IDENTIFIABILITY PROBLEMS IN SYSTEMS BIOLOGY

organized by
Marisa Eisenberg and Nicolette Meshkat

Workshop Summary

Overview

The focus of this workshop was on identifiability methods and problems, particularly for biological systems. Identifiability is the problem of determining which unknown parameters of a mathematical model can be determined from known input-output data (i.e. from particular variables which are measured/observed). In the structural identifiability problem, data is assumed to be perfect, i.e. noise-free and of any time duration required. Structural identifiability is a necessary condition for the numerical or "practical" identifiability problem, which is the parameter estimation problem for real and often noisy data. A lack of appreciation of parameter identifiability and uncertainty has been pointed to on numerous occasions as hindering the progress of mathematical modeling in biology.

In this workshop, we brought together a wide range of researchers, including experts in different identifiability methods and their application, researchers focused on related methods such as sensitivity analysis and information geometry, and mathematical biology experts facing biological questions that require identifiability analysis. We focused on several important problems that addressed both the theoretical and application-driven sides of identifibaility in mathematical modelling. We held a moderated problem generation session to identify important questions for the field of identifiability, worked in small groups on several of these problems, and finally held a perspectives and future directions session to discuss the future of the identifiability community and identify several key conjectures, problems, and open questions in the area of identifiability analysis of dynamic systems models. Below we describe some of the outcomes of this workshop, which include progress on several open problems, new collaborations, and the beginning of a review paper collecting the ideas and perspectives presented at the workshop and sharing them with the broader identifiability community.

Presentations

We had presentations from 13 speakers. Here is the summary of what was presented:

Marisa Eisenberg: Introduction to identifiability and tutorial on numerical identifiability analysis

Nicolette Meshkat: Structural identifiability and open problems

Julio Banga: Mathematical modelling and experimental design

Alejandro Villaverde: Observability-identifiability rank condition and software demo

Mats Jirstrand: Identifiability using exact arithmetic rank technique and software demo

Ariel Cintron-Arias: Sensitivity analysis and linear difference equation model

Anne Shiu: Chemical reaction networks

Gleb Pogudin: Software demonstration of SIAN

Miranda Teboh-Ewungkem: Malaria models
Harsh Jain: Cancer models
Mark Transtrum: The "sloppiness" problem
Emilie Dufresne: Identifiable scaling reparametrization using scaling symmetries
Andrew Brouwer: Identifiability of multistage clonal expansion (MSCE) models

Afternoon break-out sessions

Identifiability of time-varying parameters.
This group looked at identifiability of time-varying parameters and observation of unknown inputs. They subdivided identifiability into four categories: (1) global, (2) finite/local, (3) finite-dimensional unidentifiable, and (4) unidentifiable. They determined that Jacobian tests can be used for the last three categories. For the finite-dimensional unidentifiable case, they used sampling and posed the question: can sampling turn, for example, category (3) to category (2)? Their preliminary results showed that knowing the specific form of polynomials can help.

Time-varying inputs vs. multiple constant inputs.
This group looked at the equivalence between identifiability with known time-varying inputs vs. that of multiple, constant inputs. They examined how many experiments were needed to ensure identifiability and determined that a necessary condition is that one must modify the model with as many replicates of x, u, y as the number of experiments. They then posed the question of when this is possible. They examined a case study of the linear 2-compartment model. This was identifiable with a single experiment with time varying input. They concluded that distinct initial conditions across experiments lead to unidentifiability, but if initial conditions are the same across experiments, this leads to identifiability. Likewise, if one has a model that is fully observed, identifiability will be equivalent between time varying inputs and sufficiently many multiple constant inputs. They then tried to generalize in terms of the number of states or inputs and connect this with work from Sontag ($2r + 1$ result). Future work involves exploration of the structure of the model and more input-output mappings.

