Nonlinear Observers

Laurent Praly

***Abstract*—Observers are objects delivering estimation of variables which cannot be directly measured. The access to such *hidden* variables is made possible by combining modeling and measurements. But this is bringing face to face real world and its abstraction with, as a result, the need for dealing with uncertainties and approximations leading to difficulties in implementation and convergence.**

***Index Terms*—Estimation, distinguishability, detectability.**

I. Observation problem and its solutions

A. The context

Observers are answers to the question of estimating, from observed/measured/empirical variables, denoted y, and delivered by sensors equipping a real world system, some "theoretical" variables, called hidden variables in this text, denoted z, which are involved in a mathematical model related to this system. The measured variables make what is called the a posteriori information on the hidden variables, whereas the model is part of the a priori information. Because a model cannot fit exactly a system, introduction of uncertainties is mandatory.

Typically this model describing the link between hidden and measured variables is made of three components:

- *a dynamic model* describes the dynamics/evolution[1]:

$$\dot{x}(t) = f(x(t), t, \delta^s(t)) \quad \text{resp.} \quad x_{k+1} = f_k(x_k, \delta^s_k) \ , \quad (1)$$

  where t, in the continuous case, or k, in the discrete case, is an evolution parameter, called time is this text, x is a state, assumed finite dimensional in this text, and δ^s represents the uncertainties in the state dynamics. Any possible known inputs is represented here by the time-dependence of f.

- *a sensor model* relates state and measured variables:

$$y(t) = h(x(t), t, \delta^m(t)) \qquad \text{resp.} \qquad y_k = h_k(x_k, \delta^m_k) \qquad (2)$$

  with δ^m representing the uncertainties in the measurements.

- a model which relates state and hidden variables:

$$z(t) = \digamma(x, t, \delta^h(t)) \qquad \text{resp.} \qquad z_k = \digamma_k(x_k, \delta^h_k) \ . \qquad (3)$$

  where again δ^h represents the uncertainties in the hidden variables.

In a deterministic setting, the a priori information on the uncertainties $(\delta^s, \delta^m, \delta^h)$ may be that the values of δ^s, δ^m and δ^h are unknown but belong to known sets Δ^s, Δ^m and Δ^h. Namely we have:

$$\begin{array}{ll} & \delta^s(t) \in \Delta^s(t) \ , \ \delta^m(t) \in \Delta^m(t) \ , \ \delta^h(t) \in \Delta^h(t) \ , \\ \text{respectively} & \delta^s_k \in \Delta^s_k \ , \ \delta^m_k \in \Delta^m_k \ , \ \delta^h_k \in \Delta^h_k \ . \end{array} \qquad (4)$$

In a stochastic setting and more specifically in a Bayesian approach, it may be that δ^s, δ^m and δ^h are unknown realization of stochastic processes for which we know the probability distributions.

Similarly we may also know a priori that we have:

$$\begin{array}{lll} & x(t) \in \mathcal{X}(t) \quad , & z(t) \in \mathcal{Z}(t) \\ \text{respectively} & x_k \in \mathcal{X}_k \quad , & z_k \in \mathcal{Z}_k \end{array} \qquad (5)$$

where the sets $\mathcal{X}$ and $\mathcal{Z}$ are known or we may have a priori probability distribution for x and z.

In this context, the a priori information is the data of the functions f, h and $\digamma$, of the sets Δ^s, Δ^m, Δ^h or the corresponding probability distribution and so maybe also of the sets $\mathcal{X}$ and $\mathcal{Z}$ or the corresponding a priori probability distribution.

In the next section, we state the observation problem and give the solutions which are direct consequences of the deterministic and stochastic setting given above. This will allow us to see that an observer is actually a dynamical system with the measurements as inputs and the estimate as output. But approximations in the implementation of these solutions, not knowing how to initialize, …may lead to convergence problems even when the uncertainties disappear. The second part of this text is devoted to this convergence topic.

