Reducing Computational Load in Moving-Horizon Observers using Partial explicit Map Inversion

Oumayma Omar and Mazen Alamir

Abstract—This paper proposes a novel approach that may help reducing the computational burden when using the Nonlinear Moving-Horizon Observer (MHO) technique. This technique is often used to reconstruct unmeasured quantities of a dynamic system. The latter may include both the state and the vector of parameters that are involved in the system model. The proposed approach is based on the use of partial explicit inversion maps that express a part of the problem unknowns as a function of the remaining one. The Moving-Horizon Observers can therefore concentrate on the latter reduced dimensional unknown vector. The paper shows how explicit inversion maps can be derived based on a recently developed graphical-signature-based technique. Two illustrative examples are proposed to show the efficiency of the proposed solution.

I. INTRODUCTION

It is needless to say that the problem of reconstructing unknown quantities from a limited set of measurement remains a key open problem in nonlinear systems theory. The reason is that the solution of the above problem underlines almost all classical dynamic systems paradigms such as measurement-feedback control, diagnosis, supervision and fault tolerant control to cite but a limited number of issues.

Algorithms that achieve this task are called observers. As far as nonlinear systems are concerned, many observation techniques have been developed during the last four decades. This includes high-gain observers [5], sliding-modes observers [12], Moving-Horizon Observers (MHO) [8] and naturally, the widely used Extended-Kalman-Filter (EKF) observer. Excellent reviews of nonlinear observer design techniques can be found in [11] and [4].

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 a non convex optimization problem 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). The on-line solution of this optimization problem can be cumbersome to an extent that may questions the feasibility of the whole approach, at least for a family of problems that needs high updating rates.

The aim of the present paper is to address this specific complexity reduction problem by using partial explicit map inversion. This is made possible when a part of the unknown vector can be made explicitly dependent on the remaining part. It is suggested that such a map inversion can be performed using graphical signature based classification tool that has been recently developed in a series of papers [13], [14], [15]. By doing so, the dimension of the decision variable to be updated on-line can be strongly reduced and the real-time implementability of the MHO algorithm is enhanced.

The paper is organized as follows: First, the key idea is clearly stated in section II where the central role played by the inversion map is underlined. Section III recalls the graphical signature-based classification framework and states the necessary conditions that are needed for it to be used in the inversion map derivation. The whole solution is then illustrated in section IV by two illustrative examples in order to make concrete the set of concepts invoked throughout the paper. Finally, section V concludes the paper and gives a road map for future investigations.

II. THE KEY IDEA

Let us consider a nonlinear system that admits the following general model:

\[ x(k + i) = X(i, x(k), p, u) \]

(1)

where \( x \in \mathbb{R}^n \) is the state vector, \( u \) is the control profile that is applied during the time interval \([k\tau, (k+i)\tau]\) where \( \tau \) is some sampling period. The vector \( p \in \mathbb{R}^n \) gathers the model unknown parameters. Note that (1) may be obtained based on physical modeling involving Ordinary Differential Equations (ODEs), Partial Differential Equations (PDEs) or a more involved hybrid nonlinear model. It is assumed that measurements \( y(k) := h(x(k), p) \in \mathbb{R}^{n_y} \) are periodically acquired and stored over some time window of length \( T = N_O \cdot \tau \) to form the measurement data given by:

\[ Y_m(k) := \begin{pmatrix} y(k) - 1 \\ \vdots \\ y(k - N_O) \end{pmatrix} \in \mathbb{R}^{N_O \cdot n_y} \]

(2)

The use of Moving-Horizon Observer (MHO) for simultaneous state and parameter estimation is based on the
implicit assumption according to which there exists a map 
\( \Psi : \mathbb{R}^{N_O \times n_p} \rightarrow \mathbb{R}^n \times \mathbb{R}^{n_T} \) such that:

\[
(x(k), p) = \Psi(Y_m(k))
\]  

(3)

