

## Existence of a unique quasi-stationary distribution in stochastic reaction networks

Mads Christian Hansen\*      Carsten Wiuf\*

### Abstract

In the setting of stochastic dynamical systems that eventually go extinct, the quasi-stationary distributions are useful to understand the long-term behavior of a system before evanescence. For a broad class of applicable continuous-time Markov processes on countably infinite state spaces, known as reaction networks, we introduce the inferred notion of absorbing and endorsed sets, and obtain sufficient conditions for the existence and uniqueness of a quasi-stationary distribution within each such endorsed set. In particular, we obtain sufficient conditions for the existence of a globally attracting quasi-stationary distribution in the space of probability measures on the set of endorsed states. Furthermore, under these conditions, the convergence from any initial distribution to the quasi-stationary distribution is exponential in the total variation norm.

**Keywords:** quasi-stationary distribution; reaction network; continuous time Markov process.

**AMS MSC 2010:** 60J27; 60J28; 60B10; 80A30; 92C42; 92E20.

Submitted to EJP on April 9, 2019, final version accepted on March 9, 2020.

Supersedes arXiv:1808.06833.

## 1 Introduction

We may think of reaction networks in generality as a natural framework for representing systems of transformational interactions of entities [48]. The set of entities (species) may in principle be of any nature, and specifying not just which ones interact (stoichiometry and reactions) but also quantifying how frequent they interact (kinetics), we obtain the dynamical system of a reaction network. Examples abound in biochemistry, where the language originated, however the true power of this approach is the ability to model diverse processes such as found in biological [5, 7], medical [4], social [15], computational [13], economical [49], ecological [43] or epidemiological [36] contexts.

Lack of complete information in complex systems often introduces, at least for modeling purposes, some degree of stochasticity. A stochastic description can thus be

---

\*University of Copenhagen, Denmark.

E-mail: madschansen@gmail.com, wiuf@math.ku.dk

seen not as an alternative to the deterministic approach, but as a more complete one [41]. Indeed, the deterministic model solution is an approximation of the solution for the stochastic model, improving with the system size, and in general only remaining valid on finite time intervals [28]. Thus, the long-term behavior of a given reaction network may depend crucially on whether it is modeled deterministically or stochastically [23]. In particular, the possibility of extinction, which is a widely occurring phenomenon in nature, may sometimes only be captured by the latter [26]. As a consequence, the counterpart to a stable stationary solution in the deterministically modeled system is not generally a stationary distribution of the corresponding stochastic model. Instead, a so-called quasi-stationary distribution, which is a stationary measure when conditioned on the process not going extinct, has shown to be the natural object of study. A concise overview of the history and current state of this field can be found in [46], while [40] contains a comprehensive bibliography on quasi-stationary distributions and related work.

From a modeling standpoint, when the copy-numbers of interacting entities are low and reaction rates are slow, it is important to recognize that the individual reaction steps occur discretely and are separated by time intervals of random length [5]. This is for example the case at the cellular level [17], where stochastic effects resulting from these small numbers may be physiologically significant [13]. Furthermore, stochastic variations inherent to the system may in general be beneficial for identifying system parameters [35]. If the system under study has been running for a long time, and if the only available knowledge about the system is that it has not reached extinction, then we can often interpret the quasi-stationary distribution, if it exists and is unique, as the likely distribution of the state variable [36].

Consider a right-continuous time-homogenous Markov process  $(X_t : t \geq 0)$  [42], that evolves in a domain  $D \subseteq \mathbb{R}^d$ , wherein there is a set of absorbing states, a “trap”,  $A \subset D$ . The process is absorbed, also referred to as being killed, when it hits the set of absorbing states, implying  $X_t \in A$  for all  $t \geq \tau_A$ , where  $\tau_A = \inf\{t \geq 0 : X_t \in A\}$  is the hitting time of  $A$ . As we are interested in the process before reaching  $A$ , there is no loss of generality in assuming  $X_t = X_{t \wedge \tau_A}$ . We refer to the complement,

$$E := D \setminus A,$$

as the set of endorsed states. For any probability distribution,  $\mu$ , on  $E$ , we let  $\mathbb{P}_\mu$  and  $\mathbb{E}_\mu$  be the probability and expectation respectively, associated with the process  $(X_t : t \geq 0)$ , initially distributed with respect to  $\mu$ . For any  $x \in E$ , we let  $\mathbb{P}_x = \mathbb{P}_{\delta_x}$  and  $\mathbb{E}_x = \mathbb{E}_{\delta_x}$ . Under suitable conditions, the process hits the absorbing set almost surely (a.s.), that is  $\mathbb{P}_x(\tau_A < \infty) = 1$  for all  $x \in E$ , and we investigate the behavior of the process before being absorbed [12].

**Definition 1.1.** A probability measure  $\nu$  on  $E$  is called a quasi-stationary distribution (QSD) for the process  $(X_t : t \geq 0)$  absorbed at  $A$ , if for every measurable set  $B \subseteq E$

$$\mathbb{P}_\nu(X_t \in B \mid t < \tau_A) = \nu(B), \quad t \geq 0,$$

or equivalently, if there exists a probability measure  $\mu$  on  $E$  such that

$$\lim_{t \rightarrow \infty} \mathbb{P}_\mu(X_t \in B \mid t < \tau_A) = \nu(B),$$

in which case we also say that  $\nu$  is a quasi-limiting distribution.

We refer to [30] for a proof of the equivalence of quasi-limiting and quasi-stationary distributions. Existence and uniqueness of a QSD on a finite state space is well known [12, Chapter 3], and it is given by the normalized left Perron-Frobenius eigenvector of

the transition rates matrix restricted to  $E$ . For the infinite dimensional case, most work has been carried out for birth-death processes in one dimension [46], where classification results yielding information about the set of QSDs exist [44].

In the present paper, we will focus on a special case of multidimensional processes on countable infinite state spaces which can be viewed as reaction networks. We will prove as the main result in Theorem 5.1 and Corollary 5.2 sufficient conditions for the existence of a unique globally attracting QSD in the space of probability distributions on  $E$ , equipped with the total variation norm,  $\|\cdot\|_{TV}$ . Recall that this norm may be defined as [39]

$$\|\mu\|_{TV} = 2 \sup_{B \subseteq E} |\mu(B)|.$$

Thus, informally, the metric associated to this norm is the largest possible difference between the probabilities that two probability distributions can assign to the same event. Our result and approach is based on [9] which in turn relies on the following recent main discovery [8, Theorem 2.1].

**Theorem 1.2.** *The following are equivalent*

- *There exists a probability measure  $\nu$  on  $E$  and two constants  $C, \gamma > 0$  such that, for all initial distributions  $\mu$  on  $E$ ,*

$$\|\mathbb{P}_\mu(X_t \in \cdot \mid t < \tau_A) - \nu(\cdot)\|_{TV} \leq Ce^{-\gamma t}, \quad \forall t \geq 0.$$

- *There exists a probability measure  $\nu$  on  $E$  such that*

(A1) *there exists  $t_0, c_1 > 0$  such that for all  $x \in E$ ,*

$$\mathbb{P}_x(X_{t_0} \in \cdot \mid t_0 < \tau_A) \geq c_1 \nu(\cdot),$$

(A2) *there exists  $c_2 > 0$  such that for all  $x \in E$  and  $t \geq 0$ ,*

$$\mathbb{P}_\nu(t < \tau_A) \geq c_2 \mathbb{P}_x(t < \tau_A).$$

More precisely, using Foster-Lyapunov theory [31, 32], a series of assumptions on the process  $(X_t: t \geq 0)$  was shown in [9, Cor. 2.8] to be sufficient for (A1) and (A2) to hold. This result was then applied to a particular case of multidimensional birth-death processes, giving sufficient conditions, in terms of the parameters of the process, for the existence and uniqueness of a QSD [9, Theorem 3.1]. Here, we extend this latter part of the approach of [9], not just to a larger set of parameter values in the birth-death process case, but to the *much* broader class of stochastic processes known as stochastic reaction networks.

The outset of the paper is as follows. In Section 2, we introduce the setup and notation of reaction network theory, and define the central inferred notions of endorsed and absorbing states for this class of processes. Section 3 contains the terminology and main assumptions that we shall use throughout the paper. We then move on in Section 4, to prove that the processes associated with stochastic reaction networks do indeed satisfy all the required assumptions made by [9, Corollary 2.8]. Section 5 contains the main result, Theorem 5.1. Finally, we give some examples in Section 6, illustrating the applicability of the results.

## 2 Reaction network setup

Denote the real numbers by  $\mathbb{R}$ , the integers by  $\mathbb{Z}$ , the natural numbers by  $\mathbb{N} = \{1, 2, \dots\}$  and the nonnegative integers by  $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ . Further, for any set,  $B$ , let

$|B|$  denote its cardinality and denote by  $\mathbb{1}_B: D \rightarrow \{0, 1\}$  the indicator function of a subset  $B \subseteq D$ .

A reaction network is a triple  $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ , where  $\mathcal{S}$  is a finite ordered set of species<sup>1</sup>,  $\mathcal{C}$  is a finite set of complexes, consisting of linear combinations over  $\mathbb{N}_0$  of the species, and  $\mathcal{R} \subset \mathcal{C} \times \mathcal{C}$  is an irreflexive relation on  $\mathcal{C}$ , referred to as the set of reactions [3, 19, 22].

We define the dimension of the reaction network,  $d = |\mathcal{S}|$ . Any species  $S_i \in \mathcal{S}$  can be identified with the unit vector  $e_i \in \mathbb{N}_0^d$ , thus any complex  $y \in \mathcal{C}$  can be identified with a vector in  $\mathbb{N}_0^d$ . It is customary to denote an element  $(y_k, y'_k) \in \mathcal{R}$  by  $y_k \rightarrow y'_k \in \mathcal{R}$  in which case we refer to  $y_k$  as the source complex and to  $y'_k$  as the product complex of reaction  $k$ . We may thus write  $\mathcal{R} = \{y_k \rightarrow y'_k: k = 1, \dots, r\}$ . Employing a standard, although slight abuse of, notation, we identify  $\mathcal{S} = \{S_1, \dots, S_d\}$  with the set  $\{1, \dots, d\}$  and  $\mathcal{R}$  with  $\{1, \dots, r\}$ . We write the  $k$ 'th reaction with the notation

$$\sum_{i \in \mathcal{S}} y_{ki} S_i \rightarrow \sum_{i \in \mathcal{S}} y'_{ki} S_i,$$

where  $y_{ki} = (y_k)_i$  and  $y'_{ki} = (y'_k)_i$  are the stoichiometric coefficients associated with the source and product complexes of reaction  $k$ , respectively. Define the reaction vectors  $\xi_k = y'_k - y_k$  and the stoichiometric matrix

$$\Xi = (\xi_1 \ \xi_2 \ \dots \ \xi_r) \in \mathbb{N}_0^{d \times r}.$$

The order of reaction  $k$  is the sum of the stoichiometric coefficients of the source complex,  $\sum_{i \in \mathcal{S}} y_{ki}$ . Finally, we define the maximum of a vector over the set  $\mathcal{R}$ ,  $x = \max_{k \in \mathcal{R}} y_k$ , as the entry-wise maximum,  $x_i = \max_{k \in \mathcal{R}} y_{ki}$ .

A set of reactions  $\mathcal{R}$  induces a set of complexes and a set of species, namely the complexes and species that appear in the reactions. We will assume that a reaction network is always given in this way by  $\mathcal{R}$ , and one may then completely describe a reaction network in terms of its reaction graph, whose nodes are the complexes and whose directed edges are the reactions. This concise description will be employed in the rest of the paper. To avoid trivialities, we assume  $\mathcal{R} \neq \emptyset$ .

For each reaction we specify an intensity function  $\lambda_k: \mathbb{N}_0^d \rightarrow [0, \infty)$ ,  $k \in \mathcal{R}$ , which satisfies the stoichiometric admissibility condition:

$$\lambda_k(x) > 0 \quad \Leftrightarrow \quad x \geq y_k,$$

where we use the usual vector inequality notation;  $x \geq y$  if  $x_i \geq y_i$  for all  $i \in \mathcal{S}$ . Thus, reactions are only allowed to take place whenever the copy-numbers of each species in the current state is at least as great as those of the corresponding source complex. A widely used example is stochastic mass-action kinetics given by

$$\lambda_k(x) = \alpha_k \prod_{i=1}^d y_{ki}! \binom{x}{y_k} = \alpha_k \prod_{i=1}^d \frac{x_i!}{(x_i - y_{ki})!},$$

for some reaction rate constants  $\alpha_k > 0$  [3]. The idea is that the rate is proportional to the number of distinct subsets of the molecules present that can form the input of the reaction. It reflects the assumption that the system is well-stirred [3]. Other examples include power law kinetics or generalized mass-action kinetics [2, 25, 34]. A particular choice of such rate functions constitute a stochastic kinetics  $\lambda = (\lambda_1, \dots, \lambda_r)$  for the reaction network  $\mathcal{N}$ , and the pair  $(\mathcal{N}, \lambda)$  is referred to as a stochastic reaction system, or simply a reaction network with kinetics  $\lambda$ .

<sup>1</sup>The terminology "species" is standard, although one may equally think of them as general entities or agents.

We may then specify the stochastic process  $(X_t: t \geq 0)$  on the state space  $D := \mathbb{N}_0^d$  related to the reaction system  $(\mathcal{N}, \lambda)$ . Let  $X_t$  be the vector in  $\mathbb{N}_0^d$  whose entries are the species counts at time  $t$ . If reaction  $y_k \rightarrow y'_k$  occurs at time  $t$ , then the new state is  $X_t = X_{t-} + y'_k - y_k = X_{t-} + \xi_k$ , where  $X_{t-}$  denotes the previous state. The stochastic process then follows,

$$X_t = X_0 + \sum_{k \in \mathcal{R}} Y_k \left( \int_0^t \lambda_k(X_s) ds \right) \xi_k, \quad (2.1)$$

where  $Y_k$  are independent and identically distributed unit-rate Poisson processes [3, 18, 37]. This stochastic equation is referred to as a random time change representation. We assume throughout the paper that the process is non-explosive, so that the process is well defined. Assumption 3.4, though, will imply non-explosiveness.

## 2.1 The state space

To define the set of endorsed states and absorbing states in the setting of stochastic reaction networks, we recall some terminology from stochastic processes. We say that there is a path from  $x$  to  $y$ , denoted  $x \mapsto y$ , if there exists  $t \geq 0$  such that  $\mathbb{P}_x(X_t = y) > 0$ . We extend this notion to sets as follows;  $B_1 \mapsto B_2$  if there exists  $x \in B_1$  and  $y \in B_2$  such that  $x \mapsto y$ . Finally, we introduce the region of large copy numbers, where all reactions may take place, defined as

$$R = \{x \in \mathbb{N}_0^d \mid \lambda_k(x) > 0 \forall k \in \mathcal{R}\}.$$

Any network satisfies  $R \neq \emptyset$ . Indeed, by the stoichiometric compatibility condition,  $\{x \in D \mid x \geq M'\} \subseteq R$  where  $M' = \max_{k \in \mathcal{R}} y_k \in \mathbb{N}_0^d$ . Letting  $D_E = \{x \in D \mid x \mapsto R\}$ , we may decompose the state space into a disjoint union

$$D = D_E \sqcup D_A.$$

A state space  $D$  is irreducible if for all  $x, y \in D$  we have  $\mathbb{P}_x(X_{t_1} = y) > 0$  and  $\mathbb{P}_y(X_{t_2} = x) > 0$  for some  $t_1, t_2 > 0$  [23]. Thus,  $D$  is irreducible if for all  $x, y \in D$  there exists a path  $x \mapsto y$ . Irreducibility induces a class structure on the state space [37], and we denote the classes by  $\mathcal{I}_1, \mathcal{I}_2, \dots$  (potentially infinitely many). Let  $\mathcal{J}$  denote the set of irreducible classes. Obviously, either  $\mathcal{I}_i \subseteq D_E$  or  $\mathcal{I}_i \subseteq D_A$ ,  $i \geq 1$ .

