Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Direct identification of optimal filters for LPV systems

C. Novara, F. Ruiz, M. Milanese

***Abstract*— Direct identification of filters for Linear Parameter Varying (LPV) systems is considered. In the literature on filter design, the system whose state has to be estimated is usually assumed known. However, in most applications, this assumption does not hold, and a two-step procedure is adopted: 1) an LPV model is identified from a set of noise-corrupted data; 2) on the basis of the identified model, an LPV Kalman filter is designed. In this paper, the idea of directly identifying the LPV filter from data is investigated. In previous works by the authors, it has been shown that the direct identification may be more convenient than the two-step design. In some of these works, optimal filter design techniques for time invariant systems have been developed. In the present paper, an approach for the direct identification of optimal filters for LPV systems is proposed. The approach is developed within a Set Membership framework and optimality refers to minimizing the worst-case estimation error.**

I. Introduction

Consider the following discrete time Linear Parameter Varying (LPV) system:

$$\begin{aligned} x^{t+1} &= A\left(\widetilde{p}^t\right) x^t + B_u\left(\widetilde{p}^t\right) \widetilde{u}^t + B_\lambda\left(\widetilde{p}^t\right) \lambda^t \\ \widetilde{y}^t &= C_y\left(\widetilde{p}^t\right) x^t + D_{yu}\left(\widetilde{p}^t\right) \widetilde{u}^t + D_{y\lambda}\left(\widetilde{p}^t\right) \lambda^t \\ v^t &= C_v\left(\widetilde{p}^t\right) x^t + D_{vu}\left(\widetilde{p}^t\right) \widetilde{u}^t \end{aligned} \tag{1}$$

where $x^t \in \mathbb{R}^{n_x}$ is the state of the system, $\widetilde{u}^t \in \mathbb{R}^{n_u}$ is a known input, $\lambda^t \in \mathbb{R}^{n_\lambda}$ is an unknown noise, $\widetilde{y}^t \in \mathbb{R}^{n_y}$ and $v^t \in \mathbb{R}$ are outputs, and $\widetilde{p}^t \in \mathbb{R}^{n_p}$ is a time-varying vector of parameters. The tilde indicates the quantities 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 $k \leq t$.

A huge literature exists on minimum variance filter design, assuming that the system (1) is known (see e.g. [8], [10], [4]). However, in most practical situations, the system (1) is not known, and a two-step procedure is usually adopted:
1) a model of system (1) is identified from the available data $\widetilde{p}^t, \widetilde{u}^t, \widetilde{y}^t, \widetilde{v}^t$, $t \in [0, T-1]$;
2) on the basis of the identified model, an LPV filter is designed which, using as inputs $\widetilde{p}^t, \widetilde{u}^t, \widetilde{y}^t$, gives an estimate of v^t for $t \geq T$ (see e.g. [8]).
Note that except for cases where $C_v\left(\widetilde{p}^t\right)$ and $D_{vu}\left(\widetilde{p}^t\right)$ are actually known, measurements of v^t have to be performed.

In [12] a new approach has been proposed. This approach is based on the direct identification from the available data of a filter which, using as inputs $\widetilde{p}^t, \widetilde{u}^t, \widetilde{y}^t$, 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 [12] within a statistical framework in both the cases of Linear Time Invariant (LTI) and Nonlinear Time Invariant system. 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 [13], a Set Membership (SM) framework has been considered, and an optimal filter identification method for LTI systems has been proposed.

In the present paper, we propose a filter identification method for LPV systems. As in [13] we adopt a SM framework. The noises are assumed unknown but bounded. A method for the direct design of optimal filters is proposed, where optimality refers to the minimization of a suitable worst-case estimation error. The proposed approach is quite general. Indeed, we show that the LPV Kalman filter is obtained as a particular case of the SM framework considered here.

A simulation example regarding the estimation of vehicles yaw rate is presented to show the effectiveness of the proposed approach.

II. 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. The approach presented here is basic to the direct filter identification method presented in the next section, where the system (1) is assumed unknown.

Let us suppose that:
- The matrices $A(\widetilde{p}^t)$, $B_u(\widetilde{p}^t)$, $B_\lambda(\widetilde{p}^t)$, $C_y(\widetilde{p}^t)$, $D_{yu}(\widetilde{p}^t)$, $D_{y\lambda}(\widetilde{p}^t)$, $C_v(\widetilde{p}^t)$, $D_{vu}(\widetilde{p}^t)$ are known for all t.
- $(A^t(\widetilde{p}^t), C_y^t(\widetilde{p}^t))$ is n-step observable (see e.g. [16] for the definition of n-step observability).
- The noise λ^t is not known.
- Measurements $\widetilde{p}^t, \widetilde{u}^t, \widetilde{y}^t$ are available for any time t.
- The output v^t is not measured.

