



HAL
open science

Identifiability and Observability in Epidemiological Models - a survey -

Nik Cunniffe, Frédéric Hamelin, Abderrahman Iggidr, Alain Rapaport,
Gauthier Sallet

► **To cite this version:**

Nik Cunniffe, Frédéric Hamelin, Abderrahman Iggidr, Alain Rapaport, Gauthier Sallet. Identifiability and Observability in Epidemiological Models - a survey -. 2023. hal-02995562v3

HAL Id: hal-02995562

<https://hal.science/hal-02995562v3>

Preprint submitted on 25 May 2023

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.



Distributed under a Creative Commons Attribution 4.0 International License

N. Cunniffe, F. Hamelin, A. Iggidr,
A. Rapaport, G. Sallet

Identifiability and Observability in Epidemiological Models

– a survey –

May 19, 2023

Springer Nature

The World requires at least ten years to understand a new idea, however important or simple it may be.

Ronald Ross (1902 Nobel Prize).

Preface

In Mathematical Epidemiology, many papers present the following structure:

- a model is proposed,
- some parameters are given, extracted from literature,
- remaining unknown parameters are estimated by fitting the model to some observed data.

Fitting is done usually by using an optimization algorithm with the use for example of a least square method or a maximum likelihood estimation. To validate the parameters estimation, one can use noisy synthetic simulated data obtained from the model for given values of the parameters, to check that the algorithm is able to reconstruct from the data the values of these parameters with accuracy.

One objective of this book is to show that this procedure is not always safe and that an examination of the identifiability of parameters is a prerequisite before a numerical determination of parameters. We will review different methods to study identifiability and observability and then consider the problem of numerical identifiability. Our touchstone will be the most famous, however simple, model in Mathematical Epidemiology, the SIR model of Kermack and Mckendrick [122]. This model gets a renewed attention with the COVID-19 pandemic [79, 178]. Parameter identifiability analysis addresses the problem of which unknown parameters of an ODE model can uniquely be recovered from observed data. We will show that, even for very simple models, identifiability is far from being guaranteed.

The problem of identifiability for epidemiological models is relatively rarely addressed. For instance, a research in the Mathematical Reviews of the American Mathematical Society¹ en 2020 with `epid* AND identifiability` gives only 4 papers, while `epidem* AND parameter` returns 68 publications. Only a small part of the later publications address the problem of identifiability. In particular, the following publications consider the problem of identifiability in epidemiological models: [12, 16, 27, 46, 47, 80, 84, 112, 132, 147, 153, 163, 182, 181, 203, 204, 205,

¹ <https://mathscinet.ams.org/mathscinet>

210, 221, 223]. However the majority of these papers is published elsewhere than in Biomathematics journals.

The question of observability, i.e. the ability to reconstruct state variables of the model from measurements, is often considered separately from the problem of identifiability. Either model parameters are known, or the identifiability analysis is performed prior to the study of observability. Indeed, the concepts of identifiability and observability are closely related, as we show in this book. However, for certain models, it is possible to reconstruct state variables with observers, while the model is not identifiable. In other situations, we show that considering jointly identifiability and observability with observers can be a way to solve the identifiability problem. This is another illustration of the interest of observers. This is why we shall dedicate a fair part of this monograph reviewing the concept of observers and their practical constructions in epidemiology.

This book is aimed at scientists, researchers, graduate students, who use or develop mathematical models for epidemiology, and who are not yet familiar with the concepts of control science (detectability, observability, observers) applied to this field.

Cambridge,
Rennes,
Metz,
Montpellier,

Nik Cunniffe
Frédéric Hamelin
Abderrahman Iggidr, Gauthier Sallet
Alain Rapaport
May 2023

Acknowledgements

The authors are deeply grateful to P.-A. Bliman, C. Lobry, J. Harmand, T. Sari, M. Sofonea, M. Souza and many other colleagues or students, for exchanges and fruitful discussions that gave them the willingness to write this monograph.

Contents

1	Introduction	1
1.1	Definitions	1
1.2	Historical notes	7
1.3	Identifiability in mathematical epidemiology	7
1.4	The concept of observers	7
2	Mathematical foundations	9
2.1	Preliminaries	9
2.2	Observability	10
2.3	About identifiability	19
2.4	Identifiability does not necessarily require observability	21
2.5	Identifiability via decoupled variables	22
3	Analysis of the Kermack-McKendrick model	27
3.1	History	27
3.2	The different forms of the SIR model	27
3.3	Observability and identifiability of the SIR model	28
3.3.1	The SIR model when observing a ratio of the infected population	29
3.3.2	The SIR model when observing the incidence	34
4	Observers synthesis	37
4.1	Introduction	37
4.2	Observers with linear error dynamics	40
4.3	Observers for systems with Lipschitz non-linearity	42
4.4	Observers via decoupled variables	45
4.5	Reduced-order observers	46
4.6	The high-gain observer for nonlinear systems	48
4.7	Discussion	52

5	Practical and numerical considerations	55
5.1	Practical identifiability	55
5.1.1	Rationale for using sensitivity analysis	56
5.1.2	Observed system	56
5.1.3	Sensitivity analysis	56
5.1.4	Ordinary Least Squares	57
5.1.5	Confidence intervals	58
5.1.6	Computing the sensitivity matrix	58
5.1.7	Some case studies	60
5.1.8	Discussion	63
5.2	Observers in practice	63
5.2.1	Observers with linear assignable error dynamics	64
5.2.2	About observers with Lipchitz non-linearity	67
5.2.3	Observers with asymptotic convergence	67
5.2.4	Observers with partially assignable error dynamics	68
5.2.5	High gain observer	69
5.3	A case study : An observer to estimate state and parameter	71
A	Proofs of some useful lemmas	77
A.1	Proof of Lemma 4.1	77
A.2	Proof of Lemma 4.2	80
A.3	Proof of Theorem 4.1	81
B	Implementation of the “Boarding School” example	83
B.1	Derivation of the Fisher Information Matrix	83
B.2	Numerical implementation	85
C	Implementation of the “Plague in Bombay” example	89
C.1	Derivation of the Fisher Information Matrix	89
C.2	Numerical implementation	91
D	Generalized Least Squares	97
	References	99

Chapter 1

Introduction

Abstract We introduce the basic concepts and present the structure of the book.

1.1 Definitions

The question of parameter identifiability originates from control theory and is related to observability and controllability concepts [189]. The first appearance is in Kalman [119] and is now sixty years old. Identifiability is related to observability: the observability of a model is the ability to reconstruct the state of a system from the observation (we give a precise definition below). In the language of dynamical systems with inputs and outputs, which is the standard paradigm in control systems theory, an input-output relation is defined. Typically, models take the form

$$\begin{cases} \frac{dx(t)}{dt} = f(x(t), u(t)), \\ y(t) = h(x(t)) \end{cases}$$

where $u(t)$ and $y(t)$ are respectively "input" and "output" vectors at time t , and the vector $x(t)$ represents the internal variables of the model at time t . The inputs, represented by the function $u(\cdot)$, also called "controls" or "control variables" are considered as known. The outputs, represented by the function $y(\cdot)$, are called "observations" or "measurements" and are also considered as known. For simplicity, we will only consider systems without control (which is a peculiar case where the controls take values in a singleton...). When controls are known, with more information, observability/identifiability is sometimes easier. These problems have rarely been considered for uncontrolled systems whereas many methods have been developed for controlled systems. To be more precise, let us consider a dynamical system defined in a open domain $\mathcal{D} \subset \mathbb{R}^n$

$$\Sigma : \begin{cases} \dot{x}(t) = f(x(t)), & x(0) = x_0 \in \mathcal{D}, \\ y(t) = h(x(t)), \end{cases} \quad (1.1)$$

where we have denoted $\dot{x}(t) = \frac{dx(t)}{dt}$. The ordinary differential equation (ODE) $\dot{x} = f(x)$ is the dynamics and x is called the state of the system. To avoid technical details we will assume in all the book that for any initial condition x_0 in \mathcal{D} , there exists a unique solution denoted $x(t, x_0)$ such that $x(0, x_0) = x_0$ and

$$\frac{dx(t, x_0)}{dt} = f(x(t, x_0)).$$

Moreover, we will assume that this solution $x(t, x_0)$ is defined for any time $t \geq 0$, and we shall consider a connected subset Ω of \mathcal{D} of non-empty interior that is a positively invariant, which means that for any initial condition $x_0 \in \Omega$, the solution $x(t, x_0)$ belongs to this set for any $t \geq 0$. This is often the case with epidemiological models for which the state vector x naturally evolves in a compact and connected invariant set with non-negative vectors. This situation is also often encountered in biological systems. Throughout the manuscript, Ω will then denote the state space.

The output (or ‘‘observation’’) of the system is given by $h(x)$ where h is a differentiable function $h : x \in \Omega \subset \mathbb{R}^n \mapsto h(x) \in Y \subset \mathbb{R}^m$. The set Y is the output space. We will denote by $h(t, x_0)$ or $y(t, x_0)$ the observation at time t for an initial condition x_0 .

Definition 1.1 (Observability) The system (1.1) is observable on Ω if for two distinct initial states x_0, x'_0 in Ω , there exists a time $t \geq 0$ such that

$$h(x(t, x_0)) \neq h(x(t, x'_0))$$

This is equivalent to state

$$\left\{ h(x(t, x_0)) = h(x(t, x'_0)), \forall t \geq 0 \right\} \Rightarrow x_0 = x'_0.$$

Two states x_0, x'_0 in Ω are called indistinguishable if we have

$$h(x(t, x_0)) = h(x(t, x'_0)), \quad t \geq 0$$

Indistinguishability means that it is impossible to distinguish the evolution of the system, from two distinct initial conditions, by considering only the observation $y(\cdot)$.

Note that the observability property may depend on the choice of the (positively) invariant set Ω . Consider for instance the system

$$\begin{cases} \dot{x}_a = -x_a x_b \\ \dot{x}_b = x_a x_b \\ y = x_b \end{cases}$$

It is not observable for $\Omega = \{x_a \geq 0, x_b \geq 0, x_a + x_b \leq 1\}$ because all initial conditions of the form $(x_a, 0)$ cannot be distinguished, while it is observable for $\Omega = \{x_a > 0, x_b > 0, x_a + x_b \leq 1\}$. If two solutions $(x_a(\cdot), x_b(\cdot)), (x'_a(\cdot), x'_b(\cdot))$ generate the same output $y(\cdot)$, one should have

$$\begin{cases} y(t) = x_b(t) = x'_b(t), \\ \dot{y}(t) = x_a(t)y(t) = x'_a(t)y(t), \end{cases} \quad t \geq 0$$

but solutions in Ω are such that $y(t) \neq 0$ at any $t \geq 0$, which implies $x_a(t) = x'_a(t)$ and thus $(x_a(0), x_b(0)) = (x'_a(0), x'_b(0))$.

Sometimes, we shall simply say that "the system is observable", when $\Omega = \mathcal{D}$ or when there is no ambiguity about the choice of Ω .

Now we consider a system depending on a parameters vector $\theta \in \Theta \subset \mathbb{R}^p$ (sometimes simply called the "parameter θ ")

$$\begin{cases} \dot{x}(t) = f(x(t), \theta), x(0) = x_0, \\ y(t) = h(x(t), \theta). \end{cases} \quad (1.2)$$

We denote by $x(t, x_0, \theta)$ the solution of (1.2) for an initial condition x_0 . Then, the definition of observability needs to be adapted as follows

Definition 1.2 (Observability for parameterized systems) The system (1.2) is observable on Ω if whatever are $\theta \in \Theta$ and two distinct initial conditions x_0, x'_0 in Ω , there exists a time $t \geq 0$ such that

$$h(x(t, x_0, \theta)) \neq h(x(t, x'_0, \theta))$$

or equivalently

$$\left\{ h(x(t, x_0, \theta)) = h(x(t, x'_0, \theta)), \forall t \geq 0 \right\} \Rightarrow x_0 = x'_0.$$

Identifiability is the ability to recover the unknown parameter from the observation, when the initial condition is know.

Definition 1.3 (Identifiability) Given an initial state x_0 in Ω , system (1.2) is said to be identifiable if for any distinct θ_1, θ_2 in Θ , there exists $t \geq 0$ such that

$$h(x(t, x_0, \theta_1)) \neq h(x(t, x_0, \theta_2)).$$

The concepts of observability and identifiability are similar. Consider the augmented system which consists in adding the parameter θ as part of the (augmented) state vector with a null dynamics:

$$\begin{cases} \dot{x}(t) = f(x(t), \theta), \\ \dot{\theta} = 0, \\ y(t) = h(x(t), \theta). \end{cases} \quad (1.3)$$

Proposition 1.1 *If the system (1.3) is observable on $\Omega \times \Theta$, then system (1.2) is observable on Ω and identifiable for any $x_0 \in \Omega$.*

Proof If system (1.3) is observable then we have the property: $h(x(t, (x_0, \theta_1))) = h(x(t, (x'_0, \theta_2)))$ for all $t \geq 0$ implies $(x_0, \theta_1) = (x'_0, \theta_2)$, which implies $x_0 = x'_0$ and $\theta_1 = \theta_2$. \square

The converse is not true as it is illustrated by the following example:

$$\begin{cases} \dot{x}_1 = x_1 + \theta x_2 \\ \dot{x}_2 = 0 \\ y = x_1 \end{cases} \quad (1.4)$$

System (1.4) is observable and identifiable.

$$\begin{cases} \dot{x}_1 = x_1 + \theta x_2 \\ \dot{x}_2 = 0 \\ \dot{\theta} = 0 \\ y = x_1 \end{cases} \quad (1.5)$$

System (1.5), whose state is (x_1, x_2, θ) , is not observable since the output is the same for initial conditions (x_1, x_2, θ) and $(\bar{x}_1, \bar{x}_2, \bar{\theta})$ satisfying $x_1 = \bar{x}_1$ and $x_2\theta = \bar{x}_2\bar{\theta}$.

At several places in the following, we shall consider this augmented dynamics.

Actually, for an epidemiological model it is unlikely to know the initial condition and it has long been recognized that initial conditions play a role in identifying the parameters [73, 134, 202, 223, 184].

What we have called identifiability is also known as *structural identifiability*. This expression has been coined by R. Bellman and K.J. Åström [26] in 1970. This is to stress that identifiability depends only on the dynamics and the observation, under ideal conditions of noise-free observations and error-free model. This is a mathematical and *a priori* problem [112].

Sometimes, it happens that the concept of observability is too strong in the sense that a system in the form (1.1) does not satisfies the observability property but it is still

possible to know part of the initial condition. A relaxed definition of observability is as follows.

Definition 1.4 (Partial observability) Let φ be a smooth map from Ω to $\varphi(\Omega) \subset \mathbb{R}^{n'}$. We say that the system (1.1) is partially observable on Ω with respect to φ if for two initial states x_0, x'_0 in Ω with $\varphi(x_0) \neq \varphi(x'_0)$, there exists a time $t \geq 0$ such that

$$h(x(t, x_0)) \neq h(x(t, x'_0)).$$

Two states x_0, x'_0 in Ω are called indistinguishable with respect to φ if one has $\varphi(x_0) \neq \varphi(x'_0)$ and

$$h(x(t, x_0)) = h(x(t, x'_0)), \quad t \geq 0$$

Alternatively, we shall say the function φ is observable on Ω .

In a similar way, one can relax the definition of identifiability as follows.

Definition 1.5 (Partial Identifiability) Let ϑ be a smooth map from Θ to $\vartheta(\Theta) \subset \mathbb{R}^{p'}$. Given an initial state x_0 in Ω , system (1.2) is said to be identifiable with respect to ϑ if for any θ_1, θ_2 in Θ with $\vartheta(\theta_1) \neq \vartheta(\theta_2)$, there exists $t \geq 0$ such that

$$h(x(t, x_0, \theta_1)) \neq h(x(t, x_0, \theta_2)).$$

Alternatively, we shall say the parameter function ϑ is identifiable.

Example 1.1 Consider the system in \mathbb{R}^3

$$\begin{cases} \dot{x}_a = -\alpha x_b \\ \dot{x}_b = -\beta x_c \\ \dot{x}_c = \gamma \end{cases}$$

where α, β, γ are parameters. Solutions of this system can be made explicit:

$$\begin{aligned} x_a(t) &= x_a(0) - \alpha x_b(0)t + \frac{\alpha\beta}{2}x_c(0)t^2 + \frac{\alpha\beta\gamma}{6}t^3, \\ x_b(t) &= x_b(0) - \beta x_c(0)t - \frac{\beta\gamma}{2}t^2, \\ x_c(t) &= x_c(0) + \gamma t. \end{aligned}$$

Here, we consider $\Omega = \mathcal{D} = \mathbb{R}^3$ and $\Theta = \{(\alpha, \beta, \gamma); \alpha > 0, \beta > 0, \gamma > 0\}$. Let us show that the system is observable for the observation map $h(x_a, x_b, x_c) = x_a$. We consider two initial conditions $(x_a(0), x_b(0), x_c(0)), (x'_a(0), x'_b(0), x'_c(0))$ such that their respective solutions verify $x_a(t) = x'_a(t)$ for any $t \geq 0$. From the expression of the solution $x_a(\cdot)$, this implies that one has

$$-\alpha x_b(0)t + \frac{\alpha\beta}{2}x_c(0)t^2 = -\alpha x'_b(0)t + \frac{\alpha\beta}{2}x'_c(0)t^2$$

for any $t \geq 0$, and for $t > 0$ one gets by differentiating with respect to t

$$-\alpha x_b(0) + \frac{\alpha\beta}{2}x_c(0)t = -\alpha x'_b(0) + \frac{\alpha\beta}{2}x'_c(0)t$$

When $t > 0$ tends to 0, one obtains $x_b(0) = x'_b(0)$ and then $x_c(0) = x'_c(0)$. Therefore, the two initial conditions coincide and there is no indistinguishable distinct initial condition. The system is thus observable.

Consider now the observation map $h(x_a, x_b, x_c) = x_b$. When two initial conditions are such that their solutions verify $x_b(t) = x'_b(t)$, one gets from the expression of the solution $x_b(\cdot)$

$$-\beta x_c(0)t = -\beta x'_c(0)t$$

for any $t > 0$ and thus $x_c(t) = x'_c(t)$. However, having $x_a(0) \neq x'_b(0)$ with $x_b(t) = x'_b(t)$ and $x_c(t) = x'_c(t)$ gives exactly the same solution $x_b(\cdot), x_c(\cdot)$. The system is not observable but it is partially observable with respect to the map

$$\varphi(x_a, x_b, x_c) = (x_b, x_c).$$

Let us now study the identifiability of the system, first for the observation map $h(x_a, x_b, x_c) = x_a$. For a given initial condition with $x_b(0) \neq 0$ and $x_c(0) \neq 0$, if two sets of parameters $(\alpha_1, \beta_1, \gamma_1), (\alpha_2, \beta_2, \gamma_2)$ give the same solution $x_a(\cdot)$, one has

$$-\alpha_1 x_b(0)t + \frac{\alpha_1 \beta_1}{2} x_c(0)t^2 + \frac{\alpha_1 \beta_1 \gamma_1}{6} t^3 = -\alpha_2 x_b(0)t + \frac{\alpha_2 \beta_2}{2} x_c(0)t^2 + \frac{\alpha_2 \beta_2 \gamma_2}{6} t^3$$

for any $t \geq 0$, and for $t > 0$ one gets

$$-\alpha_1 x_b(0) + \frac{\alpha_1 \beta_1}{2} x_c(0)t + \frac{\alpha_1 \beta_1 \gamma_1}{6} t^2 = -\alpha_2 x_b(0) + \frac{\alpha_2 \beta_2}{2} x_c(0)t + \frac{\alpha_2 \beta_2 \gamma_2}{6} t^2$$

When $t > 0$ tends to 0, one obtains $\alpha_1 = \alpha_2$ and thus one has

$$\frac{\alpha_1 \beta_1}{2} x_c(0)t + \frac{\alpha_1 \beta_1 \gamma_1}{6} t^2 = \frac{\alpha_2 \beta_2}{2} x_c(0)t + \frac{\alpha_2 \beta_2 \gamma_2}{6} t^2$$

for any $t > 0$. In a similar way, one can show that one has $\alpha_1 \beta_1 = \alpha_2 \beta_2$ and then $\alpha_1 \beta_1 \gamma_1 = \alpha_2 \beta_2 \gamma_2$, which proves that one has $\beta_1 = \beta_2$ and $\gamma_1 = \gamma_2$ (remind that parameters are assumed to be non null). The system is thus identifiable.

When the observation map is $h(x_a, x_b, x_c) = x_b$, having the same output $x_b(\cdot)$ implies

$$-\beta_1 x_c(0)t - \frac{\beta_1 \gamma_1}{2} t^2 = -\beta_2 x_c(0)t - \frac{\beta_2 \gamma_2}{2} t^2$$

for any $t \geq 0$. As before, one obtains $\beta_1 = \beta_2$ and $\gamma_1 = \gamma_2$ but having $\alpha_1 \neq \alpha_2$ with $\beta_1 = \beta_2$ and $\gamma_1 = \gamma_2$ provides the same output $x_b(\cdot)$. Indeed the dynamics of x_b and x_c are decoupled from the one of x_a , where parameter α appears only. The system is not identifiable but is partially identifiable with respect to the map $\vartheta(\alpha, \beta, \gamma) = (\beta, \gamma)$.

1.2 Historical notes

The observability concept has been introduced by Kalman [119] in the sixties for linear systems. For nonlinear systems, observability has been characterized circa the seventies [97, 102]. The definition is given by Hermann and Krener in the framework of differential geometry. Identifiability and structural identifiability has been introduced in compartmental analysis in 1970 by Bellman and Åström [26] in a paper that appeared in a bio-mathematics journal. The problem of identifiability is now addressed in text-books [133, 216, 214, 213]. Numerical identifiability of linear control system is implemented in softwares such as `Matlab` and `Scilab`.

Identifiability of nonlinear systems has been addressed in different context and the first systematic approach is by Tunali and Tarn in 1987 [202] in the differential geometry framework. The introduction of the concepts of differential algebra in control theory is due to Fliess around 1990 [73, 74, 86] followed by Glad [88, 134]. Identifiability is a general problem which has received different names depending on the community:

- observation, identification,
- data assimilation,
- inverse problem,
- parameters estimation.

“Data assimilation” is mainly used in meteorology and oceanography [127, 197]. A direct (as opposed to inverse) problem is considering a model which, when introducing an input, gives an observed output. The parameters are considered as known. Conversely the “inverse problem” is to reconstruct the parameters from the knowledge of the output [199]. Finally, “parameters estimation” is used in the probability and statistics domains [5, 32, 128, 130, 201, 157, 177].

1.3 Identifiability in mathematical epidemiology

Identifiability is well known in bio-mathematics from the seventies, as already mentioned with the paper of Bellman and Åström [26]. However, considering identifiability in mathematical epidemiology is relatively recent [203, 163, 223, 182, 147, 80, 84]. The first paper, to our knowledge, considering identifiability of an intra-host model of HIV is by Xia and Moog [223], and has been published in 2003 in a journal of automatic control.

1.4 The concept of observers

The construction of an *observer* is based on an estimation approach different from statistical methods: it consists of determining a dynamical system (called an “ob-

server”) whose input is the vector $y(\cdot)$ of measures acquired over time, and whose state is an estimate $\hat{x}(t)$ of the (unknown) true state $x(t)$ of the system at time t .

An observer estimates $x(t)$ continuously over time and without anticipation, in the sense that the estimate $\hat{x}(t)$ is updated at each instant t through its dynamics as measurement $y(t)$ is available, without requiring the knowledge of any future measurement. This is why an observer is sometimes also called a “software sensor”. Since the estimate $\hat{x}(t)$ is given by the solution of a system of differential equations, the main idea behind an observer is the paradigm of *integrating instead of differentiating* the signal $y(\cdot)$. Note that although an observer is primarily devoted to state estimation, an observer can also aim to reconstructing simultaneously state and parameters, when some parameters are unknown (in this case a parameter vector p is simply considered a part of the system dynamics with $\dot{p} = 0$).

The most well-known observer is the so-called *Luenberger observer* [135] that is recalled in Chapter 4, and that has inspired most of the existing observers (several ones are discussed in Chapter 4). However, observers are yet relatively unpopular in Mathematical Epidemiology, comparatively to other application domains (such as mechanics, aeronautics, automobile, *etc*). The aim of the present review is also to promote the development and use of observers for epidemiological models.

Chapter 4 presents the theoretical background of observers construction and their convergence as estimators based on the model equations, independently of the quality of real data. In a complementary way, Chapter 5 discusses some implementation issues when observers are used with real world data that could be corrupted with noise.

Chapter 2

Mathematical foundations

Abstract We give mathematical characterizations and properties of observability and identifiability.

2.1 Preliminaries

Here and in all the following chapters, we shall consider that the maps f and g that define the system

$$\begin{cases} \dot{x} = f(x), & x \in \Omega \subset \mathbb{R}^n, \\ y = h(x) \in Y \subset \mathbb{R}^m \end{cases} \quad (2.1)$$

are analytic at any point $x \in \Omega$. We shall denote by $x(t, x_0)$ the solution of $\dot{x} = f(x)$ for the initial condition $x(0) = x_0$.

We recall that a function $\varphi : \mathcal{D} \mapsto \mathbb{R}$ is analytic on a open domain \mathcal{D} of \mathbb{R}^n (we will write $\varphi \in C^\omega(\mathcal{D}, \mathbb{R})$) if it is C^∞ (i.e. infinitely differentiable) and its Taylor series locally converge, that is for any x_0 in \mathcal{D} there exists a neighborhood \mathcal{V} of x_0 in \mathcal{D} such that

$$\varphi(x) = \lim_{n \rightarrow +\infty} \sum_{k=0}^n \frac{\varphi^{(k)}(x_0)}{k!} (x - x_0)^k, \quad x \in \mathcal{V}.$$

Indeed, up to our knowledge, the great majority of epidemiological model in the literature are analytic.

The map $f : \Omega \rightarrow \mathbb{R}^n$ is also called a *vector field* on Ω , as its image for each $x \in \Omega$ is a (velocity) vector in \mathbb{R}^n .

Given a C^∞ function $g : \mathbb{R}^n \mapsto \mathbb{R}$, the classical definition of Lie derivative of g with respect to the vector field f is given by

$$\mathcal{L}_f(g)(x) = \left. \frac{d}{dt} g(x(t, x)) \right|_{t=0} = \langle \nabla g(x) | f(x) \rangle, \quad x \in \Omega \quad (2.2)$$

where ∇g is the gradient of g and $\langle | \rangle$ the inner product of \mathbb{R}^n . For $k > 1$, we define by induction

$$\mathcal{L}_f^k(g)(x) = \mathcal{L}_f(\mathcal{L}_f^{k-1}(g))(x)$$

Note that $\mathcal{L}_f^k(h)(x)$ is the value of the k -th time derivative at time 0 of the output $y(t) = h(x(t, x))$ along the solution of system (2.1) with initial state x , but considered as function of x :

$$\begin{aligned} \mathcal{L}_f(h)(x) &= \dot{y}(x), \\ \mathcal{L}_f^2(h)(x) &= \ddot{y}(x), \\ &\vdots \\ \mathcal{L}_f^k(h)(x) &= y^{(k)}(x) \end{aligned}$$

For a C^∞ vector-valued function $g : \mathbb{R}^n \mapsto \mathbb{R}^q$, the Lie derivative $\mathcal{L}_f(g)(x)$ at point x is the vector of Lie derivatives of each component g_j ($j = 1 \cdots q$)

$$\mathcal{L}_f(g)(x) = \begin{bmatrix} \mathcal{L}_f(g_1)(x) \\ \vdots \\ \mathcal{L}_f(g_q)(x) \end{bmatrix}$$

When f and g are linear, that is $f(x) = Ax$ and $g(x) = Cx$, then $\mathcal{L}_f^k(g)(x)$ is equal to $CA^k x$.

We shall denote $^\top$ the transposition operator for a vector or a matrix.

2.2 Observability

In this section, we introduce the *observation space* and the concept of *local observability* that can be checked with the help of differential calculus in terms of a (local) *rank condition* to be checked. For linear systems, this rank condition depends only on the matrices defining the system and the output and ensures the observability property as defined in Section 1.1. For analytic nonlinear systems, we shall see that this rank condition depends on the vector field and the output function as well as on the initial condition x_0 and only ensures a "local" observability of the system on a neighborhood of x_0 .

The observability definition 1.1 given in Chapter 1 states a global observability property. It also can be formulated as follows.

Definition 2.1 (Observability) The analytic system (2.1) is observable if for any initial conditions x_1 and x_2 , for any $T > 0$, one has: $h(x(t, x_1)) = h(x(t, x_2))$ for all $t \in [0, T]$ implies $x_1 = x_2$. This is equivalent to say that the map:

$$\begin{aligned} \Omega &\rightarrow C^\omega([0, T], \mathbb{R}^m) \\ x_0 &\mapsto \left\{ \begin{array}{l} [0, T] \rightarrow \mathbb{R}^m \\ t \mapsto h(x(t, x_0)) \end{array} \right\} \end{aligned}$$

is injective.

Observability means that the initial state x_0 (and therefore the trajectory starting from this initial state) can be uniquely determined by the knowledge of the data of the output $y(\cdot)$ on any nontrivial time interval.

The components of the observation map h are denoted by $h = (h_1, \dots, h_m)$. Each h_i is a C^∞ function from the state space \mathbb{R}^n to \mathbb{R} .

Definition 2.2 ([102]) The observation space \mathcal{O} of (2.1) is the subspace of the vector space $C^\infty(\mathbb{R}^n, \mathbb{R})$ containing h_i and invariant by the Lie derivative \mathcal{L}_f .

The observation space is generated by the different Lie derivatives of the h_i :

$$\begin{aligned} \mathcal{O} &= \text{span}_{\mathbb{R}} \left\{ \mathcal{L}_f^k h_i : i = 1, \dots, m, k \in \mathbb{N} \right\} \\ &= \text{span}_{\mathbb{R}} \left\{ h_i, \mathcal{L}_f h_i, \mathcal{L}_f^2 h_i, \dots : i = 1, \dots, m \right\}. \end{aligned}$$

The observation space \mathcal{O} contains the observation function (also called the output function) and all derivatives of the output function along the system trajectories. For a linear system: $f(x) = Ax$ and $y = h(x) = Cx$, the observation space is generated by the n functions

$$Cx, CAx, \dots, CA^{n-1}x.$$

We recall that thanks to Cayley-Hamilton Theorem A^k for $k \geq n$ is a linear combination of A^q with $q \leq n - 1$.

We have the following result that relates the observation space to the observability property.

Theorem 2.1 For an analytic system (i.e., f and h are analytic functions) the observability is equivalent to the separation of the points of the state space \mathbb{R}^n by \mathcal{O} i.e., if $x_1 \neq x_2$ there exists $g \in \mathcal{O}$ such that $g(x_1) \neq g(x_2)$.

Proof By analyticity we have

$$y(t, x_0) = h(x(t, x_0)) = \sum_{k \geq 0} \left(\frac{d^k}{ds^k} h(x(s, x_0)) \Big|_{s=0} \right) \frac{t^k}{k!},$$

but, by induction we have the following relation

$$\frac{d^k}{ds^k} h(x(s, x_0)) \Big|_{s=0} = \mathcal{L}_f^k h(x_0).$$

Then a necessary and sufficient condition to distinguish $x_1 \neq x_2$ is that there exists k such that

$$\mathcal{L}_f^k . h(x_1) \neq \mathcal{L}_f^k . h(x_2).$$

□

Example 2.1 (Application to a virus dynamics model) We consider a simple model of an HIV-1 infection [164]:

$$\begin{cases} \dot{T} = \Lambda - \mu_T T - \beta VT, \\ \dot{T}^* = \beta VT - \delta T^*, \\ \dot{V} = r \delta T^* - c V, \\ y = V, \end{cases} \quad (2.3)$$

where T , T^* , V denote the concentrations of uninfected (healthy) and infected host cells, and freevirions, respectively. The rate of infection is given by βVT , with β being the infection rate constant. The parameters δ and c are the removal rates of the infected cells and virus particles respectively. The healthy cells (T) are produced at a rate Λ and μ_T is the death rate per T cell. It is assumed that on average each productively infected cell produces r (a positive integer) virions during its lifetime, so the per-capita rate of viral production for an infected cell is given by $r \delta$. All parameters are positive. We assume that the measurement of the viral load is available.

