Algebraic models and finite dynamical systems

Matthew Macauley

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

Algebraic Systems Biology

Motivation

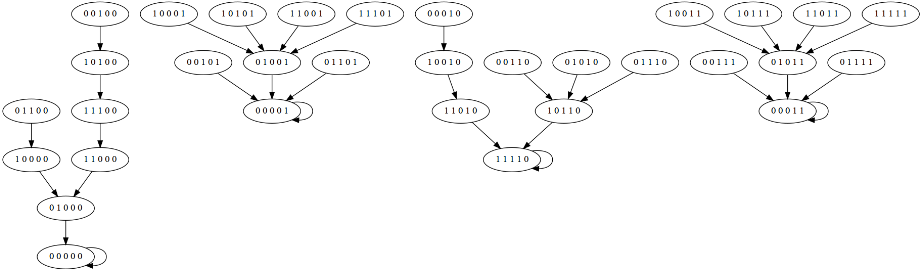
Recall our toy model for the *lac* operon, with $(x_1, x_2, x_3, x_4, x_5) = (M, E, L, L_e, G_e)$.

$$(f_1, f_2, f_3, f_4, f_5) \quad = \quad \big(\overline{x_5} \wedge (x_3 \vee x_4), \quad x_1, \quad \overline{x_5} \wedge \big[(x_2 \wedge x_4) \vee (\overline{x_2} \wedge x_3)\big], \quad x_4, \quad x_5\big)$$

If we update these functions synchronously, we get a dynamical system map

$$f \colon \mathbb{F}_2^5 \longrightarrow \mathbb{F}_2^5, \qquad x := (x_1, x_2, x_3, x_4, x_5) \longmapsto \big(f_1(x),\ f_2(x),\ f_3(x),\ f_4(x),\ f_5(x)\big).$$

This can be visualized by the (synchronous) state space graph:

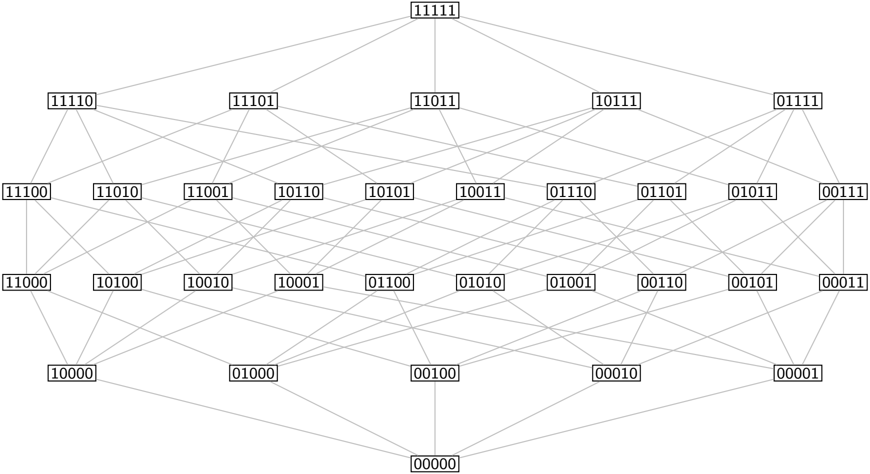


In this section, we'll formalize and study this, along with the asynchronous version.

Motivation

The asynchronous automaton is defined by updating the functions individually.

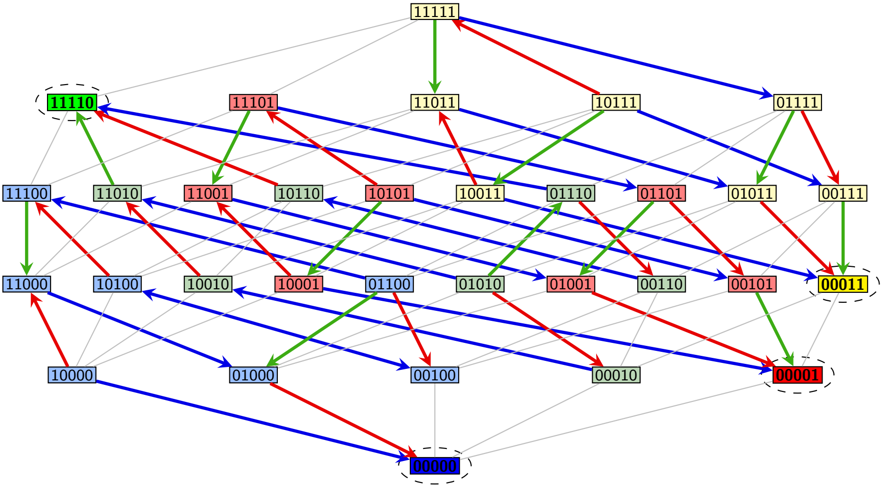
It lives on the skeleton of a Boolean lattice.



Motivation

Here is the asynchronous automaton of the following Boolean model:

$$(f_1, f_2, f_3, f_4, f_5) \quad = \quad \left(\overline{x_5} \wedge (x_3 \vee x_4), \quad x_1, \quad \overline{x_5} \wedge \left[(x_2 \wedge x_4) \vee (\overline{x_2} \wedge x_3)\right], \quad x_4, \quad x_5\right)$$



Attractors of Boolean models

Informally, an attractor is a collection of states that:

- are connected
- from which the system (if unperturbed) will never leave.

In the (synchronous) state space of a Boolean model, this is just a periodic cycle.

In the asynchronous automaton, an attractor is a terminal strongly connected component.

Biologically, attractors often correspond to

- steady-states, e.g., expression vs. non-expression of an operon,
- phenotypes, e.g., differentiated cell types,
- oscillations, e.g., cell cycles or biological rhythms.

Informally, the basin of attraction consists of the attractor, and all states that lead into it.

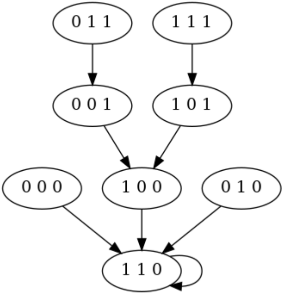
This all extends naturally to, e.g., ternary and logical models.

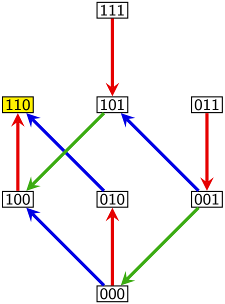
Attractors: synchronous vs. asynchronous dynamics

Consider a Boolean model:

$$(f_1, f_2, f_3) = \left(x_1 \vee \overline{x_2} \vee \overline{x_3},\;\; \overline{x_3},\;\; x_2 \wedge x_3\right).$$

The synchronous state space and asynchronous automaton are below.





There is one attractor: the fixed point $(1, 1, 0) \in \mathbb{F}_2^3$.

