Symbolic computation to determine parameter regions for multistationarity in models of the MAPK network

Matthew England - Coventry University

SYMBIONT Meeting
Bonn, Germany 23–23 March 2018

Partially supported by EU H2020 project SC² (712689).
Outline

1. Introduction
   - MAPK Network
   - Symbolic Methods

2. Results
   - Semi-algebraic descriptions of multistationarity region
   - Symbolic vs Numeric Grid Sampling
Outline

1. Introduction
   - MAPK Network
   - Symbolic Methods

2. Results
   - Semi-algebraic descriptions of multistationarity region
   - Symbolic vs Numeric Grid Sampling
Overview

- We aim to identify regions of parameter space with multi-stationarity (multiple steady states) of a biological network.

- Specifically, we consider the Mitogen-Activated Protein Kinases (MAPK) cascade. We have results for two models (#26 and #28 in the Biomodles Database\(^1\)).

- When the chemical reactions are modelled by mass action kinetics, then mathematically the task is to identify positive real solutions of a parametrised system of polynomials.

- In contrast to most of the literature on the topic, we work with methods from Symbolic Computation (where values are exact rather than floating point).

\(^1\)http://www.ebi.ac.uk/biomodels-main/
Motivation

Why multistationarity?

- Instrumental to cellular memory and cell differentiation during development or regeneration of multicellular organisms.
- Used by micro organisms in survival strategies.

Why symbolic methods?

- Numerical methods observed to give incorrect results at certain points in parameter space.
- Symbolic methods have the scope to give semi-algebraic descriptions of parameter space: the exact solution.
Case Study: Model 26

From: www.ebi.ac.uk/biомodels-main/BIOMD0000000026

\[
\begin{align*}
\dot{x}_1 &= k_2 x_6 + k_{15} x_{11} - k_1 x_1 x_4 - k_{16} x_1 x_5 \\
\dot{x}_2 &= k_3 x_6 + k_5 x_7 + k_{10} x_9 + k_{13} x_{10} - x_2 x_5 (k_{11} + k_{12}) - k_4 x_2 x_4 \\
\dot{x}_3 &= k_6 x_7 + k_8 x_8 - k_7 x_3 x_5 \\
\dot{x}_4 &= x_6 (k_2 + k_3) + x_7 (k_5 + k_6) - k_1 x_1 x_4 - k_4 x_2 x_4 \\
\dot{x}_5 &= k_8 x_8 + k_{10} x_9 + k_{13} x_{10} + k_{15} x_{11} - \\
&\quad x_2 x_5 (k_{11} + k_{12}) - k_7 x_3 x_5 - k_{16} x_1 x_5 \\
\dot{x}_6 &= k_1 x_1 x_4 - x_6 (k_2 + k_3) \\
\dot{x}_7 &= k_4 x_2 x_4 - x_7 (k_5 + k_6) \\
\dot{x}_8 &= k_7 x_3 x_5 - x_8 (k_8 + k_9) \\
\dot{x}_9 &= k_9 x_8 - k_{10} x_9 + k_{11} x_2 x_5 \\
\dot{x}_{10} &= k_{12} x_2 x_5 - x_{10} (k_{13} + k_{14}) \\
\dot{x}_{11} &= k_{14} x_{10} - k_{15} x_{11} + k_{16} x_1 x_5
\end{align*}
\]

11 differential equations

11 variables

16 parameters
Rate Constants

The biomodels database also gives us meaningful values for the rate constants.

- Some are measured accurately:

\[ k_1 = 0.02, \quad k_3 = 0.01, \quad k_4 = 0.032, \]
\[ k_7 = 0.045, \quad k_9 = 0.092, \quad k_{11} = 0.01, \]
\[ k_{12} = 0.01, \quad k_{15} = 0.086, \quad k_{16} = 0.0011. \]

- Others are estimated with confidence:

\[ k_2 = 1, \quad k_5 = 1, \quad k_6 = 15, \quad k_8 = 1, \]
\[ k_{10} = 1, \quad k_{13} = 1, \quad k_{14} = 0.5. \]
Linear Conservation Constraints

Three further Linear Conservation Constraints may be derived, introducing three further constant parameters.

\[ x_5 + x_8 + x_9 + x_{10} + x_{11} = k_{17} \]
\[ x_4 + x_6 + x_7 = k_{18} \]
\[ x_1 + x_2 + x_3 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} = k_{19} \]

We work with some realistic values for these new parameters:

\[ k_{17} = 100, \quad k_{18} = 50, \quad k_{19} \in \{200, 500\}. \]

