] u(t) - [a - a_{\mathrm{M}}(t)] y_{\mathrm{M}}(t) \tag{102}$$

which for $\tilde{K}(t) = K - K_{\mathrm{M}}(t) = 0$, $\tilde{a}(t) = a - a_{\mathrm{M}}(t) = 0$ and $a > 0$ has the trajectory $\lim_{t \to \infty} e^*(t) = 0$ as a stable equilibrium position. $\tilde{K}$ and $\tilde{a}$ are the parameter errors, which should vanish fully in the adapted state. The adjustment must now be designed so that the above trajectory

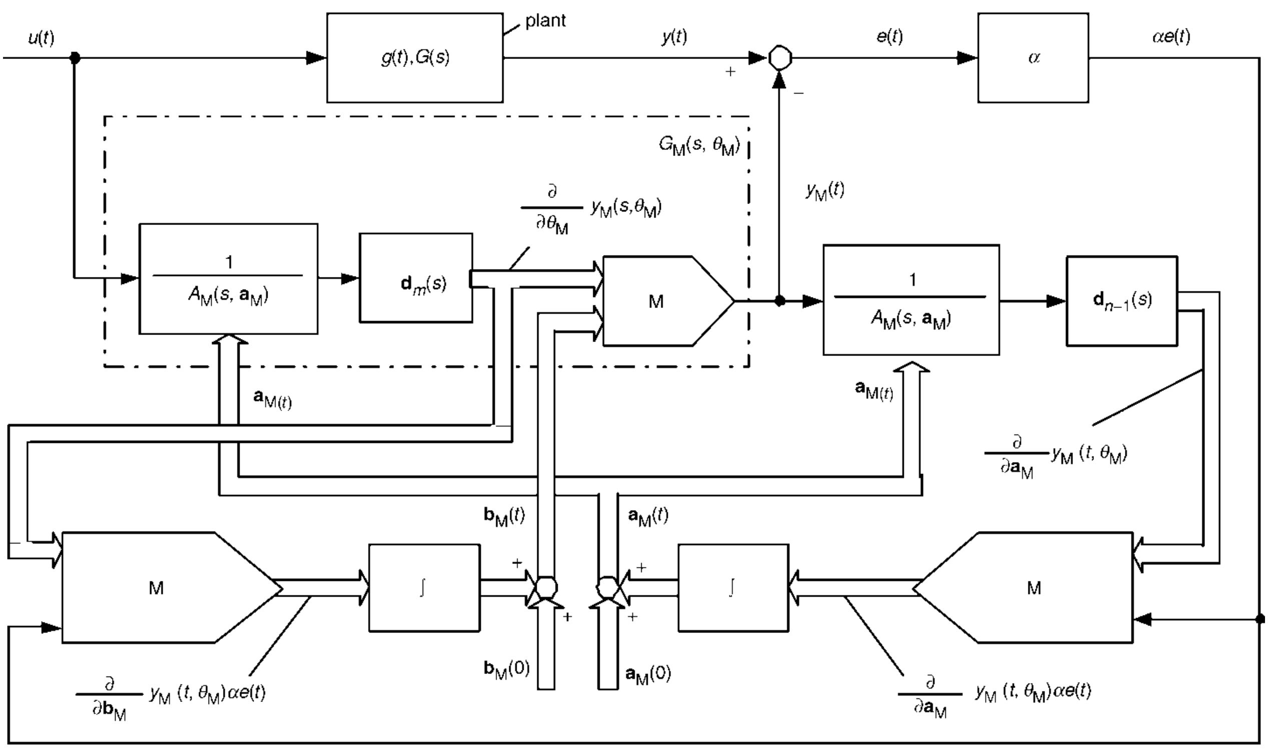


Fig. 10 *General realisation of parallel SRAM approach by sensitivity filters for SISO system*

$(e^*(t)=0,\ \tilde{K}=0,\ \tilde{a}=0)$ is globally asymptotically stable in the steady state. A possible Lyapunov function, see e.g. [107], has the quadratic form

$$V(e^*, \tilde{K}, \tilde{a}) = \frac{1}{2}e^{*2}(t) + \frac{1}{2\alpha}\tilde{K}^2(t) + \frac{1}{2\beta}\tilde{a}^2(t) \qquad (103)$$

whose time derivative after inserting $\dot{e}^*(t)$ from (102) is given by

$$\dot{V}(e^*, \tilde{K}, \tilde{a}, t) = -ae^{*2}(t) + \tilde{K}(t)\left[u(t)e^*(t) + \frac{1}{\alpha}\dot{\tilde{K}}(t)\right] - \tilde{a}(t)\left[y_M(t)e^*(t) - \frac{1}{\beta}\dot{\tilde{a}}(t)\right] \qquad (104)$$

The first term on the RHS of (104) is negative-definite for a stable dynamical system. $\dot{V}$ is then certainly negative-definite, if

$$u(t)e^*(t) + \frac{1}{\alpha}\dot{\tilde{K}}(t) = 0 \qquad (105)$$

and

$$y_M(t)e^*(t) - \frac{1}{\beta}\dot{\tilde{a}}(t) = 0 \qquad (106)$$

With $\dot{\tilde{K}} = -\dot{K}_M$ and $\dot{\tilde{a}} = -\dot{a}_M$, (105) and (106) through integration directly provide the adaptation relations

$$K_M(t) = K_M(0) + \alpha\int_0^t u(\tau)e(\tau)\mathrm{d}\tau \qquad (107)$$

and

$$a_M(t) = a_M(0) - \beta\int_0^t y_M(\tau)e(\tau)\mathrm{d}\tau \qquad (108)$$

The adaptation law is globally asymptotically stable for $\alpha > 0$ and $\beta > 0$.

6.3.2 General design method for series-parallel model approach: For identification of systems of higher order the series-parallel SRAM approach in state-space representation seems to be advantageous [108]. The series-parallel SRAM approach is realised by a combination of two partial models, one arranged in series and the other parallel to the system to be identified, as shown in Fig. 11 for a SISO system. The system to be identified is described by the state-space representation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{b}u(t) \qquad (109)$$

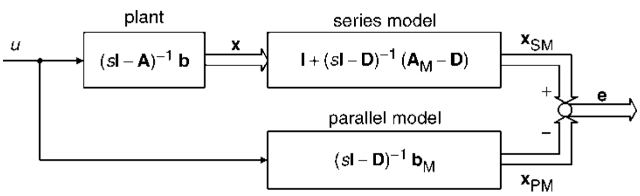


Fig. 11 *Series-parallel SRAM approach for SISO system in state-space representation*

where matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$ are unknown, and $\boldsymbol{x}(t)$ is measurable. In both partial models which are of the same order as the system to be identified, the free selectable system matrix $\boldsymbol{D}$ is fixed and has stable eigenvalues. In the series model, the matrix $\boldsymbol{A}_M$ and, in the parallel model, the vector $\boldsymbol{b}_M$, respectively, contain the parameters to be adapted such that the state error vector

$$\boldsymbol{e}(t) = \boldsymbol{x}_{SM}(t) - \boldsymbol{x}_{PM}(t) \qquad (110)$$

vanishes asymptotically. From Fig. 11, it follows that in the frequency domain directly

$$\mathrm{L}\{\boldsymbol{e}(t)\} = \boldsymbol{E}(s) = \left\{[\boldsymbol{I} - (s\boldsymbol{I} - \boldsymbol{D})^{-1}(\boldsymbol{A}_M - \boldsymbol{D})] \times (s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{b} - (s\boldsymbol{I} - \boldsymbol{D})^{-1}\boldsymbol{b}_M\right\}U(s) \qquad (111)$$

From this equation we obtain for $\boldsymbol{E}(s) = \boldsymbol{0}$ the steady-state (adapted) case:
(i) $\boldsymbol{b}_M = \boldsymbol{b}$ and (ii) $\boldsymbol{A}_M = \boldsymbol{A}$.

It is easy to show that the block diagram structure of Fig. 11 can also be transformed to that presented in Fig. 12, and is thus described by the state space representation

$$\dot{\boldsymbol{x}}_M(t) = \boldsymbol{D}\boldsymbol{x}_M(t) + \boldsymbol{b}_M(t)u(t) + [\boldsymbol{A}_M(t) - \boldsymbol{D}]\boldsymbol{x}(t) \qquad (112)$$

which, with the state error vector $\boldsymbol{e}(t) = \boldsymbol{x}(t) - \boldsymbol{x}_M(t)$, can also be represented by

$$\dot{\boldsymbol{x}}_M(t) = \boldsymbol{A}_M(t)\boldsymbol{x}(t) + \boldsymbol{b}_M(t)u(t) - \boldsymbol{D}\boldsymbol{e}(t) \qquad (113)$$

Subtracting (113) from (109) and considering the state error vector according to (110) and the parameter errors $\tilde{\boldsymbol{A}}(t) = \boldsymbol{A} - \boldsymbol{A}_M(t)$ and $\tilde{\boldsymbol{b}}(t) = \boldsymbol{b} - \boldsymbol{b}_M(t)$, finally the error vector differential equation follows as

$$\dot{\boldsymbol{e}}(t) = \boldsymbol{D}\boldsymbol{e}(t) + \tilde{\boldsymbol{A}}(t)\boldsymbol{x}(t) + \tilde{\boldsymbol{b}}(t)u(t) \qquad (114)$$

If the steady state $(\boldsymbol{e} = \boldsymbol{0},\ \tilde{\boldsymbol{A}} = \boldsymbol{0},\ \tilde{\boldsymbol{b}} = \boldsymbol{0})$ of the overall system should be globally asymptotically stable, then a possible Lyapunov function is to be found, which guarantees stability in the whole space spanned by the elements of $\boldsymbol{e}$, $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{b}}$. In view of this we take here the quadratic form

$$V(\boldsymbol{e}, \tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}}, t) = \frac{1}{2}\mathrm{tr}(\tilde{\boldsymbol{A}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{A}}^{-1}\tilde{\boldsymbol{A}}) + \frac{1}{2}\tilde{\boldsymbol{b}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{b}}^{-1}\tilde{\boldsymbol{b}} + \frac{1}{2}\boldsymbol{e}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{e} \qquad (115)$$

with positive definite symmetric weighting matrices $\boldsymbol{P}_{\mathrm{A}}$, $\boldsymbol{P}_{\mathrm{b}}$ and $\boldsymbol{P}$. The time derivative of this relation is given by

$$\dot{V}(\boldsymbol{e}, \tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}}, t) = \mathrm{tr}(\tilde{\boldsymbol{A}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{A}}^{-1}\dot{\tilde{\boldsymbol{A}}}) + \tilde{\boldsymbol{b}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{b}}^{-1}\dot{\tilde{\boldsymbol{b}}} + \frac{1}{2}\dot{\boldsymbol{e}}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{e} + \frac{1}{2}\boldsymbol{e}^{\mathrm{T}}\boldsymbol{P}\dot{\boldsymbol{e}} \qquad (116)$$

With (114) and the derivatives $\dot{\tilde{\boldsymbol{A}}}(t) = -\dot{\boldsymbol{A}}_M(t)$ and $\dot{\tilde{\boldsymbol{b}}}(t) = -\dot{\boldsymbol{b}}_M(t)$, it follows that after some simplifications from (116),

$$\dot{V}(\boldsymbol{e}, \tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}}, t) = \mathrm{tr}[\tilde{\boldsymbol{A}}^{\mathrm{T}}(\boldsymbol{P}\boldsymbol{e}\boldsymbol{x}^{\mathrm{T}} - \boldsymbol{P}_{\mathrm{A}}^{-1}\dot{\boldsymbol{A}}_M)] + \tilde{\boldsymbol{b}}^{\mathrm{T}}(\boldsymbol{P}\boldsymbol{e}u - \boldsymbol{P}_{\mathrm{b}}^{-1}\dot{\boldsymbol{b}}_M) + \frac{1}{2}\boldsymbol{e}^{\mathrm{T}}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{P} + \boldsymbol{P}\boldsymbol{D})\boldsymbol{e} \qquad (117)$$

The derivative is negative-definite if on the RHS of (117) the third term is negative-definite and the first and the second terms vanish, i.e. we should have

$$\boldsymbol{D}^{\mathrm{T}}\boldsymbol{P} + \boldsymbol{P}\boldsymbol{D} = -\boldsymbol{Q} \qquad (118)$$

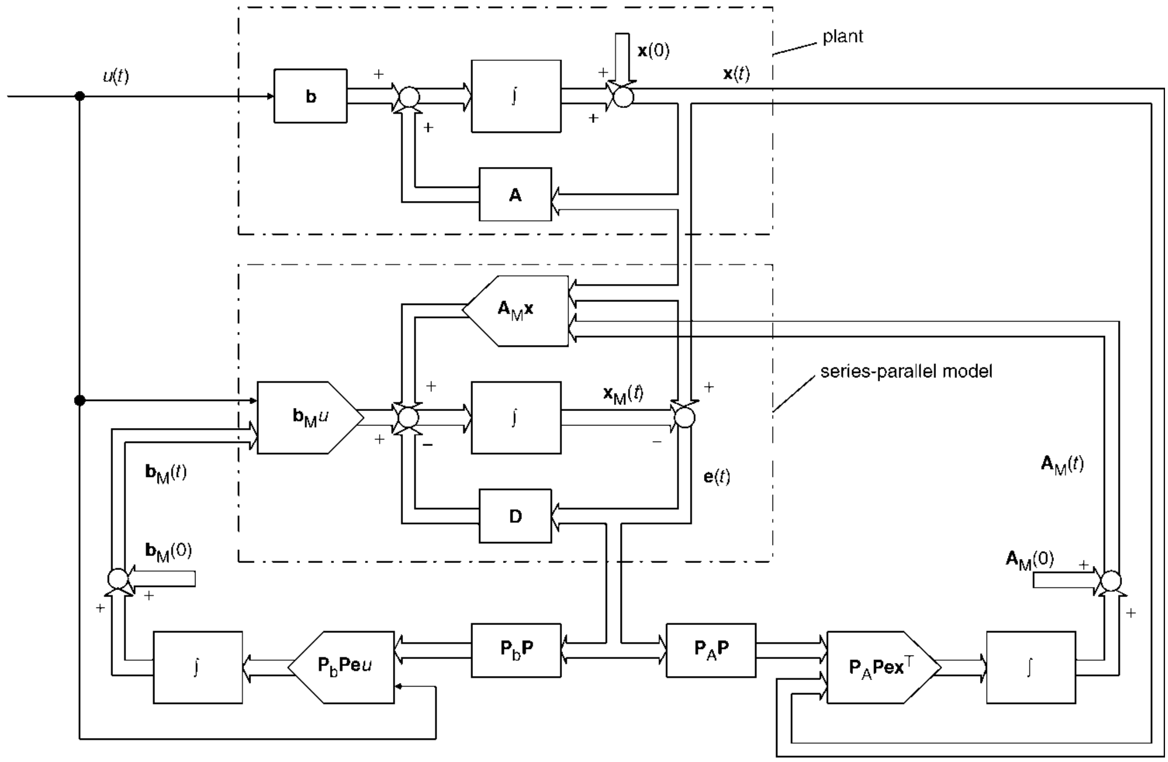


Fig. 12 *Block diagram for identification of linear SISO system by series-parallel SRAM using Liapunov's stability theory*

for an arbitrary symmetric and positive-definite matrix $\boldsymbol{Q}$ and further

$$\boldsymbol{Pex}^{\mathrm{T}} - \boldsymbol{P}_{\mathrm{A}}^{-1}\dot{\boldsymbol{A}}_{\mathrm{M}} = \boldsymbol{0} \tag{119}$$

$$\boldsymbol{Peu} - \boldsymbol{P}_{\mathrm{A}}^{-1}\dot{\boldsymbol{b}}_{\mathrm{M}} = \boldsymbol{0} \tag{120}$$

If $\boldsymbol{P}$ is a solution of the matrix Lyapunov equation (118), then from (119) and (120) through integration we directly obtain the adaptation laws for the model parameters as

$$\boldsymbol{A}_{\mathrm{M}}(t) = \boldsymbol{A}_{\mathrm{M}}(0) + \boldsymbol{P}_{\mathrm{A}}\boldsymbol{P}\int_0^t \boldsymbol{e}(\tau)\boldsymbol{x}^{\mathrm{T}}(\tau)\mathrm{d}\tau \tag{121a}$$

and

$$\boldsymbol{b}_{\mathrm{M}}(t) = \boldsymbol{b}_{\mathrm{M}}(0) + \boldsymbol{P}_{\mathrm{b}}\boldsymbol{P}\int_0^t \boldsymbol{e}(\tau)u(\tau)\mathrm{d}\tau \tag{121b}$$

Fig. 12 shows the corresponding block diagram for this identification method for which stability is well guaranteed. Considering special canonical forms, the adaptation law can be further simplified [10]. It should be mentioned that a stable SRAM approach for system identification can also be obtained by application of the concept of hyper-stability [106].

