



A systematic approach for doing an a priori identifiability study of dynamical nonlinear models

Nathalie Verdière, Sébastien Orange

► To cite this version:

Nathalie Verdière, Sébastien Orange. A systematic approach for doing an a priori identifiability study of dynamical nonlinear models. *Mathematical Biosciences*, 2019, 308, pp.105-113. 10.1016/j.mbs.2018.12.012 . hal-01993292

HAL Id: hal-01993292

<https://hal.science/hal-01993292>

Submitted on 20 Feb 2019

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.

A systematic approach for doing an *a priori* identifiability study of dynamical nonlinear models.

April 22, 2018

N. Verdière¹ and S. Orange¹

Abstract

This paper presents a method for investigating, through an automatic procedure, the (lack of) structural identifiability of dynamical models parameters. This method takes into account constraints on parameters and returns parameters whose estimations turn unidentifiable parameters into identifiable ones. It is based on i) an equivalence between an extension of the notion of identifiability and the existence of solutions of algebraic systems, ii) the use of symbolic computations for testing their existence. This method is described in details and is applied to two examples, the last one involving 12 parameters.

Keywords : Relative identifiability; Nonlinear models; Algorithm; Parameters

1 Introduction

Over the past several years, the study of identifiability of parametrized models has generated a growing interest in many scientific domains [34, 42]. Indeed, when a new model is developed, some parameters may not be directly accessible from the experiments. For estimating them, some numerical procedures using real noisy data have been developed [40, 41]. The *numerical identifiability* study guarantees that from these real data correspond exactly one parameter vector. However, the model structure may not be identifiable and, in that case, the output of the numerical procedure may not return the correct parameter vector since several or an infinite number of parameter values can well describe the input-output data. To study this identifiability property, a theoretical work is in general done in a 'best-case' version of this question, that is from noise-free data. This property, called *structural identifiability*, is also inherent to the structure of the model and constitutes a necessary condition for the numerical identifiability propriety.

¹Normandie Univ, UNIHAVRE, LMAH, FR-CNRS-3335, ISCN, 76600 Le Havre, France
(Corresponding authors: Nathalie.Verdier@univ-lehavre.fr, Sebastien.Orange@univ-lehavre.fr)

Several methods for studying structural identifiability can be found in the literature (see [7, 8, 13, 14, 16, 21, 25, 26, 35, 39, 43]). The identifiability of a model depends on the measured inputs, outputs but, as it was noticed in [14, 38, 41], also on initial conditions or, more generally, on constraints on parameters. When the model is unidentifiable, it is common to search and use identifiable combinations for this purpose through numerical/statistical approaches (See [6, 11, 17, 27, 44]), symbolic-numerical methods (See. [10, 12]) and algebraic approaches (See. [1, 9, 15, 19, 20, 28, 29, 30, 31, 32, 41]). In the last papers, structural identifiability of the model is done from an input-output formulation. In particular, authors in [20, 28, 30, 32] focus on identifiable combinations of parameters in order to reparametrize unidentifiable ODE models into identifiable ones. From the combinations processing, the (un)identifiability of some parameters can arise. In [9, 15, 19, 20, 28, 29, 30, 31, 32, 41], the authors propose to handle the parameter combinations in using Groebner bases. Some of these works had lead to the generation of software tools for performing structural identifiability analysis. For example, the authors in [31] developed an accessible web application named COMBOS which, from model output, provides identifiable combinations of parameters. From paper [1], parameter identifiability had been studied via minimal observable sets and a software tool for performing structural identifiability analysis had been developed.

In this paper, we focus on the (lack of) structural identifiability of parameters which potentially leads to the structural identifiability of the model. Indeed, biological models can contain meaningful unidentifiable parameters, important for the experimenter, that a reparametrization may annihilate. In order to obtain their identifiability, a tree giving the key parameters, potentially accessible from measurements, may serve the modeler. For that purpose, we propose a method and an algorithm based on the use of input-output polynomial coefficients. The algorithm determines a set of parameter lists. Each of these lists gives a set of key unidentifiable parameters whose determination leads to the structural identifiability of some particular parameters and, eventually, the structural identifiability of the model. In order to integrate constraints on parameters, our work is based on semialgebraic tools and not only on Groebner basis as the works cited previously. Considering these constraints permits to deal with the local identifiability of a model.

The first contribution of this paper, presented in Section 2, is a theoretical one. Since the identifiability of a parameter can depend on the identifiability of some other parameters, the classical identifiability definition has to be extended. For this purpose, we propose the notion of *relative identifiability*, that is identifiability of a given parameter when some other parameters are supposed to be known. Then, we generalize a result established in [14] for classical identifiability. This result links output trajectories and solutions of algebraic systems built on input-output polynomials. More precisely, we give an equivalence between the relative identifiability of a parameter and the emptiness of the solution set of equations and/or inequalities system. The latter can be tested with computer algebra softwares. This theoretical approach permits to obtain an automatized approach for studying the (relative) identifiability of the model parameters.

The second contribution, presented in section 3, is an algorithm based on this equivalence. It returns lists of parameters, each of these lists giving a set of

key parameters in order to obtain the identifiability of one or some parameters and eventually the identifiability of the model. Our algorithm can take into account the definition domain of the parameters and, possibly, some algebraic relations, for example, those provided by initial conditions. This new tool can help an experimenter to elaborate strategies about parameters that should be estimated to obtain the identifiability of the model. Notice that the equivalence we establish in section 2 ensures that the set of lists provided by our algorithm is exhaustive for the considered model. In section 4, our algorithm is applied to two examples. The first one revisits the identifiability of microbial growth in a batch reactor, the second one concerns an epidemiological model and involves 12 parameters.

2 Relative identifiability

2.1 Class of systems

All along this paper, we consider nonlinear dynamical parametrized models (controlled or uncontrolled) of the following form:

$$\Gamma^\Theta \begin{cases} \dot{x}(t, \Theta) = g(x(t, \Theta), u(t), \Theta), \\ y(t, \Theta) = h(x(t, \Theta), u(t), \Theta), \end{cases}$$

where:

- $x(t, \Theta) \in \mathbb{R}^n$ denotes the state variables and $y(t, \Theta) \in \mathbb{R}^s$ the outputs respectively,
- the functions g and h are real, rational and analytic ² on M , where M is an open subset of \mathbb{R}^n such that $x(t, \Theta) \in M$ for every $t \in [t_0, T]$. T is a finite or infinite time bound,
- u is the r -dimensional input vector made of smooth functions (infinitely differentiable),
- the vector of parameters $\Theta = (\theta_1, \dots, \theta_m)$ belongs to a subset \mathcal{D} of $\subseteq \mathbb{R}^m$ where \mathcal{D} is an, a priori, known set of admissible parameters.

When an initial condition $x_0 \in \mathbb{R}^n$ is given, the supplementary equation $x(0, \Theta) = x_0$ can be added to System Γ^Θ .

2.2 Admissible parameters and semialgebraic sets

In many models, it is natural to work only on a subset \mathcal{C} of the set of admissible parameters \mathcal{D} when considering additional constraints on the parameter vector such as initial conditions. It may be necessary either to ensure the existence of solutions or because of the nature of the problem. For example, in trophic chains studies, number of preys eaten by predators can not be greater than the total population of preys and in mechanic, a mass can not be negative.

