

Computationally Efficient, Approximate Moving Horizon State Estimation for Nonlinear Systems

A. Alessandri* M. Baglietto** G. Battistelli*** V. Zavala****

* *Department of Production Engineering, Thermoenergetics, and Mathematical Models, DIPTeM–University of Genoa, P.le Kennedy Pad. D, 16129 Genoa, Italy (e-mail: alessandri@diptem.unige.it).*

** *Department of Communications, Computer and System Sciences, DIST–University of Genoa, Via Opera Pia 13, 16145 Genoa, Italy (e-mail: mbaglietto@dist.unige.it).*

*** *Dipartimento di Sistemi e Informatica, Università degli Studi di Firenze, Via di S. Marta 3, 50139 Firenze, Italy (e-mail: battistelli@dsi.unifi.it).*

**** *Mathematics and Computer Science Division, Argonne National Laboratory, 9700 South Cass Ave, Bldg. 240, Argonne, IL 60439, USA (e-mail: vzavala@mcs.anl.gov).*

Abstract: Moving horizon estimation for discrete-time nonlinear systems is addressed by using fast optimization algorithms for which stability results under general conditions are ensured. The solution of the on-line moving horizon estimation problem is obtained by using the sampling time to solve a reference problem with model-predicted measurements while waiting for the next measurement. In order to correct the resulting solution, a quick nonlinear programming sensitivity calculation is accomplished as soon as the new measurement becomes available. The stability properties of such moving horizon estimation algorithm is proved under general conditions, which make the overall approach suitable for real settings with strong nonlinearities. Preliminary simulation results confirm the effectiveness of the proposed method.

1. INTRODUCTION

Moving horizon (MH) estimation is often associated with model predictive control (MPC) since such problems are dual, as pointed out by Goodwin et al. [2005]. An MH estimator is valuable in its own right as well as in the solution of more complicated tasks such as output feedback. In a common version of MH estimation, we are given a system description with dynamic and measurement equations and a sliding-window algorithm that estimates the state variables at each time step by using only a batch of the last measures. In performing the estimation, one may need to account also for constraints on the state variables when it is known that they belong to some subsets of the state space. Similarly, predictive control is generated by minimizing a forward cost function, thus generating a feedback control action that accounts for its future effects by possibly taking into accounts constraints on both state and input.

Ideas on what was later called MH estimation first were presented by Jazwinski [1968]. An approach to the design of asymptotic state observers was proposed by Moraal and Grizzle [1995] that result from the numerical solution of the measurement inversion problem by Newton's method. Zimmer [1994] and Alamir [1999] developed similar optimization-based techniques to construct estimators for continuous-time dynamic systems. Michalska

and Mayne [1995] proposed an MH observer for nonlinear continuous-time systems that performs estimation at discrete-time instants by approximately minimizing an integral error defined on the preceding time window. More recently, research on MH estimation has been successfully applied to linear systems (Rao et al. [2001], Alessandri et al. [2003, 2004, 2005a]), hybrid systems (Ferrari-Trecate et al. [2003], Alessandri et al. [2005b]), and nonlinear systems (Rao et al. [2003], Alessandri et al. [2008a]).

The MH estimation scheme proposed by Rao et al. [2003] 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 is provided. Unfortunately, such an 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 by Alessandri et al. [1999] with the possibility of admitting a certain error in the minimization of the cost function. Moreover, the computation required to design the resulting filter can be carried out off-line by using approximate optimal estimation functions, typically implemented via neural networks. These results were improved later by Alessandri et al. [2008a] who accounted for the simultaneous presence of system and measurement noises. In addition, the conditions that guarantee the stability of the estimation error

* Preprint ANL/MCS-P1718-0110

were relaxed and the essentially local results of Alessandri et al. [1999] were extended to regional stability.

It is well established that the minimization problem involved in MH estimation cannot be solved on-line when the dimension of the state vector is large. Therefore, there exists the need of accomplishing this task as much as possible off line Diehl et al. [2009]. A possible approach relies on the use of nonlinear approximate optimal estimation functions to be chosen off-line (see, for an introduction, the references reported in Alessandri et al. [2008a]). An alternative idea presented in this paper involves using fast optimization techniques for MH estimation recently proposed by Kraus et al. [2005], Zavala et al. [2008, 2007], Zavala and Biegler [2009]. Here, the idea is to use the sampling time to solve a reference MH problem by using model-predicted measurements while waiting for the next measurement and to correct the reference solution by performing a quick nonlinear programming (NLP) sensitivity calculation as soon as the measurement becomes available. These NLP sensitivity-based estimators are able to accommodate large-scale models in on-line environments while dramatically reducing the feedback delay of MPC controllers. The estimators presented by Zavala et al. [2008, 2007], Zavala and Biegler [2009] use the interior point NLP solver IPOPT (Wächter and Biegler [2006]), which is able to exploit the sparse structure of MH problems automatically at the linear algebra level. This provides an efficient approach to solve the reference MH problem between sampling times. A drawback of sensitivity-based estimators is that specialized sensitivity capabilities need to be implemented inside the NLP solver which is a non-trivial and often impossible task because the majority of state-of-the-art NLP solvers are proprietary. From a practical point of view, it is thus desired that fast MH estimation formulations be implemented without the need of sensitivity capabilities. Another drawback of sensitivity-based estimators is that their stability properties might deteriorate in the presence of large levels of noise as a result of increasing sensitivity errors. Therefore, other more robust formulations and a deeper understanding of their stability properties are needed.

