Polynomial Superlevel Set Representation of the Multistationarity Region of Chemical Reaction Networks

AmirHosein Sadeghimanesh
(Joint work with Matthew England)

Coventry University

a.h.sadeghimanesh@gmail.com
ad6397@coventry.ac.uk

SYMBIONT - 2022.03.15

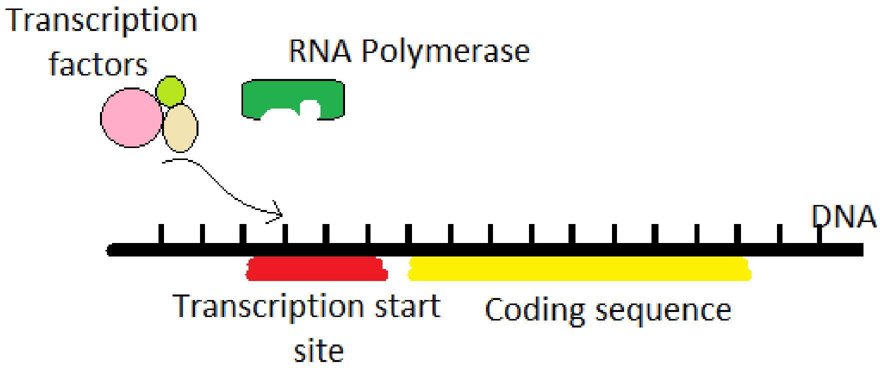
Authors supported by EPSRC grant EP/T015748/1 (The DEWCAD Project)

Introduction

- Reaction Network and Multistationarity.
- Representation of parameter regions.
- Superlevel sets.
- Superlevel set representation of the multistationarity region.
- Comparison with the other representations.
- More algorithms and discussions.

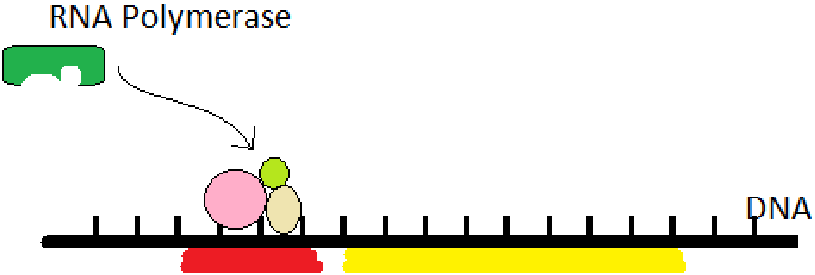
Gene regulatory networks

How does a gene get expressed?

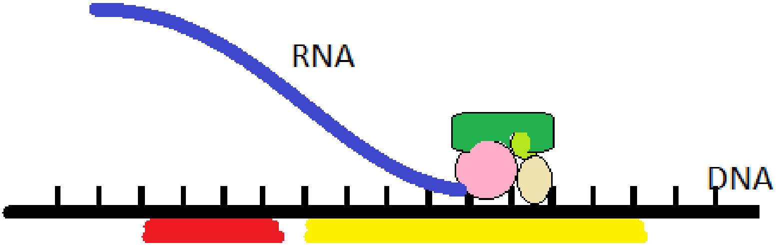


Gene regulatory networks

How does a gene get expressed?

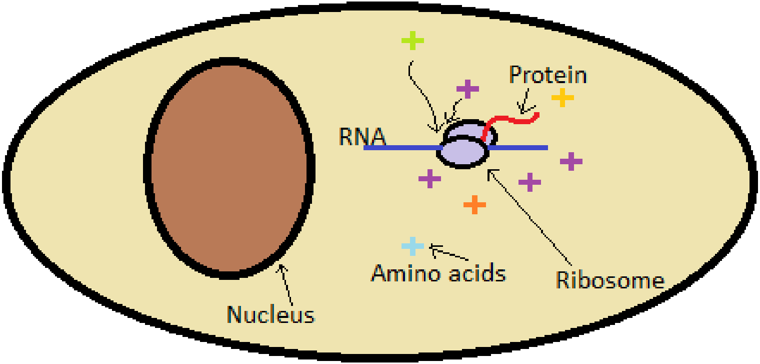


How does a gene get expressed?