Let us show that system (2.3) is observable on the set $\{(T, T^*, V) \in \mathbb{R}^3 : T \geq 0, T^* \geq 0, V > 0\}$. We shall prove that the observation space associated to System (2.3) separates the points of set $\{(T, T^*, V)^\top \in \mathbb{R}^3 : T \geq 0, T^* \geq 0, V > 0\}$.

Here we have $x = (T, T^*, V)^\top$ and

$$f(x) = \begin{bmatrix} \Lambda - \mu_T T - \beta VT \\ \beta VT - \delta T^* \\ r \delta T^* - c V \end{bmatrix}$$

and the measurable output $y = h(x) = V$. Computing the Lie-derivatives of the output gives

$$\mathcal{L}_f h(x) = \langle \nabla h(x) | f(x) \rangle = \left\langle \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \middle| \begin{bmatrix} \Lambda - \mu_T T - \beta VT \\ \beta VT - \delta T^* \\ r \delta T^* - c V \end{bmatrix} \right\rangle = r \delta T^* - c V$$

and

$$\begin{aligned} \mathcal{L}_f^2 h(x) &= \langle \nabla \mathcal{L}_f h(x) | f(x) \rangle = \left\langle \begin{bmatrix} 0 \\ r \delta \\ -c \end{bmatrix} \middle| \begin{bmatrix} \Lambda - \mu_T T - \beta VT \\ \beta VT - \delta T^* \\ r \delta T^* - c V \end{bmatrix} \right\rangle \\ &= r \delta \beta T V - r \delta (c + \delta) T^* + c^2 V. \end{aligned}$$

The explicit expression of $\mathcal{L}_f^3 h(x)$ is quite long but it can be easily shown that $\mathcal{L}_f^3 h(x)$ can be expressed as a function of $h(x)$, $\mathcal{L}_f h(x)$, $\mathcal{L}_f^2 h(x)$ and the parameters of the system. So to study observability of System (2.3), one does not need to compute Lie-derivative of order higher than 2.

The observation space \mathcal{O} contains the functions

$$\begin{aligned} g_1(x) &= V, \\ g_2(x) &= r \delta T^* - c V, \\ g_3(x) &= r \delta \beta T V - r \delta (c + \delta) T^* + c^2 V. \end{aligned}$$

It is easy to show

$$\left\{ g_1(x) = g_1(\bar{x}), g_2(x) = g_2(\bar{x}), g_3(x) = g_3(\bar{x}) \right\} \implies V = \bar{V}, T^* = \bar{T}^*$$

and $T = \bar{T}$ if $V \neq 0$. Hence, these functions separate the points of $\mathbb{R}_+^3 \setminus \{(T, T^*, 0)\}$ but they do not separate points of the form $(T, T^*, 0)$ and $(\bar{T}, \bar{T}^*, 0)$ with $\bar{T} \neq T$. Thus System (2.3) is observable on the set $\{(T, T^*, V) \in \mathbb{R}^3 : T \geq 0, T^* \geq 0, V > 0\}$.

Applying Theorem 2.1 to linear systems

$$\begin{cases} \dot{x} = Ax, & x \in \mathbb{R}^n, \\ y = Cx, & y \in \mathbb{R}^m, \end{cases} \quad (\Sigma_L)$$

allows to obtain a simple algebraic necessary and sufficient condition for observability.

Proposition 2.1 *The linear system (Σ_L) is observable if and only if the observability matrix*

$$O_{(C,A)} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \quad (2.4)$$

is of full rank, i.e., $\text{rank } O_{(C,A)} = n$.

Indeed, using Theorem 2.1, we have the following successive equivalences:

$$\begin{aligned}
& \text{The linear system } (\Sigma_L) \text{ is not observable} \\
& \iff \\
& \text{Its observation space } \mathcal{O}_{\Sigma_L} \text{ does not separate the points of the state space } \mathbb{R}^n \\
& \iff \\
& \exists x_1 \neq x_2 : \forall g \in \mathcal{O}_{\Sigma_L}, g(x_1) = g(x_2) \\
& \iff \\
& \exists x_1 \neq x_2 : \forall i \in \{0, 1, \dots, n-1\}, CA^i x_1 = CA^i x_2 \\
& \iff \\
& \exists x_1 \neq x_2 : x_1 - x_2 \in \text{Ker } CA^i, \forall i \in \{0, 1, \dots, n-1\} \\
& \iff \\
& x_1 - x_2 \in \text{ker } C \cap \text{ker } CA \cap \text{ker } CA^2 \cap \dots \cap \text{ker } CA^{n-1} \\
& \iff \\
& \text{ker } \mathcal{O}_{(C,A)} \neq \{0\} \\
& \iff \\
& \text{rank } \mathcal{O}_{(C,A)} < n.
\end{aligned}$$

The result of Theorem 2.1 can be also reformulated as follows. For the sake of writing simplicity, we consider real scalar output. Extension to vector output is straightforward. Let us define, for $k \in \mathbb{N}$, the map: $x \mapsto H_k(x) = (h(x), \mathcal{L}_f h(x), \dots, \mathcal{L}_f^k h(x))^\top$. Then we have the following characterization of observability :

Proposition 2.2 [110] *Suppose System (2.1) is analytic. Then it is observable if and only if $H_k(x_1) = H_k(x_2)$, for all $k \in \mathbb{N}$, imply that $x_1 = x_2$.*

This is equivalent to say that the following real analytic mapping (from the state space to an infinite dimensional space)

$$x \mapsto H_\infty(x) = (h(x), \mathcal{L}_f h(x), \dots, \mathcal{L}_f^k h(x), \dots)^\top$$

is injective.

Remark 2.1 In general there is no value of k to stop. This is illustrated by the following example [110]:

$$\begin{cases} \dot{x} = -x, & x \in \mathbb{R} \\ y = h(x) = x \prod_{i=1}^{\infty} (1 - e^{x^2 - i^2})^i \end{cases} \quad (2.5)$$

It has been proved in [110] that

- h is analytic on \mathbb{R} ,
- system (2.5) is observable,
- $\forall k \in \mathbb{N}^*$, the equation $H_k(x) = H_k(0)$ has countably infinite solutions $x = 0$ and $x = \pm(k+i)$, $i = 0, 1, 2, \dots$.

Observability can also be checked using the following proposition.

Proposition 2.3 *If any state $x \in \Omega$ can be expressed as a function of the observation y and its time derivatives $y^{(k)} = \mathcal{L}_f^k h(x)$, that is there exists a map ϕ such that $x = \phi\left(h(x), \mathcal{L}_f h(x), \dots, \mathcal{L}_f^k h(x), \dots\right)$, then System (2.1) is observable on Ω .*

Proof Let $x_1 \in \Omega$ and $x_2 \in \Omega$ be such that $H_\infty(x_1) = H_\infty(x_2)$. Then, applying ϕ , we obtain $x_1 = x_2$ which proves that map H_∞ is injective on Ω . \square

Example 2.2 Consider the SIR model of Kermack-McKendrick [122]

$$\begin{cases} \dot{S} = -\beta \frac{S}{N} I, \\ \dot{I} = \beta \frac{S}{N} I - \gamma I, \\ \dot{R} = \gamma I \end{cases} \quad (2.6)$$

for which the parameters β, γ and N are assumed to be known, where the total population $N = S + I + R$ is constant since $\dot{N} = \dot{S} + \dot{I} + \dot{R} = 0$. We assume that the recovery $y = \gamma I$ is observed. Here we have

$$x = \begin{pmatrix} S \\ I \\ R \end{pmatrix}, \quad f(x) = \begin{pmatrix} -\beta \frac{S}{N} I \\ \beta \frac{S}{N} I - \gamma I \\ \gamma I \end{pmatrix}, \quad h(x) = \gamma I$$

Clearly, the solutions of the system (3.4) evolve in the positively invariant set (which makes biological sense)

$$\Omega = \{(S, I, R) \mid S > 0, I > 0, R > 0, S + I + R = N\}$$

on which one has $y \neq 0$. Then we have the relations (we can divide by y)

$$S = \frac{N}{\beta} \frac{\dot{y} + \gamma y}{y}, \quad I = \frac{y}{\gamma}, \quad R = N - \frac{N}{\beta} \frac{\dot{y} + \gamma y}{y} - \frac{y}{\gamma}$$

Therefore, the state vector x can be expressed as a function of y and \dot{y} . The system is thus observable in Ω . This model is more thoroughly studied in Chapter 3.

The observability (equivalent) conditions given by Theorem 2.1 and Proposition 2.2 are hard to test, since they involve an infinite number of analytic equations and therefore checking the condition that \mathcal{O} separates points can be quite a formidable task. This is one of the reasons for which different concepts of local observability have been introduced. One of them is the local weak observability [102]. As we will see, the advantage of the local weak observability as compared to other types of observability is that it is easy to indicate for it simple sufficient conditions of an algebraic nature.

Definition 2.3

- The system (2.1) Σ is said locally observable if, for any x_0 , for any open set U containing x_0 , x_0 is distinguishable from all the points of U for the restricted system $\Sigma|U$.
- The system (2.1) is weakly observable at x if there exists an open neighborhood U of x such that the only point in U which is indistinguishable from x is x itself. The system (2.1) is weakly observable if it is weakly observable at every $x \in M$.
- The system (2.1) Σ is locally weakly observable if for any x_0 there exists an open set U containing x_0 , such that for any neighborhood V with $x_0 \in V \subset U$, x_0 is distinguishable for $\Sigma|V$ from all the points of V .

Intuitively, a system is locally weakly observable if one can instantaneously distinguish each point from its neighbors. The local weak observability can be characterized as follows.

Definition 2.4 [194] Let \mathcal{O} be the observation space of System (2.1), we define

$$d\mathcal{O} = \{d\psi \mid \psi \in \mathcal{O}\}$$

where $d\psi(x)$ is the differential of ψ at x .

Definition 2.5 A system Σ is said to satisfy the observability rank condition (ORC) at x if the dimension of $d\mathcal{O}$ at x satisfies

$$\dim(d\mathcal{O}(x)) = n$$

where $d\mathcal{O}$ is generated by the gradients of the $\mathcal{L}_f^k h$.

Theorem 2.2 (Hermann-Krener [102]) *If the analytic system (2.1) satisfies the observability rank condition (ORC) at x_0 then Σ is locally weakly observable at x_0 .*

Proof Since $\dim(d\mathcal{O}(x_0)) = n$, there exists n functions $\varphi_1, \dots, \varphi_n \in \mathcal{O}$ such that the gradients $d\varphi_1(x_0), \dots, d\varphi_n(x_0)$ are linearly independent. Therefore the function $\Phi : x \mapsto (\varphi_1(x), \dots, \varphi_n(x))$ has a non-singular Jacobian in x_0 . As a consequence, from the Inverse Function Theorem, there exists an open set U containing x_0 where Φ is a bijection.

On any open set $V \subset U$ suppose that we have $h(x(t, x_0)) = h(x(t, x_1))$ for $t \in [0, T]$. Then, from the fact that f and h are analytic and

$$\left. \frac{d^k}{dt^k} h(x(t, x_0)) \right|_{t=0} = \mathcal{L}_f^k h(x_0),$$

we have

$$h(x(t, x_0)) - h(x(t, x_1)) = \sum_{k \geq 0} \frac{t^k}{k!} \left(\mathcal{L}_f^k h(x_0) - \mathcal{L}_f^k h(x_1) \right) = 0, \text{ for } t \in [0, T].$$

This implies $\mathcal{L}_f^k h(x_0) - \mathcal{L}_f^k h(x_1) = 0$ for all $k \geq 0$ which implies that $\varphi_i(x_0) = \varphi_i(x_1)$ for all $i = 1, \dots, n$ since each φ_i is a linear combination of the $\mathcal{L}_f^k h$. Therefore $\Phi(x_0) = \Phi(x_1)$ and hence $x_1 = x_0$ since Φ is a bijection. This proves that x_0 is distinguishable from all points of V and hence the analytic system (2.1) is locally weakly observable at x_0 . \square

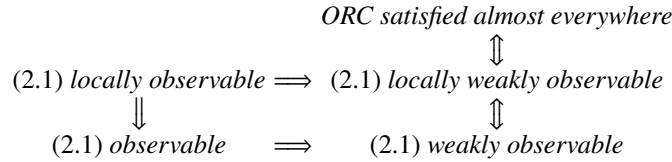
If the observability rank condition is satisfied everywhere the system is locally weakly observable.

A converse result has been proved in [102]:

Proposition 2.4 [102, 42] *If the system is locally weakly observable then the rank condition is satisfied almost everywhere, i.e., in an open dense subset of the state space.*

To summarize, we have for analytic systems:

Proposition 2.5 [102, 42] *For the analytic system (2.1), the relationships between the various observability concepts are given in the following diagram*



Remark 2.2 For linear systems, the five properties are equivalent.

Remark 2.3 The analytic system $\dot{x} = 0, y = x^3$ is observable on \mathbb{R} and weakly locally observable but the ORC is not satisfied at $x = 0$. However it is satisfied on $\mathbb{R} \setminus \{0\}$ and so it is satisfied almost everywhere.

Remark 2.4 If an analytic system is observable then the observability rank condition (ORC) is satisfied almost everywhere. It must be noticed that the converse is not true: the analytic system

$$\begin{cases} \dot{x} = 1, x \in \mathbb{R} \\ y = (\sin x, \cos x) \in \mathbb{R}^2 \end{cases}$$

satisfies the ORC at any $x \in \mathbb{R}$ but it is not observable because the states x and $x + 2k\pi$ are indistinguishable.

It must be emphasized that the study of the observability of analytic systems requires either to deal with an infinite number of analytic equations or to compute the dimension of the linear space generated by the gradients of all Lie derivatives of the output. In general, there is no bound on the number of Lie derivatives necessary to conclude, as seen in example (2.5).

Remark 2.5 For linear systems $\dot{x} = Ax$, $y = Cx$, all the definitions of observability are equivalent to having the *observability matrix*

$$O_{(C,A)} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}$$

of full rank (see for instance [117]).

The observability analysis can also be a way to choose the right sensor, as illustrated on the following example.

Example 2.3 Consider a population model structured in five age classes, whose population sizes are

- x_1 for juveniles,
- x_2 for subadults capable of reproduction when adults,
- x_3 for subadults not capable of reproduction when adults,
- x_4 for adults capable of reproduction,
- x_5 for adults not capable of reproduction,

and the dynamics is

$$\begin{cases} \dot{x}_1 = -\alpha x_1 + \beta x_4 \\ \dot{x}_2 = \frac{\alpha}{2} x_1 - \alpha x_2 - m_1 x_2 \\ \dot{x}_3 = \frac{\alpha}{2} x_1 - \alpha x_3 - m_1 x_3 \\ \dot{x}_4 = \alpha x_2 - m_2 x_4 \\ \dot{x}_5 = \alpha x_3 - m_2 x_5 \end{cases}$$

(where α is an aging rate, m_1 , m_2 are mortality rates, and β is a fecundity rate). If only one sub-population x_i can be targeted for measurement, one can easily check that the only possibility for the system to be observable is to measure the variable $y = x_5$. This system is of the form $\dot{x} = Ax$, $y = Cx$ with

$$A = \begin{bmatrix} \overbrace{\begin{bmatrix} -\alpha & 0 & 0 & \beta \end{bmatrix}}^{A_4} & 0 \\ \alpha/2 & -\alpha - m_1 & 0 & 0 \\ \alpha/2 & 0 & -\alpha - m_1 & 0 \\ 0 & \alpha & 0 & -m_2 \\ 0 & 0 & \alpha & 0 & -m_2 \end{bmatrix}$$

that is of the form

$$\dot{x} = \begin{bmatrix} A_4 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \\ \alpha x_3 - m_2 x_5 \end{bmatrix}, \quad y = Cx$$

From this, we see that for any integer $k \geq 1$, A^k is of the form

$$\begin{bmatrix} & & & & 0 \\ & & & & 0 \\ & & A_4^k & & 0 \\ & & & & 0 \\ * & * & * & * & (-m_2)^k \end{bmatrix}.$$

Therefore, if $C = [c_1 \ c_2 \ c_3 \ c_4 \ 0]$ then $CA^k = (* \ * \ * \ * \ 0)$ and hence the observability matrix will be of rank ≤ 4 and so the system is not observable if the output does not depend on x_5 .

Now, if $C = [0 \ 0 \ 0 \ 0 \ 1]$ then the observability matrix defined by (2.4) is given by:

$$O_{(C,A)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \alpha & 0 & -m_2 \\ \frac{\alpha^2}{2} & 0 & -\alpha(m_1 + m_2 + \alpha) & 0 & m_2^2 \\ -\frac{\alpha^2(2\alpha + m_1 + m_2)}{2} & 0 & h_{43} & \frac{\beta \alpha^2}{2} & -m_2^3 \\ h_{51} & \frac{\alpha^3 \beta}{2} & h_{53} & -\frac{\beta \alpha^2(2\alpha + m_1 + 2m_2)}{2} & m_2^4 \end{bmatrix}$$

with

$$h_{43} = (\alpha^2 + (2m_1 + m_2)\alpha + m_1^2 + m_1 m_2 + m_2^2)\alpha,$$

$$h_{51} = \left(\frac{3\alpha^2}{2} + \frac{(3m_1 + 2m_2)\alpha}{2} + \frac{m_1^2}{2} + \frac{m_1 m_2}{2} + \frac{m_2^2}{2} \right) \alpha^2,$$

$$h_{53} = -\alpha(m_1 + m_2 + \alpha)(\alpha^2 + 2\alpha m_1 + m_1^2 + m_2^2).$$

One has $\det O_{(C,A)} = -\frac{\alpha^8 \beta^2}{8} \neq 0$. Hence $O_{(C,A)}$ is of full rank which proves the observability of the system when $y = x_5$.

2.3 About identifiability

Since very often the initial conditions are not known, or partially known, we will consider in the following the problem of joined identifiability and observability, considering the augmented system (1.3). Note that identifiability-only problems are a special case in which $y = x$ (this is why we consider here the more general case of joined identifiability and observability).

Consider a parametrized system

$$\begin{cases} \dot{x}(t) = f(x(t), \theta), & x(0) = x_0 \\ y(t) = h(x(t), \theta) \end{cases} \quad (2.7)$$

with $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and $\theta \in \Theta \subset \mathbb{R}^p$ and assume that system (2.7) is observable for any known value of the parameter $\theta \in \Theta$. Let us denote the observability map parameterized by θ

$$H_k(x, \theta) = \begin{bmatrix} h(x, \theta) \\ \mathcal{L}_f(h)(x, \theta) \\ \vdots \\ \mathcal{L}_f^{k-1}(h)(x, \theta) \end{bmatrix}$$

When there exists an integer $k > 1$ and a map $\Phi_\theta : \mathbb{R}^n \mapsto H_k(\mathbb{R}^n, \theta)$ parameterized by θ , such that the output differential equation

$$y^{(k)}(t) = \Phi_\theta(y(t), \dot{y}(t), \dots, y^{(k-1)}(t)), \quad t \geq 0$$

as an unique solution in $H_k(\mathbb{R}^n, \theta)$, for any $\theta \in \Theta$, one may study the dependency of the map Φ_θ with respect to θ to ensure the identifiability of the system. In particular, when f and h are polynomials with coefficients parameterized by θ , the map Φ_θ is a polynomial with coefficients $c_i(\theta)$ where $c : \Theta \mapsto \mathbb{R}^v$ for some v . The injectivity of the map c is clearly a necessary condition to have identifiability, but this is not sufficient as one can see on the following example.

Example 2.4 Consider the system in \mathbb{R}_+^2

$$\begin{cases} \dot{x}_1 = -\theta_1 x_1 + \theta_2 x_2 \\ \dot{x}_2 = -\theta_2 x_2 \\ y = x_1. \end{cases} \quad (2.8)$$

where $\theta \in \Theta = \mathbb{R} \times \mathbb{R}^*$ is the unknown vector of parameters. One has

$$H_1(x, \theta) = \begin{bmatrix} x_1 \\ -\theta_1 x_1 + \theta_2 x_2 \end{bmatrix}$$

which is invertible w.r.t. x on \mathbb{R}_+^2 for any $\theta \in \Theta$

$$H_1^{-1}(z, \theta) = \begin{bmatrix} z_1 \\ \frac{z_2 + \theta_1 z_1}{\theta_2} \end{bmatrix}$$

The system is thus observable for any $\theta \in \Theta$. Moreover, one has

$$\ddot{y} = -\theta_1 \dot{y} - \theta_1^2 x_2 = -\theta_2 \dot{y} - \theta_1 (\dot{y} + \theta_2 y)$$

that is

$$\ddot{y} = \Phi_{\theta}(y, y') := -(\theta_1 + \theta_2) \dot{y} - \theta_1 \theta_2 y$$

Clearly the application $\theta \mapsto (\theta_1 + \theta_2, \theta_1 \theta_2)$ is injective on Θ , but the system is not identifiable if $x_2(0) = 0$: the solution verifies $x_2(t) = 0$ for any $t > 0$ and x_1 is solution of $\dot{x}_1 = -\theta_1 x_1$ independently of the value of θ_2 .

A natural (and usual) way to ensure the joined observability-identifiability property is to consider the augmented state:

$$\tilde{x} = \begin{bmatrix} x \\ \theta \end{bmatrix} \in \mathbb{R}^n \times \Theta$$

and require the observability of the extended dynamics

$$\begin{cases} \frac{d}{dt} \tilde{x}(t) = \tilde{f}(\tilde{x}(t)) := f(x(t), \theta), & \tilde{x}(0) = (x(0), \theta) \\ y(t) = \tilde{h}(\tilde{x}(t)) := h(x(t), \theta) \end{cases}$$

and we fall back on a problem of pure observability.

Remark 2.6 A theoretical answer to the problem of parameters reconstruction when measuring the whole state x in \mathbb{R}^n has been given by D. Aeyels [4, 3] and E. Sontag [192] in different form. For an (analytic) system with r parameters, it is *generically* sufficient to choose $2r + 1$ measures at different times to distinguish two different states (the term *generically* means here that for any system excepted for a non dense subset of systems among all the analytic systems in \mathbb{R}^n).

2.4 Identifiability does not necessarily require observability

Let us stress that identifiability does not necessarily imply observability. It can happen that the knowledge of the output function $y(\cdot)$ allows to reconstruct uniquely the set of parameters, but not necessarily the state variables of the system. We give below an example of such a situation.

Example 2.5 The following "academic" model is identifiable but not observable.

$$\begin{cases} \dot{x}_1 = -\alpha (x_1 + x_2) \\ \dot{x}_2 = \alpha (x_1 - x_2) \\ y = \frac{1}{2} (x_1^2 + x_2^2) \end{cases} \quad (2.9)$$

One immediately gets

$$\dot{y} = -\alpha y.$$

For (unknown) initial conditions $(x_1(0), x_2(0)) \neq (0, 0)$, one has $y(0) > 0$ and y is thus a positive function. Then one obtains $\alpha = -\dot{y}/y$: the system is identifiable on $\mathbb{R}^2 \setminus \{0\}$.

Compute now further derivatives: $\ddot{y} = \alpha^2 y, \dots, y^{(p)} = (-1)^p \alpha^p y$. Formally, one gets

$$\text{Jac} [h, \mathcal{L}_f h, \mathcal{L}_f^2 h, \dots, \mathcal{L}_f^p h] = \begin{bmatrix} x_1 & -\alpha x_1 & \alpha^2 x_1 & \dots & (-1)^p \alpha^p x_1 \\ x_2 & -\alpha x_2 & \alpha^2 x_2 & \dots & (-1)^p \alpha^p x_2 \\ 0 & -y & 2\alpha y & \dots & \alpha(-1)^p \alpha^{p-1} y \end{bmatrix},$$

which is of rank 2 for any $(x_1, x_2) \in \mathbb{R}^2 \setminus \{0\}$ and for any positive integer p . The parameter α is identifiable, but the system is not observable. Consider another solution $(\xi_1(\cdot), \xi_2(\cdot))$ for the initial condition $\xi_1(0) = -x_1(0), \xi_2(0) = -x_2(0)$. One can straightforwardly check that this solution verifies $\xi_1(t) = -x_1(t), \xi_2(t) = -x_2(t)$ with the same output $y(t)$ for any $t \geq 0$. Therefore, these two solutions cannot be distinguished, which shows that the system is not observable.

Another way to show the non-observability is to remark that we have

$$y(t) = e^{-\alpha t} y(0) = \frac{1}{2} e^{-\alpha t} (x_1(0)^2 + x_2(0)^2) = \frac{1}{2} e^{-\alpha t} (\bar{x}_1(0)^2 + \bar{x}_2(0)^2)$$

for any $(\bar{x}_1(0), \bar{x}_2(0)) \in S_0$ where S_0 is the circle centered at the origin with radius $r_0 = \sqrt{x_1(0)^2 + x_2(0)^2}$. Thus the output cannot distinguish the solutions emanating from different points of this circle. Hence System (2.9) is non observable.

This can also be proved by remarking that one has

$$\mathcal{L}_f^p h(x) = (-1)^p \alpha^p h(x) = \frac{(-1)^p \alpha^p}{2} (x_1^2 + x_2^2).$$

This means that the associated observation space is generated by the function $x_1^2 + x_2^2$ that does not separate the points of \mathbb{R}^2 and hence, according to Theorem (2.1), system (2.9) is non observable.

2.5 Identifiability via decoupled variables

Sometimes it is difficult to prove identifiability of a model in the original set of coordinates. Let us show the interest of considering other variables that possess "good" properties. We consider a system in \mathbb{R}^n parameterized by $\theta \in \Theta \subset \mathbb{R}^p$ of the following form

$$\begin{cases} \dot{x} = f(x, \theta), & x(0) = x_0 \in X \\ y = h(x) \end{cases} \quad (2.10)$$

where $X \subset \mathbb{R}^n$ is positively invariant for any $\theta \in \Theta$.

Proposition 2.6 [30] *Assume that the following properties hold.*

1. *The map f verifies*

$$\forall \theta_1, \theta_2 \in \Theta, \forall x \in X, \quad \theta_1 \neq \theta_2 \implies f(x, \theta_1) \neq f(x, \theta_2). \quad (2.11)$$

2. *There exist smooth maps g, \tilde{g} and l such that*

a. *for any solution $x(\cdot)$ of (2.10) in X , $w(t) := g(x(t), y(t)) \in W \subset \mathbb{R}^m$ verifies*

$$\dot{w}(t) = l(w(t), y(t)), \quad t \geq 0 \quad (2.12)$$

b. *for any $x \in X$, one has*

$$w = g(x, h(x)) \iff x = \tilde{g}(w, h(x)) \quad (2.13)$$

Then the system (1.1) is identifiable over Θ for any initial condition in X .

Proof Fix $x_0 \in X$ and denote by $x_\theta(\cdot)$ the solution of (2.10) for the parameter $\theta \in \Theta$. Consider θ_1, θ_2 in Θ that give the same output function: $h(x_{\theta_1}(t)) = h(x_{\theta_2}(t)) = y(t)$ for any $t \geq 0$.

Let $w(\cdot)$ be the (unique) solution of the Cauchy problem

$$\dot{w} = l(w, y(t)), \quad w(0) = g(x_0, h(x_0)).$$

Then, one has

$$w(t) = g(x_{\theta_i}(t), y(t)) \quad t \geq 0, \quad i = 1, 2,$$

and from property (2.13), one obtains

$$\tilde{g}(w(t), y(t)) = x_{\theta_i}(t) \quad t \geq 0, \quad i = 1, 2,$$

that is $x_{\theta_1}(t) = x_{\theta_2}(t)$ for any $t > 0$. Therefore, for any $t > 0$ one has also $\dot{x}_{\theta_1}(t) = \dot{x}_{\theta_2}(t)$ or $f(x(t), \theta_1) = f(x(t), \theta_2)$ where $x(t) = x_{\theta_1}(t) = x_{\theta_2}(t)$. Finally, from condition (2.11), we deduce that one has necessarily $\theta_1 = \theta_2$, which shows the identifiability of the system. \square

This result states that when a system is identifiable when measuring the whole state x (condition (2.11)), and there exists a variable w whose dynamics is *decoupled* in the sense that it depends on w and y only (condition (2.12)) such that the map $(x, y) \mapsto w$ is invertible with respect to x (condition (2.13)), then the system is identifiable when measuring y only. Up to our knowledge, this approach has not been deployed in the literature. Let us illustrate this result on an intra-host model for malaria infection [28].

Example 2.6 (Malaria model) The state vector is $x = (S, I_1, \dots, I_5, M)^\top$ in \mathbb{R}_+^7 , where S is the concentration of uninfected erythrocytes in the blood, I_i are the concentrations of infected erythrocytes in different age classes, and M is the concentration of free merozoites. The dynamics is given by the following system

$$\begin{cases} \dot{S} = \Lambda - \mu_S S - \beta SM, \\ \dot{I}_1 = \beta SM - (\gamma_1 + \mu_1) I_1, \\ \dot{I}_2 = \gamma_1 I_1 - (\gamma_2 + \mu_2) I_2, \\ \vdots \\ \dot{I}_5 = \gamma_4 I_4 - (\gamma_5 + \mu_5) I_5, \\ \dot{M} = r \gamma_5 I_5 - \mu_M M - \beta SM, \end{cases} \quad (2.14)$$

where the different parameters are

- Λ : recruitment of the healthy red blood cells (RBC).
- β : rate of infection of RBC by merozoites.
- μ_S : natural death rate of healthy cells.
- μ_i : natural death rate of i -th stage of infected cells.
- γ_i : transition rate from i -th stage to $(i + 1)$ -th stage of infected cells.
- r : number of merozoites released by the late stage of infected cells.
- μ_M : natural death rate of merozoites.

The two first stages of infected erythrocytes (I_1 and I_2) correspond to the concentration of free circulating parasitized erythrocytes than can be observed (seen on peripheral blood smears). Typically, the quantity

$$y(t) = h(x(t)) = I_1(t) + I_2(t)$$

is measured at any time t . Among parameters in (2.14), most of them (μ_i , γ_i , and r) are known or at least widely accepted by the community, but the infection rate β , which is playing a crucial role, is unknown and cannot be estimated by biological considerations. Let us then write the dynamics (2.14) as $\dot{x} = f(x, \beta)$. It takes the form

$$\begin{cases} \dot{x} = f(x, \beta) := Ax + \beta SM E + \Lambda e_1, \\ y = Cx \end{cases} \quad (2.15)$$

with

$$A = \begin{bmatrix} -\mu_S & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\gamma_1 - \mu_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma_1 & -\gamma_2 - \mu_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_2 & -\gamma_3 - \mu_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_3 & -\gamma_4 - \mu_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \gamma_4 & -\gamma_5 - \mu_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & r\gamma_5 & -\mu_M \end{bmatrix},$$

$$E = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \quad e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad C = [0 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0]$$

Due to the dimension of the dynamics, it is not easy to check the identifiability of the parameter β . However, on the domain $X = (\mathbb{R}_+ \setminus \{0\})^7$, one has $SM \neq 0$, which implies the property

$$f_1(x, \beta_1) = f_1(x, \beta_2) \Rightarrow \beta_1 = \beta_2, \quad x \in X.$$

For the parameter $\theta = \beta$, condition (2.11) of Proposition 2.6 is thus fulfilled. Note that one has $ECE = E$. Therefore one can consider the variable

$$w = g(x, y) := x - Ey = (I - EC)x$$

whose dynamics is independent of the non-linear term βSM :

$$\dot{w} = l(w, y) := \bar{A}w + \bar{A}Ey + \Lambda e_1$$

where we posit $\bar{A} = A - ECA$. Given w and y , the state x is then given by

$$x = \tilde{g}(w, y) := w + Ey.$$

Conditions (2.12) and (2.13) of Proposition 2.6 are thus also satisfied, which allows to conclude without any other calculation that the parameter β is identifiable.