However, the confidence in these estimates is not as high as the others. Ideally we would treat all three of these as symbolic parameters.
To identify regions of multistationarity it suffices to count real solutions of an integer polynomial system:

- Replacing the left hand sides of Model 26 by 0;
- Supplementing with the linear conservation constraints;
- Substituting for values of parameters (all but $k_{17}, k_{18}, k_{19}$ ideally);
- Converting to rationals and multiplying up to integers.
-Appending positivity constraints on all variables and free parameters.
Algebraic System of Interest II

\[0 = -200x_1x_4 - 11x_1x_5 + 860x_{11} + 10000x_6,\]
\[0 = -16x_2x_4 - 10x_2x_5 + 500x_{10} + 5x_6 + 500x_7 + 500x_9,\]
\[0 = -9x_3x_5 + 3000x_7 + 200x_8,\]
\[0 = -10x_1x_4 - 16x_2x_4 + 505x_6 + 8000x_7,\]
\[0 = -11x_1x_5 - 200x_2x_5 - 450x_3x_5 + 10000x_{10} + 860x_{11} + 10000x_8 + 10000x_9,\]
\[0 = 2x_1x_4 - 101x_6,\]
\[0 = 4x_2x_4 - 2000x_7,\]
\[0 = 45x_3x_5 - 1092x_8,\]
\[0 = 5x_2x_5 + 46x_8 - 500x_9,\]
\[0 = x_2x_5 - 150x_{10},\]
\[0 = 11x_1x_5 + 5000x_{10} - 860x_{11},\]
\[0 = -k_{17} + x_{10} + x_{11} + x_5 + x_8 + x_9,\]
\[0 = -k_{18} + x_4 + x_6 + x_7,\]
\[0 = -k_{19} + x_1 + x_{10} + x_{11} + x_2 + x_3 + x_6 + x_7 + x_8 + x_9,\]
\[0 < x_1, \ldots, 0 < x_{11}, 0 < k_{17}, 0 < k_{18}, 0 < k_{19}.\]
Algebraic System of Interest II

\[
\begin{align*}
0 &= -200x_1x_4 - 11x_1x_5 + 860x_{11} + 10000x_6, \\
0 &= -16x_2x_4 - 10x_2x_5 + 500x_{10} + 5x_6 + 500x_7 + 500x_9, \\
0 &= -9x_3x_5 + 3000x_7 + 200x_8, \\
0 &= -10x_1x_4 - 16x_2x_4 + 505x_6 + 8000x_7, \\
0 &= -11x_1x_5 - 200x_2x_5 - 450x_3x_5 + 10000x_{10} + 860x_{11} + 10000x_8 + 10000x_9, \\
0 &= 2x_1x_4 - 101x_6, \\
0 &= 4x_2x_4 - 2000x_7, \\
0 &= 45x_3x_5 - 1092x_8, \\
0 &= 5x_2x_5 + 46x_8 - 500x_9, \\
0 &= x_2x_5 - 150x_{10}, \\
0 &= 11x_1x_5 + 5000x_{10} - 860x_{11}, \\
0 &= -k_{17} + x_{10} + x_{11} + x_5 + x_8 + x_9, \\
0 &= -k_{18} + x_4 + x_6 + x_7, \\
0 &= -k_{19} + x_1 + x_{10} + x_{11} + x_2 + x_3 + x_6 + x_7 + x_8 + x_9, \\
0 &< x_1, \ldots, 0 < x_{11}, 0 < k_{17}, 0 < k_{18}, 0 < k_{19}.
\end{align*}
\]

14 polynomial equations
14 positivity conditions
11 variables
3 parameters
14 symbolic indeterminates

Matthew England
Symbolic computation for models of the MAPK network
Outline

1. Introduction
   - MAPK Network
   - Symbolic Methods

2. Results
   - Semi-algebraic descriptions of multistationarity region
   - Symbolic vs Numeric Grid Sampling
What symbolic methods do we use?

Tools designed for studying real solutions of polynomial systems (i.e. including inequalities and inequations - not just ideals).


- **Virtual Substitution** (VS). Invented by Weispfenning in the 1980s. Leading implementation in **Redlog**.

- **Lazy Real Triangularize** (LRT). Recent work by Chen et al. Implemented in the **RegularChains** Library for **Maple**.

