Available online at www.sciencedirect.com
SCIENCE DIRECT®
Automatica 41 (2005) 2019–2032
automatica
www.elsevier.com/locate/automatica

H_∞ set membership identification: A survey☆

Mario Milanese, Michele Taragna*

Dipartimento di Automatica e Informatica, Politecnico di Torino, Corso Duca degli Abruzzi, 24-10129 Torino, Italy

Received 16 February 2004; received in revised form 19 July 2004; accepted 28 July 2005
Available online 4 October 2005

Abstract

Robustness had become in past years a central issue in system and control theory, focusing the attention of researchers from the study of a single model to the investigation of a set of models, described by a set of perturbations of a "nominal" model. Such a set, often indicated as an uncertainty model set or model set for short, has to be suitably constructed to describe the inherent uncertainty about the system under consideration and to be used for analysis and design purposes. H_∞ identification methods deliver uncertainty model sets in a suitable form to be used by well-established robust design techniques, based on H_∞ or μ optimization methods. The literature on H_∞ identification is now very extensive. In this paper, some of the most relevant contributions related to assumption validation, evaluation of bounds on unmodeled dynamics, convergence analysis and optimality properties of linear, two-stage and interpolatory algorithms are surveyed from a deterministic point of view.

Keywords: Robust identification; Robustness; Uncertainty models; Identification for control; Unmodeled dynamics

1. Introduction

A quite general problem appearing in many scientific and technical fields is to make some kind of inference on a dynamical system S^o, starting from some general information on it and from a finite number of noisy measurements. Typical examples of inference are smoothing, filtering, prediction, control design, decision making, fault detection, diagnosis, etc. The usual approach is to estimate a model M of S^o and to make the inference on M. Due to the unavoidable discrepancies between the identified model M and the actual system S^o, it is of paramount importance to evaluate the inference error, i.e. the error in making inference on M instead of S^o.

Consider in particular the control design problem. Typically, the system to be controlled is not completely known and a control law has to be designed, able to drive the plant to reach, if possible, given performance specifications. The classical approach consists in building a mathematical model of the plant, on the basis of the available information (priors and measurements), and then designing a control that meets the desired performance specifications for the identified model. However, this way it is not taken into account that any identified model is only an approximation of the actual system. Indeed, the performances actually achievable on the plant may be very poor, according to the size of the modeling error, and even the closed loop stability may be missed. In order to face these problems, robustness had become in past years a central issue in system and control theory, focusing the attention of the researchers from the study of a single model M to the investigation of a set of models $\mathscr{M}$. Such a set, often indicated as *uncertainty model set* or model set for short, has to be suitably constructed to describe the inherent uncertainty about the system under consideration and to be used for analysis and design purposes. Typically, model sets are described by a set of perturbations of a "nominal" model. Thus, in the control problem, the inference error corresponds to the difference between the performances

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in reviewed form by Associate Editor Brett Ninness under the direction of Editor Manfred Morari. This research was supported in part by funds of Ministero dell'Università e della Ricerca Scientifica e Tecnologica under the Project "Robustness and optimization techniques for control of uncertain systems".

* Corresponding author.
E-mail addresses: mario.milanese@polito.it (M. Milanese), michele.taragna@polito.it (M. Taragna).

0005-1098/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2005.07.007

predicted on the nominal model and those actually achievable on the real system. Since the real system is not known, the actually achievable performances are unknown and then those guaranteed for all systems belonging to the uncertainty model set are used instead.

Since data are corrupted by noise and provide only limited information, no finite bound on the inference error can be derived if no prior assumption on S^o and on noise is made. In particular, some information on S^o is required, e.g. by assuming that S^o belongs to some subset K of dynamical systems. Different theories have been developed, according to different assumptions on the set K and the noise.

Classical statistical identification theory gives deep and extensive results for the case that K is a set of parametric models $M(p)$, $p \in \Re^q$, and that noise is stochastic with known p.d.f., possibly filtered by a parametric noise model. As a matter of fact, in most practical applications there exists no p^o such that $S^o = M(p^o)$ and the problem arises of considering that only approximated models can be estimated and that the effects of unmodeled dynamics have to be accounted for. This appears to be a formidable problem. Some results are available in statistical identification and learning literature, essentially related to asymptotic analysis, see, e.g., Ljung and Yuan (1985), Vidyasagar (1996), and Ljung and Guo (1997). Indeed, evaluation of the identification accuracy with a finite number of samples is of great importance. Set membership (SM) identification methods have been developed in the last 20 years in order to deal with unmodeled dynamics and finite samples, see, e.g., Milanese and Vicino (1991), Smith and Dahleh (1994), Mäkilä, Partington, and Gustafsson (1995), Ninness and Goodwin (1995), Milanese, Norton, Piet-Lahanier, and Walter (1996), Partington (1997), Giarré, Milanese, and Taragna (1997), Garulli, Tesi, and Vicino (1999), Chen and Gu (2000) and the references therein.

H_∞ identification, where the modeling error is measured by the H_∞ norm, is among the most investigated SM methods in the literature. Indeed, H_∞ identification methods deliver uncertainty model sets in a suitable form to be used by well-established robust design techniques, based on H_∞ or μ optimization methods. The literature on H_∞ identification is now very extensive, considering time and/or frequency domain measurements and under different prior assumptions. In this paper, some of the most relevant contributions related to assumption validation, evaluation of bounds on unmodeled dynamics, convergence analysis and optimality properties are surveyed from a deterministic point of view.

Needless to say, any overview paper reflects the authors' outlook on the topic, and this paper is not an exception. Moreover, it is not the authors' ambition to provide a complete coverage of the literature, due to space limitations and uniformity of exposition. In particular, for model set validation results under both deterministic and stochastic frameworks, closely related to the assumption validation problem investigated in this paper, the interested reader can refer, e.g., to Smith and Doyle (1992), Zhou and Kimura (1993), Poolla, Khargonekar, Tikku, Krause, and Nagpal (1994), Chen (1997), Zhou (2001), Zhou, Wang, and Sun (2002) and the references cited there. Other important issues are not reviewed as well, such as the structure of unfalsified models (see, e.g., Zhou & Kimura, 1995; Zhou, 1998, 2000), the sample complexity (see, e.g., Harrison, Ward, & Gamble, 1996; Chen & Gu, 2002), the probabilistic and mixed probabilistic–deterministic approaches to model set identification (see, e.g., Goodwin, Gevers, & Ninness, 1992; de Vries & Van den Hof, 1995; Hakvoort & Van den Hof, 1997; Milanese & Taragna, 1999; Ljung, 1999; Reinelt, Garulli, & Ljung, 2002), the derivation of reduced order model sets (see, e.g., Zhou & Kimura, 1995; Beck, Doyle, & Glover, 1996; Jikuya & Kimura, 1998, 1999; Andersson, Rantzer, & Beck, 1999), the experiment design, the input selection, the presence of outliers (Mäkilä & Partington, 2000), mixed time-frequency measurements. Moreover, only pointwise bounds on errors are considered, while other classes of bounded noise, e.g. ℓ_2 or ℓ_1 bounded noise, have been investigated. Also, we are aware that w.p.1 we have missed to reference some important contribution, and we apologize in advance with the authors. Finally, it must be remarked that some of the results reported in the paper are specific H_∞ instances of other more general referenced results.

2. Prior and experimental information

In this section, the main concepts of SM H_∞ identification are introduced. Causal, discrete-time, single-input single-output, linear time-invariant, ℓ_2 BIBO stable, possibly distributed parameter, dynamical systems are considered. Any such system S is uniquely determined by its impulse response $h^S \doteq \{h_k^S\}_{k=0}^\infty$, whose power-series representation of the transfer function $S(z) \doteq \sum_{k=0}^\infty h_k^S z^k$ is an element of the Hardy space $\mathscr{H}_\infty(\mathbf{D})$ defined as

$$\mathscr{H}_\infty(\mathbf{D}) \doteq \left\{ f : \mathbf{D} \to \mathbf{C} \,|\, f \text{ analytic in } \mathbf{D} \text{ and } \|f\|_\infty \doteq \sup_{z\in\mathbf{D}} |f(z)| < \infty \right\},$$

where $\mathbf{D} \doteq \{z \in \mathbf{C} : |z| < 1\}$ is the open unit disk. Note that $S(z)$ as defined here merely denotes the standard z-transform of h^S evaluated at z^{-1}.

Let $S^o \in \mathscr{H}_\infty(\mathbf{D})$ be the actual plant to be identified using both experimental information and prior information (or assumptions).

2.1. Experimental information

The noisy measurements are represented as

$$y^N = F_N(S^o) + e^N,$$

where $y^N = [y_0 \dots y_{N-1}]^{\mathrm{T}} \in \Re^N$ is a known vector depending on the actual measurements, F_N is a known operator called "information operator" indicating how the measurements depend on S^o, and $e^N \in \Re^N$ is an unknown vector representing the measurement noise. The following experimental settings are considered.

- Time domain measurements of N samples of the output y of the system S^o, initially at rest, fed by a known one-sided input u ($u_\ell = 0$ for $\ell < 0$, $u_0 \neq 0$):

$$y_\ell = \sum_{k=0}^{\ell} h_k^{S^o} u_{\ell-k} + e_\ell, \quad \ell = 0, \ldots, N-1. \tag{1}$$

In this case, the information operator is given by $F_N(S) = F_N h^S$, where

$$F_N = [\, T_u \quad 0_{N\times\infty}\,] \in \Re^{N\times\infty} \tag{2}$$

and $T_u \in \Re^{N\times N}$ is the lower triangular Toeplitz matrix formed by the input vector $u^N = [u_0 \ldots u_{N-1}]^{\mathrm{T}}$.

