Reconstructing Chemical Reaction Networks by Solving Boolean Polynomial Systems

Extended Abstract

Chenqi Mou and Wei Niu

1. Introduction

Graphs are useful to model chemical reaction networks to reveal their structural properties [3]. In chemical reaction networks, there are species and reactions where the chemical reactions take species as the input and output. Accordingly, there are two kinds of graphs to represent the structures of the species and reactions respectively in a chemical reaction network, namely the S-graph and R-graph [2, 11]. These two kinds of graphs have been successfully applied to study conditions of multistability or oscillations [10], to compare metabolic network [11], to explore large chemical networks [8], and etc.

An S-graph is a directed graph that consists of all the species in a chemical reaction network as vertexes and edges defined in the following way: there is an edge from one species S_i to another S_j if there exists some reaction which outputs S_j from S_i. Similarly, the directed graph R-graph consists of all the reactions occurred in a chemical reaction network and edges defined in the following way: there is an edge from one reaction R_i to another R_j if there exists some species which is the output of R_i and the input of R_j.

There exist tools for analyzing the S- and R-graphs, but both of the graphs contain only partial information of the network, and thus it is desired to have one graph which includes the species, reactions, and their interrelationship. Indeed, we have the SR-graph to capture all the information: the species and reactions are regarded as two kinds of vertexes and there exists an edge from a species (reps. reaction) to a reaction (resp. species) if the species is the input (reps. output) of the reaction [9].

Clearly SR-graphs can be decomposed into S- and R-graphs, and on the other hand, one may need to reconstruct the SR-graphs from the given S- and R-graphs. The existence of such SR-graphs, called the Compound-Reaction-Reconstruction (short as CRR hereafter) problem, has been studied in [4]. In that paper the CRR problem is proved to be NP-complete, and it is formulated into a Boolean satisfiability problem and further solved for theoretical and practical instances with existing solvers for satisfiability, satisfiability modulo theory, and integer linear programming.

In this abstract, the CRR problem is generalized to that of reconstructing all the possible SR-graphs which are composed of given S- and R-graphs. The solution of this problem can reveal more information of the potential SR-graphs involved. We show how to reduce this problem to solving structured Boolean polynomial systems and perform preliminary experiments.

M. Kosta and T. Sturm, eds. Proceedings of the MACIS 2013, Nanning, China, December 11-13, 2013.

2. Problem statement

For an SR-graph of m species $S_1, \ldots, S_m$ and n reactions $R_1, \ldots, R_n$, we denote the sets of input and output species of some reaction R_k by $I(R_k)$ and $O(R_k)$. Then one sees that any SR-graph can be characterized by the following pair of Boolean matrices $\mathbf{E}$ and $\mathbf{P}$ of sizes $m \times n$ and $n \times m$ respectively, where the entries $\mathbf{E}_{i,k}$ and $\mathbf{P}_{k,j}$ of the matrices are defined as

$$\mathbf{E}_{i,k} := \begin{cases} 1, & S_i \in I(R_k) \\ 0, & \text{Otherwise} \end{cases}, \quad \mathbf{P}_{k,j} := \begin{cases} 1, & S_j \in O(R_k) \\ 0, & \text{Otherwise} \end{cases} \quad (1 \leq i, j \leq m, 1 \leq k \leq n).$$

This is because all the input and output species of any reaction in the considered chemical reaction network are recorded in the two matrices.

