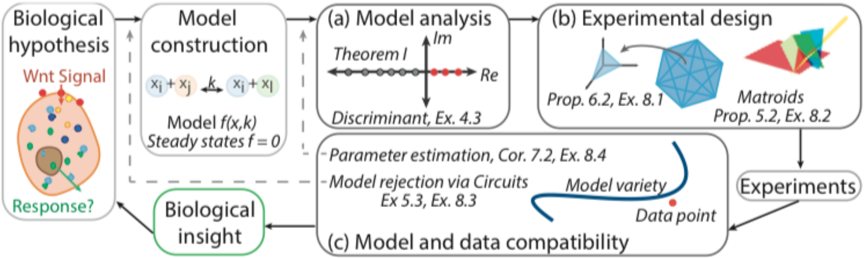
Identifiability of linear compartment models

Anne Shiu
Texas A&M University

ICERM
15 November 2018



From *Algebraic Systems Biology: A Case Study for the Wnt Pathway*

(Elizabeth Gross, Heather Harrington, Zvi Rosen, Bernd Sturmfels 2016).

Outline

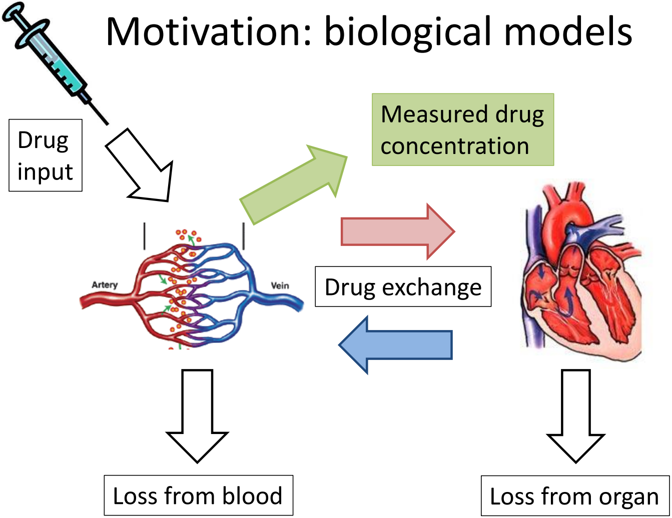
- Introduction: Linear compartment models
- Identifiability (via differential algebra)
- The singular locus

Joint work with
Elizabeth Gross, Heather Harrington, and Nicolette Meshkat

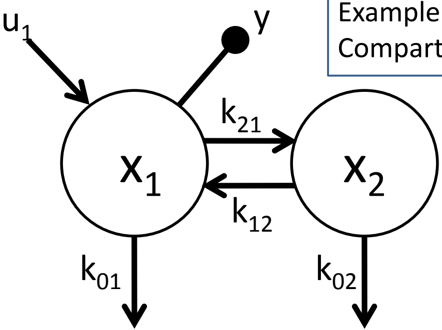
`arXiv:1709.10013` and `arXiv:1810.05575`

Introduction

Motivation: biological models



Compartment model



Example: Linear 2-Compartment Model

$$\dot{x}_1 = -(k_{01} + k_{21})x_1 + k_{12}x_2 + u_1$$
$$\dot{x}_2 = k_{21}x_1 - (k_{02} + k_{12})x_2$$
$$y = x_1$$

Structural identifiability: Recover parameters k_{ij} from perfect input-output data $u_1(t)$ and $y(t)$? (Bellman & Astrom 1970)

Identifiability via differential algebra[1]:
Which models are identifiable?

[1] Ljung and Glad 1994

Input-output equations

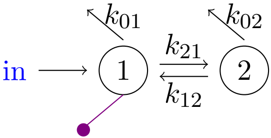
- *Setup:* a linear compartment model
- m = number of compartments
- **Input-output equation**: an equation that holds along any solution of the ODEs,

Input-output equations

- *Setup:* a linear compartment model
- m = number of compartments
- **Input-output equation**: an equation that holds along any solution of the ODEs, involving only input variables u_i and output variables y_i (and parameters k_{ij}), and their derivatives

