On the existence of KKL observers with nonlinear contracting dynamics (Long Version)

Victor Pachy, Vincent Andrieu, Pauline Bernard, Lucas Brivadis, Laurent Praly

On the existence of KKL observers with nonlinear contracting dynamics (long version)

Victor Pachy * **Vincent Andrieu** **,[1] **Pauline Bernard** ***
Lucas Brivadis **** **Laurent Praly** ***

* *Mines Paris, Université PSL, 60 boulevard Saint-Michel, Paris, France (e-mail: victor.pachy@etu.minesparis.psl.eu).*
** *Univ. Lyon, Université Claude Bernard Lyon 1, CNRS, LAGEPP UMR 5007, 43 bd du 11 novembre 1918, F-69100 Villeurbanne, France (e-mail: vincent.andrieu@gmail.com)*
*** *Centre Automatique et Systèmes, Mines Paris, Université PSL, 60 boulevard Saint-Michel, Paris, France (e-mail: (firstname).(lastname)@minesparis.psl.eu)*
**** *Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, 91190, Gif-sur-Yvette, France (e-mail: lucas.brivadis@centralesupelec.fr)*

Abstract: KKL (Kazantzis-Kravaris/Luenberger) observers are based on the idea of immersing a given nonlinear system into a target system that is a *linear* stable filter of the measured output. In the present paper, we extend this theory by allowing this target system to be a *nonlinear* contracting filter of the output. We prove, under a differential observability condition, the existence of these new KKL observers. We motivate their introduction by showing numerically the possibility of combining convergence speed and robustness to noise, unlike what is known for linear filtering.

1. INTRODUCTION.

We consider nonlinear systems of the form

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

where x lying in $\mathbb{R}^n$ is the state of the system, y lying in $\mathbb{R}$ is the measured output, and f and h are smooth maps.

The synthesis of state observers for such systems is a major topic in control theory, and numerous methods have been developed over the years. Interested readers can refer to Bernard et al. (2022) for an overview of existing methods. Among these methods, the so-called KKL approach has recently garnered significant attention due to its generality and the weak assumptions required ensuring its existence. This methodology, originating from the seminal work of Luenberger (1964) for linear systems, has been extended for nonlinear systems first locally in Kazantzis and Kravaris (1998); Shoshitaishvili (1990) and then globally in Kreisselmeier and Engel (2003); Andrieu and Praly (2006); Brivadis et al. (2023). The strategy consists in finding a positive integer m and a nonlinear mapping $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^m$ such that if (x, y) is a solution of (1), then $\boldsymbol{z} = \boldsymbol{T}(x)$ satisfies

$$\dot{\boldsymbol{z}} = A\boldsymbol{z} + By, \tag{2}$$

where $A \in \mathbb{R}^{n\times m}$ is a Hurwitz matrix and $B \in \mathbb{R}^m$ is a vector. In that case, $\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{z} - \boldsymbol{T}(x)) = A(\boldsymbol{z} - \boldsymbol{T}(x))$, hence an asymptotic exponential approximation of $\boldsymbol{T}(x)$ can be obtained by running the linear filter of the output (2). Then, if $\boldsymbol{T}$ admits a uniformly continuous left-inverse $\boldsymbol{T}^{\mathrm{inv}} : \mathbb{R}^m \to \mathbb{R}^n$, an observer of x can be defined as $\hat{x} = \boldsymbol{T}^{\mathrm{inv}}(z)$ and on has $\lim_{t\to+\infty}(\hat{x}(t) - x(t)) = 0$.

Different kind of sufficient conditions for the existence of such mappings $\boldsymbol{T}$ and $\boldsymbol{T}^{\mathrm{inv}}$ are given in Andrieu and Praly (2006).

In this paper, we propose to extend the class of admissible filters of the output for the design of the observer. More precisely, we replace (2) by

$$\dot{\boldsymbol{z}} = k\boldsymbol{\sigma}(\boldsymbol{z}, y) \tag{3}$$

where k is a positive real number and $\boldsymbol{\sigma}$ ensures that (3) has *exponentially contracting* dynamics (in a sense to be defined), ensuring in particular that the distance between any pair of solutions sharing the same y exponentially decreases towards 0. This is a natural extension, since, similar to the linear case, if one finds a mapping T such that $\boldsymbol{z} = \boldsymbol{T}(x)$ is solution to (3), an asymptotic exponential estimation of $\boldsymbol{T}(x)$ can be obtained, by running (3) from any initial condition. Then, as with the linear filter (2), if $\boldsymbol{T}$ admits a left-inverse $\boldsymbol{T}^{\mathrm{inv}} : \mathbb{R}^m \to \mathbb{R}^n$, an observer can be defined as

$$\hat{x} = \boldsymbol{T}^{\mathrm{inv}}(\boldsymbol{z}) \tag{4}$$

and one has $\lim_{t\to+\infty}(\hat{x}(t) - x(t)) = 0$ provided that $\boldsymbol{T}^{\mathrm{inv}}$ is uniformly continuous.

This study is motivated by three main reasons: nonlinear filters may (i) give access to better observer performance,

[1] The research of this author was funded in whole or in part by the National Research Agency (ANR) under the project titled "ANR-23-CE48-0006-02."

for instance allowing to combine convergence speed and robustness to noise or increase robustness to model uncertainties; (ii) bridge the gap between the fields of observer design and recursive neural networks in machine learning; (iii) give more flexibility in the research of an analytical expression of the map $\boldsymbol{T}$.

Regarding item (i), while it is well-known that in the context of linear dynamics with Gaussian noise, the use of linear correction gains, as provided by a Kalman filter, is optimal (Kalman and Bucy (1961)), the same cannot be said when departing from this context. For instance, it is typically interesting to introduce nonlinear phenomena such as saturations in the case of sporadic noise, or, dead zones in the case of high-frequency and low-amplitude noise Tarbouriech et al. (2022). Furthermore, the interest in non-linear observers for achieving robustness to model errors is typically highlighted in the context of homogeneous observers or sliding mode observers (see for instance Levant (2003)). Moreover, convergence speed and robustness to noise are usually antinomic, unless varying/switching gains are introduced or even switches among different observers, depending on the size of the output error, but with thresholds that may be hard to tune Petri et al. (2023); Chong et al. (2015); Esfandiari and Shakarami (2019). Our goal in considering (3) is to allow such performance, without requesting any canonical form of the system (1), or having any online threshold to handle.

This also brings us to item (ii) since it is well-known in the machine learning community that the addition of nonlinear terms can enhance the expressiveness of neural networks and improve performance. Actually the resemblance between KKL observers – where the system output is fed to a series of filters and the estimate is recovered from those internal states via a nonlinear map to be found – and recursive neural networks (RNNs) is uncanny (see Janny et al. (2021) for a comparison). In fact, when no explicit expression of the map $\boldsymbol{T}$ is available, it has been proposed to learn an approximate model of it via neural networks in Ramos et al. (2020); Niazi et al. (2023); Buisson-Fenet et al. (2023); Peralez and Nadri (2021); Janny et al. (2021). Allowing nonlinear contractions in KKL could pave the road towards understanding this similarity and providing theoretical foundations to the convergence of RNNs as well as guidelines in terms of dimensions.

However, all those methods require a significant (offline) computational load and suffer from a curse of dimensionality, sometimes at the expense of asymptotic precision on the online estimate. Our third goal (iii) is thus to enlarge the class of systems for which an *explicit* expression of the transformation $\boldsymbol{T}$ can be obtained by providing more flexibility in the target dynamics (3).

From a theoretical point of view, the *existence* of an immersion $\boldsymbol{T}$ of any system (1) into dynamics (3) is typically guaranteed, for instance exploiting the theory of uniformly convergent systems in Pavlov et al. (2004). The challenge rather lies in proving its *injectivity* under observability conditions. This paper achieves a first step by showing this injectivity for a particular structure of (3), consisting of m sufficiently fast parallel filters and under an assumption of strong differential observability of order m.

The article is structured as follows. Firstly, we present a general theorem that establishes the existence of a Lipschitz injective application $\boldsymbol{T}$ which, when applied on the state trajectory is a specific solution to the nonlinear filter equation for sufficintly large parameter k. The next section provides a demonstration, presented as a series of propositions, with proofs provided in the Appendix. In the following section, we offer a preliminary illustration of these results within a robustness context. Lastly, the conclusion is provided.

Notation. For a subset $\mathcal{O}$ of $\mathbb{R}^n$, we denote by $\mathcal{O}+\delta$ the set $\mathcal{O}+\delta = \{x \in \mathbb{R}^n | \ \exists x_0 \in \mathcal{O}, |x-x_0| \leq \delta\}$. When needed, to exhibit the dependency on initial conditions $x \in \mathbb{R}^n$, we denote, when defined, $X(x,t)$ the (unique) solution to (1a) at time t, initialized at x. By $L_f h$ we denote the Lie derivative of h alongside the vector field f, i.e., $L_f h(x) = \frac{\partial h}{\partial x}(x) f(x)$, for all $x \in \mathbb{R}^n$. Given a map $g : \mathbb{R}^n \to \mathbb{R}^p$, we define $\Delta g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^p$ by $\Delta g(x_a, x_b) = g(x_a) - g(x_b)$ for $(x_a, x_b) \in \mathbb{R}^n \times \mathbb{R}^n$, and we say that g is (k-)Lipschitz injective on $S \subset \mathbb{R}^n$, if there exists $k > 0$ such that for all $(x_a, x_b) \in S \times S$, $|\Delta g(x_a, x_b)| \geq k|x_a - x_b|$. Given a polynomial in powers of $\frac{1}{\lambda}$ in the form $p(\lambda) = \sum_{j=0}^{\ell} \frac{p_j}{\lambda^j}$ and given an integer m, the notation $[p(\lambda)]_{\leq i}^m$ (resp. $[p(\lambda)]_{=i}^m$) represents the polynomial in powers of $\frac{1}{\lambda}$ obtained by keeping only the monomials of power $\frac{1}{\lambda^j}$ with $j \in \{0, \dots, i\}$ (resp. $j = i$) in $p(\lambda)^m$.

