united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

SMR1327/4

Summer School on Mathematical Control Theory

(3 - 28 September 2001)

A Course on Observability

J.P. Gauthier
Laboratoire d'Analyse Appliquée et Optimisation
Université de Bourgogne
B.P. 47870
21078 Dijon
France

A Course on Observability.

J.P. Gauthier

Author address:

UNIVERSITÉ DE BOURGOGNE, LABORATOIRE D'ANALYSE APPLIQUÉE ET OPTIMISATION, BP 47870, 21078-DIJON, FRANCE

ABSTRACT. The purpose of this course is to present a set of concepts and results on observability, on the problem of synthesis of observers, and on dynamic output stabilization (using observers).

Most of the results presented here come from the book [15], and from the paper [6].

In this text, in general, we give no proof of the results: we just present and discuss definitions, state results and explain some practical methodology. However, the ideas of the proofs will be explained during the course.

For the last part of the course, (mostly the paper [6]), we give detailed proofs. These proofs, that are computational, contain some proofs on the construction of observers, that are presented in the Chapter 5.

Prevention on place (and time) does not allow us to present many real applications: there will be only one. But academic examples will be discussed along the course. They correspond to the exercises in the text. All the exercises have a detailed solution in the book [15].

I dedicate this course to Andreas Baader.

Contents

CHAPTER 1

Observability concepts.

In this chapter, we will state and explain the various definitions of observability that will be used in the course.

1. Systems under consideration.

We are concerned with general nonlinear systems of the form:

$$(\Sigma)\quad \begin{array}{l} \frac{dx}{dt} = f(x,u), \\ y = h(x,u), \end{array} \tag{1}$$

typically denoted by Σ, where x, **the state**, belongs to X, a $n-$ dimensional, connected, Hausdorff, paracompact differentiable manifold, y, **the output**, takes values in R^{d_y}, u, **the control variable**, takes values in $U \subset R^{d_u}$. For the sake of simplicity of the exposition, we take $U = R^{d_u}$ or $U = I^{d_u}$, where $I \subset R$ is a closed interval. But typically, U could be any closed submanifold of R^{d_u} with a boundary, with nonempty interior, and possibly with corners. Unless explicitly stated, X has no boundary.

The set of typical systems will be denoted by $S = F \times H$, where F is the set of $u-$parametrized vector fields f, and H is the set of functions h. In general, **except explicit mention of the contrary**, f and h are C^∞. But, depending on the context, we will have to consider also analytic systems (C^ω), or C^r systems, for some $r \in \mathcal{N}$. Thus, if necessary, the required degree of differentiability will be stated, but, in most cases, the notations will remain S, F, H.

The simplest case is the case when U is empty, the so called "**uncontrolled case**". In that situation, we will be able to prove more results than in the general case.

Usually, in practical situations, the output function h of the system does not depend on u. Unfortunately, from the theoretical point of view, this assumption is very awkward. Making it leads to clumsy statements. For that reason, we will currently assume that h depends on the control u.

2. Infinitesimal and Uniform Infinitesimal Observability.

The space of control functions under consideration will just be the space $L^\infty[U]$ of all measurable bounded, $U-$ valued functions $u : [0, T_u[\to U$, defined on semi-open intervals $[0, T_u[$ depending on u. The space of our output functions will be the space $L[R^{d_y}]$ of all measurable functions $y : [0, T_y[\to R^{d_y}$, defined on the semi-open intervals $[0, T_y[$. Usually, input and output functions are defined on closed intervals. But this is irrelevant. The following considerations led us to work with semi-open intervals: for any input $\hat{u} \in L^\infty[U]$, and any initial state x_0, the maximal solution of the Cauchy problem for positive times:

$$\frac{d\hat{x}}{dt} = f(\hat{x}(t), \hat{u}(t)), \hat{x}(0) = x_0,$$

is defined on a semi-open interval $[0, e(\hat{u}, x_0)[$, where $0 < e(\hat{u}, x_0) \leq T_{\hat{u}}$. If $e(\hat{u}, x_0) < T_{\hat{u}}$, then, $e(\hat{u}, x_0)$ is the positive escape-time of x_0 for the time dependant vector field $f(., \hat{u}(t))$. It is well known that, for all $\hat{u} \in L^\infty[U]$, the function $x_0 \to e(\hat{u}, x_0) \in \bar{R}_+^*$ is lower semi-continuous. ($\bar{R}_+^* = \{a | 0 < a \leq \infty\}$).

DEFINITION 2.1. *The input-output mapping P of Σ is defined as follows:*

$$P : L^\infty[U] \times X \to L[R^{d_y}], (\hat{u}, x_0) \to P(\hat{u}, x_0),$$

where $P(\hat{u}, x_0)$ is the function $\hat{y} : [0, e(\hat{u}, x_0)[\to R^{d_y}$ defined by

$$\hat{y}(t) = h(\hat{x}(t), \hat{u}(t)).$$

The mapping $P_{\hat{u}} : X \to L[R^{d_y}]$ is $P_{\hat{u}}(x_0) = P(\hat{u}, x_0)$.

DEFINITION 2.2. [1] *A system is called **observable** if for any triple $(\hat{u}, x_1, x_2) \in L^\infty[U] \times X \times X$, $x_1 \neq x_2$, the set of all $t \in [0, \min(e(\hat{u}, x_1), e(\hat{u}, x_2))[$ such that $P(\hat{u}, x_1)(t) \neq P(\hat{u}, x_2)(t)$ has positive measure.*

Now, we define the "first variation" of Σ, or the "lift of Σ on TX". The mapping $f : X \times U \to TX$ induces the partial tangent mapping $T_X f : TX \times U \to TTX$ (tangent bundle of TX). Then, if ω denotes the canonical involution of TTX, (see [1]), $\omega \circ T_X f$ defines a parametrized vector field on TX, also denoted by $T_X f$. Similarly, the function $h : X \times U \to R^{d_y}$ has a differential $d_X h : TX \times U \to R^{d_y}$. The first variation of Σ is the input-output system:

$$(T\Sigma) \begin{cases} \frac{d\xi}{dt} = T_X f(\xi, u) = T_X f_u(\xi), \\ \eta = d_X h(\xi, u) = d_X h_u(\xi). \end{cases} \tag{2}$$

Its input-output mapping is denoted by dP, and the trajectories of (1) and (2) are related as follows:

If $\xi : [0, T_\xi[\to TX$ is a trajectory of (2) associated with the input $\hat{u}$, the projection $\pi(\xi) : [0, T_\xi[\to X$ is a trajectory of Σ associated with the same input. Conversely, if $\varphi_t(x_0, \hat{u}) : [0, e(\hat{u}, x_0)[\to X$ is the trajectory of Σ starting from x_0 for the input $\hat{u}$, the map $x \to \varphi_\tau(x, \hat{u})$ is a diffeomorphism from a neighbourhood of x_0 onto its image, for all $\tau \in [0, e(\hat{u}, x_0)[$. Let $T_X \varphi_\tau : T_{x_0} X \to T_z X$, $z = \varphi_\tau(x_0, \hat{u})$ be its tangent mapping. Then, for all $\xi_0 \in T_{x_0} X$:

$$e_{T\Sigma}(\hat{u}, \xi_0) = e_\Sigma(\hat{u}, \pi(\xi_0)) = e_\Sigma(\hat{u}, x_0),$$

and, for almost all $\tau \in [0, e(\hat{u}, x_0)[$:

$$dP(\hat{u}, \xi_0)(\tau) = d_X h(T_X \varphi_\tau(\hat{u}, \xi_0), \hat{u}(\tau)) = d_X(P^\tau_{\Sigma, \hat{u}})(\xi_0). \tag{3}$$

The right-hand side of these equalities (3) is the differential of the function $P^\tau_{\Sigma, \hat{u}} : V \to R^{d_y}$, where V is the open set:

$V = \{x \in X | 0 < \tau < e(\hat{u}, x)\}$, and $P^\tau_{\Sigma, \hat{u}}(x) = P(\hat{u}, x)(\tau)$.

---

[1] In nonlinear control theory, the notion of observabillity defined here, is usually referred to as "uniform observability". Let us stress that it is just the old basic observability notion used for linear systems.

For any $a > 0$, let $L^\infty_{loc}([0,a[;R^{d_y})$ denote the space of measurable functions $v : [0,a[\to R^{d_y}$ which are locally in L^∞. For all $\hat{u} \in L^\infty(U)$, $x_0 \in X$, the restriction of dP to $\{\hat{u}\} \times T_{x_0}X$ defines a linear mapping:

$$dP_{\hat{u},x_0} : T_{x_0}X \to L^\infty_{loc}([0, e(\hat{u},x_0)[;R^{d_y}),$$

$$dP_{\hat{u},x_0}(\xi_0)(t) = dP(\hat{u},\xi_0)(t). \tag{4}$$

DEFINITION 2.3. *The system Σ is called infinitesimally observable at $(\hat{u},x) \in L^\infty[U] \times X$ if the linear mapping $dP_{\hat{u},x}$ is injective. It is called infinitesimally observable at $\hat{u} \in L^\infty[U]$ if it is infinitesimally observable at all pairs $(\hat{u},x)$, $x \in X$, and* ***uniformly infinitesimally observable*** *if it is infinitesimally observable at all $\hat{u} \in L^\infty[U]$.*

REMARK 2.1. *In view of the relation 3 above, the fact that the system is infinitesimally observable at $\hat{u} \in L^\infty[U]$ means that the mapping $P_{\hat{u}} : X \to L[R^{d_y}]$ is an immersion of X into $L[R^{d_y}]$ (as was stated, $P_{\hat{u}}$ is differentiable in the following sense: we know that $e(\hat{u},x) \geq e(\hat{u},x_0) - \varepsilon$ in a neighbourhood U_ε of x_0. Then $P_{\hat{u}}$ is differentiable in the classical sense from U_ε into $L^\infty([0,e(\hat{u},x_0)-\varepsilon];R^{d_y})$. $P_{\hat{u}}$ is an immersion in the sense that these differential maps are injective).*

This notion of uniform infinitesimal observability is the one which makes sense in practice, when $d_y \leq d_u$. In most of the examples from the real life we know of, when $d_y \leq d_u$, the system is uniformly infinitesimally observable.

3. The Canonical Flag of Distributions.

In this section, we assume that $d_y = 1$. As above, set: $h_u(x) = h(x,u)$, $f_u(x) = f(x,u)$.

Associated to the system Σ, there is a family of flags $\{D_0(u) \supset D_1(u) \supset ... \supset D_{n-1}(u)\}$ of distributions on X (parametrized by **the value** $u \in U$ of the control):

$D_0(u) = \mathrm{Ker}(\, d_X h_u(x))$, where d_X denotes again the differential with respect to the x variable only. For $0 \leq k < n-1$:

$$D_{k+1}(u) = D_k(u) \cap Ker(d_X(L^{k+1}_{f_u}(h_u))),$$

where L_{f_u} is the Lie derivative operator on X, w.r.t. the vector field f_u. Let us set:

$$D(u) = \{D_0(u) \supset D_1(u) \supset ... \supset D_{n-1}(u)\}. \tag{5}$$

This $u-$dependant flag of distributions is not **regular** in general (i.e. $D_i(u)$ has not constant rank $n-i-1$).

DEFINITION 3.1. *the flag $D(u)$ is called "**the canonical flag**" associated to Σ. In the case where the flag $D(u)$ is* ***regular and independent of*** *u (notation: $\partial_u D(u) = 0$), the canonical flag is said to be* ***uniform.***

The case where $D(u)$ is uniform will be specially important in Chapter 2. In fact, this case will characterize uniform infinitesimal observability.

Note: Here, for us, a distribution D is just a subset of TX, the intersection of which with each tangent plane T_xX is a vector subspace of T_xX. Once the flag $D(u)$ defined here is regular, the distributions $D_i(u)$ are smooth distributions in the usual sense.

4. The Phase Variable Representation.

Here $L^k_{f_u}(h_u)$ denotes the d_y-tuple of functions, the components of which are $L^k_{f_u}(h^i_u)$, where h^i_u is the i^{th} component of $h_u = h(.,u)$. We consider control functions that are sufficiently continuously differentiable only: k times, say.

Consider $R^{(k-1)d_u} = R^{d_u} \times ... \times R^{d_u}$ ($k-1$ times) and $R^{kd_y} = R^{d_y} \times ... \times R^{d_y}$ (k times). We denote the components of $v \in R^{(k-1)d_u}$ by $(v', ..., v^{(k-1)})$ and the components of $y \in R^{kd_y}$ by $(y, y', ..., y^{(k-1)})$.

DEFINITION 4.1. *(Valid for X with corners). There exist smooth mappings Φ^Σ_k and $S\Phi^\Sigma_k$ (the notation $S\Phi^\Sigma_k$ stands for "**suspension** of Φ^Σ_k"):*

$$(6) \qquad \begin{aligned} \Phi^\Sigma_k &: X \times U \times R^{(k-1)d_u} \to R^{kd_y}, \\ \Phi^\Sigma_k &: (x_0, u, u', ..., u^{(k-1)}) \to (y, y', ..., y^{(k-1)}), \end{aligned}$$

$$(7) \qquad \begin{aligned} S\Phi^\Sigma_k &: X \times U \times R^{(k-1)d_u} \to R^{kd_y} \times R^{kd_u}, \\ S\Phi^\Sigma_k &: (x_0, u, u', ..., u^{(k-1)}) \to (y, y', ..., y^{(k-1)}, u, u', ..., u^{(k-1)}), \end{aligned}$$

which are polynomial in the variables $(u', ..., u^{(k-1)})$, and smooth in (x_0, u), such that if $(\hat{x}, \hat{u}) : [0, T_{\hat{u}}[\to X \times U$ is a semi-trajectory of our system Σ starting at x_0, and $t \to y(t) = h(\hat{x}(t), \hat{u}(t))$ is the corresponding output trajectory, then the j^{th} derivative $y^{(j)}(0)$ of $y(t)$ at time 0 is the j^{th} block-component of $\Phi^\Sigma_k(x_0, u(0), \frac{du}{dt}(0), ..., \frac{d^j u}{dt^j}(0))$.

Let us say that the system Σ has the "**phase variable property of order** k", denoted by $PH(k)$, if, for all $x_0 \in X$ and $u(.)$ $k-$ times differentiable:

$$(8) \qquad y^{(k)} = \breve{H}(S\Phi^\Sigma_k(x_0, u, u', ..., u^{(k-1)}), u^{(k)}),$$

for some smooth (C^∞) function $\breve{H} : R^{kd_y} \times R^{(k+1)d_u} \to R^{d_y}$. Notice that if such a function does exist, **it is not unique** in general.

If one denotes temporarily by C^∞_k the ring of smooth functions $g : R^{kdy+(k+1)du} \to R$, the property $PH(k)$ means that the components $y^{(k)}_i$ of $y^{(k)}$ belong to the ring $\Re^\sharp_k$, pull back of C^∞_k by the mapping $S\bar{\Phi}^\Sigma_k : \Re^\sharp_k = (S\bar{\Phi}^\Sigma_k)^* C^\infty_k$, where:

$$\begin{aligned} S\bar{\Phi}^\Sigma_k &= S\Phi^\Sigma_k \times Id^{d_u}, \\ S\bar{\Phi}^\Sigma_k(x, u, u', ..., u^{(k-1)}, u^{(k)}) &= (S\Phi^\Sigma_k(x, u, u', ..., u^{(k-1)}), u^{(k)}). \end{aligned}$$

Then, we can consider the differential system Σ_k, on R^{kd_y} :

$$(9) \qquad (\Sigma_k) \begin{cases} \dot{z}_1 = z_2, ..., \dot{z}_{k-1} = z_k, \\ \dot{z}_k = \breve{H}(z_1, ..., z_k, u(t), ..., \frac{d^k u}{dt^k}(t)). \end{cases}$$

This differential system Σ_k is called a "**phase variable representation**" of Σ. It has the following property, consequence of the uniqueness of the solutions of smooth O.D.E.'s. For any C^k function u, Φ^Σ_k maps the trajectories $x(t)$ of Σ associated with $u(t)$ into the corresponding trajectories of Σ_k : if $x(t)$ is a trajectory of Σ corresponding to $u(t)$, then the curve $t \to \Phi^\Sigma_k(x(t), u(t), u'(t), ..., u^{(k-1)}(t))$ is the trajectory of Σ_k corresponding to $u(t)$, starting from $\Phi^\Sigma_k(x(0), u(0), u'(0), ..., u^{(k-1)}(0))$. In particular, the output trajectory $t \to y(t)$ is mapped into $t \to z_1(t)$, where z_1 denotes the first d_y components of the state z of Σ_k.

A very important particular case where the property $PH(k)$ holds is the following: assume that the map $S\Phi_k^\Sigma$ is an **injective immersion.** For any open relatively compact subset $\Omega \subset X$, let us consider the restriction $S\Phi_k^{\Sigma,\Omega} = (S\Phi_k^\Sigma)_{|\Omega \times U \times R^{(k-1)d_u}}$. If I_Ω denotes the image of $S\Phi_k^{\Sigma,\Omega}$, $I_\Omega \subset R^{kd_y} \times R^{kd_u}$, then, $y^{(k)}(x,u,u',...,u^{(k)})$ defines a function $\breve{h}$ on $I_\Omega \times R^{d_u}$ and easy arguments using partitions of unity show that we can extend this function smoothly to all of $R^{kd_y} \times R^{(k+1)d_u}$. We leave temporarily this simple fact for the reader to show: in Chapter 4, we will state a slightly stronger (but not harder) result.

Denoting this extension of $\breve{h}$ by $\breve{H}$, we get a phase variable representation of order k for Σ restricted to Ω.

As we shall see, there are other interesting cases where the map $S\Phi_k^\Sigma$ is only injective, but $PH(k)$, the phase variable property of order k, holds for Σ, for some k. This situation will be studied in Chapter 4.

Strongly related to this phase variable property, are the notions of "differential observability", and "strong differential observability".

5. Differential Observability and Strong Differential Observability.

Differential observability just means injectivity, of the map $S\Phi_k^\Sigma$. Strong differential observability will mean that **moreover** $S\Phi_k^\Sigma$ is an immersion. Let us relate precisely these notions to the notion of a "**dynamical extension** of Σ".

The control functions u are assumed sufficiently smooth. We can consider the N^{th} dynamical extension Σ^N of Σ, and the N^{th} dynamical extension f^N of f, defined as follows. f^N is just the vector field on $X \times U \times R^{(N-1)d_u} = X \times U \times (R^{d_u} \times ... \times R^{d_u})$, $(N-1$ factors $R^{d_u})$,

$$(10) \qquad f^N(x,u^{(0)},....,u^{(N-1)}) = \sum_{i=1}^{n} f_i(x,u^{(0)})\frac{\partial}{\partial x_i} + \sum_{i=0}^{N-2}\sum_{j=1}^{d_u} u_j^{(i+1)}\frac{\partial}{\partial u_j^{(i)}}.$$

Moreover if we set $b^N = (b_i^N)$, with $b_i^N = \frac{\partial}{\partial u_i^{(N-1)}}$, and $u^{(N)} = (u_i^{(N)})$, the "new control variable",

$$u^{(N)}b^N = \sum_{i=1}^{d_u} u_i^{(N)}b_i^N,$$

then,

DEFINITION 5.1. *the N^{th} dynamical extension $\Sigma^N = (F^N, \breve{h})$ of Σ, is just the control system on $X \times U \times R^{(N-1)d_u}$ with control variable $u^{(N)} \in R^{d_u}$, parametrized vector field $F^N = f^N + u^{(N)}b^N$, and observation function $\breve{h} = (h(x,u^{(0)}),u^{(0)})$.*

REMARK 5.1. *If $U = I^{d_u}$, $I \neq R$, the state space of Σ^N has corners.*

In fact, Σ^N is just the system we get by adding to the state variables the $N-1$ first derivatives of the inputs. The N^{th} derivative is the new control. The observations are the observations of Σ and the control variables u_i denoted here by $u_i^{(0)}$, $1 \leq i \leq d_u$, to stress that the function is the zeroth derivative of itself. Also, $u^{(0)}$, (resp. $u^{(j)}$) denotes the vector with components $u_i^{(0)}$ (resp. $u_i^{(j)}$) , $1 \leq i \leq d_u$.

Let us set $\underline{u}_N = (u^{(0)}, ... , u^{(N-1)})$, and more generally, for a smooth function $y(t)$, with successive derivatives $y^{(i)}(0)$ at $t = 0$, $\underline{y}_N = (y(0), y'(0), ... , y^{(N-1)}(0))$.

The maps $\Phi_N^\Sigma, S\Phi_N^\Sigma$ have already been defined in Section 4. It will also be important to make the system Σ vary in the set S of systems. Hence we will have to consider the following maps:

$$
\begin{aligned}
(11)\quad S\Phi_N \quad &: \quad X \times U \times R^{(N-1)d_u} \times S \to R^{Nd_y} \times R^{Nd_u}, \\
(x, \underline{u}_N, \Sigma) \quad &\to \quad (h(x,u^{(0)}), L_{f^N} h(x, \underline{u}_2), ..., (L_{f^N})^{N-1} h(x, \underline{u}_N), \underline{u}_N) = (\underline{y}_N, \underline{u}_N) \\
S\Phi_N^\Sigma \quad &: \quad X \times U \times R^{(N-1)d_u} \to R^{Nd_y} \times R^{Nd_u}, \\
(x, \underline{u}_N) \quad &\to \quad S\Phi_N(x, \underline{u}_N, \Sigma),
\end{aligned}
$$

and:

$$
\begin{aligned}
(12)\qquad \Phi_N \quad &: \quad X \times U \times R^{(N-1)d_u} \times S \to R^{Nd_y}, \\
(x, \underline{u}_N, \Sigma) \quad &\to \quad (h(x,u^{(0)}), L_{f^N} h(x, \underline{u}_2), ..., (L_{f^N})^{N-1} h(x, \underline{u}_N)) = \underline{y}_N, \\
\Phi_N^\Sigma \quad &: \quad X \times U \times R^{(N-1)d_u} \to R^{Nd_y}, \\
(x, \underline{u}_N) \quad &\to \quad \Phi_N(x, \underline{u}_N, \Sigma).
\end{aligned}
$$

$$S\Phi_N(x, \underline{u}_N, \Sigma) = (\Phi_N(x, \underline{u}_N, \Sigma), \underline{u}_N) = S\Phi_N^\Sigma(x, \underline{u}_N) = (\Phi_N^\Sigma(x, \underline{u}_N), \underline{u}_N).$$

DEFINITION 5.2. *Σ is said **differentially observable of order N**, if $S\Phi_N^\Sigma$ is an injective mapping. Σ is said **strongly differentially observable**, if $S\Phi_N^\Sigma$ is an injective immersion.*

As we mentioned in Section 4, if Σ is strongly differentially observable of order N, then, Σ possesses also the phase variable property $PH(N)$, when restricted to Ω, where Ω is any open relatively compact subset of X.

The reason for these definitions is that, **when $d_y > d_u$, strong differential observability** is easily tractable for the purpose of construction of observer systems. Moreover, roughly speaking, **it is a generic property**. Therefore, it is a relevant definition in that case. This is the subject of Chapter 3. The motivation to consider differential (not strong) observability is that it is the most general concept adapted to the study of dynamic output stabilization (Chapter 6).

6. The trivial foliation.

Associated to Σ, there is a subspace Θ^Σ of the space $C^\infty(X)$ of smooth functions $h : X \to R$:

Θ^Σ is the smallest subspace of $C^\infty(X)$ containing the components h_u^i of $h_u = h(., u)$, for all $u \in U$, which is closed under Lie differentiation on X, with respect to all of the vector fields $f_v = f(., v)$, $v \in U$. It is the real vector subspace of $C^\infty(X)$ generated by the functions $(L_{f_{u_r}})^{k_r}(L_{f_{u_{r-1}}})^{k_{r-1}} ... (L_{f_{u_1}})^{k_1}(h_{u_0})$, for $u_0, ..., u_r \in U$, where L denotes the Lie derivative operator on X.

Θ^Σ is called the "**observation space**" of Σ. The space $d_X\Theta^\Sigma$ of differentials (w.r.t. x) of elements of Θ^Σ, defines a codistribution which is in general singular. The distribution Δ_Σ annihilated by $d_X\Theta^\Sigma$ is called the "**trivial distribution**" associated to Σ. The level sets of Θ^Σ (i.e. the intersections of the level sets of elements of Θ^Σ) define the associated foliation, called the "**trivial foliation**" associated to Σ.

These notions are classical ([**16**]). The reason why we call this foliation the "trivial foliation" is that it is actually trivial (in the sense that the leaves are zero-dimensional), for generic systems. This is also true for systems that are uniformly infinitesimally observable, as our results will show. However, **it is worth to point out that:**

THEOREM 6.1. *([**16**]) The rank of Δ_Σ is constant along the (positive or negative time) trajectories of Σ, in the analytic case.*

Hence, as soon as the analytic system Σ is **controllable** in the weak sense of the transitivity of its Lie algebra (see Appendix 7 below), **then the distribution Δ_Σ is regular.** In particular, the leaves of the trivial foliation (the level sets of Θ^Σ) are submanifolds of the same dimension.

Now, let us consider the case where Δ_Σ is regular, nontrivial, and not necessarily analytic. By Theorem 6.1, it is always regular in the analytic controllable case.

EXERCISE 6.1. *show that Δ_Σ is preserved by the dynamics of Σ. (i.e. for any control function $u(.) \in L^\infty[U]$, $T_X\varphi_t(.,u)$ maps $\Delta_\Sigma(x_0)$ onto $\Delta_\Sigma(\varphi_t(x_0,u))$).*

Since $h(.,u)$ is constant on the leaves of Δ_Σ, for two distinct initial conditions, sufficiently close in the same leaf, the corresponding output trajectories coincide for t small enough, whatever the control function.

In particular, Σ is not observable, for any fixed value of the control function, even if restricted to small open sets: for each control, one can find couples of points, arbitrarily close, that are not distinguished by the observations, for small times.

The following simple fact is important. We leave it as an exercise.

EXERCISE 6.2. *In the case $U = \varnothing$, show the following:*

(13) *–If (iff in the C^ω case) Θ^Σ separates the points on X, then, Σ is observable.*

There is an alternative way to define the distribution Δ_Σ **in the analytic case**, which will be of interest, together with Theorem 6.1 in Chapter 4:

Let us first define the vector subspace Ξ^Σ of $H = C^\infty(X \times U)$, as follows:

Ξ^Σ is the smallest real vector subspace of H which contains the components h^i of h, and which is closed under the action of the Lie derivatives L_{f_u} on X, and with respect to the derivations $\partial_j = \frac{\partial}{\partial u_j}$, $j = 1, ..., d_u$. Ξ^Σ is generated by functions of the form:

$$L_{f_u}^{k_1}(\partial_{j_1})^{s_1} L_{f_u}^{k_2}(\partial_{j_2})^{s_2} ..... L_{f_u}^{k_r}(\partial_{j_r})^{s_r} h_{i,u}, \qquad k_i, s_i \geq 0. \tag{14}$$

Fixing $u \in U$, we obtain the vector subspace $\Xi^\Sigma(u) \subset C^\infty(X)$, and the space $d_X\Xi^\Sigma(u)$ of differentials of the elements of $\Xi^\Sigma(u)$, with respect to the x variable only. We call $\bar{\Delta}_\Sigma(u)$ the distribution annihilated by $d_X\Xi^\Sigma(u)$.

