MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **95** (2026) 565–587
ISSN: 0340–6253
doi: 10.46793/match.95-3.29125

Polynomial Time Reachable Sequence Computation in Discrete State Reaction Networks

Gergely Szlobodnyik[a,*], **Gábor Szederkényi**[a,b]

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

[b] *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 November 5, 2025)

Abstract

In this paper a computationally efficient solution is proposed for the reachable sequence computation problem in the context of sub and superconservative discrete state Chemical Reaction Network (CRN) subclasses. It is shown that a minimal-length reachable state transition sequence and reaction vector sequence can be determined in polynomial time, assuming that the reachability relation holds. A constructive proof is provided in which a computational procedure is composed to find a reachable state transition sequence, provided a pair of initial and target states. In the studied subclasses of discrete state CRNs it is known that the reachability problem – as a decision problem – can be decided in polynomial time by solving a Linear Program formulated by means of the CRN state equation. Leveraging the LP relaxation, the constructed procedure formulates the computational problem of feasible sequence calculation by iteratively solving Linear Programs related to the decision problem of CRN reachability.

*Corresponding author.

1 Introduction

Formal models of Chemical Reaction Networks (CRNs) serve as a fundamental mathematical framework for modeling and analyzing the dynamic behavior and structural properties of a diverse range of networked systems, including classical chemical and biochemical processes, gene regulatory networks, protein-protein interactions, population dynamics, epidemiological systems and synthetic biological circuits [1–3, 18, 20]. In scenarios where the interacting species are assumed to be well-mixed and present in large molecular counts in the environment, their dynamical behavior can be modeled using Ordinary Differential Equations (ODEs) of continuous variables [2, 3]. These continuous-state models can effectively capture the dynamical behavior of homogeneous systems; however, they are often inadequate when the molecular counts are low (e.g. in the order of magnitude of 100 molecules per species). Under such conditions, stochastic effects become significant, necessitating the use of discrete-state models to accurately describe the dynamics [4, 11]. The formal mathematical model of these discrete-state systems is referred to as discrete Chemical Reaction Networks (d-CRNs).

The relationship between the structural properties and dynamical behavior of CRNs is a commonly studied field of mathematical and computational chemistry. In the case of continuous-state CRNs, it is proven that the same set of ODEs might be realized by structurally different reaction networks [7, 8]. The structural non-uniqueness implies that identical dynamical behavior can emerge from different interaction topologies. In the case of d-CRNs, reachability plays an important role in the analysis of qualitative behavior. Provided a pair of initial and target states, the reachability problem determines whether the target state can be reached from the given initial state via a finite sequence of reactions (state transitions) [13,14]. The reachability relation is determined by the stoichiometry and the reaction network structure and is independent of the specific reaction rates (propensities) in the integro-differential equations governing the discrete system's dynamics. The reachability analysis provides a foundation for investigating system dynamics, including the study of extinction

events where the count of a particular species irreversibly drops to zero. Extinction events are proven to be determined by the structural properties of the underlying network [9, 10, 14].

The reachability problem also has significant implications for the design and implementation of synthetic biological circuits. In particular, the gate implementability problem, which seeks to construct reaction networks capable of realizing specific logical functions, can be reformulated as a reachability problem [16]. This reformulation enables the systematic design of d-CRNs that emulate Boolean logic gates through carefully engineered chemical reactions.

A central focus in the analysis of d-CRN reachability is the computation of reachable sequences, which returns a specific sequence of reactions driving the d-CRN from a prescribed initial state to the target. These computations are essential for elucidating certain dynamic behaviors, such as extinction, and have practical applications in the design of synthetic biological systems [17]. For instance, identifying reachable sequences of minimum length enables the prediction and control of state transitions, which is critical for ensuring robustness in synthetic circuits.

In this paper the reachability problem of d-CRNs is studied. We provide a computational procedure based on which a minimum length reachable state transition sequence can be computed. Provided a pair of initial and target states, it is shown that a reachable state transition sequence of minimum length can be determined in polynomial time in the studied subclasses of d-CRNs. The constructed computational procedure provides theoretical guarantee for reachable sequence computation in well-defined sub-classes of sub-and superconservative reaction network structures.

2 Preliminaries

2.1 Mathematical notations

This section summarizes the formal mathematical notations used in the paper.

$\mathbb{R}$	the set of real numbers,
$\mathbb{Z}$	the set of integer numbers,
$\mathbb{N}$	the set of natural numbers including 0,
$\emptyset$	empty set,
$\mathbb{P}^k_{\geq 0}$	the set of non-negative k-dimensional from the set $\mathbb{P}$,
$A^{n \times m}$	the set of matrices from a set A of n rows and m columns,
$A^\top$	the transpose of the matrix A,
1^l	a vector of dimension l with ones at each entry,
$0^{l \times k}$	a matrix of dimension $(l \times k)$ with zero entries,
$[M]_{ij}$	the entry of M with row index i and column index j,
$X \succeq Y$	element-wise comparison of the vectors X and Y,
$\bigoplus$	concatenation operator,
$BREAK$	command that terminates a loop at any iteration

Table 1. Mathematical Notations

2.2 Discrete state chemical reaction networks

This section discusses the mathematical preliminaries and formal definitions with respect to discrete state Chemical Reaction Networks. Definitions and notations follow our previous work with respect to the study of the reachability decision problem in sub-and superconservative d-CRNs [13, 14].

Definition 1. [13, 14] A Chemical Reaction Network on a discrete state space (d-CRN) of l reactions, m complexes and n species is defined by triplet $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$.

