# The network inference problem

Matthew Macauley

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

Algebraic Systems Biology

# Inferring gene regulatory networks

The inference of gene regulatory networks (GRNs) is an important and challenging problem in systems biology.

Expression data can be obtained from tools such as DNA microarrays (old) or RNA-Seq.

Once the data is obtained, the problem becomes:

*What is the wiring diagram*?

There are a number of approaches to answer this, such as:

- Correlation networks
- Regression-based methods
- Information theoretical scores (mutual information)
- Gaussian graphical models
- Bayesian networks (classic and dynamic)
- Differential equation methods
- Algebraic methods

This last technique will be the focus of this section.

# Inferring gene regulatory networks

The following papers survey existing techniques and software for inferring biological networks.

- Huynh-Thu, V. A., & Sanguinetti, G. (2019). Gene regulatory network inference: an introductory survey. *Gene regulatory networks: Methods and protocols*, 1-23.
- Saint-Antoine, M. M., & Singh, A. (2020). Network inference in systems biology: recent developments, challenges, and applications. *Curr. Opin. Biotechnol.* **63**, 89-98.
- Pušnik, Ž., Mraz, M., Zimic, N., & Moškon, M. (2022). Review and assessment of Boolean approaches for inference of gene regulatory networks. *Heliyon*, **8**(8), e10222.

The approach in this lecture will be to use computational algebra.

The data is encoded algebraically by squarefree monomial ideals, which can be described combinatorially by simplicial complexes.

The mathematics behind the scenes is called Stanley–Reisner theory.

- Francisco, C. A., Mermin, J., & Schweig, J. (2014). A survey of Stanley–Reisner theory. In *Connections Between Algebra, Combinatorics, and Geometry*, pp. 209-234. Springer New York.

# Inferring gene regulatory networks

The questions of which variables depend on which others is inherently algebraic.

The unsigned wiring diagram can be inferred using Stanley–Reisner theory.

- Jarrah, A. S., Laubenbacher, R., Stigler, B., & Stillman, M. (2007). Reverse-engineering of polynomial dynamical systems. *Adv. Appl. Math.*, **39**(4), 477-489.

This was later extended to signed wiring diagrams.

- Veliz-Cuba, A. (2012). An algebraic approach to reverse engineering finite dynamical systems arising from biology. *SIAM J. Appl. Dyn. Syst.*, **11**(1), 31-48.
- Veliz-Cuba, A., Newsome-Slade, V., & Dimitrova, E. S. (2024). A unified approach to reverse engineering and data selection for unique network identification. *SIAM J. Appl. Dyn. Syst.* **23**(1), 592-615.

The following generalizes the previous approaches to when the data is no longer discretized.

- Harrington, H. A., Stillman, M., & Veliz-Cuba, A. (2024). Algebraic network reconstruction of discrete dynamical systems. *Adv. Appl. Math.*, **161**, 102760.

# Inferring wiring diagrams of Boolean models

Suppose a Boolean model $(f_1, f_2, f_3)$ has the following (partial) state space.

$$001 \longrightarrow 101 \longrightarrow 111 \longrightarrow 110 \longrightarrow 010 \longrightarrow 000 \qquad 100 \qquad 011$$

## Main question

What are the possible wiring digrams?

For example, given the following data, *which variables must depend on which variables?*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $x$ | 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 |
| $f_1(x,y,z)$ | ? | 0 | ? | 0 | 1 | ? | 1 | 1 |
| $f_2(x,y,z)$ | ? | 0 | ? | 1 | 0 | ? | 1 | 1 |
| $f_3(x,y,z)$ | ? | 0 | ? | 0 | 1 | ? | 1 | 0 |

Note that we can treat each function $f_1$, $f_2$, $f_3$ separately.

## First question

What are the possible variable dependencies of $f : \mathbb{F}_2^n \to \mathbb{F}_2$, given partial information?

# Inferring wiring diagrams of a single Boolean function

Suppose we have an unknown Boolean function $f_i \colon \mathbb{F}_2^3 \to \mathbb{F}_2$ that satisfies:

$$f_i(1,1,1) = 0, \qquad f_i(0,0,0) = 0, \qquad f_i(1,1,0) = 1.$$

In other words, its truth table looks like

| $x_1x_2x_3$ | 111 | 110 | 101 | 100 | 011 | 010 | 001 | 000 |
|---|---|---|---|---|---|---|---|---|
| $f_i(x)$ | 0 | 1 | ? | ? | ? | ? | ? | 0 |

## Goals

1. Reverse engineering the wiring diagram: *Which sets of variables can $f_i$ depend on?*
2. Reverse engineering the model space: *Characterize all functions that "fit this data".*
3. Model selection: *What is the "best fit" function?*

We'll study the first question in this lecture.

Recall how different types of interactions are indicated in the wiring diagram:

| $f_j = x_i \wedge x_k$ | $f_j = \overline{x_i} \wedge x_k$ | $f_j = x_i + x_k$ |
|---|---|---|
| $x_i \longrightarrow x_j$ | $x_i \longrightarrow\!\!\!\mid\; x_j$ | $x_i \longrightarrow\!\!\!\mid\; x_j$ (arrow and bar) |
| "$x_i$ activates $x_j$" | "$x_i$ inhibits $x_j$" | "$x_i$ affects $x_j$ positively & negatively" |

# Unate functions

Consider the following unknown Boolean function:

| $x_1x_2x_3$ | 111 | 110 | 101 | 100 | 011 | 010 | 001 | 000 |
|---|---|---|---|---|---|---|---|---|
| $f_i(x)$ | 0 | 1 | ? | ? | ? | ? | ? | 0 |

There are $2^8 = 256$ truth tables, and of these, $2^{8-3} = 32$ fit this data.

Not all of these functions are *biologically meaningful*.

## Definition

A Boolean function $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ is unate if no variable $x_i$ and its negation $\overline{x_i}$ both appear.

## Examples

- Conjunctions: $f = x_{i_1} \wedge \cdots \wedge x_{i_k}$.
- Disjunctions: $f = x_{i_1} \vee \cdots \vee x_{i_k}$.
- AND-NOT functions: $f = x \wedge \overline{y} \wedge z$.
- OR-NOT functions: $f = x \vee \overline{y} \vee \overline{z}$.
- Others: $f = x \wedge (\overline{y} \vee z)$.

## Fact

Most functions that appear in models of molecular networks are unate.

# Min-sets

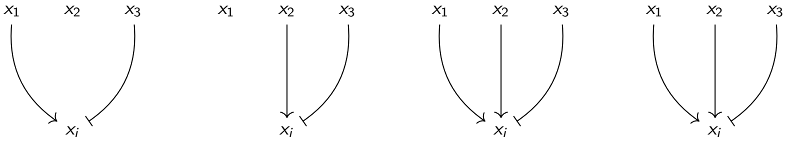
Recall the following unknown Boolean function:

| $x_1x_2x_3$ | 111 | 110 | 101 | 100 | 011 | 010 | 001 | 000 |
|---|---|---|---|---|---|---|---|---|
| $f_i(x)$ | 0 | 1 | ? | ? | ? | ? | ? | 0 |

Of the 256 Boolean functions on 3 variables, $2^{8-3} = 32$ fit this data, and only 4 are unate:

$$x_1 \wedge \overline{x_3}, \qquad x_2 \wedge \overline{x_3}, \qquad x_1 \wedge x_2 \wedge \overline{x_3}, \qquad (x_1 \vee x_2) \wedge \overline{x_3}.$$

The wiring diagrams of these functions are shown below, expressed several different ways.




$(1, 0, -1)$ $\quad$ $(0, 1, -1)$ $\quad$ $(1, 1, -1)$ $\quad$ $(1, 1, -1)$

$\{x_1, \overline{x_3}\}$ $\quad$ $\{x_2, \overline{x_3}\}$ $\quad$ $\{x_1, x_2, \overline{x_3}\}$ $\quad$ $\{x_1, x_2, \overline{x_3}\}$

We will call the minimal wiring diagrams (e.g., the first two) min-sets. If we retain the signs of the interactions, we call them signed min-sets.

# Finding min-sets using computational algebra

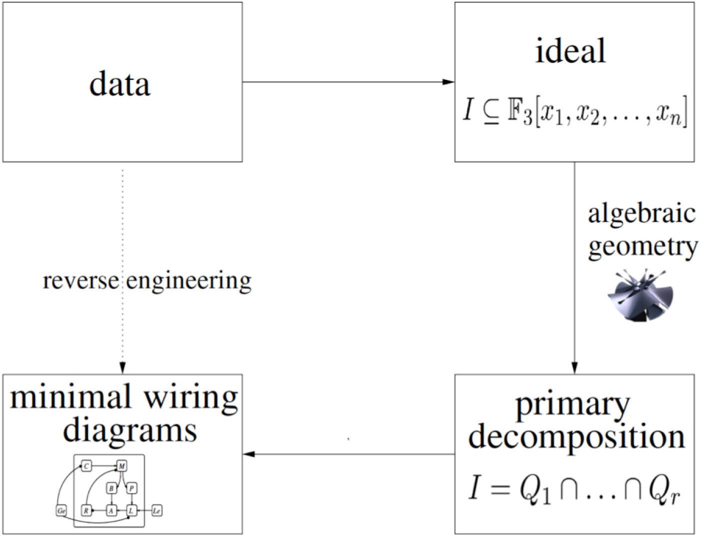



Figure: Image courtesy of Alan Veliz-Cuba.