Otherwise, the simultaneous estimation problem would be theoretically infeasible. However, the existence of such a map does not mean that it can be explicitly available for observer design. That is the reason why MHO design relies on the implicit reconstruction of \( \Psi \) through the solution of the following optimization problem:

\[
\begin{align*}
(\hat{x}(k - N_O), \hat{p}) := \\
\arg\min_{(x,p)} \sum_{i=0}^{N_O-1} \|y(k - N_O + i) - Y(i, \xi, p, u)\|^2
\end{align*}
\]  

(4)

leading to the state estimate given by

\[ \hat{x}(k) = X(N_O, \hat{x}(k - N_O), \hat{p}, u) \]

Note that the optimization problem (4) has to be solved at each sampling period which may represent a challenging task in some cases. Despite this potential problem, softwares for MHO design are by now quite frequently proposed [3], [16], [1] with encouraging successful results. Nevertheless, it remains a fact that the MHO implementation can be made easier if the underlying optimization problem can be simplified by some off-line preliminary investigation. Such a scheme is proposed in the remainder of the present paper.

A. Partial Map Inversion

In order to simplify the notation, let us use \( z := (\xi, p) \in \mathbb{R}^{n+np} \) to denote the decision variable in the optimization problem (4). Using this notation, the straightforward notation \( X(i, z, u) \) and \( Y(i, z, u) \) can be used to denote the trajectories of the state and the output starting from the initial conditions and the parameter vector value both contained in \( z \), according to the original notation \( X(i, (\xi, p), u) \) already used in (1). Moreover, the optimization problem (4) can be shortly rewritten as follows:

\[ \hat{z} := \arg\min_{z} J(z | Y_m(k), u) \]  

(5)

Now assume that the following assumption holds:

Assumption 1:

There exists a partition \( z = (z^{(1)}, z^{(2)}) \) with \( (z^{(1)}, z^{(2)}) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \) such that for any given pair \((z^{(2)}, u)\), there exists an explicitly invertible map taking the following form:

\[
T_{(z^{(2)}, u)} : \mathbb{R}^{n_1} \rightarrow \mathbb{R}^{n_1}
\]

\[
z^{(1)} \sim S\left(\{Y(i, Z(z^{(1)}, z^{(2)}), u)\}_{i=0}^{N_O-1}\right)
\]  

(6)

where \( S \) is a function that maps the sequence of measurements corresponding to the initial state and the vector of parameters defined by \((z^{(1)}, z^{(2)})\) and the control profile \( u \) into \( \mathbb{R}^{n_1} \).

More precisely, given \((z^{(2)}, u)\), for each \( z^{(1)} \), the image of \( T_{(z^{(2)}, u)} \) is constructed according to the following steps:

1) Compute \( z = Z(z^{(1)}, z^{(2)}) \)
2) Compute the output trajectory \( \{Y(i, z, u)\}_{i=0}^{N_O-1} \)
3) Apply the map \( S \)

The key point in Assumption 1 lies in the invertibility of the map \( T_{(z^{(2)}, u)} \) for any pair \((z^{(2)}, u)\) since this clearly leads to the following result:

**Proposition 1:**

If Assumption 1 holds, then, the \((n + n_p)\)-dimensional optimization problem (5) can be solved by solving the \((n + n_p - n_1)\)-dimensional problem given by:

\[ \hat{z}^{(2)} := \arg\min_{z^{(2)}} J_r(z^{(2)} | Y_m(k), u) \]  

(7)

where

\[ J_r(z^{(2)} | Y_m(k), u) := J(\hat{T}^{-1}_{(z^{(2)}, u)}(S(Y_m(k))), z^{(2)}) | Y_m(k), u) \]  

(8)

where \( T^{-1}_{(z^{(2)}, u)}(S(Y_m(k))) \) is nothing but the expected value of \( z^{(1)} \) if \( z^{(2)} \) takes the true value that led to the measurement data \( Y_m(k) \).

