On state-estimation in weakly-observable scenarios and implicitly regularized observers

Andreu Cecilia, Ramon Costa-Castelló

***Abstract*— This work proposes a framework to design observers for systems that present low observability. It is shown that, in these scenarios, the estimation problem becomes ill-posed, which drastically limits the performance of standard observers, specially in the presence of noise. Consequently, this paper presents a method to design an observer that optimizes some potential function to be defined by the designer. This allows to implicitly regularize the estimation and recover a well-posed problem. The proposed technique is validated in a set of weakly-observable systems and the performance is compared with a common Kalman filter-like observer.**

I. Introduction

Economical and technical constraints commonly limit the number of sensors that can be introduced in a system. In this context, the observer design problem naturally emerges as it is required to obtain some unmeasured internal information from the measured outputs and inputs.

Prior to the observer design, it is necessary to study the possibility of reconstructing the unmeasured states through the measured signals, in the so-called observability analysis [1], [2]. This step is commonly limited to a "yes-or-no" analysis, which does not provide any quantitative potential function on how observable is the system [1], [3]. It is interesting to consider the degree of observability, as it is related to the observer energy that is required to distinguish between states [4], which has a direct impact in the noise sensitivity of the estimator.

It is well known that noise sensitivity is one of the factors that limits the performance of observer-based control applications [5]. For this reason, it is convenient to select and place the sensors in order to maximize observability [6]. Nonetheless, application constraints may limit the achievable system observability [7], [8], [9], [10]. In this context, it is of interest to develop observers that can operate in low-observability scenarios.

In this work, it is shown that the source of the conflict in low-observability systems is that the state-estimation becomes an ill-posed inverse problem as a consequence of the bad conditioning of the observability gramian. Consequently, it is proposed a framework to implicitly regularize [11] the estimation problem, which allows to have a reliable estimation even for ill-conditioned observability gramians.

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) and by the project DOVELAR (ref. RTI2018-096001-B-C32).

A. Cecilia and R. Costa-Castelló are with the Institut de Robòtica i Informòtica 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*)

The remainder of this paper is organized as follows. Section II formalizes the problem of estimation in weakly-observable scenarios and introduces how regularization may improve the performance of the observer. Section III presents the proposed implicitly regularized observer. Section IV validates the proposed technique in a set of weakly-observable systems in which a standard Kalman filter fails to estimate the states. Finally, some conclusions are drawn in Section V.

II. Problem Formulation

Consider a linear time varying (LTV) system, the dynamics of which are depicted by:

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}(t)\mathbf{x} + \mathbf{B}(t)\mathbf{u}\\ \mathbf{y} &= \mathbf{C}(t)\mathbf{x} + \mathbf{v}\end{aligned} \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n, \mathbf{u} \in \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^q$. The signals $\mathbf{y}$ and $\mathbf{u}$ are measured. Moreover, the matrices $\mathbf{A}(t) \in \mathbb{R}^{n\times n}, \mathbf{B}(t) \in \mathbb{R}^{n\times p}$ and $\mathbf{C}(t) \in \mathbb{R}^{q\times n}$ are assumed to be known and ensure a unique solution for every input, $\mathbf{u}$, and initial condition $\mathbf{x}(0)$. The factor $\mathbf{v}$ depicts unknown high-frequency sensor noise.

The objective is to design an observer that generates an estimation of the states, $\hat{\mathbf{x}}$, such that $\lim_{x\to\infty} \|\mathbf{x} - \hat{\mathbf{x}}\| = 0$. The main difference with the standard state-estimation problem is that, here, it is considered that some modes of the system are nearly unobservable, which are commonly referred to as *weakly-observable modes* [8][1]. This concept can be formalized through the observability gramian. Consider the gramian for a time $t \in [t_0, t_1]$:

$$\mathbf{M}(t, t_0) \triangleq \int_{t_0}^{t} \boldsymbol{\phi}(\tau, t_0)^{\mathsf{T}}\mathbf{C}(t)^{\mathsf{T}}\mathbf{C}(t)\boldsymbol{\phi}(\tau, t_0)d\tau, \tag{2}$$

where $\boldsymbol{\phi}(t, t_0)$ is the fundamental matrix computed as:

$$\dot{\boldsymbol{\phi}}(t, t_0) = \mathbf{A}(t)\boldsymbol{\phi}(t, t_0),\ \ \boldsymbol{\phi}(t_0, t_0) = \mathbf{I}_n \tag{3}$$

where $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ is the identity matrix, so that:

$$\mathbf{x}(t) = \boldsymbol{\phi}(t, t_0)\mathbf{x}(t_0).$$

Definition 2.1: System (1) is *completely observable* if there are some positive constants μ_1, μ_2 and T such that [12]:

$$\mu_1\mathbf{I} \leq \mathbf{M}(t, t_0) \leq \mu_2\mathbf{I}, \quad t \geq T. \tag{4}$$

[1]Not to confuse this idea with the concept of weak observability for nonlinear systems. In the nonlinear context, weak observability refers to the possibility of distinguish a state $\mathbf{x}$ from other states in a neighbourhood [3].