This example illustrates the possible interest of exploiting conjointly identifiability and observability to solve the identifiability problem.

Chapter 3

Analysis of the Kermack-McKendrick model

Abstract This chapter is dedicated to the study of the well-known SIR model.

3.1 History

The SIR model of Kermack and McKendrick [122] is certainly one of the most famous model in Epidemiology. It is given and studied in all of the classic books of Mathematical Epidemiology. This model appears in the book of Bailey, which is probably the first book in Mathematical Epidemiology. Some examples can be found in [7, 35, 36, 37, 60, 131, 140, 150]. The figure, in the original paper, fitting the model to plague data in Bombay during the 1906 year, is one of the most famous pictures in Epidemiology. A research with SIR in MathScinet returns 11 106 articles.

In the quoted books the SIR model is fitted to data in the following ways:

- in [35, 36, 37] the model is fitted to the plague in Eyam (in the year 1666);
- in [60] the model is fitted to an influenza epidemic in England and Wales;
- in [131] a fitting is done with simulated noisy data;
- in [140, 35], in a chapter devoted to fitting epidemiological models to data, a SIR model is fitted to an influenza outbreak in an English boarding school.

More recently two publications [17, 137] revisit the fit of the Kermack-McKendrick SIR model to the plague in Bombay.

As already mentioned, before attempting to adjust parameters, an identifiability analysis should be performed.

3.2 The different forms of the SIR model

The original model [122] is

$$\begin{cases} \dot{S} = -\tilde{\beta} S I \\ \dot{I} = \tilde{\beta} S I - \gamma I \\ \dot{R} = \gamma I \end{cases} \quad (3.1)$$

where S, I, R represent respectively the numbers of susceptible, infectious and removed individuals.

This model can also be found in a slightly different form

$$\begin{cases} \dot{S} = -\beta \frac{S}{N} I \\ \dot{I} = \beta \frac{S}{N} I - \gamma I \\ \dot{R} = \gamma I \end{cases} \quad (3.2)$$

where $N = S + I + R$ is the total population. Obviously, one can pass from one model to the other though $\tilde{\beta} = \beta/N$. Both models are mathematically equivalent as long as N is a constant. However, we stress that identifying $\tilde{\beta}$ does not allow one to estimate the parameters β and N separately. For instance, estimating $\tilde{\beta}$ and γ only (without knowing N or β) does not allow one to estimate the basic reproduction number:

$$\mathcal{R}_0 = \frac{\tilde{\beta} N}{\gamma} = \frac{\beta}{\gamma}.$$

3.3 Observability and identifiability of the SIR model

Quite surprisingly, the observability and identifiability of the original Kermack-McKendrick SIR model has not been studied much, although this model is commonly used to model epidemics.

Interestingly, the observability and identifiability of the SIR model with births and deaths, constant population, and an observation $y = kI$, has been first studied in 2005 [84]:

$$\begin{cases} \dot{S} = \mu N - \beta \frac{S}{N} I - \mu S \\ \dot{I} = \beta \frac{S}{N} I - (\gamma + \mu) I \\ \dot{R} = \gamma I - \mu R \end{cases} \quad (3.3)$$

where μ is the renewal rate of the population. The article [84] concludes that the system is neither observable nor identifiable.

In [204] the identifiability of (3.2) is addressed assuming (i) that the initial conditions (and therefore $N = S(0) + I(0) + R(0)$) are known, (ii) observing $y = kI$ with $k = 1$, and using only the input-output relation to conclude. Under assumptions (i) and (ii), the identifiability is quite immediate, as we shall see, but of limited interest.

3.3.1 The SIR model when observing a ratio of the infected population

Here we study the observability-identifiability property of the SIR model

$$\begin{cases} \dot{S} = -\beta \frac{S}{N} I \\ \dot{I} = \beta \frac{S}{N} I - \gamma I \\ y = k I \end{cases} \quad (3.4)$$

The dynamics of R has been omitted since $R = N - S - I$. The observation is $y = k I$, in other words only a fraction of the infectious individuals are observed. This situation is used for example in [137, 178]. In general, the values of N and k are also not known. We have the following result.

Theorem 3.1 *System (3.4) is neither observable, nor identifiable.*

Remark 3.1 Theorem 3.1 can be obtained from [84] by setting $\mu = 0$. However we provide a short and elementary proof.

Proof System (3.4) is obviously not observable on the invariant set $I = 0$. Therefore we study the observability-identifiability properties of System (3.4) on the following positively invariant open set

$$\mathcal{D} = \{(S, I) \mid S > 0, I > 0, S + I < N\}.$$

We will show that there exist a couple of distinct initial condition and distinct parameters that generate the same output. The computation of successive time derivatives of the output y gives:

$$\begin{cases} y = kI, \\ \dot{y} = y \left(\frac{\beta S}{N} - \gamma \right), \\ \ddot{y} = -\frac{\beta^2 S}{kN^2} y^2 + \frac{\dot{y}^2}{y}, \\ \dddot{y} = (-y\ddot{y} + \dot{y}^2) \frac{\beta}{kN} - \frac{3\dot{y}^3}{y^2} + \frac{4\dot{y}\ddot{y}}{y}. \end{cases} \quad (3.5)$$

One can observe that the third derivative of y is expressed as a function of lower derivatives of y of and parameter $\frac{\beta}{kN}$ only. This property remains true for any further derivative of y . This implies that if one considers two sets of initial condition and parameters such that the expressions of y , \dot{y} , \ddot{y} and $\frac{\beta}{kN}$ coincide, then any further derivative also coincides. By analyticity of the solutions of the system as well as of the corresponding outputs, we deduce that their outputs are the same for any time

$t > 0$. More precisely, take two different initial conditions $(S_0, I_0) \neq (\bar{S}_0, \bar{I}_0)$ and sets of parameters $(N, k, \gamma, \beta) \neq (\bar{N}, \bar{k}, \bar{\gamma}, \bar{\beta})$ such that

$$\begin{aligned} kI_0 &= \bar{k}\bar{I}_0, \\ \frac{\beta}{kN} &= \frac{\bar{\beta}}{\bar{k}\bar{N}}, \\ \frac{\beta S_0}{N} - \gamma &= \frac{\bar{\beta}\bar{S}_0}{\bar{N}} - \bar{\gamma}, \\ \frac{\beta^2 S_0}{kN^2} &= \frac{\bar{\beta}^2 \bar{S}_0}{\bar{k}\bar{N}^2} \end{aligned}$$

then $x_0 = (S_0, I_0, N, k, \gamma, \beta)$ and $\bar{x}_0 = (\bar{S}_0, \bar{I}_0, \bar{N}, \bar{k}, \bar{\gamma}, \bar{\beta})$ generate the same output function $y(\cdot)$. \square

Therefore, system (3.4) is not observable nor identifiable. However, it is *partially* observable (respectively *partially* identifiable) in the sense of Definitions 1.4 and 1.5 that is some functions of the state variables (respectively of the parameters) are observable (respectively identifiable). This is made precise in the following proposition.

Proposition 3.1 *System (3.4) satisfies the following properties.*

- i. The state functions kI , $\frac{\beta S}{N}$ and kS are observable.
- ii. The parameters γ and $\frac{\beta}{kN}$ are identifiable.

Proof From (3.5), and since one has $-y\ddot{y} + \dot{y}^2 = \frac{\beta^2 k^2 I^3 S}{N^2} > 0$, we get

$$\begin{cases} kI = y, \\ \frac{\beta S}{N} - \gamma = \frac{\dot{y}}{y}, \\ \frac{\beta^2 S}{kN^2} = \frac{\dot{y}^2}{y^3} - \frac{\ddot{y}}{y^2}, \\ \frac{\beta}{kN} = \frac{\ddot{y} - \frac{4\dot{y}\ddot{y}}{y} + \frac{3\dot{y}^3}{y^2}}{-y\ddot{y} + \dot{y}^2}. \end{cases}$$

From these expressions, we deduce

$$\begin{aligned}\frac{\beta S}{N} &= \frac{(y\ddot{y} - \dot{y}^2)^2}{y(\ddot{y}y^2 - 4\dot{y}\ddot{y}y + 3\dot{y}^3)}, \\ \gamma &= \frac{(y\ddot{y} - \dot{y}^2)^2}{y(\ddot{y}y^2 - 4\dot{y}\ddot{y}y + 3\dot{y}^3)} + \frac{\dot{y}}{y}, \\ kS &= \frac{\frac{\beta S}{N}}{\frac{\beta}{kN}} = \frac{(-y\ddot{y} + \dot{y}^2)^3 y}{(y^2\ddot{y} - 4\dot{y}\ddot{y}y + 3\dot{y}^3)^2}.\end{aligned}$$

Thus the parameters γ and $\frac{\beta}{kN}$ are identifiable and the state functions kI , $\frac{\beta S}{N}$ and kS are observable. \square

Remark 3.2 Since the total population N (constant for the model considered) is often considered known in epidemiology or epidemic modeling (e.g. from the Census Bureau), Proposition 3.1 implies that the quantity β/k is identifiable. However β and k are not identifiable independently which means that there are infinitely many combinations of β and k values for which the model produce the same observable output.

As a consequence of Proposition 3.1, we have also the following properties.

Corollary 3.1 *For system (3.4), one has*

- i. *If $k = \gamma$, then the state variables S and I are observable, and the parameters γ , $\frac{\beta}{N}$ are identifiable.*
- ii. *If N is known and if $k = 1$ or $k = \gamma$ then the system is identifiable and observable.*

Remark 3.3 One could believe at the first look that if $k = \gamma$ but with N unknown, then (3.1) is observable, but this is wrong. Certainly S and I are observable, but R is not observable. Indeed the output $y = kI$ and state I , solution of system (3.1), do not depend on the variable R since R has no influence on the two first equations of (3.1). So, the output generated by system (3.1) will be the same for any initial conditions (S_0, I_0, R_0) and (S_0, I_0, \bar{R}_0) even if $R_0 \neq \bar{R}_0$. Therefore the value of $N = S + I + R$ is inaccessible. As a consequence, $\mathcal{R}_0 = \beta N / \gamma$ is not identifiable.

Remark 3.4 The second point of Corollary 3.1 can be also proved by considering the "augmented" system (defined in Chapter 1 by Equation (1.3)) and using Proposition 1.1. Indeed, when N is known and $k = \gamma$, consider the "augmented" SIR model as follows

$$\begin{cases} \dot{S} = -\beta \frac{S}{N} I \\ \dot{I} = \beta \frac{S}{N} I - \gamma I \\ \dot{\beta} = 0 \\ \dot{\gamma} = 0 \\ y = \gamma I \end{cases} \quad (3.6)$$

Posit $x = \begin{bmatrix} S \\ I \\ \beta \\ \gamma \end{bmatrix}$, $f(x) = \begin{bmatrix} -\beta \frac{S}{N} I \\ \beta \frac{S}{N} I - \gamma I \\ 0 \\ 0 \end{bmatrix}$ and $h(x) = \gamma I$. We have

$$\begin{aligned} \mathcal{L}_f h(x) &= \langle \nabla h(x) | f(x) \rangle = \left\langle \begin{bmatrix} 0 \\ \gamma \\ 0 \\ I \end{bmatrix} \middle| \begin{bmatrix} -\beta \frac{S}{N} I \\ \beta \frac{S}{N} I - \gamma I \\ 0 \\ 0 \end{bmatrix} \right\rangle \\ &= \gamma \left(\beta \frac{S}{N} I - \gamma I \right) = \gamma \left(\beta \frac{S}{N} - \gamma \right) I = \left(\beta \frac{S}{N} - \gamma \right) h(x). \end{aligned}$$

$$\begin{aligned} \mathcal{L}_f^2 h(x) &= \langle \nabla \mathcal{L}_f h(x) | f(x) \rangle = \left\langle \begin{bmatrix} \gamma \frac{\beta}{N} I \\ \gamma \left(\beta \frac{S}{N} - \gamma \right) \\ \gamma \frac{S}{N} I \\ \beta \frac{S}{N} I \end{bmatrix} \middle| \begin{bmatrix} -\beta \frac{S}{N} I \\ \beta \frac{S}{N} I - \gamma I \\ 0 \\ 0 \end{bmatrix} \right\rangle \\ &= -\gamma \frac{\beta^2}{N^2} SI^2 + \gamma \left(\beta \frac{S}{N} - \gamma \right)^2 I = -\gamma \frac{\beta^2}{N^2} SI^2 + \left(\beta \frac{S}{N} - \gamma \right) \mathcal{L}_f h(x) \\ &= -\frac{\beta^2}{N^2} SI h(x) + \left(\beta \frac{S}{N} - \gamma \right) \mathcal{L}_f h(x). \end{aligned}$$

and computing the third Lie-derivative of h gives the expression

$$\begin{aligned} \mathcal{L}_f^3 h(x) &= -\frac{\beta^2}{N^2} SI \left(\frac{\beta S}{N} - \gamma - \frac{\beta I}{N} \right) h(x) - 2 \frac{\beta^2}{N^2} SI \mathcal{L}_f h(x) + \left(\frac{\beta S}{N} - \gamma \right) \mathcal{L}_f^2 h(x) \\ \mathcal{L}_f^3 h(x) &= \frac{(S(I^2 - 4IS + S^2)\beta^3 + (4I - 3S)S\gamma N\beta^2 + 3N^2S\beta\gamma^2 - N^3\gamma^3)\gamma I}{N^3}. \end{aligned}$$

We show that the map

$$x \mapsto H_3(x) = (h(x), \mathcal{L}_f h(x), \mathcal{L}_f^2 h(x), \mathcal{L}_f^3 h(x))^\top \quad (3.7)$$

is injective. Let x and \bar{x} two elements of the state space $\Omega = \{(S, I, \beta, \gamma) \in \mathbb{R}^4 : S > 0, I > 0, \beta > 0, \gamma > 0\}$. Suppose that $H_3(x) = H_3(\bar{x})$. Then, we have the following successive implications:

$$h(x) = h(\bar{x}), \mathcal{L}_f(h)(x) = \mathcal{L}_f(h)(\bar{x}), \mathcal{L}_f^2(h)(x) = \mathcal{L}_f^2(h)(\bar{x}) \implies$$

$$\begin{cases} \gamma I = \bar{\gamma} \bar{I} \\ \left(\beta \frac{S}{N} - \gamma\right) h(x) = \left(\bar{\beta} \frac{\bar{S}}{N} - \bar{\gamma}\right) h(\bar{x}) \\ -\frac{\beta^2}{N^2} SI h(x) + \left(\beta \frac{S}{N} - \gamma\right) \mathcal{L}_f h(x) = -\frac{\bar{\beta}^2}{N^2} \bar{S} \bar{I} h(\bar{x}) + \left(\bar{\beta} \frac{\bar{S}}{N} - \bar{\gamma}\right) \mathcal{L}_f h(\bar{x}) \end{cases}$$

implies

$$\begin{aligned} & \begin{cases} \gamma I = \bar{\gamma} \bar{I} \\ \left(\beta \frac{S}{N} - \gamma\right) = \left(\bar{\beta} \frac{\bar{S}}{N} - \bar{\gamma}\right) \\ \frac{\beta^2}{N^2} SI = \frac{\bar{\beta}^2}{N^2} \bar{S} \bar{I} \end{cases} \implies \begin{cases} \gamma I = \bar{\gamma} \bar{I} \\ \left(\beta \frac{S}{N} - \gamma\right) = \left(\bar{\beta} \frac{\bar{S}}{N} - \bar{\gamma}\right) \\ \beta^2 SI = \bar{\beta}^2 \bar{S} \bar{I} \end{cases} \\ & \implies \begin{cases} \gamma I = \bar{\gamma} \bar{I} \\ \beta \frac{S}{N} - \gamma = \bar{\beta} \frac{\bar{S}}{N} - \bar{\gamma} \\ \beta^2 S = \bar{\beta}^2 \frac{\gamma}{\bar{\gamma}} \bar{S} \end{cases} \implies \begin{cases} \gamma I = \bar{\gamma} \bar{I} \\ \beta \frac{S}{N} - \gamma = \frac{\beta^2 \bar{\gamma}}{\bar{\beta} \gamma} \frac{S}{N} - \bar{\gamma} \\ \beta^2 S = \bar{\beta}^2 \frac{\gamma}{\bar{\gamma}} \bar{S} \end{cases} \end{aligned}$$

Reporting in $\mathcal{L}_f^3(h)(x) = \mathcal{L}_f^3(h)(\bar{x})$, we obtain

$$\begin{aligned} & -\frac{\beta^2}{N^2} SI \left(\frac{\beta S}{N} - \gamma - \frac{\beta I}{N}\right) h(x) - 2\frac{\beta^2}{N^2} SI \mathcal{L}_f h(x) + \left(\frac{\beta S}{N} - \gamma\right) \mathcal{L}_f^2 h(x) \\ & = -\frac{\beta^2}{N^2} SI \left(\frac{\beta S}{N} - \gamma - \frac{\bar{\beta} \bar{I}}{N}\right) h(x) - 2\frac{\beta^2}{N^2} SI \mathcal{L}_f h(x) + \left(\frac{\beta S}{N} - \gamma\right) \mathcal{L}_f^2 h(x) \end{aligned}$$

This implies

$$\frac{\beta I}{N} = \frac{\bar{\beta} \bar{I}}{N} \text{ and } \beta I = \bar{\beta} \bar{I}$$

With $\gamma I = \bar{\gamma} \bar{I}$, we obtain $\bar{\beta} = \frac{\bar{\gamma}}{\gamma} \beta$. Reporting in $\beta \frac{S}{N} - \gamma = \frac{\beta^2 \bar{\gamma}}{\bar{\beta} \gamma} \frac{S}{N} - \bar{\gamma}$, we obtain $\bar{\gamma} = \gamma$. We then deduce $\bar{I} = I$, $\bar{\beta} = \beta$, and $\bar{S} = S$, and thus $x = \bar{x}$. We have shown that the map H_3 (3.7) is injective which proves that the augmented system (3.6) is observable which implies, thanks to Proposition 1.1, that System (3.4) is observable and identifiable when the total (constant) population N is known and $k = \gamma$.

3.3.2 The SIR model when observing the incidence

Quite often, observations of new cases per unit time or incidence are available. We study how this changes the observability and identifiability of the SIR model.

We thus consider the system where the observation is given by

$$y = k\beta \frac{SI}{N}. \quad (3.8)$$

The system under consideration is then

$$\begin{cases} \dot{S} = -\beta \frac{S}{N} I, \\ \dot{I} = \beta \frac{S}{N} I - \gamma I, \\ y = k\beta \frac{SI}{N}. \end{cases} \quad (3.9)$$

This problem has been addressed for the SIR model with demography for constant population in [84]. Identifiability with known initial conditions for (3.9) is also considered in [204] using input-output relations.

Theorem 3.2 *The system (3.9) with incidence observation is neither observable, nor identifiable.*

Proof We proceed as we did for the proof of Theorem 3.1. The computation of successive time derivatives of the output y gives:

$$\begin{cases} y = k\beta \frac{SI}{N}, \\ \dot{y} = \left(\frac{\beta S}{N} - \frac{\beta I}{N} - \gamma \right) y, \\ \ddot{y} = \frac{\dot{y}^2}{y} + \left(-\frac{2\beta}{kN} y + \frac{\gamma\beta I}{N} \right) y, \\ y^{(3)} = \frac{2\dot{y}\ddot{y}}{y} - \frac{\dot{y}^3}{y^2} - 4\frac{\beta}{kN} y\dot{y} + \frac{\gamma\beta I}{N} \dot{y} + \frac{\gamma\beta}{N} y \left(\frac{y}{k} - \gamma I \right) \\ = \frac{2\dot{y}\ddot{y}}{y} - \frac{\dot{y}^3}{y^2} + \frac{\beta}{kN} (-4y\dot{y} + \gamma y^2) + \frac{\gamma\beta I}{N} (\dot{y} - \gamma y) \end{cases} \quad (3.10)$$

From the expression of \ddot{y} , one takes

$$\frac{\gamma\beta I}{N} = \frac{\ddot{y}}{y} - \frac{\dot{y}^2}{y^2} + 2\frac{\beta}{kN} y$$

and reporting in the expression of $y^{(3)}$ gives

$$\begin{aligned}
y^{(3)} &= \frac{2\dot{y}\ddot{y}}{y} - \frac{\dot{y}^3}{y^2} + \frac{\beta}{kN} (-4y\dot{y} + \gamma y^2) + \left(\frac{\ddot{y}}{y} - \frac{\dot{y}^2}{y^2} + 2\frac{\beta}{kN}y\right)(\dot{y} - \gamma y) \\
&= \frac{2\dot{y}\ddot{y}}{y} - \frac{\dot{y}^3}{y^2} + \left(\frac{\ddot{y}}{y} - \frac{\dot{y}^2}{y^2}\right)(\dot{y} - \gamma y) + \frac{\beta}{kN} (-4y\dot{y} + \gamma y^2 + 2y(\dot{y} - \gamma y)) \\
&= \frac{2\dot{y}\ddot{y}}{y} - \frac{\dot{y}^3}{y^2} + \left(\frac{\ddot{y}}{y} - \frac{\dot{y}^2}{y^2}\right)(\dot{y} - \gamma y) + \frac{\beta}{kN} (-2y\dot{y} - \gamma y^2) \\
&= \frac{\ddot{y}\dot{y}}{y} + \left(\frac{2\dot{y}}{y} - \gamma\right) \frac{(y\ddot{y} - \dot{y}^2)}{y} - \frac{\beta}{kN} (\gamma y + 2\dot{y}) y.
\end{aligned}$$

To summarize, the successive derivatives are

$$\begin{cases}
y = k\beta \frac{SI}{N}, \\
\dot{y} = \left(\frac{\beta S}{N} - \frac{\beta I}{N} - \gamma\right) y, \\
\ddot{y} = \frac{\dot{y}^2}{y} + \left(-\frac{2\beta}{kN}y + \frac{\gamma\beta I}{N}\right) y, \\
y^{(3)} = \frac{\ddot{y}\dot{y}}{y} + \left(\frac{2\dot{y}}{y} - \gamma\right) \frac{(y\ddot{y} - \dot{y}^2)}{y} - \frac{\beta}{kN} (\gamma y + 2\dot{y}) y, \\
y^{(3)} = \frac{3y\dot{y}\ddot{y} - 2\dot{y}^3}{y^2} + \frac{\dot{y}^2 - y\ddot{y}}{y} \gamma - \frac{\beta}{kN} (\gamma y + 2\dot{y}) y
\end{cases} \quad (3.11)$$

This shows that all the derivatives of y of order $q \geq 3$ can be expressed as functions of derivatives $y^{(r)}$, $r = 0, \dots, q-1$ and parameters $\gamma, \frac{\beta}{kN}$, only. We deduce that for two different $x = (S_0, I_0, N, k, \gamma, \beta)$ and $\bar{x} = (\bar{S}_0, \bar{I}_0, \bar{N}, \bar{k}, \bar{\gamma}, \bar{\beta})$ such that

$$\gamma = \bar{\gamma} \quad \text{and} \quad \frac{\beta}{kN} = \frac{\bar{\beta}}{\bar{k}\bar{N}}$$

and derivatives $y^{(r)}(0)$, $r = 0 \dots 2$, coincide, then their output $y(t)$ is the same for any $t > 0$ (by analyticity of the system). Notice that one can write

$$y = k\beta \frac{SI}{N} = \frac{\beta I}{N} \frac{\beta S}{N} \left(\frac{\beta}{kN}\right)^{-1}$$

Therefore, from expressions (3.11), $y^{(r)}(0)$, $r \in \{0, 1, 2\}$ coincide under the single condition

$$\frac{\beta I_0}{N} = \frac{\bar{\beta} \bar{I}_0}{\bar{N}}, \quad \text{and} \quad \frac{\beta S_0}{N} = \frac{\bar{\beta} \bar{S}_0}{\bar{N}}.$$

We conclude that the system is neither observable, nor identifiable. \square

Proposition 3.2 *If the parameters N , γ and k are known than System (3.9) is observable and identifiable.*

Proof Thanks to Proposition 1.1, to prove that System (3.9) is observable and identifiable, it is sufficient to prove that the following (augmented) system whose state is (S, I, β) :

$$\begin{cases} \dot{S} = -\beta \frac{S}{N} I \\ \dot{I} = \beta \frac{S}{N} I - \gamma I \\ \dot{\beta} = 0 \\ y = k\beta \frac{SI}{N} \end{cases} \quad (3.12)$$

is observable. From the expression of $y^{(3)}$, we obtain

$$\beta = \frac{\gamma(\dot{y}^2 - y\ddot{y})y + (3\dot{y}\ddot{y} - yy^{(3)})y - 2\dot{y}^3}{\gamma y^4 + 2\dot{y}y^3} kN.$$

Reporting in \ddot{y} , we obtain I as a function of N, γ, k , and the successive derivatives of y . The state variable S is obtained by reporting the expressions of β and I in the expression of y . Thus System (3.12) satisfies the condition of Proposition 2.3 and therefore it is observable on $\{(S, I, \beta) \in \mathbb{R}^3 : S > 0, I > 0, S + I < N, \beta > 0\}$. \square

Remark 3.5 If it is rather the cumulative incidence that is observed, that is

$$y(t) = k \int_0^t \beta \frac{S(\tau)I(\tau)}{N} d\tau.$$

the system is also not observable or identifiable, and the same results as in Theorem 3.2 are available. Indeed all the k -th derivatives of y (for $k \geq 1$) correspond to the $k - 1$ -th derivatives of the observation (3.8).

Chapter 4

Observers synthesis

Abstract We expose the concept of observers and review different techniques to build observers.

4.1 Introduction

As in the previous chapters we consider an observed system

$$\begin{cases} \dot{x} = f(x), & x \in \Omega \subset \mathbb{R}^n, \\ y = h(x) \in Y \subset \mathbb{R}^m \end{cases} \quad (4.1)$$

and denote by $x(t, x_0)$ the solution of $\dot{x} = f(x)$ for the initial condition $x(0) = x_0$.

So far we have studied observability as a property ensuring that knowledge of a measured “signal” $y(\cdot)$ results in the uniqueness of the initial condition x_0 . In this chapter we will address the *state estimation problem* which consist in obtaining an estimate $\hat{x}(t)$ of the state $x(t)$ of the system at time t , with the knowledge of the output $y(\cdot)$ up to time t .

We have for addressing this problem different possibilities

The first one is *given the output $y(\cdot)$ up to time t , find the possible initial condition x_0 which produces the same output $y(\cdot)$ up to time t* . Then, the estimate of $x(t)$ is given by the solution $x(t, x_0)$. The uniqueness of this problem is ascertained by the observability property of the system. This approach leads to the resolution of a *minimization problem*.

$$\min_{x_0} \int_0^t \| h(x(x_0, s)) - y(s) \|^2 ds.$$

In other words, we look, by simulating the system for different initial conditions, to the “best” one. The drawback of this approach is related to the difficulties of the non-

linear minimization algorithms (existence of different local minimal, convergence speed...). Once one obtains an estimation \hat{x}_0 of the solution of this minimization problem, the estimation of the state $x(t)$ is given by $\hat{x}(t) = x(\hat{x}_0, t)$.

A second approach is to differentiate the available outputs a number of times and then combine these derivatives appropriately to obtain the state vector. Formally, when the system is observable and we know (perfectly) enough derivatives of $y(\cdot)$ at a given time t , one just has to invert the map $\phi \mapsto (h(x), \mathcal{L}_f h(x), \dots)$ at $(y(t), \dot{y}(t), \ddot{y}(t), \dots)$ to reconstruct the state variable $x(t)$. In practice, it is known that numerically calculating derivatives on the raw signal data is imprecise and sensitive to measurement noise, especially if several successive derivatives have to be determined. It is generally preferable to use a "filter" to smooth the data. For instance, polynomial functions or splines can approach with some regularity the measurements obtained over time, on which the derivative calculations can be performed before the inversion operation. This is an approximation method that does not guarantee an exact solution, and whose accuracy can be strongly influenced by the sensitivity of the solutions of the system with respect to the derivatives of the output.

The last approach is to look for a dynamical system whose "inputs" are the output of the observed system y , and whose "output" is an estimate \hat{x} of the state of the original system, illustrated on the diagram of Figure 4.1. Such dynamical system is called an observer and is classical in the theory of control.

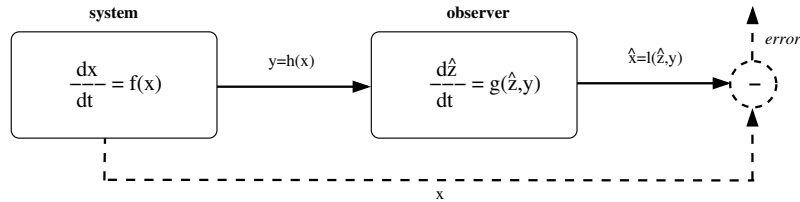


Fig. 4.1 The observer concept

Definition 4.1 An observer for system (4.1) is a input-output system of the form

$$\begin{cases} \dot{\hat{z}} = g(\hat{z}, y(t)), & \hat{z} \in Z \subset \mathbb{R}^{n_z} \\ \hat{x} = l(\hat{z}, y(t)) \end{cases} \quad (4.2)$$

such that $\hat{x}(t)$ is an asymptotic estimate of $x(t)$ satisfying

$$\lim_{t \rightarrow +\infty} \|\hat{x}(t) - x(t)\| = 0$$

for any initial condition $(x(0), \hat{z}(0)) \in \Omega \times Z$.

In this definition, note that n_z is not necessarily equal to n . It can be less than n for *reduced-order* observers (see section 4.5 below) or larger than n (see Remark 4.3 later on). Let us also underline that this amounts to require the convergence of the solution of the coupled dynamics

$$\begin{cases} \dot{\hat{x}} = f(\hat{x}) \\ \dot{\hat{z}} = g(\hat{z}, h(\hat{x})) \end{cases} \Rightarrow \lim_{t \rightarrow +\infty} l(\hat{z}(t), h(\hat{x}(t))) - x(t) = 0$$

for any initial condition in $\Omega \times Z$.

In this Chapter, we study the construction of observers and their theoretical convergence, without considering their practical performances in presence of measurements noise. This point will be addressed with more practical considerations in Chapter 5.

Many observers are indeed of the form

$$\dot{\hat{x}} = f(\hat{x}) + G(h(\hat{x}) - y(t))$$

where G is a constant matrix (i.e. of the form (4.2) with $\hat{z} = \hat{x}$). Such observers are often called *Luenberger observers* [135]. Note that this construction consists in a copy of the original dynamics f plus a correcting term which depends on the *innovation* term $h(\hat{x}) - y(t)$, that is defined as the difference between the expected output $\hat{y} = h(\hat{x}(t))$ if the true state was $\hat{x}(t)$ and the effective measured output $y(t)$. Therefore, if $\hat{x}(t)$ and $x(t)$ are equal at a certain time t , it will remain identical at any future time. The main point is that the matrix G , often called the *gains matrix*, has to be chosen such that the estimation error $\hat{x}(t) - x(t)$ does converge to 0, possibly fast. When f and h are linear and the system is observable, the theory of linear automatic control teaches that there always exists G such that the convergence speed of the estimator can be chosen arbitrarily fast (see e.g. [9]). Obviously epidemiological models are rarely linear. However, looking for an Luenberger observer is often a first trial before considering more sophisticated estimators. Indeed, we shall see that for certain nonlinear dynamics, such observers do the job and in other cases, observers can be inspired from this form.

In this chapter, we do not pretend to present an exhaustive review of all the possible kinds of observers that exist in the literature (we make some comments at the end of the chapter). We focus on the most general classes of observers, for which the proof of the convergence can be shown, under some assumptions, and that can be derived in a systematic way from the equations of the model, once these assumptions are fulfilled. We begin by some simple cases of observers for particular dynamics exploiting properties of linear systems, and then we consider a more general nonlinear framework. The numerical implementation of these observers is illustrated in Chapter 5.2.

