Transition graph decomposition for complex balanced reaction networks with non-mass-action kinetics

Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 19(8): 7649–7668.
DOI: 10.3934/mbe.2022359
Received: 19 February 2022
Revised: 14 April 2022
Accepted: 18 April 2022
Published: 24 May 2022

Research article

Transition graph decomposition for complex balanced reaction networks with non-mass-action kinetics

Daniele Cappelletti[1,*] **and Badal Joshi**[2]

[1] DISMA-Dipartimento di Scienze Matematiche "G.L. Lagrange", Politecnico di Torino, Torino, Italy

[2] Department of Mathematics, California State University San Marcos, San Marcos, USA

* **Correspondence:** Email: daniele.cappelletti@polito.it.

Abstract: Reaction networks are widely used models to describe biochemical processes. Stochastic fluctuations in the counts of biological macromolecules have amplified consequences due to their small population sizes. This makes it necessary to favor stochastic, discrete population, continuous time models. The stationary distributions provide snapshots of the model behavior at the stationary regime, and as such finding their expression in terms of the model parameters is of great interest. The aim of the present paper is to describe when the stationary distributions of the original model, whose state space is potentially infinite, coincide exactly with the stationary distributions of the process truncated to finite subsets of states, up to a normalizing constant. The finite subsets of states we identify are called *copies* and are inspired by the modular topology of reaction network models. With such a choice we prove a novel graphical characterization of the concept of complex balancing for stochastic models of reaction networks. The results of the paper hold for the commonly used mass-action kinetics but are not restricted to it, and are in fact stated for more general setting.

Keywords: reaction networks; stochastic models; state space truncation; stationary distribution

1. Introduction

Reaction networks are mathematical models for studying the evolution of biochemical systems, widely used in applications. Formally, reaction networks are made of a set of species whose populations evolve in time guided by a set of reactions such as the ones in the example below:

$$A + B \underset{\kappa_2}{\overset{\kappa_1}{\rightleftharpoons}} 2C \ , \ A \underset{\kappa_4}{\overset{\kappa_3}{\rightleftharpoons}} B. \tag{1.1}$$

Each species has a discrete number of molecules whose counts change as the molecules react or are produced from reactions. In many circumstances, the counts are sufficiently high so that the granularity

may be ignored and it suffices to model species concentrations rather than counts. However, several biological macromolecules such as proteins and enzymes are present only in small numbers (~ 100) and it becomes necessary to both track individual numbers and be mindful of stochastic fluctuations in those numbers. Due to such considerations, $\mathbb{Z}^n$ is the natural state space for the stochastic model (a continuous-time Markov chain) that we consider here. For readable references on stochastic reaction network models and connections with applications, see [1, 2].

The stationary distribution of a stochastic system, assuming it exists, describes the state of the system over a long period of time. So in a sense it is an attracting steady state of the dynamic model. Having an analytic expression of the stationary distribution in terms of the model parameters is useful for various reasons: first of all, the expected level and the fluctuations of different proteins in stationary biological systems is of great interest. Controlling the stationary regime is also attracting an increasing interest, given the novel technological possibilities of changing the cell DNA and, hence, the rate and the form of the biological reactions occurring inside. Notable examples of how the emerging field of synthetic biology focuses on the control of stationary regimes are given in [3–5]. Finally, the approximation of complex biological systems with reaction rates spanning over different orders of magnitudes can be performed if the stationary distribution of the faster subsystem is known [6]. The study of multiscale biological system is of crucial importance and is recognized as one of the key ingredients to unlock the function of genetic and cellular compositions in the Perspectives in Mathematical Biology individuated by the European Society for Mathematical and Theoretical Biology (ESMTB) [7, 8].

The aim of the present paper is twofold: first, typical computational techniques to calculate an approximation of the stationary distribution of interest consists in truncating the (potentially infinite) state space to a smaller finite set, and calculate the stationary distribution of the process restricted to it [9, 10]. The approximation error is often hard to calculate. A natural question is whether the same stationary distribution of the full model is also a stationary distribution of the restricted one; if that is the case then the approximation error only concerns the normalizing constant, but not the ratio of the distribution in different states. In the present paper we consider restrictions of the state space to the images of *copies* (or union thereof). A copy of a reaction network is an embedding of the reaction graph into the state space. Figure 1 illustrates how the state space of

$$\begin{array}{ccc} & A+B & \\ \nearrow & & \searrow \\ 0 & \longleftarrow & A \end{array} \tag{1.2}$$

can be decomposed into copies.

We prove that the process restricted to arbitrary unions of copies has the same stationary measure of the original one only in the case of complex balancing. In the case of complex balanced systems, an exact formula for the stationary distributions is already known [11], however the general line of thought can be pursued for other types of restrictions different from copies. In fact, in future work we will use this idea to efficiently calculate exact analytical expressions of stationary distributions that have not appeared in the existing literature.

Secondly, the paper focuses on the connection between stationary distributions and graphical symmetries. This perspective is based on links between steady states and graphical features of the model which have been fruitfully and extensively studied for deterministic models of reaction networks [12–20]. Connections between graphical symmetries of the model and the shape of its stationary distribution have been successfully unlocked in several recent papers [11, 21–25]. In this work, we

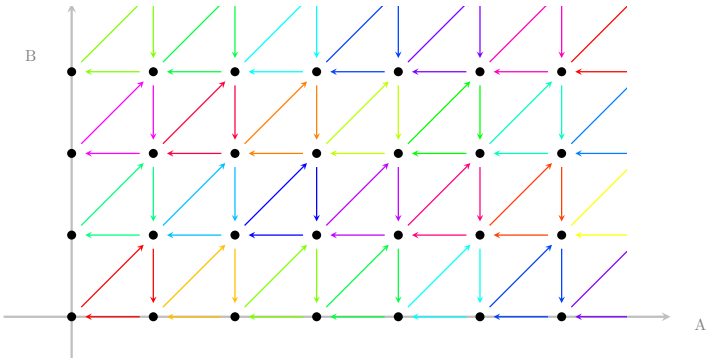

Figure 1. The figures depicts the transition graph of the continuous time Markov chain associated with the reaction network (1.2). Different colors indicate different *copies*, which are different embedding of the graph (1.2) in the state space $\mathbb{Z}^2_{\geq 0}$.

define a new type of graphical symmetry (balancing of copies, as defined below) and we show that it characterizes the notion of complex balancing under general assumptions.

While our study is valid for the commonly used mass-action kinetics, the scope of our results is not restricted to such kinetics but is much broader. Some general results can be stated for *any admissible choice of reaction rates*, and others hold for a generalization of mass-action kinetics that we call *product-form kinetics*, considered for example in [11, 26–28].

2. Background

2.1. Notation

Let $\mathbb{R}$, $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_{>0}$ represent the reals, the non-negative reals and the positive reals, respectively. Let $\mathbb{Z}$, $\mathbb{Z}_{\geq 0}$ and $\mathbb{Z}_{>0}$ represent the integers, the non-negative integers and the positive integers, respectively. For $v \in \mathbb{R}^n$, $\|v\|_1 = |v_1 + \ldots + v_n|$. For $v, w \in \mathbb{R}^n$, $v \leq w$ ($v < w$) means that $v_i \leq w_i$ ($v_i < w_i$) for all $i \in \{1, \ldots, n\}$. For $v, w \in \mathbb{R}^n$, we define

$$1_{\{v \leq w\}} = \begin{cases} 1 & , \quad v \leq w \\ 0 & , \quad \text{otherwise.} \end{cases}$$

If $v > 0$ then v is said to be positive. Similarly, if $v \geq 0$ then v is said to be nonnegative. Finally, $\operatorname{supp} v$ denotes the index set of the non-zero components. For example, if $v = (0, 1, 1)$ then $\operatorname{supp} v = \{2, 3\}$.

If $x \in \mathbb{R}^n_{\geq 0}$ and $v \in \mathbb{Z}^n_{\geq 0}$, we define

$$x^v = \prod_{i=1}^{n} x_i^{v_i}, \quad \text{and} \quad v! = \prod_{i=1}^{n} v_i!,$$

with the conventions that $0! = 1$ and $0^0 = 1$.

2.2. Reaction networks

A reaction network is a triple $\mathcal{G} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$, where $\mathcal{S}$ is a set of n species, $\mathcal{C}$ is a set of m complexes, and $\mathcal{R} \subseteq \mathcal{C} \times \mathcal{C}$ is a set of r reactions, such that $(y, y) \notin \mathcal{R}$ for all $y \in \mathcal{C}$. The complexes are linear combinations of species over $\mathbb{N}$, identified as vectors in $\mathbb{R}^n$. A reaction $(y, y') \in \mathcal{R}$ is denoted by $y \to y'$, and the vector $y' - y$ is the corresponding *reaction vector*. We require that every species has a nonzero coordinate in at least one complex and that every complex appears in at least one reaction. With this convention, there are no "redundant" species or complexes and $\mathcal{G}$ is uniquely determined by $\mathcal{R}$. In (1.1), there are $n = 3$ species (A, B, C), $m = 4$ complexes $(A + B, 2C, A, B)$, and $r = 4$ reactions.