The aim is to find a filter of the form

$$\widehat{v}^t = f\left(\widetilde{\mathbf{p}}^t\right) \widetilde{\mathbf{w}}^t \tag{2}$$

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"

The authors are with Dipartimento di Automatica e Informatica, Politecnico di Torino, Italy. E-mail: carlo.novara@polito.it, fredy.ruiz@polito.it, mario.milanese@polito.it

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

where $\widetilde{\mathbf{w}}^t \doteq \left[\widetilde{w}^t;...;\widetilde{w}^{t-m+1}\right]$, $\widetilde{w}^t = [\widetilde{y}^t;\widetilde{u}^t]$, $\widetilde{\mathbf{p}}^t \doteq \left[\widetilde{p}^t;...;\widetilde{p}^{t-m+1}\right]$, $f\left(\widetilde{\mathbf{p}}^t\right) \doteq [f^0\left(\widetilde{\mathbf{p}}^t\right),...,f^{m-1}\left(\widetilde{\mathbf{p}}^t\right)]$, with "small" estimation error $|v^t - \widehat{v}^t|$.

The notation $[...;...;...]$ indicates vertical concatenation, the notation $[...,...,...]$ indicates horizontal concatenation. The notation $f^t \equiv f\left(\widetilde{\mathbf{p}}^t\right)$ will be used when not necessary to explicit the dependence on the parameters $\widetilde{\mathbf{p}}^t$.

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: $\left\|\boldsymbol{\lambda}^t\right\|_p \leq \delta,\ t = 0,1,...$ where $\boldsymbol{\lambda}^t \doteq \left[\lambda^t;...;\lambda^{t-m+1}\right]$ ∎

Here $\|\cdot\|_p$ is either a p-norm:

$$\|\boldsymbol{\lambda}\|_p \doteq \left[\sum_{t=0}^{m-1}\sum_{i=1}^{n_\lambda}\left|\lambda_i^t\right|^p\right]^{\frac{1}{p}}, 0 < p < \infty,$$

$$\|\boldsymbol{\lambda}\|_\infty \doteq \max_{t=0,..,m-1}\ \max_{i=1,..,n_\lambda}\left|\lambda_i^t\right|$$

or the pow-norm

$$\|\boldsymbol{\lambda}\|_{pow} \doteq \sqrt{\frac{1}{m}\sum_{t=0}^{m-1}\sum_{i=1}^{n_\lambda}\left(\lambda_i^t\right)^2}.$$

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

The estimation error of the filter (2) is given by $|v^t - \widehat{v}^t| = |v^t - f^t\widetilde{\mathbf{w}}^t|$. We are interested in a filter with uniform performances with respect to the input $\widetilde{\mathbf{u}}^t$, we thus consider the error $\sup_{\|\widetilde{\mathbf{u}}^t\|_p=1}|v^t - f^t\widetilde{\mathbf{w}}^t|$. 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:

$$EF(f^t) \doteq \sup_{\|\boldsymbol{\lambda}^t\|_p\leq\delta}\ \sup_{\|\widetilde{\mathbf{u}}^t\|_p=1}\left|v^t - f^t\widetilde{\mathbf{w}}^t\right|.$$

∎

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

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

$$EF(\mathfrak{f}^t) \doteq \inf_{f\in\mathcal{F}} EF(f^t), \quad \forall t.$$

∎

We look for optimal filters $\mathfrak{f}$ within the following set of systems:

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

$\mathcal{K}(m,L,\rho)$ 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 order to derive an optimal filter, we need the following preliminary result. This result uses the definition of n-step observability of [16].

Lemma 1: Let the system (1) be n-step observable. Then the following relations hold:

$$\begin{array}{c} g_y^t\widetilde{\mathbf{y}}^t + g_u^t\widetilde{\mathbf{u}}^t = g_\lambda^t\boldsymbol{\lambda}^t \\ v^t = g_{vy}^t\widetilde{\mathbf{y}}^t + g_{vu}^t\widetilde{\mathbf{u}}^t + g_{v\lambda}^t\boldsymbol{\lambda}^t \end{array} \tag{3}$$

where $\widetilde{\mathbf{y}}^t \in \mathbb{R}^{nn_y\times 1}$, $\widetilde{\mathbf{u}}^t \in \mathbb{R}^{nn_u\times 1}$, $\boldsymbol{\lambda}^t \in \mathbb{R}^{nn_\lambda\times 1}$, $g_y^t \in \mathbb{R}^{n_y\times nn_y}$, $g_u^t \in \mathbb{R}^{n_y\times nn_u}$, $g_\lambda^t \in \mathbb{R}^{n_y\times nn_\lambda}$, $g_{vy}^t \in \mathbb{R}^{1\times nn_y}$, $g_{vu}^t \in \mathbb{R}^{1\times nn_u}$, $g_{v\lambda}^t \in \mathbb{R}^{1\times nn_\lambda}$.
The matrices/vectors g_λ^t are functions of n past values of $\widetilde{p}^t$: $g_*^t \equiv g_*\left(\widetilde{p}^t,...,\widetilde{p}^{t-n+1}\right)$, where $*$ stands for λ, y, u, vy, vu or $v\lambda$.