THEOREM 6.2. *a)$\Delta_\Sigma \subset \bar{\Delta}_\Sigma(u)$,*
b)In the analytic case, $\bar{\Delta}_\Sigma(u)$ is independent of u and $\Delta_\Sigma = \bar{\Delta}_\Sigma(u)$.

The point of interest, used in Chapter 4, will be that, in the analytic controllable case, $\bar{\Delta}_\Sigma(u) = \Delta_\Sigma$ is a regular distribution. This is a consequence of Theorem 6.1 and Theorem 6.2 just above.

EXERCISE 6.3. *Find a C^∞ example for which $\bar{\Delta}_\Sigma(u) \neq \Delta_\Sigma$ for some u.*

7. Appendix: weak controllability.

DEFINITION 7.1. *A system Σ being given, with state space X, the Lie subalgebra of the Lie algebra of smooth vector fields on X, generated by the vector fields f_u, $(f_u(x) = f(x,u))$, is called the Lie algebra of Σ, and is denoted by $Lie(\Sigma)$.*

The Lie algebra $Lie(\Sigma)$ defines an involutive (possibly singular) distribution on X.

DEFINITION 7.2. *Let a system Σ be given, with state space X. The system Σ is said weakly controllable, if the Lie Algebra $Lie(\Sigma)$ is transitive on X, i.e. $\dim(Lie(\Sigma)(x))$, the dimension of $Lie(\Sigma)$ evaluated at x, as a vector subspace of T_xX, is equal to $n = \dim(X)$, for all $x \in X$.*

The following facts are standard, and are used in the text. They come from the classical "Frobenius Theorem", "Chow Theorem" and "Hermann-Nagano Theorem".

-(1) If a system is weakly controllable, than, the **accessibility set** $A_\Sigma(x_0)$ of $x_0 \in X$, i.e. the set of points that can be joined from x_0 by some trajectory of Σ, **in positive time**, has nonempty interior in X, for all $x_0 \in X$.

-(2) If a system is weakly controllable, then, the **orbit** $O_\Sigma(x_0)$ of $x_0 \in X$, i.e. the set of points that can be joined to x_0 by some continuous curve which is a concatenation of trajectories of Σ **in positive or negative time**, is equal to X, for all $x_0 \in X$.

-(3) If moreover Σ is analytic, the statements (1), (2) above are "if and only if".

-(4) If Σ is C^∞, not weakly controllable, but the distribution $Lie(\Sigma)$ on X has constant rank or if Σ is analytic, then, $A_\Sigma(x_0)$ has nonempty relative interior in the orbit $O_\Sigma(x_0)$, which is just the integral leaf through x_0 of the distribution $Lie(\Sigma)$.

-(5) In the statements (1), (2),(3),(4) above, it is possible to restrict to piecewise constant control functions.

CHAPTER 2

The case $d_y \leq d_u$.

We will treat only the case $d_y = 1$, $d_u \geq 1$. General results for the case $d_u \geq d_y > 1$ are more difficult to obtain. However, the chapter 7 shows a nontrivial practical example where $d_y = d_u = 2$, that is uniformly infinitesimally observable.

In this chapter, except in the first section, we assume that $d_y = 1$, $d_u \geq 1$, and everything is analytic.

We characterize analytic systems that are uniformly infinitesimally observable when restricted to an open dense subset of X. The necessary and sufficient condition is that $\partial_u D(u) = 0$, i.e. **the canonical flag is uniform.** This condition $\partial_u D(u) = 0$ is extremely restrictive, and is not preserved by small perturbations of the system.

The analyticity assumption with respect to the x variable is made for purely technical reasons. It can certainly be removed to get similar results: see for instance Exercise 4.6 below.

On the other hand, analyticity with respect to u **is essential.** It is possible to obtain results in the nonanalytic case, but they will be **weaker** in the following sense: to have uniform infinitesimal observability, a certain condition has to hold on an open dense subset of X uniformly in u. In the nonanalytic case, we can only show that this condition has to hold on an open dense subset of $X \times U$. **This is much weaker.** The reasons for the "much weaker" result are: to prove the analytic case, we use the permanence properties of projections of semialgebraic or subanalytic sets.

Again d_X (resp. d_U) denotes the differential with respect to the x variables (resp. u variables) only.

1. Relation between observability and infinitesimal observability.

The relation is stated in the following theorem (also valid for $d_y > 1$):

THEOREM 1.1. *(i) For any system Σ and any input $\hat{u}$, the set $\theta(\hat{u})$ of states $x \in X$ such that Σ is infinitesimally observable at $(\hat{u}, x)$ is open in X (could be empty, of course).*

(ii) If Σ is observable for an input $\hat{u}$, then $\theta(\hat{u})$ is everywhere dense in X.

(iii) If Σ is infinitesimally observable at $(\hat{u}, x)$, then there exists an open neighbourhood V of x such that the restriction $P_{\Sigma,\hat{u}}|V$ is injective (i.e. Σ restricted to V is observable for the input $\hat{u}$).

In the remaining of the chapter, $d_y = 1$.

2. Normal form for a uniform canonical flag.

We assume that the canonical flag associated to the system Σ is uniform. We will show first that it is equivalent that Σ can be put everywhere locally in a certain normal form, called the **observability canonical form.**

THEOREM 2.1. *Σ has a uniform canonical flag if and only if, for all $x_0 \in X$, there is a coordinate neighbourhood of x_0, $(V_{x_0}, x^0, ..., x^{n-1})$, such that in these coordinates, Σ can be written as follows:*

$$
\begin{aligned}
(15) \quad \frac{dx^0}{dt} &= f_0(x^0, x^1, u), ..., \frac{dx^i}{dt} = f_i(x^0, x^1, ..., x^{i+1}, u), ..., \\
\frac{dx^{n-2}}{dt} &= f_{n-2}(x^0, x^1, ..., x^{n-1}, u), \frac{dx^{n-1}}{dt} = f_{n-1}(x^0, x^1, ..., x^{n-1}, u), \\
y &= h(x^0, u), \text{ and } \forall (x,u) \in V_{x_0} \times U, \ \frac{\partial h}{\partial x^0}(x^0, u) \neq 0, \\
& \quad \frac{\partial f_0}{\partial x^1}(x^0, x^1, u) \neq 0, ..., \frac{\partial f_{n-2}}{\partial x^{n-1}}(x^0, ..., x^{n-1}, u) \neq 0.
\end{aligned}
$$

Let us set $h_u^j(x) = h^j(x,u) = L_{f_u}^j h_u(x)$.

COROLLARY 2.2. *Σ has a uniform canonical flag if and only if, for all $x_0 \in X$, for all $v \in U$, there exists an open neighbourhood $V_{x_0,v}$ of x_0, such that the functions $x^0 = h_v^0 | V_{x_0,v}, x^1 = h_v^1 | V_{x_0,v}$ ...., $x^{n-1} = h_v^{n-1} | V_{x_0,v}$, form a coordinate system on $V_{x_0,v}$, and on $U \times V_{x_0,v}$, each h^i is a function of u, $x^0, ..., x^i$ only, $0 \leq i \leq n-1$.*

3. Characterization of uniform infinitesimal observability.

The first observation that can be made is the following:

THEOREM 3.1. *Assume that Σ is such that its canonical flag is uniform. Then, $\forall x_0 \in X$, there is an open neighbourhood V_{x_0} of x_0 such that the restriction $\Sigma_{|V_{x_0}}$ of the system Σ to V_{x_0} is observable and uniformly infinitesimally observable.*

The main result in this chapter is that, conversely, **the uniformity condition on the canonical flag is a necessary condition** for uniform infinitesimal observability, at least on an open dense subset of X.

Let us point out the following fact about this result: it is true "almost everywhere" with respect to X, but it is **global** with respect to U.

This is the hard part to prove. If one is interested with a result true almost everywhere with respect to both x and u, **the proof is much easier.**

Before proceeding, let us make the following standing assumptions:

either,

(H_1) $U = I^{d_u}$, $I \subset R$ is a compact interval, and the system is analytic,

or,

(H_2) $U = R^{d_u}$, f and h are algebraic with respect to u.

Let $\tilde{M}$ be the subset of $U \times X$:

$$\tilde{M} = \{(u,x) | d_X h_u^0(x) \wedge ... \wedge d_X h_u^{n-1}(x) = 0\}.$$

Let M be its projection on X. Then:

THEOREM 3.2. *Assume either* (H_1) *or* (H_2) *and that* Σ *is uniformly infinitesimally observable. Then:*

1. The set M *is a subanalytic (resp. semianalytic in case of* (H_2)*) set* ***of codimension at least*** 1. *In the case* (H_1), M *is closed. In any case, denote by* $\bar{M}$ *its closure,*

2. The restriction $\Sigma_{|X\setminus\bar{M}}$ *of* Σ *to* $X\setminus\bar{M}$ *has a uniform canonical flag.*

Let us give some comments, examples, and state some complementary results.

4. Complements.

4.1. Exercises:

EXERCISE 4.1. *Let* Σ *be a system with uniform canonical flag. Show that* Σ *is strongly differentially observable of order* n, *and hence has the phase variable property of order* n, *when restricted to sufficiently small open subsets of* X.

EXERCISE 4.2. *Show that the (uncontrolled) system on* R^2 :

$$\dot{x}_1 = x_2, \dot{x}_2 = 0,\ y = (x_1)^3,$$

is observable on R^2, *but not infinitesimally observable at* $x_0 = 0$.

EXERCISE 4.3. ***The output function does not depend on*** u *and* $n = \dim(X) = 2$. *Assume that we work in the class of systems such that* h *does not depend on* u, $u \in I^{d_u}$, I *compact. Fix* $x^0 \in X$.

1. Show that the property that Σ *has a uniform canonical flag in a neighbourhood of* x^0 *is stable under* C^2- *small perturbations of* Σ.

2. Show that, if $n > 2$, *this is not true.*

EXERCISE 4.4. *In the class of control affine systems (i.e.* $\dot{x} = f(x) + u\ g(x)$, $y = h(x), U = R$*), show that the result 1 of Exercise 4.3 is false.*

EXERCISE 4.5. *Show that the system on* $X = R$:

$$\begin{aligned} \dot{x} &= 1, \\ y &= \frac{1}{2}(x - \frac{\sin(2x(1+u^2)^{\frac{1}{2}})}{2(1+u^2)^{\frac{1}{2}}}) + x\sin^2 u,\ u \in R \end{aligned}$$

is uniformly infinitesimally observable, but Theorem 3.2 is false (U *is* ***not compact****).*

4.2. Control affine systems. We consider the control affine analytic systems (with single control, to simplify):

$$(\Sigma_A)\begin{cases} \dot{x} = f(x) + u\ g(x), \\ y = h(x),\ u \in R. \end{cases} \tag{16}$$

In that case, there is a stronger statement than Theorem 3.2, which is much easier to prove. Consider the mapping $\Phi : X \to R^n$, $\Phi(x) = (h(x), L_f h(x), ..., L_f^{n-1}h(x))$.

The set of points $x \in X$ at which Φ is not a local diffeomorphism, i.e. $d_X h(x) \wedge d_X L_f h(x) \wedge ... \wedge d_X L_f^{n-1}h(x) = 0$ is an analytic subset, closed in X, denoted by M.

THEOREM 4.1. *1. If Σ_A is* ***observable****, then M has codimension 1 at least, and, on each open subset $Y \subset X \backslash M$ such that the restriction $\Phi_{|Y}$ is a diffeomorphism, $\Phi_{|Y}$ maps $\Sigma_{A|Y}$ into a system $\bar{\Sigma}_A$ of the form:*

$$(17)\quad (\bar{\Sigma}_A)\ \dot{x} = \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ . \\ . \\ . \\ \dot{x}_{n-1} \\ \dot{x}_n \end{pmatrix} = \begin{pmatrix} x_2 \\ x_3 \\ . \\ . \\ . \\ x_n \\ \varphi(x) \end{pmatrix} + u \begin{pmatrix} g_1(x_1) \\ g_2(x_1, x_2) \\ . \\ . \\ . \\ g_{n-1}(x_1, ..., x_{n-1}) \\ g_n(x) \end{pmatrix},$$

$$y = x_1.$$

2. Conversely, if Ω is an open subset of R^n on which the system Σ has the form $\bar{\Sigma}_A$, then, the restriction $\Sigma_{|\Omega}$ is observable.

The proof is simple, and contains the basic idea for the proof of Theorem 3.2.

EXERCISE 4.6. *Give a statement and a proof of Theorem 4.1 in the C^∞ case.*

4.3. Bilinear systems (single output). Bilinear systems are systems on $X = R^n$, that are control affine **and** state affine:

$$(18)\qquad (\mathcal{B}) \begin{cases} \dot{x} = Ax + u\,(Bx + b), \\ y = Cx, \end{cases}$$

where $A : R^n \to R^n$, $B : R^n \to R^n$ are linear, $b \in R^n$, $C \in (R^n)^*$.

For these bilinear systems, the previous result, Theorem 4.1, can be made much stronger.

EXERCISE 4.7. *show the following theorem:*

THEOREM 4.2. *The single-output bilinear system $(\mathcal{B})$ is* ***observable*** *if and only if it has the following form in the (linear) coordinate system $x^* = (Cx, CAx, ..., CA^{n-1}x)$: $\dot{x}^* = \bar{A}x^* + u\ (\bar{B}x^* + \bar{b})$, $y = \bar{C}x$, where $\bar{C} = (1, 0, ..., 0)$, $\bar{B}$ is lower triangular, and $\bar{A}$ is a "companion matrix":*

$$(19)\qquad \bar{A} = \begin{pmatrix} 0, 1, 0, ..., 0 \\ 0, 0, 1, 0, ..., 0 \\ . \\ . \\ 0, ..., 0, 1 \\ a_1, ..., a_n \end{pmatrix}.$$

The bilinear systems (single output or not) play a very special role from the point of view of the observability property: the *initial−state → output−trajectory* mapping is an affine mapping. In fact, the control function being known, they are just linear time-dependent systems. Therefore, for instance, the observer problem can be solved jus using the linear theory.

A very important result is stated in the following exercise:

EXERCISE 4.8. *(Fliess-Kupka theorem in the analytic case.) 1. Define a reasonable notion of an immersion of a system into another one.*

2. Show that a control affine analytic system can be immersed into a bilinear one if and only if its observation space Θ^{Σ} is finite-dimensional. (See Chapter 1, Section 6 , for the definition of the observation space.)

This result (2., Exercise 4.8) is not very difficult to prove. The original result, in the paper [**10**], is a similar theorem in the C^{∞} case, the proof of which is not that easy.

CHAPTER 3

The case $d_y > d_u$.

We refer to the notations of Chapter 1, Sections 4 , 5.

The main purpose in this chapter is to show that, in this case, the picture is completely reversed: Roughly speaking, the observability becomes a generic property. More precisely, the **"strong differential observability property of order $2n+1$"** (in the sense of the definition 5.2, Chapter 1) **is generic** (in the Baire sense only: it is an open problem to prove that the set of strongly differentially observable systems contains an open dense set of systems. If we were able to show this openness property, some deep technical complications could be avoided in the proof of the other results).

Another very important result is the following:

Observability (for all L^∞ inputs) is a dense property, that is, any system can be approximated by an observable one.

In the case where X is compact, strong differential observability means that $S\Phi_N^\Sigma$ is an embedding, or $\Phi_N^\Sigma(.,\underline{u}_N)$ is an embedding for all $\underline{u}_N$. Therefore, the set of systems such that $S\Phi_N^\Sigma$ is an embedding, is generic, if $N \geq 2n+1$.

Of course, there is no chance to prove such a general result if X is not compact:

EXERCISE 0.9. *Show that, even among ordinary smooth mappings between finite dimensional manifolds X, Y, embeddings may not be dense, whatever the dimension $N = \dim(Y)$ with respect to $n = \dim(X)$.*

(For hints, see [**18**], page 54.)

The reason for the fact stated in Exercise 0.9 is that embeddings are proper mappings. For our practical purposes (synthesis of observers, output stabilization), we don't need that the fundamental mapping $S\Phi_N^\Sigma$ be proper. It is sufficient for it to be an injective immersion.

Hence, all the genericity results we prove are true for a noncompact X and for the Whitney topology. But for the sake of simplicity, we shall assume in this chapter that X **is a compact manifold**, and leave all generalizations to the reader as exercises.

Also, we will assume that X is an **analytic manifold.** One should be conscious of the fact that this is not a restriction: any C^∞ manifold possesses a compatible C^ω structure (see [**18**], page 66).

Again, in this chapter, $U = I^{d_u}$, where I is a closed bounded interval. Since we make an extensive use of subanalytic sets and their properties, this compactness assumption cannot be relaxed.

1. Definitions, notations.

The systems under consideration are of the form:

$$(\Sigma)\ \frac{dx}{dt} = f(x, u^{(0)})\ ;\ y = h(x, u^{(0)}) \tag{20}$$

or

$$(\Sigma)\ \frac{dx}{dt} = f(x, u^{(0)})\ ;\ y = h(x), \tag{21}$$

in order to take into account the more practical cases where the output function h does not depend on u : the proofs of the genericity results in that case are not different from the proofs in the general case where h depends on u, but the results do not follow from the results in the general case.

In agreement with the notations introduced in Section 5, Chapter 1, we shall use the notation $u^{(0)}$ for the control variable.

We will assume that f and h are at least C^r w.r.t. $(x, u^{(0)})$, **for r large enough.** We shall endow the set of systems with the topology of C^r uniform convergence over $X \times U$. The set of systems with this topology will be denoted by S^r. In the particular case where h does not depend on u, it will be denoted by $S^{0,r}$. Then, $S^r = F^r \times H^r$, $S^{0,r} = F^r \times H^{0,r}$, where F^r denotes the set of $u-$parametrized vector fields f over X, that are C^r with respect to both x and u. Also, H^r denotes the set of C^r maps $h : X \times U \to R^{d_y}$, and $H^{0,r}$ denotes the set of C^r maps $h : X \to R^{d_y}$. The spaces F^r, H^r, $H^{0,r}$ will also be endowed with the C^r topology.

2. Statement of our differential observability results.

Our results are the following theorems, that hold for $r > 0$, **large enough**:

THEOREM 2.1. *The set of systems such that $S\Phi_N^\Sigma$ is an immersion, **contains an open dense** subset of S^r (resp. $S^{0,r}$), for $N \geq 2n$.*

THEOREM 2.2. *The set of systems such that $S\Phi_N^\Sigma$ is an embedding, (i.e. Σ is strongly differentially observable) contains a **residual** subset of S^r (resp. $S^{0,r}$), for $N \geq 2n+1$.*

A bound $B > 0$ on the derivatives of the controls being given, denote by I_B the interval $[-B, B]$.

THEOREM 2.3. *The set of systems such that the restriction of $S\Phi_N^\Sigma$ to $X \times U \times I_B^{(k-1)d_u}$ is an embedding, **is open, dense** in S^r (resp. $S^{0,r}$), for $N \geq 2n+1$.*

THEOREM 2.4. *(X analytic) The set of **analytic** systems such that $S\Phi_N^\Sigma$ is an embedding, is **dense** in S^r (resp. $S^{0,r}$), for $N \geq 2n+1$.*

Now, we shall give several examples showing that all these theorems are false when $d_y = d_u = 1$. In all the cases, $X = S^1$, the circle, and $U = [-1, 1]$.

Consider:

$$(\Sigma^1)\begin{cases} \dot{\theta} = 1, \\ y = \varphi_1(\theta) + \varphi_2(\theta)u, \end{cases}$$

with the assumption $(H) : \varphi_1'(\theta_0) = 0, \varphi_1''(\theta_0) \neq 0,\ \varphi_2'(\theta_0) \neq 0$.

One should observe that the condition (H) is stable under small perturbations and holds if $\theta_0 = 0, \varphi_1(\theta) = \cos(\theta), \varphi_2(\theta) = \sin(\theta)$.

EXERCISE 2.1. *Show that, if* $\theta = \theta_0$, *taking* $u^{(0)} = 0$, *we can compute* $u^{(1)}, \ldots,$ $u^{(N)}$ *satisfying the equation:*

$$d_\theta \begin{pmatrix} y \\ \dot{y} \\ . \\ . \\ y^{(N)} \end{pmatrix} = 0,$$

and show that there exists an open neighbourhood $\mathcal{U}$ *of* Σ^1 *(*C^2 *open in* S^∞*), such that* $S\Phi_k^\Sigma$ *is not an immersion for any* k, *for* $\Sigma \in \mathcal{U}$.

This is a counterexample of Theorems 2.1, 2.2 when $d_y \leq d_u$. It is also a counterexample of Theorem 2.4.

A better example is the following system Σ_ε^1, for ε small:

$$(\Sigma_\varepsilon^1) \begin{cases} \dot{\theta} = 1, \\ y = \varepsilon\varphi_1(\theta) + \varphi_2(\theta)u, \end{cases}$$

with the same assumption (H) on θ_0, φ_1 and φ_2. Chose an arbitrary integer $k > 0$ and a real $B > 0$.

EXERCISE 2.2. *Show that there is an* ε_0 *sufficiently small so that, for the system* $\Sigma_{\varepsilon_0}^1$ *and* ***for a*** C^k ***neighbourhood*** $\mathcal{V}$ ***of*** $\Sigma_{\varepsilon_0}^1$ ***in*** S^∞, *there is a* θ_0 *and a point* $u^{(0)}, u^{(1)}, \ldots, u^{(k-1)}$ *such that* $S\Phi_k^\Sigma$ *is not immersive at* $\theta_0, u^{(0)}, u^{(1)}, \ldots, u^{(k-1)}, u^{(0)} \in U$, $u^{(i)} \in I_B$, $1 \leq i \leq k-1$, $\Sigma \in \mathcal{V}$.

This is a counterexample to Theorem 2.3.

Using the same typology, we can also construct an example showing that, **if** $d_y > d_u$, **the set of systems** Σ **such that** $S\Phi_N^\Sigma$ **is an immersion is not open, for any** N. Consider the system:

$$(\Sigma_\varepsilon^2) \begin{cases} \dot{\theta} = 1, \\ y_1 = \varphi_1(\theta) + \varepsilon\varphi_2(\theta)u, \\ y_2 = 0, \end{cases}$$

with $\varphi_1(\theta) = \cos(\theta)$, $\varphi_2(\theta) = \sin(\theta)$. At $\theta_0 = 0$, the assumption (H) is satisfied.

EXERCISE 2.3. *Show that, for* $\varepsilon = 0, S\Phi_2^{\Sigma_\varepsilon^2}$ *is an immersion. For* $\varepsilon \neq 0$, ε *small,* $S\Phi_k^{\Sigma_\varepsilon^2}$ *is not an immersion for any* k, *using the same reasoning as in the previous examples.*

EXERCISE 2.4. *Show that the mapping:*
$S^r \to C^{r-N+1}(\, X \times U \times R^{(N-1)d_u}, R^{Nd_y} \times U \times R^{(N-1)d_u})$,
$\Sigma \to S\Phi_N^\Sigma$,
is not continuous for the Whitney topology (over $C^{r-k+1}(X \times U \times R^{(N-1)d_u}, R^{kd_y} \times U \times R^{(N-1)d_u})$).

REMARK 2.1. *Theorem 2.4 is not a consequence of Theorem 2.2: this theorems does not prove that there is an open dense subset of systems satisfying* (F).

3. Equivalence between "observability" and "observability for smooth inputs".

In this section, **we consider analytic systems**, and we show that, for these systems, $C^\omega-$ observability (i.e. observability for all C^ω inputs) implies observability (i.e. observability for all L^∞ inputs). In fact, these notions are equivalent. This result will be crucial in order to prove our final "approximation theorem" 4.1 in the next section 4.[1]

The proof of this result is rather hard, and makes a deep use of subanalytic sets, their projections, and their tangent objects.

EXERCISE 3.1. *Show that, in the C^∞ case, observability and C^ω observability are not equivalent properties.*

THEOREM 3.1. *For an **analytic** system Σ, (either $\Sigma \in S^\omega$ or $\Sigma \in S^{0,\omega}$), the following properties are equivalent:*

(i) Σ is observable for all L^∞ inputs,

(ii) Σ is observable for all C^ω inputs.

4. The approximation theorem.

Recall that, if Σ, (analytic), is as in the previous section 2, such that $S\Phi_k^\Sigma$ is an embedding, then Σ is observable for all C^ω inputs (Σ is observable for all C^k inputs, which is stronger):

A C^k input $u(t)$ being given on some interval $[0, \Theta]$, and x_1, x_2, $x_1 \neq x_2$ being given initial conditions, assume that the corresponding outputs are equal on some time subinterval $[0, \tau]$. Then their $k-1$ first derivatives at time zero are also equal, and they, with the $u^{(j)}(0)$, are just the components of $S\Phi_k^\Sigma$ by definition. This is impossible since $S\Phi_k^\Sigma$ is injective. Σ is observable for u.

In fact, the fact that $S\Phi_k^\Sigma$ is an embedding (strong differential observability) expresses that $P_{\Sigma,u}$, the *initial − state → output − trajectory* mapping, is an embedding for all of the considered $k-$times differentiable inputs u. Although, Σ observable only means that $P_{\Sigma,u}$ is injective, but for all L^∞ inputs u.

The results of the sections 2, 3, show that, for $d_y > d_u$:

1. Any C_r system Σ^0 can be approximated by an analytic one Σ^1, which is observable for all C^ω inputs (for all C^{2n+1} inputs): Theorem 2.4,

2. Σ^1 is in fact observable (for all L^∞ inputs): Theorem 3.1.

Therefore:

THEOREM 4.1. ***(Approximation by observable analytic systems).*** *Any system $\Sigma^0 \in S^r$ (resp. $S^{0,r}$), r sufficiently large, can be approximated by an observable one $\Sigma^1 \in S^r$ (resp. $S^{0,r}$) (observable for all L^∞ inputs), that moreover can be chosen analytic and such that $S\Phi_k^{\Sigma^1}$ is an embedding, for some k.*

5. Complements.

The two following results are important:

The first one concerns uncontrolled systems, i.e. Σ is of the form:

$$(\Sigma_{uc}) \quad \frac{dx}{dt} = f(x), \; y = h(x).$$

[1]There is a related result in the paper [**34**], based upon desingularization techniques.

X is again assumed to be compact. Then, the following theorem holds.

