Computing Core Reactions of Uncertain Polynomial Kinetic Systems

Zoltán A. Tuza[1] and Gábor Szederkényi[1,2]
[1]Pázmány Péter Catholic University, Faculty of Information Technology and Bionics,
Práter u. 50/a, H-1083 Budapest, Hungary
[2]Institute for Computer Science and Control, Hungarian Academy of Sciences,
Kende u. 13-17, H-1111 Budapest, Hungary
e-mail: {tuza.zoltan,szederkenyi}@itk.ppke.hu

***Abstract*— Kinetic systems form a wide nonlinear system class with good descriptive power that can efficiently be used for the dynamical modeling of non-negative models emerging not only in (bio)chemistry but in other important scientific and engineering fields as well. The directed graph structure assigned to kinetic models give us important information about the qualitative dynamical properties of the system. In this paper we extend the previous results for computing structurally invariant directed edges (called core reactions) for uncertain kinetic polynomial models, where the uncertainty is represented as a multi-dimensional interval in the space of monomial coefficients. We show that the computation can be put into the framework of linear programming. Using illustrative examples we demonstrate the properties of the computed structures and the potential application of the method in the support of structural identification of biochemical networks.**

Keywords: *Biologically inspired systems, Nonlinear systems, Optimisation*

I. Introduction

Kinetic dynamical models come from chemistry, but their range of applicability reaches far beyond (bio)chemical models as they are suitable to describe all important dynamical phenomena such as stability/instability and multiplicity of equilibria, limit cycles or even chaos [18]. Kinetic models can effectively be used in the description of numerous natural and technological processes with nonnegative state variables such as population and disease dynamics, compartmental models, or certain transportation phenomena. Transformation of general nonlinear systems into polynomial kinetic form is also possible under mild conditions [14]. Biochemical systems obeying the mass action law can be described by nonlinear polynomial ODEs where there are strict relations between the monomial exponents and coefficients guaranteeing nonnegativity of the solutions (in case of nonnegative initial conditions), and giving rise to a weighted directed graph structure called reaction graph [3], [18]. In this graph, the participating chemical complexes are the nodes in the network and the reactions which transform complexes into each other are represented by weighted directed edges. The reaction rates are directly proportional to the edge weights.

The reconstruction of reaction network structure from measurement data and prior information is an important and intensively studied area [15], [17]. In certain chemical and biological problem statements it is generally assumed that the participating chemical species and the possible chemical complexes of the participating species are known, i.e. the node set is fixed in a graph representation [1]. Hence, an important remaining task is to determine the existence and the rate coefficients of the reactions between the participating complexes. This task unfortunately turns out to be computationally intractable as the number of complexes increase. The problem is well studied in the field of model and parameter estimation, often called as network reconstruction or inference.

It is known that different reaction network structures (weighted directed graphs) may belong to exactly the same dynamics, therefore, the structural and parametric identification problem is generally not uniquely solvable without additional prior knowledge on the network structure, even if we have full and perfect measurements [2], [11]. In certain applications, the assumption of network sparsity may improve the solvability of the inference problem [1], [17]. However, in general, sparse structures corresponding to a given kinetic dynamics are not unique [10]. Therefore, we would like to further analyze the most "certain" structural elements of the network. These are called core-reactions [11], and have the property that they are present in any reaction graph structure (realization) that is compatible with a given kinetic dynamics. The aim of this paper is to extend the previous results in [11] on computing core reactions of kinetic systems to the case when there are uncertainties in the model. These uncertainties will be modeled as intervals for the coefficients of the monomials in the ODEs, similarly to the approach that was used for the estimation of fluxes of metabolic networks in e.g. [8].

II. Kinetic system modeling

A. ODE description

The kinetic models studied in this paper are given in the following polynomial form:

$$\dot{x} = M \cdot \psi(x), \tag{1}$$

where $x \in \mathbb{R}^n_+$, $\mathbb{R}_+$ denotes the non-negative real numbers, $M \in \mathbb{R}^{n \times p}$ and $\psi(x)$ is a monomial-type vector mapping which is defined as

$$\psi_j(x) = \prod_{i=1}^{n} x_i^{\alpha_{ij}}, \quad j = 1, \ldots, p \tag{2}$$

with $\alpha_{ij} \in \mathbb{N}$. In order to define a kinetic system, the following relation has to be fulfilled between M and α [18]:

$$\alpha_{ij} \geq 1 \text{ for any } i,j \text{ for which } M_{ij} < 0. \tag{3}$$

From a parameter estimation point of view, it is worth to mention that the model (1) is linear in the coefficients contained in M, hence many parameter estimation techniques (e.g. the ones based on least squares) can be used in network reconstruction. However, inference often remains a challenging task because of poor excitation of the dynamics, bad measurement quality or the lack of structural identifiability. [15].

For computation purposes, we will use an appropriate factorization of (1) as follows. Let us define $Y \in \mathbb{N}^{n\times m}$ as the complex composition matrix of the system. Additionally, $A_k \in \mathbb{R}^{m\times m}$ is a special compartmental matrix, the so-called Kirchhoff matrix belonging to the system. A_k is defined as:

$$[A_k]_{ij} = \begin{cases} k_{ji} & \text{if } i \neq j \\ -\sum\limits_{l=1, l\neq i}^{m} k_{il} & \text{if } i = j, \end{cases} \tag{4}$$

where $k_{ij} \geq 0 \ \forall i,j$. With the help of these two matrices, we can write (1) as

$$\dot{x} = Y \cdot A_k \cdot \varphi(x), \tag{5}$$

where

$$\varphi_j(x) = \prod_{i=1}^{n} x_i^{Y_{ij}} \tag{6}$$

Note that the monomial sets contained in ψ and φ are not necessarily identical, although the right hand sides of eqs. (1) and (5) are equal. This reflects the fact that monomials corresponding to product complexes in the chemical reaction network do not appear in the kinetic differential equations.