Input-output equations

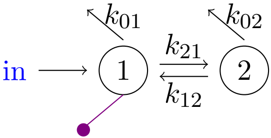
- *Setup:* a linear compartment model
- m = number of compartments
- **Input-output equation**: an equation that holds along any solution of the ODEs, involving only input variables u_i and output variables y_i (and parameters k_{ij}), and their derivatives
- Example, continued:



$$y_1^{(2)} + (k_{01} + k_{02} + k_{12} + k_{21})\, y_1' + (k_{01}k_{12} + k_{01}k_{02} + k_{02}k_{21})\, y_1 = (k_{02} + k_{12})\, u_1$$

Input-output equations

- *Setup:* a linear compartment model
- m = number of compartments
- **Input-output equation**: an equation that holds along any solution of the ODEs, involving only input variables u_i and output variables y_i (and parameters k_{ij}), and their derivatives
- Example, continued:



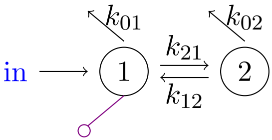
$$y_1^{(2)} + (k_{01} + k_{02} + k_{12} + k_{21})\, y_1' + (k_{01}k_{12} + k_{01}k_{02} + k_{02}k_{21})\, y_1 \;=\; (k_{02} + k_{12})\, u_1$$

- *Input-output equations* come from the elimination ideal:

$$\langle \text{ differential eqns., output eqns. } y_i = x_j \text{ , their } m \text{ derivatives } \rangle$$

$$\cap \;\; \mathbb{C}(k_{ij})[u_i^{(k)}, y_i^{(k)}]$$

Input-output equations, continued

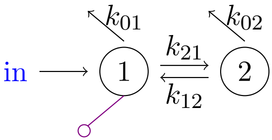


$$A = \begin{pmatrix} -k_{01} - k_{21} & k_{12} \\ k_{21} & -k_{02} - k_{12} \end{pmatrix} \qquad x'(t) = Ax(t) + u(t)$$

- **Proposition** (Meshkat, Sullivant, Eisenberg 2015):
  For a linear compartment model with input and output in compartment-1 only, the input-output equation is:

$$\det(\partial I - A) y_1 = \det\left((\partial I - A)_{11}\right) u_1 \ .$$

Input-output equations, continued



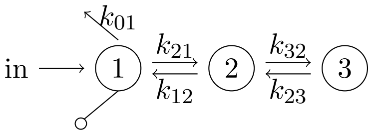
$$A = \begin{pmatrix} -k_{01} - k_{21} & k_{12} \\ k_{21} & -k_{02} - k_{12} \end{pmatrix} \qquad x'(t) = Ax(t) + u(t)$$

- **Proposition** (Meshkat, Sullivant, Eisenberg 2015):
  For a linear compartment model with input and output in compartment-1 only, the input-output equation is:

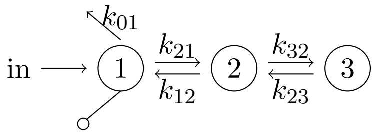
$$\det(\partial I - A) y_1 = \det\left((\partial I - A)_{11}\right) u_1 .$$

- Proof uses Cramer's Rule and Laplace expansion

Input-output equations, continued



Input-output equations, continued

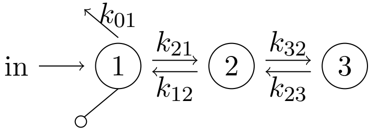


$$\det(\partial I - A) y_1 \quad = \quad \det\left((\partial I - A)_{11}\right) u_1$$

$$\det \begin{pmatrix} d/dt + k_{01} + k_{21} & -k_{12} & 0 \\ -k_{21} & d/dt + k_{12} + k_{32} & -k_{23} \\ 0 & -k_{32} & d/dt + k_{23} \end{pmatrix} y_1$$

$$= \quad \det \begin{pmatrix} d/dt + k_{12} + k_{32} & -k_{23} \\ -k_{32} & d/dt + k_{23} \end{pmatrix} u_1$$

