SIAN: software for structural identifiability analysis of ODE models

Hoon Hong[*], Alexey Ovchinnikov[†], Gleb Pogudin[‡], and Chee Yap[§]

Abstract

Biological processes are often modeled by ordinary differential equations with unknown parameters. The unknown parameters are usually estimated from experimental data. In some cases, due to the structure of the model, this estimation problem does not have a unique solution even in the case of continuous noise-free data. It is therefore desirable to check the uniqueness a priori before carrying out actual experiment. We present a new software SIAN (Structural Identifiability ANalyser) that does this. Our software can tackle problems that could not be tackled by previously developed packages.

1 Introduction

Ordinary differential equations (ODEs) with unknown parameters are widely used for modeling biological processes and phenomena. One is often interested in the values of these parameters due to their importance, as, e.g., they may represent key biological mechanisms or targets for intervention. A standard way to find the values of the parameters from experimental data is to find the parameter values that fit the data with minimal error, typically framed from a statistical perspective as maximum likelihood or Bayesian inference.

However, it might happen that, due to the structure of the model, it is impossible to recover the value of a parameter of interest from the data even assuming the ideal case of a continuous noise-free data. If this is the case, then regardless of the chosen data fitting approach, it is impossible to guarantee that it will find the correct parameter value. As we will see, this structural property can be assessed *a priori* without conducting (often costly) experiments. Thus, a crucial first step to any parameter estimation problem is to check whether the parameter of interest is *structurally globally identifiable*, i.e., the parameter value can be recovered uniquely from the data under the assumption that the data is continuous and noise-free. We explain the notion of global identifiability in more detail in Section 3. For a formal definition and illustrating examples, we refer to [Hong et al., 2018, Section 2].

We present SIAN (Structural Identifiability ANalyser), our new software for assessing identifiability for ODE models, based on the algorithm developed and rigorously justified in [Hong et al., 2018].

2 Existing software for structural identifiability

Assessing global identifiability is a challenging problem. Hence a weaker notion called "local identifiability" was introduced and tackled first. "Local" indicates that a parameter can be identified locally (in some neighborhood). For a polynomial system, it is the same as saying that a parameter can be identified up to finitely many options. There are fast and reliable software packages for assessing local identifiability such as ObservabilityTest [Sedoglavic, 2002] and EAR [Karlsson et al., 2012].

[*]hong@ncsu.edu, Department of Mathematics, North Carolina State University, Raleigh, USA

[†]aovchinnikov@qc.cuny.edu, Department of Mathematics, CUNY Queens College and Ph.D. Programs in Mathematics and Computer Science, CUNY Graudate Center, New York, USA

[‡]pogudin@cims.nyu.edu, Courant Institute of Mathematical Sciences, New York University

[§]yap@cs.nyu.edu, Courant Institute of Mathematical Sciences, New York University

However, even relatively simple real-life systems can involve locally but not globally identifiable parameters (see [Thomaseth and Saccomani, 2018, Section 4], [Norton, 1982], and Supplementary Materials A.1). Thus, it is highly desirable to have software that could assess global identifiability. There has been significant progress in this direction:

- packages DAISY [Bellu et al., 2007] and COMBOS [Meshkat et al., 2014] are based on the approach via input-output equations and can check global identifiability for systems with the "solvability" property (see [Hong et al., 2018, Example 6] for a discussion).
- GenSSI 2.0 package [Ligon et al., 2018] is based on the generating series approach and checks global identifiability conditionally on extra input, the truncation order (for a discussion how the truncation order affects the output of the algorithm, see [Hong et al., 2018, Example 7]).

3 Features

We present SIAN, software written in MAPLE, that has the following input-output specification.

Input. A **system** Σ of the form

$$\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{u}), \\ \boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{u}), \\ \boldsymbol{x}(0) = \boldsymbol{x}^*, \end{cases} \tag{1}$$

- $\boldsymbol{x}$ is a vector of state variables,
- $\boldsymbol{u}$ is a vector of input (control) variables to be chosen by an experimenter,
- $\boldsymbol{y}$ is a vector of output variables,
- $\boldsymbol{\mu}$ and $\boldsymbol{x}^*$ are vectors of unknown scalar parameters and unknown initial conditions, respectively,
- $\boldsymbol{f}$ and $\boldsymbol{g}$ are vectors of rational functions in $\boldsymbol{x}$, $\boldsymbol{\mu}$, and $\boldsymbol{u}$ with complex coefficients (other types of functions can also be handled, see Supplementary Material A.2)

and **a real number** $0 < p < 1$, the user-specified probability of correctness of the result. That is, SIAN is a Monte Carlo randomized algorithm, see [Motwani and Raghavan, 1995, Chapter 1.2].

Output. For every $\theta \in \boldsymbol{\mu} \cup \boldsymbol{x}^*$, SIAN assigns one of the following labels:

- **Globally identifiable**: for almost every solution of (1), every solution of (1) with the same $\boldsymbol{u}$-component and $\boldsymbol{y}$-component has the same value of θ.
- **Locally but not globally identifiable**: for almost every solution of (1), among the solutions of (1) with the same $\boldsymbol{u}$-component and $\boldsymbol{y}$-component, there are only finitely many possible values of θ.
- **Not identifiable**: for almost every solution of (1), among the solutions of (1) with the same $\boldsymbol{u}$-component and $\boldsymbol{y}$-component, there are infinitely many possible values of θ.

