JOURNAL OF COMPLEXITY 4, 73–85 (1988)

Conditionally Optimal Algorithms and Estimation of Reduced Order Models

B. Z. KACEWICZ

Institute of Informatics, University of Warsaw, Warsaw, Poland

AND

M. MILANESE AND A. VICINO

Dipartimento di Automatica e Informatica, Politecnico di Torino, Torino, Italy

The paper presents some extensions of the optimality results obtained in previous work on algorithms used in the field of system identification in the light of information-based complexity. In particular, a class of conditional algorithms is defined by means of a restriction on the space of solution elements and a corresponding conditional worst case error is introduced. We define conditional central algorithms and show their optimality. A conditional central algorithm is then constructed by modifying a projection algorithm and obtaining in this way a conditional projection algorithm. This algorithm is shown to enjoy local optimality properties with reference to the problem element space within the class of conditionally correct algorithms. Finally, it is shown how these results can be used to handle the problem of reduced order model estimation.

1. INTRODUCTION

In past years it has been shown that information-based complexity (Traub and Woźniakowski, 1980; Traub *et al.*, 1983) may prove very useful in dealing with problems in systems and control areas such as system identification, parameter and state estimation, prediction (Milanese and Tempo, 1985; Milanese *et al.*, 1986; Kacewicz *et al.*, 1986; Vicino *et al.*, 1987). The results obtained so far, as well as most of the results derived by means of other approaches (e.g., based on statistical estimation theory), can be applied under the so-called "standard conditions" (Genesio and Milanese, 1979). These conditions essentially state

0885-064X/88 $3.00

that the unknown "true system" which generates the available information belongs to the class of mathematical models selected to represent the actual data. Despite the fact that this assumption is seldom verified in practical applications, only recently some effort has been made to deal with this problem (see the references in Genesio and Milanese (1979) for further details).

This paper is a first attempt at solving the problem within the information-based complexity setting and can be outlined as follows. One is interested in evaluating the image $S(f)$ of an element f belonging to a linear space F under an operator S. It is assumed that S mapping F into a linear space G is known, while f is not known but only approximate information y about f is given as $y = N(f) + \eta$. N, called the information operator, maps F into Y and η belongs to a bounded set of Y. An approximation of $S(f)$ is derived by applying a suitable operator φ (called the algorithm) to the actual information y. Algorithms with optimality properties are sought among all algorithms mapping Y into a *given subset* G_0 *of* G, where G_0 may not contain $S(f)$, as usually considered in the literature. The generalization introduced here permits us to deal with problems under "nonstandard conditions."

For example, in system parameter estimation problems F consists of a class of admissible dynamic systems with some unknown parameters and G is the set of system parameters to be estimated. For many reasons (e.g., computational complexity, numerical efficiency, complexity of the eventual control of the system) only a subset F_0 of simpler systems is often considered for estimation purposes. Systems belonging to F_0, called approximating models, have in general a smaller number of unknown parameters whose possible values are a subset G_0 of the overall parameter set G. Then any estimation technique using approximating models can be considered as an algorithm mapping Y into G_0.

In this paper some results are presented which generalize results on optimality of central and least-squares algorithms (LSA) previously derived for the case $G_0 = G$ (Kacewicz *et al.*, 1986). In particular, it is shown that the most natural and widely used extension of least-squares algorithms may not preserve any of the interesting optimality properties derived for least-squares algorithms under standard conditions; a different extension is then proposed which preserves some of those properties. Although these preliminary results hold under quite restrictive conditions, they are shown to apply to meaningful examples of parameter estimation with approximating models.

