



## **Continuous-discrete time observer design for Lipschitz systems with sampled measurements (long version)**

Thach Ngoc Dinh, Vincent Andrieu, Madiha Nadri, Ulysse Serres

### **► To cite this version:**

Thach Ngoc Dinh, Vincent Andrieu, Madiha Nadri, Ulysse Serres. Continuous-discrete time observer design for Lipschitz systems with sampled measurements (long version). 2014. <hal-00944153>

**HAL Id: hal-00944153**

**<https://hal.science/hal-00944153v1>**

Submitted on 10 Feb 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Continuous-discrete time observer design for Lipschitz systems with sampled measurements (long version)

Thach Ngoc Dinh<sup>\*†</sup>, Vincent Andrieu<sup>‡</sup>, Madiha Nadri<sup>‡</sup>, Ulysse Serres<sup>‡</sup>

## Abstract

This paper concerns observers design for Lipschitz nonlinear systems with sampled output. Using reachability analysis, an upper approximation of the attainable set is given. When this approximation is formulated in terms of a convex combination of linear mappings, a sufficient condition is given in terms of linear matrix inequalities which can be solved employing a linear matrix inequalities solver. This novel approach seems to be an efficient tool to solve the problem of observer synthesis for a class of Lipschitz systems of small dimensions.

## Index Terms

Continuous discrete-time observers, reachable sets, Pontryagin Maximum Principle, LMI.

## I. INTRODUCTION

In this paper, we consider the state estimation problem for a class of continuous time Lipschitz systems with discrete time measurements. More precisely, the aim of our study is to design a continuous-discrete state observer. The use of this type of algorithm to estimate the state

<sup>\*</sup>EPI INRIA DISCO, Laboratoire des Signaux et Systèmes, L2S - CNRS - Supélec, 3 rue Joliot-Curie, 91192 Gif sur Yvette cedex, France; email: [thach.dinh@lss.supelec.fr](mailto:thach.dinh@lss.supelec.fr)

<sup>†</sup>Centre de Robotique, MINES ParisTech, 75272 Paris, France; email: [thach.dinh@mines-paristech.fr](mailto:thach.dinh@mines-paristech.fr)

<sup>‡</sup>Université de Lyon, F-69622 Lyon, France; Université Lyon 1, Villeurbanne; LAGEP, UMR CNRS 5007, 43 bd du 11 novembre 1918, 69100 Villeurbanne, France; email: <https://sites.google.com/site/vincentandrieu> & [nadri@lagep.univ-lyon1.fr](mailto:nadri@lagep.univ-lyon1.fr) & [ulysse.serres@univ-lyon1.fr](mailto:ulysse.serres@univ-lyon1.fr)

This work was supported by ANR LIMICOS contract number 12 BS03 005 01

of nonlinear systems has already been investigated in the literature. It can be traced back to Jazwinski who introduced the continuous-discrete Kalman filter to solve a filtering problem for stochastic continuous-discrete time systems (see [9]). Inspired by this approach, the popular high-gain observer introduced in [8] has been adapted to the continuous-discrete context in [7]. Since then, different approaches, have been investigated by different authors. In [4], the robustness of an observer with respect to time discretization is studied. In [10, 1, 11] observers are designed from an output predictor. Some other approach based on time delayed techniques have also been considered in [17].

In our study we focus on the preliminary work of [7], and consider the case in which the continuous-discrete observer is obtained in two steps:

- i) when no measurement is available, the state estimate is computed by integrating the model;
- ii) when a measurement occurs, the observer makes an impulsive correction on the state estimate.

Note that in [7] the correction gain of this impulsive correction is obtained by integrating a continuous-discrete time Riccati equation. However, in the following, inspired by [16], we will consider a constant correction term.

In most of the works cited above, the asymptotic convergence of the estimate to the state is obtained by dominating the Lipschitz nonlinearities with high-gain techniques. However, as this is now well understood there is a trade-off between the high-gain parameter and the measurement step size. This can lead to restrictive design conditions on the sampling measurement time (see for instance [2]) which may prevent the use of such technique in practice.

Recently, a new observer design methodology for Lipschitz nonlinear systems with continuous time measurements has been introduced in [21]. In this approach, it is shown that the differential equation satisfied by the estimation error can be rewritten in the form of a linear parameter varying system (LPV). Hence, the convergence to zero of the estimation error can be obtained by solving some specific linear matrix inequalities (see Section II for a brief summary of this approach). The aim of our paper is to extend the approach presented in [21] to the discrete time measurement case. In the adopted strategy, the main problem is decomposed into two subproblems:

- i) *Computation of an upper approximation of the reachable set for a bilinear system.* This set characterizes the possible expansion of the estimation error between two measurements

when the estimate is given by integrating the model.

- ii) *Construction of a correction term  $K$  ensuring the convergence to zero of a quadratic error Lyapunov function.* As in [21], this step is performed through linear matrix inequalities (LMIs) techniques.

In this paper, we address the first problem by considering systems with specific structure: upper triangular nonlinearities (as in the preliminary version of this work in [3]), and in observability canonical form. This allows us to obtain a constructive approach to the synthesis of observers for a wide class of Lipschitz nonlinear systems while avoiding standard high-gain approach. Consequently, we hope to build observers with larger sampling period than those obtained by usual techniques. However, the large size of the linear matrix inequality restricts the use of this approach only to low dimensional systems.

The paper is organized as follows. In Section II, after having defined the considered class of systems (in Subsection II-A), some preliminary results and in particular the approach of [21] for continuous time measurements are recalled (in Subsection II-B). An approach of observer design based on reachability set analysis is presented in Subsection II-C. Section III-A, is devoted to the study of a first particular class of systems: feedforward systems. Section III-B concerns uniformly observable systems. Simulations of two academic examples are given in Section IV to illustrate the methodology proposed in this paper. *Some notations:*

- $M \in \mathbb{R}^{m \times n}$  means that  $M$  is a  $m \times n$  matrix.
- If  $M \in \mathbb{R}^{m \times n}$ ,  $M' \in \mathbb{R}^{n \times m}$  denotes the transpose matrix.
- A symmetric matrix  $M \in \mathbb{R}^{n \times n}$  is positive definite (resp. negative definite) if for all vectors  $v \in \mathbb{R}^n$ ,  $v'Mv > 0$  (resp.  $v'Mv < 0$ ). We then write  $M > 0$  (resp.  $M < 0$ );
- The dot  $\dot{\cdot}$  denotes the derivative with respect to time.
- $|v|$  denotes the usual Euclidean norm of  $x \in \mathbb{R}^n$ , and  $\langle v, w \rangle$  the inner product of two such vectors.
- If  $N$  is a positive integer, we denote by  $C_0^1(\mathbb{R}_+, \mathbb{R}^N)$  the set of mappings from  $\mathbb{R}_+$  to  $\mathbb{R}^N$  which are  $C^1$  at zero.

## II. PRELIMINARIES

### A. Problem statement

The class of nonlinear systems under consideration is described by the following differential equation

$$\dot{x}(t) = Ax(t) + \phi(x(t), u(t)), \quad (1)$$

where  $x \in \mathbb{R}^n$  is the state variable,  $u : \mathbb{R} \rightarrow \mathbb{R}^p$  is a known input,  $A$  is a matrix in  $\mathbb{R}^{n \times n}$  and  $\phi : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^n$  is a  $C^1$  globally Lipschitz (uniformly in the input) function. In other words, the following assumption is made.

**Assumption 1.** *For every  $i, j$  in  $\{1, \dots, n\}$ , there exists a real number  $c_{ij} \geq 0$  such that*

$$\left| \frac{\partial \phi_i}{\partial x_j}(x, u) \right| \leq c_{ij}, \quad \forall (x, u) \in \mathbb{R}^n \times \mathbb{R}^p.$$

The state  $x$  of system (1) is accessible via discrete time measurements given as a sequence of  $m$  dimensional real vectors  $(y_k)_{k \in \mathbb{N}}$  of the form

$$y_k = Cx(t_k), \quad (2)$$

where  $C$  is a real matrix in  $\mathbb{R}^{m \times n}$  and  $(t_k)_{k \in \mathbb{N}}$  is a sequence of positive real numbers defined by  $t_{k+1} = t_k + \delta$ ,  $\delta > 0$  representing the sampling measurement time.

The main objective of this work is to design a global observer for system (1) which gives an estimate  $\hat{x}$  that converges asymptotically to  $x$  from the knowledge of the output  $y_k$  given in (2).

Inspired by [7] and [16], the analysis is restricted to a specific class of continuous-discrete time observers defined by the hybrid system

$$\begin{cases} \dot{\hat{x}}(t) = A\hat{x}(t) + \phi(\hat{x}(t), u(t)), & t \in [t_k, t_{k+1}), \\ \hat{x}(t_k) = \hat{x}(t_k^-) + K(y_k - C\hat{x}(t_k^-)), \end{cases} \quad (3)$$

where

$$\hat{x}(t_k^-) = \lim_{t \rightarrow t_k, t < t_k} \hat{x}(t). \quad (4)$$

The estimation problem consists in determining a gain  $K$  such that the estimation error  $e(t) = x(t) - \hat{x}(t)$  converges asymptotically toward zero. The proposed approach is based on a result obtained in [21] which is recalled in the following section.