The assigned labels are correct with probability at least p.

We would like to emphasize the following **extra features**:

- SIAN is parallellizable and can take advantage of a multicore computing environment.
- SIAN assesses not only the identifiability of the model, but checks individual identifiability of every parameter.
- SIAN can assess identifiability of the parameters appearing in the system and the initial condition. Identifiability of initial conditions is often referred to as observability.

4 Performance and Applications

In this section, we compare our software with the existing software tools for assessing global identifiability, namely COMBOS, DAISY, and GenSSI (see Section 2). All of the benchmark problems are listed in Supplementary Material B. The source code of the benchmarks problems for COMBOS, DAISY and GenSSI used for the comparison is included into the Supplementary Data. The source code for the benchmark problems for SIAN is available at https://github.com/pogudingleb/SIAN/tree/master/examples.

We use a computer with 96 CPUs, 2.4 GHz and CentOS 6.9 (Linux). The runtimes in Table 1 are the elapsed time. SIAN was run on MAPLE 2017 with the probability of correctness $p = 0.99$, GenSSI 2.0 was run on Matlab R2017a, and we used DAISY 1.9.

Table 1: Runtimes (in minutes) on benchmark problems

Example	GenSSI 2.0	COMBOS	DAISY	SIAN
Chemical Reaction	$*$	$**$	$> 6{,}000$	< 1
HIV	$> 12{,}000$	$**$	$> 6{,}600$	< 1
SIRS w/ forcing	$> 12{,}000$	$**$	$> 6{,}600$	< 1
Cholera	$*$	85	30	3
Protein complex	$> 12{,}000$	$**$	$> 6{,}600$	47
Pharmacokinetics	$> 12{,}000$	$**$	$> 7{,}800$	962

$*$: GenSSI 2.0 returns "*Warning: Unable to find explicit solution.*"

$**$: COMBOS returns "*Model may have been entered incorrectly or cannot be solved with COMBOS algorithms.*"

Acknowledgements

The authors are grateful to the CCiS at Queens College and CIMS NYU for the computational resources and to Julio Banga, Marisa Eisenberg, Nikki Meshkat, and Maria Pia Saccomani for useful discussions.

Funding

This work was supported by the National Science Foundation [CCF-1563942, CCF-1564132, CCF-1319632, DMS-1760448, CCF-1708884]; National Security Agency [#H98230-18-1-0016]; and City University of New York [PSC-CUNY #69827-0047, #60098-00 48].

References

G. Bellu, M. P. Saccomani, S. Audoly, and L. DAngió. DAISY: A new software tool to test global identifiability of biological and physiological systems. *Computer Methods and Programs in Biomedicine*, 88(1):52–61, 2007. URL http://dx.doi.org/10.1016/j.cmpb.2007.07.002.

H. Hong, A. Ovchinnikov, G. Pogudin, and C. Yap. Global identifiability of differential models. preprint, 2018. URL http://arxiv.org/abs/1801.08112.

J. Karlsson, M. Anguelova, and M. Jirstrand. An efficient method for structural identifiability analysis of large dynamic systems*. *IFAC Proceedings Volumes*, 45(16):941 – 946, 2012. URL https://doi.org/10.3182/20120711-3-BE-2027.00381.

T. Ligon, F. Fröhlich, O. T. Chiş, J. Banga, E. Balsa-Canto, and J. Hasenauer. GenSSI 2.0: multi-experiment structural identifiability analysis of SBML models. *Bioinformatics*, 34:1421–1423, 2018. URL http://dx.doi.org/10.1093/bioinformatics/btx735.

N. Meshkat, C. E.-z. Kuo, and J. DiStefano, III. On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and COMBOS: A novel web implementation. *PLOS ONE*, 9: 1–14, 10 2014. URL https://doi.org/10.1371/journal.pone.0110261.

R. Motwani and P. Raghavan. *Randomized algorithms*. Cambridge University Press, 1995.

J. Norton. An investigation of the sources of nonuniqueness in deterministic identifiability. *Mathematical Biosciences*, 60(1):89–108, 1982. URL https://doi.org/10.1016/0025-5564(82)90033-5.

A. Sedoglavic. A probabilistic algorithm to test local algebraic observability in polynomial time. *Journal of Symbolic Computation*, 33(5):735 – 755, 2002. URL http://dx.doi.org/10.1006/jsco.2002.0532.

K. Thomaseth and M. P. Saccomani. Local identifiability analysis of nonlinear ODE models: How to determine all candidate solutions. *IFAC-PapersOnLine*, 51(2):529–534, 2018. URL https://doi.org/10.1016/j.ifacol.2018.03.089.

Supplementary materials
SIAN: software for structural identifiability analysis of ODE models

Hoon Hong, Alexey Ovchinnikov, Gleb Pogudin, and Chee Yap

This document is structured as follows:

- Section A contains three illustrating examples mentioned in the text of the paper.
- Section B contains descriptions of the benchmark problems used in Section 4 of the paper.