Input-output equations, continued



$$\det(\partial I - A)y_1 \;=\; \det\left((\partial I - A)_{11}\right) u_1$$

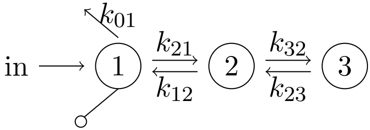
$$\det\begin{pmatrix} d/dt + k_{01} + k_{21} & -k_{12} & 0 \\ -k_{21} & d/dt + k_{12} + k_{32} & -k_{23} \\ 0 & -k_{32} & d/dt + k_{23} \end{pmatrix} y_1$$

$$= \;\det\begin{pmatrix} d/dt + k_{12} + k_{32} & -k_{23} \\ -k_{32} & d/dt + k_{23} \end{pmatrix} u_1$$

... expands to the *input-output equation*:

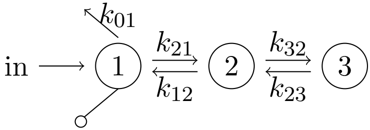
$$y_1^{(3)} + (k_{01} + k_{12} + k_{21} + k_{23} + k_{32})\, y_1^{(2)}$$
$$+ (k_{01}k_{12} + k_{01}k_{23} + k_{01}k_{32} + k_{12}k_{23} + k_{21}k_{23} + k_{21}k_{32})\, y_1' + (k_{01}k_{12}k_{23})\, y_1$$
$$= \; u_1^{(2)} + (k_{12} + k_{23} + k_{32})\, u_1' + (k_{12}k_{23})\, u_1 \;.$$

Coefficients of input-output equations



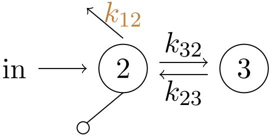
$$
\begin{aligned}
y_1^{(3)} &+ (k_{01} + k_{12} + k_{21} + k_{23} + k_{32})\, y_1^{(2)} \\
&+ (k_{01}k_{12} + k_{01}k_{23} + k_{01}k_{32} + k_{12}k_{23} + k_{21}k_{23} + k_{21}k_{32})\, y_1' + (k_{01}k_{12}k_{23})\, y_1 \\
&\quad = \; u_1^{(2)} + (k_{12} + k_{23} + k_{32})\, u_1' + (k_{12}k_{23})\, u_1 \; .
\end{aligned}
$$

Coefficients of input-output equations



$$
\begin{aligned}
y_1^{(3)} &+ (k_{01} + k_{12} + k_{21} + k_{23} + k_{32})\, y_1^{(2)} \\
&+ (k_{01}k_{12} + k_{01}k_{23} + k_{01}k_{32} + k_{12}k_{23} + k_{21}k_{23} + k_{21}k_{32})\, y_1' + (k_{01}k_{12}k_{23})\, y_1 \\
&= u_1^{(2)} + (k_{12} + k_{23} + k_{32})\, u_1' + (k_{12}k_{23})\, u_1 \;.
\end{aligned}
$$

- coefficient of $y_1^{(i)}$ corresponds to forests with $(3-i)$ edges and ≤ 1 outgoing edge per compartment
- coefficient of $u_1^{(i)}$ corresponds to $(n-i-1)$-edge forests:



- **Thm 1**: The coefficients correspond to forests in model.

Identifiability

$$
\begin{aligned}
y_1^{(3)} &+ (k_{01} + k_{12} + k_{21} + k_{23} + k_{32})\, y_1^{(2)} \\
&+ (k_{01}k_{12} + k_{01}k_{23} + k_{01}k_{32} + k_{12}k_{23} + k_{21}k_{23} + k_{21}k_{32})\, y_1' + (k_{01}k_{12}k_{23})\, y_1 \\
&\quad = \; u_1^{(2)} + (k_{12} + k_{23} + k_{32})\, u_1' + (k_{12}k_{23})\, u_1 \; .
\end{aligned}
$$

- (Generic, local) *identifiability*: can the parameters k_{ij} be recovered from coefficients of input-output equations?

