MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **81** (2019) 705-736

ISSN 0340 - 6253

Reachability Analysis of Subconservative Discrete Chemical Reaction Networks

Gergely Szlobodnyik[1], Gábor Szederkényi[1,*], Matthew D. Johnston[2]

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

[2] *Department of Mathematics, San Jose State University, One Washington Square, San Jose, CA 95192, USA*

(Received May 15, 2018)

Abstract

In this paper, we give a computational solution for the reachability problem of subconservative discrete chemical reaction networks (d-CRNs), namely whether there exists a valid state transition (reaction) sequence between a given initial and a target state. Using subconservativity, we characterize the reachable set of the d-CRN with well-defined simplexes. Moreover, upper bounds are derived for the possible length of cycle-free state transition sequences. We show that the reachability and the related coverability problem in the case of subconservative d-CRNs can be decided in polynomial time by tracing them back to fixed dimensional integer programming (IP) feasibility problems over a bounded integer lattice. The proposed computation model is also employed for determining feasible series of reactions between given (sets of) states. We also show that if the rank of the stoichiometric matrix is less than or equal to 2, then the reachability problem is equivalent to the existence of a non-negative integer solution of the corresponding state equation.

1 Introduction

Chemical Reaction Network theory (CRNT) studies both the dynamical behavior and structural properties of chemical reaction networks (CRNs). In the case of large molecule numbers and appropriate physico–chemical conditions, it is commonly assumed that CRNs obey the law of mass action. Consequently, the systems can be characterized by a subclass of non-negative polynomial systems known as kinetic systems [1–5]. For this system class, the dynamical behavior including special dynamical patterns, such as oscillations,

*corresponding author

multiplicity, bifurcation, structural properties and the relationship between them have been extensively studied [6–9]. Effective polynomial time algorithms have also been designed to identify network structures that are capable of exhibiting these dynamical behaviors [10–15].

However, if the molecular counts of a CRN are low (typically smaller than approximately 40-60 molecules per species according to the literature), then the deterministic differential equation description of kinetic systems often does not give us satisfactory results [16, 18]. This holds for example in the case of genetic circuits [20]. Therefore, it becomes important to choose another mathematical model tracking the molecular counts for each species. One can use discrete CRN (d-CRN) models for which the state variables are the integer molecular counts and the state space of the system is a subset of the integer lattice of the non-negative orthant. In this approach both the reaction vectors and state variables are modeled as vectors with integer entries (which will be simply referred to as integer vectors in the paper). Given the instantaneous molecular counts, a d-CRN can evolve along any of its reactions if the minimal amount of molecular counts for each required species are given for firing. If a particular reaction occurs, then the states must be updated according to the net gain and loss of the reaction. The only assumption of d-CRN models about the dynamical behavior of the system is that at a certain time instant at most one reaction can occur. Although there are no additional assumptions pertaining to the dynamical behavior of the underlying reaction network , d-CRNs combined with continuous time Markov chain models are commonly used to simulate the trajectories [18, 19]. In such a way we arrive to continuous time, discrete state stochastic models.

The dynamical properties of d-CRNs have been extensively studied in the literature [21, 33, 35, 36]. Kurtz has shown that the infinite volume limit of the Markov chain-based model is the deterministic solution on finite time intervals, but it is not necessarily the case for infinite ones [16, 17, 21]. It is known, that the long-term qualitative dynamical behavior predicted by the deterministic continuous state and stochastic discrete state models may be significantly different [21,22]. Hence, it is important to examine the qualitative dynamical behavior of the stochastic systems modeled by discrete state continuous time Markov chains. An important part of this qualitative analysis is the so-called reachability and coverability analysis of states which are far from the solution of the deterministic continuous state model. A question related to reachability is the existence of extinction events [21]. A sufficient condition of extinction events has recently been identified for subconservative CRNs [22], and a MILP-based computational approach has

also been published in an accompanying paper [23].

In this paper we study the reachability problem of (sub)conservative d-CRNs by means of their structural properties. An important consequence of subconservativity is the boundedness of the state space, which is extensively used. We reformulate the problem of reachability as an IP feasibility problem using the discrete state equation and additional constraints to form a bounded convex set. It is known from the literature that checking whether a bounded convex set contains integer points can be performed in polynomial time in fixed dimension [24, 26]. Moreover, all the integer points can also be enumerated in polynomial time, assuming fixed number of dimensions [27, 28]. Hence, the proposed IP feasibility-based computational approach decides a given reachability problem in polynomial time, since the dimension of the problem is determined by the initial and target states and the structure of the examined CRN, which are fixed. This computational approach can be easily extended to coverability analysis, where the task is to find at least one feasible state which is coordinate-wise larger than a prescribed one. The main result of this paper related to the reachability analysis is Proposition 4 in Section 3. Beyond the feasibility analysis, the system of equations and inequalities can be extended by a linear cost function and additional constraints so that an appropriate sequence of state transitions is determined if it exists. For example, a state transition sequence can be determined between a prescribed initial and a target state for which the length of the path or the agglomeration of toxic secondary products is minimal.

2 Notations and mathematical background

In this section we summarize the notations, definitions and mathematical background of discrete state Chemical Reaction Networks. The general notations used in this paper are the following:

$\mathbb{R}$	the set of real numbers
$\mathbb{Z}$	the set of integer numbers
$\mathbb{Z}_{\geq 0}$	the set of non-negative integer numbers
$\mathbb{T}^n$	the set of n-dimensional vectors over the set $\mathbb{T}$
$\{0,1\}^l$	the set of l-dimensional binary vectors (all the entries are equal to 0 or 1)
$[A]_{ij}$	the entry of matrix A with row index i and column index j
$\mathcal{Q}(\mathcal{O})$	the number of points with non-negative integer coordinates in a bounded set $\mathcal{O} \subset \mathbb{R}^n$
$a \prec b$	for $a, b \in \mathbb{R}^n$, $a_i < b_i$ for $i = 1, \ldots, n$
$a \preceq b$	for $a, b \in \mathbb{R}^n$, $a_i \leq b_i$ for $i = 1, \ldots, n$
$0^{n \times m}$	a zero matrix of dimension $n \times m$

Table 1. Notations

2.1 Chemical Reaction Networks with discrete state space

A discrete state Chemical Reaction Network (d-CRN) can be described by a triple $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ such that:

$$\begin{aligned}
\mathcal{S} &= \{s_i \mid i \in \{1, \ldots, n\}\} \\
\mathcal{C} &= \{y_j = \sum_{i=1}^{n} \alpha_{ji} s_i \mid \alpha_{ji} \in \mathbb{Z}_{\geq 0},\ j \in \{1, \ldots, m\},\ i \in \{1, \ldots, n\}\} \\
\mathcal{R} &= \{(y_i, y_j) \subset \mathcal{C} \times \mathcal{C} \mid i \neq j\}
\end{aligned}$$

where s_i is the i'th species, y_j is the j'th complex and the ordered pair (y_i, y_j) represents the reaction $y_i \to y_j$ of the system. If a reaction $(y_i, y_j) \in \mathcal{R}$ exists, then y_i and y_j are called source and product complexes, respectively.

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

$$\overline{y}_j = [\alpha_{j1}\ \alpha_{j2}\ \ldots\ \alpha_{jn}]^\top$$

For the reaction $(y_i, y_j) \in \mathcal{R}$ the reaction vector $r_{ij} = \overline{y}_j - \overline{y}_i \in \mathbb{Z}^n$ tracks the net gain and loss of the species. In the sequel we will assume that for the reaction vectors a given order is defined: $r_1,\ r_2\ \ldots\ , r_l$, where $|\mathcal{R}| = l$. The associated stoichiometric matrix Γ of the system is composed of the reaction vectors:

$$\Gamma = [r_1\ r_2\ \ldots\ r_l]$$

For the sake of simplicity, we will use the notation r_i for both the reactions of the system and the associated reaction vectors. Furthermore, the notation y_i will represent the complexes of the system and the vectors containing the stoichiometric coefficients, as well. The molecular count of each species of the system at time t is given by a state vector $X(t) \in \mathbb{Z}^n_{\geq 0}$. The state evolution of the system is given by the following discrete state equation:

$$X(t) = X(0) + \Gamma N(t) \tag{1}$$

where $X(0)$ is the initial state vector and $N(t) = [N_1(t)\ N_2(t)\ \ldots\ N_l(t)]^\top \in \mathbb{Z}^l_{\geq 0}$ such that $N_k(t) \in \mathbb{Z}_{\geq 0}$ keeps track of the number of times the k'th reaction has occurred until time t.

Note that the above model has no assumptions about the probability of the state transitions.

Example 2.1. *Let us consider the CRN depicted in Figure 1. This CRN is characterized by the following triple* $(\mathcal{S}, \mathcal{C}, \mathcal{R})$*:*

$$\mathcal{S} = \{A,\ B,\ E\}$$
$$\mathcal{C} = \{A + E,\ B + E,\ E\}$$
$$\mathcal{R} = \{A + E \to B + E,\ B + E \to A + E,\ A + E \to E,\ B + E \to E\}$$

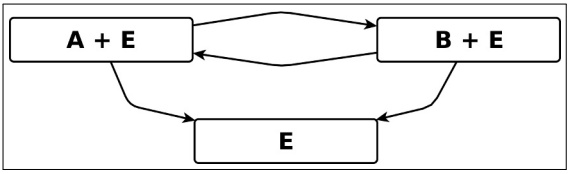

Figure 1. Introductory example of CRN's.