2. Basic Definitions and Notations

Let F be a finite dimensional linear space over the real field. Let S be a fixed operator, called the *solution operator*, from F into G

$$S : F \to G, \tag{1}$$

where G is a linear finite dimensional normed space over the real field. We are interested in approximating an element $S(f) \in G$ knowing only some perturbed information on $f \in F$. We consider an *information operator* N mapping F into a linear finite dimensional normed space Y over the real field

$$N : F \to Y, \tag{2}$$

where $\dim y \geq \dim F$. It is assumed that information $N(f)$ about f is not known exactly, but only perturbed information y is available such that

$$y = N(f) + \eta, \tag{3}$$

where η is an unknown but norm bounded vector

$$\|y - N(f)\| = \|\eta\| \leq \rho \tag{4}$$

for a given nonnegative ρ. We will also assume that information is complete, i.e., N is a one-to-one mapping. These assumptions, already introduced in (Milanese and Tempo, 1985; Milanese *et al.*, 1986) mean that problem uncertainty is due only to the information error term η; in fact, they hold true in most problems normally met in the system identification research area in which we are interested (see Milanese and Tempo, 1985; Kacewicz *et al.*, 1986; Milanese *et al.*, 1986; Vicino *et al.*, 1987).

Let G_0 be subset of G. We consider the class Φ^r of (*restricted*) *algorithms* consisting of all operators φ mapping Y into G_0:

$$\varphi : Y \to G_0. \tag{5}$$

The operator φ, applied to an information y provides an approximation $\varphi(y) \in G_0$ of the true solution element $S(f)$ belonging to G. An algorithm will be also referred to as an *estimator* of $S(f)$ (Kacewicz *et al.*, 1986). In this paper we generalize some of the optimality results reported in (Kacewicz *et al.*, 1986), where it is assumed that $G_0 = G$.

We now recall the concept of approximation errors. We introduce the three sets

$$E_Y(f) = \{y \in Y : \|y - N(f)\| \leq \rho\} \tag{6}$$

$$E_F(y) = \{f \in F : \|y - N(f)\| \leq \rho\} \tag{7}$$

$$E_G(y) = S(E_F(y)). \tag{8}$$

We shall assume that the set $E_F(y)$ including all problem elements compatible with a given information is nonempty, i.e., y belongs to a subset Y_0 of Y given by

$$Y_0 = \{y \in Y : E_F(y) \neq \emptyset\}. \tag{9}$$

Again, this hypothesis is normally satisfied in well-posed estimation problems.

We now associate the definitions of approximation errors and optimality to sets (6) and (7). We define an *F-local error* $e_F(\varphi, f)$ of an algorithm φ as

$$e_F(\varphi, f) = \sup_{y \in E_Y(f)} \|S(f) - \varphi(y)\|, \qquad f \in F. \tag{10}$$

An algorithm φ^* is called *F-strongly optimal* in a given class Φ if $\varphi^* \in \Phi$ and

$$e_F(\varphi^*, f) \leq e_F(\varphi, f) \qquad \forall\, f \in F, \forall\, \varphi \in \Phi. \tag{11}$$

In a similar way, for each $y \in Y_0$ we define a *Y- local error* $e_Y(\varphi, y)$ of an algorithm φ as

$$e_Y(\varphi, y) = \sup_{f \in E_F(y)} \|S(f) - \varphi(y)\|. \tag{12}$$

An algorithm φ^* is called Y-strongly optimal in a class Φ if $\varphi^* \in \Phi$ and

$$e_Y(\varphi^*, y) \leq e_Y(\varphi, y) \qquad \forall\, y \in Y_0, \forall\, \varphi \in \Phi. \tag{13}$$

The *(global) error* of an algorithm φ is defined as

$$e(\varphi) = \sup_{f \in F} e_F(\varphi, f) (= \sup_{y \in Y_0} e_Y(\varphi, y)). \tag{14}$$

An algorithm $\varphi_0 \in \Phi$ is called *(globally) optimal* in a class Φ if it minimizes the (global) error $e(\varphi)$.