# Monomials

We will learn how to reverse-engineer wirgram diagrams using computational algebra.

We will encode the partial data using ideals of polynomials rings generated by square-free monomials.

There is a beautiful relationship between square-free monomial ideals and a combinatoral object called a simplicial complex.

The min-sets can be found by taking the primary decomposition of the ideal.

## Notation

Every monomial can be written as $cx^\alpha$, where $x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ and $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geq 0}$.

## Example

Consider the following polynomial in $\mathbb{F}_3[x_1, x_2, x_3, x_4]$, written several different ways:

$$f = x_1^3 x_2 x_4^2 + 2x_1 x_4^5 = x_1^3 x_2^1 x_3^0 x_4^2 + 2x_1^1 x_2^0 x_3^0 x_4^5 = x^{(3,1,0,2)} + 2x^{(1,0,0,5)}.$$

# Monomial ideals

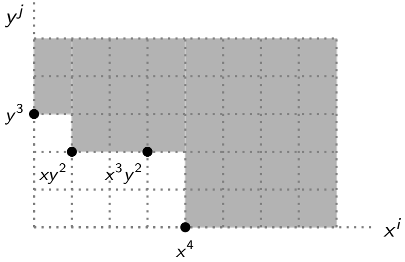
## Definition

A monomial ideal $I \leq \mathbb{F}[x_1, \dots, x_n]$ is an ideal generated by monomials.

## Proposition (exercise)

Let $\mathcal{M}(I)$ be the set of monomials in $I$. If $I$ is a monomial ideal, then $I = \langle \mathcal{M}(I) \rangle$.

Monomial ideals can be visualized by a staircase diagram. Here is an example for the monomial ideal $I = \langle y^3, xy^2, x^3y^2, x^4 \rangle$.




*Question*: Are any of these monomials not needed to generate $I$?

# Square-free monomial ideals

## Definition

A monomial $x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ is square-free if each $\alpha_i \in \{0,1\}$.

A square-free monomial ideal is any ideal generated by square-free monomials.

The exponent vector $\alpha = (\alpha_1, \ldots, \alpha_n)$ of a square-free monomial $x^\alpha$ canonically determines a subset of $[n] = \{1, \ldots, n\}$.

## Notations

- Given $x^\alpha$, we may speak of $\alpha$ as a *subset* of $[n]$ rather than a vector.
- We will write subsets as strings, e.g., $xz$ for $\{x, z\}$.

## Key property

Let $I$ be a square-free monomial ideal of $\mathbb{F}[x_1 \ldots, x_n]$, and $\alpha, \beta \subseteq [n]$. Then

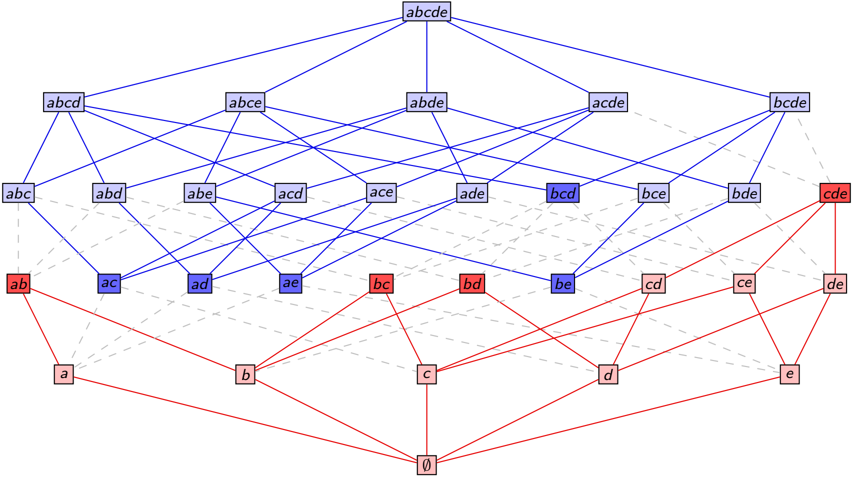
$$x^\alpha \in I \quad \text{and} \quad x^\beta \in I \quad \Longrightarrow \quad x^{\alpha\cup\beta} \in I,$$

$$x^\alpha \notin I \quad \text{and} \quad x^\beta \notin I \quad \Longrightarrow \quad x^{\alpha\cap\beta} \notin I.$$

# Square-free monomial ideals

Consider the ideal $I = \langle ac, ad, ae, bcd, be \rangle \subseteq \mathbb{F}[a, b, c, d, e]$.

The squarefree monomials in $I$ are blue. Those *not* in $I$ are red.



# Simplicial complexes

## Definition

A simplicial complex over a finite set $X$ is a collection $\Delta$ of subsets of $X$, closed under taking subsets. That is,

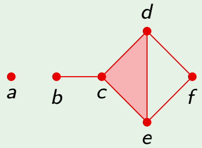
$$\beta \in \Delta \quad \text{and} \quad \alpha \subset \beta \quad \Longrightarrow \quad \alpha \in \Delta.$$

Elements in $\Delta$ are called simplices or faces.

## Example 6

$X = \{a, b, c, d, e, f\}$

$\Delta = \{\emptyset, a, b, c, d, e, f, bc, cd, ce, de, cde, df, ef\}$




A $k$-dimensional face (size-$(k+1)$ subset) is called a $k$-face. For small $k$, we also say that a:

- 0-face is a vertex, or node,
- 1-face is an edge,
- 2-face is a triangle,
- 3-face is a (solid) triangular pyramid.

# Simplicial complexes

We will often be interested in the non-faces of a simplicial complex, i.e., $\Delta^c := 2^X \setminus \Delta$.

## Key property

Let $\Delta$ be a simplicial complex.

(i) Faces of $\Delta$ are closed under intersection: $\alpha, \beta \in \Delta \quad \Rightarrow \quad \alpha \cap \beta \in \Delta$.

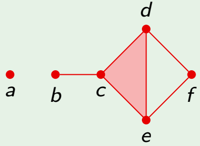
(ii) Non-faces of $\Delta$ are closed under unions: $\alpha, \beta \in \Delta^c \quad \Rightarrow \quad \alpha \cup \beta \in \Delta^c$.

## Remark

- $\Delta$ is determined by its maximal faces.
- $\Delta^c$ is determined by its minimal non-faces.

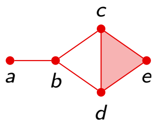
## Example 6 (continued)

- 14 faces in $\Delta = \{\emptyset, a, b, c, d, e, f, bc, cd, ce, de, cde, df, ef\}$.
- Maximal faces: $a$, $bc$, $cde$, $df$, $ef$.
- 50 non-faces in $\Delta^c$.
- Minimal non-faces: $ab$, $ac$, $ad$, $ae$, $af$, $bd$, $be$, $bf$, $cf$, $def$.

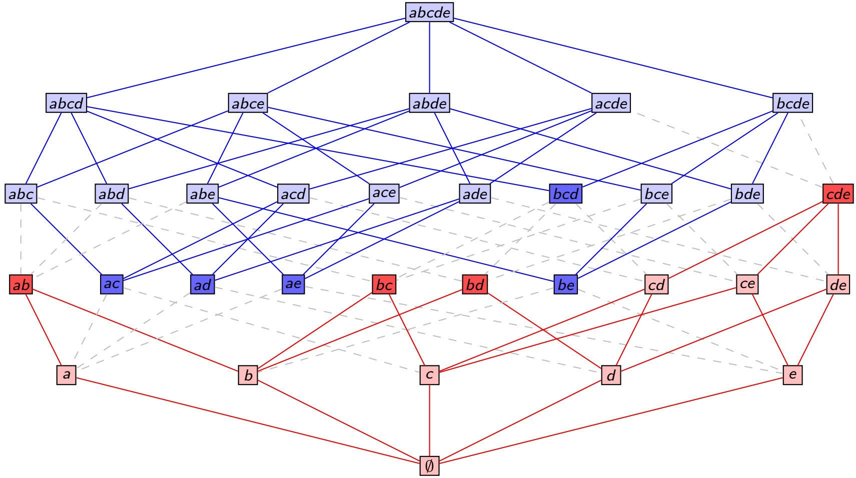


# Simplicial complexes

Consider the simplicial complex $\Delta$:




The faces in $\Delta$ are red. The non-faces are blue.