$\mathcal{S}$ defines the set of species:

$$\mathcal{S} = \Big\{ s_i \mid i = 1, \dots n \Big\}. \tag{1}$$

$\mathcal{C}$ defines the set of complexes:

$$\mathcal{C} = \Big\{ y_j = \sum_{i=1}^{n} \alpha_{ji} s_i \mid s_i \in \mathcal{S},\ \alpha_{ji} \in \mathbb{Z}_{\geq 0},\ i = 1, \dots n,\ j = 1 \dots m \Big\}. \tag{2}$$

$\mathcal{R}$ gives the set of reactions:

$$\mathcal{R} = \left\{ r_k = y_{source(r_k)} \to y_{product(r_k)} \mid y_{source(r_k)},\ y_{product(r_k)} \in \mathcal{C},\ k = 1, \dots l \right\} \tag{3}$$

In the sequel we assume a fixed ordering of the species, complexes and reaction.

The integer α_{ji} encodes the stoichiometric coefficient of the species $s_i \in \mathcal{S}$ in the complex $c_j \in \mathcal{C}$.

For any reaction $r \in \mathcal{R}$, we introduce the notation $r = y_{source(r)} \to y_{product(r)}$ where $y_{source(r)}$ and $y_{product(r)}$ denote the source complex and the product complex, respectively.

For the complex $y_j \in \mathcal{C}$ $j = 1,\ \dots\ m$, we employ the following vector representation:

$$\overline{y}_j = [\alpha_{j1}\ \alpha_{j2}\ \dots\ \alpha_{jn}\ ]^\top. \tag{4}$$

Leveraging the above notation, the reaction vector $r_{ij} \in \mathbb{Z}^n_{\geq 0}$ representation is used for each $r \in \mathcal{R}$ as follow:

$$r_{ij} = \overline{y}_j - \overline{y}_i, \tag{5}$$

with $y_i \in \mathcal{C}$ and $y_j \in \mathcal{C}$. Throughout the paper we employ r_k to denote both the k'th reaction of $\mathcal{R}$ and its reaction vector representation.

A d-CRN $\mathcal{N}$ can be equivalently represented as a directed graph $G_{\mathcal{N}} = G_{\mathcal{N}}(V, E)$, where the vertices correspond to the complexes, and the directed edges represent the reactions. In this representation, edges originate from the source complex and terminate at the product complex of the corresponding reaction. Formally, for a d-CRN $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$, the reaction graph is defined as $G_{\mathcal{N}} = G_{\mathcal{N}}(V, E)$, where $V = \mathcal{C}$ and $E = \mathcal{R}$. Furthermore, for each reaction $r \in \mathcal{R}$, expressed as $r = y_i \to y_j$, there exists a unique edge $e \in E$ such that e originates from the vertex corresponding to y_i and points to the vertex associated with y_j.

Throughout the paper the terms structure and topology are used interchangeably to refer to the reaction network graph $G_{\mathcal{N}}$ of a d-CRN $\mathcal{N}$. Note that $G = G_{\mathcal{N}}$ is unique with respect to a reaction network $\mathcal{N}$.

We assume that there is no loop reaction in $G_{\mathcal{N}}$, that is there is no $r \in \mathcal{R}$ for which $y = c \to c$ for some $c \in \mathcal{C}$. In addition, isolated complexes are not allowed, that is there is no complex $y \in \mathcal{C}$ for which there is no reaction $r \in \mathcal{R}$ with $y = source(r)$ or $y = target(r)$.

Now we introduce the tools and mathematical framework for the algebraic representation of d-CRNs.

Definition 2. [13, 14] For a d-CRN $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ the stoichiometric matrix $\Gamma_{\mathcal{N}} \in \mathbb{Z}^{n \times l}$ is defined as follows:

$$\Gamma_{\mathcal{N}} = [r_1 \ r_2 \ \dots \ r_l]. \tag{6}$$

In addition, We use the notation $\Gamma^+_{\mathcal{N}}$ as follows:

$$\Gamma^+_{\mathcal{N}} = [\overline{y}^+_{r_1} \ \dots \ \overline{y}^+_{r_l}], \tag{7}$$

where $\overline{y}^+_{r_k}$ denotes the vector representation of the product complex of the reaction r_k, $[\overline{y}^+_{r_k}]_i$ is the stoichiometric coefficient of the s_i.

Equivalently, $\Gamma^-_{\mathcal{N}}$ is used as follows:

$$\Gamma^-_{\mathcal{N}} = [\overline{y}^-_{r_1} \ \dots \ \overline{y}^-_{r_l}], \tag{8}$$

where $\overline{y}^-_{r_k}$ is the vector representation of the source complex of $r_k \in \mathcal{R}$.

Then for any $\mathcal{N}$ the following matrix equation holds

$$\Gamma_{\mathcal{N}} = \Gamma^+_{\mathcal{N}} - \Gamma^-_{\mathcal{N}}. \tag{9}$$

Note that the stoichiometric matrix encodes the net species count change in teh reactions, in particular, $[\Gamma]_{ij}$ determines the net species count change of $s_i \in \mathcal{S}$ in the reaction $r_j \in \mathcal{R}$.

Definition 3. [13,14] For a d-CRN $\mathcal{N}$, the state vector $X \in \mathbb{Z}^n_{\geq 0}$ is defined so that $[X]_i$ denotes the molecular multiplicity (count) of the ith species in state X, formally:

$$[X]_i : S_i \to \mathbb{Z}_{\geq 0}, \text{ for } i = 1, \ \dots \ n, \ s_1 \in \mathcal{S}. \tag{10}$$

Now we introduce the state equation of d-CRNs which is used to algebraically describe the state transitions or equivalently the firing of the reactions.

