Robust Estimation and Filtering in the Presence of Unknown but Bounded Noise

ROBERTO TEMPO

CENS
Consiglio Nazionale delle Ricerche (Italy)
and
Department of Computer Science, Columbia University

October 1986

CUCS-223-86

This work was supported in part by a NATO-CNR advanced fellowship and by funds of CENS-CNR

Abstract. In this paper optimal algorithms for robust estimation and filtering are constructed. No statistical assumption is supposed available or used and the noise is considered a deterministic variable unknown but bounded belonging to a set descibed by a norm. Previous results obtained for complete (one-to-one) and approximate information [1] are now extended to partial and approximate information. This information seems useful in dealing with dynamic systems not completely identifiable and/or with two different sources of noise, for example process and measurement noise. For different norms characterizing the noise, optimal algorithms (in a min-max sense) are shown. In particular for Hilbert norms a linear optimal algorithm is the well-known minimum variance estimator. For l_∞ and l_1 norms optimal algorithms, computable by linear programming, are presented. Applications to time series prediction and parameter estimation of nonidentifiable dynamic systems are shown.

State estimation is formalized in the context of the general theory. Assuming an exponential smoothing of the bounds of the noise it is proved that, for stable systems, the uncertainty of the state is aymptotically bounded. Then the results of the previous sections provide computable algorithms for this problem. Two application examples are shown: Leontief models and Markov chains.

I. Introduction.

In recent years an alternative approach to the classical statistical one has been developed for system identification. Generally, in estimation theory the uncertainty is considered as a random variable described by a certain density function. In the worst case approach no statistical assumption is available or used and the noise is supposed a deterministic variable belonging to a certain set [1]-[5]. A relationship between optimal estimators for statistical and worst case can be found in [6].

In this paper we address the problem of estimation and filtering when the uncertainty is assumed unknown but bounded belonging to a set described by a norm. We extend previous results [1] obtained for complete (one-to-one) and approximate information to the case of partial (not necessarily one-to-one) and approximate information . The use of partial and contaminated information seems useful for two kinds of problems in system identification. First, for problems in which two different sources of noise should be considered, for example process and observation noise of a dynamic system. Second, for problems in which the parameters are not completely identifiable from the output measurements and some a priori information is given.

We construct optimal algorithms defined in the sense of information-based complexity (see [1]-[4], [7], [8]). Now we briefly discuss the subject of this theory. We are interested in approximating $Sf \in Z$ where S is a linear mapping $S : F \to Z$, and an element f belongs to a certain subset $K \subset F$ representing the a priori information available about it (f and S are called, respectively, problem element and solution operator). The element f is not exactly known but only approximate information $h = Nf + \rho, h \in H$ is given, where N is called information operator and the noise ρ is unknown but bounded by a given constant ϵ. An approximation to Sf is given by an operator φ (called algorithm or estimator) operating on the information h. Optimal algorithms minimize the maximal distance between the actual solution Sf and the computed solution $\varphi(h)$ for the worst problem element f and for the worst information h. The error of an optimal algorithm is called intrinsic error or radius of information. Strongly optimal algorithm are optimal algorithms for any fixed information h. Formal definitions and notations are presented in Section II.

In Section III we present optimal algorithms for some norms describing the uncertainty of the problem elements f and of the noise ρ. First we consider the uncertainty bounded by Hilbert norms. This problem is studied in [8]: the linear optimal algorithm presented has the same structure as the maximum a posteriori estimator for Gaussian distributions, but depends on a 'smoothing factor' whose practical computation is not an easy task. In this section we show that Gauss-Markov estimator is an optimal algorithm if the bound ϵ on the noise ρ is sufficiently small (corresponding to the statistical case when the knowledge on the covariance matrix of the noise is essential). If ϵ is large (equivalent to a poor knowledge of the covariance matrix) an optimal algorithm is 0. Then we turn our attention to other norms in the measurement space. For l_∞ norm a strongly optimal algorithm, computable by linear programming is derived in [1]. Here, with the same norm we show an optimal algorithm using the restrictive information given by the active constraints of the intrinsic error. Note that these constraints correspond to the optimal information minimizing the intrinsic error, in the sense defined in [7]. For l_1 norms we construct a strongly optimal algorithm computable by linear programming. Finally, we present two examples showing the applications to time series prediction and parameter estimation of nonidentifiable systems.

In Section IV we formalize the state estimation of a dynamic system in the general context of the theory. The problem of recusive state estimation in the presence of unknown but bonded noise is studied in [5], [9], [10]. Unfortunately when instantaneous constraints are considered, only approximate solutions can be easily computed. In fact, in general, the uncertainty set cannot be characterized by a finite set of numbers, even in the case when l_2 norms are considered. Furthermore, the uncertainty set may be too small asymptotically, with respect to the actual one, as in statistical case, especially in the presence of modeling errors. For these reasons we consider a different approach. We differently weight new and old data introducing an exponential smoothing of the bounds of the noise. We prove that, for stable systems, the intrinsic error representing the uncertainty of the solution, is asymptotically finite. By neglecting the higher order powers, a computable solution can be obtained by applying the results of Section III. Finally we show two application examples: dynamic Leontief models and finite Markov chains.

II. Definitions and Notations

This section provides formal definitions and notations used in the paper.

