Optimal Estimation Theory for Dynamic Systems with Set Membership Uncertainty: An Overview

M. Milanese and A. Vicino

ABSTRACT

In many problems, such as linear and nonlinear regressions, parameter and state estimation of dynamic systems, state space and time series prediction, interpolation, smoothing, and functions approximation, one has to evaluate some unknown variable using available data. The data are always associated with some uncertainty and it is necessary to evaluate how this uncertainty affects the estimated variables. Typically, the problem is approached assuming a probabilistic description of uncertainty and applying statistical estimation theory. An interesting alternative, referred to as set membership or unknown but bounded (UBB) error description, has been investigated since the late 60s. In this approach, uncertainty is described by an additive noise which is known only to have given integral (typically l_1 or l_2) or componentwise (l_∞) bounds. In this chapter the main results of this theory are

---

M. Milanese • Dipartimento di Automatica e Informatica, Politecnico di Torino, 10129 Torino, Italy.
A. Vicino • Facoltà di Ingegneria, Università degli Studi di Siena, 53100 Siena, Italy.

Bounding Approaches to System Identification, edited by M. Milanese *et al.* Plenum Press, New York, 1996.

reviewed, with special attention to the most recent advances obtained in the case of componentwise bounds.

2.1. INTRODUCTION

Estimation theory is concerned with the problem of evaluating some unknown variables depending on given data (often obtained by measurements on a real process). Available data are always known with some uncertainty and it is necessary to evaluate how this uncertainty affects the estimated variables.

Obviously the solution of the problem depends on the type of assumptions made about uncertainty. The cases most investigated so far are unquestionably related to the assumption that uncertainty is given by an additive random noise with a (partially) known probability density function (pdf).

However, in many situations the very random nature of uncertainty may be questionable. For example, the real process generating the actual data may be very complex (large scale, nonlinear, and time varying) so that only simplified models can be practically used in the estimation process. The residuals of the estimated model have a component due to deterministic structural errors. Treating them as purely random variables may lead to unsatisfactory results.

An interesting alternative approach, set membership or unknown but bounded UBB error description has been pioneered by the work of Witsenhausen and Schweppe in the late 60s.[1,2,3] In this approach, uncertainty is described by means of an additive noise which is known only to have given bounds. The motivation for this approach is that in many practical cases the UBB error description is more realistic and less demanding than the statistical description. However, despite the appeal of its features, the UBB approach is not widely used yet. Until the early 80s, reasonable results and algorithms had been obtained only for uncertainty bounds of integral type (mainly l_2), while in practical applications componentwise bounds (l_∞) are mainly of interest.

Real advances have been obtained in the last few years for the componentwise bounds case, leading to theoretical results and algorithms which can be properly applied to practical problems where the use of statistical techniques is questionable.

The purpose of this chapter is to review these results and to present them in a unified framework, in order to contribute the present state of the art in the field and simulate further basic and applied researches.

2.2. PROBLEM FORMULATION

In this section a general framework is formulated such that the main results in the literature can be presented in a unifying view. Such formulation can be sketched as follows.[4,5]

We have a problem element λ (for example a dynamic system or a time function) and a function $S(\lambda)$ of this problem element (for example some parameter of the dynamic system or particular value of the time function) is to be evaluated. Suppose λ is not known exactly, but there is some information on it. In particular assume that it is an element of a set K of possible problem elements and that some function $F(\lambda)$ is measured. Moreover, suppose that exact measurements are not available and actual measurements y are corrupted by some error ρ.

The estimation problem is to find an estimator ϕ providing an approximation $\phi(y) \approx S(\lambda)$ using the available data y and evaluating some measure of such approximation. A geometric sketch is shown in Fig 2.1.

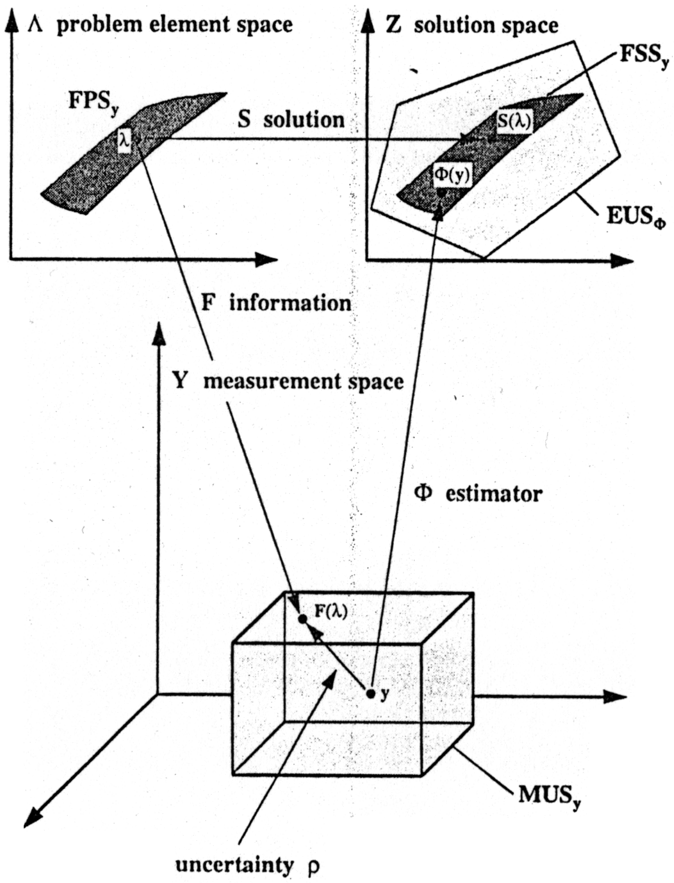

FIGURE 2.1. Generalized estimation problem.

2.2.1. Spaces and Operators

Let Λ be a linear normed n-dimensional space over the real field. Consider a given operator S, called a solution operator, which maps Λ into Z

$$S : \Lambda \rightarrow Z \tag{2.1}$$

where Z is a linear normed l-dimensional space over the real field. The aim is to estimate an element $S(\lambda)$ of the space Z, knowing approximate information about the element λ. Suppose that two kinds of information may be available. The first one (often referred to as *a priori* information) is expressed by assuming that $\lambda \in K$, where K is a subset of Λ. In most cases K is given as

$$K = \{\lambda \in \Lambda;\ \|R(\lambda - \lambda_0)\| \leq 1\} \tag{2.2}$$

where R is a linear operator and λ_0 is a known problem element. The second kind of information is usually provided by the knowledge of some function $F(\lambda)$, where F, called an information operator, maps Λ into a linear normed m-dimensional space Y

$$F : \Lambda \rightarrow Y. \tag{2.3}$$

Spaces Λ, Z, Y are called problem element, solution and measurement spaces respectively. In the following, unless otherwise specified, assume that Λ and Z are equipped with l_∞ norms and Y is equipped with an l_∞^w norm.[1]

In general, due to the presence of noise, exact information $F(\lambda)$ about λ is not available and only perturbed information y is given. In this context, uncertainty is assumed to be additive, i.e.,

$$y = F(\lambda) + \rho \tag{2.4}$$

where the error term ρ is unknown, but bounded by some given positive number ε

$$\|\rho\| \leq \varepsilon \tag{2.5}$$

Note that if an l_∞^w norm in measurement space Y is used, componentwise bounds with different values on every measurement can be treated.

