**Read**: Chapter 1: Mechanisms of gene regulation: Boolean network models of the lactose operon in *Escherichia coli*, by R. Robeva, B. Kirkwood, and R. Davis, pages 1–35.

**Do**: Create an account on the Sage Math Cloud (`http://www.sagemath.com`).

**Exercises**.

1. Given a Boolean network $(f_1, f_2, f_3, f_4, f_5)$, suppose that $\{x_1+x_5,\ x_3+1,\ x_2+x_5+1,\ x_4\}$ is a Gröbner basis of the ideal $I = \langle f_i + x_i \mid i = 1, \ldots, 5\rangle$. What, if anything, can you deduce from this?

2. Consider the following system of polynomial equations:

$$\begin{aligned} x^2 + y^2 + xyz &= 1 \\ x^2 + y + z^2 &= 0 \\ x - z &= 0 \end{aligned}$$

   To compute a Gröbner basis for this system over $\mathbb{Q}$, type the following commands into Sage, one-by-one, and press Shift+Enter after each one:

```
P.<x,y,z> = PolynomialRing(QQ, 3, order='lex'); P
I = ideal(x^2+y^2+x*y*z-1, x^2+y+z^2, x-z); I
B = I.groebner_basis(); B
```

   (a) For the system above, use the Gröbner basis you just computed to write a simpler system of polynomial equations that has the same set of solutions. Solve that system *by hand* (it's not hard) to find all solutions to the original system.

   (b) Next, solve the original system but over the binary field, $\mathbb{F}_2 = \{0, 1\}$. For this, you need to replace `QQ` with `GF(2)` in the Sage code.

   (c) Now, solve the original system but over the ternary field, $\mathbb{F}_3 = \{0, 1, 2\}$.

3. Repeat the previous problem for this system of polynomial equations:

$$\begin{aligned} x^2y - z^3 &= 0 \\ 2xy - 4z &= 1 \\ z - y^2 &= 0 \\ x^3 - 4yz &= 0 \end{aligned}$$

4. Consider the following simple model of the *lac* operon:

$$\begin{aligned} f_M &= \overline{R} & \qquad f_R &= \overline{A} \\ f_P &= M & \qquad f_A &= L \wedge B \\ f_B &= M & \qquad f_L &= P \end{aligned}$$

   For this problem, make the convention that $(x_1, x_2, x_3, x_4, x_5, x_6) = (M, P, B, R, A, L)$.

(a) Justify each function in a single sentence. What other assumptions are made in this model? (E.g., presence or absense of external lactose and glucose?)

(b) Write each function as a polynomial over $\mathbb{F}_2 = \{0, 1\}$. Then, write out the system of equations $\{f_i + x_i = 0,\ i = 1, \ldots, 6\}$, whose solutions are the fixed points of the Boolean network.

(c) Go into Sage and type the following command:

```
P.<x1,x2,x3,x4,x5,x6> = PolynomialRing(GF(2), 6, order='lex'); P
```

Now, define an ideal $I$ generated by the six polynomials, $f_i + x_i$, from Part (b). Use Sage to compute the Gröbner basis of this ideal.

(d) The Gröbner basis describes a simpler system of equations with the same solutions as the original. Write out this system and then solve it by hand to determine the fixed points of the Boolean network.

(e) Compute the entire phase space of your model with the help of either the *Analysis of Dynamic Algebraic Models* (ADAM) toolbox, at `http://adam.plantsimlab.org/`, or *TURING: Algorithms for computation with finite dynamical systems*, at `http://www.discretedynamics.org`. Are there any periodic points that are not fixed points?

5. Consider a Boolean model of the *lac* operon, based on five variables: mRNA ($M$), $\beta$-galactosidase ($B$), allolactose ($A$), intracellular lactose ($L$), and *lac* permease ($P$), and the following transition functions:

$$\begin{aligned} f_M &= A \\ f_B &= M \\ f_A &= A \vee (L \wedge B) \\ f_L &= P \vee (L \wedge \overline{B}) \\ f_P &= M \end{aligned}$$

This model does not have any parameters – it assumes that extracellular lactose is always available and extracellular glucose is always unavailable, and thus it is only able to describe the behavior of the system under the conditions.

(a) Sketch the wiring diagram for this model.

(b) Print out or draw the state space for this model using ADAM or TURING.

(c) Use Sage to find the fixed points. *Hint*: There are three of them.

(d) Explain why two of the fixed points are biologically reasonable, and why the third fixed point does not make sense biologically.

(e) Since the dynamics do not accurately reflect the behavior of the biological system it is meant to model, something is wrong. For each function, decide if it accurately reflects the underlying biology and/or the model assumptions.

(f) Propose a modification of the transition functions aimed at eliminating the biologically infeasible fixed point. Give the rationale for your modification and specify the biological mechanism or model assumptions that justify the change.