Let F be a linear normed n-dimensional space over the real field and let K be the unit ball in F defined by

$$K = \{f \in F : ||f|| \leq 1\} \ . \tag{1}$$

Consider a given linear operator S, called a *solution operator*, which maps F into Z

$$S : F \rightarrow Z$$

where Z is a linear normed r-dimensional space over the real field. Our aim is to estimate an element Sf of the Z space knowing approximate and partial information about the element $f \in K$.

Define a linear operator N, called *information operator* , which maps F into a linear normed m-dimensional space H

$$N : F \rightarrow H.$$

In this paper we do not assume that N is a complete information, i.e., N is not necessarily a one-to-one mapping. This means that the problem element f may be not identifiable from the knowledge of Nf. In general, in the presence of noise, exact information Nf about f is not available and only perturbed information h is given. In this context we assume that the uncertainty of information $\rho = h - Nf$ is unknown but bounded

$$||h - Nf|| \leq \epsilon \tag{2}$$

where ϵ is some fixed number.

An *algorithm* φ is a mapping (in general nonlinear) from H into Z

$$\varphi : H \to Z$$

i.e., it provides an approximation $\varphi(h)$ of Sf using perturbed and a priori information. Such an algorithm will be often called estimator.

Define the following sets

$$T(h) = \{f \in K : \|h - Nf\| \leq \epsilon\} \tag{3}$$

and

$$H_0 = \{h \in H : T(h) \neq \emptyset\} \tag{4}$$

where $\emptyset$ is the empty set. The set H_0 represents the set of all approximate information h compatible with the information Nf, the bound on the noise ϵ and the subset K. For each approximate information $h \in H_0$ we define a *local error $E(\varphi,h)$ of an algorithm* φ as

$$E(\varphi, h) = \sup_{f \in T(h)} \|Sf - \varphi(h)\| \quad \forall h \in H_0 \ . \tag{5}$$

An algorithm φ_s is called *strongly optimal* if

$$E(\varphi_s, h) = \inf_{\varphi} E(\varphi, h) \quad \forall h \in H_0 \ . \tag{6}$$

The strong optimality is a meaningful property in estimation problems as it ensures the minimum uncertainty between the actual solution Sf and the estimated solution $\varphi(h)$ for the worst element f belonging to the set $T(h)$ for any fixed approximate information h. Note that strongly optimal algorithms map h into the center of a minimal ball containing the set $S\{T(h)\} \, \forall h \in H_0$. For this reason they are often called *central algorithms* (see [1]-[4]).

The *global error $E(\varphi)$ of an algorithm* φ is defined as

$$E(\varphi) = \sup_{h \in H_0} E(\varphi, h) \ . \tag{7}$$

An algorithm φ_o is called *(globally) optimal* if

$$E(\varphi_o) = \inf_{\varphi} E(\varphi) \ . \tag{8}$$

The minimal global error $E(\varphi_o)$ is called *intrinsic error* or *radius* of information.

In the following, by the subscript $_i$ and the superscript ^{T} we will denote the i-th row and the transpose of a matrix or a vector; thus $(\varphi(h))_i$ denotes the i-th component of the vector $\varphi(h)$.

III. Optimal Estimators for Linear Problems

In this section optimal and strongly optimal algorithms for linear problems (N and S linear) with partial information for some norms characterizing input and output noise are presented.

1. Measurement space equipped with Hilbert norms.

This problem is studied in [8] when F, H and Z are Hilbert spaces, not necessarily of finite dimensions. Let F and H be equipped with the following Hilbert norms

$$||f||^2 = f^T V f \ , \quad V = V^T > 0 \tag{9}$$

and

$$||h||^2 = h^T R h \ , \quad R = R^T > 0. \tag{10}$$

Then, based on [8], one can derive the optimal algorithm of the form

$$\varphi_o(h) = S\big(N^T R N + \frac{\mu \epsilon^2}{1-\mu} V\big)^{-1} N^T R h \tag{11}$$

where μ is a parameter between zero and one, which is solution of the equation

$$\sup_{f \in T(0)} ||Sf|| = \sup_{\mu ||f||^2 + (1-\mu)\epsilon^{-2} ||Nf||^2 \leq 1} ||Sf|| \ . \tag{12}$$

Note that (11) is the maximum a posteriori estimator (minimum variance and minimum absolute error) of Sf when f and ρ are independent random variables normally distributed with mean value and covariance matrix given by

$$M_f = M_\rho = 0 \quad \Sigma_f = \Big(\frac{\mu\epsilon^2}{1-\mu}V\Big)^{-1} \quad \Sigma_\rho = R^{-1} \ .$$

Unfortunately, from a practical point of view, the computation of the smoothing coefficient μ, i.e., the solution of the equation (12) is not an easy task. However, for some specific cases the parameter μ can be easily found. We consider two such cases depending on the size of the bound ϵ of the noise.

Define the set $\Delta(\epsilon)$ as

$$\Delta(\epsilon) = \{f \in F : ||Nf|| \leq \epsilon\} \ . \tag{13}$$

THEOREM 1.

Let F, H and Z be Hilbert spaces. Let H be equipped with the norm (10). Then a linear optimal algorithm is given by

$$\varphi_o(h) = \begin{cases} S(N^TRN)^{-1}N^TRh & \text{if } \Delta(\epsilon) \subset K \\ 0 & \text{if } \Delta(\epsilon) \supseteq K \ . \end{cases} \tag{14}$$

PROOF:

It is easy to prove that $\mu = 0$ for the first case, $\Delta(\epsilon) \subset K$ and $\mu = 1$ for the second one, $\Delta(\epsilon) \supseteq K$.