Afterwards, we suppose that these constraints on Θ can be formulated by the mean of algebraic equations and/or inequalities. This leads naturally to consider semialgebraic sets:

²These assumptions are not restrictive since lots of models can be reduced to a rational and analytic model by variable change (see. [2])

Definition 1. (See [3]) *A set of real solutions of a finite set of polynomial equations and/or polynomial inequalities is called a semialgebraic set.*

The introduction of semialgebraic sets enables us to use efficient computer algebra libraries developed for manipulating this kind of mathematical object like the Maple packages `Raglib` (See [18]) or `SemiAlgebraicSetTools` (See also [5, 45], for example).

From now on, we suppose that:

- $C(\Theta)$ denotes the set of all algebraic equations and inequalities verified by the components of the parameter vector $\Theta \in \mathcal{D}$ of the model
- the set \mathcal{C} is the semialgebraic set defined by $C(\Theta)$.

2.3 Structural relative identifiability of parameters

Structural identifiability is a model property in the case of fault-free data/noise depending strongly on the model inputs and outputs but also on the parameter constraints. Given an input-output set of measurements, the structural identifiability is a *a priori* study ensuring the uniqueness of the parameter vector Θ .

More formally, let us recall classical structural identifiability definitions (See [2, 28, 30], for example).

Definition 2. 1. *Let $\Theta \in \mathcal{D}$ be a parameter vector and let us consider the System Γ^θ . If $y(t, \Theta) = y(t, \tilde{\Theta})$, with $\tilde{\Theta} \in \mathcal{D}$, has*

- *only one solution $\Theta = \tilde{\Theta}$, the parameter vector Θ is said globally identifiable,*
 - *finitely many distinct solutions then Θ is said locally identifiable,*
 - *infinitely many solutions then Θ is said unidentifiable.*
2. *A component θ_i of Θ is globally identifiable if for all $\tilde{\Theta} \in \mathcal{D}$ such that $y(t, \Theta) = y(t, \tilde{\Theta})$ we have $\theta_i = \tilde{\theta}_i$.*
3. *The model Γ^θ is said globally identifiable if, for all parameter vector Θ , the model is globally identifiable.*

Afterward, we extend the notion of structural identifiability of *one component* of Θ in introducing the definition of *relative identifiability*. This definition is centered on the identifiability of one particular parameter when some other parameters, and eventually none, are supposed to be identifiable. This definition appears naturally when one wants to determine which parameters should be estimated in order to turn one or some given parameters into identifiable one(s) and, eventually, the model into an identifiable one.

Definition 3. 1. *The parameter $\theta_{a_{r+1}}$ is said to be relatively identifiable with respect to a set of parameters $\{\theta_{a_1}, \dots, \theta_{a_r}\}$ if there exists an input u such that, for all $\tilde{\Theta} = (\tilde{\theta}_1, \dots, \tilde{\theta}_m) \in \mathcal{D}$ whose components satisfy $\tilde{\theta}_{a_1} = \theta_{a_1}, \dots, \tilde{\theta}_{a_r} = \theta_{a_r}$, the equation $y(t, \Theta) = y(t, \tilde{\Theta})$ implies $\theta_{a_{r+1}} = \tilde{\theta}_{a_{r+1}}$.*

2. The relative identifiability study of system Γ^Θ is the determination of the relative identifiability of any parameter relatively to any set of parameters taken among $\theta_1, \dots, \theta_m$.

By definition, $\theta_{a_{r+1}}$ is not relatively identifiable with respect to a set of parameters $\{\theta_{a_1}, \dots, \theta_{a_r}\}$ (resp. not identifiable) if, for all input u , there exists $\tilde{\Theta} = (\tilde{\theta}_1, \dots, \tilde{\theta}_m) \in \mathcal{D}$ whose components satisfy $\tilde{\theta}_{a_1} = \theta_{a_1}, \dots, \tilde{\theta}_{a_r} = \theta_{a_r}, \theta_{a_{r+1}} = \tilde{\theta}_{a_{r+1}}$ and $y(t, \Theta) = y(t, \tilde{\Theta})$.

Remark 1. 1. If the parameter $\theta_{a_{r+1}}$ is relatively identifiable with respect to the empty set of parameters, we recover the notion of the identifiability of this parameter. Consequently, identifiability of System Γ^Θ can be seen as a particular case of the relative identifiability of any component of Θ .

2. Definitions 2 and 3 can be extended to the local case by imposing that $\tilde{\Theta}$ belongs to a neighborhood $\nu(\Theta) \subset \mathcal{D}$ of Θ instead of \mathcal{D} .

The following proposition is a direct consequence of Definition 3.

Proposition 1. If θ_i is relative identifiable with respect to an eventually empty subset $\mathcal{P} \subset \{\theta_1, \dots, \theta_m\}$ then θ_i is relatively identifiable with respect to any subset of $\{\theta_1, \dots, \theta_m\}$ containing \mathcal{P} .

In the next section, we show how to reduce the relative identifiability study to a semialgebraic set problem.

2.4 A differential algebra method to perform relative identifiability analysis

2.4.1 Computation of the exhaustive summary

The input-output approach consists in this paper, from system Γ^Θ and the Rosenfeld-Groebner algorithm, to obtain specific differential polynomials of the following form

$$P_i(y, u, \Theta) = m_{i,0}(y, u) + \sum_{j=1}^{n_i} c_{i,j}(\Theta) m_{i,j}(y, u) = 0, \quad i = 1, \dots, s \quad (1)$$

where $(c_{i,j}(\Theta))_{1 \leq j \leq n_i}$ are rational in Θ such that $c_{i,j}(\Theta) \neq c_{i,l}(\Theta)$ if $j \neq l$, $(m_{i,j}(y, u))_{1 \leq j \leq n_i}$ are differential polynomials relatively to y and u and $m_{i,0}(y, u) \neq 0$.

The computation of these polynomials does not require any consideration of the admissible parameter set $\mathcal{D} \subseteq \mathbb{R}^m$. Moreover, the number of input-output polynomials, detailed in [14, 41], is equal to the number of outputs of System (1). Polynomials P_i are classically known as *input-output polynomials* and in the literature, the sequence $(c_{i,j}(\Theta))_{1 \leq i \leq s, 1 \leq j \leq n_i}$ is called the *exhaustive summary*.

By construction of the input-output polynomials, the functional determinant given by the Wronskian

$$\Delta P_i(y, u) = \begin{vmatrix} m_{i,1}(y, u) & \dots & m_{i,n_i}(y, u) \\ m_{i,1}(y, u)^{(1)} & \dots & m_{i,n_i}(y, u)^{(1)} \\ \vdots & & \vdots \\ m_{i,1}(y, u)^{(n_i-1)} & \dots & m_{i,n_i}(y, u)^{(n_i-1)} \end{vmatrix} \quad (2)$$

is nonsingular for a generic solution according to [22].

2.4.2 Characterization of the relative identifiability of a parameter

In [15], the *a priori* structural identifiability study was brought back to the injectivity study of the function

$$\begin{aligned} \phi: \mathcal{D} &\longrightarrow \mathbb{R}^N \\ \Theta &\longmapsto (c_{i,j}(\Theta))_{1 \leq i \leq s, 1 \leq j \leq n_i} \end{aligned}$$

where $N = \sum_{i=1}^s n_i$.

In the first version of the method described in [15], initial conditions were ignored and few developments were proposed to consider them (see [14, 38, 41]).

In this paper, in order to take into account initial conditions or supplementary constraints on parameters, we restrict the domain of ϕ to a subset \mathcal{C} of the set of admissible parameters \mathcal{D} (see Section 2.2).

The next proposition extends, to the case of relative identifiability, the link given in [15, 41] between identifiability and result summary.