7 Parameter identification of linear CT models in frequency domain

In many practical situations, parameter estimation in the frequency domain is of considerable interest. Information on the system behaviour under the influence of periodic test signals is the main input to the related algorithms. Identification in frequency domain will be discussed in this Section with reference to continuous-time linear systems in the environment shown in Fig. 13, where $u(t)$ and $y(t)$, are measurable input and output signals respectively, and the output signal is corrupted by the coloured unmeasurable noise signal $n(t)$. The system is represented by the frequency-response characteristic (Nyquist plot or its equivalent Bode plot):

$$G_0(\mathrm{j}\omega) = \frac{Y(\mathrm{j}\omega)}{U(\mathrm{j}\omega)} = R_0(\omega) + \mathrm{j}I_0(\omega) \tag{122}$$

This information may be available at discrete frequency values $\omega = \omega_i$, distributed, for example, logarithmically over the frequency range of interest $\omega_{\min} \le \omega \le \omega_{\max}$, from computation or by measurement. $U(\mathrm{j}\omega)$ and $Y(\mathrm{j}\omega)$ are the Fourier transforms of $u(t)$ and $y(t)$, respectively. Measurements of $G_0(\mathrm{j}\omega_i)$ for discrete values of ω_i covering the whole frequency range can be obtained either

(*a*) from a discrete set of measured input-output spectra $U(\omega_i)$ and $Y(\omega_i)$, for $i = 0, 1, 2, \ldots, N$, or

(*b*) from direct excitation by periodic test signals, or indirectly from arbitrary time-domain measurements of $u(t)$ and $y(t)$, which, by Fourier transformation [10], convert the information into frequency domain, i.e.

$$G_0(\mathrm{j}\omega_i) \simeq \frac{y_0 - (1/\omega_i T_{\mathrm{S}})\sum_{\nu=0}^{L} p_\nu \mathrm{e}^{-\mathrm{j}(\omega_i \nu T_{\mathrm{S}} - \pi/2)}}{u_0 - (1/\omega_i T'_{\mathrm{S}})\sum_{\mu=0}^{M} q_\mu \mathrm{e}^{-\mathrm{j}(\omega_i \mu T'_{\mathrm{S}} - \pi/2)}} \tag{123}$$

where $p_0 = y_1 - y_0$, $p_\nu = y_{\nu-1} - 2y_\nu + y_{\nu+1}$, $\nu = 1, 2, \ldots, L$, $q_0 = u_1 - u_0$ and $q_\nu = u_{\mu-1} - 2u_\mu + u_{\mu+1}$, $\mu = 1, 2, \ldots, M$. The input and output signals are sampled as u_0,

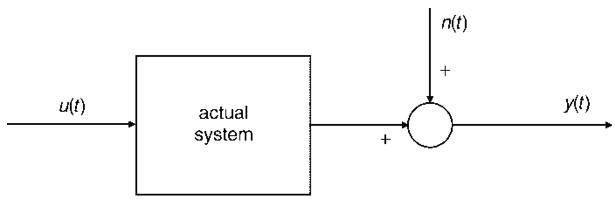


Fig. 13 *Identification environment*

$u_1, \ldots, u_{M+1}$ and $y_0, y_1, \ldots, u_{L+1}$ in equal time intervals of T_S' and T_S, respectively.

Let us assume that the system is modelled by

$$Y_M(s) = G_M(s, \boldsymbol{\theta})U(s) + G_n(s, \boldsymbol{\theta})W(s) \tag{124}$$

as shown in Fig. 14, where the transfer function $G_n(s, \boldsymbol{\theta})$ is identical to $H(s)$, in (2c), and represents the model for the stochastic component with coloured noise $n(t)$, $w(t)$ is normally distributed zero-mean white noise, and the deterministic model part is described by the rational transfer function

$$G_M(s, \boldsymbol{\theta}) = \frac{Y_S(s)}{U(s)} = \frac{b_0 + b_1 s + \cdots + b_n s^n}{a_0 + a_1 s + \cdots + a_n s^n} = \frac{B(s, \boldsymbol{\theta})}{A(s, \boldsymbol{\theta})} \tag{125}$$

The parameter vector is organised as

$$\boldsymbol{\theta} = [b_0 \quad b_1 \quad \cdots \quad b_n \;\vdots\; a_0 \quad a_1 \quad \cdots \quad a_n]^T \tag{126}$$

The problem of identification consists in estimating the vector $\boldsymbol{\theta}$ of the real parameters a_i and b_i $(i = 0, 1, \ldots, n)$. Usually, the Laplace transform of the output error [10], in the more general form compared to Fig. 7,

$$E^*(s, \boldsymbol{\theta}) = G_n^{-1}(s, \boldsymbol{\theta})[Y(s) - Y_S(s, \boldsymbol{\theta})] \tag{127}$$

as defined in Fig. 15, is introduced for this purpose. The best approximation for $G_0(j\omega)$ by the 'model' $G_M(j\omega, \boldsymbol{\theta})$ is then obtained by minimising the quadratic cost function of the model output error $e^*(t, \boldsymbol{\theta})$, i.e.

$$J(\boldsymbol{\theta}) = \boldsymbol{e}^{*T}\boldsymbol{e}^* \tag{128}$$

where $\boldsymbol{e}^* = [e^*(0) \; e^*(1) \; \cdots \; e^*(L)]^T$, whose elements are $L+1$ samples in equidistant intervals T_S of $e^*(kT_S, \boldsymbol{\theta}) = e^*(k, \boldsymbol{\theta})$. The estimated parameter vector is finally computed from

$$\hat{\boldsymbol{\theta}}_L = \arg\min_{\theta}\left[\sum_{k=0}^{L} e^{*2}(k, \boldsymbol{\theta})\right] \tag{129}$$

or, using Parseval's theorem in the frequency domain,

$$\hat{\boldsymbol{\theta}}_L = \arg\min_{\theta}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} |E^*(j\omega, \boldsymbol{\theta})|^2 d\omega\right] \tag{130}$$

where the Fourier transform of $\boldsymbol{e}^*(t, \boldsymbol{\theta})$ follows from (127), for $s = j\omega$, as

$$\begin{aligned} E^*(j\omega, \boldsymbol{\theta}) &= \int_{-\infty}^{\infty} e^*(t, \boldsymbol{\theta})e^{-j\omega t}dt \\ &= G_n^{-1}(j\omega, \boldsymbol{\theta})[Y(j\omega) - Y_S(j\omega, \boldsymbol{\theta})] \end{aligned} \tag{131a}$$

or using (125)

$$E^*(j\omega, \boldsymbol{\theta}) = G_n^{-1}(j\omega, \boldsymbol{\theta})[Y(j\omega) - G_M(j\omega, \boldsymbol{\theta})U(j\omega)] \tag{131b}$$

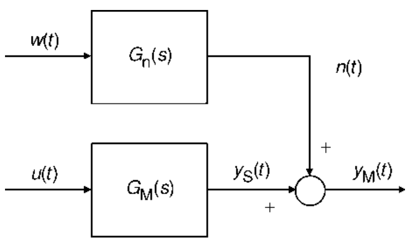


Fig. 14 *Model of actual system*

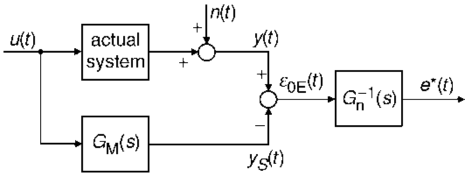


Fig. 15 *Definition of general form of output error e*(t)*

For the real signal $e^*(k, \boldsymbol{\theta})$, which has been introduced here, for the numerical treatment, the following Parseval's formula also holds directly:

$$\sum_{k=0}^{L} e^{*2}(k, \boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} E^{*2}(j\omega_i, \boldsymbol{\theta}) \tag{132}$$

Introducing (132) into (129) and replacing $E^*(j\omega_i, \boldsymbol{\theta})$ by (131b), for discrete frequencies $\omega = \omega_i$ logarithmically distributed over the whole frequency range $0 < \omega < \infty$, leads to

$$\begin{aligned} \hat{\boldsymbol{\theta}}_L = \arg\min_{\theta}\Bigg[\frac{1}{N}\sum_{i=1}^{N} &\frac{1}{|G_n(j\omega_i, \boldsymbol{\theta})|^2} \\ &\times |Y(j\omega_i) - G_M(j\omega_i, \boldsymbol{\theta})U(j\omega_i)|^2\Bigg] \end{aligned} \tag{133}$$

By introducing (122) in (133), we obtain

$$\hat{\boldsymbol{\theta}}_L = \arg\min_{\theta}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{|U(j\omega_i)|^2}{|G_n(j\omega_i, \boldsymbol{\theta})|^2}|G_0(j\omega_i) - G_M(j\omega_i, \boldsymbol{\theta})|^2\right] \tag{134}$$

If the $G_n(j\omega)$ for the stochastic component in (124) is independent of $\boldsymbol{\theta}$, then the parameter estimation by (133) will be consistent. This is usually the case if $G_n(j\omega)$ is given and fixed [109], but very often the parameters of the noise model are unknown and, therefore, have to be included in the unknown parameter vector $\boldsymbol{\theta}$ to be estimated.

It has been shown [10] that (134) leads to a nonlinear least-squares problem. Often Newton-Gauss algorithms have been used to minimise (134) [110], where the convergence region of the Newton-Gauss algorithm can be enlarged by using a Levenberg-Marquardt version [111]. Many techniques have been suggested in a linear least-squares framework to tackle the aforementioned nonlinear least-squares problem. Surveys dealing with this problem are given, for example, in [10, 110, 112].

An alternative to the least-squares estimation according to (134) provides the maximum likelihood (ML) estimation method [2, 110, 113]. For this method, the negative logarithm of the likelihood function becomes

$$\begin{aligned} L_L(\boldsymbol{\theta}) = \sum_{k=1}^{N}\Bigg\{\log|G_n(j\omega_i, \boldsymbol{\theta})|^2 + \frac{1}{\sigma_{e^*}^2}\frac{|U(j\omega_i)|^2}{|G_n(j\omega_i, \boldsymbol{\theta})|^2}|G_0(j\omega_i) \\ - G_M(j\omega_i, \boldsymbol{\theta})|^2\Bigg\} + N\log\sigma_{e^*}^2 \end{aligned} \tag{135}$$

where $\sigma_{e^*}^2$ is the variance of e^* according to Fig. 15 being white Gaussian noise with zero mean. The estimate of the parameter vector is then given by

$$\hat{\boldsymbol{\theta}}_L = \arg\min_{\theta} L_L(\boldsymbol{\theta}) \tag{136a}$$

or minimising (136a) with respect to $\sigma_{e^*}^2$ provides

$$\hat{\boldsymbol{\theta}}_L = \arg\min_{\theta}\left[N\log J(\boldsymbol{\theta}) + \sum_{i=1}^{N}\log|G_n(j\omega_i, \boldsymbol{\theta})|^2\right] \tag{136b}$$

where

$$J(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \frac{|U(\mathrm{j}\omega_i)|^2}{|G_{\mathrm{n}}(\mathrm{j}\omega_i, \boldsymbol{\theta})|^2} |G_0(\mathrm{j}\omega_i) - G_{\mathrm{M}}(\mathrm{j}\omega_i, \boldsymbol{\theta})|^2 \quad (137)$$

An estimate of $\sigma_{e^*}^2$ can be obtained from [10] as

$$\hat{\sigma}_{e^*}^2 = \frac{2}{N} J(\hat{\boldsymbol{\theta}}_N) \quad (138)$$

Compared to (134) the ML estimator according to (136*b*) includes the extra term

$$\sum_{i=1}^{N} \log |G_{\mathrm{n}}(\mathrm{j}\omega_i, \boldsymbol{\theta})|^2 \quad (139)$$

which is not essential if the noise model is given or known, as it is often assumed. It can be shown easily that $\hat{\boldsymbol{\theta}}_N = \hat{\boldsymbol{\theta}}_L(N)$ according to how (136*b*) converges to its true value as $N \to \infty$. However, for finite N the ML estimate $\hat{\boldsymbol{\theta}}_L(N)$ is finite-sample biased.

Owing to its numerical treatment the ML-method is one of the most efficient identification methods. The solution of (136*b*) is obtained by numerical optimisation. Several approaches have been proposed for the optimisation problem, using the first and second derivative of $J(\boldsymbol{\theta})$, i.e. $\boldsymbol{J}_\theta$ and $\boldsymbol{J}_{\theta\theta}$, with respect to $\boldsymbol{\theta}$. For example, in the case of the Newton-Raphson algorithm, the iterative solution for a new estimate is given by [109]

$$\hat{\boldsymbol{\theta}}_N(\nu + 1) = \hat{\boldsymbol{\theta}}_N(\nu) - \beta\{\boldsymbol{J}_{\theta\theta}[\hat{\boldsymbol{\theta}}_N(\nu)]\}^{-1} \boldsymbol{J}_\theta[\hat{\boldsymbol{\theta}}_N(\nu)] \quad (140)$$

where β is a constant factor. The numerical effort is justified, as, at the same time, the variance of the parameter estimation, for example for the parameter θ_i,

$$\hat{\sigma}_{\theta_i}^2 = \hat{\sigma}_{e^*}^2 [\boldsymbol{J}_{\theta\theta}]_{ii}^{-1} \quad (141)$$

can be obtained using (138) and the Hessian matrix $\boldsymbol{J}_{\theta\theta}$.

Up to this point we assumed the 'true' input $u(t)$ to be measurable (observable). This, however, can result in a biased parameter estimation when the real measured input $u_m(t)$ is corrupted by a noise signal $n_2(t)$, as shown in Fig. 16. This assumption can be circumvented by applying the errors-in-variable (EV) model structure [3, 114], or the instrumental variable (IV) method [2, 10]. In the frequency-domain identification, this EV method was extensively applied, e.g. [109]. It has been shown that estimators based on this model structure are consistent when the 'true' spectral density matrix of the I/O noise is known *a priori* [114]. In [115], a logarithmic LS estimator was analysed and shown that, even when the variances of the measured data are unknown, this estimator remains 'practically' consistent.

This EV method is also appropriate for the identification of a system in a feedback loop. It has been shown [116] that this problem can be solved by the EV method if the covariance matrices of the noise signals, $n_1(t)$ and $n_2(t)$, are known *a priori* or can be replaced by their sample covariance matrices, calculated from a small number of independent repeated experiments. The estimates thus obtained are still strongly consistent for a large enough set of data points in each experiment.

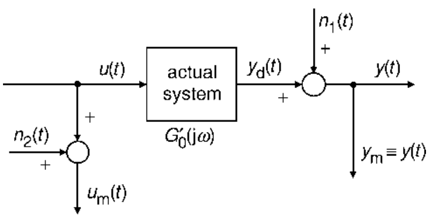


Fig. 16 *Errors-in-variable model structure*

In practice, identification of a continuous-time linear dynamic SISO-system usually starts with a discrete set of measurements of the input and output signals, sampled at equal time intervals, as already mentioned in the preceding text. It is well known that the discrete measurements do not contain all the information about the continuous-time signals unless additional assumptions are made. In [117], two very important assumptions have been considered: (i) the zero-order hold (ZOH) assumption and (ii) the band limited (BL) assumption. In case (i), it is assumed that the excitation signal $u(k)$ remains constant during the sampling interval T_{S}. In case (ii), it is assumed that the sampled signals, $u(k)$ and $y(k)$, each have limited bandwidth. Both assumptions lead to an exact description of the continuous-time system. However, it should be noted that the obtained models are only valid, if their signals obey the corresponding assumptions. In [117], it is shown that discrete-time modelling based in ZOH excitations should be used with care. On the other hand, the BL assumption offers specific advantages in many applications, especially when applying periodic excitation signals.

Usually, for system identification in time domain, large data sets are required. The record length is given by

$$T_{\mathrm{m}} = L T_{\mathrm{S}} \quad (142)$$

where L is the number of data points, and T_{S} is the sampling time. The sampling frequency ω_{S} should be selected according to Shannon's sampling theorem

$$\omega_{\mathrm{S}} = \frac{2\pi}{T_{\mathrm{S}}} \geq 2\omega_{\max} \quad (143)$$

where $\omega_{\max}$ is the frequency of the band-limited input or output signal above which the signal spectrum vanishes. The lowest frequency of interest $\omega_{\min}$ determines the minimal record length

$$T_{\mathrm{m}} > \frac{2\pi}{\omega_{\min}} \quad (144)$$

From (142)–(144), it is clear that, for the minimum number of data points in the time domain,

$$L \geq 2\omega_{\max}/\omega_{\min} \quad (145)$$

So if the frequency range of interest $\omega_{\min} \leq \omega \leq \omega_{\max}$ covers, for example, 3 decades, it is necessary that $L = 2000$ data points.

The large data sets, usually necessary in time domain, can be replaced for identification in frequency domain by a considerably reduced set of approximately logarithmically distributed frequency points ω_i covering the frequency range of interest [117, 118]. Decimating data of higher frequencies involves averaging over neighbouring frequencies, and it also helps in reducing high-frequency noise. This can be done if especially periodic excitation signals are employed.

Periodic input signals for system identification offer many advantages, compared to random and transient ones. These signals have discrete power spectra, and, consequently, it is possible to select only those frequencies ω_i of the outgoing signal $y(t)$ that have been injected, and discard others due to noise. Thus a considerable improvement of the signal-to-noise ratio of the measurements is possible.

As the frequency response data, such as Fourier coefficients at different frequencies ω_i, may be obtained from different experiments, an experimental simplification is possible by combining data from different experiments, which is not so easily possible in time-domain identification.

Besides other periodic test signals, the multi-sine test signals [117, 119] turn out to be the most flexible, providing optimal properties. These are band-limited signals consisting of an arbitrary sum of harmonically related sinusoidal terms

$$u(t) = \sum_{k=1}^{F} A_k \sin(m_k \omega_0 t + \varphi_k); \; m_k \in [1, 2, \ldots] \qquad (146)$$

where ω_0 is the basic angular frequency, A_k are the amplitudes and φ_k the phases. The main advantage of such test signals is that their total spectral power can be concentrated within a specific frequency range. However, their spectral power can be completely arbitrary, because m_k can be any integer. In addition, the amplitudes and phases must be carefully selected to minimise the crest factor CF:

$$CF = \frac{\text{peak value}|u(t)|_{\max}}{\text{effective value } u(t)_{\text{rms}}} \qquad (147)$$

of the signal $u(t)$. The motivation for minimising CF is to maximise the input-power subject avoiding input actuator saturation. There are several methods for minimising the value of CF. A small value of CF indicates a small noise-to-signal ratio of the measurements. Finally it should also be mentioned that, unlike in time-domain identification, no initial state estimation, but only steady-state conditions are necessary, if frequency-domain estimation is applied in combination with periodic excitation [110].

Besides the parameter estimation problem, there exists the problem of determining the order n of the model transfer function in (125). A number of methods developed and successfully applied for time-domain identification can also be applied for order determination in frequency-domain identification as well [110].

There exists a problem with the parameter estimation of high-order systems based on (134) or (136*b*), numerical problems in calculating the coefficients a_i and b_i in (125) may occur. This numerical ill-conditioning problem can, however, be solved in the frequency domain through the use of orthogonal polynomials. These polynomials have to be orthogonal on the discrete set of measured frequency points and depend on the available measurements [120].

There exist two MATLAB toolboxes [121, 122], which now offer a reliable set of routines for all identification steps in the frequency domain.