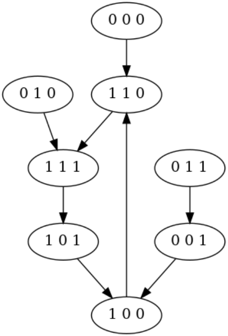
Fact: Fixed points do not depend on the update schedule. (*Why?*)

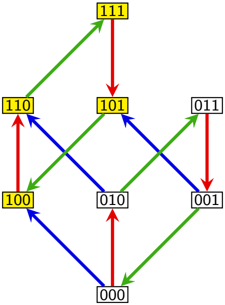
Attractors: Synchronous vs. asynchronous dynamics

Let's modify the previous example by changing $f_3 = x_2 \wedge x_3$ to $f_3 = x_2$:

$$(f_1, f_2, f_3) = \big(x_1 \vee \overline{x_2} \vee \overline{x_3},\ \ \overline{x_3},\ \ x_2\big).$$

The synchronous state space and asynchronous automaton are below.





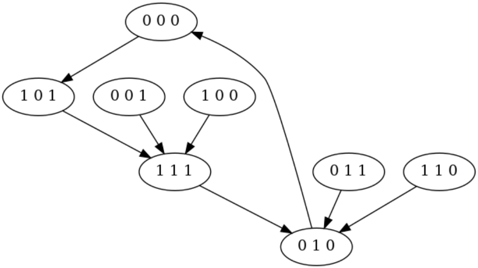
In both cases, there is one attractor: a 4-cycle.

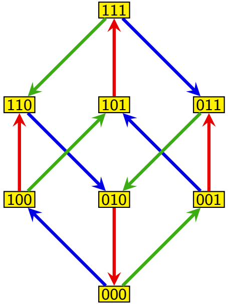
Attractors: Synchronous vs. asynchronous dynamics

Consider a Boolean model:

$$(f_1, f_2, f_3) = (\overline{x_2},\ \ x_1 \vee x_3,\ \ \overline{x_2}).$$

The synchronous state space and asynchronous automaton are below.





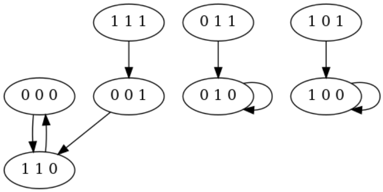
- **Synchronous**. There is one attractor: a 4-cycle.
- **Asynchronous**. There is one complex attractor of size 8.

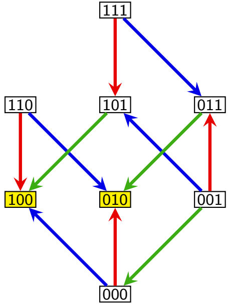
Attractors: Synchronous vs. asynchronous dynamics

The *number* of attractors depends on the update scheme. Consider a Boolean model:

$$(f_1, f_2, f_3) = \left(\overline{x_2},\ \ \overline{x_1},\ \ x_1 \wedge x_2 \wedge x_3\right).$$

The synchronous state space and asynchronous automaton are below.



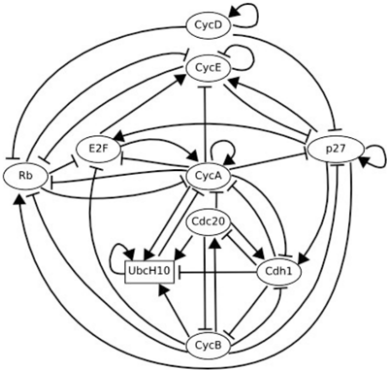


- **Synchronous**. There are three attractors: a 2-cycle, and two fixed points.
- **Asynchronous**. There is two attractors, both fixed points.

A Boolean model of the mammalian cell cycle

The following Boolean model was proposed in Fauré et al. (2006).

$$
\begin{aligned}
f_{CycD} &= CycD \\
f_{Rb} &= (\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycACycB}) \\
&\quad \vee (p27 \wedge \overline{CycD} \wedge \overline{CycB}) \\
f_{E2F} &= (\overline{Rb} \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge \overline{Rb} \wedge \overline{CycB}) \\
f_{CycE} &= E2F \wedge \overline{Rb} \\
f_{CycA} &= (E2F \wedge \overline{Rb} \wedge \overline{Cdc20} \wedge \overline{Cdh2 \wedge Ubc}) \\
&\quad \vee (CycA \wedge \overline{Rb} \wedge \overline{Cdc20} \wedge \overline{Cdh2 \wedge Ubc}) \\
f_{p27} &= (\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB}) \\
&\quad \vee (p27 \wedge \overline{CycE \wedge CycA} \wedge \overline{CycB} \wedge \overline{CycD}) \\
f_{Cdc20} &= CycB \\
f_{Cdh1} &= \overline{Cdh1} \vee (Cdh1 \wedge Ubc \wedge (Cdc20 \vee CycA \vee CycB)) \\
f_{CycB} &= \overline{Cdc20} \wedge \overline{Cdh1}
\end{aligned}
$$



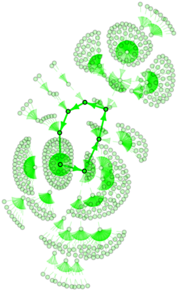
It is a Boolean version of an ODE model published by Novak and Tyson (2004).

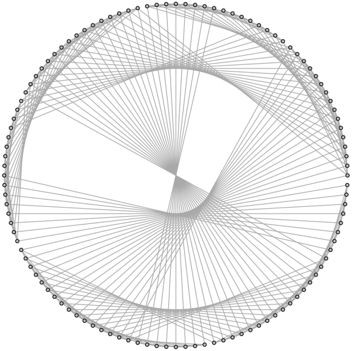
This is also the running example in the BoolNet online documentation and vignette.

A Boolean model of the mammalian cell cycle (Fauré et al., 2006)

There are two attractors in both the synchronous phase space and asynchronous automaton.

- **synchronous**: a fixed point and a 7-cycle.
- **asynchronous**: a fixed point, and a 112-node complex attractor.





Fields and rings, informally

A field is a set where we can add, subtract, multiply, and divide (except by zero).

In other words, we can do arithmetic, and the distributive law holds: $a(b+c) = ab + ac$.

We've seen that every n-variable Boolean function is a polynomial.

This holds more generally.

Fact

If K is a finite field, then every function $f\colon K^n \to K$ is a multivariate polynomial.

This allows us to frame problems involving Boolean (and ternary, etc.) networks in terms of algebraic geometry.

There is a rich toolbox of computational algebra to analyze these problems.

A ring is a set where we can add, subtract, multiply, but not necessarily divide. The distributive law also holds.

The most common rings we will see are $\mathbb{Z}$, and sets of polynomials, e.g., $K[x]$ or $K[x, y, z]$.

Fields, formally

Definition

A set $\mathbb{F}$ containing $1 \neq 0$ with addition and multiplication operations is a field if the following three conditions hold:

- $\mathbb{F}$ is an abelian group under addition.
- $\mathbb{F} \setminus \{0\}$ is an abelian group under multiplication.
- The distributive law holds: $a(b + c) = ab + ac$.