Proposition 2. *The parameter $\theta_{a_{r+1}}$ is relatively identifiable with respect to the eventually empty set of parameters $\{\theta_{a_1}, \dots, \theta_{a_r}\}$ if and only if, for any $\Theta = (\theta_j)_{j=1, \dots, m} \in \mathbb{R}^m$ and $\tilde{\Theta} = (\tilde{\theta}_j)_{j=1, \dots, m} \in \mathbb{R}^m$:*

$$\left\{ \begin{array}{l} \Theta \in \mathcal{C} \\ \tilde{\Theta} \in \mathcal{C} \\ \theta_{a_1} = \tilde{\theta}_{a_1} \\ \vdots \\ \theta_{a_r} = \tilde{\theta}_{a_r} \\ \phi(\Theta) = \phi(\tilde{\Theta}) \end{array} \right. \Rightarrow \theta_{a_{r+1}} = \tilde{\theta}_{a_{r+1}} \quad (3)$$

Proof- Sufficiency Suppose that there exists $\tilde{\Theta} \in \mathcal{C}$ such that $\theta_{a_1} = \tilde{\theta}_{a_1}, \dots, \theta_{a_r} = \tilde{\theta}_{a_r}$ and $y(t, \Theta) = y(t, \tilde{\Theta})$ for all $t \in [0, T]$. The last equality implies that for all $j \in \mathbb{N}$, $y^{(j)}(t, \Theta) = y^{(j)}(t, \tilde{\Theta})$ for all $t \in [0, T]$. By definition of the input-output polynomials $P_i(y, u, \Theta)$ and $P_i(y, u, \tilde{\Theta})$, the difference $P_i(y, u, \Theta) - P_i(y, u, \tilde{\Theta}) = \sum_{k=1}^{n_i} (c_{i,k}(\Theta) - c_{i,k}(\tilde{\Theta})) m_{k,i}(y, u)$ is equal to 0. Since the functional determinant of the last difference is not identically equal to zero, we have $c_{i,k}(\Theta) - c_{i,k}(\tilde{\Theta}) = 0$ which proves that the exhaustive summaries $\Phi(\Theta)$ and $\Phi(\tilde{\Theta})$ are equal. According to (3), $\theta_{a_{r+1}} = \tilde{\theta}_{a_{r+1}}$ and $\theta_{a_{r+1}}$ is relatively identifiable with respect to the set $\{\theta_{a_1}, \dots, \theta_{a_r}\}$.

Necessity By contrapositive. Let us consider two vectors of parameters $\Theta =$

$(\theta_1 \dots, \theta_{a_m})$ and $\tilde{\Theta} = (\tilde{\theta}_1 \dots, \tilde{\theta}_{a_m})$ such that $\theta_{a_1} = \tilde{\theta}_{a_1}, \dots, \theta_{a_r} = \tilde{\theta}_{a_r}, \theta_{a_{r+1}} \neq \tilde{\theta}_{a_{r+1}}$ and $c_{i,j}(\Theta) = c_{i,j}(\tilde{\Theta})$. These equalities imply that, for all $i \in \{1, \dots, s\}$, $P_i(y, u, \Theta) = P_i(y, u, \tilde{\Theta})$. Therefore, the input-output polynomials, and consequently their sets of solutions, are equal. According to Definition 3, $\theta_{a_{r+1}}$ is not relatively identifiable with respect to the set $\{\theta_{a_1}, \dots, \theta_{a_r}\}$.

Remark 2. Proposition 2 can be extended to a local version if we suppose that $\tilde{\Theta}$ belongs to a semialgebraic neighborhood \mathcal{V}_Θ of Θ : if \mathcal{V}_Θ is the solution set of a set $V_\Theta(\tilde{\Theta})$ of polynomial equations and inequalities³ then it is sufficient to substitute $\tilde{\Theta} \in \mathcal{C}$ in Implication (3) with $\tilde{\Theta} \in \mathcal{C} \cup \mathcal{V}_\Theta$.

In the case of an identifiability study, some approaches had been proposed in the literature to solve the real system $\phi(\Theta) = \phi(\tilde{\Theta})$ [4, 15, 37, 41]. But these tools use Groebner basis algorithms which are not adapted for the resolution of general real algebraic systems. Moreover, the eventual inequalities satisfied by parameters can not be considered by these approaches whereas the identifiability result may depend heavily on them.

In this paper, we propose an automatic procedure to prove Implication (3) including the parameter constraints. This procedure is based on semialgebraic tools. Some of these tools, already implemented in computer algebra systems have never been used to certify the structural identifiability of models.

From the system composing the left side of Implication (3), it is possible to construct a system composed of polynomial equations and inequalities defining a semialgebraic set. Indeed, \mathcal{C} is assumed to be a semialgebraic set and system $c_{i,j}(\Theta) = c_{i,j}(\tilde{\Theta})$, $i = 1, \dots, s$, $j = 1, \dots, n_i$ is equivalent to an algebraic system by handling the numerators and the denominators of these rational fractions in order to obtain polynomial equations. Polynomial inequations corresponding to non vanishing denominators are then added to $C(\Theta)$ and $C(\tilde{\Theta})$.

Let us define, for $\Theta = (\theta_1, \dots, \theta_m) \in \mathbb{R}^m$, $\tilde{\Theta} = (\tilde{\theta}_1, \dots, \tilde{\theta}_m) \in \mathbb{R}^m$ and $0 \leq r \leq n$, the following set of equations and inequalities:

$$\begin{aligned} S_{\theta_{a_1}, \dots, \theta_{a_r}} = & C(\Theta) \cup C(\tilde{\Theta}) \cup \{\theta_{a_i} = \tilde{\theta}_{a_i} \mid 1 \leq i \leq r\} \\ & \cup \{c_{i,j}(\Theta) = c_{i,j}(\tilde{\Theta}) \mid 1 \leq i \leq s, 1 \leq j \leq n_i\}. \end{aligned}$$

In order to use semialgebraic tools for proving Implication (3), the following corollary is based on the classical tool of computer algebra techniques called the Rabinowitsch trick (See [3, 36]): to show that an algebraic system has a solution with $u \neq 0$, where u is one of its indeterminates, this trick consists in adding the equation $uv - 1 = 0$ to the system, where v is a new indeterminate, and to prove that the new system has a solution. By this way, computer algebra packages testing the emptiness of semialgebraic sets can be used such as `HasRealSolution` of the Maple package `Raglib` [18] for example.

Corollary 1. Let $(\theta_1, \dots, \theta_m)$ and $(\tilde{\theta}_1, \dots, \tilde{\theta}_m)$ in \mathcal{C} . The following conditions are equivalent:

1. the parameter $\theta_{a_{r+1}}$ of System Γ^Θ is relatively identifiable with respect to the eventually empty set of parameters $\{\theta_{a_1}, \dots, \theta_{a_r}\}$;

³For example, algebraic constraints defining $V_\Theta(\tilde{\Theta})$ can be of the form $-\epsilon_i < \theta_{a_i} - \tilde{\theta}_{a_i} < \epsilon_i$ for some given non negative ϵ_i .

2. the system $S_{\theta_{a_1}, \dots, \theta_{a_r}} \cup \left\{ v \left(\theta_{a_{r+1}} - \tilde{\theta}_{a_{r+1}} \right) - 1 = 0 \right\}$, where v is a new variable, has no real solution $(\theta_{a_1}, \dots, \theta_{a_m}, \tilde{\theta}_{a_1}, \dots, \tilde{\theta}_{a_m}, v) \in \mathbb{R}^{2m+1}$.