Identifiability of stochastic models.
This group started by developing definitions of structural identifiability of stochastic models, based on how much information would be available from the data. They broke the problem into three cases: (1) full distribution information, i.e. complete knowledge of the distribution of trajectories, (2) partial information, i.e. dynamics of e.g. means or other analogous parameters, and (3) single information, i.e. observation of one trajectory. They then tested some examples, including chemical reaction networks. They tested examples of extending the input-output equation approach to stochastic models using either probability generating functions or the master equation. They found an example of a 1-compartment model which is unidentifiable as an ODE model, but identifiable when implemented stochastically. The ODE form of the model is given by:

$$\begin{aligned} \dot{x} &= k_1 u(t) - k_2 x \\ y &= k_3 x \end{aligned}$$

This model is structurally unidentifiable (which can be shown e.g. via the differential algebra approach [ljung1994global,audoly2001global]), with identifiable combination $k_1 k_3$. However, when implemented as a discrete, stochastic Markov model (in which case we take $k_3 \in \mathbb{Z}$), one can see that k_3 is identifiable, as all observed values of y will be multiples of k_3 (in particular, the smallest non-zero value of y will be k_3). Thus, the full model will be identifiable in this case (this can also be shown directly using chemical master equations). The group also showed that this model would be identifiable if the trajectories of the mean and variance were observed. The group then examined several other compartmental and chemical reaction network models, to evaluate when their ODE and stochastic implementations differed in identifiability.

Identifiability of difference equations.
This group started by examining the college admissions study from the talk given by Ariel Cintron-Arias, as well as some general models used for modeling movement and growth in ecological systems (e.g. Leslie models). They started with a homogeneous model and applied the differential algebra approach to determine input-output equations (analogously to the ODE case). They then applied it to a non-homogeneous model, which didn't modify the problem much, however the matrix in this problem was nilpotent so the standard input-output equation method wasn't useful for this problem. They then moved to a different method of considering observables. They constructed a system of equations (4 observables, 12 states) and used measurements at time points and solved individually for the parameters.

Hybrid switching systems.
This group examined piecewise linear ODEs, which can be interpreted as discretely many top cells in phase space. These discontinuous ODEs can be obtained as the limiting case of various biological switches, such as Hill functions. In the Hill function case, one can take the exponent to infinity and it becomes steeper, and the limit is a step function. The requirements for identifiability are intuitive, but steep. They determined one needs identifiability of the parameters for every piecewise portion of the ODE system, i.e. on every top cell. One also needs an orbit segment that is transverse to every co-dimension 1 space. They then posed the question: in the case where the switch time is unknown, can we estimate (identify) where the switch is happening? They determined that a likely necessary condition is that one must observe every state variable (or have every state variable be observable). They examined an ODE model with a Hill function if the exponents are unknown. They concluded that even if one observes all the state variables, it is not identifiable and one can't immediately tell where the switch happens.

Permuting inputs and outputs.
This group examined two classes of linear compartmental models, known as mammillary and catenary models. Mammillary models are models with a single central compartment and other compartments branching off in a star pattern. Catenary models have a sequence of compartments in a chain. When there is a single input and single output and leak in the same compartment, these two classes of models are known to be identifiable. This group asked the question: what happens if we permute the input and output compartments? They determined that identifiability holds as long as the input and output compartment are no more than one edge length away. They also proved a combinatorial formula for the coefficients of the input-output equation in the case of single input/output/leak. Additionally,

they examined unidentifiability of strongly connected linear compartmental models and determined sufficient conditions for unidentifiability in terms of the number of parameters and the length of the shortest path from input to output compartments.

