

# Reduction of Boolean network models

Matthew Macauley

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

Math 4500, Spring 2022

## Motivating example: Th-cell differentiation

White blood cells or *leukocytes* are in the immune system and fight diseases and infections.

One subtype are the *lymphocytes*, which includes the natural killer (NK) cells, B cells, and T cells, all which have different cellular functions.

The T-cells circulate throughout our bodies in the lymph fluid, looking for cellular abnormalities, infections, and diseases.

Helper T-cells (Th-cells) are a certain type of T-cells. They begin as *naïve*, or *Th0 cells*, and then differentiate into one of two phenotypes:

1. Type 1 are the *Th1 cells* which fight intracellular bacteria and protozoa.
2. Type 2 are the *Th2 cells* which fight extracellular parasites.

Malfunctions of immune responses involving Th1 phenotypes can result in autoimmune diseases, whereas malfunctions involving Th2 phenotypes can result in allergic reactions.

The biochemical signals that determine Th1 and Th2 differentiation act as a bistable switch, which permits either GATA3 or T-bet to be expressed, but not both. This was modeled using a 23-node Boolean network in Mendoza et. al (2006).



## Analyzing the T helper cell model

The 23-node Th-cell Boolean model has a state space of size  $2^{23} = 8,388,608$  nodes.

A Boolean model of the segment polarity genes in *Drosophila melanogaster* has 60 nodes, and a state space of size  $2^{60} \approx 1.15 \times 10^{18}$ .

In the previous lecture, we modeled time-delays and dilution & degradation by adding a number of Booleans variables.

This can causes the state space to grow enormously, though in many cases, this shouldn't affect the qualitative nature of the dynamics.

In this lecture, we'll see how large Boolean models can be “reduced” to much smaller models in a way that preserves the key feature such as fixed points.

## A motivating example

### Toy model of the *lac* operon

$$f_M = \bar{R}$$

$$f_P = M$$

$$f_B = M$$

$$f_R = \bar{A}$$

$$f_A = L \wedge B$$

$$f_L = P$$

$R$  inhibits mRNA transcription

$P$  is translated from *lac* mRNA

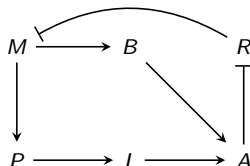
$B$  is translated from *lac* mRNA

$A$  inactivates the repressor protein

$A$  is synthesized by lactose and  $\beta$ -galactosidase

*lac* permease transports lactose into the cell

Here is the **wiring diagram**:



We won't show the state space because it's large (64 nodes), but it has two fixed points, both of which are biologically reasonable:

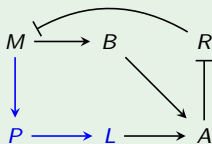
$$(M, P, B, R, A, L) = (0, 0, 0, 1, 0, 0) \quad \text{and} \quad (1, 1, 1, 0, 1, 1).$$

Our goal is to “reduce” this model in a way that in some senses, *preserves the fixed points*.

## A motivating example (cont.)

Let's eliminate  $P$

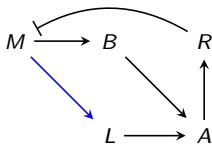
$$\begin{aligned}f_M &= \bar{R} \\f_P &= M \\f_B &= M \\f_R &= \bar{A} \\f_A &= L \wedge B \\f_L &= P\end{aligned}$$



At equilibrium,  $P(t) = P(t+1) = f_P(x(t)) = M(t)$ .

Thus, we can replace every instance of  $P$  with  $M$ :

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



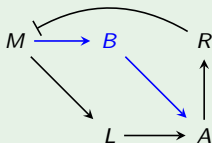
This reduced network has two fixed points:  $(M, B, R, A, L) = (0, 0, 1, 0, 0), (1, 1, 0, 1, 1)$ .

Since  $P = M$ , we can recover the fixed points of the original network.

## A motivating example (cont.)

Let's eliminate  $B$

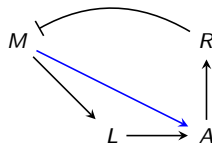
$$\begin{aligned}f_M &= \bar{R} \\f_B &= M \\f_R &= \bar{A} \\f_A &= L \wedge B \\f_L &= M\end{aligned}$$



At equilibrium,  $B = f_B = M$ , so we can replace every instance of  $B$  with  $M$ .

