A design principle in biochemical reaction networks based on realization theory

Bassam Bamieh
University of California, Santa Barbara

IMA, April ’08
Outline

• Some special features of the differential equations of biochemical reaction networks

• Why are there so many reactant species?

• Possible connection with a nonlinear realization problem

• A conjecture and a problem formulation

• A hack for the solution
Some features of the equations

- Biochemical reaction networks seem to always contain terms with special forms
- Many reactant species, but "function" appears to be described by a small subset
- e.g. Circadian Oscillation in Drosophila:

\[
\dot{x}_4 = v_3p \frac{x_3}{k_{3p} + x_3} - v_{4p} \frac{x_4}{k_{4p} + x_4} - k_3 x_4 x_8 + k_4 x_9
\]

\[
- v_{dp} \frac{x_4}{K_{dp} + x_4} - k_{d} x_4
\]

\[
\dot{x}_5 = K_{iT}^n \frac{v_{st}}{K_{iT}^n + x_{10}} - v_{mt} \frac{x_5}{k_{mt} + x_5} - k_{d} x_5
\]

- Each special form ⇐⇒ Specific biochemical reaction mechanism
A model reduction question

Given a biochemical reaction network, say

$$\frac{d}{dt} \begin{bmatrix} x_1 \\ \vdots \\ x_{30} \end{bmatrix} = f(x_1, \ldots, x_{30})$$ (1)

If function is described by a small number of states, say $x_1, x_{12}$, does there exist a reduced model

$$\frac{d}{dt} \begin{bmatrix} x_1 \\ x_{12} \end{bmatrix} = g(x_1, x_{12})$$ (2)

such that

behavior of $(x_1, x_{12})$ in (1) $\iff$ behavior of $(x_1, x_{12})$ in (2)
A Similar Question in Reverse

Given a low order system

\[ \dot{z} = g(z), \]

\(z\): *small number of variables*, \(g\): *possibly complicated function*,

can this system be imbedded in

\[ \dot{X} = F(X), \]

\(X\): *possibly large number of variables*, \(F\) has *special form*?

\(\{z_1, \ldots, z_r\} \subset \{X_1, \ldots, X_n\}\)

embedded: Every trajectory of \(z\) is part of a trajectory of \(X\)

A nonlinear realization problem
Implementation with *Linear Elements: Carleman Linearization*

**Example:** \[ \dot{z} = g(z) = z^2 \]

**Define:** \( X_1 \) := \( z \), \( X_2 \) := \( z^2 \), \ldots, \( X_n \) := \( z^n \), \ldots

\[
\begin{bmatrix}
X_1 \\
X_2 \\
X_3 \\
\vdots \\
X_4
\end{bmatrix}
= 
\begin{bmatrix}
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & 0 \\
0 & 0 & 0 & 3 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 4
\end{bmatrix}
\begin{bmatrix}
X_1 \\
X_2 \\
X_3 \\
\vdots \\
X_4
\end{bmatrix}, \quad \text{i.e.} \quad \dot{\mathbf{x}} = A \mathbf{x}
\]

*The original system is imbedded in this linear system*

Easily generalizable to vector \( z \) and analytic \( g \) using

\[ X(i_1, \ldots, i_n) := z^{i_1} \cdots z^{i_n} \]
Implementation with *Quadratic* Elements

Consider: \[ \dot{z} = g(z), \]

Can embed in a larger system

\[ \dot{x} = Q(x) \]

where \( Q \) is *Quadratic* \( g \) polynomial \( \Rightarrow x \) finite \( x, Q \) non-unique

(Starkl & del Re, ACC, CDC '03)

Note:

- Can approximate any system with \( Q \) a neural network
- Many other possibilities for the building blocks of \( Q \)
Example: $\dot{z} = z^4$

Define $x_i := z^i$

\[
\begin{align*}
\dot{x}_1 & = x_4 \\
\dot{x}_4 & = 4z^3 \dot{z} = 4z^3 z^4 = 4x_3 x_4 \\
\dot{x}_3 & = 3z^2 z^4 = 3x_2 x_4 = 3x_3^2 \\
\dot{x}_2 & = 2z z^4 = 2x_1 x_4 
\end{align*}
\]

Choice of $Z$ and RHS are non-unique!

Procedure terminates in finite number of steps
Realization (Implementation) Issues

Goal: Realize given dynamics \( \dot{z} = g(z) \)
with a given set of motifs (building blocks)

- **Choice of new intermediate variables**
  - Monomials \( x_i := z^i \)
  - Orthogonal Polynomials \( x_i := J_i(z) \)
  - ...  

- **Available building blocks, e.g.**
  - Quadratic terms \( k_{ij}x_ix_j \), restrictions on reaction rates \( k_{ij} \)
  - Hill terms \( \frac{x_i}{K+x_i} \)

- **Original dynamics are an invariant manifold of** \( \dot{z} = g(z) \)
  - for implementation: must be a *stable* invariant manifold
A Design Principle

- Design dynamics for prescribed function, e.g. as $p$ increases, oscillator $\rightarrow$ bistable switch

- Design vector field $f$ for dynamics $\dot{z} = f(z, p)$
  - Can generally be done with 2 or 3 states.
  - but $f$ is not realizable with given motifs

- Apply an expansion procedure to imbed dynamics in a larger system

$$\dot{z} = g(z, p)$$
  - $g$ constructed from available motifs
Why so many intermediate reactant species?

One answer:
To realize function using building blocks available with biochemistry

In other words:
Biological function may be easily described by \( \dot{z} = g(z) \), w/ small \( z \)

But \( g \) may contain terms not realizable with biochemical reactions
The Reverse Embedding Problem in General

Given a large system

\[ \dot{x} = f(x), \]

with specified building blocks (allowable reaction terms).

Did it arise from the “expansion” of a smaller system

\[ \dot{z} = g(z), \]

where \( z \) a subset of \( x \)?

Possibly much more interesting than forward embedding problems which are hard to employ
A Hack for the Embedding Problem

- Start with original large system

\[ \dot{x} = f(x) \]

- Simulate to produce trajectories “representative” of typical bio-function

- Identify a subset of states (relabel them \( z \)) that describe bio-function

- Use simulation data to fit a model

\[ \dot{z} = g(z), \]

where \( g \) is parametrized, e.g. polynomial of some order.
Reformulation as Linear Regression

For polynomial $g$, fitting can be made into linear regression. Define

$$
\begin{align*}
    z^{(1)} &= z \\
    z^{(2)} &= z \otimes z \\
    z^{(3)} &= z \otimes z^{(2)} \\
    &\vdots
\end{align*}
$$

(3)

Then

$$
\begin{align*}
    g(z) &= G Z, \\
    Z &:= [1 \ z' \ z^{(2)'} \ldots \ z^{(m)'}]',
\end{align*}
$$

Coefficients of polynomial $g$ are entries of matrix $G$

Fitting trajectories to $\dot{z} = GZ \iff$ Linear regression for entries of $G$
Example: Circadian Oscillation in Drosophila

Original model has 10 species

Function is described by 2: PER & TIM

Reduced model in example shown: 2 states, $g$ of degree 5

Original PER/TIM (solid)
Reduced PER/TIM (dots)
Vector field of the reduced model
Limit cycle trajectory of original system  
(Bamieh & Giarre, ACC ’07)
• Hack is easily generalizable to include parameters and/or inputs

\[ \dot{z} = g(z, p, u) \]

• Difficult to compare reduced and original model

• Need better quantification of reduction error or comparison of behavior of original and reduced

• The original forward and reverse mathematical problems are open