Examples

- The following sets are fields: $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{F}_p := \mathbb{Z}_p$ (prime p).
- The following sets are *not* fields: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Z}_n$ (composite n).

In this course, we will mostly deal with finite fields.

Proposition (exercise)

1. If I is an ideal of a commutative ring R, then R/I is a field iff I is maximal.
2. Any finite integral domain is a field.

Finite fields

Definition

Let $\mathbb{F}$ be a finite field. The *characteristic* of $\mathbb{F}$, denoted $\text{char}(\mathbb{F})$, is the smallest positive integer n for which $n1 := \underbrace{1 + 1 + \cdots + 1}_{n \text{ times}} = 0$.

Remarks

- It is elementary to show that $\text{char}(\mathbb{F})$ must be prime.
- $\mathbb{F}$ contains $\mathbb{F}_p = \{0, 1, \ldots, p-1\}$ as a subfield.
- $\mathbb{F}$ is a vector space over $\mathbb{F}_p$. Therefore, $|\mathbb{F}| = p^k$ for some $k \in \mathbb{Z}$.

Proposition

If K and L are finite fields with $K \subseteq L$ and $|K| = p^m$ and $|L| = p^n$, then m divides n.

Proof (sketch)

We have $\mathbb{F}_p \subseteq K \subseteq L$. Then L is not only a $\mathbb{F}_p$-vector space, but also a K-vector space.

Let $x_1, \ldots, x_k$ be a basis for L over K. Every $x \in L$ can be written uniquely as $x = a_1 x_1 + \cdots + a_k x_k$. Now count elements. $\square$

Finite fields

We know that:

- $\mathbb{Z}_p$ is a field iff p is prime,
- finite integral domains are fields,
- every finite field has order p^k.

But *what do these "other" finite fields look like?*

Let $R = \mathbb{F}_2[x]$ be the polynomial ring over $\mathbb{F}_2$. (Note: we can ignore all negative signs.)

The polynomial $f(x) = x^2 + x + 1$ is irreducible over $\mathbb{F}_2$ because it does not have a root. (Note that $f(0) = f(1) = 1 \neq 0$.)

Consider the ideal $I = \langle x^2 + x + 1 \rangle = \big\{ (x^2 + x + 1)h(x) \mid h \in \mathbb{F}_2[x] \big\}$.

In the quotient ring R/I, we have $x^2 + x + 1 = 0$, or equivalently, $x^2 = -x - 1 = x + 1$.

The quotient has only 4 elements:

$$0 + I\,, \qquad 1 + I\,, \qquad x + I\,, \qquad (x+1) + I\,.$$

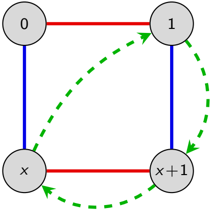
As with the quotient group (or ring) $\mathbb{Z}/n\mathbb{Z}$, we usually drop the "I", and just write

$$R/I = \mathbb{F}_2[x]/\langle x^2 + x + 1 \rangle \cong \{0,\ 1,\ x,\ x+1\}\,.$$

It is easy to check that this is a field!

The finite field of order 4

Here is a Cayley graph, and the Cayley tables for $R/I = \mathbb{F}_2[x]/\langle x^2+x+1\rangle$:



$+$	0	1	x	$x+1$
0	0	1	x	$x+1$
1	1	0	$x+1$	x
x	x	$x+1$	0	1
$x+1$	$x+1$	x	1	0

$\times$	1	x	$x+1$
1	1	x	$x+1$
x	x	$x+1$	1
$x+1$	$x+1$	1	x

Theorem

There exists a finite field $\mathbb{F}_q$ of order q, which is unique up to isomorphism, iff $q = p^k$ for some prime p. If $k > 1$, then this field is isomorphic to the quotient ring

$$\mathbb{F}_p[x]/\langle f\rangle ,$$

where f is *any* irreducible polynomial of degree k.

Much of the error correcting techniques in coding theory are built using mathematics over $\mathbb{F}_{2^8} = \mathbb{F}_{256}$. This is what allows your DVD to play despite scratches.

Finite fields

Here is the finite field of order 8: $\mathbb{F}_8 \cong R/I = \mathbb{F}_2[x]/\langle x^3+x+1\rangle$:

$+$	0	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
0	0	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
1	1	0	$x+1$	x	x^2+1	x^2	x^2+x+1	x^2+x
x	x	$x+1$	0	1	x^2+x	x^2+x+1	x^2	x^2+1
$x+1$	$x+1$	x	1	0	x^2+x+1	x^2+x	x^2+1	x^2
x^2	x^2	x^2+1	x^2+x	x^2+x+1	0	1	x	$x+1$
x^2+1	x^2+1	x^2	x^2+x+1	x^2+x	1	0	$x+1$	x
x^2+x	x^2+x	x^2+x+1	x^2	x^2+1	x	$x+1$	0	1
x^2+x+1	x^2+x+1	x^2+x	x^2+1	x^2	$x+1$	x	1	0

$\times$	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
1	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
x	x	x^2	x^2+x	$x+1$	1	x^2+x+1	x^2+1
$x+1$	$x+1$	x^2+x	x^2+1	x^2+x+1	x^2	1	x
x^2	x^2	$x+1$	x^2+x+1	x^2+x	x	x^2+1	1
x^2+1	x^2+1	1	x^2	x	x^2+x+1	$x+1$	x^2+x
x^2+x	x^2+x	x^2+x+1	1	x^2+1	$x+1$	x	x^2
x^2+x+1	x^2+x+1	x^2+1	x	1	x^2+x	x^2	$x+1$

Notice how $\mathbb{F}_2 = \{0, 1\}$ arises is a subfield, but not $\mathbb{F}_4$. (*Why?*)

Finite fields and ordering

Fields like $\mathbb{Q}$ and $\mathbb{R}$ are totally ordered: there is a natural $\leq$ operation that respects the field operations:

$$a \leq b \quad \Rightarrow \quad a + c \leq b + c, \qquad \text{and} \qquad a, b \geq 0 \quad \Rightarrow \quad ab \geq 0.$$

It is well-known that $\mathbb{C}$ cannot be totally ordered.

Proposition

Finite fields cannot be totally ordered.

In an algebraic model over $\mathbb{F}_p$, it is generally assumed that $0 < 1 < 2 < \cdots < p-1$ in $\mathbb{F}_p$.

This is generally harmless; just note that this is not an "actual" total order.

There is no canonical way, official or not, to "order" $\mathbb{F}_4 = \{0, 1, a, b\}$.

Remark

Though non-prime finite fields are generally not used in algebraic models, most of the results in the section hold for general finite fields.

Throughout, assume that $\mathbb{F}$ is a finite field of order $q = p^k$.

Polynomials vs. functions over finite fields

Let $\mathbb{F}$ be a field of order $q = p^k$. Every $f \in \mathbb{F}[x]$ defines a function $\mathbb{F} \to \mathbb{F}$, by $c \mapsto f(c)$.

