What is Algebraic Biology?

Matthew Macauley

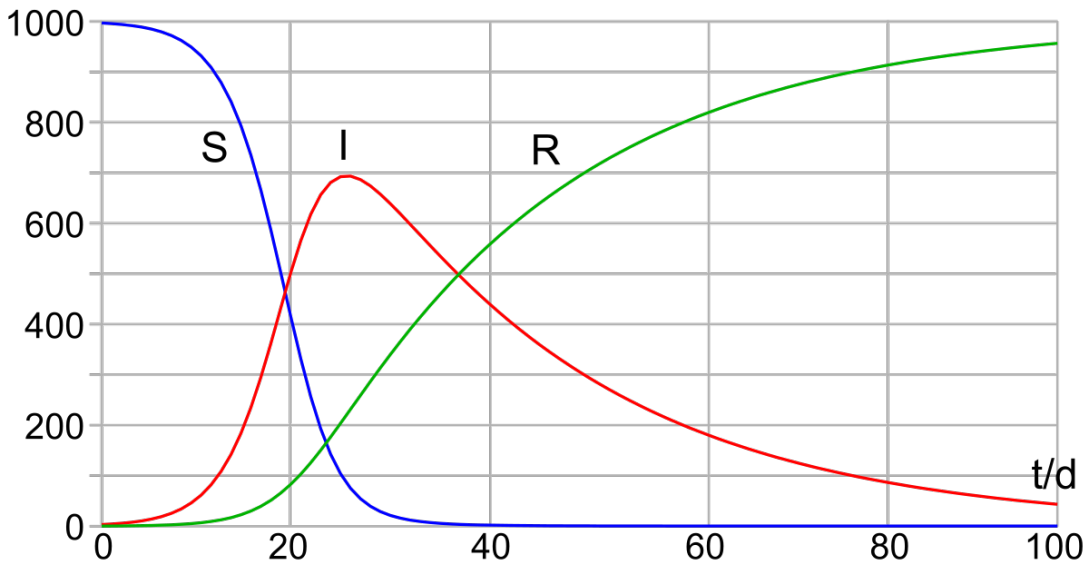
Department of Mathematical Sciences
Clemson University
http://www.math.clemson.edu/~macaule/

Abram Gannibal Project
November 18, 2020

What does algebra have to do with biology?

