49th IEEE Conference on Decision and Control
December 15-17, 2010
Hilton Atlanta Hotel, Atlanta, GA, USA

Advances in Moving Horizon Estimation for Nonlinear Systems

Angelo Alessandri, Marco Baglietto, Giorgio Battistelli, and Victor Zavala

Abstract— **In the last decade, moving horizon estimation (MHE) has emerged as a powerful technique for tackling the problem of estimating the state of a dynamic system in the presence of nonlinearities and disturbances. MHE is based on the idea of minimizing an estimation cost function defined on a sliding window composed of a finite number of time stages. The cost function is usually made up of two contributions: a prediction error computed on a recent batch of inputs and outputs; an arrival cost that serves the purpose of summarizing the past data. However, the diffusion of such techniques has been hampered by: i) the difficulty in choosing the arrival cost so as to ensure stability of the overall estimation scheme; ii) the request of an adequate computational effort on line.**

In this paper, both problems are addressed and possible solutions are proposed. First, by means of a novel stability analysis, it is constructively shown that under very general observability conditions a quadratic arrival cost is sufficient to ensure the stability of the estimation error provided that the weight matrix is adequately chosen. Second, a novel approximate MHE algorithm is proposed that is based on nonlinear programming sensitivity calculations. The approximate MHE algorithm has the same stability properties of the optimal one which make the overall approach suitable to be applied in real settings. Preliminary simulation results confirm the effectiveness of proposed method.

I. Introduction

First ideas on moving horizon estimation (MHE) date back to the sixties ([1]), motivated by the intrinsic robustness of such technique. MHE is based on the idea of minimizing an estimation cost function defined on a sliding window involving a finite number of time stages. The cost function is usually made up of two contributions: a prediction error computed on a recent batch of inputs and outputs; an arrival cost that serves the purpose of summarizing the past data.

An approach to the design of asymptotic state observers was proposed in [2] that results from the numerical solution of the measurement inversion problem via the Newton's method. In [3] and [4], a similar optimization-based techniques has been developed to construct estimators for continuous-time dynamic systems. In [5], an MH observer for nonlinear continuous-time systems was proposed that performs estimation at discrete-time instants by approximately minimizing an integral error defined on the preceding time window. More recently, advances have been obtained for MHE applied to linear systems ([6], [7], [8], [9]), hybrid systems ([10], [11]), and nonlinear systems ([12], [13], [14]).

The MH estimation scheme proposed in [13] allows one to explicitly take into account possible constraints on the system and requires the solution of a nonlinear programming problem at each time step. Moreover, a sufficient condition for the non-divergence of the estimation error in the presence of bounded noises was provided. Unfortunately, such approach requires the exact on-line minimization of a nonlinear cost function, thus reducing the practical possibility of applications. In order to overcome this drawback, a method was proposed in [12] with the possibility of admitting a certain error in the minimization of the cost function. Later on, these results were improved in [14], where the simultaneous presence of system and measurement noises was accounted for. In addition, more general conditions that guarantee the regional stability of the estimation error were presented.

In this paper, we extend the results of [14]. More specifically, the minimization of a more general cost function is addressed, that involves estimates of all the state vectors in the observation window. In this way, the estimation of the current state can be obtained without incurring delays nor resorting to the propagation of the nominal system. The observability conditions required in [14], based on a suitable $\mathcal{K}$-function, were recently demonstrated to be very general in that they are equivalent to uniform observability conditions based on the injectivity of the "observation map" or full rankness of its Jacobian (see [15], [16] and references therein). Here such conditions are made less restrictive by removing the requirement for a "finite sensitiveness" of the $\mathcal{K}$-function. Even in this more general case, stability of the proposed estimation technique can be proved and explicit bounds on the estimation error are guaranteed provided that a quadratic arrival cost is adopted and its weight matrix is adequately chosen.

One reason for the limited diffusion of MHE techniques is the request of an adequate computational effort on line. To cope with this, in [12], an approach was proposed that relies on the use of nonlinear approximate optimal estimation functions able to provide on-line estimates of the state variables. Such functions are implemented via neural networks with parameters determined off line. Further improvements to such methodology are reported in [17]. An alternative strategy to deal with the computational complexity is based on the use of the fast optimization techniques recently proposed for both MHE and model predictive control in [18], [19], [20], [21]. The idea is to make use of the sampling

A. Alessandri is with the Department of Production Engineering, Thermoenergetics, and Mathematical Models, DIPTEM–University of Genoa, P.le Kennedy Pad. D, 16129 Genova, Italy alessandri@diptem.unige.it

M. Baglietto is with the Department of Communications, Computer and System Sciences, DIST–University of Genoa, Via Opera Pia 13, 16145 Genova, Italy mbaglietto@dist.unige.it

G. Battistelli is with Dipartimento di Sistemi e Informatica, Università degli Studi di Firenze, Via di S. Marta 3, 50139 Firenze, Italy battistelli@dsi.unifi.it

V. Zavala is with the Mathematics and Computer Science Division Argonne National Laboratory 9700 South Cass Ave, Bldg. 221 Argonne, IL 60439, USA vzavala@mcs.anl.gov

time to solve a reference problem only with the available prediction of the state, while waiting for next and more precise information that allows one to correct the reference solution by performing a quick nonlinear programming (NLP) sensitivity calculation. These NLP sensitivity-based controllers and estimators are able to accommodate large-scale models in on-line environments, while dramatically reducing dangerous feedback delays.

Fast optimization techniques such as those described in [19], [20], [21] are based on the interior point NLP solver IPOPT (see [22]), which is able to exploit the sparse structure of MHE problems automatically at the linear algebra level. This provides an efficient approach to solve the reference problem in between sampling times.

Here, an approach to MHE that combines the reduced computational requirements of the sensitivity-based methods presented in [19], [20], [21], [23] with the stability guarantees of the techniques developed by [14] is presented.

The paper is organized as follows. In Section II, the considered MHE problem is formulated with particular attention to the definition of the arrival cost. The stability properties of the resulting MH estimation algorithms are proved in Section III in relation with the observability requirements. A general framework to find approximate solutions is described in Section IV. Finally, preliminary simulation results are presented in Section V.

II. PROBLEM STATEMENT

Let us consider a dynamic system described by the discrete-time equations

$$x_{t+1} = f(x_t, u_t) + w_t , \tag{1a}$$

$$y_t = h(x_t) + v_t , \tag{1b}$$

for $t = 0, 1, \ldots,$ where $x_t \in \mathbb{R}^n$ is the state vector (the initial state x_0 is unknown) and $u_t \in \mathbb{R}^m$ is the control vector. The vector $w_t \in \mathbb{R}^n$ is an additive disturbance affecting the system dynamics. The state vector is observed through the measurement equation (1b), where $y_t \in \mathbb{R}^p$ is the observation vector and $v_t \in \mathbb{R}^p$ is a measurement noise vector. We assume the statistics of x_0, w_t, and v_t to be unknown, and consider them as deterministic variables of unknown character.