**Lemma 2.1.** *The pair  $(\mathcal{J}, \preceq)$ , where  $\preceq$  is given by*

$$\mathcal{I}_j \preceq \mathcal{I}_i \Leftrightarrow \mathcal{I}_i \mapsto \mathcal{I}_j, \quad i, j \geq 1,$$

*is a well defined poset. The irreflexive kernel  $(\mathcal{J}, \prec)$  gives a well defined strict poset.*

*Proof.* Since all elements  $\mathcal{I} \in \mathcal{J}$  are irreducible, there exists a path between any two points in  $\mathcal{I}$  hence  $\mathcal{I} \preceq \mathcal{I}$  yielding the relation reflexive.

Suppose  $\mathcal{I}_i \preceq \mathcal{I}_j$  and  $\mathcal{I}_j \preceq \mathcal{I}_i$  for some  $i, j \geq 1$ . Let  $x \in \mathcal{I}_i$  and  $y \in \mathcal{I}_j$  be given. By assumption, we may find a path from  $x$  to some  $z_2 \in \mathcal{I}_j$ , and by irreducibility of  $\mathcal{I}_j$  there is a path from  $z_2$  to  $y$ . Similarly, we may by assumption find a path from  $y$  to some  $z_1 \in \mathcal{I}_i$  and by irreducibility of  $\mathcal{I}_i$  a path from  $z_1$  to  $x$ . As  $x, y$  were arbitrary, we conclude that there exists a path between any two points in  $\mathcal{I}_i \cup \mathcal{I}_j$  hence  $\mathcal{I}_i = \mathcal{I}_j$ , yielding the relation antisymmetric.

Finally, suppose  $\mathcal{I}_k \preceq \mathcal{I}_j$  and  $\mathcal{I}_j \preceq \mathcal{I}_i$  for some  $i, j, k \geq 1$ . Then there exists a path from some  $x \in \mathcal{I}_i$  to some  $z_1 \in \mathcal{I}_j$  and a path from some  $z_2 \in \mathcal{I}_j$  to some  $y \in \mathcal{I}_k$ . By irreducibility of  $\mathcal{I}_j$  there is a path from  $z_1$  to  $z_2$ , and concatenation of the three paths yield one from  $x$  to  $y$ . We conclude that  $\mathcal{I}_k \preceq \mathcal{I}_i$ , hence the relation is transitive.  $\square$

A similar ordering has been considered in [45]. However, their further analysis rests on the setting of discrete time, rendering the approach insufficient for stochastic reaction networks. To exploit the graphical structure induced by  $\preceq$ , define the marked directed acyclic graph  $\mathcal{D} = (\mathcal{I}, \mathcal{E})$  as follows. The set of directed edges is

$$\mathcal{E} = \{(\mathcal{I}_i, \mathcal{I}_j) \in \mathcal{I}^2 \mid \mathcal{I}_j \prec \mathcal{I}_i, i, j \geq 1\},$$

while the marking  $\mathcal{I} = \mathcal{I}_A \sqcup \mathcal{I}_E$  is given by

$$\mathcal{I}_E = \{\mathcal{I} \in \mathcal{I} \mid \mathcal{I} \subseteq D_E\}, \quad \mathcal{I}_A = \{\mathcal{I} \in \mathcal{I} \mid \mathcal{I} \subseteq D_A\}.$$

Let  $\mathcal{V}_1, \mathcal{V}_2, \dots$  denote the vertex set of the respective connected components of the induced subgraph  $\mathcal{D}[\mathcal{I}_E]$ , the graph with vertex set  $\mathcal{I}_E$  and edges from  $\mathcal{E}$  with start and end nodes in  $\mathcal{I}_E$ .

**Definition 2.2.** The endorsed sets and absorbing sets are defined, respectively, by

$$E_n = \bigcup_{\mathcal{I} \in \mathcal{V}_n} \mathcal{I}, \quad A_n = \bigcup_{\mathcal{I} \in \mathcal{I}_A: E_n \mapsto \mathcal{I}} \mathcal{I}, \quad n \geq 1.$$

The corresponding state space is defined by  $D_n = E_n \sqcup A_n$ .

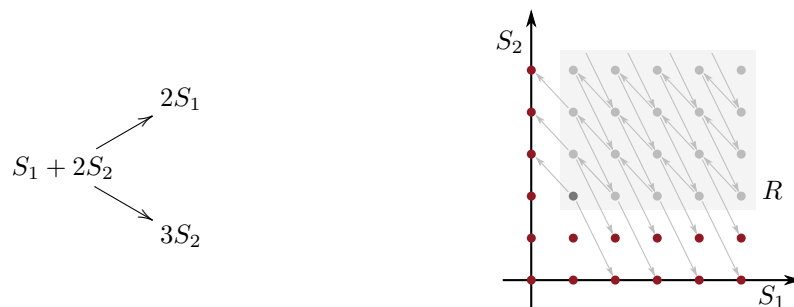


Figure 1: Left: The reaction graph of a stochastic reaction network. Right: The state space with the region  $R$  in shaded grey. Points in  $D_A$  are marked red.

As  $D_E$  is non-empty, the existence of at least one endorsed set is guaranteed. By construction, the endorsed sets are disjoint, their union is  $D_E$  and their number,  $N_E$ , may in general be countable infinite. Furthermore, any absorbing set is confined to a subset of  $\{x \in \mathbb{N}_0^d \mid x \not\geq M\}$ , lying “close” to the boundary of  $\mathbb{N}_0^d$ . If extinction is possible from an endorsed class  $E_n$  then this absorption will take place in  $A_n$ . Note, however, that the set  $\mathcal{I}_A$  may in general be empty, in which case no absorbing set exist. To further illuminate the structure of the endorsed sets, we provide the following classification result.

**Proposition 2.3.** For  $M \in \mathbb{N}^d$  sufficiently large, the set  $\{x \in D_E \mid x \geq M\}$  intersects

- (i) finitely many endorsed sets if and only if  $\text{rank } \Xi = d$ .
- (ii) a single endorsed set if and only if  $\text{span}_{\mathbb{Z}} \Xi = \mathbb{Z}^d$ .

*Proof.* (i) Suppose first that  $\text{rank } \Xi < d$ . Let  $x \in D_E$ . Then  $x \in E_1$ , say and there exists a  $y \in D_E$  such that  $y \notin (x + \text{span}_{\mathbb{R}} \Xi)$ . In particular,  $y \in E_2$  where  $E_1 \neq E_2$ . This procedure can be repeated indefinitely, yielding infinitely many endorsed sets.

Now, suppose  $\text{rank}(\Xi) = d$ . Then there exist  $a_{ik} \in \mathbb{Z}$  such that  $\sum_{k \in \mathcal{R}} a_{ik} \xi_k$  is in the interior of the  $i$ th orthant of  $\mathbb{R}^d$ ,  $i = 1, \dots, 2^d$ . Write

$$\alpha_i = \sum_{k \in \mathcal{R}} a_{ik} \xi_k = \alpha_i^+ - \alpha_i^-,$$

where  $\alpha_i^+$  is the sum over the terms with  $\text{sign}(a_{ik}) = +$ , and similarly  $\alpha_i^-$  is *minus* the sum over the terms with  $\text{sign}(a_{ik}) = -$ . It follows that, given  $x \in \mathbb{N}^d$  large enough, say  $x \geq M_i$ , then

$$x - \alpha_i^+ \mapsto x - \alpha_i, \quad x - \alpha_i^+ \mapsto x,$$

and  $x - (x - \alpha_i) = \alpha_i$ . Hence,  $x - \alpha_i$  and  $x$  belong to the same endorsed set.

Choose  $K \in \mathbb{N}$  such that  $K \geq 2 \max(|\alpha_{ij}| : i = 1, \dots, 2^d, j = 1, \dots, d)$  and  $M = \max(M_i : i = 1, \dots, d)$ . Define  $B_k = \{x \in \mathbb{Z}^d | M \leq x \leq M + k(1, \dots, 1)\}$ , for  $k \geq K$ . Then the same finitely many endorsed sets intersect  $B_k$  for any  $k \geq K$ . The proof is by induction. Clearly, the number of endorsed sets intersecting  $B_K$  is finite as the number of elements of  $B_k$  is finite. Assume all  $B_{k'}$ ,  $k \geq k' \geq K$ , are intersected by the same finitely many endorsed sets and consider  $k' = k + 1$ . Clearly,  $B_{k+1}$  is intersected by at least the same endorsed sets, so we only need to check that there are not more. For any  $x \in B_{k+1} \setminus B_k$  there is an orthant  $i \in \{1, \dots, 2^d\}$  such that  $x - \alpha_i \in B_k$ , by construction of  $\alpha_i$  (and  $K$ ). (Note that the orthant will depend on  $x \in B_{k+1} \setminus B_k$  and that there might be more than one orthant with the property.) Hence,  $x$  and  $x - \alpha_i$  belong to the same endorsed set as  $x \geq M \geq M_i$ .

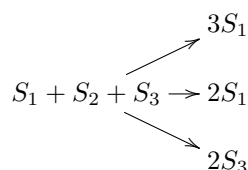
(ii) Finally, if there is one endorsed class only, then it must be that  $\text{span}_{\mathbb{Z}} \Xi = \mathbb{Z}^d$ . Oppositely, if  $\text{span}_{\mathbb{Z}} \Xi = \mathbb{Z}^d$ , then there are  $b_{ik} \in \mathbb{Z}$  such that

$$e_i = \sum_{k \in \mathcal{R}} b_{ik} \xi_k = e_i^+ - e_i^-,$$

where  $e_i$  is the  $i$ th unit vector,  $i = 1, \dots, d$ , and  $e_i^+, e_i^-$  are defined similarly to  $\alpha_i^+, \alpha_i^-$  above. By choosing  $M$  large enough it follows that for  $x \geq M$ , the elements  $x$  and  $x - e_i$  are in the same endorsed class.  $\square$

Note that, in particular, a reaction network whose associated stochastic process is a birth-death process, that is, a process where for each  $i = 1, \dots, d$  either  $e_i \in \mathcal{R}$  or  $-e_i \in \mathcal{R}$ , have a single endorsed set for  $x$  sufficiently large. In practice, one may find the endorsed sets by picking  $x \in R$  and adding states by a backtracking algorithm [38]. Verification of  $\text{span}_{\mathbb{Z}} \Xi = \mathbb{Z}^d$  can be done by calculation of the Hermite normal form [38].

One may suspect that Proposition 2.3 could be strengthened to hold on the entire set  $D_E$ . This is only partially true. Consider as an example the three-dimensional reaction network given by the reaction graph



It follows that  $\text{rank} \Xi = 3$ . However,  $D_E = \mathbb{N}^3$ ,  $D_A = \mathbb{N}_0^3 \setminus \mathbb{N}^3$ , and each singleton  $\{(1, m, 1)\}$ ,  $m \geq 1$ , constitutes its own endorsed set. Thus, in the generic picture, close to the absorbing set, there may be infinitely many endorsed sets. We do, however, have the following corollary.

**Corollary 2.4.** *If  $d \leq 2$ , there are finitely many endorsed sets if and only if  $\text{rank} \Xi = d$ .*

*Proof.* We only need to prove that if  $\text{rank} \Xi = d$  then there are finitely many endorsed sets. For this, it suffices to prove that at most finitely many  $x \in D_E$  do not have an undirected walk to a point  $z \geq M$ , where  $M$  is as in Proposition 2.3. Indeed, if such a path exists, then by definition  $x$  belongs to one of finitely many endorsed sets. With the remaining set being finite, the total number of endorsed sets is therefore finite.

Now, let  $x \in D_E$  be given. By the definition of endorsed sets, there exists a path  $x \mapsto y$  with  $\lambda_k(y) > 0$  for all  $k \in \mathcal{R}$ . As  $\text{rank } \Xi = d$ , for each  $1 \leq j \leq d$ , there exists  $k(j) \in \mathcal{R}$  such that  $\langle e_j, \xi_{k(j)} \rangle \neq 0$ . Keeping the  $j$ th coordinate fixed and increasing the possible other if necessary, thus arriving at a point  $y'$  with  $y_j = y'_j$  and  $y_i \leq y'_i$  for  $i \neq j$ , by the stoichiometric compatibility condition we may by repeated use of reaction  $k(j)$  find a path  $y' \mapsto z$  or  $z \mapsto y'$  with  $z \geq M$ . Repeating the argument for the possible remaining coordinate, we conclude that there exists an  $M' \in \mathbb{N}^d$  such that if  $x \not\prec M'$  then  $x$  has an undirected walk to a point  $z \geq M$ . As the set  $\{x \in D_E \mid x < M'\}$  is finite, this concludes the proof.  $\square$

One may easily verify, that the second part of Proposition 2.3 can also be extended for  $d = 1$ . Indeed, for any point  $x \in D_E$  there exists a reaction  $k \in \mathcal{R}$  such that  $\lambda_k(x) > 0$  and either  $x + \xi_k > x$  or  $x + \xi_k < x$ . Otherwise  $x \in D_A$ . Consequently, there is a point  $z \geq M$  for any  $M \in D_E$  such that either  $x \mapsto z$  or  $z \mapsto x$ . However, note that the network in Figure 1 shows that this result does not hold in the case  $d = 2$ .

An endorsed set,  $E_n$ ,  $n \geq 1$ , is only irreducible if it consists of a single irreducible class. If there is more than one irreducible class in  $E_n$ , then we need that there is a smallest one to ensure uniqueness of a QSD.

**Assumption 2.5.** For a given endorsed class  $E_n$ ,  $n \geq 1$ , we assume:

- (i)  $E_n$  contains a unique minimal irreducible class,  $\mathcal{I}_{\min}^n$ .
- (ii) if  $A_n \neq \emptyset$  then  $\mathcal{I}_{\min}^n \mapsto A_n$ .

We shall see that Assumption 2.5(i) is equivalent to a more technical property of the state space, which is necessary for our results to hold. Thus no generality is lost in having Assumption 2.5(i).

Networks without any minimal class exists, for example  $\emptyset \rightarrow S_1$ , which does not have an absorbing set either. Furthermore, networks with more than one minimal class also exist, for example,  $S_1 + S_2 \rightarrow \emptyset$ ,  $S_2 \rightarrow \emptyset$ . Thus Assumption 2.5(i) is indeed not superfluous. We believe Assumption 2.5(ii) is always met if Assumption 2.5(i) is. It ensures that one may always reach the absorbing set, if it is non-empty.