THEOREM 5.1. *The set of uncontrolled systems Σ_{uc} that are strongly differentially observable, of order $N = 2n + 1$, (i.e. Φ_N^Σ is an embedding) is open, dense.*

EXERCISE 5.1. *Prove Theorem 5.1.*

This is easy, as a consequence of the main theórems in this chapter. For a direct proof, see [**11**].

The second important result concerns the class of control affine systems. These systems are very common in practice. Recall that they are of the form:

$$(\Sigma_{ca}) \quad \frac{dx}{dt} = f(x) + \sum_{i=1}^{d_u} g_i(x)u_i, \; y = h(x).$$

The result is:

THEOREM 5.2. *The theorems 2.1, 2.2, 2.3, 2.4 , are all true in the class of control affine systems.*

EXERCISE 5.2. *prove Theorem 5.2.*

This exercise is not that easy, although the general idea of the proof is the same. This has been done in [**2**].

Also, the following interesting result holds: X is again an analytic compact manifold. Let us say that **a vector field f on X is observable** if there exists a continuous function $h : X \to R$, such that $\Sigma = (f, h)$ is observable.

THEOREM 5.3. *(1) An **analytic** vector field f is observable iff it has only **isolated singularities.***

(2) If (1) holds, then, the set of analytic maps h such that $\Sigma = (f, h)$ is observable, is dense in H^r.

EXERCISE 5.3. *Prove Theorem 5.3.*

This is not an easy exercise. For hints, see [**22**].

CHAPTER 4

Singular state-output mappings.

In the two previous chapters, all $initial-state \rightarrow output-trajectory$ mappings are regular in some sense: either the system has a uniform canonical flag, and it is also, at least locally, strongly differentially observable (see Exercise 4.1, Chapter 2), or, in the case where $d_y > d_u$, systems are generically strongly differentially observable of some order.

In both cases, the $initial-state \rightarrow output-trajectory$ mapping is an immersion in some sense, and as a consequence, the systems have the phase variable property.

It can happen that the $initial-state \rightarrow output-trajectory$ mapping is not an immersion, but that nevertheless, the system possesses the phase variable property of some order.

It is interesting to study these singular situations since, for observation or output stabilization, **only the phase variable property matters** , as will be clear in the next chapters. This is the purpose of this chapter.

The uncontrolled case is very different from the controlled one. We will show that, **in the uncontrolled analytic case**, a reasonable assumption is that the map Φ_N^Σ **is a finite mapping** for some N. Unfortunately, in this case, there is no C^∞ version of our results.

In both cases (controlled and uncontrolled), the first step of the study is **local** (at the level of germs of systems). Afterwards, assuming observability (injectivity), the phase variable representations can be **glued** together using a partition of unity.

On the other hand, **in the controlled case, we don't need the analyticity assumption.** But, for the sake of simplicity of the exposition, we let it stand.

1. Assumptions, definitions.

Here, we consider only **analytic** systems, of the form:

$$(\Sigma)\ \frac{dx}{dt} = f(x,u),\ y = h(x,u), \text{(controlled case), or,}$$
$$(\Sigma)\ \frac{dx}{dt} = f(x),\ y = h(x), \text{(uncontrolled case).}$$

As a first step, we will consider germs of such systems at a point $(x_0, u^{(0)}) \in X \times U$ (controlled case), or $x_0 \in X$ (uncontrolled case). In this chapter, U is not assumed to be compact. In most cases, for global considerations, we will consider that $U = R^{d_u}$.

Notations.

Again in this section the value $u = u^{(0)}$ of the control plays a role different than the higher order derivatives $u^{(i)}$, $i \geq 1$. Hence, we introduce the following notations:

Given a $N-$jet $\underline{f}_{N+1} = (f^{(0)}, f^{(1)}, ..., f^{(N)})$ of a curve f in a Euclidean space R^m, we will write:

$$\begin{aligned} \tilde{f}_N &= (f^{(1)}, ..., f^{(N)}), \\ \underline{f}_{N+1} &= (f^{(0)}, \tilde{f}_N). \end{aligned} \tag{22}$$

We use this notation for the case of infinite jets ($N = \infty$), and we drop the subscript $N = \infty$ ($\underline{f}_\infty = \underline{f}$, $\tilde{f}_\infty = \tilde{f}$).

Let us define the restricted mappings $\Phi^\Sigma_{N,\tilde{u}_{N-1}} : X \times U \to R^{Nd_y}$ and $S\Phi^\Sigma_{N,\tilde{u}_{N-1}} : X \times U \to R^{Nd_y} \times R^{d_u}$,

$$\begin{cases} \Phi^\Sigma_{N,\tilde{u}_{N-1}}(x_0, u^{(0)}) = \Phi^\Sigma_N(x_0, u^{(0)}, \tilde{u}_{N-1}), \\ S\Phi^\Sigma_{N,\tilde{u}_{N-1}}(x_0, u^{(0)}) = (\Phi^\Sigma_N(x_0, u^{(0)}, \tilde{u}_{N-1}), u^{(0)}). \end{cases} \tag{23}$$

Let O_{x_0} be the ring of germs of analytic functions at $x_0 \in R^d$.

Let $\Re$ be a subring of O_{x_0}, $x_0 \in R^d$. For $u_0 \in R^p$, $\Re\{u; u_0\}$ will denote the ring of germs at $(x_0, u_0) \in R^d \times R^p$ of analytic mappings of the form $G(u, \varphi_1(x), ..., \varphi_r(x))$, for G analytic at $(u_0, \varphi_1(x_0), ..., \varphi_r(x_0))$, and for any finite subset $\{\varphi_1, ..., \varphi_r\} \subset \Re$. If $\Re$ is an analytic algebra (in the sense of [29], for instance), then $\Re\{u; u_0\}$ is also an analytic algebra.

Rings of functions:

We have to consider several rings of (germs of) analytic functions attached to the germ of a system Σ at a point. There are different definitions for the controlled and uncontrolled case. Let us fix a point $x_0 \in X$, and an infinite jet $(u_0^{(0)}, \widetilde{u_0}) = \underline{u_0}$.

Let us define the rings $\Re_N(x_0)$, or $\Re_N(x_0, \underline{u_0}_N), \hat{\Re}_N(x_0, \underline{u_0}_N)$.

1) In the uncontrolled case:

$$\Re_N(x_0) = (\Phi^\Sigma_N)^*(O_{y_0}), \tag{24}$$

the pull back by Φ^Σ_N of the ring O_{y_0} of germs of analytic real valued functions $\varphi(y, \tilde{y}_{N-1})$ at the point $y_0 = \Phi^\Sigma_N(x_0)$, i.e.,

$$\Re_N = \{G \circ \Phi^\Sigma_N(x) | G \text{ is an analytic germ at } y_0 = \Phi^\Sigma_N(x_0)\}. \tag{25}$$

2) In the controlled case:

$$\begin{aligned} \Re_N(x_0, \underline{u_0}_N) &= (S\Phi^\Sigma_{N,\tilde{u}_{0,N-1}})^*(O_{y_0}), \\ \hat{\Re}_N(x_0, \underline{u_0}_N) &= (S\Phi^\Sigma_N)^*(O_{z_0}), \end{aligned}$$

where $y_0 = S\Phi^\Sigma_{N,\tilde{u}_{0,N-1}}(x_0, u_0^{(0)})$ and $z_0 = S\Phi^\Sigma_N(x_0, u_0^{(0)}, \widetilde{u_0}_{N-1})$.

If there is no ambiguity about the choice of $x_0, (u_0^{(0)}, \widetilde{u_0}_{N-1}) = \underline{u_0}_N$, we will write $\Re_N, \Re_N, \hat{\Re}_N$ in place of $\Re_N(x_0)$, $\Re_N(x_0, \underline{u_0}_N)$, $\hat{\Re}_N(x_0, \underline{u_0}_N)$.

For each N, $\hat{\Re}_N$ can be canonically identified to a subring of $\hat{\Re}_{N+1}$: $\hat{\Re}_N \subset \hat{\Re}_{N+1}$.

In both the controlled and the uncontrolled case, $\Re_N$ and $\hat{\Re}_N$ are Noetherian rings. They form increasing sequences:

$$\begin{aligned} &\dots \subset \Re_N \subset \Re_{N+1} \subset \dots \subset O_{x_0} \text{ or } O_{(x_0,u_0^{(0)})}, \\ &\dots \subset \hat{\Re}_N \subset \hat{\Re}_{N+1} \subset \dots \end{aligned} \tag{26}$$

In the controlled case, $\check{\Re}$ will denote the ring of germs of analytic mappings of the form $G(u,\varphi_1,\dots,\varphi_p)$ at the point $(x_0,u_0^{(0)})$, for any positive integer p and for functions φ_i of the form

$$\varphi_i = L_{f_u}^{k_1}(\partial_{j_1})^{s_1} L_{f_u}^{k_2}(\partial_{j_2})^{s_2} \dots .. L_{f_u}^{k_r}(\partial_{j_r})^{s_r} h,\ k_l, s_l \geq 0, \tag{27}$$

(see Formula (14) in the definition of Ξ^Σ, Chapter 1). Recall that ∂_j denotes the derivation with respect to the j^{th} control variable.

Obviously, $\check{\Re}$ is closed under the action of the derivations L_{f_u} and $\partial_j, 1 \leq j \leq d_u$. Also,

$$\Re_N \subset \check{\Re} \text{ for all } N. \tag{28}$$

$\check{\Re}_N$ will denote the ring of germs of analytic mappings of the form $G(u,\varphi_1,\dots,\varphi_p)$ at the point $(x_0,u_0^{(0)})$, for all functions φ_i of the form (27) above, with $\sum k_i + \sum s_i \leq N-1$.

REMARK 1.1. *(1) The ring $\hat{\Re}_N\{u^{(N)},u_0^{(N)}\}$ is exactly the ring of germs at a point of (analytic) elements of the rings $\Re_N^\sharp$, defined in Chapter 1, Section 4,*

(2)the ring $\check{\Re}$ is just the ring of analytic germs at $(x_0,u_0^{(0)})$ generated by the germs of the elements of the space Ξ^Σ, plus the control variables, (Ξ^Σ has been defined in Chapter 1, Section 6).

2. The ascending chain property.

DEFINITION 2.1. *A germ of analytic system Σ (at the point x_0 or at the point $(x_0,u_0^{(0)},\widetilde{u_0})$) satisfies the "ascending chain property of order N", denoted by $ACP(N)$, if:*

uncontrolled case: *$\Re_j = \Re_N$ for $j \geq N$,*

Controlled case: *$\hat{\Re}_{j+1} = \hat{\Re}_j\{u^{(j)};u_0^{(j)}\}$ for $j \geq N$.*

Convention: For simplicity, we will say that a vector function belongs to $\Re_N$ or $\hat{\Re}_N$ if each of its components does.

The next two lemmas show the relation between the ascending chain property $ACP(N)$ and the phase variable property $PH(N)$. The phase variable property $PH(N)$ has been defined in Chapter 1 for systems. For **germs of** analytic systems, the definition is similar and left to the reader.

LEMMA 2.1. *Σ satisfies the $ACP(N)$ at some point iff $\Re_{N+1} = \Re_N$, (resp. $\hat{\Re}_{N+1} = \hat{\Re}_N\{u^{(N)};u_0^{(N)}\}$) in the uncontrolled (resp. controlled) case.*

LEMMA 2.2. *Each of the following two conditions is necessary and sufficient for Σ to satisfy the $ACP(N)$:*

(i) $y^{(N)} = \Psi_N(y^{(0)},\tilde{y}_{N-1},u^{(0)},\tilde{u}_N)$ for some analytic function Ψ_N (locally defined in a neighbourhood of $(S\Phi_N^\Sigma(x_0,u_0^{(0)},\widetilde{u_0}_{N-1}),u_0^{(N)})$;

(ii) $y^{(j)} = \Psi_j(y^{(0)},\tilde{y}_{N-1},u^{(0)},\tilde{u}_j)$ for some analytic function Ψ_j locally defined and for all $j \geq N$.

In the uncontrolled case, the conditions (i) and (ii) of Lemma 2.2 give: (i) $y^{(N)} = \Psi_N(y^{(0)}, \tilde{y}_{N-1})$, and (ii) $y^{(j)} = \Psi_j(y^{(0)}, \tilde{y}_{N-1})$.

REMARK 2.1. *A priori condition (i) is necessary for Σ to satisfy the $ACP(N)$, and condition (ii) is sufficient.*

REMARK 2.2. *The second (resp. first) condition of Lemma 2.2 is equivalent to the phase variable property $PH(j)$ of any order $j \geq N$ (resp. the phase variable property $PH(N)$ of order N).*

3. The key lemma.

Here, we show a lemma about the ascending chain property, which will be used later on.

Let $(f_j, j > 0)$ be a sequence of analytic germs: $(X, x_0) \to (Y, f_j(x_0))$. X, Y are analytic manifolds. As we did previously in a particular case, we can associate a sequence of rings $\Re_j$ to the sequence (f_j), in the following way: we denote by $\Phi_j : X \to Y^j$ the map $\Phi_j(x) = (f_1(x), ..., f_j(x))$, and by $\Re_j$:

$$\Re_j = (\Phi_j)^*(O_{\Phi_j(x_0)}),$$

the pull back by the map Φ_j of the ring $O_{\Phi_j(x_0)}$ of germs of analytic maps at the point $\Phi_j(x_0)$. Clearly, again we have:

$$... \subset \Re_j \subset \Re_{j+1} \subset ... \subset O_{x_0}.$$

DEFINITION 3.1. *We say that the sequence (f_j) satisfies the $ACP(N)$ at x_0 if $\Re_j = \Re_N$ for $j \geq N$.*

DEFINITION 3.2. ***(of finite multiplicity)*** *$F : X \to Y$ has finite multiplicity at x_0 if $O_{x_0}/[F^*(\mathbf{m}(O_{y_0})).O_{x_0}]$ has finite dimension as a real vector space. Here $\mathbf{m}(O_{y_0})$ is the ideal of germs of analytic functions at (Y, y_0), $y_0 = F(x_0)$, which are zero at y_0.*

The dimension is the multiplicity.

There is a simple and convenient criterion for a germ to be of finite multiplicity:

F has finite multiplicity at x_0 iff there is an integer $r > 0$ such that:

$$[\mathbf{m}(O_{x_0})]^r \subset F^*(\mathbf{m}(O_{y_0})).O_{x_0}. \tag{29}$$

Therefore, to check that F has finite multiplicity at $x_0 = 0$, $(F : R^n \to Y)$, it is sufficient to check that $x_i^{r_i}$ belongs to $F^*(\mathbf{m}(O_{y_0})).O_{x_0}$ for some positive integers r_i, $i = 1, ..., n$.

Let $\Phi_N : (X, x_0) \to Y^N$, $\Phi_N = (f_1, ..., f_N)$, where the $f_i : (X, x_0) \to Y$ are germs of mappings at x_0. A "prolongation of Φ_N" is an arbitrary sequence $(\hat{f}_j)$ of germs of mappings, $\hat{f}_j : (X, x_0) \to Y$, such that $\hat{f}_j = f_j$ for $j \leq N$.

We have the following key lemma:

LEMMA 3.1. *(Jouan, [23]) The following properties are equivalent:*

(i) All prolongations of Φ_N satisfy the $ACP(k)$ for some $k \geq N$, (k depends on the prolongation),

(ii) Φ_N has finite multiplicity.

Why analyticity?

The proof of this lemma essentially follows from the **Weierstrass Preparation Theorem** (see [29]).

If we assume that our mappings are C^∞, then Lemma 3.1 is false. A counterexample is provided below.

The statement $(i) \Longrightarrow (ii)$ of the lemma is still valid in the C^∞ case but the statement $(ii) \Longrightarrow (i)$ is not: The proof of $(ii) \Longrightarrow (i)$ breaks down because $\Re_N$ is not in general Noetherian. We shall construct a smooth map Φ_2 on R^2 with finite multiplicity, and a smooth prolongation of it, which does not satisfy the $ACP(k)$ for any k. This will imply that $\Re_N$, in this counterexample, is not Noetherian.

Here, we show a smooth map Φ_2 on R^2, with finite multiplicity, and a smooth prolongation which does not satisfy the $ACP(k)$ for any k.

Counterexample: Let W be the "Weierstrass manifold", $W = \{(x_0, x_1, t) | t^2 + x_1 t + x_0 = 0\} \subset R^3$, and $\Pi : W \to R^2$, $(x_0, x_1, t) \to (x_0, x_1)$. Certainly, W is a smooth manifold. Set $f_0 = x_0$, $f_1 = x_1$ and $f_n = g_n(x_0, x_1)t$, with the sequence g_n constructed as follows.

Consider on R^2 the polar coordinates (r, θ), and the vector field X:

$$X = -r\frac{\partial}{\partial r} + \frac{\partial}{\partial \theta}.$$

We can construct some "spiraloid" disjoint subsets S_n of R^2 as follows: we pick an interval $I_1 = [a_1, b_1] \subset R_+^*$, I_1 small enough for $S_1 \cap \{x_1 = 0\} \neq \{x_1 = 0\}$, where S_1 is the union set of all trajectories of X passing through the points $(x^0, 0)$, $x^0 \in I_1$. Now, we chose a second interval $I_2 = [a_2, b_2]$, with $0 < a_2 < b_2 < a_1$ and $I_2 \cap S_1 \cap \{x_1 = 0\} = \varnothing$, and construct the set S_2 as the union set of all the trajectories of X through $(I_2, 0)$. Iterating the construction, we get S_n. We chose g_n in such a way that its support is $Int(S_n)$, the interior of S_n. This is possible since the complement of this set is closed, and since, given any closed set, there exists a C^∞ function having this set as zero set.

The multiplicity of $F = (f_0, f_1)$, $F : W \to R^2$ is finite at $(0, 0, 0)$ (it is 2).

We show that the sequence f_n does not satisfy the $ACP(k)$ for any k. For this, we work in an arbitrary small ball B centered at $(0, 0, 0)$ in W. We assume that $f_{n+1} = \Psi(f_0, f_1, ..., f_n)$ on B for some smooth Ψ. By construction, if $p = \Pi p'$ $\in Int(S_{n+1})$, then $f_{n+1}(p') = \Psi(f_0(p'), f_1(p'), 0, ...0)$. Let D be the discriminant set of W, i.e., $D = \{(x_0, x_1, t) | x_0 = \frac{1}{4}x_1^2\}$. We consider $c = (c_0, c_1)$, $c \in \Pi B \cap Int(S_{n+1}) \cap \Pi D$, $c' \in B \cap \Pi^{-1}(c)$, and a sequence (p'_k) in $\Pi^{-1}(Int(S_{n+1}))\backslash D$ such that $\lim_{k\to\infty} p'_k = c'$, and we set $p_k = \Pi p'_k$.

By definition, we have:

$$g_{n+1}(x_0, x_1)t = \Psi(x_0, x_1) \text{ on } Int(S_{n+1}).$$

Differentiating, we get:

$$\frac{\partial g_{n+1}}{\partial x_0}(x_0, x_1)t + g_{n+1}(x_0, x_1)\frac{\partial t}{\partial x_0} = \frac{\partial \Psi}{\partial x_0}(x_0, x_1),$$

which should hold at p_k,

$$\frac{\partial g_{n+1}}{\partial x_0}(p_k)t(p_k) + g_{n+1}(p_k)\frac{\partial t}{\partial x_0}(p_k) = \frac{\partial \Psi}{\partial x_0}(p_k).$$

Taking the limit when $k \to \infty$, we get:

$$\frac{\partial g_{n+1}}{\partial x_0}(c)t(c) + g_{n+1}(c)\frac{\partial t}{\partial x_0}(c) = \frac{\partial \Psi}{\partial x_0}(c),$$

where $g_{n+1}(c)$ is different from zero, and $t(c) = -\frac{c_1}{2}$. Hence, $\frac{\partial t}{\partial x_0}(c)$ is well defined.

But, $t^2 + x_1 t + x_0 = 0$ implies $\frac{\partial t}{\partial x_0} = -\frac{1}{2t+x_1}$. At c, $x_1 = c_1$, $t(c) = -\frac{c_1}{2}$, and $2t + x_1 = 0$. A contradiction.

Hence, despite the fact that the multiplicity is finite, the equality:

$$f_{n+1} = \Psi(f_0, ..., f_n),$$

never holds.

The main consequence of this lemma 3.1 is the following theorem.

THEOREM 3.2. *Let Σ be an uncontrolled system. Let $x_0 \in X$ be fixed. If for some k, Φ_k^Σ has finite multiplicity at x_0, then, Σ satisfies the $ACP(N)$ at x_0 for some $N \geq k$, and by Lemma 2.2, $y^{(N)} = \Psi_N(y^{(0)}, \tilde{y}_{N-1})$ for some analytic function Ψ_N defined in a neighbourhood of $\Phi_N^\Sigma(x_0)$.*

EXAMPLE 3.1. *$X = R$, $y = h(x)$, $\dot{x} = f(x)$, x_0 is arbitrary, h is nonconstant.*

In that case, of course, the notion of multiplicity is equivalent to the usual natural notion of multiplicity of a smooth function of a single variable. The multiplicity is always finite because h is nonconstant, and hence for some N, we have (locally): $y^{(N)} = \Psi_N(y^{(0)}, \tilde{y}_{N-1})$.

EXAMPLE 3.2. *$X = R^2$, $y = x_1$, $\dot{x}_1 = x_2^3$, $\dot{x}_2 = f(x_1, x_2)$, $x_0 = (0,0)$. This system is observable, and by our criterion, the multiplicity is finite. Hence, for some N, we have also: $y^{(N)} = \Psi_N(y^{(0)}, \tilde{y}_{N-1})$.*

Of course, it can happen that $\Phi_N = (f_1, ..., f_N)$ does not have finite multiplicity, but some particular prolongations $(\hat{f}_r)$ satisfy the $ACP(k)$ for some k. (just take the prolongation by the zero sequence for instance).

For uncontrolled systems, there are many other interesting examples where the $ACP(N)$ holds for some N, but the multiplicity is not finite. A case where the $ACP(N)$ holds everytime is the following:

EXERCISE 3.1. ***(Linear systems observed polynomially).*** *$X = R^n$, $y = p(x)$ is a polynomial, $\dot{x} = Ax$ is a linear vector field. Show that the $ACP(N)$ holds for some N. (compare with Exercise 4.8, Chapter 2).*

EXERCISE 3.2. $(\Sigma) : X = R^2$, $y = x_1(x_1^2 + x_2^2)$, $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

1. Show that Σ is observable.

By the previous exercise, the $ACP(N)$ holds: $y^{(2)} = -y^{(0)}$.

2. Show that the multiplicity is infinite. (For hints, see [**21**]*).*

The following important theorem is a consequence of more general results in the controlled case. It is a **globalization** of the previous theorem 3.2.

THEOREM 3.3. ***(Globalization of the $ACP(N)$ in the uncontrolled case).*** *Assume that X is compact, Σ is observable, and for each $x_0 \in X$, there is a N*

(depending on x_0, could be), such that Φ_N^Σ has finite multiplicity at x_0. Then there is a k and a C^∞ function φ, defined and compactly supported on R^{kd_y}, such that:

$$y^{(k)}(x) = \varphi(y^{(0)}(x), \tilde{y}_{k-1}(x)), \text{ for all } x \in X,$$

i.e., Σ satisfies $PH(k)$, the phase variable property of order k, globally on X.

4. The $ACP(N)$ in the controlled case.

As stated in Theorem 3.2, the local $ACP(N)$ holds in the uncontrolled case, as soon as there is a k such that Φ_k^Σ has finite multiplicity at the point under consideration. As we shall see, in the controlled case, the situation is not so clear cut.

Here, as we said, **the C^ω assumption can be relaxed.** Let us keep it for simplicity of exposition, anyway. But, in Chapter 6, we will use the results in the case where the systems are C^∞.

The main local result is the following.

THEOREM 4.1. *A point x_0 and an infinite jet $\underline{u_0}=(u_0^{(0)}, \widetilde{u_0})$ are fixed. Σ satisfies the $ACP(N)$ iff $\check{\Re} = \Re_N$. Moreover, in that case, $\Re_N = \Re_{N+1}$, $\Re_N \subset \hat{\Re}_N$.*

REMARK 4.1. *(i) If the $ACP(N)$ is true at $(x_0, u_0^{(0)}, \widetilde{u_0})$ for some $\widetilde{u_0}$, then,*

$$\check{\Re}_N = \check{\Re}_{N+1} \text{ at } (x_0, u_0^{(0)}),$$

(ii) this condition $\check{\Re}_N = \check{\Re}_{N+1}$ is implied by the fact that $\check{\Re}_{N_0}$ has finite multiplicity for some N_0 (in the sense that the map with components the generators of $\check{\Re}_{N_0}$, given by Formula (27), has finite multiplicity).

EXERCISE 4.1. $(\Sigma) : X = R^2$, $U = R$, $y = x_1$,

$$\begin{aligned} \dot{x}_1 &= x_2^3 - x_1, \\ \dot{x}_2 &= x_2^8 + x_2^4 u. \end{aligned}$$

We work at $x_0 = (0,0)$, and $u_0^{(0)} = 0$.

Show that:

1. Σ is observable,

2. $\check{\Re} = \Re_4 = \{G(x_1, x_2^3, x_2^{10}, x_2^{17}, u)\}$,

so that the $ACP(4)$ holds (in fact it holds as soon as $(x_0)_2 = 0$, and the $ACP(2)$ holds everywhere else).

3. Show that actually, $y^{(4)}$ can be written as a polynomial:

$$y^{(4)} = P(y^{(0)}, y^{(1)}, y^{(2)}, y^{(3)}, u^{(0)}, u^{(1)}, u^{(2)}),$$

and compute P.

5. Globalization.

We assume that Σ is given, and we fix a compact subset K of X. We also assume that Σ is differentially observable of order N. We denote by $\Phi_N^{\Sigma,K}$ the following mapping:

uncontrolled case: $\Phi_N^{\Sigma,K}$ is the restriction to K of Φ_N^Σ,

$\Phi_N^{\Sigma,K} : K \to \Phi_N^{\Sigma,K}(K) \subset R^{Nd_y}$,

controlled case: $\Phi_N^{\Sigma,K}$ is the restriction to $K \times U \times R^{(N-1)d_u}$ of $S\Phi_N^\Sigma$,

$\Phi_N^{\Sigma,K} : K \times U \times R^{(N-1)d_u} \to \Phi_N^{\Sigma,K}(K \times U \times R^{(N-1)d_u}) \subset R^{Nd_y} \times R^{Nd_u}$.

LEMMA 5.1. *$\Phi_N^{\Sigma,K}$ is a homeomorphism onto its image, which is closed.*

Comments:

(1) Lemma 5.1 shows that, as soon as Σ is differentially observable, x can be expressed on K as a continuous function φ_N^K of $(y^{(0)}, \tilde{y}_{N-1}, u^{(0)}, \tilde{u}_{N-1})$.

