What is Algebraic Biology?

Matthew Macauley

Department of Mathematical Sciences
Clemson University
http://www.math.clemson.edu/~macaule/

Algebraic Biology

A brief history of mathematical biology

- **1202**: Leonardo of Pisa models a population of rabbits with a numerical sequence in his book *Liber Abaci*.
- **1760**: Daniel Bernoulli modeled the spread of smallpox with differential equations.
- **1798**: Thomas Malthus used exponential growth models in *An Essay on the Principle of Population*.
- **1836**: Pierre Francois Verhulst developed the logistic growth model.
- **1879**: First use of mathematics in evolutionary ecology, by Fritz Müller
- **1901**: First use of the term *theoretical biology*, by Johannes Reinke.
- **1926**: Lotka-Volterra predator-prey models published.
- **1927**: SIR model of infectious disease spread proposed by Kermack and McKendrick.
- **1953**: Discovery of the structure of DNA by Watson, Crick, and Franklin.
- **1990**: Human genome prodject launched.
- **1995**: First paper publised using the term *microarray*.
- **2003**: Human genome prodject declared complete.
- **2022**: Last gapless assembly of the human genome complete.

The future of mathematical biology

PLOS BIOLOGY

PLoS Biol. 2004 Dec; 2(12): e439.
Published online 2004 Dec 14. doi: 10.1371/journal.pbio.0020439

PMCID: PMC535574
PMID: 15597117

Mathematics Is Biology's Next Microscope, Only Better; Biology Is Mathematics' Next Physics, Only Better

Joel E Cohen

Can Biology Lead to New Theorems?

Feature article

CAN BIOLOGY LEAD TO NEW THEOREMS?

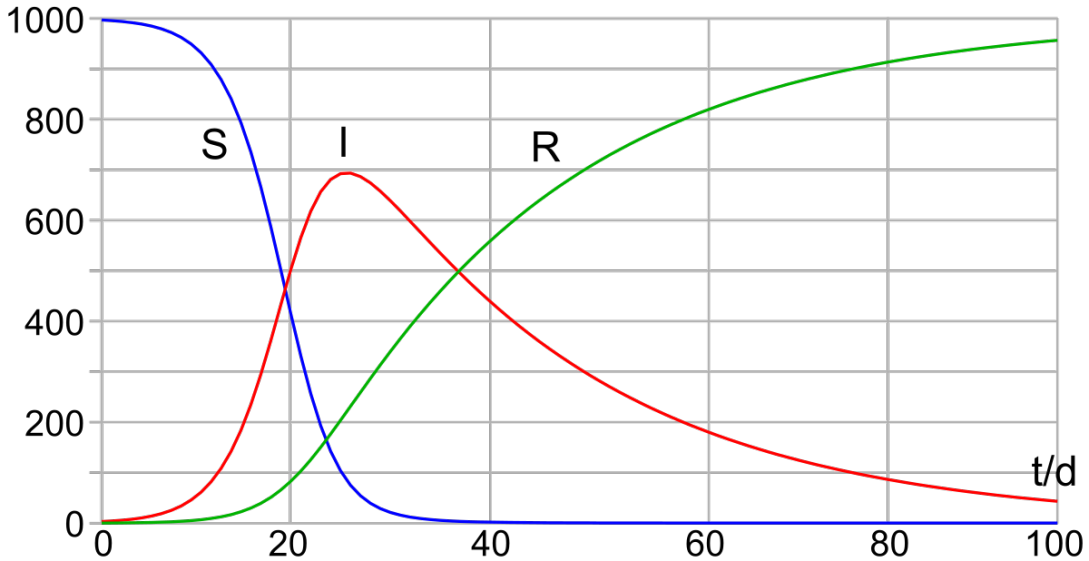
BERND STURMFELS

ABSTRACT. This article argues for an affirmative answer to the question in the title. In future interactions between mathematics and biology, both fields will contribute to each other, and, in particular, research in the life sciences will inspire new theorems in "pure" mathematics. This point is illustrated by a snapshot of four recent contributions from biology to geometry, combinatorics and algebra.

What does algebra have to do with biology?

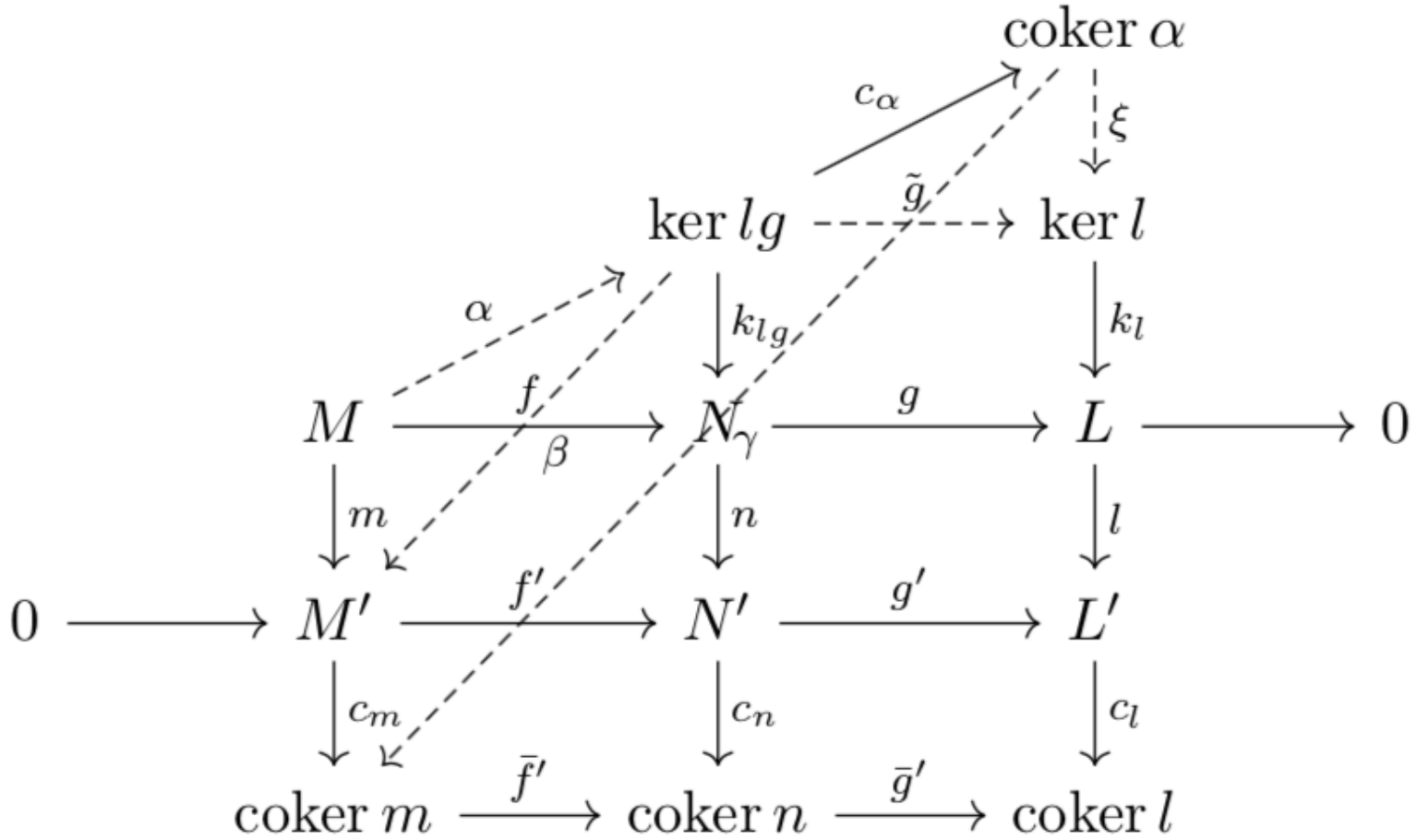
Usually, when we think of mathematical biology, we think of models such as this:

$$\frac{dS}{dt} = -\alpha SI$$
$$\frac{dI}{dt} = \alpha SI - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$



What does algebra have to do with biology?

Whereas algebra might remind us more of this:

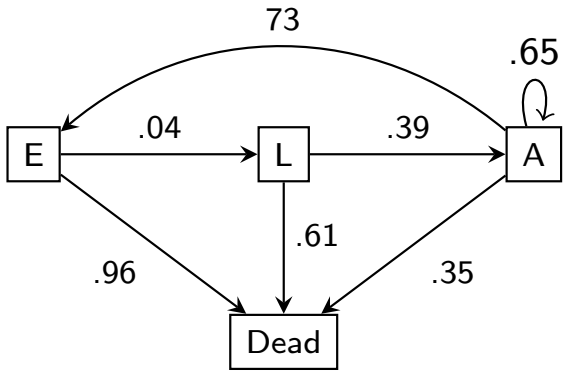


How could these two topics possibly be related??

What does algebra have to do with biology?

Linear algebra is fundamental to mathematical biology.

