

## Research Article

# Reachability Analysis of Low-Order Discrete State Reaction Networks Obeying Conservation Laws

Gergely Szlobodnyik <sup>1</sup> and Gábor Szederkényi<sup>1,2</sup>

<sup>1</sup>Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Práter u. 50/a, H-1083 Budapest, Hungary

<sup>2</sup>Systems and Control Laboratory, Institute for Computer Science and Control (MTA SZTAKI) of the Hungarian Academy of Sciences, Kende u. 13-17, H-1111 Budapest, Hungary

Correspondence should be addressed to Gergely Szlobodnyik; szlger91@gmail.com

Received 5 October 2018; Revised 1 February 2019; Accepted 25 February 2019; Published 26 March 2019

Academic Editor: Yan-Ling Wei

Copyright © 2019 Gergely Szlobodnyik and Gábor Szederkényi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper we study the reachability problem of sub- and superconservative discrete state chemical reaction networks (d-CRNs). It is known that a subconservative network has bounded reachable state space, while that of a superconservative one is unbounded. The reachability problem of superconservative reaction networks is traced back to the reachability of subconservative ones. We consider network structures composed of reactions having at most one input and one output species beyond the possible catalyzers. We give a proof that, assuming all the reactions are charged in the initial and target states, the reachability problems of sub- and superconservative reaction networks are equivalent to the existence of nonnegative integer solution of the corresponding d-CRN state equations. Using this result, the reachability problem is reformulated as an Integer Linear Programming (ILP) feasibility problem. Therefore, the number of feasible trajectories satisfying the reachability relation can be counted in polynomial time in the number of species and in the distance of initial and target states, assuming fixed number of reactions in the system.

## 1. Introduction

Employing deterministic ordinary differential equation systems to characterize the dynamical behavior of complex networks of chemically interacting components (species) is a widely used approach in mathematical and computational systems biology [1–3]. Such a continuous state modeling approach assumes high molecular count of species and their homogeneous (well-mixed) distribution in the surrounding media [4]. However, in several (bio)chemically interesting systems, such as some enzymatic and genetic networks, the molecular count of different species is relatively low (e.g., < 100 molecules) [4–6] implying that the assumption of homogeneous species distribution does not hold [7, 8]. Hence it is necessary to introduce a discrete state model capable of keeping track of the individual molecular counts in order to properly characterize the qualitative dynamical behavior of (bio)chemical networks of species with low number of

molecules [9, 10]. There exist several mathematical models describing the state evolution of discrete state chemical reactions networks, such as Markov chain models [8, 10] and stochastic Petri nets [11].

In the context of chemical reaction networks of several interacting components, in order to completely characterize the system it is needed to simultaneously study the dynamical behavior and the underlying network structure as well. Moreover, it is also important to examine how the dynamical behavior and the network structure are related to each other, and how we can predict the dynamical behavior (e.g., in the form of possible state space trajectories) and be aware of the underlying network structure. For continuous state reaction networks obeying the law of mass action, it is recognized that the network structure (i.e., topology) is not necessarily unique; i.e., the same system of differential equations can be generated by different network topologies (different sets of interactions among the given species) [12–15].

TABLE 1: Notations.

$\mathbb{R}$	the set of real numbers
$\mathbb{Z}$	the set of integer numbers
$\mathbb{Z}_{\geq 0}$	the set of non-negative integer numbers
$\mathbb{T}^{n \times m}$	the set of $(n \times m)$ -dimensional vectors over the set $\mathbb{T}$
$0^{n \times m}$	a zero matrix of dimension $n \times m$
$1^{n \times m}$	a matrix of dimension $n \times m$ for which all the entries are equal to 1
$\{0, 1\}^{n \times m}$	the set of $(n \times m)$ -dimensional binary vectors (all the entries are equal to 0 or 1)
$\{-1, 0, 1\}^{n \times m}$	the set of $(n \times m)$ -dimensional vectors composed of the entries $-1, 0, 1$
$[A]_{i,:}$	the $i$ th row of the matrix $A$
$a < b$	for $a, b \in \mathbb{R}^n$ , $a_i < b_i$ for $i = 1, \dots, n$
$a \leq b$	for $a, b \in \mathbb{R}^n$ , $a_i \leq b_i$ for $i = 1, \dots, n$
$\sigma_X$	an ordered sequence of states
$\sigma_r$	an ordered sequence of reaction vectors
$\sigma_S$	an ordered sequence of species
$\sigma_C$	an ordered sequence of complexes

In the case of discrete state reaction networks the so-called reachability is a strictly related problem to the dynamical behavior; namely, is it possible to reach a prescribed target state from a given initial one through a finite sequence of transition (reactions)? It is known that the reachability relation between any pair of nonnegative initial and target states is determined by the network structure itself. Through the reachability analysis several problems of great importance can be analyzed; one of them having high interest is the existence of so-called extinction events: the existence of trajectories resulting in the irreversible extinction of some species from the system. It has been shown that under some conditions on the network structure a discrete state chemical reaction network exhibits an extinction event from any point of its state space [9, 16, 17]. The properties of recurrence (the ability of returning to any initial state) and irreducibility (the ability of reaching any state from any other one) are also examined in the context of discrete state reaction networks [18, 19].

The mathematical model of discrete state chemical reaction networks is equivalent to an important model of theoretical computer science, namely, the so-called vector addition systems with states (VASS) or equivalently Petri nets [20, 21]. Hence the discrete chemical reaction network reachability problem is equivalent to the extensively studied problem of vector addition system (VAS) reachability. The VAS reachability problem is known to be decidable [22–25], and for the space complexity we have EXSPACE lower bound [26]. Unfortunately, contrary to the proven polynomial time complexity of reachability of rate independent continuous state chemical reaction networks [21], in the case of discrete state reaction networks it is not known whether there exists an algorithm of primitive-recursive time complexity deciding this problem [27].

The aim of this paper is to study of the reachability problem of sub- and superconservative d-CRNs. We make use of the relation between the sub- and superconservative properties. In Propositions 15 and 17, we give necessary and sufficient conditions on the network structure and the initial

and target states under which the reachability is equivalent to the nonnegative integer solution of the d-CRN state equation. Then these results in Corollaries 16 and 18 are extended to a subclass of superconservative d-CRNs.

The paper is organized as follows. In Section 2 the necessary mathematical notations and concepts of Chemical Reaction Network Theory (CRNT) are introduced. Section 3 discusses the classes of sub- and superconservative d-CRNs and their duality as well. In Section 4 the reachability problem of sub- and superconservative d-CRNs is examined. Firstly the case of low state space-dimensional d-CRNs is discussed, followed by the extension to the general case when the dimension of the state space is arbitrarily high. In Section 5 our findings are illustrated in a representative example.

## 2. Notations and Mathematical Background

In Table 1 we summarize the notations and concepts of discrete chemical reaction networks which will be extensively used later.

*2.1. Discrete State Chemical Reaction Networks.* A discrete state Chemical Reaction Network (d-CRN) with  $n$  species,  $m$  complexes, and  $l$  reactions is a triple  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  so that

$$\begin{aligned}
 \mathcal{S} &= \{s_i \mid i = 1, \dots, n\} \\
 \mathcal{C} &= \left\{ y_j = \sum_{i=1}^n \alpha_{ji} s_i \mid s_i \in \mathcal{S}, \alpha_{ji} \in \mathbb{Z}_{\geq 0}, i = 1, \dots, n, j \right. \\
 &\quad \left. = 1, \dots, m \right\} \\
 \mathcal{R} &= \left\{ r_v = y_{source(r_v)} \right. \\
 &\quad \left. \longrightarrow y_{product(r_v)} \mid y_{source(r_v)}, y_{product(r_v)} \in \mathcal{C}, v \right. \\
 &\quad \left. = 1, \dots, l \right\}
 \end{aligned} \tag{1}$$

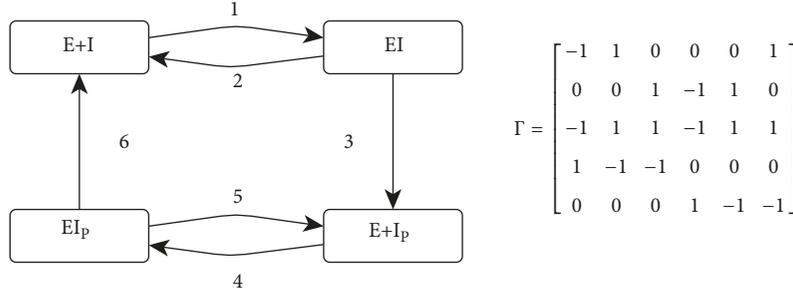


FIGURE 1: A discrete state chemical reaction network. *Left*: reaction network structure. The nodes and directed edges represent the complexes and the reactions, respectively. The numbers on the edges denote a fixed ordering of the reactions. *Right*: the stoichiometric matrix associated with the system, i.e.  $[\Gamma]_{ij}$  is the net change in the number of the  $i$ 'th species upon occurring the  $j$ 'th reaction.

where  $s_i$  is the  $i$ 'th species,  $y_j$  is the  $j$ 'th complex, and  $r_v$  is the  $v$ 'th reaction of the network. Moreover,  $\alpha_{ji}$  is the stoichiometric coefficient of the  $i$ 'th species in the  $j$ 'th complex. For a reaction  $r_v = y_{source(r_v)} \rightarrow y_{product(r_v)}$  of  $\mathcal{R}$ ,  $y_{source(r_v)}$  and  $y_{product(r_v)}$  are the source complex and the product complex, respectively.