8 Subspace identification or CT state-space models

8.1 State-space CT model identification problem

To give a better insight, we will discuss, in the following, the deterministic case for the sake of simplicity. Let us assume that N measurements of the I/O vectors $\boldsymbol{u}(t) \in \Re^r$ and $\boldsymbol{y}(t) \in \Re^m$ are given. The identification problem then consists in estimating the order n of the state vector $\boldsymbol{x}(t) \in \Re^n$ and the elements of the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ of the state-space representation:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t) \qquad (148)$$

$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t) + \boldsymbol{D}\boldsymbol{u}(t) \qquad (149)$$

with corresponding dimensions.

8.1.1 Subspace model identification (SMI) representation: By introducing up to $i-1$ derivatives of $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$ for $i > n$ at N different time instants, $t = \{t_1, t_2, \ldots, t_N\}$, not necessarily equidistant, but i and N sufficiently large, the state-space representation according to (148) and (149) can be transformed to the matrix equation, denoted also as the data equation,

$$\boldsymbol{Y}_i = \boldsymbol{T}_{1,i}\boldsymbol{X} + \boldsymbol{T}_{2,i}\boldsymbol{U}_i \qquad (150)$$

which is the basis of the SMI approach, where

$$\boldsymbol{X} = [\boldsymbol{x}(t_1) \quad \boldsymbol{x}(t_2) \quad \cdots \quad \boldsymbol{x}(t_N)] \qquad (151)$$

$$\boldsymbol{U}_i = \begin{bmatrix} \boldsymbol{u}(t_1) & \cdots & \boldsymbol{u}(t_N) \\ \dot{\boldsymbol{u}}(t_1) & & \dot{\boldsymbol{u}}(t_N) \\ \vdots & & \vdots \\ \overset{(i-1)}{\boldsymbol{u}}(t_1) & & \overset{(i-1)}{\boldsymbol{u}}(t_N) \end{bmatrix} \qquad (152)$$

$\overset{r}{(\cdot)} \hat{=} \mathrm{d}^r/\mathrm{d}t^r(\cdot)$, $\boldsymbol{Y}_i$ identical in structure to $\boldsymbol{U}_i$ but $\boldsymbol{u}(t)$ replaced by $\boldsymbol{y}(t)$,

$$\boldsymbol{T}_{1,i} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{CA} \\ \vdots \\ \boldsymbol{CA}^{i-1} \end{bmatrix} \in \Re^{mi \times n} \qquad (153)$$

$$\boldsymbol{T}_{2,i} = \begin{bmatrix} \boldsymbol{D} & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \boldsymbol{CB} & \boldsymbol{D} & & \boldsymbol{O} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{CA}^{i-2}\boldsymbol{B} & \boldsymbol{CA}^{i-3}\boldsymbol{B} & \cdots \; \boldsymbol{CB} & \boldsymbol{D} \end{bmatrix} \in \Re^{mi \times ri} \qquad (154)$$

$\boldsymbol{T}_{1,i}$ and $\boldsymbol{T}_{2,i}$ are denoted as extended ($i > n$) observability and extended controllability matrices, respectively. As the derivatives of $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$ in the data matrices $\boldsymbol{U}_i$ and $\boldsymbol{Y}_i$ are not directly measurable, they must be replaced either by approximations, or new signals have to be generated by introducing a signal-preprocessing operator R_{LD}, as shown in Fig. 6. (There are several manifestations of R_{LD} (MMF: method of multiple filters; PMF: Poisson moment functionals; GHOF: general hybrid orthogonal functions; PCBF: piecewise-constant basis functions; CBF: continuous basis functions; WF: Walsh functions; HF: Haar functions; BPF: block-pulse functions; HMF: Hartley modulating functions; FMF: Fourier modulating functions).)

The latter can be obtained for the SMI approach by several methods. In [123] linear filters are used to obtain a CT data equation. A similar approach was applied in [124]. A Laguerre filter approach was introduced in [125–127] and a distribution-theoretic approach was proposed to describe derivatives in the stochastically disturbed case. The approach described in [44] makes use of Poisson moment functional (PMF) filtering. SMI methods had been originally proposed for discrete-time (DT) systems [128–132], and their extension to CT systems was first mentioned in [123] and subsequently refined in several contributions, as for example [44, 124–127, 133, 134]. The basic idea behind SMI algorithms is that they are able to retrieve system-related matrices $\hat{\boldsymbol{A}}$, $\hat{\boldsymbol{B}}$, $\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{D}}$ as subspaces of projected data matrices [132].

8.1.2 Different families of SMI implementations: There are three different possibilities for SMI implementations:

(*a*) multivariable-output-error state-space (MOESP) approach [135],

(b) numerical algorithms for subspace state-space systems identification (N4SID) [136], and
(c) canonical variate analysis (CVA) [129].

Without going into detail, it should be mentioned that all three types of algorithms differ much concerning their computational requirements. Although the CVA approach has a significant lower number of floating point operations, the total CPU time is much higher than for the other two. Also a combination of the MOESP and N4SID algorithms is possible and provides a number of advantages [126]. The basic idea of the SMI technique will be outlined in the following Section using the N4SID combined with the PMF-filtering approach for realising the aforementioned signal-preprocessing operator $R_{\rm LD}$.

8.2 N4SID-PMF approach

8.2.1 PMF signal preprocessing: As already mentioned in Section 3, the ith PMF transform of a signal $x(t)$ at time instant t is given generally by the ith Poisson moment functional (PMF):

$$M_i\{x(t)\} = \int_0^t x(\tau)p_i(t-\tau)\mathrm{d}\tau, \quad i = 0, 1, 2, \ldots \tag{155}$$

where

$$p_i(t) = \frac{t^i}{i!}e^{-\lambda t} = \mathsf{L}^{-1}\left\{\frac{1}{(s+\lambda)^{i+1}}\right\} \tag{156}$$

is the Poisson impulse response function, and $\lambda^{-1} > 0$ is the time constant of the corresponding lowpass filter. If $x(t)$ is the input of a cascaded chain of $(n+1)$ identical lowpass filters, each described by $1/(s+\lambda)$, then all $M_i\{x(t)\}$ are directly measurable at time t from the n output stages of these filters, which all together form the Poisson filter chain. Notice that the PMFs of derivatives of $x(t)$ can be expressed as linear combinations of those of the original function itself. When neglecting all initial conditions of $x(t)$ and its derivatives after some time τ, then, for example, for $i = 2$, we have the following:

$$\begin{bmatrix} M_2\{x(t)\} \\ M_2\{\dot{x}(t)\} \\ M_2\{\ddot{x}(t)\} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -\lambda & 1 & 0 \\ \lambda^2 & -2\lambda & 1 \end{bmatrix} \begin{bmatrix} M_2\{x(t)\} \\ M_1\{x(t)\} \\ M_0\{x(t)\} \end{bmatrix} \tag{157}$$

This result can be generalised by

$$\boldsymbol{v}_i(t) = \boldsymbol{D}_i(t)\boldsymbol{v}_i^*(t) \tag{158}$$

where

$$\boldsymbol{v}_i(t) = [M_i\{x(t)\} \quad M_i\{\dot{x}(t)\} \quad \cdots \quad M_i\{x^{(i)}(t)\}]^{\rm T}$$
$$\boldsymbol{v}_i^*(t) = [M_i\{x(t)\} \quad M_{i-1}\{x(t)\} \quad \cdots \quad M_0\{x(t)\}]^{\rm T}$$
$$\boldsymbol{D}_i = \boldsymbol{D}_i(\lambda) \in \Re^{(i+1)\times(i+1)}$$

The state-space representation of the PMF-filter chain follows directly from (155), (156) and (158) as

$$\dot{\boldsymbol{v}}_i^*(t) = \begin{bmatrix} -\lambda & 1 & \cdots & 0 \\ 0 & -\lambda & \ddots & \vdots \\ \vdots & & \ddots & 1 \\ 0 & \cdots & 0 & -\lambda \end{bmatrix} \boldsymbol{v}_i^* + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} x(t) \tag{159}$$

To estimate the derivatives in the matrices $\boldsymbol{U}_i$ and $\boldsymbol{Y}_i$, we apply the PMF transformation (155) element-wise to the matrix (150) as follows:

$$M_i\{\boldsymbol{Y}_i\} = \boldsymbol{T}_{1,i}M_i\{\boldsymbol{X}\} + \boldsymbol{T}_{2,i}M_i\{\boldsymbol{U}_i\} \tag{160a}$$

or

$$\boldsymbol{Y}_i^* = \boldsymbol{T}_{1,i}M_i\{\boldsymbol{X}\} + \boldsymbol{T}_{2,i}\boldsymbol{U}_i^* \tag{160b}$$

where now the ‘filtered’ I/O matrices $\boldsymbol{U}_i^*$ and $\boldsymbol{Y}_i^*$ are known, as they are only dependent on the filtered I/O signals. The equivalence of (160b) and (150) is only given if the corresponding matrices have the same rank. Furthermore, for numerical treatment (159), and thus (160b), have to be discretised using a ZOH on the input. Moreover, to avoid problems with nonzero initial conditions [44], $\boldsymbol{U}_i^*$ and $\boldsymbol{Y}_i^*$ have to be reduced by eliminating their first m columns.

8.2.2 Estimation of the extended observability matrix $\boldsymbol{T}_{1i}$: If the filtered matrices $\boldsymbol{U}_i^*$ and $\boldsymbol{Y}_i^*$ are discretised, reduced and then denoted as $\boldsymbol{U}_i^{**}$ and $\boldsymbol{Y}_i^{**}$, respectively, and arranged in the following form, then a QR-factorisation can be performed:

$$\begin{bmatrix} \boldsymbol{U}_i^{**} \\ \boldsymbol{Y}_i^{**} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{0} \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}_1 \\ \boldsymbol{Q}_2 \end{bmatrix} \tag{161}$$

Under the condition of persistently exciting input signals

$$\mathrm{E}\{\boldsymbol{U}_i\ \boldsymbol{U}_i^{\rm T}\} > \boldsymbol{0} \tag{162}$$

the initial state is zero and i is chosen larger than the system order ($i > n$), the column space of $\boldsymbol{R}_{22}$ equals that of $\boldsymbol{T}_{1,i}$ [135]. The computation of $\boldsymbol{R}_{22}$ via its singular value decomposition (SVD) is as follows:

$$\boldsymbol{R}_{22} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\rm T} = [\boldsymbol{U}_n \;\vdots\; \boldsymbol{U}_0] \begin{bmatrix} \boldsymbol{S}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_0 \end{bmatrix} [\boldsymbol{V}_n \quad \boldsymbol{V}_0]^{\rm T} \tag{163}$$

and provides an approximation of the order n of the system by the number $\hat{n}$ of nonzero singular values of $\boldsymbol{S}$:

$$\hat{n} = \text{rank } \boldsymbol{S} \tag{164}$$

and the first $\hat{n}$ left singular vectors of $\boldsymbol{U}$, gathered in $\boldsymbol{U}_n$, determine the column space as an approximation of the extended observability matrix (153):

$$\hat{\boldsymbol{T}}_{1,i} = \boldsymbol{U}\boldsymbol{S}^{1/2} \tag{165}$$

8.2.3 Estimation of the state model matrices A and C: If we take $\boldsymbol{U}_1$ as the upper $(i-1)$ m rows of $\boldsymbol{U}_n$ and $\boldsymbol{U}_2$ the lower $(i-1)$ m rows of $\boldsymbol{U}_n$, then the approximations of $\boldsymbol{A}$ and $\boldsymbol{C}$ are as follows:

$$\hat{\boldsymbol{A}} = \boldsymbol{U}_1^{\dagger}\boldsymbol{U}_2 \tag{166}$$

with the pseudo-inverse $\boldsymbol{U}_1^{\dagger} = \boldsymbol{U}_1^{\rm T}(\boldsymbol{U}_1\ \boldsymbol{U}_1^{\rm T})^{-1}$, and

$$\hat{\boldsymbol{C}} = \text{the upper } m \text{ rows of } \boldsymbol{U}_n \tag{167}$$

*8.2.4 Estimation of the state model matrices **B** and **D**:* With the estimated matrices $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{C}}$, the matrices $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{C}}$ included in $\boldsymbol{T}_{2,i}$ can be computed in the deterministic case uniquely. In the ‘filtered’ data equation (160b) the term $\boldsymbol{T}_{1,i}$ can be removed by left multiplication with its orthogonal being of full rank and satisfying

(i) $\boldsymbol{T}_{1,i}^{\perp}\boldsymbol{T}_{1,i} = \boldsymbol{0}, \boldsymbol{T}_{1,i}^{\perp} \in \Re^{(mi-n)\times mi}$ and
(ii) $\boldsymbol{T}_{1,i}^{\perp} = \mathbf{I} - \boldsymbol{T}_{1,i}^{\rm T}(\boldsymbol{T}_{1,i}\boldsymbol{T}_{1,i}^{\rm T})^{-1}\boldsymbol{T}_{1,i}$

This leads to

$$\boldsymbol{T}_{1,i}^{\perp} \boldsymbol{Y}_i^* = \boldsymbol{T}_{1,i}^{\perp} \boldsymbol{T}_{2,i} \boldsymbol{U}_i^* \tag{168}$$

For easy extraction of the desired matrices $\boldsymbol{B}$ and $\boldsymbol{C}$, (168) is postmultiplied by the pseudo-inverse $\boldsymbol{U}_i^{*\dagger}$, from its right-hand side,

$$\boldsymbol{T}_{1,i}^{\perp} \boldsymbol{Y}_i^* \boldsymbol{U}_i^{*\dagger} = \boldsymbol{T}_{1,i}^{\perp} \boldsymbol{T}_{2,i} \tag{169}$$

and for simpler notation the left-hand side is denoted by the matrix

$$\boldsymbol{K} = [\boldsymbol{K}_1 \quad \boldsymbol{K}_2 \quad \cdots \quad \boldsymbol{K}_i]$$

and $\boldsymbol{T}_{1,i}^{\perp}$ is replaced by

$$\boldsymbol{L} = [\boldsymbol{L}_1 \quad \boldsymbol{L}_2 \quad \cdots \quad \boldsymbol{L}_i]$$

Thus, (169) can be rewritten as

$$\begin{bmatrix} \boldsymbol{K}_1 \\ \boldsymbol{K}_2 \\ \vdots \\ \boldsymbol{K}_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{L}_1 & \boldsymbol{L}_2 & \cdots & \boldsymbol{L}_{i-1} & \boldsymbol{L}_i \\ \boldsymbol{L}_2 & \boldsymbol{L}_3 & \cdots & \boldsymbol{L}_i & \boldsymbol{0} \\ \vdots & & & & \\ \boldsymbol{L}_i & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_{\mathrm{m}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_{1,i} \end{bmatrix} \begin{bmatrix} \boldsymbol{D} \\ \boldsymbol{B} \end{bmatrix} \tag{170}$$

which represents an overdetermined system of linear equations that can be solved for $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{D}}$ using, for example, the LS approach. The numerical application and the limits of this algorithm are discussed in detail in [44]. It should be mentioned that the method described here is directly applicable also for the case of an additional white zero-mean perturbation, with variance σ^2 acting on the output signal. Further cases, where (i) the output is disturbed by a coloured noise signal, or (ii) additionally to (i) the state equation has an additive white noise term, or (iii) additionally to (ii) also the input is disturbed by a measurement noise (including the closed-loop identification problem), are discussed in [123, 124].

Summarising it can be stated that the SMI approach differs significantly from the previously discussed identification methods. For example, no explicit cost function has to be minimised. The order determination directly obtained from the measured data is performed very simply. Furthermore, the SVD algorithm allows a numerically stable solution, based mainly on geometrical properties of signal spaces. In particular, the identification of MIMO systems can easily be handled by the SMI approach.

9 Identification of nonlinear CT systems

Nonlinear CT models frequently arise if a real system is derived from first principles, such as the laws of physics or chemistry. If the resulting model is structurally known, but contains unknown parameters, it is denoted by a 'grey-box' model. The importance of estimating the parameters for such models is stressed by many authors, as for example [137] and [138]. Contrary to linear models, nonlinear models in the time-domain can only be discretised approximately, therefore one of the most natural choices is a continuous-discrete model, i.e. a model with continuous-time dynamics and a discrete-time measurement equation. Other approaches in the frequency domain aim at, for example, transforming the nonlinear differential equation into an algebraic equation which can be linear or nonlinear in the parameters to be estimated and solved by different methods. Figure 17 gives a survey about the major structure of nonlinear CT models available for the solution of the identification problem. This structure can be categorised into the three major groups of (i) nonparametric models, (ii) parametric models and (iii) semiparametric models [139]. In addition, the dynamic behaviour of a nonlinear system can be described by a linear multimodel [140], consisting of different linear models for different operating points covering the entire range of operation. In special cases, even nonlinear multimodels are recommended [141]. However, this category of multimodels will, as well as the semiparametric artificial neural network models [142] and fuzzy models [143], not be discussed here.

9.1 Nonparametric nonlinear models

This class of models includes the functional series approaches and the generalised-frequency-response-function (GFRF) method.