For example, the following function $\mathbb{F}_5 \to \mathbb{F}_5$ is defined by the polynomial $f(x) = x^2 \in \mathbb{F}_5[x]$:

x	0	1	2	3	4
$f(x)$	0	1	4	4	1

This is called its truth table. There are exactly q^q functions $\mathbb{F} \to \mathbb{F}$. (*Why?*)

However, the set $\mathbb{F}[x]$ is infinite. For example, polynomials in $\mathbb{F}_5[x]$ include:

$$3, \qquad x^2+1, \qquad 2x^4+x, \qquad x^2, \qquad x^6, \qquad 3x^4+x^3+4x^2+4, \quad \dots$$

Thus, different polynomials can give the same function. For example, over $\mathbb{F}_2$, both x^2 and x define the same function.

Remark

The multiplicative group $\mathbb{F}^* := \mathbb{F} \setminus \{0\}$ is cyclic of order $q-1$. Thus, $a^q = a$ for all $a \in \mathbb{F}$.

This means that x^q and x define the same function over $\mathbb{F}_q$.

Polynomials vs. functions over finite fields

There are q^q functions $\mathbb{F} \to \mathbb{F}$, where $|\mathbb{F}| = q$.

Since x^q and x are the same function, every element in the quotient ring $\mathbb{F}[x]/I$, where $I = \langle x^q - x \rangle$, defines a function.

That is, there is a well-defined (1-to-1) mapping

$$\mathbb{F}[x]/I \longrightarrow \{\text{functions } \mathbb{F} \to \mathbb{F}\}, \qquad \overline{f} \longmapsto \Big\{ c \mapsto f(c) \Big\}.$$

Elements in the quotient ring $\mathbb{F}[x]/I$, where $I = \langle x^q - x \rangle$, have the form

$$a_{q-1}x^{q-1} + \cdots + a_1 x + a_0, \qquad a_i \in \mathbb{F}.$$

There are clearly q^q elements in $\mathbb{F}[x]/I$.

Thus, the function above is a bijection.

Summary

- Elements in the (infinite) ring $\mathbb{F}[x]$ are polynomials over $\mathbb{F}$.
- Elements in the (finite) quotient ring $\mathbb{F}[x]/\langle x^q - x \rangle$ are functions $\mathbb{F} \to \mathbb{F}$.

Multivariate polynomials as truth tables

- Every Boolean function on 3 variables (x, y, and z) can be written uniquely as

x	0	0	1	1	0	0	1	1
y	0	1	0	1	0	1	0	1
z	0	0	0	0	1	1	1	1
$f(x,y,z)$	a_{000}	a_{010}	a_{100}	a_{110}	a_{001}	a_{011}	a_{101}	a_{111}

Thus, there are $2^{(2^3)} = 2^8 = 256$ functions $\mathbb{F}_2^3 \to \mathbb{F}_2$.

- Every ternary function on 2 variables (x and y) can be written uniquely as

x	0	0	0	1	1	1	2	2	2
y	0	1	2	0	1	2	0	1	2
$f(x,y)$	a_{00}	a_{10}	a_{20}	a_{10}	a_{11}	a_{12}	a_{20}	a_{21}	a_{22}

Thus, there are $3^{(3^2)} = 3^9 = 19683$ functions $\mathbb{F}_3^2 \to \mathbb{F}_2$.

- Every function on 2 variables (x and y) over $\mathbb{F}_5$ can be written uniquely as

x	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
y	0	0	0	0	0	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4
$f(x,y)$	a_{00}	a_{10}	a_{20}	a_{30}	a_{40}	a_{01}	a_{11}	a_{21}	a_{31}	a_{41}	a_{02}	a_{12}	a_{22}	a_{32}	a_{42}	a_{03}	a_{13}	a_{23}	a_{33}	a_{43}	a_{04}	a_{14}	a_{24}	a_{34}	a_{44}

Thus, there are $5^{(5^2)} = 5^{25} \approx 2.980 \times 10^{17}$ functions $\mathbb{F}_5^2 \to \mathbb{F}_2$.

Multivariate functions as polynomials

- Every Boolean function on 3 variables (x, y, and z) can be written uniquely as

$$a_{000} + a_{100}x + a_{010}y + a_{001}z + a_{110}xy + a_{101}xz + a_{011}yz + a_{111}xyz, \qquad a_{ijk} \in \mathbb{F}_2.$$

  Thus, there are $2^{(2^3)} = 2^8 = 256$ functions $\mathbb{F}_2^3 \to \mathbb{F}_2$.

- Every ternary function on 2 variables (x and y) can be written uniquely as

$$a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2 + a_{21}x^2y + a_{12}xy^2 + a_{22}x^2y^2, \qquad a_{ij} \in \mathbb{F}_3.$$

  Thus, there are $3^{(3^2)} = 3^9 = 19683$ functions $\mathbb{F}_3^2 \to \mathbb{F}_2$.

- Every function on 2 variables (x and y) over $\mathbb{F}_5$ can be written uniquely as

$$\begin{aligned} &a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2 + a_{3}x^3 + a_{21}x^2y + a_{12}xy^2 + a_{30}y^3 \\ &\quad + a_{40}x^4 + a_{31}x^3y + a_{22}x^2y^2 + a_{13}xy^3 + a_{04}y^4 + a_{41}x^4y + a_{32}x^3y^2 + a_{23}x^2y^3 + a_{14}xy^4 \\ &\quad + a_{42}x^4y^2 + a_{33}x^3y^3 + a_{24}x^2y^4 + a_{43}x^4y^3 + a_{34}x^3y^4 + a_{44}x^4y^4, \qquad a_{ij} \in \mathbb{F}_5. \end{aligned}$$

  Thus, there are $5^{(5^2)} = 5^{25} \approx 2.980 \times 10^{17}$ functions $\mathbb{F}_5^2 \to \mathbb{F}_2$.

Notice how these are all elements in the size-$q^{(q^n)}$ quotient ring

$$\mathbb{F}_q[x_1, \ldots, x_n]/\langle x_1^q - x_1, \ldots, x_n^q - x_n \rangle.$$

Multivariate polynomials vs. functions over finite fields

Let $\mathbb{F}$ be a field of order $q = p^k$. Every $f \in \mathbb{F}[x_1, \dots, x_n]$ defines a function

$$\mathbb{F}^n \longrightarrow \mathbb{F}, \qquad (c_1, \dots, c_n) \longmapsto f(c_1, \dots, c_n).$$

For example, the following function $\mathbb{F}_3^2 \to \mathbb{F}_3$ is defined by $f(x, y) = x^2y + 1 \in \mathbb{F}_3[x]$:

x	0	0	0	1	1	1	2	2	2
y	0	1	2	0	1	2	0	1	2
$f(x, y)$	1	1	1	1	2	0	1	2	0

By counting these truth tables, we see that there are exactly $q^{(q^n)}$ functions $\mathbb{F}^n \to \mathbb{F}$.

