KKL set-valued observers for non-observable systems

Valentin Alleaume * **Pauline Bernard** *

* *Centre Automatique et Systèmes, Mines Paris, Université PSL, 60 bd Saint-Michel, 75272, Paris, France, (emails : valentin.alleaume@minesparis.psl.eu, pauline.bernard@minesparis.psl.eu)*

Abstract: KKL observer design consists in transforming the system dynamics into a filter of the output, which admits a trivial observer, and left-inverting the transformation to recover an estimate of the state in the system coordinates. This left-inversion is typically guaranteed under a backward-distinguishability condition. In this paper, instead, we demonstrate how this KKL approach may also be applied without any such observability assumption. We show that there exist appropriate choices of the filter such that any filter solution asymptotically contains the full information about the state *indistinguishable class*, namely the set of points generating the same output. Then, we investigate the existence of a set-valued left-inverse allowing to estimate asymptotically this indistinguishable class, in the Hausdorff sense. We prove that the estimate tends to be included asymptotically in the indistinguishable classes of the limit points of the system solution. Finally, we provide a numerical example illustrating this convergence.

Keywords: KKL observer design, indistinguishability

1. INTRODUCTION

We consider an autonomous dynamical system

$$\dot{x} = f(x) \quad , \quad y = h(x) \tag{1}$$

with state $x \in \mathbb{R}^{n_x}$, output $y \in \mathbb{R}^{n_y}$, $f : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ locally Lipschitz and $h : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ continuous. We assume that the trajectories of interest remain in a compact subset $\mathcal{X} \subset \mathbb{R}^{n_x}$.

A standard question is the so-called observation problem: at each time $t \geq 0$, knowing the past history of the output y on $[0, t]$, can we produce an estimate $\hat{x}(t)$ of the state $x(t)$ such that the error $\hat{x}(t) - x(t)$ asymptotically goes to zero? Among many possible methods, a possible path is the design of an *observer*, namely a dynamical system

$$\dot{z} = F(z, y) \quad , \quad \hat{x} = \mathcal{T}(z, y) \tag{2}$$

fed with the known signal y that asymptotically estimates the real system state. Such designs are reviewed in Bernard et al. (2022). However, for an observer to exist, *detectability* or *observability* properties are typically required to ensure that the output carries enough information to determine the system state uniquely. In this paper, instead, we are interested in building an algorithm extracting from the output y all the possible information about the state x, which may not be uniquely determined without observability properties.

1.1 KKL observer design with distinguishability

When the system is observable, one of the possible routes towards an observer is the *nonlinear Luenberger* or *Kazantzis-Kravaris-Luenberger (KKL)* design. Its idea originates from Luenberger (1964) where he first introduced his observer for single-output linear systems

$$\dot{x} = Fx \quad , \quad y = Hx$$

with observable pair $(F, H) \in \mathbb{R}^{n_x \times n_x} \times \mathbb{R}^{1 \times n_x}$. Indeed, he showed that, for any controllable pair $(A, B) \in \mathbb{R}^{n_x \times n_x} \times \mathbb{R}^{n_x}$ with A Hurwitz and having no common eigenvalue with F, there exists a linear invertible change of coordinates $T \in \mathbb{R}^{n_x \times n_x}$ such that $z = Tx$ follows the dynamics

$$\dot{z} = Az + By \ . \tag{3}$$

Then, since A is Hurwitz, implementing (3) for any initial condition asymptotically provides an estimate of Tx; thus, an estimate of x can be recovered through $\hat{x} = T^{-1}z$.

It turns out that this procedure extends to nonlinear systems. Indeed, the idea is to look for a (nonlinear) change of coordinates $T : \mathbb{R}^{n_x} \to \mathbb{R}^{n_z}$ with $n_z \geq n_x$ such that $z = T(x)$ is governed by (3) in the new coordinates. In other words, T is chosen such that

$$\frac{\partial T}{\partial x}(x) f(x) = AT(x) + Bh(x) \quad \forall x \in \mathcal{X} \ . \tag{4}$$

Then, if the map T is injective and admits a (uniformly) continuous left-inverse $T^{\text{inv}} : \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$, an estimate of x is obtained with $\hat{x} = T^{\text{inv}}(z)$ for any solution z to (3).

The existence of the map T was first considered in Shoshitaishvili (1990), Kazantzis and Kravaris (1998) and Krener and Xiao (2001) in the analytic context and around an equilibrium point. Then, the localness was dropped following another perspective in Kreisselmeier and Engel (2003), where a global existence result was proposed based on a strong observability assumption which unfortunately did not provide an indication on the necessary dimension n_z of the pair (A, B). This problem was solved in Andrieu and Praly (2006) by proving the existence of the injective map T under a weak *backward-distinguishability* condition, for A complex diagonal of dimension $n_x + 1$, with a generic choice of $n_x + 1$ *distinct* complex eigenvalues and recently in Brivadis et al. (2022) for *almost any* real controllable pair (A, B) of dimension $n_z = 2n_x + 1$ with A diagonalizable. In this latter paper, an existence result was also provided in a different paradigm, where, instead of picking a sufficient number of distinct eigenvalues, the filter (3) is picked triangular with a single eigenvalue of sufficiently large multiplicity, but without any indication about the necessary dimension n_z.