According to our previous description of the S- and R-graphs, the corresponding S- and R-graphs of the chemical reaction network represented by the SR-graph above can be formulated as $m \times m$ and $n \times n$ Boolean matrices $\mathbf{S}$ and $\mathbf{R}$, where the entries

$$\mathbf{S}_{i,j} := \begin{cases} 1, & \exists R_k \text{ s.t. } S_i \in I(R_k) \text{ and } S_j \in O(R_k) \\ 0, & \text{Otherwise} \end{cases} \quad (1 \leq i, j \leq m, 1 \leq k \leq n),$$

$$\mathbf{R}_{k,l} := \begin{cases} 1, & \exists S_i \text{ s.t. } S_i \in O(R_k) \text{ and } S_i \in I(R_k) \\ 0, & \text{Otherwise} \end{cases} \quad (1 \leq k, l \leq n, 1 \leq i \leq m).$$

For a chemical reaction network, its SR-graph, S-graph and R-graph are interrelated. Their relationship, reflected in the matrices $\mathbf{E}$, $\mathbf{P}$, $\mathbf{S}$, and $\mathbf{R}$ with Boolean operators $\vee$ and $\wedge$, is as follows [4]:

$$\mathbf{S}_{i,j} = \bigwedge_{k=1,\ldots,n} (\mathbf{E}_{i,k} \vee \mathbf{P}_{k,j}), \qquad \mathbf{R}_{k,l} = \bigwedge_{i=1,\ldots,m} (\mathbf{P}_{k,i} \vee \mathbf{E}_{i,l}). \tag{2.1}$$

The problem of our interest is to reconstruct the SR-graph of a chemical reaction network from the S- and R-graphs, or in the matrix language we defined above, to compute $\mathbf{E}$ and $\mathbf{P}$ from $\mathbf{S}$ and $\mathbf{R}$ such that the equations (2.1) hold. The existence of such $\mathbf{E}$ and $\mathbf{P}$ is considered in the CRR problem, studied in [4].

In this abstract, we extend the CRR problem to compute all the possible matrices $\mathbf{E}$ and $\mathbf{P}$ if they exist. This problem is to reconstruct all the possible chemical reaction networks from given species and reactions with the relationship in both species and reactions known. Clearly this problem is stronger than the CRR problem, and we call it the CRR$^+$ problem.

CRR$^+$($\mathbf{S}, \mathbf{R}$) problem: Given two Boolean matrices $\mathbf{S}$ and $\mathbf{R}$ of sizes $m \times m$ and $n \times n$ respectively, compute all the pairs of Boolean matrices $\mathbf{E}$ and $\mathbf{P}$ of sizes $m \times n$ and $n \times m$ such that the equations (2.1) hold.

The CRR$^+$ problem is of interest because once it is solved, all the possible SR-graphs will be known, and this may lead further refinement. For example, these SR-graphs may be tested with chemical experiments to find potential chemical reactions. Obviously this step is not feasible with merely the solution of the CRR problem.

3. Reduction to polynomial system solving

Suppose that $\mathbf{S}$ and $\mathbf{R}$ are two known Boolean matrices of sizes $m \times m$ and $n \times n$, and $\mathbf{E}$ and $\mathbf{P}$ are unknown Boolean matrices of sizes $m \times n$ and $n \times m$. Now consider the Boolean polynomial ring $\mathbb{F}_2[\mathbf{x}] := \mathbb{F}_2[\mathbf{E}_{1,1}, \ldots, \mathbf{E}_{m,n}, \mathbf{P}_{1,1}, \ldots, \mathbf{P}_{n,m}]$ with all the $2mn$ entries $\mathbf{E}_{i,j}$ and $\mathbf{P}_{k,l}$ $(1 \leq i, l \leq m, 1 \leq j, k \leq n)$ of $\mathbf{E}$ and $\mathbf{P}$ as variables. By using the rules $x \wedge y = x \cdot y$ and $x \vee y = x + y + x \cdot y$, where in the right hands the operators $+$ and $\cdot$ are respectively the addition and multiplication in $\mathbb{F}_2[x, y]$, we can rewrite the equations (2.1) in the form of Boolean algebra to polynomial equations in $\mathbb{F}_2[\mathbf{x}]$.