Definition 4. [13,14] Let us consider a state vector $X_0 \in \mathbb{Z}^n_{\geq 0}$, the state transitions of the d-CRN are governed by the discrete state equation:

$$X' = X_0 + \Gamma_{\mathcal{N}} c, \tag{11}$$

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

Definition 5. [13,14] Let us consider a d-CRN $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ of n species, m complexes and l reactions.

1. A complex $y \in \mathcal{C}$ is said to be charged at a state $X \in \mathbb{Z}^n_{\geq 0}$, if $X \succeq \overline{y}$, where $\overline{y}$ is the vector representation of y. A reaction $r \in \mathcal{R}$ is said to be charged at a state X if its source complex is charged.

2. A reaction vector sequence σ_r is an ordered set of reactions $\sigma_r = r_{\nu_{\mathcal{R}}(1)} \ \dots \ r_{\nu_{\mathcal{R}}(v)}$ with $v \in \mathbb{Z}_{\geq 1}$ and $\nu_{\mathcal{R}}$ is an index function mapping the index $k \in \mathbb{N}$ to the index of the respective reaction, given a fixed ordering of the reactions in $\mathcal{R}$.

3. A state $X_1 \in \mathbb{Z}^n_{\geq 0}$ reacts to the state $X_2 \in \mathbb{Z}^n_{\geq 0}$ with respect to $\mathcal{N}$, if there exists a reaction $r \in \mathcal{R}$ so that $X_1 + r = X_2$. The following notation is employed: $X_1 \rightarrow_{\mathcal{N}} X_2$.

4. A state (transition) sequence σ_X is an ordered sequence of states $\sigma_X = X_1 \dots \ X_v$ with $v \in \mathbb{Z}_{\geq 1}$ so that $X_i \rightarrow X_{i+1}$ for $i = 1, \ \dots \ v-1$.

5. Consider two states, $X_0, \ X' \in \mathbb{Z}^n_{\geq 0}$. We say that X' can be reached (reachable) from the state X_0 ($X_0 \rightsquigarrow_{\mathcal{N}} X'$), if there exists a sequence states $\sigma_X = X_{\nu(1)} \dots \ X_{\nu(v)}$ for which $X_{\nu(1)} = X_0$, $X_{\nu(v)} = X'$ and for all $X \in \sigma_X$, $X \succeq 0^n$.

6. Consider two states, $X_0,\ X' \in \mathbb{Z}^n_{\geq 0}$. It is said that the state X' is coverable from the state X_0, if there exists a state $\hat{X} \in \mathbb{Z}^n_{\geq 0}$ for which the reachability relation holds $X_0 \rightsquigarrow_{\mathcal{N}} \hat{X}$ and the following relation is satisfied $X' \succeq \hat{X}$.

7. A species $s \in \mathcal{S}$ is said to be a catalyzer with respect to a reaction $r \in \mathcal{R}$ if $r = s + s_1 \to s + s_2$ for some $s_1,\ s_2 \in \mathcal{S},\ s_1 \neq s_2$.

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

Let us consider a d-CRN $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ and a pair of states $X_0, X' \in \mathbb{Z}^n_{\geq 0}$. The reachability relation $X_0 \rightsquigarrow_{\mathcal{N}} X'$ requires the following necessary condition:

$$X_0 + \Gamma_{\mathcal{N}} c = X',\ c \in \mathbb{Z}^l_{\geq 0}. \tag{12}$$

That is the existence of a non-negative integer $c \in \mathbb{Z}^l_{\geq 0}$ to the state equation is a necessary condition to the reachability relation.

However, the existence of such a non-negative integer solution c does not, in general, guarantee that the reachability relation holds [15].

In this work, we utilize results on d-CRN reachability obtained for specific subclasses of reaction network structures (topologies). For these subclasses, the reachability problem, formulated as a decision problem, has been shown to be equivalent to the existence of a non-negative integer solution to the d-CRN state equation [13, 14].

Next we introduce the definitions of sub-and superconservativity.

Definition 6. [13, 14] Let us consider a d-CRN $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ with $\Gamma_{\mathcal{N}} \in \mathbb{Z}^{n \times l}$. $\mathcal{N}$ is said to be subconservative (superconservative), if there exists $z \in \mathbb{R}^l_{>0}$ for which $z^\top \Gamma_{\mathcal{N}} \leq 0^{1 \times l}$ ($z^\top \Gamma_{\mathcal{N}} \geq 0^{1 \times l}$).

Definition 7. [13, 14] Let us consider a d-CRN $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ with $\Gamma_{\mathcal{N}} \in \mathbb{Z}^{n \times l}$. $\mathcal{N}$ is called superconservative, if there exists a vector $z \in \mathbb{R}^l_{>0}$ for which $z^\top \Gamma_{\mathcal{N}} \geq 0^{1 \times l}$.

Definition 8. [13, 14] A d-CRN $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$, with a stoichiometric matrix $\Gamma_{\mathcal{N}} \in \mathbb{Z}^{n\times l}$, is defined as *conservative* if there exists a vector $z \in \mathbb{R}^l_{>0}$ such that the condition

$$z^\top \Gamma_{\mathcal{N}} = 0^{1\times l}$$

is satisfied.

In the contexts subconservativity, superconservaitvity and conservativity, the vector z is called conservation vector.

Subconservativity and superconservativity are topological (structural) properties of discrete-state chemical reaction networks, they are determined by the reaction vectors that constitute the stoichiometric matrix. The subclass of conservative network structures (topologies) within d-CRNs represents a special case of sub- and superconservativity. In these networks, for any initial state $X_0 \in \mathbb{Z}^n_{\geq 0}$, all reachable states lie within an at most $(n-1)$-dimensional hyperplane.