The distinguishability property assumed in Andrieu and Praly (2006); Brivadis et al. (2022) and guaranteeing the existence of an observer basically says that any distinct states x_a, x_b in $\mathcal{X}$ can be distinguished from their respective past output. In this paper, we are interested in the state estimation problem for (1) when no such distinguishability is assumed.

1.2 KKL observer design without distinguishability

If distinct solutions to (1) produce the same output, namely, they are indistinguishable, without asymptotically tending to each other, there is no hope to design an observer producing an asymptotic estimate of x. However, we could imagine to have an algorithm producing at each time a set guaranteed to *converge* asymptotically – for a certain distance to be defined – to the indistinguishable class of x. In other words, build a *set-valued observer*.

The interest for such an algorithm comes from the application of electrical machines where nonlinear models producing *finite* numbers of indistinguishable states have been exhibited (see, e.g., Moreno et al. (2017); Verrelli et al. (2018) or Bernard and Praly (2021) in the context of induction motors and permanent magnet synchronous motors (PMSM) respectively) Note that this phenomenon could not appear in a linear context where non observability necessarily implies an infinite number of indistinguishable states.

Indeed, when the number of possible indistinguishable states is finite at each time, one could hope to design an observer estimating this finite number of possibilities. This was done in Moreno et al. (2017); Moreno and Besançon (2017) through sliding mode tools for systems that can be written in an "observable-like" form, but which, due to unobservability, involves a set valued map that is considered as an unknown input in the design. Instead, in Bernard and Praly (2021), the KKL route is employed on a particular application featuring a PMSM with unknown resistance. Indeed, it is shown that undistinguishable trajectories exist, but always less than six, and that there exists a map $T : \mathbb{R}^{n_x} \to \mathbb{R}^{n_z}$ transforming the dynamics into (3) whose inversion enables to reconstruct all the possible states.

In this paper, we theoretize this approach in a general context and investigate whether the KKL paradigm may be used for non observable systems. We show in Section 2 that there exists appropriate choices of the pair (A, B) such that any solution $t \mapsto z(t)$ to (3) asymptotically contains the full information about the *indistinguishable class* of x, namely the set of points generating the same output as x. Then, after studying continuity and convergence of set-valued maps in Section 3, we investigate in Section 4 the existence of a set-valued map $T^{\text{inv}} : \mathbb{R}^{n_z} \rightrightarrows \mathbb{R}^{n_x}$ such that $T^{\text{inv}}(z)$ converges to the indistinguishable class of x. Finally, a numerical example is provided in Section 5.

Remark 1. Another paradigm exploiting KKL design for non observable systems is the so-called *functional* observer problem in Spirito et al. (2022), where the full state is not necessarily observable but only a function of the state $q(x)$ is estimated. This does not lead to set-valued observers since $q(x)$ is assumed observable and is thus reconstructed uniquely by the observer.

1.3 Notations

We denote $X(x, t)$ the solution of (1) initialized at x at time 0 and evaluated at time t, and $(\sigma^-(x), \sigma^+(x))$ its maximal domain of definition. For $x \in \mathbb{R}^{n_x}$ and $\varepsilon > 0$, we denote $\mathcal{B}(x, \varepsilon)$ the open ball centered at x and with radius ε. The identity matrix of dimension m is denoted I_m.

2. INJECTIVITY WITH RESPECT TO INDISTINGUISHABLE CLASS

Following the KKL methodology, we propose to look for a map $T : \mathbb{R}^{n_x} \to \mathbb{R}^{n_z}$ transforming the dynamics (1) into (3), namely solution to (4). This first step is the same as in (Andrieu and Praly, 2006, Theorem 1) in the observable context and can be simplified as follows when $\mathcal{X}$ is compact.

Lemma 1. Pick $n_z \in \mathbb{N}$. There exists $\rho > 0$ such that for any Hurwitz matrix $A \in \mathbb{R}^{n_z \times n_z}$ and for any $B \in \mathbb{R}^{n_z \times n_y}$ with $A + \rho I_{n_z}$ Hurwitz, there exists a C^1 map $T : \mathbb{R}^{n_x} \to \mathbb{R}^{n_z}$ such that (4) holds.

But then, without observability/distinguishability we cannot hope to prove injectivity of T on $\mathcal{X}$, namely the fact that for all $(x_a, x_b) \in \mathcal{X} \times \mathcal{X}$,

$$T(x_a) = T(x_b) \quad \Longrightarrow \quad x_a = x_b \, . \tag{5}$$

Instead, we would like to prove that $T(x)$ contains all the possible distinguishable information about x.

Definition 1. Two states $x_a, x_b \in \mathbb{R}^{n_x}$ are said backward-indistinguishable, which we denote $x_a \underset{\mathcal{I}}{\sim} x_b$, if

$$h(X(x_a, t)) = h(X(x_b, t)) \quad \forall t \in (\max\{\sigma^-(x_a), \sigma^-(x_b)\}, 0] \, .$$

Given $\mathcal{O} \subseteq \mathbb{R}^{n_x}$, for a state $x \in \mathbb{R}^{n_x}$, we then denote

$$\mathcal{I}_{\mathcal{O}}(x) = \{x' \in \mathcal{O}, \; x' \underset{\mathcal{I}}{\sim} x\} \tag{6}$$

its backward indistinguishable class in $\mathcal{O}$.

