# OBSERVATION AND IDENTIFICATION TOOLS FOR NONLINEAR SYSTEMS. APPLICATION TO A FLUID CATALYTIC CRACKER.

ERIC BUSVELLE, JEAN-PAUL GAUTHIER

Abstract. In this paper, we recall general methodologies we developed for observation and identification in nonlinear systems theory, and we show how they can be applied to real practical problems.

In a previous paper, we introduced a filter which is intermediate between the extended Kalman filter in its standard version and its high-gain version, and we applied it to certain observation problems. But we were missing some important cases. Here, we show how to treat these cases.

We also apply the same technique in the context of our identifiability theory.

As non academic illustrations, we treat a problem of observation and a problem of identification, for a fluid catalytic cracker (FCC). This FCC unit is one of the most crucial from the economic point of view, in petroleum industry.

Submitted for publication to **The International Journal of Control**

Authors' address: LE2I, UMR CNRS 5158, Université de Bourgogne, Aile des Sciences de l'Ingénieur BP 47870 - 21078 Dijon Cedex, France
Phone + 33 (0)3-80-39-58-38
Fax + 33 (0)3-80-39-58-90
e-mail: busvelle@u-bourgogne.fr, gauthier@u-bourgogne.fr

## 1. Introduction

1.1. **The observation and identification problems:** In this paper, we address the problems of observation and identification of general nonlinear control systems.

By observation, we mean reconstruction of the state trajectory of the system, on the basis of some "observed data", produced by output functions. Roughly speaking, we say that a system is observable if this reconstruction is possible. A device realizing the observation task is called an "observer". Usually, these devices are realized under the guise of a differential system, fed by the observed data.

The problem of identification is a bit different: very often, practical control systems depend on some functions, (with physical meaning), that are not well known, and that have to be determined on the basis of experiments.

If $x$ denotes the state of the system, if $x \mapsto \varphi(x)$ is the unknown function, and $y(t)$ is the observed data at time $t$, the identification problem is the problem of reconstructing the piece of the graph of $\varphi(\cdot)$, visited during the experiment. That is, for an experiment of duration $T$, we want to determine the couples $(x(t), \varphi(x(t)))$, for all $t \in [0, T]$, using only the observed data $\{y(t), t \in [0, T]\}$. We say that a system is identifiable if this is possible, whatever the experiment.

*Date*: September 20, 2004.

An identifier is a device performing this task. We will be interested with "on-line identifiers" only, i.e. identifiers that estimate the graph of $\varphi$ simultaneously to the experiment.

The two problems, of observation and identification, are of course strongly connected.

### 1.2. **Our previous results.**

1.2.1. *Observability and observers.* In the book [8], a general observability theory has been exposed, together with a methodology for constructing observer systems. (These observers we construct are called "high gain" observers).

Otherwise, there is a practical tool, used for long by engineers, to construct observers for nonlinear systems (in a stochastic context): the extended Kalman filter (EKF). The idea is just to use the classical equations of the linear Kalman filter, and to apply them to the linearization of the system along the **estimate** trajectory. This is not a well defined procedure (since we linearize along the trajectory we are just estimating, and not along the real trajectory). Nevertheless, and despite a lack of theoretical justification, the EKF gives often very good results.

One version of our high-gain construction in [8] is connected with the EKF.

In the paper [5], we propose an observer system that has the advantages of both approaches: in presence of big disturbances, or "state jumps", it has the good properties of the high gain observers, to recover the state of the system arbitrarily fast. On the contrary, when the estimation error is small, it behaves exactly as the EKF (with good performances in front of noise).

1.2.2. *Identifiability and identifiers.* In the paper [4], we establish the main results of an identification theory and we propose an observer-based strategy for identification.

A remarkable (but not surprising) fact is that the identification theory is perfectly parallel to the observation theory. To compare, let us consider, **for the observation problem, single input systems** only, and **for the identification problem, the case of a single unknown function** to be identified, and **no control**. Then, for observability:

-If the number of outputs is two or more, systems are generically observable.

-If there is only one output, then, observability is a nongeneric property, so strong that it can be characterized by a very rigid normal form.

For identifiability:

-If the number of outputs is three ore more, then, identifiability is a generic property,

-If there is only one or two outputs, then, identifiability is a nongeneric property, so strong that it can be characterized by 4 very rigid normal forms.

1.3. **Purpose of the paper.** Our purpose here is twofold:

First, we want to recall, compare and summarize the main results of both theories of observability and identifiability. We want also to present our observers, and specially the final uppermentionned observer, mixing the high-gain construction with the EKF.

Therefore, we will summarize the results of [8], [4], [5].

At the same time, we will give some improvement of our method, allowing to apply it to certain important cases that we missed in our previous works.

Second, we want to show, through a practical nonacademic application, that these theories are really applicable in practice, and give very good results, both for the observation and the identification problem.

1.4. **The application.** The process we consider is a FCC unit, i.e. a Fluid Catalytic Cracker process, used in petroleum industry to convert heavy petroleum residues into gasoline.

There is at least one such FCC in any refinery, and it is a very strategic process, from the economic point of view (may be the most one).

It is a highly nonlinear process, rather hard to control.

As usual in petroleum industry, a few measurements are available (mostly temperature, pressure and flowrate measurements).

We will apply our observation methodology to recover the state of the system on the basis of these measurements, and specially, our purpose is to reconstruct a crucial variable inside the system: the Carbon Conradson factor.

Also, the FCC model depends on a certain "oxygen reaction rate" function, which is very important, and is in practice not well known. Also, due to the degree of simplification of the model under consideration, even if this "oxygen reaction rate" function is known in theory, it has to be adapted.

We will apply our identification procedure to estimate this function.

1.5. **Organization of the paper.** First, in the remaining of this introduction, we will fix precisely the systems under consideration, the classes of controls we consider, and we will define a few notions that are crucial for our work.

Section 2 will be devoted to the observability theory: we state the definitions of the various notions of observability we use, and we give a summary of the main observability results.

In Section 3, we do the same for our identification theory.

In Section 4, we present our results on the construction of observers, and on practical observer-based identification. We also present certain important simple improvement of our methodology.

Section 5 is devoted to the application to the Fluid Catalytic Cracker unit.

1.6. **Prerequisites.** Here, the basic terminology is the standard one in "geometric nonlinear control theory", as may be found in one of the best–sellers for engineers available today. The most understandable and reasonable one is [11] in our opinion.

As in this book, the mathematical background concerns differential calculus, differential geometry up to integrability of distributions. Words as "immersions", "embeddings", "tangent bundle", ... belong to this basic vocabulary.

Besides this, we use some terminology from differential topology, semi–algebraic and subanalytic sets, stratifications, transversality theory, jet spaces. Words as "Whitney topology" are pronounced.

In general, we explain inside the paper the minimum needed for intuitive understanding of the concepts and results. But in the paper, almost nothing is proven. Therefore, we hope that this contribution will be accessible to people with low mathematical background.

If one is interested in detailed proofs, or more precise statements, he should consult the papers [4, 5], and the books [1, 8]. The appendix in [8] is specially recommended. The book [10] is also a useful reasonable mathematical reference.

### 1.7. **Notations, systems under consideration.**

1.7.1. *Conventions.* All along the paper, for a smooth real-valued function $(x, y) \mapsto h(x, y)$, or for a smooth mapping $(x, y) \mapsto f(x, y)$, the notation $d_x h$ (resp. $D_x f$) means the differential of $h$ (resp. the tangent mapping to $f$) w.r.t. the variable $x$ only. In coordinates, they are represented by the Jacobian matrices of $h, f$ w.r.t. $x$ only.

Also, all along the paper, the notation $A'$ means the transpose of the matrix $A$.

1.7.2. *Systems.* We will consider general finite dimensional nonlinear controlled systems, of the form

$$(\Sigma^1) \left\{ \begin{array}{l} \dot{x} = f(x, u), \\ y = h(x, u), \end{array} \right. \tag{1.1}$$

or,

$$(\Sigma^2) \left\{ \begin{array}{l} \dot{x} = f(x, u, \varphi(x)), \\ y = h(x, u, \varphi(x)), \end{array} \right. \tag{1.2}$$

where the state $x \in \mathbb{R}^n$, or more generally to a $n$-dimensional analytic differentiable manifold $X$, where $u$ denotes the control variable , $u \in U$, some "regular" compact subset of $\mathbb{R}^p$ of dimension $p$ (precisely subanalytic, a polyhedron for instance) with nonempty interior, and $y \in \mathbb{R}^m$ is the output variable. In $\Sigma^1$ (resp. $\Sigma^2$), $f$ is a $u$–parametrized (resp. $(u, \varphi)$–parametrized) smooth vector field and the observation mapping $h$ is a smooth mapping $X \times U \to \mathbb{R}^{d_y}$ (resp. $X \times U \times \mathbb{I} \to \mathbb{R}^{d_y}$). The function $\varphi$ in $\Sigma^2$ is an unknown function to be identified. In this paper, we will restrict to the case where it is $\mathbb{I}$-valued, where $\mathbb{I}$ is a compact subinterval of $\mathbb{R}$ (i.e., we consider only the case of a single function to be identified) but the theory we developed in [4], clearly has extensions to higher dimension.

In practice, very often, this function $\varphi$ represents some "physical characteristic" inside the system, that has to be determined on the basis of experiments. It may happen that $\varphi$ does not depend on the whole state $x$ of the system, but only on some projection $\pi(x)$ ($\pi : X \to Z$ a known fixed smooth function).

**Remark 1.** *Usually, output functions do not depend on the controls. It might seem only a (practically void) mathematical assumption to consider this dependence. Unfortunately, assuming the non-dependence would lead to more complicated unnatural statements, for many of our results below.*

1.7.3. *Topologies.* The sets of systems $S^1 = \{\Sigma^1 = (f, h)\}$, $S^2 = \{\Sigma^2 = (f, h)\}$, are given the $C^\infty$ Whitney topology, (which is the relevant topology for our considerations). Basic neighborhoods of a system $\Sigma = (f, h)$ in the $C^j$ Whitney topology are determined by the data of functions $\varepsilon(z) > 0$, and are formed by the systems $\Sigma' = (f', h')$ such that all the partial derivatives, w.r.t. all variables, up to order $j$ of $(f' - f, h' - h)$, have norm at $z = (x, u)$ (resp. $z = (x, u, \varphi)$) smaller than $\varepsilon(z)$. The $C^\infty$ Whitney topology is generated by these open sets for all $j$. Because of the fact that $\varepsilon$ depends on $z$, this topology "controls" the behavior at infinity of the systems. The counterpart of this fact is that (unless $X \times U \times \mathbb{I}$ is compact), it is not metrizable. Nevertheless, it has the nice "Baire property" that a countable intersection of open dense subsets is still dense. A subset of $S^1, S^2$ is called **residual** if it is such a countable intersection of open dense subsets. A subset is said generic if it contains a residual subset. (Contrarily to the way it sounds, "residual" means very big).

1.7.4. *Controls and outputs.* Control and output functions $u(\cdot)$, $y(\cdot)$ of systems $\Sigma^1$ will be defined on semi-open intervals $[0, T_{u(\cdot)}[$ depending on the control. May be $T_{u(\cdot)} = +\infty$. Controls $u(\cdot)$ and outputs $y(\cdot)$ are measurable functions, bounded on any compact subinterval $[0, T] \subset [0, T_{u(\cdot)}[$. The space of such functions is denoted by $L^\infty(U)$, (resp. $L^\infty(\mathbb{R}^m)$)[1].

For systems $\Sigma^2 = (f(x, u, \varphi), h(x, u, \varphi)) \in S^2$, it will be convenient to consider $\varphi$ not as a function of $x$, but as an extra input function $\varphi(t)$ of the time. Then, we will often consider in (1.2) that $\varphi = \varphi(t)$, $\varphi(\cdot) \in L^\infty(\mathbb{I})$.

1.7.5. *State-output, input-state-output mappings and their "first variations".* For a system $\Sigma \in S^1$, the control $u(\cdot)$ being fixed, we may consider the **state-output mapping** $PX_{\Sigma,u}$, which to the initial condition $x_0$ associates the output trajectory $y(\cdot) \in L^\infty(\mathbb{R}^m)$.

$$PX_{\Sigma,u} : x_0 \mapsto y(\cdot).$$

If $u(\cdot) \in L^\infty[0, T_u[$, it may happen that the output is only defined on $[0, T_y[$, $T_y < T_u$, i.e. up to the "**explosion time**" $T_y$. In that case, $\lim_{t \to T_y} x(t) = \infty$, (obviously, here, $x(\cdot)$ is the state trajectory corresponding to initial condition $x_0$ and control $u(\cdot)$). For all $T_0 < T_y$, there is a neighborhood $V_{x_0} \times V_u \subset X \times L^\infty[0, T_0]$, such that for all $(x_0, u(\cdot)) \in V_{x_0} \times V_u$, the corresponding trajectories $t \mapsto x(t), y(t)$ are well defined on $[0, T_0]$.

Moreover, the mapping $PX_{\Sigma,u}$ (defined on a neighborhood of $x_0$ as we said) is differentiable with respect to $x_0$. Let $TPX_{\Sigma,u}|x_0$ be this differential (it is well defined as a linear mapping from $T_{x_0}X$ to $L^\infty([0, T_0], \mathbb{R}^m)$, $T_0 < T_y$).

In fact, this differential $TPX_{\Sigma,u}|x_0$ is also the state-output mapping of another system $TX_\Sigma$ , called the **first (state) variation** of $\Sigma$, with state space $TX$, the tangent bundle of $X$ (or on $\mathbb{R}^n \times \mathbb{R}^n$)):

$$(TX_\Sigma) \begin{cases} \dot{x} = f(x, u), \\ \dot{\xi} = D_x f(x, u)\xi, \\ \hat{y} = d_x h(x, u)\xi. \end{cases} \tag{1.3}$$

Here, $(x, \xi) \in TX$ (or $\mathbb{R}^n \times \mathbb{R}^n$) is the state of $TX_\Sigma$.

[1] In fact, a $y(\cdot) \in L^\infty(\mathbb{R}^m)$ is an element of the usual space $L^\infty_{loc}([0, T_y[, \mathbb{R}^m)$ but the interval $[0, T_y[$ depends on the element $y(\cdot)$. The same applies later to elements $\eta(\cdot)$ of $L^\infty(\mathbb{R})$.

Let $PTX_{\Sigma,u}$ denote the state-output map of $TX_\Sigma$. Then, the linear mapping $\xi \mapsto PTX_{\Sigma,u}(x_0,\xi)$ is the differential at $x_0$, $TPX_{\Sigma,u}|x_0$, of the state-output mapping .

Now, for the purpose of explaining the main features of our identification theory, we will define the **input-state-output mapping**, and the first (input-state-output) variation of $\Sigma^2$ in a similar way.

In fact, in our case, for systems $\Sigma^2 \in S^2$, we will be led to consider $\varphi$ as an extra control function, as we said, the (eventual) usual control $u(\cdot)$ remaining fixed.

Then, the input-state-output mapping is the mapping $PX\mathbb{I}_{\Sigma,u}$ (or $PX\mathbb{I}_\Sigma$ if there is no control $u$)

$$PX\mathbb{I}_{\Sigma,u} : (x_0, \varphi(\cdot)) \mapsto y(\cdot),$$

which to the initial state $x_0$ and the extra control function $\varphi(\cdot)$ associates the output function.

Assume that $PX\mathbb{I}_{\Sigma,u} : D \subset X \times L^\infty(\mathbb{I}) \to L^\infty(\mathbb{R}^m)$, is defined at a point $(x_0,\varphi_0)$ on the time interval $[0, T_y[$. Then, for all $T < T_y$, it is defined on an open neighborhood of $(x_0,\varphi_0)$ in $X \times L^\infty([0,T],\mathbb{I})$, and it is differentiable in the Frechet sense at $(x_0,\varphi_0)$ on $X \times L^\infty([0,T],\mathbb{I})$. The differential is denoted by $T_{(x_0,\varphi_0)}PX\mathbb{I}_{\Sigma,u}$.

The **first (input-state-output) variation** of $\Sigma^2$ is the system $TX\mathbb{I}_{\Sigma,u}$,

$$(TX\mathbb{I}_{\Sigma,u}) : \begin{cases} \dot{x} = f(x,u,\varphi_0), \\ \dot{\xi} = D_x f(x,u,\varphi_0)\xi + D_\varphi f(x,u,\varphi_0)\eta \\ \hat{y} = d_x h(x,u,\varphi_0)\xi + d_\varphi h(x,u,\varphi_0)\eta, \end{cases} \tag{1.4}$$

with (variational) control $\eta \in L^\infty(\mathbb{R})$.

If we take initial conditions $(x_0,\xi_0) \in TX$, and control functions $\varphi_0(\cdot)$, $u(\cdot)$ as above, then, the input-state-output mapping $PTX\mathbb{I}_{\Sigma,u}$ of $TX\mathbb{I}_{\Sigma,u}$ is the mapping:

$$(\xi_0, \eta(\cdot)) \mapsto \hat{y}(\cdot),$$
$$T_{x_0}X \times L^\infty[\mathbb{R}] \to L^\infty[\mathbb{R}^m].$$

This mapping also coincides (on the small enough finite time intervals $[0,T]$ considered above) with the tangent mapping $T_{(x_0,\varphi_0)}PX\mathbb{I}_{\Sigma,u}$.

1.7.6. *$k$-jet extensions of state-output and input-state-output mappings.* Let us consider $k$-jets $j^k\hat{\varphi}$, $j^k\hat{u}$, of smooth functions $\hat{\varphi}$, $\hat{u}$ at $t = 0$,

$$\hat{\varphi} : [0,\varepsilon[ \to \mathbb{I},\ \hat{u} : [0,\varepsilon[ \to U,$$
$$j^k\hat{\varphi} = (\hat{\varphi}(0), \hat{\varphi}'(0), ..., \hat{\varphi}^{(k-1)}(0)), \quad j^k\hat{u} = (\hat{u}(0), \hat{u}'(0), ..., \hat{u}^{(k-1)}(0)).$$

Then, for any $x_0 \in X$, the corresponding $k$-jet $j^k\hat{y} = (\hat{y}(0), \hat{y}'(0), ..., \hat{y}^{(k-1)}(0))$ is well defined, in such a way that the (extension to $k$-jets) mappings

$$\Phi_k^{\Sigma^1} : (x_0, j^k\hat{u}) \mapsto j^k\hat{y};\quad X \times U \times \mathbb{R}^{(k-1)p} \to \mathbb{R}^{km},$$
$$\Phi_k^{\Sigma^2} : (x_0, j^k\hat{u}, j^k\hat{\varphi}) \mapsto j^k\hat{y};\quad X \times (U \times \mathbb{R}^{(k-1)p}) \times (\mathbb{I} \times \mathbb{R}^{(k-1)}) \to \mathbb{R}^{km},$$

are continuous. We call these mappings $\Phi_k^{\Sigma^1}$, $\Phi_k^{\Sigma^2}$ the **$k$-jets state-output mappings, (resp. $k$-jets input-state output mappings)** associated to $\Sigma^1 \in S^1$ (resp. $\Sigma^2 \in S^2$).

