A Convex Optimization Approach to Synthesizing Bounded Complexity ℓ^∞ Filters

Franco Blanchini and Mario Sznaier.

Abstract

This paper considers the worst–case estimation problem in the presence of unknown but bounded noise. Contrary to stochastic approaches, the goal here is to confine the estimation error within a bounded set. Previous work dealing with the problem has shown that the complexity of estimators based upon the idea of constructing the state consistency set (e.g. the set of all states consistent with the a-priori information and experimental data) cannot be bounded a-priori, and can, in principle, continuously increase with time. To avoid this difficulty in this paper we propose a class of bounded complexity filters, based upon the idea of confining r–length error sequences (rather than states) to hyperrectangles. The main result of the paper shows that this can be accomplished by using Linear Time Invariant (LTI) filters of order no larger than r. Further, synthesizing these filters reduces to a combination of convex optimization and line search. Finally, as we show here, these filters are globally optimal in a restricted information scenario, where accurate information about past trajectories is traded-off against filter complexity.

I. INTRODUCTION

Classical stochastic estimation methods are not well suited for situations where it is of interest to obtain hard bounds on estimation errors or where the only information available on exogenous disturbances is a bound on a suitable norm (or, alternatively, a set-membership characterization). These cases can be handled by resorting to a deterministic, unknown-but-bounded approach where the goal is to design an estimator that minimizes, in a suitable sense, the worst case estimation error due to exogenous inputs only known to belong to a given set. Initial work in this area dates back to the early 70's [13], [3], where it was shown that in the case of ℓ^2 bounded exogenous disturbances, the set of states consistent with the experimental observations is an ellipsoid whose center and covariance matrix can be recursively obtained via a Kalman–filter like estimator. Unfortunately, this is no longer the case for point-wise in time (e.g. ℓ^∞ like) constraints on the disturbance. In this case, even constraining the disturbances to belong to an ellipsoid at each point in time does not lead to easily characterizable consistency sets for the states, although these sets can be conservatively overbounded by an ellipsoid.

Worst case estimation in the presence of ℓ^∞ bounded disturbances was studied in [9], [11], [18] (see also the survey [10]). The main result of these papers shows that pointwise optimal estimators can be obtained as the product of a subset of past measurements and a (time varying) gain. Both the gain and the set of relevant measurements result from solving a linear programming

Franco Blanchini has been partially supported by the Italian Ministry for University and Research and by the Italian National Council of Research (C.N.R.). M. Sznaier has been partially supported by NSF grants IIS–0713003, ECCS-0901433, and AFOSR grant FA9550–09–0253.

Corresponding author,Dept. of Mathematics and Computer Science, University of Udine, Italy. (blanchini@uniud.it)

Dept. of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA (msznaier@coe.neu.edu)

optimization problem. However, this optimization problem involves all past measurements. Thus, the complexity of these estimators grows with time. In the case of stable systems, given $\epsilon > 0$, ϵ–suboptimal approximations can be found by simply dropping all measurements older than an a-priori pre–computable horizon $N(\epsilon)$. Still, guaranteeing a small approximation error requires large values of N (see [18] for details.) Moreover, the filter is non–recursive, in the sense that current estimates are obtained by solving an LP problem that involves all available information, rather than by propagating past estimates.

The use of nonlinear recursive filters was proposed in [17], where the idea is to bound the set of possible states consistent with the output observations by a set whose center is propagated recursively and whose shape can be found by solving (at each instant) an optimization problem. Still, the complexity of the resulting observer is potentially high and its sub-optimality properties hard to ascertain.

A semi–recursive algorithm was proposed in [21]. In the case of known initial conditions, the optimal ℓ^∞ estimation problem is reduced to an ℓ^1 model matching problem [2], [6], [5] that can be solved (with arbitrary precision) by using the techniques in [5]. The case of unknown initial conditions is handled by first pre–computing an horizon N after which the estimation error due to these initial conditions falls below a pre-specified error level ϵ. The complete, semi–recursive estimator is obtained by using a non–recursive pointwise optimal estimator similar to that in [18] for the first N-1 time steps, switching afterwards to the recursive ℓ^1 estimator. Since this estimator is based on solving a 2–block ℓ^1 model matching problem its complexity (and hence that of the overall estimator) cannot be bounded a priori.

An alternative approach involves set–valued observers [15], [16], where pointwise optimal estimators are obtained by recursively applying the Fourier–Motzkin algorithm to construct a polyhedral set guaranteed to contain the states of the plant. An ℓ^∞ point–wise optimal estimator is then obtained from these sets, by simply using as estimate of the unknown output z the center z_c of the set of all output values compatible with the present set estimate of the state. However, propagation of these estimates is not recursive, e.g. $z_c(k+1)$ cannot be directly constructed from the past estimates $z_c(k-i)$. Moreover, in principle the complexity of the estimator (measured in terms of the number of hyperplanes defining the set observer) is not bounded a-priori and increases with time.

Motivated by the high complexity entailed in the approaches above, the goal of this paper is to synthesize *fixed* order recursive filters for systems subject to ℓ^∞ bounded disturbances, with guaranteed worst case estimation error. The main idea underlying the proposed approach is to, rather than attempting to confine the state of the system to a given set, to simply confine the estimation error to hyperrectangles. Intuitively, this amounts to willingly dropping information in return for obtaining bounded complexity filters. Our main results show that the problem of synthesizing bounded complexity filters that confine the error to the tightest possible hyperrectangle, for a set of suitable initial conditions, can be reduced to a combination linear programming/line search. For initial conditions outside this set, the estimation error converges, in finite time, to the design value. Moreover, as we show in the paper, these filters are globally optimal, that is, they achieve the lowest possible worst-case estimation error amongst all filters restricted to this simplified information scenario.

II. Preliminaries

A. Notation

For ease of reference, the notation used in the paper is summarized below:

$\|y\|_\infty$	∞ norm of the vector $y \in R^n$: $\|y\|_\infty \doteq \max_i \|y\|_i$.
$M(j,:)$	j^{th} row of the matrix M.
$\|M\|_1$	$\infty \to \infty$ induced norm of matrix $M \in R^{n\times m}$: $\|M\|_1 \doteq max_i \sum_j \|M_{ij}\|$
ℓ_n^1, ℓ_n^∞	extended Banach spaces of vector valued real sequences $\{y\}_0^\infty \in R^n$ equipped with the norms $\|y\|_{\ell^1} \doteq \sum_{i=0}^\infty \|y_i\|_\infty$ and $\|y\|_{\ell^\infty} \doteq \sup_i \|y_i\|_\infty$, respectively.
$\mathcal{B}\ell^1$, $\mathcal{B}\ell^\infty$	unit balls in ℓ^1, ℓ^∞.
$\mathcal{B}\ell^\infty(\mu)$	scaled unit ball in ℓ_n^∞. Given $\mu \doteq \left[\mu_1 \dots \mu_n\right]$ $$\mathcal{B}\ell^\infty(\mu) \doteq \{e \in \ell_n^\infty \colon e_i/\mu_i \in \mathcal{B}\ell^\infty\}$$
$\|G\|_{\ell^\infty\to\ell^\infty}$	ℓ^∞ to ℓ^∞ induced norm of the operator $G : \ell^\infty \to \ell^\infty$, e.g. $\|G\|_{\ell^\infty\to\ell^\infty} \doteq \sup_{y\neq 0} \frac{\|Gy\|_{\ell^\infty}}{\|y\|_{\ell^\infty}}$
$Y(\lambda)$	λ–transform of a sequence $\{y_k\}_0^\infty$ $Y(\lambda) \doteq \sum_{i=0}^\infty y_k \lambda^k$