In other words, $\mathcal{I}_{\mathcal{O}}(x)$ contains all the states in $\mathcal{O}$ that cannot be distinguished from x based on the knowledge of the past output. To ease the notation, we will omit the mention of $\mathcal{O}$ when $\mathcal{O} = \mathbb{R}^{n_x}$.

Note that for analytic systems, $t \mapsto X(x, t)$ is analytic in time, and therefore, equality of outputs during an arbitrarily short amount of time implies equality of outputs on the whole interval of definition. In other words, two solutions $t \mapsto X(x_a, t)$ and $X(x_b, t)$ are indistinguishable, i.e., have the same output, on some interval of time is equivalent to them being indistinguishable on their whole interval of definition, and equivalently, $X(x_a, t) \underset{\mathcal{I}}{\sim} X(x_b, t)$ at all times.

The following result shows the existence of a map $T : \mathbb{R}^{n_x} \to \mathbb{R}^{n_z}$ transforming the dynamics (1) into (3) such that T distinguishes the distinguishable states.

Theorem 1. Assume there exists an open bounded set $\mathcal{O}$, containing $\mathcal{X}$, that is backward invariant by f. Denote $n_0 = 2n_x + 1$ and $n_z = (2n_x + 1)n_y$. There exists $\rho > 0$ such that for almost any pair $(A_0, B_0) \in \mathbb{R}^{n_0 \times n_0} \times \mathbb{R}^{n_0}$ with $A_0 + \rho I_{n_0}$ Hurwitz, there exists $T : \mathbb{R}^{n_x} \to \mathbb{R}^{n_z}$ verifying (4) with

$$A = I_{n_y} \otimes A_0 \quad , \quad B = I_{n_y} \otimes B_0$$

and for all $(x_a, x_b) \in \mathcal{O} \times \mathcal{O}$,

$$T(x_a) = T(x_b) \quad \Longleftrightarrow \quad x_a \underset{\mathcal{I}}{\sim} x_b \ . \tag{7}$$

Remark 2. The "almost any" comes from the fact that the pair (A, B) is picked controllable, with A diagonalizable, and with almost any $p_\mathbb{C}$ eigenvalues in $\mathbb{C}$ and $p_\mathbb{R}$ eigenvalues in $\mathbb{R}$, all distinct with real part smaller than $-\rho$, such that $2p_\mathbb{C} + p_\mathbb{R} \geq 2n + 1$.

Remark 3. It is always possible to make the set $\mathcal{O}$ backward invariant by modifying f outside of $\mathcal{O}$. However, then, (7) holds for the modified system. It follows that $T(x_a) = T(x_b)$ ensures equality of the outputs of the original system only as long as $X(x_a, t)$ and $X(x_b, t)$ remain in the set where f has not been modified. Note however that for any $t \mapsto x(t)$ evolving in the compact set $\mathcal{X}$, its omega-limit set is compact and backward invariant. It means that the values taken by T at those limit points are not impacted by the modification of f.

Proof. We adapt the proof of Theorem 3.4 in Brivadis et al. (2022) by removing the "backward-distinguishability" assumption, and replacing the injectivity of T by (7). Indeed, we first note that the proof follows in the same way if Proposition D.1 holds with T_{diag} verifying (7) instead of being injective. Digging into the proof of Proposition D.1, we see that T_{diag} is of the form

$$T_{\text{diag}}(x) = (T_0(\lambda_1, x), \ldots, T_0(\lambda_{2n_x - \ell + 1}, x))$$

where ℓ is the number of independent complex eigenvalues of A_0, $(\lambda_i) \in \mathbb{C}_\rho^\ell \times \mathbb{R}_\rho^{2(n_x - \ell)+1}$ where

$$\mathbb{R}_\rho = \{\lambda \in \mathbb{R} : \lambda < -\rho\} \ , \ \mathbb{C}_\rho = \{\lambda \in \mathbb{C} : \text{Re}(\lambda) < -\rho\} \ ,$$

and by backward-invariance of $\mathcal{O}$, for $(\lambda, x) \in \mathbb{C}_\rho \times \mathcal{O}$,

$$T_0(\lambda, x) = \int_{-\infty}^0 e^{-\lambda s} h(X(x, s)) ds \ .$$

From this expression, it is clear that for all $(x_a, x_b) \in \mathcal{O} \times \mathcal{O}$, $x_a \underset{\mathcal{I}}{\sim} x_b$ implies $T_{\text{diag}}(x_a) = T_{\text{diag}}(x_b)$. Then, the converse of this implication is obtained by following Section D.2.3, but applying Lemma D.3 to

$$\Upsilon = \{(x_a, x_b) \in \mathcal{O}^2, \exists t < 0, h(X(x_a, t) \neq h(X(x_b, t))\}$$

instead of

$$\Upsilon_{\text{old}} = \{(x_a, x_b) \in \mathcal{O}^2, x_a \neq x_b\} \ .$$

Indeed, by continuity of h and of $(x, s) \mapsto X(x, s)$, Υ is open and by definition, for all $(x_a, x_b) \in \Upsilon$, there exists $s < 0$ such that $\Delta(x_a, x_b, s) \neq 0$, where

$$\Delta(x_a, x_b, s) = \exp(\rho s)(h(X(x_a, s) - h(X(x_b, s)) \ .$$

Lemma D.3 then allows to say that for almost any $(\lambda_i) \in \mathbb{C}_\rho^\ell \times \mathbb{R}_\rho^{2(n_x - \ell)+1}$, and for all $(x_a, x_b) \in \mathcal{O}^2$, $T_{\text{diag}}(x_a) = T_{\text{diag}}(x_b)$ implies $(x_a, x_b) \notin \Upsilon$, which is equivalent to $x_a \underset{\mathcal{I}}{\sim} x_b$. In other words, (7) holds for T_{diag}. ■