Definition 2.2 accommodates the general case where uniqueness of a QSD does not necessarily hold, in which case the support of the QSD may stretch the entire endorsed set rather than, as we shall see, the unique minimal irreducible class. We remark that rather than investigating an entire endorsed set, one may be interested in a particular irreducible component, say  $\mathcal{I}$ . Letting the state space be  $D = E \sqcup A$  where

$$E = \mathcal{I}, \quad A = \bigcup_{\mathcal{J} \in \mathcal{J}_A: E \mapsto \mathcal{J}} \mathcal{J},$$

the theory to be developed in this paper applies to this case as well.

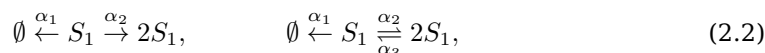
As an illuminating example consider the generalized death process  $mS_1 \rightarrow \emptyset$  with  $m \in \mathbb{N}$ , where each point in the state space  $D = \mathbb{N}_0$  constitutes its own irreducible class. Here, the endorsed and absorbing sets are  $E_n = \{n + pm - 1 \mid p \in \mathbb{N}\}$  and  $A_n = \{n - 1\}$  respectively, for  $n = 1, \dots, m$ , and Assumption 2.5 is satisfied for all  $n$ . Thus,  $D_E = \{m, m + 1, \dots\}$  and  $D_A = \{0, \dots, m - 1\}$ . It is known that in the simple death case,  $m = 1$ , uniqueness does not hold on  $D_E = \mathbb{N}$ . Indeed, there is a continuum of QSDs with support larger than  $\{1\}$ , the unique minimal class [20].



Figure 2: State space of the reaction network  $2S_1 \rightarrow \emptyset$ . There are two endorsed sets.



In the setting of birth-death processes in one dimension, which has an infinite state space, it is known that there will be either none, a unique or a continuum of QSDs [44]. Consider the two reaction networks



endowed with mass-action kinetics. In both cases we conclude, according to Definition 2.2, that the set of absorbing states and the set of endorsed states are

$$D_A = A_1 = \{0\}, \quad D_E = E_1 = \{1, 2, \dots\},$$

respectively. For the network on the left in (2.2), assuming  $\alpha_1 > \alpha_2$ , there is a continuum of QSDs on  $E_1$ , while for the network on the right, there is a unique QSD on  $E_1$ , for all parameter values [30]. This fits well with our result – the necessity of having reactions of order higher than one to ensure uniqueness permeates to higher dimensions.

### 3 Extension of arguments

In this and the following sections, we shall simply use the notation  $E$  to refer to a single endorsed set, with corresponding non-empty absorbing set  $A$ , when there is no ambiguity. Further, as existence and uniqueness is known on finite state spaces, we shall assume without loss of generality that  $E$  is countably infinite. We make the following definitions inspired by [9] and [31].

**Definition 3.1.** For any vector  $v \in \mathbb{N}^d$ , we define a corresponding function  $\langle v, \cdot \rangle: \mathbb{Z}^d \rightarrow \mathbb{Z}$  given by the standard inner product

$$\langle v, x \rangle = x \cdot v.$$

This function may in general take negative values, however, when restricting  $\langle v, \cdot \rangle$  to  $E$ , one obtains a norm-like function, [31]. Choosing  $v = (1, \dots, 1)$  we recover the function used in [9]. In general, we shall choose  $v \in \mathbb{N}^d$  based on the particular reaction network at hand, and will in the following consider it fixed. For  $n \in \mathbb{N}$ , define the sets

$$O_n = \{x \in E: \langle v, x \rangle \leq n\}.$$

which, irrespectively of  $v$ , are compact subsets of  $E$ , satisfying  $O_n \subseteq O_{n+1}$  and  $E = \bigcup_{n \in \mathbb{N}} O_n$ . We denote the first hitting time of  $A$ , the first hitting time of  $O_n$  and the first exit time of  $O_n$  by

$$\tau_A = \inf\{t \geq 0: X_t \in A\}, \quad \tau_n = \inf\{t \geq 0: X_t \in O_n\}, \quad T_n = \inf\{t \geq 0: X_t \notin O_n\},$$

respectively. Note that all of these are stopping times and might be infinite. As we will be concerned with the application of unbounded functions serving the purpose of a Lyapunov function, we introduce the weakened generator,  $L$ , for the Markov process [31, 8].

**Definition 3.2.** A measurable function  $W: D \rightarrow \mathbb{R}$  belongs to the domain  $\mathcal{D}(L)$  of the weakened generator  $L$  of  $(X_t: t \geq 0)$  if there exists a measurable function  $U: E \rightarrow \mathbb{R}$  such that, for all  $n \in \mathbb{N}$ ,  $t \geq 0$  and  $x \in E$

$$\mathbb{E}_x W(X_{t \wedge T_n}) = W(x) + \mathbb{E}_x \left( \int_0^{t \wedge T_n} U(X_s) ds \right),$$

and

$$\mathbb{E}_x \left| \int_0^{t \wedge T_n} U(X_s) ds \right| < \infty, \quad (3.1)$$

and we define  $LW = U$  on  $E$  and  $LW \equiv 0$  on  $A$ .

As the state space of interest is always countable, all functions  $f: D \rightarrow \mathbb{R}$  are measurable. Moreover, as the state space is discrete and  $O_n$  is finite for all  $n \in \mathbb{N}$ , all functions  $f: D \rightarrow \mathbb{R}$  are in the domain of the weakened generator,  $\mathcal{D}(L)$  [31]. In particular,  $\mathbb{E}_x W(X_{t \wedge T_n})$  is well-defined and finite. Generally, the weakened and the infinitesimal generator need not agree, and the infinitesimal generator may not exist [31].

However, if  $f$  is bounded, then they do agree. In particular,  $\mathbb{E}_x |f(X_t)| < \infty$  and it follows that as  $t \rightarrow 0$ ,

$$\begin{aligned} \mathbb{E}_x f(X_t) &= \left( \sum_{k \in \mathcal{R}} f(x + \xi_k) \mathbb{P}_x(X_t = x + \xi_k) \right) + f(x) \mathbb{P}_x(X_t = x) + o(t) \\ &= \left( \sum_{k \in \mathcal{R}} f(x + \xi_k) \lambda_k(x) t + o(t) \right) + f(x) \left( 1 - \sum_{k \in \mathcal{R}} \lambda_k(x) t + o(t) \right) + o(t) \\ &= \sum_{k \in \mathcal{R}} \lambda_k(x) (f(x + \xi_k) - f(x)) t + f(x) + o(t). \end{aligned}$$

Hence  $\mathbb{E}_x f(X_t)$  is differentiable and from the fundamental theorem of calculus we conclude that the weakened generator coincides with the (weak) infinitesimal generator [31],

$$\hat{L}f(x) = \lim_{t \rightarrow 0} \frac{\mathbb{E}_x f(X_t) - f(x)}{t} = \sum_{k \in \mathcal{R}} \lambda_k(x) (f(x + \xi_k) - f(x)),$$

for  $x \in E$ .

Moreover, setting  $W(x) = \langle v, x \rangle$  as in Definition 3.1, it follows from the Poisson characterization of the process (2.1), that

$$\begin{aligned} \mathbb{E}_x W(X_{t \wedge T_n}) &= \mathbb{E}_x \langle v, X_{t \wedge T_n} \rangle = \mathbb{E}_x \langle v, X_0 \rangle + \mathbb{E}_x \left\langle v, \sum_{k \in \mathcal{R}} Y_k \left( \int_0^{t \wedge T_n} \lambda_k(X_s) ds \right) \xi_k \right\rangle \\ &= W(x) + \mathbb{E}_x \left( \int_0^{t \wedge T_n} \sum_{k \in \mathcal{R}} \lambda_k(X_s) \langle v, \xi_k \rangle ds \right), \end{aligned}$$

such that

$$LW(x) = \sum_{k \in \mathcal{R}} \lambda_k(x) \langle v, \xi_k \rangle, \quad x \in E. \quad (3.2)$$

Note that (3.1) is fulfilled as  $O_n$  is finite.

**Definition 3.3.** Define the functions  $d_v, d^v: \mathbb{N} \rightarrow \mathbb{R}$  by

$$\begin{aligned} d_v(n) &= - \max_{x \in E, \langle v, x \rangle = n} \sum_{k=1}^r \lambda_k(x) \langle v, \xi_k \rangle \mathbb{1}_E(x + \xi_k), \\ d^v(n) &= \max_{x \in E, \langle v, x \rangle = n} n \sum_{k=1}^r \lambda_k(x) \mathbb{1}_A(x + \xi_k). \end{aligned}$$

All networks, for which extinction is possible, have the property that there exists a  $v \in \mathbb{N}^d$  such that  $\langle v, \xi_k \rangle \leq 0$  for some reaction  $k \in \mathcal{R}$ . Indeed, suppose that  $\langle v, \xi_k \rangle > 0$  for all  $v \in \mathbb{N}^d$  and  $k \in \mathcal{R}$ . Then  $\xi_k \in \mathbb{N}_0^d$  for all  $k \in \mathcal{R}$ , hence  $x + \xi_k \geq x$  for any  $x \in D$ . In particular, if  $x \in E$  then  $x + \xi_k \in E$  and we conclude, from the observation that any absorbing set is confined to a subset of  $\{x \in \mathbb{N}_0^d \mid x \not\geq M\}$  for some  $M$  sufficiently large,

that the process is not absorbed. By contraposition the desired claim holds. Note that for fixed  $n \in \mathbb{N}$ , the set  $\{x \in E, \langle v, x \rangle = n\}$  might be empty, thus we define

$$\mathcal{N} = \{n \in \mathbb{N} \mid \exists x \in E: \langle v, x \rangle = n\},$$

and make the following central assumption.

**Assumption 3.4.** *There exists  $v \in \mathbb{N}^d$  and  $\eta > 0$ ,  $N \in \mathbb{N}$ , such that, for  $n \geq N$ ,*

$$d_v(n) \geq \eta d^v(n), \quad (\text{i})$$

*and, with the limit being taken over  $\mathcal{N} \subseteq \mathbb{N}$ ,*

$$\lim_{n \rightarrow \infty} \frac{d_v(n)}{n^{1+\eta}} = \infty. \quad (\text{ii})$$

We note that, as  $d^v(n)$  is always non-negative, this assumption assures that  $d_v(n)$  is non-negative for  $n$  sufficiently large. We shall see that this assumption further ensures the ability to “come down from infinity” in finite time.

## 4 Verifying assumptions

We start by introducing some notation and definitions from [9] for ease of reference.

**Definition 4.1.** *A couple  $(V, \varphi)$  of measurable functions  $V$  and  $\varphi$  from  $D = E \cup A$  to  $\mathbb{R}$  is an admissible couple of functions if*

- (i)  *$V$  and  $\varphi$  are bounded and nonnegative on  $D$ , positive on  $E$ , satisfy  $V(x) = \varphi(x) = 0$  for all  $x \in A$ , and further*

$$\inf_{x \in E} \frac{V(x)}{\varphi(x)} > 0.$$

- (ii) *For all sequences  $(x_p)_{p \geq 1}$  in  $E$  such that  $\{p \in \mathbb{N} : x_p \in O_n\}$  is finite for all  $n \geq 1$ ,*

$$\lim_{p \rightarrow \infty} \frac{V(x_p)}{\varphi(x_p)} = \infty, \quad \text{and} \quad \lim_{n \rightarrow \infty} V(X_{T_n}) = 0 \quad \mathbb{P}_x\text{-a.s. for all } x \in E.$$

- (iii)  *$LV$  is bounded from above and  $L\varphi$  is bounded from below.*

The definition of a couple of admissible functions in [9] further requires that  $V$  and  $\varphi$  belong to the domain of the weakened infinitesimal generator of  $(X_t : t \geq 0)$ . However, since any function  $f : \mathbb{N}_0^d \rightarrow \mathbb{R}$  is in this domain for discrete state spaces [31], the requirement is automatically satisfied. Furthermore, as  $V, \varphi$  are bounded, the infinitesimal generator  $\widehat{L}$  is defined hereon and agrees with the weakened generator  $L$ .

The question of extinction has recently attracted much attention on its own [26]. Therefore, we provide the following proposition which renders an explicit criterion for when the stochastic process associated to a stochastic reaction network goes extinct almost surely. The assumption in the proposition is weaker than Assumption 3.4.

**Proposition 4.2.** *Under Assumption 2.5, with  $A \neq \emptyset$ , the process  $(X_t : t \geq 0)$  is absorbed  $\mathbb{P}_x$ -a.s. for all  $x \in E$  if  $d_v(n) > \zeta \frac{d^v(n)}{n}$  for  $n$  sufficiently large, where*

$$\zeta = \max_{k \in \mathcal{R}_A} \langle v, \xi_k \rangle, \quad \mathcal{R}_A = \{k \in \mathcal{R} : (E + \xi_k) \cap A \neq \emptyset\}.$$

*Proof.* Define the norm-like function  $W(x) = \langle v, x \rangle$  on  $\mathbb{N}_0^d$  as in Definition 3.1, and let  $L$  be the weakened infinitesimal generator of  $(X_t : t \geq 0)$ . It follows from (3.2) and

the assumption that  $d_v(n) > C \frac{d^v(n)}{n}$  for  $n$  sufficiently large that for each  $x \in E$  with  $\langle v, x \rangle = n$ ,

$$\begin{aligned} LW(x) &= \sum_{k=1}^r \lambda_k(x) \langle v, \xi_k \rangle \\ &= \sum_{k=1}^r \lambda_k(x) \langle v, \xi_k \rangle \mathbb{1}_E(x + \xi_k) + \sum_{k=1}^r \lambda_k(x) \langle v, \xi_k \rangle \mathbb{1}_A(x + \xi_k) \\ &\leq \max_{x' \in E, \langle v, x' \rangle = n} \sum_{k=1}^r \lambda_k(x') \langle v, \xi_k \rangle \mathbb{1}_E(x' + \xi_k) + \max_{x' \in E, \langle v, x' \rangle = n} \sum_{k=1}^r \lambda_k(x') \langle v, \xi_k \rangle \mathbb{1}_A(x' + \xi_k) \\ &\leq -d_v(n) + \zeta \max_{x' \in E, \langle v, x' \rangle = n} \sum_{k=1}^r \lambda_k(x') \mathbb{1}_A(x' + \xi_k) = -d_v(n) + \frac{\zeta}{n} d^v(n) < 0, \end{aligned} \quad (4.1)$$

for  $n$  sufficiently large. In particular, there exists an  $N \in \mathbb{N}$  such that for  $n = \langle v, x \rangle \geq N$ , we have  $LW(x) < 0$ , hence, setting  $M = \max_{x \in E: 1 \leq \langle v, x \rangle \leq N} \{0, LW(x)\}$ , yields

$$LW(x) \leq M \cdot \mathbb{1}_{O_N}(x), \quad x \in E.$$

Since  $O_N$  is compact, we may apply [31, Theorem 3.1] to conclude that the process  $(X_t: t \geq 0)$  is non-evanescent, that is,

$$\mathbb{P}_x \left( \langle v, X_t \rangle \xrightarrow{t \rightarrow \infty} \infty \right) = 0, \quad x \in E. \quad (4.2)$$