Let $\Delta(\epsilon) \subset K$. Note that this condition implies that N is one-to-one. Under this assumption we obtain

$$\sup_{f \in T(0)} ||Sf|| = \sup_{||Nf|| \leq \epsilon} ||Sf|| \ . \tag{15}$$

From this we conclude that $\mu = 0$ is a solution of (12) and Gauss-Markov is an optimal algorithm.

If $\Delta(\epsilon) \supseteq K$ we get

$$\sup_{f \in T(0)} ||Sf|| = \sup_{||f|| \leq 1} ||Sf|| \ . \tag{16}$$

Thus, $\mu = 1$ is now a solution of (12). Then, for μ tending to one, the algorithm φ_o in (11) tends to zero. Hence, the zero algorithm is optimal as claimed.

∎

The interest of Theorem 1 lies in the fact that, if $\Delta(\epsilon) \subset K$, a linear optimal algorithm is the Gauss-Markov estimator. Note that if N is one-to-one and $K \equiv F$ Gauss-Markov is strongly optimal (see [11]). Note that $\Delta(\epsilon) \subset K$ is equivalent to $||N^{-1}|| > \epsilon^{-1}$.

If the knowledge on the noise is weak, i.e. $K \subseteq \Delta(\epsilon)$ (corresponding to the statistical case when a poor knowledge of the covariance matrix of the noise is available), the optimal algorithm is 0, as in the statistical case. Observe that $\Delta(\epsilon) \supseteq K$ is equivalent to $||N|| \leq \epsilon$.

Optimal algorithms are now presented under different assumptions on the norms.

2. Measurement space equipped with $l\infty$ norms.

In [1] a solution is given when H and Z are equipped with l^w_∞ weighted norms[1] and K is a piecewise linear, convex set. The nonlinear strongly optimal algorithm presented can be easily computed by means of linear programming. In [6] it is shown that the same algorithm is a maximum likelihood estimators for uniform distribution of the noise when the information is complete and $K \equiv F$. Note that the use of the l^w_∞ weighted norm in the measurement space allows one to handle situations in which the noise of every measurement may be differently bounded.

In this section we present a linear optimal algorithm: this algorithm is constructed using the restrictive information given by the *active constraints* of the intrinsic error. Consider the case when S is a linear functional and K is a convex balanced set. In this case the intrinsic error is given by (see [12])

$$E(\varphi_o) = \sup_{f \in T(0)} |Sf| \ . \tag{17}$$

[1]The norm $||h||^w_\infty$ is defined as

$$||h||^w_\infty = \max_i w_i |h_i| \quad w_i > 0 \ .$$

Define an $(n+m)$ by n matrix P and an $(n+m)$-dimensional vector d such that

$$P = [I \ , \ N]^T \tag{18}$$

$$d = [0 \ , \ h]^T \tag{19}$$

where 0 is the null vector. If K is the unit ball and w a suitable weight $T(h)$ can be written as

$$T(h) = \{f \in F : \|Pf - d\|_\infty^w \leq \epsilon\} \ . \tag{20}$$

If $w_1 = \ldots = w_n = \epsilon$, using the definition of l_∞^w norm we obtain

$$T(0) = \left\{f \in F : |P_i f| \leq \frac{\epsilon}{w_i} \ ; i = 1, \ldots, (m+n)\right\} \ . \tag{21}$$

Then the intrinsic error is

$$E(\varphi_o) = \sup_{\substack{f \in F : |P_i f| \leq \frac{\epsilon}{w_i} \\ i = 1, \ldots, (m+n)}} Sf = \sup_{\substack{f \in F : |P_i f| \leq \frac{\epsilon}{w_i} \\ i = t_1, \ldots, t_n}} Sf \ . \tag{22}$$

The constraints $t_1, \ldots, t_n$ are called the active constraints of the linear programming problem (22).

Define the n by n matrix $\widetilde{P}$ containing the rows $t_1, \ldots, t_n$ of the matrix P

$$\widetilde{P} = [P_{t_1}, \ldots, P_{t_n}]^T \tag{23}$$

and the n-dimensional vector $\widetilde{d}$ containing the active constraints $t_1, \ldots, t_n$ of the vector d

$$\widetilde{d} = [d_{t_1}, \ldots, d_{t_n}]^T \ . \tag{24}$$

The information $\widetilde{d}$ is the optimal information in the sense defined in [7]. Note that at most n measurements out of m should be available and are used by the following optimal algorithm.

THEOREM 2.

Let F and H be linear normed spaces equipped with l_∞^w norms and S a linear functional. If rank $\widetilde{P} = n$ then the algorithm

$$\varphi_o(h) = S\widetilde{P}^{-1}\widetilde{d} \tag{25}$$

is linear and optimal.