Definition 2.2: System (1) is *weakly-observable* if it is completely observable and the minimum singular value of the observability gramian, $\underline{\sigma}(\mathbf{M}(t, t_0))$, is close to zero [4].

The concept of complete observability depicts that the states of the system can be distinguished one from the other. Nonetheless, it does not give any insight on how "easy" is to distinguish between said states. In such context, the minimum singular value of $\mathbf{M}(t, t_0)$ may be used as an index of how close are some modes of the system to unobservability [4].

Although weakly-observable modes are still "observable" from a theoretical point of view, they require more "energy" in order to distinguish between states, which results in the necessity of higher observer gains. It is well known that the presence of sensor noise induces an upper bound on the acceptable observer gain [5]. Moreover, high gains may induce undesirable transient behaviours [13]. Therefore, weakly-observable modes may be practically unobservable in the presence of noise. This fact is exemplified in the following subsection.

A. State reconstruction in weakly-observable systems

Consider a system depicted by (1) in a time interval $\tau \in [t_0, t_1]$. Without loss of generality, assume that $\mathbf{u} = 0$. The evolution of the output signal $\mathbf{y}$ is depicted by:

$$\mathbf{y}(\tau) = \mathbf{C}(\tau)\boldsymbol{\phi}(\tau, t)\mathbf{x}(t) \quad \forall \tau \in [t_0, t_1]. \tag{5}$$

If one pre-multiplies (5) by $\boldsymbol{\phi}(\tau, t)^\intercal \mathbf{C}(\tau)^\intercal$ and integrates from t_0 to t the following is obtained:

$$\mathbf{M}(t, t_0)\mathbf{x}(t) = \int_{t_0}^{t} \boldsymbol{\phi}(\tau, t)^\intercal \mathbf{C}(\tau)^\intercal \mathbf{y}(\tau) d\tau$$

Therefore, the state at time t can be reconstructed by solving:

$$\hat{\mathbf{x}}(t) = \min_{\hat{\mathbf{x}}(t)} \left\| \mathbf{M}(t, t_0)\hat{\mathbf{x}}(t) - \mathbf{z}(t, t_0) \right\| \tag{6}$$

where $\mathbf{z}(t, t_0) = \int_{t_0}^{t} \phi(\tau, t)^\intercal \mathbf{C}(\tau)^\intercal \mathbf{y}(\tau) d\tau$, which clearly results in:

$$\mathbf{x}(t) = \mathbf{M}(t, t_0)^{-1}\mathbf{z}(t, t_0). \tag{7}$$

Remark 2.1: The iterative computation of (7) recovers the well-known Kalman filter-like observer of the form [12]:

$$\begin{aligned} \dot{\hat{\mathbf{x}}} &= \mathbf{A}(t)\hat{\mathbf{x}} + \mathbf{B}(t)\mathbf{u} + \mathbf{P}(t)^{-1}\mathbf{C}(t)^\intercal(\mathbf{y} - \mathbf{C}(t)\hat{\mathbf{x}}) \\ \dot{\mathbf{P}}(t) &= -\mathbf{P}(t)\mathbf{A}(t) - \mathbf{A}(t)^\intercal \mathbf{P}(t) + \mathbf{C}(t)^\intercal \mathbf{C}(t), \\ \mathbf{P}(0) &= \mathbf{P}(0)^\intercal > 0. \end{aligned} \tag{8}$$

where $\boldsymbol{P}$ converges to the observability gramian $\boldsymbol{M}(t, t_0)$.

Naturally, the precision of the inversion in (7) and its performance in the presence of noise is limited by the minimum singular value and the condition number of the observability gramian.

The low accuracy and convergence rate in weak observability scenarios has motivated the modification of (8) in order to improve the performance of the observer. In relation to the convergence rate, a common approach is based on resetting the covariance matrix $\mathbf{P}$, in which $\mathbf{P}(t_r) = \mathbf{P}(0)$ at some time t_r [14], which prevents $\mathbf{P}$ to become excessively large. Nonetheless, the reset may result in large estimation transients after t_r. A second approach is based on including some kind of forgetting factor in $\dot{\mathbf{P}}$ of (8) [15], which increases the importance of recent data over past data. This usually results in higher convergence rates. Nonetheless, the forgetting rate needs to be consistent with the noise statistics [16] and large forgetting rates may unstabilize the $\mathbf{P}$ dynamics. In relation to the measurement noise, it is common to estimate the noise statistics and adapt the observer accordingly [17].

In any case, such modifications only address the symptoms of the problem, but do not directly undertake on the source of the conflict, that is, the ill-conditioning of the observability gramian. In this direction, better conditioning may be achieved by implementing a proper model order reduction [7], [8], [18], [19], which only retains the most observable modes. Nonetheless, this approach may eliminate weakly-observable modes that may be of interest for the application. Or the result may still be weakly-observable.

This work approaches the problem from a different perspective. Instead of increasing the conditioning of the observability gramian, it is improved the conditioning of the optimization problem (6) by introducing the proper regularization.