### B. An approach for continuous time measurements

In [21], the authors addressed the problem of observer design for system (1) where the output is considered as a continuous time function and given by

$$y(t) = Cx(t), \quad \forall t > 0. \quad (5)$$

In the rest of the paper, we extend this procedure when the measurement is discrete in time.

Let  $\mathcal{R} \subset \mathbb{R}^{n \times n}$  be the set of matrices defined as

$$\mathcal{R} = \{R \in \mathbb{R}^{n \times n} \mid R_{ij} = A_{ij} \pm c_{ij}, \quad \forall i, j = 1, \dots, n\}.$$

Note that  $\mathcal{R}$  is composed of  $2^\rho$  elements where  $\rho$  is the number of  $c_{ij} \neq 0$ .

One of the results obtained in [21] can be summarized by the following theorem.

**Theorem II.1** ([21]). *Assume that Assumption 1 is satisfied for system (1). If there exist a symmetric positive definite (SPD) matrix  $P$  in  $\mathbb{R}^{n \times n}$  and a matrix  $L$  in  $\mathbb{R}^{n \times m}$  such that the following matrix inequalities hold:*

$$R'P + PR - C'L' - LC < 0, \quad \forall R \in \mathcal{R}, \quad (6)$$

then the system

$$\dot{\hat{x}}(t) = A\hat{x}(t) + \phi(\hat{x}(t), u(t)) + P^{-1}L(y(t) - C\hat{x}(t)), \quad (7)$$

is an asymptotic observer for system (1) where  $y$  is the continuous time measurement given by (5), i.e.  $\lim_{t \rightarrow +\infty} |\hat{x}(t) - x(t)| = 0$ .

### C. An approach based on reachability analysis

The approach of observer design proposed in this work, is based on a reachable set computation. To develop this approach, let us consider system (1) and observer (3). The estimation error  $e = \hat{x} - x$  is solution to

$$\begin{cases} \dot{e}(t) = Ae(t) + \Delta\phi(\hat{x}(t), u(t), e(t)), & t \in [k\delta, (k+1)\delta), \\ e(k\delta) = (\text{Id} - KC)e(k\delta^-), \end{cases} \quad (8)$$

where the notation (4) is used and  $\Delta\phi$  is the continuous function defined as

$$\Delta\phi(\hat{x}, u, e) = \phi(\hat{x}, u) - \phi(\hat{x} - e, u).$$

Employing the mean value theorem, it yields the existence of  $n$  functions  $z_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $i = 1, \dots, n$  such that the components of the function  $\Delta\phi$  satisfy

$$\Delta\phi_i(\hat{x}, u, e) = \frac{\partial\phi_i}{\partial x}(z_i(\hat{x}, e), u)e, \quad i = 1, \dots, n. \quad (9)$$

Hence, the error  $e(t)$  satisfies the following equation

$$\dot{e}(t) = Ae(t) + V(t)e(t), \quad (10)$$

where  $t \in [k\delta, (k+1)\delta)$  and the elements  $V_{ij}(t) = \frac{\partial \phi_i}{\partial x_j}(z_i(\hat{x}(t), e(t)), u(t))$  of the matrix  $V(t)$  satisfy

$$|V_{ij}(t)| \leq c_{ij}, \quad \forall i, j \in \{1, \dots, n\}. \quad (11)$$

In our approach, we consider the error equation (10) as a bilinear control system where the elements  $V_{ij}$  are bounded control inputs. The proposed observer should converge for all possible values of  $V_{ij}$  satisfying (11). One way to formalize this problem is to introduce the notion of attainable set in finite time.

Let  $\mathcal{A}_\delta(e_0)$  denote the attainable set from  $e_0$  at time  $\delta \geq 0$  of system (10), i.e.,

$$\mathcal{A}_\delta(e_0) = \{e(\delta) \mid \delta \geq 0, e(\cdot) \text{ is solution to (10) with } e(0) = e_0\}.$$

The following theorem uses the set  $\mathcal{A}_\delta(e)$  to give a condition (formulated as LMIs) guaranteeing the convergence of a continuous-discrete observer (3) for system (1)-(2).

**Theorem II.2** (Sufficient condition for observer design). *Let Assumption 1 hold for system (1) and let  $\delta$  be the sampling measurement time. Assume that there exist a finite set  $\mathcal{M}$  of matrix functions (mappings from  $\mathbb{R}_+$  into  $\mathbb{R}^{n \times n}$ ), a positive definite matrix  $P$  in  $\mathbb{R}^{n \times n}$  and a vector  $W$  in  $\mathbb{R}^{n \times m}$  such that <sup>1</sup>*

$$\mathcal{A}_\delta(e) \subset \text{Conv}\{M(\delta)e, M \in \mathcal{M}\}, \quad \forall e \in \mathbb{R}^n, \quad (12)$$

and

$$\begin{bmatrix} P & M(\delta)'(P - C'W') \\ (P - WC)M(\delta) & P \end{bmatrix} > 0, \quad \forall M \in \mathcal{M}. \quad (13)$$

Then, for  $K = P^{-1}W$ , the estimation error given by the observer (3) converges asymptotically to zero.

*Proof.* For  $M \in \mathcal{M}$  define  $N_M = (\text{Id} - KC)M$ . On the one hand, equation (13) and the Schur complement leads to

$$N_M'PN_M - P < 0, \quad \forall M \in \mathcal{M}.$$

Hence, for every  $M \in \mathcal{M}$  we have

$$N_M'PN_M \leq (1 - \kappa_M)P,$$

<sup>1</sup>Conv denotes the convex closure.

where  $0 < \kappa_M < 1$  is defined as

$$\kappa_M = \frac{\lambda_{\min}(P - N'_M P N_M)}{\lambda_{\max}(P)},$$

with  $\lambda_{\max}(\cdot)$  and  $\lambda_{\min}(\cdot)$  being the largest and the smallest eigenvalue respectively. Set  $\kappa = \min_{M \in \mathcal{M}} \{\kappa_M\}$ . Since  $\mathcal{M}$  is finite,  $\kappa \in (0, 1)$ . Consequently,

$$N'_M P N_M \leq (1 - \kappa)P, \quad \forall M \in \mathcal{M}.$$

On the other hand, take  $k \in \mathbb{N}$ . The error  $e(\cdot)$  being solution of (10), it follows from Assumption 1 and the inclusion (12) that

$$e(t_{k+1}^-) \in \text{Conv}\{M(\delta)e(t_k), M \in \mathcal{M}\}.$$

With the discrete dynamics of the error in (8), it yields

$$e(t_{k+1}) \in \text{Conv}\{N_M(\delta)e(t_k), M \in \mathcal{M}\}.$$

In other words, there exist  $M_1, \dots, M_\ell$  in  $\mathcal{M}$  and  $\alpha_1, \dots, \alpha_\ell$  in  $\mathbb{R}_+$ , with  $\sum_{i=1}^\ell \alpha_i = 1$  such that

$$e(t_{k+1}) = \sum_{i=1}^\ell \alpha_i N_{M_i} e(t_k).$$

We have

$$\begin{aligned} e(t_{k+1})' P e(t_{k+1}) &= \sum_{i=1}^\ell \sum_{j=1}^\ell \alpha_i \alpha_j e(t_k)' N'_{M_i} P N_{M_j} e(t_k) \\ &\leq \frac{1}{2} \sum_{i=1}^\ell \sum_{j=1}^\ell \alpha_i \alpha_j e(t_k)' \left[ N'_{M_i} P N_{M_i} \right. \\ &\quad \left. + N'_{M_j} P N_{M_j} \right] e(t_k) \\ &= \sum_{i=1}^\ell \sum_{j=1}^\ell \alpha_i \alpha_j e(t_k)' N'_{M_i} P N_{M_i} e(t_k) \\ &\leq (1 - \kappa) e(t_k)' P e(t_k) \\ &\leq (1 - \kappa)^{k+1} e(0)' P e(0). \end{aligned} \tag{14}$$

According to (10), we have for all  $t \in [t_{k+1}, t_{k+2})$ ,

$$\overline{e(t)' P e(t)} = e(t)' (P(A + V(t)) + (A + V(t))' P) e(t).$$



Assumption 1 implies that for all  $t$ ,

$$\sum_{i=1}^n (A_{ij} + V_{ij}(t))e(t) \in \text{Conv}\{Re(t), R \in \mathcal{R}\}.$$

Thus,  $\overline{e(t)'Pe(t)} \leq \eta e(t)'Pe(t)$ , for all  $t \in [t_{k+1}, t_{k+2})$ , where  $\eta$  is the positive real number defined as

$$\eta = \frac{\max_{R \in \mathcal{R}} \{\lambda_{\max}(PR + R'P), 0\}}{\lambda_{\min}(P)}.$$

Using (14), the following holds

$$e(t)'Pe(t) \leq (1 - \kappa)^{k+1} \exp(\eta\delta) e(0)'Pe(0),$$

$$\forall t \in [t_{k+1}, t_{k+2}).$$

Since  $0 < \kappa < 1$ , the function  $t \mapsto e(t)'Pe(t)$  goes to zero as  $t$  goes to infinity. The function  $e \mapsto e'Pe$  being proper and positive definite, it follows that  $e(t)$  converges to zero which ends the proof. ■

#### D. Some remarks on the approach of Theorem II.2

The first step of the proposed approach is the computation of an approximation of the attainable set  $\mathcal{A}_\delta(e)$  for a bilinear control system. Reachability analysis has received numerous attentions in the literature; for instance, in [5], the author analyzes the geometry of the reachable set of bilinear systems. In [19], the author gave sufficient conditions guaranteeing that the reachable set of a bilinear controllable system is convex. If some results on the characterization of the reachable set are now available for low dimension systems (see for instance the recent result in [14]) its characterization is still an open problem in general.

