Arbitrarily Fast Robust KKL Observer for Nonlinear Time-varying Discrete Systems

Gia Quoc Bao Tran, Pauline Bernard

Arbitrarily Fast Robust KKL Observer for Nonlinear Time-varying Discrete Systems

Gia Quoc Bao Tran, *Graduate Student Member, IEEE* and Pauline Bernard

***Abstract*— This work presents the Kazantzis-Kravaris/Luenberger (KKL) observer design for nonlinear time-varying discrete systems. We first give sufficient results on the existence of a sequence of functions $(\mathsf{T}_k)_{k\in\mathbb{N}}$ transforming the given system dynamics into an exponentially stable filter of the output in some other target coordinates, where an observer is directly designed. Then, we prove that under uniform Lipschitz backward distinguishability, the maps $(\mathsf{T}_k)_{k\in\mathbb{N}}$ become uniformly Lipschitz injective after a certain time, if the target dynamics is pushed sufficiently fast. This leads to an arbitrarily fast discrete observer, which exhibits similarities with the famous high-gain observer for continuous-time systems. Input-to-state stability of the estimation error with respect to uncertainties, input disturbances, and measurement noise is then shown. Next, under the milder backward distinguishability, we show the injectivity of the maps $(\mathsf{T}_k)_{k\in\mathbb{N}}$ after a certain time for a generic choice of the target filter dynamics. Examples including a discretized permanent magnet synchronous motor (PMSM) illustrate the proposed observer.**

***Index Terms*— KKL observer, discrete systems, time-varying systems.**

I. Introduction

OBSERVERS are algorithms developed for estimating the state of dynamical systems from their known outputs and inputs. Among many existing routes [1], the Kazantzis-Kravaris/Luenberger (KKL) observers [2]–[5] are of interest in nonlinear observer design thanks to their beautiful theory revolving around Coron's Lemma [2], [6]. These consist in transforming the system dynamics (of dimension n_x) into an exponentially stable filter of the output in some new coordinates (referred to as the target coordinates, of dimension $n_z \geq n_x$), where an observer readily exists, and inverting this transformation to recover the estimate of the state in the original coordinates. This design then translates into the following three main questions:

- Under what conditions does such a transformation *exist*?
- Under what conditions is the transformation *uniformly injective*?
- How to find an *explicit* and *implementable* expression of this transformation, and more importantly, of its left inverse?

The injectivity property is indeed needed to find a left inverse of the transformation and thus guarantee stability and convergence in the system coordinates. The two main questions about existence and injectivity have been answered in the literature for several classes of systems. Initially, David Luenberger proposed this method for linear time-invariant (LTI) continuous systems in [7]—he showed that an invertible linear transformation into a stable filter of the output always exists as long as the given system is observable and the eigenvalues of the filter are picked different from those of the system. Several attempts were then made to extend this theory to nonlinear continuous systems. The existence of a nonlinear transformation was first considered in [8]–[10] in the analytic context and around an equilibrium point. Then, the localness was dropped following another perspective in [11] where a global existence result was proposed based on a strong observability assumption which unfortunately did not provide an indication of the necessary dimension of the filter. This problem was solved in [2] by proving the existence of an injective transformation under a mild *backward distinguishability* condition, for complex-valued filters of dimension n_x+1, with almost any choice of n_x+1 distinct *complex* eigenvalues and recently in [12] for almost any *real* diagonalizable filter of dimension $2n_x+1$, both applied to each output. Stronger *uniform* injectivity results were also obtained under *differential observability* conditions, in the case where the eigenvalues of the filter are pushed sufficiently fast [3]. In parallel, this KKL paradigm was also developed for non-autonomous continuous systems [4] and for autonomous discrete systems [5], under similar *backward distinguishability* and *differential observability* conditions. Existing KKL observer results for various system classes are in Table I at the end of this paper.

Regarding the third question about a constructive design, an explicit and exploitable expression of the transformation can be found in particular contexts such as parameter identification [13] or state/parameter estimation for electrical machines [14], [15]. But when an implementable expression for the transformation or its left inverse is not available, numerical approximation methods based on neural networks are being developed as in [16]–[19]. This aspect being a research direction in its own right, here we leave it aside and focus instead on the questions of existence and injectivity of the transformation in the context of nonlinear time-varying discrete systems.

In this case, assuming invertibility of the dynamics, we show that there exists a sequence of transformations transforming the dynamics into a discrete stable filter of the output. Under an appropriate *uniform Lipschitz backward distinguishability* property, this sequence of transformations is shown to become

The authors are with Centre Automatique et Systèmes (CAS), Mines Paris, Université PSL, 75006 Paris, France (e-mail: {gia-quoc-bao.tran,pauline.bernard}@minesparis.psl.eu).

uniformly Lipschitz injective when the filter has an appropriate dimension and is pushed sufficiently fast. Our observer combines two main features. First, it provides an *arbitrarily fast* convergence of the estimation error in the system coordinates, as soon as allowed by the distinguishability condition. Second, this KKL design allows us to filter the output and provides input-to-state stability (ISS) of the estimation error, with an explicit strict ISS Lyapunov function. Such a design may thus be seen as a discrete counterpart of the celebrated *high-gain observer* for continuous-time systems [20], which as far as we know does not exist for discrete systems (apart from discretizations of continuous high-gain observers [21]). Reviewing in more detail the literature on discrete-time estimators, our *uniform Lipschitz backward distinguishability* condition is the same as in [22, Definitions 3 and 4]. It requires that for some $m \in \mathbb{N}$, the map between a state and its m past outputs is *uniformly Lipschitz injective*. Such a property is widely exploited in the literature, including moving horizon state estimators [23]–[25] (known for their robustness with respect to modeling uncertainties and numerical errors [26]), or discrete (dead-beat) estimators based on the left inversion of this observability map, such as [27] with Newton algorithms, which provide instantaneous estimation as soon as enough output information is gathered, but have no filtering effects against measurement noise.

Forgetting about the condition of uniformity (in time), this distinguishability property was shown to be generic for $m = 2n_x + 1$ in [28] (and the references therein) when the number of outputs is larger than the number of inputs. Note that relaxing further the Lipschitzness and the uniformity in m leads to a weaker distinguishability condition similar to [29, Definition 3], which we show guarantees injectivity of the KKL transformations, but not *uniform* injectivity, thus preventing us from stating any convergence result.

In the linear context, the *uniform Lipschitz backward distinguishability* turns out to coincide with Kalman's well-known *uniform complete observability*. Under this assumption, [30], [31] show asymptotic stability "in the large" of the widely used discrete Kalman filter, in the stochastic and deterministic context respectively. The discrete KKL design proposed in this paper thus constitutes an alternative to the discrete Kalman filter for linear systems. Despite a larger dimension, its advantage mainly lies in its explicit quadratic ISS strict Lyapunov function, which facilitates the robustness analysis, unlike in [30], [31] where the Lyapunov function is not strict and decreases over a certain finite number of steps. More importantly, the KKL design extends to nonlinear systems and guarantees (semi-)global asymptotic stability. On the contrary, the extended Kalman filter/observer for nonlinear systems typically provides only *local* convergence, assuming the uniform complete observability condition holds on the linearization of the dynamics *along the estimate* [32]–[34]. Unfortunately, this kind of assumption typically introduces a loop in the analysis, since the estimation error must remain small to guarantee observability along the estimate, which is in turn needed to keep the error small. This loop is broken in [32] but the analysis remains inherently local. Note also that those papers do not mention any explicit stability guarantees.

Other local designs have been proposed for general discrete systems as in [35] or based on local linearization techniques [36], [37]. In terms of global designs, some LMI-based approaches have been developed for discrete normal forms with Lipschitz nonlinearities as in [38]. But to the best of our knowledge, there do not exist systematic global observer designs for general discrete systems. To further highlight our contribution, we are not aware of any other discrete observer design that can be both arbitrarily fast and robust at the same time. The KKL design we propose in this paper does not assume any particular form for the system dynamics and provides a systematic arbitrarily fast robust observer design under only an appropriate distinguishability condition on the system. Lastly, note that although they both rely on transforming the given dynamics into linear dynamics with output injection, the crucial difference between KKL designs and linearization techniques [36], [37], [39] is that the former does not require a linear output map in the new coordinates (in fact, we do not need to express this output map in the new coordinates), thus leading to much more generic results as the class of systems where the method is applicable is much wider.

This paper is organized as follows. The KKL observer design problem is stated in Section II. Then, sufficient conditions for the existence of a sequence of maps $(T_k)_{k\in\mathbb{N}}$ transforming the dynamics into a filter of the output are presented in Section III. Then, Section IV shows uniform Lipschitz injectivity of $(T_k)_{k\in\mathbb{N}}$ under the uniform Lipschitz backward distinguishability, which allows us to obtain an arbitrarily fast observer in discrete time. Section V shows injectivity of the maps $(T_k)_{k\in\mathbb{N}}$ under a weaker backward distinguishability, but without any convergence guarantee of the estimation error. Lastly, Section VI gives examples including the case of linear time-varying systems and a permanent magnet synchronous motor (PMSM) illustrating the interest of using discrete KKL design for discretized continuous systems.

Notations: Let $\mathbb{R}$ (resp. $\mathbb{N}$) denote the set of real numbers (resp. natural numbers, i.e., $\{0, 1, 2, \ldots\}$). $\mathbb{R}_{\geq 0} = [0, +\infty)$ while $\mathbb{R}_{>0} = (0, +\infty)$ and $\mathbb{N}_{>0} = \mathbb{N} \setminus \{0\}$. $\mathbb{R}^{m\times n}$ (resp. $\mathbb{C}^{m\times n}$) is the set of real (resp. complex) $(m \times n)$-dimensional matrices. For a set E, let $\mathrm{cl}(E)$ be its closure and $E + \sigma$ be the set of points that lie within the distance $\sigma \in \mathbb{R}_{>0}$ from a point in E. Let $\Re(z)$ and $\Im(z)$ denote the real and imaginary parts of the complex variable z. Given a vector norm denoted $|\cdot|$, we denote $\|\cdot\|$ as the induced matrix norm. Let $\mathrm{eig}(W)$ be the set of eigenvalues of the matrix W. For a sequence $(x_k)_{k\in\mathbb{N}}$ of vectors in $\mathbb{R}^m$ indexed by the discrete time $k \in \mathbb{N}$, x_k is the vector at time k, while $x_{i,k}$ denotes its i^{th} component at time k. A function $\rho : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is class-$\mathcal{K}$ if ρ is continuous, $\rho(0) = 0$, and ρ is strictly increasing. A function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is class-$\mathcal{KL}$ if for all $r \in \mathbb{R}_{\geq 0}, \beta(\cdot, r)$ is class-$\mathcal{K}$ and for all $s \in \mathbb{R}_{\geq 0}$, $\beta(s, \cdot)$ is decreasing and $\lim_{r\to+\infty} \beta(s, r) = 0$. For two functions f and g, $f \circ g$ is their composition, namely for all x in the domain of g, $g(x)$ is in the domain of f and $(f \circ g)(x) = f(g(x))$. The left inverse f^* of the map f on the set $\mathcal{X}$ is one such that $f^*(f(x)) = x$ for all $x \in \mathcal{X}$. $A \otimes B$ is the Kronecker product of matrices A and B. Last, for $x \in \mathbb{R}^m$, $B_r(x)$ denotes the open ball of radius $r > 0$ centered at x.

II. Problem Statement

Consider the nonlinear time-varying discrete system

$$x_{k+1} = f_k(x_k), \quad y_k = h_k(x_k), \tag{1}$$

where $f_k : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ and $h_k : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ are the dynamics and output maps, $x_k \in \mathbb{R}^{n_x}$ is the state, and $y_k \in \mathbb{R}^{n_y}$ is the output at discrete time k.

Remark 1: Any system of the form

$$x_{k+1} = f_k(x_k, u_k), \quad y_k = h_k(x_k, u_k), \tag{2}$$

where the input $u_k \in \mathbb{R}^{n_u}$ is a known trajectory of time, can be put into form (1) with the maps $(f_k, h_k)_{k\in\mathbb{N}}$ depending on a particular sequence of inputs $(u_k)_{k\in\mathbb{N}}$. The results of this paper thus depend on this sequence of inputs, but some can be made *uniform* with respect to a family of $(u_k)_{k\in\mathbb{N}}$, if the corresponding assumptions also hold uniformly in the inputs.

Assumption 1: The solutions of (1) of interest are initialized in a set $\mathcal{X}_0$ and remain in a compact set $\mathcal{X} \supseteq \mathcal{X}_0$ in positive time.[1]

The KKL observer design consists in seeking a sequence of nonlinear maps $(T_k)_{k\in\mathbb{N}}$, with $T_k : \mathbb{R}^{n_x} \to \mathbb{R}^{n_z}$, transforming the dynamics (1) into an LTI discrete filter of the output, i.e., such that $z_k = T_k(x_k)$ verifies

$$z_{k+1} = Az_k + By_k, \tag{3}$$

where $A \in \mathbb{R}^{n_z\times n_z}$ is Schur and $B \in \mathbb{R}^{n_z\times n_y}$ such that (A, B) is controllable. In other words, we look for $(T_k)_{k\in\mathbb{N}}$ satisfying for all $k \in \mathbb{N}$,

$$T_{k+1}(x_{k+1}) = AT_k(x_k) + Bh_k(x_k) \tag{4}$$

along solutions to (1) remaining in $\mathcal{X}$. A sufficient condition for that is to have for all $k \in \mathbb{N}$,

$$(T_{k+1} \circ f_k)(x) = AT_k(x) + Bh_k(x), \quad \forall x \in \mathcal{X} : f_k(x) \in \mathcal{X}. \tag{5}$$

The observer in the z-coordinates is then made of a simple filter of the output

$$\hat{z}_{k+1} = A\hat{z}_k + By_k, \tag{6}$$

since the estimation error then verifies $(z_{k+1}-\hat{z}_{k+1}) = A(z_k-\hat{z}_k)$, which is exponentially stable. The following Theorem 1 then shows that if the sequence $(T_k)_{k\in\mathbb{N}}$ to (5) is *uniformly injective* after a certain time (as in (8) below), it admits a sequence of left inverses $(T^*_k)_{k\in\mathbb{N}}$, with $T^*_k : \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$, such that the observer

$$\hat{z}_{k+1} = A\hat{z}_k + By_k, \quad \hat{x}_k = T^*_k(\hat{z}_k), \tag{7}$$

initialized as $\hat{z}_0 \in T_0(\mathcal{X})$, provides an asymptotic estimate $\hat{x}_k \in \mathbb{R}^{n_x}$ of x_k and the estimation error in the x-coordinates is asymptotically stable (as in (9) below). The goal of this paper is then to provide sufficient conditions to guarantee the existence of such a sequence of maps $(T_k)_{k\in\mathbb{N}}$.

Theorem 1: Assume there exists $(T_k)_{k\in\mathbb{N}}$ satisfying (5) with T_0 continuous on $\mathcal{X}$ and $(T_k)_{k\in\mathbb{N}}$ is uniformly injective after a time, i.e., there exist a *concave* class-$\mathcal{K}$ function ρ and $k^\star \in \mathbb{N}$ such that for all $k \geq k^\star$ and for all $(x_a, x_b) \in \mathcal{X}\times\mathcal{X}$,

$$|x_a - x_b| \leq \rho(|T_k(x_a) - T_k(x_b)|). \tag{8}$$

Then, there exists $(T^*_k)_{k\in\mathbb{N}}$ and a class-$\mathcal{KL}$ function β such that for any solution $k \mapsto x_k$ of (1) with $x_0 \in \mathcal{X}_0$ and any solution $k \mapsto \hat{z}_k$ of (7) with $\hat{z}_0 \in T_0(\mathcal{X})$ and input $y_k = h_k(x_k)$, we have

$$|x_k - \hat{x}_k| \leq \beta(|x_0 - \hat{x}_0|, k). \tag{9}$$

Remark 2: In this paper, the concavity assumption of ρ is not restrictive because we will achieve, in Theorem 3, uniform Lipschitz injectivity of $(T_k)_{k\in\mathbb{N}}$ characterized by a linear ρ. In general, this assumption can also be dropped if there exists a compact set $\mathcal{Z} \subset \mathbb{R}^{n_z}$ such that for all $k \geq k^\star$, $T_k(\mathcal{X}) \subseteq \mathcal{Z}$.