To ease the presentation we deal only with the discrete time case in Section I-C, and the continuous time case in Section I-D and part II.

B. The observation problem

Let $X^{\delta^s}(x, t, s)$, respectively $X^{\delta^s}_l(x, k)$, denote a solution of (1) at time s, respectively l, going through x at time t, respectively k, and under the action of δ^s.

Observation problem: At each time t, respectively k, given the function $s \in]t - T, t] \mapsto y(s)$, respectively the sequence $l \in \{k - K, \dots, k\} \mapsto y_l$, find an estimation $\hat{z}(t)$, respectively $\hat{z}_k$, of $z(t)$, respectively z_k, satisfying:

$$\hat{z}(t) = \digamma(\hat{x}(t), t, \delta^h(t)) \qquad \textit{resp.} \qquad \hat{z}_k = \digamma_k(\hat{x}_k, \delta^h_k) \ .$$

where $\hat{x}(t)$, respectively $\hat{x}_k$, is to be found as a solution of:

$$\begin{array}{c} \hat{x}(t) \in \mathcal{X}(t) \ , \\ y(s) = h(X^{\delta^s}(\hat{x}(t), t, s), s, \delta^m(s)) \quad \forall s \in]t - T, t] \ , \end{array}$$

respectively

$$\begin{array}{c} \hat{x}_k \in \mathcal{X}_k \ , \\ y_l = h_l(X^{\delta^s}_l(\hat{x}_k, k), \delta^m_l) \qquad \forall l \in \{k - K, \dots, k\} \end{array}$$

and where the time functions δ^s, δ^m and δ^h must agree with the a priori (deterministic/stochastic) information or minimized in some way.

[1] $\dot{x}$ denotes the time derivative $\frac{dx}{dt}$.

In this statement T, respectively K, quantify the time window length or memory length during which we record the measurement. The accumulation with time of measurements, together with the model equations (1) to (3) and the assumptions on $(\delta^s, \delta^m, \delta^h)$, give a redundancy of data compared with the number of unknowns that the hidden variables are. This is why it may be possible to solve this observation problem.

To simplify the following presentation, we restrict our attention on the case[2] where the hidden variables are actually the full model state, i.e.

$$z = \mathcal{U}(x) = x \ .$$

C. Set valued and conditional probability valued observers

Conceptually the answer to this problem is easy at least when the memory increases with time ($\dot{T}(t) = 1$ resp. $K_{k+1} = K_k + 1$) leading to an infinite non fading memory. It consists in starting from all what the a priori information makes possible and to eliminate what is not consistent with the a posteriori information. In the set valued observer setting, in the discrete time case, this gives the following observer. To ease its reading, we underline the data given by the a priori information. It requires the introduction of two sets ξ_k, and $\xi_{k|k-1}$ which are updated at each time k when a new measurement y_k is made available. ξ_k is the set which x_k is guaranteed to belong to at time k, knowing all the measurements up to time k, and $\xi_{k|k-1}$ is the same but with measurements known up to time $k-1$.

Set valued observer:

Initialization: $\xi_0 = \underline{\mathcal{X}_0}$

At each time k:

prediction (flowing) $\xi_{k|k-1} = \underline{f_{k-1}}(\xi_{k-1}, \underline{\Delta^s_{k-1}})$

restriction (consistency) $\xi_k = \left\{ x \in \left(\xi_{k|k-1} \bigcap \underline{\mathcal{X}_k}\right) : y_k \in \underline{h_k}(x, \underline{\Delta^m_k}) \right\}$

estimation $\widehat{x}_k \in \xi_k$

A key feature here is that *this observer has a state* ξ_k – a set – *and is a dynamical system in the form:*

$$\xi_{k+1} = \varphi_k(\xi_k, y_k) \quad , \qquad \widehat{x}_k \in \xi_k$$

with y *as input and* $\widehat{x}$ *as output* which is not single valued. Important also, the initial condition of the state ξ is given by the a priori information.