Thus, we can replace every instance of  $P$  with  $M$ :

$$\begin{aligned}f_M &= \bar{R} \\f_B &= \cancel{M} \\f_R &= \bar{A} \\f_A &= L \wedge M \\f_L &= M\end{aligned}$$



This reduced network has two fixed points:  $(M, R, A, L) = (0, 1, 0, 0)$ ,  $(1, 0, 1, 1)$ .

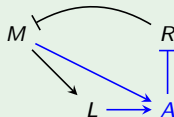
Since  $B = P = M$ , we can recover the fixed points of the original network.

$$(M, P, B, R, A, L) = (M, M, M, \bar{A}, L \wedge M, M) = (0, 0, 0, 1, 0, 0), \quad \text{and} \quad (1, 1, 1, 0, 1, 1).$$

## A motivating example (cont.)

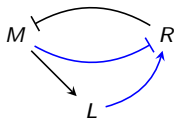
### Let's eliminate $A$

$$\begin{aligned}f_M &= \overline{R} \\f_R &= \overline{A} \\f_A &= L \wedge M \\f_L &= M\end{aligned}$$



At equilibrium,  $A = f_A = L \wedge M$ , so we can replace every instance of  $A$  with  $L \wedge M$ :

$$\begin{aligned}f_M &= \overline{R} \\f_R &= \overline{L \wedge M} = \overline{L} \vee \overline{M} \\f_A &= L \wedge M \\f_L &= M\end{aligned}$$



There are two fixed points of this reduced network:  $(M, R, L) = (0, 1, 0)$ ,  $(1, 0, 1)$ .

Since  $B = P = M$ ,  $A = L \wedge M$ , we can recover the fixed points of the original network by *back-substituting*.

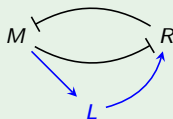
$$(M, P, B, R, A, L) = (M, M, M, R, L \wedge M, L) = (0, 0, 0, 1, 0, 0), \quad \text{and} \quad (1, 1, 1, 0, 1, 1).$$



## A motivating example (cont.)

Let's eliminate  $L$

$$\begin{aligned}f_M &= \bar{R} \\f_R &= \bar{L} \vee \bar{M} \\f_L &= M\end{aligned}$$



At equilibrium,  $L = f_L = M$ , so we can replace every instance of  $L$  with  $M$ :

$$\begin{aligned}f_M &= \bar{R} \\f_R &= \bar{M} \vee \bar{M} = \bar{M} \\f_L &= M\end{aligned}$$



There are two fixed points of this reduced network  $(M, R) = (0, 1), (1, 0)$ .

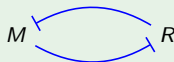
Since  $L = B = P = M$  and  $A = L \wedge M = M$ , we can recover the steady-states of the original network by *back-substituting*.

$$(M, P, B, R, A, L) = (M, M, M, R, M, M) = (0, 0, 0, 1, 0, 0), \quad \text{and} \quad (1, 1, 1, 0, 1, 1).$$

## A motivating example (cont.)

Let's eliminate  $R$

$$\begin{aligned}f_M &= \bar{R} \\ f_R &= \bar{M}\end{aligned}$$



At equilibrium,  $R = f_R = \bar{M}$ , so we can replace every instance of  $R$  with  $\bar{M}$ :

$$\begin{aligned}f_M &= \bar{\bar{M}} = M \\ f_R &= \bar{M}\end{aligned}$$



There are two fixed points of this reduced network  $M = 0, 1$ .

Since  $L = B = P = M$  and  $A = L \wedge M = M$ , we can recover the steady-states of the original network by *back-substituting*.

$$(M, P, B, R, A, L) = (M, M, M, \bar{M}, M, M) = (0, 0, 0, 1, 0, 0), \quad \text{and} \quad (1, 1, 1, 0, 1, 1).$$

# General reduction

## Reduction steps

1. Simplify the Boolean functions and wiring diagram.
  - 1.1 Reduce / simplify Boolean expressions using Boolean algebra.
  - 1.2 Remove unnecessary edges from the wiring diagram.
2. Delete vertices  $x_i$  with no self-loop (equivalently,  $f_{x_i}$  doesn't depend on  $x_i$ ), by doing the following:
  - 2.1 For all vertices  $y$  such that  $x_i \rightarrow y$ , substitute  $f_{x_i}$  into  $x_i$ :

$$f_y(x_1, \dots, \underbrace{\dots x_i \dots}_{\text{pos. } x_i}, \dots, x_n) \quad \text{becomes} \quad f_y((x_1, \dots, \underbrace{\dots f_{x_i} \dots}_{\text{pos. } x_i}, \dots, x_n).$$

- 2.2 Replace edges  $v \rightarrow x_i \rightarrow y$  by  $v \rightarrow y$  and remove  $x_i$  (and all edges to/from  $x_i$ ).