(2) If $X = R^n$, then, by Urysohn's lemma, φ_N^K can be extended to a continuous function defined on all of $R^{Nd_y} \times R^{Nd_u}$, hence x can be written as a continuous function, defined on all of $R^{Nd_y} \times R^{Nd_u}$:

$$x = \varphi_N^K(y^{(0)}, \tilde{y}_{N-1}, u^{(0)}, \tilde{u}_{N-1}).$$

(3) If $X = R^n$, (or if X is not R^n but φ_N^K is globally defined and continuous on $R^{Nd_y} \times R^{Nd_u}$), the classical assumption that φ_N^K is smooth is equivalent to the **strong** differential observability assumption (in restriction to K) . It is much stronger than the differential observability assumption made here. Of course, it implies (in the uncontrolled case) that the multiplicity is finite. Actually, the multiplicity is one.

It is the case in Chapters 2 and 3: in both chapters, strong differential observability holds (by assumption in Chapter 3, and as a consequence of uniform infinitesimal observability in Chapter 2).

EXERCISE 5.1. *Consider the system Σ :*

$$(\Sigma)\ \dot{x} = 1, y = \cos(x) + \cos(\alpha x),\ x \in R,$$

where α is an irrational number.

1. Show that Σ is observable,

2. Show that the observation space Θ^Σ of Σ is finite dimensional. Compare with Exercise 3.1.

So that the $ACP(N)$ holds for some N.

3. Show that x cannot be expressed as a continuous function of $(y^{(0)}, \tilde{y}_M)$, whatever M.

EXERCISE 5.2. *Show that, in the example 3.2 (uncontrolled):*

1. Σ is differentially observable of order 2,

2. depending on the choice of f, it can happen that Φ_N^Σ is an immersion for some N, or Φ_N^Σ is not an immersion for any N.

EXERCISE 5.3. *Show that, in Exercise 4.1:*

1. Σ is differentially observable of order 2,

2. $S\Phi_N^\Sigma$ is not an immersion for any N.

The main result is the following:

THEOREM 5.2. *Assume that Σ satisfies the $ACP(N)$ at each point, and is differentially observable of order N. Consider K, any fixed compact subset of X. Then, there exists a C^∞ function φ_N^K, **compactly supported w.r.t.** $(y^{(0)}, \tilde{y}_{N-1})$, such that:*

$$y^{(N)} = \varphi_N^K(y^{(0)}, \tilde{y}_{N-1}, u^{(0)}, \tilde{u}_N), \tag{30}$$

for all $x \in K$, all $u, \tilde{u}_N$. That is, Σ satisfies $PH(N)$, the phase variable property of order N, in restriction to K.

Compactly supported w.r.t. $(y^{(0)}, \tilde{y}_{N-1})$ means that, for any K', a compact subset of $U \times R^{Nd_u}$, φ_N^K restricted to $R^{Nd_y} \times K'$ is compactly supported. (It is not equivalent that φ_N^K is compactly supported, for all fixed $u, \tilde{u}_N$).

The following corollary will be also used in Chapter 6.

COROLLARY 5.3. *Theorem 5.2 is valid not only for $y^{(N)}$, but for any function α in $\hat{\Re}_{N+1} = \hat{\Re}_N\{u^{(N)}; u_0^{(N)}\}$ (i.e. the germs of α at each point belong to $\hat{\Re}_{N+1}$). It is true also for any function α in $\hat{\Re}_N$, and in that case:*

$$\alpha = \varphi_N^K(y^{(0)}, \tilde{y}_{N-1}, u^{(0)}, \tilde{u}_{N-1}), \tag{31}$$

for all $x \in K$, all $(u^{(0)}, \tilde{u}_{N-1})$.

EXAMPLE 5.1. *Example 3.2 and Exercise 4.1 (see also the exercises 5.2, 5.3) satisfy, in the uncontrolled and controlled cases, the assumptions of Theorem 5.2. In the case of Exercise 4.1, it has been already stated that Formula (30) holds globally, (for φ_N^K a certain polynomial).*

EXERCISE 5.4. *(single output, $d_y = 1$). Let Σ be a system with uniform canonical flag. Remember that Σ satisfies the phase variable property $PH(n)$ in restriction to small neighbourhoods of each point in X (Exercise 4.1, Chapter 2). Assume that moreover Σ is differentially observable of some order N. Show that Σ satisfies the $PH(N)$ (in restriction to any compact subset K of X).*

6. The controllable case.

Let us assume that Σ is controllable, in the usual weak sense of the transitivity of its Lie algebra (see Appendix 7, Chapter 1). We will use the theorems 6.1, 6.2, of Chapter 1, which in this case allow us to conclude that the **trivial foliation Δ_Σ is regular** , and equal to the foliation $\bar{\Delta}_\Sigma$, as was already stated just before Theorem 6.2 in Chapter 1.

Assume that these foliations are not trivial (i.e. their dimension is strictly > 0). Then, as was also stated in Chapter 1, Section 6, Σ cannot be observable, for any fixed input. Hence, Σ cannot be differentially observable.

The consequence in the (analytic) controlled case, **if Σ is controllable**, is that **this part of the theory is void** : assume that, as in the assumptions of Theorem 5.2, Σ satisfies the $ACP(N)$ and is differentially observable. Then, the "trivial foliation" has to be trivial, which implies that $\Xi^\Sigma(u)$ has full rank everywhere, hence $rank(d_X\check{\Re}) = n$. By Theorem 4.1, the $ACP(N)$ holds iff $\check{\Re} = \Re_N$, and in that case, $\Re_N \subset \hat{\Re}_N$. Therefore, $rank(d_X\hat{\Re}_N) = n$, and $S\Phi_N^\Sigma$ is an immersion. In fact, we are back to the situation of Chapters 2, or 3, where Σ is strongly differentially observable.

EXERCISE 6.1. *Study the controllability (in the weak sense) for the system of Exercise 4.1.*

CHAPTER 5

Observers: the high-gain construction.

The subject of this chapter is **observers**. The purpose of an observer is to obtain information about the state of the system, from the observed data.

In Section 1 of this chapter, we are going to discuss the concept of observers. The main ingredient of that concept is the notion of "estimation". There is no completely satisfactory definition of estimation. For that reason, we have to present several definitions of an observer, each having its domain of application.

We shall explain these different definitions of observers, and point out the relations between them.

In the remainder of the chapter, we shall construct explicitly several types of observers. The fundamental idea behind all of these constructions is to use the classical observers for linear systems and to kill the nonlinearities by an appropriate **time rescaling**.

The construction we present, and its variations, provide explicit, efficient, and robust algorithms for state estimation. It is closely related to the results of the three previous chapters and it applies in all the cases dealt with in these chapters.

Our observers can be used for several purposes:
a)-state estimation in itself,
b)-dynamic output stabilization.
They will be used in the next chapter 6 for the purpose b).

There are several problems in defining an observer. First, there is no good definition of a state observer when **the state space X is not compact**. A second difficulty is the "peak phenomenon", explained below, for observers with arbitrary exponential decay.

Finally, let us point out that our construction of observers is related to nonlinear filtering theory. But this is beyond the scope of this book. A good reference for this relation is [8].

In this chapter, we will make the following basic assumption: the system (Σ) is **differentially observable** of a certain order $N \geq 1$.

1. Definition of observer systems and comments.

An observer system is a system Σ_O, the inputs of which consist of the "**observed data**" of Σ, i.e. the inputs of Σ , their derivatives, and the outputs of Σ. The task of the observer is the estimation of the state of Σ.

Let us make a few remarks. The inputs are selected by the "operator" of the system. In particular, he can chose them differentiable, and then, their derivatives are known. On the other hand, it is hard to estimate the derivatives of the outputs from the knowledge of them. For this reason, we strictly avoid any use of these

derivatives in our theory: actually, the first problem we shall deal with, will be to estimate the derivatives of the outputs.

Because Σ is differentially observable, **for sufficiently smooth controls**, estimating the state is equivalent to estimate the $N-1$ first derivatives of the outputs, $\tilde{y}_{N-1}$.

We denote by $U^{r,B}$ the set of inputs $u : [0,\infty[\to U = I^{d_u}$, such that u is $r-1$ times continuously differentiable, its r^{th} derivative belongs to $L^\infty([0,\infty[; R^{d_u})$ and all the derivatives up to order r are bounded by $B > 0$. $U^{0,B}$ is just the subset of $L^\infty([0,\infty[; U)$ formed by the $u(.)$ that are bounded by B. (Here, $U = I^{d_u}$ is not necessarily compact: $I = R$ is possible), $||.||$ **is the canonical Euclidean norm on R^{d_y} or on R^{Nd_y}.**

1.1. Output observers.

1.1.1. *Definitions.* We use again the notation $\tilde{u}_r = (u^{(1)}, ..., u^{(r)})$, introduced in Chapter 4, Section 1.

DEFINITION 1.1. *An $U^{r,B}$ output observer of Σ, relative to Ω is a system $(\Sigma_{Oy}^{r,B,\Omega})$, $r \geq N$:*

$$(\Sigma_{Oy}^{r,B,\Omega}) \begin{cases} \frac{dz}{dt} = F(z, u^{(0)}(t), \tilde{u}_r(t), y^{(0)}(t)), \\ \eta(t) = \mathcal{H}(z(t), u^{(0)}(t), \tilde{u}_r(t), y^{(0)}(t)), \end{cases} \tag{32}$$

on the d_Z dimensional manifold Z, where $\Omega \subset X$ is open, where η, the output, belongs to R^{Nd_y}, F and $\mathcal{H}$ are C^∞, and satisfy the following condition:

for all $u \in U^{r,B}$, for all $x_0 \in \Omega$, such that the corresponding semi trajectory of Σ, $x(t, x_0)$, is defined on $[0, +\infty[$ and stays in Ω, for all $z_0 \in Z$, the output $\eta(t)$ is well defined and,

$$\lim_{t \to +\infty} ||\eta(t) - \underline{y}_N(t)|| = 0. \tag{33}$$

REMARK 1.1. *We will be mostly interested in two cases: X is compact and $\Omega = X$, or X is noncompact but Ω is relatively compact.*

Definition 1.2 below strengthens Definition 1.1.

DEFINITION 1.2. *An **exponential** $U^{r,B}$ output observer of Σ, relative to Ω, is a one parameter family of output observers of Σ, depending on the real parameter $\alpha > 0$, with state manifold Z, independent of α, which satisfies the following condition (34), strengthening (33):*

for any $\bar{K}$, a compact subset of Z, for all $z_0 \in \bar{K}$, for all $x_0 \in \Omega$, for all $u \in U^{r,B}$:

$$||\eta(t) - \underline{y}_N(t)|| \leq k(\alpha) e^{-\alpha t} ||\eta(0) - \underline{y}_N(0)||, \tag{34}$$

as long as $x(t, x_0)$ stays in Ω , where $k : R_+ \to R_+$ is a function of polynomial growth, depending on $\bar{K}$ in general.

Such a one parameter family will be typically denoted by $(\Sigma_{Oye}^{r,B,\Omega})$.

REMARK 1.2. *The fact that $k(\alpha)$ has polynomial growth warrants that, if α is large enough, **the estimate can be made arbitrarily close to the real value in arbitrary short time.***

1.1.2. *Comment.* Let us assume that $(\Sigma_{Oy}^{r,B,\Omega})$ is an "output observer" and $r = N$. Since Σ is differentially observable of order N, we obtain an estimation $\hat{x}(t)$ of the state $x(t)$ of Σ as follows.

For Ω' open, $cl(\Omega) \subset \Omega'$, denote by $E(t, z_0)$ the set:

(35)
$$\{x^* \in cl(\Omega') |\ ||\eta(t) - \Phi_N^\Sigma(x^*, u(t), \tilde{u}_{N-1}(t))|| = \inf_{x \in \Omega'} ||\eta(t) - \Phi_N^\Sigma(x, u(t), \tilde{u}_{N-1}(t))||\}.$$

If Ω is relatively compact, then Ω' can be taken relatively compact. In that case, this set $E(t, z_0)$ is not empty, and for any metric d on X, compatible with its topology, $\lim_{t \to +\infty} d(E(t, z_0), x(t)) = 0$. **If Ω is not relatively compact, this is not true.**

EXERCISE 1.1. *In this situation, show that* $\lim_{t \to +\infty} \#(E(t, z_0)) = 1$, *if moreover* Σ *is* ***strongly*** *differentially observable (of order N), for a trajectory $x(t, x_0)$ that stays in Ω for all $t \geq 0$.*

1.1.3. *The observability distance.* The following could be a way to overcome the problems linked to the noncompactness of X or Ω: one should try to construct a canonical distance over X, related to the observability properties. This canonical distance could then be used in the definition of observers.

A trivial way to do this is to define the distance on X :

$$d_O(x, y) = \sup_{||u^{(0)}||, ||\tilde{u}_{N-1}|| \leq B} ||\Phi_N^\Sigma(x, u^{(0)}, \tilde{u}_{N-1}) - \Phi_N^\Sigma(y, u^{(0)}, \tilde{u}_{N-1})||. \tag{36}$$

EXERCISE 1.2. *Show that (36) actually defines a distance over X.*

Unfortunately, this distance is not compatible with the topology of X in general:

EXERCISE 1.3. *Consider the system of Exercise 5.1, Chapter 4, (uncontrolled case). Show that the observability distance is not compatible with the topology of X.*

Moreover, this distance is not very natural because it depends on both B (the bound on the input and its derivatives) and on N (the degree of differential observability).

There is a special case where the situation is better: if X is compact, and if Σ is analytic, uncontrolled and just observable, then by the proof of Theorem 3.3 in the previous chapter 4, there is an N such that Σ is differentially observable of order N. Taking the smallest such N, we get a canonical observability distance, in that case. Unfortunately, this is not very interesting, because X is compact.

EXERCISE 1.4. *Show that this distance is compatible with the topology of X.*

1.2. State observers.

1.2.1. *Definitions.*

DEFINITION 1.3. *An $U^{r,B}$ state observer of Σ, relative to Ω, is a system* $(\Sigma_{Ox}^{r,B,\Omega})$:

$$\frac{dz}{dt} = F(z, u^{(0)}, \tilde{u}_r, y^{(0)}), \eta = \mathcal{H}(z, u^{(0)}, \tilde{u}_r, y^{(0)}), \tag{37}$$

on the d_Z dimensional manifold Z, where $\Omega \subset X$ is open, η, the output, belongs to X, F and $\mathcal{H}$ are C^∞, and satisfy the condition:

for all $u \in U^{r,B}$, for all $x_0 \in \Omega$, such that the corresponding semi trajectory of Σ, $x(t,x_0)$, is defined for $t \in [0,+\infty[$ and stays in Ω , for all $z_0 \in Z$, the output $\eta(t,z_0)$ is well defined and,

$$\lim_{t\to+\infty} d(\eta(t,z_0),x(t,x_0)) = 0, \tag{38}$$

where d is any metric compatible with the topology of X.

Again, **this definition makes sense if Ω is relatively compact only.**

DEFINITION 1.4. *An exponential $U^{r,B}$ state observer of Σ, relative to Ω, typically denoted by $(\Sigma^{r,B,\Omega}_{Oxe})$, is a one parameter family of state observers for Σ, depending on the real parameter $\alpha > 0$ (on the same manifold Z), which satisfies the following condition (39), strengthening (38):*

for any compact $\bar{K} \subset Z$, for any Riemannian distance d on X, there exists $a > 0$, and $k : R+ \to R+$ with polynomial growth, k and a depending on d and $\bar{K}$, such that:

for all $x_0 \in \Omega$, for all $u \in U^{r,B}$, for all $z_0 \in \bar{K}$,

$$Inf[a, d(\eta(t,z_0),x(t,x_0))] \leq k(\alpha)e^{-\alpha t}d(\eta(0,z_0),x_0), \tag{39}$$

as long as $x(t,x_0)$ stays in Ω .

It is important to note that, in the preceding definition, one can replace "there exists $a > 0$" by "for all a, $0 < a \leq a_0$".

Again, if X is not compact, and Ω is not relatively compact, the inequality (39) also does not make sense: all Riemannian metrics are not equivalent, hence the inequality (39) cannot hold for all Riemannian metrics.

REMARK 1.3. *There is no hope to have a reasonable theory if we ask that the condition (39) in Definition 1.4 to be valid for any distance on X (compatible with the topology of X), even if Ω is relatively compact. But for Riemannian distances, everything is fine, since differentiable mappings between Riemannian manifolds are locally Lipschitz.*

REMARK 1.4. *The inequality (39) is more complicated than the inequality (34), due to the "**peak phenomenon**", well known to engineers (and to control theoretists). In fact, it happens already in the linear theory. We explain it now.*

1.2.2. *Peak phenomenon.* Assume that Ω is relatively compact. Let d be a given Riemannian metric, and assume that the following inequality is satisfied, instead of (39):

$$d(\eta(t,z_0).x(t,x_0)) \leq k_d(\alpha)e^{-\alpha t}d(\eta(0,z_0),x_0), \tag{40}$$

where k_d has polynomial growth.

(40) cannot hold for all Riemannian metrics on X. This is due to the "*peak phenomenon*":

It can happen that there exists a trajectory $x(t,x_0)$ of Σ, with $x_0 \in \Omega$, and a function $\alpha \in R^*_+ \to t_\alpha \in R_+$, which tends to zero as α tends to $+\infty$, and such that, if $\eta_\alpha(t,z_0)$ denotes a corresponding trajectory of the observer, $\eta_\alpha(t_\alpha,z_0) \to \infty$ as α tends to $+\infty$.

One can construct a Riemannian metric on X such that for the associated distance function δ on X. $\delta(\eta_\alpha(t_\alpha,z_0),\ x(t_\alpha,x_0))$ tends to $+\infty$ faster than any power of α.

The "peak phenomenon" already occurs in the linear theory for the classical Luenberger's or Kalman's observers.

Of course, estimations of x that do not belong to Ω are irrelevant, but this is unimportant: the only important point is that relevant estimations of x are obtained in arbitrary short time, for α large enough.

1.2.3. *Consistency of our definition of an exponential state observer.*

EXERCISE 1.5. *Show that, on a manifold X, and on any compact set $C \subset X$, the distances induced on C by Riemannian distances are all equivalent.*

LEMMA 1.1. *If Ω is relatively compact, then Definition 1.4 is independent of the choice of the Riemannian metric d on X.*

1.2.4. *Alternative definitions of an exponential state observer.* We give now two other **apparently different** definitions, but more tractable than Definition 1.4.

DEFINITION 1.5. *An exponential $U^{r,B}$ state observer of Σ, relative to Ω, is a one parameter family of state observers for Σ, depending on the real parameter $\alpha > 0$ (on the same manifold Z independent of α), which satisfies the following condition:*

***There exists a Riemannian metric** d such that relation (40) holds, for all $x_0 \in \Omega$, for all $\bar{K}$ compact, for all $z_0 \in \bar{K}$, for all $u \in U^{r,B}$, as long as $x(t, x_0)$ stays in Ω. The function k_d has polynomial growth and depends on $\bar{K}$.*

DEFINITION 1.6. *An exponential $U^{r,B}$ state observer of Σ, relative to Ω, is a one parameter family of state observers for Σ, depending on the real parameter $\alpha > 0$ (on the same manifold Z independent of α), which satisfies the following condition:*

***There exists a Riemannian metric** d such that relation (40) holds, for all $x_0 \in \Omega$, for all $z_0 \in Z$, for all $u \in U^{r,B}$, as long as $x(t, x_0)$ stays in Ω. The function k_d has polynomial growth.*

Of course, Definition 1.6 is strictly contained in Definition 1.5. On the other hand, if Ω is relatively compact, by Lemma 1.1, Definition 1.5 is itself contained in Definition 1.4. But, in fact, we have:

PROPOSITION 1.2. *Definitions 1.5, 1.4 are equivalent.*

The only motive to introduce the equivalent definition 1.4 is that it is independent of any special Riemannian metric over X.

The observers that we will construct will be of two types, as in the classical linear theory: 1) Luenberger type, 2) Kalman's type.

For the Luenberger case, the statement of Definition 1.6 is valid, and in the Kalman's case, the statement of Definition 1.5 is valid. This is true both for the classical linear theory and our nonlinear theory.

1.3. Relations between state observers and output observers. (1) Assume that $\Sigma_{Oy}^{r,B,\Omega}$ is an **output** observer. If $X = R^n$, if Ω is relatively compact, then, by Lemma 5.1 and the comment just after, in the previous chapter 4, there is a continuous function $\varphi_N^\Omega : R^{Nd_y} \times R^{Nd_u} \to X$, such that $x = \varphi_N^\Omega(y, \tilde{y}_{N-1}, u, \tilde{u}_{N-1})$. This function provides a continuous single-valued estimation $\hat{x}(t)$ of $x(t, x_0) : \hat{x}(t) = \varphi_N^\Omega(\eta(t), u(t), \tilde{u}_{N-1}(t))$. If $|.|$ denotes any norm on $X = R^n$, one has:

$$\lim_{t \to +\infty} |x(t) - \hat{x}(t)| = 0. \tag{41}$$

Therefore, in that case, the output observer $\Sigma_{Oy}^{r,B,\Omega}$ allows to construct an $U^{r,B}$ state observer, relative to Ω .

(2) The assumptions are the same as in (1) but Σ is **strongly** differentially observable of order N, and (34) holds. Then, the function φ_N^Ω can be taken smooth, compactly supported, and if d is any Riemannian distance over X, $d(x(t), \hat{x}(t)) \leq k_d(\alpha)\ e^{-\alpha t}$ for some function k_d with polynomial growth , depending on d, Ω , $\bar{K}$.

(3) Assume that $\Sigma_{Ox}^{r',B,\Omega}$, $r' \geq 0$ is a **state** observer. Then, a fortiori, it is an $U^{r,B}$ state observer for some r_0, for all $r \geq r_0 \geq N$. We know that Ω is relatively compact. We can use the mapping Φ_r^Σ in order to construct an $U^{r,B}$ output observer as follows: we can replace Φ_r^Σ by a smooth (C^∞) mapping $\tilde{\Phi}_r^\Sigma$, which is constant outside a compact set, and the restriction of which to $\Omega \times V$ coincides with Φ_r^Σ. Here, V is the set of $(r-1)$ jets at $t=0$ of control functions $u(t)$, the $r-1$ first derivatives of which are bounded by B ($V = (U \cap (I_B)^{d_u}) \times (I_B)^{(r-1)d_u}$). Taking the composition $\tilde{\mathcal{H}}$ of this mapping $\tilde{\Phi}_r^\Sigma$ with $\mathcal{H}$, as the output mapping of the observer, we get an $U^{r,B}$ output observer relative to Ω.

If $\Sigma_{Ox}^{r,B,\Omega}$ is exponential, then: 1) $||\tilde{\Phi}_r^\Sigma(\eta(t,z_0), \underline{u}_r(t)) - \tilde{\Phi}_r^\Sigma(x(t,x_0), \underline{u}_r(t))|| \leq \lambda d(\eta(t,z_0), x(t,x_0))$ for λ large enough. 2) $||\tilde{\Phi}_r^\Sigma(\eta(t,z_0), \underline{u}_r(t) - \tilde{\Phi}_r^\Sigma(x(t,x_0), \underline{u}_r(t))|| \leq M$ because $\tilde{\Phi}_r^\Sigma$ is single-valued outside a compact. Hence, for λ large enough:

$$||\tilde{\Phi}_r^\Sigma(\eta(t,z_0), \underline{u}_r(t)) - \tilde{\Phi}_r^\Sigma(x(t,x_0), \underline{u}_r(t))|| \leq \lambda\ Inf(d(\eta(t,z_0), x(t,x_0)), \frac{M}{\lambda}),$$

$$||\tilde{\Phi}_r^\Sigma(\eta(t,z_0), \underline{u}_r(t)) - \tilde{\Phi}_r^\Sigma(x(t,x_0), \underline{u}_r(t))|| \leq \lambda\ k(\alpha) e^{-\alpha t} d(\eta(0,z_0), x_0).$$

EXERCISE 1.6. *($d_y > d_u$, $r \geq 2n+1$.) If moreover Σ is **strongly differentially observable** of order r, prove that $\tilde{\Phi}_r^\Sigma$ can be chosen so that, $d(\eta(0,z_0), x_0) \leq \gamma\ ||\tilde{\Phi}_r^\Sigma(\eta(0,z_0), \underline{u}_r(0)) - \tilde{\Phi}_r^\Sigma(x(0,x_0), \underline{u}_r(0))||$, for all $z_0 \in \bar{K}$, all $x_0 \in \Omega$, where γ depends on the compact $\bar{K}$. This shows that the modified observer is an $U^{r,B}$ exponential output observer.*

2. The "high gain construction".

2.1. Discussion about the "high-gain construction". The "high-gain construction" is a general way to construct either state or output $U^{r,B}$ observers, that are moreover **exponential**. Before explaining this construction, we want to point out a certain number of facts, concerning the results in this chapter.

1. Systems with a phase variable representation: for the systems appearing in Chapters 3, 4, we obtain a phase variable representation of a certain order N. As we shall see, our "high-gain observers" work for systems in the phase variable representation, for C^N controls. Hence they apply to these general classes of systems.

2. Systems with a uniform canonical flag: (the single output controlled case of Chapter 2). In that case, as we know (see Exercise 5.4, Chapter 4), Σ satisfies the phase variable property of some order, either locally or in restriction to arbitrarily large compact sets if moreover Σ is differentially observable.

Hence, the construction also applies for sufficiently smooth inputs. **But there is a stronger result: If Σ has the observability canonical form** (15) of

Chapter 2, then, **our observers work also for arbitrary L^∞ inputs,** i.e. they are $U^{0,B}$ observers.

3. The high gain construction has mainly two versions: referring to the terminology of linear systems theory, the first version is in the "Luenberger style" and the second version is in the "Kalman filter style" (a deterministic version of the Kalman filter).

4. There are versions of the "high gain observer" that are "continuous-continuous", and others that are "continuous-discrete": continuous- continuous means that the observer equation are ODE'S, and observations are continuous functions of time. In the continuous-discrete version, which is more realistic, the observer equations are ODE'S with jumps, and observations are sampled.

2.2. The "Luenberger style" observer. This section will concern uniformly infinitesimally observable systems.

Let us assume that $X = R^n$ and our system Σ, **analytic**, has the observability canonical form (15) **globally**. Recall that this canonical form exists locally as soon as Σ has a uniform canonical flag.

Let us denote by $\underline{x}_i$ the vector $(x^0, ..., x^i)$. The following two additional assumptions will be crucial.

(A_1) Each of the maps f_i, $i = 0, ..., n-1$, is globally Lipschitz w.r.t. $\underline{x}_i$, uniformly with respect to u and x^{i+1},

(A_2) there exists two real α, β, $0 < \alpha < \beta$, such that

$$\alpha \leq |\frac{\partial h}{\partial x^0}| \leq \beta, \tag{42}$$

$$\alpha \leq |\frac{\partial f_i}{\partial x^{i+1}}| \leq \beta, 0 \leq i \leq n-2.$$

In fact, these assumptions can be automatically satisfied, as soon as one is interested only in the trajectories that stay in a given compact convex set $\Gamma \subset X = R^n$, as shows the following exercise.