In the sequel, we will associate to a scalar ARMA model of the form

$$y(k) = -\sum_{i=1}^{n} a_i y(k-i) + \sum_{i=0}^{m} b_i v(k-i);\ n \geq m \tag{1}$$

its λ–transform representation[1]

$$y(\lambda) = \frac{\sum_{i=0}^n b_i \lambda^i}{\sum_{i=0}^m a_i \lambda^i} v(\lambda) \tag{2}$$

B. Definitions and Preliminary Results

The notion of equalized performance, introduced in [4] (see also [12]) will play a key role in obtaining bounded complexity filters.

Definition 2.1: Consider an LTI plant described by a model of the form (1). Given $r \geq n$, the plant is said to achieve an equalized r–performance level μ if, whenever the input and output sequences $\{v\}, \{y\}$ satisfy $|v(t)| \leq 1$ and $|y(t)| \leq \mu$ for all $t = k, k-1, \dots, k-r+1$, then $\|y(k+1)\| \leq \mu$ (thus $\|y(k+i)\| \leq \mu$, for $i > 0$). In particular the case $r = n$ will be simply referred to as equalized performance.

As shown in [4], only superstable plants (in the sense of [12]) achieve (finite) equalized performance. However, any stable plant achieves finite equalized r–performance for some large enough r. Further, if a SISO plant achieves r–performance μ for some finite r, the it achieves r'–performance μ for any $r' > r$.

Next, we recall, for ease of reference, some properties concerning the relationship between equalized performance and the ℓ^∞ induced norm.

Lemma 2.1 ([4]): Given a stable, LTI SISO plant $y(\lambda) = G(\lambda)v(\lambda)$, as in (2) with finite r-equalized performance $\mu(r_o)$ for some $r_o \geq n$, the following holds:

1) $\|G\|_{\ell^\infty\to\ell^\infty} \leq \mu(r_o)$, with the equality holding for finite impulse response (FIR) plants.
2) $\mu(r) \downarrow \|G\|_{\ell^\infty\to\ell^\infty}$.

[1] this corresponds to setting $\lambda = 1/z$ in the usual z-transform representation.

C. *Why equalized filtering?*

As already shown in [14], [17] recursive set valued observers based upon the idea of propagating a set known to contain the (unknown) state of the plant have high complexity. To avoid this difficulty, in this paper, rather than attempting to confine the state, we will work directly with the estimation error and attempt to design a filter such that, if at some time instant t_o the past r values of the error are "captured" in an r-hyperrectangle, then this property will hold for all $t > t_o$ and all $||v||_{\ell^\infty} \leq 1, \|w\|_{\ell^\infty} \leq 1$. Further, we are interested in synthesizing the tightest hyperrectangle satisfying this property. The main result of this paper shows that this can be accomplished by reducing the problem to an equalized performance one. Moreover, contrary to the controller design case considered in [4], in the filtering case the results are easily extended to MIMO systems by simply considering a collection of component–wise filters.

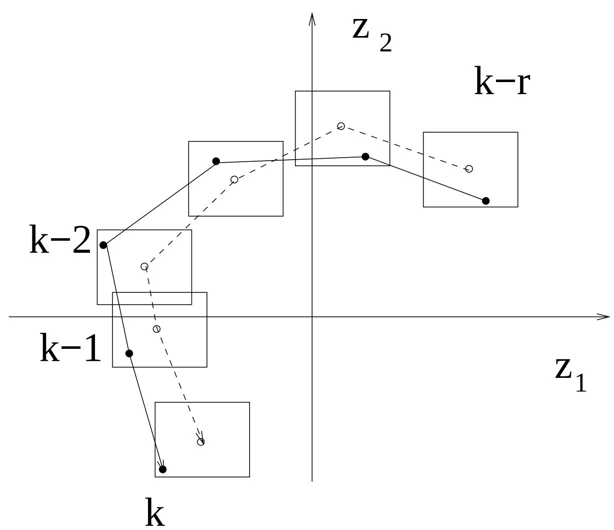

Fig. 1. The equalized filtering idea: full dots are the true trajectory the empty dots are the estimated trajectory

III. Problem setup and preliminary results

Consider an LTI plant subject to ℓ^∞ bounded disturbances, with state space realization:

$$\begin{array}{rcl} x_{k+1} & = & Ax_k + Bv_k \\ z_k & = & Hx_k \\ y_k & = & Cx_k + Dw_k \end{array} \tag{3}$$

or with λ–transform representation

$$z(\lambda) = \frac{M(\lambda)}{d(\lambda)} v(\lambda) \tag{4}$$

$$y(\lambda) = \frac{N(\lambda)}{d(\lambda)} v(\lambda) + Dw(\lambda) \tag{5}$$

where $z \in \mathbb{R}^s$, $y \in \mathbb{R}^q$, $v \in \mathbb{R}^p$ and $w \in \mathbb{R}^q$ represent the output to be estimated, the measurements available to the filter, process and measurement noise, respectively, and where

$d(\lambda) = det(I - \lambda A)$. Note that we have assumed that the plant is strictly proper with respect to the input v. This assumption is made for notational simplicity and can be removed at the price of a more involved notation in the subsequent development. For the time being, we will also assume that z is a scalar, but this assumption will be relaxed later. Our goal is to design a filter of the form:

$$\hat{z}(\lambda) = \frac{B(\lambda)}{a(\lambda)} y(\lambda) \tag{6}$$

such that the estimation error

$$e(\lambda) = z(\lambda) - \hat{z}(\lambda) \tag{7}$$

is confined to an hyperrectangle. The complete filtering scheme is illustrated in Fig. 2. In the

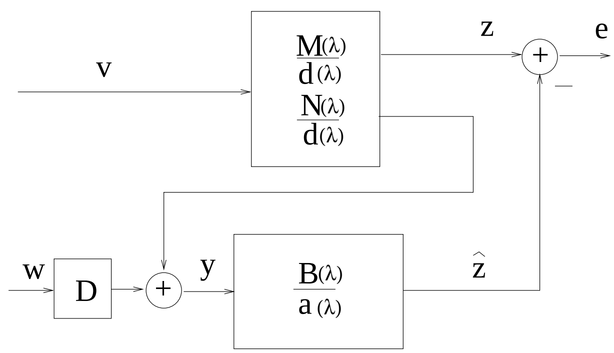

Fig. 2. The filtering scheme.

sequel, we will limit our attention to filters that belong to the class of generalized Luenberger observers, defined as follows:

Definition 3.1 ([7]): A system of the form

$$\xi_{k+1} = P\xi_k + Ly_k \tag{8}$$

$$\hat{x}_k = Q\xi_k + Ry_k \tag{9}$$

$$\hat{z}_k = H\hat{x}_k \tag{10}$$

is a generalized Luenberger (state) observer for system (3) if P is a stable matrix and $\hat{x}_k - x_k \rightarrow 0$ as $k \rightarrow \infty$, when $w(k) \equiv 0$ and $v(k) \equiv 0$.