Proof–1. \Rightarrow 2. By contrapositive. If system $S_{\theta_{a_1}, \dots, \theta_{a_r}} \cup \left\{ v \left(\theta_{a_{r+1}} - \tilde{\theta}_{a_{r+1}} \right) - 1 = 0 \right\}$, where v is a new variable, has a real solution, then this solution satisfies $\theta_{a_{r+1}} - \tilde{\theta}_{a_{r+1}} \neq 0$. According to Proposition 2, the parameter $\theta_{a_{r+1}}$ is not relatively identifiable with respect to $\{\theta_{a_1}, \dots, \theta_{a_r}\}$.

2. \Rightarrow 1. By contrapositive. If $\theta_{a_{r+1}}$ of System $S_{\theta_{a_1}, \dots, \theta_{a_r}}$ is not relatively identifiable with respect to $\{\theta_{a_1}, \dots, \theta_{a_r}\}$, there exist two solutions $(\theta_{a_1}, \dots, \theta_{a_m})$ and $(\theta_{a_1}, \dots, \theta_{a_r}, \tilde{\theta}_{a_{r+1}}, \dots, \tilde{\theta}_{a_m})$ of this system such that $\theta_{a_{r+1}} \neq \tilde{\theta}_{a_{r+1}}$ (See Proposition 2). This inequality is equivalent to the existence of a real v such that $v \left(\theta_{a_{r+1}} - \tilde{\theta}_{a_{r+1}} \right) - 1 = 0$. Consequently, the $2m + 1$ t-uplet of real $(\theta_{a_1}, \dots, \theta_{a_m}, \theta_{a_1}, \dots, \theta_{a_r}, \tilde{\theta}_{a_{r+1}}, \dots, \tilde{\theta}_{a_m}, v)$ is a solution of system $S_{\theta_{a_1}, \dots, \theta_{a_r}} \cup \left\{ v \left(\theta_{a_{r+1}} - \tilde{\theta}_{a_{r+1}} \right) - 1 = 0 \right\}$.

Remark 3. In the same way as in Proposition 2, Corollary 1 can be extended to a local version by substituting $C(\Theta)$ with $C(\Theta) \cup V_{\Theta}(\Theta)$ in this corollary where $V_{\Theta}(\Theta)$ is defined in Remark 2.

3 Relative identifiability algorithm

In this section, we present our algorithm for the relative identifiability study of parameters of System Γ^{Θ} .

3.1 Useful concepts

In order to describe our algorithm and its outputs, we recall the following notations and definitions of computer science.

Notation 1. $[\theta_{a_1}, \dots, \theta_{a_r}]$ is the list⁴ of r distinct elements $(1 \leq r \leq m)$ taken among $\{\theta_1, \dots, \theta_m\}$.

Definition 4. 1. Let $[\theta_{a_1}, \dots, \theta_{a_r}]$ be a list of r distinct elements of $\{\theta_1, \dots, \theta_m\}$. The empty list and any list $[\theta_{a_1}, \dots, \theta_{a_i}]$ $(1 \leq i \leq r - 1)$ are called prefix of $[\theta_{a_1}, \dots, \theta_{a_r}]$.

2. Let l_1 and l_2 be two lists. The concatenated list $l_1 \text{ cat } l_2$ is the list obtained by joining the two lists l_1 and l_2 end-to-end.

In order to distinguish identifiable parameters to non identifiable ones in a list, we introduce the notation hereafter.

Notation 2. Let $l = [\theta_{a_1}, \dots, \theta_{a_r}]$ be a list of r elements taken among $\{\theta_1, \dots, \theta_m\}$. If the parameter θ_{a_1} is not identifiable, it will be written \emptyset_{a_1} in the list l . More generally, if θ_{a_i} $(2 \leq i \leq r)$ is not relatively identifiable with respect to $\{\theta_{a_1}, \dots, \theta_{a_{i-1}}\}$, we write it \emptyset_{a_i} in l .

⁴In computer science, a list is a finite **sequence** of elements of a set.

From now on, in order to lighten the text, a parameter will be said relatively identifiable with respect to a list of parameters if it is relatively identifiable with respect to the corresponding set of parameters.

3.2 Relative identifiability study

A way to obtain an exploitable output for an algorithm is to use lists of parameters rather than sets and naturally to consider a tree traversal algorithm in order to avoid useless emptiness tests of semialgebraic sets and also redundant information in this output. Such an algorithm consists in constructing all the possible lists of $\theta_1, \dots, \theta_m$ and in indicating non identifiable parameters of these lists. This approach leads to the following definition which is the output of our algorithm.

Definition 5. *The identifiability tree \mathcal{T} of Γ^Θ is the set of all possible lists of parameters taken among $\{\theta_1, \dots, \theta_m\}$ such that, for all list $b \in \mathcal{T}$ and any prefix p of b , p is followed in b by an identifiable parameter relatively to p if there exists one.*

Note that if lists of \mathcal{T} have a common prefix, tests performed for this prefix will be computed only once. Moreover, in order to avoid some useless tests, the following remarks are used in our algorithm.

Two important remarks to improve the efficiency of the identifiability tree algorithm.

Remark A *(Consequence of Proposition 1) If a prefix p of a list of \mathcal{T} is followed by successive identifiable parameters, these identifiable parameters can be permuted in this list since it refers to the same information: each of them are relatively identifiable with respect to p .*

Remark B *If lists completing a given prefix p_1 into elements of \mathcal{T} are known, these lists complete any permutation of p_1 into elements of \mathcal{T} . Indeed, the relative identifiability of a parameter relatively to a prefix p depends only on the set of parameters appearing in p .*

This last remark leads naturally to a pre-order tree traversal (see [23]) in order to complete prefixes with lists of \mathcal{T} already computed. By this way, some emptiness tests of semialgebraic sets can be avoided which decrease the global cost of the computation of \mathcal{T} essentially due to these tests.

3.3 Algorithm for studying the relative identifiability

In this section, we describe our recursive algorithm called `IdentifiabilityTree` which returns the identifiability tree \mathcal{T} . The complete pseudo-code of this algorithm is the object of the appendix of this paper. Identifiability tests are realized by applying Corollary 1.

This algorithm takes as input any prefix p of a list of \mathcal{T} and computes all the lists of \mathcal{T} admitting p as prefix. Lists of \mathcal{T} produced by the algorithm are

stored in the set \mathcal{T}' . In particular, the entire identifiability tree is returned by the call `IdentifiabilityTree(p)`, with $p = []$ and \mathcal{T}' initialized to \emptyset .

During any recursive call, the prefix p is completed first with all the identifiable parameters with respect to p in any order (see Remark A) and, after, by each of the non identifiable parameters. Generated prefixes are then completed by recursive calls of the algorithm. When it is possible, Remark B is used to complete prefixes using \mathcal{T}' avoiding useless semialgebraic emptiness tests.

Let us summarize the different steps of the algorithm applied to a prefix p .

`IdentifiabilityTree(p)` _____

1. While there exists some identifiable parameters relatively to p do
 - 1.1 Complete p with an identifiable parameter θ ;
 - 1.2 Check whether Remark B. can be applied to complete p cat $[\theta]$ into new lists of the identifiability tree. If it is the case, add these new lists to \mathcal{T}' , stop this While loop and do not proceed to step 2.
 2. If there exists some non identifiable parameters relatively to p , for each of these parameters, do
 - 2.1 Complete p with a non identifiable parameter \emptyset ;
 - 2.2 If Remark B. can be applied to complete p cat $[\emptyset]$ into new lists of the identifiability tree, add these new lists to \mathcal{T}' . Otherwise, do a recursive call of the algorithm with p cat $[\emptyset]$ as input.
-

The following proposition gives the complexity of our algorithm.

