An overview of system modeling and identification

Gérard Favier

An overview of system modeling and identification

Gérard Favier

Laboratoire I3S, University of Nice Sophia Antipolis, CNRS, Les Algorithmes- Bât. Euclide B,
2000 Route des lucioles, B.P. 121 - 06903 Sophia Antipolis Cedex, France
favier@i3s.unice.fr

1 Introduction

System identification consists in building mathematical models of dynamical systems from experimental data. Such a methodology was mainly developed for designing model-based control systems. More generally, parameter estimation is at the heart of many signal processing applications aiming to extract information from signals, like radar, sonar, seismic, speech, communication, or biomedical (EEG, ECG, EMG) signals. Nowadays, dynamical models and identification methods play an important role in most of disciplines such as automatic control, signal processing, physics, economics, medicine, biology, ecology, seismology, etc. In this plenary talk, an overview of the principal models and identification methods will be presented.

Since the pioneering works of Gauss and Legendre who introduced the least squares (LS) method, at the end of the eighteenth century, in the field of astronomy to predict the motion of planets and comets from telescopic measurements, quite a lot of papers and books on the parameter estimation problem have been published, so that it is not easy to get a general view of available model identification methods. Due to time limitation, some restrictive choices have been made as summarized below, with the attempt to have the best coverage as possible.

First, we will answer to the following basic questions:

- Why to use a model?
- How to classify the models?
- How to build a model?

Then, several model-based applications will be briefly discussed, with particular emphasis on simulation/synthesis, filtering, prediction/forecasting, interpolation/smoothing, source separation, adaptive equalization and adaptive control. It is important to note that depending on the considered application, a model is also called filter, predictor, interpolator, mixture, channel, equalizer or controller.

In a second part of the talk, we will present different classes of models.

Discrete-time deterministic linear models will be first considered, both in state-space (SS) and input-output (I/O) forms. Then, several discrete-time stochastic linear models will be presented in a unified way. These models can be viewed as

- either linear dynamical filters allowing the generation/synthesis, analysis and classification of random signals, as it is the case of the autoregressive (AR), moving average (MA) and ARMA models,
- or linear models with a random additive noise that represents measurement noise, external disturbances, and modeling errors, like ARX, ARMAX, and ARARX models.

Real-life systems being often time-varying or/and nonlinear in nature, two other classes of models are currently used: linear parameter-varying (LPV) and nonlinear (NL) models. These two classes will be described, with a focus on Volterra and block-structured (Wiener, Hammerstein, Wiener-Hammerstein...) models.

LPV models are suited to modeling linear time-varying (LTV) systems, the dynamics of which are functions of a measurable, time-varying parameter vector $p(t)$, called the scheduling variable. They can also be used for representing nonlinear systems linearized along the trajectory $p(t)$. LPV models can be viewed as intermediate descriptions between linear time invariant (LTI) models and nonlinear time-varying models.

Nonlinear models are very useful for various application areas including chemical and biochemical processes (distillation columns, polymerization reactors, biochemical reactors, see [10]), hydraulic plants, pneumatic valves, radio-over-fiber (RoF) wireless communication systems (due to the optical/electrical (O/E) conversion), high power amplifiers (HPA) in satellite communications, loudspeakers, active noise cancellation systems, physiological systems [44], vibrating structures and more generally mechatronic systems like robots [34].

Block oriented NL models are composed of a concatenation of LTI dynamic subsystems and static NL subsystems. The linear subsystems are generally parametric (transfer functions, SS representations, I/O models), whereas the NL subsystems may be with memory or memoryless. The different subsystems are interconnected in series or parallel.