An algorithm ϕ is an operator (in general nonlinear) from Y into Z:

$$\phi : Y \rightarrow Z \tag{2.6}$$

i.e., it provides an approximation $\phi(y)$ of $S(\lambda)$ using the available data y. Such an algorithm is also called an estimator.

[1]The l_∞^w norm is defined as $\|y\|_\infty^w = \max_i \{w_i |y_i|,\ w_i > 0$

Some examples are now presented in order to show how specific estimation problems fit into this general framework.

2.2.2. Example 1: Parameter Estimation of ARX Models

Consider the ARX model

$$y_k = \sum_{i=1}^{p} \nu_i y_{k-i} + \sum_{i=1}^{q} \theta_i u_{k-i} + \rho_k \tag{2.7}$$

where y_k is a scalar output, u_k is a known scalar input and ρ_k is an unknown but bounded error such that

$$|\rho_k| \leq \varepsilon_k, \quad \forall k \tag{2.8}$$

Suppose that m values $[y_1, \ldots, y_m]$ are known and the aim is to estimate parameters $[\nu_i, \theta_i]$. For the sake of simplicity suppose that $p \geq q$. Λ can be defined as the $(p + q)$-dimensional space with elements

$$\lambda = [\nu_1, \ldots, \nu_p, \theta_1, \ldots, \theta_q]^T. \tag{2.9}$$

If no *a priori* knowledge on parameter λ is available, then $K = \Lambda$.

Z is the $(p + q)$-dimensional space with elements $z = \lambda$, so that $S(\lambda)$ is identity. Y is the $(m - p)$-dimensional space with elements $y = [y_{p+1}, \ldots, y_m]^T$, and consequently $F(\lambda)$ is linear and is given by

$$F(\lambda) = \begin{bmatrix} y_p & \cdots & y_1 & u_p & \cdots & u_{p+1-q} \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ y_{m-1} & \cdots & y_{m-p} & u_{m-1} & \cdots & u_{m-q} \end{bmatrix} \lambda. \tag{2.10}$$

2.2.3. Example 2: State Estimation of Linear Dynamic Systems

Consider the problem of estimating the state of the following discrete, linear, time invariant dynamic system

$$\begin{cases} x_{k+1} = Ax_k + Bu_k \\ y_k \quad = Cx_k + \rho_k \end{cases} \tag{2.11}$$

where x_k, y_k, u_k and ρ_k are the state, observation, process noise and observation noise vectors respectively; A, B and C are given matrices. For the sake of simplicity, suppose that x is l-dimensional and y, u, and ρ are scalar variables.

Assume that the samples of process and observation noise are unknown but bounded

$$|u_k| \le U_k, \quad \forall k \tag{2.12}$$

$$|\rho_k| \le \varepsilon_k, \quad \forall k. \tag{2.13}$$

Suppose that m values $[y_1, ..., y_m]$ are known and the aim is to estimate x_m. Λ can be defined as the $l + m - 1$-dimensional space with elements

$$\lambda = [x_1^T, u_1, \ldots, u_{m-1}]^T. \tag{2.14}$$

If no *a priori* information on the initial state x_1 is available, K is defined by

$$K = \{\lambda \in \Lambda;\ |u_j| \le U_j, j = 1, \ldots, m-1\} \tag{2.15}$$

Z is the l-dimensional space with elements $z = x_m$. Y is the m-dimensional space with elements $y = [y_1, ..., y_m]^T$. Standard computation of solutions of the set of difference Eq. (2.11) shows that the solution and information operator are linear and are given by

$$S(\lambda) = [A^{m-1}, A^{m-2}B, \quad , AB, B]\lambda \tag{2.16}$$

$$F(\lambda) = \begin{bmatrix} C & 0 & \cdots & 0 & 0 \\ CA & CB & \cdots & 0 & 0 \\ CA^2 & CAB & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ CA^{m-2} & CA^{m-3}B & \cdots & CB & 0 \\ CA^{m-1} & CA^{m-2}B & \cdots & CAB & CB \end{bmatrix} \lambda. \tag{2.17}$$

2.2.4. Example 3: Parameter Estimation of Multiexponential Models

Consider the multiexponential model

$$y(t) = \sum_{i=1}^{l} \mu_i e^{-\nu_i t} + \rho(t) \tag{2.18}$$

where μ_i and ν_i are unknown real parameters and $\rho(t)$ is unknown but bounded by a given $\varepsilon(t)$

$$|\rho(t)| \le \varepsilon(t). \tag{2.19}$$

Suppose that m values $[y(t_1), \cdots, y(t_m)]$ are known and the aim is to estimate parameters μ_i and ν_i, $i = 1, \ldots, l$.

By setting $\xi_i = e^{-\nu_i}$, , l, the space Λ is taken as the $2l$-dimensional space with elements

$$\lambda = [\mu_1, \quad ., \mu_l, \xi_1, \quad ., \xi_l]^T. \tag{2.20}$$

$S(\lambda)$ can be taken as the identity operator. In this way, estimation of variables ξ_i is considered instead of ν_i. Original variables can be obtained by logarithmic transformation.

Y is defined as the m-dimensional space with elements $y = [y(t_1), \ldots, y(t_m)]^T$. Then, information operator $F(\lambda)$ becomes the polynomial function

$$\begin{bmatrix} F_1(\lambda) \\ \cdots \\ F_m(\lambda) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{l} \mu_i \xi_i^{t_1} \\ \cdots \\ \sum_{i=1}^{l} \mu_i \xi_i^{t_m} \end{bmatrix} \tag{2.21}$$

2.2.5. Example 4: Multistep Prediction with ARX Models

Consider the ARX model Eq. (2.7) and suppose that the aim is to estimate y_{m+h} when past values $[y_1, ..., y_m]$ are measured (h-step ahead prediction problem). For the sake of simplicity, consider the case $h = 2$.