Proof: From the uniform injectivity of $(T_k)_{k\in\mathbb{N}}$ in (8), there exists a sequence of left inverse maps $(T^{-1}_k)_{k\in\mathbb{N}} : T_k(\mathcal{X}) \to \mathbb{R}^{n_x}$ such that for all $k \geq k^\star$,

- For all $x \in \mathcal{X}$, $T^{-1}_k(T_k(x)) = x$;
- For all $(z_a, z_b) \in T_k(\mathcal{X}) \times T_k(\mathcal{X})$, $|T^{-1}_k(z_a) - T^{-1}_k(z_b)| \leq \rho(|z_a - z_b|)$.

Applying [40] component-wise, $(T^{-1}_k)_{k\in\mathbb{N}}$ can be extended into a sequence of left inverse maps $(T^*_k)_{k\in\mathbb{N}} : \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ such that there exists $c_1 \in \mathbb{R}_{>0}$ such that for all $k \geq k^\star$,

- For all $x \in \mathcal{X}$, $T^*_k(T_k(x)) = x$;
- For all $(z_a, z_b) \in \mathbb{R}^{n_z}\times\mathbb{R}^{n_z}$, $|T^*_k(z_a)-T^*_k(z_b)| \leq c_1\rho(|z_a-z_b|)$.

It follows that for all $k \geq k^\star$,

$$\begin{aligned}|T^*_k(T_k(x_k)) - T^*_k(\hat{z}_k)| &= |x_k - \hat{x}_k| \\ &\leq c_1\rho(|T_k(x_k) - \hat{z}_k|) \\ &\leq c_1\rho(c_2c_3^k|T_0(x_0) - \hat{z}_0|),\end{aligned}$$

for some $c_2 \in \mathbb{R}_{>0}$ and $c_3 \in (0, 1)$ thanks to the exponential stability in the z-coordinates given by $(z_{k+1} - \hat{z}_{k+1}) = A(z_k - \hat{z}_k)$. Pick $\hat{x}_0 \in \mathcal{X}$ such that $\hat{z}_0 = T_0(\hat{x}_0)$. Because T_0 is continuous on the compact set $\mathcal{X}$, it is also uniformly continuous on $\mathcal{X}$, meaning that there exists a class-$\mathcal{K}$ function ρ_0 such that for any $x_0 \in \mathcal{X}_0$ and $\hat{x}_0 \in \mathcal{X}$, $|T_0(x_0) - \hat{z}_0| = |T_0(x_0) - T_0(\hat{x}_0)| \leq \rho_0(|x_0 - \hat{x}_0|)$. Finally, we get

$$|x_k - \hat{x}_k| \leq c_1\rho(c_2c_3^k\rho_0(|x_0 - \hat{x}_0|)),$$

which is a class-$\mathcal{KL}$ function in $|x_0 - \hat{x}_0|$ and k. ■

The *uniform* injectivity of $(T_k)_{k\in\mathbb{N}}$ as in (8) is thus sufficient to guarantee asymptotic stability of the estimation error. The following academic example shows that it is not necessary, but the injectivity of each map T_k alone, without uniformity in k, can sometimes be insufficient to ensure convergence.

Example 1: Consider the first-order time-varying system

$$x_{k+1} = x_k, \quad y_k = h_kx_k, \tag{10}$$

where $h_k \in \mathbb{R}_{\geq 0}$. We see that the output enables us to reconstruct the constant state x_k as soon as $h_k \neq 0$ for some k. Let us try to build a KKL observer. Thanks to the dynamics being linear, we look for a transformation of the form $T_k(x) = m_kx$, where $(m_k)_{k\in\mathbb{N}}$ is a sequence of scalars

[1] This is much milder than requiring that $\mathcal{X}$ is forward invariant, which means that all trajectories initialized in $\mathcal{X}$, including the ones we are not interested in, remain in $\mathcal{X}$.

to be found so that (5) holds. Picking $\lambda \in (0,1)$, this is achieved if for all $k \in \mathbb{N}$,

$$m_{k+1} = \lambda m_k + h_k,$$

of which the solution is

$$m_k = \lambda^k m_0 + \sum_{j=0}^{k-1} \lambda^{k-j-1} h_j$$

for some initial m_0. As long as $m_0 > 0$, the m_k are always positive for $k > 0$ so that each map T_k is injective. However, if h_k vanishes asymptotically, m_k decays to zero as k increases, and the sequence $(T_k)_{k\in\mathbb{N}}$ is not uniformly injective. We get

$$\begin{aligned} |x_k - \hat{x}_k| &= \frac{1}{m_k}|z_k - \hat{z}_k| = \frac{\lambda^k}{m_k}|z_0 - \hat{z}_0| \\ &= \frac{\lambda^k}{\lambda^k m_0 + \sum_{j=0}^{k-1} \lambda^{k-j-1} h_j}|h_0 x_0 - h_0 \hat{x}_0| \\ &= \frac{h_0}{m_0 + \sum_{j=0}^{k-1} \frac{h_j}{\lambda^{j+1}}}|x_0 - \hat{x}_0|. \end{aligned}$$

Consider the first case where for some $k^\star \in \mathbb{N}_{>0}$,

$$h_k = \begin{cases} 1 & \text{if } k \leq k^\star \\ 0 & \text{if } k > k^\star, \end{cases} \tag{11}$$

then, $|x_k - \hat{x}_k|$ does not converge to zero. The reason is that even though each map T_k is injective at each k, $(T_k)_{k\in\mathbb{N}}$ becomes less and less injective over time. Consider another case where $h_k = h_0 \epsilon^k$ for some constants $h_0 \in \mathbb{R}_{>0}$ and $\epsilon \in (0,1)$, so the system is instantaneously observable at each k, but "less and less" over time. We have

$$\begin{aligned} |x_k - \hat{x}_k| &= \frac{h_0}{m_0 + \frac{h_0}{\lambda}\sum_{j=0}^{k-1}\left(\frac{\epsilon}{\lambda}\right)^j}|x_0 - \hat{x}_0| \\ &= \frac{h_0}{m_0 + \frac{h_0}{\epsilon-\lambda}\left(\left(\frac{\epsilon}{\lambda}\right)^k - 1\right)}|x_0 - \hat{x}_0| \end{aligned}$$

so that if we choose $\lambda < \epsilon$, the error converges to zero asymptotically. Furthermore, if we initialize $(m_k)_{k\in\mathbb{N}}$ as $m_0 = \frac{h_0}{\epsilon-\lambda} > 0$ (note that $(h_k)_{k\in\mathbb{N}}$ is known), we even get exponential stability of the error as

$$|x_k - \hat{x}_k| = (\epsilon - \lambda)\left(\frac{\lambda}{\epsilon}\right)^k |x_0 - \hat{x}_0|.$$

This estimation can also be made arbitrarily fast by keeping pushing λ smaller. Therefore, uniform injectivity is a sufficient condition according to Theorem 1, but it is not necessary. Convergence, stability, as well as other properties, could still happen without uniformity in k, but it is not guaranteed.

In this work, we provide sufficient conditions to guarantee:

- Existence of $(T_k)_{k\in\mathbb{N}}$ satisfying (5) in Section III;
- Uniform Lipschitz injectivity of $(T_k)_{k\in\mathbb{N}}$ after a certain time in Section IV;
- Injectivity of each T_k after a certain time in Section V.

Actually, in Section IV, we achieve a stronger asymptotic property than (9): we show *exponential* stability of the estimation error in the x-coordinates, namely, there exist $c_1 \in \mathbb{R}_{>0}$, $c_2 \in (0,1)$, and $k^\star \in \mathbb{N}$ such that for all $k \geq k^\star$,

$$|x_k - \hat{x}_k| \leq c_1 c_2^k |x_0 - \hat{x}_0|. \tag{12}$$

Such a property is achieved by strengthening the uniform injectivity of $(T_k)_{k\in\mathbb{N}}$ in (8) into uniform *Lipschitz* injectivity and the continuity of T_0 into *Lipschitz* continuity (with ρ and ρ_0 linear). This stronger result enables us to obtain a discrete observer with arbitrarily fast robust convergence as soon as allowed by the distinguishability property. More precisely, for any desired convergence rate $c_2^\star \in (0,1)$, there exists a choice of (A,B) such that (12) is satisfied with $c_2 \leq c_2^\star$. Note that decreasing c_2 typically leads to an increase in c_1, namely, we get a discrete-time equivalence of the *peaking* phenomenon typically encountered in the high-gain observers in continuous time [20]. Also, such a design allows for robustness against disturbances/uncertainties and filtering of measurement noise.

III. Existence of $(T_k)_{k\in\mathbb{N}}$

This part studies the sufficient conditions for the existence of $(T_k)_{k\in\mathbb{N}}$ satisfying (5). It is established under the following assumption.

Assumption 2: For all $k \in \mathbb{N}$, f_k is invertible and its inverse function f_k^{-1} is defined on $\mathbb{R}^{n_x}$.

Remark 3: While invertibility is for now required globally, since the solutions of interest are known to remain in $\mathcal{X}$, it may be possible to modify the maps $(f_k)_{k\in\mathbb{N}}$ (and so $(f_k^{-1})_{k\in\mathbb{N}}$) outside of the set $\mathcal{X}$, while still keeping the observability property mentioned below (see Section IV-D).

Such an assumption is common in discrete observers, such as [5], [27], [37] or in the Kalman literature [30]–[32], [41] and concerns a wide class of systems. For instance, discrete dynamics that are discretizations of continuous dynamics take the form $x_{k+1} = x_k + \Delta t_k \Phi(x_k, t_k)$, which is close to identity for sufficiently small sampling times Δt_k, and therefore invertible. The physical meaning of this assumption is that a given current state has only one possible past. Such invertibility of the dynamics allows us to go back and forth in discrete time and access states at different times, according to

$$\begin{aligned} x_{k+n} &= (f_{k+n-1} \circ f_{k+n-2} \circ \ldots \circ f_k)(x_k), \\ x_{k-n} &= (f_{k-n}^{-1} \circ f_{k-(n-1)}^{-1} \circ \ldots \circ f_{k-1}^{-1})(x_k), \end{aligned}$$

for $k, n \in \mathbb{N}$. Under this invertibility assumption, Theorem 2 gives existence results for the function sequence $(T_k)_{k\in\mathbb{N}}$.

Theorem 2: Under Assumption 2, given any $T_0 : \mathbb{R}^{n_x} \to \mathbb{R}^{n_z}$, the sequence $(T_k)_{k\in\mathbb{N}}$ such that each $T_k : \mathbb{R}^{n_x} \to \mathbb{R}^{n_z}$ is given by

$$\begin{aligned} T_k(x) &= A^k(T_0 \circ f_0^{-1} \circ f_1^{-1} \circ \ldots \circ f_{k-1}^{-1})(x) \\ &\quad + \sum_{j=0}^{k-1} A^{k-j-1} B(h_j \circ f_j^{-1} \circ f_{j+1}^{-1} \circ \ldots \circ f_{k-1}^{-1})(x) \end{aligned} \tag{13}$$

verifies (5). Furthermore, for any other solution $(T'_k)_{k\in\mathbb{N}}$ to (5), for all $x \in \mathcal{X}$ such that $f_{k-1}^{-1}(x) \in \mathcal{X}$ and for all $k \in \mathbb{N}_{>0}$, we have $T'_k(x) = T_k(x)$ with T_k defined in (13) for $T_0 := T'_0$.

Proof: To start, notice that under Assumption 2, (5) is verified if and only if

$$T_k(x) = A(T_{k-1} \circ f_{k-1}^{-1})(x) + B(h_{k-1} \circ f_{k-1}^{-1})(x), \tag{14}$$

for all $x \in \mathcal{X}$ such that $f_{k-1}^{-1}(x) \in \mathcal{X}$ and for all $k \in \mathbb{N}_{>0}$. We next show by induction that (14) is equivalent to (13) for all $k \in \mathbb{N}_{>0}$ and for all $x \in \mathcal{X}$ such that $f_{k-1}^{-1}(x) \in \mathcal{X}$. This is trivial for $k = 1$. Then, assuming (14) is equivalent to (13) for a given k and for all $x \in \mathcal{X}$ such that $f_{k-1}^{-1}(x) \in \mathcal{X}$, we have

$$
\begin{aligned}
&A(T_k \circ f_k^{-1})(x) + B(h_k \circ f_k^{-1})(x) = \\
&A\bigg(A^k(T_0 \circ f_0^{-1} \circ f_1^{-1} \circ \ldots \circ f_{k-1}^{-1})(f_k^{-1}(x)) \\
&+ \sum_{j=0}^{k-1} A^{k-j-1} B(h_j \circ f_j^{-1} \circ f_{j+1}^{-1} \circ \ldots \circ f_{k-1}^{-1})(f_k^{-1}(x))\bigg) \\
&+ B(h_k \circ f_k^{-1})(x) = \\
&AA^k(T_0 \circ f_0^{-1} \circ f_1^{-1} \circ \ldots \circ f_{k-1}^{-1} \circ f_k^{-1})(x) \\
&+ \bigg(A \sum_{j=0}^{k-1} A^{k-j-1} B(h_j \circ f_j^{-1} \circ f_{j+1}^{-1} \circ \ldots \circ f_{k-1}^{-1} \circ f_k^{-1})(x) \\
&+ B(h_k \circ f_k^{-1})(x)\bigg) = \\
&A^{k+1}(T_0 \circ f_0^{-1} \circ f_1^{-1} \circ \ldots \circ f_{k-1}^{-1} \circ f_k^{-1})(x) \\
&+ \sum_{j=0}^{k+1-1} A^{k+1-j-1} B(h_j \circ f_j^{-1} \circ f_{j+1}^{-1} \circ \ldots \circ f_{k-1}^{-1} \circ f_k^{-1})(x).
\end{aligned}
$$

Therefore, by mathematical induction, both expressions are equivalent for all $x \in \mathcal{X}$ such that $f_{k-1}^{-1}(x) \in \mathcal{X}$ and for all $k \in \mathbb{N}_{>0}$. Finally, on $\mathbb{R}^{n_x}$, (13) satisfies (5) analytically. ■

Example 2: Consider the class of (1) with linear dynamics and polynomial output

$$x_{k+1} = F_k x_k, \quad y_k = H_k P_d(x_k), \tag{15}$$

where $(F_k)_{k\in\mathbb{N}} \in \mathbb{R}^{n_x \times n_x}$ and $(H_k)_{k\in\mathbb{N}} \in \mathbb{R}^{n_y \times n_d}$ are sequences of matrices and $P_d : \mathbb{R}^{n_x} \to \mathbb{R}^{n_d}$ is a vector of n_d monomials with degrees less than or equal to d. We then look for $(T_k)_{k\in\mathbb{N}}$ of the form

$$T_k(x) = M_k P_d(x).$$

Since $P_d(F_k(x))$ contains polynomials of x of order less than or equal to d, there exists $(D_k)_{k\in\mathbb{N}} \in \mathbb{R}^{n_d \times n_d}$ such that

$$P_d(F_k x) = D_k P_d(x).$$

Therefore, we have $T_{k+1}(x_{k+1}) = M_{k+1} P_d(x_{k+1}) = M_{k+1} P_d(F_k x_k) = M_{k+1} D_k P_d(x_k)$ and (5) holds if

$$M_{k+1} D_k = A M_k + B H_k. \tag{16}$$

If $(D_k)_{k\in\mathbb{N}}$ is invertible for all $k \in \mathbb{N}$, it can be proven by mathematical induction that (16) admits the unique solution

$$M_k = A^k M_0 \prod_{j=0}^{k-1} D_j^{-1} + \sum_{j=0}^{k-1} A^{k-j-1} B H_j \prod_{q=j}^{k-1} D_q^{-1},$$

for all $k \in \mathbb{N}_{>0}$, initialized as M_0. So $(T_k)_{k\in\mathbb{N}}$ is of the form

$$T_k(x) = \left(A^k M_0 \prod_{j=0}^{k-1} D_j^{-1} + \sum_{j=0}^{k-1} A^{k-j-1} B H_j \prod_{q=j}^{k-1} D_q^{-1}\right) P_d(x). \tag{17}$$

The particular case where the system is fully linear, namely with $P_d(\cdot)$ identity, is detailed below in Section VI-A.

