A combined Moving Horizon and Direct Virtual Sensor approach for constrained nonlinear estimation ⋆

Lorenzo Fagiano [a,b], Carlo Novara [a]

[a] *Dipartimento di Automatica e Informatica, Politecnico di Torino, Torino, Italy*

[b] *Department of Mechanical Engineering, University of California at Santa Barbara, Santa Barbara, CA, USA*

Abstract

This paper presents a novel approach to design estimators for nonlinear systems. The approach is based on a combination of linear Moving Horizon Estimation (MHE) and Direct Virtual Sensor (DVS) techniques, and allows the design of estimators with guaranteed stability, which can account for convex constraints on the variables to be estimated. It is also shown that the designed estimators are optimal, in the sense that they give minimal worst-case estimation error, on the basis of the available finite number of noise-corrupted data, with respect to an ideal MHE filter (obtained by assuming exact knowledge of the system dynamics and of the global solution of the related nonlinear program). The approach is tested on a nonlinear mass-spring-damper system.

Key words: Nonlinear estimation; Moving Horizon Estimation; Direct Virtual Sensors; Nonlinear approximation; Set Membership approximation

1 Introduction

In this paper, we study the problem of estimating a variable of interest v^t in a nonlinear discrete-time dynamical system. The variable v^t is assumed to be a nonlinear function of the system state x^t and input u^t, and it can be subject to constraints. Estimation problems for nonlinear systems are in general very difficult [12], [3]. The common approach is to obtain approximate solutions such as extended Kalman filters, [10], [8], [14], unscented Kalman filters, [11], ensemble filters, [7], or particle filters, [21], [6], [13]. However, no optimality properties are usually guaranteed by these approximations, even the stability of the estimation error is often not ensured. One of the few filtering techniques that are able to effectively cope with this kind of issues is Moving Horizon Estimation (MHE) (see e.g. [20], [1]). In MHE, at each time step t, an estimate $\widehat{v}^t$ of v^t is computed, by solving a constrained optimization problem, which involves the simulation of a system model and the optimization with respect to an estimate of the initial state some τ steps in the past. Such an optimization procedure is repeated at each time step, in a moving horizon (or receding horizon) fashion [20]. Interesting features of MHE are the possibility to treat nonlinear models, and to include constraints in the formulation. By suitably designing the cost function of the underlying optimization problem, stability of the estimation error can be also guaranteed [1,20]. However, it has to be noted that MHE relies heavily on the knowledge of a system model, but this model may result to be inaccurate, with consequent degradation of the estimation accuracy. Moreover, when nonlinear models are used, the resulting optimization problem is in general not convex, and finding a global minimum may thus involve a high computational complexity. On the other hand, when used with a linearized model, the MHE optimization problem is convex and it can be efficiently solved, and the resulting filter (named "convex MHE" here) usually gives very good performance when the system state is close to the operating point chosen for linearization.

In this paper, we propose a method that allows one to improve the performance of a convex MHE, when the system operating conditions are different from the ones pertaining to the linearized model embedded in the MHE filter itself. At the same time, the method is able to exploit the good accuracy that the convex MHE achieves when the underlying linearized model is accurate. This new technique can be applied to any convex MHE and it is based on the concept of Direct Virtual Sensor (DVS), i.e. a filtering algorithm derived directly from a finite number of measured data, without using a system model [17], [18]. Such data are assumed to be collected during an initial set of experiments on the system, in which also the variable to be estimated v^t is measured, in addition to the outputs y^t and inputs u^t. It must

⋆ This paper was not presented at any IFAC meeting. Corresponding author L. Fagiano.

Email addresses: lorenzo.fagiano@polito.it (Lorenzo Fagiano), carlo.novara@polito.it (Carlo Novara).

Preprint submitted to Automatica
Received July 23, 2012 13:46:38 PST

be remarked that, except for particular cases (i.e. when the function from x^t and u^t to v^t is exactly known), in most filtering approaches some measurements not only of y^t and u^t, but also of v^t are necessary in order to identify a system model, to be used for filter design.

The resulting filtering algorithm, named Improved MHE (IMHE), is the sum of the convex MHE and of a DVS part, which compensates the mismatch between the MHE and the behavior of the real nonlinear system. The IMHE is stable by construction and it is able to account for bounds on the variable to be estimated. Moreover, we show that, under mild assumptions, the IMHE results to be an optimal approximation of an "ideal" MHE (i.e. a MHE obtained by assuming that an exact model of the system dynamics is available, and that the global optimum of the MHE optimization problem can be computed). In particular, optimality of the IMHE is established in the sense of minimal worst-case estimation error with respect to an ideal MHE. The features of this new technique are illustrated in a simulation example, where the problem of estimating the state of a nonlinear mass-spring-damper system is considered.

2 Problem Formulation

Consider a discrete-time nonlinear system described in state-space form:

$$\begin{aligned} x^{t+1} &= F\left(x^t, \widetilde{u}^t, w^t\right) \\ \widetilde{y}^t &= H_y\left(x^t, \widetilde{u}^t, w^t\right) \\ v^t &= H_v\left(x^t, \widetilde{u}^t, w^t\right) \end{aligned} \tag{1}$$

where $t \in \mathbb{N}$ is the discrete time variable, $x^t \in \mathbb{R}^{n_x}$ is the system state, $\widetilde{u}^t \in \mathbb{R}^{n_u}$ is the measured input, $\widetilde{y}^t \in \mathbb{R}^{n_y}$ is a measured output, $w^t \in \mathbb{R}^{n_w}$ is an unmeasured disturbance, and $v^t \in \mathbb{R}^{n_v}$ is an unmeasured variable to be estimated. Note that in (1) the disturbance w is a vector that includes both process disturbance and measurement noise.

In the following, a sequence of input values starting from time step t_1 up to time step t_2 will be denoted by $\tilde{U}_{t_1}^{t_2} = \{\tilde{u}^t\}_{t=t_1}^{t=t_2}$. Likewise, $\tilde{Y}_{t_1}^{t_2}$ and $W_{t_1}^{t_2}$ denote sequences of outputs and disturbances. The predicted trajectory of the state of system (1) at time step t obtained by starting from the state x^{t-j} at time step $t-j$ and by applying given sequences $\tilde{U}_{t-j}^{t-1}$ and disturbances W_{t-j}^{t-1} is indicated as $x(t, t-j, x^{t-j}, \tilde{U}_{t-j}^{t-1}, W_{t-j}^{t-1})$, while the disturbance-free predicted trajectory (i.e. $W_{t-j}^{t-1} = 0$) is denoted by $x(t, t-j, x^{t-j}, \tilde{U}_{t-j}^{t-1})$. The predicted output at time step t starting from the state x^{t-j} at time step $t-j$ and applying given sequences of inputs $\tilde{U}_{t-j}^{t}$ and W_{t-j}^{t} is denoted by $y(t, t-j, x^{t-j}, \tilde{U}_{t-j}^{t}, W_{t-j}^{t})$, while $y(t, t-j, x^{t-j}, \tilde{U}_{t-j}^{t})$ denotes the disturbance-free predicted output. Similarly, $v(t, t-j, x^{t-j}, \tilde{U}_{t-j}^{t}, W_{t-j}^{t-1})$ and $v(t, t-j, x^{t-j}, \tilde{U}_{t-j}^{t})$ denote the predicted unmeasured variable v in the presence and in the absence of disturbances, respectively.

