A New Identification Framework For Off-line Computation of Moving-Horizon Observers

Mazen Alamir

Abstract—In this paper, a new nonlinear identification framework is proposed to address the issue of off-line computation of moving-horizon observer estimate. The proposed structure merges the advantages of nonlinear approximators with the efficient computation of constrained quadratic programming problems. A bound on the estimation error is proposed and the efficiency of the resulting scheme is illustrated using two state estimation examples.

Index Terms—Nonlinear Moving-Horizon Observers, Nonlinear Identification, Off-Line Computation, State Estimation, Ecoli.

I. INTRODUCTION

State estimation is a key issue in nonlinear systems control and diagnosis. Algorithms that achieve this task are called observers. These algorithms attempt to reconstruct the evolution of the state vector by using the only measured (and generally noisy) quantities. As far as nonlinear systems are concerned, many observation techniques have been developed during the last four decades. This includes high-gain observers [7], sliding-modes observers [15], Moving-Horizon Observers (MHOs) [11] and naturally, the widely used Extended-Kalman-Filter (EKF). Excellent reviews of nonlinear observer design techniques can be found in [14] and [6].

Amongst all possible observer design alternatives, MHO technique has witnessed an increasing interest these last years because of its ability to handle constraints and to fully exploit precise and generally nonlinear models of the dynamic processes under study. This observer requires on-line solution of non convex optimization problems in which the cost function is the integral output prediction error while the decision variable is the set of unknown quantities to be recovered (state and unknown parameter vectors).

Despite encouraging recent advances on the issue of real-time computation of MHOs (see [3], [5] and the references therein for a recent survey on the topics), the highly demanding on-line computation involved may question the feasibility of the algorithm when system needing high sampling rates are involved or when the use of highly involved optimization software is prohibited by real-life context. When the combination of such obstacles and the absence of mathematical structure that renders impossible the use of analytical observers arises, something has to be done in order to achieve the estimation task.

M. Alamir is with CNRS, Gipsa-lab, Control Systems Department, University of Grenoble. 11 rue des Mathématiques, Domaine Universitaire, 38402 Saint Martin d'Hres, France. Email:mazen.alamir@grenoble-inp.fr. http://www.mazenalamir.fr

The ideas proposed in this paper follow a suggestion made by [1] aiming at identifying off-line the relationship between the sequence of (input/output) measurements and the corresponding initial state at the beginning of the moving observation window. By doing so, the cumbersome on-line optimization step involved in Moving-Horizon Observers [11] (MHOs) can be avoided. Concrete implementation of such an idea using Neural Networks (NNs) identification structure has been proposed in [2]. As pointed out in [1], NNs structures, like all standard nonlinear approximators offer high approximation capabilities at the price of non convex optimization schemes which generally suffer from convergence and computation time issues. In the present paper, a novel nonlinear identification scheme is proposed, which after a suitable change in the decision variables, can be solved using constrained quadratic programming. Such a feature is crucial if this optimization problem lies in the inner loop of a whole static game optimization formulation (such as the one proposed in [1]) that may be needed to assess the convergence of the resulting MHO. More precisely, the solution of the static game is needed to compute the set of approximator parameters that achieves sufficiently small maximum error over all possible initial states (and initial estimation error). The price to pay in order to gain computational efficiency is a theoretical restriction of the class of situations that can be addressed as the proposed identification structure possesses no universal approximation property. For this reason, the comparison between the two identification schemes is problem-dependent.

It is important to underline that in the scheme proposed in [1], [2] a set of non convex optimization problems are solved off-line (using different initial states and different initial state estimation errors) in order to built the learning data for the identification step. Instead, the scheme of the present paper avoids the use of such non convex optimization by focusing on the model-based state/output relationship that can be discovered using only system simulation to provide the learning data followed by an identification step using specific class of nonlinear structures. This so identified function gives the output-related guess of the state which can then be amended by a model related term in a rather trivial way in order to recover a kind of classical measurement/model confidence trade-off. By doing so, one can avoid the risk of local minima that may corrupt the quality of the learning data.

The paper is organized as follows. First, section II defines the state estimation-related identification problem. Section III describes the proposed nonlinear identification setting

and gives a bound on the state estimation error. The way the learning data used in the identification scheme is built is shown in section IV. Illustrative examples are given in section V while section VI summarizes the contribution of the paper and suggests hints for further investigations.