Proof. Minor modifications of the proof of Lemma 1 in [12]. ∎

The representation (3) is used to derive an optimal filter. Consider the following optimization problem:

$$h_o^t = \arg\min_{h\in\mathcal{K}(\widehat{m},L_h,\rho_h)}\left\|[g_{v\lambda}^t, \mathbf{0}] - hG_\lambda^t\right\|_{ip} \tag{4}$$

where $\widehat{m} = m - n + 1$,

$$G_\lambda^t \doteq \begin{bmatrix} g_\lambda^{t,0} & \cdots & g_\lambda^{t,n-1} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & g_\lambda^{t-1,0} & \cdots & g_\lambda^{t-1,n-1} & \mathbf{0} & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \tag{5}$$

$g_\lambda^t = [g_\lambda^{t,0};...;g_\lambda^{t,n-1}]$, $\mathbf{0}$ indicates a zero matrix of suitable dimension, and $\|\cdot\|_{ip}$ is the induced norm

$$\|g\|_{ip} \doteq \sup_{\|\boldsymbol{\lambda}\|_p=1}|g\boldsymbol{\lambda}|. \tag{6}$$

Note that the optimization problem (4) is convex for any norm $\|\cdot\|_{ip}$. Indeed, a norm is a convex function of its argument. The argument is a linear function of h. Therefore $\|[g_{v\lambda}^t;\mathbf{0}] - hG_\lambda^t\|_{ip}$ is a convex function of h (see e.g. [5]). Moreover, the constraint $h \in \mathcal{K}(\widehat{m},L_h,\rho_h)$ can be written as a set of linear inequalities. It follows that the optimization problem (4) is convex. Note also that $\widehat{m}$, m and n are fixed for all t, then, as new data arrive, the complexity of the optimization problem (4) does not grow.

Let us define the following filter:

$$v_o^t = f_o\left(\widetilde{\mathbf{p}}^t\right)\widetilde{\mathbf{w}}^t = f_{oy}\left(\widetilde{\mathbf{p}}^t\right)\widetilde{\mathbf{y}}^t + f_{ou}\left(\widetilde{\mathbf{p}}^t\right)\widetilde{\mathbf{u}}^t \tag{7}$$

$$\begin{array}{c} f_{oy}\left(\widetilde{\mathbf{p}}^t\right) \doteq \left[g_{vy}^t, \mathbf{0}\right] + h_o^tG_y^t \\ f_{ou}\left(\widetilde{\mathbf{p}}^t\right) \doteq \left[g_{vu}^t, \mathbf{0}\right] + h_o^tG_u^t \end{array}$$

where $\widetilde{\mathbf{w}}^t = \left[\widetilde{w}^t;...;\widetilde{w}^{t-m+1}\right]$, $\widetilde{\mathbf{y}}^t = \left[\widetilde{y}^t;...;\widetilde{y}^{t-m+1}\right]$, $\widetilde{\mathbf{u}}^t = \left[\widetilde{u}^t;...;\widetilde{u}^{t-m+1}\right]$, and G_y^t, G_u^t are defined from g_y^t, g_u^t analogously to G_λ^t in (5).

The main idea behind the derivation of this filter is the following: $h_o^tG_\lambda^t$ is an approximation of $[g_{v\lambda}^t, \mathbf{0}]$, see (4). Therefore, $h_o^tG_\lambda^t[\boldsymbol{\lambda}^t;\mathbf{0}]$ is an estimate of the signal $g_{v\lambda}^t\boldsymbol{\lambda}^t$. This estimate can be obtained from the first of equations (3), and used in the second one to estimate v^t. If some m and some h exist such that $[g_{v\lambda}^t, \mathbf{0}] = hG_\lambda^t$, then the filter (7) provides the exact estimate of v^t. If this is not the case, the length m can be suitably chosen to obtain a satisfactory noise filtering.

Theorem 1: The filter f_o is optimal within the set $\mathbf{K}(m,L,\rho)$ with $L,\rho<\infty$. The worst case estimation error of f_o is given by:

$$EF(f_o(\widetilde{\mathbf{p}}^t)) = \delta \left\| g_o^t \right\|_{ip}, \quad t=0,1,... \tag{8}$$

where $g_o^t \doteq [g_{v\lambda}^t, \mathbf{0}] - h_o^t G_\lambda^t$.

Proof. See [14]. ■

Remarks

1. The filter length m and the constraints on the exponential decay L_h, ρ_h are parameters of the filter design. In particular, L_h, ρ_h allow us to: 1) guarantee the BIBO 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 BIBO stability of optimal filter $f_o(\widetilde{\mathbf{p}}^t)$ is guaranteed if 1) $m<\infty$ or 2) $m=\infty$, $L_h<\infty$, $\rho_h<1$. Indeed, g_{vy}^t, $g_{v\lambda}^t$, g_{yy}^t, g_{vu}^t, g_{yu}^t are all FIR, and thus stable systems. It follows that: in the case 1), $f_o(\widetilde{\mathbf{p}}^t)$ is also a FIR, and thus a stable system; in the case 2), since $\rho_h<1$, h_o^t is a stable system, which implies that $f_o(\widetilde{\mathbf{p}}^t)$ is a stable system too. BIBO stability of $f_o(\widetilde{\mathbf{p}}^t)$ directly follows from BIBO stability of f_o^t, $\forall t$, see e.g. [16].

