Combinatorial and numerical methods for studying global dynamics in gene regulatory networks

Shane Kepley
VU Amsterdam

VU Mathematics Colloquium
October 20, 2021

Gene regulatory networks
Hill function models
Combinatorial dynamics
Returning to ODE models
Analysis of the Toggle Switch

Outline

Gene regulatory networks

- A gene regulatory network (GRN) is a collection of genes which mutually regulate their own gene expression levels or protein production.
- We represent a GRN as a directed graph. Each node represents a gene or protein and the edges between them describe (transcriptional) regulation.
- $A \rightarrow B$ ($A \dashv B$) means A activates (represses) B (i.e. more A increases (decreases) production rate of B).
- Examples:
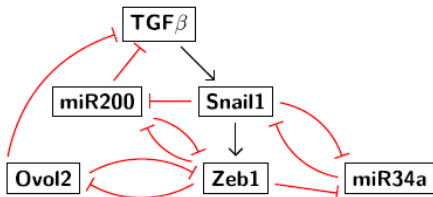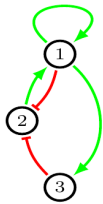
A more interesting example

- A modern GRN of "known" p53 interactions (G. Chen, M. Cairelli, et al, PLoS CompBio 2014).
- The apparent complexity of a typical GRN actually promotes robust dynamics.
- Central question: What dynamics are imposed by the topology of the GRN?
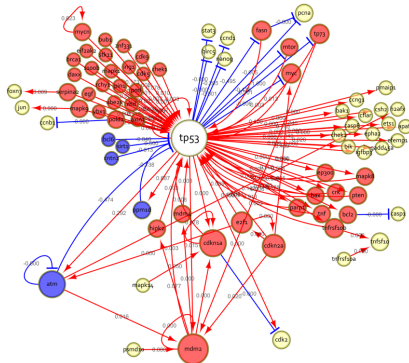
Main challenge

No first principles. How do we solve $\dot{x} = \boxed{?}$

1. Define a framework for transcribing GRNs into nonlinear ODEs with relatively simple response functions for each interaction.
2. These ODE models have very high parameter dimension.
3. Compute "combinatorial dynamics" to obtain a coarse but global description of the dynamics for all parameters.
4. Use the combinatorial dynamics to identify subsets of the parameter space with interesting dynamics.
5. Return to the ODE model: Apply numerical methods restricted to these parameter subsets to study dynamics classically (i.e. orbits, invariant sets, bifurcations, etc).

Gene regulatory networks
Hill function models
Combinatorial dynamics
Returning to ODE models
Analysis of the Toggle Switch

Outline

From networks to ODEs

- For a GRN with N genes we consider a nonlinear ODE:

$$\dot{x} = f(x, \lambda), \quad x \in X := [0, \infty)^N, \quad \lambda \in \Lambda := (0, \infty)^M.$$

- For $1 \leq i \leq N$, the variable x_i denotes the protein concentration or gene expression level of node i.

- Assume that each protein or gene degrades at a rate proportional to its own concentration/expression level, i.e.

$$\dot{x}_i = -\gamma_i x_i + \text{nonlinear/inhomogeneous production},$$

where the *decay rate* is $\gamma_i > 0$, which is unknown.

- For transcriptional interactions the biology suggests the functional form

$$\dot{x}_i = -\gamma_i x_i + \sum_{\substack{\text{interaction} \\ \text{groups}}} \prod_{\substack{\text{interactions} \\ \text{with } x_i}} g_{i,j}(x_j, \lambda).$$

Modeling the nonlinear production

Definition

A polynomial, $p \in \mathbb{R}[z_1, \ldots, z_N]$ is called an interaction function if it has the form

$$p = \prod_{m=1}^{q} p_m, \qquad \text{where} \quad p_m = \sum_{j \in I_m} z_j,$$

and $\{I_1, \ldots, I_q\}$ is an integer partition of $\{1, \ldots, N\}$. The linear polynomial p_m is called the mth summand of p.