3. Conditional Central and Conditional Least-Squares Algorithms

A particularly interesting class of algorithms which has been extensively studied both from theoretical and constructive points of view

(Traub and Woźniakowski, 1980; Milanese and Tempo, 1985; Milanese *et al.* 1986) is that of central algorithms. More precisely, in the above-mentioned references the case of $G_0 = G$ has always been considered. We now generalize the concept of a central algorithm by defining a conditional central algorithm. Let us define a *conditional center* of the set $E_G(y)$ with respect to the set G_0 as the point $c(y) \in G_0$ such that

$$\sup_{\hat{g} \in E_G(y)} \|c(y) - \hat{g}\| = \inf_{g \in G_0} \sup_{\hat{g} \in E_G(y)} \|g - \hat{g}\|. \tag{15}$$

We now assume that $E_G(y)$ has a conditional center $c(y)$ for any $y \in Y_0$.

DEFINITION 1. A *conditional central algorithm* φ_c is defined by

$$\varphi_c(y) = c(y), \qquad \forall\, y \in Y_0. \tag{16}$$

As it can be easily observed, when $G_0 = G$ (15) defines the classical Chebychev center of $E_G(y)$ and (16) the classical central algorithm. In that case it has been shown that a central algorithm is Y-strongly optimal (and globally optimal) in the class of all algorithms. It can be readily proven that the same result still holds when $G_0 \subset G$.

THEOREM 1. *A conditional central algorithm φ_c is Y-strongly and globally optimal in the class Φ^r.*

Proof. From the definition of conditional central algorithms (16) and (15) it easily follows that

$$\begin{aligned} e_Y(\varphi, y) &= \sup_{f \in E_F(y)} \|S(f) - \varphi(y)\| \geq \inf_{g \in G_0} \sup_{f \in E_F(y)} \|S(f) - g\| \\ &= e_Y(\varphi_c, y), \qquad \forall\, y \in Y_0 \end{aligned} \tag{17}$$

which proves the theorem. ∎

In general the computation of conditional central algorithms is not an easy task. To the authors knowledge, the most extensive results on characterization of conditional centers can be found in the works of Laurent and Pham-Dinh-Tuan (1970), where necessary and sufficient conditions are given for a point to be a conditional center in the case where G_0 has linear variety. Unfortunately, these conditions cannot easily be checked even for those particular assumptions which allow for a simple computation of central algorithms in the usual case $G_0 = G$ (Milanese *et al.*, 1986; Kacewicz *et al.*, 1986).

Another appealing class of algorithms which already has been investigated in (Kacewicz *et al.*, 1986) in the case $G_0 = G$ is that of projection algorithms. These algorithms are very frequently used in the field of system identification especially when a Hilbert norm is used in the information space Y, in which case the class of least-squares algorithms is obtained. For this reason we now extend the concept of least-squares algorithms to the case $G_0 \subset G$.

First, we recall the definition of least-squares algorithms. Let Y be equipped with a Hilbert norm $\|\cdot\|_H$.

DEFINITION 2. Let $y \in Y$, and f_y be such that

$$\|N(f_y) - y\|_H = \inf_{f \in F} \|N(f) - y\|_H. \tag{18}$$

A *least-squares algorithm* φ_{LS} is defined as

$$\varphi_{LS}(y) = S(f_y). \tag{19}$$

The concept of LSA may be extended in several ways. We introduce an extension which leads to what we call a *conditional least-squares algorithm* (CLSA) and which takes into account the way of restricting from G to G_0. We suppose that an operator R mapping G into G_0 is given. We call it a *restriction operator*.

DEFINITION 3. For a given restriction operator $R : G \rightarrow G_0$, a conditional least-squares algorithm φ_{CLS} is defined as

$$\varphi_{CLS}(y) = R(S(f_y)), \qquad y \in Y, \tag{20}$$

where f_y is defined by (18).

It must be observed that what is usually done in the context of estimation theory when dealing with reduced order systems corresponds to a different type of extension of LSA which we call the *reduced least-squares algorithm* (RLSA).

