

# Polynomial Time Reachability Analysis in Discrete State Chemical Reaction Networks Obeying Conservation Laws

Gergely Szlobodnyik<sup>a,\*</sup>, Gábor Szederkényi<sup>a,b</sup>

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

<sup>b</sup> *Systems and Control Laboratory, ELKH Institute for Computer Science  
and Control (SZTAKI), Kende u. 13-17, H-1111 Budapest, Hungary*  
szlobodnyik.gergely@itk.ppke.hu, szederkenyi.gabor@itk.ppke.hu

(Received April 5, 2022)

## Abstract

In this paper the reachability problem of discrete state Chemical Reaction Networks (d-CRNs) is studied. We consider sub-classes of sub- and superconservative d-CRN network structures and prove that the reachability relation can be decided in polynomial time. We make use of the result that in the studied d-CRN sub-classes, the reachability relation is equivalent to the existence of a non-negative integer solution of the d-CRN state equation. The equivalence implies the reformulation of the reachability problem as integer linear programming decision problem. We show that in the studied classes of d-CRN structures, the state equation has a totally unimodular coefficient matrix. As the reachability relation is equivalent to the non-negative integer solution of the state equation, the resulting integer programming decision program can be relaxed to a simple linear program having polynomial time complexity. Hence, in the studied sub-classes of sub- and superconservative reaction network structures, the reachability relation can be decided in polynomial time and the number of continuous decision variables is equal to the number of reactions of the d-CRN.

---

\*Corresponding author.

---

# 1 Introduction

Formal mathematical models of Chemical Reaction Networks (CRNs) are commonly used to model the dynamical behavior and structural properties of a wide range of networked systems such as classical chemical and biochemical reactions, gene regulatory networks, protein-protein interactions, population dynamics, or epidemiological systems [1, 5]. Assuming high molecular counts and well mixed (homogeneous) distribution of the interacting species, the dynamical behavior of CRNs is commonly described by ordinary differential equations (ODEs) of continuous variables [2–5]. In many cases, however, the molecular count of the different species can be low (e.g., smaller than 100 molecules per species). In this case continuous ODEs are not suitable for modeling of the underlying dynamical behavior. Discrete state models are introduced as means for characterizing the quantitative dynamical behavior of CRNs [6]. As the state space of the resulting models is discrete, these are called discrete (state) Chemical Reaction Networks (d-CRNs). There are multiple formal models used to describe the dynamical behavior of d-CRNs, such as Markov chains or Petri nets [7–9].

In this paper we study the structural properties of d-CRNs and their implications on the dynamical behavior. It is well-known that in order to completely characterize a CRN, it is required to simultaneously study the dynamical behavior and network structure. In the case of continuous (state) CRNs, the same ODE model can be obtained by structurally (topologically) different reaction networks [10, 11]. That is, the same dynamical behavior can be the result of different sets of interactions between a set of species. Considering d-CRNs, the reachability problem is important to understand and predict dynamical behavior. Given a pair of non-negative initial and target states, is it possible to reach the target state from the initial one by means of the available set of reactions? Through the reachability analysis, specific dynamical properties can be studied such as the existence of extinction events. An extinction event is the irreversible zeroing of the count of a species. It is proven that under network structure-related conditions, an extinction event can occur from any initial state of

---

the discrete state space of the d-CRN [12, 13]. We note that the reachability relation (for any pair of initial and target states) is determined by the d-CRN's network structure and it is independent of the integro-differential equations characterizing the evolution of the discrete state variables. In addition to the structure-dynamics relations, we note that the reachability problem is equivalent to the gate implementability problem, i.e. the implementability of synthetic biological circuits can be reformulated as reachability problems [29].

The formal mathematical models of d-CRNs are equivalent to well-studied models of theoretical computer science, namely Petri Nets and Vector Addition Systems with States (VASS) [14, 15]. The d-CRN reachability problem is equivalent to the VASS reachability problem. The VASS reachability is proven to be decidable [16–19] with EXSPACE lower bound in space complexity [20]. In the case of continuous state chemical reaction networks, the reachability problem has polynomial time complexity [21]. However, for the class of d-CRNs it is an open problem whether there exists an algorithm with primitive recursive time complexity that can decide the problem [22]. In the general case, the d-CRN reachability problem can be formulated as an integer programming (IP) decision problem [23]. For specific sub-classes of d-CRN structures, it was proven that the reachability relation can be decided with relaxed time complexity using the d-CRN state equation as the constraint set of the problem [23, 24].

The aim of this paper is to provide theoretically guaranteed bounds on the computational complexity of the d-CRN reachability problem for specific sub-classes of d-CRN topologies. Furthermore, a linear programming approach is also described to decide the reachability problem with polynomial time complexity. We prove that for sub-classes of sub- and superconservative reaction network structures the NP-hard problem of d-CRN reachability can be reformulated as linear programming problem of polynomial complexity. In [24] necessary and sufficient conditions are obtained for which the d-CRN reachability relation is equivalent to the non-negative integer solution of the d-CRN state equation. Using this result, the authors expressed the reachability problem as integer decision problem for which well-decoupled time complexity can be obtained by means

of the Lenstra and Barvinok algorithms. In this paper it is shown that for the same sub-classes of d-CRN topologies, the integer problem has a totally unimodular constraint matrix. It is known that for totally unimodular matrices the linear integer programming problem can be relaxed to a linear optimization problem of continuous variables that has polynomial time complexity [30]. Making use of this result we can relax the NP-hard integer decision problem to polynomial time linear programming problem.