- Informally, an interaction function is a "product of sums". Interactions with similar mechanisms are combined in the same summand.
- Interaction functions provide control over how the contributions from multiple regulators are combined.
- Example: $p = (z_1 + z_2)(z_3 + z_4)z_5$ where

$$q = 3, \; N = 5, \; I_1 = \{1, 2\}, \; I_2 = \{3, 4\}, \; I_3 = \{5\}.$$

Definition

Let $H^+ : [0, \infty) \to (0, \infty)$ denote the activating Hill function defined by the formula

$$H^+(x) := \ell + \delta \frac{x^d}{\theta^d + x^d}.$$

$H^- : [0, \infty) \to (0, \infty)$ denotes the repressing Hill function defined by the formula

$$H^-(x) := \ell + \delta \frac{\theta^d}{\theta^d + x^d}$$

- H^* depends on 4 positive real parameters, $\{\ell, \delta, \theta, d\}$.
- d is called the *Hill coefficient* (or Hill exponent).
- Aliases: Holling type II and type III (ecology), Michaelis-Menten or Monod ($d = 1$) functions.

- H^+ is increasing and H^- is decreasing.
- $H^+(0) = \ell = \lim_{x \to \infty} H^-(x)$.
- $H^-(0) = \ell + \delta = \lim_{x \to \infty} H^+(x)$.
- $H^*(\theta) = \ell + \frac{1}{2}\delta$ (half saturation).
- Increasing d produces a steeper response.
- H^* is hyperbolic when $1 \le d \le 2$ and sigmoidal for $d > 2$.

Hill ODE models
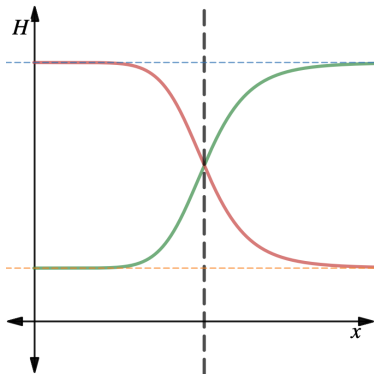
Definition

A Hill model is an ODE of the form

$$\dot{x} = f(x, \lambda) = -\Gamma x + \mathcal{H}(x, \lambda),$$

where

$$\Gamma := \begin{pmatrix} \gamma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \gamma_N \end{pmatrix}, \qquad \mathcal{H}_i(x) = p_i\big(H_{i,1}^*(x_1) \dots, H_{i,N}^*(x_N)\big),$$

p_i is an interaction function, and $H_{i,j}^*$ is a Hill function, $1 \leq i, j \leq N$.

- If there is a directed edge from node j to node i, then $H_{i,j}^*$ contributes 4 parameters to the model, $\{\ell_{i,j}, \delta_{i,j}, \theta_{i,j}, d_{i,j}\}$.
- If $j \dashv i$ then $H_{i,j}^* = H^-$ and if $j \rightarrow i$ then $H_{i,j}^* = H^+$.
- Parameter space dimension: $M = N + 4 \cdot \#E$.
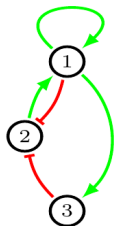
Example: the Toggle Switch

$$\dot{x}_1 = -\gamma_1 x_1 + \ell_{1,2} + \delta_{1,2} \frac{\theta_{1,2}^{d_{1,2}}}{\theta_{1,2}^{d_{1,2}} + x_2^{d_{1,2}}} = -\gamma_1 x_1 + H_{1,2}^-(x_2)$$

$$\dot{x}_2 = -\gamma_2 x_2 + \ell_{2,1} + \delta_{2,1} \frac{\theta_{2,1}^{d_{2,1}}}{\theta_{2,1}^{d_{2,1}} + x_1^{d_{2,1}}} = -\gamma_2 x_2 + H_{2,1}^-(x_1)$$

$x = (x_1, x_2) \in [0, \infty)^2.$

$\lambda = (\gamma_1, \ell_{1,2}, \delta_{1,2}, \theta_{1,2}, d_{1,2}, \gamma_2, \ell_{2,1}, \delta_{2,1}, \theta_{2,1}, d_{2,1}) \in (0, \infty)^{10}.$

