Genes, Rings, and Modeling Stress

Brandy Stigler
Virginia Bioinformatics Institute
Department of Mathematics
Virginia Tech

September 21, 2004
What is Mathematical Biology

A branch of mathematics devoted to the development of theory/framework (algorithms and tools) for analyzing biological systems

- Mathematical modeling
Mathematical Tools

- Continuum mechanics, kinetic theory, stochastic processes, system theory
- Linear algebra, computational algebra
- Algebraic statistics/design of experiments, Bayesian statistics, probability theory
- Graph theory, knot theory, lattice theory
- Numerical analysis
Gene Regulatory Networks

Input signals → Gene regulatory network component → Primary outputs = Changed RNA and protein complements → Terminal outputs = Changed cell behaviors and structures

feedback circuitry

http://doegenomestolife.org
Protein-protein Interaction Networks

http://www.nature.com/ng/journal/v35/n2/images/ng1003-118-F1.gif
Cell Cycle

http://www.dundee.ac.uk/cellcyclegenetics/aspnet.htm
Epidemiological Dynamics

http://www.vu.union.edu/~robleec/math134/equation_overall.htm
# A (Brief) History of MB

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1917</td>
<td>D’Arcy Thompson</td>
</tr>
<tr>
<td></td>
<td><em>On Growth and Form</em>: animals and plants can only be studied using mathematics</td>
</tr>
<tr>
<td>Early 1900s</td>
<td>K. Pearson and J. Blakeman</td>
</tr>
<tr>
<td>1944</td>
<td>Alston Householder</td>
</tr>
<tr>
<td>1947</td>
<td>Nicolas Rashevsky</td>
</tr>
<tr>
<td>1963</td>
<td>Hodgkin and Huxley</td>
</tr>
<tr>
<td>1970</td>
<td>Bernhard Katz</td>
</tr>
<tr>
<td></td>
<td>Nobel Prize for advances in neuromuscular interaction</td>
</tr>
<tr>
<td></td>
<td>Nobel Prize for advances in the chemistry of nerve conduction</td>
</tr>
<tr>
<td></td>
<td>Foundations of PDE models in biology</td>
</tr>
<tr>
<td></td>
<td>Foundations of mathematical neuroscience along with Herbert Landahl</td>
</tr>
</tbody>
</table>
What isn’t Mathematical Biology

Bioinformatics (computational biology)

1. The science of developing and utilizing databases and algorithms to accelerate and enhance biological research (NIEHS, 2002)

2. The science of determining the words, grammar, sentences, and meaning of the sequences of letters (A,C,G,T) produced from genomic research (Wikipedia, 2004)

Data mining and analysis of genomic data, sequence alignment, protein structure prediction, systems biology, protein-protein interactions
[B]ioinformatics…
“…seems to focus almost exclusively on specific algorithms that can be applied to large molecular biological data sets…”
…[whereas] mathematical biology…
“…includes things of theoretical interest which are not necessarily algorithmic, not necessarily molecular in nature, and are not necessarily useful in analyzing collected data.”

http://www.colorbasepair.com/what_is_bioinformatics.html

“Biologists can be divided into two classes: experimentalists who observe things that cannot be explained, and theoreticians who explain things that cannot be observed.”

Aharon Katzir-Katchalsky or George Oster, 1982
Where it’s at

• Centre for Mathematical Biology at the University of Oxford
  – Spatial and spatiotemporal pattern formation
  – Cancer therapy
  – Behavior dynamics in ecology
• Mathematical Biosciences Institute at Ohio State University
  – To develop mathematical theories, statistical methods, and computational algorithms for the solution of fundamental problems in the biosciences
• Virginia Bioinformatics Institute
  – COPASI (Complex Pathway Simulator)
  – Metabolic engineering of vitamin C
  – Infectious diseases
• University graduate programs
  – NCSU Biomathematics Graduate Program
  – UCLA Biomathematics PhD Program
  – VT Genetics, Bioinformatics, and Computational Biology
Computational Algebra in MB

We will apply computational algebra to model biochemical networks, in particular gene regulatory networks.

http://chem1.eng.wayne.edu/~yhuang/BioModeling.htm
Let $x_1, \ldots, x_n$ be nodes in the network, each with states in a finite set $X$.

**Finite dynamical system (FDS):**

$$f_i(x_1, \ldots, x_n) : X^n \to X, \ i = 1, \ldots, n \quad (\text{transition function})$$

$$f = (f_1, \ldots, f_n) \ 	ext{where} \ f(x_1, \ldots, x_n) : X^n \to X^n$$

**Dynamics:**

For any $x \in X^n$, dynamics = iterates of $f$

$$f(x), f^2(x) := f(f(x)), \ldots$$
Model Assumptions

• Gene regulatory networks can be modeled as FDS.
  – Gene $\Rightarrow$ variable
  – Dynamics of a gene $\Rightarrow$ transition function.

• If $|X| = p$ (prime), then $X$ is a finite field ($k$).

*Theorem*: Every function $f : k^n \rightarrow k$ is a polynomial function.

