Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Direct design of optimal filters from data

C. Novara * **F. Ruiz** * **M. Milanese** *

* *Politecnico di Torino, Italy*
e-mail: carlo.novara@polito.it, fredy.ruiz@polito.it, mario.milanese@polito.it

Abstract: In the literature on filter design, the system whose state has to be estimated is usually assumed known. However, in most practical situations, this assumption does not hold, and a two-step procedure is adopted: 1) a model is identified from a set of noise-corrupted data; 2) on the basis of the identified model, a Kalman filter is designed. In this paper, the idea of directly identifying the filter from data is investigated. In previous works by the authors, it has been shown that the direct identification of the filter may be more convenient than the two-step design. In this paper, an approach for the direct design of optimal filters is proposed, where optimality refers to the minimization of a suitable worst-case estimation error. It is also shown that the Kalman filter is a particular case of the proposed approach.

Keywords: Filter identification; optimality; worst-case estimation; Set Membership identification.

1. INTRODUCTION

Consider the following discrete-time LTI system:

$$\begin{aligned} x^{t+1} &= Ax^t + B_u\widetilde{u}^t + B_\lambda\lambda^t \\ \widetilde{y}^t &= C_y x^t + D_{yu}\widetilde{u}^t + D_{y\lambda}\lambda^t \\ v^t &= C_v x^t + D_{vu}\widetilde{u}^t \end{aligned} \tag{1}$$

where x is the state of the system, $\widetilde{u}$ is a known input, λ is an unknown noise, $\widetilde{y}$ and v are outputs. All the signals are vector-valued except v: $v^t \in \mathbb{R}$. The tilde indicates the variables which are measured.

The aim of filtering is to obtain a (possibly optimal in some sense) estimate $\widehat{v}^t$ of v^t using the measurements $\widetilde{u}^k, \widetilde{y}^k$ for $k \leq t$.

Within the statistical setting, a huge literature exists on minimum variance filter design, assuming that the system (1) is known (see e.g. Gelb (1974); Maybeck (1979)). However, in most practical situations, the system (1) is not known. Then, a two-step procedure is usually adopted:
1) a model of system (1) is identified from the available data $\widetilde{u}^t, \widetilde{y}^t, \widetilde{v}^t$, $t \in [0, T-1]$ (see e.g. Ljung (1999));
2) on the basis of the identified model, a filter is designed which, using as input $(\widetilde{u}^t, \widetilde{y}^t)$, gives an estimate of v^t for $t \geq T$ (see e.g. Gelb (1974)).
Note that except for cases where C_v and D_{vu} are actually known, measurements of v^t have to be performed.

In Milanese et al. (2006) a new approach has been proposed. This approach is based on the direct identification from the available data $\widetilde{u}^t, \widetilde{y}^t, \widetilde{v}^t$, $t \in [0, T-1]$, of a filter which, using as inputs $(\widetilde{u}^t, \widetilde{y}^t)$, gives an estimate of v^t for $t \geq T$. The identified filter is called *direct filter* or *direct virtual sensor* and can be used when the actual sensor is no longer available.

The advantages of the direct approach have been shown in Milanese et al. (2006) within a statistical framework. In that paper, it has been proven that even in the most favorable situations, e.g. no modeling errors and the minimum variance filter is actually computable, the two-step procedure based on Kalman filter design perform no better than the direct approach. More importantly, in the presence of modeling errors, the directly identified filter, although not absolutely optimal, is the minimum variance estimator among the selected approximating filter class. A similar feature is not ensured by the two-step filter, whose performance deterioration caused by modeling errors may be significantly larger.

In the present paper, the direct approach is analyzed within a Set Membership (SM) framework. The noises are assumed unknown but bounded in ℓ_p-norm. A method for the direct design of optimal filters is proposed, where optimality refers to the minimization of a suitable worst-case estimation error, measured by an ℓ_r-norm. The proposed approach is quite general. Indeed, the statistical setting is a particular case of the SM framework considered here: for $p = r = 2$ and for white noises, the direct approach provides the Kalman filter, which is the optimal estimator in the statistical setting.

A simulation example is presented to show the effectiveness of the proposed approach.

2. OPTIMAL FILTERS FOR KNOWN SYSTEM

In this section, we introduce an approach to the filtering problem for the case that the system (1) is known. A wide literature exists on this problem, giving solutions for both

⋆ This work has been partly supported by Ministero dell'Università e della Ricerca of Italy, under the National Projects "Advanced control and identification techniques for innovative applications" and "Control of advanced systems of transmission, suspension, steering and braking for the management of the vehicle dynamics".

the case of statistical noise, see e.g. Gelb (1974); Maybeck (1979), and the case of unknown but bounded noise, see e.g. Nagpal and Khargonekar (1991); Shaked and Theodor (1992). The approach presented here is basic to the direct filter design method presented in the next section, where the system (1) is assumed unknown.

Let us suppose that:

- The matrices A, B_u, B_λ, C_y, D_{yu}, $D_{y\lambda}$, C_v, D_{vu} are known.