The space Λ can be defined as the $p + q + 2$-dimensional space with elements

$$\lambda = [\nu_1, \ldots, \nu_p, \theta_1, \ldots, \theta_q, \rho_{m+1}, \rho_{m+2}]^T. \tag{2.22}$$

If no *a priori* knowledge on parameter λ is available, K is given by

$$K = \{\lambda \in \Lambda;\ |\rho_{m+1}| \le \varepsilon_{m+1}, |\rho_{m+2}| \le \varepsilon_{m+2}\}. \tag{2.23}$$

Z is the 1-dimensional space with elements $z = y_{m+2}$ and consequently $S(\lambda)$ is the polynomial function given by

$$\begin{aligned} S(\lambda) = {} & (\nu_1\nu_1 + \nu_2)y_m + (\nu_1\nu_2 + \nu_3)y_{m-1} + \cdots + \nu_p y_{m-p+1} \\ & + \theta_1 u_{m+1} + (\nu_1\theta_1 + \theta_2)u_m + \cdots + \theta_q u_{m-q+1} \\ & + \nu_1\rho_{m+1} + \rho_{m+2}. \end{aligned} \tag{2.24}$$

Y is an $(m - p)$ dimensional space with elements $y = [y_{p+1}, ..., y_m]^T$ and $F(\lambda)$ is linear and given by

$$F(\lambda) = \begin{bmatrix} y_p & & y_1 & u_p & & u_{p+1-q} & 0 & 0 \\ \cdots & & \cdots & \cdots & & \cdots & \cdots & \cdots \\ y_{m-1} & & y_{m-p} & u_{m-1} & & u_{m-q} & 0 & 0 \end{bmatrix} \lambda. \tag{2.25}$$

2.3. MAIN DEFINITIONS AND CONCEPTS

This section provides definitions of the main sets involved in the theory, optimality concepts used to evaluate estimator's performances, and types of estimators investigated.

2.3.1. Relevant Sets

The following sets play key roles in set membership estimation theory: measurement uncertainty set:

$$MUS_y = \{\tilde{y} \in Y : \|\tilde{y} - y\|_\infty^w \leq \varepsilon\} \tag{2.26}$$

estimate uncertainty set (for a given estimator ϕ);

$$EUS_\phi = \phi(MUS_y) \tag{2.27}$$

feasible problem elements set;

$$FPS_y = \{\lambda \in K : \|y = F(\lambda)\|_\infty^w \leq \varepsilon\} \tag{2.28}$$

and feasible solutions set

$$FSS_y = S(FPS_y). \tag{2.29}$$

Note the difference between EUS_ϕ and FSS_y. The former depends on the particular estimator ϕ used and gives all possible estimated values that could be obtained for all possible measurements consistent with the actual measurement y and the given error bounds. The latter depends only on the problem setting and gives all possible values which are consistent with the available information on the problem.

In the literature on parameter estimation, where problem element λ represents the parameters to be estimated and $S(\lambda)$ is identity (see Section 2.2.2), FPS_y coincides with FSS_y and has been given also different names such as feasible parameters set, membership-set estimate and likelihood set.

An exact description of FSS_y or EUS_ϕ is in general not simple, since they may be very complex sets (e.g. non-convex, not simply connected). For this reason, approximate descriptions are often looked for, using simply shaped sets like boxes or ellipsoids containing (outer bounding) or contained in (inner bounding) the set of interest (see Fig. 2.2). In particular minimum volume outer box (MOB) or ellipsoid (MOE) and maximum volume inner box (MIB) or ellipsoid (MIE) are of interest.

Information of great practical interest is also provided by the values uncertainty intervals (VUI) and estimate uncertainty intervals (EUI), giving the maximum ranges of possible variations of the feasible and the estimated values, respectively. The VUIs are defined as

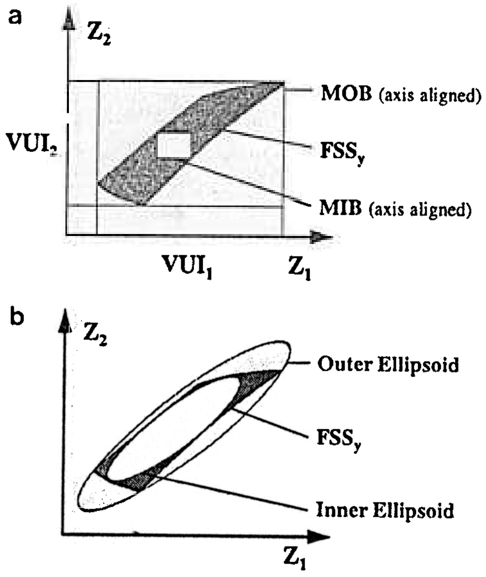

FIGURE 2.2. (a) Box and (b) ellipsoid inner- and outer-bounding.

$$VUI_i = [z_i^m, z_i^M] \quad i = 1, \ldots, l, \tag{2.30}$$

where

$$z_i^m = \inf_{z \in FSS_y} z_i = \inf_{\lambda \in FPS_y} S_i(\lambda) \quad i = 1, \quad , l$$

and

$$z_i^M = \sup_{z \in FSS_y} z_i = \sup_{\lambda \in FPS_y} S_i(\lambda) \quad i = 1, \ldots, l. \tag{2.31}$$

Note that the VUIs are the sizes (along coordinate axis) of the axis aligned box of minimal volume containing FSS_y (see Fig. 2.2).

The EUIs are defined in the same way substituting EUS_ϕ for FSS_y.

2.3.2. Errors and Optimality Concepts

Algorithm performance is measured according to the following errors:

Y-local error $E_y^\varepsilon(\phi)$, where

$$E_y^\varepsilon(\phi) = \sup_{\lambda \in FPS_y} \|S(\lambda) - \phi(y)\| \tag{2.32}$$

Λ-local error $E_\lambda^\varepsilon(\phi)$, where

$$E_\lambda^\varepsilon(\phi) = \sup_{y \in MUS_{F(\lambda)}} \|S(\lambda) - \phi(y)\| \tag{2.33}$$

and global error $E^{\varepsilon}(\phi)$

$$E^{\varepsilon}(\phi) = \sup_{y \in Y} E^{\varepsilon}_{y}(\phi) = \sup_{\lambda \in \Lambda} E^{\varepsilon}_{\lambda}(\phi). \tag{2.34}$$

Dependence on ε is dropped out in subsequent notation, except when necessary.

Algorithms minimizing these types of errors are indicated respectively as Y-locally, Λ-locally and globally optimal.

Notice that Y-local optimality is of particular interest in system identification problems, where a set of measurements y is available and one wants to determine the best achievable estimate $S(\lambda)$ for each possible y using an algorithm $\phi(y)$. Also Λ-local optimality is a particularly meaningful property in estimation problems, since it ensures the minimum uncertainty of the estimates for the worst measurement y, for any possible element $\lambda \in K$.

Y- and Λ-local optimality are stronger properties than global optimality, as can be seen from Eq. (2.34). For example, a Y-locally optimal algorithm minimizes the local error $E_y(\phi)$ for all data y, while a globally optimal algorithm minimizes the global error $E(\phi)$ only for the worst data. In other words, a Y-locally optimal algorithm is also globally optimal, while the converse is not necessarily true.

2.3.3. Classes of Estimators

Some classes of estimators whose properties have been investigated in the literature are now introduced.

