Maple Application for Structural Identifiability Analysis of ODE models

Ilia Ilmer
Department of Computer Science
CUNY Graduate Center
New York, NY, USA,
iilmer@gradcenter.cuny.edu

Alexey Ovchinnikov
Department of Mathematics
CUNY Queens College
New York, NY, USA,
aovchinnikov@qc.cuny.edu

Gleb Pogudin
LIX, CNRS,
École Polytechnique,
Institute Polytechnique de Paris
Paris, France
gleb.pogudin@polytechnique.edu

Abstract

Structural identifiability properties of models of ordinary differential equations help one assess if the parameter's value can be recovered from experimental data. This theoretical property can be queried without the need for data collection and is determined with help of differential algebraic tools. We present a web-based Structural Identifiability Toolbox that rigorously uncovers identifiability properties of individual parameters of ODE systems as well as their functions (also called identifiable combinations) using the apparatus of differential algebra. The application requires no installation and is readily available at https://maple.cloud/app/6509768948056064/

1 Introduction

Let us begin by defining a model in a state-space form.

Definition 1 (Model in the state-space form)**.** A model in *the state-space form* is a system

$$\boldsymbol{\Sigma} := \begin{cases} \mathbf{x}' & = \mathbf{f}(\mathbf{x},\ \boldsymbol{\mu},\ \mathbf{u}), \\ \mathbf{y} & = \mathbf{g}(\mathbf{x},\ \boldsymbol{\mu},\ \mathbf{u}), \\ \mathbf{x}(0) & = \mathbf{x}^*, \end{cases}$$

where $\mathbf{f} = (f_1, \ldots, f_n)$ and $\mathbf{g} = (g_1, \ldots, g_n)$ with $f_i = f_i(\mathbf{x},\ \boldsymbol{\mu},\ \mathbf{u})$, $g_i = g_i(\mathbf{x},\ \boldsymbol{\mu},\ \mathbf{u})$ are rational functions over the field of complex numbers $\mathbb{C}$.

The vector $\mathbf{x} = (x_1, \ldots, x_n)$ represents the time-dependent state variables and $\mathbf{x}'$ represents the derivative. The vector-function $\mathbf{u} = (u_1, \ldots, u_s)$ represents the input variables. The m-vector $\mathbf{y} = (y_1, \ldots, y_n)$ represents the output variables. The vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_\lambda)$ represents the parameters and $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)$ defines initial conditions of the model.

Let us (informally) introduce the concept of identifiability. We say that a parameter is *locally* (respectively, *globally*) identifiable if, from given inputs and outputs of an ODE model, the parameter's value can be recovered up to finitely many values (respectively, such value is unique). If there are infinitely many

such values, the parameter is said to be *non-identifiable*. In this case, we may mitigate the issue by considering all possible identifiable combinations of model's parameters. Such combinations can be obtained from single or multiple experiments; thus we call this type of identifiability single- or multi-experiment identifiability, respectively (see below). The former implies that parameter combinations can already be found from one experiment, while the latter implies that there exists β such that β experiments are sufficient. We refer to [12, Definition 7] and to [13, Definition 2.7] for definitions of single- and multi-experiment identifiability, respectively. Consider an illustrative example of a model that is single-experiment globally identifiable:

$$\begin{cases} x' = ax, \\ y = x. \end{cases}$$

Parameter a is identifiable from a single experiment as $a = \frac{y'(0)}{y(0)}$. On the other hand, consider another toy example:

$$\begin{cases} x_1' = 0, \\ y_1 = x_1,\ y_2 = ax_1 + b. \end{cases}$$

Note that one cannot identify a, b from just one experiment since there are infinitely many pairs a, b that yield the same output. On the other hand, with two experiments, we get

$$\begin{cases} x_{1,1}' = x_{2,1}' = 0, \\ y_{1,1} = x_{1,1},\ y_{1,2} = ax_{1,1} + b, \\ y_{2,1} = x_{2,1},\ y_{2,2} = ax_{2,1} + b, \end{cases}$$

which yields $a = \frac{y_{2,2} - y_{1,2}}{y_{2,1} - y_{1,1}},\ b = \frac{y_{1,1}y_{2,2} - y_{1,2}}{y_{2,1}y_{1,2} - y_{1,1}}$. Therefore $a,\ b$ are identifiable from 2 experiments.