The associated stoichiometric matrix composed of the reaction vectors of the system is the following:

$$\Gamma = \begin{bmatrix} -1 & 1 & -1 & 0 \\ 1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

In the sequel we do not write out the triple $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ if it is clear from the figure. In these cases the fixed order of species, complexes and reactions will be considered as given by the stoichiometric matrix Γ.

The definitions of the remaining part of this section are adopted from [22].

Definition 1. Let us consider a d-CRN. We say that:

1. A complex $y \in \mathcal{C}$ is **charged** at state $X \in \mathbb{Z}^n_{\geq 0}$ if $X_i \geq y_i$ for all $i \in \{1, 2 \ldots n\}$.

2. A reaction $r \in \mathcal{R}$ is charged if its respective source complex is charged.

3. A state $X \in \mathbb{Z}^n_{\geq 0}$ **reacts** to a state $X' \in \mathbb{Z}^n_{\geq 0}$ (denoted by $X \to X'$) if there exists a reaction $r \in \mathcal{R}$ such that r is charged at state X and $X + r = X'$.

4. A **path** σ is a finite sequence of non-negative ordered states $X_1,\ X_2,\ \ldots\ , X_p$ for which $X_1 \to X_2 \to \ldots \to X_{p-1} \to X_p$.

5. A state $X' \in \mathbb{Z}^n_{\geq 0}$ is **reachable** from a state $X \in \mathbb{Z}^n_{\geq 0}$ (denoted by $X \rightsquigarrow X'$) if there exists a path in the state space so that $X = X_{\nu(1)} \to X_{\nu(2)} \to \ldots \to X_{\nu(l)} = X'$.

6. A set of states $\mathcal{T} \subset \mathbb{Z}^n_{\geq 0}$ is said to be **coverable** from a state $X \in \mathbb{Z}^n_{\geq 0}$ if there exists a state $X' \in \mathbb{Z}^n_{\geq 0}$ for which $X \rightsquigarrow X'$ and $X' \geq X''$ for all $X'' \in \mathcal{T}$.

Since the set of reachable states depend on the initial state, we introduce the definition of the reachable state space as follows:

Definition 2. Let us consider an arbitrary d-CRN with a given initial state X_0. The **reachable state space** $\mathcal{Z}_{X_0}$ of the system corresponding to X_0 is the set of non-negative discrete states reachable from X_0.

From now on, we will call $\mathcal{Z}_{X_0}$ the state space of the system. Note that in order to fire a reaction $r \in \mathcal{R}$ at state $X \in \mathbb{Z}^n_{\geq 0}$, the source complex of r is required to be charged. The definition of a charged complex is related to the concept of coverability: given the initial state $X_0 \in \mathbb{Z}^n_{\geq 0}$ and a complex $y \in \mathcal{C}$, if there exists a state $X' \in \mathbb{Z}^n_{\geq 0}$ where the complex y is charged and for which $X_0 \rightsquigarrow X'$, then the set of states $\{X \mid X \in \mathbb{Z}^n_{\geq 0},\ X \leq y\}$ is coverable from X_0.

In the context of complexes and reactions the so-called recurrency and transiency are also defined [22]:

Definition 3. Let us consider a d-CRN. We say that:

1. A complex $y \in \mathcal{C}$ is **strongly recurrent** from a state $X \in \mathbb{Z}^n_{\geq 0}$, if $X \rightsquigarrow Y$ implies that there exists a state $Z \in \mathbb{Z}^n_{\geq 0}$ such that $Y \rightsquigarrow Z$ and y is charged at Z, otherwise y is called **weakly transient** from X.
2. A complex $y \in \mathcal{C}$ is **weakly recurrent** from $X \in \mathbb{Z}^n_{\geq 0}$ if there exists a state $Y \in \mathbb{Z}^n_{\geq 0}$ such that $X \rightsquigarrow Y$ and y is strongly recurrent from Y, otherwise y is **strongly transient** from X.

We also introduce the so-called extinction events, which are related to transiency and the reaction network structure [22].

Definition 4. Let us consider a discrete state CRN. We say that the CRN exhibits an **extinction event** on $\mathcal{C}' \subseteq \mathcal{C}$ from $X_0 \in \mathbb{Z}^m_{\geq 0}$, if every complex $y \in \mathcal{C}'$ is strongly transient from X_0.

In the sequel we will extensively use the following definitions in the analysis of d-CRNs:

Definition 5. Let us consider an arbitrary CRN $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ where $\mathcal{R} = \{r_1, \ \dots \ , r_l\}$. The **stoichiometric subspace S** of this system is the space spanned by the reaction vectors:

$$\mathbf{S} = span(r_1, \ \dots \ , r_l)$$

Definition 6. Let us consider an arbitrary CRN $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ and a non-negative initial state $X_0 \in \mathbb{Z}^n_{\geq 0}$. The **non-negative stoichiometric compatibility class** associated to X_0 is:

$$\mathbf{C}_{X_0} = (X_0 + \mathbf{S}) \cap \mathbb{R}^n_{\geq 0}$$

where $\mathbf{S}$ is the stoichiometric subspace of the system.

Note that the above definitions of the stoichiometric subspace and the non-negative stoichiometric compatibility classes are not restricted to the case of discrete state CRNs.

Clearly, we can write for the state space of a d-CRN that

$$\mathcal{Z}_{X_0} \subseteq \mathbf{C}_{X_0} \cap \mathbb{Z}^n_{\geq 0}$$

2.2 Mathematical optimization background

In this section we review some important aspects of mathematical programming which are extensively used in our study. An Integer Program (IP) can be defined as follows:

$$IP \begin{cases} \min\limits_x \{c^\top x\} \\ \quad s.t. \\ Ax \leq b \\ x \geq 0^{n \times 1} \\ x \in \mathbb{Z}^n \end{cases} \tag{2}$$

If the solution itself is not important only its existence, then we get a so-called feasibility problem (FP): given the polytope $P \subset \mathbb{R}^n$ defined by the inequalities $Ax \leq b$, $x \geq 0^{n \times 1}$ and the lattice $\Lambda \subseteq \mathbb{Z}^n$ of integer vectors. We have to decide whether $P \cap \Lambda = \emptyset$ holds. If the intersection is not empty then we say that the problem is feasible. Formally the feasibility problem can be expressed as follows:

$$FP \begin{cases} P \subset \mathbb{R}^n \\ \Lambda \subseteq \mathbb{Z}^n \\ P \cap \Lambda \stackrel{?}{=} \emptyset \end{cases} \tag{3}$$

While an IP is generally NP-hard [30], the associated FP in fixed dimension n can be computed in polynomial time using Lenstra's algorithm [24]. It is also possible to algorithmically determine the number of feasible points by the so-called Barvinok's algorithm [27] for which there exists an implementation in the software package LatteE [28]. These algorithms are based on the LenstraLenstraLovász basis reduction method [25].

3 Computational solution for reachability analysis

3.1 Problem statement

We are interested in the following problems related to reachability and coverability:

1. Given: an initial state $X_0 \in \mathbb{Z}^n_{\geq 0}$ and a target state $X^{'} \in \mathbb{Z}^n_{\geq 0}$. Can we find a path ensuring $X_0 \rightsquigarrow X^{'}$?
2. Given: an initial state $X_0 \in \mathbb{Z}^n_{\geq 0}$ and a set of target states $\mathcal{T} \subset \mathbb{Z}^n_{\geq 0}$. Can we find a state $X' \in \mathbb{Z}^n_{\geq 0}$ for which $X_0 \rightsquigarrow X^{'}$ and $X^{'} \succeq T$ hold for all $T \in \mathcal{T}$?

The above problems can be important both in the analysis and synthesis of chemical reaction networks. It can be crucial to analyze the existence of states of interest which are reachable from the initial state of a CRN. For example an undesired state can express the agglomeration of toxic species and/or the fact that a reaction producing toxic species is able to fire (i.e. the source complex is charged). Note that we consider paths between X_0 and $X^{'}$ without directed cycles, since the existence of those does not affect reachability.

We note that the time instant when the target state is reached is not important in this study, but only the reachability itself, hence in the notation of the target state $X^{'}$ the time variable does not appear.

3.2 Constraint formulation

Consider a discrete state CRN with stoichiometric matrix Γ. Let us denote the initial state by X_0. If a given state $X^{'}$ is reachable from the initial state X_0, then there exists a – not necessarily unique – vector $c \in \mathbb{Z}^m_{\geq 0}$ satisfying the Diophantine equation below:

$$X_0 + \Gamma c = X^{'} \tag{4}$$

This can also be used to check whether from the initial state we can reach a target state where a complex $y \in \mathcal{C}$ is charged:

$$X_0 + \Gamma c \geq y \tag{5}$$

If the state space of a CRN is bounded, the following inequalities also hold:

$$X_0 + \Gamma c \leq X_{max} \tag{6}$$

where $[X_{max}]_i \geq [X]_i$ for all $i \in \{1, \ldots, n\}$ and $X \in \mathcal{Z}_{X_0}$.

The inequalities described above do not guarantee that there exists a path $X_0 = X_{\nu(1)} \rightarrow X_{\nu(2)} \rightarrow \ldots \rightarrow X_{\nu(l)} = X^{'}$ represented by c which is valid in a (bio)chemical

sense. It is possible that there exists a state transition along the sequence of reactions where the source complex of the firing reaction is not charged at the actual state X_k, therefore, for the succeeding state $[X_{k+1}]_i < 0$ for some $i \in \{1, \dots, n\}$. This problem can be solved by using further constraints.