For each complex  $y_j \in \mathcal{C}$ ,  $j \in \{1, \dots, m\}$ , the stoichiometric coefficients of the species can be represented as a vector of the following form:

$$\bar{y}_j = [\alpha_{j1} \ \alpha_{j2} \ \dots \ \alpha_{jn}]^T \quad (2)$$

For each  $r \in \mathcal{R}$ , a reaction vector  $r_{ij} \in \mathbb{Z}^n$  can be associated with the track of the net molecular count changes of the species upon firing the reaction:

$$r_{ij} = \bar{y}_j - \bar{y}_i \quad (3)$$

so that  $y_j$  and  $y_i$  are the corresponding source and product complexes of  $r$ . In the sequel the notation  $r_i$  will be used for denoting both the  $i$ 'th reaction of the d-CRN and the associated reaction vector as well. We will also assume that for all the examined d-CRNs a fixed order of the reaction vectors is given; i.e., an order  $r_1, r_2, \dots, r_l$  is fixed and  $l = |\mathcal{R}|$ .

A d-CRN can also be represented by a directed graph  $G = G(V, E)$  such that the vertices and edges correspond to the complexes and the reactions, respectively, i.e.,

$$V = \mathcal{C} \quad (4)$$

$$E = \mathcal{R}. \quad (5)$$

The direction of the edges is determined by the reactions of  $\mathcal{R}$ , so that if  $y_i \rightarrow y_j \in \mathcal{R}$ , then there exists an edge  $e \in E$  from the vertex representing  $y_i$  to the vertex of  $y_j$ . For each edge a weight corresponding to the reaction rate constant (also called intensity or propensity) corresponding to the respective reaction can also be associated.

Beyond the above representations it is also possible to describe a d-CRN in an algebraic way by means of its unique stoichiometric matrix.

*Definition 1.* Let us consider a d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ . The stoichiometric matrix  $\Gamma \in \mathbb{Z}^{n \times l}$  of  $\mathcal{N}$  is defined as follows:

$$\Gamma = [r_1 \ \dots \ r_l] \quad (6)$$

Note that  $[\Gamma]_{ij}$  encodes the net molecule count change on species  $s_i$  upon the occurrence of reaction  $r_j$ . Besides  $\Gamma$  we also define the following matrices:

$$\Gamma^+ = [\bar{y}_{r_1}^+ \ \dots \ \bar{y}_{r_l}^+]^T \quad (7)$$

$$\Gamma^- = [\bar{y}_{r_1}^- \ \dots \ \bar{y}_{r_l}^-]^T \quad (8)$$

where  $\bar{y}_{r_i}^+$  denotes the vector form of the product complex belonging to reaction  $r_i$  while  $\bar{y}_{r_j}^-$  represents the vector of the source complex associated with reaction  $r_j$ . The relation among the above defined matrices is as follows:

$$\Gamma = \Gamma^+ - \Gamma^- \quad (9)$$

*Example 2.* Let us consider the d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  depicted in Figure 1.  $\mathcal{N}$  characterizes a simple network of a bifunctional enzyme  $E$  having both phosphorylation and dephosphorylation activities on species  $I$  and  $I_p$ , respectively. This network is characterized by the following sets.

$$\begin{aligned} \mathcal{S} &= \{I, I_p, E, EI, EI_p\} \\ \mathcal{C} &= \{I + E, EI, I_p + E, EI_p\} \\ \mathcal{R} &= \{E + I \rightarrow EI, IE \rightarrow E + I, EI \rightarrow I_p + E, E \\ &\quad + I_p \rightarrow EI_p, EI_p \rightarrow E + I_p, EI_p \rightarrow E + I\} \end{aligned} \quad (10)$$

We fix the order of species and reactions as they are listed in the above sets.

$\mathcal{N}$  has no information on the probabilities of the reactions, but at any given time instant at most one reaction can occur.

The molecular count of each species of a d-CRN at any time  $t \geq 0$  is given by its state vector  $X(t) \in \mathbb{Z}_{\geq 0}^n$  and the time evolution of the system is characterized by the following discrete state equation:

$$X(t) = X(0) + \Gamma N(t) \quad (11)$$

where  $X(0)$  is the state vector belonging to the initial time instant and  $N(t) = [N_1(t), N_2(t), \dots, N_m(t)]^T \in \mathbb{Z}_{\geq 0}^l$  such

that  $N_k(t) \in \mathbb{Z}_{\geq 0}$  stores the number of occurrences of the  $k$ 'th reaction up to time  $t$ . We note that  $N(t)$  is typically modeled as some point process [8, 10].

For our further analysis the time instants when the reactions have occurred are not of interest, but only the order of reactions; therefore we abandon the notation of time  $t$  in the formulas.

*Definition 3.* Let us consider a d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ . It is said that:

- (1) a species  $s \in \mathcal{S}$  is a *catalyzer* of a reaction  $r \in \mathcal{R}$  if it has the form of  $r = s + s_1 \rightarrow s + s_2$  with  $s_1, s_2 \in \mathcal{S}$ ,
- (2) a complex  $y \in \mathcal{C}$  is *charged* at state  $X \in \mathbb{Z}_{\geq 0}^n$  if  $X \geq y$ ,
- (3) a reaction  $r \in \mathcal{R}$  is charged if its respective source complex is charged,
- (4) a state  $X \in \mathbb{Z}_{\geq 0}^n$  *reacts* to a state  $X' \in \mathbb{Z}_{\geq 0}^n$  (denoted by  $X \rightarrow X'$ ) if there exists a reaction  $r \in \mathcal{R}$  such that  $r$  is charged at state  $X$  and  $X + r = X'$ ,
- (5) a *reaction (vector) sequence*  $\sigma_r$  is an ordered set of reaction vectors  $\sigma_r = r_1 \dots r_v$  where  $r_i \in \mathcal{R}$ ,  $i = 1, \dots, v$ ,
- (6) a *state transition sequence*  $\sigma_X$  is an ordered set states  $X_0, X_1, \dots, X_p$  so that  $X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_{p-1} \rightarrow X_p$ ,
- (7) a state  $X' \in \mathbb{Z}_{\geq 0}^n$  is *reachable* from a state  $X \in \mathbb{Z}_{\geq 0}^n$  (denoted by  $X \rightsquigarrow_{\mathcal{N}} X'$ ) if there exists a directed path in the state space so that  $X = X_{\nu(1)} \rightarrow X_{\nu(2)} \rightarrow \dots \rightarrow X_{\nu(\nu)} = X'$ .

Considering a state transition sequence  $\sigma_X = X_0 X_1 \dots X_{p-1} X_p$ , we call  $X_0$  and  $X_p$  the initial and target states, respectively, while  $X_i$  for  $i \in \{1, \dots, p-1\}$  are called transition states of  $\sigma_X$ .

The condition that a reaction  $r \in \mathcal{R}$  is charged at state  $X \in \mathbb{Z}_{\geq 0}^n$  can be expressed by the inequality  $X \geq \bar{y}_r$ . For a reaction sequence  $\sigma_r$  a state transition sequence  $\sigma_X = X_0 X_1 \dots X_p$  can be uniquely associated so that

$$X_j = X_{j-1} + r_j, \quad j \in \{1, \dots, p\} \quad (12)$$

where the initial state  $X_0$  is assumed to be given. A state transition sequence  $\sigma_X$  is said to be *admissible* if  $X_i \geq r_{i+1}$  for  $X_i \in \sigma_X$ ,  $i \in \{0, \dots, p-1\}$ ; moreover, we say that a reaction sequence  $\sigma_r$  is admissible if the corresponding state transition sequence is admissible.

From the reachability of a state  $X' \in \mathbb{Z}_{\geq 0}^n$  from an initial state  $X_0 \in \mathbb{Z}_{\geq 0}^n$ , it follows that the following equation has a nonnegative integer solution  $c \in \mathbb{Z}_{\geq 0}^l$ :

$$X' = X_0 + \Gamma c \quad (13)$$

where  $[c]_i$  encodes the number of occurrences for reaction  $r_i \in \mathcal{R}$  for  $i \in \{1, \dots, l\}$ . However, it is important to note that from the existence of a nonnegative integer solution  $c$  of (13), the reachability relation  $X_0 \rightsquigarrow_{\mathcal{N}} X'$  does not necessarily follow.

We note that  $c$  of (13) corresponds to  $N(t)$  of (11). Since a solution  $c \in \mathbb{Z}_{\geq 0}^l$  of (13) encodes the number of occurrences for each reaction in a fixed order, the following equality is fulfilled:

$$\Gamma c = \sum_{i=1}^h r_i \quad (14)$$

where  $h = \sum_{i=1}^l [c]_i$  and  $r_i \in \mathcal{R}$  for  $i \in \{1, \dots, h\}$ . When we want to emphasize that a reaction vector sequence is encoded by a particular  $c \in \mathbb{Z}_{\geq 0}^l$ , we will use the notation  $\sigma_r^c = r_1, \dots, r_h$  and the state transition sequence determined by  $\sigma_r^c$  will be denoted by  $\sigma_X^c$ .

*Definition 4.* Let us consider a d-CRN  $\mathcal{N}$  with stoichiometric matrix  $\Gamma \in \mathbb{Z}^{n \times l}$  and an initial state  $X_0 \in \mathbb{Z}_{\geq 0}^n$ . The *reachable state space*  $\text{Reach}(\mathcal{N}, X_0)$  of  $\mathcal{N}$  with initial state  $X_0$  is the set of nonnegative discrete states reachable from  $X_0$ .

$$\text{Reach}(\mathcal{N}, X_0) = \{X \mid X \in \mathbb{Z}_{\geq 0}^n, X_0 \rightsquigarrow_{\mathcal{N}} X\} \quad (15)$$

### 3. Integer Linear Programming

In this section some relevant concepts of mathematical programming that will be extensively employed later are briefly reviewed. An Integer Linear Programming (ILP) instance can be formulated as follows:

$$\text{ILP} \begin{cases} \min_x \{a^\top x\} \\ \text{subject to} \\ Ax \leq b \\ x \in \mathbb{Z}^n \end{cases} \quad (16)$$

where  $x$  is the  $n$ -dimensional vector of decision variables while  $a \in \mathbb{Z}^n$ ,  $A \in \mathbb{Z}^{m \times n}$ , and  $b \in \mathbb{Z}^m$  are fixed coefficients. Generally, the above ILP computational problem is known to be NP-hard, which may highly confine our ability to efficiently solve problems of integers in high dimension.