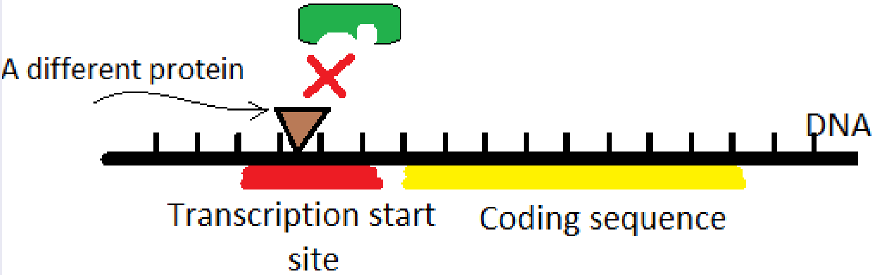


Gene regulatory networks

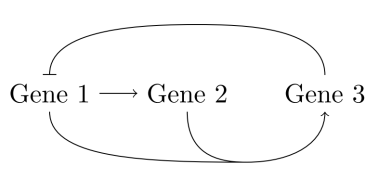
How does a gene get expressed?



How does a gene get expressed?

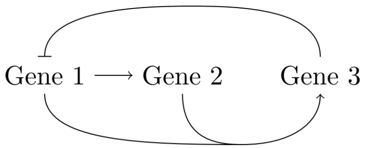


An example of gene regulatory network



$$\frac{d[A](t)}{dt} = k_{A,\max} \cdot \frac{1}{1 + k_{A,C}[C](t)} - k_{A,d}[A](t)$$
$$\frac{d[B](t)}{dt} = k_{B,\max} \cdot \frac{k_{B,A}[A](t)}{1 + k_{B,A}[A](t)} - k_{B,d}[B](t)$$
$$\frac{d[C](t)}{dt} = k_{C,\max} \cdot \frac{k_{C,A}[A](t)}{1 + k_{C,A}[A](t)} \cdot \frac{k_{C,B}[B](t)}{1 + k_{C,B}[B](t)} - k_{C,d}[C](t)$$

An example of gene regulatory network

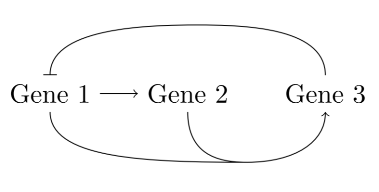


$$\dot{x_1} = k_1 \cdot \frac{1}{1 + k_2 x_3} - k_3 x_1$$

$$\dot{x_2} = k_4 \cdot \frac{k_5 x_1}{1 + k_5 x_1} - k_6 x_2$$

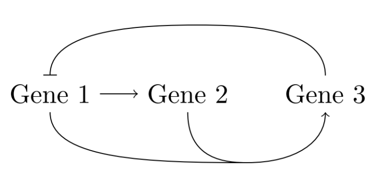
$$\dot{x_3} = k_7 \cdot \frac{k_8 x_1}{1 + k_8 x_1} \cdot \frac{k_9 x_2}{1 + k_9 x_2} - k_{10} x_3$$

An example of gene regulatory network



$$\begin{aligned}
k_1 \cdot \frac{1}{1 + k_2 x_3} - k_3 x_1 &= 0 \\
k_4 \cdot \frac{k_5 x_1}{1 + k_5 x_1} - k_6 x_2 &= 0 \\
k_7 \cdot \frac{k_8 x_1}{1 + k_8 x_1} \cdot \frac{k_9 x_2}{1 + k_9 x_2} - k_{10} x_3 &= 0
\end{aligned}$$

An example of gene regulatory network



$$
\begin{aligned}
k_1 \cdot \frac{1}{1 + k_2 x_3} - k_3 x_1 &= 0 \\
k_4 \cdot \frac{k_5 x_1}{1 + k_5 x_1} - k_6 x_2 &= 0 \\
k_7 \cdot \frac{k_8 x_1}{1 + k_8 x_1} \cdot \frac{k_9 x_2}{1 + k_9 x_2} - k_{10} x_3 &= 0
\end{aligned}
$$

This is an open network because of the degradation reactions.