- Measurements of real and imaginary part of the complex-valued samples of the system frequency response $S^o(\omega_k)$, $k = 1, \ldots, N/2$:

$$\begin{aligned} y_{2k-2} &= \Re\mathrm{e}(S^o(\omega_k)) + e_{2k-2},\\ y_{2k-1} &= \Im\mathrm{m}(S^o(\omega_k)) + e_{2k-1}. \end{aligned} \tag{3}$$

In this case, the information operator is given by $F_N(S) = F_N h^S$, where

$$F_N = [\, \Omega^{\mathrm{T}}(\omega_1) \quad \cdots \quad \Omega^{\mathrm{T}}(\omega_{N/2})\,]^{\mathrm{T}} \in \Re^{N\times\infty}, \tag{4}$$

$$\Omega(\omega) = \begin{bmatrix} \Omega_1(\omega) \\ \Omega_2(\omega) \end{bmatrix} = \begin{bmatrix} \Re\mathrm{e}(\Psi(\omega)) \\ \Im\mathrm{m}(\Psi(\omega)) \end{bmatrix} \in \Re^{2\times\infty}, \tag{5}$$

$$\Psi(\omega) = [\, 1 \quad \mathrm{e}^{\mathrm{j}\omega} \quad \mathrm{e}^{\mathrm{j}2\omega} \quad \cdots\,] \in \mathbf{C}^{1\times\infty}. \tag{6}$$

- Mixed time and frequency domain measurements, consisting of both (1) and (3), can be considered as well. In such a case, the information operator can be obtained by stacking together the information operators (2) and (4).

2.2. Prior information on plant S^o

Assumption on S^o is typically given as $S^o \in K \subset \mathscr{H}_\infty(\mathbf{D})$, where K is a nonfinitely parametrized subset of dynamical systems. The following subsets K investigated in the literature are considered in the paper:

(1) $K_{\rho,L}^{(1)} \doteq \{S \in \mathscr{H}_\infty(\mathbf{D}) : \sup_{z\in\overline{\mathbf{D}}_\rho} |S(z)| \leqslant L\}$ with $L > 0$, $\rho > 1$ and $\overline{\mathbf{D}}_\rho \doteq \{z \in \mathbf{C} : |z| \leqslant \rho\}$ the closed disk of radius ρ.

(2) $K_{\rho,L}^{(2)} \doteq \{S \in \mathscr{H}_\infty(\mathbf{D}) : |h_k^S| \leqslant L\rho^{-k},\ \forall k \geqslant 0\}$ with $L > 0$ and $\rho > 1$.

(3) $K_\gamma^{(3)} \doteq \{S \in \mathscr{H}_\infty(\mathbf{D}) : \sup_{z\in\mathbf{D}} |\frac{\mathrm{d}S(z)}{\mathrm{d}z}| \leqslant \gamma\}$.

(4) $K_\gamma^{(4)} \doteq \{S \in \mathscr{H}_\infty(\mathbf{D}) : \sup_{\omega\in[0,\pi]} |\frac{\mathrm{d}S_{\mathrm{R/I}}(\omega)}{\mathrm{d}\omega}| \leqslant \gamma\}$, where the notation $S_{\mathrm{R}}(\omega) \doteq \Re\mathrm{e}[S(\mathrm{e}^{-\mathrm{j}\omega})]$ and $S_{\mathrm{I}}(\omega) \doteq \Im\mathrm{m}[S(\mathrm{e}^{-\mathrm{j}\omega})]$ is used. The notation $(\cdot)_{\mathrm{R/I}}$ indicates both $(\cdot)_{\mathrm{R}}$ and $(\cdot)_{\mathrm{I}}$.

The following result, derived by standard properties of analytic functions, may prove useful to understand the relations among these sets.

Result 1.

(i) $K_{\rho,L}^{(1)} \subset K_{\rho,L}^{(2)}$,

(ii) $K_{\rho,L}^{(2)} \subset K_\gamma^{(3)}$ for $\gamma = \frac{L\rho}{(\rho-1)^2}$,

(iii) $K_\gamma^{(3)} \subset K_\gamma^{(4)}$.

It has to be pointed out that the computational complexity of the assumption validation and derivation of almost optimal algorithms is highly dependent on the assumed subset K. In fact, as shown in Sections 3 and 7 and discussed in Section 8, the choice of K has a direct effect on the number of data that can be reasonably processed and on the tightness of the identified model sets.

2.3. Prior information on noise e^N

$$e^N \in \mathscr{B}_e = \{\tilde{e}^N = [\tilde{e}_0 \cdots \tilde{e}_{N-1}]^{\mathrm{T}} \in \Re^N : \|A\tilde{e}^N\|_\infty^{W_e} \leqslant \varepsilon\}, \tag{7}$$

where

$$\|A\tilde{e}^N\|_\infty^{W_e} = \|W_e^{-1}A\tilde{e}^N\|_\infty = \max_{0\leqslant k\leqslant l-1} w_{e,k}^{-1} |(A\tilde{e}^N)_k|,$$

$A \in \Re^{l\times N}$ is a given matrix of rank N and $W_e = \mathrm{diag}(w_{e,0}, \ldots, w_{e,l-1}) \in \Re^{l\times l}$ is a given weighting matrix with $w_{e,k} > 0\ \forall k$.

By taking $A = W_e = I_{N\times N}$, such an assumption accommodates for constant magnitude bound on noise: $|\tilde{e}_k| \leqslant \varepsilon\ \forall k$. By suitably choosing W_e, it is possible to consider noise bounds dependent on k: $|\tilde{e}_k| \leqslant w_{e,k}\varepsilon\ \forall k$, to account e.g. for relative measurement errors. By suitably choosing A, it is possible to account for information on deterministic uncorrelation properties of the noise, see, e.g., Hakvoort and Van den Hof (1995), Paganini (1996), Venkatesh and Dahleh (1997).

3. Validation of prior assumptions

As typical in any identification theory, the problem of checking the validity of prior assumptions arises. The only thing that can be actually done is to check if prior assumptions are invalidated by the data, evaluating if no system exists consistent with data and assumptions. To this end it is useful to consider the Feasible Systems Set, often called *unfalsified systems set*, i.e. the set of all systems consistent with prior information and measured data.

Definition 1 (*Feasible Systems Set*).

$$\begin{aligned} &FSS(K, \mathscr{B}_e, F_N, y^N)\\ &\quad \doteq \{S \in K : y^N = F_N(S) + \tilde{e}^N,\ \tilde{e}^N \in \mathscr{B}_e\}. \end{aligned}$$

Thus, prior assumptions are invalidated if *FSS* is empty. However, it is usual to introduce the concept of prior assumption validation as consistency with the measured data, i.e. *FSS* is not empty.

Definition 2 (*Validation of prior assumptions*). Prior assumptions are considered *validated* if $FSS \neq \emptyset$.

Note that the fact that the prior assumptions are consistent with the present data does not exclude that they may be invalidated by future data.

The following results show how to validate different prior assumptions.

Result 2 (*Chen, Nett, & Fan, 1992, 1995*). Given $N/2$ frequency-domain measurements, the prior assumptions $S^o \in K^{(1)}_{\rho,L}$ and $e^N \in \mathscr{B}_e$ with $A = W_e = I_{N\times N}$ are validated if and only if there is a vector $\eta^N \in \mathscr{B}_e$ such that

$$\begin{bmatrix} -Q^{-1} & -L^{-1}(D_y - D_\eta)^H \\ -L^{-1}(D_y - D_\eta) & -Q \end{bmatrix} \leqslant 0, \tag{8}$$

where $D_y \doteq \mathrm{diag}(\{y_{2k-2} + jy_{2k-1}\}_{k=1}^{N/2})$, $D_\eta \doteq \mathrm{diag}(\{\eta_{2k-2} + j\eta_{2k-1}\}_{k=1}^{N/2})$ and $Q \doteq [\frac{1}{1-\rho^{-2}z_i\bar{z}_l}]$, with $z_k = \mathrm{e}^{\mathrm{j}\omega_k}$, for $i, l, k = 1, \ldots, N/2$.

Result 3 (*Chen & Nett, 1995*). Given N time-domain measurements, the prior assumptions $S^o \in K^{(1)}_{\rho,L}$ and $e^N \in \mathscr{B}_e$ with $A = W_e = I_{N\times N}$ are validated if and only if there is a vector $\eta^N \in \mathscr{B}_e$ such that

$$\begin{bmatrix} -T_u^{\mathrm{T}} D_\rho^2 T_u & -(T_y - T_\eta)^{\mathrm{T}} \\ -(T_y - T_\eta) & -L^2 D_\rho^{-2} \end{bmatrix} \leqslant 0 \tag{9}$$

with $D_\rho \doteq \mathrm{diag}(1, \rho, \ldots, \rho^{N-1})$ and T_u, T_y and T_η given by the lower triangular Toeplitz matrices associated with u^N, y^N and η^N, respectively.

These two validation results for $K = K^{(1)}_{\rho,L}$ are based on two different interpolation techniques, Nevanlinna–Pick (see, e.g., Ball, Gohberg, & Rodman, 1990) and Carathéodory–Fejér (see, e.g., Rosenblum & Rovnyak, 1985), respectively. The interpolation conditions are converted into the LMI problems (8) and (9), which appear to be computationally more efficient.

Let S^n be the FIR_n system with impulse response $h^{S^n} \doteq \{h_0^S, h_1^S, \ldots, h_{n-1}^S, 0, \ldots\}$.

Result 4 (*Milanese & Taragna, 2000*). Given time or frequency domain measurements, let n be a given positive integer. Conditions for validating prior assumptions $S^o \in K^{(2)}_{\rho,L}$ and $e^N \in \mathscr{B}_e$ are as follows:

(i) $\varepsilon^* \leqslant \varepsilon$ is a sufficient condition, being ε^* solution of the problem:

$$\varepsilon^* = \min_{\tilde{\varepsilon}, S^n: \begin{cases} S^n \in K^{(2)}_{\rho,L}, \\ \|A(y^N - F_N h^{S^n})\|_\infty^{W_e} \leqslant \tilde{\varepsilon}. \end{cases}} \tilde{\varepsilon}. \tag{10}$$

(ii) $\varepsilon^* \leqslant \varepsilon + \varepsilon_n \|W_e^{-1}A\|_{\infty,\infty}$ is a necessary condition, with $\varepsilon_n = \frac{L\rho^n}{1-\rho}$ and $\|W_e^{-1}A\|_{\infty,\infty} = \max_{0\leqslant i\leqslant l-1} \sum_{k=0}^{N-1} |(W_e^{-1}A)_{ik}|$.