2. MAIN RESULT

We assume (i) the system solutions of interest, initialized in some set $\mathcal{X}_0 \subset \mathbb{R}^n$, are defined on $\mathbb{R}_{\geq 0}$ and remain in a compact set $\mathcal{X} \subset \mathbb{R}^n$, and (ii) the system (1) is *strongly differentially observable* of order m on $\mathcal{X}$, as detailed next.

Assumption 1. There exists a compact set $\mathcal{X} \subset \mathbb{R}^n$ such that for all solution of (1) such that $x(0) \in \mathcal{X}_0$ then $x(t) \in \mathcal{X}$ for all $t \geq 0$.

Assumption 2. There exists an integer $m \geq 1$ such that the map $\boldsymbol{H}_m : \mathbb{R}^n \to \mathbb{R}^m$ defined by:

$$\boldsymbol{H}_m : x \mapsto \begin{pmatrix} h(x) \\ L_f h(x) \\ \vdots \\ L_f^{m-1} h(x) \end{pmatrix} \tag{5}$$

is $\underline{k}_H$-Lipschitz injective on $\mathcal{X}$ for some constant $\underline{k}_H > 0$.

The latter assumption is guaranteed as soon as $\boldsymbol{H}_m$ is injective and its jacobian is full-rank on $\mathcal{X}$, i.e., when $\boldsymbol{H}_m$ is an injective immersion on $\mathcal{X}$.

In this paper, we show that the system (1) can then be transformed through a left-invertible change of coordinates into a contracting filter (3) with $\boldsymbol{\sigma}$ consisting of m parallel contracting filters of the output in the form

$$\boldsymbol{\sigma}(z,y) = \begin{pmatrix} \lambda_0 \sigma(z_0, y) \\ \vdots \\ \lambda_{m-1} \sigma(z_{m-1}, y) \end{pmatrix} \tag{6}$$

where $\lambda_0, \dots, \lambda_{m-1}$ are positive distinct scalars, and $\sigma : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ satisfying:

Assumption 3. The map $\sigma : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is of class C^{m+1} and verifies for all $(z, y) \in \mathbb{R} \times \mathbb{R}$,

$$0 < \gamma \leq \left| \frac{\partial \sigma}{\partial y}(z, y) \right| \tag{7a}$$

$$-\beta \leq \frac{\partial \sigma}{\partial z}(z, y) \leq -\alpha < 0 \tag{7b}$$

for some constants $\alpha, \beta, \gamma > 0$.

It can be noticed that the map $\boldsymbol{\sigma}$ defined in (6) then verifies

$$\frac{\partial \boldsymbol{\sigma}}{\partial \boldsymbol{z}}(\boldsymbol{z}, y) + \frac{\partial \boldsymbol{\sigma}}{\partial \boldsymbol{z}}(\boldsymbol{z}, y)^\top < -\mu \, \mathrm{I}_m \, , \forall (\boldsymbol{z}, y) \in \mathbb{R}^m \times \mathbb{R}, \tag{8}$$

where $\mu = -\alpha \min_{0 \leq i \leq m-1} \lambda_i > 0$ which according to (Pavlov et al., 2004, Theorem 1), guarantees that the filter (3) is *uniformly convergent* in the sense of Pavlov et al. (2004). From this, we show the following result.

Theorem 1. Under Assumptions 1, 2 and 3, for all m-uplet $(\lambda_0, \ldots, \lambda_{m-1})$ of distinct positive scalars, there exists $k^* > 0$ such that for all $k > k^*$, there exists a map $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^m$ that is Lipschitz injective on $\mathcal{X}$ and such that, for all solution $t \mapsto x(t)$ of (1a) with initial condition in $\mathcal{X}_0$, $t \mapsto \boldsymbol{T}(x(t))$ is solution to (3) with $t \mapsto y(t)$ given by (1b) and $\boldsymbol{\sigma}$ defined in (6).

This result is proved in Section 3. Under the conditions of the above theorem, we deduce from the injectivity of $\boldsymbol{T}$ that there exists a continuous map $\boldsymbol{T}^{\mathrm{inv}} : \mathbb{R}^m \to \mathbb{R}^n$ that is a left-inverse of $\boldsymbol{T}$ on $\mathcal{X}$, namely

$$\boldsymbol{T}^{\mathrm{inv}}(\boldsymbol{T}(x)) = x \quad \forall x \in \mathcal{X} \, .$$

For any such left-inverse, any solution (x, z) of (1)-(3) initialized in $\mathcal{X}_0 \times \mathbb{R}^m$, with $\boldsymbol{\sigma}$ defined in (6), verifies $\lim_{t \to +\infty}(\hat{x}(t) - x(t)) = 0$, with $\hat{x}$ given by (4) (see for instance (Brivadis et al., 2023, Theorem 1.1)). Note that from the *Lipschitz*-injectity of $\boldsymbol{T}$, $\boldsymbol{T}^{\mathrm{inv}}$ can even be picked globally Lipschitz, providing *exponential* convergence of the estimation error as in Andrieu (2014).

3. PROOF

3.1 Some preliminaries

For each y in $\mathbb{R}$, the function $z \mapsto \sigma(z, y)$ is decreasing and is a bijection from $\mathbb{R}$ to $\mathbb{R}$ from Assumption 3. Hence, there exists a unique map $\psi : \mathbb{R} \to \mathbb{R}$ such that for all $y \in \mathbb{R}$,

$$\sigma(\psi(y), y) = 0. \tag{9}$$

Moreover, according to the implicit function theorem, $\psi \in C^{m+1}(\mathbb{R}, \mathbb{R})$ since $\sigma \in C^{m+1}(\mathbb{R}, \mathbb{R})$.

Also, it can be observed that, from Assumption 1, the solutions to (1a) initialized in $\mathcal{X}_0$ coincide in positive time with that of the modified dynamics

$$\dot{x} = \breve{f}(x) = \chi(x) f(x), \tag{10}$$

with χ a smooth function such that

$$\chi(x) = 1 \text{ if } x \in \mathcal{X}, \qquad \chi(x) = 0 \text{ if } x \notin \mathcal{X} + \delta_u$$

for some $\delta_u > 0$. Besides, Assumptions 1 and 2 still hold for (10) with output map (1b). In the rest of the proof, we thus assume without loss of generality that $f = \breve{f}$ and we consider that Assumption 1, whenever used, ensures (forward and backward) invariance of the compact set $\mathcal{X} + \delta_u$ as well as completeness and boundedness of all solutions.

3.2 Construction of $\boldsymbol{T}$

Proposition 1. Under Assumptions 1 and 3, for all m-uplet $(\lambda_0, \ldots, \lambda_{m-1})$ of distinct positive scalars, and for all $k > 0$ there exists $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^m$ such that, for all solution $t \mapsto x(t)$ of (1a) with initial condition in $\mathbb{R}^n$, $t \mapsto \boldsymbol{T}(x(t))$ is solution to (3) with $t \mapsto y(t)$ given by (1b) and $\boldsymbol{\sigma}$ defined in (6).

Proof. With Assumption 1, for all $x \in \mathbb{R}^n$, the function

$$s \mapsto y(s) := h(X(x, s))$$

is defined, smooth, and bounded for all positive time. According to (Pavlov et al., 2004, Theorem 1), Assumption 3 guarantees that for all x in $\mathbb{R}^n$, the system

$$\dot{z} = \lambda \sigma(z, y) \tag{11}$$

with state $z \in \mathbb{R}$ and $\lambda > 0$ admits a unique bounded solution $t \mapsto \bar{Z}(t, y, \lambda)$ defined on $\mathbb{R}$, which is uniformly globally asymptotically stable. For all $x \in \mathbb{R}^n$ and all $\lambda > 0$, we then define

$$T(x, \lambda) = \bar{Z}(0, s \mapsto h(X(x, s)), \lambda), \tag{12}$$

and for all $\lambda_0, \ldots, \lambda_{m-1} > 0$, and all $k > 0$,

$$\boldsymbol{T}(x) = \begin{pmatrix} T(x, k\lambda_0) \\ T(x, k\lambda_1) \\ \vdots \\ T(x, k\lambda_{m-1}) \end{pmatrix}. \tag{13}$$

We then show that for any $x \in \mathbb{R}^n$, $t \mapsto \boldsymbol{T}(X(x, t))$ is solution to (3). To do so, it is sufficient to show that for any $\lambda, k > 0$, $T(X(x, t), k\lambda) = \bar{Z}(t, s \mapsto h(X(x, s)), k\lambda)$ for all $t \in \mathbb{R}$. For all t in $\mathbb{R}$, we have

$$\begin{aligned} T(X(x, t), k\lambda) &= \bar{Z}(0, s \mapsto h(X(X(x, t), s)), k\lambda) \\ &= \bar{Z}(0, s \mapsto h(X(x, t + s)), k\lambda). \end{aligned}$$