In this paper, state estimation is addressed within a MHE framework [7], [9], [12], [14]. More specifically, at any time $t = N, N+1, \ldots$, the objective is to derive estimates of the state vectors $x_{t-N}, \ldots, x_t$ on the basis of the *information vector*[1]

$$I_t^{(N)} \triangleq \text{col}\,(y_{t-N,t}, u_{t-N,t-1}) \tag{2}$$

where $N+1$ measurements and N input vectors are collected within a "sliding window" $[t-N, t]$. Hereafter, $\hat{x}_{t-N|t}, \ldots, \hat{x}_{t|t}$ will denote the estimates (to be made at time t) of $x_{t-N}, \ldots, x_t$, respectively.

As we have assumed the statistics of the disturbances and of the initial continuous state to be unknown, a natural criterion to derive the estimator consists in resorting to a least-squares approach. Toward this end, at any time $t = N, N+1, \ldots$, the minimization of the following cost function can be addressed:[2]

$$\begin{aligned} &J_t^{(N)}\left(\hat{x}_{t-N,t|t}, I_t^{(N)}\right) \\ &= \Gamma_{t-N}\left(\hat{x}_{t-N|t}\right) + \sum_{i=t-N}^{t-1} \left\| \hat{x}_{i+1|t} - f(\hat{x}_{i|t}, u_i) \right\|_Q^2 \\ &+ \sum_{i=t-N}^{t} \left\| y_i - h(\hat{x}_{i|t}) \right\|_R^2 \end{aligned} \tag{3}$$

where the initial penalty function $\Gamma_{t-N}(\cdot)$ is assumed non-negative and the matrices Q, and R are assumed to be positive definite and can be regarded as design parameters. The first term is known in the MHE literature as *arrival cost* [6], [13] and serves the purpose of summarizing the past data, $y_{0,t-N-1}$ and $u_{0,t-N-1}$, not explicitly accounted for in the objective function. The second contribution, weighted by the matrix Q, takes into account the evolution of the state in terms of the state equation (1). Finally, the third term, weighted by the matrix R, penalizes the distances of the "expected output" (based on the state estimates) from the actual measurements.

It should be evident that the form of the arrival cost in (3) plays a crucial role in determining the behavior and the performance of the overall estimation scheme. In this connection, a first possibility [6], [13] would consists in choosing each function $\Gamma_{t-N}(\cdot)$ as the *true arrival cost*, i.e., the arrival cost obtained by imposing that minimization of the MH cost (3) at time t corresponds to minimization of the *full-information* cost

$$\begin{aligned} J_t^{\text{FI}}\left(\hat{x}_{0,t|t}, I_t^{(t)}\right) &= \Gamma_0\left(\hat{x}_{0|t}\right) + \sum_{i=0}^{t-1} \left\| \hat{x}_{i+1|t} - f(\hat{x}_{i|t}, u_i) \right\|_Q^2 \\ &+ \sum_{i=0}^{t} \left\| y_i - h(\hat{x}_{i|t}) \right\|_R^2 . \end{aligned}$$

It is quite straightforward to show that, under mild assumptions, such a choice always leads to a stable estimation error dynamics. Unfortunately, an algebraic expression for the true arrival cost exists only in few particular cases (e.g., the linear unconstrained case [6]). Therefore, when the system is nonlinear or constrained, some approximation has to be used. For instance, in [6], [13], it is shown that stability of the estimation error dynamics can be preserved provided that the approximate arrival cost is somehow bounded from above by the true one (this condition corresponds to requiring that the approximate arrival cost should not add additional information not specified in the data). While, from the theoretical point of view, such a choice allows for a clean stability analysis, from a practical perspective its applicability is severely limited by the fact that (with the notable exception

[1] In general, given a sequence $\{z_t\}$ and two time instants $0 \le t_1 \le t_2$, we shall use the notation $z_{t_1,t_2} \triangleq \text{col}(z_{t_1}, \ldots, z_{t_2})$.

[2] Throughout the paper, given a symmetric positive definite matrix M and a vector z, $\|z\|_M$ denotes the weighted norm of z, $\|z\|_M \triangleq (z^\top M z)^{1/2}$.

of the linear constrained case [6]) no constructive method is available for the determination of the approximate arrival cost.

A second possibility [12], [14] consists in assigning to the arrival cost a fixed structure that penalizes the distance of the estimate $\hat{x}_{t-N|t}$ of the state at the beginning of the sliding window from some prediction $\bar{x}_{t-N}$. With this respect, a natural choice is the quadratic arrival cost

$$\Gamma_{t-N}\left(\hat{x}_{t-N|t}\right) = \left\|\hat{x}_{t-N|t} - \bar{x}_{t-N}\right\|_P^2 , \quad P > 0 \quad (4)$$

with the prediction $\bar{x}_{t-N}$ chosen as the estimate of x_{t-N} made at the previous time instant $t-1$ (the vector $\bar{x}_0$ denotes an a-priori prediction of x_0). In Section III, it will be shown that, under quite general assumptions, such a simple arrival cost is sufficient to ensure the stability of the estimation error dynamics provided that the weight matrix P is adequately chosen. Hereafter, the MH cost with the quadratic arrival cost (4) will be denoted by $J^{(N)}\left(\hat{x}_{t-N,t|t}, \bar{x}_{t-N}, I_t^{(N)}\right)$.

In the lines of [13], [14], the following preliminary assumptions are needed.

A1. The sets $\mathcal{W}$, $\mathcal{V}$, and $\mathcal{U}$ where w_t, v_t and u_t (respectively) take their values are compact sets, with $0 \in \mathcal{W}$ and $0 \in \mathcal{V}$.

A2. The initial state x_0 and the control sequence $\{u_t\}$ are such that, for any possible sequence of disturbances $\{w_t\}$, the system trajectory $\{x_t\}$ lies in a compact set $\mathcal{X}$.