Next we recall a characterization of the class of the generalized observers.

Lemma 3.1: The system (8)–(10) is a generalized observer for (3) iff P is stable and there exists a full column rank matrix T such that

$$TA - LC = PT, \tag{11}$$

$$QT + RC = I, \tag{12}$$

Proof: See [7], [19]. ■

Remark 3.1: The standard (n–order) Luenberger observer corresponds to the choice $T = I$ and $R = 0$. Selecting a "tall" T matrix leads to a higher order observer, with additional degrees of freedom that can be used to optimize performance.

Next we show that restricting the filter to be a generalized observer imposes a constraint on its structure.

Lemma 3.2: If the filter (6) is a generalized state observer for system (3), then the polynomial matrices $M(\lambda)$ (of dimension $1 \times p$), $N(\lambda)$ (of dimension $q \times p$), $B(\lambda)$ (of dimension $1 \times d$) and the polynomials $a(\lambda)$ and $d(\lambda)$ satisfy the following condition:

$$M(\lambda)a(\lambda) - B(\lambda)N(\lambda) = C(\lambda)d(\lambda) \tag{13}$$

for some polynomial matrix $C(\lambda)$.

Proof: From equations (8)–(12) it follows that:

$$\begin{array}{rcl} [Tx - \xi]_{k+1} & = & TAx_k - P\xi_k - L(Cx_k + Dw_k) + TBv_k \\ & = & P[Tx_k - \xi_k] + TBv_k - LDw_k \\ x_k - \hat{x}_k & = & x_k - Q\xi_k - RCx_k - RDw_k \\ & = & Q[Tx_k - \xi_k] - RDw_k \end{array} \tag{14}$$

Consider now the change of variables $\eta = x$ and $\theta = [Tx - \xi]$. In term of these variables the state space representation of the combined plant–filter system is given by:

$$\begin{array}{rcl} \begin{bmatrix} \eta_{k+1} \\ \theta_{k+1} \end{bmatrix} & = & \begin{bmatrix} A & 0 \\ 0 & P \end{bmatrix} \begin{bmatrix} \eta_k \\ \theta_k \end{bmatrix} + \begin{bmatrix} B & 0 \\ TB & -LD \end{bmatrix} \begin{bmatrix} v_k \\ w_k \end{bmatrix} \\ e_k & = & \begin{bmatrix} 0 & HQ \end{bmatrix} \begin{bmatrix} \eta_k \\ \theta_k \end{bmatrix} + \begin{bmatrix} 0 & -HRD \end{bmatrix} \begin{bmatrix} v_k \\ w_k \end{bmatrix} \end{array}$$

Thus η is unobservable from e. Hence, the modes of A are canceled in the transfer function $T_{e,\eta}$. From (4)–(7) it follows that:

$$\begin{array}{rcl} e(\lambda) & = & \left[\dfrac{M(\lambda)}{d(\lambda)} - \dfrac{B(\lambda)}{a(\lambda)}\dfrac{N(\lambda)}{d(\lambda)}\right] v(\lambda) + \left[\dfrac{B(\lambda)}{a(\lambda)}\right] Dw(\lambda) \\ & = & \left[\dfrac{M(\lambda)a(\lambda) - B(\lambda)N(\lambda)}{a(\lambda)d(\lambda)}\right] v(\lambda) + \left[\dfrac{B(\lambda)}{a(\lambda)}\right] Dw(\lambda) \end{array}$$

Since $d(\lambda) = det(I - A\lambda)$, the cancellation of the modes of A in $T_{e\eta}$ implies that $M(\lambda)a(\lambda) - B(\lambda)N(\lambda)$ has $d(\lambda)$ as a factor, precisely what (13) states. ■

Since we are interested in generalized observer like filters, in the sequel we will limit our attention to polynomial matrices satisfying (13) for some $C(\lambda)$. It is trivial to show that in this case the estimation error is governed by the equation

$$e(\lambda) = \frac{C(\lambda)}{a(\lambda)} v(\lambda) + \frac{B(\lambda)}{a(\lambda)} Dw(\lambda) \tag{15}$$

IV. Equalized performance filtering

We are now in the position to formally state the equalized–performance filtering problem.

Problem 4.1: Given an integer $r \geq n$ and $\mu > 0$ find a filter of the form (6) of order r satisfying the constraint (13) and such that $a(\lambda)$ is stable (i.e. all its poles are outside the unit circle) and

$$\begin{array}{l} |e_{k-j}| \leq \mu, j = 1, 2, \ldots, r \Rightarrow |e_t| \leq \mu \\ \text{for all } t \geq k \text{ and all sequences } v, w \in \mathcal{B}\ell^\infty \end{array} \tag{16}$$

Note that the problem above does not explicitly make any assumptions on x_o, the initial conditions of the plant. As we will show later, if the plant achieves an equalized performance level $\mu < \infty$, then there exist a set of initial conditions $\mathcal{X}_o(\mu)$ such that if $x_o \in \mathcal{X}_o(\mu)$ then $|e_k| \leq \mu$ for all k. For initial conditions outside this set, the condition will be satisfied after a finite number of steps.

Theorem 4.1: An r^{th} order filter of the form (6) with:

$$\begin{aligned} a(\lambda) &= 1 + a_1\lambda + \cdots + a_r\lambda^r \\ B(\lambda) &= B_0 + B_1\lambda + \cdots + B_r\lambda^r \\ C(\lambda) &= C_0 + C_1\lambda + \cdots + C_r\lambda^r \end{aligned}$$

solves Problem 4.1 above if and only if

$$\begin{aligned} \mu\|[a_1 \ a_2 \ldots a_r]\|_1 &+ \|[C_0 \ C_1 \ldots C_r]\|_1 \\ &+ \|B_0 D \ \ldots B_r D\|_1 \leq \mu \end{aligned} \tag{17}$$

Proof: From (15) the ARMA model relating the signals e, v, w is given by

$$e_k = -\sum_{i=1}^{r} a_i e_{k-i} + \sum_{i=0}^{r} C_i v_{k-i} + \sum_{i=0}^{r} B_i D w_{k-i} \tag{18}$$