A reaction network $\mathcal{G} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ can be viewed as a graph with node set $\mathcal{C}$ and edge set $\mathcal{R}$ in a natural manner. We will frequently use the viewpoint of a reaction network as a graph in the rest of the paper.

A reaction network $\mathcal{G}$ is *weakly reversible* if every reaction $y \to y' \in \mathcal{R}$ is contained in a closed directed path. Moreover, $\mathcal{G}$ is *reversible* if for any reaction $y \to y' \in \mathcal{R}$, $y' \to y$ is in $\mathcal{R}$. It is clear that each reversible reaction network is also weakly reversible. As an example, the network in (1.1) is reversible, and therefore weakly reversible.

The *stoichiometric subspace* of $\mathcal{G}$ is the linear subspace of $\mathbb{R}^n$ generated by the reaction vectors, namely

$$S = \text{span}(y' - y | y \to y' \in \mathcal{R}).$$

For $v \in \mathbb{R}^n$, the sets $(v + S) \cap \mathbb{R}^n_{\geq 0}$ are called the *stoichiometric compatibility classes* of $\mathcal{G}$.

2.3. Reaction systems

We will consider dynamics of a reaction network with n species both on $\mathbb{R}^n_{\geq 0}$ and $\mathbb{Z}^n_{\geq 0}$. $\mathbb{R}^n_{\geq 0}$ is the usual underlying state space for deterministic models, while for classic stochastic models the state space is $\mathbb{Z}^n_{\geq 0}$. It is worth mentioning that dynamics of a reaction network can also be modeled via stochastic differential equations, an example of stochastic models whose underlying state space is $\mathbb{R}^n_{\geq 0}$ [29–32]. However we do not consider this case in this paper.

2.3.1. Deterministic dynamics

Let $\mathcal{G}$ be a reaction network. We want to associate each reaction $y \to y' \in \mathcal{R}$ with a *rate function* $\lambda_{y \to y'}$, whose domain is $\mathbb{R}^n_{\geq 0}$. Formally, we define as *deterministic kinetics* the following correspondence between reactions and rate functions:

$$\Lambda : (y \to y') \mapsto \lambda_{y \to y'}$$

It is required that for each $y \to y' \in \mathcal{R}$

$$\lambda_{y \to y'}(x) > 0 \quad \text{only if } \operatorname{supp} y \subseteq \operatorname{supp} x. \tag{2.1}$$

The pair $(\mathcal{G}, \Lambda)$ is called *continuous reaction system*. In the deterministic context, the evolution of the species concentrations $z(t) \in \mathbb{R}^n_{\geq 0}$ is determined by the system of ODEs

$$\frac{dz}{dt} = \sum_{y \to y' \in \mathcal{R}} (y' - y) \lambda_{y \to y'}(z). \tag{2.2}$$

Condition (2.1) implies that any solution to (2.2) is non-negative at all times for which it is defined. The solution $z(t)$ is also confined to its *stoichiometric compatibility class*:

$$z(t) \in (z(0) + S) \cap \mathbb{R}^n_{\geq 0}.$$

A state $c \in \mathbb{R}^n$ is said to be a *steady state* of a continuous reaction system $(\mathcal{G}, \Lambda)$ if

$$\sum_{y \to y' \in \mathcal{R}} (y' - y)\lambda_{y \to y'}(c) = 0.$$

2.3.2. Stochastic dynamics

Let $\mathcal{G}$ be a reaction network. We now want to associate each reaction $y \to y' \in \mathcal{R}$ with a *rate function* $\lambda_{y \to y'}$, whose domain is $\mathbb{Z}^n_{\geq 0}$. Similarly to before, we define as *stochastic kinetics* the following correspondence between reactions and rate functions:

$$\Lambda : (y \to y') \mapsto \lambda_{y \to y'}.$$

We require

$$\lambda_{y \to y'}(x) > 0 \quad \text{only if } x \geq y. \tag{2.3}$$

The pair $(\mathcal{G}, \Lambda)$ is called *stochastic reaction system*. In this setting, a state $x = (x_1, \ldots, x_n) \in \mathbb{Z}^n_{\geq 0}$ represents the counts x_i of each species $i = 1, \ldots, n$. $X(t)$ represents the state of the system at time t, and is considered to be a continuous-time Markov chain with transition rate from state x to state x' given by

$$q(x, x') = \sum_{\substack{y \to y' \in \mathcal{R} \\ y' - y = x' - x}} \lambda_{y \to y'}(x).$$

Equivalently, each rate function described the rate of occurrence of the associated reaction, and whenever a reaction $y \to y'$ takes place the process $X(\cdot)$ moves from the current state x to $x + y' - y$. Condition (2.3) forces the process $X(\cdot)$ to the positive orthant.

We say that $x' \in \mathbb{Z}^n_{\geq 0}$ is *accessible* from $x \in \mathbb{Z}^n_{\geq 0}$, if there is a sequence of states $(x = u_0, u_1, \ldots, u_{n-1}, u_n = x')$ such that for each consecutive pair of states (u_i, u_{i+1}), $0 \leq i \leq n - 1$, we have $q(u_i, u_{i+1}) > 0$. A non-empty set $\Gamma \subseteq \mathbb{Z}^n_{\geq 0}$ is an *closed irreducible set* of $(\mathcal{G}, \Lambda)$ if for all $x \in \Gamma$ and all $u \in \mathbb{Z}^n_{\geq 0}$, u is accessible from x if and only if $u \in \Gamma$ [33]. A probability distribution π is a stationary distribution for a continuous-time Markov chain if for all $t \in \mathbb{R}_{\geq 0}$ and $x \in \mathbb{Z}^n_{\geq 0}$

$$P(X(t) = x | X(0) \sim \pi) = \pi(x).$$

By standard theory on Markov chains, the support of π (that is, the largest set of states where π has a positive value) is a union of closed irreducible sets [33]. Moreover, by standard theory on Markov chains, if the continuous-time Markov chain is non-explosive (in the sense of [33]) then a probability distribution π is stationary if and only if for all states x

$$\pi(x) \sum_{y \to y' \in \mathcal{R}} \lambda_{y \to y'}(x) = \sum_{y \to y' \in \mathcal{R}} \pi(x + y - y')\lambda_{y \to y'}(x + y - y'). \tag{2.4}$$

We further give the definitions below.

Definition 2.1. Let $(\mathcal{G}, \Lambda)$ be a discrete reaction system and let μ be a measure on $\mathbb{Z}^n_{\geq 0}$.

1) μ is said to be a *stationary measure* if for all $x \in \mathbb{Z}^n_{\geq 0}$

$$\mu(x) \sum_{y \to y' \in \mathcal{R}} \lambda_{y \to y'}(x) = \sum_{y \to y' \in \mathcal{R}} \mu(x + y - y') \lambda_{y \to y'}(x + y - y'). \tag{2.5}$$

2) μ is said to be a *σ-finite measure* if $\mu(x) < \infty$ for all $x \in \mathbb{Z}^n_{\geq 0}$. π is said to be a *finite measure* if $\sum_{x \in \mathbb{Z}^n_{\geq 0}} \mu(x) < \infty$.

Note that based on the definitions used in this paper, if a stationary measure is a probability distribution is not necessarily a stationary distribution, unless the model is non-explosive. Non-explosiveness of the process can be assessed if it is *complex balanced*, as stated in Theorem 3.3 below.

2.3.3. Mass-action kinetics

An important choice of kinetics is *mass-action kinetics*.

Definition 2.2. Consider a reaction network $\mathcal{G} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$.

1) A *deterministic mass-action system* is a continuous reaction system $(\mathcal{G}, \Lambda^D_\kappa)$ with

$$K^D_\kappa(y \to y')(z) = \kappa_{y \to y'} z^y$$

for some constants $\kappa_{y \to y'} \in \mathbb{R}_{>0}$, called *rate constants*.

2) A *stochastic mass-action system* is a discrete reaction system $(\mathcal{G}, K^S_\kappa)$ with

$$K^S_\kappa(y \to y')(x) = \kappa_{y \to y'} \frac{x!}{(x-y)!} 1_{\{x \geq y\}} \tag{2.6}$$

for some constants $\kappa_{y \to y'} \in \mathbb{R}_{>0}$, called *rate constants*.