Let us consider the following assumptions.

Assumption 1 *At any step t the disturbance w^t and the variable v^t are contained inside convex compact sets $\mathcal{W} \in \mathbb{R}^{n_w}$ and $\mathcal{V} \in \mathbb{R}^{n_v}$ respectively. Moreover, the set $\mathcal{V}$ is of the form $\mathcal{V} = \{v \in \mathbb{R}^{n_v} : \underline{v} \preceq v \preceq \overline{v}\}$, where $\overline{v}, \underline{v} \in \mathbb{R}^{n_v}$ and the symbol $\preceq$ denotes element-wise inequalities.* □

The sets $\mathcal{W}$ and $\mathcal{V}$ are usually chosen on the basis of the available physical insight of the system.

Assumption 2 *The system (1) is uniformly observable. That is, for any two state values x_1^t, x_2^t there exist a finite number of time steps N_o and a $\mathcal{K}_\infty$-function ζ such that, for any given sequence of inputs $\overline{\tilde{U}}_t^{t+N_o-1}$, $\zeta(\|x_1^t - x_2^t\|) \leq \sum_{j=0}^{N_o-1} \|y(t+j, t, x_1^t, \overline{\tilde{U}}_t^{t+j}) - y(t+j, t, x_2^t, \overline{\tilde{U}}_t^{t+j})\|$ for some norm $\|\cdot\|$ (see e.g. [20]).* □

We recall that a continuous function $\zeta : \mathbb{R}^+ \to \mathbb{R}^+$ is a $\mathcal{K}_\infty$-function if it is strictly monotone increasing, $\zeta(0) = 0$, $\zeta(z) > 0$ for any $z \neq 0$ and $\lim_{z\to\infty} = \zeta(z) = \infty$.

Within the described framework, the problem we consider can be stated as follows:

Problem 1: Find a function f (named "estimator", "estimation algorithm" or "filter") that computes, at each time step t, an estimate $\widehat{v}^t \approx v^t$ such that $\widehat{v}^t \in \mathcal{V}$, whose estimation error $e^t = v^t - \widehat{v}^t$ is bounded in some norm and possibly minimal with respect to a suitable optimization criterion. □

Obviously, if $v^t = x^t$ the described problem is equivalent to a state estimation problem, but in general one could be interested in estimating also other system variables. Due to the presence of constraints on v, it is not easy to solve **Problem 1** even in the case of a linear system (i.e. with linear functions F, H_y and H_v). Moreover, the presence of nonlinearities further increases the difficulty of **Problem 1**.

3 Moving Horizon Estimation

3.1 Nonlinear and convex Moving Horizon Estimators

Most of the design techniques employed in the literature to address **Problem 1** rely on the knowledge of the system equations (1), of an initial estimate $\overline{x}^{t-\tau}$ of the system state at a suitable time $t-\tau$ and, finally, of given sequences of past measured input and output values, $\tilde{Y}_{t-\tau}^t$ and $\tilde{U}_{t-\tau}^t$ respectively, up to a finite number $\tau+1$ of past time steps. Among such design techniques, Moving Horizon Estimation (MHE) is one of the most promising due to its capability to take into account explicitly system nonlinearities and constraints [1,20,9,3]. In MHE, a cost function of the following form is considered:

$$J(\widehat{x}^{t-\tau}, \tilde{U}_{t-\tau}^t, W_{t-\tau}^t, \tilde{Y}_{t-\tau}^t, \overline{x}^{t-\tau}) = \sum_{j=0}^{\tau} \mathcal{L}(e_y^{t-\tau+j}, w^{t-\tau+j}) + \Phi(\overline{x}^{t-\tau}, \widehat{x}^{t-\tau}), \tag{2}$$

where the output error $e_y^{t-\tau+j}$, $j = 0, \ldots, \tau$ is defined as: $e_y^{t-\tau+j} \doteq \tilde{y}^{t-\tau+j} - y(t-\tau+j, t-\tau, \widehat{x}^{t-\tau}, \tilde{U}_{t-\tau}^{t-\tau+j}, W_{t-\tau}^{t-\tau+j})$.

In (2), the initial state guess $\overline{x}^{t-\tau}$ and the sequences $\tilde{Y}_{t-\tau}^t$, $\tilde{U}_{t-\tau}^t$ of measured outputs and inputs are known parameters in the optimization, while the initial state estimate

Preprint submitted to Automatica
Received July 23, 2012 13:46:38 PST

$\widehat{x}^{t-\tau}$ and the disturbance sequence $W_{t-\tau}^{t}$ are optimization variables. The length $N=\tau+1$ of $\tilde{U}_{t-\tau}^{t}$ and of $\tilde{Y}_{t-\tau}^{t}$ is a design parameter, as well as the stage cost function $\mathcal{L}(\cdot,\cdot)$ and the initial cost function $\Phi(\cdot,\cdot)$. Then, **Problem 1** is cast in a numerical optimization framework:

$$\min_{\widehat{x}^{t-\tau},W_{t-\tau}^{t}} \quad J(\widehat{x}^{t-\tau},\tilde{U}_{t-\tau}^{t},W_{t-\tau}^{t},\tilde{Y}_{t-\tau}^{t},\overline{x}^{t-\tau}) \tag{3a}$$

subject to

$$v(t-\tau+j,t-\tau,\widehat{x}^{t-\tau},\tilde{U}_{t-\tau}^{t-\tau+j},W_{t-\tau}^{t-\tau+j})\in\mathcal{V}, \quad \forall j\in[0,\tau]$$

$$w^{t-\tau+j}\in W,\ \forall j\in[0,\tau]. \tag{3b}$$

If a solution $(\widehat{x}^{t-\tau*},W_{t-\tau}^{t*})$ to (3) is found, the estimate $\widehat{v}^{\text{MHE}}$ is computed as:

$$\widehat{v}^{t,\text{MHE}}=v(t,t-\tau,\widehat{x}^{t-\tau*},\tilde{U}_{t-\tau}^{t},W_{t-\tau}^{t*}). \tag{4}$$

Finally, problem (3) is solved at each time step after having updated the sequences $\tilde{Y}_{t-\tau}^{t}$ and $\tilde{U}_{t-\tau}^{t}$ with new measurements and taking the initial state guess as $\overline{x}^{t-\tau}=x(t-\tau,t-\tau-1,\widehat{x}^{t-\tau-1*},\tilde{U}_{t-\tau-1}^{t-\tau},W_{t-\tau-1}^{t-\tau*})$, according to a moving horizon strategy [20].