In our talk, finite-dimensional discrete-time Volterra models, also called truncated Volterra series expansions, that allow to approximate any fading memory nonlinear system with an arbitrary precision, as shown in [5], will be considered in more detail; see [15], and [25]. The importance of such models for applications is due to the fact that they represent a direct nonlinear extension of the very popular finite impulse response (FIR) linear model, with guaranteed stability in the bounded-input bounded-output (BIBO) sense, and they have the advantage to be linear in their parameters, the kernel coefficients [53]. Moreover, they are interpretable in terms of multidimensional convolutions which makes easy the derivation of their z-transform and Fourier transform representations [45]. The main drawback of these models concerns their parametric complexity implying the need to estimate a huge number of parameters. So, several complexity reduction approaches for Volterra models have been developed using symmetrization or triangularization of Volterra kernels, or also their expansion on orthogonal bases (like Laguerre, Kautz and GOB bases), or their Parafac decomposition. These two last ap-

proaches lead to the so-called Volterra-Laguerre and Volterra-Parafac models; [6], [7], [18], [39], [40]

The third part of the talk will be concerned with the model identification problem which is strongly related to statistics, stochastic processes (also called random signals), time series analysis, signal processing, information/detection/estimation theories, optimization, linear algebra, machine learning, and more recently with compressive sensing, also known as compressive sampling and sparse sampling, that consists in finding sparse solutions to an underdetermined system of linear equations, i.e. characterized by more unknowns than equations [9]. The increasing interest in this last topic is due to the fact that most signals are sparse, i.e. with many coefficients close to or equal to zero, as it is the case of images for instance.

A complete identification procedure is composed of six main steps [42]:

- Experiment design,
- Input-output measurements,
- Model structure choice,
- Structure parameter determination,
- Model parameter estimation,
- Model validation.

This procedure is generally iterative in the sense that, starting from a priori information about the system to be identified, the different above-listed sub-problems are successively addressed with the need to revise some choices until the model is validated.

The experiment design consists in making several choices (optimal design of input sequence, i.e. excitation signal, sampling period, etc.).

The determination of the structure parameters is an important problem, for which several tests are proposed in the literature, the most popular ones being the Akaike information criterion [1], abbreviated to AIC, and the minimum description length (MDL) criterion of Rissanen [50].

The model validation consists in testing whether the model is valid, e.g. ”good enough”. Such tests are based on a priori information about the system, the intended utilization of the model, its fitness to real I/O data, etc.

The parameter estimation algorithms are depending on two choices:

- The cost function (criterion) to be optimized.
- The optimization method or the algorithm utilized to compute the optimal solution.

Among the most popular parameter estimation methods, we can cite:

- The LS variants like the weighted least squares (WLS) method, with the Gauss-Markov estimator as particular case that is a best linear unbiased estimator (BLUE) under certain assumptions, the generalized least squares (GLS) method, introduced

by Aitken (1935), the extended least squares (ELS) and the total least squares (TLS) methods [57].
- The maximum-likelihood (ML) method, introduced by Gauss and Laplace, and then popularized by Fisher at the beginning of the 19th century, which consists in maximizing the probability density of the observations conditioned upon the parameters.
- The maximum a posteriori (MAP) method which consists in maximizing the probability density function of the unknown random parameters conditioned upon the knowledge of measured signals.
- The minimum mean squared error (MMSE) estimation method.
- The M-estimation methods, introduced by Huber (1964) in the context of robust statistics [30], and more generally robust identification methods in presence of outliers in the data or small errors in the hypotheses [4], [24].
- The instrumental variable (IV) methods [55], [59].
- The bounding approach that consists in determining a feasible set (also called membership set) for the parameters or states when the additive output error is assumed to be bounded; see [47] for a review of bounding techniques both for linear and nonlinear models; in [17], a review and a comparison of ellipsoidal outer bounding algorithms are made.
- The subspace methods whose the concept was introduced with the MUSIC (MUltiple SIgnal Classification) algorithm, which is a super resolution technique for array processing [54]. They are also strongly connected with deterministic and stochastic realization theories [13], [14], [29], [32].

Many books discuss estimation theory and system identification. In the field of engineering sciences, we can cite the fundamental contributions of [2], [3], [11], [12], [51]. See also [21], [23], [28], [33], [41], [42], [43], [46], [49], [56], [60] for linear systems, and [10], [15], [22], [26], [31], [45], [48] for nonlinear systems.