- (A, C_y) is observable.

- The noise λ is not known.

- Measurements $\widetilde{u}^t, \widetilde{y}^t$ are available for any time t.

- The output v is not measured.

Let us recall the definition of p-norm:

$$\begin{aligned} \|\lambda\|_p &\doteq \left[\sum_{t=0}^{\tau-1}\sum_{i=1}^{n_\lambda}\left|\lambda_i^t\right|^p\right]^{\frac{1}{p}}, p<\infty \\ \|\lambda\|_\infty &\doteq \max_{t=0,..,\tau-1}\max_{i=1,..,n_\lambda}\left|\lambda_i^t\right| \end{aligned} \tag{2}$$

and of pow-norm:

$$\|\lambda\|_{pow} \doteq \sqrt{\frac{1}{\tau}\sum_{t=0}^{\tau-1}\sum_{i=1}^{n_\lambda}(\lambda_i^t)^2} \tag{3}$$

Note that, for $\tau \to \infty$, these norms become the ℓ_p-norm and the ℓ_{pow}-semi-norm, respectively.

The aim is to find a filter of the form

$$\widehat{v}^t = f\widetilde{\mathbf{w}}^t, \quad t = 0, 1, ..., \tau - 1 \tag{4}$$

$$\begin{aligned} &\widetilde{\mathbf{w}}^t \doteq \left[\widetilde{w}^t; \widetilde{w}^{t-1}; ...; \widetilde{w}^{t-m+1}\right] \in \mathbb{R}^{mn_w\times 1} \\ &\widetilde{w}^t = [\widetilde{y}^t; \widetilde{u}^t] \in \mathbb{R}^{n_w\times 1} \\ &f \doteq [f^0, f^1, ..., f^{m-1}] \in \mathbb{R}^{1\times mn_w}, f^k \in \mathbb{R}^{1\times n_w} \end{aligned}$$

with "small" <u>estimation error</u> $\|v - \widehat{v}\|_r$. Here $r \in \{1, 2, ..., \infty, pow\}$, $v = [v^0; v^1; ...; v^{\tau-1}]$, $\widetilde{v} = [\widetilde{v}^0; \widetilde{v}^1; ...; \widetilde{v}^{\tau-1}]$. The estimation time horizon τ can be either finite or infinite. The notation $[..., ..., ...]$ is used to indicate a row vector, the notation $[...; ...; ...]$ to indicate a column vector.

Since the measurements are noise-corrupted no finite bound on the estimation error can be derived if no assumptions are made on the noise λ. We assume that this noise is bounded as follows.

Assumption on λ: $\|\lambda\|_p \le \delta, \quad p \in \{1, 2, ..., \infty, pow\}$ ■

The estimation error of the filter (4) is given by $\|v - \widehat{v}\|_r = \|v - f * \widetilde{w}\|_r$, where $*$ indicates the convolution product. We are interested in a filter with uniform performances with respect to the sequence $\widetilde{w}$, we thus consider the error $\sup_{\|\widetilde{w}\|_p=1}\|v - f * \widetilde{w}\|_r$. This error is not known, since v depends on λ, which is not known. However, the tightest bound on it is given by the following worst-case error.

Definition 1. Worst-case estimation error of a filter f:

$$EF(f) \doteq \sup_{\|\lambda\|_p\le\delta}\ \sup_{\|\widetilde{w}\|_p=1}\|v - f * \widetilde{w}\|_r$$

■

Looking for filters that minimize this error, leads to the following optimality concepts. Let $\mathcal{F}$ a set of asymptotically stable filters.

Definition 2. A filter $\mathfrak{f}$ is optimal within the filter set $\mathcal{F}$ if:

$$EF(\mathfrak{f}) \doteq \inf_{f\in\mathcal{F}} EF(f)$$

■

We look for optimal filter within the following set of systems:

$$\begin{aligned} \mathcal{K}(m, L, \rho) \doteq \{g = [g^0, g^1, ..., g^{m-1}], g^t \in \mathbb{R}^{1\times n_w} : \\ \left\|g^t\right\|_\infty \le L\rho^t, t = 0, 1, ..., m-1\} \end{aligned}$$

where $\|g^t\|_\infty \doteq \max_{i=1,..,n_w}|g_i^t|$. This is the set of all LTI systems with impulse response of length m and of exponential decay L, ρ. If $m < \infty$, $\mathcal{K}(m, L, \rho)$ is a set Finite Impulse Response (FIR) systems, otherwise it is a set of Infinite Impulse Response (IIR) systems.

In Milanese et al. (2006) it is shown that, if (A, C_y) is observable, then the system (1) can be represented as

$$\begin{aligned} g_{yy} * \widetilde{y} + g_{yu} * \widetilde{u} = g_{y\lambda} * \lambda \\ v = g_{vy} * \widetilde{y} + g_{vu} * \widetilde{u} + g_{v\lambda} * \lambda \end{aligned} \tag{5}$$

where g_{yy}, g_{yu}, $g_{y\lambda}$, g_{vy}, g_{vu}, $g_{v\lambda} \in \mathcal{K}(n_x, L, \rho)$, $n_x \doteq \dim(x^t)$, and $L, \rho < \infty$. In Milanese et al. (2006) it is shown how g_{yy}, ..., $g_{v\lambda}$ can be computed from (1).