- Population (Leslie) matrix:



E = Egg, L = Larvae, A = Adult

$$\begin{bmatrix} 0 & 0 & 73 \\ .04 & 0 & 0 \\ .65 & .39 & 0 \end{bmatrix} \begin{bmatrix} E_t \\ L_t \\ A_t \end{bmatrix} = \begin{bmatrix} E_{t+1} \\ L_{t+1} \\ A_{t+1} \end{bmatrix}.$$

- Competing species (Lotka-Volterra equations):

$$\begin{aligned} P' &= P(1 - P - Q) \\ Q' &= Q(.75 - Q - .5P) \end{aligned} \qquad \begin{bmatrix} X' \\ Y' \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 0 & .25 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} - \begin{bmatrix} X^2 + XY \\ .5XY + Y^2 \end{bmatrix},$$

via changing variables $(X, Y) = (P - 1, Q)$.

What does algebra have to do with biology?

Linear algebra, the study of linear polynomials and their solutions.

Analyzing nonlinear polynomials and their solutions is much more complex.

It involves fields such as algebraic geometry and computational algebra.

Though these themes are not as ubiquitous in biology as linear algebra is, they arise in a number of biological problems.

Algebraic Biology is the subfield that encompasses these problems, and the new mathematics that they spawn.

Examples of where biological problems where nonlinear algebra arises:

1. Biochemical reaction networks
2. Algebraic statistics and data science
3. Place fields in neuroscience
4. Boolean models of molecular networks

I'll also discuss some new (pure) mathematics that has arisen from these biological problems.

Linear algebra vs. nonlinear algebra

A vector space is a:

- set V of vectors (e.g., $\mathbb{R}^n$)
- field K of scalars (e.g., $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$)

that is closed under addition, subtraction, and scalar multiplication of vectors.

Many concepts in nonlinear algebra have simple linear algebra analogues. For example,

- the subspace of V spanned by $v_1, \ldots, v_k$ is the set

$$\text{Span}(v_1, \ldots, v_k) = \{a_1 v_1 + \cdots + a_k v_k \mid a_i \in K\}.$$

- the ideal of $R = \mathbb{F}[x]$ generated by polynomials $f_1, \ldots, f_k$ is the set

$$\langle f_1, \ldots, f_k \rangle = \{a_1(x) f_1(x) + \cdots + a_k(x) f_k(x) \mid a_i(x) \in R\}.$$

nonlinear algebra concept	linear algebra concept
polynomial ring $R = K[x_1, \ldots, x_n]$	vector space V
ideal $I \leqslant R$	subspace $W \leqslant V$
Gröbner basis $\mathcal{G}$	"nice" vector space basis $\mathcal{B}$
algebraic variety	solution space

Some general resources

Books

- U. Alon. *An introduction to systems biology: design principles of biological circuits*. 2nd edition. CRC press, 2019.
- D. Cox. Applications of polynomial systems. 2020.
- H. A. Harrington, M. Omar, and M. Wright. *Algebraic and Geometric Methods in Discrete Mathematics*, volume 685. American Mathematical Society, 2017.
- N. Jonoska and M. Saito. *Discrete and Topological Models in Molecular Biology*. Springer, 2013.
- R. Robeva. *Algebraic and Discrete Mathematical Methods for Modern Biology*. Elsevier, 2015.
- R. Robeva and T. Hodge. *Mathematical Concepts and Methods in Modern Biology: Using Modern Discrete Models*. Academic Press, 2013.
- R. Robeva and M. Macauley. *Algebraic and Combinatorial Computational Biology*. Elsevier, 2018.

Articles

- R. Laubenbacher and B. Sturmfels. Computer algebra in systems biology. *Amer. Math. Monthly*, pages 882–891, 2009.
- M. Macauley and R. Robeva. Algebraic models, pseudomonomials, and inverse problems in algebraic biology. *Lett. Biomath.*, 7(1):81–104, 2020.
- M. Macauley and N. Youngs. The case for algebraic biology: from research to education. *Bull. Math. Biol.*, 82(115), 2020.
- B. Sturmfels. Can biology lead to new theorems? *Annual report of the Clay Mathematics Institute*, pages 13–26, 2005.

1. Biochemical reaction networks

Karin Gatermann (1961–2005) introduced computational algebra tools to reaction networks.

Assuming the law of mass-action kinetics,

$$A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C, \qquad A \xrightarrow{k_3} 2B,$$

leads to the following system of ODEs:

$$\begin{aligned} x_1' &= -k_1 x_1 x_2 - k_3 x_1 + k_2 x_3 \\ x_2' &= -k_1 x_1 x_2 + k_2 x_3 + 2k_3 x_1 \\ x_3' &= k_1 x_1 x_2 - k_2 x_3. \end{aligned}$$

The steady-state solutions (set each $x_i' = 0$) are an algebraic variety.

Gatermann introduced complex-balancing (toric) dynamical systems. Steady-states are toric varieties.

Global attractor conjecture

For a complex-balanced system, each equilibria $c \in \mathbb{R}^N_{>0}$ is globally asymptotically stable relative to the interior of its compatibility class.

1. Biochemical reaction networks

Research goals and open-ended questions

Key idea

What does the polynomial algebra tells us about the dynamics of the ODEs?

Persistence conjecture (Feinberg, 1987)

Every weakly reversible mass-action kinetics ODE is persistent, regardless of the rate constants.

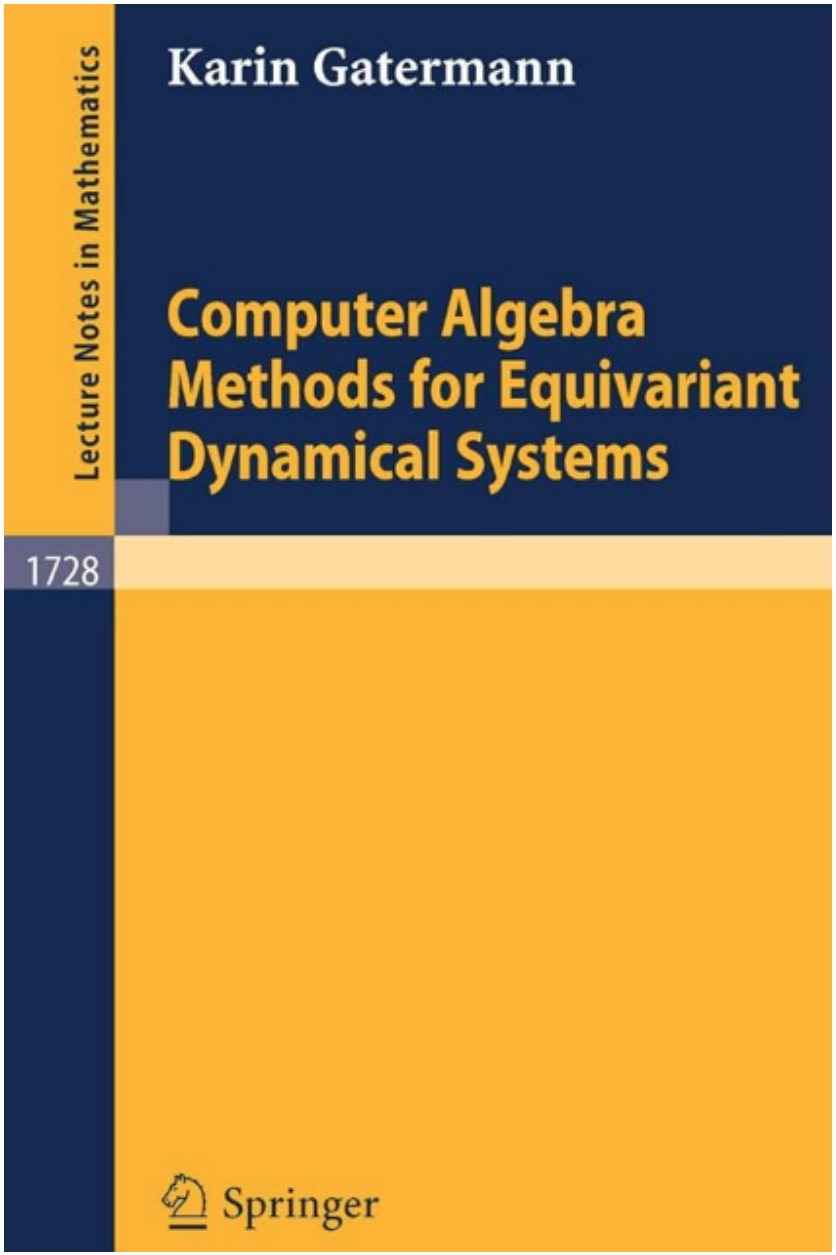
Permanence conjecture (stronger)