First we introduce the notation c_{max} for the upper bound of c (i.e. $c \leq c_{max}$). By summing up the entries of c_{max}, the overall number of reactions is given:

$$K = \sum_{i=1}^{l} [c_{max}]_i$$

While the entries of c_{max} correspond to upper bounds for the maximal number of occurrences of each reaction firing along a directed cycle-free path from X_0 to X', K is equal to the associated upper bound for the number of reactions along a directed cycle-free path. Using the above notations we introduce the following decomposition:

$$c = \sum_{j=1}^{K} v_j \tag{7}$$

$$v_j \in \{0,1\}^l \qquad j = 1, \dots, K \tag{8}$$

$$\sum_{i=1}^{l} [v_j]_i \leq 1 \qquad j = 1, \dots, K \tag{9}$$

where the binary vector v_j, $j \in \{1, \dots, K\}$ represents the reaction occurring in the network in the j-th time step, and $[v_j]_i$ denotes the i-th coordinate of v_j. Therefore, $[v_j]_i = 1$ means that the reaction r_i is firing in the j-th time step. It can be seen from Eqs. (7)-(9) that 'empty' reactions with $v_j = 0^{l \times 1}$ are technically allowed in the computations. Note that to apply the above decomposition in practice, we need to find an appropriate upper bound $c_{max} \in \mathbb{Z}^l$ for c such that $[c_{max}]_i < \infty$ for $i \in \{1, \dots, l\}$. Based on the decomposition in Eq. (7), the k'th state can be expressed as follows:

$$X_k = X_0 + \Gamma \sum_{i=1}^{k} v_i$$

The requirement for the state variable X_k is that the source complex of the forthcoming reaction – represented by v_{k+1} – is charged, which can be expressed by the following inequality:

$$X_0 + \Gamma \sum_{i=0}^{k} v_i \geq \Gamma_S v_{k+1} \qquad k = 0, \dots, K-1 \tag{10}$$

where $v_0 = 0^{l \times 1}$ and the columns of $\Gamma_S \in \mathbb{Z}_{\geq 0}^{n \times l}$ contain the stoichiometric coefficients of the source complexes of each reaction, ordered in the same way as the columns of Γ. This

means that $[\Gamma_S]_{ij}$ is the stoichiometric coefficient of the i-th species in the source complex of the j-th reaction for $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, l\}$.

The reachability and coverability problems of d-CRNs can be expressed by Eqs. (4), (6)-(10), and Eqs. (5)-(10), respectively. It can be seen that Eqs. (4)-(10) contain linear equalities and inequalities with integer unknowns in terms of v_j, $j = 1, \dots, K$. Therefore, the computational tasks of reachability and coverability can be written as integer programming (IP) feasibility problems of fixed dimensions and can be solved in polynomial time [24, 27].

3.3 Bounds for the length of reaction sequences

The number of possible reaction steps is an important factor that affects the practical computability of the reachability problem. Therefore, this subsection is devoted to computing upper bounds for the length of reaction sequences in the subconservative and conservative cases.

3.3.1 Subconservative case

Definition 7. Consider a CRN having stoichiometric matrix Γ. The system is called conservative (subconservative) if there exists a vector $z \in \mathbb{R}^n_{>0}$ for which $z^\top \Gamma = 0^{1\times m}$ ($z^\top \Gamma \leq 0^{1\times m}$). z is called the conservation vector.

One can see, that in *Example 2.1.1* the CRN is conservative with conservation vector $z = [1\ 1\ 1]^\top$. Note that if a CRN is (sub)conservative then the conservation vector is not unique, since it can be scaled by an arbitrary positive constant.

We note that the conservativity of d-CRNs is related to P-invariance (also called S-invariance) of Petri nets and this structural (network structure-related) property was previously considered in the context of CRN theory [22, 35, 36]. In this section, it is assumed that all the examined CRNs are subconservative and there exists at least one reaction producing at least one molecule of at least one species. It is known that a (sub)conservative CRN has a finite state space [1, 34]. Based on the proof of finiteness of subconservative CRNs' state space, it is possible to compute an upper bound for the coordinates of the reachable states [1].

Lemma 1. Let us consider a subconservative d-CRN with initial state $X_0 \in \mathbb{Z}^n_{\geq 0}$ and conservation vector $z \in \mathbb{R}^n_{>0}$. Then for all $X' \in \mathcal{Z}_{X_0}$ states the following general upper bound holds:

$$[X']_j \leq \frac{z^\top X_0}{\zeta} \qquad j \in \{1, \dots, n\}$$

where $\zeta = \min_{j\in\{1,\ldots,n\}} \{z_j\}$.

Proof.

According to the subconservativity:

$$\exists z \in \mathbb{R}^n_{>0}: \quad z^\top \Gamma \leq 0^{1\times m}$$

Let us take an arbitrary $X^{'} \in \mathcal{Z}_{X_0}$. Since $X^{'}$ is reachable from X_0, there exists a non-negative finite linear combination of the reaction vectors $r_1, \ \ldots \ r_m$ for which:

$$X^{'} = X_0 + a_1 r_1 + \ \ldots \ + a_m r_m$$

Let us take the following dot product:

$$z^\top (X^{'} - X_0) = z^\top (a_1 r_1 + \ \ldots \ + a_m r_m) =$$
$$a_1 z^\top r_1 + \ \ldots \ + a_m z^\top r_m \leq 0$$

Note that the dot product of the conservation vector z and an arbitrary reaction vector will be non-positive. From the above inequality:

$$z^\top X^{'} \leq z^\top X_0 \quad \Longrightarrow \quad 0 \leq \sum_{i=1}^{n} z_j X_j^{'} \leq z^\top X_0 = M$$

Let us define

$$\zeta = \min_{j\in\{1, \ \ldots \ n\}} z_j > 0$$

Then

$$0 \leq \zeta \sum_{i=1}^{n} X_i^{'} \leq \sum_{i=1}^{n} z_i X_i^{'} \leq z^\top X_0 = M$$

From the above inequality we can derive the following upper bound for $X_j^{'}$:

$$0 \leq X_j^{'} \leq \frac{M}{\zeta}$$

■

The above general bound can be tightened to an element-wise upper bound along each dimension (see, [34]). For convenience, we give our own proof for this bound in the next proposition.

Proposition 1. Let us consider a subconservative d-CRN having initial state $X_0 \in \mathbb{Z}^n_{\geq 0}$ and conservation vector $z \in \mathbb{R}^n_{>0}$. Then for all $X^{'} \in \mathcal{Z}_{X_0}$, the following element-wise upper bound holds:

$$[X^{'}]_j \leq \frac{z^\top X_0}{z_j} \qquad j \in \ \{1, \ldots, n\} \tag{11}$$

Proof. (Indirect)
Let us assume that there exists a state $X^{'}$ *which is reachable from* X_0 *and for some* $j \in \{1 \ \dots \ , n\}$*:*

$$[X^{'}]_j > \frac{z^\top X_0}{z_j}$$

Since $X^{'}$ *is reachable from* X_0 *according to Lemma 1 we have that*

$$\frac{z^\top X^{'}}{\zeta} \leq \frac{z^\top X_0}{\zeta}$$

otherwise the maximal coordinate value of the states reachable from $X^{'}$ *would be higher than* $\frac{z^\top X_0}{\zeta}$*, but this would also mean that the maximal coordinate value of the states reachable from* X_0 *is strictly higher than* $x_{max} = \frac{z^\top X_0}{\zeta}$*. From the above inequality and the indirect assumption we get:*

$$z^\top X_0 \geq \sum_{i=1}^{n} z_i[X^{'}]_i = \sum_{\substack{i=1 \\ i \neq j}}^{n} z_i[X^{'}]_i + z_j[X^{'}]_j > \sum_{\substack{i=1 \\ i \neq j}}^{n} z_i[X^{'}]_i + z^\top X_0$$

from which it follows that

$$\sum_{\substack{i=1 \\ i \neq j}}^{n} z_i[X^{'}]_i < 0$$

Since $z_j > 0$ *for all* $j \in \{1, \ \dots \ , n\}$*, this implies that* $[X^{'}]_j < 0$ *for some* j*, which is a contradiction. Therefore, the bound* (11) *holds.* ■

For determining the above upper bound, we have to compute a conservation vector z. This can be done, e.g. by solving the following LP minimization:

$$\begin{aligned} &\min \sum_{j=1}^{n} z_j \\ &s.t. \\ &z^\top \Gamma \leq 0^{1 \times m} \\ &z \succeq 0^{m \times 1} + \varepsilon^{m \times 1}, \ \varepsilon \succ 0^{m \times 1} \end{aligned}$$

Given the initial state $X_0 \in \mathbb{Z}^n_{\geq 0}$, one can derive an n-dimensional hyperrectangle $\mathcal{H}_{X_0}$ containing all the states $X \in \mathcal{Z}_{X_0}$. One corner point of $\mathcal{H}_{X_0}$ is $0^{n \times 1}$ and the farthest point from $0^{n \times 1}$ is X_{max} which is defined as

$$[X_{max}]_j = \frac{z^\top X_0}{z_j}, \qquad j \in \{1, \dots, n\} \tag{12}$$

By means of the non-negative integer points of $\mathcal{H}_{X_0}$, a conservative upper bound can be derived for the maximal length of directed cycle-free paths (i.e. the number of firing

reactions) from X_0 to $X^{'}$ of a subconservative d-CRN:

$$\sum_{i=1}^{l} [c_{max}]_i \leq \prod_{i=1}^{n} ([X_{max}]_i + 1) \tag{13}$$

The above inequality can be used to complete the feasibility problem of reachability and coverability defined by Eqs. ((4), (6)-(10)) and Eqs. ((5) - (10)).

