Determining biochemical reaction network structures for kinetic polynomial models with uncertain coefficients

Gábor Szederkényi[*,†], Z. A. Tuza[†] and Katalin M. Hangos[**,‡]

[*]*Process Control Research Group, Computer and Automation Research Institute, Hungarian Academy of Sciences. Kende u. 13-17, H-1111 Budapest, Hungary, e-mail: szeder@scl.sztaki.hu*
[†]*Faculty of Information Technology, Pázmány Péter Catholic University, Práter u. 50/a, H-1083 Budapest, Hungary*
[**]*Process Control Research Group, Computer and Automation Research Institute, Hungarian Academy of Sciences. Kende u. 13-17, H-1111 Budapest, Hungary*
[‡]*Dept. of Electrical Engineering and Information Systems, Faculty of Information Technology, University of Pannonia, Egyetem u. 10, H-8200 Veszprém, Hungary*

Abstract. A numerical method is proposed in this paper for the computation of dense and sparse reaction network structures for kinetic polynomial models with uncertain parameters represented as intervals. The problem is traced back to mixed integer linear programming.

Keywords: chemical reaction networks, network structure, dynamical systems, optimization
PACS: 05.20.Dd, 05.10.-a, 02.60.Pn

INTRODUCTION

An important subclass of nonnegative polynomial dynamical systems is the set of chemical reaction network (CRN) models obeying the mass-action law [1]. Such networks can be used to describe pure chemical reactions, but they are also widely used to model the dynamics of intracellular processes, metabolic or cell signalling pathways [2]. Thus, CRNs are able to describe key mechanisms both in industrial processes and living systems. Being able to produce all the important qualitative dynamical properties like stable and unstable equilibria, multiple equilibria, bifurcation phenomena, oscillatory and even chaotic behaviour [3], CRN models are suitable for describing basically any smooth bounded nonnegative dynamics [4]. This fact explains that CRNs have attracted significant attention not only among chemists but in numerous other fields such as physics, or even pure and applied mathematics where nonlinear dynamical systems are considered [5, 6, 7].

It has been known for long that different reaction networks can produce exactly the same kinetic differential equations [8, 1]. This phenomenon is also called *macro-equivalence*. In our terminology, CRNs with different parametrization are called *dynamically equivalent*, if they give rise to the same ODEs. A possible CRN (with a certain structure and parametrization) having a given dynamics will be called a *realization* of that dynamics. Naturally, the phenomenon of dynamical equivalence has an important impact on the identifiability of reaction rate constants: if a kinetic dynamics have different dynamically equivalent CRN realizations, then the model, where the parameter set to be estimated consists of all reaction rate coefficients, cannot be structurally identifiable [9] that is a fundamental obstacle to effective network inference. Moreover, since many important conditions on the qualitative properties of CRN dynamics are realization-dependent, it is worth examining whether there exists a dynamically equivalent (or sufficiently 'similar') realization that guarantees certain properties of the corresponding dynamics that are not directly recognizable from the initial CRN or from the corresponding differential equations. Such an approach also gives us valuable help to set up meaningful constraints that guarantee the structural uniqueness and identifiability of the examined network structure [10].

Several optimization-based algorithms have been published lately for the computation of dynamically equivalent or linearly conjugate CRN structures with prescribed chemically and/or dynamically relevant properties [11, 12, 13, 14, 15]. The aim of this paper is to introduce interval methods to extend the above mentioned results in order to be able to scan through an interval of parameters (and thus through a family of models) instead of requiring exact dynamical equivalence. The practical motivation behind this problem is that if we have a kinetic polynomial ODE model estimated from concentration measurements, then the exact values of the coefficients are not precisely known.

In the present case, this uncertainty will be simply represented in the form of parameter intervals similarly to the study on metabolic networks in [16].

