ODEbase: an extensible database providing algebraic properties of dynamical systems

Christoph Lüders, Hassan Errami, Matthias Neidhardt,
Satya S. Samal and Andreas Weber

University of Bonn

CASC 2019, August 27th, Moscow

SYMBIONT PROJECT

Overview

1. What is it good for?
2. Searching & properties
3. Generation of ODEs and properties
4. Classification of ODEs
5. Implementation
6. Live (!) demonstration

Mission statement

Web page: http://odebase.cs.uni-bonn.de

ODEbase ...

- is an extensible database for algebraic properties of ODEs,
- allows storing and retrieval of parameterized ODEs and differential-algebraic equations,
- allows storing of arbitrary properties of these ODEs,
- is publicly available on the Internet,
- can be searched, sorted, filtered and downloaded by the user.

It is a database for the community, and especially the SYMBIONT project, to present results of their computations of ODE systems.

Furthermore, it is a repository when models are needed for benchmarking and testing or formulation of hypotheses (experimental mathematics).

Data sources

As the name suggests, ODEbase can store any ODE system.

- Currently, all models in ODEbase come from the BioModels database (Le Novère et al., 2006) and are chemical reaction networks (CRNs).
- Other data sources are available and we are planning to incorporate them as well, e.g., KEGG and MetaCyc.
- There is no limitation to CRNs. Possible systems include:
  - models of neurons,
  - epidemic models,
  - others, e.g., models from physics are possible to store as well.

Filtering and sorting

The database can be accessed through a Web interface. To ease access to models and their properties, filters and sorting are available.

- Models can be filtered for a variety of criteria:
  - Numerical values (may be filtered as just one value ("a"), from-to ("a–b"), minimum value ("a–") and maximum value ("–a")):
    - dimension, i.e., number of variables,
    - number of reactions,
    - rank of the stoichiometric matrix,
    - deficiency.
  - Three-valued fields (yes/no/don't care): rational, polynomial, mass-action, Gröbner basis availability
- Models can be sorted by these properties as well.
- Additional filters are easy to implement.

Which properties?

For the currently saved CRNs, the following properties are computed and stored:

- vector field,
- stoichiometric matrix,
- kinetic matrix (Gatermann et al., 2005),
- linear conservation laws (Schuster and Höfer, 1991),
- standard parameters,
- mapping of variable & parameter names to names used in the model,
- deficiency (Feinberg, 1995, § 2),
- Web link to the model source,
- Web links to additional information, e.g., papers about the model,
- Gröbner basis of the vector field.

Which properties are still coming?

We're currently working to incorporate these information:

- tropical prevariety, computed with PtCut,
- numerical solutions for fixpoints, computed with Bertini and PHCpack,
- toricity of systems,
- more SYMBIONT results, e.g., necessary conditions for multi-stationarity based on graph-theoretic methods.

SBML Parser: SBMLode

Models from the BioModels database are read and their ODEs and other data are extracted and saved. This is done with SBMLode, a parser for SBML.

- The BioModels database contains descriptions in SBML, an XML-style description language. It is used to describe species, chemical reactions, stoichiometry, and reaction kinetics. It can be used to simulate a chemical reaction network.
- To get the ODEs from a BioModel, we use the reaction kinetics and the stoichiometry.
- SBMLode computes linear conservation laws.
- For some systems no ODEs can be extracted.

Generation of the ODEs

- Read parameters, compartments, rules and reactions.
- For each reaction, read stoichiometry, kinetic law and its parameters.
- Build stoichiometric matrix S and check for mass-action kinetics.
- For each species i, check boundary condition and if row S_i is zero. If either is true, it doesn't interact with the CRN.

  Otherwise, add for each reaction j a term $S_{ij} \cdot \ell$, where ℓ is the reactions' kinetic law. Divide the whole sum by the compartment size. This is the rate of change for species i.

  The rate of change is "computed" just by string processing to keep the term structure unchanged.
- Compute conservation laws, deficiency and rank of S.

What doesn't work

SBML allows for complex modeling, this can cause problems. Currently (25-Jul-2019), there are 735 curated models in the BioModels database.

At the present time, SBMLode can *not* handle:

- function definitions (can be supported?): 201 models,
- rate rules (might be supported at a later time): 137 models,
- time dependency: 48 models,
- use of `delay()` or `piecewise()`: 36 models,
- model could not be converted within 1h: 7 models,
- various problems (can likely be fixed): 23 models.