The resulting MHE, when the system (1) is nonlinear, is named here "nonlinear MHE". If the dynamical system underlying the optimization problem (3) is linear and the functions $\mathcal{L}(\cdot,\cdot)$ and $\Phi(\cdot,\cdot)$ are chosen to be convex, the optimization problem to be solved results to be convex with respect both to the optimization variables $\widehat{x}^{t-\tau},W_{t-\tau}^{t}$ and to the parameters $\overline{x}^{t-\tau}$, $\tilde{U}_{t-\tau}^{t}$, $\tilde{Y}_{t-\tau}^{t}$. In this case, the resulting MHE is named here "convex MHE".

Although nonlinear MHE is potentially a powerful approach whose use is increasing, some issues are still open. One of these issues is that the nonlinear program (NLP) (3) is in general non-convex. In this case, finding the global minimum of (3) may be extremely hard, whereas local minima of this function, which are easier to be found, may lead to poor estimates and/or "jumps" in the estimated variable between two subsequent time steps.

Another important problem, shared by the MHE approach with all the other model-based design methods (e.g. Extended Kalman Filter) is that the system (1) in most practical situations is not known and a model of it is used instead, identified from a set of measured data $\widetilde{u}^t,\widetilde{y}^t,\widetilde{v}^t$, $t=1,2,\ldots,L$. However, only approximate models can be identified from real noise-corrupted data, and an estimator designed from an approximate model may display very large estimation errors or even instability when applied to the true system [17].

On the other hand, in the case of a convex MHE the optimization problem (3) can be solved efficiently. In many practical cases some information on the dynamics of the system is available, as well as on the values of the related physical parameters, and a linearized system model, around some nominal operating condition $(\overline{x},\,\overline{u},\,\overline{w})$ of interest, can be obtained. In these cases, convex MHEs are able to provide a quite accurate estimate around the operating point considered for linearization, and to account for constraints on the variable to be estimated.

The novel idea of this paper is to exploit the advantages of a convex MHE designed by using a linear system model, thus obtaining a good estimate $\widehat{v}^{t,MHE}\simeq v^t$ in a neighborhood of the system nominal operating condition $(\overline{x},\,\overline{u},\,\overline{w})$, and to correct the MHE estimate outside this neighborhood by means of a so-called Direct Virtual Sensor (DVS) approach (see e.g. [18]), based on Nonlinear Set Membership (NSM) function approximation theory. The resulting estimator is thus able to improve, when necessary, the MHE estimate, and guarantees in any case suitable stability and optimality properties. It has to be noted that the features of (nonlinear) MHE and DVS have been studied in the context of an automotive application in [4], where a simple comparison has been carried out, without exploring the idea of actually combining the two approaches. Indeed, in [4] it was highlighted that a plain DVS approach may need a quite high memory usage and a large amount of data to be collected in initial experiments, in order to achieve good accuracy, while the MHE can be sensitive to model uncertainty and give rise to computational issues when the resulting optimization problem is not convex. The combined technique proposed here aims to overcome these issues, hence obtaining an estimator that needs lower memory usage than a plain DVS, and avoids non-convexity and uncertainty issues that may arise in nonlinear MHE.

Before introducing the new approach, it is now useful to make some more considerations on the structure of a stable MHE, either nonlinear or convex, and on the regularity properties of a convex MHE.

3.2 Structural properties of MHE estimators

Once the design parameters N, $\mathcal{L}(\cdot,\cdot)$ and $\Phi(\cdot,\cdot)$ have been chosen, the MHE Algorithm can be regarded to as a function f_τ^{MHE} whose arguments are the initial state guess $\overline{x}^{t-\tau}$ and the measured sequences $\tilde{Y}_{t-\tau}^{t}$ and $\tilde{U}_{t-\tau}^{t}$:

$$\widehat{v}^{t,\text{MHE}}=f_\tau^{\text{MHE}}(\tilde{Y}_{t-\tau}^{t},\tilde{U}_{t-\tau}^{t},\overline{x}^{t-\tau}). \tag{5}$$

Moreover, from step 2) of the MHE Algorithm, it can be noted that at each time step t the initial state guess $\overline{x}^{t-\tau}$ is a function of the sequences $\tilde{Y}_{t-\tau-1}^{t-1}$ and $\tilde{U}_{t-\tau-1}^{t-1}$ and of the previous initial state guess $\overline{x}^{t-\tau-1}$, i.e.: $\overline{x}^{t-\tau}=g(\tilde{Y}_{t-\tau-1}^{t-1},\tilde{U}_{t-\tau-1}^{t-1},\overline{x}^{t-\tau-1})$.

Then, the estimate (5) can be also expressed as

$$\widehat{v}^{t,\text{MHE}}=f_{\tau+1}^{\text{MHE}}(\tilde{Y}_{t-\tau-1}^{t},\tilde{U}_{t-\tau-1}^{t},\overline{x}^{t-\tau-1}), \tag{6}$$

where $f_{\tau+1}^{\text{MHE}}(\tilde{Y}_{t-\tau-1}^{t},\tilde{U}_{t-\tau-1}^{t},\overline{x}^{t-\tau-1})=f_\tau^{\text{MHE}}(\tilde{Y}_{t-\tau}^{t},\tilde{U}_{t-\tau}^{t},g(\tilde{Y}_{t-\tau-1}^{t-1},\tilde{U}_{t-\tau-1}^{t-1},\overline{x}^{t-\tau-1}))$. Thus, assuming that the MHE algorithm (5) is set up at time step $t_0+\tau$ with an initial state guess $\overline{x}^{t_0}$, the estimate $\widehat{v}^{t,MHE}$ at the generic time t can be expressed as

$$\widehat{v}^{t,\text{MHE}}=f_{t-t_0}^{\text{MHE}}(\tilde{Y}_{t_0}^{t},\tilde{U}_{t_0}^{t},\overline{x}^{t_0}), \tag{7}$$

Preprint submitted to Automatica
Received July 23, 2012 13:46:38 PST

where $f^{\text{MHE}}_{t-t_0}$ is the function given by the recursive application of (6). We now consider the following assumption on the filter $f^{\text{MHE}}_{t-t_0}$:

Assumption 3 *The estimator $f^{MHE}_{t-t_0}$ is asymptotically stable. That is, for any $\overline{x}^{t_0,1}$ and $\overline{x}^{t_0,2}$, it holds that $\|f^{MHE}_{t-t_0}(\tilde{Y}^t_{t_0}, \tilde{U}^t_{t_0}, \overline{x}^{t_0,1}) - f^{MHE}_{t-t_0}(\tilde{Y}^t_{t_0}, \tilde{U}^t_{t_0}, \overline{x}^{t_0,2})\|_2 \leq M\,\rho^{(t-t_0)}\|\overline{x}^{t_0,1} - \overline{x}^{t_0,2}\|_2$, for some $M \in (0,\infty)$ and $\rho \in (0,1)$.* □