## 2. Observation theory

2.1. **Definitions .** In this section, the relevant set of systems is $S^1$, i.e. there is no unknown function $\varphi$.

We summarize the main observability results of the observation theory developed in [8].

We are not very precise in definitions and results with the explosion times, intervals of definitions of input and output functions, but everything is natural, and details can be found in [8].

**Definition 1.** *The system $\Sigma^1 = (f_1, h_1)$ is said uniformly **observable**, or just **observable**, w.r.t. a certain class $\mathcal{C}$ of inputs ($L^\infty(U)$ in most cases) if, for each $u(\cdot) \in \mathcal{C}$, the state output mapping $PX_{\Sigma,u}$ is injective.*

Observability means that we can reconstruct the complete information about the system (i.e. the full state trajectory $x(\cdot)$), from the knowledge of the input-output data $(u(\cdot), y(\cdot))$.

Injectivity is not a very tractable property, since it is not stable (even for standard mappings between finite dimensional spaces -example: $x \mapsto x^3, \mathbb{R} \to \mathbb{R}$ which may lose injectivity under perturbation by arbitrary small functions) Therefore, in order to state results, we need a few other definitions.

Notice also that, the bigger the class $\mathcal{C}$, the more restrictive observability property. For example, $L^\infty(U)$-observability is very strong, and implies observability in the $C^k$ class, $k = 0, .., \infty, \omega$. A major property of $C^\omega$ (analytic) systems is the following, that expresses that in fact, no matter the class:

**Theorem 1.** [8, page 56] *For $C^\omega$ systems, $C^\omega$ observability implies $L^\infty$ observability.*

This theorem is in fact very hard to prove.

For usual smooth mappings between finite dimensional spaces, a way to make the injectivity property stable is to add the requirement of infinitesimal injectivity (i.e. injectivity of all the tangent mappings). This is done for the study of differential mappings in differential topology. In the same spirit, let us define uniform infinitesimal observability.

**Definition 2.** *System $\Sigma^1$ is said uniformly infinitesimally observable if, for each $u(\cdot) \in L^\infty(U)$, each $x_0 \in X$, all the tangent mappings $TPX_{\Sigma^1,u}|x_0$ are injective. By Section 1.7.5, it is equivalent to require that the state-output mappings $PTX_{\Sigma^1,u}$ of the first variation of $\Sigma^1$ are injective.*

Another way to be more effective is to "pass to $k$-jets":

**Definition 3.** *System $\Sigma^1$ is said differentially observable (of order $k$) if for all $j^k\hat{u}$, the extension to $k$-jets mapping $\Phi_k^{\Sigma^1} : x_0 \mapsto j^k\hat{y}$; $X \to \mathbb{R}^{km}$ is injective.*

Again, this definition will become more effective if one adds an "infinitesimal injectivity" requirement:

**Definition 4.** *System $\Sigma^1$ is said strongly differentially observable (of order $k$) if for all $j^k\hat{u}$, the extension to $k$-jets mapping $\Phi_{k,j^k\hat{u}}^{\Sigma^1} : x_0 \mapsto j^k\hat{y}$; $X \to \mathbb{R}^{km}$ is an injective immersion (immersion means that all the tangent mappings $T_{x_0}\Phi_{k,j^k\hat{u}}^{\Sigma^1}$ to this map, have full rank $n$ at each point).*

Clearly, strong differential observability implies differential observability, which implies observability for the $C^\infty$ class, which -for analytic systems- implies $L^\infty$-observability by Theorem 1.

It is also a consequence of the theory that -for analytic systems- uniform infinitesimal observability implies observability of the restrictions of $\Sigma^1$ to small open subsets of $X$, the union of which is dense in $X$ (but this is a priori non-obvious).

2.2. **The generic case.** We consider here the case where $m > p$ (i.e. the number of outputs is strictly larger than the number of inputs). Then in that case, strong differential observability is generic:

**Theorem 2.** [8, page 40]. *a). The set of systems that are strongly differentially observable of order $2n+1$ is residual in $S^1$;*

*b) The set of **analytic** strongly differentially observable systems (of order $2n+1$) that are moreover $L^\infty$-observable is dense in $S^1$.*

For people who know about these topics, it could seem that b) is a consequence of a) and of Theorem 1, using some general result of "approximation of smooth by analytic". It is not at all the case, and this part b) is difficult in itself.

This theorem has a nice consequence:

**Theorem 3.** *The following is a generic (residual) property on $S^1$ : Set $k = 2n+1$.*

*For all sufficiently smooth $u(\cdot)$, set $j^k u(t) = (u(t), \dot{u}(t), ..., u^{(k-1)}(t))$. Chose an arbitrarily large relatively compact open subset $\Gamma$ of $X$, and an arbitrary bound on $u, \dot{u}, ..., u^{(k)}$, the control and its first $k$ derivatives. Then the mappings $\Phi^{\Sigma^1}_{k,j^k u}$ : $x(t) \mapsto (y, \dot{y}(t), ..., y^{(k-1)}(t))$ are smooth injective immersions that map the trajectories of the system $\Sigma^1$ (restricted to $\Gamma$) to the trajectories of the following system:*

$$
\begin{aligned}
y &= z_1, \\
\dot{z}_1 &= z_2, \\
&\cdot \\
&\cdot \\
\dot{z}_{k-1} &= z_k, \\
\dot{z}_k &= \varphi_K(z_1, ..., z_k, u, \dot{u}, ..., u^{(k)}).
\end{aligned}
\tag{2.1}
$$

A system under the form (2.1) is called a "phase-variable representation" (of order $k$). It means that, in restriction to a compact subset $K$ of $X$,

$$y^{(k)} = \varphi_K(y, ..., y^{(k-1)}, u, \dot{u}, ..., u^{(k)}).$$

Theorem 3 claims that, generically, in restriction to (arbitrarily large) compact subset, a system $\Sigma^1 \in S^1$ can be embedded into a phase-variable one, and the state $x(t)$ of $\Sigma^1$ can be recovered from the state $z(t)$ of (2.1), by the inverse mapping of $\Phi^{\Sigma^1}_{k,j^k u}$, which is also smooth.

Of course, if we consider $(u, \dot{u}, ..., u^{(k)}) = v$ as the control, systems of the form (2.1), $y^{(k)} = \varphi_K(y, ..., y^{(k-1)}, v)$, are observable, strongly differentially observable, uniformly infinitesimally observable. Hence, if $m > p$, (and for sufficiently smooth inputs), generic systems are subsystems of other systems that are very good from the point of view of observability.

2.3. **The nongeneric case $m \leq p$.**

2.3.1. *The canonical flag.* In this discussion, we will restrict to the case where $m = 1,\ \ p \geq 1$. Results for $m > 1$ are less clear. We will restrict also to analytic systems in $S^1$, but this is a purely technical assumption that can be avoided.

Associated to $\Sigma = (f, h) \in S^1$, we may define the **canonical flag** $D(u)$ of distributions as follows:

$$D(u) = \{D^0(u) \supset D^1(u) \supset ... \supset D^{n-1}(u)\}, \tag{2.2}$$
$$D^0(u) = Ker(d_x h), \quad D^{k+1}(u) = D^k(u) \cap Ker(d_x L_f^{k+1} h),$$

where $L_f h$ is the Lie derivative of the function $h$ w.r.t. the vector field $f$, the control $u$ being considered as fixed.

The flag $D(u)$ is a flag of possibly singular distributions, depending on the value of the control $u$.

If the distributions $D^i(u)$ have constant rank $n - i - 1$ and are independent of $u$, then, the canonical flag $D(u)$ is said to be **uniform**.

**Theorem 4.** [8, page 22] *The system $\Sigma$ has a uniform canonical flag if and only if, for all $x^0 \in X$, there is a coordinate neighborhood of $x^0$, $(V_{x^0}, x)$, such that, in these coordinates, the system $\Sigma_{|V_{x^0}}$ ($\Sigma$ restricted to $V_{x^0}$) can be written as:*

$$\begin{aligned} y &= h(x_1, u); \\ \dot{x}_1 &= f_1(x_1, x_2, u), \\ \dot{x}_2 &= f_2(x_1, x_2, x_3, u), \\ &\vdots \\ \dot{x}_{n-1} &= f_{n-1}(x_1, x_2, .., x_n, u), \\ \dot{x}_n &= f_n(x_1, x_2, ..., x_n, u), \end{aligned} \tag{2.3}$$

*where moreover*

$$\frac{\partial h}{\partial x_1} \text{ and } \frac{\partial f_i}{\partial x_{i+1}}, i = 1, .., n-1, \tag{2.4}$$

*are never zero on $V_{x_0} \times U$.*

The property to have a uniform canonical flag is highly non-generic (it has codimension $\infty$).

It is easily seen from this normal form that, if a system $\Sigma$ has a uniform canonical flag, then, when restricted to neighborhoods $V_{x_0} \times U$ where it is under the normal form (2.3, 2.4), it is **infinitesimally observable, observable, and differentially observable of order** $n$.

2.3.2. *Characterization of uniform infinitesimal observability.* The main result is the following: a necessary condition for a system $\Sigma$ to be uniformly infinitesimally observable, is that, on an open-dense subset of $X$, it has a uniform canonical flag.

**Theorem 5.** *If $\Sigma$ is uniformly infinitesimally observable, then, on the complement of a subanalytic subset of $X$ of codimension 1, $\Sigma$ has a uniform canonical flag.*

In other term, the canonical form (2.3, 2.4) characterizes uniform infinitesimal observability.

2.3.3. *Control affine case.* In the control affine case, where $\Sigma$ can be written:

$$\begin{aligned} y &= h(x); \\ \dot{x} &= f(x) + \sum_{i=1}^{p} g_i(x) u_i, \end{aligned} \tag{2.5}$$

there is a stronger result. Set $\Phi = (h, L_f h, ..., L_f^{n-1} h)$, $\Phi : X \to \mathbb{R}^n$. First, it is an elementary exercise to show that, if $\Sigma$ is observable, then $\Phi$ has to have maximum rank $n$ on an open dense subset $V$ of $X$. Then, consider any subset $W \subset X$ in restriction to which $\Phi$ is a diffeomorphism.

**Theorem 6.** *Assume that $\Sigma$ is observable. Then, the restriction $\Phi_{|W}$ maps $\Sigma$ into a system of the form:*

$$\begin{aligned} y &= x_1; \\ \dot{x}_1 &= x_2 + \sum_{i=1}^{p} g_{1,i}(x_1) u_i, \\ \dot{x}_2 &= x_3 + \sum_{i=1}^{p} g_{2,i}(x_1, x_2) u_i, \\ &\cdot \\ &\cdot \\ \dot{x}_{n-1} &= x_n + \sum_{i=1}^{p} g_{n-1,i}(x_1, x_2, .., x_{n-1}) u_i, \\ \dot{x}_n &= \psi(x) + \sum_{i=1}^{p} g_{n,i}(x_1, x_2, .., x_{n-1}, x_n) u_i. \end{aligned} \tag{2.6}$$

*Conversely, if a system is under the form (2.6) on an open subset $\Omega \subset \mathbb{R}^n$, then it is observable.*

This normal form (2.6) is of course a special case of the uniform infinitesimal observability canonical form (2.3, 2.4).

Notice that both theorems 5, 6 have a global character: they are local almost everywhere w.r.t. $x$, but global w.r.t. $u$. For this reason, proof of Theorem 5 is not that easy.

## 3. Identification theory

We will restrict to the uncontrolled case, i.e. our systems $\Sigma^3 = (f, h) \in S^3$ are of the form:

$$\Sigma^3 : \left\{ \begin{array}{l} y = h(x, \varphi(x)); \\ \dot{x} = f(x, \varphi(x)), \end{array} \right. \tag{3.1}$$

with $\varphi : X \to \mathbb{I}$, i.e. there is no control, and a single function to identify. These are the results presented in [4]. But now, we already have results for more general systems, with controls, and with several $\varphi'$s.

The results of our identifiability theory are very comparable to the results of Section 2 above.

3.1. **Definitions.** We will give several definitions of identifiability, starting form a general natural one, but not very tractable. In our definitions, as we said, $\varphi(\cdot)$ will not be considered as a function of $x$, but as an extra input, function of the time $t$. Some of these definitions are with respect to the class of functions $\varphi(\cdot)$ that are measurable bounded only (although, $\varphi(x(t))$ is smooth, and even analytic in $t$ if $\Sigma^3$ is analytic). This choice (of a largest class of $\varphi'$s) is in fact justified by the following property: if a system -analytic- is identifiable (or uniformly infinitesimally identifiable) in the sense defined below, for $C^\omega$ inputs $\varphi(t)$, then it is also identifiable in the same sense, for general $L^\infty$ inputs. This is discussed in details in our paper [4], and in the paper to appear [3].

In presence of controls $u$ (for systems in $S^2$, which we do not address here), the natural class for the $\varphi$'s as functions of $t$ is the class of absolutely continuous functions: $\varphi(x)$ is smooth and $x(t)$ is absolutely continuous.

**Definition 5.** *The system $\Sigma \in S^3$ is said identifiable at $y(\cdot) \in C^\infty[0, T_y[$, if there is at most a single couple $(x_0, \varphi(\cdot))$, with $\varphi(\cdot) \in C^\infty[0, T_y[$, such that, for all $t \in [0, T_y[$,*

$$PX\mathbb{I}_\Sigma(x_0, \varphi)(t) = y(t).$$

*$\Sigma$ is said identifiable if it is identifiable at all $y(\cdot) \in C^\infty[0, T_y[$.*

In other terms, $\Sigma$ is identifiable if its input-state-output mapping is injective.

Now, let us consider $\Phi_k^\Sigma : X \times \mathbb{I} \times \mathbb{R}^{(k-1)} \to \mathbb{R}^{km}$, the $k$-jet input-state output mapping of $\Sigma$, $(x_0, j^k\hat{\varphi}) \mapsto j^k\hat{y}$.

**Definition 6.** *The system $\Sigma$ is said differentially identifiable of order $k$, if,*

$$\Phi_k^\Sigma(x_0^1, j^k\hat{\varphi}^1) = \Phi_k^\Sigma(x_0^2, j^k\hat{\varphi}^2)$$

*implies that $(x_0^1, \hat{\varphi}^1(0)) = (x_0^2, \hat{\varphi}^2(0))$.*

This property is weaker than the injectivity of $\Phi_k^\Sigma$. It means that all couples (initial state, value of $\varphi$) are distinguished between them by the observations and their $k-1$ first derivatives. But, it may happen that certain couples $(x_0, j^k\hat{\varphi})$ are not distinguished between them.

For the purpose of getting a genericity result similar to Theorem 2 for observability, this is the adequate notion. (one could think that the injectivity of $\Phi_k^\Sigma$ is the right notion for this purpose, but it is never generic).

Also, the following is more or less obvious:

**Theorem 7.** *Differential identifiability at some order implies identifiability.*

Now, we will define the infinitesimal notion of identifiability.

We consider $TX\mathbb{I}_\Sigma$, the first input-state-output variation of $\Sigma \in S^3$, and its input-state-output map $PTX\mathbb{I}_\Sigma$,

$$(\xi_0, \eta(\cdot)) \mapsto \hat{y}(\cdot),$$
$$T_{x_0}X \times L^\infty[\mathbb{R}] \to L^\infty[\mathbb{R}^m].$$

It is equivalent to consider the tangent mapping $T_{(x_0,\varphi_0)}PX\mathbb{I}_\Sigma$ of the input-state-output mapping $PX\mathbb{I}_\Sigma$ of $\Sigma$.

**Definition 7.** *$\Sigma$ is said uniformly infinitesimally identifiable if, for all $(x_0, \varphi_0(\cdot)) \in X \times L^\infty[\mathbb{I}]$, the tangent mapping $PTX\mathbb{I}_\Sigma$ is injective (as a mapping $T_{x_0}X \times L^\infty([0,t], \mathbb{R}) \to L^\infty([0,t], \mathbb{R}^m)$, for all $t < T_{y_0}$, where $y_0(\cdot) \in L^\infty(\mathbb{R}^m)$ is defined on $[0, T_{y_0}[$).*

That is, uniform infinitesimal identifiability means that all the tangent mappings to the input-state-output mapping are injective.

It will be a consequence of the theory that (in all the cases under consideration) uniform infinitesimal identifiability implies identifiability of the restrictions of the system to certain small open subsets of $X \times \mathbb{I}$, the union of which is dense in $X \times \mathbb{I}$.

3.2. **The generic case.** We have the fundamental following result, comparable to Theorem 2 for observability.

**Theorem 8.** *If the number of outputs $m$ is larger or equal to 3, then, differential identifiability of order $2n+1$ is a generic property. In particular, identifiability is a generic property.*

Of course, this theorem is false if $m = 1, 2$. On the contrary, identifiability becomes a property of infinite codimension.

3.3. **The nongeneric cases $m = 1, 2$.** Again here, and also for purely technical reasons, we consider systems that are analytic only.

3.3.1. *The single output case.* We denote by $L_f$ (or $L_{f_\varphi}$, when $\varphi$ is fixed) the Lie-derivative operator on $X$. Also, $f_\varphi$ denotes the vector field $f(.,\varphi)$, for $\varphi \in \mathbb{I}$, and $h_\varphi : X \to \mathbb{R}$ is the map $h(.,\varphi)$.

**Theorem 9.** *If $\Sigma$ is uniformly infinitesimally identifiable, then, there is a subanalytic closed subset $Z$ of $X$, of codimension 1 at least, such that on the open dense set $X \backslash Z$, the following two equivalent properties 1 and 2 below hold:*

*1.a.* $\frac{\partial}{\partial \varphi}\left\{(L_{f_\varphi})^k h_\varphi\right\} \equiv 0$, *for* $k = 0, ..., n-1$, *b.* $\frac{\partial}{\partial \varphi}\left\{(L_{f_\varphi})^n h_\varphi\right\} \neq 0$ *(in the sense that it* ***never*** *vanishes), c.* $d_x h_\varphi \wedge ... \wedge d_x L_{f_\varphi}^{n-1} h_\varphi \neq 0$,

*2. any $x_0 \in X \backslash Z$ has a coordinate neighborhood $(x_1, ...., x_n,\ V_{x_0})$, $V_{x_0} \subset X \backslash Z$ in which $\Sigma$ (restricted to $V_{x_0}$) can be written:*

$$\left\{\begin{array}{c} \dot{x}_1 = x_2, \\ \dot{x}_2 = x_3, \\ . \\ . \\ \dot{x}_{n-1} = x_n, \\ \dot{x}_n = \psi(x, \varphi); \\ y = x_1; \end{array}\right. \tag{3.2}$$

*where $\frac{\partial}{\partial \varphi}\psi(x,\varphi)$ never vanishes.*

This theorem has the following pseudo-converse:

**Theorem 10.** *Assume that $\Sigma$ meets the equivalent conditions of the previous theorem.*

*Then, any $x_0$ has a neighborhood $V_{x_0}$ such that the restriction $\Sigma_{|V_{x_0}}$ of $\Sigma$ to $V_{x_0}$ is uniformly infinitesimally identifiable,* ***identifiable and differentially identifiable of order*** *$n+1$.*

Notice that, again, Theorem 9 has a global character: it is almost everywhere on $X$, but it is global with respect to $\varphi \in \mathbb{I}$.