9.1.1 Functional series models: These models had been already discussed in detail in a former survey paper [144]. The Volterra series is an intuitively satisfying description for nonlinear systems, which generalises the concept of the I/O representation of a linear system in the form of a convolution integral,

$$y(t) = \int_0^t g_1(\tau_1) u(t - \tau_1) \mathrm{d}\tau_1$$

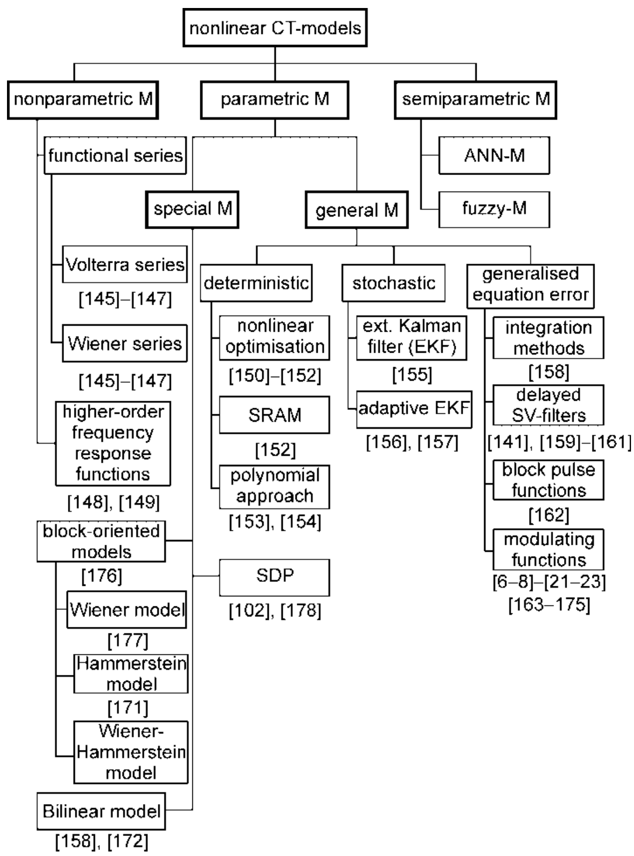


Fig. 17 *Major structure of nonlinear continuous-time models*

to the representation of nonlinear systems

$$\begin{aligned} y(t) &= \sum_{i=1}^{\infty} y_i(t) \\ &= \sum_{i=1}^{\infty} \int_0^t \int_0^t \cdots \int_0^t g_i(\tau_1, \tau_2, \ldots, \tau_i) \\ &\quad \times \prod_{j=1}^{i} \{u(t-\tau_j)\mathrm{d}\tau_j\} \end{aligned} \tag{171}$$

where $g_i(\tau_1, \ldots, \tau_i)$ is an impulse response or weighting function or the Volterra kernel of ith degree. The aim of identification is to obtain (nonparametric) values for the Volterra kernels g_i over the desired time range by, for example, numerical deconvolution of (171). Besides the classical phase-plane description, the nonparametric system description in the form of this Volterra series model was one of the earliest approximation of nonlinear systems introduced for the first time in 1958 by Wiener [145]. The model (171) can be extended to the stochastically disturbed case. Wiener introduced, for the case of white-noise input, a similar nth-order functional, however the kernels of this Wiener series model are not equal to the Volterra kernels [144].

Many different methods of calculating the Volterra and Wiener kernels have been proposed. However, the excessive computations associated with both models, due to the large number of parameters to be estimated in the second- and higher-order kernels, have resulted in very few applications of these techniques. Even though some minor improvements have been proposed [146, 147] for the numerical treatment during the last few years, this modelling framework has not become very popular.

9.1.2 Frequency-domain models: If a nonlinear stable system in the neighbourhood of an equilibrium point is described by a truncated Volterra series as in (171) with a finite summation, i.e. ∞ is replaced by N, where N denotes the degree of the nonlinearity, and transformed to the frequency domain by Fourier transformation, then the output can be expressed as

$$Y(\mathrm{j}\omega) = \sum_{i=1}^{N} Y_i(\mathrm{j}\omega) \quad \forall\omega \tag{172}$$

where

$$\begin{aligned} Y_i(\mathrm{j}\omega) &= \frac{1/\sqrt{i}}{(2\pi)^{i-1}} \int_{\omega_1+\omega_2+\cdots+\omega_i=\omega} G_i(\mathrm{j}\omega_1, \ldots, \mathrm{j}\omega_i) \\ &\quad \times \prod_{\mu=1}^{i} U(\mathrm{j}\omega_\mu)\mathrm{d}\sigma_\omega \end{aligned} \tag{173}$$

where $Y_i(\mathrm{j}\omega)$ and $U(\mathrm{j}\omega)$ represent the Fourier transforms of $y_i(t)$ and $u(t)$, the integral expression denotes the integration over the i-dimensional hyperplane ω and $G_i(\mathrm{j}\omega_i, \ldots, \mathrm{j}\omega_2)$ is defined as the ith-order GFRF [148]. It should be mentioned that the nonlinear frequency-domain outputs $Y_i(\mathrm{j}\omega)$ depend on the association of variables of each degree of the system nonlinearities, which induces mixing and intermodulation effects of the input frequencies to produce outputs at new frequency models [149]. GFRFs can be obtained, for example, by fitting a nonlinear ARMAX model (NARMAX) to the system and then mapping this model into the frequency domain. In [148] a new algorithm for solving this problem is proposed which produces unbiased estimation for a stochastically disturbed system. The nonlinear model can be constructed sequentially by building in the linear model forms, followed by the quadratic terms and so on. At each stage, the significance of each candidate model term is tested and only relevant model terms are included, thus a nonlinear differential equation model can be obtained component-wise.

9.2 Parametric nonlinear models

This class of models can be categorised into general and special models.

9.2.1 General parametric nonlinear models: In this category we will discuss deterministic as well as stochastic approaches and methods based on the (generalised) equation error.

(*a*) *Deterministic approaches:* Three different types will be presented in the following.

(i) *Nonlinear optimisation techniques:* The identification problem can be formulated as a minimisation of an error criterion based, for example, on the error between the model and the measured output data of the system to be identified. This leads usually to a nonlinear parametric optimisation problem, that must be solved iteratively. Different optimisation methods can be applied [150, 151]. If the nonlinearity is not known, the nonlinear function can be approximated by a polynomial whose coefficients are also parameters to be identified.
(ii) *Nonlinear SRAM-approach:* In [152] a specific system identification method is applied, for joint state and parameter estimation of nonlinear CT systems based on the parameter adjustment of a simulator model with the same dynamic structure as the plant. The identification is performed online. The parameter adjustment is of SRAM type and based on extremum searching of a multimodel nonlinear function using the Newton's searching method. Only the deterministic case is considered.
(iii) *Polynomial approach:* In [153] a series of piecewise multiple general orthogonal polynomials (PMGOP) is introduced for the parameter identification of a specific class of nonlinear CT systems. A recursive implementation of the method is possible. In [154] a nonlinear CT model is identified from process data through normal form theory. The model structure, in general, contains a low-order-polynomial vector-field characteristic of the particular system. Parameters for this structure are found through a nonlinear LS algorithm.

(*b*) *Stochastic approaches:* Let us assume a nonlinear stochastically disturbed time-varying system to be described by the state-space equation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}[\boldsymbol{x}(t), \boldsymbol{u}(t), \hat{\boldsymbol{\theta}}(t)] + \boldsymbol{v}(t) \tag{174}$$

and its measurement equation, at discrete time instants t_k,

$$\boldsymbol{y}(t_k) = \boldsymbol{h}[\boldsymbol{x}(t_s), \boldsymbol{u}(t_k), \boldsymbol{\theta}(t_k)] + \boldsymbol{w}(t_k) \tag{175}$$

where $\boldsymbol{v}(t)$ and $\boldsymbol{w}(t)$ are white noise signals, then the aim of the stochastic approaches for system identification from sampled I/O data consists in finding a continuous-discrete model description in the form of a continuous-time state equation

$$\dot{\hat{\boldsymbol{x}}}(t) = \hat{\boldsymbol{f}}[\hat{\boldsymbol{x}}(t), \boldsymbol{u}(t), \hat{\boldsymbol{\theta}}(t)] \tag{176}$$

and a discrete-time output equation

$$\hat{\boldsymbol{y}}(t_{k+1}) = \boldsymbol{h}[\hat{\boldsymbol{x}}(t_{k+1}^-), \boldsymbol{u}(t_{k+1}), \boldsymbol{\theta}(t_k)] \tag{177}$$

where the superscript '−' indicates the time instant just before a measurement is taken. Generally, both the structure and the parameter vector $\boldsymbol{\theta}(t)$ have to be determined. Once the structure is determined, parameter estimation algorithms can be used to find the unknown values of the system parameters from measurements of the I/O signals. Often it is desired that these parameters are estimated recursively in real time during normal operation of the system. This occurs, for example, if the parameters are time-varying and need to be monitored online. A general approach that subsumes a variety of recursive parameter estimation algorithms is the recursive prediction error method [155]. This method, which is applicable to a wide class of models, starts from formulating a predictor for a given model. This predictor is then adjusted such that with the prediction error

$$\boldsymbol{e}(t_{k+1}) = \boldsymbol{y}(t_{k+1}) - \hat{\boldsymbol{y}}(t_{k+1}) \tag{178}$$

a quadratic cost function

$$J(t_{k+1}, \boldsymbol{\theta}) = \mathrm{E}\{\tfrac{1}{2}\boldsymbol{e}^{\mathrm{T}}(t_{k+1}, \boldsymbol{\theta})\boldsymbol{W}(t_{k+1})\boldsymbol{e}(t_{k+1}, \boldsymbol{\theta})\} \tag{179}$$

is minimised, where $\boldsymbol{W}$ represents a sequence of weighting matrices. There are several algorithms to solve this problem.

Quite commonly, a gradient-based approach is used to adjust the parameter estimates. This requires the gradient of the cost functional, and, consequently, the gradient of the prediction error, and also the gradient of the prediction error covariance matrix, although the latter is often neglected. These gradients can be obtained from a so-called sensitivity model. However, the derivation of such a sensitivity model is extremely complicated even for the simplest nonlinear filter, the extended Kalman filter (EKF). This complication arises because the filter gain depends on the system parameters. For linear systems, this problem can be alleviated by using an innovations model. The innovations model corresponds to the stationary Kalman filter with a constant gain. This filter gain can then be independently parameterised and estimated [155]. From a theoretical point of view, this constant-gain assumption is only justifiable for linear systems. Nevertheless, it has also been applied to nonlinear systems. In [156], sensitivity models for four popular continuous-discrete approximate nonlinear filters are developed. The derivation of these sensitivity models is based on a reformulation of the higher-order filters reported in [157] and makes heavy use of matrix differential calculus. With the sensitivity models of these filters, a general adaptive filtering algorithm has been developed. Practical applications showed that already the simplest adaptive filter, the adaptive extended Kalman filter, gave sufficient results and it was found unnecessary to use higher-order filters. The potential of this method has also been shown for failure detection and is due to its fast and accurate convergence rate.

(*c*) *Methods based on (generalised) equation error (EE/GEE):* The following considerations refer to identification methods based on the minimisation of EE/GEE as defined in Fig. 8. Various approaches belong to this general category:

(i) *Integration methods:* Into this group fits the newly introduced 'reinitialised partial moments function' approach, which has been applied to some nonlinear systems and especially to bilinear ones [158]. The I/O data are sampled and at any sampling step reinitialised partial moments are used. The method is based on the minimisation of a quadratic cost function of the equation error over a fixed observation interval. The original iterative offline Marquardt algorithm is transformed to a recursive version, under the assumption that the next estimate of the actual parameter vector is located in the near neighbourhood of the estimate of the previous sampling step.

(ii) *Delayed state-variable (SV) filters:* The basic idea of this approach was originally proposed in [159]. However, its implementation was limited, due to the deficiency of the selected delay filters, therefore a new method [160] has been proposed for identifying nonlinear CT systems from sampled data records, based on special SV filters and coupled with an orthogonal LS estimation. The main idea consits in the special choice of delay or transportation lag filters. The result of passing a signal through a nonlinear function and then passing this result through a transportation lag filter is exactly the same when the nonlinear function and the transportation lag filter are changed. Application of Butterworth filter group delays equalised with two second-order all-pass filters provides a good approximation to the transportation lag device. The delay-filtered input as well as output signals and their associated higher-order derivatives can be used for identification of the unknown system parameters using an orthogonal LS estimator. For better determining the cutoff frequency of the filters a RBF network model is proposed in [161], and its structure is properly obtained by applying a genetic algorithm.

(iii) *Block-pulse functions:* In [162] an interesting approximation algorithm for parameter identification of systems governed by nonlinear differential equations, by means of a block-pulse operator (BPO) is described. The basic idea of this BPO method is to convert the problem of parameter identification in the original function space into an equivalent approximate identification problem in the image space of the BPO, where the algortithm for solving the approximation problem may be simplified considerably.

(iv) *Modulating functions (MF):* As already mentioned in Section 3.1, the idea of the MF approach is to convert a differential equation involving I/O signals on a specified time interval into a sequence of algebraic equations. The idea was motivated by Laplace and Fourier transformation. This method offers two distinct advantages for linear as well as nonlinear CT systems over other identification methods: first, they allow for arbitrary initial conditions and, secondly, avoid approximation of time derivatives from noisy signals. For the identification of nonlinear CT systems two types of MFs have been proposed: Fourier MFs [21, 22, 163–167] and Hartley MFs [6–8, 168–175].

Both these classes of MFs are especially appropriate for the identification of integrable and convolvable types of nonlinear CT systems. Integrable nonlinear CT systems are described by models of the type

$$\sum_{i=0}^{n_1} \sum_{j=0}^{n_2} a_{ij} \frac{\mathrm{d}^i}{\mathrm{d}t^i}\{f_j[u(t), y(t)]\} = 0 \tag{180}$$

where $f_j(u, y)$ denotes a known differentiable function of u and y. Any of the coefficients a_{ij} may be set to unity for normalisation. In the absence of structural information regarding f_j, these can, for example, be chosen as multinomials in u and y. Several physical systems fall into this category. One example is the celebrated forced van der Pol's oscillator. As this category of models is directly integrable, all the methods of parameter estimation for linear systems can be used, albeit with the additional computation of f_j.

A further generalisation of this is given for convolvable nonlinear CT systems by the model

$$\sum_{i=0}^{n_1}\sum_{j=0}^{n_1}\sum_{k=0}^{n_1} a_{ijk} g_k[u(t), y(t)] \frac{d^i}{dt^i}\{f_j[u(t), y(t)]\} = 0 \quad (181)$$

Here, $g_k(u, y)$ is another known function of u and y. By using identities related to multiple derivatives of functions of u and y, many of the possible nonlinear terms involving these functions and their derivatives can be expressed in this form. This is the most general form of nonlinear differential equations that has till now been used for parameter estimation. Only the methods employing Fourier and Hartley modulating functions have been successful in solving this problem without the need to estimate unknown initial conditions [166, 168].

Thus, the initial step in identifying a nonlinear CT system is to rearrange a model in the form of (180) or (181), which can describe a large number of physical systems. Then it is possible to specify a cost function $J(\boldsymbol{\theta}) \geq 0$ using the FMF or HMF method for a given I/O-data set over the observation time interval $[0, T]$, and minimising $J(\boldsymbol{\theta})$ will lead to a one-shot or a newly developed batch scheme recursive LS estimation [21, 22, 173, 175]of the parameter vector $\hat{\boldsymbol{\theta}}$. The basic procedure in the case of HMFs is the introduction of the modulating function

$$\phi_m(t) = \sum_{i=0}^{n} (-1)^i \binom{n}{i} \operatorname{cas}(n+m-i)\omega_0 t, \quad 0 < t \leq T \quad (182)$$

where $\operatorname{cas}(\omega t) = \cos(\omega t) + \sin(\omega t)$, $m = 0, \pm 1, \ldots$ is referred to as the modulating frequency index, $\omega_0 = 2\pi/T$ which plays the role of a resolving frequency, and $\phi_m(t)$ is a member of the family of an nth-order HMF if $\phi_m(t)$ is sufficiently smooth, and the two-point boundary conditions given by

$$\begin{cases} \phi_m^{(l)}(t) \text{ exists} & \text{for all } l = 0, 1, \ldots, n-1 \text{ and} \\ \phi_m^{(l)}(t) = 0 & \text{for } t = 0 \text{ and } t = T \end{cases} \quad (183)$$

are satisfied for each m, where $\phi_m^{(l)}(t)$ is the lth derivative of a member of a family of MF $\{\phi_m(t)\}$ and n describes the highest derivative of the system under consideration. This modulating function is closely related to the FMF. Compared to the very efficient FMF method, the HMF method still has the important additional advantages that the HMFs are real-valued and the Hartley spectra can be computed efficiently with the help of fast algorithms for the Hartley integral transformation. This new methodology is applicable to a large class of nonlinear continuous-time systems by defining a set of HMFs for characterising the continuous process signals. The Hartley transform (HT) of a continuous signal $\xi(t)$ is defined by

$$\begin{aligned} H_\xi(\omega) &= \int_{-\infty}^{\infty} \xi(t)\operatorname{cas}(\omega t)dt \\ &= \int_{-\infty}^{\infty} \xi(t)\cos(\omega t)dt + \int_{-\infty}^{\infty} \xi(t)\sin(\omega t)dt \\ &= H_{\xi_{even}}(\omega) + H_{\xi_{odd}}(\omega) \end{aligned} \quad (184)$$

where $\operatorname{cas}(\omega t) = \cos(\omega t) + \sin(\omega t) = \sqrt{2}\sin(\omega t + \pi/4)$, $H_{\xi_{even}}(\omega)$ and $H_{\xi_{odd}}(\omega)$ are the even and odd parts of the HT of a signal $\xi(t)$, respectively. The relationship between the even and odd parts of the HT can be expressed as

$$\begin{bmatrix} H_{\xi_{even}}(\omega) \\ H_{\xi_{odd}}(\omega) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} H_\xi(\omega) \\ H_\xi(-\omega) \end{bmatrix} \quad \text{or}$$
$$\begin{bmatrix} H_\xi(\omega) \\ H_\xi(-\omega) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} H_{\xi_{even}}(\omega) \\ H_{\xi_{odd}}(\omega) \end{bmatrix} \quad (185)$$

The inverse HT is given by

$$\xi(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H_\xi(\omega)\operatorname{cas}(\omega t)d\omega \quad (186)$$

In general, modulation of a simple function or signal $\xi(t)$ consists in multiplying by a HMF $\phi_m(t)$ and integrating over $[0, T]$, i.e. $\int_0^T \xi(t)\phi_m(t)dt$; and this manipulation has coined the name Hartley spectrum of a function or signal $\xi(t)$. Thus, the mth HMF spectral component $\bar{H}_\xi(m\omega_0)$ of a simple function or signal $\xi(t)$ is defined by

$$\begin{aligned} \bar{H}_\xi(m\omega_0) &= \int_0^T \xi(t)\phi_m(t)dt \\ &= \sum_{j=0}^{n}(-1)^i\binom{n}{i}\int_0^T \xi(t)\operatorname{cas}[(n+m-i)\omega_0 t]dt \\ &= \sum_{j=0}^{n}(-1)^i\binom{n}{i} H_\xi[(n+m-i)\omega_0] \end{aligned} \quad (187)$$

If $\xi^{(\nu)}(t)$, $\nu = 1, 2, \ldots, n$ is the νth derivative of $\xi(t)$, then its corresponding Hartley modulated spectrum is given by

$$\begin{aligned} \bar{H}_\xi^{(\nu)}(m\omega_0) &= \sum_{i=0}^{n}(-1)^i\binom{n}{i}(n+m-i)^\nu \omega_0^\nu \operatorname{cas}'\left(\frac{\nu\pi}{2}\right) \\ &\quad \times H_\xi[(-1)^\nu(n+m-i)\omega_0] \end{aligned} \quad (188)$$

where $\operatorname{cas}'(\nu\pi/2) = \cos(\nu\pi/2) - \sin(\nu\pi/2)$, $\bar{H}_\xi^{(\nu)}(m\omega_0)$ is the mth HMF spectral component of the νth derivative of $\xi(t)$, and $H_\xi(m\omega_0)$ is the Hartley transform of $\xi(t)$. The Hartley spectrum of the νth derivative of the signal $\xi^{(\nu)}(t)$, $\nu = 1, 2, \ldots, n$ can be computed analytically by repeatedly applying integration by parts until all the derivatives of the signal shift to a known value of $\phi_m(t)$. Similar rules as in (187) and (188) can be established for products of functions and their derivatives, as for example $\xi_1(t)\xi_2^{(\nu)}(t)$ [168]; then the spectrum for such a product is given by

$$\begin{aligned} \bar{H}_{\xi_1,\xi_2}^{0,\nu}(m\omega_0) &= E_{\xi_1}(m\omega_0) * \bar{H}_{\xi_2}^{(\nu)}(m\omega_0) \\ &\quad + O_{\xi_1}(m\omega_0) * \bar{H}_{\xi_2}^{(\nu)}(m\omega_0) \\ &= H_{\xi_1}(m\omega_0) \otimes \bar{H}_{\xi_2}^{(\nu)}(m\omega_0) \end{aligned} \quad (189)$$

where

$$E_{\xi_1}(m\omega_0) = \frac{1}{2}[H_{\xi_1}(m\omega_0) + H_{\xi_1}(-m\omega_0)] \quad \text{and}$$

$$O_{\xi_1}(m\omega_0) = \frac{1}{2}[H_{\xi_1}(m\omega_0) - H_{\xi_1}(-m\omega_0)]$$

are the even and odd parts of $H_{\xi_1}(m\omega_0)$, respectively, and the operator $\otimes$ symbol represents the two convolutions in short form for simplicity.