Thanks to (4), we know that for any initial condition, solutions to (3) converge to $T(x)$. Since T is not injective we cannot deduce a unique $\hat{x}$ converging to x by left-inversion of T. However, exploiting (7), we may hope to estimate the class of indistinguishability of x, or some elements of it, by looking for the set of pre-images of z by T. Indeed, (7) says that for all $x \in \mathcal{O}$ and for any subset $\mathcal{S}$ of $\mathcal{O}$, the pre-image set in $\mathcal{S}$, corresponds to the set of indistinguishable states in $\mathcal{S}$, i.e,

$$T^{-1}(T(x)) \cap \mathcal{S} = \mathcal{I}_\mathcal{S}(x) \ . \tag{8}$$

In particular, for the given compact set $\mathcal{X}$, this allows to consider a set-valued map $T^{\text{inv}} : \mathbb{R}^{n_z} \rightrightarrows \mathbb{R}^{n_x}$ verifying

$$T^{\text{inv}}(T(x)) = \mathcal{I}_\mathcal{X}(x) \qquad \forall x \in \mathcal{X} \ , \tag{9}$$

for instance,

$$T^{\text{inv}}(z) = \underset{x_s \in \mathcal{X}}{\operatorname{argmin}} \|T(x_s) - z\| \ . \tag{10}$$

Since z converges to $T(x)$, it is natural to study the convergence properties of $T^{\text{inv}}(z)$ towards $\mathcal{I}_\mathcal{X}(x)$.

3. CONTINUITY AND CONVERGENCE OF SET-VALUED MAPS

3.1 Definitions

In order to compare $T^{\text{inv}}(z)$ and $\mathcal{I}_\mathcal{X}(x)$, we consider the Hausdorff distance $d_\mathcal{H}$ defined as follows (see for instance Aubin and Frankowska (2009)).

Definition 2. Given two subsets $\mathcal{S}_a$ and $\mathcal{S}_b$ of $\mathbb{R}^{n_x}$, the Hausdorff distance is defined as

$$d_\mathcal{H}(\mathcal{S}_a, \mathcal{S}_b) := \max\left\{\delta(\mathcal{S}_a, \mathcal{S}_b), \delta(\mathcal{S}_b, \mathcal{S}_a)\right\}$$

where

$$\delta(\mathcal{S}_a, \mathcal{S}_b) := \sup_{x_a \in \mathcal{S}_a} d(x_a, \mathcal{S}_b) = \sup_{x_a \in \mathcal{S}_a} \inf_{x_b \in \mathcal{S}_b} d(x_a, x_b) \ .$$

Note that, for closed sets, we have

$$\delta(\mathcal{S}_a, \mathcal{S}_b) = 0 \Longleftrightarrow \mathcal{S}_a \subseteq \mathcal{S}_b \tag{11a}$$

$$d_\mathcal{H}(\mathcal{S}_a, \mathcal{S}_b) = 0 \Longleftrightarrow \mathcal{S}_a = \mathcal{S}_b \tag{11b}$$

Considering now a set-valued map $T^{\text{inv}} : \mathbb{R}^{n_z} \rightrightarrows \mathbb{R}^{n_x}$, evaluating limits of quantities of the type $d_\mathcal{H}(T^{\text{inv}}(z_a), T^{\text{inv}}(z_b))$ typically requires *(Hausdorff) continuity*, which contains two concepts : upper semicontinuity (usc) and lower semicontinuity (lsc) (see (Aubin and Cellina, 1984, Chapter 1)).

Definition 3. A set-valued function T^{inv} is said to be upper semicontinuous at $z^\star$ if for any open neighbourhood V containing $T^{\text{inv}}(z^\star)$ there exists a neighbourhood W of $z^\star$ such that for all $z \in W$, $T^{\text{inv}}(z) \subseteq V$. This is equivalent to

$$\lim_{z \to z^\star} \delta(T^{\text{inv}}(z), T^{\text{inv}}(z^\star)) = 0 \ . \tag{12}$$

Definition 4. A set-valued function is said to be lower semicontinuous at $z^\star$ if for any $x \in T^{\text{inv}}(z^\star)$ and any neighborhood V of x, there exists a neighborhood W of $z^\star$ such that for all $z \in W$, $V \cap T^{\text{inv}}(z) \neq \emptyset$. This is equivalent to

$$\lim_{z \to z^\star} \delta(T^{\text{inv}}(z^\star), T^{\text{inv}}(z)) = 0 \ . \tag{13}$$

The following result is a direct consequence of the Maximum Theorem (see (Aubin and Cellina, 1984, Chap 1, Sec 2, Theorem 6)).

Lemma 2. The map T^{inv} defined in (10) is upper semicontinuous.