For the convenience of the reader while navigating between the main paper and the Supplementary materials, we recall that SIAN, software written in MAPLE, has the following input-output specification.

Input. A **system** Σ of the form

$$\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{u}), \\ \boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{u}), \\ \boldsymbol{x}(0) = \boldsymbol{x}^*, \end{cases} \tag{1}$$

where

- $\boldsymbol{x}$ is a vector of state variables,
- $\boldsymbol{u}$ is a vector of input (control) variables to be chosen by an experimenter,
- $\boldsymbol{y}$ is a vector of output variables,
- $\boldsymbol{\mu}$ and $\boldsymbol{x}^*$ are vectors of unknown scalar parameters and unknown initial conditions, respectively,
- $\boldsymbol{f}$ and $\boldsymbol{g}$ are vectors of rational functions in $\boldsymbol{x}$, $\boldsymbol{\mu}$, and $\boldsymbol{u}$ with complex coefficients (other types of functions can also be handled, see Section A.2)

and **a real number** $0 < p < 1$, the user-specified probability of correctness of the result. That is, SIAN is a Monte Carlo randomized algorithm, see [10, Chapter 1.2].

Output. For every $\theta \in \boldsymbol{\mu} \cup \boldsymbol{x}^*$, the program assigns one of the following labels:

- **Globally identifiable**: for almost every solution of (1), every solution of (1) with the same $\boldsymbol{u}$-component and $\boldsymbol{y}$-component has the same value of θ.
- **Locally but not globally identifiable**: for almost every solution of (1), among the solutions of (1) with the same $\boldsymbol{u}$-component and $\boldsymbol{y}$-component, there are only finitely many possible values of θ.
- **Not identifiable**: for almost every solution of (1), among the solutions of (1) with the same $\boldsymbol{u}$-component and $\boldsymbol{y}$-component, there are infinitely many possible values of θ.

The assigned labels are correct with probability at least p.

A Illustrating examples

A.1 Chemical reaction

Purpose of the example

- to show that locally but not globally identifiable parameters appear even in small systems arising in real-life systems;
- to illustrate how one could take into account the possibility of having some of the parameters
  - unknown at the stage of creating the model but
  - become directly known (measured) while performing the experiment.

System and discussion. Consider the following consecutive reaction scheme with three species A, B, and C:

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C.$$

Then the amounts x_A, x_B, and x_C of species evolve according to the following system of differential equations

$$\begin{cases} \dot{x}_A = -k_1 x_A, \\ \dot{x}_B = k_1 x_A - k_2 x_B, \\ \dot{x}_C = k_2 x_B. \end{cases} \tag{2}$$

We assume that, in the experiment, we can observe the amount x_C and a combination $\varepsilon_A x_A + \varepsilon_B x_B + \varepsilon_C x_C$ (where ε_A, ε_B, and ε_C are parameters), which may represent absorbance, conductivity, or ligand release [12, p. 701]. This gives two outputs $y_1 = x_C$ and $y_2 = \varepsilon_A x_A + \varepsilon_B x_B + \varepsilon_C x_C$.

In addition to this, we are also given [12, p. 701] that the values of the parameters ε_A and ε_C will become known at the experiment stage but are unknown at the modeling stage. We can encode this within our framework by considering ε_A and ε_C as *observable functions (outputs) with zero derivative*. In total, we arrive at the following system:

$$\begin{cases} \dot{x}_A = -k_1 x_A, \\ \dot{x}_B = k_1 x_A - k_2 x_B, \\ \dot{x}_C = k_2 x_B, \\ \dot{\varepsilon}_A = 0, \\ \dot{\varepsilon}_C = 0, \\ y_1 = x_C, \\ y_2 = \varepsilon_A x_A + \varepsilon_B x_B + \varepsilon_C x_C, \\ y_3 = \varepsilon_A, \\ y_4 = \varepsilon_C, \end{cases} \tag{3}$$

where $\boldsymbol{x} = (x_A, x_B, x_C, \varepsilon_A, \varepsilon_C)$, $\boldsymbol{y} = (y_1, y_2, y_3, y_4)$, $\boldsymbol{\mu} = (k_1, k_2, \varepsilon_B)$, and $\boldsymbol{x}^* = (x_A^*, x_B^*, x_C^*, \varepsilon_A^*, \varepsilon_C^*)$.

Results Our software outputs that all the parameters $\boldsymbol{\mu}$ and initial values $\boldsymbol{x}^*$ are locally identifiable, but only x_C^*, ε_A^*, and ε_C^* are globally identifiable. In fact, one can show that the set $\{k_1, k_2\}$ can be always found but any of these two numbers can be either k_1 or k_2 [12, Equation (1.3)]. In the literature, this phenomenon is referred to as slow-fast ambiguity.

Source code: https://github.com/pogudingleb/SIAN/blob/master/examples/SlowFast.mpl.

Remark Applying SIAN to

$$\begin{cases} \dot{x}_A = -k_1 x_A, \\ \dot{x}_B = k_1 x_A - k_2 x_B, \\ \dot{x}_C = k_2 x_B, \\ y_1 = x_C, \\ y_2 = \varepsilon_A x_A + \varepsilon_B x_B + \varepsilon_C x_C, \end{cases}$$

one can show that the assumption that ε_A and ε_C can be measured separately is redundant: they both are globally identifiable even just from y_1 and y_2.