Identification methods can be classified in different ways depending on:

- The class of systems to be identified (linear/nonlinear, SISO/MIMO, ...).
- The assumptions made about model uncertainties (probabilistic description/unknown but bounded (UBB) error description).
- The domain of the used information (time-domain/frequency domain).
- The experiment configuration (open-loop/closed-loop).
- The I/O data processing (non iterative/iterative, non recursive/recursive or non adaptive/adaptive).
- The knowledge or not of the input signals (supervised/unsupervised or blind approaches).

Concerning this last point, we have to note that, unlike control applications for which the input signals are optimized, and therefore measured, in such a way that the system to be controlled has a desired behaviour, most of signal processing applications are characterized by the fact that input signals can not be measured, as it is the case of seismic, astronomical, audio, digital communication, or biomedical applications. That leads to unsupervised or blind identification methods. Such blind approaches are used for blind

seismic deconvolution, blind channel equalization in communications, or blind source separation in a more general context of MIMO (multi input - multi output) systems like antenna arrays in underwater acoustics, electrode arrays in electroencephalography or electrocardiography, audio mixtures etc. See [8].

Adaptive estimation is closely linked with adaptive filtering [27], [52], [58]. Adaptive modeling and processing are very useful for a wide variety of applications like adaptive control, adaptive noise cancelling, adaptive channel equalization, adaptive receiving arrays for various types of signals (seismic, acoustic, electromagnetic), among many others.

In our talk, we will begin by presenting the cost functions generally considered for parameter estimation.

Then, two standard optimization methods will be introduced: The Newton-Raphson method, with its variant the Levenberg-Marquardt algorithm, and the Gradient descent method.

Five parameter estimation methods will be presented for discrete-time Volterra models:

- The MMSE method, that leads to the Wiener solution,
- The non recursive deterministic LS method,
- The recursive least squares (RLS) method,
- The least mean squares (LMS) algorithm,
- The normalized least mean squares (NLMS) algorithm.

Other identification methods for Volterra systems can be found in [10], [37].

In the case of a Pth-order Volterra system and an i.i.d. (independently and identically distributed) input signal, we will give the necessary and sufficient condition for system identifiability using the MMSE solution. Moreover, in the case of a second-order Volterra system, we will give the decoupling conditions for the parameter estimation of the linear and quadratic kernels. The necessary condition to be satisfied by the step size for ensuring the LMS algorithm convergence will be also given.

As already mentioned, we have to note that a tensor-based approach has been recently proposed for reducing the parametric complexity of Volterra models, leading to the so-called Volterra-Parafac models [18]. Tensor-based methods have also been developed for blind identification of SISO and MIMO linear convolutive systems, and for block structured nonlinear system identification [16], [19], [20], [35], [36], [37], [38].

Some simulation results will be presented to illustrate the behaviour and to compare the performance of the (RLS, LMS, NLMS) algorithms and the (EKF, LMS, NLMS) algorithms for the parameter estimation of linear FIR models and Volterra-Parafac models, respectively.

Finally, our plenary talk will be concluded by giving several topics to be addressed in future research works.