3.3.2. *The two-output case.* Let us first state the results in a non invariant way.

**Theorem 11.** *If $\Sigma$ is uniformly infinitesimally identifiable, then, there is an open-dense subanalytic subset $\tilde{U}$ of $X \times \mathbb{I}$, such that each point $(x_0, \varphi_0)$ of $X \times \mathbb{I}$, has a neighborhood $V_{x_0} \times \mathbb{I}_{\varphi_0}$, and coordinates $x$ on $V_{x_0}$ such that the system $\Sigma$ restricted to $V_{x_0} \times \mathbb{I}_{\varphi_0}$, denoted by $\Sigma_{|V_{x_0} \times \mathbb{I}_{\varphi_0}}$, has one of the three following normal forms:*

*-**type 1 normal form:** (in that case, $n > 2k$)*

$$
\begin{aligned}
y_1 = x_1,\ y_2 = x_2,&\\
\dot{x}_1 = x_3,\ \dot{x}_2 = x_4,&\\
..&\\
\dot{x}_{2k-3} = x_{2k-1},\ \dot{x}_{2k-2} = x_{2k},&\\
\dot{x}_{2k-1} = f_{2k-1}(x_1, ..., x_{2k+1}),&\\
\dot{x}_{2k} = x_{2k+1},&\\
..&\\
\dot{x}_{n-1} = x_n,&\\
\dot{x}_n = f_n(x, \varphi),\ \textit{with } \frac{\partial f_n}{\partial \varphi} \neq 0 \textit{ (never vanishes),}&
\end{aligned}
\tag{3.3}
$$

*-**type 2 normal form:***

$$
\begin{aligned}
y_1 = x_1,\ y_2 = x_2,&\\
\dot{x}_1 = x_3,\ \dot{x}_2 = x_4,&\\
..&\\
\dot{x}_{2r-3} = x_{2r-1},\ \dot{x}_{2r-2} = x_{2r},&\\
\dot{x}_{2r-1} = \Phi(x, \varphi),\ \dot{x}_{2r} = F_{2r}(x_1..., x_{2r+1}, \Phi(x, \varphi)),&\\
\dot{x}_{2r+1} = F_{2r+1}(x_1..., x_{2r+2}, \Phi(x, \varphi)),&\\
..&\\
\dot{x}_{n-1} = F_{n-1}(x, \Phi(x, \varphi)),&\\
\dot{x}_n = F_n(x, \varphi),&
\end{aligned}
\tag{3.4}
$$

*with $\frac{\partial \Phi}{\partial \varphi} \neq 0$, $\frac{\partial F_{2r}}{\partial x_{2r+1}} \neq 0, ...., \frac{\partial F_{n-1}}{\partial x_n} \neq 0$,*

*-**type 3 normal form:***

$$
\begin{aligned}
y_1 = x_1,\ y_2 = x_2,&\\
\dot{x}_1 = x_3,\ \dot{x}_2 = x_4,&\\
..&\\
\dot{x}_{n-3} = x_{n-1},\ \dot{x}_{n-2} = x_n,&\\
\dot{x}_{n-1} = f_{n-1}(x, \varphi),\ \dot{x}_n = f_n(x, \varphi),&
\end{aligned}
\tag{3.5}
$$

*where ($\frac{\partial f_{n-1}}{\partial \varphi}$ , $\frac{\partial f_n}{\partial \varphi}$) never vanishes.*

Notice that the type 2 normal form 3.4 is very comparable to the observability normal form (2.3, 2.4).

Now, we will give the intrinsic characterization of the conditions “type 1, type 2, type 3”.

We define two integers $r$ and $k$, attached to a two-output system $\Sigma \in S^3$. The first one $r$ is called the order of the system. It is the first integer such that $d_\varphi L_f^r h$ does not vanish identically on $X \times \mathbb{I}$.

Set $h = (h_1, h_2)$.

Let $N(l)$ be the rank at generic points of $X \times \mathbb{I}$ of the family $E_l$ of one-forms on $X$:

$$E_l = \{d_x h_i, d_x L_f h_i, ..., d_x L_f^{l-1} h_i, i = 1, 2\}.$$

Set $N(0) = 0$.

This set of generic points $U_l$, is the intersection of the open sets $\tilde{U}_i$, $i \leq l$, where $E_i$ has maximal rank. $U_l$ is semianalytic, open and dense in $X \times \mathbb{I}$. Moreover, $U_{l+1} \subset U_l$.

It is easy to check that $N(l)$ increases strictly by steps of 2, up to $l \stackrel{\text{def.}}{=} k$, and after, (eventually), it increases by steps of 1 up to $l \stackrel{\text{def.}}{=} l_M$, $N(l_M) \leq n$.

It may happen that $k = 0$, i.e. $N(1) = 1$.

**Lemma 1.** *If $\Sigma$ is uniformly infinitesimally identifiable, then, $N(l_M) = n$ and $r \leq l_M$.*

**Definition 8.** *A system $\Sigma$ is **regular** if $N(l_M) = n$ and $r \leq l_M$.*

Lemma 1 says that, if a system is uniformly infinitesimally identifiable, then it is regular. From now on, in this section, we will assume that systems $\Sigma$ under consideration are regular.

The integer $k$ is the first with the following properties:

$$d_x h_1 \wedge d_x h_2 \wedge d_x L_f h_1 \wedge ... \wedge d_x L_f^k h_1 \wedge d_x L_f^k h_2 \equiv 0, \quad \text{but}$$

$$d_x h_1 \wedge d_x h_2 \wedge d_x L_f h_1 \wedge ... \wedge d_x L_f^{k-1} h_1 \wedge d_x L_f^{k-1} h_2 \neq 0 \text{ (not identically zero).}$$

If $r = k$, there are three possibilities:

**A**. $n = 2k$;

**B.**

**B.1.**

$$d_x h_1 \wedge d_x h_2 \wedge d_x L_f h_1 \wedge ... \wedge d_x L_f^{k-1} h_2 \wedge d_x L_f^k h_1 \neq 0$$

(hence $n > 2k$) and $d_\varphi L_f^k h_2 \neq 0$; or,

**B.2.**

$$d_x h_1 \wedge d_x h_2 \wedge d_x L_f h_1 \wedge ... \wedge d_x L_f^{k-1} h_2 \wedge d_x L_f^k h_2 \neq 0$$

(hence $n > 2k$) and $d_\varphi L_f^k h_1 \neq 0$;

**C.**

**C.1**

$$d_x h_1 \wedge d_x h_2 \wedge d_x L_f h_1 \wedge ... \wedge d_x L_f^{k-1} h_2 \wedge d_x L_f^k h_1 \neq 0$$

(hence $n > 2k$) and $d_\varphi L_f^k h_2 \equiv 0$, $d_x h_1 \wedge d_x h_2 \wedge d_x L_f h_1 \wedge ... \wedge d_x L_f^{k-1} h_2 \wedge d_x L_f^k h_2 \equiv 0$, or

**C.2**

$$d_x h_1 \wedge d_x h_2 \wedge d_x L_f h_1 \wedge ... \wedge d_x L_f^{k-1} h_2 \wedge d_x L_f^k h_2 \neq 0$$

(hence $n > 2k$) and $d_\varphi L_f^k h_1 \equiv 0$, $d_x h_1 \wedge d_x h_2 \wedge d_x L_f h_1 \wedge ... \wedge d_x L_f^{k-1} h_2 \wedge d_x L_f^k h_1 \equiv 0$

**Definition 9.** *Let $\Sigma$ be a **regular** system. We say that $\Sigma$ has:*
*-**type 1** if $r > k$, or $r = k$ but **C**. is satisfied,*
*-**type 2** if $r < k$, or $r = k$ but **B**. is satisfied,*
*-**type 3** if $r = k$ and **A**. is satisfied.*

**Lemma 2.** *Types 1, 2 and 3 exhaust the class of regular systems, and form a partition of this class.*

**Type 2 regular systems:**

For a regular system of type 2, eventually interchanging the role of $h_1$, $h_2$, we can assume that $d_\varphi L_f^r h_2(x,\varphi) \neq 0$. In a neighborhood of a point $(x_0, \varphi_0) \in U_{l_M}$, such that $L_f^r h_2(x_0,\varphi_0) = u_0$ and $d_\varphi L_f^r h_2(x_0,\varphi_0) \neq 0$, there is an analytic function $\Phi^*(x,u)$, such that $L_f^r h_2(x, \Phi^*(x,u)) = u$. Let us consider the "auxiliary system" $\Sigma_A$ :

$$\Sigma_A \begin{cases} \dot{x} = f(x, \Phi^*(x,\tilde{\varphi})) = F(x,\tilde{\varphi}) \\ y = h(x, \Phi^*(x,\tilde{\varphi})) = H(x,\tilde{\varphi}). \end{cases}$$

This system is well defined and intrinsic, over an open set $V_{x_0} \times V_{u_0} \subset X \times \mathbb{R}$.

By construction, the integer $r$ (the order) associated with this auxiliary system is the same as the one of the given system $\Sigma$.

Moreover, the following flags $D$ and $D^A$ of integrable distributions over $V_{x_0}$ :

$$D_0(x) = T_x X,\ D_1(x) = Ker(d_x h(x)), ... D_r(x) = D_{r-1}(x) \cap Ker(d_x L_f^{r-1} h(x)),$$
$$D = \{D_0 \supset D_1 \supset ... \supset D_r\};$$

and

$$D_0^A(x) = T_x X,\ D_1^A(x) = Ker(d_x H(x)), ... D_r^A(x) = D_{r-1}^A(x) \cap Ker(d_x L_F^{r-1} H(x)),$$
$$D^A = \{D_0^A \supset D_1^A \supset ... \supset D_r^A\},$$

are equal.

Let us "prolong" the auxiliary flag $D^A$, in the following way:

$$D_{r+1}^A(x,\tilde{\varphi}) = D_r^A(x) \cap Ker(d_x L_F^r H_1(x,\tilde{\varphi})),$$
$$D_{i+1}^A(x,\tilde{\varphi}) = D_i^A(x) \cap Ker(d_x L_F^i H_1(x,\tilde{\varphi})),$$
$$D^A(\tilde{\varphi}) = \{D_0^A \supset D_1^A \supset .. \supset D_r^A \supset D_{r+1}^A(\tilde{\varphi}) \supset .. \supset D_l^A(\tilde{\varphi}) = D_{l+1}^A(\tilde{\varphi})\},$$

where $l$ is the first integer such that $D_l^A(x,\tilde{\varphi}) = D_{l+1}^A(x,\tilde{\varphi})$ at generic points.

**Definition 10.** *The auxiliary flag $D^A(\tilde{\varphi})$ is **regular** on an open subset $U \subset X \times \mathbb{I}$, if $D_l^A(\tilde{\varphi}) = \{0\}$, and all the other $D_i^A(\tilde{\varphi})$ have constant rank first $n - 2i$ $(i \leq r)$, second $n - r - i$ $(r < i < l)$, third, $0$ $(i \geq l = n - r)$; on this open set.*

**Definition 11.** *The auxiliary flag $D^A(\tilde{\varphi})$ is **uniform** on an open subset $U \subset X \times \mathbb{I}$, if it is regular, and independent of $\tilde{\varphi}$.*

This property of having a uniform auxiliary flag (for identifiability) is the equivalent of having a uniform canonical flag for observability: it is the necessary (and almost sufficient condition) for uniform infinitesimal identifiability (type 2):

**Theorem 12.** ***(Normal form for a uniform auxiliary flag)*** *A system $\Sigma$ has a uniform auxiliary flag around $(x_0, \varphi_0)$, iff there is a neighborhood $V_{x_0} \times \mathbb{I}_{\varphi_0}$ of $(x_0, \varphi_0)$, and coordinates on $V_{x_0}$ such that $\Sigma$ can be written:*

$$\begin{aligned}
y_1 = x_1,\ y_2 = x_2,\\
\dot{x}_1 = x_3,\ \dot{x}_2 = x_4,\\
..\\
\dot{x}_{2r-3} = x_{2r-1},\ \dot{x}_{2r-2} = x_{2r},\\
\dot{x}_{2r-1} = \Phi(x,\varphi),\ \dot{x}_{2r} = F_{2r}(x_1...,x_{2r+1},\Phi(x,\varphi)),\\
\dot{x}_{2r+1} = F_{2r+1}(x_1...,x_{2r+2},\Phi(x,\varphi)),\\
..\\
\dot{x}_{n-1} = F_{n-1}(x,\Phi(x,\varphi)),\\
\dot{x}_n = F_n(x,\varphi),
\end{aligned}$$

*with* $\frac{\partial \Phi}{\partial \varphi} \neq 0$, $\frac{\partial F_{2r}}{\partial x_{2r+1}} \neq 0, ...., \frac{\partial F_{n-1}}{\partial x_n} \neq 0$.

**Theorem 13.** ***(intrinsic result in the 2-output case)*** *If $\Sigma$ is uniformly infinitesimally identifiable, (**hence regular**), then, there is an open-dense subanalytic subset $\tilde{U}$ of $X \times \mathbb{I}$, such that at each point $(x_0, \varphi_0)$ of $\tilde{U}$, $\Sigma$ has the following properties, on a neighborhood of $(x_0, \varphi_0)$:*

*-If $\Sigma$ has type 2, the auxiliary flag is uniform,*

*-If $\Sigma$ has type 1, then, $N(r) = n$.*

This theorem is in fact equivalent to Theorem 11.

These two equivalent theorems (Theorems 13, 11) have a weak converse:

**Theorem 14.** *Assume that $\Sigma$ satisfies the equivalent conditions of theorems 13, 11, on some subset $V_{x_0} \times \mathbb{I}_{\varphi_0}$ of $X \times \mathbb{I}$ (so that, taking $V_{x_0}$, $\mathbb{I}_{\varphi_0}$ small enough, the restriction $\Sigma_{|V_{x_0} \times \mathbb{I}_{\varphi_0}}$ has one of the three normal forms above on $V_{x_0} \times \mathbb{I}_{\varphi_0}$). Then, in case type 1, type 2, (normal forms 3.3, 3.4) $\Sigma_{|V_{x_0} \times \mathbb{I}_{\varphi_0}}$ is uniformly infinitesimally identifiable **and identifiable**. In case type 3 (normal form 3.5), this is also true, eventually restricting the neighborhoods $V_{x_0}$, $\mathbb{I}_{\varphi_0}$.*

Also, in the special case of type 1, there is a stronger result:

**Theorem 15.** *Assume $\Sigma$ is uniformly infinitesimally identifiable, (**hence regular**). Assume that $\Sigma$ has type 1. Then, there is an open-dense subanalytic subset $\tilde{X}$ of $X$, such that each point $x_0$ of $\tilde{X}$, has a neighborhood $V_{x_0}$, and coordinates $x$ on $V_{x_0}$ such that the system $\Sigma$ restricted to $V_{x_0} \times \mathbb{I}$, denoted by $\Sigma_{|V_{x_0}}$, has the normal form 3.3 (globally over $V_{x_0} \times \mathbb{I}$). Conversely, if it is the case, then, the restriction $\Sigma_{|V_{x_0}}$ is uniformly infinitesimally identifiable **and identifiable**.*

## 4. Observer and identifier design

4.1. **Observer design.** Here, mainly, we recall the results of [8], and of the paper [5]. We add an improvement that makes the strategy proposed in [5] effective for uniformly infinitesimally observable systems, i.e. systems in normal form (2.3, 2.4). This improvement uses a crucial observation of Hammouri et al., in [9].

4.1.1. *Luenberger-type high-gain observers.* We present the basic construction of high gain observers. It works for general uniformly infinitesimally observable systems in normal form (2.3, 2.4). As a consequence, it works also for control affine systems in the normal form (2.6), or for systems in phase-variable representation (2.1), (therefore, it works in all cases -generic or not- of "observable" systems exhibited by the theory). In the 2 last cases, the construction is more explicit than in the case of Normal-form (2.3, 2.4).

Let us consider a system on $\mathbb{R}^n$, which is globally under the form (2.3, 2.4), or under the affine normal form (2.6), or under the phase-variable form:

$$\begin{aligned} y = Cx = x_1, \\ \dot{x}_1 = x_2, ...., \dot{x}_{N-1} = x_N, \\ \dot{x}_N = \psi(x, u); \end{aligned} \tag{4.1}$$

In that case, we may have several outputs (i.e. $m > 1$, each $x_i \in \mathbb{R}^m$, $n = Nm$), and the control in the normal form (2.1), that was $(u, \dot{u}, ..., u^{(N)})$ is now denoted by $u$. This practically means that the observer system will be fed not only by the control, but also by certain of its derivatives.

We will make the following technical assumptions, that will be discussed below:
**Assumptions (A):**
**Case of Normal-form** (2.3, 2.4):
A1. $0 < \alpha \leq \frac{\partial h}{\partial x_1} \leq \beta$, $0 < \alpha \leq \frac{\partial f_i}{\partial x_{i+1}} \leq \beta$, $i = 1, ..., n-1$
A2. Each of the maps $f_i$, $i = 1, ..., n-1$, is globally Lipschitz w.r.t. $(x_1, ..., x_i))$, uniformly w.r.t. $u$ and $x_{i+1}$.
A3. The control $u$ is bounded,
**Case of Normal-forms** (2.6) **and** (4.1):
A1. The control $u$ (and a certain number of first derivatives of $u$, if any) is bounded
A2. All the functions $\psi$, $g_{i,j}$ appearing in the normal forms are compactly supported.

**Comment 1 about these assumptions**:
The assumption that $u$ is bounded, seems to be a non avoidable requirement. In fact, since we assumed $U$ compact, it is automatically satisfied. We recall it here for clarity. Also, in the case of a phase variable representation 2.1, it means that certain derivatives of $u$ have to be bounded.