Section IV shows two examples with \( n = 2, n_p = 1 \) and \( n_1 = 2 \) meaning that an original 3-dimensional optimization problem can be replaced by a scalar optimization problem, after an appropriate inversion map is found off-line.

Needless to say that finding an explicitly invertible map like the one invoked in Assumption 1, is generally an extremely hard task. The main contribution of the present paper is to attract the reader attention to an approximate inversion technique that the authors found efficient on a variety of examples of which only two are shown here for the lack of space. The next section reminds the signature-based classification tool that has been recently developed and successfully applied (see [13], [14], [15]). This tool is in the heart of the proposed approximate partial inversion map invoked in the above discussion since it is used to define the map \( S \) invoked in Assumption 1.

III. A GRAPHICAL SIGNATURE-BASED TOOL FOR MAP INVERSION

A. Definitions

Let \((z^{(2)}, u)\) be given; we consider the set \( Z^{(1)} \subset \mathbb{R}^{n_1} \) of admissible values of \( z^{(1)} \) (say a hypercube of \( \mathbb{R}^{n_1} \)). Note that the map:

\[ z^{(1)} \sim \{Y(i, Z(z^{(1)}, z^{(2)}), u)\}_{i=0}^{N_O-1} \]  

(9)

associates to each value \( z^{(1)} \in Z^{(1)} \) an output sequence of length \( N_O \). Now let’s take \( N = N_O - M \) (for some integer \( M \)) and assume that there is a map \( P : \mathbb{R}^{N+1} \rightarrow \mathbb{R}^2 \) that associates to each output sequence \( Y \) of length \( N + 1 \), a
specific point \( P(Y) \) in \( \mathbb{R}^2 \) (see Fig. 1). Applying this map to the \( M \) output sequences given by:

\[
\{ Y^{(i)}(z^{(1)}) := \begin{pmatrix} Y(N + i, \ldots) \\ \vdots \\ Y(i, \ldots) \end{pmatrix} \in \mathbb{R}^{N+1} \}_{i=0}^{M-1}
\]

(10)
yields a 2D graphical signature (including \( M \) points) in the 2D plane. This signature is denoted hereafter as follows:

\[
S(z^{(1)}) := \left\{ \left\{ \chi_i(z^{(1)}) \right\} = P(Y^{(i)}(z^{(1)})) \in \mathbb{R}^2 \right\}_{i=0}^{M-1}
\]

(11)
The function \( P \) that is needed to generate this signature is detailed in Appendix A.

To summarize, each measurement data contained in the observation window of length \( N_D \) can be represented by a 2D signature containing \( M \) points. As a matter of fact, the integer \( N = N_D - M \geq 1 \) is a free parameter that can be chosen to obtain different signatures (see Appendix A.), consequently, one may write \( S(N)(z^{(1)}) \) to explicitly express this fact.

**D. Definition of the invertible map \( T_{(z^{(2)}, u)}(\cdot) \)**

For a given pair \( (z^{(2)}, u) \), a typical definition of \( T_{(z^{(2)}, u)}(z^{(1)}) \) invoked in Assumption 1 is given by:

\[
T_{(z^{(2)}, u)}(z^{(1)}) := \begin{pmatrix} r_1(S^{N_1}(z^{(1)})) \\ \vdots \\ r_{n_1}(S^{N_{n_1}}(z^{(1)})) \end{pmatrix}
\]

(12)
where \( c_i := (S^{(N_i)}, r_i), i = 1, \ldots, n_1 \) are \( n_1 \) conveniently chosen coordinates making (12) invertible over the admissible set \( \mathbb{Z}^{(1)} \). The choice of these coordinates is done by visual inspection based on the investigation of different candidate values for \( N \) and \( r \). The visual inspection is highly facilitated by the 2D character of the signatures. The examples proposed in the next section make easier the understanding of the proposed framework since concrete situations, signatures, properties and coordinates can be examined.