B. Explicitly regularized state-estimation

In weak observability scenarios, the state-estimation problem becomes an ill-posed inverse problem. In such context, it is common to explicitly regularize the optimization problem [20]. Indeed, instead of (6), consider the L_2 regularized optimization:

$$\hat{\mathbf{x}}(t) = \min_{\hat{\mathbf{x}}(t)} \left\| \mathbf{M}(t, t_0)\hat{\mathbf{x}}(t) - \mathbf{z}(t, t_0) \right\| + \hat{\mathbf{x}}^\intercal \boldsymbol{\Pi} \hat{\mathbf{x}}. \tag{9}$$

The solution of (9) is:

$$\hat{\mathbf{x}} = \left(\mathbf{M}(t, t_0)^\intercal \mathbf{M}(t, t_0) + \boldsymbol{\Pi} \right)^{-1} \mathbf{M}(t, t_0)^\intercal \mathbf{z}(t, t_0). \tag{10}$$

In this case, even a small value $\boldsymbol{\Pi}$ improves the conditioning of the problem and significantly reduces the noise sensitivity of the observer. This fact motivates the inclusion of regularization factors in the observer design.

However, there are some concerns that deprives the direct implementation of (9).

- Unlike standard optimization problems, regularization factors cannot be implemented in observers without having an impact in stability. As the inclusion of the factor $\boldsymbol{\Pi}$ introduces a bias in the state estimation.
- The optimal parameter $\boldsymbol{\Pi}$, in the sense of minimizing the bias and covariance of (9), depends on the unknown states. This fact was proven in [21] for the parameter-estimation case.
- There may be alternative regularizations that are more beneficial for the considered problem.

For this reason, this work proposes an alternative approach in order to regularize the problem (6). First, the state-estimation problem is transformed into an equivalent parameter-estimation problem, following recent ideas in parameter-based observer design [22]. Second, based on recent results of adaptive control and natural descent-like adaptation [23] [11], the resulting parameters are estimated through a modified recursive least squares that respects an underlying Riemannian geometry to be specified. This will allow to have an implicit regularization of the state-estimation problem that, under some conditions, recover a well-posed estimation problem.

III. Implicitly Regularized Observer

This subsection will present the insights related to the observer design.

A. Parameter-based observer

The aim of this subsection is to transform the original state-estimation problem into an equivalent parameter-estimation problem [22]. Consider the following dynamics:

$$\dot{\boldsymbol{\xi}} = \mathbf{A}(t)\boldsymbol{\xi} + \mathbf{B}(t)\mathbf{u}. \tag{11}$$

Define the signal $\mathbf{e}_x = \mathbf{x} - \boldsymbol{\xi}$, the dynamics of which are depicted by:

$$\dot{\mathbf{e}}_x = \mathbf{A}(t)\mathbf{e}_x.$$

Consequently, the time evolution of the signal, $\mathbf{e}_x$, can be computed through the fundamental matrix (3) as follows:

$$\mathbf{e}_x(t) = \boldsymbol{\phi}(t,0)\mathbf{e}_x(0). \tag{12}$$

The key idea of the parameter-based observer is to take $\mathbf{e}_x(0) = \boldsymbol{\theta}$ as a constant parameter to be estimated, $\hat{\boldsymbol{\theta}}$ [22]. Then, the state can be reconstructed as follows:

$$\hat{\mathbf{x}}(t) = \xi(t) + \boldsymbol{\phi}(t,0)\hat{\boldsymbol{\theta}}.$$

Now, consider the variable $\tilde{\mathbf{y}}$ as:

$$\tilde{\mathbf{y}} \triangleq \mathbf{y} - \mathbf{C}\boldsymbol{\xi}.$$

Then, the following signal is obtained:

$$\tilde{\mathbf{y}} = \boldsymbol{\Psi}\boldsymbol{\theta} \tag{13}$$

where $\boldsymbol{\Psi} = \mathbf{C}\boldsymbol{\phi}(t,0)$.

Notice that the equation in (13) can be though as a linear regression in the parameters, thus, the unknown parameters can be estimated through linear identification algorithms. Indeed, the least-squares method results in the following parameter-estimation [24]:

$$\hat{\boldsymbol{\theta}}(t) = \left[\int_0^t \boldsymbol{\Psi}\boldsymbol{\Psi}^\intercal d\tau\right]^{-1} \int_0^t \boldsymbol{\Psi}^\intercal \tilde{\mathbf{y}} d\tau. \tag{14}$$

It is remarkable that the first integral in the right-hand side of (14) is the observability gramian (2). Therefore, in the proposed parameter-based state-estimation context, the least-squares algorithm performance is also limited by the conditioning of the observability gramian.

Following the insights presented in Section II. Next subsection will propose a modified parameter-estimation that is consistent with an underlying potential function to be defined. This will allow to have implicit regularization that improves the conditioning of the state-estimation problem with provable performance and convergence of the observer.