Note that Assumptions A1 and A2 are quite reasonable from a practical point of view when considering the state estimation problem for a physical system: it is very typical that the state variables as well as the exogenous inputs are bounded in some way. Since, under assumption A2, at every time step $t = 0, 1, \ldots$, the state x_t falls within the set $\mathcal{X}$, it is natural to include the constraints

$$\hat{x}_{i|t} \in \mathcal{X} , \quad i = t-N, \ldots, t \quad (5)$$

when addressing the minimization of the cost functional (3). Further, when also the sets $\mathcal{W}$ and $\mathcal{V}$ are known, the additional constraints

$$\hat{x}_{i+1|t} - f(\hat{x}_{i|t}, u_i) \in \mathcal{W} , \qquad i = t-N, \ldots, t-1 \quad (6)$$

$$y_i - h(\hat{x}_{i|t}) \in \mathcal{V} , \qquad i = t-N, \ldots, t \quad (7)$$

can be imposed. It is worth noting that the state constraints (5) are in general crucial for ensuring the applicability of the approach (in fact, most nonlinear programming algorithms are designed to work only for bounded solution sets). On the contrary, the constraint sets (6) and (7) turn out to be unessential not only for the applicability of the approach but also for the stability analysis (as will be shown in the next section). Moreover, in some applications, it may be difficult to know an upper bound on the exogenous inputs (for example due to the presence of outliers). In view of the above discussion, only the state constraints (5) will be taken into account in the subsequent problem formulation.

The following algorithm can now be stated.

Algorithm 1 (ε-optimal MHE). Given an a-priori prediction $\bar{x}_0$ and a desired accuracy $\varepsilon > 0$, at any time $t = N, N+1, \ldots$

1) find some state estimates $\hat{x}^{\varepsilon}_{t-N|t} \in \mathcal{X}, \ldots, \hat{x}^{\varepsilon}_{t|t} \in \mathcal{X}$ such that

$$J^{(N)}(\hat{x}^{\varepsilon}_{t-N,t|t}, \bar{x}_{t-N}, I_t^{(N)}) \leq J^{(N)}(\hat{x}_{t-N,t|t}, \bar{x}_{t-N}, I_t^{(N)}) + \varepsilon \quad (8)$$

for any $\hat{x}_{t-N|t} \in \mathcal{X}, \ldots, \hat{x}_{t|t} \in \mathcal{X}$;

2) set

$$\bar{x}_{t-N+1} = \hat{x}^{\varepsilon}_{t-N+1|t} .$$

The positive real ε represents the desired accuracy in the minimization of the MH cost. In fact, condition (8) amounts to requiring that the cost related to the "approximate estimates" $\hat{x}^{\varepsilon}_{t-N,t|t}$ should not exceed the *optimal cost*

$$\min_{\hat{x}_{t-N,t|t} \in \mathcal{X}^{N+1}} J^{(N)}(\hat{x}_{t-N,t|t}, \bar{x}_{t-N}, I_t^{(N)})$$

by more than ε. Of course, a compromise must be accepted when selecting such a quantity. In fact, the smaller the ε, the better the expected estimation, but a larger computational effort is required to guarantee such an accuracy. In practical situations, a large enough ε must be considered so that the optimization process may end in each sampling period.

III. Stability of the estimation error

In order to study the stability properties of the proposed MHE algorithm, some preliminary definitions and assumptions are needed.

First of all, let $\text{co}(\mathcal{X})$ be the convex closure of $\mathcal{X}$. Then, the following smoothness requirement on system (1) can be introduced.

A3. The functions f and h are $\mathcal{C}^2$ functions with respect to x on $\text{co}(\mathcal{X})$ for every $u \in \mathcal{U}$.

Let the *observation map* on a window of length $N+1$ be defined as

$$F^{(N)}\left(x_{t-N}, u_{t-N,t-1}, w_{t-N,t-1}\right) \triangleq \begin{bmatrix} h\left(x_{t-N}\right) \\ h \circ f^{u_{t-N},w_{t-N}}(x_{t-N}) \\ \vdots \\ h \circ f^{u_{t-1},w_{t-1}} \circ \cdots \circ f^{u_{t-N},w_{t-N}}(x_{t-N}) \end{bmatrix},$$

where "$\circ$" denotes function composition and $f^{u_i,w_i}(x_i) \triangleq f\left(x_i, u_i\right) + w_i$. Notice that, by exploiting such a definition, one can write

$$y_{t-N,t} = F^{(N)}\left(x_{t-N}, u_{t-N,t-1}, w_{t-N,t-1}\right) + v_{t-N,t} .$$

Further, by defining the estimated system disturbances as

$$\hat{w}_{i|t} \triangleq \hat{x}_{i+1|t} - f(\hat{x}_{i|t}, u_i) , \quad i = t-N, \ldots, t-1 ,$$

the MH cost with quadratic arrival cost can be rewritten as

$$\begin{aligned} & J_t\left(\hat{x}_{t-N,t|t}, \bar{x}_{t-N}, I_t^{(N)}\right) \\ & \quad = \left\|\hat{x}_{t-N|t} - \bar{x}_{t-N}\right\|_P^2 + \left\|\hat{w}_{t-N,t-1|t}\right\|_{\bar{Q}}^2 \\ & \quad + \left\| y_{t-N,t} - F^{(N)}\left(\hat{x}_{t-N|t}, u_{t-N,t-1}, \hat{w}_{t-N,t-1|t}\right)\right\|_{\bar{R}}^2 \end{aligned} \tag{9}$$

where

$$\bar{Q} \triangleq \mathbb{I}_N \otimes Q\,, \quad \bar{R} \triangleq \mathbb{I}_{N+1} \otimes R$$

with $\otimes$ the Kronecker product and $\mathbb{I}_N$ the $N \times N$ identity matrix.

We are now able to introduce the following observability definitions.

Definition 1: System (1) is said to be observable in $N+1$ steps if there exists a K-function[3] $\varphi(\cdot)$, such that

$$\varphi\left(\|x_1 - x_2\|^2\right) \leq \left\|F^{(N)}(x_1, u, 0) - F^{(N)}(x_2, u, 0)\right\|^2, \tag{10}$$

$\forall\, x_1,\, x_2 \in \mathcal{X}$ and $\forall\, u \in \mathcal{U}^N$.