Thus, if $|e_{k-i}| \leq \mu$ and $i = 1, 2, \ldots, r$, $v, w \in \mathcal{B}\ell^\infty$ then

$$\begin{aligned} |e_k| &= |-\sum_{i=1}^{r} a_i e_{k-i} + \sum_{i=0}^{r} C_i v_{k-i} + \sum_{i=0}^{r} B_i D w_{k-i}| \\ &\leq \sum_{i=1}^{r} |a_i||e_{k-i}| + \left\| [C_0 \ C_1 \ldots C_r] \begin{bmatrix} v_{k-1} \\ : \\ v_{k-r} \end{bmatrix} \right\|_\infty + \left\| [B_0 D \ B_1 D \ldots B_r D] \begin{bmatrix} w_{k-1} \\ : \\ w_{k-r} \end{bmatrix} \right\|_\infty \\ &\leq \sum_{i=1}^{r} |a_i|\mu + \|[C_0 \ C_1 \ldots C_r]\|_1 + \|[B_0 D \ B_1 D \ldots B_r] D\|_1 \leq \mu \end{aligned}$$

Therefore the condition is sufficient. To prove necessity, start by rewriting (18) as

$$\begin{aligned} |e(k)| &= |[\mu a_1 \ldots \mu a_r \ C_0 \ldots C_r \ B_0 D \ldots B_r D]x| \\ &\doteq |\Xi x| \end{aligned}$$

where $x \doteq \begin{bmatrix} e_{k-1}/\mu \ldots e_{k-r}/\mu & v_k \ldots v_{k-r} w_k \ldots & w_{k-r} \end{bmatrix}^T$. From the hypothesis it follows that x is an arbitrary element of $\mathcal{B}\ell^\infty$. Hence

$$\sup_{\|x\|_\infty \leq 1} |e_k| \leq \mu \iff \|\Xi\|_1 \leq \mu$$

or, equivalently,

$$|e(k)| = |\mu a_1| + |\mu a_2| + \ldots |\mu a_r| + \|[C_0 \ C_1 \ldots C_r]\|_1 + \|[B_0 D \ B_1 D \ldots B_r D]\|_1 \leq \mu$$

which proves necessity. To conclude the proof, we need to establish that condition (17) implies stability of the filter. This follows immediately from the fact that it implies

$$\|[a_1 \ a_2 \ldots a_r]\|_1 = \sum_{i=1}^{r} |a_i| = \rho < 1$$

■

Remark 4.1: In the sequel we will refer to filters satisfying (17) as r–equalized filters, with performance μ or, whenever clear from the context, simply as equalized filters. Moreover, the value of μ such that the equality holds in (17) will be referred to as the optimal equalized filtering level, and denoted by μ_{opt}. As we will show in section VII, μ_{opt} is indeed the radius of information (under the precise conditions stated there) and thus the lowest possible worst-case error attainable by any filter.

V. Optimal fixed–order synthesis

As shown next, the main advantage of the proposed approach is that fixed complexity equalized filters can be efficiently synthesized via a combination of convex optimization and line search. To establish this, begin by rewriting (13) as the following linear constraint in the variables a_k, B_k and C_k:

$$\mathcal{M}\mathcal{A} - \mathcal{B}\begin{bmatrix} N_0 \\ N_1 \\ \vdots \\ N_l \end{bmatrix} = \mathcal{C}\mathcal{D}$$

where

$$\mathcal{M} \doteq \begin{bmatrix} M_0 & 0 & \dots & 0 & 0 \\ M_1 & M_0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ M_r & M_{r-1} & \dots & M_1 & M_o \\ \vdots & \vdots & \dots & \vdots & \vdots \\ M_n & M_{n-1} & \dots & M_{n-r+1} & M_{n-r} \\ 0 & M_n & M_{n-1} & \vdots & \vdots \\ \vdots & \vdots & 0 & M_n & M_{n-1} \\ 0 & 0 & \dots & 0 & M_n \end{bmatrix}$$

$$\mathcal{A} \doteq \text{block-diag}\{a, a, \dots, a\},\ a \doteq \begin{bmatrix} 1 & a_1 & \dots & a_r \end{bmatrix}^T$$

$$\mathcal{B} \doteq \begin{bmatrix} B_0 & 0 & \dots & 0 & 0 \\ B_1 & B_0 & \dots & 0 & 0 \\ \vdots & B_1 & \dots & B_0 & 0 \\ B_r & \vdots & \dots & B_1 & B_0 \\ 0 & B_r & \dots & \vdots & \vdots \\ 0 & \vdots & \dots & B_r & B_{r-1} \\ 0 & 0 & \dots & 0 & B_r \end{bmatrix},\ \mathcal{C} \doteq \begin{bmatrix} C_0 & 0 & \dots & 0 & 0 \\ C_1 & C_0 & \dots & 0 & 0 \\ \vdots & C_1 & \dots & C_0 & 0 \\ C_r & \vdots & \dots & C_1 & C_0 \\ 0 & C_r & \dots & \vdots & \vdots \\ 0 & \vdots & \dots & C_r & C_{r-1} \\ 0 & 0 & \dots & 0 & C_r \end{bmatrix},$$

$$\mathcal{D} \doteq \text{block} - \text{diag}\{d, d, \dots, d\},\ d \doteq \begin{bmatrix} 1 & d_1 & \dots & d_r \end{bmatrix}^T$$

Since for a fixed μ, (17) is also convex in a_k, B_k and C_k, it follows that establishing feasibility of (13)–(17) reduces to a convex problem. Finally, the optimal filter (and its associated optimal filtering error μ_{opt}) can be found via bisection. These observations are summarized in the following algorithm.

Algorithm 5.1: 0.- Select $\mu > 0$, tolerances ϵ and δ, and set $\mu^- = 0$.

1.- Solve the feasibility problem (13)–(17). If it is unfeasible set $\mu = 2\mu$ and go to step 1, else set $\mu^+ = \mu$.
2.- Solve the feasibility problem for $\mu = (\mu^+ + \mu^-)/2$.
3.- If it is feasible, set $\mu^+ = \mu$ else set $\mu^- = \mu$.
4.- If $\mu^+ - \mu^- < \delta$ then set $\mu_{opt} = \mu$ and STOP, else go to step 3.

A. The multi–output case

In the previous sections we considered the case where z, the quantity to be estimated, is a scalar. The main result of this section shows that these results can be extended to the multiple outputs case, $z \in R^s$ by simply considering an array of single–output filters, each of which estimates one of the components of z. To this effect, we begin by extending the definition of equalized filtering performance to the multi–output case.

Definition 5.1: The filter (6) with error $\hat{z} - z = e \in R^s$ is said to achieve a vector equalized performance level $\mu \doteq \left[\mu_1, \mu_2, \ldots \mu_s\right]$ if it is stable and:

$$\begin{array}{l} e_{k-j} \in \mathcal{B}\ell^\infty(\mu), j = 1, 2, \ldots, r \Rightarrow e_k \in \mathcal{B}\ell^\infty(\mu); \\ \text{for all sequences } v, w \in \mathcal{B}\ell^\infty \end{array} \tag{19}$$

The next result shows that vector equalized performance is equivalent to component–wise scalar equalized performance.