In the stochastic setting, following the Bayesian paradigm, the observer has the same structure but with the state ξ_k being a conditional probability. See [17, Theorem 6.4] or [11, Table 2.1]. In that setting too the observer is not a single state; it is the (a posteriori) conditional probability of the random variable x_k given the a priori information and the sequence of measurements $l \in \{k-K, \dots, k\} \mapsto y_l$.

Comments

Implementation: For the time being, except for very specific cases (Kalman filter, …) the set valued and the conditional probability valued observers remain conceptual since we do not know how to manipulate numerically sets and probability laws. Their implementation requires approximations. For instance, see [21], [27] for the set case and [4], [10], [11], [17] for the conditional probability case.

Need of finite or infinite but fading memory: In these observers, model states x which are consistent with the a priori information but do not agree with the a posteriori information are eliminated (set intersection or probability product). But once a point is eliminated, this is for ever. As a consequence if there is, at some time, a misfit between a priori and a posteriori information, it is mistakenly propagated in future times. A way to round this problem is to keep the information memory finite or infinite but fading. In particular, with fixed length memory, consistent points which were disregarded due to measurements which are no more in the memory are reintroduced. This says also that observers should not be sensitive to their initial condition.

Not single valued estimate. The observers introduced above realize a lossless data compression with extracting and preserving all what concerns the hidden variables in the redundant data given by a priori and a posteriori information. But this "lossless compression" answer is not single valued (set valued or conditional probability valued) as a result of taking uncertainties into account. Actually, to get a single valued answer, *the observation problem must be complemented by making precise for what the estimation is made.* For instance we may want to select the most likely or the average or more generally some cost-minimizing estimate $\widehat{x}$ among all the possible ones given by ξ. In this way we obtain an observer giving a single valued estimate:

$$\xi_{k+1} = \varphi_k(\xi_k, y_k) \quad , \qquad \widehat{x}_k = \tau_k(\xi_k)$$

respectively

$$\dot{\xi}(t) = \varphi(\xi(t), y(t), t) \quad , \qquad \widehat{x}(t) = \tau(\xi(t), t) \qquad (6)$$

But then, in general, we loose information and in particular we have no idea on the confidence level this estimate has. Also, since the function τ, at least, encodes for what the estimate $\widehat{x}$ is used, for different uses, different functions τ may be needed.

D. An optimization approach

A short-cut to obtain directly an observer giving a single valued estimate is to design it by trading off among a priori and a posteriori information (see [13, pages 7-10], [1], …). For example, in the continuous time case, we can select the estimate $\widehat{x}(t)$ among the minimizers (in x) of:

$$\mathcal{C}(\{s \mapsto \delta^s(s)\}, x, t) = \int_{-\infty}^{t} C\left(\delta^s(s), y(s), X^{\delta^s}(x,t,s), s\right) ds$$

where $X^{\delta^s}(x,t,s)$ is still the notation for a solution to (1) and $\{s \mapsto \delta^s(s)\}$, representing the unmodelled effect on the dynamics, is among the arguments for the minimization. The infinitesimal cost C is chosen to take non negative values and be such that $C(0, h(x,s), x, s)$ is zero. For instance, it can be:

$$C(\delta^s, y, x, s) = \|\delta^s\|^2_x + d_y(y, h(x,s))^2$$

[2]When z differs from x, observers are called functional observers.

where $\|.\|_x$ is a norm at the point x and d_y is a distance in the measurement space. In the same spirit, instead of optimization, a minimax approach can be followed. See for instance [8], [6, Chapter 7], [26].