The first class is related to the idea of taking the Chebicheff center of FSS_y as estimate of $S(\lambda)$. The center of FSS_y, $c(FSS_y)$, and the corresponding radius, $rad(FSS_y)$, are defined by

$$\sup_{z \in FSS_y} \|c(FSS_y) - z\| = \inf_{\tilde{z} \in Z} \sup_{z \in FSS_y} \|\tilde{z} - z\| = rad(FSS_y). \tag{2.35}$$

A central estimator ϕ^c is defined as

$$\phi^c(y) = c(FSS_y) \tag{2.36}$$

The second class includes estimators analogous to unbiased estimators in statistical theory, which give exact values if applied to exact data.

An estimator ϕ is correct if

$$\phi(F(\lambda)) = S(\lambda) \quad \forall \lambda \in \Lambda. \tag{2.37}$$

Such a class is meaningful only for $l \leq m$, that is, when there are at least as many measurements as variables to be estimated (the typical situation in estimation practice). This class contains most of the commonly used estimators, such as projection estimators.

A projection estimator ϕ^p is defined as

$$\phi^p(y) = S(\lambda_y) \tag{2.38}$$

where $\lambda_y \in \Lambda$ is such that

$$\|y - F(\lambda_y)\| = \inf_{\lambda \in \Lambda} \|y - F(\lambda)\|. \tag{2.39}$$

The most widely investigated and used estimators in this class are least square estimators (ϕ^{LS}), which are projection estimators when an l_2 norm is used in space Y. Least-absolute values and least-maximum value estimators have been also considered in the literature, which are projection estimators when l_1 and l_∞ norms are respectively used in space Y.

In the next sections the results available in the literature regarding the following aspects are reviewed: existence and characterization of estimators, optimal with respect to some of the optimality concepts introduced previously; actual computation of the derived optimal estimators; evaluation of the errors of optimal and of projection estimators; and exact or approximate description of feasible sets FPS_y, FSS_y and estimate uncertainty set EUS_ϕ. Whenever possible, a statistical counterpart of the presented results is indicated, based on the analogy:

Y-local optimality $\Leftrightarrow$ minimum variance optimality

$FSS_y \Leftrightarrow$ minimum variance estimate pdf

$EUS_\phi \Leftrightarrow$ estimate pdf

EUI's $\Leftrightarrow$ estimate confidence intervals

VUIs $\Leftrightarrow$ Cramer-Rao lower bound confidence intervals.

2.4. NONLINEAR PROBLEMS

A first important result is related to the existence of a Y-locally optimal estimator. No general results are available for Λ-locally optimal estimators. This result also shows that the minimum Y-local error is given by the radius of FSS_y.

Result 1. [4,6] A central estimator ϕ^c is Y-locally optimal

$$E_y(\phi^c) \leq E_y(\phi) \quad \forall y \in Y, \forall \phi \tag{2.40}$$

Its Y-local error is

$$E_y(\phi^c) = rad(FSS_y) \tag{2.41}$$

This result holds for any norm in Λ, Z, Y. □

It can be considered as the counterpart of the conditional mean theorem in statistics. As with conditional mean estimators, central estimators are in general difficult to compute. The computation of ϕ^c involves the knowledge of FSS_y, which may be a very complex set (nonconvex, not simply connected).

Several approaches have been proposed to describe FSS_y, mainly in papers related to dynamic system parameter estimation. In Ref. 7 a random sample of parameters is generated by a Monte Carlo technique, and Eqs. (2.4 and 2.5) are used to check if they belong to FSS_y. Global optimization methods based on random search are used in Refs. 8 and 9 to construct the boundary of FSS_y. In Ref. 8 projections of FSS_y onto coordinate one-dimensional or two-dimensional subspaces are looked for. In Ref. 9 intersections of the boundary of FSS_y with bundles of straight lines centered at points inside FSS_y are searched. The optimization methods used in these papers converge in probability to the global maximum or minimum of interest. However, this convergence property is not very useful in practice, because no estimate is given of the distance of the achieved solution from the global solution. Moreover, all these approaches suffer the curse of dimensionality. These reasons motivate the interest in looking for less detailed but more easily computable information on FSS_y.

An important result in this direction is that the computation of ϕ^c and of its Y-local error do not require the exact knowledge of FSS_y, but only of the VUIs.

Result 2.[5] The center $c(FSS_y)$ can be computed as

$$c_i(FSS_y) = (z_i^M + z_i^m)/2 \quad i = 1, \ldots, l \tag{2.42}$$

The radius $rad(FSS_y)$ can be computed as

$$rad(FSS_y) = \max_i(z_i^M - z_i^m)/2 \tag{2.43}$$

where z_i^m and z_i^M are given by Eq. (2.31). □

Result 2 states that the computation of a central algorithm and of minimum Y-local error is equivalent to the computation of the VUIs, requiring the solution of only $2l$ optimization problems of the type Eq. (2.31).

Equation (2.31) problems are in general not convex, exhibiting local extrema. Any of the general global optimization algorithms available in the literature give approximate solutions converging to the exact ones only in probability and, more seriously, they do *not* provide any assessment on how far the approximate solution is from the correct one.

If $S(\lambda)$ and $F(\lambda)$ are polynomial functions, specific global algorithms exist, for obtaining better results.

Result 3.[10] If $S(\lambda)$ and $F(\lambda)$ are polynomial, algorithms exist converging with certainty to global extrema of Eq. (2.31).

Under the assumptions of Result 3, Eq. (2.31) are polynomial optimization problems, in the sense that both cost functions and constraints are polynomials in λ. Polynomial problems are in general nonconvex and may admit local extrema.[11] Nevertheless, if all the variables are strictly positive (in which case the term signomial problems is used), an algorithm is available to find a global maximum.[10,12,13] The underlying idea of this algorithm is to construct a sequence of convex problems approximating the original problem iteratively better. In this way, the algorithm generates a sequence of lower and upper bounds of the global extremum, converging monotonically to it. If the sign of some of the variables is not known, it is possible to reduce a polynomial problem to a signomial problem by setting these variables as the difference of strictly positive new variables.

The hypothesis of Result 2 covers large classes of problems, as shown in examples (2.2.2–2.2.5). The implication is that an optimal estimator and its error can be exactly computed for several nonlinear problems of practical interest. No analogous result is available in the statistical context.

Most of the papers in the literature focus on studying FSS_y, while very few results are available on EUS_ϕ. For any correct estimator, FSS_y is an inner bounding set of EUS_ϕ.[14]

Result 4.[14] If ϕ is correct then

$$FSS_y \subseteq EUS_\phi \quad \forall\, y \in Y \tag{2.44}$$

□

Hence, for correct estimators the VUIs are lower bounds of the EUIs, that is,

$$VUI_i \subseteq EUI_i \quad i = 1, \ldots, l \tag{2.45}$$

Consider the properties of projection estimators. In general they are not optimal with respect to any of the three considered type of errors.[15] However they are almost Y-locally optimal (within a factor 2) as shown by the following result.