B. Directed graph structure

We can associate a graph representation to (5). A kinetic system equipped with this graph structure will be called a Chemical Reaction Network (CRN) as it is described in e.g. [3]. In this representation, a CRN is characterized by three sets:

1) $\mathcal{S} = \{X_1, \ldots, X_n\}$ is the set of *species* or chemical substances.
2) $\mathcal{C} = \{C_1, \ldots, C_m\}$ is the set of *complexes*. Formally, the complexes are represented as linear combinations of the species, i.e.
$$C_i = \sum_{j=1}^{n} \beta_{ij} X_j, \qquad i = 1, \ldots, m, \tag{7}$$
where β_{ij} are nonnegative integers and are called the *stoichiometric coefficients*.
3) $\mathcal{R} = \{(C_i, C_j) \mid C_i, C_j \in \mathcal{C},\ i \neq j,$ and C_i is transformed to C_j in the CRN$\}$ is the set of *reactions*.
The reaction $(C_i, C_j) \in \mathcal{R}$ will be denoted as $C_i \to C_j$. Moreover, a positive weight, the *reaction rate coefficient* denoted by k_{ij} is assigned to each reaction $C_i \to C_j$.

From the above description, a directed graph structure (called the reaction graph) can be directly constructed, where the nodes and weighted directed edges represent complexes and reactions, respectively. It is important to remark that loop edges are not allowed in the reaction graph, i.e. $i \neq j$ for all directed edges $C_i \to C_j$.

The relation between the ODE model (5) and the reaction graph is the following. The state vector x contains the species concentrations. $Y_{ij} = \beta_{ji}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$, and $[A_k]_{ij} = k_{ji}$ is the reaction rate coefficient corresponding to the reaction $C_j \to C_i$. $[A_k]_{ij} = 0$ means that the reaction $C_j \to C_i$ does not occur in the CRN. It is clear that matrices Y and A_k encode the stoichiometric composition and the weighted directed graph of a CRN, respectively, and they are sufficient to write the kinetic dynamics (1) belonging to a reaction graph.

C. Dynamical equivalence of kinetic systems

It has been known since at least the 1970's that multiple different structures/parametrizations of a CRN can generate exactly the same dynamics of the concentrations [4], [18]. This phenomenon is called *macro-equivalence* or *dynamical equivalence*. However, the exact geometric conditions of macro-equivalence were not studied until relatively recently in [2]. Naturally, the phenomenon of dynamical equivalence may hamper the parameter identification process, since multiple structures can explain equally well the modeled dynamics [11].

Mathematically, dynamical equivalence means that the factorization in (5) is non-unique. Therefore, the matrix pair (Y, A_k), where Y is a complex composition matrix and A_k is a Kirchhoff matrix, is called a *dynamically equivalent realization* of the kinetic system (1), if $M \cdot \psi(x) = Y \cdot A_k \cdot \varphi(x)$ $\forall x \in \mathbb{R}^n_+$. We note that a given kinetic dynamics can generally be represented using different complex sets. However, there exists a simple procedure described in [5] that generates a possible dynamically equivalent realization called the canonical structure for any kinetic polynomial model. From now on, we assume that the set of complexes is known and fixed, therefore, all dynamical equivalent realizations can be characterized by the equation

$$Y \cdot A_k = M. \tag{8}$$

Clearly, if $A_k^{(1)}$ and $A_k^{(2)}$ give dynamically equivalent realizations with fixed Y, then $A_k^{(3)} = \frac{A_k^{(1)} + A_k^{(2)}}{2}$ also gives a valid dynamically equivalent realization with Y. Therefore, a kinetic system with different dynamically equivalent realizations has infinitely many dynamically equivalent realizations.

A reaction $C_i \longrightarrow C_j$ of a CRN is called a *core reaction*, if it is present in any dynamically equivalent realization of a kinetic system. Core reactions can be determined using LP-based optimization methods [11].

D. Introductory example

To give an example of a polynomial kinetic system we use a biomolecular model reported in [9]. In this example a dynamical model of positive feedback motif is used, the biomolecular mechanism behind the model and the detailed explanation for the dynamical behavior can be found in the original paper [9].

This model tracks the concentrations of a monomer protein (x_1), its dimer form (x_2), also its mRNA (x_5). The promoter which regulates the production of the protein is also included in the model (x_3 is the unoccupied promoter and x_4 is the occupied promoter). The protein (x_1) is made at a basal rate given by the concentration of x_4 then, this protein forms a dimer (x_2) and binds back to the promoter, forming an occupied promoter (x_4). This loop acts as a positive feedback and accelerates the production of the protein. The kinetic polynomial system representing this dynamics is given as

$$\begin{aligned}
\dot{x}_1 &= -2k_1x_1^2 + 2k_2x_2 + k_9x_5 - k_8x_1 \\
\dot{x}_2 &= k_1x_1^2 - k_2x_2 - k_3x_2x_3 + k_4x_4 \\
\dot{x}_3 &= -k_3x_2x_3 + k_4x_4 \\
\dot{x}_4 &= k_3x_2x_3 - k_4x_4 \\
\dot{x}_5 &= k_5x_4 + k_6x_3 - k_7x_5.
\end{aligned} \tag{9}$$