Definition 2: System (1) is said to be observable in $N+1$ steps with *finite sensitivity* $1/\delta$ if it is observable in $N+1$ steps and, in addition, the K-function $\varphi(\cdot)$ satisfies the following condition

$$\delta \triangleq \inf_{x_1, x_2 \in \mathcal{X};\, x_1 \neq x_2} \frac{\varphi\left(\|x_1 - x_2\|^2\right)}{\|x_1 - x_2\|^2} > 0\,. \tag{11}$$

The observabilty definition expressed by (10) has been widely used in the framework of nonlinear state estimation in both the discrete-time and the continuous-time settings (see for example [2], [4], [24], [13], [14]). The additional condition (11) was introduced in [14] by taking into account the sensitivity of the inverse mapping from the noise-free observations $y = F^{(N)}(x, u, 0)$ to the state x. In fact, the fulfillment of condition (11) ensures that small variations in the observation vector y always correspond to small variations of the state vector x. Recently [15], [16], it has been shown that such definitions are quite general and natural in that, under mild assumptions: i) Definition 1 is equivalent to the injectivity of the mapping from the state x to the noise-free observations $y = F^{(N)}(x, u, 0)$; ii) Definition 2 is equivalent to the observability rank condition

$$\text{rank}\frac{\partial F^{(N)}(x, u, 0)}{\partial x} = n\,, \quad \forall x \in \mathcal{X},\, \forall u \in \mathcal{U}^N;$$

iii) taking into account observation windows of fixed length is not restrictive.

Let us denote by k_f an upper bound on the Lipschitz constant of $f(x, u)$ with respect to x on $\mathcal{X}$ for every $u \in U$. Further, let

$$r_w \triangleq \max_{w \in \mathcal{W}} \|w\|^2\,, \quad r_v \triangleq \max_{v \in \mathcal{V}} \|v\|^2\,.$$

[3]Recall that a function $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ is a K-function if it is continuous, strictly monotone increasing, and such that $\varphi(0) = 0$.

Finally, for the sake of presentation simplicity, suppose that the matrix P is diagonal, i.e., $P = p\,\mathbb{I}_n$ with p a positive real (notice that this can be done without loss of generality because the general case can be subsumed to this special one by means of a suitable linear change of variables). Then, the following stability result can be stated.

Theorem 1: Suppose that assumptions A1-A3 are satisfied and that system (1) is observable in $N+1$ steps with finite sensitivity $1/\delta$. Further, let the estimates $\hat{x}^{\varepsilon}_{t-N|t}$ and $\hat{w}^{\varepsilon}_{t-N,t-1|t}$, $t = N, N+1, \ldots$, be computed recursively via Algorithm 1. Then, the following upper bounds hold

$$\|x_{t-N} - \hat{x}^{\varepsilon}_{t-N|t}\|^2 \leq \xi_{t-N} \tag{12}$$

$$\left\|\hat{w}^{\varepsilon}_{t-N,t-1|t}\right\|^2 \leq \omega_{t-N} \tag{13}$$

where the sequences $\{\xi_t\}$ and $\{\omega_t\}$ are generated by the linear system

$$\begin{bmatrix} \omega_{t+1} \\ \xi_{t+1} \end{bmatrix} = A(p, \delta) \begin{bmatrix} \omega_t \\ \xi_t \end{bmatrix} + B(p, \delta) \begin{bmatrix} r_w \\ r_v \end{bmatrix} + C(p, \delta)\,\varepsilon \tag{14}$$

with

$$\begin{aligned} A(p, \delta) &\triangleq \begin{bmatrix} c_1\, p & c_1\, k_f\, p \\ c_2\, p/(p + c_3\,\delta) & c_2\, k_f\, p/(p + c_3\,\delta) \end{bmatrix}, \\ B(p, \delta) &\triangleq \begin{bmatrix} c_4\, p + c_5 & c_6 \\ (c_7\, p + c_8)/(p + c_3\,\delta) & c_9/(p + c_3\,\delta) \end{bmatrix}, \\ C(p, \delta) &\triangleq \begin{bmatrix} c_{10} \\ c_{11}/(p + c_3\,\delta) \end{bmatrix}, \end{aligned}$$

and c_i, $i = 1, \ldots, 11$ are suitable positive constants.

□

Some remarks on Theorem 1 are in order. First, it is worth pointing out that, from (12) and (13), one can easily derive also an upper bound on the estimation error $\|x_t - \hat{x}^{\varepsilon}_{t|t}\|^2$ by simple Lipschitz arguments. Further, it can be seen that (12) and (13) still hold, without any modification, even when the additional constraints (6) and (7) are included in the state estimation algorithm. Finally, noting that the eigenvalues of the matrix $A(p, \delta)$ are 0 and

$$a(p, \delta) \triangleq c_1\, p + k_f\, c_2\, p/(p + c_3\,\delta)\,,$$

the following corollary to Theorem 1 can be readily stated.

Corollary 1: Under the same assumptions of Theorem 1, if the design parameter p is selected such that

$$a(p, \delta) < 1\,, \tag{15}$$

then the dynamics of system (14) is asympotically stable and the upper bounding sequences $\{\omega_t\}$ and $\{\xi_t\}$ converge exponentially to the asymptotic values $\omega_\infty(p, \delta)$ and $\xi_\infty(p, \delta)$, respectively, obtained as

$$\begin{bmatrix} \omega_\infty(p, \delta) \\ \xi_\infty(p, \delta) \end{bmatrix} \triangleq (\mathbb{I}_2 - A(p, \delta))^{-1} \left(B(p, \delta) \begin{bmatrix} r_w \\ r_v \end{bmatrix} + C(p, \delta)\varepsilon\right). \tag{16}$$

□

Since $a(0,\delta) = 0$, it is immediate to see that condition (15) can be easily satisfied by imposing that the positive weight p does not exceed a certain stability threshold $p_{\max}$. For instance, it can be seen that the smaller is k_f (i.e., the more contractive is the system) and the larger is δ (i.e., the more observable is the system) the wider is the range of values of p that satisfy condition (15).

When system (1) is noise-free (i.e., $\mathcal{W} = \{0\}$ and $\mathcal{V} = \{0\}$) and the accuracy ε is set to zero, the asymptotic upper bounds $\omega_\infty(p,\delta)$ and $\xi_\infty(p,\delta)$ turn out to be equal to zero. Thus, in this case, the MHE algorithm yields an exponential observer, provided that the stability condition (15) is satisfied. In the other cases, the asymptotic upper bounds grow linearly with the amplitude of the noises (i.e., r_w and r_v) as well as the with the accuracy ε.

In some cases, the finite sensitivity requirement expressed by equation (11) may not hold even if the system itself is observable. Of course, the results of Theorem 1 and Corollary 1 could still be applied also in these situations by setting $\delta = 0$. However, the resulting stability condition would become quite restrictive and, possibly, unfeasible for any choice of the design parameter p. Nevertheless, by following a different approach, it is all the same possible to derive more meaningful upper bounds and stability conditions. In this connection, the basic idea consists in excluding a certain interval $[0,m)$ (with m a positive real) from the calculus of the infimum in condition (11). Then one can define the quantity

$$\delta^{(m)} \triangleq \min_{x_1,x_2\in\mathcal{X},\|x_1-x_2\|^2\geq m} \frac{\varphi\left(\|x_1-x_2\|^2\right)}{\|x_1-x_2\|^2} \tag{17}$$

which is always greater than 0 for any $m > 0$ provided that system (1) is observable according to Definition 1. Thanks to such an arrangement, the following result can be stated.