(iii) $\varepsilon^* \leqslant \varepsilon$ is a necessary and sufficient condition for the case of time domain data, if $n \geqslant N$ is chosen.

Note that problem (10) can be solved by linear programming techniques. By choosing n sufficiently large, the "gap" between sufficient and necessary conditions can be made arbitrarily small.

Result 5 (*Milanese, Novara, & Taragna, 2001*). Given $N/2$ frequency-domain measurements, conditions for validating prior assumptions $S^o \in K^{(4)}_\gamma$ and $e^N \in \mathscr{B}_e$ with $A = I_{N\times N}$ are:

(i) $\xi_{\mathrm{R/I},k} \geqslant \underline{h}_{\mathrm{R/I},k}$, for $k = 1, \ldots, N/2$, is a necessary condition, where

$$\xi_{\mathrm{R/I},k} \doteq \min_{l=1,\ldots,N/2} (\overline{h}_{\mathrm{R/I},l} + \gamma|\omega_l - \omega_k|),$$

$$\underline{h}_{\mathrm{R},k} \doteq y_{2k-2} - w_{e,2k-2}\varepsilon, \quad \underline{h}_{\mathrm{I},k} \doteq y_{2k-1} - w_{e,2k-1}\varepsilon,$$

$$\overline{h}_{\mathrm{R},k} \doteq y_{2k-2} + w_{e,2k-2}\varepsilon, \quad \overline{h}_{\mathrm{I},k} \doteq y_{2k-1} + w_{e,2k-1}\varepsilon.$$

(ii) $\xi_{\mathrm{R/I},k} > \underline{h}_{\mathrm{R/I},k}$, for $k = 1, \ldots, N/2$, is a sufficient condition.

These latter validation conditions can be easily checked by straightforward computation.

4. Identification algorithms, model sets and identification errors

The $FSS(K, \mathscr{B}_e, F_N, y^N)$ summarizes the overall information on the system to be identified, i.e. prior assumptions on system and noise $(K, \mathscr{B}_e)$ and information coming from experimental data (F_N, y^N), thus describing the uncertainty about the system to be identified. If prior assumptions are "true", *FSS* includes S^o and, in the line with the robustness paradigm, control should be designed to be robust versus such an uncertainty model set. Some results on the *FSS* structure and the parameterization of all the unfalsified models for weekly corrupted plants can be found, e.g., in Zhou and Kimura (1995), Zhou (1998, 2000). However, *FSS* is in general not represented in a suitable form to be used by robust control design techniques, and model sets with such a property have to be looked for. In order to be consistent with robust control design philosophy, *model sets including the set of unfalsified systems* have to be looked for. This is formalized by the following definition.

Definition 3 (*Model set*). A set of models $\mathscr{M} \subseteq \mathscr{H}_\infty(\mathbf{D})$ is called a *model set* for S^o if:

$$\mathscr{M} \supseteq FSS.$$

In this paper, additive model sets of the following form are considered:

$$\mathscr{M}(M, W) = \{M + \Delta : |\Delta(\omega)| \leqslant W(\omega), \quad \forall\omega \in [0, 2\pi]\}, \tag{11}$$

where M is called the nominal model. In this case, the following result is an immediate consequence of Definition 3.

Result 6. For a given nominal model M:

- $\mathscr{M}(M, W)$ is a model set for S^o if and only if:

$$W(\omega) \geqslant W^*(\omega, M) \doteq \sup_{S \in FSS} |S(\omega) - M(\omega)|, \quad \forall \omega \in [0, 2\pi].$$

- $\mathscr{M}(M, W^*)$ is the smallest model set for S^o of the form (11), i.e. for any model set $\mathscr{M}(M, W)$ it results:

$$\mathscr{M}(M, W^*) \subseteq \mathscr{M}(M, W).$$

The nominal model can be obtained by some identification algorithm, i.e. an operator ϕ mapping the available information, represented by the quadruple $(K, \mathscr{B}_e, F_N, y^N)$, into a model $M \in \mathscr{H}_\infty(\mathbf{D})$:

$$\phi(K, \mathscr{B}_e, F_N, y^N) = M.$$

For notational simplicity, the dependence on y^N only will be usually evidenced.

Some of the main features of an identification algorithm can be summarized as follows.

Definition 4 (*Linear/nonlinear, untuned/tuned, interpolatory algorithm*).

- An algorithm ϕ is said to be *linear* if it is a linear function of the data y^N; otherwise, it is said to be *nonlinear*.
- An algorithm ϕ is said to be *untuned* if it does not depend on plant and noise information, i.e. if $\phi(K, \mathscr{B}_e, F_N, y^N)$ is actually not dependent on the constants involved in K and $\mathscr{B}_e$ definitions; otherwise, it is said to be *tuned*.
- An algorithm ϕ^{I} is said to be *interpolatory* if it always gives models consistent with prior information and measured data, i.e. if $M^{\mathrm{I}} = \phi^{\mathrm{I}}(y^N) \in FSS$.

Given an algorithm ϕ, the error $\|S^o - \phi(y^N)\|_\infty$ cannot be exactly known. The tightest upper bound on this error *for given data record* is $\sup_{S \in FSS} \|S - \phi(y^N)\|_\infty$, while *for any possible system and noise* is $\sup_{S \in K} \sup_{\tilde{e}^N \in \mathscr{B}_e} \|S - \phi(F_N(S) + \tilde{e}^N)\|_\infty$. This motivates the definition of the following two identification errors.

Definition 5 (*Local and global identification errors*).

- The *local identification error* of the algorithm ϕ and of the identified model $M = \phi(y^N)$ is

$$E_l(\phi) = E(M) = \sup_{S \in FSS} \|S - M\|_\infty.$$

- The *global identification error* of the algorithm ϕ is

$$E_g(\phi) = \sup_{S \in K} \sup_{\tilde{e}^N \in \mathscr{B}_e} \|S - \phi(F_N(S) + \tilde{e}^N)\|_\infty.$$

Note that $E_g(\phi) \geqslant E_l(\phi)$, since the following result holds.

Result 7 (*Milanese & Tempo, 1985*).

$$E_g(\phi) = \sup_{y^N} E_l(\phi(y^N)).$$

The local error $E_l(\phi)$, contrary to the global error $E_g(\phi)$, is not worst-case with respect to the noise. This fact has important implications in optimality and convergence properties, as shown in the next sections.

In Section 7, bounds on $E_g(\phi)$ for different prior assumptions and algorithms are reported.

For a given identification algorithm ϕ, providing the model $M = \phi(y^N)$, the evaluation of local error $E_l(\phi)$ is important because it represents the tightest bound on model error $\|S^o - M\|_\infty$. Since $E_l(\phi) = \sup_\omega \sup_{S \in FSS} |S(\omega) - \hat{M}(\omega)| = \sup_\omega W^*(\omega, M)$, evaluation of $E_l(\phi)$ can be made by computing $W^*(\omega, M)$ for a sufficiently coarse set of frequencies. The following result shows how to compute $W^*(\omega, M)$ in the case that $K = K^{(2)}_{\rho,L}$.

Result 8 (*Milanese & Taragna, 2002*). Assume time or frequency domain measurements and $K = K^{(2)}_{\rho,L}$. Let $m \geqslant 3$ and n be such that there exists a FIR_n system $S^n \in FSS$. Then, for given model M:

$$\underline{W}^n_m(\omega, M) \leqslant W^*(\omega, M) \leqslant \overline{W}^n_m(\omega, M),$$

$$\lim_{n,m \to \infty} \overline{W}^n_m(\omega, M) = \lim_{n,m \to \infty} \underline{W}^n_m(\omega, M) = W^*(\omega, M),$$

where

$$\overline{W}^n_m(\omega, M) = \max_{k=1,\ldots,m} \|M(\omega) - \overline{v_k}(\omega)\|_2 + \frac{L\rho^n}{1-\rho},$$

$$\underline{W}^n_m(\omega, M) = \max_{k=1,\ldots,m} \|M(\omega) - \underline{t_k}(\omega)\|_2,$$

$$\overline{v_k}(\omega) = \begin{bmatrix} \sin(s_k) & \cos(s_k) \\ \sin(s_{k+1}) & \cos(s_{k+1}) \end{bmatrix}^{-1} \times \begin{bmatrix} [\sin(s_k) & \cos(s_k)]\overline{t_k}(\omega) \\ [\sin(s_{k+1}) & \cos(s_{k+1})]\overline{t_{k+1}}(\omega) \end{bmatrix}, \tag{12}$$

$$s_k = 2\pi k/m,$$

$$\overline{t_k}(\omega) = \Omega(\omega) \operatorname*{argmin}_{S^n \in \overline{FSS}^n} [-\Omega_1(\omega)\sin(s_k) + \Omega_2(\omega)\cos(s_k)] h^{S^n} \in \Re^2, \tag{13}$$

$$\underline{t_k}(\omega) = \Omega(\omega) \operatorname*{argmin}_{S^n \in \underline{FSS}^n} [-\Omega_1(\omega)\sin(s_k) + \Omega_2(\omega)\cos(s_k)] h^{S^n} \in \Re^2, \tag{14}$$

$$\overline{FSS}^n = \{S^n \in K^{(2)}_{\rho,L} : \|(W_e + \widetilde{W})^{-1} A(y^N - F_N h^{S^n})\|_\infty \leqslant \varepsilon\},$$

$$\underline{FSS}^n = \{S^n \in K^{(2)}_{\rho,L} : \|W_e^{-1} A(y^N - F_N h^{S^n})\|_\infty \leqslant \varepsilon\},$$

$$\widetilde{W} = \frac{L\rho^n}{1-\rho} \operatorname{diag}(\|a_1\|_1, \ldots, \|a_l\|_1) \in \Re^{l \times l},$$

with $a_\ell = \ell$th row of A.

Note that the optimization problems (13) and (14) are linear programs. Moreover, the value of n as required by the result can

always be found, provided that prior assumptions are validated, i.e. $\underline{FSS}^n \neq \emptyset$.