However, if the value of the decision vector that minimizes (or maximizes) the prescribed objective function is not important for us, but only the existence of a  $x \in \mathbb{Z}^n$  vector satisfying the set of specified constraints, then the problem is called ILP feasibility problem.

$$\text{FP} \begin{cases} P = \{x \mid Ax \leq b, A \in \mathbb{Z}^{m \times n}, b \in \mathbb{Z}^m, x \in \mathbb{R}^n\} \\ P \cap \mathbb{Z}^n \stackrel{?}{=} \emptyset \end{cases} \quad (17)$$

An ILP feasibility problem, as a decision problem, addresses the question of whether the polytope  $P$  contains an integer lattice point, formally  $P \cap \mathbb{Z}^n \stackrel{?}{=} \emptyset$ . While a FP is also known to be NP-hard, it has well-decoupled time complexity with respect to the number of variables, the number of constraints, and the maximum of the absolute values of the entries of  $A$  and  $b$ . Therefore, a feasibility problem of the form (17), assuming fixed dimension  $n$ , can be decided in polynomial time in the number of constraints  $m$  and the maximum of

the absolute values of the coefficients  $A$  and  $b$  by means of the Lenstra algorithm [28, 29]. Moreover, the number of integer lattice points in  $P$  can also be numerated in polynomial time in  $m$  and the maximum of the absolute value of the coefficients using Barvinok's integer lattice point counting algorithm [30–33]. We note that for the Barvinok algorithm there exists an effective implementation called LattE [34].

#### 4. Sub- and Superconservative d-CRNs

We define conservativity and subconservativity in the same way as they were introduced, e.g. in [4, 16].

**Definition 5.** A d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  having stoichiometric matrix  $\Gamma \in \mathbb{Z}^{n \times l}$  is called *subconservative* (*superconservative*) if there exists a strictly positive vector  $z \in \mathbb{R}_{>0}^n$  for which  $z^\top \Gamma \leq 0^{1 \times l}$  ( $z^\top \Gamma \geq 0^{1 \times l}$ ) holds. The vector  $z$  is called a *conservation vector*.

An important property related to subconservativity is the strong boundedness which is defined as follows.

**Definition 6.** A d-CRN  $\mathcal{N}$  is said to be *strongly bounded* if, for any  $X_0 \in \mathbb{Z}_{\geq 0}^n$  initial state, the reachable state space  $\text{Reach}(\mathcal{N}, X_0)$  is bounded.

The subconservative property of the reaction network structure is a necessary and sufficient condition of strong boundedness [16, 35].

**Proposition 7** (see [35]). *Let us consider a d-CRN  $\mathcal{N}$ . The following propositions are equivalent:*

- (1)  $\mathcal{N}$  is subconservative,
- (2)  $\mathcal{N}$  is strongly bounded.

As a special case covered by the intersection of sub- and superconservativity, we can define the conservative property as well.

**Definition 8.** Let us consider a d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  with stoichiometric matrix  $\Gamma \in \mathbb{Z}^{n \times l}$ . The d-CRN  $\mathcal{N}$  is said to be *conservative* if there exists a vector  $z \in \mathbb{R}_{>0}^n$  satisfying the matrix equation  $z^\top \Gamma = 0^{1 \times l}$ .

We note that the above structural properties can be easily decided in polynomial time by means of an LP of the following form:

$$\begin{aligned} \min \quad & \sum_{j=1}^n z_j \\ \text{s.t.} \quad & \\ & z^\top \Gamma \leq 0^{1 \times l} \quad (\text{or } z^\top \Gamma \geq 0^{1 \times l}) \\ & z \geq 0^{n \times 1} + \varepsilon^{n \times 1}, \quad \varepsilon > 0^{n \times 1} \end{aligned} \quad (18)$$

The relationship between sub- and superconservativity can be expressed by the following proposition.

**Proposition 9.** *A d-CRN  $\mathcal{N}$  with stoichiometric matrix  $\Gamma_{\mathcal{N}} \in \mathbb{Z}^{n \times l}$  is subconservative if and only if the d-CRN  $\mathcal{N}'$  with stoichiometric matrix  $\Gamma_{\mathcal{N}'} = -\Gamma_{\mathcal{N}}$  is superconservative.*

*Proof.*

$$z^\top \Gamma \leq 0^{1 \times m} \iff z^\top (-\Gamma) \geq 0^{1 \times l} \quad (19)$$

□

We note that  $-\Gamma_{\mathcal{N}}$  means the change of the direction of each reaction in the d-CRN  $\mathcal{N}$  of stoichiometric matrix  $\Gamma_{\mathcal{N}}$ .

**Example 10.** Figure 2 depicts two d-CRNs: a subconservative and a superconservative reaction network structure. Indeed, these networks are counterparts that can be easily transformed to each other by changing the sign of the entries in the stoichiometric matrices. Such a transformation results in the change of the direction of the edges in the reaction network.

From Proposition 9 it follows that, instead of the reachability problem of a superconservative network structure, one can consider an equivalent subconservative d-CRN reachability problem as is discussed in Proposition 11.

**Proposition 11.** *Let us consider a subconservative d-CRN  $\mathcal{N}$  characterized by the matrices  $\Gamma_{\mathcal{N}} = \Gamma$ ,  $\Gamma_{\mathcal{N}}^+ = \Gamma^+$  and a superconservative d-CRN  $\mathcal{N}'$  with matrices  $\Gamma_{\mathcal{N}'} = -\Gamma$ ,  $\Gamma_{\mathcal{N}'}^- = \Gamma^+$ . Let us take an initial state  $X_0 \in \mathbb{Z}_{\geq 0}^n$  and a target state  $X' \in \mathbb{Z}_{\geq 0}^n$ . Then the reachability  $X_0 \rightsquigarrow_{\mathcal{N}} X'$  holds if and only if  $X' \rightsquigarrow_{\mathcal{N}'} X_0$  also holds.*

*Proof.*

- (1)  $X_0 \rightsquigarrow_{\mathcal{N}} X' \implies \exists c \in \mathbb{Z}_{\geq 0}^l$  such that  $X_0 + \Gamma c = X'$  which is equivalent to  $X' + (-\Gamma)c = X_0$ .

From  $X_0 \rightsquigarrow_{\mathcal{N}} X'$  it follows that the solution  $c \in \mathbb{Z}_{\geq 0}^l$  can be decomposed to an admissible reaction vector sequence  $\sigma_r^c = r_1^c \dots r_h^c$ ,  $h = \sum_{i=1}^l [c]_i$ ; i.e., all the states of  $\sigma_X^c$  determined by  $\sigma_r^c$  are composed of nonnegative entries. Then, by reversing  $\sigma_X^c$ , we obtain a nonnegative state transition sequence  $\hat{\sigma}_X^c$  from  $X'$  to  $X_0$  which is uniquely determined by means of the reaction vector sequence  $\hat{\sigma}_r^c = -r_h^c \dots -r_1^c$ .

It is also needed to show that  $\hat{\sigma}_r^c$  is an admissible reaction sequence. This can be done as follows: for each state  $X \in \sigma_X^c \setminus X_0$  there exists a reaction  $r \in \sigma_r^c$  so that upon firing  $r$  the resulting state is  $X$ , from which it follows that  $X \geq \bar{y}_r^+$ ; moreover, considering the reversed reaction sequence  $\hat{\sigma}_r^c$ , the reaction vector that will occur at state  $X$  is  $-r \in \hat{\sigma}_r^c$  which is charged at  $X$  even if  $X \geq \bar{y}_r^+$ .

Then the admissibility of  $\hat{\sigma}_r^c$  follows.

- (2) The proof for the other direction  $X' \rightsquigarrow_{\mathcal{N}'} X_0$  works analogously as above.

□

The importance of Proposition 11 is that the reachability problem of a superconservative d-CRN of unbounded reachable state space can be easily traced back to the reachability

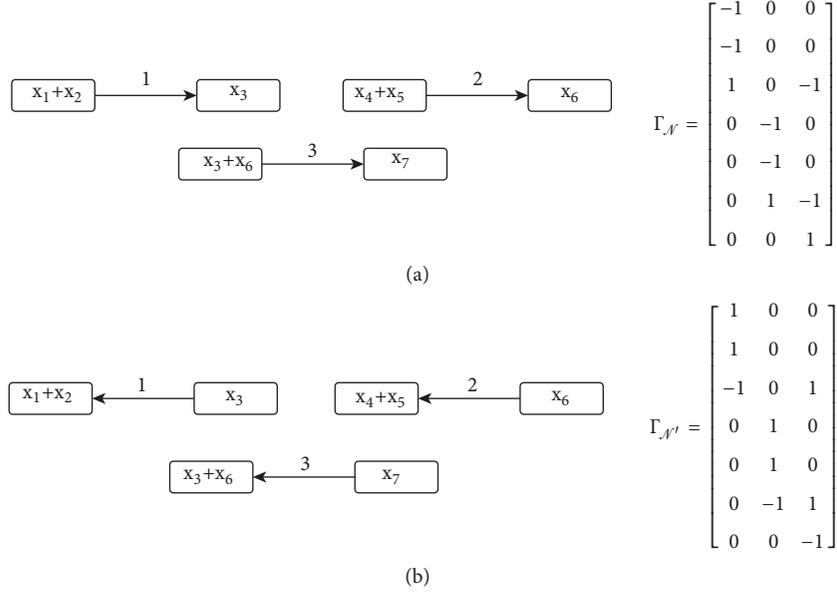


FIGURE 2: A pair of sub- and superconservative reaction network structures denoted by  $\mathcal{N}$  and  $\mathcal{N}'$ , respectively. The ordering of the reactions are denoted by the numbers on the edges of the graphs. The two networks can be transformed to each other by changing the sign of the entries in their stoichiometric matrices. (a) Subconservative d-CRN. (b) Superconservative d-CRN.

problem of a d-CRN of bounded reachable state space which can make the original decision problem computationally tractable.