In this work, we present an approach to MH estimation that combines the reduced computational requirements of the sensitivity-based methods presented in Zavala et al. [2008, 2007], Zavala and Biegler [2009], Zavala [2009] with the stability guarantees of the techniques developed by Alessandri et al. [2008a]. To this end, in Section 2 the problem is formulated. A general framework to find approximate solutions is described in Section 3. The stability properties of the resulting MH estimation algorithms are proved in Section 4. Preliminary simulation results are presented in Section 5.

2. PROBLEM STATEMENT AND PRELIMINARY RESULTS

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

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

$$y_t = h(x_t) + \eta_t, \quad (1b)$$

for $t = 0, 1, \dots$, 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 $\xi_t \in \mathbb{R}^n$ 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 $\eta_t \in \mathbb{R}^p$ is a measurement noise vector. We assume the statistics of x_0 , ξ_t , and η_t to be unknown, and we consider them as deterministic variables of unknown character that take their values from known compact sets.

We adopt the estimation scheme described in Alessandri et al. [2008a], which is based on an MH strategy: At any time $t = N, N + 1, \dots$, the estimate $\hat{x}_{t-N,t}$ of the state vector x_{t-N} is obtained on the basis of a *prediction* \bar{x}_{t-N} of the state x_{t-N} and of the *information vector*

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

where $N + 1$ measurements and N input vectors are collected within a “sliding window” $[t - N, t]$. The prediction \bar{x}_{t-N} is obtained from the estimate $\hat{x}_{t-N-1,t-1}$ by applying the function f , that is,

$$\bar{x}_{t-N} = f(\hat{x}_{t-N-1,t-1}, u_{t-N-1}), \quad t = N + 1, N + 2, \dots$$

The vector \bar{x}_0 denotes an a priori prediction of x_0 .

In the lines of Alessandri et al. [2008a] we make the following assumptions.

- A1.** The sets Ξ , H , and U , where ξ_t , η_t and u_t respectively take their values, are compact sets, with $0 \in \Xi$ and $0 \in H$.
- A2.** The initial state x_0 and the control sequence $\{u_t\}$ are such that, for any possible sequence of disturbances $\{\xi_t\}$, the system trajectory $\{x_t\}$ lies in a compact set X .

Since, under Assumption A2, at every time step $t = 0, 1, \dots$, the state x_t falls within the set X , the condition $\hat{x}_{t-N,t} \in X$ could be considered as a further constraint in our estimation scheme. In general, fulfilling such a constraint when applying a mathematical programming procedure is a hard task. In order to mitigate this problem, a compact convex outer-approximation of X will be considered and denoted by \mathcal{X} . Then, the constraint

$$\hat{x}_{t-N,t} \in \mathcal{X} \quad (3)$$

will be enforced.

The following assumption is also needed

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

Of course, the fulfillment of constraint (3), together with Assumption A3, ensures that the prediction \bar{x}_{t-N} belongs to the compact set $\bar{\mathcal{X}} \triangleq f(\mathcal{X}, U)$ for every $t = N, N + 1, \dots$ (the a priori prediction \bar{x}_0 is chosen inside the set $\bar{\mathcal{X}}$).

Let Y and $\mathcal{I}^{(N)}$ denote the sets wherein the vectors y_t and $I_t^{(N)}$, respectively, take their values. One can immediately see that

$$Y = h(\mathcal{X}) + H$$

and $\mathcal{I}^{(N)} \subseteq Y^{N+1} \times U^N$ (here and in the following, the set summation is intended in the sense of Minkowski). Further, a more precise characterization of the form of the set $\mathcal{I}^{(N)}$ can be given by defining the function¹

$$F^{(N)}(x_{t-N}, u_{t-N}^{t-1}, \xi_{t-N}^{t-1}) \triangleq \begin{bmatrix} h(x_{t-N}) \\ h \circ f^{u_{t-N}, \xi_{t-N}}(x_{t-N}) \\ \vdots \\ h \circ f^{u_{t-1}, \xi_{t-1}} \circ \dots \circ f^{u_{t-N}, \xi_{t-N}}(x_{t-N}) \end{bmatrix},$$

where “ \circ ” denotes function composition and $f^{u_i, \xi_i}(x_i) \triangleq f(x_i, u_i) + \xi_i$. In fact, by exploiting such a definition one can write

$$y_{t-N}^t = F^{(N)}(x_{t-N}, u_{t-N}^{t-1}, \xi_{t-N}^{t-1}) + \eta_{t-N}^t.$$

This, in turn, yields

$$\mathcal{I}^{(N)} = \{I = \text{col}(y, u) : u \in U^N, y \in F^{(N)}(\mathcal{X}, u, \Xi^N) + H^{N+1}\}.$$

Following a least-squares approach, in Alessandri et al. [2008a] the minimization of the following cost function² is addressed:

$$\begin{aligned} J^{(N)}(\hat{x}_{t-N,t}, \bar{x}_{t-N}, I_t^{(N)}) \\ &= \mu \|\hat{x}_{t-N,t} - \bar{x}_{t-N}\|^2 + \sum_{i=t-N}^t \|y_i - h(\hat{x}_{i,t})\|^2 \\ &= \mu \|\hat{x}_{t-N,t} - \bar{x}_{t-N}\|^2 + \left\| y_{t-N}^t - F^{(N)}(\hat{x}_{t-N,t}, u_{t-N}^{t-1}, 0) \right\|^2 \end{aligned} \quad (4)$$

where the estimates $\hat{x}_{t-N+1,t}, \dots, \hat{x}_{t,t}$ are generated here by $\hat{x}_{t-N,t}$ through the noise-free dynamics, that is,

$$\hat{x}_{i+1,t} = f(\hat{x}_{i,t}, u_i), \quad i = t-N, \dots, t-1. \quad (5)$$

The last equality in (4) follows from recursive application of (5). The positive scalar μ expresses our belief in the prediction \bar{x}_{t-N} with respect to the observation model.