CAD is necessary, and theoretically sufficient to solve the problem, but used alone is computationally infeasible. We found success when combining with either VS / LRT and pre-processing input.
Recent work by speaker and co-authors

  *A Case Study on the Parametric Occurrence of Multiple Steady States.*

  *Symbolic Versus Numerical Computation and Visualization of Parameter Regions for Multistationarity of Biological Networks.*
  Pre-processing; 3d symbolic grid sampling for Models 26+28.

  *Identifying the Parametric Occurrence of Multiple Steady States for some Biological Networks.* Submitted to Journal, 2017.
What is a CAD?

A CAD is:

- **decomposition** meaning a partition of $\mathbb{R}^n$ into connected subsets called *cells*;
- **(semi)-algebraic** meaning that each cell can be defined by a sequence of polynomial equations and inequalities.
- **cylindrical** meaning the cells are arranged in a useful manner - their projections (relative to a given variable ordering) are either equal or disjoint.

Produced from a set of polynomials so each has constant sign (positive, zero or negative) in each cell (thus truth of overall system also constant).

What is a CAD?

A CAD is:

- a decomposition meaning a partition of $\mathbb{R}^n$ into connected subsets called cells;
- (semi)-algebraic meaning that each cell can be defined by a sequence of polynomial equations and inequalities.
- cylindrical meaning the cells are arranged in a useful manner - their projections (relative to a given variable ordering) are either equal or disjoint.

Produced from a set of polynomials so each has constant sign (positive, zero or negative) in each cell (thus truth of overall system also constant).

E.g. A CAD produced for the polynomial $f := x^2 + y^2 - 1$ will commonly have 13 cells. Note how:

- cells are stacked in cylinders over the same portions of the x-axis to give **cylindricity**;
- the cells break over the polynomial graph to give **sign-invariance**.
E.g. A CAD produced for the polynomial \( f := x^2 + y^2 - 1 \) will commonly have 13 cells. Note how:

- cells are stacked in cylinders over the same portions of the \( x \)-axis to give **cylindricity**;
- the cells break over the polynomial graph to give **sign-invariance**.
E.g. A CAD produced for the polynomial $f := x^2 + y^2 - 1$ will commonly have 13 cells. Note how:

- cells are stacked in cylinders over the same portions of the $x$-axis to give cylindricity;
- the cells break over the polynomial graph to give sign-invariance.
E.g. A CAD produced for the polynomial
\[ f := x^2 + y^2 - 1 \]
will commonly have 13 cells. Note how:

- cells are stacked in cylinders over the same portions of the \( x \)-axis to give cylindricity;
- the cells break over the polynomial graph to give sign-invariance.
CAD: 2d Example

E.g. A CAD produced for the polynomial \( f := x^2 + y^2 - 1 \) will commonly have 13 cells. Note how:

- cells are stacked in cylinders over the same portions of the \( x \)-axis to give cylin
dricity;
- the cells break over the polynomial graph to give sign-invariance.
E.g. A CAD produced for the polynomial \( f := x^2 + y^2 - 1 \) will commonly have 13 cells. Note how:

- cells are stacked in cylinders over the same portions of the \( x \)-axis to give **cylindricity**;
- the cells break over the polynomial graph to give **sign-invariance**.
E.g. A CAD produced for the polynomial 
\( f := x^2 + y^2 - 1 \) will commonly have 13 cells. Note how:

- cells are stacked in cylinders over the same portions of the \( x \)-axis to give cylindricity;
- the cells break over the polynomial graph to give sign-invariance.
CAD: 2d Example

E.g. A CAD produced for the polynomial $f := x^2 + y^2 - 1$ will commonly have 13 cells. Note how:

- cells are stacked in cylinders over the same portions of the $x$-axis to give cylindricity;
- the cells break over the polynomial graph to give sign-invariance.
E.g. A CAD produced for the polynomial $f := x^2 + y^2 - 1$ will commonly have 13 cells. Note how:

- cells are stacked in cylinders over the same portions of the $x$-axis to give cylindricity;
- the cells break over the polynomial graph to give sign-invariance.
CAD: 2d Example

E.g. A CAD produced for the polynomial $f := x^2 + y^2 - 1$ will commonly have 13 cells. Note how:

- cells are stacked in cylinders over the same portions of the $x$-axis to give **cylindricity**;
- the cells break over the polynomial graph to give **sign-invariance**.
What is a Real Triangularization?