4.2 Observers with linear error dynamics

Consider systems of the form

$$\begin{cases} \dot{x} = Ax + \Phi(t, y) \\ y = Cx \end{cases} \quad (4.3)$$

where $x \in \Omega$ and $y \in \mathbb{R}$. We are looking for observers in the Luenberger form

$$\dot{\hat{x}} = A\hat{x} + \Phi(t, y(t)) + G(C\hat{x} - y(t)), \quad \hat{x} \in \mathbb{R}^n \quad (4.4)$$

for which the error vector $e(t) = \hat{x}(t) - x(t)$ is solution of the linear dynamics

$$\dot{e} = (A + GC)e \quad (4.5)$$

The choice of the *gains* vector G providing a convergence of $e(t)$ to 0 comes directly from *the poles placement* technique of the theory of linear systems that we recall below. Let us underline that although Ω is assumed to be positively invariant by the dynamics (4.3), there is no reason for Ω to be invariant by (4.4) because of the additive correction term. This is why we consider the dynamics (4.4) in whole \mathbb{R}^n .

The following lemma gives a key result to construct the matrix G , which is well-known and often used in automatic control.

Lemma 4.1 *Let A be a square matrix of size n and C be a line vector of length n . If the observability matrix O defined in (2.4) is of full rank, then for any set $\Lambda = \{\lambda_1, \dots, \lambda_n\}$ of n real or complex two-by-two conjugate numbers, there is a G vector of size n such that*

$$Sp(A + GC) = \Lambda.$$

Specifically, if

$$\pi_A(\xi) = \xi^n + a_1\xi^{n-1} + \dots + a_{n-1}\xi + a_n$$

is the characteristic polynomial of A , then one has

$$G = P \left[a_{n+1-i} + (-1)^{n-i} \sigma_{n+1-i}(\Lambda) \right]_{i=1}^n$$

where

$$P = O^{-1} \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ 1 \end{bmatrix} [I \ A \ \dots \ A^{n-1}]$$

and the σ_k designate the symmetric functions of the roots

$$\sigma_k(\Lambda) = \sum_{1 \leq i_1 \leq \dots \leq i_k \leq n} \lambda_{i_1} \lambda_{i_2} \dots \lambda_{i_k} \quad (4.6)$$

A proof of Lemma 4.1 is given in Appendix A.1.

Remark. This result can be generalized to vectorial observations, i.e. for matrices C with $m > 1$ rows and n columns.

Finally, by choosing numbers λ_i with negative real parts, one can make the convergence of the error given by the exponentially decreasing dynamics (4.5) as fast as desired. The observer (4.4) is thus adjustable. It should be noted that when the difference $\hat{z}_1(t) - y(t)$ (usually called "innovation") becomes and remains close to 0, the trajectories of the observer follow those of the system: we can then consider that the observer has practically converged. The innovation is thus very useful in practice because it provides information on the current stage of convergence of the estimate.

We illustrate this technique on a population model with age classes.

Example 4.1 Let us consider a population structured in three stages: young, subadult and adult, of stocks x_1, x_2, x_3 respectively. It is assumed that only adults x_3 can reproduce, giving birth to young x_1 :

$$\begin{cases} \dot{x}_1 = -a_1x_1 - m_1x_1 + r(t, x_3) \\ \dot{x}_2 = a_1x_1 - a_2x_2 - m_2x_2 \\ \dot{x}_3 = a_2x_2 - m_3x_3 \\ y = x_3 \end{cases} \quad (4.7)$$

The coefficients a_i are the transition rates between age classes, m_i are the mortality rates of each class, and $r(\cdot)$ is the reproductive function (usually non-linear and seasonally dependent), for example

$$r(t, x_3) = \frac{\bar{r}(t)x_3}{k + x_3}, \quad \bar{r}(t) \in [\bar{r}_{min}, \bar{r}_{max}].$$

Here, it is assumed that only the size of the adult class is measured over time. The aim is to estimate the stocks of larvae and subadults over time. The model (4.7) is of the form (4.3) with $\Omega = \mathbb{R}_+^3$, where we posed

$$A = \begin{bmatrix} -(a_1 + m_1) & 0 & 0 \\ a_1 & -(a_2 + m_2) & 0 \\ 0 & a_2 & -m_3 + 0 \end{bmatrix}, \quad C = [0 \ 0 \ 0 \ 1]$$

and

$$\Phi(t, x_3) = \begin{bmatrix} 0 \\ r(t, x_3) \end{bmatrix}$$

One can check that the observability matrix \mathcal{O} is full rank, or alternatively directly check that the system is observable. Indeed, we obtain by using the expression \dot{x}_3 :

$$x_2 = \frac{\dot{y} + m_3y}{a_2},$$

then with the expression \dot{x}_2 :

$$x_1 = \frac{\dot{x}_2 + (a_2 + m_2)x_2}{m_1} = \frac{\ddot{y} + (a_2 + m_2)m_3y}{a_2m_2}.$$

Therefore, the following system in \mathbb{R}^3

$$\begin{cases} \dot{\hat{x}}_1 = -a_1\hat{x}_1 - m_1\hat{x}_1 + r(t, y(t)) + G_1(\hat{x}_3 - y(t)) \\ \dot{\hat{x}}_2 = a_1\hat{x}_1 - a_2\hat{x}_2 - m_2\hat{x}_2 + G_2(\hat{x}_3 - y(t)) \\ \dot{\hat{x}}_3 = a_2\hat{x}_2 - m_3\hat{x}_3 + G_3(\hat{x}_3 - y(t)) \end{cases} \quad (4.8)$$

with well chosen gains G_1, G_2, G_3 is an observer for the dynamics (4.7), with an exponential convergence.

In the next section, we study a more general kind of non-linearity.

4.3 Observers for systems with Lipschitz non-linearity

Let us consider a system of the form

$$\begin{cases} \dot{x} = Ax + \phi(x) \\ y = Cx \end{cases} \quad (4.9)$$

where

1. the pair (A, C) is observable i.e. the observability matrix O defined in (2.4) is of full rank,
2. the map ϕ is globally Lipschitz on \mathbb{R}^n , with a Lipschitz constant ℓ .

For this non-linear system, consider an observer in Luenberger form

$$\dot{\hat{x}} = A\hat{x} + \phi(\hat{x}) + G(C\hat{x} - y(t)) \quad (4.10)$$

We recall the Lyapunov Theorem [161], which is useful in many situations.

Theorem 4.1 *Given a symmetric positive definite matrix Q , there exists a unique symmetric definite positive matrix P satisfying $M^\top P + PM + Q = 0$ if and only if the linear system $\dot{x} = Mx$ is globally exponentially stable.*

A proof of this Theorem is given in Appendix A.3. Then, one has the following result

Proposition 4.1 *Take G such that $A + GC$ is Hurwitz¹, then if ℓ is small enough, (4.10) is an exponential observer of (4.9), in the sense that there exist $\alpha > 0$ and $\beta > 0$ such that*

¹ A matrix is Hurwitz if all its eigenvalue have negative real parts.

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

for any initial condition $x(0)$, $\hat{x}(0)$.

Proof Accordingly to Theorem 4.1, there exists an unique symmetric definite matrix P satisfying

$$(A + GC)^\top P + P(A + GC) + I = 0 \quad (4.11)$$

(where I is the identity matrix). Let us consider the candidate Lyapunov function

$$V(e) = e^\top P e$$

and the time function $v(t) = V(e(t))$. One gets

$$\begin{aligned} \dot{v} &= e^\top \left((A + GC)^\top P + P(A + GC) \right) e + 2e^\top P(\phi(\hat{x}) - \phi(x)) \\ &= -\|e\|_2^2 + 2e^\top P(\phi(\hat{x}) - \phi(x)) \\ &\leq -\|e\|_2^2 + 2\ell \|P\|_2 \|e\|_2^2 = -(1 - 2\ell \|P\|_2) \|e\|_2^2 \end{aligned}$$

If the Lipschitz constant ℓ is small, one has

$$1 - 2\ell \|P\|_2 > 0. \quad (4.12)$$

On the other hand, as P is a positive definite matrix, one has

$$\lambda_{\min} \|e\|_2 \leq V(e) \leq \lambda_{\max} \|e\|_2$$

where λ_{\min} , λ_{\max} are the smallest and largest eigenvalues of P . Then one obtains $\dot{v} \leq -2\beta v$ where

$$\beta = \frac{1 - 2\ell \|P\|_2}{2\lambda_{\min}} > 0$$

which implies

$$v(t) \leq v(0) e^{-2\beta t} \Rightarrow \|e(t)\|^2 \leq \alpha^2 \|e(0)\|^2 e^{-2\beta t}$$

with $\alpha = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$. Finally, one gets

$$\|e(t)\| \leq \alpha \|e(0)\| e^{-\beta t}, \quad t \geq 0$$

which proves that the estimation error e converges exponentially to 0. \square

There exist other techniques to design a gain G based on the Riccati equation rather than the Lyapunov equation (4.11), but that are more technical and thus out of the scope of this book (we refer to [2] for interested readers).

Remark 4.1 If the map ϕ can be written as $\phi(x) = \varphi(h(x), x) = \varphi(y, x)$, where φ is globally Lipschitz w.r.t. x uniformly in y , then one can consider the observer

$$\dot{\hat{x}} = A\hat{x} + \varphi(y(t), \hat{x}) + G(C\hat{x} - y(t)) \quad (4.13)$$

which generalizes the observers (4.4) and (4.10).

Example 4.2 We consider the "SIRS" model [125], that is the SIR model with loss of immunity, assuming that the size of the recovered population is tracked over time

$$\begin{cases} \dot{S} = -\beta SI + \mu R \\ \dot{I} = \beta SI - \gamma I \\ \dot{R} = \gamma I \\ y = R \end{cases}$$

Here S , I and R denote the densities of susceptible, infected and recovered populations, so that one has $S + I + R = 1$ at any time. We assume that all the parameters are known, and we aim at estimating S and I . Measuring R is equivalent to measuring $S + I$, and we rewrite this model as follows

$$\begin{cases} \dot{S} = -\beta I + \varphi(S, I) + \mu(1 - S - I) \\ \dot{I} = \beta I - \varphi(S, I) - \gamma I \\ y = S + I \end{cases}$$

where we posit

$$\varphi(S, I) = \beta(1 - S)I$$

If we consider the initial stage of an epidemics, variables I and S are close to 0 and 1 respectively, that is on a domain $\Delta := \{1 - \varepsilon < S < 1; 0 < I < \varepsilon\}$ for some $\varepsilon > 0$. This domain is not invariant by the dynamics but we shall consider the estimation problem on a time windows for which the solution stays in this set Δ . One can check that the function φ is Lipschitz with a constant $\ell = 2\varepsilon$ on Δ , and can consider an extension of φ outside Δ with the same Lipschitz constant ℓ but on all \mathbb{R}^2 , for instance

$$\tilde{\varphi}(I, S) = \varphi(\text{sat}_{[0, \varepsilon]} \text{sat}_{[0, \varepsilon]}(I))$$

where $\text{sat}_{[a, b]}$ denotes the saturation function

$$\text{sat}_{[a, b]}(x) = \max(a, \min(b, x)).$$

On Δ , the system can be written as follows

$$\begin{aligned} \frac{d}{dt} \begin{bmatrix} S \\ I \end{bmatrix} &= \underbrace{\begin{bmatrix} -\mu & -\beta - \mu \\ 0 & \beta - \gamma \end{bmatrix}}_A \begin{bmatrix} S \\ I \end{bmatrix} + \underbrace{\begin{bmatrix} \tilde{\varphi}(S, I) \\ -\tilde{\varphi}(S, I) \end{bmatrix}}_{\phi(S, I)} \\ y &= \underbrace{\begin{bmatrix} 1 & 1 \end{bmatrix}}_C \begin{bmatrix} S \\ I \end{bmatrix} \end{aligned} \quad (4.14)$$

One can check that the pair (A, C) is observable, as the matrix

$$\begin{bmatrix} C \\ CA \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -\mu & -\mu - \gamma \end{bmatrix}$$

is full rank. The system (4.14) is well defined on the whole \mathbb{R}^2 , and ϕ is a globally Lipschitz map, with a Lipschitz constant equal to 2ℓ . We are thus in position to apply Proposition 4.1: if ε is small enough, the system

$$\frac{d}{dt} \begin{bmatrix} \hat{S} \\ \hat{I} \end{bmatrix} = A \begin{bmatrix} \hat{S} \\ \hat{I} \end{bmatrix} + \phi(\hat{S}, \hat{I}) + G(\hat{S} + \hat{I} - y(t))$$

where G is such that $A + GC$ is Hurwitz, is an exponential observer of (4.14).

In Section 5.2.2, we discuss the applicability of this observer with numerical values.

It may happen that the estimation error of an observer is only partially assignable, as we shall see in the next example.

4.4 Observers via decoupled variables

Consider, as in Section 2.5, systems

$$\begin{cases} \dot{x} = f(x), & x \in \Omega \\ y = h(x) \end{cases}$$

for which there exists a change of coordinates $x \in \Omega \mapsto w = g(x, h(x)) \in \mathcal{W} \subset \mathbb{R}^{n'}$ such that one has

$$\dot{w}(t) = l(w(t), h(x(t))), \quad t \geq 0$$

for any solution $x(\cdot)$ in Ω , with the properties

1. $\{w = g(x, h(x)) \iff x = \tilde{g}(w, h(x))\}, (x, w) \in \Omega \times \mathcal{W}$
2. $\{w = g(x, h(x)) \implies h(x) = k(w)\}, (x, w) \in \Omega \times \mathcal{W}$

where g, \tilde{g}, l and k are smooth maps.

Then, one can look for an observer $\hat{w}(\cdot)$ of the system

$$\begin{cases} \dot{w} = l(w, h(x)) \\ y = k(w) \end{cases}$$

and take, as an estimator of $x(\cdot)$

$$\hat{x}(t) = \tilde{g}(\hat{w}(t), y(t)). \quad (4.15)$$

There is an advantage of considering such a change of coordinates when the maps g , \tilde{g} , l and k are independent of a parameter θ present in the expression of f , as in Proposition 2.6 of Section 2.5. However, note that the estimator (4.15) does not filter the measurement $y(\cdot)$ and might be sensitive to noise.

Let us illustrate this approach on the malaria model (Example 2.6).

Example 4.3 Consider the model (2.14) of Example 2.6. With the variable $w = x - Ey = (I - EC)x$ in \mathbb{R}^7 , the dynamics (2.15) is independent of the unknown non-linear term βSM :

$$\begin{cases} \dot{w} = \bar{A}w + \bar{A}Ey + \Lambda e_1 \\ y = Cx \end{cases}$$

and one can consider the following observer for system (2.15) in Luenberger form

$$\begin{cases} \dot{\hat{w}} &= \bar{A}\hat{w} + \bar{A}Ey + \Lambda e_1 + L(y(t) - C(\hat{w} + Ey(t))) \\ &= (\bar{A} - LC)\hat{w} + (L + (\bar{A} - LC)E)y(t) + \Lambda e_1, \\ \hat{x}(t) &= \hat{w}(t) + E y(t). \end{cases} \quad (4.16)$$

where L is a gains vector in \mathbb{R}^7 to be chosen. The dynamics of the error $e(t) = \hat{x}(t) - x(t)$ is given by

$$\dot{e} = (\bar{A} - LC)e$$

Note that one has $C\bar{A} = 0$. Therefore the rank of the observability matrix of the pair (\bar{A}, C) is equal to one, and 0 is an eigenvalue of \bar{A} . The choice of L allows then to assign only one eigenvalue of $\bar{A} - LC$, equal to $-(L_2 + L_3)$, the other eigenvalues remaining negative. Therefore (4.16) is an observer for system (2.15) with exponential convergence, that does not use the unknown parameter β .

Remark 4.2 Differently to observers of Section 4.2, one cannot expect a convergence speed of the observer (4.16) faster than the dynamics (2.14), because the error dynamics is not completely assignable. However, the convergence is exponential. This is illustrated with numerical simulations in Section 5.2.4.

4.5 Reduced-order observers

A typical situation is when one can operate a state decomposition when $m < n$, as follows

1. decompose the state vector x (may be at the price of a change of variables) as

$$x = \begin{bmatrix} y \\ x_u \end{bmatrix}$$

where $x_u \in \mathbb{R}^{n-m}$ represent the *unmeasured* variables,

2. look for an auxiliary variable (that we called z) whose dynamics is independent of x_u

$$\dot{z} = g(z, y(t)), \quad z \in \mathbb{R}^q$$

(for some q) and asymptotically stable (that is any solution $z(\cdot)$ converges to 0 when $t \rightarrow +\infty$), and such that x_u can be globally expressed as

$$x_u = l(z, y)$$

where l is a smooth map (say C^1).

Then, the dynamics

$$\begin{cases} \dot{\hat{z}} = g(\hat{z}, y(t)) \\ \hat{x}_u = l(\hat{z}, y(t)) \end{cases}$$

is an *asymptotic observer*, whose error convergence $\hat{x}_u - x_u$ is simply provided by the asymptotic convergence of \hat{z}_z to 0, whatever is the initial condition $\hat{z}(0) \in \mathbb{R}^q$. When the convergence speed of an estimator cannot be adjusted, it is usually called an *asymptotic observer*, differently to the previous section for which the error convergence can be made arbitrarily fast. Note that differently to the previous section, these observers have no tuning parameters and are not driven by innovation terms. These estimators are *reduced-order* observers when the variable z is of lower dimension than x (i.e. when $q < n$). An interest for such observers is that it can possess robustness features when the maps g and l are independent of some terms or parameters of the dynamics $\dot{x} = f(x)$. Let us illustrate this feature on the Kermack-McKendrick model with fluctuating rates.

Example 4.4 We consider the SIR model with birth and death terms

$$\begin{cases} \dot{S} = -\beta(t)SI + \nu N - \mu S \\ \dot{I} = \beta(t)SI - \rho(t)I - \mu I \\ \dot{R} = \rho(t)I - \mu R \end{cases} \quad (4.17)$$

where parameters β and ρ fluctuate unpredictably over time. We assume, for simplicity, that the birth rate ν is equal to the death rate μ , so that the total population remains constant of size $N = S + I + R$ (assumed to be known). Let us suppose that the size of the infected population is monitored over time as well as the number of new cured individuals, which amounts considering that the observation vector at time t is

$$y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} I(t) \\ \rho(t)I(t) \end{bmatrix}.$$

Stocks of classes S and R are not initially known. Then, the system

$$\begin{cases} \dot{Z} = \nu N - y_2(t) - \mu Z \\ \hat{S} = Z - y_1(t) \\ \hat{R} = N - Z \end{cases} \quad (4.18)$$

is an observer allowing to estimate S and R without knowing $\beta(\cdot)$ and $\rho(\cdot)$. Indeed, the dynamics of the estimators verifies

$$\begin{aligned}\frac{d}{dt}(\hat{S} - S) &= \dot{Z} - \dot{y}_1 - \dot{S} = -\mu(Z - S - I) = -\mu(\hat{S} - S) \\ \frac{d}{dt}(\hat{R} - R) &= -\dot{Z} - \dot{R} = -\nu N + \mu(Z + R) = -\mu(\hat{R} - R)\end{aligned}$$

which ensures the convergence of the \hat{S} and \hat{R} estimates. Note that the internal dynamics of the observer is here of smaller dimension than the system, and that the estimate of the unmeasured state variable S is a function of the internal state Z of the observer and the observation y_1 . The speed of convergence of this observer is not adjustable, but it has the advantage of being perfectly robust to any (unknown) variations of the terms $\beta(\cdot)$ and $\rho(\cdot)$. This is illustrated with numerical simulations in Section 5.2.3.

4.6 The high-gain observer for nonlinear systems

In the two previous examples, the dynamics of the estimation error was linear. For an observable non-linear system, the existence of an observer whose estimation error is linear is not guaranteed. This is a difficult problem. However, one can consider the (nonlinear) *observability canonical form* in \mathbb{R}^n [92] (given here for a scalar output i.e. for $m = 1$)

$$\dot{z} = F(z) := \underbrace{\begin{bmatrix} 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ & \ddots & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix}}_A z + \underbrace{\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}}_B \psi(z) \quad (4.19)$$

$$y = \underbrace{[1 \ 0 \ \cdots \ 0]}_C z \quad (4.20)$$

where the function ψ is Lipschitz on \mathbb{R}^n . Then, one can show that there exists an observer of the Luenberger form

$$\dot{\hat{z}} = F(\hat{z}) + G(C\hat{z} - y(t)) \quad (4.21)$$

with exponential convergence when G is a well-chosen gains vector. When an observable system

$$\begin{cases} \dot{x} = f(x), & x \in \mathbb{R}^n \\ y = h(x), & y \in \mathbb{R} \end{cases}$$

is not in normal form, but the application

$$\phi_n(x) = \begin{bmatrix} h(x) \\ \mathcal{L}_f h(x) \\ \vdots \\ \mathcal{L}_f^{n-1} h(x) \end{bmatrix}$$

is a diffeomorphism² from \mathbb{R}^n into \mathbb{R}^n and the function

$$\psi(z) := \mathcal{L}_f^m h \circ \phi_n^{-1}(z)$$

is Lipschitz on \mathbb{R}^n , then the observer (4.21) can be written in the x coordinates as follows

$$\dot{\hat{x}} = f(\hat{x}) + [J\phi_n(\hat{x})]^{-1}G(h(\hat{x}) - y(t))$$

where $J\phi_n(x)$ denotes the Jacobian matrix of ϕ at x . The observer preserves the Luenberger structure but with variable gains.

Let us first note that the pair (A, C) as defined in (4.19)-(4.20) is observable. Indeed, we have $O = Id$. Thus, according to the Lemma 4.1, one can freely assign the spectrum of $A + GC$ by choosing the vector G . We show now how to choose the eigenvalues of $A + GC$ to ensure the convergence of the non-linear observer (4.21). To do this, we begin by giving some properties of the Vandermonde matrices

$$V_{\lambda_1, \dots, \lambda_n} := \begin{bmatrix} \lambda_1^{n-1} & \lambda_1^{n-2} & \dots & \lambda_1 & 1 \\ \lambda_2^{n-1} & \lambda_2^{n-2} & \dots & \lambda_2 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ \lambda_n^{n-1} & \lambda_n^{n-2} & \dots & \lambda_n & 1 \end{bmatrix}$$

related to the normal form.

Lemma 4.2 *Let $\Lambda = \{\lambda_1, \dots, \lambda_n\}$ be a set of n distinct real numbers and G a vector such that*

$$Sp(A + GC) = \Lambda := \{\lambda_1, \dots, \lambda_n\}$$

Then

$$V_{\lambda_1, \dots, \lambda_n} (A + GC) V_{\lambda_1, \dots, \lambda_n}^{-1} = \begin{bmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \lambda_n \end{bmatrix}$$

² a diffeomorphism is an invertible map such that both the map and its inverse are differentiable.

Moreover, for any numbers $c > 0$ and $\theta > 0$, there exist $\lambda_n < \lambda_{n-1} < \dots < \lambda_1 < 0$ such that

$$\lambda_1 + c \|V_{\lambda_1(\theta), \dots, \lambda_n(\theta)}^{-1}\|_{\infty} = -\theta$$

The proof of Lemma 4.2 is given in Appendix A.2.

We are now ready to show the convergence of the observer (4.21) in coordinates z , for a gains vector G such that $A + GC$ has n distinct eigenvalues $\lambda_1, \dots, \lambda_n$ of negative real values. Denote the error $e = \hat{z} - z$. We have

$$\dot{e} = (A + GC)e + B(\psi(\hat{z}) - \psi(z))$$

Let $\xi = Ve$ where V designates the Vandermonde matrix $V_{\lambda_1, \dots, \lambda_n}$. Thanks to Lemma 4.2, we obtain

$$\dot{\xi} = \Delta\xi + VB(\psi(\hat{z}) - \psi(z))$$

where Δ is the diagonal matrix $\text{diag}(\lambda_1, \dots, \lambda_n)$. By multiplying on the left by ξ^T , one obtains

$$\begin{aligned} \xi^T \dot{\xi} &= \xi^T \Delta \xi + \xi^T VB(\psi(\hat{z}) - \psi(z)) \\ &\leq \lambda_1 \|\xi\|^2 + \|\xi\| \sqrt{n} |\psi(\hat{z}) - \psi(z)| \\ &\leq \lambda_1 \|\xi\|^2 + \|\xi\| \sqrt{n} L \|e\| \\ &\leq (\lambda_1 + \sqrt{n} L \|V^{-1}\|_{\infty}) \|\xi\|^2 \end{aligned}$$

where L is the Lipschitz constant of ψ . Thus the norm of ξ verifies

$$\|\xi(t)\| \leq \|\xi(0)\| + \int_0^t (\lambda_1 + \sqrt{n} L \|V^{-1}\|_{\infty}) \|\xi(\tau)\| d\tau$$

and by Gronwall's Lemma, we obtain

$$\|\xi(t)\| \leq \|\xi(0)\| e^{(\lambda_1 + \sqrt{n} L \|V^{-1}\|_{\infty})t}$$

Finally, for any $\theta > 0$, Lemma 4.2 gives the existence of numbers $\lambda_n < \lambda_{n-1} < \dots < \lambda_1 < 0$ such that $\|\xi(t)\| \leq \|\xi(0)\| e^{-\theta t}$ for any $t > 0$, which guarantees the exponential convergence of the error e to 0.

The observer (4.21) with the gains vector G_{θ} is called *high gain observer* [92], because the value of θ must be "sufficiently" large, and its successive powers might take large values.

Remark 4.3 In practice, the function ψ is not necessarily globally Lipschitz on \mathbb{R}^n , and even not properly defined outside Ω , while the observer (4.21) needs to be defined on whole \mathbb{R}^n . Nevertheless, if there exists a compact subset K of Ω that is forwardly invariant by the dynamics (4.19), one can consider an extension of ψ outside K that is globally Lipschitz on \mathbb{R}^n and define then the observer on whole \mathbb{R}^n (see for instance [100, 167]). This is illustrated on Example 4.5 below. It could also happen that for observable systems the map ϕ_n is not injective (and thus cannot be a diffeomorphism), but $\phi_{\tilde{n}}$ is injective for $\tilde{n} > n$. Then, it is theoretically possible to

embed the system in $\mathbb{R}^{\tilde{n}}$, that is to write the dynamics with a state vector in dimension $\tilde{n} > \bar{n}$ and then to construct a high gain observers in $\mathbb{R}^{\tilde{n}}$. Such construction is beyond the present book (see [167] for some techniques to build such extensions).

Let us now illustrate this construction on the Kermack-McKendrick model.

Example 4.5 We consider the classical SIR model

$$\begin{cases} \dot{S} = -\beta SI \\ \dot{I} = \beta SI - \rho I \\ \dot{R} = \rho I \end{cases} \quad (4.22)$$

where the parameters β and ρ are known, and suppose that the only observation is the cumulative number of recovered individuals since a time t_0

$$y(t) = \int_{t_0}^t \rho I(\tau) d\tau.$$

It is also assumed that the size of the total population $N = S + I + R$ is known. To put the system in canonical form, we write

$$\begin{aligned} z_1 &= y \\ z_2 &= \dot{z}_1 = \rho I \\ z_3 &= \dot{z}_2 = (\beta S - \rho)z_2 \end{aligned}$$

and

$$\dot{z}_3 = (\beta S - \rho)z_3 - \beta(\beta SI)z_2 = \psi(z) := \frac{z_3^2}{z_2} - \frac{\beta}{\rho} \left(\frac{z_3}{z_2} + \rho \right) z_2^2.$$

We can then reconstruct the I , S and R stocks from the variables z as follows

$$\begin{aligned} S &= \frac{1}{\beta} \left(\frac{z_3}{z_2} + \rho \right) \\ I &= \frac{z_2}{\rho} \\ R &= N - \frac{1}{\beta} \left(\frac{z_3}{z_2} + \rho \right) - \frac{z_2}{\rho} \end{aligned}$$

Note that ψ is not globally Lipschitz on \mathbb{R}^3 , and has a singularity at $z_2 = 0$. Nevertheless, we notice that the term z_3/z_2 can be framed as follows

$$\frac{z_3}{z_2} = \beta S - \rho \in [-\rho, \beta - \rho]$$

and that one has

$$\dot{z}_3 \in [-\rho^3 N - \rho\beta^2 N^3, \rho(\beta - \rho)N].$$

We can therefore consider the expression

$$\tilde{\psi}(z) = \text{sat}_{[-\rho^3 N - \rho \beta^2 N^3, \rho(\beta - \rho)N]} \left(\text{sat}_{[-\rho, \beta - \rho]} \left(\frac{z_3}{z_2} \right) z_3 - \frac{\beta}{\rho} z_3 z_2 - \beta z_2^2 \right)$$

instead of $\psi(z)$, where $\text{sat}_{[\cdot]}$ denotes the saturation function

$$\text{sat}_{[a,b]}(x) = \max(a, \min(b, x)).$$

Finally, we choose the gains G_i of the observer such that $Sp(A+GC) = \{\lambda_1, \dots, \lambda_n\}$ with $\lambda_n < \lambda_{n-1} < \dots < \lambda_1 < 0$ and λ_1 *enough* negative. This amounts to take $G_i = \sigma_i(\{\lambda_1, \dots, \lambda_n\})$. One thus obtains the internal dynamics of the observer

$$\begin{cases} \dot{\hat{z}}_1 = \hat{z}_2 + (\lambda_1 + \lambda_2 + \lambda_3)(\hat{z}_1 - y(t)) \\ \dot{\hat{z}}_2 = \hat{z}_3 + (\lambda_1 \lambda_2 + \lambda_1 \lambda_3 + \lambda_2 \lambda_3)(\hat{z}_1 - y(t)) \\ \dot{\hat{z}}_3 = \tilde{\psi}(\hat{z}) + \lambda_1 \lambda_2 \lambda_3 (\hat{z}_1 - y(t)) \end{cases} \quad (4.23)$$

and the estimators

$$\begin{cases} \hat{S} = \frac{1}{\beta} \left(\text{sat}_{[-\rho, \beta - \rho]} \left(\frac{\hat{z}_3}{\hat{z}_2} \right) + \rho \right) \\ \hat{I} = \frac{\hat{z}_2}{\rho} \\ \hat{R} = N - \hat{S} - \hat{I} \end{cases} \quad (4.24)$$

4.7 Discussion

The construction of an observer can avoid in certain situations to study the identifiability. For instance, in Sections 4.4 and 4.5, the observers do not require the knowledge of all the parameters of the model, and even of some functions involved in the model. This is why such observers are also called *unknown-inputs observers*. The theory of unknown-inputs observers has been mainly studied for linear systems [107, 45, 152]. Very few general results are available for nonlinear systems (this research field is today largely open).

The existence of observers without the possibility of fixing arbitrarily the speed of convergence, as in Examples 4.3 and 4.4, is connected to the property of *detectability* (see for instance [148, 10]), which is a weaker than observability: a system

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

is *detectable* (in Ω) if for any pair of solutions $x^a(\cdot), x^b(\cdot)$ in Ω , one has

$$\{h(x^a(t)) = h(x^b(t)), t \geq 0\} \Rightarrow \lim_{t \rightarrow +\infty} x^a(t) - x^b(t) = 0.$$

For linear dynamics $\dot{x} = Ax, y = Cx$ that are not observable, there exists a *Kalman decomposition* [119] i.e. an invertible matrix P such that

$$PAP^{-1} = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}, \quad CP^{-1} = [C_1 \ 0]$$

with $A_{11} \in \mathcal{M}_{l \times l}, C_1 \in \mathcal{M}_{1 \times l}$ where $l < n$ is the rank of the observability matrix O recalled in (2.4), such that the subsystem $\dot{z} = A_{11}z, y = C_1z$ is observable. Then, the system is detectable when the matrix A_{22} is Hurwitz. This is exactly the case of Example 4.3 with $l = 1$.

Let us mention the more recent technique proposed by Kazantzis and Kravaris for obtaining Luenberger like observers for non-linear systems. It consists in looking at a (non-linear) change of coordinates as a diffeomorphism $T: x \mapsto \zeta = T(x)$ such that the dynamics of the variable ζ writes

$$\dot{\zeta} = A\zeta + B(h(T^{-1}(x)))$$

where A is a Hurwitz matrix and B a smooth map. Then,

$$\dot{z} = AZ + B(y), \quad \hat{x} = T^{-1}(z)$$

is a natural observer, whose speed of convergence is given by the spectrum of the matrix A . However, the map T has to be found as a solution of the partial derivative equation

$$\frac{\partial T}{\partial x}(x)f(x) = AT(x) + B(h(x))$$

and has to be Lipschitz and invertible for \hat{x} to be well defined and converging. This is a difficult problem to solve (see [126, 11]), still open in general. This is why we have not presented this method.