Proposition 3 (Complexity). *The number of emptiness tests of semialgebraic sets performed by our algorithm is bounded by $(2m - \nu + 2)2^{\nu-1}$ where m (resp. ν) is the number of parameters (resp. of non identifiable parameters) of System S .*

Proof. Let \mathcal{T}_r the set of prefixes of minimal length of lists of \mathcal{T} admitting exactly r non identifiable parameters ($0 \leq r \leq \nu$). Let \mathcal{C}_r be the subset of \mathcal{T}_r composed of prefixes for which the non identifiable parameters requires exactly r emptiness tests of semialgebraic sets.

For each prefix of \mathcal{C}_r constructed by the algorithm, at least $r!$ prefixes of \mathcal{T}_r are completed into lists of \mathcal{T} using Remark B that is without any tests. Since there is at most $\nu(\nu - 1) \cdots (\nu - (r - 1))$ prefixes, up to a permutation of non identifiable parameters, \mathcal{C}_r contains at most $\binom{\nu}{r}$ prefixes.

This implies that the number of tests needed for the last non identifiable parameters appearing in the lists of \mathcal{C}_r is bounded by $\binom{\nu}{r}$ and, consequently, that the number of tests performed by our algorithm for the determination of the non identifiable parameters is bounded by $\sum_{r=0}^{\nu} \binom{\nu}{r} = 2^{\nu}$.

Moreover, for each prefix p of \mathcal{T}_r , there are $m - r$ identifiable parameters in the list of \mathcal{T} corresponding to p . Consequently, at most $(m - r)\binom{\nu}{r}$ tests are performed for the identifiable parameters of lists of \mathcal{T} completing lists of \mathcal{T}_r . This implies that the number of tests realized by our algorithm for the

identifiable parameters is bounded by

$$\sum_{r=0}^{\nu} (m-r) \binom{\nu}{r} = (2m-\nu)2^{\nu-1}.$$

Therefore, the number of test realized by our algorithm for computing the identifiability tree is bounded by $(2m-\nu+2)2^{\nu-1}$. \square

3.4 Steps of the method

The step for studying relative identifiability of System Γ^{Θ} are summed up below:

Step 1. Determination of the set of conditions $C(\Theta)$, including eventually the initial conditions.

Step 2. Elimination of unobservable state variables using the Rosenfeld-Groebner algorithm in order to obtain the input output polynomials of the system.

Step 3. Construction of the exhaustive summary.

Step 4. Computation of the relative identifiability tree by our algorithm.

4 Examples

The examples presented in this section had been treated by an implementation of our algorithm in the computer algebra system Maple. In order to test relative identifiability using Corollary 1, emptiness tests of semialgebraic sets had been realized with the function `IsEmpty` of the Maple package `SemiAlgebraicSetTools`.

The first example revisits the identifiability of the model introduced by [24] and describes the microbial growth in a batch reactor. The identifiability tree study shows that, for this example, initial conditions do not change the identifiability of the model contrary to parameter constraints. Moreover, with a supplementary observation, we deduce the identifiability of the model from the output of our algorithm.

The second example of this section, dealing with the epidemiological model taken from [33], is more complex in term of number of parameters. Despite the cost of the computation, the size of the exhaustive summary and implicitly the number of algebraic polynomials involved in the input-output polynomial, our algorithm provides the identifiability tree of this model.

4.1 Microbial growth in a batch reactor

The model of a microbial growth in a batch reactor studied by [8, 19] is considered. It is known as the first unidentifiable real-life nonlinear model from the only measure of concentration of microorganisms. We propose to revisit this model in introducing a supplementary parameter: the *maintenance factor* m . It corresponds to the mass of substrate that one unit of biomass requires for non-growth functions in one unit time.

The dynamic model for the growth process is also governed by the two following equations:

$$\Gamma^\Theta \begin{cases} \dot{x}(t) = \frac{\mu s(t) x(t)}{K_S + s(t)} - mY x(t), \\ \dot{s}(t) = \frac{-\mu s(t) x(t)}{Y(K_S + s(t))}, \\ y = x. \end{cases} \quad (4)$$

and where x is the concentration of microorganisms, s the concentration of growth-limiting substrate, μ the maximum velocity of the reaction, K_S the Michaelis-Menten constant and Y the yield coefficient. Due to the nature of the parameters, μ , K_S , m and Y are supposed to be positive reals. Since it is easier to measure $x(t)$ than $s(t)$ within the reactor, we consider that the measured output is $y = x$. The model studied in [8] is a particular case in setting $K_d = mY$.

We now apply the different steps of the method proposed at Section 3.4.

Step 1. The significations of the parameters lead to the set of conditions $C(\Theta) = \{\mu > 0, K_S > 0, m > 0, Y > 0\}$.

Step 2. The Rosenfeld-Groebner algorithm applied to System 4 with the elimination order induced by $y \prec x \prec s$ returns the following input-output polynomial:

$$(Y^3 m^3 - 2Y^2 m^2 \mu + Y m \mu^2) y^3 + (3Y^2 m^2 - 4Y m \mu + \mu^2) y^2 \dot{y} + \mu K_S Y (y \ddot{y} - \dot{y}^2) + (3Y m - 2\mu) y \dot{y}^2 + \dot{y}^3 = 0.$$

Step 3. From this differential polynomial, we obtain the exhaustive summary

$$(Y^3 m^3 - 2Y^2 m^2 \mu + Y m \mu^2, 3Y^2 m^2 - 4Y m \mu + \mu^2, \mu K_S Y, 3Y m - 2\mu)$$

which is also the image of $\Theta = (\mu, K_S, m, Y)$ by ϕ (See Section 2.4.2).

Step 4. The relative identifiability tree can then be computed by Algorithm `IdentifiabilityTree`. The algorithm runs through the full tree of Figure 1.

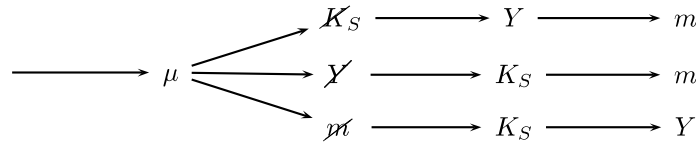


Figure 1: Tree transversal realized by Algorithm `IdentifiabilityTree`

The output of the algorithm is $\mathcal{T} = \{[\mu, K_S, Y, m], [\mu, Y, K_S, m], [\mu, m, K_S, Y]\}$. These lists show that μ is identifiable and that none of the three other parameters are identifiable. Remark that if one parameter among K_S , m or Y is known then the system is identifiable.

Remark 4. 1. In this example, the role of the constraints imposed on the parameters plays a crucial role for establishing the identifiability of System 4. Indeed, if constraints $C(\Theta)$ are not considered, i.e. if $C(\Theta) = \emptyset$, we obtain

$$\mathcal{T} = \{[\mu, K_S, Y, m], [\mu, K_S, m, Y], [\mu, Y, K_S, m], [\mu, Y, m, K_S], [\mu, m, K_S, Y], [\mu, m, Y, K_S]\}.$$

Contrary to the previous situation, the tree parameters have to be identified to turn the model into an identifiable one.

2. In order to evaluate the role of initial conditions, we introduce the new unknown parameters x_0, s_0, xp_0, sp_0 corresponding respectively to $x(0), s(0), \dot{x}(0), \dot{s}(0)$ and we add to $C(\Theta)$ equations obtained from System (4) evaluated at $t = 0$.

