Library-based adaptive observation through a sparsity-promoting adaptive observer

Andreu Cecilia, Ramon Costa-Castelló

***Abstract*— This paper proposes an adaptive observer for a class of nonlinear system with linear parametrization. The main novelty of the technique is that the regressor vector is considered to be unknown. Instead, a library of candidate non-linear functions is implemented, which transforms the original parameter vector into a new one that is characterized by being sparse. In such problem, it is shown that standard adaptive observers cannot recover the original vector due to a lack of persistence of excitation. Instead, a parameter-adaptation with an implicit l_1 regularization is implemented. It is shown that this new observer can recover the parameter vector under standard assumptions of sparse signal recovery. The results are validated in a numerical simulation.**

I. Introduction

During observer design, it is reasonable to expect significant discrepancy between the mathematical model and the real system. For this reason, observers should be robustified in front of model uncertainty. A common approach to achieve such robustness is based on modelling the uncertainty as a linear combination of a known regressor function, $\boldsymbol{\phi}$, and a vector of constant unknown parameters, $\boldsymbol{\theta}$. Then, combine the observer with a recursive identification algorithm that estimates the parameters, $\boldsymbol{\theta}$, and decouples the state-estimation from the uncertainty [1][2]. This approach allows to exactly cancel the effect of the uncertainty, which can potentially out-perform standard robust observers, that are limited by a well-known trade-off between bandwidth and noise rejection [3].

Nevertheless, the state-estimation and parameter-estimation decoupling is a fragile operation that can fail whenever the regressor vector, $\boldsymbol{\phi}$, is not exactly known ([4] Section 3.3.5). Moreover, as the uncertainty is related with the unmodelled parts of the system, it is common that the functions that compose $\boldsymbol{\phi}$ are unknown. This fact explains the fragility that is usual in adaptive observer techniques when they are implemented in practice. This fragility motivates the modification of the parameter-adaptation in order to increase its robustness, e.g. through the σ-modification or parameter projection [5]. However, these modifications aggravates the accuracy of the estimation.

In such context, it is more reasonable to expect that the regressor vector, $\boldsymbol{\phi}$ is not known, but there is some prior information on the type of functions that compose, $\boldsymbol{\phi}$. Therefore, it is possible to design a library of non-linear function candidates, $\boldsymbol{\Theta}$, to model the uncertainty. Then, ideally, the adaptive observer selects the best combination of functions that explains the uncertainty.

The main conflict in such approach is that, in general, most functions from the library will not be present in the real system. Therefore, the library is over-complete, which makes the parameter-estimation an ill-posed inverse problem, i.e. there are an infinite number of solutions that can explain the uncertainty in the observed system's trajectory. Consequently, the adaptive algorithm will not converge to the true parameter vector.

The key observation is that for many systems, the regressor vector, ϕ, consists on only a few terms. Therefore, only a sparse selection of non-linear functions will be active in the designed library, $\boldsymbol{\Theta}$. The presence of an ill-posed inverse problem and these hints of sparsity motivates the implementation of a l_1 regularization [6], as it is commonly used in least absolute shrinkage and selection operator (LASSO) regression. However, even though this type of optimization has been deeply studied in the context of system identification [7], regularization in general cannot be directly implemented in adaptive control. The robustness and stability of adaptive observers is commonly proved by constructing a Lyapunov function and designing a parameter-adaptation dynamics that cancels the factors that depend on the unknown parameters, $\boldsymbol{\theta}$. The addition of a l_1 penalization term on the parameter-adaptation prevents this cancellation and may destabilize the observer's dynamics [8].

For this reason, this work proposes exploiting recent results in natural gradient-based adaptive control [9], which has been shown to provide state and parameter-estimation decoupling with an implicit regularization in the parameter-estimation. The adaptive algorithm will be designed to promote sparsity on the parameter solution and it will be shown that the true parameter can be recovered under some assumptions that are common in the sparse signal recovery field [6].

The remaining of this paper is as follows. In Section II, the library-based adaptive observation problem will be formally formulated. Section III, shows that standard parameter-adaptation algorithms are insufficient for the considered problem. Section IV presents an unbiased adaptive observer that promotes sparsity in the parameter-estimation. Section V establishes some structural conditions on the library for the adaptive algorithm to recover the true regressor vector. The algorithm is validated in a numerical simulation in Section VI. Finally, some conclusions are drawn in Section VII.

This work has been partially funded by the Spanish State Research Agency through the María de Maeztu Seal of Excellence to IRI (MDM-2016-0656), by the project DOVELAR (ref. RTI2018-096001-B-C32) and by the PTI FLOWBAT 2021 project (ref. 642 201980E101).

A. Cecilia and R. Costa-Castelló are with the Institut de Robòtica i Informàtica Industrial, CSIC-UPC Llorens i Artigas 4-6, 08028 Barcelona, Spain (e-mail: andreu.cecilia@upc.edu and ramon.costa@upc.edu). (*Corresponding author: Andreu Cecilia*)

II. Problem Formulation

This work considers a MIMO nonlinear system of the form

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x},\mathbf{u}) + \mathbf{B}\boldsymbol{\phi}(\mathbf{x},\mathbf{u})\boldsymbol{\theta} \\ \mathbf{y} &= \mathbf{C}\mathbf{x} + \mathbf{v}\end{aligned} \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$ are the states, $\mathbf{y} \in \mathbb{R}^m$ is the measured output and $\boldsymbol{\theta} \in \mathbb{R}^p$ is a vector of unknown constant parameters. The matrices $\mathbf{B} \in \mathbb{R}^{n\times s}$ and $\mathbf{C} \in \mathbb{R}^{m\times n}$ are assumed to be known. The functions $\mathbf{f} \in \mathbf{R}^{n\times 1}$ and $\boldsymbol{\phi} \in \mathbb{R}^{s\times p}$ are assumed to be Lipschitz. The factor $\mathbf{v} \in \mathbb{R}^m$ depicts high-frequency sensor noise, which is assumed to be upper-bounded by a positive constant v_2 as $\|\mathbf{v}\|_2 \leq v_2$.