However, T^{inv} is a priori not lower-semicontinuous, so that Hausdorff continuity and Hausdorff convergence of $T^{\text{inv}}(z)$ to $\mathcal{I}_{\mathcal{X}}(x)$ is not guaranteed. However, exploiting upper-semicontinuity, i.e., (12), and (11a), one could hope to have asymptotically at least some inclusion of $T^{\text{inv}}(z)$ into $\mathcal{I}_{\mathcal{X}}(x)$, or the reverse. The following counter-examples show that this is not even guaranteed when both z and x move with time.

3.2 Counter-examples

Consider the map

$$T(x) = -x^5 + 2x^3 - x \tag{14}$$

on the domain $\mathcal{X} = [-1.5, 1.5]$. Its plot is given on Figure 1. Imagine that the system solution $t \mapsto x(t)$ tends to $x^* = 1$ asymptotically. Then, any solution $t \mapsto z(t)$ converges to $T(x^*) = 0$ marked by the orange dashed line on Figure 1. We have $\mathcal{I}_{\mathcal{X}}(x^*) = T^{\text{inv}}(T(x^*)) = \{-1, 0, 1\}$ marked by the red dots. Now let us compare $T^{\text{inv}}(z(t))$ and $\mathcal{I}_{\mathcal{X}}(x(t)) = T^{\text{inv}}(T(x(t)))$. On the one hand, $T^{\text{inv}}(T(x(t))$ equals after a certain time a set with three elements marked by the three green dots below the orange dashed line, which asymptotically converge to the set $\{0, 1\}$. On the other hand, if z converges to zero from above, $T^{\text{inv}}(z(t))$ equals after a certain time a set with three elements marked by the three green dots above the orange dashed line, which asymptotically converge to the set $\{-1, 0\}$. It follows that $T^{\text{inv}}(z(t))$ tends to be included asymptotically in $\mathcal{I}_{\mathcal{X}}(x^*)$, i.e.,

$$\lim_{t\to\infty} \delta(T^{\text{inv}}(z(t)), \mathcal{I}_{\mathcal{X}}(x^*)) = 0$$

which illustrates the upper semicontinuity of T^{inv}. This asymptotic inclusion in the limit indistinguishable set of x^* is formalized and extended in Theorem 2 below. However, we do not have the Hausdorff convergence, i.e.,

$$\lim_{t\to\infty} d_{\mathcal{H}}(T^{\text{inv}}(z(t)), \mathcal{I}_{\mathcal{X}}(x^*)) \neq 0$$

since $1 \in \mathcal{I}_{\mathcal{X}}(x^*)$ will never be "visible" in $T^{\text{inv}}(z(t))$.

But now if we compare $T^{\text{inv}}(z(t))$ and $\mathcal{I}_{\mathcal{X}}(x(t))$ dynamically throughout time, we neither have

$$\lim_{t\to\infty} \delta(T^{\text{inv}}(z(t)), \mathcal{I}_{\mathcal{X}}(x(t)) = 0$$

nor

$$\lim_{t\to\infty} \delta(\mathcal{I}_{\mathcal{X}}(x(t)), T^{\text{inv}}(z(t))) = 0$$

and therefore even less

$$\lim_{t\to\infty} d_{\mathcal{H}}(T^{\text{inv}}(z(t)), \mathcal{I}_{\mathcal{X}}(x(t)) = 0 \ .$$

In other words, we can state a convergence result only with respect to the limit indistinguishable set $\mathcal{I}_{\mathcal{X}}(x^*)$.

Note that this example seems to suggest at least some asymptotic "intersection" between $T^{\text{inv}}(z(t))$ and $\mathcal{I}_{\mathcal{X}}(x(t))$, namely the fact that

$$\lim_{t\to\infty} \inf_{x' \in T^{\text{inv}}(z(t))} d(x', \mathcal{I}_{\mathcal{X}}(x(t))) = 0 \ ,$$

but this is not even ensured as illustrated on Figure 2 for a particular map T exhibiting a plateau. Indeed, if $t \mapsto x(t)$ tends to -0.5 from the left, $t \mapsto T(x(t))$ tends to zero from below so that $t \mapsto \mathcal{I}_{\mathcal{X}}(x(t))$ tends to $\{-0.5\}$, while if $t \mapsto z(t)$ tends to zero from above $T^{\text{inv}}(z(t))$ tends to $\{0.5\}$.

Fig. 1. Counter-example to asymptotic inclusion between $T^{\text{inv}}(z(t))$ and $\mathcal{I}(x(t)) = T^{\text{inv}}(T(x(t)))$

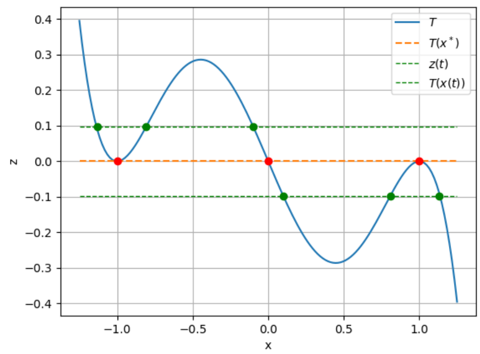
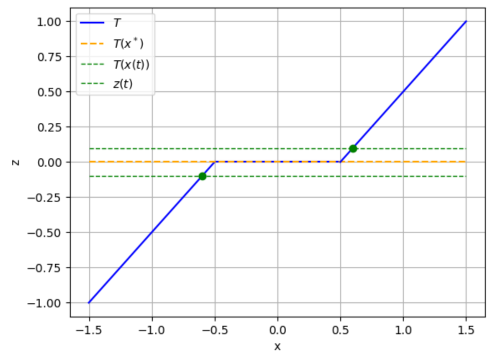