A **Real Triangularization** is a decomposition of a polynomial system into finitely many regular semi-algebraic systems. These are the real counterparts of the well studied *regular chains*. Such decompositions are always possible. Key features shown in next example. Details here:


The paper described an algorithm to produce them, and a **Lazy** variant which produced the highest dimension component and unevaluated function calls which if evaluated and output appended would give the full solution.
Consider the generic equation of degree two.

```
> R := PolynomialRing( [x, c, b, a] );
> sys := [ a x^2 + b x + c = 0 ]
```

Compute a triangular decomposition of the 4-variable hypersurface it defines.

```
> dec := RealTriangularize(sys, R) : Display(dec, R);
```

\[
\begin{cases}
  a x^2 + b x + c = 0 \\
  -4 a c + b^2 > 0 \text{ and } a \neq 0 \\
\end{cases}
\begin{cases}
  2 a x + b = 0 \\
  4 a c - b^2 = 0 \\
  a \neq 0 \\
\end{cases}
\begin{cases}
  b x + c = 0 \\
  a = 0 \\
  b \neq 0 \\
\end{cases}
\begin{cases}
  c = 0 \\
  b = 0 \\
  a = 0 \\
\end{cases}
\]
Consider the generic equation of degree two.

\[
\begin{align*}
& > R := \text{PolynomialRing}([x, c, b, a]); \\
& \quad \text{sys} := \left[ a x^2 + b x + c = 0 \right] \\
& \quad R := \text{polynomial\_ring} \\
& \quad \text{sys} := \left[ a x^2 + b x + c = 0 \right]
\end{align*}
\]

Compute a triangular decomposition of the 4-variable hypersurface it defines.

\[
> \text{dec} := \text{RealTriangularize}(\text{sys}, R) : \text{Display}(\text{dec}, R);
\]

\[
\begin{align*}
& \left\{ \\
& \quad a x^2 + b x + c = 0 \\
& \quad -4 a c + b^2 > 0 \text{ and } a \neq 0 \\
& \right\}, \\
& \quad \left[ \begin{array}{c}
2 a x + b = 0 \\
4 a c - b^2 = 0 \\
a \neq 0
\end{array} \right], \\
& \quad \left[ \begin{array}{c}
b x + c = 0 \\
a = 0 \\
b \neq 0
\end{array} \right], \\
& \quad \left[ \begin{array}{c}
c = 0 \\
b = 0 \\
a = 0
\end{array} \right]
\end{align*}
\]

\[
> \text{LazyRealTriangularize}(\text{sys}, R, \text{output = piecewise})
\]

\[
\begin{align*}
& \left\{ \left[ a x^2 + b x + c = 0 \right] \right\}, \\
& \quad 0 < -4 a c + b^2 \text{ and } a \neq 0 \\
& \quad a = 0 \\
& \quad -4 a c + b^2 = 0 \\
& \quad \text{otherwise}
\end{align*}
\]
In the case where we set one of \( \{ k_{17}, k_{18}, k_{19} \} \) to a constant we can obtain a full semi-algebraic solution, i.e., descriptions of those regions in two parameter space where multi-stationarity can occur. Descriptions are:

- **Clear and easy to read**: Set of cells, each cell described by sequence of polynomial equations or inequalities. These are cylindrical. E.g. if over \((k_{17}, k_{18})\) then \(k_{17}\) is a constant (possibly algebraic) or an interval; and \(k_{18}\) either \(k_{18} = f(k_{17})\) or \(\ell(k_{17}) < k_{18} < u(k_{17})\).

- **But large**: too large to put in slide (or even paper). Polynomials of high degree, a great many of such cells. Scope for simplification work here.

In the three parameter case we have identification of solution via 3d grid sampling.
Outline

1. Introduction
   - MAPK Network
   - Symbolic Methods

2. Results
   - Semi-algebraic descriptions of multistationarity region
   - Symbolic vs Numeric Grid Sampling
MAPK models have remarkably low total degrees with many linear monomials. This promoted idea of pre-processing MAPK input with essentially **Gaussian Elimination** (GE): in the sense of solving single suitable equations with respect to some variable and substituting the corresponding solution into the system.

Parametric GE requires case distinctions but here positivity conditions cancelled out the necessary case distinctions. **Key question**: does this generalise Models 26 and 28?