**Comment 2 about these assumptions**:
The other assumptions can be always realized, provided that we restrict the observation problem to a compact subset $K \subset X$ (which means that we want to estimate the state $x(t)$ of $\Sigma$ as long as the trajectory $x(t)$ remains in $K$ only).
Indeed:
1. in the case of Normal forms (2.6) and (4.1), all the functions under consideration can be multiplied by a smooth cut-off function, which is equal to one on $K$;
2. in the case of Normal-form (2.3, 2.4), $h$ and $f$ can be smoothly prolonged outside $K$, for they satisfy A1, A2. (This last point is not immediate, and it is shown in [8]).

Under these assumptions, let us consider another system of the form:

$$\frac{d\hat{x}}{dt} = f(\hat{x}, u) - \Delta_\theta \Omega(h(\hat{x}, u) - y(t)); \tag{4.2}$$

where $\Omega$ is a certain constant $n \times m$ matrix, $\theta$ is a real parameter, and $\Delta_\theta$ is the block-diagonal matrix:

$$\Delta_\theta = \text{block-diag}(\theta Id_m, ..., \theta^N Id_m).$$

Here, $Id_m$ denotes the $m \times m$ identity matrix, and $m = 1$ in other cases than (4.1).

**Theorem 16.** *There is an $\Omega$ with the following properties: for all $\hat{\beta} > 0$, there exists a $\theta$ (large enough), such that, whatever $\hat{x}(0)$, the solution $\hat{x}(t)$ of (4.2) satisfies:*

$$||\hat{x}(t) - x(t)|| \leq k(\hat{\beta}) e^{-\hat{\beta} t} ||\hat{x}(0) - x(0)||,$$

*as long as $x(t)$ remains in $K$. The function $k(\cdot)$ has polynomial growth.*

**Remark 2.** *The function $k$ having polynomial growth, it implies that the "estimation error" $||\hat{x}(t) - x(t)||$ can be made arbitrarily small in arbitrary short time (a polynomial against an exponential).*

The construction of the matrix $\Omega$ is not so hard in the case (2.3, 2.4), See [8]. It is specially simple in the cases (2.6), (4.1): any $\Omega$ such that $(A - \Omega C')$ is Hurwitz does the job (where $(A, C)$ is the canonical linear system in Brunowsky form).

This very simple "luenberger-type" observer shows already good performances in many cases. Since it is high-gain ($\theta$ is large, then the "correction gain" $\Delta_\theta \Omega$ might be big), it may be sensitive to noise.

4.1.2. *Kalman-filter type high-gain observers.* The high-gain extended Kalman filter is another solution of high-gain type. Since it is high gain, it might be also sensitive to noise (although, very often, it works well in practice). It is related to the classical extended Kalman filter, and it can be shown that (in a stochastic context), it is a nonlinear filter with bounded variance (See [6]).

In [8] and in [5], it is applied to a less general class of nonlinear systems than the Luenberger high-gain observer. In Section 4.1.4 below, we will show how to apply it to the same class of systems. But in this section, let us stay at the level of the results of [8] and [5].

We assume that the system $\Sigma \in S^1$ is (on $\mathbb{R}^n$) in normal form (2.6), or more generally in the following (multi-output) normal form:

$$\begin{aligned} y = Cx = x_1, \\ \dot{x}_1 = Ax + b(x, u), \end{aligned} \tag{4.3}$$

where $C : \mathbb{R}^n \to \mathbb{R}^m$, $C = (Id_m, 0, ..., 0)$, where $A$ is the $Nm \times Nm = n \times n$ block-antishift matrix:

$$A = \begin{pmatrix} 0 & Id_m & 0 & . & . & 0 \\ 0 & 0 & Id_m & 0 & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & . & . & . & 0 & Id_m \\ 0 & 0 & . & . & . & 0 \end{pmatrix},$$

and where $x = (x_1, ..., x_N)$, $x_i \in \mathbb{R}^m$, and $b(x, u)$ is lower block-triangular: the $i^{th}$ component $b_i$ depends only on $(x_1, ..., x_i)$.

Notice that this normal form includes the case (2.1), (4.1) of a phase-variable representation.

The only case of observable systems which is not covered by this normal form is the "uniform infinitesimal observability normal form" (2.3, 2.4). But, as we said, we will remedy to this in Section 4.1.4.

Again, we need additional technical assumptions:

**Assumption B**:

B1. the components $(x, u) \mapsto b_i(x, u)$ are compactly supported with respect to all their respective arguments;

B2. $u$ is bounded (void assumption if $u \in U$ which is assumed to be compact, but non void assumption if $u = (u, \dot{u}, ..., u^{(k)})$ for a phase-variable representation).

**Remark 3.** *As in the previous section 4.1.1, the assumption B1 can be realized by smooth prolongation of $b$ out of any compact subset $K \subset X \times U$. In that case, the observer we construct will work as long as the trajectories $x(t)$ of the system remain inside $K$.*

Consider the Kalman-type equations:

$$\begin{aligned} \frac{dS}{dt} &= -(A + D_{\hat{x}} b(\hat{x}, u))' S - S(A + D_{\hat{x}} b(\hat{x}, u)) + C' r^{-1} C - S Q_\theta S, \\ \frac{d\hat{x}}{dt} &= A\hat{x} + b(\hat{x}, u) - S^{-1} C' r^{-1} (C\hat{x} - y). \end{aligned} \tag{4.4}$$

Here, $Q_\theta = \theta^2 \Delta^{-1} Q \Delta^{-1}$, $\Delta = \text{block-diag}(Id_m, \frac{1}{\theta} Id_m, ..., (\frac{1}{\theta})^{N-1} Id_m)$. Here, $\theta$ is a real parameter. The matrix $S$, as usual, lies in the cone of symmetric positive definite matrices. This equation is called the "high-gain extended Kalman filter". If $\theta = 1$, it is just the standard Extended Kalman filter equation, and in a stochastic context, $Q$ and $r$ (both symmetric positive definite) are the covariance matrices of the state and output noise respectively.

**Theorem 17.** *Let $x(t)$, $t \geq 0$ be a semi-trajectory of (4.3). Let $\theta$ be large enough. Let $\hat{x}(0)$, $S(0)$ (positive definite) be initial conditions for (4.4). Let $\hat{x}(t)$, $S(t)$ be the corresponding semi-trajectories ($t \geq 0$). Then:*

*a. $S(t)$ remains in a compact subset of the cone of positive definite symmetric matrices,*

*b. for all $\beta > 0$, there is a $\theta$ (large), such that:*

$$||\hat{x}(t) - x(t)|| \leq H(S(0)) k(\beta) e^{-\beta t} ||\hat{x}(0) - x(0)||,$$

*where $H$ is a smooth mapping, and $k$ is a function with polynomial growth.*

Again, the exponential against the polynomial ensure that the estimation error, $||\hat{x}(t) - x(t)||$ can be made arbitrarily small in arbitrary short time.

4.1.3. *A mixed solution.* We present here the version of the extended Kalman filter proposed in [5], that mixes the standard extended Kalman filter and the high-gain extended Kalman filter of the previous section.

We obtain an observer with the following properties:

a) in presence of big disturbances (state jumps), it has the high-gain behavior: the estimation error can be made arbitrarily small very quickly;

b) when the error is small, it behaves as the regular extended Kalman filter (good performances w.r.t. noise).

This observer behaves extremely well in practice, as we will show in Section 5 on a non-academic example, and it can be also used for identification, as explained in section 4.2.

We consider exactly the same class of systems on $\mathbb{R}^n$, globally in normal form (4.3), as in the previous section 4.1.2, and meeting assumption **B** of this section. Again, the compact support assumption for $b$, if not realized, can be obtained by smoothly modifying $b$ out of a large compact subset of $X$.

Then, the equations of the observer are:

$$\begin{aligned} \frac{dS}{dt} &= -(A + D_{\hat{x}} b(\hat{x}, u))' S - S(A + D_{\hat{x}} b(\hat{x}, u)) + C' r^{-1} C - S Q_\theta S, \\ \frac{d\hat{x}}{dt} &= A\hat{x} + b(\hat{x}, u) - S^{-1} C' r^{-1} (C\hat{x} - y), \\ \frac{d\theta}{dt} &= \lambda(1 - \theta). \end{aligned} \tag{4.5}$$

Now, $\theta$ is not constant anymore, hence we have an initial condition $\theta_0$, and $\lambda$ is a positive real, to be chosen not too large as we shall see.

Intuitively, this observer should behave like that:

a. When $t \to +\infty$, $\theta \to 1$, and then, the behavior of the observer for large $t$ is the same as the one of the usual extended Kalman filter (known as good w.r.t. noise in practice [2]).

b. On the contrary, if $\theta_0$ is large, when $t$ is small, the behavior is the one of the high-gain EKF: the error can be made arbitrarily small in arbitrary short time.

This is true, as stated in the following theorem:

**Theorem 18.** *1. For all $0 \leq \lambda \leq \lambda_0$, ($\lambda_0$ a certain positive real), for all $\theta_0$ large enough, (depending on $\lambda$), for all $S_0 \geq c\, Id$, for all $\bar{K} \subset \mathbb{R}^n$, $\bar{K}$ a compact subset, for all $\varepsilon_0 = \hat{x}_0 - x_0$, $\varepsilon_0 \in \bar{K}$, the following estimation holds, for all $\tau \geq 0$.*

***Long term estimation:***

$$\begin{aligned} ||\varepsilon(\tau)||^2 &\leq R(\lambda, c) e^{-a\ \tau} ||\varepsilon_0||^2 \Lambda(\theta_0, \tau, \lambda), \\ \Lambda(\theta_0, \tau, \lambda), &= {\theta_0}^{2(n-1)+\frac{a}{\lambda}} e^{-\frac{a}{\lambda}\theta_0(1-e^{-\lambda\tau})}, \end{aligned} \tag{4.6}$$

*where $a > 0$. $c \mapsto R(\lambda, c)$ is a decreasing function.*

*2. Set $\theta(T) = 1 + (\theta_0 - 1)e^{-\lambda T}$. For all $T > 0$, $\tau \leq T$, for all $\theta_0 \geq \bar{\theta}_0$, $\bar{\theta}_0$ a certain positive real depending on $\lambda T$, for $a_1, a_2$ certain positive constants, and with $H(c)$, a*

[2] No more than in practice: there is no theoretical result (but negative) on the stochastic behavior of the EKF, even for small noise. See [16].

*certain decreasing positive function of c, where* $S(0) \geq c\ Id$*, the following estimation holds.*

***Short term estimation:***

$$||\varepsilon(\tau)||^2 \leq \theta(\tau)^{2(n-1)} H(c) e^{-(a_1\theta(T)-a_2)\tau} ||\varepsilon(0)||^2. \tag{4.7}$$

This theorem has been proved in [5].

**Comments.**

a. Note that the function $\Lambda(\theta_0, \tau, \lambda)$ is a decreasing function of $\tau$, and that, for all $\tau > 0$, $\lambda > 0$, $\Lambda(\theta_0, \tau, \lambda)$ can be made arbitrarily small, increasing $\theta_0$.

b. This means that, provided that $\lambda$ is smaller than a certain constant $\lambda_0$, and $\theta_0$ is large in front of $\lambda$, the estimation error goes exponentially to zero, and can be made arbitrarily small in arbitrary short time.

c. The asymptotic behavior of the observer is the one of the extended Kalman filter,

d. The "short term behavior" is the one of the "high gain extended Kalman filter".

**Practical implementation:**

The problem with this observer is that its behavior is "time dependant": for small time, it is the high-gain EKF, and for large time it is the ordinary EKF.

A way to remedy this evil is to proceed as follows.

We consider a one parameter family $\{O_\tau, \tau \geq 0\}$ of observers of type (4.5), indexed by the time, each of them starting from $S_0$, $\theta_0$, at the current time $\tau$. In fact, in practice, it will be sufficient to consider, at current time $t$, a slipping window of time, $[t-T, t[$, and a finite set of observers $\{O_{t_i},\ t-T \leq t_i \leq t\}$, with $t_i =$ largest multiple of $\frac{T}{N}$ smaller than $t-(i-1)\frac{T}{N}$, $i = 1, ..., N$.

As usual, we call the term $I(\tau) = \hat{y}(\tau) - y(\tau)$, (the difference at time $\tau$ between the estimate output and the real output), the **"innovation".** Here, for each observer $O_{t_i}$,we have an innovation $I_{t_i}(\tau)$.

Our suggestion (very natural and very simple), is to take as the estimate of the state, the estimation given by the observer $O_{t_i}$ that minimizes a norm of the innovation.

Let us analyze what will be the effect of this procedure in a deterministic setting:

1. Let us assume that there is no "jump" of the state. Then, clearly, the best estimation will be given by the "oldest" observer in the window, $O_{t_N}$. Then, the error will be given by the "long term" and "short term" estimates at time $T$ :

$$||\varepsilon(\tau+T)||^2 \leq R(\lambda, c) e^{-a\ T} ||\varepsilon(\tau)||^2 \Lambda(\theta_0, T, \lambda),$$
$$||\varepsilon(\tau+T)||^2 \leq \theta(T)^{2(n-1)} H(c) e^{-(a_1\theta(T)-a_2)T} ||\varepsilon(\tau)||^2.$$

a. If $T$ is large enough, the asymptotic behavior will be the one of the "extended Kalman filter".

b. At the beginning, the transient is the one of the HGEKF.

c. the error can be made arbitrarily small in arbitrary short time, provided that $\theta_0$ is large enough.

2. If at a certain time we have a "jump" of the state, then, the innovation of the "old observers" will become large. The "youngest" one will be chosen, and the transient will be the same as the one of the HGEKF, first, and of the EKF, after $T$.

4.1.4. *The case of the uniform infinitesimal observability normal form (2.3).* In fact, the observer presented in the previous section can be also applied to uniformly infinitesimally observable systems, under the general normal form (2.3, 2.4), provided that the control functions $u(\cdot)$ are (globally on $[0, +\infty[$) Lipschitz functions of the time. We will prove this now.

A first remark is that systems under this normal form (2.3, 2.4) are strongly differentially observable of order $n$ (this is easily checked). A consequence is that they have a phase variable representation of order $n$ (of type 2.1). This is a way to apply the results of the previous section, but this method involves complicated changes of coordinates, and above all, it uses the derivatives of the control $\dot{u}, ..., u^{(n)\cdot}$.

In fact, there is a way to transform systems in normal form (2.3, 2.4) into systems in normal form (4.3), by using only $\dot{u}$, the first derivative of the control.

Consider such a system $\Sigma$ on $\mathbb{R}^n$, and set:

$$z = \Phi_u(x) = (h(x,u), L_f h(x,u), ..., L_f^{n-1} h(x,u)). \tag{4.8}$$

Let $K \subset \mathbb{R}^n$ be any fixed open relatively compact subset. As previously in this paper, we deal with semi-trajectories of $\Sigma$ that remain in $K$, only. It follows from (2.4) that, for all $u \in U$, $\Phi_u$ is an injective immersion (this is easily checked by induction on the components of $\Phi_u$). Therefore, $\Phi_u$ is a $u$-dependent diffeomorphism from $K$ onto its image. Consider the image $\Sigma'$ of the system $\Sigma_{|K}$ ($\Sigma$ restricted to $K$) by the time dependant diffeomorphism $\Phi_u$. It is of the form:

$$\begin{aligned} y &= z_1, \\ \dot{z} &= F(z,u) + G(z,u)\dot{u}, \end{aligned} \tag{4.9}$$

and moreover, it is in fact in the form (4.3):

$$\begin{aligned} y &= z_1, \\ \dot{z} &= Az + \bar{G}(z,u,\dot{u}), \end{aligned} \tag{4.10}$$

where $A$ is the antishift matrix, and where $\bar{G}$ is smooth and depends in a triangular way of $z$ : $\bar{G}_1(x_1, u, \dot{u}), ...$

There is the very small difficulty that $\bar{G}$ is not defined on the whole of $\mathbb{R}^n \times U \times \mathbb{R}^p$, and then, has to be smoothly prolonged to $\mathbb{R}^n \times U \times \Omega$, where $\Omega$ is a compact subset of $\mathbb{R}^p$ ($\Omega$ : set of values of $\dot{u}$, which we may take compact as will be justified below). Moreover, this prolongation can be taken with compact support, in order to meet assumption **B1** of the two previous sections.

Consider any semi-trajectory $x(\cdot) : [0, T[$ of $\Sigma$ (possibly $T = +\infty$), corresponding to a Lipschitz control $u$, with Lipschitz constant $K_u$. Then $u$ is absolutely continuous, and its derivative is bounded by $K_u$ (which explains that we may take $\Omega$ compact: $\dot{u}$ is bounded by the Lipschitz constant of $u$). Consider $Z(t) = \Phi_{u(t)}(x(t))$.

Then, since $x, u$ are absolutely continuous, $Z(\cdot)$ is also absolutely continuous. But, by construction, $Z(t)$ satisfies almost everywhere (4.10) and (4.9), provided

that $x(t)$ remains in $K$, which we assume. Then, $Z(\cdot)$ is the unique absolutely continuous solution of (4.10), corresponding to the controls $u, \dot{u}$, and the initial condition $z_0 = \Phi_{u(0)}(x_0)$.

We have shown the following:

**Lemma 3.** *For Lipschitz controls, (with given Lipschitz constant), semi-trajectories of $\Sigma$ that remain in $K$ are mapped by $\Phi_u$ in the semi-trajectories of the systems (4.9, 4.10).*

Then, at the price of using the first derivative of $u$, the EKF and the "mixed" observer of the previous section 4.1.3, can be used for general uniformly infinitesimally observable systems.

This observation, in fact, comes from the paper [9] (where it is not stated precisely, but the idea is present).

4.2. **Identifier design.** In this section, we will remain in the context of Section 3, i.e. systems without controls. Nevertheless, the application considered in Section 5 will concern systems with controls, but the effect of the controls will be transparent, since the systems are in the normal forms studied previously (but control-dependent).

The following simple single-output example shows that there is no chance to avoid approximate differentiation for effective identification:

$$\begin{aligned} y &= x_1, \\ \dot{x}_1 &= x_2, .., \dot{x}_{n-1} = x_n, \\ \dot{x}_n &= \varphi(x). \end{aligned} \tag{4.11}$$

In fact, in this example, identifying (i.e. **reconstructing the piece of the graph of $\varphi$ visited during the experiment**) is just **equivalent** to differentiate the output $n$ times.

Also, we will not consider the generic case $m \geq 3$, but only the cases corresponding to $m = 1$ (normal form 3.2) and $m = 2$, systems of type 1,2,3.

Our basic idea is the same in all cases, and leads to **the use of the nonlinear observers** developed in the previous section 4.1: we assume, along the trajectories visited, **a local model for** $\varphi$. For instance, a simple local model is: $\varphi^{(k)} = 0$.