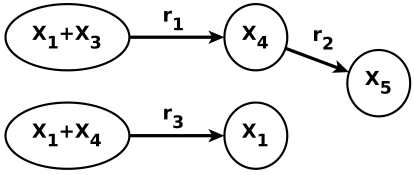
Example 1. Two d-CRNs with sub-and superconservative property are depicted in Figure 1. Sub-and superconservative networks can be transformed to each other by reversing the signs of the entries in their stoichiometric matrices. By changing the signs of the entries in a stoichiometric matrix one can revere the direction of the reactions in a reaction network graph $G_{\mathcal{N}}$.

In the sequel, we make use of the definition of totally unimodular matrices [19]:

Definition 9. [13, 14] A matrix A is totally unimodular if each subdeterminant of A is 0, $+1$, or -1.

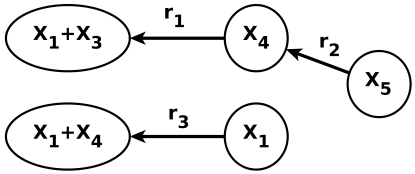
We note that in a totally unimodular matrix all the entries are 0 or ± 1.

Totally unimodular matrices has particular importance in mathematical optimization. Let us consider an Integer Linear Program (ILP) of the form:



$$\Gamma_{\mathcal{N}} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \\ 1 & -1 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

(a)



$$\Gamma_{\mathcal{N}'} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ -1 & 1 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

(b)

Figure 1. A pair of a sub-and superconservative d-CRNs. The reaction network graphs 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.

$$ILP \begin{cases} \min_x \{a^\top x\} \\ \text{subject to} \\ Ax \leq b \\ x \in \mathbb{Z}^n \end{cases} \tag{13}$$

It is known that if the coefficient matrix A in Eq. (13) is totally unimodular and b is an integer vector, then by relaxing the integer constraint on the decision variable vector x, the resulting linear program provides optimal solution for the ILP [19].

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.

If $A \in \mathbb{Z}^{k\times p}$ is a totally unimodular matrix and $I \in \mathbb{Z}^{p\times p}$ is an identity matrix, then it can be easily shown that the following block matrix is totally unimodular:

$$\left[\frac{A}{I}\right]. \tag{14}$$

3 Overview of d-CRN reachability

In this section we review the relevant d-CRN reachability results from the scientific literature. In general, the d-CRN reachability problem is NP-hard and may require additional conditions to solve. As shown in [13,14], for d-CRNs with stoichiometric matrices satisfying total unimodularity, the NP-hard reachability decision problem can be transformed into an LP feasibility problem. This transformation hinges on the structural properties of the reaction network, particularly when:

1. Each reaction modifies at most one species.
2. The reaction network is sub-or superconservative.
3. The stoichiometric matrix $\Gamma_{\mathcal{N}}$ contains entries only in the set $\{0, \pm 1\}$.

In the above case, the integer feasibility of the state equation implies actual reachability, enabling polynomial-time decision making.

The function $M = M(\Gamma^-)$ is extensively used in the next sections [12,13]:

$$[M(\Gamma^-)]_i = max\Big\{[\Gamma^-]_{ij} :\ j = 1,\dots\ l\Big\}, \qquad i = 1,\dots\ n. \tag{15}$$

Note that for any state $X \in \mathbb{Z}^n_{\geq 0}$, the relation $X \succeq M(\Gamma^-_{\mathcal{N}})$ indicates that each $r \in \mathcal{R}$ reaction is charged at X.

The following proposition is extensively used in the main result of the paper.

Proposition 2. *[14] 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* $r \in \mathcal{R}$, $\sum_{i=1}^{n} [\overline{y}^{+}]_i \leq 1$ *and* $\sum_{i=1}^{n} [\overline{y}^{-}]_i = 1$. *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.* (15). *Then the reachability relation* $X_0 \rightsquigarrow_{\mathcal{N}} X'$ *can be decided in polynomial time.*

Proposition 2 provides a constructive proof in which it is shown that the polynomial relaxation makes use of the discrete state equation of the d-CRN. Since the stoichiometric matrix of the specified sub-and superconservative d-CRN subclasses are shown to be totally unimodular, it follows that the the discrete state equation (Eq. (12)) based IP can be relaxed to be a Linear Program.

4 Main results

This section discusses the main result of the paper. Making use of the relaxed time complexity result for reachability analysis in Proposition 2, we prove that the reachable sequence computation problem has polynomial time complexity. A constructive proof is given in which a computational procedure is discussed for reachable state transition sequence computation. The constructed algorithm is of polynomial time complexity.

Proposition 3. *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* $r \in \mathcal{R}$, $\sum_{i=1}^{n} [\overline{y}^{+}]_i \leq 1$ *and* $\sum_{i=1}^{n} [\overline{y}^{-}]_i = 1$. *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.* (15). *Assuming* $X_0 \rightsquigarrow_{\mathcal{N}} X'$, *an admissible state transition sequence* $\sigma_X = X_0 X_1 \ \ldots \ X'$ *(and reaction sequence* σ_r*) can be determined in polynomial time.*

Proof. Constructive.

A constructive proof is provided in which an algorithm is composed for computing a feasible σ_X state transition sequence and σ_r reaction se-

quence. First a set of auxiliary procedures is introduced. Next the algorithm **ComputeReachableSequence** is introduced for computing a pair of feasible sequences for σ_X and σ_r so that the reachability relation $X_0 \rightsquigarrow_{\mathcal{N}} X'$ is satisfied. Finally, theoretical guarantees are provided for the correctness of **ComputeReachableSequence** .