Equations (5) show that the relations between the variables of system (1) can be represented by means of "short" (of orders less or equal than the order of system (1)) FIR systems. This representation is used to derive an optimal filter. Consider the following optimization problem:

$$h_o \doteq \arg\min_{h\in\mathcal{K}(m_h, L_h, \rho_h)}\|g_{v\lambda} * (1 - h * g_{y\lambda})\|_{r,p} \tag{6}$$

where $\|\cdot\|_{r,p}$ is the induced norm

$$\|g\|_{r,p} \doteq \sup_{\|u\|_p=1}\|g * u\|_r \tag{7}$$

Note that the optimization problem (6) is <u>convex</u> for any norm $\|\cdot\|_{r,p}$. Indeed, a norm is a convex function of its argument. The argument is a linear function of h. Therefore $\|g_{v\lambda} * (1 - h * g_{y\lambda})\|_{r,p}$ is a convex function of h (see e.g. Boyd and Vandenberghe (2004)). Moreover, the constraint $h \in \mathcal{K}(m_h, L_h, \rho_h)$ can be written as a set of linear inequalities. It follows that the optimization problem (6) is convex.

Let us define the following filter:

$$v_o^t = f_o\widetilde{\mathbf{w}}^t, \quad t = 0, 1, ... \tag{8}$$

$$\begin{aligned} &\widetilde{\mathbf{w}}^t \in \mathbb{R}^{mn_w\times 1},\ f_o \in \mathbb{R}^{1\times mn_w},\ f_o^k \doteq [f_{oy}^k, f_{ou}^k] \in \mathbb{R}^{1\times n_w} \\ &f_{oy} \doteq g_{vy} + g_{v\lambda} * h_o * g_{yy}, \quad f_{ou} \doteq g_{vu} + g_{v\lambda} * h_o * g_{yu} \end{aligned}$$

Theorem 1. The filter f_o is optimal within the set $\mathcal{K}(m, L, \rho)$, where $m = m_h + 2n_x$ and for some $L, \rho < \infty$. The worst case estimation error of f_o is given by:

$$EF(f_o) = \delta\left\|g_o\right\|_{r,p} \tag{9}$$

where $g_o \doteq g_{v\lambda} * (1 - h_o * g_{y\lambda})$.

Proof. First, let us prove that $EF(f_o) = \delta\left\|g_o\right\|_{r,p}$. From (5) we have

$$\begin{aligned} v - f_o * \widetilde{w} &= g_{vy} * \widetilde{y} + g_{vu} * \widetilde{u} + g_{v\lambda} * \lambda \\ -g_{vy} * \widetilde{y} - g_{v\lambda} * h_o * g_{yy} * \widetilde{y} &- g_{vu} * \widetilde{u} - g_{v\lambda} * h_o * g_{yu} * \widetilde{u} \\ &= g_{v\lambda} * \lambda - g_{v\lambda} * h_o * (g_{yy} * \widetilde{y} + g_{yu} * \widetilde{u}) \\ &= g_{v\lambda} * \lambda - g_{v\lambda} * h_o * g_{y\lambda} * \lambda = g_o * \lambda \end{aligned}$$

Then $\|v - f_o * \widetilde{w}\|_r = \|g_o * \lambda\|_r$. It follows that

$$EF(f_o) = \sup_{\|\lambda\|_p \le \delta} \|g_o * \lambda\|_r = \delta \|g_o\|_{r,p}$$

Consider a generic filter f. Hence $v - f * \widetilde{w} = (\widehat{g} - f) * \widetilde{w} + g_{v\lambda} * \lambda$, where $\widehat{g} = [g_{vy}, g_{vu}]$. The worst-case estimation error of f is

$$EF(f) = \sup_{\|\lambda\|_p \le \delta} \sup_{\|\widetilde{w}\|_p = 1} \|(\widehat{g} - f) * \widetilde{w} + g_{v\lambda} * \lambda\|_r$$

Taking a $\widetilde{w}$ such that $(\widehat{g} - f) * \widetilde{w} = 0$, we have $EF(f) \ge \sup_{\|\lambda\|_p \le \delta} \|g_{v\lambda} * \lambda\|_r = \sup_{\|\lambda\|_p \le \delta} \|g_{v\lambda} * (1 - 0 * g_{y\lambda})\|_r$. From (6) we have

$$\begin{aligned} EF(f) &\ge \sup_{\|\lambda\|_p \le \delta} \|g_{v\lambda} * (1 - 0 * g_{y\lambda}) * \lambda\|_r \\ \ge \sup_{\|\lambda\|_p \le \delta} &\|g_{v\lambda} * (1 - h_o * g_{y\lambda}) * \lambda\|_r = \delta \|g_o\|_{r,p} = EF(f_o) \end{aligned}$$

where the last inequality follows from (6). This holds for any filter of the form (4), therefore f_o is an optimal filter. Moreover, it is easy to see from (8) that $m = m_h + 2n_x$. ■

Remarks

1. The constraints on the exponential decay L, ρ are parameters of the filter design. They allow to: 1) guarantee the asymptotic stability of the optimal filter f_o (see the next remark); 2) choose the speed of response of the optimal filter f_o.