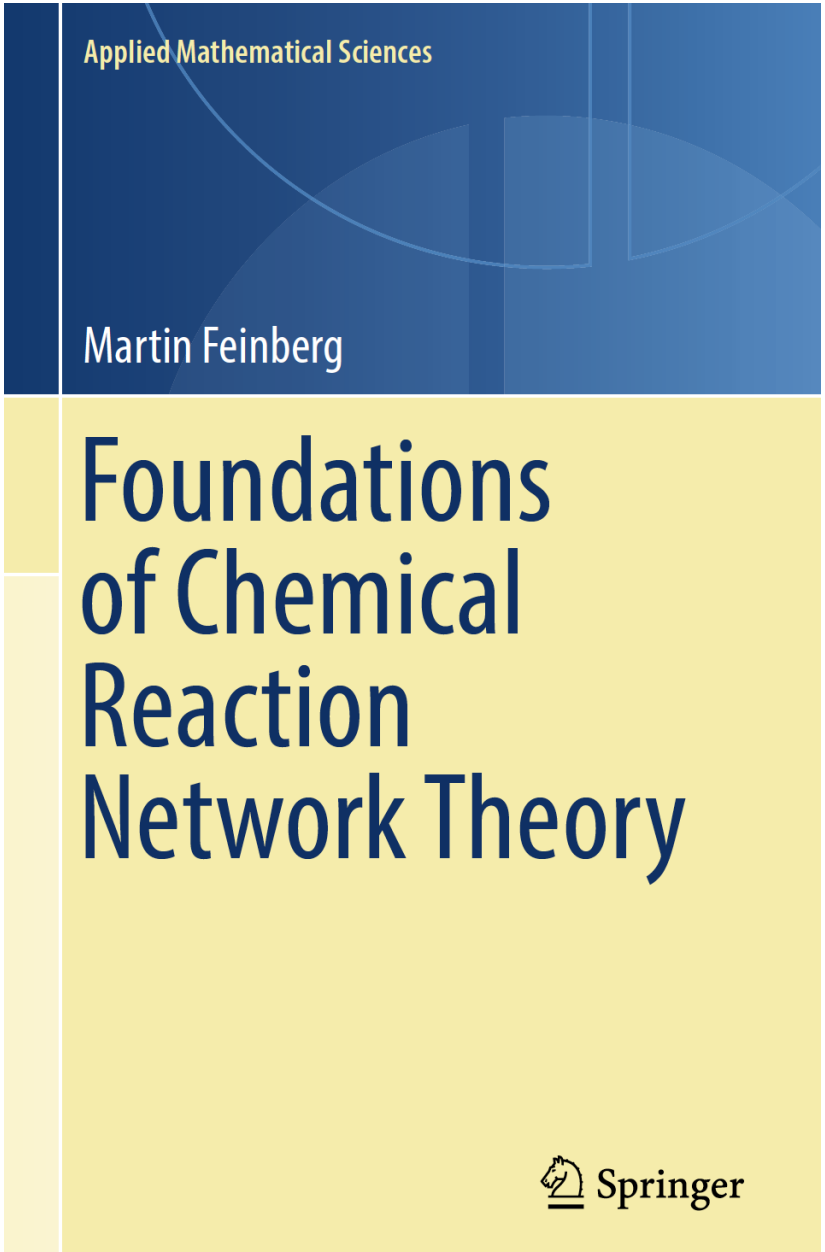
Every endotactic reaction network is permanent, regardless of rate constants.

Global attractor conjecture (weaker)

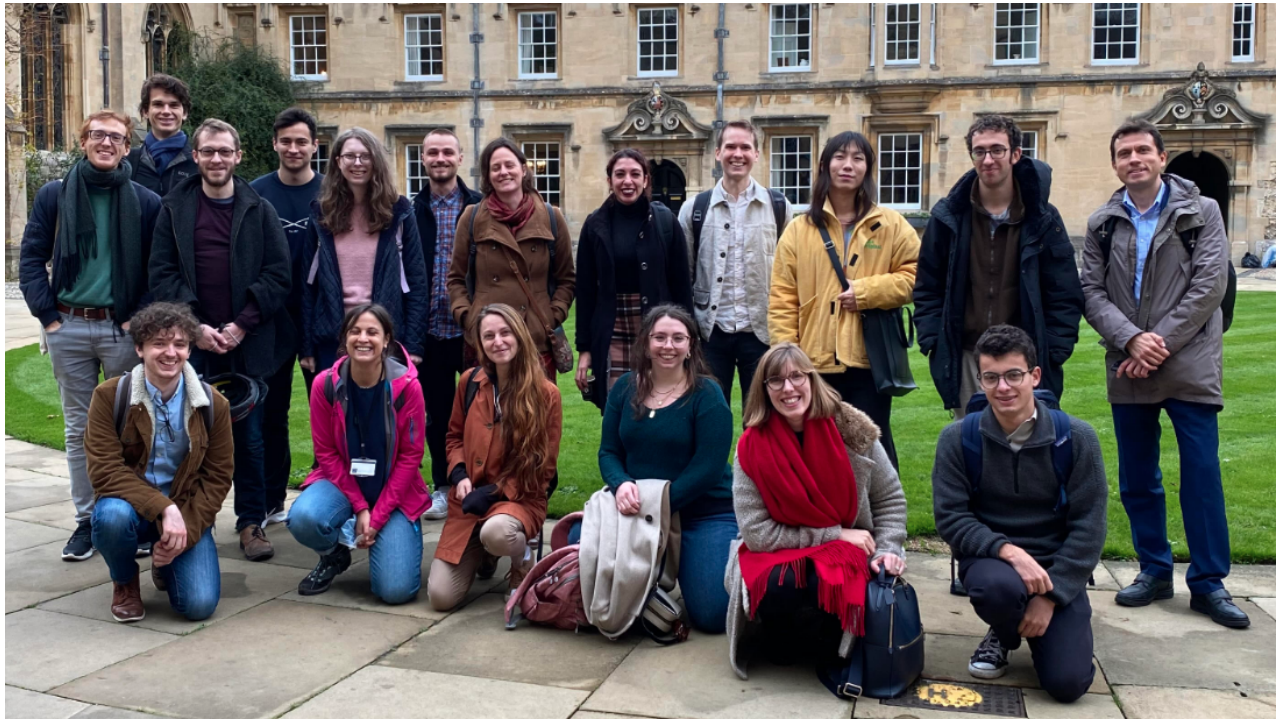
For a complex-balanced system, each equilibria $c \in \mathbb{R}^N_{>0}$ is globally asymptotically stable relative to the interior of its compatibility class.

1. Chemical reaction network books: 2000 and 2019





1. Algebraic systems biology at Oxford (Heather Harrington)



Algebraic Systems Biology

Much of our research is motivated by applications. We develop models and methods to study primarily biological and chemical systems; however, our work is also applied towards engineering, medical, physical and social problems. Such analysis often requires working with data.

Our research group uses mathematical and statistical techniques including numerical algebraic geometry, Bayesian statistics, computational topology, differential equations, linear algebra, network science, and optimisation, in order to solve interdisciplinary problems. Our research interests include applied algebraic geometry, algebraic statistics, dynamical systems, topological data analysis, cellular signaling, chemical reaction network theory, mathematical and systems biology.

The research group is led by Heather Harrington. See our members page for more details. We are mathematicians working at the interface of theoretical, applied, and data science.

1. Chemical reaction networks

Books

- M. Feinberg. *Foundations of Chemical Reaction Network Theory*. Springer, 2019.
- K. Gatermann. *Computer Algebra Methods for Equivariant Dynamical Systems*. Springer, 2000.

Articles

- C. Conradi, M. Mincheva, and A. Shiu. Emergence of oscillations in a mixed-mechanism phosphorylation system. *Bull. Math. Biol.*, 81(6):1829–1852, 2019.
- G. Craciun, A. Dickenstein, A. Shiu, and B. Sturmfels. Toric dynamical systems. *J. Symb. Comput.*, 44(11):1551–1565, 2009.
- G. Craciun and M. Feinberg. Multiple equilibria in complex chemical reaction networks: I. the injectivity property. *SIAM J. Appl. Math.*, 65(5):1526–1546, 2005.
- G. Craciun and M. Feinberg. Multiple equilibria in complex chemical reaction networks: Ii. the species-reaction graph. *SIAM J. Appl. Math.*, 66(4):1321–1338, 2006.
- G. Craciun, F. Nazarov, and C. Pantea. Persistence and permanence of mass-action and power-law dynamical systems. *SIAM J. Appl. Math.*, 73(1):305–329, 2013.
- E. Gross, H. A. Harrington, Z. Rosen, and B. Sturmfels. Algebraic systems biology: a case study for the wnt pathway. *Bull. Math. Biol.*, 78(1):21–51, 2016.
- A. Shiu and B. Sturmfels. Siphons in chemical reaction networks. *Bull. Math. Biol.*, 72(6):1448–1463, 2010.

Surveys / book chapters

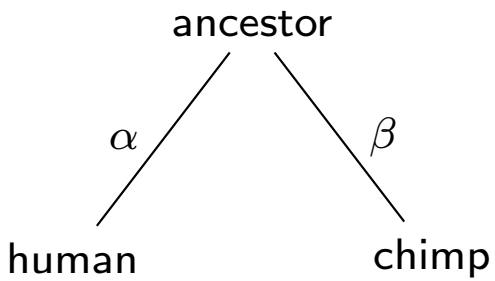
- C. Pantea, A. Gupta, J. B. Rawlings, and G. Craciun. The QSSA in chemical kinetics: as taught and as practiced. In *Discrete and Topological Models in Molecular Biology*, pages 419–442. Springer, 2014.
- K. Conradi and C. Pantea. Multistationarity in biochemical networks: results, analysis, and examples. In *Algebraic and Combinatorial Computational Biology*, pages 279–317. Elsevier, 2018.

2. Phylogenetics and algebraic statistics

Consider a simple evolutionary relationship of two species and their most common ancestor.

Fix a particular base in the genome at a site that all three species share in a mutual alignment.

Under the Jukes-Cantor model of evolution, the probability of a mutation at that site is a constant.