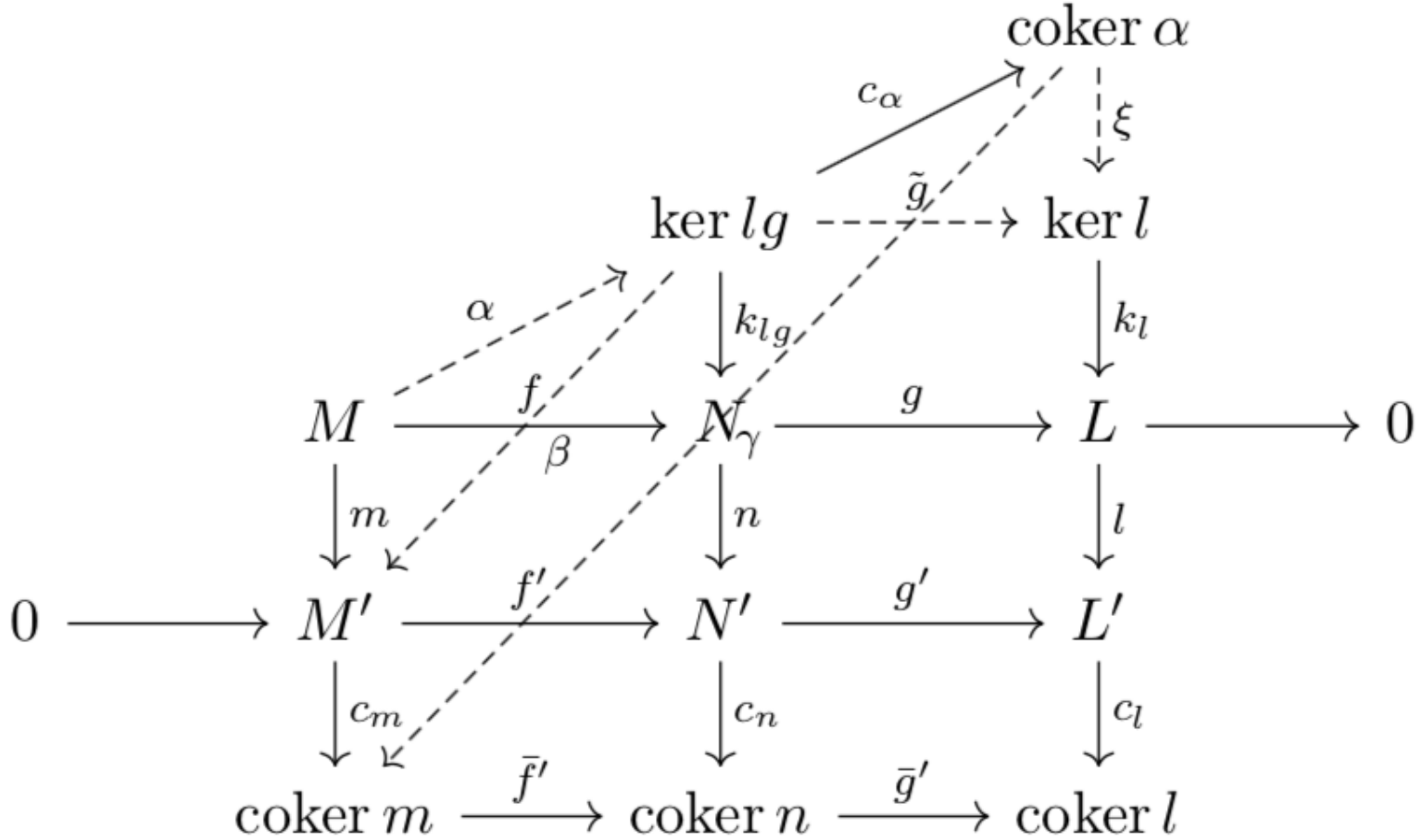
Usually, when we think of mathematical biology, we think of models such as this:

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



What does algebra have to do with biology?

Whereas algebra might remind us more of this:

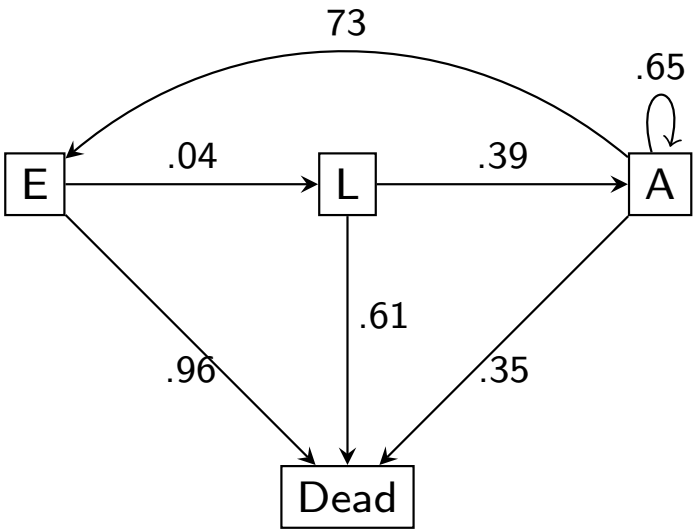


How could these two topics possibly be related??

What does algebra have to do with biology?

We all know that linear algebra is fundamental to mathematical biology.

Consider the following example, of a structured population of Eggs, Larvae, and Adults.



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

This is one example of many, of how linear differential or difference equations can model natural phenomena.

What does algebra have to do with biology?

Linear algebra also arises when approximating non-linear models, a process called linearization.

For example, consider the following Lotka-Volterra equations that model two competing species:

$$P' = P(1 - P - Q)$$
$$Q' = Q(.75 - Q - .5P)$$

Since $(1, 0)$ is a steady-state, we can change variables $(X, Y) = (P - 1, Q - 0)$, and get the system

$$\begin{bmatrix} X' \\ Y' \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 0 & .25 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} - \begin{bmatrix} X^2 + XY \\ .5XY + Y^2 \end{bmatrix}.$$

For $(X, Y) \approx (1, 0)$, the non-linear terms are negligible. The *linearized* system is thus

$$\begin{bmatrix} X' \\ Y' \end{bmatrix} \approx \begin{bmatrix} -1 & -1 \\ 0 & .25 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}.$$

What does algebra have to do with biology?