There is a range of existing identifiability software packages such as [1, 8, 10, 14, 16]. We refer to [4, 8] for an overview of available software. Structural Identifiability analyzer (SIAN) from [7] is, to the best of our knowledge, typically most efficient in terms of runtime and resources for parameter global identifiability problems, see [7, Table 1]. For parameter combinations, there are web applications COMBOS and its recent development, COMBOS 2, [9, 11]. The main disadvantage is limited efficiency of the underlying algorithm of COMBOS [7, Table 1]. For other packages, the lack of all-in-one approach without efficiency compromise and requirement of proprietary or unpopular dependencies makes it less convenient for the end-user.

We present a toolbox that solves these problems as a web-based application written in Maple and runnable from a web-browser. Our application is capable of assessing individual identifiability properties using SIAN [7] as well as finding all single- and multi-experiment identifiable combinations of parameters via algorithms and implementations from [12, 13].

2 Computational efficiency

The are two parts of our application, one is responsible for individual parameter identifiability properties and the other assesses identifiability of parameter combinations. The first part utilizes SIAN [7, 8], which transforms the input ODE model into a system of polynomial equations, to categorize individual parameters and initial conditions of the input model as *globally*, *locally-not-globally*, and *non-identifiable*. It relies on computation of a Gröbner basis for global identifiability and rank of Jacobian for local identifiability properties. In addition to individual identifiability properties, we report the multiplicities of locally identifiable parameters. This is achieved by performing a change of variable ordering from graded reverse lexicographic to the elimination ordering using Gröbner walk procedure. By considering each locally-not-globally identifiable parameter, we report the degree of the polynomial in this parameter only having eliminated other variables via Gröbner walk. Note that, while this result is probabilistic due to the Monte Carlo nature of SIAN, the resulting probability will be different from the prescribed one for SIAN.

The second part is based on the results of [12, Theorem 11] and [12, Theorem 21]. To find all identifiable combinations, we use the implementation from [12] which relies on the Rosenfeld-Gröbner algorithm of BLAD package [2] (included in MAPLE as `DifferentialAlgebra` package [3]) for computing characteristic sets used in input-output equations (see [12, Definition 8]). We add an option to refine the bound on the number of experiments, which is achieved by changing the ranking in the Rosenfeld-Gröbner procedure for different variable orderings (see Example 2 below).

Computing single- and multi-experiment identifiable combinations relies additionally on simplifying the generators. In the process of computing this simplification, we utilize Gröbner walk procedure converting the total degree lexicographic order to pure lexicographic ordering. For instance, the computation for the ODE model from DAISY [1, Example 5], `https://daisy.dei.unipd.it/`:

$$\begin{cases} x_1' = -(a_{21} + a_{31} + a_{01})x_1 + a_{12}x_2 + a_{13}x_3 + u(t), \\ x_2' = a_{21}x_1 - a_{12}x_2, \\ x_3' = a_{31}x_1 - a_{13}x_3, \\ y = x_1 \end{cases}$$

finishes the multi-experiment identifiability check in under 1 minute while if the lexicographic ordering is used directly without the Gröbner walk, the process does not finish in reasonable time.

To further maximize the speedup, we take advantage of the output of each algorithm that we use. Concretely, if all parameters are named globally identifiable by SIAN, then this is also interpreted as both single- and multi-experiment identifiability with bound 1. Since SIAN is a Monte Carlo algorithm with user-specified correctness probability p, the single- and multi-experiment identifiability results are to be interpreted as Monte Carlo with the same probability of correctness.

If such bypass is not used, the single- and multi-experiment results are reported *deterministically*. In this case, additional speedup is achieved if the multi-experiment check (which runs first) reports bound equal to 1, which allows us to avoid extra computations for single-experiment queries (which are typically more time-consuming). We can also achieve a speedup for bound refining via single-experiment identifiability. Indeed, if a bound is reported as $\beta > 1$ but the single-experiment identifiable combinations are identical to those identifiable from at most β experiments, we automatically refine the bound to be $\beta = 1$.