## 2 Modeling and computation background

### 2.1 Discrete state chemical reaction networks

**Definition 1.** A discrete state Chemical Reaction Network (d-CRN) of  $n$  species,  $m$  complexes and  $l$  reactions is a 3-tuple  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  so that

1.  $\mathcal{S} = \{s_i \mid i = 1, \dots, n\}$
2.  $\mathcal{C} = \{y_j = \sum_{i=1}^n \alpha_{ji} s_i \mid s_i \in \mathcal{S}, \alpha_{ji} \in \mathbb{Z}_{\geq 0}, i = 1, \dots, n, j = 1 \dots m\}$
3.  $\mathcal{R} = \{r_k = y_{source(r_k)} \rightarrow y_{product(r_k)} \mid y_{source(r_k)}, y_{product(r_k)} \in \mathcal{C}, k = 1, \dots, l\}$

where  $s_i$  denotes the  $i$ th species,  $y_j$  is the  $j$ th complex and  $r_k$  is the  $k$ th reaction of  $\mathcal{N}$ , provided a fixed ordering of the entries in  $\mathcal{S}, \mathcal{C}$  and  $\mathcal{R}$ . In the sequel it is assumed that there is a fixed ordering of the species, complexes and reactions. The non-negative integer  $\alpha_{ji}$  denotes the stoichiometric coefficient of the  $i$ th species in the  $j$ th complex. Given a reaction  $r \in \mathcal{R}$ ,  $r = y_{source(r)} \rightarrow y_{product(r)}$ ,  $y_{source(r)}$  and  $y_{product(r)}$  denote the source complex and the product complex, respectively.

For each  $y_j \in \mathcal{C}$  complex, we introduce its vector representation  $\bar{y}_j$  as follows

$$\bar{y}_j = [\alpha_{j1} \ \alpha_{j2} \ \dots \ \alpha_{jn}]^T \quad (1)$$

Using the above defined encoding of the complexes, for each  $r \in \mathcal{R}$  we define the reaction vectors

$$r_{ij} = \bar{y}_j - \bar{y}_i \quad (2)$$

where  $y_i$  and  $y_j$  are the respective source and product complexes of  $r$ . Note that  $r_{ij}$  encodes the net molecular count changes as a result of the firing (occurrence) of the reaction between complexes  $y_i$  and  $y_j$ . For the sake of convenience, from now on we use the notation  $r_k$  to denote both the  $k$ 'th reaction and its respective reaction vector representation. For each  $r \in \mathcal{R}$  reaction we can assign a real-valued reaction rate constant (intensity, propensity) as  $k_r$ .

Note that in the above definition of d-CRNs  $\mathcal{S}$ ,  $\mathcal{C}$  and  $\mathcal{R}$  are sets in a mathematical sense. This implies that the different species and complexes are unique. In addition, we assume that there are no reactions in which the product and the source complexes are the same, that is loop reactions of the form  $C \rightarrow C$  for some  $C \in \mathcal{C}$  are not allowed. Isolated complexes, that is complexes not taking part in any reactions, are also disallowed.

A d-CRN can also be described in the form of a weighted directed graph  $G$ . The nodes are the complexes and the directed edges denote the reactions between complexes so that a directed edge points from the source complex to the product complex of the respective reaction. Formally, the network representation can be written as  $G = (V, E)$ , where

$$\begin{aligned} V &= \mathcal{C} \\ E &= \mathcal{R} \end{aligned}$$

so that if  $r \in \mathcal{R}$  for  $r = y_i \rightarrow y_j$ , then  $\exists e \in E$  so that  $e$  points from the vertex corresponding to  $y_i$  to the vertex of  $y_j$ . For each edge  $e \in E$  we assign a weight that is equal to the reaction rate constant of the reaction represented by  $e$ . We will use the terms ‘structure’ and ‘topology’ interchangeably to denote the reaction network graph.

**Definition 2.** Let us consider a d-CRN  $\mathcal{N}$  with reaction vectors  $r_1, r_2, \dots, r_l$ . The stoichiometric matrix  $\Gamma_{\mathcal{N}}$  is defined as

$$\Gamma_{\mathcal{N}} = [r_1 \ r_2 \ \dots \ r_l] \tag{3}$$

We also define  $\Gamma_{\mathcal{N}}^+$  and  $\Gamma_{\mathcal{N}}^-$  as follow

$$\Gamma_{\mathcal{N}}^+ = [\bar{y}_{r_1}^+ \ \dots \ \bar{y}_{r_l}^+] \tag{4}$$

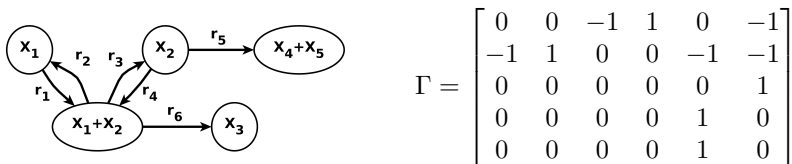
where  $\bar{y}_{r_k}^+$  for  $k = 1, 2 \dots l$  denotes the vector representation of the product complex of the reaction  $r_k$ , that is  $[\bar{y}_{r_k}^+]_i$  encodes the stoichiometric coefficient of the  $i$ 'th species for  $i = 1, 2 \dots n$  in reaction  $r_k$ .

$$\Gamma_{\mathcal{N}}^- = [\bar{y}_{r_1}^- \ \dots \ \bar{y}_{r_l}^-] \quad (5)$$

where  $\bar{y}_{r_k}^-$  denotes the vector representation of the source complex of  $r_k$   $k = 1, 2 \dots l$ .

For an arbitrary d-CRN  $\mathcal{N}$  the following equality holds

$$\Gamma_{\mathcal{N}} = \Gamma_{\mathcal{N}}^+ - \Gamma_{\mathcal{N}}^- \quad (6)$$



**Figure 1.** A discrete state chemical reaction network (d-CRN). **Left:** reaction network structure (topology). The nodes represented the complexes ( $V = \mathcal{C}$ ) while the edges indicate the reactions between complexes ( $E = \mathcal{R}$ ). Note that the directed edges point from the node of the source complex to the node of the product complex of the underlying reaction. By the numbers on the edges we indicate a fixed ordering of the reactions. **Right:** the stoichiometric matrix of the d-CRN.

Let us consider an integer initial state vector  $X_0 \in \mathbb{Z}_{\geq 0}^n$ . Clearly,  $[X_0]_i$  encodes the molecule count of the  $i$ 'th species in the initial state for  $i = 1, 2 \dots n$ . Any state transition of the d-CRN can be algebraically described by the d-CRN's discrete state equation

$$X' = X_0 + \Gamma c \quad (7)$$

where  $X' \in \mathbb{Z}_{\geq 0}^n$  is a state vector and  $c \in \mathbb{Z}_{\geq 0}^l$  encodes the occurrences of the reactions along a state transition sequence from  $X_0$  to  $X'$ , that is  $[c]_k$  encodes the number of times the  $k$ 'th reaction occurred (fired) along a state transition sequence.