Theorem 2: Suppose that assumptions A1-A3 are satisfied and that system (1) is observable in $N+1$ steps. Further, let the estimates $\hat{x}^{\varepsilon}_{t-N|t}$ and $\hat{w}^{\varepsilon}_{t-N,t-1|t}$, $t = N, N+1, \ldots$, be computed recursively via Algorithm 1. Then, for any $m > 0$, the following upper bounds hold

$$\|x_{t-N} - \hat{x}^{\varepsilon}_{t-N|t}\|^2 \leq \xi^{(m)}_{t-N} \tag{18}$$

$$\left\|\hat{w}^{\varepsilon}_{t-N,t-1|t}\right\|^2 \leq \omega^{(m)}_{t-N} \tag{19}$$

where the sequences $\{\xi^{(m)}_t\}$ and $\{\omega^{(m)}_t\}$ are generated by the nonlinear system

$$\begin{aligned} \begin{bmatrix} \omega^{(m)}_{t+1} \\ \overline{\xi}^{(m)}_{t+1} \end{bmatrix} &= A(p,\delta^{(m)}) \begin{bmatrix} \omega^{(m)}_t \\ \xi^{(m)}_t \end{bmatrix} + B(p,\delta^{(m)}) \begin{bmatrix} r_w \\ r_v \end{bmatrix} \\ &\quad + C(p,\delta^{(m)})\,\varepsilon\,, \\ \xi^{(m)}_{t+1} &= \max\left\{m\,,\,\overline{\xi}^{(m)}_{t+1}\right\}. \end{aligned} \tag{20}$$

□

Notice that, in the limit $m \to 0$, the novel bounding sequences of Theorem 2 tend to those of Theorem 1 (in fact, $\delta^{(0)} = \delta$). As to the asymptotic behavior of such bounding sequences, the following result holds.

Theorem 3: Let the same assumptions of Theorem 2 hold and let the design parameter p be such that

$$a(p,\delta^{(m)}) < 1\,. \tag{21}$$

Then, system (20) admits a globally exponentially stable equilibrium point.

In particular, considering the functions $\omega_\infty(\cdot,\cdot)$ and $\xi_\infty(\cdot,\cdot)$ defined in (16), one of the following two possibilities occur:

i. if $m \leq \xi_\infty(p,\delta^{(m)})$,

$$\lim_{t\to\infty} \xi^{(m)}_t = \xi_\infty(p,\delta^{(m)})\,,$$
$$\lim_{t\to\infty} \omega^{(m)}_t = \omega_\infty(p,\delta^{(m)})\,;$$

ii. if $m > \xi_\infty(p,\delta^{(m)})$,

$$\lim_{t\to\infty} \xi^{(m)}_t = m\,,$$
$$\lim_{t\to\infty} \omega^{(m)}_t = \frac{1}{1-c_1\,p} \times (c_1\,k_f\,m + (c_4\,p + c_5)r_w + c_6\,r_v + c_{10}\,\varepsilon)\,.$$

□

Such a stability result overcomes the limitations of Corollary 1. In fact, for an observable system the scalar $\delta^{(m)}$ is always strictly positive (even if the finite sensitivity condition (11) does not hold) and consequently the stability condition (21) is always feasible, i.e., for any given $m > 0$ there always exist a threshold $p^{(m)}_{\max} > 0$ such that condition (21) holds for any choice of the design parameter p in the interval $[0, p^{(m)}_{\max})$.

It is also worth noting that, actually, Theorem 3 yields a family of asymptotic upper bounds, one for each choice of m. In this connection, since by construction $\delta^{(m)}$ increases with m, it can be verified that $\xi_\infty(p,\delta^{(m)})$ decreases with m when the value of the design parameter p is fixed. Thus, provided that the behavior of $\delta^{(m)}$ as a function of m is known (either analytically or by numerical tabulation), the best upper bound among such a family can be obtained by choosing m as the unique solution of

$$m = \xi_\infty(p,\delta^{(m)})\,.$$

Notice that this strategy can be applied also when $\delta > 0$ in order to obtain less conservative upper bounds (this can be useful when the sensitivity $1/\delta$ of the observability map is finite but very large, i.e., the parameter δ is close to zero).

IV. One-step ahead moving-horizon estimation

As a matter of fact, even allowing for a certain error ε, the minimization involved in Algorithm 1 cannot be completed instantaneously. This would induce a delay in the estimation process that can be problematic in some applications, e.g., when MHE is used in connection with a state feedback controller. A possibility for facing such a drawback consists in the use of parameterized functions (to be optimized off line) to approximate the estimation function obtained applying Algorithm 1. Such a possibility was considered in [12] and, more recently, in [17]. Here, a different approach is adopted.

First of all, in order to derive an approximate MHE algorithm providing an estimate of the continuous state almost in real-time, let us suppose that the MH cost can be minimized with accuracy $\varepsilon = 0$ in less than one sampling interval. It is pointed out that this simplifying assumption is adopted only for the sake of clarity, however the strategy proposed hereafter actually ensures the stability of the estimation error dynamics for a generic accuracy ε. Further, the case when the delay is greater than one sampling interval can be dealt with in a similar way by means of straightforward modifications.

Of course, this amounts to assuming that, at each time t, the most recent available estimates are the optimal estimates $\hat{x}^{\circ}_{t-N-1,t-1|t-1}$ that solve the optimization problem

$$\min_{\hat{x}_{t-N-1,t-1|t-1} \in \mathcal{X}^{N+1}} J^{(N)}\left(\hat{x}_{t-N-1,t-1|t-1}, \bar{x}_{t-N-1}, I^{(N)}_{t-1}\right) . \tag{22}$$

Then, one can exploit the following fact.