A.2 Ruminal lipolysis

Purpose of the example is to show how one can handle the case in which the right-hand side of some of the equations is not a rational function of the parameters.

System and discussion. The following model of ruminal lipolysis was considered in [9, Equations (1-5)], and its identifiability was discussed in [11, Supplementary Material S2].

$$\begin{cases} \dot{x}_1 = -\frac{k_1 x_1}{k_2 + x_1} e^{-k_3 t}, \\ \dot{x}_2 = \frac{2k_1 x_1}{3(k_2 + x_1)} e^{-k_3 t} - k_4 x_2, \\ \dot{x}_3 = \frac{1}{2} k_4 x_2 - k_4 x_3, \\ \dot{x}_4 = \frac{k_1 x_1}{3(k_2 + x_1)} e^{-k_3 t} + \frac{1}{2} k_4 x_2 + k_4 x_3, \\ y_1 = x_1, \\ y_2 = x_2 + x_3, \\ y_3 = x_4, \end{cases} \tag{4}$$

where $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$, $\boldsymbol{y} = (y_1, y_2, y_3)$, $\boldsymbol{\mu} = (k_1, k_2, k_3, k_4)$, and $\boldsymbol{x}^* = (x_1^*, x_2^*, x_3^*, x_4^*)$.

The right-hand side of some of the equations involve an exponential function. Let us denote $k_1 e^{-k_3 t}$ by x_5. By replacing all occurences of $k_1 e^{-k_3 t}$ by x_5 and adding an extra equation $\dot{x}_5 = -k_3 x_5$, system (4) can be written using just rational functions as follows

$$\begin{cases} \dot{x}_1 = -\frac{x_1 x_5}{k_2 + x_1}, \\ \dot{x}_2 = \frac{2x_1 x_5}{3(k_2 + x_1)} - k_4 x_2, \\ \dot{x}_3 = \frac{1}{2} k_4 x_2 - k_4 x_3, \\ \dot{x}_4 = \frac{x_1 x_5}{3(k_2 + x_1)} + \frac{1}{2} k_4 x_2 + k_4 x_3, \\ \dot{x}_5 = -k_3 x_5, \\ y_1 = x_1, \\ y_2 = x_2 + x_3, \\ y_3 = x_4, \end{cases} \tag{5}$$

where $\boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5)$, $\boldsymbol{y} = (y_1, y_2, y_3)$, $\boldsymbol{\mu} = (k_2, k_3, k_4)$, and $\boldsymbol{x}^* = (x_1^*, x_2^*, x_3^*, x_4^*, k_1)$.

Models (4) and (5) have the same set $\boldsymbol{\mu} \cup \boldsymbol{x}^*$ with the only difference that k_1 has been moved from $\boldsymbol{\mu}$ to $\boldsymbol{x}^*$. One can see that in the sense of identifiability models (4) and (5) are equivalent. Similar change of variables is possible in many cases when the right-hand side of some of the equations is not a rational function (see also Section A.3).

Results All the parameters and initial conditions of (5) (and, consequently, (4)) are globally identifiable.

Source code: https://github.com/pogudingleb/SIAN/blob/master/examples/Lipolysis.mpl.

A.3 Goodwin oscillator

Purpose of the example

- to show that locally but not globally identifiable parameters appear even in small systems arising in real-life systems;
- to show how one can handle the case in which the right-hand side of some of the equations is not a rational function of the parameters.

System and discussion The following model describes the oscillations in enzyme kinetics [6] and has been already used as a benchmark for software for identifiability analysis in [3, Case 1].

$$\begin{cases} \dot{x}_1 = -bx_1 + \frac{a}{A+x_3^\sigma}, \\ \dot{x}_2 = \alpha x_1 - \beta x_2, \\ \dot{x}_3 = \gamma x_2 - \delta x_3, \\ y_1 = x_1, \end{cases} \tag{6}$$

where $\boldsymbol{x} = (x_1, x_2, x_3)$, $\boldsymbol{u} = \varnothing$, $\boldsymbol{y} = (y_1)$, $\boldsymbol{\mu} = (a, A, b, \alpha, \beta, \gamma, \delta, \sigma)$, $\boldsymbol{x}^* = (x_1^*, x_2^*, x_3^*)$.