B. Natural gradient descent-like parameter-estimation

In the scenario of weak observability with noise, there are multiple state values that are consistent with the measured outputs. Specifically, the parameter-estimation in (14) converges to a set:

$$\boldsymbol{\Omega} \triangleq \{\hat{\boldsymbol{\theta}} \mid \|\tilde{\mathbf{y}} - \boldsymbol{\Psi}\hat{\boldsymbol{\theta}}\| \leq \gamma(\mathbf{v})\}, \tag{15}$$

where γ is a positive function.

Then, we can stablish the following result.

Lemma 3.1: Assume that the system (1) is completely observable. Consider a strongly convex function $\psi(\cdot)$ and the natural gradient-like adaptation law:

$$\begin{aligned} \dot{\hat{\boldsymbol{\theta}}} &= \left[\nabla^2\psi(\hat{\boldsymbol{\theta}})\right]^{-1} \mathbf{P}\boldsymbol{\Psi}^\intercal \left(\tilde{\mathbf{y}} - \boldsymbol{\Psi}\hat{\boldsymbol{\theta}} + \mathbf{v}\right) \\ \dot{\mathbf{P}} &= -\mathbf{P}\boldsymbol{\Psi}^\intercal\boldsymbol{\Psi}\mathbf{P}, \quad \mathbf{P}(0) = \mathbf{P}(0)^\intercal > 0. \end{aligned} \tag{16}$$

Moreover, consider that $\hat{\boldsymbol{\theta}}(0) = \min_{\boldsymbol{\theta}\in\mathbb{R}^q} \psi(\boldsymbol{\theta})$. Then, the parameter-estimation, $\hat{\boldsymbol{\theta}}$, converges to $\min_{\boldsymbol{\theta}\in\boldsymbol{\Omega}} \psi(\boldsymbol{\theta})$.

Proof: This proof is based on the results of [11]. The first step of the proof consists in showing that (16) converges to Ω. Define the Bregman divergence of a strictly convex function ψ as [25]:

$$d_\psi(\mathbf{a}||\mathbf{b}) \triangleq \psi(\mathbf{a}) - \psi(\mathbf{b}) - (\mathbf{a}-\mathbf{b})^\intercal \nabla\psi(\mathbf{b}).$$

Consider the radially unbounded Lyapunov function:

$$V = d(\boldsymbol{\theta}||\hat{\boldsymbol{\theta}}). \tag{17}$$

Define $\mathbf{e}_\theta \triangleq \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$. The derivative of (17) satisfies the following:

$$\begin{aligned} \dot{V} &= \mathbf{e}_\theta^\intercal \nabla^2\psi(\hat{\boldsymbol{\theta}})\dot{\hat{\boldsymbol{\theta}}} \\ &= -\mathbf{e}_\theta^\intercal \nabla^2\psi(\hat{\boldsymbol{\theta}})\left[\nabla^2\psi(\hat{\boldsymbol{\theta}})\right]^{-1} \mathbf{P}\boldsymbol{\Psi}^\intercal\left(\tilde{\mathbf{y}} - \boldsymbol{\Psi}\hat{\boldsymbol{\theta}} + \mathbf{v}\right) \\ &\leq -\alpha_1 \mathbf{e}_\theta^\intercal \boldsymbol{\Psi}^\intercal \left(\tilde{\mathbf{y}} - \boldsymbol{\Psi}\hat{\boldsymbol{\theta}}\right) + \alpha_2 \mathbf{e}_\theta^\intercal \boldsymbol{\Psi}^\intercal \mathbf{v} \\ &\leq -\alpha_1 \left(\tilde{\mathbf{y}} - \boldsymbol{\Psi}\hat{\boldsymbol{\theta}}\right)^\intercal \left(\tilde{\mathbf{y}} - \boldsymbol{\Psi}\hat{\boldsymbol{\theta}}\right) + \alpha_2 \left(\tilde{\mathbf{y}} - \boldsymbol{\Psi}\hat{\boldsymbol{\theta}}\right)^\intercal \mathbf{v} \end{aligned}$$

where

$$\alpha_1 \mathbf{I} \leq \mathbf{P} \leq \alpha_2 \mathbf{I}.$$

By means of (4) it can be shown that for $t > T$, $\alpha_1 > 0$ and $\alpha_2 < \infty$. Therefore, the system is stable in a input-to-state sense [26] with respect to the noise, which proves the convergence to the set Ω.

Second, it will be shown that the gradient-like dynamics (16) implicitly optimize the convex function $\psi(\hat{\boldsymbol{\theta}})$.