In effect: ODEs are stored for 323 models.

Floating point problems

- libSBML is used to access SBML data. It returns numerical values in stoichiometry, parameters, compartment sizes and initial values as IEEE floats.
- Since SBML is a text format, values are denoted in base-10. Yet, today's computers are base-2. This leads to *representation error*. E.g., `0.1` can not be represented exactly, whereas `0.5` can.
- To avoid that, SBMLode uses XML to query values as strings and represents them as exact fractions.
- Kinetic laws can contain float literals. When used in computations, this can lead to *rounding errors*. E.g., (from BM 161):
  `0.0016611295681063 1 * 0.0016611295681063 1 =`
  `0.000027593514420370556`, instead of the correct
  `0.00002759351442037055927594618161`.
  Float literals are replaced by fractions as well.

Classification of ODE systems

ODEbase classifies systems in several classes:

- rational: terms are of the form $\sum_i P_i/Q_i$, where P_i, Q_i are polynomials,
- polynomial: terms are polynomials,
- mass-action: terms are polynomials and are of the form $k\boldsymbol{x}^{S_{\cdot j}}$, where $S_{\cdot j}$ is column j of the stoichiometric matrix and k is some constant.

The classes are proper supersets of one another:
rational $\supsetneq$ polynomial $\supsetneq$ mass-action.

- 232 models are rational.
- 137 models are polynomial.
- Only 38 models have mass-action kinetics!

Classification as mass-action kinetics

According to (Feinberg, 1995, § 3) a *mass-action reaction* j has a rate function of the form

$$r_j = k \prod_i x_i^{S_{ij}},$$

with k the rate constant, x_i the concentration of reactant species i, and S the stoichiometric matrix.

SBML allows for reactions to be reversible or one-way, but only offers a single rate function. A reversible reaction is split into its positive, resp. negative monomials for the forward, resp. backward reaction.

To determine if a rate function has mass-action kinetics, its monomials are divided by $\prod_i x_i^{S_{ij}}$ (for each direction if reversible).

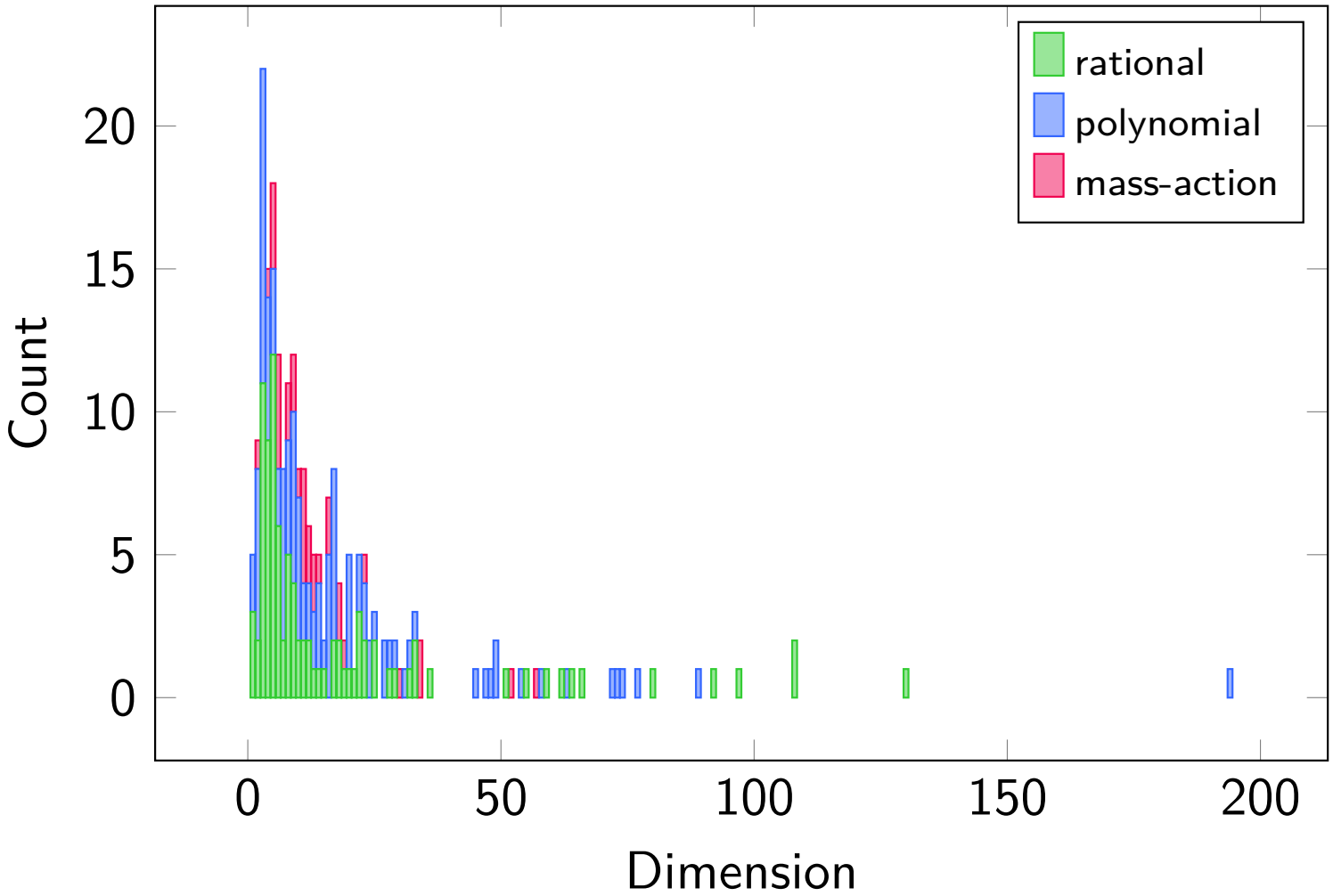
If the result after division and cancellation contains only parameters and, in case of a reversible reaction, the analog is true for the backward reaction, the reaction is considered to have mass-action kinetics.