To bring system (6) to the form (1), we introduce a new parameter c and a new state variable x_4 defined by

$$c = \frac{A}{a}, \quad x_4 = \frac{x_3^\sigma}{a}. \tag{7}$$

Then the first equation in (6) can be rewritten as $\dot{x}_1 = -bx_1 + \frac{1}{c+x_4}$, and an equation for x_4 can be derived as follows

$$\dot{x_4} = \frac{1}{a}\sigma \dot{x}_3 x_3^{\sigma-1} = \sigma \frac{x_3^\sigma}{a} \cdot \frac{\dot{x}_3}{x_3} = \sigma x_4 \frac{\gamma x_2 - \delta x_3}{x_3}.$$

Thus, we can rewrite (6) using just rational functions as

$$\begin{cases} \dot{x}_1 = -bx_1 + \frac{1}{c+x_4}, \\ \dot{x}_2 = \alpha x_1 - \beta x_2, \\ \dot{x}_3 = \gamma x_2 - \delta x_3, \\ \dot{x}_4 = \sigma x_4 \frac{\gamma x_2 - \delta x_3}{x_3}, \\ y_1 = x_1. \end{cases} \tag{8}$$

Here we have

- $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$,
- $\boldsymbol{u} = \varnothing$,
- $\boldsymbol{y} = (y_1)$,
- $\boldsymbol{\mu} = (b, c, \alpha, \beta, \gamma, \delta, \sigma)$,

- $\boldsymbol{x}^* = (x_1^*, x_2^*, x_3^*, x_4^*)$.

Our computations show that in the system (8)

- b, c, σ, x_1^*, and x_4^* are globally identifiable,
- β and δ are locally but not globally identifiable,
- and α, γ, x_2^*, and x_3^* are non-identifiable.

Now we recall that, from (7), we have

$$c = \frac{A}{a} \quad \text{and} \quad x_4^* = \frac{(x_3^*)^\sigma}{a}.$$

If a were locally identifiable, then the global identifiability of x_4^* and σ would imply that x_3^* is locally identifiable. Therefore, since x_3^* is non-identifiable , a is non-identifiable. Together with the global identifiability of c, the non-identifiability of a yields the non-identifiability of A. To sum up, the result of our identifiability analysis of (6) is the following:

- b, σ, and x_1^* are globally identifiable,
- β and δ are locally but not globally identifiable,
- and $a, A, \alpha, \gamma, x_2^*$, and x_3^* are non-identifiable.

Once it is known that one of the parameters is not globally identifiable, one might want to understand where does the non-uniqueness come from and what to do about it. Possible options include:

- Among the possible parameter values, all but one violate some extra constraints coming from biology and can simply be discarded at the data fitting stage.
- The non-uniqueness of the parameter value might arise from a flaw in the model that should be remedied by redesigning the model.
- The non-uniqueness has its own biological meaning, for example, it might indicate the existence of several distinct "regimes" of the model (see, for example, Section A.1). This biological meaning can be further used to identify the value of the parameter uniquely.

A natural step towards understanding the nature of the non-uniqueness of a parameter is to find a change of variables and parameters that leaves the outputs unchanged but changes the value of the parameter. In the case of locally but not globally identifiable parameters β and δ in (8), one such change of variables and parameters is the following:

$$\begin{aligned} &x_1 \to x_1, \quad x_2 \to x_2 + \frac{\beta - \delta}{\gamma} x_3, \quad x_3 \to x_3, \quad x_4 \to x_4, \quad b \to b, \\ &c \to c, \quad \alpha \to \alpha, \quad \beta \to \delta, \quad \gamma \to \gamma, \quad \delta \to \beta, \quad \sigma \to \sigma. \end{aligned} \tag{9}$$

One can verify that (9) preserves the output of (8) by a direct computation. Below we show one way to derive (9) using our software.

1. From the intermediate results of the computation done by SIAN, we can extract that the pair of values $\{\beta, \delta\}$ is identifiable but it is impossible, based on the observations, to find out which of these two numbers is the value of β and which one of them is the value of δ.

2. We try to find which of the state variables and/or parameters can be assumed to be known without making β and δ globally identifiable. Using SIAN, one can verify (in a couple of seconds) that adding extra outputs $y_2 = x_3$, $y_3 = \alpha$, and $y_4 = \gamma$ does not make β and γ globally identifiable.

3. Thus, there exists a change of variables and parameters that swaps β and δ and leaves everything except for β, δ, and x_2 unchanged. We can find the new function $\tilde{x}_2$ by looking at the third equation in (8) before and after the change of variables and parameters

$$\dot{x}_3 = \gamma x_2 - \delta x_3,$$
$$\dot{x}_3 = \gamma \tilde{x}_2 - \beta x_3.$$

A direct computation shows that

$$\tilde{x}_2 = x_2 + \frac{\beta - \delta}{\gamma} x_3.$$

Thus, we arrive at (9).

Results

- b, σ, and x_1^* are globally identifiable,
- β and δ are locally but not globally identifiable,
- and $a, A, \alpha, \gamma, x_2^*$, and x_3^* are non-identifiable.

Source code: https://github.com/pogudingleb/SIAN/blob/master/examples/Goodwin.mpl.

B Benchmarks

Section 4 of the paper compares performance of SIAN, GenSSI 2.0, COMBOS, and DAISY. For the convenience of the reader, we reproduce the table with the runtimes (see Table 1). The purpose of this section is to describe the used benchmark problems. The source files of the benchmark problems for GenSSI 2.0, COMBOS, and DAISY are available in the Supplementary Data at https://cs.nyu.edu/~pogudin/SupplementaryData.zip.