If, on the other hand, the bound is $\beta > 1$ and single- and multi-experiment identifiable functions are different, we can run a different refinement procedure. This refinement is performed by permuting variables with a user-specified number of permutations. For each permutation, we obtain a different ranking for Rosenfeld-Gröbner procedure and we retain the smallest bound value that we obtain this way.

3 Examples

Example 1: Multi-experiment Check Bypass via SIAN

Below is an example of a mixed-mechanism network [5], where the state functions $x_i(t), i = 1, \dots, 6$ are concentrations and the parameters $k_i, i = 1, \dots, 6$ are rate constants. The functions y_1, y_2 are the outputs.

$$\begin{cases} x_1' = -k_1x_1x_2 + k_2x_4 + k_4x_6, \\ x_2' = k_1x_1x_2 + k_2x_4 + k_3x_4, \\ x_3' = k_3x_4 + k_5x_6 - k_6x_3x_5, \\ x_4' = k_1x_1x_2 - k_2x_4 - k_3x_4, \\ x_5' = k_4x_6 + k_5x_6 - k_6x_3x_5, \\ x_6' = -k_4x_6 - k_5x_6 + k_6x_3x_5, \\ y_1 = x_3,\ y_2 = x_2. \end{cases}$$

The global and local identifiability for all parameters is returned in under 4 seconds. This is used to conclude that multi-experiment identifiable combinations with the bound of 1 are the parameters themselves. If we turn off such bypass, the multi-experiment identifiable combinations $k_1, k_3, k_5, k_6, \frac{-k_2k_4+k_3k_5}{k_2+k_3}, k_2-k_3$ with bound 1 are returned in 433 seconds.

Example 2: Refining Multi-Experiment Identifiability Bound (slow-fast ambiguity in a chemical reaction network)

Consider the following system based on a kinetic reaction $A \xrightarrow{k_1} B \xrightarrow{k_2} C$ from [15] with an extra output equation y_2:

$$\begin{cases} x'_A = -k_1 x_A, \\ x'_B = k_1 x_A - k_2 x_B, \\ x'_C = k_2 x_B, \\ e'_A = e'_C = 0, \\ y_1 = e_A x_A + e_B x_B + e_C x_C,\ y_2 = x_C,\ y_3 = e_A,\ y_4 = e_C. \end{cases}$$

Global identifiability is reported only for initial conditions $x_C(0),\ e_A(0),\ e_C(0)$, while everything else is locally identifiable (with probability p). Upon checking identifiable parameter combinations, we observe single-experiment identifiability for $k_1k_2,\ k_1+k_2$. This implies that the parameters k_1 and k_2 are identifiable up to a permutation, so it is possible to infer the reaction rates from an experiment but not which rate corresponds to which reaction. For parameters $e_B,\ k_1,\ k_2$ we obtain multi-experiment identifiability with bound of 3 experiments. We then try to refine the bound with default number of refining attempts being 4. As a result, the new bound for the number of experiments is 2. This is done via different orderings of outputs y_1, y_2 in the procedure.

To illustrate this point in another way, we can tell SIAN to consider multiple copies of the system. We observed that the refined bound for parameters e_B, k_1, k_2 was 2. Setting number of copies to 2, SIAN yields global identifiability of e_B, k_1, k_2, supporting the observation above. Without using the bypass option in the search for combinations, we observe that the application still returns e_B, k_1, k_2 as identifiable with more than 1 experiments, however, single-experiment check overwrites this result, making the bound 1.

4 Discussion and future directions

The structural identifiability toolbox we present here is available online for efficient identifiability analysis of ODE systems. It is capable of fast assessment of individual parameter identifiability properties in Monte Carlo fashion with a user-specified correctness probability as well as assessing deterministically single- and multi-experiment identifiable functions of parameters. By taking advantage of the output information from underlying algorithms of the application, we increase the efficiency of the application via bypassing some time-consuming computations (in some of such bypasses, deterministic result is replaced with a result correct with user-specified probability). However, some computational bottlenecks still need to be resolved:

1. At the time of input-output equation calculation, it is of interest to investigate potential improvements coming from the variable ordering in the Rosenfeld-Gröbner procedure.