The parameter values are

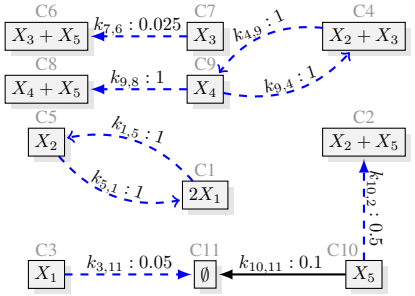
$$\begin{aligned}
k_1 &= 1\ k_2 = 1\ k_3 = 1\ k_4 = 1\ k_5 = 1 \\
k_6 &= 0.025\ k_7 = 0.1\ k_8 = 0.05\ k_9 = 0.5.
\end{aligned} \tag{10}$$

Then, the set of ODEs can be encoded in M and $\psi(x)$:

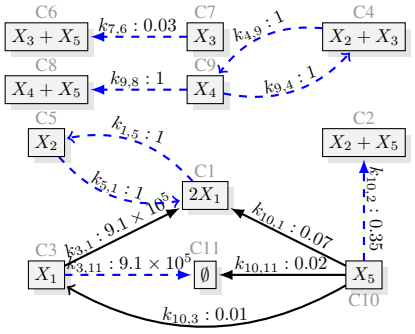
$$M = \begin{bmatrix}
-2k_1 & 0 & -k_8 & 0 & 2k_2 & 0 & 0 & 0 & 0 & k_9 \\
k_1 & 0 & 0 & -k_3 & -k_2 & 0 & 0 & 0 & k_4 & 0 \\
0 & 0 & 0 & -k_3 & 0 & 0 & 0 & 0 & k_4 & 0 \\
0 & 0 & 0 & k_3 & 0 & 0 & 0 & 0 & -k_4 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_5 & -k_7
\end{bmatrix} \tag{11}$$

$$\psi(x) = \begin{bmatrix} x_1^2 & x_1x_5 & x_1 & x_2x_3 & x_2 & x_3x_5 & x_3 & x_4x_5 & x_4 & x_5 \end{bmatrix}^T$$

Figure 1 shows two dynamically equivalent realizations of (9). The core reactions which are structurally invariant under dynamical equivalence are shown with dashed blue edges while non-core reactions are shown with black edges.



(a) A reaction graph representation of the positive feedback motif based on (9)



(b) This reaction graph shows the maximum number of reaction possible within the boundaries of dynamical equivalence. The graph was calculated via an optimization algorithm reported in [12].

Fig. 1: Two dynamically equivalent realizations of the positive feedback motif. The dashed blue edges represent the core-reactions which are structurally invariant under dynamical equivalence, i.e. the edge exist in all dynamically equivalent realization, but the weight on the might be different. The black edges are the non-core reactions. The reader can notice that the graph on panel (a) is a subgraph of the graph on panel (b), this property was proven in [6].

III. Uncertain polynomial kinetic systems

A. Parametric uncertainty of polynomial kinetic systems

The uncertainty of kinetic models will be modeled by intervals of the monomial coefficients. Thus, a family of kinetic systems is obtained that can easily be represented as

$$\begin{aligned}
\dot{x} &= M \cdot \psi(x) \\
[M_l]_{ij} &\le [M]_{ij} \le [M_u]_{ij}
\end{aligned} \tag{12}$$

where matrices $M_l \in \mathbb{R}^{n\times m}$, $M_u \in \mathbb{R}^{n\times m}$ and $M \in \mathbb{R}^{n\times m}$ are kinetic matrices defined in Equation (3). This formalism defines a family of kinetic models and just like in the case of the regular polynomial kinetic systems, an uncertain polynomial kinetic system is completely characterized by the matrix triplet Y,M_l,M_u [13].

With these upper and lower bounds we can represent the certainty of the system's parameters individually. These bounds can represent uncertainty of the parameters due to temperature change, measurement error, poor excitation of the dynamics, etc. In case of parameter estimation it can represent the parameter uncertainty, hence the quality of parameter estimation can be calculated via confidence intervals which are directly applicable in Equation (12).

B. Calculating Core reactions of Uncertain Polynomial Kinetic systems

We can take advantage of the fact that if core reactions exist, then they are present in any dynamically equivalent realization. Therefore, they are essential part of the system's structure to produce the dynamics described by matrix pair Y, M.

Formally, a reaction $(C_i, C_j) \in \mathcal{R}, i, j \in \{1, \ldots, m\}$ is a core reaction if and only if for any dynamically equivalent realization (Y, A_k), $[A_k]_{ji} > 0$ holds. This condition can be

translated as a constraint set in a linear program as it was introduced in [11].

1) Algorithm for computing core reactions: Let us assume that we have an uncertain kinetic system given by matrices Y, M_l and M_u, then the following linear program (LP) can be formulated

$$\min \textstyle\sum_{ij} E_{ij}[A_k]_{ij} \tag{13}$$

s.t.

$$\textstyle\sum_{i=1}^{m} [A_k]_{ij} = 0, \quad j = 1, \ldots, m \tag{14}$$

$$[A_k]_{ij} \geq 0, i, j = 1, \ldots, m, \ i \neq j \tag{15}$$

$$Y \cdot A_k - M = 0 \tag{16}$$

$$[M_l]_{ij} \leq [M]_{ij} \leq [M_u]_{ij} \tag{17}$$

where entries of matrices A_k and M are both optimization variables of the optimization problem. The binary matrix $E \in \{0,1\}^{m \times m}$ selects the elements of A_k that are part of the cost function.

The first two constrains (Eqs. 14 and 15) force A_k to be a Kirchhoff matrix (as defined in Eq. (4)). The constraint in Equation 16 enforces dynamical equivalence (see Eq. (8) for the definition). Finally, the last type of constraints (Eq. 17) set up element-wise the lower and upper bounds for all entries of the kinetic matrix M.

The LP defined in Equation (13) is called as `DynEqLP` in the pseudo code below and this LP needs 4 inputs for operation, such as Y, M_l, M_u and E and calculates a realization characterized by matrices A_k and M.