Note that this assumption is not restrictive since the MHE estimator stability can be guaranteed by suitably choosing the design parameters N, $\mathcal{L}(\cdot,\cdot)$, $\Phi(\cdot,\cdot)$ (see e.g. [1] and the references therein). If the MHE estimator is asymptotically stable, then for any (arbitrarily small) $\mu > 0$ there exists a sufficiently large number of time steps m such that, for any two initial state guesses $\overline{x}^{t_0}_1$, $\overline{x}^{t_0}_2$, it holds that $\|f^{\text{MHE}}_m(\tilde{Y}^{t_0+m}_{t_0}, \tilde{U}^{t_0+m}_{t_0}, \overline{x}^{t_0}_1) - f^{\text{MHE}}_m(\tilde{Y}^{t_0+m}_{t_0}, \tilde{U}^{t_0+m}_{t_0}, \overline{x}^{t_0}_2)\| \leq \mu$. That is, the effect of the initial condition $\overline{x}^{t_0}$ tends to fade away [20]. Thus, with arbitrarily good precision, after a suitable (possibly large) number $m \geq N+1$ of time steps it can be considered that the estimate $\widehat{v}^{t_0+m}$ depends only on the sequences $\tilde{Y}^{t_0+m}_{t_0}$, $\tilde{U}^{t_0+m}_{t_0}$ and not on the initial state $\overline{x}^{t_0}$. Therefore, in general a stable MHE algorithm can be expressed as a Nonlinear Finite Impulse Response (NFIR) estimator f^{MHE}_o plus a "small" truncation error e^t_{trunc}:

$$f^{\text{MHE}}_{t-t_0}(\tilde{Y}^t_{t_0}, \tilde{U}^t_{t_0}, \overline{x}^{t_0}) = f^{\text{MHE}}_o(\tilde{Y}^t_{t-m}, \tilde{U}^t_{t-m}) + e^t_{trunc}. \quad (8)$$

The estimator f^{MHE}_o (8) can be seen as an "ideal" MHE, i.e. obtained by assuming that an exact model of the system dynamics is available, and that the global optimum of the MHE optimization problem can be computed. Clearly, f^{MHE}_o is not known in general, therefore we aim to derive an approximation of it. This task will be carried out in Section 4.

Furthermore, strictly convex MHEs also enjoy a continuity property. In fact, problem (3) can be regarded to as a parametric optimization problem $\mathcal{P}(s,\theta)$, with optimization variable $s = (\widehat{x}^{t-\tau}, W^t_{t-\tau})$ and parameters $\theta = (\tilde{U}^t_{t-\tau}, \tilde{Y}^t_{t-\tau}, \overline{x}^{t-\tau})$. Now, in the context of convex multi-parametric programming it has been shown that the optimizer $s^*(\theta) = \arg\min \mathcal{P}(s,\theta)$ is a continuous (in general non-smooth) function of θ [2]. Thus, being the system linear, the estimate $\widehat{v}^{t,MHE}$ (7) results to be a continuous function of $\overline{x}^{t-\tau}$, $\tilde{U}^t_{t-\tau}$, $\tilde{Y}^t_{t-\tau}$. The parameter $\overline{x}^{t-\tau}$ is, on its turn, a continuous function of $\overline{x}^{t-\tau-1}$, $\tilde{U}^{t-1}_{t-\tau-1}$, $\tilde{Y}^{t-1}_{t-\tau-1}$ and so on backward in time. Therefore, in the case of convex MHEs, the function $f^{\text{MHE}}_{t-t_0}(\tilde{Y}^t_{t_0}, \tilde{U}^t_{t_0}, \overline{x}^{t_0})$ and the FIR filter $f^{\text{MHE}}_o(\tilde{Y}^t_{t-m}, \tilde{U}^t_{t-m})$ in (8) are continuous with respect to their arguments.

4 Improved MHE via Direct Virtual Sensor techniques

In this Section, the case $v^t \in \mathbb{R}$ is considered for simplicity of notation. Multi-dimensional variables can be treated by deriving an estimator for each component of the vector v^t separately. Suppose that a linear model of the system (1) is available, of the form:

$$\begin{aligned} x^{t+1} &= Ax^t + B_u\widetilde{u}^t + B_w w^t \\ \widetilde{y}^t &= C_{y,x}x^t + D_{y,x}\widetilde{u}^t + D_{y,w}w^t \\ v^t &= C_{v,x}x^t + D_{v,u}\widetilde{u}^t + D_{v,w}w^t \end{aligned} \quad (9)$$

The model (9) can be obtained either by linearization or identified from experimental data. Let $\widehat{f}^{\text{MHE}}(\widetilde{\boldsymbol{\varphi}}^t_{t-m})$, $\widetilde{\boldsymbol{\varphi}}^t_{t-m} \doteq (\tilde{Y}^t_{t-m}, \tilde{U}^t_{t-m})$ be a stable convex MHE designed on the basis of this linear model, and let us define the following *residue function*: $\Delta^{\text{MHE}}(\widetilde{\boldsymbol{\varphi}}^t_{t-m}) \doteq f^{\text{MHE}}_o(\widetilde{\boldsymbol{\varphi}}^t_{t-m}) - \widehat{f}^{\text{MHE}}(\widetilde{\boldsymbol{\varphi}}^t_{t-m})$, where f^{MHE}_o is the unknown nonlinear MHE estimator defined in (8). The approach proposed here is to identify, directly from a set of data generated by the system (1), an approximation $\widehat{\Delta}$ of Δ^{MHE}, and to obtain an Improved Moving Horizon Estimator (IMHE) $\widehat{f}^{\text{MHE}}(\widetilde{\boldsymbol{\varphi}}^t_{t-m}) + \widehat{\Delta}(\widetilde{\boldsymbol{\varphi}}^t_{t-m})$, approximation of $f^{\text{MHE}}_o(\widetilde{\boldsymbol{\varphi}}^t_{t-m})$, giving accurate estimates even when the system (1) is not operating in linearity conditions. The following problem is thus considered:

Problem 2: Suppose that a set of data D of fixed length L has been generated in an initial set of experiments: $D = \{\widetilde{u}^t, \widetilde{y}^t, \widetilde{v}^t, t = 1, 2, \ldots, L\}$, where $\widetilde{v}^t \doteq v^t + \xi^t$ is the measured value of v^t corrupted by the noise ξ^t. Then, find an estimator of the form $\widehat{v}^t = \widehat{f}(\widetilde{\boldsymbol{\varphi}}^t_{t-m}) = \widehat{f}^{\text{MHE}}(\widetilde{\boldsymbol{\varphi}}^t_{t-m}) + \widehat{\Delta}(\widetilde{\boldsymbol{\varphi}}^t_{t-m})$, $t > L$, with estimation error minimal with respect to a suitable criterion. □