It is also possible to improve the upper bound given in Eq. (13) by making use of subconservativity of a CRN.

Proposition 2. Let us consider a subconservative d-CRN having a conservation vector $z \in \mathbb{R}^n_{>0}$ and initial state $X_0 \in \mathbb{Z}^n_{\geq 0}$. Introduce the notation $X_{max} \in \mathbb{Z}^n_{\geq 0}$ according to Eq. (12) for the vector containing the maximal coordinate values of the reachable states along each dimension. Then any state $X^{'}$ reachable from X_0 is an element of the simplex $\Sigma_{X_{max}}$ defined by X_{max}:

$$\Sigma_{X_{max}} = \left\{ x \in \mathbb{R}^n_{\geq 0} \mid \sum_{i=1}^{n} \frac{x_i}{[X_{max}]_i} \leq 1 \right\} \tag{14}$$

Proof. Let us substitute X_0 into the equation of the above defined simplex (14)

$$\sum_{i=1}^{n} \frac{[X_0]_i}{[X_{max}]_i} = \sum_{i=1}^{n} \frac{[X_0]_i z_i}{z^\top X_0} = \frac{1}{z^\top X_0} \sum_{i=1}^{n} [X_0]_i z_i = 1$$

Let us assume that there exists a state $X^{'}$ such that $X_0 \rightsquigarrow X^{'}$ and $\sum_{i=1}^{n} \frac{[X^{'}]_i}{[X_{max}]_i} > 1$ (i.e. $X^{'}$ is out of $\Sigma_{X_{max}}$ and reachable from X_0). Then the following holds:

$$\frac{z^\top X^{'}}{z_i} \leq \frac{z^\top X_0}{z_i}$$

from which we get

$$\sum_{i=1}^{n} \frac{[X^{'}]_i}{[X_{max}]_i} \leq \sum_{i=1}^{n} \frac{[X_0]_i}{[X_{max}]_i} = 1$$

This is a contradiction. ■

It can be seen that Eq. (14) explicitly contains the non-zero extreme points of the simplex as the entries of X_{max}. We note that $\Sigma_{X_{max}}$ can equivalently be defined using the initial state X_0 and the conservation vector z as

$$\Sigma_{X_{max}} = \left\{ x \in \mathbb{R}^n_{\geq 0} \;\middle|\; z^\top x \leq z^\top X_0 \right\} \tag{15}$$

Due to the subconservativity, instead of the hyperrectangle $\mathcal{H}_{X_0}$, $\Sigma_{X_{max}}$ can be used to construct a sharper upper bound for the number of transitions along a directed cycle-free path.

The number of non-negative integer points $\mathcal{Q}(\Sigma)$ of an n-dimensional simplex Σ defined by the points $[a_1\ 0\ \ldots\ 0]^\top$, $[0\ a_2\ \ldots\ 0]^\top$, $\ldots$, $[0\ 0\ \ldots\ a_n]^\top$, $[0\ 0\ \ldots\ 0]^\top$ can be bounded by the following expression [32]:

$$\mathcal{Q}(\Sigma) \leq \frac{1}{n!}(a_1(1+a)-1)(a_2(1+a)-1)\ \ldots\ (a_n(1+a)-1) \tag{16}$$

where $a = \frac{1}{a_1} + \frac{1}{a_2} + \ldots + \frac{1}{a_n}$, $a_i \geq 1$ $i \in \{1, \ldots, n\}$ and $n \geq 3$. Thus, if $n \geq 3$, the number of non-negative integer points in $\Sigma_{X_{max}}$ is bounded as follows:

$$\mathcal{Q}(\Sigma_{X_{max}}) \leq \frac{1}{n!}\prod_{i=1}^{n}\Big[\frac{1}{z_i}(z^\top X_0 + \sum_{j=1}^{n} z_j) - 1\Big] \tag{17}$$

Furthermore, if $z_i = z_j$ $i, j \in \{1, \ldots, n\}$, then Eq. (17) is simplified as follows:

$$\mathcal{Q}(\Sigma_{X_{max}}) \leq \frac{1}{n!}\Big(\frac{z^\top X_0 - \zeta}{\zeta} + n\Big)^n \tag{18}$$

Using the above inequalities, the new upper bound for the number of reactions along a directed cycle-free path is as follows:

$$\sum_{i=1}^{m} [c_{max}]_i \leq \mathcal{Q}(\Sigma_{X_{max}}) \tag{19}$$

The result of *Proposition 2* can be extended as follows.

Proposition 3. Let us consider a subconservative d-CRN with conservation vector $z \in \mathbb{R}^m_{>0}$ and non-zero initial state $X_0 \in \mathbb{Z}^n_{\geq 0}$. Let us define the maximal coordinate values along each dimension by the vector X_{max} according to (12). Consider an arbitrary non-negative state $X \in \Sigma_{X_{max}}$ and $X^* \notin \Sigma_{X_{max}}$. Then X^* is not reachable from X.

Proof. (Indirect)

Let us assume that there exists a state $X^* \in \mathbb{Z}^n_{\geq 0}$ *such that* $X^* \notin \Sigma_{X_{max}}$ *and* $X \rightsquigarrow X^*$ *where* $X \in \Sigma_{X_{max}}$. *Then*

$$\sum_{i=1}^{n} \frac{[X^*]_i}{[X_{max}]_i} = \sum_{i=1}^{n} \frac{[X^*]_i z_i}{z^\top X_0} > 1$$

from which we get

$$z^\top X^* > z^\top X_0$$

Since X^* *is reachable from* X*:* $\exists c \in \mathbb{Z}^l_{\geq 0}$ *for which*

$$X^* = X + \Gamma c$$

Let us multiply both sides by $z^\top$*:*

$$z^\top (X^* - X) = z^\top \Gamma c$$

Since $X \in \Sigma_{X_{max}}$, *the inequality* $z^\top X \leq z^\top X_0$ *holds which implies that* $z^\top (X^* - X) > 0$ *while* $z^\top \Gamma c < 0$ *which is a contradiction.* ■

Given the target state $X^{'} \in \mathbb{Z}^n_{\geq 0}$ to be reached from a predefined non-zero initial state $X_0 \in \mathbb{Z}^n_{\geq 0}$, consider the non-zero state $X^{''} = X^{'} - v$ where all the entries of v are equal to 0 except for some $j \in \{1, \ \ldots \ , n\}$ for which $v_j = 1$. By means of $X^{''}$ we define the following simplex:

$$\Sigma_{X^{''}_{max}} = \left\{ x \in \mathbb{R}^n_{\geq 0} \mid \sum_{i=1}^{n} \frac{x_i}{[X^{''}_{max}]_i} \leq 1 \right\} \tag{20}$$

where $[X^{''}_{max}]_i = \frac{z^\top X^{''}}{z_i}$ for $i \in \{1, \ldots, n\}$. According to *Proposition 4*, all the reachable states are out of the simplex $\Sigma_{X^{''}_{max}}$, thus one can reduce the bound of the maximal length along directed cycle-free paths from X_0 to $X^{'}$ by the number of non-zero integer points of $\Sigma_{X^{''}_{max}}$:

$$\sum_{i=1}^{m} [c_{max}]_i \leq K^{sub} = \mathcal{Q}(\Sigma_{X_{max}}) - \mathcal{Q}(\Sigma_{X^{''}_{max}}) \tag{21}$$

As a special case, let us consider a subconservative d-CRN $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ for which $r_i \in \mathbb{Z}^n_{\leq 0}$ for all $i \in \{1, \ldots, m\}$. In this case the farthest point of the hyperrectangle $\mathcal{H}_{X_0}$ from $0^{n\times 1}$ is determined by X_0, i.e. $X_{max} = X_0$. Since there is no reaction producing new molecules, for all $X^{'} \in \mathbb{R}^n_{>0}$, $[X^{'}]_i > [X_0]_i$ for some $i \in \{1, \ldots, n\}$, $X^{'}$ is not reachable from X_0. Hence the maximal length of directed cycle-free paths can be bounded by the following term:

$$\sum_{i=1}^{m} [c_{max}]_i \leq \mathcal{Q}(\mathcal{H}_{X_0}) - \left(\mathcal{Q}(\mathcal{H}_{X^{'}}) - 1 \right) \tag{22}$$

3.3.2 Conservative case

In this section we consider conservative systems, i.e CRNs for which $\exists z \in \mathbb{R}^n_{>0}$ such that $z^\top \Gamma = 0$. Note that the conservativity of an n-dimensional CRN can be checked in polynomial time through the following LP problem:

$$\begin{aligned} &\min \sum_{j=1}^{n} z_j \\ &s.t. \\ &z^\top \Gamma = 0^{1\times l} \\ &z \succeq \varepsilon^{l\times 1}, \ \varepsilon \succ 0^{l\times 1} \end{aligned}$$

Due to the scalability of z, the choice of ε is arbitrary.

From the definition of conservativity it follows that $\mathbf{C}_{X_0}$ is a closed bounded hypersurface, hence it can be projected to a simplex of dimension $g = rank(\Gamma)$.