Other approaches consider non-smooth observers. In particular, the following observer proposed by Levant [129]

$$\begin{cases} \dot{\hat{z}}_1 = z_2 - K_1 L^{\frac{1}{3}} |z_1 - y|^{\frac{2}{3}} \text{sign}(z_1 - y) \\ \dot{\hat{z}}_2 = z_3 - K_2 L^{\frac{1}{2}} |z_1 - y|^{\frac{1}{2}} \text{sign}(z_1 - y) \\ \dot{\hat{z}}_3 = -K_3 L \text{sign}(z_1 - y) \end{cases}$$

reconstructs theoretically \dot{y} and \ddot{y} in finite time from the measurement y , provided that the solutions of the original system are bounded. However, such observers and more generally sliding-mode observers [196], are extremely sensitive to measurement noise. Although well employed for mechanical or electro-mechanical systems, we believe that there are not very well suited to epidemiological models.

Chapter 5

Practical and numerical considerations

Abstract The objective of this chapter is to show how to put theory in practice, illustrated in some cases studies.

5.1 Practical identifiability

Till now we have studied structural identifiability/observability . While structural identifiability is a property of the model structure, given a set of outputs, practical identifiability is related to the actual data. In particular, it depends on the amount of information contained in the data.

A model can be structurally identifiable, but still be practically unidentifiable due to poor data quality, e.g., bad signal-to-noise ratio, errors in measurement or sparse sampling [170]. Structural identifiability means that parameters are identifiable with ideal (continuous, noise-free) data. While structural identifiability is a prerequisite for parameter identification, it does not guarantee that parameters are practically identifiable with a finite number of noisy data points.

Moreover, parameter estimation requires using numerical optimization algorithms. The distance, for the problem considered, to the nearest ill-posed problem, [66, 104], i.e., the conditioning of the problem, can challenge the convergence of algorithms. Another source of practical unidentifiability is the lack of information from the data, i.e., the signal from the data does not satisfy the persistence of excitation [134]. This is the case when the observation is near an equilibrium [22].

In this section we use sensitivity analysis and results from asymptotic statistical theory to study practical identifiability. We refer to previous surveys and papers on the topic [51, 20, 21, 61, 23]. Our purpose here is to give an intuitive account of these techniques.

5.1.1 Rationale for using sensitivity analysis

Practical identifiability is often assessed in terms of confidence intervals on parameters [219]. Confidence intervals can be derived from the Fisher Information Matrix (FIM) [32]. More specifically, the covariance matrix (Σ) of the estimated parameters may be approximated as the inverse of the FIM. The diagonal elements of $\Sigma \approx \text{FIM}^{-1}$ correspond to the variance of the parameter estimates. Their square-roots (the standard deviations) give confidence intervals on the parameters, thus providing information on practical identifiability.

In the least-squares framework, the Fisher Information Matrix can be expressed in terms of sensitivity matrices, that we define below.

5.1.2 Observed system

We consider that the initial condition x_0 is unknown. Unless otherwise specified, the term “parameter” now refers to both the parameter θ and the initial condition x_0 , i.e. $\Theta = (\theta, x_0)$. We make explicit the dependence of the state variables x and y on Θ to clarify the following derivations:

$$\begin{cases} \dot{x}(t, \Theta) = f(x(t, \Theta), \theta), & x(0, \Theta) = x_0, \\ \dot{\theta}(t) = 0, & \theta(0) = \theta, \\ y(t, \Theta) = h(x(t, \Theta), \theta). \end{cases} \quad (5.1)$$

with $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and $\theta \in \mathbb{R}^p$.

5.1.3 Sensitivity analysis

We wish to quantify how the observed variable $y(t, \Theta)$ changes for a small parameter variation $\Delta\Theta$.

We denote the Jacobian of the observation $y(t, \Theta)$ with respect to the parameter Θ as

$$\chi(t, \Theta) = \frac{\partial y}{\partial \Theta}(t, \Theta).$$

This $m \times (n + p)$ matrix is called the sensitivity matrix.

By linearization (first-order Taylor approximation), one can write

$$\Delta y(t, \Theta) = \chi(t, \Theta) \Delta\Theta.$$

Side remarks

Reid [173] defined a parameter vector as “sensitivity identifiable” if the above equation can be solved uniquely for $\Delta\Theta$. This linear problem is well known: if χ has maximal rank then the solution is given by means of the Moore-Penrose pseudo-inverse $\chi^+ = (\chi^T \chi)^{-1} \chi^T$:

$$\Delta\Theta = \chi^+(t, \Theta) \Delta y(t, \Theta).$$

It is also well known [93] that the sensitivity of this solution is ruled by the condition number $\kappa_2(\chi) = \sigma_{\max}/\sigma_{\min}$ with σ_{\max} and σ_{\min} respectively the greatest and smallest singular value of χ (which are the square roots of the corresponding eigenvalues of $\chi^T \chi$).

5.1.4 Ordinary Least Squares

Now we consider a set of M observations $Y_i, i = 1, \dots, M$, that have been obtained at times t_i . We assume that the observation is given by

$$Y_i = y(t_i, \Theta) + \mathcal{E}_i,$$

with the error \mathcal{E}_i assumed to be a random variable satisfying the following assumptions:

- the errors \mathcal{E}_i have mean zero $E[\mathcal{E}_i] = 0$;
- the errors have a constant variance $\text{var}(\mathcal{E}_i) = \sigma^2$;
- the errors are independent and identically distributed.

The Fisher Information Matrix, for the preceding defined observations, is defined as

$$\text{FIM}(\Theta, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^M \chi(t_i, \Theta)^T \chi(t_i, \Theta). \quad (5.2)$$

Solving the ordinary least square (OLS) equations gives an estimator $\hat{\Theta}_{\text{OLS}}$ of the parameter Θ :

$$\hat{\Theta}_{\text{OLS}} = \arg \min_{\Theta} \sum_{i=1}^M (Y_i - y(t_i, \Theta))^2 \quad (5.3)$$

Even though the error’s distribution is not specified, asymptotic statistical theory can be used to approximate the mean and variance of the estimated Θ (a random variable) [185, 23]: the bias-adjusted approximation for σ^2 (with $n+p$ “parameters”) is

$$\hat{\sigma}_{\text{OLS}}^2 = \frac{1}{M - (n + p)} \sum_{i=1}^M \left(Y_i - y(t_i, \hat{\Theta}_{\text{OLS}}) \right)^2. \quad (5.4)$$

5.1.5 Confidence intervals

The above approximation of the error variance can be used to further approximate the parameter covariance matrix Σ :

$$\hat{\Sigma} := \left[\text{FIM} \left(\hat{\Theta}, \hat{\sigma}_{\text{OLS}}^2 \right) \right]^{-1}. \quad (5.5)$$

The standard error (SE) for $\hat{\Theta}_{\text{OLS}}$ can be approximated by taking the square roots of the diagonal elements of the covariance matrix Σ : for all $k = 1, \dots, n + p$,

$$\text{SE} \left(\hat{\Theta}_{\text{OLS}}(k) \right) = \sqrt{\hat{\Sigma}_{kk}}. \quad (5.6)$$

Finally, to compute the 95% confidence interval for the k -th component of the parameter vector $\hat{\Theta}_{\text{OLS}}$ with $n + p$ “parameters”, one may use the Student’s t -distribution with $M - (n + p)$ degrees of freedom: letting

$$\zeta(k) = t_{0.025}^{M-(n+p)} \times \text{SE} \left(\hat{\Theta}_{\text{OLS}}(k) \right),$$

the confidence interval is defined as

$$\hat{\Theta}_{\text{OLS}}(k) - \zeta(k) < \hat{\Theta}_{\text{OLS}}(k) < \hat{\Theta}_{\text{OLS}}(k) + \zeta(k).$$

From these formulas it appears that the conditioning of the Fisher Information Matrix plays an essential role. Huge confidence intervals give indications about the practicality of the identification.

5.1.6 Computing the sensitivity matrix

The sensitivity matrix $\chi(t, \Theta)$, with $\Theta = (\theta, x_0)$, is obtained by integrating an ODE. The components of the ODE to be integrated depend on whether one differentiates with respect to θ or x_0 .

Differentiating with respect to θ

The first part of the ODE is given by

$$\frac{\partial y}{\partial \theta}(t, \theta, x_0) = \frac{d}{dt} h(x(t, \theta, x_0), \theta) = \frac{\partial h}{\partial x} \frac{\partial x}{\partial \theta}(t, x_0, \theta) + \frac{\partial h}{\partial \theta}(x(t, \theta, x_0), \theta).$$

The Jacobian $\frac{\partial h}{\partial x}$ is a $m \times n$ matrix while $\frac{\partial h}{\partial \theta}$ is a $m \times p$ matrix.

We then have to compute the $n \times p$ matrix

$$z(t, \Theta) := \frac{\partial x}{\partial \theta}(t, \theta, x_0) = \frac{\partial x}{\partial \theta}(t, \Theta).$$

Let $A(t, \Theta)$ and $B(t, \Theta)$ be the following time-dependent $n \times n$ and $n \times p$ matrices, respectively:

$$A(t, \Theta) := \frac{\partial f}{\partial x}(x(t, \Theta), \theta),$$

and

$$B(t, \Theta) := \frac{\partial f}{\partial \theta}(x(t, \Theta), \theta).$$

It is well known [101] that $z(t)$ is the solution of the linear matrix equation:

$$\dot{z}(t, \Theta) = A(t, \Theta) z(t, \Theta) + B(t, \Theta),$$

with the initial condition $z(0, \Theta) = 0_{n \times p}$ (a zero matrix of size $n \times p$).

Differentiating with respect to x_0

The second part of the ODE is given by

$$\frac{\partial y}{\partial x_0}(t, \theta, x_0) = \frac{\partial}{\partial x_0} h(x(t, \theta, x_0), \theta) = \frac{\partial h}{\partial x} \frac{\partial x}{\partial x_0}(t, \theta, x_0).$$

Let

$$w(t, \Theta) := \frac{\partial x}{\partial x_0}(t, \theta, x_0) = \frac{\partial x}{\partial x_0}(t, \Theta).$$

Based on the same reference [101], $w(t, \Theta)$ is solution of the linear matrix ODE

$$\dot{w}(t, \Theta) = A(t, \Theta) w(t, \Theta),$$

with the initial condition $w(0, \Theta) = \text{Id}_{n \times n}$ (the identity matrix of size n).

Full system

To summarize, one has to solve the following system in dimension $n^2 + np + n$

$$\begin{cases} \dot{x}(t, \Theta) = f(x(t, \Theta), \theta), & x(0, \Theta) = x_0, \\ \dot{z}(t, \Theta) = A(t, \Theta) z(t, \Theta) + B(t, \Theta), & z(0, \Theta) = 0_{n \times p}, \\ \dot{w}(t, \Theta) = A(t, \Theta) w(t, \Theta), & w(0, \Theta) = \text{Id}_{n \times n} \end{cases} \quad (5.7)$$

with $A(t, \Theta) = \frac{\partial f}{\partial x}(x(t, \Theta), \theta)$ and $B(t, \Theta) = \frac{\partial f}{\partial \theta}(x(t, \Theta), \theta)$.

For large systems, the computation of the different Jacobians can be prohibitive, in this case automatic differentiation software has to be used.

5.1.7 Some case studies

In this section, we consider two classical examples as case studies. These examples have been used in many books of mathematical epidemiology, e.g. [150].

Case 1. Influenza in a boarding school

Our first example is an outbreak of influenza in a United Kingdom boarding school which occurred in 1978 [44]. In [150] the parameters β, γ are identified by an unspecified “best-fit” algorithm. A more complete analysis is done in [35], where the analysis is done using sensitivity analysis and asymptotic statistical theory. In [140] the same example is considered. Different sources exist for the data [36, 35, 65] with small differences.

Using the figure in [44] and the *Plot Digitizer* software, we got an approximation of the data. It was reported that $N = 763$, and the conditions at the start of this outbreak were $S_0 = 762$ and $I_0 = 1$. We used the following data, in which time t is in day and $I(t)$ denotes the number of infectious people at time t .

t	0	1	2	3	4	5	6	7	8	9	10	11	12	13
$I(t)$	1	6	26	73	222	293	258	237	191	124	68	26	10	3

Specifically, we considered model (3.4) with $k = 1$ (all infectious are assumed to be observed). We obtained the following OLS estimation, as given by the Scilab software:

$$\beta \approx 1.96, \quad \gamma \approx 0.475$$

(see the numerical code in Appendix B). The fit is shown in Figure 5.1.

We computed the 95% confidence intervals using the formulas given in the preceding section: (5.2), (5.3), (5.4), (5.5), (5.6), up to a few changes due to the fact that the initial conditions $x_0 = (S_0, I_0)$ are assumed to be known in this example (see Appendix B). We find:

$$\begin{aligned} \beta &\approx 1.96 \pm 0.073, \\ \gamma &\approx 0.475 \pm 0.04. \end{aligned}$$

One can obtain approximately the same results using the likelihood profile method to compute confidence intervals [32], assuming normally distributed errors. However,

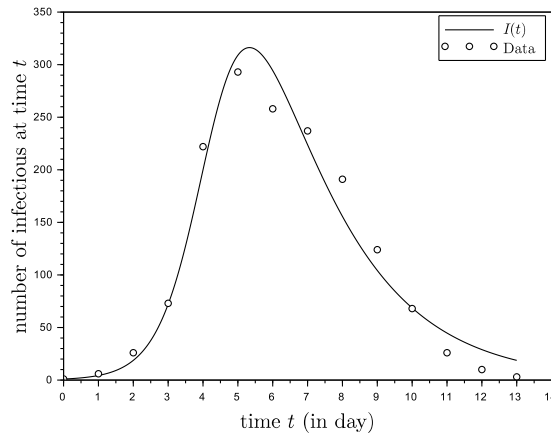


Fig. 5.1 Boarding School example.

it is well known that the profile method quickly becomes impractical for model with more than two parameters [32], which is the rule rather than the exception, as will be the case in the following example. This is why we stick to the FIM method. Lastly, we note that the condition number of the FIM is approximately equal to 3.80 in this example.

Case 2. Plague in Bombay

Our second example is the Bombay Plague of 1905–1906 [122]. We collected the data from [59, Table IX], over the same period as [122] (Dec. 17 to Jul. 21). The form of the data is presented in the following table, in which time t is in week, and $\dot{R}(t)$ denotes the number of death per week at time t .

t	0	1	2	3	4	5	6	...	24	25	26	27	28	29	30
$\dot{R}(t)$	8	10	12	16	24	48	51	...	106	64	46	35	27	28	24

We consider that the number of death per week is the same as $\dot{R}(t) = \gamma I(t)$, meaning that all infections lead to death, which is a reasonable assumption in this context [17]. Therefore, we consider model (3.4) with $k = \gamma$. In this example, not only the parameters β and γ , but also the size of the population, N , as well as the initial conditions, S_0 and I_0 , are unknown [17]. According to Theorem 3.1 (and Remark 3.3), the model is neither observable nor identifiable. However, the model is

partly identifiable in the sense that S_0 , I_0 , γ and $\tilde{\beta} = \beta/N$ are structurally identifiable. Starting from an arbitrary initial guess, we obtained the following OLS estimation, as given by the Scilab software:

$$\tilde{\beta} \approx 0.0000855, \quad \gamma \approx 3.72, \quad S_0 \approx 4.81 \cdot 10^4, \quad I_0 \approx 1.42,$$

(see the numerical code in Appendix C). The fit is shown in Figure 5.2.

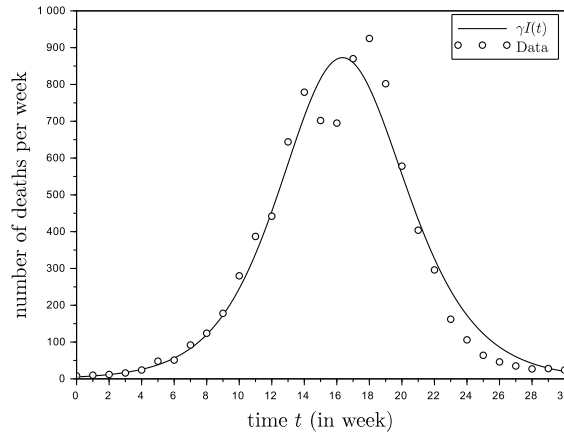


Fig. 5.2 Plague in Bombay example.

Proceeding as in the previous example, we obtained the following 95%-confidence intervals:

$$\tilde{\beta} \approx 0.0000855 \pm 0.00157,$$

$$\gamma \approx 3.72 \pm 25,$$

$$S_0 \approx 4.81 \cdot 10^4 \pm 6 \cdot 10^4,$$

$$I_0 \approx 1.42 \pm 34.$$

The confidence intervals are huge, which means that we can have absolutely no confidence in the estimated values of the parameters, even though the fit looks good and these parameters are structurally identifiable in principle. In practice, one can show that many other and very different combinations of the parameters can yield approximately the same fit. Note that if we did as if the initial conditions were known, the confidence intervals on $\tilde{\beta}$ and γ would be reasonable, as in the previous example. In this example, the condition number of the FIM is approximately 9.14×10^{24} , meaning that the problem is “sloppy” [48]; note however that while it is usual that a model is both sloppy and practically non-identifiable, this is not always the case [58]. Altogether, we can conclude that there is a severe practical identifiability issue in this classical example.

5.1.8 Discussion

The SIR model of Kermack-McKendrick has been studied in a series of papers [52, 51, 53, 40] where the problem of observability/identifiability is approached from the statistical point view: addressing parameter identifiability by exploiting properties of both the sensitivity matrix and uncertainty quantification in the form of standard errors. In this series of papers, structural observability and identifiability were not explicitly addressed. For example in [53] the authors identify $(S_0, I_0, \beta/N, \gamma)$ based on incidence observations - akin to equation (3.5) with $k = 1$ - which we have proved to be structurally identifiable (see Theorem ??). Similarly in [40] the authors seek to identify $(S_0, I_0, \beta, \gamma)$ which, with N known, are structurally identifiable in principle. However the authors encounter practical identifiability issues. This is a typical example of strictly practical unidentifiability (as in our second example, the Plague in Bombay).

Although a structural observability and identifiability analysis should be done as a prerequisite to a practical identifiability analysis, it does not suffice. Moreover, when doing practical identifiability analyses, the error structure of the data should be considered. For instance, sensitivity analyses can be extended to non-constant error variance through Generalized Least Squares (GLS), which makes it possible to test different ways of weighting errors [52]. Appendix D shows the principle of GLS and how to compute the Fisher Information Matrix in this case. We also provide an online example¹ based on the “Influenza in a Boarding School” data (since the “Plague in Bombay” example generates numerical optimization issues related to its practical non-identifiability).

An additional issue may occur when the output signal is not sufficiently informative (i.e., not *persistently exciting* [134]). For example when the data correspond to states near unobservability, e.g., near an equilibrium. In those cases, one has to wait to have data sufficiently far from equilibrium.

To conclude, the problem of observability and identifiability, either structural or practical, is far from being simple, even in relatively simple SIR models with seemingly good quality data [52]. Of course, the more complex the model, the more parameters there are to identify, the more serious the problem of identifiability.

5.2 Observers in practice

In this section, we show how the various observers presented in Section 4 behave in practice, and the role of the tuning parameters. Up to now, we have assumed the measurements to be perfect i.e. not tainted with any noise. Since integration has good “averaging” properties, an observer is expected to filter noise or inaccuracies in the measurements. However, we will see that the filtering capacity of an observer is

¹ <https://github.com/nikcunniffe/Identifiability>

related to his convergence speed, which often leads to a “precision-speed” dilemma in the choice of the observer or his settings.

Let us underline that when identifiability/observability cannot be proved theoretically or is too difficult to be proven analytically, one can still look for an observer and study its asymptotic convergence, theoretically or numerically.

5.2.1 Observers with linear assignable error dynamics

We illustrate the observer (4.8) of the age-structured model (4.7) on simulations, for the following values of the parameters.

a_1	a_2	m_1	m_2	m_3	k	\bar{r}_{min}	\bar{r}_{max}
0.1	0.1	0.05	0.07	0.07	1.0	0.9	1.1

The following code is used to compute the gain vector G for a set of desired eigenvalues.

```
a1=0.1;a2=0.1;m1=0.05;m2=0.07;m3=0.07;
Sp=0.3*[-0.3,-0.33,-0.36];
A=[-a1-m1,0,0;
a1,-a2-m2,0;
0,a2,-m3];
C=[0,0,1];
B=[0;0;1];
Obs=[C;C*A;C*A*A];
L=inv(Obs)*B;
P=[L,A*L,A*A*L];
Abar=inv(P)*A*P;
sigma=coeff(poly(Sp,'x'));
G=P*(-sigma(1:3)'-Abar(:,3));
```

Figure 5.3 shows convergence for a moderately negative spectrum, while Figure 5.4 shows the acceleration of convergence obtained for a spectrum located further to the left in the complex plane. For the same choice of gains, Figures 5.5 and 5.6 show the effect of noise on the $y(\cdot)$ measurements. It can be seen that a faster convergence is more sensitive to noise and loses accuracy. In practice, one often has to make a compromise for the choice of the observer’s setting.

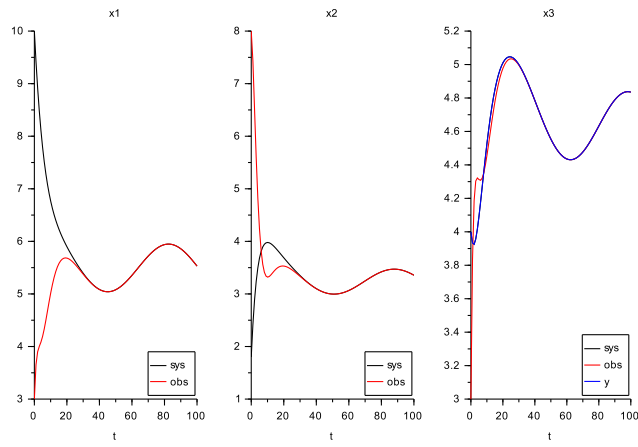


Fig. 5.3 $Sp(A + GC) = \{-0.3, -0.33, -0.36\}$ without measurement noise

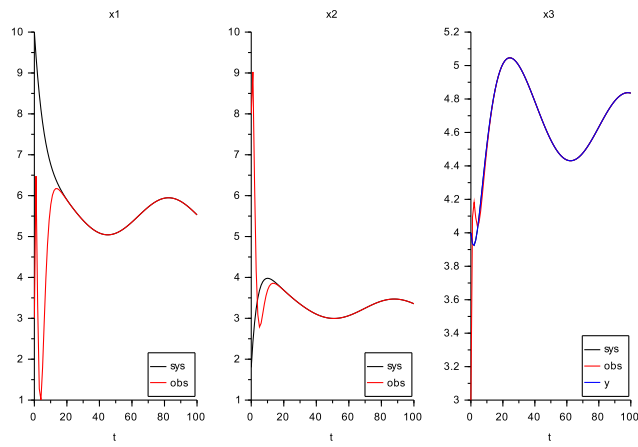


Fig. 5.4 $Sp(A + GC) = \{-0.6, -0.66, -0.72\}$ without measurement noise

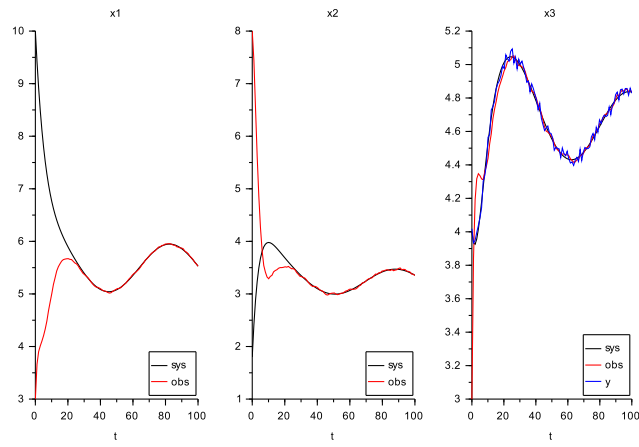


Fig. 5.5 $Sp(A + GC) = \{-0.3, -0.33, -0.36\}$ with measurement noise

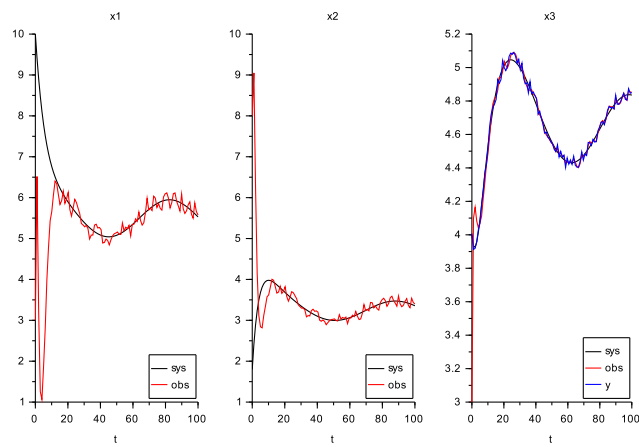


Fig. 5.6 $Sp(A + GC) = \{-0.6, -0.66, -0.72\}$ with measurement noise

5.2.2 About observers with Lipchitz non-linearity

Let us illustrate the result on Exemple 4.2 with the following numerical values of the parameters

β	γ	μ
0.13	0.1	0.05

The gain vector G such that $Sp(A + GC) = \{-1, -2\}$, and the corresponding matrix P solution of the Lyapunov equation (4.11) can be numerically computed (using, for instance, scilab) as

$$G = \begin{bmatrix} -23.889 \\ 20.909 \end{bmatrix} \quad P = \begin{bmatrix} 73.135667 & 83.710112 \\ 83.710112 & 96.199374 \end{bmatrix}$$

with

$$\text{norm}(P) = 169.16821$$

for which the condition (4.12) gives $\varepsilon \leq 0.001477$. As an illustration, for a total population size of one billion, this gives $I \leq 1477$, which is a very small number..

This example shows that this technique is not well suited to epidemiological models such as the SIRS one, because it requires a too small Lipschitz constant of the non linear terms. However, we consider useful to have exposed this known approach and shown its drawback.

5.2.3 Observers with asymptotic convergence

We illustrate on simulations the behavior of the asymptotic observer (4.18) of the SIR model with fluctuating rates (4.17), for the following values of the parameters.

β	ρ	ν	μ	N
0.4 ± 0.08	0.2 ± 0.04	0.05	0.05	1000

Here β and ρ are functions of time chosen randomly in between the bounds given in the table. Figures 5.7 and 5.8 show that the observer has a convergence relatively insensitive to measurement noise, but the speed of convergence is slow because the exponential decay of the error is equal to μ , which is not adjustable.

Unlike the observers in previous sections, let us underline that the present observer is not based on a $\hat{y} - y$ innovation. Therefore, one is not informed of the quality of the estimate over time, which is a *price to pay* to have a observer insensitive to unknown variations of the epidemic parameters β, ρ .

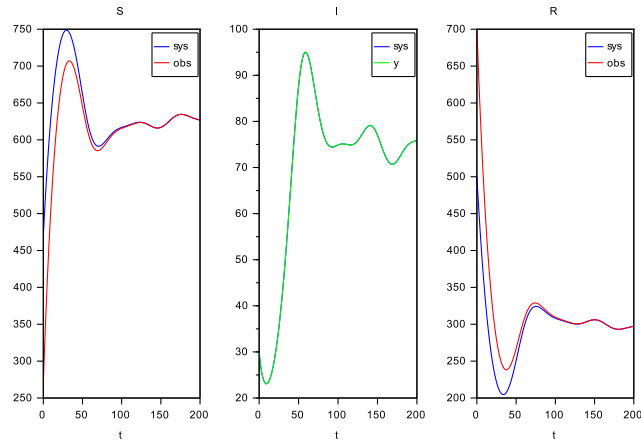


Fig. 5.7 The high gain observer without measurement noise

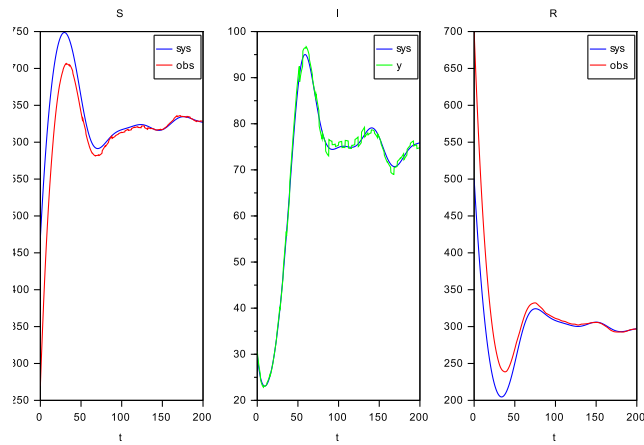


Fig. 5.8 The high gain with measurement noise

5.2.4 Observers with partially assignable error dynamics

The observer given in (4.16) for the intra-host malaria model (2.14) is illustrated here on real data, as one can see for instance on Figure 5.9.

As already mentioned in Remark 4.2, the speed of convergence of this observer cannot be tuned as fast as desired. However, this is quite satisfactory in practice.

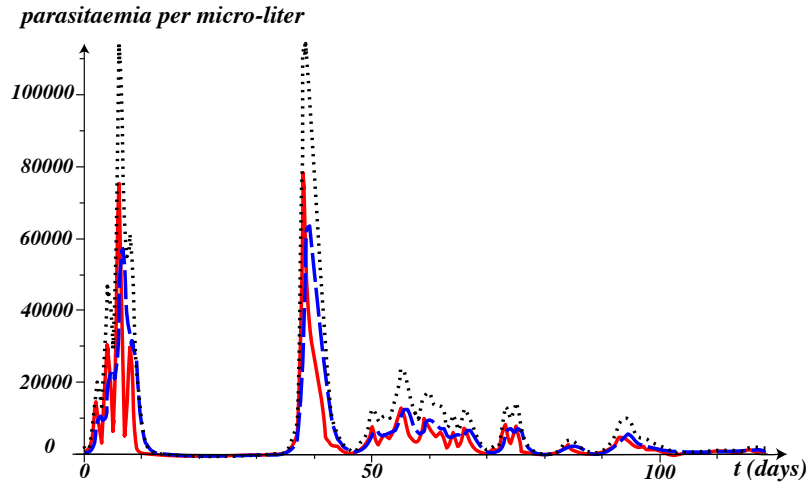


Fig. 5.9 Example of patient S1204: Measures (data) of peripheral parasitaemia are plotted with red solid line. The estimations delivered by the observer (4.16) are plotted with blue dashed line for the estimated sequestered parasitaemia, and with black dotted line for the estimated total parasitaemia. The gain used is $L = (0, 5, 5, 0, 0, 0, 0)^T$. $Sp(\bar{A} - LC) = \{-10, -\mu_S, -\mu_M, -\mu_5 - \gamma_5, -\mu_4 - \gamma_4, -\mu_3 - \gamma_3, -\mu_2 - \gamma_2 - \gamma_1\}$.

Let us also underline that the observer does not require the reconstruction of the parameter β , although this parameter is identifiable (see Section 2.5). This is a strength of this observer, because the parameter β could switch or fluctuate with time.