II. STATE ESTIMATION AS AN IDENTIFICATION PROBLEM

Let us consider a nonlinear system given by:

$$x(k + 1) = f(x(k), u(k)) \quad ; \quad y(k) = h(x(k), u(k)) \quad (1)$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^{n_u}$ and $y \in \mathbb{R}^{n_y}$ represent the state, the measured input and the measured output vectors respectively. The integer k refers to the sampling instant $k\tau$ for some sampling period $\tau > 0$. Regardless of the algorithm that may be used to reconstruct the state vector using the measured quantities, the implicit assumption that underlines the possibility of state reconstruction is that there is an integer $N \in \mathbb{N}$ and a map $\mathcal{F}$ such that the following approximation holds:

$$x(k) \approx \mathcal{F}(Z(k)) \quad (2)$$

where $Z(k)$ is the regressor built up with the past measured quantities according to:

$$Z(k) := \begin{pmatrix} \tilde{y}(k) \\ \tilde{u}(k) \end{pmatrix} \in \mathbb{R}^{N(n_u + n_y) =: n_z} \quad (3)$$

where: $\tilde{y}(k) := (y^T(k), \cdots, y^T(k - N + 1))^T$ and $\tilde{u}(k) := (u^T(k), \cdots, u^T(k - N + 1))^T$. Note that in the absence of measurement noise, the implicit definition of the map $\mathcal{F}$ involved in (2) is given by the solution of the following optimization problem:

$$\mathcal{F}(Z(k)) := X(k, k - N + 1, x^*, \tilde{u}(k)) \quad (4)$$
$$x^* := \arg \min_x J(x, Z(k)) :=$$
$$\sum_{i=0}^{N-1} \|y(k - i) - Y(k - i, k - N + 1, x, \tilde{u}(k))\|^2 \quad (5)$$

where $X(j, k - N + 1, x, \tilde{u}(k))$ [resp. $Y(j, k - N + 1, x, \tilde{u}(k))$] refers to the model-based predicted value at instant j of the state [resp. output] when the state at instant $k - N + 1$ is equal to x and when the sequence of controls defined by $\tilde{u}(k)$ is applied on the time interval $[k - N + 1, k]$.

It results that if one obtains through off-line computations a good approximation of the map $\mathcal{F}$ involved in (2), then the measurement-related part of the MHOs would be obtained without on-line optimization. This is obviously a static identification problem. More precisely, since this identification is to be obtained for each component of the state vector one needs to reconstruct, the basic generic problem one has to solve is the one consisting of finding a nonlinear map F that links a scalar quantity r (a component of the state vector) to a regression vector Z, namely:

$$r \approx F(Z) \quad ; \quad r \in \mathbb{R} \quad ; \quad Z \in \mathbb{R}^{n_z} \quad (6)$$

Remark 1: Note that, as suggested in [1] the cost function involved in (5) may contain a regularization term $\|x - \bar{x}\|^2$

where $\bar{x}$ stands for the predicted value of the decision variable based on the state space model and the last estimate. This enables a regularization of the estimation process and enhance more stability of the iterates. Such regularization can be decoupled from the identification process by introducing afterward the following final estimation which represents a trade-off between the measurement related estimation and the model related estimation based on the previous solution:

$$\hat{x} = \lambda \mathcal{F}(Z) + (1 - \lambda)\bar{x} \quad (7)$$

where $\mathcal{F}(Z)$ is the measurement-based identified part while $\bar{x}$ is the model predicted part based on the past value of the estimation. In the sequel, we focus on $\mathcal{F}(Z)$ involved in (7) .

III. A NONLINEAR IDENTIFICATION FRAMEWORK

Identification of nonlinear relationships is an open issue. Many frameworks have been proposed including neural networks, Wiener-Hammerstein, Volterra Series based formulations [9] to cite but few possibilities. While offering high approximation capabilities, nonlinear approximators need non convex optimization in order to compute the approximator parameters. Such optimization problems suffer from computation time issue and the presence of local minima that may prevent the solver from reaching the global minimizers. In this paper, a nonlinear approximator is proposed that can be computed through constrained QP formulation at the price of lesser universality.

More precisely, in this paper, attention is focused on nonlinear maps F that takes the following form:

$$F(Z) := \Gamma^{-1}(L^T Z) \quad ; \quad \Gamma(\cdot) \quad \text{strictly increasing} \quad (8)$$

The existence of Γ^{-1} is guaranteed by the strict monotonicity of Γ. Note that putting together (6) and (8) enables the identification problem to be re-formulated as the one of finding:

- the vector $L \in \mathbb{R}^{n_z}$ and
- the monotonic increasing function Γ such that the following approximation holds:

$$\Gamma(r(k)) \approx L^T Z(k) \quad (9)$$

Remark 2: Note that (8) is a nonlinear parameterization as one has to find both Γ^{-1} and L which operate nonlinearly. Moreover, while this structure is adapted to the derivation of efficiently solvable constrained QP problems, it is not universal in the sense that an arbitrary function $F(Z)$ cannot necessarily be represented using the structure (8). Only the class of functions F for which there exists a linear combination of the components of Z that maps to Z through a monotonic function is eligible. This property is impossible to check a priori, but the efficiency of the QP problem solvers makes it easy to check even for very rich parameterization. If the residual is still too high, then the system is surely out of this class since failure cannot result from local minima as it is the case in standard nonlinear approximators computation.