Many generalizations of mass-action kinetics have been proposed both for stochastic and deterministic models. In the present paper, we will study the following generalization for stochastic models, considered for example in [11, 26–28]. We define the *stochastic product-form kinetics* as the discrete kinetics

$$K^S_{\kappa,\theta}(y \to y')(x) = \kappa_{y \to y'} \prod_{i=1}^{n} \prod_{j=0}^{y_i - 1} \theta_i(x_i - j), \tag{2.7}$$

for some constants $\kappa_{y \to y'} \in \mathbb{R}_{>0}$, called *rate constants*, and some functions $\theta_i \colon \mathbb{Z} \to \mathbb{R}_{\geq 0}$ satisfying $\theta_i(m) = 0$ if and only if $m \leq 0$. A stochastic reaction system $(\mathcal{G}, K^S_{\kappa,\theta})$ will be called *stochastic product-form system*.

Note that if $\theta_i(m) = m 1_{\{m \geq 0\}}$ for all i, then $K^S_{\kappa,\theta} = K^S_\kappa$. Finally, we give the following definition:

Definition 2.3. A stochastic product-form system is called *non-saturating* if $\lim_{m \to \infty} \theta_i(m) = \infty$ for all $1 \leq i \leq n$.

3. Complex balancing

The definition of complex balancing dates back to [12, 13], where deterministic models were of interest. The definition is the following:

Definition 3.1. Let $(\mathcal{G}, \Lambda)$ be a continuous reaction system. A state $c \in \mathbb{R}^n_{\geq 0}$ is *complex balanced* if for all complexes $y \in \mathcal{C}$ we have

$$\sum_{y' \in \mathcal{C} : y \to y' \in \mathcal{R}} \lambda_{y \to y'}(c) = \sum_{y' \in \mathcal{C} : y' \to y \in \mathcal{R}} \lambda_{y' \to y}(c).$$

In words, c is complex balanced if the sum of rates of reactions "entering" any complex y is equal to the sum of the rates of reactions "exiting" from y, calculated in c. It is known that every complex balanced state is a steady state, while the opposite is not true in general [12]. The following are classical results from [12, 13]:

Theorem 3.1. *Let $(\mathcal{G}, K^D_\kappa)$ be a deterministic mass-action system. If a positive complex balanced state c exists, then $\mathcal{G}$ is weakly reversible, all positive steady states are complex balanced, and exactly one positive steady state exists within each stoichiometric compatibility class.*

Theorem 3.1 implies that the positive steady states of a deterministic mass-action system are either all complex balanced, or none of them is. Moreover, in the former case non-positive steady states are necessarily complex balanced as well as shown in [21, Theorem 4]. Mass- action systems are called *complex balanced* if at least one positive steady state is complex balanced, which by Theorem 3.1 is equivalent to the existence of at least a positive steady state and to all of them being complex balanced.

In [11] the first connection between complex balancing and stochastic dynamics is performed, and the following result is proven.

Theorem 3.2. *Let $\mathcal{G}$ be a reaction network. Assume the deterministic mass-action system $(\mathcal{G}, K^D_\kappa)$ is complex balanced and $c \in \mathbb{R}^n_{>0}$ is a positive complex balanced state. Then the stochastic reaction system $(\mathcal{G}, K^S_{\kappa,\theta})$ with $K^S_{\kappa,\theta}$ defined as in* (2.7) *and the same rate constants as $(\mathcal{G}, K^D_\kappa)$ has a stationary measure of the form*

$$\mu(x) = c^x \prod_{i=1}^{n} \prod_{j=1}^{x_i} \frac{1}{\theta_i(j)} \tag{3.1}$$

If $K^S_{\kappa,\theta}$ reduced to mass-action kinetics, then (3.1) becomes proportional to a product-form Poisson distribution, which is a stationary distribution because the process is non-explosive as proven in [34]. In [26, Theorem 4.1] mild conditions implying finiteness of (3.1) and non-explosiveness of the model are derived:

Theorem 3.3. *Let $\mathcal{G}$ be a reaction network. Assume the deterministic mass-action system $(\mathcal{G}, K^D_\kappa)$ is complex balanced and $c \in \mathbb{R}^n_{>0}$ is a positive complex balanced state. Consider a stochastic non-saturating product-form system $(\mathcal{G}, K^S_{\kappa,\theta})$ with the same rate constants as $(\mathcal{G}, K^D_\kappa)$. Then, the associated process $X(\cdot)$ is non-explosive and the measure* (3.1) *is finite.*

Theorem 3.1, non-explosiveness of the process and finiteness of (3.1) imply that (3.1) is proportional to a stationary distribution of $X(\cdot)$. In [21] a parallel theory on complex balancing is developed for stochastic models. A complex balanced measure is defined as follows.

Definition 3.2. Let $(\mathcal{G}, \Lambda)$ be a discrete reaction system. A measure μ with support in $\mathbb{Z}^n_{\geq 0}$ is *complex balanced* if for all complexes $y \in C$ and all states $x \in \mathbb{Z}^n_{\geq 0}$ we have

$$\sum_{y' \in C\,:\, y \to y' \in \mathcal{R}} \lambda_{y \to y'}(x)\mu(x) = \sum_{y' \in C\,:\, y' \to y \in \mathcal{R}} \lambda_{y' \to y}(x + y' - y)\mu(x + y' - y). \tag{3.2}$$

In words, a measure is complex balanced if, at any state x, the sum of the ingoing fluxes of reactions "entering" any complex y is equal to the sum of the outgoing fluxes of reactions "exiting" from y. It is not difficult to check that a complex balanced measure is necessary stationary, while the converse does not hold in general. Definition 3.2 was given for probability distributions in [21] and then extended to measures in [23]. In particular, it is proven in [23, Proposition 4.13] that under the hypothesis of mass-action a σ-finite complex balanced measure is finite, hence normalizable to a probability distribution. A strong connection between complex balancing in stochastic and deterministic models exists, as stated below.

Theorem 3.4. *A stochastic non-saturating product-form system $(\mathcal{G}, K^S_{\kappa,\theta})$ has a σ-finite, positive, complex balanced measure if and only if the deterministic mass-action system $(\mathcal{G}, K^D_\kappa)$ with the same choice of rate constants is complex balanced.*

Theorem 3.4 is proven in [21] in the case of stochastic mass-action kinetics. The proof for the more general result above is given in Section 6.

4. Main results

Here we state and prove our main results. Perhaps it is worth mentioning that our results hold not only for mass-conserving systems with only finitely many accessible states but also for more general systems where the number of accessible states may be infinite.

Definition 4.1. Let $\mathcal{G}$ be a reaction network. A function $f : C \to \mathbb{Z}^n_{\geq 0}$ is a *copy of* $\mathcal{G}$ if for every $y \to y' \in \mathcal{R}$, we have $f(y') - f(y) = y' - y$. We further say that a copy f is *active* if every $y \to y' \in \mathcal{R}$ is active at $f(y)$.

Denote by $f(C)$ the image of f. $\mathcal{G}$ induces a directed graph on $f(C)$ in a natural manner by associating a directed edge from $f(y)$ to $f(y')$ whenever $y \to y' \in \mathcal{R}$. With this understanding, the inclusion copy $\iota : C \hookrightarrow \mathbb{Z}^n_{\geq 0}$ corresponds to the geometrically embedded graph of $\mathcal{G}$ defined in [35]. We state here a lemma that will be useful to prove our main results.

Lemma 4.1. *Let $(\mathcal{G}, \Lambda)$ be a stochastic reaction system. Let $y \to y'$ be a reaction active at a state x. Then, there exists an injective copy f of $\mathcal{G}$ with $f(y) = x$. Moreover, if Λ is such that $\lambda_{y \to y'}(\tilde{x}) > 0$ if and only if $\tilde{x} \geq y$, then there exists an injective, active copy f of $\mathcal{G}$ with $f(y) = x$.*

Proof. Let $h = x - y$. Since $y \to y'$ is active at x, it follows from (2.3) that $h \in \mathbb{Z}^n_{\geq 0}$. Consider the function $f : C \to \mathbb{Z}^n_{\geq 0}$ defined by $f(\widetilde{y}) = h + \widetilde{y}$. We have $f(y) = h + y = x$. Moreover, for all two complexes $\widetilde{y}, \widetilde{y}' \in C$ we have $f(\widetilde{y}') - f(\widetilde{y}) = \widetilde{y}' - \widetilde{y}$, so f is a copy of $\mathcal{G}$ and is injective. Moreover, if $\lambda_{y \to y'}(\tilde{x}) > 0$ if and only if $\tilde{x} \geq y$ then f is active. Hence, the proof is concluded.

Definition 4.2. Let $(\mathcal{G}, \Lambda)$ be a stochastic reaction system. Let f be a copy of $\mathcal{G}$ and let μ be a measure on $\mathbb{Z}^n_{\geq 0}$.