2. Field intersection procedure relies on prime decomposition of polynomial ideals [12, Algorithm 2] and can be computationally expensive. We ask if this can be replaced by, for instance, regular chains.

3. To take advantage of parallel computing and absence of 4 GB memory limits of `DifferentialAlgebra` in Maple, one could use Differential Thomas decomposition [6] in place of Rosenfeld-Gröbner for characteristic sets. This is currently a work in progress.

Acknowledgements

This work was partially supported by the NSF grants CCF-1564132, CCF-1563942, DMS-1760448, DMS-1853650, and DMS-1853482, and by the Paris Ile-de-France Region.

References

[1] G. Bellu, M. P. Saccomani, S. Audoly, and L. D'Angiò. "DAISY: A new software tool to test global identifiability of biological and physiological systems". In: *Computer methods and programs in biomedicine* 88.1 (2007), pp. 52–61.

[2] F. Boulier. *The BLAD libraries*. http://cristal.univ-lille.fr/ boulier/BLAD/. 2004.

[3] F. Boulier, D. Lazard, F. Ollivier, and M. Petitot. "Computing representations for radicals of finitely generated differential ideals". In: *Applicable Algebra in Engineering, Communication and Computing* 20.1 (2009), p. 73.

[4] O.-T. Chiş, J. R. Banga, and E. Balsa-Canto. "Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods". In: *PLoS ONE* 6.11 (2011), e27755.

[5] C. Conradi and A. Shiu. "Dynamics of Posttranslational Modification Systems: Recent Progress and Future Directions". In: *Biophysical Journal* 114.3 (2018), pp. 507–515.

[6] V. P. Gerdt, M. Lange-Hegermann, and D. Robertz. "The MAPLE package TDDS for computing Thomas decompositions of systems of nonlinear PDEs". In: *Computer Physics Communications* 234 (2019), pp. 202–215.

[7] H. Hong, A. Ovchinnikov, G. Pogudin, and C. Yap. "SIAN: software for structural identifiability analysis of ODE models". In: *Bioinformatics* 35.16 (2019), pp. 2873–2874.

[8] H. Hong, A. Ovchinnikov, G. Pogudin, and C. Yap. "Global identifiability of differential models". In: *Communications on Pure and Applied Mathematics* 73.9 (2020), pp. 1831–1879.

[9] A. Kalami Yazdi, M. Nadjafikhah, and J. Distefano III. "COMBOS2: an algorithm to the input–output equations of dynamic biosystems via Gaussian elimination". In: *Journal of Taibah University for Science* 14.1 (2020), pp. 896–907.

[10] T. S. Ligon et al. "GenSSI 2.0: multi-experiment structural identifiability analysis of SBML models". In: *Bioinformatics* 34.8 (2018), pp. 1421–1423.

[11] 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". In: *PLoS One* 9.10 (2014), e110261.

[12] A. Ovchinnikov, A. Pillay, G. Pogudin, and T. Scanlon. "Computing all identifiable functions for ODE models". In: *arXiv preprint arXiv:2004.07774* (2020).

[13] A. Ovchinnikov, A. Pillay, G. Pogudin, and T. Scanlon. "Multi-experiment parameter identifiability of ODEs and model theory". In: *arXiv preprint arXiv:2011.10868* (2020).

[14] M. P. Saccomani, G. Bellu, S. Audoly, and L. D'Angiò. "A new version of DAISY to test structural identifiability of biological models". In: *International conference on computational methods in systems biology*. Springer. 2019, pp. 329–334.

[15] S. Vajda and H. Rabitz. "Identifiability and distinguishability of first-order reaction systems". In: *The Journal of Physical Chemistry* 92.3 (1988), pp. 701–707.

[16] A. F. Villaverde, A. Barreiro, and A. Papachristodoulou. "Structural identifiability of dynamic systems biology models". In: *PLoS computational biology* 12.10 (2016), e1005153.