2. The asymptotic stability of optimal filter f_o is guaranteed if 1) $m < \infty$ or 2) $m = \infty$, $L < \infty$, $\rho < 1$. Indeed, g_{vy}, $g_{v\lambda}$, g_{yy}, g_{vu}, g_{yu} are all FIR, and thus stable systems. It follows that: in the case 1), f_o is also a FIR, and thus a stable system; in the case 2), since $\rho < 1$, h_o is a stable system, which implies that f_o is a stable system too.

3. If $g_{y\lambda}$ is invertible (as a system), then h_o is an approximation of $g_{y\lambda}^{-1}$ (here $g_{y\lambda}^{-1}$ indicates the inverse of system $g_{y\lambda}$). As $m \to \infty$, we have that $h_o \to g_{y\lambda}^{-1}$, $EF(f_o) \to 0$, $v_o^t \to v^t$.

4. The filter f_o is not polarized: If there is no noise, i.e. $B_\lambda = 0$, $D_{y\lambda} = 0$, then $g_{y\lambda} = 0$, $g_{v\lambda} = 0$, $EF(f_o) = 0$, $v_o = f_o * \widetilde{w} = g_{vy} * \widetilde{y} + g_{vu} * \widetilde{u} = v$.

5. Reduced order filters can be obtained by suitably approximating the filter f_o. This approximation can be performed in the frequency domain, using e.g. $\mathcal{H}_2$ or $\mathcal{H}_\infty$ approximation techniques, or in the time domain, using a set of noise-free input-output data generated by f_o.

6. The present approach, based on the unknown but bounded noise framework, is quite general. Indeed, consider the case of infinite-length signals and systems, and choosing $\rho < 1$:

a. $(r, p) = (2, 2)$: f_o is an $\mathcal{H}_\infty$ filter, see e.g. (Dahleh and Diaz-Bobillo (1995)).

b. $(r, p) = (\infty, 2)$: f_o is an $\mathcal{H}_2$ filter, see e.g. (Dahleh and Diaz-Bobillo (1995)).

c. $(r, p) = (\infty, 2)$ and λ white: f_o is a minimum variance filter, see e.g. (Dahleh and Diaz-Bobillo (1995)). It follows that, for m large enough, f_o is the steady-state Kalman filter of system (1). ■

3. DIRECT DESIGN OF OPTIMAL FILTERS FROM DATA

Let us suppose that:

- The matrices A, B_u, B_λ, C_y, D_{yu}, $D_{y\lambda}$, C_v, D_{vu} are not known and thus the optimal filter f_o in (8) is not known.

- (A, C_y) is observable.

- The noise λ is not known.

- Measurements $\widetilde{u}^t, \widetilde{y}^t$ are available for any time t.

- Noise-corrupted measurements $\widetilde{v}^t$ of v^t are available for $t \in [0, T-1]$.

The problem is to estimate the variable v^t, for $t \ge T$.

In this section, we consider an approach based on the direct identification from the available data $\widetilde{w}^t, \widetilde{v}^t$, $t \in [0, T-1]$, of a filter which, using as inputs $\widetilde{w}^t$, gives an estimate of v^t, for $t \ge T$.

Consider that a set of noise-corrupted measurements $\widetilde{v}^t$ and $\widetilde{w}^t$, $t \in [0, T-1]$ is available. Then:

$$\widetilde{v}^t = f_o \widetilde{w}^t + d^t, \; t = 0, 1, .., T-1 \tag{10}$$

where $d^t = \widetilde{v}^t - v_o^t$. Note that the noise term d^t is composed of two contributions: $d^t = \widetilde{v}^t - v^t + v^t - v_o^t = \xi^t + e_o^t$, where $\xi^t = \widetilde{v}^t - v^t$ is the noise on the measure of v^t, and $e_o^t = v^t - v_o^t$ is the estimation error of the optimal filter f_o.

The aim is to identify a filter of the form:

$$\widehat{v}^t = \widehat{f} \widetilde{\mathbf{w}}^t, \quad t \ge T \tag{11}$$

with "small" estimation error $\|v - \widehat{v}\|_r$, where $r \in \{1, 2, ..., \infty, pow\}$, $v = [v^T; v^{T+1}; ...; v^{T+\tau-1}]$, $\widetilde{v} = [\widetilde{v}^T; \widetilde{v}^{T+1}; ...; \widetilde{v}^{T+\tau-1}]$.

In order to identify such a filter, we look for an optimal approximation of the filter f_o. Then, we show that the optimal approximation provides an optimal estimate.