Result 5.[15] A projection algorithm ϕ^P is such that

$$E_y(\phi^P) \le 2\, rad(FSS_y) \le 2E_y(\phi) \quad \forall\, y \in Y,\ \forall\, \phi \tag{2.46}$$

□

Projection estimators enjoy interesting properties of robustness with respect to inexact knowledge of the uncertainty bound ε. Central estimators are not robust in such a sense: a central algorithm computed supposing that $\varepsilon = \varepsilon_0$ may not be optimal if the actual ε is different. A central estimate $\phi^c(y)$ may not even belong to the actual FSS_y and its Y-local error $E_y(\phi^c)$ may be greater than 2 $rad(FSS_y)$.

On the contrary, projection estimators are robustly almost Y-locally optimal, independent of the volume of ε, as shown by the next result.

Result 6.[16] Let ϕ^P be the projection estimator. Then

$$E_y^\varepsilon(\phi^P) \le 2\, rad(FSS_y) \le 2E_y^\varepsilon(\phi) \quad \forall y \in Y, \forall \phi, \forall \varepsilon \tag{2.47}$$

□

Projection estimators also have nice properties in a statistical context. For example, an l_2-projection estimator is the maximum likelihood estimator (MLE) if noise ρ is supposed to be gaussian; an l_1-projection estimator is the MLE if noise is supposed to have a Laplace pdf; an l_∞-projection estimator is the MLE if noise is uniformly distributed. Projection estimators l_1 and l_∞ also have interesting robustness properties with respect to uncertainty in the pdf's knowledge.[17,18,19]

2.5. LINEAR PROBLEMS

Consider the case in which $S(\lambda)$ and $F(\lambda)$ are linear. In this case, Eq. (2.4) is written as

$$y = \mathcal{A}\lambda + \rho \tag{2.48}$$

where $\mathcal{A}$ is a matrix of dimension (m, n).

These assumptions are restrictive, but include cases of practical interest such as parameter estimation of linear regressions, parameter estimation of ARMA models with polynomial trends and harmonic components, state estimation of dynamic systems, and time series forecasting. Moreover, if uncertainty bounds are not too large, linear theory can be used for a first approximate analysis using some linearization techniques.

From Result 1 a central estimator is Y-locally and globally optimal. In the linear case it is also correct and Λ-locally optimal in the class of correct estimators, as shown in the next result.

Result 7.[20] ϕ^c is Y-locally optimal:

$$E_y(\phi^c) \le E_y(\phi) \quad \forall\, \phi \tag{2.49}$$

ϕ^c is a Λ-locally optimal (among correct estimators)

$$E_\lambda(\phi^c) \le E_\lambda(\phi) \quad \forall\, \lambda \in K, \forall\, \phi \text{ correct} \tag{2.50}$$

□

In Ref. (15) it is proven that Result 7 holds for any norm in Y.

Under the present assumptions, FSS_y and FPS_y are polytopes. Then from Result 2 it follows that ϕ^c and its Y-local error $E_y(\phi^c)$ can be obtained by solving the $2l$ linear programming problems of Eq. (2.31).

A linear estimator can be computed, which is correct, globally optimal, and Λ-locally optimal within the class of correct estimators. This gives a complete solution to the linear case, representing the counterpart of the Gauss-Markov theory in statistical estimation.

Result 8.[5,14] Let $K = \Lambda$ and $m \ge n$. Then there exists a linear estimator H^* that is correct and globally optimal

$$E(H^*) \le E(\phi) \quad \forall\, \phi \tag{2.51}$$

The linear estimator H^* is Λ-locally optimal (among correct estimators)

$$E_\lambda(H^*) \le E_\lambda(\phi) \quad \forall\, \lambda \in \Lambda, \forall\, \phi \text{ correct} \tag{2.52}$$

Its errors are

$$E(H^*) = E_\lambda(H^*) = E_{\mathcal{A}\lambda}(H^*) = rad(FSS_{\mathcal{A}\lambda}) \quad \forall\, \lambda \in \Lambda \tag{2.53}$$

□

Estimator H^* can be computed from the knowledge of the active constraints of the linear programming problems of Eq. (2.31) with $y = 0$.[5,14]

In case that an l_2-norm is used in Y, H^* can be computed by least squares. Under this assumption, the least squares estimator is linear and correct, robustly Y-locally optimal and Λ-locally optimal within the class of correct estimators, as shown by the next result.

Result 9.[15] Let $K = \Lambda$, $m \ge n$ and Y be a Hilbert space. Let ϕ^{LS} be the projection (least square) estimator. Then:

ϕ^{LS} is central, linear, correct and robustly Y-locally optimal

$$E_y^\varepsilon(\phi^{LS}) \le E_y^\varepsilon(\phi) \quad \forall\, y \in Y, \forall\, \phi, \forall\, \varepsilon \tag{2.54}$$

ϕ^{LS} is Λ-locally optimal (among correct estimators)

$$E_\lambda(\phi^{LS}) \le E_\lambda(\phi) \quad \forall\, \lambda \in \Lambda, \forall\, \phi \text{ correct} \tag{2.55}$$

□

The sets FPS_y, FSS_y and EUS_ϕ, (for linear ϕ), are polytopes described by the sets of linear inequalities appearing in Eqs. (2.27–2.29). This is not the simplest way to describe them (for example, many linear inequalities may not concur to defining the boundary of the polytope) and simpler descriptions could be of interest. One way of characterizing a polytope P is through its vertices. Algorithms exist which allow one to compute recursively the vertices of a polytope P_k, defined by the first k measurements, from the knowledge of P_{k-1} and the k-th measurement.[21,22,23,24] The number of vertices may be relatively smaller than the theoretical maximum. For example, Monte Carlo simulations on ARMA models parameter estimation,[23] have shown that the mean number of vertices of FSS_y is approximately constant for $m > 50$. For $l = 4$ and $l = 5$, for example, they are approximately 50 and 150, respectively.

Polytopes can be represented alternatively by describing their faces. This representation is used to derive a recursive algorithm.[25] This approach seems more involved than the previous one, but it also allows the recursive computation of an outer bounding polytope with a fixed number of faces, leading to an approximating description of the polytope of interest by means of a polytope of prescribed complexity.

The most investigated approaches to approximate description of polytopes are through ellipsoids and boxes for the case of parameter estimation, where the polytope of interest is the feasible parameter set.

A recursive algorithm for outer bounding ellipsoid computation has been proposed in.[26] The underlying idea is as follows.

Let OE_{k-1} be the outer ellipsoid bounding P_{k-1}. Let R_k be the feasible parameter set corresponding only to the k-th measurement

$$R_k = \{\lambda \in \Lambda : y_k - \varepsilon_k \leq a_k^T\lambda \leq y_k + \varepsilon_k\} \tag{2.56}$$

where a_k^T is the k-th row of $\mathcal{A}$.

Clearly $P_k \subseteq OE_{k-1} \cap R_k$. OE_k is computed as the minimal volume ellipsoid containing $OE_{k-1} \cap R_k$, and then containing P_k also.