DEFINITION 3'. Let F_0 be a subset of F such that

$$F_0 = \{f \in F : S(f) \in G_0). \tag{21}$$

Let $f_y^0 \in F_0$ be such that

$$\|N(f_y^0) - y\|_H = \inf_{f \in F_0} \|N(f) - y\|_H. \tag{22}$$

A reduced least-squares algorithm is defined as

$$\varphi_{\mathrm{RLS}}(y) = S(f_y^0). \tag{23}$$

Though widely used, RLSA does not preserve any of the interesting optimality properties of LSA given by Kacewicz *et al.* (1986). We recall that LSA's have been shown to be central, Y-strongly optimal in the class of all algorithms and F-strongly optimal in a class of algorithms of practical interest called "correct" algorithms. In the following we show that a CLSA partially preserves these properties.

First, we prove that a CLSA is F-strongly optimal in a class of algorithms called "conditionally correct" algorithms.

DEFINITION 4. For a given restriction operator $R : G \to G_0$, an algorithm φ is said to be *conditionally correct* if

$$\varphi(N(f)) = R(S(f)) \qquad \forall\, f \in F. \tag{24}$$

Whenever a conditionally correct algorithm is applied to exact information, it gives the exact solution restricted to G_0 according to R. The class of conditionally correct algorithms will be denoted by Φ^{cc}.

Remark 1. If $G_0 = G$ and $R = I_d$ (identity operator), Φ^{cc} coincides with the class of correct algorithms considered by Kacewicz *et al.* (1986).

Remark 2. Φ^{cc} includes algorithms obtained by first using classical estimation algorithms such as least-squares, least absolute values, min–max, and then applying the restriction operator to the obtained estimate; in particular $\varphi_{\mathrm{CLS}} \in \Phi^{cc}$.

Remark 3. One of the most interesting choices for R is the projection operator on G_0 according to the norm defined in G. In this case, the class Φ^{cc} becomes particularly meaningful, because a conditionally correct algorithm provides a minimum estimation error when applied to exact information ($\rho = 0$).

THEOREM 2. *If N is a linear operator and Y is a Hilbert space, then φ_{CLS} is F-strongly optimal in the class Φ^{cc}.*

Proof. For each problem element $f \in F$ we have that

$$\begin{aligned} e_F(\varphi_{\mathrm{CLS}}, f) &= \sup_{y\,:\,\|y-Nf\|_{\mathrm{H}}\le\rho} \|S(f) - \varphi_{\mathrm{CLS}}(y)\| \\ &= \sup_{y\,:\,\|y-Nf\|_{\mathrm{H}}\le\rho} \|S(f) - R(S(f_y))\| \\ &= \sup_{f_y\,:\,\|Nf_y-Nf\|_{\mathrm{H}}\le\rho} \|S(f) - R(S(f_y))\|. \end{aligned} \tag{25}$$

On the other hand, for any conditionally correct algorithm φ

$$e_F(\varphi, f) = \sup_{y : \|y - Nf\|_H \le \rho} \|S(f) - \varphi(y)\|$$

$$\ge \sup_{\hat{f} : \|N\hat{f} - Nf\|_H \le \rho} \|S(f) - \varphi(N\hat{f})\|$$

$$= \sup_{\hat{f} : \|N\hat{f} - Nf\|_H \le \rho} \|S(f) - R(S(\hat{f}))\|. \tag{26}$$

From (25) and (26) it follows that

$$e_F(\varphi_{\text{CLS}}, f) \le e_F(\varphi, f), \qquad \forall\, f \in F, \forall\, \varphi \in \Phi^{\text{cc}} \tag{27}$$

which proves the theorem. ■

We now look for conditions ensuring Y-strong optimality of a conditional least-squares algorithm. The problem is equivalent to studying the conditional centrality of φ_{CLS} which, as previously noted, may not be an easy task. In the following we prove that φ_{CLS} is Y-strongly optimal under conditions that are rather restrictive from a purely mathematical point of view but which nevertheless are certainly of interest in problems of simplified model parameter estimation (see the example in the next section).