However, $(T_k)_{k\in\mathbb{N}}$, even if it exists, may not be injective and may thus be unusable for observer design. Sufficient conditions guaranteeing injectivity are analyzed next.

IV. An Arbitrarily Fast Robust Discrete Observer from Uniform Lipschitz Backward Distinguishability

This part shows that the uniform Lipschitz injectivity of $(T_k)_{k\in\mathbb{N}}$ is obtained after a certain time under *uniform Lipschitz backward distinguishability* if the target dynamics are pushed sufficiently fast. This leads to an arbitrarily fast robust discrete observer as soon as allowed by distinguishability.

A. Uniform Lipschitz Injectivity of $(\mathsf{T}_k)_{k\in\mathbb{N}}$ from Uniform Lipschitz Backward Distinguishability

In this part, A is chosen of the form $\gamma\tilde{A}$ with $\tilde{A}$ Schur, and $\gamma \in (0,1]$ sufficiently small to ensure *uniformly Lipschitz injectivity* of $(T_k)_{k\in\mathbb{N}}$ after a certain time. This is done under the following distinguishability condition.

Definition 1: The system (1) is uniformly Lipschitz backward distinguishable on a set $\mathcal{X}$ if for each output y_i, $i \in \{1, 2, \ldots, n_y\}$, there exists $m_i \in \mathbb{N}_{>0}$ such that for all $k \geq \overline{m} := \max_i m_i$, the sequence of backward distinguishability maps $(\mathcal{O}_k^{bw})_{k\in\mathbb{N}}$ defined as

$$\mathcal{O}_k^{bw}(x) = (\mathcal{O}_{1,k}^{bw}(x), \mathcal{O}_{2,k}^{bw}(x), \ldots, \mathcal{O}_{n_y,k}^{bw}(x)),$$

where $\mathcal{O}_{i,k}^{bw}(x) \in \mathbb{R}^{m_i}$ is defined as

$$
\mathcal{O}_{i,k}^{bw}(x) = \begin{pmatrix}
(h_{i,k-1} \circ f_{k-1}^{-1})(x) \\
(h_{i,k-2} \circ f_{k-2}^{-1} \circ f_{k-1}^{-1})(x) \\
\ldots \\
(h_{i,k-(m_i-1)} \circ f_{k-(m_i-1)}^{-1} \circ \ldots \circ f_{k-1}^{-1})(x) \\
(h_{i,k-m_i} \circ f_{k-m_i}^{-1} \circ f_{k-(m_i-1)}^{-1} \circ \ldots \circ f_{k-1}^{-1})(x)
\end{pmatrix},
$$

is uniformly Lipschitz injective on $\mathcal{X}$, i.e., there exists $c_o \in \mathbb{R}_{>0}$ such that for all $k \geq \overline{m}$ and for all $(x_a, x_b) \in \mathcal{X} \times \mathcal{X}$,

$$|\mathcal{O}_k^{bw}(x_a) - \mathcal{O}_k^{bw}(x_b)| \geq c_o |x_a - x_b|.$$

Intuitively, the concatenation of a sufficient number m_i of the past outputs determines uniquely and uniformly the current state (and equivalently the trajectory as well). Equivalent kinds of uniform observability are assumed in [42, Theorem 4.1] and [4, Theorem 2] for autonomous and time-varying continuous-time systems respectively, leading to similar results with arbitrarily fast convergence of the estimation error.

Remark 4: While the condition in Definition 1 is what is required and assumed later for the proof, in practice it is not always easy to obtain the closed forms of the inverse maps of f_k in $(\mathcal{O}_k^{bw})_{k\in\mathbb{N}}$. Actually, this condition is satisfied with $m_i = m$ for all $i \in \{1, 2, \ldots, n_y\}$ if both of the following conditions are satisfied.

- There exists $m \in \mathbb{N}_{>0}$ such that there exists $c_{o'} \in \mathbb{R}_{>0}$ such that for all $k \in \mathbb{N}$ and for all $(x_a, x_b) \in \mathcal{X} \times \mathcal{X}$,

$$|\mathcal{O}_k^{fw}(x_a) - \mathcal{O}_k^{fw}(x_b)| \geq c_{o'} |x_a - x_b|,$$

where the sequence of forward distinguishability functions $(\mathcal{O}_k^{fw})_{k\in\mathbb{N}}$ is defined as

$$\mathcal{O}_k^{fw}(x) = (\mathcal{O}_{1,k}^{fw}(x), \mathcal{O}_{2,k}^{fw}(x), \dots, \mathcal{O}_{n_y,k}^{fw}(x)),$$

where $\mathcal{O}_{i,k}^{fw}(x) \in \mathbb{R}^m$ is defined as

$$\mathcal{O}_{i,k}^{fw}(x) = \begin{pmatrix} h_{i,k}(x) \\ (h_{i,k+1}\circ f_k)(x) \\ \dots \\ (h_{i,k+(m-2)}\circ f_{k+(m-3)}\circ\ldots\circ f_k)(x) \\ (h_{i,k+(m-1)}\circ f_{k+(m-2)}\circ f_{k+(m-3)}\circ\ldots\circ f_k)(x) \end{pmatrix};$$

- The sequence of inverses $(f_k^{-1})_{k\in\mathbb{N}}$ is uniformly Lipschitz injective, i.e., there exists $\underline{c}_f \in \mathbb{R}_{>0}$ such that for all $k \in \mathbb{N}$ and for all $(x_a, x_b) \in \mathbb{R}^{n_x}\times\mathbb{R}^{n_x}$,

$$|f_k^{-1}(x_a) - f_k^{-1}(x_b)| \geq \underline{c}_f|x_a - x_b|.$$

Indeed, from the two conditions above, we have for all $k \in \mathbb{N}$ and for all $(x_a, x_b) \in \mathcal{X}\times\mathcal{X}$,

$$\begin{aligned}
&|\mathcal{O}_k^{bw}(x_a) - \mathcal{O}_k^{bw}(x_b)| \\
&= |\mathcal{O}_{k-m}^{fw}((f_{k-m}^{-1}\circ f_{k-(m-1)}^{-1}\circ\ldots\circ f_{k-1}^{-1})(x_a)) \\
&\quad - \mathcal{O}_{k-m}^{fw}((f_{k-m}^{-1}\circ f_{k-(m-1)}^{-1}\circ\ldots\circ f_{k-1}^{-1})(x_b))| \\
&\geq c_{o'}|(f_{k-m}^{-1}\circ f_{k-(m-1)}^{-1}\circ\ldots\circ f_{k-1}^{-1})(x_a) \\
&\quad - (f_{k-m}^{-1}\circ f_{k-(m-1)}^{-1}\circ\ldots\circ f_{k-1}^{-1})(x_b)| \\
&\geq c_{o'}\underline{c}_f^m|x_a - x_b| := c_o|x_a - x_b|,
\end{aligned}$$

by letting $c_o = c_{o'}\underline{c}_f^m$. Checking uniform Lipschitz backward distinguishability using $(\mathcal{O}_k^{fw})_{k\in\mathbb{N}}$ is much more convenient than $(\mathcal{O}_k^{bw})_{k\in\mathbb{N}}$ since the forward maps $(f_k)_{k\in\mathbb{N}}$ are available.

For our uniform Lipschitz injectivity result, we make the following assumptions.

Assumption 3: We assume that:

(A3.1) The sequences $(f_k^{-1})_{k\in\mathbb{N}}$ and $(h_k)_{k\in\mathbb{N}}$ are uniformly Lipschitz, i.e., there exist positive scalars c_f and c_h such that for all $k \in \mathbb{N}$ and for all $(x_a, x_b) \in \mathbb{R}^{n_x}\times\mathbb{R}^{n_x}$,

$$|f_k^{-1}(x_a) - f_k^{-1}(x_b)| \leq c_f|x_a - x_b|,$$
$$|h_k(x_a) - h_k(x_b)| \leq c_h|x_a - x_b|;$$

(A3.2) The system (1) is uniformly Lipschitz backward distinguishable on $\mathcal{X}$ for some $m_i \in \mathbb{N}_{>0}$, $i \in \{1, 2, \dots, n_y\}$.

Remark 5: Assumption (A3.1) requires global uniform Lipschitzness of the sequences $(f_k^{-1})_{k\in\mathbb{N}}$ and $(h_k)_{k\in\mathbb{N}}$. Its relaxation into uniform Lipschitzness over a compact set is analyzed in Section IV-D. Note that for a linear time-varying system, Assumption (A3.1) is reduced to uniform boundedness of the dynamics and output matrices (see Section VI-A).

The following theorem then shows uniform Lipschitz injectivity of $(T_k)_{k\in\mathbb{N}}$ after a certain time.

Theorem 3: Suppose Assumptions 1, 2, and 3 hold. Define $n_z = \sum_{i=1}^{n_y} m_i$. Consider a globally Lipschitz[2] map $T_0 : \mathbb{R}^{n_x} \to \mathbb{R}^{n_z}$, and for each $i \in \{1, 2, \dots, n_y\}$, a controllable pair $(\tilde{A}_i, \tilde{B}_i) \in \mathbb{R}^{m_i\times m_i}\times\mathbb{R}^{m_i}$ with $\tilde{A}_i$ Schur. Then, there exists $\gamma^\star \in \mathbb{R}_{>0}$ such that for any $0 < \gamma < \gamma^\star$, there exists $k^\star \in \mathbb{N}$ such that the sequence $(T_k)_{k\in\mathbb{N}}$ defined in (13) with

$$A = \gamma\,\mathrm{diag}(\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_{n_y}) \in \mathbb{R}^{n_z\times n_z}, \tag{18a}$$
$$B = \mathrm{diag}(\tilde{B}_1, \tilde{B}_2, \dots, \tilde{B}_{n_y}) \in \mathbb{R}^{n_z\times n_y}, \tag{18b}$$

and initialized as T_0, is uniformly Lipschitz injective on $\mathcal{X}$ for all $k \geq k^\star$, where $\gamma^\star$ and $k^\star$ are defined in the proof. More precisely, there exists $c \in \mathbb{R}_{>0}$ (independent of γ) such that for all $k \geq k^\star$ and all $(x_a, x_b) \in \mathcal{X}\times\mathcal{X}$, we have

$$|T_k(x_a) - T_k(x_b)| \geq c\gamma^{\overline{m}-1}|x_a - x_b|, \tag{19}$$

where $\overline{m} := \max_i m_i$.

Proof: First, pick γ small enough for $\gamma\,\mathrm{diag}(\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_{n_y})$ to be Schur, namely $\gamma\max_i\max|\mathrm{eig}(\tilde{A}_i)| < 1$. Consider a solution $(T_k)_{k\in\mathbb{N}}$ of (5) for (A, B) given in (18). Then,

$$T_k(x) = (T_{1,k}(x), T_{2,k}(x), \dots, T_{i,k}(x), \dots, T_{n_y,k}(x))$$

where for each $i \in \{1, 2, \dots, n_y\}$, $(T_{i,k})_{k\in\mathbb{N}}$ is solution to (5) with (A, B) replaced by $(\gamma\tilde{A}_i, \tilde{B}_i)$. Therefore, Theorem 2 applies to each $(T_{i,k})_{k\in\mathbb{N}}$. It follows that for each $i \in \{1, 2, \dots, n_y\}$, for all $k \geq m_i$, and for all $(x_a, x_b) \in \mathcal{X}\times\mathcal{X}$, $T_{i,k}(x_a) - T_{i,k}(x_b)$ can be written as the sum of three parts

$$\begin{aligned}
T_{i,k}(x_a) - T_{i,k}(x_b) &= \big(\mathcal{I}_{i,k}(x_a) - \mathcal{I}_{i,k}(x_b)\big) \\
&\quad + \big(\mathcal{T}_{i,k}(x_a) - \mathcal{T}_{i,k}(x_b)\big) + \big(\mathcal{R}_{i,k}(x_a) - \mathcal{R}_{i,k}(x_b)\big)
\end{aligned}$$

where

$$\begin{aligned}
&\mathcal{I}_{i,k}(x_a) - \mathcal{I}_{i,k}(x_b) \\
&= (\gamma\tilde{A}_i)^k\bigg((T_0\circ f_0^{-1}\circ f_1^{-1}\circ\ldots\circ f_{k-1}^{-1})(x_a) \\
&\quad - (T_0\circ f_0^{-1}\circ f_1^{-1}\circ\ldots\circ f_{k-1}^{-1})(x_b)\bigg),
\end{aligned}$$

$$\begin{aligned}
&\mathcal{R}_{i,k}(x_a) - \mathcal{R}_{i,k}(x_b) \\
&= \sum_{j=0}^{k-m_i-1}(\gamma\tilde{A}_i)^{k-j-1}\tilde{B}_i\bigg((h_{i,j}\circ f_j^{-1}\circ f_{j+1}^{-1}\circ\ldots\circ f_{k-1}^{-1})(x_a) \\
&\quad - (h_{i,j}\circ f_j^{-1}\circ f_{j+1}^{-1}\circ\ldots\circ f_{k-1}^{-1})(x_b)\bigg),
\end{aligned}$$

$$\begin{aligned}
&\mathcal{T}_{i,k}(x_a) - \mathcal{T}_{i,k}(x_b) \\
&= \sum_{j=k-m_i}^{k-1}(\gamma\tilde{A}_i)^{k-j-1}\tilde{B}_i\bigg((h_{i,j}\circ f_j^{-1}\circ f_{j+1}^{-1}\circ\ldots\circ f_{k-1}^{-1})(x_a) \\
&\quad - (h_{i,j}\circ f_j^{-1}\circ f_{j+1}^{-1}\circ\ldots\circ f_{k-1}^{-1})(x_b)\bigg) \\
&= \mathcal{D}_i(\gamma)\mathcal{C}_i(\mathcal{O}_{i,k}^{bw}(x_a) - \mathcal{O}_{i,k}^{bw}(x_b)),
\end{aligned}$$

where $\mathcal{D}_i(\gamma) = \mathrm{diag}(1, \gamma, \gamma^2, \dots, \gamma^{m_i-1})$ and $\mathcal{C}_i = \begin{pmatrix}\tilde{B}_i & \tilde{A}_i\tilde{B}_i & \tilde{A}_i^2\tilde{B}_i & \dots & \tilde{A}_i^{m_i}\tilde{B}_i\end{pmatrix}$ is the controllability matrix of the pair $(\tilde{A}_i, \tilde{B}_i)$. Now, we will establish bounds on each of the three parts. As $(T_k)_{k\in\mathbb{N}}$ is initialized globally Lipschitz, there exists $c_T \in \mathbb{R}_{\geq 0}$ such that for all $(x_a, x_b) \in \mathbb{R}^{n_x}\times\mathbb{R}^{n_x}$, $|T_0(x_a) - T_0(x_b)| \leq c_T|x_a - x_b|$. Exploiting

[2]This is only a constraint on how to initialize $(T_k)_{k\in\mathbb{N}}$, which should not impact estimation since this will be forgotten and at the initial time we do not have uniform Lipschitz backward distinguishability anyway; in fact, it is suggested to choose T_0 identically zero when possible.