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

4. The present approach, based on the unknown but bounded noise framework, is quite general.

Indeed, the LPV Kalman Filter (KF) can be recast in this framework. Consider the set of filters $\mathbf{K}(m,L,\rho)$ where $L<\infty$, $\rho<1$, and $m\to\infty$. Clearly, the KF belongs to $\mathbf{K}(m,L,\rho)$, since this set includes all stable Linear Time Varying systems. Moreover, the KF is a minimum variance filter. On the other hand, in the proof of Theorem 1 it is shown that $v^t - v_o^t = g_o^t \boldsymbol{\lambda}^{t,m}$. If $\boldsymbol{\lambda}^{t,m}$ is white, then

$$\mathrm{Var}\left(v^t - v_o^t\right) = \left\| g_o^t \right\|_{i2} \mathrm{Var}\left(\boldsymbol{\lambda}^{t,m}\right)$$

as shown e.g. in [7]. From (4) it follows that f_o is a minimum variance filter too. Therefore, it can be concluded that f_o is a representation of the KF.

Another interesting case is when the induced norm $\|\cdot\|_{i\infty}$ is used. In this case, f_o gives the minimum absolute error $|v^t - v_o^t|$. The filter f_o is thus an optimal ℓ_1 filter, since it minimizes the induced norm from ℓ_∞ to ℓ_∞, see e.g. [7]. ■

III. Direct identification of optimal filters from data

Let us suppose that:

- The matrices $A(\widetilde{p}^t)$, $B_u(\widetilde{p}^t)$, $B_\lambda(\widetilde{p}^t)$, $C_y(\widetilde{p}^t)$, $D_{yu}(\widetilde{p}^t)$, $D_{y\lambda}(\widetilde{p}^t)$, $C_v(\widetilde{p}^t)$, $D_{vu}(\widetilde{p}^t)$ are not known and thus the optimal filter (7) is not known.
- $(A(\widetilde{p}^t), C_y(\widetilde{p}^t))$ is n-step observable.
- The noise λ^t is not known.
- Measurements $\widetilde{p}^t, \widetilde{u}^t, \widetilde{y}^t$ 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 $t\geq T$.

In this section, we consider an approach based on the direct identification from the available data $\widetilde{p}^t, \widetilde{w}^t, \widetilde{v}^t$, $t\in[0,T-1]$, of a filter which, using as inputs $\widetilde{p}^t, \widetilde{u}^t, \widetilde{y}^t$, gives an estimate of v^t, for $t\geq T$. From (7), we have

$$\widetilde{v}^t = f_o(\widetilde{\mathbf{p}}^t)\widetilde{\mathbf{w}}^t + d^t, \; t=0,1,..,T-1 \tag{9}$$

where the term $d^t = \widetilde{v}^t - v_o^t$ accounts for the fact that the data are corrupted by noise. Note that 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 $f_o(\widetilde{p}^t)\widetilde{\mathbf{w}}^t$.

The aim is to identify a filter of the form:

$$\widehat{v}^t = \widehat{F}(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t), \quad t\geq T \tag{10}$$

with "small" <u>estimation error</u> $|v^t - \widehat{v}^t|$.

In order to identify such a filter, we look for an optimal approximation of the filter $F_o(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t) \doteq f_o(\widetilde{\mathbf{p}}^t)\widetilde{\mathbf{w}}^t$. Then, we show that the optimal approximation provides an optimal estimate.

Let us first consider the problem of identifying a function $\widehat{F}$ that approximates F_o with "small" <u>identification error</u> $\left\|F_o - \widehat{F}\right\|_q$, where $\|\cdot\|_q$ is an L_q norm:

$\|f\|_q \doteq \left[\int_{W_r} |f(r)|^q \, dr\right]^{\frac{1}{q}}$, $0<q<\infty$,
$\|f\|_\infty \doteq \text{ess-sup}_{r\in W_r} |f(r)|$,
W_r is a compact convex set.