It is thus enough to show that for each $t \in \mathbb{R}$,

$$\bar{Z}(0, s \mapsto h(X(x, t + s)), k\lambda) = \bar{Z}(t, s \mapsto h(X(x, s)), k\lambda). \tag{14}$$

Let $t \in \mathbb{R}$. For any bounded $\tau \mapsto y(\tau)$, the system

$$\frac{\mathrm{d}z}{\mathrm{d}\tau}(\tau) = k\lambda \sigma(z(\tau), y(\tau)), \quad \forall \tau \in \mathbb{R}, \tag{15}$$

is uniformly convergent in the sense of Pavlov et al. (2004). Therefore, it admits a unique bounded solution defined on $\mathbb{R}$ that is $\tau \mapsto \bar{Z}(\tau, s \mapsto y(s), k\lambda)$. Consider the map $z_t : \tau \mapsto \bar{Z}(\tau + t, s \mapsto h(X(x, s)), k\lambda)$. It is bounded on $\mathbb{R}$ and verifies, by definition of $\bar{Z}$,

$$\frac{\mathrm{d}z_t}{\mathrm{d}\tau}(\tau) = k\lambda \sigma(\bar{Z}(\tau + t, s \mapsto h(X(x, s)), k\lambda), h(X(x, t + \tau))).$$

So by uniqueness of the bounded solution of (15),

$$z_t(\tau) = \bar{Z}(\tau, s \mapsto h(X(x, t + s)), k\lambda).$$

Taking $\tau = 0$ yields (14) and concludes the proof. □

3.3 Construction of an approximation of $\boldsymbol{T}$ which is Lipschitz injective

To prove Theorem 1, we need to show that the mapping $\boldsymbol{T}$ obtained from Proposition 1 is Lipschitz injective on $\mathcal{X}$ for sufficiently large k. The idea of the proof is to use an approximation of the components of $\boldsymbol{T}$ (i.e. $T(\cdot, k\lambda_i)$), and more precisely a development in powers of $\frac{1}{k}$. The approximation we consider is obtained from a C^1 mapping

$\boldsymbol{\phi} : \mathbb{R}^n \to \mathbb{R}^m$, with $\boldsymbol{\phi} = (\phi_0, \dots, \phi_{m-1})$ where ϕ_ℓ are of class C^{m-i+1} and are defined recursively for x in $\mathbb{R}^n$ as

$$\phi_0(x) = \psi(h(x)), \tag{16}$$

and for $\ell \in \{1, \dots, m-1\}$:

$$\phi_\ell(x) = \kappa(h(x))\Big[L_f\phi_{\ell-1}(x) - \sum_{j=1}^{\ell} \frac{1}{j!}\frac{\partial^j \sigma}{\partial z^j}(\psi(h(x)), h(x)) \times \sum_{\substack{\ell_1+\cdots+\ell_j=\ell \\ 1\le\ell_1,\dots,\ell_j\le\ell-1}} \phi_{\ell_1}(x)\cdots\phi_{\ell_j}(x)\Big], \tag{17}$$

where

$$\kappa(y) = \left(\frac{\partial \sigma}{\partial z}(\psi(y), y)\right)^{-1}. \tag{18}$$

Notice that the definition of $\phi_\ell(x)$ for $\ell \geq 1$ involves $\phi_i(x)$ with $i \in \{1, \cdots, \ell-1\}$ only and is independent from λ. The approximation of $T(x, \lambda)$ is then defined as

$$T_a(x, \lambda) = \sum_{\ell=0}^{m-1} \frac{\phi_\ell(x)}{\lambda^\ell}. \tag{19}$$

From there, given $\boldsymbol{\lambda} = (\lambda_0, \dots, \lambda_{m-1})$ and given $k > 1$, we thus get from (13) the following approximation of $\boldsymbol{T}$

$$\boldsymbol{T_a}(x) = \mathcal{V}K^{-1}\boldsymbol{\phi}(x), \tag{20}$$

where $K = \operatorname{diag}\left(1, \dots, k^{m-1}\right)$ and $\mathcal{V}$ is the Vandermonde matrix

$$\mathcal{V} = \begin{pmatrix} 1 & \lambda_0^{-1} & \dots & \lambda_0^{-(m-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_{m-1}^{-1} & \dots & \lambda_{m-1}^{-(m-1)} \end{pmatrix}. \tag{21}$$

Clearly, injectivity of $\boldsymbol{T_a}$ may be deduced from the injectivity of $\boldsymbol{\phi}$ as soon as the λ_i's are all distinct. Actually, in the following proposition, it is shown that with Assumption 2, injectivity of $\boldsymbol{\phi}$ is ensured on $\mathcal{X}$.

Proposition 2. Under Assumption 2, $\boldsymbol{\phi}$ is Lipschitz injective in $\mathcal{X}$. In other word, there exists a positive real number $\underline{k}_{\boldsymbol{\phi}}$ such that for all (x_a, x_b) in $\mathcal{X}^2$

$$|\boldsymbol{\phi}(x_a) - \boldsymbol{\phi}(x_b)| \geq \underline{k}_{\boldsymbol{\phi}} |x_a - x_b|. \tag{22}$$

The proof of Proposition 2 can be found in Section A.5. A direct consequence is that for all m-uplet $(\lambda_0, \dots, \lambda_{m-1})$ of positive distinct scalars, all $k > 1$, all (x_a, x_b) in $\mathcal{X}^2$ the following inequality is satisfied

$$|\boldsymbol{T_a}(x_a) - \boldsymbol{T_a}(x_b)| \geq \frac{\underline{k}_{\boldsymbol{\phi}}}{k^{m-1}\|\mathcal{V}^{-1}\|}|x_a - x_b|, \tag{23}$$

since $\|K\| \leq k^{m-1}$, which establishes that $\boldsymbol{T_a}$ is Lipschitz injective in $\mathcal{X}$.

3.4 T_a is an approximation of order λ^{-m} of T

The reason for stating that T_a is an approximation of T will be shown in this section. Indeed, we demonstrate that the difference between these two functions is of the order λ^{-m}. In Proposition 1, we have shown that for any $\lambda_0, \dots, \lambda_{m-1} > 0$, any $k > 0$ and any $x \in \mathbb{R}^n$, $t \mapsto \boldsymbol{T}(X(x,t))$ is solution to (3). From the definitions of $\boldsymbol{\sigma}$ and $\boldsymbol{T}$ in (6) and (13), we deduce that T is supposed to verify, for all $x \in \mathbb{R}^n$ and all $\lambda > 0$,

$$L_f T(x, \lambda) = \lambda\sigma(T(x, \lambda), h(x)). \tag{24}$$

In order to establish that T_a approximates T, we first study the error in the partial differential equation (24) induced by this approximation.

Proposition 3. Let $\omega : \mathbb{R}^n \times \mathbb{R}_{>0} \to \mathbb{R}$ be defined as

$$\omega(x, \lambda) = \frac{L_f T_a(x, \lambda) - \lambda\sigma(T_a(x, \lambda), h(x))}{\lambda}$$

with T_a defined in (19). Under Assumptions 1 and 3 there exist two positive real numbers $\bar{\omega}$ and k_ω such that

$$\forall x \in \mathcal{X} + \delta_u, \forall \lambda > 1, \quad |\omega(x, \lambda)| \leq \frac{\bar{\omega}}{\lambda^m}, \tag{25}$$

and for all $(x_a, x_b) \in (\mathcal{X} + \delta_u)^2, \forall \lambda > 1$,

$$|\omega(x_a, \lambda) - \omega(x_b, \lambda)| \leq \frac{k_\omega}{\lambda^m}|x_a - x_b|. \tag{26}$$

The proof of Proposition 3 can be found in Section A.1. We introduce R the difference between T and its approximation T_a:

$$R(x, \lambda) = T(x, \lambda) - T_a(x, \lambda). \tag{27}$$

Employing the bounds obtained in Proposition 3, the following two propositions establish that R is bounded and is Lipschitz *with order* $\frac{1}{\lambda^m}$:

Proposition 4. Under Assumption 1 and 3,

$$\forall x \in \mathcal{X} + \delta_u, \forall \lambda > 1, \quad |R(x, \lambda)| \leq \frac{\bar{\omega}}{\alpha\lambda^m}. \tag{28}$$

Proposition 5. Under Assumption 1 and 3, there exist $\lambda^* > 1$ and a positive real number k_R such that for all (x_a, x_b) in $(\mathcal{X} + \delta_u)^2$ for all $\lambda > \lambda^*$,

$$|R(x_a, \lambda) - R(x_b, \lambda)| \leq \frac{k_R}{\lambda^m}|x_a - x_b|. \tag{29}$$

The proof of Proposition 4 and 5 can be found in Section A.2 and Section A.4.

3.5 Proof of Theorem 1

For $k > 1$ and for a given m-uplet $\boldsymbol{\lambda}$ of distinct positive real numbers, let $\boldsymbol{T}$ be given by Proposition 1. For all solution $t \mapsto x(t)$ of (1a) with initial condition in $\mathcal{X}_0$, $t \mapsto \boldsymbol{T}(x(t))$ is solution to (3) with $t \mapsto y(t)$ given by (1b) and $\boldsymbol{\sigma}$ defined in (6). With $\boldsymbol{T_a}$ given previously, we can introduce the mapping $\boldsymbol{R}$ as

$$\boldsymbol{T}(x) = \boldsymbol{T_a}(x) + \boldsymbol{R}(x), \tag{30}$$

with

$$\boldsymbol{R}(x) = (R(x, k\lambda_0), \dots, R(x, k\lambda_{m-1})).$$

With Proposition 2, equation (23) and Proposition (29), we conclude that for all k such that $k\lambda_i > \lambda^*$ for all $i = 0, \dots, m-1$,

$$|\Delta\boldsymbol{T}(x_a, x_b)| \geq \frac{1}{k^{m-1}}\left(\frac{\underline{k}_{\boldsymbol{\phi}}}{\|\mathcal{V}^{-1}\|} - \frac{1}{k}\sum_{i=0}^{m-1}\frac{k_R}{\lambda_i^m}\right)|x_a - x_b|.$$

This proves that there exists $k^* > 0$ such that, for any $k \geq k^*$, $x \mapsto \boldsymbol{T}(x)$ given in (13) is Lipschitz injective on $\mathcal{X}$ which concludes the proof of Theorem 1.