However, the set $\mathbb{F}[x_1, \dots, x_n]$ is infinite. For example, polynomials in $\mathbb{F}_3[x, y]$ include:

$$2, \qquad x^2 + 1, \qquad 2x^4 + xy, \qquad x + y^3, \qquad x + y, \qquad xy + x^2y^2 + 2, \quad \dots$$

As before, different polynomials can give the same function. For example, over $\mathbb{F}_3$, both x_i^3 and x_i define the same function.

More generally x_i^q and x_i define the same function over $\mathbb{F}_q$.

Multivariate polynomials vs. functions over finite fields

Let $|\mathbb{F}| = q$. Since x_i^q and x_i are the same function, every element in the quotient ring $\mathbb{F}[x_1, \ldots, x_n]/I$, where $I = \langle x_1^q - x_1, \ldots, x_n^q - x_n \rangle$, defines a function.

That is, there is a well-defined (1-to-1) mapping

$$\mathbb{F}[x_1, \ldots, x_n]/I \longrightarrow \{\text{functions } \mathbb{F}^n \to \mathbb{F}\}, \qquad f + I \longmapsto \Big\{c \mapsto f(c)\Big\}.$$

Elements in the quotient ring $\mathbb{F}[x_1, \ldots, x_n]/I$ are sums of monomials with each exponent from $0, \ldots, q-1$:

$$f = \sum c_\alpha x^\alpha, \qquad x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}, \qquad \alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}_q^n, \qquad c_\alpha \in \mathbb{F}$$

For example, in $\mathbb{F}_3[x, y]/I$, each element can be uniquely written as

$$c_{00} + c_{10}x + c_{01}y + c_{20}x^2 + c_{11}xy + c_{02}y^2 + c_{21}x^2y + c_{12}xy^2 + c_{22}x^2y^2.$$

Since there are q^n monomials, there are $q^{(q^n)}$ elements in $\mathbb{F}[x_1, \ldots, x_n]/I$, so the map above is bijective.

Summary

- Elements in the (infinite) ring $\mathbb{F}[x_1, \ldots, x_n]$ are polynomials over $\mathbb{F}$.
- Elements in the (finite) quotient ring $\mathbb{F}[x_1, \ldots, x_n]/\langle x_i^q - x_i, \forall i \rangle$ are functions $\mathbb{F}^n \to \mathbb{F}$.

A familiar example: Boolean functions

There are several standard ways to write a Boolean function $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$.

1. As a logical expression, using $\wedge$, $\vee$, and $\neg$ (or $\overline{}$, !, etc.)
2. As a "square-free" polynomial in $\mathbb{F}[x_1, \ldots, x_n]/\langle x_1^2 - x_1, \ldots, x_n^2 - x_n \rangle$
3. As a truth table.

Boolean operation	logical form	polynomial form
AND	$z = x \wedge y$	$z = xy$
OR	$z = x \vee y$	$z = x + y + xy$
NOT	$z = \overline{x}$	$z = 1 + x$
XOR	$z = x \oplus y = (x \wedge \overline{y}) \vee (\overline{x} \wedge y)$	$z = x + y$

Example

The following are three different ways to express the function that outputs 0 if $x = y = z = 1$, and 1 otherwise.

- $f(x, y, z) = \overline{x \wedge y \wedge z}$
- $f(x, y, z) = 1 + xyz$
-

x	1	1	1	1	0	0	0	0
y	1	1	0	0	1	1	0	0
z	1	0	1	0	1	0	1	0
$f(x, y, z)$	0	1	1	1	1	1	1	1

Boolean networks

Classically, a Boolean network (BN) is an n-tuple $f=(f_1,\dots,f_n)$ of Boolean functions, where $f_i\colon \mathbb{F}_2^n \to \mathbb{F}_2$. This defines a finite dynamical system (FDS) map

$$f\colon \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^n, \qquad x=(x_1,\dots,x_n)\longmapsto (f_1(x),\dots,f_n(x)).$$

Any function from a finite set to itself can be described by a directed graph with every node having out-degree 1. For a BN, this is called the *phase space*, or *state space*.

Definition

The phase space of a BN is the digraph with vertex set $\mathbb{F}_2^n$ and edges $\big\{(x,f(x)) \mid x\in\mathbb{F}_2^n\big\}$.

Proposition

Every function $f\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ is the phase space of a Boolean network $f=(f_1,\dots,f_n)$.

Proof

Clearly, every BN defines a function $\mathbb{F}_2^n \to \mathbb{F}_2^n$. We want to prove the converse. It suffices to show that these sets have the same cardinality.

To count functions $\mathbb{F}_2^n \to \mathbb{F}_2^n$, we count phase spaces. Each of the 2^n nodes has 1 out-going edge, and 2^n destinations. Thus, there are $(2^n)^{2^n}=2^{n2^n}$ phase spaces.

To count BNs: there are $2^{(2^n)}$ choices for each f_i, and so $(2^{(2^n)})^n=2^{n2^n}$ possible BNs. $\square$

Boolean networks: an example

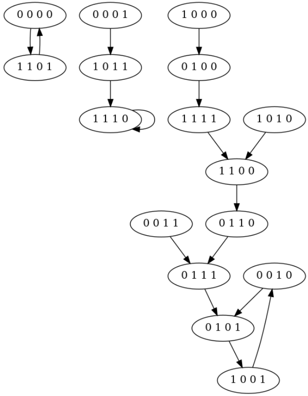
Consider the following Boolean model, where $x + y = x \text{ XOR } y = (x \wedge \overline{y}) \vee (\overline{x} \wedge y)$:

$$(f_1, f_2, f_3, f_4) = \big(x_1 + \overline{x_3},\quad x_3 \vee \overline{x_4},\quad x_2 + x_4,\quad \overline{x_1}\big).$$

The state space has:

- Three basins of attraction (connected components)
- Three attractors (cycles):
  - One 3-cycle
  - One 2-cycle
  - One fixed point (1-cycle)
- Six periodic states
- Ten transient states.

We will leave it as an exercise to formalize these definitions.



Remark

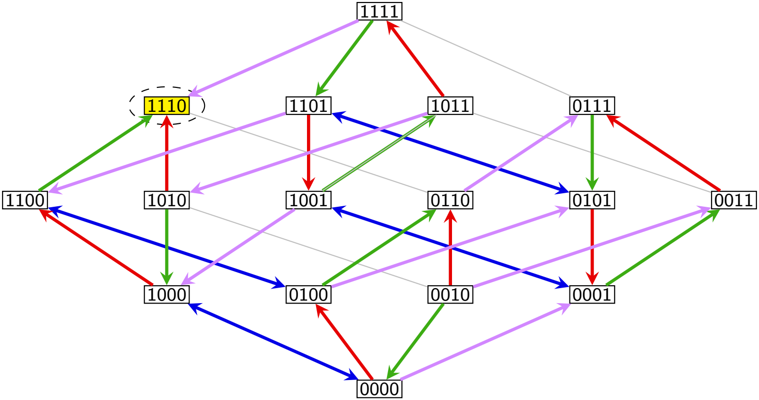
Some sources consider fixed points to be cyclic attractors; others do not.