The objective here is to design an observer algorithm that achieves an estimation of the states, $\hat{\mathbf{x}}$, and the unknown parameters, $\hat{\boldsymbol{\theta}}$, such that as $t \to \infty$, the estimation errors satisfy, $\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \to \varepsilon_1$ and $\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|_2 \to \varepsilon_2$, where ε_1 and ε_2 are small positive constants.

The system is assumed to satisfy the following observer matching condition

Assumption 2.1: All the unknown parameters appear in the first derivative of the output, i.e.

$$rank(\mathbf{CB}) = rank(\mathbf{B}). \tag{2}$$

The main difference with respect to common adaptive observer problems appears by considering the following scenario. It is considered that the regressor vector $\boldsymbol{\phi} \in \mathbb{R}^{s\times p}$ is unknown. Instead, it is assumed that the designer constructs a matrix, $\boldsymbol{\Theta}(\mathbf{x},\mathbf{u}) \in \mathbb{R}^{s\times q}$ with $q > p$, of candidate linearly independent non-linear functions.

The function $\boldsymbol{\Theta}$ is assumed to be Lipschitz and satisfy the following assumption

Assumption 2.2: Define $\boldsymbol{\phi}_w(\mathbf{x},\mathbf{u}) \in \mathbb{R}^{s\times 1}$ for $w = 1, ..., q$, as the column vector functions that compose $\boldsymbol{\Theta}$. Each column vector of the regressor, $\phi(\mathbf{x},\mathbf{u})$, can be computed as a scaled vector $c\boldsymbol{\phi}_w(\mathbf{x},\mathbf{u})$ for some w, where $c \in \mathbb{R}$.

Assumption 2.2 allows to rewrite system (1) in the following equivalent form

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x},\mathbf{u}) + \mathbf{B}\boldsymbol{\Theta}(\mathbf{x},\mathbf{u})\boldsymbol{\theta}_s \\ \mathbf{y} &= \mathbf{C}\mathbf{x} + \mathbf{v}\end{aligned} \tag{3}$$

where $\boldsymbol{\theta}_s \in \mathbb{R}^q$ is a vector of unknown parameters, in which p elements are equal to the elements in the original parameter vector, $\boldsymbol{\theta}$, and $q - p$ elements are zero.

The matrix $\boldsymbol{\Theta}$ is going to be referred as a *library* of candidate non-linear functions. In this library, the column vectors, $\boldsymbol{\phi}_w(\mathbf{x},\mathbf{u})$, are defined as *inactive* if they do not compose the original regressor vector $\boldsymbol{\phi}$, otherwise, are defined as *active*. The elements of the parameter vector, $\boldsymbol{\theta}_s$, associated with inactive functions will be zero, otherwise, will be non-zero. The objective of the adaptation dynamics is to determine the active vectors and accurately estimating the unknown parameter vector $\boldsymbol{\theta}_s$. By considering the equivalent system (3), the original problem has been transformed into the design of an observer algorithm that achieves an estimation of the states, $\hat{\mathbf{x}}$, and the unknown parameters, $\hat{\boldsymbol{\theta}}_s$, such that as $t \to \infty$, the estimation errors satisfy, $\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \to \varepsilon_1$ and $\|\boldsymbol{\theta}_s - \hat{\boldsymbol{\theta}}_s\|_2 \to \varepsilon_2$, where ε_1 and ε_2 are small positive constants.

As it will be presented in the next section, standard adaptive observer techniques allow to solve the state-estimation objective. However, it will be shown that said results cannot achieve the parameter-estimation objective, as the overparametrization, in general, prevents the satisfaction of the necessary persistence of excitation condition.

III. Standard Adaptive Observer Performance

In order to ease the read, this section and the following will consider the noiseless case, i.e. when $\mathbf{v} = 0$. This assumption does not modify the results obtained in the following sections.

Assume that there is an observer that achieves a state-estimation for the case where the parameters, $\boldsymbol{\theta}_s$, are known. Specifically, let the observer dynamics be depicted by

$$\dot{\hat{\mathbf{x}}} = \mathbf{g}(\hat{\mathbf{x}},\mathbf{u},\mathbf{y}) + \mathbf{B}\boldsymbol{\Theta}(\hat{\mathbf{x}},\mathbf{u})\boldsymbol{\theta}_s. \tag{4}$$

where $\mathbf{g}(\hat{\mathbf{x}},\mathbf{u},\mathbf{y}) \in \mathbb{R}^{n\times 1}$ is a vector function that is designed to make the state-estimation error, $\mathbf{e}_x \triangleq \mathbf{x} - \hat{\mathbf{x}}$, converge to zero. Specifically, the vector function $\mathbf{g}(\hat{\mathbf{x}},\mathbf{u},\mathbf{y})$ is defined such that there is a differentiable Lyapunov function $V_x(\mathbf{e}_x) = \frac{1}{2}\mathbf{e}_x^\intercal \mathbf{P}\mathbf{e}_x$, with $\mathbf{P} = \mathbf{P}^\intercal > 0$ that satisfies

$$\begin{aligned}&\alpha_1(\mathbf{e}_x) \leq V_x(\mathbf{e}_x) \leq \alpha_2(\mathbf{e}_x) \\ &\frac{\partial V_x}{\partial \mathbf{e}_x}\dot{\mathbf{e}}_x \leq -k\alpha_3(\mathbf{e}_x)\end{aligned} \tag{5}$$

where $\alpha_i \; for \; i = 1, .., 3$ are positive definite functions and k is a positive constant.