Assumption (A3.1), we thus have for all $i \in \{1, 2, \dots, n_y\}$, for all $k \geq m_i$, and for all $(x_a, x_b) \in \mathcal{X} \times \mathcal{X}$,

$$|\mathcal{I}_{i,k}(x_a) - \mathcal{I}_{i,k}(x_b)| \leq c_T(\gamma\|\tilde{A}_i\|c_f)^k|x_a - x_b|.$$

Then, for γ such that $\gamma \max_i \|\tilde{A}_i\|c_f < 1$, exploiting Assumption (A3.1), we have for all $i \in \{1, 2, \dots, n_y\}$, for all $k \geq m_i$, and for all $(x_a, x_b) \in \mathcal{X} \times \mathcal{X}$,

$$\begin{aligned}
&|\mathcal{R}_{i,k}(x_a) - \mathcal{R}_{i,k}(x_b)| \\
&\leq \sum_{j=0}^{k-m_i-1} (\gamma\|\tilde{A}_i\|)^{k-j-1}\|\tilde{B}_i\|c_h c_f^{k-j}|x_a - x_b| \\
&= \|\tilde{B}_i\|c_h c_f \frac{(\gamma\|\tilde{A}_i\|c_f)^{m_i}}{1 - \gamma\|\tilde{A}_i\|c_f}(1 - (\gamma\|\tilde{A}_i\|c_f)^{k-m_i-1})|x_a - x_b| \\
&\leq \|\tilde{B}_i\|c_h c_f \frac{(\gamma\|\tilde{A}_i\|c_f)^{m_i}}{1 - \gamma\|\tilde{A}_i\|c_f}|x_a - x_b|.
\end{aligned}$$

As the pairs $(\tilde{A}_i, \tilde{B}_i) \in \mathbb{R}^{m_i \times m_i} \times \mathbb{R}^{m_i}$ are controllable, there exists $c_c \in \mathbb{R}_{>0}$ such that $\|\mathcal{C}_i\| \geq c_c > 0$ for all $i \in \{1, 2, \dots, n_y\}$. Next, from Assumption 3, we deduce that for all $i \in \{1, 2, \dots, n_y\}$, for all $k \geq m_i$, and for all $(x_a, x_b) \in \mathcal{X} \times \mathcal{X}$,

$$|\mathcal{T}_{i,k}(x_a) - \mathcal{T}_{i,k}(x_b)| \geq \gamma^{m_i-1}c_c|\mathcal{O}^{bw}_{i,k}(x_a) - \mathcal{O}^{bw}_{i,k}(x_b)|.$$

Therefore, for all $i \in \{1, 2, \dots, n_y\}$, for all $k \geq m_i$, and for all $(x_a, x_b) \in \mathcal{X} \times \mathcal{X}$,

$$\begin{aligned}
|T_{i,k}(x_a) - T_{i,k}(x_b)| &\geq |\mathcal{T}_{i,k}(x_a) - \mathcal{T}_{i,k}(x_b)| - |\mathcal{R}_{i,k}(x_a) \\
&\quad - \mathcal{R}_{i,k}(x_b)| - |\mathcal{I}_{i,k}(x_a) - \mathcal{I}_{i,k}(x_b)| \\
&\geq \gamma^{m_i-1}c_c|\mathcal{O}^{bw}_{i,k}(x_a) - \mathcal{O}^{bw}_{i,k}(x_b)| \\
&\quad - \|\tilde{B}_i\|c_h c_f \frac{(\gamma\|\tilde{A}_i\|c_f)^{m_i}}{1 - \gamma\|\tilde{A}_i\|c_f}|x_a - x_b| \\
&\quad - c_T(\gamma\|\tilde{A}_i\|c_f)^k|x_a - x_b| \\
&\geq \gamma^{m_i-1}\Bigg(c_c|\mathcal{O}^{bw}_{i,k}(x_a) - \mathcal{O}^{bw}_{i,k}(x_b)| \\
&\quad - \|\tilde{B}_i\|c_h c_f \frac{\gamma(\|\tilde{A}_i\|c_f)^{m_i}}{1 - \gamma\|\tilde{A}_i\|c_f}|x_a - x_b| \\
&\quad - c_T\gamma^{k-m_i+1}(\|\tilde{A}_i\|c_f)^k|x_a - x_b|\Bigg).
\end{aligned}$$

Now, if we concatenate the outputs, depending on the norm, there exists a constant $c_N \in \mathbb{R}_{>0}$ such that for all $k \geq \overline{m}$ and for all $(x_a, x_b) \in \mathcal{X} \times \mathcal{X}$, we have since $\gamma \in (0, 1]$ and thanks to Assumption (A3.2),

$$\begin{aligned}
&|T_k(x_a) - T_k(x_b)| \geq \\
&c_N\gamma^{\overline{m}-1}\Bigg(c_c c_o - \max_i \|\tilde{B}_i\|c_h c_f \frac{\gamma \max_i((\|\tilde{A}_i\|c_f)^{m_i})}{1 - \gamma \max_i \|\tilde{A}_i\|c_f} \\
&\qquad - c_T\gamma^{k-\overline{m}+1}(\max_i \|\tilde{A}_i\|c_f)^k\Bigg)|x_a - x_b|.
\end{aligned}$$

If we select $\gamma \in (0, 1]$ such that

$$0 < \gamma < \gamma^\star = \min\Bigg\{\frac{1}{\max_i \max|\operatorname{eig}(\tilde{A}_i)|}, \frac{1}{\max_i \|\tilde{A}_i\|c_f}, \frac{c_c c_o}{\max_i \|\tilde{A}_i\|c_f c_c c_o + \max_i \|\tilde{B}_i\|c_h c_f \max_i((\|\tilde{A}_i\|c_f)^{m_i})}\Bigg\},$$

then with γ fixed, for all $k \geq k^\star$ where $k^\star = \overline{m}$ if $c_T = 0$ and

$$k^\star = \max\left\{\overline{m}, \left\lfloor \frac{(\overline{m} - 1)\ln\gamma + \ln\tilde{c} - \ln c_T}{\ln\gamma + \ln(\max_i \|\tilde{A}_i\|c_f)} + 1 \right\rfloor\right\},$$

where $\tilde{c} = c_c c_o - \max_i \|\tilde{B}_i\|c_h c_f \frac{\gamma \max_i((\|\tilde{A}_i\|c_f)^{m_i})}{1-\gamma \max_i \|\tilde{A}_i\|c_f}$ if $c_T > 0$, there exists a constant $c \in \mathbb{R}_{>0}$ (independent of γ) such that for all $(x_a, x_b) \in \mathcal{X} \times \mathcal{X}$, we have (19). ■

Remark 6: This is a high-gain result in discrete time since we have to push the (discrete) dynamics sufficiently fast, namely take γ sufficiently small, to guarantee uniform Lipschitz injectivity of $(T_k)_{k\in\mathbb{N}}$. However, as γ is picked closer to zero, the coefficient $\frac{1}{c\gamma^{\overline{m}-1}}$ quantifying the injectivity of $(T_k)_{k\in\mathbb{N}}$ in (19) increases, making $(T_k)_{k\in\mathbb{N}}$ "less (but still) uniformly Lipschitz injective". We also observe that:

- If c_o is close to zero, i.e., the system (1) is "less uniformly Lipschitz backward distinguishable", the upper bound
$$\frac{c_c c_o}{\max_i \|\tilde{A}_i\|c_f c_c c_o + \max_i \|\tilde{B}_i\|c_h c_f \max_i((\|\tilde{A}_i\|c_f)^{m_i})}$$
on γ is reduced, which means we have to pick γ closer to zero to guarantee uniform Lipschitz injectivity of $(T_k)_{k\in\mathbb{N}}$;
- As γ is picked closer to zero, the quantity
$$\left\lfloor \frac{(\overline{m} - 1)\ln\gamma + \ln\tilde{c} - \ln c_T}{\ln\gamma + \ln(\max_i \|\tilde{A}_i\|c_f)} + 1 \right\rfloor$$
approaches $\overline{m}$, so $k^\star = \overline{m}$, which means $(T_k)_{k\in\mathbb{N}}$ becomes uniformly Lipschitz injective right after we have uniform Lipschitz backward distinguishability, namely in $\overline{m}$ steps. Also, the discontinuity of $k^\star$ in c_T reflexes the time dependence of the injectivity of $(T_k)_{k\in\mathbb{N}}$. Indeed, if $c_T = 0$, then the uniform Lipschitz injectivity of $(T_k)_{k\in\mathbb{N}}$ is achieved as soon as we get uniform Lipschitz backward distinguishability, so it is independent of time. For $c_T > 0$, we will have to wait some time until the terms $(\mathcal{I}_k)_{k\in\mathbb{N}}$ become dominated. Therefore, this injectivity is time-dependent.

Example 3: Consider the system in Example 1. We have

$$\mathcal{O}^{bw}_k(x) = \begin{pmatrix} h_{k-1}x \\ h_{k-2}x \\ \dots \\ h_{k-m}x \end{pmatrix} = \begin{pmatrix} h_{k-1} \\ h_{k-2} \\ \dots \\ h_{k-m} \end{pmatrix} x := \mathcal{H}^{bw}_k x,$$

which is not uniformly Lipschitz injective since the $\|\mathcal{H}^{bw}_k\|$ part in

$$|\mathcal{O}^{bw}_k(x_a) - \mathcal{O}^{bw}_k(x_b)| = \|\mathcal{H}^{bw}_k\||x_a - x_b|$$

cannot be lower bounded by any positive constant uniformly in k for any m. Therefore, this example does not fall into the context of Theorem 3.

B. Arbitrarily Fast Observer Design

According to the proof of Theorem 1, once $(T_k)_{k\in\mathbb{N}}$ has become uniformly Lipschitz injective on $\mathcal{X}$ following Theorem 3, there exists a sequence of left inverse maps $(T^*_k)_{k\in\mathbb{N}} : \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ and $c' \in \mathbb{R}_{>0}$ such that

$$T^*_k(T_k(x)) = x, \quad \forall k \geq k^\star, \quad \forall x \in \mathcal{X}, \tag{20a}$$

$$|T_k^*(z_a) - T_k^*(z_b)| \leq \frac{c'}{c\gamma^{\overline{m}-1}}|z_a - z_b|,$$
$$\forall k \geq k^\star, \quad \forall (z_a, z_b) \in \mathbb{R}^{n_z} \times \mathbb{R}^{n_z}. \quad (20b)$$

Exploiting Lipschitzness, the result of Theorem 1 can thus be strengthened as follows, obtaining exponential asymptotic stability of the estimation error in the x-coordinates, and an arbitrarily fast discrete observer.

Corollary 1: Under the assumptions of Theorem 3, consider A and B of the form (18) with $\gamma < \gamma^\star$, $(T_k)_{k\in\mathbb{N}}$, and $k^\star$ provided by Theorem 3. Then, there exist $(T_k^*)_{k\in\mathbb{N}}$ and $\overline{c} \in \mathbb{R}_{>0}$ such that for any solution $k \mapsto x_k$ of (1) with $x_0 \in \mathcal{X}_0$ and any solution $k \mapsto \hat{z}_k$ of (7) with[3] $\hat{z}_0 \in T_0(\mathcal{X})$ and input $y_k = h_k(x_k)$,

$$|x_k - \hat{x}_k| \leq \frac{\overline{c}(\gamma\|\tilde{A}\|)^k}{\gamma^{\overline{m}-1}}|x_0 - \hat{x}_0|, \quad \forall k \geq k^\star. \quad (21)$$

Corollary 1 shows that the observer (7) can be made arbitrarily fast after $(T_k)_{k\in\mathbb{N}}$ has become uniformly Lipschitz injective, by picking γ closer to zero. Indeed, compared with (12), the error in the x-coordinates is exponentially stable with $c_1 = \frac{\overline{c}}{\gamma^{\overline{m}-1}}$ and $c_2 = \gamma\|\tilde{A}\|$. For any desired convergence rate $c_2^\star \in (0,1)$, by picking $\gamma \leq \min\left\{\frac{c_2^\star}{\|\tilde{A}\|}, \gamma^\star\right\}$, we achieve $c_2 \leq c_2^\star$. Note though that this typically increases c_1, because if $c_2 \leq c_2^\star$ then $c_1 \geq c_1^\star = \frac{\overline{c}\|\tilde{A}\|^{\overline{m}-1}}{(c_2^\star)^{\overline{m}-1}}$. We thus recover here a discrete-time version of the well-known *peaking* behavior in continuous-time high-gain designs [20]. This observer is illustrated in Section VI.

Remark 7: While we assume in Assumption (A3.1) that the maps $(f_k^{-1})_{k\in\mathbb{N}}$ and $(h_k)_{k\in\mathbb{N}}$ are uniformly Lipschitz, namely the Lipschitz constants c_f and c_h are the same for all k, we can instead consider sequences of Lipschitz constants $(c_{f,k})_{k\in\mathbb{N}}$ and $(c_{h,k})_{k\in\mathbb{N}}$ providing that there are positive scalars $\overline{c}_f$ and $\overline{c}_h$ such that for all $k \in \mathbb{N}$, $c_{f,k} \leq \overline{c}_f$ and $c_{h,k} \leq \overline{c}_h$. Assumption (A3.1) then holds with $c_f = \overline{c}_f$ and $c_h = \overline{c}_h$. These upper bounds prevent an asymptotic loss of Lipschitzness (when $(c_{f,k})_{k\in\mathbb{N}}$ and $(c_{h,k})_{k\in\mathbb{N}}$ diverge to infinity). Similarly, in Assumption (A3.2), we can consider a sequence $(c_{o,k})_{k\in\mathbb{N}}$ lower bounded by $\underline{c}_o > 0$ (to prevent an asymptotic loss of observability). Indeed, this allows us to update dynamically $\gamma \in (0,1]$ at each iteration k, as follows

$$0 < \gamma_k < \mu\gamma_k^\star = \mu \min\left\{\frac{1}{\max_i \max|\operatorname{eig}(\tilde{A}_i)|}, \frac{1}{\max_i \|\tilde{A}_i\|c_{f,k}}, \frac{c_c c_{o,k}}{\max_i \|\tilde{A}_i\|c_{f,k}c_c c_{o,k} + \max_i \|\tilde{B}_i\|c_{h,k}c_{f,k}\max_i((\|\tilde{A}_i\|c_{f,k})^{m_i})}\right\},$$

for some constant $\mu \in (0,1)$. The role of μ is to prevent $(\gamma_k)_{k\in\mathbb{N}}$ from converging asymptotically to $(\gamma_k^\star)_{k\in\mathbb{N}}$, which cannot converge to zero thanks to the upper bounds $\overline{c}_f$ and $\overline{c}_h$. Indeed, this could prevent convergence/injectivity. The interest of allowing γ to vary is that, at some time when we have a lot of observability (large $c_{o,k}$) or Lipschitzness (small $c_{f,k}$ or $c_{h,k}$), we can afford to let γ_k increase while still keeping convergence, thus decreasing the peaking (or the noise amplification, see next Section IV-C) caused by a too fast observer (see Section VI-B for illustrations). Finally, we can pick a time-varying target filter in the z-coordinates, provided that the properties are uniform with respect to this variation. For instance, it was observed on a continuous-time motor [43], without any rigorous proof, that performance can be improved if the eigenvalues of the filter are adapted to the motor speed.

Remark 8: If T_0 is taken constant (or even identically zero) meaning that $c_T = 0$, then for any initial condition $x_0 \in \mathcal{X}_0$ of the system and $\hat{z}_0$ of the observer, we have $\hat{z}_0 = T_0(x_0)$. This leads to $\hat{z}_k = T_k(x_k)$ for all $k \in \mathbb{N}$ and so $\hat{x}_k = x_k$ for all $k \geq k^\star$. Therefore, we have finite-time convergence.