If an ellipsoid OE_k is defined by its centers λ_k^c and positive definite matrix Σ_k according to

$$OE_k = \{\lambda \in \Lambda : (\lambda - \lambda_k^c)^T \Sigma_k^{-1} (\lambda - \lambda_k^c) \leq 1\} \tag{2.57}$$

the following recursive algorithm has been obtained.

Result 10.[26] The ellipsoid OE_k can be computed by the recursion

$$\lambda_k^c = \lambda_{k-1}^c + \frac{\sigma_k V_k a_k \nu_k}{\varepsilon_k^2} \tag{2.58}$$

$$\Sigma_k = (1 + \sigma_k - \frac{\sigma_k \nu_k^2}{\varepsilon_k^2 + \sigma_k \mu_k}) V_k \tag{2.59}$$

where

$$V_k = \Sigma_{k-1} - \frac{\sigma_k \Sigma_{k-1} a_k a_k^T \Sigma_{k-1}}{\varepsilon_k^2 + \sigma_k \mu_k} \tag{2.60}$$

$$\nu_k = y_k - a_k^T \lambda_{k-1} \tag{2.61}$$

$$\mu_k = a_k^T \Sigma_{k-1} a_k \tag{2.62}$$

and σ_k is the positive solution of the equation

$$(l-1)\mu_k^2\sigma_k^2 + [(2l-1)\varepsilon_k^2 - \mu_k + \nu_k^2]\mu_k\sigma_k + \varepsilon_k^2[l(\varepsilon_k^2 - \nu_k^2) - \mu_k] = 0 \tag{2.63}$$

if a positive solution exists, otherwise $\sigma_k = 0$. □

Computational complexity of this algorithm and slight modifications for implementation on a systolic architecture can be found.[27] A modification of this

algorithm with data-dependent updating and forgetting factor has been proposed.[28]

A similar approach can be used for the recursive computation of inner bounding ellipsoids.[29,30] Let IE_{k-1} the inner bounding ellipsoid contained in P_{k-1}. Then IE_k is chosen as the maximal volume ellipsoid such that

$$IE_k \subseteq IE_{k-1} \cap R_k \subseteq P_k \tag{2.64}$$

The resulting recursive algorithm is much as for the outer bounding ellipsoid and is not reported here.

The main drawback of these recursive algorithms is that they do *not* give the minimal and maximal volume ellipsoids bounding the feasible parameter set.[29,31] This is true also for improved versions of the algorithm.[31,32] Since IE_k has an unfortunate tendency to shrink rapidly and vanish,[30] the inclusion $IE_k \subset P_k \subset OE_k$ in practice may not give any reasonable information on the looseness of bound OE_k.

A nonrecursive solution to the problem of finding the minimal volume outer ellipsoid contained in FPS_y (MOE_{FPS}) and the maximal volume inner ellipsoid contained in FPS_y (MIE_{FPS}), has been proposed.[33,34] The solution for MIE_{FPS} is given by the following result.

Result 11.[33] The MIE_{FPS} has center λ^{c*} and matrix Σ^* solution of

$$\max \ \det(\Sigma) \tag{2.65}$$

subject to

$$\lambda^c,\Sigma: \begin{cases} (u_i^T\lambda^c + c_i)^2 - u_i^T\Sigma u_i \geq 0, i = 1, \ldots, 2m \\ u_i^T\lambda^c + c_i \geq 0, i = 1, \ldots, 2m \\ \Sigma_i > 0, i = 1, \ldots, n \end{cases}$$

where Σ_i, $i = 1, \ldots, n$ are the principal minors of Σ, and matrix $U \in \mathbb{R}^{(2m,n)}$ (with rows denoted by u_i^T) and vector $c \in \mathbb{R}^{2m}$ are given by

$$U = [A^T | -A^T]^T; c = [-y_-^T | y_+^T]^T \tag{2.66}$$

$$y_- = [y_1 - \varepsilon w_1, y_2 - \varepsilon w_2, \ldots, y_m - \varepsilon w_m]^T$$

$$y_+ = [y_1 + \varepsilon w_1, y_2 + \varepsilon w_2, \ldots, y_m + \varepsilon w_m]^T \tag{2.67}$$

□

Equation (2.65) is a polynomial optimization problem and can be solved by use of signomial programming.[10] The solution of Eq. (2.65) may be computationally cumbersome, even for a few parameters. Then less general but simpler solutions may be of interest. For example, the maximum ellipsoid of given shape may be sought. Consider that Σ is given except for a scale factor (for example the shape of the outer ellipsoid given by Result 10 can be used). In such a case, Eq. (2.65) reduces to a linear programming problem with $(n + 1)$ variables and $(2m + 1)$ constraints.

The solution for MOE_{FPS} also can be obtained by solving a suitable polynomial problem.[34] Unfortunately, the computational complexity is high for the general case, and does not reduce, as for MIE_{FPS}, if restricted classes of ellipsoids are considered.

For the computation of extremal volume inner and outer boxes definitions are as follows.

A box is defined as:

$$B(\lambda^c, l, R) = \{\lambda \in \Lambda : \|R(\lambda - \lambda^c)\|_\infty^l \leq 1\} \tag{2.68}$$

where R is an orthonormal matrix. The box is described by its center λ^c, axis lengths l_i and rotation R. If $R = I$ the box is aligned with coordinate axis.

A solution to the problem of finding the minimal volume outer box contained in FPS_y (MOB_{FPS}) is provided in Ref. 34 as solution of a suitable polynomial problem. Its computational complexity is high, unless the rotation of the box is given. In such a case the problem can be reduced to a linear programming problem. In particular, if $R = I$ the axis-aligned MOB_{FSS} can be computed directly from Eq. (2.31). This requires the solution of $2l$ linear programming problems with n variables and $2m$ inequalities constraints.

The solution to the problem of finding the maximal volume inner box contained in FPS_y (MIB_{FPS}) is provided by the following result.

Result 12.[33] The MIB_{FPS} has center λ^{c*}, axis length l^* and rotation R^* solution of

$$\max \prod_{i=1}^{n} l_i \tag{2.69}$$

subject to

$$\lambda^c, l, R : \begin{cases} l_i > 0, i = 1, \ldots, n \\ (u_i^T R^T \lambda^c + c_i) - \Sigma_{j=1}^n l_j |u_{ij}| \geq 0, i = 1, \ldots, 2m \\ R^T R = I \end{cases} \qquad \square$$

Equation (2.69) is a polynomial optimization problem which can be solved by use of signomial programming. If matrix rotation R is fixed, Eq. (2.69) reduces to a convex problem with $2n$ variables and $(2m + n)$ linear constraints, which can be efficiently solved by means of normally available convex programming algorithms. If axis length l is also fixed except for a scale factor (i.e., the maximum box of a given shape is sought), Eq. (2.69) reduces to a linear programming problem with $(n + 1)$ variables and $(2m + 1)$ constraints.