4. ILLUSTRATION

In this part, we show how a KKL observer with *nonlinear* filter dynamics may allow to obtain simultaneously fast convergence and robustness to measurement noise. For

that, we compare its performance to two KKL observers with *linear* filter dynamics, one *fast* and the other *slow*

$$\dot{z} = \lambda a_{\text{fast}}(z - y), \quad \dot{z} = \lambda a_{\text{slow}}(z - y), \tag{31}$$

where $\lambda > 0$, and $a_{\text{fast}} < a_{\text{slow}} < 0$. When choosing a family $(\lambda_i)_{1\leq i\leq m}$, the corresponding observers are given by (2) with $A_{\text{fast}} = a_{\text{fast}}\,\text{diag}(\lambda_i)_{1\leq i\leq m}$ and $B_{\text{fast}} = -a_{\text{fast}}(1,\dots,1)^\top$ (resp. $A_{\text{slow}} = a_{\text{slow}}\,\text{diag}(\lambda_i)_{1\leq i\leq m}$ and $B_{\text{slow}} = -a_{\text{slow}}(1,\dots,1)^\top$). It is well-known that when tuning the parameters of linear filters, a compromise has to be found between speed of convergence and robustness to noise. More precisely, modulo the nonlinear left-inversion of $\boldsymbol{T}$, KKL observers built with fast (resp. slow) linear filters usually exhibit fast (resp. slow) convergence properties but poor (resp. good) robustness to measurement noise.

On the other hand, building a KKL observer based on nonlinear filters allows to consider dynamics of the form

$$\dot{z} = \lambda(a_{\text{fast}}(z - y) + (a_{\text{slow}} - a_{\text{fast}})\tanh(z - y)) \tag{32}$$

which can be checked to verify Assumption 3. When choosing a family $(\lambda_i)_{1\leq i\leq m}$, the corresponding observer is given by (3) for $k = 1$ and the corresponding contraction $\boldsymbol{\sigma}$ as in (6). The motivation is the following: On the one hand, for large values of $|z - y|$, the nonlinear dynamics behave as the *fast* linear dynamics. Therefore, a fast convergence is expected. On the other hand, for small values of $|z - y|$, the nonlinear dynamics behave as the *slow* linear dynamics. Therefore, the same robustness with respect to small perturbations of the measurement y is expected. In other words, the nonlinear KKL observers is expected to take the best of both worlds.

For simulations, we consider a nonlinear Duffing oscillator

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -0.2x_1 - x_1^3 \end{cases}, \quad y = x_1, \tag{33}$$

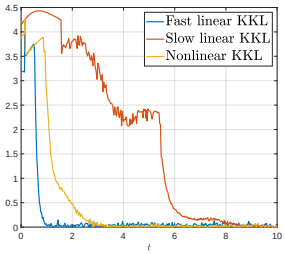
which verifies Assumptions 1 and 2 with $\mathcal{X}_0 = [-2, 2]^2$ and $m = 2$. We pick $\lambda_1 = -2$, $\lambda_2 = -4$, $\lambda_3 = -6$, $a_{\text{fast}} = 5$ and $a_{\text{slow}} = 0.5$. For each KKL observer, we create a dataset of points $(x, \boldsymbol{z})$ approximating $(x, \boldsymbol{T}(x))$, with the map $\boldsymbol{T}$ such that the image of solutions to (33) by $\boldsymbol{T}$ is solution to the corresponding observer dynamics (2) or (3). This is done by simulating the interconnection of (33) with (2) or (3) from a grid of 200×200 initial conditions in $\mathcal{X}_0 \times \{0\}$, and storing the obtained pairs $(x, \boldsymbol{z})$ after a time $t_l := 20/\min\{|\lambda_i|\}$, needed for the filters to "forget" their initial condition. Values of $\boldsymbol{T}(x)$ for a given x and of $\boldsymbol{T}^{\text{inv}}(\boldsymbol{z})$ for a given $\boldsymbol{z}$ can then be obtained by fetching the closest point in the dataset.

We then propose two simulation scenarios: 1) we initialize all observers with the same initial error randomly picked so that $|\boldsymbol{z}(0) - \boldsymbol{T}(x(0))| = 100$ and we do not add measurement noise, and 2) we initialize all observers to their correct initial condition $\boldsymbol{T}(x(0))$ and we add a sinusoidal measurement noise ($\nu(t) = 0.1\sin(10t)$). The first (resp. second) scenario aims at comparing the convergence times (resp. the impact of measurement noise). The numerical results for a particular choice of $x(0)$ are provided in Figure 1. As expected, the KKL observer with nonlinear dynamics converges as fast as the one with fast linear dynamics and seems almost as robust to noise than the one with slow linear dynamics. Minimum, maximum and mean convergence time and gain with respect to noise for

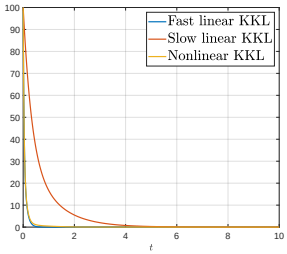
		Fast linear	Slow linear	Nonlinear
Convergence time	Min	0.78	6.29	1.52
	Max	0.87	7.72	3.15
	Mean	0.83	6.79	2.27
Gain w.r.t. noise	Min	7.23	0.94	1.56
	Max	8.07	1.53	2.37
	Mean	7.57	1.15	1.95

Table 1. Comparison of KKL observers with slow/fast linear dynamics and KKL observer with nonlinear dynamics in terms of convergence time and gain with respect to noise of $|\hat{x} - x|$.

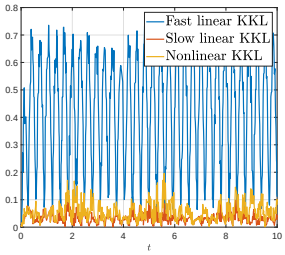
100 random $x(0)$ are given in Table 1. The convergence time is computed in Scenario 1 as the first time after which the error remains below tolerance thresholds, determined by the precision of the approximation of $\boldsymbol{T}$ and $\boldsymbol{T}^{\text{inv}}$. The robustness to measurement noise is quantified in Scenario 2 by dividing the 2-norm of the steady state error by the amplitude of the noise. We refer the reader to Brivadis et al. (2024) to experiment the simulations.



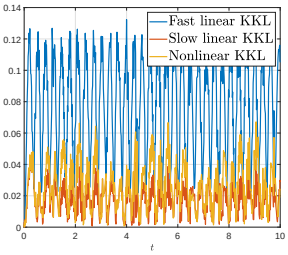
(a) Scenario 1: $|\hat{x}(t) - x(t)|$



(b) Scenario 1: $|\boldsymbol{z}(t) - \boldsymbol{T}(x(t))|$



(c) Scenario 2: $|\hat{x}(t) - x(t)|$



(d) Scenario 2: $|\boldsymbol{z}(t) - \boldsymbol{T}(x(t))|$

Fig. 1. Estimation error in x and z coordinates from two KKL observers with linear dynamics (2) with slow and fast pairs $(A_{\text{slow}}, B_{\text{slow}})$ and $(A_{\text{fast}}, B_{\text{fast}})$ defined from (31), and by a KKL observer with nonlinear dynamics (3) defined from (32). Scenario 1: initial error but no noise; Scenario 2 : no initial error but noise.

5. CONCLUSION

In this article, we have presented a KKL-type observer with non-linear dynamics. In the scenario where the system is differentially observable of order m, we have demonstrated the existence of such an estimation algorithm when the observer's dynamics are structured as m non-linear filters operating in parallel, provided that their dynamics are sufficiently fast. Through a simplified illustration, we have highlighted a potential application of this technique to obtain a better qualitative behavior of the estimate. In a more general context, demonstrating the existence of such an observer for a broader class of contractions would

be highly intriguing. However, the techniques employed in this article may not readily adapt to this more general framework.

REFERENCES

Andrieu, V. (2014). Convergence speed of nonlinear Luenberger observers. *SIAM Journal on Control and Optimization*, 52(5), 2831–2856.

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.

Bernard, P., Andrieu, V., and Astolfi, D. (2022). Observer design for continuous-time dynamical systems. *Annual Reviews in Control*, 53, 224–248.

Brivadis, L., Andrieu, V., Bernard, P., and Serres, U. (2023). Further remarks on KKL observers. *Systems & Control Letters*, 172, 105429.

Brivadis, L., Bernard, P., and Andrieu, V. (2024). KKL observers with nonlinear dynamics. Available at `https://github.com/paulinebernard/KKL-with-Nonlinear-Dynamics`.

Buisson-Fenet, M., Bahr, L., Morgenthaler, V., and Meglio, F.D. (2023). Towards gain tuning for numerical KKL observers. *IFAC-PapersOnLine*, 56(2), 4061–4067. 22nd IFAC World Congress.