Linear algebra, the study of linear polynomials and their solutions, is a fundamental pillar of mathematical biology.

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.

In the rest of this lecture, we'll see four examples of biological problems where nonlinear algebra arises:

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

Then we'll discuss new (pure) mathematics that has arisen from these biological problems.

Linear algebra vs. nonlinear algebra

I will assume that everyone is familiar with the concept of a vector space, which 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

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

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

Consider a simple biochemical reaction, where A, B, and C are molecular species:

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

The constants k_1, k_2, and k_3 represent reaction rates.

Let $x_1(t)$, $x_2(t)$, and $x_3(t)$ denote concentrations of A, B, and C. The assumption of the laws of mass-action kinetics leads to the following system of ODEs:

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

To find the steady-states, set each $x_i' = 0$ and solve the system.

Biologically, we only care about solutions in the non-negative orthant of $\mathbb{R}^3$. However, polynomials are easier to study over $\mathbb{C}$.

For each fixed choice of parameters, the solutions form an algebraic variety in $\mathbb{C}^3$.

This can be found by computing a Gröbner basis of the ideal

$$I = \langle -k_1 x_1 x_2 - k_3 x_1 + k_2 x_3,\ -k_1 x_1 x_2 + k_2 x_3 + 2k_3 x_1,\ k_1 x_1 x_2 - k_2 x_3 \rangle.$$

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. Biochemical 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. Boolean models of molecular networks