However, the novelty of the studied problematic is that the exact computation of this set is not needed. As a matter of fact, only an upper approximation in terms of the matrix functions  $M \in \mathcal{M}$  as expressed in (12) is required. As it will be seen in the remaining part of the paper, for uniformly observable systems, an upper approximation can be explicitly given. Hence, for these two classes of systems our observer design strategy can be performed.

Given the matrix function set  $\mathcal{M}$ , the second step of the design is to solve the linear matrix inequality (13). In fact, the usual detectability property is embedded in this inequality. For instance, it is a necessary condition that the couple  $(\exp(A\delta), C)$  is detectable for this inequality

to have a solution. Note however that inequality (13) is much stronger than this detectability condition since all Lipschitz nonlinearities have to be taken into account.

Moreover, note that this condition is not necessary for the existence of a continuous discrete observer. Indeed, as in all high-gain based designs, the nonlinearities are considered in the design as *disturbances*. For instance, the trivial Lipschitz system with no measurement defined on  $\mathbb{R}$  as

$$\dot{x} = -\frac{x}{1+|x|}, \quad y = 0, \quad (15)$$

defines a contraction (see [13]) and consequently a simple copy of the system

$$\dot{\hat{x}} = -\frac{\hat{x}}{1+|\hat{x}|},$$

defines an asymptotic state observer and fits in the class of observers considered in this note (the impulsive correction is simply made with  $K = 0$ ). However, our approach cannot apply to system (15) since the associated linear part (which is null) is not detectable.

In fact, it can be shown that if Theorem II.1 applies for the continuous time measurement case, then for a small sampling measurement time, the proposed approach can be applied provided that the elements of the matrix functions set  $\mathcal{M}$  satisfy some local properties. Indeed, the link between the two matrix inequalities (6) and (13) can be expressed as follows.

**Proposition II.1** (Local properties of matrices in  $\mathcal{M}$ ). *Assume that there exist  $P$  and  $L$  such that the matrix inequality (6) holds for a given set of matrices  $\mathcal{R}$ . If the set  $\mathcal{M}$  of matrix functions is such that every function  $M \in \mathcal{M}$  is  $C^1$  at time  $t = 0$  and satisfies*

$$M(0) = \text{Id}, \quad \dot{M}(0) \in \mathcal{R}, \quad (16)$$

*then, for all  $\delta$  small enough, the matrix inequality (13) is satisfied with the same matrix  $P$  and for  $W = \delta L$ .*

*Proof of Proposition II.1.* Let  $M$  be in  $\mathcal{M}$  and let  $S$  be the matrix defined by,

$$S = M'(\text{Id} - P^{-1}WC)'P(\text{Id} - P^{-1}WC)M - P.$$

We have

$$\begin{aligned} S &= (M - \text{Id})'PM + P(M - \text{Id}) \\ &\quad + M'(-WC - C'W' + C'W'P^{-1}WC)M. \end{aligned}$$

Taking  $W = \delta L$  yields

$$\frac{S(\delta)}{\delta} = \frac{M(\delta)' - \text{Id}}{\delta} PM(\delta) + P \frac{M(\delta) - \text{Id}}{\delta} + M(\delta)'(-LC - C'L' + \delta C'L'P^{-1}LC)M(\delta),$$

which, using the fact that  $M$  is a  $C^1$  matrix function and that  $M(0) = \text{Id}$ , implies

$$\lim_{\delta \rightarrow 0} \frac{S(\delta)}{\delta} = P\dot{M}(0) + \dot{M}(0)'P - LC - C'L'.$$

Note that with (16) and the fact that (6) is satisfied for all matrices in  $\mathcal{R}$ , it yields that  $\lim_{\delta \rightarrow 0} S(\delta)/\delta < 0$  for all  $M \in \mathcal{M}$ .

This implies that for all sufficiently small positive  $\delta$  the inequality  $S(\delta) < 0$  holds. With the Schur complement, it shows that the matrix inequality (13) is satisfied for sufficiently small  $\delta$ . ■

### III. SOME SPECIFIC STRUCTURES ON THE SYSTEM

From Proposition II.1, it follows that a good upper approximation of the reachable set  $\mathcal{A}_\delta(e_0)$  in terms of the matrix functions  $M \in \mathcal{M}$  should be those which satisfy (16). As it will be seen in this section, this is indeed the case when considering some particular class of Lipschitz nonlinear systems: The class of feedforward systems is considered in Subsection III-A. Uniformly observable systems are considered in Subsection III-B.

#### A. Case of feedforward systems

The approach presented in the previous section can be applied when considering a specific class of Lipschitz nonlinear systems (1). Indeed, when the matrix  $A$  and the function  $\phi$  have an upper triangular structure, the computation of the compact set  $\mathcal{M}$  of matrix functions involved in the procedure, can be made explicitly.

In this subsection we consider the case in which the matrix  $A$  and the function  $\phi$  satisfies the following assumption.

**Assumption 2** (Feedforward systems). *The matrices  $A$  and  $V$  in (10) are upper triangular. In other words,  $A_{ij} = c_{ij} = 0$  for all  $i, j \in \{1, \dots, n\}$  such that  $j < i$ .*

Under Assumption 2, the following result can be stated.

**Theorem III.1** (Case of feedforward systems). *Consider system (10) with control constraint (11) where the matrix  $A$  and the positive real numbers  $c_{ij}$  satisfy Assumption 2. Then, there exists a set  $\mathcal{M} \subset C_0^1(\mathbb{R}_+, \mathbb{R} \times \mathbb{R})$  of matrix functions such that for all  $\delta \geq 0$ , the inclusion (12) is satisfied. Moreover, the conditions (16) are satisfied for all  $M$  in  $\mathcal{M}$ .*

Theorem III.1 allows to construct an observer which estimates asymptotically the system state if the matrix inequality (13) is satisfied. Since conditions (16) are satisfied for all  $M$  in  $\mathcal{M}$ , we know from Proposition II.1 that this matrix inequality has a solution provided that the LMI (6) has one and  $\delta$  is sufficiently small. Note however that as it will be seen in the proof of Theorem III.1 the set  $\mathcal{M}$  may be composed of  $2^{\frac{n(n+1)}{2}}$  matrices. This implies that the associated linear matrix inequality may be of very high dimension. Hence, this approach may fail to be applied for high dimensional systems. which can be described by the following lemma the proof of which is given in Appendix A at the end of the paper.

**Lemma III.1** (Iterative design of  $\mathcal{M}$ ). *Let  $\mathcal{M}^{n_y}$  be a subset of  $C_0^1(\mathbb{R}_+, \mathbb{R}^{n_y \times n_y})$  such that  $M^*(0) = \text{Id}$  for all  $M^*$  in  $\mathcal{M}^{n_y}$ . Consider the following control system with  $(x, y)$  in  $\mathbb{R} \times \mathbb{R}^{n_y}$  defined for all positive time  $t$  by:*

$$\dot{x}(t) = u(t)x(t) + \langle v(t), y(t) \rangle, \quad (u, v) \in \mathcal{U} \times \Omega, \quad (17)$$

$$y(t) \in \text{Conv}\{M^*(t)y_0, M^* \in \mathcal{M}^{n_y}\}, \quad (18)$$

$$x(0) = x_0 \quad (19)$$

where  $\mathcal{U} = [u_{\min}, u_{\max}] \subset \mathbb{R}$  and  $\Omega = \prod_{j=1}^{n_y} [v_{j \min}, v_{j \max}] \subset \mathbb{R}^{n_y}$  are the control sets. Then, there exists a set of  $C^1$  matrix functions  $\mathcal{M}^{1+n_y} \subset C^1(\mathbb{R}_+, \mathbb{R}^{(1+n_y) \times (1+n_y)})$  such that,

1) for all positive time  $t$ ,

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \in \text{Conv} \left\{ M(t) \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}, M \in \mathcal{M}^{1+n_y} \right\}.$$

2) for every  $M$  in  $\mathcal{M}^{1+n_y}$ ,  $M(0) = \text{Id}$ .

3) for every  $M$  in  $\mathcal{M}^{1+n_y}$ , there exist  $M^*$  in  $\mathcal{M}^{n_y}$  and  $(u^*, v^*) \in \partial(\mathcal{U} \times \Omega)$ <sup>2</sup> such that

$$\dot{M}(0) = \begin{bmatrix} u^* & v^* \\ 0 & \dot{M}^*(0) \end{bmatrix}. \quad (20)$$

<sup>2</sup> $\partial(\mathcal{U} \times \Omega)$  denotes the boundary of  $\mathcal{U} \times \Omega$

Using this lemma, we give below the iterative procedure for the construction of  $\mathcal{M}$  and prove Theorem III.1.

*Proof of Theorem III.1.* We consider the error system (10) with Assumption 2.