This does not mean, at the end, that we will identify $\varphi$ as a polynomial in $t$: the question is not that this polynomial models the function $\varphi$ globally as a function of $t$, but only locally, on reasonable time intervals (reasonable w.r.t. the performances of the observer that we will use).

4.2.1. *The single output case.* Let us consider a system $\Sigma$ in the identifiability normal form 3.2. Adding the local model for $\varphi$, we get the system:

$$\begin{aligned} y &= x_1, \\ \dot{x}_1 &= x_2, ..., \dot{x}_{n-1} = x_n, \\ \dot{x}_n &= \Psi(x, \varphi_1), \dot{\varphi}_1 = \varphi_2, ..., \dot{\varphi}_{k-1} = \varphi_k, \dot{\varphi}_k = 0, \end{aligned} \tag{4.12}$$

$$\frac{\partial \Psi}{\partial \varphi_1} \neq 0 \quad \text{(never vanishes)}. \tag{4.13}$$

This is a system on $\mathbb{R}^{n+k}$, which is not controlled (however, for the considerations that follow, $\Psi$ could depend on a control $u$), and this system is under the normal form (2.3, 2.4).

Therefore, we may apply the high gain Luenberger observer, or we may apply the trick of the previous section 4.1.4. Then, for instance, the observer of Section 4.1.3 may be applied to this system. It will provide estimations of $x(t), \varphi(t)$, that is, just an estimation of the piece of the graph of $\varphi$ visited during the experiment. (It provides also estimations of $\dot{\varphi}, ..., \varphi^{(k)}$, which we don't care about).

4.2.2. *The two-output case.* The cases of normal forms (3.3), (3.4), (3.5), corresponding to Type 1 to 3 systems can be treated in a similar way to the single-output case, with some more or less easy adaptations of the methods of the previous sections.

Let us just consider one example: the case of a Type 2 system, with $r = 0$, which is very illustrative, and goes directly back to the observation problem for uniformly infinitesimally observable systems:

$$\begin{aligned} y_1 &= \Phi(x,\varphi), \quad y_2 = h(x_1, \Phi(x,\varphi)), \\ \dot{x}_1 &= F_1(x_1, x_2, \Phi(x,\varphi)), \\ &.. \\ \dot{x}_{n-1} &= F_{n-1}(x, \Phi(x,\varphi)), \\ \dot{x}_n &= F_n(x,\varphi), \end{aligned} \tag{4.14}$$

with $\frac{\partial \Phi}{\partial \varphi} \neq 0$, $\frac{\partial F_1}{\partial x_2} \neq 0, ...., \frac{\partial F_{n-1}}{\partial x_n} \neq 0$, $\frac{\partial h}{\partial x_1} \neq 0$.

In that case, Let us set $\Phi(x,\varphi) = \Phi(t)$ $(= y_1(t))$. Then, forgetting about $y_1$, and since $\Phi_x(\varphi)$ is an invertible function of $\varphi$ for $x$ fixed, we have the system:

$$\begin{aligned} y_2 &= h(x_1, \Phi(t)) \\ \dot{x}_1 &= F_1(x_1, x_2, \Phi(t)), \\ &.. \\ \dot{x}_{n-1} &= F_{n-1}(x, \Phi(t)), \\ \dot{x}_n &= F_n(x, \Phi_x^{-1}(\Phi(t))). \end{aligned} \tag{4.15}$$

The function $\Phi(t)$ being known, the identification problem is just the observation problem for this new system, which is in uniform infinitesimal observability normal form: having an estimate $\hat{x}(t)$ of the state, we get an estimate of $\varphi(t)$ by $\hat{\varphi}(t) = \Phi^{-1}_{\hat{x}(t)}(y_1(t))$.

Then, we may apply the high-gain Luenberger observer (4.2), or at the price of a single (approximate) differentiation of $y_1$, we may apply again the observer of section 4.1.3.

## 5. Application to a fluid catalytic cracker

A fluid catalytic cracker (FCC) is a process used in refineries to produce gasoline from heavy petroleum residues.

The well known model of FCC used in this paper is adapted from [12]. It has been used by several authors as an example of a highly nonlinear system with a lot of strong interactions, see for instance [13, 14]. This model has already been used by the first author, to study the performances of high gain observers, in [2].

A FCC unit, as depicted on Figure 1, is composed of a reactor and a regenerator. The heavy petroleum residues feed the FCC by the bottom of the reactor (called the riser). Long molecules are broken thanks to a catalyst, that circulates between the reactor and the regenerator. When the long molecules are broken, carbon is produced (coke), and is fixed on the catalyst, that becomes dirty. In the regenerator, the coke on the catalyst is burned, regenerating the catalyst. The cracking reactions inside the riser and the reactor are endothermic, and use the heat produced when regenerating the catalyst. This "heat flowrate" is driven from the regenerator to the riser of the reactor, by the flow of regenerated catalyst.

Therefore it is clear that the thermal balance between the reactor and the regenerator results in a strong coupling between these two parts, that affects crucially the behavior of the unit.

5.1. **Description of the model, and purpose of the study.** Our model is closely related to the Kurihara model [12], as described in [14] for dynamic optimization. The Kurihara FCC model has also been used in [2] in order to estimate the "carbon Conradson factor" $F_{cf}$, an important parameter from the point of view of operation: it characterizes in some sense the propensity of the catalyst to become inefficient. In particular, it becomes worse when the catalyst is old, and it is not very well known. **Our first objective here, will be also the estimation of this parameter** $F_{cf}$.

But here, we will use less measurements than in [2]: indeed, we will only use temperature measurements, from both the reactor and the regenerator. We will not use the measurement of the carbon concentration in the regenerator, which was supposed to be reconstructed in [2] thanks to a measurement of the concentration of oxygen in flue gas.

Another problem is that the local model of oxygen combustion in the regenerator is not very accurate (may be not very well known, in fact). Therefore, the purpose of the second part of this study will be to **identify this model.**

The control of the FCC unit is performed via **two control variables**, namely the "air flowrate" ($R_{ai}$) (air of combustion of the coke on the catalyst, inside the regenerator) and the "catalyst circulation rate" ($R_c$).

The Kurihara model is really a two time-scales model: the evolutions of the rate of carbon in the reactor and in the regenerator are modelled by three differential equations. However, one of these equations (modelling the catalytic carbon balance) has a very short time-constant with respect to other dynamics inside the system. Here, we have replaced this differential equation by an algebraic one. Comparisons of solutions between both models lead to very similar results.

Another simplification that we make is the following: We assume steady-state for the catalyst flowrates between the reactor and the regenerator. These two simplifications are analyzed from a chemical-engineering point of view in [13]. In this paper ([13]), authors use the model to illustrate the fact that the FCC may admit multiple equilibria.

In fact, these two modifications of the Kurihara FCC model and the local model of oxygen burning in the regenerator are the main differences between [12] and [13] FCC models.

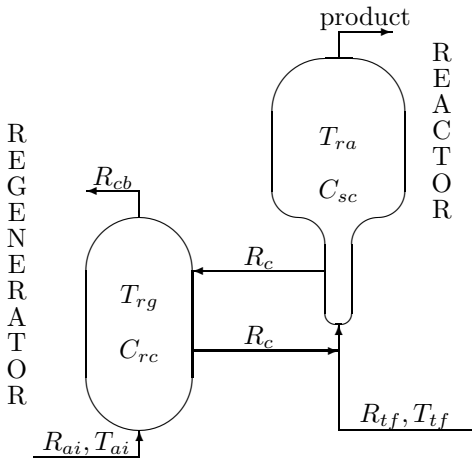


FIGURE 1. FCC Unit

As we shall see, in the case of temperature measurements only, this simplified model is observable, even if $F_{cf}$ is considered as an unknown parameter, and added to the system as a constant state variable (indeed, this parameter varies very slowly: it represents the long-term degradation of the catalyst, as we said).

Finally, the model we consider consists of a set of four differential equations representing the evolution of the reactor temperature (5.1), regenerator temperature (5.8), carbon concentration on spent catalyst (5.4) and carbon concentration on regenerated catalyst (5.13).

5.1.1. *Reactor model.*
Temperature in the reactor:

$$S_c H_{ra} \dot{T}_{ra} = S_c R_c \left(T_{rg} - T_{ra}\right) + S_{tf} R_{tf} \left(T_{tf} - T_{ra}\right) - \Delta H_{fv} R_{tf} - \Delta H_{cr} R_{tf} C_{tf} \tag{5.1}$$

$$C_{tf} = \frac{R_{cr}}{R_{cr} + R_{tf}} \tag{5.2}$$

$$R_{cr} = \frac{\sqrt{k_{cr} R_c P_{ra} H_{ra}}}{10\, C_{rc}^{0.12}} \exp\left(-\frac{1}{2} \frac{A_{cr}}{R T_{ra}}\right), \tag{5.3}$$

with:

- Reactor operating conditions $T_{ra}|_{t=0} = 775\,\mathrm{K}$, $H_{ra} = 1.85\,10^{-4}\,\mathrm{kg}$, $P_{ra} = 211.7\,\mathrm{kPa}$,
- Feed properties $R_{tf} = 41\,\mathrm{kg\,/\,s}$, $T_{tf} = 492.8\,\mathrm{K}$, $S_{tf} = 3140\,\mathrm{J\,/\,(kg\,.\,K)}$,
- Catalyst recirculation $R_c|_{t=0} = 290\,\mathrm{kg\,/\,s}$, $0 < R_c^{\min} \leq R_c \leq R_c^{\max}$, $S_c = 1047\,\mathrm{J\,/\,(kg\,.\,K)}$,
- Heat constants $\Delta H_{cr} = 4.65\,10^5\,\mathrm{J\,/\,kg}$, $\Delta H_{fv} = 1.74\,10^5\,\mathrm{J\,/\,kg}$, $\Delta H_{rg} = 3.02\,10^7\,\mathrm{J\,/\,kg}$,
- $k_{cr} = 25.96\ \mathrm{kPa}^{-1}\,\mathrm{s}^{-1}$, $A_{cr} = 83.8\,10^3\,\mathrm{J\,/\,mol}$
- $R = 8.314\,\mathrm{J\,/\,(mol\,.\,K)}$

Carbon concentration on spent catalyst in the reactor:

$$H_{ra}\dot{C}_{sc} = R_c\,(C_{rc} - C_{sc}) + 100\,R_{cf} \tag{5.4}$$

$$R_{cf} = R_{cc} + R_{ad} \tag{5.5}$$

$$R_{cc} = \frac{\sqrt{k_{cc}R_cP_{ra}H_{ra}}}{10\,C_{rc}^{0.03}}\exp\left(-\frac{1}{2}\frac{A_{cc}}{RT_{ra}}\right) \tag{5.6}$$

$$R_{ad} = F_{cf}R_{tf} \tag{5.7}$$

with:

- $C_{sc}|_{t=0} = 1.2$
- $k_{cc} = 2.66\,10^{-4}\,\mathrm{kPa}^{-1}\,\mathrm{s}^{-1}$, $A_{cc} = 4.18\,10^4\,\mathrm{J}\,/\,\mathrm{mol}$

5.1.2. *Regenerator model.*
Temperature in the regenerator:

$$S_cH_{rg}\dot{T}_{rg} = S_cR_c\,(T_{ra} - T_{rg}) + S_{ai}R_{ai}\,(T_{ai} - T_{rg}) + \Delta H_{rg}R_{cb} \tag{5.8}$$

$$R_{cb} = \frac{R_{ai}}{242}\,(21 - O_{fg}) \tag{5.9}$$

$$O_{fg} = 21\exp\left(\frac{-\frac{P_{rg}H_{rg}}{R_{ai}}}{\frac{1}{K_{od}} + \frac{1}{K_{or}C_{rc}}}\right) \tag{5.10}$$

$$K_{od} = 6.34\,10^{-9}R_{ai}^2 \tag{5.11}$$

$$K_{or} = 1.16\,10^{-5}\exp\left(\frac{A_{or}}{R\left(\frac{1}{866.7} - \frac{1}{T_{rg}}\right)}\right) \tag{5.12}$$

with:

- Regenerator operating conditions $T_{rg}|_{t=0} = 943\,\mathrm{K}$, $H_{rg} = 1.53\,10^5\,\mathrm{kg}$, $P_{rg} = 254.4\,\mathrm{kPa}$,
- Air properties $R_{ai}|_{t=0} = 26\,\mathrm{kg}\,/\,\mathrm{s}$, $0 < R_{ai}^{\min} \leq R_{ai} \leq R_{ai}^{\max}$, $T_{ai} = 394\,\mathrm{K}$, $S_{ai} = 1130\,\mathrm{J}\,/\,(\mathrm{kg}\,.\,\mathrm{K})$
- $A_{or} = 1.47\,10^5\,\mathrm{J}\,/\,\mathrm{mol}$

Carbon concentration on regenerated catalyst in the regenerator:

$$H_{rg}\dot{C}_{rc} = R_c\,(C_{sc} - C_{rc}) - 100\,R_{cb} \tag{5.13}$$

with $C_{rc}|_{t=0} = 0.3$

5.2. **Estimation of the Carbon Conradson factor.** Here, we will use this model to estimate unmeasured state variables. We will first study the observability, and, after positive answer to this question, we will apply the observer construction described above. Moreover, in the next section, we will assume that a part of the model is unknown and we will show that the unknown function is identifiable. Therefore the system can be transformed to a certain observability canonical form, similar to the type 2 normal form 3.4.

In fact, it can also be put under a form such that our "mixed" observer construction of Section 4.1.3, can be applied in order to estimate simultaneously both the state variables and unknown function.

However, this section is devoted to observation only: we will assume that our knowledge-based model is perfectly known and we will use it to **estimate unknown**

**state variables** ($C_{rc}$, $C_{sc}$ and $F_{cf}$) thanks to **temperature measurements** ($T_{ra}$ and $T_{rg}$).

5.2.1. *Observer construction.* First of all, an elementary analysis shows that the system is observable and infinitesimally observable: Indeed, since $T_{ra}$ is measured, the derivative $\dot{T}_{ra}$ allows to compute $C_{rc}$ from $C_{tf}$. Then $\dot{C}_{rc}$ gives $C_{sc}$. Finally, $\dot{C}_{sc}$ gives $R_{cf}$ and hence $F_{cf}$. Moreover, the function $C_{rc} \mapsto C_{tf}$ has nonvanishing derivative on any interval of the form $[\varepsilon, +\infty[$, $\varepsilon > 0$ (the other variables being fixed and positive). Hence, the system is globally observable and **uniformly infinitesimally observable**. The fact that $C_{rc}$ belongs to an interval of the form $[\varepsilon, +\infty[$ will be shown later (Proposition 1)

Notice that the **measurement $T_{rg}$ is not used** here. We will use it in the next section in order to **identify the unknown function.**

This analysis of infinitesimal observability shows us that the system may already be written under the (single output, as we said) observability canonical form 2.3, 2.4: by the theory, this can be done at least locally (see [8] and Section 2.3).

Nevertheless, in order to apply our mixed high-gain extended Kalman filter construction 4.1.3, we have to apply the trick presented in Section 4.1.4. To do this, we have to find a coordinate-change, such that the system in new coordinates is under the canonical form 4.10, as explained in Section 4.1.4. To do this, we will consider successive time derivatives of the output, and at each step, we will obtain a new coordinate, corresponding to a new variable obtained by derivation.

Let us consider first the reactor temperature. Its time derivative is given by the right hand side of 5.1. Since both $T_{ra}$ and $T_{rg}$ are outputs, the new information about the state, provided by $\dot{T}_{ra}$, is the value of $C_{tf}$ which is a function of $T_{ra}$ and $C_{rc}$. Let us observe that:

$$C_{tf} = \frac{R_{cr}}{R_{cr} + R_{tf}} = 1 - R_{tf} \frac{1}{R_{cr} + R_{tf}},$$

so that,

$$\dot{T}_{ra} = \frac{\Delta H_{cr} R_{tf}^2}{S_c H_{ra}} \frac{1}{R_{cr} + R_{tf}} + \frac{1}{S_c H_{ra}} \left( S_c R_c \left( T_{rg} - T_{ra} \right) \right.$$
$$\left. + S_{tf} R_{tf} \left( T_{tf} - T_{ra} \right) - \left( \Delta H_{fv} + \Delta H_{cr} \right) R_{tf} \right)$$

Denoting the two measured state variables $x_1 = T_{ra}$ and $x_5 = T_{rg}$, the first control variable $u_1 = R_c$ and defining the new state variable $x_2 = \frac{1}{R_{cr}+R_{tf}}$ then

$$\dot{x}_1 = a_2 x_2 + g_1 \left( x_1, T_{rg}, u_1 \right) \tag{5.14}$$

with $a_2 = \frac{1}{S_c H_{ra}} \Delta H_{cr} R_{tf}^2$.

It is clear that (other variables being fixed), the function $C_{rc} \mapsto \frac{1}{R_{tf}+R_{cr}(C_{rc})}$ is a diffeomorphism from any open interval $]\varepsilon, +\infty[$, $\varepsilon > 0$, to its image. Practically, $C_{rc}$ is the concentration of carbon on spent catalyst hence it is a positive variable, and it can be assumed to have a strictly positive lower bound.

In fact, there is more than that. We have **mathematical coherence** of the model with this property, in the sense that the domains $\{C_{rc} > \varepsilon\}$, are **positively invariant under the dynamics**. Let us check this property now.