EXERCISE 2.1. *Assume that $X = R^n$, $\Gamma \subset X$ is the closure of an open, relatively compact, convex subset of X, and Σ has the normal form (15), (it is sufficient that it has this normal form only on Γ).*

Show that, for all $B > 0$, Σ can be extended smoothly (C^∞) outside of $\Gamma \times V_B$, so that the assumptions A_1, A_2, are satisfied (globally) on $X \times U$. (Here, $V_B = \{u \in U;\ |u| \leq B\}$).

In order to prove the main result of this section, the following technical lemma is crucial:

LEMMA 2.1. *Consider time-dependant real matrices $A(t)$ and $C(t)$:*

$$A(t) = \begin{pmatrix} 0, \varphi_2(t), ..., 0 \\ 0, 0, \varphi_3(t), ..., 0 \\ \cdot \\ \cdot \\ 0, 0, ..., 0, \varphi_n(t) \\ 0, 0, ..., 0, 0 \end{pmatrix}, \ C(t) = (\varphi_1(t), 0, ..., 0).$$

$A(t)$ is $n \times n$, and $C(t)$ is $1 \times n$. Assume that there are two real constant α, β, such that:

$$0 < \alpha < \beta, \alpha < \varphi_i(t) < \beta, 1 \leq i \leq n.$$

Then, there is a real $\lambda > 0$, a vector $\bar{K} \in R^n$, and a symmetric positive definite $n \times n$ matrix S, λ and S depending on α, β only, such that:

$$(A(t) - \bar{K}C(t))'S + S(A(t) - \bar{K}C(t)) \leq -\lambda\, Id.$$

Here, $(A(t) - \bar{K}C(t))'$ means transpose of $(A(t) - \bar{K}C(t))$ and $\leq$ is the (partial) ordering of symmetric matrices defined by the cone of symmetric positive semi-definite matrices.

Now, let us define the dynamics of our "observer system" Σ_O as follows:

$$\frac{d\hat{x}}{dt} = f(\hat{x}, u) - K_\theta(h(\hat{x}, u) - y), \tag{43}$$

where $K_\theta = \Delta_\theta K$, $\Delta_\theta = diag(\theta, \theta^2, ..., \theta^n)$ for $\theta > 1$, and K (together with S and λ) comes from Lemma 2.1, relative to α, β, in the assumption A_2, (42).

THEOREM 2.2. *For any $a > 0$, there is a $\theta > 1$ (large enough) such that:*

$$\forall (x_0, \hat{x}_0) \in X \times X, ||\hat{x}(t) - x(t)|| \leq k(a)e^{-at}||\hat{x}_0 - x_0||, \tag{44}$$

for some polynomial k, of degree n, where $\hat{x}(t)$ and $x(t)$ denote the solutions at time t of the observer system Σ_O and the system Σ, with respective initial conditions $\hat{x}_0$, x_0.

COROLLARY 2.3. *For all $B > 0$, for any relatively compact $\Omega \subset X$, the system Σ_O given by Formula (43) is an $U^{0,B}$ exponential state observer for Σ, relative to Ω .*

COROLLARY 2.4. *Let Ω be an open relatively compact convex subset of $X = R^n$. Assume that the restriction $\Sigma_{|cl(\Omega)}$ is* ***globally*** *in the observability canonical form (15). Then, for all $B > 0$, there is a $U^{0,B}$ exponential state observer for Σ, relative to Ω .*

Observation: in both corollaries, the polynomial $k(a)$ is independent of the compact set $\bar{K}$ in the definition of the observer.

Corollary 2.3 is an immediate consequence of Theorem 2.2 and of the definition of a state observer. Corollary 2.4 is a consequence of Exercise 2.1 and of Theorem 2.2.

2.3. The case of a phase-variable representation. Here, d_y is arbitrary.

Let us assume that we have a system in the phase-variable representation, $y^{(N)} = \varphi(y^{(0)}, \tilde{y}_{N-1}, u^{(0)}, \tilde{u}_N)$, then, the previous construction can be adapted to obtain an exponential $U^{N,B}$ output observer, at the cost of an additional assumption:

(A_3) φ is compactly supported w.r.t. $(y^{(0)}, \tilde{y}_{N-1})$. (Remember that this means that for any K', a compact subset of $U \times R^{Nd_u}$, the restriction of φ to $R^{Nd_y} \times K'$ is compactly supported).

As we know, this assumption is satisfied in many situations, for instance:

-In Chapter 4, in the situation of Theorem 5.2,

-In particular, in the case where $d_y > d_u$, we have generically a phase variable representation. If we restrict Σ to a compact subset $\bar{\Omega}$ of X, Theorem 5.2 gives us a phase-variable representation for Σ, with the additional property (A_3).

Let us denote by A the (Nd_y, d_y) block-antishift matrix:

$$A = \begin{pmatrix} 0, Id_{d_y}, 0, ..., 0 \\ 0, 0, Id_{d_y}, 0, ..., 0 \\ 0, \\ 0, 0, ..., 0, Id_{d_y} \\ 0, 0, ..., 0, 0 \end{pmatrix}.$$

$\hat{z}$ denotes a typical element of R^{Nd_y}. Then $\frac{d\hat{z}}{dt} = A\hat{z}$ is the linear ODE on R^{Nd_y} corresponding to the vector field:

$$A(\hat{z}) = \sum_{\substack{i=1,..,N-1 \\ j=1,d_y}} \{\hat{z}_{id_y+j} \frac{\partial}{\partial \hat{z}_{(i-1)d_y+j}}\}.$$

$b(\hat{z}, u^{(0)}, \tilde{u}_N)$ is a vector field on R^{Nd_y}, depending on $u^{(0)}, \tilde{u}_N$, defined as follows:

$$b(\hat{z}, u^{(0)}, \tilde{u}_N)) = \sum_{j=1}^{d_y} \varphi_j(\hat{z}, u^{(0)}, \tilde{u}_N) \frac{\partial}{\partial \hat{z}_{(N-1)d_y+j}}.$$

$C : R^{Nd_y} \to R^{d_y}$ is the linear mapping with matrix:

$$C = (Id_{d_y}, 0, ..., 0),$$

i.e. $C(\hat{z})$ denotes the vector of R^{d_y} the components of which are the d_y first components of $\hat{z}$.

DEFINITION 2.1. *A square matrix A is called stable if all its eigenvalues have strictly negative real parts.*

Consider the system Σ_O, on R^{Nd_y}, with output $\eta \in R^{Nd_y}$,

$$(\Sigma_O) \quad \frac{d\hat{z}}{dt} = A(\hat{z}) + b(\hat{z}, u^{(0)}, \tilde{u}_N) - K_\theta(C(\hat{z}) - y(t)), \eta = \hat{z}, \tag{45}$$

where $\theta > 1$ is a given real, $K_\theta = \Delta_\theta K$, Δ_θ is the block diagonal matrix:

$$\Delta_\theta = Block - diag(\theta Id_{d_y}, \theta^2 Id_{d_y}, ..., \theta^N Id_{d_y}),$$

and K is such that the matrix $A - KC$ is a stable matrix.

EXERCISE 2.2. *Prove that such a K does exist.*

THEOREM 2.5. *Σ_O is an exponential $U^{N,B}$ output observer relative to an arbitrarily large relatively compact $\Omega \subset X$ (if the assumption A_3 is satisfied. If not, one has first to modify the mapping φ outside $S\Phi_N^\Sigma(\Omega \times R^{Nd_u})$ in order that A_3 be satisfied. In that case, the observer system depends on Ω).*

Observation: again, the function k in the definition of the exponential output observer does not depend on the compact $\bar{K}$ (in which the observer is initialized). This will not be the case any more in the next paragraph.

2.4. The "extended Kalman filter style" construction.

2.4.1. *Introduction and main result.* No prerequisites about the Kalman filter are required to understand this section. We remain in the deterministic setting, and complete proofs of all results can be given in a very elementary way. Let us just add a few "comments", about the"extended Kalman filter".

The "extended Kalman filter" (for simplicity denoted by "E.K.F.") applies the linear time dependant version of the Kalman filter, to the linearized system along the **estimate of the trajectory**. If it was along the real trajectory, then, the procedure would be perfectly well defined. But, it has to be along the **estimate** of the trajectory, since the real trajectory is unknown: the purpose of the filter is precisely to estimate it.

Because of this, it is an easy exercise to check that the equations of the "extended Kalman filter" are not intrinsic. They depend on the coordinate system. They were introduced by the engineers, and they perform very well in practice, because they take the noise into account.

Our point of view in this section is like that:

1) We use special coordinates, for instance, the special coordinates of the uniform observability canonical form (17), in the single output **control affine** case, or, the coordinates of a phase variable representation in other cases. These coordinates are essentially uniquely defined, hence, **the extended Kalman filter written in these coordinates, becomes a well defined object,**

2) it is possible to adapt the high gain construction shown in the previous section in order that the equations of the E.K.F. in the special coordinates give the same results as in the previous section (arbitrary exponential convergence of the estimation error).

The main difference with the Luenberger style version, is that the **correction term** "K_θ" is not constant: it is computed as a function of the information appearing at the current time t. We have observed in the applications that the E.K.F. performs very well in practice, probably for this reason.

Let us present this construction in the single-output, **control-affine** case: this last requirement seems essential. We make the following assumption:

a1) Σ is globally in the normal form (17).

This is true in several situations, for example if Σ is observable and if we make one of the following assumptions (a2), (a3).

a2) $\Phi = (h, L_f h, ..., L_f^{n-1} h)$ is a global diffeomorphism,

or, weaker,

a3) in restriction to Γ, the closure of an open relatively compact subset of X, Φ is a diffeomorphism.

Remember that the observability assumption implies that Φ should be almost everywhere a local diffeomorphism from X into R^n, and that, in the coordinates defined by Φ, the system Σ has to be in the normal form (17). (See Theorem 4.1, Chapter 2). In the case of the assumption a3), all the functions φ and g_i in the normal form (17) can be extended to all of R^n so that they are smooth and compactly supported w.r.t. all their arguments, and g_i depends on $(x_1, ..., x_i)$ only.

Let us recall the normal form (17):

$$\dot{x} = \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ . \\ . \\ . \\ \dot{x}_{n-1} \\ \dot{x}_n \end{pmatrix} = \begin{pmatrix} x_2 \\ x_3 \\ . \\ . \\ . \\ x_n \\ \varphi(x) \end{pmatrix} + u \begin{pmatrix} g_1(x_1) \\ g_2(x_1, x_2) \\ . \\ . \\ . \\ g_{n-1}(x_1, ..., x_{n-1}) \\ g_n(x) \end{pmatrix}$$

$$y = x_1$$

Let us assume moreover that:

a4) φ and g_i are globally Lipschitz. In the case a3), this will be automatically true, by what we just said.

Denote again by A the antishift matrix:

$$A = \begin{pmatrix} 0,1,0,....,0 \\ ... \\ ... \\ 0,.....0,1 \\ 0,0,.....,0,0 \end{pmatrix},$$

C is the linear form over R^n with matrix $C = (1,0,.....,0)$. Let us rewrite the normal form (17) in matrix notations as follows:

$$\dot{x} = Ax + b(x,u),\ y = Cx, \tag{46}$$

where b_i, the i^{th} component of b depends only on $\underline{x}_i = (x_1, ..., x_i)$ and u.

Let Q be a given symmetric positive definite $n \times n$ matrix. r, θ are positive real numbers, $\Delta_\theta = diag(1, \frac{1}{\theta}, ..., (\frac{1}{\theta})^{n-1})$. Let $b^*(x,u)$ denote the Jacobian matrix of $b(x,u)$ w.r.t. x. Set $Q_\theta = \theta^2(\Delta_\theta)^{-1}Q(\Delta_\theta)^{-1}$. The following equations:

$$(i)\ \frac{dz}{dt} = Az + b(z,u) - S(t)^{-1}C'r^{-1}(Cz - y(t)), \tag{47}$$

$$(ii)\frac{dS}{dt} = -(A + b^*(z,u))'S - S(A + b^*(z,u)) + C'r^{-1}C - SQ_\theta S,$$

$$\eta = z, \tag{48}$$

define what is called the "extended Kalman filter" for our system (46), (Q_θ and r are analogous to the covariance matrices of the state noise and the output noise in the stochastic context).

The following theorem holds:

THEOREM 2.6. *Under the assumptions (a1), (a4), for $\theta > 1$, for all $T > 0$, the extended Kalman filter (47) satisfies, for $t \geq \frac{T}{\theta}$:*

$$||z(t) - x(t)|| \leq \theta^{n-1}k(T)\ ||z(\frac{T}{\theta}) - x(\frac{T}{\theta})||\ e^{-(\theta\omega(T) - \mu(T))(t - \frac{T}{\theta})}, \tag{49}$$

for some positive continuous functions $k(T), \omega(T), \mu(T)$.

COROLLARY 2.7. *Under the assumptions (a1), (a4), for any open relatively compact $\Omega \subset X = R^n$, for any $B > 0$, the extended Kalman filter is an exponential $U^{0,B}$ state observer, relative to Ω.*

In fact, this theorem and this corollary generalize to the case of a multi output system, having a phase variable representation. Let us assume that Σ has the phase variable representation $y^{(N)} = \varphi(y, \tilde{y}_{N-1}, u, \tilde{u}_N)$, and that φ is compactly supported w.r.t. $(y, \tilde{y}_{N-1})$, as in Section 2.3.

We consider systems on R^{Np}, of the general form:

$$\frac{dx}{dt} = A_{N,p}x + b(x,u), \tag{50}$$

where p is an integer, $A_{N,p}$ is the $(Np, p)-$ antishift matrix:

$$A_{N,p=}\begin{pmatrix} 0, Id_p, 0, ..., 0 \\ . \\ . \\ 0, ..., 0, Id_p \\ 0, ..., 0 \end{pmatrix},$$

and where $\frac{\partial b_i}{\partial x_j} \equiv 0$ if, for some integer k :

$$kp < i \leq (k+1)p, \text{ and } j > (k+1)p,$$

and all the functions $b_i(x,u)$ are compactly supported w.r.t. their x arguments.

Clearly, this form **includes the normal form (17)**, but also the systems with p outputs, that are in the **phase variable representation** (in this case, u **in (50) denotes not only the control, but its N first derivatives**).

Let us consider the same "extended Kalman filter" equations:

$$\begin{cases} (i)\ \frac{dz}{dt} = A_{N,p}z + b(z,u) - S(t)^{-1}C'r^{-1}(Cz - y(t)), \\ (ii)\frac{dS}{dt} = -(A_{N,p} + b^*(z,u))'S - S(A_{N,p} + b^*(z,u)) + \\ \qquad C'r^{-1}C - SQ_\theta S, \\ \eta = z, \end{cases} \tag{51}$$

where $C = (Id_p, 0, ..., 0)$, $Q_\theta = \theta^2\Delta^{-1}Q\Delta^{-1}$, $\Delta = BlockDiag(Id_p, \frac{1}{\theta}Id_p, ..., (\frac{1}{\theta})^{N-1}Id_p)$. $b^*(z,u)$ is again the Jacobian matrix of $b(z,u)$ w.r.t. z.

As in the case of our "Luenberger type" observers, we have:

THEOREM 2.8. *Theorem 2.6 holds also for systems of the form (50).*

COROLLARY 2.9. *If a system Σ has a phase variable representation of order N, then, for all open relatively compact subsets $\Omega \subset X$, for all $B > 0$, the system (51) is an exponential $U^{N,B}$ output observer for Σ, relative to Ω.*

Corollary 2.9 is just a restatement of Corollary 2.7 in this new context.

The following theorem is crucial for the proof of these results.

THEOREM 2.10. *If S_0 is positive definite, then, the solution $S(t)$ of the Riccati equation (51, (ii)) is well defined and positive definite for all $t \geq 0$. For all $T > 0$, there are constants $0 < \gamma < \delta$, depending on T, B, Q, r only (not on S_0 !) such that, for $t \geq T$:*

$$\gamma Id_{Np} \leq S(t) \leq \delta Id_{Np}.$$

This is classical, but one has to be very careful: all original versions of statements and proofs of this theorem are wrong. In particular, the following classical inequality is false:

$$(\frac{\alpha_2}{1 + \alpha_2\beta_1})Id_{Np} \leq P(t) \leq Id_{Np}(\frac{1}{\alpha_1} + \beta_2),$$

where α_1, β_1 are the bounds on the Gramm observability matrix, and α_2, β_2 are the bounds on the Gramm controllability matrix.

2.4.2. *The continuous-discrete version of the High-gain extended Kalman filter.* This is a more realistic version of the previous high-gain observer: Observations are sampled.

As the continuous high gain extended Kalman filter, it applies to systems that are in the normal form (50), in restriction to compact sets. In particular, it applies to all systems that have a phase variable representation for sufficiently smooth controls, and to control affine systems that have a uniform canonical flag, for general L^∞ controls.

For the statement of our result, let us make exactly the same assumptions as in the previous section 2.4.1. For simplicity in exposition, let us consider the **single output** case only.

Let us chose a time step δt, small enough. The equations of the continuous-discrete version $\Sigma_{Oc.d.}$ of our "extended Kalman filter" are, for $t \in [(k-1)\delta t, k\delta t[$:

(Prediction step:)

$$(52) \qquad \begin{aligned} (i)\ \frac{dz}{dt} &= Az + b(z,u), \\ (ii)\frac{dS}{dt} &= -(A + b^*(z,u))'S - S(A + b^*(z,u)) - SQ_\theta S, \end{aligned}$$

and at time $k\delta t$:

(innovation step:)

$$(53) \qquad \begin{aligned} (i)\ z_k(+) &= z_k(-) - S_k(+)^{-1}C'r^{-1}\delta t(Cz_k(-) - y_k), \\ (ii)\ S_k(+) &= S_k(-) + C'r^{-1}C\delta t. \end{aligned}$$

The assumptions being the same as for Theorem 2.6, Corollary 2.7, we have:

THEOREM 2.11. *For all $T > 0$, there are two positive constants θ_0, μ, such that, for all δt small enough, $\theta > \theta_0$, $\theta\ \delta t < \mu$, one has, for all $t > T$:*

$$||z(t) - x(t)|| \leq \bar{k}\theta^{n-1}e^{-(\lambda\theta-\omega)(t-\frac{T}{\theta})}||z(\frac{T}{\theta}) - x(\frac{T}{\theta})||,$$

for some positive constants $\bar{k}, \lambda, \omega$.

This is the continuous-discrete analog of Formula (49). Hence it is possible to state the continuous-discrete analogs of the other corollaries in the previous section. We leave this to the reader.

CHAPTER 6

Dynamic Output Stabilization.

Using the results of the previous chapters, we can derive a constructive method to solve the following problem. We are given a system Σ:

$$(\Sigma) \begin{cases} \dot{x} = f(x,u), \\ y = h(x,u), \end{cases}$$

$u \in U = I^{d_u}$, where I is a closed interval of R (possibly unbounded). This system is assumed to have an equilibrium point x_0 which is asymptotically stabilizable by smooth, $U-$valued state feedback, that is: there is a smooth function $\alpha(x)$, such that x_0 is an asymptotically stable equilibrium of the vector field $\dot{x} = f(x, \alpha(x))$.

The problem is the following: is it possible to stabilize asymptotically by using not the state information (as does the state feedback $\alpha(x)$), but by using only the output information. As usual in this type of problem, we avoid to differentiate the outputs (since, from the physical point of view, we would have to differentiate the noise, or the measurement errors, which is not reasonable).

We will be interested only by the behaviour of Σ within the basin of attraction of the equilibrium x_0, hence, we can restrict X to this basin of attraction and assume that $X = R^n$ (see [38]).

The basic idea, coming from the linear theory, is to construct a state observer, and to control the system using the feedback α evaluated on the estimate $\hat{x}$ of the state.

1) We will show that this is possible for the systems of Chapter 2, for which a uniform canonical flag exists, and we can construct an exponential state observer, by the previous chapter.

2) In the general case where we have a phase variable representation only, i.e. for systems of Chapters 3, 4, we will show that this is possible by using exponential output observers, but the construction is a bit more sophisticated.

In fact, we will not be able to cover (as in the linear case) the whole original basin of attraction: we will obtain asymptotic stability within **arbitrarily large compact sets** contained in this basin of attraction, only.

At the end of the chapter, we will say a few words about a situation in which the results can be improved from a practical point of view: our theorems depend very essentially on the high-gain construction. Moreover, the output stabilization is "twice high gain" (in a sense that will be clear later on: first, we need high gain for the observer to estimate exponentially, second, this exponential rate of the observer has to be very large). It is important to understand that, in some situations that are very common in practice, only exponential convergence of the estimation of the state is required, but the exponential rate can be small.

1. The case of a uniform canonical flag.

We will make the assumptions of Section 2.2 of the previous chapter 5: $X = R^n$, and our system is globally in the normal form (15). We know already that we can modify the normal form outside any open ball B^0 for the Lipschitz conditions $(A1)$, $(A2)$ of this chapter 5 to be satisfied globally over X. In the proof of the main theorem 1.1 below, the semi-trajectories of Σ under consideration will not leave a compact subset, denoted below by D_{m+1}. This justifies making these assumptions.

We will consider first the "Luenberger type" observer. The same task can be performed by the "Kalman type" observer, but it applies to the control affine case only, and the proof is more complicated, as we shall see.

1.1. Semi global asymptotic stabilizability. The most convenient notion to be handled in this chapter is not "global asymptotic stabilizability", but the weaker "**semi-global asymptotic stabilizability**", that we define immediately.

Notation: In order to shorten certain statements, let us say that a vector field on X is "**asymptotically stable at $x_0 \in X$ within a compact set $\Gamma \subset X$**" if x_0 is an asymptotically stable equilibrium point and the basin of attraction of x_0 contains Γ.

DEFINITION 1.1. *We say that the (unobserved) system Σ on X, $\dot{x} = f(x,u)$ is semi-globally asymptotically stabilizable at (x_0, u_0) if, for each compact $\Gamma \subset X$, there is a smooth feedback $\alpha_\Gamma : X \to Int(U)$, $\alpha_\Gamma(x_0) = u_0$, such that the vector field on X:*

$$\dot{x} = f(x, \alpha_\Gamma(x)), \tag{54}$$

is asymptotically stable at x_0 within Γ.

We make the assumption that $X = R^n$, but it may not be clear that the state space X of a semi-globally asymptotically stabilizable Σ should be R^n. But in fact, this is true: by definition, any compact subset $\Gamma \subset X$ is contained in the basin of attraction of x_0 for a certain smooth vector field. By the results of [**38**], such a basin of attraction is diffeomorphic to R^n.

Then, since we assumed X paracompact, the Brown-Stallings theorem gives the result.

Comment: It does not follow immediately from [**38**] that the basin of attraction of the origin, for an asymptotically stable vector field, is diffeomorphic to R^n, since it is assumed in the paper that the state space is R^n. But, one can modify slightly the arguments to make them work for a general (paracompact) manifold.

1.2. Stabilization with the Luenberger-type observer. We assume (reparametrization) that $x = 0$, $u = 0$ is the equilibrium point under consideration, and the smooth stabilizing feedbacks are $\alpha_\Gamma(x)$, i.e. $\alpha_\Gamma(0) = 0$, and $x = 0$ is an asymptotically stable (within Γ) equilibrium point of the vector field:

$$\dot{x} = f(x, \alpha_\Gamma(x)). \tag{55}$$

Now Γ is fixed, together with the corresponding α_Γ. The basin of attraction of zero is B_Γ, $\Gamma \subset B_\Gamma$.

We have to study the following system on $X \times X$:

$$\left\{ \begin{array}{c} \dot{x} = f(x, \tilde{\alpha}(z)), \\ \dot{z} = f(z, \tilde{\alpha}(z)) - K_\theta(h(z, \tilde{\alpha}(z)) - y), \\ y = h(x, \tilde{\alpha}(z)), \end{array} \right. \tag{56}$$

where K_θ is as in Theorem 2.2 of Chapter 5. The purpose is to show that, if θ is large enough, $(0,0)$ is an asymptotically stable equilibrium of (56), and the basin of attraction of $(0,0)$ can be made arbitrarily large in $B_\Gamma \times X$, by increasing θ. Here, $\tilde{\alpha}$ is a certain other smooth feedback, depending on the compact Γ, that we construct now:

Using the inverse Lyapunov theorems, we can find a smooth, proper strict Lyapunov function $V : B_\Gamma \to R_+$, for the vector field (55), $V(0) = 0$. This means that, along the trajectories of Σ, $\frac{dV(x)}{dt} < 0$ (except for $x = 0$). The function V reaches its maximum m over Γ. Let us consider $D_{m+1} = \{x | V(x) \leq m + 1\}$, and let us replace α_Γ by $\tilde{\alpha}$ such that:

1) $\tilde{\alpha} = \alpha_\Gamma$ on D_{m+1},

2) $\tilde{\alpha}$ is smooth, compactly supported, with values in $Int(U)$ (we have already assumed above that, by translation in the U space, $0 \in Int(U)$).

Let us also set $B = \sup_{x \in R^n} ||\tilde{\alpha}(x)||$. The following theorem holds:

THEOREM 1.1. *Given arbitrary compact sets $\Gamma, \Gamma' \subset X$, if θ is large enough, then, (56) is asymptotically stable at the origin within $\Gamma \times \Gamma'$.*

Comment: This theorem means that, provided that we know a compact set Γ where the system starts, and provided that we modify the stabilizing feedback at infinity, we can just plug the estimate of the state given by our "state observer" into the feedback, and the resulting system is asymptotically stable at the origin. Moreover, the semi trajectories starting from $\Gamma \times \Gamma'$ tend to the origin. That is, **we can asymptotically stabilize at the origin via the observer, using observations only.** This can be done **within arbitrarily large compact sets.**

As it appears in the proof, it is very important that the observer is exponential, with arbitrary exponential decay: we need the exponential decay for local asymptotic stabilization, but we also need an arbitrarily large exponential in order to estimate the state very quickly.