C. Robust and Input-to-state Stability of the Error

In this part, we now study the robust stability (in the sense of [44]) and ISS properties [45] of the observer given by Corollary 1. Suppose the system has dynamics (1) with some disturbance/uncertainty v_k and a measurement with noise w_k:

$$x_{k+1} = f_k(x_k) + v_k, \quad y_k = h_k(x_k) + w_k. \quad (22)$$

Then, if the pair $(f_k)_{k\in\mathbb{N}}$, $(h_k)_{k\in\mathbb{N}}$ verifies the conditions of Theorem 3, we know that there exists a sequence of left inverses $(T_k^*)_{k\in\mathbb{N}}$ for $k \geq k^\star$ that verifies (20). However, in practice, following for instance [16], such maps are only approximately known. Theorem 4 then shows the robustness of the estimation error in the x-coordinates with respect to all those uncertainties.

Theorem 4: Under the assumptions of Theorem 3, consider A and B of the form (18) with $\gamma < \gamma^\star$, $(T_k)_{k\in\mathbb{N}}$, and $k^\star$ provided by Theorem 3, and $(T_k^*)_{k\in\mathbb{N}}$ provided by Corollary 1. Consider an approximation $(\tilde{T}_k^*)_{k\in\mathbb{N}}$ of $(T_k^*)_{k\in\mathbb{N}}$ and $\delta \in \mathbb{R}_{>0}$ such that

$$|\tilde{T}_k^*(z) - T_k^*(z)| \leq \delta, \quad \forall z \in \mathbb{R}^{n_z}. \quad (23)$$

Then, there exist positive scalars $\overline{c}$, $\overline{c}_v$, and $\overline{c}_w$ (independent of γ) such that for any solution to the system (22) with $x_0 \in \mathcal{X}_0$ and any solution to

$$\hat{z}_{k+1} = \gamma A\hat{z}_k + By_k, \quad \hat{x}_k = \tilde{T}_k^*(\hat{z}_k), \quad (24)$$

initialized as $\hat{z}_0 \in T_0(\mathcal{X})$, we have for all $k \geq k^\star$,

$$|x_k - \hat{x}_k| \leq \frac{\overline{c}(\gamma\|\tilde{A}\|)^k}{\gamma^{\overline{m}-1}}|x_0 - \hat{x}_0| + \frac{1}{\gamma^{\overline{m}-1}}\sum_{j=0}^{k-1}(\gamma^\star\|\tilde{A}\|)^{k-j-1}(\overline{c}_v v_j + \overline{c}_w w_j) + \delta. \quad (25)$$

Proof: First, we prove that $(T_k)_{k\in\mathbb{N}}$ provided by Theorem 3 is uniformly Lipschitz. Indeed, from Assumption (A3.1), we have for all $k \in \mathbb{N}$ and for all $(x_a, x_b) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$,

$$\begin{aligned}
&|T_k(x_a) - T_k(x_b)| \\
&\leq c_T(\gamma \max_i \|\tilde{A}_i\|c_f)^k|x_a - x_b| \\
&\quad + \sum_{j=0}^{k-1}(\gamma \max_i \|\tilde{A}_i\|)^{k-j-1}\max_i \|\tilde{B}_i\|c_h c_f^{k-j}|x_a - x_b| \\
&\leq c_T|x_a - x_b| \\
&\quad + \max_i \|\tilde{B}_i\|c_h c_f \frac{1 - (\gamma\max_i \|\tilde{A}_i\|c_f)^{k-1}}{1 - \gamma\max_i \|\tilde{A}_i\|c_f}|x_a - x_b| \\
&\leq \left(c_T + \frac{\max_i \|\tilde{B}_i\|c_h c_f}{1 - \gamma\max_i \|\tilde{A}_i\|c_f}\right)|x_a - x_b| \\
&:= c_L|x_a - x_b|.
\end{aligned}$$

[3] It is intuitive to initialize $\hat{z}_0$ in the image of the known set $\mathcal{X}$. If T_0 is globally Lipschitz as in Theorem 3, then $\hat{x}_0$ can be anywhere in $\mathbb{R}^{n_x}$.

We now prove the robust stability and ISS properties. Consider a solution to the system (22) with $x_0 \in \mathcal{X}_0$ and a solution to (24) with $z_0 \in T_0(\mathcal{X})$. Denoting $z_k = T_k(x_k)$, we write the dynamics in the z-coordinates as

$$\begin{aligned} z_{k+1} &= T_{k+1}(f_k(x_k) + v_k) \\ &= T_{k+1}(f_k(x_k)) + T_{k+1}(f_k(x_k) + v_k) - T_{k+1}(f_k(x_k)) \\ &= \gamma\tilde{A}T_k(x_k) + Bh_k(x_k) + T_{k+1}(f_k(x_k) + v_k) \\ &\quad - T_{k+1}(f_k(x_k)) \\ &= \gamma\tilde{A}T_k(x_k) + B(y_k - w_k) + T_{k+1}(f_k(x_k) + v_k) \\ &\quad - T_{k+1}(f_k(x_k)) \\ &= \gamma\tilde{A}z_k + By_k + T_{k+1}(f_k(x_k) + v_k) - T_{k+1}(f_k(x_k)) \\ &\quad - Bw_k. \end{aligned}$$

Because $(T_k)_{k\in\mathbb{N}}$ is uniformly Lipschitz, we have for all $k \in \mathbb{N}$,

$$|T_{k+1}(f_k(x) + v_k) - T_{k+1}(f_k(x))| \leq c_L|v_k|.$$

According to (24), we get for all $k \in \mathbb{N}_{>0}$,

$$\begin{aligned} &z_k - \hat{z}_k = (\gamma\tilde{A})^k(z_0 - \hat{z}_0) \\ &+\sum_{j=0}^{k-1}(\gamma\tilde{A})^{k-j-1}(T_{j+1}(f_j(x_j)+v_j)-T_{j+1}(f_j(x_j))-Bw_j). \end{aligned}$$

Therefore, we have for all $k \geq k^\star$,

$$\begin{aligned} |x_k - \hat{x}_k| &= |T_k^*(z_k) - \tilde{T}_k^*(\hat{z}_k)| \\ &\leq |T_k^*(z_k) - T_k^*(\hat{z}_k)| + \delta \\ &\leq \frac{c'}{c\gamma^{\overline{m}-1}}|z_k - \hat{z}_k| + \delta \\ &\leq \frac{c'(\gamma\|\tilde{A}\|)^k}{c\gamma^{\overline{m}-1}}|z_0 - \hat{z}_0| \\ &\quad + \frac{c'}{c\gamma^{\overline{m}-1}}\sum_{j=0}^{k-1}(\gamma\|\tilde{A}\|)^{k-j-1}|T_{j+1}(f_j(x_j) + v_j) \\ &\quad - T_{j+1}(f_j(x_j)) - Bw_j| + \delta \\ &\leq \frac{c'(\gamma\|\tilde{A}\|)^k}{c\gamma^{\overline{m}-1}}|T_0(x_0) - T_0(\hat{x}_0)| \\ &\quad + \frac{c'}{c\gamma^{\overline{m}-1}}\sum_{j=0}^{k-1}(\gamma^\star\|\tilde{A}\|)^{k-j-1}(c_L v_j + \|B\|w_j) + \delta. \end{aligned}$$

This concludes the proof. ■

Remark 9: Theorem 4 shows that the estimation error in the x-coordinates is robustly stable with respect to the disturbance/uncertainty v_k as well as the noise w_k and it is ISS with respect to the approximation error δ. The former property, defined in [44], is stronger than the ISS one defined in [45].

Note that it is the exponential stability (rather than asymptotic stability) of the estimation error that provides the ISS with respect to disturbances and measurement noise. We also see from (25) that accelerating the convergence by pushing γ closer to zero will worsen the effect of the disturbances and noise, but not that of the approximation of the inverse transformation.

D. Saturating the Inverse Maps to Relax Assumption 3

In Assumption (A3.1), we require that the map sequences $(f_k^{-1})_{k\in\mathbb{N}}$ and $(h_k)_{k\in\mathbb{N}}$ are *globally* uniformly Lipschitz, which is due to the fact that we do not have backward invariance of the sequence on $\mathcal{X}$. Here, we would like to study how to relax that into a *local* requirement on a certain bounded set, without losing Assumption (A3.2).

Let us assume that, given the m_i of Assumption (A3.2), there exists a large enough positive scalar σ_d such that for all $x \in \mathcal{X}$ and for all $k \geq \overline{m} := \max_i m_i$, all the pre-images $f_{k-1}^{-1}(x)$, $(f_{k-2}^{-1} \circ f_{k-1}^{-1})(x)$, up to $(f_{k-\overline{m}}^{-1} \circ f_{k-(\overline{m}-1)}^{-1} \circ \ldots \circ f_{k-1}^{-1})(x)$ are in $\mathcal{X} + \sigma_d$. This means that we can change $(f_k^{-1})_{k\in\mathbb{N}}$ as we want outside of $\mathcal{X} + \sigma_d$ without altering Assumption (A3.2) (and without altering the system dynamics on the set $\mathcal{X}$ where the solutions of interest evolve).

Now, for any $\sigma_c > \sigma_d$, let us consider a saturating function $\chi : \mathbb{R}^{n_x} \to \mathbb{R}$ defined as

$$\chi(x) = \begin{cases} 1 & \text{if } x \in \mathcal{X} + \sigma_d \\ g(x) & \text{if } x \in (\mathcal{X} + \sigma_c) \setminus (\mathcal{X} + \sigma_d) \\ 0 & \text{if } x \notin \mathcal{X} + \sigma_c, \end{cases} \tag{26}$$

where g is any locally Lipschitz function such that χ is locally Lipschitz. We then define $(f_k^\dagger)_{k\in\mathbb{N}} : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ as

$$f_k^\dagger(x) = \chi(x)f_k^{-1}(x) + (1 - \chi(x))x. \tag{27}$$

The set

$$\mathcal{I} = (\mathcal{X} + \sigma_c) \cup \left(\bigcup_{k\in\mathbb{N}} f_k^\dagger(\mathcal{X} + \sigma_c)\right) \subset \mathbb{R}^{n_x}$$

illustrated in Figure 1 is backward invariant with respect to $(f_k^\dagger)_{k\in\mathbb{N}}$. Indeed, pick any $x \in \mathcal{I}$ and any $k \in \mathbb{N}$. Then, either $x \in \mathcal{X} + \sigma_c$ and thus $f_k^\dagger(x) \in \mathcal{I}$, or $x \notin \mathcal{X} + \sigma_c$ and then $\chi(x) = 0$ and $f_k^\dagger(x) = x \in \mathcal{I}$. It follows that all the requirements of global uniform Lipschitzness of $(f_k^{-1})_{k\in\mathbb{N}}$, $(h_k)_{k\in\mathbb{N}}$, and T_0 as in Assumption 3 can be replaced by uniform Lipschitzness on this backward invariant set $\mathcal{I}$, by replacing $(f_k^{-1})_{k\in\mathbb{N}}$ with $(f_k^\dagger)_{k\in\mathbb{N}}$ defined in (27) in all the equations. Similarly, in Remark 4, we can check uniform Lipschitz backward distinguishability using $(\mathcal{O}_k^{fw})_{k\in\mathbb{N}}$ instead of $(\mathcal{O}_k^{bw})_{k\in\mathbb{N}}$ if $(f_k^{-1})_{k\in\mathbb{N}}$ is uniformly Lipschitz injective on $\mathcal{I}$. Actually, even the invertibility of each f_k as in Assumption 2 may only be required on $\mathcal{I}$.

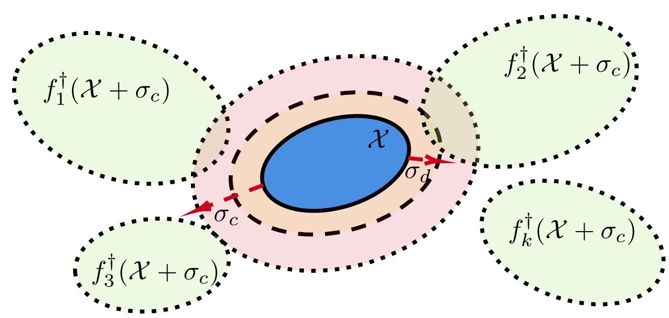

Fig. 1. Illustration of the backward invariant set $\mathcal{I}$ (the union of all).

In particular, $\mathcal{I}$ is bounded if and only if the sequence of sets $(f_k^\dagger(\mathcal{X} + \sigma_c))_{k\in\mathbb{N}}$ is *uniformly* bounded, which is guaranteed if $(f_k^{-1})_{k\in\mathbb{N}}$ is uniformly bounded on $\mathcal{X}+\sigma_c$. In this case, all those assumptions become much more favorable.

Remark 10: In the case of a discretization, $f_k(x) = x + \Delta t\Phi(x,t_k)$, where either Δt is very small or the function Φ is uniformly bounded (like in the PMSM example below in Section VI-B), then the maps $(f_k^\dagger)_{k\in\mathbb{N}}$ are close to identity and there is a good chance that the sets $(f_k^\dagger(\mathcal{X}+\sigma_c))_{k\in\mathbb{N}}$ should be close to $\mathcal{X}+\sigma_c$, which is known, and that $\mathcal{I}$ should be bounded.

V. Injectivity from Backward Distinguishability

In this part, we show the injectivity of $(T_k)_{k\in\mathbb{N}}$ after a certain time from *non-uniform* and *non-Lipschitz* backward distinguishability only. Note that, as illustrated in Section II, non-uniform injectivity can sometimes be insufficient to guarantee the asymptotic convergence of the observer.

Definition 2: The system (1) is backward distinguishable on a set $\mathcal{X}$ after time $k^\star$ if there exist an open set $\mathcal{O}$ containing $\mathrm{cl}(\mathcal{X})$ and $k^\star\in\mathbb{N}$ such that for each $k\geq k^\star$, for all $(x_a,x_b)\in\mathcal{O}\times\mathcal{O}$ with $x_a\neq x_b$, there exists a $j_k\in\{0,1,\ldots,k-1\}$ such that

$$(h_{j_k}\circ f_{j_k}^{-1}\circ f_{j_k+1}^{-1}\circ\ldots\circ f_{k-1}^{-1})(x_a)\neq (h_{j_k}\circ f_{j_k}^{-1}\circ f_{j_k+1}^{-1}\circ\ldots\circ f_{k-1}^{-1})(x_b).$$

In words, this means that given two different state values at a time k, there exists at least one instant in the past where their corresponding outputs have been different. Note that this is much lighter than the uniform Lipschitz backward distinguishability of Section IV—no uniformity of the sequence $(j_k)_{k\in\mathbb{N}}$ is required with respect to k nor to the pair (x_a,x_b). Therefore, this is one of the weakest forms of observability we may consider. For our injectivity result, we then make the following assumptions.

Assumption 4: We assume that:

(A4.1) For all $k\in\mathbb{N}$, the functions f_k^{-1} and h_k are C^1;

(A4.2) There exists $k^\star\in\mathbb{N}$ such that the system (1) is backward distinguishable on $\mathcal{X}$ after time $k^\star$.

Theorem 5 then gives injectivity results for $(T_k)_{k\in\mathbb{N}}$, with $T_0=0$ and for a generic choice of (A,B) of sufficient dimension. Its proof is based on the generalized Coron's Lemma developed recently in [12].

Theorem 5: Under Assumptions 1, 2, and 4, there exists a set $\mathcal{M}$ of zero Lebesgue measure in $\mathbb{R}^{(2n_x+1)\times(2n_x+1)}\times\mathbb{R}^{2n_x+1}$ such that for any pair $(\tilde{A},\tilde{B})\in(\mathbb{R}^{(2n_x+1)\times(2n_x+1)}\times\mathbb{R}^{2n_x+1})\setminus\mathcal{M}$ with $\tilde{A}$ Schur and any $k\geq k^\star$, the sequence of functions $(T_k)_{k\in\mathbb{N}}$ defined in (13) for

$$A = I_{n_y}\otimes\tilde{A}\in\mathbb{R}^{(2n_x+1)n_y\times(2n_x+1)n_y}, \tag{28a}$$

$$B = I_{n_y}\otimes\tilde{B}\in\mathbb{R}^{(2n_x+1)n_y\times n_y}, \tag{28b}$$

and initialized as $T_0=0$, is injective on $\mathcal{X}$.