BACKGROUND

Chemical reaction network representation and dynamical equivalence

Following the standard way [17], we characterize CRNs using the following three sets:

1. $\mathscr{S} = \{X_1, \ldots, X_n\}$ is the set of *species* or chemical substances.
2. $\mathscr{C} = \{C_1, \ldots, C_m\}$ is the set of *complexes*. Formally, the complexes are represented as linear combinations of the species, i.e. $C_i = \sum_{j=1}^{n} \alpha_{ij} X_j, \qquad i = 1, \ldots, m$, where α_{ij} are nonnegative integers and are called the *stoichiometric coefficients*.
3. $\mathscr{R} = \{(C_i, C_j) \mid C_i, C_j \in \mathscr{C},$ and C_i is transformed to C_j in the CRN$\}$ is the set of *reactions*. The relation $(C_i, C_j) \in \mathscr{R}$ will be denoted as $C_i \to C_j$. Moreover, a nonnegative weight, the *reaction rate coefficient* denoted by k_{ij} is assigned to each reaction $C_i \to C_j$. According to our convention, if the reaction $C_i \to C_j$ is not present in the CRN then $k_{ij} = 0$.

Given the above three sets, a weighted directed graph can be built $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, where $\mathscr{V}$ denotes the vertices which are the complexes of the reaction network, $\mathscr{V} = \{C_1, C_2, \ldots, C_m\}$. The set $\mathscr{E}$ contains the directed edges representing the reactions between the complexes, i.e. $(C_i, C_j) \in \mathscr{E}$ if reaction $C_i \to C_j$ occurs. For each edge (C_i, C_j) of the CRN a positive weight k_{ij} is assigned as reaction rate for the reaction $C_i \to C_j$.

To describe the time-evolution of species concentration, we use the following widely applied factorization of the kinetic ODEs [17]:

$$\dot{x} = Y \cdot A_k \cdot \Psi(x) = M \cdot \Psi(x), \tag{1}$$

where $x = [X] \in \mathbb{R}^n$ is the concentration vector of the species from $\mathscr{S}$. $Y \in \mathbb{R}^{n \times m}$ is called the complex composition matrix where the jth column encodes the composition of the C_j complex, namely $Y_{i,j} = \alpha_{j,i}$. The $A_k \in \mathbb{R}^{m \times m}$ matrix stores the structure and parameters of the reaction graph. A_k is a column conservation matrix (also called the *Kirchhoff matrix* of the CRN) and it is defined as follows:

$$[A_k]_{i,j} = \begin{cases} -\sum_{l=1, l \neq i}^{m} k_{il}, & \text{if} \quad i = j \\ k_{ji}, & \text{if} \quad i \neq j. \end{cases} \tag{2}$$

The last term in the differential equation is a monomial-type vector mapping defined by

$$\psi_j(x) = \prod_{i=1}^{n} x_i^{Y_{i,j}}, \quad j = 1, \ldots, m. \tag{3}$$

A set of polynomial differential equations is called *kinetic* if it can be written in the form (1) with a matrix Y containing nonnegative integer entries, and a Kirchhoff matrix A_k.

If the set of applicable chemical complexes is previously given and fixed (which we assume in the paper), then the condition for dynamical equivalence of two networks is simply given by $Y \cdot A_k^{(1)} = Y \cdot A_k^{(2)}$, where $A_k^{(1)}$ and $A_k^{(2)}$ denote the Kirchhoff matrices of the first and second network, respectively.

HANDLING PARAMETER INTERVALS USING OPTIMIZATION

The starting point is that we have a kinetic ODE system $\dot{x} = M \cdot \Psi(x)$ where the coefficients of the monomials (i.e. the elements of the M matrix) are not fixed, but they can belong to previously defined intervals. This way, a family of polynomial models is given within which we can search for possible CRN structures with prescribed properties. In this extended abstract, we are concentrating on determining the so-called *dense* and *sparse* realizations, containing the maximal and minimal number of chemical reactions (i.e. nonzero reaction rate coefficients). As we will show, this task can be written as a mixed integer linear programming (MILP) problem.

The linear constraints corresponding to mass action dynamics can be written as

$$Y \cdot A_k = M \tag{4}$$

$$\sum_{i=1}^{m} [A_k]_{ij} = 0, \;\; j = 1, \ldots, m \tag{5}$$

$$M_{ij}^{min} \leq M_{ij} \leq M_{ij}^{max} \tag{6}$$

$$[A_k^{min}]_{ij} \leq [A_k]_{ij} \leq [A_k^{max}]_{ij}, \tag{7}$$

where $M^{min}, M^{max} \in \mathbb{R}^{n \times m}$ and $A_k^{min}, A_k^{max} \in \mathbb{R}^{m \times m}$ contain the lower and upper bounds for the elements of M and A_k, respectively. The decision variables in (4)-(7) are the elements of M and the off-diagonal elements of A_k, while (4) expresses the constraints corresponding to mass action dynamics, (5) represents the Kirchhoff property of A_k, and (6), (7) contain the bound (interval) constraints of the ODE coefficients and the reaction rate coefficients, respectively. Note that a bounded interval-type constraint for the elements of A_k is only necessary technically if the final optimization problem contains integer variables.