PROOF:

Using (20) the local error of the algorithm (25) is

$$E(S\widetilde{P}^{-1}\widetilde{d}, h) = \sup_{f:\|Pf-d\|_\infty^w \leq \epsilon} |Sf - S\widetilde{P}^{-1}\widetilde{d}| \ . \tag{26}$$

Since

$$\{f \in \|Pf - d\|_\infty^w \leq \epsilon\} \subseteq \left\{f \in \|\widetilde{P}f - \widetilde{d}\|_\infty^w \leq \epsilon\right\} \tag{27}$$

it results that

$$E(S\widetilde{P}^{-1}\widetilde{d}, h) \leq \sup_{f:\|\widetilde{P}f-\widetilde{d}\|_\infty^w \leq \epsilon} |Sf - S\widetilde{P}^{-1}\widetilde{d}| =$$

$$= \sup_{f:\|\widetilde{P}(f-\widetilde{P}^{-1}\widetilde{d})\|_\infty^w \leq \epsilon} |S(f - \widetilde{P}^{-1}\widetilde{d})| = \sup_{f:\|\widetilde{P}f\|_\infty^w \leq \epsilon} |Sf| \ . \tag{28}$$

Using the definitions of active constraints (22) and of matrix $\widetilde{P}$ (23) we get

$$E(S\widetilde{P}^{-1}\widetilde{d}, h) \leq E(\varphi_o) \ . \tag{29}$$

Since the opposite inequality is obvious the proof is complete.

∎

In [1] the properties of a similar algorithm are studied for complete and approximate information. Remark that the condition about the invertibility of the matrix $\widetilde{P}$ is not restrictive in practice; in fact if rank $\widetilde{P} < n$ the intrinsic error is not finite.

REMARK.

The previous theorem can be easily extended when S is a linear operator and Z is equipped with l_∞ norm. In this case the active constraints of the problem can be computed componentwise and a linear optimal algorithm is the rectangular matrix r by m having row by row the optimal solution of each component of Sf obtained by Theorem 2.

∎

3. MEASUREMENT SPACE EQUIPPED WITH l_1 NORMS.

In this section we present a strongly optimal algorithm when the measurement noise is described by l_1^w norms[1]

Define the following matrix W of size 2^{m-1} by m containing the weights w_i

$$W = \begin{pmatrix} w_1 & w_2 & \dots & w_{m-1} & w_m \\ w_1 & w_2 & \dots & w_{m-1} & -w_m \\ w_1 & w_2 & \dots & -w_{m-1} & w_m \\ w_1 & w_2 & \dots & -w_{m-1} & -w_m \\ \dots & \dots & \dots & \dots & \\ \dots & \dots & \dots & \dots & \\ w_1 & -w_2 & \dots & -w_{m-1} & -w_m \end{pmatrix} . \tag{30}$$

THEOREM 3.

Let H and Z be linear normed spaces equipped with l_1^w and l_∞ norms and K be the unit ball. Then the algorithm given by

$$\varphi_s(h)_i = \frac{1}{2} \sup_{\substack{f \in K : |W_i(Nf-h)| \le \epsilon \\ i=1,\dots,2^{m-1}}} S_i f + \frac{1}{2} \inf_{\substack{f \in K : |W_i(Nf-h)| \le \epsilon \\ i=1,\dots,2^{m-1}}} S_i f \quad i = 1, \dots, r \tag{31}$$

[1]The norm l_1^w is defined as

$$||h||_1^w = \sum_{i=1}^m w_i |h_i| \quad w_i > 0 .$$

is strongly optimal.

PROOF:

Using the definition of the l_1^w norm in the space H we get

$$||Nf - h||_1^w = \sum_{i=1}^{m} w_i |N_i f - h_i| = \sum_{i=1}^{m} |w_i(N_i f - h_i)| \leq \epsilon \tag{32}$$

which is equivalent to

$$\begin{aligned} W_i(Nf - h) \leq \epsilon \quad i = 1, \ldots, 2^{m-1} \\ -W_i(Nf - h) \leq \epsilon \quad i = 1, \ldots, 2^{m-1} \end{aligned} \tag{33}$$

and

$$|W_i(Nf - h)| \leq \epsilon \quad i = 1, \ldots, 2^{m-1} \ . \tag{34}$$

A strongly optimal algorithm is given by the center of the set $S\{T(h)\}$ (see [1]-[4]). If the space Z is equipped with l_∞ norm the center can be computed componentwise. Then relation (31) follows.

∎

If K is defined by linear inequalities (i.e., the space F is equipped with l_∞ or l_1 norms) the computation of the previous algorithm is a linear programming problem, although the number of constraints is at least $(2^m + 2n)$. The previous algorithm can be efficiently computed in many practical problems when the number of measurements is limited, for example, in nonidentifiable dynamic systems. Nevertheless, in many other cases, since the structure of the matrix (30) is very particular, the number of constraints may be greatly reduced.

Now we present two examples showing the practical application of the theory previously developed to problems of time series prediction and parameter estimation of nonidentifiable dynamic systems.

EXAMPLE 1. *(Time Series Prediction)*

Time series prediction is the estimation of a future value of a time function knowing the observed values during a previous time interval. The problem element f is a discrete time function of the form

$$f(k) = \sum_{i=1}^{l} a_i g_i(k) + u(k) \quad k = 1, \ldots, (m+1)$$

where l is a fixed integer, $g_i(k)$ are given discrete functions, a_i arbitrary real unknown coefficients and $u(k)$ are samples of the process noise. In this case, since the exact knowledge of $Nf = \{f(1), \ldots, f(m)\}$ is available, the information is partial and exact. Our aim is to approximate the problem element $Sf = f(m+1)$ by an algorithm φ, called in this case optimal error predictor, operating on the information Nf. A complete solution of this problem is provided by the results of Section III, according to the norm used to describe the process noise.