In the algorithm **ComputeReachableSequence** , the following subprocedures are employed:

DecideReachability $(X_0, X', \mathcal{N})$: decides the reachability problem of $X_0 \rightsquigarrow_{\mathcal{N}} X'$. Provided the assumptions of the Proposition with respect to the d-CRN $\mathcal{N}$, the reachability relation can be decided in polynomial time based on the following Linear Program [13, 14]:

$$LP_{DecideReachability} \begin{cases} \min_c \{a^\top c\} \\ \text{subject to} \\ \Gamma_{\mathcal{N}} c = X' - X_0 \\ c \in \mathbb{R}^l_{\geq 0} \end{cases} \tag{16}$$

Note that $a \in \mathbb{Z}^l_{>0}$ is arbitrary with finite entries. In the Linear Program of Eq. (16) c denotes the vector of decision variables. The existence of a feasible solution of the LP implies that the reachability relation $X_0 \rightsquigarrow X'$ holds for some reaction vector sequence σ_r and $[c]_i$ encodes the number of occurrence of $r_i \in \mathcal{R}$ in σ_r for $i = 1, \ \ldots \ l$.

Since $\Gamma_{\mathcal{N}}$ is unimodular, Proposition 2 guarantees that feasibility of $LP_{DecideReachability}$ implies that the reachability relation holds and it can be decided in polynomial time.

The procedure returns a tuple $(d, \hat{\mathcal{R}})$ where d is a boolean variable. $d = TRUE$ if and only if the reachability relation $X_0 \rightsquigarrow_{\mathcal{N}} X'$ holds, otherwise $d = FALSE$.

For $d = TRUE$, $\hat{\mathcal{R}}$ is a multiset of the reactions so that there exists a σ_r ordering of the reactions in $\hat{\mathcal{R}}$ for which an admissible $\sigma_X = X_0 \ \ldots \ X'$ state transition sequence can be realized.

For $d = FALSE$, $\hat{\mathcal{R}} = \emptyset$.

DecideConstReachability $(X_0, X', \mathcal{N}, \hat{\mathcal{R}})$: it decides the reachability problem of $X_0 \rightsquigarrow_{\mathcal{N}} X'$, given the constraint that the multiset $\hat{\mathcal{R}}$ prescribes the firing reactions with their respective multiplicities. Note that $\hat{\mathcal{R}}$ does not prescribe the order of the reactions, but a multiset of the reactions so that for each $r \in \hat{\mathcal{R}}$, the multiplicity of r in $\hat{\mathcal{R}}$ equals to the number of times r fires. The procedure decides the reachability problem by means of the following Linear Program:

$$LP_{DecideConstReachability} \begin{cases} \min_c \{a^\top c\} \\ \text{subject to} \\ \Gamma_{\mathcal{N}} c = X' - X_0 \\ Ic = \hat{c} \\ c \in \mathbb{R}^l_{\geq 0} \end{cases} \tag{17}$$

where $I \in \mathbb{Z}^{l \times l}$ is an identity matrix, and $\hat{c} \in \mathbb{Z}^l$ encodes the multiplicities of the reactions in $\hat{\mathcal{R}}$, provided the fixed ordering of the reactions in $\mathcal{R}$:

$$\hat{c} = \left[count(r_1, \hat{\mathcal{R}}) \; count(r_2, \hat{\mathcal{R}}) \; \ldots \; count(r_l, \hat{\mathcal{R}}) \right]^\top \tag{18}$$

where $count(r_i, \hat{\mathcal{R}})$ denotes the multiplicity of r_i in the multiset $\hat{\mathcal{R}}$ for $i = 1, \ldots l$. $a \in \mathbb{Z}^l_{>0}$ is arbitrary with finite entries in the LP of Eq. (17).

The procedure returns a boolean variable d so that $d = TRUE$ if and only if the reachability relation $X_0 \rightsquigarrow_{\mathcal{N}} X'$ is satisfied, given a constrained set of reactions $\hat{\mathcal{R}}$, otherwise $d = FALSE$.

$d = TRUE$ if and only if there exists a σ_r ordering of the reactions in the multiset $\hat{\mathcal{R}}$ for which an admissible $\sigma_X = X_0 \; \ldots \; X'$ state transition sequence can be realized.

The pseudo-code for computing a feasible state transition sequence σ_X and reaction sequence σ_r can be given as follows:

Algorithm 1 ComputeReachableSequence $(X_0, X', \mathcal{N})$

```
(d, R̂) ← DecideReachability(X_0, X', N)
if d = TRUE then
    L ← Σ_{i=1}^{l} count(r_i, R̂)
    X_currentn ← X_0
    σ_X ← X_0
    σ_r ← ∅
    for i = 1 ... L do
        for r ∈ set(R̂) do
            R̂_candidate ← R̂/r
            X_candidate ← X_current + r
            d ← DecideConstReachability(X_candidate, X', N, R̂_candidate)
            if d = TRUE then
                R̂ ← R̂_candidate
                X_current ← X_candidate
                σ_X ← σ_X ⊕ X_current
                σ_r ← σ_r ⊕ r
                BREAK
            end if
        end for
    end for
    return (σ_X, σ_r)
else
    return ∅ //Reachability does not hold.
end if
```

Now we prove that the procedure **ComputeReachableSequence** is well defined, that is it finds an admissible state transition sequence $\sigma_x = X_0 X_1 \ \ldots \ X'$ in polynomial time in the number of species (dimensionality of the state space) of the d-CRN. There are 2 cases.

1. If the reachability relation does not hold, then the procedure returns $\emptyset$ after the execution of **DecideReachability**$(X_0, X', \mathcal{N})$. The unreachability can be decided in polynomial time.