It is straightforward to compute the probability that (human,chimp)=(A,C):

$$P(AC) = P\left(\begin{smallmatrix} & A & \\ A & & C \end{smallmatrix}\right) + P\left(\begin{smallmatrix} & G & \\ A & & C \end{smallmatrix}\right) + P\left(\begin{smallmatrix} & C & \\ A & & C \end{smallmatrix}\right) + P\left(\begin{smallmatrix} & T & \\ A & & C \end{smallmatrix}\right)$$

$$= \frac{1}{4}(1-3\alpha)\beta + \frac{1}{4}\alpha\beta + \frac{1}{4}\alpha(1-3\beta) + \frac{1}{4}\alpha\beta = \frac{1}{4}(\alpha+\beta-\alpha\beta).$$

Similarly, $P(AA) = \frac{1}{4}(1-3\alpha)(1-3\beta) + \frac{3}{4}\alpha\beta = 3\alpha\beta + \frac{1}{4}(1-3\alpha-3\beta)$.

The space of possible probabilities can be described by a mapping

$$\varphi\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^{16}, \qquad \varphi\colon (\alpha,\beta) \longmapsto \big(P(AA), P(AC), \ldots, P(TT)\big).$$

For an n-leaf tree with $m = 2n-2$ edges, we get a map $\varphi\colon \mathbb{R}^m \to \mathbb{R}^{4^n}$.

The intersection of $\mathrm{Im}(\varphi)$, with the $d = 4^n - 1$ dimensional simplex Δ_d is the *phylogenetic model*, $\mathcal{M}_T \subseteq \mathbb{R}^{4^n}$.

The polynomials that vanish on $\mathcal{M}_T$ is called the ideal of phylogenetic invariants,

$$I_T = I_T(\mathcal{M}_T) = \big\{ f \in \mathbb{R}[x_1, \ldots, x_{4^n}] \mid f(p) = 0, \quad \text{for all } p \in \mathcal{M}_T \big\}.$$

The points that vanish on all polynomials in the ideal I_T is called the phylogenetic variety of T:

$$V_T = V_T(I_T) = \big\{ p \in \mathbb{R}^{4^n} \mid f(p) = 0, \quad \text{for all } f \in I_T \big\}.$$

2. Phylogenetics and algebraic statistics

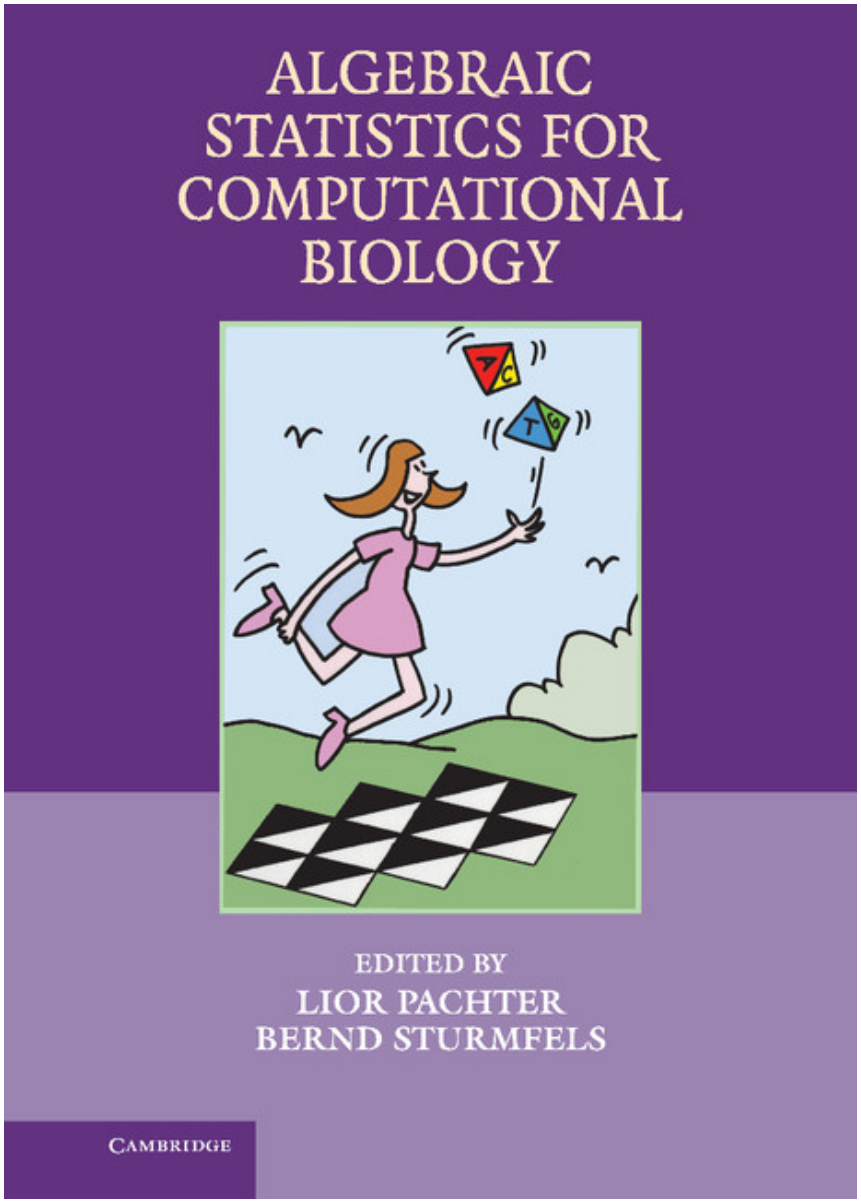
Books

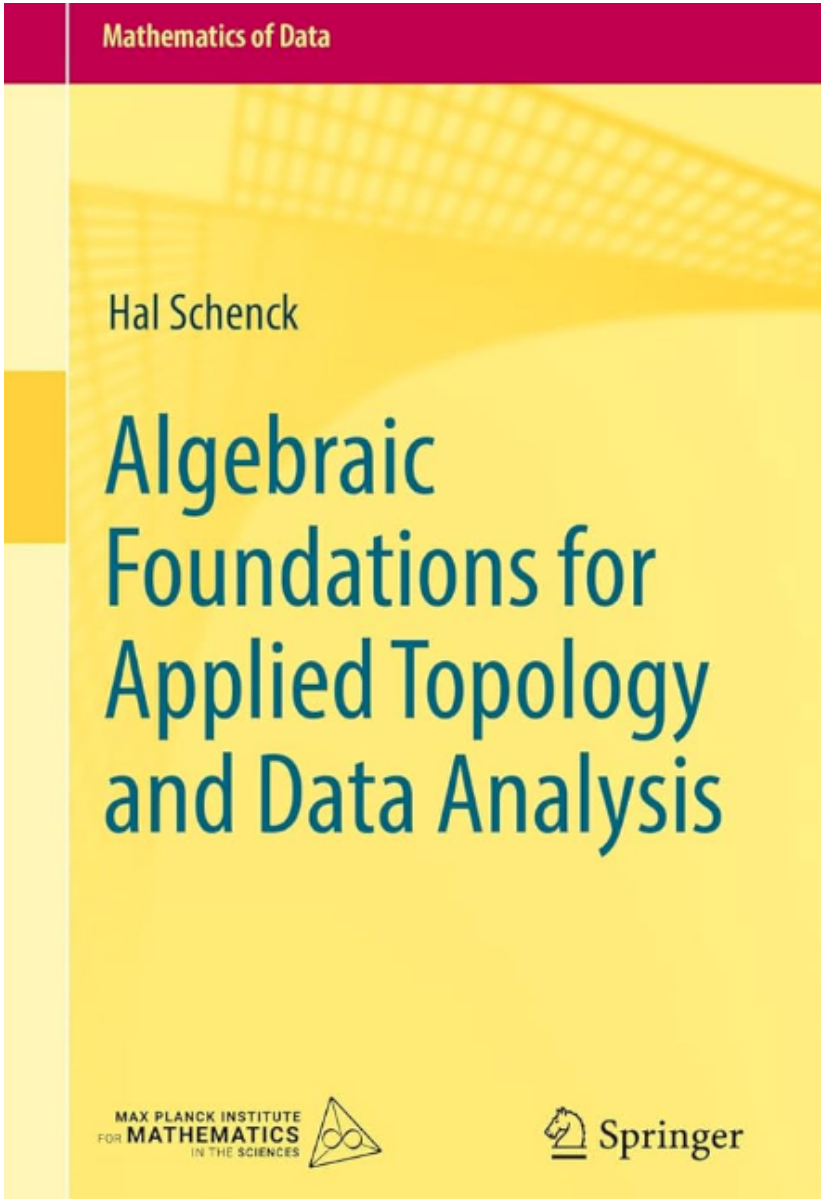
- D. Maclagan and B. Sturmfels. *Introduction to Tropical Geometry*, volume 161. American Mathematical Soc., 2015.
- L. Pachter and B. Sturmfels. *Algebraic Statistics for Computational Biology*, volume 13. Cambridge University Press, 2005.
- G. Pistone, E. Riccomagno, and H. P. Wynn. *Algebraic Statistics: Computational Commutative Algebra in Statistics*. Chapman and Hall/CRC, 2000.
- R. Rabadán and A. J. Blumberg. *Topological Data Analysis for Genomics and Evolution: Topology in Biology*.
- S. Sullivant. *Algebraic Statistics*, volume 194. American Mathematical Soc., 2018.