**IV. ILLUSTRATIVE EXAMPLES**

In section II we have clearly explained the inversion scheme that helps to reduce the complexity of optimisation problem invoked when using the MHO algorithm. In this section we will focus on the steps followed to find the invertible map (12); we will demonstrate how to find an appropriate graphical signature and how to choose a relevant system of coordinates. This one will help in finding a part of unknown variables depicted in (5) as off-line step of resolution.

**A. Example 1**

Let us first consider the reversible reaction [10]:

\[
2A \rightleftharpoons B
\]

(13)
with stoichiometric matrix \( \nu = (-2, 1) \) and reaction rate \( \bar{r} = \bar{r}C_A^2 \). The state and measurement vectors are respectively given by \( x := (A, B, \beta)^T \) and \( y = x_1 + x_2 \). Assuming a perfect gas and a perfectly mixed isothermal reactor, the system equations write:

\[
\dot{x} = f(x) = \nu^T \bar{r} = \left( -\frac{2\beta x_1^2}{\beta x_1^2} \right)
\]

(14)
\[
y = x_1 + x_2
\]

(15)
which, in discrete time form leads to a system similar to (1) with two states \( x_1, x_2 \) and one parameter \( \beta \) that is assumed here to be unknown.

The framework proposed in section II is applied to this example using the following definitions:

- \( z^{(1)} := (x_1, x_2)^T \), \( n_1 = 2 \), \( z^{(2)} = \beta \)

- \( \mathbb{Z}^{(1)} := [0, 3] \times [0, 3], N_D = 11, \tau = 0.01 \)

Let us define the following discrete subset of \( \mathbb{Z}^{(1)} \) that may be obtained using uniform grids on the admissible intervals for \( x_1 \) and \( x_2 \), namely:

\[
\mathbb{Z}_d^{(1)} := \{x_{11}, \ldots, x_{15}\} \times \{x_{21}, \ldots, x_{25}\}
\]
This is clearly a discrete set containing 25 elements. To these 25 elements of $\mathbb{Z}_d^{(1)}$ correspond 25 signatures

$$\left\{ S^N(z^{(1)}) \right\}_{z^{(1)} \in \mathbb{Z}_d^{(1)}}$$

that can be defined once a given value $N < N_0$ and $\bar{\beta}$ are chosen.

Fig. 2 shows these 25 signatures viewed differently on the two subplots a) and b). More precisely Fig 2.a) shows these 25 signatures with all those sharing the same value $x_1$ plotted using the same color. For instance, the five signatures in green on Fig. 2.a) are obtained for the set of values:

$$\mathcal{E}_{i_1} := \left\{ (x_{i_1} = 0, x_{2j}) \right\}_{j=1}^5$$

Similarly, Fig. 2b) regroups the signatures based on the values of $x_2$. For instance, all the signatures plotted in green are those corresponding to the set of elements in $\mathbb{Z}_d^{(1)}$ given by:

$$\mathcal{E}_{i_2} := \left\{ (x_{1j}, x_{2j} = 3) \right\}_{j=1}^5$$

Visual inspection of the signatures depicted in Fig. 2 shows clearly that once the signature $S(z^{(1)})$ is plotted, it is possible to identify $z^{(1)} = (x_1, x_2)^T$ using the two following steps:

1) First determine $x_1$ based on the signature property

$$r_1(S) := \min_{i \in \{1, \ldots, M\}} \chi_i$$

This can be obtained by interpolating the information on Fig. 2.a), indeed:

$$\forall \epsilon \in \mathcal{E}_{i_1}, \quad r_1(S) \approx -3.5$$

$$\forall \epsilon \in \mathcal{E}_{i_2}, \quad r_1(S) \approx -1.5$$

and so on. For a new value of $r_1$, one can easily associate the corresponding value of $x_1$.