Note that the filter $F_o(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t)$ is a function of the regressor $\widetilde{\mathbf{w}}^t$ and of the parameters vector $\widetilde{\mathbf{p}}^t$. The dependence on $\widetilde{\mathbf{w}}^t$ is linear, the dependence on $\widetilde{\mathbf{p}}^t$ is nonlinear. Two main approaches can thus be adopted to approximate $F_o(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t)$. The first is to use an approximating function of the general nonlinear form $\widehat{F}(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t)$ without using this structural information. The second is to use an approximating function of the same form as the filter to approximate: $\widehat{F}(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t) = \widehat{f}(\widetilde{\mathbf{p}}^t)\widetilde{\mathbf{w}}^t$, see e.g. [1], [15], [3]. The first approach is simpler, from both the mathematical and the programming points of view, but it does not use the information on the function structure. This may lead to computational complexity problems and local minima problems (in the case of non-convex estimation algorithms). The second one, using the information on the function structure, allows us to reduce such problems, but it is more complicate. In this paper, we introduce some optimality notions and results which hold for filters of the general form (10), then we show how to identify a filter with the same structure of F_o.

Consider then a filter of the form (10). Clearly, the identification error $\left\|F_o - \widehat{F}\right\|_q$ is not known, since the function F_o and the noise d are not known. In order to guarantee a bound on $\left\|F_o - \widehat{F}\right\|_q$ some assumptions on F_o and on noise d have to be made. In this paper, we follow the Nonlinear Set Membership (NSM) "local" approach of [11], and take the following assumptions:

Assumptions on F_o:

$$F_o \in \mathbf{F}(\gamma) \doteq \{F : F = \widehat{F} + F_\Delta, F_\Delta \in \mathcal{F}(\gamma)\}$$
$$\mathcal{F}(\gamma) \doteq \{g \in C^1(W_r), \|g'(\mathbf{p},\mathbf{w})\| \leq \gamma, \forall [\mathbf{p},\mathbf{w}] \in W_r\}.$$

F_Δ is called *residue function* and g' is the gradient of g.

Assumption on d: $\|d\|_\infty \le \varepsilon$, $d \doteq \left[d^0;...;d^{T-1}\right]$. ■

In [11] a method for choosing proper values of γ and δ in such a way that the prior assumptions are validated is given.

A key role in the Set Membership 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 is

$$FSS^T \doteq \left\{F \in \mathbf{F}(\gamma) : \left|\widetilde{v}^t - F(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t)\right| \le \varepsilon, t \in [0, T-1]\right\}.$$

■

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 $F_o \in FSS^T$.

Let us assume 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 $\mathfrak{f}$ is optimal in identification if

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

■

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:

$$\widehat{v}^t = F_c(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t) = \widehat{F}(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t) + f_\Delta(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t), \quad t \ge T \quad (11)$$

$$\begin{aligned}
&f_\Delta(\mathbf{p}, \mathbf{w}) \doteq \tfrac{1}{2}\left[\underline{f}(\mathbf{p}, \mathbf{w}) + \overline{f}(\mathbf{p}, \mathbf{w})\right]\\
&\overline{f}(\mathbf{p}, \mathbf{w}) \doteq \min_{t=0,..,T-1}\left(\overline{h}^t + \gamma \left\|[\mathbf{p}, \mathbf{w}] - [\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t]\right\|\right)\\
&\underline{f}(\mathbf{p}, \mathbf{w}) \doteq \max_{t=0,..,T-1}\left(\underline{h}^t - \gamma \left\|[\mathbf{p}, \mathbf{w}] - [\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t]\right\|\right)\\
&\overline{h}^t \doteq \widetilde{v}^t - \widehat{F}(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t) + \varepsilon, \ \underline{h}^t \doteq \widetilde{v}^t - \widehat{F}(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t) - \varepsilon.
\end{aligned}$$

Theorem 2: The direct filter F_c is optimal in identification, for any L_q norm. 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 [11]. ■

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. We now show that F_c is the filter that, without using the knowledge of the system (1), provides the best estimate of the variable v.

Let us consider the estimation error $\left|v^t - \widehat{F}(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t)\right|$. This error can be written as $\left|e_o^t + F_o(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t) - \widehat{F}(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t)\right|$, where $e_o^t = v^t - F_o(\widetilde{p}^t, \widetilde{\mathbf{w}}^t)$. 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 $\left|e_o^t\right| \le \delta_o^t \doteq \delta \left\|g_o^t\right\|_{ip}$, see (8). Here, we assume to know this bound:

Assumption on e_o: $\left|e_o^t\right| \le \delta_o^t$. ■

The tightest bound on $\left|e_o^t + F_o(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t) - \widehat{F}(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t)\right|$ is thus given by the following worst-case error.