**Definition 3.** Let us consider a d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  and an index function  $\nu$  that assigns an index to each reaction:  $\nu : \mathcal{R} \mapsto \mathbb{N}$ .

1. A species  $s \in \mathcal{S}$  is called a catalyzer of a reaction  $r \in \mathcal{R}$ , if  $r = s + s_1 \rightarrow s + s_2$  for some  $s_1, s_2 \in \mathcal{S}$ .
2. A complex  $y \in \mathcal{C}$  is said to be charged at a state  $X \in \mathbb{Z}_{\geq 0}^n$ , if  $X \succeq \bar{y}$ . A reaction  $r \in \mathcal{R}$  is said to be charged at a state  $X$  if its respective source complex is charged.
3. A state  $X \in \mathbb{Z}_{\geq 0}^n$  reacts to a state  $X' \in \mathbb{Z}_{\geq 0}^n$  ( $X \rightarrow X'$ ) if there exists a reaction  $r \in \mathcal{R}$  so that  $X + r = X'$ .
4. A reaction vector sequence  $\sigma_r$  is an ordered set of reactions  $\sigma_r = r_{\nu(1)} \dots r_{\nu(v)}$  with  $v \in \mathbb{Z}_{\geq 1}$ .
5. A state (transition) sequence  $\sigma_X$  is an ordered sequence of states  $\sigma_X = X_{\nu(1)} \dots X_{\nu(v)}$  with  $v \in \mathbb{Z}_{\geq 1}$  so that  $X_{\nu(i)} \rightarrow X_{\nu(i+1)}$  for  $i = 1, \dots, v - 1$ .
6. A state  $X'$  is said to be reachable from a state  $X_0$  ( $X_0 \rightsquigarrow_{\mathcal{N}} X'$ ), if there exists a state transition sequence  $\sigma_X = X_{\nu(1)} \dots X_{\nu(v)}$  so that  $X_{\nu(1)} = X_0$ ,  $X_{\nu(v)} = X'$  and for all  $X \in \sigma_X$ ,  $X \succeq 0^n$ .

For any state transition sequence  $\sigma_X = X_1 \dots X_v$ ,  $X_0$  is called the initial state,  $X'$  is the target state and  $X_i$  for  $i = 2, \dots, v - 1$  are called transition states. A state transition sequence  $\sigma_X$  is said to be admissible if  $X \succeq 0^n$  for all  $X \in \sigma_X$ . We note that the definition of reachability is restricted to admissible state transition sequences.

For any non-negative pair  $X_0, X'$ , the existence of a non-negative  $c \in \mathbb{Z}_{\geq 0}^l$  solution of the d-CRN state equation Eq. (7) does not imply the reachability  $X_0 \rightsquigarrow_{\mathcal{N}} X'$ . The existence of a  $c$  non-negative integer vector guarantees that the system can be driven from the initial state  $X_0$  to the target  $X'$ . However, it is not guaranteed that there exists a trajectory in which all the transition states are non-negative. If negative transition states occur in all the possible trajectories, then the reachability relation does not follow.

Next we introduce the definition of a Net System to emphasize the importance of initial states in d-CRNs.

**Definition 4.** A Chemical Reaction Network System (Net System) is defined by a tuple  $(\mathcal{N}, X_0)$  where  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  is an arbitrary d-CRN and  $X_0$  is a non-negative initial state of  $\mathcal{N}$ .

**Definition 5. Reachable state space**

Let consider a Net System  $(\mathcal{N}, X_0)$ . The reachable state space  $Reach(\mathcal{N}, X_0)$  associated to  $(\mathcal{N}, X_0)$  is the set of all the non-negative states reachable from  $X_0$  by the d-CRN  $\mathcal{N}$ , formally:

$$Reach(\mathcal{N}, X_0) = \left\{ X \mid X_0 \rightsquigarrow_{\mathcal{N}} X \right\} \quad (8)$$

No we introduce the definitions of sub-and superconservativity.

**Definition 6.** A d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  with stoichiometric matrix  $\Gamma_{\mathcal{N}} \in \mathbb{Z}^{n \times l}$  is said to be subconservative (superconservative), if there exists a strictly positive vector  $z \in \mathbb{R}_{>0}^l$  so that  $z^{\top} \Gamma_{\mathcal{N}} \leq 0^{1 \times l}$  ( $z^{\top} \Gamma_{\mathcal{N}} \geq 0^{1 \times l}$ ). Then the vector  $z$  is called conservation vector.

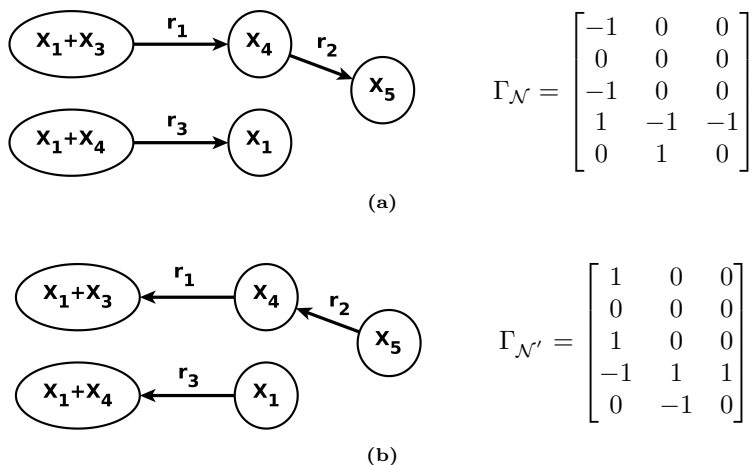
**Example 1.** Figure 2 depicts two reaction networks with sub-and super-conservative property. Note that the networks can be transformed to each other by reversing the signs of the entries in their stoichiometric matrices. Geometrically, changing the signs of the entries in a stoichiometric matrix means reversing the direction of the encoded reaction network structure.