Define the discrete time jump chain  $(Y_n: n \in \mathbb{N}_0)$  by  $Y_n = X_{J_n}$ , where  $J_0, J_1, \dots$  denote the jump times of  $(X_t: t \geq 0)$  given by

$$J_0 = 0, \quad J_{n+1} = \inf\{t \geq J_n : X_t \neq X_{J_n}\}.$$

Let  $B_m = \{Y_n \in O_m \text{ i.o.}\}$  and  $F = \{Y_n \in A \text{ i.o.}\} = \{Y_n \in A \text{ for some } n\}$ , where the last equality follows from  $A$  being an absorbing set. By Assumption 2.5, all states in  $O_m \subseteq E$  have a shortest path to  $A$  (there is a path to the minimal irreducible class, and then to  $A$ , for any  $x \in O_m$ ), which has some positive probability. For each state  $x \in O_m$ , let  $b_x$  be the probability of this shortest path, and define  $\beta_m = \min_{x \in O_m} b_x$ . As  $O_m$  is compact,  $\beta_m > 0$ . It follows that for each  $n \in \mathbb{N}_0$  the conditioned process fulfils

$$\mathbb{P} \left( \bigcup_{k=1}^{\infty} (Y_{n+k} \in A) \mid Y_n \in O_m \right) \geq \beta_m > 0.$$

By [16, Theorem 2.3] we get

$$\mathbb{P}_y(B_m \setminus F) = 0, \quad (4.3)$$

with  $y = Y_0 = X_0 = x$ , for any  $m \in \mathbb{N}$ . Now, the complement of the event  $\bigcup_{m=0}^{\infty} B_m$  is the event  $G \cup F$ , where  $G = \left\{ \langle v, Y_n \rangle \xrightarrow{n \rightarrow \infty} \infty \right\}$ . As  $B_m$  is an increasing sequence of events in  $m$ , we obtain by monotone convergence and (4.3) that

$$1 = \mathbb{P}_y \left( G \cup F \cup \bigcup_m B_m \right) = \lim_{m \rightarrow \infty} \mathbb{P}_y(G \cup F \cup B_m) = \lim_{m \rightarrow \infty} \mathbb{P}_y(G \cup F) = \mathbb{P}_y(G \cup F).$$

Thus,  $(Y_n: n \in \mathbb{N}_0)$  either tends to infinity or is eventually absorbed in  $A$ . The same holds for the full process  $(X_t: t \geq 0)$ , and by (4.2) we conclude that  $\mathbb{P}_x(G) = 0$  hence  $\mathbb{P}_x(\tau_A < \infty) = \mathbb{P}_x(X_t \in A \text{ for some } t) = 1$ . In particular, we also have that

$$\lim_{n \rightarrow \infty} T_n = \tau_A,$$

thus the process is regularly absorbed, by definition.  $\square$

Note that  $\zeta$  in Proposition 4.2 may be negative thus Assumption 3.4(i) is stronger and immediately provides the same conclusion of almost sure absorption of the process. Further, as we shall see in the next proposition, Assumption 3.4(ii) assures that the expected magnitude of  $X_t$ , in the form of  $\langle v, X_t \rangle$  given  $X_0 = x$ , is uniformly bounded in  $x \in E$  for any  $t > 0$ . This, in turn, implies that the time of “coming down from infinity” is finite, which is closely related to the uniqueness of QSDs. This is where  $\eta > 0$  is required.

**Proposition 4.3.** *Under Assumptions 2.5-3.4 with  $A \neq \emptyset$ , the process  $(X_t, t \geq 0)$  satisfies*

$$\tau_A = \lim_{n \rightarrow \infty} T_n < \infty \quad \mathbb{P}_x\text{-a.s. for all } x \in E,$$

*in particular, the process is absorbed  $\mathbb{P}_x$ -a.s. Further,  $\sup_{x \in E} \mathbb{E}_x \langle v, X_t \rangle < \infty$  for any  $t > 0$ .*

*Proof.* By Assumption 3.4(i), it follows, applying the same notation as in Proposition 4.2, that

$$d_v(n) \geq \eta d^v(n) > \zeta \frac{d^v(n)}{n},$$

for  $n$  sufficiently large and  $\zeta$  as in the proposition. Thus by Proposition 4.2, the process  $(X_t: t \geq 0)$  satisfies

$$\tau_A = \lim_{n \rightarrow \infty} T_n < \infty,$$

$\mathbb{P}_x$ -a.s. for all  $x \in E$ . Hence the process is regularly absorbed.

The second claim is apparently a ‘classical result’ [8] but we are not aware of a proof in the literature, hence we provide one here. Let  $W(x) = \langle v, x \rangle$  on  $\mathbb{N}_0^d$  as in Definition 3.1. It follows from (4.1) of Proposition 4.2 and Assumption 3.4(i) that with  $\langle v, x \rangle = n$ ,

$$\begin{aligned} LW(x) &\leq -d_v(n) + \zeta \max_{x' \in E, \langle v, x' \rangle = n} \sum_{k=1}^r \lambda_k(x') \mathbb{1}_A(x' + \xi_k) \\ &\leq -d_v(n) + \zeta \frac{1}{n} d^v(n) \leq -d_v(n) + \frac{\zeta}{n\eta} d_v(n) = -\left(1 - \frac{\zeta}{n\eta}\right) d_v(n). \end{aligned}$$

It follows, under Assumption 3.4(ii), that

$$\frac{LW(x)}{W(x)^{1+\eta}} \leq -\left(1 - \frac{\zeta}{n\eta}\right) \frac{d_v(n)}{n^{1+\eta}} \rightarrow -\infty,$$

as  $n = \langle v, x \rangle \rightarrow \infty$  in  $\mathcal{N}$ , in which case  $1 - \zeta/(n\eta)$  becomes positive. Hence, there exist constants  $D_1, D_2 > 0$  such that

$$LW(x) \leq D_2 - D_1 W(x)^{1+\eta}, \quad \text{for all } x \in E. \quad (4.4)$$

Since we have  $\sum_{k \in \mathcal{R}} \lambda_k(x) < \infty$  for each  $x \in D$ , it follows from [3, p. 12] and the equivalence of the weakened and infinitesimal generators on  $W$  that

$$W(X_t) - W(0) - \int_0^t LW(X_s) ds$$

is a martingale. Thus, by the martingale property we find

$$\mathbb{E}_x W(X_t) = W(x) + \int_0^t \mathbb{E}_x (LW(X_s)) ds, \quad (4.5)$$

which is a form of Dynkin's formula [27]. Using the bound (4.4) combined with Jensen's inequality we obtain, upon differentiation of (4.5),

$$\frac{d}{dt} \mathbb{E}_x W(X_t) = \mathbb{E}_x (LW(X_t)) \leq \mathbb{E}_x (D_2 - D_1 W(X_t)^{1+\eta}) \leq D_2 - D_1 (\mathbb{E}_x W(X_t))^{1+\eta}.$$

Define  $f_x(t) = \mathbb{E}_x W(X_t)$  and choose  $D_3 > D_2$ . Consider the associated differential equations

$$g'_{x,\epsilon}(t) = D_3 - D_1 g_{x,\epsilon}(t)^{1+\eta}, \quad g_{x,\epsilon}(0) = W(x) + \epsilon \quad (4.6)$$

$$h'_x(t) = -D_1 h_x(t)^{1+\eta}, \quad h_x(0) = W(x). \quad (4.7)$$

Define  $F(t, z) = D_3 - D_1 z^{1+\eta}$ . It then follows that

$$\begin{aligned} f'_x(t) &< F(t, f_x(t)), & f_x(0) &= W(x), \\ h'_x(t) &< F(t, h_x(t)), & h_x(0) &= W(x), \\ g'_{x,\epsilon}(t) &= F(t, g_{x,\epsilon}(t)), & g_{x,\epsilon}(0) &= W(x) + \epsilon. \end{aligned}$$

By Petrovitsch' theorem [33, p. 316] applied twice and the fact that solutions to the ordinary differential equations are continuous in the initial value, we conclude that

$$f_x(t) \leq g_x(t), \quad h_x(t) \leq g_x(t) \quad t \in [0, T], \quad (4.8)$$

where  $g_x(t) = \lim_{\epsilon \rightarrow 0} g_{x,\epsilon}(t)$ . The solution to the initial value problem (4.6) cannot be given in explicit form, however, the associated simpler differential equation (4.7) does have an explicit solution for  $\eta > 0$ , given by

$$h_x(t) = \frac{1}{(D_1 \eta t + W(x)^{-\eta})^{1/\eta}}, \quad t \geq 0.$$

In order to use this to bound  $f_x(t)$ , define the function

$$k_t(x) = D_3 t + h_x(t).$$

Then, using (4.8), we have

$$g'_x(t) = D_3 - D_1 g_x(t)^{1+\eta} \leq D_3 - D_1 h_x(t)^{1+\eta} = D_3 + h'_x(t) = k'_x(t).$$

Since  $g_x(0) = W(x) = k_x(0)$  it follows that  $g_x(t) \leq k_x(t)$  and we infer that

$$\sup_{x \in E} \mathbb{E}_x W(X_t) = \sup_{x \in E} f_x(t) \leq \sup_{x \in E} g_x(t) \leq \sup_{x \in E} k_x(t) < \infty,$$

for all fixed  $t > 0$  as desired.  $\square$

Inspired by [9], we now make the following definition, which allows us to generalize their method of proof for the subsequent lemma.

**Definition 4.4.** Let  $v \in \mathbb{N}^d$  and  $\alpha, \beta > 1$ . Define  $V: \mathbb{N}_0^d \rightarrow \mathbb{R}$  and  $\varphi: \mathbb{N}_0^d \rightarrow \mathbb{R}$  by

$$V(x) = \mathbb{1}_E(x) \sum_{j=1}^{\langle v, x \rangle} \frac{1}{j^\alpha}, \quad \varphi(x) = \mathbb{1}_E(x) \sum_{j=\langle v, x \rangle + 1}^{\infty} \frac{1}{j^\beta}.$$

**Lemma 4.5.** Under Assumption 3.4, for suitable choices of  $\alpha, \beta > 1$ , the pair  $(V, \varphi)$  satisfies

(a)  $V, \varphi$  are bounded.

(b) There exists an integer  $n$  and a constant  $C \geq 0$ , such that

$$-L\varphi \leq C \mathbb{1}_{O_n}.$$

(c) There exists constants  $\epsilon, C' > 0$  and  $C'' \geq 0$ , such that

$$LV + C' \frac{V^{1+\epsilon}}{\varphi^\epsilon} \leq C'' \varphi.$$

*Proof.* As  $v \in \mathbb{N}^d$ , it follows that  $\langle v, x \rangle \in \mathbb{N}_0$  for all  $x \in E$ . In particular,  $V(x), \varphi(x) \leq \sum_{j=1}^{\infty} \frac{1}{j^p}$  with  $p > 1$  which is a convergent hyperharmonic series. This proves (a). Thinking in terms of Riemann sums, using that  $j \mapsto j^{-\beta}$  is decreasing in  $j$  for  $\beta > 1$ , we obtain the bound

$$\sum_{j=\langle v, x \rangle + 1}^{\infty} \frac{1}{j^\beta} \leq \int_{\langle v, x \rangle}^{\infty} \frac{1}{y^\beta} dy = \frac{\langle v, x \rangle^{1-\beta}}{\beta - 1}.$$

Exploiting the linearity of  $\langle v, \cdot \rangle$ , we may now follow the method of [9]. Consider for  $x \in E$  with  $n = \langle v, x \rangle$ , and  $L$  the weakened generator which coincides with the infinitesimal generator since  $\varphi$  is bounded,

$$\begin{aligned} L\varphi(x) &= \sum_{k=1}^r \lambda_k(x) (\varphi(x + \xi_k) - \varphi(x)) \\ &= \sum_{k=1}^r \lambda_k(x) \mathbb{1}_E(x + \xi_k) (\varphi(x + \xi_k) - \varphi(x)) + \sum_{k=1}^r \lambda_k(x) \mathbb{1}_A(x + \xi_k) (\varphi(x + \xi_k) - \varphi(x)) \\ &= \sum_{k=1}^r \lambda_k(x) \mathbb{1}_E(x + \xi_k) \left( \sum_{j=\langle v, x + \xi_k \rangle + 1}^{\infty} \frac{1}{j^\beta} - \sum_{j=\langle v, x \rangle + 1}^{\infty} \frac{1}{j^\beta} \right) \\ &\quad - \sum_{k=1}^r \lambda_k(x) \mathbb{1}_A(x + \xi_k) \sum_{j=\langle v, x \rangle + 1}^{\infty} \frac{1}{j^\beta} \\ &= - \sum_{k=1}^r \lambda_k(x) \mathbb{1}_E(x + \xi_k) \mathbb{1}_{(0, \infty)}(\langle v, \xi_k \rangle) \sum_{i=1}^{\langle v, \xi_k \rangle} \frac{1}{(\langle v, x \rangle + i)^\beta} \\ &\quad + \sum_{k=1}^r \lambda_k(x) \mathbb{1}_E(x + \xi_k) \mathbb{1}_{(-\infty, 0)}(\langle v, \xi_k \rangle) \sum_{i=\langle v, \xi_k \rangle}^{-1} \frac{1}{(\langle v, x \rangle + i + 1)^\beta} \\ &\quad - \sum_{k=1}^r \lambda_k(x) \mathbb{1}_A(x + \xi_k) \sum_{j=\langle v, x \rangle + 1}^{\infty} \frac{1}{j^\beta} \\ &\geq \frac{1}{\langle v, x \rangle^\beta} \left( - \sum_{k=1}^r \lambda_k(x) \mathbb{1}_E(x + \xi_k) \mathbb{1}_{(0, \infty)}(\langle v, \xi_k \rangle) \langle v, \xi_k \rangle \right. \\ &\quad \left. - \sum_{k=1}^r \lambda_k(x) \mathbb{1}_E(x + \xi_k) \mathbb{1}_{(-\infty, 0)}(\langle v, \xi_k \rangle) \langle v, \xi_k \rangle \right) - \sum_{k=1}^r \lambda_k(x) \mathbb{1}_A(x + \xi_k) \frac{\langle v, x \rangle^{1-\beta}}{\beta - 1} \\ &= \frac{1}{\langle v, x \rangle^\beta} \left( - \sum_{k=1}^r \lambda_k(x) \mathbb{1}_E(x + \xi_k) \langle v, \xi_k \rangle - \sum_{k=1}^r \lambda_k(x) \mathbb{1}_A(x + \xi_k) \frac{\langle v, x \rangle}{\beta - 1} \right) \\ &\geq \frac{1}{\langle v, x \rangle^\beta} \left( - \max_{x' \in E, \langle v, x' \rangle = n} \sum_{k=1}^r \lambda_k(x') \langle v, \xi_k \rangle \mathbb{1}_E(x' + \xi_k) \right. \\ &\quad \left. - \max_{x' \in E, \langle v, x' \rangle = n} \sum_{k=1}^r \lambda_k(x') \mathbb{1}_A(x' + \xi_k) \frac{\langle v, x' \rangle}{\beta - 1} \right) \\ &= \frac{1}{n^\beta} \left( d_v(n) - \frac{d^v(n)}{\beta - 1} \right). \end{aligned}$$

Using Assumption 3.4(i), we can choose  $\beta > 1$  large enough such that  $L\varphi(x) \geq 0$  for all  $n = \langle v, x \rangle$  sufficiently large. In particular, condition (b) is satisfied.