Articles

- M. Casanellas and J. A. Rhodes. Algebraic methods in phylogenetics. *Bull. Math. Biol.*, 81:313–315, 2019.
- J. Chifman and L. Kubatko. Quartet inference from SNP data under the coalescent model. *Bioinformatics*, 30(23):3317–3324, 2014.
- P. Diaconis, B. Sturmfels, et al. Algebraic algorithms for sampling from conditional distributions. *Ann. Stat.*, 26(1):363–397, 1998.
- J. Fernández-Sánchez and M. Casanellas. Invariant versus classical quartet inference when evolution is heterogeneous across sites and lineages. *Syst. Biol.*, 65(2):280–291, 2016.
- L. Pachter and B. Sturmfels. The mathematics of phylogenomics. *SIAM Rev.*, 49(1):3–31, 2007.

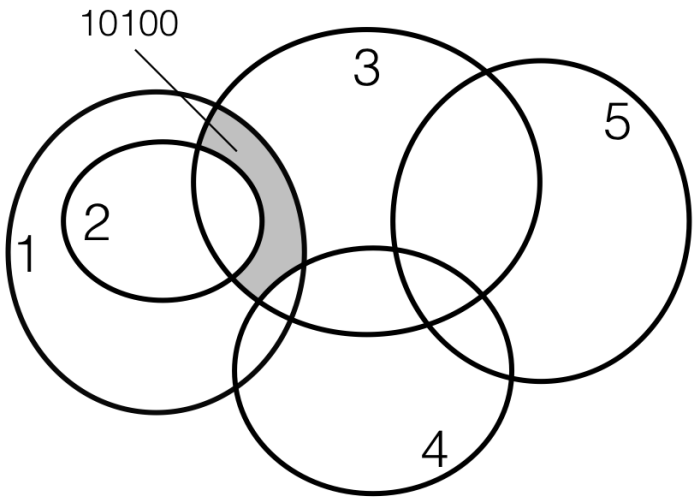
2. Algebra, topology, and statistics





3. Place fields in neuroscience

Experiments have shown that neurons called place cells fire based on an animal's location.



The place fields $\mathcal{U} = \{U_1, U_2, U_3, U_4, U_5\}$ define the neural code

$$\mathcal{C} = \{00000, 10000, 11000, 10100, 11100, 10010, 10110, 00100, 00110, 00101, 00111, 00010, 00011, 00001\}?$$

Motivating question

Given a neural code, reconstruct the place fields, if possible.

The shaded region is encoded by the pseudomonomial called its characteristic polynomial:

$$\chi_c(x) = x_1(x_2 - 1)x_3(x_4 - 1)(x_5 - 1) = x_1\overline{x_2}x_3\overline{x_4}\,\overline{x_5} = \begin{cases} 1 & x = c \\ 0 & x \neq c. \end{cases}$$

We can encode $\mathcal{C}$ and $\mathcal{U}$ by ideals in $\mathbb{F}_2[x_1, \ldots, x_5]$ involving these and other polynomials.

3. Place fields in neuroscience

Another interesting question

Given a neural code, can it be realized by a collection of open convex place fields?

For example, the code

$$\mathcal{C} = \{000, 100, 010, 101, 110, 011\}$$

cannot be realized by open convex place fields.

Many of these questions can be approached algebraically. Every code $\mathcal{C}$ has a vanishing ideal,

$$I_{\mathcal{C}} = \big\{ f \in \mathbb{F}_2[x_1, \ldots, x_n] \mid f(\mathrm{c}) = 0 \text{ for all } \mathrm{c} \in \mathcal{C} \big\},$$

and a neural ideal, generated by the characteristic polynomials of the non-code words:

$$J_{\mathcal{C}} = \big(\chi_{\mathrm{n}}(\mathrm{x}) \mid \mathrm{n} \notin \mathcal{C}\big), \qquad \text{where} \quad \chi_{\mathrm{n}}(\mathrm{x}) = \begin{cases} 1 & \mathrm{x} = \mathrm{n} \\ 0 & \mathrm{x} \neq \mathrm{n}. \end{cases}$$

These ideals are related by

$$I_{\mathcal{C}} = J_{\mathcal{C}} + \mathcal{B} = \big(\{\chi_{\mathrm{n}}(\mathrm{x}) \mid \mathrm{n} \notin \mathcal{C}\} \cup \{x_i^2 - x_i \mid i = 1, \ldots, n\}\big).$$

3. Place fields in neuroscience

Geometric & combinatorial features encoded algebraically are receptive field (RF) relationships.

geometric	**combinatorial**	**algebraic**
$U_1 \cap U_2 \neq \varnothing$	$c_1 = c_2 = 1$ for some $c \in \mathcal{C}$	$(x_1 - 1)(1 - x_2) \notin J_{\mathcal{C}}$
$U_1 \subseteq U_2$	$c_1 = 1 \Rightarrow c_2 = 1$	$x_1(1 - x_2) \in J_{\mathcal{C}}$

Definition

The set of minimal pseudomonomials in $J_{\mathcal{C}}$ is called the canonical form of $J_{\mathcal{C}}$.

The canonical form $\mathrm{CF}(J_{\mathcal{C}})$ can be computed from the primary decomposition of $J_{\mathcal{C}}$.

Example

Suppose $U_1 \subseteq U_2 \subseteq U_3$. The neural code is $\mathcal{C} = \{000, 111, 011, 001\}$ and the neural ideal is

$$J_{\mathcal{C}} = \big((1-x_1)x_2(1-x_3),\ x_1(1-x_2)(1-x_3),\ x_1(1-x_2)x_3,\ x_1x_2(1-x_3)\big) = \big(x_1(1-x_2),\ x_2(1-x_3)\big),$$

and its canonical form $\mathrm{CF}(J_{\mathcal{C}}) = \big\{x_1(1 - x_2),\ x_2(1 - x_3),\ x_1(1 - x_3)\big\}$.

3. Place fields in neuroscience

Research goals and open-ended questions

Key idea

How does the algebra encode the geometric and topological properties of the place fields?

- How are certain combinatorial relationships (e.g., intersections, subsets, etc.) encoded algebraically?
- What are necessary and sufficient conditions for a neural code to be convex?
- Given a code, what is the smallest dimension where it can be realized?
- How can one construct the canonical form?
- How do properties of the vanishing and neural ideals tells us about the place fields?
- Explore the theory of pseudomonomial ideals.

3. Algebraic neuroscience at Brown University (Carina Curto)

ICERM

Home Programs Your Visit Videos About Support ICERM

Neural Coding and Combinatorics

Oct 30 - Nov 3, 2023

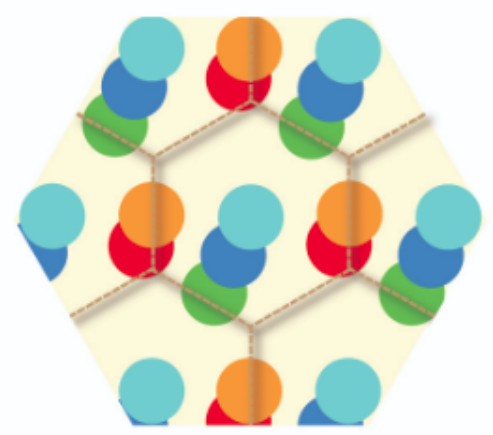
Workshop Overview | Workshop Participants | Workshop Schedule | Reimburse Me | Semester Workshops

Organizing Committee

- Zachary Kilpatrick
  University of Colorado Boulder
- Tatyana Sharpee
  Salk Institute
- Katie Morrison
  University of Northern Colorado
- Nora Youngs
  Colby College
- Elad Schneidman
  Weizmann Institute of Science

Abstract

Cracking the neural code is one of the longstanding questions in neuroscience. How does the activity of populations of neurons represent stimuli and perform neural computations? Decades of theoretical and experimental work have provided valuable clues about the principles of neural coding, as well as descriptive understandings of various neural codes. This raises a number of mathematical questions touching on algebra, combinatorics, probability, and geometry. This workshop will explore questions that arise from sensory perception and processing in olfactory, auditory, and visual coding, as well as properties of place field codes and grid cell codes, mechanisms for decoding population activity, and the role of noise and correlations. These questions may be tackled with techniques from information theory, mathematical coding theory, combinatorial commutative algebra, hyperplane arrangements, oriented matroids, convex geometry, statistical mechanics, and more.



3. Place fields in neuroscience

Surveys (book chapters, graduate theses)

- C. Curto, A. Veliz-Cuba, and N. Youngs. Analysis of combinatorial neural codes: An algebraic approach. In *Algebraic and Combinatorial Computational Biology*, pages 213–240. Elsevier, 2018.
- S. A. Tsiorintsoa. *Pseudo-monomials in algebraic biology.* MSc thesis. AIMS South Africa, 2018.
- N. Youngs. *The neural ring: using algebraic geometry to analyze neural codes*. PhD thesis, University of Nebraska, Lincoln, 2014.