Let us consider the projection $\mathcal{P} : \mathbb{R}^n \to \mathbb{R}^g$ for which the transformation matrix is denoted by $T \in \mathbb{Z}^{n\times g}$. All the integer points of $\mathbf{C}_{X_0}$ – including all possible states reachable

from X_0 (i.e. $\forall\ X \in \mathcal{Z}_{X_0}$) – will also be integer ones after applying the transformation, i.e.:

$$X \in \mathbb{Z}^n_{\geq 0}\ and\ X \in \mathbf{C}_{X_0} \Rightarrow \mathcal{P}(X) = TX \in \mathbb{Z}^g_{\geq 0}$$

Hence we can consider the integer point enumeration problem in the projected space, instead of that of $\mathbf{C}_{X_0}$. It is known that $\mathcal{P}(\mathbf{C}_{X_0})$ is an g-dimensional simplex of some points $[p_1\ 0\ \ldots\ 0]^\top, [0\ p_2\ \ldots\ 0]^\top,\ \ldots, [0\ 0\ \ldots\ p_g]^\top, [0\ 0\ \ldots\ 0]^\top$, where $p_1,\ p_2, \ldots, p_g \in \mathbb{R}_{\geq 0}$. To bound the integer points – by means of Eq. (16) – the exact values of $p_i,\ i \in \{1,\ 2,\ \ldots\ , g\}$ are needed which can be easily bounded from above by X_{max}. In this way, an upper bound for the length of directed cycle-free paths for conservative d-CRNs can be given as:

$$\sum_{i=1}^{m} [c_{max}]_i \leq K^{con} = \frac{1}{g!} \prod_{i=1}^{g} \Big[\frac{1}{z_i} (z^\top X_0 + \sum_{j=1}^{g} z_j) - 1 \Big] \tag{23}$$

3.4 Computational solution of the reachability problem

Proposition 4. Consider a subconservative discrete state CRN $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ of dimension $n \geq 3$. Let us denote the initial state of the system by $X_0 \in \mathbb{Z}^n_{\geq 0}$. Then the following problems can be decided in polynomial time:

(P_1) Let $X' \in \mathbb{Z}^n_{\geq 0}$ be an arbitrary target state. Is X' reachable from X_0?

(P_2) Let $y \in \mathcal{C}$ be an arbitrary complex. Is it possible to reach a state $X' \in \mathbb{Z}^n_{\geq 0}$ from X_0 in such a way that y is charged at X'?

Proof. (Constructive)
The above questions can be answered by the following IP feasibility problems:

(P_1) *Consider the following feasibility problem:*

$$\begin{cases} X_0 + \Gamma c = X' \\ v_j \in \{0,1\}^l \qquad j = 1, \ldots, K \\ \sum_{i=1}^{l} [v_j]_i \leq 1 \qquad j = 1, \ldots, K \\ X_0 + \Gamma \sum_{i=1}^{k} v_i \geq \Gamma_S v_{k+1} \quad k = 1, \ldots, K-1 \\ \sum_{i=1}^{l} [v_j]_i \leq \sum_{i=1}^{l} [v_{j+1}]_i \quad j = 1, \ldots, K-1 \end{cases} \tag{24}$$

(P_2) *Consider the feasibility problem defined by the inequalities and equalities below:*

$$\begin{cases} X_0 + \Gamma c \geq y \\ v_j \in \{0,1\}^l \qquad j = 1, \ldots, K \\ \sum_{i=1}^{l} [v_j]_i \leq 1 \qquad j = 1, \ldots, K \\ X_0 + \Gamma \sum_{i=1}^{k} v_i \geq \Gamma_S v_{k+1} \quad k = 1, \ldots, K-1 \\ \sum_{i=1}^{l} [v_j]_i \leq \sum_{i=1}^{l} [v_{j+1}]_i \quad j = 1, \ldots, K-1 \end{cases} \tag{25}$$

where K is given as K^{sub} in (21) *for subconservative and as K^{con} in* (23) *for conservative CRNs. The lattice of feasible solutions in both cases is represented by the vectors v_j, $j = 1, \ldots, K$. The number of decision variables equals to $K \cdot l$. To check the feasibility of the above problems one has to find an integer lattice point in the feasibility regions defined by the inequalities and equalities. This can be determined in polynomial time using the Lenstra's algorithm, given the fixed dimension $K \cdot l$ of the problems.* ■

It is important to note that, according to Eqs. (21) and (23), K is not polynomial in n, but n is known to be constant for a given CRN. Furthermore K can be greater than the minimally required number of steps for reaching a prescribed target state which may imply the appearance of zero vectors v_j in the solution for some $j \in \{1, \ldots, K\}$. The position of zero vectors among the non-zero ones does not affect the solution but makes redundancy, hence the following ordering constraints are introduced to exclude additional feasible permutations of the same set of reactions:

$$\sum_{i=1}^{l} [v_j]_i \leq \sum_{i=1}^{l} [v_{j+1}]_i \quad j = 1, \ldots, K-1 \tag{26}$$

Once the feasibility problems (24) and (25) are equipped with the above inequality constraints, then each feasible solution represents a distinct path between the initial and target states, hence Barvinok's algorithm can be used to count them. Moreover, the feasibility problems can be easily extended with further linear constraints on the decision variables to decide the feasibility in constrained convex sets while maintaining polynomial time complexity.

We emphasize that the above feasibility problems can be easily equipped with an appropriate linear objective function to form an integer program. For example:

$$\min \left\{ \sum_{j=1}^{K} \sum_{i=1}^{l} [v_j]_i \right\}$$

can be applied to find a path with minimum length from the initial state to target states if at least one exists.

3.5 Reachability in low dimensions

In this section we consider subconservative d-CRNs for which the non-negative stoichiometric compatibility class can be mapped into an at most 2-dimensional simplex. We introduce a distinguished state X^m for which:

$$[X^m]_i = \begin{cases} [M]_i + \max\left\{\left|[r_j]_i\right| : j = 1,\ldots,l\right\} & \text{, if } \exists j\ [r_j]_i < 0 \\ [M]_i & \text{, otherwise} \end{cases} \qquad i = 1,\ldots,n \tag{27}$$

where $M \in \mathbb{Z}^n_{\geq 0}$ is defined as

$$[M]_i = max\left\{[\Gamma_S]_{ij} : j = 1,\ldots,l\right\} \qquad i = 1,\ldots,n \tag{28}$$

and $\Gamma_S \in \mathbb{Z}^{n\times l}_{\geq 0}$ is defined in Eq. (10).

Proposition 5. Let us consider a 2-dimensional subconservative d-CRN with stoichiometric matrix Γ and conservation vector $z \in \mathbb{R}^n_{>0}$. We consider an initial state $X_0 \in \mathbb{Z}^n_{\geq 0}$ and a target state $X' \in \mathbb{Z}^n_{\geq 0}$ such that $X' \succeq X^m$ and $X_0 \succeq X^m$ hold where X^m is characterized by Eq. (27). Then the state $X' \in \mathbb{Z}^n_{\geq 0}$ is reachable from X_0 if and only if the equation

$$\Gamma c = X' - X_0 \tag{29}$$

has a non-negative integer solution c.

Proof.

1. $X_0 \rightsquigarrow X' \implies \exists c \in \mathbb{Z}^l_{\geq 0} : \Gamma c = X' - X_0$
   From the definition of reachability the existence of a nonnegative solution c follows.

2. $X_0 \rightsquigarrow X' \impliedby \exists c \in \mathbb{Z}^l_{\geq 0} : \Gamma c = X' - X_0$ *We can rewrite Eq.* (29) *as*

$$X_0 - M + \Gamma c = X' - M \tag{30}$$

   where M is defined according to Eq. (28) *and we have that $X_0 - M \succeq 0$ and $X' - M \succeq 0$. Clearly, the solution c determines the number of occurrences for each reaction. Let us consider any sequence of k reactions determined by c: $\sigma_r = r_{\nu(1)},\ldots,r_{\nu(k)}$, where $k = \sum_{i=1}^m [c]_i$. If we do not take into account the chargedness of the source complexes (i.e., considering the reaction vectors as simple transition rules without any constraints on their executability) it is possible to reorder the elements of σ_r so*

that all the transition states along the path from $X_1 = X_0 - M$ to $X_{k+1} = X^{'} - M$ are non-negative. The following algorithm returns such a valid permutation of σ_r having non-negative transition states:

Algorithm 1

1: **procedure** REORDER($X_0,\ [r_{\nu(1)},\ r_{\nu(2)},\ \ldots\ , r_{\nu(k)}],\ M$)
2: $X_{current} \leftarrow X_0 - M$
3: **for** $i = 1$ to k **do**
4: **if** $[X_{current} + r_{\nu(i)}]_1 < 0$ **or** $[X_{current} + r_{\nu(i)}]_2 < 0$ **then**
5: *Choose a reaction vector $r_{\nu(j)}$, $i < j \leq k$ for which $X_{current} + r_{\nu(j)} \succeq 0^{2\times 1}$*
6: $r^{'} \leftarrow r_{\nu(i)}$
7: $r_{\nu(i)} \leftarrow r_{\nu(j)}$
8: $r_{\nu(j)} \leftarrow r^{'}$
9: **end if**
10: $X_{current} \leftarrow X_{current} + r_{\nu(i)}$
11: **end for**
12: **return** $[r_{\nu(1)},\ r_{\nu(2)},\ \ldots\ , r_{\nu(k)}]$
13: **end procedure**

The correctness of Algorithm 1 can be proved as follows:

Let us consider the operation of Algorithm 1 with input reaction ordering σ_r. Let us denote the reordered reaction sequence after iteration i of the loop in lines 3–11 of the pseudocode of Algorithm 1 by σ_r^i. Let us denote the jth transition state corresponding to σ_r^i by X_j^i, for $i = 1, \ldots, k$ and $j = 2, \ldots, k$. Furthermore, let the ordered set of states corresponding to σ_r^i be denoted by σ_X^i for $i = 1, \ldots, k$. Clearly, σ_r^i and σ_r^j for $i,\ j \in \{1, \ldots, k\},\ i \neq j$ are not necessarily different. Let us assume that there exists a nonnegative intermediate state X_j^{j-1} for some $j \in \{2, \ldots, k-1\}$ along the state transition sequence σ_X^{j-1} for which the forthcoming state has a negative coordinate, i.e. $[X_{j+1}^{j-1}]_d < 0$ for some $d \in \{1, 2\}$. Since the target state to be reached is a non-negative one, it follows that there exists a reaction in the set $\{r_{\nu(j+1)}, \ldots, r_{\nu(k)}\}$ increasing the state variable along coordinate d. Let us assume that all the reactions increasing the state variable along coordinate d decrease the other coordinate $d^{'}$ of the state to a negative value, i.e. bypassing a state with a negative coordinate is not possible. By the construction of X^m in Eq. (27), this assumption implies that the current state X_j^{j-1} is an interior point of the polyhedron $\mathcal{H}$ which is defined as

$$\mathcal{H} = \{\ x \in \mathbb{R}^n_{\geq 0}\ |\ x \prec X^m - M\}. \tag{31}$$

However, the polyhedron $\mathcal{H}$ is contained in the simplex Σ_{X^m} defined as

$$\Sigma_{X^m} = \left\{ x \in \mathbb{R}^n_{\geq 0} \mid z^\top x \leq z^\top (X^m - M) \right\}. \tag{32}$$

In this case, according to Proposition 2 which holds due to the subconservativity of the network, it is not possible to reach a state X for which $z^\top(X^m - M) \leq z^\top X$. But this is a contradiction, since arbitrary permutation of the input ordering σ_r results in the same non-negative target state X_{k+1}, given the initial state X_1. Then the correctness of Algorithm 1 follows.

The reaction vector sequence $\sigma_r^k = r_{\nu(1)},\ r_{\nu(2)}, \ldots, r_{\nu(k)}$ returned by Algorithm 1 uniquely determines a non-negative state transition sequence:

$$\sigma_X^k = X_1,\ X_2^k, \ldots, X_k^k,\ X_{k+1}. \tag{33}$$

By adding M to each state, σ_X^k can be shifted back to

$$\sigma_X^{'} = X_0,\ X_2^k + M, \ldots, X_k^k + M,\ X^{'}. \tag{34}$$

For the state transition sequence $\sigma_X^{'}$ all the states are non-negative, and furthermore, all reactions are charged at each state. Hence $\sigma_X^{'}$ represents a valid executable state transition sequence from X_0 to $X^{'}$, i.e. $X_0 \rightsquigarrow X^{'}$ holds.

■

Corollary 1. Let us consider a conservative d-CRN $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ of arbitrary dimension with stoichiometric matrix Γ and assume that $rank(\Gamma) \leq 2$. Consider an initial state $X_0 \in \mathbb{Z}^n_{\geq 0}$ and a target state $X^{'} \in \mathbb{Z}^n_{\geq 0}$ for which we assume that $X^{'} \succeq X^m$ and $X_0 \succeq X^m$ where X^m is defined according to (27). Then

$$X_0 \rightsquigarrow X^{'} \quad \Longleftrightarrow \quad \exists c \in \mathbb{Z}^l_{\geq 0} : \quad \Gamma c = X^{'} - X_0$$

Proof.

Assume that $rank(\Gamma) \leq 2$. The state space of a conservative d-CRN for which $rank(\Gamma) \leq 2$ can be mapped into the state space of an at most 2-dimensional subconservative one, because the set of reachable states of an n-dimensional conservative d-CRN resides in a $rank(\Gamma)$-dimensional linear manifold. Then for the projected simplex Proposition 5 holds.

■

Note that both *Proposition* 5 and *Corollary* 1 work for arbitrary choice of the conservation vector z.

The results stated in *Proposition* 5 and *Corollary* 1 substantially mitigate the computational complexity of the optimization frameworks (24) and (25), since in these cases the decomposition of c into sum of distinct vectors v_j, $j = 1, \ldots, K$ and determining the upper bound K are not needed. Hence the respective feasibility problem to be analyzed is simplified as follows:

$$\begin{cases} \Gamma c = X^{'} - X_0 \\ c \in \mathbb{Z}^{l}_{\geq 0} \end{cases} \tag{35}$$

where the decision variable is the vector c which can be involved into an objective function if needed, e.g. for finding shortest path from X_0 to $X^{'}$ we have

$$\min \sum_{j=1}^{l} c_j$$

4 Examples

In this section we illustrate our methods on two examples. Beyond the IP feasibility approach, we equipped Eqs. (24) and Eqs. (25) with a linear objective function of the form $\sum_{j=1}^{m} c_j$. The resulting integer programs are capable of finding state transition sequences from the initial state X_0 to a prescribed target state $X^{'}$ (or a set of well-defined target states $\mathcal{X}^{'}$). We also implemented the next reaction method [29] – as it was presented in [18] – to simulate the stochastic behavior of the studied CRNs. We note again that it is also possible to count the number of feasible solutions using Barvinok's algorithm for which an effective implementation exists [28].

The algorithms were implemented in Python 2.7 and the applied mathematical optimization solver was Gurobi [31]. All the computations were performed on a Lenovo P51s workstation having two 2.70GHz i7-7500U CPUs and 32GB RAM (DDR4 2133 MHz).

4.1 Example 1: A conservative CRN showing extinction events

Let us consider the CRN taken from [22] and shown in Fig. 2.

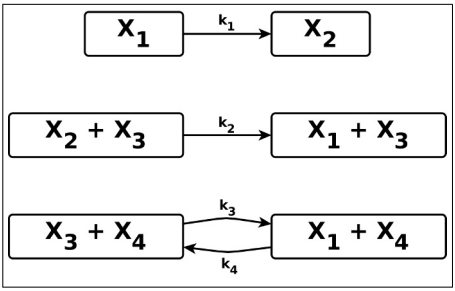

Figure 2. Reaction network structure of *Example 1*

The associated stoichiometric matrix is

$$\Gamma = \begin{bmatrix} -1 & 1 & 1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Based on Γ, it is easy to see that the above CRN is conservative with a possible conservation vector $z = [1,\ 1,\ 1,\ 1]^\top$. Assuming deterministic mass action kinetics, one can describe this CRN with the following system of differential equations:

$$\begin{aligned} \frac{dx_1(t)}{dt} &= -k_1 x_1(t) + k_2 x_2(t) x_3(t) + k_3 x_3(t) x_4(t) - k_4 x_1(t) x_4(t) \\ \frac{dx_2(t)}{dt} &= k_1 x_1(t) - k_2 x_2(t) x_3(t) \\ \frac{dx_3(t)}{dt} &= -k_3 x_3(t) x_4(t) + k_4 x_1(t) x_4(t) \\ \frac{dx_4(t)}{dt} &= 0 \end{aligned}$$

where $x_i(t)$, $i \in \{1,\ 2,\ 3,\ 4\}$ denotes the concentration of the ith species at time t and $k_i,\ i \in \{1,\ 2,\ 3,\ 4\}$ is the reaction rate constant associated to reaction r_i.

The time evolution of the molecular counts in the case of the stochastic d-CRN is given by the following state equations:

$$\begin{aligned} X_1(t) &= X_1(0) - Y_1(k_1 \int_0^t X_1(\tau)d\tau) + Y_2(k_2 \int_0^t X_2(\tau)X_3(\tau)d\tau) - Y_3(k_3 \int_0^t X_3(\tau)X_4(\tau)d\tau) \\ X_2(t) &= X_2(0) + Y_1(k_1 \int_0^t X_1(\tau)d\tau) - Y_2(k_2 \int_0^\tau X_2(\tau)X_3(\tau)d\tau) \\ X_3(t) &= X_3(0) - Y_3(k_3 \int_0^t X_3(\tau)X_4(\tau)d\tau) + Y_4(k_4 \int_0^t X_1(\tau)X_4(\tau)d\tau) \\ X_4(t) &= X_4(0) \end{aligned}$$

where $X_i(t)$, $i \in \{1, \dots, 4\}$ denotes the molecular count for the ith species, and Y_j for $j \in \{1, \dots, 4\}$ are independent unit rate (unit intensity) homogenous Poisson processes.

For a general initial state $X_0 = [K\ L\ M\ N]^\top \in \mathbb{Z}^4_{\geq 0}$, the target state $X' = [0\ (K + L + M)\ 0\ N]^\top$ is reachable which gives an extinction event, since X_2 depletes both X_1 and X_3 which are necessary for firing all of the reactions. In the sequel we justify this reachability argument by employing the proposed IP framework.