Fig. 2. Counter-example to asymptotic intersection between $T^{\text{inv}}(z(t))$ and $\mathcal{I}(x(t)) = T^{\text{inv}}(T(x(t)))$



4. CONVERGENCE OF KKL OBSERVER

As we have seen, the fact that T^{inv} may not be Hausdorff continuous heavily restricts the kind of convergence result we can state in the KKL context. However, we can still state an interesting asymptotic property if we no longer consider the distance of $T^{\text{inv}}(z(t))$ to the current indistinguishable set $\mathcal{I}_{\mathcal{X}}(x(t))$, but rather to the indistinguishable set of a limit point of x, as suggested by the examples above.

Theorem 2. Assume there exist a C^1 map $T : \mathbb{R}^{n_x} \to \mathbb{R}^{n_z}$ verifying (4) and an upper semicontinuous map $T^{\text{inv}} : \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ verifying (9). Consider a trajectory $t \mapsto x(t)$ of (1) in the compact set $\mathcal{X}$ and $x^\star$ a limit-point of $t \mapsto x(t)$. Then, for every time sequence $(t_k)_{k\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} t_k = +\infty \quad , \quad \lim_{k\to\infty} x(t_k) = x^\star$$

and any solution $t \mapsto z(t)$ of (3), we have

$$\lim_{k\to\infty} \delta(T^{\text{inv}}(z(t_k)), \mathcal{I}_{\mathcal{X}}(x^\star)) = 0 \tag{15}$$

In other words, $T^{\text{inv}}(z(t))$ tends to be included asymptotically in the indistinguishable sets of the limit points of x. In particular, if the system solution $t \mapsto x(t)$ converges to a periodic limit-cycle $t \mapsto x^\star(t)$ with period τ, then we have for any solution $t \mapsto z(t)$ of (4) and for any $t^\star \in [0, \tau]$,

$$\lim_{n\to\infty} \delta(T^{\text{inv}}(z(t^\star + n\tau)), \mathcal{I}_{\mathcal{X}}(x^\star(t^\star))) = 0 \ .$$

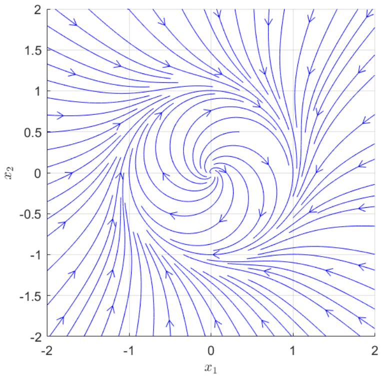


Fig. 3. Phase portrait of $\dot{x} = f(x)$ with f defined in (16).

Proof. First, $x^\star$ exists and is in $\mathcal{X}$ since $\mathcal{X}$ is compact. Then, according to (4), $t \mapsto T(x(t))$ is solution to (3). Since A is Hurwitz, $z(t) - T(x(t))$ tends to zero asymptotically, and therefore, by continuity of T and definition of (t_k), $\lim_{k\to\infty} z(t_k) = T(x^\star)$. Let $\alpha = T(x^\star)$, $\varepsilon > 0$ and the open set

$$V_\varepsilon = \bigcup_{x \in \mathcal{I}_\mathcal{X}(x^\star)} \mathcal{B}(x, \varepsilon)$$

By upper semicontinuity of T^{inv}, since $T^{\text{inv}}(\alpha) = \mathcal{I}_\mathcal{X}(x^\star)$ according to (9), there exists a neighborhood W of α such that for all $z \in W$, $T^{\text{inv}}(z) \subseteq V_\varepsilon$. Since $(z(t_k))$ converges to α, we know that there exists $k_\varepsilon \in \mathbb{N}$ such that for $k > k_\varepsilon$, $z(t_k) \in W$. Therefore, for any $k > k_\varepsilon$, $T^{\text{inv}}(z(t_k)) \subseteq V_\varepsilon$ and, by definition of V_ε, $\delta(T^{\text{inv}}(z(t_k)), T^{\text{inv}}(\alpha)) < \varepsilon$. The result follows by making ε go to zero. ■

5. NUMERICAL EXAMPLE

Consider

$$f(x) = \begin{pmatrix} x_2 + x_1(1 - (x_1^2 + x_2^2)) \\ -x_1 + x_2(1 - (x_1^2 + x_2^2)) \end{pmatrix}$$
$$h(x) = \begin{pmatrix} x_1^2 - x_2^2 \\ 2x_1x_2 \end{pmatrix} \tag{16}$$

and a set of interest $\mathcal{X} = \overline{\mathcal{B}(0, 1.7)}$, where the solutions to be estimated evolve. The phase portrait of the corresponding dynamics are given in Figure 3. It is easily seen that if $t \mapsto x(t)$ is a solution of $\dot{x} = f(x)$, then $-x$ is too. Since $h(x) = h(-x)$, x and $-x$ are thus indistinguishable and we actually have for any $x \in \mathbb{R}^2$, $\mathcal{I}_\mathcal{X}(x) = \{x, -x\}$.