A three node GRN example

$$f(x,\lambda) = -\underbrace{\begin{pmatrix} \gamma_1 & 0 & 0 \\ 0 & \gamma_2 & 0 \\ 0 & 0 & \gamma_3 \end{pmatrix}}_{\text{linear decay}} x + \underbrace{\begin{pmatrix} H_{1,1}^+(x_1) + H_{2,1}^+(x_2) \\ H_{1,2}^-(x_1) H_{3,2}^-(x_3) \\ H_{1,3}^+(x_1) \end{pmatrix}}_{\text{nonlinear production}}.$$

$x = (x_1, x_2, x_3) \in [0, \infty)^3.$

$\lambda = \big(\gamma_1, \ell_{1,1}, \delta_{1,1}, \theta_{1,1}, d_{1,1}, \ell_{2,1}, \delta_{2,1}, \theta_{2,1}, d_{2,1}, \gamma_2, \ell_{1,2}, \delta_{1,2}, \theta_{1,2}, d_{1,2},$
$\quad \ell_{3,2}, \delta_{3,2}, \theta_{3,2}, d_{3,2}, \gamma_3, \ell_{1,3}, \delta_{1,3}, \theta_{1,3}, d_{1,3}\big) \in (0, \infty)^{23}.$

Gene regulatory networks
Hill function models
Combinatorial dynamics
Returning to ODE models
Analysis of the Toggle Switch

Outline

- Main idea: When Λ is high dimensional or parameters are difficult to estimate it is impractical to study invariant sets.
- Give up (temporarily) on studying invariant sets (e.g. orbits, equilibria, invariant manifolds, connections, etc).
- Study robust dynamical descriptions (e.g. attractor/repeller pairs, isolating blocks, Morse decompositions, Conley indices, etc).
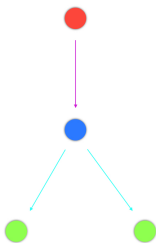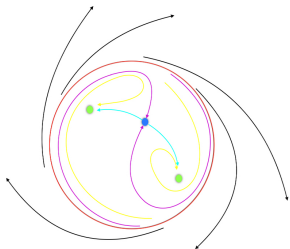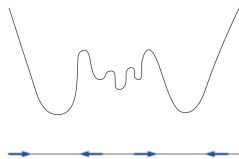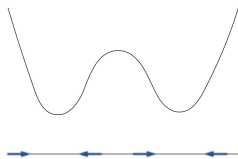
What we think we see.
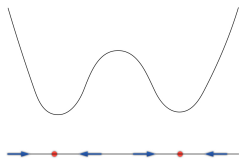
What we actually see.

Model coarseness and robustness

- Informally, *combinatorial dynamics* refers to broadly observable dynamic phenotypes which depend only on the topology of the GRN.

- Let x_I denote the vector of state variables which regulate x_i (i.e. source nodes for edges incoming to node i). We imagine a system of ODEs of the form

$$x_i = -\gamma_i x_i + f_i(x_I, \lambda)$$

where f_i is some unknown (monotone?) function.

- Combinatorial dynamics are coarse descriptions which are relatively insensitive to the choice of f_i or λ.

- These take the form of computable dynamical invariants for equivalence classes of models $[f_i]$ which are robust over large subsets of parameters.

- The dynamics computed are rigorous for all models in $[f_i]$.

Dynamic Signatures Generated by Regulatory Networks (DSGRN)

- DSGRN is a computational platform for computing global combinatorial dynamics for GRN models.
- DSGRN uses a *simple model* for which the following can be done explicitly, efficiently, and rigorously.
    1. Discretize X: Compute a coarse description (Morse graph) of the gradient-like dynamics on X for an arbitrary parameter.
    2. Discretize Λ: Decompose Λ into open, semi-algebraic subsets on which the Morse graph is constant.
- Both steps are highly nontrivial.
- The combinatorial description of the global dynamics is a collection of parameter regions which partition Λ, and a Morse graph associated to each region.
- The combinatorial dynamics computed using DSGRN are valid for infinite Hill coefficients (i.e. $d_{i,j} \to \infty$ for all i, j).

