Noname manuscript No.
(will be inserted by the editor)

On the role of algebra in models in molecular biology

Elisenda Feliu

Received: date / Accepted: date

Classical reference books in mathematical biology, e.g. [7], illustrate how, in its origins, this once emerging field essentially relied on small models, and their thorough analysis employed a suit of advanced techniques from dynamical systems theory. It was often feasible to pursue such analysis without fixing the value of model parameters, thereby obtaining a full picture of the possible behaviors of the model. In many cases, the purpose of the model was to provide a qualitative understanding of a phenomenon, which not necessarily needed to fit exactly with observational data.

As models have become larger and more complex, and with the increasing availability of data, in particular in molecular biology, a standard approach to analyze mathematical models has been to first gain some insight about suitable parameter values, for example via estimation or extrapolating from related species, and then employ numerical methods to simulate the models. In this way, a precise description of the system of interest could be obtained. A problem arises when parameters are unidentifiable, or cannot be determined with the desired precision, or when we need to take into account that parameter values typically fluctuate, are specific to the individual, and depend on the environment. Then we are back to the original problem of understanding the model in a larger region of the parameter space. As the complexity of the models forbid detailed hands-on analyses, model inspection is often achieved through a combination of parameter sampling and numerical simulation.

Parallel to this development, some theories have centered around systems of ordinary differential equations that model the concentration of species in an interaction network in time. Although these models are typically associated with chemical and biochemical reactions, the formalism fits as well models in

EF acknowledges funding from the Independent Research Fund of Denmark.

Elisenda Feliu
Department of Mathematical Sciences, University of Copenhagen.
Universitetsparken 5, 2100 Copenhagen, Denmark
E-mail: efeliu@math.ku.dk

ecology, like the Lotka-Volterra model, or in epidemiology: All these have in common that the interactions among entities drive the changes of the system.

The origin of these theories goes mainly back to the 70'ies and 80'ies, with, to name a few, the work of Feinberg, Horn and Jackson leading to what is known as Chemical Reaction Network Theory (CRNT) [6]; Vol'pert [8]; and Clarke, leading to Stoichiometric Network Analysis [2]. Common to these theories is the search for easy-to-apply methods concerning dynamics, by relying on the structure of the interaction network and assumptions about the rates of the interactions. This has lead to simple (but powerful) theorems on number of steady states and their stability, to give some examples.

These theories are at one end of the spectrum of the level of abstraction of mathematical models. At the left end of the spectrum we find models that are fitted to real data and provide detailed quantitative information of a specific system under study; and at the right end of the spectrum, we find general theories aimed at studying classes of models that share some particularities and their qualitative properties. Moving from left to right we go from models whose goal is to represent reality in detail, to models seeking to identify underlying principles. Although at first sight one might think that the left region of the spectrum is the one that really matters in practical scenarios, a throughout qualitative analysis of families of models can be valuable to guide experimental design, to support conclusions of fitted models, and can be helpful in synthetic biology. Furthermore, it is advantageous, and at the core of mathematics, to rely on general theories when studying specific models.

Algebra and Interaction Networks

In recent years, these old theories about interaction networks from the 70'ies, mainly CRNT, have been revised and further developed under the umbrella of computational algebra and algebraic geometry. The reason behind this, is that models arising from interaction networks typically involve polynomials and rational functions (quotients of polynomials). Prominent examples are the mass-action assumption, yielding polynomial differential equations, or Michaelis-Menten type kinetics, yielding models with rational functions. In this case, the steady states of a model are the solutions to a system of polynomial equations, which is the object of computational algebra and algebraic geometry. Furthermore, computational algebra is well suited to systems with unspecified parameters, after choosing the right coefficient field. It can for example find relations that hold for all parameter values at steady state, or find descriptions of the steady states by means of a simple parametrization. However, two main drawbacks prevent these methods to stand out: the high computational cost, and the fact that the restriction of the steady states to positive values causes nice results from the theory of polynomial equations to fail. For example, any real polynomial of degree n has exactly n complex roots counted with multiplicity, but only some generic upper bounds can be given for the number of real and positive real roots.

Progress within this area has focused on solving these challenges by exploiting the fact that the polynomials under study arise from interaction net-

works. Through a close interplay with real algebraic geometry, this had lead to numerous strategies to count the number of positive steady states and even understand the parameter space in that respect [3]. More recently, similar ideas are being applied to study stability and bifurcations, as these, via the Routh-Hurwitz criterion, are also expressed in algebraic terms. In general, whenever the question of interest can be reduced to understanding the solutions to a system of polynomial equalities and inequalities, then computational algebra might well be the right theory to call.

Integrating the whole spectrum

Mathematical biology, and applied mathematics in general, is witnessing how theory traditionally belonging to the realm of pure mathematics is finding its place in the study of mathematical models. This certainly applies to algebraic geometry, but also to other disciplines like topology. Despite the broad range of existing theories to analyze mathematical models in molecular biology from different perspectives, the preferred choice often involves numerical simulations combined with parameter inference or parameter sampling. This is presumably driven by the numerous existing tools that address this end of the spectrum, e.g. [1], while we lack proper dissemination and computational tools that cover the rest of the spectrum, and facilitate the access to users without a suitable mathematical background. With few exceptions [4, 5], the latter is partially a consequence of the notable challenges involved with providing easy-to-use black-box implementations of that end of the spectrum. But without these, much of the valuable theory currently being developed to analyze families of models at once, will remain a curiosity and its potential use in real applications will be overlooked.

As methods, tools and theories are constantly being developed to understand the overwhelmingly-complex systems of interacting elements, it would be desirable to have integrative platforms where users, these being experimental biologists or theoreticians, can dissect models at all possible levels.

References

1. COPASI: Biochemical System Simulator. URL http://copasi.org/
2. Clarke, B.L.: Stability of Complex Reaction Networks, *Advances in Chemical Physics*, vol. 43. John Wiley & Sons, Inc., Hoboken, NJ, USA (1980)
3. Dickenstein, A.: Biochemical reaction networks: an invitation for algebraic geometers. In: Mathematical Congress of the Americas, *Contemp. Math.*, vol. 656, pp. 65–83. Amer. Math. Soc., Providence, RI (2016). DOI 10.1090/conm/656/13076
4. Donnell, P., Banaji, M., Marginean, A., Pantea, C.: CoNtRol: an open source framework for the analysis of chemical reaction networks. Bioinformatics **30**(11), 1633–1634 (2014)
5. Ellison, P., Feinberg, M., Ji, H., Knight, D.: Chemical reaction network toolbox, version 2.2. Available online at http://www.crnt.osu.edu/CRNTWin (2012)
6. Feinberg, M.: Foundations of Chemical Reaction Network Theory, *Applied Mathematical Sciences*, vol. 202. Springer International Publishing (2019). DOI 10.1007/978-3-030-03858-8
7. Murray, J.D.: Mathematical Biology: I. An introduction, *Interdisciplinary Applied Mathematics*, vol. 17, third edn. Springer (2002)
8. Vol'pert, A.I., Hudjaev, S.I.: Analysis in classes of discontinuous functions and equations of mathematical physics, *Mechanics: Analysis*, vol. 8. Springer Netherlands (1985)