Similarly, as  $V$  is bounded, we find for  $x \in E$  with  $\langle x, v \rangle = n$ ,

$$\begin{aligned}
 LV(x) &= \sum_{k=1}^r \lambda_k(x) (V(x + \xi_k) - V(x)) \\
 &= \sum_{k=1}^r \lambda_k(x) \mathbb{1}_E(x + \xi_k) \left( \sum_{j=1}^{\langle v, x + \xi_k \rangle} \frac{1}{j^\alpha} - \sum_{j=1}^{\langle v, x \rangle} \frac{1}{j^\alpha} \right) - \sum_{k=1}^r \lambda_k(x) \mathbb{1}_A(x + \xi_k) \sum_{j=1}^{\langle v, x \rangle} \frac{1}{j^\alpha} \\
 &\leq \sum_{k=1}^r \lambda_k(x) \mathbb{1}_E(x + \xi_k) \mathbb{1}_{(0, \infty)}(\langle v, \xi_k \rangle) \sum_{j=\langle v, x \rangle + 1}^{\langle v, x + \xi_k \rangle} \frac{1}{j^\alpha} \\
 &\quad - \sum_{k=1}^r \lambda_k(x) \mathbb{1}_E(x + \xi_k) \mathbb{1}_{(-\infty, 0)}(\langle v, \xi_k \rangle) \sum_{j=\langle v, x + \xi_k \rangle + 1}^{\langle v, x \rangle} \frac{1}{j^\alpha} \\
 &= \sum_{k=1}^r \lambda_k(x) \mathbb{1}_E(x + \xi_k) \mathbb{1}_{(0, \infty)}(\langle v, \xi_k \rangle) \sum_{j=1}^{\langle v, \xi_k \rangle} \frac{1}{(j + \langle v, x \rangle)^\alpha} \\
 &\quad - \sum_{k=1}^r \lambda_k(x) \mathbb{1}_E(x + \xi_k) \mathbb{1}_{(-\infty, 0)}(\langle v, \xi_k \rangle) \sum_{j=\langle v, \xi_k \rangle}^{-1} \frac{1}{(j + \langle v, x \rangle + 1)^\alpha} \\
 &\leq \sum_{k=1}^r \lambda_k(x) \mathbb{1}_E(x + \xi_k) \mathbb{1}_{(0, \infty)}(\langle v, \xi_k \rangle) \frac{\langle v, \xi_k \rangle}{\langle v, x \rangle^\alpha} \\
 &\quad + \sum_{k=1}^r \lambda_k(x) \mathbb{1}_E(x + \xi_k) \mathbb{1}_{(-\infty, 0)}(\langle v, \xi_k \rangle) \frac{\langle v, \xi_k \rangle}{\langle v, x \rangle^\alpha} \\
 &\leq \frac{1}{\langle v, x \rangle^\alpha} \sum_{k=1}^r \lambda_k(x) \langle v, \xi_k \rangle \mathbb{1}_E(x + \xi_k) \\
 &\leq \frac{1}{\langle v, x \rangle^\alpha} \max_{x' \in E, \langle v, x' \rangle = n} \sum_{k=1}^r \lambda_k(x') \langle v, \xi_k \rangle \mathbb{1}_E(x' + \xi_k) = -\frac{d_v(n)}{n^\alpha}.
 \end{aligned}$$

Note that by treating  $V(x)$  as a lower Riemann sum, for  $x \in E$ ,

$$V(x) \leq \sum_{j=1}^{\infty} \frac{1}{j^\alpha} = 1 + \sum_{j=2}^{\infty} \frac{1}{j^\alpha} \leq 1 + \int_1^{\infty} \frac{1}{x^\alpha} dx = \frac{\alpha}{\alpha - 1},$$

and similarly, treating  $\varphi(x)$  as an upper Riemann sum,

$$\varphi(x) = \sum_{j=\langle v, x \rangle + 1}^{\infty} \frac{1}{j^\beta} \geq \int_{\langle v, x \rangle + 1}^{\infty} \frac{1}{x^\beta} dx = \frac{(1 + \langle v, x \rangle)^{1-\beta}}{\beta - 1} \geq \frac{\langle v, x \rangle^{1-\beta}}{2(\beta - 1)},$$

with the last inequality holding for  $\langle v, x \rangle$  sufficiently large, using that for  $\beta > 1$ ,

$$\frac{(1 + \langle v, x \rangle)^{1-\beta}}{\langle v, x \rangle^{1-\beta}} = \left( \frac{1}{\langle v, x \rangle} + 1 \right)^{1-\beta} \rightarrow 1 > \frac{1}{2}, \quad \text{for } n = \langle v, x \rangle \rightarrow \infty.$$

We infer that

$$LV(x) + \frac{V^{1+\epsilon}(x)}{\varphi^\epsilon(x)} \leq -\frac{d_v(n)}{n^\alpha} + Cn^{\epsilon(\beta-1)},$$



where  $C = [\alpha/(\alpha - 1)]^{1+\epsilon}[2(\beta - 1)]^\epsilon$ . Note that by definition  $\varphi(x) > 0$  for  $x \in E$  hence, choosing  $\alpha = 1 + \eta/2$  and  $\epsilon = \eta/[2(\beta - 1)]$ , we get

$$LV(x) + \frac{V^{1+\epsilon}(x)}{\varphi^\epsilon(x)} \leq -\frac{d_v(n)}{n^{1+\eta/2}} + Cn^{\eta/2} = \left(C - \frac{d_v(n)}{n^{1+\eta}}\right)n^{\eta/2}.$$

Using Assumption 3.4(ii), the first term becomes negative for  $\langle v, x \rangle$  sufficiently large. In other words  $LV(x) + \frac{V^{1+\epsilon}(x)}{\varphi^\epsilon(x)} \leq 0$  for  $x \notin O_n$  with  $n$  sufficiently large. Since by definition  $\varphi(x) \geq 0$ , condition (c) holds.  $\square$

**Lemma 4.6.** *Under Assumptions 2.5-3.4, the pair  $(V, \varphi)$  is an admissible couple of functions.*

*Proof.* Choose  $\alpha, \beta \in \mathbb{R}$  such that the conclusions of Lemma 4.5 hold. In particular,  $\alpha, \beta > 1$  hence the functions  $V$  and  $\varphi$  are bounded, non-negative on  $E \cup A$  and positive on  $E$ , and by definition  $V(x) = \varphi(x) = 0$  for  $x \in A$ . Furthermore,  $\inf_{x \in E} V(x) > 0$ , hence by non-negativity of  $\varphi$ ,

$$\inf_{x \in E} \frac{V(x)}{\varphi(x)} > 0,$$

and Definition 4.1(1) is fulfilled. Let  $(x_p)_{p \geq 1}$  be any sequence in  $E$  such that the set  $\{p \in \mathbb{N} : x_p \in O_n\}$  is finite for all  $n \geq 1$ . Then  $\langle v, x_p \rangle \rightarrow \infty$  as  $p \rightarrow \infty$ , hence

$$\lim_{p \rightarrow \infty} \frac{V(x_p)}{\varphi(x_p)} = \lim_{p \rightarrow \infty} \frac{\sum_{j=1}^{\langle v, x_p \rangle} \frac{1}{j^\alpha}}{\sum_{j=\langle v, x_p \rangle+1}^{\infty} \frac{1}{j^\beta}} = \infty.$$

Furthermore, since the process is regularly absorbed by Proposition 4.3, we have

$$\lim_{n \rightarrow \infty} V(X_{T_n}) = V(X_{\tau_A}) = 0 \quad \mathbb{P}_x\text{-a.s.},$$

hence Definition 4.1(2) is fulfilled. Finally, from Lemma 4.5 and the fact that  $V$  and  $\varphi$  are both bounded functions, it follows that  $LV$  is bounded from above and  $L\varphi$  is bounded from below. This concludes the proof.  $\square$

#### 4.1 Lemmas

**Lemma 4.7.** *Assumption 2.5(i) is equivalent to the following: There exists  $n_0 \in \mathbb{N}$ ,  $\theta_0, \theta_1, a_1 > 0$  and a probability measure  $\nu$  on  $E$  such that, for all  $x \in O_{n_0}$  and all  $s \in [\theta_0, \theta_0 + \theta_1]$ ,*

$$\mathbb{P}_x(X_s \in \cdot) \geq a_1 \nu, \quad (4.9)$$

and in addition, for all  $n \geq n_0$ , there exists  $s_n \geq 0$  such that

$$\inf_{x \in O_n} \mathbb{P}_x(X_{s_n} \in O_{n_0}) > 0. \quad (4.10)$$

*Proof.* We first prove that Assumption 2.5 implies the existence of such constants and probability measure. Let  $\mathcal{I}_{\min}$  be the unique minimal irreducible class contained in  $E$ . Set

$$n_0 = \inf\{n \in \mathbb{N} : O_n \cap \mathcal{I}_{\min} \neq \emptyset\} < \infty.$$

Pick  $z \in \mathcal{I}_{\min} \cap O_{n_0}$  arbitrarily and let  $\nu = \delta_z$ . Pick arbitrary  $\theta_0, \theta_1 > 0$  and let

$$a_1 = \inf_{s \in [\theta_0, \theta_0 + \theta_1], x \in O_{n_0}} \mathbb{P}_x(X_s = z).$$

Note that for all  $n \in \mathbb{N}$  the set  $O_n$  is finite and any  $x \in O_n \subset E$  has a path to  $\mathcal{I}_{\min}$  and thus, by irreducibility, to  $z$ . By continuity of  $\mathbb{P}_x(X_\cdot = z)$ , we conclude that  $a_1 > 0$  and choosing  $s_n = \theta_0 > 0$  for all  $n$ , the desired holds.

For the reverse direction, we first prove uniqueness of the minimal irreducible class in the endorsed set  $E$ . Suppose for contradiction that  $\mathcal{I}_i \neq \mathcal{I}_j$  are irreducible minimal classes in  $E$ . Let  $x_1 \in \mathcal{I}_i$  and  $x_2 \in \mathcal{I}_j$ . Then there is a path  $x_1 \mapsto y_1$  and a path  $x_2 \mapsto y_2$  with  $y_1, y_2 \in O_{n_0}$ . Indeed, if  $\mathcal{I}_\ell \cap O_{n_0} \neq \emptyset$  for  $\ell = i, j$  we may simply choose  $y_\ell = x_\ell$ ,  $\ell = 1, 2$ . Otherwise,  $x_1, x_2$  are in some  $O_{n_1}$  and  $O_{n_2}$  respectively with  $n_1, n_2 > n_0$ , hence by (4.10), there exist paths as described. By (4.9), there exist  $\theta_0, \theta_1, a_1 > 0$  and a probability measure  $\nu$  on  $E$  such that, for all  $y \in O_{n_0}$  and all  $s \in [\theta_1, \theta_1 + \theta_0]$ ,

$$\mathbb{P}_y(X_s \in \cdot) \geq a_1 \nu.$$

As  $\nu$  is a probability measure on a countable space, there exists some  $z \in E$  such that  $\nu(\{z\}) > 0$ . But then  $\mathbb{P}_y(X_s = z) > 0$  for all  $y \in O_{n_0}$ . We conclude that there exist paths  $y_1 \mapsto z$  and  $y_2 \mapsto z$ . By minimality of  $\mathcal{I}_i$  and  $\mathcal{I}_j$ , we conclude that  $z \in \mathcal{I}_i \cap \mathcal{I}_j$  which is a contradiction. This proves the uniqueness in Assumption 2.5(i).

We now prove existence. Suppose for contradiction that no minimal class exists. Then, for all  $\mathcal{I} \subseteq E$  there exists  $\mathcal{J} \subseteq E$  such that  $\mathcal{I} \rightarrow \mathcal{J}$ . Repeating the argument results in an infinite path

$$\mathcal{J}_1 \rightarrow \mathcal{J}_2 \rightarrow \mathcal{J}_3 \rightarrow \dots \quad (4.11)$$

In the case where there are only finitely many irreducible classes in  $E$ , there must exist an  $i \in \mathbb{N}$  such that  $\mathcal{J}_1 \rightarrow \mathcal{J}_i = \mathcal{J}_1$  which contradicts the lack of cycles in  $\mathcal{D}$ .

Suppose therefore that there are infinitely many classes. Given  $n_0$ , the set  $O_{n_0}$  is finite. Thus it intersects at most finitely many irreducible classes. We conclude that in the infinite path (4.11), there are infinitely many  $i \in \mathbb{N}$  such that

$$\mathcal{I}_i \cap O_{n_0} = \emptyset, \quad \mathcal{I}_i \cap O_{n_i} \neq \emptyset,$$

for some  $n_i > n_0$ . However, by (4.10) each of these classes have a path to  $O_{n_0}$ . This implies the existence of at least one irreducible class intersecting  $O_{n_0}$  which appears more than once in the infinite path (4.11). This creates a cycle in  $\mathcal{D}$  which is a contradiction.  $\square$

**Lemma 4.8.** *Under Assumptions 2.5-3.4, for all  $\lambda > 0$ , there exists  $n \geq 1$  such that*

$$\sup_{x \in E} \mathbb{E}_x \left( e^{\lambda(\tau_n \wedge \tau_A)} \right) < \infty. \quad (4.12)$$

*Proof.* It follows from Proposition 4.3 that for all  $t > 0$  there exists a constant  $M > 0$  such that  $\sup_{x \in E} \mathbb{E}_x \langle v, X_t \rangle < M$ . Let  $\tau_{n,A} = \tau_n \wedge \tau_A$ . Then, for  $\epsilon > 0$

$$\mathbb{E}_x \langle v, X_\epsilon \rangle = \sum_{n=0}^{\infty} \mathbb{P}_x(\langle v, X_\epsilon \rangle > n) \geq \sum_{n=0}^{\infty} \mathbb{P}_x(\tau_{n,A} > \epsilon).$$

Suppose for contradiction that  $\mathbb{P}_x(\tau_{n,A} > \epsilon)$  does not converge uniformly in  $x$  to 0 as  $n \rightarrow \infty$ . Then for any  $\delta > 0$ ,  $\sup_{x \in E} \mathbb{P}_x(\tau_{n,A} > \epsilon) > \delta$  for infinitely many  $n \in \mathbb{N}$ . Thus, choosing  $\delta = \epsilon$ , there exists a sequence  $(x_i, n_i)_{i \geq 1}$  with  $\langle v, x_i \rangle > n_i$  such that  $\mathbb{P}_{x_i}(\tau_{n_i,A} > \epsilon) > \epsilon$  and  $n_i > n_{i-1}$ . But then, noting that  $\mathbb{P}_x(\tau_{n-1,A} > \epsilon) > \mathbb{P}_x(\tau_{n,A} > \epsilon)$  for all  $n \in \mathbb{N}$ , we obtain

$$\mathbb{E}_{x_i} \langle v, X_\epsilon \rangle \geq \sum_{n=0}^{\infty} \mathbb{P}_{x_i}(\tau_{n,A} > \epsilon) \geq \sum_{n=0}^{n_i} \mathbb{P}_{x_i}(\tau_{n,A} > \epsilon) \geq n_i \mathbb{P}_{x_i}(\tau_{n_i,A} > \epsilon) > n_i \epsilon.$$

Letting  $i \rightarrow \infty$  we would have  $M > \lim_{i \rightarrow \infty} \mathbb{E}_{x_i} \langle v, X_\epsilon \rangle > \lim_{i \rightarrow \infty} n_i \epsilon = \infty$ , which is a contradiction. We conclude that  $\sup_{x \in E} \mathbb{P}_x(\tau_{n,A} > \epsilon) \leq \epsilon$  for  $n \geq N(\epsilon)$ .