Articles

- J. Cruz, C. Giusti, V. Itskov, and B. Kronholm. On open and closed convex codes. *Discrete Comput. Geom.*, 61(2):247–270, 2019.
- C. Curto, E. Gross, J. Jeffries, K. Morrison, M. Omar, Z. Rosen, A. Shiu, and N. Youngs. What makes a neural code convex? *SIAM J. Appl. Alg. Geom.*, 1(1):222–238, 2017.
- C. Curto, E. Gross, J. Jeffries, K. Morrison, Z. Rosen, A. Shiu, and N. Youngs. Algebraic signatures of convex and non-convex codes. *J. Pure Appl. Alg.*, 223(9):3919–3940, 2019.
- C. Curto, V. Itskov, K. Morrison, Z. Roth, and J. L. Walker. Combinatorial neural codes from a mathematical coding theory perspective. *Neural Computation*, 25(7):1891–1925, 2013.
- C. Curto, V. Itskov, A. Veliz-Cuba, and N. Youngs. The neural ring: an algebraic tool for analyzing the intrinsic structure of neural codes. *Bull. Math. Biol.*, 75(9):1571–1611, 2013.
- R. Garcia, L. D. García Puente, R. Kruse, J. Liu, D. Miyata, E. Petersen, K. Phillipson, and A. Shiu. Gröbner bases of neural ideals. *Int. J. Algebr. Comput.*, 28(04):553–571, 2018.
- S. Gunturkun, J. Jeffries, and J. Sun. Polarization of neural rings. *arXiv:1706.08559*, 2017.
- A. Kunin, C. Lienkaemper, and Z. Rosen. Oriented matroids and combinatorial neural codes. *arXiv:2002.03542*, 2020.
- E. Petersen, N. Youngs, R. Kruse, D. Miyata, R. Garcia, and L. D. G. Puente. Neural ideals in sagemath. In *International Congress on Mathematical Software*, pages 182–190. Springer, 2018.
- A. Ruys de Perez, Laura Felicia Matusevich, and Anne Shiu. Neural codes and the factor complex. *Adv. Appl. Math.*, 114:101977, 2020.
- N. Youngs. Neural ideal: a Matlab package for computing canonical forms, 2015. `https://github.com/nebneuron/neural-ideal/`.

4. Boolean models of molecular networks

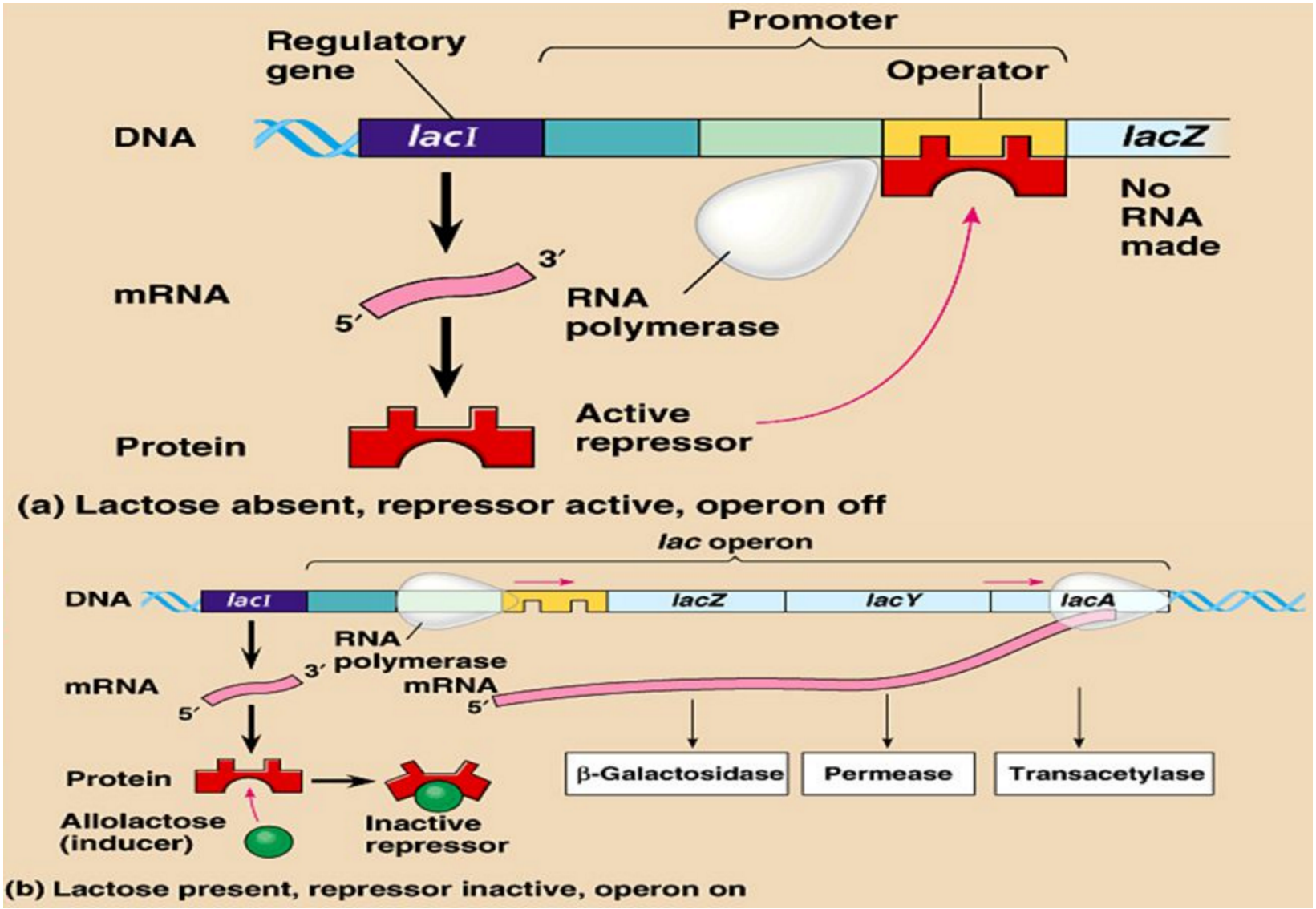


Figure: The lactose operon in *E. coli*

4. Delay differential equation models of molecular networks

- The repressor protein binds to allolactose:

$$R + nA \underset{1}{\overset{K_1}{\rightleftharpoons}} RA_n, \qquad \frac{d[RA_n]}{dt} = K_1[R][A]^n - [RA_n]$$

- *The repressor protein binds to the operator region if there is no allolactose*:

$$O + R \underset{1}{\overset{K_2}{\rightleftharpoons}} OR \qquad \frac{d[OR]}{dt} = K_2[O][R] - [OR].$$

- β-galactosidase converts *Lactose into allolactose*:

$$L + B \rightleftharpoons LB \longrightarrow A + B, \qquad \frac{d[A]}{dt} = \alpha_A \frac{[B][L]}{K_L + [L]}$$

- β-galactosidase cleaves *Allolactose into glucose and galactose*

$$A + B \rightleftharpoons AB \longrightarrow B + Glu + Gal, \qquad \frac{d[A]}{dt} = -\alpha_G \frac{[B][A]}{K_A + [A]}$$

- *lac* permease transports lactose **into** the cell:

$$L_e + P \rightleftharpoons PL_e \longrightarrow P + L, \qquad \frac{d[L]}{dt} = \alpha_L \frac{[P][L_e]}{K_{L_e} + [L_e]}$$

- *lac* permease transports lactose **out** of the cell:

$$L + P \rightleftharpoons PL \longrightarrow P + L_e, \qquad \frac{d[L]}{dt} = -\alpha_{L_e} \frac{[P][L]}{K_{L'} + [L]}$$

4. Delay differential equation models of molecular networks

This leads to the following nonlinear system of ODEs:

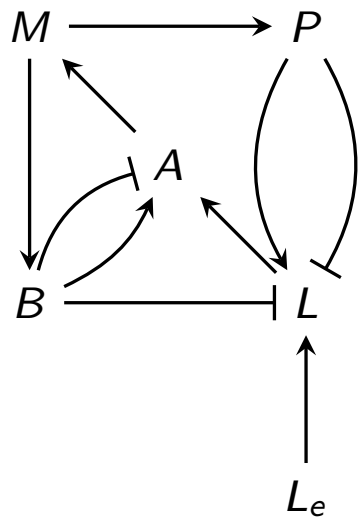
$$\frac{dM}{dt} = \alpha_M \frac{1 + K_1(e^{-\mu\tau_M} A_{\tau_M})^n}{K + K_1(e^{-\mu\tau_M} A_{\tau_M})^n} + \Gamma_0 - \widetilde{\gamma}_M M$$

$$\frac{dB}{dt} = \alpha_B e^{-\mu\tau_B} M_{\tau_B} - \widetilde{\gamma}_B B$$

$$\frac{dA}{dt} = \alpha_A B \frac{L}{K_L + L} - \beta_A B \frac{A}{K_A + A} - \widetilde{\gamma}_A A$$