An example of a closed network

$$2\,\mathrm{O_2}^- + 2\,\mathrm{H}^+ \xrightarrow{k} \mathrm{O_2} + \mathrm{H_2O_2}$$

$$\begin{aligned}
\dot{x}_1 &= -2kx_1^2x_2^2,\\
\dot{x}_2 &= -2kx_1^2x_2^2,\\
\dot{x}_3 &= kx_1^2x_2^2,\\
\dot{x}_4 &= kx_1^2x_2^2.
\end{aligned}$$

An example of a closed network

$$2\,\mathrm{O_2}^- + 2\,\mathrm{H}^+ \xrightarrow{k} \mathrm{O_2} + \mathrm{H_2O_2}$$

$$\begin{aligned}
\dot{x_1} &= -2kx_1^2x_2^2,\\
\dot{x_2} - \dot{x_1} &= 0,\\
\dot{x_1} + 2\dot{x_3} &= 0,\\
\dot{x_1} + 2\dot{x_4} &= 0.
\end{aligned}$$

An example of a closed network

$$2\,\mathrm{O_2}^- + 2\,\mathrm{H}^+ \xrightarrow{k} \mathrm{O_2} + \mathrm{H_2O_2}$$

$$\begin{aligned}\dot{x}_1 &= -2kx_1^2x_2^2,\\ x_2 - x_1 &= T_1,\\ x_1 + 2x_3 &= T_2,\\ x_1 + 2x_4 &= T_3.\end{aligned}$$

A closed network has at least one conservation law, for example since nothing is allowed to go out or come in, the total mass should be conserved.

An example of a closed network

$$2\,\mathrm{O_2}^- + 2\,\mathrm{H}^+ \xrightarrow{k} \mathrm{O_2} + \mathrm{H_2O_2}$$

$$\begin{aligned} \dot{x}_1 &= -2kx_1^2x_2^2, \\ x_2 + 2x_3 &= T_1, \\ x_1 + 2x_3 &= T_2, \\ x_1 + 2x_4 &= T_3. \end{aligned}$$

It is preferred to have positive coefficients in the conservation laws. Which is always possible. For example by summing the relation with a suitable scalar multiple of the previous mentioned trivial relation.

An example of a partially open network

$$X \xrightarrow{k_1} X + P$$

$$P \xrightarrow{k_2} 0$$

$$2P \underset{k_4}{\overset{k_3}{\rightleftharpoons}} PP$$

$$X + PP \underset{k_6}{\overset{k_5}{\rightleftharpoons}} XPP$$

$$XPP \xrightarrow{k_7} XPP + P$$

$$\begin{aligned}
\frac{dx_1}{dt} &= -k_5 x_1 x_3 + k_6 x_4, \\
\frac{dx_2}{dt} &= k_1 x_1 - k_2 x_2 - 2k_3 x_2^2 + 2k_4 x_3 + k_7 x_4, \\
\frac{dx_3}{dt} &= k_3 x_2^2 - k_4 x_3 - k_5 x_1 x_3 + k_6 x_4, \\
\frac{dx_4}{dt} &= k_5 x_1 x_3 - k_6 x_4.
\end{aligned}$$