The following algorithm can be stated.

Optimal MH estimator (Algorithm E^o). Given an a priori prediction \bar{x}_0 , at any time $t = N, N+1, \dots$

- (1) find a solution $\hat{x}_{t-N|t}^\circ$ to the NLP problem

$$\begin{aligned} &\text{minimize } J^{(N)}(\hat{x}_{t-N|t}, \bar{x}_{t-N}, I_t^{(N)}) \\ &\text{subject to } \hat{x}_{t-N|t} \in \mathcal{X} \end{aligned} \quad (6)$$

- (2) compute the optimal estimate sequence

$$\hat{x}_{i+1|t}^\circ \triangleq f(\hat{x}_{i|t}^\circ, u_i) \quad \text{for } i = t-N, t-N+1, \dots, t-1$$

and extract the estimate $\hat{x}_{t|t}^\circ$ of the current state;

¹ In general, given a sequence $\{z_t\}$, we shall use the notation $z_t^{t+t'} \triangleq \text{col}(z_t, \dots, z_{t+t'})$.

² Throughout the paper, for a generic vector v , we denote its Euclidean norm by $\|v\|$. For a generic matrix M , we shall denote its induced Euclidean norm by $\|M\|$.

(3) set

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

3. APPROXIMATE SOLUTION

In fact, the minimization involved in Algorithm E^o cannot be completed instantaneously. This would induce a delay in the estimation process. One possibility for facing such a drawback consists in the use of parametrized functions (to be optimized off-line) to approximate the optimal estimation function obtained by applying algorithm E^o. Such a possibility was considered by Alessandri et al. [1999] and, more recently, by Alessandri et al. [2008b]. Here a different approach is adopted.

To derive an approximate MHE algorithm providing an estimate of the continuous state almost in real time, one can proceed as follows. Suppose that at each time t , the most recent available optimal estimate is $\hat{x}_{t-N-1|t-1}^\circ$ (of course, this amounts to assuming that problem (6) can be solved to the desired accuracy in less than one sample time). Then, one can exploit the following fact.

Proposition 1. Let $\hat{x}_{t-N-1|t-1}^\circ$ be a solution of problem (6) at time $t-1$. Then, $\hat{x}_{t-N-1|t-1}^\circ$ is also a solution of the NLP problem

$$\begin{aligned} &\text{minimize } J^{(N+1)}(\hat{x}_{t-N-1|t-1}, \bar{x}_{t-N-1}, \bar{I}_t^{(N+1)}) \\ &\text{subject to } \hat{x}_{t-N-1|t-1} \in \mathcal{X}, \end{aligned} \quad (7)$$

where

$$\begin{aligned} \bar{I}_t^{(N+1)} &\triangleq \text{col}(u_{t-N-1}^{t-1}, y_{t-N-1}^{t-1}, \bar{y}_t), \\ \bar{y}_t &\triangleq h \circ f(\hat{x}_{t-1|t-1}^\circ, u_{t-1}). \end{aligned}$$

□

In light of Proposition 1, one can construct a linear update formula of the form

$$\hat{x}_{t-N-1|t}^\circ = \hat{x}_{t-N-1|t-1}^\circ + K_t(y_t - \bar{y}_t). \quad (8)$$

To this end, suppose that the optimal estimate $\hat{x}_{t-N-1|t-1}^\circ$ corresponds to a strict *isolated* minimizer of $J^{(N+1)}(\hat{x}_{t-N-1|t-1}, \bar{x}_{t-N-1}, \bar{I}_t^{(N+1)})$. As is well known, a strict isolated minimizer satisfies the so-called strong second-order sufficient conditions (SSOC) (see Nocedal and Wright [1999]), which we present here in the context of MHE.

Lemma 1. (SSOC Conditions) Let $J^{(N+1)}(\hat{x}, \bar{x}, \bar{I})$ be a C^2 function w.r.t. \hat{x} in a neighborhood of \hat{x}° , with \hat{x}° not lying at the boundary of the set \mathcal{X} . If $\nabla_{\hat{x}} J^{(N+1)}(\hat{x}^\circ, \bar{x}, \bar{I}) = 0$ and $\nabla_{z,z} J^{(N+1)}(\hat{x}^\circ, \bar{x}, \bar{I})$ is positive definite, then \hat{x}° is a strict isolated minimizer.

The requirement of $J^{(N+1)}(\hat{x}, \bar{x}, \bar{I})$ being a C^2 function follows from Assumption A3. A detailed SSOC analysis in the context of MHE can be found in Chapters 3 and 6 in Zavala [2008].