(g) Draw the wiring diagram and print and state space of your modified model. Use either the ADAM or TURING software.

(h) Analyze your model. How many fixed points are there? Do they all correspond to biologically realistic situations? Note that there should be no limit cycles of size $k \geq 2$.

6. An alternative to using Sage to compute the fixed points is to use *Macaulay2*, an open source software system for computational algebraic geometry and commutative algebra. This can be accessed online at `http://web.macaulay2.com`, or from within Sage, by typing:

```
%default_mode macaulay2
```

The following three commands in Macaulay2 perform the same task as the three Sage commands from Exercise 1. Adding a semi-colon at the end of each line will suppress the output.

```
P = QQ[x,y,z]
I = ideal(x^2+y^2+x*y*z-1, x^2+y+z^2, x-z)
G = gens gb I
```

For each of the previous problems, convert your Sage code into Macaulay2 code and check that you get the same results. Finite fields are entered as e.g., `ZZ/2` instead of `GF(2)`.

**Summary of relevant literature**.

Francois Jacob and Jacques Monod won the 1965 Nobel Prize in medicine for discovering the *lac* operon [JPSM60]. This has since been modeled with differential equations in [BM85], [YM03], in [YSHM04]. More recently, the first Boolean network model of it was published in [VCS11].

Boolean networks were first proposed as models of gene regulatory networks in 1969 by Stuart Kauffman [Kau69]. A few years later, René Thomas developed the framework of "logical models" [Tho73], and his collaborators have been studying them ever since. See also [Td90] for a book on the topic. These are like Boolean networks but the main differences are that the functions are updated asynchronously, and the set of states of the nodes can be different, e.g., $\{0, 1, 2\}$ or $\{0, 1, 2, 3\}$.

The paper [DJ02] contains a nice survey of modeling gene regulatory networks and [Alb04] contains a survey of Boolean network modeling. The paper [LS09] from the *Mathematical Monthly* gives an overview of the new field of *Algebraic Biology*.

Freely available online software for include the *Analysis of Dynamical and Algebraic Models* (ADAM) [HBG+11], which has since been improved by the crowd-sourced TURING [HL16]. The software *Gene Interaction Network simulation* (GINsim) [CNT12] is a tool for logical networks.

# References


[Alb04] R. Albert. Boolean modeling of genetic regulatory networks. In *Complex Networks*, pages 459–481. Springer, 2004.

[BM85] S. Busenberg and J. Mahaffy. Interaction of spatial diffusion and delays in models of genetic control by repression. *J. Math. Biol.*, 22(3):313–333, 1985.

[CNT12] C. Chaouiya, A. Naldi, and D. Thieffry. Logical modelling of gene regulatory networks with GINsim. *Bacterial Molecular Networks: Methods and Protocols*, pages 463–479, 2012.

[DJ02] H. De Jong. Modeling and simulation of genetic regulatory systems: a literature review. *J. Comp. Biol.*, 9(1):67–103, 2002.

[HBG+11] F. Hinkelmann, M. Brandon, B. Guang, R. McNeill, G. Blekherman, A. Veliz-Cuba, and R. Laubenbacher. ADAM: analysis of discrete models of biological systems using computer algebra. *BMC Bioinformatics*, 12(1):295, 2011.

[HL16] A. Honsy and R. Laubenbacher. TURING: Algorithms for computation with finite dynamical systems. Published electronically at `http://www.discretedynamics.org/`, 2016.

[JPSM60] F. Jacob, D. Perrin, C. Sánchez, and J. Monod. L'opéron: groupe de gènes à expression coordonnée par un opérateur. *C.R. Acad. Sci.*, 250:1727–1729, 1960.

[Kau69] S.A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.*, 22(3):437–467, 1969.

[LS09] R. Laubenbacher and B. Sturmfels. Computer algebra in systems biology. *Amer. Math. Monthly*, pages 882–891, 2009.

[Td90] R. Thomas and R. d'Ari. *Biological feedback*. CRC press, 1990.

[Tho73] R. Thomas. Boolean formalization of genetic control circuits. *J. Theor. Biol.*, 42(3):563–585, 1973.

[VCS11] A. Veliz-Cuba and B. Stigler. Boolean models can explain bistability in the lac operon. *Journal of Computational Biology*, 18(6):783–794, 2011.

[YM03] N. Yildirim and M.C. Mackey. Feedback regulation in the lactose operon: a mathematical modeling study and comparison with experimental data. *Biophysical J.*, 84(5):2841–2851, 2003.

[YSHM04] N. Yildirim, M. Santillan, D. Horike, and M.C. Mackey. Dynamics and bistability in a reduced model of the lac operon. *Chaos*, 14(2):279–292, 2004.