Let us first consider the problem of identifying a system $\widehat{f}$ which approximates f_o with "small" identification error $\left\| f_o - \widehat{f} \right\|_q$, where $\|\cdot\|_q$ is a q-norm (2).

Whatever identification method is used, no finite bound on this error can be guaranteed, unless some assumptions are made on f_o and on noise d. In this paper, we follow the Set Membership (SM) approach, see e.g. Milanese and Vicino (1991), Milanese et al. (1996), Partington (1997), Chen and Gu (2000), and take the following assumptions.

Assumptions on f_o: $f_o \in \mathcal{K}(m, L, \rho)$

Assumption on d: $\|d\|_p \le \varepsilon$, $p \in \{2, \infty, pow\}$ ■

A key role in the SM framework is played by the Feasible Systems Set, often called "unfalsified systems set", i.e. the set of all systems consistent with prior information and measured data.

Definition 3. The Feasible Systems Set FSS^T is

$$FSS^T \doteq \left\{ f \in \mathcal{K}(m, L, \rho) : \|\widetilde{v} - f * \widetilde{w}\|_p \leq \varepsilon \right\}$$

■

The Feasible Systems Set FSS^T summarizes all the information (measured data and prior information on f_o and noise d) that is available up to time $T-1$. An important property in order to derive optimal estimates is that, if prior assumptions are true, then $f_o \in FSS^T$.

The identification error of a direct filter of the form (11) is given by $\left\| f_o - \widehat{f} \right\|_q$. This error is not known, since f_o is not known. It is only known that $f_o \in FSS^T$. Therefore, the tightest bound on $\left\| f_o - \widehat{f} \right\|_q$ is given by the following worst-case error.

Definition 4. Worst-case identification error of a direct filter $\widehat{f}$:

$$E_q(\widehat{f}) \doteq \sup_{f \in FSS^T} \left\| f - \widehat{f} \right\|_q$$

■

Looking for direct filters that minimize the worst-case error, leads to the following optimality concept.

Definition 5. A direct filter $\widehat{\mathfrak{f}}$ is optimal in identification if

$$E_q(\widehat{\mathfrak{f}}) = \inf_{\widehat{f}} E_q(\widehat{f}) = r_I$$

■

The quantity r_I, called radius of information, gives the minimal identification error that can be guaranteed by any estimate based on the available information up to time $T-1$.

Let us define the direct filter:

$$\widehat{v}_c^t = f_c \widetilde{\mathbf{w}}^t, \quad t \geq T \tag{12}$$

$$f_c \doteq \frac{1}{2}\left(\underline{f} + \overline{f}\right)$$

where the components $\underline{f}_i^k$ and $\overline{f}_i^k$ of $\underline{f}$ and $\overline{f}$ are obtained by means of the following <u>convex</u> optimization problems:

$$\underline{f}_i^k = \min f_i^k, \quad \overline{f}_i^k = \max f_i^i$$

subject to:

$$\|\widetilde{v} - f * \widetilde{w}\|_p \leq \varepsilon$$

$$\left| f_i^k \right| \leq L\rho^k, \quad k = 0, ..., m-1, \quad i = 1, ..., n_w$$

Theorem 2. The direct filter f_c is optimal in identification, for any q-norm $\|\cdot\|_q$. The worst-case identification error of f_c is given by:

$$E_q(f_c) = \frac{1}{2}\left\|\overline{f} - \underline{f}\right\|_q = r_I$$

Proof. See Milanese and Tempo (1985).

■

According to this result, the direct filter f_c is the best approximation of the filter f_o. Moreover, f_o is the filter which, using the knowledge of the system (1), provides the best estimate of the variable v (within the filter class $\mathcal{K}(m, L, \rho)$). We now show that f_c is the filter which, without using the knowledge of the system (1), provides the best estimate of the variable v.

Let us consider the estimation error $\sup_{\|\widetilde{w}\|_p = 1} \left\| v - \widehat{f} * \widetilde{w} \right\|_r$, which can be written as $\sup_{\|\widetilde{w}\|_p = 1} \left\| e_o + f_o * \widetilde{w} - \widehat{f} * \widetilde{w} \right\|_r$, where $e_o = v - f_o \widetilde{w}$. This error is not known, since f_o and e_o are not known. It is only known that $f_o \in FSS^T$ and that e_o is bounded as $\|e_o\|_r \leq \delta_o \doteq \delta \|g_o\|_{r,p}$, see (9). Here, we assume to know this bound.

Assumption on e_o: $\|e_o\|_r \leq \delta_o, \quad r \in \{1, 2, ..., \infty, pow\}$ ■

The tightest bound on $\sup_{\|\widetilde{w}\|_p = 1} \left\| e_o + f * \widetilde{w} - \widehat{f} * \widetilde{w} \right\|_r$ is thus given by the following worst-case error.