The $d(\boldsymbol{\theta}||\hat{\boldsymbol{\theta}})$ factor presents the following time derivative:

$$\begin{aligned} \frac{d}{dt} d_\psi(\boldsymbol{\theta}||\hat{\boldsymbol{\theta}}) &= -\left(\frac{d}{dt}\nabla\psi(\hat{\boldsymbol{\theta}})\right)^\intercal \mathbf{e}_\theta = \\ &\quad -\left(\tilde{\mathbf{y}} - \boldsymbol{\Psi}\hat{\boldsymbol{\theta}} + \mathbf{v}\right)^\intercal \boldsymbol{\Psi}\mathbf{P}\mathbf{e}_\theta. \end{aligned} \tag{18}$$

The integration of (18) results in:

$$d_\psi(\boldsymbol{\theta}||\hat{\boldsymbol{\theta}}(0)) = d_\psi(\boldsymbol{\theta}||\hat{\boldsymbol{\theta}}(t)) + \int_0^t \left(\tilde{\mathbf{y}} - \boldsymbol{\Psi}\hat{\boldsymbol{\theta}} + \mathbf{v}\right)^\intercal \boldsymbol{\Psi}\mathbf{P}\mathbf{e}_\theta d\tau.$$

Previously, it has been shown that the parameter-estimation converges to the set $\hat{\boldsymbol{\theta}}(t) \in \boldsymbol{\Omega}$. Consequently, one can show that in the limit $t \to \infty$, or $\boldsymbol{\theta} \in \boldsymbol{\Omega}$ and $\hat{\boldsymbol{\theta}}(\infty) \in \boldsymbol{\Omega}$, the next condition is satisfied:

$$\begin{aligned} d_\psi(\boldsymbol{\theta}||\hat{\boldsymbol{\theta}}(0)) &= d_\psi(\boldsymbol{\theta}||\hat{\boldsymbol{\theta}}(\infty)) \\ &+ \int_0^\infty \left(\tilde{\mathbf{y}} - \boldsymbol{\Psi}\hat{\boldsymbol{\theta}} + \mathbf{v}\right)^\intercal \mathbf{P}\left(\tilde{\mathbf{y}} - \boldsymbol{\Psi}\hat{\boldsymbol{\theta}}\right) d\tau. \end{aligned} \quad (19)$$

Only the first term in the right side of (19) depends on $\boldsymbol{\theta}$. Consequently, the minimum value over $\boldsymbol{\theta}$ of the right hand side of (19) is found at $\hat{\boldsymbol{\theta}}(\infty)$. Moreover, the minimum of the left-hand side factor is naturally obtained in $arg \min_{\boldsymbol{\theta}\in\boldsymbol{\Omega}} d_\psi(\boldsymbol{\theta}||\hat{\boldsymbol{\theta}}(0))$. Therefore,

$$\hat{\boldsymbol{\theta}}(\infty) = \min_{\boldsymbol{\theta}\in\boldsymbol{\Omega}} d_\psi(\boldsymbol{\theta}||\hat{\boldsymbol{\theta}}(0)). \quad (20)$$

If the observer states are initialized in $\hat{\boldsymbol{\theta}}(0) = \min_{\boldsymbol{\theta}\in\mathbb{R}^q} \psi(\boldsymbol{\theta})$, the equation in (20) becomes:

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

■

Lemma 3.1 establishes that for a completely observable system and in the absence of noise, the parameter-estimation dynamics (16) converges to the true value. In the weakly-observable scenario with noise, the estimation converges to the value $\hat{\boldsymbol{\theta}} \in \Omega$ that minimizes the convex function $\psi(\cdot)$ to be defined. Intuitively, the function $\psi(\cdot)$ is a potential function that can be implemented to exploit additional prior information to ensure that the state-estimation problem is well-posed. Some examples of prior information that can be exploited in this context may be physical consistency conditions of the system [23], physical bounds on the states [27], non-negativeness of the system [28] or an L_p norm of the states to be minimized [11]. The numerical example of this work focuses on the latter. In such case, the factor $\left[\nabla^2\psi(\hat{\boldsymbol{\theta}})\right]^{-1}$ in (16) may not be known for some p norms. Consequently, the potential function ψ is implemented as, $\frac{1}{2}\|\hat{\boldsymbol{\theta}}\|_p^2$. The benefit of this modification is that the Jacobian of a squared p norm has an analytical inverse, which allows the adaptation (16) to be computed as follows [29],

$$\dot{\mathbf{w}} = \mathbf{P}\boldsymbol{\Psi}^\intercal\left(\tilde{\mathbf{y}} - \boldsymbol{\Psi}\hat{\boldsymbol{\theta}}\right) \quad (21)$$

$$\hat{\theta}_i = \|\mathbf{w}\|_d^{2-d}|w_i|^{d-1} sign(w_i) \ for \ i = 1, ..., q \quad (22)$$

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

IV. Numerical Simulation

The benefits of the approach have been validated in a set of synthetic examples.

A. Example 1: State and parameter-estimation

Consider the autonomous LTV system of the form (1) with the state vector $\mathbf{x} = [x_1, x_2]^\intercal$ and the matrices:

$$\mathbf{A}(t) = \begin{bmatrix} 0 & e^{-0.9t} \\ 0 & 0 \end{bmatrix}; \mathbf{b}(t) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \quad \mathbf{c} = \begin{bmatrix} 1 & 0 \end{bmatrix}^\intercal \quad (23)$$

The measured output is corrupted with high-frequency noise of variance $1.974 \cdot 10^{-4}$.