1.3. Stabilization with the high-gain E.K.F. Assume that we are in the control affine case. Then, we could try to use the "high gain extended Kalman filter" exactly in the same way (see Theorem 2.6, Corollary 2.7, Chapter 5). **There are several additional difficulties, but the same result holds:**

THEOREM 1.2. *Replacing the "Luenberger High-gain observer" by the "High-gain extended Kalman filter", Theorem 1.1 is still valid, i.e. : for any triple of compact subsets $\Gamma, \Gamma', \Gamma'', \Gamma \times \Gamma' \times \Gamma'' \subset R^n \times R^n \times S_n(+)$, for θ large enough, the "High gain extended Kalman filter" coupled with the system Σ to which the feedback control $\tilde{\alpha}(z)$ is applied , is asymptotically stable within $\Gamma \times \Gamma' \times \Gamma''$ at $(0, 0, S_\infty)$.*

EXERCISE 1.1. *Give the proof of Theorem 1.2.*

We can also use the continuous-discrete version of the high gain extended Kalman filter:

EXERCISE 1.2. *State and prove a version of Theorem 1.2 using the continuous-discrete version of the high-gain EKF.*

2. The general case of a phase variable representation.

2.1. Preliminaries. We will now deal with the case of Chapters 3, 4 , where the system has a phase variable representation of a certain order. This happens generically if $d_y > d_u$ (at least for bounded and sufficiently differentiable controls, but it will be the case here).

The systems of the previous paragraph have a phase variable representation, as we have noticed already. So that, one could ask: why the considerations in the previous paragraph? The answer is that the procedure for stabilizing (asymptotically) via output information **is much less complicated in that case.** In particular, now, we will have to deal with a certain number of successive derivatives of the inputs, because we have only $U^{N,B}$ output observers. In the previous paragraph, we had a $U^{0,B}$ state observer, hence, these problems did not appear.

We will obtain the result with the high-gain Luenberger output observer. Generalizations to the case of the high-gain extended Kalman filters (either continuous-continuous or continuous-discrete) can be done in the same way as in the proof of Theorem 1.2 above. We will leave these generalizations as exercises.

2.1.1. *Rings of C^∞ functions.* Recall that, in Chapter 4, we introduced several rings of (germs of) analytic functions: $\Re_N, \hat{\Re}_N, \bar{\Re}_N, \tilde{\Re}$. Recall that $\hat{\Re}_N$ was just the pull back ring: $\hat{\Re}_N = (S\Phi_N^\Sigma)^*(O_{z_0})$, where $z_0 = S\Phi_N^\Sigma(x_0, u_0^{(0)}, \widetilde{u_0}_{N-1})$. We will consider the C^∞ analogs $\tilde{\Re}_N, \bar{\Re}_N, \tilde{\Re}$ of the rings $\hat{\Re}_N, \Re_N, \tilde{\Re}$. i.e.

$$\tilde{\Re}_N(x_0, u^{(0)}, \tilde{u}_{N-1}) = \{G \circ S\Phi_N^\Sigma\},$$

where G varies over the germs of C^∞ functions at the point $S\Phi_N^\Sigma(x_0, u^{(0)}, \tilde{u}_{N-1})$, and

$$\bar{\Re}_N(x_0, u^{(0)}, \tilde{u}_{N-1}) = \{G \circ S\Phi_{N,\tilde{u}_{N-1}}^\Sigma\},$$

where G varies over the germs of C^∞ functions at the point $S\Phi_{N,\tilde{u}_{N-1}}^\Sigma(x_0, u^{(0)})$.

$\tilde{\Re}(x_0, u^{(0)})$, or simply $\tilde{\Re}$, if there is no ambiguity, will be the ring of germs of functions of the form

$$G(u, \varphi_1, ..., \varphi_p)$$

at the point $(x_0, u^{(0)})$, where G is C^∞ and all the functions φ_i are of the form:

$$\varphi_i = L_{f_u}^{k_1}(\partial_{j_1})^{s_1} L_{f_u}^{k_2}(\partial_{j_2})^{s_2} ..... L_{f_u}^{k_r}(\partial_{j_r})^{s_r} h.$$

(Again, $\partial_j = (\frac{\partial}{\partial u_j})$).

Recall that, in the analytic case, the condition $ACP(N)$ is equivalent to $\tilde{\Re} = \Re_N \subset \hat{\Re}_N$, by Theorem 4.1, Chapter 4. Of course, if $ACP(N)$ holds, then, a fortiori,

$$\tilde{\Re} = \bar{\Re}_N \subset \tilde{\Re}_N : \tag{57}$$

the above generators u, φ_i of $\tilde{\Re}$ belong to $\Re_N \subset \hat{\Re}_N$, hence C^∞ functions of them belong to $\bar{\Re}_N$. Therefore $\tilde{\Re} \subset \bar{\Re}_N$. Using the same reasoning, $\tilde{\Re} \subset \bar{\Re}_N \subset \tilde{\Re}_N$.

Also, by definition, $\bar{\Re}_N \subset \tilde{\Re}$.

Moreover, the analyticity assumption plays no role in this result, so that, it is also true for a C^∞ system Σ that $ACP(N)$ is equivalent to the condition (57).

2.1.2. *Assumptions.* We will start with the most general situation, that is, we assume that our system Σ satisfies the assumptions of Theorem 5.2, Chapter 4, i.e.:

(H_1)-Σ satisfies $ACP(N)$ at each point,

(H_2)-Σ is differentially observable of order N.

In particular, if Σ is strongly differentially observable of order N (the "generic" situation of Chapter 3), these assumptions are satisfied.

For the same reasons as above, we assume also that $X = R^n$.

In this chapter, Σ is semi globally asymptotically stabilizable at ($x_0 = 0$, $u_0 = 0$). We will have to make an additional assumption (H_3), relative to the stabilizing feedbacks α_Γ :

(H_3)- The germs of the $\alpha_{\Gamma,j}(.)$ at x_1, $j = 1, ..., d_u$, belong to $\bar{\Re}_N(x_1, u^{(0)}, \tilde{u}_{N-1})$, for all $x_1 \in X = R^n$, for all $(u^{(0)}, \tilde{u}_{N-1}) \in U \times R^{(N-1)d_u}$. Equivalently, these germs belong to $\tilde{\Re}$, by virtue of (57).

2.1.3. *Comments.* 1) This assumption (H_3), (together with (H_1), (H_2)), **is automatically satisfied if Σ is strongly differentially observable of order N** : in that case, the ring $\bar{\Re}_N$ is just the ring of germs of smooth functions at $(x_1, u^{(0)})$.

EXERCISE 2.1. *Prove this last statement.*

2) We have defined the C^∞ analogs $\tilde{\Re}_N, \bar{\Re}_N, \tilde{\Re}$ of our rings $\hat{\Re}_N, \Re_N, \check{\Re}$ for the following reasons: first, in this section, we want to deal with C^∞ systems (recall that the results of Chapter 4 are valid also in the C^∞ case). Second, it could happen that, **even if Σ is analytic**, and asymptotically stabilizable, then it is asymptotically stabilizable by a feedback which is only C^∞.

3) At this point, it is important to say a few words about U. In the previous section, we assumed that $U = I^{d_u}$, where I is a closed interval of R. Assume that I is not equal to R. Then, we can find a diffeomorphism $\Psi : Int(I) \to R$. The rings above are intrinsic objects, that do not depend on coordinates on U. So that $ACP(N)$ does not depend on a change of variable over u. On the same way, the assumption (H_2) is intrinsic. Also, the fact that the germ of $\alpha_{\Gamma,j}$ belongs to $\bar{\Re}_N$ at each point is intrinsic.

Therefore, since our stabilizing feedbacks α_Γ take their values in the interior of U, we see that **we can replace u by v, $v_i = \Psi(u_i)$ and assume that $I = R$.** This is what we will do in the remainder of this section.

2.1.4. *A crucial lemma.* In order to state our main theorem in this section, we need a preliminary result:

LEMMA 2.1. *Assume that Σ is given, and that (H_1), (H_3) are satisfied. Then, (H_1), (H_3) are also satisfied for Σ^r, the r^{th} dynamical extension of Σ.*

The definition of the r^{th} dynamical extension of Σ is given in Chapter 1, Definition 5.1.

COROLLARY 2.2. *If Σ satisfies $(H_1), (H3)$, then Σ^N is stabilizable within any arbitrary compact set Γ with a feedback α_Γ that belongs to $\bar{\Re}_N(\Sigma)$ (the germs of which belong to $\tilde{\Re}_N(\Sigma)$).*

2.2. Output stabilization again. First, we will consider the Luenberger version of the output observer.

We set $M = N(d_y+d_u)$, where N is such that Σ satisfies the condition $ACP(N)$ and is differentially observable of order N. $R^M = R^{N(d_y+d_u)}$.

We chose $\Gamma \subset X$, $\Gamma' \subset R^{Nd_y}$, $\Gamma'' \subset R^{Nd_u}$, three arbitrary compact sets. We denote, (as in the previous chapters) by $A_{m,p}$ the (mp,p) block-antishift matrix, i.e. $A_{m,p} : R^{mp} \to R^{mp}$,

$$A_{m,p} = \begin{pmatrix} 0, Id_p, 0, \ldots\ldots\ldots, 0 \\ \cdot \\ \cdot \\ 0, \ldots\ldots\ldots\ldots, 0, Id_p \\ 0, \ldots\ldots\ldots\ldots\ldots\ldots, 0 \end{pmatrix}.$$

Also, $C_{m,p} : R^{mp} \to R^p$ and $b_{m,p} : R^p \to R^{mp}$ denote the matrices:

$$C_{m,p} = (Id_p, 0, \ldots\ldots, 0),$$

$$b_{m,p} = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ Id_p \end{pmatrix}.$$

We take a feedback $\alpha^N_{\Gamma\times\Gamma''}$, given by Corollary 2.2, that stabilizes the N^{th} dynamical extension Σ^N of Σ within $\Gamma \times \Gamma''$. Recall that Σ^N is given by:

$$\begin{cases} \dot{x} = f(x,u), \\ \dot{\omega} = A_{N,d_u}\omega + b_{N,d_u}u_N, \end{cases} \tag{58}$$

with the notation $\omega = (u^{(0)}, \tilde{u}_{N-1})$.

Then, the feedback system:

$$\begin{cases} \dot{x} = f(x,u^{(0)}), \\ \dot{\omega} = A_{N,d_u}\omega + b_{N,d_u}\alpha^N_{\Gamma\times\Gamma''}(x,\omega), \end{cases} \tag{59}$$

is asymptotically stable at $(x_0,0)$ within $\Gamma\times\Gamma''$. Let $\tilde{B}$ denote the basin of attraction of $(x_0,0)$ for (59).

Let V be a proper Lyapunov function for this vector field on $\tilde{B}$. The function V has a maximum m over $\Gamma \times \Gamma''$. Setting $D_k = \{s|V(s) \leq k\}$, $k \geq 0$, let us consider D_m, D_{m+1}. Then, $\Gamma \times \Gamma'' \subset D_m \subset Int(D_{m+1})$.

Using the C^∞ version of Corollary 5.3, Chapter 4, we can find a C^∞ function α defined on all of R^M such that:

$$\alpha^N_{\Gamma\times\Gamma''}(x,u^{(0)},\tilde{u}_{N-1}) = \alpha(S\Phi^\Sigma_N(x,u^{(0)},\tilde{u}_{N-1})), \tag{60}$$

for all $(x,u^{(0)},\tilde{u}_{N-1}) \in D_{m+1}$, and moreover, α can be taken compactly supported.

Hence, α reaches its maximum over R^M. So do $u^{(0)}$, $u^{(i)}$, $1 \leq i \leq N-1$ over D_{m+1}. Let B be the maximum of these maxima. Let $\tilde{\Gamma}$ be the image of D_{m+1} by the projection $\Pi_1 : X \times R^{Nd_u} \to X$. The set $\tilde{\Gamma}$ is compact.

We consider the $\varphi^{\tilde{\Gamma}}_N$ given by Theorem 5.2, Chapter 4, applied to Σ and $\tilde{\Gamma}$, i.e.:

$$y^{(N)} = \varphi^{\tilde{\Gamma}}_N(y^{(0)},\tilde{y}_{N-1},u^{(0)},\tilde{u}_N),$$

for all $x \in \tilde{\Gamma}$, all $u^{(0)}$, $\tilde{u}_N$.

We can "now couple" the feedback $\alpha^N_{\Gamma\times\Gamma''}$ with the $U^{N,B}$ output observer in its Luenberger form (Formula (45), Section 2.3, Chapter 5).

Let us write the equation of the full "coupled" system over $X \times R^M$:

$$
\begin{aligned}
(61)\quad (i)\ \dot{x} &= f(x, u^{(0)}),\\
(ii)\ \dot{\omega} &= A_{N,d_u}\omega + b_{N,d_u}\alpha(z,\omega),\\
(iii)\ \dot{z} &= (A_{N,d_y} - K_\theta C_{N,d_y})z + K_\theta h(x,u^{(0)})^{\cdot} + b_{N,d_y}\varphi_N^{\bar{\Gamma}}(z,\omega,\alpha(z,\omega)).
\end{aligned}
$$

Our result will be the following, as expected:

THEOREM 2.3. *Assumptions (H_1), (H_2), (H_3) are made. For any $\Gamma' \subset R^{Nd_y}$, for θ large enough, the system (61) is asymptotically stable within $\Gamma \times \Gamma'' \times \Gamma'$ at $(x_0, 0, 0)$.*

This theorem means that, in all the cases we have dealt with in the previous chapters, we can stabilize asymptotically, using output informations only, within arbitrarily large compact sets, as soon as we can stabilize asymptotically within compact sets by smooth state feedback.

In particular, if the system Σ is strongly differentially observable of some order N, which is generic if $d_y > d_u$, and if Σ is smooth state feedback stabilizable, this theorem applies.

EXERCISE 2.2. *Consider the system Σ of Exercise 4.1, Chapter 4:*
$X = R^2$, $U = R$, $y = h(x) = x_1$,

$$
\begin{aligned}
\dot{x}_1 &= x_2^3 - x_1,\\
\dot{x}_2 &= x_2^8 + x_2^4 u.
\end{aligned}
$$

1. Show that the feedback $u_r(x) = -r\ (x_2)^3$, $r > 0$, stabilizes asymptotically Σ at $(0,0)$. Show that the basin of attraction is $b_r = \{x |\ x_2 < r\}$.

2. Show that the previous theorem applies, but Σ is not strongly differentially observable, of any order.

THEOREM 2.4. *The statement of Theorem 2.3 also holds when we replace the Luenberger version of the output observer by its extended Kalman filter version (Corollary 2.9, Chapter 5).*

EXERCISE 2.3. *Give a proof of Theorem 2.4.*

EXERCISE 2.4. *State and prove a version of Theorem 2.4, using the high-gain extended Kalman filter in its continuous-discrete form (Theorem 2.11, Chapter 5).*

3. Complements:

3.1. Systems with positively invariant compact state spaces. This is a situation which seems to be very common in practice. In particular, it will appear in the first application of the next chapter. We make the additional assumption:

(H_4) : (i) The system Σ is such that the state space is not $X = R^n$, but a certain **relatively compact** open subset $\Omega \subset R^n$. We assume that Σ is also defined and smooth on the boundary $\partial\Omega$, and the closure $Cl(\Omega)$ is positively invariant for the dynamics of Σ whatever the control $u(.)$, with values in U,

(ii) the state feedback is asymptotically stabilizing within $Cl(\Omega)$.

Observation: $(H_4, (i))$ implies that Ω also is positively invariant for the dynamics of Σ.

PROPOSITION 3.1. *In the case where the assumption* (H_4) *holds, all the theorems in the two previous sections in this chapter are true with a refinement: as soon as the observers are exponential , **with any rate of decay of the error**, the full system controller+observer is asymptotically stable within* $Cl(\Omega) \times \hat{X}$, *where* $\hat{X}$ *is the state space of the observer.*

EXERCISE 3.1. *Give details of the previous proof, specially in the case of the high gain extended Kalman filter.*

CHAPTER 7

Something Between the HGEKF and the EKF.

1. Introduction, systems under consideration

1.1. Systems under consideration. We consider nonlinear systems of the following form (62), on $\mathbb{R}^n$. The control space U, is a closed subset of $\mathbb{R}^d$. **Only for simplicity of the exposition of the proof of the main result**, the observation is taken to be single-valued: it is a $u-$ dependant linear form on $\mathbb{R}^n$.

$$\frac{dx}{d\tau} = A(u)x + b(x,u), \qquad y = C(u)x. \tag{62}$$

$A(u)$, $C(u)$ are matrices:

$$C(u) = (a_1(u), 0, ..., 0),$$

$$A(u) \begin{pmatrix} 0, a_2(u), 0, ..., 0 \\ 0, 0, a_3(u), 0, ..., 0 \\ \cdot \\ \cdot \\ 0, ..., 0, a_n(u) \\ 0, ..., 0 \end{pmatrix}.$$

where $a_i(.)$, $i = 1, ..., n$, are positive smooth functions, bounded from above and from below:

$$0 < a_m \leq a_i(u) \leq a_M.$$

Also, $b(x,u)$is a smooth, $u-$dependant vector field, depending triangularly on x and compactly supported:

$$b = b_1(x_1,u)\frac{\partial}{\partial x_1} + b_2(x_1,x_2,u)\frac{\partial}{\partial x_2} + ... + b_n(x_1,...,x_n,u)\frac{\partial}{\partial x_n}.$$

This assumption corresponds in fact to a special case of an (observable and) uniformly infinitesimally observable system. For instance, observable control affine systems of the form (17), or systems with a phase variable representation (46), are special subcases of this normal form. The assumption $0 < a_m \leq a_i(u) \leq a_M$ is the analog of assumption (42) in the general "uniformly infinitesimally observable" case.

The reason for this form in this chapter is the practical application we present below (a distillation column, Section 3, see also the notes 1 and 2 below). It is a multi output generalisation of this normal form (62). All the results we prove here generalize easily to this multi-output case. We leave this generalization to the reader.

Also, we leave to the reader the generalization to the general case of systems of the form (50), which is also straightforward: multi-output systems with a phase variable representation.

With the same justification as in Chapter 5, the compact support assumption for $b(x,u)$ can be made, eventually modifying b outside an arbitrarily large compact set.

We stress again that here, **the single output assumption can be removed everywhere.**

1.2. Presentation of the results. Our purpose herein is **to construct observers**, for the observable systems (62) described above.

In fact, for these systems, several types of nonlinear observers can be constructed. We will focus on two types of construction that both turn around the "extended Kalman filter", in either its deterministic or its stochastic form:

1. **First construction:** The Extended Kalman Filter itself,
2. **Second construction:** The High Gain Extended Kalman Filter,
3. **Our construction in this chapter:** a mixing of 1. and 2.

Let us just give some details now, to explain where we want to go.

1. The extended Kalman Filter.

It is known that, **under observability conditions**, the Extended Kalman filter, has good ("local") properties:

(i) In its deterministic form, it is a local observer in the following sense. For sufficiently small initial error on the estimate of the state, the estimation error converges exponentially to zero. The prototype of these results can be found in [**3**] for instance.

For our systems (62), with the assumptions of Section 1, it is not hard to check that they are uniformly infinitesimally observable, and hence, the linearized systems along any trajectory are uniformly observable, (in the classical sense of the linear theory, and with uniform bounds on the Gramm observability matrices). Therefore, this result applies.

(ii) In its stochastic form, except for the linear case, where the EKF is the "optimal" filter, there is no general theoretical result that applies. Even for good observable systems in our normal form (62), for small noise, small initial variance and dimension 1: there is a counterexample of such a system, in [**30**] for instance, where the EKF doesn't work at all.

Nevertheless, despite the lack of these theoretical justifications, people use it in practice for nonlinear filtering and it may give very good results (even for systems that have much weaker observability properties than those considered here).

In the application of our techniques, presented in section 3 below, we will show a (family of) practical examples which is very interesting because, it seems that, the results of [**30**] on the EKF for small noise, apply in general, and that the "small parameter" has a physical interpretation.

We will not say more about that because this is beyond the scope of this course. But it is one more justification of the use of our method developed here to this application.

2. The High Gain Extended Kalman Filter.

As it was explained in Chapter 5, we consider the equations of the extended Kalman filter, in which the "covariance matrix Q" depends on a real parameter θ, $\theta \geq 1$, in the following way:

$$Q_{ij} = \theta^{i+j+1} Q^0_{i,j}.$$

For $\theta = 1$, it is exactly the EKF. For θ large enough, it is what we called here the "High gain extended Kalman filter".

(i) In the deterministic setting, as we have shown, the estimation error has **arbitrarily large exponential decay** (depending on θ). This holds **whatever the initial error is, (that is, this is a global result).**

(ii) In the stochastic setting, it is a nonlinear filter with "bounded variance" (the variance is bounded in θ^n, which is not that good, but it is bounded anyway). ([8], for instance).

3. What we want to do in this chapter.

The idea in this paper is the **very simple** following one: we give the parameter θ in the HGEKF an exponential decay from θ_0 large, to 1.

What is expected, (and what happens) is the following:

(i) The beginning of the transient of the estimation error is the one of the high gain extended Kalman observer: there is an exponential decay that can be made arbitrarily large.

(ii) There is a global exponential decay of the estimation error (but, of course, it cannot be controlled).

(iii) The asymptotic behavior is the one of the standard "extended Kalman filter", (that people like in practice, as stated above).

Our main result, Theorem 2.1 in Section 2 proves (i) and (ii). The proof is more or less an improvement of the proof of convergence of the high gain Kalman observer, chapter 5. (In particular, it contains the proof of the results of chapter 5).

Of course, this construction has a terminal defect: it is time dependant. In deterministic terms, it will work for large initial estimation errors, but not for big "jumps" of the state at intermediate times. In the section 2.3, we propose a very simple practical way to make the observer "recursive".

In the section 3, we show the application of this procedure to a binary distillation column in which the "quality of the feed" is unknown, an subject to large changes. It was already noticed in the book [**15**] that this application is a nontrivial nice application of the observability theory, and of high gain observers.

Here, it is even much more convincing: when the feed changes, (a big "state jump"), the behavior of the observer is the one of a high-gain observer: recovering arbitrarily fast the quality of the feed, and when the feed does not move, the asymptotic behavior of the observer is the one of the extended Kalman filter, almost optimal with respect to small noise in that case (but we do not prove anything about this optimality in this paper).

For first applications of "high gain observers" to distillation columns, see [**36**], [**37**].

Note 1. The reasons for which we make the matrix $A(u)$ depend on u in the normal form (62) may look not clear, because, in all the cases described above, it doesn't.

In fact, the only reason to consider this dependance is the following: the formal computations we do in the proof of our main result, work for that type of systems. Moreover, in the application we describe in Section 3, the matrix A actually does depend on u.

Note 2. In that case were a_i depends on u, the following should also be noticed: even the high gain version of the extended Kalman filter is much better in practice than the "high gain Luenberger observer" mentioned above: the high gain observers both kill the nonlinearities contained in the vector field b. But the extended Kalman filter takes into account the variations of u, through the matrix $A(u)$. The standard high gain observers in Luenberger form don't do this. This is the case in the application, Section 3 below.

2. Statement and proof of the theoretical result

The observer we propose, is based upon the High gain extended Kalman filter of Chapter 5.

2.1. The observer and the statement of the theorem. The equation of the observer is:

$$\left\{\begin{array}{c} (i)\ \frac{dz}{d\tau} = A(u)z + b(z,u) - S(t)^{-1}C'r^{-1}(Cz - y(t)), \\ (ii)\frac{dS}{d\tau} = -(A(u) + b^*(z,u))'S - S(A(u) + b^*(z,u)) + \\ C'r^{-1}C - SQ_\theta S, \\ \frac{d\theta}{d\tau} = \lambda(1-\theta), \end{array}\right. \tag{63}$$

where $C = (a_1(u), 0, ..., 0)$, $Q_\theta = \theta^2\Delta^{-1}Q\Delta^{-1}$, $\Delta = diag(1, \frac{1}{\theta}, ..., (\frac{1}{\theta})^{n-1})$. Here, $b^*(z,u)$ denotes the Jacobian matrix of $b(z,u)$ w.r.t. z, and r, λ are positive scalars. Q is a symmetric positive definite matrix.

Comments:

1. Q, r, in the stochastic context, are the covariances of the state noise and output noise respectively.

2. If $\lambda = 0$ and $\theta_0 = 1$, or if $\lambda > 0$, but t is large, this is exactly the (deterministic version of) the extended Kalman filter.

3. If θ_0 is large, and if $\tau \leq T$, then, this equation is almost the equation of the high gain extended Kalman filter with gain $\theta(T)$. Hence, for $\tau \leq T$, setting $\varepsilon(\tau) = z(\tau) - x(\tau)$, ($\varepsilon$ is the **estimation error**), we can expect the following, for θ_0 large enough in front of T:

$$||\varepsilon(\tau)||^2 \leq \theta(\tau)^{2(n-1)}H(c)e^{-(a_1\theta(T)-a_2)\tau}||\varepsilon(0)||^2. \tag{64}$$

Here, a_1, a_2 are positive constants, $H(c)$ is a decreasing positive function of c, where $S(0) \geq c\ Id$. Also, $\theta(T) = 1 + (\theta_0 - 1)e^{-\lambda T}$.

In particular, this implies that the error $\varepsilon(t)$ **can be made arbitrarily small, in arbitrarily short time, increasing** θ_0. For θ constant, this is the behavior of the "high gain extended Kalman filter. We will prove it below for θ nonconstant.

Our main result herein will be the following:

THEOREM 2.1. *1. For all* $0 \leq \lambda \leq \lambda_0$, ($\lambda_0 = \frac{Q_m\alpha}{4(n-2)}$, *where* $Q \geq Q_m Id$ *and* α *comes from Lemma 2.2 below), for all* θ_0 *large enough, depending on* λ, *for all*

$S_0 \geq c\ Id$, for all $K \subset \mathbb{R}^n$, K a compact subset, for all $\varepsilon_0 = z_0 - x_0$, $\varepsilon_0 \in K$, the following estimation holds, for all $\tau \geq 0$:

$$||\varepsilon(\tau)||^2 \leq R(\lambda, c)e^{-a\ \tau}||\varepsilon_0||^2\Lambda(\theta_0, \tau, \lambda), \tag{65}$$

$$\Lambda(\theta_0, \tau, \lambda), = {\theta_0}^{2(n-1)+\frac{a}{\lambda}} e^{-\frac{a}{\lambda}\theta_0(1-e^{-\lambda\tau})},$$

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

2. Moreover the short term estimate (64) holds for all $T > 0$, $\tau \leq T$, for all $\theta_0 \geq \bar{\theta}_0$, $\bar{\theta}_0 = e^{\lambda T}(\frac{L'}{Q_m \alpha} - 1) + 1$, where L' is the sup of the partial derivatives of b w.r.t. x.

Comments.

a. Note that the function $\Lambda(\theta_0, \tau, \lambda)$ is a decreasing function of τ, and that, for all $\tau > 0$, $\lambda > 0$, $\Lambda(\theta_0, \tau, \lambda)$ can be made arbitrarily small, increasing θ_0.

b. This means that, provided that λ is smaller than a certain constant λ_0, and θ_0 is large in front of λ, 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".