2. If the reachability relation holds, then **DecideReachability**$(X_0, X', \mathcal{N})$ returns a multiset $\hat{\mathcal{R}}$ in polynomial time. $\hat{\mathcal{R}}$ contains the reactions firing along an admissible $\sigma_X = X_0 \ \ldots \ X'$ state transition sequence. As $\hat{\mathcal{R}}$ is a multiset, it contains

the reactions with their respective multiplicities (number of occurrences) along σ_X.

Assuming that $X_0 \rightsquigarrow_{\mathcal{N}} X'$ holds, the procedure has L number of iterations where L denotes the number of firings of all the reactions in $\hat{\mathcal{R}}$. In each iteration, a reaction r is extracted from $\hat{\mathcal{R}}$ and the sequences σ_X and σ_r are extended.

Now we show by induction that for each $1 \leq i \leq L$, there exists an $r \in \hat{\mathcal{R}}$ for which the procedure

$$DecideConstReachability(X_{candidate}, X', \mathcal{N}, \hat{\mathcal{R}}_{candidate}) \tag{19}$$

returns $d = TRUE$ boolean response, provided that $\hat{\mathcal{R}}_{candidate} = \hat{\mathcal{R}}/r$.

Induction

(a) $i = 1$

DecideConstReachability solves the Linear Program described by Eq (17). The constraints of the LP can be formulated as follows:

$$\left[\frac{\Gamma_{\mathcal{N}}}{I}\right] c = \left[\frac{X' - X_{current}}{\hat{c}}\right] \tag{20}$$

If $\Gamma_{\mathcal{N}}$ is totally unimodular, then it follows that the following block matrix is totally unimodular:

$$\left[\frac{\Gamma_{\mathcal{N}}}{I}\right] \tag{21}$$

The unimodularity of the block matrix in Eq. (21) implies that the optimal solution $c \in \mathbb{R}^{l}_{\geq 0}$ satisfying Eq. (20) in the LP (17) must be an integer vector.

For $X_{current} = X_0$ the reachability relation $X_0 \rightsquigarrow_{\Gamma_{\mathcal{N}}} X'$ holds. Then it follows that $\exists r \in \hat{\mathcal{R}}$ for which $X_{candidate} \rightsquigarrow_{\Gamma_{\mathcal{N}}} X'$ with $X_{candidate} = X_{current} + r$ and the reachability can be realized by

a reaction sequence σ_r so that σ_r is an ordered sequence of the reactions of the multiset $\hat{\mathcal{R}}_{candidate} = \hat{\mathcal{R}}/r$. Then the Linear Program of Eq. (17) in *DecideConstReachability* is solved in polynomial time with the input $(X_{current}, X', \mathcal{N}, \hat{\mathcal{R}}_{candidate})$.

In the LP the constraint set of Eq. (20) can be decomposed as follows:

$$\Gamma_{\mathcal{N}} c = X' - X_{candidate} \tag{22}$$

$$\Gamma_{\mathcal{N}} c = \hat{c} \tag{23}$$

The optimal solution $c \in \mathbb{R}^l_{\geq 0}$ of the LP is proven to be integer and it satisfies Eqs. (22) and (23). (22) implies that the reachability relation holds. (23) implies that c encodes the multiset $\hat{\mathcal{R}}$ and there exists a σ_r ordering of the reactions so that it drives the system from $X_{candidate}$ to X'.

For $i = 1$, the multiset $\hat{\mathcal{R}}$ is updated so that $\hat{\mathcal{R}} = \hat{\mathcal{R}}_{candidate}$, this way the following equation holds: $|\hat{\mathcal{R}}| = L - 1$.

(b) *inductive assumption* $i = k$

Let us assume that for $i = k$ there exists an $r \in \hat{\mathcal{R}}$ for which the procedure *DecideConstReachability* returns $d = TRUE$, given that $\hat{\mathcal{R}}_{candidate} = \hat{\mathcal{R}}/r$. Let us assume that by setting $\hat{\mathcal{R}} = \hat{\mathcal{R}}_{candidate}$, we have that $|\hat{\mathcal{R}}| = L - k$.

(c) For $i = k + 1$

Let us consider $\hat{\mathcal{R}}$ and $X_{current}$ at iteration $i = k+1$. It is given by the inductive assumption for $i = k$ that $X_{current} \rightsquigarrow_{\mathcal{N}} X'$ holds with an ordered sequence of the reactions in the multiset $\hat{\mathcal{R}}$. The reachability relation $X_{current} \rightsquigarrow_{\mathcal{N}} X'$ implies that there exists $r \in \hat{\mathcal{R}}$ for which $X_{candidate} \rightsquigarrow_{\mathcal{N}} X'$ holds with $X_{candidate} = X_{current} + r$ and the entries of $\hat{\mathcal{R}}_{candidate} = \hat{\mathcal{R}}/r$ can be ordered in a reaction sequence σ_r so that it defines an admissible state transition sequence $\sigma_x = X_{candidate} \ \ldots \ X'$. Then LP of Eq. (17) in *DecideConstReachability* with parame-

ters ($X_{candidate,X',\mathcal{N},\hat{\mathcal{R}}_{candidate}}$) is satisfied. The constraint matrix (21) in the LP is totally unimodular, this way the optimal solution $c \in \mathbb{R}^l_{\geq 0}$ is integer.

The procedure sets $\hat{\mathcal{R}} = \hat{\mathcal{R}}_{candidate}$, hence we have that $|\hat{\mathcal{R}}| = L - (i-1)$.