In order to fall into the scope of Theorem 1, we need to make an open set $\mathcal{O}$, containing $\mathcal{X}$, backward invariant for the dynamics, while preserving the dynamics on $\mathcal{X}$ and the indistinguishable sets. This is done by considering $\check{f}(x) = \chi(\|x\|)f(x)$, where $\chi : \mathbb{R} \to \mathbb{R}$ is a C^1 map, equal to 1 when $\|x\| \le 2$ and 0 when $\|x\| \ge 10$.

Inspired by Theorem 1, we should pick a pair (A, B) of dimension $(2n+1)n_y = 10$. However, to reduce dimension, following Remark 2, we may try to pick $p_\mathbb{C} = n + 1$ eigenvalues in $\mathbb{R}$ (included in $\mathbb{C}$) leading to a real pair (A, B) of dimension $(n+1)n_y = 6$, aware that this choice might be in the zero-measure set that does not ensure (7). For the simulations, we choose $A = -\operatorname{diag}(3, 4, 5, 6, 7, 8)$ and $B = I_2 \otimes \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, which is not exactly the form recommended by Theorem 1 and shows the flexibility of the result.

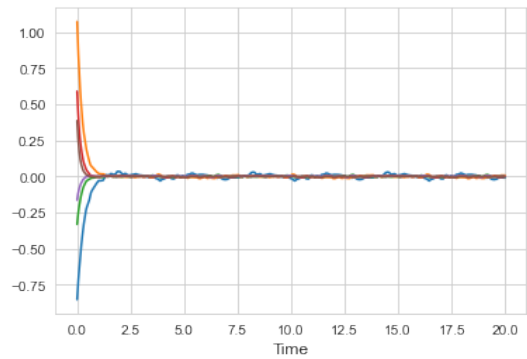
5.1 Numerical computation of T

In order to compute numerically a map T satisfying (4), we use the algorithm developed in Ramos et al. (2020) and the associated toolbox Buisson-Fenet et al. (2022):

(1) Create a grid of x_0 points covering $\mathcal{X}$.
(2) Simulate $\dot{x} = \check{f}(x)$ backward initialized with x_0 during a time t_b sufficiently large compared to the slowest eigenvalue of A. This gives a new grid of points $x_{0,b}$. Note here the interest of using $\check{f}$ instead of f, which explodes in finite backward time.
(3) Simulate forward during t_b units of time the augmented system (x, z) made of (1)-(3) and initialized at $(x_{0,b}, 0)$.
(4) We obtain a grid of regression points (x_0, z_0), where z_0 is an approximation of $T(x_0)$ if t_b is sufficiently large for (3) to "forget" its initial condition.
(5) Fit a neural network model of T.

Applying this method, we obtain an approximation of a map $T : \mathbb{R}^2 \to \mathbb{R}^6$ such that for any solution $t \mapsto x(t)$ to (1) remaining in $\mathcal{X}$, any solution $t \mapsto z(t)$ of (3) converges to $T(x)$ as illustrated on Figure 4. We use a data set with 10^5 points uniformly chosen in $\mathcal{X}$ and a neural network with 5 hidden layers of 50 neurons, ReLU activation function, learning rate 0.001, weight decay 10^{-6}, and scheduler with factor 0.1, patience 3 and threshold 0.0001. Note that, as expected from Theorem 1, the mean relative error between $T(x)$ and $T(-x)$ is low (under 3%).

Fig. 4. Convergence of the error $z - T(x)$



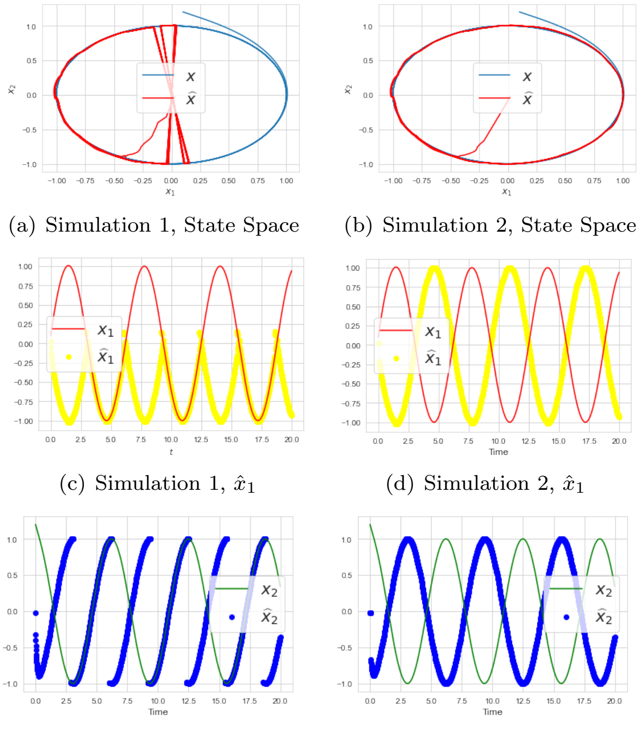
5.2 Inversion of T

The next step is to compute numerically the pre-image of T in order to reconstruct online at least one element of the indistinguishable class of x.