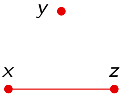


# Example 3a

Let $X = \{x, y, z\}$, and consider the following simplicial complex $\Delta$ and ideal in $\mathbb{F}[x, y, z]$.

Faces: $\Delta = \{\emptyset, x, y, z, xz\}$ (maximal: $y$, $xz$)

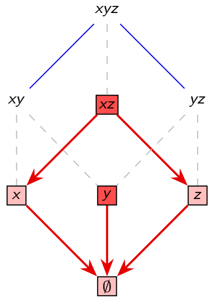
Non-faces: $\Delta^c = \{xy, yz, xyz\}$ (minimal: $xy$, $yz$)




$$I_{\Delta^c} = \langle xy, yz\rangle$$

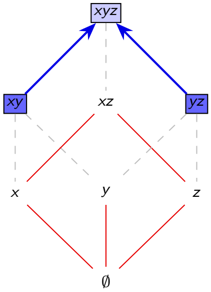
$$= \Big\{ \underbrace{xy \cdot h_1(x,y,z) + yz \cdot h_2(x,y,z)}_{y\left(x\cdot h_1(x,y,z)+z\cdot h_2(x,y,z)\right)\in\langle y\rangle\cap\langle x,z\rangle} : h_1, h_2 \in R\Big\} = \langle y\rangle \cap \langle x, z\rangle$$

"*primary decomposition*"



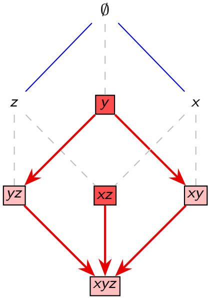

Faces $\Delta$

Monomials *not* in $I_{\Delta^c}$




Non-faces $\Delta^c$

Monomials in $I_{\Delta^c}$




Complements of faces in $\Delta$

Primary components are darker

# Alexander duality

We've seen the following bijection, called **Alexander duality**:

- Every square-free monomial ideal $I$ defines a canonical simplicial complex, $\Delta_{I^c}$.
- Every simplicial complex $\Delta$ defines a canonical square-free monomial ideal $I_{\Delta^c}$.

**Example 6** (contin.): Consider the square-free monomial ideal $I$ in $\mathbb{F}[a, b, c, d, e, f]$:

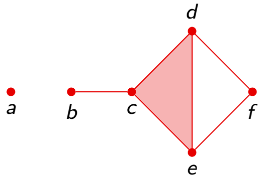
$$I = \langle ab, ac, ad, ae, af, bd, be, bf, cf, def \rangle.$$

The monomials *not* in $I$ are closed under intersection, and so they form a simplicial complex

$X = \{a, b, c, d, e, f\}$

$\Delta_{I^c} = \{\emptyset, a, b, c, d, e, f, bc, cd, ce, de, cde, df, ef\}$

Minimal non-faces: $ab$, $ac$, $ad$, $ae$, $af$, $bd$, $be$, $bf$, $cf$, $def$




Note that $\Delta_{I^c}$ is determined by its maximal faces: $a$, $bc$, $cde$, $df$, $ef$.

Complements of max'l faces: $bcdef$, $adef$, $abf$, $abce$, $abcd$, so the primary decomposion is

$$I = \langle b, c, d, e, f\rangle \cap \langle a, d, e, f\rangle \cap \langle a, b, f\rangle \cap \langle a, b, c, e\rangle \cap \langle a, b, c, d\rangle;$$

# Summary so far

## Key property

A square-free monomial ideal $I$ is completely determined by the subsets $\alpha$ for which $x^\alpha \in I$.

- If $\alpha \subseteq \beta$ and $x^\alpha \in I$, then $x^\beta \in I$.
- If $\alpha \subseteq \beta$ and $x^\beta \notin I$, then $x^\alpha \notin I$.

In other words,

(i) As subsets, exponents of square-free monomials *in I* are closed under unions.

(ii) As subsets, exponents of square-free monomials *not in I* are closed under intersections.

## Key ideal

We can describe a square-free monomial ideal $I$ combinatorially as a collection of subsets, closed under intersections, and vice-versa.

These subsets have two interpretations, one algebraic and one combinatorial.

- *algebraically*: the monomials $x^\alpha$ not in $I$;
- *combinatorially*: the faces $\alpha$ of a simplicial complex, that we will denote by $\Delta_{I^c}$.

# Alexander duality, formalized

## Definition

Given a squarefree monomial ideal $I$ in $\mathbb{F}[x_1, \ldots, x_n]$, define the simplicial complex

$$\Delta_{I^c} := \{\alpha \mid x^\alpha \notin I\}.$$

Given a simplicial complex $\Delta$, define a square-free monomial ideal

$$I_{\Delta^c} := \langle x^\alpha \mid \alpha \notin \Delta \rangle.$$

This is called the Stanley-Reisner ideal of $\Delta$.

## Theorem

The correspondence $I \mapsto \Delta_{I^c}$ and $\Delta \mapsto I_{\Delta^c}$ is a bijection between:

(i) simplicial complexes on $[n] = \{1, \ldots, n\}$,

(ii) square-free monomial ideals in $\mathbb{F}[x_1, \ldots, x_n]$.

This correspondence is called **Alexander duality**.

# Primary decomposition (motivation)

In grade-school, everybody learns how to factor integers into products of primes, e.g.,

$$6 = 2 \cdot 3, \qquad \text{and} \qquad 45 = 3^2 \cdot 5.$$

Ideals in the ring $R = \mathbb{Z}$ behave similarly.

Since $\mathbb{Z}$ is a principal ideal domain (PID), every ideal has the form $I = \langle a \rangle$ for some $a \in \mathbb{Z}$.

Every ideal $I$ can be written as an intersection of primary ideals. For example,

$$\langle 6 \rangle = \langle 2 \rangle \cap \langle 3 \rangle, \qquad \text{and} \qquad \langle 45 \rangle = \langle 9 \rangle \cap \langle 5 \rangle.$$

This is called a primary decomposition of the ideal.

Note that there is no way to further break up $\langle 9 \rangle$ into an expression involving $\langle 3 \rangle$.

Ideals of the form $I = \langle p \rangle$ for a prime $p$ are called prime ideals and those of the form $I = \langle p^k \rangle$ are called primary ideals.

These concepts and this construction holds in a much larger class of commutative rings than just $\mathbb{Z}$.

# Primary decomposition

## Defintion

Let $I$ be an ideal of a commutative ring $R$.

- $I$ is a prime ideal if $fg \in I$ implies either $f \in I$ or $g \in I$.
- $I$ is a primary ideal if $fg \in I$ implies either $f \in I$ or $g^k \in I$ for some $k \in \mathbb{N}$.

## Example

Consider the ring $R = \mathbb{Z}$.

- The prime ideals (excluding 0 and $\mathbb{Z}$) are of the form $I = \langle p \rangle$ for some prime $p$.
- The primary ideals (excluding 0 and $\mathbb{Z}$) are of the form $I = \langle p^k \rangle$ for $k \in \mathbb{N}$.

The following theorem can be thought of as a way to "factor" ideals in polynomial rings, much like how integers can be factored into primes.

## Lasker-Noether Theorem

Every ideal $I$ of $\mathbb{F}[x_1, \dots, x_n]$ can be written as $I = \bigcap_{i=1}^{r} \mathfrak{p}_i$, where $\mathfrak{p}_i$ is a primary ideal. We call this a primary decomposition of $I$. The $\mathfrak{p}_i$ are called primary components.

In general, primary decompositions are hard to compute and need not be unique. But for square-free monomial ideals, they have a simple combinatorial description.