Theorem 5.1: A filter $F\colon y \in \ell^\infty_n \to \hat{z} \in \ell^\infty_s$ achieves a vector equalized performance level μ iff each component $F_i\colon y \in \ell^\infty_n \to \hat{z}^i \in \ell^\infty$ achieves scalar equalized performance (in the sense of (16)) μ_i, where $\hat{z}^i$ denotes the i^{th} component of $\hat{z}$.

Proof: Clearly, if each component F_i achieves an equalized performance level μ_i, the overall filter F obtained by stacking each component satisfies the conditions in Definition 5.1. Conversely, assume that the filter F satisfies (19). Note that the multiple–output version of the filter (18) can be written in terms of its h component as follows

$$e^h_k = -\sum_{i=1}^{r} a^h_i e^h_{k-i} + \sum_{i=0}^{r} C^h_i v_{k-i} + \sum_{i=0}^{r} B^h_i D w_{k-i} \tag{20}$$

and that the error terms e^h_{k-i}, $i = 1, 2, \ldots, r$ can be initialized independently in each "partial filter". Assume now that for a given i the corresponding mapping F_i does not satisfy (16). Then, it is easily seen that initializing all the other variables e^j, $j \neq h$, to $e^j_{k-i} = 0$, $i = 1, 2, \ldots, r$ leads to violation of (19). ∎

Hence, optimal MIMO filters can be synthesized by simply applying Algorithm 5.1 componentwise.

VI. Filter initialization

In this section we consider the problem of filter initialization. The main result shows that, given an initial set of r measurements, $\mathbf{y} \doteq \left[y_o, y_1, \ldots, y_{r-1}\right]$ there exists a finite performance level μ and a filter initial condition ξ_o such that the estimation error satisfies $e_k \in \mathcal{B}\ell^\infty(\mu)$ for all k. Let $\mathcal{K}_o$, T_y and T^j_z denote the r^{th} order Kalman observability matrix of the system (3) and the Toeplitz operators mapping v to y and the j^{th} component of z, respectively, e.g.

$$\mathcal{K}_o = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{r-1} \end{bmatrix} \quad T_y = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ CB & 0 & 0 & \dots & 0 \\ CAB & CB & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ CA^{r-1}B & CA^{r-2}B & \dots & CB & 0 \end{bmatrix}$$

$$T_z^j \doteq \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ H(j,:)B & 0 & 0 & \dots & 0 \\ H(j,:)AB & H(j,:)B & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ H(j,:)A^{r-1}B & H(j,:)A^{r-2}B & \dots & H(j,:)B & 0 \end{bmatrix}$$

then, the filter can be initialized (after r measurements) as follows (see Fig. 3).

1) For $k = 0, \dots, r-1$ and $j = 1, \dots, s$ compute:

$$\begin{aligned} z_k^{j,+} &\doteq \max_{x,v,w} H(j,:)A^{k-1}\mathbf{x} + T_z^j(k,:)\mathbf{v} \\ z_k^{j,-} &\doteq \min_{x,v,w} H(j,:)A^{k-1}\mathbf{x} + T_z^r(k,:)\mathbf{v} \end{aligned} \tag{21}$$

$$\begin{aligned} &\text{subject to: } \mathbf{x} \in \tilde{\mathcal{X}}_o \\ &\mathbf{y} = T_y\mathbf{v} + \mathcal{D}\mathbf{w},\ \mathbf{w}, \mathbf{v} \in \mathcal{B}\ell_\infty \end{aligned}$$

where $\mathcal{D} \doteq \text{diag}\{D^T\}$ and where $\tilde{\mathcal{X}}_o$ is a set known to contain the initial condition (if no information is available then $\tilde{\mathcal{X}}_o = R^n$).

2) Define:

$$\begin{aligned} z_k^{j,c} &\doteq \frac{z_k^{j,+} + z_k^{j,-}}{2}, \\ \mu_k^j &\doteq \frac{1}{2}|z_k^{j,+} - z_k^{j,-}| \end{aligned} \tag{22}$$

3) Let $\mu^{init,j} = \max_{0 \le t \le r-1} \{\mu_t^j\}$ and choose a filter initial condition such that the first r filter estimates are $\hat{z}_t = z_t^{j,c},\ t = 0, \dots, r-1$. This is always feasible, since the order of the filter is precisely r.

Note that if $\mathcal{X}_o$ is convex, then the optimization problem (21) is convex Further, if $\mathcal{X}_o$ is a polytope, this problems reduces to LP. Thus z^+, z^-, and z^c above can be found efficiently. [2].

Since $\| \begin{bmatrix} a_1 \dots a_r \end{bmatrix} \|_1 < 1$ by construction, it can be easily shown that if (17) holds for some $\tilde{\mu}$, then it also holds for all $\mu \ge \tilde{\mu}$, and so does (16). It follows that if $\mu^{init,j} \le \mu_{opt}^j$ the optimal equalized performance level in (17), then the filter (6), with the initialization above, achieves optimal equalized performance level μ_{opt}^j for all $t \ge r$. On the other hand, as we show next, if $\mu^{init,j} > \mu_{opt}^j$, then the worst case ℓ^∞ estimation error is bounded above by $\mu^{init,j}$ and converges, in a finite number of steps, to μ_{opt}^j.

Theorem 6.1: Consider a filter of the form (6), with the initialization above. Then, for all t, $|e_t| \le \mu^{init}$. Moreover, given $\mu \ge \mu_{opt}$ satisfying (17), for any plant and filter initial condition pairs $\{x_o, \xi_o\}$ there exists a finite time $T(x_o, \xi_o, \mu)$ such that for all $t > T$, $|e_t| \le \mu$.

[2]The estimate z_c can be thought off as a smoothing problem equivalent of the central estimator introduced in [14] or, equivalently, the pointwise optimal estimators proposed in [18].

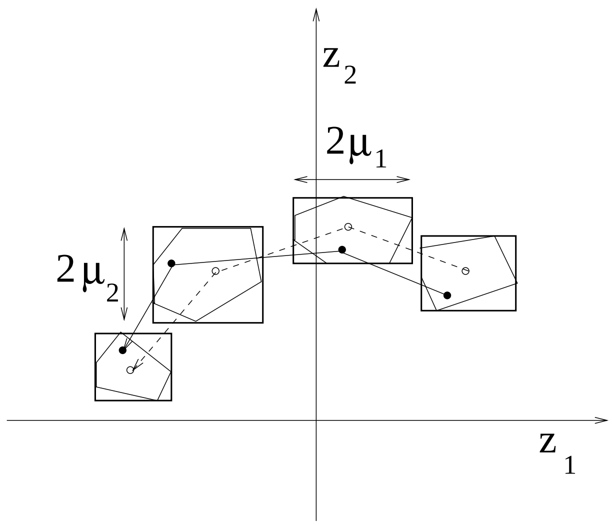

Fig. 3. The filter initialization

Proof: Consider the "Lyapunov like" function

$$\psi_k \doteq \max_{i=1,\dots,r} \quad |e_{k-i}|$$

From(18) we have that:

$$\begin{aligned} |e_k| &\leq \sum_{i=1}^{r} |a_i|\psi_k + \|[C_0 \dots C_r]\|_1 + \|[B_0 \dots B_r]D\|_1 \\ &= \sum_{i=1}^{r} |a_i|\psi + \|[C_0 \dots C_r]\|_1 + \|[B_0 \dots B_r]D\|_1 \\ &+ \sum_{i=1}^{r} |a_i|(\psi_k - \mu) \\ &< \mu + (\psi_k - \mu)\sum_{i=1}^{r} |a_i| = \mu + (\psi_k - \mu)\rho \end{aligned}$$

with $\rho < 1$. Thus

$$|e_k| < \rho\psi_k + (1-\rho)\mu \leq \max\{\mu, \psi_k\}$$

Hence the sequence ψ_k is non increasing as long as $\psi_k > \mu$. Since by construction $|e_k| \leq \mu^{init}, 0 \leq t \leq r-1$, it follows that $|e_t| \leq \mu^{init}$ for all t. Further, due to the strict inequality, the subsequence ψ_{kr} (r is the parameter in Definition 2.1) is strictly decreasing (since each value generated after $|e_k|$ is strictly smaller than ψ_k). Indeed,

$$\psi_{(k+1)r} < \max\{\rho\psi_{kr} + (1-\rho)\mu, \mu\}$$

This equation implies that the subsequence ψ_{kr} and (thus ψ_k) converges to μ from above, and it can be easily seen that this convergence occurs in finite time, e.g. $\psi_t = \mu$ for t $>$ some T. The fact that $|e_t| \leq \mu,\ t > T$ follows immediately follows from the definition of ψ. ■

VII. Optimality Considerations

As briefly outlined in section II-C, the intuition underlying this paper is to avoid the high (potentially infinite) complexity entailed in set valued observers by purposedly dropping information. That is, as illustrated in Figure 1, rather than keeping the full information about the past, an r^{th} order equalized filter only "remembers" that for the past r time instants, the quantity to be estimated was confined to a hyperrectangle of size μ_{opt}, centered around the past filter estimates, $\hat{z}_{k-j},\ j = 0, \ldots, r$. Thus, a legitimate question is whether other filters, operating on this same restricted information, can yield better performance, measured in terms of the worst-case estimation error. The main result of this section shows that this is not the case. Indeed, equalized filters are globally optimal, over the set of all causal filters.

A. Background on Information Based Complexity

In this section we recall some key results from Information Based Complexity (IBC) required to establish optimality of the equalized filters. For simplicity, we consider the case of bounded LTI operators in ℓ^∞. A general treatment can be found for instance in the book [20].

Let K denote a set in ℓ^∞ and consider two linear operators[3] $S_y, S_z \colon \ell^\infty \to \ell^\infty$. In this context, the estimation problem can be stated as, given an element $f_o \in K$, find an estimate $\hat{z}$ of $z \doteq S_z f_o$ using noisy experimental information $y = S_y f_o + \eta$, where the noise η is only known to belong to some bounded set $\mathcal{N} \subset \ell^\infty$. Note that in general S_y is not invertible, and thus it is not possible to recover f_o, even in the absence of noise. This is related to the concept of consistency set, defined as:

$$\mathcal{T}(y) \doteq \{f \in K \colon y = S_y f + \eta \text{ for some } \eta \in \mathcal{N}\} \tag{23}$$

that is, the set of all possible elements in K that could have generated the observed data.

Given an estimation algorithm $\hat{z} = \mathcal{A}(y)$ (not necessarily linear), it is of interest to compute its worst case approximation error. For a given measurement y, the *local* error $\epsilon(y, \mathcal{A})$ of a given algorithm $\mathcal{A}$ is defined as the worst case distance between the true quantity z and its estimate $\mathcal{A}(y)$. Since all the elements f that could have generated y belong to $\mathcal{T}(y)$, it follows that:

$$\epsilon(\mathcal{A}, y) \doteq \sup_{f \in \mathcal{T}(y)} \|S_z f - \mathcal{A}(y)\|_\infty \tag{24}$$

Similarly, the global error $\epsilon(\mathcal{A})$ of an algorithm $\mathcal{A}$ is defined as the worst possible case over all possible measurement sequences, that is:

$$\epsilon(\mathcal{A}) \doteq \sup_y \epsilon(\mathcal{A}, y) \tag{25}$$

Definition 1: An algorithm $\mathcal{A}_o(.)$ is said to be globally optimal if $\epsilon(\mathcal{A}_o) = \inf_{\mathcal{A}} \epsilon(\mathcal{A})$.
The minimum global error $\epsilon(\mathcal{A}_o)$ is called the radius of information, $r(I)$. It provides a lower bound on achievable performance, since no estimation algorithm can have smaller global worst case error. Further, it is a standard fact in IBC that, under mild conditions, $r(I)$ can be explicitly computed. This result, quoted below for ease of reference, will be used to establish that equalized filters are globally optimal.

[3]In the IBC literature, S_y and S_z are usually referred to as the information and solution operators, respectively.

Theorem 7.1: [8] Assume that the sets K and $\mathcal{N}$ are convex, balanced. Then

$$r(I) = \sup_{f \in \mathcal{T}(0)} \|S_z f\|_\infty$$

Note that this result shows that, for the purpose of estimation, the worst–case trajectory is the one that yields identically zero measurements, e.g. $y = 0$.

B. Optimality of equalized filters

Next we use the results above to establish global optimality of the equalized filters when operating on a simplified information scenario. This result will be established by using Theorem 7.1 to compute $r(I)$ and show that $r(I) = \mu_{opt}$. Since the error achieved by the equalized filter satisfies $|e| \leq \mu_{opt}$, optimality follows. In the sequel we consider, for simplicity, the case where z is a scalar. Since as shown in section V-A multiple output filters can be obtained by simply stacking single output ones, the extension of the result to vector outputs is trivial.

Consider again the combination system–filter (4)–(6), illustrated in Figure 2. The simplified information scenario considered here (estimation error bounded by an hyperrectangle centered at $\hat{z}_k$) amounts to replacing the measurement y by the "noisy" measurements of z given by

$$\hat{z}_t = z_t + e_t, \; t = k, k-1, \ldots, k-r, \; |e_t| \leq \mu_{opt} \tag{26}$$

where $\hat{z}_t$ is the output of the filter (6).