The general form of the l.h.s of (9) that is used hereafter is the one given by:

$$\Gamma(r) \quad := \quad \left[B\big(\frac{r - r_{min}}{r_{max} - r_{min}} \big) \right] \cdot \mu$$

$$=: \quad \left[B(\eta(r)) \right] \cdot \mu = \sum_{j=1}^{n_b} \mu_j \left[B^{(j)}(\eta(r)) \right] \quad (10)$$

where r_{min} and r_{max} are the minimum and maximum values of r over the learning data (see section IV) while $B(\eta)$ is a function basis that is hereafter defined according to:

$$\left\{ B^{(j)} \right\}_{j=1}^{n_b} := \left\{ 1 \right\} \cap \left\{ B_1^{(i)} \right\}_{i=1}^{n_m - 1} \cap \left\{ B_2^{(i)} \right\}_{i=1}^{n_m} \quad (11)$$

where the number of functions in the set is $n_b = 2 n_m$ while the functions $B_1^{(i)}$ and $B_2^{(i)}$ are defined by:

$$B_1^{(i)} := (1 + \alpha_i) \frac{\eta}{1 + \alpha_i \eta} \quad ; \quad B_2^{(i)} := \frac{\eta}{1 + \alpha_i - \alpha_i \eta}$$

The coefficients α_i are given by $\alpha_i := \exp\big(\beta(1 - i) \big) - 1$. Note that many other function basis can be used although the author's experience suggests that the function basis proposed above is rather appropriate given the monotonicity character of the targeted nonlinear map.

Now by combining (9)-(10) , it follows that the unknowns L and μ may be obtained by solving the following least squares problem:

$$\min_{L, \mu} \sum_{(r, Z) \in \mathcal{E}} \| B(\eta(r))\mu - Z^T L \|^2$$

$$\text{under the constraints (15)} \quad (12)$$

where $\mathcal{E}$ is the learning data including a large number of instantiations of the pairs (r, Z). Note however that the least squares minimization invoked in (12) has to be done taking into account the following constraints:

1) **Γ must be strictly increasing**. This leads to the following inequality constraint on the parameter vector μ:

$$\forall \eta \in [0, 1] \quad \left[\frac{dB}{d\eta}(\eta) \right]\mu \geq \varepsilon \quad (13)$$

for some a priori chosen lower bound of the derivative $\varepsilon > 0$.
2) The following **normalization integral constraint has to be satisfied** in order to avoid trivial zero values solution ($L = 0$, $\mu = 0$):

$$\left[\int_0^1 B(\eta)d\eta \right] \cdot \mu = \frac{1}{2} \left[r_{min} + r_{max} \right] \quad (14)$$

The two sets of constraints (13) and (14) are affine in the decision variable μ. Therefore by considering a sequence of values $0 = \eta_1 < \eta_2 < \cdots < \eta_q = 1$, these constraints can be approximated by the following matrix inequalities:

$$[A_{ineq}]\mu \leq B_{ineq} \in \mathbb{R}^q \quad ; \quad [A_{eq}]\mu = B_{eq} \in \mathbb{R} \quad (15)$$

This can be done by using a finite difference or a collocation scheme. This is because for any given finite number n_b of functions, it can be shown that there is a sufficiently high number of collocation points q such that the satisfaction of (13) at the η_i's implies the satisfaction of $\frac{dB}{d\eta}(\eta) \geq \varepsilon/2$ for

all η. Moreover, one can achieve any prescribed degree of precision in the satisfaction of (14) which is anyway a loosely defined normalization constraint.

To summarize, the quadratic function involved in (12) in the decision variables L and μ is minimized under the linear constraints (15) to obtain the optimal parameters L_{opt} and μ_{opt}.
Note that there is no loss in generality in taking Γ strictly increasing since it suffices to take L of opposite sign to make (9) valid for a strictly decreasing map.