This problem has been studied in detail in [13] and [14]. Furthermore, in [14] a wide numerical comparison between autoregressive, mixed harmonic, ARMA, ARIMA, GMDH, bilinear, subset bilinear, subset autoregressive and optimal error predictors is presented using three well-known time series, namely Wölf Sunspot Numbers, Annual Canadian Lynx Data and Australian Births.

∎

EXAMPLE 2. *(Parameter Estimation of Nonidentifiable Dynamic Systems)*

In this case the problem element $f = f(\lambda, k)$ is the input-output pair of a linear discrete dynamic system with unknown parameter vector λ. The information is the knowledge of $f(\lambda, k)$ at different times k possibly corrupted by additive measurement noise and the solution is the estimation of λ. Since $f(\lambda, k)$ may be nonidentifiable from the knowledge of the output measurements the information is partial and possibly contaminated. The a priori information acquired about the parameters λ is taken into account in the set K. In this case the results of Section III can be used. The problem of parameter estimation of dynamic identifiable systems in the presence of measurement noise unknown but bounded and described by l_∞ norm is studied in [1] and [15].

IV. STATE ESTIMATION OF LINEAR SYSTEMS IN THE PRESENCE OF UNKNOWN BUT BOUNDED NOISE

In this section we apply the general theory previously developed to state estimation of linear systems when process and observation noise are unknown but bounded.

Consider the following discrete, linear, time invariant dynamic system represented in state variable form

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) + \eta(k) \end{aligned} \tag{35}$$

where $x(k)$ is the r-dimensional state vector; $y(k)$ is the q-dimensional observation vector; $u(k)$ and $\eta(k)$ are the p-dimensional process noise and the q-dimensional observation noise vector; A, B and C are given matrices with compatible dimensions.

At time k the problem element f is defined as

$$f = [x(0), u(0), \ldots, u(k)]^T \ . \tag{36}$$

It results that the space F of the elements f is a linear space of dimensions $n = r + p(k+1)$. The subset K contains the information about the initial conditions (if available) and the process noise and is defined by

$$K = \{f \in F : \|x(0)\| \leq \delta; \|u(i)\| \leq \tau, i = 1, \ldots, k\} \ . \tag{37}$$

Since our goal is the estimation of the state $x(k+1)$ the solution operator is given by

$$Sf = x(k+1) \tag{38}$$

where $x(k+1)$ can be written in terms of the variables $x(0)$, $u(0)$, ..., $u(k)$ as

$$x(k+1) = A^{k+1}x(0) + \sum_{i=0}^{k} A^i B u(k-i) \; . \tag{39}$$

The space Z of the solution Sf is linear and r-dimensional. At time k the information operator N is given by

$$Nf = [Cx(0), Cx(1), \ldots, Cx(k)]^T =$$

$$= \left[Cx(0), CAx(0) + CBu(0), \ldots, CA^k x(0) + C\sum_{i=0}^{k-1} A^i B u(k-i-1) \right]^T \; . \tag{40}$$

In this case the exact information (40) is corrupted by observation noise and only the approximate information h is available

$$h = [y(0), \ldots, y(k)]^T = Nf + [\eta(0), \ldots, \eta(k)]^T \; . \tag{41}$$

The space H is linear and of dimensions $m = q(k+1)$. According to (2) we assume that the uncertainty of information

$$\rho = [\eta(0), \ldots, \eta(k)]^T \tag{42}$$

is unknown but bounded by a constant ϵ. An algorithm φ operating on the information (41), available at step k, provides an estimate $\hat{x}(k+1)$ of the state $x(k+1)$

$$\varphi(h) = \hat{x}(k+1) \; . \tag{43}$$

In particular we are looking for locally and globally optimal algorithms minimizing the uncertainty of the estimate of the state measured with a suitable norm

$$\|Sf - \varphi(h)\| = \|x(k+1) - \hat{x}(k+1)\| \tag{44}$$

according to the min-max criteria defined in (6) and (8).

It is to remark that the strong optimality property may be very useful for this problem. In fact it guarantees the minimization of the uncertainty (44) for the worst problem element f when the information h is fixed. The intrinsic error represents the minimum uncertainty (44) for the worst h and f, i.e., the minimum global error.

REMARK.

The previous formulation can be generalized to include a measurable input vector $v(k)$, i.e.,

$$x(k+1) = Ax(k) + Bu(k) + v(k)$$

and a dynamic time-variant system with matrices $A(k)$, $B(k)$ and $C(k)$.

∎

REMARK.

The condition of observability of the dynamic system is necessary such that the set $T(h)$ is bounded and the radius is finite, at least when no reliable information about the initial condition is available, as frequently happens. In this case $\delta = \infty$ in the set (37) and we must assume that

$$rank\left[C, CA, \ldots, CA^{n-1}\right]^T = n$$

in order to ensure that the term $A^{k+1}x(0)$ in formula (39) could be recovered.

∎

The problem of recursive parameter and state estimation of a dynamic sytem in the presence of unknown but bounded noise is addressed in [9], [10], [16]. In particular in [10] a recursive algorithm, having the same structure as the Kalman filter, is given. Unfortunately, the solution suggested is sharp only if energy-type constraints on the noise are considered. In general, in the case of instantaneous constraints the resulting uncertainty set cannot be easily described. The approximate solution proposed is to bound the set of possible states compatible with the given observations by an ellipsoid. Equations for the center and for the weighting matrix are given.