TABLE 1. Timings for solving $F = 0$

m, n	P	Density	#Var	$\#F$	Time	#Solutions
8	0.9	3.13/15.63	128	940	0.27	0
8	0.9	9.38/9.38	128	940	36.77	0
8	0.9	3.12/9.38	128	968	>1000	unknown
9	0.9	11.11/6.17	162	1346	8.25	0
9	0.9	12.35/6.17	162	1338	0.62	0
9	0.9	9.88/8.64	162	1338	>1000	unknown
10	0.9	10/8	200	1838	1.21	0
10	0.9	9/12	200	1811	1.17	0
11	0.9	14.05/10.74	242	2362	2.17	0
5	0.95	8/8	50	234	0.06	296
5	0.95	4/8	50	238	0.70	7759

Furthermore, if $\mathbf{S}_{i,j} = 1$ then according to the equations (2.1) and the above rewriting rules, we will have a polynomial of degree $2n$ in $2n$ variables; otherwise

$$\mathbf{S}_{i,j} = \bigwedge_{k=1,\ldots,n} (\mathbf{E}_{i,k} \vee \mathbf{P}_{k,j}) = 0,$$

and thus we have n quadratic equations $\mathbf{E}_{i,k} \cdot \mathbf{P}_{k,j} = 0$ $(k = 1, \ldots, n)$. The case for $\mathbf{R}_{i,j}$ can be analyzed in a similar way. As one can expect, the structure of the derived Boolean polynomial equation set is dependent on the number of zeros in the matrices $\mathbf{S}$ and $\mathbf{R}$, namely their sparsity.

Let p and q $(0 \leq p \leq m^2, 0 \leq q \leq n^2)$ be the numbers of zeros in the matrices $\mathbf{S}$ and $\mathbf{R}$ respectively, and $F \subseteq \mathbb{F}_2[\mathbf{x}]$ such that $F = 0$ is the Boolean polynomial equation set derived from $\mathbf{S}$ and $\mathbf{R}$ with the equations (2.1). We call an equation $f \in F$ of type s (resp. r) if it is derived from (2.1) with some $\mathbf{S}_{i,j}$ (resp. $\mathbf{R}_{i,j}$) equal to 1; and otherwise it is said to be of type 0. Then it is easy to know that F consists of $np + mq$ equations of type 0, $m^2 - p$ ones of type s, and $n^2 - q$ ones of type r. The total number of equations is $m^2 + n^2 + (n-1)p + (m-1)q \geq 2mn$, and thus $F = 0$ is overdefined (number of variables being $2mn$). In particular, the sparser the matrices $\mathbf{S}$ and $\mathbf{R}$ are, the more overdefined $F = 0$ will be.

To solve $F = 0$ for $F \subseteq \mathbb{F}_2[\mathbf{x}]$, one may choose to use tools like Gröbner bases [1] and triangular sets [12]. In particular, the F_4 algorithm for computing Gröbner bases is well acknowledged for its efficiency [5], and there also exist specific algorithms for computing Boolean triangular decomposition, e.g., [7].

4. Experimental results

In fact in practical chemical reaction networks, the matrices $\mathbf{S}$ and $\mathbf{R}$ for the S- and R-graphs are sparse. For example, the percentages of zeros in the above matrices in a real-world instance in [4, Section 5.2] are all above 96%. Hence in our preliminary experiments the matrices $\mathbf{S}$ and $\mathbf{R}$ are set sparse. In addition, we choose $m{=}n$ and construct the Boolean matrices $\mathbf{S}$ and $\mathbf{R}$ randomly with the probability P for each of their entries to be 0.

In Table 1 we provide the timings (in seconds) for selected instances in our experiments, which are performed using F_4 algorithm in Magma 2.17-1 under Scientific Linux OS release 5.5 on 8 Intel(R) Xeon(R) CPUs E5420 at 2.50 GHz with 20.55G RAM. The columns "Density", "#Var", "$\#F$", "Time", and "#Solutions" in the table mean respectively the actual percentage of nonzero entries in the random matrices $\mathbf{S}$ and $\mathbf{R}$, the number of variables, the number of polynomials in F, the time to compute the Gröbner basis, and the number of solutions of $F = 0$.