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

$$ED(\widehat{F}, t) \doteq \sup_{F \in FSS^T} \sup_{|e^t| \le \delta_o^t} \left|e^t + F(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t) - \widehat{F}(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t)\right|.$$

■

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

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

$$ED(\mathfrak{f}, t) = \inf_{\widehat{F}} ED(\widehat{F}, t), \quad \forall t \ge T.$$

■

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

$$ED(F_c, t) = \delta_o^t + \frac{1}{2}\left[\overline{f}(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t) - \underline{f}(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t)\right].$$

Proof. See [14]. ■

Up to now, an approximating function of the general nonlinear form $\widehat{F}(\widetilde{p}^t, \widetilde{\mathbf{w}}^t)$ has been considered. No information on the structure of the filter F_o has been used. As mentioned at the beginning of this section, this may lead to computational complexity and local minima problems when estimating $\widehat{F}$. Using an approximating function of the same form as F_o may help to significantly reduce these problems. We here consider the following 4-layers neural network structure:

- layer 1: $\varphi_{1i} = \sigma\left(\beta_{1i}\mathbf{p} + \mu_{1i}\right)$, $i = 1, ..., \alpha$
- layer 2: $\varphi_{2i} = \sum_{j=1}^{\alpha} \beta_{2ij}\varphi_{1j} + \mu_{2i}$, $i = 1, ..., m$
- layer 3: $\varphi_{3i} = \varphi_{2i}\mathbf{w}^i$, $i = 1, ..., m$
- layer 4: $\varphi_4 = \sum_{i=1}^{m} \varphi_{3i}$

where $\beta_1, \mu_1, \beta_2, \mu_2$ are parameters to estimate, α is the number of neurons of the network, and $\sigma(x) = 2/(1 + e^{-2x}) - 1$ is a sigmoidal basis function.

Clearly, this neural network defines a function of the form

$$\varphi_4 = F_{nn}(\mathbf{p}, \mathbf{w}) = f_{nn}(\mathbf{p})\mathbf{w} \quad (12)$$

where $f_{nn}(\mathbf{p}) = [f_{nn}^1(\mathbf{p}), ..., f_{nn}^m(\mathbf{p})]$, $f_{nn}^i(\mathbf{p}) = \varphi_{2i} = \sum_j \beta_{2ij}\sigma\left(\beta_{1j}\mathbf{p} + \mu_{1j}\right) + \mu_{2i}$.

The main motivation for using neural networks is that they are characterized by strong approximation properties in high dimensional spaces [2]. Moreover, there exist very efficient (though not convex) algorithms for the estimation of the parameters $\beta_1, \mu_1, \beta_2, \mu_2$ (see e.g. [9]).

For the solution of the estimation problem considered in this section, we propose the following optimal direct filter:

$$\widehat{v}^t = F_c(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t) = F_a(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t) + f_\Delta(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t), \quad t \ge T \quad (13)$$

where F_a is a neural network of the form (12) identified from the data $\widetilde{p}^t, \widetilde{\mathbf{w}}^t, \widetilde{v}^t$, $t \in [0, T-1]$ and f_Δ is defined in (11) with $\widehat{F} = F_a$.

To simplify the computation in on-line applications, the following filter may also be used:

$$\widehat{v}^t = F_a(\widetilde{\mathbf{p}}^t, \widetilde{\mathbf{w}}^t), \quad t \geq T \tag{14}$$

provided that the correction function f_Δ is small for all $(\widetilde{p}^t, \widetilde{\mathbf{w}}^t) \in W_r$. The neglectability of f_Δ can be checked off-line, using the measured data. If the filters (13) and (14) have comparable accuracy, a confirmation is obtained that F_a is a good approximation of F_o. On the other hand, the filter (13) may give improvements over the filter (14) in the case that: 1) the neural network estimation algorithm got stuck in a local minimum, 2) the number of neurons α has not been properly chosen, 3) the basis function σ is not suitable.

As final remark of this section, we note that the filters considered are Nonlinear Finite Impulse Response (NFIR) systems. A NFIR system is guaranteed to be BIBO stable, but it can be characterized by a large order, and thus by a large number of parameters. If it is of interest to obtain a low order filter, the approach proposed in this section can be applied to identify an optimal direct filter in auto-regressive form

$$\widehat{v}^t = a(\widetilde{\mathbf{p}}^{t-1})\widehat{\mathbf{v}}^{t-1} + b(\widetilde{\mathbf{p}}^t)\widetilde{\mathbf{w}}^t. \tag{15}$$

An optimal filter of this form exists if the NFIR optimal filter (7) admits an LPV state-space realization, see e.g. [16]. The main advantage of the auto-regressive representation is that a significantly lower order (number of parameters) is required. A drawback is that the auto-regressive term $a(\widetilde{p}^t)\widehat{\mathbf{v}}^t$ may give instability problems.

IV. Example

An example of filter design for vehicles yaw rate is presented.

The knowledge of such a variable is used by Vehicle Dynamics Control (VDC) systems to improve stability of the vehicle motion in emergency situations. In order to generate the required control actions, commercially available VDC systems use the values of yaw rate, lateral acceleration and vehicle longitudinal velocity, measured by appropriate sensors. The cost of the yaw rate sensors alone amounts to about 1/3 of the overall cost of the VDC system. Thus, the availability of an accurate yaw rate virtual sensor could allow a significant reduction in the VDC systems production costs and consequently leading to a larger diffusion of active safety systems on commercial cars, even in the segments B and C cars. The problem of estimating the yaw rate of a real car is here approached by means of the direct filter design method presented in Section III.