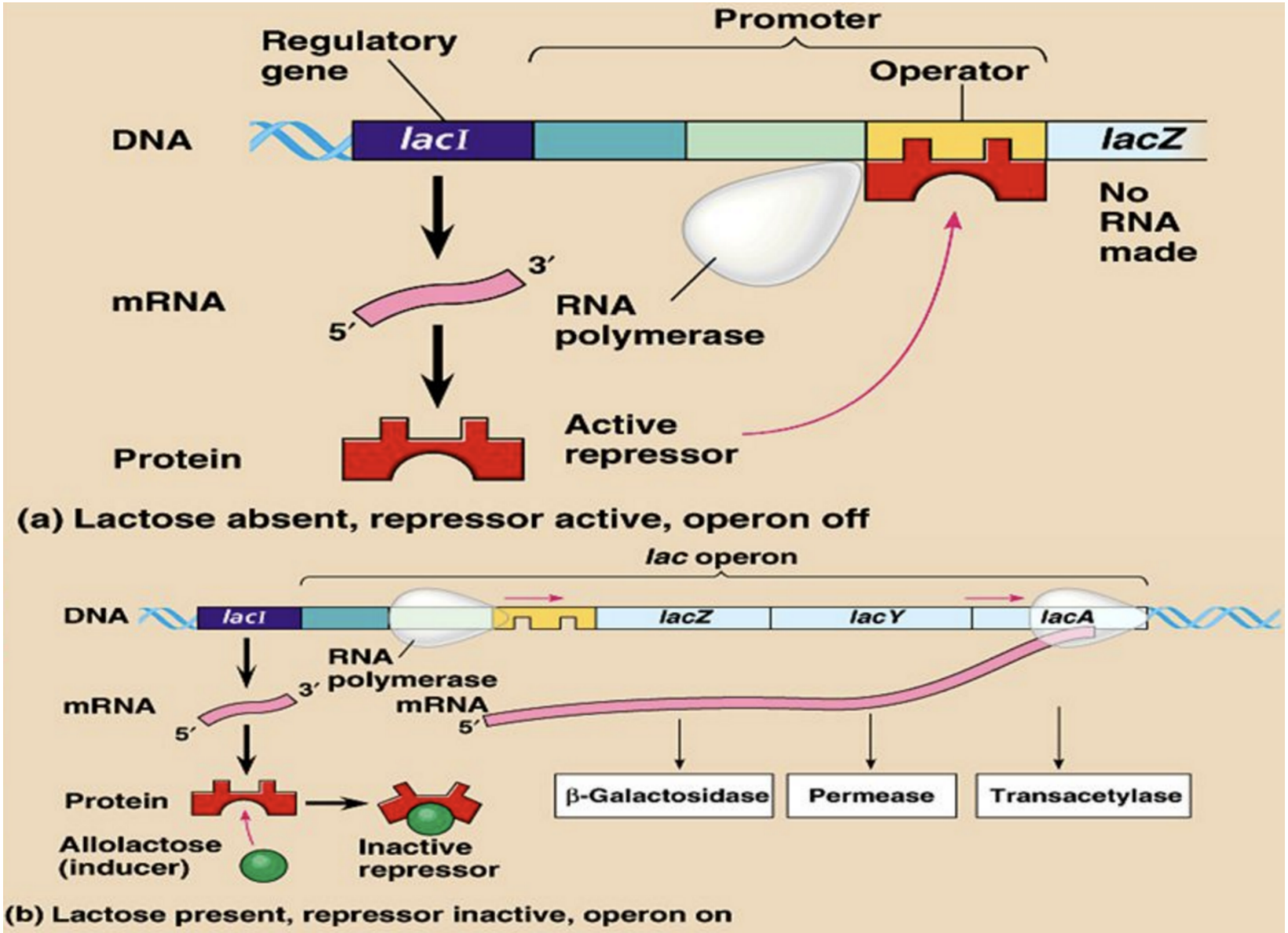


Figure: The lactose operon in *E. coli*

2. Boolean models of molecular networks

The following is a Boolean model of the lactose (*lac*) operon in *E. coli*:

$$\begin{aligned} x_1(t+1) &= x_3(t) \\ x_2(t+1) &= x_1(t) \\ x_3(t+1) &= (x_2(t) \wedge L_m) \vee L \vee (x_3(t) \wedge \overline{x_2(t)}) \end{aligned}$$

Time is discretized, and $x_1(t)$, $x_2(t)$, and $x_3(t)$ represent mRNA, β-galactosidase, and allolactose.

There are two parameters (constants):

- L = concentration of extracellular lactose (high)
- L_m = concentration of extracellular lactose (at least medium)

To find the fixed points, we work over the quotient ring of Boolean functions:

$$\mathbb{F}_2[x_1, x_2, x_3, L, L_m]/\langle x_1^2 - x_1,\ x_2^2 - x_2,\ x_3^2 - x_3,\ L^2 - L,\ L_m^2 - L_m \rangle$$

The above system, in polynomial form, becomes the algebraic model

$$\begin{aligned} f_1 &= x_3 \\ f_2 &= x_1 \\ f_3 &= x_3 + (1 + x_3 + x_2x_3 + x_2L_m)L + x_2(x_3 + L_m) \end{aligned}$$

To find the fixed points, we need to solve the system $\{f_i - x_i = 0 \mid i = 1, 2, 3\}$.

2. Boolean models of molecular networks

The solution to the system of equations $\{f_i - x_i = 0 \mid i = 1, 2, 3\}$ is the variety of the ideal

$$I = \langle f_i - x_i \mid i = 1, 2, 3 \rangle = \langle x_3 + x_1,\ x_1 + x_2,\ (1 + x_3 + x_2x_3 + x_2L_m)L + x_2(x_3 + L_m) \rangle.$$

A computer algebra package easily computes a Gröbner basis to be

$$\mathcal{G} = \{x_1 + x_3,\ x_2 + x_3,\ (1 + L_m)x_3 + (1 + L_m)L,\ (1 + x_3)L\}.$$

Key idea

The system $\{g_i = 0 \mid i = 1, \ldots, 4\}$ has the same solutions as $\{f_i - x_i = 0 \mid i = 1, 2, 3\}$.

Let's solve this (by hand) for the initial condition $(L, L_m) = (0, 1)$, i.e., medium lactose levels.

2. Boolean models of molecular networks

Research goals and open-ended questions

Key idea

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

- Given a gene regulatory network, how can we models 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?

2. 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.

2. 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.