As an example, such observer could be found following the insights presented in [10].

Now, consider the case where the parameters are not exactly known, and only an estimation, $\hat{\boldsymbol{\theta}}_s$, is given. In such case, the state-estimation error does not converge to zero, and the derivative of the Lyapunov function (5) is upper-bounded as

$$\frac{\partial V_x}{\partial \mathbf{e}_x}\dot{\mathbf{e}}_x \leq -k\alpha_3(\mathbf{e}_x) + \mathbf{e}_x^\intercal \mathbf{P}\mathbf{B}\boldsymbol{\Theta}(\hat{\mathbf{x}},\mathbf{u})\mathbf{e}_{\theta_s}, \tag{6}$$

where $\mathbf{e}_{\theta_s} \triangleq \boldsymbol{\theta}_s - \hat{\boldsymbol{\theta}}_s$.

The structure of the second term in the right side of (6), motivates the design of the following gradient descent-like parameter-estimation [2]

$$\dot{\hat{\boldsymbol{\theta}}}_s = \boldsymbol{\Theta}^\intercal(\hat{\mathbf{x}},\mathbf{u})\mathbf{M}(\mathbf{y} - \mathbf{C}\hat{\mathbf{x}}) \tag{7}$$

where $\mathbf{M}$ is designed to fulfil

$$\mathbf{B}^\intercal \mathbf{P} = \mathbf{MC}. \tag{8}$$

Remark 3.1: Equation (8) can always be solved if the matching condition (2) is satisfied. In the case of unmatched parameters, condition (2) can be relaxed by the use of high-gain observers [11].

The motivation behind parameter-adaptation (7) is that it cancels the effect of the unknown parameters in the derivative of the following composite Lyapunov function

$$V_{x,\theta_s}(\mathbf{e}_x, \mathbf{e}_{\theta_s}) = V_x(\mathbf{e}_x) + \mathbf{e}_{\theta_s}^\intercal \mathbf{e}_{\theta_s}. \tag{9}$$

Specifically, taking into account (7) and (8), the function $V_{x,\theta_s}(\mathbf{e}_x, \mathbf{e}_{\theta_s})$ is upper-bounded by

$$\dot{V}_{x,\theta_s} \leq -k\alpha_3(\mathbf{e}_x) + \mathbf{e}_x^\intercal \mathbf{PB\Theta}(\hat{\mathbf{x}}, \mathbf{u})\mathbf{e}_{\theta_s} - \mathbf{e}_{\theta_s}^\intercal \mathbf{\Theta}^\intercal(\hat{\mathbf{x}}, \mathbf{u})\mathbf{MCe}_x$$
$$= -k\alpha_3(\mathbf{e}_x). \tag{10}$$

The bound (10) proves two facts. First, the state-estimation error, $\mathbf{e}_x$, converges to zero independently of the parameter-estimation error $\mathbf{e}_{\theta_s}$. This result can be understood as a sort of separation principle between the state-estimation and the parameter-estimation. Second, the parameter-estimation error converges to zero if the system satisfies the persistence of excitation condition [12]. The proof of these results is included in Appendix I.

This work expands on the parameter-estimation result. In practice, the parameter-adaptation dynamics (7) converges to a set of possible solutions depending on the dimension of the time-varying null-space of $\mathbf{B\Theta}(\mathbf{x}, \mathbf{u})$, i.e. if the null-space's dimension is greater than zero, the problem can be solved by an infinite number of parameter vectors. The major conflict that arises in the considered problem is that, as $dim(\boldsymbol{\theta}_s) > dim(\boldsymbol{\theta})$, even if the null-space dimension of the original regressor vector, $\mathbf{B}\boldsymbol{\phi}(\mathbf{x}, \mathbf{u})$, is zero, no conclusion can be drawn for the null-space of $\mathbf{B\Theta}(\mathbf{x}, \mathbf{u})$. Moreover, in general, no trajectory of the original system (1) can ensure that the null-space of $\mathbf{B\Theta}(\mathbf{x}, \mathbf{u})$ is zero-dimensional, not even if the original system (1) is persistently excited.

This result shows that, in most cases, the considered overparametrized problem presents an infinite amount of solutions and standard parameter-adaptation schemes will converge to some solution that cannot be ensured to be the one appearing in the original system. The aim of the following sections is to present an alternative parameter-adaptation dynamics that, under some assumptions on the library, $\mathbf{\Theta}(\mathbf{x}, \mathbf{u})$, can always ensure the convergence of the parameter-estimation error if the original regressor, $\mathbf{B}\boldsymbol{\phi}(\mathbf{x}, \mathbf{u})$, is persistently exciting.