Theorem 7.2: For the system (4) with measurements(26), the radius of information $r(I) = \mu_{opt}$

Proof: The first step towards computing $r(I)$ using Theorem 7.1 is to compute the consistency set $\mathcal{T}(0)$. From its definition and equations (4) and (26) it follows that

$$\mathcal{T}(0) \doteq \{z_t \colon \quad \begin{array}{l} z = \frac{M}{d} v \text{ and } 0 = z_t + e_t; t = k, k-1, \ldots, k-r \\ \text{for some sequences } \|v\|_\infty \leq 1, \; \|e\|_\infty \leq \mu_{opt}\} \end{array} \tag{27}$$

Next, note that if $\hat{z}_t = 0, \; t = k, \ldots, k-r$, then, from (6) it follows that $B(\lambda) y_t = 0$. Hence the measurement noise sequence w in (5) satisfies: $B(\lambda) D w + B(\lambda) \hat{y} = 0$, where $\hat{y}$ denotes the output of the system $\frac{N(\lambda)}{d(\lambda)}$ to the input v, when initialized with the same initial conditions that generated the output z. In turn, when combined with (13) and (4), this implies that

$$0 = -a(\lambda) z - C(\lambda) v + B(\lambda) D w \tag{28}$$

Further, note that the sets $\|v\|_\infty \leq 1$ and $\|e\|_\infty \leq \mu_{opt}$ are convex, balanced. Direct application of Theorem 7.1 yields:

$$\begin{array}{rcl} r(I) & = & \sup_{z \in \mathcal{T}(0)} |z_n| \\ & = & sup |z_n| \\ & & \text{subject to (28)}, \; |z_t| \leq \mu_{opt}, \; t = n-1, \ldots, n-r, \; \|v\|_\infty \leq 1, \; \|w\|_\infty \leq 1 \\ & = & \mu_{opt} \end{array} \tag{29}$$

where the last equality follows from Theorem 4.1 and the definition of μ_{opt}. ■

VIII. Connection with existing results

In this section we briefly comment on the connection with existing approaches. The initialization procedure described in section VI is equivalent to using the set valued observers proposed in [14], [15] for the first r steps, switching afterwards to the filter (6). In this sense, the proposed algorithm resembles the approach in [21], where pointwise optimal estimators are used until the ℓ^∞–induced filter reduces the error due to the unknown initial conditions below a given tolerance ϵ, switching then to the latter filter. However, the approach proposed here differs in several aspects, in addition to its ability to fix, a–priori, the complexity of the filter. Specifically, (i) the set–valued filters are used for a fixed horizon (equal to the order r of the equalized filter), as opposed to a problem–dependent horizon, and (ii) the filter (6) is switched on after the estimation error sequence has been driven to a hyperrectangle of size μ, the optimal worst–case estimation error, rather than below the tolerance ϵ (typically $\mu \gg \epsilon$).

IX. Illustrative examples

In this section we illustrate the proposed approach with several examples.

Example 1: Consider the following second order plant:

$$\frac{M(\lambda)}{d(\lambda)} = \frac{\lambda^2}{1} \quad \frac{N(\lambda)}{d(\lambda)} = \frac{(1-0.5\lambda)(1-2\lambda)}{1}$$

and assume that the process and measurement noise satisfy $|v_k| \leq \gamma$ and $|w_k| \leq \beta$, respectively. In the sequel, we analyze optimal filter behavior as a function of these parameters. For $0 < \gamma < 2$ the optimal equalized estimate is $\hat{z} = 0$, (e.g. zero filter), with $\mu_{opt} = \gamma$.

For $\beta = 1$, $\gamma = 2$, the problem admits multiple solutions, amongst them $\frac{b(\lambda)}{a(\lambda)} = 0$ and

$$\frac{a(\lambda)}{b(\lambda)} = \frac{-0.1517 - 0.4870\lambda - 0.1609\lambda^2 - 0.0464\lambda^3}{1.0000 - 0.0656\lambda - 0.0450\lambda^2 - 0.0464\lambda^3}$$

Finally, the case $\beta = 1$, and $\gamma > 2$ seems to yield, independently of γ, the following filter:

$$\frac{a(\lambda)}{b(\lambda)} = \frac{-0.2463 - 0.6158\lambda - 0.2933\lambda^2 - 0.1173\lambda^3}{1.0000 - 0.1173\lambda^3}$$

with poles at 0.4895 and $-0.2448 \pm j0.4239$. The corresponding equalized cost is given by the following piecewise affine function of γ:

$$\mu_{opt}(\gamma) = \begin{cases} \gamma & \text{for } \; 0 < \gamma < 2 \\ 2 + \kappa(\gamma - 2) & \text{for } \; 2 < \gamma \end{cases}$$

with $\kappa \approx 0.28$.

For comparison we considered the $\mathcal{H}_2$–optimal filter whose state–space realization is

$$A_{h_2} = \begin{bmatrix} 0.3700 & 0.0750 \\ -0.2430 & 0.6076 \end{bmatrix} \quad B_{h_2} = \begin{bmatrix} -0.3700 \\ 0.2430 \end{bmatrix}$$

$$C_{h_2} = \begin{bmatrix} 1.1777 & -0.4443 \end{bmatrix} \quad D_{h_2} = [-0.1777] \, .$$

We report in Fig. 4 the frequency response of the equalized and the $\mathcal{H}_2$–optimal filters and in Fig. 5 the simulations in the presence of random piecewise–constant noise. We see that the equalized filter has a slightly sharper cut–off effect. The noise is randomly generated by taking $v = \pm\gamma$

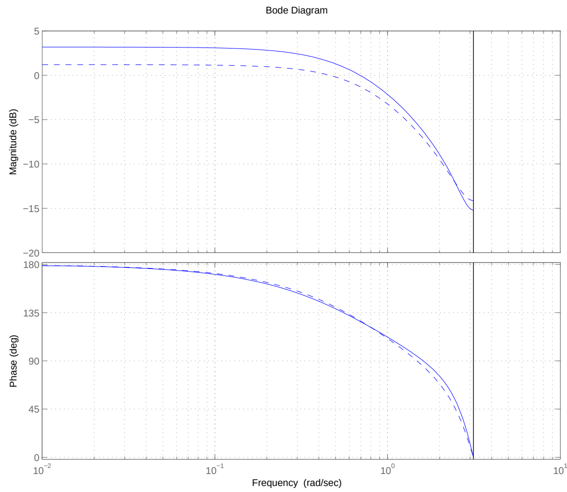

Fig. 4. The equalized filter (plain) versus the $\mathcal{H}_2$ filter (dashed) frequency response

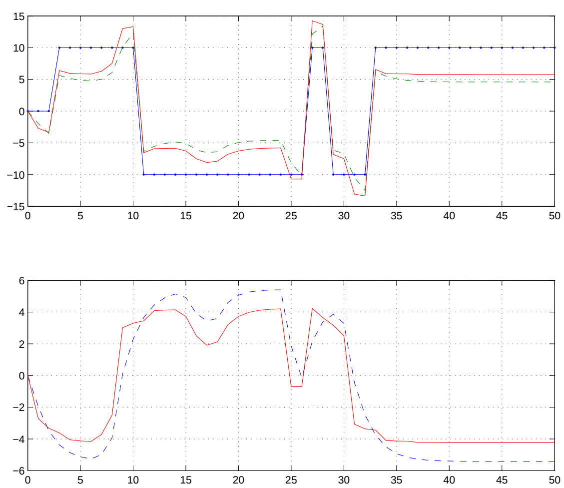

Fig. 5. The equalized filter versus the $\mathcal{H}_2$ filter: Top figure: the reference (dotted), the $\mathcal{H}_2$ filter output (dashed), the equalized output (plain); Bottom figure: the $\mathcal{H}_2$ filter error (dashed), the equalized filter error (plain)

and $w = \pm\beta$ and randomly changing sign with probability $\pi = 1/10$ at each instant. We took $\gamma = 10$ and $\beta = 1$ in this simulation. It is apparent that the error produced by the equalized is always smaller than the $\mathcal{H}_2$ filter. Several experiments with randomly generated sequences show that for this example we have roughly a 30% improvement on the worst case error.