Definition 6. Worst-case estimation error of a direct filter $\widehat{f}$:

$$ED_r(\widehat{f}) \doteq \sup_{f \in FSS^T} \sup_{\|e\|_r \leq \delta_o} \sup_{\|\widetilde{w}\|_p = 1} \left\| e + f * \widetilde{w} - \widehat{f} * \widetilde{w} \right\|_r$$

■

Looking for estimates that minimize the worst-case error, leads to the following optimality concept.

Definition 7. A direct filter $\widehat{\mathfrak{f}}$ is optimal in estimation if

$$ED_r(\widehat{\mathfrak{f}}) = \inf_{\widehat{f}} ED_r(\widehat{f})$$

■

Theorem 3. The direct filter f_c is optimal in estimation. The worst-case estimation error of f_c is given by:

$$ED_r(f_c) = \delta_o + r_I$$

Proof. Consider that:

$$\sup_{\|e\|_r \leq \delta_o} \sup_{\|\widetilde{w}\|_p = 1} \left\| e + f * \widetilde{w} - \widehat{f} * \widetilde{w} \right\|_r$$
$$= \sup_{\|e\|_r \leq \delta_o} \sup_{\|\widetilde{w}\|_p = 1} \left(\|e\|_r + \left\| \left(f - \widehat{f}\right) * \widetilde{w} \right\|_r \right)$$
$$= \sup_{\|e\|_r \leq \delta_o} \|e\|_r + \sup_{\|\widetilde{w}\|_p = 1} \left\| \left(f - \widehat{f}\right) * \widetilde{w} \right\|_r = \delta_o + \left\| f - \widehat{f} \right\|_{r,p}$$

where $\|\cdot\|_{r,p}$ is the induced norm (7). Then:

$$ED_r(\widehat{f}) = \sup_{f \in FSS^T} \left(\delta_o + \left\| f - \widehat{f} \right\|_{r,p} \right)$$
$$= \delta_o + \sup_{f \in FSS^T} \left\| f - \widehat{f} \right\|_{r,p}$$

Since f_c is a central estimate of f_o, it follows that

$$\sup_{f \in FSS^T} \left| f\widetilde{\mathbf{w}}^t - f_c \widetilde{\mathbf{w}}^t \right| \leq \sup_{f \in FSS^T} \left| f\widetilde{\mathbf{w}}^t - \widehat{f}\widetilde{\mathbf{w}}^t \right|, \forall \widetilde{\mathbf{w}}^t, \widehat{f}$$

This implies that

$$\sup_{f \in FSS^T} \|(f - f_c) * \widetilde{w}\|_r \leq \sup_{f \in FSS^T} \left\| \left(f - \widehat{f}\right) * \widetilde{w} \right\|_r, \forall \widetilde{w}, \widehat{f}$$

which implies that

$$\sup_{f \in FSS^T} \|f - f_c\|_{r,p} \leq \sup_{f \in FSS^T} \left\| f - \widehat{f} \right\|_{r,p}, \forall \widehat{f}$$

Therefore, we have

$$\inf_{\widehat{f}} ED_r(\widehat{f}) = \delta_o + \inf_{\widehat{f}} \sup_{f \in FSS^T} \left\| f - \widehat{f} \right\|_{r,p}$$
$$= \delta_o + \sup_{f \in FSS^T} \|f - f_c\|_{r,p} = ED_r(f_c) = \delta_o + r_I$$

■

Remarks

1. The filter f_c is not polarized: If $d = 0$, then, $f_c = f_o$, $\widehat{v}_c = v_o$. If also $\lambda = 0$, then $\widehat{v}_c = v_o = v$.

2. Let f_{ls} be the least-squares estimate of f_o. If $p = 2$ and $f_{ls} \in \mathcal{K}(m, L, \rho)$ then $f_c = f_{ls}$, see e.g. Kacewicz et al. (1986). On the other side, if d is white and u is quasi-stationary, then f_{ls} converges to f_o as $T \to \infty$, Ljung (1999). It follows that, under the conditions of Remark 6.c in Section 2, f_c tends to the steady-state Kalman filter of system (1) as $T \to \infty$.

3. Reduced order filters can be obtained by suitably approximating the filter f_c. This approximation can be performed in the frequency domain, using e.g. $\mathcal{H}_2$ or $\mathcal{H}_\infty$ approximation techniques, or in the time domain, using a set of noise-free input-output data generated by f_c. ■

4. EXAMPLE

An example of filter design for an automotive problem is presented. The vertical dynamics of a vehicle with controlled suspensions can be approximated as a fourth order linear system, known as a quarter-car model, formed by the sprung(chassis) mass M_c and the unsprung (wheel) mass M_w, connected by the suspension spring K_c and damper β_c to one another, and to the ground by the tire stiffness K_w. The usual instrumentation for this system is an accelerometer measuring the chassis vertical acceleration and the objective is to recover the differential speed, between chassis and wheel, for control purposes. The system is described by the following set of equations:

$$\begin{aligned} M_c\ddot{x}_c &= \widetilde{u} - K_c\,(x_c - x_w) - \beta_c(\dot{x}_c - \dot{x}_w) \\ M_w\ddot{x}_w &= -\widetilde{u} + K_c\,(x_c - x_w) - K_w\,(x_w - x_r) \\ &\quad +\beta_c(\dot{x}_c - \dot{x}_w) \end{aligned} \tag{13}$$

where $\ddot{x}_c$, $\dot{x}_c$ and x_c are the chassis vertical acceleration, speed and position respectively, $\ddot{x}_w$, $\dot{x}_w$ and x_w are the wheel vertical acceleration, speed and position respectively, x_r is the road profile, and $\widetilde{u}$ is the active suspension force.