The lateral dynamics of a vehicle can be described by a fourth order LPV model, known as a single track model, see e.g. [6]. The vehicle is modeled as a rigid body of mass m and moment of inertia J_z around the vertical axis. The state vector is formed by the yaw rate $\dot{\psi}(t)$, the side-slip angle $\beta(t)$, the front axle lateral force $F_f(t)$ and the rear axle lateral force $F_r(t)$. The manipulated input is the steering angle $\delta(t)$ and $d(t)$ is an unknown disturbance torque. The constant parameters involved in the model are the distances between the front and rear axles, a and b, respectively; the cornering stiffness of the axles c_f and c_l, and the tire relaxation lengths l_f and l_r. p is the vehicle speed, and corresponds to the time varying parameter.

It is assumed that an accelerometer provides a measurement of the lateral acceleration of the vehicle a_y, and the objective is to recover the yaw rate for stability control purposes.

The system is described by the following equations set:

$$\begin{array}{rcl} mp(t)\dot{\beta}(t) & = & F_f(t) + F_r(t) - mp(t)\dot{\psi}(t) \\ J_z\ddot{\psi} & = & aF_f(t) - bF_r(t) + d(t) \\ l_f/p(t)\dot{F}_f(t) & = & -c_f\left(\beta(t) + a\dot{\psi}(t)/p(t) + F_f(t) + \delta(t)\right) \\ l_r/p(t)\dot{F}_r(t) & = & -c_r\left(\beta(t) - b\dot{\psi}(t)/p(t) + F_r(t)\right) \\ a_y(t) & = & \left(F_f(t) + F_r(t)\right)/m. \end{array} \tag{16}$$

A set of 2500 data has been generated from a discrete time equivalent of the single track model, using a sample time $T_s = 2ms$, with parameters $m = 432.8\ kg$, $J_z = 2697\ kgm^2$, $a = 1.13\ m$, $b = 1.57\ m$, $l_f = 1\ m$, $l_r = 1\ m$, $c_f = 76000\ Nm/rad$, $c_r = 95000\ Nm/rad$. The system is driven by a random Gaussian steer angle filtered to a maximum band of 25 Hz, $d(t)$ is white noise of standard deviation 10 and $p(t)$ is a sum of sinusoids, with maximum frequency 2.5 Hz, taking values between 20 m/s and 40 m/s.

The data set has been partitioned in two sets:

$$D_m = \{(p^t, \delta^t, a_y^t, \dot{\psi}^t), t = 0, \ldots, T-1\}$$

$$D_s = \{(p^t, \delta^t, a_y^t), t = T, \ldots, N-1\}$$

with $T = 1500$ and $N = 2500$.

In this example, the problem of estimating the variable $v^t = \dot{\psi}(T_s t)$, $t = T$, ..., $N-1$ using the measurements $\widetilde{y}^t = a_y(T_s t)$, $\widetilde{u}^t = \delta(T_s t)$, $\widetilde{p}^t = p(T_s t)$, $t = 0$, ..., $N-1$, $\widetilde{v}^t = \dot{\psi}(T_s t) + \xi^t$, $t = 0$, ..., $T-1$, is considered, where ξ^t is a white noise of standard deviation $1e-3$.

Optimal filters of the form (7) and of lengths $m = [4, 5, \ldots, 50]$ have been designed for a discrete equivalent of model (16), solving the problem (4) for $p = 2$. These filters are called OF.

Optimal direct filters of the form (11) and of lengths $m = [10, 20, \ldots, 50]$, identified using data set D_m, have been designed. Two approaches have been considered.

In the first, the optimal filter F_c has been designed using a function $\widehat{F}$ of the general nonlinear form. A two layer neural network with 5 neurons in the hidden layer and sigmoidal activation function has been taken. The filters obtained by this approach are called UF.

In the second, the optimal filter F_c has been designed using a function $\widehat{F} = F_a$, where F_a is a neural network of the structured form (12), with 3 neurons in the first layer and sigmoidal activation function. The filters obtained by this approach are called SF.

All the filters have been applied to the set D_s. The RMS errors obtained on the set D_s by the OF filters are reported

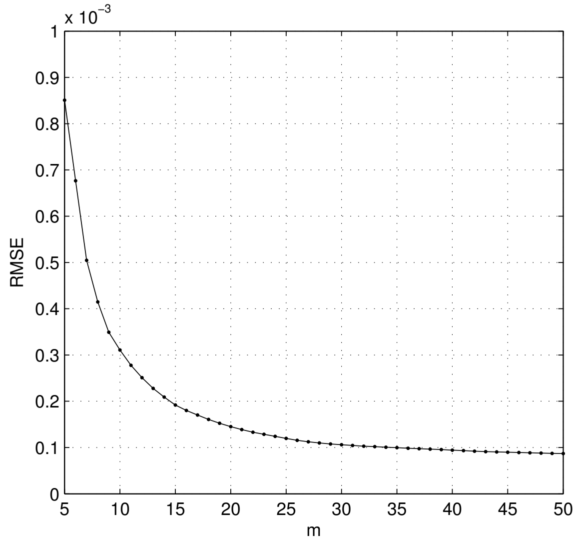

Fig. 1. RMS errors of the optimal filters OF.