Chong, M.S., Nešić, D., Postoyan, R., and Kuhlmann, L. (2015). Parameter and state estimation of nonlinear systems using a multi-observer under the supervisory framework. *IEEE Transactions on Automatic Control*, 60(9), 2336–2349.

Esfandiari, K. and Shakarami, M. (2019). Bank of high-gain observers in output feedback control: Robustness analysis against measurement noise. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(4), 2476–2487.

Janny, S., Andrieu, V., Nadri, M., and Wolf, C. (2021). Deep KKL: Data-driven output prediction for non-linear systems. In *2021 60th IEEE Conference on Decision and Control (CDC)*, 4376–4381. IEEE.

Kalman, R. and Bucy, R. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 108, 83–95.

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).

Levant, A. (2003). Higher-order sliding modes, differentiation and output-feedback control. *International journal of Control*, 76(9-10), 924–941.

Luenberger, D.G. (1964). Observing the state of a linear system. *IEEE Transactions on Military Electronics*, 8(2), 74–80.

Niazi, M.U.B., Cao, J., Sun, X., Das, A., and Johansson, K.H. (2023). Learning-based design of Luenberger observers for autonomous nonlinear systems. In *2023 American Control Conference (ACC)*, 3048–3055.

Pachy, V., Andrieu, V., Bernard, P., Brivadis, L., and Praly, L. (2024). On the existence of KKL observers with nonlinear contracting dynamics. *MICNON 2024*.

Pavlov, A., van de Wouw, N., and Nijmeijer, H. (2004). The uniform global output regulation problem. In *IEEE Conference on Decision and Control (CDC)*, volume 5, 4921–4926 Vol.5.

Peralez, J. and Nadri, M. (2021). Deep learning-based Luenberger observer design for discrete-time nonlinear systems. In *2021 60th IEEE Conference on Decision and Control (CDC)*, 4370–4375.

Petri, E., Postoyan, R., Astolfi, D., Nesic, D., and Andrieu, V. (2023). Hybrid multi-observer for improving estimation performance. *arXiv preprint arXiv:2303.06936*.

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

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

Tarbouriech, S., Alessandri, A., Astolfi, D., and Zaccarian, L. (2022). Lmi-based stubborn and dead-zone redesign in linear dynamic output feedback. *IEEE Control Systems Letters*, 7, 187–192.

Appendix A. PROOFS

This part of the paper is not in the version which has been published in Pachy et al. (2024).

In this section, we give the technical proofs employed to get the main result. Note that to simplify the presentation, we use the following notation: $\sigma^{(j)}(x) = \frac{\partial^j \sigma}{\partial z^j}(\psi(h(x)), h(x))$.

A.1 Proof of Proposition 3

First of all, we have the following technical lemma whose proof can be found in Appendix A.3.

Lemma 1. For each $x \in \mathbb{R}^n$ and $\lambda > 0$,

$$L_f T_a(x,\lambda) = \lambda \sum_{j=1}^{m-1} \frac{\sigma^{(j)}(x)}{j!} \left[\sum_{\ell=1}^{m-1} \frac{\phi_\ell(x)}{\lambda^\ell} \right] \Bigg|_{\leq m-1}^{j} + \frac{1}{\lambda^{m-1}} L_f \phi_{m-1}(x). \quad \text{(A.1)}$$

Hence, with (18), (19), (16) and (A.1), it yields

$$\omega(x,\lambda) = \sum_{i=1}^{m-1} \frac{\sigma^{(i)}(x)}{i!} \left[\sum_{\ell=1}^{m-1} \frac{\phi_\ell(x)}{\lambda^\ell} \right] \Bigg|_{\leq m-1}^{i} - \sigma\left(\psi(h(x)) + \sum_{\ell=1}^{m-1} \frac{\phi_\ell(x)}{\lambda^\ell}, h(x) \right) + \frac{1}{\lambda^m} L_f \phi_{m-1}(x)$$

As σ is C^{m+1}, using (9) and a Taylor expansion of σ around $(\psi(h(x)), h(x))$,

$$\omega(x,\lambda) = \frac{\omega_0(x,\lambda)}{\lambda^m} - G(x,\lambda) + \frac{1}{\lambda^m} L_f \phi_{m-1}(x)$$

where

$$\omega_0(x,\lambda) = \frac{1}{(m-1)!} \int_0^1 \left[(1-s)^{m-1} \frac{\partial^m \sigma}{\partial z^m} \left(\psi(h(x)) + s \sum_{\ell=1}^{m-1} \frac{\phi_\ell(x)}{\lambda^\ell}, h(x) \right) \right] \mathrm{d}s \left[\sum_{\ell=1}^{m-1} \frac{\phi_\ell(x)}{\lambda^{\ell-1}} \right]^m$$

$$G(x,\lambda) = \sum_{i=1}^{m-1} \frac{\sigma^{(i)}(x)}{i!} \left(\left[\sum_{\ell=1}^{m-1} \frac{\phi_\ell(x)}{\lambda^\ell} \right]^i - \left[\sum_{\ell=1}^{m-1} \frac{\phi_\ell(x)}{\lambda^\ell} \right] \Bigg|_{\leq m-1}^{i} \right)$$

Note that G equals 0 if $i = 1$ and equals

$$\frac{1}{\lambda^m}\sum_{j=0}^{i(m-1)}\frac{1}{\lambda^j}\tilde{\phi}_{ij}(x)$$

if $i > 1$, with $\tilde{\phi}_{ij}$ being a certain polynomial of the ϕ_ℓ for $\ell = 1, \dots, m-1$, therefore this gives

$$\omega(x,\lambda) = \frac{\omega_0(x,\lambda) + L_f\phi_{m-1}(x)}{\lambda^m} + \frac{1}{\lambda^m}\sum_{i=2}^{m-1}\frac{\sigma^{(i)}(x)}{i!}\sum_{j=0}^{i(m-1)}\frac{\tilde{\phi}_{ij}(x)}{\lambda^j}. \quad \text{(A.2)}$$

Therefore, using triangular inequality, for $x \in \mathcal{X} + \delta_u$ and $\lambda > 0$,

$$|\lambda^m||\omega(x,\lambda)| \le |\omega_0(x,\lambda)| + |L_f\phi_{m-1}(x)| + \sum_{i=2}^{m-1}\frac{1}{i!}\left|\sigma^{(i)}(x)\right|\sum_{j=0}^{i(m-1)}\frac{1}{\lambda^j}|\tilde{\phi}_{ij}(x)|. \quad \text{(A.3)}$$

As σ and ψ are C^{m+1}, $\sigma^{(i)}$ is continuous, and therefore bounded on the compact set $\mathcal{X}+\delta_u$ for all i in $\{1,\dots,m\}$. Moreover, $\tilde{\phi}_{kj}$ is a polynomial of the continuous functions ϕ_ℓ, and therefore is bounded on the compact set $\mathcal{X}+\delta_u$. As ϕ_{m-1} is C^2, $L_f\phi_{m-1}$ is also bounded on $\mathcal{X}+\delta_u$. Moreover, there exists $\delta_\psi > 0$ such that $\left|\sum_{\ell=1}^{m-1}\frac{\phi_\ell(x)}{\lambda^\ell}\right| \le \delta_\psi$ for all $x \in \mathcal{X} + \delta_u$ and all $\lambda > 1$. By continuity of $\frac{\partial^m\sigma}{\partial z^m}$, it is bounded on $(\psi(h(\mathcal{X}+\delta_u)) + \delta_\psi) \times h(\mathcal{X}+\delta_u)$, uniformly in $\lambda > 1$. We can conclude that there exists $\bar{\omega} > 0$, independent from λ, such that (25) holds. Note also that ω is C^1 and so are all the previously listed maps. So similarly, it yields,

$$\lambda^m\left|\frac{\partial\omega}{\partial x}(x,\lambda)\right| \le \left|\frac{\partial\omega_0}{\partial x}(x,\lambda)\right| + \left|\frac{\partial L_f\phi_{m-1}}{\partial x}(x)\right| + \sum_{i=2}^{m-1}\frac{1}{i!}\left|\frac{\partial\sigma^{(i)}}{\partial x}(x)\right|\sum_{j=0}^{i(m-1)}\frac{1}{\lambda^j}\left|\frac{\partial\tilde{\phi}_{ij}}{\partial x}(x)\right|. \quad \text{(A.4)}$$

Hence, there exists a positive real number k'_ω, such that for all $\lambda > 1$ and $x \in \mathcal{X} + \delta_u$, it yields

$$\lambda^m\left|\frac{\partial\omega}{\partial x}(x,\lambda)\right| \le k'_\omega.$$

Inequality (26) is obtained for some $k_\omega > k'_\omega$.