From this result it follows that

$$\underline{E}_m^n(M) \doteq \sup_{0\leqslant\omega\leqslant 2\pi} \underline{W}_m^n(\omega, M) \leqslant E(M) \leqslant \sup_{0\leqslant\omega\leqslant 2\pi} \overline{W}_m^n(\omega, M) \doteq \overline{E}_m^n(M),$$

$$\lim_{n,m\to\infty} \overline{E}_m^n(M) = \lim_{n,m\to\infty} \underline{E}_m^n(M) = E(M).$$

The computation of $W^*(\omega, M)$ allows also one to derive, for given nominal model M, the smallest model set guaranteeing to contain the Feasible Systems Set, according to Result 6.

5. Optimality properties

Algorithms minimizing the global and local identification errors lead to the following optimality concepts.

Definition 6 (*Algorithm local and global optimality*).

- An algorithm ϕ^* is called ℓ-*optimal* if for all $K, \mathscr{B}_e, F_N, y^N$:

$$E_l(\phi^*) = \inf_{\phi} E_l(\phi(K, \mathscr{B}_e, F_N, y^N)) \doteq r(K, \mathscr{B}_e, F_N, y^N),$$

  $r(K, \mathscr{B}_e, F_N, y^N)$ is called *local radius of information*.
- An algorithm ϕ^g is called *g-optimal* if for all $K, \mathscr{B}_e, F_N, y^N$:

$$E_g(\phi^g) = \inf_{\phi} E_g(\phi(K, \mathscr{B}_e, F_N, y^N)) \doteq R(K, \mathscr{B}_e, F_N)$$

  $R(K, \mathscr{B}_e, F_N)$ is called *global radius of information*.

Local optimality is a stronger optimality concept than global optimality. In fact, if an algorithm is ℓ-optimal, then it is g-optimal, but the converse implication is not true.

It is also useful to define the optimality of identified model $M = \phi(y^N)$ as follows:

Definition 7 (*Optimal model*). A model $M = \phi(y^N)$ is called *optimal* if for the given $K, \mathscr{B}_e, F_N, y^N$:

$$E(M) = \inf_{\widetilde{M}\in\mathscr{H}_\infty(\mathbf{D})} E(\widetilde{M}) = r(K, \mathscr{B}_e, F_N, y^N).$$

A basic result in IBC relates ℓ-optimal algorithms, optimal model and the H_∞ Chebicheff center M^c of *FSS*, defined as

$$M^c = \arg \inf_{\widetilde{M}\in\mathscr{H}_\infty(\mathbf{D})} \sup_{S\in FSS} \|S - \widetilde{M}\|_\infty.$$

Result 9 (*Traub, Wasilkowski, & Woźniakowski, 1988*). If M^c exists, then it is an optimal model and the algorithm $\phi^c(y^N) = M^c$, called *central*, is an ℓ-optimal algorithm.

Note that central algorithm ϕ^c is g-optimal, but there exist other g-optimal algorithms. In particular, while ϕ^c is nonlinear, linear g-optimal algorithms exist, as given by the following result.

Result 10 (*Marchuk & Osipenko, 1975; Milanese & Tempo, 1985*). There exist linear g-optimal algorithms.

Computing central or linear g-optimal algorithms is not known in the present H_∞ setting. This motivates the interest in deriving algorithms having lower complexity, at the expense of some degradation in the accuracy of the identified model set. The following definition is introduced to give a measure of such a degradation of the local error of a given algorithm with respect to the minimal error obtained by a central algorithm.

Definition 8 (*Algorithm deviation*). The *deviation* $dev(\phi)$ of the algorithm ϕ is

$$dev(\phi) = \sup_{y^N} E_l(\phi(y^N)) \Big/ r(y^N).$$

Note that $dev(\phi) \geqslant 1\ \forall\phi$ and $dev(\phi^c) = 1$. Higher values of $dev(\phi)$ mean worse identified models when any possible set of measurements is processed.

The following result shows that in the present H_∞ setting, linear algorithms, though possibly g-optimal, may give large degradation of the local error with respect to the minimal error obtained by a central algorithm.

Result 11 (*Traub et al., 1988; Kon & Tempo, 1989*). No linear algorithm with finite deviation exists.

The question arises if it is possible to derive computable algorithms with finite and possibly "small" deviation. This question is answered by the following result.

Result 12 (*Traub et al., 1988*). For any interpolatory algorithm ϕ^I it holds that

$$dev(\phi^I) \leqslant 2.$$

For this reason interpolatory algorithms are often called *almost optimal*. Methods for computing interpolatory algorithms for different prior assumptions are presented in Section 7.

A given algorithm ϕ, by processing *any possible information* $K, \mathscr{B}_e, F_N, y^N$, gives an identified model $M = \phi(y^N)$ for which the ratio $E(M)/r$ is bounded as

$$E(M)/r \leqslant dev(\phi).$$

However, *for given information* $K, \mathscr{B}_e, F_N, y^N$, the *actual ratio* $E(M)/r$ may be significantly lower than $dev(\phi)$. Then, for given identified model, it is of interest to evaluate the actual value of this ratio, called model optimality level.

Definition 9 (*Model optimality level*). The *optimality level* $\alpha(M)$ of a model $M = \phi(y^N)$ is

$$\alpha(M) = E(M)/r.$$

Note that $\alpha(M) \geqslant 1\ \forall M \in \mathscr{H}_\infty(\mathbf{D})$ and $\alpha(M^c) = 1$. The model optimality level is actually a measure of the degradation

of the identification accuracy: the higher $\alpha(M)$ is, the worse is the model M meant to be, on the basis of the available information.

A result on evaluation of model optimality level is available for $K = K^{(2)}_{\rho,L}$.

Let $V \subset \Re^2$ be a polytope. Its radius $r_2[V]$ in the Euclidean norm is defined as $r_2[V] = \inf_{s\in\Re^2} \sup_{v\in V} \|s - v\|_2$ and can be easily computed via standard algorithms available in computational geometry literature.

Result 13 (*Milanese & Taragna, 2002*). Assume time or frequency domain measurements and $K = K^{(2)}_{\rho,L}$. Let $m \geqslant 3$ and n be such that there exists a FIR_n system $S^n \in FSS$. Then, for given model M:

$$\underline{\alpha}^n_m(M) \leqslant \alpha(M) \leqslant \overline{\alpha}^n_m(M),$$

where

$$\underline{\alpha}^n_m(M) = \max\left\{1, \underline{E}^n_m(M) \Big/ \left[\sup_{0\leqslant\omega\leqslant 2\pi} r_2[VO^n_m(\omega)] + \frac{L\rho^n}{1-\rho}\right]\right\},$$

$$\overline{\alpha}^n_m(M) = \overline{E}^n_m(M) / \sup_{0\leqslant\omega\leqslant 2\pi} r_2[VI^n_m(\omega)]$$

with $VO^n_m(\omega)$ and $VI^n_m(\omega)$ the convex hulls of points $\overline{v_k}(\omega)$ and $\underline{t_k}(\omega)$, $k = 1, \ldots, m$, defined in (12) and (14), respectively.

Polytopes $VO^n_m(\omega)$ and $VI^n_m(\omega)$ are outer and inner convergent approximations of the value set $V(\omega)$, the set in the complex plane of $S(e^{-j\omega})$ for all $S \in FSS$.

6. Convergence properties

In order to investigate algorithm convergence when applied to any $S \in K$ and any $e^N \in \mathscr{B}_e$, conditions for convergence of the global error to zero as $N \to \infty$ are looked for. In general, this convergence cannot hold, unless $\varepsilon \to 0$, as shown by the following result.

Result 14 (*Traub et al., 1988*).

$$E_g(\phi) \geqslant \sup_{S\in FSS(K,\mathscr{B}_e,F_N,0)} \|S\|_\infty > 0, \quad \forall N.$$

Different kinds of algorithm convergence can be defined.

Definition 10 (*Algorithm convergence and robust convergence*).

- An algorithm ϕ is said to be *convergent* if:

$$\lim_{\varepsilon\to 0} \lim_{N\to\infty} E_g(\phi) = 0.$$

- An algorithm ϕ is said to be *robustly convergent* if it converges regardless of a priori information.

By Definition 4, tuned algorithms are not robustly convergent.

The following strong negative result holds for linear algorithms.

Result 15 (*Partington, 1992*). No robustly convergent linear algorithm exists.

This implies that any untuned linear algorithm is not convergent. In particular, least squares algorithms are not convergent, since they are linear and untuned, i.e. not dependent on prior assumptions on system and noise.

Convergent algorithms can be obtained by interpolation, as shown by the next result.

Result 16 (*Chen & Gu, 2000*). Any interpolatory algorithm is convergent.

In Section 7, interpolatory algorithms are presented for the different prior assumptions. These algorithms are tuned and then they are not robustly convergent.

Two questions may be of interest:

- Do there exist convergent linear algorithms?
- Do there exist robustly convergent algorithms?

The answer is affirmative for both questions, as shown in Section 7 where convergent tuned linear and robustly convergent nonlinear algorithms are presented.

In contrast to the global error $E_g(\phi)$, the local error $E_l(\phi)$ may converge to zero for finite values of ε, under suitable deterministic uncorrelation assumptions on noise and for suitable inputs. This may happen because the local error $E_l(\phi)$, contrary to the global error $E_g(\phi)$, is not the worst case with respect to the noise, and then can account for information on its uncorrelation properties. In particular, by suitably choosing A and W_e, the noise set $\mathscr{B}_e$ can arbitrarily approximate the set:

$$\widetilde{\mathscr{B}}_e = \left\{e^N \in \Re^N : \sup_\omega \left|\sum_{k=0}^{N-1} e_k e^{j\omega k}\right| N^{-\alpha} \leqslant \varepsilon, \quad \alpha > 1/2\right\},$$

which is composed of deterministic counterparts of uncorrelated noise. For example, sequences of i.i.d. bounded random variables asymptotically belong to $\widetilde{\mathscr{B}}_e$ with probability 1, see, e.g. Hakvoort and Van den Hof (1995) and Venkatesh and Dahleh (1997). If noise in time domain experiments belongs to this set, the *FSS* asymptotically shrinks to a singleton for any ε, as shown by the following result.