Identifiability

$$y_1^{(3)} + (k_{01}+k_{12}+k_{21}+k_{23}+k_{32})\, y_1^{(2)}$$
$$+ (k_{01}k_{12}+k_{01}k_{23}+k_{01}k_{32}+k_{12}k_{23}+k_{21}k_{23}+k_{21}k_{32})\, y_1' + (k_{01}k_{12}k_{23})\, y_1$$
$$= u_1^{(2)} + (k_{12}+k_{23}+k_{32})\, u_1' + (k_{12}k_{23})\, u_1 \ .$$

- (Generic, local) *identifiability*: can the parameters k_{ij} be recovered from coefficients of input-output equations?

$$\mathbb{R}^5 \to \mathbb{R}^5$$
$$(k_{01}, k_{12}, k_{21}, k_{23}, k_{32}) \mapsto (k_{01}+k_{12}+k_{21}+k_{23}+k_{32}, \ \ldots \ )$$

- Solve directly, or use ...
- **Proposition** (Meshkat, Sullivant, Eisenberg 2015): Identifiable $\Leftrightarrow$ Jacobian matrix of coefficient map has (full) rank = number of parameters

Identifiability

$$y_1^{(3)} + (k_{01} + k_{12} + k_{21} + k_{23} + k_{32})\, y_1^{(2)}$$
$$+ (k_{01}k_{12} + k_{01}k_{23} + k_{01}k_{32} + k_{12}k_{23} + k_{21}k_{23} + k_{21}k_{32})\, y_1' + (k_{01}k_{12}k_{23})\, y_1$$
$$= u_1^{(2)} + (k_{12} + k_{23} + k_{32})\, u_1' + (k_{12}k_{23})\, u_1 \ .$$

- (Generic, local) *identifiability*: can the parameters k_{ij} be recovered from coefficients of input-output equations?

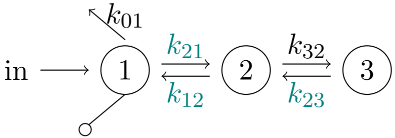
$$\mathbb{R}^5 \rightarrow \mathbb{R}^5$$
$$(k_{01}, k_{12}, k_{21}, k_{23}, k_{32}) \mapsto (k_{01} + k_{12} + k_{21} + k_{23} + k_{32}, \ \ldots \ )$$

- Solve directly, or use ...
- **Proposition** (Meshkat, Sullivant, Eisenberg 2015): Identifiable ⇔ Jacobian matrix of coefficient map has (full) rank = number of parameters *generically*

The singular locus

Definition

- Focus on the *non-identifiable* parameters:
  the **singular locus** is where the Jacobian matrix of coefficient map is rank-deficient.
- Example, continued:



The equation of the singular locus is:

$$\det \mathrm{Jac} \;=\; k_{12}^2 k_{21} k_{23} \;=\; 0\ .$$

Identifiable submodels

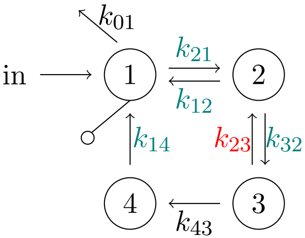
- *Motivation*: drug targets
- **Thm 2**: Let $\mathcal{M}$ be an identifiable linear compartment model, with singular-locus equation f. Let $\widetilde{\mathcal{M}}$ be obtained from $\mathcal{M}$ by deleting edges $\mathcal{I}$.

  If $f \notin \langle k_{ji} \mid (i,j) \in \mathcal{I} \rangle$, then $\widetilde{\mathcal{M}}$ is identifiable.

Identifiable submodels

- *Motivation*: drug targets
- **Thm 2**: Let $\mathcal{M}$ be an identifiable linear compartment model, with singular-locus equation f. Let $\widetilde{\mathcal{M}}$ be obtained from $\mathcal{M}$ by deleting edges $\mathcal{I}$.

  If $f \notin \langle k_{ji} \mid (i,j) \in \mathcal{I} \rangle$, then $\widetilde{\mathcal{M}}$ is identifiable.
- Example:

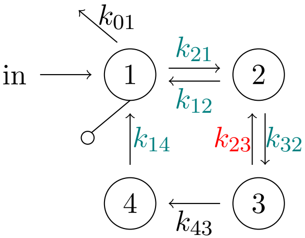


$$f \;=\; k_{12}k_{14}k_{21}^2k_{32}(k_{12}k_{14} - k_{14}^2 - \ldots)(k_{12}k_{23} + k_{12}k_{43} + k_{32}k_{43}) \;.$$

Identifiable submodels

- *Motivation*: drug targets
- **Thm 2**: Let $\mathcal{M}$ be an identifiable linear compartment model, with singular-locus equation f. Let $\widetilde{\mathcal{M}}$ be obtained from $\mathcal{M}$ by deleting edges $\mathcal{I}$.