With x fixed, the minimization of $\mathcal{C}$ is an infinite horizon optimal control problem in reverse time. Solving on line this problem is extremely difficult and again approximations are needed. We do not go on with this approach, but we remark that, under extra assumptions, the observer we obtain following this approach can also be implemented in the form of a dynamical system (6) but with the specificity that *the estimate $\widehat{x}$ is part of the observer state ξ and its dynamics are a copy of the undisturbed model with a correction term which is zero when the estimated state reproduce the measurement.* Namely we get:

$$\dot{\widehat{x}}(t) = f(\widehat{x}(t), t, 0) + E\Big(\{\sigma \mapsto y(\sigma)\}, \widehat{x}(t), y(t), t\Big)$$

where E is zero when $h(\widehat{x}(t), t) = y(t)$. But, as opposed to what we saw in the previous section, the initial condition for the part $\hat{x}$ of the observer state is unknown. Hence we encounter again the need for the observer to forget its initial condition.

II. Convergent Observers

We have mentioned that often an observer can be implemented as a dynamical system, but without knowing necessarily how to initialize it. Also approximation is involved both in its design and its implementation. So, at least when it gives a single valued estimate, we are facing the problem of convergence of this estimate to the "true" value, at least when there is no uncertainties. We concentrate now our attention on the study of this convergence, but, to simplify, in the continuous time case only.

Let the model and observer dynamics be:

$$\dot{x}(t) = f(x(t), t) \quad , \qquad y(t) = h(x(t), t) \tag{7}$$

$$\dot{\xi} = \varphi(\xi(t), y(t), t) \quad , \qquad \widehat{x}(t) = \tau(\xi(t), y(t), t) \tag{8}$$

with the observer state ξ of finite dimension m. We denote by $(X(x,t,s), \Xi((x,\xi),t,s))$ a solution of (7)-(8).

Since we are dealing with convergence, the focus is on what is going on when the time becomes very large and in particular on the set Ω of model states which are accumulation points of some solution. Specifically we are interested in the stability properties of the set

$$\mathfrak{Z}(t) = \Big\{(x,\xi) : x \in \Omega \quad \& \quad x = \tau(\xi, h(x,t), t)\Big\}$$

which is contained in the zero estimation error set associated with the given model-observer pair.

Definition 1 (Convergent observer): We say the observer (8) is convergent if, for each t, there exists a set $\mathfrak{Z}_a(t) \subset \mathfrak{Z}(t)$, such that, on the domain of existence of the solution, a distance between the point $(X(x,t,s), \Xi((x,\xi),t,s))$ and the set $\mathfrak{Z}_a(s)$ is upperbounded by a real function $s \mapsto \beta^c_{x,\xi,t}(s)$, maybe dependent on (x,ξ,t), with non negative values, strictly decreasing and going to zero as s goes to infinity.

A. Necessary conditions for observer convergence

1) No restriction on τ:

It is possible to prove that, *if the observer is convergent then,*

Necessity of detectability: When h and τ are uniformly continuous in x and ξ respectively, the estimate $\widehat{x}$ does converge to the model state x. In this case, two solutions of the model (7) which produce the same measurement must converge to each other. This is an asymptotic distinguishability property called detectability.
If we are interested, not only in the asymptotic behavior, but also in the transient (as for output feedback) a property stronger than detectability is needed. In particular instantaneous distinguishability (see Section II-B2) is necessary if we want to be able to impose the decay rate of the function $\beta^c_{x,\xi,t}$.

Necessity of $m \geq n-p$: For each t, there exists a subset $\mathcal{X}_a(t)$ of Ω, supposed to collect the model states which can be asymptotically estimated, and such that we can associate, to each of its point x, a set $\tau^i(x,t)$ allowing us to redefine the set $\mathfrak{Z}_a(t)$ as:

$$\mathfrak{Z}_a(t) = \big\{(x,\xi) : x \in \mathcal{X}_a(t) \quad \& \quad \xi \in \tau^i(x,t)\big\} \ .$$

This implies that, for each t and each x in $\mathcal{X}_a(t)$, there is a point ξ satisfying:

$$x = \tau(\xi, h(x,t), t) \ . \tag{9}$$

This is a surjectivity property of the function τ but of a special kind since $h(x,t)$ is an argument of τ. We say that, for each t, *the function τ is surjective to $\mathcal{X}_a(t)$ given h.* In a "generic" situation this property requires the dimension m of the observer state ξ to be larger or equal to the dimension n of the model state x minus the dimension p of the measurement y.