IV. Sparsity-Promoting Adaptive Observer

Let $\mathbf{s}$ be a signal generated as

$$\mathbf{s} \triangleq \mathbf{B}\boldsymbol{\phi}(\mathbf{x}, \mathbf{u})\boldsymbol{\theta}. \tag{11}$$

As the state and parameter-estimations are decoupled, the adaptation dynamics are actively solving the inverse problem

$$\mathbf{s} = \mathbf{B\Theta}(\mathbf{x}, \mathbf{u})\hat{\boldsymbol{\theta}}_s. \tag{12}$$

In general, as the regressor vector $\mathbf{B\Theta}(\mathbf{x}, \mathbf{u})$ is not persistently excited, the problem (12) is ill-posed. In such situation, the common approach is to introduce a regularization term [13]. The key observation is that the parameter vector, $\boldsymbol{\theta}_s$, will have many null values, making it sparse in the considered space of non-linear function candidates. Consequently, a natural approach is to introduce a regularization term that promotes sparsity in the solution and can be shown to converge to the true parameter vector through sparse signal recovery theory [14]. Specifically, this work considers the l_1 regularization term, $\|\hat{\boldsymbol{\theta}}_s\|_1$, which penalizes non-sparse parameter vectors.

It should be remarked that, because the adaptive observer technique is based on a decoupling between state and parameter-estimation (the cancellation in equation (10)), the regularization term cannot be directly introduced as a penalty term within a minimization criterion. An alternative strategy is based on including a regularization term in the parameter-adaptation (7), that regularizes the inverse problem, and a secondary regularization term in the state-estimation dynamics (4), that ensures the state/parameter decoupling [8]. However, this approach induces significant bias in the state and parameter-estimation. The bias can be reduced by incorporating prior knowledge of the parameter value in the adaptive scheme. Nonetheless, the sub-set of active non-linear functions is unknown in the considered problem. Thus, the prior knowledge of $\boldsymbol{\theta}_s$ is insufficient for an adequate state and parameter-estimation, which may lead to erroneous sub-set selection.

Lee et al. [15] proposed a modification of classic adaptive controllers by substituting the gradient-like adaptation by a natural gradient, so the resulting adaptation law respects an underlying Riemannian geometry to be specified. Based on this result, Boffi and Slotine [9] showed that it is possible to design an adaptive observer for output dependent regressor vectors that preserves the state/parameter-estimation decoupling while inducing an implicit regularization in the parameter-estimation. The idea is to substitute the quadratic term $\mathbf{e}_{\theta_s}^\intercal \mathbf{e}_{\theta_s}$ of the composite Lyapunov function (9) with the Bregman divergence of a strictly convex function ψ [16],

$$d_\psi(\boldsymbol{\theta}_s || \hat{\boldsymbol{\theta}}_s) \triangleq \psi(\boldsymbol{\theta}_s) - \psi(\hat{\boldsymbol{\theta}}_s) - (\boldsymbol{\theta}_s - \hat{\boldsymbol{\theta}}_s)^\intercal \nabla \psi(\hat{\boldsymbol{\theta}}_s).$$

Consequently, the following Lyapunov function is obtained

$$V_\psi = V_x(\mathbf{e}_x) + d_\psi(\boldsymbol{\theta}_s || \hat{\boldsymbol{\theta}}_s). \tag{13}$$

Furthermore, define the set of possible parameter solutions for the inverse problem (12) as

$$\mathbf{\Omega} \triangleq \{\hat{\boldsymbol{\theta}}_s \in \mathbb{R}^q |\ \mathbf{s} = \mathbf{B\Theta}(\mathbf{x}, \mathbf{u})\hat{\boldsymbol{\theta}}_s, \quad \forall t > t_s\},$$

where s is defined in (11) and t_s is a positive value. Then, we can stablish the following result.

Lemma 4.1: Consider the system (3) without noise, the state observer (4) and the natural gradient-like adaptation law

$$\dot{\hat{\boldsymbol{\theta}}}_s = \left[\nabla^2 \psi(\hat{\boldsymbol{\theta}}_s)\right]^{-1} \mathbf{\Theta}^\intercal(\hat{\mathbf{x}}, \mathbf{u})\mathbf{M}(\mathbf{y} - \mathbf{C}\hat{\mathbf{x}}). \tag{14}$$

Then, the state-estimation error, $\boldsymbol{e}_x$, converges to zero; and, consequently, converges to a sufficiently small positive value, i.e. $\|\boldsymbol{e}_x\| \leq \varepsilon$ for all time $t > t_s$.

Moreover, consider that $\hat{\boldsymbol{\theta}}_s(t_s) = \min_{\boldsymbol{\theta}_s \in \mathbb{R}^q} \psi(\boldsymbol{\theta}_s)$. Then, the parameter-estimation, $\hat{\boldsymbol{\theta}}_s$, converges to $\min_{\boldsymbol{\theta}_s \in \mathbf{\Omega}} \psi(\boldsymbol{\theta}_s)$.

The results of Lemma 4.1 can be interpreted as follows. First, the state-estimation converges to zero independently of the parameter-estimation. Thus, the state-estimation is decoupled from the parameter-estimation. This is a stronger result than the one obtained in [8], where the presence of the regularization term introduced a bias in the state and parameter-estimation. Second, even though the regressor vector, $\mathbf{B\Theta}(\mathbf{x},\mathbf{u})$ is not persistently exciting, the parameter-estimation converges to a value that depends on the strictly convex function ψ.

This work proposes implementing the l_1 norm as the function ψ in order to promote sparsity in the parameter-estimation solution. Specifically, $\psi = \|\hat{\boldsymbol{\theta}}_s\|_1$. Next section will show that, under some structural assumptions on the library $\mathbf{\Theta}$, the parameter-adaptation (14) is able to identify the non-zero components of $\boldsymbol{\theta}_s$, and converge to its true value if the original regressor is persistently exciting.

Remark 4.1: The l_1 norm is not strictly convex, which prevents V_ψ to be a Lyapunov function. For this reason, ψ is implemented as the norm $\|\hat{\boldsymbol{\theta}}_s\|_p$ with $p = 1+\epsilon$, where ϵ is small. This is the closest strictly convex norm to the l_1 regularization.