## 5. Reachability Analysis

**5.1. Low-Dimensional Case.** In this section the case of low-dimensional ( $\text{rank}(\Gamma) \leq 2$ ) subconservative d-CRNs is considered. We state a modified version of Proposition 5 of [36] where the conditions on the initial and target states are less strict. Then we extend the result to superconservative d-CRNs.

In order to discuss low-dimensional reachability problems, we introduce a distinguished state  $M = M(\Gamma^-)$  as follows:

$$[M(\Gamma^-)]_i = \max \{[\Gamma^-]_{ij} : j = 1, \dots, l\} \quad i = 1, \dots, n. \quad (20)$$

Here  $\Gamma^-$  is defined by (8). Note that the set  $\{X \mid X \in \mathbb{Z}_{\geq 0}^n, X \geq M\}$  contains all the states where each reaction is charged.

**Proposition 12.** *Let us consider a subconservative d-CRN  $\mathcal{N}$  with stoichiometric matrix  $\Gamma \in \{-1, 0, 1\}^{n \times l}$  and  $\Gamma^- \in \{0, 1\}^{n \times l}$ . Assume that  $\text{rank}(\Gamma) \leq 2$ . We consider an initial state  $X_0 \in \mathbb{Z}_{\geq 0}^n$  and a target state  $X' \in \mathbb{Z}_{\geq 0}^n$  such that  $X_0 \geq M$  and  $X' \geq M$  hold where  $M = M(\Gamma^-)$  is defined by (20). Then the state  $X'$  is reachable from  $X_0$  through a state transition sequence  $\sigma_X = X_0 X_1 \dots X'$  for which,  $\forall X \in \sigma_X, X \geq M$  if and only if the equation*

$$\Gamma c = X' - X_0 \quad (21)$$

has a nonnegative integer solution  $c$ .

*Proof.*

- (1) If  $X'$  is reachable from  $X_0$  through an admissible state transition sequence  $\sigma_X$ , then it follows that a solution  $c \in \mathbb{Z}_{\geq 0}^l$  exists.
- (2) Assume that there exists  $c \in \mathbb{Z}_{\geq 0}^l$  such that  $X_0 + \Gamma c = X'$  holds. Let us consider any reaction vector decomposition  $\sigma_r = r_{v(1)} \dots r_{v(h)}$  of  $c$  where  $\sum_{j=1}^h r_{v(j)} = c$  and  $\sum_{j=1}^l [c]_j = h$ . We show that Algorithm 1 returns a permutation of  $\sigma_r$ , so that for all the transition states  $X$  the inequality  $X \geq M$  holds.

Let us assume that there exists a transition state  $X_i, X_i \geq M$ , so that the forthcoming state  $X_{i+1}$  satisfies the inequality  $[X_{i+1}]_d < [M]_d$  for some  $d \in \{1, 2\}$ . For the target state  $X'$  to be reached the inequality  $X' \geq M$  holds; hence there exists a reaction increasing the state variable along the coordinate  $d$ . Let us assume that all the reactions increasing the state variable along  $X_i$  decrease the other coordinate  $d'$  so that the resulting forthcoming state  $X_{i+1}$  satisfies the inequality  $[X_{i+1}]_{d'} < [M]_{d'}$ . Then  $X_i = M$  holds. Now there are two different cases:

(P<sub>1</sub>) If  $X' = M$ , then Algorithm 1 terminates, and the correctness follows.

(P<sub>2</sub>) If  $X' \neq M$ , then the subconservativity of  $\mathcal{N}$  implies that it is not possible to reach a state  $X, X \geq M, X \neq M$ ; i.e.,  $X'$  is not reachable from  $X_i$ . This is contradiction, since arbitrary permutation of the initial ordering  $\sigma_r$  results in the same target state  $X'$ , given the initial state  $X_0$ . Then the correctness of Algorithm 1 follows.

□

```

1: procedure REORDER( $X_0 [r_{\nu(1)} \ r_{\nu(2)} \ \dots \ r_{\nu(h)}], M$ )
2:    $X_{current} \leftarrow X_0$ 
3:   for  $i = 1$  to  $h$  do
4:     if  $X_{current} = X'$  then
5:       return  $[r_{\nu(1)} \ r_{\nu(2)} \ \dots \ r_{\nu(h)}]$ 
6:     end if
7:     if  $[X_{current} + r_{\nu(i)}]_l < [M]_l$  for some  $l \in \{1, \dots, n\}$  then
8:       Choose a transition vector  $r_{\nu(j)}$ ,  $i < j \leq h$  so that
9:          $X_{current} + r_{\nu(j)} \geq M$ 
10:         $r' \leftarrow r_{\nu(i)}$ 
11:         $r_{\nu(i)} \leftarrow r_{\nu(j)}$ 
12:         $r_{\nu(j)} \leftarrow r'$ 
13:      end if
14:       $X_{current} \leftarrow X_{current} + r_{\nu(i)}$ 
15:    end for
16:  return  $[r_{\nu(1)} \ r_{\nu(2)} \ \dots \ r_{\nu(h)}]$ 
17: end procedure

```

ALGORITHM 1

Algorithm 1 can be easily extended to the class of superconservative reaction networks.

**Corollary 13.** *Let us consider a superconservative  $d$ -CRN  $\mathcal{N}$  with stoichiometric matrix  $\Gamma \in \{-1, 0, 1\}^{n \times l}$  and  $\Gamma^- \in 0, 1^{n \times l}$ . Assume that  $\text{rank}(\Gamma) \leq 2$  holds and consider an initial state  $X_0 \in \mathbb{Z}_{\geq 0}^n$  and a target state  $X' \in \mathbb{Z}_{\geq 0}^n$  for which  $X_0 \geq M$  and  $X' \geq M$  hold where  $M$  is defined by (20). Then the state  $X' \in \mathbb{Z}_{\geq 0}^n$  is reachable from  $X_0$  if and only if the equation*

$$\Gamma c = X' - X_0 \quad (22)$$

has a nonnegative integer solution  $c$ .

*Proof.* According to Proposition 11 we can consider a subconservative  $d$ -CRN  $\mathcal{N}'$  of stoichiometric matrix  $-\Gamma$  and take the reachability problem  $X' \rightsquigarrow_{\mathcal{N}'} X_0$ . Then Proposition 7 can be applied.  $\square$

**5.2. Sub- and Superconservative  $d$ -CRNs of Arbitrary High State Space Dimension.** In this section the reachability problem of arbitrary high-dimensional sub- and superconservative  $d$ -CRNs is considered. Firstly we examine network structures composed of reactions having at most one input and one output species. It is shown by an inductive proof that, under some auxiliary condition, the reachability relation  $X_0 \rightsquigarrow_{\mathcal{N}} X'$  is equivalent to the existence of a  $c \in \mathbb{Z}_{\geq 0}^l$  solution of the  $d$ -CRN state equation  $X_0 + \Gamma c = X'$ . Then, according to the relation between sub- and superconservative reaction network structures, this result is generalized to a subclass of superconservative  $d$ -CRNs as well. We also extend the results to  $d$ -CRNs containing second-order reactions by allowing catalyzer species.

Firstly, we adopt the following necessary and sufficient condition of reachability from the theory of Petri nets (see Theorem 16, [37]) which will be extensively used in the sequel.

**Lemma 14.** *Let us consider a  $d$ -CRN  $\mathcal{N}$  with stoichiometric matrix  $\Gamma \in \{-1, 0, 1\}^{n \times l}$  such that for all  $r \in \mathcal{R}$  reactions  $\sum_{i=1}^n [y^+]_i \leq 1$  and  $\sum_{i=1}^n [y^-]_i = 1$  holds. Assume that the reaction network of  $\mathcal{N}$  does not contain directed cycle (i.e.,  $\mathcal{N}$  has an acyclic network structure). Consider two states  $X_0, X' \in \mathbb{Z}_{\geq 0}^n$ . Then the reachability relation  $X_0 \rightsquigarrow_{\mathcal{N}} X'$  holds if and only if there exists  $c \in \mathbb{Z}_{\geq 0}^l$  vector satisfying the state equation  $X_0 + \Gamma c = X'$ .*

Now we can state the result on the reachability of subconservative  $d$ -CRNs composed of reaction having at most one input and one output species.