Computation of \mathcal{T} shows that in any list, the parameter μ is identifiable and either

- two parameters among x_0, s_0, sp_0, xp_0 are not identifiable and exactly one parameter among K_s, Y and m is not identifiable or
- three parameters among x_0, s_0, sp_0, xp_0 are not identifiable.

The output of our algorithm gives all the possible choices involving model parameters and initial conditions in order to obtain identifiability of any parameter without considering their feasibility.

3. In the same way, without considering the initial conditions, when outputs x and s are supposed to be observed, the identifiability tree is reduced to $\{[K_s, Y, m, \mu]\}$. This means that the model is identifiable.
4. The authors in [8, 19] prove that in the case $K_d = mY$, the parameters and parameter combinations $x(0), K_d mY, s(0)/Y$ and $s(0)/K_s$ are identifiable, while individual parameters $m, Y, s(0)$ and K_s are unidentifiable. Furthermore, using the parameter combinations, they deduced that, if $s(0)$ or K_s were known, then all the parameters become identifiable. With the help of Algorithm `IdentifiabilityTree`, we obtain the same result.

4.2 Chikungunya model

The following model Γ^Θ is an epidemiological model describing the propagation of the Chikungunya disease to human population proposed in [33]. In [46], an identifiability study had been done on this model assuming that some parameters are known. In this section, we assume that none of the parameters are known and we propose to find lists of key parameters which turn Γ^Θ into an identifiable model.

$$\Gamma^\Theta \left\{ \begin{array}{l} \Gamma_1^\Theta \left\{ \begin{array}{l} \dot{E}(t) = bA(t) \left(1 - \frac{E(t)}{k_E} \right) - (s + d)E(t) \\ \dot{L}(t) = sE(t) \left(1 - \frac{L(t)}{k_L} \right) - (s_L + d_L)L(t) \\ \dot{A}(t) = s_L L(t) - d_m A(t) \end{array} \right. \\ \\ \Gamma_2^\Theta \left\{ \begin{array}{l} \dot{S}_H(t) = -(b_H + \beta_H I_m(t)) S_H(t) + b_H \\ \dot{I}_H(t) = \beta_H I_m(t) S_H(t) - (\gamma + b_H) I_H(t) \\ \dot{I}_m(t) = - \left(s_L \frac{L(t)}{A(t)} + \beta_m I_H(t) \right) I_m(t) + \beta_m I_H(t) \end{array} \right. \\ \\ y_1 = L, y_2 = S_H, y_3 = I_H \end{array} \right.$$

Let us explain briefly this system. Equations appearing in Γ_1^Θ correspond to the three biological steps in the life cycle of the mosquito transmitting the

disease : E is the number of eggs, L the number of pupae and E the number of adult females. Assuming that the number of larvae can be counted weekly by biologists, this variable is considered as one of the measured variable of the model. In System Γ_2^Θ , S_H corresponds to the number of humans susceptible to be infected, I_H to the number of infected humans and I_M to infected mosquitoes. In Γ_2^Θ , S_H and I_H can be supposed to be measured variables.

As in [46], the model outputs are defined as the number of pupae, $y_1 = L$, the number of humans susceptible to be infected $y_2 = S_H$ and the number of infected humans $y_3 = I_H$.

The unknown parameter vector is $\Theta = (k_E, k_L, b_H, \beta_H, \beta_m, d_L, d_m, \gamma, b, d, s, s_L)$ where

- b is the intrinsic rate of eggs, s (resp. s_L) is the transfer rate from E to L (resp. from L to A);
- k_E (resp. k_L) is the carrying capacity of E (resp. carrying capacity of L);
- d , d_L and d_m are the rates of natural deaths for eggs, larvae and adults;
- b_H the human birth;
- γ is the transfer rate between infected humans and recovered humans;
- β_H (resp. β_M) is the infectious contact rate between susceptible humans and vectors (resp. susceptible mosquitoes and humans).

Naturally, real parameters of System Γ^Θ are supposed to be positive which leads to set $C(\Theta) = \{b_H > 0, \beta_H > 0, \gamma > 0, s_L > 0, \beta_m > 0, d > 0, d_L > 0, d_M > 0, b > 0, s > 0, k_E > 0, k_L > 0, b_H > 0\}$.

In order to apply the Rosenfeld-Groebner algorithm to system Γ^Θ , the last equation of Γ_2^Θ is multiplied by function A . This algorithm used with the elimination order induced by

$$[y_1, y_2, y_3] \prec [S_H, I_H, L, I_M, E, A]$$

returns three differential equations linking unknown parameters and functions y_1, y_2, y_3 . One of these equations does not have a constant coefficient c_{ij} and therefore its coefficients can be estimated uniquely up to a multiplicative constant from experimental values (see [41]). This equation is divided by one of these coefficients, $k_E k_L$, to obtain a polynomial of the form (1).

From these three input-output polynomials written in the form (1), we obtain an exhaustive summary containing 74 algebraic polynomials depending on the parameters. Our algorithm applied to the system obtained from $C(\Theta)$ and function ϕ returns the following set⁵:

⁵Despite the size of the semialgebraic system, the time needed for the computation of the identifiability tree is 1634 s with Intel(R) Core(TM) i7 2.5 GHz processor with 8 GO of RAM.

$$\begin{aligned}
& \{[k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{k_E}, d, s, \cancel{b}, d_L, s_L], [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{k_E}, d, s, \cancel{b}, d_L, s, s_L], \\
& [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{b}, \cancel{d}, k_E, d_L, s, s_L], [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{b}, \cancel{d}, k_E, d, s, s_L], \\
& [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{b}, \cancel{s}, k_E, d, d_L, s_L], [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{b}, \cancel{s_L}, k_E, d, d_L, s], \\
& [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{d}, k_E, s, \cancel{b}, d_L, s_L], [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{d}, k_E, s, \cancel{d_L}, b, s_L], \\
& [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{d}, k_E, s, \cancel{s_L}, b, d_L], [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{d_L}, s_L, \cancel{b}, k_E, d, s], \\
& [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{d_L}, s_L, \cancel{b}, k_E, b, d], [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{d_L}, s_L, \cancel{b}, k_E, b, s], \\
& [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{s}, k_E, d, \cancel{d_L}, b, s_L], [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{s}, k_E, d, \cancel{s_L}, b, d_L], \\
& [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{s_L}, d_L, \cancel{k_E}, b, d, s], [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{s_L}, d_L, \cancel{b}, k_E, d, s], \\
& [k_L, b_H, \beta_H, \beta_m, d_m, \gamma, \cancel{s_L}, d_L, \cancel{b}, k_E, b, d]\}
\end{aligned}$$

For the given set of outputs, y_1 , y_2 and y_3 , this set of lists shows that $k_L, b_H, \beta_H, \beta_m, d_m, \gamma$ are identifiable and indicates also which pairs of other parameters have to be estimated to turn the model into an identifiable one. Many other information can be extracted from these lists. For example, if an experimenter wants an estimation of the unknown unidentifiable parameter k_E from the model and from the estimation of only one parameter, these lists show that he must estimate either d or s .

5 Conclusion

The introduction of the *relative identifiability* definition permits to elaborate strategies for studying parameters identifiability of nonlinear models. The proposed algorithm does not require any experimental values of inputs and outputs and gives a mean for studying *a priori* relative identifiability of differential models. This method enables the user to take into account inequalities satisfied by parameters and initial conditions.

This method is composed of two steps. The first one consists in the determination of algebraic relations between the outputs, the inputs and the unknown parameters by a classical elimination method used in differential algebra. The second step consists in using the algorithm proposed in this paper in order to determine the identifiability of any parameter relatively to any set of other parameters by testing the emptiness of semialgebraic sets.