An example of a partially open network

$$X \xrightarrow{k_1} X + P$$

$$P \xrightarrow{k_2} 0$$

$$2P \underset{k_4}{\overset{k_3}{\rightleftharpoons}} PP$$

$$X + PP \underset{k_6}{\overset{k_5}{\rightleftharpoons}} XPP$$

$$XPP \xrightarrow{k_7} XPP + P$$

$$\begin{aligned}
&-k_5x_1x_3 + k_6x_4 = 0,\\
&k_1x_1 - k_2x_2 - 2k_3x_2^2 + 2k_4x_3 + k_7x_4 = 0,\\
&k_3x_2^2 - k_4x_3 - k_5x_1x_3 + k_6x_4 = 0,\\
&x_1 + x_4 = k_8.
\end{aligned}$$

Multistationarity

Definition

Consider a network with n species. Replace redundant steady state equations by conservation laws if there exist any. Let k stands for the vector of constants of both the reaction rates and conservation laws and be of the size r. A network is called *multistationary* over $B \subseteq \mathbb{R}^r$ if there exists a $k \in B$ such that $f_k(x) = 0$ has more than one solution in $\mathbb{R}^n_{>0}$.

Definition

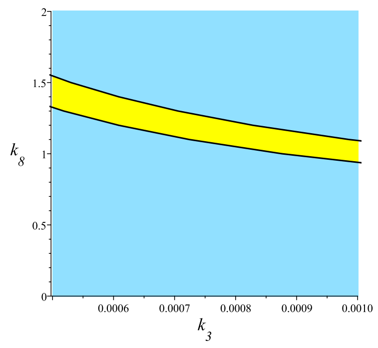
Consider a network with n species. Replace redundant steady state equations by conservation laws if there exist any. Let k stands for the vector of constants of both the reaction rates and conservation laws and be of the size r. A network is called *multistationary* over $B \subseteq \mathbb{R}^r$ if there exists a $k \in B$ such that $f_k(x) = 0$ has more than one solution in $\mathbb{R}^n_{>0}$.

The set $\{k \in B \mid \#\big(f_k^{-1}(0) \cap \mathbb{R}^n_{>0}\big) \geq 2\}$ is called the **multistationarity region** of the network.

Finding, representing and storing mult. region

CAD representation

How to find? Use CAD of the parameter space with respect to the discriminant variety.
How to represent? As a semialgebraic set.
How to store? Saving a list of CAD cells.



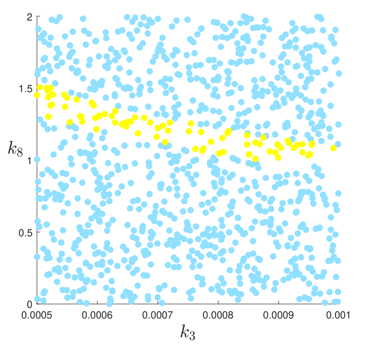
Finding, representing and storing mult. region

Sampling representation

How to find? Solve the system at many sample points.

How to represent? With a finite subset of it.

How to store? Saving a list of points.

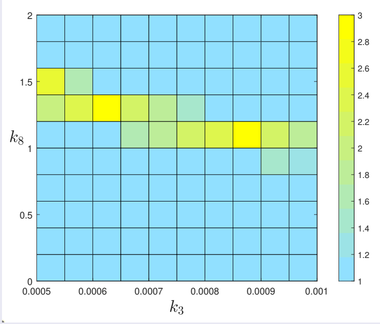


Rectangular representation

How to find? Partition the region to smaller regions. Solve the system for many sample points from the subregions and associate the average number to each subregion.

How to represent? With a union of hyperrectangles.

How to store? Saving a list of hyperrrectangles.



Finding, representing and storing mult. region

Rectangular representation

Remark: One can get the rectangular representation via Kac-Rice integrals. See reference no. 3.

Superlevel sets

Definition

Consider an arbitrary function $f\colon \mathbb{R}^n \to \mathbb{R}$. For a given $u \in \mathbb{R}$ a *superlevel set* of f is the set of the form

$$U_u(f) = \{x \in \mathbb{R}^n \mid f(x) \geq u\}.$$

When $u = 1$ we drop the index and only write $U(f)$. Naturally, a *polynomial superlevel set* is a superlevel set of a polynomial.