5.2.5 High gain observer

The non-linear observer (4.23)-(4.23) of the classical SIR model (4.22) is illustrated on simulations for the following values

β	ρ	N
0.4	0.1	10000

where the y cumulative measures were made discretely every day (rounded to the nearest integer). In order to obtain a time-continuous $y(\cdot)$ signal, we performed an interpolation by cubic splines. Figure 5.10 shows the convergence of the observer for the eigenvalues $\{-2, -2.2, -2.4\}$. We also simulated the observer when the measurements are corrupted by random counting errors up to ± 5 individuals per day (see Figure 5.11). As for the adjustable observer in Section 5.2.1, these simulations show

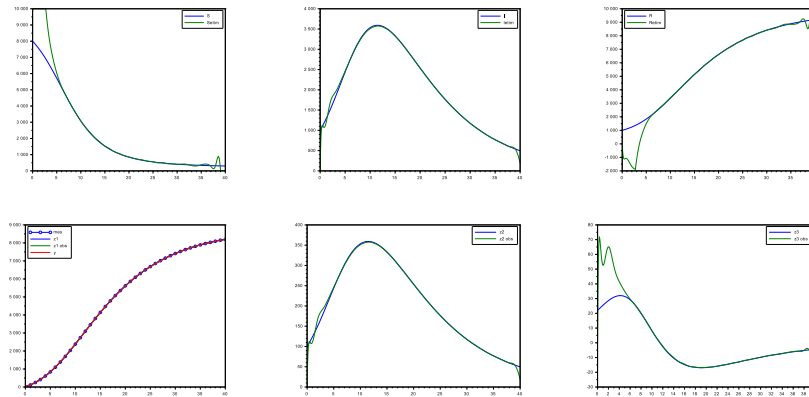


Fig. 5.10 Observer simulations (top: variables S , I , R and their estimates; bottom: coordinates z_i with measurements points on the left)

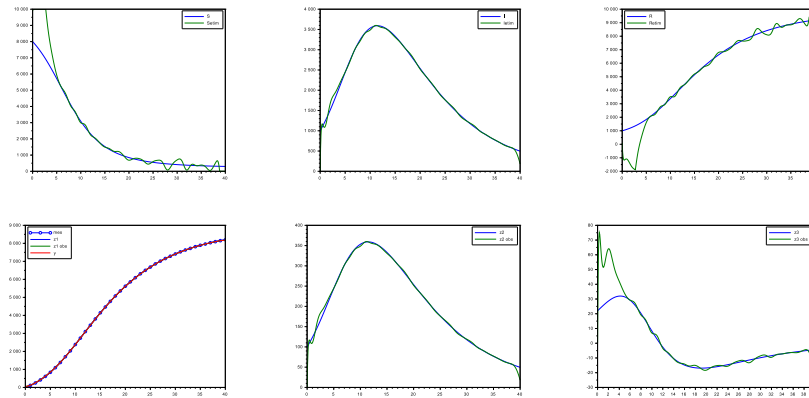


Fig. 5.11 Observer simulations with noisy measurements

the dilemma *accuracy versus speed* of the estimation in presence of measurement noise.

We end this section by showing an example for which an observer is used to reconstruct state and parameter simultaneously.

5.3 A case study : An observer to estimate state and parameter

Consider the following simple bilharzia transmission model [25]:

$$\begin{cases} \dot{w} = -\gamma w + as \\ \dot{s} = b(1-s)w - \mu s \end{cases} \quad (5.8)$$

where w is the number of female schistosomes (worms) *per* single host and s the proportion of infected snails. The female schistosomes *per* single host decays at a per capita rate γ and are replenished at a rate a . The latter process is proportional to the proportion of infected snails s . The proportion of susceptible snails $1-s$ are infected through an indirect contact with schistosomes that are excreted from hosts w at rate b , and naturally die at a rate μ .

The number of female schistosomes (worms) *per* single host can be measured using urine or faeces samples. Therefore, we can assume that the measurable output is $y(t) = w(t)$.

It is easy to show that the compact set

$$\Omega = \{(w, s) \in \mathbb{R}^2 \mid 0 \leq w \leq \frac{a}{\gamma}; 0 \leq s \leq 1\} \quad (5.9)$$

is a positively invariant set for System (5.8). The measurable output $y(t)$ satisfies the equation $\dot{y} = -\gamma y + as$. Hence, $y(t) > 0$ for all $t \geq 0$ if $y(0) > 0$.

We assume that the ecosystem is in an endemic situation which means that the basic reproduction number $\mathcal{R}_0 = \frac{ab}{\gamma\mu} > 1$ and implies [25] that $y(t)$ converges to $\bar{y} = \frac{a}{\gamma} \left(1 - \frac{\gamma\mu}{ab}\right) > 0$ as time goes to infinity.

A crucial problem in epidemic models is the estimation of the transmission parameters. For the schistosomiasis model, it is the parameter a which represents the snail-host infection rate that is difficult to estimate [25].

In order to study the observability and identifiability of model (5.8), we consider the augmented system by adding the unknown parameter a to the augmented state $x = (w, s, a)^T$:

$$\begin{cases} \dot{x} = \begin{bmatrix} \dot{w} \\ \dot{s} \\ \dot{a} \end{bmatrix} = \begin{bmatrix} -\gamma w + as \\ b(1-s)w - \mu s \\ 0 \end{bmatrix} = f(x) \\ y = w = h(x) \end{cases} \quad (5.10)$$

One determines

$$\dot{y} = \mathcal{L}_f h(x) = \langle \nabla h(x) | f(x) \rangle = -\gamma w + as$$

$$\begin{aligned} \ddot{y} = \mathcal{L}_f^2 h(x) &= \langle \nabla \mathcal{L}_f h(x) | f(x) \rangle = \left\langle \begin{bmatrix} -\gamma \\ a \\ s \end{bmatrix} \middle| \begin{bmatrix} -\gamma w + as \\ b(1-s)w - \mu s \\ 0 \end{bmatrix} \right\rangle \\ &= -\gamma(-\gamma w + as) + a(b(1-s)w - \mu s). \end{aligned}$$

The map

$$\begin{aligned} H_3 : \{0 < w \leq \frac{a}{\gamma}; 0 < s < 1; a > 0\} &\rightarrow \mathbb{R}^3 \\ x &\mapsto H_3(x) = \left(h(x), \mathcal{L}_f h(x), \mathcal{L}_f^2 h(x) \right) \end{aligned}$$

is injective. Indeed

$$\begin{aligned} H_3(\bar{x}) = H_3(x) &\Rightarrow \begin{cases} \bar{w} = w, \\ -\gamma\bar{w} + \bar{a}\bar{s} = -\gamma w + as \\ \bar{a}(b(1-\bar{s})\bar{w} - \mu\bar{s}) = a(b(1-s)w - \mu s) \end{cases} \\ \Rightarrow \begin{cases} \bar{w} = w, \\ \bar{a}\bar{s} = as, \\ \bar{a}b(1-\bar{s})\bar{w} = ab(1-s)w \end{cases} &\Rightarrow \begin{cases} \bar{w} = w, \\ \bar{a}\bar{s} = as, \\ \bar{a} = a \text{ (since } b > 0, w > 0, \bar{w} > 0, s < 1, \bar{s} < 1). \end{cases} \\ &\Rightarrow \begin{cases} \bar{w} = w, \\ \bar{s} = s \text{ (since } a > 0), \\ \bar{a} = a. \end{cases} \Rightarrow \bar{x} = x. \end{aligned}$$

Therefore, thanks to Proposition 2.2, the augmented system (5.10) is observable. Using chap 1 Proposition 1.1, we deduce that Model (5.8) is observable and identifiable. Moreover, the state variables w and s as well as the parameter a can be expressed as rational functions of y , \dot{y} and \ddot{y} as follows:

$$\begin{cases} w = y, \\ s = \frac{(\gamma y + \dot{y})b}{(b + \mu)\gamma y + (b + \gamma + \mu)\dot{y} + \ddot{y}}, \\ a = \frac{(b + \mu)\gamma y + (b + \gamma + \mu)\dot{y} + \ddot{y}}{b} = \frac{\gamma y + \dot{y}}{s}. \end{cases}$$

Now, we shall built an observer that will allow to estimate the unmeasured state variable (here it is $s(t)$) as well as the unknown parameter a . To this end, we perform the following change of coordinates:

$$z_1 = w, \quad z_2 = -\gamma w + x_2 a, \quad z_3 = a.$$

One has then

$$\begin{cases} \dot{z}_1 = -\gamma z_1 + z_3 s = -\gamma z_1 + z_3 \frac{z_2 + \gamma z_1}{z_3} = z_2 \\ \dot{z}_2 = -\gamma(-\gamma z_1 + z_3 s) + (b(1-s)w - \mu s)z_3 = -\gamma z_2 + (b(1-s)w - \mu s)z_3 \\ \quad = -\gamma z_2 + (b(1 - \frac{z_2 + \gamma z_1}{z_3})z_1 - \mu \frac{z_2 + \gamma z_1}{z_3})z_3 \\ \quad = -\gamma z_2 + b(z_3 - z_2 - \gamma z_1)z_1 - \mu z_2 - \mu \gamma z_1 \\ \dot{z}_3 = 0 \end{cases}$$

Note that, since $y = z_1$, the dynamics of z_2 can be written as

$$\dot{z}_2 = -\gamma(\mu + by)z_1 - (\mu + \gamma + by)z_2 + byz_3$$

Therefore, the dynamics takes the form

$$\begin{cases} \dot{z}(t) = A(y) z(t), \\ y = C_0 z, \end{cases} \quad (5.11)$$

where

$$A(y) = \begin{bmatrix} 0 & 1 & 0 \\ -\gamma(\mu + by) & -(\mu + \gamma + by) & by \\ 0 & 0 & 0 \end{bmatrix}, \quad C_0 = [1 \ 0 \ 0]$$

For any fixed $y > 0$, the corresponding observability matrix is

$$\mathcal{O}_{(C_0, A)} = \begin{bmatrix} C_0 \\ C_0 A \\ C_0 A^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\gamma(by + \mu) & -by - \gamma - \mu & by \end{bmatrix}$$

that is of full rank if $y \neq 0$. Therefore, by the pole-shifting theorem (see [220, page 61]), it is possible to find a y -dependent gain $K(y)$ such that $sp(A(y) - K(y)C_0) = \{-\lambda_1, -\lambda_2, -\lambda_3\}$, where $sp(M(y))$ denotes the spectrum of $M(y)$ and λ_i are any positive real numbers. This gain $K(y)$ can be computed using for instance Ackermann's formula (see [14] page 382):

$$K(y) = \prod_{i=1}^3 (A(y) + \lambda_i I_3) \cdot \mathcal{O}_{(C_0, A)}^{-1} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

where I_3 is the 3×3 identity matrix.

$$K(y) = \begin{bmatrix} K_1(y) \\ K_2(y) \\ K_3(y) \end{bmatrix} = \begin{bmatrix} -(\gamma + \mu) + \lambda_1 + \lambda_2 + \lambda_3 - by \\ \gamma^2 + \mu^2 + \mu\gamma - (\mu + \gamma)(\lambda_1 + \lambda_2 + \lambda_3) + (\lambda_2 + \lambda_3)\lambda_1 + \lambda_2\lambda_3 \\ + b^2y^2 - (\lambda_1 + \lambda_2 + \lambda_3 - \gamma - 2\mu)by \\ \frac{\lambda_1\lambda_2\lambda_3}{by} \end{bmatrix}$$

The gain $K(y)$ is well defined since $y(t) > 0$ for all $t \geq 0$. An observer is then given by:

$$\dot{\hat{z}}(t) = A(y)\hat{z}(t) - K(y(t))(C_0\hat{z}(t) - y(t)). \quad (5.12)$$

In coordinates:

$$\begin{cases} \dot{\hat{z}}_1 = \hat{z}_2 - (\hat{z}_1 - y)K_1(y) \\ \dot{\hat{z}}_2 = -\mu\gamma\hat{z}_1 - (\mu + \gamma)\hat{z}_2 + b(\hat{z}_3 - \hat{z}_2 - \gamma\hat{z}_1)y - (\hat{z}_1 - y)K_2(y) \\ \dot{\hat{z}}_3 = -(\hat{z}_1 - y)K_3(y) \end{cases} \quad (5.13)$$

The error equation

$$\dot{e}(t) = (A(y) - K(y)C_0)e(t) = M(y)e(t). \quad (5.14)$$

The eigenvalues of the matrix $M(y)$ are $-\lambda_1$, $-\lambda_2$ and $-\lambda_3$. It has been proved in [29] that the error e converges exponentially fast to zero (the proof being quite long is omitted here), which gives the exponential convergence of the observer (5.12).

Finally, let us illustrate this observer on numerical simulations. We have taken the following values of the parameters

β	μ	a	b
0.05	0.04	2	0.01

The initial conditions are $(w(0), s(0)) = (3, 0.3)$ and $(\hat{z}_1(0), \hat{z}_2(0), \hat{z}_3(0)) = (1, 0.1, 0.2)$. The set of eigenvalues of the matrix $M(y)$ have been chosen to be $\{-0.4, -1.4, -2.4\}$. Figure 5.12 shows the convergence of the estimation of the unmeasured proportion of infected snails $s(t)$

$$\hat{s}(t) = \frac{\hat{z}_2(t) + \gamma\hat{z}_1(t)}{\hat{z}_3(t)}$$

and of unknown parameter a

$$\hat{a}(t) = \hat{z}_3(t)$$

delivered by the observer (5.13).

The same simulations have been conducted in Scilab with measurement noise

$$y = x(5) + 0.1 * \text{grand}(1, 1, "nor", 1, 1)$$

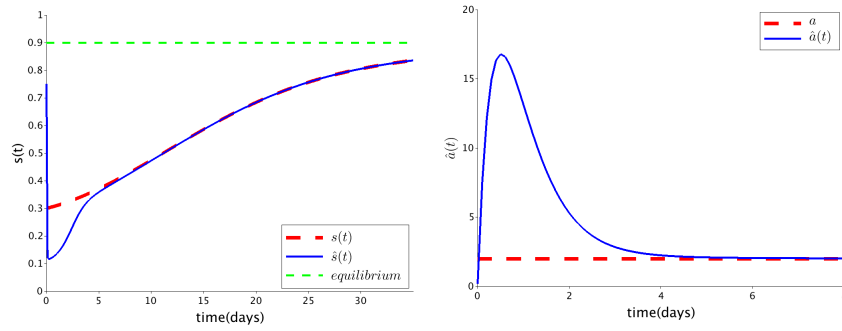


Fig. 5.12 Dynamics of proportion of infected snails $s(t)$ with its estimate $\hat{s}(t)$ (left) and of the estimation $\hat{a}(t)$ of parameter a (right). The green dashed line corresponds to the steady state value of s .

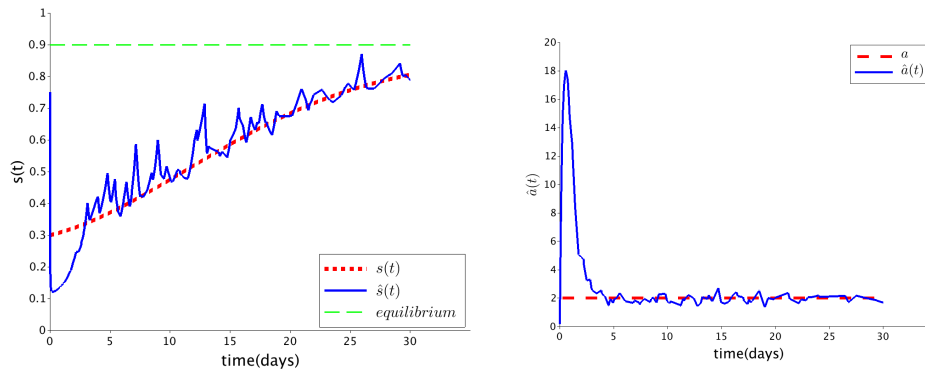


Fig. 5.13 Simulations with measurement noise.

Figure 5.13 shows that for this choice of gains, the estimations are hekkled but follow quite well the unknown state and parameter.

Appendix A

Proofs of some useful lemmas

A.1 Proof of Lemma 4.1

The proof is adapted from [9].

Let us first consider pairs (\bar{A}, \bar{C}) of the canonical form known as Brunovsky's form

$$\bar{A} = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 & -a_n \\ 1 & 0 & \cdots & \cdots & 0 & -a_{n-1} \\ \cdot & \ddots & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & 0 & -a_2 \\ 0 & \cdots & \cdots & 0 & 1 & -a_1 \end{bmatrix}, \quad \bar{C} = [0 \cdots 0 1]$$

where the a_i are any numbers. Their observability matrices are lower triangular:

$$\bar{O} = \begin{bmatrix} & & & & & 1 \\ & & & & & \mathbf{0} \\ & & & & & \cdot \\ & & & & & \cdot \\ & & & & & \star \\ 1 & & & & & \end{bmatrix}$$

therefore invertible. It is easy to see that the characteristic polynomial of the matrix \bar{A} is given by

$$\pi_{\bar{A}}(\xi) = \xi^n + a_1 \xi^{n-1} + \cdots + a_{n-1} \xi + a_n$$

Indeed, if X is a left eigenvector of \bar{A} for an eigenvalue λ (possibly complex), $X\bar{A} = \lambda X$ gives

$$\begin{aligned}
X_2 &= \lambda X_1, \\
X_3 &= \lambda X_2 = \lambda^2 X_1, \\
&\vdots \\
X_n &= \lambda X_{n-1} = \lambda^{n-1} X_1, \\
-a_n X_1 - a_{n-1} X_2 - \cdots - a_1 X_n &= \lambda X_n.
\end{aligned}$$

Thus the line vector X is of the form

$$X = [1 \ \lambda \ \lambda^2 \ \cdots \ \lambda^{n-1}] X_1 \quad \text{with } X_1 \neq 0$$

and λ verifies

$$(\lambda^n + a_1 \lambda^{n-1} + a_2 \lambda^{n-2} + \cdots + a_{n-1} \lambda + a_n) X_1 = 0.$$

Since X_1 is non-zero, we deduce that the eigenvalues are roots of the polynomial

$$\lambda^n + a_1 \lambda^{n-1} + a_2 \lambda^{n-2} + \cdots + a_{n-1} \lambda + a_n = 0$$

which is of degree n and whose coefficient of λ^n is equal to 1.

The characteristic polynomial of the matrix $\bar{A} + \bar{G}\bar{C}$, where \bar{G} is a vector of \mathbb{R}^n with elements denoted \bar{g}_i , is written as follows

$$\pi_{\bar{A} + \bar{G}\bar{C}}(\xi) = \xi^n + (a_1 - \bar{g}_n) \xi^{n-1} + \cdots + (a_{n-1} - \bar{g}_2) \xi + (a_n - \bar{g}_1)$$

Thus, one can arbitrarily choose the n coefficients of this polynomial by choosing the n elements of \bar{G} , and thus freely assign the spectrum of the matrix $\bar{A} + \bar{G}\bar{C}$. For any set $\Lambda = \{\lambda_1, \dots, \lambda_n\}$ of n real or complex numbers two by two conjugates, one has just to identify the coefficients of the polynomial $\pi_{\bar{A} + \bar{G}\bar{C}}$ with those of

$$\prod_{i=1}^n (\xi - \lambda_i) = \xi^n + \sum_{k=1}^n (-1)^k \sigma_k(\Lambda) \xi^{n-k}$$

Thus, we obtain

$$\bar{g}_i = a_{n+1-i} + (-1)^{n-i} \sigma_{n+1-i}(\Lambda), \quad i = 1 \cdots n$$

Let us now show that for any pair (A, C) such that O is full rank, there is an invertible P matrix such that $P^{-1}AP = \bar{A}$ and $CP = \bar{C}$, where the pair (\bar{A}, \bar{C}) is in the Brunovsky's form. Consider the vector

$$L = O^{-1} \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ 1 \end{bmatrix} \Rightarrow \begin{cases} CA^k L = 0, & k = 0 \cdots n-2 \\ CA^{n-1} L = 1 \end{cases}$$

and the matrix consisting of the concatenation of the columns

$$P = [L \ AL \ \cdots \ A^{n-1}L]$$

We have

$$OL = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ 1 \end{bmatrix}, \quad OAL = \begin{bmatrix} CAL \\ CA^2L \\ \vdots \\ CA^{n-1}L \\ CA^nL \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ 1 \\ \star \end{bmatrix}, \quad OA^2L = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \star \\ \star \end{bmatrix}, \dots$$

up to

$$OA^{n-1}L = \begin{bmatrix} 1 \\ \star \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}$$

Thus the OP matrix is of the form

$$OP = \begin{bmatrix} & & & 1 \\ \mathbf{0} & \ddots & & \\ & & \star & \\ 1 & & & \end{bmatrix}$$

which shows that P is indeed an invertible matrix. Finally, the columns of the AP matrix are

$$AP = [AL \ A^2L \ \cdots \ A^{n-1}L \ A^nL]$$

Its first $n - 1$ columns are written as follows

$$[AL \ A^2L \ \cdots \ A^{n-1}L] = P \begin{bmatrix} 0 \ \cdots \\ 1 \ 0 \ \cdots \\ 0 \ \ddots \ \ddots \\ \vdots \quad 1 \ 0 \\ 0 \ \cdots \ 0 \ 1 \end{bmatrix}$$

By Cayley-Hamilton's Theorem, we have $\pi_A(A) = 0$, which allows us to write the last column of AP as

$$A^n L = -a_n L - a_{n-1} A L - \cdots - a_1 A^{n-1} L = P \begin{bmatrix} -a_n \\ -a_{n-1} \\ \vdots \\ -a_1 \end{bmatrix}$$

which shows that we have $PA = \bar{A}P$. Finally, one gets

$$CP = [CL \quad CAL \quad \cdots \quad CA^{n-1}] = [0 \quad \cdots \quad 0 \quad 1] = \bar{C}$$

Then, for a vector \bar{G} such that $S_p(\bar{A} + \bar{G}\bar{C}) = \Lambda$, we have $S_p(P^{-1}(\bar{A} + \bar{G}\bar{C})P) = \Lambda$, where $P^{-1}(\bar{A} + \bar{G}\bar{C})P = A + P\bar{G}C$. We conclude that for the gain vector $G = P\bar{G}$, we have $S_p(A + GC) = \Lambda$.

A.2 Proof of Lemma 4.2

The proof is adapted from [83].

Let X be a left eigenvector of $A+GC$ for the eigenvalue λ_i . By writing $X(A+GC) = \lambda_i X$, we obtain the $n - 1$ inequalities.

$$\begin{aligned} X_1 &= \lambda_i X_2 \\ X_2 &= \lambda_i X_3 \\ &\vdots \\ X_{n-1} &= \lambda_i X_n \end{aligned}$$

Thus X_n is necessarily non-zero and can be taken equal to 1, which gives

$$X = [\lambda_i^{n-1} \quad \lambda_i^{n-2} \quad \cdots \quad \lambda_i \quad 1]$$

We then obtain the n rows of the matrix $V_{\lambda_1, \dots, \lambda_n}$, which defines a matrix of change of basis that diagonalizes the matrix $A + GC$.

Now let's show how to determine the inverse of $V_{\lambda_1, \dots, \lambda_n}$. Let w_{ij} be the coefficients of $V_{\lambda_1, \dots, \lambda_n}^{-1}$. The equality $(V_{\lambda_1, \dots, \lambda_n}^{-1})(V_{\lambda_1, \dots, \lambda_n}) = Id$ gives

$$\sum_{k=1}^n w_{kj} \lambda_i^{nk} = \delta_{ij} := \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j \end{cases} \quad (\text{A.1})$$

For each j in $\{1, \dots, n\}$, let's consider the polynomial

$$P_j(X) = \sum_{k=1}^n w_{kj} X^{n-k} \quad (\text{A.2})$$

The conditions (A.1) amount to write $P_j(\lambda_i) = \delta_{ij}$, i.e. the polynomial P_j has $n - 1$ roots λ_i for $i \neq j$ and $P_j(\lambda_j)$ is equal to 1. So it has the following expression

$$P_j(X) = \prod_{k \neq j} \frac{X - \lambda_k}{\lambda_j - \lambda_k}$$

By identifying its coefficients with those of the expression (A.2), we obtain

$$w_{ij} = (-1)^{i-1} \frac{\sigma_{i-1}(\Lambda \setminus \{\lambda_j\})}{\prod_{k \neq j} \lambda_j - \lambda_k} \quad (\text{A.3})$$

where the σ_k are the symmetric functions defined in (4.6).

Let

$$\varphi(\lambda_1, \dots, \lambda_n) = \lambda_1 + c \|V_{\lambda_1, \dots, \lambda_n}^{-1}\|_{\infty} + \theta$$

The expression (A.3) shows that the norm $\|V_{\lambda_1, \dots, \lambda_n}^{-1}\|_{\infty}$ becomes arbitrarily large when $\lambda_i - \lambda_j$ approaches 0 (for $i \neq j$), which ensures the existence of numbers $\lambda_n < \lambda_{n-1} < \dots < \lambda_1 < 0$ such as $\varphi(\lambda_1, \dots, \lambda_n) > 0$. For $\lambda_i = -\alpha^i$ ($i = 1, \dots, n$), we obtain, for any j

$$\lim_{\alpha \rightarrow +\infty} w_{ij} = \begin{cases} 0 & i < n \\ 1 & i = n \end{cases}$$

and $\|V_{-\alpha, -\alpha^2, \dots, -\alpha^n}^{-1}\|_{\infty}$ thus tends towards 1 when α tends towards $+\infty$, which shows the existence of numbers $\lambda_n < \lambda_{n-1} < \dots < \lambda_1 < 0$ such that $\varphi(\lambda_1, \dots, \lambda_n) < 0$. Finally, by continuity of φ , we deduce the existence of $\lambda_n < \lambda_{n-1} < \dots < \lambda_1 < 0$ such that $\varphi(\lambda_1, \dots, \lambda_n) = 0$.

A.3 Proof of Theorem 4.1

The proof is adapted from [118].

Let P be a symmetric positive definite matrix satisfying $M^T P + PM + Q = 0$ where Q is a symmetric positive definite matrix. We consider the Lyapunov function

$$V(x) = \|x\|_P = x^T P x$$

whose time derivative along solutions of $\dot{x} = Mx$ is

$$\frac{d}{dt} V(x(t)) = -x(t)^T Q x(t) \leq - \underbrace{\frac{\lambda_{\min}(Q)}{\lambda_{\max}(P)}}_{\beta} x(t)^T P x(t)$$

(where λ_{\min} , λ_{\max} denote the smallest and largest real eigenvalues of a symmetric matrix). Then, one has

$$\frac{d}{dt}V(x(t)) \leq -\beta V(x(t))$$

where $\beta > 0$. We deduce that $t \mapsto \|x(t)\|_P$ converges exponentially to 0.

Conversely, when M is Hurwitz, one can consider the symmetric matrix

$$P = \lim_{t \rightarrow +\infty} \int_0^t e^{\tau M^\top} Q e^{\tau M} d\tau$$

(which exists as a sum of terms $t^k e^{\lambda_i t}$ where $\operatorname{Re} \lambda_i < 0$). Then, one has

$$\begin{aligned} M^\top P + PM &= \lim_{t \rightarrow +\infty} \int_0^t M^\top e^{\tau M^\top} Q e^{\tau M} + e^{\tau M^\top} Q e^{\tau M} M d\tau \\ &= \lim_{t \rightarrow +\infty} \int_0^t \frac{d}{d\tau} \left(e^{\tau M^\top} Q e^{\tau M} \right) d\tau \\ &= \lim_{t \rightarrow +\infty} e^{t M^\top} Q e^{t M} - Q \\ &= -Q \end{aligned}$$

Let us show that P is necessarily definite positive. If there exists $u \neq 0$ such that $Pu = 0$, then one has

$$u^\top Pu = 0 = u^\top \left(\int_0^\infty e^{\tau M^\top} Q e^{\tau M} d\tau \right) u = \int_0^\infty \|Q^{1/2} e^{\tau M} u\|^2 d\tau$$

where $Q^{1/2}$ is the positive definite matrix such that $Q^{1/2} Q^{1/2} = Q$. Therefore, u has to be null which shows that P is non singular.

Finally, if there exists another symmetric definite positive matrix \tilde{P} which satisfies $M^\top \tilde{P} + \tilde{P} M = -Q$, then one has

$$M^\top (\tilde{P} - P) + (\tilde{P} - P) M = 0$$

But then

$$e^{t M^\top} (M^\top (\tilde{P} - P) + (\tilde{P} - P) M) e^{t M} = \frac{d}{dt} \left(e^{t M^\top} (\tilde{P} - P) e^{t M} \right) = 0$$

i.e. $t \mapsto e^{t M^\top} (\tilde{P} - P) e^{t M}$ is a constant function. In particular,

$$e^{0 M^\top} (\tilde{P} - P) e^{0 M} = \tilde{P} - P = \lim_{t \rightarrow +\infty} e^{t M^\top} (\tilde{P} - P) e^{t M} = 0$$

which gives $\tilde{P} = P$.

Appendix B

Implementation of the “Boarding School” example

B.1 Derivation of the Fisher Information Matrix

In this example, we consider $x \in \mathbb{R}^n$ with $n = 2$, $y \in \mathbb{R}^m$ with $m = 1$, i.e.,

$$x(t) = \begin{bmatrix} S(t) \\ I(t) \end{bmatrix}, \quad y(t) = I(t),$$

and $\theta \in \mathbb{R}^p$ with $p = 2$, i.e.,

$$\theta = \begin{bmatrix} \beta \\ \gamma \end{bmatrix}.$$

We consider the following model, equivalent to model (3.4) with $k = 1$:

$$\begin{cases} \dot{x} = f(x, \theta) = \begin{bmatrix} \dot{S} \\ \dot{I} \end{bmatrix} = \begin{bmatrix} f_S(S, I, \theta) \\ f_I(S, I, \theta) \end{bmatrix} = \begin{bmatrix} -\beta SI/N \\ \beta SI/N - \gamma I \end{bmatrix}, & x(0) = x_0 = \begin{bmatrix} S_0 \\ I_0 \end{bmatrix}, \\ y = h(x, \theta) = I. \end{cases}$$

We disregard $\Theta = (\theta, x_0)$ since the initial conditions are assumed to be known in this example. The Jacobian of the observation with respect to the parameter θ is:

$$\chi(t, \theta) = \frac{\partial y}{\partial \theta}(t) = \frac{\partial h}{\partial x} \frac{\partial x}{\partial \theta}(t),$$

since in this example,

$$\frac{\partial h}{\partial \theta}(t) = 0.$$

This Jacobian has dimension $m \times p = 1 \times 2$. We have

$$\frac{\partial h}{\partial x} = \begin{bmatrix} \frac{\partial h}{\partial S} & \frac{\partial h}{\partial I} \end{bmatrix} = [0 \ 1]$$

and

$$z = \frac{\partial x}{\partial \theta} = \begin{bmatrix} \frac{\partial S}{\partial \beta} & \frac{\partial S}{\partial \gamma} \\ \frac{\partial I}{\partial \beta} & \frac{\partial I}{\partial \gamma} \end{bmatrix}.$$

This yields

$$\chi = \frac{\partial h}{\partial x} \frac{\partial x}{\partial \theta} = \begin{bmatrix} \frac{\partial I}{\partial \beta} & \frac{\partial I}{\partial \gamma} \end{bmatrix}.$$

Let $\{t_i\}$, $i = 0, 1, 2, \dots, M$, be the sampling times. Fisher’s Information Matrix is defined as:

$$\text{FIM}(\theta, \sigma) = \frac{1}{\sigma^2} \sum_{i=1}^M \chi(t_i, \theta)^\top \chi(t_i, \theta),$$

where σ^2 is defined as the sum of the squared error (SSE) divided with $M - p$ instead of $M - (n + p)$ as in equation (5.4), since the initial conditions are assumed to be known in this example.