ALGORITHM 1: The goal of the algorithm is to find core reactions of an Uncertain Kinetic System characterized by matrix triplet (Y,M_l,M_u). The Algorithm returns with the set of core reactions, $\mathcal{R}_c$.

$D \leftarrow 0_m$ // empty matrix;
$E \leftarrow 1_m - I_m$ // only off-diagonals are non-zero ;
$A_k \leftarrow 0_m$ // empty matrix ;
while *true* **do**
  $A_k \leftarrow$ DynEqLP(Y,M_l,M_u,E);
  $D \leftarrow E$;
  $E \leftarrow$ (PositiveElements(A_k) & D) // & is an element-wise logical and ;
  if $D = E$ **then**
    break;
  end
end
$\mathcal{R}_C \leftarrow \emptyset$;
foreach $(i, j) \in$ *PositiveElements(E)* **do** ;
$Z \leftarrow \mathbf{0}_{m \times m}$;
$Z(i, j) \leftarrow 1$;
$A_k \leftarrow$ DynEqLP(Y,M_l,M_u,Z) ;
if $A_k(i, j) > 0$ **then**
  $\mathcal{R}_C \leftarrow \mathcal{R}_C \cup (C_j \rightarrow C_i)$;
end
;
return $\mathcal{R}_C$

In Algorithm 1, the procedure *PositiveElements* finds the nonzero elements of the argument and gives back a binary matrix containing ones where the original matrix elements are larger then zero.

Algorithm 1 starts with binary matrix E where the off-diagonal elements of A_k are ones and the rest are zero. In the first iteration the `DynEqLP` tries to minimize the sum of off-diagonal elements of A_k. After that, in each iteration the zero off-diagonal elements are excluded from the cost function until no more off-diagonal elements can be excluded, i.e. matrix E remains the same between two iterations. In the final stage, for each of the remaining non-zero off-diagonal elements (encoded in matrix E) the `DynEqLP` is executed where only one element from matrix E is minimized. If this element remains non-zero after the optimization, then this reaction (edge) is part of the core reaction set.

In the worst case, the Algorithm 1 has to solve the `DynEqLP` $2 + (m^2 - 1)$ times. Hence, this is a polynomial time algorithm to find core reactions of an uncertain polynomial kinetic system. In each iteration the number of (decision) variables of `DynEqLP` is $nm + m^2 - m$. Furthermore, this LP has $2m$ equality constraints and $2nm + m^2$ inequality constraints.

The proposed algorithm for calculation of core reactions outperforms the previously applied one in two areas. First, the previous algorithm, detailed in [11], is based on systematic exclusion of reactions and testing for the feasibility of the resulting linear program, which might take palace for different reasons than exclusion of a reaction. Second, the previous algorithm has to check systematically each off-diagonal element of A_k to establish the core reaction set. By comparison, this can only happen as the worst case of the proposed algorithm.

IV. Examples

A. Example of an Uncertain Kinetic System

In this this example, we revisit the positive feedback motif from Section II-D and investigate that how increasing intervals around matrix M affect the core reaction set, $\mathcal{R}_C$.

The positive feedback motif model is encoded with matrix pair (Y,A_k) as

$$Y = \begin{bmatrix} 2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \tag{18}$$

and the only non-zero off-diagonal elements of A_k are

$$\begin{aligned} [A_k]_{5,1} &= k_1, \ [A_k]_{1,5} = k_2, \ [A_k]_{9,4} = k_3, \\ [A_k]_{4,9} &= k_4, \ [A_k]_{8,9} = k_5, \ [A_k]_{6,7} = k_6, \\ [A_k]_{11,10} &= k_7, \ [A_k]_{11,3} = k_8, \ [A_k]_{2,10} = k_9. \end{aligned} \tag{19}$$

First, the lower and upper bounds are set as $M_l = M_u = M$ (i.e. there is no parametric uncertainty in the system) and with Algorithm 1 we calculated that this system has $|\mathcal{R}_C| = 8$ core reactions and these are

$$\begin{aligned} &C_1 \rightarrow C_5, C_3 \rightarrow C_{11}, \ C_4 \rightarrow C_9, \ C_5 \rightarrow C_1 \\ &C_7 \rightarrow C_6, C_9 \rightarrow C_4, \ C_9 \rightarrow C_8, \ C_{10} \rightarrow C_2 \end{aligned} \tag{20}$$

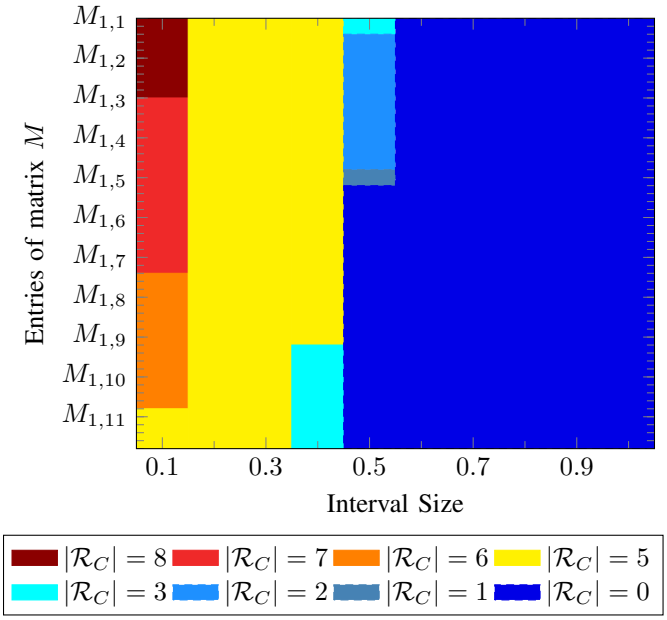


Fig. 2: Figure shows how the increasing intervals around M is affecting the number of core reactions inside the interval. The vertical axis lists the elements of M and the horizontal axis shows the accumulation of the interval size around the values of M. The calculation start at the top left corner and goes down along the horizontal axis, then current interval gets increased and the calculation start at the top in the next column along the vertical axis. Each color represents the size of the core reaction set within the interval. In each iteration, the applied step size was 0.1.

and they are shown as blue dashed edges on Figure 1.