Superlevel sets

Theorem [reference no. 2]

Let $B \subseteq \mathbb{R}^n$ be a compact set and K a closed subset of B. For $d \in \mathbb{N}$ define

$$S_d = \{p \in P_d \mid p \geq 0 \text{ on } B,\ p \geq 1 \text{ on } K\}.$$

Then there exists a polynomial $p_d \in S_d$ such that

$$\int_B p_d(x)dx = \inf \Big\{ \int_B p(x)dx \mid p \in S_d \Big\}.$$

Furthermore $\lim_{d\to\infty} \mathrm{Vol}(U(p_d) - K) = 0$.

How to compute a superlevel set

If $p(x) = \sum_{\alpha \in \mathbb{N}_d^n} c_\alpha x^\alpha$ where $\mathbb{N}_d^n$ is the set of $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}_{\geq 0}^n$ such that $\sum_{i=1}^n \alpha_i \leq d$, then

$$\int_B p(x)dx = \int_B \Big(\sum_{\alpha \in \mathbb{N}_d^n} c_\alpha x^\alpha \Big) dx = \sum_{\alpha \in \mathbb{N}_d^n} c_\alpha \int_B x^\alpha dx = \sum_{\alpha \in \mathbb{N}_d^n} \Big(\int_B x^\alpha dx \Big) c_\alpha.$$

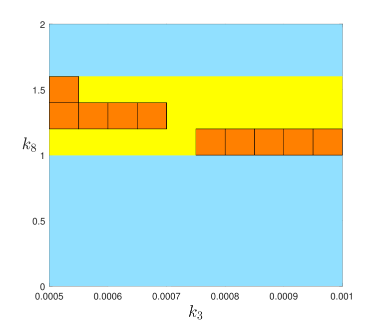
Let B be a hyperrectangle defined by $\prod_{j=1}^n [a_j, b_j]$. Then $p(x)$ being positive on B can be guaranteed by finding sum of squares polynomials s_0, s_1, ..., s_n such that

$$p(x) - \sum_{j=1}^n s_j(x)(x_j - a_j)(b_j - x_j) = s_0(x).$$

How to compute a superlevel set

Implementation

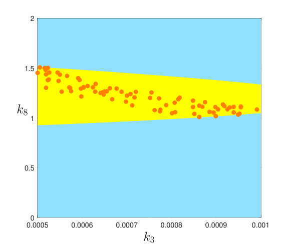
A PSS (Polynomial Superlevel Set) representation can be obtained by solving a minimization problem with linear target function and SOS (Sum of Squares polynomial) constraints using `YALMIP` + `SeDuMi` packages of `MatLab`.

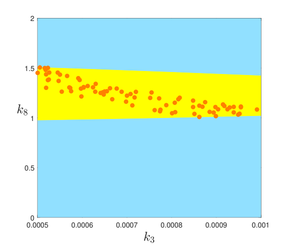


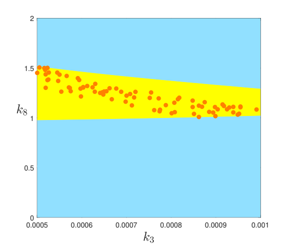
How to compute a superlevel set

Implementation

Instead of a rectangular representation, one can use a sampling representation and convert the SOS constraints to linear constraints or a mixture of the two.