Table 1: Runtimes (in minutes) on benchmark problems

Example	GenSSI 2.0	COMBOS	DAISY	SIAN
Chemical Reaction (B.1)	$*$	$**$	$> 6{,}000$	< 1
HIV (B.2)	$> 12{,}000$	$**$	$> 6{,}600$	< 1
SIRS w/ forcing (B.3)	$> 12{,}000$	$**$	$> 6{,}600$	< 1
Cholera (B.4)	$*$	85	30	3
Protein complex (B.5)	$> 12{,}000$	$**$	$> 6{,}600$	47
Pharmacokinetics (B.6)	$> 12{,}000$	$**$	$> 7{,}800$	962

$*$: GenSSI 2.0 returns "*Warning: Unable to find explicit solution.*"
$**$: COMBOS returns "*Model may have been entered incorrectly or cannot be solved with COMBOS algorithms.*"

All the results presented in the rest of the section are computed with probability of correctness $p = 0.99$.

B.1 Chemical Reaction

System The following system of ODEs corresponds to a chemical reaction network [4, Eq. 3.4], which is a reduced fully processive, n-site phosphorylation network.

$$\begin{cases} \dot{x}_1 &= -\mu_1 x_1 x_2 + \mu_2 x_4 + \mu_4 x_6, \\ \dot{x}_2 &= -\mu_1 x_1 x_2 + \mu_2 x_4 + \mu_3 x_4, \\ \dot{x}_3 &= \mu_3 x_4 + \mu_5 x_6 - \mu_6 x_3 x_5, \\ \dot{x}_4 &= \mu_1 x_1 x_2 - \mu_2 x_4 - \mu_3 x_4, \\ \dot{x}_5 &= \mu_4 x_6 + \mu_5 x_6 - \mu_6 x_3 x_5, \\ \dot{x}_6 &= -\mu_4 x_6 - \mu_5 x_6 + \mu_6 x_3 x_5, \\ y_1 &= x_2, \\ y_2 &= x_3 \end{cases}$$

Here we have

- $\boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5, x_6)$,
- $\boldsymbol{u} = \varnothing$,
- $\boldsymbol{y} = (y_1, y_2)$,
- $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6)$,
- $\boldsymbol{x}^* = (x_1^*, x_2^*, x_3^*, x_4^*, x_5^*, x_6^*)$.

Source code

- SIAN: https://github.com/pogudingleb/SIAN/blob/master/examples/ChemicalReactionNetwork.mpl.
- DAISY: file *DAISY/ChemicalReactionNetwork.txt* in the Supplementary Data.
- COMBOS: file *COMBOS/ChemicalReactionNetwork.txt* in the Supplementary Data.
- GenSSI 2.0: file *GenSSI2/CRN.m* in the Supplementary Data.

Result All the parameters $\boldsymbol{\mu}$ and initial conditions $\boldsymbol{x}^*$ are globally identifiable.

B.2 HIV

System Consider the following model of HIV [13, Equation (6)] that describes immune impairment dynamics.

$$\begin{cases} \dot{x} &= \lambda - dx - \beta xv, \\ \dot{y} &= \beta xv - ay, \\ \dot{v} &= ky - uv, \\ \dot{w} &= czyw - c * q * yw - bw, \\ \dot{z} &= cqyw - hz, \\ y_1 &= w, \\ y_2 &= z. \end{cases}$$

Here we have

- $\boldsymbol{x} = (x, y, v, w, z)$,
- $\boldsymbol{u} = \varnothing$,
- $\boldsymbol{y} = (y_1, y_2)$,
- $\boldsymbol{\mu} = (\beta, \lambda, a, c, d, h, k, q, u)$,
- $\boldsymbol{x}^* = (x^*, y^*, v^*, w^*, z^*)$.

Source code

- SIAN: https://github.com/pogudingleb/SIAN/blob/master/examples/HIV2.mpl.
- DAISY: file *DAISY/HIV2.txt* in the Supplementary Data.
- COMBOS: file *COMBOS/HIV2.txt* in the Supplementary Data.
- GenSSI 2.0: file *GenSSI2/HIV2.m* in the Supplementary Data.

Results a, b, d, h, q, u, w^* and z^* are globally identifiable, $\beta, \lambda, c, k, x^*, y^*$ and v^* are non-identifiable.

B.3 SIRS with forcing

System The following model is an extension of the SIRS model that incorporates the seasonal nature of transmission of RSV [2, Equations (7-11)].

$$\begin{cases} \dot{s} &= \mu - \mu s - b_0(1 + b_1 x_1) i s + g r, \\ \dot{i} &= b_0(1 + b_1 x_1) i s - (\nu + \mu) i, \\ \dot{r} &= \nu i - (\mu + g) r, \\ \dot{x_1} &= -M x_2, \\ \dot{x_2} &= M x_1, \\ y_1 &= i, \\ y_2 &= r. \end{cases}$$

Here we have

- $\boldsymbol{x} = (s, i, r, x_1, x_2)$,
- $\boldsymbol{u} = \varnothing$,
- $\boldsymbol{y} = (y_1, y_2)$,
- $\boldsymbol{\mu} = (\mu, \nu, b_0, b_1, g, M)$,
- $\boldsymbol{x}^* = (s^*, i^*, r^*, x_1^*, x_2^*)$.