# Ideals and varieties

## Definition

Given an ideal $I \leq \mathbb{F}[x_1, \ldots, x_n]$, the variety of $I$ is its set of common zeros:

$$V(I) := \{x \in \mathbb{F}^n : f(x) = 0 \text{ for all } f \in I\}.$$

The ideal generated by a variety $V \subseteq \mathbb{F}^n$ is

$$I(V) := \left\{f \in \mathbb{F}[x_1, \ldots, x_n] \mid f(v) = 0,\ \forall v \in V\right\}.$$

## Proposition

For any two varieties $V_1$ and $V_2$ in $\mathbb{F}^n$,

$$I(V_1 \cup V_2) = I(V_1) \cap I(V_2).$$

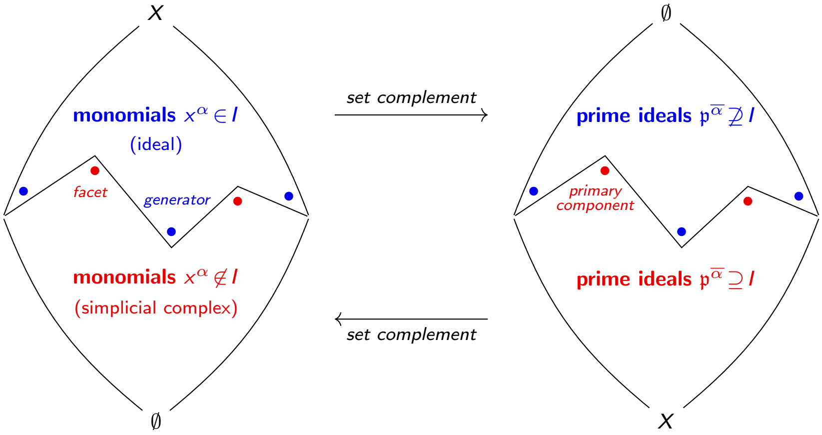
For any $\alpha \subseteq [n]$, define $\mathrm{p}^\alpha = \langle x_i : i \in \alpha \rangle$ and $\mathrm{p}^{\overline{\alpha}} = \mathrm{p}^{[n]-\alpha} = \langle x_i : i \notin \alpha \rangle$. Both are prime.

## Theorem

Let $\Delta$ be a simplicial complex over $[n]$. The Stanley-Reisner ideal of $\Delta$ in $R = \mathbb{F}[x_1, \ldots, x_n]$ is

$$I_{\Delta^c} = \bigcap_{\alpha \in \Delta} \mathrm{p}^{\overline{\alpha}} = \bigcap_{\substack{\alpha \in \Delta \\ \text{maximal}}} \mathrm{p}^{\overline{\alpha}}.$$

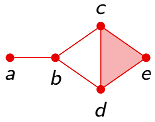
# A summary of Stanley-Reisner theory



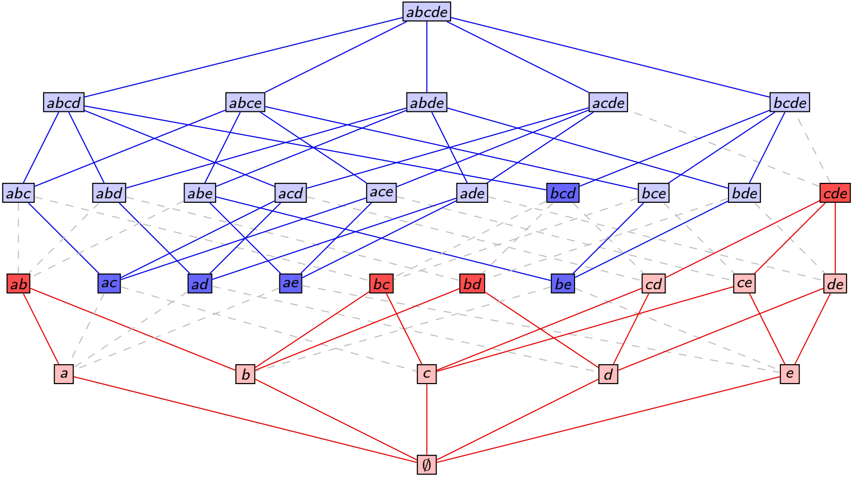

- Alexander duality are the bijections $I \longmapsto \Delta_{I^c}$ and $\Delta \mapsto I_{\Delta^c}$.
- The primary decomposition of a square-free monomial ideal is: $I_{\Delta^c} = \bigcap_{\alpha \in \Delta} \mathfrak{p}^{\overline{\alpha}} = \bigcap_{\substack{\alpha \in \Delta \\ \text{maximal}}} \mathfrak{p}^{\overline{\alpha}}$.

# Example 5, revisited

Consider the simplicial complex $\Delta$:




The ideal is $I_{\Delta^c} = \langle ac, ad, ae, bcd, be \rangle = \langle c, d, e \rangle \cap \langle a, d, e \rangle \cap \langle a, c, e \rangle \cap \langle a, b \rangle$.



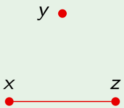
# Computing the primary decomposition

## Example 3a (revisted)

Faces: $\Delta = \{\emptyset, x, y, z, xz\}$   (maximal: $y$, $xz$)

Non-faces: $\Delta^c = \{xy, yz, xyz\}$   (minimal: $xy$, $yz$)

$I_{\Delta^c} = \langle xy, yz\rangle = \langle x, z\rangle \cap \langle y\rangle$



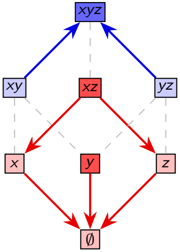

The primary decomposition of $I_{\Delta^c}$ is generated by the complements of the 5 faces in $\Delta$.

This is *not* the *set complement* of $\Delta$, i.e., the 3 non-faces $\Delta^c = \{xy, yz, xyz\}$, but rather,

$$\{\overline{\emptyset}, \overline{x}, \overline{z}, \overline{y}, \overline{xz}\} = \{xyz, yz, xy, xz, y\}.$$

By the previous theorem, the primary decomposition of $I_{\Delta^c}$ is

$$\begin{aligned} I_{\Delta^c} = \langle xy, yz\rangle = \bigcap_{\alpha\in\Delta} \mathfrak{p}^{\overline{\alpha}} &= \mathfrak{p}^{\overline{\emptyset}} \cap \mathfrak{p}^{\overline{x}} \cap \mathfrak{p}^{\overline{z}} \cap \mathfrak{p}^{\overline{y}} \cap \mathfrak{p}^{\overline{xz}} \\ &= \mathfrak{p}^{xyz} \cap \mathfrak{p}^{yz} \cap \mathfrak{p}^{xy} \cap \mathfrak{p}^{xz} \cap \mathfrak{p}^{y} \\ &= \underbrace{\langle x, y, z\rangle \cap \langle y, z\rangle \cap \langle x, y\rangle}_{\text{unnecessary}} \cap \langle x, z\rangle \cap \langle y\rangle \\ &= \langle x, z\rangle \cap \langle y\rangle = \bigcap_{\substack{\alpha\in\Delta \\ \text{maximal}}} \mathfrak{p}^{\overline{\alpha}}. \end{aligned}$$



# Applying Stanley-Reisner theory to algebraic models

Now, we are ready to use Stanley-Reisner theory to infer wiring diagrams.

Here is a summary of the process:

1. Consider every pair of input vectors that give a different output.
2. For each pair, take the monomial $x^\alpha$, where $\alpha \subseteq [n]$ is the set where the entries differ.
3. These generate an ideal. The primary decomposition encodes all minimal wiring diagrams.

## Simplification

We can consider each coordinate independently.

This is best seen with an example. Consider the following Boolean model $f = (f_1, f_2, f_3)$.

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


Thus, we will consider a function $f : \mathbb{F}^n \to \mathbb{F}$ with partial data, and attempt to infer its wiring diagram.

# Data and model spaces

Let $f\colon \mathbb{F}^n \to \mathbb{F}$ be a function, where $\mathbb{F} = \mathbb{F}_p$.

## Definition

Consider a set

$$\mathcal{D} = \big\{(\mathrm{s}_1, t_1), \dots, (\mathrm{s}_m, t_m)\big\}, \qquad \mathrm{s}_i \in \mathbb{F}^n, \quad t_i \in \mathbb{F}$$

of input-output pairs, all $\mathrm{s}_i$ distinct. We call such a set data, and say that $f$ fits the data $\mathcal{D}$ if

$$f(\mathrm{s}_i) = f(s_{i1}, \dots, s_{in}) = t_i, \qquad \text{for all } i = 1, \dots, m.$$

The model space of $\mathcal{D}$ is the set $\mathrm{Mod}(\mathcal{D})$ of all functions that fit the data, i.e.,

$$\mathrm{Mod}(\mathcal{D}) = \big\{f\colon \mathbb{F}^n \to \mathbb{F} \mid f(\mathrm{s}_i) = t_i \text{ for all } i = 1, \dots, m\big\}.$$

For any $f$ in $\mathrm{Mod}(\mathcal{D})$, the support of $f$, denoted $\mathrm{supp}(f)$, is the set of variables on which $f$ depends.

Under a slight abuse of notation, we can think of the support as a subset of $\{x_1, \dots, x_n\}$ or as a subset $\alpha \subseteq [n] = \{1, \dots, n\}$.

Either way, we can write $\mathrm{supp}(f)$ as a string.

# Disposable and non-disposable sets

## Definition

With respect to a set $\mathcal{D}$ of data, a set $\alpha \subseteq [n]$ is:

- disposable if there is some $f \in \text{Mod}(\mathcal{D})$ for which $\text{supp}(f) \cap \alpha = \emptyset$.
- otherwise, it is non-disposable.

## Remark

Let $\mathcal{D}$ be a set of data, and $\alpha, \beta \subseteq [n]$.