- f is *active with respect to* (Λ, μ) if for all $y \to y' \in \mathcal{R}$

$$\mu(f(y)) \sum_{\substack{\widetilde{y}\to\widetilde{y}'\in\mathcal{R}:\\ f(\widetilde{y})=f(y), f(\widetilde{y}')=f(y')}} \lambda_{\widetilde{y}\to\widetilde{y}'}(f(y)) > 0$$

- f is *node balanced with respect to* (Λ, μ) if for every $x \in f(C)$

$$\mu(x) \sum_{y\to y'\in\mathcal{R}: f(y)=x} \lambda_{y\to y'}(x) = \sum_{y'\to y\in\mathcal{R}: f(y)=x} \mu(f(y'))\lambda_{y'\to y}(f(y')) \tag{4.1}$$

To understand the meaning of node balancing, let $W_f(\cdot)$ be the continuous-time Markov chain whose transition graph is the one induced f, and the transition rates are given by

$$q(x, x') = \sum_{\substack{y\to y'\in\mathcal{R}:\\ f(y)=x, f(y')=x'}} \lambda_{y\to y'}(x).$$

Then, the following holds.

Proposition 4.2. *Let $(\mathcal{G}, \Lambda)$ be a stochastic reaction system. Let f be a copy of $\mathcal{G}$ and let μ be a σ-finite measure of $\mathbb{Z}^n_{\geq 0}$. Then, there exists a stationary distribution of $W_f(\cdot)$ proportional to the restriction of μ to $f(C)$ if and only if f is node balanced with respect to (Λ, μ).*

Proof. The state space of W_f is $f(C)$, which is finite. Then, there exists a stationary distribution proportional to the restriction of μ to $f(C)$ if and only if the latter is a stationary measure, which is equivalent to (4.1). The proof is then concluded.

Proposition 4.2 implies that if a copy f is node balanced with respect to (Λ, π), then the restriction of π to $f(C)$ is stationary for W_f. If π is a stationary distribution of $X(\cdot)$, then π is stationary for both $X(\cdot)$ and $W_f(\cdot)$. The same holds for union of copies, as expressed in the following corollary.

Corollary 4.3. *Let $(\mathcal{G}, \Lambda)$ be a stochastic reaction system and let μ be a σ-finite measure of $\mathbb{Z}^n_{\geq 0}$. For some $h \geq 1$, let $\Phi = \{f_1, \ldots, f_h\}$ be a set of node balanced copies of $\mathcal{G}$ with respect to (Λ, μ). Then, the continuous-time Markov chain on state space $\mathbb{Z}^n_{\geq 0}$ with transition rates*

$$q(x, x') = \sum_{i=1}^{h} \sum_{\substack{y\to y'\in\mathcal{R}:\\ f_i(y)=x, f_i(y')=x'}} \lambda_{y\to y'}(x), \tag{4.2}$$

has a stationary distribution $\pi = \gamma\mu$ where γ is a positive real constant.

Proof. The proof is concluded by simply noting that, due to Proposition 4.2, for every state x and any $1 \leq i \leq h$ we have

$$\sum_{x'\in\mathbb{Z}^n_{\geq 0}} \mu(x) \sum_{\substack{y\to y'\in\mathcal{R}:\\ f_i(y)=x, f_i(y')=x'}} \lambda_{y\to y'}(x) = \sum_{x'\in\mathbb{Z}^n_{\geq 0}} \mu(x') \sum_{\substack{y\to y'\in\mathcal{R}:\\ f_i(y)=x', f_i(y')=x}} \lambda_{y\to y'}(x').$$

Hence, by summing both sides over i we obtain that μ is a stationary measure, hence proportional to a stationary distribution by classic theory of Markov chains.

Corollary 4.3 implies that the stationary distributions of the full model can be found, up to a normalizing constant, by studying the stationary distributions of the finite continuous time Markov chain with rates 4.2, as long as the copies are node balanced with respect to them. As already mentioned in the introduction, the node balancing of the copies is intimately related to complex balancing, in a way made precise by the results below.

Theorem 4.4. *Let $(\mathcal{G}, \Lambda)$ be a stochastic reaction system and let μ be a measure. Then the following are equivalent:*

(1) every injective copy of $\mathcal{G}$ is node balanced with respect to (Λ, μ);

(2) μ is a complex balanced measure;

(3) every copy of $\mathcal{G}$ is node balanced with respect to (Λ, μ).

Proof. We will prove that (1) implies (2), which implies (3), which implies (1).

(1) $\implies$ (2)**.** Fix $y \in C$ and $x \in \mathbb{Z}^n_{\geq 0}$. If $x \not\geq y$ then $\lambda_{y\to y'}(x) = 0$ for all reactions $y \to y' \in \mathcal{R}$ by (2.3). Moreover, for all reactions $y' \to y \in \mathcal{R}$ we have $x + y' - y \not\geq y'$ which implies $\lambda_{y'\to y}(x + y' - y) = 0$ and again $\lambda_{y\to y'}(x) = 0$ by (2.3). It follows that (3.2) holds. If $x \geq y$ then by Lemma 4.1 there exists an injective copy f of $\mathcal{G}$ with $f(x) = y$, which is node balanced with respect to (Λ, μ). By injectivity, (4.1) becomes (3.2). In conclusion, (3.2) holds for every $y \in C$ and every $x \in \mathbb{Z}^n_{\geq 0}$, and (2) is proven.

(2) $\implies$ (3)**.** Let f be a copy of $\mathcal{G}$ and let $x \in f(C)$. Then, summing both sides of (3.2) over the complexes y with $f(y) = x$ we obtain (4.1), so (3) holds.

(3) $\implies$ (1)**.** This is trivially true as injective copies of $\mathcal{G}$ are copies of $\mathcal{G}$.

With more assumptions on the form of the kinetics, a finite number of node balanced copies with respect to (Λ, μ) are sufficient to imply complex balancing of μ, as detailed below.

Theorem 4.5. *Let $(\mathcal{G}, K^S_{\kappa,\theta})$ be a stochastic reaction system and let μ be a positive stationary measure. Then there exist positive and finite constants M_1, M_2 such that the following are equivalent:*

1) all the injective copies of $\mathcal{G}$ that intersect the cube $[0, M_1]^n$ are node balanced with respect to $(K^S_{\kappa,\theta}, \mu)$;

2) a closed irreducible set Γ containing a state x with $\|x\| \geq M_2$ is such that all copies f of $\mathcal{G}$ with $f(C) \subseteq \Gamma$ are node balanced with respect to $(K^S_{\kappa,\theta}, \mu)$;

3) μ is a complex balanced measure.

The above theorem states that if enough, finitely many copies are node balanced, then the measure μ is complex balanced. The same holds even if all the copies in a single, large enough, closed irreducible set are node balanced. The proof is given in Section 6. In the following, if f is a copy of $\mathcal{G}$ and $v \in \mathbb{Z}^n$, we denote by $f + v$ the function defined by $(f + v)(y) = f(y) + v$ for all $y \in C$. Note that for $v \in \mathbb{Z}^n_{\geq 0}$, $f + v$ is a copy of $\mathcal{G}$ while this may not always hold for general $v \in \mathbb{Z}^n$.

Theorem 4.6. *Let $(\mathcal{G}, K^S_{\kappa,\theta})$ be a stochastic product-form system and let μ be of the form* (3.1)*. Then the following are equivalent:*

(1) there is an active, injective, node balanced copy of $\mathcal{G}$ with respect to $(K^S_{\kappa,\theta}, \mu)$;

(2) μ is a complex balanced measure.