2) The subplot Fig 2.b) clearly shows that there is a monotonic map defined by:

$$x_2 = X_2(\eta_{\min} \mid x_1)$$

that gives $x_2$ knowing $x_1$ and the property

$$\eta_{\min} = r_2(S) := \min_{i \in \{1, \ldots, M\}} \eta_i$$

This is because once $x_1$ is given [thanks to $r_1(S)$], once knows that the tail of the signature lies approximately on a vertical line with a height that is directly linked to the value of $x_2$ in a bijective way.

The above discussion completely defines the invertible map (12) using the signature defined by $N = 2$ and the two signature properties $r_1$ and $r_2$. Note that it can be checked that this procedure holds for any value of $\beta \in [0.1, 0.35]$.

Once the invertible map is obtained, the scalar dimensional problem (8) can be defined in the scalar decision variable $\beta$. Fig. 2c) shows the shape of the cost function $J_r(\beta \mid Y_m(k))$ for different instants $k$ when the value $\beta = 0.15$ is used to produce the measurement vector $Y_m(k)$. This clearly shows that solving this scalar optimization problem is quite easier than solving the original 3-dimensional problem.
B. Example 2:

In order to illustrate the above ideas on another concrete example, let us consider the dynamical model of recombinant Escherichia Coli strain ([2], [7], [9], [6]). This model is a mass balance model describing the pure recombinant microbial batch culture of E. Coli strain X growing on the limiting substrate glycerol S while yielding a final intracellular product β-galactosidase protei P:

\[
\begin{align*}
\dot{X} &= \mu(S)X - k_d \exp(-\frac{k_p}{P})X \\
\dot{S} &= -y_s \mu(S)X - k_m X \\
\dot{P} &= \frac{y_p \mu(S)}{I + k_l}X - k_d \exp(-\frac{k_p}{P})P
\end{align*}
\]

(16) (17) (18)

where \( \mu \) is the growth rate that is modeled using classical Monod-type relation:

\[
\mu(S) = \frac{\mu_m S}{k_s + S}
\]

in which \( \mu_m \) 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. \( y_s, y_p \) are identified yield coefficients. \( I \) stands for the arabinose inducer that is assumed to be constant (no degradation).

The output measurement is the light produced by the bioluminescence that is defined by the following expression:

\[
L = y_l \mu(S) \frac{I}{I + k_l}XP
\]

(19)

In a series of recent works ([2], [7], [9]), the parameters of the above model have been identified and the resulting model has been experimentally validated using Micro-fermentor test-bed. The resulting set of values are given on table I.

Here, the framework proposed in section II is applied to this system using the following definitions:

- \( z^{(1)} := (x_1, x_3)^T, n_1 = 2, z^{(2)} = x_2 \)
- \( Z^{(1)} := [0.01, 0.09] \times [0.05, 0.15], N_O = 11, \tau = 0.01 \)

Let us define the following discrete subset of \( Z^{(1)} \) that may be obtained using uniform grids on the admissible intervals for \( x_1 \) and \( x_3 \), namely:

\[
Z_d^{(1)} := \{x_{11}, \ldots, x_{15}\} \times \{x_{31}, \ldots, x_{35}\}
\]

The same road map that has been used in example 1 is conducted here, namely, the signatures associated to the 25 values of \( z^{(1)} \) contained in \( Z_d^{(1)} \) are plotted for visual inspection.

Fig. 3 shows these signatures viewed differently in the two subplots a) and b). More precisely, Fig. 3a) [resp. (b)] show these 25 signatures with all those sharing the same value \( x_1 \) [resp. \( x_3 \) ] plotted using the same color.