Identifiability of $\mathcal{R}_0$*.*

This group focused on understanding the identifiability of $\mathcal{R}_0$, a key bifurcation parameter used in a wide range of infectious disease models. This problem was motivated by the fact that in early epidemic settings (or other settings near the disease free equilibrium), the dynamics of infectious disease systems are given by exponential growth with a rate $\mathcal{R}_0$. This means that from early epidemic data, we would expect to always be able to identify $\mathcal{R}_0$ (at least structurally). The group began by reviewing the next-generation matrix method for calculating $\mathcal{R}_0$ [van2002reproduction], and developing a list of common model and measurement equation types to consider. They then used the next-generation approach to calculate $\mathcal{R}_0$ for a range of examples, and assessed its identifiability using two differential algebra methods (the input-output equation method and the approach of Sedoglavic, 2002 [sedoglavic2002probabilistic]). The group conjectured that sufficient conditions for identifiability of $\mathcal{R}_0$ might be: a constant total population which is either known or consistent with the definition of $\mathcal{R}_0$, with all infectious compartments measured (using a measurement equation based on prevalence, i.e. $y = kI$, where k is a reporting rate parameter), then $\mathcal{R}_0$ is identifiable. The group also examined the FV^{-1} matrix used in the next generation approach (of which the spectral radius is $\mathcal{R}_0$), to see whether the entries, characteristic polynomial, and other features of the matrix are identifiable. They found examples in which the entries of FV^{-1} were not identifiable, but conjectured that the coefficients of the characteristic polynomial may be identifiable when $\mathcal{R}_0$ is identifiable. The group also considered whether bifurcation parameters in general are identifiable (in broader contexts than infectious disease models)—because these parameters control qualitative changes to the system dynamics, one might expect that they would be identifiable if the model outputs were such that these qualitative changes were observed, however further examination is needed.

Proving identifiability for classes of models using different methods.

This group compared the differential algebra approach [ljung1994global, audoly2001global] to the observability-identifiability rank condition [villaverde2016structural] for testing structural identifiability of classes of models. In general, the rank condition often has significant speed advantages to the differential algebra approach when used for evaluating specific models, however the differential algebra approach is more often used to prove identifiability for general classes of models. The differential algebra approach relies on forming an input-output equation, which can be cumbersome for large models, so this group set out to try the observability-identifiability rank condition on a simple linear 2-compartment model to see if the approach proved fruitful. They concluded that, even for a simple linear 2-compartment model with 2 states and 3 parameters, the 5 by 5 matrix that results from the observability-identifiability rank condition is not easier to work with than the 3 by 3 Jacobian of the coefficient mapping obtained from the input-output equations using the differential algebra approach. Their conclusion was that it may be possible to use the observability-identifiability rank condition to prove identifiability results for adding a parameter to an identifiable model to see if the resulting model is still identifiable, but that using it to prove identifiability results for general classes of models did not seem fruitful compared to the differential algebra approach.

Implicit initial conditions.
This group investigated the implications of implicit initial conditions given by the steady state equations on the relations between derivatives of outputs and states/parameters (i.e., the Lie-derivatives). The solvability of these relations is what lies behind the rank criterion on the observability matrix (via the inverse function theorem). Steady state (with $u = 0$) implies $f_i = 0$ and hence the Lie derivatives along f (at $f = 0$) only consist of the "u-terms":

$$L_f = \sum_{i=0}^{\infty} u(i+1) \frac{\partial}{\partial u(i)}.$$

This group looked at some small examples to see what implications this could have on the rank conditions.

Collaborations started/Future work

Several groups have started collaborations that will continue beyond this workshop and we expect several papers to ensue. We also discussed starting work on a review paper which collects the ideas and perspectives presented at the workshop and shares them with the broader identifiability community.

Because the workshop was such a success, the group has decided to make this workshop the first meeting in a newly-formed biennial meeting on identifiability.

Bibliography

[ljung1994global] Ljung L, Glad T. On global identifiability for arbitrary model parametrizations. Automatica. 1994;30(2):265–276.

[audoly2001global] Audoly S, Bellu G, D'Angio L, Saccomani MP, Cobelli C. Global identifiability of nonlinear models of biological systems. IEEE Transactions on biomedical engineering. 2001;48(1):55–65.

[van2002reproduction] Van den Driessche P, Watmough J. Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission. Mathematical biosciences. 2002;180(1-2):29–48.

[sedoglavic2002probabilistic] Sedoglavic A. A probabilistic algorithm to test local algebraic observability in polynomial time. Journal of Symbolic Computation. 2002;33(5):735–755.

[villaverde2016structural] Villaverde AF, Barreiro A, Papachristodoulou A. Structural identifiability of dynamic systems biology models. PLoS computational biology. 2016;12(10):e1005153.