In our experiments with random matrices, we find that most polynomial systems $F = 0$ have no solution. This is consistent with the experimental results in [4] on the existence of $\mathbf{E}$ and $\mathbf{P}$ in the CRR problem chosen randomly. This could be due to the fact that when $\mathbf{S}$ and $\mathbf{R}$ are sparse, F is

much overdefined. But we do find some random instances when $F = 0$ has solutions, and in this case the number of solutions turns out to be large, which means that there exist many SR-graphs having the same S- and R-graphs.

As regards for the efficiency, based on our preliminary experiments and those in [4, Section 5.1], we have to admit that without optimization, experimentally the method in this abstract is not comparable to the method by solving Boolean satisfiability problems in [4]. This may be because that we are indeed trying to solve the CRR^+ one, more difficult than the CRR problem, and that the polynomial system F has not been simplified based on its structure. Deeper analyses are still on-going work.

5. Concluding remarks

In this extended abstract we study the problem of reconstructing all possible SR-graphs for chemical reaction networks with given S- and R-graphs by reducing it to solving Boolean polynomial systems. Since the constructed polynomial set is quite specially structured, one can expect some further optimizations to simplify it, so that the computational efficiency is improved. In addition, specific complexity analyses on such structured polynomial systems may be done (like that in [6]) to gain a better understanding on solving the CRR^+ problem with polynomial systems.

The authors would like to thank the reviewers for their helpful comments and suggestion.

References

[1] B. Buchberger. *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal*. PhD thesis, Universität Innsbruck, Austria, 1965.

[2] G. Craciun and M. Feinberg. Multiple equilibria in complex chemical reaction networks: II. the species-reaction graph. *SIAM J. Appl. Math.*, 66(4):1321–1338, 2006.

[3] M. Domijan and M. Kirkilionis. Graph theory and qualitative analysis of reaction networks. *Netw. Heterog. Media*, 3(2):295–322, 2008.

[4] R. Fagerberg, C. Flamm, D. Merkle, P. Peters, and P.F. Stadler. On the complexity of reconstructing chemical reaction networks. *Math. Comput. Sci.*, 7(3):275–292, 2013.

[5] J.-C. Faugère. A new efficient algorithm for computing Gröbner bases (F_4). *J. Pure Appl. Algebra*, 139(1–3):61–88, 1999.

[6] J.-C. Faugère, M. Safey El Din, and P.-J. Spaenlehauer. Gröbner bases of bihomogeneous ideals generated by polynomials of bidegree (1,1): Algorithms and complexity. *J. Symb. Comput.*, 46(4):406–437, 2011.

[7] X.-S. Gao and Z. Huang. Characteristic set algorithms for equation solving in finite fields. *J. Symb. Comput.*, 47(6):655–679, 2012.

[8] S. Klamt, U.-U. Haus, and F. Theis. Hypergraphs and cellular networks. *PLoS Comput. Biol.*, 5(5):e1000385, 2009.

[9] M. Mincheva and M.R. Roussel. Graph-theoretic methods for the analysis of chemical and biochemical networks. I. multistability and oscillations in ordinary differential equation models. *J. Math. Biol.*, 55(1):61–86, 2007.

[10] C. Soulé. Graphic requirements for multistationarity. *ComPlexUs*, 1:123–133, 2003.

[11] A. Wagner and D.A. Fell. The small world inside large metabolic networks. *Proc. R. Soc. London, Ser. B: Biol. Sci.*, 268(1478):1803–1810, 2001.

[12] D. Wang. *Elimination Methods*. Springer, Wien, New York, 2001.

Chenqi Mou
LMIB–School of Mathematics and System Sciences
Beihang University, Beijing 100191, China
e-mail: chenqi.mou@buaa.edu.cn

Wei Niu
École Centrale Pékin
Beihang University, Beijing 100191, China
e-mail: wei.niu@buaa.edu.cn