It should be remarked that the inverse of the Hessian in (14), may not be computable for all p norms. For this reason, the function ψ is implemented as the squared p norm $\frac{1}{2}\|\hat{\boldsymbol{\theta}}\|_p^2$, with p close to one. The motivation behind such decision is that the Jacobian of the squared p norm has an analytical inverse, which allows the adaptation (14) to be implemented through the following dynamics,

$$\dot{\mathbf{w}} = \mathbf{\Theta}^\intercal(\hat{\mathbf{x}},\mathbf{u})\mathbf{M}(\mathbf{y}-\mathbf{C}\hat{\mathbf{x}}) \tag{15}$$

$$\hat{\theta}_{s,i} = \|\mathbf{w}\|_d^{2-d}|w_i|^{d-1}sign(w_i) \text{ for } i = 1,...,q \tag{16}$$

where $\mathbf{w} \triangleq \nabla\psi(\hat{\boldsymbol{\theta}}_s)$, $\|\cdot\|_d$ is the d-norm and $\frac{1}{d}+\frac{1}{p}=1$.

V. Performance of the l_1 Minimization

The aim of this section is to study under which circumstances, the parameter-adaptation is capable of estimating the non-zero components of $\boldsymbol{\theta}_s$ and its values. In order to generalize the results, this section will consider the case when there is noise in the measured signal, i.e. $\mathbf{v} \neq 0$.

The state-estimation converges to an ultimate bound that depends on the sensor noise's 2-norm upper bound, v_2. Then, as the equation $\dot{\mathbf{e}}_x$ is Lipschitz, it is possible to see that the 2-norm of the function $\mathbf{B\Theta}(\mathbf{x},\mathbf{u})\mathbf{e}_{\theta_s}$ is ultimately bounded by a function $\gamma(v_2)$. Therefore, $\forall t > t_s$ where t_s is a sufficiently large positive value, the natural gradient-like parameter-adaptation (15)-(16) is solving the following optimization problem

$$arg\min_{\hat{\boldsymbol{\theta}}_s} \|\hat{\boldsymbol{\theta}}_s\|_1 \quad s.t.\ \|s - \mathbf{B\Theta}(\mathbf{x},\mathbf{u})\hat{\boldsymbol{\theta}}_s\|_2 \le \gamma(v_2)\ \forall t > t_s. \tag{17}$$

The error-constrained l_1 minimization (17) has been deeply studied in the computer science field in the context of sparse signal recovery problems [6], where the recovery of the original vector $\boldsymbol{\theta}$ can be proved by taking some assumptions on the library $\mathbf{B\Theta}(\mathbf{x},\mathbf{u})$. In particular, the recovery can be proved by studying the mutual incoherence of the library, i.e. if the pairwise correlations among the columns of $\mathbf{B\Theta}(\mathbf{x},\mathbf{u})$ are small, the sparsest solution of (17) is unique [6]. Then, if the original vector is persistently exciting, the sparsest solution is in the span of $\boldsymbol{\theta}$.

VI. Numerical Simulation

The statements presented in this document are exemplified through a numerical simulation, in which the performance of a standard adaptive observer and the proposed sparsity-promoting observer are compared.

This section considers the autonomous Van der Pol oscillator, the dynamics of which are depicted by the following expression

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{B}\boldsymbol{\phi}(\mathbf{x})\boldsymbol{\theta} \tag{18}$$

where $\mathbf{x} = [x_1, x_2]^\intercal$, $\boldsymbol{\theta} = [-0.2, 1, -0.3]^\intercal$ and

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} x_2 \\ 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{\phi}(\mathbf{x}) = \left[x_1, x_2, x_1^2 x_2\right].$$

The system is initialized at $\mathbf{x}(0) = [1, 0]^\intercal$, which produces the oscillatory dynamics that makes the vector $\mathbf{B}\boldsymbol{\phi}(\mathbf{x})$ persistently exciting.

It is assumed that the vector $\boldsymbol{\phi}(\mathbf{x})$ is unknown. Instead, the following library of non-linear function candidates is considered

$$\begin{aligned}\mathbf{\Theta}(\mathbf{x}) = [&1,\ x_1,\ x_2,\ x_1x_2,\ x_1^2x_2,\ x_1x_2^2,\ x_1^2x_2^2,\ x_1^2,\ x_2^2,\\ &sin(x_1),\ sin(x_2),\ cos(x_1),\ cos(x_2)\\ &sin(x_1)cos(x_2),\ cos(x_1)sin(x_2)].\end{aligned}$$

This library satisfies Assumption 2.2. Thus, system (18) can be rewritten as follows

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{B\Theta}(\mathbf{x})\boldsymbol{\theta}_s$$

where $\boldsymbol{\theta}_s = [\theta_{s,1}, ..., \theta_{s,15}]^\intercal$ is a vector of zeros except for

$$\theta_{s,2} = -0.2, \quad \theta_{s,3} = 1, \quad \theta_{s,5} = -0.3.$$

Finally, it is assumed that the states are unknown, and only the signal $y = \mathbf{c}^\intercal\mathbf{x} = x_1 + x_2$ is being measured.

The objective is to design an adaptive observer that through the measurement of y can reconstruct the unknown state, $\mathbf{x}$, and parameters, $\boldsymbol{\theta}_s$. The parameter-estimation should be close enough to the true value so it is possible to identify the active non-linear functions of the library, $\mathbf{\Theta}$.