2) τ is injective given h:

We consider now the case where the observer has been designed with a function τ which is injective given h, namely we have the following implication, when x is in $\mathcal{X}_a(t)$,

$$\Big[\tau(\xi_1, h(x,t), t) = \tau(\xi_2, h(x,t), t) \quad \& \quad \xi_1 \in \tau^i(x,t)\Big] \implies \xi_1 = \xi_2 \ .$$

In a "generic" situation, this property together with the surjectivity given h, implies that the dimension m of the observer state ξ should be between $n-p$ and n.

If a convergent observer has a such a function τ, then $(x,t) \mapsto \tau^i(x,t)$, which is (of course) a (single valued) function, admits a Lie derivative[3] $L_f\tau^i$ satisfying:

$$L_f\tau^i(x,t) = \varphi(\tau^i(x,t), h(x,t), t) \qquad \forall x \in \mathcal{X}_a(t) \tag{10}$$

This says (very approximatively) that φ is nothing but the image of the vector field f, under the change of coordinates $(x,t) \mapsto (\tau^i(x,t), t)$ but again all this given h. As partly obtained in the optimization approach, the observer dynamics are then a copy of the model dynamics with maybe a correction

[3] $L_f\tau^i(x,t) = \lim_{dt \to 0} \dfrac{\tau^i(X(x,t,t+dt), t+dt) - \tau^i(x,t)}{dt}$

term which is zero when the estimated state reproduce the measurement.

If moreover the functions h and τ are uniformly continuous in x and ξ respectively, then, given ξ_1 and ξ_2 a distance between $\Xi((x,\xi_1),t,s)$ and $\Xi((x,\xi_2),t,s)$ goes to zero as s goes to infinity. This property is related to what was called extreme stability (see [28]) in the 50's and 60's and is called incremental stability today (see [3]). It holds when, with denoting by $\Xi^y(\xi,t,s)$ the solution at time s of the observer dynamics :

$$\dot{\xi}(t) \;=\; \varphi(\xi(t),y(t),t)$$

going through ξ at time t and under the action of y, *the flow* $\xi \mapsto \Xi^y(\xi,t,s)$ *is a strict contraction*[4] *for each* $s>t$ *or, at least, if a distance between any two solutions* $\Xi^y(\xi_1,t,s)$ *and* $\Xi^y(\xi_2,t,s)$*, with the same input* y*, converges to* 0.

B. Sufficient conditions

Knowing now how a convergent observer should look like, we move to a quick description of some such observers.

1) Observers based on contraction: Since the flow generated by the observer should be a contraction, we may start its design by picking the function φ as:

$$\dot{\xi}(t) \;=\; \varphi(\xi(t),y(t),t) \;=\; A\,\xi(t) \;+\; B(y(t),t)$$

where A, not related to f, is a matrix whose eigen values have strictly negative real part. Under weak restriction, there exists a function τ^i satisfying (10), namely:

$$L_f\tau^i(x,t) \;=\; A\,\tau^i(x,t) \;+\; B(h(x,t),t)\ . \qquad (11)$$

To obtain a convergent observer it is then sufficient that there exists a (uniformly continuous) function τ satisfying:

$$x \;=\; \tau(\tau^i(x,t),h(x,t),t)$$

For this to be possible, the function τ^i should be injective given h. This injectivity holds when the observer state has dimension $m \geq 2(n+1)$, the model is distinguishable and provided the eigen values of A have a sufficiently negative real part and are not in a set of zero Lebesgue measure.