(i) If $\alpha$ and $\beta$ are disposable with respect to $\mathcal{D}$, then so is $\alpha \cap \beta$.

(ii) If $\alpha$ and $\beta$ are non-disposable with respect to $\mathcal{D}$, then so is $\alpha \cup \beta$.

## Key point

Let $\mathcal{D}$ be a set of data, and $\alpha, \beta \subseteq [n]$.

(i) The disposable sets form a simplicial complex $\Delta_{\mathcal{D}}$.

(ii) The non-disposable sets generate an ideal

$$I_{\Delta_{\mathcal{D}}^c} = \langle \alpha \mid \alpha \text{ is non-disposable} \rangle \subseteq \mathbb{F}[x_1, \ldots, x_n],$$

called the ideal of non-disposable sets.

# Feasible, infeasible, and min-sets

## Definition

With respect to a set $\mathcal{D}$ of data, a set $\alpha \subseteq [n]$ is:

- feasible if there is there is some $f \in \text{Mod}(\mathcal{D})$ for which $\text{supp}(f) \subseteq \alpha$.
- otherwise, it is infeasible.

## Remarks

- A set $\alpha$ is feasible iff its complement $\overline{\alpha} := [n] - \alpha$ is disposable.
- These are *not* opposite concepts; $\alpha$ can be both feasible and disposable, or neither.

## Key point

Let $\mathcal{D}$ be a set of data, and $\alpha, \beta \subseteq [n]$.

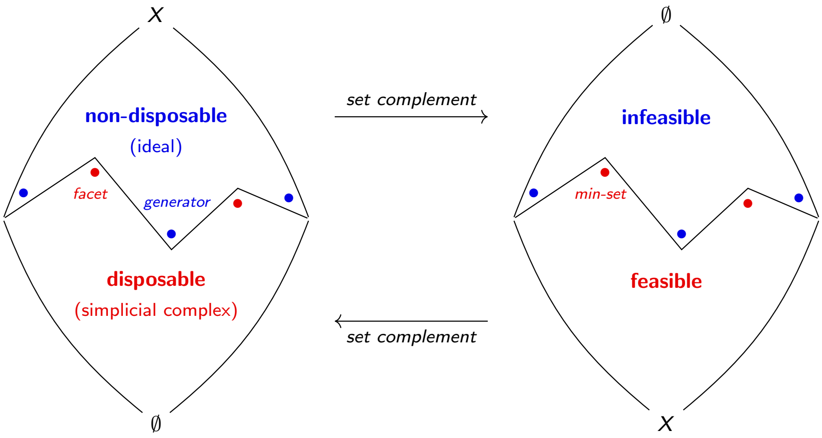
(i) If $\alpha$ and $\beta$ are feasible with respect to $\mathcal{D}$, then so is $\alpha \cup \beta$.

(ii) If $\alpha$ and $\beta$ are infeasible with respect to $\mathcal{D}$, then so is $\alpha \cap \beta$.

## Definition

A minimal feasible set $\alpha \subseteq [n]$ of $\mathcal{D}$ is called a **min-set**. Equivalently, its complement $\overline{\alpha} := [n] - \alpha$ is a maximal disposable set.

# A summary of applying Stanley-Reisner theory to network inference




- Alexander duality are the bijections $I_{\mathcal{D}} \longmapsto \Delta_{I_{\mathcal{D}}^c}$ and $\Delta_{\mathcal{D}} \mapsto I_{\Delta_{\mathcal{D}}^c}$.
- The primary decomposition of the ideal of non-disposable sets is: $I_{\Delta_{\mathcal{D}}^c} = \bigcap_{\overline{\alpha} \text{ is a min-set}} \mathfrak{p}^{\overline{\alpha}}$

# Extracting non-disposable sets from the data

## Theorem (Alexander duality)

There is a bijective correspondence between:

- the simplicial complex $\Delta_{\mathcal{D}}$ of disposable sets,
- the square-free monomial ideal $I_{\Delta^c_{\mathcal{D}}}$ in $\mathbb{F}[x_1, \dots, x_n]$ of non-disposable sets.

For each pair $(\mathsf{s}, t), (\mathsf{s}', t') \in \mathcal{D}$, the coordinates in which the inputs differ can be encoded by the monomial

$$m(\mathsf{s}, \mathsf{s}') := \prod_{s_i \neq s'_i} x_i.$$

By construction, if $t \neq t'$, then $\mathrm{supp}(m(\mathsf{s}, \mathsf{s}'))$ *must* be non-disposable.

## Theorem

The ideal of non-disposable sets is the ideal in $\mathbb{F}_2[x_1, \dots, x_n]$ defined by

$$I_{\Delta^c_{\mathcal{D}}} = \langle m(\mathsf{s}, \mathsf{s}') \mid t \neq t' \rangle = \bigcap_{\substack{\alpha \in \Delta^c_{\mathcal{D}} \\ \text{maximal}}} \mathsf{p}^{\overline{\alpha}} = \bigcap_{\substack{\overline{\alpha} \text{ is a} \\ \text{min-set}}} \mathsf{p}^{\overline{\alpha}}.$$

That is, the generators of the primary components of $I_{\Delta^c_{D}}$ are the min-sets of $\mathcal{D}$.

# Example 3a (continued)

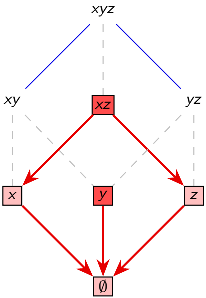
Consider a Boolean function $f\colon \mathbb{F}_2^3 \to \mathbb{F}_2$ with the following partial data:

| $xyz$ | 101 | 000 | 110 |
|---|---|---|---|
| $f(x,y,z)$ | 0 | 0 | 1 |

Using our notation, the *data* $\mathcal{D}$, grouped by output value, is

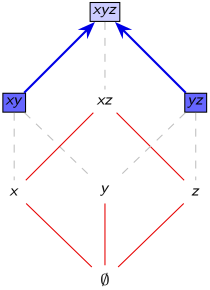
$$\mathcal{D} = \big\{(s_1,t_1),(s_2,t_2),(s_3,t_3)\big\} = \big\{(101,0),(000,0),(110,1)\big\}.$$

Since $t_1 = t_2 \neq t_3$, we compute $m(s_1,s_3) = yz$ and $m(s_2,s_3) = xy$.



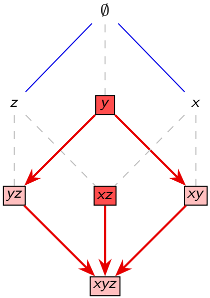

Disposable sets $\Delta_{\mathcal{D}}^c$

Monomials *not* in $I_{\Delta^c}$




Non-disposable sets $\Delta_{\mathcal{D}}^c$

Monomials in $I_{\Delta_{\mathcal{D}}^c}$




Feasible sets of $\Delta$

Min-sets are darker

# Example 3b

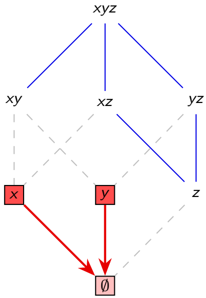
Consider a Boolean function $f\colon \mathbb{F}_2^3 \to \mathbb{F}_2$ with the following partial data:

| $xyz$ | 111 | 000 | 110 |
|---|---|---|---|
| $f(x,y,z)$ | 0 | 0 | 1 |

Using our notation, the *data* $\mathcal{D}$, grouped by output value, is

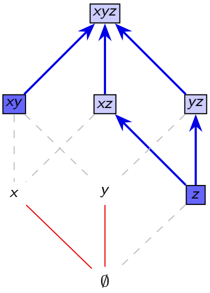
$$\mathcal{D} = \{(\mathrm{s}_1, t_1), (\mathrm{s}_2, t_2), (\mathrm{s}_3, t_3)\} = \{(111, 0), (000, 0), (110, 1)\}.$$

Since $t_1 = t_2 \neq t_3$, we compute $m(\mathrm{s}_1, \mathrm{s}_3) = z$ and $m(\mathrm{s}_2, \mathrm{s}_3) = xy$.