Discretizing state space

- Partition X into top cells with transverse faces.
- Build a graph representation of the dynamics (state transition graph) and compute its lattice of attractors.
- Attractor lattices are robust and their algebraic structure encodes topological (and therefore dynamical) properties. (W. Kalies, K. Mischaikow, R.C.A.M. Vandervorst).
- Compute the associated Morse graph which is a robust description of the coarse dynamical phenotype.
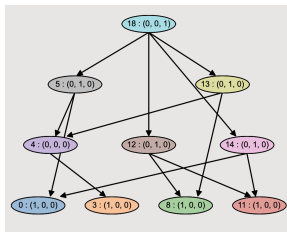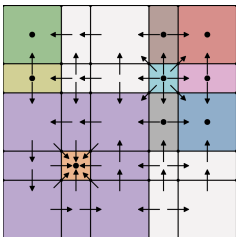
Descretizing parameter space

- A combinatorial parameter space decomposition (PSD) is a partition of Λ into subsets on which the Morse graph is constant.
- For the simple model these subsets are semi-algebraic sets given by explicit inequalities.
- Unfortunately, there are a lot of these subsets, most of them are empty, and there is no efficient algorithm for determining whether an arbitrary semi-algebraic set is nonempty.
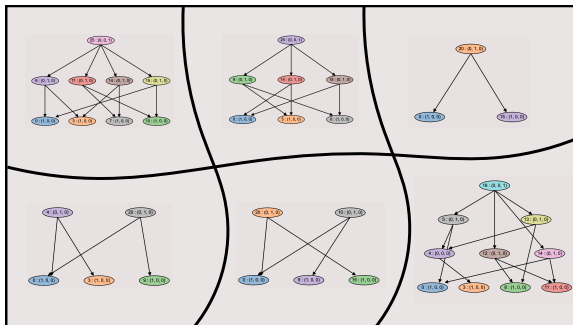
Connections to Algebraic Geometry and Order Theory

- Consider a semi-algebraic set $\Xi \subset \mathbb{R}^d$, and a collection of polynomials $\mathcal{P} := \{p_1, \ldots, p_K\} \subset \mathbb{R}[x_1, \ldots, x_d]$.

- Equip $\mathcal{P}$ with a partial order such that if $p \prec q$, then

$$p(\xi) < q(\xi) \quad \text{for all } \xi \in \Xi.$$

- A permutation $\sigma \in S_K$ defines a linear extension of this partial order, $\prec_\sigma$, satisfying

$$p_{\sigma(1)} \prec_\sigma p_{\sigma(2)} \prec_\sigma \cdots \prec_\sigma p_{\sigma(K)}.$$

- We define the *realizable set* associated to σ by

$$\Xi_\sigma := \{\xi \in \Xi : p_{\sigma(k)}(\xi) < p_{\sigma(k+1)}(\xi) \text{ for all } 1 \leq k \leq K - 1\}.$$

## The algebraically constrained linear extension problem (ACLEP)

Given $(\mathcal{P}, \prec, \Xi)$, rigorously compute the set of all permutations with nonempty realizable sets.

Relationship between ACLEP and PSD

- Computing a combinatorial PSD for the simple model is equivalent to a specific instance of the ACLEP.

- Surprisingly, these instances of the ACLEP can be solved "efficiently" (S.K., K. Mischaikow, L. Zhang, 2020 SIAGA).

- This problem only needs to be solved once for each distinct configuration of incoming edges to a GRN node and the results stored in a database.

- The current version of DSGRN has rigorous solutions stored for a node with up to 5 incoming (transcriptional) edges with any interaction function except $p = (z_1 + z_2 + z_3)(z_4 + z_5)$.

- We have also solved PSD problems for "non-transcriptional" regulation modeling ubiquitination and phosphorylation (B. Cummins, M. Gameiro, T. Gedeon, S.K., K. Mischaikow, 2021).