Result 17 (*Venkatesh & Dahleh, 1997*). Let measurements be in time domain and $e^N \in \widetilde{\mathscr{B}}_e$. Then, an input sequence u can be found such that:

$$\lim_{N\to\infty} r(y^N) = 0.$$

Looking for convergence to zero of local error for given ε leads to the following convergence concept.

Definition 11 (*Strong convergence*). An algorithm ϕ is said to be *strongly convergent* if:

$$\lim_{N\to\infty} E_l(\phi) = 0.$$

Results 17 and 12 imply that, in case of deterministic uncorrelated noise belonging to $\widetilde{\mathscr{B}}_e$, an input sequence u can be found such that any central and interpolatory algorithm using such an input is strongly convergent.

7. Identification algorithm properties

In this section, the main algorithms available in literature are reviewed, starting from the simplest ones (linear algorithms) and finishing with the most performing ones (nonlinear interpolatory algorithms). The features analyzed are convergence, tightness in error evaluation, order of the identified model, computational complexity.

7.1. Linear algorithms

The linear algorithms operate linearly on the experimental data. *Untuned* linear algorithms, based on least squares or polynomial approximation techniques (see, e.g., Parker & Bitmead, 1987), are independent of the prior information available on system and noise and, as a consequence of Result 15, they cannot be convergent. Indeed, their global identification error may be divergent for finite ε (Akçay & Hjalmarsson, 1994; Partington & Mäkilä, 1995; Akçay & Ninness, 1998).

Convergent *tuned* linear algorithms have been obtained based on least squares optimization, with constraints or penalty terms depending on plant and noise prior information (Gu & Khargonekar, 1992b; Helmicki, Jacobson, & Nett, 1993; Gu, Chu, & Kim, 1994).

The following result is obtained by minimizing least squares with a quadratic penalty term.

Result 18 (*Helmicki et al., 1993; Gu et al., 1994*). Assume $S^o \in K^{(1)}_{\rho,L}$, $e^N \in \mathscr{B}_e$ with $A = W_e = I_{N\times N}$ and $N/2$ equispaced frequency-domain measurements. Then, the linear algorithm:

$$\hat{M}(z) = \sum_{k=0}^{n(N)-1} q_k^* z^k, \quad q_k^* = \frac{c_k(y^N)}{1 + (\varepsilon/L + \rho^{-n(N)})^2 \rho^{2k}}$$

with $c_k(y^N) = \frac{2}{N}\sum_{l=1}^{N/2}(y_{2l-2} + j y_{2l-1})(\mathrm{e}^{\mathrm{j}4\pi/N})^{(l-1)k}$ the inverse DFT coefficients of y^N, has global error which, for $N/2 \geqslant n(N) > 0$, is bounded as:

$$E_g(\phi) \leqslant L\rho^{-n(N)} + (1+\sqrt{2})L\sqrt{\frac{\rho+1}{\rho-1}}\,(\varepsilon/L + \rho^{-n(N)})^{1/2}.$$

If $\lim_{N\to\infty} n(N) = \infty$, the algorithm is convergent.

The result can be extended to nonuniformly spaced frequency-domain measurements.

Linear algorithms are simple and easy to be computed, but have some important drawbacks. In particular, they cannot be robustly convergent (Result 15) and do not have finite deviation (Result 11), i.e. give identified models whose optimality properties can be arbitrarily bad. In order to have robust convergence or finite deviation, it is necessary to resort to more sophisticated nonlinear algorithms.

7.2. Nonlinear "two-stage" algorithms

In order to overcome the robust convergence limitations of linear algorithms, nonlinear untuned algorithms have been derived in case of frequency-domain measurements, performing the following "two-step" procedure:

- *Stage* 1: A noncausal preliminary model $\hat{M}^{(0)} \in \mathscr{L}_\infty$ is derived through a "untuned" linear algorithm performing a bilateral interpolation in $\mathscr{L}_\infty$ by means of trigonometric polynomials

$$\hat{M}^{(0)}(z) = \sum_{k=-n+1}^{n-1} w_{k,n}\, c_k(y^N) z^k,$$

  where $\{w_{k,n}\}_{k=0}^{n-1}$ is a weighting (or *window*) sequence independent of prior information.
- *Stage* 2: The identified model is chosen as the best (causal) approximation of $\hat{M}^{(0)}$ in $\mathscr{H}_\infty(\mathbf{D})$, by solving the nonlinear Nehari approximation problem

$$\hat{M}(z) = \operatorname*{argmin}_{M\in\mathscr{H}_\infty(\mathbf{D})} \|\hat{M}^{(0)} - M\|_\infty.$$

The solution is given by Nehari's theorem (Nehari, 1957):

$$\hat{M}(z) = \hat{M}^{(0)}(z) - \bar{\sigma}\frac{\sum_{k=1}^{n-1}\psi_{n-k}z^k}{z^{n-1}\sum_{k=1}^{n-1}\zeta_k z^k},$$

where $\zeta = [\zeta_1, \ldots, \zeta_{n-1}]^{\mathrm{T}}$ and $\psi = [\psi_1, \ldots, \psi_{n-1}]^{\mathrm{T}}$ are the right and left singular vectors of the Hankel matrix associated to the coefficients $w_{k,n}\, c_k(y^N)$, $k = -1, -2, \ldots, -n+1$, and $\bar{\sigma}$ is the corresponding maximum singular value.

The two-stage algorithms proposed in the literature differ from one another in the first step, since the approximation in $\mathscr{L}_\infty$ can be performed using different weighting sequences, even symmetric with respect to k (i.e., sinc-square, triangular, cosine, trapezoidal windows) and truncated for $k \geqslant n$.

Result 19 (*Helmicki, Jacobson, & Nett, 1991; Gu & Khargonekar, 1992a,b; Partington, 1992*). Assume $S^o \in K^{(1)}_{\rho,L}$, $e^N \in \mathscr{B}_e$ with $A = W_e = I_{N\times N}$ and $N/2$ equispaced frequency-domain measurements. Then, the global identification error of

a two-stage algorithm is bounded as follows:

(i) If $w_{k,n} = \sin(2k\pi/N)^2/(2k\pi/N)^2$ for $|k| < n$ and $w_{k,n} = 0$ for $|k| \geqslant n$ (sinc-square window), then

$$E_g(\phi) \leqslant 2\left[\min\left(\frac{8L\pi}{N(\rho-1)}, \frac{4L\pi^2(\rho+1)}{N^2(\rho-1)^2}\right) + \frac{N^2(L+\varepsilon)}{2n\pi^2} + \varepsilon\right].$$

(ii) If $w_{k,n} = 1 - \frac{|k|}{n}$ for $|k| < n$ and $w_{k,n} = 0$ for $|k| \geqslant n$ (triangular window), then, for any $m \leqslant \frac{N+2}{4}$

$$E_g(\phi) \leqslant 2\left[\varepsilon + 2\frac{L}{\rho^m} + \frac{L\rho(1-\rho^{-m})}{n(\rho-1)^2}\right].$$

(iii) If $w_{k,n} = \cos(\frac{k\pi}{2n+1})$ for $|k| < n$ and $w_{k,n} = 0$ for $|k| \geqslant n$ (cosine window), then, for any $m \leqslant \frac{N+2}{4}$

$$E_g(\phi) \leqslant 2\left[\frac{L}{\rho^m} + (\pi-1)\left(\varepsilon + \frac{L}{\rho^m}\right) + \frac{\pi^2 L\rho(1+\rho)}{8n^2(\rho-1)^3}\right].$$

(iv) If $w_{k,n}$ is a trapezoidal window defined by

$$w_{k,n} = \begin{cases} 1 + k/n & -n < k < -1, \\ 1 & 0 \leqslant k \leqslant m-1, \\ 1 - \dfrac{k-m+1}{n} & m \leqslant k \leqslant n+m-1, \\ 0 & \text{elsewhere,} \end{cases}$$

where $m + n \leqslant N/2 + 1$, then

$$E_g(\phi) \leqslant \sqrt{\frac{2N}{n}}\varepsilon + \left(1 + \sqrt{\frac{2N}{n}}\right)\frac{L}{\rho^m}.$$

From the above result it turns out that, to achieve robust convergence, the number n of inverse DFT samples to be computed (and successively smoothed) varies significantly according to the chosen window sequence. With the sinc-square window, n must be such that $\lim_{N\to\infty} N^2/n(N) = 0$: this condition is met, for example, for any choice of n such that $n(N) = \mathcal{O}(N^3)$. In the other cases, the identification error approaches 0 as $1/n$, $1/n^2$ and $\sqrt{N/n}$, respectively; then, in the first two cases it is only requested that $n \to \infty$, while in the latter case n can be chosen such that $n = n(N) = \mathcal{O}(N^2)$. Since the order of the identified model $\hat{M}(z)$ is $2n-3$, it can be well understood the benefit obtained by suitably smoothing the inverse DFT coefficients of data. At the same time, it can be observed the trade-off between the rate of convergence of the identification error bound as a function of the number of data and the magnitude of the bound on the worst-case error of the approximation algorithm. It turns out that, similar to what happens in classical statistical spectral analysis, while sophisticated windows providing fast convergence on the algorithm can be found, they necessarily imply worse errors than simpler windows for a small number of data.

To summarize the positive features of "two-stage" nonlinear algorithms, they are robustly convergent if the weighting sequence $\{w_{k,n}\}$ is even symmetric with respect to k, truncated for $k \geqslant n$ and thus independent of the prior information. Their identification error bound and model order are highly dependent on the chosen window sequence. The computational complexity of these algorithms is relatively small. As main drawbacks, their deviation is unknown and, even more relevant, the identified model may not belong to the set *FSS* of systems consistent with the overall priors available on the system to be identified.