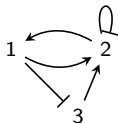
## Exercise

In Step 2.2 above, how should you replace edges of the form:

- $v \rightarrow x_i \rightarrow y$
- $v \rightarrow x_i \rightarrow y$
- $v \rightarrow x_i \rightarrow y$

## General reduction: an example

Consider the Boolean network  $f(x) = (x_2, (x_1 \wedge x_3) \vee \bar{x}_2, \bar{x}_1)$ .



Let's remove  $x_3 = \bar{x}_1$ . The new Boolean functions are

$$\begin{aligned}h_1(x_1, x_2) &= f_1(x_1, x_2, x_3) = f_1(x_1, x_2, \bar{x}_1) = x_2, \\h_2(x_1, x_2) &= f_2(x_1, x_2, x_3) = f_2(x_1, x_2, \bar{x}_1) = (x_1 \wedge \bar{x}_1) \vee \bar{x}_2\end{aligned}$$

However,  $x_1 \wedge \bar{x}_1 = 0$ , and so

$$h_2(x_1, x_2) = (x_1 \wedge \bar{x}_1) \vee \bar{x}_2 = 0 \vee \bar{x}_2 = \bar{x}_2.$$

The reduced Boolean network is thus  $h(x_1, x_2) = (x_2, \bar{x}_2)$



To find the fixed points, we must solve the system  $h_i = x_i$  for  $i = 1, 2$ :

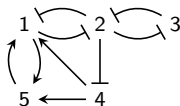
$$\begin{cases} h_1(x_1, x_2) = x_2 = x_1 \\ h_2(x_1, x_2) = \bar{x}_2 = x_2. \end{cases}$$

Since  $x_2 \neq \bar{x}_2$ , there are no fixed points in the reduced BN, and thus none in the original BN.

## General reduction: an example

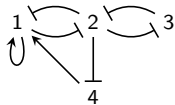
Consider the Boolean network:

$$f = (x_5 \vee \bar{x}_2 \vee x_4, \bar{x}_1 \wedge \bar{x}_3, \bar{x}_2, \bar{x}_2, x_1 \vee x_4).$$



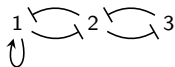
Remove  $x_5 = x_1 \vee x_4$ :

$$\begin{aligned} f &= ((x_1 \vee x_4) \vee \bar{x}_2 \vee x_4, \bar{x}_1 \wedge \bar{x}_3, \bar{x}_2, \bar{x}_2) \\ &= (x_1 \vee \bar{x}_2 \vee x_4, \bar{x}_1 \wedge \bar{x}_3, \bar{x}_2, \bar{x}_2). \end{aligned}$$



Remove  $x_4 = \bar{x}_2$ :

$$f = (x_1 \vee \bar{x}_2 \vee \bar{x}_2, \bar{x}_1 \wedge \bar{x}_3, \bar{x}_2) = (x_1 \vee \bar{x}_2, \bar{x}_1 \wedge \bar{x}_3, \bar{x}_2)$$



Remove  $x_3 = \bar{x}_2$ :

$$f = (x_1 \vee \bar{x}_2, \bar{x}_1 \wedge \bar{\bar{x}}_2) = (x_1 \vee \bar{x}_2, \bar{x}_1 \wedge x_2)$$



This yields the system:

$$\begin{cases} h_1(x_1, x_2) = x_1 \vee \bar{x}_2 \\ h_2(x_1, x_2) = \bar{x}_1 \wedge x_2 \\ x_3 = \bar{x}_2 \\ x_4 = \bar{x}_2 \\ x_5 = x_1 \vee x_4 \end{cases}$$