Boolean networks: an example

Here is the same Boolean model, but under an asynchronous update:

$$(f_1, f_2, f_3, f_4) \quad = \quad \left(x_1 + \overline{x_3}, \quad x_3 \vee \overline{x_4}, \quad x_2 + x_4, \quad \overline{x_1}\right).$$

Notice how the larger limit cycles disappear; there is only one attractor.



Boolean models as polynomials

Every directed graph with $V = \mathbb{F}_2^n$ with uniform out-degree 1 is the phase space of some Boolean model $(f_1, \dots, f_n)$.

Each function $f_i \colon \mathbb{F}_2^n \to \mathbb{F}_2$ lies in the quotient ring $\mathbb{F}_2[x_1, \dots, x_n]/\langle x_1^2 - x_1, \dots, x_n^2 - x_n \rangle$.

Summary

There are natural bijections between the following sets of size 2^{n2^n}:

(i) Boolean models $(f_1, \dots, f_n)$, where $f_i \colon \mathbb{F}_2^n \to \mathbb{F}$.

(ii) Phase space graphs, i.e., functions $f \colon \mathbb{F}_2^n \to \mathbb{F}_2^n$.

(iii) Elements in the direct product $(R/I) \times \cdots \times (R/I)$ of quotient rings, where

$$R = \mathbb{F}_2[x_1, \dots, x_n] \qquad \text{and} \qquad I = \langle x_1^2 - x_1, \dots, x_n^2 - x_n \rangle.$$

Natural question

Given a function $f \colon \mathbb{F}_2^n \to \mathbb{F}_2^n$, how can we find the individual local functions $f_1, \dots, f_n$ for which

$$f \colon (x_1, \dots, x_n) \longmapsto (f_1(x), \dots, f_n(x))?$$

Algebraic models and FDSs

We just saw how every function $f = \mathbb{F}_2^n \to \mathbb{F}_2^n$ can be written as an n-tuple of "square-free" polynomials over $\mathbb{F}_2$:

$$f = (f_1, \dots, f_n), \qquad f_i \in \mathbb{F}_2[x_1, \dots, x_n]/\langle x_1^2 - x_1, \dots, x_n^2 - x_n \rangle.$$

This carries over to generic finite fields, but we will carefully re-define things first.

Definition

Let $\mathbb{F}$ be a finite field. An algebraic model over $\mathbb{F}$ is an n-tuple of functions $f = (f_1, \dots, f_n)$, where each $f_i \colon \mathbb{F}^n \to \mathbb{F}$.

Definition

Every algebraic model $f = (f_1, \dots, f_n)$ over $\mathbb{F}$ defines a finite dynamical system (FDS), by iterating the map

$$f \colon \mathbb{F}^n \longrightarrow \mathbb{F}^n, \qquad x = (x_1, \dots, x_n) \longmapsto (f_1(x), \dots, f_n(x)).$$

Remark

A classical Boolean network (BN) is just an algebraic model over $\mathbb{F}_2$.

Algebraic models and FDSs

Let $\mathbb{F}$ be a finite field of order $q = p^k$. Recall that

$$R/I = \mathbb{F}[x_1, \dots, x_n]/\langle x_1^q - x_1, \dots, x_n^q - x_n \rangle$$

is the set of functions $\mathbb{F}^n \to \mathbb{F}$.

Remark

Every algebraic model $f = (f_1, \dots, f_n)$ can be associated with an element in $(R/I) \times \cdots \times (R/I)$.

Recall that there are $q^{(q^n)}$ elements in R/I.

Summary

(i) There are $q^{(nq^n)}$ algebraic models $(f_1, \dots, f_n)$ over $\mathbb{F}$.

(ii) There are $q^{(nq^n)}$ functions $\mathbb{F}^n \to \mathbb{F}^n$ (i.e., FDS maps, or phase spaces).

In other words, there is a natural bijection between these sets.

Said differently every function $\mathbb{F}^n \to \mathbb{F}^n$ is indeed the finite dynamical system (FDS) map (i.e., phase space) of an algebraic model $(f_1, \dots, f_n)$ over $\mathbb{F}$.

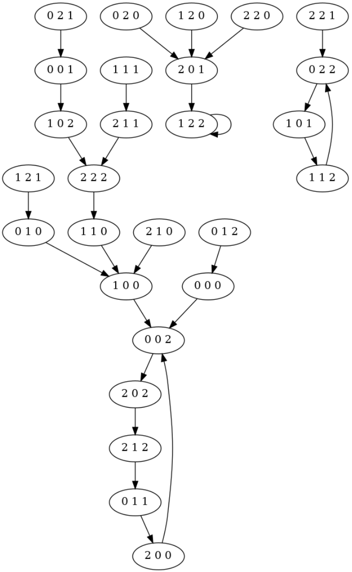
Algebraic models: a ternary example

Consider the following ternary model:

$$(f_1, f_2, f_3) = \big(x_2 + x_3,\ x_1x_3,\ 2 + x_2(x_1x_3 + 1)\big).$$

The state space has:

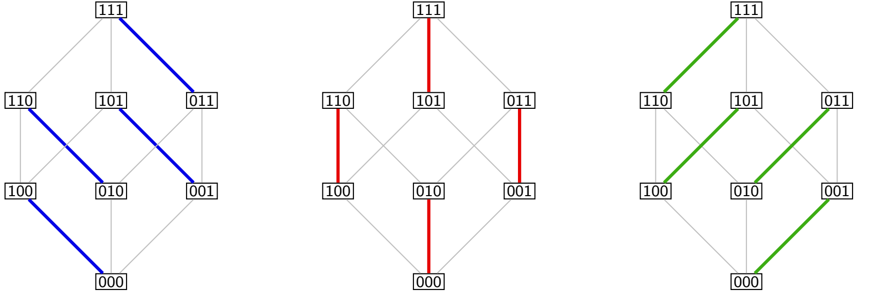
- Three basins of attraction (connected components)
- Three attractors (cycles):
  - One 5-cycle
  - One 3-cycle
  - One fixed point (1-cycle)
- 9 periodic states
- 18 transient states.



Composing functions asynchronously

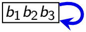
Consider a Boolean model (f_1, f_2, f_3), where $f_i \in \mathbb{F}_2^3 \to \mathbb{F}_2$.

Suppose only one f_i is applied at a time. Then the only state transitions are between nodes that differ in one bit.



f_1 *can only change the 1st bit* f_2 *can only change the 2nd bit* f_3 *can only change the 3rd bit*

Moreover, upon applying f_i from any node $x \in \mathbb{F}_2^3$, there are only two possibilities:

- f_i fixes the i^{th} bit: 

- f_i flips the i^{th} bit. 

Asynchronous Boolean networks

Consider a Boolean network $f = (f_1, \ldots, f_n)$.

Composing the functions synchronously defines the FDS map $f\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$.