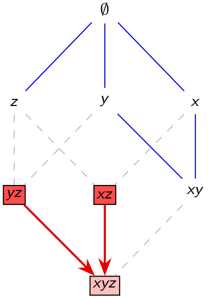

Disposable sets $\Delta_{\mathcal{D}}^c$

Monomials *not* in $I_{\Delta^c}$




Non-disposable sets $\Delta_{\mathcal{D}}^c$

Monomials in $I_{\Delta_{\mathcal{D}}^c}$




Feasible sets of $\Delta$

Min-sets are darker

# Summary so far

The following table summarizes the correspondence between the combinatorial structures in the network inference problem to Stanley-Reisner theory and Alexander duality.

| Inferring wiring diagrams from data | Stanley-Reisner theory |
|---|---|
| Disposable sets of $\mathcal{D}$ | Faces of the simplicial complex $\Delta_{\mathcal{D}}$ |
| Non-disposable sets of $\mathcal{D}$ | The non-faces, $\Delta^c_{\mathcal{D}}$ |
| The ideal $\langle m(\mathrm{s},\mathrm{s}') \mid t \neq t' \rangle$ of non-disposable sets | The Stanley-Reisner ideal $I_{\Delta^c_{\mathcal{D}}}$ |
| Feasible sets of $\mathcal{D}$ | Complements of faces of $\Delta_{\mathcal{D}}$ |
| Min-sets of $\mathcal{D}$ | Complements of max'l faces of $\Delta_{\mathcal{D}}$ $\leftrightarrow$ primary components of $I_{\Delta^c_{\mathcal{D}}}$ |

# Min-sets over non-Boolean fields

Consider a function $f\colon \mathbb{F}_5^5 \to \mathbb{F}_5$ with the following partial data:

$$(s_1, t_1) = (01210, 0),$$
$$(s_2, t_2) = (01211, 0),$$
$$(s_3, t_3) = (01214, 1),$$
$$(s_4, t_4) = (30000, 3),$$
$$(s_5, t_5) = (11113, 4).$$

The monomials $m(s_i, s_j)$ are:

$$m(s_1, s_4) = x_1x_2x_3x_4,$$
$$m(s_1, s_5) = m(s_2, s_5) = m(s_3, s_5) = x_1x_3x_5,$$
$$m(s_2, s_4) = m(s_3, s_4) = m(s_4, s_5) = x_1x_2x_3x_4x_5,$$
$$m(s_1, s_3) = m(s_2, s_3) = x_5.$$

The ideal of non-disposable sets in $\mathbb{F}_2[x_1, x_2, x_3, x_4, x_5]$ is

$$I_{\Delta^c_{\mathcal{D}}} = \langle m(s_i, s_j) \mid t_i \neq t_j \rangle = \langle x_1x_2x_3x_4x_5,\ x_1x_3x_5,\ x_1x_2x_3x_4,\ x_5 \rangle = \langle x_1x_2x_3x_4,\ x_5 \rangle.$$

We can compute the primary decomposition in Macaulay2:

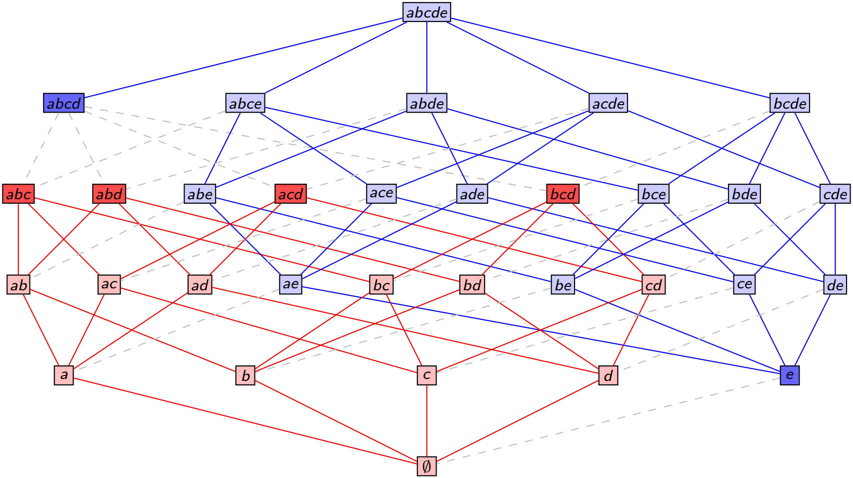
```
R = QQ[x1,x2,x3,x4,x5];
I_nonDisp = ideal(x5, x1*x2*x3*x4);
primaryDecomposition I_nonDisp
```

Output: `{ideal (x1, x5), ideal(x2, x5), ideal(x3, x5), ideal(x4, x5)}`

- Primary decomposition: $I_{\Delta^c_{\mathcal{D}}} = \langle x_1, x_5 \rangle \cap \langle x_2, x_5 \rangle \cap \langle x_3, x_5 \rangle \cap \langle x_4, x_5 \rangle$.
- Unsigned min-sets: $\{x_1, x_5\}$, $\{x_2, x_5\}$, $\{x_3, x_5\}$, $\{x_4, x_5\}$.

# Min-sets over non-Boolean fields

Suppose the ideal of non-disposable sets is $I_{\Delta_{\mathcal{D}}^c} = \langle abcd, e \rangle$.




- Maximum disposable sets: $abc$, $abd$, $acd$, $bcd$.
- Primary decomposition: $I_{\Delta_{\mathcal{D}}^c} = \langle d, e \rangle \cap \langle c, e \rangle \cap \langle b, e \rangle \cap \langle a, e \rangle$.

# Finding signed min-sets of algebraic models

Consider a set of data (i.e., input-output pairs) with all $\mathsf{s}_i$ distinct:

$$\mathcal{D} = \{(\mathsf{s}_1, t_1), \dots, (\mathsf{s}_m, t_m)\}, \qquad \mathsf{s}_i \in \mathbb{F}^n, \quad t_i \in \mathbb{F}.$$

Last time: For each pair $(\mathsf{s}, t), (\mathsf{s}', t') \in \mathcal{D}$, define the monomial $m(\mathsf{s}, \mathsf{s}') := \prod_{s_i \neq s_i'} x_i$.

This time: For each coordinate $i$ that $\mathsf{s}$ and $\mathsf{s}'$ differ, include:

- $x_i$ if the interaction is positive: $\text{sign}(s_i - s_i') = \text{sign}(t_i - t_i')$,
- $\overline{x_i} = x_i + 1$ if the interaction is negative: $\text{sign}(s_i - s_i') \neq \text{sign}(t_i - t_i')$.

Specifically, define the pseudomonomial

$$p(\mathsf{s}, \mathsf{s}') := \prod_{\text{pos. } i} x_i \prod_{\text{neg. } i} \overline{x_i}.$$

**Theorem**

The ideal of signed non-disposable sets in $\mathbb{F}[x_1, \dots, x_n]$ is defined by

$$J_{\Delta^c_{\mathcal{D}}} = \langle p(\mathsf{s}_i, \mathsf{s}_j) \mid t_i \neq t_j \rangle.$$

The primary components of $J_{\Delta^c_{\mathcal{D}}}$ give the signed min-sets.

# Example 3a (revisited)

Consider a Boolean function $f\colon \mathbb{F}_2^3 \to \mathbb{F}_2$ with the following partial data:

| $xyz$ | 111 | 000 | 110 |
|---|---|---|---|
| $f(x,y,z)$ | 0 | 0 | 1 |

The *data* is $\mathcal{D} = \{(s_1,t_1),(s_2,t_2),(s_3,t_3)\} = \{(111,0),(000,0),(110,1)\}$.

Note that

$$p(s_1,s_3) = \overline{z}, \qquad p(s_2,s_3) = xy.$$

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

$$J_{\Delta_{\mathcal{D}}^c} = \langle p(s_1,s_3),\ p(s_2,s_3)\rangle = \langle \overline{z},\ xy\rangle.$$

The following Macaulay2 commands compute the primary decomposition of $J_{\Delta_{\mathcal{D}}^c}$:

```
R = ZZ/2[x,y,z];
J_nonDisp = ideal(z+1, x*y);
primaryDecomposition J_nonDisp
```

Output: `{ideal (z + 1, y), ideal (z + 1, x)}`

- Primary decomposition: $J_{\Delta_{\mathcal{D}}^c} = \langle x, z+1\rangle \cap \langle y, z+1\rangle$.
- Signed min-sets: $\{x,\overline{z}\}$, $\{y,\overline{z}\}$.