In this paper we consider a different approach to the problem. According to previous formulation if the bounds ϵ and τ are constant and independent of k the algorithms given in Section III use all the available data, including all observations $y(0), \ldots, y(k)$ and a priori information about the process noise $u(0), \ldots, u(k)$ contained in the set K. In this case the uncertainty for k large can be too small with respect to the true one, especially when the dynamic system (35) is just a good representation of the actual system and the matrices A, B and C are not exacly known. This problem is particularly relevant when the bounds ϵ and τ are small. A similar problem arise when a statistical approach is used: the error covariance matrix becomes very small and the new observations available are ignored. This problem, generally called divergence between the estimate $\hat{x}(k+1)$ and the state $x(k+1)$, is relevant in the presence of modelling errors in statistical filtering theory (see [17]). In order to avoid these drawbacks the choice of different models of information, weighting differently new and old data, seems suitable. This technique, generally called exponential smoothing (or exponential age weighting , see [17] and [18]), well known and frequently used in statistical contexts, is now introduced in the case of unknown but bounded noise.

Assume that the following exponential bounds on process and observation noise are given

$$\|u(i)\| \leq \tau \exp(\frac{k-i}{\alpha}) \quad \alpha > 0 \quad i = 0, \ldots, k \tag{45}$$

$$\|\eta(i)\| \leq \epsilon \exp(\frac{k-i}{\beta}) \quad \beta > 0 \quad i = 0, \ldots, k \; . \tag{46}$$

In practice the number of observation to be processed is constant when k increases and the lenght of the memory can be suitably chosen by means of the constants α and β. However in this case we must guarantee that with this model of noise a solution can be computed even for $k \to \infty$. The next theorem shows that if the norm of the matrix A is sufficiently small the state $x(k+1)$ is asymptotically bounded.

THEOREM 4.

Let (45) and (46) be the bounds on process and measurement noise. If $\|A\| \exp(1/\alpha) < 1$ *then*

$$\limsup_{k\to\infty} \|x(k+1)\| \leq \frac{\tau\|B\|}{1-\exp(1/\alpha)\|A\|} \, . \tag{47}$$

PROOF:

Let us consider the solution operator given by (38) and (39)

$$\|x(k+1)\| \leq \|A^{k+1}x(0)\| + \|\sum_{i=0}^{k} A^i Bu(k-i)\| \, . \tag{48}$$

Since $\|A\| < 1$ by hypothesis the term $\|A^{k+1}x(0)\|$ vanishes for $k \to \infty$. Furthermore, using (45), (48) and the properties of the matrix norm we get

$$\limsup_{k\to\infty} \|x(k+1)\| \leq \tau \sum_{i=0}^{\infty} \exp(i/\alpha)\|A^i B\| \leq \tau\|B\| \sum_{i=0}^{\infty} \exp(i/\alpha)\|A^i\| \leq$$

$$\leq \tau\|B\| \sum_{i=0}^{\infty} \exp(i/\alpha)\|A\|^i = \tau\|B\| \sum_{i=0}^{\infty} \left(\exp(1/\alpha)\|A\|\right)^i \, . \tag{49}$$

The proof is complete by observing that the geometric series (49) is convergent if $\exp(1/\alpha)\|A\| < 1$.

∎

REMARK.

From Theorem 4 it follows that the radius is asymptotically finite if $\|A\| < 1$ and the smoothing factor α is appropriately chosen. Note that $\|A\|$ is strictly related to the spectral radius $\sigma(A)$ by means of the following formula

$$\sigma(A) \leq \|A\| \, .$$

Note that the previous inequality is sharp for symmetric matrices and the second norm. From a practical point of view it is enough to require the stability of the dynamic system and choose a norm such that $||A|| < 1$. The constant α should be appropriately chosen: if the largest eigenvalue of A is close to the unit circle then $\exp(1/\alpha)$ must be very small and a long memory should be considered.

Note that the condition $\sigma(A) < 1$ is verified for nonnegative matrices (i.e., matrices with nonnegative entries a_{ij}) for which $\sum_{i=1}^{n} a_{ij} < 1 \quad j = 1, \ldots, n$. Nonnegative matrices of this kind are frequently used in many application areas, for example in economic problems (see [19]).

∎

REMARK.

Using the classical statistical techniques the optimal estimation (in a minimum variance sense) of the state can be computed without the requirement of the stability of the system, under the assumption of complete observability and controllability. Nevertheless, in practical applications, the modelling errors may cause great problems. In fact, in the presence of parameter variations, nonlinearities, neglected unstable dynamics the requirement of the stability of the actual system is necessary in order to avoid the divergence of the algorithm (see [17] and [18] for further details on the subject).

∎

Note that, using the exponential smoothing the number of constraints on the noise is, in practice, constant when k increases. On the other hand the number of samples of the input noise $u(i)$ increases when k increases. Thus, in order to reduce the computational burden instead to approximate Sf we look for an algorithm approximating an easier solution $\widetilde{S}f$ obtained by neglecting the higher power in (39)

$$\widetilde{S}f = A^{k+1}x(0) + \sum_{i=0}^{j} A^{i}Bu(j-i) \quad , \quad j < k \; . \tag{50}$$

The number of terms in (50), i.e. j, should be chosen in order to guarantee an approximation γ required

$$\limsup_{k\to\infty} || \sum_{i=j+1}^{k} A^i B u(k-i) || \leq \gamma \,. \tag{51}$$

The constant γ should be chosen according to precision and computational requirements. It is to remark that γ may be sufficiently small since in (50) just the power of high order are neglected. Next theorem present a relation to compute j in (51).