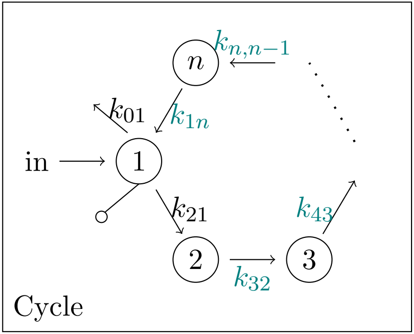
  If $f \notin \langle k_{ji} \mid (i,j) \in \mathcal{I} \rangle$, then $\widetilde{\mathcal{M}}$ is identifiable.
- Example:

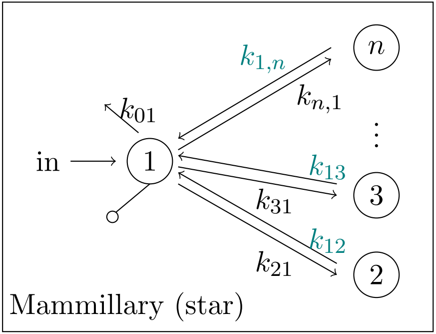


$$f = k_{12}k_{14}k_{21}^2k_{32}(k_{12}k_{14} - k_{14}^2 - \dots)(k_{12}k_{23} + k_{12}k_{43} + k_{32}k_{43}) .$$

- *Converse is false*: deleting k_{12} and k_{23} is identifiable!

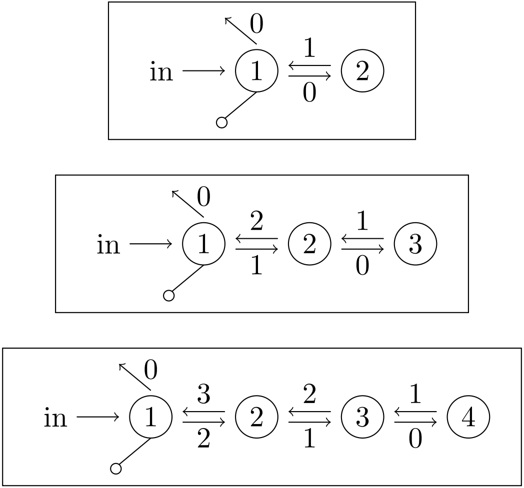
Cycle and mammillary models



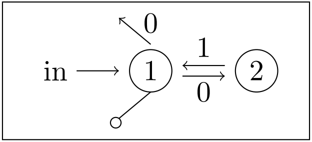


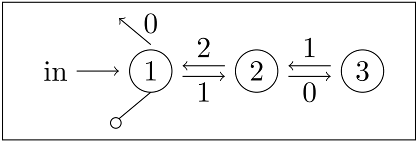
- **Thm 3**:
  - The singular-locus equation for the Cycle model is $k_{32}k_{43}\dots k_{n,n-1}k_{1,n}\prod_{2\leq i<j\leq n}(k_{i+1,i}-k_{j+1,j})$.
  - The singular-locus equation for the Mammillary model is $k_{12}k_{13}\dots k_{1,n}\prod_{2\leq i<j\leq n}(k_{1i}-k_{1j})^2$.