In order to track whether an individual reaction rate coefficient is zero or not, we introduce the binary variables δ_{ij} for $i, j = 1, \ldots, m, \;\; i \neq j$. Then the logical constraint $\delta_{ij} = 1 \;\Leftrightarrow\; k_{ij} > 0$ (where the symbol '$\Leftrightarrow$' denotes 'if and only if') can be translated to the following linear inequalities [18]:

$$0 \leq [A_k]_{ij} - \varepsilon \delta_{ij}, \;\; i, j = 1, \ldots, m, \;\; i \neq j \tag{8}$$

$$0 \leq -[A_k]_{ij} + [A_k^{max}]_{ij} \delta_{ij}, \;\; i, j = 1, \ldots, m, \;\; i \neq j \tag{9}$$

where $\varepsilon \geq 0$ is a small nonnegative threshold for treating reaction rate coefficients as zero or not. Finally, determining the dense or sparse CRN structures leads to the maximization or minimization of the following objective function, respectively:

$$F(\delta) = \sum_{\substack{i,j \\ i \neq j}} \delta_{ij} \tag{10}$$

EXAMPLE

Let us consider the following kinetic ODEs showing possible bistable behaviour taken from [19] and also examined in [20]:

$$\begin{aligned} \dot{x}_1 &= p_1 x_2 - p_2 x_1^2 - p_3 x_1 x_2 - p_4 x_1 \\ \dot{x}_2 &= p_5 x_1^2 - p_6 x_2, \end{aligned} \tag{11}$$

where all the parameters $p_1, \ldots, p_6$ are positive real numbers. Using the algorithm described in [21], it is possible to give the so-called 'canonic' reaction graph of (11) that is shown in Fig. 1.a. In the figure, the complex denoted by 0 is the 'zero complex' for which all the α coefficients are zero (see e.g. [17] for the details). First, let us fix the parameters of (11) randomly as: $p_1 = 1$, $p_2 = 2$, $p_3 = 1.5$, $p_4 = 0.7$, $p_5 = 1.3$, $p_6 = 2.2$. It is visible from subplots 1.b and 1.c that

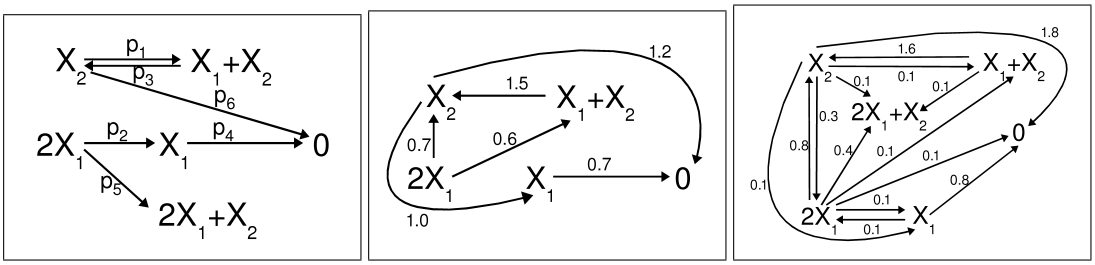

FIGURE 1. **(a)** Canonic realization of eq. (11), **(b)** a possible sparse realization of eq. (11) with fixed parameters, **(c)** dense realization of eq. (11) with fixed parameters

the dynamically equivalent sparse and dense reaction structures for the fixed parameter set contain 6 and 13 reactions, respectively.