Computing Fisher’s Information Matrix

Let

$$A(t) = \frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_S}{\partial S} & \frac{\partial f_S}{\partial I} \\ \frac{\partial f_I}{\partial S} & \frac{\partial f_I}{\partial I} \end{bmatrix} = \begin{bmatrix} -\frac{\beta I}{N} & -\frac{\beta S}{N} \\ \frac{\beta I}{N} & \frac{\beta S}{N} - \gamma \end{bmatrix}$$

and

$$B(t) = \frac{\partial f}{\partial \theta} = \begin{bmatrix} \frac{\partial f_S}{\partial \beta} & \frac{\partial f_S}{\partial \gamma} \\ \frac{\partial f_I}{\partial \beta} & \frac{\partial f_I}{\partial \gamma} \end{bmatrix} = \begin{bmatrix} -\frac{SI}{N} & 0 \\ \frac{SI}{N} & -I \end{bmatrix}.$$

The matrix z can be computed by numerically solving the following system of ODE’s:

$$\begin{cases} \dot{x} = f(x, \theta), & x(0) = x_0, \\ \dot{z} = Az + B, & z(0) = 0_{n \times p} \end{cases}$$

which is a subsystem of (5.7) since the initial conditions are assumed to be known in this example (i.e., we disregard w). In the following code, the entries of x and z are indexed in this way:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad z = \begin{bmatrix} z_3 & z_5 \\ z_4 & z_6 \end{bmatrix},$$

leading to

$$\chi = \begin{bmatrix} z_4 & z_6 \end{bmatrix}.$$

B.2 Numerical implementation

The code has been written with the Scilab language and executed under SCILAB 6.0.0¹. It consists in a function for identifying β and γ and using the `lsqrsolve` function which implements the Levenberg-Marquard algorithm to perform ordinary least squares. We could have chosen the `fminsearch` function which is an implementation of the Nelder-Mead algorithm, but this gives exactly the same results. For solving ODE's, Scilab uses the `lsoda` solver of ODEPACK. It automatically selects between non-stiff predictor-corrector Adams method and stiff Backward Differentiation Formula (BDF) method. It uses non-stiff method initially and dynamically monitors data in order to decide which method to use.

We define the following functions in the Scilab environment:

```
function [kguess_n,SSE]=identifKMK(OBS,T,kguess,N)
// kguess_n = [BETA;GAMMA]
t0=T(1);m=length(T);
[x,SSE]=lsqrsolve(kguess,errorKmcK,m,[1.d-8,1.d-8,1.d-5,1d9,0,100]);
nbparam=length(kguess)
kguess_n=x;
x0=[N-OBS(1);OBS(1)];
sol=ode(x0,t0,T,list(KmcK,x(1),x(2)));
kguess_n=kguess_n(:);
SSE=sum(SSE.^2)
xset("window",1)
sol1=ode(x0,t0,T(1):0.01:T($),list(KmcK,x(1),x(2)));
clf
plot((T(1):0.01:T($))',sol1(2,:))
plot(T',OBS,'ro')
endfunction

function y=errorKmcK(k,m)
x0=[N-OBS(1);OBS(1)];x0=x0(:);
BETA=k(1);GAMMA=k(2);
sol=ode(x0,t0,T,list(KmcK,BETA,GAMMA));
predic=sol(2,:);
predic=predic(:);
OBS=OBS(:);
y=OBS-predic;
endfunction

function xdot=KmcK(t,x,BETA,GAMMA,N)
xdot=[-BETA/N*x(2),0;BETA/N*x(2),-GAMMA]*x
endfunction
```

¹ <https://www.scilab.org/>

Then, the Scilab session goes like this

```
[commandchars=\\{\}]
--> load('databoarding')
ans =
T
--> OBS=dataBSFlu

OBS =

column 1 to 8

1.   6.   26.   73.   222.   293.   258.   237.

column 9 to 14

191.   124.   68.   26.   10.   3.

--> M=length(OBS);
--> T=0:M-1;
--> N=763;
--> beta0=2;gamma0=0.5;param=[beta0,gamma0];
--> p=length(param);
--> [param,SSE]=identifKMK(OBS,T,param,N)

SSE =

4892.6472

param =

1.9605032
0.4751562

--> sigma2=SSE/(M-p)

sigma2 =

407.72060
```

```
--> BETA=param(1);GAMMA=param(2);
```

We then compute confidence intervals using the formulas (5.2), (5.3), (5.4), (5.5), (5.6).

```
[commandchars=\\{\}]
function FIM=fimKmcK(x0,T,BETA,GAMMA,sigma2)
// Compute the sensitivity matrix
x0=x0(:);t0=T(1);
X0=[x0;0;0;0;0]
sol=ode(X0,t0,T,list(JKmcK,BETA,GAMMA));
M=sol([4,6],:);
FIM=M*M'./sigma2;
endfunction

function xdot=JKmcK(t,x,BETA,GAMMA)
xdot(1)=-BETA*x(1)*x(2)/N
xdot(2)=BETA*x(1)*x(2)/N-GAMMA*x(2)
xdot(3)=-BETA*x(2)*x(3)/N-BETA*x(1)*x(4)/N-x(1)*x(2)/N
xdot(4)=BETA*x(2)*x(3)/N+(BETA*x(1)/N-GAMMA)*x(4)+x(1)*x(2)/N
xdot(5)=-BETA*x(2)*x(5)/N-BETA*x(1)*x(6)/N
xdot(6)=BETA*x(2)*x(5)/N+(BETA*x(1)/N-GAMMA)*x(6)-x(2)
endfunction
```

Then the Scilab session is

```
[commandchars=\\{\}]

--> x0=[N-OBS(1); OBS(1)];

--> FIM=fimKmcK(x0,T,BETA,GAMMA)
FIM =

974.5073   -523.73985
-523.73985   3132.2047

--> cond(FIM)
ans =

3.8082403

--> CovMAT=inv(FIM)
CovMAT =
```

```
0.0011275  0.0001885  
0.0001885  0.0003508
```

```
--> t=cdf("T",M-p,0.975,0.025)
```

```
t =
```

```
2.1788128
```

```
--> confBETA=t*sqrt(CovMAT(1,1))
```

```
confBETA =
```

```
0.0731602
```

```
--> confGAMMA=t*sqrt(CovMAT(2,2))
```

```
confGAMMA =
```

```
0.0408077
```

Appendix C

Implementation of the “Plague in Bombay” example

C.1 Derivation of the Fisher Information Matrix

In this example, we consider $x \in \mathbb{R}^n$ with $n = 2$, $y \in \mathbb{R}^m$ with $m = 1$, i.e.,

$$x(t) = \begin{bmatrix} S(t) \\ I(t) \end{bmatrix}, \quad y(t) = I(t),$$

and $\theta \in \mathbb{R}^p$ with $p = 2$, i.e.,

$$\tilde{\beta} = \frac{\beta}{N}, \quad \theta = \begin{bmatrix} \tilde{\beta} \\ \gamma \end{bmatrix}.$$

We consider the following model, equivalent to model (3.4) with $k = \gamma$:

$$\begin{cases} \dot{x} = f(x, \theta) = \begin{bmatrix} \dot{S} \\ \dot{I} \end{bmatrix} = \begin{bmatrix} f_S(S, I, \theta) \\ f_I(S, I, \theta) \end{bmatrix} = \begin{bmatrix} -\tilde{\beta}SI \\ \tilde{\beta}SI - \gamma I \end{bmatrix}, & x(0) = x_0 = \begin{bmatrix} S_0 \\ I_0 \end{bmatrix}, \\ y = h(x, \theta) = \gamma I. \end{cases}$$

We consider $\Theta = (\theta, x_0)$ since the initial conditions are assumed to be unknown in this example. The Jacobian of the observation with respect to the parameter Θ is:

$$\chi(t, \Theta) = \frac{\partial y}{\partial \Theta}(t) = \frac{\partial h}{\partial x} \frac{\partial x}{\partial \Theta}(t) + \frac{\partial h}{\partial \Theta}(t),$$

This Jacobian has dimension $m \times (p + n) = 1 \times 4$. We have

$$\begin{aligned} \frac{\partial h}{\partial x} &= \begin{bmatrix} \frac{\partial h}{\partial S} & \frac{\partial h}{\partial I} \end{bmatrix} = [0 \ \gamma], \\ \frac{\partial h}{\partial \Theta} &= \begin{bmatrix} \frac{\partial h}{\partial \tilde{\beta}} & \frac{\partial h}{\partial \gamma} & \frac{\partial h}{\partial S_0} & \frac{\partial h}{\partial I_0} \end{bmatrix} = [0 \ I \ 0 \ 0], \end{aligned}$$

and

$$\frac{\partial x}{\partial \Theta} = \begin{bmatrix} \frac{\partial S}{\partial \tilde{\beta}} & \frac{\partial S}{\partial \gamma} & \frac{\partial S}{\partial S_0} & \frac{\partial S}{\partial I_0} \\ \frac{\partial I}{\partial \tilde{\beta}} & \frac{\partial I}{\partial \gamma} & \frac{\partial I}{\partial S_0} & \frac{\partial I}{\partial I_0} \end{bmatrix}.$$

This yields

$$\frac{\partial h}{\partial x} \frac{\partial x}{\partial \tilde{\theta}} = \gamma \begin{bmatrix} \frac{\partial I}{\partial \tilde{\beta}} & \frac{\partial I}{\partial \gamma} & \frac{\partial I}{\partial S_0} & \frac{\partial I}{\partial I_0} \end{bmatrix}.$$

Therefore,

$$\chi = \gamma \begin{bmatrix} \frac{\partial I}{\partial \tilde{\beta}} & \frac{\partial I}{\partial \gamma} + I & \frac{\partial I}{\partial S_0} & \frac{\partial I}{\partial I_0} \end{bmatrix}.$$

Let $\{t_i\}$, $i = 0, 1, 2, \dots, M$, be the sampling times. Fisher’s Information Matrix is defined as:

$$\text{FIM}(\Theta, \sigma) = \frac{1}{\sigma^2} \sum_{i=1}^M \chi(t_i, \Theta)^\top \chi(t_i, \Theta),$$

where σ^2 is defined as in equation (5.4), since the initial conditions are unknown in this example.

Computing Fisher’s Information Matrix

We make the following decomposition:

$$\frac{\partial x}{\partial \Theta} = \begin{bmatrix} \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial x_0} \end{bmatrix} = [z \ w],$$

with

$$z = \frac{\partial x}{\partial \theta} = \begin{bmatrix} \frac{\partial S}{\partial \tilde{\beta}} & \frac{\partial S}{\partial \gamma} \\ \frac{\partial I}{\partial \tilde{\beta}} & \frac{\partial I}{\partial \gamma} \end{bmatrix} \quad \text{and} \quad w = \frac{\partial x}{\partial x_0} = \begin{bmatrix} \frac{\partial S}{\partial S_0} & \frac{\partial S}{\partial I_0} \\ \frac{\partial I}{\partial S_0} & \frac{\partial I}{\partial I_0} \end{bmatrix}.$$

Letting

$$A(t) = \frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_S}{\partial S} & \frac{\partial f_S}{\partial I} \\ \frac{\partial f_I}{\partial S} & \frac{\partial f_I}{\partial I} \end{bmatrix} = \begin{bmatrix} -\tilde{\beta}I & -\tilde{\beta}S \\ \tilde{\beta}I & \tilde{\beta}S - \gamma \end{bmatrix},$$

and

$$B(t) = \frac{\partial f}{\partial \theta} = \begin{bmatrix} \frac{\partial f_S}{\partial \tilde{\beta}} & \frac{\partial f_S}{\partial \gamma} \\ \frac{\partial f_I}{\partial \tilde{\beta}} & \frac{\partial f_I}{\partial \gamma} \end{bmatrix} = \begin{bmatrix} -SI & 0 \\ SI & -I \end{bmatrix}.$$

The FIM can be computed via numerically solving the following system of ODE’s:

$$\begin{cases} \dot{x} = f(x, \theta), & x(0) = x_0, \\ \dot{z} = Az + B, & z(0) = 0_{n \times p}, \\ \dot{w} = Aw, & w(0) = \text{Id}_{n \times n}, \end{cases}$$

which repeats equation (5.7). In the following code, the entries of x , z , and w are indexed in this way:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad z = \begin{pmatrix} z_3 & z_5 \\ z_4 & z_6 \end{pmatrix}, \quad w = \begin{bmatrix} w_7 & w_9 \\ w_8 & w_{10} \end{bmatrix},$$

leading to

$$\chi = \gamma \begin{bmatrix} z_4 & z_6 + x_2 & w_8 & w_{10} \end{bmatrix}.$$

C.2 Numerical implementation

Although the code is very similar the one provided in the previous example (Appendix B), we provide it for convenience, as it required a number of small changes.

We define the following functions in the Scilab environment:

```
function [kguess_n ,SSE]=identifKMK(OBS,T,kguess)
t0=T(1);m=length(T);
[x,SSE]=lsqrsolve(kguess,errorKmcK,m,[1.d-8,1.d-8,1.d-5,1d9,0,100]);
nbparam=length(kguess);
kguess_n=x;
x0=[kguess_n(3);kguess_n(4)];
sol=ode(x0,t0,T,list(KmcK,x(1),x(2)));
kguess_n=kguess_n(:);
SSE=sum(SSE.^2);
xset("window",1);
sol1=ode(x0,t0,T(1):0.01:T($),list(KmcK ,x(1),x(2)));
clf;
plot((T(1):0.01:T($))',kguess_n(2)*sol1(2,:),'k');
plot(T',OBS ,'ko')
legend(["$\Large \gamma I(t)$", "$\Large \mbox{Data}$"])
ylabel("$\Large \mbox{number of deaths per week}$","fontsize",3);
xlabel("$\Large \mbox{time }t\mbox{ (in week)}$", "fontsize",3);
endfunction

function y=errorKmcK(k,m)
x0=[k(3);k(4)];
x0=x0(:);
B=k(1);
GAMMA=k(2);
sol=ode(x0,t0,T,list(KmcK,B,GAMMA));
```

```

predic=GAMMA*sol(2,:);
predic=predic(:);
OBS=OBS(:);
y=OBS-predic;
endfunction

function xdot=KmcK(t,x,B,GAMMA)
xdot=[-B*x(2),0; B*x(2),-GAMMA]*x
endfunction

```

Then, the Scilab session goes like this

```

[commandchars=\\{\}]
--> load('databombay')
ans =
T
--> OBS=dataBSFlu

OBS =

column 1 to 9

8.   10.   12.   16.   24.   48.   51.   92.   124.

column 10 to 16

178.   280.   387.   442.   644.   779.   702.

column 17 to 23

695.   870.   925.   802.   578.   404.   296.

column 24 to 31

162.   106.   64.   46.   35.   27.   28.   24.

--> M=length(OBS);

--> T=0:M-1;

--> S0=15000;I0=7;gamma0=0.6;b0=8e-5;//Initial guesses

--> param=[b0,gamma0,S0,I0];//With b=beta/N

--> p=length(param);

```

```

--> [param,SSE]=identifKMK(OBS,T,param,N)

SSE =

106336.49

param =

0.0000855
3.7161743
48113.13
1.4213612

--> n=2;sigma2=SSE/(M-(n+p))

sigma2 =

4253.4597

--> B=param(1);GAMMA=param(2);

```

We then compute confidence intervals using the formulas (5.2), (5.3), (5.4), (5.5), (5.6).

```

[commandchars=\\{\}]
function FIM=fimKmcK(x0,T,B,GAMMA,sigma2)
// Compute the sensitivity matrix
x0=x0(:);t0=T(1);
X0=[x0;zeros(4,1);eye(2,2)(:)];
sol=ode(X0,t0,T,list(JKmcK,B,GAMMA));
M=GAMMA*sol([4,6,8,10],:);
M(2,:)=M(2,)+GAMMA*sol(2,:);
FIM=M*M'./sigma2;
endfunction

function xdot=JKmcK(t,x,B,GAMMA)
xdot (1)=-B*x(1)*x(2);
xdot (2)=B*x(1)*x(2)-GAMMA*x(2);

xdot (3)=-B*x(2)*x(3)-B*x(1)*x(4)-x(1)*x(2);
xdot (4)=B*x(2)*x(3)+(B*x(1)-GAMMA)*x(4)+x(1)*x(2);
xdot (5)=-B*x(2)*x(5)-B*x(1)*x(6);
xdot (6)=B*x(2)*x(5)+(B*x(1)-GAMMA)*x(6)-x(2);

```



```

xdot (7)=-B*x(2)*x(7)-B*x(1)*x(8);
xdot (8)=B*x(2)*x(7)+(B*x(1)-GAMMA)*x(8);
xdot (9)=-B*x(2)*x(9)-B*x(1)*x(10);
xdot (10)=B*x(2)*x(9)+(B*x(1)-GAMMA)*x(10);
endfunction

```

Then the Scilab session is

```

[commandchars=\\{\}]

--> x0=[param(3);param(4)];

--> FIM=fimKmcK(x0,T,B,GAMMA)
FIM =

1.100D+14  -2.117D+09  196924.7  92108921.
-2.117D+09  40885.251  -3.7835134  -1826.2845
196924.7  -3.7835134  0.0003533  0.1603199
92108921.  -1826.2845  0.1603199  118.22324

--> cond(FIM)
ans =

9.141D+24

--> CovMAT=inv(FIM)
Warning: Matrix is close to singular or badly scaled.
CovMAT =

0.00000006  0.0093978  -220.96547  -0.0128142
0.0093978  150.37071  -3535242.5  -204.997
-220.96547  -3535242.5  8.313D+10  4820745.
-0.0128142  -204.997  4820745.  279.63056

--> t=cdf("T",M-(n+p),0.975,0.025)
t =

2.0595386

--> confB=t*sqrt(CovMAT(1,1))
confB =

0.0015784

--> confGAMMA=t*sqrt(CovMAT(2,2))
confGAMMA =

```

25.255243

```
--> confS0=t*sqrt(CovMAT(3,3))
```

```
confS0 =
```

593794.26

```
--> confI0=t*sqrt(CovMAT(4,4))
```

```
confI0 =
```

34.439929

Appendix D

Generalized Least Squares

With Ordinary Least Squares, constant variance has been assumed which may be not appropriate for some data. A relative error, i.e., when the error is assumed to be proportional to the size of the measurement, is an assumption that might be reasonable when counting individuals in a population.

In this case we assume that the observation are [21, 53, 23, 40]:

$$Y_i = y(t_i, \Theta) + y(t_i, \Theta)^\rho \mathcal{E}_i,$$

with $\Theta = (\theta, x_0)$.

The criterion to be minimized is

$$\mathcal{J}(\Theta) = \sum_{i=1}^N w_i [Y_i - y(t_i, \Theta)]^2.$$

The values of the weights (w_i) depend on the value of the model and are not known. The process is carried with an iterated re-weighted least squares:

1. Estimate $\hat{\Theta}_0$ with an OLS step ($\rho = 0$): $w_i = 1$ for all $i = 1, \dots, N$
2. Set $\rho = 1$ (for instance) and $w_i = 1/[y(t_i, \hat{\Theta}_0)]^{2\rho}$ for all $i = 1, \dots, N$
3. Form $\mathcal{J}(\Theta)$ with these w_i and estimate

$$\hat{\Theta}_1 = \arg \min_{\Theta} \mathcal{J}(\Theta)$$

4. Continue the procedure till the estimates $\hat{\Theta}_k$ and $\hat{\Theta}_{k+1}$ are sufficiently close to each other, to obtain $\hat{\Theta}_{\text{GLS}}$.

With $\hat{\Theta}_{\text{GLS}}$, as in the OLS case, we can obtain the covariance matrix (Σ) of the estimated parameters, approximated as the inverse Fisher Information Matrix with weights:

$$\text{FIM}(\hat{\Theta}_{\text{GLS}}, \hat{\sigma}_{\text{GLS}}^2) = \frac{1}{\hat{\sigma}_{\text{GLS}}^2} \sum_{i=1}^N \frac{1}{y(t_i, \hat{\Theta}_{\text{GLS}})^{2\rho}} \frac{\partial y}{\partial \Theta}(t_i, \hat{\Theta}_{\text{GLS}})^\top \frac{\partial y}{\partial \Theta}(t_i, \hat{\Theta}_{\text{GLS}}),$$

with

$$\hat{\sigma}_{\text{GLS}}^2 = \frac{1}{N-p} \sum_{i=1}^N \frac{1}{y(t_i, \hat{\Theta}_{\text{GLS}})^{2\rho}} [Y_i - y(t_i, \hat{\Theta}_{\text{GLS}})]^2.$$

We then obtain

$$\hat{\Sigma}_{\text{GLS}} = [\text{FIM}(\hat{\Theta}_{\text{GLS}}, \hat{\sigma}_{\text{GLS}}^2)]^{-1},$$

The square roots of the diagonal elements of the approximation of the covariance matrix $\hat{\Sigma}_{\text{GLS}}$ give the standard errors.

References

1. A. ABDELHEDI, D. BOUTAT, L. SBITA, R. TAMI, AND D.-Y. LIU, *Observer design for a class of nonlinear piecewise systems. Application to an epidemic model with treatment*, Math. Biosci., 271 (2016), pp. 128–135.
2. C. ABOKY, G. SALLET, AND J.C. VIVALDA, *Observers for Lipschitz nonlinear systems* Int. J. Contr., 75 (3) (2002), pp. 204–212.
3. D. AEYELS, *Generic observability of differentiable systems*, SIAM J. Control Optim., 19 (1981), pp. 595–603.
4. D. AEYELS, *On the number of samples necessary to achieve observability*, Systems Control Lett., 1 (1981/82), pp. 92–94.
5. H. AKAIKE, *A new look at the statistical model identification*, IEEE Trans. Autom. Control, 19 (1976), pp. 716–723.
6. S. ALONSO-QUESADA, M. DE LA SEN, R. P. AGARWAL, AND A. IBEAS, *An observer-based vaccination control law for an SEIR epidemic model based on feedback linearization techniques for nonlinear systems*, Adv. Difference Equ., (2012), pp. 2012:161, 32.
7. R. M. ANDERSON AND R. M. MAY, *Infectious Diseases of Humans. Dynamics and Control*, Oxford science publications, 1991.
8. R. M. ANDERSON, R. M. MAY, AND S. GUPTA, *Non-linear phenomena in host-parasite interactions.*, Parasitology, 99 Suppl (1989), pp. S59–79.
9. B. ANDRÉA-NOVELL AND M. DE LARA, *Control Theory for Engineers: A Primer.*, Springer, Berlin, Heidelberg (2013).
10. V. ANDRIEU, G. BESANÇON AND U. SERRES, *Observability necessary conditions for the existence of observers.* IEEE Conference on Decision and Control, Florence, Italy (2013).
11. V. ANDRIEU AND L. PRALY, *On the Existence of a Kazantzis–Kravaris/Luenberger Observer*, SIAM Journal on Control and Optimization 45(2) (2006), pp.432–456.
12. M. ANGUELOVA, *Observability and identifiability of nonlinear systems with applications in biology*, PhD thesis, Chalmers University of Technology and Göteborg University, 2007.
13. M. ANGUELOVA, J. KARLSSON, AND M. JIRSTRAND, *Minimal output sets for identifiability.*, Math Biosci, 239 (2012), pp. 139–153.
14. P. J. ANTSAKLIS AND A. N. MICHEL *State Feedback and State Observers*, pages 351–410. Birkhäuser Boston, Boston, MA, 2007.
15. M. ASHYRALIYEV, Y. FOMEKONG-NANFACK, J. A. KAANDORP, AND J. G. BLOM, *Systems biology: parameter estimation for biochemical models.*, FEBS J, 276 (2009), pp. 886–902.
16. 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 (2001), pp. 55–65.
17. N. BACAER, *The model of Kermack and McKendrick for the plague epidemic in Bombay and the type reproduction number with seasonality*, J. Math. Biol., 64 (2012), pp. 403–422.
18. E. BALSACANTO AND J. R. BANGA, *Amigo, a toolbox for advanced model identification in systems biology using global optimization.*, Bioinformatics, 27 (2011), pp. 2311–2313.

19. E. BALSÁ-CANTO, D. HENRIQUES, A. GÁBOR, AND J. R. BANGA, *Amigo2, a toolbox for dynamic modeling, optimization and control in systems biology.*, Bioinformatics, 32 (2016), pp. 3357–3359.
20. H. T. BANKS, A. CINTRÓN-ARIAS, AND F. KAPPEL, *Parameter selection methods in inverse problem formulation*, in Mathematical modeling and validation in physiology, vol. 2064 of Lecture Notes in Math., Springer, Heidelberg, 2013, pp. 43–73.
21. H. T. BANKS, M. DAVIDIAN, J. SAMUELS, AND K. SUTTON, *An inverse problem statistical methodology summary*, in Mathematical and Statistical Estimation Approaches in Epidemiology, G. Chowell, ed., Springer, 2009, pp. 249–302.
22. H. T. BANKS, S. L. ERNSTBERGER, AND S. L. GROVE, *Standard errors and confidence intervals in inverse problems: sensitivity and associated pitfalls*, J. Inverse Ill-Posed Probl., 15 (2007), pp. 1–18.
23. H. T. BANKS, S. HU, AND W. C. THOMPSON, *Modeling and inverse problems in the presence of uncertainty*, Monographs and Research Notes in Mathematics, CRC Press, Boca Raton, FL, 2014.
24. H. T. BANKS AND H. T. TRAN, *Mathematical and experimental modeling of physical and biological processes*, Textbooks in Mathematics, CRC Press, Boca Raton, FL, 2009. With 1 CD-ROM (Windows, Macintosh and UNIX).
25. A. D. BARBOUR, *Macdonald's model and the transmission of bilharzia.*, Trans R Soc Trop Med Hyg, 72 (1978), pp. 6–15.
26. R. BELLMAN, AND K. J. ÅSTRÖM, *On structural identifiability*, Math. Biosci., 7 (1970), pp. 329–339.
27. G. BELLU, M. P. SACCOMANI, S. AUDOLY, AND L. D'ANGIO, *Daisy: a new software tool to test global identifiability of biological and physiological systems.*, Comput Methods Programs Biomed, 88 (2007), pp. 52–61.
28. D. BICHARA, N. COZIC, AND A. IGGIDR, *On the estimation of sequestered infected erythrocytes in plasmodium falciparum malaria patients*, Math. Biosci. Eng., 11 (2014), pp. 741–759.
29. D. BICHARA, A. GUIRO, A. IGGIDR, AND D. NGOM. State and parameter estimation for a class of schistosomiasis models. *Mathematical Biosciences*, 315, 2019.
30. D. BICHARA, A. IGGIDR, M. OUMOUN, A. RAPAPORT, AND G. SALLET. Identifiability and Observability via decoupled variables: Application to a malaria intra-host model. In *Proc. the 22nd Triennial IFAC World Congress*, Yokohama, Japan, 2023.
31. N. A. BODUNOV, S. A. KOLBINA, AND S. Y. PILYUGIN, *Locally parameter identifiable systems are generic*, Vestnik St. Petersburg Univ. Math., 45 (2012), pp. 61–64.
32. B. M. BOLKER, *Ecological models and data in R*. Princeton University Press (2008).
33. ———, *Prevalence of locally parameter identifiable systems*, Vestnik St. Petersburg Univ. Math., 48 (2015), pp. 204–208.
34. F. BOULIER, F. OLLIVIER, D. LAZARD, AND M. PETITOT, *Computing representations for radicals of finitely generated differential ideals*, Appl. Algebra Engrg. Comm. Comput., 20 (2009), pp. 73–121.
35. F. BRAUER AND C. CASTILLO-CHÁVEZ, *Mathematical models in population biology and epidemiology*, vol. 40 of Texts in Applied Mathematics, Springer-Verlag, New York, 2001.
36. F. BRAUER, C. CASTILLO-CHAVEZ, AND Z. FENG, *Mathematical models in epidemiology*, vol. 69 of Texts in Applied Mathematics, Springer, New York, 2019. With a foreword by Simon Levin.
37. F. BRAUER, J. WU, AND P. VAN DEN DRIESSCHE, eds., *Mathematical Epidemiology*, no. 1945 in Lectures Notes in Math., Springer-Verlag, 2008.
38. R. BRUN, P. REICHERT, AND KUNSCH, *Practical identifiability analysis of large environmental simulation models.*, Water Resources Research, 37 (2001), pp. 1015–1030.
39. B. N. CANTÓ, C. COLL, AND E. SÁNCHEZ, *Estimation of parameters in a structured SIR model*, Adv. Difference Equ., (2017), pp. Paper No. 33, 13.
40. A. CAPALDI, S. BEHREND, B. BERMAN, J. SMITH, J. WRIGHT, AND A. L. LLOYD, *Parameter estimation and uncertainty quantification for an epidemic model.*, Math Biosci Eng, 9 (2012), pp. 553–576.

41. M. A. CAPISTRAN, M. A. MORELES, AND B. LARA, *Parameter estimation of some epidemic models. The case of recurrent epidemics caused by respiratory syncytial virus*, Bull. Math. Biol., 71 (2009), pp. 1890–1901.
42. J. L. Casti. Recent developments and future perspectives in nonlinear system theory. *SIAM Review*, 24(3):301–331, 1982.
43. F. CELLE, J.-P. GAUTHIER, D. KAZAKOS, AND G. SALLET, *Synthesis of nonlinear observers: a harmonic-analysis approach*, Math. Systems Theory, 22 (1989), pp. 291–322.
44. C. D. S. CENTER, *Influenza in a boarding school*, BMJ, 1 (1978), p. 587.
45. W. CHEN AND M. SAIF, *Unknown input observer design for a class of nonlinear systems: an LMI approach*, American Control Conference, Minneapolis, MN, USA (2006).
46. O. CHİŞ, J. R. BANGA, AND E. BALSACANTO, *Genssi: a software toolbox for structural identifiability analysis of biological models.*, Bioinformatics, 27 (2011), pp. 2610–2611.
47. O.-T. CHIS, J. R. BANGA, AND E. BALSACANTO, *Structural identifiability of systems biology models: a critical comparison of methods.*, PLoS One, 6 (2011), p. e27755.
48. O.-T. CHIS, A. F. VILLAVARDE, J. R. BANGA, AND E. BALSACANTO, *On the relationship between sloppiness and identifiability*, Mathematical biosciences, 282, 147–161, 2016.
49. O.-T. CHIS, A. F. VILLAVARDE, J. R. BANGA, AND E. BALSACANTO, *On the relationship between sloppiness and identifiability*, Math. Biosci., 282 (2016), pp. 147–161.
50. I.-C. CHOU AND E. O. VOIT, *Recent developments in parameter estimation and structure identification of biochemical and genomic systems*, Math. Biosci., 219 (2009), pp. 57–83.
51. A. CINTRÓN-ARIAS, H. T. BANKS, A. CAPALDI, AND A. L. LLOYD, *A sensitivity matrix based methodology for inverse problem formulation*, J. Inverse Ill-Posed Probl., 17 (2009), pp. 545–564.
52. A. CINTRÓN-ARIAS, C. CASTILLO-CHÁVEZ, L. M. A. BETTENCOURT, A. LLOYD, AND H. T. BANKS, *The estimation of the effective reproductive number from disease outbreak data*, Tech. Rep. CRSC-TR08, Center for Research in Scientific Computation, 2008.
53. A. CINTRÓN-ARIAS, C. CASTILLO-CHÁVEZ, L. M. A. BETTENCOURT, A. L. LLOYD, AND H. T. BANKS, *The estimation of the effective reproductive number from disease outbreak data*, Math. Biosci. Eng., 6 (2009), pp. 261–282.
54. C. COBELLI AND J. J. R. DiSTEFANO, *Parameter and structural identifiability concepts and ambiguities: a critical review and analysis.*, Am J Physiol, 239 (1980), pp. R7–24.
55. C. COBELLI, A. LEPSCHY, AND G. ROMANIN-JACUR, *Identifiability of compartmental systems and related structural properties*, Math. Biosci., 44 (1979), pp. 1–18.
56. C. COBELLI AND G. ROMANIN-JACUR, *Controllability, Observability and Structural Identifiability of Multi Input and Multi Output Biological Compartmental Systems*, IEEE Trans Biomed Eng, 23 (1976), pp. 93–100.
57. ———, *On the structural identifiability of biological compartmental systems in a general input-output configuration*, Math. Biosci., 30 (1976), pp. 139–151.
58. D.J. COLE, *Parameter redundancy and identifiability*. Chapman and Hall/CRC, 2020.
59. COMMISSION, *Xxii-epidemiological observations made by the commission in bombay city*, J. Hyg (Lond), 7 (1907), pp. 724–798.
60. D. DALEY AND J. GANI, *Epidemic modelling : An introduction*, Cambridge University Press, 1999.
61. M. DAVIDIAN AND D. GILTINAN, *Nonlinear Models for Repeated Measurement Data.*, Chapman and Hall, Boca Raton, 1995.
62. M. C. DE JONG, O. DIEKMANN, AND H. HEESTERBEEK, *How does transmission of infection depend on population size ?*, in How does transmission of infection depend on population size? By: De Jong, Mart C. M.; Diekmann, Odo; Heesterbeek, Hans Conference: NATO Advanced Research Workshop on Epidemic Models Location: Cambridge, England, Publications of the Newton Institute, 1993, pp. 84–94.
63. M. DE LA SEN, A. IBEAS, AND S. ALONSO-QUESADA, *Observer-based vaccination strategy for a true mass action SEIR epidemic model with potential estimation of all the populations*, Discrete Dyn. Nat. Soc., (2011), pp. Art. ID 743067, 19.
64. P. DE LEENHEER AND H. L. SMITH, *Virus dynamics: A global analysis.*, SIAM J. Appl. Math., 63 (2003), pp. 1313–1327.