# Signed min-sets over non-Boolean fields

Let's compute the pseudomonomials for our previous example of $f\colon \mathbb{F}_5^5 \to \mathbb{F}_5$ with data:

$$\begin{aligned}
(s_1,t_1) &= (01210,0), & p(s_1,s_3) &= p(s_2,s_3) = x_5,\\
(s_2,t_2) &= (01211,0), & p(s_3,s_5) &= x_1\overline{x_3}\,\overline{x_5},\\
(s_3,t_3) &= (01214,1), & p(s_1,s_4) &= x_1\overline{x_2}\,\overline{x_3}\,\overline{x_4},\\
(s_4,t_4) &= (30000,3), & p(s_1,s_5) &= p(s_2,s_5) = x_1\overline{x_3}x_5,\\
(s_5,t_5) &= (11113,4). & p(s_4,s_5) &= \overline{x_1}x_2x_3x_4x_5,\\
& & p(s_2,s_4) &= p(s_3,s_4) = x_1\overline{x_2}\,\overline{x_3}\,\overline{x_4}\,\overline{x_5}.
\end{aligned}$$

The last three are redundant. The ideal of signed non-disposable sets in $\mathbb{F}[x_1,x_2,x_3,x_4,x_5]$ is

$$J_{\Delta^c_{\mathcal{D}}} = \langle p(s_i,s_j) \mid t_i \neq t_j \rangle = \langle x_5,\ x_1\overline{x_3}\,\overline{x_5},\ x_1\overline{x_2}\,\overline{x_3}\,\overline{x_4} \rangle.$$

We can compute the primary decomposition in Macaulay2:

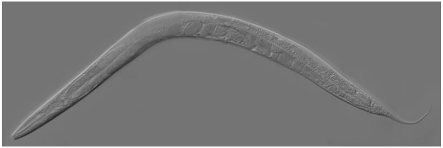
```
R = QQ[x1,x2,x3,x4,x5];
J_nonDisp = ideal(x5, x1*(x3+1)*(x5+1), x1*(x2+1)*(x3+1)*(x4+1));
primaryDecomposition J_nonDisp
```

Output: `{ideal (x5, x3+1), ideal(x5, x1)}`

- Primary decomposition: $J_{\Delta^c_{\mathcal{D}}} = \langle x_1, x_5 \rangle \cap \langle x_3+1, x_5 \rangle$.
- Signed min-sets: $\{x_1, x_5\}$, $\{\overline{x_3}, x_5\}$.

# Application to a real gene network

*Caenorhabditis elegans* is a microscopic roundworm and common organism in biology.



It was the first multicellular organism to have its full genome sequenced, and its nervous system (*connectome*) completely mapped. The latter consists of just 302 neurons and $\approx 7000$ synapses.

In 2012, Stigler & Chamberlin studied a network with 20 genes involved in embryonal development of *C. elegans*.

They discretized data from two time series, $s_1, \dots, s_{10}$ and $u_1, \dots, u_{10}$, to 7 states, i.e., $s_i, u_i \in \mathbb{F}_7^{20}$.

The $i^{\text{th}}$ input state is $s_i$ and the $i^{\text{th}}$ output state is $t_i = f(s_i) = s_{i+1}$, where $f \colon \mathbb{F}_7^{20} \to \mathbb{F}_7^{20}$ is the FDS map of an unknown algebraic model over $\mathbb{F}_7$. Similarly, $v_i = f(u_i) = u_{i+1}$.

# Time-series data

Note that the 20 points in $\mathbb{F}_7^{20}$ in two time series describe 18 input-output pairs.

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | $x_{11}$ | $x_{12}$ | $x_{13}$ | $x_{14}$ | $x_{15}$ | $x_{16}$ | $x_{17}$ | $x_{18}$ | $x_{19}$ | $x_{20}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $s_1$ | 4 | 6 | 5 | 0 | 3 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| $s_2 = t_1$ | 3 | 6 | 5 | 0 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| $s_3 = t_2$ | 1 | 3 | 1 | 0 | 2 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| $s_4 = t_3$ | 1 | 3 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 2 | 1 |
| $s_5 = t_4$ | 0 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 2 | 1 |
| $s_6 = t_5$ | 0 | 2 | 1 | 4 | 6 | 4 | 1 | 3 | 1 | 1 | 0 | 0 | 1 | 2 | 0 | 1 | 0 | 1 | 1 | 1 |
| $s_7 = t_6$ | 0 | 3 | 1 | 6 | 5 | 5 | 1 | 4 | 2 | 1 | 0 | 0 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 0 |
| $s_8 = t_7$ | 1 | 3 | 1 | 4 | 2 | 6 | 1 | 4 | 2 | 3 | 1 | 1 | 3 | 2 | 4 | 4 | 0 | 3 | 3 | 0 |
| $s_9 = t_8$ | 1 | 3 | 1 | 6 | 2 | 5 | 1 | 5 | 1 | 5 | 2 | 5 | 6 | 2 | 5 | 5 | 0 | 4 | 4 | 0 |
| $t_9$ | 0 | 2 | 1 | 4 | 2 | 3 | 1 | 3 | 1 | 4 | 1 | 3 | 4 | 2 | 5 | 3 | 1 | 5 | 5 | 2 |

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | $x_{11}$ | $x_{12}$ | $x_{13}$ | $x_{14}$ | $x_{15}$ | $x_{16}$ | $x_{17}$ | $x_{18}$ | $x_{19}$ | $x_{20}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $u_1$ | 4 | 3 | 3 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| $u_2 = v_1$ | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| $u_3 = v_2$ | 5 | 3 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| $u_4 = v_3$ | 4 | 4 | 3 | 0 | 2 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 |
| $u_5 = v_4$ | 1 | 2 | 1 | 1 | 2 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 2 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| $u_6 = v_5$ | 2 | 3 | 1 | 2 | 4 | 2 | 2 | 2 | 3 | 1 | 0 | 0 | 2 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| $u_7 = v_6$ | 5 | 3 | 1 | 3 | 2 | 2 | 3 | 3 | 5 | 2 | 0 | 1 | 2 | 3 | 1 | 1 | 1 | 1 | 0 | 1 |
| $u_8 = v_7$ | 6 | 5 | 6 | 5 | 4 | 5 | 6 | 4 | 6 | 1 | 0 | 4 | 2 | 2 | 3 | 2 | 1 | 2 | 2 | 0 |
| $u_9 = v_8$ | 3 | 3 | 1 | 4 | 2 | 2 | 4 | 2 | 4 | 3 | 0 | 4 | 5 | 0 | 3 | 2 | 2 | 2 | 4 | 0 |
| $v_9$ | 4 | 5 | 4 | 6 | 2 | 3 | 5 | 6 | 2 | 6 | 2 | 6 | 5 | 2 | 6 | 6 | 1 | 6 | 6 | 3 |

# Application to a real gene network

## Goal

Reconstruct a wiring diagram for the subnetwork of three genes responsible for body wall (mesodermal) tissue development.

| Gene | Variable | Muscle Type |
|---|---|---|
| *hlh-1* | $x_8$ | skeletal |
| *hnd-1* | $x_{18}$ | cardiac |
| *unc-120* | $x_{19}$ | cardiac, smooth, skeletal |

These genes are known to be regulated by the maternally controlled *pal-1* genes.

Though all three regulate a single tissue type in *C. elegans*, some vertebrates have homologous transcription factors related to these genes that regulate three different muscle types.

Understanding their regulatory interactions has implications in human muscle development and disease.

For each gene $j$ of interest ($j = 8, 18, 19$), we extract a set $\mathcal{D}_j$ of data. For example, the data for the *hlh-1* gene is

$$\mathcal{D}_8 = \{(\mathsf{s}_1, t_{18}), (\mathsf{s}_2, t_{28}), \dots, (\mathsf{s}_9, t_{98}), (\mathsf{u}_1, v_{18}), (\mathsf{u}_2, v_{28}), \dots, (\mathsf{u}_9, v_{98})\}.$$

The ideal of non-disposable sets for the *hlh-1* gene is

$$I_{\mathcal{D}_8^c} = \langle \{m(\mathsf{s}_i, \mathsf{s}_j) \mid t_{i8} \neq t_{j8}\} \cup \{m(\mathsf{u}_i, \mathsf{u}_j) \mid v_{i8} \neq v_{j8}\} \cup \{m(\mathsf{s}_i, \mathsf{u}_j) \mid t_{i8} \neq v_{j8}\} \rangle.$$