A set of 7000 data has been generated from the quarter-car model, with $M_c = 432.8kg$, $M_w = 40kg$, $K_c = 17200N/m$, $K_w = 200000N/m$, $\beta_c = 3000Ns/m$, recorded with a sample time of $T_s = 1/512s$. The system is driven by a "On-Off Sky-Hook" control law as $\widetilde{u}$ and a Pavé road profile as x_r (see e.g. Canale et al. (2006)).

The data-set has been partitioned in two sets:

$$D_m = \{(\widetilde{y}^t, \widetilde{u}^t, \widetilde{v}^t), t = 0, \ldots, T-1\}$$
$$D_s = \{(\widetilde{y}^t, \widetilde{u}^t), t = T, \ldots, N-1\}$$

with $T = 3500$ and $N = 7000$.

In this example, the problem of estimating the variable $v^t = \dot{x}_c(T_s t) - \dot{x}_w(T_s t)$, $t = T$, ..., $N-1$ using the measurements $\widetilde{y}^t = \ddot{x}_c^t(T_s t) + \lambda^t$, $\widetilde{u}^t = \widetilde{u}(T_s t)$, $t = 0$, ..., $N-1$, $\widetilde{v}^t = v^t + \xi^t$, $t = 0$, ..., $T-1$, is considered. The noises λ and ξ are i.i.d. Gaussian noises of zero mean and standard deviation 0.038 and 0.01 respectively. Note that the maximum amplitudes of $\widetilde{y}$ and $\widetilde{v}$ are about 4 and 0.8 respectively, and the signal to noise ratios are 57 and 50 respectively.

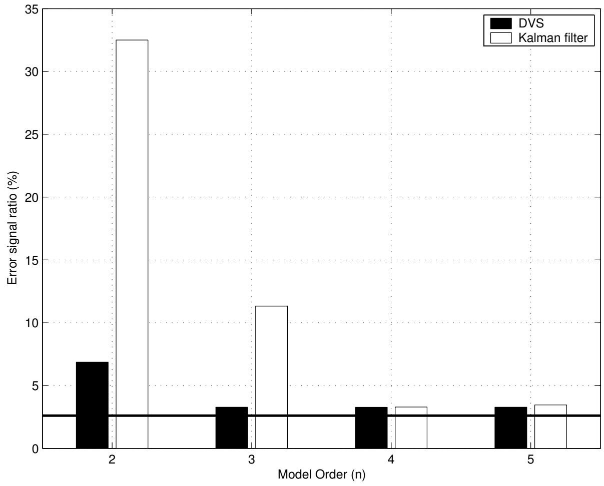

Fig. 1. Automotive example: Error signal radio on set D_s for virtual sensors of different orders.

As standard procedure, LTI models of the form $[\widehat{y}^t, \widehat{v}^t]^\dagger = \hat{M}(u^t, \lambda^t, \xi^t)$ have been identified on the set D_m using the prediction error method and state space model structures of different orders $n_K = 2, \ldots, 5$. For each model $\hat{M}$, a Kalman filter $\widehat{v}_K^t = \hat{K}(\widetilde{y}^t, \widetilde{u}^t)$ has been designed.

In the direct design methodology, the optimal filter $\widehat{v}_c^t = f_c\widetilde{\mathbf{w}}^t$ has been identified on the set D_m assuming $f_o \in \mathcal{K}(148, 0.175, 0.965)$ and $\|d\|_2 \leq 0.08$. Moreover, reduced order filters have been obtained by approximation of f_c in the time domain.

All the filters have been applied to the set D_s. The error-signal ratio obtained on the set D_s by the filters are reported in Figure 1. The horizontal black line corresponds to the performance of the filter f_c, the black columns to the performance of the reduced order direct filters (DVS), the white columns to the performance of the Kalman filters. It can be noted that all the direct filters offer satisfactory performances even in the presence of undermodeling. The two-step Kalman filters turn out to be very sensitive to unmodelled dynamics: when the order of the model set used in the first step is lower than 4 (the order of the true system) the resulting estimators have a poor performance.

In figure 2 the estimates provided by the filter f_c and the reduced order filter of order 2 are compared to the true signal v on a portion of the set D_s. It can be observed that the three lines are nearly overlapping. In figure 3 the estimates provided by the Kalman filters of orders 4 and 2 are compared to the true signal v on a portion of the set D_s. It can be observed that, while the estimate of the 4-th order Kalman filter nearly overlaps the true signal, the estimate of the 2-nd order Kalman filter is quite inaccurate.