The satisfaction of SSOC also has implications on the sensitivity of the solution to perturbations on the problem data \bar{I} around a reference solution $\hat{x}^\circ(\bar{I})$ (Basu and Bresler [2000b]). To explore this, we use the following well-known result, adapted to our context.

Theorem 1. (NLP Sensitivity) Fiacco [1983, 1976]. If a nominal solution $\hat{x}^\circ(\bar{I}')$ satisfies SSOC, then the following hold:

- For \bar{I} in a neighborhood of \bar{I}' there exists a unique, continuous, and differentiable vector function $\hat{x}^\circ(\bar{I})$ that is a strict isolated minimizer satisfying SSOC.
- The optimal cost is locally Lipschitz in a neighborhood of \bar{I}' .

From these results, provided that SSOC holds at $\hat{x}_{t-N-1|t-1}^\circ$ for the cost $J^{(N+1)}(x_{t-N-1|t-1}, \bar{x}_{t-N-1}, \bar{I}_t^{(N+1)})$, the gain K_t can be chosen on the basis of a first-order Taylor expansion. For instance, exploiting the implicit function theorem, one has

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

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

$$J^{(N+1)}(x_{t-N-1|t}^\epsilon, \bar{x}_{t-N-1}, \bar{I}_t^{(N+1)}) \\ \leq J^{(N+1)}(x_{t-N-1|t-1}^\circ, \bar{x}_{t-N-1}, \bar{I}_t^{(N+1)}). \quad (9)$$

The foregoing discussion leads to the following approximate MHE algorithm.

Approximate MH estimator (Algorithm E^ε). Given an a priori prediction \bar{x}_0 , at any time $t = N, N + 1, \dots$

- (1) if $t = N$, then go to Step 6; else go to Step 2;
- (2) if SSOC holds at $\hat{x}_{t-N-1|t-1}^\circ$, then compute the updated estimate $\hat{x}_{t-N-1|t}^\circ$ as in (8) and go to Step 3; else set $\hat{x}_{t-N-1|t}^+ = \hat{x}_{t-N-1|t-1}^\circ$ and go to Step 4;
- (3) if inequality (9) holds, then set $\hat{x}_{t-N-1|t}^\epsilon = \hat{x}_{t-N-1|t}^\circ$; else set $\hat{x}_{t-N-1|t}^+ = \hat{x}_{t-N-1|t-1}^\circ$;

- (4) compute the approximate estimate sequence

$$\hat{x}_{i+1|t}^+ \triangleq f(\hat{x}_{i|t}^+, u_i) \quad \text{for } i = t - N - 1, t - N, \dots, t - 1$$

and extract the estimate $\hat{x}_{t|t}^+$ of the current state;

- (5) set

$$\bar{x}_{t-N} = f(\hat{x}_{t-N-1|t-1}^\circ, u_{t-N-1}) \quad (10)$$

- (6) find a solution $\hat{x}_{t-N|t}^\circ$ to the NLP problem (6).

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

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, in the next section we will show that stability of the estimation error $x_t - \hat{x}_{t|t}^+$ is ensured under the same assumptions of the optimal MHE algorithm.

Remark 3. Notice that in Step 5, the prediction is computed from the optimal estimate $\hat{x}_{t-N-1|t-1}^\circ$ instead that from $\hat{x}_{t-N-1|t}^+$. Of course, since $\hat{x}_{t-N-1|t}^+$ is believed to be a better estimate than $\hat{x}_{t-N-1|t-1}^\circ$, it would be natural to replace (10) with

$$\bar{x}_{t-N} = f(\hat{x}_{t-N-1|t}^+, u_{t-N-1}). \quad (11)$$

For such a modified algorithm, stability can still be ensured, but the condition on the parameter μ becomes more stringent (see Remark 6).

In the following, the approximate MH estimation algorithm obtained by replacing (10) with (11) will be denoted as Algorithm E⁺.

4. STABILITY OF THE ESTIMATION ERROR

In order to study the stability properties of the proposed approximate MHE algorithms, the following observability assumption is made.

A4. System (1) is \mathcal{X} -observable in $N + 1$ steps. That is, there exists a K -function³ $\varphi(\cdot)$, such that

$$\varphi(\|x_1 - x_2\|^2) \leq \left\| F^{(N)}(x_1, u, 0) - F^{(N)}(x_2, u, 0) \right\|^2, \quad (12)$$

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

The observability definition expressed by (12) has been widely used in the framework of nonlinear state estimation in both the discrete-time and the continuous-time settings (see, for example, Moraal and Grizzle [1995], Alamir [1999], Alamir and Cavillo-Corona [2002], Rao et al. [2003], Alessandri et al. [2008a]). Recently, Hanba [2009] has shown that such a definition is quite general in that taking

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

into account observation windows of fixed length is not restrictive. We point out that (12) implies that

$$\varphi(\|x_1 - x_2\|^2) \leq \left\| F^{(M)}(x_1, u, 0) - F^{(M)}(x_2, u, 0) \right\|^2 \quad (13)$$

also holds for any observation window length M greater than N .

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 k_h an upper bound on the Lipschitz constant of $h(x)$. Then the following preliminary result can be stated.