Unfortunately, we are facing again a possible difficulty in the implementation since an expression for a function τ^i satisfying (11) is needed and the function $\tau:(\xi,y,t)\mapsto \hat{x}(t)$ is known implicitly only as:

$$\xi \;=\; \tau^i(\hat{x}(t),t) \quad .$$

See [2], [20], [24].

2) Observers based on instantaneous distinguishability: Instantaneous distinguishability means that we can distinguish as quickly as we want two model states by looking at the paths of the measurements they generate. A sufficient condition to have this property can be obtained by looking at the Taylor expansion in s of $h(X(x,t,s),s)$. Indeed, we have:

$$h(X(x,t,s),s) \;=\; \sum_{i=0}^{m-1} h_i(x,t)\,\frac{(s-t)^i}{i!} \;+\; o((s-t)^{m-1})$$

[4]See [18] for a bibliography on contraction.

where h_i is a function obtained recursively as

$$h_0(x,t) \;=\; h(x,t)$$
$$h_{i+1}(x,t) \;=\; \dot{\overbrace{h_i(x,t)}} \;=\; \frac{\partial h_i}{\partial x}(x,t)f(x,t) \;+\; \frac{\partial h_i}{\partial t}(x,t)\ .$$

If there exists an integer m such that, in some uniform way with respect to t, the function

$$x \;\mapsto\; H_m(x,t) = (h_0(x,t)\,,\,\dots\,,\,h_{m-1}(x,t))$$

is injective then we do have instantaneous distinguishability. We say the system is differentially observable of order m when this injectivity property holds. When a system has such a property, the model state space has a very specific structure as discussed in [16, Section 1.9]. It means that we can reconstruct x from the knowledge of y and its $m-1$ first time derivatives, i. e. there exists a function Φ such that we have:

$$x \;=\; \Phi\left(H_m(x,t),t\right)\ .$$

This way, we are left with estimating the derivatives of y. This can be done as follows. With the notation $\eta_i \;=\; h_{i-1}(x,t)$, we obtain:

$$\dot{\eta}(t) \;=\; F\,\eta + G\,h_m\left(\Phi(\eta(t),t)\,,\,t\right)$$

where

$$F\,\eta = (\eta_2\,,\,\dots\,,\,\eta_m\,,\,0)\ ,\quad G \;=\; (0\,,\,\dots\,,\,0\,,\,1)\ .$$

When the last term on the right hand side is Lipschitz, we can find a convergent observer in the form:

$$\begin{aligned}\dot{\xi}(t) \;&=\; F\,\xi(t) + G\,h_m\left(\widehat{x}(t),t\right) + K(y(t)-\xi_1(t))\ ,\\ \widehat{x}(t) \;&=\; \tau\left(\xi(t),t\right)\ ,\end{aligned}$$

with ξ being actually an estimation of η and where K is a constant matrix and τ is a modified version of Φ keeping the estimated state in its a priori given set $\mathcal{X}(t)$.

This is the high gain observer paradigm. See [15], [25]. The implementation difficulty is in the function $\widehat{\Phi}$, not to mention sensitivity to measurement uncertainty.

3) Observers with τ bijective given h:

a) Case where τ is the identity function: A convergent observer whose function τ is the identity has the following form:

$$\dot{\xi} \;=\; f(\xi,t) \;+\; E\big(\{\sigma\mapsto y(\sigma)\},\xi(t),y(t),t\big)\ ,\ \widehat{x}(t) \;=\; \xi(t)\ . \qquad (12)$$

The only piece remaining to be designed is the correction term E. It has to ensure convergence and may be also other properties like symmetry preserving (see [9]).