Without loss of generality, let $F = R^n$, $Y = R^m$ $(m \ge n)$ and $G = R^p$. Consider the following hypotheses.

H1. *Assume that Y is equipped with a Hilbert norm, G is equipped with an 1_∞ norm and S and N are linear operators.*

H2. *Assume that $G_0 \subset G$ is a linear variety defined by k real numbers* $a_1, a_2, \ldots, a_k$, $0 \le k \le p$

$G_0 = \{(g_1, g_2, \ldots, g_p)\colon g_{i_1} = a_1, g_{i_2} = a_2,$
$\ldots, g_{i_k} = a_k$, $i_1, i_2, \ldots, i_k$ fixed and such that $1 \le i_j \le p$ and $i_j \ne i_1$ for $j \ne 1\}$. (28)

H3. *R is the projection operator on G_0 in the Euclidean norm, i.e.,*

$$R(g) = g_0, \qquad g \in G, \tag{29a}$$

where g_0 is such that

$$\|g - g_0\|_2 = \min_{\hat{g} \in G_0} \|g - \hat{g}\|_2. \tag{29b}$$

Under this last condition the conditional least-squares estimator becomes as follows. Let f_y be defined as in (18) and g_y be such that

$$\|g_y - S(f_y)\|_2 = \min_{g \in G_0} \|g - S(f_y)\|_2, \qquad g_y \in G_0 \tag{30a}$$

then

$$\varphi_{\text{CLS}}(y) = g_y. \tag{30b}$$

Now, we will show that the algorithm (30) is Y-strongly optimal in the class of all algorithms. In order to do that we have to state a preliminary lemma. We observe that under the hypotheses on S and N, the following extrema are finite and they are achieved at some points belonging to $E_F(y)$ ($y \in Y_0$):

$$\begin{aligned} \alpha_i &= \inf_{f \in E_F(y)} (Sf)_i \\ \beta_i &= \sup_{f \in E_F(y)} (Sf)_i \end{aligned} \qquad i = 1, \ldots, p. \tag{31}$$

In (31) the subscript i denotes the ith component of a vector.

LEMMA 1. *Under the hypotheses $H1$, $H2$, and $H3$ we have that*

$$\text{(i)}\quad (\varphi_{\text{CLS}}(y))_i = \begin{cases} a_j, & \text{if } i = i_j, \text{ for some } j = 1, 2, \ldots, k \\ (Sf_y)_i, & \text{if } i \neq i_j, \text{ for } j = 1, 2, \ldots, k \end{cases} \tag{32}$$

$$\text{(ii)}\qquad (Sf_y)_i = \frac{\alpha_i + \beta_i}{2}, \qquad \text{for } i \neq i_j,\ j = 1, 2, \ldots, k. \tag{33}$$

The proof of (i) follows easily from the definition of φ_{CLS} (30) and the hypothesis H3. The proof of (ii) follows from the fact that, due to the Hilbert norm in the space Y and the linearity of N, $E_G(y)$ is a centrally symmetric set with Sf_y as the symmetry center. Thus, Sf_y is also a Chebychev center of $E_G(y)$ (see, e.g., Kacewicz *et al.* (1986)) and hence its coordinates are given by (33).

We can now state Theorem 3.

THEOREM 3. *If hypotheses $H1$, $H2$, and $H3$ hold, then φ_{CLS} is a conditionally central algorithm and therefore it is Y-strongly optimal in the class Φ^{r}.*

Proof. Let r be the Y-local error of φ_{CLS}

$$r = e_Y(\varphi_{CLS}, y) = \sup_{f \in E_F(y)} \|Sf - g_y\|_\infty. \tag{34}$$