The objective is to design an observer for the estimation of the unknown states. The conflict is that, although the system is completely observable, the system observability reduces with time and the estimation problem becomes ill-posed, i.e. $\lim_{t_0\to\infty} \underline{\sigma}(\mathbf{M}(t, t_0)) = 0$. This fact is the result of the observability study included in the Appendix I.

As the system is completely observable, it is convenient to estimate the states through a Kalman filter of the form (8). However, as the system loses observability with time, the "pure" Kalman filter cannot converge to the true value. A reasonable strategy could be based on including a forgetting rate-like factor in the $\mathbf{P}$ dynamics of (8) [15]. Specifically, consider the following modified $\mathbf{P}$ dynamics:

$$\dot{\mathbf{P}} = -\mu\mathbf{P} - \mathbf{P}(t)\mathbf{A}(t) - \mathbf{A}(t)^\intercal\mathbf{P}(t) + \mathbf{C}(t)^\intercal\mathbf{C}(t). \quad (24)$$

where $\mu > 0$.

The convergence rate of the state-estimation can be improved by incrementing the factor μ, which increments the gain of the observer [15]. The aim of this modification is to achieve a somewhat reliable estimation before the system loses observability. Nonetheless, as explained in Section II, noise and the smallest singular value of the observability gramian (2) induce an upper bound in the acceptable observer gain. Therefore, this strategy is not applicable in weakly-observable systems. Indeed, in the considered problem, a Kalman filter with factor $\mu = 0.9$ does not converge to the true value of x_2 and is practically useless in the presence of noise, see Figure 1.

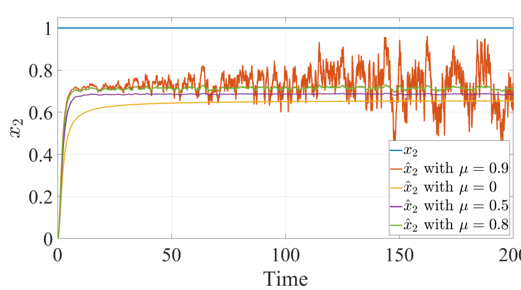

Fig. 1. True evolution of the state x_2 and Kalman filter estimation with multiple factors μ.

Alternatively, one can exploit some properties of the system to design a potential function $\psi(\cdot)$ that regularizes the estimation problem. In the considered system, an observer can be initialized as $\mathbf{C}\hat{\mathbf{x}}(0) = \mathbf{C}\mathbf{x}(0)$. In such case, the following is satisfied for (13):

$$\mathbf{C}\boldsymbol{\theta} = 0. \quad (25)$$

Condition (25) means that, in the parameter-based observer approach, under the proper observer initialization, the parameter vector to be estimated presents multiple zeros. Following ideas of compressed sensing and sparse signal recovery [30][31], it can be shown that an L_1 regularization factor improves the conditioning of an inverse problem if the parameter vector to be estimated is somewhat sparse. Consequently, an implicitly regularized observer that optimizes the potential function $\psi(\hat{\boldsymbol{\theta}}) = \|\hat{\boldsymbol{\theta}}\|_1$, where $\|\cdot\|_1$ is the 1-norm, should present better convergence rate and estimation accuracy.

Remark 4.1: Notice that the 1-norm is not strictly convex, which prevents (17) to be a Lyapunov function. For this reason, ψ is implemented as the norm $\frac{1}{2}\|\hat{\boldsymbol{\theta}}\|_p^2$ with $p = 1+\epsilon$, where ϵ is a small positive constant. This is the closest strictly convex function to the 1-norm.

In Figure 2 it is depicted the estimation through a Kalman filter-like observer with $\mu = 0.8$ and the estimation of an implicitly regularized observer with potential function $\frac{1}{2}\|\hat{\boldsymbol{\theta}}\|_{1.1}$ computed through (21)-(22). During the large observability transient, the implicitly regularized observer presents a similar convergence rate as the Kalman filter-like observer. This fact validates the performance of the observer in non-weakly-observable scenarios. Moreover, during the weak observability time range, the proposed observer still maintains its convergence to the true value without relying on high observer gains, consequently, it also presents significant lower noise sensitivity.

Furthermore, the structure of the observer (16), allows to introduce common modifications in least squares estimation. For example, it is possible to introduce a forgetting rate-like factor in the $\mathbf{P}$ dynamics in (16) as [24]:

$$\dot{\mathbf{P}} = \mu\mathbf{P} - \mathbf{P}\boldsymbol{\Psi}^\intercal\boldsymbol{\Psi}\mathbf{P}, \quad \mathbf{P}(0) = \mathbf{P}(0)^\intercal > 0,\ \mu > 0$$

in order to improve the convergence rate of observer at a cost of higher noise sensitivity. The benefits of this modification are shown in Figure 2.