Note that sub-and superconservativity are topological (structural) properties of d-CRNs, they are independent of the reaction rates. As a special case we can introduce the class of conservative reaction networks.

**Definition 7.** A d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  of stoichiometric matrix  $\Gamma_{\mathcal{N}} \in \mathbb{Z}^{n \times l}$  is called conservative if there exists a vector  $z \in \mathbb{R}_{>0}^l$  for which  $z^{\top} \Gamma_{\mathcal{N}} = 0^{1 \times l}$  holds.

The reachable state space  $Reach(\mathcal{N}, X_0)$  for a conservative d-CRN  $\mathcal{N}$  and arbitrary initial state  $X_0 \in \mathbb{Z}_{\geq 0}^n$  is located in on at most  $(n - 1)$ -dimensional hyperplane, which is called the conservativity hyperplane of the CRN.





**Figure 2.** A sub and a superconservative d-CRN. The fixed ordering of the reactions is noted on the edges of the reaction graphs. The presented reaction networks can be transformed to each other by reversing the signs of the entries of the stoichiometric matrix. **a)** subconservative d-CRN. **b)** superconservative d-CRN.

We note that CRNs with conservation laws compose widely studied subclasses of reaction network structures. Conservative CRNs are considered with both continuous and discrete state space in the literature [25–28].

## 2.2 Integer linear programming

In this section we briefly review the integer linear programming problem, the corresponding feasibility problem and their relaxation. We will extensively use these concepts as solution approaches for the d-CRN reachability problem.

An integer linear programming (ILP) problem is formulated as follows:

$$\text{ILP} \begin{cases} \min_x \{a^\top x\} \\ \text{subject to} \\ Ax \leq b \\ x \in \mathbb{Z}^n \end{cases} \quad (9)$$

where  $x$  is the vector of integer-valued decision variables,  $a \in \mathbb{R}^n$  is the coefficient vector of the objective function,  $A \in \mathbb{R}^{n \times k}$ ,  $b \in \mathbb{R}^k$  are constraint variables for some  $k \in \mathbb{Z}$ . Generally, an ILP problem is NP-hard, which restricts us to solve high-dimensional ILPs with efficient running time [30].

If the decision vector providing the extreme value for a linear problem is not important, just the existence of a feasible solution, then we obtain an ILP feasibility problem:

$$FP \left\{ \begin{array}{l} P = \{x \mid Ax \leq b, A \in \mathbb{Z}^{m \times n}, b \in \mathbb{Z}^m, x \in \mathbb{R}^n\} \\ P \cap \mathbb{Z}^n \stackrel{?}{=} \emptyset \end{array} \right. \quad (10)$$

The decision problem addresses the existence problem of an integer point in the polytopic constraint set defined by  $P$ . The feasibility problem is known to be NP-hard, however, it has well-decoupled time complexity with respect to the number of constraints  $k$  and the maximum of absolute values of the coefficients in  $A$  and  $b$ . Assuming fixed decision vector dimension  $n$ , the feasibility problem  $FP$  can be solved in polynomial time in the number of constraints  $k$  and the maximal absolute value of the constraints by the Lenstra algorithm [31, 32]. In addition, it is also possible to count the number of feasible integer points in the constraint set  $P$  using the Barvinok lattice point counting algorithm [33–36]. We note that there exists an efficient implementation of the Barvinok algorithm [37].

The FP problem in Eq. (10) can be viewed as a potential relaxation approach to an NP-hard ILP in situations where only the knowledge of the existence of an integer solution is enough. By making use of specific structural properties in an ILP problem, additional relaxation can be obtained. Total unimodularity – as a property of the coefficient matrix  $A$  – provides relaxation for the respective NP-hard ILP problem. Let us introduce the definition of totally unimodular matrices [30]:

**Definition 8.** A matrix  $A$  is totally unimodular if each sub-determinant of  $A$  is 0, +1, or -1.

Note that in a totally unimodular matrix all the entries are 0 or  $\pm 1$ .

If the coefficient matrix  $A$  in Eq. (9) is totally unimodular and  $b$  is an integer vector, then by relaxing the integer constraint on the decision variable vector  $x$ , the solution of the resulting linear program is proven to be optimal for the ILP [30].

The following proposition provides condition on the coefficient matrix to be totally unimodular:

**Proposition 1.** *A matrix  $A$  is totally unimodular, if there are no more than two non-zero entries in each column and the rows can be partitioned into two sets  $l_1$  and  $l_2$  so that:*

1. *if a column has two non-zero entries with the same sign, then they are in different partitions,  $l_1$  and  $l_2$ ;*
2. *if a column has two non-zero entries of different signs, then they are in the same partition, either  $l_1$  or  $l_2$ .*

### 3 Overview of the literature of d-CRN reachability

In this section we review the literature of the computational problem of d-CRN reachability. We will focus on the special classes of sub-and super-conservative reaction network structures.

The d-CRN state equation Eq. (7) is an IP decision problem with respect to  $c$ , provided the initial state  $X_0$ , the target state  $X'$  and the stoichiometric matrix  $\Gamma$ . The existence of a non-negative vector  $c$  satisfying the state equation is a necessary condition for the d-CRN reachability. This section overviews sub-classes of sub-and superconservative reaction network structures for which the existence of a non-negative integer  $c$  solution of the state equation is a sufficient and necessary condition of reachability. The importance of the latter condition is that the number of integer decision variables in Eq. (7) is given by the number of reactions of the d-CRN, that is the dimension of  $c$ . We note that in the general case the decision problem of d-CRN reachability requires auxiliary conditions to be added to the state equation and the number of decision variables is not

limited by the number of reactions in the system. The general NP-hard IP of d-CRN reachability can be relaxed to the form of IP decision problem. The relaxed problem has a reduced number of decision variables compared to that of the general IP formulation.

We employ the following auxiliary vector-valued function  $M = M(\Gamma^-)$  [23, 24]:

$$[M(\Gamma^-)]_i = \max\{[\Gamma^-]_{ij} : j = 1, \dots, l\}, \quad i = 1, \dots, n. \quad (11)$$

We note that the particular importance of the function  $M(\cdot)$  defined in Eq. (11) is that for any d-CRN  $\mathcal{N}$  and state  $X \in \mathbb{Z}_{\geq 0}^n$ ,  $X \succeq M(\Gamma_{\mathcal{N}}^-)$  implies that all the reactions of  $\mathcal{N}$  are charged at  $X$ .