The estimator $\widehat{f}$ is selected within the following set of functions:

$$\begin{aligned} \mathcal{F}(\gamma, m) \doteq \{&\widehat{f}^{\text{MHE}} + \Delta : \left|\Delta(\boldsymbol{\varphi}^t_{t-m}) - \Delta(\widehat{\boldsymbol{\varphi}}^t_{t-m})\right| \\ &\leq \gamma \left\|\boldsymbol{\varphi}^t_{t-m} - \widehat{\boldsymbol{\varphi}}^t_{t-m}\right\|_\infty, \forall \boldsymbol{\varphi}^t_{t-m}, \widehat{\boldsymbol{\varphi}}^t_{t-m} \in \Phi\} \end{aligned} \quad (10)$$

where $\|\cdot\|_\infty$ is the ℓ_∞ norm, $\gamma \geq 0$ is the Lipschitz constant, and the regressor domain Φ is a bounded convex subset of $\mathbb{R}^{(m+1)(n_y+n_u)}$. The required estimator $\widehat{f}$ is thus given by the convex MHE $\widehat{f}^{\text{MHE}}$ plus a function which is Lipschitz continuous on Φ. Since the term $\widehat{f}^{\text{MHE}}$ is also continuous, as discussed in Section 3.2, all the functions belonging to the set $\mathcal{F}(\gamma, m)$ result to be continuous, too. The motivation for considering the set (10) is to allow the definition of a reasonable optimality criterion for **Problem 2**. Ideally, one would aim at finding the best possible approximation $\widehat{f}$ of the MHE filter f^{MHE}_o. However, since no information is available on f^{MHE}_o, it is impossible even to compute the approximation error. Yet, by restricting our attention to the set of continuous estimators (10), we can aim at finding the "best" approximation $\widehat{f}$ of an optimal estimator f_o defined as follows:

Definition 1 *(Optimal continuous estimator)*
An estimator $f_o \doteq \arg\min_{f \in \mathcal{F}(\gamma,m)} \left\|f^{MHE}_o - f\right\|_\infty$, where $\|f\|_\infty \doteq \operatorname{ess\,sup}_{\boldsymbol{\varphi} \in \Phi} |f(\boldsymbol{\varphi})|$, is named an optimal continuous

Preprint submitted to Automatica
Received July 23, 2012 13:46:38 PST

estimator, and it gives the best approximation, within the set $\mathcal{F}(\gamma, m)$, of the MHE estimator f_o^{MHE} in (8). □

Note that, if $f_o^{\text{MHE}} \in \mathcal{F}(\gamma, m)$, then it obviously holds that $f_o = f_o^{\text{MHE}}$. Otherwise, f_o is anyway the best approximation of the ideal MHE filter within the function set (10).

Since $f_o \in \mathcal{F}(\gamma, m)$, it has a NFIR structure and it is therefore stable by construction. Then, it can be noted that the estimation error $e_o^t \doteq v^t - f_o\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)$ is bounded as $|e_o^t| \le \delta_o,\ \forall t$, for some $\delta_o < \infty$. Moreover, the measurement noise $\xi^t \doteq \tilde{v}^t - v^t$ has to be also taken into account. To this end, we consider the following assumption:

Assumption 4 *The measurement noise on the variable v^t, $t \in [1, L]$ is bounded as $|\xi^t| \le \delta_\xi,\ \forall t \in [1, L]$.* □

An overall bound on the error between the measured values $\tilde{v}^t$, and the corresponding estimated values $f_o\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)$ can be derived: $\left|\tilde{v}^t - f_o\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)\right| \le |\tilde{v}^t - v^t| + \left|v^t - f_o\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)\right| \le \delta_\xi + \delta_o \doteq \varepsilon,\ \forall t \in [m+1, L]$. This bound, combined with the information that $f_o \in \mathcal{F}(\gamma, m)$ (10), allows us to define the following Feasible Estimator Set:

Definition 2 *(Feasible Estimator Set)*
The Feasible Estimator Set FES is: $FES \doteq \{f \in \mathcal{F}(m,\gamma) : \left|\tilde{v}^t - f\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)\right| \le \varepsilon,\ t \in [m+1, L]\}$. □

Remark 1 *According to this definition, FES is the smallest set guaranteed to contain f_o. Note that the values of the bound ε and of the Lipschitz constant γ can be "optimally" chosen by means of the validation procedure in [15]. The value of δ_o is not required for the design of the optimal estimator presented in the following.* □

For any approximation $\widehat{f} \approx f_o$, the estimation error $v^t - \widehat{f}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)$ is bounded as: $|v^t - \widehat{f}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)|$ $= |v^t - f_o\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) + f_o\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) - \widehat{f}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)|$ $= |e_o^t + f_o\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) - \widehat{f}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)| \le \delta_o + |f_o\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) - \widehat{f}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)|$, where the quantity $\left|f_o\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) - \widehat{f}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)\right|$ is the bias between the estimator $\widehat{f}$ and the optimal continuous estimator f_o. Due to the fact that FES is the tightest set guaranteed to contain f_o (see Remark 1), the tightest worst-case bound on the bias $\left|f_o - \widehat{f}\right|$ is clearly given by $\sup_{f \in FES}\left|f\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) - \widehat{f}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)\right|$, leading to the following definition of worst-case estimation error.

Definition 3 *(Worst-case estimation error)*
The worst-case estimation error ED of a DVS $\widehat{f}$ is:

$$ED\left(\widehat{f}, t\right) \doteq \delta_o + \sup_{f \in FES}\left|f\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) - \widehat{f}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)\right|. \quad \square \tag{11}$$

Looking for an estimator that minimizes this error, leads to the following optimality concept for **Problem 2**:

Definition 4 *(Optimal estimator)*
An estimator f^* is optimal if $ED(f^*, t) = \inf_f ED(f, t)$, $\forall t$. □

Remark 2 *The difference between the optimality concepts in Definitions 1 and 4 lies in the fact that the former is concerned with the minimization of the L_∞ norm of the error in an ideal setting, i.e. if both functions f_o^{MHE} and Δ^{MHE} were available, while the latter is concerned with the minimization of the worst-case L_∞ norm of the error in a real settings, where function f_o^{MHE} is not known and only a finite number of noise-corrupted measurements of the residue Δ^{MHE} are available. Hence, the optimality concept in Definition 1 defines a quantity that can not be computed in practice, while the optimality concept in Definition 4 is actually met by the combined estimator proposed in this paper.* □