2. Boolean models of molecular networks

Articles (algebraic aspects)

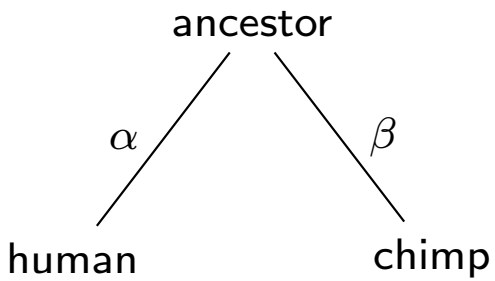
- E.S. Dimitrova, A.S. Jarrah, R. Laubenbacher, and B. Stigler. A Gröbner fan method for biochemical network modeling. In *Proc. Internat. Symposium Symb. Algebraic Comput.*, pages 122–126. ACM, 2007.
- F. Hinkelmann and R. Laubenbacher. Boolean models of bistable biological systems. *Discrete Cont. 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.

3. 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):

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

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

3. Phylogenetics and algebraic statistics

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:

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

3. Phylogenetics and algebraic statistics

Books

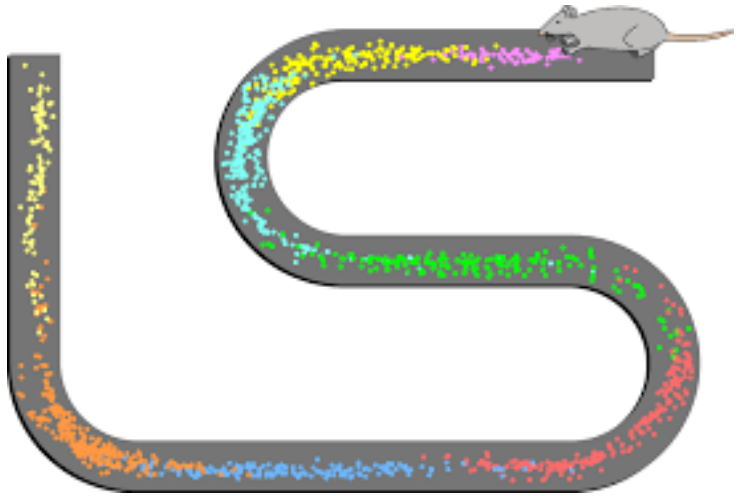
- D. Maclagan and B. Sturmfels. *Introduction to Tropical Geometry*, volume 161. American Mathematical Soc., 2015.
- L. Pachter and B. Sturmfels. *Algebraic Statistics for Computational Biology*, volume 13. Cambridge University Press, 2005.
- G. Pistone, E. Riccomagno, and H. P. Wynn. *Algebraic Statistics: 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.

4. Place fields in neuroscience

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



As an animal moves around, different subsets of neurons fire. The region that causes a specific neuron to fire is its place field.

We can encode which neurons fire with a binary string. For example, $\mathbf{c} = 10100$ means neurons 1 and 3 fire, and neurons 2, 4, and 5 are silent.

We can encode this with a pseudomonomial in $\mathbb{F}_2[x_1, x_2, x_3, x_4, x_5]$ called its characteristic polynomial:

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

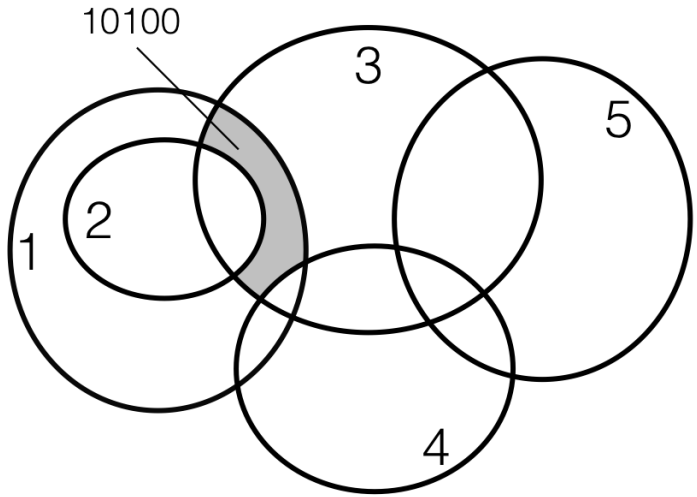
4. Place fields in neuroscience

Motivating question

Given a collection of binary strings called a neural code, reconstruct the place fields.