Visual inspection of the signatures depicted in Fig. 3 shows clearly that once a signature is plotted, it is possible to identify \( z^{(1)} = (x_1, x_3)^T \) using the following steps:

1) First determine \( x_3 \) based on the signature property

\[
r_1(S) := \text{curvilinear coordinate of the} \ \text{lowest point of the signature} \ \text{in the curvilinear coordinates of Fig 3.a)}
\]

This can be obtained by interpolating the information on Fig. 3a)

2) The subplot Fig .3b) clearly shows that there is a monotonic map defined by:

\[
x_3 = X_3(\eta_{min} \mid x_1)
\]

that gives \( x_3 \) knowing \( x_1 \) and the property

\[
\eta_{min} = r_2(S) := \min_{i \in \{1, \ldots, M\}} \eta_i
\]

\( \eta_{min} \) is the energy required for normal upkeep and repair.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
parameter & Values & Units \\
\hline
\( \mu_m \) & 0.49 & \( h^{-1} \) \\
\( k_s \) & 0.06 & \( g/l \) \\
\( k_p \) & 0.047 & \( g/l \) \\
\( k_d \) & 0.005 & \( g/l \) \\
\( k_m \) & 0.21 & \( h^{-1} \) \\
\( k_l \) & 0.03 & \( g/l \) \\
\( y_s \) & 0.75 & \( g \ \text{cell}/g \ \text{glycerol} \) \\
\( y_p \) & 0.32 & \( g \ \text{protein}/\beta - \text{galactosidase} \) \\
\( y_l \) & 17.6 & \( U/\beta - \text{galactosidase} \) \\
\hline
\end{tabular}
\caption{Identified Parameters for the dynamic model.}
\end{table}
This is because once $x_1$ is given [thanks to $r_1(S)$], one knows that the tail of the signature lies approximately on a vertical line with horizontal position determined by the value of $x_3$ in a bijective way.

The above discussion completely defines the invertible map (12) using the signature defined by $N = 4$ and the two signature properties $r_1$ and $r_2$. Once the invertible map is obtained, the scalar dimensional problem (8) can be defined in the scalar decision variable $x_2$ which leads to a scalar optimization quite easier than solving the original 3-dimensional problem.

V. CONCLUSIONS AND FUTURE WORKS

In this paper, a novel approach has been proposed that aims at reducing the on-line computational burden associated to the moving horizon observer. This is done by off-line derivation of explicit expression of a part of the unknown vector as a function of the other part and the vector of past measurements that are available for MHO computation. The concepts have been validated through two concrete examples showing the potential efficiency of the proposed approach.

Note however that although the general framework involves the presence of controlled input in the system model, no such example have been yet studied. Future work include such investigated together with an analysis of the robustness of the proposed approach in the presence of measurement noise that may alter the quality of the invertible map. Finally, the use of the proposed scheme as a tool for the analysis of the feasibility of the inverse problem as well as general nonlinear observability may also be a promising

APPENDIX

A. Expression of the map $P_0(Y)$

Given $Y \in \mathbb{R}^{N+1}$, the following map has been suggested in [13]:

$$ P_0(Y) = \Phi_0(Y) + \lambda(Y) \cdot \left[ \Phi_1(Y) - \Phi_0(Y) \right] $$

(20)

$$ \Phi_0(Y) = \frac{1}{2N} \sum_{j=1}^{N} \Psi_j(Y) $$

(21)

$$ \Phi_1(Y) = \frac{1}{2N} \sum_{j=1}^{N} Z_j \Psi_j(Y) $$

(22)

where

$$ \Psi_j(Y) := \left[ (1 + Z_i)Q_i + (Z_i - 1)Q_i \right] $$

$$ Z_i := \frac{Y_i}{\eta_n \max_{j=1}^{N} Y_i + 1} $$

$$ Q_i := \left( \cos\left(\frac{2\pi(i-1)}{N}\right), \sin\left(\frac{2\pi(i-1)}{N}\right) \right)^T $$

$$ \lambda(Y) := \frac{Y_{N+1}}{\eta_n \max_{j=1}^{N} Y_i + 1} - \frac{1}{N} \sum_{i=1}^{N} Z_i $$

The vector of signature parameter $q$ include the integer $N$, the normalization coefficient $\eta_n \in \{0, 1\}$ and the under-sampling integer that is not mentioned in the equation for simplicity.

REFERENCES