2.2. Proof of Theorem 2.1.

2.2.1. *Preparation for the proof.* Let us recall that:

$$\theta(\tau) = 1 + (\theta_0 - 1)e^{-\lambda\tau}, \tag{66}$$

and let us set $F = diag(0, 1, 2, ..., n-1)$. Then:

$$\begin{aligned} \frac{d(\frac{1}{\theta})}{d\tau} &= -\frac{\lambda(1-\theta)}{\theta^2}, \\ \frac{d\Delta}{d\tau} &= -F\Delta\frac{\lambda(1-\theta)}{\theta}, \\ \frac{d\Delta^{-1}}{d\tau} &= F\Delta^{-1}\frac{\lambda(1-\theta)}{\theta}. \end{aligned} \tag{67}$$

The equations under consideration are:

$$\begin{aligned} &(i)\ \tfrac{d\varepsilon}{d\tau} = A(u)\varepsilon + b(z,u) - b(x,u) - S(t)^{-1}C'r^{-1}C\varepsilon, \\ &(ii)\tfrac{dS}{d\tau} = -(A(u) + b^*(z,u))'S - S(A(u) + b^*(z,u)) + C'r^{-1}C - SQ_\theta S, \\ &(iii)\ \tfrac{d\theta}{d\tau} = \lambda(1-\theta). \end{aligned} \tag{68}$$

We make the following changes of variables, with $P = S^{-1}$:

$$\tilde{x} = \Delta x, \tilde{z} = \Delta z, \varepsilon = z - x, \tilde{\varepsilon} = \Delta\varepsilon, \tilde{S} = \theta\Delta^{-1}S\Delta^{-1}, \tag{69}$$

$$\tilde{P} = \tilde{S}^{-1} = \frac{1}{\theta}\Delta P\Delta, \tilde{b}(z) = \Delta b(\Delta^{-1}z), \tilde{b}^*(z) = \Delta b^*(\Delta^{-1}z)\Delta^{-1}.$$

Remark : It should be noted that the Lipschitz constant of $\tilde{b}$ is the same as the one of b, and the maximum of $||\tilde{b}^*||$ is the same as the one of $||b^*||$ (recall that the component b_i of b is compactly supported with respect to all of its arguments $(x_1, ..., x_i, u)$, and that $\theta \geq 1$).

An obvious computation gives:

$$\frac{d}{d\tau}(\tilde{\varepsilon}) = \theta[(A - \tilde{P}C'r^{-1}C)\tilde{\varepsilon} + \frac{1}{\theta}(\tilde{b}(\tilde{z}) - \tilde{b}(\tilde{x})) - \frac{\lambda(1-\theta)}{\theta^2}F\tilde{\varepsilon}], \tag{70}$$

$$\begin{aligned}\frac{d}{d\tau}(\tilde{S}) = \theta[&-(A + \frac{1}{\theta}\tilde{b}^*(\tilde{z}) - (\frac{Id}{2} + F)\frac{\lambda(1-\theta)}{\theta^2})'\tilde{S} \\ &- \tilde{S}(A + \frac{1}{\theta}\tilde{b}^*(\tilde{z}) - (\frac{Id}{2} + F)\frac{\lambda(1-\theta)}{\theta^2}) + C'r^{-1}C - \tilde{S}Q\tilde{S}],\end{aligned} \tag{71}$$

$$\frac{d\theta}{d\tau} = \lambda(1-\theta).$$

Important comment. At this place, we used the observability properties: the normal form (62) is crucial in the computation above.

Now, we can make a time rescaling. We set:

$$dt = \theta(\tau)d\tau, \text{ or } t = \int_0^\tau \theta(v)dv,$$

$$\tilde{\varepsilon}(\tau) = \bar{\varepsilon}(t), \tilde{S}(\tau) = \bar{S}(t), \tilde{P}(\tau) = \bar{P}(t), \theta(\tau) = \bar{\theta}(t),$$

to get the final set of equations:

$$(i)\ \frac{d}{dt}(\bar{\varepsilon}) = [(A - \bar{P}C'r^{-1}C)\bar{\varepsilon} + \frac{1}{\bar{\theta}}(\tilde{b}(\bar{z}) - \tilde{b}(\bar{x})) - \frac{\lambda(1-\bar{\theta})}{\bar{\theta}^2}F\bar{\varepsilon}], \tag{72}$$

$$\begin{aligned}(ii)\ \frac{d}{dt}(\bar{S}) = [&-(A + \frac{1}{\bar{\theta}}\tilde{b}^*(\bar{z}) - (\frac{Id}{2} + F)\frac{\lambda(1-\bar{\theta})}{\bar{\theta}^2})'\bar{S} \\ &- \bar{S}(A + \frac{1}{\bar{\theta}}\tilde{b}^*(\bar{z}) - (\frac{Id}{2} + F)\frac{\lambda(1-\bar{\theta})}{\bar{\theta}^2}) + C'r^{-1}C - \bar{S}Q\bar{S}],\end{aligned}$$

$$(iii)\ \frac{d\bar{\theta}}{dt} = \lambda\frac{(1-\bar{\theta})}{\bar{\theta}}.$$

First, there are some classical results allowing to bound the solutions of the Ricatti equation (72), (ii), for $\theta_0 > 1$, and $\lambda < 1$. To apply these results, one has to notice that the linear time dependant systems:

$$\begin{aligned}\frac{dx}{dt} &= (A(u(t)) + \frac{1}{\bar{\theta}}\tilde{b}^*(\bar{z}) - (\frac{Id}{2} + F)\frac{\lambda(1-\bar{\theta})}{\bar{\theta}^2})x(t), \\ y &= C(u(t))x(t),\end{aligned}$$

are uniformly observable (in the sense of linear systems), for all bounded measurable functions $a_i(u(t))$, $\tilde{b}^*_{i,j}(\bar{z}(t)), \bar{\theta}(t)$, with $a_M \geq a_i \geq a_m > 0$. Precisely, we have:

LEMMA 2.2. *If the functions $a_i(u(t))$, $|\tilde{b}^*_{i,j}(\bar{z}(t))|$, $\bar{\theta}(t)$, are all smaller than $a_M > 0$, and if $a_i(u(t)) > a_m > 0$, (which is the case by our assumptions), if $0 \leq \lambda \leq 1$, and $1 < \bar{\theta}(t)$ then, the solution of the Ricatti equation 72, (ii), satisfies the following inequality,*

$$\alpha\ Id \leq S(t) \leq \beta\ Id,$$

*for all $T_0 > 0$, for all $t \geq T_0$, where α and β depend on T_0, a_m, a_M **(but do not depend on c, $\bar{S}_0 \geq c\ Id$!)***

Straightforward computations with (72) give:

$$\frac{d}{dt}(\bar{\varepsilon}(t)'\bar{S}(t)\bar{\varepsilon}(t)) \leq -Q_m\ \bar{\varepsilon}'\bar{S}(t)^2\bar{\varepsilon} + 2\bar{\varepsilon}'\bar{S}(t)(\frac{1}{\bar{\theta}}(\tilde{b}(\bar{z}) - \tilde{b}(\bar{x}) - \tilde{b}^*(\bar{z})\bar{\varepsilon})) + \frac{\lambda(1-\bar{\theta})}{\bar{\theta}^2}\bar{\varepsilon}'\bar{S}(t)\bar{\varepsilon}, \tag{73}$$

where $Q \geq Q_m\ Id$.

In particular, if $t \geq T_0$, with α given by Lemma 2.2, this gives:

$$\frac{d}{dt}(\bar{\varepsilon}(t)'\bar{S}(t)\bar{\varepsilon}(t)) \leq -(Q_m\alpha + \frac{\lambda(\bar{\theta}-1)}{\bar{\theta}^2})\ \bar{\varepsilon}'\bar{S}(t)\bar{\varepsilon} + 2\bar{\varepsilon}'\bar{S}(t)(\frac{1}{\bar{\theta}}(\tilde{b}(\bar{z}) - \tilde{b}(\bar{x}) - \tilde{b}^*(\bar{z})\bar{\varepsilon})). \tag{74}$$

Using this equation, and again Lemma 2.2, we will now prove the theorem.

2.2.2. *Proof of the short term estimation 64.* This proof is in two steps. We will first prove an estimation for $T \geq t \geq T_0 > 0$, and after for $t \leq T_0$. Gluing them together, we get the short term estimation (64).

Step 1, $T \geq t \geq T_0$.

Straightforward computations using (74), Lemma 2.2 and the remark in Section 2.2.1 give:

$$\bar{\varepsilon}(t)'\bar{S}(t)\bar{\varepsilon}(t) \leq \bar{\varepsilon}(T_0)'\bar{S}(T_0)\bar{\varepsilon}(T_0)e^{-(Q_m\alpha - \frac{L'}{\theta(T)})(t-T_0)}. \tag{75}$$

Therefore $\bar{\varepsilon}(t)'\bar{S}(t)\bar{\varepsilon}(t) \leq \beta||\bar{\varepsilon}(T_0)||^2 e^{-(Q_m\alpha - \frac{L'}{\theta(T)})(t-T_0)}$, and finally:

$$T \geq t \geq T_0 : \quad ||\bar{\varepsilon}(t)||^2 \leq \frac{\beta}{\alpha}e^{-(Q_m\alpha - \frac{L'}{\theta(T)})(t-T_0)}||\bar{\varepsilon}(T_0)||^2. \tag{76}$$

Step 2, $t \leq T_0$.

We need a more straightforward estimation here. A very rough one is obtained just using Gronwall's identity. For certain $s, k > 0$, we have:

$$||\bar{P}(t)|| \leq (||\bar{P}(0)|| + k)e^{sT_0}. \tag{77}$$

We assume that $S(0) = S_0$ lies in the compact set: $c\ Id \leq S_0 \leq d\ Id$. As a consequence, $P(0) \leq \frac{1}{c}Id$.

By the equation (72), we have, for $t \leq T_0$: $\frac{d}{dt}(\bar{\varepsilon}) = (A - \bar{P}C'r^{-1}C)\bar{\varepsilon} + \frac{1}{\bar{\theta}}(\tilde{b}(\bar{z}) - \tilde{b}(\bar{x})) - \frac{\lambda(1-\theta)}{\bar{\theta}^2}F\bar{\varepsilon}$, hence:

$$||\bar{\varepsilon}(t)||^2 \leq ||\bar{\varepsilon}(0)||^2 + \int_0^t ||\bar{\varepsilon}(\tau)||^2(2||A|| + 2||C||^2||r^{-1}||\ ||\bar{P}|| + \frac{f}{\bar{\theta}})d\tau,$$

and by 77, we know that $||\bar{P}(t)|| \leq \varphi_1(T_0) + ||\bar{P}_0||\varphi_2(T_0)$. Then, since $\bar{P}_0 = \frac{1}{\theta_0}\Delta P_0\Delta(0)$, $\theta_0 > 1$, $||\bar{P}(t)|| \leq \varphi_1(T_0) + ||P_0||\varphi_2(T_0) \leq \varphi_1(T_0) + \frac{1}{c}\varphi_2(T_0) = \varphi(T_0, c)$.

$$||\bar{\varepsilon}(t)||^2 \leq ||\bar{\varepsilon}(0)||^2 + \bar{\nu}(T_0, c)\int_0^t ||\bar{\varepsilon}(\tau)||^2 d\tau,$$

and $\bar{\nu}(T_0, c)$ is a positive **decreasing** function of c.

Gronwall's inequality implies that:

$$||\bar{\varepsilon}(t)||^2 \leq \Psi(T_0, c)||\bar{\varepsilon}(0)||^2,$$

with: $\Psi(T_0,c)=e^{\nu T_0}$, $\Psi(T_0,c)$ is also a decreasing function of c.
In particular, $||\bar{\varepsilon}(T_0)||^2 \leq \Psi(T_0,c)||\bar{\varepsilon}(0)||^2$. Plugging this in (76), we get:

$$||\bar{\varepsilon}(t)||^2 \leq \frac{\beta}{\alpha}e^{-(Q_m\alpha-\frac{L'}{\theta(T)})(t-T_0)}\Psi(T_0,c)||\bar{\varepsilon}(0)||^2, \text{ for } T\geq t\geq T_0. \tag{78}$$

Hence, for $T\geq t\geq T_0$,

$$||\bar{\varepsilon}(t)||^2 \leq \frac{\beta}{\alpha}e^{-(Q_m\alpha-\frac{L'}{\theta(T)})t}e^{Q_m\alpha T_0}\Psi(T_0,c)||\bar{\varepsilon}(0)||^2. \tag{79}$$

Going back to $t\leq T_0$, we have:

$$\begin{aligned}||\bar{\varepsilon}(t)||^2 &\leq \Psi(T_0,c)||\bar{\varepsilon}(0)||^2 \leq \Psi(T_0,c)\frac{\beta}{\alpha}||\bar{\varepsilon}(0)||^2\\ &\leq \frac{\beta}{\alpha}e^{-(Q_m\alpha-\frac{L'}{\theta(T)})t}e^{Q_m\alpha T_0}\Psi(T_0,c)||\bar{\varepsilon}(0)||^2,\end{aligned}$$

Hence, in all cases (either $t\leq T_0$ or $T_0\leq t$), we have:

$$||\bar{\varepsilon}(t)||^2 \leq H(T_0,c)e^{-(Q_m\alpha-\frac{L'}{\theta(T)})t}||\bar{\varepsilon}(0)||^2,\ 0\leq t\leq T, \tag{80}$$

with $H(T_0,c)=\frac{\beta}{\alpha}\Psi(T_0,c)e^{Q_m\alpha T_0}$, a decreasing function of c. Therefore, going back to the initial time τ, since $t=\int_0^\tau\theta(v)dv$, and $t\leq T$, then, $\tau\leq\tau(T)$, and $t\geq\theta(\tau(T))\tau$:

$$||\tilde{\varepsilon}(\tau)||^2 \leq H(T_0,c)e^{-(Q_m\alpha\theta(\tau(T))-L')\tau}||\tilde{\varepsilon}(0)||^2, \tau(T)\geq\tau\geq 0,$$

if $\tilde{C}=Q_m\alpha\theta(\tau(T))-L'>0$, which is implied by

$$\theta_0 > e^{\lambda\tau(T)}(\frac{L'}{Q_m\alpha}-1)+1, \tag{81}$$

indeed, if (81) holds, since $\theta(\tau(T))=\bar{\theta}(T)=1+(\theta_0-1)e^{-\lambda\tau(T)}>\frac{L'}{Q_m\alpha}$.

Since $\varepsilon=\Delta^{-1}\tilde{\varepsilon}$, and $\theta>1$, $||\varepsilon(\tau)||^2\leq||(\Delta^{-1})||^2||\tilde{\varepsilon}(\tau)||^2\leq\theta^{2(n-1)}||\tilde{\varepsilon}(\tau)||^2$, we get, for all $\tau_0\geq\tau\geq 0$:

$$\begin{aligned}||\varepsilon(\tau)||^2 &\leq \theta^{2(n-1)}(\tau)H(T_0,c)e^{-(Q_m\alpha\theta(\tau_0)-L')\tau}||\varepsilon(0)||^2,\\ &\text{for } \theta_0 > e^{\lambda\tau_0}(\frac{L'}{Q_m\alpha}-1)+1,\end{aligned}$$

or equivalently, $\theta(\tau_0)>\dfrac{L'}{Q_m\alpha}$.

$H(T_0,c)$ is a decreasing function of c.

This is the short term estimation (64). If $\lambda=0$, it gives the standard high gain estimation.

2.2.3. *proof of the long term estimation.* Going back to (74), and using Lemma 4.2, in Section 4, we get, for all λ, $0\leq\lambda<1$, $t\geq T_0$,

$$\frac{d}{dt}(\bar{\varepsilon}(t)'\bar{S}(t)\bar{\varepsilon}(t)) \leq -k_1\ \bar{\varepsilon}'\bar{S}(t)\bar{\varepsilon}+k_2\ \bar{\theta}(t)^{(n-2)}||\bar{S}||\ ||\bar{\varepsilon}||^3,$$

where $k_1=Q_m\alpha$, k_2 is a positive constant.

Lemma 2.2, applied to the Riccati equation in (72), implies:

$$\frac{d}{dt}(\bar{\varepsilon}(t)'\bar{S}(t)\bar{\varepsilon}(t)) \leq -k_1\ \bar{\varepsilon}'\bar{S}(t)\bar{\varepsilon}+k_2'\ \bar{\theta}^{(n-2)}\ ||\bar{\varepsilon}(t)'\bar{S}(t)\bar{\varepsilon}(t)||^{\frac{3}{2}}, \tag{82}$$

for another positive constant k_2'.

Now, we apply Lemma 4.1, in Section 4, to get that, for $t \geq T \geq T_0$:

$$\bar{\varepsilon}(t)'\bar{S}(t)\bar{\varepsilon}(t) \leq 4e^{-k_1(t-T)}\bar{\varepsilon}(T)'\bar{S}(T)\bar{\varepsilon}(T), \tag{83}$$

as soon as

$$(\mathfrak{P})\ \bar{\varepsilon}(T)'\bar{S}(T)\bar{\varepsilon}(T)\bar{\theta}(T)^{2(n-2)} \leq \frac{(k_1)^2}{4(k_2')^2}.$$

Setting, $q = \bar{\varepsilon}(T)'\bar{S}(T)\bar{\varepsilon}(T)\bar{\theta}(T)^{2(n-2)}$, let us use the short term estimation (80). It gives $q \leq \beta H(T_0, c)e^{-(Q_m\alpha - \frac{L'}{\bar{\theta}(T)})T}||\bar{\varepsilon}(0)||^2\bar{\theta}(T)^{2(n-2)}$,

$$q \leq \beta H(T_0, c)e^{-(Q_m\alpha - \frac{L'}{\bar{\theta}(T)})T}||\bar{\varepsilon}(0)||^2\theta_0^{2(n-2)}.$$

If :

$$\theta_0 \geq e^{\lambda T}(\frac{2L'}{Q_m\alpha} - 1) + 1, \tag{84}$$

then $\frac{Q_m\alpha}{L'} - \frac{1}{\bar{\theta}(T)} \geq \frac{Q_m\alpha}{2L'}$. Indeed, in that case, $\bar{\theta}(T) \geq \theta(T) = 1 + (\theta_0 - 1)e^{-\lambda T} \geq \frac{2L'}{Q_m\alpha}$.

Then, let us chose $T = T^* = Log(\frac{\theta_0 - 1}{\frac{2L'}{Q_m\alpha} - 1})^{\frac{1}{\lambda}} \geq T_0$ (in order to get the equality in (84)). This is possible, since we can assume from the very beginning that $\frac{2L'}{Q_m\alpha} - 1 > 0$ (we can increase L' for this) and $\frac{\theta_0 - 1}{\frac{2L'}{Q_m\alpha} - 1} > e^{T_0} > e^{\lambda T_0}$ (we can take θ_0 large enough).

$$\begin{aligned} q &\leq \beta H(T_0, c)(\frac{\frac{2L'}{Q_m\alpha} - 1}{\theta_0 - 1})^{\frac{Q_m\alpha}{2\lambda}}||\bar{\varepsilon}(0)||^2\theta_0^{2(n-2)} \\ &\leq \beta H(T_0, c)||\bar{\varepsilon}(0)||^2(2(\frac{2L'}{Q_m\alpha} - 1))^{\frac{Q_m\alpha}{2\lambda}}\theta_0{}^{2(n-2)-\frac{Q_m\alpha}{2\lambda}}. \end{aligned}$$

Then, if:

$$\lambda < \frac{Q_m\alpha}{4(n-2)}, \tag{85}$$

for θ_0 large enough, for $||\varepsilon_0||$ bounded, q is arbitrarily small.

This means that the property $(\mathfrak{P})$ above is met at $T = T^*(\theta_0, \lambda)$, as soon as λ satisfies (85) and θ_0 is large enough.

In that case, (83) above holds, for $t \geq T^*$ $(\geq T_0)$:

$$\begin{aligned} \bar{\varepsilon}(t)'\bar{S}(t)\bar{\varepsilon}(t) &\leq 4e^{-k_1(t-T^*)}\bar{\varepsilon}(T^*)'\bar{S}(T^*)\bar{\varepsilon}(T^*), \\ &\leq 4e^{-k_1 t}(\frac{\theta_0 - 1}{\frac{2L'}{Q_m\alpha} - 1})^{\frac{k_1}{\lambda}}\bar{\varepsilon}(T^*)'\bar{S}(T^*)\bar{\varepsilon}(T^*). \end{aligned}$$

This implies, with (80):

$$\begin{aligned} ||\bar{\varepsilon}(t)||^2 &\leq 4\frac{\beta}{\alpha}e^{-k_1 t}(\frac{\theta_0 - 1}{\frac{2L'}{Q_m\alpha} - 1})^{\frac{k_1}{\lambda}}||\bar{\varepsilon}(T^*)||^2, \\ &\leq 4\frac{\beta}{\alpha}H(T_0, c)e^{L'T^*}e^{-k_1 t}(\frac{\theta_0 - 1}{\frac{2L'}{Q_m\alpha} - 1})^{\frac{k_1}{\lambda}}||\varepsilon_0||^2, \\ &\leq 4\frac{\beta}{\alpha}H(T_0, c)e^{-k_1 t}(\frac{\theta_0 - 1}{\frac{2L'}{Q_m\alpha} - 1})^{\frac{k_1 + L'}{\lambda}}||\varepsilon_0||^2, \end{aligned}$$

for $t \geq T^*$ $(\geq T_0)$.

For $t \leq T^*$, using (80), and the fact that $k_1 = Q_m \alpha$:

$$||\bar{\varepsilon}(t)||^2 \leq H(T_0, c) e^{-k_1 t} e^{L' t} ||\varepsilon_0||^2,$$
$$\leq H(T_0, c) e^{-k_1 t} 4 \frac{\beta}{\alpha} \left(\frac{\theta_0 - 1}{\frac{2L'}{Q_m \alpha} - 1}\right)^{\frac{k_1 + L'}{\lambda}} ||\varepsilon_0||^2,$$

because $\frac{\theta_0 - 1}{\frac{2L'}{Q_m \alpha} - 1} > 1$.

Therefore, for all $t \geq 0$:

$$||\bar{\varepsilon}(t)||^2 \leq 4 \frac{\beta}{\alpha} H(T_0, c) e^{-k_1 t} \left(\frac{\theta_0 - 1}{\frac{2L'}{Q_m \alpha} - 1}\right)^{\frac{k_1 + L'}{\lambda}} ||\varepsilon_0||^2,$$
$$\leq \tilde{H}(T_0, c, \lambda) e^{-k_1 t} \theta_0^{\frac{k_1 + L'}{\lambda}} ||\varepsilon_0||^2,$$

where $\tilde{H}$ is a decreasing function of c. Hence:

$$||\tilde{\varepsilon}(\tau)||^2 \leq \tilde{H}(T_0, c, \lambda) e^{-k_1 t} \theta_0^{\frac{k_1 + L'}{\lambda}} ||\varepsilon_0||^2,$$

and, with $t = \tau + \frac{\theta_0 - 1}{\lambda}(1 - e^{-\lambda \tau})$,

$$||\tilde{\varepsilon}(\tau)||^2 \leq \tilde{H}(T_0, c, \lambda) e^{-k_1 \tau} \theta_0^{\frac{k_1 + L'}{\lambda}} e^{-k_1 \frac{\theta_0 - 1}{\lambda}(1 - e^{-\lambda \tau})} ||\varepsilon_0||^2.$$

Finally,

$$||\varepsilon(\tau)||^2 \leq \bar{H}(T_0, c, \lambda) e^{-k_1 \tau} ||\varepsilon_0||^2 \theta_0^{\frac{k_1 + L'}{\lambda} + 2(n-1)} e^{-k_1 \frac{\theta_0}{\lambda}(1 - e^{-\lambda \tau})},$$

where $\bar{H}(T_0, c, \lambda)$ is a decreasing function of c.

This is the long term estimation. It holds as soon as λ satisfies (85), and for θ_0 large, depending on λ.

2.3. Practical implementation: making the observer "recursive". We consider a one parameter family $\{O_\tau, \tau \geq 0\}$ of observers of type (63), indexed by the time, each of them starting from S_0, θ_0, at the current time τ. In fact, in practice, it will be sufficient to consider, at time τ, a slipping window of time, $[\tau - T, \tau[$, and a finite set of observers $\{O_{t_i},\ \tau - T \leq t_i \leq \tau\}$, with $t_i = \tau - i\frac{T}{N}$, $i = 1, ..., N$.

As usual, we call the term $I(\tau) = \hat{y}(\tau) - y(\tau)$, (the difference at time τ 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 the absolute value 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 θ_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.

This looks very promising. We show on an example in the next section, that it works very well.

3. Application: observation of a binary distillation column

3.1. The constant molar overflow model. The model we consider is the classical "constant molar overflow" (CMO) model. It is one of the most simple distillation models, and it is used by many process engineers (for instance, even in its static form, it is used for simple short-cut distillation calculations).

Since everything here follows from the very special "tridiagonal" structure of this model, and since any reasonable distillation model possesses such a structure, all what we do in this paper can certainly be extended to more precise distillation models.

The equations are based upon:

a. a thermodynamical relation describing the thermal equilibria for each tray.

b. Material balances on each plate.

Thermal balance on each plate is replaced by the "Lewis hypotheses", that lead to the fact that the liquid and vapor flowrates along the column are constant in the "stripping" (above the feed) and "rectification" (below the feed) zones. For justification of these assumptions, see [**19**].

The equations of this model are:

Total condenser:

$$H_1 \frac{dx_1}{dt} = (V + FV)(y_2 - x_1). \tag{86}$$

Rectifying section, $j = 2, \cdots, f-1$:

$$H_j \frac{dx_j}{dt} = L(x_{j-1} - x_j) + (V + FV)(y_{j+1} - y_j). \tag{87}$$

Feed tray:

$$\begin{aligned} H_f \frac{dx_f}{dt} &= FL(Z_F - x_f) + FV(k(Z_F) - y_f) \\ &\quad + L(x_{f-1} - x_f) + V(y_{f+1} - y_f). \end{aligned} \tag{88}$$

Stripping section, $j = f+1, \cdots, n-1$:

$$H_j \frac{dx_j}{dt} = (L + FL)(x_{j-1} - x_j) + V(y_{j+1} - y_j). \tag{89}$$

Bottom of the column:

$$H_n \frac{dx_n}{dt} = (L + FL)(x_{n-1} - x_n) + V(x_n - y_n). \tag{90}$$

The parameters have the following physical meaning:

n	number of trays,
f	index of the feed tray,
H_j	liquid hold up on the j^{th} tray (a geometric constant),
x_j	liquid composition on the j^{th} tray,
y_j	vapor composition on the j^{th} tray,
FL, FV, L, V	feed (liquid and vapor), reflux and vapor flow,
Z_F	feed composition (molar fraction of light component in feed).

On each tray the liquid and vapor compositions, x_j and y_j, are linked by the liquid/vapor equilibrium law $y_j = k(x_j)$. We assume that the function k is monotonic, *i.e.* we do not consider azeotropic distillation.