Proposition 1: Let the estimates $\hat{x}^{\circ}_{t-N-1,t-1|t-1}$ be a solution of the optimization problem (22) and consider the one-step ahead optimal prediction

$$\hat{x}^{\circ}_{t|t-1} = f(\hat{x}^{\circ}_{t-1|t-1}, u_{t-1}) .$$

Then, the estimates $\hat{x}^{\circ}_{t-N-1,t|t-1}$ are also a solution of the optimization problem

$$\min_{\hat{x}_{t-N-1,t|t} \in \mathcal{X}^{N+2}} J^{(N+1)}\left(\hat{x}_{t-N-1,t|t}, \bar{x}_{t-N-1}, \overline{I}^{(N+1)}_{t}\right) \tag{23}$$

where

$$\overline{I}^{(N+1)}_{t} \triangleq \mathrm{col}\left(u_{t-N-1,t-1}, y_{t-N-1,t-1}, \bar{y}_t\right) ,$$
$$\bar{y}_t \triangleq h(\hat{x}^{\circ}_{t|t-1}) .$$

□

In the light of Proposition 1, one can exploit well-known results on the sensitivity of NLP problems with respect to variations in the problem data and update the currently available estimates $\hat{x}^{\circ}_{t-N-1,t-1|t-1}$ on the basis of the difference between the predicted measurement $\bar{y}_t$ and the true one y_t.

To this end, suppose that the estimates $\hat{x}^{\circ}_{t-N-1,t|t-1}$ correspond to a strict *isolated* minimizer of (23) satisfying the so-called strong second-order sufficient conditions (SSOC) (see [25]). As well known, when the estimates $\hat{x}^{\circ}_{t-N-1,t|t-1}$ do not lie at the boundary of the set $\mathcal{X}^{N+2}$, such conditions take the form

$$\frac{\partial J^{(N+1)}}{\partial \hat{x}_{t-N-1,t|t}}\left(\hat{x}^{\circ}_{t-N-1,t|t-1}, \bar{x}_{t-N-1}, \overline{I}^{(N+1)}_{t}\right) = 0 , \tag{24}$$

$$\frac{\partial^2 J^{(N+1)}}{\left(\partial \hat{x}_{t-N-1,t|t}\right)^2}\left(\hat{x}^{\circ}_{t-N-1,t|t-1}, \bar{x}_{t-N-1}, \overline{I}^{(N+1)}_{t}\right) > 0 . \tag{25}$$

For a detailed SSOC analysis in the context of MHE (also discussing the case wherein the optimal estimates lie at the boundary of $\mathcal{X}^{N+2}$), the interested reader is referred to Chapters 3 and 6 in [26].

The satisfaction of SSOC also has implications on the sensitivity of the solution to perturbations on the problem data around the reference solution $\hat{x}^{\circ}_{t-N-1,t|t-1}$. To explore this, we use the following well-known result, adapted to our context.

Theorem 4: (NLP Sensitivity) [27], [28]. Suppose that the SSOC (24) and (25) hold at $\hat{x}^{\circ}_{t-N-1,t|t-1}$ for the cost $J^{(N+1)}(\hat{x}_{t-N-1,t|t}, \bar{x}_{t-N-1}, \overline{I}^{(N+1)}_{t})$, then the following hold:

- There exists a unique, continuous and differentiable vector function $a^{\circ}(\cdot)$ such that,
  - $a^{\circ}(\bar{y}_t) = \hat{x}^{\circ}_{t-N-1,t|t-1}$;
  - for any y_t in a neighborhood of $\bar{y}_t$, $a^{\circ}(y_t)$ is a strict isolated minimizer satisfying SSOC.
- The optimal cost is locally Lipschitz in a neighborhood of $\bar{y}_t$.

□

In the light of Theorem 4, one can construct a linear update formula of the form

$$\hat{x}^{u}_{t-N-1,t|t} = \hat{x}^{\circ}_{t-N-1,t|t-1} + K_t(y_t - \bar{y}_t) \tag{26}$$

by means of a first-order Taylor expansion of $a^{\circ}(\cdot)$ around $\bar{y}_t$. For instance, exploiting the implicit function theorem, the gain K_t turns out to be

$$K_t = \left\{ \frac{\partial^2 J^{(N+1)}}{(\partial \hat{x}_{t-N-1,t|t-1})^2}\left(\hat{x}^{\circ}_{t-N-1,t|t-1}, \bar{x}_{t-N-1}, \overline{I}^{(N+1)}_{t}\right) \right\}^{-1} \times \frac{\partial^2 J^{(N+1)}}{\partial \hat{x}_{t-N-1,t|t-1} \partial y_t}\left(\hat{x}^{\circ}_{t-N-1,t|t-1}, \bar{x}_{t-N-1}, \overline{I}^{(N+1)}_{t}\right) .$$

Of course, such a first-order Taylor expansion provides a good approximation for the minimizer of $J^{(N+1)}(x_{t-N-1,t|t-1}, \bar{x}_{t-N-1}, I^{(N+1)}_{t})$ only in a neighborhood of $\bar{y}_t$. Conversely, for large values of $y_t - \bar{y}_t$, the linear update formula (26) may even lead to a worse state estimate than the original one. However, while it is not possible to directly compare the two estimates $\hat{x}^{u}_{t-N-1,t|t}$ and $\hat{x}^{\circ}_{t-N-1,t|t-1}$ being the true states $x_{t-N-1,t}$ unknown, it is still possible to make an indirect comparison by means of cost $J^{(N+1)}(\cdot)$. For instance, one can argue that $\hat{x}^{u}_{t-N-1,t|t}$ represents an improvement with respect to $\hat{x}^{\circ}_{t-N-1,t|t-1}$ only if

$$J^{(N+1)}(x^{u}_{t-N-1,t|t}, \bar{x}_{t-N-1}, I^{(N+1)}_{t}) \leq J^{(N+1)}(x^{\circ}_{t-N-1,t|t-1}, \bar{x}_{t-N-1}, I^{(N+1)}_{t}) . \tag{27}$$

The foregoing discussion leads to the following one-step ahead MHE algorithm.

Algorithm 2 (One-step ahead MHE). Given an a-priori prediction $\bar{x}_0$, at any time $t = N, N+1, \ldots$

1) if $t = N$, then go to step 5;
   else go to step 2;
2) if the SSOC (24) and (25) hold at $\hat{x}^{\circ}_{t-N-1,t|t-1}$, then compute the updated estimates $\hat{x}^{u}_{t-N-1,t|t}$ as in (26) and go to step 3;
   else set $\hat{x}^{u}_{t-N-1,t|t} = \hat{x}^{\circ}_{t-N-1,t|t-1}$ and go to step 4;
3) if inequality (27) holds, then set $\hat{x}^{+}_{t-N-1,t|t} = \hat{x}^{u}_{t-N-1,t|t}$;
   else set $\hat{x}^{+}_{t-N-1,t|t} = \hat{x}^{\circ}_{t-N-1,t|t-1}$;
4) set
$$\bar{x}_{t-N} = \hat{x}^{\circ}_{t-N|t-1}\,; \tag{28}$$
5) find a solution $\hat{x}^{\circ}_{t-N,t|t}$ to the NLP problem
$$\min_{\hat{x}_{t-N,t|t} \in \mathcal{X}^{N+1}} J^{(N)}\left(\hat{x}_{t-N,t|t}, \bar{x}_{t-N}, I_t^{(N)}\right)\,.$$

Remark 1: The estimates $\hat{x}^{u}_{t-N-1,t|t}$ are understood to belong to the set $\mathcal{X}^{N+2}$. If it is not the case, then one can either project $\hat{x}^{u}_{t-N-1,t|t}$ onto $\mathcal{X}^{N+2}$ (provided that this is computationally feasible) or consider the test in step 3 automatically failed and set $\hat{x}^{+}_{t-N-1,t|t} = \hat{x}^{\circ}_{t-N-1,t|t-1}$.