5. CONCLUSIONS

A Set Membership filtering approach for linear systems has been proposed. A method for the design of optimal filters from data has been presented and it has been proved that the optimal filter minimizes the worst-case error for any ℓ_r norm used to measure the estimation error. An example of filter design for an automotive problem has been presented

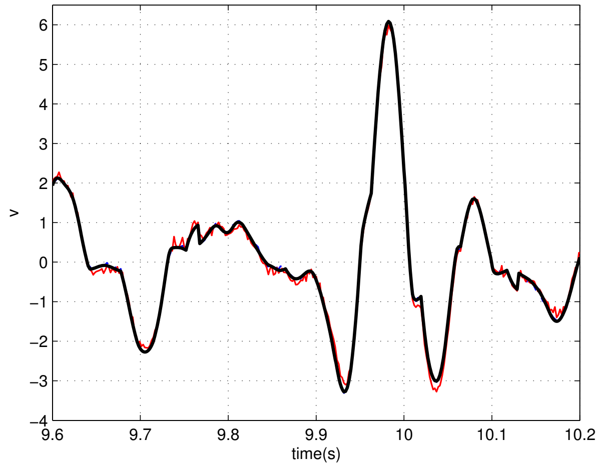

Fig. 2. Direct filters estimation. Bold line (black): true signal. Dashed line (blue): estimate provided by f_c. Thin line (red): estimate provided by the direct filter of order 2.

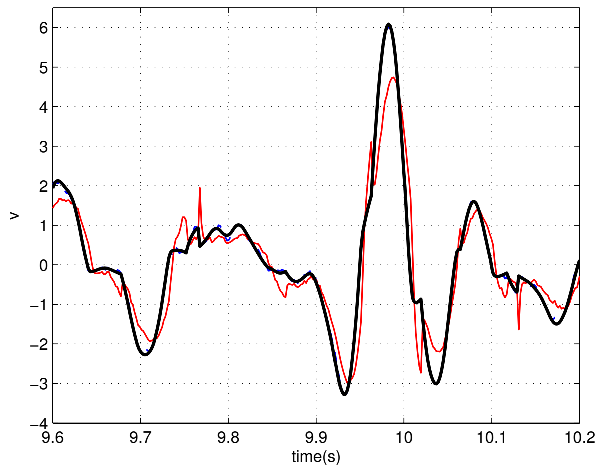

Fig. 3. Kalman filters estimation. Bold line (black): true signal. Dashed line (blue): estimate provided by the filter of order 4. Thin line (red): estimate provided by the filter of order 2.

to demonstrate the capabilities of the proposed approach and its advantages over two-step approaches. Possible extensions of the presented results are the use of of LTI filters to nonlinear systems and the application of the proposed framework to the design of optimal nonlinear filters for nonlinear systems.

REFERENCES

S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

M. Canale, M. Milanese, and C. Novara. Semiactive suspension control using fast model predictive techniques. *IEEE Transactions on Control Systems Technology*, 14 (6):1034–1047, 2006.

J. Chen and G. Gu. *Control-Oriented System Identification: An H_∞ Approach.* John Wiley & Sons, New York, 2000.

M. Dahleh and J. Diaz-Bobillo. *Control of Uncertain Systems, a Linear Programming Approach.* Prentice Hall, Englewood Cliffs, New Jersey, 1995.

A. Gelb. *Applied Optimal Estimation.* The MIT Press., Cambridge, Mass., 1974.

B.Z. Kacewicz, M. Milanese, R. Tempo, and A. Vicino. Optimality of central and projection algorithms for bounded uncertainty. *Systems and Control Letters*, 8: 161–171, 1986.

L. Ljung. *System Identification: Theory for the User.* Prentice Hall, Upper Saddle River, N.J., 1999.

P. S. Maybeck. *Stochastic Models, Estimation, and Control.* Mathematics in Science and Engineering, Academic Press, 1979.

M. Milanese, J. Norton, H Piet Lahanier, and E. Walter. *Bounding Approaches to System Identification.* Plenum Press, 1996.

M. Milanese, C. Novara, K. Hsu, and K. Poolla. Filter design from data: direct vs. two-step approaches. In *American Control Conference*, Minneapolis, Minnesota, USA, 2006.

M. Milanese and R. Tempo. Optimal algorithms theory for robust estimation and prediction. *IEEE Transaction on Automatic Control*, 30:730–738, 1985.

M. Milanese and A. Vicino. Optimal algorithms estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica*, 27:997–1009, 1991.

K. M. Nagpal and P. P. Khargonekar. Filtering and amoothing in an $\mathcal{H}_\infty$ setting. *IEEE Transactions on Automatic Control*, 36:152–166, 1991.

J. R. Partington. *Interpolation, Identification and Sampling*, volume 17. Clarendon Press - Oxford, New York, 1997.

Y. Shaked and Y. Theodor. $\mathcal{H}_\infty$-optimal estimation: A tutorial. In *Proc. of the 31st IEEE Conference on Decision and Control*, Tucson Arizona, USA, 1992.