Each of the equations is relative to a tray. It just expresses the accumulation of the liquid on the corresponding tray, and the thermodynamical equilibrium.

The condenser and the bottom of the column are assimilated to tray 1 and tray n respectively. The state of the model is the liquid composition profile of the more volatile component on each tray, denoted by (x_j).

The top and bottom product compositions x_1 and x_n are the two observed variables. In practice, they are also the two variables that one wants to control: they are the "qualities" of the products going out of the column.

The two control variables are the reflux flow-rate L and the vapor flow-rate V.

There are also two disturbances to be counteracted:

a. changes in the feed rate $F = FL + FV$. In general this is a "measured disturbance", (a flowrate measurement),

b. the in-feed composition Z_F. In general, it is unknown, and it is practically very expensive to "observe it". Moreover, it may change brutally. We will consider this feed composition Z_F as an unknown (constant) state variable. When Z_F changes, the consequence is **a jump of the state of the system.**

The qualitative properties of this model are very nice (see [**15**], [**33**], [**32**]):

a. For positive control variables L and V, (negative doesn't physically makes sense), the "physical" domain $D = [O, 1]^n$ is positively invariant under the dynamics. This means that all the state variables x_j remain between 0 and 1.

b. In the domain D, all other variables (than the x_i's and the y_i's) being constant, **there is a single equilibrium, which is globally asymptotically stable.**

c. It has very nice observability properties, as will be discussed later on.

Our goal in this section is to construct an estimator of the state x, and more specifically of the feed composition Z_F, by using the results of the previous sections.

3.2. Observability of the model and synthesis of the observer. A complete analysis of observability and observer synthesis has been carried out in [**15**] in the general case. It happens that, even if the feed is considered as an unknown state variable (meeting the equation $\frac{dZ_F}{dt} = 0$), the model is observable in the strongest possible sense. In particular, as we shall see, it can be put in a normal form similar to (62).

Our purpose here is just to apply the observer described in the previous sections. Hence, we will fix a special case of distillation column. But all what we show works in general. We will chose:

- $n = 5$ and $f = 3$,
- The function k is a diffeomorphism from $[0,1]$ into itself and is given by,

$$k(x) = \frac{\alpha x}{1 + (\alpha - 1)x}.$$

  Here α is the "relative volatility" of the mixture. It is a physical parameter larger than 1 (but close to 1). The closer to one, the most difficult distillation. If $\alpha = 1$, the two products are thermodynamically identical, and cannot be distillated (the model is not controllable).
- Let us observe that k is a diffeomorphism from $\left]-\frac{1}{\alpha-1}, +\infty\right[$ to $\left]-\infty, \frac{\alpha}{\alpha-1}\right[$.
- The feed is assumed to enter the column at its "bubble point". As a consequence, $F = FL$.

Let us make the following change of state variables: $\xi_1 = x_1$, $\xi_2 = k(x_2)$, $\xi_3 = x_3$, $\xi_4 = x_4$, $\xi_5 = x_5$ and $\xi_6 = Z_F$.

Then, the system can be rewritten as:

$$(91) \quad \begin{cases} H_1 \frac{d\xi_1}{dt} &= V(\xi_2 - \xi_1), \\ H_2 \frac{d\xi_2}{dt} &= k'\left(k^{-1}(\xi_2)\right)\left(L(\xi_1 - k^{-1}(\xi_2)) + V(k(\xi_3) - \xi_2)\right), \\ H_3 \frac{d\xi_3}{dt} &= F(\xi_6 - \xi_3) + L(k^{-1}(\xi_2) - \xi_3) + V(k(\xi_4) - k(\xi_3)), \\ H_4 \frac{d\xi_4}{dt} &= (L+F)(\xi_3 - \xi_4) + V(k(\xi_5) - k(\xi_4)), \\ H_5 \frac{d\xi_5}{dt} &= (L+F)(\xi_4 - \xi_5) + V(\xi_5 - k(\xi_5)), \\ H_6 \frac{d\xi_6}{dt} &= 0, \end{cases}$$

or:

$$(92) \quad \frac{d\xi_t}{dt} = A(L,V)\,\xi_t + \widetilde{b}(L,V,\xi_t),$$

where,

$$A(L,V) = \begin{pmatrix} 0 & \frac{V}{H_1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{F}{H_3} \\ 0 & 0 & \frac{L+F}{H_4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{L+F}{H_5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

and,

$$\widetilde{b}(L,V,\xi) = \begin{pmatrix} -\frac{V}{H_1}\xi_1 \\ k'\left(k^{-1}(\xi_2)\right)\left(L(\xi_1 - k^{-1}(\xi_2)) + V(k(\xi_3) - \xi_2)\right)/H_2 \\ \left(-F\xi_3 + L(k^{-1}(\xi_2) - \xi_3) + V(k(\xi_4) - k(\xi_3))\right)/H_3 \\ \left(-(L+F)\xi_4 + V(k(\xi_5) - k(\xi_4))\right)/H_4 \\ \left(-(L+F)\xi_5 + V(\xi_5 - k(\xi_5))\right)/H_5 \\ 0 \end{pmatrix},$$

$$= \begin{pmatrix} \widetilde{b}_1(V,\xi_1) \\ \widetilde{b}_2(L,V,\xi_1,\ldots,\xi_5) \\ \widetilde{b}_3(L,V,\xi_3,\xi_4,\xi_5) \\ \widetilde{b}_4(L,V;\xi_4,\xi_5) \\ \widetilde{b}_5(L,V,\xi_5) \\ 0 \end{pmatrix}.$$

The observations are then given by

$$y = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \xi = C\xi.$$

Now, since in fact the only pertinent **(and positively invariant)** part of the state space is $D' = [0,1]^6$, we can manage the things for $\widetilde{b}$ be compactly supported, as in section 1, and unchanged on D'. Let us change $\widetilde{b}(L,V,\xi)$ in the following way outside $[0,1]^6$: replace $\widetilde{b}(L,V,\xi)$ by $b(L,V,\xi) = \widetilde{b}(L,V,\Phi(\xi))$ where $\Phi(\xi_1,\ldots,\xi_6) = (\varphi(\xi_1),\ldots,\varphi(\xi_6))$ and $\varphi(\xi)$ is any C^∞ function from $\mathbb{R}$ to $[0,1]$ equal to one in $[0,1]$ and equal to zero outside $\left]-\frac{1}{\alpha-\frac{1}{2}}, \frac{\alpha}{\alpha-\frac{1}{2}}\right[$. This modification does not change the "physical trajectories".

Our system has the property to be *observable for any input*, as soon as the control variables L and V are > 0. Here, we assume that L, V are bounded from below (and from above) by > 0 constants:

$$L_M \geq L(t) \geq \varepsilon_1 > 0, \quad V_M \geq V(t) \geq \varepsilon_2 > 0.$$

This assumption is the analog of the assumption $0 < a_m \leq a_i(u) \leq a_M$, in section 1. It is a realistic requirement from the physical point of view.

To finish, let us point out the fact that we are in **case 1** of Chapter 2 above (i.e. the nongeneric case): The number of observations is equal to the number of control variables (it is 2).

Due to these observability properties, we will be able to apply the observer of the previous section 2.3. In fact, it will be an adaptation of the results of section 2, Theorem 2.1, to this multi-output case.

We leave the reader to check (this is really straightforward) that all the reasoning in the proof of Theorem 2.1 can be strictly repeated, and that the statements of this theorem are valid for the distillation column.

Of course, in practice, we didn't compute the theoretical bounds λ_0 and $\theta_0(\lambda)$. We have just got some values for them by experimentation. Also, the number N of

"parallel" observers, and the "sampling times" t_i of section 2.3 have been chosen experimentally.

Finally, the state of our observer is the collection of the states of N independent observers $(z_i, S_i, \theta_i)_{i=1,\dots,N}$. Each observer is a set of three equations of the following form:

$$\begin{cases} \frac{dz}{dt} &= A(u)\,z + b(u,z) - S(t)^{-1} C^T R_\theta^{-1} (Cz - y(t)) \\ \frac{dS}{dt} &= -(A(u) + b^*(z,u))' S - S(A(u) + b^*(z,u)) + C' R_\theta^{-1} C - S Q_\theta S \\ \frac{d\theta}{dt} &= \lambda(1-\theta) \end{cases}$$

where $u = (L, V)$.

Due to the multi-output structure, with "Brunovsky-like" blocks of different dimensions (4 and 2), a way to make the proof of Theorem 2.1 work, is to take a matrix R depending also on θ, as shown below. This could be avoided by increasing the dimension of the state as explained in [**15**].

It is not hard to check that a good choice is to set:

$$\Delta = \mathrm{diag}\left(\frac{1}{\theta^2}, \frac{1}{\theta^3}, \frac{1}{\theta^2}, \frac{1}{\theta}, 1, \frac{1}{\theta^3}\right)$$

with $Q_\theta = \theta^2 \Delta^{-1} Q \Delta^{-1}$ and $R_\theta = \left(C\Delta^{-1}C'\right) R \left(C\Delta^{-1}C'\right)$.

In practice, we have chosen $N = 5$ observers, and we have taken a regular sampling $\frac{T}{N}$. That is to say, at each time step $k\frac{T}{N}$, the oldest observer is replaced by a new one (with $\theta = \theta_0$ and a new guess of state and covariance matrix). At the beginning of the simulation, we chose an initial value θ_0 of θ for each observer, such that the i^{th} observer has $\theta_i = 1 + e^{-\lambda \frac{(i-1)T}{N}} (\theta_0 - 1)$, see figure 3, where "crosses" represent reinitializations.

We have implemented our observer as described in the previous section. Since the state has dimension 6, each observer requires to solve 28 ordinary differential equations (for the state, the Riccati matrix, and the very simple equation for θ). Finally, our observer is a set of 140 ODE's. We have solved it in conjunction with the model (6 equations) using LSODAR from ODEPACK ([**17**]), without taking into account the possibility of decoupling these equations (which are indeed equivalent to five systems of 34 equations, including the model into each system). A simulation of 3 hours of real time takes about 40 seconds on a Pentium III machine.

3.3. Simulation results. We have chosen the following constant parameters:

- Hold-up $H_1 = 40$, $H_j = 10$ for $j = 2, 3, 4$ and $H_5 = 80$,
- Relative volatility $\alpha = 2$.

We have applied the following scenario:

- During the simulation, the state noise is simulated by the sum of several sine functions at some random frequencies representing a band limited noise with an amplitude of 10^{-8} before the time $t_2 = 116\,\mathrm{mn}\,40\,\mathrm{s}$ and 10^{-2} after this time,

- Moreover, at time $t_1 = 66\,\mathrm{mn}\,40\,\mathrm{s}$, we simulate a step in the feed quality Z_F from 0.45 to 0.60. Hence we can consider that there is no perturbation before time t_1, where a large "jump of the state" occurs,

- after that, nothing happens until time t_2 where a periodic perturbation on Z_F is applied.

We have also added a measurement noise at some random high frequencies and with amplitude of 10^{-2}. The effect of noise can be seen on Figure 1 (top and bottom lines).

To make the simulation more realistic, we have applied a very simple controller, which calculates the inputs L and V in order to regulate top and bottom qualities at a reasonable level (that is, 73% for the top quality and 23% for the bottom quality).

As we said already, the parameters of the observers where tuned in order to obtain good performances, and not caring about the theoretical bounds.

Practically, we have used $\theta_0 = 10$, $\frac{T}{N} = 10\,\text{mn}$ and $\lambda = \frac{1}{600}\,\text{s}^{-1}$, in such a way that the time of life of an observer is $T = 50\,\text{mn}$, and then an old observer has $\theta \approx 1.16$. Also, there is always an observer with $\theta > 4.3$ which is running.

Finally, R is equal to 10^{-2} times the 2×2–identity matrix and Q is 10^{-9} times the 6×6–identity matrix.

First of all, the behaviour of the observer is very good during the unmodelled transient as well as during smooth operation, see Figure 1: top and bottom quality measurements are plotted, as well as the unknown feed quality, each curve being represented by a continuous line. The overall estimation of the feed quality, corresponding to the estimation of the feed quality provided by the observer with the **smallest innovation,** is represented by a dashed line. It is very close to the actual feed quality.

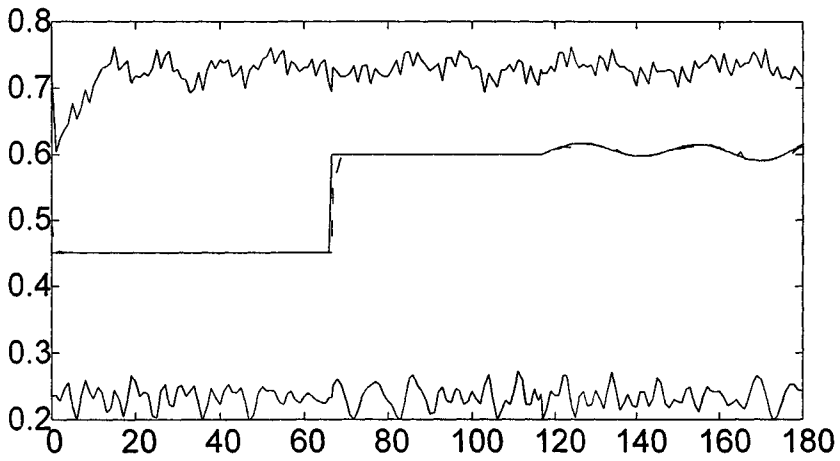

FIGURE 1. Measured output and estimation of the feed quality.

A more accurate plot is presented on Figure 2 where we have only shown the relative estimation error of the feed quality. The estimation provided by the best observer (in our sense, that is to say, the observer with minimal innovation) is the continuous line. The crosses represent the estimation of Z_F provided by other

observers every minute. One can see that our criteria on the innovation to select the right observer is a good choice, at least in this case.

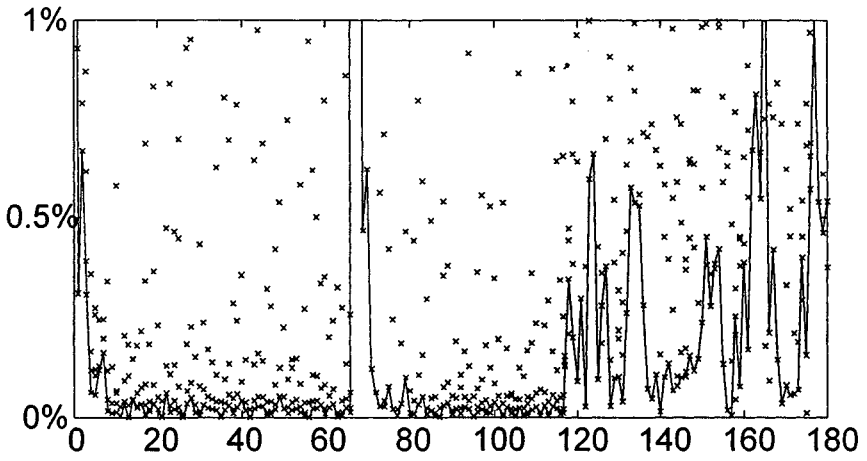

FIGURE 2. Relative error between the actual feed quality and its estimation by the selected observer (continuous line) and the others.

Moreover, the behavior of the observer is very close to what we expected from the theoretical results:

- When no perturbation arises, the best observer (that is to say the observer with the smallest innovation) is the one with the smallest value of θ *i.e.* the oldest observer which is also the observer which is the closest to the pure extended Kalman observer.

-If a large perturbation occurs (such as the feed change at time $t_1 = 66\,\text{mn}\,40\,\text{s}$), the best observer becomes the youngest one, *i.e.* the observer with the highest θ.

-Of course, small perturbations are well corrected by oldest or intermediate observers. This is very clear on the figure 4.

Our conclusion, from these simulations, is that even if the use of several observers in parallel requires the introduction of new tuning parameters (θ_0, λ, N and T), the choice of these new parameters is very easy, due to their very clear effect on the results.

>From a practical point of view, θ_0, λ, N and T have to be chosen such that at any time, there is an HGEKF and an EKF-like observer running at the same time, that is to say such that $1 + e^{-\lambda \frac{T}{N}} (\theta_0 - 1)$ is large enough (to ensure that at least one observer is a HGEKF) and such that $1 + e^{-\lambda T} (\theta_0 - 1)$ is close to 1.

Also, an important point, for people that are used to tune Kalman's observers, is that the choice of the Q and R matrices is less crucial than with a single observer which has to be tuned in order to be efficient both with and without perturbations.

Moreover, this approach allows us to obtain a diagnosis of abnormal behavior: if the smallest innovation is provided by the last reinitialized observer then one can conclude that the model has encountered a perturbation. If this happen for a

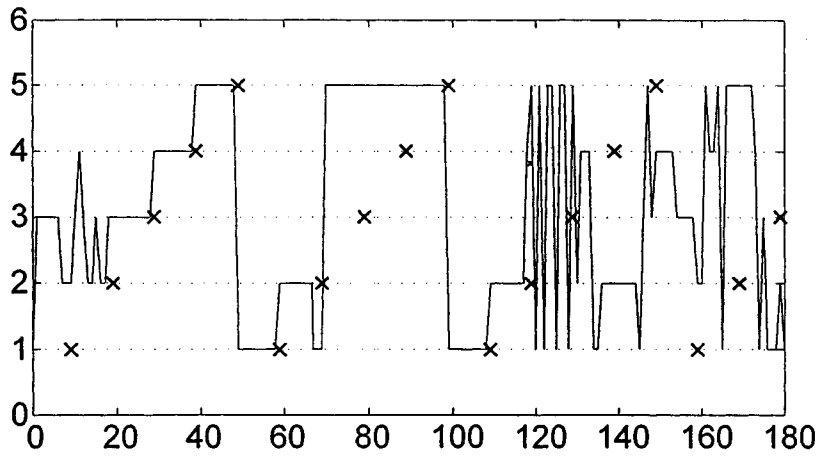

FIGURE 3. The 5 observers. Time of reinitialization of each observer (×), and the best one (continuous line).

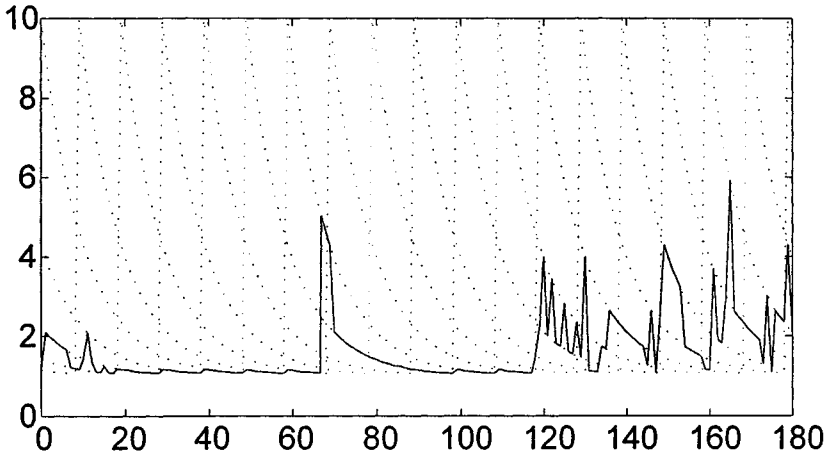

FIGURE 4. Various values of θ versus time (dotted lines), and best observer (continuous line).

long time then one can conclude that the model has some difficulties to deal with certain unmodelled perturbations. Indeed, the scenario that we have applied in our simulations can be easily deduced from the figure 4.

4. Appendix. Technical lemmas

LEMMA 4.1. *Let* $\{x(t) > 0,\, t \geq 0\} \subset \mathbb{R}^n$ *be absolutely continuous, and satisfying:*

$$\frac{dx}{dt} \leq -\lambda x + kx\sqrt{x},$$

for almost all $t > 0$, *for* $\lambda, k > 0$. *Then, as soon as* $x(0) < \frac{\lambda^2}{4k^2}$, $x(t) \leq 4x(0)e^{-t\lambda}$.

LEMMA 4.2. *Let* $B = \tilde{b}(z) - \tilde{b}(x) - \tilde{b}^*(z)\varepsilon$ *be as in Section 2:* $\varepsilon = z - x$, $\tilde{b}(x) = \Delta b(\Delta^{-1}x)$, $\tilde{b}^*(z) = \Delta b^*(\Delta^{-1}x)\Delta^{-1}$, *where* $b^*(x)$ *is the Jacobian matrix of* b *at* x, *and where* b *is compactly supported.* $\Delta = diag(1, \frac{1}{\theta}, ..., \frac{1}{\theta^{n-1}})$, $\theta \geq 1$. *Then,* $||B|| \leq K\ \theta^{n-1}||\varepsilon||^2$, *for some* $K > 0$.

Bibliography

[1] R. ABRAHAM, J. ROBBIN, Transversal mappings and flows ; W.A. Benjamin, Inc., 1967.

[2] M. BALDE, P. JOUAN, Observability of control affine systems, ESAIM/COCV, Vol. 3, pp. 345-359, 1998.

[3] J.S. BARAS, A. BENSOUSSAN, M.R. JAMES, "*Dynamic observers as asymptotic limits of recursive filters: special cases*", SIAM J. Appl. Math., **48**, (1988), 1147–1158.

[4] N. BOURBAKI, Eléments de Mathématiques, Topologie générale, livre III, Actualités Scientifiques et Industrielles, 1142, Hermann, Paris, 1961.

[5] R. BUCY, P. JOSEPH, Filtering for stochastic processes with applications to guidance, Chelsea publishing company, 1968, second edition, 1987.

[6] E. BUSVELLE, J.P. GAUTHIER, High-gain and non high-gain observes for nonlinear systems, december 2000, to appear in "Proceedings of GCTAA 2000, in honor of Vel. Jurdjevic", World Scientific.

[7] J. CARR, Applications of centre manifold theory, Appl. Math. Sci. 35, Springer Verlag, 1981.

[8] F. DEZA, Contribution to the synthesis of exponential observers, Phd thesis, INSA de Rouen, France, June 1991.

[9] F.DEZA, E.BUSVELLE, J.P.GAUTHIER, High-gain estimation for nonlinear systems, Systems and Control Letters 18, pp. 295-299, 1992.

[10] M.FLIESS, I.KUPKA, A finiteness criterion for nonlinear input-output differential systems, SIAM Journal Contr. and Opt., 21, pp. 721-728, 1983.

[11] J.P. GAUTHIER, H. HAMMOURI, I. KUPKA, Observers for nonlinear systems; IEEE CDC Conference, december, 1991, pp. 1483-1489; Brighton, England.

[12] J.P GAUTHIER, H. HAMMOURI, S. OTHMAN, A simple observer for nonlinear systems. IEEE Trans. Aut. Control, 37, pp. 875-880, 1992.

[13] J.P. GAUTHIER, I. KUPKA, Observability and observers for nonlinear systems. SIAM Journal on Control, vol. 32, N° 4, pp. 975-994, 1994.

[14] J.P. GAUTHIER, I. KUPKA, Observability for systems with more outputs than inputs. Mathematische Zeitschrift, 223, pp. 47-78, 1996.

[15] J.P.GAUTHIER, I. KUPKA, Deterministic Observation Theory and Applications, Cambridge University Press, 2001.

[16] R. HERMANN and all., Nonlinear controllability and observability, IEEE Trans. Aut. Control, AC-22, pp. 728-740, 1977.

[17] A. C. HINDMARSCH, *Odepack, a systematized collection of ode solvers*, in scientific computing, r. s. Stepleman et al. (eds.), North-Holland, Amsterdam, 1983, pp. 55-64

[18] M.W. HIRSCH, Differential Topology, Springer-Verlag, Graduate texts in maths, 1976.

[19] C.D. HOLLAND, Multicomponent Distillation, Englewood Cliffs, New-Jersey, USA: Prentice Hall, 1963.

[20] A. JASWINSKY, Stochastic processes and filtering theory, Academic Press, New York, 1970.

[21] P. JOUAN, Singularités des systèmes non linéaires, observabilité et observateurs, PHD thesis, Université de Rouen, 1995.

[22] P. JOUAN, Observability of real analytic vector fields on a compact manifold, Systems and control letters 26, pp. 87-93, 1995.

[23] P. JOUAN, J.P. GAUTHIER, Finite singularities of nonlinear systems. Output stabilization, observability and observers. Journal of Dynamical and Control Systems, vol. 2, N° 2, 1996, pp. 255-288.

[24] J. KURZWEIL, On the inversion of Lyapunov's second theorem, On stability of motion, Transl. Am. Math. Soc. pp. 19-77, 1956.

[25] J. LASALLE, S. LEFSCHETZ, "Stability by Lyapunov's Direct Method with Applications", New York: Academic Press, 1961.
[26] S. LEFSCHETZ, "Ordinary Differential Equations: Geometric Theory", J. Wiley, Intersciences, 1963.
[27] D.G. LUENBERGER, Observers for multivariable systems ; IEEE Trans. Aut. Control 11, 1966, pp. 190-197.
[28] J.L.MASSERA, Contribution to stability theory, Annals of Maths 64, pp. 182-206, 1956.
[29] R. NARASIMHAN, Introduction to the theory of analytic spaces, Springer Verlag, Lecture Notes in Mathematics 25, 1966.
[30] J. PICARD, "*Efficiency of the extended Kalman filter for nonlinear systems with small noise*", SIAM J. Appl. Math., **51**, No3, (1991), 843-885.
[31] N. ROUCHE, P. HABETS, M. LALOY, Stability theory by Lyapunov's direct method, Lecture notes in applied mathematical sciences, Vol. 22, New York: Springer Verlag.
[32] H.H. ROSENBROCK, A Lyapunov function with applications to some nonlinear physical systems, Automatica, 1, pp. 31-53, 1962.
[33] P. ROUCHON, *Simulation dynamique et commande non linéaire des colonnes à distiller*, Thèse de l'école des mines de Paris, 1990
[34] H.J. SUSSMANN, Trajectory regularity and real analyticity, some recent results, Proceedings of 25th CDC conference, Athens, Greece, dec 1986.
[35] H.J. SUSSMANN, Single input observability of continuous time systems, Mathematical Systems Theory, 12, N.4, pp. 371-393,1979.
[36] F. VIEL, Stabilité des systèmes non linéaires controlés par retour d'état estimé. Application aux réacteurs de polymérisation et aux colonnes à distiller, Thèse de l'université de Rouen, 1994.
[37] F. VIEL, E. BUSVELLE, J.P. GAUTHIER, A stable control structure for binary distillation columns, International Journal on Control, Vol 67, N° 4, pp. 475-505, 1997.
[38] F.W. WILSON Jr , The structure of the level surfaces of a Lyapunov function, Journ. Diff. Equ. 3, pp. 323-329, 1967.
[39] H. WHITNEY, Analytic extensions of differentiable functions defined in closed sets, Trans. Am. Math. Soc. 36, pp. 63-89, 1934.
[40] O. ZARISKI, P. SAMUEL, Commutative Algebra, Van Nostrand Company, 1958.