Now, given  $t > 0$  we can choose  $0 < \epsilon < 1$  such that  $t = p\epsilon$  for some  $p \in \mathbb{N}$ . Applying the Markov property we get, for  $n \geq N(\epsilon)$ ,

$$\begin{aligned} \mathbb{P}_x(\tau_{n,A} \geq t) &= \mathbb{P}_x(\tau_{n,A} \geq p\epsilon) \\ &= \sum_{y \in D} \mathbb{P}_x(\tau_{n,A} > p\epsilon \mid \tau_{n,A} > (p-1)\epsilon, X_{(p-1)\epsilon} = y) \mathbb{P}_x(\tau_{n,A} > (p-1)\epsilon, X_{(p-1)\epsilon} = y) \\ &= \sum_{y \in D} \mathbb{P}_y(\tau_{n,A} > \epsilon) \mathbb{P}_x(\tau_{n,A} > (p-1)\epsilon, X_{(p-1)\epsilon} = y) \leq \epsilon \mathbb{P}_x(\tau_{n,A} > (p-1)\epsilon) \leq \epsilon^p, \end{aligned}$$

hence we conclude, as  $t < p$ , that

$$\sup_{x \in E} \mathbb{P}_x(\tau_{n,A} \geq t) \leq \epsilon^t.$$

Since  $e^{\lambda(\tau_n \wedge \tau_A)}$  is non-negative, by choosing  $\epsilon < e^{-\lambda} < 1$  we get

$$\begin{aligned} \sup_{x \in E} \mathbb{E}_x \left( e^{\lambda(\tau_n \wedge \tau_A)} \right) &= \sup_{x \in E} \int_0^\infty \mathbb{P}_x \left( e^{\lambda(\tau_n \wedge \tau_A)} \geq t \right) dt \leq 1 + \sup_{x \in E} \int_1^\infty \mathbb{P}_x \left( \tau_{n,A} \geq \frac{\ln t}{\lambda} \right) dt \\ &\leq 1 + \int_1^\infty \epsilon^{\ln t / \lambda} dt = 1 + \int_0^\infty (e\epsilon^{1/\lambda})^u du = 1 + \frac{1}{1 + \ln(\epsilon)/\lambda} < \infty, \end{aligned}$$

thus (4.12) holds as desired.  $\square$

**Lemma 4.9.** *For all  $n \geq 0$ , there exists a constant  $C_n$  such that, for all  $t \geq 0$ ,*

$$\sup_{x \in O_n} \mathbb{P}_x(t < \tau_A) \leq C_n \inf_{x \in O_n} \mathbb{P}_x(t < \tau_A).$$

Further, with  $V$  from Definition 4.4, there exist constants  $r_0, p_0 > 0$  such that for  $n$  sufficiently large,

$$\mathbb{P}_x(r_0 < \tau_A) \leq p_0 V(x), \quad \text{for all } x \in E \setminus O_n.$$

*Proof.* Let  $n \geq 0$  be given. If  $O_n$  is empty, then the statement is vacuously true for any  $C_n$ . If  $O_n$  is non-empty, then as  $\sup_{x \in O_n} \mathbb{P}_x(t < \tau_A) \leq 1$  and  $\inf_{x \in O_n} \mathbb{P}_x(t < \tau_A) > 0$  since  $O_n$  is finite, we may simply choose

$$C_n = \frac{\sup_{x \in O_n} \mathbb{P}_x(t < \tau_A)}{\inf_{x \in O_n} \mathbb{P}_x(t < \tau_A)} < \infty.$$

To see the second claim, note that letting

$$p_0 = \frac{1}{\inf_{x \in E} V(x)} > 0,$$

we have for any  $r_0 > 0$ ,

$$\mathbb{P}_x(r_0 < \tau_A) \leq 1 = p_0 \inf_{x \in E} V(x) \leq p_0 V(x),$$

for all  $x \in E$ , and the desired holds.  $\square$

## 5 The main result

We are now ready to state and prove the main result of the paper. In the case where only a single endorsed set,  $E$ , is considered, we find that the unique QSD hereon is in fact globally attracting in the space of probability measures on  $E$ .

**Theorem 5.1.** *A reaction network  $(\mathcal{N}, \lambda)$  with associated stochastic process  $(X_t: t \geq 0)$  on  $D = E \sqcup A$ , with  $A \neq \emptyset$  and satisfying Assumption 2.5-3.4, admits a unique quasi-stationary distribution  $\nu$ . Further, there exist constants  $C, \gamma > 0$  such that, for all probability measures  $\mu$  on  $E$ ,*

$$\|\mathbb{P}_\mu(X_t \in \cdot | t < \tau_A) - \nu\|_{TV} \leq Ce^{-\gamma t}, \quad t \geq 0.$$

*Proof.* By Proposition 4.3, the process is regularly absorbed, and by Lemma 4.6 the pair  $(V, \varphi)$  given in Definition 4.4 is admissible, satisfying conditions (a) and (b) from Lemma 4.5. The result now follows from Lemma 4.7-4.9 together with [9, Cor. 2.8].  $\square$

In the case where  $A = \emptyset$ , and hence  $d^v(n) = 0$ , this indicates that Assumption 3.4(ii) might be sufficient for uniform exponential ergodicity. Note that this assumption implies that

$$L\langle v, x \rangle = \sum_{k=1}^r \lambda_k(x) \langle v, \xi_k \rangle \leq \max_{x' \in E, \langle v, x' \rangle = n} \sum_{k=1}^r \lambda_k(x') \langle v, \xi_k \rangle = -d_v(n) \leq -cn^{1+\eta} = -c\langle v, x \rangle^{1+\eta},$$

for  $n$  larger than some  $N \in \mathbb{N}$  and  $\eta > 0$ . It has been claimed that the weaker assumption corresponding to  $\eta \geq 0$  is sufficient [23, Proposition 4], however, this is not the case (the proof in [23] only implies exponential ergodicity). Indeed, the subcritical linear birth-death process is exponentially ergodic but not uniformly so [1].

**Corollary 5.2.** *Let  $\nu$  be the unique quasi-stationary distribution on  $E$  and  $\mathcal{I}_{\min}$  the unique minimal class of  $E$ . Then  $\text{supp } \nu = \mathcal{I}_{\min}$ .*

*Proof.* Suppose for contradiction that  $\text{supp } \nu \not\subseteq \mathcal{I}_{\min}$ . Then there exists a point  $y \in \mathcal{I} \neq \mathcal{I}_{\min}$  for which  $\nu(\{y\}) > 0$ , where  $\mathcal{I}$  is an irreducible class of  $E$ . As  $\nu$  is globally attracting in  $\mathcal{P}(E)$ , the space of probability distributions on  $E$ , it follows by definition that

$$\lim_{t \rightarrow \infty} \mathbb{P}_\mu(X_t \in B | t < \tau_A) = \nu(B),$$

for any  $\mu \in \mathcal{P}(E)$  and any measurable set  $B \subseteq E$ . In particular, letting  $\mu = \delta_z$  with  $z \in \mathcal{I}_{\min}$  yields, by minimality

$$0 = \lim_{t \rightarrow \infty} \mathbb{P}_{\delta_z}(X_t = y | t < \tau_A) = \nu(\{y\}) > 0,$$

which is a contradiction. We conclude that  $\text{supp } \nu \subseteq \mathcal{I}_{\min}$ . In particular, there exists  $x' \in \mathcal{I}_{\min}$  such that  $\nu(\{x'\}) > 0$ . Further, since  $\mathcal{I}_{\min}$  is irreducible,  $\mathbb{P}_x(X_t = y) > 0$  for all  $x, y \in \mathcal{I}_{\min}$ . As  $\nu$  is a QSD, it follows from [12, p. 48] that

$$e^{-\theta t} \nu(y) = \sum_{x \in E} \nu(x) \mathbb{P}_x(X_t = y) = \sum_{x \in \mathcal{I}_{\min}} \nu(x) \mathbb{P}_x(X_t = y) > \nu(x') \mathbb{P}_{x'}(X_t = y) > 0.$$

for some  $\theta > 0$  and all  $t \geq 0$ ,  $y \in \mathcal{I}_{\min}$ . Consequently,  $\text{supp } \nu \supseteq \mathcal{I}_{\min}$  which in turn implies  $\text{supp } \nu = \mathcal{I}_{\min}$  as desired.  $\square$

We remark that if  $A = \emptyset$  for an endorsed set  $E$ , then the definitions of quasi-stationary distribution and stationary distribution on  $E$  are equal,  $\tau_A = \infty$  and all the previous proofs go through without changes. In particular, it follows from Corollary 5.2 that any stationary distribution is supported by the unique minimal irreducible class (which exists by assumption) as well.

**Corollary 5.3.** *A one-species reaction network  $(\mathcal{N}, \lambda)$  has  $\gcd(\xi_1, \dots, \xi_r) < \infty$  endorsed sets with countable infinite number of states. If, in addition, the kinetics  $\lambda$  is stochastic mass-action, the highest order of a reaction  $R \geq 2$ , and*

$$\sum_{k \in \mathcal{R} : \max(y_{ki} | i=1, \dots, d) = R} \lambda_k(x) \xi_k < 0, \quad \text{for } x \text{ sufficiently large,}$$

*then each endorsed set has a unique QSD.*

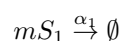
*Proof.* It follows similarly to the proof of Proposition 2.3 that for  $x$  large enough  $x$  and  $x + z$  are in the same endorsed set if and only if  $z = kg$ ,  $k \in \mathbb{Z}$ , is a multiple of  $g = \gcd(\xi_1, \dots, \xi_r)$ .

As there is only one species, we may without loss of generality take  $v = 1$ . It follows that  $d_v(n) = -\sum_{k \in \mathcal{R}} \lambda_k(n) \xi_k > 0$  for  $n$  sufficiently large and  $d_v(n) = \mathcal{O}(n^a)$  with  $a \geq 2$  by assumption. Furthermore,  $d^v(n) = 0$  for  $n$  sufficiently large. Thus Assumption 3.4 is satisfied. We infer the desired by Theorem 5.1.  $\square$

## 6 Examples

In this section, the main theorems and their applicability are illustrated through a series of examples. In particular, we show explicitly how the results of [9] are extended.

### Example 6.1.



As discussed in Section 2, the endorsed sets and corresponding absorbing sets are

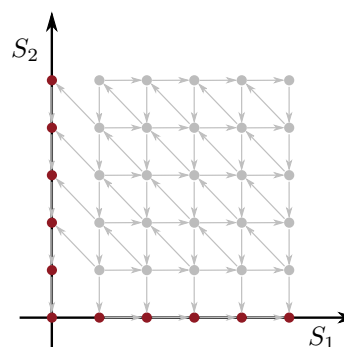
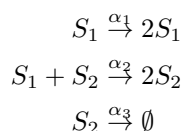
$$E_i = \{i + pm - 1 \mid p \in \mathbb{N}\}, \quad A_i = \{i - 1\}, \quad \text{for } i = 1, \dots, m,$$

respectively. These endorsed sets are evidently not irreducible. However,  $\{i + m - 1\}$  is the unique minimal irreducible class in  $E_i$  from which one may jump directly to  $A_i$ , thus Assumption 2.5 is satisfied. Further, assuming mass-action kinetics,

$$\begin{aligned} d_v(n) &= \alpha_1 n(n-1) \dots (n-m+1) m \mathbb{1}_E(n-m) = \mathcal{O}(n^m), \\ d^v(n) &= \mathcal{O}(1), \end{aligned}$$

for  $n$  sufficiently large, hence we conclude by Theorem 5.1 that there exists a unique QSD on each  $E_i$  if  $m \geq 2$ . Further, in this case, for any initial distribution,  $\mu$ , on the full set of endorsed states,  $D_E = \{m, m+1, \dots\}$ , the measure  $\mathbb{P}_\mu(X_t \in \cdot \mid t < \tau_A)$  tends to  $\nu_\mu$  exponentially fast for  $t \rightarrow \infty$ .

### Example 6.2 (Lotka-Volterra).



The Lotka-Volterra system describing competitive and predator-prey interactions has been of interest for approximately a century [29, 51]. In the stochastic description of the model, we find  $R = \mathbb{N}^2$  hence it follows that the state space can be divided into the endorsed and absorbing sets given by

$$D_E = E = \mathbb{N}^2, \quad D_A = A = \mathbb{N}_0^2 \setminus \mathbb{N}^2,$$

respectively. Using mass-action kinetics, it follows that for  $v = (v_1, v_2) \in \mathbb{N}^2$ ,

$$\begin{aligned} d_v(n) &= - \max_{x \in E, \langle v, x \rangle = n} \sum_{k=1}^r \lambda_k(x) \langle v, \xi_k \rangle \mathbb{1}_E(x + \xi_k) \\ &= - \max_{x \in E, \langle v, x \rangle = n} (\alpha_1 v_1 x_1 + (v_2 - v_1) \alpha_2 x_1 x_2 \mathbb{1}_{\{2,3,\dots\}}(x_1) - v_2 \alpha_3 x_2 \mathbb{1}_{\{2,3,\dots\}}(x_2)). \end{aligned}$$

Letting  $(v_1, v_2) = (\ell, 1)$  yields  $d_v(n) = -\ell \alpha_1(n-1) + (\ell-1) \alpha_2(n-1)$ . Thus, choosing  $\ell$  sufficiently large,  $d_v(n) = \mathcal{O}(n)$  and  $d_v(n) > 0$  for  $n$  sufficiently large, provided that  $\alpha_2 > \alpha_1$ . Note also that  $\max_{k \in \mathcal{R}_A} \langle v, \xi_k \rangle < 0$  hence by Proposition 4.2 the process is  $\mathbb{P}_x$ -a.s. absorbed for all  $x \in E$  if  $\alpha_2 > \alpha_1$ . However,  $d^v(n) = \mathcal{O}(n^2)$ , hence Assumption 3.4(i) is not satisfied, and one can not apply Theorem 5.1.