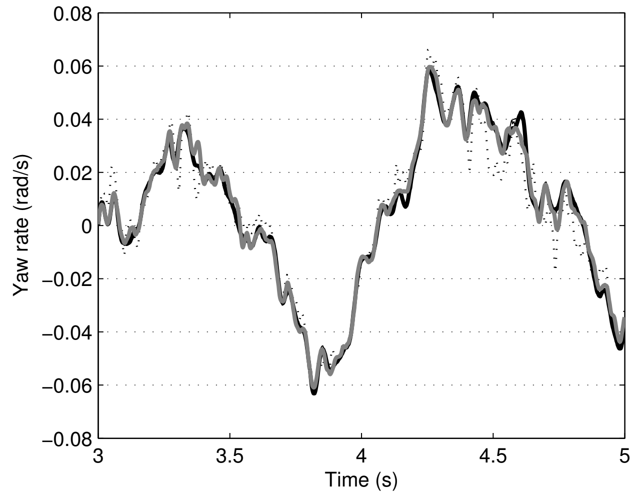

Fig. 2. Model-based and direct filters estimation. Bold: true signal. Thin: estimate provided by the OF filter with $m = 30$. Grey: estimate provided by the SF filter with $m = 30$. Dotted: estimate provided by the UF filter with $m = 30$.

in Figure 1. The RMS errors obtained on the set D_s by the direct filters UF and SF are reported in Table 1. The estimates provided by the filter OF and the direct filter SF for $m = 30$ are compared to the true signal v in figure 2.

From these results, we can see that the estimates provided by the optimal filters OF are very accurate, despite the large amplitude of the noise d. The estimates of the direct filters UF and SF, though presenting deterioration with respect those of the filters OF, are still satisfactory. Clearly, this deterioration was expected, since 1) the direct filters do not use the knowledge of the system, 2) the measurements of v are corrupted by the noise ξ (this noise does not affect the estimates of the filters OF), 3) a relatively small number of data has been used to identify the direct filters.

It can finally be observed that the filters SF are more accurate than the filters UF. This shows that using the information on the LPV structure, as done for the filters of the form (13), allows sensible improvements of the estimates.

m	10	20	30	40	50
UF	0.0103	0.0097	0.0069	0.0046	0.0122
SF	0.0051	0.0032	0.0026	0.0024	0.0041

Table 1: RMS errors of the optimal direct filters.

V. Conclusions

Direct identification of filters for LPV systems has been considered. Within a Set Membership framework, a method for the design of optimal filters from data has been presented. An example of filter design for an automotive problem has been presented to demonstrate the capabilities of the proposed approach.

References

[1] B. Bamieh and L. Giarre. Identification of linear parameter varying models. *International Journal of Robust and Nonlinear Control*, 12/9:841–853, 2002.

[2] A. R. Barron. Universal approximation bounds for superposition of a sigmoidal function. *IEEE Transaction on Information Theory*, 39:930–945, 1993.

[3] G. Belforte and P. Gay. Optimal worst case estimation for LPV-FIR models with bounded errors. *Systems and Control Letters*, 53:259–268, 2004.

[4] F. Blanchini and S. Miani. Stabilization of LPV systems: state feedback, state estimation and duality. *SIAM Journal on Contr. and Opt.*, 32(1):76–97, 2003.

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

[6] M. Canale, L. Fagiano, M. Milanese, and P. Borodani. Robust vehicle yaw control using an active differential and imc techniques. *Control Engineering Practice*, 15/8:923–941, 2007.

[7] M. Dahleh and J. Diaz-Bobillo. *Control of uncertain systems, a linear programming approach*. Prentice Hall, Englewood Cliffs, New Jersey, 1995.

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

[9] J. Hertz, A. Krogh, and G. Palmer. *Introduction to the theory of neural computation*. Addison - Wesley, Santa Fe Institute studies in the sciences of complexity, Reading (Mass.), 1991.

[10] P. S. Maybeck. *Stochastic models, estimation, and control*. Mathematics in Science and Engineering, Academic Press, 1979.

[11] M. Milanese and C. Novara. Set membership identification of nonlinear systems. *Automatica*, 40/6:957–975, 2004.

[12] 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.

[13] C. Novara, F. Ruiz, and M. Milanese. Direct design of optimal filters from data. In *17th IFAC World Congress*, Seoul, Korea, 2008.

[14] C. Novara, M. Canale and M. Milanese, "Direct identification of optimal filters for LPV systems", Politecnico di Torino Internal Report, 2008.

[15] F. Previdi and M. Lovera. Identification of non-linear parametrically varying models using separable least squares. *International Journal of Control*, 77(16):1382 – 1392, 2004.

[16] W. J. Rugh. *Linear System Theory*. Prentice Hall, 1996.