- Top 14 hysteretic switch designs out of 19,683 three node GRNs (M. Gameiro, T. Gedeon, S.K., K. Mischaikow, PLoS CompBio 2020)
- Synthetic Biology labs at UW and UC Santa Barbara are actually building these designs.
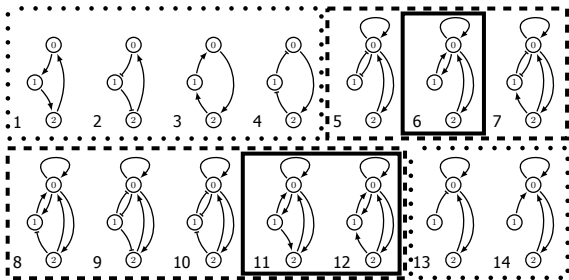
Gene regulatory networks
Hill function models
Combinatorial dynamics
Returning to ODE models
Analysis of the Toggle Switch

Outline

1. Gene regulatory networks

2. Hill function models

3. Combinatorial dynamics

4. Returning to ODE models

5. Analysis of the Toggle Switch

The HillCont library

- *HillCont*: An open source Python library for numerical analysis of Hill models (coauthored with Elena Queirolo).
- Implement Hill function ODE models given only GRN topology for arbitrary state/parameter dimension.
- Efficient evaluation of Hill models and the following derivatives

$$D_x f, D_\lambda f, D_{xx} f, D_{\lambda x} f, D_{\lambda\lambda} f, D_{xxx} f, D_{\lambda xx} f, D_{\lambda\lambda x} f.$$

via automatic differentiation (i.e. without formulas or calculus).

- Main features:
  1. Automatically search for equilibria, saddle-node bifurcations, Hopf bifurcations, and pitchfork bifurcations.
  2. Constrained parameter optimization and/or continuation along equilibria/bifurcations.
  3. Rigorous enclosures of equilibria for biologically relevant GRNs via the "bootstrap map".

The bootstrap map

Definition

A continuous function $g \colon [0, \infty)^N \to (0, \infty)^N$ has a monotone factorization if for each $i = 1, \ldots, N$, g_i factors as

$$g_i(x) = g_i^+(x) g_i^-(x) \quad \text{for all } x \in [0, \infty)^N,$$

where $g_i^+ \colon [0, \infty)^N \to (0, \infty)^N$ is bounded and strictly increasing and $g_i^- \colon [0, \infty)^N \to (0, \infty)^N$ is bounded and strictly decreasing as a function of each variable $x_1, \ldots, x_N$.

Definition

If $f(x) = -\Gamma x + \mathcal{H}(x)$ and $\mathcal{H}$ has a monotone factorization, then the bootstrap map for f is $\Phi \colon \mathbb{R}^{2N} \to \mathbb{R}^{2N}$ defined by the formula

$$\Phi_i(\alpha, \beta) = \frac{1}{\gamma_i} \mathcal{H}_i^+(\alpha) \mathcal{H}_i^-(\beta) \quad \text{and} \quad \Phi_{N+i}(\alpha, \beta) = \frac{1}{\gamma_i} \mathcal{H}_i^+(\beta) \mathcal{H}_i^-(\alpha),$$

where $\alpha, \beta \in \mathbb{R}^N$ and $i = 1, \ldots, N$.

Rigorous equilibrium enclosures

Theorem (S.K., K. Mischaikow, E. Queirolo)

Consider f and Φ as previously defined and assume that
$\liminf_{\|x\| \to \infty} g_i^-(x) > 0$ *for all $i = 1, \ldots, N$. Then, the following*
are true.

(i) $x \in [0, \infty)^N$ *is a zero of f if and only if $(x, x) \in [0, \infty^{2N})$ is a*
fixed point of Φ.