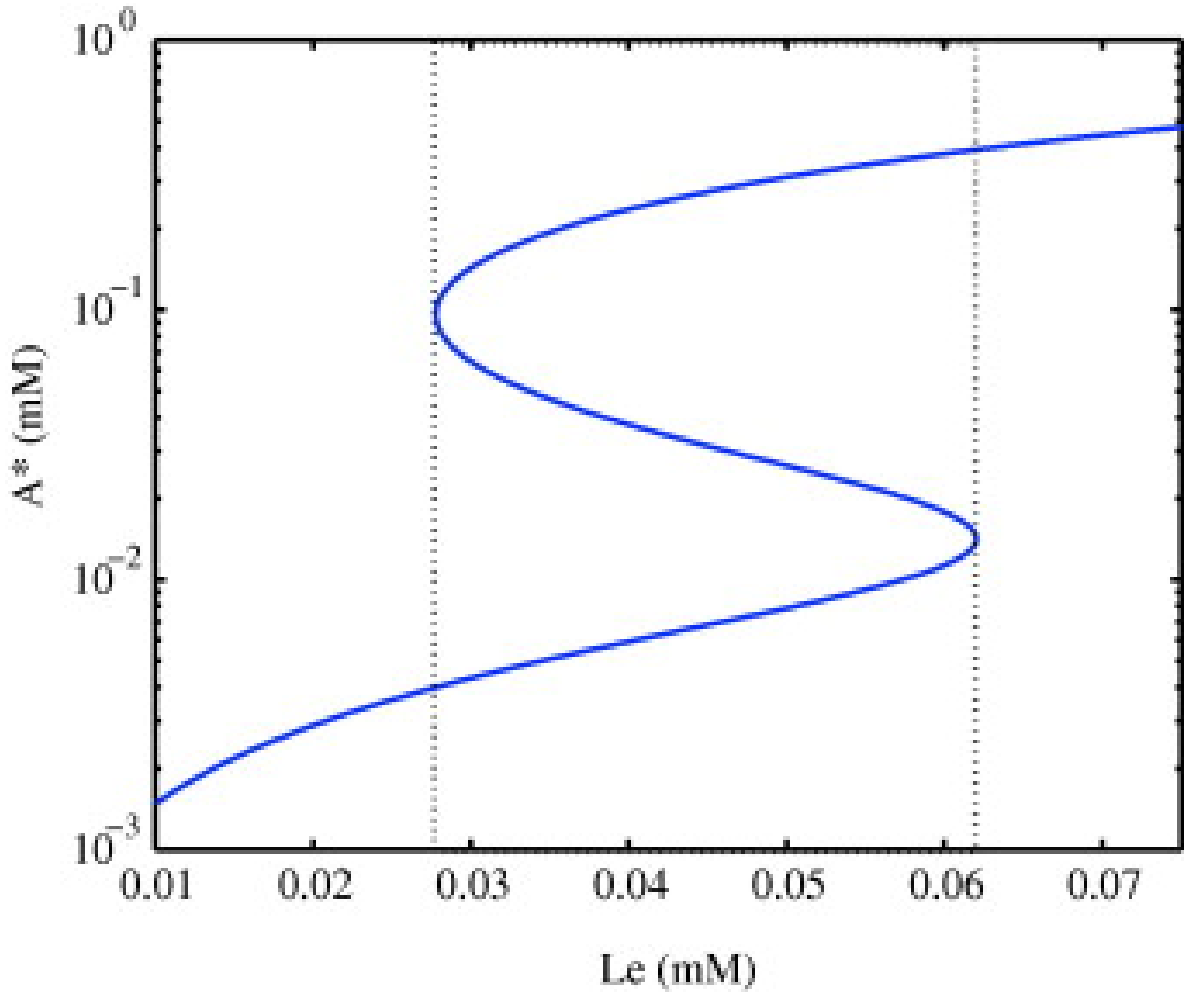
$$\frac{dP}{dt} = \alpha_P e^{-\mu(\tau_B + \tau_P)} M_{\tau_B + \tau_P} - \widetilde{\gamma}_P P$$

$$\frac{dL}{dt} = \alpha_L P \frac{L_e}{K_{L_e} + L_e} - \beta_{L_e} P \frac{L}{K_{L_e} + L} - \alpha_A B \frac{L}{K_L + L} - \widetilde{\gamma}_L L$$



4. Delay differential equation models and bistability

The above ODE model captures the bistability of the *lac* operon.

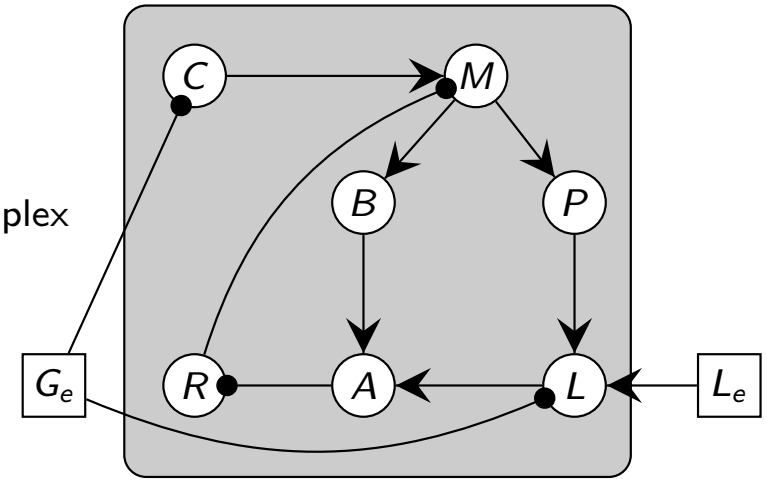


Can a Boolean model exhibit this as well?

4. Boolean models of molecular networks

The following Boolean network model was published in Veliz-Cuba / Stigler (2011).

$M =$ mRNA
$P =$ *lac* permease
$B =$ β-galactosidase
$C =$ cAMP-CAP complex
$R =$ repressor protein
$L =$ lactose
$A =$ allolactose
$G =$ glucose



$$\begin{cases} f_M = \overline{R} \wedge \overline{R_m} \wedge C \\ f_P = M \\ f_B = M \\ f_C = \overline{G_e} \\ f_R = \overline{A} \wedge \overline{A_m} \\ f_{R_m} = (\overline{A} \wedge \overline{A_m}) \vee R \\ f_A = L \wedge B \\ f_{A_m} = L \vee L_m \\ f_L = \overline{G_e} \wedge P \wedge L_e \\ f_{L_m} = \overline{G_e} \wedge ((L_{em} \wedge P) \vee L_e) \end{cases}$$

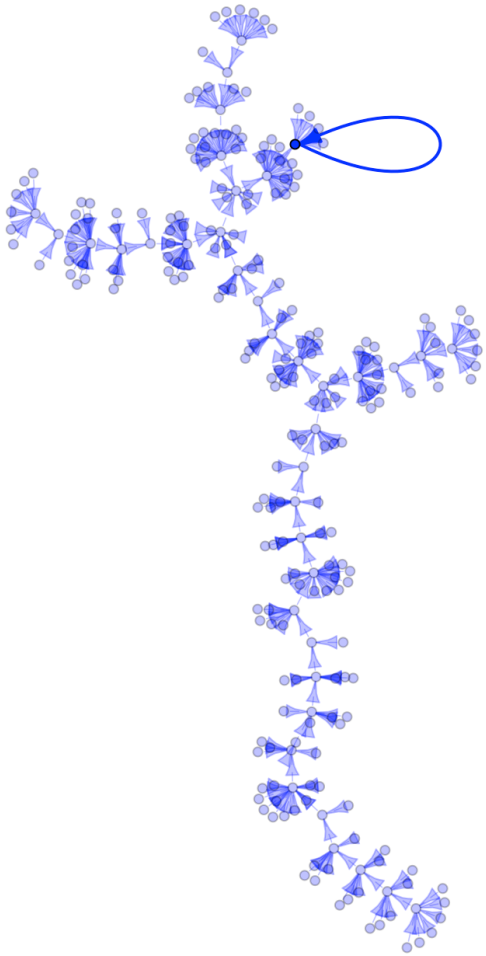
- The 'e' subscript means *extracellular*
- The 'm' subscript means *(at least) medium* levels

To validate this model, we need to analyze it for

$$(G_e, L_e, L_{em}) = (0,0,0),\ (0,0,1),\ (0,1,1),\ (1,0,0),\ (1,0,1),\ (1,1,1).$$

4. Dynamics of the Boolean model, and bistability

Here is the phase space with $(G_e, L_e, L_{em}) = (0, 0, 1)$, generated with BoolNet in R.



```
> print(getBasinOfAttraction(lacAttractorsBistable,2))
     State      Next state  Attr. basin  # trans. to attr.
1101001001 =>   1111000101            2                  1
1111001001 =>   1111000101            2                  1
1101000101 =>   1111000101            2                  1
1111000101 =>   1111000101            2                  0
1101001101 =>   1111000101            2                  1
1111001101 =>   1111000101            2                  1

Genes are encoded in the following order: M P B C R Rm A Am L Lm
```

4. Boolean models of molecular networks

We can write a Boolean model as polynomials in $\mathbb{F}_2[x_1, \ldots, x_{10}]$.

$$\begin{cases} f_1 = x_4(x_5+1)(x_6+1) \\ f_2 = x_1 \\ f_3 = x_1 \\ f_4 = G_e + 1 \\ f_5 = (x_7+1)(x_8+1) \\ f_6 = x_5 + (x_7+1)(x_8+1) + x_5(x_7+1)(x_8+1) \\ f_7 = x_3 x_9 \\ f_8 = x_9 + x_{10} + x_9 x_{10} \\ f_9 = x_2(G_e+1)L_e \\ f_{10} = (x_2 L_{em} + L_e + x_2 L_{em} L_e)(G_e+1) \end{cases}$$

The steady-states are the algebraic variety of the set

$$\{f_1 + x_1, \ldots, f_{10} + x_{10}\}.$$

We can now use computational algebraic tools to analyze it.