**Proposition 15.** *Let us consider a subconservative  $d$ -CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  of stoichiometric matrix  $\Gamma \in \{-1, 0, 1\}^{n \times l}$  and  $\Gamma^- \in \{0, 1\}^{n \times l}$  for which  $\mathcal{C} = \mathcal{S} \cup \{0\}$ . Assume that for all  $r \in \mathcal{R}$  reactions  $\sum_{i=1}^n [y^+]_i \leq 1$  and  $\sum_{i=1}^n [y^-]_i = 1$  hold. Let us consider two states  $X_0, X' \in \mathbb{Z}_{\geq 0}^n$  so that  $X_0 \geq M$  and  $X' \geq M$  hold where  $M = M(\Gamma^-)$  is defined by (20). Then the reachability relation  $X_0 \rightsquigarrow_{\mathcal{N}} X'$  holds if and only if there exists a vector  $c \in \mathbb{Z}_{\geq 0}^l$  satisfying the state equation  $X_0 + \Gamma c = X'$ .*

*Proof.*

$$(1) X_0 \rightsquigarrow_{\mathcal{N}} X' \implies \exists c \in \mathbb{Z}_{\geq 0}^l : X_0 + \Gamma c = X'$$

By the definition of reachability it is guaranteed that the state equation is satisfied with some  $c \in \mathbb{Z}_{\geq 0}^l$ .

$$(2) X_0 \rightsquigarrow_{\mathcal{N}} X' \iff \exists c \in \mathbb{Z}_{\geq 0}^l : X_0 + \Gamma c = X'$$

For this side an inductive proof is employed.

(a)  $k = 2$

If a  $d$ -CRN is 2-dimensional, according to Proposition 12, the existence of a solution  $c \in \mathbb{Z}_{\geq 0}^l$  of the state equation implies that the reachability relation holds.

(b) Inductive assumption

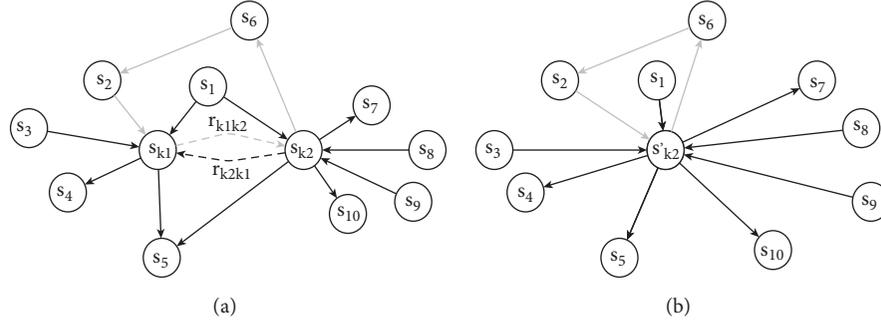


FIGURE 3: Graphical explanation of how the reaction network structure of  $\mathcal{N}'$  in the proof of Proposition 15 is constructed. (a) Reaction network structure of an  $n$ -dimensional d-CRN  $\mathcal{N}$ . (b) Reaction network structure of  $\mathcal{N}'$  resulting from merging the species  $s_{k_1}$  and  $s_{k_2}$  of  $\mathcal{N}$  along their shared reaction  $r_{k_1 k_2}$  (and reverse counterpart reaction  $r_{k_2 k_1}$ ). Note that by merging  $s_{k_1}$  and  $s_{k_2}$  we obtain a stoichiometric matrix  $\Gamma'$  having redundant reactions (e.g.,  $(s_1, s_{k_1})$ ,  $(s_1, s_{k_2})$  resulting in  $(s_1, s'_{k_1})$ ,  $(s_1, s'_{k_1})$ ) and zero reaction vectors (i.e., self-loops on  $s'_{k_1}$ ), but they are omitted in (b). A directed cycle on which the chosen reaction  $r_{k_1 k_2}$  lies is depicted in gray.

For  $k = n - 1$  we assume that the reachability relation  $X_0 \rightsquigarrow_{\mathcal{N}} X'$  holds.

(c)  $k = n$

We have two different cases with respect to the existence of directed cycles.

If the reaction network has no directed cycle, then the reachability relation  $X_0 \rightsquigarrow_{\mathcal{N}} X'$  is guaranteed by Lemma 14.

Assume that the reaction network contains at least one directed cycle

$$\sigma_S = s_{\nu(1)} \dots s_{\nu(h)} \quad (23)$$

where  $h \leq n$ ,  $s_{\nu(1)} = s_{\nu(h)}$  and  $s_{\nu(i)} \neq s_{\nu(j)}$  for  $i, j \in \{1, \dots, h\}$ ,  $i \neq j$ . Note again that  $\mathcal{C} = \mathcal{S} \cup \{0\}$ , and hence  $\sigma_S$  can be considered as a directed cycle of complexes in the reaction network (i.e.,  $\sigma_S = \sigma_C = s_{\nu(1)} \dots s_{\nu(h)}$ ). Let us consider an arbitrary  $r_{k_1 k_2} \in \mathcal{R}$  reaction defined between some  $s_{k_1}, s_{k_2} \in \sigma_S$ , i.e.  $r_{k_1 k_2} = s_{k_1} \longrightarrow s_{k_2}$ .

Now we construct a d-CRN  $\mathcal{N}' = (\mathcal{S}', \mathcal{C}', \mathcal{R}')$  from the stoichiometric matrix  $\Gamma \in \{-1, 0, 1\}^{(n-1) \times l}$  and  $\Gamma^- \in \{0, 1\}^{(n-1) \times l}$  as follows:

$$[\Gamma']_{i,:} = \begin{cases} [\Gamma]_{i,:}, & i < k_{\max}, i \neq k_{\min}, \\ [\Gamma]_{k_{\min},:} + [\Gamma]_{k_{\max},:}, & i = k_{\min}, \\ [\Gamma]_{i+1,:}, & k_{\max} \leq i \leq n-1, \end{cases} \quad (24)$$

and

$$[\Gamma^{-'}]_{i,:} = \begin{cases} [\Gamma^-]_{i,:}, & i < k_{\max}, i \neq k_{\min}, \\ [\Gamma^-]_{k_{\min},:} + [\Gamma^-]_{k_{\max},:}, & i = k_{\min}, \\ [\Gamma^-]_{i+1,:}, & k_{\max} \leq i \leq n-1. \end{cases} \quad (25)$$

Here  $k_{\min} = \min\{k_1, k_2\}$  and  $k_{\max} = \max\{k_1, k_2\}$ . This way we obtained a d-CRN  $\mathcal{N}'$  satisfying the assumptions of the proposition. Figure 3 gives an illustrative example of how  $\mathcal{N}'$  is constructed. Now we assign to each  $r' \in \mathcal{R}'$  the ordered pair of source complex and product complex of  $r \in \mathcal{R}$  from which it is obtained. In such a way every reaction of  $\mathcal{N}'$  is uniquely described by an ordered pair  $(r', r) \in \mathcal{R}' \times \mathcal{R}$ . Then by the mapping  $P((r', r)) = r$  one can uniquely determine the reaction  $r \in \mathcal{R}$  from which  $r' \in \mathcal{R}'$  is derived.

Let us construct the states  $X_0^m \in \mathbb{Z}_{\geq 0}^{n-1}$  and  $X'^m \in \mathbb{Z}_{\geq 0}^{n-1}$  as follows:

$$[X_0^m]_i = \begin{cases} [X_0]_i, & i < k_{\max}, i \neq k_{\min}, \\ [X_0]_{k_{\min}} + [X_0]_{k_{\max}}, & i = k_{\min}, \\ [X_0]_{i+1}, & k_{\max} \leq i \leq n-1. \end{cases} \quad (26)$$

$$[X'^m]_i = \begin{cases} [X']_i, & i < k_{\max}, i \neq k_{\min}, \\ [X']_{k_{\min}} + [X']_{k_{\max}}, & i = k_{\min}, \\ [X']_{i+1}, & k_{\max} \leq i \leq n-1. \end{cases} \quad (27)$$

Then we have that  $X_0^m \geq M(\Gamma^-)$  and  $X'^m \geq M(\Gamma^{-'})$ , and hence the  $(n-1)$ -dimensional d-CRN  $\mathcal{N}'$  with the initial and final states  $X_0^m$  and  $X'^m$  satisfies the assumptions of the proposition. From  $X_0 + \Gamma c = X'$  we have that  $X_0^m + \Gamma' c = X'^m$  holds; hence, according to the  $(n-1)$ -dimensional inductive assumption, the reachability relation

$$X_0^m \rightsquigarrow_{\mathcal{N}'} X'^m \quad (28)$$

follows.

Let us consider an admissible reaction vector sequence  $\sigma_r'$  associated with relation (28). Since for each  $r' \in \mathcal{R}'$  we associated the reaction  $r \in \mathcal{R}$  from which  $r'$  is obtained, making use of the mapping  $P: \mathcal{R}' \times \mathcal{R} \rightarrow \mathcal{R}$ , we can consider the reaction vector sequence  $\sigma_r$  ( $r \in \mathcal{R} \forall r \in \sigma_r$ ) uniquely determined by  $\sigma_r'$ . We start from  $X_0$  and modify the state variable  $X \in \mathbb{Z}_{\geq 0}^n$  according to the reaction vector sequence  $\sigma_r$ . We may get to two invalid cases:

(C<sub>1</sub>)  $[X]_{k_2} = 0$ , but the source complex of the forthcoming reaction  $r_{current} \in \sigma_r$  is  $s_{k_2}$ . Then, according to the  $(n-1)$ -dimensional reachability, it is guaranteed that  $s_{k_1}$  is charged at the current state  $X$ . Let us insert  $r_{k_1 k_2}$  into  $\sigma_r$  before the current reaction  $r_{current}$ .

(C<sub>2</sub>)  $[X]_{k_1} = 0$ , but the source complex of the forthcoming reaction  $r_{current} \in \sigma_r$  is  $s_{k_1}$ . Then, according to the  $(n-1)$ -dimensional reachability, it is guaranteed that  $s_{k_2}$  is charged at the current state  $X$ . It is known that  $s_{k_1}$  can be reached from  $s_{k_2}$  along a reaction vector sequence  $\sigma_r^*$  in the reaction network of  $\mathcal{N}$ . Let us insert  $\sigma_r^*$  into  $\sigma_r$  before the current reaction  $r_{current}$ .