A.2 Proof of Proposition 4

The evaluation $R(X(x,t),\lambda)$ of R along any solution initiated at $x \in \mathcal{X} + \delta_u$ satisfies

$$\frac{d}{dt}R(X(x,t),\lambda) = \lambda[\sigma(T_a(X(x,t),\lambda) + R(X(x,t),\lambda), h(X(x,t))) - \sigma(T_a(X(x,t),\lambda), h(X(x,t)))] - \lambda\omega(X(x,t),\lambda)$$

With the condition (7b) it yields for all $x \in \mathcal{X} + \delta_u$

$$R(x,\lambda)\big[\sigma(T(x,\lambda) + R(x,\lambda), h(x)) - \sigma(T(x,\lambda), h(x))\big] \le -\alpha R(x,\lambda)^2.$$

Since moreover, $\mathcal{X} + \delta_u$ is negatively invariant along the flow and with Young's inequality, for all $\varepsilon > 0$ it gives

$$\begin{aligned}\frac{d}{dt}R(X(x,t),\lambda)^2 &\le -2\lambda\alpha R(X(x,t),\lambda)^2 + \frac{\lambda}{\varepsilon}\omega(X(x,t),\lambda)^2 + \lambda\varepsilon R(X(x,t),\lambda)^2 \\ &= -\lambda\alpha R(X(x,t),\lambda)^2 + \frac{\lambda}{\alpha}\omega(X(x,t),\lambda)^2\end{aligned}$$

where we chose $\varepsilon = \alpha$ for the last equality to hold. By multiplying each side by $e^{\lambda\alpha t}$ and rearranging the terms, we obtain

$$\frac{d}{dt}(R(X(x,t),\lambda)^2 e^{\lambda\alpha t}) \le \frac{\lambda}{\alpha}\omega(X(x,t),\lambda)^2 e^{\lambda\alpha t}. \quad \text{(A.5)}$$

Let $\delta > 0$. Integrating between $-\delta$ and 0 and using (25), we get for any $x \in \mathcal{X} + \delta_u$ and any $\lambda > 1$,

$$\begin{aligned}R(x,\lambda)^2 &\le R(X(x,-\delta),\lambda)^2 e^{-\lambda\alpha\delta} + \frac{\lambda}{\alpha}\int_{-\delta}^{0}\omega(X(x,s),\lambda)^2 e^{\lambda\alpha s}\mathrm{d}s \\ &\le R(X(x,-\delta),\lambda)^2 e^{-\lambda\alpha\delta} + \frac{\lambda}{\alpha}\frac{\bar{\omega}^2}{\lambda^{2m}}\int_{-\delta}^{0}e^{\lambda\alpha s}\mathrm{d}s \\ &= R(X(x,-\delta),\lambda)^2 e^{-\lambda\alpha\delta} + \frac{\lambda\bar{\omega}^2(1-e^{-\lambda\alpha\delta})}{\alpha^2\lambda^{2m+1}}\end{aligned}$$

$\delta \mapsto X(x,-\delta)$ is bounded, so with the definition of T in (12), $\delta \mapsto T(X(x,-\delta),\lambda)$ is bounded, and, by continuity of ϕ_ℓ for each $\ell \in \{0,\dots,m-1\}$, $\delta \mapsto R(X(x,-\delta),\lambda)$ is also bounded. Letting δ go to $+\infty$, it yields

$$\forall\lambda > 1, \forall x \in \mathcal{X} + \delta_u, \quad R(x,\lambda)^2 \le \frac{\bar{\omega}^2}{\alpha^2\lambda^{2m}}.$$

We finally get (28).

A.3 Proof of Lemma 1

Define $T_{a,i}(x,\lambda) = \sum_{j=0}^{i-1}\frac{\phi_j(x)}{\lambda^j}$, so that $T_{a,m} = T_a$. The proof of Lemma 1 is obtained by recursion. We are going to prove the following property for $i \ge 1$:

$\underline{\mathcal{P}(i)}$: For all $x \in \mathbb{R}^n$ and $\lambda > 0$, $T_{a,i}$ verifies

$$L_fT_{a,i}(x,\lambda) = \lambda\sum_{j=1}^{i-1}\frac{\sigma^{(j)}(x)}{j!}\left[\sum_{\ell=1}^{i-1}\frac{\phi_\ell(x)}{\lambda^\ell}\right]\Bigg|_{\le i-1}^{j} + \frac{L_f\phi_{i-1}(x)}{\lambda^{i-1}}. \quad \text{(A.6)}$$

Clearly, $L_fT_{a,1} = L_f\phi_0$ so the property is true for $i = 1$. Assume $\mathcal{P}(i)$ is true for a certain $i \ge 1$. Let us show that $\mathcal{P}(i+1)$ holds. We have

$$\begin{aligned}L_fT_{a,i+1}(x,\lambda) &= \sum_{\ell=0}^{i}\frac{L_f\phi_\ell(x)}{\lambda^\ell} \\ &= L_fT_{a,i}(x,\lambda) + \frac{L_f\phi_i(x)}{\lambda^i} \\ &= \lambda\sum_{j=1}^{i-1}\frac{\sigma^{(j)}(x)}{j!}\left[\sum_{\ell=1}^{i-1}\frac{\phi_\ell(x)}{\lambda^\ell}\right]\Bigg|_{\le i-1}^{j} + \frac{L_f\phi_{i-1}(x)}{\lambda^{i-1}} + \frac{L_f\phi_i(x)}{\lambda^i}.\end{aligned}$$

On another hand, we have

$$
\begin{aligned}
&\lambda\sum_{j=1}^{i}\frac{\sigma^{(j)}(x)}{j!}\left[\sum_{\ell=1}^{i-1}\frac{\phi_\ell(x)}{\lambda^\ell}+\frac{\phi_i(x)}{\lambda^i}\right]\Bigg|_{\le i}^{j}+\frac{L_f\phi_i(x)}{\lambda^i}\\
&=\lambda\Bigg[\sum_{j=1}^{i-1}\frac{\sigma^{(j)}(x)}{j!}\left[\sum_{\ell=1}^{i-1}\frac{\phi_\ell(x)}{\lambda^\ell}\right]\Bigg|_{\le i}^{j}+\frac{\sigma^{(1)}(x)}{\lambda^i}\phi_i(x)\\
&\quad+\frac{\sigma^{(i)}(x)}{i!}\left[\sum_{\ell=1}^{i-1}\frac{1}{\lambda^\ell}\phi_\ell(x)\right]\Bigg|_{\le i}^{i}\Bigg]+\frac{L_f\phi_i(x)}{\lambda^i}\\
&=\lambda\Bigg[\sum_{j=1}^{i-1}\frac{\sigma^{(j)}(x)}{j!}\left[\sum_{\ell=1}^{i-1}\frac{\phi_\ell(x)}{\lambda^\ell}\right]\Bigg|_{\le i-1}\\
&\quad+\sum_{j=1}^{i-1}\frac{\sigma^{(j)}(x)}{j!}\left[\sum_{\ell=1}^{i-1}\frac{\phi_\ell(x)}{\lambda^\ell}\right]\Bigg|_{=i}^{j}\\
&\quad+\sigma^{(1)}(x)\frac{\phi_i(x)}{\lambda^i}+\frac{1}{\lambda^i}\frac{\sigma^{(i)}(x)}{i!}(\phi_1(x))^i\Bigg]+\frac{L_f\phi_i(x)}{\lambda^i}\\
&=L_fT_{a,i+1}(x,\lambda)+\lambda\Bigg[\sum_{j=1}^{i-1}\frac{\sigma^{(j)}(x)}{j!}\left[\sum_{\ell=1}^{i-1}\frac{\phi_\ell(x)}{\lambda^\ell}\right]\Bigg|_{=i}^{j}\\
&\quad+\sigma^{(1)}(x)\frac{\phi_i(x)}{\lambda^i}+\frac{1}{\lambda^i}\frac{\sigma^{(i)}(x)}{i!}(\phi_1(x))^i\Bigg]-\frac{L_f\phi_{i-1}(x)}{\lambda^{i-1}}.
\end{aligned}
$$

Therefore, with ϕ_i defined in (17) and noting that

$$
\left[\sum_{\ell=1}^{i-1}\frac{\phi_\ell(x)}{\lambda^\ell}\right]\Bigg|_{=i}^{j}=\frac{1}{\lambda^i}\sum_{\substack{\ell_1+\cdots+\ell_j=i\\1\le\ell_1,\ldots,\ell_j\le i-1}}\phi_{\ell_1}(x)\cdots\phi_{\ell_j}(x)
$$

it yields the result.

A.4 Proof of Proposition 5

To simplify the readability of this proof, in the following the dependencies on λ have been removed.

Note that given $(z_a, r_a, y_a, z_b, r_b, y_b)$ in $\mathbb{R}^6$, the function σ being C^2, we have

$$
\sigma(z_a+r_a,y_a)-\sigma(z_a,y_a)=\left[\int_0^1\frac{\partial\sigma}{\partial z}(z_a+\theta r_a,y_a)\mathrm{d}\theta\right]r_a
$$

and

$$
\begin{aligned}
&(\sigma(z_a+r_a,y_a)-\sigma(z_a,y_a))-(\sigma(z_b+r_b,y_b)-\sigma(z_b,y_b))\\
&=\left[\int_0^1\frac{\partial\sigma}{\partial z}(z_a+\theta r_a,y_a)\mathrm{d}\theta\right]r_a-\left[\int_0^1\frac{\partial\sigma}{\partial z}(z_b+\theta r_b,y_b)\mathrm{d}\theta\right]r_b\\
&=\left[\int_0^1\frac{\partial\sigma}{\partial z}(z_a+\theta r_a,y_a)\mathrm{d}\theta\right][r_a-r_b]\\
&\quad+\left[\int_0^1\frac{\partial\sigma}{\partial z}(z_a+\theta r_a,y_a)-\frac{\partial\sigma}{\partial z}(z_b+\theta r_b,y_b)\mathrm{d}\theta\right]r_b\\
&=\left[\int_0^1\frac{\partial\sigma}{\partial z}(z_a+\theta r_a,y_a)\mathrm{d}\theta\right][r_a-r_b]\\
&+\left(\int_0^1\left[\int_0^1\frac{\partial^2\sigma}{\partial z^2}(z(\theta,\rho),y(\rho))(z_a-z_b+\theta(r_a-r_b))\mathrm{d}\rho\right]\mathrm{d}\theta\right)r_b\\
&\qquad+\left(\int_0^1\left[\int_0^1\frac{\partial^2\sigma}{\partial z\partial y}(z(\theta,\rho),y(\rho))(y_a-y_b)\mathrm{d}\rho\right]\mathrm{d}\theta\right)r_b
\end{aligned}
$$

where

$$
\begin{aligned}
z(\theta,\rho)&=z_b+\theta r_b+\rho[z_a-z_b+\theta(r_a-r_b)]\\
y(\rho)&=y_b+\rho(y_a-y_b).
\end{aligned}
$$