Lemma 2. Suppose that Assumptions A1, A2, A3, and A4 are satisfied. Moreover, suppose that the K -function φ , defined in Assumption A4, satisfies the following condition

$$\delta \triangleq \inf_{x_1, x_2 \in \mathcal{X}; x_1 \neq x_2} \frac{\varphi(\|x_1 - x_2\|^2)}{\|x_1 - x_2\|^2} > 0. \quad (14)$$

Further, let $\hat{x}_{t-N-1|t-1}^\circ$, $\hat{x}_{t-N-1|t}^+$, and $\hat{x}_{t|t}^+$ be computed as in Algorithm E $^\epsilon$ or E $^+$ from a prediction \bar{x}_{t-N-1} .

Then, at any time $t = N + 1, N + 2, \dots$ the following inequalities hold:

$$\begin{aligned} & \left\| x_{t-N-1} - \hat{x}_{t-N-1|t-1}^\circ \right\|^2 \\ & \leq \alpha^\circ(\mu) \|x_{t-N-1} - \bar{x}_{t-N-1}\|^2 + \beta^\circ \end{aligned} \quad (15)$$

$$\begin{aligned} & \left\| x_{t-N-1} - \hat{x}_{t-N-1|t}^+ \right\|^2 \\ & \leq \alpha^+(\mu) \|x_{t-N-1} - \bar{x}_{t-N-1}\|^2 + \beta^+ \end{aligned} \quad (16)$$

$$\left\| x_t - \hat{x}_{t|t}^+ \right\|^2 \leq \kappa \left\| x_{t-N-1} - \hat{x}_{t-N-1|t}^+ \right\|^2 + \gamma, \quad (17)$$

where

$$\alpha^\circ(\mu) \triangleq \frac{4\mu}{\mu + \delta}, \quad \alpha^+(\mu) \triangleq \frac{4\mu}{(\mu + \delta)^2} [4k_h^2 \kappa + \delta + \mu],$$

$$\kappa = (2k_f^2)^{N+1},$$

and $\beta^\circ, \beta^+, \gamma$ are suitable scalars. \square

Inequality (15) can be derived following the lines of Theorem 1 of Alessandri et al. [2008a]. Inequality (16) can be obtained with similar arguments by noting that, thanks to the test (9) on the updated cost, one has

$$\begin{aligned} & J^{(N+1)}(\hat{x}_{t-N-1|t}^+, \bar{x}_{t-N-1}, I_t^{(N+1)}) \\ & \leq J^{(N+1)}(\hat{x}_{t-N-1|t-1}^\circ, \bar{x}_{t-N-1}, I_t^{(N+1)}) \\ & = J^{(N)}(\hat{x}_{t-N-1|t-1}^\circ, \bar{x}_{t-N-1}, I_{t-1}^{(N)}) + \|y_t - \bar{y}_t\|^2. \end{aligned}$$

Inequality (17) descends directly from repeated application of the Lipschitz inequality.

Remark 4. Since the set \mathcal{X} is compact, condition (14) turns out to be equivalent to

$$\lim_{\|x_1 - x_2\| \rightarrow 0^+} \frac{\varphi(\|x_1 - x_2\|)}{\|x_1 - x_2\|} > 0.$$

Recalling that the K -function φ has to satisfy inequality (12), one immediately sees that the fulfillment of condition

(14) ensures that small variations in the observation vector $F^{(N)}(x_t, u_{t-N}^{t-1}, 0)$ always correspond to small variations of the state vector x_t . Such a requirement is typical in the nonlinear programming literature, where a nonlinear least-squares problem of the form

$$\min_{\hat{x}} \left\| y_{t-N}^t - F^{(N)}(\hat{x}, u_{t-N}^{t-1}, 0) \right\|^2 \quad (18)$$

is said to be *stable* if the mapping from the observations vector y_{t-N}^t to the global minimum of the cost function is sufficiently smooth (see Fiacco [1983], Basu and Bresler [2000a]). Basu and Bresler [2000a] show that a sufficient condition for the stability of the least-squares problem (18) is that the gradient matrix $\partial F^{(N)}(x, u_{t-N}^{t-1}, 0)/\partial x$ has full column rank.

Remark 5. We point out that the additive constants β°, β^+ , and γ depend continuously on the ‘‘amplitudes’’ of the noises (i.e., on the radii of the sets Ξ and H). Moreover, if system (1) is noise-free (i.e., $\Xi = \{0\}$ and $H = \{0\}$), then such constants turn out to be zero.

In view of Lemma 2, the stability of the proposed approximate MHE algorithms can be analyzed. To this end, noting that for Algorithm E $^\epsilon$ the prediction update (10) yields

$$\begin{aligned} & \left\| x_{t-N-1} - \bar{x}_{t-N-1} \right\|^2 \\ & \leq 2k_f^2 \left\| x_{t-N-2} - \hat{x}_{t-N-2|t-2}^\circ \right\|^2 + 2r_\xi^2 \end{aligned}$$

with $r_\xi = \sup_{\xi \in \Xi} \|\xi\|$, the following stability result can be stated.

Theorem 2. Suppose that Assumptions A1, A2, A3, and A4 are satisfied. Moreover suppose that condition (14) is satisfied. Further let $\hat{x}_{t|t}^+$ be computed recursively according to Algorithm E $^\epsilon$.