Let us now define the estimator $\widehat{v}^t = f_c\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right),\ t > L$, where

$$\begin{aligned} & f_c\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) \\ & \doteq \widehat{f}^{\text{MHE}}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) + \tfrac{1}{2}\left[\overline{\Delta}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) + \underline{\Delta}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)\right] \end{aligned} \tag{12}$$

$$\begin{aligned} \overline{\Delta}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) &\doteq \min\left[\overline{v} - \widehat{f}^{\text{MHE}}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t), \overline{\Lambda}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)\right] \\ \underline{\Delta}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) &\doteq \max\left[\underline{v} - \widehat{f}^{\text{MHE}}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t), \underline{\Lambda}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)\right] \\ \overline{\Lambda}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) &\doteq \min_{k \in [m+1, L]}\left(\overline{h}^k + \gamma\left\|\widetilde{\boldsymbol{\varphi}}_{t-m}^t - \widetilde{\boldsymbol{\varphi}}_{k-m}^k\right\|_\infty\right) \\ \underline{\Lambda}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) &\doteq \max_{k \in [m+1, L]}\left(\underline{h}^k - \gamma\left\|\widetilde{\boldsymbol{\varphi}}_{t-m}^t - \widetilde{\boldsymbol{\varphi}}_{k-m}^k\right\|_\infty\right) \end{aligned} \tag{13}$$

and $\overline{h}^k \doteq \tilde{v}^k - \widehat{f}^{\text{MHE}}(\widetilde{\boldsymbol{\varphi}}_{k-m}^k) + \varepsilon$, $\underline{h}^k \doteq \tilde{v}^k - \widehat{f}^{\text{MHE}}(\widetilde{\boldsymbol{\varphi}}_{k-m}^k) - \varepsilon$.
The evaluation of $\overline{\Delta}(\boldsymbol{\varphi})$ for given $\boldsymbol{\varphi}$ is very simple and can be performed as follows: 1) compute the value of each function $\overline{h}^k + \gamma\left\|\boldsymbol{\varphi} - \widetilde{\boldsymbol{\varphi}}_{k-m}^k\right\|_\infty$, $k = m+1, 2, ..., L$, at the given point $\boldsymbol{\varphi}$; 2) obtain $\overline{\Lambda}(\boldsymbol{\varphi})$ by taking the minimum among these $L - m$ values, 3) take the minimum between the computed value and the quantity $\overline{v} - \widehat{f}^{\text{MHE}}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)$. The evaluation of $\underline{\Delta}(\boldsymbol{\varphi})$ can be performed in a similar way.

The amount of memory required by the estimator f_c grows only linearly with L and m. Moreover, it has been shown in [19] that the accuracy of an asymptotically stable filter does not significantly deteriorate when the regressor $\widetilde{\boldsymbol{\varphi}}_{t-m}^t$ has high dimension.

Due to their NFIR structure, both the contributions $\widehat{f}^{\text{MHE}}$ and $\frac{1}{2}\left[\overline{\Delta} + \underline{\Delta}\right]$ are asymptotically stable by construction. It follows that f_c is also asymptotically stable by construction, for any regressor length $m > 0$.

Theorem 1 *Let Assumptions 1-4 hold. Then:*
i) The filter f_c is optimal.
ii) The following bounds on v^t hold: $\widehat{f}^{MHE}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) + \underline{\Delta}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) - \delta_o \le v^t \le \widehat{f}^{MHE}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) + \overline{\Delta}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) + \delta_o$.

Preprint submitted to Automatica
Received July 23, 2012 13:46:38 PST

iii) The worst-case estimation error of f_c is given by $ED\left(f_c, t\right) = \delta_o + \frac{1}{2}\left[\overline{\Delta}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) - \underline{\Delta}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)\right]$.
iv) The constraints are satisfied: $\underline{v} \leq f_c\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) \leq \overline{v}$.

Proof. Let us define the following functions: $\overline{f}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) \doteq \widehat{f}^{\text{MHE}} + \overline{\Delta}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)$, $\underline{f}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) \doteq \widehat{f}^{\text{MHE}} + \underline{\Delta}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)$. From the Lipschitz continuity property in (10), following the same line of the proof of Theorem 2 in [5], it can be shown that

$$\begin{aligned}\inf_{f\in FES} f(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) &= \underline{f}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) \leq f_o(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) \\ &\leq \overline{f}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) = \sup_{f\in FES} f(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)\end{aligned} \tag{14}$$

for any $\widetilde{\boldsymbol{\varphi}}_{t-m}^t \in \Phi$. That is, $\overline{f}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)$ and $\underline{f}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)$ are the tightest upper and lower bound of $f_o(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)$, for any $\widetilde{\boldsymbol{\varphi}}_{t-m}^t \in \Phi$. Clearly, f_c is the mean of these optimal bounds. It follows that, for any $\widetilde{\boldsymbol{\varphi}}_{t-m}^t \in \Phi$, $f_c(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)$ is the central value of the interval $[\underline{f}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t), \overline{f}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)]$, and thus

$$\begin{aligned}&\sup_{f\in FES}\left|f(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) - f_c(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)\right| \\ &= \inf_{\widehat{f}} \sup_{f\in FES}\left|f(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) - \widehat{f}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)\right| \\ &= \tfrac{1}{2}\left[\overline{f}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) - \underline{f}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)\right],\ \forall \widetilde{\boldsymbol{\varphi}}_{t-m}^t \in \Phi.\end{aligned} \tag{15}$$

From (11) and (15), it follows: $ED(f_c, t) = \delta_o + \sup_{f\in FES}\left|f(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) - f_c(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)\right|$ $= \delta_o + \inf_{\widehat{f}} \sup_{f\in FES}\left|f(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) - \widehat{f}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)\right|$. Since δ_o does not depend on the estimate $\widehat{f}$, we have $ED(f_c, t) = \inf_{\widehat{f}}\left(\delta_o + \sup_{f\in FES}\left|f(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) - \widehat{f}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)\right|\right)$ $= \inf_{\widehat{f}} ED\left(\widehat{f}, t\right)$. which shows that f_c is an optimal DVS, proving claim i).

Claim ii) follows from (14), by considering that $v^t - f_o\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) \doteq e_o^t$ and $|e_o^t| \leq \delta_o$.

Moreover, from (15) it follows also that the worst-case estimation error of f_c is given by $ED(f_c, t) = \delta_o + \sup_{f\in FES}\left|f(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) - f_c(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)\right| = \delta_o + \frac{1}{2}[\overline{f}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) - \underline{f}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)] = \delta_o + \frac{1}{2}\left[\overline{\Delta}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) - \underline{\Delta}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)\right]$, $\forall \widetilde{\boldsymbol{\varphi}}_{t-m}^t \in \Phi$, which proves claim iii).