Following the insights presented in [10], it is possible to prove that the observer dynamics

$$\dot{\hat{\mathbf{x}}} = \mathbf{f}(\hat{\mathbf{x}}) + \mathbf{B}\boldsymbol{\phi}(\hat{\mathbf{x}})\hat{\boldsymbol{\theta}} + \mathbf{l}(y - \mathbf{c}^\intercal\hat{\mathbf{x}}) \tag{19}$$

with $\mathbf{l} = [-0.314,\ 3.156]^\intercal$, satisfy the Lyapunov condition (5).

Consequently, taking into account the insights presented in Section III, a natural approach to solve the concerned estimation problem is to couple the observer dynamics (19) with the parameter-adaptation (7) and design $M = 0.4781$ in order to satisfy (8).

Such adaptive observer, with adaptation (7), provides a relatively accurate estimation of the states. However, the standard adaptation dynamics cannot identify the active non-linear functions of $\mathbf{\Theta}$ and the estimation $\hat{\boldsymbol{\theta}}_s$ does not converge to the true value. This fact can be seen in Figure 1, where it is depicted the parameter-estimation evolution for the standard adaptive observer in the noiseless case ($\mathbf{v} = 0$). It is noticeable

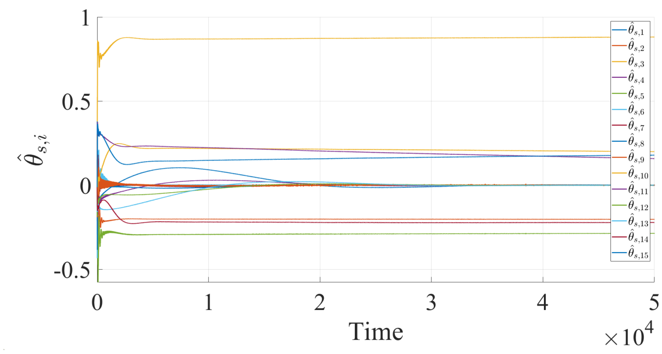

Fig. 1. parameter-estimation evolution through the standard adaptive observer. No sensor noise is considered $\mathbf{v} = 0$

that there are parameters, which are not $\theta_{s,2}, \theta_{s,3}$ or $\theta_{s,5}$, that converge to a non-zero value. Moreover, the parameters $\theta_{s,2}, \theta_{s,3}$ or $\theta_{s,5}$ do not converge to its true value.

Even though, the original regressor vector $\mathbf{B}\phi(\mathbf{x})$ is persistently excited, the considered library regressor, $\mathbf{B}\mathbf{\Theta}(\mathbf{x})$, is not. This fact exemplifies the standard adaptive observer's inability to recover the active functions of $\mathbf{\Theta}$. Nonetheless, the parameter vector, $\boldsymbol{\theta}_s$, is somewhat sparse. Thus, it is reasonable to think that the proposed sparsity-promoting observer should out-perform the standard adaptive observer, even in the presence of sensor noise.

Indeed, a second simulation has been designed for the same Van der Pol system. In this simulation the sensor is corrupted with Gaussian noise with variance 0.01. Moreover, the parameter-adaptation (7) is substituted by the natural gradient-like adaptation (15)-(16) with $p = 1.1$. As it can be seen, the sparsity-promoting adaptive observer achieves an accurate estimation of the system states, even in the presence of significant sensor noise, Figure 2.

Furthermore, in Figure 3 it is depicted the evolution of the parameter-estimation for the natural gradient-like adaptation (15)-(16). It can be seen that the algorithm recovers the active nonlinear functions of $\mathbf{\Theta}$, as the parameters related to the non-active functions converge to zero. Moreover, the relative error[1] between the estimation of $\theta_{s,2}, \theta_{s,3}, \theta_{s,5}$ and its true value converges below the 2%. There are some "spikes" in the parameter-estimation evolution that increase this error. These spikes are induced by the sudden changes in the excitation of the library, $\mathbf{\Theta}$, due to the abrupt oscillations of x_2 during the state evolution, see Figure 2. These spikes could be reduced by normalizing $\mathbf{B}\mathbf{\Theta}(\mathbf{x})$. However, in order to make the example more clear, this normalization step has been obviated.

VII. Conclusions

This work has presented a new approach for the design of adaptive observers for linearly parametrized systems. Instead of relying on a single regressor vector, which may be unknown, the observer implements a library of candidate non-linear functions and the adaptive algorithm selects the

[1]The relative error $[\%]$ between x and $\hat{x}$ is computed as $\dfrac{\|x - \hat{x}\|}{\|x\|} \cdot 100$

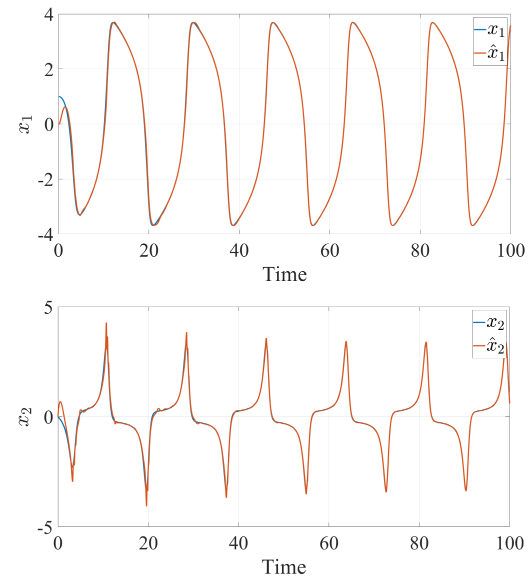

Fig. 2. Evolution of the system's states and sparsity-promoting adaptive observer estimation.