4. Reverse-engineering Boolean functions

Consider an unknown Boolean function $f : \mathbb{F}_2^3 \to \mathbb{F}_2$ satisfying:

xyz	111	000	110
$f(x, y, z)$	0	0	1

We encode this *data* as

$$\mathcal{D} = \{(s_1, t_1), (s_2, t_2), (s_3, t_3)\} = \{(111, 0), (000, 0), (110, 1)\}.$$

For each pair with different outputs, we get a pseudomonomial:

$$p(s_1, s_3) = z - (\text{sign}(s_{33} - s_{13})) = z + 1, \qquad p(s_2, s_3) = (x - 1)(y - 1).$$

The ideal of signed non-disposable sets for $\mathcal{D}$ is

$$J_{\Delta_{\mathcal{D}}^c} = \langle p(s_1, s_3),\ p(s_2, s_3) \rangle = \langle z + 1,\ (x - 1)(y - 1) \rangle.$$

We compute the primary decomposition of $J_{\Delta_{\mathcal{D}}^c}$:

```
R = ZZ/3[x,y,z];
J_nonDisp = ideal(z+1, (x-1)*(y-1));
primaryDecomposition J_nonDisp
```

Output: `{ideal (z + 1, y - 1), ideal (z + 1, x - 1)}`

- Primary decomposition: $J_{\Delta_{\mathcal{D}}^c} = \langle x - 1, z + 1 \rangle \cap \langle y - 1, z + 1 \rangle$.
- Signed min-sets: $\{x, \overline{z}\}$ and $\{y, \overline{z}\}$.

4. Reverse-engineering Boolean functions

Consider an unknown Boolean function $f : \mathbb{F}_2^3 \to \mathbb{F}_2$ satisfying:

xyz	111	000	110
$f(x, y, z)$	0	0	1

Any such function must depend on (at least) $\{x, \overline{z}\}$ or on $\{y, \overline{z}\}$.

Theorem

The ideal of signed non-disposable sets is the ideal in $\mathbb{F}_3[x_1, \ldots, x_n]$ defined by

$$J_{\Delta_{\mathcal{D}}^c} = \left\langle p(s_i, s_j) \mid i < j,\ t_i \neq t_j \right\rangle.$$

But what if we want to find all such functions?

The model space is a coset of the vanishing ideal:

$$\mathrm{mod}(\mathcal{D}) = f + I = \left\{ f + h \mid h \in I \right\}.$$

4. Pseudomonomials

Pseudo-monomials in algebraic biology

Sandra Annie Tsiorintsoa (sandra@aims.ac.za)
African Institute for Mathematical Sciences (AIMS)

Supervised by: Dr. Matthew Macauley
Clemson University, South Carolina

24 May 2018

Submitted in partial fulfillment of a structured masters degree at AIMS South Africa

AIMS | African Institute for Mathematical Sciences SOUTH AFRICA

4. Algebraic systems biology at Univ. Florida (Reinhard Laubenbacher)

4. Boolean models of molecular networks

Research goals and open-ended questions

Key idea

What does the polynomial algebra tell us about the dynamics of the algebraic model?

- Given a gene regulatory network, how can we model it with polynomials?
- How we can determine whether are there only fixed points?
- How we we reduce a large model while perserving its key features? (e.g., fixed points, limit cycles, etc.)
- How can we characterize stability of the dynamics of an algebraic model?
- *How can we reverse-engineer a model given partial data?*
- What can we say about the dynamics if we restrict to a particular class of functions?
- How does the update order (synchronous, asynchronous, block-sequential) determine the dynamics?

4. Boolean models of molecular networks

Books

- Y. Crama and P. L. Hammer. *Boolean Models and Methods in Mathematics, Computer Science, and Engineering*, volume 2. Cambridge University Press, 2010.
- I. Shmulevich and E.R. Dougherty. *Probabilistic Boolean networks: the modeling and control of gene regulatory networks*. SIAM, 2010.
- R. Thomas and R. d'Ari. *Biological feedback*. CRC press, 1990 (updated 2006).

Book chapters

- R. Robeva. *Algebraic and Discrete Mathematical Methods for Modern Biology*. Elsevier, 2015. Chapters 1–6.
- R. Robeva and T. Hodge. *Mathematical Concepts and Methods in Modern Biology: Using Modern Discrete Models*. Academic Press, 2013. Chapters 4–7.
- R. Robeva and M. Macauley. *Algebraic and Combinatorial Computational Biology*. Elsevier, 2018. Chapters 4–6.

Survey articles

- B. Drossel. *Random Boolean networks*, Chapter 3, pages 69–110. Wiley-VCH Verlag GmbH & Co., Weinheim, Germany, 2009.
- R.-S. Wang, A. Saadatpour, and R. Albert. Boolean modeling in systems biology: an overview of methodology and applications. *Phys. Biol.*, 9(5):055001, 2012.
- F. Hinkelmann, D. Murrugarra, A. S. Jarrah, and R. Laubenbacher. A mathematical framework for agent based models of complex biological networks. *Bulletin of mathematical biology*, 73(7):1583–1602, 2011.
- A. Gonzalez, A. Naldi, L. Sanchez, D. Thieffry, and C. Chaouiya. GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks. *Biosystems*, 84(2):91–100, 2006.
- D. Thieffry and M. Kaufman. Prologue to the special issue of JTB dedicated to the memory of René Thomas (1928-2017): A journey through biological circuits, logical puzzles and complex dynamics. *J. Theor. Biol.*, 474:42, 2019.

4. Boolean models of molecular networks

Articles (biological modeling)

- R. Albert and H.G. Othmer. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. *J. Theor. Biol.*, 223(1):1–18, 2003.
- E.E. Allen, J.S. Fetrow, L.W. Daniel, S.J. Thomas, and D.J John. Algebraic dependency models of protein signal transduction networks from time-series data. *J. Theor. Biol.*, 238(2):317–330, 2006.
- M. Davidich and S. Bornholdt. Boolean network model predicts cell cycle sequence of fission yeast. *PloS ONE*, 3(2):e1672, 2008.
- A. Fauré, A. Naldi, C. Chaouiya, and D. Thieffry. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics*, 22(14):e124–e131, 2006.
- A. Jenkins and M. Macauley. Bistability and asynchrony in a Boolean model of the L-arabinose operon in *Escherichia coli*. *Bull. Math Biol.*, 79(8):1778–1795, 2017.
- S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein. Random Boolean network models and the yeast transcriptional network. *Proc. Natl. Acad. Sci.*, 100(25):14796–14799, 2003.
- L. Mendoza, D. Thieffry, and E. Alvarez-Buylla. Genetic control of flower morphogenesis in Arabidopsis thaliana: a logical analysis. *Bioinformatics*, 15(7):593–606, 1999.
- B. Stigler and H. Chamberlin. A regulatory network modeled from wild-type gene expression data guides functional predictions in caenorhabditis elegans development. *BMC Syst. Biol.*, 6(1):77, 2012.
- A. Veliz-Cuba and B. Stigler. Boolean models can explain bistability in the *lac* operon. *J. Comp. Biol.*, 18(6):783–794, 2011.

4. Boolean models of molecular networks

Articles (algebraic aspects)

- E.S. Dimitrova, A.S. Jarrah, R. Laubenbacher, and B. Stigler. A Gröbner fan method for biochemical network modeling. In *Proc. Internat. Symposium Symb. Algebraic Comput.*, pages 122–126. ACM, 2007.
- F. Hinkelmann and R. Laubenbacher. Boolean models of bistable biological systems. *Discrete Cont. Dyn. Sys. Ser. S*, 4(6):1443–1456, 2011.
- A.S. Jarrah, R. Laubenbacher, B. Stigler, and M. Stillman. Reverse-engineering of polynomial dynamical systems. *Adv. Appl. Math.*, 39(4):477–489, 2007.
- R. Laubenbacher and B. Stigler. A computational algebra approach to the reverse engineering of gene regulatory networks. *J. Theor. Biol.*, 229(4):523–537, 2004.
- D. Murrugarra and E.S. Dimitrova. Molecular network control through Boolean canalization. *EURASIP J. Bioinformatics Sys. Biol.*, 2015(1):1–8, 2015.
- D. Murrugarra, A. Veliz-Cuba, B. Aguilar, and R. Laubenbacher. Identification of control targets in Boolean molecular network models via computational algebra. *BMC Systems Biology*, 10(1):94, 2016.
- L. Sordo Vieira, R.C. Laubenbacher, and D. Murrugarra. Control of intracellular molecular networks using algebraic methods. *Bull Math Biol*, 82(2), 2020.
- A. Veliz-Cuba. An algebraic approach to reverse engineering finite dynamical systems arising from biology. *SIAM Journal on Applied Dynamical Systems*, 11(1):31–48, 2012.
- A. Veliz-Cuba, B. Aguilar, F. Hinkelmann, and R. Laubenbacher. Steady state analysis of Boolean molecular network models via model reduction and computational algebra. *BMC Bioinformatics*, 15(1):221, 2014.
- A. Veliz-Cuba and R. Laubenbacher. On the computation of fixed points in Boolean networks. *J. Appl. Math. Comput.*, 39(1-2):145–153, 2012.