The following result on the estimation error is straightforward:
Proposition 1: Let $\mathbb{X} \subset \mathbb{R}^n$ be a subset of the state space to which belong all states of interest. Suppose that the following conditions are satisfied:
1) The admissible control sequences are bounded (i.e. $\tilde{u} \in \tilde{\mathbb{U}}$) and lead for any $x^{(0)} \in \mathbb{X}$ to a bounded sequence of output ($\tilde{y} \in \tilde{\mathbb{Y}}$)
2) There is an upper bound $\varepsilon_x > 0$ on the identification residual:

$$\sup_{(x^0, \tilde{u}) \in \mathbb{X} \times \tilde{\mathbb{U}}} \left\| x^{(N-1)} - \mathcal{F}\big(Z(x^{(0)}, \tilde{u})\big) \right\| \leq \varepsilon_x \quad (16)$$

where the short notation $x^{(N-1)} = X(N-1, 0, x^{(0)}, \tilde{u})$ is used while $Z(x^{(0)}, \tilde{u})$ is the measurement regressor that is built by simulating the system model starting from the initial state $x^{(0)}$ and using the control sequence $\tilde{u}$.
3) For any initial state $x^{(0)}$, the combined effect of noise and model mismatches on the regressor Z is bounded according to:

$$\| Z_{\text{reel}}(x^{(0)}, \tilde{u}) - Z(x^{(0)}, \tilde{u}) \|_\infty \leq \gamma \quad (17)$$

where Z_{reel} denotes the real measurement matrix (which is not obtained by simulation) that may be obtained on the real system starting from $x^{(0)}$ and under the control sequence $\tilde{u}$.

then the estimation error satisfies:

$$\| \hat{x} - x \| \leq \varepsilon_x + O(\gamma) \quad (18)$$

PROOF. First of all, note that assumption 2) implies that all nominal measurement regressors Z of interest belong to some compact set $\mathbb{Z}$. Using the assumptions (16)-(17), one clearly has:

$$\left\| x^{(N-1)} - \hat{x} \right\| \; := \; \left\| x^{(N-1)} - \mathcal{F}\big(Z_{\text{reel}}(x^{(0)}, \tilde{u})\big) \right\|$$

$$\leq \| x^{(N-1)} - \mathcal{F}(Z) \| + \| \mathcal{F}(Z) - \mathcal{F}(z_{\text{reel}}) \|$$

$$\leq \varepsilon_x + \max_{Z \in \mathcal{Z}} \left\| \frac{\partial \mathcal{F}}{\partial Z}(Z) \right\| \cdot \| Z - Z_{\text{reel}} \|$$

$$\leq \varepsilon_x + \sqrt{n_z} \cdot M(\gamma) \cdot \gamma$$

where $\mathcal{Z} := \mathbb{Z} + B(0, \gamma)$ and where $M(\gamma)$ is the maximum value of the continuous (by construction) map $\frac{\partial \mathcal{F}}{\partial Z}$ over the compact set $\mathcal{Z}$. $\square$

Remark 3: Note that the upper bound ε_x involved in (16) can be used as a cost function $\varepsilon_x(N, n_m, \beta)$ to be minimized in the decision variable (N, n_m, β) which are the parameters

of the approximator. This obviously results in a static game in which the identification step appears in the inner loop. This strengthens the relevance of having easy to solve identification problem through the constrained QP formulation. Note also that the value of ε_x reflects to which extent the map linking the regressor Z and the initial state is far from the set of maps described by the structure (8).

IV. BUILDING THE LEARNING DATA

Assume without loss of generality that one is focused on the identification problem associated to the estimation of $r = x_i$ for some $i \in \{1, \ldots, n\}$. Let us also assume that the set of relevant values of the state vector x is contained in some hypercube, namely:

$$x \in \mathbb{X} := \Pi_{i=1}^n [x_i^{min}, x_i^{max}] \qquad (19)$$

The learning data set $\mathcal{E}$ involved in the definition of the least squares problem (12) is obtained through the following steps:
1) First, a set of initial states $\{x_0^{(j)}\}_{j=1}^{n_g}$ is chosen. This can be obtained using uniform grid on each interval $[x_i^{min}, x_i^{max}]$ of possible values of the i-th component of the state.
2) A set of n_g control profiles $\{\tilde{u}^{(j)}\}_{j=1}^{n_g}$; $\tilde{u}^{(j)} \in \mathbb{R}^{N_s \cdot n_u}$ is also generated. Each profile defines a control sequence over $N_s > N$ sampling periods.
3) For each pair $(x_0^{(j)}, \tilde{u}^{(j)})$, the system is simulated over N_s sampling periods in order to generate the corresponding state profiles:

$$\left\{ X(k, 0, x_0^{(j)}, \tilde{u}^{(j)}) \quad , \quad k \in \{1, \ldots, N_s\} \right\}_{j=1}^{n_g} \qquad (20)$$