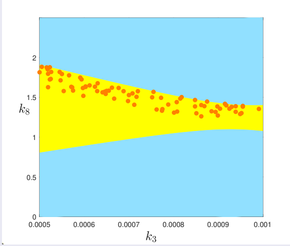


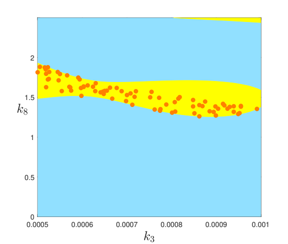


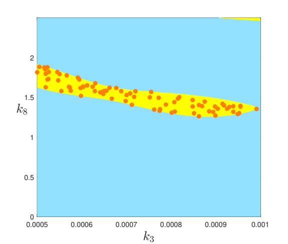


Be careful when using YALMIP

YALMIP usually finishes the computations with a message '*Numerical problems (SeDuMi)*'. One way to help YALMIP is to rescale the problem to the unit cube. Compare the following results with the previous ones.







Finding, representing and storing mult. region

Pros and cons of each method

- CAD has high complexity, only applicable for very small examples.
- Hard to infer geometry of the mult. region from sampling and rectangular representations for dimensions higher than 3.
- Saving a single polynomial may take less space than a large set of points or large set of hyperrectangles and may be easier to use for membership test of a point to the region, or finding distance of a point from the boundary of the region.

Using bisection approaches

Do we need to have all hyperrectangles in a rectangular representation to be of the same size?

Using bisection approaches

Consider the following example. This system has 1 or 3 positive solutions for a generic choice of a parameter point.

$$
\begin{aligned}
X_1 \xrightarrow{k_1} X_2 &\xrightarrow{k_2} X_3 \xrightarrow{k_3} X_4 \\
X_3 + X_5 &\xrightarrow{k_4} X_1 + X_6 \\
X_4 + X_5 &\xrightarrow{k_5} X_2 + X_6 \\
X_6 &\xrightarrow{k_6} X_5
\end{aligned}
$$

$$
\begin{aligned}
&k_4x_3x_5 - k_1x_1 = 0 \\
&k_5x_4x_5 + k_1x_1 - k_2x_2 = 0 \\
&-k_4x_3x_5 + k_2x_2 - k_3x_3 = 0 \\
&-k_4x_3x_5 - k_5x_4x_5 + k_6x_6 = 0 \\
&x_1 + x_2 + x_3 + x_4 - k_7 = 0 \\
&x_5 + x_6 - k_8 = 0
\end{aligned}
$$

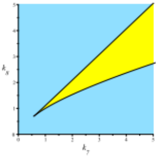
Using bisection approaches

$$
\begin{aligned}
X_1 \xrightarrow{k_1} X_2 \xrightarrow{k_2} X_3 \xrightarrow{k_3} X_4 \\
X_3 + X_5 \xrightarrow{k_4} X_1 + X_6 \\
X_4 + X_5 \xrightarrow{k_5} X_2 + X_6 \\
X_6 \xrightarrow{k_6} X_5
\end{aligned}
$$

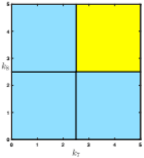
(a)

$$
\begin{aligned}
k_4 x_3 x_5 - k_1 x_1 &= 0 \\
k_5 x_4 x_5 + k_1 x_1 - k_2 x_2 &= 0 \\
-k_4 x_3 x_5 + k_2 x_2 - k_3 x_3 &= 0 \\
-k_4 x_3 x_5 - k_5 x_4 x_5 + k_6 x_6 &= 0 \\
x_1 + x_2 + x_3 + x_4 - k_7 &= 0 \\
x_5 + x_6 - k_8 &= 0
\end{aligned}
$$

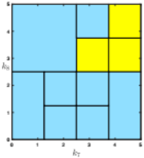
(b)



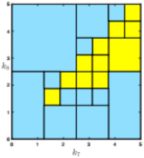
(c)



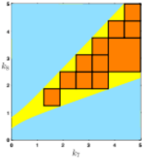
(d)



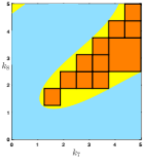
(e)



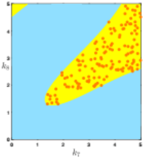
(f)



(g)



(h)



(i)