The following proposition provides conditions on the d-CRN topology and the pair of initial and target states under which the non-negative integer solution of the discrete state equation is a sufficient and necessary condition of the reachability relation  $X_0 \rightsquigarrow_{\mathcal{N}} X'$ .

**Proposition 2.** [24] *Let us consider a subconservative or superconservative d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  with stoichiometric matrix  $\Gamma \in \{-1, 0, 1\}^{n \times l}$  and  $\Gamma^- \in \{0, 1\}^{n \times l}$  and  $\mathcal{C} = \mathcal{S} \cup \{\emptyset\}$ . Let us assume that for each  $r \in \mathcal{R}$ ,  $\sum_{i=1}^n [\bar{y}^+]_i \leq 1$  and  $\sum_{i=1}^n [\bar{y}^-]_i = 1$ . Let us consider two arbitrary states,  $X_0, X' \in \mathbb{Z}_{\geq 0}^n$  so that  $X_0 \succeq M$ ,  $X' \succeq M$  where  $M = M(\Gamma^-)$  is defined by Eq. (11) Then the reachability relation  $X_0 \rightsquigarrow_{\mathcal{N}} X'$  hold if and only if there exists a vector  $c \in \mathbb{Z}_{\geq 0}^l$  satisfying the dCRN state equation  $X_0 + \Gamma_{\mathcal{N}} c = X'$ .*

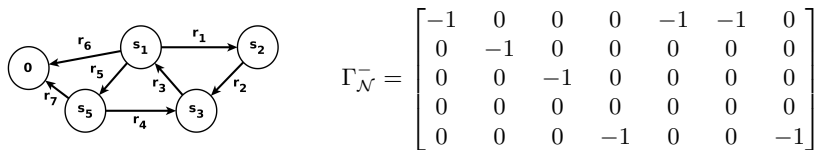
Proposition 2 leads to the following integer feasibility problem [24]:

$$\begin{cases} \Gamma_{\mathcal{N}} c = X' - X_0 \\ c \in \mathbb{Z}_{\geq 0}^l \end{cases} \quad (12)$$

The ILP of Eq. (12) can be solved using the implementation of Barvinok algorithm called LattE [37].

**Example 2.** Consider the reaction network in Fig. 3. For any pair of states,  $X_0, X' \in \mathbb{Z}_{\geq 0}^n$ ,  $X_0 \succeq M(\Gamma_{\mathcal{N}}^-)$ ,  $X' \succeq M(\Gamma_{\mathcal{N}}^-)$ , the existence of a non-negative integer solution  $c \in \mathbb{Z}_{\geq 0}^l$  of the respective state equation

$X_0 + \Gamma_{\mathcal{N}}c = X'$  implies that the reachability relation  $X_0 \rightsquigarrow_{\mathcal{N}} X'$  holds. Note that  $0$  denotes the zero complex, that is the reaction  $s_1 \rightarrow 0$  results in the reduction of the species  $s_1$  with one molecule.



**Figure 3.** A subconservative d-CRN  $\mathcal{N}$  satisfying the conditions of Proposition 2.

Proposition 2 can be extended to a wider range of reaction network structures by allowing catalyzer species in the reactions as follows.

**Proposition 3.** [24] *Let us consider a subconservative d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  with stoichiometric matrix  $\Gamma_{\mathcal{N}} \in \{-1, 0, 1\}^{n \times l}$  and  $\Gamma^- \in \{0, 1\}^{n \times l}$ . Assume that for each  $r \in \mathcal{R}$ :*

1.  $r = s_1 \rightarrow s_2$  for  $s_1, s_2 \in \mathcal{S}$ ,  $s_1 \neq s_2$ ,  $s_1 \neq 0$ , or
2.  $r = s + s_1 \rightarrow s + s_2$  for  $s, s_1, s_2 \in \mathcal{S}$ ,  $s \neq s_1 \neq s_2$ ,  $s \neq 0$ ,  $s_1 \neq 0$  and  $s$  is not consumed by any reaction  $r \in \mathcal{R}$ .

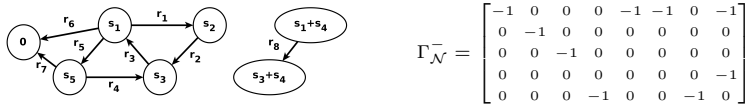
*Let us consider a pair of states  $X_0, X' \in \mathbb{Z}_{\geq 0}^n$  so that  $x_0 \succeq M$  and  $X' \succeq M$  where  $M = M(\Gamma^-)$  is defined by Eq. (11). The relation  $X_0 \rightsquigarrow_{\mathcal{N}} X'$  holds if and only if there exists a non-negative integer solution  $c \in \mathbb{Z}_{\geq 0}^l$  for which  $X_0 + \Gamma_{\mathcal{N}}c = X'$  is satisfied.*

**Proposition 4.** [24] *Let us consider a superconservative d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  with stoichiometric matrix  $\Gamma_{\mathcal{N}} \in \{-1, 0, 1\}^{n \times l}$  and  $\Gamma^- \in \{0, 1\}^{n \times l}$ . Assume that for each  $r \in \mathcal{R}$ :*

1.  $r = s_1 \rightarrow s_2$  for  $s_1, s_2 \in \mathcal{S}$ ,  $s_1 \neq s_2$ ,  $s_2 \neq 0$ , or
2.  $r = s + s_1 \rightarrow s + s_2$  for  $s, s_1, s_2 \in \mathcal{S}$ ,  $s \neq s_1 \neq s_2$ ,  $s \neq 0$ ,  $s_2 \neq 0$  and  $s$  is not consumed by any reaction  $r \in \mathcal{R}$ .

Let us consider a pair of states  $X_0, X' \in \mathbb{Z}_{\geq 0}^n$  so that  $x_0 \succeq M$  and  $X' \succeq M$  where  $M = M(\Gamma^-)$  is defined by Eq. (11). The relation  $X_0 \rightsquigarrow_{\mathcal{N}} X'$  holds if and only if there exists a non-negative integer solution  $c \in \mathbb{Z}_{\geq 0}^l$  for which  $X_0 + \Gamma_{\mathcal{N}}c = X'$  is satisfied.