The output of this algorithm can be used, in particular, to determine which parameters must be known to make the model identifiable for the given set of observations or outputs. A future work consists in establishing the link between the output of the algorithm presented in this paper and the parameter combinations studied in [28, 30, 31].

References

- [1] M. Anguelova, J. Karlsson, and M. Jirstrand. Minimal output sets for identifiability. *Mathematical Biosciences*, 239(1):139153, 2012.
- [2] S. Audoly, G. Bellu, L. D'Angio, M. P. Saccomani, and C. Cobelli. Global identifiability of nonlinear models of biological systems. *IEEE Trans. Biomed. Eng.*, 48:55–65, 2001.

- [3] S.S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in Real Algebraic Geometry (Algorithms and Computation in Mathematics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- [4] G. Bellu, M. Pia Saccomani, S. Audoly, and L. D’Angiò. DAISY: A new software tool to test global identifiability of biological and physiological systems. *Computer Methods and Programs in Biomedicine*, 28 (1):52—61, 2007.
- [5] C. W. Brown. QEPCAD B: A program for computing with semi-algebraic sets using CADs. *SIGSAM BULLETIN*, 37:97–108, 2003.
- [6] R. Brun, M. Khni, H. Siegrist, W. Gujer, and P. Reichert. Practical identifiability of ASM2d parameters systematic selection and tuning of parameter subsets. *Water Research*, 36(16):41134127, 2002.
- [7] M.J. Chapman, K.R. Godfrey, M.J. Chappell, and N.D. Evans. Structural identifiability of nonlinear systems using linear/nonlinear splitting. *Int. J. Control*, 76:209–216, 2003.
- [8] M.J. Chappell and K.R. Godfrey. Structural identifiability of the parameters of a nonlinear batch reactor model. *Math. Biosci*, 108:245–251, 1992.
- [9] M.J. Chappell and R. Gunn. A procedure for generating locally identifiable reparametrisations of unidentifiable non-linear systems by the similarity transformation approach. 148:21–41, 03 1998.
- [10] R. Choquet and D.J. Cole. A hybrid symbolic-numerical method for determining model structure. *Mathematical Biosciences*, 236(2):117125, 2012.
- [11] A. Cintrón-Arias, H. Banks, A. Capaldi, and A. Lloyd. A sensitivity matrix based methodology for inverse problem formulation. 17:1–20, 08 2009.
- [12] D. Cole, B. Morgan, E. Catchpole, and B. Hubbarn. Parameter redundancy in markrecovery models. *Biometrical Journal*, 54(2):117125, 2012.
- [13] L. Denis-Vidal and G. Joly-Blanchard. Equivalence an identifiability analysis of uncontrolled nonlinear dynamical systems. *Automatica*, 40:287–292, 2004.
- [14] L. Denis-Vidal, G. Joly-Blanchard, and C. Noiret. Some effective approaches to check identifiability of uncontrolled nonlinear systems. *Mathematics and Computers in Simulation*, 57:35–44, 2001.
- [15] L. Denis-Vidal, G. Joly-Blanchard, C. Noiret, and M. Petitot. An algorithm to test identifiability of non-linear systems. In *Proceedings of 5th IFAC NOLCOS*, volume 7, pages 174–178, St Petersburg, Russia, 2001.
- [16] M. Diop and M. Fliess. Nonlinear observability, identifiability and persistent trajectories. In *Proc. 30th CDC*, pages 714–719, Brighton, UK, 1991.
- [17] M.C. Eisenberg and M.A.L. Hayashi. Determining identifiable parameter combinations using subset profiling. *Mathematical Biosciences*, 256:116126, 2014.

- [18] M.S. El Din. RAGLib: A library for real solving polynomial systems of equations and inequalities. 2007.
- [19] N.D. Evans, M.J. Chapman, M.J. Chappell, and K.R. Godfrey. Identifiability of uncontrolled nonlinear rational systems. *Automatica*, (38):1799–1805, 2002.
- [20] N.D. Evans and M.J. Chappell. Extensions to a procedure for generating locally identifiable reparameterisations of unidentifiable systems. *Math. Biosci.*, (168):137–159, 2000.
- [21] M. Fliess and T. Glad. An algebraic approach to linear and nonlinear control. In *Essays on control: perspectives in the theory and its application*, volume 7, pages 223–267, Cambridge, MA, Birkhauser, 1993.
- [22] S.T. Glad. Nonlinear input output relations and identifiability, 02 1992.
- [23] S. Harris and J. Ross. *Beginning algorithms*. Wiley Indianapolis, IN, 2005.
- [24] A. Holmberg and J. Ranta. Procedures for parameter and state estimation of microbial growth process models. *Automatica*, 18(2):181–193, 1982.
- [25] D.L. Janzén, M. Jirstrand, M.J. Chappell, and N.D. Evans. Three novel approaches to structural identifiability analysis in mixed-effects models. *Computer Methods and Programs in Biomedicine*, In Press, Corrected Proof, Available online 6 May 2016, 2016.
- [26] L. Ljung and T. Glad. On global identifiability for arbitrary model parametrizations. *Automatica*, 30:265–276, 1994.
- [27] K. McLean, S. Wu, and K. McAuley. Mean-squared-error methods for selecting optimal parameter subsets for estimation. *Industrial & Engineering Chemistry Research*, 51(17):6105–6115, 2012.
- [28] N. Meshkat, C. Anderson, and J.J. DiStefano. Finding identifiable parameter combinations in nonlinear ODE models and the rational reparameterization of their inputoutput equations. *Mathematical Biosciences*, 233(1):1931, 2011.
- [29] N. Meshkat, C. Anderson, and J.J. DiStefano III. Alternative to ritt’s pseudodivision for finding the input-output equations in algebraic structural identifiability analysis. 03 2012.
- [30] N. Meshkat, M. Eisenberg, and J.J. DiStefano. An algorithm for finding globally identifiable parameter combinations of nonlinear ODE models using Grbner Bases. *Mathematical Biosciences*, 222(2):6172, 2009.
- [31] N. Meshkat, C. Kuo, and J.J. DiStefano. On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and combos: A novel web implementation. In *PloS one*, 2014.
- [32] N. Meshkat and S. Sullivant. Identifiable reparametrizations of linear compartment models. *Journal of Symbolic Computation*, 63:4667, 2014.