**Strategy for optimal GE**: Draw a graph, where vertices are variables and edges indicate multiplication between variables within some monomial. Then Gauss-eliminate a maximum independent set, which is the complement of a minimum vertex cover.
Model 26: Pre-processing

\[
\overline{\psi} = x_5 > 0 \land x_4 > 0 \land k_{19} > 0 \land k_{18} > 0 \land k_{17} > 0
\]

\[
\land 1062444k_{18}x_4^2x_5 + 23478000k_{18}x_4^2 + 1153450k_{18}x_4x_5^2 + 2967000k_{18}x_4x_5 \\
+ 638825k_{18}x_5^3 + 49944500k_{18}x_5^2 - 5934k_{19}x_4^2x_5 - 989000k_{19}x_4x_5^2 \\
- 1062444x_4^3x_5 - 23478000x_4^3 - 1153450x_4^2x_5^2 - 2967000x_4x_5 \\
- 638825x_4x_5^3 - 49944500x_4x_5^2 = 0
\]

\[
\land 1062444k_{17}x_4^2x_5 + 23478000k_{17}x_4^2 + 1153450k_{17}x_4x_5^2 + 2967000k_{17}x_4x_5 \\
+ 638825k_{17}x_5^3 + 49944500k_{17}x_5^2 - 1056510k_{19}x_4^2x_5 - 164450k_{19}x_4x_5^2 \\
- 638825k_{19}x_5^3 - 1062444x_4^2x_5^2 - 23478000x_4^2x_5 - 1153450x_4x_5^3 \\
- 2967000x_4x_5^2 - 638825x_5^4 - 49944500x_5^3 = 0.
\]
We now assume one parameter is set to constant \((k_{18} = 50)\). We start with LRT. In 5 seconds we obtain:

**Component:** Same positivity conditions and two equations:
- one equation does not involve \(x_5\) at all (triangular);
- the other is only linear in \(x_4\);
- but total degrees and number of terms higher than \(\overline{\psi}\).

**Unevaluated Calls (Lazy)**
- two evaluate trivially to NULL;
- the other two define solutions on graphs of two polynomials in \((k_{17}, k_{19})\)-space. These are blind spots for our solution.
To finish we use CAD technology:

- Perform CAD projection for the equation without $x_5$ and those forming positivity conditions.
- Build Open CAD of $(k_{17}, k_{19})$-space for them (meaning full dimensional cells only - so further graphs forming blind spots for solution)
- Identify cells in upper quadrant of interest.
- Isolate and calculate real roots of equational polynomial in each.
- Check corresponding roots in $x_5$ are positive.
- If three positive real roots then mark as multistationarity region.

We hence identify the region as 35 semi-algebraic cells in 17 seconds.
Outline

1. Introduction
   - MAPK Network
   - Symbolic Methods

2. Results
   - Semi-algebraic descriptions of multistationarity region
   - Symbolic vs Numeric Grid Sampling
Grid Sampling

We may use grid-sampling to get an understanding of the parameter region in three dimensions. We considered two approaches:

1. **Symbolic**: Iteratively applying RT + CAD with no free parameters.
2. **Numeric**: Using the homotopy solver **BERTINI**.

Comparison:

- For Model 26 the symbolic method actually computed faster than the numeric (unexpected). However, for a larger system (# 28) this was reversed.
- However, in all cases the symbolic methods avoided errors present in the numerical ones.
Grid Sampling Comparison
High Sampling Density
3D Sampling

3D Maple Point Plot produced grid sampling on Biomod-26

Convex Hull of the bistable points

Matthew England
Symbolic computation for models of the MAPK network
Things we did not have time to discuss today:

- Results also for Model # 28.
- Alternative approach combining VS + CAD.

Conclusions:

- Problems like MAPK would have been until recently out of the scope of symbolic methods. But by combining the latest approaches progress is possible.
- At meeting in Berlin last year we reported only one-parameter solution - we are now confident of two. Three parameters?
- Combination of symbolic and numeric approach leads to better grid sampling.
Final Thoughts

Things we did not have time to discuss today:
- Results also for Model \# 28.
- Alternative approach combining VS + CAD.

Conclusions:
- Problems like MAPK would have been until recently out of the scope of symbolic methods. But by combining the latest approaches progress is possible.
- At meeting in Berlin last year we reported only one-parameter solution - we are now confident of two. Three parameters?
- Combination of symbolic and numeric approach leads to better grid sampling.
Publications


_A Case Study on the Parametric Occurrence of Multiple Steady States._


_Symbolic Versus Numerical Computation and Visualization of Parameter Regions for Multistationarity of Biological Networks._

R. Bradford, et al.

_ Identifying the Parametric Occurrence of Multiple Steady States for some Biological Networks._ Submitted to Journal, 2017.

First two are published (and preprints freely available on Arxiv). If you want a copy of the third just email: Matthew.England@coventry.ac.uk
Further Information

Contact Details

Matthew.England@coventry.ac.uk

Slides available to download from my website:
http://computing.coventry.ac.uk/~mengland/index.html

Thanks for listening!