We know that for some $f^* \in E_F(y)$ and some $1 \le i \le p$

$$r = |(Sf^*)_i - (g_y)_i|. \tag{35}$$

Let $g \in G_0$; with reference to (35) there are two possible cases. In the first one, if $i = i_j$ for some $j = 1, 2, \ldots, k$, then

$$g_i = (g_y)_i = a_j$$

and we have that

$$r = |(Sf^*)_i - g_i| \le \|Sf^* - g\|_\infty. \tag{36}$$

In the second case, if $i \ne i_j$ for $j = 1, 2, \ldots, k$, then from Lemma 1 it follows that

$$(g_y)_i = \frac{\alpha_i + \beta_i}{2} \tag{37}$$

and either $(Sf^*)_i = \alpha_i$ or $(Sf^*)_i = \beta_i$ with $r = (\beta_i - \alpha_i)/2$. Thus, if we denote by $\bar{f}$ the point of $E_F(y)$ symmetric to f^*, i.e., $\bar{f} = 2f_y - f^*$, we have

$$r \le \max(|(Sf^*)_i - g_i|, |(S\bar{f})_i - g_i|), \qquad \forall\, g_i. \tag{38}$$

It follows from (36) and (38) that in both cases we have

$$r \le \sup_{f \in E_F(y)} \|Sf - g\|_\infty, \qquad \forall\, g \in G_0 \tag{39}$$

which proves the theorem. ■

The following result easily follows from Theorems 1, 2, and 3.

COROLLARY. *If $H1$, $H2$, and $H3$ hold, then φ_{CLS} is Y-strongly optimal in the class Φ^r and F-strongly optimal in the class Φ^{cc}.*

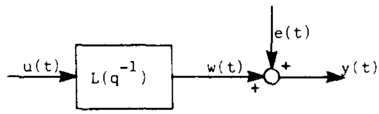

FIGURE 1

4. A PARAMETER ESTIMATION PROBLEM

Now we report an example from the system identification area to clarify the meaning of conditional estimators and of the optimality properties of conditional least-squares estimators.

Consider the linear discrete time dynamic system shown in Fig. 1. In Fig. 1 the signals $u(t)$, $w(t)$, $e(t)$, and $y(t)$ are assumed to be discrete time functions; $u(t)$ is the system input, $w(t)$ is the system output, $e(t)$ is measurement noise, and $y(t)$ is the corrupted system output; $L(q^{-1})$ is a linear operator which for a large class of systems (called autoregressive systems) takes the form

$$L(q^{-1}) = \frac{1}{A(q^{-1})}, \tag{40}$$

where $A(q^{-1})$ is a polynomial of order n in the unit delay shift operator q^{-1}[1]

$$A(q^{-1}) = 1 - \sum_{i=1}^{n} a_i q^{-i}. \tag{41}$$

The parameter estimation problem consists of evaluating the unknown system parameters a_i, assuming that the quantities $u(t)$ and $y(t)$ are known for a given set of time points $[0, 1, \ldots, N]$. The difference equation describing the system of Fig. 1 is

$$y(t) = w(t) + e(t) = \sum_{i=1}^{n} a_i y(t-i) + u(t) + e(t) + \sum_{i=1}^{n} a_i e(t-i), \qquad t = n+1, \ldots, N. \tag{42}$$

[1] The operator q^{-i} is such that $q^{-i}y(t) = y(t-i)$.

In (42) the values $y(0), \ldots, y(n)$ are assumed to be the system initial conditions. Now, setting

$$\tilde{y}(t) = y(t) - u(t) \tag{43}$$

$$\varepsilon(t) = e(t) - \sum_{i=1}^{n} a_i e(t-i) \tag{44}$$

we obtain the equation

$$\tilde{y}(t) = \sum_{i=1}^{n} a_i y(t-i) + \varepsilon(t), \qquad t = n+1, \ldots, N, \tag{45}$$