The inductive proof guarantees that the iterative procedure consumes all the reactions of the multiset $\hat{\mathcal{R}}$ so that in each iteration a suitable $r \in \hat{\mathcal{R}}$ is identified. The number of reactions in the multiset $\hat{\mathcal{R}}$ is monotonously decreasing along the iterations. Hence it is proven that all the L reactions of $\hat{\mathcal{R}}$ are consumed in L number of iterative steps.

It is proven that for each i, $X_{current} \rightsquigarrow_{\mathcal{N}} X'$. Then we have that for $i = L$, $X_{current} \rightsquigarrow_{\mathcal{N}} X'$ with $\hat{\mathcal{R}} = \{r\}$ before the update on $\hat{\mathcal{R}}$. Then we have that $X_{current} + r = X'$. ■

Remark. We note the following points with respect to Proposition 3 and its proof:

1. The proof is constructive, an algorithm is provided for determining an admissble state transition sequence σ_X and reaction sequence σ_r.

2. The constructed proof shows that the proposed procedure is of polynomial time complexity. Under the conditions of Proposition 3, the procedures **DecideReachability** and **DecideConstReachability** have polynomial time complexity with respect to the state space dimensionality n which is equialent to the number of species. This way, given a pair of initial and target states, the time complexity of the reachable sequence computation algorithm is calculated as follows: $\Big(1 + L \cdot |\mathcal{R}|\Big) \cdot \mathcal{P}(n)$, where L is the total number of reactions with multiplicities that are executed, $|\mathcal{R}|$ denotes the number of reactions in the d-CRN and $\mathcal{P}(n)$ is a polynomial function with $n = |\mathcal{S}|$.

Proposition 4. *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* $r \in \mathcal{R}$,

$\sum_{i=1}^{n} [\overline{y}^{+}]_i \leq 1$ *and* $\sum_{i=1}^{n} [\overline{y}^{-}]_i = 1$. *Let us consider two arbitrary states,* $X_0,\ X' \in \mathbb{Z}^n_{\geq 0}$ *so that* $X_0 \succeq M$, $X' \succeq M$ *where* $M = M(\Gamma^{-})$ *is defined by Eq.* (15). *Assuming* $X_0 \rightsquigarrow_{\mathcal{N}} X'$, *a shortest (minimal-length) state transition sequence* $\sigma_X = X_0 X_1 \ \ldots \ X'$ *(and reaction sequence* σ_r*) can be determined in polynomial time.*

Proof. Algorithm 1 provides a reaction vector sequence and state transition sequence from X_0 to X' in polynomial time. Let us assume that the procedures **DecideReachability** and **DecideConstReachability** are configured so that the vectors $a \in \mathbb{Z}^l_{\geq 0}$ of Eqs. (16) and (17) satisfy the following equation

$$a = 1^l. \tag{24}$$

Then the Linear Program defined by Eq.(16) returns a decision vector $c \in \mathbb{Z}^l_{\geq}$ for which the sum of the entries is minimized. That is c encodes the reactions incorporated in a minimal-length reaction vector sequence. Then Algorithm 1 returns an admissible reaction vector sequence from the multiset of reactions provided by c. This way a minimal-length reaction vector sequence is determined in polynomial time. ■

5 Example and performance evaluation

The computational experiments were performed on a workstation containing 12 Intel i7-8750H vCPUs (2.20GHz, 7.6 GiB). In the implementation we used Python 3.10 and the Gurobi optimization package [21].

To illustrate the algorithm, we apply it to a monomolecular conservative reaction network described in [20]. In Fig. 2 the respective d-CRN is depicted. The network has a totally unimodular stoichiometric matrix, ensuring that our assumptions for polynomial-time computability are satisfied. We note that the same example was used in our previous papers analyzing the reachability decision problem of d-CRNs [13, 14]. The associated stochiometric matrix is given by Eq. (25).

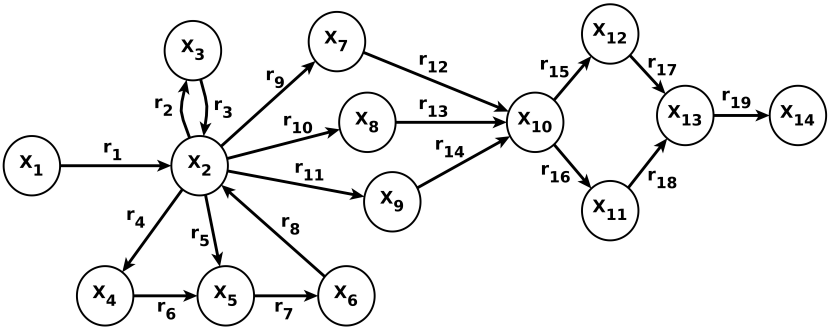


Figure 2. A conservative d-CRN of a monomolecular reaction network. The d-CRN has totally unimodular stoichometric matrix. It satisfies the conditions of Proposition 3, this way a reachable state transition sequence can be computed in polynomial time.

$$\Gamma_{\mathcal{N}} = \begin{bmatrix}
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -1 & 1 & -1 & -1 & 0 & 0 & 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} \tag{25}$$

In order to demonstrate the power of Proposition 3, the following computational experiment was executed. We iterated from $i = 10$ to $i = 500$. For each i, 20 state transition sequences σ_X with initial state X_0 were generated so that $|\sigma_X| = i$. Note that the sequences were generated randomly, it was not necessary to apply the Gillespie or other related stochastic simulation algorithm, since we needed only feasible random reaction sequences. The following value of X_0 was assumed in the experiment:

$$X_0 = \Big[100\ 100\ 100\ 100\ 100\ 100\ 100\ 100\ 100\ 100\ 100\ 100\ 100\ 100\ 100\Big]^{\top}. \tag{26}$$

For each σ_X generated, we took the last state X_{target} of the sequence. Making use of the reachable sequence computation algorithm provided in the proof of Proposition 3, we determined a state transition sequence $\overline{\sigma}_X =$

$X_0 \ \ldots \ X_{\text{target}}$. For each execution of the reachable sequence computation algorithm, we measured its execution time. The average execution time is depicted in Figure 3.