**Step 0: initialization step.** At the first step of the iterative procedure, the dynamics of the last component  $e_n$  of the error in system (10) is considered. According to Assumption 2, this component satisfies

$$\dot{e}_n = [A_{nn} + V_{nn}]e_n,$$

with the control constraint  $|V_{nn}| \leq c_{nn}$ . The solutions of this system are given by

$$e_n(t) = \exp \left( \int_0^t (A_{nn} + V_{nn}(s)) ds \right) e_n(0).$$

Hence,

$$e_n(t) \in \text{Conv} \left\{ M(t)e_n(0), M \in \mathcal{M}^1 \right\},$$

where  $\mathcal{M}^1 = \{\exp([A_{nn} - c_{nn}]t), \exp([A_{nn} + c_{nn}]t)\}$ . Note that, for  $M \in \mathcal{M}^1$ ,

$$M(0) = \text{Id}, \quad \dot{M}(0) \in \{A_{nn} - c_{nn}, A_{nn} + c_{nn}\}.$$

**Step  $\ell$  ( $\ell \leq n$ ):** Apply Lemma III.1 with

$$\begin{aligned} x &= e_\ell, \quad y = (e_{(\ell+1)}, \dots, e_n), \\ u &= A_{\ell\ell} + V_{\ell\ell}, \quad v = (A_{\ell(\ell+1)} + V_{\ell(\ell+1)}, \dots, A_{\ell n} + V_{\ell n}), \\ \mathcal{U} &= [A_{\ell\ell} - c_{\ell\ell}, A_{\ell\ell} + c_{\ell\ell}], \quad \Omega = \prod_{j=1}^{n-\ell} [A_{\ell(\ell+j)} - c_{\ell(\ell+j)}, A_{\ell(\ell+j)} + c_{\ell(\ell+j)}]. \end{aligned}$$

It is also assumed that the previous step yield the existence of a set of matrix functions  $\mathcal{M}^{n-\ell}$  such that (18) is satisfied ( $n_y = n - \ell$ ). According to Assumption 2 equation (17) is satisfied. Consequently, Lemma III.1 and the structure of the matrix functions in (20) yield the existence a set of matrix functions  $\mathcal{M}^{1+(n-\ell)}$  which satisfies items 1, 2 and 3 (of Lemma III.1).

Finally, at Step  $n$  of this iterative design, the result is obtained. ■

### B. Case of system in observability canonical form

Another context of interest is the one in which the nonlinear model (1) is given in observability canonical form. In this section, we assume the following assumption.

**Assumption 3** (Observability canonical systems). *The matrix  $A$  and the positive real number  $c_{ij}$  are such that*

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \\ 0 & \dots & \dots & 0 & 1 \\ 0 & \dots & \dots & \dots & 0 \end{bmatrix},$$

$$c_{ij} = 0, \quad \forall (i, j) \in \{1, \dots, n-1\} \times \{1, \dots, n\}.$$

In this particular context, we have  $V'(t) = (0, \dots, 0, v(t))$ , where  $v = (v_1, \dots, v_n)$ . Consequently, we may rewrite system (10) with  $c_j := c_{nj}$  as

$$\begin{cases} \dot{e}_j = e_{j+1}, & j = 1, \dots, n-1 \\ \dot{e}_n = \langle v, e \rangle, & v \in \prod_{j=1}^n [-c_j, c_j], \end{cases} \quad (21)$$

1) *Statement of the main result:* In this framework, the following result is established.

**Theorem III.2** (Case of observability canonical systems). *Let Assumptions 1 and 3 hold for system (1). Then, there exist  $\delta^* > 0$  and a set  $\mathcal{M} \subset C_0^1(\mathbb{R}_+, \mathbb{R}^{n \times n})$  of matrix functions such that for all  $\delta \in [0, \delta^*]$ , the inclusion (12) is satisfied. Moreover, the conditions (16) are satisfied for all  $M$  in  $\mathcal{M}$ .*

Before proving Theorem III.2 let us make a few remark.

As Theorem III.1, Theorem III.2 allows to construct an observer which estimates asymptotically the system state if the matrix inequality (13) is satisfied. Since conditions (16) are satisfied for all  $M$  in  $\mathcal{M}$  we know from Proposition II.1 that this matrix inequality has a solution provided that the LMI (6) has one and  $\delta$  is sufficiently small.

Note that when  $C = [1, \dots, 0]$ , (6) has a solution which is the well-known high-gain observer (see [8]). Hence, Theorem III.2 result gives a continuous discrete version of the high-gain

observer. This result is not new since the system is in observable form and based on the results presented in [16] we know that there exists an observer for sampling-time  $\delta$  small enough. Moreover, not all systems considered in [16] can be addressed since general lower triangular nonlinearities are not allowed.

The interest of Theorem III.2 lies in its constructive proof and in the fact that the approach is not based on high-gain technics. Hence, we expect that  $\delta$  may be chosen larger than the one allowed employing the high-gain approach of [16] (see [2] for a study of the limitation of the usual high-gain approach).

However, this approach may be difficult to apply due to the fact that for systems of high dimension ( $n$  large), the set  $\mathcal{M}$  may have  $2^{n^2}$  elements.

2) *Preliminaries for the proof of Theorem III.2:* The idea of the proof of Theorem III.2 is to approximate the attainable set  $\mathcal{A}_\delta(e)$  by an  $n$ -dimensional rectangle (a direct product of  $n$  intervals). This approximation is obtained solving some optimal control problems.

Indeed, the next proposition which is proved employing Pontryagin's maximum principle gives a way to build such a rectangle from the solutions of the two following Lipschitz dynamical systems:

$$\dot{e} = F^+(e), \quad F^+(e) = \left[ e_2, \dots, e_n, \sum_{j=1}^n c_j |e_j| \right]', \quad (22)$$

$$\dot{e} = F^-(e), \quad F^-(e) = \left[ e_2, \dots, e_n, -\sum_{j=1}^n c_j |e_j| \right]'. \quad (23)$$

Let  $e^+(e_0, \cdot)$  (resp.  $e^-(e_0, \cdot)$ ) denote the solution of (22) (resp. (23)) emanating from a point  $e_0$  at time 0 and let  $T(e_0, \delta)$  be the  $n$ -dimensional rectangle defined by

$$T(e_0, \delta) = \prod_{j=1}^n [e_j^-(e_0, \delta), e_j^+(e_0, \delta)] \subset \mathbb{R}^n.$$

**Proposition III.2.** *There exists  $\delta^* > 0$  such that for every  $e_0 \in \mathbb{R}^n$  and every  $\delta \in [0, \delta^*]$ ,  $\mathcal{A}_\delta(e_0)$  is contained in  $T(e_0, \delta)$ .*

Proposition III.2 establishes that the solutions to systems (22) and (23) are bounding state trajectories for system (21). Note that this result can be related to Müller's theorem (see [15])

but is however different and has the advantage to be more realizable in practice since the bounding systems are autonomous.

The proof of Proposition III.2 is obtained solving  $2n$  Mayer problems of the form:

$$\begin{cases} \text{Minimize } \langle a, x(T) \rangle \\ \text{Subject to } \dot{x} = f(x, v), \quad v \in \Omega \\ x(0) = x_0 \end{cases} \quad (24)$$

where  $a \in \mathbb{R}^n$  is a constant vector,  $f$  is a smooth functions, and  $\Omega \subset \mathbb{R}^n$  is compact.

The main tool to solve the Mayer problem (24) is the Pontryagin's Maximum Principle (PMP) arising in optimal control. We refer the reader to [20] for a general version of PMP adapted to the resolution of Mayer problems. The next theorem is a version of PMP that we state in our own context only.

**Theorem III.3** (PMP for problem (24)). *Consider the control system (21). Let  $H : \mathbb{R}^n \times \mathbb{R}^n \times \Omega \rightarrow \mathbb{R}$  denote the control dependent Hamiltonian function defined by  $H(x, p, v) = \langle p, f(x, v) \rangle$ . If the pair  $(x(\cdot), v(\cdot)) : [0, T] \rightarrow \mathbb{R}^n \times \Omega$  is a (local) minimizer for problem (24), then there exist a Lipschitz covector  $p(\cdot) : t \in [0, T] \mapsto p(t) \in \mathbb{R}^n$  and a constant  $\lambda \leq 0$  such that the pair  $(p(\cdot), \lambda)$  is never trivial and for a.e.  $t \in [0, T]$ :*

- i.  $\dot{x}(t) = \frac{\partial H}{\partial p}(x(t), p(t), v(t)),$
- ii.  $\dot{p}(t) = -\frac{\partial H}{\partial x}(x(t), p(t), v(t)),$
- iii.  $H(x(t), p(t), v(t)) = \max_{v \in \Omega} H(x, p, w),$
- iv.  $p(T) = \lambda a.$

**Remark III.3.** • *The PMP is just a necessary condition for optimality. A trajectory  $x(\cdot)$  (resp. a pair  $(x(\cdot), p(\cdot))$ ) satisfying the conditions given by the PMP is said to be an extremal (resp. an extremal pair). An extremal corresponding to  $\lambda = 0$  is said to be an abnormal extremal, otherwise we call it a normal extremal.*

- *Notice that a Mayer problem of the form (24) do not admit abnormal extremals. Indeed, otherwise, due to the transversality condition iv and equation ii (which is linear), the pair  $(p(t), \lambda)$  would be trivial (everywhere) which contradicts the PMP. Consequently,  $p(t)$  is never zero as soon as the constant vector  $a \neq 0$ .*



We are now ready to prove Proposition III.2.