We can also compose them asynchronously. For each function f_i, define the function

$$F_i\colon \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^n, \qquad x = (x_1, \ldots, x_i, \ldots, x_n) \longmapsto (x_1, \ldots, f_i(x), \ldots, x_n).$$

Rather than a canonical dynamical system map, this defines a finite state automaton.

Definition

The asynchronous automaton of $(f_1, \ldots, f_n)$ is the digraph with vertex set $\mathbb{F}_2^n$ and edges

$$\big\{(x, F_i(x)) \mid i = 1, \ldots, n;\ x \in \mathbb{F}_2^n\big\}.$$

Remarks

- Clearly, this graph has $n \cdot 2^n$ edges, though self-loops are often omitted.
- Every non-loop edge connect two vertices that differ in exactly one bit. That is, all non-loops are of the form $(x, x + e_i)$, where e_i is the i^{th} standard unit basis vector.
- It is elementary to extend this concept from BNs to algebraic models over finite fields.

The asynchronous automaton of an algebraic model

Recall: every function $\mathbb{F}^n \to \mathbb{F}^n$ can be realized as the FDS map (i.e., phase space) of an algebraic model over $\mathbb{F}$.

Similarly, every digraph with vertex set $\mathbb{F}^n$ that "could be" the asynchronous automaton of an algebraic model, is one.

Theorem

Let $G = (\mathbb{F}^n, E)$ be a digraph with the following local property (definition):

For every $x \in \mathbb{F}^n$ and $i = 1, \dots, n$: E contains exactly one edge of the form $(x, x + ke_i)$, where $k \in \mathbb{F}$ (possibly a self-loop)

Then G is the asynchronous automaton of some algebraic model $(f_1, \dots, f_n)$ over $\mathbb{F}$.

Proof

It suffices to show there there are $q^{(nq^n)}$ digraphs $G = (\mathbb{F}^n, E)$ with the "local property".

Each of the q^n nodes $x \in \mathbb{F}^n$ has n out-going edges (including loops). Each edge has q possible destinations: $x + ke_i$ for $k \in \mathbb{F}$.

This gives q^n choices at each node, for all q^n nodes, for $(q^n)^{q^n} = q^{(nq^n)}$ graphs in total. □

Algebraic models over general finite fields: synchronous vs. asynchronous

Let $\mathbb{F}$ be a finite field of order $q = p^k$. The following quotient ring has cardinality $q^{(q^n)}$:

$$R/I = \mathbb{F}[x_1, \dots, x_n]/\langle x_1^q - x_1, \dots, x_n^q - x_n \rangle,$$

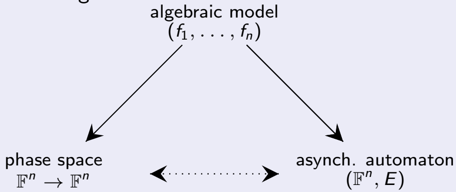
Summary (updated)

Each of the following sets have cardinality $q^{(nq^n)}$:

- algebraic models $(f_1, \dots, f_n)$ over $\mathbb{F}$.
- elements of $(R/I) \times \cdots \times (R/I)$. [$n$ copies]
- synchronous phase spaces, i.e., FDS maps $\mathbb{F}^n \to \mathbb{F}^n$.
- asynchronous automata: a digraph $G = (\mathbb{F}^n, E)$ with the "local property".

Open-ended question

Better understand the following:



Phase space vs. asynchronous automaton

The phase space of an algebraic model $f = (f_1, \ldots, f_n)$ has two types of nodes:

- *transient points*: $f^k(x) \neq x$ for all $k \geq 1$.
- *periodic points*: $f^k(x) = x$ for some $k \geq 1$. ($k = 1$: *fixed point*)

Thus, the phase space consists of periodic cycles and directed paths leading into these cycles.

The asynchronous automaton of $f = (f_1, \ldots, f_n)$ can be more complicated.

For $x, y \in \mathbb{F}^n$, define $x \sim y$ iff there is a directed path from x to y and from y to x.

The resulting equivalence classes are the strongly connected components (SCC) of the phase space. An SCC is terminal if it has no out-going edges from it.

A point $x \in \mathbb{F}^n$:

- *is transient* if it is not in a terminal SCC.
- *lies on a cyclic attractor* if its terminal SCC is a chordless k-cycle ($k = 1$: *fixed point*).
- *lies on a complex attractor* otherwise.

Proposition

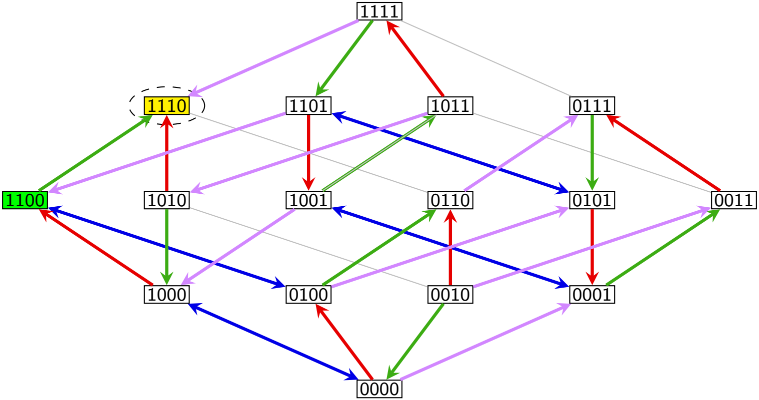
The fixed points of an algebraic model are the same under synchronous and asynchronous update.

Strongly connected components

Let's revisit a previous example, the asynchronous automaton of

$$(f_1, f_2, f_3, f_4) \quad = \quad \big(x_1 + \overline{x_3}, \quad x_3 \vee \overline{x_4}, \quad x_2 + x_4, \quad \overline{x_1}\big).$$

There are 3 strongly connected components, colored below.



Wiring diagrams

A function $f_j\colon \mathbb{F}^n \to \mathbb{F}$ depends on x_i if for some $x \in \mathbb{F}^n$ and $k \in \mathbb{F}$,

$$f_j(x) \neq f_j(x + ke_i),$$

where $e_i \in \mathbb{F}^n$ is the i^{th} standard unit basis vector.

Definition

The wiring diagram of an algebraic model $(f_1, \dots, f_n)$ over $\mathbb{F}$ is a directed graph G with vertex set $x_1, \dots, x_n$ (or just $1, \dots, n$) and a directed edge (x_i, x_j) if f_j depends on x_i.

If $\mathbb{F} = \mathbb{F}_p$, then an edge $x_i \longrightarrow x_j$ is positive if $a \leq b$ implies

$$f_j(x_1, \dots, x_{i-1}, a, x_{i+1}, \dots, x_n) \leq f_j(x_1, \dots, x_{i-1}, b, x_{i+1}, \dots, x_n)$$

and negative if the second inequality is reversed.

Negative edges are denoted with circles or blunt arrows instead of traditional arrowheads.