Example 2: Next, we consider the case of a plant with poles on the stability boundary[4]:

$$\frac{M(\lambda)}{d(\lambda)} = \frac{\lambda^2}{1-\lambda^2} \quad \frac{N(\lambda)}{d(\lambda)} = \frac{(1-0.5\lambda)(1-2\lambda)}{1-\lambda^2}$$

In this case, the optimal equalized filter corresponding to $\beta = 1$, $\gamma = 8$ and $r = 3$ is given by:

$$\frac{b(\lambda)}{a(\lambda)} = \frac{-0.3268 - 0.8171\lambda - 0.3891\lambda^2 - 0.1556\lambda^3}{1 - 0.1556\lambda^3}$$

and achieves an equalized performance level $\mu_{opt} = 5.1$

X. Conclusion and discussion

Filtering in the presence of unknown but bounded noise aims at confining the estimation error within a bounded set. Previous work dealing with the problem, based on constructing first the consistency set for the states of the plant (e.g. the set of states compatible with both a-priori assumptions and experimental measurements), led to filters whose complexity can be arbitrarily large, and potentially grows online. Overbounding these sets (using for instance ellipsoids or the approach in [13], [17]), leads to conservative filters with hard to ascertain optimality properties. The receding horizon approach to filtering (see for instance [1]) requires the solution of non–trivial optimization problems online.

To avoid these difficulties, in this paper we propose a different approach, based on the idea of *equalized* performance, first introduced in [4] in the context of suboptimal ℓ^1 controller design. The main idea is to, rather than attempting to find bounded complexity sets that contain the consistency set, work directly with r-length estimation error sequences, confining them to the tightest possible hyperrectangle. As shown in the paper, this can be achieved with an r^{th} order LTI filter, whose coefficients can be found via convex optimization. Further, as opposed to the control case, multiple outputs can be readily handled by simply considering a collection of scalar filters. Intuitively, the approach proposed here can be thought of as trading accurate information about past trajectories against filter complexity: one obtains recursive, bounded complexity filters by only remembering that past estimation errors were confined to an hyperrectangle. As shown in the paper, equalized filters are indeed globally optimal when operating in this simplified scenario.

These results were illustrated with some simple examples. An intriguing fact borne out of these examples is that while in the context of control design the optimal equalized closed loop was almost always "near dead–beat" (e.g. "almost zero" closed–loop poles) the estimation error equation governing the filtering error does not exhibit this feature.

Research is currently underway seeking to extend the results presented in this paper to switched, piecewise linear systems.

References

[1] A. Alessandri, M. Baglietto, and G. Battistelli. "Receding-horizon estimation for discrete-time linear system", IEEE Trans. Automat Control, Vol. 46, no. 3, pp. 473–478, 2003.

[2] N. E. Barabanov and O. N. Granichin, "Optimal controller for a linear plant with bounded noise", Autom. Remote Control, Vol. 45, no, 5. pp. 578–584, 1984.

[3] Bertsekas, D. P. and Rhodes, I. B.,"Recursive state estimation for a set-membership description of uncertainty", IEEE Trans. Automat Control, Vol. 16, pp. 117–128, 1971.

[4] Note that, due to these poles, this case cannot be handled by the approach in [21].

[4] Blanchini, F. and Sznaier, M., "A convex optimization approach to fixed–order controller design for disturbance rejection for SISO systems", IEEE Transactions on Automatic Control, Vol. AC-45, no. 4, pp. 784-789, April 2000.

[5] M. A. Dahleh and I. D. Diaz-Bobillo, "*Control of Uncertain Systems: A Linear Programming Approach*", Prentice-Hall, Englewood Cliffs, New Jersey, USA, 1995.

[6] M. A. Dahleh and J. B. Pearson, "l^1–optimal feedback controllers for MIMO discrete–time system", IEEE Trans. Automat Control, Vol, 32, no. 4, pp. 314–322, 1987.

[7] Luenberger, D. "An introduction to observers," IEEE Trans. Automat. Control, vol. 16, no. 6, pp. 596–602, 1971.

[8] Micchelli, C. A., and Rivlin, T. J., "A Survey of Optimal Recovey," in Optimal Recovery in Approximation Theory, C. A. Micchelli and T. J. Rivlin Eds., Plenum, N.Y., 1977, pp. 1–54.

[9] Milanese, M. and Tempo, R., "Optimal algorithms theory for robust estimation and prediction", IEEE Trans. Automat Control, Vol. 30, no. 5, pp. 730–738, 1985.

[10] Milanese, M. and Vicino, A. "Optimal estimation theory for dynamic systems with set membership uncertainty: An overview," Automatica, Vol. 27, pp. 997–1009, 1991.

[11] Milanese, M. and Belforte G., "Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors: Linear families of models and estimators", IEEE Trans. Automat Control, Vol. 27, no. 2, pp. 408–414, Apr. 1982.

[12] Polyak, B. T. and Halpern, M. E., "The use of a new optimization criterion for discrete-time feedback controller design," Proc. 38^{th} IEEE CDC, Tucson, AZ, Dec. 1999, pp. 894–899.

[13] Schweppe, F. C., "Recursive state estimation: Unknown but bounded errors and system inputs," IEEE Trans. Automat. Contr., Vol. 13, pp. 22–28, 1968.

[14] Shamma, J. S. and Tu, E. "Optimality of Set-Valued Observers for Linear Systems," Proc. 34^{th} IEEE CDC, New Orleans, LA, Dec. 1995, pp. 2081–2086.

[15] J. S. Shamma and K. Y. Tu, "Approximate set-valued observers for nonlinear systems", IEEE Trans. Automat Control, Vol. 42, no. 5, pp. 648–658, 1997.

[16] Shamma, J. and Tu, KY, "Set–valued observer and Optimal Disturbance Rejection", IEEE Trans. Automat. Control, Vol. 44, no.2, pp. 253–164, Feb. 1999.

[17] Stoorvogel, A. A., "ℓ^1 state estimation for linear systems using non-linear observers,", Proc. 35^{th} IEEE CDC, Kobe, Japan, pp. 2407–2411, Dec. 1996.

[18] Tempo, R., "Robust estimation and filtering in the presence of bounded noise," IEEE Trans. Aut. Control, Vol. 33, no. 9, pp. 864–867, 1988.

[19] Tsui, C., "A new design approach to unknown input observers," IEEE Trans. Automat. Control, Vol. 41, pp. 464–468, 1996.

[20] Traub, J. B., Wozniakowski, H., and Wasilkowski, G. W., *Information Based Complexity,* Academic Press, 1988.

[21] Voulgaris, P., "On optimal ℓ^∞ to ℓ^∞ filtering", Automatica, Vol. 31, pp. 489–495, 1995.