References

1. H. Akaike. A new look at the statistical model identification. *IEEE Tr. on Automatic Control*, 19:716 – 723, 1974.
2. K. J. Aström and Bohlin T. Numerical identification of linear dynamic systems for normal operating records. In *Proc. 2nd IFAC Symp. on Theory of self-adaptive systems*, pages 96–111, Teddington, 1965.
3. G. E. P. Box and G. M. Jenkins. *Time series analysis: Forecasting, and control*. Holden-Day, San Francisco, 1970.
4. G.E.P. Box and G.C. Tiao. A bayesian approach to some outlier problems. *Biometrika*, 55(1):119–129, 1968.
5. S. Boyd and L. O. Chua. Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Tr. on Circuits and Systems*, CAS-32(11):1150–1161, Nov. 1985.
6. R. J. Campello, W. C. Amaral, and G. Favier. A note on the optimal expansion of Volterra models using Laguerre functions. *Automatica*, 42(4):689–693, 2006.
7. R. J. Campello, G. Favier, and W. C. Amaral. Optimal expansions of discrete-time Volterra models using Laguerre functions. *Automatica*, 40(5):815–822, 2004.
8. P. Comon and C. Jutten. *Handbook of blind source separation. Independent component analysis and applications*. Academic Press, Elsevier, 2010.
9. D. L. Donoho. Compressed sensing. *IEEE Tr. on Information Theory*, 52(4):1289–1306, 2006.
10. F. J. Doyle III, R. K. Pearson, and B. A. Ogunnaike. *Identification and control using Volterra models*. Springer, 2002.
11. P. Eykhoff. Some fundamental aspects of process-parameter estimation. *IEEE Tr. on Automatic Control*, 8:347–357, 1963.
12. P. Eykhoff. *System identification. Parameter and state estimation*. John Wiley & Sons, 1974.
13. P. Faurre, M. Clerget, and F. Germain. *Opérateurs rationnels positifs*. Dunod, 1979.
14. G. Favier. *Filtrage, modélisation et identification de systèmes linéaires stochastiques à temps discret*. Editions du CNRS, Paris, 1982.
15. G. Favier. *Signaux aléatoires : Modélisation, estimation, détection*, chapter 7: Estimation paramétrique de modèles entrée-sortie, pages 225–329. M. Guglielmi, Lavoisier, Paris, 2004.
16. G. Favier. Nonlinear system modeling and identification using tensor approaches. In *10th Int. Conf. on Sciences and Techniques of Automatic control and computer engineering (STA'2009)*, Hammamet, Tunisia, December 2009.
17. G. Favier and L.V. R. Arruda. *Bounding approaches to system identification*, chapter 4: Review and comparison of ellipsoidal bounding algorithms, pages 43–68. Plenum Press, 1996.
18. G. Favier and T. Bouilloc. Identification de modèles de Volterra basée sur la décomposition PARAFAC de leurs noyaux et le filtre de Kalman étendu. *Traitement du Signal*, 27(1):27–51, 2010.
19. C. E. R. Fernandes, G. Favier, and J. C. M. Mota. Blind channel identification algorithms based on the PARAFAC decomposition of cumulant tensors: The single and multiuser cases. *Signal Processing*, 88(6):1382–1401, June 2008.
20. C. E. R. Fernandes, G. Favier, and J. C. M. Mota. PARAFAC -based blind identification of convolutive MIMO linear systems. In *Proc. of the 15th IFAC Symp. on System Identification (SYSID)*, pages 1704 – 1709, Saint-Malo, France, July 2009.