Definition

A function $f_j\colon \mathbb{F}^n \to \mathbb{F}$ is unate (or monotone) if every edge in the wiring diagram is either positive or negative.

Wiring diagrams in Boolean networks

- A positive edge $x_i \longrightarrow x_j$ represents a situation where i activates j.

  Examples.

  - $f_j = x_i \wedge y$: $\quad 0 = f_j(x_i = 0, y) \leq f_j(x_i = 1, y) \leq 1.$
  - $f_j = x_i \vee y$: $\quad 0 \leq f_j(x_i = 0, y) \leq f_j(x_i = 1, y) = 1.$

- A negative edge $x_i \longrightarrow\!\!| \; x_j$ represents a situation where i inhibits j.

  Examples.

  - $f_j = \overline{x_i} \wedge y$: $\quad 1 \geq f_j(x_i = 0, y) \geq f_j(x_i = 1, y) = 0.$
  - $f_j = \overline{x_i} \vee y$: $\quad 1 = f_j(x_i = 0, y) \geq f_j(x_i = 1, y) \geq 0.$

- We can write $x_i \longrightarrow\!\!\!\!\!| \; x_j$ for edges that are neither positive nor negative:

  Example. (The logical "XOR" function):

  - $f_j = x_i + y = (x_i \wedge \overline{y}) \vee (\overline{x_i} \wedge y)$:

    $0 = f_j(x_i = 0, y = 0) < f_j(x_i = 1, y = 0) = 1$

    $1 = f_j(x_i = 0, y = 1) > f_j(x_i = 1, y = 1) = 0$

Most edges in Boolean network models are either positive or negative because most biological interactions are either simple activations or inhibitions.

Enumerating Boolean networks

Motivating question

Recall our 9-node Boolean network model of the *lac* operon. For all 4 initial conditions $(G_e, L_e) \in \mathbb{F}_2^2$, the phase space had exactly 1 fixed point that made biological sense.

What are the chances that this would have happened purely by coincidence?

To answer this, we need to count the number of Boolean networks, as well as those that have just that one fixed point.

Recall

Every graph $G = (\mathbb{F}^n, E)$ with uniform out-degree 1 is the phase space of some algebraic model $(f_1, \dots, f_n)$ over $\mathbb{F}$.

Corollary

Start with a phase space with vertex set $\mathbb{F}_2^n$. Remove k edges. There are exactly 2^{nk} algebraic models that "fit this data".

Proof

The tail of each "missing edge" is a state $x \in \mathbb{F}_2^n$, and there are 2^n possibile destinations $x \to y$ when replacing it. □

An example

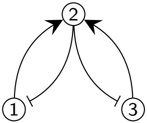
Exercise (easy)

How many Boolean networks contain the 4-cycle $000 \to 101 \to 111 \to 010 \to 000$ in their phase space?

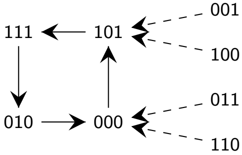
Here is one example:

$$\begin{cases} f_1 = \overline{x_2} \\ f_2 = x_1 \wedge x_3 \\ f_3 = \overline{x_2} \end{cases}$$

Functions



Wiring diagram



Phase space

Suppose we remove all of the "dashed edges." Then we can replace each one 8 different ways. Thus, there are $8^4 = 4096$ possibilities.

Exercise (harder)

How many Boolean networks contain a 4-cycle in their phase space? What if we require that there is additionally *only one connected component*?

Counting algebraic models

Theorem

There are $q^{(nq^n)}$ algebraic models on n nodes. Of these:

(a) $q^n!$ have a phase space consisting of a length-q^n chain of transient points.

(b) $q^n!$ are invertible (i.e., have no transient points).

(c) $(q^n - 1)!$ are invertible with a phase space consisting of a single cycle.

(d) $(q^n - 1)^{q^n}$ have no fixed points.

(e) $(q^n)^{q^n - 1}$ have a single connected component and fixed point.

(f) $(q^n + 1)^{q^n - 1}$ have only fixed points (i.e., no longer periodic cycles).

As an example, the number of Boolean networks (that is, $q = 2$) on n nodes with various properties is shown below.

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
total BNs	256	1.678×10^{7}	1.845×10^{19}	1.462×10^{48}	3.940×10^{115}
invertible	24	40320	2.092×10^{13}	2.631×10^{35}	1.269×10^{89}
single big cycle	6	5040	1.308×10^{12}	8.223×10^{33}	1.983×10^{87}
no fixed points	81	5.765×10^{6}	6.568×10^{18}	5.291×10^{47}	1.438×10^{115}
1 component & f.p.	64	2.097×10^{6}	1.153×10^{18}	4.567×10^{46}	6.157×10^{113}
only fixed points	125	4.782×10^{6}	2.862×10^{18}	1.189×10^{47}	1.635×10^{114}

Counting algebraic models

Theorem

There are $q^{(nq^n)}$ algebraic models on n nodes. Of these:

(a) $q^n!$ have a phase space consisting of a length-q^n chain of transient points.

(b) $q^n!$ are invertible (i.e., have no transient points).

(c) $(q^n - 1)!$ are invertible with a phase space consisting of a single cycle.

(d) $(q^n - 1)^{q^n}$ have no fixed points.

(e) $(q^n)^{q^n - 1}$ have a single connected component and fixed point.

(f) $(q^n + 1)^{q^n - 1}$ have only fixed points (i.e., no longer periodic cycles).

Proof (sketch)

(a)–(d) are elementary counting arguments.

(e) is just the number labeled rooted trees on q^n nodes.

For (f), use a bijection between phase spaces and labeled unrooted trees on $q^n + 1$ nodes. □

Cayley's formula (and corollaries)

- $\#\{$labeled unrooted trees on n nodes$\} = n^{n-2}$.
- $\#\{$labeled rooted trees on n nodes$\} = n^{n-1}$.
- The number of labeled forests on n labeled vertices is $(n+1)^{n-1}$.

Back to the *lac* operon

Motivating question

Recall our 9-node Boolean network model of the *lac* operon. For all 4 initial conditions $(G_e, L_e) \in \mathbb{F}_2^2$, the phase space had exactly 1 fixed point that made biological sense.

What are the chances that this would have happened purely by coincidence?

There are $(2^9)^{(2^9)} = 512^{512} \approx 1.400 \times 10^{1387}$ Boolean networks on 9 nodes.

Of these, $(2^9)^{2^9-1} = 512^{511} \approx 2.735 \times 10^{1384}$ have a single component and fixed point.

Of these, $(2^9)^{2^9-2} = 512^{510} \approx 5.342 \times 10^{1381}$ have the "correct" fixed point.

In other words, 1 in 262,141 Boolean networks on n nodes have this property.

Thus, the probability that each $(G_e, L_e) \in \mathbb{F}_2^2$ would yield such a phase space purely by chance is approximately

$$\left(\frac{1}{262{,}141}\right)^4 \approx 2.118 \times 10^{-22}.$$