By modifying  $\sigma_r$  according to the above discussed cases (C<sub>1</sub>) and (C<sub>2</sub>), we obtain an admissible reaction vector sequence  $\sigma_{r_{mod}}$  with respect to the reachability relation

$$X_0 \rightsquigarrow_{\mathcal{N}'} X^* \quad (29)$$

where  $X^* \geq 0^n$ ,  $[X^*]_i = [X']_i$  for  $i \in \{1, \dots, n\}$ ,  $i \neq k_1$  and  $i \neq k_2$ ; moreover,  $[X^*]_{k_1} + [X^*]_{k_2} = X'$ . According to the assumptions  $\mathcal{N}$  contains directed paths both from  $s_{k_1}$  to  $s_{k_2}$  and from  $s_{k_2}$  to  $s_{k_1}$ ; hence the reachability relation  $X^* \rightsquigarrow_{\mathcal{N}'} X'$  follows. Then, due to the transitivity of the relation  $\rightsquigarrow_{\mathcal{N}'}$ , we have that  $X_0 \rightsquigarrow_{\mathcal{N}'} X'$  also holds.  $\square$

Proposition 15 can be extended to the case of superconservative d-CRNs.

**Corollary 16.** *Let us consider a superconservative d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  with stoichiometric matrix  $\Gamma \in \{-1, 0, 1\}^{n \times l}$  and  $\Gamma^- \in \{0, 1\}^{n \times l}$  for which  $\mathcal{C} = \mathcal{S}$ . Assume that for all  $r \in \mathcal{R}$  reactions  $\sum_{i=1}^n [\bar{y}^+]_i = 1$  and  $\sum_{i=1}^n [\bar{y}^-]_i \leq 1$  hold. Let us consider two states  $X_0, X' \in \mathbb{Z}_{\geq 0}^n$  so that  $X_0 \geq M$  and  $X' \geq M$  hold where  $M = M(\Gamma^-)$  is defined by (20). Then the reachability relation  $X_0 \rightsquigarrow_{\mathcal{N}'} X'$  holds if and only if there exists a vector  $c \in \mathbb{Z}_{\geq 0}^l$  satisfying the state equation  $X_0 + \Gamma c = X'$ .*

*Proof.* By changing the sign of the entries in the stoichiometric matrix  $\Gamma$ , we get a subconservative d-CRN  $\mathcal{N}'$  of stoichiometric matrix  $-\Gamma$ . Then we can consider the reachability problem  $X' \overset{?}{\rightsquigarrow}_{\mathcal{N}'} X_0$ .  $\square$

We can extend Proposition 15 by allowing the restricted application of catalyzer species as follows.

**Proposition 17.** *Let us consider a subconservative d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  of stoichiometric matrix  $\Gamma \in \{-1, 0, 1\}^{n \times l}$  and  $\Gamma^- \in \{0, 1\}^{n \times l}$ . Assume that for each reaction  $r$ :*

- (1)  $r = s_1 \rightarrow s_2$  for some  $s_1, s_2 \in \mathcal{S}$ ,  $s_1 \neq s_2$ ,  $s_1 \neq \mathbf{0}$ , OR
- (2)  $r = s + s_1 \rightarrow s + s_2$  where  $s, s_1, s_2 \in \mathcal{S}$ ,  $s \neq s_1 \neq s_2$ ,  $s \neq \mathbf{0}$ ,  $s_1 \neq \mathbf{0}$  and  $\forall r' \in \mathcal{R}r'$  does not consume  $s$ .

*Let us consider two states  $X_0, X' \in \mathbb{Z}_{\geq 0}^n$  for which  $X_0 \geq M$  and  $X' \geq M$  where  $M = M(\Gamma^-)$  is defined by (20). Then the reachability relation  $X_0 \rightsquigarrow_{\mathcal{N}'} X'$  holds if and only if there exists a vector  $c \in \mathbb{Z}_{\geq 0}^l$  satisfying the state equation  $X_0 + \Gamma c = X'$ .*

*Proof.*

$$(1) X_0 \rightsquigarrow_{\mathcal{N}'} X' \implies X_0 + \Gamma c = X'$$

It follows from the definition of reachability.

$$(2) X_0 + \Gamma c = X' \implies X_0 \rightsquigarrow_{\mathcal{N}'} X'$$

Since in the initial state  $X_0$  the number of each catalyzer molecule is higher than or equal to 1 and there is no reaction in  $\mathcal{N}$  consuming a catalyzer species, it follows that for each state reachable from  $X_0$  the number of each catalyzer molecule is higher than or equal to 1. Let us remove all the catalyzer species of  $\mathcal{N}$  from the reactions where they act as a catalyzer; i.e., for each  $r \in \mathcal{R}$  of the form  $r = s + s_1 \rightarrow s + s_2$  we erase the catalyzer  $s$  to obtain  $r' = s_1 \rightarrow s_2$ . In such a way a d-CRN  $\mathcal{N}'$  is obtained so that for each  $X \geq M$ ,  $X_0 \rightsquigarrow_{\mathcal{N}'} X$  iff  $X_0 \rightsquigarrow_{\mathcal{N}'} X$ .  $\mathcal{N}'$  satisfies the conditions of Proposition 15; hence the reachability relation  $X_0 \rightsquigarrow_{\mathcal{N}'} X'$  holds implying that  $X_0 \rightsquigarrow_{\mathcal{N}'} X'$  also holds.  $\square$

According to the duality of the sub- and superconservativity properties, we can extend Proposition 17 to the case of superconservative d-CRNs.

**Corollary 18.** *Let us consider a superconservative d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  of stoichiometric matrix  $\Gamma \in \{-1, 0, 1\}^{n \times l}$  and  $\Gamma^- \in \{0, 1\}^{n \times l}$ . Assume that for each reaction  $r$*

- (1)  $r = s_1 \rightarrow s_2$  for some  $s_1, s_2 \in \mathcal{S}$ ,  $s_1 \neq s_2$ ,  $s_2 \neq \mathbf{0}$ , OR
- (2)  $r = s + s_1 \rightarrow s + s_2$  where  $s, s_1, s_2 \in \mathcal{S}$ ,  $s \neq s_1 \neq s_2$ ,  $s \neq \mathbf{0}$ ,  $s_2 \neq \mathbf{0}$  and  $\forall r' \in \mathcal{R}r'$  does not produce  $s$ .

*Let us consider two states  $X_0, X' \in \mathbb{Z}_{\geq 0}^n$  for which  $X_0 \geq M$  and  $X' \geq M$  where  $M = M(\Gamma^-)$  is defined by (20). Then the reachability relation  $X_0 \rightsquigarrow_{\mathcal{N}'} X'$  holds if and only if there exists a vector  $c \in \mathbb{Z}_{\geq 0}^l$  satisfying the state equation  $X_0 + \Gamma c = X'$ .*

*Proof.* By changing the sign of the entries in the stoichiometric matrix  $\Gamma$ , we obtain a subconservative d-CRN  $\mathcal{N}'$  of stoichiometric matrix  $-\Gamma$  satisfying the conditions of Proposition 17. We can consider the reachability problem  $X' \overset{?}{\rightsquigarrow}_{\mathcal{N}'} X_0$ .  $\square$

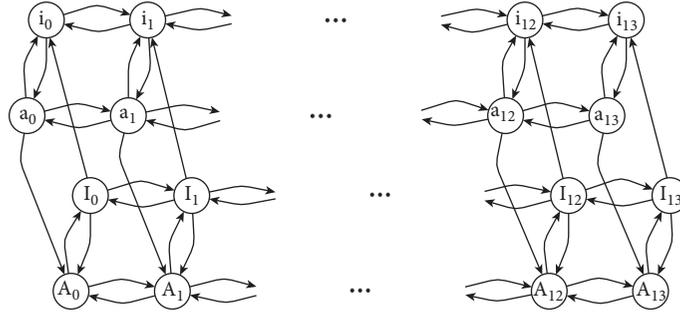


FIGURE 4: Conformational switch model of NFAT1 [38]. Lower case letters denote the protein located in the cytoplasm while upper case letters refer to the protein in the nucleus.  $a_j, A_j$  and  $i_j, I_j$  for  $j = 0, \dots, 13$  denote the active and inactive proteins, respectively. Lower indices denote the number of phosphorylated residues.

By the above corollary, any reachability problem on a superconservative d-CRN satisfying the conditions of Corollary 18 can be easily traced back to that of a subconservative network; hence the problem is equivalent to finding a  $c \in \mathbb{Z}_{\geq 0}^l$  solution for the respective d-CRN state equation.

The reaction network class covered by the above statements might be beneficial in modeling first- and second-order (bio)chemical reaction networks. For a representative example, see Example 19 below. We also note that any mass action type chemical reaction network can be dynamically described by an appropriately constructed reaction network containing at most second-order reactions [39]. Moreover, the hypergraph representation of chemical reaction networks (see, e.g., [40]) is helpful for checking the conditions of Proposition 17.

*Example 19.* Nuclear factors of activated T-cells (NFAT) are proteins that can exist in highly phosphorylated states [38]. They act as transcription factors; i.e., they have regulatory role in transcription. NFAT1, which is a member of the NFAT family, has 13 residues that can be dephosphorylated upon stimulation. NFAT1 has two different states: active and inactive. The transition between active and inactive states of the protein is regulated by the level of phosphorylation such that the higher the level of phosphorylation is, the lower the rate of transition becomes from inactive state to the active one and vice versa. Phosphorylation and dephosphorylation are achieved by a kinase and calcineurin, respectively. In the mathematical model the activities of kinase and calcineurin are modeled as rate constants; hence the respective reactions can be considered as first-order ones. The protein might be located in the cytoplasm or the nucleus of the cell. Cytoplasmic active NFAT1 is imported to the nucleus, while inactive NFAT1 of the nucleus is exported back to the cytoplasm.

