An Algebraic Approach to Modeling Biological Systems

Brandy Stigler
Virginia Bioinformatics Institute
Department of Mathematics
Virginia Tech
Biological Motivation

- Understand dynamics in a gene regulatory network.
- Construct mathematical models based on experimental data.

http://plantsci.nottingham.ac.uk/undergraduate/PlantSciBiol.html
Reverse-engineering Problem

Let $T = \{ s_1, \ldots, s_t \} \subset \mathbb{R}^n$ be a time series of experimental data for a network of $n$ genes, where $s_i = (s_{i1}, \ldots, s_{in})$.

**Problem**: Find a biologically relevant function $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that $F(s_i) = s_{i+1}$, for each $s_i \in T$, $i = 1, \ldots, t-1$.

**Caveats**

1. Low resolution of experimental data
2. Experimental data are noisy.

$\implies$ Move to a finite state set
Finite Dynamical Systems

Let $X$ be a finite set.

**Finite dynamical system (FDS)**

$$F = (f_1, \ldots, f_n) : X^n \rightarrow X^n \text{ where}$$

$f_i(x_1, \ldots, x_n)$ is a *local transition function*

$F(x), F^2(x) := F(F(x)), \ldots$ generate the *dynamics* of $F$
Gene regulatory networks can be modeled as FDS.
- Gene $\mapsto$ variable
- Dynamics of a gene $\mapsto$ transition function.

**Theorem:** Let $k$ be a finite field. Then every function $f : k^n \rightarrow k$ is a polynomial.

The FDS $F : k^n \rightarrow k^n$ is a polynomial model.
Reverse-engineering Problem

Let $T = \{s_1, \ldots, s_t\} \subset k^n$ be a time series of states in a finite field $k$.

**Problem**: Find a biologically relevant polynomial model

$$F : k^n \rightarrow k^n$$

such that $F(s_i) = s_{i+1}$,

for each $s_i \in T$.

**Strategy**: Find all such polynomial models; then use biological information to select a candidate model.
As a Matter of Fact…

Suppose $F = (f_1, \ldots, f_n)$ and $G = (g_1, \ldots, g_n)$ are two polynomial models that interpolate $T = \{s_1, \ldots, s_t\}$.

Then $(f_j - g_j)(s_i) = 0$, for each $s_i, j$.

So $f_j - g_j \in I(\{s_1, \ldots, s_{t-1}\})$. 
Ring-theoretic Solution

\[ T = \{ s_1, \ldots, s_t \} \]

\[ R = k[x_1, \ldots, x_n] / \langle x_1^{p} - x_1, \ldots, x_n^{p} - x_n \rangle \]  
ring of poly. functions

\[ I_j = \langle x_1 - s_{j1}, \ldots, x_n - s_{jn} \rangle \]  
ideal of \( s_j \)

\[ I_1, \ldots, I_{t-1} \]  
comaximal ideals of \( R \)

\[ \varphi : R \longrightarrow R/I_1 \times \cdots \times R/I_{t-1} \]  
ring homomorphism

\[ \{ s_{1i}, \ldots, s_{ti} \} \]  
state transitions, gene \( i \)

\[ (s_{2i} + I_1, \ldots, s_{ti} + I_{t-1}) \]  
image of \( \varphi \)

By the Chinese Remainder Theorem,
\[ \exists f_i(x) \in R \text{ with } f_i(x) = s_{j+1,i} \]  
(unique mod. \( ker \varphi = \cap I_j \).)

Virginia Tech
The Algorithm

<table>
<thead>
<tr>
<th>Time</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.6104</td>
<td>1.2042</td>
<td>1.0072</td>
</tr>
<tr>
<td>2</td>
<td>1.7073</td>
<td>1.3252</td>
<td>1.0185</td>
</tr>
<tr>
<td>3</td>
<td>1.7254</td>
<td>1.4118</td>
<td>1.0336</td>
</tr>
<tr>
<td>4</td>
<td>1.7011</td>
<td>1.4616</td>
<td>1.0508</td>
</tr>
<tr>
<td>5</td>
<td>1.6601</td>
<td>1.4814</td>
<td>1.0685</td>
</tr>
</tbody>
</table>

Input
Time series data

Step 1
Discretize
The Algorithm

\[ f_1 = x_1^2 x_3 - x_1^2 + x_1 x_3 \]
\[ f_2 = -x_1^2 x_3 + x_1^2 - x_1 x_3 - x_1 + 1 \]
\[ f_3 = -x_1^2 x_3 - x_1^2 - x_1 x_3 + x_1 + 1 \]

**Step 2**

Find a particular solution for each \( i \):

\[ f_j (x) = \sum_{i=1}^{i-1} s_{i+1,j} r_i (x) \]

\[ r_k (x) = \begin{cases} 
1, & \text{if } x = s_k \\
0, & \text{otherwise} \end{cases} \]

**Step 3**

Construct \( I = \{ h(x_1, ..., x_n) : h(s_i) = 0 \} \).

\[ I = \bigcap I_i = \bigcap < x_1 - s_{i_1}, ..., x_n - s_{i_n} > \]

\[ I = < x_1 + x_2 - 1, x_2 x_3 - x_3^2 + x_2 - x_1, x_2^2 - x_3^2 + x_2 - x_3 > \]
The Algorithm

\[ g_1 = -x_3^2 + x_3 \]
\[ g_2 = x_3^2 - x_3 + 1 \]
\[ g_3 = -x_3^2 + x_2 + 1 \]

Output
Select a polynomial model.

\[ g_i = \overline{gb(i)} \]

• The model is *minimal*: it contains no terms that vanish on the time series.

• If we don’t have more information, we can include more data.

What kind, how much?
**Knock-out Time Series**

**Definition:** $T = \{s_1, \ldots, s_t\}$ is a knock-out time series if there is $1 \leq i \leq t$ such that

$$s_{1i} = \cdots = s_{ti} = 0.$$

To generate a knock-out time series for gene $i$,

$$f_i = 0.$$
Proof of Concept

- Segment polarity network in *D. melanogaster* (Albert and Othmer, 2004)
- Using published functions, we generated (Boolean) wildtype and knock-out time series.

**Results** (Laubenbacher and Stigler, 2004)
- Wildtype: 32% of the interactions (links)
- Wildtype **AND** knock-out: 84% of the interactions.
The Goal

To determine if and why knock-out time series give better results.

Strategy
Discover relationships between the data and (the Gröbner basis of the) the ideal of points.
- The ideal is constructed directly from the data.
- Polynomial models differ by an element of the ideal.

Conjecture: Data must “span” enough of the state space.
\[ R = \mathbb{Z}_2[x_1, x_2, x_3]/ \langle x_i^2 - x_i \rangle \]

\[ f_1 = x_1 + x_2 \]
\[ f_2 = x_2 + x_3 \]
\[ f_3 = x_1 + x_3 \]
Characterizing the Ideal

Determine the leading terms of the generating (Gröbner basis) elements.

**Results**

- The ideal of 2 points will have generators of degree 1 (in $\mathbb{Z}_p$).
- The ideal of 3 points will have generators of degree at most 2 (in $\mathbb{Z}_2$).
- The ideal $I(V(\mathbb{Z}_2^n - (1,\ldots,1))) = \langle x_1x_2\cdots x_n \rangle$. 
Applied Discrete Mathematics Group

**Director**
R. Laubenbacher

**Research Associate**
A. Jarrah

**Graduate Research Assistants**
O. Colon-Reyes   J. McGee
M. Colon-Velez (GBCB) D. Potter
E. Delgado-Eckert P. Vera-Licona
E. Dimitrova