The reduced system  $(h_1, h_2)$  has 2 fixed points:

$$(x_1, x_2) = (1, 0), (0, 1).$$

Thus, the original system has two fixed points:

$$(x_1, x_2, x_3, x_4, x_5) = (1, 0, 1, 1, 1), (0, 1, 0, 0, 0).$$

## Boolean model reduction in Macaulay2

We want polynomials in variables  $x_1, \dots, x_5$ , over the field  $\mathbb{F}_2$ , and  $x_i^2 = x_i$ :

```
R = ZZ/2[x1,x2,x3,x4,x5] / ideal(x1^2-x1, x2^2-x2, x3^2-x3, x4^2-x4, x5^2-x5);
```

For convenience, let's define  $a|b := a + b + ab$  and  $a\&b := a * b$ :

```
RingElement | RingElement :=(x,y)->x+y+x*y;  
RingElement & RingElement :=(x,y)->x*y;
```

Input the Boolean network  $f = (f_1, f_2, f_3, f_4, f_5) = (x_5 \vee \bar{x}_2 \vee x_4, \bar{x}_1 \wedge \bar{x}_3, \bar{x}_2, \bar{x}_2, x_1 \vee x_4)$ :

```
f1 = x5 | (1+x2) | x4;  
f2 = (1+x1) & (1+x3);  
f3 = 1+x2;  
f4 = 1+x2;  
f5 = x1 | x4;
```

Typing  $(f1, f2, f3, f4, f5)$  gives the following output:

```
(x2x4x5+x2x4+x2x5+x2+1, x1x3+x1+x3+1, x2+1, x2+1, x1x4+x1+x4)
```

## Boolean model reduction in Macaulay2

Let's eliminate  $x_5 = x_1 \vee x_4$ , which appears in the function  $f_1$ .

```
f1=sub(f1,{x5=>f5});
```

Now, typing  $(f_1, f_2, f_3, f_4)$  gives the following output:

```
(x1x2x4+x1x2+x2x4+x2+1, x1x3+x1+x3+1, x2+1, x2+1)
```

Let's eliminate  $x_4 = \overline{x_2}$ , which appears in the function  $f_1$ .

```
f1=sub(f1,{x4=>f4});
```

Typing  $(f_1, f_2, f_3)$  gives the following output:

```
(x1x2+x2+1, x1x3+x1+x3+1, x2+1)
```

## Boolean model reduction in Macaulay2

Let's eliminate  $x_3 = \overline{x_2}$ , which appears in the function  $f_2$ .

```
f2=sub(f1,{x3=>f3});
```

Typing `(f1,f2)` gives the following output:

```
(x1x2+x2+1, x1x2+x2+1)
```

`gens gb ideal(f1+x2, f2+x2)` gives the following output:

```
(x1+x2+1)
```

which means that  $x_1 = \overline{x_2}$ .

Back-substituting  $x_3 = \overline{x_2}$ , and  $x_4 = \overline{x_2}$ , and  $x_5 = x_1 \vee x_4 = \overline{x_2} \vee \overline{x_2} = \overline{x_2}$  gives

$$(x_1, x_2, x_3, x_4, x_5) = (\overline{x_2}, x_2, \overline{x_2}, \overline{x_2}, \overline{x_2})$$

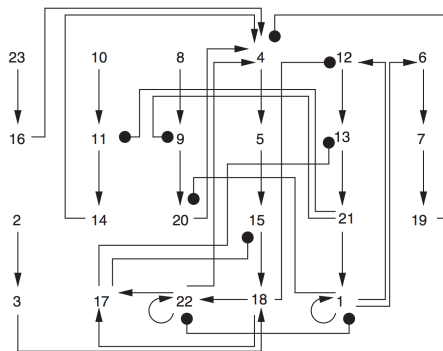
and thus we have two fixed points of our original Boolean model:

$$(x_1, x_2, x_3, x_4, x_5) = (1, 0, 1, 1, 1) \quad \text{and} \quad (0, 1, 0, 0, 0).$$