(ii) *Define $\left(\alpha^{(0)}, \beta^{(0)}\right) \in \mathbb{R}^{2N}$ coordinate wise by*

$$\alpha_i^{(0)} := \frac{1}{\gamma_i} g_i^+(0) \liminf_{\|x\| \to \infty} g_i^-(x), \quad \beta_i^{(0)} := \frac{1}{\gamma_i} \limsup_{\|x\| \to \infty} g_i^+(x) g_i^-(0)$$

for $i = 1, \ldots, N$. Iteratively, define $(\alpha^{n+1}, \beta^{n+1}) = \Phi(\alpha^n, \beta^n)$
for $n \geq 1$. Then $(\hat{\alpha}, \hat{\beta}) := \lim_{n \to \infty} (\alpha^n, \beta^n)$ exists and $(\hat{\alpha}, \hat{\beta})$ is a
fixed point of Φ.

(iii) *If $f(\hat{x}) = 0$, then*

$$\hat{x} \in \hat{R} := \prod_{i=1}^{N} [\hat{\alpha}_i, \hat{\beta}_i].$$

Gene regulatory networks
Hill function models
Combinatorial dynamics
Returning to ODE models
Analysis of the Toggle Switch

Outline

1. Gene regulatory networks

2. Hill function models

3. Combinatorial dynamics

4. Returning to ODE models
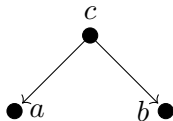
5. Analysis of the Toggle Switch

DSGRN analysis of the Toggle Switch

- $\Lambda = (0, \infty)^{10}$ is partitioned into 9 semi-algebraic subsets $\{R_1, \ldots, R_9\}$ with constant Morse graphs.

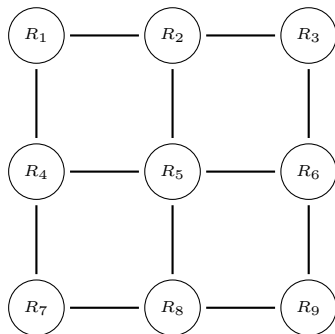- Example: R_5 is defined by the inequalities

$$\ell_{1,2} < \gamma_1 \theta_{2,1} < \ell_{1,2} + \delta_{1,2}$$
$$\ell_{2,1} < \gamma_2 \theta_{1,2} < \ell_{2,1} + \delta_{2,1}$$

and has associated Morse graph

- The remaining 8 regions are have Morse graphs with a unique minimal attractor of "FP" type.

Predicting ODE dynamics from combinatorial dynamics

- The Toggle Switch Hill model is

$$\dot{x} = f(x, \lambda) = \begin{pmatrix} -\gamma_1 & 0 \\ 0 & -\gamma_2 \end{pmatrix} x + \begin{pmatrix} H_{1,2}^-(x_2) \\ H_{2,1}^-(x_1) \end{pmatrix}$$

  with $\lambda = (\gamma_1, \ell_{1,2}, \delta_{1,2}, \theta_{1,2}, d_{1,2}, \gamma_2, \ell_{2,1}, \delta_{2,1}, \theta_{2,1}, d_{2,1})$.

- We identify both Hill coefficients (i.e. $d_{1,2} = d_{2,1} = d$).
- At $d = 1$, f has a unique globally attracting stable equilibrium.
- The DSGRN analysis holds in the limit $d \to \infty$.
- Assume f has no codimension-1 bifurcations other than saddle-node bifurcations.
- If the combinatorial dynamics were a perfect predictor of the ODE dynamics, then we expect to find saddle-node bifurcations when continuing parameters in R_5.
- Using *HillCont* we can quantify the correlation between the combinatorial and ODE dynamics with statistical analysis.

Constructive saddle-node bifurcation theorem

Theorem (J.B. van den Berg, M. Gameiro, J.P. Lessard, J.D. Mireles James, K. Mischaikow)

Let $g : X \times \mathbb{R} \to TX$ be a one parameter family of vector fields. Define $G : \mathbb{R}^{2N+1} \to \mathbb{R}^{2N+1}$ by the formula

$$G(x, v, s) := \begin{pmatrix} g(x, s) \\ D_x g(x, s) v \\ v^T v - 1 \end{pmatrix} \qquad x, v \in \mathbb{R}^N, s \in \mathbb{R}.$$