Moreover, for all (x_a, x_b) in $(\mathcal{X}+\delta_u)^2$,

$$
\begin{aligned}
&\frac{d}{dt}(R(X(x_a,t))-R(X(x_b,t)))\\
&=\lambda\Big[\sigma(T_a(X(x_a,t))+R(X(x_a,t)),h(X(x_a,t)))\\
&\qquad-\sigma(T_a(X(x_a,t)),h(X(x_a,t)))\\
&\quad-\big(\sigma(T_a(X(x_b,t))+R(X(x_b,t)),h(X(x_b,t)))\\
&\qquad-\sigma(T_a(X(x_b,t))),h(X(x_b,t)))\big)\Big]\\
&\quad-\lambda\big[\omega(X(x_a,t))-\omega(X(x_b,t))\big].
\end{aligned}
$$

On another hand, with Proposition 4, there exists a positive real number $B_{\sigma,2}$ such that for all (x_a, x_b) in $(\mathcal{X}+\delta_u)^2$ and for all $\lambda>1$, and (θ,ρ) in $[0,1]^2$,

$$
\begin{aligned}
\left|\frac{\partial^2\sigma}{\partial z^2}(z(\theta,\rho),y(\rho))\right|&\le B_{\sigma,2}\\
\left|\frac{\partial^2\sigma}{\partial z\partial y}(z(\theta,\rho),y(\rho))\right|&\le B_{\sigma,2}
\end{aligned}
$$

with $z_a=T_a(x_a)$, $z_b=T_a(x_b)$, $r_a=R(x_a)$, $r_b=R(x_b)$, $y_a=h(x_a)$, and $y_b=h(x_b)$. Hence, since $\mathcal{X}+\delta_u$ is invariant along the flow and with (7a), we get

$$
\begin{aligned}
&\frac{d}{dt}\Delta R(X(x_a,t),X(x_b,t))^2\\
&\le-2\lambda(\alpha-B_{\sigma,2}|R(X(x_b,t))|)\Delta R(X(x_a,t),X(x_b,t))^2\\
&\quad+2\lambda B_{\sigma,2}|R(X(x_b,t))|\\
&\qquad S(X(x_a,t),X(x_b,t))|\Delta R(X(x_a,t),X(x_b,t))|\\
&\quad+2\lambda|\Delta\omega(X(x_a,t),X(x_b,t))||\Delta R(X(x_a,t),X(x_b,t))|,
\end{aligned}
$$

where $S=|\Delta T_a|+|\Delta h|$. Using Young's inequality twice, we obtain for some ε and ε' strictly positive:

$$
\begin{aligned}
&\frac{d}{dt}\Delta R(X(x_a,t),X(x_b,t))^2\\
&\le-2\lambda\left(\alpha-\frac{\varepsilon'}{2}-B_{\sigma,2}|R(X(x_b,t))|\left(1+\frac{\varepsilon}{2}\right)\right)\\
&\qquad\times\Delta R(X(x_a,t),X(x_b,t))^2\\
&\qquad+\frac{\lambda}{\varepsilon}B_{\sigma,2}|R(X(x_b,t))|S(X(x_a,t),X(x_b,t))^2\\
&\qquad+\frac{\lambda}{\varepsilon'}\Delta\omega(X(x_a,t),X(x_b,t))^2
\end{aligned}
$$

Then, using Proposition 4,

$$
\begin{aligned}
&\frac{d}{dt}\Delta R(X(x_a,t),X(x_b,t))^2\\
&\le-2\lambda\left(\alpha-\frac{\varepsilon'}{2}-B_{\sigma,2}\frac{\bar\omega}{\alpha\lambda^m}\left(1+\frac{\varepsilon}{2}\right)\right)\\
&\qquad\times\Delta R(X(x_a,t),X(x_b,t))^2\\
&\qquad+\frac{\bar\omega\lambda B_{\sigma,2}}{\alpha\varepsilon\lambda^m}S(X(x_a,t),X(x_b,t))^2\\
&\qquad+\frac{\lambda}{\varepsilon'}\Delta\omega(X(x_a,t),X(x_b,t))^2.
\end{aligned}
$$

Choosing $\varepsilon=\frac{\alpha^2}{2B_{\sigma,2}\bar\omega}\lambda^m$ and $\varepsilon'=\frac{\alpha}{2}$ gives

$$
\begin{aligned}
&\frac{d}{dt}\Delta R(X(x_a,t),X(x_b,t))^2\\
&\leq -\frac{\lambda}{2}\left(\alpha - 4B_{\sigma,2}\frac{\bar\omega}{\alpha\lambda^m}\right)\Delta R(X(x_a,t),X(x_b,t))^2\\
&\quad + \frac{\bar\omega^2 B_{\sigma,2}\lambda}{\alpha^3\lambda^{2m}} S(X(x_a,t),X(x_b,t))^2\\
&\quad + \frac{2\lambda}{\alpha}\Delta\omega(X(x_a,t),X(x_b,t))^2.
\end{aligned}
$$

If we multiply by $\exp\left(\frac{\lambda}{2}\left(\alpha - 4B_{\sigma,2}\frac{\bar\omega}{\alpha\lambda^m}\right)t\right) \equiv \exp(\Theta t)$ on both sides, we obtain (by ommitting the dependency on λ of Θ for now)

$$
\begin{aligned}
&\frac{d}{dt}\Big(\Delta R(X(x_a,t),X(x_b,t))^2\exp(\Theta t)\Big)\\
&\leq\Big(\frac{\bar\omega^2 B_{\sigma,2}\lambda}{\alpha^3\lambda^{2m}} S(X(x_a,t),X(x_b,t))^2\\
&\quad + \frac{2\lambda}{\alpha}\Delta\omega(X(x_a,t),X(x_b,t))^2\Big)\exp(\Theta t)
\end{aligned}
$$

We then integrate between $-\delta < 0$ and 0, for $(x_a,x_b) \in (\mathcal{X}+\delta_u)^2$ and $\lambda > 1$,

$$
\begin{aligned}
&\Delta R(x_a,x_b)^2\\
&\leq \Delta R(X(x_a,-\delta),X(x_b,-\delta))^2\exp(-\Theta\delta)\\
&\quad + \int_{-\delta}^0 \frac{\bar\omega^2 B_{\sigma,2}\lambda}{\alpha^3\lambda^{2m}} S(X(x_a,s),X(x_b,s))^2\exp(\Theta s)\mathrm{d}s\\
&\quad + \int_{-\delta}^0 \frac{2\lambda}{\alpha}\Delta\omega(X(x_a,s),X(x_b,s))^2\exp(\Theta s)\mathrm{d}s
\end{aligned}
$$

Since by Proposition 3, $s \mapsto \Delta\omega(X(x_a,s),X(x_b,s))^2$ is bounded and $s \mapsto \Delta T(X(x_a,s),X(x_b,s))^2$ is also bounded, it yields that for all $\lambda > \left(4B_{\sigma,2}\frac{\bar\omega}{\alpha^2}\right)^{1/m}$ the former integrals are well defined when δ goes to $+\infty$. Moreover, since by Proposition 4 and the invariance of $(\mathcal{X}+\delta_u)$, $\delta \mapsto \Delta R(X(x_a,-\delta),X(x_b,-\delta))^2$ is bounded it yields

$$
\begin{aligned}
&\Delta R(x_a,x_b)^2\\
&\leq \int_{-\infty}^0 \frac{\bar\omega^2 B_{\sigma,2}\lambda}{\alpha^3\lambda^{2m}} S(X(x_a,s),X(x_b,s))^2\exp(\Theta s)\mathrm{d}s\\
&\quad + \int_{-\infty}^0 \frac{2\lambda}{\alpha}\Delta\omega(X(x_a,s),X(x_b,s))^2\exp(\Theta s)\mathrm{d}s.
\end{aligned}
$$

Also, the functions h and ϕ_ℓ being C^1, by definition of T_a in (19), there exists k_S such that for all $\lambda > 1$ and all (x_a,x_b) in $(\mathcal{X}+\delta_u)^2$,

$$
S(x_a,x_b) \leq k_S|x_a - x_b|.
$$

Using Proposition 3 with $\lambda > \max\{\left(4B_{\sigma,2}\frac{\bar\omega}{\alpha^2}\right)^{1/m}, 1\}$ we can write

$$
\begin{aligned}
&\Delta R(x_a,x_b)^2\\
&\leq \frac{\bar\omega^2 B_{\sigma,2}\lambda}{\alpha^3\lambda^{2m}}k_S^2\int_{-\infty}^0 |X(x_a,s)-X(x_b,s)|^2\exp(\Theta s)\mathrm{d}s\\
&\quad + \frac{2\lambda}{\alpha}\frac{k_\omega^2}{\lambda^{2m}}\int_{-\infty}^0 |X(x_a,s)-X(x_b,s)|^2\exp(\Theta s)\mathrm{d}s.
\end{aligned}
\tag{A.7}
$$

