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 either the Sage Notebook (`http://www.sagemath.org/`) or on the Clemson Sage Server (`https://sage.math.clemson.edu:34567`). Email me your username and I will share with you a Sage worksheet containing the 9-variable Boolean model of the *lac* operon that appears below in Problem 3.

1. Consider the first model for the *lac* operon that we saw:

$$\begin{aligned} f_M &= \overline{G_e} \wedge (L \vee L_e) \\ f_E &= M \\ f_L &= \overline{G_e} \wedge ((E \wedge L_e) \vee (L \wedge \overline{E})). \end{aligned}$$

   Convert these equations into polynomials over $\mathbb{F}_2 := \{0, 1\}$. Recall that $x \wedge y = xy$, $x \vee y = x + y + xy$, and $\overline{x} = 1 + x$. Fully simplify your answers.

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, type the following commands into Sage, one-by-one, and press Shift+Enter after each one:

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

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

3. Repeat the steps of 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. Recall the model of Stigler and Veliz-Cuba for the *lac* operon, where a global system state is a 9-varible Boolean vector: $(M, P, B, C, R, A, A_\ell, L, L_\ell) = (x_1, x_2, \ldots, x_9)$:

$$\begin{array}{ll} f_M = \overline{R} \wedge C & \\ f_P = M & f_B = M \\ f_C = \overline{G_e} & f_R = \overline{A} \wedge \overline{A_\ell} \\ f_A = L \wedge B & f_{A_\ell} = A \vee L \vee L_\ell \\ f_L = \overline{G_e} \wedge P \wedge L_e & f_{L_\ell} = \overline{G_e} \wedge (L \vee L_e) \end{array}$$

   Give a well-written one sentence justification for each function. For example, $f_M = \overline{R} \wedge C$ could be: "mRNA is produced if the *lac* repressor protein is absent and the concentration of the catabolite activator protein CAP is high."

5. Do *one* of the following problems:

   (A) In the 9-variable Boolean model from the previous problem, there is no variable to represent the cAMP receptor protein *cmp*. Could you justify this decision? Propose a new model with one additional variable C_{AMP} that represents this. Use Gröbener bases and Sage to compute the fixed points of this model for each of the four possibilities for (L_e, G_e), external lactose and glucose. Are they biologically feasible? Do you notice any qualitative difference between this model and the original?

   (B) Modify the model from the previous problem by removing the variables L_ℓ and A_ℓ, which represent low concentrations of intracellular lactose and glucose, respectively. Use Gröbener bases and Sage to compute the fixed points of this model for each of the four possibilities for (L_e, G_e), external lactose and glucose. Do the fixed points make sense biologically? Since the state space has size $2^7 = 128$, the Discrete Visualizer of Dynamics program (`http://dvd.vbi.vt.edu/`) should be able to handle it. Are there any periodic cycles that are not fixed points? Do you notice any shortcomings of this new model?

   (C) Recall the first and most basic model of the *lac* operon that we saw, which is also in Problem 1:

$$f_M = \overline{G_e} \wedge (L \vee L_e), \qquad f_E = M, \qquad f_L = \overline{G_e} \wedge ((E \wedge L_e) \vee (L \wedge \overline{E})).$$

   Instead of introducing variables L_ℓ and A_ℓ that represent low concentration levels of *intracellular* lactose and allolactose (and thus allowing zero, low, and high levels in your model), propose a model that has variables $L_{e,\ell}$ and $G_{e,\ell}$ representing low concentration levels of *external* lactose and glucose. You can incorporate other elements if you want. Use Gröbener bases and Sage to compute the fixed points of this model. Are you results biologically reasonable? If not, explain why.