For demonstration of the method, let us consider the following continuous-time nonlinear dynamic system with integrable and convolvable terms given by

$$\ddot{y}(t) = -d_1\dot{y}(t) - d_2 y(t) - d_3 y^2(t)\dot{y}(t) + d_4 y(t)u(t) + d_5 u(t)\dot{y}(t) + d_6 u(t) \qquad (190)$$

where d_i, $i = 1, 2, \ldots, 6$ are the parameters. Equation (190) contains integrable (the 3rd term) and convolvable (the 4th and 5th term of the right-hand side) nonlinear differential terms. The problem is to identify the parameters d_i based on the I/O-data records over a finite interval $T = NT_S$. Modulating (190) leads to the HMF model of the nonlinear system

$$\bar{H}_y^{(2)}(m\omega_0) = -d_1\bar{H}_y^{(1)}(m\omega_0) - d_2\bar{H}_y(m\omega_0) - d_3\frac{1}{2}\bar{H}_{y^3}^{(1)}(m\omega_0) + d_4 H_y(m\omega_0)\otimes\bar{H}_u(m\omega_0) + d_5 H_u(m\omega_0) \otimes\bar{H}_y^{(1)}(m\omega_0) + d_6\bar{H}_u(m\omega_0) \qquad (191)$$

which is linear in the parameters. Letting $z(m\omega_0) = \bar{H}_y^{(2)}(m\omega_0)$, assuming an equation error $\varepsilon(m\omega_0)$ and rearranging the terms of (191), it can be rewritten as a regression equation in the 'frequency' domain

$$z(m\omega_0) = \boldsymbol{\varphi}^{\mathrm{T}}(m\omega_0)\boldsymbol{\theta} + \varepsilon(m\omega_0) \qquad (192)$$

with

$$\boldsymbol{\varphi}^{\mathrm{T}}(m\omega_0) = [-\bar{H}_y^{(1)}(m\omega_0) - \bar{H}_y(m\omega_0) - \frac{1}{3}\bar{H}_{y^3}^{(1)}(m\omega_0)H_y(m\omega_0) \otimes\bar{H}_u(m\omega_0)H_u(m\omega_0)\otimes\bar{H}_y^{(1)}(m\omega_0)\bar{H}_u(m\omega_0)]$$

and

$$\boldsymbol{\theta}^{\mathrm{T}} = [d_1 \;\; d_2 \;\; d_3 \;\; d_4 \;\; d_5 \;\; d_6]$$

Let a sequence of observations be made for $m = 0, \pm 1, \ldots, \pm M$, then $(2M + 1)$ regression equations can be represented as a vector equation and minimising a frequency-weighted cost function with respect to the unknown parameter vector $\boldsymbol{\theta}$ leads to the frequency-weighted-least-squares (FWLS) estimate. Introducing the RMS normalised error

$$\|\Delta\boldsymbol{\theta}\| = \left\{\frac{1}{n_\theta}\sum_{i=1}^{n_\theta}[(\hat{\theta}_i - \theta_{i\text{true}})/\theta_{i\text{true}}]^2\right\}^{1/2} \qquad (193)$$

where n_θ is the number of estimated parameters and taking $M = 12$, $N = 1025$, $T = 0.16$ s and $u(t) = 0.25(\cos 0.4\pi t + \sin 0.2\pi t)$, then, even for a superimposed output noise with NSR = 60%, the value of the RMS is below 0.1, which shows the robustness of this approach, which had been considerably improved in [169–175] in comparison to the early investigations in [168] by computational changes.

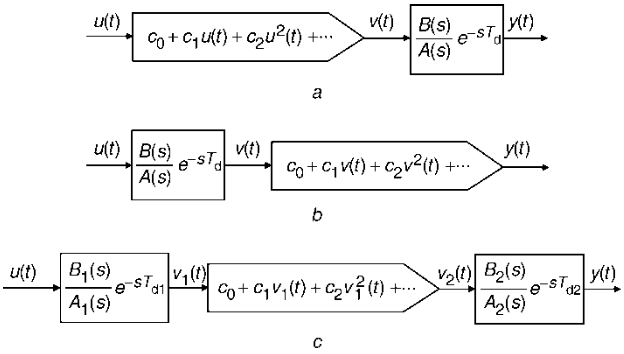


Fig. 18 *Schemes of undisturbed simple block-oriented CT models*

a Hammerstein
b Wiener
c Wiener-Hammerstein

9.2.2 Special parametric nonlinear models: Under this category of models we will briefly discuss block-oriented nonlinear models and bilinear models.

(*a*) *Block-oriented models:* These models are characterised by a blockwise separation into linear dynamic and nonlinear static partial systems [176]. Mainly three basic stuctures, presented in Fig. 18, are of interest:

(i) Hammerstein models,
(ii) Wiener models, and
(iii) Wiener-Hammerstein models.

These models may include a manifold of disturbances of deterministic or stochastic type, which are not shown in Fig. 18.

Even though these models have been applied widely in the identification of DT systems the corresponding modelling techniques have not become very popular for CT systems. There are only a few identification approaches dealing with CT types of Hammerstein [171] and Wiener systems [177]. In the first case, the HMF approach has been applied and, in the second case, the resulting identification problem is a nonparametric one, i.e. the impulse response of the linear subsystem is recovered by a correlation method, while the nonlinear characteristic is estimated through nonparametric kernel regression.

(*b*) *Bilinear model structure:* In the past years, much attention has been focused to the identification of bilinear systems using DT approaches. Compared to these efforts, there has been relatively less work done for the parameter identification of CT bilinear model structures, using data records of I/O signals. A time-invariant bilinear SISO system can be represented in the state space form as

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{b}u(t) + \boldsymbol{N}\boldsymbol{x}(t)u(t) \qquad (194)$$

$$y(t) = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}(t) + \varepsilon(t) \qquad (195)$$

which, by replacing the state variables $x_i(t)$ by their corresponding I/O signals, can easily be written in a I/O-differential standard form, which allows directly the application of the HMF method for identifying the system parameters involved in $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{c}^{\mathrm{T}}$ and $\boldsymbol{N}$ [172, 175]. Another approach based on an integration method is described in [158].

9.3 Signal-dependent parameter (SDP) models

Very often in engineering practice it is not completely clear whether the dynamic behaviour of a technical system is actually nonlinear or can also be described by a linear model with signal dependent parameters. Such a dependence on external (measurable) signals, characterised by the signal vector $\boldsymbol{s}(t)$, can occur, for example, at set-point changes, load changes, changes of raw materials or new initial conditions etc. of the plant. The continuous-time plant model can be described by the time-variable parameter (TVP) structure

$$y(t) = \hat{y}[\boldsymbol{s}(t)] + n(t)$$

where

$$\hat{y}[\boldsymbol{s}(t)] = -\sum_{i=1}^{n} a_i[\boldsymbol{s}(t)] \frac{\mathrm{d}^i}{\mathrm{d}t^i} y(t) + \sum_{j=0}^{n} b_j[\boldsymbol{s}(t)] \frac{\mathrm{d}^j}{\mathrm{d}t^j} u(t)$$

and the noise signal is assumed to be obtained by filtering a $(0, \sigma_0)$-distributed white noise signal $\varepsilon(t)$ by the filter transfer function $1/\sum_{i=1}^{n} a_i(\boldsymbol{s})s^i$. If $\boldsymbol{s}(t)$ actually constitutes a stochastic variable, then the system is truly nonlinear and likely to exhibit severe nonlinear behaviour, which cannot be approximated in a simple TVP manner [102]. Then a more powerful SDP modelling method must be applied [178] involving the nonparametric identification of the signal dependency using recursive methods of TVP estimation which allow for rapid parametric change. These methods are based on fixed-interval-smoothing (FIS) algorithms which provide the signal dependencies in graphical form. Parametrisation of these graphs can be obtained finally by curve filtering based on LS methods or using neural or neurofuzzy networks. The identified structural form of the model can be converted into a nonlinear stochastic state-space form. The final constant parameter estimation is obtained by applying LS or ML methods and prediction error decomposition leading finally to a parametrically efficient nonlinear model representation.

Finally, it should be mentioned that much work has still to be invested into the enlargement of the existing toolboxes for the identification of nonlinear systems [179].

10 A comparative study of DT and CT methods

Several studies assessed a number of CT methods [180–182], [184–186]. In the recent years, two fundamentally different time-domain approaches to the problem of obtaining a CT model of a naturally CT system from its sampled input/output data are compared:

(i) *indirect approach:* this involves two stages. At first a DT model for the original CT system is obtained by applying the well-established DT methods and the DT model is then transformed into CT form;
(ii) *direct approach:* in this approach a CT model is obtained straightaway using well-known CT methods

The results of extensive comparative studies [180, 181] using the SID and CONTSID toolboxes are now available on a system where the following transfer function has been simulated with a pseudorandom-binary signal (PRBS) as input $u(t)$:

$$G_0(s) = \frac{1600 - 6400s}{1600 + 416s + 408s^2 + 5s^3 + s^4} \qquad (196)$$

Two types of input signals are considered: a PRBS of maximum length, respecting the zero-order hold assumption and multisine signals, respecting the bandlimited assumption. These signals are generated to excite the system in its bandwidth.

Multisine: The input signal is chosen as the sum of five sinusoidal signals

$$u(t) = \sin(t) + \sin(1.9t) + \sin(2.1t) + \sin(18t) + \sin(22t)$$

The observation time is set to $T = 75$ s. Owing to the two sampling periods, the input signal has 1500 or 7500 samples.

PRBS of maximum length: The characteristics of the signal, whose amplitude switches between -1 and $+1$, are the following: the number of stages of the shift register is set to $n_s = 10$, the clock period is set to $n_p = 10$, which makes a number of points $N = 7161$ for a sampling period setting of 10 ms. With $n_s = 9$ and $n_p = 3$, we have $N = 1533$ for a sampling period setting of 50 ms.

10.1 Types of measurement noise

The data generating system is given by the following relations:

$$y_u(t) = G_0(p)u(t)$$
$$y(t_k) = y_u(t_k) + v(t_k)$$

The following types of measurement noise $v(t_k)$ are considered:

(*a*) DT white noise

$$v(t_k) = e(t_k)$$

(*b*) DT coloured noise (ARMA process noise)

$$v(t_k) = \frac{0.2236q^{-1} - 0.1630q^{-2}}{1 - 1.8906q^{-1} + 0.9512q^{-2}} e(t_k)$$

where q^{-1} is the backward shift operator, and $e(t_k)$ is a zero-mean independent identically distributed (IID) Gaussian sequence. The variance σ_e^2 of $e(t_k)$ is adjusted to obtain a desired signal-to-noise ratio ($SNR = 10$ or 0 dB). The SNR is defined as

$$SNR = 10 \log \frac{P_{y_u}}{P_v}$$

where P_v represents the average power of the zero-mean additive noise on the system output (e.g. the variance), while P_{y_u} denotes the average power of the noise-free output fluctuations.

Two sets of input-output data are gathered, one with sampling time $T_s = 50$ ms and another with $T_s = 10$ ms. The latter sampling rate is rapid relative to the critical frequencies of the simulated system. In a practical identification scenario in which the critical frequencies of the system are not usually well known, the tendency to sample input output data at a high rate is natural in the interest of preserving the signal content.

10.2 Identification techniques

The CT model identification is carried out with the help of the following techniques provided in the CONTSID toolbox:

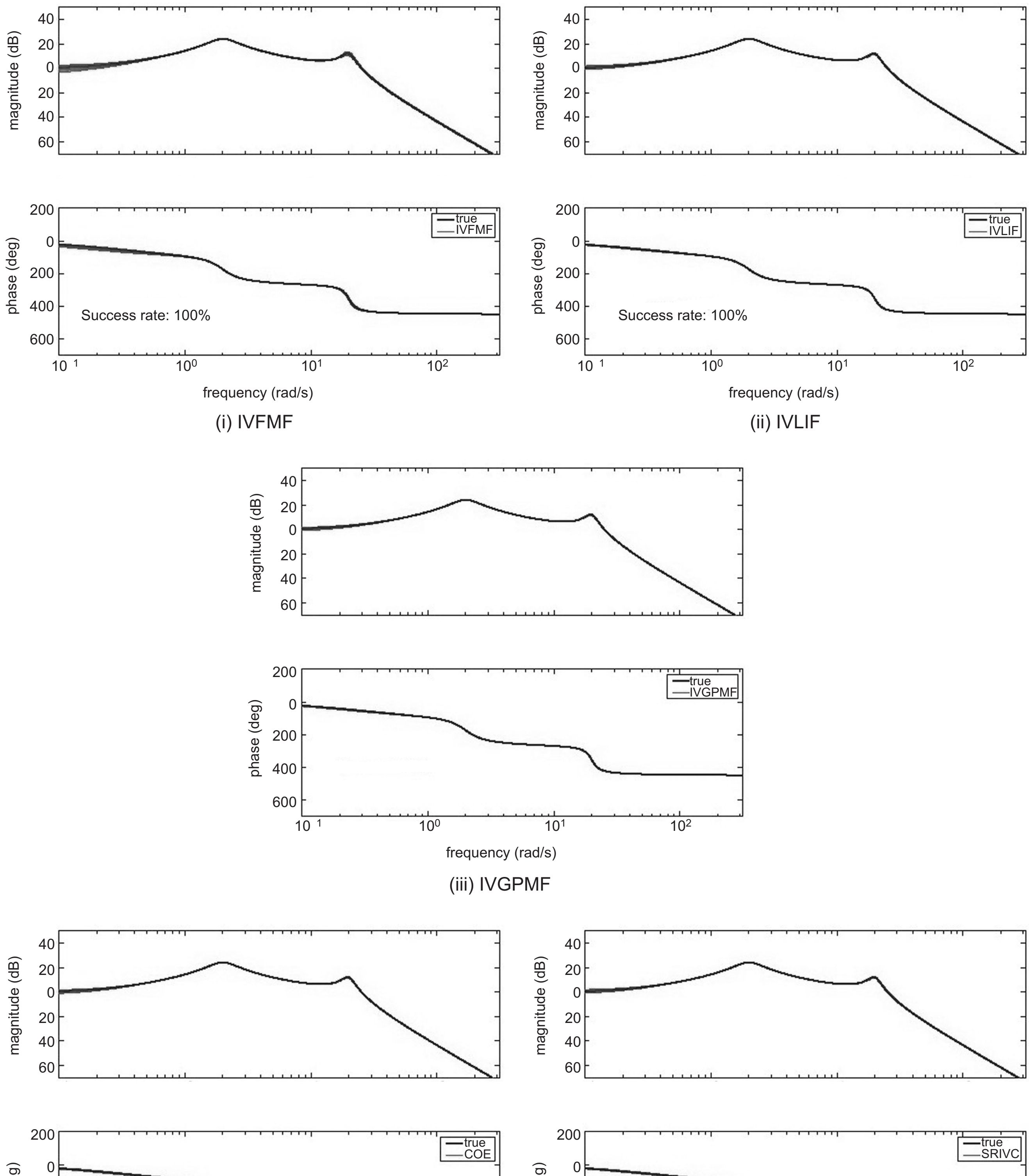
(*a*) generalised Poisson moment functionals (GPMF),
(*b*) Fourier modulating functions (FMF), and
(*c*) linear integrating filter (LIF).