Remark 11: Actually, the pair $(\tilde{A},\tilde{B})$ is chosen controllable and with $\tilde{A}$ diagonalizable, which is true for almost any such pair in $\mathbb{R}^{(2n_x+1)\times(2n_x+1)}\times\mathbb{R}^{2n_x+1}$.

Proof: Recall that the set $\mathcal{M}_{ND}$ of pairs of real matrices $(\tilde{A},\tilde{B})$ where $\tilde{A}$ is non-diagonalizable in $\mathbb{C}$ and the set $\mathcal{M}_{NC}$ of uncontrollable pairs of real matrices $(\tilde{A},\tilde{B})$ are both of zero measure in $\mathbb{R}^{(2n_x+1)\times(2n_x+1)}\times\mathbb{R}^{2n_x+1}$. Indeed, they are the zero-locus of non-identically zero polynomials: the discriminant of the characteristic polynomial for the former and the determinant of the controllability matrix for the latter. Now, consider $(\tilde{A},\tilde{B})$ in $\mathbb{R}^{(2n_x+1)\times(2n_x+1)}\times\mathbb{R}^{2n_x+1}$ controllable with $\tilde{A}$ Schur and diagonalizable in $\mathbb{C}$. Consider the maps $(T_k)_{k\in\mathbb{N}}$ defined in (13) with $T_0=0$, which can be written as

$$T_k(x) = \sum_{j=0}^{k-1}A^{k-j-1}B(h_j\circ f_j^{-1}\circ f_{j+1}^{-1}\circ\ldots\circ f_{k-1}^{-1})(x), \tag{29}$$

with (A,B) defined in (28). Define

$$\tilde{A}_{\mathrm{real}} = \mathrm{diag}(\Lambda_1,\Lambda_2,\ldots,\Lambda_l,\lambda_{l+1},\lambda_{l+2},\ldots,\lambda_{2n_x-l+1}),$$

$$\tilde{B}_{\mathrm{real}} = \begin{pmatrix} B_1\\ B_2\\ \ldots\\ B_{2n_x-l+1}\end{pmatrix},$$

where

$$\Lambda_i = \begin{pmatrix}\Re\lambda_i & -\Im(\lambda_i)\\ \Im(\lambda_i) & \Re(\lambda_i)\end{pmatrix},$$

$$B_i = \begin{cases}\begin{pmatrix}1\\0\end{pmatrix} & i\in\{1,2,\ldots,l\}\\ 1 & i\in\{l+1,l+2,\ldots,2n_x-l+1\},\end{cases}$$

where $l\in\{0,1,\ldots,n_x\}$ is the number of complex non-real eigenvalues of $\tilde{A}$ (that come in pairs of conjugates since $\tilde{A}$ is real). As shown in [12, Appendix B.1], there exists an invertible matrix $\tilde{P}\in\mathbb{R}^{(2n_x+1)\times(2n_x+1)}$ such that

$$\tilde{A}_{\mathrm{real}} = \tilde{P}^{-1}\tilde{A}\tilde{P},\quad \tilde{B}_{\mathrm{real}} = \tilde{P}^{-1}\tilde{B}.$$

First, since $\tilde{P}$ is invertible, the injectivity of the maps $(T_k)_{k\in\mathbb{N}}$ in (29) is implied by the injectivity of the maps $(T_{\mathrm{real},k})_{k\in\mathbb{N}}$ defined as

$$T_{\mathrm{real},k}(x) = (I_{n_y}\otimes\tilde{P}^{-1})T_k(x).$$

We have

$$\begin{aligned}
&T_{\mathrm{real},k}(x)\\
&= (I_{n_y}\otimes\tilde{P}^{-1})T_k(x)\\
&= (I_{n_y}\otimes\tilde{P}^{-1})\sum_{j=0}^{k-1}A^{k-j-1}B(h_j\circ f_j^{-1}\circ f_{j+1}^{-1}\circ\ldots\circ f_{k-1}^{-1})(x)\\
&= (I_{n_y}\otimes\tilde{P}^{-1})\sum_{j=0}^{k-1}(I_{n_y}\otimes\tilde{A})^{k-j-1}(I_{n_y}\otimes\tilde{B})\\
&\qquad (h_j\circ f_j^{-1}\circ f_{j+1}^{-1}\circ\ldots\circ f_{k-1}^{-1})(x)\\
&= \sum_{j=0}^{k-1}(I_{n_y}\otimes(\tilde{P}^{-1}\tilde{A}\tilde{P}))^{k-j-1}(I_{n_y}\otimes(\tilde{P}^{-1}\tilde{B}))\\
&\qquad (h_j\circ f_j^{-1}\circ f_{j+1}^{-1}\circ\ldots\circ f_{k-1}^{-1})(x)\\
&= \sum_{j=0}^{k-1}A_{\mathrm{real}}^{k-j-1}B_{\mathrm{real}}(h_j\circ f_j^{-1}\circ f_{j+1}^{-1}\circ\ldots\circ f_{k-1}^{-1})(x),
\end{aligned}$$

with the pair $(A_{\mathrm{real}},B_{\mathrm{real}})$ defined as

$$A_{\mathrm{real}} = I_{n_y}\otimes\tilde{A}_{\mathrm{real}},\quad B_{\mathrm{real}} = I_{n_y}\otimes\tilde{B}_{\mathrm{real}}.$$

Second, we prove the injectivity of the maps $(T_{\text{real},k})_{k\in\mathbb{N}}$. Define now the open sets $\Upsilon = \{(x_a, x_b) \in \mathcal{O}\times\mathcal{O} : x_a \neq x_b\}$ and $\Lambda_l = (B_1(0))^l \times (-1,1)^{2n_x-l+1}$. For $\lambda \in \mathbb{C}$, define the map $\mathcal{T}_{\lambda,k}$ as

$$T_{\lambda,k}(x) = \sum_{j=0}^{k-1} \lambda^{k-j-1}(h_j \circ f_j^{-1} \circ f_{j+1}^{-1} \circ \ldots \circ f_{k-1}^{-1})(x).$$

With the structure of $\tilde{A}_{\text{real}}$ and $\tilde{B}_{\text{real}}$, the functions $(T_{\text{real},k})_{k\in\mathbb{N}}$ can be written up to a permutation as

$$\begin{aligned} T_{\text{real},k}(x) = (&\Re(T_{\lambda_1,k}(x)), \Im(T_{\lambda_1,k}(x)), \ldots, \Re(T_{\lambda_l,k}(x)),\\ &\Im(T_{\lambda_l,k}(x)), T_{\lambda_{l+1},k}(x), \ldots, T_{\lambda_{2n_x-l+1},k}(x)). \end{aligned}$$

It follows that proving the injectivity of $T_{\text{real},k}$ for some $(\lambda_1, \lambda_2, \ldots, \lambda_{2n_x-l+1}) \in \Lambda_l$ is equivalent to proving the injectivity of

$$T_{\text{complex},k}(x) = (T_{\lambda_1,k}(x), T_{\lambda_2,k}(x), \ldots, T_{\lambda_{2n_x-l+1},k}(x)).$$

We now prove that this is guaranteed for all $k \geq k^\star$ and for almost any choice of $(\lambda_1, \lambda_2, \ldots, \lambda_{2n_x-l+1}) \in \Lambda_l$ in the Lebesgue measure sense. For that, we define the sets

$$\Theta_i = \begin{cases} B_1(0) & i \in \{1,2,\ldots,l\} \\ (-1,1) & i \in \{l+1, l+2, \ldots, 2n_x - l + 1\}, \end{cases}$$

the counters

$$p_i = \begin{cases} 2 & i \in \{1,2,\ldots,l\} \\ 1 & i \in \{l+1, l+2, \ldots, 2n_x - l + 1\}, \end{cases}$$

and the functions $g_{i,k} : \Upsilon \times \Theta_i \to \mathbb{R}$ for $i \in \{1,2,\ldots,l\}$ and $g_{i,k} : \Upsilon \times \Theta_i \to \mathbb{C}$ for $i \in \{l+1, l+2, \ldots, 2n_x - l + 1\}$, by

$$\begin{aligned} g_{i,k}((x_a,x_b),\lambda) &= \mathcal{T}_{\lambda,k}(x_a) - \mathcal{T}_{\lambda,k}(x_b) \\ &= \sum_{j=0}^{k-1} \lambda^{k-j-1}\big((h_j \circ f_j^{-1} \circ f_{j+1}^{-1} \circ \ldots \circ f_{k-1}^{-1})(x_a) \\ &\qquad - (h_j \circ f_j^{-1} \circ f_{j+1}^{-1} \circ \ldots \circ f_{k-1}^{-1})(x_b)\big). \end{aligned}$$

Now, we check the conditions for the generalized Coron's Lemma in [12, Lemma B.3]. For any $l \in \{0, 1, \ldots, n_x\}$,

- For $i \in \{1,2,\ldots,l\}$, we see that $g_{i,k}((x_a,x_b),\cdot)$ is holomorphic on $B_1(0)$ for all $(x_a,x_b) \in \Upsilon$. From the chain rule, under Assumption (A4.1), for all $\lambda \in B_1(0)$, $g_{i,k}(\cdot,\lambda)$ is C^1 on Υ for each $\lambda \in B_1(0)$ because it is a finite composition of C^1 functions;
- For $i \in \{l+1, l+2, \ldots, 2n_x - l + 1\}$, as $g_{i,k}((x_a,x_b),\cdot)$ is a polynomial, it is C^∞ on $(-1,1)$ for all $(x_a,x_b) \in \Upsilon$. From the chain rule, under Assumption (A4.1), for all $\lambda \in (-1,1)$ and for all $j \in \mathbb{N}$, the maps $\frac{\partial^j g_{i,k}}{\partial \lambda^j}(\cdot,\lambda)$ are C^1 on Υ because they are finite compositions of C^1 functions.

We then show that under Assumption (A4.2), for all $i \in \{1,2,\ldots,2n_x - l + 1\}$, $g_{i,k}((x_a,x_b),\cdot)$ cannot be identically zero on Θ_i. Take $(x_a,x_b) \in \Upsilon$ and take $k \in \mathbb{N} : k \geq k^\star$ and assume $g_{i,k}((x_a,x_b),\lambda) = 0$ for all $\lambda \in \Theta_i$. By uniqueness of polynomials, for all $j \in \{0,1,\ldots,k-1\}$, we have $(h_j \circ f_j^{-1} \circ f_{j+1}^{-1} \circ \ldots \circ f_{k-1}^{-1})(x_a) = (h_j \circ f_j^{-1} \circ f_{j+1}^{-1} \circ \ldots \circ f_{k-1}^{-1})(x_b)$, which contradicts Assumption (A4.2). From the generalized Coron's Lemma [12] applied at each $l \in \{0,1,\ldots,n_x\}$ and each $k \geq k^\star$, since $\sum_{i=1}^{2n_x-l+1} p_i = 2n_x + 1$, the set

$$\begin{aligned} \mathcal{E}_{l,k} = \bigcup_{(x_a,x_b)\in\Upsilon} \Big\{&(\lambda_1,\lambda_2,\ldots,\lambda_{2n_x-l+1}) \in \Lambda_l \mid \\ &\forall i \in \{1,2,\ldots,2n_x-l+1\}, g_{i,k}((x_a,x_b),\lambda_i) = 0\Big\}, \end{aligned}$$

which is literally the set of eigenvalues in Λ_l making $T_{\text{complex},k}$ (at each k) non-injective, has zero Lebesgue measure. Then, from [12, Lemma B.2], the set

$$\begin{aligned} \mathcal{M}_{l,k} = \{(\tilde{A},\tilde{B}) \in \mathbb{R}^{(2n_x+1)\times(2n_x+1)} \times \mathbb{R}^{2n_x+1},& \\ \tilde{A} \text{ has the eigenvalues in } \mathcal{E}_{l,k}\}& \end{aligned}$$

also has zero measure. Now, recall that the *countable* union of infinitely many zero Lebesgue measure sets also has zero Lebesgue measure [46]. Therefore, the set

$$\mathcal{M} = \mathcal{M}_{ND} \cup \mathcal{M}_{NC} \cup \bigcup_{\substack{k\in\mathbb{N} \\ l\in\{0,1,\ldots,n_x\}}} \mathcal{M}_{l,k}$$

also has zero Lebesgue measure. ■

It is interesting to see that this injectivity result is proven differently from the continuous-time case in [4], due to the different nature of time. Indeed, the continuous time t belongs to the open uncountable set $[0,+\infty)$, so the result in [4, Theorem 3] is proven with Coron's Lemma applied only once to a set Υ that contains time. However, the discrete time k belongs to $\mathbb{N}$, which is not open but countable, so the generalized Coron's Lemma is here applied separately at each instant k, and the result is then obtained for the whole time domain by the countable union of zero-measure sets.

Example 4: Consider the system in Example 1. It verifies the backward distinguishability condition in Assumption (A4.2) as long as there exists k such that $h_k \neq 0$. Therefore, Theorem 5 applies with $T_0 = 0$ (so $m_0 = 0$): there exists a sequence of injective maps $(T_k)_{k\in\mathbb{N}}$ (from a certain time) transforming the dynamics into a form (3).

This result only ensures the injectivity of each map T_k after a certain time, without any uniformity in k, which may impair convergence, as seen in Example 1. However, we saw in Example 1 that injectivity alone can still suffice in some cases. Therefore, if we initialize $T_0 = 0$, the observer may still work under backward distinguishability only, which is a very mild observability condition.

Remark 12: In general, solutions to (5) taking the form (13) is written as $T_k(x) = \mathcal{I}_k(x) + \mathcal{T}_k(x)$ where $\mathcal{I}_k(x) = A^k(T_0 \circ f_0^{-1} \circ f_1^{-1} \circ \ldots \circ f_{k-1}^{-1})(x)$ and $\mathcal{T}_k(x) = \sum_{j=0}^{k-1} A^{k-j-1}B(h_j \circ f_j^{-1} \circ f_{j+1}^{-1} \circ \ldots \circ f_{k-1}^{-1})(x)$. In Theorem 5, we prove the injectivity of $(T_k)_{k\in\mathbb{N}}$ for all $k \geq k^\star$ assuming $T_0 = 0$, namely the injectivity of $(\mathcal{T}_k)_{k\in\mathbb{N}}$. Therefore, it is advised to initialize $(T_k)_{k\in\mathbb{N}}$ such that T_0 is identically zero, if possible. In a stronger case, if $(\mathcal{T}_k)_{k\in\mathbb{N}}$ is uniformly injective, i.e., there exist a class-$\mathcal{K}$ function κ and $l \in \mathbb{R}_{>0}$ such that for all $k \geq k^\star$ and for all $(x_a,x_b) \in \mathcal{X}\times\mathcal{X}$,

$$|\mathcal{T}_k(x_a) - \mathcal{T}_k(x_b)| \geq l\kappa(|x_a - x_b|),$$

and if for all $k \geq k^\star$ and for all $(x_a, x_b) \in \mathcal{X} \times \mathcal{X}$,

$$|T_0(x_a) - T_0(x_b)| \leq \kappa(|x_a - x_b|),$$

then

$$\begin{aligned} |T_k(x_a) - T_k(x_b)| &= |\mathcal{I}_k(x_a) - \mathcal{I}_k(x_b) + \mathcal{T}_k(x_a) - \mathcal{T}_k(x_b)| \\ &\geq |\mathcal{T}_k(x_a) - \mathcal{T}_k(x_b)| - |\mathcal{I}_k(x_a) - \mathcal{I}_k(x_b)| \\ &\geq (l - 2\|A\|^k)\kappa(|x_a - x_b|), \end{aligned}$$

which implies that $(T_k)_{k\in\mathbb{N}}$ becomes uniformly injective after a certain time. This, as seen in Theorem 1, is sufficient for an asymptotic observer assuming that the inverse map of κ is concave, whose dynamics unfortunately cannot be assigned arbitrarily fast.