**Example 3.** Consider the reaction network in Fig. 4. It is visible that by reversing the edges of the reaction graph we obtain a superconservative reaction network structure that satisfies Proposition 4.



**Figure 4.** A subconservative d-CRN  $\mathcal{N}$  for which the conditions of Proposition 3 hold.

## 4 Main results

In this section we extend the d-CRN reachability results discussed in the previous section. We show that the ILP problem of Eq. (12) can be relaxed to a linear program running in polynomial time under the same conditions on the initial and target states and the reaction network structure. This way a computational method is provided for the d-CRN reachability problem with polynomial time complexity and theoretical guarantee.

**Proposition 5.** Let us consider a subconservative or superconservative d-CRN  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$  with stoichiometric matrix  $\Gamma_{\mathcal{N}} \in \{-1, 0, 1\}^{n \times l}$  and  $\Gamma_{\mathcal{N}}^- \in \{0, 1\}^{n \times l}$  and  $\mathcal{C} = \mathcal{S} \cup \{\emptyset\}$ . Let us assume that for each  $r \in \mathcal{R}$ ,  $\sum_{i=1}^n [\bar{y}^+]_i \leq 1$  and  $\sum_{i=1}^n [\bar{y}^-]_i = 1$ . Let us consider two arbitrary states,  $X_0, X' \in \mathbb{Z}_{\geq 0}^n$  so that  $X_0 \succeq M$ ,  $X' \succeq M$  where  $M = M(\Gamma^-)$  is defined by Eq. (11). Then the reachability relation  $X_0 \rightsquigarrow_{\mathcal{N}} X'$  can be decided in polynomial time.

*Proof.* The conditions imply that the existence of a  $c \in \mathbb{Z}_{\geq 0}^l$  for which  $X_0 + \Gamma_{\mathcal{N}}c = X'$  holds is a sufficient and necessary condition of the reachability relation  $X_0 \rightsquigarrow_{\mathcal{N}} X'$ .

We make use of Proposition 1 to show that the stoichiometric matrix  $\Gamma_{\mathcal{N}}$  is totally unimodular. The following propositions hold for  $\Gamma_{\mathcal{N}}$ :

1. Every entry of  $\Gamma_{\mathcal{N}}$  is 0, 1 or  $-1$ .
2. Every column of  $\Gamma_{\mathcal{N}}$  contains at most 2 non-zero entries.
3. The rows of  $\Gamma_{\mathcal{N}}$  can be partitioned into two disjoint subsets  $\mathcal{S}_1$  and  $\mathcal{S}_2$  so that:
  - (a) if two entries in a column of  $\Gamma_{\mathcal{N}}$  have the same sign, then one is in  $\mathcal{S}_1$  while the other one is in  $\mathcal{S}_2$ ;
  - (b) if two entries in a column of  $\Gamma_{\mathcal{N}}$  have the opposite sign, then they are in the same subset  $\mathcal{S}_1$  or  $\mathcal{S}_2$ .

Clearly, each entry in  $\Gamma_{\mathcal{N}}$  equals to  $+1$ ,  $-1$  or  $0$ . Each reaction consumes at most one species and produces at most another one, that is each column of  $\Gamma_{\mathcal{N}}$  has at most 2 non-zero entries. Multiple entries of the same sign in a column of  $\Gamma_{\mathcal{N}}$  would imply that different species are consumed or produced by a reaction, but this is not possible in the considered class of reaction networks. Considering any column of  $\Gamma_{\mathcal{N}}$ , the (at most) two rows containing non-zero entries must be in the same set ( $\mathcal{S}_1$  or  $\mathcal{S}_2$ ). If the reaction network graph is connected, then all the rows are put in the same set. Let us assume that the reaction network graph is not connected. In this case there exist 2 or more linkage classes in the reaction network graph. Note that the linkage classes are not necessarily strongly connected. The linkage classes cover disjoint sets of rows in  $\Gamma_{\mathcal{N}}$  and for each linkage class we can choose arbitrarily, either  $\mathcal{S}_1$  or  $\mathcal{S}_2$ , irrespective of the other linkage classes. Note that empty rows are not possible as we assume that isolated species are not allowed.

The above proof implies that  $\Gamma_{\mathcal{N}}$  is guaranteed to be totally unimodular [38]. Clearly, for a totally unimodular matrix  $\Gamma_{\mathcal{N}}$  the following LP provides an optimal integer solution:

$$\begin{cases} \min_c \{a^\top c\} \\ \text{subject to} \\ \Gamma_{\mathcal{N}} c = X' - X_0 \\ c \in \mathbb{R}_{\geq 0}^l \end{cases} \quad (13)$$

for any  $a \in \mathbb{R}^l$ . ■

The practical importance of Proposition 5 is that the reachability relation can be decided by a linear program, which has polynomial time complexity with respect to the state space dimensionality ( $n$ ). The IP feasibility formulation employed the Lenstra and Barvonok algorithms and assumed fixed state space dimensionality to obtain polynomial time complexity in terms of the number of reactions and the maximum absolute value entry of the constraint matrices. However, the IP feasibility approach was exponential in the state space dimensionality (number of species).

Proposition 5 provides theoretical guarantees that the feasibility of the relaxed LP implies reachability, while the infeasibility implies that the reachability relation does not hold.

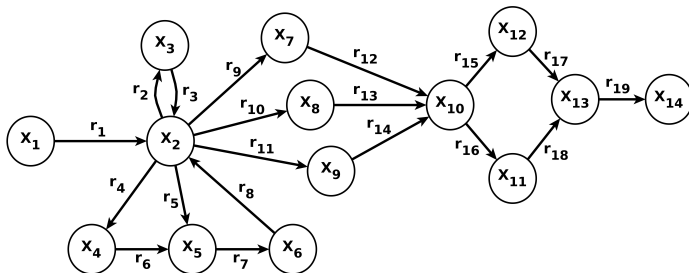
We can naturally extend Proposition 3 and Proposition 4 by making use of the totally unimodular property of the stoichiometric matrix  $\Gamma_{\mathcal{N}}$  similarly as we obtained Proposition 5. Finally, we note that by linear programming, the infeasibility can be decided with high accuracy. However, in the case of ILP problems, the infeasibility returned by a solver conveys lower accuracy.