Proof. In what follows we will use that, by substituting the rate functions with (2.7) and μ with (3.1) and by simplifying, we get that for all $y' \to y'' \in \mathcal{R}$ and for all $\tilde{x} \in \mathbb{Z}^n_{\geq 0}$ with $\tilde{x} \geq y'$

$$\mu(\tilde{x})\lambda_{y'\to y''}(\tilde{x}) = c^{\tilde{x}}\kappa_{y'\to y''}\prod_{i=1}^{n}\prod_{j=1}^{x_i-y'_i}\frac{1}{\theta_i(j)}. \tag{4.3}$$

(1) $\Longrightarrow$ (2)**.** Fix $y \in C$ and let x be such that $f(x) = y$. By injectivity and node balancing of f we have

$$\mu(x)\sum_{y'\in C\,:\,y\to y'\in\mathcal{R}}\lambda_{y\to y'}(x) = \sum_{y'\in C\,:\,y'\to y\in\mathcal{R}}\mu(x+y'-y)\lambda_{y'\to y}(x+y'-y),$$

which by (4.3) becomes

$$\sum_{y'\in C\,:\,y\to y'\in\mathcal{R}}\kappa_{y\to y'} = \sum_{y'\in C\,:\,y'\to y\in\mathcal{R}}c^{y'-y}\kappa_{y'\to y}, \tag{4.4}$$

which is in turn equivalent to (3.2) for mass-action kinetics and hence proves that $(\mathcal{G}, K^D_\kappa)$ is complex balanced. Moreover, by using (4.3) again we have that for all $\tilde{x} \in \mathbb{Z}^n_{\geq 0}$ with $\tilde{x} \geq y$

$$\begin{aligned}\sum_{y'\in C\,:\,y'\to y\in\mathcal{R}}\mu(\tilde{x}+y'-y)\lambda_{y'\to y}(\tilde{x}+y'-y) &= \sum_{y'\in C\,:\,y'\to y\in\mathcal{R}}c^{\tilde{x}+y'-y}\kappa_{y'\to y}\prod_{i=1}^{n}\prod_{j=1}^{\tilde{x}_i-y_i}\frac{1}{\theta_i(j)}\\ &= \sum_{y'\in C\,:\,y\to y'\in\mathcal{R}}c^{\tilde{x}}\kappa_{y\to y'}\prod_{i=1}^{n}\prod_{j=1}^{\tilde{x}_i-y_i}\frac{1}{\theta_i(j)}\\ &= \sum_{y'\in C\,:\,y'\to y\in\mathcal{R}}\mu(\tilde{x})\lambda_{y\to y'}(\tilde{x}),\end{aligned}$$

where the second equality follows from (4.3). Thus, (2) is proven.

(2) $\Longrightarrow$ (1)**.** The existence of an injective, active copy follows from Lemma 4.1, and the fact that is node balanced follows from Theorem 4.4.

Our last result only concerns mass-action models.

Theorem 4.7. *Let $(\mathcal{G}, K^S_\kappa)$ be a stochastic mass-action system and let μ be of the form* (3.1)*. Then the following are equivalent:*

(1) there is a copy f of $\mathcal{G}$ such that $f + v$ is node balanced with respect to $(K^S_{\kappa,\theta}, \mu)$ for all $v \in \mathbb{Z}^n_{\geq 0}$;

(2) there is an active copy f of $\mathcal{G}$ such that $f + v$ is node balanced with respect to $(K^S_{\kappa,\theta}, \mu)$ for all $v \in \Xi$ where $\Xi \subseteq \mathbb{Z}^n_{\geq 0}$ is such that the only polynomial of degree at most $d = \max\{\|y\|_1 : y \to y' \in \mathcal{R}\}$ that vanishes on Ξ is the zero polynomial;

(3) μ is a complex balanced measure.

Proof. We will prove that (1) implies (2), which implies (3), which implies (1). Similarly to what done in the proof of Theorem 4.6, we first note that by substituting the rate functions with (2.6) and μ with (3.1) and by simplifying, we get that for all $y' \to y'' \in \mathcal{R}$ and for all $\tilde{x} \in \mathbb{Z}^n_{\geq 0}$ with $\tilde{x} \geq y'$

$$\mu(\tilde{x})\lambda_{y'\to y''}(\tilde{x}) = c^{\tilde{x}}\kappa_{y'\to y''}\frac{1}{(x-y')!}. \tag{4.5}$$

(1) $\Longrightarrow$ (2)**.** This simply follows from the existence of a set Ξ as described. A finite set Ξ can always be constructed, but to prove (2) it is enough to note that Ξ can be chosen as $\mathbb{Z}^n_{\geq 0}$, since the only polynomial vanishing on $\mathbb{Z}^n_{\geq 0}$ is the zero polynomial.

(2) $\Longrightarrow$ (3)**.** Let x be in $f(C)$. By definition of node balancing we have

$$\mu(x+v)\sum_{y\to y'\in\mathcal{R}:\, f(y)=x}\lambda_{y\to y'}(x+v) = \sum_{y'\to y\in\mathcal{R}:\, f(y)=x}\mu(x+v+y'-y)\lambda_{y'\to y}(x+v+y'-y)$$

for all $v \in \Xi$. By multiplying both sides by $(x+v)!$ and by applying (4.5) we get

$$\sum_{y\in C\,:\, f(y)=x}\frac{(x+v)!}{(x+v-y)!}\left(\sum_{y'\in C\,:\, y\to y'\in\mathcal{R}}\kappa_{y\to y'} - \sum_{y'\in C\,:\, y'\to y\in\mathcal{R}} c^{y'-y}\kappa_{y'\to y}\right) = 0.$$

Note that the above is a polynomial of degree at most d in the variable v and it vanishes for all $v \in \Xi$. Hence, it is the zero polynomial and as a consequence $(\mathcal{G}, K^D_\kappa)$ is complex balanced with complex balanced steady state c. Then, by using (4.5) again we have that for all $\tilde{x} \in \mathbb{Z}^n_{\geq 0}$ with $\tilde{x} \geq y$

$$\begin{aligned}\sum_{y'\in C\,:\, y'\to y\in\mathcal{R}}\mu(\tilde{x}+y'-y)\lambda_{y'\to y}(\tilde{x}+y'-y) &= \sum_{y'\in C\,:\, y'\to y\in\mathcal{R}} c^{\tilde{x}+y'-y}\kappa_{y'\to y}\frac{1}{(\tilde{x}-y)!}\\ &= \sum_{y'\in C\,:\, y\to y'\in\mathcal{R}} c^{\tilde{x}}\kappa_{y\to y'}\frac{1}{(\tilde{x}-y)!}\\ &= \sum_{y'\in C\,:\, y'\to y\in\mathcal{R}}\mu(\tilde{x})\lambda_{y\to y'}(\tilde{x}),\end{aligned}$$

where the second equality follows from (4.5). Thus, (2) is proven.

(3) $\Longrightarrow$ (1)**.** This follows from Theorem 4.4.

We give here an example showing that condition 1 alone does not imply μ is complex balanced, unless μ is assumed to be of the form (3.1).

Example 4.1. Consider the stochastic mass-action system

$$0 \xrightarrow{\kappa_1} A\ ,\ 3A \xrightarrow{\kappa_2} 2A,$$

where the reaction rates have been written above the corresponding reactions. Let π be the unique stationary distribution, whose support is in $\{m \in \mathbb{Z} : m \geq 2\}$. Since the model is a birth and death chain, π is detailed balanced: for all $m \geq 2$ we have $\pi(m)q(m, m+1) = \pi(m+1)q(m+1, m)$ [33]. Specifically, we have

$$\pi(m)\lambda_{0\to A}(m) = \pi(m+1)\lambda_{3A\to 2A}(m+1) \tag{4.6}$$

The deterministically modeled system cannot be complex balanced by Theorem 3.1 because it is not weakly reversible. By [21, Corollary 19], π cannot be complex balanced. However, it follows from (4.6) that the copy defined by $f(0) = f(2A) = 2$ and $f(A) = f(3A) = 3$ is node balanced with respect to (K_κ^S, π), and the same holds for all the copies $f + v$ with $v \in \mathbb{Z}_{\geq 0}$.

5. Discussion

We end by recapitulating the consequences of our results. The truncation of the state space, as already discussed in the Introduction, is a widely used tool to approximate the stationary distribution of stochastic models. A truncation is specified in two steps: the first is a description of the restricted state space and the second is the definition of the constrained process. In particular, a prescription for the allowed transitions of the process on the boundary of the truncated space needs to be given. As an example, in [9] the outgoing transitions from the truncated state space are redirected to a designated state within the truncated state space. Many other strategies have been studied, as discussed in the review [10]. While truncation methods normally give an approximation to the stationary distribution of the original process, our interest lies in identifying the situations where the approximation is exact, except for a normalizing constant.

In the context of this paper, we consider as a truncation a finite set of copies $\{f_1, \ldots, f_h\}$, as described in (4.2). By Corollary 4.3, if the copies $f_i : 1 \leq i \leq h$ are node balanced with respect to the kinetics and a measure μ, then μ is a stationary distribution of the truncated process, up to normalization. We are interested in understanding whether μ is also equal to a stationary distribution of the original process, up to normalization. A consequence of Theorem 4.5 is that this happens if the truncated state space is large enough and the kinetics have product form, a generalization of mass-action kinetics. Unfortunately, in this case and still by the same theorem, we have that μ is a complex balanced measure. Hence, we knew already that μ is proportional to (3.1) by Theorems 3.2 and 3.4, so the application of truncation methods is not useful. In future works, we will identify exact approximation via different truncation methods related to copies and avail the exact stationary distributions of models that are not complex balanced.

The second important consequence of our results, mainly from a theoretical point of view, is that we have a further characterization of complex balancing given by Theorems 4.4–4.7.