65. G. DE VRIES, T. HILLEN, M. LEWIS, J. MÜLLER, AND B. SCHÖNFISCH, *A course in mathematical biology*, vol. 12 of Mathematical Modeling and Computation, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006. Quantitative modeling with mathematical and computational methods.
66. J. W. DEMMEL, *On condition numbers and the distance to the nearest ill-posed problem*, Numer. Math., 51 (1987), pp. 251–289.
67. L. DENIS-VIDAL AND G. JOLY-BLANCHARD, *A necessary condition, a sufficient condition for structural identifiability*, in proceedings 36th CDC, 1997, pp. 1289–1290.
68. L. DENIS-VIDAL AND G. JOLY-BLANCHARD, *An easy to check criterion for (un)identifiability of uncontrolled systems and its applications*, IEEE Trans. Automat. Control, 45 (2000), pp. 768–771.
69. L. DENIS-VIDAL, G. JOLY-BLANCHARD, AND C. NOIRET, *Some effective approaches to check the identifiability of uncontrolled nonlinear systems*, Math. Comput. Simulation, 57 (2001), pp. 35–44.
70. M. DIABY, A. IGGIDR, AND M. SY, *Observer design for a schistosomiasis model*, Math. Biosci., 269 (2015), pp. 17–29.
71. S. DIOP, *Elimination in control theory*, Math. Control Signals Systems, 4 (1991), pp. 17–32.
72. ———, *Differential-algebraic decision methods and some applications to system theory*, Theoret. Comput. Sci., 98 (1992), pp. 137–161. Second Workshop on Algebraic and Computer-theoretic Aspects of Formal Power Series (Paris, 1990).
73. S. DIOP AND M. FLIESS, *Nonlinear observability, identifiability, and persistent trajectories*, in proceedings 36th IEEE-CDC, 1991, pp. 714–719.
74. S. DIOP AND M. FLIESS, *On nonlinear observability*, in Proceedings EEC91, vol. 1, Hermès, 1991, pp. 154–211.
75. S. Diop and Y. Wang. Equivalence between algebraic observability and local generic observability. In *Proceedings of 32nd IEEE Conference on Decision and Control*. IEEE, 1993.
76. J. DISTEFANO AND C. COBELLI, *On parameter and structural identifiability : nonunique observability/reconstructibility for identifiable systems, other ambiguities and new definitions*, IEEE Trans. Autom. Control, 25 (1980), pp. 830–833.
77. J. J. R. DISTEFANO, *Dynamic systems Biology modeling and application*, Academic Press, 2013.
78. S. DOBRE, T. BASTOGNE, C. PROFETA, M. BARBERI-HEYOB, AND A. RICHARD, *Limits of variance-based sensitivity analysis for non-identifiability testing in high dimensional dynamic models*, Automatica J. IFAC, 48 (2012), pp. 2740–2749.
79. A. DUCROT, P. MAGAL, T. NGUYEN, AND G. F. WEBB, *Identifying the number of unreported cases in SIR epidemic models*, Math. Med. Biol., 37 (2020), pp. 243–261.
80. M. C. EISENBERG, S. L. ROBERTSON, AND J. H. TIEN, *Identifiability and estimation of multiple transmission pathways in cholera and waterborne disease.*, J Theor Biol, 324 (2013), pp. 84–102.
81. N. D. EVANS, M. J. CHAPMAN, M. J. CHAPPELL, AND K. R. GODFREY, *Identifiability of uncontrolled nonlinear rational systems*, Automatica J. IFAC, 38 (2002), pp. 1799–1805.
82. N. D. EVANS, M. J. CHAPPELL, M. J. CHAPMAN, AND K. R. GODFREY, *Structural indistinguishability between uncontrolled (autonomous) nonlinear analytic systems*, Automatica J. IFAC, 40 (2004), pp. 1947–1953.
83. G. CICCARELLA, M. DALA MAARA AND A. GERMANI, *A Luenberger-like observer for nonlinear systems*, Inter. J. of Control, 57 (1993), pp. 537–556.
84. N. D. EVANS, L. J. WHITE, M. J. CHAPMAN, K. R. GODFREY, AND M. J. CHAPPELL, *The structural identifiability of the susceptible infected recovered model with seasonal forcing*, Math. Biosci., 194 (2005), pp. 175–197.
85. M. FLIESS, *Quelques définitions de la théorie des systèmes à la lumière des corps différentiels*, C. R. Acad. Sci. Paris Sér. I Math., 304 (1987), pp. 91–93.
86. ———, *Nonlinear control theory and differential algebra*, in Modelling and adaptive control (Sopron, 1986), vol. 105 of Lect. Notes Control Inf. Sci., Springer, Berlin, 1988, pp. 134–145.
87. ———, *Automatique et corps différentiels*, Forum Math., 1 (1989), pp. 227–238.

88. M. FLIESS AND T. GLAD, *Essays on Control*, no. 8 in Springer sciences, Springer, 1993, ch. An Algebraic Approach to Linear and Nonlinear Control, pp. 223–267.
89. M. FLIESS AND H. SIRA-RAMIREZ, *Reconstructeurs d’Etats*, CRAS, 338 (2004), pp. 91–96.
90. A. GÁBOR, A. F. VILLAVARDE, AND J. R. BANGA, *Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems.*, BMC Syst Biol, 11 (2017), p. 54.
91. J. P. GAUTHIER AND I. A. K. KUPKA, *Observability for systems with more outputs than inputs and asymptotic observers*, Math. Z., 223 (1996), pp. 47–78.
92. J. P. GAUTHIER AND I. A. K. KUPKA, *Deterministic Observation Theory and Applications*, Cambridge University Press (2001).
93. G. H. GOLUB AND C. VAN LOAN, *Matrix Computations*, The John Hopkins Univesity Press, 1989.
94. M. B. GRAVENOR, A. L. LLOYD, P. G. KREMSNER, M. A. MISSINOU, M. ENGLISH, K. MARSH, AND D. KWIATKOWSKI, *A model for estimating total parasite load in falciparum malaria patients.*, J Theor Biol, 217 (2002), pp. 137–48.
95. M. B. GRAVENOR, A. R. MCLEAN, AND D. KWIATKOWSKI, *The regulation of malaria parasitaemia: parameter estimates for a population model.*, Parasitology, 110 (Pt 2) (1995), pp. 115–22.
96. M. B. GRAVENOR, M. B. VAN HENS BROEK, AND D. KWIATKOWSKI, *Estimating sequestered parasite population dynamics in cerebral malaria.*, Proc Natl Acad Sci U S A, 95 (1998), pp. 7620–4.
97. E. W. GRIFFITH AND K. S. P. KUMAR, *On the observability of nonlinear systems. I*, J. Math. Anal. Appl., 35 (1971), pp. 135–147.
98. J. GUILLAUME, J. JAKEMAN, S. MARSILI-LINELLI, M. ASHER, P. BRUNNER, B. CROKE, M. HILL, A. JAKEMAN, K. KEESAMN, S. RAZAVI, AND J. STIGTER, *Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose*, Environ. Model. Softw., 119 (2019), pp. 418–432.
99. K. P. HADELER, *Parameter identification in epidemic models*, Math. Biosci., 229 (2011), pp. 185–189.
100. H. HAMMOURI, J.-P. GAUTHIER AND S. OTHMAN, *A simple observer for nonlinear systems applications to bioreactors*, IEEE Trans. Aut. Control, 37 (6) (1992), pp. 875–880.
101. P. HARTMAN, *Ordinary differential equations*, vol. 38 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002. Corrected reprint of the second (1982) edition [Birkhäuser, Boston, MA; MR0658490 (83e:34002)], With a foreword by Peter Bates.
102. R. HERMANN AND A. J. KRENER, *Nonlinear controllability and observability.*, IEEE Trans. Autom. Control, 22 (1977), pp. 728–740.
103. H. W. HETHCOTE AND J. W. VAN ARK, *Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation, and immunization programs*, Math. Biosci., 84 (1987), pp. 85–118.
104. N. J. HIGHAM, *Matrix nearness problems and applications*, in Applications of matrix theory (Bradford, 1988), vol. 22 of Inst. Math. Appl. Conf. Ser. New Ser., Oxford Univ. Press, New York, 1989, pp. 1–27.
105. H. HONG, A. OVCHINNIKOV, G. POGUDIN, AND C. YAP, *SIAN: a tool for assessing structural identifiability of parametric ODEs*, ACM Commun. Comput. Algebra, 53 (2019), pp. 37–40.
106. H. HONG, A. OVCHINNIKOV, G. POGUDIN, AND C. YAP, *Global identifiability of differential models*, Comm. Pure Appl. Math., 73 (2020), pp. 1831–1879.
107. M. HOU AND P. C. MÜLLER, *Design of observers for linear systems with unknown inputs*, IEEE Trans. Automat. Control, 37 (1992), pp. 871–875.
108. S. HUG, A. RAUE, J. HASENAUER, J. BACHMANN, U. KLINGMÜLLER, J. TIMMER, AND F. J. THEIS, *High-dimensional Bayesian parameter estimation: case study for a model of JAK2/STAT5 signaling*, Math. Biosci., 246 (2013), pp. 293–304.
109. A. IBEAS, M. DE LA SEN, S. ALONSO-QUESADA, AND I. ZAMANI, *Stability analysis and observer design for discrete-time SEIR epidemic models*, Adv. Difference Equ., (2015), pp. 2015:122, 21.

110. Y. Inouye. On the observability of autonomous nonlinear systems. *Journal of Mathematical Analysis and Applications*, 60(1):236–247, aug 1977.
111. J. A. JACQUEZ, *Compartmental analysis in Biology and Medicine*, BioMedware, 1996.
112. J. A. JACQUEZ AND P. GREIF, *Numerical parameter identifiability and estimability : Integrating identifiability, estimability and optimal sampling design.*, Math. Biosci., 77 (1985), pp. 201–277.
113. D. L. I. JANZEN, L. BERGENHOLM, M. JIRSTRAND, J. PARKINSON, J. YATES, N. D. EVANS, AND M. J. CHAPPELL, *Parameter identifiability of fundamental pharmacodynamic models.*, Front Physiol, 7 (2016), p. 590.
114. D. L. I. JANZEN, M. JIRSTRAND, M. J. CHAPPELL, AND N. D. EVANS, *Three novel approaches to structural identifiability analysis in mixed-effects models.*, Comput Methods Programs Biomed, (2016).
115. M. JIRSTRAND, *Algebraic Methods for Modeling and Design in Control*, PhD thesis, Linkoping Studies in Science and Technology, Linkoping University, 1996.
116. G. JOLY-BLANCHARD AND L. DENIS-VIDAL, *Some remarks about an identifiability result of nonlinear systems*, Automatica, 34 (1998), pp. 1151–1152.
117. T. KAILATH, *Linear systems*, Prentice-Hall (1990).
118. H. KHALIL, *Nonlinear systems*, Prentice-Hall (1996).
119. R. E. KALMAN, *Mathematical description of linear dynamical systems*, J. SIAM Control Ser. A, 1 (1963), pp. 152–192 (1963).
120. J. KARLSSON, M. ANGUELOVA, AND M. JIRSTRAND, *An efficient Method for Structural identifiability Analysis of Large Dynamic Systems*, in 16th IFAC Symposium on system identification, IFAC, 2012, pp. 941–946.
121. J. T. KEMPER, *Identification of silent infections in SIR epidemics*, Bull. Math. Biol., 43 (1981), pp. 249–257.
122. W. KERMACK AND A. MCKENDRICK, *A contribution to the mathematical theory of epidemics*, Proc. R. Soc., A115 (1927), pp. 700–721.
123. E. R. KOLCHIN, *Differential algebra and algebraic groups*, Academic Press, New York-London, 1973. Pure and Applied Mathematics, Vol. 54.
124. T. C. KOOPMANS AND O. REIERSØ L, *The identification of structural characteristics*, Ann. Math. Statistics, 21 (1950), pp. 165–181.
125. A. KORBEINIKOV AND G. C. WAKE, *Lyapunov functions and global stability for SIR, SIRS, and SIS epidemiological models*, Applied Mathematics Letters, 151(8) (2002), pp. 955–960.
126. N. KAZANTZIS AND C. KRAVARIS, *Nonlinear observer design using Lyapunov’s auxiliary theorem*, Systems and Control Letters 34 (1998), pp. 241–247.
127. F.-X. LE DIMET AND O. TALAGRAND, *Variational algorithms for analysis and assimilation of meteorological observations : theoretical aspects*, Tellus, 38A (1986), pp. 97–110.
128. J. LEANDER, T. LUNDH, AND M. JIRSTRAND, *Stochastic differential equations as a tool to regularize the parameter estimation problem for continuous time dynamical systems given discrete time measurements*, Math. Biosci., 251 (2014), pp. 54–62.
129. A. LEVANT, *Higher-order sliding modes, differentiation and output-feedback control*, International Journal on Automatic, 76, (9/10) (2003), pp. 924–941.
130. M. LI, J. DUSHOFF, AND B. M. BOLKER, *Fitting mechanistic epidemic models to data: a comparison of simple Markov chain Monte Carlo approaches*, Stat. Methods Med. Res., 27 (2018), pp. 1956–1967.
131. M. Y. LI, *An introduction to mathematical modeling of infectious diseases*, vol. 2 of Mathematics of Planet Earth, Springer, Cham, 2018.
132. J. LINTUSAARI, M. U. GUTMANN, S. KASKI, AND J. CORANDER, *On the identifiability of transmission dynamic models for infectious diseases.*, Genetics, 202 (2016), pp. 911–918.
133. L. LJUNG, *System identification : Theory for the user*, Prentice Hall, 1999.
134. L. LJUNG AND T. GLAD, *On global identifiability for arbitrary model parametrizations*, Automatica J. IFAC, 30 (1994), pp. 265–276.
135. D.G. LUENBERGER, *An Introduction to Observers*, IEEE Transactions on Automatic Control, 16, (1971), pp. 596–602.

136. G. MACDONALD, *The dynamics of helminth infections, with special reference to schistosomes.*, Trans R Soc Trop Med Hyg, 59 (1965), pp. 489–506.
137. P. MAGAL AND G. WEBB, *The parameter identification problem for SIR epidemic models: identifying unreported cases*, J. Math. Biol., 77 (2018), pp. 1629–1648.
138. G. MARGARIA, E. RICCOMAGNO, M. J. CHAPPELL, AND H. P. WYNN, *Differential algebra methods for the study of the structural identifiability of rational function state-space models in the biosciences*, Math. Biosci., 174 (2001), pp. 1–26.
139. G. MARGARIA, E. RICCOMAGNO, AND L. J. WHITE, *Structural identifiability analysis of some highly structured families of state space models using differential algebra*, J. Math. Biol., 49 (2004), pp. 433–454.
140. M. MARTCHEVA, *An introduction to mathematical epidemiology*, vol. 61 of Texts in Applied Mathematics, Springer, New York, 2015.
141. H. MCCALLUM, N. BARLOW, AND J. HONE, *How should pathogen transmission be modelled?*, Trends Ecol Evol, 16 (2001), pp. 295–300.
142. N. MESHKAT, C. ANDERSON, AND J. J. DiSTEFANO, III, *Alternative to Ritt’s pseudodivision for finding the input-output equations of multi-output models*, Math. Biosci., 239 (2012), pp. 117–123.
143. N. MESHKAT, M. EISENBERG, AND J. J. DiSTEFANO, III, *An algorithm for finding globally identifiable parameter combinations of nonlinear ODE models using Gröbner bases*, Math. Biosci., 222 (2009), pp. 61–72.
144. N. MESHKAT, C. E.-z. KUO, AND J. R. DiSTEFANO, *On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and combos: a novel web implementation.*, PLoS One, 9 (2014), p. e110261.
145. N. MESHKAT, Z. ROSEN, AND S. SULLIVANT, *Algebraic tools for the analysis of state space models*, in The 50th anniversary of Gröbner bases, vol. 77 of Adv. Stud. Pure Math., Math. Soc. Japan, Tokyo, 2018, pp. 171–205.
146. N. C. MESHKAT, *A Differential Algebra Method for Eliminating Unidentifiability*, PhD thesis, UCLA, 2011. Thesis (Ph.D.)—University of California, Los Angeles.
147. H. MIAO, X. XIA, A. S. PERELSON, AND H. WU, *On identifiability of nonlinear ODE models and applications in viral dynamics*, SIAM Rev., 53 (2011), pp. 3–39.
148. J. MORENO AND D. DOCHAIN, *Global observability and detectability analysis of uncertain reaction systems*, International Journal of Control, 81(7) (2008), pp. 1062–1070.
149. T. G. MÜLLER, N. NOYKOVA, M. GYLLENBERG, AND J. TIMMER, *Parameter identification in dynamical models of anaerobic waste water treatment*, Math. Biosci., 177/178 (2002), pp. 147–160. Deterministic and stochastic modeling of biointeraction (West Lafayette, IN, 2000).
150. J. MURRAY, *Mathematical Biology I: An introduction*, vol. 17 of Interdisciplinary Applied Mathematics, Springer-Verlag, 2002.
151. M. NAKATSUI, A. SEDOGLAVIC, F. LEMAIRE, F. BOULIER, A. ÜRGÜPLÜ, AND K. HORIMOTO, *A general procedure for accurate parameter estimation in dynamic systems using new estimation errors*, in Algebraic and numeric biology, vol. 6479 of Lecture Notes in Comput. Sci., Springer, Heidelberg, 2012, pp. 149–166.
152. S. NAZARI, *The Unknown Input Observer and its Advantages with Examples*, preprint arXiv:1504.07300 (2015).
153. V. K. NGUYEN, S. C. BINDER, A. BOIANELLI, M. MEYER-HERMANN, AND E. A. HERNANDEZ-VARGAS, *Ebola virus infection modeling and identifiability problems.*, Front Microbiol, 6 (2015), p. 257.
154. M. A. NOWAK AND R. M. MAY, *virus dynamics. Mathematical principles of immunology and virology*, Oxford University Press, 2000.
155. F. OLLIVIER, *Le problème de l’identifiabilité structurelle globale : étude théorique, méthodes effectives et bornes de complexité.*, PhD thesis, Ecole Polytechnique, 1990.
156. F. OLLIVIER, *Standard bases of differential ideals*, in Applied algebra, algebraic algorithms and error-correcting codes (Tokyo, 1990), vol. 508 of Lecture Notes in Comput. Sci., Springer, Berlin, 1991, pp. 304–321.

157. P. D. O'NEILL, *A tutorial introduction to bayesian inference for stochastic epidemic models using markov chain monte carlo methods*, *Mathematical Biosciences*, 180 (2002), pp. 103–114.
158. A. OVCHINNIKOV, A. PILLAY, G. POGUDIN, AND T. SCANLON, *Multi-experiment parameter identifiability of odes and model theory*, 2020.
159. S. PANT, *Information sensitivity functions to assess parameter information gain and identifiability of dynamical systems.*, *J R Soc Interface*, 15 (2018).
160. S. PANT AND D. LOMBARDI, *An information-theoretic approach to assess practical identifiability of parametric dynamical systems*, *Math. Biosci.*, 268 (2015), pp. 66–79.
161. P.C. PARKS, *A. M. Lyapunov's stability theory—100 years on*, *IMA Journal of Mathematical Control and Information*, 9(4) (1992), pp. 275–303.
162. M. PEIFER AND J. TIMMER, *Parameter estimation in ordinary differential equations for biochemical processes using the method of multiple shooting.*, *IET Syst. Biol.*, (2007).
163. A. PERASSO, B. LAROCHE, Y. CHITOUR, AND S. TOUZEAU, *Identifiability analysis of an epidemiological model in a structured population*, *J. Math. Anal. Appl.*, 374 (2011), pp. 154–165.
164. A. S. PERELSON AND P. W. NELSON, *Mathematical analysis of HIV-1 dynamics in vivo*, *SIAM Rev.*, 41-1(1999), pp. 3–44.
165. H. POHJANPALO, *System identifiability based on the power series expansion of the solution*, *Math. Biosci.*, 41 (1978), pp. 21–33.
166. M. RAISSI, N. RAMEZANI, AND P. SESHAIYER, *On parameter estimation approaches for predicting disease transmission through optimization, deep learning and statistical inference methods*, *Lett. Biomath.*, 6 (2019), p. 26.
167. A. RAPAPORT AND A. MALOUM, *Design of exponential observers for nonlinear systems by embedding*, *International Journal of Robust and Nonlinear Control*, 14 (2004), pp. 273–288.
168. A. RAUE, V. BECKER, U. KLINGMÜLLER, AND J. TIMMER, *Identifiability and observability analysis for experimental design in nonlinear dynamical models*, *Chaos*, 20 (2010), pp. 045105, 8.
169. A. RAUE, J. KARLSSON, M. P. SACCOMANI, M. JIRSTRAND, AND J. TIMMER, *Comparison of approaches for parameter identifiability analysis of biological systems.*, *Bioinformatics*, 30 (2014), pp. 1440–1448.
170. A. RAUE, C. KREUTZ, T. MAIWALD, J. BACHMANN, M. SCHILLING, U. KLINGMULLER, AND J. TIMMER, *Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood.*, *Bioinformatics*, 25 (2009), pp. 1923–1929.
171. A. RAUE, C. KREUTZ, T. MAIWALD, U. KLINGMULLER, AND J. TIMMER, *Addressing parameter identifiability by model-based experimentation.*, *IET Syst Biol*, 5 (2011), pp. 120–130.
172. A. RAUE, B. STEIERT, M. SCHELKER, C. KREUTZ, T. MAIWALD, H. HASS, J. VANLIER, C. TÖNSING, L. ADLUNG, R. ENGESSER, W. MADER, T. HEINEMANN, J. HASENAUER, M. SCHILLING, T. HÖFER, E. KLIPP, F. THEIS, U. KLINGMÜLLER, B. SCHÖBERL, AND J. TIMMER, *Data2dynamics: a modeling environment tailored to parameter estimation in dynamical systems.*, *Bioinformatics*, 31 (2015), pp. 3558–3560.
173. J. REID, *Structural identifiability in linear-time invariant systems*, *IEEE Trans. Autom. Control*, 22 (1977), pp. 242–246.
174. J. F. RITT, *Differential algebra*, *Dover Publications, Inc.*, New York, 1966.
175. M. G. ROBERTS, *An epidemic model with noisy parameters*, *Math. Biosci.*, 287 (2017), pp. 36–41.
176. C. RODA, B. VARUGHESE, D. HAN, AND M. LI, *Why is it difficult to accurately predict the covid-19 epidemic?*, preprint, (2020).
177. W. C. RODA, *Bayesian inference for dynamical systems.*, *Infect Dis Model*, 5 (2020), pp. 221–232.
178. W. C. RODA, M. B. VARUGHESE, D. HAN, AND M. Y. LI, *Why is it difficult to accurately predict the covid-19 epidemic?*, *Infect Dis Model*, 5 (2020), pp. 271–281.
179. M. RODRIGUEZ-FERNANDEZ, J. A. EGEA, AND J. R. BANGA, *Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems.*, *BMC Bioinformatics*, 7 (2006), p. 483.
180. R. ROSS, *The prevention of malaria*, *John Murray*, 1911.

181. M. SACCOMANI, S. AUDOLY, G. BELLU, AND L. D'ANGIÒ, *Parameter identifiability of nonlinear biological systems*, in Positive systems (Rome, 2003), vol. 294 of Lect. Notes Control Inf. Sci., Springer, Berlin, 2003, pp. 87–93.
182. M. P. SACCOMANI, *An effective automatic procedure for testing parameter identifiability of HIV/AIDS models*, Bull. Math. Biol., 73 (2011), pp. 1734–1753.
183. M. P. SACCOMANI, *Structural vs practical identifiability in system biology*, in IWWBIO 2013, 2013, pp. 305–313.
184. M. P. SACCOMANI, S. AUDOLY, AND L. D'ANGIÒ, *Parameter identifiability of nonlinear systems: the role of initial conditions*, Automatica J. IFAC, 39 (2003), pp. 619–632.
185. G. A. F. SEBER AND C. J. WILD, *Nonlinear regression*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1989.
186. A. SEDOGLAVIC, *A probabilistic algorithm to test local algebraic observability in polynomial time*, J. Symbolic Comput., 33 (2002), pp. 735–755. Computer algebra (London, ON, 2001).
187. A. SOLONEN, H. HAARIO, J. M. TCHUENCHE, AND H. RWEZAURA, *Studying the identifiability of epidemiological models using MCMC*, Int. J. Biomath., 6 (2013), pp. 1350008, 18.
188. E. D. SONTAG, *On the length of inputs necessary in order to identify a deterministic linear system*, IEEE Trans. Automat. Control, 25 (1980), pp. 120–121.
189. E. D. SONTAG, *Mathematical control theory, deterministic finite dimensional systems*, no. 6 in Texts in Applied Mathematics, Springer-Verlag, 1990.
190. E. D. SONTAG, *Spaces of observables in nonlinear control*, in Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994), Birkhäuser, Basel, 1995, pp. 1532–1545.
191. ———, *Critical points for least-squares problems involving certain analytic functions, with applications to sigmoidal nets*, Adv. Comput. Math., 5 (1996), pp. 245–268.
192. E. D. SONTAG, *For differential equations with r parameters, $2r + 1$ experiments are enough for identification*, J. Nonlinear Sci., 12 (2002), pp. 553–583.
193. E. D. SONTAG, *Dynamic compensation, parameter identifiability, and equivariances.*, PLoS Comput Biol, 13 (2017), p. e1005447.
194. E. D. SONTAG AND Y. WANG, *I/o equations for nonlinear systems and observation spaces*, in Proc. IEEE Conf. Decision and Control, Brighton, UK., 1991, IEEE Publications, 1991, pages 720–725, 1991., 1991, pp. 720–725.
195. R. C. SPEAR AND A. HUBBARD, *Parameter estimation and site-specific calibration of disease transmission models.*, Adv Exp Med Biol, 673 (2010), pp. 99–111.
196. S. SPURGEON, *Sliding mode observers: a survey*, International Journal of Systems Science, 39(8) (2008), pp. 751–764.
197. O. TALAGRAND, *On the mathematics of data assimilation*, Tellus, 33 (1981), pp. 321–339.
198. R. TAMI, D. BOUTAT, G. ZHENG, AND F. KRATZ, *Parameters and state estimation for dengue epidemic model*, in 2014 European control conference (ECC^o, 2014, pp. 528–533.
199. A. TARANTOLA, *Inverse problem theory and methods for model parameter estimation*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.
200. K. THOMASETH AND C. COBELLI, *Generalized sensitivity functions in physiological system identification.*, Ann Biomed Eng, 27 (1999), pp. 607–616.
201. C. TÖNSING, J. TIMMER, AND C. KREUTZ, *Profile likelihood-based analyses of infectious disease models*, Stat. Methods Med. Res., 27 (2018), pp. 1979–1998.
202. E. T. TUNALI AND T. J. TARN, *New results for identifiability of nonlinear systems*, IEEE Trans. Automat. Control, 32 (1987), pp. 146–154.
203. N. TUNCER, H. GULBUDAK, V. L. CANNATARO, AND M. MARTCHEVA, *Structural and practical identifiability issues of immuno-epidemiological vector-host models with application to Rift Valley Fever*, Bull. Math. Biol., 78 (2016), pp. 1796–1827.
204. N. TUNCER AND T. T. LE, *Structural and practical identifiability analysis of outbreak models*, Math. Biosci., 299 (2018), pp. 1–18.
205. N. TUNCER, M. MARCOTHEVA, B. LABARRE, AND S. PAYOUTE, *Structural and practical identifiability analysis of Zika epidemiological models*, Bull. Math. Biol., 80 (2018), pp. 2209–2241.

206. S. VAJDA, K. R. GODFREY, AND H. RABITZ, *Similarity transformation approach to identifiability analysis of nonlinear compartmental models*, Math. Biosci., 93 (1989), pp. 217–248.
207. J. X. VELASCO-HERNÁNDEZ, J. A. GARCÍA, AND D. KIRSCHNER, *Remarks on modeling host-viral dynamics and treatment.*, in Mathematical approaches for emerging and reemerging infectious diseases: An introduction. Proceedings of a tutorial Introduction to epidemiology and immunology, Castillo-Chavez, ed., Springer. IMA Vol. Math. Appl. 125, 2002, pp. 287–308.
208. A. VILLAVERDE, *Observability and structural identifiability of nonlinear biological systems*, Complexity, 2019 (2019).
209. A. VILLAVERDE AND J. BANGA, *Reverse engineering and identification in systems biology: strategies, perspectives and challenges*, J. R. Soc. Interface, 11 (2014).
210. A. F. VILLAVERDE, A. BARREIRO, AND A. PAPACHRISTODOULOU, *Structural identifiability of dynamic systems biology models.*, PLoS Comput Biol, 12 (2016), p. e1005153.
211. A. F. VILLAVERDE, N. TSIANTIS, AND J. R. BANGA, *Full observability and estimation of unknown inputs, states and parameters of nonlinear biological models.*, J R Soc Interface, 16 (2019), p. 20190043.
212. E. WALTER, *Identifiability of state space models*, vol. 46 of Lecture Notes in Biomathematics, Springer-Verlag, Berlin-New York, 1982. With applications to transformation systems.
213. E. WALTER AND Y. LECOURTIER, *Global approaches to identifiability testing for linear and nonlinear state space models*, Math. Comput. Simulation, 24 (1982), pp. 472–482.
214. E. WALTER AND L. PRONZATO, *Identification de Modèles Paramétriques à partir de Données Expérimentales*, Masson, 1994.
215. E. WALTER AND L. PRONZATO, *Identifiabilities and nonlinearities*, in Nonlinear systems, Vol. 1, Chapman and Hall, London, 1995, pp. 111–143.
216. E. WALTER AND L. PRONZATO, *Identification of parametric models*, Communications and Control Engineering Series, Springer-Verlag, Berlin; Masson, Paris, 1997. From experimental data, Translated from the 1994 French original and revised by the authors, with the help of John Norton.
217. Y. WANG AND E. D. SONTAG, *On two definitions of observation spaces.*, Syst. Control Lett., 13 (1989), pp. 279–289.
218. Y. WANG AND E. D. SONTAG, *Orders of input/output differential equations and state-space dimensions*, SIAM J. Control Optim., 33 (1995), pp. 1102–1126.
219. F.-G. WIELAND, A. L. HAUBER, M. ROSENBLATT, C. TÖNSING, AND J. TIMMER, *On structural and practical identifiability*, Curr Opin Syst Biol, 25 (2021), pp. 60–69.
220. W. Wonham. *Linear multivariable control: a geometric approach. 2nd ed.* Springer, 1979.
221. H. WU, H. ZHU, H. MIAO, AND A. S. PERELSON, *Parameter identifiability and estimation of HIV/AIDS dynamic models*, Bull. Math. Biol., 70 (2008), pp. 785–799.
222. X. XIA, *Estimation of HIV/AIDS parameters*, Automatica J. IFAC, 39 (2003), pp. 1983–1988.
223. X. XIA AND C. H. MOOG, *Identifiability of nonlinear systems with application to HIV/AIDS models*, IEEE Trans. Automat. Control, 48 (2003), pp. 330–336.
224. J. W. T. YATES, R. D. O. JONES, M. WALKER, AND S. Y. A. CHEUNG, *Structural identifiability and indistinguishability of compartmental models.*, Expert Opin Drug Metab Toxicol, 5 (2009), pp. 295–302.