The DT model identification was carried out with the help of the SID toolbox using IV4, N4SID, OE and PEM methods. The number of experiment N_{exp} is set to 200 in each trial which involves noisy data. The results are

Table 3: Summary of results of Monte Carlo simulation

Simulation conditions			SID toolbox methods				
T_s	Input	Noise	Criterion	IV4	N4SID	OE	PEM
10 ms	PRBS	White 10 dB	η_i	33	63	92	92
			MSE	7.7	1.0×10^{1}	1.4	7.7
			σ_{SE}	2.4	6.3	2.6	2.6
			$E_{\hat{G}_i}$	5.60×10^{4}	4.21×10^{4}	1.64×10^{4}	4.31×10^{4}
			$E_{\hat{\phi}_i}$	1.88×10^{7}	1.80×10^{7}	3.55×10^{6}	1.77×10^{7}
	Multisine	White 10 dB 8&9	η_i	13	52	97	99
			MSE	1.3×10^{2}	3.1×10^{3}	1.7×10^{1}	6.9×10^{4}
			σ_{SE}	1.3×10^{2}	1.3×10^{4}	4.7×10^{1}	9.7×10^{5}
			$E_{\hat{G}_i}$	3.98×10^{5}	4.38×10^{5}	1.76×10^{5}	3.56×10^{5}
			$E_{\hat{\phi}_i}$	2.10×10^{6}	6.69×10^{5}	2.22×10^{6}	7.42×10^{6}
		White 0 dB	η_i	16	60	98	99
			MSE	1.7×10^{2}	1.1×10^{3}	1.4×10^{9}	1.2×10^{5}
			σ_{SE}	9.7×10^{1}	2.0×10^{3}	2.010	1.7×10^{6}
			$E_{\hat{G}_i}$	4.28×10^{5}	6.28×10^{5}	3.75×10^{5}	3.85×10^{5}
			$E_{\hat{\phi}_i}$	1.89×10^{7}	1.16×10^{6}	1.47×10^{7}	9.60×10^{6}
		Coloured 10 dB	η_i	20	80	100	100
			MSE	9.0×10^{1}	1.2×10^{2}	2.9	5.4×10^{2}
			σ_{SE}	1.6×10^{1}	2.2×10^{1}	2.2×10^{1}	7.5×10^{3}
			$E_{\hat{G}_i}$	2.72×10^{5}	2.93×10^{5}	7.75×10^{4}	1.21×10^{5}
			$E_{\hat{\phi}_i}$	5.37×10^{6}	1.96×10^{6}	6.92×10^{5}	6.95×10^{5}
		Coloured 10 dB	η_i	29	81	98	99
			MSE	1.1×10^{2}	1.2×10^{2}	6.5	1.4×10^{1}
			σ_{SE}	8.3×10^{1}	1.5×10^{1}	2.9×10^{1}	3.7×10^{1}
			$E_{\hat{G}_i}$	3.27×10^{5}	3.17×10^{5}	1.01×10^{5}	1.96×10^{5}
			$E_{\hat{\phi}_i}$	5.39×10^{6}	3.40×10^{6}	8.40×10^{5}	1.55×10^{6}
50 ms	PRBS	White 10 dB	η_i	48	100	94	92
			MSE	2.6×10^{1}	5.8	2.4	1.1×10^{1}
			σ_{SE}	2.8×10^{1}	1.2×10^{1}	4.8	8.1
			$E_{\hat{G}_i}$	2.09×10^{4}	2.95×10^{3}	2.67×10^{3}	1.15×10^{4}
			$E_{\hat{\phi}_i}$	1.64×10^{7}	2.22×10^{5}	3.64×10^{6}	1.19×10^{7}
	Multisine	White 10 dB	η_i	49	69	97	99
			MSE	4.8×10^{1}	1.0×10^{1}	2.5	6.4×10^{1}
			σ_{SE}	2.3×10^{1}	1.7×10^{1}	1.1×10^{1}	6.2×10^{1}
			$E_{\hat{G}_i}$	6.69×10^{4}	2.79×10^{4}	1.16×10^{4}	6.16×10^{4}
			$E_{\hat{\phi}_i}$	9.50×10^{5}	3.32×10^{5}	4.38×10^{5}	1.66×10^{6}

Simulation conditions			CONTSID toolbox methods				
T_s	Input	Noise	IVFMF	IVLIF	IVGPMF	COE	SRIVC
10 ms	PRBS	White 10 dB	100	100	100	100	100
			1.7×10^{-1}	4.7×10^{-2}	3.0×10^{-3}	8.4×10^{-4}	8.4×10^{-4}
			1.0×10^{-1}	1.6×10^{-2}	2.0×10^{-3}	4.5×10^{-4}	4.5×10^{-4}
			5.61×10^{2}	3.78×10^{2}	12.1	2.41	2.41
			1.82×10^{5}	1.70×10^{4}	2.89×10^{2}	56.6×10^{1}	56.6
	Multisine	White 10 dB 8&9	100	100	100	100	100
			1.9×10^{-1}	1.2×10^{-1}	1.4×10^{-1}	1.3×10^{-1}	6.3×10^{-2}
			8.3×10^{-2}	3.9×10^{-2}	3.8×10^{-2}	3.7×10^{-2}	1.3×10^{-2}
			29.3	12.3	7.99	7.37	13.1
			6.23×10^{2}	2.65×10^{2}	1.70×10^{2}	1.64×10^{2}	3.01×10^{2}
		White 0 dB	100	100	100	100	100
			1.0	2.1	4.2×10^{-1}	3.4×10^{-1}	2.7×10^{-1}
			5.5×10^{-1}	3.5	2.0×10^{-1}	1.6×10^{-1}	1.3×10^{-1}
			2.92×10^{2}	9.96×10^{2}	90.9	72.6	77.1
			6.36×10^{3}	8.58×10^{3}	2.09×10^{3}	1.79×10^{3}	1.84×10^{3}
		Coloured 10 dB	100	100	100	100	100
			3.3×10^{-1}	2.9×10^{-1}	3.2×10^{-1}	2.1×10^{-1}	1.7×10^{-1}
			2.5×10^{-1}	1.8×10^{-1}	2.3×10^{-1}	1.1×10^{-1}	1.2×10^{-1}
			41.4	66.6	61.2	21.0	30.4
			9.46×10^{2}	7.81×10^{2}	7.77×10^{2}	4.36×10^{2}	6.86×10^{2}
		Coloured 10 dB	93	92	96	100	100
			1.5×10^{1}	1.2×10^{1}	6.0	1.9	1.5
			9.9×10^{1}	2.7×10^{1}	2.2×10^{1}	6.2	1.4
			2.49×10^{3}	4.65×10^{3}	1.68×10^{3}	5.76×10^{2}	5.67×10^{2}
			2.81×10^{4}	8.00×10^{4}	1.53×10^{4}	1.48×10^{4}	1.04×10^{4}
50 ms	PRBS	White 10 dB	100	100	100	100	100
			4.4	1.0	4.5×10^{-2}	8.4×10^{-3}	8.4×10^{-3}
			1.6	2.1×10^{-1}	2.3×10^{-2}	4.8×10^{-3}	4.8×10^{-3}
			4.10×10^{3}	7.92×10^{2}	80.0	14.8	14.8
			9.63×10^{5}	3.38×10^{5}	2.20×10^{3}	3.93×10^{2}	3.93×10^{2}
	Multisine	White 10 dB	100	100	100	100	100
			2.5	2.2	2.2	2.5	9.1×10^{-1}
			5.9×10^{-1}	7.5×10^{-1}	3.5×10^{-1}	3.1×10^{-1}	6.4×10^{-2}
			1.49×10^{2}	3.55×10^{2}	52.1×10^{1}	40.7	2.34×10^{2}
			3.94×10^{3}	1.13×10^{4}	1.34×10^{3}	9.38×10^{2}	8.48×10^{3}



(i) IVFMF

(ii) IVLIF

(iii) IVGPMF

(iv) COE

(v) SRIVC

(*a*)

Fig. 19 *Continued*

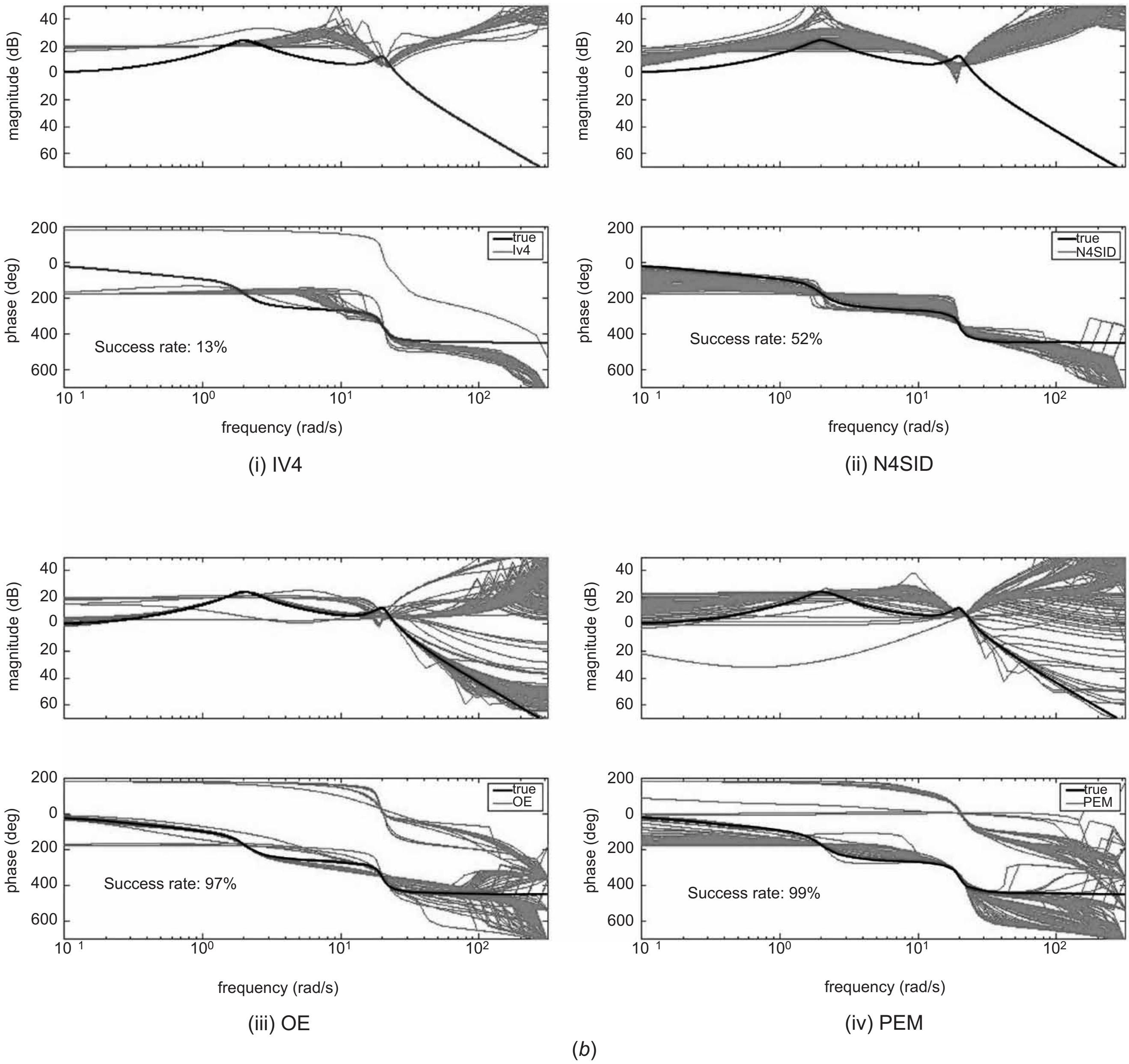


Fig. 19 *Bode plots of identified models*
(*a*) Direct methods
(*b*) Indirect methods

summarised in Table 3 where η_i = percentage of successful runs (i.e. where the identified model is stable):

$$MSE = \frac{1}{N_{\exp}} \sum_{i=1}^{N_{\exp}} SE(i); \quad SE = \frac{1}{N} \sum_{k=1}^{N} \varepsilon^2(t_k);$$

$$\varepsilon(t_k) = y_u(t_k) - \hat{y}_u(t_k);$$

$$\sigma_{SE}^2 = \frac{1}{N_{\exp}} \sum_{i=1}^{N_{\exp}} (SE(i) - MSE)^2;$$

$$E_{\hat{G}_i} = \frac{1}{N_{\exp}} \sum_{i=1}^{N_{\exp}} \sum_{k=1}^{N_\omega} (G_0(\omega_k) - G_i(\omega_k))^2;$$

$$E_{\hat{\phi}_i} = \frac{1}{N_{\exp}} \sum_{i=1}^{N_{\exp}} \sum_{k=1}^{N_\omega} (\phi_0(\omega_k) - \phi_i(\omega_k))^2$$

Fig. 19 shows the Bode plots of the identified models.

10.3 Salient observations

(i) The stability rate SR shows that the direct methods are highly reliable, as all the estimated models are stable. On the other hand, the indirect methods result in a high percentage of models that are unstable, as is evident from the performance of IV4 and N4SID algorithms, and the situation is aggravated by rapidly sampled data. Even if the estimated DT model is stable, it has a higher AMSE value, i.e. the estimated model differs significantly from the actual system.
(ii) All the three CT methods in the direct approach required the same computational effort as the IV4 algorithm of the indirect approach. In the other cases of indirect estimation the computational time is 3–5 times higher.
(iii) Even in experiments using normally sampled data and additive noise, it is clear that the stable estimated DT models deviate considerably from the actual system.
(iv) The methods of the indirect approach were improved when the data were decimated and prefiltered. However, the direct methods showed still better performance.

(v) In the multisine input case the performance of the indirect methods improves only when the number of sinusoidal components increases, this is understandable because of the larger number of parameters of the DT model in the indirect approach.

For many people working in the field of system identification, the choice between the direct and indirect approaches may seem trivial, but this study clearly shows that this is not so; the direct approach outperforms the indirect one in many respects. While those who work mainly with DT methods seem to be of the conviction that the indirect approach cannot be inferior to the direct one, a follow-up investigation [183] confirmed the results of [180] in a discussion of the initialisation aspects of the indirect methods which are related to the poor performance of the indirect methods.

11 Conclusion

This paper has attempted to provide a continuous-time perspective of the problem of system identification. The focus is particularly on lumped linear and nonlinear models and on those developments that followed earlier surveys by other authors, especially for linear [184] and nonlinear [144] CT models and by the present authors themselves [5–8]. The various CT approaches have been outlined in a unified framework and the significance of the CT models of physical systems has been discussed, in general, and with respect to control engineering applications, in particular. A generalised framework for linear estimation that is based on Markov parameters and time moments has been presented. Recent developments in identification of nonlinear systems are also surveyed. It is hoped that these will soon be included in the respective toolboxes.

A summary of the results of the identification experiment with a simulated model using some DT and CT methods is presented. These results suggest that the DT model estimation may be adequate in some situations, but, in general, if the conditions of the identification experiment are not adequately in favour of DT methods, the results may not be reliable in the sense that the resultant models may be unstable, or even if they are stable they may not be accurate. Rapid sampling, which is natural for several reasons, accentuates these phenomena. Estimation of CT models is free from all these problems and it assures stable and more accurate models, particularly with rapidly sampled data. The CT and DT methods must therefore be integrated, complementing each other to provide a wider set of comprehensive tools with greater choice of options to the system identification community, to assure dependable methods and acceptable results in a wide variety of circumstances. It is time now to take appropriate steps towards integration of the DT identification tools [180] with those of CT identification [185, 186] leading to a comprehensive tool kit for identification of lumped linear and nonlinear systems. It will be even more desirable to include the various methods for distributed parameter systems to render the toolbox truly comprehensive, and the authors hope that this will take place in the future.

12 Acknowledgment

The authors would like to thank the referees for their constructive comments. G.P. Rao is grateful to the A. von Humboldt Foundation for the support to his visit to Ruhr University, Bochum, in September 2004, to complete the writing of this contribution.