4) The data described above enables, for each pair $(j, k) \in \{1, \ldots, n_g\} \times \{N, \ldots, N_s\}$ to obtain:
- two sequences $\tilde{y}^{(j)}(k)$, $\tilde{u}^{(j)}(k)$ that defines $Z^{(j)}(k)$ according to (3).
- the corresponding value of $r^{(j)}(k)$ according to

$$r^{(j)}(k) := X\left(N, 0, X(k - N, x_0^{(j)}, \tilde{u}^{(j)}), \tilde{u}^{(j)}\right)$$

Finally, the learning data set $\mathcal{E}$ involved in the definition of the least squares problem (12) is given by:

$$\mathcal{E} := \left\{ (r^{(j)}(k), Z^{(j)}(k)) \right\}_{(j,k) \in \{1, \ldots, n_g\} \times \{N, \ldots, N_s\}} \qquad (21)$$

which is obviously a discrete set of cardinality n_E that is given by:

$$n_E := n_g \cdot (N_s - N + 1) \qquad (22)$$

V. ILLUSTRATIVE EXAMPLES

In this section, illustrative examples are proposed in order to give concrete instantiations of the different steps of the proposed framework.

A. Example 1

Let us consider the famous Van der Pol oscillator which is governed by:

$$\dot{x}_1 = -x_2 \quad ; \quad \dot{x}_2 = 4x_1 + (1 - x_1^2)x_2$$
$$y := x_1 + x_2 + w \qquad (23)$$

where w is a measurement noise assumed here to be white, Gaussian with variance $\sigma = 0.2$. The basic sampling period is taken equal to $\tau = 0.1$ sec. The bounds on the state components leading to the definition of the subset $\mathbb{X}$ are given by $\mathbb{X} := [-2, +2] \times [-5, +5]$ which obviously *enhances* the nonlinear character of the resulting identification problem (as 2^2 is not negligible when compared to 1 in the expression of $\dot{x}_2$). The learning data has been defined using a uniform grid containing $n_g = 4^2 = 16$ initial states. For each state, the system is simulated during $N_s = 100$ sampling period (10 sec) which, according to (22) leads to a learning data $\mathcal{E}$ of cardinality $n_E = 16 \times (100 - 10 + 1) = 1456$. Note that since there are two states x_1 and x_2 to be estimated, two identification problems are solved and two nonlinear maps $F^{(1)}$ and $F^{(2)}$ are obtained for the reconstruction of these two states according to $\hat{x}_i(k) = F^{(i)}(Z(k))$, $i \in \{1, 2\}$ The identification parameters $N = 10$ and $n_m = 5$ are used which leads to an observation horizon of $N \times \tau = 1$ sec and a functional basis containing $n_b = 10$ elements.

The quality of the resulting matching between the identified and the *true* values are shown in the upper subplots of Figure 1 (left plots). Note that the identified values are obtained using an intentionally noised simulation data which enables one to inject the noise already in the identification process resolving the trade-off at this early stage. The lower subplots of Figure 1 (Left) shows the gradient of the resulting nonlinear functions $\Gamma^{(1)}$ and $\Gamma^{(2)}$ that are involved in (8) for the two components of the state respectively. One may note the highly nonlinear character of the map $\Gamma^{(1)}$ in particular. An estimation scenario is shown in Figure 1 (right plots) where the initial state of the system and the observer are respectively given by $x(0) = (2, 0.5)^T$ and $\hat{x}(0) = (1, 0.7)^T$ and where a new generation of the measurement noise is used (different from the ones used to construct the learning data set). Note that the correction of the observer begins only when data is obtained that covers the observer horizon length. This means that the first correction occurs at $t = (N - 1)\tau = 0.9$ sec. It is worth emphasizing here that the observer design is done using only the output measurement related correction in order to concentrate on the contribution of the present paper. It goes without saying that a balanced estimation in which the output-related estimation and the state equation related estimation can be implemented following (7) or more generally the standard ideas of [8], [13] and the references therein.