In Buisson-Fenet et al. (2022), written in the spirit of observer design where T is injective, this is done by learning a neural model of a left-inverse T^{inv} of T such that $x_0 \approx T^{\text{inv}}(z_0)$ on the previously obtained grid. However, due to the non-injectivity of T, we cannot exploit such a method here. Indeed, the minimization cost will be flawed by the presence of distinct targets for the same input. For instance, if a given z_0 should be mapped both to x_0 and $-x_0$, then, the algorithm typically finds

$$T^{\text{inv}}(z_0) = \operatorname*{argmin}_{\hat{x}} \{\|\hat{x} - x_0\|^2 + \|\hat{x} - (-x_0)\|^2\} = 0$$

Fig. 5. Simulation results in the x-coordinates via gradient descent algorithm.



(a) Simulation 1, State Space (b) Simulation 2, State Space

(c) Simulation 1, $\hat{x}_1$ (d) Simulation 2, $\hat{x}_1$

(e) Simulation 1, $\hat{x}_2$ (f) Simulation 2, $\hat{x}_2$

everywhere. An alternative would be to learn T^{inv} on a reduced data set where only one pre-image of each z_0 is kept. But this raises regularity and threshold issues. Instead, we implement a gradient descent algorithm online, solving (10) at each time step. Results of simulations appear on Figure 5.

As expected, once z has converged to $T(x)$, the algorithm gives as estimate either x or $-x$, with possible jumps depending on the initialization of the optimization (see Simulation 1). A "warm start" of the optimization around the previously found estimate allows to follow a continuous selection in $T^{\text{inv}}(z(t))$ and avoid those jumps (see Simulation 2). This does not ensure to follow the trajectory x, but a potential indistinguishable trajectory, namely $-x$ here. Note that more efficient algorithms could be implemented to follow the optimum through time, or better yet, reconstruct both candidates x and $-x$.

6. CONCLUSION

Even without any observability assumption, the KKL approach is well-suited to estimate information about the state, up to its indistinguishable class. We studied which kind of convergence could be expected from the estimator. The advantage of such an approach is that no particular normal form is needed unlike high-gain or sliding-mode based methods in Moreno et al. (2017); Moreno and Besançon (2017). Further work includes improving the obtained convergence theorem and finding better sufficient conditions guaranteeing Hausdorff continuity. In particular, the apparent link between critical points of T and lack of lower semicontinuity of T^{inv} could be studied.

REFERENCES

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

Aubin, J.P. and Cellina, A. (1984). *Differential Inclusions: Set-Valued Maps and Viability Theory*, volume 264 of *Grundlehren Der Mathematischen Wissenschaften*. Springer Berlin Heidelberg.

Aubin, J.P. and Frankowska, H. (2009). *Set-Valued Analysis*. Birkhäuser Boston.

Bernard, P. and Praly, L. (2021). 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.

Bernard, P., Andrieu, V., and Astolfi, D. (2022). Observer design for continuous-time dynamical systems. *Annual Reviews in Control*, S1367578821000948.

Brivadis, L., Andrieu, V., Bernard, P., and Serres, U. (2022). Further remarks on KKL. *Submitted to Systems and Control Letters, Available online at* `https://hal.archives-ouvertes.fr/hal-03695863`.

Buisson-Fenet, M., Bahr, L., and Meglio, F.D. (2022). Learning to observe : neural network-based KKL observers. Python toolbox available at `https://github.com/Centre-automatique-et-systemes/learn_observe_KKL.git`.

Kazantzis, N. and Kravaris, C. (1998). Nonlinear observer design using Lyapunov's auxiliary theorem. *Systems & Control Letters*, 34(5), 241–247.

Kreisselmeier, G. and Engel, R. (2003). Nonlinear observers for autonomous Lipschitz continuous systems. *IEEE Transactions on Automatic Control*, 48(3).

Krener, A.J. and Xiao, M. (2001). Nonlinear observer design in the siegel domain through coordinate changes. *IFAC Proceedings Volumes*, 34(6), 519–524.

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

Moreno, J.A., Mujica-Ortega, H., and Espinosa-Pérez, G. (2017). A global bivalued-observer for the sensorless induction motor. *IFAC World Congress*.

Moreno, J. and Besançon, G. (2017). Multivalued finite-time observers for a class of nonlinear systems. *IEEE Conference on Decision and Control*.

Ramos, L.d.C., Di Meglio, F., Morgenthaler, V., da Silva, L.F.F., and Bernard, P. (2020). Numerical design of Luenberger observers for nonlinear systems. In *2020 59th IEEE Conference on Decision and Control (CDC)*, 5435–5442. IEEE.

Shoshitaishvili, A. (1990). Singularities for projections of integral manifolds with applications to control and observation problems. *Theory of singularities and its applications*, 1, 295.

Spirito, M., Bernard, P., and Marconi, L. (2022). On the existence of KKL functional observers. *American Control Conference*.

Verrelli, C., Lorenzani, E., Fornari, R., Mengoni, M., and Zarri, L. (2018). Steady-state speed sensor fault detection in induction motors with uncertain parameters: A matter of algebraic equations. *Control Engineering Practice*, 80, 125–137.