The matrix M is defined by Equations (18) and (19) has 5×11 elements. In each step we symmetrically increase the interval around one element of M, then the invariant reaction set is calculated with the help of Algorithm 1. Meanwhile the kinetic property of matrix M is ensured by checking the sign constraints of M, defined by Equation (3). The result is summarized on Figure 2, in the top left corner there are 8 core reactions in the system (dark read area) as we increase the bounding box around M the number of core reactions gradually decrease. Finally, the core reaction set becomes empty (dark blue area). It should be noted that in each step the previous interval is a subset of the current interval, hence the core reaction set can either remain the same or become a subset of its previous one.

B. Network reconstruction example

Using a chemical reaction network example, we highlight a possible field of application of the algorithm presented in Section III-B. In (bio)chemical modeling and many other fields, the goal of the network reconstruction is—in most of the cases—to find the sparsest network describing the measured dynamics. Therefore, our goal is to find the structurally invariant elements that are characteristic for the dynamics while the parameters of the network are uncertain.

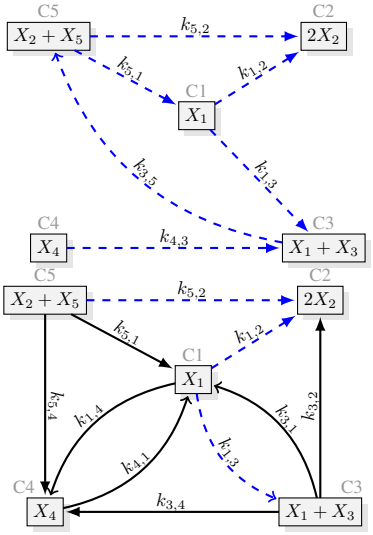


Fig. 3: Comparison of the original network from [1] (top) and the network given by parameter estimation (bottom). The core reactions in each case are shown with blue dashed edges.

First let us investigate a benchmark example from the literature which is based on Figure 2 in [1] and encoded as

$$Y = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{21}$$

and

$$A_k = \begin{bmatrix} -1.163 & 0 & 0 & 0 & 0.8492 \\ 0.3386 & 0 & 0 & 0 & 0.4290 \\ 0.8244 & 0 & -0.7364 & 0.5631 & 0 \\ 0 & 0 & 0 & -0.5631 & 0 \\ 0 & 0 & 0.7364 & 0 & -1.2782 \end{bmatrix} \tag{22}$$

As a first step we establish the core reaction set for the dynamics represented by (Y,A_k) and shown on the top panel of Figure 3. The Algorithm 1 with boundary matrices $M_u = M_l = Y \cdot A_k$ tells us that all six reactions in the original network are core reactions (blue dashed edges on the top panel of Figure 3), hence this is the sparsest and also the only sparse realization of the given dynamics.

The goal of the parameter estimation is to obtain an interval model in the form of Equation (13) from time series measurements of the species concentration. During the parameter estimation procedure, we will exploit the fact that the type of kinetic systems we use in this paper is linear in parameters. Finally, the parametric uncertainty will be characterized by the covariance matrix of the estimator to calculate the core reaction set of the interval model.

1) Parameter Estimation Procedure: We utilize the discrete Least Squares framework and for that we discretize Equation (1). Taking sufficiently small intervals between

samples we apply the forward Euler method

$$x_i(k) = x_i(k-1) + hM_{i,\cdot}\,\psi_i(x(k-1)) \quad k = 2...T_{end},\ i = 1,\dots,n \tag{23}$$

where x_i is the ith state variable from Equation (1), $\psi_i(x)$ is the ith element of vector mapping ψ and vector $M_{i,\cdot}$ is the ith row of matrix M. Finally, the h is the sampling time, which is $h = 0.1\ sec$ in our case.

We assume that we can measure all state variables, then let us define the outputs as

$$y_i(k) := x_i(k) - x_i(k-1),\ i = 1,\dots,n$$

and model, which is used for the the parameter estimation

$$y_i(k,\theta) = \theta_{i,\cdot}^T\,\varphi(k)$$

where $\theta_{i,\cdot}$ is the ith row of M and

$$\varphi(k) = \left[\psi_1(x(k-1))\ \psi_2(x(k-1))\ \psi_m(x(k-1))\right]^T.$$

Further, let us define matrix $R \in \mathbb{R}^{m\times m}$ and vector $d_i \in \mathbb{R}^m$

$$R = \frac{1}{N}\sum_{k=2}^{N}\varphi(k)\varphi(k)^T \tag{24}$$

$$d_i = \frac{1}{N}\sum_{k=2}^{N}\varphi(k)y_i(k). \tag{25}$$

Then, the parameter vector θ_i can be calculated using a closed form as

$$\hat{\theta}_i = R^{-1}d_i. \tag{26}$$

The regression matrix R depends on all of the state variables, hence each row of M can be calculated in one step. By defining

$$C = \text{diag}(R,\dots,R) \tag{27}$$

where $C \in \mathbb{R}^{n\cdot m\times n\cdot m}$ is block diagonal matrix and

$$d = \left[d_1\ \cdots\ d_i\ \cdots\ d_n\right]^T \tag{28}$$

then, we can calculate all the elements of M in one optimization step. Let us define parameter vector $\theta = \text{row}(M)^T$ which is the transpose of the row expansion of M. The row expansion of M is defined as $\text{row}(M) = \left[M_{1,\cdot}, M_{2,\cdot},\dots,M_{n,\cdot}\right]$ where $M_{i,\cdot}$ denotes the ith row of matrix M.