*Proof of Proposition III.2.* First of all, notice that the result is obvious for  $e_0 = 0$  since for every  $\delta$  the attainable set of system (21) at  $e_0 = 0$  is  $\mathcal{A}_\delta(0) = \{0\}$ . Suppose from now that  $e_0 \neq 0$ .

Let  $a_1, \dots, a_n$  denote the canonical basis vectors of  $\mathbb{R}^n$ , and let  $\Omega = \prod_{j=1}^n [-c_j, c_j]$ .

The proof is based upon the resolution of the  $2n$  Mayer's problems  $(\mathbf{P}_{k,\epsilon,\delta})$  with  $k \in \{1, \dots, n\}$ ,  $\epsilon \in \{-, +\}$ ,  $\delta > 0$  defined as follows:

$$(\mathbf{P}_{k,\epsilon,\delta}) \quad \begin{cases} \text{Minimize } \langle \epsilon a_k, e(\delta) \rangle \\ \text{Subject to (21)} \\ e(0) = e_0 \end{cases}$$

We note that the solution of each of these problems allow us to find the minimum and maximum of each component of the solutions of the error system (21). Consequently, to show Proposition III.2, it suffices to prove the existence of a  $\delta^* > 0$  such that, for every  $\delta < \delta^*$ , and every  $k = 1, \dots, n$  we have

$$\min_{v(\cdot) \in \mathbb{L}^\infty(\mathbb{R}_+, \Omega)} \langle \epsilon a_k, e(\delta) \rangle = e_k^\epsilon(\delta), \quad (25)$$

where  $e_k^\epsilon(\delta)$  is the solution of (22) or (23) depending on the value of  $\epsilon$ . The rest of the proof is devoted to the resolution of the  $2n$  optimal control problems  $(\mathbf{P}_{k,\epsilon,\delta})$  to prove the existence of such a  $\delta^*$ .

Let us apply Theorem III.3 to problem  $(\mathbf{P}_{k,\epsilon,\delta})$ . The control dependent Hamiltonian associated to (21) reads

$$H(e, p, v) = \sum_{j=1}^{n-1} p_j e_{j+1} + p_n \sum_{j=1}^n e_j v_j, \quad p \in \mathbb{R}^n.$$

Assume that  $(e(\cdot), v(\cdot))$  is an extremal pair associated with the minimization problem  $(\mathbf{P}_{k,\epsilon,\delta})$ , then, according to Theorem III.3, for a.e.  $t \in [0, \delta]$ : the adjoint system reads

$$\begin{cases} \dot{p}_1 = -p_n v_1 \\ \dot{p}_j = -p_{j-1} - p_n v_j, \quad j = 2, \dots, n, \end{cases} \quad (26)$$

and the maximality reads

$$\begin{aligned} H(e(t), p(t), v(t)) &= \max_{w \in \Omega} H(e(t), p(t), w) \\ &= \sum_{j=1}^{n-1} p_j(t) e_{j+1}(t) + \sum_{j=1}^n c_j |p_n(t) e_j(t)|. \end{aligned} \quad (27)$$

Moreover, the following transversality condition holds:

$$p(\delta) = -\epsilon a_k. \quad (28)$$

Notice that, according to Remark III.3, the adjoint vector  $p(\cdot)$  is nontrivial. Hence,  $p_n(\cdot)$  cannot vanish on an interval. Indeed, otherwise, since (26), we get also  $p_{n-1}(t) = -\dot{p}_n(t) = 0$  for every  $t$  in this interval. Then iteratively  $p(\cdot) = 0$  in this interval.

Similarly, for  $j = 1, \dots, n$ ,  $e_j(\cdot)$  cannot vanish on an interval. Indeed, otherwise, from (21) we get also  $e_{j+1}(t) = \dot{e}_j(t) = 0$  for  $t$  in this interval and iteratively, we obtain  $e_\ell(\cdot) = 0$  for  $\ell = j, \dots, n$ . If  $j \geq 2$ , we have  $e_1^{(j)}(\cdot) = 0$  and therefore  $e_1, \dots, e_{j-1}$  are polynomials of degree  $j-2, \dots, 0$  respectively. On the other hand, in this interval we have  $\dot{e}_n(t) = \sum_{\ell=1}^{j-1} v_\ell(t) e_\ell(t) = 0$ . Thus, with (27) we get that, for almost all  $t$  in this interval

$$p_n(t) \sum_{\ell=1}^{j-1} e_\ell(t) v_\ell(t) = |p_n(t)| \sum_{\ell=1}^{j-1} c_\ell |e_\ell(t)| = 0.$$

Now, since the  $e_j(\cdot)$ s are polynomials of different degrees and  $p_n(\cdot)$  cannot vanish on an interval, we conclude that  $e_1(t) = \dots = e_{j-1}(t) = 0$  for a.e.  $t$ . Consequently,  $e(t) = 0$ , which is a contradiction with  $e_0 \neq 0$  since the system (21) is linear.

Summing up, we obtain that  $p_n(\cdot)e_j(\cdot)$ ,  $j = 1, \dots, n$  cannot vanish on an interval, which implies that for almost all  $t$  in  $\mathbb{R}_+$

$$v_j(t) = |c_j| \text{sign}(p_n(t)e_j(t)).$$

Note that if we show that there exists a time  $\delta^* > 0$  such that  $p_n$  does not change its sign if  $t < \delta^*$ , then with the transversality condition we get that

$$v_j(t) = |c_j| \text{sign}(p_n(\delta^*)) \text{sign}(e_j(t)) = -\epsilon |c_j| \text{sign}(e_j(t)) \quad \forall t \in [0, \delta^*]. \quad (29)$$

In this case equation (25) is thus obtained and the proof of Proposition (III.2) is complete. Consequently, to finish the proof it remains to show that  $p_n(\cdot)$  does not change its sign if  $t$  is small enough. This property is obtained from the following technical lemma, the proof of which is given in Appendix B.

**Lemma III.4.** *There exist a  $\delta_1$  positive and a positive coefficient  $d$  such that for every  $t$  in  $[0, \delta_1]$  and for every  $\delta > 0$ , we have  $|p_n(\delta - t)| < d t^{n-k}$ .*

From Lemma III.4, (26) and using the fact that  $|v_1(t)| \leq c_1$  we get,

$$|\dot{p}_1(\delta - t)| \leq c_1 d t^{n-k}, \quad \forall t \in [0, \delta_1].$$

Integrating this inequality, we obtain

$$|p_1(\delta - t)| \leq \frac{c_1 d}{n - k + 1} t^{n-k+1} + |p_1(\delta)|, \quad \forall t \in [0, \delta_1].$$

By integrating successively the previous inequality, and using the transversality conditions (28), it yields that there exists a positive number  $d_{k-1}$  such that:

$$|p_{k-1}(\delta - t)| \leq d_{k-1} t^{n-k+1}, \quad \forall t \in [0, \delta_1]. \quad (30)$$

On the other hand, we have the following inequality for all  $t$  in  $[0, \delta_1]$ :

$$p_k(\delta - t) \geq p_k(\delta) - \int_0^t |p_{k-1}(\delta - s)| + c_k |p_n(\delta - s)| ds.$$

From (28) and (30), it follows that for  $t$  in  $[0, \delta_1]$ :

$$p_k(\delta - t) \geq 1 - d_{k-1}(\delta - t)^{n-k+2} - c_k d t^{n-k} \geq 1 - d_k t^{n-k},$$

where  $d_k$  is positive. Now, by integrating the previous inequality, we obtain for all  $t$  in  $[0, \delta_1]$ :

$$p_{k+1}(\delta - t) \geq t - d_k t^{n-k+1} - c_{k+1} d t^{n-k} \geq t - d_{k+1} t^{n-k},$$

where  $d_{k+1}$  is positive. Proceeding in the same manner successively, we obtain a positive constant  $d_{n-1}$  such that

$$p_{n-1}(\delta - t) \geq t^{n-k-1} - d_{n-1} t^{n-k}.$$

Comparing the two degrees of monomials of the last equation, we obtain the existence of a time  $\delta_2$  in  $[0, \delta_1]$  such that

$$p_{n-1}(\delta - t) > 0, \quad \forall t \in [0, \delta_2].$$

Finally, we have  $\dot{p}_n \geq v_n p_n + p_{n-1}$  and  $p_n(\delta) = 0$ . Thus, we get for every  $t$  in  $[0, \delta_2]$ ,

$$p_n(\delta - t) = \int_\delta^{\delta-t} \exp\left(\int_s^{\delta-t} v_n(r) dr\right) p_{n-1}(s) ds > 0.$$

Proceeding in the same way, one infers that there exists a sufficiently small time  $\delta^*$ , such that for every  $t$  in  $[0, \delta^*]$  and every  $k \in \{1, \dots, n\}$ ,  $p_n(\delta - t)$  is positive (resp. negative) if  $\epsilon = -$  (resp.  $\epsilon = +$ ). From the structure of minimizing control, given in (29), it is concluded that  $\delta < \delta^*$ . Hence,  $v_k(t) = -\epsilon |e_k(t)|$ . ■

3) *Proof of Theorem III.2:* With the approximation of the attainable set  $\mathcal{A}_\delta(e_0)$  by the rectangle  $T(e_0, \delta)$  (see Proposition III.2), we can now give the proof of Theorem III.2. In order to do that, let us define the Clarke's gradient in the direction of  $x_j$  (denoted by  $\partial_C^j f(x)$ ) of a vector function  $f$  to be the *generalized gradient*<sup>3</sup> of the function  $x_j \mapsto f(x_1, \dots, x_n)$ .