7.3. Interpolatory algorithms

These nonlinear tuned algorithms identify models belonging to the *FSS*:

$$\phi^{\mathrm{I}}(K, \mathscr{B}_e, F_N, y^N) = \hat{M}^{\mathrm{I}} \in FSS$$

and they are able to interpolate the experimental data in an approximated way, taking explicitly into account the available prior information. From Result 12, their deviation is not greater than 2 and for this reason they are often called "almost optimal" or "2-optimal". Moreover, these interpolatory algorithms are convergent but not robustly, since they are tuned.

In general, a two-step procedure is carried out:

- *Step* 1: Validation of prior information.
- *Step* 2: Identification of a model $\hat{M}^{\mathrm{I}} \in FSS$ by means of nonlinear interpolation techniques.

Result 20 (*Gu, Xiong, & Zhou, 1993; Chen et al., 1995*). Assume $S^o \in K^{(1)}_{\rho,L}$, $e^N \in \mathscr{B}_e$ with $A = W_e = I_{N\times N}$ and $N/2$ equispaced frequency-domain measurements. Then, an interpolatory algorithm is given by the following procedure:

- *Step* 1: Find a solution $\eta^N \in \mathscr{B}_e$ of the consistency problem represented by the LMI (8) in Result 2.
- *Step* 2: By means of the standard Nevanlinna–Pick's algorithm, build a function $\hat{M}(z) \in K^{(1)}_{\rho,L}$ interpolating $\tilde{y}^N \doteq y^N + \eta^N$ and use it as the identified model.

The global identification error is bounded by

$$E_g(\phi) \leqslant 2\varepsilon\sqrt{\frac{l+m}{l-m}} + 2L\rho^{-(2m+1)}\left(1 + \sqrt{\frac{l+m}{l-m}}\right),$$

where l, m are arbitrary integers satisfying $0 < m < l \leqslant N/2 - m$.

Result 21 (*Chen & Nett, 1995*). Assume $S^o \in K^{(1)}_{\rho,L}$, $e^N \in \mathscr{B}_e$ with $A = W_e = I_{N\times N}$ and time domain measurements. Then, an interpolatory algorithm is given by the following procedure:

- *Step* 1: Find a solution $\eta^N \in \mathscr{B}_e$ of the consistency problem represented by the LMI (9) in Result 3.
- *Step* 2: By means of the standard Carathéodory–Fejér procedure, build a function $\hat{M}(z) \in K^{(1)}_{\rho,L}$ interpolating $\tilde{y}^N \doteq y^N + \eta^N$ and use it as the identified model.

The global identification error is bounded by

$$E_g(\phi) \leqslant 2\left[\sum_{k=0}^{N-1} \min\left\{\varepsilon \sum_{i=0}^{k} |\tau_i|, \frac{L}{\rho^k}\right\} + \frac{L}{\rho^{N-1}(\rho-1)}\right],$$

where $\{\tau_k\}_{k=0}^{N-1}$ are the elements of the first column of T_u^{-1}.

The order of the identified model is equal to the number of data N, except in singular cases where the order may be lower. Under the computational point of view, these two interpolatory algorithms are very much burdensome when the number of data is high. Moreover, the Pick's matrix may easily result to be ill-conditioned, while similar comments hold for the time domain algorithm. In order to significantly reduce the computational complexity for the case that the model order n is selected to be substantially lower than N, suboptimal algorithms have been proposed, which are simpler but are not interpolatory and produce identified models with a typically small quantifiable increment in the identification error, see, e.g., Gu (1994), Gu and Chen (2001).

Result 22 (*Milanese & Taragna, 2001*). Given time or frequency domain data and $K = K_{\rho,L}^{(2)}$, let n be a positive integer such that there exists a FIR_n system $S^n \in FSS$. Then, for given positive integers m and q, an interpolatory algorithm is $\phi^{no}(y^N) = M_n^{no}$, where M_n^{no} is the FIR_n model whose impulse response is obtained as solution of the problem:

$$h^{M_n^{no}} = \arg \min_{S^n \in FSS} \|s^* - \Omega^* h^{S^n}\|_\infty, \tag{15}$$

where

$$s^* = \begin{bmatrix} s^*(\widetilde{\omega}_1) \\ \vdots \\ s^*(\widetilde{\omega}_q) \end{bmatrix}, \quad s^*(\omega) = \begin{bmatrix} \max\limits_{s=\left[\begin{smallmatrix} s_1 \\ s_2 \end{smallmatrix}\right] \in VO_m^n(\omega)} s_1 \\ \max\limits_{s=\left[\begin{smallmatrix} s_1 \\ s_2 \end{smallmatrix}\right] \in VO_m^n(\omega)} s_2 \\ \min\limits_{s=\left[\begin{smallmatrix} s_1 \\ s_2 \end{smallmatrix}\right] \in VO_m^n(\omega)} s_1 \\ \min\limits_{s=\left[\begin{smallmatrix} s_1 \\ s_2 \end{smallmatrix}\right] \in VO_m^n(\omega)} s_2 \end{bmatrix},$$

$$\Omega^* = \begin{bmatrix} \Omega^*(\widetilde{\omega}_1) \\ \vdots \\ \Omega^*(\widetilde{\omega}_q) \end{bmatrix}, \quad \Omega^*(\omega) = \begin{bmatrix} \Omega^n(\omega) \\ \Omega^n(\omega) \end{bmatrix}$$

with $\widetilde{\omega}_k \in [0, \pi]$ for $k = 1, \ldots, q$; VO_m^n the convex hull of points $\overline{v_k}(\omega)$, $k = 1, \ldots, m$, defined in (12) in Result 8; $\Omega^n(\omega)$ given by the first n columns of the matrix $\Omega(\omega)$ defined in (5).

Solution of problem (15) can be performed by linear programming. Since ϕ^{no} is interpolatory, it follows from Result 12 that $\alpha(M_n^{no}) \leqslant 2$. Indeed, the actual value of $\alpha(M_n^{no})$ can be evaluated from Result 13 and it can be expected to be near to 1, since model M_n^{no} is derived as an approximation of the optimal model M^c.

Result 23 (*Glaum, Lin, & Zames, 1996*). Given $N/2$ equispaced frequency-domain measurements, $K = K_\gamma^{(3)}$ and $A = W_e = I_{N\times N}$, an interpolatory algorithm is $\phi^I(y^N) = M_n$, where M_n is the FIR_n model whose impulse response is obtained as solution of the problem:

$$h^{M_n} = \arg \min_{S^n \in FSS} \max_{\theta \in [0,\pi)} \left| \sum_{k=0}^{n-1} k h_k^{S^n} e^{jk\theta} \right|$$

$$\text{s.t.} \left| \sum_{k=0}^{n-1} h_k^{S^n} e^{j2k\pi\ell/N} - (y_{2\ell-2} + j y_{2\ell-1}) \right|$$

$$\leqslant \varepsilon + \gamma\left(\frac{1}{n} + \sqrt{\frac{\pi}{2nN}}\right), \quad \ell \in \left[0, \frac{N}{2} - 1\right].$$

The global identification error is bounded by

$$E_g(\phi) \leqslant 2\varepsilon + \gamma\left(\frac{1}{n} + \frac{\pi}{N} + \sqrt{\frac{\pi}{nN}}\right).$$

The above polynomial minimization problem can be solved by standard convex optimization methods.

Result 24 (*Milanese et al., 2001*). Given $N/2$ frequency-domain data, $K = K_\gamma^{(4)}$ and $A = I_{N\times N}$, let n be a positive integer such that there exists a FIR_n system $S^n \in FSS$. Then, for given positive integer q, an interpolatory algorithm is $\phi^*(y^N) = M_n^*$, where M_n^* is the FIR_n model whose impulse response is obtained as solution of the problem:

$$h^* = \arg \min_{S^n \in FSS} \|s^* - \Omega^* h^{S^n}\|_\infty, \tag{16}$$

where

$$s^* = [\Re e[S^{no}(\widetilde{\omega}_1)], \Im m[S^{no}(\widetilde{\omega}_1)], \ldots, \Re e[S^{no}(\widetilde{\omega}_q)], \Im m[S^{no}(\widetilde{\omega}_q)]]^T,$$

$$S^{no}(\omega) = \tfrac{1}{2}\{\overline{S}_R^*(\omega) + \underline{S}_R^*(\omega) + j[\overline{S}_I^*(\omega) + \underline{S}_I^*(\omega)]\},$$

$$\underline{S}_{R/I}^*(\omega) = \max_{k=1,\ldots,N/2} (\underline{h}_{R/I,k} - \gamma|\omega - \omega_k|),$$

$$\overline{S}_{R/I}^*(\omega) = \min_{k=1,\ldots,N/2} (\overline{h}_{R/I,k} + \gamma|\omega - \omega_k|),$$

$$\Omega^* = \begin{bmatrix} \Omega^n(\widetilde{\omega}_1) \\ \vdots \\ \Omega^n(\widetilde{\omega}_q) \end{bmatrix},$$

with $\widetilde{\omega}_k \in [0, \pi]$ for $k = 1, \ldots, q$; $\underline{h}_{R/I,k}$ and $\overline{h}_{R/I,k}$ defined as in Result 5; $\Omega^n(\omega)$ defined as in Result 22.

The local identification error is bounded by

$$E_l(\phi) \leqslant \frac{1}{2} \sup_{0 \leqslant \omega \leqslant 2\pi} \sqrt{[\overline{S}_R^*(\omega) - \underline{S}_R^*(\omega)]^2 + [\overline{S}_I^*(\omega) - \underline{S}_I^*(\omega)]^2}$$

$$\leqslant \sqrt{2} E_l(\phi).$$

Solution of problem (16) can be obtained by linear programming.

8. Discussion

Some main features of the presented results are now discussed from a user point of view.

The main ingredients on which the different results are built are as follows:

- *type of experiment*: time domain (Eqs. (1)–(2)), frequency domain (Eqs. (3)–(6));
- *type of algorithm*: linear or nonlinear, tuned or untuned, interpolatory (Definition 4);
- *type of prior information on plant*, i.e., type of subset K such that $S^o \in K$;
- *type of prior information on noise*. In this survey, for the reasons discussed in Section 1, methods assuming only pointwise bounded noise as described by Eq. (7) are presented. Thus, the main distinction between the methods is their ability in dealing with $A \neq I_{N\times N}$, allowing one to account for information on deterministic uncorrelation properties of noise.