Source code

- SIAN: https://github.com/pogudingleb/SIAN/blob/master/examples/SIRSForced.mpl.
- DAISY: file *DAISY/SIRSForced.txt* in the Supplementary Data.
- COMBOS: file *COMBOS/SIRSForced.txt* in the Supplementary Data.
- GenSSI 2.0: file *GenSSI2/SIRSForced.m* in the Supplementary Data.

Results $b_0, g, \mu, \nu, s^*, i^*, r^*$ are globally identifiable, M is locally identifiable, but not globally identifiable, and b_1, x_1^*, x_2^* are non-identifiable.

B.4 Cholera

System The following version of SIWR is an extension of the SIR model, see [7, Eq. 3]:

$$\begin{cases} \dot{s} & = \mu - \beta_I si - \beta_W sw - \mu s + \alpha r, \\ \dot{i} & = \beta_W sw + \beta_I si - \gamma i - \mu i, \\ \dot{w} & = \xi(i - w), \\ \dot{r} & = \gamma i - \mu r - \alpha r, \\ y_1 & = \kappa i, \\ y_2 & = s + i + r, \end{cases}$$

where s, i, and r stand for the fractions of the population that are susceptible, infectious, and recovered, respectively. The variable w represents the concentration of the bacteria in the environment. Here we have

- $\boldsymbol{x} = (s, i, w, r)$,
- $\boldsymbol{u} = \varnothing$,
- $\boldsymbol{y} = (y_1, y_2)$,
- $\boldsymbol{\mu} = (\mu, \beta_I, \beta_W, \alpha, \gamma, \xi, \kappa)$,
- $\boldsymbol{x}^* = (s^*, i^*, w^*, r^*)$.

Source code

- SIAN: https://github.com/pogudingleb/SIAN/blob/master/examples/Cholera.mpl.
- DAISY: file *DAISY/Cholera.txt* in the Supplementary Data.
- COMBOS: file *COMBOS/Cholera.txt* in the Supplementary Data.
- GenSSI 2.0: file *GenSSI2/Cholera.m* in the Supplementary Data.

Result All the parameters $\boldsymbol{\mu}$ and initial conditions $\boldsymbol{x}^*$ are globally identifiable.

B.5 Protein Complex (NFκB)

Consider the model of NFκB regulatory module proposed in [8] (see also [1] and [3, Case 6]) defined by the following system [3, Equation 27]

$$\begin{cases}
\dot{x}_1 = k_{prod} - k_{deg}x_1 - k_1x_1u,\\
\dot{x}_2 = -k_3x_2 - k_{deg}x_2 - a_2x_2x_{10} + t_1x_4 - a_3x_2x_{13} + t_2x_5 + (k_1x_1 - k_2x_2x_8)u,\\
\dot{x}_3 = k_3x_2 - k_{deg}x_3 + k_2x_2x_8u,\\
\dot{x}_4 = a_2x_2x_{10} - t_1x_4,\\
\dot{x}_5 = a_3x_2x_{13} - t_2x_5,\\
\dot{x}_6 = c_{6a}x_{13} - a_1x_6x_{10} + t_2x_5 - i_1x_6,\\
\dot{x}_7 = i_1k_vx_6 - a_1x_{11}x_7,\\
\dot{x}_8 = c_4x_9 - c_5x_8,\\
\dot{x}_9 = c_2 + c_1x_7 - c_3x_9,\\
\dot{x}_{10} = -a_2x_2x_{10} - a_1x_{10}x_6 + c_{4a}x_{12} - c_{5a}x_{10} - i_{1a}x_{10} + e_{1a}x_{11},\\
\dot{x}_{11} = -a_1x_{11}x_7 + i_{1a}k_vx_{10} - e_{1a}k_vx_{11},\\
\dot{x}_{12} = c_{2a} + c_{1a}x_7 - c_{3a}x_{12},\\
\dot{x}_{13} = a_1x_{10}x_6 - c_{6a}x_{13} - a_3x_2x_{13} + e_{2a}x_{14},\\
\dot{x}_{14} = a_1x_{11}x_7 - e_{2a}k_vx_{14},\\
\dot{x}_{15} = c_{2c} + c_{1c}x_7 - c_{3c}x_{15},\\
y_1 = x_2,\\
y_2 = x_{10} + x_{13},\\
y_3 = x_9,\\
y_4 = x_1 + x_2 + x_3,\\
y_5 = x_7,\\
y_6 = x_{12}
\end{cases} \tag{10}$$

The values of all the parameters except $t_1, t_2, c_{3a}, c_{4a}, c_5, k_1, k_2, k_3, k_{prod}, k_{deg}, i_1, e_{2a}, i_{1a}$ are known from the existing literature (see [1, Table 1]). Here we have

- $\mathbf{x} = (x_1, x_2, \ldots, x_{15})$,
- $\mathbf{u} = (u)$,
- $\mathbf{y} = (y_1, y_2, \ldots, y_6)$,
- $\mu = t_1, t_2, c_{3a}, c_{4a}, c_5, k_1, k_2, k_3, k_{prod}, k_{deg}, i_1, e_{2a}, i_{1a}$,
- $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_{15}^*)$.