6. Proofs

6.1. Deficiency theory

The main idea behind the proofs in this section is based upon classical notions of deficiency theory which we will briefly introduce here. We start with giving the definition of deficiency, first given in [13].

Definition 6.1. The *deficiency* of a reaction network $\mathcal{G} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ is the number $\delta = m - \ell - \dim S$, where ℓ is the number of connected components of the directed graph $(\mathcal{C}, \mathcal{R})$ and S is the stoichiometric subspace of $\mathcal{G}$.

The following geometric interpretation will be used in our proofs: let $\{e_y\}_{y\in\mathcal{C}}$ be the canonical basis of $\mathbb{R}^m$, where each coordinate is univocally associated with a complex. Further, define $d_{y\to y'} = e_{y'} - e_y$ for $y \to y' \in \mathcal{R}$. Let $D = \operatorname{span}(d_{y\to y'} : y \to y' \in \mathcal{R})$. It is proven in [12, 13] that $\dim D = m - \ell$. Then, the space D is linearly isomorphic to S if and only if $\delta = 0$. Specifically, consider the homomorphism

$$\begin{aligned} \varphi\colon \quad \mathbb{R}^m &\to \mathbb{R}^n \\ e_y &\mapsto y. \end{aligned} \tag{6.1}$$

For $y \to y' \in \mathcal{R}$, we have $\varphi(d_{y\to y'}) = y' - y$ and $\varphi_{|D}\colon D \to S$ is thus a surjective homomorphism. Therefore,

$$\dim \operatorname{Ker} \varphi_{|D} = \dim D - s = m - \ell - s = \delta, \tag{6.2}$$

which implies that $\varphi_{|D}$ is an isomorphism if and only if $\delta = 0$. It further follows that the deficiency is a non-negative number.

The following classical result is proven in [12, 13].

Theorem 6.1. *Let $(\mathcal{G}, K_\kappa^D)$ be a deterministic mass-action system. If $\mathcal{G}$ is weakly reversible and its deficiency is zero, then $(\mathcal{G}, K_\kappa^D)$ is complex balanced for any choice of rate constants.*

6.2. A new model

Consider a stochastic reaction system $(\mathcal{G}, K^S_{\kappa,\theta})$ with $\mathcal{G} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ and $K^S_{\kappa,\theta}$ as in (2.7). We will follow an idea first proposed in [21], and consider the reaction network $\tilde{G} = (\tilde{\mathcal{S}}, \tilde{\mathcal{C}}, \tilde{\mathcal{R}})$ with

$$\tilde{\mathcal{S}} = \mathcal{S} \cup \{A_y : y \in \mathcal{C}\}, \quad \tilde{\mathcal{C}} = \{y + A_y : y \in \mathcal{C}\}, \quad \text{and} \quad \tilde{\mathcal{R}} = \{y + A_y \to y' + A_{y'} : y \to y' \in \mathcal{R}\}. \tag{6.3}$$

It follows from the construction above that $\tilde{\mathcal{G}}$ is weakly reversible if and only if $\mathcal{G}$ is weakly reversible. Moreover, the following holds.

Lemma 6.2. *The deficiency of $\tilde{\mathcal{G}}$ as defined in* (6.3) *is zero.*

Proof. Consider the linear homomorphism φ as defined in (6.1), for the reaction network $\tilde{G}$. Since the vectors associated with the complexes $y + A_y$ are linear independent, φ is an isomorphism and as a consequence the deficiency of $\tilde{G}$ is 0.

We associate $\tilde{G}$ with a general mass-action kinetics $K^S_{\kappa,\tilde{\theta}}$ such that the rate of any reaction $y + A_y \to y' + A_{y'} \in \tilde{\mathcal{R}}$ is of the form

$$K^S_{\kappa,\tilde{\theta}}(y + A_y \to y' + A_{y'})(x, u) = K^S_{\kappa,\theta}(y \to y')(x) u_{A_y} \quad \text{for all } (x, u) \in \mathbb{Z}^n_{\geq 0} \times \mathbb{Z}^m_{\geq 0}.$$

The following is the key result of this paper, on which the proof of the other main theorems are based on.

Theorem 6.3. *Let $(\mathcal{G}, K^S_{\kappa,\theta})$ be a stochastic reaction system with $K^S_{\kappa,\theta}$ as in* (2.7)*. Let μ be a σ-finite measure on $\mathbb{Z}^n_{\geq 0}$. Define Υ as the set of pairs $(x, y) \in \mathbb{Z}^n_{\geq 0} \times \mathcal{C}$ such that there exists an active, injective copy f of $\mathcal{G}$ that is node balanced with respect to $(K^S_{\kappa,\theta}, \mu)$, satisfies $f(y) = x$, and fulfils $f(\mathcal{C}) \subseteq \operatorname{supp} \mu$. Let $\hat{\mathcal{R}}$ be a subset of reactions whose reaction vectors form a basis of S. Assume that*

1) there exists $\hat{x} \in \mathbb{Z}^n_{\geq 0}$ *such that* $(\hat{x}, y) \in \Upsilon$ *for all* $y \in C$*;*

2) for all $y^\star \to y^{\star\star} \in \mathcal{R} \setminus \hat{\mathcal{R}}$*, there exist a sequence of (potentially repeated) reactions* $\{y_i \to y_i'\}_{i=1}^h$ *contained in* $\{y^\star \to y^{\star\star}\} \cup \hat{\mathcal{R}}$ *such that*

(a) $y^\star \to y^{\star\star} = y_1 \to y_1'$*;*

(b) $\sum_{i=1}^h (y_i' - y_i) = 0$*;*

(c) for all $j = 0, 1, 2, \ldots, h-1$ *we have*

$$\left(\hat{x} + \sum_{i=1}^{j} (y_i' - y_i), y_{j+1}\right) \in \Upsilon.$$

Then, the deterministic mass-action system $(\mathcal{G}, K_\kappa^D)$ *is complex balanced.*

Proof. Consider $(x, e_y) \in \mathbb{Z}^n_{\geq 0} \times \mathbb{Z}^m_{\geq 0}$. If $x \geq y$, then any reaction of the form $y + A_y \to y' + A_{y'}$ can take place and move the process from (x, e_y) to $(x + y' - y, e_{y'})$. From $x \geq y$ it follows that $x + y' - y \geq y'$ hence at $(x + y' - y, e_{y'})$ any reaction of the form $y' + A_{y'} \to y'' + A_{y''}$ can take place and move the process to $(x + y'' - y, e_{y''})$, and so on. Since $\mathcal{G}$ is weakly reversible it is possible to eventually return to the original state (x, e_y), and the closed irreducible set containing (x, e_y) is precisely given by $\{(x + \tilde{y} - y, e_{\tilde{y}}) \,:\, y \in C_y\}$, where C_y denotes the connected component of $(C, \mathcal{R})$ containing y. We denote the closed irreducible set containing (x, e_y) by $\Gamma_{(x,e_y)}$. Of course the notation is not bijective: for example if $x \geq y$ then $\Gamma_{(x,e_y)} = \Gamma_{(x+y'-y,e_{y'})}$. If $(x, y) \in \Upsilon$, then be definition there exists an active, injective copy f of $\mathcal{G}$ that is node balanced with respect to $(K^S_{\kappa,\theta}, \mu)$ and that satisfies $f(y) = x$. Hence, by definition of node balancing, for all $\tilde{y} \in C_y$ we have

$$\sum_{y' \in C : \tilde{y} \to y' \in \mathcal{R}} \mu(x + \tilde{y} - y) \lambda_{\tilde{y} \to y'}(x + \tilde{y} - y) = \sum_{y' \in C : y' \to \tilde{y} \in \mathcal{R}} \mu(x + y' - y) \lambda_{y' \to \tilde{y}}(x + y' - y),$$

which implies

$$\sum_{y' \in C : \tilde{y} + A_{\tilde{y}} \to y' + A_{y'} \in \tilde{\mathcal{R}}} \mu(x + \tilde{y} - y) \lambda_{\tilde{y} + A_{\tilde{y}} \to y' + A_{y'}}(x + \tilde{y} - y, e_{\tilde{y}})$$
$$= \sum_{y' \in C : y' + A_{y'} \to \tilde{y} + A_{\tilde{y}} \in \tilde{\mathcal{R}}} \mu(x + y' - y) \lambda_{y' + A_{y'} \to \tilde{y} + A_{\tilde{y}}}(x + y' - y, e_{y'}).$$

Since $\Gamma_{(x,e_y)}$ is finite, it follows from the equation above that μ restricted to $\Gamma_{(x,e_y)}$ is proportional to the unique stationary distribution of $(\tilde{\mathcal{G}}, K^S_{\kappa,\tilde{\theta}})$ with support $\Gamma_{(x,e_y)}$. Since $\tilde{\mathcal{G}}$ has deficiency zero by Lemma 6.2 and is weakly reversible because $\mathcal{G}$ is weakly reversible, by Theorems 6.1 and 3.1 we have

$$\mu(x + \tilde{y} - y) = M_{(x,e_y)} \tilde{c}_{A_y} c^x \prod_{j=1}^{x_i} \frac{1}{\theta_i(j)}, \tag{6.4}$$

where $(c, \tilde{c}) \in \mathbb{R}^n_{>0} \times \mathbb{R}^m_{>0}$ is a complex balanced steady state of $(\tilde{\mathcal{G}}, K_\kappa^D)$ and $M_{(x,e_y)}$ is a proportionality constant depending on the closed irreducible set $\Gamma_{(x,e_y)}$, hence the notation is not bijective and $M_{(x,e_y)} = M_{(x+\tilde{y}-y,e_{\tilde{y}})}$ for all $\tilde{y} \in C_y$.