VI. Examples

A. System with Linear Dynamics and Output

Consider a linear time-varying discrete systems of form

$$x_{k+1} = F_k x_k, \quad y_k = H_k x_k. \tag{30}$$

A linear transformation $x_k \mapsto z_k = T_k x_k$ into (3) can be found with the sequence of matrices $(T_k)_{k\in\mathbb{N}}$ satisfying

$$T_{k+1}F_k = AT_k + BH_k,$$

initialized as T_0. Under invertibility of the sequence $(F_k)_{k\in\mathbb{N}}$, it is defined by the closed form

$$T_k = A^k T_0 \prod_{j=0}^{k-1} F_j^{-1} + \sum_{j=0}^{k-1} A^{k-j-1} B H_j \prod_{q=j}^{k-1} F_q^{-1},$$

for all $k \in \mathbb{N}_{>0}$. Then, provided each T_k is full-rank, and thus left-invertible (see below), the KKL observer takes the form

$$\begin{cases} \hat{z}_{k+1} = A\hat{z}_k + By_k \\ T_{k+1} = AT_kF_k^{-1} + BH_kF_k^{-1} \end{cases}, \quad \hat{x}_k = T_k^* \hat{z}_k \tag{31}$$

where T_k^* is a left inverse of T_k.

The system (30) is uniformly Lipschitz backward distinguishable (see Definition 1) if and only if there exists $m \in \mathbb{N}_{>0}$ such that there exists $c_o \in \mathbb{R}_{>0}$ such that for all $k \geq m$, the backward distinguishability matrix

$$\mathcal{O}_k^{bw} = \begin{pmatrix} H_{k-1}F_{k-1}^{-1} \\ H_{k-2}F_{k-2}^{-1}F_{k-1}^{-1} \\ \dots \\ H_{k-(m-1)}F_{k-(m-1)}^{-1}\dots F_{k-1}^{-1} \\ H_{k-m}F_{k-m}^{-1}F_{k-(m-1)}^{-1}\dots F_{k-1}^{-1} \end{pmatrix}$$

verifies

$$\mathcal{O}_k^{bw\top}\mathcal{O}_k^{bw} \geq c_o I > 0. \tag{32}$$

Alternatively, under uniform boundedness of $(F_k)_{k\in\mathbb{N}}$, we can use the forward version similar to the one in Remark 4. According to Theorem 3, under (32) and the uniform boundedness of $(F_k, H_k)_{k\in\mathbb{N}}$, picking A sufficiently fast of dimension m, there exist $c_t \in \mathbb{R}_{>0}$ and $k^\star \in \mathbb{N}$ such that for all $k \geq k^\star$, $T_k^\top T_k \geq c_t I > 0$, namely $(T_k)_{k\in\mathbb{N}}$ is (uniformly) left-invertible for k sufficiently large. Therefore, (31) is implementable and provides arbitrarily fast robust exponentially stable estimation for (30).

Interestingly, (32) coincides with the *uniform complete observability* condition required by the Kalman filter (see [30, Condition (13)] or [31, Definition 3]), namely there exist $m \in \mathbb{N}_{>0}$ and $c_o \in \mathbb{R}_{>0}$ such that for all $k \geq m$,

$$\sum_{j=k-m}^{k-1} F_{k-1}^{-1\top} F_{k-2}^{-1\top} \dots F_j^{-1\top} H_j^\top H_j F_j^{-1} \dots F_{k-2}^{-1} F_{k-1}^{-1} \geq c_o I > 0.$$

It is thus interesting to compare both designs. In terms of dimensions, the complexity of the Kalman filter is $\frac{n_x(n_x+1)}{2} + n_x$, while that of the KKL observer is $(mn_y)^2 + mn_y$ with $mn_y \geq n_x$ (or $\sum_{i=1}^{n_y} m_i$ instead of mn_y if the observability multiplicities m_i are considered in (32)). Therefore, the Kalman filter is advantageous in dimension compared to the KKL observer. However, the advantage of the latter (besides being applicable in the nonlinear context) is that, there exists a strict ISS Lyapunov function $V_k : \mathbb{R}^{n_x} \to \mathbb{R}_{\geq 0}$ of quadratic form

$$V_k(x_k) = (x_k - \hat{x}_k)^\top T_k^\top P T_k (x_k - \hat{x}_k),$$

where $P \in \mathbb{R}^{n_z \times n_z}$ is a positive definite solution to $A^\top PA - P < 0$, and verifying

$$\alpha x_k^\top x_k \leq V_k(x_k), \quad \forall k \geq k^\star,$$

for some $\alpha \in \mathbb{R}_{>0}$ independent of k. Exponential ISS of the estimation error can thus be proven with an explicit quadratic Lyapunov function, unlike the discrete Kalman filter [30] whose Lyapunov function is not strict.

B. Permanent Magnet Synchronous Motor

Consider a permanent magnet synchronous motor (PMSM) with the reproduced model [47]

$$\dot{x} = u - Ri, \quad y = |x - Li|^2 - \Phi^2 = 0, \tag{33}$$

where $x \in \mathbb{R}^2$ is the electromagnetic flux (in Vs); the voltages u (in V) and currents i (in A) are inputs in $\mathbb{R}^2$; the resistance $R = 1.45$ (Ω), the inductance $L = 0.0121$ (H), and the flux $\Phi = 0.1994$ (Vs) are constant parameters. Here in this example, the value of the output y is always zero. Even though this system has linear dynamics, its quadratic output map renders observer design very challenging and thus necessitates thorough studies [14], [15], [43], [47], [48].

We see that the function

$$H(x, u, i, \dot{u}, \frac{di}{dt}, \frac{d^2i}{dt^2}) = \begin{pmatrix} |x - Li|^2 - \Phi^2 \\ 2\eta^\top(x - Li) \\ 2\dot{\eta}^\top(x - Li) + 2\eta^\top\eta \end{pmatrix},$$

describing the output and its two first time derivatives, with $\eta = u - Ri + L\frac{di}{dt}$, is uniformly Lipschitz injective if there exists $c_\eta \in \mathbb{R}_{>0}$ such that

$$\begin{pmatrix} \eta^\top \\ \dot{\eta}^\top \end{pmatrix}^\top \begin{pmatrix} \eta^\top \\ \dot{\eta}^\top \end{pmatrix} \geq c_\eta I > 0. \tag{34}$$

It can be shown that this property holds if the motor speed is uniformly bounded away from zero [48]. Following [48], a continuous-time KKL observer with a sufficiently fast continuous pair (A, B) of dimension 3 can be designed for this system. But actually, in practice, the input signals u and i are only known at specific sampling times, typically related to the PWM. Two paths are then possible:

- Design a *continuous* KKL observer for the continuous model and then discretize it at the sampling rate; or
- Build a discretized model of the system at the sampling rate and design a *discrete* KKL observer for this discrete model.

Intuitively, both paths should be equivalent for small sampling times Δt. However, for a PMSM discretized at the PWM, discretization errors are significant at high speeds and we illustrate here the great interest of following the second path. Indeed, it offers the crucial advantage of using an appropriate discretization, adapted to the physics of the system, which is not the case in the first path where physical insight is much trickier to exploit for the observer discretization.

1) Discrete KKL Observer with Euler's Method:

One way to discretize the PMSM is by using Euler's method

$$x_{k+1} = x_k + \Delta t(u_k - Ri_k), \quad y_k = |x_k - Li_k|^2 - \Phi^2 = 0. \tag{35}$$

Let us now verify the assumptions needed for observer design, more particularly those required by Theorem 3.

- Assumption 1: The solutions of (35), when injected with sinusoidal inputs, are also sine waves, so they remain in a compact set in positive time;
- Assumption 2: The dynamics map of (35), with $(u_k)_{k\in\mathbb{N}}$ and $(i_k)_{k\in\mathbb{N}}$ known, is invertible;
- Assumption 3: First, uniform Lipschitzness of the inverse dynamics and output maps of (35) holds since the inputs $(u_k)_{k\in\mathbb{N}}$ and $(i_k)_{k\in\mathbb{N}}$ are uniformly bounded and solutions remain in a compact set. Second, the uniform Lipschitz backward distinguishability is very hard to check analytically in discrete time because it involves the inversion of the dynamics. We thus use its continuous-time version related to (34) to argue that the equivalent property should hold in discrete time if the sampling period Δt is sufficiently small.

Guided by Example 2 and the knowledge that a KKL observer of dimension 3 exists in continuous time, we look for a transformation of the form

$$z_k = T_k(x_k) = a_k|x_k|^2 + b_k x_k + c_k \in \mathbb{R}^3,$$

where

$$\begin{aligned} a_k &= \begin{pmatrix} a_{1,k} & a_{2,k} & a_{3,k} \end{pmatrix}^\top \in \mathbb{R}^3, \\ b_k &= \begin{pmatrix} b_{1,k} & b_{2,k} & b_{3,k} \end{pmatrix}^\top \in \mathbb{R}^{3\times 2}, \\ c_k &= \begin{pmatrix} c_{1,k} & c_{2,k} & c_{3,k} \end{pmatrix}^\top \in \mathbb{R}^3. \end{aligned}$$

Note that each $b_{i,k}$, $i = 1, 2, 3$ is a vector in $\mathbb{R}^2$. With $A \in \mathbb{R}^{3\times 3}$ Schur and the pair (A, B) controllable, z_k is solution to (3) if

$$\begin{aligned} a_{k+1} &= Aa_k + B, \\ b_{k+1} &= Ab_k - 2\Delta t a_{k+1}(u_k - Ri_k)^\top - 2LBi_k^\top, \\ c_{k+1} &= Ac_k - \Delta t^2 a_{k+1}|u_k - Ri_k|^2 \\ &\qquad - \Delta t b_{k+1}(u_k - Ri_k) + B(L^2|i_k|^2 - \Phi^2). \end{aligned} \tag{36}$$

Note that a_k can be picked constant equal to $(I - A)^{-1}B$. Because $y_k = 0$ for all k, z_k converges to zero exponentially fast and it is straightforward to pick for instance the particular solution $\hat{z}_k = 0$ for the observer. Then, the estimate is obtained by solving $T_k(\hat{x}_k) = \hat{z}_k = 0$, namely

$$\hat{x}_k = -\begin{pmatrix} a_{1,k}b_{2,k} - a_{2,k}b_{1,k} \\ a_{1,k}b_{3,k} - a_{3,k}b_{1,k} \end{pmatrix}^{-1} \begin{pmatrix} a_{1,k}c_{2,k} - a_{2,k}c_{1,k} \\ a_{1,k}c_{3,k} - a_{3,k}c_{1,k} \end{pmatrix}. \tag{37}$$

2) Discrete KKL Observer with Rotation Correction:

According to [49], a more appropriate method to discretize the PMSM (33) taking into account its rotating dynamics is

$$\begin{aligned} x_{k+1} &= x_k + \Delta t\Omega_k(u_k - Ri_k)\operatorname{sinc}(\varphi_k), \\ y_k &= |x_k - Li_k|^2 - \Phi^2 = 0, \end{aligned} \tag{38}$$

where $\Omega_k = \begin{pmatrix} \cos(\varphi_k) & -\sin(\varphi_k) \\ \sin(\varphi_k) & \cos(\varphi_k) \end{pmatrix}$ and $\varphi_k = \frac{\hat{\omega}_k \Delta t}{2}$ where

$$\hat{\omega}_k = \operatorname{sign}((u_k - u_{k-1})^\top u_{k-1}) \frac{|u_k - u_{k-1}|}{\Delta t|u_k|}$$

is the estimate of the motor's rotation speed that is approximately the same for $(u_k)_{k\in\mathbb{N}}$, $(i_k)_{k\in\mathbb{N}}$, and $(x_k)_{k\in\mathbb{N}}$, assuming that this speed does not vary too fast. Notice that when $\hat{\omega}_k = 0$ for all k (no rotation), we recover Euler's discretized version in (35). We also see that (38), with the inputs $(u_k)_{k\in\mathbb{N}}$ and $(i_k)_{k\in\mathbb{N}}$ being sinusoidal, satisfies all the assumptions required by Theorem 3.

Keeping the same pair (A, B), we get this time

$$\begin{aligned} a_{k+1} &= Aa_k + B, \\ b_{k+1} &= Ab_k - 2\Delta t a_{k+1}(\Omega_k(u_k - Ri_k)\operatorname{sinc}(\varphi_k))^\top - 2LBi_k^\top, \\ c_{k+1} &= Ac_k - \Delta t^2 a_{k+1}\operatorname{sinc}^2(\varphi_k)|\Omega_k(u_k - Ri_k)|^2 \\ &\quad - \Delta t b_{k+1}\Omega_k(u_k - Ri_k)\operatorname{sinc}(\varphi_k) + B(L^2|i_k|^2 - \Phi^2), \end{aligned} \tag{39}$$

with a_k still possibly constant equal to $(I - A)^{-1}B$, and the estimate is still obtained with (37).

3) Comparison of Performance:

Due to space constraints, we only show the estimation error for one of the two state components, the other one being similar. In Figure 2, the estimation errors with respect to the continuous-time trajectory of (33) are compared among the three cases: 1) A continuous KKL observer designed following [47] and discretized using Euler's method; 2) A discrete KKL observer designed based on the Euler discretization of (33); 3) A discrete KKL observer designed based on the discretization of (33) with rotation correction. From here, we draw two important lessons: 1) It seems better to design a discrete observer from a discretized model than to discretize a continuous observer already designed; 2) The numerical errors due to incorrect discretization may be reduced by taking into account the system's physics in the discrete model.

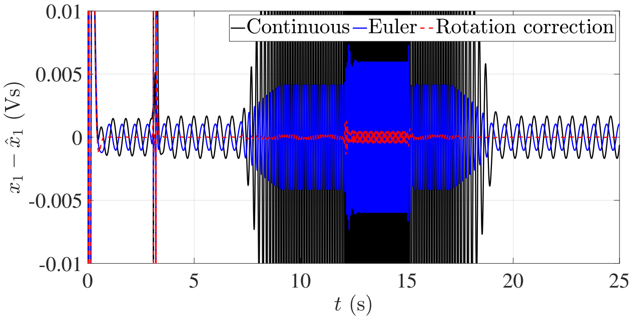

Fig. 2. Estimation error by: 1) The continuous KKL observer from [47] with $A_c = -\mathrm{diag}(10, 44, 80)$ discretized at $\Delta t = 0.001$ (s) (Euler); 2) The discrete KKL observer (36), (37); 3) The discrete KKL observer with rotation correction (39), (37), for $A = e^{\Delta t A_c}$ and $B = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}^\top$.

Simulations with multiple choices of γ in the rotation correction case are compared in Figure 3. It is observed that a smaller γ gives a faster convergence, but a more serious amplification of numerical noise, which is coherent with the robustness results in Theorem 4. However, in the region of too high rotating speeds, the three designs tend to perform the same, since the discretized model becomes less appropriate, which is something the observers cannot deal with. Last, it is interesting to notice that in this application case, as we choose $\hat{z}_k = 0$ for the observer in the z-coordinates, it is indeed the transformation $(T_k)_{k\in\mathbb{N}}$ that serves to provide the estimation.

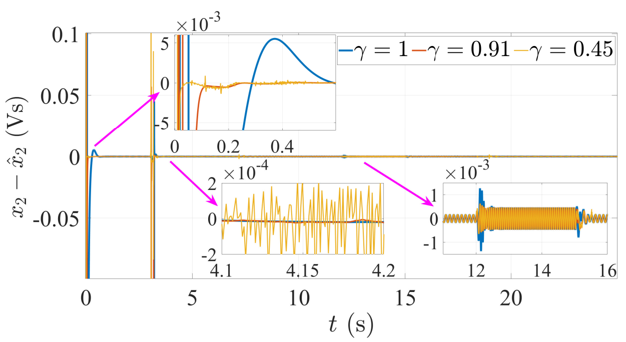

Fig. 3. Comparison among different choices of γ in the rotation correction case.