Finally, claim iv) holds since by construction (13) we have $\underline{v} - \widehat{f}^{\text{MHE}}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) \leq \frac{1}{2}\left[\overline{\Delta}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) + \underline{\Delta}\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right)\right], \leq \overline{v} - \widehat{f}^{\text{MHE}}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)$, hence $\underline{v} \leq f_c\left(\widetilde{\boldsymbol{\varphi}}_{t-m}^t\right) \leq \overline{v}$. □

Remark 3 *In many practical situations, the set of data D is fixed and the IMHE estimator has to be designed from this set. In other situations, experiments on the system whose variables have to be estimated can be performed. In these situations, several experiments should be carried out if possible, considering different initial conditions and different input sequences, in order to obtain a sufficiently informative data set D (a method for assessing the degree of information of a given data set can be found in [16], which allows us to evaluate the impact of the filer design data on the estimation quality).* □

Remark 4 *Suppose that* $\frac{1}{2}[\overline{\Delta}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t) - \underline{\Delta}(\widetilde{\boldsymbol{\varphi}}_{t-m}^t)] \ll \delta_o$ *for all t, see claim iii) of Theorem 1. This condition is met if a sufficiently informative data set D is available. In this case, we have that* $ED\left(f_c, t\right) \cong \delta_o$, *and* $f_c \cong f_o$, *i.e. the IMHE estimator is close to an optimal continuous estimator. If in addition* $f_o^{MHE} \in \mathcal{F}(\gamma, m)$, *then* $f_c \cong f_o^{MHE}$, *i.e. the IMHE estimator is close to an ideal MHE.* □

The Improved Moving Horizon Estimation Algorithm is the following:

IMHE Algorithm

(1) At time step t, update the sequences $\tilde{Y}_{t-\tau}^t$ and $\tilde{U}_{t-\tau}^t$ with the measured variables $\tilde{y}^t$, $\tilde{u}^t$;
(2) compute the convex MHE estimate by using (4) with the linear model (9);
(3) update the value of $\widetilde{\boldsymbol{\varphi}}_{t-m}^t$ and compute the IMHE estimate according to (12);
(4) repeat the procedure from step (1) by setting $t = t+1$.

5 Simulation example

Consider the equations of motion of the nonlinear mass-spring-damper system:

$$\begin{aligned}\dot{x}_1(t) &= x_2(t) \\ \dot{x}_2(t) &= -\beta(x_1(t))x_2(t) + \kappa(x_1(t))x_1(t) + u(t) \\ y(t) &= x_1(t) + w(t)\end{aligned} \tag{16}$$

where $u(t)$ is the input force in N, $w(t)$, $|w(t)| \leq 0.025$ is a uniformly distributed disturbance acting on the output, $x_1(t)$ is the mass position in m, $x_2(t)$ is the speed in m/s, and$\kappa(x_1) = a_0\,\exp(a_1\,x_1) + a_2$, $\beta(x_1) = a_0\,\exp(a_3\,x_1) + a_2$. The parameter values are $a_0 = 0.7$, $a_1 = -1$, $a_2 = 0.3$, $a_3 = -2$. Moreover, the speed $x_2(t)$ is mechanically saturated between ± 1 m/s:

$$x_2(t) \in [-1, 1],\ \forall t. \tag{17}$$

The origin is a globally asymptotically stable fixed point and the system is input-to-state stable, so that experiments can be carried out in open-loop.
A first experiment has been performed, assuming that all the states can be measured, to identify a second-order, discrete time LTI model of the system (16). This model is of the form (9), with sampling time $t_s = 0.05$ s. A uniformly distributed random input $u(t)$ with amplitude 0.4 N, plus a sequence of zero-mean square wave signals and sinusoids of increasing amplitudes, from 0.4 N to 0.6 N, has been injected for

Preprint submitted to Automatica
Received July 23, 2012 13:46:38 PST

1000 s to the system. A uniform random noise with amplitude 0.025 has been added to all of the measured quantities, i.e. $\tilde{x}^t$, $\tilde{u}^t$, $\tilde{y}^t$. The model matrices A, B_u, $C_{y,x}$ and $D_{y,w}$, identified via least squares, are $A = \begin{bmatrix} 0.988 & 0.043 \\ 0.927 & -0.48 \end{bmatrix}$, $B_u = \begin{bmatrix} 0.009 \\ 0.053 \end{bmatrix}$, $C_{y,x} = \begin{bmatrix} 1 & 0 \end{bmatrix}$, $D_{y,w} = 1$, while the other matrices in (9) are equal to zero. It can be noted that the numerical values of the system's matrices are close to the linearized and discretized equations of the system (16) at $x \simeq 0$. Therefore, a good accuracy of the MHE filter in a neighborhood of this operating condition is expected. The cost function used in the MHE has been designed according to (2), with $N = 3$ and

$$\begin{aligned} \mathcal{L}(e_y, w) &= Qe_y^2 + Rw^2 \\ \Phi(\widehat{x}, \overline{x}) &= (\widehat{x} - \overline{x})^T Q_x (\widehat{x} - \overline{x}), \end{aligned} \tag{18}$$

where

$$Q = 10,\ R = 1,\ Q_x = \begin{bmatrix} 1 & 0 \\ 0 & 0.01 \end{bmatrix} \tag{19}$$

and $(\widehat{x} - \overline{x})^T$ indicates the transpose of $(\widehat{x} - \overline{x})$. Moreover, the constraint on $x_2(t)$ has been included in the optimization problem (3), which results to be a quadratic program.
A second 1000-s-long experiment has been carried out to collect the data $\tilde{x}^t$, $\widehat{x}^{t,\text{MHE}}$, and the related values of the regressor $\widetilde{\boldsymbol{\varphi}}_{t-m}^t$, with $m = 6$. The values of N, Q, R, Q_x, m, have been tuned by trial and error procedures in order to achieve good performance of the MHE filter and of the related DVS correction. For each state x_1^t and x_2^t, an IMHE estimator of the form (12) has been designed. The related parameters γ_1, ε_1 and γ_2, ε_2 have been estimated according to the guidelines given in [15]. In particular, the values $\gamma_1 = 10^{-7}, \varepsilon_1 = 0.07, \gamma_2 = 4.37$ and $\varepsilon_2 = 0.07$ have been chosen. Moreover, the regressor $\widetilde{\boldsymbol{\varphi}}_{t-m}^t$ has been scaled in order to adapt to the properties of the collected data (see [15] for more details). Note that the quite low value of γ_1 indicates that the estimation errors on this variables have low variability with respect to $\widetilde{\boldsymbol{\varphi}}_{t-m}^t$. This is reasonable, since the first variable is directly measured and the related estimation error is practically negligible and caused by the measurement noise only. On the other hand, a higher estimation error occurs for the second state variable, so that the DVS technique presented in this paper can actually provide a significant improvement.
The designed MHE estimator and its improved version, IMHE, have been tested in a third experiment, performed by injecting square waves of varying amplitudes to the system. Also in this experiment, all the measured quantities have been corrupted by a uniform random measurement noise with amplitude 0.025. Note that the considered noise amplitude is quite large, corresponding to a noise-to-signal ratio of about 10% in average.

A first square wave with amplitude equal to 0.5 N has been used to test the estimators in linear system operating conditions. Other square waves, with growing amplitudes up to 2.5 N, have been used to test the filters when nonlinearities are gradually predominant. The results, in terms of bias and Root Mean Square Error (RMSE, i.e. $e_{1,2}^{\text{RMSE}} \doteq \sqrt{\left(\sum_{t=0}^{N_s - 1} (x_{1,2}^t - \hat{x}_{1,2}^t)^2 \right) / N_s}$ where N_s is the number of simulated time steps), are reported in Tables 1 and 2, respectively, where they are also compared to the performance given by an ideal nonlinear MHE filter (NMHE), obtained by using the exact knowledge of the system's dynamics and by solving the resulting NLP (3) with $\tau = 10$ and functions $\mathcal{L}$ and Φ as in (18)-(19). These design parameters have been chosen through a trial and error procedure.