**Proposition 1.** *Certain domains* $\{C_{rc} > \varepsilon\}$, *for* $\varepsilon$ *small, are positively invariant*

*Proof.* Let us consider both temperatures. Let us assume that $T_{ra} = 273\,\mathrm{K}$ and $T_{rg} \geq 273\,\mathrm{K}$ then using (5.1),

$$\begin{aligned}\frac{1}{R_{tf}} S_c H_{ra} \dot{T}_{ra} &\geq S_{tf}\,(T_{tf} - 273) - (\Delta H_{fv} + \Delta H_{cr}) \\ &= 3140\,(492.8 - 273) - 1.74\,10^5 - 4.65\,10^5 > 0\end{aligned}$$

and also if $T_{ra} \geq 273\,\mathrm{K}$ and $T_{rg} = 273\,\mathrm{K}$ then using (5.8),

$$S_c H_{rg} \dot{T}_{rg} \geq S_a R_{ai}\,(T_{ai} - 273) > 0.$$

Hence $T_{ra}$ and $T_{rg}$ are bounded from below by $273\,\mathrm{K}$. Then we will prove that there exist $\varepsilon_1$ such that $C_{cs} \geq C_{rc} + \varepsilon_1$. Using (5.4,5.13),

$$\frac{d}{dt}\,(C_{sc} - C_{rc}) = -R_c \left(\frac{1}{H_{ra}} + \frac{1}{H_{rg}}\right)(C_{sc} - C_{rc}) + \frac{100}{H_{ra}} R_{cf} + \frac{100}{H_{rg}} R_{cb}$$

but $R_{cf} = R_{cc} + R_{ad}$ and since $0 < R_c^{\min} \leq R_c \leq R_c^{\max}$, $R_{cc}$ can be bounded from below by a positive decreasing function of $C_{rc}$ (recall that $T_{ra} \geq 273\,\mathrm{K}$). Moreover, $R_{cb}$ can be bounded from below by a positive increasing function of $C_{rc}$ (using $T_{rg} \geq 273\,\mathrm{K}$ and $0 < R_{ai}^{\min} \leq R_{ai} \leq R_{ai}^{\max}$ in (5.9) to (5.12)). Therefore, there exist $\varepsilon_1$ such that $\frac{100}{H_{ra}} R_{cf} + \frac{100}{H_{rg}} R_{cb} \geq \varepsilon_1 R_c^{\max} \left(\frac{1}{H_{ra}} + \frac{1}{H_{rg}}\right)$ hence

$$\frac{d}{dt}\,(C_{sc} - C_{rc} - \varepsilon_1) \geq -R_c \left(\frac{1}{H_{ra}} + \frac{1}{H_{rg}}\right)(C_{sc} - C_{rc} - \varepsilon_1)$$

We can chose $\varepsilon_1$ such that $C_{sc} - C_{rc}|_{t=0} \geq \varepsilon_1$ and so $C_{sc} - C_{rc} \geq \varepsilon_1$ along the trajectory.

Finally

$$\begin{aligned}H_{rg} \dot{C}_{rc} &= R_c\,(C_{cs} - C_{rc}) - 100\,R_{cb} \\ &\geq R_c^{\min} \varepsilon_1 - 100\,R_{cb}\end{aligned}$$

and since $R_{cb} \overset{C_{rc} \to 0}{\longrightarrow} 0$ there exist $\varepsilon_2$ such that $R_{cb} \leq \frac{1}{100} R_c^{\min} \varepsilon_1$ if $C_{rc} \leq \varepsilon_2$ hence $H_{rg} \dot{C}_{rc} \geq 0$ for $C_{rc} = \varepsilon_2$ and therefore, $C_{rc}$ is bounded from below by $\varepsilon_2$. □

Due to this analysis, the diffeomorphism $C_{rc} \mapsto x_2$ can be smoothly prolonged to all of $\mathbb{R}$ without changing the trajectories on the physical domain. This property could be important, in order to reach the assumption B that is needed for the construction of our observer (Section 4.1.2). In fact, in practice, we observe in simulations that our estimations of $C_{rc}$ never vanish. Hence, we just don't make any prolongation, and in fact, we don't use this property, which nevertheless is crucial from the theoretical point of view.

Let us remark that our change of variable already depends on the control variable $R_c$ since $x_2 = \frac{1}{R_{cr} + R_{tf}}$ and $R_{cr}$ depends explicitly on $R_c$. Then, as expected, the first derivative of the control will appear after the coordinate-change.

We have to calculate the derivative of $x_2$ with respect to time:

$$\begin{aligned}\dot{x}_2 &= \frac{1}{(R_{cr}+R_{tf})^2}\left(-\dot{R}_{cr}\right)\\ &= -x_2^2\left(\frac{dR_{cr}}{dC_{rc}}\dot{C}_{rc}+\frac{dR_{cr}}{dT_{ra}}\dot{T}_{ra}+\frac{dR_{cr}}{dR_c}\dot{R}_c\right)\\ &= -x_2^2\,(-0.12)\,\frac{R_{cr}}{C_{rc}}\frac{R_c}{H_{rg}}C_{sc}+g_2\,(x_1,x_2,x_5,u_1,\dot{u}_1,u_2)\\ &= a_3x_3+g_2\,(x_1,x_2,x_5,u_1,\dot{u}_1,u_2)\,, \qquad (5.15)\end{aligned}$$

with $x_3 = x_2^2\dfrac{R_{cr}}{C_{rc}}R_cC_{sc}$ and $a_3 = \dfrac{0.12}{H_{rg}}$.

To finish, differentiating once more, we obtain:

$$\begin{aligned}\dot{x}_3 &= \frac{x_3}{C_{sc}}\frac{100\,R_{tf}F_{cf}}{H_{ra}}+g_3\,(x_1,x_2,x_3,x_5,u,\dot{u})\\ &= a_4x_4+g_3\,(x_1,x_2,x_3,x_5,u,\dot{u})\,, \qquad (5.16)\end{aligned}$$

where $x_4 = x_3\dfrac{F_{cf}}{C_{sc}} = x_2^2\dfrac{R_{cr}}{C_{rc}}R_cF_{cf}$ and $a_4 = \dfrac{100\,R_{tf}}{H_{ra}}$.

Finally, our (control depending) change of coordinates is

$$\psi_u\,(T_{ra},C_{rc},C_{sc},F_{cf},T_{rg}) = (x_1,x_2,x_3,x_4,x_5)$$

Here $u$ denotes the control variables $u = (R_c, R_{ai})$ but $\psi_u$ does not depend explicitly on $R_{ai}$.

Again, a certain number of "theoretical precautionary measures" have to be taken: it is easy to check that $\psi_u$ is a smooth function in the interior of the physical domain (positive variables). Therefore, it would be possible to prolong $\psi$ outside this domain in order that it becomes (smoothly) everywhere defined. In practice, we don't do this: we do not prolong $\psi_u$ but we keep in mind that our simulation results are justified only if the state variables remain in the physical domain.

The only problem that may occur -and that occurs in practice- concerns the fact that, **temporarily**, **the estimations of** $C_{rc}$ may vanish or become negative, which may have unpleasant consequences in the other equations. To palliate this difficulty, we just introduce a smooth cut-off function $\chi_\varepsilon$ such that $\chi_\varepsilon$ is one to one from $\mathbb{R}$ into $\left]\frac{\varepsilon}{2},+\infty\right[$ and $\chi_\varepsilon(z) = z$ if $z > \varepsilon$. Then we replace $C_{rc}$ in (5.3) by $\chi_\varepsilon(C_{rc})$ with $\varepsilon$ small enough. This is just an artificial way to correct irrelevant estimations of $C_{rc}$.

Finally, our system is equivalent up to a diffeomorphism to the following system

$$\begin{aligned} y_1 &= x_1, \quad y_2 = x_5\\ \dot{x}_1 &= a_2x_2+g_1\,(x_1,x_5,u)\\ \dot{x}_2 &= a_3x_3+g_2\,(x_1,x_2,x_5,u,\dot{u})\\ \dot{x}_3 &= a_4x_4+g_3\,(x_1,x_2,x_3,x_5,u,\dot{u})\\ \dot{x}_4 &= g_4\,(x)\end{aligned}$$

Then we may apply our "mixed" extended Kalman filter to this system. To do this, we don't care about the theoretical bounds of the tuning parameters $\theta$, $\lambda$ that come from the theory. We just tune these parameters in order to obtain reasonable practical performances: increasing $\theta$ results in better performances in high-gain mode, decreasing $\lambda$ makes the filtering mode (performances w.r.t. noise) be good a long time after the occurrence of large perturbations.

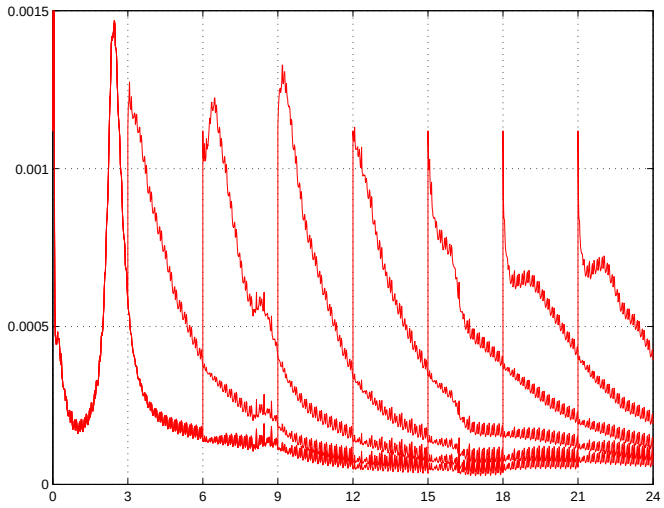


FIGURE 2. Euclidian norm of the gain

5.2.2. *Tuning of parameters.* As we explained, we want to build an observer mixing the good properties of the high-gain EKF with respect to large unmodelled disturbances and the good properties of the classical EKF with respect to noise. In order to achieve this goal, we have to tune each parameter very carefully. let us now explain how to achieve this goal.

(1) As a first step, we just use a single classical EKF (that is $\theta = 1$) and we tune $Q$ and $r$ – in (4.5) – in order to obtain best possible performances with respect to measurement noise. During this first step, we do not simulate disturbances and, of course, we initialize our EKF at the right value of the state. Nevertheless, we choose $Q$ and $r$ such that the EKF reaches also good performances when $\theta$ is slightly larger than 1 (for instance $\theta \approx 2$). As a consequence, when several observers will be working together, a number of them will reach good performances with respect to noise (those among them for which $\theta$ will be close to 1), similar to performances of a classical EKF.

(2) The second step is to tune the high-gain EKF. We use the same matrices $Q$ and $r$ as in the first step and we use $\theta_0$ and $\lambda$ to achieve our purposes, that is fast convergence. We have simulated several "physical" disturbances. Since the rate of convergence might be theoretically arbitrary, we have simulated a very small noise, and asked for a – fast but – reasonable convergence. During this step, we keep in mind that the high-gain EKF should become a classical EKF as fast as possible. Therefore, despite the fact that $\lambda$ should be small enough to ensure exponential convergence, it should not be too small. The price to pay for a too small $\lambda$ will be **a large number of observers**. As in the first step, we check that the performances are reasonable even for slightly lower values than $\theta_0$. We will denote by $\theta_1$

the minimal value of $\theta$ which ensures good performance in presence of disturbances.

(3) The last step consists of using several observers as explained in Section 4.1.3. Since at each current time, we want to have at least an observer working in "high-gain mode", and another one working in "filtering mode", the number of observers will depend on the values obtained for $\theta_0$, $\theta_1$ and $\lambda$. The time between two consecutive initializations of an observer will be the time necessary for $\theta(t)$ to reach $\theta_1$ starting from $\theta_0$ and satisfying $\dot{\theta} = \lambda(1-\theta)$. The number of observers will be high enough in order to be sure that at any time, there exist observers with a current value of $\theta$ almost equal to 1 (at least less than 2, according to first step). It can be useful to plot informations concerning the actual gain value versus the time in order to check that the gain is actually high enough when $\theta$ is high.

Indeed, since the gain is obtained as the solution of a Ricatti equation, and since this equation itself depends on an exponentially converging parameter, it is not at all obvious to get intuition of how the gain is varying. We decided to plot the Euclidian norm of the correction applied to the state, (which is equal to the correction gain times a normalized innovation (equal to 1 K) ), see figure 2. Here, the ratio between high gain and non-high-gain is approximately equal to 15.

Practically, we used 5 observers running in parallel, each of them with a lifetime equal to 15 h (hence we initialize a new observer each 3 h). This "3 hours" is comparable to the average response time of the FCC to perturbations. The initial value of $\theta$ has been set to $\theta_0 = 10$ and we have set $\lambda = 0.27\,\mathrm{h}^{-1}$, such that at any time, there is an observer with a corresponding value of $\theta$ greater than 2. The value $\theta_1 = 5$ looks sufficient to ensure convergence of a high-gain observer for any initial condition (from simulation results).

All simulations below were performed using Simulink® [15].

5.2.3. *Discussion of numerical results.* Figures 3 and 4 represent respectively $T_{ra}$ and $T_{rg}$ measurements. Figures 5, 6 and 7 represent the (unknown) state variables $C_{rc}$, $C_{sc}$ and $F_{cf}$ respectively. We also plotted on figures 3 to 7 the estimation provided by the best observer. On figure 8, we have plotted five curves corresponding to the value of $\theta$ for each observer. Hence, it is simply five exponentially decreasing functions, each of them obtained from the others by a time shift. On the same figure, we have also plotted the value of $\theta$ corresponding to the best observer at current time (see below).

We simulate some measurement-noise on each temperature. We simulate also a disturbance consisting of a ramp on $F_{cf}$, starting at $t = 10\,\mathrm{h}$ from $F_{cf} = 0$ to $F_{cf} = 5.6\,10^{-4}$ at $t = 12\,\mathrm{h}$. This unmodelled and unmeasured disturbance is larger than realistic actual disturbances: the coke formation factor usually varies very slowly. The tracking of the parameter is fast and accurate (figure 7).

Moreover, figure 8 shows that our multiple high-gain extended Kalman filter does exactly what we expected from the theory. Indeed, we have represented on figure 8 the behavior of $\theta(t)$ for each of our five observers (thin lines). At each time, the value of $\theta(t)$ corresponding to the observer with smallest innovation is plotted with a thick line. So it is clear which kind of observer (high-gain or classical extended Kalman filter) has the best performance at each time. At the

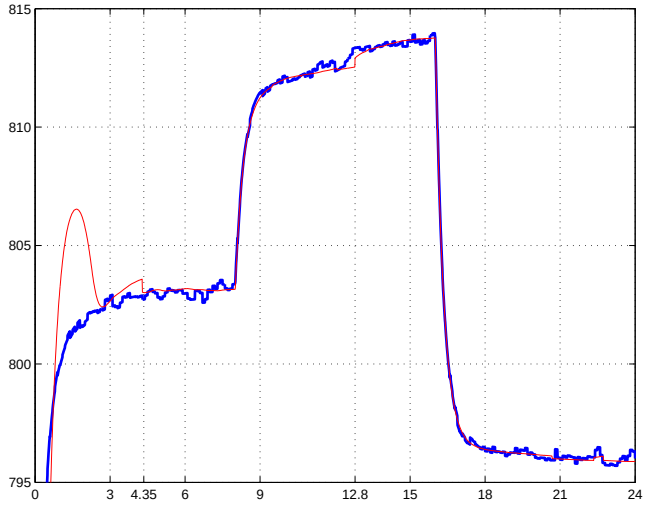


FIGURE 3. $T_{ra}$

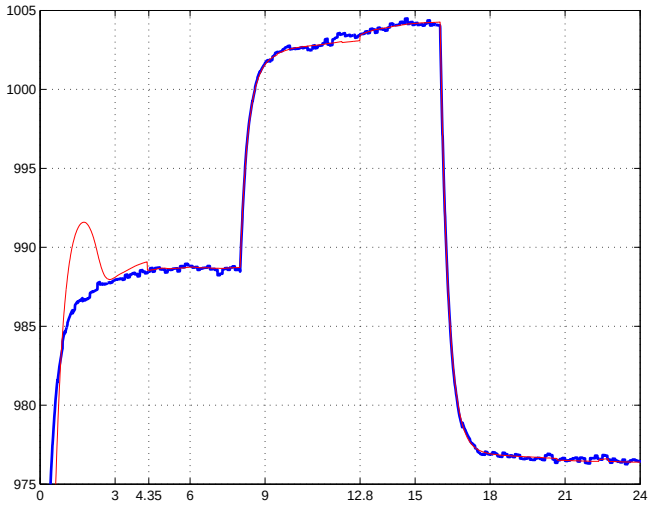


FIGURE 4. $T_{rg}$

beginning, each observer is a high-gain observer. Then, since the model is accurate (no disturbances) the observer with the smallest value of $\theta(t)$ becomes more robust to measurement noise and so has the smallest innovation. When suddenly $F_{cf}(t)$ begins to vary according to a ramp, the observer with highest gain becomes more

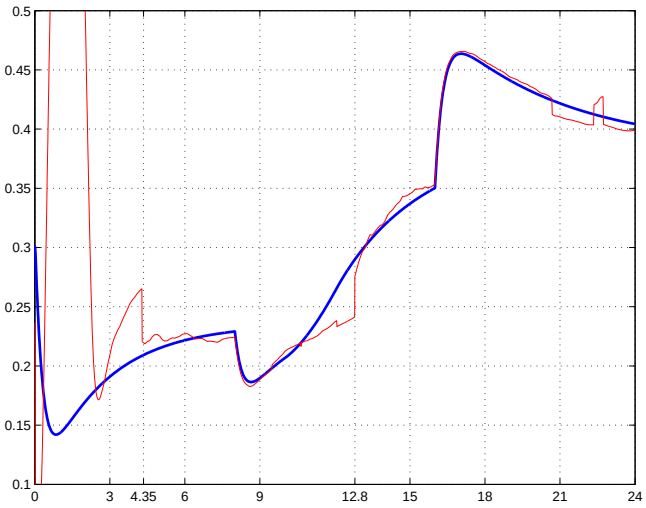


FIGURE 5. $C_{rc}$

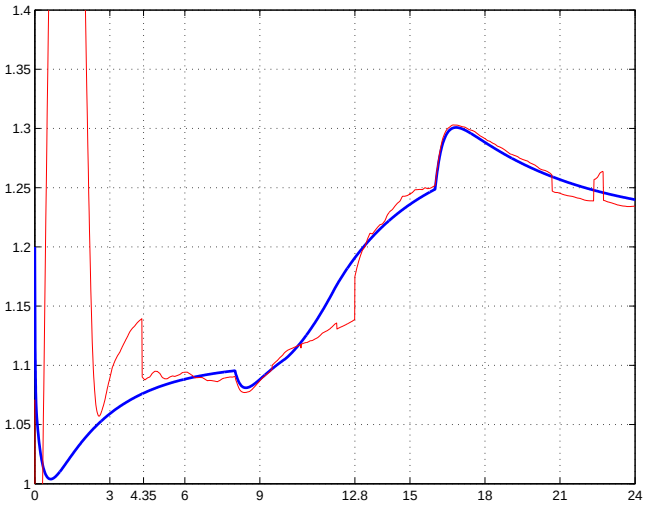


FIGURE 6. $C_{sc}$

accurate and this behavior illustrates the well-known ability of high-gain observers to track the state in presence of unmodelled disturbances. When $F_{cf}(t)$ stabilizes and is correctly estimated, the classical extended Kalman filter becomes better again, thanks to its good (optimal) local properties.