**Example 4.** In this example an illustrative reaction network example is provided from the literature of Chemical Reaction Networks [39]. Fig. 5 depicts a conservative reaction network structure  $\mathcal{N}$  of 14 metabolites. The stoichiometric matrix associated to the system is given by Eq. (14). A suitable conservation vector for the d-CRN is given by  $z = 1^n$ :  $z^\top \Gamma_{\mathcal{N}} = 0^{1 \times l}$ . Clearly, for any non-negative initial state, the reachable state space  $Reach(\mathcal{N}, X_0)$  is an  $(n - 1)$ -dimensional hyperplane.

$\mathcal{N}$  has a monomolecular reaction network structure with totally unimodular stoichiometric matrix. This implies that Proposition 5 can be



applied. For any pair of state vectors  $X_0, X' \in \mathbb{Z}_{\geq 0}^n$ ,  $X_0, X' \succeq M(\Gamma_{\mathcal{N}}^-)$ ,  $M(\Gamma_{\mathcal{N}}^-) = 1^n$ , the reachability relation  $X_0 \rightsquigarrow X'$  can be validated in polynomial time by running a linear program of the form of Eq. (13) with some non-zero  $a \in \mathbb{R}^l$ .



**Figure 5.** A conservative d-CRN  $\mathcal{N}$  for which Proposition 5 holds.  $\mathcal{N}$  is a monomolecular reaction network with totally unimodular stoichiometric matrix. This implies that the generally NP-hard problem of deciding the reachability relation can be relaxed to a linear program with guaranteed polynomial time complexity.

$$\Gamma_{\mathcal{N}} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & -1 & -1 & 0 & 0 & 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (14)$$

Table 1 depicts the computational comparison of different algorithms proposed for deciding the d-CRN reachability. Eq. (12) and Eq. (13) were implemented in Python 3.7. using the Gurobi solver [40]. We used the LattE implementation of the Barvinok's algorithm. Note that Proposition

5 provides theoretical guarantee that the LP formulation of Eq. (13) has integer solution. Clearly, the feasibility of (12) and Eq. (13) implies the d-CRN reachability. In case **A**, the LP formulation can decide the reachability problem within half of the time of the IP solution.

Case	Continuous	Integer	Barvinok
A	0.1 [s]	0.23 [s]	>60 [min]
B	0.01 [s]	0.01 [s]	>60 [min]
C	<0.01 [s]	<0.01 [s]	>60 [min]

**Table 1.** Computational comparison of the different methods proposed for deciding the d-CRN reachability. **Continuous** represents the LP given by Eq. (13), **Integer** denotes the integer program Eq. (12) and **LattE** stands for the implementation of the Barvinok’s algorithm [37]. For each case  $X_0 = 200^{14 \times 1}$ . In Case **A**,  $[X_0]_1 = 5$ ,  $[X_1]_4 = 395$ , otherwise  $[X_i] = 200$  for  $i = 2, \dots, 13$ . In case **B**,  $[X_0]_1 = 100$ ,  $[X_1]_4 = 300$ , otherwise  $[X_i] = 200$  for  $i = 2, \dots, 13$ . Finally, in case **C**,  $[X_0]_1 = 195$ ,  $[X_1]_4 = 205$ , otherwise  $[X_i] = 200$  for  $i = 2, \dots, 13$ .

## 5 Summary

In this paper the reachability problem of discrete state Chemical Reaction Networks (d-CRNs) was studied. Given a reaction network  $\mathcal{N}$  and a pair of initial and target states  $X_0, X' \in \mathbb{Z}_{\geq 0}^n$ ,  $X'$  is said to be reachable from  $X_0$  ( $X_0 \rightsquigarrow_{\mathcal{N}} X'$ ), if there exists a non-negative trajectory in the state space along which the system can be driven to  $X'$  from  $X_0$  using the available set of reactions of the reaction network  $\mathcal{N}$ . We assumed that the reaction networks are sub-or superconservative with additional constraints on their structure. We used the result that the reachability relation is equivalent to the non-negative integer solution of the d-CRN state equation in the considered sub-classes of reaction network structures [24]. We proved that in the studied sub-classes of d-CRNs, the stoichiometric matrix is guaranteed to be totally unimodular. As the reachability problem can be decided by validating the existence of a non-negative integer solution of the d-CRN state equation, an IP decision problem can be applied to solve it. Making use of the totally unimodular property of the stoichiometric matrix, we relaxed the IP decision problem to a linear programming problem of con-

tinuous variables. This way it is proved that in the considered sub-classes of sub- and superconservative d-CRNs, the reachability relation can be decided in polynomial time by formulating a linear programming problem in which the number of decision variables equals to the number of reactions in the d-CRN. Clearly, a feasible solution of the resulting linear program implies the reachability relation. An example of a metabolic network taken from the literature was provided to illustrate the obtained result.

**Acknowledgment:** This project was partially supported by the National Research, Development and Innovation Office (NKFIH) through grants no. 131545 and the Thematic Excellence Programme (TKP2020-NKA-11). Gergely Szlobodnyik is supported by the UNKP-21-4 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund.

## References

- [1] M. Feinberg, *Foundations of Chemical Reaction Network Theory*, Springer, 2019.
- [2] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman & Hall, 2007.
- [3] W. Liu, *Introduction to Modeling Biological Cellular Control Systems*, Springer, 2012.
- [4] D. Angeli, A tutorial on chemical reaction network dynamics, *Eur. J. Control* **15** (2009) 398–406.
- [5] P. Érdi, J. Tóth, *Mathematical Models of Chemical Reactions. Theory and Applications of Deterministic and Stochastic Models*, Manchester Univ. Press, Manchester, 1989.
- [6] S. Smith, R. Grima, Single-cell variability in multicellular organisms, *Nature Commun.* **9** (2018) #345.
- [7] D. F. Anderson, T. G. Kurtz, *Stochastic Analysis of Biochemical Systems*, Springer, Berlin, 2015.
- [8] D. F. Anderson, T. G. Kurtz, Continuous time Markov chain models for chemical reaction networks, in: H. Koepl, G. Setti, M. di Bernardo, D. Densmore (Eds.), *Design and Analysis of Biomolecular*

---

*Circuits: Engineering Approaches to Systems and Synthetic Biology*, Springer, Berlin, 2011. pp. 3–42.