Suppose $\hat{u} = (\hat{x}, \hat{v}, \hat{s}) \in \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}$ satisfies
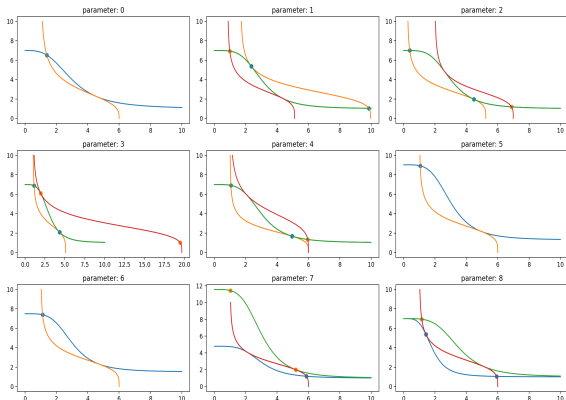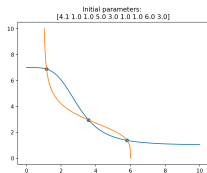
1. $G(\hat{u}) = 0$.

2. $D_u G(\hat{u})$ is an isomorphism.

3. *Every nonzero eigenvalue of $D_x g(\hat{x}, \hat{s})$ has nonzero real part.*

Then, g undergoes a saddle-node bifurcation at $(\hat{x}, \hat{s})$ and

$$\ker D_x g(\hat{x}, \hat{s}) = \operatorname{span}(\{\hat{v}\}).$$

Gene regulatory networks
Hill function models
Combinatorial dynamics
Returning to ODE models
Analysis of the Toggle Switch

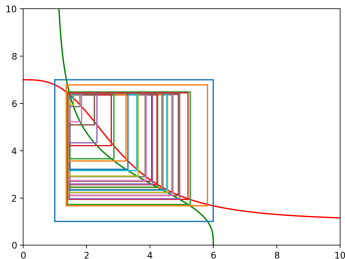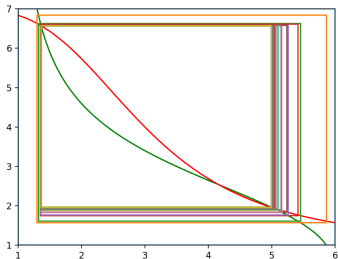# Searching for saddle-node bifurcations for $\lambda \in R_5$

Rigorous equilibria for the Toggle Switch

Theorem (S.K., K. Mischaikow, E. Queirolo)

Let $\Phi : \mathbb{R}^4 \to \mathbb{R}^4$ be the bootstrap map for Toggle Switch. Suppose the bootstrap iteration converges to $(\hat{\alpha}, \hat{\beta}) \in \mathbb{R}^2 \times \mathbb{R}^2$ and let $\hat{R} := [\hat{\alpha}_1, \hat{\beta}_1] \times [\hat{\alpha}_2, \hat{\beta}_2]$. Then, either $\hat{R}$ is degenerate and f is monostable with unique equilibrium $\hat{x} = (\hat{\alpha}_1, \hat{\alpha}_2)$, or f is bistable with stable equilibria at the corners of $\hat{R}$ with coordinates

$$\hat{x}_1 = (\hat{\alpha}_1, \hat{\beta}_2), \qquad \hat{x}_2 = (\hat{\beta}_1, \hat{\alpha}_2).$$

Statistical analysis of combinatorial predictions

χ^2 hypothesis testing

We test the null hypothesis that $\lambda \in R_5$ is independent of the existence of a saddle-node bifurcation along the parameterized path $\gamma : [1, \infty] \to \Lambda$ defined by

$$\gamma(s) = (\gamma_1, \ell_{1,2}, \delta_{1,2}, \theta_{1,2}, s, \gamma_2, \ell_{2,1}, \delta_{2,1}, \theta_{2,1}, s).$$

- Contingency table for 1,000 randomly chosen parameters:

	SNB found	No SNB found	Bisection fail	Bad parity
$\lambda \in R_5$	86	0	30	5
$\lambda \notin R_5$	2	875	0	2

- χ^2 test statistic: 968.19
- p-value: $p \approx 1.4 \times 10^{-209}$

Thank you for your attention!