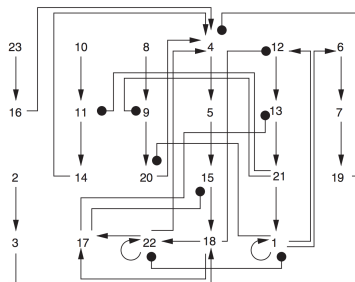


## Finding the fixed points of the T helper cell model

$x_1 = \text{GATA3}$	$f_1 = (x_1 \vee x_{21}) \wedge \overline{x_{22}}$
$x_2 = \text{IFN-}\beta$	$f_2 = 0$
$x_3 = \text{IFN-}\beta\text{R}$	$f_3 = x_2$
$x_4 = \text{IFN-}\gamma$	$f_4 = (x_{14} \vee x_{16} \vee x_{20} \vee x_{22}) \wedge \overline{x_{19}}$
$x_5 = \text{IFN-}\gamma\text{R}$	$f_5 = x_4$
$x_6 = \text{IL-10}$	$f_6 = x_1$
$x_7 = \text{IL-10R}$	$f_7 = x_6$
$x_8 = \text{IL-12}$	$f_8 = 0$
$x_9 = \text{IL-12R}$	$f_9 = x_8 \wedge \overline{x_{21}}$
$x_{10} = \text{IL-18}$	$f_{10} = 0$
$x_{11} = \text{IL-18R}$	$f_{11} = x_{10} \wedge \overline{x_{21}}$
$x_{12} = \text{IL-4}$	$f_{12} = x_1 \wedge \overline{x_{18}}$
$x_{13} = \text{IL-4R}$	$f_{13} = x_{12} \wedge \overline{x_{17}}$
$x_{14} = \text{IRAK}$	$f_{14} = x_{11}$
$x_{15} = \text{JAK1}$	$f_{15} = x_5 \wedge \overline{x_{17}}$
$x_{16} = \text{NFAT}$	$f_{16} = x_{23}$
$x_{17} = \text{SOCS1}$	$f_{17} = x_{18} \vee x_{22}$
$x_{18} = \text{STAT1}$	$f_{18} = x_3 \vee x_{15}$
$x_{19} = \text{STAT3}$	$f_{19} = x_7$
$x_{20} = \text{STAT4}$	$f_{20} = x_9 \wedge \overline{x_1}$
$x_{21} = \text{STAT6}$	$f_{21} = x_{13}$
$x_{22} = \text{T-bet}$	$f_{22} = (x_{18} \vee x_{22}) \wedge \overline{x_1}$
$x_{23} = \text{TCR}$	$f_{23} = 0$



## Reducing the T helper cell model



If we eliminate the variables in the order,  $x_{23}, x_{21}, x_{20}, \dots$ , we get

Variable	Boolean function	Polynomial function
$x_1 = \text{GATA3}$	$h_1(x_1, x_{22}) = x_1 \wedge \overline{x_{22}}$	$h_1(x_1, x_{22}) = x_1 x_{22} + x_1$
$x_{22} = \text{T-bet}$	$h_{22}(x_1, x_{22}) = \overline{x_1} \wedge x_{22}$	$h_{22} = x_1 x_{22} + x_{22}$

There are three fixed points:

- (0, 0): GATA3 and T-bet are inactive, the “signature” of Th0 cells.
- (0, 1): Only T-bet is active, the signature of Th1 cells.
- (1, 0): Only GATA3 is active, the signature of Th2-cells.

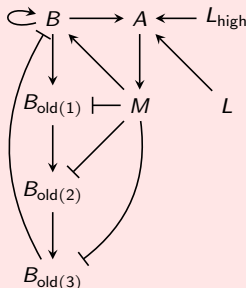
## Application: Modeling time delays and degradation & dilution

In the last lecture, we saw how to add Boolean variables to model time delays and loss of concentration due to degradation / dilution.

Consider the following model of the *lac* operon (slightly modified from last lecture) that assumes that  $\beta$ -galactosidase takes several time-steps to degrade.

### Example model

$$\left\{ \begin{array}{l} f_M = A \\ f_A = (B \wedge L) \vee L_{\text{high}} \\ f_B = M \vee (B \wedge \overline{B_{\text{old}(3)}}) \\ f_{B_{\text{old}(1)}} = \overline{M} \wedge B \\ f_{B_{\text{old}(2)}} = \overline{M} \wedge B_{\text{old}(1)} \\ f_{B_{\text{old}(3)}} = \overline{M} \wedge B_{\text{old}(2)} \end{array} \right.$$



Do you see why the precise number of  $B_{\text{old}(i)}$  variables is unimportant, regarding the number and qualitative nature of the fixed points?