2.6. OTHER TYPES OF RESULTS

This section briefly recalls papers on topics related to set membership estimation theory, such as experiment design, estimation with reduced order models, and uncertainty in the information operator. Almost all these papers consider linear problems.

2.6.1. Experiment Design

In the previous sections information operator $\mathcal{A}$ is supposed to be given. In some practical application it is possible to choose among different information operators $\mathcal{A}$ (optimal information problem). For example it may be possible to choose the sampling times at which measurements are taken of the input and the output of the dynamic system to be identified. Then a natural choice is the one minimizing the error $E_\lambda(\phi^c)$. In Ref. 35 some results are given for the case in which information is provided by sampling.[35] In Ref. 20 similar results are derived for more general classes of information. In this paper it is also shown that the optimal sampling times can be chosen *a priori*, and no improvements can be obtained by means of more sophisticated sampling schemes.[20] The optimal sampling problem is approached through p-widths theory.[36]

Another criterion is to minimize the volume of FPS_y.[29] In Ref. 29 a recursive selection procedure is given based on heuristics to avoid poor choices without guaranteeing the best. Characterization is given of the minimum number of sampling times assuring minimum volume of the feasible parameter set FPS_y for $y = \mathcal{A}\lambda$ in Ref. 37.

2.6.2. Reduced Order Models

In the previous sections, it is supposed that the structure of the problem is known, for example the number of autoregressive and moving-average terms for an ARMA model. In many cases, however, the structure of the problem and in particular the dimension of space Λ is not known and must be evaluated from the available information (order determination problems). Some methods are analogous to methods widely used for order determination in statistical contexts,[38,39] such as the principal component analysis and singular value decomposition. A method is also proposed, based on the expected behavior of FPS_y for overparameterized and underparameterized structures.

A second important problem is how estimation algorithms can take into account that only approximating structures are used. The usual approach in statistical contexts is to ignore the deterministic nature of modeling errors, and eventually discard badly approximating structures with residuals evidently not satisfying the assumed statistical hypotheses. In the UBB approach, modeling errors can be taken into account in a more natural way, since it is possible to evaluate bounds on such modeling errors.[40,41,42] A deeper analysis considers explicitly that using approximating structures corresponds to restricting the analysis to a subset $K \subset \Lambda$ *not* containing the "true" problem element λ.[43] In this paper, the concept of conditionally central estimator is introduced as an extension of a central estimator, and it is shown to be Y-locally optimal. The same paper shows that there are two possible ways of extending least squares estimators. The first one corresponds to what is usually done (more or less explicitly) when dealing with reduced order models. However, this estimator does not preserve any of the interesting optimality properties of least squares estimators. A second type of extension is introduced, which is shown to have interesting Λ-locally and Y-locally optimality properties.

2.6.3. Uncertain Information Operator

In some papers the case in which information matrix $\mathcal{A}$ is not exactly known is studied. In particular, perturbation of the type $\mathcal{A} = \mathcal{A}_o + \Delta\mathcal{A}$ has been considered, where $\mathcal{A}_o$ is given and $\Delta\mathcal{A}$ is not known but bounded. A modification of the recursive algorithm for outer ellipsoid bounding reported in Result 10 is proposed.[44] Two different extensions of FPS_y are considered in Refs. 45 and 46. FPS_y^1 is defined in Ref. 45 by considering that Eq. (2.28) holds for all $\Delta\mathcal{A}$ and is described by a set of $m2^{n+1}$ linear inequalities. In Ref. 46, FPS_y^2 is defined by considering that Eq. (2.28) holds for some $\Delta\mathcal{A}$, and the problem of finding the corresponding MOB by means of suitable linear programming problems is also discussed.

2.7. APPLICATIONS

The UBB approach is now beginning to be used in various application fields. Some papers report applications to real word problems arising in biology,[47,48] pharmacokinetics,[14,49] time series filtering and prediction,[50,51] economics,[52] chemistry,[53] image processing,[54] ecology,[55,56] measurement,[8,57,58] tracking,[59] and speech processing.[27,60,61,62]

Application of set membership estimation theory has also been investigated in the context of identification for robust and adaptive control design,[28,63,64,41,42,65] and in Chapters 27–30 of this volume.

2.8. CONCLUSIONS

In this chapter an outline of the main results in the area of estimation theory for set membership uncertainty has been presented. The main emphasis of the paper is on the following aspects: existence and characterization of worst-case optimal estimators; actual computation of the derived optimal estimators; evaluation of the errors of optimal and of other widely used estimators (least squares, least absolute values, least maximum values); exact or approximate description of feasible parameter and solution sets. A quick reference to less assessed topics such as experiment design, reduced-order modeling, and more general error models are also made in the paper.

Some general considerations may be drawn from this overview.

Concerning linear problems, real advances have been done in the last decade. As a result, properties of estimators and exact or approximate description of feasible parameter and solution sets can be considered subjects with reasonably well understood and usable solutions. In fact, many of the available algorithms have been used for several applications in different real world problems.

Concerning nonlinear problems, in spite of the work done in the last few years, much more remains to be done. Some algorithms for computing exact parameter or solution uncertainty intervals have been proposed. They work reasonably well on problems with a limited number of measurements and parameters. However, their behavior in more complex situations has not been deeply investigated yet.

Several basic problems remain open and need a thorough investigation, both for linear and nonlinear problems, for example the topological properties of the feasible parameter set as a function of the nonlinearity and uncertainty structures, inner and outer bounding for the nonlinear case, the effects of model approximations, the interaction of set membership estimation theory and robust or adaptive control.

ACKNOWLEDGMENTS. This work was partially supported by Ministero della Università e della Ricerca Scientifica and EC project SIMONET.

REFERENCES

1. H. S. Witsenhausen, *IEEE Trans. Autom. Control* **AC-13**, 556 (1968).
2. F. C. Schweppe, *IEEE Trans. Autom. Control* **AC-13**, 22 (1968).
3. F. C. Schweppe, *Uncertain Dynamic Systems*, Prentice Hall, Englewood Cliffs, NJ (1973).
4. J. F. Traub and H. Wozniakowski, *A General Theory of Optimal Algorithms*. Academic Press, New York (1980).
5. M. Milanese and R. Tempo, *IEEE Trans. Autom. Contr.* **AC-30**, 730 (1985).
6. C. A. Micchelli and T. J. Rivlin, in: *A Survey of Optimal Recovery* (C.A. Micchelli and T. J. Rivlin, eds.) *Optimal Estimation in Approximation Theory*, pp. 1–54; Plenum Press, New York (1977).
7. K. Keesman and G. van Straten, *Proceedings of the 12th IMACS World Congress*, Paris (1988).
8. M. K. Smit, *Measurement* **1**, 181 (1983).
9. E. Walter and H. Piet-Lahanier, *Proceedings of the 25th IEEE Conference on Decision and Control*, Athens (1986).
10. M. Milanese and A. Vicino, *Automatica* **27**, 403 (1991).
11. J. G. Ecker, *SIAM Review* **1**, 339 (1980).
12. J. E. Falk, Tech. Rep. T-274, George Washington University, Washington DC (1973).