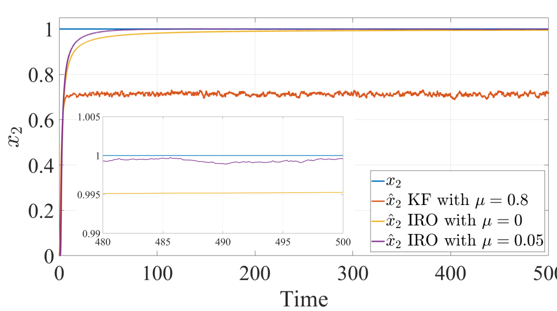

Fig. 2. True evolution of the state x_2, Kalman filter (KF) estimation with factor $\mu = 0.8$ and implicitly regularized observer (IRO) estimation with multiple forgetting rates μ.

B. Example 2: Slow and fast dynamics

Another example of a weakly-observable system appear in systems with fast and slow modes. Specifically, when only the slow modes are being measured. In such context, the condition number of the observability gramian may become significantly large, which affect the conditioning of the state-estimation problem. As an example, consider a linear time invariant system with:

$$\mathbf{A} = \begin{bmatrix} -0.1 & 0.4 & 0 & 0 \\ 0 & 0 & 0.2345 & 0 \\ 0 & -5.24 & -4.65 & 2.62 \\ 1 & 0 & 0 & -10 \end{bmatrix}; \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix};$$

$$\mathbf{c} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^\intercal \tag{26}$$

The measured output is corrupted with high-frequency noise of variance 0.01.

The faster eigenvalue of $\mathbf{A}$ is -10 and the slowest is -0.0724, thus, the system states evolves in two clearly different time-scales.

In this system, the rank of the observability map is 4, which proves that the system is observable. Nonetheless, the minimum singular value of the observability gramian is $1.1544 \cdot 10^{-9}$. Consequently, in the considered scenario, the gain of the standard Kalman filter (8), $\mathbf{P}(t)^{-1}\mathbf{C}^\intercal$, presents a component of the order 10^9, which prevents the implementation of this type of observer.

Alternatively, it is possible to design an implicitly regularized observer that exploits the sparsity condition (25), as explained in the last example. In Fig. 3 it is depicted the evolution of the observer estimation error. It can be seen that the observer achieves a fairly accurate estimation even in the presence of significant sensor noise.

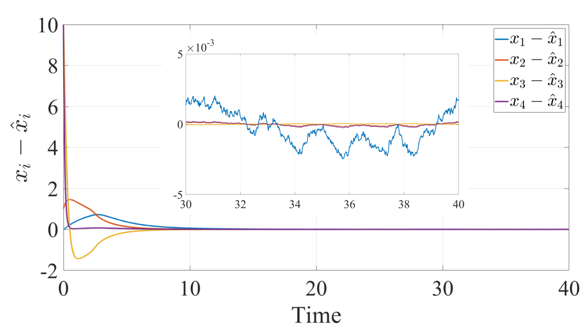

Fig. 3. Estimation errors evolution of the implicitly regularized observer with $p = 1.1$ and $\mu = 0.2$.

V. Conclusions

This work has proposed an observer design framework to address the state-estimation problem in weakly-observable systems. The proposed technique is based in an adaptation law that optimizes a potential function to be defined by the designer. This allows to implicitly regularize the estimation and, under some conditions, recover a well-posed problem. The proposed technique is based on, first, transforming the state-estimation problem in an equivalent parameter-estimation problem. Second, estimate the parameters with a natural gradient descent-like adaptation that respects the underlying Riemannian geometry related with the designed potential function. The results have been validated in a set of synthetic systems with low observability and the technique has been compared with a standard Kalman filter-like observer. It has been shown that the proposed observer presents positive results by exploiting a sparsity property of the observer problem.

APPENDIX I
OBSERVABILITY STUDY

The fundamental matrix of system (23) in the interval $[t_0, t]$ is:

$$\phi(t,t_0) = \begin{bmatrix} 1 & -\frac{1}{0.9}\left(e^{-0.9t} - e^{-0.9t_0}\right) \\ 0 & 1 \end{bmatrix}.$$

Consequently, the observability gramian (2) of the system is:

$$\mathbf{M}(t,t_0) = \begin{bmatrix} t - t_0 & m_{12}(t)e^{-0.9t_0} \\ m_{12}(t)e^{-0.9t_0} & m_{22}(t)e^{-1.8t_0} \end{bmatrix}.$$

where

$$m_{12}(t) = \frac{1}{0.9}\left[t - t_0 - \frac{1}{0.9}\left(1 - e^{-0.9(t-t_0)}\right)\right]$$
$$m_{22}(t) = \frac{1}{0.9^2}\bigg(- \frac{1}{1.8}e^{-1.8(t-t_0)} + \frac{1}{0.45}e^{-0.9(t-t_0)} + \frac{1}{1.8} - \frac{1}{0.45} + t - t_0\bigg).$$

The eigenvalues of the observability gramian are:

$$\lambda_{1,2} = \frac{1}{2}\bigg[t + m_{22}e^{-1.8t_0} \pm \sqrt{(t - t_0 - m_{22}e^{-1.8t_0}) + 4m_{12}^2 e^{-1.8t_0}}\bigg]$$

The eigenvalues are positive for all t_0 and $t > t_0$. Thus, the system is completely observable. However, it can be seen that:

$$\lim_{t_0\to\infty} \lambda_{min} = 0, \quad \lim_{t_0\to\infty} \lambda_{max} = t - t_0.$$

Therefore, the system becomes weakly-observable with time.