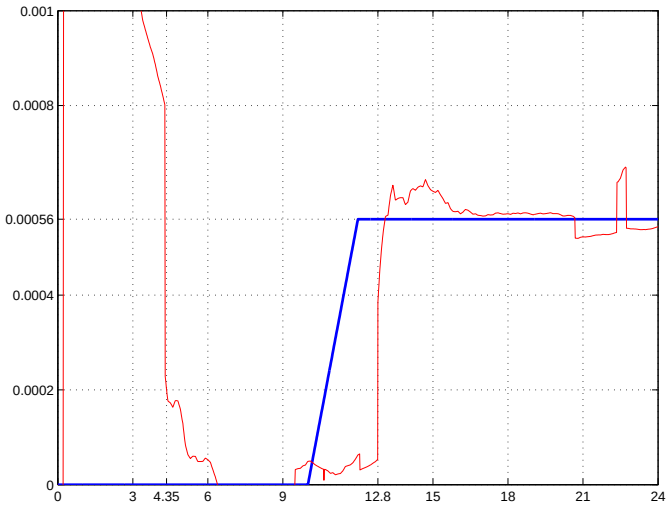


FIGURE 7. $F_{cf}$

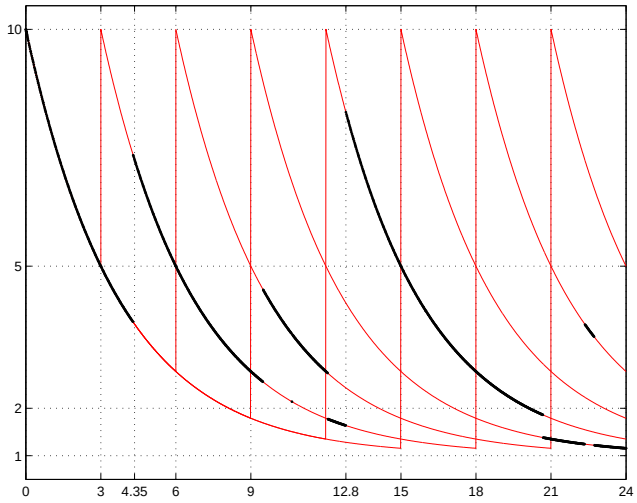


FIGURE 8. $\theta(t)$ and the best observer versus time

## 5.3. **Identification of reaction rate model of oxygen.**

5.3.1. *Identifiability analysis.* Now, we will assume that the function describing the reaction rate of oxygen is unknown. As a matter of fact, this part of the model

is usually very dependant from the FCC unit under consideration, and/or more simply, from the author of the paper.

Indeed, the main difference between the model described in [14] and the simplest one in [13] concerns the oxygen reaction rate model $O_{fg}$ which gives the rate $R_{cb}$ of burning carbon. Therefore, **we will assume that** $K_{or}$ (in Formula 5.12) **is now an unknown function of** $T_{rg}$. We use the same approach as in the previous section to check the (infinitesimal) identifiability of the system. Since $K_{or}$ is now considered as an unknown function, we will use $T_{rg}$ to estimate it:

Mainly, by differentiation, $\dot{T}_{rg}$ allows to compute $R_{cb}$ and $R_{cb}$ gives $O_{fg}$. Here, $O_{fg}$ is a nonlinear function (5.10) of our unknown function $K_{or}$. This function is not defined at $K_{or} = 0$. But, in restriction to the set of values that are physically relevant, it is bijective. Hence, it is possible, and useful to modify it outside a "relevant set", as we shall see below.

Therefore, let us set $K_{or} = \varphi(T_{rg})$, and $R_{cb} = \Phi(C_{rc}, R_{ai}, \varphi(T_{rg}))$ so that (5.15) becomes:

$$\dot{x}_2 = a_3 x_3 + g_2(x_1, x_2, x_5, \Phi(x_2, u, \varphi(x_5)), u, \dot{u}).$$

Starting from this point, it would be easy to transform directly the FCC system into Identifiability canonical form. But, in order to illustrate the intrinsic characterizations given in Section 3 (details in the paper [4]), we will apply the theory to the FCC model.

At first, we have to calculate the two indices, $k$ and $r$, as defined in section 3.3.2. Briefly, let us consider a system of the form:

$$\begin{cases} \dot{x} &= f(x, \varphi) \\ y_1 &= h_1(x, \varphi) \\ y_2 &= h_2(x, \varphi), \end{cases}$$

and recall that:

- $k$ is the first index such that the respective ranks $(N_l)_{l=0,1,\ldots}$ of the family $(E_l)_{l=0,1,\ldots}$ of one-forms
$$E_l = \text{span}\left(\{d_x h_i, d_x L_f h_i, \ldots, d_x L_f^{l-1} h_i, i = 1, 2\}\right)$$
are such that $N_k = 2k$ and $N_{k+1} < 2k + 2$;
- $r$ is the order of the system with respect to $\varphi$ that is the first index such that $d_\varphi L_f^r h_1$ or $d_\varphi L_f^r h_2$ does not vanish identically.

In the FCC case, $h_1(x, \varphi) = T_{ra} = x_1$ and $h_2(x, \varphi) = T_{rg} = x_5$, hence $E_1 = \text{span}(\{dT_{ra}, dT_{rg}\})$. But since the only new state variable appearing in $T_{ra}$ and $T_{rg}$ is $C_{rc}$, then $E_2 = \text{span}(\{dT_{ra}, dT_{rg}, dC_{rc}\})$. As a consequence, $N_1 = 2$ and $N_2 = 3$ therefore $k = 1$. Moreover, thanks to our observability analysis, let us observe that $N_3 = 4$ and $N_4 = 5 = n$.

**Remark 4.** *The order of the system is* $r = 1$ *since* $L_f h_2$ *is a function of* $O_{fg}$ *and therefore of the unknown function* $K_{or}$ *(*$d_\varphi L_f h_2 \neq 0$*). Nevertheless,* $d_\varphi L_f h_1 \equiv 0$.

This remark will be used for the construction of the observer

The system is regular (see Section 3.3.2, or [4]), but we have to look further in order do decide if the system is of type 1, 2 or 3. Here, since $d_x h_1 \wedge d_x h_2 \wedge d_x L_f h_1 \neq 0$ then **hypothesis B.1.** is satisfied hence the system has **type 2**.

More precisely, it may be written under the following form:

$$
\begin{array}{lcllcl}
y_1 & = & x_1 & y_2 & = & x_5 \\
\dot{x}_1 & = & F_1\left(x_1, x_2, x_5, u\right) & \dot{x}_5 & = & \Phi\left(x, \varphi\right) \\
\dot{x}_2 & = & F_2\left(x_1, x_2, x_3, x_5, \Phi\left(x, \varphi\right), u\right) & & & \\
\dot{x}_3 & = & F_3\left(x_1, x_2, x_3, x_4, x_5, \Phi\left(x, \varphi\right), u\right) & & & \\
\dot{x}_4 & = & F_4\left(x, \varphi, u\right) & & &
\end{array}
\tag{5.17}
$$

with $\frac{\partial \Phi}{\partial \varphi} \neq 0$, $\frac{\partial F_1}{\partial x_2} \neq 0$, $\frac{\partial F_2}{\partial x_3} \neq 0$, $\frac{\partial F_3}{\partial x_4} \neq 0$.

Notice that this canonical form has a particularity: $\dot{x}_1$ does not depend of $\Phi$ since $d_\varphi L_f h_1 \equiv 0$.

5.3.2. *Model for identification.* We will estimate the function $\Phi$. In order to construct our exponentially converging observer of "mixed" high-gain extended Kalman filter type, we will use a "local" second order model for $\Phi$ that is to say we will assume that $\frac{d^3\Phi}{dt^3} = 0$ (locally: on reasonable time intervals, $\Phi$ is accurately approximated by a 2rd order polynomial, which doesn't mean that it is globally a third order polynomial).

Hence we add three state variables to the original state variables, $x_6 = \Phi\ (= R_{cb})$, $x_7 = \dot{\Phi}$ and $x_8 = \ddot{\Phi}$ such that $\dot{x}_6 = x_7$, $\dot{x}_7 = x_8$ and $\dot{x}_8 = 0$.

It has to be noticed that this local model becomes not valid when the control variable $R_{ai}$ has jumps, since $R_{ai}$ appears both in $R_{cb}$ as a function of $O_{fg}$ and in $O_{fg}$ as a nonlinear function of $K_{or}$. We will tune the model in order to obtain very fast convergence so that this kind of disturbance has no large effect on estimation.

In order to retrieve $\varphi$ from $\Phi$, the function $\varphi \mapsto \Phi\left(x, \varphi\right)$ should be one to one. More precisely, the following function

$$
z \mapsto \exp\left(\frac{-a}{b + \frac{c}{z}}\right),
$$

has to be modified out of the set of "physical values", in such a way that it becomes a smooth diffeomorphism for any positive value of the parameters $a$, $b$ and $c$, bounded from below. We leave the reader to check that this function can be smoothly modified outside any domain $\{A > z > 0, A \text{ large}\}$ in order to obtain a global diffeomorphism from $\mathbb{R}$ to $\mathbb{R}$.

Moreover, we observed that, during simulations, the observer internal variables remain in the physical state domain except at the very beginning of the simulation, so that, again, (and this seems to be a general fact in that type of applications), our carefulness is not justified here. Brute force computation appears to be enough, in most cases.

Finally, in agreement with the theory developed in the first part of this paper, our system can be written globally,

$$
\begin{array}{lcllcl}
y_1 & = & x_1 & y_2 & = & x_5 \\
\dot{x}_1 & = & a_2 x_2 + g_1\left(x_1, x_5, u\right) & \dot{x}_5 & = & x_6 + g_5\left(x_1, x_5, u\right) \\
\dot{x}_2 & = & a_3 x_3 + g_2\left(x_1, x_2, x_5, x_6, u, \dot{u}\right) & \dot{x}_6 & = & x_7 \\
\dot{x}_3 & = & a_4 x_4 + g_3\left(x_1, x_2, x_3, x_5, u, \dot{u}\right) & \dot{x}_7 & = & x_8 \\
\dot{x}_4 & = & g_4\left(x\right). & \dot{x}_8 & = & 0
\end{array}
\tag{5.18}
$$

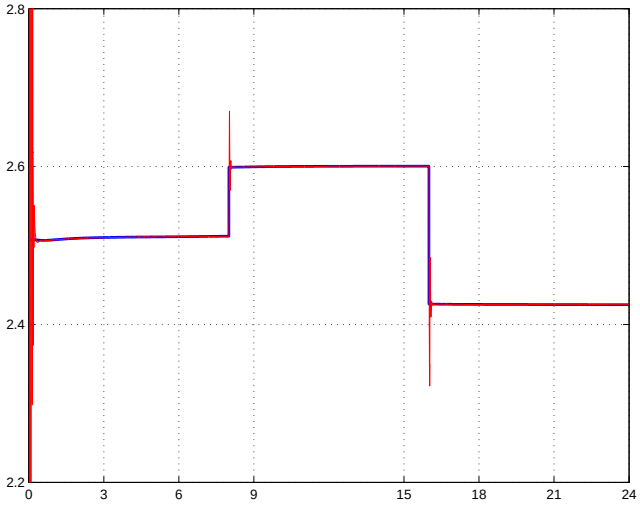


FIGURE 9. $R_{cb}$

5.3.3. *Identification algorithm.* As we said, the system (5.17) is clearly in the type 2 identifiability normal form. However, we cannot apply our (mixed Kalman) observer directly on the system (5.18) because $x_6$ appears in $\dot{x}_2$ which is not allowed in our method. To overcome this (small) difficulty, we decided to proceed as follows.

We consider the two following systems $(\Sigma_1)$ and $(\Sigma_2)$ independently

$$(\Sigma_2)\begin{cases} y_2 &= \xi_1 \\ \dot{\xi}_1 &= \alpha_1(u)\,\xi_1 + \alpha_2\xi_2 + \alpha_3(u)\,y_1 + \beta(u) \\ \dot{\xi}_2 &= \xi_3 \\ \dot{\xi}_3 &= \xi_4 \\ \dot{\xi}_4 &= 0, \end{cases}$$

where the equation of $\dot{\xi}_1$ comes immediately from (5.8), and

$$(\Sigma_1)\begin{cases} y_1 &= x_1 \\ \dot{x}_1 &= a_2x_2 + g_1(x_1, y_2, u) \\ \dot{x}_2 &= a_3x_3 + g_2(x_1, x_2, \xi_1, \xi_2, u, \dot{u}) \\ \dot{x}_3 &= a_4x_4 + g_3(x_1, x_2, x_3, \xi_1, u, \dot{u}) \\ \dot{x}_4 &= g_4(x) \end{cases}$$

The observer of Subsystem $(\Sigma_2)$ will be mainly an approximate derivator of $\xi_2 = \Phi$ $(= R_{cb})$.

But $(\Sigma_2)$ is very simple: it is linear, time dependant. This will allow us to "filter" the output $T_{rg}$ and its derivative, using a **standard linear Kalman filter**. This filter will provide an **accurate estimation** of $\xi_1 = T_{rg}$ and $\xi_2 = \Phi$.

We will use directly this estimation of $(T_{rg}, \Phi)$ inside $(\Sigma_1)$, to which we will apply the (mixed) observer of Section 4.1.3, just considering $\xi_1$ and $\xi_2$ as new inputs.

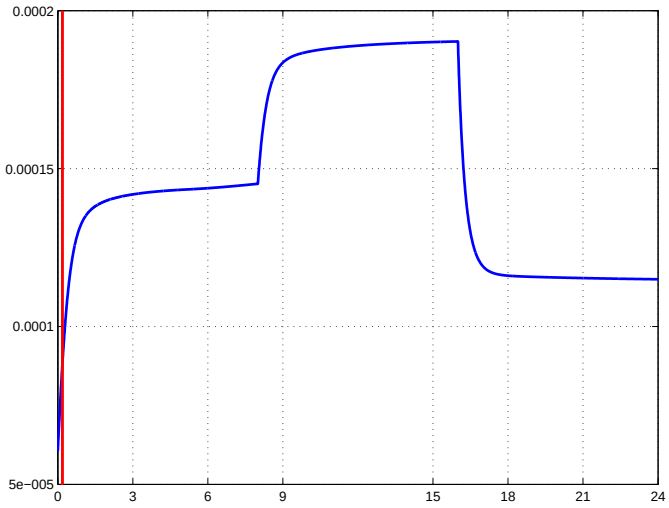


FIGURE 10. $K_{or}$

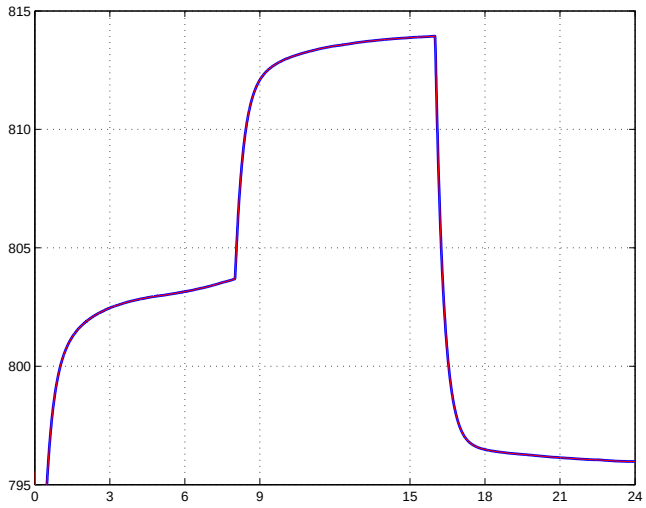


FIGURE 11. $T_{ra}$

**Remark 5.** *For general results concerning that type of cascade systems, and on the way to apply exponentially converging high-gain observers (mostly Luenberger-type) to cascade systems, see* [17].

5.3.4. *Simulation.* We used exactly the same scenario as in Section 5.2.

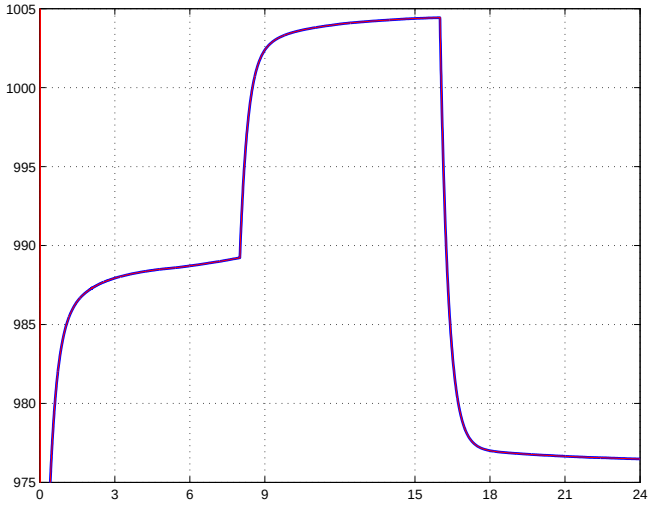


FIGURE 12. $T_{rg}$

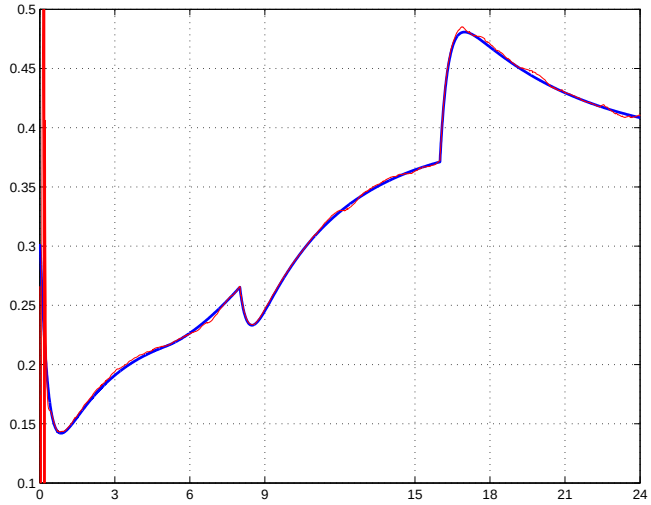


FIGURE 13. $C_{rc}$

Estimation of $R_{cb}$ is not a hard task, even if it requires one derivative of $y_2 = T_{rg}$. But the estimation of $K_{or}$ requires a very good estimation of $R_{cb}$: this is due to the high "sensitivity" of the nonlinear function $R_{cb}(K_{or})$ (i.e. large values of the derivative of $R_{cb}$ w.r.t. $K_{or}$).