Remark 2: The main advantage of such an algorithm is that the estimation recursion is unaffected by the linear update formula. Further, thanks to the test concerning the cost improvement, it can be seen that stability of the estimation error $x_t - \hat{x}^{+}_{t|t}$ is ensured under the very same assumptions of the ε-optimal MHE algorithm of Section II.

Remark 3: Notice that in step 4, the prediction is obtained from the estimates $\hat{x}^{\circ}_{t-N-1,t|t-1}$ instead that from $\hat{x}^{+}_{t-N-1,t|t}$. Of course, since $\hat{x}^{+}_{t-N-1,t|t}$ are believed to be better estimates than $\hat{x}^{\circ}_{t-N-1,t|t-1}$, it would be natural to replace (28) with
$$\bar{x}_{t-N} = \hat{x}^{+}_{t-N|t-1}\,. \tag{29}$$
Even for such a modified algorithm stability could still be ensured, but the details are not presented here due to space contraints.

V. Numerical Case Study

In this section, we illustrate the effect of numerical errors on the performance of the MH estimators described in this paper. We consider a simulated MHE scenario on the nonlinear continuous stirred tank reactor described in [29]:
$$\frac{dx^{(1)}}{d\tau} = \frac{x^{(1)}(\tau) - 1}{\theta} + k_0 x^{(1)}(\tau)\exp\left[\frac{-E_a}{x^{(2)}(\tau)}\right] + w^{(1)}(\tau)$$
$$\frac{dx^{(2)}}{d\tau} = \frac{x^{(2)}(\tau) - x_f^{(2)}}{\theta} - k_0 x^{(1)}(\tau)\exp\left[\frac{-E_a}{x^{(2)}(\tau)}\right]$$
$$+ \alpha\, u(\tau)(x^{(2)}(\tau) - x_{cw}^{(2)}) + w^{(2)}(\tau),$$
where the parameters were chosen as follows: $\theta = 100$ $x_{cw}^{(2)} = 0.38$, $x_f^{(2)} = 0.395$, $E_a = 5$, $\alpha = 1.95 \times 10^4$, and $k_0 = 300$.

The system involves two states: $x = [x^{(1)}, x^{(2)}]^{\top}$ corresponding to the concentration and temperature, respectively, one control u corresponding to the cooling water flowrate and two process noise sequences $w = [w^{(1)}, w^{(2)}]^{\top}$. The continuous-time model is transformed into a discrete-time form through an implicit Euler discretization scheme. The temperature is used as the measured output ($y_t = x_t^{(2)}$) to infer the concentration $x_t^{(1)}$. We use batch data generated from a simulated closed-loop feedback control scenario with Gaussian process noise sequences with standard deviations $\sigma_{w^{(1)}} = 0.01$, $\sigma_{w^{(2)}} = 0.05$. The resulting temperatures are then corrupted with a Gaussian noise with standard deviation $\sigma_y = 0.0125$ to simulate measurement noise. We use $\bar{x}_0 = [0.15,\ 0.15]$ as the a-priori prediction for the initial state x_0 and a regularization penalty $P = 30\ \mathbb{I}_2$. The estimators are simulated over 250 time steps.

In the top graph of Figure 1, we compare the estimates obtained by Algorithm 1 and by the same algorithm where the second term in (3), weighting the "estimated system noises" is omitted (denote it as "Algorithm 1 No Est"). In the latter algorithm (see [14]), only the state at the beginning of the observation window is estimated and corresponds to setting the estimated system noises to zero. An observation window $N = 3$ is used. As can be observed, the use of the term weighting the estimates of process disturbances improves the performance. This is particularly noticeable at the beginning of the trajectory. In the bottom graph of the same figure we compare the performance of Algorithm 1 and of Algorithm 2. Note that the performance of both algorithms is nearly identical despite the large levels of noise and the short observation window.

In Figure 2 we illustrate the performance of Algorithm 2 for increasing observation horizons. Here, we compare the mean sum of squared errors (SSE) with respect to Algorithm 1 of the state estimates $x_t^{(1)^{\epsilon}}$ generated with Algorithm 2 *with and without* considering the term weighting the estimated process noise in the cost (3). The mean SSE is defined as $\frac{1}{250}\sum_{t=1}^{250}(x_t^{(1)\epsilon} - x_t^{(1)\circ})^2$. We can see that the performance of Algorithm 2 approaches that of the optimal Algorithm 1 when process noise is estimated. This is due to the fact that more accurate step-ahead predictions can be obtained when noise sequences are explicitely accounted. This in turn reduces the sensitivity errors. Note also that the effect of noise is appreciable at short horizons but this effect dies out quickly as the horizon is increased. In this study, $N = 3$ is sufficient to obtain close to optimal performance. Finally, it is interesting to observe that, as the horizon is increased, Algorithm 2 achieves close to optimal performance.

References

[1] A. H. Jazwinski, "Limited memory optimal filtering," *IEEE Transactions on Automatic Control*, vol. 13, no. 5, pp. 558–563, 1968.

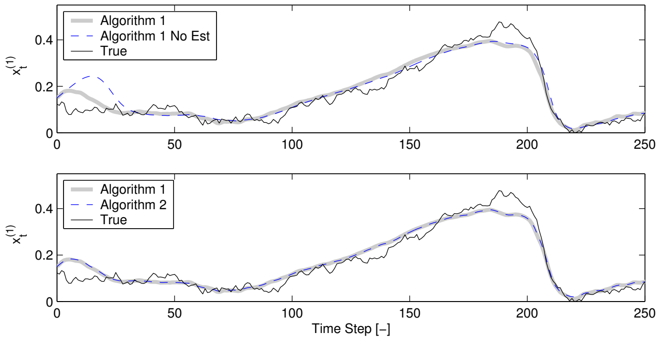

Fig. 1. Performance of Algorithm 1 with and without estimation of process noise disturbances (top). Performance of Algorithm 1 and 2 with estimation of process noise disturbances (bottom).