Now, let us allow $p_1,\ldots,p_6$ to be in the interval $(0,3]$. By minimizing the objective function (10), we obtain the reaction structure shown in Fig. 2. It is apparent that this CRN realization contains only 4 reactions. The corresponding parameter values are $p_1 = 3,\ p_2 = 1,\ p_3 = 2.5,\ p_4 = 1.5,\ p_5 = 1,\ p_6 = 1.5$. Therefore, we found a parameter combination in a set of models that allows us to realize the original kinetic model with the minimal number of reactions. In other words, (11) cannot be described by less than 4 reactions if the model parameters are in the above defined interval. It is interesting to mention, that it is mathematically proved in [19] that the structure shown in Fig. 2 is the smallest chemical reaction system with bistability, which coincides with our result. It is also important to remark that the suggested optimization method is suitable for handling significantly larger networks than this example (possibly consisting of several hundred complexes).

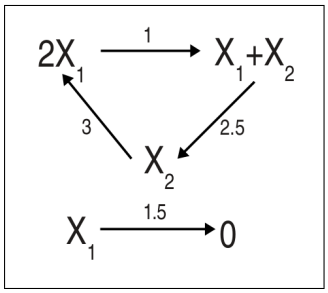

FIGURE 2. The sparsest reaction structure realizing (11) using intervals for $p_1,\ldots,p_6$

ACKNOWLEDGMENTS

This research has been supported by the Hungarian National Research Fund through grants no. K83440 and NF104706. The first and second authors were also supported by the projects TÁMOP-4.2.1.B-11/2/KMR-2011-0002 and TÁMOP-4.2.2/B-10/1-2010-0014.

REFERENCES

1. F. Horn, and R. Jackson, *Archive for Rational Mechanics and Analysis* **47**, 81–116 (1972).
2. J. Haag, A. Wouver, and P. Bogaerts, *Mathematical Biosciences* **193**, 25–49 (2005).
3. I. R. Epstein, and J. A. Pojman, *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves,Patterns and Chaos (Topics in Physical Chemistry)*, Oxford University Press, 1998.
4. 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.
5. E. Sontag, *IEEE Transactions on Automatic Control* **46**, 1028–1047 (2001).
6. D. Angeli, *European Journal of Control* **15**, 398–406 (2009).
7. V. Chellaboina, S. P. Bhat, W. M. Haddad, and D. S. Bernstein, *IEEE Control Systems Magazine* **29**, 60–78 (2009).
8. F. J. Krambeck, *Archive for Rational Mechanics and Analysis* **38**, 317–347 (1970).
9. G. Craciun, and C. Pantea, *Journal of Mathematical Chemistry* **44**, 244–259 (2008).
10. G. Szederkényi, J. R. Banga, and A. A. Alonso, *BMC Systems Biology* **5**, 177 (2011).
11. G. Szederkényi, *Journal of Mathematical Chemistry* **47**, 551–568 (2010), URL `http://www.springerlink.com`.
12. G. Szederkényi, and K. M. Hangos, *Journal of Mathematical Chemistry* **49**, 1163–1179 (2011).
13. G. Szederkényi, K. M. Hangos, and T. Péni, *MATCH Commun. Math. Comput. Chem.* **65**, 309–332 (2011).
14. G. Szederkényi, K. M. Hangos, and Z. Tuza, *MATCH Commun. Math. Comput. Chem.* **67**, 193–212 (2012), URL `http://arxiv.org/abs/1103.4741`.
15. M. D. Johnston, D. Siegel, and G. Szederkényi, *Journal of Mathematical Chemistry* **50**, 274–288 (2012).
16. F. Llaneras, and J. Picó, *Journal of Theoretical Biology* **246**, 290–308 (2007).
17. M. Feinberg, *Lectures on chemical reaction networks*, Notes of lectures given at the Mathematics Research Center, University of Wisconsin, 1979.
18. A. Bemporad, and M. Morari, *Automatica* **35**, 407–427 (1999).
19. T. Wilhelm, *BMC Systems Biology* **3**, 90 (2009).
20. G. Szederkényi, "Computing reaction kinetic realizations of positive nonlinear systems using mixed integer programming," in *8th IFAC Symposium on Nonlinear Control Systems - NOLCOS 2010, Bologna, Italy, 1-3 September*, 2010, p. ThP04.2.
21. V. Hárs, and J. Tóth, "On the inverse problem of reaction kinetics," in *Qualitative Theory of Differential Equations*, edited by M. Farkas, and L. Hatvani, North-Holland, Amsterdam, 1981, vol. 30 of *Coll. Math. Soc. J. Bolyai*, pp. 363–379.