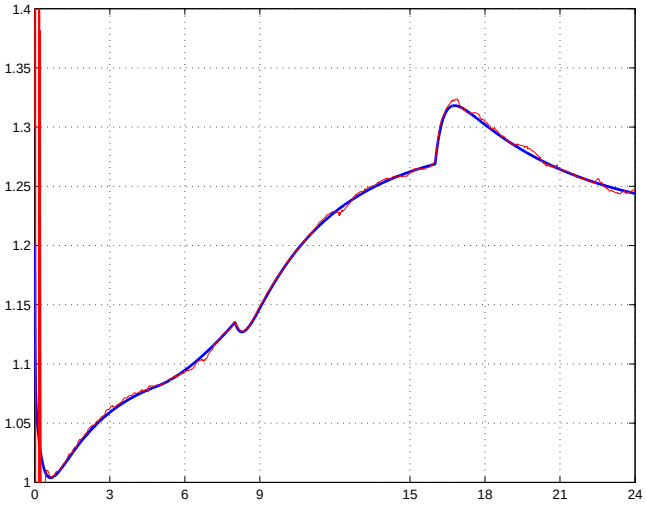


FIGURE 14. $C_{sc}$

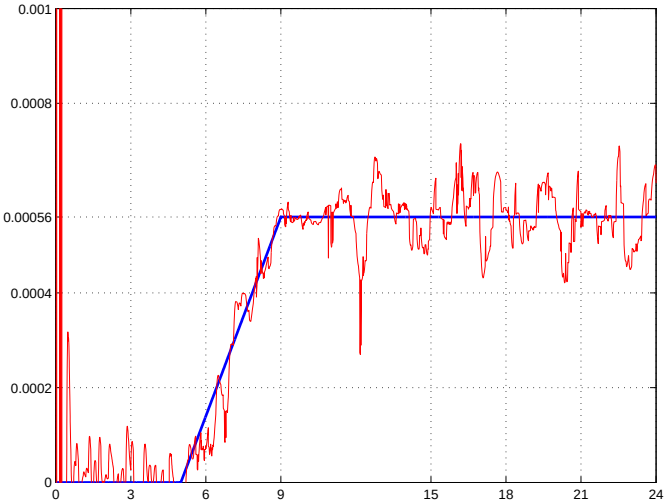


FIGURE 15. $F_{cf}$

The model – supposed to be unknown – used for simulation is the equation (5.12), coming from [14]. Figures 11 to 15 represent the state estimation.

The noise being not too large, estimation of $F_{cf}$ is not hard. Therefore, we were able to tune our observers for they estimate faster than in Section 5.2. The result

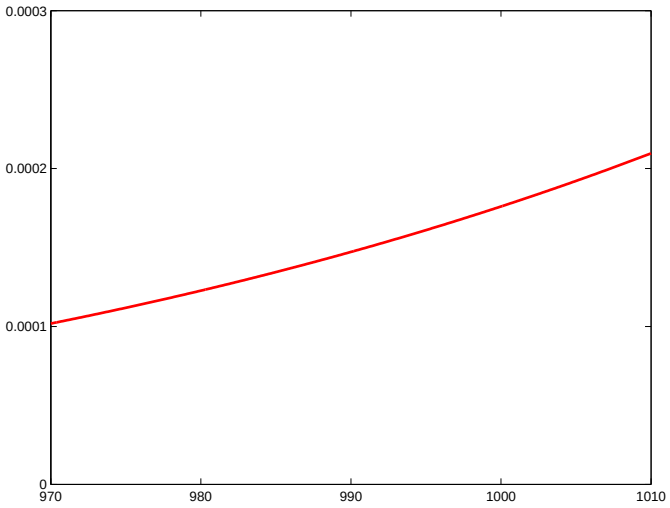


FIGURE 16. $K_{or}$ as a function of $T_{rg}$, $t = 0\,\mathrm{h}$

of this new choice of parameters appears clearly on figure 15, where one can see the very quick estimation of $F_{cf}$ (comparing to figure 7).

We also plotted the estimation (coming from a standard linear Kalman filter, as we said) of $R_{cb}$ on figure 9 and the estimation of $K_{or}$ resulting from the estimation of $R_{cb}$, on figure 10.

We have plotted the results of the identification at time 0 h, 1 h, 9 h and 17 h on figures 16 to 19. The continuous line is the actual function $K_{or}$ versus $T_{rg}$. The dotted line represents the estimation of $K_{or}$ as a function of $T_{rg}$. The figure 16 represents the unknown function to be identified.

On figure 17, we have no information about this function for temperatures larger than 985 K, because $T_{rg}(t)$ does not pass beyond 985 K. However, the function has been identified between 970 K and 985 K. This interval represents the range of regenerator temperatures during the first hour of operation, as it can be seen on figure 12. After 9 h, the function has been identified between 970 K and 1005 K (figure 18). At the end of the simulation (figure 19), the function has been also identified between 970 K and 1005 K, that is for each values reached by $T_{rg}$ during the simulation. It can be seen on figure 19 that there are two values for which $K_{or}(T_{rg})$ is very badly estimated (around $T_{rg} = 990\,\mathrm{K}$ and $T_{rg} = 1005\,\mathrm{K}$). These values correspond to discontinuities of $R_{ai}$ : the estimation of $R_{cb}$ becomes bad during a short transient.

An accurate estimation of the unknown function can be obtained by some regularization of the informations collected during simulation. We just applied some outliers removal procedure followed by some smoothing procedure, in order to obtain the estimation shown on figure 20.

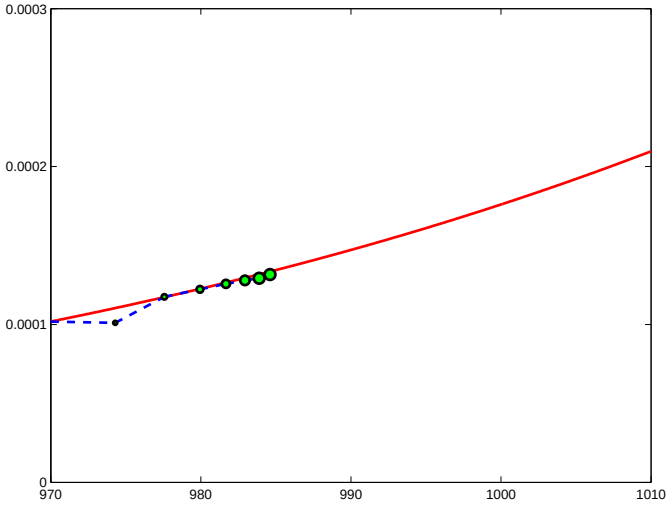


FIGURE 17. $K_{or}$ as a function of $T_{rg}$, $t = 1\,\mathrm{h}$

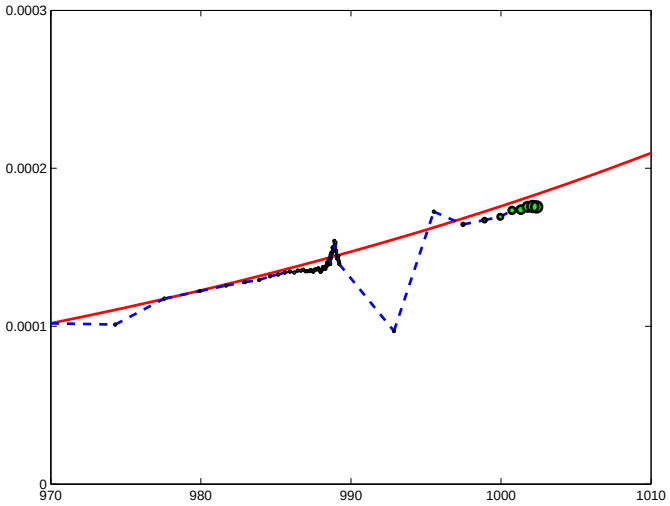


FIGURE 18. $K_{or}$ as a function of $T_{rg}$, $t = 9\,\mathrm{h}$

## 6. Comments, conclusion, and future research

6.1. **Conclusion.** In this paper, we have explained and compared our theoretical results for observability and identification.

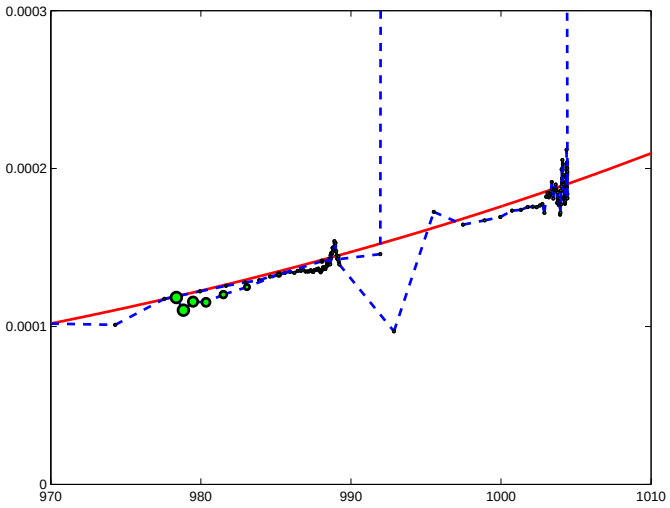


FIGURE 19. $K_{or}$ as a function of $T_{rg}$, $t = 17\,\mathrm{h}$

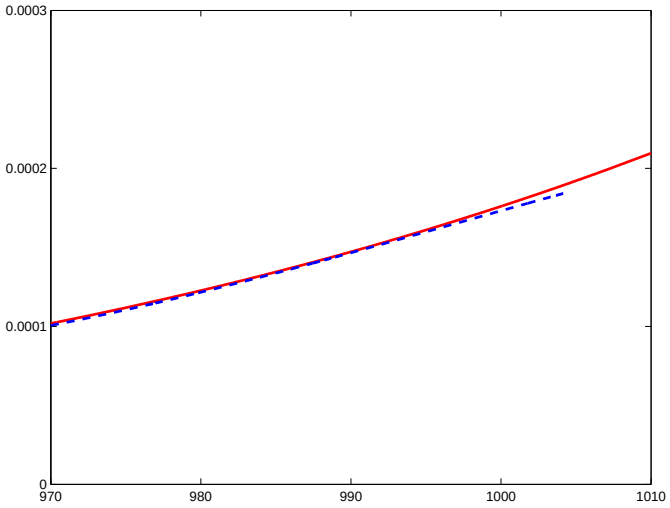


FIGURE 20. $K_{or}$ as a function of $T_{rg}$ after smoothing

These theories are parallel, leading to the same type of observability (resp. identifiability) results. In particular, systems fall in two classes: those that are generically observable (resp. identifiable) corresponding to a large number of outputs, and those with a small number of outputs, for which the observability property

(resp. identifiability) is on the contrary completely nongeneric (an infinite codimension property, in fact). For all the observable (resp. identifiable) systems, we have developped a practical observer (resp. identifier), based upon the classical "extended Kalman filter". We have shown a non–academic practical application of our theory, to a system from petroleum industry.

Except a few technical improvements, this paper contains no new general result: we try simply to explain our methodology, and to illustrate it.

Now, after exposition of these results, we would like to focus on some "philosophical question".

6.2. **Functional identification versus parametric identification.** Our approach to identification, in this paper, is "**functional identification**" (identification of functions). This concept has to be faced with the concept of "**parametric identification**". In fact, as a final step of our "functional identification" analysis, we perform parametric identification via a "local model" of the unknown function $\varphi(t)$. This local model is in general polynomial as explained in section 4.2 but, everything that we do is firstly justified by a "functional identifiability" property.

By "parametric identification", we mean parametrizing the unknown function by a finite number of parameters, in order to transform the problem into a finite dimensional observation problem, or simply into a finite dimensional (nonlinear) optimisation problem.

Parametric identification has been used for long in practice by many people: typing on a reasonable data–base (Zentralblatt for instance) the words "parametric identification", we get a lot of references (118), in which "parametric identification" is understood in this sense.

It is also a very classical approach to use the extended Kalman filter exactly in the way we do it, to perform observation and parametric identification. There are so many references, for practical examples of this approach, that we cannot be exhaustive, and hence we will not cite anyone.

What we want to say here is the following statement, that we feel extremely important: *without precise identifiability analysis, parametric identifiability may be very dangerous !*

We explain this now on the basis of a naive example, but in fact, the idea is not naive at all.

6.3. **An example.** For the sake of simplicity in the exposition, we have chosen a very academic example. Consider

$$(\Sigma)\begin{cases} \dot{x} &= \varphi(x) \\ y &= x+\varphi(x) \end{cases}, x \in \mathbb{R}$$

This system is of course not identifiable (neither identifiable, nor differentially identifiable, nor infinitesimally identifiable). In fact, in that case, all these notions are equivalent because, if we set $\hat{\varphi}(t) = \varphi(x(t))$, the mapping

$$P_\Sigma : (x_0, \hat{\varphi}(\cdot)) \mapsto y(\cdot)$$

is linear.

Let us make the exemple more concrete: set $\psi(\varphi) = -\varphi - \log(\varphi - 1)$, $\psi : ]1, +\infty[ \to \mathbb{R}$ is a diffeomorphism onto $\mathbb{R}$. Set $\tilde{\varphi}(x) = \psi^{-1}(x)$, $\tilde{\varphi} : \mathbb{R} \to ]1, +\infty[$, and finally set $\varphi^*(x) = \tilde{\varphi}(x + \psi(2))$.

Consider $\Sigma_1$ (resp. $\Sigma_2$) the system $\Sigma$ corresponding to the choice $\varphi(\cdot) = 1$ (resp. $\varphi(\cdot) = \varphi^*(\cdot)$). Direct analysis shows that both systems $\Sigma_1$ and $\Sigma_2$ are very well defined: they are "forward–complete", i.e. all trajectories $t \mapsto x(t)$ are defined on full intervals of the form $[0, +\infty]$.

Consider for $\Sigma_1$ (resp. $\Sigma_2$) the choice $x_0 = 1$ (resp. $x_0 = 0$). An elementary computation shows that both outputs $y_1$ and $y_2$ are equal: $y_1(t) = y_2(t) = t + 2$ for all $t \geq 0$. Then $\Sigma$ is not identifiable (in any of the three senses).

Another less explicit but simpler way to see what happens is to consider the kernel $K$ of the (linear) input–state-output map $P_\Sigma$ and to check that $K$ is the one–dimensional set $K = \{(x_0, -\exp(-t)\, x_0), x_0 \in \mathbb{R}\}$. Because of this, the system $\Sigma$ is in fact **identifiable in the class of polynomial functions** $\hat{\varphi}(t)$ **!**

In this example, our methodology for identification becomes specially natural: if we assume a "local model" of polynomial type: $\frac{d^k \hat{\varphi}(t)}{dt^k} = 0$ for an arbitrary $k$, we get a linear system, observable, with extended state $\left(x, \varphi(0), \dot{\varphi}(0), \ldots, \varphi^{(k-1)}(0)\right)$. Then, it is not even necessary to use a high gain–non high gain extended Kalman filter: the standard linear Kalman filter will give perfect estimation of $\hat{\varphi}(t)$. Would the system be $\Sigma_1$ or $\Sigma_2$, we would get $\hat{\varphi}(t) = 1$ for all $t$. With small noise, the estimation will be almost perfect, giving confidence in the result. But, if the system is $\Sigma_2$, this result is in fact completely false !

In our approach, the step of "parametric identification" occurs in the same way as in this example, when we specify the polynomial local model for $\hat{\varphi}(\cdot)$. But, the "functional identifiability" assumption prevents from the occurence of the phenomenon described in the example.

6.4. **Future work.** First, the general situation (more than a single function to identify) has to be studied. We already have two kind of results on this subject:

(1) Identifiability is generic if the number of outputs is larger than or equal to the number of controls plus the number of functions to identify plus two;
(2) There is a finite number of identifiability normal forms in the nongeneric case (this is much harder than in the case of a single function).

Theses points will be the purpose of a forthcoming paper.

Besides this study, as for observability, we have results concerning normal forms in the generic situation (in that case, systems can be globally embedded into systems in a very special canonical form). This study is the purpose of [3].

Finally of course, several other applications of our theories and techniques have to be developped.

## References

[1] Abraham, R., and Robbin, J., 1967, Transversal Mappings and Flows, W. A. Benjamin, Inc.

[2] Busvelle, E., Deza, F., Mokrani, F., and Hapiak, S. 1992, Application of a nonlinear observer to a fluid catalytic cracker, NOLCOS'92, Juin 1992 Bordeaux, France

[3] Busvelle, E., and Gauthier, J.-P., 2004, New results on identifiability of nonlinear systems, 2nd IFAC Symposium on System, Structure and Control, December 8–10, Oaxaca, Mexico, to appear.

[4] Busvelle, E., and Gauthier, J.-P., 2003, On determining unknown functions in differential systems, with an application to biological reactors., *COCV*, **9**, 509-552.

[5] Busvelle, E., and Gauthier, J.-P., 2002, High-Gain and Non High-Gain Observers for nonlinear systems, In Contemporary Trends in Nonlinear Geometric Control Theory, (World Scientific, Anzaldo-Meneses, Bonnard, Gauthier, Monroy-Perez, editors), pp. 257-286.
[6] Deza, F., 1991, Contribution to the synthesis of nonlinear observers, PHD Thesis, INSA de Rouen, France.
[7] Gauthier, J.-P., Hammouri, H., and Othman, S., 1992, A simple observer for nonlinear systems. Applications to Bioreactors. *IEEE Trans. Aut. Control*, **37**, 875-880.
[8] Gauthier, J.-P., and Kupka, I., 2001, *Deterministic Observation Theory and Applications* (Cambridge University Press).
[9] Hammouri, H., and Farza, M., 2003, Nonlinear observers for locally uniformly observable systems, *COCV*, **9**, 343-352.
[10] Hirsh, M. W., 1976, Differential topology, Springer–Verlag, Graduate Texts in Maths.
[11] Isidori, A., 1995, Nonlinear Control Systems, Springer–Verlag, Communications and Control Engineering Series, Third edition
[12] Kurihara, H., 1967, Optimal control of fluid catalytic cracking processes, Sc. D. Dissertation, M.I.T. Cambridge, MA.
[13] Lee, W., and Kugelman, A. M., 1973, Number of steady–state operating points and local stability of open–loop fluid catalytic cracker, *Ind. Eng. Chem. Process Des. Develop.*, **12**, N° 2, 197–204.
[14] McFarlane, R. C., and Bacon, D. W., 1989, Adaptive optimizing control of multivariable constrained chemical processes. 2. Application studies, *Ind. Eng. Chem. Res.*, **28**, N° 12, 1834–1845.
[15] Matlab® and Simulink®, The MathWorks, Inc., http://www.mathtools.net/MATLAB/-index.html
[16] Picard, J., 1991, Efficiency of the extended Kalman filter for nonlinear systems with small noise, *SIAM Journal on Appl. Math.*, **51**, 3, 843-885.
[17] Shim, H., Son, Y. I., and Seo, J. H., 2001, Semi-global observer for multi-output nonlinear systems, *Systems Control Lett.* **42** 233–244.

LE2I, UMR CNRS 5158,, Université de Bourgogne, Aile des Sciences de l'Ingénieur, BP 47870 - 21078 Dijon Cedex, France
*E-mail address*: busvelle@u-bourgogne.fr, gauthier@u-bourgogne.fr