Then, parameter vector θ can be calculated as

$$\hat{\theta} = \underset{\theta\in\Theta}{\text{argmin}}\ \frac{1}{2}||C\theta - d||_2^2. \tag{29}$$

To ensure that $\hat{\theta}$ represents a proper kinetic matrix, the set Θ denotes the possible parameter vectors where the sign condition of M is fulfilled (see Equation (3) for details). Finally, the estimated matrix M is given as $\hat{\theta}/h$.

Estimation of matrix A_k: It is possible to improve the parameter estimation by exploiting the fact that with fixed matrix Y, the matrix M can be written as $Y \cdot A_k$. Hence, we can estimate A_k directly. For that, we define

$$\bar{Y} = \begin{bmatrix} Y_{1,\cdot} & 0 & \cdots & 0 \\ 0 & Y_{1,\cdot} & \cdots & 0 \\ 0 & 0 & \cdots & Y_{1,\cdot} \\ Y_{2,\cdot} & 0 & \cdots & 0 \\ 0 & Y_{2,\cdot} & \cdots & 0 \\ 0 & 0 & \cdots & Y_{2,\cdot} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{n,\cdot} & 0 & \cdots & 0 \\ 0 & Y_{n,\cdot} & \cdots & 0 \\ 0 & 0 & \cdots & Y_{n,\cdot} \end{bmatrix} \quad \bar{A}_k = \begin{bmatrix} [A_k]_{1,\cdot}^T \\ \vdots \\ [A_k]_{i,\cdot}^T \\ \vdots \\ [A_k]_{m,\cdot}^T \end{bmatrix} \tag{30}$$

where $Y_{i,\cdot}$ denotes the ith row of matrix Y and vector $[A_k]_{j,\cdot}$ is the jth row of matrix A_k.

Then, we formulate the following constrained Least Squares optimization problem

$$\begin{gathered} \min_{\bar{A}_k}\ \tfrac{1}{2}||C\bar{Y}\bar{A}_k - d||_2^2 \\ \text{s.t.} \\ \textstyle\sum_{i=1}^{m}[A_k]_{ij} = 0, \quad j = 1,\dots,m \\ [A_k]_{ij} \geq 0, i, j = 1,\dots,m,\ i \neq j. \end{gathered} \tag{31}$$

Finally, the optimal value of $\bar{A}_k$ is denoted as $\hat{\bar{A}}_k$. In this way we can have an estimate matrix M as $\bar{Y}\hat{\bar{A}}_k/h$, but imposing a set of constraints on the elements of A_k.

2) Results: Now, we attempt to (partially) restore the entries of matrix A_k which was defined in Equation (22) by applying the Least Squares framework from Subsection IV-B.1. For this, we assume that we can measure all state variables of the kinetic system and to each state variable 5% additive Gaussian noise is added. A dataset with 25 different initial vectors is generated with uniform Latin hypercube sampling. Using the optimization defined in Equation (31), we calculate the graph structure from the noisy dataset as follows

$$A_k^{est} = \begin{bmatrix} -2.3 & 0 & 0.46 & 1.51 & 0.07 \\ 1.1 & 0 & 0.06 & 0 & 0.09 \\ 0.92 & 0 & -0.58 & 0 & 0 \\ 0.26 & 0 & 0.065 & -1.51 & 0.047 \\ 0 & 0 & 0 & 0 & -0.21 \end{bmatrix}. \tag{32}$$

The graph encoded by Equation (32) can be seen on the bottom panel of Figure 3.

Thanks to the factorization of M in Equation (31), we have an estimate of matrix M as

$$\hat{M} = \bar{Y}\hat{\bar{A}}_k/h. \tag{33}$$

The parametric uncertainty around matrix M is extracted from the estimator's covariance matrix. It is calculated as $Cov(\hat{\theta}) = C^{-1}\cdot\sigma^2$ where σ^2 is the standard deviation of the measurement noise. Then, for each element of the estimated matrix M, we establish the 95% confidence interval and cut

off intervals that contradicts the sign constraints of matrix M as follows

$$M_l = \hat{M} - 1.96\sqrt{\text{diag}(Cov(\hat{\theta}))} \quad (34)$$

$$[M_l]_i = \begin{cases} \max([M_l]_i, 0) & \text{if } [\text{row}(Y)]_i > 0 \\ [M_l]_i & \text{otherwsie} \end{cases} \quad i = 1, \ldots, n \cdot m \quad (35)$$

$$M_u = \hat{M} + 1.96\sqrt{\text{diag}(Cov(\hat{\theta}))} \quad (36)$$

where $\text{diag}(Cov(\hat{\theta}))$ selects the diagonal elements of the covariance matrix $Cov(\hat{\theta})$.

These modified confidence intervals define the interval model as it was introduced in Equation (13), and we apply Algorithm 1 on matrices (Y,M_l,M_u) as input. It took only 6 iterations for the algorithm to establish the core reaction set, compare to the previous algorithm which would require 20 iteration to complete the same task, it is more than three times speed up.

The result shows that 3 reactions are in the core reaction set, namely $C_1 \to C_2, C_1 \to C_3$ and $C_5 \to C_2$ (blue dashed edges on the bottom panel of Figure 3). This information can be utilized during the further refinement of the network reconstruction, e.g. in case of a new dataset, these reactions can be constrained as fixed parts of the graph. Even though the Least Squares approach performed poorly on the reconstruction of the original network structure, this performance was enough to recover a subset of the original core reaction set. That shows an important feature of this approach, namely it is capable of recovering the core reaction set, even if the original network is only partially restored, which is often the case with other network reconstruction approaches [1], [16]. It should be noted that the covariance matrix of the contained estimation can be calculated directly using optimization, e.g. in [7], which might improve the results.