Since f is Lipschitz on $(\mathcal{X}+\delta_u)$, there exists $k_f > 0$ such that for all $s < 0$ and (x_a,x_b) in $(\mathcal{X}+\delta_u)^2$,

$$
\begin{aligned}
&|X(x_a,s)-X(x_b,s)|\\
&\qquad \leq |x_a - x_b| + \int_s^0 k_f|X(x_a,s)-X(x_b,s)|\mathrm{d}s.
\end{aligned}
$$

Grönwall's lemma gives for $t < 0$,

$$
|X(x_a,t)-X(x_b,t)| \leq |x_a - x_b|\exp(-k_f t)
$$

and reinjecting this in (A.7),

$$
\begin{aligned}
&\Delta R(x_a,x_b)^2\\
&\leq \frac{\bar\omega^2 B_{\sigma,2}\lambda k_S^2}{\alpha^3\lambda^{2m}}|x_a-x_b|^2\int_{-\infty}^0 \exp\left(\left[\Theta - 2k_f\right]s\right)\mathrm{d}s\\
&\quad + \frac{2\lambda}{\alpha}\frac{k_\omega^2}{\lambda^{2m}}|x_a-x_b|^2\int_{-\infty}^0 \exp\left(\left[\Theta - 2k_f\right]s\right)\mathrm{d}s.
\end{aligned}
$$

Let $\lambda^* > 1$ be large enough such that

$$
\frac{\lambda}{2}\left(\alpha - 4B_{\sigma,2}\frac{\bar\omega}{\alpha\lambda^m}\right) - 2k_f > 0.
$$

Then, for all $\lambda > \lambda^*$, the integrals converge and we get

$$
\int_{-\infty}^0 \exp\left(\left[\Theta - 2k_f\right]s\right)\mathrm{d}s = \frac{1}{\Theta - 2k_f}.
$$

By integrating the last identity into the previous one and taking the square root, we have for any $(x_a,x_b) \in \mathcal{X}^2$ and any $\lambda > \lambda^*$,

$$
|\Delta R(x_a,x_b)| \leq \frac{1}{\lambda^m}\left(\frac{\bar\omega^2 B_{\sigma,2}\lambda k_S^2\alpha^2 + 2\lambda k_\omega^2}{\alpha^3(\Theta - 2k_f)}\right)^{1/2}|x_a - x_b|.
$$

Which can be rewritten as (29) for some positive real number k_R.

A.5 Proof of Proposition 2

We now establish Lipschitz injectivity of $\boldsymbol{\phi}$. For ϕ_0, under Assumption 3, we have

$$
\left|\frac{d\psi}{dy}(y)\right| > \left|\frac{\frac{\partial\sigma}{\partial y}(\psi(y),y)}{\frac{\partial\sigma}{\partial z}(\psi(y),y)}\right| \geq \frac{\gamma}{\beta} \equiv \mu_0 > 0. \tag{A.8}
$$

So we deduce that

$$
\forall (x_a,x_b) \in \mathcal{X}^2, \quad |\Delta\phi_0(x_a,x_b)| \geq \mu_0|\Delta h(x_a,x_b)|.
$$

Then, an immediate recursion from (16)-(17) shows that there exist C^1 functions $P_1,\dots,P_{m-1}$ such that for all $x \in \mathcal{X}+\delta_u$, $\lambda > 0$, and $\ell \in \{1,\dots,m-1\}$, ϕ_ℓ is written as

$$
\phi_\ell(x) = \kappa(h(x))^\ell\frac{d\psi}{dy}(h(x))L_f^\ell h(x) + P_\ell(\boldsymbol{H}_\ell(x)).
$$

where

$$
\boldsymbol{H}_\ell = (h,\dots,L_f^{\ell-1}h).
$$

To simplify the expressions, when there is no ambiguity, we write $g(h) = g\circ h$ for the composition of functions.

For $\ell \geq 1$ and $(x_a,x_b) \in \mathcal{X}^2$, the triangular inequality gives

$$
\begin{aligned}
|\Delta\phi_\ell(x_a,x_b)| \geq &\left|\Delta(\kappa^\ell(h)\frac{d\psi}{dy}(h)L_f^\ell h)(x_a,x_b))\right|\\
&-|\Delta P_\ell(\boldsymbol{H}_\ell)(x_a,x_b)|.
\end{aligned}
$$

The first term can be decomposed this way

$$
\begin{aligned}
&\left|\Delta(\kappa^\ell(h)\frac{d\psi}{dy}(h)L_f^\ell h)(x_a,x_b)\right| \\
&\geq \left|\left(\kappa^\ell \frac{d\psi}{dy}\right)(h(x_a))\Delta L_f^\ell h(x_a,x_b)\right| \\
&\quad - \left|L_f^\ell(h(x_b))\kappa(h(x_a))^\ell \Delta\left(\frac{d\psi}{dy}\right)(h(x_a),h(x_b))\right| \\
&\quad - \left|L_f^\ell(h(x_b))\frac{d\psi}{dy}(h(x_b))\Delta\kappa^\ell(h(x_a),h(x_b))\right|.
\end{aligned}
$$

We want to lower bound the first term, and upper bound the other two. The output $h(x)$ is bounded for x in the compact $\mathcal{X}$, and as h and f are regular enough and x lies in a compact, $L_f^\ell h(x)$ is bounded too. Moreover, κ is C^m and $\frac{d\psi}{dy}$ is C^m, therefore they are bounded and Lipschitz on the compact set $h(\mathcal{X})$. It follows that κ^l is Lipschitz on the compact set $h(\mathcal{X})$. Moreover, using (A.8) and the definition of κ with Assumption 3, we get for all $(x_a,x_b)\in\mathcal{X}^2$

$$
\begin{aligned}
\left|\Delta(\kappa^\ell(h)\frac{d\psi}{dy}(h)L_f^\ell h)(x_a,x_b))\right| \geq \frac{\mu_0}{\beta}|\Delta L_f^\ell h(x_a,x_b))| \\
-\tilde{\rho}_\ell|\Delta h(x_a,x_b)|
\end{aligned}
$$

for some $\tilde{\rho}_\ell \geq 0$.

Moreover, P_ℓ is C^1 on the compact set $H_\ell(\mathcal{X})$, so we have

$$
\forall (x_a,x_b)\in\mathcal{X}^2, \quad |\Delta P_\ell(H_\ell)(x_a,x_b))| \leq k_{P_\ell}|\Delta H_\ell(x_a,x_b)|
$$

for some $k_{P_\ell}>0$. So we finally get

$$
\begin{aligned}
\forall (x_a,x_b)\in\mathcal{X}^2, \quad |\Delta\phi_\ell(x_a,x_b)| \geq \mu_\ell|\Delta L_f^\ell(h(x_a),h(x_b))| \\
-\rho_\ell\sum_{i=0}^{\ell-1}|\Delta L_f^i(h(x_a),h(x_b))|
\end{aligned}
$$

for some $\mu_\ell>0$ and $\rho_\ell\geq 0$.

To conclude the proof, we show the Lipschitz injectivity of $D\boldsymbol{\phi}$ with $D=\mathrm{diag}(d_0,d_1,\dots,d_{m-1})$ for some $d_i>0$ to be picked. We compute

$$
\begin{aligned}
&|\Delta(D\boldsymbol{\phi})(x_a,x_b)| = \sum_{\ell=0}^{m-1} d_\ell|\Delta\phi_\ell(x_a,x_b)| \\
&\geq \sum_{\ell=0}^{m-1} d_\ell\left(\mu_\ell|\Delta L_f^\ell h(x_a,x_b)| - \rho_\ell\sum_{i=0}^{\ell-1}|\Delta L_f^i h(x_a,x_b))|\right) \\
&= \sum_{\ell=0}^{m-1}(d_\ell\mu_\ell - \sum_{i=\ell+1}^{m-1} d_i\rho_i)|\Delta L_f^\ell h(x_a,x_b))|.
\end{aligned}
$$

We want to fix $d_\ell\mu_\ell - \sum_{i=\ell+1}^{m-1} d_i\rho_i = 1$ for all $\ell=0,\dots,m-1$. This is equivalent to

$$
\begin{pmatrix}
\mu_0 & -\rho_1 & \cdots & \cdots & -\rho_{m-1} \\
0 & \mu_1 & -\rho_2 & \cdots & -\rho_{m-1} \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & 0 & \ddots & \ddots & -\rho_{m-1} \\
0 & \cdots & \cdots & 0 & \mu_{m-1}
\end{pmatrix}
\begin{pmatrix} d_0 \\ \vdots \\ d_{m-1} \end{pmatrix}
= \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.
$$

Since all μ_i are non-zero, the matrix is invertible. Hence there exists a suitable choice of d_i. Assumption (2) of differentiable observability gives for any $(x_a,x_b)\in\mathcal{X}^2$

$$
|\Delta(D\boldsymbol{\phi})(x_a,x_b)| \geq |\Delta\boldsymbol{H}_m(x_a,x_b)| \geq \underline{k}_H|x_a-x_b|.
$$

Hence, we readily get that for all $(x_a,x_b)\in\mathcal{X}^2$,

$$
|\Delta\boldsymbol{\phi}(x_a,x_b)| \geq \frac{\underline{k}_H}{\max d_i}|x_a-x_b|.
$$