21. A. Gelb. *Applied optimal estimation*. M.I.T. Press, Cambridge, 1974.
22. G. B. Giannakis and E. Serpedin. A bibliography on nonlinear system identification. *Signal Processing*, 81:533–580, 2001.
23. G. C. Goodwin and R. L. Payne. *Dynamic system identification: Experiment design and data analysis*. Academic Press, New York, 1977.
24. G.C. Goodwin, J.C. Agüero, J.S. Welsh, J.I. Yuz, G.J. Adams, and C.R. Rojas. Robust identification of process models from plant data. *Journal of Process control*, 18:810–820, 2008.
25. M. Guglielmi and G. Favier. *Signaux aléatoires : Modélisation, estimation, détection*, chapter 3: Modèles monovariables linéaires et non linéaires, pages 73–99. Lavoisier, Paris, 2004.
26. R. Haber and L. Keviczky. *Nonlinear system identification. Input-ouput modeling approach. Vol. 1: Nonlinear system parameter identification*. Kluwer Academic Publishers, 1999.
27. S. Haykin. *Adaptive filter theory*. Prentice-Hall, Englewood Cliffs, 4th edition, 2000.
28. P. S. C. Heuberger, P. M. J. Van den Hof, and B. Wahlberg. *Modelling and identification with rational orthogonal basis functions*. Springer, 2005.
29. B. L. Ho and R. E. Kalman. Effective construction of linear state-variable models from input/output functions. *Regelungstechnik*, 14(2):545–548, 1966.
30. P. J. Huber. *Robust statistics*. John Wiley, New York, 1981.
31. A. Janczak. *Identification of nonlinear systems using neural networks and polynomial models: A block-oriented approach*. Springer, 2005.
32. T. Katayama. *Subspace methods for system identification*. Springer, 2005.
33. S. M. Kay. *Fundamentals of statistical signal processing: Estimation theory*. Prentice-Hall, 1993.
34. G. Kerschen, K. Worden, A. F. Vakakis, and J. C. Golinval. Past, present and future of nonlinear system identification in structural dynamics. *Mechanical Systems and Signal Processing*, 20:505–592, 2006.
35. A. Y. Kibangou and G. Favier. Wiener-Hammerstein systems modeling using diagonal Volterra kernels coefficients. *IEEE Signal Processing Letters*, 13(6):381–384, June 2006.
36. A. Y. Kibangou and G. Favier. Identification of parallel-cascade Wiener systems using joint diagonalization of third-order Volterra kernel slices. *IEEE Signal Processing Letters*, 16(3):188–191, March 2009.
37. A. Y. Kibangou and G. Favier. Identification of fifth-order Volterra systems using i.i.d. inputs. *IET Signal Processing*, 4(3):30–44, 2010.
38. A. Y. Kibangou and G. Favier. Tensor analysis-based model structure determination and parameter estimation for block-oriented nonlinear systems. *IEEE J. of Selected Topics in Signal Processing, Special issue on Model Order Selection in Signal Processing Systems*, 4:514–525, 2010.
39. A. Y. Kibangou, G. Favier, and M. Hassani. Laguerre-Volterra filters optimization based on Laguerre spectra. *EURASIP J. on Applied Signal Processing*, 17:2874–2887, 2005.
40. A.Y. Kibangou, G. Favier, and M.M. Hassani. Selection of generalized orthonormal bases for second order Volterra filters. *Signal Processing*, 85(12):2371–2385, December 2005.
41. P. R. Kumar and P. Varaiya. *Stochastic systems: Estimation, identification and adaptive control*. Prentice-Hall, 1986.
42. L. Ljung. *System identification: Theory for the user*. Prentice-Hall, 1987.
43. L. Ljung and T. Söderström. *Theory and practice of recursive identification*. The MIT Press, 1983.
44. V. Z. Marmarelis. Nonlinear dynamic modeling of physiological systems. *Wiley-IEEE Press*, 2004.
45. V. J. Mathews and G. L. Sicuranza. Polynomial signal processing. *John Wiley & Sons*, 2000.
46. J. M. Mendel. *Lessons in digital estimation theory*. Prentice-Hall, 1987.

47. M. Milanese, J. Norton, P. Piet-Lahanier, and E. Walter. *Bounding approaches to system identification*. Plenum Press, New York, 1996.
48. O. Nelles. *Nonlinear system identification: From classical approaches to neural networks and fuzzy models*. Springer, 2001.
49. J. P. Norton. *An introduction to identification*. Academic Press, London, 1986.
50. J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.
51. A. P. Sage and J. L. Melsa. *System identification*. Academic Press, New York, 1971.
52. A. H. Sayed. *Fundamentals of adaptive filtering*. John Wiley & Sons, 2003.
53. M. Schetzen. *The Volterra and Wiener theories of nonlinear systems*. John Wiley & Sons, 1980.
54. R. O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Tr. on Antennas and Propagation*, 34(3):276–280, 1986.
55. T. Söderström and P. Stoica. *Instrumental variable methods for system identification*. Springer-Verlag, 1983.
56. T. Söderström and P. Stoica. *System identification*. Prentice-Hall, 1989.
57. S. Van Huffel and J. Vandewalle. *The total least squares problem: Computational aspects and analysis*. SIAM, 1991.
58. B. Widrow and S. D. Stearns. *Adaptive signal processing*. Prentice-Hall, Englewood Cliffs, 1985.
59. P. C. Young. Some observations on instrumental variable methods of time-series analysis. *International Journal of Control*, 23:593–612, 1976.
60. P. C. Young. *Recursive estimation and time-series analysis*. Springer-Verlag, Berlin, 1984.