Since (22) and (23) are globally Lipschitz,  $e^+(e_0, \delta)$  and  $e^-(e_0, \delta)$  are globally Lipschitz<sup>4</sup>. Consequently, the Clarke gradients  $\partial_C^j e_i^+(e, \delta)$  and  $\partial_C^j e_i^-(e, \delta)$  exist for every  $i = 1, \dots, n$ . Let  $\mathbb{S}$  denote the unit sphere of  $\mathbb{R}^n$ . Introduce the two functions

$$m_{ij}^-(\delta) = \min \bigcup_{\nu \in \mathbb{S}} \min \{ \partial_C^j e_i^+(\nu, \delta), \partial_C^j e_i^-(\nu, \delta) \},$$

$$m_{ij}^+(\delta) = \max \bigcup_{\nu \in \mathbb{S}} \max \{ \partial_C^j e_i^+(\nu, \delta), \partial_C^j e_i^-(\nu, \delta) \}.$$

Let  $\mathcal{M}$  be the set of  $2^{n^2}$  matrices  $M$  given by

$$M_{ij}(\delta) \in \{m_{ij}^-(\delta), m_{ij}^+(\delta)\}. \quad (31)$$

To complete the proof of Theorem III.2 we have to show the following two properties:

- 1) the conditions (16) are satisfied for all  $M$  in  $\mathcal{M}$ ;
- 2) there exists  $\delta^* > 0$  such that for all  $0 \leq \delta \leq \delta^*$ , the inclusion (12) is satisfied.

**About the conditions (16).** First of all, note that since  $e_i^-(\nu, 0) = e_i^+(\nu, 0) = \nu$ , it yields that  $M(0) = \text{Id}$ . For the local property of the time derivative, let  $M_{ij}^+(\delta)$  be the function defined by

$$M_{ij}^+(\delta) = \max \bigcup_{\nu \in \mathbb{S}} \partial_C^j e_i^+(\nu, \delta).$$

Note that by definition of the generalized gradient, for each  $\delta$  there exists a sequence of  $(\nu_\ell(\delta))_{\ell \in \mathbb{N}}$  converging toward  $\nu^*(\delta)$  in  $\mathbb{S}$  such that for all  $\ell$  in  $\mathbb{N}$ ,  $\frac{\partial e_i^+}{\partial e_j}(\nu_\ell(\delta), \delta)$  is well defined and

$$M_{ij}^+(\delta) = \lim_{\ell \rightarrow +\infty} \frac{\partial e_i^+}{\partial e_j}(\nu_\ell(\delta), \delta).$$

<sup>3</sup> Following F. Clarke [6], we define the generalized gradient of a scalar function  $f$  at  $x_0$  as the convex envelop of all possible limits of derivatives of  $f$  at points  $x_n \in \mathbb{R}^n$ ,  $x_n \rightarrow x_0$ . Note that, in general,  $\partial_C f(x_0)$  is a set.

<sup>4</sup>To see this, note that for  $(e_a, e_b)$  in  $\mathbb{R}^n$ , we have

$$\begin{aligned} |e^+(e_a, \delta) - e^+(e_b, \delta)| &\leq \int_0^\delta |F^+(e^+(e_a, s)) - F^+(e^+(e_b, s))| ds \\ &\leq \int_0^\delta c_{\max} |e^+(e_a, s) - e^+(e_b, s)| ds \\ &\leq \exp(c_{\max} \delta) |e_a - e_b|. \end{aligned}$$

On another hand, with  $v_j$  the unit vector with 1 at the  $j$ st component, we have for  $(\ell, \delta)$  in  $\mathbb{N} \times \mathbb{R}_+$

$$\frac{\partial e_i^+}{\partial e_j}(\nu_\ell(\delta), \delta) = \lim_{r \rightarrow 0} \frac{e_i^+(\nu_\ell(\delta) + rv_j, \delta) - e_i^+(\nu_\ell(\delta), \delta)}{r}.$$

However, we have for  $(r, \ell, \delta)$  in  $\mathbb{R}_+ \times \mathbb{N} \times \mathbb{R}_+$

$$\begin{aligned} e_i^+(\nu_\ell(\delta) + rv_j, \delta) - e_i^+(\nu_\ell(\delta), \delta) &= r \text{Id}_{ij} \\ &+ \delta [F_i^+(\nu_\ell(\delta) + rv_j) - F_i^+(\nu_\ell(\delta))] + G_i^+(\nu_\ell(\delta), r, \delta), \end{aligned}$$

where

$$\begin{aligned} G_i^+(\nu_\ell(\delta), r, \delta) &= \int_0^\delta F_i^+(e_i^+(\nu_\ell(\delta) + rv_j, s)) \\ &- F_i^+(\nu_\ell(\delta) + rv_j) - F_i^+(e_i^+(\nu_\ell(\delta), s)) \\ &+ F_i^+(\nu_\ell(\delta)) ds. \end{aligned} \quad (32)$$

Note that  $F_i^+$  being a globally Lipschitz vector field with Lipschitz constant denoted  $c_{\max}$ , it yields,

$$\begin{aligned} |G_i^+(\nu_\ell(\delta), r, \delta)| &\leq \int_0^\delta c_{\max} |e_i^+(\nu_\ell(\delta) + rv_j, s) - \nu_\ell(\delta) \\ &- rv_j - e_i^+(\nu_\ell(\delta), s) + \nu_\ell(\delta)| ds, \end{aligned} \quad (33)$$

which gives,

$$\begin{aligned} |G_i^+(\nu_\ell(\delta), r, \delta)| &\leq \int_0^\delta c_{\max} \left| \int_0^s F_i^+(e_i^+(\nu_\ell(\delta) + rv_j, \tau)) \right. \\ &\quad \left. - F_i^+(e_i^+(\nu_\ell(\delta), \tau)) d\tau \right| ds \\ &\leq c_{\max} \int_0^\delta \int_0^s |F_i^+(e_i^+(\nu_\ell(\delta) + rv_j, \tau)) \\ &\quad - F_i^+(e_i^+(\nu_\ell(\delta), \tau))| d\tau ds \\ &\leq c_{\max}^2 \int_0^\delta \int_0^s |e_i^+(\nu_\ell(\delta) + rv_j, \tau) \\ &\quad - e_i^+(\nu_\ell(\delta), \tau)| d\tau ds. \end{aligned}$$

Moreover, note that for all  $\ell$  and  $\delta$ , we have

$$\lim_{r \rightarrow 0} \frac{F_i^+(\nu_\ell(\delta) + rv_j) - F_i^+(\nu_\ell(\delta), \delta)}{r} \leq \max_{\nu \in \mathbb{S}} \bigcup \partial_C^j F_i^+(\nu).$$

This implies,

$$M_{ij}^+(\delta) \leq \text{Id}_{ij} + \delta \max_{\nu \in \mathbb{S}} \bigcup \partial_C^j F_i^+(\nu) + c_L^2 \delta^2 \bar{m}_{ij}^+(\delta),$$

where

$$\bar{m}_{ij}^+(\delta) = \max_{s \in \bigcup_{\nu \in \mathbb{S}} \partial_C^j e_i^+(\nu, \delta)} |s|.$$

This gives finally

$$\begin{aligned} \lim_{\delta \rightarrow 0} \frac{M_{ij}^+(\delta) - M_{ij}^+(0)}{\delta} &\leq \max_{\nu \in \mathbb{S}} \bigcup \partial_C^j F_i^+(\nu) \\ &\leq A_{ij} + c_{ij}. \end{aligned}$$

A similar property may be obtain for the minimum. In other words, we get

$$\lim_{\delta \rightarrow 0} \frac{M_{ij}^+(\delta) - M_{ij}^+(0)}{\delta} \geq A_{ij} - c_{ij}.$$

This implies that

$$\lim_{\delta \rightarrow 0} \frac{M_{ij}^+(\delta) - M_{ij}^+(0)}{\delta} \in \text{Conv} \{R_{ij}, R \in \mathcal{R}\},$$

where  $\mathcal{R}$  is the set of matrices defined in Section II-B. From this we conclude that the set  $\mathcal{M}$  of matrix functions defined in (31) satisfies

$$\dot{M}_{ij}(0) \subset \text{Conv}\{R_{ij}, R \in \mathcal{R}\}.$$

**About the inclusion (12).** To finish the proof of the Theorem, it remains to show that (12) is also satisfied. The systems (22) and (23) being homogeneous of degree zero, their solutions are homogeneous of degree one. We deduce that for all  $\lambda > 0$ , all  $e_0 \in \mathbb{R}^n$ ,

$$e^+(\lambda e_0, \delta) = \lambda e^+(e_0, \delta), \quad e^-(\lambda e_0, \delta) = \lambda e^-(e_0, \delta).$$