Table 1
Simulation example. Bias of the MHE, IMHE and NMHE filters with different input square waves.

Input ampl. (N)	0.5	1	2	2.5
MHE, x_1^t (10^{-3}m)	1.0	7.0	32	40
IMHE, x_1^t (10^{-3}m)	-1.7	0.6	0.8	10
NMHE, x_1^t (10^{-3}m)	0.14	0.16	-0.14	-2.3
MHE, x_2^t (10^{-3}m/s)	16	90	450	550
IMHE, x_2^t (10^{-3}m/s)	-3.8	2.0	-16	-3.4
NMHE, x_2^t (10^{-3}m/s)	1.6	3.2	-4.8	-4.0

Table 2
Simulation example. RMSE of the MHE, IMHE and NMHE filters with different input square waves.

Input ampl. (N)	0.5	1	2	2.5
MHE, x_1^t (10^{-3}m)	8.0	11	41	50
IMHE, x_1^t (10^{-3}m)	8.0	8.5	9.9	11.9
NMHE, x_1^t (10^{-3}m)	8.1	8.2	12.2	18.3
MHE, x_2^t (10^{-3}m/s)	24	100	540	660
IMHE, x_2^t (10^{-3}m/s)	15	20.1	91	59
NMHE, x_2^t (10^{-3}m/s)	24.6	24.9	24.7	26.2

The estimate of the first state variable is of little interest, since it is directly measured and all the estimators achieve very good accuracy. As regards the second state variable, it can be noted that the MHE and IMHE filters give quite similar results in linear operating conditions (i.e. for "small" input amplitudes), as expected, while in nonlinear conditions, the IMHE is able to achieve a significant improvement with respect to the MHE, and its performance are indeed quite close to those of the NMHE, which exploits full knowledge of the nonlinear dynamics. As an example, the time course of x_2^t and of its estimates $\widehat{x}_2^{t,\text{MHE}}$, $\widehat{x}_2^{t,\text{IMHE}}$ and $\widehat{x}_2^{t,\text{NMHE}}$ provided by the MHE, IMHE and NMHE filters, respectively, are shown in Fig. 1 for a input square wave with large amplitude, such that the nonlinear dynamics are evidenced. The quite poor performance of the MHE is evident, as well as the good behavior of the IMHE and NMHE. Finally, it can be also noted that the IMHE filter is able to correctly handle the constraint (17).

Preprint submitted to Automatica
Received July 23, 2012 13:46:38 PST

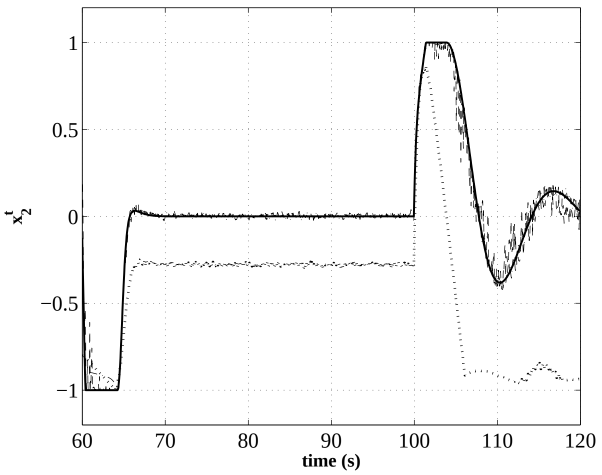

Fig. 1. Simulation example. Courses of the speed x_2^t (solid line) and of its estimates $\widehat{x}_2^{t,\mathrm{MHE}}$ (dotted line), $\widehat{x}_2^{t,\mathrm{IMHE}}$ (dashed line), $\widehat{x}_2^{t,\mathrm{NMHE}}$ (dash-dot line) with a square wave input with amplitude equal to 2.5.

Acknowledgements

This research has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n. PIOF-GA-2009-252284 - Marie Curie project "Innovative Control, Identification and Estimation Methodologies for Sustainable Energy Technologies."

References

[1] A. Alessandri, M. Baglietto, and G. Battistelli. Moving-horizon state estimation for nonlinear discrete-time systems: New stability results and approximation schemes. *Automatica*, 44:1753–1765, 2008.

[2] M. Baes, M. Diehl, and I. Necoara. Every continuous nonlinear control system can be obtained by parametric convex programming. *IEEE Transactions on Automatic Control*, 53(8):1963–1967, 2008.

[3] G. Besançon. *Nonlinear observers and applications*. Lecture notes in control and information sciences. Springer, 2007.

[4] M. Canale, L. Fagiano, and C. Novara. A direct moving horizon approach to vehicle side-slip angle estimation. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 2898 –2903, dec. 2010.

[5] M. Canale, C. Novara, and M. Milanese. NSM constrained approximation with application to fast predictive control. In *46th IEEE Conference on Decision and Control*, New Orleans, USA, 2007.

[6] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. SpringerVerlag, New York, 2001.

[7] G. Evensen. *Data assimilation: The ensamble Kalman filter*. Springer, New York, 2007.

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

[9] E. L. Haseltine and J. B. Rawlings. Critical evaluation of extended kalman filtering and moving-horizon estimation. *Industrial & Engineering Chemistry Research*, 44(8):2451–2460, 2005.

[10] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. New York: Academic, 1970.

[11] S. Julier, J. Uhlmann, and H. Durrant-Whyte. A new method for the nonlinear transformation of mean and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, 45:477–482, 2002.

[12] S.J. Julier and J.K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401 – 422, mar 2004.

[13] G. Kitagawa. Monte Carlo filter and smoother for non-gaussian nonlinear state space models. *J. Computational and Graphical Statistics*, 5(1):1–25, 1996.

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

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

[16] M. Milanese and C. Novara. Computation of local radius of information in SM-IBC identification of nonlinear systems. *Journal of Complexity*, 23:937 – 951, 2007.

[17] C. Novara, M. Milanese, E. Bitar, and K. Poolla. The filter design from data (FD2) problem: parametric-statistical approach. *International Journal of Robust and Nonlinear Control*, published on-line in 2011 (doi: 10.1002/rnc.1791).

[18] M. Milanese, C. Novara, K. Hsu, and K. Poolla. The filter design from data (FD2) problem: Nonlinear Set Membership approach. *Automatica*, 45(10):2350–2357, 2009.

[19] C. Novara, F. Ruiz, and M. Milanese. A new approach to optimal filter design for nonlinear systems. In *18th IFAC World Congress*, Milano, Italy, 2011.

[20] C. V. Rao, J. B. Rawlings, and D. Q. Mayne. Constrained state estimation for nonlinear systems: stability and moving horizon approximations. *IEEE Transactions on Automatic Control*, 48(2):246–258, 2003.

[21] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond Kalman Filter: Particle Filters and Tracking Applications*. Artech House, Boston, MA, 2004.

Preprint submitted to Automatica
Received July 23, 2012 13:46:38 PST