13 References

1 Aström, K.J., and Eykhoff, P.: 'System identification—a survey', *Automatica*, 1971, **7**, (2), pp. 123–162
2 Ljung, L.: 'System identification: theory for the user' (Prentice-Hall, Upper Saddle River, NJ, 1999)
3 Söderström, T., and Stoica, P.: 'System identification' (Prentice-Hall, Englewood Cliffs, NJ, 1989)
4 Young, P.C.: 'Parameter estimation of continuous-time models', *Automatica*, 1981, **17**, (1), pp. 23–39
5 Unbehauen, H., and Rao, G.P.: 'Continuous-time approaches to system identification—a survey', *Automatica*, 1990, **26**, (1), pp. 23–35
6 Unbehauen, H., and Rao, G.P.: 'Identification of continuous-time systems—a tutorial'. Proc. 11th IFAC Symposium on Identification (SYSID'97), Fukuoka, Japan, July 1997, pp. 1023–1049
7 Unbehauen, H., and Rao, G.P.: 'A review of identification in continuous systems', *IFAC Annu. Rev. Control*, 1998, **22**, pp. 145–171
8 Unbehauen, H., and Rao, G.P.: 'Identification of continuous systems—a survey', *Syst. Anal. Model. Simul.*, 1998, **33**, pp. 99–155
9 Saha, D.C., and Rao, G.P.: 'Identification of continuous dynamical systems—the Poisson moment functional (PMF) approach'. LNCIS, **56** (Springer Verlag, Berlin, 1983)
10 Unbehauen, H., and Rao, G.P.: 'Identification of continuous systems' (North Holland, Amsterdam, 1987)
11 Sinha, N.K., and Rao, G.P.: 'Identification of continuous systems—methodology and computer implementation', *in* Sinha, N.K., and Rao, G.P. (Eds.) (Kluwer, Dordrecht, 1991)
12 Rao, G.P.: 'Piecewise constant orthogonal functions and their application to systems and control'. LNCIS, **55** (Springer Verlag, Berlin, 1983)
13 Datta, K.B., and Mohan, B.M.: 'Orthogonal functions in systems and control' (World Scientific, Singapore, 1994)
14 Patra, A., and Rao, G.P.: 'General hybrid orthogonal functions and their applications in systems and control'. LNCIS, **213** (Springer Verlag, Berlin, 1996)
15 Zadeh, L.A.: 'From circuit theory to system theory', *Proc. Inst. Radio Eng.*, 1962, **50**, pp. 856–865
16 Goodwin, G.C., and Middleton, R.H.: 'Digital control and estimation: a unified approach' (Prentice-Hall, NJ, 1990)
17 Young, P.C., Chotai, A., and Tych, W.: 'Identification, estimation and control of continuous-time systems described by delta operator models', *in* Sinha, N.K., and Rao, G.P. (Eds.): 'Identification of continuous-time systems' (Kluwer, Dordrecht, 1991), pp. 363–418
18 Shinbrot, M.: 'On the analysis of linear and nonlinear systems', *Trans. ASME*, 1957, **79**, pp. 547–552
19 Pearson, A.E., Shen, Y., and Klein, V.: 'Application of Fourier modulating functions to parameter estimation of a multivariable linear differential system'. Proc. 10th IFAC Symp. Syst. Identification (SYSID'94), Copenhagen, Denmark, 1994, pp. 49–54
20 Skantze, F.P., Pearson, A.E., and Klein, V.: 'Parameter identification for unsteady aerodynamic systems by the modulating function technique'. Proc. 10th IFAC Symp. Syst. Identification (SYSID'94), Copenhagen, Denmark, 1994, pp. 571–576
21 Co, T.B., and Ungarala, S.: 'Batch scheme recursive parameter estimation of continuous-time systems using the modulating function method', *Automatica*, 1997, **33**, (6), pp. 1185–1191
22 Ungarala, S., and Co, T.B.: 'Time-varying system identification using modulating functions and spline models with applications to bio-processes', *Comput. Chem. Eng.*, 2000, **24**, pp. 2739–2753
23 Daniel-Berhe, S., and Unbehauen, H.: 'Batch scheme recursive Hartley modulating functions identification of nonlinear continuous-time Hammerstein model'. Proc. European Control Conf. (ECC-99), Karlsruhe, Germany, 1999, paper 1013-3
24 Fairman, F.W., and Shen, D.W.C.: 'Parameter identification for linear time-varying dynamic processes', *Proc. Inst. Electr. Eng.*, 1970, **117**, pp. 2025–2029
25 Fairman, F.W.: 'Parameter identification for a class of multivariable nonlinear processes', *Int. J. Syst. Sci.*, 1971, **1**, (3), pp. 291–296
26 Mathew, A.V., and Fairman, F.W.: 'Identification in the presence of initial conditions', *IEEE Trans. Autom. Control*, 1972, **17**, pp. 394–396
27 Rao, G.P., and Sivakumar, L.: 'Identification of deterministic time-lag systems', *IEEE Trans. Autom. Control*, 1976, **21**, pp. 527–529

28 Saha, D.C., and Rao, G.P.: 'Identification of lumped linear systems in the presence of initial conditions via Poisson moment functionals', *Int. J. Control*, 1980, **31**, (4), pp. 637–644

29 Saha, D.C., and Rao, G P.: 'Identification of lumped linear time varying parameter systems via Poisson moment functionals', *Int. J. Control*, 1980, **32**, (4), pp. 709–721

30 Saha, D.C., and Rao, G.P.: 'Identification of lumped linear systems in the presence of small unknown time delays—the Poisson moment functional approach', *Int. J. Control*, 1981, **33**, (5), pp. 945–951

31 Saha, D.C., and Rao, G.P.: 'A general algorithm for parameter identification in lumped continuous systems—the Poisson moment functional approach', *IEEE Trans. Autom. Control*, 1982, **27**, (1), pp. 223–225

32 Rao, G.P., and Sivakumar, L.: 'Parameter identification in lumped linear continuous systems in a noisy environment via Kalman filtered Poisson moment functionals', *Int. J. Control*, 1982, **35**, (3), pp. 509–519

33 Saha, D.C., and Rao, G.P.: 'Transfer function matrix identification in MIMO systems via Poisson moment functionals', *Int. J. Control*, 1982, **35**, (4), pp. 727–738

34 Rao, G.P., Saha, D.C., Rao, T.M., Bhaya, A., and Aghoramurthy, K.: 'A microprocessor based system for on line parameter identification in continuous dynamical systems', *IEEE Trans. Ind. Electron.*, 1982, **29**, (3), pp. 197–201

35 Saha, D.C., Rao, B.B.P., and Rao, G.P.: 'Structure and parameter identification in linear continuous lumped systems—the Poisson moment functional approach', *Int. J. Control*, 1982, **36**, (3), pp. 477–491

36 Mukherjee, A.K., Saha, D.C., and Rao, G.P.: 'Identification of large scale distributed parameter systems—some simplifications in the multidimensional Poisson moment functional approach', *Int. J. Sys. Sci.*, 1983, **14**, (7), pp. 777–788

37 Bapat, V.N.: 'Some extensions to Poisson moment functional based estimation of continuous-time models of dynamical systems'. PhD Thesis, Indian Institute of Technology, Kharagpur, 1993

38 Garnier, H., Sibille, P., and Spott, T.: 'Influence of the initial covariance matrix on recursive LS estimation of continuous models via generalised Poisson moment functionals'. Proc. 10th IFAC Symp. Syst. Identification (SYSID'94), Copenhagen, Denmark, 1994, pp. 3669–3674

39 Garnier, H., Sibille, P., and Richard, A.: 'Recursive least-squares estimation of continuous models via generalised Poisson moment functionals'. Proc. VIIth European Signal Processing Conf. (EUSIPCO'94), Edinburgh, Scotland, 1994, pp. 518–521

40 Garnier, H., Sibille, P., Nguyen, H.L., and Spott, T.: 'A bias-compensating least-squares method for continuous-time system identification via Poisson moment functionals'. Proc. 10th IFAC Symp. Syst. Identification (SYSID'94), Copenhagen, Denmark, 1994, pp. 3675–3680

41 Garnier, H., Sibille, P., and Nguyen, H.L.: 'A new bias-compensating least-squares method for continuous-time MIMO system identification applied to a laboratory-scale process'. Proc. 3rd IEEE Conf. Control Appl. (CCA'94), Glasgow, Scotland, 1994, pp. 1711–1716

42 Garnier, H., Sibille, P., and Richard, A.: 'Continuous-time canonical state-space model identification via Poisson moment functionals'. Proc. 34th IEEE Conf. Decision Control (CDC'95), New Orleans, USA, 1995, pp. 3004–3009

43 Garnier, H., Sibille, P., Mensler, M., and Richard, A.: 'Pilot crane identification and control in presence of friction'. Proc. 13th IFAC World Congress, San Francisco, USA, 1996, **J**, pp. 107–112

44 Bastogne, T., Garnier, H., Sibille, P., and Mensler, M.: 'Applicability of Poisson moment functionals to subspace methods for continuous-time system identification'. Proc. 11th IFAC Symp. Syst. Identification (SYSID'97), Fukuoka, Japan, 1997, **4**, pp. 1665–1670

45 Garnier, H., Sibille, P., and Bastogne, T.: 'A bias-free least-squares parameter estimator for continuous-time state-space models'. Proc. 36th IEEE Conf. Decis. Control (CDC'97), San Diego, USA, 1997, pp. 1860–1865

46 Garnier, H., Gilson, M., and Zheng, W.X.: 'A bias-eliminated least-squares method for continuous-time model identification of closed-loop systems', *Int. J. Control*, 2000, **73**, (1), pp. 38–48

47 Bastogne, T., Garnier, H., and Sibille, P.: 'A PMF-based subspace method for continuous-time model identification—application to a multivariable winding process', *Int. J. Control*, 2001, **74**, (2), pp. 118–132

48 Diamessis, J.E.: 'A new method of determining the parameters of physical systems', *Proc. IEEE*, 1965, **53**, pp. 205–206

49 Diamessis, J.E.: 'On the determination of the parameters of certain nonlinear systems', *Proc. IEEE*, 1965, **53**, pp. 319–320

50 Diamessis, J.E.: 'A new method of determining the parameters of certain time-varying systems', *Proc. IEEE*, 1965, **53**, pp. 396–397

51 Clymer, A.B.: 'Direct system synthesis by means of computers', *Trans. Amer. Inst. Electr. Eng.*, 1959, **77**, (I), pp. 798–806

52 Sagara, S., and Zhao, Z.Y.: 'Numerical integration approach to on-line identification of continuous-time systems in the presence of measurement noise'. Proc. 8th IFAC/IFORS Symp. Identification Syst. Parameter Estimation, Beijing, PR China, 1988, pp. 441–446

53 Sagara, S., and Zhao, Z.Y.: 'Recursive identification of transfer function matrix in continuous systems via linear integral filter', *Int. J. Control*, 1989, **50**, (2), pp. 457–477

54 Sagara, S., Yang, Z.J., and Wada, K.: 'Recursive identification algorithms for continuous systems using an adaptive procedure', *Int. J. Control*, 1991, **53**, (2), pp. 391–409

55 Sagara, S., Yang, Z.J., and Wada, K.: 'Identification of continuous systems using digital low-pass filtering', *Int. J. Sys. Sci.*, 1991, **22**, (7), pp. 1159–1179

56 Sagara, S., and Zhao, Z.Y.: 'Filtering approaches to identification of continuous systems'. Proc. 9th IFAC/IFORS Symp. Identification Syst. Parameter Estimation, Budapest, Hungary, 1991, pp. 305–310

57 Sagara, S., Yang, Z.J., and Wada, K.: 'Identification of continuous systems from noisy sampled input-output data'. Proc. 9th IFAC/IFORS Symp. Identification Syst. Parameter Estimation, Budapest, Hungary, 1991, pp. 603–608

58 Zhao, Z.Y., Sagara, S., and Kumamaru, K.: 'On-line identification of time delay and system parameters of continuous systems based on discrete-time measurements'. Proc. 9th IFAC/IFORS Symp. Identification Syst. Parameter Estimation, Budapest, Hungary, 1991, pp. 721–726

59 Zhao, Z.Y., Sagara, S., and Tomizuka, M.: 'A new bias compensating LS method for continuous system identification in the presence of coloured noise', *Int. J. Control*, 1992, **56**, (6), pp. 1441–1452

60 Yang, Z.J., Sagara, S., and Wada, K.: 'Parameter identification based on the Steiglitz-McBride method from noisy input-output data', *Trans. Soc. Instrum. Control Eng.*, 1992, **28**, (12), pp. 1492–1494

61 Yang, Z.J., Sagara, S., and Wada, K.: 'Identification of continuous-time systems from sampled input-output data using bias eliminating techniques', *Control Theory Adv. Tech.*, 1993, **9**, (1), pp. 53–75

62 Rao, G.P., and Srinivasan, T.: 'Remarks on "Authors reply" to "Comments on the design of piecewise constant gains for optimal control"', *IEEE Trans. Autom. Control*, 1978, **23**, (4), pp. 762–763

63 Chakravarty, B., Mandyam, N., Mukhopadhyay, S., Patra, A., and Rao, G.P.: 'Real-time parameter estimation via block pulse functions'. Proc. SICE'89, Matsuyama, Japan, 1989, pp. 1095–1098

64 Wang, S.: 'Block pulse operator method for parameter identification problems in non-linear continuous systems', *Int. J. Syst. Sci.*, 1991, **22**, (12), pp. 2441–2455

65 Dhawan, R.K., Sahai, A., Nishar, D.V., and Rao, G.P.: 'Recursive estimation of Markov parameters in linear continuous-time SISO systems via block pulse functions'. Proc. 9th IFAC/IFORS Symp. Identification Syst. Parameter Estimation, Budapest, Hungary, 1991, **2**, pp. 1495–1500

66 Wang, S.-Y.: 'On the orthogonal functions approximation method of non-linear system identification and its application in biochemical processes', *Control Theory Adv. Technol.*, 1993, **9**, (1), pp. 39–52

67 Dastych, J., and Kueper, P.: 'Identifikation zeitkontinuierlicher Systeme mit Block-Puls-Funktionen und bilinearen Transformationen', *Automatisierungstechnik*, 1995, **43**, (1), pp. 24–30

68 Jiang, Z.H., and Schaufelberger, W.: 'Block pulse functions and their application in control systems' (Springer-Verlag, Berlin, 1992)

69 Corrington, M.S.: 'Solution of differential and integral equations with Walsh functions', *IEEE Trans. Circuit Theory*, 1973, **20**, (5), pp. 470–476

70 Chen, C.F., and Hsiao, C.H.: 'Time domain synthesis via Walsh functions', *Inst. Electr. Eng. Proc.*, 1975, **122**, (5), pp. 565–570

71 Rao, G.P., and Sivakumar, L.: 'System identification via Walsh functions', *Inst. Electr. Eng. Proc.*, 1975, **122**, (10), pp. 1160–1161

72 Rao, G.P., and Tzafestas, S.G.: 'A decade of piecewise constant orthogonal functions in systems and control', *Math. Comput. Simul.*, 1985, **27**, (5), pp. 389–407

73 Rao, G.P., Unbehauen, H., Mukhopadhyay, S., and Patra, A.: 'From calculus to algebra in models of continuous-time systems'. Proc. 9th IFAC/IFORS Symp. Identification Syst. Parameter Estimation, Budapest, Hungary, 1991, **2**, pp. 1340–1346

74 Mukhopadhyay, S., Patra, A., and Rao, G.P.: 'A new class of discrete-time models for continuous-time systems', *Int. J. Control*, 1992, **55**, (5), pp. 1161–1187

75 Söderstrom, T., Fan, H., Carlsson, B., and Bigi, S.: 'Least squares parameter estimation of continuous-time ARX models from discrete-time data', *IEEE Trans. Autom. Control*, 1997, **42**, (5), pp. 659–673