The reaction network structure is depicted in Figure 4. It is visible that each reaction is of first order and there is no degradation and synthesis; hence the reaction network structure is conservative with a particular conservativity vector  $z = 1^{56}$  and Proposition 15 can be applied.

We note that a reachability problem of the discussed reaction network class without additional constraints may be determined in polynomial time [41]. However, by using

an ILP feasibility approach, the number of all distinct trajectories satisfying a prescribed reachability relation can be determined efficiently (see Remark 20), assuming the fixed number of reactions in the network. In addition, the ILP formulation can also be equipped with further linear constraints.

*Remark 20.* Let us consider a subconservative (superconservative) d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  of  $n$  species,  $m$  complexes, and  $l$  reactions. Assume that  $\mathcal{N}$  satisfies the conditions of Proposition 17 (Corollary 18). Then for any  $X_0, X' \in \mathbb{Z}_{\geq 0}^n$  initial and target states for which  $X_0 \geq M(\Gamma^-)$ ,  $X' \geq M(\Gamma^-)$  hold we have that the number of distinct trajectories  $\sigma_X$  satisfying the reachability relation  $X_0 \overset{?}{\rightsquigarrow}_{\mathcal{N}} X'$  can be determined in polynomial time in the distance of  $X_0$  and  $X'$ , given the fixed number of reactions  $l$  in the d-CRN. The explanation of this is the following. According to Proposition 17 (Corollary 18) the reachability problem  $X_0 \overset{?}{\rightsquigarrow}_{\mathcal{N}} X'$  is equivalent to the existence of a nonnegative integer solution  $c \in \mathbb{Z}_{\geq 0}^l$  of the state equation  $X_0 + \Gamma c = X'$ . In this way the reachability problem can be reformulated as an ILP feasibility problem in terms of  $c$ , and the Barvinok algorithm can be applied. Using the Barvinok algorithm in this particular case, the following complexity bounds are obtained:

- (1) exponential in the dimension of the decision variables, that is, in the number of different reactions  $l$ ,
- (2) polynomial in the number of constraints, that is, in the number of species  $n$ ,
- (3) polynomial in the maximum of the absolute values of the coefficients  $\Gamma, X' - X_0$ .

The particular importance of Remark 20 is that the time complexity of the trajectory counting problem between a prescribed pair of states is polynomial in the number of constraints and in the distance of the initial and target states even in the case of superconservative d-CRNs for which the associated reachable state space can be unbounded for any  $X_0$  initial state.

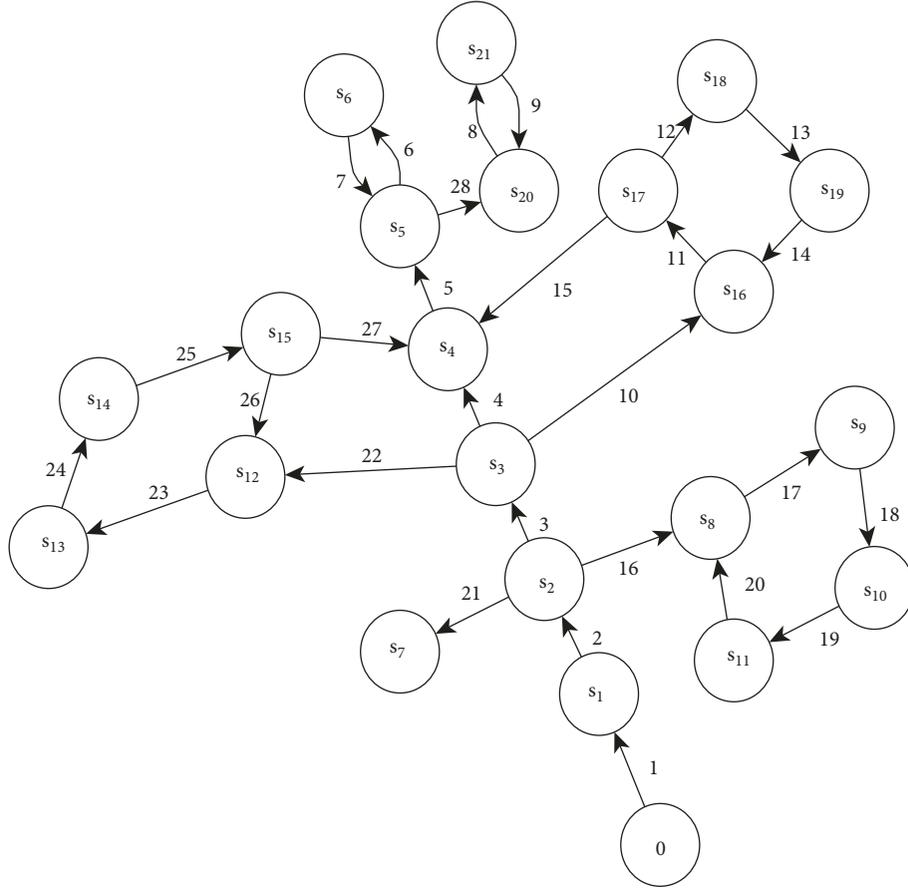


FIGURE 5: A superconservative d-CRN.  $\mathbf{0}$  indicates the zero complex and the numbers denote the indices of the reactions on which they are located. Due to the superconservativity of the network structure, the above d-CRN is unbounded for any initial state  $X_0 \in \mathbb{Z}_{\geq 0}^n$ .

## 6. Computational Example: A Superconservative d-CRN of First-Order Reactions

Let us consider the d-CRN depicted in Figure 5. This system is superconservative with a particular conservation vector  $z = 1^{21 \times 1}$  implying the unboundedness of its reachable state space regardless of the initial state  $X_0$ . Making use of the above results, the reachability problem of  $X_0 \stackrel{?}{\rightsquigarrow} X'$  for any  $X_0, X' \in \mathbb{Z}_{\geq 0}^{21}$  can be reformulated as a subconservative d-CRN reachability problem for which the boundedness of the reachable state space, i.e. structural boundedness, is guaranteed and is equivalent to the existence of a nonnegative integer solution of the respective subconservative d-CRN state equation.

As initial state we consider  $X_0$  given by (31) that was randomly generated from  $[10, 100]^{21}$ . In order to find a target state  $X'$  satisfying the reachability relation  $X_0 \rightsquigarrow X'$  we

randomly generated target states so that the number of each species was uniformly sampled from the interval  $[40, 100]$ . In the choice of the intervals from which we sample, it was taken into consideration that the discrete state model of reaction networks is typically employed in the case of low molecular counts [4, 5]. In order to decide the reachability relation between a pair of particular states  $X_0$  and  $X'$ , we need to solve the following decision problem.

$$\begin{aligned} \Gamma c &= X' - X_0 \\ c &\in \mathbb{Z}_{\geq 0}^{28} \end{aligned} \quad (30)$$

Clearly, Corollary 18 guarantees that  $\Gamma c = X' - X_0$  is satisfied with some  $c \in \mathbb{Z}_{\geq 0}^{28}$  if and only if the reachability relation  $X_0 \rightsquigarrow X'$  holds. Let us consider the following initial and final states.

$$X_0 = [56 \ 10 \ 35 \ 87 \ 66 \ 75 \ 87 \ 60 \ 60 \ 55 \ 50 \ 89 \ 58 \ 72 \ 52 \ 71 \ 48 \ 71 \ 57 \ 47 \ 68]^\top \quad (31)$$

$$X' = [50 \ 46 \ 65 \ 77 \ 88 \ 95 \ 71 \ 56 \ 59 \ 54 \ 43 \ 76 \ 55 \ 78 \ 40 \ 62 \ 51 \ 71 \ 53 \ 64 \ 91]^\top \quad (32)$$

We found that for the target state  $X'$  given by (32) the reachability relation holds. To solve the decision problems of the form (30) the LattE [34] software was used.

Now, let us examine the reachability from  $X_0$  to  $X'$  with additional constraints. One can observe that  $X'$  results in a significant increase in the number of molecules in the species  $s_5$ ,  $s_6$ ,  $s_{20}$ , and  $s_{21}$  and any trajectory from  $X_0$  to  $X'$  results in a net increase in the number of molecules. These together imply the flow of molecules from the zero complex (environment). The flow of molecules over the network from the zero complex to  $s_5$ ,  $s_6$ ,  $s_{20}$ , and  $s_{21}$  can take place through different paths. We assume that the directed paths

$$\sigma_{S1} = s_3 \ s_{12} \ s_{13} \ s_{14} \ s_{15} \ s_4, \quad (33)$$

$$\sigma_{S2} = s_3 \ s_{16} \ s_{17} \ s_4 \quad (34)$$

$$c = [112 \ 106 \ 112 \ 111 \ 118 \ 48 \ 8 \ 29 \ 12 \ 18 \ 16 \ 11 \ 11 \ 7 \ 8 \ 13 \ 17 \ 16 \ 15 \ 8 \ 17 \ 13 \ 16 \ 13 \ 19 \ 16 \ 9 \ 51]^T \quad (36)$$

For implementation purposes we employed Python 2.7 programming language and the Gurobi mathematical optimization solver [42]. A Lenovo P51s workstation with two 2.70 GHz i7-7500U CPUs and 32GB RAM (DDR4 2133 MHz) was used for all the computations.