REFERENCES

[1] J. O'Reilly, *Observers for linear systems*. Academic press, 1983, vol. 170.

[2] P. Bernard, "Observer design for nonlinear systems," 2019.

[3] R. Hermann and A. Krener, "Nonlinear controllability and observability," *IEEE Trans. Autom. Control*, vol. 22, no. 5, pp. 728–740, Oct. 1977.

[4] A. J. Krener and K. Ide, "Measures of unobservability," in *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, 2009, pp. 6401–6406.

[5] 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.

[6] S. D. Bopardikar, "A randomized approach to sensor placement with observability assurance," *Automatica*, vol. 123, p. 109340, 2021.

[7] M. S. Grewal, L. R. Weill, and A. P. Andrews, *Global positioning systems, inertial navigation, and integration*. John Wiley & Sons, 2007.

[8] F. V. Lima, M. R. Rajamani, T. A. Soderstrom, and J. B. Rawlings, "Covariance and state estimation of weakly observable systems: Application to polymerization processes," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 4, pp. 1249–1257, 2013.

[9] 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.

[10] A. Cecilia, S. Sahoo, T. Dragicevic, R. Costa-Castello, and F. Blaabjerg, "On addressing the security and stability issues due to false data injection attacks in dc microgrids an adaptive observer approach," *IEEE Transactions on Power Electronics*, 2021, to appear.

[11] N. M. Boffi and J.-J. E. Slotine, "Implicit regularization and momentum algorithms in nonlinearly parameterized adaptive control and prediction," *Neural Computation*, pp. 1–84, 2021.

[12] R. Kalman and R. Bucy, "New results in linear filtering and prediction theory," *J. Basic Eng.*, pp. 83:95–108, 1961.

[13] H. K. Khalil and L. Praly, "High-gain observers in nonlinear feedback control," *Int. J. Robust Nonlinear Control*, vol. 24, no. 6, pp. 993–1015, 2014.

[14] S. Sastry and M. Bodson, *Adaptive control: stability, convergence and robustness*. Prentice-Hall, New Jersey, 1989.

[15] H. Hammouri and J. de Leon Morales, "Observer synthesis for state-affine systems," in *29th IEEE Conference on Decision and Control*, 1990, pp. 784–785 vol.2.

[16] A. L. Bruce, A. Goel, and D. S. Bernstein, "Convergence and consistency of recursive least squares with variable-rate forgetting," *Automatica*, vol. 119, p. 109052, 2020.

[17] Y. Huang, Y. Zhang, Z. Wu, N. Li, and J. Chambers, "A novel adaptive kalman filter with inaccurate process and measurement noise covariance matrices," *IEEE Trans. Autom. Control*, vol. 63, no. 2, pp. 594–601, 2018.

[18] M. Farhood and G. E. Dullerud, "On the balanced truncation of ltv systems," *IEEE Trans. Autom. Control*, vol. 51, no. 2, pp. 315–320, 2006.

[19] A. Astolfi, G. Scarciotti, J. Simard, N. Faedo, and J. V. Ringwood, "Model reduction by moment matching: Beyond linearity a review of the last 10 years," in *2020 59th IEEE Conference on Decision and Control (CDC)*, 2020, pp. 1–16.

[20] 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.

[21] T. Chen, H. Ohlsson, and L. Ljung, "On the estimation of transfer functions, regularizations and gaussian processes—revisited," *Automatica*, vol. 48, no. 8, pp. 1525–1535, 2012.

[22] R. Ortega, A. Bobtsov, N. Nikolaev, J. Schiffer, and D. Dochain, "Generalized parameter estimation-based observers: Application to power systems and chemical–biological reactors," *Automatica*, vol. 129, p. 109635, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0005109821001552

[23] 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.

[24] P. A. Ioannou and J. Sun, *Robust adaptive control*. Prentice Hall, Upper Saddle River, NJ, 1996.

[25] 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. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0041555367900407

[26] Z. Jiang, A. Teel, and L. Praly, "Small-gain theorem for iss systems and applications," *Math. Control. Signal Sys.*, vol. 7, pp. 95–120, 06 1994.

[27] P. M. Wensing and J.-J. Slotine, "Beyond convexity—contraction and global convergence of gradient descent," *PLOS ONE*, vol. 15, no. 8, pp. 1–29, 08 2020.

[28] A. Berman, M. Neumann, R. J. Plemmons, and R. J. Stern, *Nonnegative matrices in dynamic systems*. Wiley-Interscience, 1989, vol. 3.

[29] C. Gentile, "The robustness of the p-norm algorithms," *Machine Learning*, vol. 53, pp. 265–299, 2003.

[30] 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.

[31] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of

sparse overcomplete representations in the presence of noise," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, 2006.