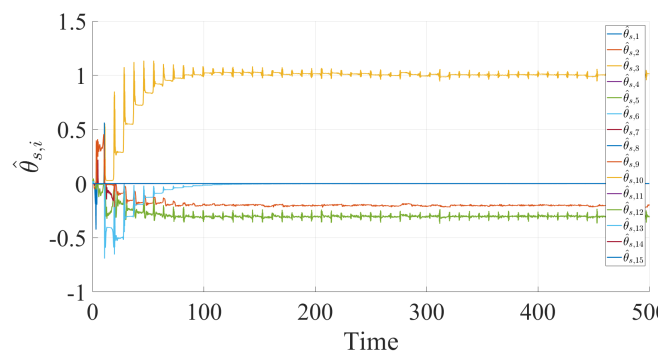

Fig. 3. parameter-estimation evolution through the sparsity-promoting adaptive observer. The estimations present some sudden spikes due to the excitation changes in the library $\mathbf{\Theta}$.

active functions that participate on the system. In general, this new regressor vector is not persistently exciting but results in a sparse unknown parameter vector. For this reason, this work proposed implementing an adaptive observer that exploits recent results in natural gradient-like adaptation [9] in order to promote sparsity in the parameter-estimation convergence. It has been shown that such adaptive observer converges to an error-constrained l_1 minimization, where the parameter-estimation error can be proved to converge through well-known sparse signal recovery theory, even in the presence of noise. The advantages of this adaptive observer have been validated through a numerical simulation.

Appendix I
Proof of Standard Adaptive Observer Convergence

The derivative of the Lyapunov function candidate (9) is negative semidefinite and $\|\mathbf{e}_x\| \to 0$ can be proved by the

Barbalat's lemma. This result shows that the state-estimation accuracy is decoupled of the parameter-estimation.

The second part of the proof shows the condition under which the parameter-estimation converges to the true value. As $\mathbf{x} \in C^1$, then

$$\int_0^\infty \dot{\mathbf{e}}_x dt = \lim_{t\to\infty} \mathbf{e}_x(t) - \mathbf{e}_x(0) = -\mathbf{e}_x(0) < \infty.$$

Moreover, as the functions $\mathbf{f}$ and $\boldsymbol{\phi}$ are Lipschitz, then $\dot{\mathbf{e}}_x$ is uniformly continuous. Therefore, by the Barbalat's lemma, $\dot{\mathbf{e}}_x \to 0$. Then, by considering the state-estimation error dynamics, the following can be deduced

$$\|\mathbf{B}\boldsymbol{\Theta}(\mathbf{x},\mathbf{u})\mathbf{e}_{\theta_s}\| \to 0.$$

From the fact that (9) is non-increasing and lower bounded by zero, it has a limit as $t \to \infty$. Therefore, $\hat{\boldsymbol{\theta}}_s$ and $\mathbf{e}_{\theta_s}$ must converge to a constant. Consequently, if the vector $\mathbf{B}\boldsymbol{\Theta}(\mathbf{x},\mathbf{u})$ is persistently exciting [12], then, the parameter-estimation error, $\|\mathbf{e}_{\theta_s}\|$, will also converge to zero.

APPENDIX II
PROOF LEMMA 4.1

First, the derivative of the Lyapunov function (13) satisfies the following

$$\begin{aligned}\dot{V}_\psi &\leq -k\alpha_3(\mathbf{e}_x) + \mathbf{e}_x^\intercal \mathbf{P}\mathbf{B}\boldsymbol{\Theta}(\hat{\mathbf{x}},\mathbf{u})\mathbf{e}_{\theta_s} - \mathbf{e}_{\theta_s}\nabla^2\psi(\hat{\boldsymbol{\theta}}_s)\dot{\hat{\boldsymbol{\theta}}}_s \\ &= k\alpha_3(\mathbf{e}_x) + \mathbf{e}_x^\intercal \mathbf{P}\mathbf{B}\boldsymbol{\Theta}(\hat{\mathbf{x}},\mathbf{u})\mathbf{e}_{\theta_s} - \mathbf{e}_{\theta_s}^\intercal \boldsymbol{\Theta}^\intercal(\hat{\mathbf{x}},\mathbf{u})\mathbf{M}\mathbf{C}\mathbf{e}_x \\ &= -k\alpha_3(\mathbf{e}_x).\end{aligned}$$

Then, according to Appendix A, $\mathbf{e}_x$ converges to zero and converges to a small positive value in a time $t_s > 0$.

The second part of the proof is based on the one presented by Boffi and Slotine in [9]. Being the main difference that this work considers a state dependent regressor vector.

Taking into account (14) and (8), the time derivative of the Bregman divergence is

$$\begin{aligned}&\frac{d}{dt}d_\psi(\boldsymbol{\theta}_s||\hat{\boldsymbol{\theta}}_s) = -\left(\frac{d}{dt}\nabla\psi(\hat{\boldsymbol{\theta}})\right)^\intercal \mathbf{e}_{\theta_s} = -\mathbf{e}_x^\intercal \mathbf{P}\mathbf{B}\boldsymbol{\Theta}(\hat{\mathbf{x}},\mathbf{u})\mathbf{e}_{\theta_s} \\ &= -\mathbf{e}_x^\intercal \mathbf{P}\mathbf{B}\boldsymbol{\Theta}(\mathbf{x},\mathbf{u})\mathbf{e}_{\theta_s} - \mathbf{e}_x^\intercal \mathbf{P}\mathbf{B}\left[\boldsymbol{\Theta}(\hat{\mathbf{x}},\mathbf{u}) - \boldsymbol{\Theta}(\mathbf{x},\mathbf{u})\right]\mathbf{e}_{\theta_s}.\end{aligned} \tag{20}$$