- [9] F. Bause, P. S. Kritzinger, *Stochastic Petri Nets: An Introduction to the Theory*, Vieweg Verlag 2. Aufl., 2002.
- [10] B. Ács, G. Szederkényi, Z. Tuza, Z. A. Tuza, Computing all possible graph structures describing linearly conjugate realizations of kinetic systems, *Comput. Phys. Commun.* **204** (2016) 11–20.
- [11] B. Ács, G. Szlobodnyik, G. Szederkényi, A computational approach to the structural analysis of uncertain kinetic systems, *Comput. Phys. Commun.* **228** (2018) 83–95.
- [12] M. D. Johnston, D. F. Anderson, G. Craciun, R. Brijder, Conditions for extinction events in chemical reaction networks with discrete state spaces, *J. Math. Biol.* **76** (2018) 1535–1558.
- [13] M. D. Johnston, A computational approach to extinction events in chemical reaction networks with discrete state spaces, *Math. Biosci.* **294** (2017) 130–142.
- [14] T. Murata, Petri nets: properties, analysis and applications, *Proceed. IEEE* **77** (1989) 541–580.
- [15] G. Memmi, G. Roucairol, Linear algebra in net theory, in: W. Brauer (Ed.), *Net Theory and Applications*, Springer, 1975, pp. 213–223.
- [16] E. W. Mayr, An algorithm for the general petri net reachability problem, *STOC* (1981) 238–246.
- [17] S. R. Kosaraju, Decidability of reachability in vector addition systems, *STOC* (1982) 267–281.
- [18] J.L. Lambert, A structure to decide reachability in Petri nets, *Theor. Comput. Sci.* **99** (1992) 79–104.
- [19] J. Leroux, Vector addition reachability problem (a simpler solution), in: G. Sutcliffe (Ed.), *The Alan Turing Centenary Conference*, Easy-Chair, 2012, pp. 214–228.
- [20] R. J. Lipton, The reachability problem requires exponential space, Tech. Rep. 62, Dep. Comput. Sci., Yale Univ., 1976.
- [21] A. Case, J. H. Lutz, D. M. Stull, Reachability problems for continuous chemical reaction networks, *Nat. Comput.* **17** (2018) 223–230.

- 
- [22] S. Schmitz, Complexity Hierarchies beyond Elementary, *ACM Trans. Comput. Theory* **8** (2016) 1–36.
- [23] G. Szlobodnyik, G. Szederkényi, M. Johnston, Reachability analysis of subconservative discrete chemical reaction networks, *MATCH Commun. Math. Comput. Chem.* **81** (2019) 705–736.
- [24] G. Szlobodnyik, G. Szederkenyi, Reachability analysis of low-order discrete state reaction networks obeying conservation laws, *Complexity* **2019** (2019) #1035974.
- [25] A. Mahdi, A. Ferragut, C. Valls, C. Wiuf, Conservation laws in biochemical reaction networks, *SIAM J. Appl. Dyn. Syst.* **16** (2017) 2213–2232.
- [26] D. Angeli, P.D. Leenheer, E. Sontag, A Petri net approach to persistence analysis in chemical reaction networks, in: I. Queinnec, S. Tarbouriech, G. Garcia, S. I. Niculescu (Eds.), *Biology and Control Theory: Current Challenges*, Springer, Berlin, 2007, pp. 181–216.
- [27] D. Angeli, On modularity and persistence of chemical reaction networks, *IEEE Conference on Decision and Control*, 2008, pp. 2650–2655.
- [28] M. D. Johnston, D. Anderson, Weak dynamic non-emptiability and persistence of chemical kinetics systems, *SIAM J. Appl. Math.* **71** (2011) 1263–1279.
- [29] A. Condon, D. Harel, J. N. Kok, A. Salomaa, E. Winfree (Eds.), *Algorithmic Bioprocesses*, Springer, 2009.
- [30] A. Schrijver, *Theory of Linear and Integer Programming*, Wiley, New York, 1999.
- [31] H. W. Lenstra, Integer programming with a fixed number of variables, *Math. Oper. Res.* **8** (1983) 538–548.
- [32] M. Grötschel, L. Lovász, A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, Springer, Berlin, 1988.
- [33] A. J. Barvinok, A polynomial time algorithm for counting integral points in polyhedra when the dimension is fixed, *Math. Oper. Res.* **19** (1994) 769–779.
- [34] M. Köppe, A primal Barvinok algorithm based on irrational decompositions, *SIAM J. Discr. Math.* **21** (2007) 220–236.

- 
- [35] S. Verdoolaege, R. Seghir, K. Beyls, V. Loechner, M. Bruynooghe, Counting integer points in parametric polytopes using Barvinok's rational functions, *Algorithmica* **48** (2007) 37–66.
- [36] J. B. Lasserre, Integer programming, Barvinok's counting algorithm and Gomory relaxations, *Oper. Res. Lett.* **32** (2004) 133–137.
- [37] J. A. De Loera, R. Hemmecke, J. Tauzer, R. Yoshida, Effective lattice point counting in rational convex polytopes, *J. Symb. Comput.* **38** (2004) 1273–1302.
- [38] I. Heller, C. B. G. Tompkins, An extension of a theorem of Dantzig's, in: H. W. Kuhn, A. W. Tucker (Eds.), *Linear Inequalities and Related Systems, Annals of Mathematics Studies*, Princeton Univ. Press, Princeton, 1956, pp. 247–254.
- [39] N. Vassena, H. Matano, Monomolecular reaction networks: Flux-influenced sets and balloons, *Math. Meth. Appl. Sci.* **40** (2017) 7722–7736.
- [40] Gurobi Optimization Inc., Gurobi Optimizer Reference Manual, [www.gurobi.com](http://www.gurobi.com), 2016.