B. Example 2

Let us consider the problem of estimating the state of a dynamic model describing the Escherichia Coli Strain. Many
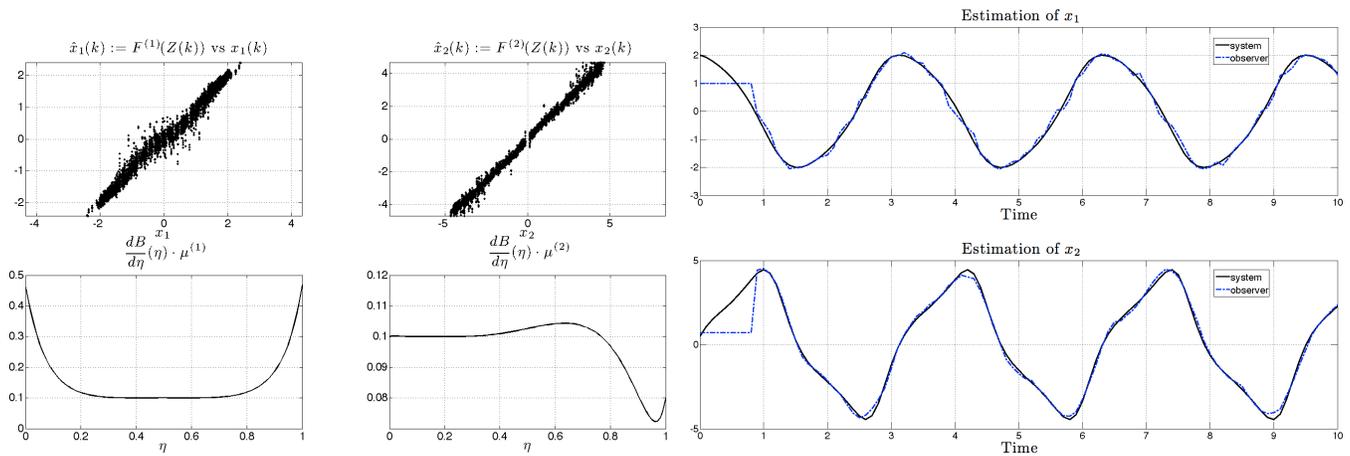
Fig. 1.   Example 1. (Left): A picture showing the results of the nonlinear identification procedure using a learning set of cardinality $n_E = 1456$. The upper plots compares the estimated states based on the identified maps $F^{(1)}$ and $F^{(2)}$ while the lower plots show the gradient of the corresponding monotonic functions $\Gamma^{(1)}$ and $\Gamma^{(2)}$ (Right) : Typical results of the state estimation using a purely output-related estimation. The observer begins once the first $N = 10$ first measurements are acquired (first correction at $t = 0.9\ sec$). This estimation is obtained while a Gaussian measurement noise w of variance $\sigma = 0.2$ is injected

knowledge-based model derivation attempts have been investigated in order to better understand the mechanisms that underline the evolution of the population [4], [10] or to develop model-based state observers [12]. The dynamic model that is commonly used in deriving dynamic state estimation involves the E. Coli strain X that grows on the limiting substrate S while yielding a final intracellular product: the β-galactosidae P. The model is given by:

$$\dot{X} = \mu(S)X - k_d \exp(-\frac{k_p}{P})X \tag{24}$$

$$\dot{S} = -y_s \mu(S)X - k_m X \tag{25}$$

$$\dot{P} = y_p \mu(S)\frac{I}{I + k_I}X - k_d \exp(-\frac{k_p}{P})P \tag{26}$$

where μ is the growth rate that is modelled using classical Monod-type relation such as $\mu(S) := \dfrac{\mu_m S}{k_s + S}$ where μ_m is the maximum specific growth rate for the cell growth (in h^{-1}). k_s is the half saturation constant; k_p and k_d are constants involved in the Arrhenius-type death kinetic that depends on P. k_m is a maintenance rate that describes the energy required for normal upkeep and repair. y_s, y_l [used in the measurement equation (27) below] and y_p are identified coefficients. I stands for the arabinose inducer that is assumed to be constant (no degradation). The output measurement vector is given by $y := (X + w_1, L + w_2)^T$ where L is the light produced by the bioluminescence that is linked to the state variables by the following expressions:

$$L = y_l \cdot \mu(S)\frac{I}{I + k_l}XP \tag{27}$$

while w_1 and w_2 are measurement noises that are taken here white, Gaussian and of variances σ_1 and σ_2 respectively. The values of the model parameters used in the sequel can be found in [12].
For this example, the framework described above can be used to construct a reduced observer. More precisely, it is shown

hereafter that there is a satisfactory solution to the underlined identification problem for any learning set that is constructed using an admissible set of initial conditions of the form $\mathbb{X}(X_0) := \{X_0\} \times [0, 5] \times [0, 0.3]$ with the following parameters $\tau = 0.2$; $N_s = 40$; $N = 3$; $n_m = 5$; $n_g = 25$. This means that for each initial value X_0 of the E. Coli strain, $n_g = 25$ simulations of the system with different initial states (sharing all the same value X_0 while having different values of P_0 and S_0) is simulated during 8 time units ($= N_s \tau = 40 \times 0.2$) hence generating a learning set of cardinality n_E given by $n_E = 25 \times (40 - 3) = 925$. Note that two identification problems are to be defined and solved using the following definition of the quantities r_1 and r_2 to be identified:

$$r_1 := P \quad ; \quad r_2 = \mu(S) \cdot X \tag{28}$$

Note that if r_1 and r_2 are well estimated then $\mu(S)$ can also be well estimated since X is assumed to be measured (reduced observer). Note also that since only $\mu(S)$ is involved in the system equation, S is necessarily estimated through $\mu(S)$.
The identification results are shown on Figure 2 for two different values of $X_0 \in \{0.5, 2\}$. One can appreciate that a good match between the estimated $\hat{r}_i$ and the simulated r_i for $i = 1, 2$ is obtained over the 925 learning set while using a rather economic parametrization ($N = 3$ and $n_m = 5$). This clearly shows that the proposed methodology enables us to perform the estimation scheme provided that the initial value of the state component X is measured at the beginning of the batch.
    Figure 3 shows typical behavior of the output-based state estimation that used the two nonlinear maps identified above. Note that the noises w_1 and w_2 that affect the measured signals used in the construction of the regressor are given by $\sigma_1 = 0.1$ and $\sigma_2 = 20$.

VI. CONCLUSION AND FUTURE WORK

In this paper, a nonlinear approximator has been proposed for a class of nonlinear relationships and has been applied

(a) Identification results for $X(0) = 0.5$

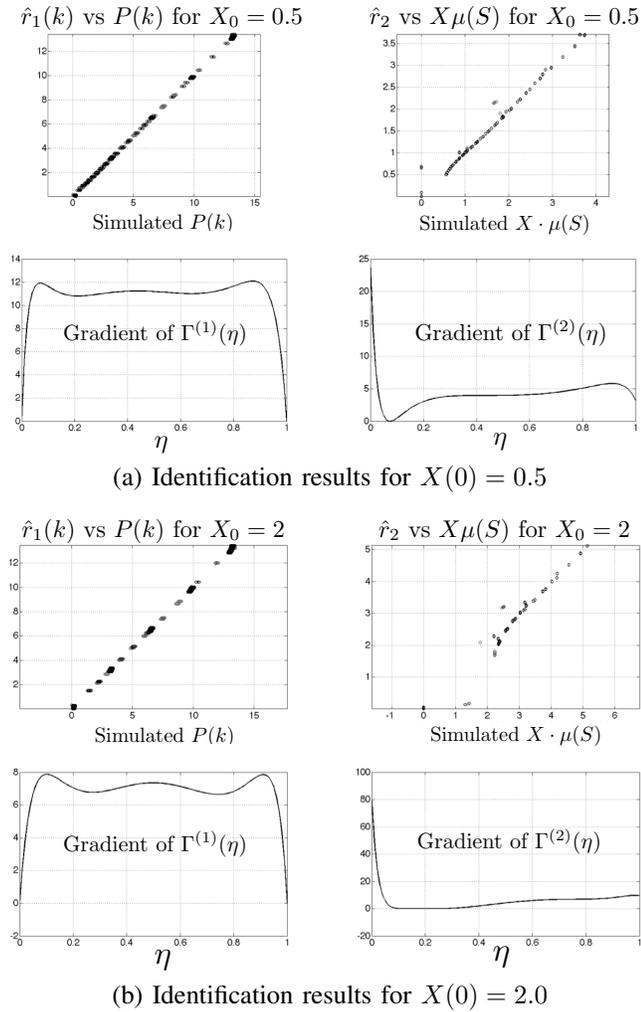(b) Identification results for $X(0) = 2.0$

Fig. 2.   Example 2. Examples of Identification results for $r_1 = P$ and $r_2 = \mu(S) \cdot X$ for two different values of the E. Coli strain $X_0 \in \{0.5, 2\}$.
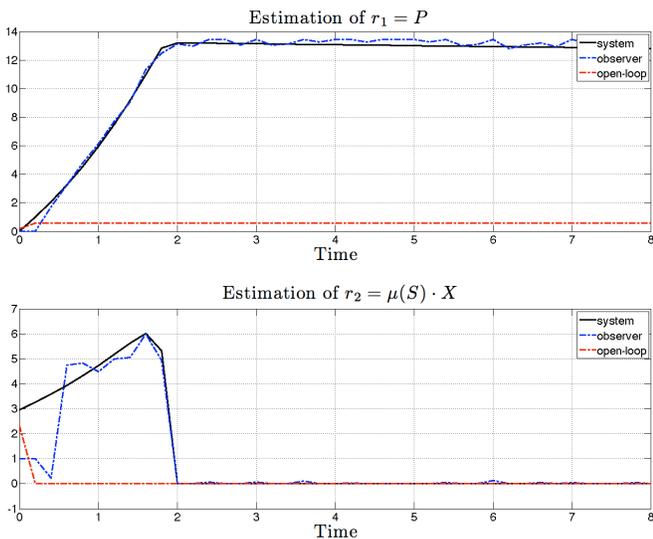