76 Eykhoff, P.: 'System identification' (Wiley, New York, 1974)

77 Diekmann, K., and Unbehauen, H.: 'Recursive identification of multiple-input multiple-output systems'. Proc. of 5th IFAC Symp. Identification Syst. Parameter Estimation, Darmstadt, Germany, 1979, pp. 423–429
78 Mukhopadhyay, S., Patra, A., and Rao, G.P.: 'Irreducible model estimation for MIMO systems', *Int. J. Control*, 1991, **53**, (1), pp. 223–253
79 Rao, G.P., Diekmann, K., and Unbehauen, H.: 'Parameter estimation in large-scale interconnected systems'. Proc. of 9th IFAC World Congress, Budapest, Hungary, 1984, **X**, pp. 118–122
80 Wahlberg, B., and Ljung, L.: 'Design variables for bias distribution in transfer function estimation', *IEEE Trans. Autom. Control*, 1986, **31**, (2), pp. 134–144
81 Wahlberg, B.: 'Laguerre and Kautz models'. Proc. 10th IFAC Symp. Identification Syst. Parameter Estimation (SYSID'94), Copenhagen, Denmark, 1994, **3**, pp. 1–12
82 Goodwin, G.C., Ninness, B., and Poor, V.: 'Choice of basis functions for continuous and discrete system modelling'. Proc. of 9th IFAC Symp. Identification Syst. Parameter Estimation, Budapest, Hungary, 1991, pp. 1179–1184
83 Dhawan, R.K., Sahai, A., Nishar, D.V., and Rao, G.P.: 'Recursive estimation of Markov parameters in linear continuous-time SISO systems via block-pulse functions'. Proc. of 9th IFAC Symp. Identification Syst. Parameter Estimation, Budapest, Hungary, 1991, pp. 1495–1500
84 Subrahmanyam, A.V.B., and Rao, G.P.: 'Identification of continuous-time SISO systems via Markov parameter estimation', *IEE Proc. D, Control Theory Appl.*, 1993, **140**, (1), pp. 1–10
85 Subrahmanyam, A.V.B., Saha, D.C., and Rao, G.P.: 'Identification of continuous-time MIMO systems via time moments', *Control Theory Adv. Technol.*, 1995, **10**, (4), pp. 1359–1378
86 Subrahmanyam, A.V.B., Saha, D.C., and Rao, G.P.: 'Irreducible continuous model identification via Markov parameter estimation', *Automatica*, 1996, **32**, (2), pp. 249–253
87 Rao, G.P., and Subrahmanyam, A.V.B.: 'Models in generalized MA form for identification of continuous-time systems', *in* Katayama, T., and Sugimoto, S. (Eds.): 'Statistical methods in control and signal processing' (Marcel Dekker, New York, 1996)
88 Gantmacher, F.R.: 'Applications of the theory of matrices' (Interscience Publishers, Inc., New York, 1959)
89 Wolovich, W.A.: 'Linear multivariable systems' (Springer-Verlag, New York, 1974)
90 Kalman, R.E.: 'Aspects of network and systems theory' (Holt, Reinhart and Winston, New York, 1970)
91 Tether, A.: 'Construction of minimum state-variable models from finite input-output data', *IEEE Trans. Autom. Control*, 1970, **15**, (4), pp. 427–436
92 Garnier, H., and Mensler, M.: 'CONTSID – A continuous-time system identification toolbox for MATLAB'. Proc. European Control Conf. (ECC-99), Karlsruhe, Germany, 1999, paper F-1013-6.
93 Mensler, M., Garnier, H., Richard, A., and Sibille, P.: 'Comparison of sixteen continuous-time system identification methods with the CONTSID toolbox'. Proc. European Control Conf. (ECC-99), Karlsruhe, Germany, 1999, paper F-1013-7.
94 Rucker, R.A.: 'Real-time system identification in the presence of noice'. Pepr. IEEE Western Electronics Convention, 1963, paper 2.9
95 Young, P.C.: 'In flight dynamic checkout—a discussion', *IEEE Trans. Aerospace Electron. Syst.*, 1964, **2**, pp. 1106–1111
96 Young, P.C.: 'The determination of the parameters of a dynamic process', *Radio Electron. Eng. J. IERE*, 1965, **29**, pp. 345–361
97 Young, P.C.: 'Process parameter estimation and self adaptive control.', *in* Hammond, P.H. (Ed.): 'Theory of self adaptive control systems' (Plenum Press, New York, 1965), pp. 118–140
98 Young, P.C.: 'Applying parameter estimation to dynamic systems: Part 2', *Control Eng.*, 1969, **16**, (11), pp. 118–124
99 Schaufelberger, W.: 'Modelladaptive systeme'. Dr.-Ing. Dissertation, ETH, Zürich, Switzerland, 1969
100 Isermann, R., Lachmann, K.-H., and Matko, D.: 'Adaptive control systems' (Prentice-Hall, NY, 1992)
101 Kohr, R.H., and Haberock, L.L.: 'An experimental determination of differential equations to describe simple nonlinear systems'. Proc. Joint Autom. Control Conf. (JACC), 1966, pp. 616–623
102 Young, P.C., and Jakeman, A.L.: 'Refined instrumental variable methods of recursive time-series analysis: Part III, Extensions', *Int. J. Control*, 1980, **31**, pp. 741–764
103 Young, P.C., Garnier, H., and Javis, A.: 'Identification of continuous-time linear and nonlinear models: A tutorial with environmental applications'. Proc. 13th IFAC Symp. Syst. Identification, Rotterdam, 2003, pp. 618–628
104 Young, P.C.: 'Comments on "On the estimation of continuous-time transfer functions"', *Int. J. Control*, 2002, **75**, pp. 693–697
105 Johansson, R.: 'Identification of continuous-time models', *IEEE Trans. Signal Process.*, 1994, **42**, (4), pp. 887–897
106 Unbehauen, H.: 'Adaptive model approaches', *in* Sinha, N., and Rao, G.P. (Eds.): 'Identification of continuous-time systems' (Kluwer Academic Publishers, Dordrecht, 1991)
107 Unbehauen, H.: 'Regelungstechnik II' (F. Vieweg & Sohn, Braunschweig, 2000, 8th edn.)
108 Pazdera, J.S., and Pottinger, H.J.: 'Linear system identification via Liapunov design techniques'. Proc. Joint Automatic Control Conf., Boulder, 1969, pp. 795–801
109 Schoukens, J., and Pintelon, R.: 'Identification of linear systems: a practical guideline to accurate modeling' (Pergamon Press, London, 1991)
110 Pintelon, R., Guillaume, P., Rolain, Y., Schoukens, J., and Van Hamme, H.: 'Parametric identification of transfer functions in the frequency domain—a survey', *IEEE Trans. Autom. Control*, 1994, **39**, (11), pp. 2245–2260
111 Fletcher, R.: 'Practical methods of optimization' (Wiley, New York, 1991)
112 Whitfield, A.: 'Transfer function synthesis using frequency response data', *Int. J. Contr.*, 1986, **43**, (5), pp. 1413–1425
113 Unbehauen, H.: 'Regelungstechnik III' (F.Vieweg & Sohn, Braunschweig, 2000, 6th edn.)
114 Anderson, B.: 'Identification of scalar errors-in-variables models with dynamics', *Automatica*, 1985, **21**, (6), pp. 709–716
115 Guillaume, P., Pintelon, R., and Schoukens, J.: 'Robust parametric transfer function estimation using complex logarithmic frequency response data'. Proc. 10th IFAC Symp. Syst. Identification, Copenhagen, Denmark, 1994, **2**, pp. 495–502
116 Schoukens, J., Pintelon, R., Vandersteen, G., and Guillaume, P.: 'Frequency-domain system identification using non-parametric noise models estimated from a small number of data sets', *Automatica*, 1997, **33**, (6), pp. 1073–1086
117 Schoukens, J., Pintelon, R., and Van Hamme, H.: 'Identification of linear dynamic systems using piecewise constant excitations: use, misuse and alternatives', *Automatica*, 1994, **30**, (7), pp. 1153–1169
118 Ljung, L.: 'Some results on identifying linear systems using frequency domain data'. Proc. 32rd IEEE Conf. on Decis. Control (CDC), San Antonio, TX, USA, 1993, pp. 3534–3538
119 Evans, C., Weiss, M., Escobet, T., Quevedo, J., Rees, D., and Jones, L.: 'Identification in the time and frequency domains: a comparison using the ECC bench-mark model'. Proc. UKACC Int. Conf. Control, University of Exeter, UK, 1966, pp. 1290–1296
120 Rolain, Y., Pintelon, R., Xu, K.Q., and Vold, H.: 'On the use of orthogonal polynomials in high order frequency domain system identification and its application to model parameter estimation'. Proc. 33rd IEEE Conf. on Decis. Control (CDC), Lake Buena Vista, FL, USA, 1994, pp. 3365–3373
121 Ljung, L.: 'System identification toolbox' (The Math-Works, Natick, MA, 1992)
122 Kollár, I.: 'Frequency domain system identification toolbox' (The Math-Works, Natick, MA, 1994)
123 Moonen, M., De Moor, B., and Vandevalle, J.: 'SVD-based subspace methods for multivariable continuous-time system identification', *in* Sinha, N., and Rao, G.P. (Eds.): 'Identification of continuous-time systems' (Kluwer Academic Publishers, Dordrecht, 1991)
124 Johansson, R., Verhaegen, M., Haverkamp, B., and Chou, C.T.: 'Correlation methods of continuous-time state-space model identification'. Proc. Europ. Control Conf. (ECC), Karlsruhe, Germany, 1999, paper F 1013-2.
125 Haverkamp, B., Verhaegen, M., Chou, C.T., and Johansson, R.: 'Continuous-time subspace model identification method using Laguerre filters'. Proc. 11th IFAC Symp. Identification, Kitakyushu, Japan, 1997, pp. 1143–1148
126 Haverkamp, R.: 'State space identification'. PhD Thesis, Technical University of Delft, The Netherlands, 2001
127 Ohsumi, A., Kameyama, K., and Yamaguchi, K.: 'Subspace-based identification for continuous-time stochastic systems via distribution-theoretic approach'. Proc. 14th IFAC World Congress, Beijing, China, 1999, pp. 223–228
128 De Moor, B., Vandevalle, J., Vanderberghe, L., and Van Mieghem, P.: 'A geometrical strategy for identification of state space models of linear multivariable systems with singular value decomposition'. Proc. IFAC-Symp. Identification Syst. Parameter Estimation, Beijing, China, 1988, pp. 700–704
129 Larimore, W.: 'Canonical variate analysis identification, filtering and adaptive control'. Proc. 29th IEEE-Conf. Decis. Control (CDC), Hawaii, 1990, pp. 596–604
130 Verhaegen, M.: 'A novel non-iterative MIMO state space model identification technique'. Proc. 9th IFAC/IFORS Symp. Identification Syst. Parameter Estimation, Budapest, Hungary, 1991, pp. 1453–1458

131 Verhaegen, M.: 'Subspace model identification, pt. 3: Analysis of the ordinary output-error state space model identification algorithm', *Intern. J. Control*, 1993, **58**, pp. 555–586
132 Van Overschee, P., and De Moor, B.: 'Subspace identifiction for linear systems' (Kluwer Academic Publishers, Dordrecht, 1996)
133 Haverkamp, B., Verhaegen, M., Chou, C.T., and Johansson, R.: 'Identification of continuous-time MIMO state space models from sampled data in the presence of process and measurement noise'. Proc. 35th IEEE Conf. Decis. Control (CDC), Kobe, Japan, 1966, pp. 1539–1544
134 Yang, Z., Sagara, S., and Wada, T.: 'Subspace model identification for continuous-time systems'. Proc. 11th IFAC Symp. Syst. Identification Parameter Estimation, Kitakyushu, Japan, 1997, pp. 1671–1676
135 Verhaegen, M.: 'Identification of the deterministic part of MIMO state space models given in innovations form from input-output data', *Automatica*, 1994, **30**, (1), pp. 61–74
136 Van Overschee, P., and De Moor, B.: 'N4SID: Subspace algorithms for the identification of combined deterministic stochastic systems', *Automatica*, 1994, **30**, (1), pp. 75–93
137 Balchen, J.: 'How have we arrived at the present state of knowledge in process control? Is there a lesson to be learned?', *J. Process Control*, 1999, **9**, pp. 101–108
138 Ljungquist, D., and Balchen, J.: 'Recursive prediction error methods for online estimation in nonlinear state-space models'. Proc. 32nd IEEE Conf. Decision and Control (CDC'93), San Antonio, USA, 1993, pp. 714–719
139 Unbehauen, H.: 'Some new trends in identification and modelling of nonlinear dynamical systems', *J. Appl. Math. Comput.*, 1996, **78**, pp. 279–297
140 Jedner, U., and Unbehauen, H.: 'Identification of a class of nonlinear systems by parameter estimation of a linear multi-model'. Proc. IMACS Symp. Model. Simul. Control Lumbed Distrib. Parameter Syst., Villeneuve, France, 1986, pp. 297–299
141 Pottmann, M., Unbehauen, H., and Seborg, D.: 'Application of a general multi-model approach for identification of highly nonlinear processes—a case study', *Int. J. Control*, 1993, **57**, pp. 97–120
142 Junge, T., and Unbehauen, H.: 'On-line identification of nonlinear systems using structurally adaptive radial basis functions'. Proc. 16th Am. Control Conf. (ACC'97), Albuquerque, NM, USA, 1997, pp. 1037–1041
143 Kortmann., P., and Unbehauen, H.: 'Structure identification of functional-type fuzzy models with application to modelling nonlinear dynamic plants', *in* Reusch, B. (Ed.): 'Computational intelligence'. LNCS 1226 (Springer Verlag, Berlin-Heidelberg, 1997)
144 Billings, S.A.: 'Identification of nonlinear systems—a survey', *IEE Proc. D, Control Theory Appl.*, 1980, **127**, (6), pp. 272–285
145 Wiener, N.: 'Nonlinear problems in random theory' (MIT Press, Cambridge, MA, 1958)
146 Hunter, I.W., and Nielson, P.M.: 'Nonlinear and time-varying identification of living muscle cell mechanics'. Proc. 11th IFAC World Congress, Tallin, USSR, 1990, pp. 104–107
147 Kurth, J.: 'Identifikation nichtlinearer Systeme mit komprimierten Volterra-Reihen' (VDI-Verlag, Düsseldorf, 1995)
148 Li, L.M., and Billings, S.A.: 'Continuous-time linear and nonlinear system identification in the frequency domain'. Research Report 711, Department Automatic Control and Systems Engineering, University of Sheffield, 1998
149 Billings, S.A., and Lang, Z.: 'Truncation of nonlinear system expansions in the frequency domain', *Int. J. Control*, 1997, **68**, (5), pp. 1019–1042
150 Rufer, D.F.: 'Implementation and properties of a method for identification of nonlinear continuous-time models'. Fachgruppe Automatik, ETH Zürich, Report 77-01, 1977
151 Lion, P.M.: 'Rapid identification of linear and nonlinear systems', *AIAA J.*, 1967, **5**, (10), pp. 1835–1842
152 Inoue, K., Kumamaru, K., Nakahashi, Y., Nakamura, N., and Uchida, M.: 'A quick identificatiuon method of continuous-time nonlinear systems and its application to power plant control'. Proc. 10th IFAC Symp. Syst. Identification Parameter Estimation (SYSID'94), Copenhagen, Denmark, 1994, pp. 319–324
153 Hu, Y.Z., and Davison, E.J.: 'A PMGOP method of parameter identification for a class of continuous nonlinear systems', *Int. J. Control*, 1995, **62**, (2), pp. 345–378
154 Read, K.N., and Ray, W.H.: 'Application of nonlinear dynamic analysis in the identification and control of nonlinear systems—I. Simple dynamics', *J. Process Control*, 1998, **8**, (1), pp. 1–15
155 Ljung, L., and Söderström, T.: 'Theory and practice of recusive identification' (MIT Press, Cambridge, MA, 1983)
156 Bohn, C., and Unbehauen, H.: 'Sensitivity models for nonlinear filters with application to recursive parameter estimation for nonlinear state-space models', *IEE Proc., Control Theory Appl.*, 2001, **148**, (2), pp. 137–145
157 Bohn, C., and Unbehauen, H.: 'The application of matrix differential calculus for derivation of simplified expressions in approximate nonlinear filter algorithms', *Automatica*, 2000, **36**, pp. 1553–1560
158 Coirault, P., Trigeassou, J.C., Gaubert, J.P., and Champenois, Q.: 'Parameter identification of an induction machine for diagnosis'. Proc. IFAC-Safe Process '97, Kingston upon Hull, UK, 1997, pp. 276–281
159 Kohr, R.H.: 'A method for determination of a differential equation model for simple nonlinear systems', *IEEE Trans. Electron. Comput.*, 1963, **12**, pp. 394–400
160 Tsang, K.M., and Billings, S.A.: 'Identification of continuous time nonlinear systems using delayed state variable filters', *Int. J. Control*, 1994, **60**, (2), pp. 159–180
161 Hachimo, T., Karube, I., Minari, Y., and Takata, H.: 'Continuous-time identification of nonlinear systems using radial basis function network model and genetic algorithm'. Proc. 12th IFAC Symp. Syst. Identification Parameter Estimation, Santa Barbara, CA, USA, 2000, paper 141
162 Wang, S.: 'Block pulse operator method for parameter identification problems in non-linear continuous systems', *Int. J. Syst. Sci.*, 1991, **22**, (12), pp. 2441–2455
163 Pearson, A.E., and Lee, F.C.: 'On the identification of polynomial input-output differential systems', *IEEE Trans. Autom. Control*, 1985, **30**, (8), pp. 778–782
164 Pearson, A.E., and Lee, F.C.: 'Parameter identification of linear differential systems via Fourier based modulating functions', *Control-Theory Adv. Technol. (C-TAT)*, 1985, **1**, pp. 239–266
165 Pearson, A.E.: 'Least squares parameter identification of nonlinear differential input-output models'. Proc. 27th IEEE Conf. Decis. Control (CDC), Austin, USA, 1988, pp. 1831–1835
166 Pearson, A.E.: 'Explicit parameter identification for a class of nonlinear input/output differential operator models'. Proc. 31st IEEE Conf. Decis. Control (CDC), Tuscon, USA, 1992, pp. 3656–3660
167 Pearson, A.E., Shen, Y., and Pan, J.Q.: 'Discrete frequency formats for linear differntial system identification'. Proc. 12th IFAC World Congress, Sydney, Australia, 1993, **VII**, pp. 143–148
168 Patra, A., and Unbehauen, H.: 'Identification of a class of nonlinear continuous-time systems using Hartley modulating functions', *Int. J. Control*, 1995, **62**, (6), pp. 1431–1451
169 Daniel-Berhe, S., and Unbehauen, H.: 'Parameter estimation of nonlinear continuous-time systems using Hartley modulating functions'. Proc. IEE UKACC Int. Conf. Control, Exeter, UK, 1996, pp. 228–233
170 Daniel-Berhe, S., and Unbehauen, H.: 'Physical parameter estimation of the nonlinear dynamics of a single link robotic manipulator with flexible joint using the HMF-method'. Proc. of 16th Am. Contr. Conf. (ACC), Albuquerque, NM, USA, 1997, pp. 1504–1508
171 Daniel-Berhe, S., and Unbehauen, H.: 'Identification of nonlinear continuous-time Hammerstein model via HMF-method'. Proc. of 36th IEEE Conf. Decis. Control (CDC), San Diego, CA, USA, 1997, pp. 2990–2995
172 Daniel-Berhe, S., and Unbehauen, H.: 'Bilinear continuous-time systems identification via Hartley-based modulating functions', *Automatica*, 1998, **34**, (4), pp. 499–503
173 Daniel-Berhe, S., and Unbehauen, H.: 'State space identification of bilinear continuous-time canonical systems via batch scheme Hartley modulating functions approach'. Proc. 37th Int. Conf. Decis. Control (CDC), Tampa, FL, USA, 1998, pp. 4482–4487
174 Daniel-Berhe, S., and Unbehauen, H.: 'Physical parameters estimation of the nonlinear continuous-time dynamics of a DC motor using HMF-method', *J. Franklin Inst.*, 1999, **336**, pp. 481–501
175 Daniel-Berhe, S.: 'Parameter identification of nonlinear continuous-time systems using the Hartley modulating functions method' (Cuvillier-Verlag, Göttingen, 1999)
176 Haber, R., and Keviczky, L.: 'Nonlinear system identification-input-output modelling approach. Vol. 1 & 2' (Kluwer, Dordrecht, The Netherlands, 1999)
177 Greblicki, W.: 'Continuous-time Wiener system identification', *IEEE Trans. Autom. Control*, 1998, **43**, (10), pp. 1488–1492
178 Young, P.C., McKenna, P., and Bruun, J.: 'Identification of non-linear stochastic systems by state dependent parameter estimation', *Int. J. Control*, 2001, **74**, pp. 1837–1857
179 N.N.: 'Identification of nonlinear systems toolbox for use with MATLAB'. Scientific Computers GmbH, 562064 Aachen, Germany, Franzstr. 107
180 Rao, G.P., and Garnier, H.: 'Numerical illustrations of the relevance of direct continuous-time model identification'. 15th Triennial IFAC

World Congress, Barcelona, Spain, 2002, CD-Rom publication (paper 2456)
181 Garnier, H., Mensler, M., and Richard, A.: 'Continuous-time model identification from sampled data. Implementation issues and performance evaluation', *Int. J. Control*, 2003, **76**, (13), pp. 1337–1357
182 Rao, G.P., and Garnier, H.: 'Identification of continuous-time systems: direct or indirect?'. Proc. 15th Int. Conf. Systeme Science, Wroclaw, Poland, 2004, **I**, pp. 66–86
183 Ljung, L.: 'Initialisation aspects for subspace and output-error identification methods'. Proc. European Control Conf., Cambridge, 2003, CD-Rom publication
184 Homssi, L., Titli, A., and Despujols, A.: 'Continuous-time process identification: comparison of eight methods and practical aspects'. Proc. 9th IFAC-Symp. Syst. Identification Parameter Estimation, Budapest, Hungary, 1991, **2**, pp. 1634–1642
185 Garnier, H., and Mensler, M.: 'Comparison of sixteen continuous-time system identification methods with the CONTSID toolbox'. Proc. European Control Conf. ECC'99, Karlsruhe, Germany, paper no. dm 3-6, 1999
186 Garnier, H., and Mensler, M.: 'The CONTSID toolbox: a matlab toolbox for continuous-time system identification'. Proc. 12th IFAC-Symp. Syst. Identification Parameter Estimation, Santa Barbara, CA, USA, 2000, pp. 851–856