Consider a vertice  $S(e_0, \delta)$  of  $T(e_0, \delta)$  and assume without loss of generality that  $S_i(e_0, \delta) = e_i^+(e_0, \delta)$ . Given  $e_0 \neq 0$  in  $\mathbb{R}^n$ , let  $\nu_0 = \frac{e_0}{|e_0|}$  in  $\mathbb{S}$ . We can also introduce  $\nu_\ell$  a sequence in  $\mathbb{R}^n$  such that  $\lim_{\ell \rightarrow +\infty} \nu_\ell = \nu_0$  and  $e^+(\nu_\ell, \delta)$  is  $C^1$  in its first argument. We have

$$e^+(e_0, \delta) = e^+(\nu_0, \delta)|e_0| = \lim_{\ell \rightarrow +\infty} e^+(\nu_\ell, \delta)|e_0|.$$

By the homogeneity property, we have,

$$e_i^+(\nu_\ell, \delta)|e_0| = |e_0| \sum_{j=1}^n \frac{\partial e_i^+}{\partial e_j}(\nu_\ell, \delta)(\nu_\ell)_j.$$

Hence, it yields

$$e_i^+(e_0, \delta) = \lim_{\ell \rightarrow +\infty} \sum_{j=1}^n \frac{\partial e_i^+}{\partial e_j}(\nu_\ell, \delta)(\nu_\ell)_j |e_0| \quad (34)$$

$$= \lim_{\ell \rightarrow +\infty} \sum_{j=1}^n \frac{\partial e_i^+}{\partial e_j}(\nu_\ell, \delta)(e_0)_j. \quad (35)$$

Consequently, from the definition of the set of matrices  $\mathcal{M}$  we obtain

$$\begin{aligned} S_i(e_0, \delta) &= e_i^+(e_0, \delta) \\ &= \lim_{\ell \rightarrow +\infty} \sum_{j=1}^n \frac{\partial e_i^+}{\partial e_j}(\nu_\ell, \delta)(\nu_\ell)_j |e_0| \\ &\in \text{Conv}\{M_i(\delta)e_0, M \in \mathcal{M}\}. \end{aligned}$$

By iterating this procedure on each component of the vertex  $S$ , it yields that

$$S(e_0, \delta) \in \text{Conv}\{M_i(\delta)e_0, M \in \mathcal{M}\}.$$

By iterating this procedure for each vertex of the multidimensional rectangle which is a convex set, finally we obtain

$$T(e_0, \delta) \subset \text{Conv}\{M(\delta)e_0, M \in \mathcal{M}\}.$$

The result is then obtained using Proposition III.2.

#### IV. ILLUSTRATIVE EXAMPLES

We consider the following simple model of a pendulum  $\dot{x}_1 = x_2$ ,  $\dot{x}_2 = \sin x_1$ , where  $x_1, x_2$  denote the angle between the pendulum and the vertical axis and the pendulum speed, respectively, and  $y_k = x_1(t_k)$ ,  $t_k = t_{k-1} + \delta$ . The associate estimation error equation is given by

$$\dot{e}_1 = e_2, \quad \dot{e}_2 = v e_1, \quad v \in [-1, 1].$$

Following the proposed approach, we get that the approximation by the rectangle denoted by  $T(e_0, \delta)$  is possible for  $\delta^* = \sqrt{3} - 1$ . By integrating solutions of systems (22) and (23), we

obtain a set  $\mathcal{M}$  of 64 matrix functions  $M$  satisfying the assumptions of Theorem III.2, where

$$\begin{aligned} M_{11} &\in \{\cos \delta, \cosh \delta\}, & M_{21} &\in \{-\sin \delta, \sinh \delta\}, \\ M_{12} &\in \{\sin \delta, \sinh \delta, \cosh \delta \tan \delta, \tanh \delta \cos \delta\}, \\ M_{22} &\in \left\{ \cos \delta, \cosh \delta, \frac{1+\sinh \delta \sin \delta}{\cos \delta}, \frac{1+\sin \delta \sinh \delta}{\cosh \delta} \right\}. \end{aligned}$$

Note that by considering the minimum and maximum of each element, we can reduce the number of matrices to 16.

We also note that the obtained set of matrices satisfies the local properties (16) of Proposition II.1. The system is in uniformly observable form, so we know that for small values of  $\delta$ , the assumptions of Theorem II.2 are satisfied.

Employing the *Yalmip* package ([12]) in *Matlab* in combination with the solver *Sedumi* ([18]), it can be checked that the LMI (13) is satisfied<sup>5</sup> for  $\delta \leq 0.668$ . The observer gain for  $\delta = 0.668$  is  $K = [-1, -1.8361]'$ . It is interesting to notice that the bound obtained is much larger than the one obtained employing the usual high-gain approach as exposed in [2].

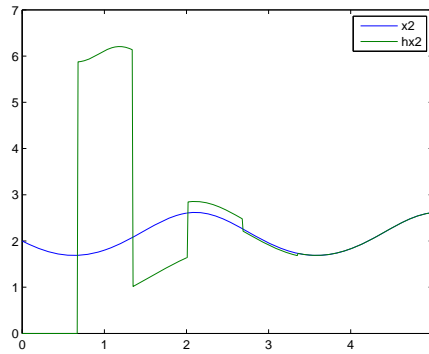


Figure 1. Evolution with time of the state component  $x_2$  and its estimation  $\hat{x}_2$ .

The figure 1 represents a simulation of the observer using an integration algorithm of the model with a semi-implicit integration step of 0.01.

## V. CONCLUSION

In this paper, the problem of designing an observer for nonlinear systems with discrete time measurements and globally Lipschitz nonlinearities is addressed. A solution based on

<sup>5</sup>The Matlab files can be downloaded from <https://sites.google.com/site/vincentandrieu/>



the synthesis of an upper approximation of a reachable set have been presented. When this approximation is given in terms of a convex combination of linear mappings, a sufficient condition of the global convergence of the proposed observer is obtained in terms of a linear matrix inequality. The good performances obtained on an illustrative example demonstrate that the proposed approach is an efficient tool.

## APPENDIX

### A. Proof of Lemma III.1

With the variation of constant method, the solution of the system (17)-(18) is given by:

$$x(t) = \exp\left(\int_0^t u(s)ds\right) x(0) + \int_0^t \exp\left(\int_s^t u(r)dr\right) \langle v(s), y(s) \rangle ds.$$

Let  $\Omega^* = \prod_{j=1}^{n_y} \{v_{j \min}, v_{j \max}\}$ . Let  $\Gamma^*$  denote the set of  $4^{n_y}$  vector functions  $(s, t) \mapsto \gamma(s, t) \in \mathbb{R}^{n_y}$  defined as follows. For each  $\gamma \in \Gamma^*$ , the  $j^{th}$  component of  $\gamma(s, t)$  (denoted by  $\gamma_j(s, t)$ ) satisfies,

$$\gamma_j(s, t) \in \left\{ \min_{(M, v) \in \mathcal{M} \times \Omega^*} \left\{ e^{(t-s)u^1} (v' M)_j(s) \right\}, \right. \\ \left. \max_{(M, v) \in \mathcal{M} \times \Omega^*} \left\{ e^{(t-s)u^2} (v' M)_j(s) \right\} \right\},$$

where  $u^1, u^2 \in \{u_{\min}, u_{\max}\}$ . Note that for every  $j$  in  $\{1, \dots, n_y\}$ , it holds:

$$(v' M)_j(s)(y_0)_j \leq \begin{cases} \left( \max_{(M, v) \in \mathcal{M} \times \Omega^*} \{(v' M)_j(s)\} \right) (y_0)_j \\ \quad \text{if } (y_0)_j \geq 0 \\ \left( \min_{(M, v) \in \mathcal{M} \times \Omega^*} \{(v' M)_j(s)\} \right) (y_0)_j \\ \quad \text{if } (y_0)_j \leq 0, \end{cases}$$

and,

$$(v' M)_j(s)(y_0)_j \geq \begin{cases} \left( \min_{(M, v) \in \mathcal{M} \times \Omega^*} \{(v' M)_j(s)\} \right) (y_0)_j \\ \quad \text{if } (y_0)_j \geq 0 \\ \left( \max_{(M, v) \in \mathcal{M} \times \Omega^*} \{(v' M)_j(s)\} \right) (y_0)_j \\ \quad \text{if } (y_0)_j \leq 0, \end{cases}$$

This implies that, for all  $y_0$  in  $\mathbb{R}^{n_y}$  there exist  $\gamma^1$  and  $\gamma^2$  in  $\Gamma^*$  such that for all  $M \in \mathcal{M}^{n_y}$  and all  $(s, t)$

$$\begin{aligned} \langle \gamma^1(s, t), y_0 \rangle &\leq \exp \left( \int_s^t u(r) dr \right) (v' M)(s) y_0 \\ &\leq \langle \gamma^2(s, t), y_0 \rangle. \end{aligned}$$

Hence,

$$\begin{aligned} \int_0^t \langle \gamma^1(s, t), y_0 \rangle ds &\leq \int_0^t \exp \left( \int_s^t u(r) dr \right) (v' M)(s) y_0 ds \\ &\leq \int_0^t \langle \gamma^2(s, t), y_0 \rangle ds, \end{aligned}$$

which gives,

$$\begin{aligned} \int_0^t \exp \left( \int_s^t u(r) dr \right) \langle v(s), y(s) \rangle ds &\in \text{Conv} \left\{ \int_0^t \langle \gamma(s, t), y_0 \rangle, \gamma \in \Gamma^* \right\}. \end{aligned}$$

Finally,  $\mathcal{M}^{1+n_y}$  is the set of  $2 \times 4^{n_y} \times \# \mathcal{M}^{n_y}$  ( $\# \mathcal{M}^{n_y}$  denotes the cardinal of  $\mathcal{M}^{n_y}$ ) matrix functions defined by

$$\mathcal{M}^{1+n_y} = \left\{ t \mapsto \begin{bmatrix} \exp(tu^*) & \int_0^t \gamma(s, t) ds \\ 0 & M^*(t) \end{bmatrix} \mid M^* \in \mathcal{M}^{n_y}, \gamma \in \Gamma^*, u^* \in \{u_{\min}, u_{\max}\} \right\}.$$

Moreover, notice that for every  $M \in \mathcal{M}^{1+n_y}$

$$\begin{aligned} \dot{M}(0) &\in \left\{ \begin{bmatrix} u^* & \gamma(0, 0) \\ 0 & \dot{M}^*(0) \end{bmatrix} \mid M^* \in \mathcal{M}^{n_y}, \gamma \in \Gamma^*, u^* \in \{u_{\min}, u_{\max}\} \right\}. \end{aligned}$$

With the definition of  $\Gamma^*$ , (20) is satisfied. This concludes the proof.