For this design, a first step is to exhibit some specific properties of the vector field f by writing it in some appropriate coordinates. For example, there may exist coordinates such that the expression of f takes the form $\mathfrak{f}(x(t),h(x,t),t)$ and the corresponding observer (12) is such that there exists a positive definite matrix P for which the function $s \mapsto (X(x,t,d)-\widehat{X}((x,\widehat{x}),t,s))'P(X(x,t,d)-\widehat{X}((x,\widehat{x}),t,s))$ is strictly decaying (if not zero). A necessary condition for this to be possible is that $\mathfrak{f}$ is monotonic tangentially to the level sets of the function h, i.e. for all (x, y, v, t) satisfying $y = h(x, t)$ and $\frac{\partial h}{\partial x}(x, t)v = 0$, we have:

$$v^T P \frac{\partial \mathfrak{f}}{\partial x}(x, y, t)\, v \;\leq\; 0 \;. \tag{13}$$

This is another way of expressing a detectability condition. This expression is a coordinate dependent. Hence the importance of choosing the coordinates properly.

When this condition is strict and uniform in t, it is sufficient to get a locally convergent observer and even a non local one when h is linear in x, i.e. $h(x, t) = H(t)x$, again a coordinate dependent condition. In this latter case the observer takes the form:

$$\begin{aligned} \dot{\xi}(t) &= \mathfrak{f}(\xi(t), y(t), t) \\ &\qquad + \ell(\xi(t))\, P^{-1} H(t)^T [y(t) - H(t)\xi(t)] \;, \\ \widehat{x}(t) &= \xi(t) \;, \end{aligned}$$

where ℓ is a real function to be chosen with sufficiently large values. If (13) is strict and uniform and holds for all v, the correction term is not needed.

There are many other results of this type, exploiting one or the other specificity of the dependence on x of the function $\mathfrak{f}$ – monotonicity, convexity, $\ldots$. See [14], [19], [22], [23], $\ldots$

b) Case where $(x, t) \mapsto (\tau^i(x, t), h(x, t), t)$ is a diffeomorphism:
At each time t we know already that the model state x we want to estimate satisfy $y(t) = h(x, t)$. So, as remarked in [20], when $(h(x, t), t)$ can be used as part of coordinates for (x, t), we need to estimate the remaining part only. This can be done if we find a function τ^i, whose values are $n - p$ dimensional, such that $(x, t) \mapsto (y, \eta, t) = (h(x, t), \tau^i(x, t), t)$ is a diffeomorphism and the flow $\eta \mapsto \eta^y(\eta, t, s)$ generated by

$$\begin{aligned} \dot{\eta}(t) &= \frac{\partial \tau^i}{\partial x}(x(t), t) f(x(t), t) \;+\; \frac{\partial \tau^i}{\partial t}(x(t), t) \;, \\ &= \varphi(\eta(t), y(t), t) \end{aligned}$$

is a strict contraction for all $s > t$. Indeed in this case the observer dynamics can be chosen as:

$$\dot{\xi}(t) \;=\; \varphi(\xi(t), y(t), t)$$

and the estimate $\widehat{x}(t)$ is obtained as solution of:

$$\tau^i(\widehat{x}(t), t) \;=\; \xi(t) \quad , \qquad h(\widehat{x}(t), t) \;=\; y(t) \;.$$

This is the reduced order observer paradigm. See for instance [7, Proposition 3.2], [12], [20, Theorem 4].

References

[1] Alamir M. (2007) Nonlinear Moving Horizon Observers: Theory and Real-Time Implementation. In Nonlinear Observers and Applications. Lecture Notes in Control and Information Sciences, Springer-Verlag.

[2] Andrieu V., Praly L. (2006) On the existence of Kazantzis-Kravaris / Luenberger Observers. SIAM Journal on Control and Optimization, 45(2):432456.

[3] Angeli D. (2002) A Lyapunov approach t incremental stability properties. IEEE Transactions on Automatic Control, Vol. 47, No. 3, 410421.