Then, the estimation error is bounded as

$$\left\| x_t - \hat{x}_{t|t}^+ \right\|^2 \leq 2k_f^2 \kappa \alpha^+(\mu) \zeta_{t-N-2}^\epsilon + 2\kappa \alpha^+(\mu) r_\xi^2 + \kappa \beta^+ + \gamma$$

where $\{\zeta_t^\epsilon\}$ is a sequence generated recursively by

$$\zeta_{t+1}^\epsilon = 2k_f^2 \alpha^\circ(\mu) \zeta_t^\epsilon + \beta^\circ + 2r_\xi^2, \quad t = 0, 1, \dots$$

$$\zeta_0^\epsilon = \left\| x_0 - \hat{x}_{0|N}^\circ \right\|^2.$$

Moreover, if μ is selected such that

$$2k_f^2 \alpha^\circ(\mu) < 1, \quad (19)$$

the bounding sequence $\{\zeta_t^\epsilon\}$ converges exponentially to the asymptotic value

$$(\beta^\circ + 2r_\xi^2)/(1 - 2k_f^2 \alpha^\circ(\mu)). \quad (20)$$

\square

Similarly, in the case of Algorithm E $^+$, since the prediction update (11) implies

$$\begin{aligned} & \left\| x_{t-N-1} - \bar{x}_{t-N-1} \right\|^2 \\ & \leq 2k_f^2 \left\| x_{t-N-2} - \hat{x}_{t-N-2|t-1}^+ \right\|^2 + 2r_\xi^2, \end{aligned}$$

the following theorem follows at once.

Theorem 3. Suppose that assumptions A1, A2, A3, and A4 are satisfied. Moreover suppose that condition (14) is satisfied. Further let $\hat{x}_{t|t}^+$ be computed recursively according to Algorithm E⁺.

Then, the estimation error is bounded as

$$\left\|x_t - \hat{x}_{t|t}^+\right\|^2 \leq \kappa \zeta_{t-N-1}^+ + \gamma$$

where $\{\zeta_t^+\}$ is a sequence generated recursively by

$$\begin{aligned} \zeta_{t+1}^\epsilon &= 2k_f^2 \alpha^+(\mu) \zeta_t^\epsilon + \beta^+ + 2r_\xi^2, \quad t = 0, 1, \dots \\ \zeta_0^\epsilon &= \left\|x_0 - \hat{x}_{0|N+1}^+\right\|^2. \end{aligned}$$

Moreover, if μ is selected such that

$$2k_f^2 \alpha^+(\mu) < 1, \quad (21)$$

the bounding sequence $\{\zeta_t^+\}$ converges exponentially to the asymptotic value

$$(\beta^+ + 2r_\xi^2)/(1 - 2k_f^2 \alpha^+(\mu)). \quad (22)$$

□

Some comments on Theorems 2 and 3 are in order.

Remark 6. Note that condition (19) can be easily satisfied for any value of k_f by imposing that the positive weight μ does not exceed a certain stability threshold μ_{\max}^ϵ . For instance, if $8k_f^2 \leq 1$, one can choose any $\mu \geq 0$ (i.e., $\mu_{\max}^\epsilon = +\infty$). Instead, when $8k_f^2 > 1$, (19) one has $\mu_{\max}^\epsilon = \delta/(8k_f^2 - 1)$. Thus, 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 μ that satisfy condition (19).

Similarly, in the case of Algorithm E⁺ and for any value of k_f there exists a *strictly* positive stability threshold μ_{\max}^+ such that condition (21) is satisfied whenever $0 \leq \mu < \mu_{\max}^+$. In this case, however, the condition on the design parameter μ becomes more stringent in that $0 < \mu_{\max}^+ \leq \mu_{\max}^\epsilon$.

In this connection, one can still adopt the improved prediction step (11) without reducing the range of feasible μ 's by modifying the test in Step 3 as follows:

$$\begin{aligned} &J^{(N+1)}(x_{t-N-1|t}^\epsilon, \bar{x}_{t-N-1}, I_t^{(N+1)}) \\ &\leq J^{(N+1)}(x_{t-N-1|t-1}^\circ, \bar{x}_{t-N-1}, I_t^{(N+1)}) - \|y_t - \bar{y}_t\|^2 + \nu, \end{aligned}$$

where ν is some positive real. One can see that the further the true measurement y_t is from the predicted on \bar{y}_t , the more difficult is to satisfy such a condition. This situation agrees with the fact that the first-order Taylor expansion yielding (8) is a good approximation of the optimal estimation function only for small values of $y_t - \bar{y}_t$.

Remark 7. In view of the considerations of Remark 5, one immediately sees that the asymptotic bounds in (20) and (22) are equal to zero when system (1) is noise-free (i.e., $\Xi = \{0\}$ and $H = \{0\}$). Thus, in this case, both the proposed approximate MHE estimators are exponential observers, provided that the stability conditions (19) and, respectively, (21) are satisfied.