Cases with more than 2 generic possibilities

Lemma

Let $B \subseteq \mathbb{R}^r$ be a hyperrectangle and $g \colon B \to \{n_1, \ldots, n_s\} \subseteq \mathbb{Z}_{\geq 0}$. Assume that $\mathbb{E}\big(g(k) \mid k \sim U(B)\big) = n_i$ for some $i \in \{1, \ldots, s\}$. Then with probability one we have that B is almost subset of $L_{n_i}(g)$ if and only if $\mathbb{E}\big(g(k) \mid k \sim q\big) = n_i$ for a randomly chosen distribution q on B with the same zero measure sets as Lebesgue measure's.

Cases with more than 2 generic possibilities

proof

Nota that;

- $$\mathbb{E}\big(g(k) \mid k \sim U(B)\big) = \sum_{i=1}^{s} n_i \frac{\mathrm{Vol}\big(B \cap L_{n_i}(g)\big)}{\mathrm{Vol}(B)},$$

- $$\mathrm{Vol}\big(B \cap L_{n_i}(g)\big) = 0 \Longleftrightarrow \forall q : \int_{B \cap L_{n_i}(g)} q(x)dx = 0.$$

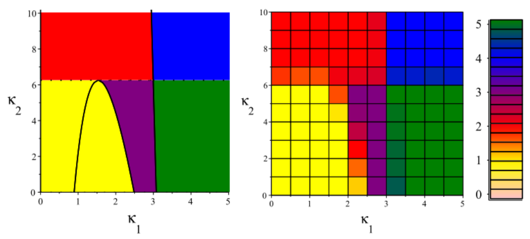
Now consider the two following sets;

$$T_1 = \{(x_1, \ldots, x_t) \in (0,1)^t \mid x_1 + \cdots + x_t = 1\},$$
$$T_2 = \{(x_1, \ldots, x_t) \in (0,1)^t \mid x_1 + \cdots + x_t = 1,\ n_{\alpha_1}x_1 + \cdots + n_{\alpha_t}x_t = n_i\},$$

where t is the number of non-zero volume regions.

Trying Lemma on an example

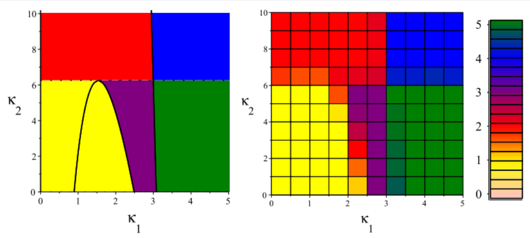
Example of section 2.3 of reference no. 3



The example is related to a parametric system of polynomial equations with 1 variable and two parameters. The possible number of positive real solutions to the system are 0, 1, 2, 3, 4 and 5. The average number of solutions on the rectangle $[2, 2.5] \times [2, 2.5]$ when parameters are coming from uniform distribution is 2. But if this rectangle is not inside the parameter region with 2 solutions. It has almost equal intersection with the regions with 1 and 3 solutions.

Trying Lemma on an example

Example of section 2.3 of reference no. 3



Using MCKR app, the expected number of solutions when parameters are equipped with truncated normal distribution with mean 2.25, and variance 0.1 is 1.8 and hence by Lemma, the fact that this rectangle is not inside the region with 2 solutions can be observed.

References

1. AmirHosein Sadeghimanesh, Matthew England, *Polynomial Superlevel Set Representation of the Multistationarity Region of Chemical Reaction Networks*, submitted, 2022.
2. Fabrizio Dabbene, Didier Henrion, Constantino Lagoa *Simple Approximations of Semialgebraic Sets and their Applications to Control*. Automatica, DOI: 10.1016/j.automatica.2016.11.021, 2017.
3. Elisenda Feliu, AmirHosein Sadeghimanesh, *Kac-Rice formulas and the number of solutions of parametrized system of polynomial equations*, arXiv: 2010.00804, 2020.

Thank you for listening.