THEOREM 5.

Let (45) and (46) be the bounds on process and measurement noise. If $||A|| \exp(1/\alpha) < 1$ *and* j *is given by*

$$\left(||A|| \exp(1/\alpha)\right)^{j+1} \leq \gamma \frac{1 - ||A|| \exp(1/\alpha)}{\tau ||B||} \tag{52}$$

then

$$\limsup_{k\to\infty} || \sum_{i=j+1}^{k} A^i B u(k-i) || \leq \gamma \,. \tag{53}$$

PROOF:

Following the line of proof of Theorem 4 (see formula (49)) we obtain

$$|| \sum_{i=j+1}^{k} A^i B u(k-i) || \leq \tau ||B|| \sum_{i=j+1}^{k} \left(\exp(1/\alpha) ||A||\right)^i \tag{54}$$

and

$$|| \sum_{i=j+1}^{k} A^i B u(k-i) || \leq \tau ||B|| \frac{\left(\exp(1/\alpha)||A||\right)^{j+1} - \left(\exp(1/\alpha)||A||\right)^{k+1}}{1 - \exp(1/\alpha)||A||} \,. \tag{55}$$

Since $\exp(1/\alpha)||A|| < 1$ for $k \to \infty$ (55) becomes

$$\limsup_{k\to\infty} || \sum_{i=j+1}^{k} A^i B u(k-i)|| \leq \tau ||B|| \frac{\left(\exp(1/\alpha)||A||\right)^{j+1}}{1-\exp(1/\alpha)||A||} \,. \tag{56}$$

If j is chosen such that

$$\left(\exp(1/\alpha)||A||\right)^{j+1} \leq \gamma \frac{1-\exp(1/\alpha)||A||}{\tau ||B||} \tag{57}$$

the right hand side of (56) is smaller than γ and relation (53) is proved.

∎

By means of easy computations formula (52) can be written as

$$(j+1)\left(1/\alpha + \ln ||A||\right) \leq \ln\left(1 - ||A|| \exp(1/\alpha)\right) + \ln \gamma - \ln \tau - \ln ||B|| \tag{58}$$

and for the approximation γ required a suitable value of j can be computed, when A, B, α, τ are given. Remark that j given by (58) is independent of k and the number of terms in the approximate solution $\widetilde{S}f$ (50) is constant when k increases.

When j and the smoothing constants α and β are chosen a complete solution to state estimation problem can be obtained by applying the results presented in Section III.

Now we show two application examples: input-output economic models and finite Markov chains.

EXAMPLE 3. *(Dynamic Leontief Models)*

In this example we describe a typical application to economics. A dynamic open Leontief model (see [20]) represents an economic situation in which the production of some industries is a consequence of the demand of market. In particular consider a situation involving n independent industries producing a single type of commodity. By $x(k)$ and $v(k)$ we denote the whole production and the demand of market at time k

$$x(k+1) = Ax(k) + v(k)$$

where the nonnegative matrix A is called Leontief input-output matrix. Since in practice the demand $v(k)$ is not exacly known we introduce the uncertainty $u(k)$ in the previous relation

$$x(k+1) = A^{k+1}x(0) + \sum_{i=0}^{k} A^i u(k-i) + \sum_{i=0}^{k} A^i v(k-i) .$$

The a priori information available, given by suitable bounds on the uncertainty $u(i)$ and on the initial production $x(0)$, is contained in the set K. The information h is given by the approximate knowledge $y(i)$ on the production $x(i)$

$$y(i) = x(i) + \eta(i) \quad i = 0, \ldots, k .$$

For a feasible Leontief model the solution Sf defined as

$$Sf = A^{k+1}x(0) + \sum_{i=0}^{k} A^i u(k-i)$$

remains bounded even for k large, since $\sigma(A) < 1$. In this case an approximation to Sf is provided by an algorithm $\varphi(h)$ and the estimated production $\hat{x}(k+1)$ is given by

$$\hat{x}(k+1) = \varphi(h) + \sum_{i=0}^{k} A^i v(k-i) .$$

An application of the optimal error predictors to dynamic Leontief models can be found in [21] when the input-output matrix is not known, the production $x(k)$ is given and the aim is the forecasting of the production $x(k+1)$. ∎

EXAMPLE 4. *(Finite Markov Chains)*

A finite Markov chain can be represented in the following form

$$x(k+1) = Ax(k)$$

where the entries a_{ij} of the matrix A are the transition probabilities between the state $x_j(k)$ and the state $x_i(k+1)$ and describe the probabilistic behavior of the Markov chain. Clearly, the coefficients a_{ij} satisfy the following relations

$$a_{ij} \geq 0 \quad i = 1, \ldots, n; \; j = 1, \ldots, n$$

$$\sum_{j=i}^{n} a_{ij} = 1 \quad i = 1, \ldots, n \; .$$

By $x_i(0)$ and $x_i(k+1)$ we denote the probability that a system is in the state i initially and after $k+1$ steps. We study the case when $x(0)$ is uncertain, i.e. bounded in a suitable way, and the aim is the estimation of $x(k+1)$. In general no information is available on the probabilities at steps $1, \ldots, k$ and the only information available is that the initial probabilities belong to a set defined by the following inequalities

$$0 \leq x_i(0) \leq \delta_i \quad i = 1, \ldots, n$$

for given positive constants $\delta_i \leq 1$.