VII. Conclusion

This work presents the KKL observer design for nonlinear time-varying discrete systems. After giving the closed form of the transformation $(T_k)_{k\in\mathbb{N}}$ into an exponentially stable filter of the measurement, we have shown how the uniform Lipschitz injectivity of this transformation is achieved after a certain time under uniform Lipschitz backward distinguishability if the target dynamics is sufficiently fast. This result provides an arbitrarily fast discrete observer that is ISS with respect to uncertainties, input disturbances, and measurement noise. For linear systems, this provides an alternative to the discrete Kalman filter, with an explicit quadratic ISS Lyapunov function. We have also shown how non-uniform injectivity of the transformations is achieved under backward distinguishability, a mild observability condition, which in some cases is enough for observer design. The example of a PMSM with sampled inputs illustrates how it may be more efficient to design a discrete KKL observer for an appropriate faithful discrete model of the system, instead of discretizing a continuous KKL observer designed for the continuous model.

A future question is how to develop or find an alternative to the generalized Coron's Lemma to obtain a uniform injectivity result possibly without Lispchitzness and arbitrarily fast convergence, typically through a uniform non-Lipschitz distinguishability property. Finally, a major drawback of KKL observers is in the numerical computation of the transformation, because closed-form expressions such as (13) are hard to achieve analytically for general nonlinear systems. To address this, numerical tools are currently being developed [16], [17].

Acknowledgment

The authors would like to thank Florent Di Meglio, Vincent Andrieu, and Lucas Brivadis for their useful remarks and fruitful discussions.

References

[1] P. Bernard, V. Andrieu, and D. Astolfi. Observer Design for Continuous-time Dynamical Systems. *Annual Reviews in Control*, 53:224–248, 2022.

[2] V. Andrieu and L. Praly. On the Existence of a Kazantzis–Kravaris/Luenberger Observer. *SIAM Journal on Control and Optimization*, 45(2):432–456, 2006.

[3] V. Andrieu. Convergence Speed of Nonlinear Luenberger Observers. *SIAM Journal on Control and Optimization*, 52(5):2831–2856, 2014.

[4] P. Bernard and V. Andrieu. Luenberger Observers for Nonautonomous Nonlinear Systems. *IEEE Transactions on Automatic Control*, 64(1):270–281, 2019.

[5] L. Brivadis, V. Andrieu, and U. Serres. Luenberger Observers for Discrete-time Nonlinear Systems. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 3435–3440, 2019.

[6] J.-M. Coron. On the Stabilization of Controllable and Observable Systems by an Output Feedback Law. *Math. Control Signals Systems*, 7(3):187–216, 1994.

[7] D. G. Luenberger. Observing the State of a Linear System. *IEEE Transactions on Military Electronics*, 8(2):74–80, 1964.

[8] A.N. Shoshitaishvili. Singularities for Projections of Integral Manifolds with Applications to Control and Observation Problems. *Theory of Singularities and its Applications*, 1:295, 1990.

[9] N. Kazantzis and C. Kravaris. Nonlinear Observer Design Using Lyapunov's Auxiliary Theorem. *Systems & Control Letters*, 34(5):241–247, 1998.

[10] A. J. Krener and M.Q. Xiao. Nonlinear Observer Design in the Siegel Domain through Coordinate Changes. *IFAC Proceedings Volumes*, 34(6):519–524, 2001.

[11] G. Kreisselmeier and R. Engel. Nonlinear Observers for Autonomous Lipschitz Continuous Systems. *IEEE Transactions on Automatic Control*, 48:451–464, 2003.

[12] L. Brivadis, V. Andrieu, P. Bernard, and U. Serres. Further Remarks on KKL Observers. *Systems & Control Letters*, 172:105429, 2023.

[13] C. Afri, V. Andrieu, L. Bako, and P. Dufour. State and Parameter Estimation: A Nonlinear Luenberger Observer Approach. *IEEE Transactions on Automatic Control*, 62(2):973–980, 2017.

[14] N. Henwood, J. Malaizé, and L. Praly. A Robust Nonlinear Luenberger Observer for the Sensorless Control of SM-PMSM : Rotor Position and Magnets Flux Estimation. *IECON Conference on IEEE Industrial Electronics Society*, 2012.

[15] P. Bernard and L. Praly. Estimation of Position and Resistance of a Sensorless PMSM : A Nonlinear Luenberger Approach for a Non-observable System. *IEEE Transactions on Automatic Control*, 66:481–496, 2021.

[16] L. da C. Ramos, F. Di Meglio, V. Morgenthaler, L. F. F. da Silva, and P. Bernard. Numerical Design of Luenberger Observers for Nonlinear Systems. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 5435–5442, 2020.

[17] M. Buisson-Fenet, L. Bahr, and F. Di Meglio. Learning to Observe : Neural Network-based KKL Observers. Python toolbox available at https://github.com/Centre-automatique-et-systemes/learn_observe_KKL.git, 2022.
[18] J. Peralez and M. Nadri. Deep Learning-based Luenberger Observer Design for Discrete-time Nonlinear Systems. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 4370–4375. IEEE, 2021.
[19] M. Buisson-Fenet, L. Bahr, and F. Di Meglio. Towards Gain Tuning for Numerical KKL Observers. Available at https://arxiv.org/abs/2204.00318, 2022.
[20] J.P. Gauthier, H. Hammouri, and S. Othman. A Simple Observer for Nonlinear Systems Applications to Bioreactors. *IEEE Transactions on Automatic Control*, 37(6):875–880, 1992.
[21] A. M. Dabroom and H. K. Khalil. Discrete-time Implementation of High-gain Observers for Numerical Differentiation. *International Journal of Robust and Nonlinear Control*, 72(17):1523–1537, 1999.
[22] S. Hanba. Further Results on the Uniform Observability of Discrete-Time Nonlinear Systems. *IEEE Transactions on Automatic Control*, 55(4):1034–1038, 2010.
[23] C.V. Rao, J.B. Rawlings, and D.Q. Mayne. Constrained State Estimation for Nonlinear Discrete-time Systems: Stability and Moving Horizon Approximations. *IEEE Transactions on Automatic Control*, 48(2):246–258, 2003.
[24] A. Alessandri, M. Baglietto, and G. Battistelli. Moving-horizon State Estimation for Nonlinear Discrete-time Systems: New Stability Results and Approximation Schemes. *Automatica*, 44(7):1753–1765, 2008.
[25] A. Alessandri and M. Gaggero. Fast Moving Horizon State Estimation for Discrete-Time Systems Using Single and Multi Iteration Descent Methods. *IEEE Transactions on Automatic Control*, 62(9):4499–4511, 2017.
[26] A. Jazwinski. Limited Memory Optimal Filtering. *IEEE Transactions on Automatic Control*, 13(5):558–563, 1968.
[27] P.E. Moraal and J.W. Grizzle. Observer Design for Nonlinear Systems with Discrete-time Measurements. *IEEE Transactions on Automatic Control*, 40(3):395–404, 1995.
[28] S. Ammar, M. Mabrouk, and J.-C. Vivalda. On the Genericity of the Differential Observability of Controlled Discrete-Time Systems. *SIAM Journal on Control and Optimization*, 46(6):2182–2198, 2007.
[29] S. Hanba. On the "Uniform" Observability of Discrete-Time Nonlinear Systems. *IEEE Transactions on Automatic Control*, 54(8):1925–1928, 2009.
[30] Jr. J. Deyst and C. Price. Conditions for Asymptotic Stability of the Discrete Minimum-variance Linear Estimator. *IEEE Transactions on Automatic Control*, 13:702–705, 1968.
[31] Q. Zhang. On Stability of the Kalman Filter for Discrete Time Output Error Systems. *Systems and Control Letters*, 107:84–91, July 2017.
[32] Y. Song and J. W. Grizzle. The Extended Kalman Filter as a Local Asymptotic Observer for Nonlinear Discrete-Time Systems. In *1992 American Control Conference*, pages 3365–3369, 1992.
[33] M. Boutayeb, H. Rafaralahy, and M. Darouach. Convergence Analysis of the Extended Kalman Filter Used as an Observer for Nonlinear Deterministic Discrete-Time Systems. *IEEE Transactions on Automatic Control*, 42(4):581–586, 1997.
[34] M. Boutayeb and D. Aubry. A Strong Tracking Extended Kalman Observer for Nonlinear Discrete-Time Systems. *IEEE Transactions on Automatic Control*, 44:1550 – 1556, 09 1999.
[35] G. Ciccarella, M. Dalla Mora, and A. Germani. Observers for Discrete-time Nonlinear Systems. *Systems & Control Letters*, 20(5):373–382, 1993.
[36] W. Lin and C. I. Byrnes. Remarks on Linearization of Discrete-time Autonomous Systems and Nonlinear Observer Design. *Systems & Control Letters*, 25(1):31–40, 1995.
[37] C. Califano, S. Monaco, and D. Normand-Cyrot. On the Observer Design in Discrete-time. *Systems & Control Letters*, 49(4):255–265, 2003.
[38] S. Ibrir. Circle-criterion Approach to Discrete-time Nonlinear Observer Design. *Automatica*, 43(8):1432–1441, 2007.
[39] S. Monaco and D. Normand-Cyrot. The Immersion under Feedback of a Multidimensional Discrete-time Non-linear System into a Linear System. *International Journal of Control*, 38:245–261, 07 1983.
[40] E. J. McShane. Extension of Range of Functions. *Bulletin of the American Mathematical Society*, 40(12):837 – 842, 1934.
[41] Z. Wang, T. N. Dinh, Q. Zhang, T. Raïssi, and Y. Shen. Fast Interval Estimation for Discrete-time Linear Systems: An L1 Optimization Method. *Automatica*, 137:110029, 2022.
[42] V. Andrieu. Convergence Speed of Nonlinear Luenberger Observers. *SIAM Journal on Control and Optimization*, 52(5):2831–2856, 2014.
[43] P. Bernard, T. Devos, A. K. Jebai, P. Martin, and L. Praly. KKL Observer Design for Sensorless Induction Motors. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 1235–1242, 2022.
[44] D. A. Allan, J. Rawlings, and A. R. Teel. Nonlinear Detectability and Incremental Input/Output-to-State Stability. *SIAM Journal on Control and Optimization*, 59(4):3017–3039, 2021.
[45] E. D. Sontag and Y. Wang. On Characterizations of the Input-to-state Stability Property. *Systems & Control Letters*, 24(5):351–359, 1995.
[46] E. M. Stein and R. Shakarchi. *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. Princeton lectures in analysis. Princeton Univ. Press, Princeton, NJ, 2005.
[47] P. Bernard. Luenberger Observers for Nonlinear Controlled Systems. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 3676–3681, 2017.
[48] P. Bernard. Luenberger Observers for Nonlinear Controlled Systems. *IEEE Conference on Decision and Control*, 2017.
[49] P. Bernard and A. K. Jebai. Robust Sensorless Estimation of the Position and Magnet Flux of PMSMs. In *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, pages 1082–1087, 2020.

Gia Quoc Bao Tran has been working at the Centre Automatique et Systèmes (CAS) of Mines Paris towards a Ph.D. in Mathematics and Control from Université PSL since 2021. His Ph.D. work "Observer Design for Hybrid Dynamical Systems" is supervised by Prof. Pauline Bernard and Prof. Florent Di Meglio at CAS and Prof. Ricardo Sanfelice at UC Santa Cruz, USA. Bao obtained his Engineering Degree in Automatic Control from Grenoble INP, Université Grenoble Alpes, France in 2021, after his Bachelor in Mechatronics at HCMUT, Vietnam in 2019. He was a research intern at GIPSA-lab, France and UTokyo, Japan in 2020 and 2021.

Pauline Bernard graduated in Applied Mathematics from Mines Paris in 2014 (formerly MINES ParisTech). She joined the Centre Automatique et Systèmes of Mines Paris and obtained her Ph.D. in Mathematics and Control from Université PSL in 2017. For her work on observer design for nonlinear systems, she obtained the European Ph.D. award on Control for Complex and Heterogeneous Systems 2018. As a post-doctoral scholar, she visited the Hybrid Systems Lab at the University California Santa Cruz, USA, and the Center for Research on Complex Automated Systems at the University of Bologna, Italy. Since 2019, she has been an associate professor at the Centre Automatique et Systèmes of Mines Paris, Université PSL, France. Her research interests cover the observation and output regulation of nonlinear and hybrid systems.

Class of systems and references	System dynamics	Observer in the z-coordinates	Equation to solve for the transformation	Closed form of the transformation	Observability assumption
Linear time-invariant continuous [7]	$\dot{x} = Fx$ $y = Hx$	$\dot{z} = Az + By$ A Hurwitz, $z \in \mathbb{R}^{n_x}$	$TF = AT + BH$	$T = \int_0^{+\infty} e^{At} B H e^{-Ft} dt$ (when defined)	Observability of (F, H)
Linear time-invariant discrete (inferred from [7])	$x_{k+1} = Fx_k$ $y_k = Hx_k$	$z_{k+1} = Az_k + By_k$ A Schur, $z \in \mathbb{R}^{n_x}$	$TF = AT + BH$	$T = \sum_{i=0}^{+\infty} A^i B H F^{-(i+1)}$ (when defined)	Observability of (F, H)
Nonlinear time-invariant continuous [2]	$\dot{x} = f(x)$ $y = h(x)$	$\dot{z} = Az + By$ A Hurwitz, $z \in \mathbb{R}^{n_z}$, $n_z \geq n_x$	$\frac{\partial T}{\partial x}(x) f(x) = AT(x) + Bh(x)$	$T(x) = \int_{-\infty}^{0} e^{-As} Bh(X(x,s)) ds$	Backward distinguishability
Nonlinear time-invariant discrete [5]	$x_{k+1} = f(x_k)$ $y_k = h(x_k)$	$z_{k+1} = Az_k + By_k$ A Schur, $z \in \mathbb{R}^{n_z}$, $n_z \geq n_x$	$T(f(x)) = AT(x) + Bh(x)$	$T(x) =$ $\sum_{i=0}^{+\infty} A^i Bh((\underbrace{f^{-1} \circ f^{-1} \circ \ldots \circ f^{-1}}_{i+1 \text{ time(s)}})(x))$	Backward distinguishability
Nonlinear non-autonomous continuous [4]	$\dot{x} = f(x,u)$ $y = h(x,u)$	$\dot{z} = Az + By$ A Hurwitz, $z \in \mathbb{R}^{n_z}$, $n_z \geq n_x$	$\frac{\partial T}{\partial x}(x,t) f(x,u(t)) + \frac{\partial T}{\partial t}(x,t)$ $= AT(x,t) + Bh(x,u(t))$	$T(x,t) = \int_0^t e^{A(t-s)} Bh(X(x,t;s;u)) ds$	Uniform Lipschitz differential observability or Backward distinguishability
Nonlinear non-autonomous discrete (this work)	$x_{k+1} = f_k(x_k)$ $y_k = h_k(x_k)$	$z_{k+1} = Az_k + By_k$ A Schur, $z \in \mathbb{R}^{n_z}$, $n_z \geq n_x$	$T_{k+1}(f_k(x))$ $= AT_k(x) + Bh_k(x)$	$T_k(x) = A^k(T_0 \circ f_0^{-1} \circ f_1^{-1} \circ \ldots \circ f_{k-1}^{-1})(x)$ $+ \sum_{j=0}^{k-1} A^{k-j-1} B(h_j \circ f_j^{-1} \circ f_{j+1}^{-1} \circ \ldots \circ f_{k-1}^{-1})(x)$	Uniform Lipschitz backward distinguishability or Backward distinguishability

TABLE I

SUMMARY OF KKL OBSERVER RESULTS.