For example, how would you construct place fields $\mathcal{U} = \{U_1, U_2, U_3, U_4, U_5\}$ that realize the code

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



The shaded region is encoded by the pseudomonomial

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

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

It also defines a simplicial complex

$$\Delta(\mathcal{C}) := \{\sigma \subseteq [n] \mid \sigma \subseteq \text{supp}(\mathbf{c}) \text{ for some } \mathbf{c} \in \mathcal{C}\}.$$

4. 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,

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

A related object is the neural ideal, which is defined by the characteristic polynomials of the non-code words:

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

These ideals are related by

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

4. Place fields in neuroscience

Combinatorial features encoded algebraically are called recptive field (RF) relationships.

For example, note that:

- if $c_1 = c_2 = 1$ for some $\mathbf{c} \in \mathcal{C}$, then $U_1 \cap U_2 \neq \varnothing$
- if $c_1 = 1$ implies $c_2 = 1$, then $U_1 \subseteq U_2$.

The corresponding RF relationships are:

- if $(x_1 - 1)(1 - x_2) \notin J_{\mathcal{C}}$, then $U_1 \cap U_2 \neq \varnothing$
- if $x_1(1 - x_2) \in J_{\mathcal{C}}$, then $U_1 \subseteq U_2$.

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 $J_{\mathcal{C}}$.

Exercise

Show that the neural ideal of $\mathcal{C} = \{000, 111, 011, 001\}$ is

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

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

4. 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?

4. 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/`.

The theory of pseudomonomials

A squarefree monomial is a polynomial of the form $f(x) = x_{i_1} x_{i_2} \cdots x_{i_k}$.

Ideals generated by these are called squarefree monomial ideals, and are well-studied.

The polynomial

$$\chi_{\mathbf{c}}(\mathbf{x}) = x_1(x_2 - 1)x_3(x_4 - 1)(x_5 - 1) = x_1\overline{x_2}x_3\overline{x_4}\,\overline{x_5}$$

is not a monomial, but we'll call it a (squarefree) pseudomonomial.

We can define the concept of a (square-free) pseudomonomial ideal.

These have been studied since 2013. And even they can be generalized.

Definition (S. Tsiorintsoa, 2018)

A pseudomonomial is a polynomial of the form

$$x_{r_i}^{a_i} := (x_1 - r_{i1})^{a_{i1}} \cdots (x_n - r_{in})^{a_{in}}, \qquad r_{in} \in \mathbb{F},\ a_{ij} \in \mathbb{Z}^n_{\geqslant 0}$$

A pseudomonomial ideal is any ideal generated by pseudomonomials.

All of this mathematical theory has been inspired from actual biological problems!

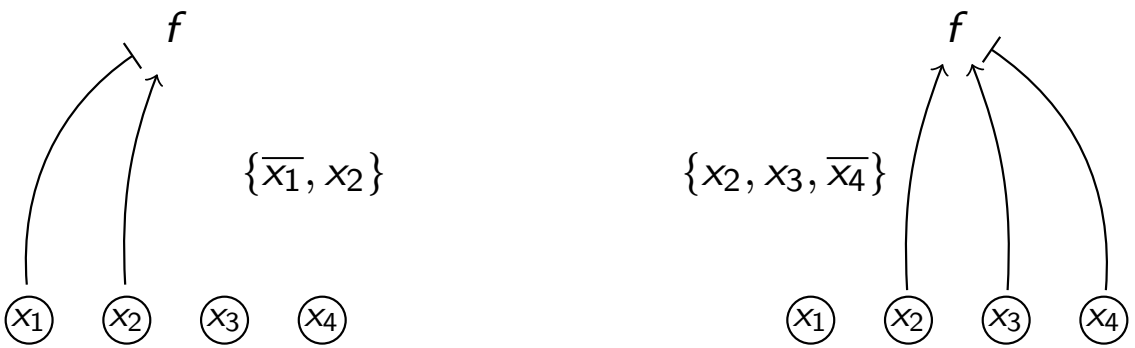
Pseudomonomials from algebraic models

There are $2^{12} = 4096$ functions $f : \mathbb{F}_2^4 \to \mathbb{F}_2$ that "*fit the data*", called the "model space".

x	1111	1110	1101	1100	1011	1010	1001	1000
$f(x)$	**0**	?	**0**	?	?	?	?	?

x	0111	0110	0101	0100	0011	0010	0001	0000
f(x)	?	**1**	?	?	?	**0**	?	?