We now want to show that there exists a vector $\hat{c} \in \mathbb{R}^n_{>0}$ such that

$$\frac{\tilde{c}_{A_{y'}}}{\tilde{c}_{A_y}} = \hat{c}^{y'-y} \quad \text{for all } y \to y' \in \mathcal{R}. \tag{6.5}$$

First, note that (6.5) holds for all $y \to y' \in \hat{\mathcal{R}}$, because it is equivalent to

$$(y'-y)^\top \log \hat{c} = \log \tilde{c}_{A_{y'}} - \log \tilde{c}_{A_y} \quad \text{for all } y \to y' \in \hat{\mathcal{R}}.$$

The latter has a solution because the reaction vectors of the reactions in $\hat{R}$ are linearly independent. If $y \to y' \in \mathcal{R} \setminus \hat{\mathcal{R}}$ then by hypothesis there exists a sequence of reactions $\{y_i \to y_i'\}_{i=1}^h$ contained in $\{y \to y'\} \cup \hat{\mathcal{R}}$ such that $y_1 \to y_1' = y \to y'$ and

$$\left(\hat{x} + \sum_{i=1}^{j}(y_i' - y_i), y_{j+1}\right) \in \Upsilon.$$

for all $j = 0, 1, 2, \ldots, h-1$. Hence, it follows from $y' - y + \sum_{i=1}^h (y_i' - y_i) = 0$ and from applying (6.4) recursively that

$$\begin{aligned}
M_{(\hat{x},y_1)} &= M_{(\hat{x}+y'-y+\sum_{i=1}^h (y_i'-y_i),y_1)} \\
&= \frac{\tilde{c}_{A_{y_{h-1}'}}}{\tilde{c}_{A_{y_1}}} M_{(\hat{x}+y'-y+\sum_{i=1}^h (y_i'-y_i),y_{h-1}')} = \frac{\tilde{c}_{A_{y_{h-1}'}}}{\tilde{c}_{A_{y_1}}} M_{(\hat{x}+y'-y+\sum_{i=1}^{h-1} (y_i'-y_i),y_{h-1})} \\
&= \frac{\tilde{c}_{A_{y_{h-1}'}} \tilde{c}_{A_{y_{h-2}'}}}{\tilde{c}_{A_{y_1}} \tilde{c}_{A_{y_{h-1}}}} M_{(\hat{x}+y'-y+\sum_{i=1}^{h-1} (y_i'-y_i),y_{h-2}')} = \frac{\tilde{c}_{A_{y_{h-1}'}} \tilde{c}_{A_{y_{h-2}'}}}{\tilde{c}_{A_{y_1}} \tilde{c}_{A_{y_{h-1}}}} M_{(\hat{x}+y'-y+\sum_{i=1}^{h-2} (y_i'-y_i),y_{h-2})} \\
&= \cdots = M_{(\hat{x},y_1)} \prod_{i=1}^{h} \frac{\tilde{c}_{A_{y_i'}}}{\tilde{c}_{A_{y_i}}}.
\end{aligned}$$

As a consequence,

$$\prod_{i=1}^{h} \frac{\tilde{c}_{A_{y_i'}}}{\tilde{c}_{A_{y_i}}} = 1.$$

If $y \to y'$ appears α times in the sequence $\{y_i \to y_i'\}_{i=1}^h$, and all other reactions $\tilde{y} \to \tilde{y}'$ appear $\beta_{\tilde{y}\to\tilde{y}'}$ times, then

$$\frac{\tilde{c}_{A_{y'}}}{\tilde{c}_{A_y}} = \prod_{\tilde{y}\to\tilde{y}' \in \hat{\mathcal{R}}} \left(\frac{\tilde{c}_{A_{\tilde{y}}}}{\tilde{c}_{A_{\tilde{y}'}}}\right)^{\frac{\beta_{\tilde{y}\to\tilde{y}'}}{\alpha}} = \hat{c}^{-\frac{1}{\alpha}\sum_{\tilde{y}\to\tilde{y}'\in\hat{\mathcal{R}}} \beta_{\tilde{y}\to\tilde{y}'}(y'-y)} = \hat{c}^{y'-y}.$$

Hence, (6.5) is proven. For all $y \in C$ we have $(\hat{x}, y) \in \Upsilon$, which implies

$$\sum_{y' \in C : y \to y' \in \mathcal{R}} \mu(\hat{x}) \lambda_{y\to y'}(\hat{x}) = \sum_{y' \in C : y' \to y \in \mathcal{R}} \mu(x + y' - y) \lambda_{y'\to\tilde{y}}(x + y' - y).$$

By substituting the expressions for the transition rates and μ (over the image of node balanced copies) and by simplifying we obtain

$$\sum_{y' \in C : y \to y' \in \mathcal{R}} c^x \kappa_{y\to y'} = \sum_{y' \in C : y' \to y \in \mathcal{R}} \frac{\tilde{c}_{A_{y'}}}{\tilde{c}_{A_y}} c^{x+y'-y} \kappa_{y'\to y}$$

which by (6.5) implies

$$\sum_{y' \in C : y \to y' \in \mathcal{R}} \kappa_{y \to y'} \prod_{i=1}^{n} (c_i \hat{c}_i)^{y_i} = \sum_{y' \in C : y' \to y \in \mathcal{R}} \kappa_{y' \to y} \prod_{i=1}^{n} (c_i \hat{c}_i)^{y'_i},$$

hence a positive complex balanced steady state exists for $(\mathcal{G}, K^D_\kappa)$, which concludes the proof.

6.3. Proof of Theorem 3.4

By Theorem 3.2 we already know that if $(\mathcal{G}, K^D_\kappa)$ is complex balanced, then there exists a σ-finite, positive, stationary measure for $(\mathcal{G}, K^S_{\kappa,\theta})$ given by μ as in (3.1). Checking that μ is complex balanced is not difficult, since by substituting the reaction rates with (2.7) and μ with (3.1), Eq (3.2) simplifies to

$$\sum_{y' \in C \,:\, y \to y' \in \mathcal{R}} \kappa_{y \to y'} = \sum_{y' \in C \,:\, y' \to y \in \mathcal{R}} c^{y'-y} \kappa_{y' \to y},$$

which holds because $(\mathcal{G}, K^D_\kappa)$ is complex balanced.

Conversely, assume a σ-finite, positive, complex balanced measure for $(\mathcal{G}, K^S_{\kappa,\theta})$ exists. Then, all copies of $\mathcal{G}$ are node balanced with respect to $(\mathcal{G}, K^S_{\kappa,\theta})$ by Theorem 4.4. Moreover, since μ is positive then for all copies f of $\mathcal{G}$ we have $f(C) \subseteq \operatorname{supp} \mu$. Hence, the proof is concluded by Theorem 6.3 by choosing $\hat{x} \in \mathbb{Z}^n_{\geq 0}$ large enough.

6.4. Proof of Theorem 4.5

It follows from Theorem 4.4 that (3) implies (1) for any M_1. Moreover, the existence of a positive complex balanced measure implies that $\mathcal{G}$ is weakly reversible [23, Theorem 4.4] hence $\mathbb{Z}^n_{\geq 0}$ is union of closed irreducible sets [36]. As a consequence, the existence of a closed irreducible set as in (2) is implied by Theorem 4.4 for any M_2.