An important class of Markov chains is that of absorbing chains (see [19]) for which the transition matrix can be permuted in the form

$$A = \begin{pmatrix} I & O \\ D & G \end{pmatrix}$$

where O is the null matrix and G is a stable matrix (i.e., $\sigma(G) < 1$). The k power of A is

$$A^k = \begin{pmatrix} I & O \\ \sum_{i=0}^{k-1} G^i D & G^k \end{pmatrix} .$$

Since G is stable we get

$$\sum_{i=0}^{\infty} G^i D = (I - G)^{-1} D .$$

A problem of major interest in this context is the estimation of the state $x(k+1)$ as $k \to \infty$. For absorbing chains, althuogh the condition $\sigma(A) < 1$ is not verified, the matrix A^k remains asymptotically bounded and the probability vector $x(k+1)$ can be recovered by means of the previous theory. ∎

V. Conclusions

In this paper problems of estimation and filtering in the presence of unknown but bounded errors are considered. In particular we have shown computable algorithms for nonidentifiable dynamic systems and/or for systems with two different sources of noise, measurement and process noise. Algorithms for some norms characterizing the uncertainty, i.e. Hilbert, l_1 and l_∞ are presented. For Hilbert norms a linear optimal algorithm is the well known minimum variance estimator. For l_∞ and l_1 norms optimal algorithms, computable by linear programming, are constructed. The problem of state estimation is formalized in the general framework of the theory. Using an exponential smoothing on the bounds of the noise it is proved that the asymptotic uncertainty of the state is bounded. Examples of dynamic Leontief models and Markov chains are presented.

Acknowledgements

I would like to thank Henryk Woźniakowski for his valuable help and comments. He made many useful suggestions that greatly improved the quality of the paper.

I would also like to thank Mark Kon for stimulating discussions on the subject.

References

[1] M. Milanese and R. Tempo, "Optimal algorithms theory for robust estimation and prediction," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 730- 738, 1985.

[2] J. F. Traub and H. Woźniakowski, *A General Theory of Optimal Algorithms.* New York: Academic, 1980.

[3] J. F. Traub, G.W. Wasilkowski and H. Woźniakowski, *Information, Uncertainty, Complexity.* Reading, MA: Addison-Weseley, 1983.

[4] C. A. Micchelli and T. J. Rivlin, "A survey of optimal recovery," in *Optimal Estimation in Approximation Theory*, C.A. Micchelli and T.J. Rivlin, Eds. New York: Plenum, 1977, pp.1-54.

[5] F. C. Schweppe, *Uncertain Dynamic Systems.* Englewood Cliffs, NJ: Prentice-Hall, 1973.

[6] R. Tempo and G. Wasilkowski, "Comparisons between maximum likelihood estimators and worst case optimal algorithms for system identification," in progress.

[7] M. Milanese, R. Tempo and A. Vicino, "Strongly optimal algorithms and optimal information in estimation problems," *J. Complexity*, vol. 2, pp. 78-94.

[8] A. Melkman and C. A. Micchelli, "Optimal estimation of linear operators in Hilbert spaces from inaccurate data," *SIAM J. Numer. Anal.*, vol. 16, pp. 87-105, 1979.

[9] B. R. Barmish and J. Sankaran "The propagation of parametric uncertainty via politopes, ". *IEEE Trans. Automat. Contr.*, vol. AC-24, pp. 346- 349, 1979.

[10] D. P. Bertzekas and I. B. Rhodes, "Recursive state estimation for a set-membership description of uncertainty," *IEEE Trans. Automat Contr.*, vol. AC-16, pp. 117-124, 1971.

[11] B. Z. Kacewicz, M. Milanese, R. Tempo and A. Vicino, "Optimality of central and projection algorithms for bounded uncertainty," to appear in *System and Control Letters*, 1986.

[12] A. G. Marchuk and K. Yu. Osipenko, " Best approximation of functions specified with an error at a finite number of points (in Russian), *Mat. Zametki*, vol. 17, pp. 359-368, 1975. (English Transl., *Mat. Notes*, vol. 17, pp. 207-212, 1975).

[13] M. Milanese, R. Tempo and A. Vicino, " Robust time series prediction by optimal

algorithms theory," in *Proc. 23th IEEE Conf. Decision Contr.*, Las Vegas, NV, 1984.

[14] A. Vicino, R. Tempo, R. Genesio and M. Milanese, "Optimal error and GMDH algorithms: a comparison with some statistical technique," to appear in *The International Journal of Forecasting*, 1987.

[15] M. Milanese and G. Belforte, "Estimation theory and uncertainty interval evaluation in presence of unknown but bounded errors: Linear families of models and estimators," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 140-142, 1982.

[16] E. Fogel and F. Huang, "On the value of information in system identification-Bounded noise case," *Automatica*, vol. 18, pp. 229-238, 1982.

[17] A. H. Jazwinski, *Stochastic Processes and Fitering Theory.* New York: Academic, 1982.

[18] P. S. Maybeck, *Stochastic Models, Estimation and Control, vol. 1.* New York: Academic, 1970.

[19] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences.* New York: Academic, 1979.

[20] W. W. Leontief, *Input-Output Economics.* London: Oxford University Press, 1966.

[21] M. Milanese, R. Tempo and A. Vicino, "Optimal error predictors for economic models," Report Dipartimento Automatica e Informatica, Politecnico di Torino (Italy), 1986.