5. NUMERICAL CASE STUDY

In this section, we illustrate the effect of numerical errors on the performance of the approximate MH estimators described in this paper and discuss some of the stability properties developed in the previous sections. We consider a simulated MHE scenario on the nonlinear continuous stirred tank reactor studied by Hicks and Ray [1971]:

$$\begin{aligned} \frac{dx^1}{d\tau} &= \frac{x^1(\tau) - 1}{\theta} + k_0 \cdot x^1(\tau) \cdot \exp\left[\frac{-E_a}{x^2(\tau)}\right] \\ \frac{dx^2}{d\tau} &= \frac{x^2(\tau) - x_f^2}{\theta} - k_0 \cdot x^1(\tau) \cdot \exp\left[\frac{-E_a}{x^2(\tau)}\right] \\ &\quad + \alpha \cdot u(\tau) \cdot (x^2(\tau) - x_{cw}^2). \end{aligned}$$

The system involves two states: $x = [x^1, x^2]$ corresponding to the concentration and temperature, respectively, and one control u corresponding to the cooling water flowrate. 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 . The model parameters are $x_{cw}^2 = 0.38$, $x_f^2 = 0.395$, $E_a = 5$, $\alpha = 1.95 \times 10^4$, and $k_0 = 300$. We use batch data generated from a simulated closed-loop feedback control scenario. The simulated temperatures are corrupted with different levels of Gaussian noise with standard deviation σ . The corrupted temperature values are used to study the effect of noise disturbances η_t . We use $\bar{x}_0 = [0.15 \ 0.15]$ as the initial reference state and a regularization penalty $\mu = 30$. The estimators are simulated over 250 time steps.

In Figure 1 we contrast the predicted \bar{y}_t and true measurements y_t corresponding to the temperature x_t^2 . These generate the perturbation $\|y_t - \bar{y}_t\|^2$ for the approximate estimators. The noise standard deviation is set to $\sigma = 0.0125$. In Figure 2 we compare the performance of Algorithms E^o, E^ε, and E⁺. The performance is shown for a horizon $N = 1$ (top graph) and $N = 10$ (bottom graph). Note that both approximate estimators remain stable despite the large level of noise and the short horizons. Note also that we have introduced a large error in the initial state estimate that leads to large deviations during the first 50 time steps. Both estimators eventually converge to the true states. The performance of Algorithm E⁺ is close to that of Algorithm E^o. In other words, the sensitivity error is negligible even for short horizons. Also interesting is the fact that, as the horizon is increased, the performance of Algorithm E^ε converges to that of Algorithm E^o. The reason is that, as the horizon is increased, the effect of the most current measurement on the state estimate becomes meaningless. This situation can be easily explained from NLP sensitivity properties and the structure of MHE problems Zavala et al. [2008]. In Figure 3 we analyze the effect of the estimation horizon and noise level on the performance of Algorithm E^ε. Here, we compare the sum of squared errors (SSE) of the state estimates x_t^1 generated with the approximate and the optimal estimators. The SSE is defined as $\sum_{t=1}^{250} (x_t^1 - x_t^{1*})^2$. As previously observed, the effect of noise is appreciable at short horizons but this effect dies out as the horizon is increased.

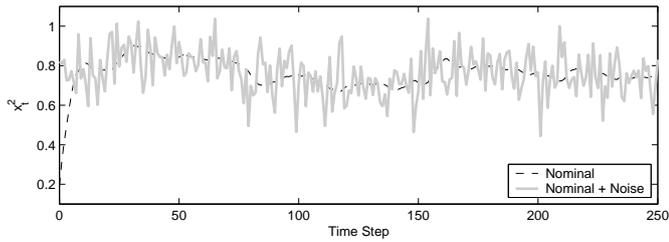


Fig. 1. Nominal and noisy output measurement.

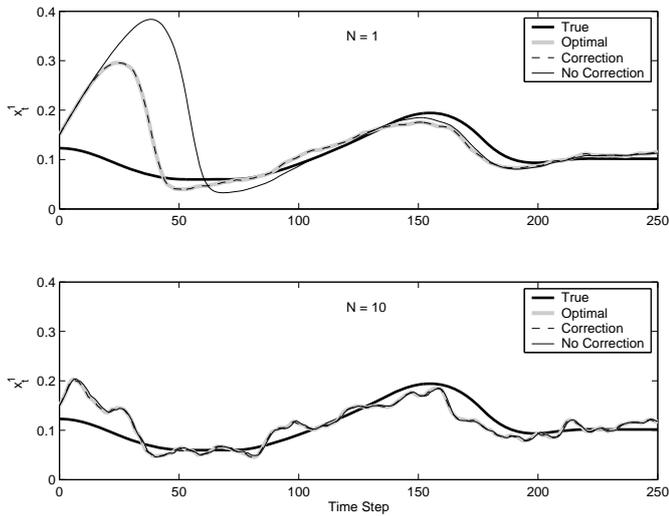


Fig. 2. Performance of Algorithms E^o (grey line), Algorithm E^ϵ (solid line), and Algorithm E^+ (dashed line).

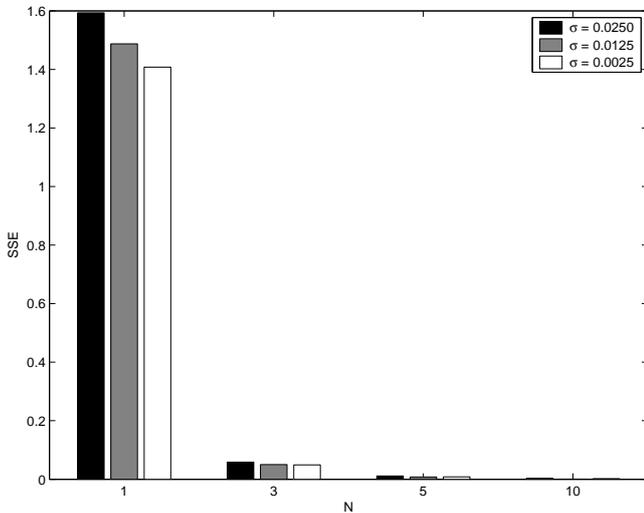


Fig. 3. Effect of noise level and horizon length on the performance of Algorithm E^ϵ .