In this example, we used the simplest possible method to show how our methods can support the network reconstruction procedure. Any type of network reconstitution can be used as long as some form of parameter uncertainty is available from the applied method or from a priori information.

C. Example for an uncertain system with non-unique sparse realizations

In this example we revisit the positive feedback motif introduced in Section II-D. Examples of multiple dynamically equivalent sparse realizations of the positive feedback motif was reported in [11] (see Figure 2 in [11]). In this model, there are 8 core reactions and a sparse structure has 9 reactions, this difference is a necessary condition to have more than one sparse structure.

We want to investigate the question of sparse non-uniqueness in case of uncertain models. Since we do not have measurement, we add roughly 10% uncertainty to each parameter. By applying Algorithm 1 on the positive feedback motif with $[M_l]_{ij} = [M]_{ij} - 0.1$ and $[M_u]_{ij} = [M]_{ij} + 0.1$ where M was given in Equation (11) yields a core reaction set with 5 elements (yellow region in Figure 2) and shown with blue dashed edges in Figure 4.

The algorithm calculated the core reaction set for this example in 10 iterations, which is only 9% of the number of iterations would be required by the previous algorithm from [11].

To calculate the sparse structure of an uncertain kinetic polynomial system we need to formulate an MILP optimization based on a method described in [13] and Equation (13)

$$\begin{gathered} \text{minimize} \sum_{i,j=1,i\neq j}^{m} \delta_{ij} \\ Y \cdot A_k - M = 0 \\ \sum_{i=1}^{m} [A_k]_{ij} = 0,\ j = 1, \ldots, m \\ [A_k]_{ij} \geq 0\ i, j = 1, \ldots, m\ i \neq j \\ 0 \leq [A_k]_{ij} - \delta_{ij}\ i, j = 1, \ldots, m, i \neq j \\ 0 \leq -[A_k]_{ij} + l_{ij}\delta_{ij} \leq 0,\ i, j = 1, \ldots, m, i \neq j \\ [M_l]_{ij} \leq [M]_{ij} \leq [M_u]_{ij}\ i, j = 1, \ldots, m. \end{gathered} \quad (37)$$

We calculate a sparse realization of the uncertain positive feedback motif by applying optimization problem defined in Equation (37). The resulting realization can be seen on the top panel of Figure 4.

The sparse realization on the top panel of Figure 4 has only one non-core reaction ($C_{10} \to C_2$), thus we can define a set denoted by $\mathcal{K}$, for reactions we want to exclude from the reaction set an uncertain sparse realization. To do that we are modifying the cost function in Equation (37) as

$$\sum_{i,j=1;i\neq j;i,j\notin\mathcal{K}}^{m} \delta_{ij}. \quad (38)$$

Let us exclude reaction $C_{10} \to C_2$ by adding it to set $\mathcal{K}$ and calculate another sparse realization of the uncertain positive feedback motif (depicted on the bottom panel of Figure 4). In this realization only reaction $C_{10} \to C_1$ is not part of the core reaction set. Again, we remove the previous elements from set $\mathcal{K}$ and add $[A_k]_{1,10}$ to it and calculate another uncertain sparse realization. The result is the same as the realization shown on the top panel of Figure 4, therefore at this level of uncertainty only two sparse realizations exist for the positive feedback motif model.

The reader can notice that by comparing Figures 1 and 4, in the later one some complexes became isolated, i.e. a complex has no incoming or outgoing edges. The explanation of this lays in the bounds around M. By comparing the corresponding M matrix of the sparse realization on the top panel of Figure 4, denoted by M^{bound}, and the current lower bound M_l, it can be seen that, the current lower bounds on matrix M allows only the columns 2,3,6,7,8 and 11 to be zero (see matrix M_l in Eq. (40)). Since we try to minimize the number of reactions in the realization, the optimization tries to push the values of matrix M toward the minimum bound (M_l). As a result of that isolated complexes may emerge.

It should be noted that there exist another case when a column in M^{bound} is zero, but the corresponding complex is not isolated. In both sparse cases on Figure 4, the complex

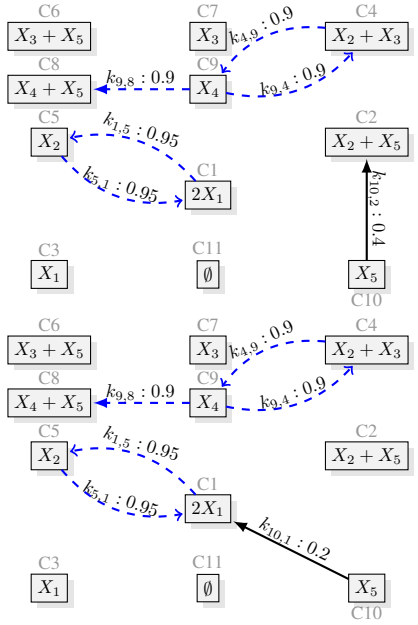


Fig. 4: Two sparse realizations of the positive feedback motif with $[M_l]_{ij} = [M]_{ij} - 0.1$ and $[M_u]_{ij} = [M]_{ij} + 0.1$.