Gröbner bases

Gröbner bases are a useful tool in algebraic geometry. Yet, they can be prohibitively hard to compute. Thus, it makes sense to save them once the work has been done.

- Thomas Sturm has computed Gröbner bases using Maple for 143 models.
- Gröbner bases can be downloaded, together with their term order.
- Gröbner bases for more systems will follow.
- Other toricity tests will follow as well.
  See the talk of Grigoriev, Sturm and Weber on Friday.

Distribution of dimension



Total of 323 models, of which 232 models are rational, of which 137 models are polynomial, of which 38 models have mass-action kinetics.

Support for various computer algebra systems

- ODEbase is designed to return its output file in formats that are readily understood by various computer algebra systems.
- Currently, SageMath and Maple are supported, other formats will follow.
- Output is generated on-the-go through string processing. Thus, it's easy to extend to new formats.

Miscellaneous:

- All data on each model can be downloaded in one .zip file, together with the original model specification in SBML.
- The URLs to all downloads will stay fixed and allow specification of model, property and format, like `http://odebase.cs.uni-bonn.de/ODEModelApp/model/odes/BIOMD0000000001/maple`.

Live demo

Here be dragons.

Implementation

- SBMLode is written in Python 3.5+ and has about 2000 lines of code. The source code will be freely available under LGPL v3. It requires libSBML, which is freely available as well.
- The Web server is written in Python 3 with Django, a Python Web framework. The source code is 450 lines of code.
- The database used is PostgreSQL, a free and open-source relational database management system.
- All software used is free and runs on both Windows and Linux.

Outlook

Extend ODEbase into multiple, orthogonal dimensions:

- more data sources,
- more properties that are computed,
- more support for SBML features,
- more CAS formats,
- improved user interface,
- more additional links to articles.

Suggestions are welcome!

End credits

Thank you for your attention!

ODEbase: http://odebase.cs.uni-bonn.de

SYMBIONT: https://www.symbiont-project.org

Special thanks to Thomas Sturm.

References:

M. Feinberg. The existence and uniqueness of steady states for a class of chemical reaction networks. *Archive for Rational Mechanics and Analysis*, 132(4):311–370, 1995. doi: 10.1007/bf00375614.

K. Gatermann, M. Eiswirth, and A. Sensse. Toric ideals and graph theory to analyze hopf bifurcations in mass action systems. *Journal of Symbolic Computation*, 40(6):1361–1382, dec 2005. doi: 10.1016/j.jsc.2005.07.002.

N. Le Novère, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, J. L. Snoep, and M. Hucka. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research*, 34(Database issue):D689–D691, Jan 2006.

S. Schuster and T. Höfer. Determining all extreme semi-positive conservation relations in chemical reaction systems: a test criterion for conservativity. *J. Chem. Soc., Faraday Trans.*, 87(16):2561–2566, 1991. doi: 10.1039/ft9918702561.