[4] Arulampalam M., Maskell S., Gordon N., Clapp T. (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking IEEE Transactions on Signal Processing, Vol. 50 , No. 2, P 174 - 188.

[5] Bain A., Crisan D. (2009) Fundamentals of stochastic filtering. Vol.60, Stochastic modelling and applied probability. Springer.

[6] Başar T., Bernhard P. (1995) H^∞ optimal control and related minimax design problems: a dynamic game approach. Revised 2nd edition Birkäuser, Boston, MA.

[7] Besançon G. (2000) Remarks on nonlinear adaptive observer design. Systems & Control Letters, Vol. 41, No. 4, 271-280.

[8] Bertsekas D., I. Rhodes I.B. (1971) On the minimax reachability of target sets and target tubes. Automatica, 7: 233-213.

[9] Bonnabel S., Martin P., Rouchon P. (2008) Symmetry-Preserving Observers IEEE Transactions on Automatic Control, Vol. 53, No. 11, 2514-2526.

[10] Bucy R., Joseph P. (1987) Filtering for stochastic processes with applications to guidance. Second edition. Chelsea publishing company.

[11] Candy J. (2009) Bayesian signal processing: Classical, modern, and particle filtering methods. Wiley series in Adaptive learning systems for signal processing, communications and control. John Wiley & Sons.

[12] Carnevale D., Karagiannis D., Astolfi A. (2008) Invariant manifold based reduced-order observer design for nonlinear systems, IEEE Transactions on Automatic Control, Vol. 53, No. 11, 26022614.

[13] Cox H. (1964) On the estimation of state variables and parameters for noisy dynamic systems. IEEE Transactions on Automatic Control, Vol.9, No. 1, 5 - 12.

[14] Fan X., M. Arcak M. (2003) Observer design for systems with multivariable monotone nonlinearities. Systems & Control Letters 50, 319 330

[15] Gauthier J.-P., Kupka I. (2001) Deterministic observation theory and applications. Cambridge University Press.

[16] Isidori A. (1995) Nonlinear Control Systems- third edition, Springer-Verlag.

[17] Jazwinski A. (2007) Stochastic processes and filtering theory. Dover Publications.

[18] Jouffroy J., (2005) Some ancestors of contraction analysis. Proceedings of the IEEE Conference on Decision and Control, pages 5450-5455.

[19] Krener A., A. Isidori (1983) Linearization by output injection and nonlinear observer. Systems & Control Letters, Vol. 3, No. 1, 47-52.

[20] Luenberger D. (1964) Observing the state of a linear system. IEEE Transactions on Military Electronics, MIL-8:7480.

[21] Milanese M. , Norton J. , Piet-Lahanier H., Walter E., editors (1996) Bounding Approaches to System Identification. Plenum Press, New York.

[22] Respondek W. , Pogromsky A., Nijmeijer H. (2004) Time scaling for observer design with linearizable error dynamics Automatica 40, 277 285

[23] Sanfelice R., Praly L. (2012) Convergence of nonlinear observers on with a riemannian metric (Part I), IEEE Transactions on Automatic Control, Vol. 57, No. 7, 1709-1722.

[24] Shoshitaishvili A. (1990) Singularities for projections of integral manifolds with applications to control and observation problems. in Advances in Soviet Mathematics, Volume 1: Theory of singularities and its applications, V.I. Arnol'd, Editor.

[25] Tornambe A. (1988) Use of asymptotic observers having high gains in the state and parameter estimation. Proceedings of the IEEE Conference on Decision and Control, pages 1791–1794.

[26] Willems J.C., (2004) Deterministic least squares Filtering. Journal of Econometrics 118 341373.

[27] Witsenhausen H. (1966) Minimax control of uncertain systems, Elec. Syst. Lab.. M.I.T.. Rep. ESL-R-269M. Cambridge, May.

[28] Yoshizawa T. (1966) Stability theory by Lyapunov's second method. The mathematical Society of Japan.