### B. Proof of Lemma III.4:

At first, we note that because of the transversality condition, we have  $p_n(\delta) = 0$ . The application  $t \mapsto p_n(t)$  being continuous, for every  $\epsilon > 0$  there exists a time  $\delta_1$  such that  $|p_n(\delta - t)| < \epsilon$  for every  $t$  in  $[0, \delta_1]$ .

Now, assume that there exist a  $C^\infty$  function  $\alpha : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  and an integer  $q_\alpha$  such that for every  $t$  in  $[0, \delta_1]$

$$\begin{aligned} |p_n(\delta - t)| &\leq \alpha(t), \\ \alpha^{(\ell)}(0) &= 0, \quad \ell = 0, \dots, q_\alpha - 1, \quad \alpha^{(q_\alpha)}(0) > 0. \end{aligned}$$

Thus, using (26) and the fact that  $|v_1(t)| \leq c_1$  we have  $|\dot{p}_1| \leq c_1|p_n|$  and it is obtained that for every  $t$  in  $[0, \delta_1]$

$$|p_1(\delta - t)| \leq c_1 \int_0^t |p_n(\delta - s)| ds + |p_1(\delta)| \leq c_1 \alpha_1(t) + |p_1(\delta)|,$$

where  $\alpha_1(t) = \int_0^t \alpha(s) ds$ . Similarly, we have  $|\dot{p}_2| \leq |p_1| + c_2|p_n|$  and it yields,

$$|p_2(\delta - t)| \leq c_1 \alpha_2(t) + |p_1(\delta)|t + c_2 \alpha_1(t) + |p_2(\delta)|, \quad \forall t \in [0, \delta_1],$$

where  $\alpha_2(t) = \int_0^t \alpha_1(s) ds$ .

Then, by integrating successively along the trajectories of the system we obtain the inequality for every  $t$  in  $[0, \delta_1]$

$$|p_{n-1}(\delta - t)| \leq \sum_{\ell=1}^{n-1} \left( c_\ell \alpha_{n-\ell}(t) + |p_\ell(\delta)| \frac{t^{n-\ell-1}}{(n-\ell-1)!} \right),$$

where the functions  $\alpha_\ell$ ,  $\ell = 1, \dots, n-1$  are defined by the following iterative procedure:

$$\alpha_\ell(t) = \int_0^t \alpha_{\ell-1}(s) ds, \quad \ell = 1, \dots, n.$$

Using the transversality conditions (28), this inequality becomes

$$|p_{n-1}(\delta - t)| \leq \sum_{\ell=1}^{n-1} c_\ell \alpha_{n-\ell}(t) + \frac{t^{n-k-1}}{(n-k-1)!}, \quad \forall t \in [0, \delta_1],$$

with  $k$  being the one considered in the proof of Proposition III.2. Therefore, using Gronwall's lemma, this gives that the last component and for every  $t$  in  $[0, \delta_1]$ :

$$\begin{aligned} |p_n(\delta - t)| &\leq \int_{\delta-t}^{\delta} e^{c_n(s+t-\delta)} |p_{n-1}(s)| ds, \\ &\leq \beta(t), \end{aligned}$$

where  $\beta : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is the  $C^\infty$  function defined by:

$$\beta(t) = e^{c_n \delta_1} \left( \sum_{\ell=1}^{n-1} \frac{c_\ell}{c_n} \alpha_{n-\ell+1}(t) + \frac{t^{n-k}}{c_n(n-k)!} \right).$$

Note that  $\beta$  satisfies

$$\beta^{(\ell)}(\delta) = 0, \quad \ell = 0, \dots, q_\beta - 1, \quad \beta^{(q_\beta)}(\delta) > 0,$$

with  $q_\beta = \min\{q_\alpha + 2, n - j\}$ . Reiterating the procedure with the function  $\beta$  instead of the function  $\alpha$ , we obtain the existence of a  $C^\infty$  function  $\alpha$  such that for every  $t$  in  $[0, \delta_1]$

$$|p_n(\delta - t)| \leq \alpha(t),$$

$$\alpha^{(\ell)}(0) = 0, \quad \ell = 0, \dots, n - k - 1, \quad \alpha^{(n-k)}(0) > 0.$$

This completes the proof of the lemma.

## REFERENCES

- [1] T. Ahmed-Ali, V. Van Assche, J. Massieu, and P. Dorleans. Continuous-discrete observer for state affine systems with sampled and delayed measurements. *Automatic Control, IEEE Transactions on*, 58(4):1085–1091, 2013.
- [2] V. Andrieu. Extended-Kalman-Filter-like observers for continuous time systems with discrete time measurements. March 2010.
- [3] V. Andrieu and M. Nadri. Observer design for lipschitz systems with discrete-time measurements. In *Proc. of the 49th IEEE Conference on Decision and Control*, pages 6522–6527. IEEE, 2010.
- [4] M. Arcak and D. Nešić. A framework for nonlinear sampled-data observer design via approximate discrete-time models and emulation. *Automatica*, 40(11):1931–1938, 2004.
- [5] R. W. Brockett. On the reachable set for bilinear system. 111:54–63, 1975.
- [6] F. Clarke. *Optimization and nonsmooth analysis*. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons Inc., New York, 1983. A Wiley-Interscience Publication.
- [7] F. Deza, E. Busvelle, JP Gauthier, and D. Rakotopara. High gain estimation for nonlinear systems. *Systems & control letters*, 18(4):295–299, 1992.

- [8] J.-P. Gauthier, H. Hammouri, and S. Othman. A simple observer for nonlinear systems applications to bioreactors. *IEEE Transactions on Automatic Control*, 37(6):875–880, 1992.
- [9] A.H. Jazwinski. *Stochastic processes and filtering theory*. Mathematics in Science and Engineering, 1970.
- [10] I. Karafyllis and C. Kravaris. From continuous-time design to sampled-data design of observers. *Automatic Control, IEEE Transactions on*, 54(9):2169–2174, 2009.
- [11] M. Lamine Fall, M. Farza, M. M’Saad, E. Pigeon, O. Gehan, and R. Mosrati. State observers for a class of continuous nonlinear systems with discrete measurements. In *Proceeding of the European Control Conference*, 2013.
- [12] J. Löfberg. Yalmip : A toolbox for modeling and optimization in MATLAB. In *Proc. of the CACSD Conference*, Taipei, Taiwan, 2004. <http://control.ee.ethz.ch/~joloef/yalmip.php>.
- [13] W. Lohmiller and J.-J. E. Slotine. Contraction analysis of non-linear distributed systems. *International Journal of Control*, 78(9):678–688, 2005.
- [14] M. Margaliot and M.S. Branicky. Nice reachability for planar bilinear control systems with applications to planar linear switched systems. *IEEE Transactions on Automatic Control*, 54(6):1430–1435, 2009.
- [15] M. Müller. Über das fundamentaltheorem in der theorie der gewöhnlichen differentialgleichungen. *Mathematische Zeitschrift*, 26(1):619–645, 1927.
- [16] M. Nadri, H. Hammouri, and R.M. Grajales. Observer design for uniformly observable systems with sampled measurements. *Automatic Control, IEEE Transactions on*, 58(3):757–762, 2013.
- [17] T. Raff, M. Kogel, and F. Allgower. Observer with sample-and-hold updating for Lipschitz nonlinear systems with nonuniformly sampled measurements. In *Proc. of American Control Conference*, pages 5254–5257, 2008.
- [18] J.F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11-12:625–653, 1999. <http://sedumi.mcmaster.ca/>.
- [19] M. V. Topunov. The convexity of the reachable set for a bilinear controllable system. *Journal of Applied Mathematics and Mechanics*, 67(5):665–670, 2003.
- [20] R. Vinter. *Optimal control*. Systems & Control: Foundations & Applications. Birkhäuser Boston Inc., Boston, MA, 2000.
- [21] A. Zemouche, M. Boutayeb, and G.I. Bara. Observers for a class of Lipschitz systems with

extension to  $H^\infty$  performance analysis. *Systems & Control Letters*, 57(1):18–27, 2008.