We consider this d-CRN with the following particular parametrization: $k_1 = 4.7$, $k_2 = 2.4$, $k_3 = 4.9$, $k_4 = 0.3$. The initial state is $X_0 = [X_1(0)\ X_2(0)\ X_3(0)\ X_4(0)]^\top = [15\ 10\ 20\ 20]^\top$. Assuming deterministic mass action kinetics, our parametrization from X_0 results in equilibrium with $x_1 \approx 12.27$, $x_2 \approx 31.98$, $x_3 \approx 0.75$ and $x_4 = 20$. Considering the discrete state system one can see that in the absence of X_1 and X_3 this CRN has no reaction which is able to fire. Employing the feasibility framework (24) one can check the reachability of such a target state, namely $X' = [0,\ 45,\ 0,\ 20]$. We solved the IP problem (24) equipped with the objective function $\sum_{j=1}^{l} c_j$. From the minimization we get that $c_{opt} = [35\ 0\ 20\ 0]^\top$, i.e. through the first and third reactions one can reach X' during 55 occurrences of these reactions. The determined discrete state transition sequence of the shortest path is depicted in Figure 5.

After running the next reaction method several times from the prescribed initial state X_0, we obtained a representative sample path reaching the critical state X'. This is shown in Figure 3. Note that the depicted state transition sequence is not the shortest path that we determined by the IP.

The feasibility approach has the advantage that it can be easily generalized to check the reachability of a set of states having prescribed properties. It is known that considering deterministic mass action kinetics, the system reaches a positive equilibrium point, for which $x_2(t) \succ x_1(t)$. Modifying the framework (24) one can check the reachability of states significantly differing from the deterministic equilibrium point. Instead of $\Gamma c = X' - X_0$ in (24), we employ the following inequalities for feasibility analysis:

$$\begin{aligned} [X_0 + \Gamma c]_1 &\geq X_2^{det} \\ [X_0 + \Gamma c]_2 &\leq X_1^{det} \\ [X_0 + \Gamma c]_3 &\leq X^{max} \end{aligned} \tag{36}$$

where $X_1^{det} = 20$, $X_2^{det} = 32$ are upper estimates for the deterministic case equilibrium values of x_1 and x_2, respectively, if the initial state is $X_0 = [13\ 32\ 8]^\top$. Based on the IP (24) equipped with the objective $\sum_{j=1}^{l} c_j$ to be minimized, we proved that there exists

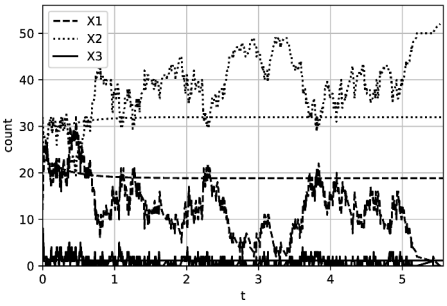

Figure 3. A sample trajectory of *Example 1* for which the finial state is equal to $X^{'} = [0,\ 45,\ 0,\ 20]^{\top}$. For each species the continuous deterministic trajectory (smooth curve) and the stochastic discrete state counterpart are depicted with the same line type. Note that the target state can also be reached by fewer occurrences of the reactions.

a reachable state for which the above conditions hold. The state transition sequence corresponding to the shortest path computed by *Algorithm 1* is depicted in Figure 5.

The existence of such a reachable state is also confirmed through simulation based on the next reaction method. A simulated sample path is depicted in Figure 4 and states of interest are denoted by an arrow: while the deterministic system approaches equilibrium, the stochastic counterpart satisfies the constraint set (36).

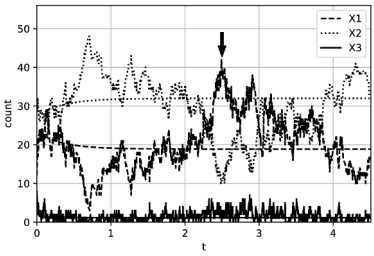

Figure 4. A sample trajectory of *Example 1* for which a subset of the states significantly differ from the equilibrium point of the deterministic mass action system. The arrow is pointing to the states where the conditions (36) hold, though the deterministic counterpart approaches positive equilibrium. The continuous deterministic (smooth) and the stochastic discrete state trajectories of the same species are depicted using the same line type.

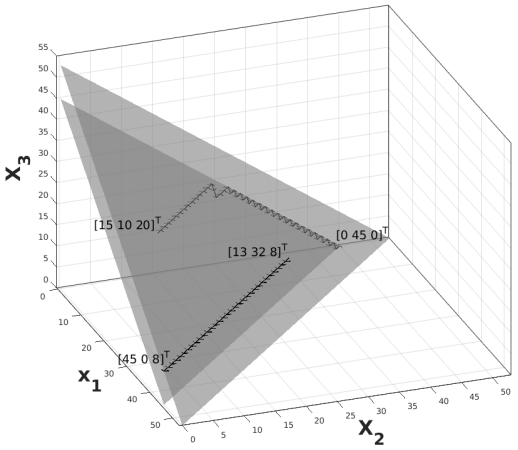

Figure 5. Minimal-length state transition sequences of *Example 1* proving the reachability cases $[13\ 32\ 8]^\top \rightsquigarrow [45\ 0\ 8]^\top$ and $[15\ 10\ 20]^\top \rightsquigarrow [0\ 45\ 20]^\top$, respectively. The conservativity surfaces (positive stoichiometric compatibility classes) associated to the initial states are denoted in gray color.

4.2 Example 2: A subconservative CRN

In the second example we consider a SIRS epidemiological model having three species and four reactions. The reaction network structure is depicted in Figure 6. The model describes the time evolution of the susceptible, infected and recovered species of a closed system. Susceptible species become infected with intensity proportional to β, while infected species get recovered with intensity proportional to γ. The death-rate of infected species is proportional to parameter μ. Recovered species get susceptible with intensity proportional to w.

Assuming large population number, such a system can be characterized by the following deterministic differential equation system:

$$\begin{aligned}
\frac{ds(t)}{dt} &= wr(t) - \beta s(t)i(t) \\
\frac{di(t)}{dt} &= \beta s(t)i(t) - \gamma i(t) - \mu i(t) \\
\frac{dr(t)}{dt} &= \gamma(t)i(t) - wr(t)
\end{aligned}$$

where s, i, and r denote the continuous number of susceptible, infected, and recovered individuals, respectively.

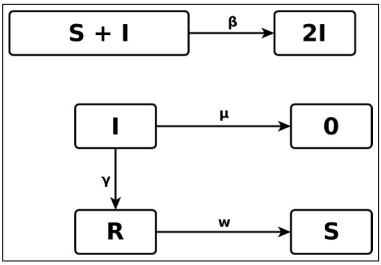

Figure 6. Reaction network structure of the epidemiological model in *Example 2*. Note that **0** denotes the zero-complex.

For the case of small population numbers, we consider the following state equations:

$$S(t) = S(0) + Y_1(w\int_0^t R(\tau)d\tau) - Y_2(\beta\int_0^t S(\tau)I(\tau)d\tau)$$
$$I(t) = S(0) + Y_2(\beta\int_0^t S(\tau)I(\tau)d\tau) - Y_3(\gamma\int_0^t I(\tau)d\tau) - Y_4(\mu\int_0^t I(\tau)d\tau)$$
$$R(t) = R(0) + Y_3(\gamma\int_0^t I(\tau)d\tau) - Y_1(w\int_0^t R(\tau)d\tau)$$

where S, I, and R are the integer numbers of susceptible, infected, and recovered individuals, respectively. Moreover, Y_i for $i \in \{1, \ldots, 4\}$ are independent unit-rate Poisson processes.

The stoichiometric matrix associated to the system is the following:

$$\Gamma = \begin{bmatrix} -1 & 0 & 0 & 1 \\ 1 & -1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

Based on Γ, it can be easily seen that the above system is a subconservative one for which a subconservativity vector is $z = [1\ 1\ 1]^\top$.

In our simulation setting the parameters had the following values: $\beta = 0.18$, $\gamma = 0.9$, $\mu = 0.05$ and $w = 0.39$, respectively. Firstly, we consider the reachability of the target state $X' = [0\ 52\ 0]^\top$ where all the individuals are infected. As initial states we choose $X_0^1 = [50\ 2\ 0]^\top$ and $X_0^2 = [10\ 26\ 10]^\top$.

Employing the IP (24) equipped with $\sum_{j=1}^{l} c_j$ we found that there exist paths from the above initial state to X'. The determined shortest paths are depicted in Figure 7 with black color. Note that these are partially overlapping reaction sequences. The

subconservativity surface which is the same for the two initial states is shown in gray. All the states reachable from X_0^1 and X_0^2 are located on the surface or below that.

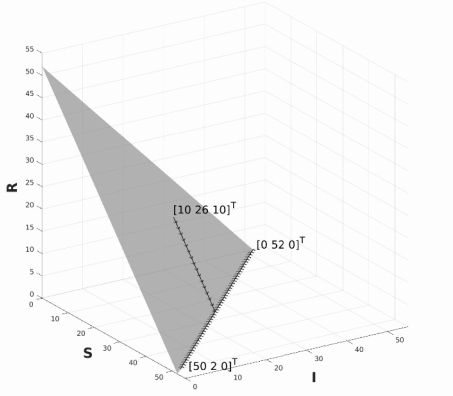

Figure 7. The bounded state space of *Example 2*. The conservativity hypersurface is denoted by gray color. Shortest state transition sequences starting from $[50\ 2\ 0]^\top$ and $[10\ 26\ 10]^\top$, respectively, and reaching $[0\ 52\ 0]^\top$ (where all the individuals are infected) are depicted in black color.