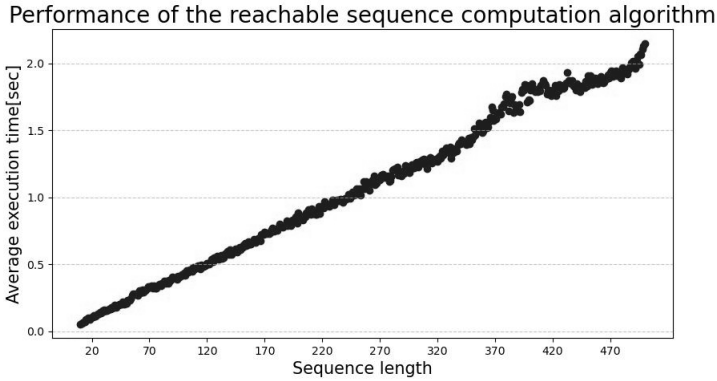


Figure 3. Average execution time [sec] of the reachable sequence computation algorithm provided in the constructive proof of Proposition 3.

6 Summary

In this paper the reachability analysis of sub-and superconservative d-CRN subclasses were studied. Leveraging on existing results of CRNs with totally unimodular stoichiometric matrix, we proved that – under structural (reaction network topology related) conditions – a reachable state transition sequence (and reaction vector sequence) can be determined in polynomial time. The results are provided with a constructive proof incorporating a polynomial time algorithm. The computational power of the proposed method was demonstrated on a computational example. Note that in the case of general d-CRN reaction network structures, the reachability decision problem is NP-hard, which might make the reachability calculation and reaction sequence calculation practically untreatable for high state space dimensionality [14].

The practical importance of Proposition 3 is that for well-defined sub- and superconservative d-CRN subclasses, the problem of reachable se-

quence computation can be solved in polynomial time. This can support the analysis of various practical problems, such as extinction events in (bio)chemical reaction networks, logical gate design in synthetic biology and verification analysis in discrete event systems.

***Acknowledgment*:** Gergely Szlobodnyik acknowledges the support of the grant PD 146397 from the National Research, Development, and Innovation Office (NKFIH). Gábor Szederkényi acknowledges the support of the grant K145934 from the National Research, Development, and Innovation Office (NKFIH).

References

[1] M. Feinberg, *Foundations of Chemical Reaction Network Theory*, Springer, 2019.

[2] D. Angeli, A tutorial on chemical reaction network dynamics, *Eur. J. Control* **15** (2009) 398–406.

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

[4] S. Smith, R. Grima, Single-cell variability in multicellular organisms, *Nature Commun.* **9** (2018) #345.

[5] D. F. Anderson, T. G. Kurtz, *Stochastic Analysis of Biochemical Systems*, Springer, Berlin, 2015.

[6] D. F. Anderson, T. G. Kurtz, Continuous time Markov chain models for chemical reaction networks, in: H. Koeppl, G. Setti, M. di Bernardo, D. Densmore (Eds.), *Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology*, Springer, Berlin, (2011) 3–42.

[7] 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.

[8] B. Ács, G. Szlobodnyik, G. Szederkényi, A computational approach to the structural analysis of uncertain kinetic systems, *Comput. Phys. Commun.* **228** (2018) 83–95

[9] 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** (2017) 1535–1558.

[10] M. D. Johnston, A computational approach to extinction events in chemical reaction networks with discrete state spaces, *Math. Biosci.* **294** (2017) 130–142.

[11] D. F. Anderson, A. S. Howells, Stochastic reaction networks within interacting compartments, *Bull. Math. Biol.* **85** (2023) #87.

[12] G. Szlobodnyik, G. Szederkényi, M. Johnston, Reachability analysis of subconservative discrete chemical reaction networks, *MATCH Commun. Math. Comput. Chem.* **81** (2019) 705–736.

[13] G. Szlobodnyik, G. Szederkényi, Reachability analysis of low-order discrete state reaction networks obeying conservation laws, *Complexity* **2019** (2019) #1035974.

[14] G. Szlobodnyik, G. Szederkényi, Polynomial time reachability analysis in discrete state chemical reaction networks obeying conservation laws, *MATCH Commun. Math. Comput. Chem.* **89** (2023) 175–196.

[15] T. Murata, Petri nets: properties, analysis and applications, *Proceed. IEEE* **77** (1989) 541–580.

[16] A. Condon, D. Harel, J. N. Kok, A. Salomaa, E. Winfree (Eds.), *Algorithmic Bioprocesses*, Springer, 2009.

[17] E. Lakatos, M. P. H. Stumpf, Control mechanisms for stochastic biochemical systems via computation of reachable sets, R. Soc. Open Sci. 4(8):160790 (2017)

[18] T. Veloz, D. Flores, Reaction network modeling of complex ecological interactions: Endosymbiosis and multilevel regulation, *Complexity* **2021** (2021) #8760937.

[19] A. Schrijver, *Theory of Linear and Integer Programming*, Wiley, 1999.

[20] N. Vassena, H. Matano, Monomolecular reaction networks: Flux-influenced sets and balloons, *Math. Meth. Appl. Sci.* **40** (2017) 7722–7736.

[21] Gurobi Optimization Inc., Gurobi Optimizer Reference Manual, `www.gurobi.com`, 2016.