Using generalized mass-action [34] for the same standard Lotka-Volterra network, we may obtain a different result. Suppose for example that

$$\lambda_1(x) = \alpha_1 x_1 x_2, \quad \lambda_2(x) = \alpha_2 x_1^4 x_2^2, \quad \lambda_3(x) = \alpha_3 x_2^3.$$

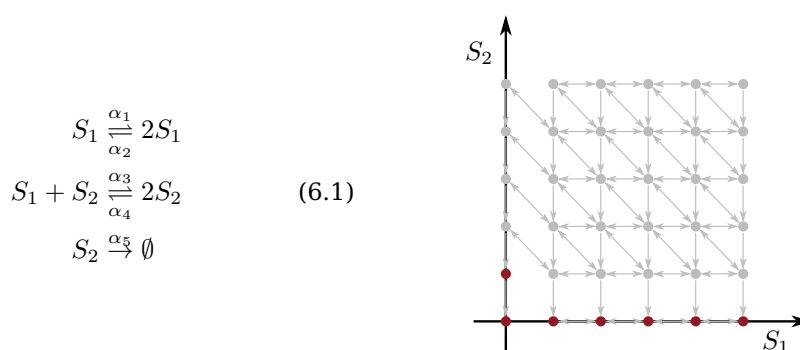
Choosing  $v = (2, 1)$ , say, we find

$$d_v(n) = - \max_{x \in E, \langle v, x \rangle = n} (2\alpha_1 x_1 x_2 - \alpha_2 x_1^4 x_2^2 \mathbb{1}_{\{2,3,\dots\}}(x_2) - \alpha_3 x_2^3 \mathbb{1}_{\{2,3,\dots\}}(x_2)) = \mathcal{O}(n^3),$$

and likewise

$$d^v(n) = \max_{x \in E, \langle v, x \rangle = n} n (\alpha_2 x_1^4 x_2^2 \mathbb{1}_{\{1\}}(x_1) + \alpha_3 x_2^3 \mathbb{1}_{\{1\}}(x_2)) = \mathcal{O}(n^3).$$

Thus Assumption 3.4 is satisfied. As  $E$  is irreducible, Assumption 2.5 is also satisfied and we conclude by Theorem 5.1 that there is a unique QSD on  $E$ .



Let us now consider the slightly altered version of the original Lotka-Volterra system using mass-action kinetics, obtained by addition of the reactions  $2S_1 \rightarrow S_1$  and  $2S_2 \rightarrow S_1 + S_2$ . In this case, there is still one endorsed set with corresponding absorbing set given by

$$E = \mathbb{N}_0 \times \mathbb{N} \setminus \{(0, 1)\}, \quad A = \mathbb{N}_0^2 \setminus E,$$

respectively.

For a general  $v = (v_1, v_2) \in \mathbb{N}^2$  it follows that  $d^v(n) = \mathcal{O}(n)$  and

$$\begin{aligned} d_v(n) &= - \max_{x \in E, \langle v, x \rangle = n} \sum_{k=1}^r \lambda_k(x) \langle v, \xi_k \rangle \mathbb{1}_E(x + \xi_k) \\ &= \min_{x \in E, \langle v, x \rangle = n} (-\alpha_1 x_1 v_1 + v_1 \alpha_2 x_1 (x_1 - 1) - (v_2 - v_1) \alpha_3 x_1 x_2 \\ &\quad + (v_2 - v_1) \alpha_4 x_2 (x_2 - 1) + v_2 \alpha_5 x_2 \mathbb{1}_{\{2,3,\dots\}}(x_2)). \end{aligned}$$

We need  $v_2 > v_1$  for the coefficient of the 4th reaction to be positive. Further, by the second derivative test,  $d_v(n) = \mathcal{O}(n^2)$  exactly when

$$4v_1 \alpha_2 (v_2 - v_1) \alpha_4 > (v_2 - v_1)^2 \alpha_3^2.$$

The set of possible  $v$ -vectors,  $\mathcal{V}$ , is therefore

$$\mathcal{V} = \left\{ v \in \mathbb{N}^2 \mid v_1 < v_2 < v_1 \left( 1 + \frac{4\alpha_2 \alpha_4}{\alpha_3^2} \right) \right\},$$

which is non-empty for any positive reaction rates. A particular choice would be  $v = (\ell, \ell + 1)$ , for  $\ell$  sufficiently large. As  $E$  is irreducible, Assumption 2.5 is satisfied and we conclude by Theorem 5.1 that the modified Lotka-Volterra system has a unique QSD on  $E$  for any reaction rates.

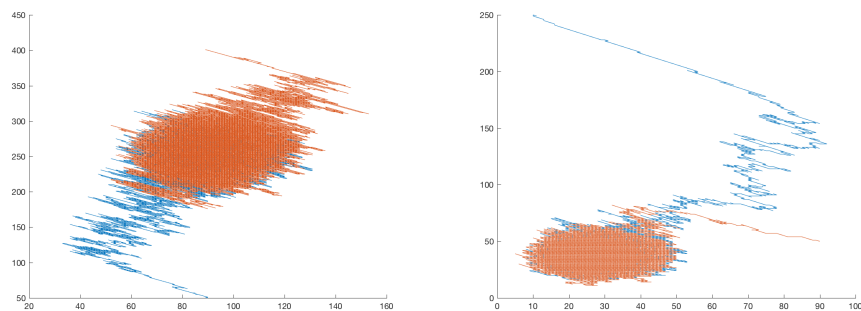
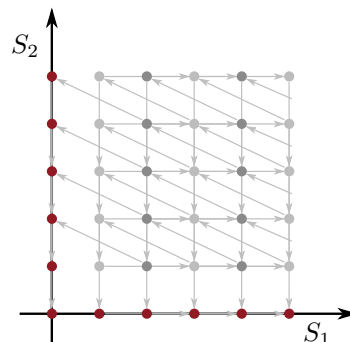
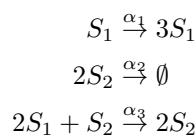


Figure 3: Two realizations of the process  $(X_t : t \geq 0)$ , associated with the reaction network (6.1), with differing  $X_0$ . Left:  $\alpha_1 = 30, \alpha_2 = 0.3, \alpha_3 = 2, \alpha_4 = 0.7, \alpha_5 = 1$ . Right:  $\alpha_1 = 20, \alpha_2 = 0.7, \alpha_3 = 1, \alpha_4 = 0.7, \alpha_5 = 1$ .

### Example 6.3.



There are two endorsed sets given by

$$\begin{aligned} E_1 &= \{x \in \mathbb{N}^2 : x_1 \equiv 1 \pmod{2}\}, \\ E_2 &= \{x \in \mathbb{N}^2 : x_1 \equiv 0 \pmod{2}\}. \end{aligned}$$

The corresponding set of absorbing sets are given by

$$\begin{aligned} A_1 &= \{x \in \mathbb{N}_0^d : x_2 = 0, x_1 \equiv 1 \pmod{2}\} \\ A_2 &= \{x \in \mathbb{N}_0^d : x_2 = 0, x_1 \equiv 0 \pmod{2}\} \cup \{x \in \mathbb{N}_0^d : x_1 = 0\}. \end{aligned}$$

Assuming mass-action kinetics, it follows that for a general  $v = (v_1, v_2) \in \mathbb{N}^2$  we obtain for  $E_1$  and  $E_2$  respectively,

$$\begin{aligned} d_v(n) &= - \max_{x \in E, \langle v, x \rangle = n} (2v_1\alpha_1x_1 - 2v_2\alpha_2x_2(x_2 - 1)\mathbb{1}_{\{3,4,\dots\}}(x_2) \\ &\quad + (-2v_1 + v_2)\alpha_3x_1(x_1 - 1)x_2), \\ d_v(n) &= - \max_{x \in E, \langle v, x \rangle = n} (2v_1\alpha_1x_1 - 2v_2\alpha_2x_2(x_2 - 1)\mathbb{1}_{\{3,4,\dots\}}(x_2) \\ &\quad + (-2v_1 + v_2)\alpha_3x_1(x_1 - 1)x_2\mathbb{1}_{\{3,4,\dots\}}(x_1)), \end{aligned}$$

which are both  $\mathcal{O}(n^2)$  exactly if  $v_2 < 2v_1$ . A particular choice would be  $v = (1, 1)$ . Further,

$$d^v(n) = \max_{x \in E, \langle v, x \rangle = n} n(\alpha_2x_2(x_2 - 1)\mathbb{1}_{\{2\}}(x_2) + \alpha_3x_1(x_1 - 1)x_2\mathbb{1}_{\{2\}}(x_1)) = 2\alpha_3n^2 = \mathcal{O}(n^2).$$

We conclude that there exists an  $0 < \eta < 1$  such that Assumption 3.4 holds. Since both  $E_1$  and  $E_2$  are irreducible, Assumption 2.5 is satisfied hence Theorem 5.1 applies regardless of the rate constants – there exists a unique QSD,  $\nu_n$ , on each  $E_n$ .

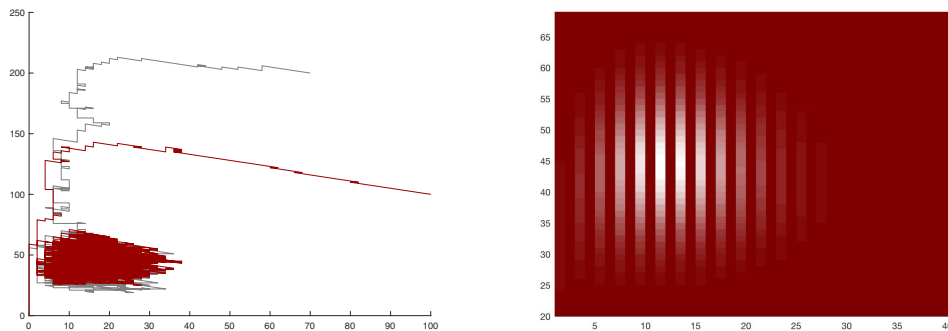


Figure 4: Left: Two realizations of  $(X_t : t \geq 0)$  with  $\alpha_1 = 300, \alpha_2 = 1, \alpha_3 = 0.5$ .  $X_0 = (70, 200)$  in grey and  $X_0 = (100, 100)$  in red. Right: Approximate density of the QSD on one of the two endorsed sets.

Note that if we had modeled the network using deterministic mass-action [22], we would have found the attracting<sup>2</sup> fixed point

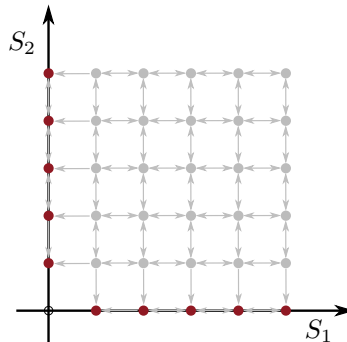
$$(x, y) = \left( \frac{(2\alpha_1\alpha_2)^{1/3}}{\alpha_3^{2/3}}, \frac{\alpha_1^{2/3}}{(2\alpha_2\alpha_3)^{1/3}} \right),$$

<sup>2</sup>One can calculate the trace of the Jacobin matrix to be  $-2\alpha_1 - (2\alpha_1\alpha_2)^{2/3}\alpha_3^{-1/3} < 0$ , making the fixed point attracting for all parameter values.



which seems to lie near the peak of the quasi-stationary distribution for the parameter values used in Figure 4. Indeed, we find the fixed point (13.39, 44.81) which is a stable spiral.

**Example 6.4** (Birth-death).



The full set of endorsed states is found to be  $D_E = \mathbb{N}^2$  which is irreducible, and the full set of absorbing states is  $D_A = \mathbb{N}_0^2 \setminus \mathbb{N}^2$ . Thus Assumption 2.5 is satisfied. For a general  $v \in \mathbb{N}^2$  we have, with  $\alpha_5 - \alpha_6 = \rho_1$ ,  $\alpha_7 - \alpha_8 = \rho_2$ ,  $\alpha_1 - \alpha_2 = \rho_3$  and  $\alpha_3 - \alpha_4 = \rho_4$ ,

$$d_v(n) = \min_{x \in E_v(v, x) = n} (v_1 \rho_3 + v_2 \rho_4) x_1 x_2 - v_1 \rho_1 x_1 (x_1 - 1) - v_2 \rho_2 x_2 (x_2 - 1).$$

The second derivative test gives the sufficient and necessary criteria for  $d_v(n)$  to be  $\mathcal{O}(n^2)$ ,

$$4v_1 v_2 \rho_1 \rho_2 - (v_1 \rho_3 + v_2 \rho_4)^2 > 0.$$

Solving for  $v_2$  we find the set of possible  $v$ -vectors,

$$\mathcal{V} = \left\{ v \in \mathbb{N}^2 \mid \frac{2\rho_1 \rho_2 - \rho_3 \rho_4 - 2\sqrt{\rho_1^2 \rho_2^2 - \rho_1 \rho_2 \rho_3 \rho_4}}{\rho_4^2} v_1 < v_2 < \frac{2\rho_1 \rho_2 - \rho_3 \rho_4 + 2\sqrt{\rho_1^2 \rho_2^2 - \rho_1 \rho_2 \rho_3 \rho_4}}{\rho_4^2} v_1 \right\},$$

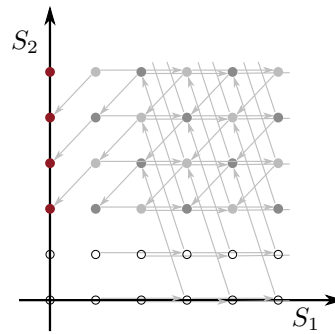
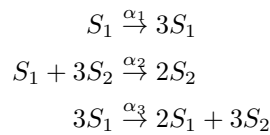
which is non-empty exactly when  $\rho_1 \rho_2 > \rho_3 \rho_4$ . Choosing parameter values as, say,



we find

$$\mathcal{V} \approx \{v \in \mathbb{N}^2 \mid 1.373v_1 < v_2 < 46.627v_1\}.$$

Note that  $v = (1, 1) \notin \mathcal{V}$ , and Theorem 5.1 would not be applicable with this choice. We have therefore extended the result of [9].