By means of the presented IP framework equipped with cost function $\sum_{i=1}^{m} c_i$ we algorithmically determined a finite reaction sequence starting from $X_0 = [50\ 2\ 0]^\top$ for which the final infected count is equal to zero meaning the complete regression of the disease. This can be achieved by introducing the following linear equality constraint:

$$X'(2) = 0 \tag{37}$$

Such a sample path simulated by the next reaction method is depicted in Figure 8 along the respective deterministic dynamical behavior. In Figure 9*a* one can see a shortest state transition sequence determined by the IP, while in Figure 9*b* the state transition sequence associated to the stochastic sample path in Figure 8 is shown.

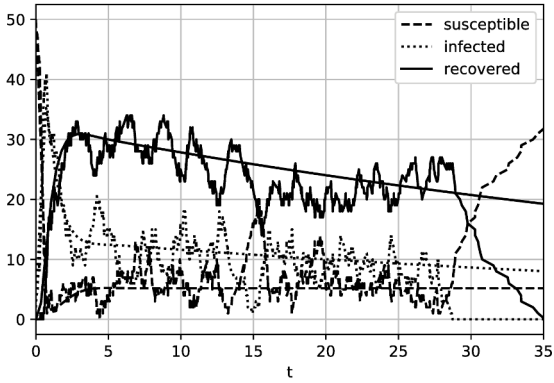

Figure 8. A sample path of *Example 2* for which the dynamical behavior significantly differs from that of the deterministic model. We use the same line type to denote the continuous deterministic (smooth) and the stochastic discrete state trajectories of the same species.

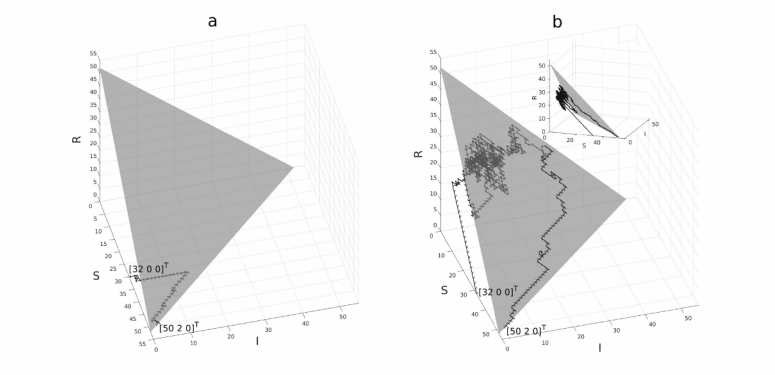

Figure 9. State transition sequences reaching the same target state $[32\ 0\ 0]^\top$ where the disease vanishes. The conservativity surface is denoted by gray color. **a)** the shortest path determined by IP, **b)** a path simulated by the next reaction method.

5 Conclusion

In this paper the reachability problem of (sub)conservative d-CRNs was studied. We characterized a bounded set containing all the reachable states along a valid state transition sequence, given the initial and target states. Upper bounds were also derived for the possible length of cycle-free state transition sequences. A computational method was proposed for reachability and coverability analysis of subconservative d-CRNs. The mathematical constraints associated to reachability and coverability can be reformulated as integer programming feasibility problems. This way the problems are reduced to finding integer lattice points in a convex polytope that is solvable by Lenstra's algorithm in polynomial time if the number of dimensions is fixed [24]. Since their state space is bounded, the subconservativity of the d-CRNs is exploited in the formulation of the feasibility problems. This approach can also be used to determine the number of acyclic paths between predefined initial and target states using Barvinok's lattice point counting algorithm [27]. Beyond the decidability, it is also possible to extend the computational model with appropriate linear objective functions and constraints to obtain integer programs so that one can determine the shortest path between two states satisfying a finite set of predefined constraints. Furthermore, it was shown that in the case of d-CRNs having a stoichiometric matrix with rank lower than equal to 2, the reachability problem is equivalent to the non-negative integer solution of the respective d-CRN state equation under some additional conditions. The corresponding proof relies on an algorithm capable of reordering any non-negative integer solution of the state equation in order to get a non-negative state transition sequence from the prescribed initial state to the target one.

Acknowledgments: The authors acknowledge the support of grants 115694 and 125739 of the National Research, Development and Innovation Office – NKFIH. The research has also been supported by the European Union, co-financed by the European Social Fund through the grant EFOP-3.6.3-VEKOP-16-2017-00002. The authors thank the anonymous reviewers for their constructive comments. The authors are also grateful to Dr. Bernadett Ács for carefully reading the manuscript.

References

[1] F. Horn, R. Jackson, General mass action kinetics, *Arch. Rat. Mech. Anal.* **47.** (1972) 81–116.

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

[3] D. Angeli, A tutorial on Chemical Reaction Network Dynamics, *Eur. J. Control* **15.** (2009) 398–406.

[4] V. Chellaboina, S. P. Bhat, W. M. Haddad, D. S. Bernstein, Modeling and analysis of mass-action kinetics nonnegativity, realizability, reducibility, and semistability., *IEEE Contr. Syst. Mag.* **29** (2009) 60–78.

[5] J. H. Espenson, *Chemical Kinetics and Reaction Mechanisms*, McGraw–Hill, Singapore, 1995.

[6] D. F. Anderson, A proof of the global attractor conjecture in the single linkage class case, *SIAM J. Appl. Math.* **71** (2011) 1487–1508.

[7] G. Szederkényi, K. M. Hangos, T. Péni, Maximal and minimal realizations of reaction kinetic systems: computation and properties, *MATCH Commun. Math. Comput. Chem.* **65** (2011) 309–332.

[8] M. D. Johnston, D. Siegel, Linear conjugacy of chemical reaction networks, *J. Math. Chem.* **49** (2011) 1263–1282.

[9] I. R. Epstein, J. A. Pojman, *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos*, Oxford Univ. Press, Oxford, 1998.

[10] B. Ács, G. Szederkényi, Z. Tuza, Z. A. Tuza, Computing all possible graph structures describing linearly conjugate realizations of kinetic systems, *Comp. Phys. Comm.* **204** (2016) 11–20.

[11] B. Ács, G. Szederkényi, Z. A. Tuza, Z. Tuza, Computing linearly conjugate weakly reversible kinetic structures using optimization and graph theory, *MATCH Commun. Math. Comput. Chem.* **74** (2015) 489–512.

[12] G. Szederkényi, Computing sparse and dense realizations of reaction kinetic systems, *J. Math. Chem.* **47** (2010) 551–568.

[13] M. D. Johnston, D. Siegel, G. Szederkényi, A linear programming approach to weak reversibility and linear conjugacy of chemical reaction networks, *J. Math. Chem* **50** (2012) 274–288.

[14] M. D. Johnston, A linear programming approach to dynamical equivalence, linear conjugacy, and the deficiency one theorem, *J. Math. Chem.* **54** (2016) 1612–1631.

[15] G. Szederkényi, B. Ács, G. Szlobodnyik, Structural analysis of kinetic systems with uncertain parameters, *IFAC–PapersOnLine* **49** (2016) 24–27.

[16] T. G. Kurtz, The relationship between stochastic and deterministic models for chemical reactions, *J. Chem. Phys.* **57** (1972) 2976–2978.

[17] T. G. Kurtz, Strong approximation theorems for density dependent Markov chains, *Stoch. Proc. Appl.* **6** (1978) 223–240.

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

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

[20] C. J. Myers, *Engineering Genetic Circuits*, CRC Press, Boca Raton, 2016.

[21] D. F. Anderson, G. A. Enciso, M. D. Johnston, Stochastic analysis of chemical reaction networks with absolute concentration robustness, *J. R. Soc. Interface* **11** (2014) 1309–1339.

[22] 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.

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

[24] H. W. Lenstra, Integer Programming with a fixed number of variables, *Math. Oper. Res.* **8** (1983) 538–548.

[25] A. K. Lenstra, H. W. Lenstra, L. Lovasz, Factoring polynomials with rational coefficients, *Math. Ann.* **261** (1982) 515–534.

[26] K. Aardal, R. Weismantel, L. A. Wolsey, Non–standard approaches to integer programming, *Discr. Appl. Math.* **123** (2002) 5–74.

[27] 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.

[28] 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.

[29] M. A. Gibson, J. Bruck, Efficient exact stochastic simulation of chemical systems with many species and many channels, *J. Phys. Chem. A* **104** (2000) 1876–1889.

[30] B. Korte, J. Vygen, *Combinatorial Optimization: Theory and Algorithms*, Springer, Berlin, 2006.

[31] Gurobi Optimization Inc., *Gurobi Optimizer Reference Manual* , www.gurobi.com, 2016.

[32] S. T. Yau, L. Zhang, An upper estimate of integral points in real simplices with an application to singularity theory, *Math. Res. Lett.* **13** (2006) 911–921.

[33] L. Paulevé, G. Craciun, H. Koeppl, Dynamical properties of discrete reaction networks, *J. Math. Biol.* **69** (2014) 55–72.

[34] G. Memmi, G. Roucairol, Linear algebra in net theory, in: W. Brauer (Ed.), *Net Theory and Applications*, Springer, Berlin, 1975, pp. 213–223.

[35] R. Brijder, Dominance and T-invariants for Petri nets and chemical reaction networks, in: A. Phillips, P. Yin (Eds.), *Proceedings of the 21st International Conference on DNA Computing and Molecular Programming (DNA 21)*, Springer, Cham, 2015, pp. 1–15.

[36] R. Brijder, Dominance and deficiency for Petri nets and chemical reaction networks, *Nat. Comput.* **16** (2017) 285–294.