- [33] D. Moulay, M.A. Aziz-Alaoui, and M. Cadivel. The chikungunya disease: Modeling, vector and transmission global dynamics. *Mathematical Biosciences*, 229(1):50–63, 2011.
- [34] A. Pironet, P.C. Dauby, J.G. Chase, P.D. Docherty, J.A. Revie, and T. Desai. Structural identifiability analysis of a cardiovascular system model. *Medical Engineering & Physics*, 38(5):433–441, 2016.
- [35] H. Pohjanpalo. System identifiability based on the power series expansion of the solution. *Math. Biosciences*, 41:21–33, 1978.
- [36] J. L. Rabinowitsch. Zum Hilbertschen Nullstellensatz. *Math. Ann.*, 102(1):520, 1930.
- [37] M.P. Saccomani, S. Audoly, G. Bellu, and L. D’Angiò. Examples of testing global identifiability of biological and biomedical models with the DAISY software. *Journal Computers in Biology and Medicine*, 40 (4):402–407, 2010.
- [38] M.P. Saccomani, S. Audoly, and L. D’Angio. Parameter identifiability of nonlinear systems: the role of initial conditions. *Automatica*, 39:619–632, 2004.
- [39] J. D. Stigter and J. Molenaar. A fast algorithm to assess local structural identifiability. *Automatica J. IFAC*, 58:118–124, 2015.
- [40] A. Tarantola. *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.
- [41] N. Verdière, L. Denis-Vidal, G. Joly-Blanchard, and D. Domurado. Identifiability and estimation of pharmacokinetic parameters for the ligands of the macrophage mannose receptor. *Int. J. Appl. Math. Comput. Sci.*, 15:517–526, 2005.
- [42] O.J. Walch and M.C. Eisenberg. Parameter identifiability and identifiable combinations in generalized Hodgkin–Huxley models. *Neurocomputing*, 199:137–143, 2016.
- [43] E. Walter and Y. Lecourtier. Global approaches to identifiability testing for linear and nonlinear state space models. *Math. and Comput.in Simul.*, 24:472–482, 1982.
- [44] S.R. Weijers and P.A. Vanrolleghem. A procedure for selecting best identifiable parameters in calibrating activated sludge model no. 1 to full-scale plant data. *Water Science and Technology*, 36(5):6979, 1997. WATERMA-TEX ’97 Systems Analysis and Computing in Water Quality Management: Towards A New Agenda.
- [45] B. Xia. DISCOVERER: A Tool for Solving Semi-algebraic Systems. *ACM Commun. Comput. Algebra*, 41(3):102–103, September 2007.
- [46] L. Denis-Vidal N. Verdière Z. Shousheng. Identifiabilitys study in a model describing the propagation of the chikungunya to the human population. In *MOSIM 2014*, Nancy,France, 5-7 November, 2014.

Appendix

The function `IdentifiabilityTree` takes as input the system

$$S = \{\phi(\Theta) = \phi(\tilde{\Theta}), C(\Theta), C(\tilde{\Theta})\}$$

and returns the identifiability tree (See Definition 5).

In order to avoid useless computations, Remark B is used in the auxiliary function `CompletionWithComputedBranches`. When this function permits to complete a prefix l into branches of the identifiability tree, the set \mathcal{T}' returned by this function contains all the branches of the identifiability tree admitting l as prefix.

```
Function CompletionWithComputedBranches( $l, \mathcal{T}$ );
.  /* Inputs: .A set  $\mathcal{T}$  of branches of the identifiability tree
.             .A list  $l$  of parameters taken among  $\{\theta_1, \dots, \theta_m\}$  */
.  /* Outputs: .A boolean HasBeenCompleted equal to True if the sets of parameters of  $l$  and of
.             one prefix of at least one element of  $\mathcal{T}$  are equal ; False otherwise.
.             . The set  $\mathcal{T}'$  containing  $\mathcal{T}$  and all the branches of the form  $l \text{ cat } s$  where  $s$  is
.             the suffix of a branch of  $\mathcal{T}$  admitting as prefix a permutation of  $l$ . */
.  HasBeenCompleted := False;
.   $\mathcal{T}'$  :=  $\mathcal{T}$  ;
.  For  $l' \in \mathcal{T}$  do
.  .  If a permutation of a prefix of  $l'$  is equal to  $l$  then
.  . .   $l''$  :=  $l \text{ cat } l'[\text{Length}(l) + 1, \dots, m]$ ;
.  . .   $\mathcal{T}'$  :=  $\mathcal{T}' \cup \{l''\}$  ;
.  . .  HasBeenCompleted := True;
.  .  End If
.  End For
.  Return [ HasBeenCompleted ;  $\mathcal{T}'$  ];
End Function ;
```

Corollary 1 is used in the next function to complete a prefix l first with identifiable parameters relatively to l . For each non identifiable parameter relatively to l , if there exists, l is completed by this non identifiable parameter. A recursive call of the function `CompletionWithComputedBranches` is then used to complete this new prefix into branches of the identifiability tree.

Note that, the function `CompletionWithComputedBranches` is called as soon as a prefix is constructed in order to complete it, if this is possible, without any semialgebraic emptiness test.

```

Function ConstructTheBranches( $S, \{\theta_1, \dots, \theta_m\}, 0., \mathcal{T}$ );
.  /* Inputs: .the system  $S$ 
.             .the set  $\{\theta_1, \dots, \theta_m\}$  of parameters of system  $S$ 
.             .a tuple  $l$  of  $\{\theta_1, \dots, \theta_m\}$ 
.             .a set of branches  $\mathcal{T}$  of the identifiability tree
.  Output: The set  $\mathcal{T}$  of all the branches of the identifiability tree already computed and
.          the one(s) admitting  $l$  as prefix
.   $SPP := \{\theta_1, \dots, \theta_m\} \setminus Set(l)$  ;
.  /*SPP is the set of possible parameters to complete the prefix  $l$  */
.   $hbc := \text{False}$  ;
.  /*hbc is set to True when  $l$  is completed by function CompletionWithComputedBranches*/
.  If  $SPP \neq \emptyset$  then
.  .  For  $\theta \in SPP$  do
.  . .  If  $hbc = \text{False}$  and  $SetOfRealSolutions(S \cup \{v * (\theta - \theta') = 0\}) = \emptyset$  then
.  . . .  /* The parameter  $\theta$  is identifiable relatively to  $l$  (See Corollary 1) */
.  . . .   $l := l \text{ cat } [\theta]$ ;
.  . . .   $[hbc, \mathcal{T}] := \text{CompletionWithComputedBranches}(l, \mathcal{T})$ ;
.  . . .  /* If Remark B can be applied,  $l$  is completed with the branches of  $\mathcal{T}$  already computed */
.  . . .   $SPP := SPP \setminus \{\theta\}$ ;
.  . . .   $S := S \cup \{\theta = \theta'\}$ ;
.  . .  End if
.  .  End for
.  End If
.  /* SPP contains now only non identifiable parameters relatively to  $l$ . */
.  If  $SPP = \emptyset$  then
.  .   $\mathcal{T} = \mathcal{T} \cup \{l\}$ 
.  .  /* In this case,  $l$  is a branch of the identifiability tree */
.  else
.  .  For  $\theta \in SPP$  do
.  . .  /* Adjoining to  $\mathcal{T}$  the set of branches admitting  $l \text{ cat } [\theta]$  as prefix*/
.  . . .   $SPP := SPP \setminus \{\theta\}$ ;
.  . . .   $[hbc, \mathcal{T}] := \text{CompletionWithComputedBranches}(l \text{ cat } [\theta], \mathcal{T})$ ;
.  . . .  If  $hbc = \text{False}$  then
.  . . . .  /*  $l \text{ cat } [\theta]$  can not be completed with the branches of  $\mathcal{T}$  already computed*/
.  . . . .   $\mathcal{T} := \mathcal{T} \cup \text{ConstructTheBranches}(S \cup \{\theta = \theta'\}, \{\theta_1, \dots, \theta_m\}, l \text{ cat } [\theta], \mathcal{T})$ ;
.  . . . .  End if
.  . .  End for
.  End If
.  Return  $\mathcal{T}$ ;
End Function ;

```

This last function returns the identifiability tree from System S and from the set of parameters $\{\theta_1, \dots, \theta_m\}$.

```

Function IdentifiabilityTree( $S, \{\theta_1, \dots, \theta_m\}$ );
.  /* Inputs: The system  $S$  and the set of all the parameters  $\{\theta_1, \dots, \theta_m\}$  */
.  /* Output: The set of branches of the identifiability tree */
.   $l := []$ ;
.   $\mathcal{T} := \{[]\}$ ;
.  Return ConstructTheBranches( $S, \{\theta_1, \dots, \theta_m\}, l, \mathcal{T}$ );
End Function ;

```