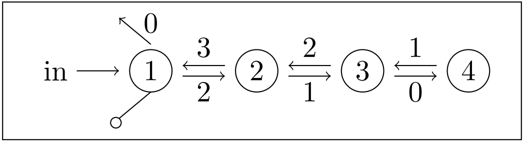
Catenary (path) models



Catenary (path) models

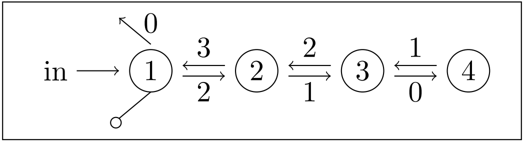


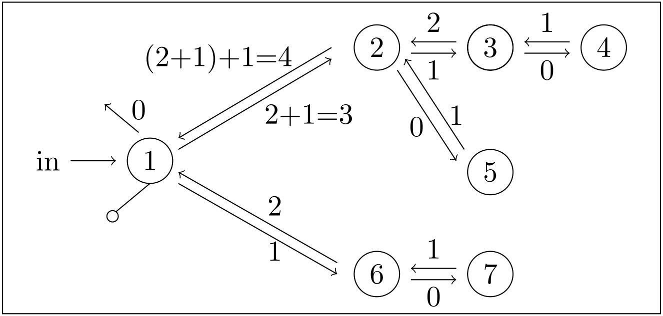




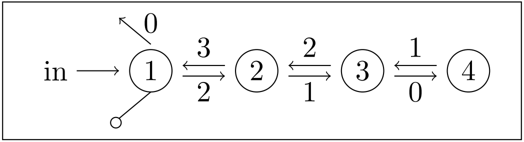
Conjecture: For **catenary models**, the exponents in the singular-locus equation generalize the pattern above.

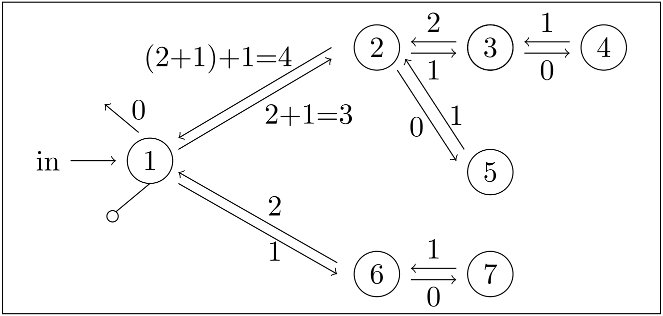
Tree conjecture





Tree conjecture





Conj.: (Hoch, Sweeney, Tung) For **tree models**, the exponents in the singular-locus equation generalize the pattern above.

Identifiable submodels (again)

- **Thm 4**: Let $\widetilde{\mathcal{M}}$ be obtained by:
  - adding a leak to a strongly connected model $\mathcal{M}$ with *no* leaks, or
  - deleting the leak from a strongly connected model $\mathcal{M}$ with input, output, and leak in *one* compartment.

  Then, if $\mathcal{M}$ is identifiable, then so is $\widetilde{\mathcal{M}}$.

[2] Can delete edges *without* making the singular-locus equation $= 0$.

Identifiable submodels (again)

- **Thm 4**: Let $\widetilde{\mathcal{M}}$ be obtained by:
  - adding a leak to a strongly connected model $\mathcal{M}$ with *no* leaks, or
  - deleting the leak from a strongly connected model $\mathcal{M}$ with input, output, and leak in *one* compartment.

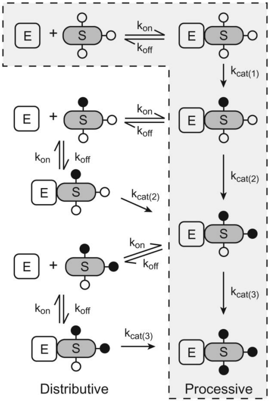
Then, if $\mathcal{M}$ is identifiable, then so is $\widetilde{\mathcal{M}}$.

Operation	**Preserves identifiability?**
Add input	Yes
Add output	Yes
Add leak	Not always (and see above)
Add edge	Not always
Delete input	Not always
Delete output	Not always
Delete leak	Open (and see above)
Delete edge	Not always (recall Thm 2[2])

[2] Can delete edges *without* making the singular-locus equation = 0.

Future work

Nonlinear models



From Processive phosphorylation: mechanism and biological importance, Patwardhan and Miller, *Cell Signal.* 2007.

Summary

The **singular locus** is an interesting mathematical object that can help us answer the question, *which linear compartment models are identifiable?*

Thank you.

Identifiability degree

- the **identifiability degree** of a model is the number of parameter vectors that match (generic) input-output data

Identifiability degree

- the **identifiability degree** of a model is the number of parameter vectors that match (generic) input-output data
- **Proposition** (Cobelli, Lepschy, Romanin Jacur 1979)

Model	**Identifiability degree**
Catenary (path)	1
Mammillary (star)	$(n-1)!$

- **Thm 5**

Model	**Identifiability degree**
Cycle	$(n-1)!$