Source code

- SIAN: https://github.com/pogudingleb/SIAN/blob/master/examples/NFkB.mpl.
- DAISY: file *DAISY/NFkB.txt* in the Supplementary Data.
- COMBOS: file *COMBOS/NFkB.txt* in the Supplementary Data.
- GenSSI 2.0: file *GenSSI2/NFkB.m* in the Supplementary Data.

Result All the parameters $\boldsymbol{\mu}$ and initial values $\boldsymbol{x}^*$ except x_{15}^* are globally identifiable. x_{15}^* is non-identifiable.

B.6 Pharmacokinetics

System This is a simplified version of a model arising in pharmacokinetics [5]:

$$\begin{cases} \dot{x}_1 = a(x_2 - x_1) - \frac{k_a V_m x_1}{k_c k_a + k_c x_3 + k_a x_1}, \\ \dot{x}_2 = a(x_1 - x_2), \\ \dot{x}_3 = b_1(x_4 - x_3) - \frac{k_c V_m x_3}{k_c k_a + k_c x_3 + k_a x_1}, \\ \dot{x}_4 = b_2(x_3 - x_4), \\ y = x_1. \end{cases} \tag{11}$$

Here we have

- $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$,
- $\boldsymbol{u} = \varnothing$,
- $\boldsymbol{y} = (y)$,
- $\boldsymbol{\mu} = (a, b_1, b_2, k_a, k_c, V_m)$,
- $\boldsymbol{x}^* = (x_1^*, x_2^*, x_3^*, x_4^*)$.

Source code

- SIAN: https://github.com/pogudingleb/SIAN/blob/master/examples/Pharm.mpl.
- DAISY: file *DAISY/Pharm.txt* in the Supplementary Data.
- COMBOS: file *COMBOS/Pharm.txt* in the Supplementary Data.
- GenSSI 2.0: file *GenSSI2/Pharmacokinetics.m* in the Supplementary Data.

Result All the parameters $\boldsymbol{\mu}$ and initial values $\boldsymbol{x}^*$ are globally identifiable.

References

[1] E. Balsa-Canto, A. A. Alonso, and J. R. Banga. An iterative identification procedure for dynamic modeling of biochemical networks. *BMC Systems Biology*, 4(11), 2010. URL https://doi.org/10.1186/1752-0509-4-11.

[2] 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. *Bulletin of Mathematical Biology*, 71:1890–1901, 2009. URL http://dx.doi.org/10.1007/s11538-009-9429-3.

[3] O.-T. Chis, J. R. Banga, and E. Balsa-Canto. Structural identifiability of systems biology models: A critical comparison of methods. *PLOS ONE*, 6(11):1–16, 11 2011. URL http://dx.doi.org/10.1371/journal.pone.0027755.

[4] C. Conradi and A. Shiu. Dynamics of post-translational modification systems: recent progress and future directions. *Biophysical Journal*, 114(3):507–515, 2018. URL https://doi.org/10.1016/j.bpj.2017.11.3787.

[5] S. Demignot and D. Domurado. Effect of prosthetic sugar groups on the pharmacokinetics of glucose-oxidase. *Drug Design and Delivery*, 1(4):333–348, 1987.

[6] B. C. Goodwin. Oscillatory behavior in enzymatic control processes. *Advances in Enzyme Regulation*, 3:425–437, 1965. URL https://doi.org/10.1016/0065-2571(65)90067-1.

[7] E. C. Lee, M. R. Kelly, B. M. Ochocki, S. M. Akinwumi, K. E. Hamre, J. H. Tien, and M. C. Eisenberg. Model distinguishability and inference robustness in mechanisms of cholera transmission and loss of immunity. *Journal of Theoretical Biology*, 420:68–81, 2017. URL http://dx.doi.org/10.1016/j.jtbi.2017.01.032.

[8] T. Lipniacki, P. Paszek, A. R. Brasier, B. Luxon, and M. Kimmel. Mathematical model of nf-κb regulatory module. *Journal of Theoretical Biology*, 228(2):195–215, 2004. URL https://doi.org/10.1016/j.jtbi.2004.01.001.

[9] P. Moate, R. Boston, T. Jenkins, and I. Lean. Kinetics of ruminal lipolysis of triacylglycerol and biohydrogenation of long-chain fatty acids: New insights from old data. *Journal of Dairy Science*, 91 (2):731–742, 2008. URL https://doi.org/10.3168/jds.2007-0398.

[10] R. Motwani and P. Raghavan. *Randomized algorithms*. Cambridge University Press, 1995.

[11] R. Muñoz-Tamayo, L. Puillet, J. Daniel, D. Sauvant, O. Martin, M. Taghipoor, and P. Blavy. Review: To be or not to be an identifiable model. Is this a relevant question in animal science modelling? *Animal*, 12(4):701–712, 2018. URL https://doi.org/10.1017/S1751731117002774.

[12] S. Vajda and H. Rabitz. Identifiability and distinguishability of first-order reaction systems. *Journal of Physical Chemistry*, 92:701–707, 1988. URL https://doi.org/10.1021/j100314a024.

[13] D. Wodarz and M. A. Nowak. Mathematical models of HIV pathogenesis and treatment. *Bioessays*, 24(12):1178–1187, 2002. URL http://dx.doi.org/10.1002/bies.10196.