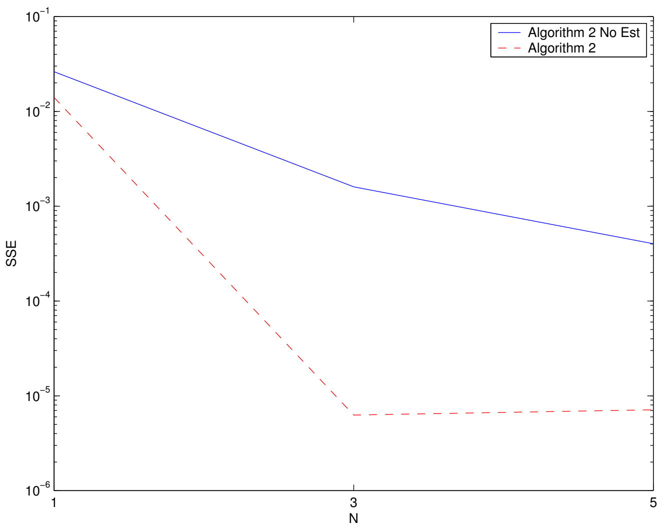

Fig. 2. Performance of Algorithm 2 as horizon is increased.

[2] P. E. Moraal and J. W. Grizzle, "Observer design for nonlinear systems with discrete–time measurements," *IEEE Transactions on Automatic Control*, vol. 40, no. 3, pp. 395–404, 1995.

[3] G. Zimmer, "State observation by on-line minimization," *International Journal of Control*, vol. 60, no. 4, pp. 595–606, 1994.

[4] M. Alamir, "Optimization based non-linear observers revisited," *International Journal of Control*, vol. 72, no. 13, pp. 1204–1217, 1999.

[5] H. Michalska and D. Q. Mayne, "Moving horizon observers and observer–based control," *IEEE Transactions on Automatic Control*, vol. 6, no. 6, pp. 995–1006, 1995.

[6] C. V. Rao, J. B. Rawlings, and J. H. Lee, "Constrained linear estimation–a moving horizon approach," *Automatica*, vol. 37, no. 10, pp. 1619–1628, 2001.

[7] A. Alessandri, M. Baglietto, and G. Battistelli, "Receding-horizon estimation for discrete-time linear systems," *IEEE Transactions on Automatic Control*, vol. 48, no. 3, pp. 473–478, 2003.

[8] ——, "On estimation error bounds for receding-horizon filters using quadratic boundedness," *IEEE Transactions on Automatic Control*, vol. 49, no. 8, pp. 1350–1355, 2004.

[9] ——, "Robust receding-horizon state estimation for uncertain discrete-time linear systems," *Systems & Control Letters*, vol. 54, no. 7, pp. 627–643, 2005.

[10] G. Ferrari-Trecate, D. Mignone, and M. Morari, "Moving-horizon estimation for hybrid systems," *IEEE Transactions on Automatic Control*, vol. 47, no. 10, pp. 1663–1676, 2003.

[11] A. Alessandri, M. Baglietto, and G. Battistelli, "Receding-horizon estimation for switching discrete-time linear systems," *IEEE Transactions on Automatic Control*, vol. 50, no. 11, pp. 1736–1748, 2005.

[12] A. Alessandri, M. Baglietto, T. Parisini, and R. Zoppoli, "A neural state estimator with bounded errors for nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 44, no. 11, pp. 2028–2042, 1999.

[13] C. V. Rao, J. B. Rawlings, and D. Q. Mayne, "Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations," *IEEE Transactions on Automatic Control*, vol. 48, no. 2, pp. 246–257, 2003.

[14] A. Alessandri, M. Baglietto, and G. Battistelli, "Moving-horizon state estimation for nonlinear discrete-time systems: New stability results and approximation schemes," *Automatica*, vol. 44, no. 7, pp. 1753–1765, 2008.

[15] S. Hanba, "On the uniform observability of discrete-time nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 54, no. 8, pp. 1925–1928, 2009.

[16] ——, "Further results on the uniform observability of discrete-time nonlinear systems," *IEEE Transactions on Automatic Control*, to appear.

[17] A. Alessandri, M. Baglietto, G. Battistelli, and R. Zoppoli, "Moving-horizon state estimation for nonlinear systems using neural networks," in *47th Conference on Decision and Control*, Cancun, MX, 2008, pp. 2557–2562.

[18] T. Kraus, P. Kuehl, L. Wirsching, H. G. Bock, and M. Diehl, "A moving horizon state estimation algorithm applied to the tennessee eastman benchmark process," in *Proceedings of IEEE Conference on Multisensor Fusion and Integration*, vol. 44, 2005, pp. 2541–2460.

[19] V. Zavala, C. Laird, and L. Biegler, "A fast moving horizon estimation algorithm based on nonlinear programming sensitivity," *Journal of Process Control*, vol. 18, pp. 876–884, 2008.

[20] ——, "A fast computational framework for large-scale moving horizon estimation," in *Proceedings of 8th International Symposium on Dynamics and Control of Process Systems, Cancun, Mexico*, 2007.

[21] V. Zavala and L. Biegler, "The advanced-step NMPC controller: optimality, stability and robustness," *Automatica*, vol. 45, no. 1, pp. 86–93, 2009.

[22] A. Wächter and L. T. Biegler, "On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming," *Mathematical Programming*, vol. 106, pp. 25–57, 2006.

[23] V. M. Zavala, "Stability analysis of an approximate scheme for moving horizon estimation," *Submitted for Publication*, 2009.

[24] M. Alamir and L. A. Cavillo-Corona, "Further results on nonlinear receding-horizon observers," *IEEE Transactions on Automatic Control*, vol. 47, no. 7, pp. 1184–1188, 2002.

[25] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer Verlag, New York, 1999.

[26] V. M. Zavala, *Computational Strategies for the Optimal Operation of Large-Scale Chemical Processes*. Ph.D. Thesis, Department of Chemical Engineering, Carnegie Mellon University, 2008.

[27] A. V. Fiacco, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*. New York: Academic, 1983, vol. 165.

[28] ——, "Sensitivity analysis for nonlinear programming using penalty methods," *Math. Programm.*, vol. 10, no. 1, pp. 287–311, 1976.

[29] G. A. Hicks and W. H. Ray, "Approximation methods for optimal control synthesis," *Can. J. Chem. Eng.*, vol. 49, pp. 522–529, 1971.