The main features on which the "goodness" of the methods is evaluated are as follows:

- *convergence*: simple, robust (Definition 10) or strong (Definition 11);
- "*a priori*" *optimality*: measured by the algorithm deviation $dev(\phi)$ (Definition 8), which gives the maximal degradation for any possible set of measurements of the local identification error guaranteed by the algorithm ϕ with respect to the minimal error achievable by an optimal algorithm;
- *tightness in error evaluation*: the identification algorithm ϕ, after processing the available data y^N, gives a model $M = \phi(y^N)$ and the local error $E_l(\phi)$ represents the tightest bound on the model error $||S^o - M||_\infty$. If this information is used e.g. for robust control design, the tighter the evaluation of $E_l(\phi)$ is, the less conservative is the design;
- *frequency shaping of uncertainty*: many methods provide only upper bounds on the identification error $E_g(\phi) \leqslant \overline{E}$. This way, model sets of the form (11) with $W(\omega) = \overline{E}$, $\forall \omega \in [0, 2\pi]$, are obtained. According to Result 6, methods able to obtain tight evaluation of $W^*(\omega, M)$ allow one to deliver smaller model sets, which in turn, if used for robust control, give rise to less conservative design;
- "*a posteriori*" *optimality*: measured by the model optimality level $\alpha(M)$ (Definition 9), which gives the actual value of the identification error degradation of the model M identified using the available data y^N with respect to the optimal model;
- *computational complexity*.

Let us now summarize the main properties of the different types of algorithms.

Linear algorithms require low computational effort, allowing one to work with very large number of data (up to several thousands), but they are not robustly convergent (Result 15), i.e., in order to guarantee convergence they have to be tuned to the prior assumptions on the system to be identified and on noise (Definition 4). A significant drawback of linear algorithms is that they have no finite deviation (Result 11), i.e., the local identification error of the identified models may be arbitrarily larger than the minimal possible one.

Nonlinear two-stage algorithms have been devised, which are robustly convergent, i.e., the convergence is guaranteed for any value of the constants appearing in the prior assumptions on the system to be identified and on noise (Result 19). Their computational effort is still relatively low, since they require, in addition to the computation of a linear untuned algorithm, the solution of Nehari problem. Thus, two-stage algorithms can process quite large amount of data (up to some thousands). No optimality property of two-stage algorithms is known. In particular, no bound on their deviation is known, so that it is unknown how far from being optimal the identified models are.

For both types of algorithms, linear and two-stage, bounds $E_g(\phi) \leqslant \overline{E}$ on their global error are provided assuming that $S^o \in K^{(1)}_{\rho,L}$ and $A = I_{N\times N}$ in noise assumption (7). These bounds are useful to prove their convergence properties. However, their tightness is unknown and, in view of Result 7, they cannot be tight bounds on the local error $E_l(\phi)$. Thus, model sets $\mathscr{M}(M, \overline{E}) = \{M + \Delta : |\Delta(\omega)| \leqslant \overline{E},\ \forall \omega \in [0, 2\pi]\}$ derived from these bounds may be largely conservative.

Nonlinear interpolatory algorithms are convergent (Result 16), but not robustly, since they are tuned to the prior assumptions on the system to be identified and on noise. However, in case of deterministic uncorrelated noise, they are strongly convergent, since from Results 17 and 12 it follows that the local error $E(M)$ of identified models converges to zero for finite values of ε. Another important property of interpolatory algorithms is that their deviation is bounded by 2 (Result 12), thus guaranteeing that also the optimality level of identified models is not greater than 2. For this reason, interpolatory algorithms are often indicated as almost-optimal, since deviation 1 is guaranteed by optimal algorithms.

In conclusion, interpolatory algorithms have excellent convergence and optimality features. Their properties in relation to the other features (computational complexity, tightness in error evaluation, frequency shaping, "a posteriori" optimality) are highly dependent on the assumed set K:

- *computational complexity*
  - If $K = K^{(1)}_{\rho,L}$, algorithms based on Nevanlinna–Pick and Carathéodory–Fejér interpolation are used for both steps, validation (Results 2 and 3) and algorithm computation (Results 20 and 21). Computational problems may arise in processing more than moderate number of data (some decades).
  - If $K = K^{(2)}_{\rho,L}$, linear programming optimization has to be performed both in validation (Result 4) and algorithm computation (Result 22). The required computational effort allows to process up to several hundreds of data.
  - If $K = K^{(3)}_{\gamma}$, convex optimization methods are used (Result 23). Also in this case several hundreds of data may be processed with a reasonable computational effort.

Table 1

Algorithm	Convergence	"A priori" optimality	Error tightness	Frequency shaping	"A posteriori" optimality	Computational complexity
Linear "untuned"	No	No	No	No	No	Low
Linear "tuned"	Simple	No	No	No	No	Low
Nonlinear "two-stage"	Robust	No	No	No	No	Low
Nonlinear "interpolatory"						
with $K = K^{(1)}_{\rho,L}$	Simple/strong[a]	Almost	No	No	No	High
with $K = K^{(2)}_{\rho,L}$	Simple/strong[a]	Almost	Yes	Yes	Yes	Medium
with $K = K^{(3)}_{\gamma}$	Simple/strong[a]	Almost	No	No	No	Medium
with $K = K^{(4)}_{\gamma}$	Simple/strong[a]	Almost	$\sqrt{2}$	No	No	Low

[a]If deterministically uncorrelated noise is assumed.

- If $K = K^{(4)}_{\gamma}$, the validation is computationally trivial (Result 5) and the algorithm computation requires the solution of one linear programming problem (Result 24). Then, very large number of data (up to several thousands) can be processed.

- *tightness in error evaluation*
  - If $K = K^{(1)}_{\rho,L}$, only bounds $E_g(\phi) \leqslant \overline{E}$ on their global error are provided (Results 20 and 21). Their tightness is unknown and, in view of Result 7, they cannot be tight bounds on the local error $E_l(\phi)$. Moreover, these bounds do not account for possible deterministic uncorrelation information on noise.
  - If $K = K^{(2)}_{\rho,L}$, the error $E(M)$ of the model M identified by the interpolatory algorithm of Result 22 can be evaluated as tightly as desired by means of Result 8, possibly taking into account noise uncorrelation properties.
  - If $K = K^{(3)}_{\gamma}$, only a bound $E_g(\phi) \leqslant \overline{E}$ on the global error is provided (Result 23). Possible deterministic uncorrelation information on noise is not accounted for.
  - If $K = K^{(4)}_{\gamma}$, an upper bound on the local error is provided, whose overbounding is not greater than $\sqrt{2}$ (Result 24). Also in this case, possible uncorrelation information on noise is not accounted for.
- *frequency shaping of uncertainty*
  Methods for tight evaluation of the frequency shaping of uncertainty of the identified model are available only for the case $K = K^{(2)}_{\rho,L}$ (Result 8).
- "*a posteriori*" *optimality*
  Methods for evaluating the optimality level of the identified model are available only for the case $K = K^{(2)}_{\rho,L}$ (Result 13).

Table 1 summarizes the main results here presented and discussed.

Acknowlegments

The authors would like to thank the anonymous reviewers and the Associate Editors for the careful reading of the manuscript and their invaluable criticisms and suggestions, which were helpful in the improvement of this paper.

References

Akçay, H., & Hjalmarsson, H. (1994). The least-squares identification of FIR systems subject to worst-case noise. *Systems & Control Letters*, *23*, 329–338.

Akçay, H., & Ninness, B. (1998). On the worst-case divergence of the least-squares algorithm. *Systems & Control Letters*, *33*, 19–24.

Andersson, L., Rantzer, A., & Beck, C. (1999). Model comparison and simplification. *International Journal of Robust and Nonlinear Control*, *9*, 157–181.

Ball, J. A., Gohberg, I., & Rodman, L. (1990). *Interpolation of rational matrix functions*. Basel: Birkhäuser.

Beck, C. L., Doyle, J., & Glover, K. (1996). Model reduction of multidimensional and uncertain systems. *IEEE Transactions on Automatic Control*, *AC-41*(10), 1466–1477.

Chen, J. (1997). Frequency-domain tests for validation of linear fractional uncertain models. *IEEE Transactions on Automatic Control*, *AC-42*(6), 748–760.

Chen, J., & Gu, G. (2000). *Control-oriented system identification: an H_∞ approach*. New York: Wiley.

Chen, J., & Gu, G. (2002). Worst-case asymptotic properties of H_∞ identification. *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, *49*(4), 437–446.

Chen, J., & Nett, C. N. (1995). The Carathéodory–Fejér problem and H_∞/ℓ_1 identification: A time domain approach. *IEEE Transactions on Automatic Control*, *AC-40*(4), 729–735.

Chen, J., Nett, C. N., & Fan, M. K. H. (1992). Worst-case system identification in H_∞: Validation of *a priori* information, essentially optimal algorithms, and error bounds. In *Proceedings of the American control conference* (pp. 251–257), Chicago, IL.

Chen, J., Nett, C. N., & Fan, M. K. H. (1995). Worst case system identification in H_∞: Validation of *a priori* information, essentially optimal algorithms, and error bounds. *IEEE Transactions on Automatic Control*, *AC-40*(7), 1260–1265.

de Vries, D. K., & Van den Hof, P. M. J. (1995). Quantification of uncertainty in transfer function estimation: A mixed probabilistic-worst-case approach. *Automatica*, *31*(4), 543–557.

Garulli, A., Tesi, A., & Vicino, A. (Eds.). (1999). *Robustness in identification and control*, *Lecture notes in control and information sciences* (Vol. 445). Godalming, UK: Springer.

Giarré, L., Milanese, M., & Taragna, M. (1997). H_∞ identification and model quality evaluation. *IEEE Transactions on Automatic Control*, *AC-42*(2), 188–199.

Glaum, M., Lin, L., & Zames, G. (1996). Optimal H_∞ approximation by systems of prescribed order using frequency response data. In *Proceedings of the 35th IEEE conference on decision and control* (pp. 2318–2321), Kobe, Japan.