where $\tilde{y}(t)$ and $y(t-i)$ are known data and $\varepsilon(t)$ represents the measurement uncertainty called equation error. Assuming that $\varepsilon(t)$ is unknown but bounded in some norm, the problem of estimating the parameters a_i can be formulated in the context of information-based complexity. More precisely, the space F is identified as the n-dimensional linear space of parameters, the operator S is the identity, and the information space Y is an $(N-n)$ dimensional space containing the corrupted information $\tilde{y}(t)$. If a Hilbert norm is used in Y (as is frequently done in system identification contexts), then we can apply the theory developed in the previous sections. The need for reduced order model estimation stems from the fact that the value of n (order of the autoregressive system) may be very high and only simplified models are looked for, containing fewer parameters. Thus, our problem consists of evaluating the parameters of reduced order models corresponding to subspaces $G_0 \subset G$ of dimension less than n and restriction matrix R described as

$$r_{ij} = \begin{cases} 1, & \text{if } i = j,\ i \leq n-k \\ 0, & \text{if } i > n-k \text{ or } i \neq j \text{ and } i \leq n-k. \end{cases} \tag{47}$$

Now, if $g = [a_1, a_2, \ldots, a_n]^T \in G$, then $g_0 = Rg = [a_1, a_2, \ldots, a_{n-k}, 0, \ldots, 0]^T \in G_0$. In these conditions, the conditional least-squares estimator φ_{CLS} defined in (30) is by Theorem 3 a Y-strongly optimal algorithm in the class of all algorithms Φ^r, provided that Y-local errors are defined in a 1_∞ norm. This result is particularly significant, because it means that the use of an ordinary least-squares estimator (Kacewicz *et al.*, 1986) for the full order system permits in one shot an optimal solution to the problem of estimating all the reduced order models for any value of k $(1 \leq k \leq n)$. It must be observed that this conclusion means that no further improvement of the Y-local estimation error can be achieved by reformulating, as is usually done, the reduced order model estimation problem as a new $(n - k)$ dimensional problem and applying an ordinary least-squares algorithm; indeed, this would mean using an RLSA. Finally, it must be noted that this fact does not contradict the optimality results on least-squares estimators given in (Kacewicz *et al.*, 1986) that refer to the case in which the space G_0 contains the true system ($\dim G_0 = n$); in our case the assumption that the space G_0 is of dimension $(n - k)$ means that G_0 does not necessarily contain the true solution element.

References

GENESIO, R., AND MILANESE, M. (1979), Methods for the selection of approximating classes of models, *in* "Proceedings, 5th IFAC Symp. Identification and System Parameter Estimation," Pergamon, Elmsford, NY.

KACEWICZ, B. Z., MILANESE, M., TEMPO, R., AND VICINO A. (1986), Optimality of central and projection algorithms for bounded uncertainty, *Syst. Control Latters* **8**, 161–171.

LAURENT, P. J., AND PHAM-DINH-TUAN (1970), Global approximation of a compact set by elements of a convex set in a normed space, *Numer. Math.* **15**, 137–150.

MILANESE, M., AND TEMPO, R. (1985), Optimal algorithms theory for robust estimation and prediction, *IEEE Trans. Automat. Control* **AC-30**, 730–738.

MILANESE, M., TEMPO, R., AND VICINO, A. (1986), Strongly optimal algorithms and optimal information in estimation problems, *J. Complexity* **2**, 78–94.

SCHWEPPE, F. C. (1973), "Uncertain Dynamic Systems," Prentice–Hall, Englewood Cliffs, NJ.

TRAUB, J. F., WASILKOWSKI, G., AND WÓZNIAKOWSKI, H. (1983), "Information, Uncertainty, Complexity," Addison–Wesley, Reading, MA.

TRAUB, J. F., AND WÓZNIAKOWSKI, H. (1980), "A General Theory of Optimal Algorithms," Academic Press, New York.

VICINO, A., TEMPO, R., GENESIO, R., AND MILANESE, M. (1987), Optimal error and GMDH predictors: A comparison with some statistical techniques, *Int. J. Forecasting* **3**, 313–328.