ACKNOWLEDGMENTS

Victor M. Zavala was supported by the Department of Energy, through Contract No. DE-AC02-06CH11357.

REFERENCES

M. Alamir. Optimization based non-linear observers revisited. *International Journal of Control*, 72(13):1204–1217, 1999.

- M. Alamir and L. A. Cavillo-Corona. Further results on nonlinear receding-horizon observers. *IEEE Transactions on Automatic Control*, 47(7):1184–1188, 2002.
- A. Alessandri, M. Baglietto, T. Parisini, and R. Zoppoli. A neural state estimator with bounded errors for nonlinear systems. *IEEE Transactions on Automatic Control*, 44(11):2028–2042, 1999.
- A. Alessandri, M. Baglietto, and G. Battistelli. Receding-horizon estimation for discrete-time linear systems. *IEEE Transactions on Automatic Control*, 48(3):473–478, 2003.
- A. Alessandri, M. Baglietto, and G. Battistelli. On estimation error bounds for receding-horizon filters using quadratic boundedness. *IEEE Transactions on Automatic Control*, 49(8):1350–1355, 2004.
- A. Alessandri, M. Baglietto, and G. Battistelli. Robust receding-horizon state estimation for uncertain discrete-time linear systems. *Systems & Control Letters*, 54(7):627–643, 2005a.
- A. Alessandri, M. Baglietto, and G. Battistelli. Receding-horizon estimation for switching discrete-time linear systems. *IEEE Transactions on Automatic Control*, 50(11):1736–1748, 2005b.
- A. Alessandri, M. Baglietto, and G. Battistelli. Moving-horizon state estimation for nonlinear discrete-time systems: New stability results and approximation schemes. *Automatica*, 44(7):1753–1765, 2008a.
- 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*, pages 2557–2562, Cancun, Mexico, 2008b.
- S. Basu and Y. Bresler. Stability of nonlinear least squares problems and the Cramer-Rao bound. *IEEE Transactions on Signal Processing*, 48(12):3426–3436, 2000a.
- S. Basu and Y. Bresler. The stability of nonlinear least squares problems and the cramer-rao bound. *IEEE Trans. Signal Proc.*, 48(12):3427–3436, 2000b.
- M. Diehl, H. J. Ferreau, and N. Haverbeke. Efficient numerical methods for nonlinear mpc and moving horizon estimation. In *Nonlinear Model Predictive Control*, pages 391–417, 2009.
- G. Ferrari-Trecate, D. Mignone, and M. Morari. Moving-horizon estimation for hybrid systems. *IEEE Transactions on Automatic Control*, 47(10):1663–1676, 2003.
- A. V. Fiacco. Sensitivity analysis for nonlinear programming using penalty methods. *Math. Programm.*, 10(1):287–311, 1976.
- A. V. Fiacco. *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, volume 165. New York: Academic, 1983.
- G. C. Goodwin, J. A. De Doná, M. M. Seron, and X. W. Zhuo. Lagrangian duality between constrained estimation and control. *Automatica*, 41:935–944, 2005.
- S. Hanba. On the uniform observability of discrete-time nonlinear systems. *IEEE Transactions on Automatic Control*, 54(8):1925–1928, 2009.
- G. A. Hicks and W. H. Ray. Approximation methods for optimal control synthesis. *Can. J. Chem. Eng.*, 49:522–529, 1971.
- A. H. Jazwinski. Limited memory optimal filtering. *IEEE Transactions on Automatic Control*, 13(5):558–563, 1968.

- 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*, volume 44, pages 2541–2460, 2005.
- H. Michalska and D. Q. Mayne. Moving horizon observers and observer-based control. *IEEE Transactions on Automatic Control*, 6(6):995–1006, 1995.
- P. E. Moraal and J. W. Grizzle. Observer design for nonlinear systems with discrete-time measurements. *IEEE Transactions on Automatic Control*, 40(3):395–404, 1995.
- J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Verlag, New York, 1999.
- C. V. Rao, J. B. Rawlings, and J. H. Lee. Constrained linear estimation—a moving horizon approach. *Automatica*, 37(10):1619–1628, 2001.
- 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*, 48(2):246–257, 2003.
- 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*, 106:25–57, 2006.
- 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.
- V. M. Zavala. Stability analysis of an approximate scheme for moving horizon estimation. *Argonne National Laboratory, Preprint ANL/MCS-P1644-0609*, 2009.
- V.M. Zavala and L.T. Biegler. The advanced-step NMPC controller: Optimality, stability and robustness. *Automatica*, 45(1):86–93, 2009.
- V.M. Zavala, C.D. Laird, and L.T. Biegler. 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.
- V.M. Zavala, C.D. Laird, and L.T. Biegler. A fast moving horizon estimation algorithm based on nonlinear programming sensitivity. *Journal of Process Control*, 18: 876–884, 2008.
- G. Zimmer. State observation by on-line minimization. *International Journal of Control*, 60(4):595–606, 1994.

The submitted manuscript has been created by the University of Chicago as Operator of Argonne National Laboratory (“Argonne”) under Contract No. DE-AC02-06CH11357 with the U.S. Department of Energy. The U.S. Government retains for itself, and others acting on its behalf, a paid-up, nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.