Fig. 3.   Example 2. Behavior of the output-based estimation of $r_1 = P$ and $r_2 = \mu(S) \cdot X$. The measurement noise variances are respectively given by $\sigma_1 = 0.1$ and $\sigma_2 = 20$.

in the context of moving-horizon observer design. The proposed scheme offers the advantage of requiring only a constrained QP problem solution and can therefore be efficiently integrated in the inner loop of a global scheme aiming at optimizing the approximator parameters.

A potential research line is to investigate a systematic computation of the optimal triplet (N, τ, β) defining the nonlinear approximator. Indeed, a convenient (*optimal*) choice of these parameters is intimately linked to the noise level as well as the uncertainty structure. A good knowledge of the latter is crucial to obtain pertinent choice of these parameters which had been found in this paper by trial and error approach.

Another research track concerns the use of sparse identification techniques in order to derive low dimensional parameter vector. This can be greatly facilitated by the availability of the Lagrange multipliers of the QP problem that underline the identification step.

REFERENCES

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

[2] A. Alessandri, M. Baglietto, G. Battistelli, and M. Gaggero. Moving-horizon state estimation for nonlinear systems using neural networks. *IEEE Transactions on Neural Networks*, 22(5):768–780, 2011.

[3] A. Alessandri, M. Baglietto, G. Battistelli, and V. Zavala. Advances in moving horizon estimation for nonlinear systems. In *Proceedings of the 49th IEEE conference on Decision and Control, Atlanta, GA, USA.*, 2010.

[4] H. J. Cha, C. F. Wu, J. Valdes, G. Rao, and W. Bentley. Observation of green fluorescent protein as a fusion partner in genetically engineered escherichia coli: monitoring protein expression and solubility. *Biotech. Bioeng.*, 76:565–574, 2000.

[5] M. Diehl, H. J. Ferreau, and N. Haverbeke. *Assessment and Future Directions of Nonlinear Model Predictive Control*, chapter Efficient Numerical Methods for Nonlinear MPC and Moving Horizon Estimation, pages 317–417. Springer-Verlag, 2009.

[6] G. Besançon (Ed). *Nonlinear Observers and Applications*. Lecture Notes in Control and Information Sciences. Springer-Verlag, 2007.

[7] J. P. Gauthier, H. Hammouri, and S. Othman. A simple observer for nonlinear systems – application to bio-reactors. IEEE Transactions on Automatic Control, 37, 1992.

[8] E. L. Hasteline and J. B. Rawlings. Critical Evaluation of Extended Kalman Filtering and Moving Horizon Estimation. *Ind. Eng. Chem.*, 44(8), 2005.

[9] F. J. Doyle III, R. K. Pearson, and B. A. Ogunnaike. *Identification and Control Using Volterra Models*. Springer-Verlag, London, 2001.

[10] J. Lee and F. Ramirez. Mathematical modeling of induced protein production by recombinant bacteria. *Biotech. Bioeng.*, 29:635–646, 1992.

[11] H. Michalska and D. Q. Mayne. Moving-horizon observers and observer-based control. *IEEE Transactions on Automatic Control*, 40:995–1006, 1995.

[12] M. Nardi, I. Trezzani, H. Hammouri, and P. Dhurjati. Mathematical model and observer for recombinant *escherichia coli* strain with biolumninscence indicator genes. In *Proceedings of the Second International Symposium on Communications, Control and Signal Processing (ISCCSP 2006)*, 2006.

[13] C. V. Rao, J. B. Rawlings, and D. Q. Mayne. State Constrained Estimation for Nonlinear Discrete-Time Systems: Stability and Moving Horizon Approximations. *IEEE Transactions on Automatic Control*, 48(2), 2003.

[14] D. Simon. *Optimal State Estimation*. Wiley Interscience, 2006.

[15] J. E. Slotine, J. K. Hedrick, and E. A. Misawa. On sliding observers for nonlinear systems. Journal of Dynamics Sysytems, Measurements and Control, 109, 1987.