Goodwin, G. C., Gevers, M., & Ninness, B. (1992). Quantifying the error in estimated transfer functions with application to model order selection. *IEEE Transactions on Automatic Control*, *AC-37*(7), 913–928.

Gu, G. (1994). Suboptimal algorithms for worst case identification in H_∞ and model validation. *IEEE Transactions on Automatic Control*, *AC-39*(8), 1657–1661.

Gu, G., & Chen, J. (2001). A nearly interpolatory algorithm for H_∞ identification with mixed time and frequency response data. *IEEE Transactions on Automatic Control*, *AC-46*(3), 464–469.

Gu, G., Chu, C. -C., & Kim, G. (1994). Linear algorithms for worst case identification in H_∞ with applications to flexible structures. In *Proceedings of the American control conference* (pp. 112–116), Baltimore, MD.

Gu, G., & Khargonekar, P. P. (1992a). A class of algorithms for identification in H_∞. *Automatica*, *28*(2), 299–312.

Gu, G., & Khargonekar, P. P. (1992b). Linear and nonlinear algorithms for identification in H_∞ with error bounds. *IEEE Transactions on Automatic Control*, *AC-37*(7), 953–963.

Gu, G., Xiong, D., & Zhou, K. (1993). Identification in H_∞ using Pick's interpolation. *Systems & Control Letters*, *20*, 263–272.

Hakvoort, R. G., & Van den Hof, P. M. J. (1995). Consistent parameter bounding identification for linearly parametrized model sets. *Automatica*, *31*(7), 957–969.

Hakvoort, R. G., & Van den Hof, P. M. J. (1997). Identification of probabilistic system uncertainty regions by explicit evaluation of bias and variance errors. *IEEE Transactions on Automatic Control*, *AC-42*(11), 1516–1528.

Harrison, K. J., Ward, J. A., & Gamble, D. K. (1996). Sample complexity of worst-case H_∞-identification. *Systems & Control Letters*, *27*(4), 255–260.

Helmicki, A. J., Jacobson, C. A., & Nett, C. N. (1991). Control oriented system identification: A worst-case/deterministic approach in H_∞. *IEEE Transactions on Automatic Control*, *AC-36*(10), 1163–1176.

Helmicki, A. J., Jacobson, C. A., & Nett, C. N. (1993). Least squares methods for H_∞ control-oriented system identification. *IEEE Transactions on Automatic Control*, *AC-38*(5), 819–826.

Jikuya, I., & Kimura, H. (1998). Reduction of model sets with inclusion. In *Proceedings of the 37th IEEE conference on decision and control* (pp. 2201–2206), Tampa, FL.

Jikuya, I., & Kimura, H. (1999). Representation and reduction of model sets. In *Proceedings of the 38th IEEE conference on decision and control* (pp. 1482–1487), Phoenix, AZ.

Kon, M. A., & Tempo, R. (1989). On linearity of spline algorithms. *Journal of Complexity*, *5*(2), 251–259.

Ljung, L. (1999). Model validation and model error modeling. In B. Wittenmark, & A. Rantzer (Eds.), *The Åström Symposium on Control* (pp. 15–42). Lund, Sweden, Studentlitteratur.

Ljung, L., & Guo, L. (1997). The role of model validation for assessing the size of the unmodeled dynamics. *IEEE Transactions on Automatic Control*, *AC-42*(9), 1230–1239.

Ljung, L., & Yuan, Z. D. (1985). Asymptotic properties of black-box identification of transfer functions. *IEEE Transactions on Automatic Control*, *AC-30*(6), 514–530.

Mäkilä, P. M., & Partington, J. R. (2000). Robustness in H_∞ identification. *Automatica*, *36*(11), 1685–1691.

Mäkilä, P. M., Partington, J. R., & Gustafsson, T. K. (1995). Worst-case control-relevant identification. *Automatica*, *31*(12), 1799–1819.

Marchuk, A. G., & Osipenko, K. Yu. (1975). Best approximation of functions specified with an error at a finite number of points. *Mathematische Zametki*, *17*, 359–368 (in Russian; English Transl., *Mathematical Notes*, Vol. 17, pp. 207–212, 1975).

Milanese, M., Norton, J., Piet-Lahanier, H., & Walter, É. (Eds.). (1996). *Bounding approaches to system identification*. New York: Plenum Press.

Milanese, M., Novara, C., & Taragna, M. (2001). "Fast" set membership H_∞ identification from frequency-domain data. In *Proceedings of European control conference ECC 2001* (pp. 1698–1703), Porto, Portugal.

Milanese, M., & Taragna, M. (1999). H_∞ identification of "soft" uncertainty models. *Systems & Control Letters*, *37*(4), 217–228.

Milanese, M., & Taragna, M. (2000). Set membership identification for H_∞ robust control design. In *Proceedings of 12th IFAC symposium on system identification SYSID 2000*, Santa Barbara, CA.

Milanese, M., & Taragna, M. (2001). Nearly optimal model sets in H_∞ identification. In *Proceedings of the European control conference 2001* (pp. 1704–1709), Porto, Portugal.

Milanese, M., & Taragna, M. (2002). Optimality, approximation, and complexity in set membership H_∞ identification. *IEEE Transactions on Automatic Control*, *AC-47*(10), 1682–1690.

Milanese, M., & Tempo, R. (1985). Optimal algorithms theory for estimation and prediction. *IEEE Transactions on Automatic Control*, *AC-30*, 730–738.

Milanese, M., & Vicino, A. (1991). Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica*, *27*(6), 997–1009.

Nehari, Z. (1957). On bounded bilinear forms. *Annals of Mathematics*, *65*, 153–162.

Ninness, B., & Goodwin, G. C. (1995). Estimation of model quality. *Automatica*, *31*(12), 1771–1797.

Paganini, F. (1996). A set-based approach for white noise modeling. *IEEE Transactions on Automatic Control*, *AC-41*(10), 1453–1465.

Parker, P. J., & Bitmead, R. R. (1987). Adaptive frequency response identification. In *Proceedings of the 26th IEEE conference on decision and control* (pp. 348–353).

Partington, J. R. (1992). Robust identification in H_∞. *Journal of Mathematical Analysis and Applications*, *166*, 428–441.

Partington, J. R. (1997). *Interpolation, identification, and sampling, London mathematical society monographs new series* (Vol. 17). Oxford, New York: Clarendon Press.

Partington, J. R., & Mäkilä, P. M. (1995). Worst-case analysis of the least-squares method and related identification methods. *Systems & Control Letters*, *24*, 193–200.

Poolla, K., Khargonekar, P., Tikku, A., Krause, J., & Nagpal, K. (1994). A time-domain approach to model validation. *IEEE Transactions on Automatic Control*, *AC-39*(5), 951–959.

Reinelt, W., Garulli, A., & Ljung, L. (2002). Comparing different approaches to model error modeling in robust identification. *Automatica*, *38*(5), 787–803.

Rosenblum, M., & Rovnyak, J. (1985). *Hardy classes and operator theory*. New York: Oxford University Press.

Smith, R. S., & Dahleh, M. (Eds.). (1994). *The modeling of uncertainty in control systems*, *Lecture notes in control and information sciences* (Vol. 192). London, UK: Springer.

Smith, R. S., & Doyle, J. C. (1992). Model validation: A connection between robust control and identification. *IEEE Transactions on Automatic Control*, *AC-37*(7), 942–952.

Traub, J. F., Wasilkowski, G. W., & Woźniakowski, H. (1988). *Information-based complexity*. New York: Academic Press.

Venkatesh, S. R., & Dahleh, M. A. (1997). Identification in the presence of classes of unmodeled dynamics and noise. *IEEE Transactions on Automatic Control*, *AC-42*(12), 1620–1635.

Vidyasagar, M. (1996). *A theory of learning and generalization with application to neural networks and control systems*. Berlin: Springer.

Zhou, T. (1998). A parameterization of all the unfalsified plant models for MIMO systems. *IEEE Transactions on Automatic Control*, *AC-43*(1), 18–30.

Zhou, T. (2000). Unfalsified model parametrization based on frequency domain noise information. *Automatica*, *36*(5), 685–696.

Zhou, T. (2001). On the consistency between an LFT described model set and frequency domain data. *IEEE Transactions on Automatic Control*, *AC-46*(12), 2001–2007.

Zhou, T., & Kimura, H. (1993). Time domain identification for robust control. *Systems & Control Letters*, *20*(3), 167–178.

Zhou, T., & Kimura, H. (1995). Structure of model uncertainty for a weakly corrupted plant. *IEEE Transactions on Automatic Control*, *AC-40*(4), 639–655.

Zhou, T., Wang, L., & Sun, Z. (2002). Closed-loop model set validation under a stochastic framework. *Automatica*, *38*(9), 1449–1461.

Mario Milanese graduated in Electronic Engineering at Politecnico di Torino in 1967. From 1968 to 1972 he was Teaching Assistant at Politecnico di Torino, from 1972 to 1980 Associate Professor of System Theory at Università di Torino. From 1980 he is Full Professor of System Theory at Politecnico di Torino. From 1982 to 1987 he was head of the Department of Automatica e Informatica at Politecnico di Torino. His research interests include robust identification, prediction and control of uncertain systems, and applications to biomedical, automotive, aerospace, financial and environmental problems. He is author of more than 200 papers in international journals and conference proceedings. He is editor of the books "Robustness in Identification and Control", Plenum Press, 1989, and "Bounding Approaches to System Identification", Plenum Press, 1996.

Michele Taragna was born in Torino (Italy) in 1964. He received the Laurea degree in Electronic Engineering (1992) and the Ph.D. in Systems Engineering (1996), both from the Politecnico di Torino (Italy). Since 1995, he has been a Research Associate at the Dipartimento di Automatica e Informatica (Control and Computer Engineering Department) of the Politecnico di Torino, where he is teaching systems and control theory as Associate Professor since 2002. His research interests include the robustness field (in particular, set-membership identification and robust control) and the interaction issues between identification and control. He is currently serving as Associate Editor on the Conference Editorial Board of the IEEE Control System Society.