Two of them are: $f = \overline{x_1} \wedge x_2$ and $f = x_2 \wedge x_3 \wedge \overline{x_4}$. Their wiring diagrams are:



Questions

- What (minimal) sets of variables does f have to depend on?
- Are these dependencies positive or negative?

Pseudomonomials from algebraic models

Instead of $\mathbb{F}_2 = \{0, 1\}$, we'll work over $\mathbb{F}_3 = \{0, 1, -1\}$.

We'll encode the data with a pseudomonomial ideal J and take its primary decomposition

$$J = \mathfrak{P}_1 \cap \cdots \cap \mathfrak{P}_\ell, \qquad \mathfrak{P}_i \text{ is a primary ideal.}$$

The is the analogue of factoring an integer into its prime powers. For example, $360 = 2^3 \cdot 3^2 \cdot 5$, and

$$360\mathbb{Z} = 8\mathbb{Z} \cap 9\mathbb{Z} \cap 5\mathbb{Z}.$$

Theorem (Veliz-Cuba, 2011)

Each primary component $\mathfrak{P}_i$ encodes a minimal signed wiring diagram.

Let's re-visit the partial data:

x	0010	1100	1111	0110
$f(x)$	0	0	0	1

For each $t_i < t_j$, we'll compute a pseudomonomial $p(\mathbf{s}_i, \mathbf{s}_j)$ encoding the sign of the change:

$$p(\mathbf{s}_1, \mathbf{s}_4) = x_2 - 1, \qquad p(\mathbf{s}_2, \mathbf{s}_4) = (x_1 + 1)(x_3 - 1), \qquad p(\mathbf{s}_3, \mathbf{s}_4) = (x_1 + 1)(x_4 + 1).$$

These define the ideal of signed non-disposable sets

$$J_{\triangle_{\mathcal{D}}^c} = \big\langle p(\mathbf{s}_1, \mathbf{s}_4),\ p(\mathbf{s}_2, \mathbf{s}_4),\ p(\mathbf{s}_3, \mathbf{s}_4) \big\rangle = \big\langle (x_2 - 1), (x_1 + 1)(x_3 - 1), (x_1 + 1)(x_4 + 1) \big\rangle$$

Signed min-sets, formally

In coordinate i, we're representing

- a positive change (activation) with $(x_i - 1)$
- a negative change (inhibition) with $(x_i + 1)$.

Encode the change in the i^{th} coordinate of $\mathbf{s}$ and $\mathbf{s}'$ as:

$$\partial_i(\mathbf{s}, \mathbf{s}') = \begin{cases} 1 & s_i < s_i' \\ -1 & s_i > s_i' \\ 0 & s_i' = s_i \end{cases}$$

Definition / theorem

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

$$J_{\triangle_{\mathcal{D}}^c} = \langle p(\mathbf{s}, \mathbf{s}') \mid t < t' \rangle \quad \text{where} \quad p(\mathbf{s}, \mathbf{s}') = \prod_{s_i \neq s_i'} \big(x_i - \partial(\mathbf{s}, \mathbf{s}')\big).$$

The primary components of $J_{\triangle_{\mathcal{D}}^c}$ are signed min-sets of $\mathcal{D}$.

Computing the signed min-sets

We can compute the primary decomposition in Macaulay2 (or Singular, Sage, etc.):

```
R = ZZ/3[x1,x2,x3,x4]
J_nonDisp = ideal(x2-1,(x1+1)*(x3-1),(x1+1)*(x4+1))
primaryDecomposition J_nonDisp
```

There are two primary components:

$$\begin{aligned} J_{\triangle_{\mathcal{D}}^c} &= \big\langle (x_2-1), (x_1+1)(x_3-1), (x_1+1)(x_4+1) \big\rangle \\ &= \big\langle x_1-1,\, x_2+1 \big\rangle \cap \big\langle x_2-1,\, x_3-1,\, x_4+1 \big\rangle. \end{aligned}$$

The signed min-sets are $\{x_1, \overline{x_2}\}$ and $\{x_2, x_3, \overline{x_4}\}$.

Thus, any function f that fits the data

x	0010	1100	1111	0110
$f(x)$	0	0	0	1

must depend:

- positively on x_1 and negatively on x_2, or
- positively on x_2 and x_3, and negatively on x_4