## 7. Conclusion

In this paper the reachability problems of sub- and super-conservative discrete state chemical reaction networks are considered. It is shown that the reachability problem of a superconservative reaction network of unbounded reachable state space can be transformed to that of a subconservative network for which the boundedness of the reachable state space is always guaranteed. Using an inductive proof we provided a set of necessary and sufficient conditions under which the equivalence between a d-CRN reachability problem and existence of nonnegative integer solution of the corresponding state equations is guaranteed. In such a way the reachability problem can be traced back to an IP-feasibility (decision) problem for which the number of decision variables is significantly lower than that employed in the literature [36]. Moreover the number of trajectories satisfying the reachability relation can also be enumerated efficiently, assuming a fixed reaction network structure. The applicability of our approach is illustrated on a high-dimensional superconservative d-CRN.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

are slow compared to the other ones; hence we wish to minimize the flow through them in order to lower their effect in  $c$ . This can be easily expressed by posing addition linear constraints on  $c$  as is done in the decision problem (35).

$$\begin{aligned} \Gamma c &= X' - X_0 \\ c &\in \mathbb{Z}_{\geq 0}^{28} \\ [c]_{15} &\leq 10 \\ [c]_{27} &\leq 10 \end{aligned} \quad (35)$$

We also determined a particular solution  $c$  by equipping (35) with the objective function  $\sum_{i=1}^{28} [c]_i$  to be minimized.

## Acknowledgments

Gergely Szlobodnyik acknowledges the support of the New National Excellence Program (ÚNKP-18-3). Gábor Szederkényi acknowledges the support of the European Union, cofinanced by the European Social Fund through the grant EFOP-3.6.3-VEKOP-16-2017-00002, and the support of the National Research, Development and Innovation Office (NKFIH) through grant no. 125739.

## References

- [1] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman & Hall, 2007.
- [2] W. Liu, *Introduction to Modeling Biological Cellular Control Systems*, Springer Science & Business Media, 2012.
- [3] D. Angeli, "A tutorial on chemical reaction network dynamics," *European Journal of Control*, vol. 15, no. 3-4, pp. 398–406, 2009.
- [4] P. érdi and J. Tóth, *Mathematical Models of Chemical Reactions. Theory and Applications of Deterministic and Stochastic Models*, Manchester Univ. Press, 1989.
- [5] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, no. 5584, pp. 1183–1186, 2002.
- [6] S. Smith and R. Grima, "Single-cell variability in multicellular organisms," *Nature Communications*, vol. 9, no. 1, 2018.
- [7] T. G. Kurtz, "The relationship between stochastic and deterministic models for chemical reactions," *The Journal of Chemical Physics*, vol. 57, no. 7, pp. 2976–2978, 1972.
- [8] D. F. Anderson and T. G. Kurtz, *Stochastic Analysis of Biochemical Systems*, Springer, Berlin, Germany, 2015.
- [9] D. F. Anderson, G. A. Enciso, and M. D. Johnston, "Stochastic analysis of chemical reaction networks with absolute concentration robustness," *Journal of the Royal Society Interface*, vol. 11, no. 90, Article ID 20130505, 2014.
- [10] D. F. Anderson and T. G. Kurtz, "Continuous time Markov chain models for chemical reaction networks," in *Design and Analysis*

- of *Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology*, H. Koepl, Ed., pp. 3–42, Springer, Berlin, Germany, 2011.
- [11] F. Bause and P. S. Kritzing, *Stochastic Petri Nets: An Introduction to the Theory*, Vieweg Verlag, 2002.
  - [12] G. Szederkenyi, K. M. Hangos, and T. Péni, “Maximal and minimal realizations of reaction kinetic systems: computation and properties,” *MATCH - Communications in Mathematical and in Computer Chemistry*, vol. 65, no. 2, pp. 309–332, 2011.
  - [13] G. Szederkényi, “Computing sparse and dense realizations of reaction kinetic systems,” *Journal of Mathematical Chemistry*, vol. 47, no. 2, pp. 551–568, 2010.
  - [14] B. Ács, G. Szederkényi, Z. Tuza, and Z. A. Tuza, “Computing all possible graph structures describing linearly conjugate realizations of kinetic systems,” *Computer Physics Communications*, vol. 204, pp. 11–20, 2016.
  - [15] B. Ács, G. Szlobodnyik, and G. Szederkényi, “A computational approach to the structural analysis of uncertain kinetic systems,” *Computer Physics Communications*, vol. 228, pp. 83–95, 2018.
  - [16] M. D. Johnston, “Conditions for extinction events in chemical reaction networks with discrete state spaces,” *Mathematical Biosciences*, vol. 294, pp. 130–142, 2017.
  - [17] M. D. Johnston, “A computational approach to extinction events in chemical reaction networks with discrete state spaces,” *Mathematical Biosciences*, vol. 294, pp. 130–142, 2017.
  - [18] L. Paulevé, G. Craciun, and H. Koepl, “Dynamical properties of discrete reaction networks,” *Journal of Mathematical Biology*, vol. 69, no. 1, pp. 55–72, 2014.
  - [19] R. Brijder, “Dominance and deficiency for Petri nets and chemical reaction networks,” *Natural Computing*, vol. 16, no. 2, pp. 285–294, 2017.
  - [20] M. Cook, D. Soloveichik, E. Winfree, and J. Bruck, “Programmability of chemical reaction networks,” in *Algorithmic bioprocesses*, Nat. Comput. Ser., pp. 543–584, Springer, Berlin, 2009.
  - [21] A. Case, J. H. Lutz, and D. M. Stull, “Reachability problems for continuous chemical reaction networks,” *Natural Computing*, vol. 17, no. 2, pp. 223–230, 2018.
  - [22] E. W. Mayr, “An algorithm for the general petri net reachability problem,” in *Proceedings of the the thirteenth annual ACM symposium*, pp. 238–246, Milwaukee, Wisconsin, USA, May 1981.
  - [23] S. R. Kosaraju, “Decidability of reachability in vector addition systems,” in *STOC*, pp. 267–28, ACM, 1982.
  - [24] J.-L. Lambert, “A structure to decide reachability in Petri nets,” *Theoretical Computer Science*, vol. 99, no. 1, pp. 79–104, 1992.
  - [25] J. Leroux, “Vector addition reachability problem (a simpler solution),” in *The Alan Turing Centenary Conference*, vol. 10, pp. 214–228, EPiC series, 2012.
  - [26] R. J. Lipton, “The reachability problem requires exponential space,” Tech. Rep 62, Department of Computer Science, Yale University, 1976.
  - [27] S. Schmitz, “Complexity hierarchies beyond elementary,” *ACM Transactions on Computation Theory*, vol. 8, no. 1, pp. 1–36, 2016.
  - [28] J. Lenstra, “Integer programming with a fixed number of variables,” *Mathematics of Operations Research*, vol. 8, no. 4, pp. 538–548, 1983.
  - [29] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, vol. 2 of *Algorithms and Combinatorics: Study and Research Texts*, Springer, Berlin, Germany, 1988.
  - [30] A. I. Barvinok, “A polynomial time algorithm for counting integral points in polyhedra when the dimension is fixed,” *Mathematics of Operations Research*, vol. 19, no. 4, pp. 769–779, 1994.
  - [31] M. Köppe, “A primal Barvinok algorithm based on irrational decompositions,” *SIAM Journal on Discrete Mathematics*, vol. 21, no. 1, pp. 220–236, 2007.
  - [32] S. Verdoolaege, R. Seghir, K. Beyls, V. Loechner, and M. Bruynooghe, “Counting integer points in parametric polytopes using Barvinok’s rational functions,” *Algorithmica. An International Journal in Computer Science*, vol. 48, no. 1, pp. 37–66, 2007.
  - [33] J. B. Lasserre, “Integer programming, Barvinok’s counting algorithm and Gomory relaxations,” *Operations Research Letters*, vol. 32, no. 2, pp. 133–137, 2004.
  - [34] J. A. De Loera, R. Hemmecke, J. Tauzer, and R. Yoshida, “Effective lattice point counting in rational convex polytopes,” *Journal of Symbolic Computation*, vol. 38, no. 4, pp. 1273–1302, 2004.
  - [35] G. Memmi and G. Roucairol, “Linear algebra in net theory,” in *Net Theory and Applications*, W. Brauer, Ed., vol. 84 of *Lecture Notes in Comput. Sci.*, pp. 213–223, Springer, Berlin-New York, 1975.
  - [36] G. Szlobodnyik, G. Szederkenyi, and M. D. Johnston, “Reachability analysis of subconservative discrete chemical reaction networks,” *MATCH Communications in Mathematical and in Computer Chemistry*, vol. 81, no. 3, pp. 705–736, 2019.
  - [37] T. Murata, “Petri nets: properties, analysis and applications,” *Proceedings of the IEEE*, vol. 77, no. 4, pp. 541–580, 1989.
  - [38] C. Salazar and T. Höfer, “Allosteric regulation of the transcription factor NFAT1 by multiple phosphorylation sites: a mathematical analysis,” *Journal of Molecular Biology*, vol. 327, no. 1, pp. 31–45, 2003.
  - [39] T. Wilhelm, “Chemical systems consisting only of elementary steps –a paradigm for nonlinear behavior,” *Journal of Mathematical Chemistry*, vol. 27, no. 1-2, pp. 71–88, 2000.
  - [40] R. Fagerberg, C. Flamm, D. Merkle, P. Peters, and P. F. Stadler, “On the complexity of reconstructing chemical reaction networks,” *Mathematics in Computer Science*, vol. 7, no. 3, pp. 275–292, 2013.
  - [41] P. Alimonti, E. Feuerstein, L. Laura, and U. Nanni, “Linear time analysis of properties of conflict-free and general Petri nets,” *Theoretical Computer Science*, vol. 412, no. 4-5, pp. 320–338, 2011.
  - [42] Gurobi Optimization, “Gurobi optimizer reference manual,” 2016, <http://www.gurobi.com>.