# The ideal of non-disposable sets for the *hlh-1* gene

$$
\begin{aligned}
I_{\mathcal{D}_8^c} = \langle\, & x_1x_2x_4x_5x_6x_7x_8x_9x_{13}x_{14},\; x_2x_3x_5x_9x_{11}x_{13}x_{14},\; x_2x_4x_6x_9x_{12}x_{13}x_{14},\; x_1x_3x_9x_{11}x_{12}x_{13}x_{14},\\
& x_1x_2x_3x_5x_7x_{11}x_{12}x_{13}x_{15},\; x_2x_3x_5x_7x_{11}x_{13}x_{14}x_{15},\; x_1x_2x_{13}x_{16},\; x_1x_2x_4x_6x_7x_8x_9x_{10}x_{14}x_{15}x_{17},\\
& x_1x_4x_6x_7x_8x_9x_{10}x_{12}x_{13}x_{14}x_{15}x_{17},\; x_1x_2x_3x_4x_5x_6x_7x_8x_9x_{12}x_{13}x_{18},\; x_1x_2x_3x_4x_5x_6x_8x_{12}x_{14}x_{18},\\
& x_1x_2x_3x_5x_6x_7x_8x_9x_{10}x_{11}x_{14}x_{16}x_{18},\; x_1x_2x_3x_5x_6x_8x_{10}x_{11}x_{14}x_{15}x_{16}x_{18},\; x_1x_2x_4x_9x_{19},\\
& x_1x_4x_5x_6x_7x_8x_9x_{13}x_{19},\; x_2x_4x_5x_6x_8x_{14}x_{19},\; x_1x_2x_4x_6x_{12}x_{13}x_{14}x_{19},\; x_1x_4x_5x_6x_8x_{12}x_{13}x_{14}x_{19},\\
& x_1x_5x_6x_7x_8x_9x_{13}x_{16}x_{19},\; x_2x_4x_6x_{12}x_{13}x_{14}x_{16}x_{19},\; x_1x_4x_5x_7x_8x_9x_{10}x_{12}x_{14}x_{15}x_{17}x_{19},\\
& x_1x_2x_3x_4x_6x_{12}x_{18}x_{19},\; x_1x_2x_3x_4x_{13}x_{14}x_{18}x_{19},\; x_4x_6x_8x_9x_{10}x_{11}x_{12}x_{13}x_{15}x_{16}x_{18}x_{19},\\
& x_1x_3x_4x_5x_6x_7x_8x_9x_{11}x_{14}x_{15}x_{16}x_{17}x_{18}x_{19},\; x_1x_6x_7x_8x_9x_{11}x_{12}x_{13}x_{14}x_{15}x_{16}x_{17}x_{18}x_{19},\\
& x_1x_4x_5x_6x_{10}x_{11}x_{12}x_{13}x_{14}x_{15}x_{16}x_{17}x_{18}x_{19},\; x_1x_5x_8x_9x_{10}x_{11}x_{12}x_{13}x_{14}x_{15}x_{16}x_{17}x_{18}x_{19},\\
& x_1x_4x_5x_6x_7x_8x_9x_{13}x_{15}x_{16}x_{17}x_{20},\; x_1x_2x_3x_4x_5x_7x_8x_{11}x_{12}x_{13}x_{18}x_{20},\\
& x_1x_3x_5x_6x_7x_8x_9x_{11}x_{14}x_{18}x_{20},\; x_1x_2x_3x_4x_5x_7x_8x_9x_{13}x_{14}x_{18}x_{20},\\
& x_1x_2x_3x_5x_6x_8x_{11}x_{14}x_{15}x_{18}x_{20},\; x_3x_4x_5x_6x_7x_8x_9x_{10}x_{13}x_{14}x_{15}x_{17}x_{18}x_{20},\\
& x_1x_2x_3x_4x_5x_6x_8x_9x_{12}x_{15}x_{16}x_{17}x_{18}x_{20},\; x_2x_4x_5x_6x_8x_9x_{15}x_{16}x_{17}x_{19}x_{20},\\
& x_2x_3x_5x_8x_9x_{11}x_{12}x_{14}x_{15}x_{19}x_{20},\; x_1x_4x_5x_6x_8x_9x_{15}x_{16}x_{17}x_{19}x_{20},\; x_2x_5x_7x_8x_{11}x_{12}x_{14}x_{19}x_{20},\\
& x_1x_3x_4x_5x_6x_7x_8x_{11}x_{13}x_{14}x_{16}x_{18}x_{19}x_{20},\; x_2x_4x_6x_8x_9x_{10}x_{11}x_{13}x_{14}x_{15}x_{16}x_{18}x_{19}x_{20},\\
& x_4x_6x_8x_{10}x_{11}x_{12}x_{13}x_{14}x_{15}x_{16}x_{18}x_{19}x_{20},\; x_1x_4x_6x_7x_8x_9x_{10}x_{11}x_{13}x_{14}x_{15}x_{16}x_{17}x_{18}x_{19}x_{20},\\
& x_1x_4x_5x_7x_9x_{10}x_{12}x_{13}x_{14}x_{15}x_{16}x_{17}x_{18}x_{19}x_{20},\; x_1x_4x_7x_8x_9x_{10}x_{12}x_{13}x_{14}x_{15}x_{16}x_{17}x_{18}x_{19}x_{20}\,\rangle.
\end{aligned}
$$

# Min-sets of the *hlh-1* gene

The primary decomposition of $I_{\mathcal{D}_8^c}$ consists of 483 primary components (min-sets). That is,

$$I_{\mathcal{D}_8^c} = \bigcap_{i=1}^{483} \mathfrak{p}_i.$$

However, it is known experimentally that *hlh-1* is controlled by the *pal-1* genes (variables $x_1, x_2, x_3$).

Therefore, we can disregard all min-sets that involve none of these variables.

This happens to be 481 of them, leaving two candidates for min-sets of *hlh-1*:

$$\{x_2, x_3, x_8, x_{18}\} \qquad \text{and} \qquad \{x_2, x_3, x_8, x_{19}\}.$$

There are two possible wiring diagrams at the *hlh-1* gene (variable $x_8$):

# Min-sets of the *hnd-1* and *unc-120* genes

Applying a similar process for the other two genes gives:

- 580 min-sets for the *hnd-1* gene,
- 498 min-sets for the *unc-120* gene.

As before, these can be drastically reduced by discarding those that do not contain any of the *pal-1* genes $(x_1, x_2, x_3)$.

Then, they are filtered so that they contain (i) as many of the variables for *hlh-1*, *hnd-1*, *unc-120* $(x_8, x_{18}, x_{19})$ as possible, and (ii) no other variables. The min-sets are:

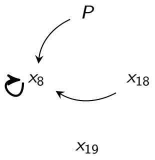
| *hlh-1* $(x_8)$ | *hnd-1* $(x_{18})$ | *unc-120* $(x_{19})$ |
|---|---|---|
| $\{x_2, x_3, x_8, x_{18}\}$ | $\{x_2, x_8, x_{18}\}$ | $\{x_2, x_3, x_8, x_{18}\}$ |
| $\{x_2, x_3, x_8, x_{19}\}$ | $\{x_2, x_8, x_{19}\}$ | $\{x_2, x_3, x_8, x_{19}\}$ |
| | $\{x_3, x_8, x_{19}\}$ | $\{x_2, x_8, x_9, x_{19}\}$ |
| | $\{x_3, x_8, x_9, x_{18}\}$ | |

Collapsing the *pal-1* variables into a single node $P$ gives the following simplified min-sets:

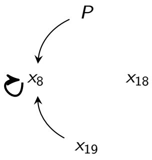
| *hlh-1* $(x_8)$ | *hnd-1* $(x_{18})$ | *unc-120* $(x_{19})$ |
|---|---|---|
| $\{P, x_8, x_{18}\}$ | $\{P, x_8, x_{18}\}$ | $\{P, x_8, x_{18}\}$ |
| $\{P, x_8, x_{19}\}$ | $\{P, x_8, x_{19}\}$ | $\{P, x_8, x_{19}\}$ |

# Minimal wiring diagrams

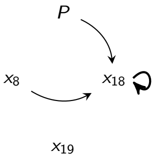
*hlh-1* :



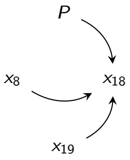

OR



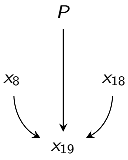

*hnd-1* :



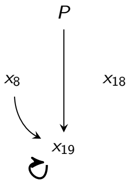

OR




*unc-120* :




OR



# Min-sets of general discrete dynamical systems

The techniques here can be used for more than just algebraic models.

In (Harrington, et al., 2024), the authors recovered the wiring diagram from sampling data from the continuous-space dynamical system $f\colon [0,1]^5 \to [0,1]^5$ defined by

$$(f_1,\ f_2,\ f_3,\ f_4,\ f_5) \ = \ \left(\tfrac{x_1}{(1+x_1)(1+x_2^2)},\quad \tfrac{1}{(1+x_1x_2)(1+x_5)},\quad \tfrac{x_1^2}{(1+x_1^2)(1+x_2)},\quad \tfrac{1}{1+x_2},\quad \tfrac{x_1x_2}{(1+x_1)(1+x_2)}\right).$$

They also applied it to two difference equation models of flour beetle populations

$$\begin{aligned} L_{n+1} &= bA_n \\ P_{n+1} &= (1-\mu_L)L_n \\ A_{n+1} &= (1-\mu_P)P_n + (1-\mu_A)A_n \end{aligned}$$

$$\begin{aligned} L_{n+1} &= bA_n e^{-c_{EA}A_n} e^{-c_{EL}L_n} \\ P_{n+1} &= (1-\mu_L)L_n \\ A_{n+1} &= P_n e^{-c_{PA}A_n} + (1-\mu_A)A_n \end{aligned}$$

and one of a fish population:

$$\begin{aligned} A_{n+1} &= \phi(k_C C_n + k_D D_n + k_E E_n) \\ B_{n+1} &= s_B B_n \\ C_{n+1} &= s_C C_n \\ D_{n+1} &= s_D D_n \\ E_{n+1} &= s_E E_n \end{aligned}$$