For all $t > t_s$, the last term in (20) is close to zero and monotonically decreasing. Consequently, can be neglected for a sufficiently large t_s. Following this line, the integration of the expression (20) $\forall t > t_s$ can be approximated as

$$d_\psi(\boldsymbol{\theta}_s||\hat{\boldsymbol{\theta}}_s(t_s)) = d_\psi(\boldsymbol{\theta}_s||\hat{\boldsymbol{\theta}}_s(t)) - \int_{t_s}^t \mathbf{e}_x^\intercal(\tau)\mathbf{P}\mathbf{B}\boldsymbol{\Theta}(\mathbf{x},\mathbf{u})\mathbf{e}_{\theta_s}(\tau)d\tau. \tag{21}$$

Appendix I has proved that $\hat{\boldsymbol{\theta}}_s$ and $\mathbf{e}_{\theta_s}$ converge to a constant. Thus, one can compute the limit as $t \to \infty$ of (21) and conclude that the following holds

$$\begin{aligned}d_\psi(\boldsymbol{\theta}_s||\hat{\boldsymbol{\theta}}_s(t_s)) &= d_\psi(\boldsymbol{\theta}_s||\hat{\boldsymbol{\theta}}_s(\infty)) \\ &+ \int_{t_s}^\infty \mathbf{e}_x^\intercal(\tau)\mathbf{P}(\mathbf{B}\boldsymbol{\Theta}(\mathbf{x},\mathbf{u})\hat{\boldsymbol{\theta}}_s(\tau) - s(\tau))d\tau.\end{aligned} \tag{22}$$

The right hand-side of (22) has a minimum over $\boldsymbol{\theta}_s$ at $\hat{\boldsymbol{\theta}}_s(\infty)$. In parallel, the minimum of the left-hand side factor is obtained in $arg \min_{\boldsymbol{\theta}_s\in\boldsymbol{\Omega}} d_\psi(\boldsymbol{\theta}_s||\hat{\boldsymbol{\theta}}_s(t_s))$. Therefore,

$$\hat{\boldsymbol{\theta}}_s(\infty) = \min_{\boldsymbol{\theta}_s\in\boldsymbol{\Omega}} d_\psi(\boldsymbol{\theta}_s||\hat{\boldsymbol{\theta}}_s(t_s)). \tag{23}$$

By considering that the parameter-estimation satisfies $\hat{\boldsymbol{\theta}}_s(t_s) = \min_{\boldsymbol{\theta}_s\in\mathbb{R}^q} \psi(\boldsymbol{\theta}_s)$, expression (23) reduces to

$$\hat{\boldsymbol{\theta}}_s(\infty) = \min_{\boldsymbol{\theta}_s\in\boldsymbol{\Omega}} \psi(\boldsymbol{\theta}_s).$$

REFERENCES

[1] D. Efimov, C. Edwards, and A. Zolghadri, "Enhancement of adaptive observer robustness applying sliding mode techniques," *Automatica*, vol. 72, pp. 53 – 56, 2016.

[2] G. Besançon, "Remarks on nonlinear adaptive observer design," *Systems & Control Letters*, vol. 41, no. 4, pp. 271 – 280, 2000.

[3] G. C. Goodwin, "A critique of observers used in the context of feedback control," in *Intelligent Robotics and Applications*, Z. Chen, A. Mendes, Y. Yan, and S. Chen, Eds. Cham: Springer International Publishing, 2018, pp. 1–24.

[4] A. Astolfi, D. Karagiannis, and R. Ortega, *Nonlinear and adaptive control with applications*. Springer Science & Business Media, 2007.

[5] U. Montanaro, R. Costa-Castelló, J. M. Olm, and C. L. Barros, "Experimental validation of a continuous-time mcsi algorithm with bounded adaptive gains," *Journal of the Franklin Institute*, vol. 356, no. 12, pp. 5881 – 5897, 2019.

[6] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.

[7] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657 – 682, 2014.

[8] Q. Zhang, F. Giri, and T. Ahmed-Ali, "Regularized adaptive observer to address deficient excitation," *IFAC-PapersOnLine*, vol. 52, no. 29, pp. 251 – 256, 2019, 13th IFAC Workshop on Adaptive and Learning Control Systems ALCOS 2019.

[9] N. M. Boffi and J.-J. E. Slotine, "Implicit Regularization and Momentum Algorithms in Nonlinearly Parameterized Adaptive Control and Prediction," *Neural Computation*, vol. 33, no. 3, pp. 590–673, 03 2021.

[10] R. Rajamani, W. Jeon, H. Movahedi, and A. Zemouche, "On the need for switched-gain observers for non-monotonic nonlinear systems," *Automatica*, vol. 114, p. 108814, 2020.

[11] A. Cecilia, M. Serra, and R. Costa-Castelló, "Nonlinear adaptive observation of the liquid water saturation in polymer electrolyte membrane fuel cells," *Journal of Power Sources*, vol. 492, p. 229641, 2021.

[12] K. S. Narendra and A. M. Annaswamy, "Persistent excitation in adaptive systems," *International Journal of Control*, 1987.

[13] A. N. Tikhonov, "On the solution of ill-posed problems and the method of regularization," in *Doklady Akademii Nauk*, vol. 151, no. 3. Russian Academy of Sciences, 1963, pp. 501–504.

[14] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 3932–3937, 2016.

[15] T. Lee, J. Kwon, and F. C. Park, "A natural adaptive control law for robot manipulators," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 1–9.

[16] L. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 3, pp. 200 – 217, 1967.