**Example 6.5.**

The stoichiometric matrix is in this example given by

$$\Xi = \begin{pmatrix} 2 & -1 & -1 \\ 0 & -1 & 3 \end{pmatrix}.$$

It follows that  $\text{rank } \Xi = 2 = d$  hence by Proposition 2.3, there is a finite number endorsed sets for  $x$  sufficiently large. Indeed, upon inspection, we find two endorsed sets given by

$$\begin{aligned}
 E_1 &= \{x \in \mathbb{N}^2 : x_2 \geq 0, x_1 + x_2 = 0 \pmod{2}\}, \\
 E_2 &= \{x \in \mathbb{N}^2 : x_2 \geq 0, x_1 + x_2 = 1 \pmod{2}\},
 \end{aligned}$$

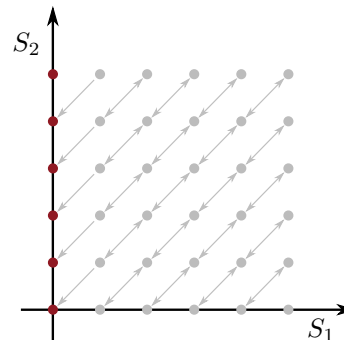
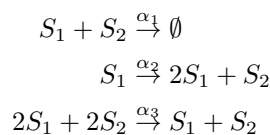
and these each have a unique minimal irreducible class hence Assumption 2.5 is satisfied. Note, however, that the sets  $E_1$  and  $E_2$  are not irreducible. Here  $D_E = \{x \in \mathbb{N}_0^2 : x_1 \geq 1\}$  and  $D_A = \{x \in \mathbb{N}_0^2 : x_1 = 0\}$ . Taking  $v = (4, 1)$ , it follows that for both classes

$$d_v(n) = - \max_{x \in E, \langle v, x \rangle = n} (8\alpha_1 x_1 - 5\alpha_2 x_1 x_2 (x_2 - 1)(x_2 - 2) \mathbb{1}_{\{2,3,\dots\}}(x_1) - \alpha_3 x_1 (x_1 - 1)(x_1 - 2)),$$

which is  $\mathcal{O}(1)$ . Furthermore,

$$d^v(n) = \max_{x \in E, \langle v, x \rangle = n} n x_1 x_2 (x_2 - 1)(x_2 - 2) \mathbb{1}_{\{1\}}(x_1) = \mathcal{O}(n^4).$$

Thus, Assumption 3.4 is not satisfied and we can not apply Theorem 5.1.

**Example 6.6.**

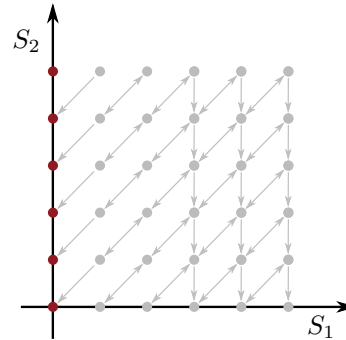
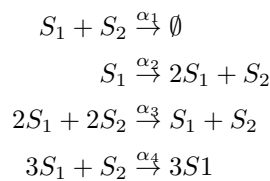
As  $\text{rank } \Xi = 1 < d$ , there are infinitely many endorsed sets. These are each irreducible. Indeed, upon inspection, we find the endorsed sets

$$E_i = \{x \in \mathbb{N}_0^2 : x_1 \geq 1, x_2 = x_1 + (-1)^i \lfloor i/2 \rfloor\}, \quad i \geq 1,$$

and the full set of absorbing states  $D_A = \{x \in \mathbb{N}_0^2 : x_1 = 0\}$ . Furthermore, some sets ( $E_i$  with  $i \equiv 1 \pmod{2}$ ) have no corresponding absorbing set while all others do. For all classes however, we find

$$d_v(n) = \mathcal{O}(n^2), \quad d^v(n) = \mathcal{O}(n),$$

hence, by Theorem 5.1, we conclude that for any initial distribution  $\mu$  with support contained in a finite subset of the endorsed sets, the measure  $\mathbb{P}_\mu(X_t \in \cdot | t < \tau_A)$  converges exponentially fast in  $t$  to  $\nu_\mu$ . Had we instead looked at the slightly altered network



then Assumption 2.5 would no longer be satisfied as no minimal irreducible class exists.

## 7 Conclusions

In this paper, we have provided sufficient conditions for the existence and uniqueness of a quasi-stationary distribution, within each endorsed set, for general stochastic reaction networks. In particular, we have provided sufficient conditions for the existence and uniqueness of a globally attracting quasi-stationary distribution in the space of probability distributions.

The requirement that for mass-action reaction systems there exists a stoichiometric coefficient strictly greater than one for each species is strong, however, it seems to be intrinsic to the problem of guaranteeing uniqueness of the QSD. Indeed, this can be seen already in 1-dimensional systems, and one can imagine that getting stuck near one of the axes would approximately reduce the multi-dimensional system to such a 1-dimensional case. It is an important question to determine sufficient conditions for just the existence of a QSD, in which case one would expect much weaker conditions to be satisfied. This is still an open problem.

The application of our results depends on the existence of a vector  $v \in \mathbb{N}^d$ , such that Assumption 3.4 holds. One would like to have easy graphical ways of guaranteeing this existence or an explicit algorithmic way of constructing such a vector. This may not be possible in general, since the problem is equivalent to determining the sign of a multivariate polynomial in a certain region of the positive orthant. However, the specific network at hand is often prone to analytical ad hoc methods, and even in lack of this, one may rely on numerical methods to find a suitable candidate for  $v$ . Another caveat is the exact calculation of the endorsed sets. Methods for finding these numerically, in the case where the intensity functions are positive in the positive orthant, exist [24, 38]. However, the problem becomes more complicated in the general case [23].

Knowing the existence of a unique QSD, one would of course like to know the explicit analytical expression for this distribution on each endorsed set. However, this seems to be a very hard problem to solve in general, and even for 1-dimensional systems it has not yet been fully resolved. Thus, so far, one is forced to apply numerical methods or rely on analytical approximations, see [21, 50].

Another problem which is important for applications, is the question of observability. Indeed, the relative sizes of the time to extinction,  $\tau_A$ , versus the time to reach the mean of the QSD,  $\tau_\nu$ , determines whether we are likely to observe the QSD or not. Only if

$\tau_A \gg \tau_\nu$  would we expect the process to behave according to the QSD [47]. Few recent results on this matter exist for limited scenarios [23], while methods exploiting the WKB approximation [6] are more general although not yet fully rigorous [10].

Finally, numerical evidence seems to suggest a strong connection between the deterministic and stochastic models of the same underlying reaction network. Indeed, for systems close to thermodynamic equilibrium, also referred to as the fluid limit [28], the modes of the QSD appear to be located near the deterministic steady states. However, far from equilibrium, the picture may be radically different. Our result can be seen as a stochastic analogue to the deterministic case of having an equilibrium point within each stoichiometric compatibility class. In this light, the QSD bridges the gap between the knowledge of extinction in the stochastic description and the existence of a stationary steady state in the deterministic setting. Future work lies in analyzing what can be inferred about the QSD from the corresponding deterministic dynamical system.

## References

- [1] W. J. Anderson, *Continuous-Time Markov Chains: An Applications-Oriented Approach*, Probability and Its Applications, Springer (1991). MR-1118840
- [2] D. F. Anderson, *Global Asymptotic Stability for a Class of Nonlinear Chemical Reactions*, SIAM J. Appl. Math., 68 (2008), pp. 1464–1476. MR-2407133
- [3] D. F. Anderson and T. G. Kurtz, *Stochastic Analysis of Biochemical Systems*, Mathematical Biosciences Institute Lecture Series, Springer (2015). MR-3363610
- [4] J. C. Anderson, E. J. Clarke, A. P. Arkin, and C. A. Voigt, *Environmentally Controlled Invasion of Cancer Cells by Engineered Bacteria*, J. of Molecular Biology, 355(4) (2006), pp. 619–627.
- [5] A. Arkin, J. Ross, and H. H. McAdams, *Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage  $\lambda$ -infected Escherichia Coli Cells*, Genetics, 149 (1998), pp. 1633–1648.
- [6] M. Assaf and B. Meerson, *WKB Theory of Large Deviations in Stochastic Populations*, J. Phys. A: Math. Theor., 50 (2017). MR-3664339
- [7] N. Barkai and S. Leibler, *Biological Rhythms: Circadian Clocks Limited by Noise*, Nature, 403 (2000), p. 267.
- [8] N. Champagnat and D. Villemonais, *Exponential Convergence to Quasi-stationary Distribution and Q-process*, Probability Theory and Related Fields, 164 (2016), pp. 243–286. MR-3449390
- [9] N. Champagnat and D. Villemonais, *Lyapunov Criteria for Uniform Convergence of Conditional Distributions of Absorbed Markov Processes*, arXiv:1704.01928 (2017).
- [10] J.-R. Chazottes, P. Collet and S. Méléard, *On Time Scales and Quasi-stationary Distributions for Multitype Birth-and-Death Processes*, arXiv:1702.05369 (2017). MR-4029154
- [11] G. Cohen, T. Mora, and O. Moreno, eds., *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, vol. 673 in Lecture Notes in Computer Science, Springer (1993). MR-1251964
- [12] P. Collet, S. Martinez, and J. S. Martin, *Quasi-stationary Distributions. Markov Chains, Diffusions and Dynamical Systems*, Probability and Its Applications, Springer (2013). MR-2986807
- [13] M. Cook, D. Soloveichik, E. Winfree, and J. Bruck, *Programmability of Chemical Reaction Networks, Chapter in Algorithmic Bioprocesses*, Springer, 2009. MR-2762062
- [14] D. Dadush, *Lattices, Convexity & Algorithms*, Lecture Notes, NYU (2013). MR-3103593
- [15] P. Dittrich and L. Winter, *Chemical Organizations in a Toy Model of the Political System*, Adv. Complex Syst., 11 (2008), pp. 609–627.
- [16] R. Durrett, *Probability: Theory and Examples*, Duxbury Press, 2nd ed. (1996). MR-1609153
- [17] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Stochastic Gene Expression in a Single Cell*, Science, 297 (2002), p. 1183.
- [18] S. N. Ethier and T. G. Kurtz, *Markov Processes*, Wiley Series in Probability and Mathematical Statistics, Wiley (1986). MR-0838085

- [19] M. Feinberg, *Lectures on Chemical Reaction Networks. Delivered at the Mathematics Research Center, Univ. Wisc.-Madison.* (1979).
- [20] A. Griffin, *Quasi-Stationary Distributions for Evolving Epidemic Models: Simulation and Characterisation*, PhD Thesis, Mathematics Institute, The University of Warwick (2016).
- [21] P. Groisman and M. Jonckheere, *Simulation of Quasi-stationary Distributions on Countable Spaces*, arXiv:1206.6712 (2012). MR-3156964
- [22] J. Gunawardena, *Chemical Reaction Network Theory for In-Silico Biologists* (2003).
- [23] A. Gupta, C. Briat, and M. Khammash, *A Scalable Computational Framework for Establishing Long-term Behavior of Stochastic Reaction Networks*, PLOS Computational Biology, 10 (2014).
- [24] A. Gupta and M. Khammash, *Determining the Long-Term Behavior of Cell Populations: A New Procedure for Detecting Ergodicity in Large Stochastic Reaction Networks*, arXiv:1312.2879 (2013).
- [25] F. J. M. Horn and R. Jackson, *General Mass Action Kinetics*, Arch. Ration. Mech. Anal., 47 (1972), pp. 81–116. MR-0400923
- [26] M. D. Johnston, D. F. Anderson, G. Craciun, and R. Brijder, *Conditions for Extinction Events in Chemical Reaction Networks with Discrete State Spaces*, Journal of Mathematical Biology, (2017), pp. 1–24. MR-3771429
- [27] O. Kallenberg, *Foundations of Modern Probability*, Probability and Its Applications, Springer, 2nd ed. (2001). MR-1876169
- [28] T. G. Kurtz, *Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes*, J. Appl. Prob., 7 (1970), pp. 49–58. MR-0254917
- [29] A. J. Lotka, *Elements of Physical Biology*, Williams and Witkins, Baltimore (1925).
- [30] S. Méléard and D. Villemonais, *Quasi-stationary Distributions and Population Processes*, Probability Surveys, 9 (2012), pp. 340–410. MR-2994898
- [31] S. P. Meyn and R. L. Tweedie, *Stability of Markovian Processes III: Foster-Lyapunov Criteria for Continuous-Time Processes*, Advances in Applied Probability, 25 (1993), pp. 518–548. MR-1234295
- [32] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Cambridge University Press, 2nd ed. (2009). MR-2509253
- [33] D. S. Mitrinovic, J. Pecaric, and A. M. Fink, *Inequalities Involving Functions and Their Integrals and Derivatives*, Mathematics and Its Applications (East European Series), Springer (1991). MR-1190927
- [34] S. Müller and G. Regensburger, *[9] Generalized Mass Action Systems: Complex Balancing Equilibria and Sign Vectors of the Stoichiometric and Kinetic-Order Subspaces*, SIAM J. Appl. Math., 72 (2012), pp. 1926–1947. MR-3024168
- [35] B. Munsky, B. Trinh, and M. Khammash, *Listening to the Noise: Random Fluctuations Reveal Gene Network Parameters*, Molecular Systems Biology, 5 (2009).
- [36] I. Nåsell, *Extinction and Quasi-stationarity in the Stochastic Logistic SIS Model*, Lecture Notes in Mathematics, Springer (2011). MR-2808078
- [37] J. R. Norris, *Markov Chains*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press (2009). MR-1600720
- [38] L. Paulevé, G. Craciun, and H. Koepl, *Dynamical Properties of Discrete Reaction Networks*, J. Math. Biol., 69 (2014), pp. 55–72. MR-3215073
- [39] D. Pollard, *Total Variation Distance between Measures* (2005).
- [40] P. K. Pollett, *Quasi-stationary Distributions: A Bibliography* (2015).
- [41] H. Qian, *Nonlinear Stochastic Dynamics of Mesoscopic Homogeneous Biochemical Reaction Systems – An Analytical Theory*, Nonlinearity, 24 (2011). MR-2793894
- [42] L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes and Martingales*, vol. 1 of Cambridge Mathematical Library, Cambridge University Press (2000). MR-1780932
- [43] M. Shakil, H. A. Wahab, M. Naeem, S. Bhatti, and M. Shahzad, *The Modeling of Predator-Prey Interactions*, Network Biology, 5(2) (2015), pp. 71–81.

- [44] E. A. van Doorn, *Quasi-stationary Distributions and Convergence to Quasi-stationarity of Birth-Death Processes*, Advances in Applied Probability, 23 (1991), pp. 683–700. MR-1133722
- [45] E. A. van Doorn and P. K. Pollett, *Quasi-stationary Distributions for Reducible Absorbing Markov Chains in Discrete Time*, Markov Processes and Relat. Fields, 15 (2009), pp. 191–204. MR-2538313
- [46] E. A. van Doorn and P. K. Pollett, *Quasi-stationary Distributions for Discrete-State Models*, European Journal of Operational Research, 230 (2013), pp. 1–14. MR-3063313
- [47] M. Vellela and H. Qian, *A Quasistationary Analysis of a Stochastic Chemical Reaction: Keizer’s Paradox*, Bulletin of Mathematical Biology, 69 (2007), pp. 1727–1746. MR-2326546
- [48] T. Veloz and P. Razeto-Barry, *Reaction Networks as a Language for Systemic Modeling: Fundamentals and Examples*, Systems, 5 (2017).
- [49] T. Veloz, P. Razeto-Barry, P. Dittrich, and A. Fajardo, *Reaction Networks and Evolutionary Game Theory*, Journal of Mathematical Biology, 68 (2014), pp. 181–206. MR-3147459
- [50] D. Villemonais, *Minimal Quasi-stationary Distributions*, arXiv:1404.6648 (2015).
- [51] V. Volterra, *Fluctuations in the Abundance of a Species Considered Mathematically*, Nature, 118 (1926), pp. 558–560.

**Acknowledgments.** The authors would like to thank Abhishek Pal Majumder for many valuable and interesting discussions on this topic. We are grateful for the reviewer’s careful reading of the manuscript and her/his valuable corrections and suggestions.