The FDS $f : k^n \rightarrow k^n$ is a *polynomial model*.
## Common Modeling Approaches

<table>
<thead>
<tr>
<th>Forward/bottom-up modeling</th>
<th>Reverse/top-down modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Starting point: known inferences</td>
<td>- Starting point: experimental data</td>
</tr>
<tr>
<td>- Decide on a modeling framework</td>
<td>- Decide on a modeling framework</td>
</tr>
<tr>
<td>- Build model from constraints</td>
<td>- Build model that fits data</td>
</tr>
</tbody>
</table>

**Benefits:** inferences built into model

**Problems:** too much information, constraints built into model

**Benefits:** no assumptions about interrelationships

**Problems:** multiple models
Reverse-engineering Problem

Let $T = \{s_1, \ldots, s_t\} \subset k^n$ be a time series of states where $s_i = (s_{i1}, s_{i2}, \ldots, s_{in})$.

Find $f : k^n \rightarrow k^n$ such that $f(s_i) = s_{i+1}$ for $i = 1, \ldots, t-1$.

Since $f = (f_1, \ldots, f_n)$, we have the reduced problem:

Find $f_j : k^n \rightarrow k$ such that $f_j(s_i) = s_{i+1,j}$ for $i = 1, \ldots, t-1$ and $j = 1, \ldots, n$. 

Virginia Tech
Chinese Remainder Theorem (CRT): 
$I_1, \ldots, I_t$ comaximal ideals of ring $R$
$g : R \to R/I_1 \times \cdots \times R/I_t$ ring homomorphism

Then,
$g$ is onto and the preimage is unique modulo
$\ker g = I_1 \times \cdots \times I_t = I_1 \cap \cdots \cap I_t$. 
Apply Theorem one node at a time.

State transitions of gene $i : \{s_{1i}, \ldots, s_{ti}\}$

Ideal of $s_j$:

$I_j = \langle x_1 - s_{j1}, \ldots, x_n - s_{jn} \rangle$ for each $j = 1, \ldots, t-1$

Image = $(s_{2i} + I_1, \ldots, s_{ti} + I_{t-1})$

By the CRT, there exists $f_i(x)$ with $f_i(x) = s_{j+1,i}$

$j = 1, \ldots, t-1$ (unique mod ker $g$).
The Algorithm

**Input**: time series \( T = \{s_1, \ldots, s_t\} \) of expression data.

1. Choose \( p = \text{char} \ k \) and discretize \( T \).

2. Find a particular solution \( f^0 = (f_1, \ldots, f_n) \) so that \( f_j(s_i) = s_{i+1,j} \).
   *(Lagrange interpolating function)*

3. Construct \( I = \{h(x_1, \ldots, x_n) : h(s_i) = 0\} \).
   *(ideal of points = vanishing polynomials)*

**Output**: all models \( f^0 + I^n := \{f^0 + h : h \in I^n\} \)

**Selection**: \( f^0_\text{red} = \text{reduction mod } I \) (term order)
Choose a term ordering.

- Order of variables matters: *variable ordering*
- Order of terms matters: *term ordering*

By changing the variable and term orders, different models result.

The remainder $f^0$ is minimal in that it contains no *vanishing terms* (terms = 0 on T).
Example of a Variable Order

Let $f(x, y) = x^2 + x + y$ and $g(x, y) = x + y$.

If $x > y$, then $f \mod g = (x^2 + x + y) \mod (x + y) = y^2$.

If $y > x$, then $f \mod g = (y + x^2 + x) \mod (y + x) = x^2$.

Role in modeling: In computing the function for a gene, the order can determine what variables the function will contain.
Example of a Term Order

Let $f(x,y) = x^2 + xy^2$.

If $x > y$ in Lex (alphabet order),
then $f = x^2 + xy^2$.

If $x > y$ in Grevlex (high degrees come first),
then $f = xy^2 + x^2$.

Role in modeling: The order may determine the degree or number of variables in a function.
# Algorithm in Action: Step 1

<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>

<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</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Algorithm in Action: Steps 2 & 3

\[ f = (f_1, f_2, f_3) \text{ in } \mathbb{Z}/3[x_1, x_2, x_3] \text{ where} \]

\[ f_1 = -x_3^2 + x_3 \]

\[ f_2 = x_3^2 - x_3 + 1 \]

\[ f_3 = -x_3^2 + x_2 + 1 \]

\[ I = < x_1 + x_2 - 1, \ x_2 x_3 - x_3^2 + x_2 - x_1, \ x_2^2 - x_3^2 + x_2 - x_3 > \]
Example of a Wiring Diagram

\[ f = (f_1, f_2, f_3) \text{ over } \mathbb{Z}/3 \text{ where} \]

\[ f_1 = -x_3^2 + x_3 + (x_1 + x_2 - 1) \]

\[ f_2 = x_3^2 - x_3 + 1 \]

\[ f_3 = -x_3^2 + x_2 + 1 \]

\[ I = \langle x_1 + x_2 - 1, x_2 x_3 - x_3^2 + x_2 - x_1, x_2^2 - x_3^2 + x_2 - x_3 \rangle \]
Yeast Project

Develop mathematical tools to model biochemical networks from experimental data.

Apply tools oxidative stress response network in *S. cerevisiae*.
- Gene expression data
- Protein concentrations
- Metabolic concentrations

Discover type and amount of data required for model:
- time series vs. single time point
- perturbations vs. wildtype

Design experiments for models of stress response.
I’m Not Stressed!!

Oxidative stress …

• is common to all aerobically grown organisms.

• results as a natural consequence of aerobic processes.

• can cause various physiological disorders: Parkinson’s disease, stomach cancers.
Project Goals

• Construct models for glutathione network
  – Polynomial model
  – Continuous (ODE) model

• Develop framework/theory for integrating both models.