Conversely, since μ is positive, it is always possible to choose M_1 large enough such that all injective copies intersecting $[0, M_1]^n$ being node balanced with respect to $(K^S_{\kappa,\theta})$ implies that the assumptions of Theorem 6.3 are satisfied. Hence, for large enough M_1 (1) implies that $(\mathcal{G}, K^D_\kappa)$ is complex balanced. In turn, this implies that μ is proportional to (3.1) on every closed irreducible set because the model is non-explosive by Theorem 3.3 (hence all stationary measures are proportional to each other on every closed irreducible set by standard Markov chain theory [33]). By substituting the reaction rates with (2.7) and μ as proportional to (3.1), Eq (3.2) simplifies to

$$\sum_{y' \in C \,:\, y \to y' \in \mathcal{R}} \kappa_{y \to y'} = \sum_{y' \in C \,:\, y' \to y \in \mathcal{R}} c^{y'-y} \kappa_{y' \to y},$$

which holds because $(\mathcal{G}, K^D_\kappa)$ is complex balanced.

Similarly, since μ is positive, all injective copies in a closed irreducible set with large enough states being node balanced with respect to $(K^S_{\kappa,\theta})$ implies that the assumptions of Theorem 6.3 are satisfied. By following the same reasoning as above, this implies (3) and concludes the proof.

Acknowledgements

DC was supported by the MIUR grant 'Dipartimenti di Eccellenza 2018-2022' (E11G18000350001). The authors acknowledge support from Politecnico di Torino through the Open Access initiative.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. D. F. Anderson, T. G. Kurtz, Continuous time markov chain models for chemical reaction networks, in *Design and analysis of biomolecular circuits*, Springer, (2011), 3–42. https://doi.org/10.1007/978-1-4419-6766-4_1
2. D. Schnoerr, G. Sanguinetti, R. Grima, Approximation and inference methods for stochastic biochemical kineticsa tutorial review, *J. Phys. A: Math. Theor.*, **50** (2017), 093001. https://doi.org/10.1088/1751-8121/aa54d9
3. S. Aoki, G. Lillacci, A. Gupta, A. Baumschlager, D. Schweingruber, M. Khammash, A universal biomolecular integral feedback controller for robust perfect adaptation, *Nature*, **570** (2019), 533–537. https://doi.org/10.1038/s41586-019-1321-1
4. T. Plesa, G. Stan, T. Ouldridge, W. Bae, Quasi-robust control of biochemical reaction networks via stochastic morphing, *J. R. Soc. Interface*, **18** (2021), 20200985. https://doi.org/10.1098/rsif.2020.0985
5. J. Kim, G. Enciso, Absolutely robust controllers for chemical reaction networks, *J. R. Soc. Interface*, **17** (2020), 20200031. https://doi.org/10.1098/rsif.2020.0031
6. H. Kang, T. Kurtz, Separation of time-scales and model reduction for stochastic reaction networks, *Ann. Appl. Probab.*, **23** (2013), 529–583. https://doi.org/10.1214/12-AAP841
7. B. Brook, S. Waters, Mathematical challenges in integrative physiology, *J. Math. Biol.*, **56** (2008), 893–896.
8. L. Preziosi, Hybrid and multiscale modelling, *J. Math. Biol.*, **53** (2006), 977–978. https://doi.org/10.1007/s00285-006-0042-x
9. A. Gupta, J. Mikelson, M. Khammash, A finite state projection algorithm for the stationary solution of the chemical master equation, *J. Chem. Phys.*, **147** (2017), 154101. https://doi.org/10.1063/1.5006484
10. J. Kuntz, P. Thomas, G. Stan, M. Barahona, Stationary distributions of continuous-time Markov chains: a review of theory and truncation-based approximations, *SIAM Rev.*, **63** (2021), 3–64. https://doi.org/10.1137/19M1289625
11. D. Anderson, G. Craciun, T. Kurtz, Product-form stationary distributions for deficiency zero chemical reaction networks, *Bull. Math. Biol.*, **72** (2010), 1947–1970. https://doi.org/10.1007/s11538-010-9517-4
12. F. Horn, R. Jackson, General mass action kinetics, *Arch. Ration. Mech. Anal.*, **47** (1972), 81–116.
13. M. Feinberg, Complex balancing in general kinetic systems, *Arch. Ration. Mech. Anal.*, **49** (1972), 187–194. https://doi.org/10.1007/BF00255665
14. G. Craciun, A. Dickenstein, A. Shiu, B. Sturmfels, Toric dynamical systems, *J. Symb. Comput.*, **44** (2009), 1551–1565. https://doi.org/10.1016/j.jsc.2008.08.006

15. M. Gopalkrishnan, E. Miller, A. Shiu, A geometric approach to the global attractor conjecture, *SIAM J. Appl. Dyn.*, **13** (2014), 758–797. https://doi.org/10.1137/130928170

16. B. Joshi, A. Shiu, Atoms of multistationarity in chemical reaction networks, *J. Math. Chem.*, **51** (2013), 153–178. https://doi.org/10.1007/s10910-012-0072-0

17. B. Joshi, G. Craciun, Reaction network motifs for static and dynamic absolute concentration robustness, preprint, arXiv:2201.08428.

18. M. Pérez Millán, A. Dickenstein, The structure of MESSI biological systems, *SIAM J. Appl. Dyn. Syst.*, **17** (2018), 1650–1682. https://doi.org/10.1137/17M1113722

19. D. Cappelletti, A. Gupta, M. Khammash, A hidden integral structure endows absolute concentration robust systems with resilience to dynamical concentration disturbances, *J. R. Soc. Interface*, **17** (2020), 20200437. https://doi.org/10.1098/rsif.2020.0437

20. G. Craciun, B. Joshi, C. Pantea, I. Tan, Multistationarity in cyclic sequestration-transmutation networks, *Bull. Math. Biol.*, **84** (2022). https://doi.org/10.1007/s11538-022-01021-7

21. D. Cappelletti, C. Wiuf, Product-form Poisson-like distributions and complex balanced reaction systems, *SIAM J. Appl. Math.*, **76** (2016), 411–432. https://doi.org/10.1137/15M1029916

22. B. Joshi, A detailed balanced reaction network is sufficient but not necessary for its Markov chain to be detailed balanced, *Discrete Contin. Dyn. Syst. Ser. B*, **20** (2015), 1077–1105. https://doi.org/10.3934/dcdsb.2015.20.1077

23. D. Cappelletti, B. Joshi, Graphically balanced equilibria and stationary measures of reaction networks, *SIAM J. Appl. Dyn. Syst.*, **17** (2018), 2146–2175. https://doi.org/10.1137/17M1153315

24. L. Hoessly, C. Mazza, Stationary distributions and condensation in autocatalytic reaction networks, *SIAM J. Appl. Math.*, **79** (2019), 1173–1196. https://doi.org/10.1137/18M1220340

25. D. F. Anderson, S. L. Cotter, Product-form stationary distributions for deficiency zero networks with non-mass action kinetics, *Bull. Math. Biol.*, **78** (2016), 2390–2407. https://doi.org/10.1007/s11538-016-0220-y

26. D. Anderson, T. Nguyen, Results on stochastic reaction networks with non-mass action kinetics, *Math. Biosci. Eng.*, **16** (2019), 2118–2140. https://www.aimspress.com/article/doi/10.3934/mbe.2019103

27. F. Kelly, *Reversibility and Stochastic Networks*, Wiley, Chichester, 1979.

28. P. Whittle, *Systems in Stochastic Equilibrium*, John Wiley & Sons, Inc., 1986.

29. T. Kurtz, Limit theorems and diffusion approximations for density dependent markov chains, in *Stochastic Systems: Modeling, Identification and Optimization, I*, Springer, 1976, 67–78.

30. T. Kurtz, Strong approximation theorems for density dependent Markov chains, *Stoch. Proc. Appl.*, **6** (1978), 223–240. https://doi.org/10.1016/0304-4149(78)90020-0

31. S. Leite, R. Williams, A constrained Langevin approximation for chemical reaction network, *Ann. Appl. Probab.*, **29** (2019), 1541–1608.

32. A. Angius, G. Balbo, M. Beccuti, E. Bibbona, A. Horvath, R. Sirovich, Approximate analysis of biological systems by hybrid switching jump diffusion, *Theor. Comput. Sci.*, **587** (2015), 49–72. https://doi.org/10.1016/j.tcs.2015.03.015

33. J. Norris, *Markov Chains*, Cambridge university press, 1998.

34. D. Anderson, D. Cappelletti, M. Koyama, T. Kurtz, Non-explosivity of stochastically modeled reaction networks that are complex balanced, *Bull. Math. Biol.*, **80** (2018), 2561–2579. https://doi.org/10.1007/s11538-018-0473-8

35. G. Craciun, Toric differential inclusions and a proof of the global attractor conjecture, preprint, arXiv:1501.02860.

36. L. Paulevé, G. Craciun, H. Koeppl, Dynamical properties of discrete reaction networks, *J. Math. Biol.*, **69** (2014), 55–72. https://doi.org/10.1007/s00285-013-0686-2

AIMS Press