13. M. Milanese and A. Vicino, in: *Robust Estimation and Exact Uncertainty Intervals Evaluation for Nonlinear Models*, of *Systems Modelling and Simulation* (S. Tzafetas, A. Eisinberg, and L. Carotenuto, eds.) Elsevier Science Publishers (1988).
14. M. Milanese and G. Belforte, *IEEE Trans. Autom. Control* **AC-27**, 408 (1982).
15. B. Z. Kacewicz, M. Milanese, R. Tempo, and A. Vicino, *Systems and Control* **8**, 161 (1986).
16. R. Tempo and G. Wasilkowski, *Systems and Control* **10**, 265 (1988).
17. R. L. Launer and G. N. Wilkinson, Eds. *Robustness in Statistics*, Academic Press, New York (1979).
18. B. T. Poljak and J. Z. Tsypkin, *Automatica* **16**, 53 (1980).
19. A. van den Bos, *Automatica* **24**, 803 (1985).
20. M. Milanese, R. Tempo, and A. Vicino, *J. Complexity* **2**, 78 (1986).
21. V. V. Kapitonenko, *Autom. Remote Control* **22**, 166 (1982).
22. H. Piet-Lahanier and E. Walter, *Proceedings of the 12th IMACS World Congress*, Paris (1988).
23. S. H. Mo and J. P. Norton, *Proceedings of the 12th IMACS World Congress*, Paris (1988).
24. H. Piet-Lahanier and E. Walter, *Proceedings of the 28th IEEE Conference on Decision and Control*, Tampa (1989).
25. V. Broman and M. J. Shensa, *Proceedings of the 12th IMACS World Congress*, Paris (1988).
26. E. Fogel and F. Huang, *Automatica* **18**, 140 (1982).
27. J. R. Deller, *IEEE Acoust., Speech, Signal Process* **6**, 4 (1989).
28. S. Dasgupta and Y. F. Huang, *IEEE Trans. Inf. Theory* **IT-33**, 383 (1987).
29. J. P. Norton, *Automatica* **23**, 497 (1987).
30. J. P. Norton, *Int. J. Control* **50**, 2623 (1989).
31. G. Belforte and B. Bona, *7th IFAC Symposium on Identification and System Parameter Estimation*, York (1985).
32. L. Pronzato, E. Walter, and H. Piet-Lahanier, *Proceedings of the 28th IEEE Conference on Decision and Control*, Tampa (1989).
33. A. Vicino and M. Milanese, *Proceedings of the 28th IEEE Conference on Decision and Control*, Tampa (1989); Also in *IEEE Trans. Autom. Control* **36**, 759 (1991).
34. A. Vicino and M. Milanese, *9th IFAC/IFORS Symposium on Identification and System Parameter Estimation*, Budapest (1991).
35. G. Belforte, B. Bona, and S. Frediani, *IEEE Tans. Autom. Control* **AC-32**, 179 (1987).
36. C. A. Micchelli, in: *Robustness in Identification and Control* (M. Milanese, R. Tempo, and A. Vicino, eds.) Plenum Press, New York (1989).
37. L. Pronzato and E. Walter, *Automatica* **25**, 383 (1989).
38. J. P. Norton, in: *Robustness in Identification and Control* (M. Milanese, R. Tempo, and A. Vicino, eds.) Plenum Press, New York (1989).
39. S. M. Veres and J. P. Norton, *Int. J. Control* **50**, 639 (1989).
40. R. Genesio and M. Milanese, *IEEE Trans. Autom. Control* **AC-21**, 118 (1976).
41. M. K. Lau, R. L. Kosut, and S. Boyd, *29th IEEE CDC*, Honolulu (1990).
42. J. M. Krause, G. Stein, and P. P. Khargonekar, *29th IEEE CDC*, Honolulu (1990).
43. B. Z. Kacewicz, M. Milanese, and A. Vicino, *J. Complex.* **4**, 73 (1988).
44. T. Clement and S. Gentil, *Proceedings of the 12th IMACS World Congress*, Paris (1988).
45. R. Tempo, B. R. Barmish, and J. Trujillo, *Proceedings of the 8th IFAC Symposium on Identification and Parameter Estimation*, Beijing (1988).
46. G. Belforte, B. Bona, and V. Cerone, *Proceedings of the 12th IMACS World Congress*, Paris (1988).
47. G. Belforte, B. Bona, and M. Milanese, *CRC J. Biomed. Eng.* **10**, 275 (1983).
48. J. P. Norton, *Proceedings of the 25th IEEE Conference on Decision and Control*, Athens (1986).
49. R. Gomeni, H. Piet-Lahanier, and E. Walter, *Proceedings of the 3rd IMEKO Congress on Measurements in Clinical Medicine*, Edinburgh (1986).
50. A. Vicino, R. Tempo, R. Genesio, and M. Milanese, *Int. J. Forecasting* **2**, 313 (1987).

51. R. Tempo, *IEEE Trans. Autom. Contr.* **AC-33**, 864 (1988).
52. M. Milanese, R. Tempo, and A. Vicino, *Int. J. Systems Sci.* **19**, 1189 (1988).
53. E. Walter, Y. Lecourtier, J. Happel, and J. Y. Kao, *AIChE J.* **32** , 1360 (1986).
54. A. Venot, L. Pronzato, E. Walter, and J. F. Lebruchec, *Automatica* **22**, 105 (1986).
55. G. Van Straten, *Appl. Math. Comp.* **17**, 459 (1985).
56. K. J. Keesman and G. Van Straten, *Proceedings of the IAWPRC Symposium on System Analysis in Water Quality Management*, Pergamon Press, New York (1987).
57. G. Belforte, B. Bona, E. Canuto, F. Donati, F. Ferraris, I. Gorini, S. Morei, M. Peisino, and S. Sartori, *Annals CIRP* **36**, 359 (1987).
58. J. W. Verhoof and M. K. Smit, *Proceedings of the 12th IMACS World Congress*, Paris (1988).
59. V. Broman and M. J. Shensa, *Proceedings of the 25th IEEE Conference on Decision and Control*, Athens (1986).
60. J. R. Deller and T. C. Luk, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing*, Dallas (1987).
61. J. R. Deller and D. Hsu, *IEEE Trans. Circ. Syst.* **CAS-34**, 782 (1987).
62. P. L. Combettes and H.J. Trussell, *Proceedings of the IEEE International Conference Acoustics, Speech and Signal Processing*, New York (1988).
63. R. L. Kosut, *Int. J. Adapt. Control Signal Proc.* **2**, 371 (1988).
64. A. Vicino, A. Tesi, and M. Milanese, *IEEE Trans. Autom. Control* **AC-35**, 835 (1990).
65. G. Fiorio, S. Malan, M. Milanese, and A. Vicino, *29th IEEE CDC*, Honolulu (1990).