C_8 is not an isolated complex. It is explained by the fact, complex C_8 is a product complex, i.e. it has only ingoing edges, which causes the associated column in the matrix M^{bound} in Equation (39) to become zero.

$$M^{bound} = \begin{bmatrix} -1.9 & 0 & 0 & 0 & 1.9 & 0 & 0 & 0 & 0 & 0.4 & 0 \\ 0.95 & 0 & 0 & -0.9 & -0.95 & 0 & 0 & 0 & 0.9 & 0 & 0 \\ 0 & 0 & 0 & -0.9 & 0 & 0 & 0 & 0 & 0.9 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 & 0 & 0 & 0 & -0.9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9 & 0 & 0 \end{bmatrix} \quad (39)$$

$$M_l = \begin{bmatrix} -2.1 & -0.1 & -0.15 & 0 & 1.9 & 0 & 0 & 0 & 0 & 0.4 & 0 \\ 0.9 & 0 & 0 & -1.1 & -1.1 & 0 & 0 & 0 & 0.9 & 0 & 0 \\ 0 & 0 & 0 & -1.1 & 0 & -0.1 & -0.1 & 0 & 0.9 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 & 0 & 0 & -0.1 & -1.1 & 0 & 0 \\ 0 & -0.1 & 0 & 0 & 0 & -0.1 & 0 & -0.1 & 0.9 & -0.2 & 0 \end{bmatrix} \quad (40)$$

This example shows that how to calculate multiple sparse realizations of a kinetic system with parametric uncertainty, exploiting the core reactions which are the only certain elements of the (uncertain) kinetic model.

V. Conclusion

We have given an effective algorithm to compute the so-called core reactions of uncertain kinetic polynomial models assuming a given complex set. The method iteratively uses linear programming steps, and it is shown that it runs in polynomial time. Several numerical examples have been given to illustrate the algorithm: firstly, the effect of parametric uncertainty on the core reaction set is analyzed, and through another example, we have highlighted how to utilize the core reaction calculation in a simple network reconstruction problem. Finally, we have shown that how to compute multiple sparse realizations of a kinetic system with parametric uncertainty exploiting the structural invariance of the core reactions.

VI. Acknowledgments

This work was partially supported by the Hungarian Scientific Research Fund under grant no. OTKA NF104706 and by KAP-1.1-14/029. We would like to thank Bernadett Ács and Attila Gábor for their suggestions and comments on the manuscript.

References

[1] E. August and A. Papachristodoulou. Efficient, sparse biological network determination. *BMC Systems Biology*, 3:25, 2009.

[2] G. Craciun and C. Pantea. Identifiability of chemical reaction networks. *Journal of Mathematical Chemistry*, 44:244–259, 2008.

[3] M. Feinberg. *Lectures on chemical reaction networks*. Notes of lectures given at the Mathematics Research Center, University of Wisconsin, 1979.

[4] F. Horn and R. Jackson. General mass action kinetics. *Archive for Rational Mechanics and Analysis*, 47:81–116, 1972.

[5] V. Hárs and J. Tóth. On the inverse problem of reaction kinetics. In M. Farkas and L. Hatvani, editors, *Qualitative Theory of Differential Equations*, volume 30 of *Coll. Math. Soc. J. Bolyai*, pages 363–379. North-Holland, Amsterdam, 1981.

[6] M. D. Johnston, D. Siegel, and G. Szederkényi. Dynamical equivalence and linear conjugacy of chemical reaction networks: new results and methods. *MATCH Commun. Math. Comput. Chem.*, accepted:to appear, 2012.

[7] Chong Kiew Liew. Inequality constrained least-squares estimation. *Journal of the American Statistical Association*, 71(355):746–751, 1976.

[8] F. Llaneras, A. Sala, and J. Pico. Dynamic estimations of metabolic fluxes with constraint-based models and possibility theory. *Journal of Process Control*, 22:1946–1955, 2012.

[9] Y. Mileyko, R. I. Joh, and J. S. Weitz. Small-scale copy number variation and large-scale changes in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 105:16659–16664, 2008.

[10] G. Szederkényi. Computing sparse and dense realizations of reaction kinetic systems. *Journal of Mathematical Chemistry*, 47:551–568, 2010.

[11] G. Szederkényi, J. R. Banga, and A. A. Alonso. Inference of complex biological networks: distinguishability issues and optimization-based solutions. *BMC Systems Biology*, 5:177, 2011.

[12] G. Szederkényi, K. M. Hangos, and T. Péni. Maximal and minimal realizations of reaction kinetic systems: computation and properties. *MATCH Commun. Math. Comput. Chem.*, 65:309–332, 2011.

[13] Gábor Szederkényi, Zoltan A Tuza, and Katalin M Hangos. Determining biochemical reaction network structures for kinetic polynomial models with uncertain coefficients. *AIP Conference Proceedings-American Institute of Physics*, 1479:2427–2430, 2012.

[14] Zoltán András Tuza, Gábor Szederkényi, Katalin M. Hangos, Antonio A. Alonso, and Julio R. Banga. Computing all sparse kinetic structures for a lorenz system using optimization. *I. J. Bifurcation and Chaos*, 23(8), 2013.

[15] Éric Walter, Luc Pronzato, and John Norton. *Identification of parametric models from experimental data*. Communications and control engineering. Springer Paris, Berlin, 1997. Translated from an updated French version : Identification de modèles paramétriques à partir de données expérimentales.

[16] Ye Yuan, Guy-Bart Stan, Sean Warnick, and Jorge Goncalves. Robust dynamical network structure reconstruction. *Automatica*, 47(6):1230 – 1235, 2011. Special Issue on Systems Biology.

[17] M. M. Zavlanos, A. A. Julius, S. P. Boyd, and G. J. Pappas. Inferring stable genetic networks from steady-state data. *Automatica*, 47(6):1113–1122, 2011.

[18] P. Érdi and J. Tóth. *Mathematical Models of Chemical Reactions. Theory and Applications of Deterministic and Stochastic Models*. Manchester University Press, Princeton University Press, Manchester, Princeton, 1989.